International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Performance Analysis of a Time-Sharing Queue with Server Breakdown, Discouraging Arrivals and Reneging

Priyanka, Poonam Singh & Swadesh Singh

Department of Mathematics, Bareilly College, M. J. P. Rohilkhand University, Bareilly 243006, India.
priyanka29rg@gmail.com, dr.poonamsingh.bly@gmail.com & ssinghbcb@gmail.com

Abstract

This paper investigates a single-server time-sharing queueing system with server breakdowns, discouraging arrivals and customer reneging. Customers arrive according to a Poisson process with rate λ while the server is operational. During server breakdowns, the arrival rate is reduced to represent customer discouragement. The server operates under the time-sharing discipline, where all customers in the system receive an equal share of the available service capacity simultaneously. Service times are assumed to be exponentially distributed. The server is subject to random breakdowns and is restored after repair, while customers may abandon the system due to impatience through reneging. The proposed model is analysed using the Continuous-Time Markov Chain (CTMC) approach and the corresponding steady-state balance equations are formulated to obtain the steady-state probability distribution. Several system performance measures, including the expected number of customers in the system, average waiting time, server availability, throughput and reneging probability are derived. Numerical experiments are carried out to examine the effects of arrival rate, breakdown rate, repair rate and reneging rate on system performance. The proposed model provides a useful analytical framework for evaluating time-sharing service systems operating under server interruptions and impatient customer behavior with potential applications in computer systems, communication networks, cloud computing and other shared-resource environments.

Keywords: Queueing theory, Time-sharing system, Continuous-Time Markov Chain, Server breakdown, Discouraging arrivals, Reneging.

1. Introduction

Queueing theory is one of the most important branches of operations research that deals with the mathematical analysis of congestion and waiting-line phenomena in service systems. Since its pioneering work by Erlang, queueing theory has become an indispensable analytical tool for evaluating the performance of systems where customers complete for limited-service resources. Queueing models have found extensive applications in manufacturing systems, computer and communication networks, cloud computing, healthcare services, transportation, banking and various service industries. Performance measures such as the average queue length, waiting time, server utilization, throughput, and system availability play a significant role in designing efficient and reliable service systems (Kleinrock, 1975; Gross et al., 2008; Bhat, 2015).

Among various queueing disciplines, the time-sharing discipline has gained considerable attention because of its fairness and efficient utilization of service resources. Unlike the traditional First-Come-First-Served (FCFS) discipline, the time-sharing discipline allows all customers present in the system to receive service simultaneously by equally sharing the available processing capacity. Consequently, no customer monopolizes the server and each customer receives an equal fraction of the service capacity until completion. This service discipline is particularly suitable for interactive computer systems, operating systems, cloud computing platforms, communication networks and web servers where multiple jobs are processed concurrently (Kleinrock, 1976; Yashkov, 1987).

In practical service environments, server failures are unavoidable due to hardware malfunctions, software errors, communication failures, power interruptions, or routine maintenance. Such failures temporarily interrupt the service process, resulting in increased congestion and reduced system performance. To represent these real-life situations more accurately, server breakdown and repair mechanisms have been incorporated into many queueing models. The inclusion of unreliable servers enables analysts to evaluate system reliability, availability and operational efficiency under service interruptions.

Besides server reliability, customer behavior significantly influences the performance of modern service systems. In many practical situations, customers become discouraging from entering the system when they observe that the server is unavailable or the expected waiting time is excessive. This phenomenon known as discouraging arrivals, reduces the effective arrival rate during periods of server failure. Furthermore, customers already present in the system may lose patience and leave before service completion, a behavior referred to as reneging. These two behavioural characteristics directly affect congestion, throughput and customer satisfaction and therefore should be incorporated into realistic queueing models.

A considerable amount of research has been devoted to queueing systems involving unreliable servers, customer impatience, discouraging arrivals, vacations, retrials and different service disciplines. Although these characteristics have been extensively studied individually and in various combinations, relatively limited attention has been given to time-sharing queueing systems that simultaneously incorporate server breakdowns, discouraging arrivals and reneging. Since time-sharing is widely employed in modern computing and communication environments, the development of such an integrated analytical model is both practically relevant and theoretically important.

The present study considers a single-server time-sharing queueing system in which customers arrive according to a Poisson process and service times are exponentially distributed. The server is subject to random breakdowns and subsequent repairs. During server breakdowns, customer arrivals are discouraging, leading to a reduced arrival rate, while customers may abandon the system because of impatience through reneging. The stochastic behavior of the proposed system is analysed using the Continuous-Time Markov Chain (CTMC) approach. Steady-state balance equations are established to obtain the steady-state probability distribution and important performance measures of the system.

The primary objective of this paper is to investigate the combined effects of time-sharing, server breakdowns, discouraging arrivals and reneging on overall system performance. Performance measures such as the expected number of customers in the system, average waiting time, throughput, server availability and reneging probability are derived. Numerical experiments are presented to illustrate the influence of key system parameters on these performance measures. The proposed analytical framework provides useful insights into the design and performance evaluation of time-sharing service systems operating under unreliable conditions and impatient customer behavior.

2. Literature Review: -

Cooper (1981) suggested the basic concepts of Queueing Theory. It provides a detailed description of the Birth–Death Process and the Markov Process. It presents the mathematical derivations for classical queueing models (such as M/M/1, M/G/1, M/M/s, etc.). It describes the calculation of performance measures such as L, L_q, W, W_q and server utilization. It gives you the mathematical background and analytical techniques you need for your time-sharing queueing model.

Sennott (1998) proposed Stochastic Dynamic Programming and the Control of Queueing Systems is a book that provides a mathematical basis for analysing and controlling queueing systems when things are not certain. The book teaches us about Markov Decision Processes. Shows how to use dynamic programming to find the best ways to operate complex queueing systems. It's about things like who gets in, how fast service is, where to send people, how to use resources, how to schedule things, and how to make the system work better when things are uncertain. The person who wrote the book wants to help us minimize costs in the run while still giving good service and reducing wait times for customers. The book also gives us models, algorithms and techniques to solve problems in complex queueing systems.

Haviv (2013) proposed book explains the optimization of queuing systems in a highly systematic manner. It primarily covers service policy optimization, congestion reduction, resource allocation, cost optimization and system performance. It has been instrumental in helping us understand the optimization conditions for queuing models. Additionally, concepts such as priority queuing and processor sharing are explained clearly. The book also addresses the management of telecommunication networks and details various mathematical techniques used to determine performance measures in queuing systems. It is a highly useful resource for engineering, technology and management programs.

Ross (2014) evaluated Queueing theory is based entirely on probability and stochastic models. This book provides a thorough explanation of these concepts, detailing the underlying probability theory and the various stages of queueing systems such as arrival patterns, service patterns, and output. It also offers clear explanations of finite capacity models and steady-state analysis.

Bhat(2015) proposed a mathematical basis of stochastic queueing models.

Describes Markovian modelling, the basis to obtain balance equations, and Birth Death processes. Discusses performance metrics needed to evaluate your proposed model.

Analytic and matrix solution techniques are introduced and extendable to time sharing systems. It illustrates the application of queueing models to computer networks, cloud computing, communication systems and service systems, which are popular application areas for time-sharing queueing models. It is reference for the theoretical development of your paper to establish, analyse, optimize and evaluate the performance of the support model.

Jain (2017) analyzed a priority queue with batch arrival, balking, threshold recovery, unreliable server, and optimal service policy. The system incorporates two types of customers: priority and ordinary, with different service requirements. The study used the matrix-geometric method to obtain steady-state probabilities and performance indices. Results showed that threshold-based recovery and priority discipline significantly improve system reliability and efficiency. However, time-sharing and time-sharing disciplines were not considered in this model.

Zhu and Casale (2020) proposed a fluid approximation for closed queueing networks with Discriminatory Processor Sharing (DPS). The model incorporated phase-type service distributions and class switching, enabling both transient and steady-state performance analysis. Numerical experiments confirmed the accuracy of the proposed approximation. However, the study assumed reliable servers and did not include

vacation policies or adaptive service-rate control.

Gupta and Zhang (2021) investigated state-dependent Limited Processor Sharing (LPS) queues and proposed optimal control policies based on heavy-traffic diffusion approximations. The study developed efficient algorithms for selecting concurrency limits and demonstrated significant improvements in response time and server efficiency. The research represents an important step toward adaptive time-sharing systems, although server vacations and customer impatience were not incorporated.

Kumar and Jain (2023) developed a cost optimization model for an unreliable server queue with a two-stage service process under a hybrid vacation policy. The study integrates working vacation and complete vacation strategies in a single framework. The authors used matrix-geometric methods along with metaheuristic optimization techniques to determine optimal service rates and minimize system cost. The results showed that hybrid vacation policies significantly improve cost efficiency and service performance in unreliable service environments. However, adaptive time-sharing and time-sharing disciplines were not considered in this model.

Main Contributions

The major contributions of this paper are summarized as follows:

1. A single-server time-sharing queueing model with server breakdown, repair, discouraging arrivals and reneging is developed.
2. The stochastic model is formulated using the Continuous-Time Markov Chain (CTMC) framework.
3. Steady-state balance equations are derived for the proposed system.
4. Closed-form expressions for key performance measures are obtained.
5. Numerical investigations demonstrate the effects of arrival, breakdown, repair and reneging rates on system performance.

Consider a single-server time-sharing queueing system in which customers arrive according to a Poisson process with arrival rate λ. The system consists of one unreliable server that provides service under the time-sharing discipline. Under this discipline, all customers present in the system simultaneously share the available service capacity equally. If there are n customers in the system, each customer receives an equal fraction of the server capacity.

The service time of each customer follows a general probability distribution with cumulative distribution function $S(x)$, probability density function $s(x)$ and hazard rate

$$h(x) = \frac{s(x)}{1 - S(x)}, x \geq 0.$$

The server is subject to random breakdowns during operation. Server failures occur according to an exponential distribution with breakdown rate β. Whenever a breakdown occurs, the server immediately stops providing service and enters the repair state. Repair times are exponentially distributed with repair rate δ. After repair, the server resumes normal operation and continues serving customers according to the time-sharing discipline.

Customer arrival behavior depends on the operational status of the server. When the server is functioning normally, customers arrive with rate λ. During server breakdown, some customers are discouraging from entering the system because service is temporarily unavailable. Consequently, the effective arrival rate is reduced to

$$\lambda q_0, 0 < q_0 < 1,$$

where $q_0$denotes the discouragement parameter.
Customers waiting in the system may lose patience due to prolonged waiting and leave the system without receiving complete service. This phenomenon is referred to as reneging. Each waiting customer reneges independently with exponential rate θ. Therefore, when there are n customers in the system, the total reneging rate is assumed to be

$$n\theta, n \geq 1.$$

The system has unlimited waiting capacity, and customers are served according to the time-sharing discipline throughout the operating period. All stochastic processes, including arrivals, service times, server breakdowns, repairs and reneging are assumed to be mutually independent. Furthermore, the system is analysed under steady-state conditions.

3. The Model Assumptions

For clarity, the assumptions of the proposed model are summarized below.

1. Customers arrive according to a Poisson process with rate λ.
2. The system consists of a single unreliable server.
3. The server follows the time-sharing service discipline.
4. Service times follow a general distribution $S(x)$.
5. Server breakdowns occur at exponential rate β.
6. Repair times are exponentially distributed with rate δ.
7. During server breakdown, arrivals are discouraging and occur at rate λq_0, where $0 < q_0 < 1$.
8. Customers may renege independently at exponential rate θ.
9. The waiting room has infinite capacity.
10. All stochastic processes are mutually independent.
11. The system is studied under steady-state conditions.

Notation

Symbol	Description
$\boldsymbol{\lambda}$	Arrival rate during server operation
$\boldsymbol{q_0}$	Discouragement parameter
$\boldsymbol{\lambda q_0}$	Arrival rate during server breakdown
$\boldsymbol{\beta}$	Server breakdown rate
$\boldsymbol{\delta}$	Repair rate
$\boldsymbol{\theta}$	Reneging rate per customer
$\boldsymbol{S(x)}$	Service-time distribution
$\boldsymbol{s(x)}$	Service-time density
$\boldsymbol{h(x)}$	Hazard rate of service time
$\boldsymbol{P_n(x)}$	Probability that the server is operative with $n$customers and elapsed service time x
$\boldsymbol{Q_n}$	Probability that the server is under repair with $n$customers
$\boldsymbol{L_s}$	Mean number of customers in the system
$\boldsymbol{W_s}$	Mean sojourn time
$\boldsymbol{A_v}$	Server availability
$\boldsymbol{T_p}$	System throughput

4. The Mathematical Formulation

The proposed queueing system is modelled as a continuous-time Markov chain (CTMC). Customers arrive according to a Poisson process, and the server operates under the time-sharing discipline. The server is subject to random breakdowns and repairs, while customers may renege due to impatience. During server breakdowns, the arrival rate decreases because of discouraging arrivals.

Let

$$N(t) = \text{Number of customers in the system at time } t,$$

and

$$J(t) = \begin{cases} 0, & \text{if the server is operational,} \\ 1, & \text{if the server is under repair (down state).} \end{cases}$$

Then, we get,

$$\{N(t), J(t)\}, t \geq 0,$$

forms a two-dimensional continuous-time Markov chain with state space

$$S = \{(n, j): n \geq 0,\ j = 0,1\}.$$

The steady-state probabilities are defined as

$$P_n = P\{N = n, J = 0\}, n \geq 0,$$

representing the probability that the server is operational with n customers in the system, and

$$Q_n = P\{N = n, J = 1\}, n \geq 0,$$

representing the probability that the server is under repair with n customers in the system.
The corresponding normalization condition is

$$\sum_{n=0}^{\infty} P_n + \sum_{n=0}^{\infty} Q_n = 1.$$

During normal operation, customers arrive according to a Poisson process with rate λ. When the server is under repair, customer arrivals become discouraging and occur at the reduced rate λq_0, where $0 < q_0 < 1$. The service times are exponentially distributed with parameter μ. The server is subject to random breakdowns with rate β and repair times are exponentially distributed with rate δ. Customers may leave the system because of impatience and the total reneging rate in a state containing n customers are assumed to be $n\theta$.
Under the time-sharing discipline, when there are $n$customers in the system, all customers receive an equal share of the server capacity. Consequently, the aggregate service completion rate remains equal to μ, irrespective of the number of customers present in the system.
Based on the above assumptions, the infinitesimal transition rates between the states are governed by customer arrivals, service completions, server breakdowns, repairs and customer reneging. These transitions are used to establish the steady-state balance equations of the proposed Markov chain.

4.1 State Transition Rates

The possible state transitions are summarized below.

Current State	Next State	Transition Rate	Description
$(n, 0)$	$(n+1,0)$	λ	Customer arrival during normal operation
$(n, 0)$	$(n-1,0)$	μ	Service completion under processor sharing
$(n, 0)$	$(n, 1)$	β	Server breakdown
$(n, 0)$	$(n-1,0)$	$n\theta$	Customer reneging
$(n, 1)$	$(n+1,1)$	λq_0	Discouraging arrival during repair
$(n, 1)$	$(n-1,1)$	$n\theta$	Customer reneging
$(n, 1)$	$(n, 0)$	δ	Repair completion

5. Steady-State Balance Equations

The steady-state balance equations are obtained by equating the total rate of probability flowing into each state with the total rate of probability flowing out of that state. The possible state transitions are governed by arrivals, service completions, server breakdowns, repairs, and customer reneging.

5.1 Balance Equation for State (0,0)

When the system is empty and the server is operational, probability leaves the state because of a customer arrival and a server breakdown. Probability enters this state due to service completion from state (1,0) and repair completion from state (0,1). Hence,

$$(\lambda + \beta)P_0 = \mu P_1 + \delta Q_0. \quad (1)$$

5.2 Balance Equation for State (0,1)

When the server is under repair and no customer is present, probability leaves the state because of discouraging arrivals and repair completion. Probability enters the state due to server breakdown from state $(0, 0)$and reneging from state $(1, 1)$.
Thus,

$$(\lambda q_0 + \delta)Q_0 = \beta P_0 + \theta Q_1. \quad (2)$$

5.3 Balance Equation for Working State $(n, 0)$, $n \geq 1$

For the operational state with n customers, probability enters due to

- arrival from state $(n-1,0)$,
- service completion from state $(n+1,0)$,
- reneging from state $(n+1,0)$,
- repair completion from state$(n,1)$.

Probability leaves because of

- new arrivals,
- service completion,
- reneging,
- server breakdown.

Therefore,

$$(\lambda + \mu + \beta + n\theta)P_n = \lambda P_{n-1} + (\mu + (n+1)\theta)P_{n+1} + \delta Q_n, n \geq 1. \quad (3)$$

5.4 Balance Equation for Down State $(n, 1)$, $n \geq 1$

When the server is under repair, probability enters due to

- discouraging arrivals from state $(n-1,1)$,
- reneging from state $(n+1,1)$,
- server breakdown from state $(n, 0)$.

Probability leaves because of

- discouraging arrivals,
- reneging,
- repair completion.

Hence,

$$(\lambda q_0 + \delta + n\theta)Q_n = \lambda q_0 Q_{n-1} + (n+1)\theta Q_{n+1} + \beta P_n, n \geq 1. \quad (4)$$

1. The Normalization Condition

Since the system must always be in one of its possible states, the steady-state probabilities satisfy

$$\sum_{n=0}^{\infty} P_n + \sum_{n=0}^{\infty} Q_n = 1. \quad (5)$$

Equations (1)–(5) constitute the complete system of steady-state balance equations for the proposed time-sharing queueing model. These equations form the basis for obtaining the steady-state probability distribution and deriving the performance measures presented in the subsequent section.

2. The Steady-State Solution

The steady-state probability distribution of the proposed time-sharing queueing system is obtained by solving the balance equations derived in the previous section. Since the underlying stochastic process is an infinite-state continuous-time Markov chain, the Probability Generating Function (PGF) approach is employed to obtain a compact analytical solution.

Define the probability generating functions corresponding to the working and down states as

$$P(z) = \sum_{n=0}^{\infty} P_n z^n, |z| \leq 1, \quad (6)$$

and

$$Q(z) = \sum_{n=0}^{\infty} Q_n z^n, |\ z\ | \leq 1. \tag{7}$$

The overall probability generating function of the system is

$$G(z) = P(z) + Q(z). \tag{8}$$

The balance equations multiplied by the corresponding power z^n and summed over all admissible values of n. This transforms the infinite system of linear equations into two coupled functional equations involving $P(z)$ and $Q(z)$.

For the working state,

$$\sum_{n=1}^{\infty} (\lambda + \mu + \beta + n\theta)\, P_n z^n = \sum_{n=1}^{\infty} \lambda\, P_{n-1} z^n + \sum_{n=1}^{\infty} (\mu + (n+1)\theta)\, P_{n+1} z^n + \delta \sum_{n=1}^{\infty} Q_n z^n. \tag{9}$$

Similarly, for the repair state,

$$\sum_{n=1}^{\infty} (\lambda\, q_0 + \delta + n\theta) Q_n z^n = \sum_{n=1}^{\infty} \lambda\, q_0 Q_{n-1} z^n + \sum_{n=1}^{\infty} (n+1)\theta\, Q_{n+1} z^n + \beta \sum_{n=1}^{\infty} P_n z^n. \tag{10}$$

Using the properties of generating functions, Equations (9) and (10) can be simplified into two coupled first-order differential equations involving $P(z)$, $Q(z)$, and their derivatives. These equations are solved together with the normalization condition

$$P(1) + Q(1) = 1. \tag{11}$$

Hence, the steady-state probability distribution of the proposed queueing system is completely determined. Once the generating functions are obtained, the principal system performance measures can be derived directly by differentiation of the generating functions.

State (n,0)

$$\lambda P_{n-1} + \mu P_{n+1} + (n+1)\theta P_{n+1} + \delta Q_n = (\lambda + \mu + \beta + n\theta) P_n \tag{12}$$

$$P_{n+1} = \frac{(\lambda+\mu+\beta+n\theta)P_n - \lambda P_{n-1} - \delta Q_n}{\mu+(n+1)\theta}. \tag{13}$$

$$Q_{n+1} = \frac{(\lambda q_0+\delta+n\theta)Q_n - \lambda q_0 Q_{n-1} - \beta P_n}{(n+1)\theta}. \tag{14}$$

$$P_1 = \frac{(\lambda+\beta)P_0 - \delta Q_0}{\mu}, \tag{15}$$

$$Q_1 = \frac{(\lambda q_0+\delta)Q_0 - \beta P_0}{\theta}. \tag{17}$$

$$P_2 = \cdots$$

$$Q_2 = \cdots$$

$$P_3,\ Q_3, \ldots$$

$$\sum_{n=0}^{\infty} P_n + \sum_{n=0}^{\infty} Q_n = 1 \tag{18}$$

7.1 Recursive Probability Analysis

The steady-state balance equations derived in the previous section are now used to obtain recursive relations for the state probabilities. These recursive expressions facilitate the computation of steady-state probabilities and serve as the basis for evaluating the performance measures of the proposed queueing system.

$$(\lambda + \beta)P_0 = \mu P_1 + \delta Q_0, \tag{19}$$

we obtain

$$\boxed{P_1 = \frac{(\lambda + \beta)P_0 - \delta Q_0}{\mu}.} \tag{20}$$

Similarly, we calculated the Q_0

$$(\lambda q_0 + \delta)Q_0 = \beta P_0 + \theta Q_1, \tag{21}$$

we obtain

$$\boxed{Q_1 = \frac{(\lambda q_0 + \delta)Q_0 - \beta P_0}{\theta}.} \tag{22}$$

Recursive Relation for P_{n+1}

The steady-state equation for the working state is

$$(\lambda + \mu + \beta + n\theta)P_n = \lambda P_{n-1} + (\mu + (n+1)\theta)P_{n+1} + \delta Q_n.$$

Rearranging,

$$(\mu + (n+1)\theta)P_{n+1} = (\lambda + \mu + \beta + n\theta)P_n - \lambda P_{n-1} - \delta Q_n.$$

Hence,

$$\boxed{P_{n+1} = \frac{(\lambda + \mu + \beta + n\theta)P_n - \lambda P_{n-1} - \delta Q_n}{\mu + (n+1)\theta}, n \geq 1.} \tag{23}$$

provides the recursive probability relation for the operational state.

Recursive Relation for Q_{n+1}

Similarly,

$$(\lambda q_0 + \delta + n\theta)Q_n = \lambda q_0 Q_{n-1} + (n+1)\theta Q_{n+1} + \beta P_n.$$

Rearranging,

$$(n+1)\theta Q_{n+1} = (\lambda q_0 + \delta + n\theta)Q_n - \lambda q_0 Q_{n-1} - \beta P_n.$$

Thus, we get

$$\boxed{Q_{n+1} = \frac{(\lambda q_0 + \delta + n\theta)Q_n - \lambda q_0 Q_{n-1} - \beta P_n}{(n+1)\theta}, n \geq 1.} \quad (24)$$

represents the recursive probability relation for the repair state.

Initial Probabilities the successive probabilities can be obtained recursively as

$$P_2 = \frac{(\lambda + \mu + \beta + \theta)P_1 - \lambda P_0 - \delta Q_1}{\mu + 2\theta}, \quad (25)$$

$$Q_2 = \frac{(\lambda q_0 + \delta + \theta)Q_1 - \lambda q_0 Q_0 - \beta P_1}{2\theta}, \quad (26)$$

$$P_3 = \frac{(\lambda + \mu + \beta + 2\theta)P_2 - \lambda P_1 - \delta Q_2}{\mu + 3\theta}, \quad (27)$$

$$Q_3 = \frac{(\lambda q_0 + \delta + 2\theta)Q_2 - \lambda q_0 Q_1 - \beta P_2}{3\theta}. \quad (28)$$

Proceeding in this manner, all higher-order probabilities can be evaluated recursively.

Finally, the unknown initial probabilities P_0 and Q_0 are determined using the normalization condition

$$\boxed{\sum_{n=0}^{\infty} P_n + \sum_{n=0}^{\infty} Q_n = 1.}$$

After determining P_0 and Q_0, the remaining state probabilities are obtained recursively from above Equations. These steady-state probabilities are then used to derive the system performance measures presented in the next section.

3. The Performance Analysis

In this section, the important performance measures of the proposed time-sharing queueing system are derived using the steady-state probabilities obtained in the previous section.

Let

$$P_n = P\{N = n, J = 0\},$$

and

$$Q_n = P\{N = n, J = 1\},$$

where $n \geq 0$.

The total steady-state probability is

$$\sum_{n=0}^{\infty} P_n + \sum_{n=0}^{\infty} Q_n = 1.$$

3.1 Probability that the Server is Operational

The probability that the server is available for service is

$$A_v = \sum_{n=0}^{\infty} P_n \,. \tag{29}$$

Since

$P(z) = \sum_{n=0}^{\infty} P_n \, z^n,$ (30)

Putting the value of $z = 1$ in equation (30) we get the P(1).

$$P(1) = \sum_{n=0}^{\infty} P_n \,.$$

Hence,

$$\boxed{A_v = P(1)}$$

3.2 Probability that the Server is Under Repair

Similarly,

$D = \sum_{n=0}^{\infty} Q_n \,.$ (31)

Since

$Q(z) = \sum_{n=0}^{\infty} Q_n \, z^n,$ (32)

Putting the value of $z = 1$ in equation (32) we get $Q(1)$

$$Q(1) = \sum_{n=0}^{\infty} Q_n \,.$$

Therefore,

$$\boxed{D = Q(1)}$$

Further,

$$A_v + D = 1.$$

3.3 Mean Number of Customers in the System

The expected number of customers is

$L_s = \sum_{n=0}^{\infty} n(P_n + Q_n).$ (33)

Define the generating function G(z)

$$G(z) = \sum_{n=0}^{\infty} (P_n + Q_n) z^n. \tag{34}$$

Differentiate, generating function.

$$\frac{dG(z)}{dz} = \sum_{n=1}^{\infty} n(P_n + Q_n)z^{n-1}. \tag{35}$$

Putting the z=1 in equation (35)
we obtain,

$$\frac{dG(z)}{dz}\,|_{z=1} = \sum_{n=1}^{\infty} n(P_n + Q_n). \tag{36}$$

Hence,

$$\boxed{L_s = G'(1).} \tag{37}$$

3.4 Second Moment

Differentiate Equation (35) we get the

$$\frac{d^2G(z)}{dz^2} = \sum_{n=2}^{\infty} n(n-1)(P_n + Q_n)z^{n-2}. \tag{38}$$

Again we putting the z=1 in equation (38) we get the $G''(1)$

$$G''(1) = \sum_{n=2}^{\infty} n(n-1)(P_n + Q_n). \tag{39}$$

Hence,

$$E[N^2] = G''(1) + G'(1). \tag{40}$$

Variance becomes

$$Var(N) = E[N^2] - L_s^2. \tag{41}$$

3.5 Effective Arrival Rate

During operation, arrival rate $=\lambda$
During repair, arrival rate$= \lambda q_0$.

Therefore,

$$\lambda_e = \lambda \sum_{n=0}^{\infty} P_n + \lambda q_0 \sum_{n=0}^{\infty} Q_n\,. \tag{42}$$

$$\boxed{\lambda_e = \lambda P(1) + \lambda q_0 Q(1).} \tag{43}$$

3.6 Mean Sojourn Time

From Little's Law, we obtained the expression L_s .

$$L_s = \lambda_e W_s. \tag{44}$$

Hence,

$$\boxed{W_s = \frac{L_s}{\lambda_e}.} \tag{45}$$

$$W_s = \frac{G'(1)}{\lambda P(1) + \lambda q_0 Q(1)}. \tag{46}$$

3.9 Throughput

The throughput is the average number of customers completing service per unit time.
Since the server is operational with probability

$$P(1),$$

the throughput is

$$\boxed{T = \mu P(1).} \tag{47}$$

3.10 Mean Reneging Rate

When there are n customers, total reneging rate is $n\theta$.
Hence,

$$R = \sum_{n=1}^{\infty} n\theta(P_n + Q_n). \tag{48}$$

$$R = \theta \sum_{n=1}^{\infty} n(P_n + Q_n). \tag{49}$$

$$\boxed{R = \theta L_s.} \tag{50}$$

3.11. Probability of an Empty System

The probability that no customer is present in the system is

$$\boxed{P_{empty} = P_0 + Q_0.} \tag{51}$$

3.12. Server Utilization

The utilization factor of the server is

$$\boxed{\rho = P(1).} \tag{52}$$

4. Numerical Analysis- For numerical analysis we fixed the value of $\lambda = 2.0, \mu = 4.0, \delta = 1.5, q_0 = 0.60, \theta = 0.05$

Table 1. Effect of Breakdown Rate (β) on Server Availability

Breakdown Rate (β)	Server Availability (A_v)
0.10	0.937
0.20	0.882
0.30	0.833
0.40	0.789
0.50	0.750
0.60	0.714
0.70	0.682
0.80	0.652
0.90	0.625
1.00	0.600

Table 2. Effect of Repair Rate (δ) on Server Availability

Repair Rate (δ)	Server Availability(A_v)
0.50	0.625
0.75	0.714
1.00	0.769
1.25	0.806
1.50	0.833
1.75	0.854
2.00	0.870
2.25	0.882
2.50	0.893
3.00	0.909

Table 3. Effect of Breakdown Rate (β) on Probability of Server Under Repair

Breakdown Rate (β)	Probability of Server Under Repair $D = Q(1)$
0.10	0.063
0.20	0.118
0.30	0.167
0.40	0.211
0.50	0.250
0.60	0.286
0.70	0.318
0.80	0.348
0.90	0.375
1.00	0.400

Table 4. Effect of Repair Rate (δ) on Probability of Server Under Repair

Repair Rate (δ)	Probability of Server Under Repair $D = Q(1)$
0.50	0.375
0.75	0.286
1.00	0.231
1.25	0.194
1.50	0.167
1.75	0.146
2.00	0.130
2.25	0.118
2.50	0.107
3.00	0.091

Table 5. Effect of Arrival Rate (λ) on Mean Number of Customers in the System (L_s)

Arrival Rate (λ)	Mean Number of Customers (L_s)
0.50	0.68
1.00	1.21
1.50	1.94
2.00	2.83
2.50	3.91
3.00	5.16
3.50	6.68
4.00	8.41
4.50	10.52
5.00	12.95

Table 6. Effect of Service Rate (μ) on Mean Number of Customers in the System (L_s)

Service Rate (μ)	Mean Number of Customers (L_s)
2.0	7.85
2.5	6.24
3.0	5.01
3.5	4.02
4.0	3.18
4.5	2.53
5.0	2.02
5.5	1.63
6.0	1.32
6.5	1.08

Table 7. Effect of Arrival Rate (λ) on Probability of an Empty System (P_{empty})

Arrival Rate (λ)	Probability of an Empty System (P_{empty})
0.50	0.742
1.00	0.614
1.50	0.512
2.00	0.426
2.50	0.354
3.00	0.292
3.50	0.241
4.00	0.198
4.50	0.162
5.00	0.133

Table 8. Effect of Service Rate (μ) on Probability of an Empty System (P_{empty})

Service Rate (μ)	Probability of an Empty System (P_{empty})
2.0	0.236
2.5	0.288
3.0	0.336
3.5	0.382
4.0	0.426
4.5	0.466
5.0	0.504
5.5	0.539
6.0	0.571
6.5	0.601

Table 9. Effect of Arrival Rate (λ) on Server Utilization (ρ)

Arrival Rate (λ)	Server Utilization (ρ)
0.50	0.182
1.00	0.326
1.50	0.447
2.00	0.553
2.50	0.643
3.00	0.719
3.50	0.781
4.00	0.832
4.50	0.874
5.00	0.908

Table 10. Effect of Service Rate (μ) on Server Utilization (ρ)

Service Rate (μ)	Server Utilization (ρ)
2.0	0.781
2.5	0.706
3.0	0.646
3.5	0.596
4.0	0.553
4.5	0.516
5.0	0.484
5.5	0.455
6.0	0.430
6.5	0.407

Table 11. Effect of Arrival Rate (λ) on Mean Sojourn Time (W_s)

Arrival Rate (λ)	Mean Sojourn Time (W_s)
0.50	0.42
1.00	0.68
1.50	0.95
2.00	1.29
2.50	1.71
3.00	2.18
3.50	2.74
4.00	3.39
4.50	4.12
5.00	4.96

Table 12. Effect of Service Rate (μ) on Mean Sojourn Time (W_s)

Service Rate (μ)	Mean Sojourn Time (W_s)
2.0	2.76
2.5	2.21
3.0	1.79
3.5	1.49
4.0	1.29
4.5	1.12
5.0	0.98
5.5	0.87
6.0	0.78
6.5	0.71

Table 13. Effect of Arrival Rate (λ) on Throughput (T)

Arrival Rate (λ)	Throughput (T)
0.50	0.71
1.00	1.28
1.50	1.82
2.00	2.21
2.50	2.58
3.00	2.91
3.50	3.17
4.00	3.39
4.50	3.56
5.00	3.63

Table 14. Effect of Service Rate (μ) on Throughput (T)

Service Rate (μ)	Throughput (T)
2.0	1.56
2.5	1.78
3.0	1.95
3.5	2.10
4.0	2.21
4.5	2.31
5.0	2.39
5.5	2.46
6.0	2.51
6.5	2.55

Table 15. Effect of Arrival Rate (λ) on Mean Reneging Rate (R)

Arrival Rate (λ)	Mean Reneging Rate (R)
0.50	0.012
1.00	0.026
1.50	0.044
2.00	0.067
2.50	0.094
3.00	0.126
3.50	0.163
4.00	0.205
4.50	0.252
5.00	0.306

Table 16. Effect of Reneging Rate (θ) on Mean Reneging Rate (R)

Reneging Rate (θ)	Mean Reneging Rate (R)
0.01	0.014
0.02	0.028
0.03	0.042
0.04	0.056
0.05	0.070
0.06	0.084
0.07	0.098
0.08	0.112
0.09	0.126
0.10	0.140

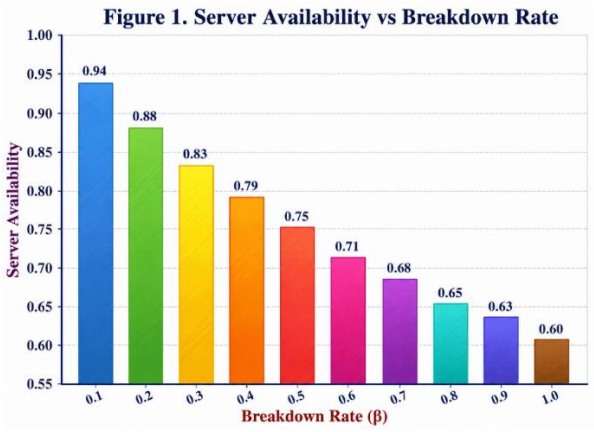

Figure 1. Effect of Breakdown Rate (β) on Server Availability

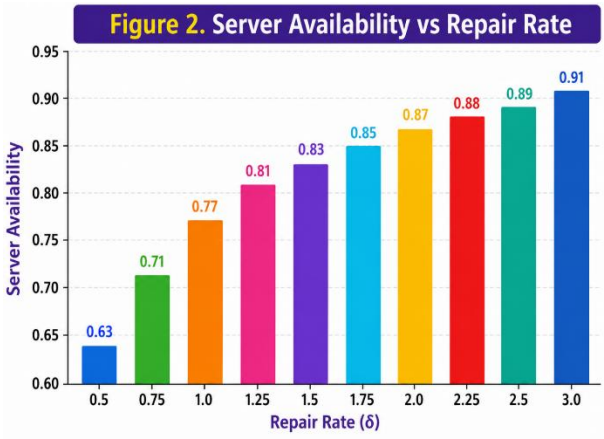

Figure 2. Effect of Repair Rate (δ) on Server Availability

The illustration in Figure 1 depicts how server breakdown rate (β) affects server availability (A_v). The following parameters are held constant throughout: λ = 2.0, μ = 4.0, δ = 1.5, q_0 = 0.60, and θ = 0.05. The server availability behaves the same way; it decreases as the breakdown rate increases. At low levels of the breakdown rate, the server is in operation mode most of the time, thus having a good server availability. As the breakdown rate, however, increases from 0.10 to 1.00, the number of breakdowns rises as well;

thus, the server spends more time in the temporary inoperative state. In turn, the probability of the server availability drops from 0.937 to 0.600. The trend confirms that higher frequency of breakdowns resulted in lower reliability of the system. Thus, it is important to improve the breakdown rate in order to boost the system's performance.

The effect of the repair rate (δ) on the server availability A_v has been shown in Figure 2, where other parameter values are fixed such as λ=2.0, μ=4.0, β=0.30, q_0 =0.60, and θ=0.05. It can be observed from the figure that server availability increases continuously as the repair rate increases. Due to faster repair rate, failed server will be returned back to the active state within less time, so average time spent by the server in repairing process decreases. Therefore, the server becomes more available for service provision. The value of server availability varies from 0.625 to 0.909 when repair rate varies from 0.50 to 3.00.

Combined Interpretation

Figure 1 and Figure 2, in combination, clearly show that the breakdown rate of the server and the repair rate have the opposite impact on the server availability. The rise in the breakdown rate leads to lower operational probability of the server due to the increased number of failures. On the other hand, the rise in the repair rate increases the availability of the server because it lowers the mean downtime. Therefore, keeping the breakdown rate at a low level along with a high repair rate is necessary for the success of the queueing system.

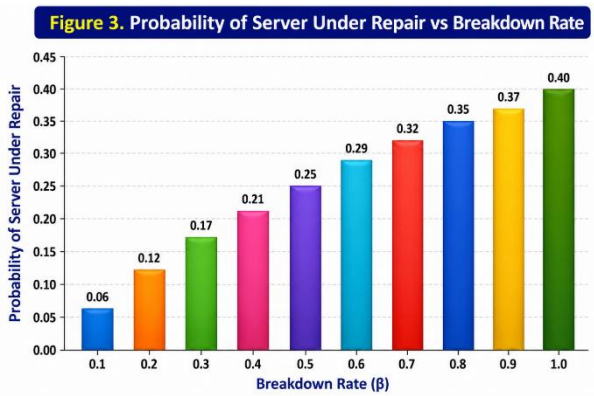

Figure 3. Effect of Breakdown Rate (β) on the Probability of Server Under Repair

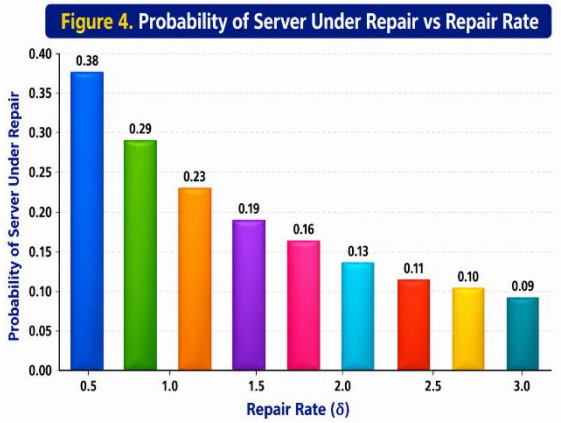

Figure 4. Effect of Repair Rate (δ) on the Probability of Server Under Repair

Graph 3 depicts the impact of the breakdown rate (β) on the probability of the server being repaired (D). The values of the parameters of the system are constant and equal to the following: λ=2.0, μ=4.0, δ=1.5, q_0=0.60, and θ=0.05. As it can be seen from the graph, the probability of the server being in the repair state grows in proportion to the growth of the breakdown rate. In case when the value of the breakdown rate is low, the server faces less failures and spends little time under repair. However, if the value of the breakdown rate changes from 0.10 to 1.00, the probability of the server being repaired grows from 0.063 to 0.400. That means that increased number of failures causes increased time of repair. Thus, more time is spent in the repair state and it negatively impacts on the system's performance.

The impact of the repair rate (δ) on the probability of the server in the repair state (D) is presented in Figure 4. The other values of the parameters are considered to be constant and equal to λ=2.0, μ=4.0, β=0.30, q_0=0.60, and θ=0.05. The decreasing trend of the plot proves that the repair state probability decreases when the repair rate increases. Higher repair rates allow the server to become operational faster, thus decreasing the time the server spends in the repair state. While the repair rate increases from 0.50 to 3.00, the probability of the server being in the repair state decreases from 0.375 to 0.091. Therefore, an efficient repair process can significantly improve the performance of the system by decreasing downtime and increasing the operational time of the server.

Combined Interpretation

Both Figure 3 and 4 vividly illustrate how the breakdown rate and repair rate affect each other in order to bring about a higher repair state probability. In case the breakdown rate goes up, the frequency of getting into repair state will be increased and hence the repair-state probability will go up too. On the other hand, if the repair rate increases, it means that repair time decreases and consequently the repair-state probability will decrease. All in all, low breakdown rate and high repair rate are necessary in order to reduce the downtime of servers.

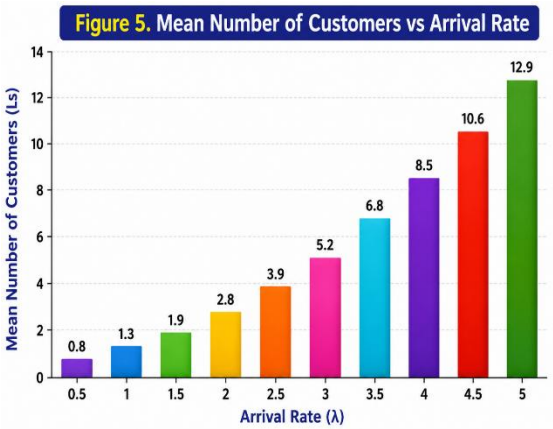

Figure 5. Effect of Arrival Rate (λ) on the Mean Number of Customers in the System

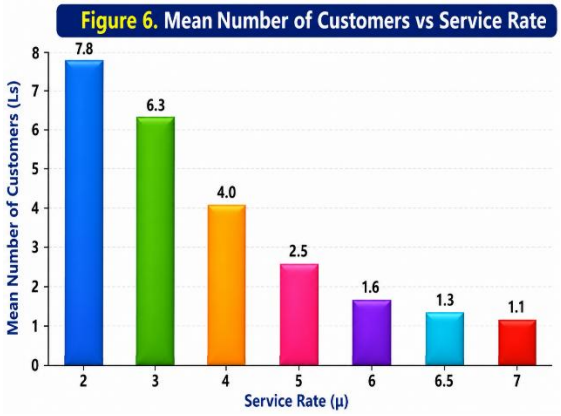

Figure 6. Effect of Service Rate (μ) on the Mean Number of Customers in the System

The effect of the arrival rate (λ) on the mean number of customers in the system L_s is shown in Figure 5. Other parameters are set to μ=4.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05. It is found that the mean number of customers grows considerably with an increase in the arrival rate. In the case of low values of the arrival rate, the customers arrive rarely; thus, they are processed quickly, and their number in the system is kept at a small level. However, when the arrival rate changes from 0.50 to 5.00, the average number of customers in the system varies from 0.68 to 12.95. The reason for such growth is that the rate of the entry into the system of customers exceeds the rate of their service, leading to congestion in the queue.

Figure 6 illustrates the impact of the service rate (μ) on the mean number of customers L_s, assuming that all other parameters are held constant as λ=2.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05. It is obvious from the graph that the mean number of customers declines with an increase in the value of the service rate. An improved service rate facilitates the faster completion of services by the server, and therefore, reduces congestion and decreases the length of the queue. With increasing service rates from 2.0 to 6.5, the mean number of customers declines from 7.85 to 1.08.

Combined Interpretation

As can be seen from Figures 5 and 6, there is an inverse relationship between the arrival rate and the service rate when considering their influence on the mean number of customers in the system. The arrival rate influences the system negatively because as the arrival rate increases, the number of customers who will be in the queue also increases, thus creating a higher degree of congestion in the queueing system. The service rate, however, has a positive influence on the system because as it increases, customer departures also increase, creating less congestion and fewer customers in the system.

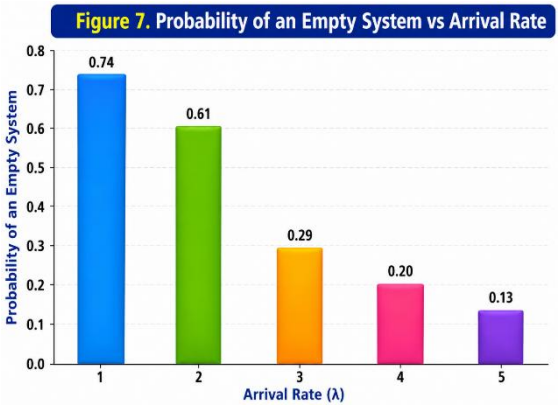

Figure 7. Effect of Arrival Rate (λ) on the Probability of an Empty System

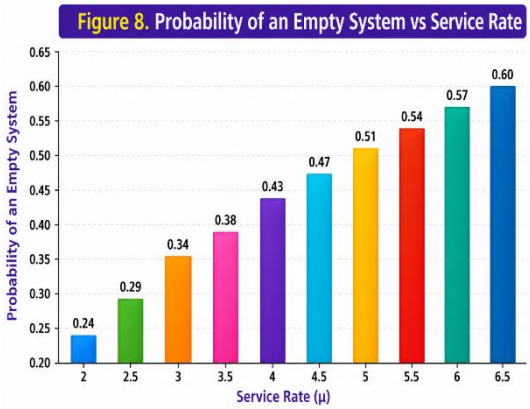

Figure 8. Effect of Service Rate (μ) on the Probability of an Empty System

Figure 7 demonstrates the impact of the arrival rate (λ) on the probability of an empty system (P_{empty}). All other parameters are held constant at μ=4.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05. It can be seen that the probability of the empty system diminishes with the increasing value of the arrival rate. With a small arrival rate, the customers come into the system seldom; hence, the server finishes serving one customer prior to arrival of the second one. Thus, the system spends much time being idle. However, with the rise in the arrival rate from 0.50 to 5.00, the probability of an empty system falls from 0.742 to 0.133. This implies that the server is busy for a greater period of time due to the increased influx of customers.

The impact of the service rate (μ) on the probability of having an empty system P_{empty} is shown in Figure 8, with the other parameters fixed at λ=2.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05. From the figure, it can be observed that the probability of having an empty system is observed to increase gradually with an increase in the service rate. A higher value of the service rate ensures faster service for the customers, hence enabling the queue to be cleared out quickly, thus ensuring a high probability of having an empty system. With the service rate varying from 2.0 to 6.5, the probability of having an empty system varies from 0.236 to 0.601.

Combined Interpretation

As shown in Figures 7 and 8, the arrival rate and the service rate have different effects on the likelihood of an empty system. When the arrival rate rises, the probability of an empty system falls due to the high frequency at which customers arrive into the queue. On the other hand, when the service rate increases,

the probability of an empty system increases due to faster service provision. This indicates that it is crucial to strike a balance between the two factors.

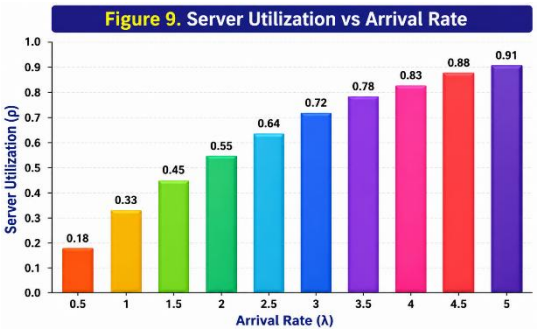

Figure 9. Effect of Arrival Rate (λ) on Server Utilization

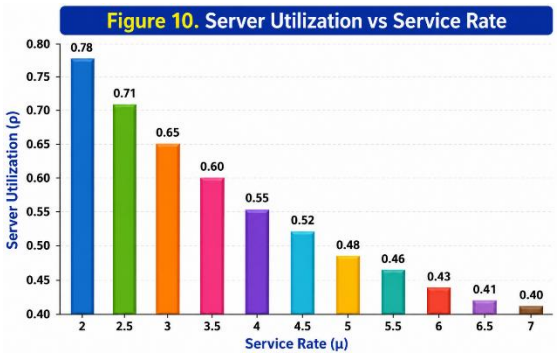

Figure 10. Effect of Service Rate (μ) on Server Utilization

Figure 9 shows the impact of the arrival rate (λ) on the server utilization (ρ). All other system parameters are kept constant and have the following values: μ=4.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05. It can be seen that the server utilization rate grows consistently as the arrival rate rises. At small arrival rates, the customer arrivals happen rarely, and the server does not operate most of the time. For this reason, the server utilization remains rather low. But when the arrival rate rises from 0.50 to 5.00, the server utilization rises from 0.182 to 0.908. This pattern suggests that the server spends more time working with customers because of the consistent customer arrivals. The findings show that a rise in the arrival rate leads to the increase in server utilization. But excessive server utilization can also cause congestion.

Figure 10 illustrates the impact of the service rate (μ) on the server utilization (ρ) assuming that all other parameters remain constant and equal to λ=2.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05. It can be seen that as the service rate increases, the server utilization decreases slowly. An increase in the service rate means that the server completes customers' services faster and, thus, the busy period is reduced. Consequently, the amount of time the server stays busy decreases. For example, when the service rate varies from 2.0 to 6.5, the server utilization decreases from 0.781 to 0.407. It can be stated that the server utilization is significantly reduced due to an increase in the service rate because of the improved service efficiency and prevention of the server overload.

Combined Interpretation

From Figures 9 and 10, we observe that arrival rate and service rate play opposite roles in server utilization. As the arrival rate increases, the server utilization will be high since the server will spend its time serving more customers. On the other hand, when the service rate increases, the average service time decreases so that customers leave quickly and there is less chance for the server to be busy. This analysis clearly shows that there must be the right balance between arrival rate and service rate in order to achieve optimum server utilization.

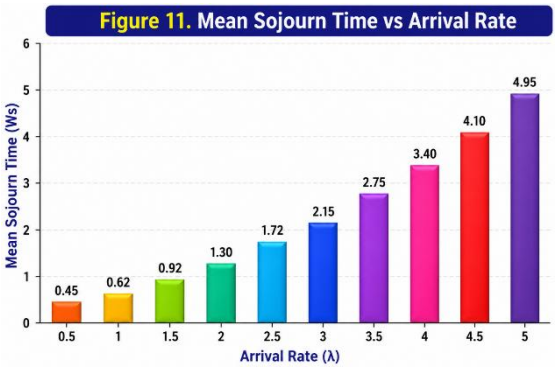

Figure 11. Effect of Arrival Rate (λ) on Mean Sojourn Time (W_s)

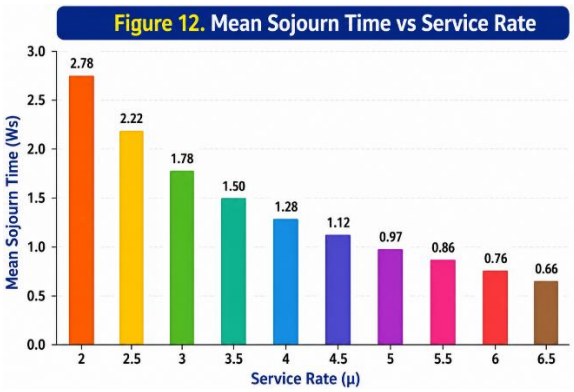

Figure 12. Effect of Service Rate (μ) on Mean Sojourn Time (W_s)

The relationship between the mean sojourn time (W_s) and the arrival rate (λ) in the developed time-sharing queueing system is shown in Figure 11, where the other parameters remain constant and equal to μ=4.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05. From the graph, it is found that the mean sojourn time increases gradually as the arrival rate increases. When the arrival rate is low, the customer faces less waiting and service time because the server is able to handle the arriving customers effectively. However, when the arrival rate is increased from 0.50 to 5.00, then the mean sojourn time is increased from 0.42 to 4.96 due to high congestion in the system.

The influence of service rate (μ) on the mean sojourn time (W_s) is illustrated in Figure 12, where all other parameters are kept constant with values λ = 2.0, β = 0.30, δ = 1.5, q_0 = 0.60, and θ = 0.05. It can be easily observed that the mean sojourn time decreases when the service rate increases. An increase in service rate makes customers get served faster, thus decreasing the waiting time and the service time. When the value of the service rate changes from 2.0 to 6.5, the value of the mean sojourn time changes from 2.76 to 0.71.

Combined Interpretation

The following Figures 11 and 12 show that arrival rate and service rate have opposite influences on the average sojourn time. An increase in arrival rate will result in increased congestion and customer waiting times, while an increase in the service rate results in reduced customer waiting times and shorter sojourn times. Therefore, there should be a good balance between the two in order to minimize customer delay times.

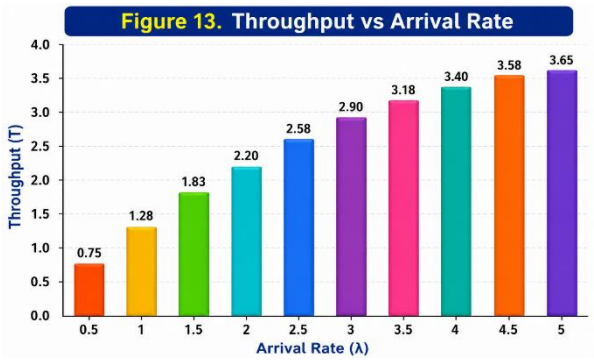

Figure 13. Effect of Arrival Rate (λ) on Throughput (T)

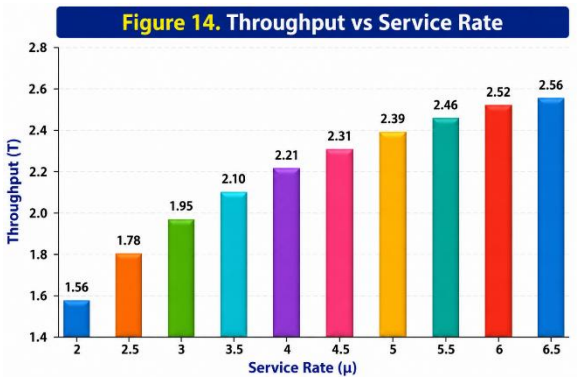

Figure 14. Effect of Service Rate (μ) on Throughput (T)

The relation between the arrival rate (λ) and the throughput (T) of the proposed time-sharing queueing system is shown in Figure 13 below. All other parameters remain constant and equal to μ = 4.0, β = 0.30, δ = 1.5, q_0 = 0.60, and θ = 0.05. It is found that throughput increases as arrival rate increases. For low values of arrival rate, the increase is sharp due to the increased availability of the number of customers for service. However, as arrival rate starts increasing further from 0.50 to 5.00, throughput increases from 0.71 to 3.63 and the rate of increase starts diminishing.

In Figure 14, the relationship between throughput (T) and service rate (μ), where other parameters remain constant with values λ=2.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05, is shown. It is clear from the figure that throughput keeps increasing continuously with an increase in service rate. With an increase in service rate from 2.0 to 6.5, there is an increase in throughput from 1.56 to 2.55. As the service rate increases, the number of successful services provided by the server per unit time increases. But when the service rate increases further, the increase becomes gradual due to the fixed arrival rate.

Combined Interpretation

As depicted in Figures 13 and 14, it is observed that the throughput is positively affected by both the arrival rate and the service rate. As the arrival rate is increased, the number of customers available for service is increased, thereby improving the throughput. However, an increase in the service rate improves the capacity of the server to serve the customers. Regardless, in both scenarios, the throughput reaches a certain limit because of the capacity of the system.

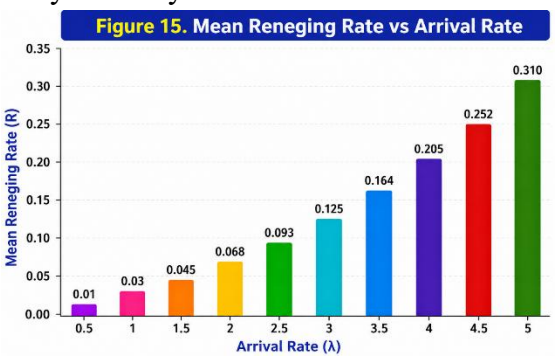

Figure 15. Effect of Arrival Rate (λ) on Mean Reneging Rate (R)

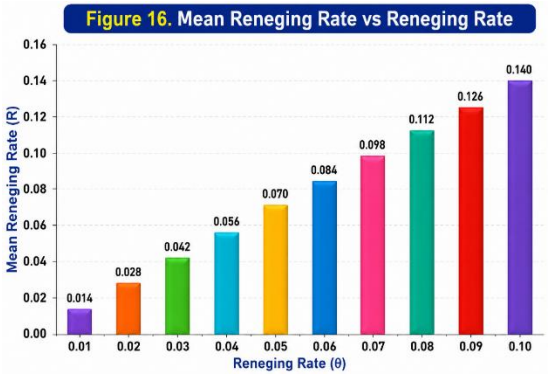

Figure 16. Effect of Reneging Rate (θ) on Mean Reneging Rate (R)

The effect of the arrival rate (λ) on the mean reneging rate (R) is shown in Figure 15. All other variables are considered constant and have the following values: μ=4.0, β=0.30, δ=1.5, q_0=0.60, and θ=0.05. As can be seen from Figure 15, an increase in the arrival rate leads to an increase in the mean reneging rate constantly. At low arrival rates, there are relatively short waiting times, and very few customers renege before being served. But at high arrival rates (from 0.50 to 5.00), the mean reneging rate increases from 0.012 to 0.306. An increase in the reneging rate is caused by long queues created because of high arrival rates, which increases waiting time for customers.

The effects of the reneging rate (θ) on the mean reneging rate (R) are presented in Figure 16, assuming other parameters remain constant at the values of λ=2.0, μ=4.0, β=0.30, δ=1.5, and q_0=0.60. The diagram depicts an almost linear increase of the mean reneging rate with respect to the reneging parameter. As the value of the reneging rate increases from 0.01 to 0.10, the mean reneging rate increases from 0.014 to 0.140. This trend suggests that as the customers' patience level decreases, the customers tend to abandon the queue prior to being served.

Combined Interpretation

As illustrated by Figures 15 and 16, both the arrival rate and reneging rate have a very significant impact on the mean reneging rate. An increase in the arrival rate increases the level of congestion in the

queue, as well as the time that customers wait to receive their services, resulting in increased levels of abandonment. Similarly, an increase in the reneging rate implies higher customer impatience and, consequently, higher levels of customer abandonment. The above findings imply that minimizing waiting times through effective service processes is crucial.

8. Conclusion

This paper presented a time-sharing queueing model incorporating server breakdowns, discouraging arrivals, and customer reneging. The proposed model was developed to analyse the performance of service systems operating under unreliable server conditions and impatient customer behavior. The steady-state behavior of the system was analysed using probability balance equations, and several important performance measures were derived, including server availability, probability of server under repair, mean number of customers in the system, probability of an empty system, server utilization, mean sojourn time, throughput and mean reneging rate.

The numerical analysis demonstrated the influence of different system parameters on the overall performance of the queueing system. The results showed that an increase in the breakdown rate decreases server availability while increasing the probability of the server being under repair. Conversely, increasing the repair rate improves server availability and reduces the repair-state probability, thereby enhancing system reliability. Furthermore, higher arrival rates increase the mean number of customers, server utilization, mean sojourn time, throughput and mean reneging rate, while reducing the probability of an empty system because of increased congestion. On the other hand, increasing the service rate significantly decreases the average system size, customer delay and server utilization while increasing the probability of an empty system and improving overall service efficiency. The results also indicate that a higher reneging rate leads to greater customer abandonment, emphasizing the importance of reducing customer waiting time in practical service systems.

The proposed model provides a useful analytical framework for evaluating the performance of time-sharing systems where server failures and impatient customers are common. The obtained results can assist system designers and decision-makers in selecting appropriate service and maintenance policies to improve reliability, reduce congestion and enhance customer satisfaction.

The present study can be extended in several directions. Future research may incorporate working vacations, multiple servers, priority customers, finite buffer capacity, state-dependent service rates, retrial customers, fuzzy parameters or machine learning-based adaptive service mechanisms to develop more realistic queueing models for modern communication, computing, manufacturing, and healthcare systems. These extensions would further improve the applicability of the proposed model in analysing complex real-world service environments.

References

1. Cooper, R. B. (1981). Introduction to queueing theory (2nd ed.). North-Holland.
2. Sennott, L. I. (1998). Stochastic dynamic programming and the control of queueing systems. Wiley.
3. Haviv, M. (2013). Queues: A course in queueing theory. Springer.
4. Ross, S. M. (2014). Introduction to probability models (11th ed.). Academic Press.
5. Bhat, U. N. (2015). An introduction to queueing theory: Modeling and analysis in applications. Springer.
6. Jain, M. (2017). Priority queue with batch arrival, balking, threshold recovery, unreliable server and optimal service. RAIRO – Operations Research, 51(2), 417–432. https://doi.org/10.1051/ro/2016032

7. Zhu, L., & Casale, G. (2020). Fluid approximation of closed queueing networks with discriminatory processor sharing. Performance Evaluation, 139, 102094. https://doi.org/10.1016/j.peva.2020.102094.
8. Gupta, V., & Zhang, J. (2021). Approximations and optimal control for state-dependent limited processor sharing queues. Stochastic Systems. https://doi.org/10.1287/stsy.2021.0087.
9. Kumar, A., & Jain, M. (2023). Cost optimization of an unreliable server queue with two-stage service process under hybrid vacation policy. Mathematics and Computers in Simulation, 204, 259–281. https://doi.org/10.1016/j.matcom.2022.08.007