International Journal of
Operations Research

International Journal of Operations Research Vol. 20, 12−24 (2023)

Navigating Customer Demand in a Failure Prone Retrial Queueing System with Predictive Maintenance

Poornima Rathinasamy[1*], Kirupa Krishnasamy[1], Saranya Karnakumar[1], and Bathmakumari Ponnusamy[1]

[1]Avinashilingam Institute for Home Science and Higher Education for Women
Bharathi Park Rd, Saibaba Colony, Coimbatore, Tamil Nadu 641043, India

Received April 2023; Revised June 2023; Accepted August 2023

Abstract: This paper examines a retrial queueing system with repair and predictive maintenance involving a single server. Customers' arrival follows a Poisson process, and if the server is available, the customer is served immediately. If the server is busy, the customer is placed in a queue known as the orbit. Once the customer is served, they exit the system. At a rate of α, the server may experience breakdowns while busy. Whenever a breakdown occurs, the server requires repair, and the customer must wait in orbit for service until the server is fixed. If the server wants to undergo predictive maintenance at a rate of ω, it will transition to an idle state, even if there are n customers in the queue. The system's performance metrics are demonstrated through numerical analysis using the supplementary variable technique metrics, such as the average number of customers in both the queue and the system. The property of stochastic decomposition has also been verified to validate the proposed model.

Keywords – Retrial queue, Single arrival, Predictive maintenance, Repair.

INTRODUCTION

A retrial queueing system is a dynamic and adaptive model used to handle customer requests in various service environments. Unlike traditional queueing systems, where customers are simply rejected when service resources are unavailable, retrial systems allow customers to reattempt service after a certain waiting period. This waiting period provides customers with an opportunity to retry, resulting in improved customer satisfaction and increased system efficiency. Retrial queueing systems have found applications in telecommunications, call centers, and various service industries where customer requests are crucial. By incorporating the concept of retrial, these systems offer a more flexible and customer-centric approach to managing queues and ensuring efficient service delivery. Falin's (1990) survey on retrial queues serves as an indispensable resource for understanding the fundamental principles, modeling techniques, and performance evaluation methods in this field. Its comprehensive coverage, clarity of presentation, and timeless relevance make it an essential reference for researchers, practitioners, and students interested in the analysis and optimization of queueing systems. While some limitations exist, Falin's survey remains a significant contribution that has laid the groundwork for further advancements in the study of retrial queues. Aissani's (1994) analysis of retrial queues with redundancy and unreliable servers offers a significant contribution to the field of queueing theory. By combining redundancy and unreliable servers, the study introduces a realistic model that can be applied to various practical scenarios. Although the work has some limitations, it serves as a foundation for further research and provides valuable insights into the performance analysis of complex queueing systems. Yang and Wu's (2019) study on the performance analysis and optimization of a retrial queue with working vacations and starting failures is a noteworthy contribution to the field of queueing theory. This comprehensive approach adds practicality to their research, as it reflects real-world scenarios encountered in various service industries.

Preventive maintenance refers to a systematic approach to performing maintenance tasks on equipment, machinery, or systems at regular intervals to prevent potential failures and prolong their operational lifespan. The primary goal of preventive maintenance is to identify and address minor issues or signs of wear and tear before they escalate into major problems, leading to costly repairs, unplanned downtime, or safety hazards. By implementing a proactive maintenance strategy, organizations can minimize the risk of equipment failure, optimize operational efficiency, and reduce overall maintenance costs. It is vital across various industries, including manufacturing, transportation, healthcare, and facility management, to ensure equipment reliability and maximize productivity (Yang *et al.*, 2019). Research on a two-phase preventive maintenance policy considering imperfect repair and postponed replacement offers valuable insights into optimizing maintenance strategies in the presence of real-world constraints.

* Corresponding author's e-mail: poornimapro10@gmail.com

http://doi.org/10.6886/IJOR.202308_20.0002

Their rigorous analytical approach, optimization techniques, and practical insights make this study a significant contribution to the field of maintenance management. Azhagappan and Deepa's (2020) study's analytical framework demonstrates a deep understanding of queueing theory principles and incorporates several important aspects of real-world queueing systems. By including system disaster repair preventive maintenance, re-service, balking, closedown, and setup times, the authors provide a comprehensive analysis that reflects the intricacies of practical queueing scenarios. Peschansky's (2020) study on the stationary characteristics of an unreliable single-server queueing system with losses and preventive maintenance offers valuable insights into the analysis and optimization of queueing systems in the presence of preventive maintenance and system unreliability. The author addresses the complex challenge of incorporating both preventive maintenance and losses into the performance evaluation of a single-server queueing system. Wang *et al.* (2020) delve into optimizing the preventive maintenance schedule to maximize system availability. Their optimization analysis helps identify the most effective maintenance strategies based on the system's operational characteristics and performance requirements. This approach enables decision-makers to allocate maintenance resources efficiently and minimize disruptions to system availability. Rykov *et al.*(2021) research on preventive maintenance of the k-out-of-n system concerning a cost-type criterion offers valuable insights into optimizing maintenance strategies while considering cost as a primary criterion. Their rigorous analytical approach, optimization techniques, and consideration of practical constraints make this study a significant contribution to the field of maintenance optimization. It provides decision-makers with a framework to develop cost-effective and efficient preventive maintenance plans, ultimately improving system reliability and minimizing maintenance costs in real-world applications. The study conducted by Wang *et al.* (2022) on reliability modeling and analysis for linear consecutive-k-out-of-n: F retrial systems with two maintenance activities is a valuable contribution to the field of reliability engineering. The authors tackle a complex scenario where the system under consideration follows a linear consecutive-k-out-of-n: F structure and incorporates retrial mechanisms along with two distinct maintenance activities. Wang (2022) studied analytical framework demonstrates a deep understanding of retrial queueing systems, preventive maintenance, and system reliability. By incorporating these factors, the authors provide a comprehensive analysis that reflects the complexities of real-world scenarios, where retrials, preventive maintenance, and unreliable service stations can significantly impact system performance and costs.

Repair refers to the process of restoring or fixing a malfunctioning or damaged object or system to its proper working condition. It involves identifying the problem or fault, diagnosing the cause, and implementing the necessary repairs or replacements to resolve the issue. Gupta and Kumar's (2021) research on the performance analysis of a retrial queueing model with working vacation, interruption, waiting server, breakdown, and repair offers valuable insights into optimizing the performance and reliability of queueing systems. Gao and Wang (2021) investigated the reliability and availability analysis of a retrial system with mixed standbys, and an unreliable repair facility is a valuable contribution to the field of system reliability engineering. The authors address the problem by considering a retrial system with multiple types of standby units and an unreliable repair facility, which better reflects real-world scenarios where system components can have varying reliability characteristics.

The focus of our research explores the challenges of managing customer demand in a failure-prone retrial queueing system with predictive maintenance. In such systems, failures can lead to customer retrials, causing delays and potential service disruptions. To mitigate these issues, predictive maintenance strategies are implemented to proactively address potential failures. The goal is to minimize customer waiting times, reduce service disruptions, and maximize system performance. Numerical analyses are conducted to evaluate the effectiveness of the proposed approach. The findings provide insights into the benefits of predictive maintenance and highlight the importance of proactive management of customer demand in failure-prone retrial queueing systems. The results can guide decision-makers in effectively allocating resources, improving service levels, and enhancing customer satisfaction.

Practical Justification of the Model

The real-world application of the retrial queuing system with repair and preventive maintenance may be seen in a customer service department for telecoms. With each call center, the system manages a large amount of consumer calls. If a server is on duty, the call is connected, and the customer's question is answered. The customer enters an orbit and makes repeated attempts to connect if the server is busy. Breakdowns may occur when the server is interacting with callers, making the agent momentarily unavailable. When technical problems or system breakdowns are fixed, the server can resume providing customer service. In addition to breakdowns, the system incorporates predictive maintenance. This indicates that the server enters an idle state to do preventative maintenance even when consumers are in orbit waiting for service. The objective of this maintenance is to stop possible problems and guarantee the server's long-term availability. The server resumes operations and continues to serve clients after the maintenance is complete. This retrial queuing system improves call center operations by promptly responding to client questions, minimizing server failures, and preventatively maintaining the server's reliability and efficacy. It improves service quality and reduces downtime, which leads to increased call center efficiency overall.

2. MODEL DESCRIPTION

In this paper, we examine a retrial queuing system with repair and predictive maintenance that utilizes a single server, which operates under an M/G/1 model. Customer arrivals are based on a Poisson process at a rate λ, and they are served according to a first-come, first-served (FCFS) queue discipline. When a customer arrives, and the server is free, they are served right away.

If the server is occupied, the customer goes into an infinite capacity orbit and repeatedly attempts to connect until the service is completed with a retrial completion rate of η. Once the customer receives service, they leave the system with a service completion rate which follows an exponential distribution with a parameter of μ. While the server is occupied or engaged, it is subject to breakdown at a rate of α, and once the repair is completed, the server resumes its operation by providing service to customers at a repair completion rate which follows an exponential distribution with a parameter of β. Additionally, even though having n customers in orbit, the server enters the idle state to undergo predictive maintenance at a rate of ω. Upon completion of the predictive maintenance with rate Ω, the server recommences operation by servicing customers.

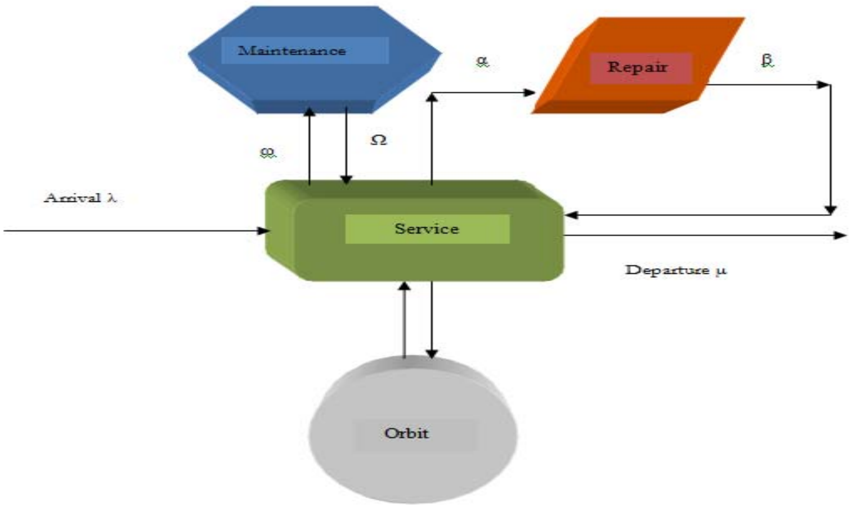

Figure 1. Structure of the model.

The time it takes for a customer to retry service after an unsuccessful attempt is defined by its distribution function $A(\wp)$, density function $a(\wp)$, Laplace-Stieltje's transform $A^*(\theta)$, and the rate at which a customer completes a retry attempt, known as the conditional completion rate $\eta(\wp) = \frac{a(\wp)}{1-A(\wp)}$. Similarly, the time it takes for a customer to receive service is characterized by its distribution function $B(\wp)$, density function $b(\wp)$, Laplace-Stieltje's transform $B^*(\theta)$, moments, μ_1 and μ_2, and its conditional completion rate $\mu(\wp) = \frac{b(\wp)}{1-B(\wp)}$. The predictive maintenance process required for the server is defined by its distribution function $W(\wp)$, density function $w(\wp)$, Laplace-Stieltje's transform $W^*(\theta)$, moments Ω_1 and Ω_2, and its conditional completion rate $\Omega(\wp) = \frac{w(\wp)}{1-W(\wp)}$. Finally, the repair time is described by its distribution function $R(\wp)$, density function $r(\wp)$, Laplace-Stieltje's transform $R^*(\wp)$, moments β_1 and β_2, and its conditional completion rate $\beta(\wp) = \frac{r(\wp)}{1-R(\wp)}$.

The expansion of the parameters which are using in this paper is mentioned below:

Parameters	Description
λ	arrival rate
η	successive retrial rate
$\mu(\wp)$	service completion rate
α	repair rate
$\beta(\wp)$	repair completion rate
ω	maintenance rate
$\Omega(\wp)$	maintenance completion rate

The distribution function, density function, Laplace Stieltje's transform, and the first two moments of retrial time, service time, repair time, and maintenance time, which are generally distributed, are given below.

Time	Distribution Function	Density Function	Laplace Stieltje's Transform	First Two Moments
Retrial	$A(\wp)$	$a(\wp)$	$A^*(\theta)$	-
Service	$B(\wp)$	$b(\wp)$	$B^*(\theta)$	μ_1, μ_2
Repair	$R(\wp)$	$r(\wp)$	$R^*(\theta)$	r_1, r_2
Maintenance	$W(\wp)$	$w(\wp)$	$W^*(\theta)$	Ω_1, Ω_2

3. STEADY-STATE EQUATION

This section obtains the steady state probability generating function of the orbit size distribution by considering the elapsed service time, elapsed predictive maintenance time, and elapsed repair time of the server as additional variables.

Define the states of the server at time t as

$$c(t) = \begin{cases} 0, & \text{when the server is idle} \\ 1, & \text{when the server is busy} \\ 2, & \text{when the server is under maintenance} \\ 3, & \text{when the server is under repair} \end{cases}$$

We define the random variable $\xi(t)$ as follows for $t \geq 0$:

(i) If $C(t) = 0$, then $\xi(t)$ indicates the elapsed retrial time,
(ii) If $C(t) = 1$, then $\xi(t)$ indicates the elapsed service time,
(iii) If $C(t) = 2$, then $\xi(t)$ indicates the elapsed maintenance time,
(iv) If $C(t) = 3$, then $\xi(t)$ indicates the elapsed repair time.

Then, the process $\{\wp(t); t \geq 0\} = \{C(t), N(t), \xi(t); t \geq 0\}$ is a Markov process, where N(t) denotes the number of customers in orbit. For the process $\{\wp(t); t \geq 0\}$, we define the following probabilities.

$$I_0(t) = P\{C(t) = 0, N(t) = 0\}$$
$$I_n(\wp, t)\,d\wp = P\{C(t) = 0, N(t) = n, \wp < \xi(t) \leq \wp + d\wp\}, n \geq 1$$
$$P_n(\wp, t)\,d\wp = P\{C(t) = 1, N(t) = n, \wp < \xi(t) \leq \wp + d\wp\}, n \geq 0$$
$$M_n(\wp, t)\,d\wp = P\{C(t) = 2, N(t) = n, \wp < \xi(t) \leq \wp + d\wp\}, n \geq 0$$
$$R_n(\wp, t)\,d\wp = P\{C(t) = 3, N(t) = n, \wp < \xi(t) \leq \wp + d\wp\}, n \geq 0$$

The following is the set of equations that governs the model using the supplementary variable technique,

$$\frac{d}{dt} I_0(t) = -(\lambda + \omega) I_0(t) + \int_0^\infty P_0(\wp, t)\mu(\wp)d\wp + \int_0^\infty M_0(\wp, t)\Omega(\wp)d\wp + \int_0^\infty R_0(\wp, t)\beta(\wp)d\wp \tag{1}$$

$$\left(\frac{d}{d\wp} + \frac{d}{dt}\right) I_n(\wp, t) = -(\lambda + \eta(\wp) + \omega) I_n(\wp, t), n \geq 1 \tag{2}$$

$$\left(\frac{d}{d\wp} + \frac{d}{dt}\right) P_n(\wp, t) = -(\lambda + \mu(\wp) + \alpha) P_n(\wp, t) + (1 - \delta_{0n})\lambda P_{n-1}(\wp, t), n \geq 0 \tag{3}$$

$$\left(\frac{d}{d\wp} + \frac{d}{dt}\right) M_n(\wp, t) = -(\lambda + \Omega(\wp)) M_n(\wp, t) + (1 - \delta_{0n})\lambda M_{n-1}(\wp, t), n \geq 0 \tag{4}$$

$$\left(\frac{d}{d\wp} + \frac{d}{dt}\right) R_n(\wp, t) = -(\lambda + \beta(\wp)) R_n(\wp, t) + (1 - \delta_{0n})\lambda R_{n-1}(\wp, t), n \geq 0 \tag{5}$$

with boundary conditions,

$$I_n(0,t)=\int_0^{\infty}P_n(\wp,t)\mu(\wp)d\wp+\int_0^{\infty}M_n(\wp,t)\Omega(\wp)d\wp+\int_0^{\infty}R_n(\wp,t)\beta(\wp)d\wp, n\geq 1 \quad (6)$$

$$P_0(0,t)=\lambda I_0(t)+\int_0^{\infty}I_1(\wp,t)\eta(\wp)d\wp \quad (7)$$

$$P_n(0,t)=\int_0^{\infty}I_{n+1}(\wp,t)\eta(\wp)d\wp+\lambda\int_0^{\infty}I_n(\wp,t)d\wp, n\geq 1 \quad (8)$$

$$M_0(0,t)=\omega I_0(t) \quad (9)$$

$$M_n(0,t)=\omega\int_0^{\infty}I_n(\wp,t)d\wp, n\geq 1 \quad (10)$$

$$R_n(0,t)=\alpha\int_0^{\infty}P_n(\wp,t)d\wp, n\geq 0 \quad (11)$$

Define the steady-state probabilities

$I_0=\lim_{t\to\infty}I_0(t)$;

$I_n(\wp)=\lim_{t\to\infty}I_n(\wp,t), \wp\geq 0, n\geq 1$;

$P_n(\wp)=\lim_{t\to\infty}P_n(\wp,t), \wp\geq 0, n\geq 0$;

$M_n(\wp)=\lim_{t\to\infty}M_n(\wp,t), \wp\geq 0, n\geq 0$;

$R_n(\wp)=\lim_{t\to\infty}R_n(\wp,t), \wp\geq 0, n\geq 0$.

By evaluating both sides of the equations at tending t towards infinity, we can obtain the steady-state equations.

$$(\lambda+\omega)I_0=\int_0^{\infty}P_0(\wp)\mu(\wp)d\wp+\int_0^{\infty}M_0(\wp)\Omega(\wp)d\wp+\int_0^{\infty}R_0(\wp)\beta(\wp)d\wp \quad (12)$$

$$\left(\frac{d}{d\wp}\right)I_n(\wp)=-(\lambda+\eta(\wp)+\omega)I_n(\wp), n\geq 1 \quad (13)$$

$$\left(\frac{d}{d\wp}\right)P_n(\wp)=-(\lambda+\mu(\wp)+\alpha)P_n(\wp)+(1-\delta_{0n})\lambda P_{n-1}(\wp), n\geq 0 \quad (14)$$

$$\left(\frac{d}{d\wp}\right)M_n(\wp)=-(\lambda+\Omega(\wp))M_n(\wp)+(1-\delta_{0n})\lambda M_{n-1}(\wp), n\geq 0 \quad (15)$$

$$\left(\frac{d}{d\wp}\right)R_n(\wp)=-(\lambda+\beta(\wp))R_n(\wp)+(1-\delta_{0n})\lambda R_{n-1}(\wp), n\geq 0 \quad (16)$$

with boundary conditions,

$$I_n(0)=\int_0^{\infty}P_n(\wp)\mu(\wp)d\wp+\int_0^{\infty}M_n(\wp)\Omega(\wp)d\wp+\int_0^{\infty}R_n(\wp)\beta(\wp)d\wp, n\geq 1 \quad (17)$$

$$P_0(0) = \lambda I_0 + \int_0^\infty I_1(\wp)\eta(\wp)d\wp \tag{18}$$

$$P_n(0) = \int_0^\infty I_{n+1}(\wp)\eta(\wp)d\wp + \lambda\int_0^\infty I_n(\wp)d\wp, n \geq 1 \tag{19}$$

$$M_0(0) = \omega I_0 \tag{20}$$

$$M_n(0) = \omega\int_0^\infty I_n(\wp)d\wp, n \geq 1 \tag{21}$$

$$R_n(0) = \alpha\int_0^\infty P_n(\wp)d\wp, n \geq 0 \tag{22}$$

The normalizing condition is

$$I_0 + \sum_{n=1}^\infty\int_0^\infty I_n(\wp)d\wp + \sum_{n=0}^\infty\int_0^\infty P_n(\wp)d\wp + \sum_{n=0}^\infty\int_0^\infty M_n(\wp)d\wp + \sum_{n=0}^\infty\int_0^\infty R_n(\wp)d\wp = 1 \tag{23}$$

Define the probability-generating functions for $|\Im| \leq 1$:

$I(\wp,\Im) = \sum_{n=1}^\infty I_n(\wp)\,\Im^n$;

$P(\wp,\Im) = \sum_{n=0}^\infty \mathrm{P}_n(\wp)\,\Im^n$;

$M(\wp,\Im) = \sum_{n=0}^\infty \mathrm{M}_n(\wp)\,\Im^n$ and

$R(\wp,\Im) = \sum_{n=0}^\infty \mathrm{R}_n(\wp)\,\Im^n$.

Multiplying equation (13) by $\Im^n$ and summing over n, we get

$$\left[\frac{d}{d\wp} + \left(\lambda + \eta(\wp) + \omega\right)\right]I(\wp,\Im) = 0 \tag{24}$$

Solving the partial differential equation (24), we get

$$I(\wp,\Im) = I(0,\Im)e^{-(\lambda+\omega)\wp}\left[1 - A(\wp)\right] \tag{25}$$

Similarly solving equations (14), (15), and (16), respectively, we obtain

$$P(\wp,\Im) = P(0,\Im)e^{-(\lambda+\alpha+\lambda\Im)\wp}\left[1 - B(\wp)\right] \tag{26}$$
$$M(\wp,\Im) = M(0,\Im)e^{-(\lambda-\lambda\Im)\wp}\left[1 - W(\wp)\right] \tag{27}$$
$$R(\wp,\Im) = R(0,\Im)e^{-(\lambda-\lambda\Im)\wp}\left[1 - R(\wp)\right] \tag{28}$$

Multiplying equations (17), (19), (21), and (22) by $\Im^n$ and summing over n, we get

$$I(0,\Im) = \int_0^\infty P(\wp,\Im)\mu(\wp)d\wp + \int_0^\infty M(\wp,\Im)\Omega(\wp)d\wp + \int_0^\infty R(\wp,\Im)\beta(\wp)d\wp - (\lambda + \omega)I_0 \tag{29}$$

$$P(0,\Im) = \frac{1}{\Im}\left[\int_0^\infty I(\wp,\Im)\eta(\wp)d\wp\right] + \lambda\left[\int_0^\infty I(\wp,\Im)d\wp + I_0\right] \tag{30}$$

$$M(0,\Im) = \omega\left[\int_0^\infty I(\wp,\Im)d\wp + I_0\right] \tag{31}$$

$$R(0,\Im)=\alpha\int_0^\infty P(\wp,\Im)d\wp \tag{32}$$

Using equation (25) in equation (30), we have

$$P(0,\Im)=\frac{1}{\Im}\left[\lambda\Im I_0+I(0,\Im)\left(A^*(a)+\frac{1}{(\lambda+\omega)}\Im\lambda\left[1-A^*(a)\right]\right)\right] \tag{33}$$

where, $a=(\lambda+\omega)$
Using equation (25) in equation (31), we have

$$M(0,\Im)=\omega\left[I_0+I(0,\Im)\frac{1}{\lambda+\omega}\left[1-A^*(a)\right]\right] \tag{34}$$

Using equation (26) in equation (32), we have

$$R(0,\Im)=\frac{\alpha}{g(\Im)}P(0,\Im)\left[1-B^*(g(\Im))\right] \tag{35}$$

Using equations (26), (27), and (28) in equation (29) and simplifying, we get

$$I(0,\Im)=P(0,\Im)B^*(g(\Im))+M(0,\Im)W^*(h(\Im))+R(0,\Im)R^*(h(\Im))-(\lambda+\omega)I_0 \tag{36}$$

Using equations (33), (34), and (35) in equation (36) and solving, we get

$$I(0,\Im)=\frac{\Im I_0(\lambda+\omega)\begin{Bmatrix}\lambda\left[(g(\Im))B^*(g(\Im))+\alpha R^*(h(\Im))\left[1-B^*(g(\Im))\right]\right]\\+g(\Im)\left[\omega W^*(h(\Im))-(\lambda+\omega)\right]\end{Bmatrix}}{D(z)} \tag{37}$$

Using equations (37) in equation (33) and simplifying, we get

$$P(0,\Im)=\frac{I_0 g(\Im)(\lambda+\omega)A^*(a)\left[\omega\left[W^*(h(\Im))-1\right]+\lambda(\Im-1)\right]}{D(\Im)} \tag{38}$$

Using equations (37) in equation (34) and simplifying, we get

$$M(0,\Im)=\frac{\omega I_0(\lambda+\omega)A^*(a)\left[\Im g(\Im)-g(\Im)B^*(g(\Im))-\alpha R^*(h(\Im))\left[1-B^*(g(\Im))\right]\right]}{D(\Im)} \tag{39}$$

Using equations (38) in equation (35) and simplifying, we get

$$R(0,\Im)=\frac{\alpha I_0(\lambda+\omega)A^*(a)\left[\omega\left[W^*(h(\Im))-1\right]+\lambda(\Im-1)\right]\left[1-B^*(g(\Im))\right]}{D(\Im)} \tag{40}$$

The function that partially generates the probability of the orbit size during server idle time is expressed as follows,

$$I(\Im)=\int_0^\infty I(\wp,\Im)d\wp$$
$$=\frac{I_0\Im\left[1-A^*(a)\right]\begin{Bmatrix}\lambda\left[g(\Im)B^*(g(\Im))+\alpha R^*(h(\Im))\left[1-B^*(g(\Im))\right]\right]\\+g(\Im)\left[\omega W^*(h(\Im))-(\lambda+\omega)\right]\end{Bmatrix}}{D(\Im)} \tag{41}$$

The function that partially generates the probability of the orbit size during service time is expressed as follows,

$$P(\Im) = \int_0^\infty P(\wp, \Im) d\wp$$
$$= \frac{I_0(\lambda+\omega)A^*(a)\left[1-B^*(g(\Im))\right]\left[\omega\left[W^*(h(\Im))-1\right]+\lambda(\Im-1)\right]}{D(\Im)} \quad (42)$$

The function that generates the probability of a partial size of the orbit when the server undergoes predictive maintenance is expressed as follows,

$$M(\Im) = \int_0^\infty M(\wp, \Im) d\wp$$
$$= \frac{I_0\omega(\lambda+\omega)A^*(a)\left[1-W^*(h(\Im))\right]\left\{\Im g(\Im) - g(\Im)B^*(g(\Im)) - \alpha R^*(h(\Im))\left[1-B^*(g(\Im))\right]\right\}}{D(\Im)h(\Im)} \quad (43)$$

The function that generates the probability of a partial size of the orbit when the server undergoes repair is expressed as follows,

$$R(\Im) = \int_0^\infty R(\wp, \Im) d\wp$$
$$= \frac{I_0\alpha(\lambda+\omega)A^*(a)\left[1-B^*(g(\Im))\right]\left[1-R^*(h(\Im))\right]\left\{\omega\left[W^*(h(\Im))-1\right]+\lambda(\Im-1)\right\}}{D(\Im)h(\Im)} \quad (44)$$

The partial probability generating function of the orbit size is given by,

$$P_q(\Im) = I_0 + I(\Im) + P(\Im) + M(\Im) + R(\Im)$$
$$= \frac{N(\Im)}{\lambda D(\Im)} \quad (45)$$

The partial probability generating function of the system size is given by,

$$P_s(\Im) = I_0 + I(\Im) + \Im P(\Im) + M(\Im) + R(\Im)$$
$$= \frac{N_1(\Im)}{h(\Im)D(\Im)} \quad (46)$$

where,
$h(\Im) = (\lambda - \lambda\Im)$
$g(\Im) = (\lambda + \alpha - \lambda\Im)$
$N(\Im) = I_0 g(\Im)(\lambda+\omega)A^*(a)\left[\lambda(\Im-1)+\omega\left[W^*(h(\Im))-1\right]\right]$

$$N_1(\Im) = I_0(\lambda+\omega)A^*(a)\left\{\left[\lambda(\Im-1)+\omega\left[W^*(h(\Im))-1\right]\right]\left\{\begin{matrix}\left[1-B^*(g(\Im))\right]\left[\Im(\lambda-\lambda\Im)+\alpha\right] \\ -g(\Im)\left[\Im - B^*(g(\Im))\right]\end{matrix}\right\}\right\}$$

$$D(\Im) = \Im g(\Im)\left[(\lambda+\omega)-\omega W^*(h(\Im))\left[1-A^*(a)\right]\right] - \left\{\begin{matrix}\left[(\lambda+\omega)A^*(a)+\lambda\Im\left[1-A^*(a)\right]\right] \\ \left[g(\Im)B^*(g(\Im))+\alpha R^*(h(\Im))\left[1-B^*(g(\Im))\right]\right]\end{matrix}\right\}$$

4. PERFORMANCE METRICS

By using the L-Hopital rule for equations (41)-(46), we derived the following performance metrics for various server states.
The probability that the server is idle during retrial time is given by,

$$I = \lim_{\Im \to 1} I(\Im)$$
$$= \frac{I_0\lambda\left[1-A^*(a)\right]\left[\lambda\left[1-B^*(\alpha)\right](1+\alpha\beta_1)+\alpha\omega\Omega_1\right]}{D'(1)} \quad (47)$$

The probability that the server is busy in service is given by,

$$P = \lim_{\Im \to 1} P(\Im) = \frac{I_0 \lambda(\lambda+\omega) A^*(a)\left[1 - B^*(\alpha)\right]\left(1+\omega\Omega_1\right)}{D'(1)} \tag{48}$$

The probability that the server undergoes predictive maintenance is given by,

$$M = \lim_{\Im \to 1} M(\Im) = \frac{I_0 \omega\Omega_1 (\lambda+\omega) A^*(a)\left[\alpha - \lambda\left[1 - B^*(\alpha)\right]\left(1+\alpha\beta_1\right)\right]}{D'(1)} \tag{49}$$

The probability that the server undergoes repair is given by,

$$R = \lim_{\Im \to 1} R(\Im) = \frac{I_0 \lambda\alpha\beta_1 (\lambda+\omega) A^*(a)\left[1 - B^*(\alpha)\right]\left(1+\omega\Omega_1\right)}{D'(1)} \tag{50}$$

Using equations (47) to (50), equation (23) become

$$I_0 = \frac{D'(1)}{\alpha(\lambda+\omega)A^*(a)(1+\omega\Omega_1)} \tag{51}$$

The mean number of customers in the orbit is given by,

$$L_q = \lim_{\Im \to 1} \frac{d}{d\Im} P_q(\Im) = \frac{D'(1)N''(1) - N'(1)D''(1)}{2\lambda D'(1)^2} \tag{52}$$

where,

$$N'(1) = I_0 \lambda\alpha(\lambda+\omega)A^*(a)\left(1+\omega\Omega_1\right)$$

$$N''(1) = I_0 \lambda^2 (\lambda+\omega)A^*(a)\left[\alpha\omega\Omega_2 - 2\left(1+\omega\Omega_1\right)\right]$$

$$D'(1) = \left\{ \begin{array}{l} -\lambda^2\left[1 - B^*(\alpha)\right] + \omega A^*(a)(\alpha-\lambda) - \lambda\alpha\omega\Omega_1\left[1 - A^*(a)\right] \\ + \lambda\omega A^*(\lambda)\left[B^*(\alpha) - \alpha\beta_1\left[1 - B^*(\alpha)\right]\right] + \lambda\alpha\left[A^*(a) - \lambda\beta_1\left[1 - B^*(\alpha)\right]\right] \end{array} \right\}$$

$$D''(1) = -\lambda \left\{ \begin{array}{l} \left[\omega A^*(a) + \lambda\right]\left\{2\left(1-\lambda\mu_1\right) + \lambda\alpha\left[\beta_2\left[1 - B^*(\alpha)\right] - 2\mu_1\beta_1\right]\right\} + \\ \left[1 - A^*(a)\right]\left\{\omega\left[\alpha\lambda\Omega_2 - 2\Omega_1(\lambda-\alpha)\right] - 2\lambda\left[B^*(\alpha) - \alpha\beta_1\left[1 - B^*(\alpha)\right]\right]\right\} \end{array} \right\}$$

The mean number of customers in the system is given by,

$$L_s = \lim_{\Im \to 1} \frac{d}{d\Im} P_s(\Im) = L_q + P \tag{53}$$

Under steady-state conditions, the average customer waiting time in the orbit (W_q) is found by known Little's formula as follows:

$$W_q = \frac{L_q}{\lambda} \tag{54}$$

Under steady-state conditions, the average customer waiting time in the system (W_s) is found by known Little's formula as follows:

$$W_s = \frac{L_s}{\lambda} \tag{55}$$

Expected orbit size when the server is inert,

$$S_I = [I'(\Im)]_{\Im=1} = \frac{D'(1)N_2''(1) - N_2'(1)D''(1)}{2D'(1)^2} \tag{56}$$

where,

$$N_2'(1) = I_0\lambda[1 - A^*(a)][\lambda[1 - B^*(\alpha)](1 + \alpha\beta_1) + \alpha\omega\Omega_1]$$

$$N_2''(1) = I_0\lambda[1 - A^*(a)]\{\lambda[1 - B^*(\alpha)][2(1 + \alpha\beta_1) + \lambda\alpha\beta_2] - 2\lambda^2\mu_1(1 + \alpha\beta_1) + \omega[\lambda\alpha\Omega_2 - 2\Omega_1(\lambda - \alpha)]\}$$

Expected orbit size when the server is busy,

$$S_P = [P'(\Im)]_{\Im=1} = \frac{D'(1)N_3''(1) - N_3'(1)D''(1)}{2D'(1)^2} \tag{57}$$

where,

$$N_3'(1) = I_0\lambda(\lambda + \omega)A^*(a)[1 - B^*(\alpha)](1 + \omega\Omega_1)$$

$$N_3''(1) = I_0\lambda^2(\lambda + \omega)A^*(a)[\omega\Omega_2[1 - B^*(\alpha)] - 2\mu_1(1 + \omega\Omega_1)]$$

Expected orbit size when the server is under maintenance

$$S_M = [M'(\Im)]_{\Im=1} = \frac{D_r''(1)N_4'''(1) - N_4''(1)D_r'''(1)}{3D_r''(1)^2} \tag{58}$$

where,

$$N_4''(1) = I_0\omega\Omega_1(\lambda + \omega)A^*(a)[\alpha - \lambda[1 - B^*(\alpha)](1 + \alpha\beta_1)]$$

$$N_4'''(1) = \lambda^2\omega I_0(\lambda + \omega)A^*(a)\begin{Bmatrix} 3\Omega_1\{2 - \lambda[2\mu_1(1 + \alpha\beta_1) + \alpha\beta_2[1 - B^*(\alpha)]]\} \\ -3\Omega_2[\alpha + \lambda[1 - B^*(\alpha)](1 + \alpha\beta_1)] - \lambda\alpha\Omega_3 \end{Bmatrix}$$

Expected orbit size when the server is under maintenance

$$S_R = [R'(\Im)]_{\Im=1} = \frac{D_r''(1)N_5'''(1) - N_5''(1)D_r'''(1)}{3D_r''(1)^2} \tag{59}$$

where,

$$N_5''(1) = I_0\lambda\alpha\beta_1(\lambda + \omega)A^*(a)[1 - B^*(\alpha)](1 + \omega\Omega_1)$$

$$N_5'''(1) = 3I_0\lambda^3\alpha(\lambda + \omega)A^*(a)\{[2\mu_1\beta_1 - \beta_2[1 - B^*(\alpha)]](1 + \omega\Omega_1) - \omega\Omega_2\beta_1[1 - B^*(\alpha)]\}$$

$$D_r = h(z)D(z)$$

$$D_r''(1) = -2\lambda D'(1)$$

$$D_r'''(1) = -3\lambda D''(1)$$

The system availability at time t is the probability that the server is either working state or in an idle state. Then under steady state condition availability of the server is shown to be

$$\text{A} = 1 - [M + R] = 1 - \left[\frac{I_0(\lambda + \omega)A^*(\lambda)\{\omega\Omega_1[\alpha - \lambda[1 - B^*(g(z))]] + \lambda\alpha\beta_1[1 - B^*(g(z))]\}}{D'(1)}\right] \tag{60}$$

The steady-state failure frequency of the server is

$$F = \alpha P = \frac{I_0\lambda\alpha(\lambda + \omega)A^*(a)[1 - B^*(\alpha)](1 + \omega\Omega_1)}{D'(1)} \tag{61}$$

5. STOCHASTIC DECOMPOSITION LAW

To validate the derived result of our proposed model by verifying the Stochastic decomposition property. Fuhrmann and Cooper have found that queueing models of the M/G/1 type with server vacations exhibit stochastic decomposition (1985). Additionally, Krishnakumar and Arivudainambi (2002) have demonstrated that stochastic decomposition holds for several retrial queueing models.

Theorem:
The product of independent random variables (one of which is the number of customers in the corresponding classical queueing system (in steady-state) at a random time, and the other of which, depending on the scheduling of the vacations) may have different probabilistic interpretations in different conditions.

Proof:
The following is a possible representation of the PGF for the stationary system size distribution (46).

$$P_s(\Im) = \left[\left\{ \frac{\begin{array}{c}\left\{-\lambda^2\left[1-B^*(\alpha)\right]+\omega(\alpha-\lambda)+\lambda\omega\left[B^*(\alpha)-\alpha\beta_1\left[1-B^*(\alpha)\right]\right]+\lambda\alpha\left[1-\lambda\beta_1\left[1-B^*(\alpha)\right]\right]\right\} \\ \left[\lambda(\Im-1)+w\left[W^*(h(\Im))-1\right]\right]\left\{\left[1-B^*(g(\Im))\right]\left[\Im(\lambda-\lambda\Im)+\alpha\right]-g(\Im)\left[\Im-B^*(g(\Im))\right]\right\}\end{array}}{\alpha(1+\omega\Omega_1)h(\Im)(\lambda+\omega)\left[\Im g(\Im)-g(\Im)B^*(g(\Im))-\alpha\left[1-B^*(g(\Im))\right]R^*(h(\Im))\right]} \right\} \times \left\{ \frac{(\lambda+\omega)A^*(a)\left[\Im g(\Im)-g(\Im)B^*(g(\Im))-\alpha\left[1-B^*(g(\Im))\right]R^*(h(\Im))\right]D'(1)}{A^*(a)D(\Im)\left\{-\lambda^2\left[1-B^*(\alpha)\right]+\omega(\alpha-\lambda)+\lambda\omega\left[B^*(\alpha)-\alpha\beta_1\left[1-B^*(\alpha)\right]\right]+\lambda\alpha\left[1-\lambda\beta_1\left[1-B^*(\alpha)\right]\right]\right\}} \right\} \right] \quad (62)$$

Let $P_s(\Im)$ be the stationary system size distribution with a single server, maintenance, and repair, which is the convolution of two independent random variables $\pi(\Im)$ and $\chi(\Im)$.

1. The system size distribution with a single server, maintenance, and repair [second term of equation (54)].
2. The conditional distribution of the number of the retrial group under the assumption that the system is inert. [First term of equation (54)].

Our system's retry definition may also be used to derive the second term.

$$\chi(\Im) = \frac{I_0 + I(\Im)}{I_0 + I}$$

From equation (54), it is quite evident that,

$$P_s(\Im) = \pi(\Im) \times \chi(\Im).$$

6. NUMERICAL ANALYSIS

The table provides numerical data that demonstrate how different system parameters affect certain performance measures, including I_0 (the probability of having no customers in the system), I (the probability of the server being idle when the system is not empty), P (the probability of the server is servicing a customer), M (the probability of the server is undergoing maintenance), R (the probability that the server is being repaired), and L_s (the average number of customers in the system).

By fixing the parameters $\eta=1$, $\mu=2$, $\alpha=3$, $\beta=2$, and $\Omega=1$ and by varying values of λ and ω, the results are calculated and are presented in Table1. The findings indicate that as the value of λ increases, M decrease while P, R, L_q, and L_s increase. In addition, when ω is increased, M, L_q, and L_s increase, and P and R remain unchanged.

Table 2 displays the experimental results obtained by setting the system parameters as follows: $\mu=1$, $\lambda=2$, $\alpha=3$, $\omega=5$, and $\Omega=1$ for different values of β and η. The findings reveal that I, R, and L_s decrease as the value of β increases, whereas Mincrease and P remain unchanged. When the value of η is increased, the performance measures I is observed to decrease, while P, M, and R remain constant, along with a rise in L_s.

Table 1: Performance measures when η=1, μ= 2, α=3, β= 2, and Ω=1 by varying values of λ and ω.

λ	Ω	P	M	R	L_q	L_s
2	3	0.2222	0.3333	0.3333	0.5621	0.7844
	4	0.2222	0.3556	0.3333	0.6318	0.8540
	5	0.2222	0.3704	0.3333	0.6841	0.9064
	6	0.2222	0.3810	0.3333	0.7241	0.9463
2.2	3	0.2444	0.2917	0.3667	1.1567	1.4011
	4	0.2444	0.3111	0.3667	1.2326	1.4770
	5	0.2444	0.3241	0.3667	1.2898	1.5343
	6	0.2444	0.3333	0.3667	1.337	1.5781
2.4	3	0.2667	0.2500	0.4000	1.8113	2.0780
	4	0.2667	0.2667	0.4000	1.8992	2.1659
	5	0.2667	0.2778	0.4000	1.9646	2.2313
	6	0.2667	0.2857	0.4000	2.0144	2.2811
2.6	3	0.2889	0.2083	0.4333	2.5276	2.8165
	4	0.2889	0.2222	0.4333	2.6325	2.9214
	5	0.2889	0.2315	0.4333	2.7090	2.9979
	6	0.2889	0.2381	0.4333	2.7666	3.0555

Table 2: Performance measures when μ= 1, λ=2, α=3,ω= 5 and Ω=1 by varying values of β_1 and η.

β	η	I	P	M	R	L_s
3	2	0.7011	0.4444	0.0926	0.4444	1.7506
	3	0.3739	0.4444	0.0926	0.4444	2.0904
	4	0.2103	0.4444	0.0926	0.4444	2.6950
	5	0.1122	0.4444	0.0926	0.4444	4.0716
3.1	2	0.6994	0.4444	0.1045	0.4301	1.7348
	3	0.3730	0.4444	0.1045	0.4301	2.0676
	4	0.2098	0.4444	0.1045	0.4301	2.6654
	5	0.1119	0.4444	0.1045	0.4301	4.0533
3.2	2	0.6978	0.4444	0.1157	0.4167	1.7194
	3	0.3721	0.4444	0.1157	0.4167	2.0455
	4	0.2093	0.4444	0.1157	0.4167	2.6362
	5	0.1116	0.4444	0.1157	0.4167	4.0339
3.3	2	0.6962	0.4444	0.1263	0.4040	1.7047
	3	0.3713	0.4444	0.1263	0.4040	2.0240
	4	0.2089	0.4444	0.1263	0.4040	2.6074
	5	0.1114	0.4444	0.1263	0.4040	4.0137

7. CONCLUSION

In this paper, we analyzed a single server retrial queue with a single phase of essential service, repair, and predictive maintenance. By using the supplementary variable technique, it is possible to determine the steady-state probability-generating function for the system and the number of customers in the system while it is idle, busy, and undergoing repair and maintenance. The average number of customers in orbit or in the system and the average waiting time of a client in orbit or the system were calculated as important performance measures and also verified by the stochastic decomposition law. Numerical examples are used to validate the analytical conclusions. Also given a particularly suitable real-life application in the telecommunications call center. When the server is in an idle state, the system incorporates predictive maintenance for the server. So it will help the server farm avoid unexpected breakdowns and save downtime. This approach ensures that the servers are operating at their optimal performance levels, improving overall system reliability and user experience. From this, we can conclude that the system is always stable by reducing the breakdown of the server. Because of this, maintenance can reduce the customer waiting time in the orbit as well as in the system. This model may be expanded by including the following ideas: postponing repairs, vacation policies, and impatient customers.

REFERENCES

1. Aissani, A. (1994). A retrial queue with redundancy and unreliable server. *Queueing Systems*, 17: 431–449. https://doi.org/10.1007/BF01158703
2. Azhagappan, A. and Deepa, T. (2020). Transient analysis of N-policy queue with system disaster repairs preventive maintenance re-service balking closedown and setup times. *Journal of Industrial & Management Optimization*, 16(3): 2843-2856. https://doi.org/10.3934/jimo.2019083
3. Falin, G. (1990) A survey of retrial queues. *Queueing Systems*, **7**,127–167. https://doi.org/10.1007/BF01158472
4. Gao, S. and Wang, J. (2021). Reliability and availability analysis of a retrial system with mixed standbys and an unreliable repair facility. *Reliability Engineering and System Safety*, 205: 107240. http://dx.doi.org/10.1016/j.ress.2020.107240
5. Gupta, P. and Kumar, N. (2021). Performance analysis of retrial queueing model with working vacation, interruption, waiting server, breakdown and repair. *Journal of Scientific Research*, 13(3): 833–844. https://doi.org/10.3329/jsr.v13i3.52546
6. Peschansky, A. (2020). Stationary Characteristics of an unreliable single-server queueing system with losses and preventive maintenance. *Autom Remote Control*, 81: 1243–1257.https://doi.org/10.1134/S0005117920070061
7. Rykov, V., Kochueva, O., and Rykov, Y. (2021). Preventive maintenance of the *k*-out-of-*n* system with respect to cost-type criterion. *Mathematics*, 9: 2798. https://doi.org/10.3390/math9212798
8. Wang, K.-H., Wu, C.-H., and Yen, T.-C. (2022). Comparative cost-benefit analysis of four retrial systems with preventive maintenance and unreliable service station. *Reliability Engineering and System Safety*, 221(3): 108342, ISSN 0951-8320. http://dx.doi.org/10.1016/j.ress.2022.108342
9. Wang, N., Hu, J., Ma, L., Xiao, B., and Liao, H. (2020). Availability analysis and preventive maintenance planning for systems with general time distributions. *Reliability Engineering and System Safety*, 201: 106993. https://doi.org/10.1016/j.ress.2020.106993
10. Wang, Y., Hu, L., Yang, L., and Li, J. (2022). Reliability modeling and analysis for linear consecutive-k-out-of-n: F retrial systems with two maintenance activities. *Reliability Engineering and System Safety*, 226: 108665. https://doi.org/10.1016/j.ress.2022.108665
11. Yang, D.-Y. and Wu, C.-H. (2019). Performance analysis and optimization of a retrial queue with working vacations and starting failures. *Mathematical and Computer Modelling of Dynamical Systems*, 25(5): 463-481.https://doi.org/10.1080/13873954.2019.1660378
12. Yang, L., Ye, Z., Lee, C.-G., Yang, S., and Peng, R. (2019). A two-phase preventive maintenance policy considering imperfect repair and postponed replacement. *European Journal of Operational Research*, 274(3): 966-977. http://dx.doi.org/10.1016/j.ejor.2018.10.049