Original Research

Cost and sensitivity analysis of transient Markovian queue with waiting server, heterogeneous vacation policy and reneged customers

Mayank Singh[1]. Madhu Jain[1*]. A. Azhagappan[2]

Received: 15 Dec 2023/ Accepted: 11 Jan 2025/ Published online: 31 Jan 2025

*Corresponding Author, madhu.jain@ma.iitr.ac.in

1- Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee-247667, India.
2- Dpartment of Mathematics, St. Anne's College of Engineering and Technology, Anna University, Panruti, Cuddalore district, 607110 Tamilnadu, India.

Abstract
The unavailability of server during complete vacation (CV) and the reduced service pace during working vacation (WV) may pose a critical challenge to the service and manufacturing organizations. This study focuses on analyzing the performance of a queueing system operating under a hybrid vacation policy while considering customer impatience. The server has the option to choose either CV or WV, immediately after finishing all the jobs. During the WV period, the server renders service in slow pace as such some impatient customers may abandon the system before served. Moreover, the joining rate of the customers is state dependent i.e. the arriving customers joins the system with different rates for the different status of the server. The transient analytical expressions for the queue size distribution using continued fraction (CF) and probability generating function (PGF) methods are established which are further used to formulate various performance metrics. The practical applicability of the system is demonstrated through numerical simulation Moreover, cost optimization is performed to obtain minimum cost and corresponding optimal service rate.

Keywords- Transient queue; Hybrid vacation policy; Reneging; Probability Generating function, Continued fraction; Cost optimization.

INTRODUCTION

Efficient utilization of idle periods is important for the improving resource optimization. This can enhance the performance and reliability of service systems. The server's idle period can be efficiently managed by optimizing service facility resources using vacation queueing models. Queueing theorists have developed these models, allowing the server to take a vacation or break once all current jobs in the system have been serviced. The vacation is defined as the server's unavailability at the service station (Doshi 1986). Sikdar and Gupta (2005) examined a batch service queueing system with a finite buffer and multiple vacations, deriving performance metrics including queue length distributions, loss probability, and mean system size. Isijola-Adakeja and Ibe (2014) formulated the queue size distribution for a Markovian model that incorporated a differentiated vacation policy. Utilizing the recursive approach, they obtained the probability distribution for the queue size. Suranga Sampath and Liu (2020) examined the impact of customer reneging behavior in a Markovian single-server queue with vacation. Azhagappan and Deepa (2022) examined the vacation interruption policy for a transient M/M/1 queueing model with differentiated vacation types. They employed analytical methods involving PGF and CF to derive transient outcomes for the system size and related performance measures. Kumar (2023) dealt with transient Markov queue with discouragements and multiple vacation policies.

They used the confluent hypergeometric function (CHF) approach to formulate the various transient analytical formulae. As industries continue to evolve, the demand for more efficient service systems has never been greater. The need for enhanced server utilization during idle periods drives the search for innovative solutions that balance both service efficiency and resource management. The concept of working vacations (WV) introduces a promising avenue to maximize server productivity while maintaining service standards. In service systems where even short periods of inactivity can lead to substantial losses, such strategies can offer a competitive advantage. An effective working vacation (WV) strategy enables the server to offer services at reduced rates during the vacation period. Servi and Finn (2002) generalized the concept of server vacation by extending the classical vacation model to allow the server to operate at varying rates during vacation periods, instead of completely halting. The motivation for this research stems from the server's utilization and cost effectiveness approach of the service providers, especially in the service systems where complete server unavailability could lead to significant delays and revenue losses. In the context of conventional WV queueing systems, it is observed that the server immediately begins a WV period upon completing the task for the last available customer in the system. Li and Tian (2007) examined the M/M/1 queue with working vacation (WV) by employing quasi-birth-and-death processes and matrix-geometric methods. They derived key performance metrics, including customer distribution and waiting time, alongside their stochastic decomposition structures.

To study M/M/1 queue having multiple WV, Sudhesh and Raj (2012) have made use of the methods of probability generating function (PGF) and continued fraction (CF) to derive the transient formula for the queue size distribution. Ma et al. (2018) examined M/M/1/WV queueing model operating under N-policy. They derived the analytical formulae to analyze the queue size distribution, mean queue length, and social welfare function. Jain et al. (2019) examined the effect of customers' feedback and discouragement on a Markov queue with working vacation. They used the technique of CF and PGF to formulate the various queueing performance metrics including expected queueing length, throughput, operational cost of the system, etc. Jain et al. (2021) developed a transient Markov queueing model that incorporates the disaster failure and multiple working vacation policy. To derive the transient queue size and various system indices, they employed the technique of CF and PGF, in conjunction with Laplace transform. Sridhar et al. (2022) studied Markovian queueing system with WV, server breakdowns, and customer impatience behaviors such as balking and reneging. By solving the transient state probabilities through differential equations, various performance indices were computed, and a sensitivity analysis was conducted to evaluate parameter impacts. Ramesh and Udayabaskaran (2023) dealt with a queueing with multi-level environment, random switch and working vacation policy, where the system transitions between N levels based on probabilistic assignments. Transient and steady-state probabilities, key performance metrics, and numerical validations are presented to explore the system's operational dynamics and efficiency.

In todays' technologically advanced world, there can be the situations in the queueing systems where the server can avail both kinds of vacations viz. CV and WV. There is a scarcity of research articles that deals with the performance modelling of queueing systems that specifically addresses the combined concepts of CV and WV together. Furthermore, the existing such research primarily focusses on the steady state analysis. Kumar and Jain (2023a) dealt with M/M/1 queueing model with hybrid vacation policy and established the formulae for the queue size distribution via matrix geometric approach. Kumar and Jain (2023b) presented a study on the M/M/1 queueing system having provision of bi-level vacation policy consisting of CV and WV. Divya and Indhira (2024) incorporated CV and WV in a queueing model with feedback customers and unreliable server. They provided various analytical formulae for the performance measures and evaluated them numerically to predict the operating behavior of the system.

TABLE 1
IMPORTANT RELEVANT CONTRIBUTIONS RELATED TO PRESENT INVESTIGATION

Authors	Specific features				
	Transient analysis	Complete vacation (CV)	Working vacation (WV)	Impatient customers	Cost Optimization
Kumar and Jain (2023a)	✗	✓	✓	✗	✓
Kumar and Jain (2023b)	✗	✓	✓	✗	✓
Deepa and Azhagappan (2021)	✓	✓	✗	✗	✗
Suranga Sampath and Liu (2020)	✓	✓	✗	✓	✗
Ammar (2017)	✓	✓	✗	✓	✗
Sudhesh et al. (2017)	✓	✗	✓	✓	✗
Yu et al. (2013)	✗	✗	✓	✓	✗
Ye and Liu (2017)	✗	✓	✓	✗	✗
Proposed Model	✓	✓	✓	✓	✓

The effect of customer impatience on performance metrics in queueing models can be observed in many congestion situations. Customers leaving before receiving service leads to both economic losses and a decline in the organization's reputation. Therefore, understanding customer impatience is crucial for improving the revenue and fostering customer trust. Ammar (2015) analyzed a Markovian queue using modified Bessel's function, incorporating vacation and impatient customers, and provided expressions for transient probabilities, expectations, and variances of jobs in the system. Vijaya Laxmi and Kassahun (2020) explored an M/M/c queueing system with variant working vacations, focusing on customer reneging during working vacation periods, and derived steady-state probabilities, performance metrics, and cost optimization using the quadratic fit search method. Jain and Singh (2020) conducted transient analysis of a Markovian feedback queue with disaster failure and customer impatience, utilizing PGF and CF. Jain et al. (2022) examined the impact of balking behavior in a Markovian queue with imperfect service and retrial attempts, employing PGF to derive analytical expressions. Kumar and Jain (2023a) proposed an M/M/1 queueing model incorporating a bi-level vacation policy and a balking process. The study derived matrix-form expressions for the queue size distributions and formulated an optimization framework for decision variables. Numerical simulations and sensitivity analyses were conducted to validate the theoretical results. The findings highlight the model's practical relevance, with conclusions outlining its novel features and potential avenues for future research.

The transient analytical results of the queueing system provide deeper insights compared to the corresponding steady-state outcomes. Analyzing transient queueing models enables more precise predictions of system behavior over time, although deriving transient solutions proves to be more complex than steady-state solutions. The scarcity of research in the transient analysis of queueing systems stems from the inherent challenges of obtaining closed-form analytical solutions. A thorough literature review uncovers a gap in studies addressing the transient behavior of queueing models that incorporate state-dependent arrival rates, CV, WV, and customer impatience. This study aims to bridge that gap by focusing on the transient analysis of a single-server Markovian queueing model, accounting for customer reneging and the option for the server to choose between a complete vacation (CV) and a working vacation (WV) when the system is empty. The proposed model has significant applications in performance modeling and cost optimization across diverse sectors such as banking, service industries, manufacturing, hospital management, cloud computing, etc. For instance, in the smartphone assembly industry, a robotic arm responsible for battery assembly may require either complete or partial maintenance after a busy period. During complete maintenance, the robotic arm remains idle, while it operates at a reduced speed during partial maintenance. Additionally, managers may become impatient due to the slower service rate, leading to jobs reneging from the system.

The primary objective of this study is to examine the queueing model under transient conditions, providing an in-depth analysis of its behavior and dynamics. To achieve this, we employ the methods of continued fraction (CF) and probability generating function (PGF) to derive transient analytical results. These methods complement each other, offering computational efficiency and mathematical tractability for solving the complex queueing models. The CF approach is particularly useful for representing transient probabilities, especially when three-term recurrence relations arise in the governing equation, enabling rapid and accurate convergence. The PGF method is utilized to succinctly represent the probability distribution of the number of customers in the system over time by transforming the problem into a more manageable algebraic form, which enhances the understanding of transient behavior in the queue size distribution. Table 1 summarizes the contributions of the present study and compares them to other relevant articles. The article is structured as follows: Section 2 presents the foundational assumptions and notations used to develop the mathematical model, along with practical justifications for the model choice. Section 3 derives expressions for the transient probabilities and performance measures under specific scenarios. Section 4 formulates the system metrics. Section 5 provides numerical simulations and cost optimization analyses to validate the theoretical findings. Finally, Section 6 discusses key insights and explores future research directions.

MODEL DESCRIPTION AND ASSUMPTIONS

This section presents the development of a Markovian queueing model incorporating impatient customers, where the server has the option to select either CV or WV when the system becomes empty. The objective of the study is to investigate the transient behavior of the system. Figure 1 provides the transition diagram, depicting the inflow and outflow rates for each state of the system. The assumptions underlying the formulation of the model in the transient state are as follows:

- Customers arrive at the system following a Poisson process, with varying arrival rates during different states: λ during normal busy (NB) states, λ_v during complete vacation (CV) states, and λ_w during working vacation (WV) states.
- The service process is exponentially distributed, with service rates μ and μ_w during NB and WV periods, respectively. The service rate during the WV period is assumed to be slower than during NB (*i.e.* $\mu > \mu_w$).
- Upon completing the last pending customer, the server may opt for either a CV or WV. The probability of entering a CV is denoted by δ, while the probability of entering a WV is $\bar{\delta} = 1-\delta$. The duration of each vacation type follows an exponential distribution, with respective means η_v^{-1} and η_w^{-1} for CV and WV.

- Customers may abandon the system during the WV period due to the slower service rate. This abandonment, or reneging, occurs if the customers' patience expires, with the impatience rate represented by θ, following an exponential distribution.
- The queueing discipline is first-come-first-served (FCFS), meaning customers are served in the order of their arrival without prioritization or preemption.
- Transitions between states are governed by rates influenced by both the number of customers in the system and the current state of the server. These rates are illustrated in the transition diagram (Figure 1), which outlines the in-flow and out-flow rates for each system state.

Let $\xi(t)$ represent the number of customers in the system and $\gamma(t)$ denotes the server's state at the time *t*. To define the system's behavior, the bivariate stochastic process $\{\xi(t),\gamma(t),t\geq 0\}$ is utilized to specify the system's states. Now, denote the various modes of the server at time *t* as,

$$\gamma(t)=\begin{cases}0;\text{ WV mode,}\\1;\text{ NB mode,}\\2;\text{ CV mode}\end{cases}.$$

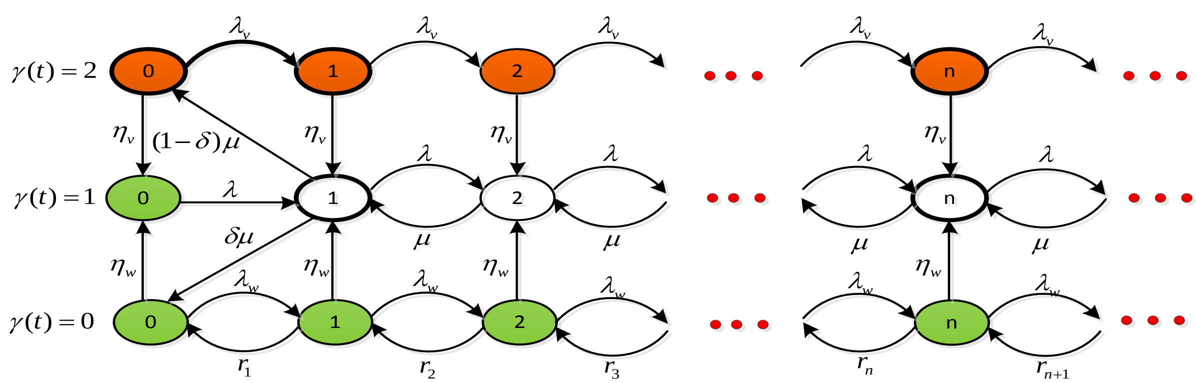


FIGURE 1
TRANSITION DIAGRAM FOR M/M/1 HYBRID VACATION MODEL WITH RENEGING

TRANSIENT PROBABILITIES

For the state space $\Omega=\{(k,n)/k=0,1,2;n=0,1,2,...\}$ of Markov chain $\{\gamma(t),\xi(t),\, t\geq 0\}$, we denote the various state probabilities for $n\geq 0$, as follows:

$$P_n(t)=P\{\gamma(t)=0,\xi(t)=n\},\ Q_n(t)=P\{\gamma(t)=1,\xi(t)=n\},\ \ R_n(t)=P\{\gamma(t)=2,\xi(t)=n\}$$

We denote Laplace transform (LT) of $P_n(t)$, $Q_n(t)$ and $\mathrm{R}_n(t)$ by $\hat{P}_n(s)$, $\hat{Q}_n(s)$ and $\hat{\mathrm{R}}_n(s)$, respectively.

For brevity of notations, we denote $r_n=\mu_w+n\theta, \Lambda_w=\lambda_w+\eta_w, \Lambda_v=\lambda_v+\eta_v$.

The forward Kolmogorov equations for the system are formulated as,

$$P_0'(t)=-\Lambda_w P_0(t)+r_1P_1(t)+\mu\delta Q_1(t), \tag{1}$$
$$P_n'(t)=-(\Lambda_w+r_n)P_n(t)+\lambda_w P_{n-1}(t)+r_{n+1}P_{n+1}(t),\ n\geq 1, \tag{2}$$
$$Q'_0(t)=-\lambda Q_0(t)+\eta_w P_0(t)+\eta_v R_0(t), \tag{3}$$
$$Q'_n(t)=-(\lambda+\mu)Q_n(t)+\lambda Q_{n-1}(t)+\mu Q_{n+1}(t)+\eta_w P_n(t)+\eta_v R_n(t),\ n\geq 1, \tag{4}$$
$$R'_0(t)=-\Lambda_v R_0(t)+\mu\bar{\delta}Q_1(t), \tag{5}$$
$$R'_n(t)=-\Lambda_v R_n(t)+\lambda_v R_{n-1}(t), n\geq 1, \tag{6}$$

The initial condition is $P_0(0)=1$.

I. Evaluation of $R_0(t)$ ***and*** $R_n(t)$

After taking LT of equation (5) and then simplifying, we get

$$\hat{R}_0(s) = \frac{\mu\bar{\delta}}{s+\Lambda_v}\hat{Q}_1(s). \tag{7}$$

Laplace inversion of (7) yields

$$R_0(t) = \mu\bar{\delta}e^{-\Lambda_v t} * Q_1(t). \tag{8}$$

Taking LT of the equation (6), and some algebraic manipulation, yields

$$\hat{R}_n(s) = \left(\frac{\lambda_v}{s+\Lambda_v}\right)^n \hat{R}_0(s), \ n \geq 1. \tag{9}$$

Laplace inversion of the above equation (9) provides

$$R_n(t) = \lambda_v{}^n e^{-\Lambda_v t}\frac{t^{n-1}}{(n-1)!} * \mu\bar{\delta}e^{-\Lambda_v t} * Q_1(t), \ n \geq 1. \tag{10}$$

II. Evaluation of $Q_0(t)$

Taking LT on both sides of equation (3), we get

$$\hat{Q}_0(s) = \frac{\eta_w}{s+\lambda}\hat{P}_0(s) + \frac{\eta_v}{s+\lambda}\hat{R}_0(s). \tag{11}$$

Laplace inversion of equation (11) yields

$$Q_0(t) = \eta_w e^{-\lambda t} * P_0(t) + \eta_v e^{-\lambda t} * R_0(t). \tag{12}$$

III. Evaluation of $Q_n(t)$

Denote the PGF of $Q_n(t)$ by

$$\mathrm{G}(t,z) = \sum_{n=1}^{\infty} Q_n(t) z^n. \tag{13}$$

By employing equation (13), equation (4) can be expressed as

$$\frac{\partial G(t,z)}{\partial t} = \Lambda(z)G(t,z) + \eta_w \sum_{n=1}^{\infty} P_n(t)z^n + \eta_v \sum_{n=1}^{\infty} R_n(t)z^n + \mu\left(1-\frac{1}{z}\right)Q_0(t). \tag{14}$$

where $\Lambda(z) = \dfrac{\lambda z^2 - (\lambda+\mu)z + \mu}{z}$.

From equation (14), we get

$$G(t,z) = \int_0^t \sum_{n=1}^{\infty}\left[\eta_w P_n(\tau) + \eta_v R_n(\tau)\right]z^n \exp\{-\alpha(\tau)+\beta(\tau)\}\,d\tau + \mu\left(1-\frac{1}{z}\right)\int_0^t Q_1(\tau)\exp\{-\alpha(\tau)+\beta(\tau)\}\,d\tau, \tag{15}$$

where $\alpha(\tau) = (\lambda+\mu)(t-\tau)$ and $\beta(\tau) = \left(\lambda z + \mu z^{-1}\right)(t-\tau)$.

Denoting $q = 2\sqrt{\lambda\mu}$, $r = \sqrt{\dfrac{\lambda}{\mu}}$ and the modified Bessel's function of the first kind of order n by $I_n(\cdot)$, we have

$$e^{\left[\left(\lambda z + \mu z^{-1}\right)t\right]} = \sum_{n=-\infty}^{\infty} (rz)^n I_n(qt). \tag{16}$$

For $n > 0$, the coefficients of z^n of equation (15) with the help of equation (16), yields.

$$Q_n(t)=\int_0^t\sum_{m=1}^{\infty}\left[\eta_w P_m(\tau)+\eta_v R_m(\tau)\right]r^{n-m}I_{n-m}(q(t-\tau))e^{-\alpha(\tau)}d\tau+\mu\int_0^t Q_0(\tau)r^n\left[I_n(q(t-\tau))-rI_{n+1}(q(t-\tau))\right]e^{-\alpha(\tau)}d\tau. \quad (17)$$

Using the identity $I_{-n}(\cdot)=I_n(\cdot)$, for negative values of n, we obtain

$$0=\int_0^t\sum_{m=1}^{\infty}\left[\eta_w P_m(\tau)+\eta_v R_m(\tau)\right]r^{-n-m}I_{n+m}(q(t-\tau))e^{-\alpha(\tau)}d\tau+\mu\int_0^t Q_0(\tau)r^{-n}\left[I_n(q(t-\tau))-rI_{n-1}(q(t-\tau))\right]e^{-\alpha(\tau)}d\tau. \quad (18)$$

Using equations (17) and (18), we obtain

$$\begin{aligned}Q_n(t)&=\int_0^t\sum_{m=1}^{\infty}\left[\eta_w P_m(\tau)+\eta_v R_m(\tau)\right]r^{n-m}\left[I_{n-m}(q(t-\tau))-I_{n+m}(q(t-\tau))\right]e^{-\alpha(\tau)}d\tau\\&\quad+\mu\int_0^t Q_0(\tau)r^{n+1}\left[I_{n-1}(q(t-\tau))-I_{n+1}(q(t-\tau))\right]e^{-\alpha(\tau)}d\tau,\quad n\geq 1.\end{aligned} \quad (19)$$

Equation (19) provides an expression for $Q_n(t)$ in terms of $P_n(t)$, $R_n(t)$ and $Q_0(t)$, for $n=1,2,3,\ldots$.

IV. Evaluation of $P_n(t)$

Laplace transforms of equations (1) and (2) yield

$$\hat{P}_0(s)=\frac{1}{(s+\Lambda_w)-r_1\dfrac{\hat{P}_1(s)}{\hat{P}_0(s)}-\mu\delta\dfrac{\hat{Q}_1(s)}{\hat{P}_0(s)}}. \quad (20)$$

and $$\frac{\hat{P}_n(s)}{\hat{P}_{n-1}(s)}=\frac{\lambda_w}{(s+\Lambda_w+r_n)-r_{n+1}\dfrac{\hat{P}_{n+1}(s)}{\hat{P}_n(s)}},\ n\geq 1. \quad (21)$$

Solving equation (21), we get (see, Lorentzen and Waadeland 1992)

$$\hat{P}_n(s)=\left(\lambda_w\theta^{-1}\right)^n\frac{1}{\prod_{i=1}^{n}\left(M(s)+i\right)}\frac{{}_1F_1\left(r_n\theta^{-1}+1;M(s)+n+1;-\lambda_w\theta^{-1}\right)}{{}_1F_1\left(r_1\theta^{-1};M(s)+1;-\lambda_w\theta^{-1}\right)}\hat{P}_0(s), \quad (22)$$

$$\hat{P}_n(s)=\hat{\phi}_n(s)\hat{P}_0(s), \quad (23)$$

where

$$M(s)=\frac{s+\eta_w+\mu_w}{\theta}$$

$$\hat{\phi}_n(s)=\left(\lambda_w\theta^{-1}\right)^n\frac{1}{\prod_{i=1}^{n}\left(M(s)+i\right)}\frac{{}_1F_1\left(r_n\theta^{-1}+1;M(s)+n+1;-\lambda_w\theta^{-1}\right)}{{}_1F_1\left(r_1\theta^{-1};M(s)+1;-\lambda_w\theta^{-1}\right)}. \quad (24)$$

Laplace inversion of equation (23) yields

$$P_n(\mathrm{t})=\phi_n(\mathrm{t})*P_0(\mathrm{t}). \quad (25)$$

V. Evaluation of $Q_1(t)$

For $n=1$, Laplace transform of equation (19) yields

$$\hat{Q}_1(s)=\frac{1}{\mu}\sum_{m=1}^{\infty}\left[\eta_w\hat{P}_m(s)+\eta_v\hat{R}_m(s)\right]\left(\hat{A}(s)\right)^m+\frac{\lambda}{\mu}\hat{A}(s)\hat{Q}_0(s), \quad (26)$$

where $d = s+\lambda+\mu$ and $\hat{A}(s) = \left(\dfrac{d-\sqrt{d^2-q^2}}{qr}\right)$. (27)

Simplifying equation (26) and using equations (7), (9), (11) and (23), we get

$$\hat{Q}_1(s) = \left[\frac{\eta_w}{\mu}\sum_{m=1}^{\infty}\hat{\phi}_m(s)\left(\hat{A}(s)\right)^m + \frac{\lambda\eta_w}{\mu(s+\lambda)}\hat{A}(s)\right]\times\sum_{k=0}^{\infty}\sum_{i=0}^{k}\binom{k}{i}\left(\frac{\lambda}{s+\lambda}\right)^{k-i}\left(\frac{\eta_v\bar{\delta}}{s+\Lambda_v}\right)^k$$

$$\times\left(\hat{A}(s)\right)^{k-i}\left[\sum_{m=1}^{\infty}\left(\frac{\lambda_v}{s+\Lambda_v}\right)^m\left(\hat{A}(s)\right)^m\right]^i\hat{P}_0(s). \quad (28)$$

Laplace inversion of equation (28) yields

$$Q_1(t) = \left[\frac{\lambda\eta_w}{\mu}\sum_{m=1}^{\infty}\phi_m(t)*\frac{1}{r^{m+1}}e^{-(\lambda+\mu)t}\left[I_{m-1}(qt)-I_{m+1}(qt)\right]+\frac{\lambda^2\eta_w}{\mu r^2}e^{-\lambda t}*e^{-(\lambda+\mu)t}\left[I_0(qt)-I_2(qt)\right]\right]$$

$$*\sum_{k=0}^{\infty}\sum_{i=0}^{k}\left(\lambda r^{-1}\right)^{k-i+1}(\eta_v\bar{\delta})^k\binom{k}{i}e^{-\lambda t}\frac{t^{k-i-1}}{(k-i-1)!}*e^{-\Lambda_v t}\frac{t^{k-1}}{(k-1)!}*e^{-(\lambda+\mu)t}\left[I_{k-i-1}(qt)-I_{k-i+1}(qt)\right]$$

$$*\left[\sum_{m=1}^{\infty}\frac{\lambda\lambda_v^m}{r^{m+1}}e^{-\Lambda_v t}\frac{t^{m-1}}{(m-1)!}*e^{-(\lambda+\mu)t}\left[I_{m-1}(qt)-I_{m+1}(qt)\right]\right]^{*i}*P_0(t). \quad (29)$$

VI. Evaluation of $P_0(t)$

For $n=1$, using equations (23) and (28) in equation (20) and performing some algebraic simplifications, we get

$$\hat{P}_0(s) = \frac{1}{(s+\Lambda_w)}\sum_{n=0}^{\infty}\sum_{j=0}^{n}\binom{n}{j}\left(\frac{r_1}{s+\Lambda_w}\right)^{n-j}\hat{\phi}_1^{n-j}(s)\left(\frac{\mu\delta}{s+\Lambda_w}\right)^j\times\left[\frac{\eta_w}{\mu}\sum_{m=1}^{\infty}\hat{\phi}_m(s)\left(\hat{A}(s)\right)^m+\frac{\lambda\eta_w}{\mu(s+\lambda)}\hat{A}(s)\right]^j$$

$$\times\left[\sum_{k=0}^{\infty}\sum_{i=0}^{k}\binom{k}{i}\left(\frac{\lambda}{s+\lambda}\right)^{k-i}\left(\frac{\eta_v\bar{\delta}}{s+\Lambda_v}\right)^k\left(\hat{A}(s)\right)^{k-i}\left[\sum_{m=1}^{\infty}\left(\frac{\lambda_v}{s+\Lambda_v}\right)^m\left(\hat{A}(s)\right)^m\right]^i\right]^j. \quad (30)$$

On taking Laplace inversion of the equation (30), we get

$$P_0(t) = e^{-\Lambda_w t}*\sum_{n=0}^{\infty}\sum_{j=0}^{n}\binom{n}{j}r_1^{n-j}e^{-\Lambda_w t}\frac{t^{n-j-1}}{(n-j-1)!}*\phi_1^{n-j}(t)*(\mu\delta)^j e^{-\Lambda_w t}\frac{t^{j-1}}{(j-1)!}$$

$$*\left[\frac{\lambda\eta_w}{\mu}\sum_{m=1}^{\infty}\phi_m(t)*\frac{1}{r^{m+1}}e^{-(\lambda+\mu)t}\left[I_{m-1}(qt)-I_{m+1}(qt)\right]+\frac{\lambda^2\eta_w}{\mu r^2}e^{-\lambda t}*e^{-(\lambda+\mu)t}\left[I_0(qt)-I_2(qt)\right]\right]^{*j}$$

$$*\left[\sum_{k=0}^{\infty}\sum_{i=0}^{k}\left(\lambda r^{-1}\right)^{k-i+1}(\eta_v\bar{\delta})^k\binom{k}{i}e^{-\lambda t}\frac{t^{k-i-1}}{(k-i-1)!}*e^{-\Lambda_v t}\frac{t^{k-1}}{(k-1)!}*e^{-(\lambda+\mu)t}\left[I_{k-i-1}(qt)-I_{k-i+1}(qt)\right]\right.$$

$$\left.*\left[\sum_{m=1}^{\infty}\frac{\lambda\lambda_v^m}{r^{m+1}}e^{-\Lambda_v t}\frac{t^{m-1}}{(m-1)!}*e^{-(\lambda+\mu)t}\left[I_{m-1}(qt)-I_{m+1}(qt)\right]\right]^{*i}\right]^{*j}. \quad (31)$$

VII. Evaluation of $\phi_n(t)$

The definition of CHF and equation (24) together yields

$$\frac{{}_1F_1\left(\frac{r_{n+1}}{\theta};\frac{s+\eta_w+r_{n+1}}{\theta};-\frac{\lambda_w}{\theta}\right)}{\prod_{i=1}^{n}\left(\frac{s+\eta_w+r_i}{\theta}\right)}=\sum_{\varepsilon=0}^{\infty}\frac{\prod_{j=1}^{\varepsilon}r_{n+j}}{\prod_{i=1}^{n+\varepsilon}\left(s+\eta_w+r_i\right)}\frac{(-\lambda_w)^{\varepsilon}}{\theta^{k-n}\varepsilon!}. \tag{32}$$

Equation (32) can be resolved into partial fraction to obtain

$$\frac{{}_1F_1\left(\frac{r_{n+1}}{\theta};\frac{s+\eta_w+r_{n+1}}{\theta};-\frac{\lambda_w}{\theta}\right)}{\prod_{i=1}^{n}\left(\frac{s+\eta_w+r_i}{\theta}\right)}=\sum_{\varepsilon=0}^{\infty}\frac{\prod_{j=1}^{\varepsilon}r_{n+j}}{\varepsilon!}\frac{(-\lambda_w)^{\varepsilon}}{\theta^{2k-1}}\times\sum_{i=1}^{n+\varepsilon}\frac{(-1)^{i-1}}{(i-1)!(n+\varepsilon-i)!}\frac{1}{s+\eta_w+r_i}. \tag{33}$$

Further,

$${}_1F_1\left(\frac{r_1}{\theta};\frac{s+\eta_w+r_1}{\theta};-\frac{\lambda_w}{\theta}\right)=\sum_{\varepsilon=0}^{\infty}\frac{\prod_{j=1}^{\varepsilon}r_j}{\prod_{i=1}^{\varepsilon}\left(s+\eta_w+r_i\right)}\frac{(-\lambda_w)^{\varepsilon}}{\theta^{k}\varepsilon!}=\sum_{\varepsilon=0}^{\infty}(-\lambda)^{\varepsilon}\hat{g}_{\varepsilon}(s),$$

where

$$\hat{g}_{\varepsilon}(s)=\frac{\prod_{j=1}^{\varepsilon}r_j}{\prod_{i=1}^{\varepsilon}\left(s+\eta_w+r_i\right)}\frac{1}{\theta^{\varepsilon}\varepsilon!};\quad \hat{g}_0(s)=1. \tag{34}$$

The partial fractions of equation (34) gives

$$\hat{g}_{\varepsilon}(s)=\frac{1}{\theta^{2\varepsilon-1}\varepsilon!}\sum_{r=1}^{\infty}\frac{\prod_{j=1}^{\varepsilon}r_j(-1)^{r-1}}{(r-1)!(\varepsilon-r)!}\frac{1}{s+\eta_w+r_r},\ for\ \varepsilon=1,2,3,\ \ldots. \tag{35}$$

By using the following identity (see Gradshteyn and Ryzhik (2007)), we obtain

$$\left[{}_1F_1\left(\frac{r_1}{\theta};\frac{s+\eta_w+r_1}{\theta};-\frac{\lambda_w}{\theta}\right)\right]^{-1}=\sum_{\varepsilon=0}^{\infty}\hat{h}_{\varepsilon}(s){\lambda_w}^{\varepsilon}, \tag{36}$$

where ${}_1F_1$ is CHF defined by

$${}_1F_1(\upsilon;\omega;z)=\sum_{l=0}^{\infty}\frac{(\upsilon)_l}{(\omega)_l}\frac{z^l}{l!},$$

Here z, υ and ω are complex and $(\upsilon)_l$, known as Pochhammer symbol, given as

$$(\upsilon)_l=\begin{cases}1, & for\ l=0\\ \upsilon(\upsilon+1)(\upsilon+2)\ldots(\upsilon+l-1), & for\ l\geq 1.\end{cases}$$

Also $\hat{h}_0(s)=1$ and for $\varepsilon=1,2,3,\ \ldots$, we have

$$\hat{h}_{\varepsilon}(s)=\begin{vmatrix} \hat{\iota}_1(s) & 1 & & \cdots & & \\ \hat{\iota}_2(s) & \hat{\iota}_1(s) & 1 & \cdots & & \\ \hat{\iota}_3(s) & \hat{\iota}_2(s) & \hat{\iota}_1(s) & \cdots & & \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \hat{\iota}_{\varepsilon-1}(s) & \hat{\iota}_{\varepsilon-2}(s) & \hat{\iota}_{\varepsilon-3}(s) & \cdots & \hat{\iota}_1(s) & 1 \\ \hat{\iota}_{\varepsilon}(s) & \hat{\iota}_{\varepsilon-1}(s) & \hat{\iota}_{\varepsilon-2}(s) & \cdots & \hat{\iota}_2(s) & \hat{\iota}_1(s) \end{vmatrix}$$

$$=\sum_{i=1}^{k}(-1)^{i-1}\hat{\iota}_i(s)\hat{h}_{\varepsilon-i}(s). \tag{37}$$

Substituting equations (28) and (30) in equation (23), we get

$$\hat{\phi}_n(s)=\lambda_w{}^n\sum_{\varepsilon=0}^{\infty}(-\lambda_w)^{\varepsilon}\frac{\prod_{j=1}^{\varepsilon}r_{n+j}}{\theta^{\varepsilon}\varepsilon!}\hat{\iota}_{n+\varepsilon}(s)\sum_{i=1}^{\infty}\lambda_w{}^i\hat{h}_i(s). \tag{38}$$

Performing Laplace inversion yields

$$\phi_n(t)=\lambda_w{}^n\sum_{\varepsilon=0}^{\infty}(-\lambda_w)^{\varepsilon}\frac{\prod_{j=1}^{\varepsilon}r_{n+j}}{\theta^{\varepsilon}\varepsilon!}\varpi_{n+\varepsilon}(t)*\sum_{i=1}^{\infty}\lambda_w{}^i h_i(t), \tag{39}$$

$$\iota_{\varepsilon}(t)=\frac{1}{\theta^{2\varepsilon-1}\varepsilon!}\sum_{r=1}^{\varepsilon}\prod_{j=1}^{\varepsilon}r_j\frac{(-1)^{r-1}}{(r-1)!(\varepsilon-r)!}e^{-(\eta_w+r_r)t},\ \varepsilon\geq 1, \tag{40}$$

$$h_{\varepsilon}(t)=\sum_{i=1}^{\varepsilon}(-1)^{i-1}\iota_i(t)*h_{\varepsilon-i}(t),\ \varepsilon=2,3,4,\ \ldots.;h_1(t)=\iota_1(t). \tag{41}$$

Here $P_n(t)$ is expressed in terms of $P_0(t)$, for $n\geq 1$ and $P_0(t)$ is computed explicitly.

VIII. Special case

When $\eta_v=0$ and $\delta=1$, then $Q_n(t)$, $P_n(t)$ and $P_0(t)$ become

$$Q_n(t)=\eta_w\int_0^t\sum_{m=1}^{\infty}P_n(\tau)r^{n-m}\left[I_{n-m}(q(t-\tau))-I_{n+m}(q(t-\tau))\right]e^{-(\lambda_w+\mu_w)(t-\tau)}d\tau$$

$$+\mu_v\int_0^t Q_0(\tau)r^{n+1}\left[I_{n-1}(q(t-\tau))-I_{n+1}(q(t-\tau))\right]e^{-(\lambda_w+\mu_w)(t-\tau)}d\tau, \tag{42}$$

$$P_n(t)=\phi_n(t)*P_0(t), \tag{43}$$

$$P_0(t)=\sum_{k=0}^{\infty}\sum_{j=0}^{k}\sum_{i=0}^{j}\binom{k}{j}\binom{j}{i}\left(\frac{r_1}{\lambda_w\eta_w}\right)^i\lambda_w{}^j\eta_w{}^k e^{-\Lambda_w t}\frac{t^k}{k!}*e^{-\lambda_w t}\frac{t^{j-i-1}}{(j-i-1)!}*\phi_1^i(t)*\frac{\lambda_w}{r^{j-i+1}}e^{-(\lambda_w+\mu_w)t}$$

$$\left[I_{j-i-1}(q(t)-I_{j-i+1}(q(t))\right]^{*(j-i)}*\left[\sum_{m=0}^{\infty}\frac{\lambda_w}{r^{m+1}}e^{-(\lambda_w+\mu_w)t}\left[I_{m-1}(q(t)-I_{m+1}(q(t))\right]*\phi_m(t)\right]^{*(k-j)}, \tag{44}$$

which coincides respectively with the results (3.7), (3.9) and (3.10) given in Sudhesh et al. (2017) when matching notations are $Q_n(t)=P_{n,1}(t)$, $P_n(t)=P_{n,0}(t)$, $\lambda_w=\lambda, \mu_1=\mu, \eta_w=\gamma, \mu_w=\mu_0, \theta=\xi$.

SYSTEM PERFORMANCE MEASURES AND COST FUNCTION

To evaluate the performance of the queueing system, several key metrics must be analyzed to predict its real-time behavior. This section derives important performance indicators, including the expected queue length $E\{\xi(\mathrm{t})\}$, the variance of the queue length $V\{\xi(t)\}$, the effective service rate (throughput) $TH(t)$, and the probabilities associated with different server states.

(i) The formula for the expected queue length is obtained as

$$E\{\xi(\mathrm{t})\} = \lambda_v \sum_{n=1}^{\infty}\int_0^t R_{n-1}(\varpi)d\varpi + \lambda \sum_{n=1}^{\infty}\int_0^t Q_{n-1}(\varpi)d\varpi - \mu \sum_{n=1}^{\infty}\int_0^t Q_n(\varpi)d\varpi + \lambda_w \sum_{n=1}^{\infty}\int_0^t P_{n-1}(\varpi)d\varpi - r_n \sum_{n=1}^{\infty}\int_0^t P_n(\varpi)d\varpi \tag{45}$$

(ii) The expression for the variance of the queue size is

$$V\{\xi(\mathrm{t})\} = E\{\xi^2(\mathrm{t})\} - \left[E\{\xi(\mathrm{t})\}\right]^2,$$

where

$$\begin{aligned} E\{\xi^2(\mathrm{t})\} = {} & \lambda_v \sum_{n=1}^{\infty}(2\mathrm{n}-1)\int_0^t R_{n-1}(\varpi)d\varpi + \lambda \sum_{n=1}^{\infty}(2\mathrm{n}-1)\int_0^t Q_{n-1}(\varpi)d\varpi - \mu\int_0^t Q_1(\varpi)d\varpi - \mu\sum_{n=1}^{\infty}(2\mathrm{n}+1)\int_0^t Q_{n+1}(\varpi)d\varpi \\ & -\theta\int_0^t P_1(\varpi)d\varpi + \lambda_w \sum_{n=1}^{\infty}(2\mathrm{n}-1)\int_0^t P_{n-1}(\varpi)d\varpi + \sum_{n=1}^{\infty}[\mu_w n^2 - \theta(\mathrm{n}+1)(2\mathrm{n}+1)]\int_0^t P_{n+1}(\varpi)d\varpi \end{aligned} \tag{46}$$

The values of the probabilities $R_n(t)$, $Q_n(t)$ and $P_n(t)$ are given in (10), (18) and (25), respectively.

(iii) The throughput TH(t) of the system is

$$TH(t) = \mu\sum_{n=0}^{\infty} Q_n(t) + \sum_{n=0}^{\infty} r_{n+1} P_n(t) \tag{47}$$

(iv) The probabilities that the sever being in *NB, WV* and *CV* modes respectively i.e., $P_{NB}(t)$, $P_{WV}(t)$ and $P_{CV}(t)$ are

$$P_{NB}(t) = \sum_{n=0}^{\infty} Q_n(t),\ \ P_{WV}(t) = \sum_{n=0}^{\infty} P_n(t),\ \ P_{CV}(t) = \sum_{n=0}^{\infty} R_n(t) \tag{48}$$

I. The cost function

The cost modelling of queueing model plays a pivotal role in determining the minimum operational expenses and optimal service rate, which collectively help to alleviate congestion in the queueing system at the lowest possible service cost. In this section, a cost function is formulated based on the model under analysis, considering the service rates μ and μ_w as decision variables for the effective cost optimization. The various expenses incurred per unit time in the various operations of the system are categorized as cost elements and are defined as follows:

C_H : Holding cost of each customer in the system.
C_{WV}: Cost associated with the server being in WV mode.
C_B : Cost associated with the server being in NB mode.
C_V : Cost associated with the server being in CV mode.
C_M : Service cost of each customer in NB mode.
C_{MW}: Service cost of each customer in WV mode.

Now the cost function is framed as,

$$TC(t;\mu,\mu_w) = C_H E\{\xi(t)\} + C_W P_{WV}(t) + C_B P_{NB}(t) + C_V P_{CV}(t) + C_M\mu + C_{MW}\mu_w \tag{49}$$

As the optimization cannot be conducted with respect to the time variable, it is necessary to establish a fixed value for time. We are fixing $t=2$ units, to proceed with the optimization process. Hence, the cost optimization problem associated with the cost function $TC(t;\mu,\mu_w)$ is formulated as:

$$TC(\mu^*,\mu_w^*)=\min_{(\mu,\mu_w)} TC(\mu,\mu_w) \tag{50}$$

where μ^* and μ_w^* denote the optimum values of the service rates μ and μ_w.

The optimization problem outlined in equation (50) is complex and non-linear in nature, making it unsolvable by analytical methods, thus necessitating the use of the direct search method for its solution. This method involves examining a set of points in the neighborhood of the current point to identify one that results in a lower value of the cost function compared to the current point.

NUMERICAL SIMULATION, COST OPTIMAZATION AND DISCUSSION

This paper aims to assess the real-time effectiveness of the model by conducting an extensive analysis of numerical simulations and implementing a cost-optimization framework. The results obtained from these quantitative analyses are then utilized to provide a deeper understanding of the managerial implications that arise from the study concluded in this paper.

I. Numerical simulation

To obtain the numerical simulation results, the computational tasks were carried out using the MATLAB R2021b software with the number of customers fixed at $n=8$ and for the default parameter as $\lambda=3,\lambda_w=2.5,\lambda_v=2,\mu=3.2,\mu_w=3,\eta_w=1,\eta_v=1,\delta=0.3,\theta=3.$

The impact of variation in system parameters is examined on the various system indices; the key observations for different indices are as follows:

(i) ***E{ξ(t)}:*** From Figures 2(i-vi), it is noticed that $E\{\xi(\text{t})\}$ increases as time grows up to t=9 and after that it becomes asymptotically constant and reveals the steady state behavior. By increasing the values of arrival rates $\lambda,\lambda_w,\lambda_v$ and WV termination rate η_w, the values of $E\{\xi(\text{t})\}$ increases as shown in the Figures 2(i-iii) and 2(v). On the other hand, we see that $E\{\xi(\text{t})\}$ decreases with the increments in δ and η_v which is clear from the Figures 2(iv) and 2(vi).

(ii) ***TH(t):*** It is clear from the Figures 3 (i) and 3 (iii-v) that *TH(t)* initially decreases as time grows up to a threshold point which is attained near about t=3 for all the four figures. Then after *TH(t)* increases before attaining the almost constant value. From Figures 3 (ii) and 3 (vi) we observe that *TH(t)* for $\delta=0.6$ and $\eta_v=4$, initially starts increasing rapidly up to a maximum point then it decreases and finally attains the constant value. Also, from the figure 3 (i-vi), it is observed that *TH(t)* increases with the increments in parameters $\mu,\delta,\lambda,\lambda_w,$ and η_v; on the other hand, *TH(t)* decreases with η_w.

(iii) ***P_{NB}(t), P_{WV}(t), P_{CV}(t) :*** From Tables 2 and 3 we notice that the $P_{NB}(t)$ increases with time but decrease by increasing the values of θ and μ_w; this may be due to the fact that as θ and μ_w grow, the number of customers in the queue will decrease and so the $P_{NB}(t)$. The value of $P_{WV}(t)$ decreases as time passes which is noticed from the Tables 2 and 3. Also from both the tables we conclude that $P_{WV}(t)$ is less sensitive with respect to the parameters θ and μ_w as the increment in the value of $P_{WV}(t)$ with respect to both of these parameters is negligible. From tables 2 and 3, it is noticed that $P_{CV}(t)$ follows the same trend as noticed for $P_{WV}(t)$ i.e., it decreases with time and remains almost constant with the increment in both the parameters θ and μ_w.

TABLE 2
IMPACT OF θ ON VARIOUS SYSTEM STATE PROABILITIES

θ	t	$P_{NB}(t)$	$P_{WV}(t)$	$P_{CV}(t)$
1	3	0.591017	0.122695	0.286288
	5	0.688015	0.093595	0.218389
	7	0.734211	0.079737	0.186052
	9	0.750539	0.074838	0.174623
3	3	0.590272	0.122918	0.286810
	5	0.684770	0.094569	0.220661
	7	0.730608	0.080818	0.188574
	9	0.747231	0.075831	0.176938
5	3	0.590571	0.122884	0.286730
	5	0.684369	0.094793	0.221180
	7	0.730025	0.081096	0.189220
	9	0.746684	0.076089	0.177540

TABLE 3
IMPACT OF μ_w ON VARIOUS SYSTEM STATE PROABILITIES

μ_w	t	$P_{NB}(t)$	$P_{WV}(t)$	$P_{CV}(t)$
1	3	0.589996	0.123001	0.287003
	5	0.685666	0.094300	0.220034
	7	0.731775	0.080467	0.187757
	9	0.748316	0.075505	0.176179
3	3	0.590272	0.122918	0.286810
	5	0.684770	0.094569	0.220661
	7	0.730608	0.080818	0.188574
	9	0.747231	0.075831	0.176938
5	3	0.590571	0.122829	0.286601
	5	0.684369	0.094689	0.220942
	7	0.730025	0.080993	0.188983
	9	0.746684	0.075995	0.177321

II. Cost Optimization

By direct search approach the optimization problem given by equation (50) is solved to obtain the minimum operation cost $TC(\mu^*, \mu_w^*)$ and the optimal service rates μ^* and μ_w^*. Furthermore, the effect of λ_w and θ on the optimality of the system is highlighted in Table 5. Two different cost sets (see Table 4) are chosen for the cost optimization purpose.
In Figures 4(i-iii) and 5(i-iii), the surface plot of the cost function is shown for both the cost sets. The data shown in Table 5 reveals that the cost function exhibits a minimum value \$121.08 for cost set I at optimal service rates $\mu^* = 3.0824$ and $\mu_w^* = 2.0969$. For cost set II, the minimum value of the cost function is \$=80.87, which is achieved when $\mu^* = 2.4$ and $\mu_w^* = 0.7864$.

III. Managerial Insights

This study offers significant insights to inform and guide managerial decision-making processes and have the potential applicability. Including option between complete vacation (CV) and working vacation (WV) in the service system, managers/system designers may chalk out the service strategies to meet the demand at an optimal cost. The system designers and managers can employ this understanding to balance between CV and WV periods of the server, thus improving the quality of service (QoS) and the system's efficiency. The analysis of the customer impatience behavior during the WV period underscores the significance of adopting proactive measures to alleviate customer attrition. These measures may include providing incentives or enhancing service quality. The sensitivity analysis done may be helpful in optimally designing the waiting area for the customers. The numerical findings of various system indices can be employed to make predictions regarding the in-flow of the customers based on the server's status within the service system. Therefore, by appropriately managing the system parameters, it is possible to reduce the waiting time of the jobs, and overall operational efficiency of the service system within the technological and cost restrictions.

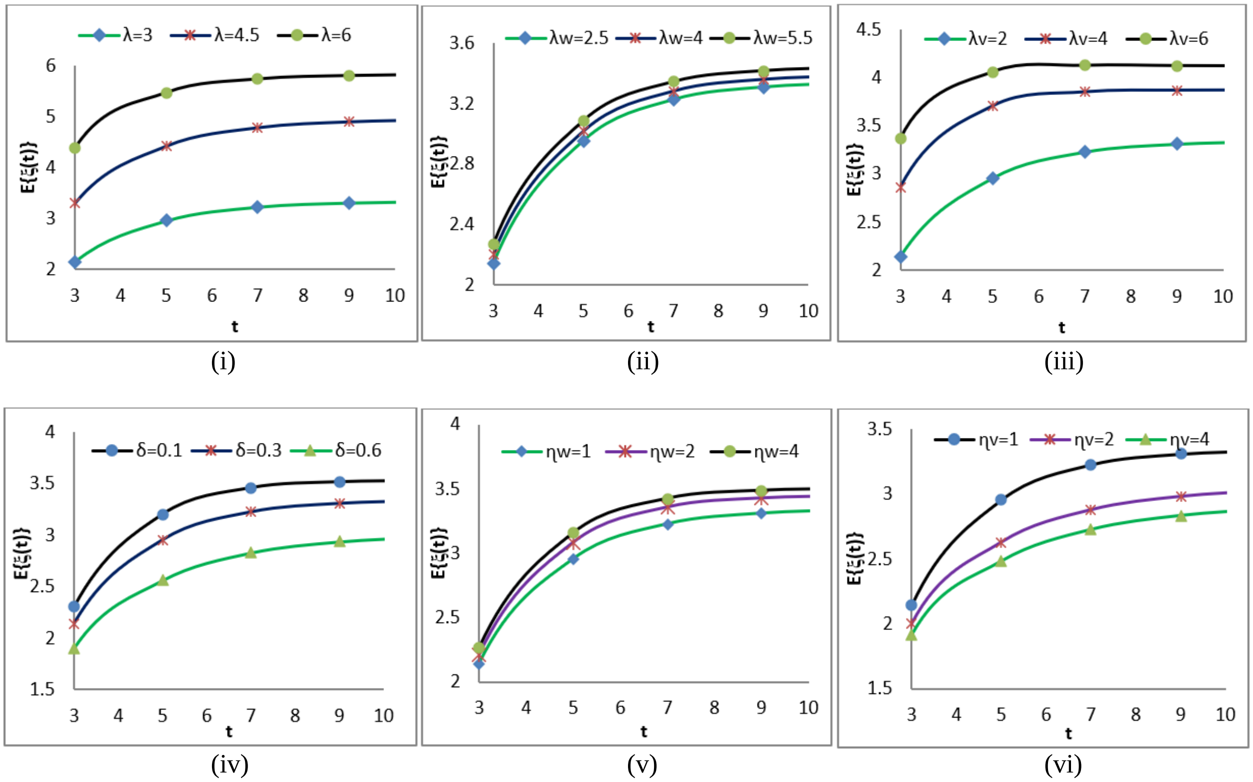


FIGURE 2

VARIATION IN $E\{\xi(\mathrm{t})\}$ WITH RESPECT TO TIME FOR VARIOUS VALUES OF (I) λ (II) λ_w (III) λ_v (IV) δ (V) η_w (VI) η_v

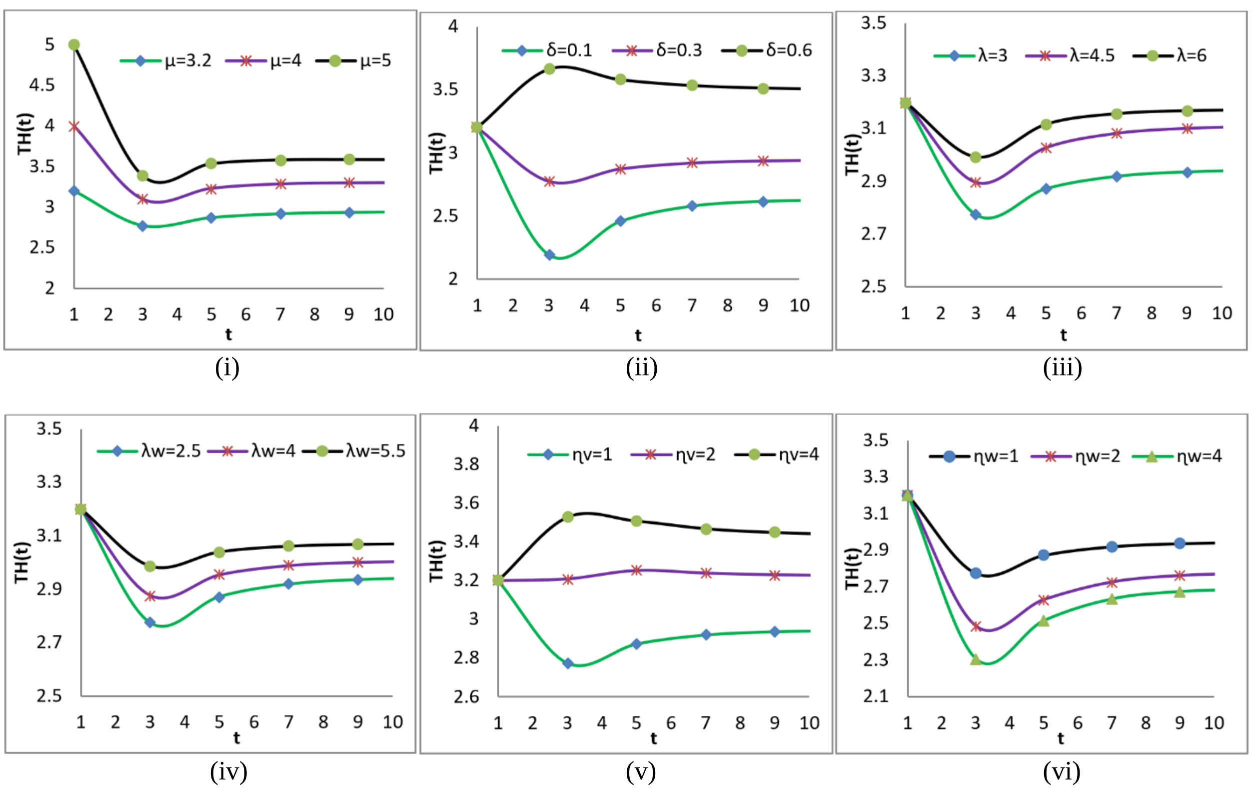
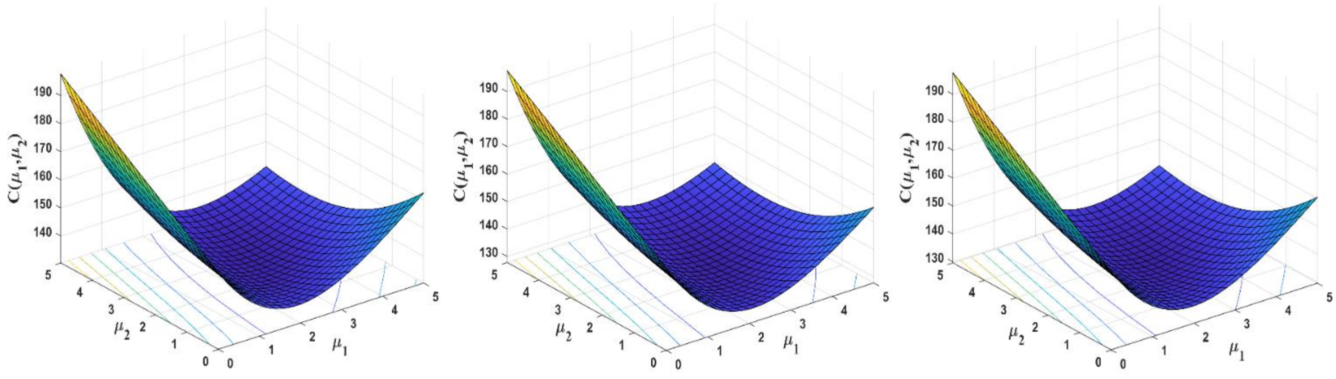


FIGURE 3

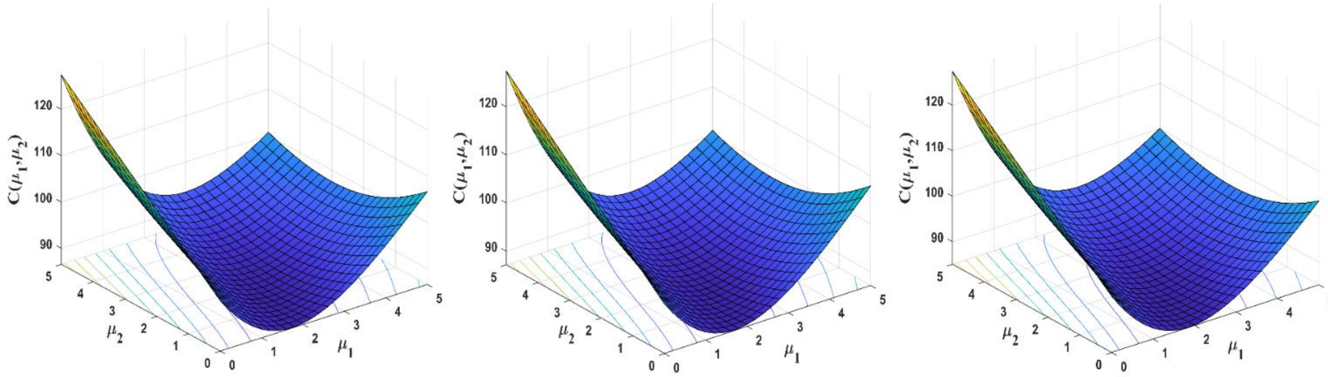
VARIATION IN *TH(T)* WITH RESPECT TO TIME FOR VARIOUS VALUES OF (I) μ (II) δ (III) λ (IV) λ_w (V) η_v (VI) η_w



(i) (ii) (iii)

FIGURE 4

TOTAL COST OF THE SYSTEM FOR COST SET I AT $\lambda_w = 2.5$ AND FOR (I) $\theta = 0.01$ (II) $\theta = 0.05$ (III) $\theta = 0.15$



(i) (ii) (iii)

FIGURE 5

TOTAL COST OF THE SYSTEM FOR COST SET II AT $\lambda_w = 2.5$ AND FOR (I) $\theta = 0.01$ (II) $\theta = 0.05$ (III) $\theta = 0.15$

TABLE 4

VARIOUS COST SETS FOR COST MINIMIZATION PROBLEM (IN $)

Cost Set I	$C_H = 25$	$C_W = 20$	$C_B = 22$	$C_V = 12$	$C_M = 10$	$C_{MW} = 8$
Cost Set II	$C_H = 15$	$C_W = 12$	$C_B = 16$	$C_V = 8$	$C_M = 8$	$C_{MW} = 6$

TABLE 5

EFFECT OF λ_w AND θ FOR BOTH THE COST SETS ON OPTIMAL VALUES OF SERVICE RATE μ^* AND μ_w^* AND COST ASSOCIATED WITH THE SYSTEM WHEN $\lambda = 3, \lambda_V = 2, \eta_W = 0.01, \eta_V = 1, \delta = 0.9$

Cost Set	(λ_w, θ)	(μ^*, μ_w^*)	$TC(\mu^*, \mu_w^*)$
Cost Set I	(2.5,0.01)	(3.0901,2.4406)	123.52
	(2.5,0.05)	(3.0880,2.3419)	122.82
	(2.5,0.15)	(3.0824,2.0969)	121.08
	(2.99,0.01)	(3.0069,2.9363)	130.43
	(2.99,0.05)	(3.0040,2.8286)	129.68
	(2.99,0.15)	(2.9964,2.5604)	127.82
Cost Set II	(2.5,0.01)	(2.4252,1.1866)	82.79
	(2.5,0.05)	(2.4185,1.0732)	82.24
	(2.5,0.15)	(2.4000,0.7864)	80.87
	(2.99,0.01)	(2.1943,1.2345)	87.17
	(2.99,0.05)	(2.1752,1.0805)	86.56
	(2.99,0.15)	(2.1032,0.6340)	85.06

CONCLUSION

Customers' impatience can significantly affect both the quality of service and the accuracy of the performance predictions in queueing systems. This impatience can lead to increased abandonment rates, which in turn affects system stability and overall service efficiency. As a result, understanding and accounting for customer behavior are crucial for optimizing queueing system performance. In many service and manufacturing/commercial systems, when the server becomes free, he may avail either CV or WV. These scenarios may encounter in telecommunication systems, programmable robotics arms, machine-dependent production, call centers, packaging systems, etc. The transient study of a Markovian queueing model, which includes the option for CV and WV, is conducted by considering the impact of customer impatience and state-dependent arrival and service rates. Based on our study, the strategic joining/reneging behavior of the customers can be predicted. Numerical simulations illustrate the feasibility of quantitatively assessing transient performance metrics which provides valuable insights into the dynamic behavior of various congestion scenarios. However, this research has certain limitations. The feature of unreliable server incorporated in the model investigated seems to be more realistic and portrays many real-world machining systems where servers are prone to failures. Moreover, the study focuses on a single-server environment, which limits its applicability to systems with multiple servers. The transient model developed focuses on a single-server environment, which can be further modified for the multiple servers. The possible extension of present work is to include Bernoulli feedback, which would make the model more realistic for many real-world systems. Moreover, the model could be extended to include bulk input or bulk service, but analysis will become very tedious.

Acknowledgements: The authors express their gratitude to the editor in chief and anonymous reviewers for their helpful and constructive feedback. Mayank Singh acknowledges the financial support provided by UGC, India through the JRF/SFR scheme (Grant code: 11-04-2016-429688) for conducting this research.

Declaration of Interest: All the author confirms that there is no conflicts of interest.

REFRENCES

[1] Doshi, B., T. (1986). Queueing systems with vacations - A survey. *Queueing Systems*, 1, 29–66. https://doi.org/10.1007/BF01149327

[2] Sikdar, K., & Gupta, U., C. (2005). The queue length distributions in the finite buffer bulk-service MAP / G / 1 queue with multiple vacations. *Top*, 13, 75–130. https://doi.org/10.1007/BF02578989

[3] Isijola-Adakeja, O., A., & Ibe, O., C. (2014). M/M/1 multiple vacation queueing systems with differentiated vacations and vacation interruptions. *IEEE Access*, 2, 1384–1395. https://doi.org/10.1109/ACCESS.2014.2372671

[4] Suranga Sampath, M., I., G., & Liu, J. (2020). Impact of customers' impatience on an M/M/1 queueing system subject to differentiated vacations with a waiting server. *Quality Technology and Quantitative Management*, 17, 125–148. https://doi.org/10.1080/16843703.2018.1555877

[5] Azhagappan, A., & Deepa, T. (2022). Transient analysis of various vacation interruption policies in a Markovian queue with differentiated multiple vacations. *International Journal of Operational Research*, 45, 492–510. https://doi.org/10.1504/IJOR.2022.128401

[6] Kumar, A. (2023). Single server multiple vacation queue with discouragement solve by confluent hypergeometric function. *Journal of Ambient Intelligence and Humanized Computing*, 14, 6411–6422. https://doi.org/10.1007/s12652-020-02467-0

[7] Servi, L., D., & Finn, S., G. (2002). M/M/1 queues with working vacations (M/M/1/WV). *Performance Evaluation*, 50, 41–52. https://doi.org/10.1016/S0166-5316(02)00057-3

[8] Li. J., & Tian, N. (2007). The M/M/1 queue with working vacations and vacation interruptions. *Journal of Systems Science and Systems Engineering*, 16, 121–127. https://doi.org/10.1007/s11518-006-5030-6

[9] Sudhesh, R., & Raj, L., F. (2012). Computational analysis of stationary and transient distribution of single server queue with working vacation. *In: Global Trends in Computing and Communication Systems. Springer*, Berlin Heidelberg, 480–489. https://doi.org/10.1007/978-3-642-29219-4_55

[10] Ma, Q-Q., Li, J-H., & Liu, W-Q. (2018). Social optimization in M/M/1 queue with working vacation and N-policy. *RAIRO - Operations Research*, 52, 439–452. https://doi.org/10.1051/ro/2017041

[11] Farahani, A., Tohidi, H., Shoja, A. (2019). An integrated optimization of quality control chart parameters and preventive maintenance using Markov chain. Advances in Production Engineering & Management. 14(1), pp 5–14. https://doi.org/10.14743/apem2019.1.307

[12] Jain, M., Rani, S., & Singh, M. (2019). Transient analysis of Markov feedback queue with working vacation and discouragement. *In: Deep K, Jain M, Salhi S (eds) Performance Prediction and Analytics of Fuzzy, Reliability and Queuing Models : Theory and Applications*, Springer Singapore, Singapore, 235–250. https://doi.org/10.1007/978-981-13-0857-4_18

[13] Tohidi, H. & Jabbari, M.M. (2012}. Measuring organizational learning capability. Procedia - Social and Behavioral Sciences, vol. 31, p.p. 428-432, 2012, https:// doi:10.1016/j.sbspro.2011.12.079

[14] Jain, M., Singh, M., & Meena, R.K. (2021). Time-dependent analytical and computational study of an m/m/1 queue with disaster failure and multiple working vacations. *In: Chadli O, Das S, Mohapatra RN, Swaminathan A (eds) Mathematical Analysis and Applications*, Springer Singapore, Singapore, 293–304. https://doi.org/10.1007/978-981-16-8177-6_21

[15] Sridhar, G., Rama Devi, V. N., & Chandan, K. (2022). Performance analysis of M/M/1 queue with working vacation, N-policy and customer impatience in transient mode. *In: AIP Conference Proceedings. American Institute of Physics Inc*. https://doi.org/10.1063/5.0108662

[16] Ramesh, A., & Udayabaskaran, S. (2023). Transient analysis of a single server queue with working vacation operating in a multi-level environment controlled by a random switch. *New Mathematics and Natural Computation*, 1–16. https://doi.org/10.1142/S1793005725500097

[17] Kumar, A., & Jain, M. (2023a). M/M/1 queue with bi-level network process and bi-level vacation policy with balking. *Communications in Statistics - Theory and Methods,* 52, 5502–5526. https://doi.org/10.1080/03610926.2021.2012197

[18] Kumar, A., & Jain, M. (2023b). 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

[19] Divya, K., & Indhira, K. (2024). Performance analysis and ANFIS computing of an unreliable Markovian feedback queueing model under a hybrid vacation policy. *Mathematics and Computers in Simulation,* 218, 403–419. https://doi.org/10.1016/j.matcom.2023.12.004

[20] Ammar, S. I. (2017). Transient solution of an M/M/1 vacation queue with a waiting server and impatient customers. *Journal of the Egyptian Mathematical Society,* 25, 337–342. https://doi.org/10.1016/j.joems.2016.09.002

[21] Sudhesh, R., Azhagappan, A., & Dharmaraja, S. (2017) Transient analysis of M/M/1 queue with working vacation, heterogeneous service and customers' impatience. *RAIRO - Operations Research,* 51, 591–606. https://doi.org/10.1051/ro/2016046

[22] Yu, X., Wu, D., & Wu, L., et al. (2013). An M/M/1 queue system with single working vacation and impatient customers. *In: Qi E, Shen J, Dou R (eds) The 19th International Conference on Industrial Engineering and Engineering Management*, Springer Berlin Heidelberg, Berlin, Heidelberg, 767–776. https://doi.org/10.1007/978-3-642-38391-5_81

[23] Ye, Q., & Liu, L. (2017). The analysis of the M/M/1 queue with two vacation policies (M/M/1/SWV+MV). *International Journal of Computer Mathematics*, 94, 115–134. https://doi.org/10.1080/00207160.2015.1091450

[24] Ammar, S. I. (2015). Transient analysis of an M/M/1 queue with impatient behavior and multiple vacations. *Applied Mathematics and Computation,* 260, 97–105. https://doi.org/10.1016/j.amc.2015.03.066

[25] Vijaya Laxmi, P., & Kassahun, T. W. (2020). Analysis of variant working vacation queue with reneging under a multi-server environment. *International Journal of Management Science and Engineering Management,* 15, 130–137. https://doi.org/10.1080/17509653.2019.1653237

[26] Jain, M., & Singh, M. (2020). Transient analysis of a Markov queueing model with feedback, discouragement and disaster. *International Journal of Applied and Computational Mathematics,* 6, 31. https://doi.org/10.1007/s40819-020-0777-x

[27] Jain, M., Dhibar, S., & Sanga, S. S. (2022). Markovian working vacation queue with imperfect service, balking and retrial. *Journal of Ambient Intelligence and Humanized Computing,* 13, 1907–1923. https://doi.org/10.1007/s12652-021-02954-y

[28] Lorentzen, L., & Waadeland, H. (1992). Continued Fractions with Applications. Studies in Computational Mathematics, vol. 3., 1st edn. North-Holland, Amsterdam

[29] Gradshteyn, I.S., & Ryzhik, I.M. (2007). Table of Integrals, Series, and Products, 7th edn. Academic Press, MA, USA