The Journal of Supercomputing (2024) 80:15429–15463
https://doi.org/10.1007/s11227-024-06068-y

Particle swarm optimization and FM/FM/1/WV retrial queues with catastrophes: application to cloud storage

Sibasish Dhibar[1] · **Madhu Jain**[1]

Accepted: 12 March 2024 / Published online: 3 April 2024

Abstract

The cloud storage service, known for its flexible and expandable nature, often has difficulties managing operating costs while ensuring dependable service and quick response times. This investigation presents a novel approach to optimizing cost efficiency in cloud storage systems by applying particle swarm optimization of the Markovian retrial queueing model in a generic setup by incorporating the working vacation and users' discouragement behavior. Some users may opt not to enter the system or join the retry pool to wait for their turn if the server is occupied. After returning from working vacation, if there is one user available for service, the server can interrupt the vacation period. The server is subject to breakdown and can be recovered after getting the repair. In the proposed model, the server is prone to catastrophes and can fail at any time, leading to the entire system breaking down, and no users being able to access it during this period. Chapman–Kolmogorov (CK) steady-state equations associated with the quasi-birth-death (QBD) process are constructed to make a mathematical design. The governing equations framed to derive the queue length distributions and various performance indices are solved using the recursive method and difference equation theory. The fuzzified parameters are used to develop the FM/FM/1/WV model, which is analyzed using a parametric nonlinear programming approach. To determine the optimal design parameters, the cost minimization problem has been done using the quasi-Newton method and particle swarm optimization. This model incorporates features such as server failures, retrials, and catastrophes, thereby reflecting the complex nature of cloud storage operations. A suitable illustration of cloud storage is taken for both classical and fuzzified models to facilitate the numerical results of performance indices and optimal decision descriptors.

Keywords Cloud storage · Retrial queue · Working vacation · Catastrophes · Fuzzy · PSO · QNM

Extended author information available on the last page of the article

1 Introduction

The swift advancement of various technologies, including cloud computing, distributed file systems, and Internet of Things (IoT) technologies, has significantly expedited the processes of data storage and computation. Consequently, intelligent IoT systems generate a significant amount of multi-source industrial data, necessitating substantial storage, and computing resources for real-time data processing and analysis. Cloud computing, with its robust computing and storage capacities, can be intricately incorporated into intelligent IoT systems. Storing these industrial data on cloud servers' aids in the continuous monitoring of industrial process conditions. As a result, data are expanding at an unprecedented rate. According to IDC forecasts, the global volume of data is expected to hit 175 zettabytes which was suggested by Xiong et al. [1]. Cloud storage systems are increasingly being used to hold vast quantities of unstructured data, such as information originating from the Internet of Things (IoT). The IoT effectively leverages contemporary information technology and sophisticated communication technology to create a range of intelligent service systems. Nonetheless, conventional relational databases fall short in maintaining high availability and performance when dealing with the extensive storage and multi-dimensional querying of IoT data, Bjeladinovic et al. [2]. Consequently, various non-relational databases that are based on cloud storage systems have gained widespread use in domains related to the IoTs, with HBase being a prominent example. Therefore, cloud storage plays a crucial role in queueing models, significantly enhancing file transfer efficiency.

In the queueing systems, waiting in the queue is a common scenario in various spheres of our daily lives such as call centers, railway ticket counters, restaurant, petrol stations, banks, shopping malls, and many more. In the congestion situations when the service providers are occupied, the customers may have to wait in the retrial orbit and come back later to retry for the service. Krishnamoorthy et al. [3] conducted an analysis of a single-server queueing model with retrial attempts and interruptions in the service. Gao et al. [4] proposed a joining strategy for Markovian system featuring with constant retrials and impatient customers. Jain and Dhibar [5] studied transient behavior of a single-server Markovian queue with retrial phenomena and imperfect service. Zhang and Wang [6] developed a logit choice model to capture customers' behaviors and examined how waiting time estimation errors affect the customers' decision-making. Shi and Liu [7] studied the joining strategies for an M/M/1 retrial queueing model based on priority features. The retrial policy plays a significant role in queuing scenarios, especially in the application of communication systems, and considerable work has been done. However, research on the impact of retrial policy in cloud storage applications is still lacking.

When the system suffers from random catastrophe, the servers may fail suddenly, and all the waiting (including in the service) customers are lost. In such system after catastrophic failure, the server returns to idle state after the repair. Numerous researchers [8–12] have focused their attention on studying the retrial

queues with unreliable servers. Bura [13] studied Markovian queueing system with catastrophes and infinite servers. Jain et al. [14] presented a literature survey on the queueing systems with an unreliable server and service interruptions. Li and Wang [15] conducted performance analysis of an M/M/1 retrial queueing system with Poisson-generated catastrophes and server breakdowns that result in the removal of all the customers in the system. Recently, M/M/1 queue with catastrophes was investigated by Souza and Rodriguez [16] to derive the state probabilities, mean and variance of the number of customers present in the system by using fractional derivative approach. Danilyuk et al. [17] studied a retrial queueing system by developing unreliable server M/M/1 model which incorporates the features of impatient customers and collisions. However, there has not been any researcher who has focused on examining the Markov model involving catastrophes server of infinite capacity retrial queue while incorporating the notions of reneging and joining strategy.

Several types of vacations are studied in queueing article to study the behavior of queueing models with vacationing server. These included vacations taken after completion of a busy period, scheduled vacations, working vacations, Bernoulli vacation, etc. Servi and Finn [18] have introduced the concept of working vacation (WV) for Markovian single-server queue with semi-vacation during which the service is rendered with slower rate in place of fully stopping the service. The WV queueing models in different contexts under a variety of features have been well studied in the past two decades. Li and Li [19] investigated Markovian model for single-server queue with WV and customers' discouragement policy. They also established the stochastic decomposition results. Ameur et al. [20] investigated performance measures of the M/M/1 retrial queue with WV, vacation interruption, and finite orbit size. They provided the sensitivity analysis to identify key parameters affecting the performance indices. Do et al. [21] studied the M/M/1 retrial queue with working vacations and a fixed retrial rate under the assumption that the customers decide about the entry based on the information upon their arrival instants. Kumar and Jain [22] analyzed an M/M/1 queueing model with WV that incorporates the bi-level network process and bi-level vacation policy. This model includes two types of customers and two types of vacation periods to deal a more realistic representation of real-world queueing scenarios. Jain et al. [23] investigated a Markovian queueing system having features of imperfect service during WV, retrial, and impatient customers. Muthusamy et al. [24] proposed a WV queueing model with retrial attempts by categorizing the customers into three classes, namely, regular, priority, and disaster. Recently, Dhibar and Jain [25] presented strategic behavior for the double orbit retrial queue with imperfect service and WV interruption. In the real world, the WV policy plays a crucial role in queuing systems, and a lot of research related to WV has been conducted. However, the application of file transfer into cloud storage, facilitated by WV and characterized by slower transfer speeds due to network constraints, has not been investigated until now.

Sometimes, in queueing scenarios, the system parameters are not expressed precisely, i.e., more suitably expressed in linguistic terms rather than by classical numbers. However, in real-world scenarios, the performance metrics may not be accurately evaluated or may fluctuate over time due to factors such as customer behavior

and uncertainty in arrivals. Consequently, it is imperative to address performance metrics uncertainty in a realistic scenario and adaptable manner. Fuzzy set theory offers a mathematical framework that utilizes linguistic terms and fuzzy numbers to manage imprecise information. The classical queueing model characterized by classical parameters seems to be more practical if it is extended to fuzzified queueing models in terms of fuzzy numbers. The pioneer researcher, who has introduced the fuzzy set theory, was Zadeh [26]. There is a few research papers in the queueing literature which dealt with fuzzy queueing models. Li and Lee [27] studied M/F/1 and FM/FM/1 models based on Zadeh's extension principle. Negi and Lee [28] investigated M/F/1 fuzzy queue which reduced to M/G/1 queue through various α -cuts level sets. In queueing literature, notable research articles which dealt with a variant of fuzzy queueing model under different assumptions are done by Chen [29], Pardo and de la Fuente [30], Chen et al. [31], Jain et al. [32], and others. Sanga and Jain [33] studied a strategic joining policy for the FM/FM/1 fuzzy queueing model and established performance measures using P-NLP approaches. Sanga and Jain [34] conducted a study that focuses on a fuzzy double orbit queueing model by incorporating the features of retrial orbit and discouragement. They derived various performance metrics for both classical and fuzzy environments. The classical single-server queueing model with retrial attempts can be expanded to encompass a fuzzy single-server retrail queueing model with customers exhibiting strategic behavior. This expanded investigation holds promise for broader applications in real-time systems and stands to benefit decision-makers by enabling the provision of enhanced services for which customers are willing to pay.

There are specific optimization methods available in queueing literature to find the cost optimization/profit amximization by setting the optimal parameter values. The particle swarm optimization (PSO) method is a metaheuristic evolutionary optimization technique and can also be used for evaluating the optimal parameters of queueing systems. The PSO was initiated by Eberhart and Kennedy [35]. Wang et al. [36] investigated a social maximization problem in Markovian single-server retrial queueing system operating under threshold N-policy. Zhou et al. [37] extended the research work of Wang et al. [36] by considering the setup time. Wang et al. [38] presented the equilibrium and socially optimal balking strategies for the customers who arrive at the retrial queueing system operating under N-policy. They improved PSO algorithm to visualize the impact of system parameters on the profit of service providers. Sumathi and Manivannan [39] proposed a stochastic model for channel selection in cognitive radio (CR) system. They used metaheuristic technique to optimize the time for data delivery in CR. Meena et al. [40] did the performance analysis of a fault tolerant machine repair system with vacationing server. Optimal service rates were obtained by employing both genetic algorithm (GA) and PSO techniques. Jain and Dhibar [41] studied strategic behavior in M/M/1 retrial queues associated with a working vacation policy, wherein the service parameters are optimized through the PSO and quasi-Newton method (QNM).

Queueing problems with different kinds of catastrophes can be noticed in various circumstances in a real-life scenario. For example, sometimes mobile network transmits significantly weak signal; therefore, the clients might be failed to transmit the packets due to network failure (catastrophes), and all packets are lost. In case, if

channel is busy, the arriving call can wait in the virtual retrial orbit. Due to impatience while waiting in the orbit, some users may leave (i.e., renege) while others may retry to get service. For the illustration purpose, we give an example related to catastrophes which may arise due to the virus in a machine (e.g., laptop, Android phone, etc.). Due to virus attack, software embedded system or computer becomes slow and finally crashes. Therefore, all jobs may be lost while waiting in the service system. The arrival of jobs starts only when the machine resumes its functionality after recovery from faults/shocks.

Table 1 summarizes the noble features incorporated in the recent relevant research articles and the proposed model. According to the best of authors' knowledge, no researcher has investigated the fuzzified retrial queueing system with WV, discouragement, and catastrophes. The main research work of the present studies invetigates the classical model and extended this model into fuzzy environment. We investigate the balking strategies for the customers who intend to get service from a single server who is capable to serve the customers with slow rate during working vacation period also. In order to ascertain the probability distribution of performance metrics, the non-homogeneous linear difference equations governing the model are formulated. Several performance indices are derived to analyze the system characteristics and further employed to investigate the sensitivity of system parameters via numerical simulation. Metaheuristic optimization techniques have been employed to determine optimal service rates and corresponding expected optimal cost. In the present study, the cost minimization using PSO and QNM approach for finding the optimum service parameters has been proposed. The noble contributions of proposed work are as follows:

(i). The steady-state distributions for the system size and other performance metrics for the concerned working vacation retrial queue are established by incorporating the disaster failure of the server and user's discouragement.
(ii). For the fuzzified model, performance metrics are obtained as explicit form's using α-cuts approach which is computationally tractable as verified by numerical simulation results.

Table 1 Summary of recent research works related to proposed model

Authors	WV	Retrial	Balking	Reneging	Disaster	Fuzzy model	Optimization
Li et al. [42]	√	√	×	×	×	×	×
Ameur et al. [20]	√	√	×	×	×	√	×
Yang and Wu [43]	√	√	×	√	×	×	PSO
Do et al. [21]	√	√	√	×	×	×	×
Jain and Sanga [44]	×	√	√	×	√	×	QNM
Jain and Rani [12]	×	√	√	√	×	×	×
Lakaour et al. [45]	×	√	×	√	√	×	×
Danilyuk et al. [17]	×	√	√	×	√	×	×
Jain et al. [23]	√	√	√	×	×	×	GA
Proposed model	√	√	√	√	√	√	PSO, QNM

(iii). The cost optimization process is executed using a metaheuristic approach, specifically the PSO technique, and a numerical strategy, namely, the QNM.
(iv). A practical implementation of the cloud data storage system for file transfer, enhanced with quality of service (QoS) features, has been presented.

The research article is arranged in different sections as follows. Section 2 describes the cost optimization techniques viz. PSO and QNM. Section 3 presents the classical queueing model and formulates the steady-state equations governing the model. The stationary probability distribution and various performance indices are derived in Sects. 4 and 5, respectively. The cost function and its optimization issues are discussed in Sect. 6. In Sect. 7, the fuzzy queueing model is studied by using the parametric nonlinear programming (P-NLP) approach. Section 8 presents the numerical illustration, sensitivity analysis, and cost optimization results. Finally, in Sect. 9, we conclude the outcomes of the research work done by mentioning the noble features and future scope.

2 Preliminaries on PSO & QNM

The focus of the present research work is to tackle the cost optimization issues for which we shall frame the cost function associated with the concern service system. In the following subsections, we present an appropriate background of the methodological aspects to design the optimal parameters using optimization techniques PSO and QNM.

2.1 Particle swarm optimization (PSO)

In 1995, Kendelly and Eberhart [35] introduced the concept of PSO which is based on the idea of birds flocking and fish schooling and is used to solve the optimization problems related to continuous nonlinear function. No matter how complex a function is, this algorithm can be successfully implemented to find a globally optimal solution. PSO deals with a population of particles, and each particle indicates a child solution, i.e., the particle's position and velocity. Moreover, the activity of each particle gives its best global position in the realizable space. The basic concepts involved in the PSO algorithm are briefly explained below.

The main objective of a PSO is to solve a multi-variable nonlinear continuous optimization problem with D decision variables. In PSO, the solution space should be viewed as D dimensional space, and each particle position is a potential solution of every considered problem. In general, $\mathrm{X}_i = \left(\chi_{i1}, \chi_{i2}, ..., \chi_{id}, ...\chi_{iD}\right)$ is defined as particle position at i and χ_{id} be the particle position at i with coordinate d. Consequently, particles move forward according to its own coordinate and step size. The velocity $V_i = \left(v_{i1}, v_{i2}, ..., v_{id}, ...v_{iD}\right)$ is randomly generated by its best solution (*pbest*)

and global best solution (*gbest*). The position χ_{id} and velocity v_{id} can be updated using the following formulae:

$$\chi_{id}^{t+1} = \chi_{id}^{t} + v_{id}^{t}$$
$$\text{and, } v_{id}^{t+1} = v_{id}^{t} + 2 \times rand(\bullet) \times \left(p_{id} - \chi_{id}^{t}\right) + 2 \times rand(\bullet) \times \left(p_{gd} - \chi_{id}^{t}\right)$$

where $rand(\bullet)$ is a number lying between 0 and 1, p_{id} denotes the best position of particle i with dimension d, and p_{gd} is the global best position of flock with coordinate d.

2.2 Quasi-Newton method (QNM)

To obtain the optimal decision results, we have used the gradient-based numerical optimization algorithm called quasi-Newton method (QNM). It can be easily implemented to evaluate the design parameters of complex nonlinear objective functions. QNM is quite often employed to solve global optimization problems with continuous decision variables.

To outline QNM, we explain the vector $\vec{X}$ corresponding to decision parameters μ and η, and get a gradient function $\vec{\nabla} TC\left(\vec{X}\right) = \left[\frac{\partial TC(\mu,\eta)}{\partial \mu} \; \frac{\partial TC(\mu,\eta)}{\partial \eta}\right]^{T}$. Let (μ^*, η^*) be the optimum value of (μ, η). First, we initialize the values of decision variables $\vec{X}_1 = [\mu_1, \eta_1]$. Then, we construct the respective gradient $\vec{\nabla}\, TC\left(\vec{X}_1\right)$.

The algorithmic steps of quasi-Newton's approach are as follows:

***Step 1*:** Fix an initial solution $\vec{X}_1 = [\mu_1, \eta_1]$, j = 1, and tolerance ε.

***Step 2*:** Calculate initial trial solution for $\vec{X}_1$, and compute $TC\left(\vec{X}_1\right)$..

***Step 3*:** Obtain gradient and Hessian matrix objective function at the point $\vec{X}_j$ and set $j = 1$. $\vec{\nabla} TC\left(\vec{X}_j\right) = \left[\frac{\partial TC(\mu,\eta)}{\partial \mu}, \frac{\partial TC(\mu,\eta)}{\partial \eta}\right]^{T}\Big|_{\vec{X}_j}$ and $H\left(\vec{X}_j\right) = \left[\begin{array}{cc} \frac{\partial^2 TC(\mu,\eta)}{\partial \mu^2} & \frac{\partial^2 TC(\mu,\eta)}{\partial \mu \partial \eta} \\ \frac{\partial^2 TC(\mu,\eta)}{\partial \mu \partial \eta} & \frac{\partial^2 TC(\mu,\eta)}{\partial \eta^2} \end{array}\right]\Bigg|_{\vec{X}_j}$.

***Step 4*:** Obtain new solution by $\vec{X}_{j+1} = \vec{X}_j - \left[\mathrm{H}\left(\vec{X}_j\right)\right]^{-1} \vec{\nabla} TC\left(\vec{X}_j\right)$.

***Step 5*:** Fix $j = j + 1$ and repeat steps (3) - (4) until $\mathrm{Max}\left(\left|\frac{\partial TC(\mu,\eta)}{\partial \mu}\right|, \left|\frac{\partial TC(\mu,\eta)}{\partial \eta}\right|\right) < \varepsilon$..

***Step 6*:** Compute the global minima of objective function at $\vec{X}_j^* = (\mu^*, \eta^*)$ using $TC(\mu^*, \eta^*) = TC(\vec{X}_j)$.

3 Proposed M/M/1/WV with disaster queueing model

This section describes a Markovian working vacation queueing model having discouragement, retrial pool, and disaster. Now, we describe the concerned model by stating various assumptions and notations and then formulate the governing equations.

We formulate a single-server retrial queueing model where arrival/service times and re-attempts/vacation duration are governed by Markovian processes. The user's discouragement behavior, i.e., balking/reneging, is considered. The arrivals of the users follow Poisson fashion with rate λ. As soon as the server becomes free, an arriving user is served immediately; otherwise, the user becomes a repeated user and joins retrial orbit. The user residing in the orbit retries for the service again and again with rate γ. If a repeated user becomes impatient, then he leaves (i.e., reneges) the system with probability $1-\nu$ (i.e., unsuccessful) without getting service or stay in the system with probability ν (i.e., successful) till getting service. During a normal busy period, the server renders service to the user following exponential distribution (Exp-D) with rate μ. Due to some unpredictable cause, the server may fail catastrophically with rate δ, and all the users in the system are lost. The repair of the server is done without delay with rate β, and the server becomes as good as before failure. After repair, the system becomes functional for rendering the service to the users. Whenever, there is no user in the system, the server switches to WV mode; the time duration of the server in WV follows EXP-D with rate θ. In the system, during WV, if there are only users present, then the server serves the user and moves to the idle state of WV busy mode. During WV, the server provides service to the users with rate $\eta(<\mu)$. Thereby, if there are users in the queue at the service completion moment of WV period, and the server returns to the idle state of normal busy mode, i.e., after completion of WV period, the server shifted to normal busy (NB) mode. The user joins the orbit with probability $q(q_v)$ depending upon the server status who may be in NB (WV) mode. The inter-arrival time, service time, repair time, catastrophes, and retrial time are mutually independent. However, for system stability, we

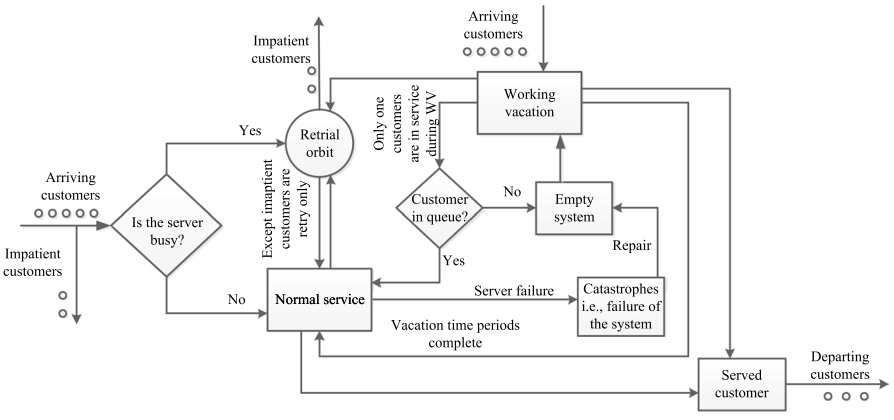


Fig. 1 A pictorial representation of the model

consider that the traffic intensity $\frac{\lambda}{\mu+\eta}$ is always less than 1. A pictorial representation of the concerned M/M/1/WV retrial queueing model with disaster, balking, or reneging behaviors of the users is shown in Fig. 1.

Let $\phi(t)$ denotes the number of users in the system including in the orbit at time t. The server status at time t is represented by random variables $\psi(t)$. We consider that (i) $\psi(t) = 0$, for the idle state of the server during WV mode, (ii) $\psi(t) = 1$, when the server is in busy state during WV mode, (iii) $\psi(t) = 2$, when the server is in idle state during NB mode, (iv) $\psi(t) = 3$, when the server is in busy state during NB mode, and (v) $\psi(t) = 4$, for the catastrophes state of the system when all the users are lost. The bivariate process $\{(\phi(t), \psi(t)),\ t \geq 0\}$ is a continuous time Markov chain (CTMC) with transition rates $e_{(n,i)(m,j)}$ from state (n, i) to (m, j). Here

$$e_{(0,0)(0,1)} = \lambda q_v,\ e_{(0,1)(0,0)} = \eta,\ e_{(0,3)(0,0)} = \mu,\ e_{(0,4)(0,0)} = \beta;$$

$$e_{(n,1)(n+1,1)} = \lambda q_v,\ n \geq 0;\ e_{(n,1)(n,2)} = \eta,\ n \geq 1;$$

$$e_{(n,1)(n,3)} = \theta,\ n \geq 0;\ e_{(n,2)(n,3)} = \lambda,\ n \geq 1;\ e_{(n,3)(n+1,3)} = \lambda q,\ n \geq 0;$$

$$e_{(n,3)(n,2)} = \mu,\ n \geq 1;\ e_{(n,2)(n-1,3)} = \gamma\nu,\ n \geq 1; e_{(n,3)(0,4)} = \delta,\ n \geq 0.$$

The transient probabilities are denoted by $P_{n,i}(t) = \text{Prob.}\{\phi(t) = n,\ \psi(t) = i\}$; the respective steady-state probabilities are $P_{n,i} = \lim_{t\to\infty} P_{n,i}(t)$.

4 Stationary probability distribution

Now, we formulate Chapman–Kolmogorov equations at steady state by using the appropriate in-flow & out-flow rates of the system states shown in Fig. 2.

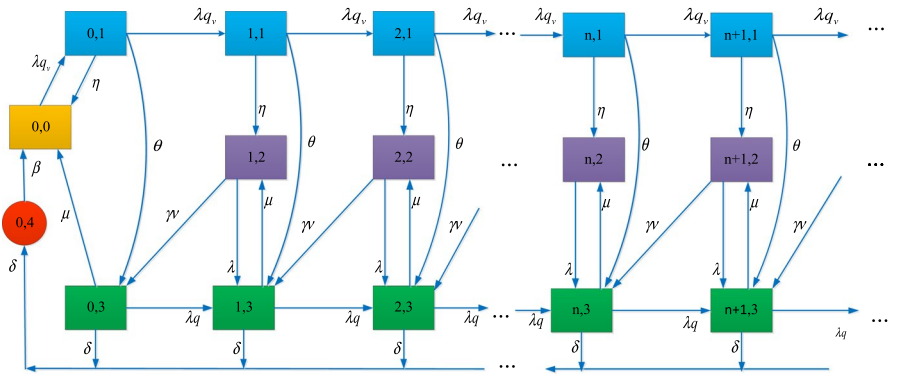


Fig. 2 Transition diagram of the model

4.1 Governing equations

By balancing the in-flow and out-flow rates, the following governing equations for the system states are framed:

$$\lambda q_v P_{0,0} = \eta P_{0,1} + \mu P_{0,3} + \beta P_{0,4} \tag{1}$$

$$\big(\lambda q_v + \eta + \theta\big) P_{0,1} = \lambda q_v P_{0,0} \tag{2}$$

$$(\lambda q + \mu + \delta) P_{0,3} = \theta P_{0,1} + \gamma \upsilon P_{1,2} \tag{3}$$

$$\big(\lambda q_v + \eta + \theta\big) P_{n,1} = \lambda q_v P_{n-1,1}, \quad n \geq 1 \tag{4}$$

$$(\gamma \upsilon + \lambda) P_{n,2} = \mu P_{n,3} + \eta P_{n,1}, \quad n \geq 1 \tag{5}$$

$$(\lambda q + \mu + \delta) P_{n,3} = \lambda P_{n,2} + \lambda q P_{n-1,3} + \theta P_{n,1} + \gamma \upsilon P_{n+1,2}, \quad n \geq 1 \tag{6}$$

$$\beta P_{0,4} = \delta \sum_{n=0}^{\infty} P_{n,3} \tag{7}$$

4.2 Evaluation of probabilities

The set of difference Eqs. (1)–(7) are solved using the recursive approach and difference equations approach [46].

Iteratively solving Eq. (4) and combined with Eq. (2), we obtain

$$P_{n,1} = \ell^{n+1} P_{0,0} \tag{8}$$

where $\ell = \frac{\lambda q_v}{\lambda q_v + \eta + \theta}$.

After some algebraic manipulations, Eqs. (5), (6), and (8) yield

$$\gamma \upsilon \mu P_{n+1,3} - \{a(\lambda q + \delta) + \mu \gamma \upsilon\} P_{n,3} + a \lambda q P_{n-1,3} = -(\lambda \eta + \gamma \upsilon \eta \ell + \theta(\lambda + \gamma \upsilon)) \ell^{n+1} P_{0,0}, \; n \geq 1 \tag{9}$$

where $a = \lambda + \gamma$.

The general solution of the homogeneous difference Eq. (9) is

$$P_{n,3} = A_1 \rho_1^n + A_2 \rho_2^n, \quad n = 1, 2, 3, \ldots \tag{10}$$

where A_1 and A_2 are constants (see Sect. 2.3 of Elaydi 2006), and ρ_1 and ρ_2 are the root of the corresponding characteristic equation given by

$$w(x) := \gamma \upsilon \mu x^2 - \{a(\lambda q + \delta) + \mu \gamma \upsilon\} x + a \lambda q = 0 \tag{11}$$

Thus,

$$\left.\begin{aligned}\rho_1 &= \frac{a(\lambda q+\delta)+\mu\gamma v-\sqrt{(a(\lambda q+\delta)+\mu\gamma v)^2-4a\gamma v\mu\lambda q}}{2\gamma v\mu}\\ \rho_2 &= \frac{a(\lambda q+\delta)+\mu\gamma v++\sqrt{(a(\lambda q+\delta)+\mu\gamma v)^2-4a\gamma v\mu\lambda q}}{2\gamma v\mu}\end{aligned}\right\}. \tag{12}$$

Since $w(0) > 0$, $w(1) < 0$ and $\lim_{x\to+\infty} w(x) = +\infty$, therefore from Bolzano's theorem, we have $\rho_1 \in (0, 1)$ and $\rho_2 \in (1, +\infty)$. Furthermore, $\sum_{n=0}^{\infty} P_{n,3} < \infty$ implies that the constant A_2 should be zero. Thus, Eq. (10) yields

$$P_{n,3} = A_1\left(\frac{a(\lambda q+\delta)+\mu\gamma v\mp\sqrt{(a(\lambda q+\delta)+\mu\gamma v)^2-4a\gamma v\mu\lambda q}}{2\gamma v\mu}\right)^n \tag{13}$$

Now, we obtain the specific solution which is of the form $x^{Spec} = K\ell^n P_{0,0}$ corresponding to nonhomogeneous part of Eq. (9). The specific value $x^{Spec} = K\ell^n P_{0,0}$ put in Eq. (9); thus, we get

$$K = -\frac{\{\lambda\eta+\gamma v\ell\eta+a\theta\}\ell^2}{\gamma v\mu\ell^2-\{a(\lambda q+\delta)+\mu\gamma v\}\ell+a\lambda q} \tag{14}$$

Now, the general solution of Eq. (9) is given by

$$P_{n,3} = A_1\rho_1^n + K\ell^n P_{0,0} \tag{15}$$

Using Eqs. (7) and (15), we have

$$P_{0,4} = \frac{\delta}{\beta}\left[\frac{A_1}{1-\rho_1}+\frac{KP_{0,0}}{1-\ell}\right] \tag{16}$$

Substituting results of Eqs. (15)–(16) into Eq. (1), we get

$$P_{00} = RA_1 \tag{17}$$

where

$$R = \frac{\{\mu(1-\rho_1)+\delta\}(1-\ell)}{(1-\rho_1)\{(\lambda q_v-\eta\ell-\mu K)(1-\ell)-\delta K\}} \tag{18}$$

Thus, Eqs. (15) and (16) yield

$$P_{n,3} = (\rho_1^n + KR\ell^n)A_1 \tag{19}$$

$$P_{0,4} = \frac{\delta}{\beta}\left[\frac{1}{1-\rho_1} + \frac{KR}{1-\ell}\right]A_1 \tag{20}$$

Taking into account Eqs. (4), (5), and using Eq. (19), we have

$$P_{n,1} = \ell^{n+1} R A_1 \tag{21}$$

$$P_{n,2} = \left[\frac{\mu}{\lambda+\gamma v}\left(\rho_1^n + KR\ell^n\right) + \frac{R\eta}{\lambda+\gamma v}\ell^{n+1}\right]A_1 \tag{22}$$

It is noticed that probability distributions given in Eqs. (19)–(22) involve the constant term A_1 which is determined by normalization condition given by which implies that

$$P_{0,0} + \sum_{n=0}^{\infty} P_{n,1} + \sum_{n=1}^{\infty} P_{n,2} + \left(1 + \frac{\delta}{\beta}\right)\sum_{n=0}^{\infty} P_{n,3} = 1$$

$$A_1 = \left[\frac{R}{(1-\ell)} + \frac{\mu\rho_1(1-\ell) + R\ell(\mu K + \ell\eta)(1-\rho_1)}{(\lambda+\gamma v)(1-\rho_1)(1-\ell)} + \frac{(\beta+\delta)\{(1-\ell) + KR(1-\rho_1)\}}{\beta(1-\rho_1)(1-\ell)}\right]^{-1} \tag{23}$$

5 Performance indices

Here, we determine performance indices using the stationary distribution derived in the previous section.

The mean number of users in the system (EL_S) and in the queue (EL_Q), respectively, are given by

(i)

$$\begin{aligned} EL_S &= \sum_{n=0}^{\infty} nP_{n,1} + \sum_{n=1}^{\infty} nP_{n,2} + \sum_{n=0}^{\infty} nP_{n,3} + \frac{\delta}{\beta}P_{0,4} \\ &= \frac{R\ell^2}{(1-\ell)^2} + \frac{\mu\rho_1(1-\ell)^2 + R\ell(1-\rho_1)^2(\mu K + \ell\eta)}{(\lambda+\gamma v)(1-\ell)^2(1-\rho_1)^2} + \frac{(\delta+\beta)\left\{\rho_1(1-\ell)^2 + KR\ell(1-\rho_1)^2\right\}}{\beta(1-\ell)^2(1-\rho_1)^2} \end{aligned} \tag{24}$$

(ii)

$$EL_Q = \sum_{n=1}^{\infty}(n-1)P_{n,0} + \sum_{n=1}^{\infty}(n-1)P_{n,2} + \sum_{n=1}^{\infty}(n-1)P_{n,3} + \frac{\delta}{\beta}P_{0,4} \tag{25}$$

Using Little's formula, we obtain the average waiting time of the users in the system (EW_S) and in the queue (EW_Q), respectively, as follows:

(iii) $$EW_S = EL_S / \lambda_{eff} \tag{26}$$

(iv) $$EW_Q = EL_Q / \lambda_{eff} \tag{27}$$

where $\lambda_{eff} = \lambda q_v \sum_{n=0}^{\infty} P_{n,1} + \lambda \sum_{n=1}^{\infty} P_{n,2} + \lambda q \sum_{n=0}^{\infty} P_{n,3}$.

The mean number of users when the server in the normal busy (NB) state and in working vacation (WV) states, respectively, are

(i) $$EL_{NB} = \sum_{n=0}^{\infty} nP_{n,3} \tag{28}$$

(ii) $$EL_{WV} = \sum_{n=0}^{\infty} nP_{n,1} \tag{29}$$

The system throughput (*TH*) and the catastrophes rate (*DR*), respectively, are obtained using

(i) $$TH = \eta \sum_{n=0}^{\infty} P_{n,1} + \mu \sum_{n=0}^{\infty} P_{n,3} \tag{30}$$

(ii) $$DR = \delta \left[\frac{1}{1-\rho_1} + \frac{KR}{1-\ell} \right] A_1 \tag{31}$$

The long-run probabilities of the server being in different states, i.e., probabilities that the working vacationing server being idle $\left(P_{IWV}\right)$ and busy $\left(P_{BWV}\right)$, the probabilities the server being free $\left(P_{FNB}\right)$ and busy $\left(P_{BNB}\right)$ during NB mode, the server being in catastrophes state $\left(P_{DR}\right)$, respectively, are evaluated using

$$\text{(i)}P_{IWV} = P_{0,0},\ \text{(ii)}P_{BWV} = \sum_{n=0}^{\infty} P_{n,1},\ \text{(iii)}P_{FNB} = \sum_{n=1}^{\infty} P_{n,2},\ \text{(iv)}P_{BNB} = \sum_{n=0}^{\infty} P_{n,3},\ \text{(v)}P_{DR} = \frac{\delta}{\beta} \sum_{n=0}^{\infty} P_{n,3} \tag{32}$$

Balking rate (*BR*) and joining rate (*JR*) of the users, respectively, are given by

(i) $$BR = q_v P_{BWV} + qP_{BNB} \tag{33}$$

(ii) $$JR = 1 - \left(q_v P_{BWV} + qP_{BNB} \right) \tag{34}$$

6 Cost optimization

In this section, we determine the optimal control parameters viz service rates (μ and η) by constructing the cost function. The several cost factors (per unit time) used in framing the cost function are:

C_{hold}: Holding cost for the users staying in the system.
C_{iwv}: Cost when the server is idle during WV mode.
C_{bwv}: Cost when the server is busy during WV mode.
C_{fnb}: Cost when the server is free during NB mode.
C_{bnb}: Cost when the server is busy during NB mode.
C_{μ}: Cost when the server rendering the service during NB mode with rate μ.
C_{η}: Cost when the server is rendering the service during WV mode with rate η.
Now, the average total cost per unit time is given by

$$TC(\mu,\ \eta) = C_{hold}EL_S + C_{iwv}P_{IWV} + C_{bwv}P_{BWV} + C_{fnb}P_{FNB} + C_{bnb}P_{BNB} + \mu C_{\mu} + \eta C_{\eta} \tag{35}$$

Our main goal is to find optimal decision parameters viz. service rates (μ^*, η^*) and associated minimum cost $TC(\mu^*, \eta^*)$ defined in Eq. (35). However, it is observed that the $TC(\mu,\ \eta)$ is highly nonlinear as such it is very tedious to minimize it analytically. Therefore, we employ the numerical technique QNM and meta-heuristic PSO for the optimization purpose.

7 Fuzzy environment of FM/FM/1/WV with unreliable queueing model

Here, we consider the trapezoidal fuzzy numbers for different classical parameters and derive the performance indices of the queueing system in the fuzzified environment. The use of fuzzy numbers rather than classical numbers is a more suitable way to create an FM/FM/1/WV model with disaster and discouragement. It is worthwhile to give the brief account of fuzzy numbers related to concerned queueing problem ([32, 47]).

Now, we construct a fuzzy FM/FM/1/WV model for the unreliable server queue, which is more practical than the classical queue M/M/1/WV model. The fuzzified parameters $\overline{\lambda}$, $\overline{\mu}$, $\overline{\eta}$, and $\overline{\beta}$ corresponding to classical parameters λ, μ, η, and β, respectively, and their corresponding MFs $\varepsilon_{\overline{\lambda}}(a_1)$, $\varepsilon_{\overline{\mu}}(a_2)$, $\varepsilon_{\overline{\eta}}(a_3)$, and $\varepsilon_{\overline{\beta}}(a_4)$, are used to develop the fuzzified model.

Assume that $\overline{\Phi}_i$ be the fuzzy set with membership function (MF) $\varepsilon_{\overline{\Phi}_i}(x_i)$. Therefore, we get

$$\overline{\Phi}_i = \left\{ \left(x_i, \varepsilon_{\overline{\Phi}_i}(x_i) \right) : x_i \in X_i \right\},\ i = 1, 2, 3,\ 4 \tag{36}$$

For, the FM/FM/1/WV model, we obtain

$$\overline{\Phi}_i = \begin{cases} \overline{\lambda}, & i = 1 \\ \overline{\mu}, & i = 2 \\ \overline{\eta}, & i = 3 \\ \overline{\beta}, & i = 4 \end{cases}$$

As $\overline{\Phi}_i$ presents the fuzzy number, therefore $\overline{EL}_S$, $\overline{EW}_S$, $\overline{TH}$, and $\overline{DR}$ are also fuzzy numbers corresponding to EL_S, EW_S, TH, and DR, respectively.

Let us denote that $f(x_1, x_2, x_3, x_4) = EL_S$. The MFs of $\overline{EL}_S$, $\overline{EW}_S$, $\overline{TH}$, and $\overline{DR}$, respectively, are given by

$$\varepsilon_{\overline{EL}_S}(x) = \sup \min \left\{ \varepsilon_{\overline{\Phi}_i}(x_i) \,:\, i = 1, 2, 3, 4 \middle| x = EL_S \right\} \tag{37}$$

$$\varepsilon_{\overline{EW}_S}(x) = \sup \min \left\{ \varepsilon_{\overline{\Phi}_i}(x_i) \,:\, i = 1, 2, 3, 4 \middle| x = EW_S \right\} \tag{38}$$

$$\varepsilon_{\overline{TH}}(x) = \sup \min \left\{ \varepsilon_{\overline{\Phi}_i}(x_i) \,:\, i = 1, 2, 3, 4 \middle| x = TH \right\} \tag{39}$$

$$\varepsilon_{\overline{DR}}(x) = \sup \min \left\{ \varepsilon_{\overline{\Phi}_i}(x_i) \,:\, i = 1, 2, 3, 4 \middle| x = DR \right\} \tag{40}$$

The α-cuts of parameters $\overline{\Phi}_i$ are defined as

$$\Phi_i(\alpha) = \left[(x_i)_\alpha^{lb}, (x_i)_\alpha^{ub} \right] = \left[\min_{x_i \in X_i} \left\{ x_i \middle| \varepsilon_{\Phi_i}(x_i) \geq \alpha \right\}, \max_{x_i \in X_i} \left\{ x_i \middle| \varepsilon_{\Phi_i}(x_i) \geq \alpha \right\} \right]$$

7.1 Parametric nonlinear programming (P-NLP)

Using extension principle, we find α-cuts of $\overline{EL}_S$, $\overline{EW}_S$, $\overline{TH}$, and $\overline{DR}$ through P-NLP approach. Thus, we develop the MFs of $\overline{EL}_S$, $\overline{EW}_S$, $\overline{TH}$, and $\overline{DR}$. It is observed that $\varepsilon_{\overline{EL}_S}(x)$ is the minimum of $\varepsilon_{\overline{\Phi}_i}(x_i)$, $i = 1, 2, 3, 4$. To obtain the MFs, at least one of the following conditions must satisfies such that $x = \overline{EL}_S$, which satisfies $\varepsilon_{\overline{EL}_S}(x) = \alpha$.

(i) $\varepsilon_{\overline{\Phi}_1}(x_1) = \alpha, \varepsilon_{\overline{\Phi}_i}(x_i) \geq \alpha, i = 2, 3, 4$
(ii) $\varepsilon_{\overline{\Phi}_2}(x_2) = \alpha, \varepsilon_{\overline{\Phi}_i}(x_i) \geq \alpha, i = 1, 3, 4$
(iii) $\varepsilon_{\overline{\Phi}_3}(x_3) = \alpha, \varepsilon_{\overline{\Phi}_i}(x_i) \geq \alpha, i = 1, 2, 4$
(iv) $\varepsilon_{\overline{\Phi}_4}(x_4) = \alpha, \varepsilon_{\overline{\Phi}_i}(x_i) \geq \alpha, i = 1, 2, 3$

From above four conditions, the lower bound $\left(x_\alpha^{lb}\right)$ and upper bound $\left(x_\alpha^{ub}\right)$ of α -cuts of $\overline{EL}_S$ are computed as:

$x_{i,\alpha}^{lb} = \min EL_S$ and $x_{i,\alpha}^{ub} = \max EL_S$, $i = 1, 2, 3, 4$.

Now, $\left[\left(x_i\right)_{\alpha_1}^{lb}, \left(x_i\right)_{\alpha_1}^{ub}\right] \subseteq \left[\left(x_i\right)_{\alpha_2}^{lb}, \left(x_i\right)_{\alpha_2}^{ub}\right]$, $i = 1$ (1) 4 for $0 < \alpha_2 < \alpha_1 \leq 1$, is nested structure w.r.t. 'α.' To compute x_α^{lb} and x_α^{ub}, where $a_{i,\alpha}^{lb} \leq a_i \leq a_{i,\alpha}^{ub}$, $i = 1, 2, 3,\ 4$, it is sufficient to determine the left hand side and right hand side of $\varepsilon_{\overline{EL_S}}(x)$ defined in (37). Therefore, we get

$$\left(EL_S\right)_\alpha^{lb} = x_\alpha^{lb} = \min EL_S \tag{41}$$

$$\left(EL_S\right)_\alpha^{ub} = x_\alpha^{ub} = \max EL_S \tag{42}$$

Since, x_α^{lb} is increasing function and x_α^{ub} is decreasing function w.r.t. 'α,' we have. $x_{\alpha_2}^{lb} \leq x_{\alpha_1}^{lb}$ and $x_{\alpha_1}^{lb} \leq x_{\alpha_2}^{ub}$ for $0 < \alpha_2 < \alpha_1 \leq 1$.

Hence, the MFs of $\overline{EL_S}$ are obtained as

$$\varepsilon_{\overline{EL_S}}(x) = \begin{cases} L_S(x), & \left(EL_S\right)_{\alpha=0}^{lb} \leq x < \left(EL_S\right)_{\alpha=1}^{ub} \\ 1, & \left(EL_S\right)_{\alpha=1}^{lb} \leq x \leq \left(EL_S\right)_{\alpha=1}^{ub} \\ R_S(x), & \left(EL_S\right)_{\alpha=1}^{ub} < x \leq \left(EL_S\right)_{\alpha=0}^{ub} \end{cases} \tag{43}$$

Here $\left(EL_S\right)_\alpha^{lb}$ and $\left(EL_S\right)_\alpha^{ub}$ cannot be computed by analytical method. Therefore, we evaluate the shape function of the MFs $\left(\varepsilon_{\overline{EL_S}}(x)\right)$ of $\overline{EL_S}$. In (43), $L_S(x)$ and $R_S(x)$ stand for left and right shape functions of $\overline{EL_S}$ and defined as $L_S(x) = \left(\left(EL_S\right)_\alpha^{l}\right)^{-1}$ and $R_S(x) = \left(\left(EL_S\right)_\alpha^{l}\right)^{-1}$, respectively.

Now, we develop the lower bounds $\left(EW_S\right)_\alpha^{lb}$ and upper bound $\left(EW_S\right)_\alpha^{ub}$ of α-cuts of $\overline{EW}_S$ as.

$\left(EW_S\right)_\alpha^{lb} = \min EW_S$, where

$$a_{i,\alpha}^{lb} \leq a_i \leq a_{i,\alpha}^{ub},\ i = 1, 2, 3,\ 4 \tag{44}$$

$\left(EW_S\right)_\alpha^{ub} = \max EW_S$, where

$$a_{i,\alpha}^{lb} \leq a_i \leq a_{i,\alpha}^{ub},\ i = 1, 2, 3,\ 4 \tag{45}$$

The corresponding MFs $\varepsilon_{\overline{EW_S}}$ of $\overline{EW}_S$ are

$$\varepsilon_{\overline{EW_S}}(x) = \begin{cases} L_W(x), & \left(EW_S\right)_{\alpha=0}^{lb} \leq x < \left(EW_S\right)_{\alpha=1}^{ub} \\ 1, & \left(EW_S\right)_{\alpha=1}^{lb} \leq x \leq \left(EW_S\right)_{\alpha=1}^{ub} \\ R_W(x), & \left(EW_S\right)_{\alpha=1}^{ub} < x \leq \left(EW_S\right)_{\alpha=0}^{ub} \end{cases} \tag{46}$$

where $\left(\left(EW_S\right)_\alpha^{lb}\right)^{-1} = L(x)$ and $\left(\left(EW_S\right)_\alpha^{ub}\right)^{-1} = R(x)$.

Similarly, MFs of throughput $\left(\overline{TH}\right)$ and catastrophes $\left(\overline{DR}\right)$ are computed as given below.

$$(TH)_{\alpha}^{lb} = \min TH_S,\ (TH)_{\alpha}^{ub} = \max TH_S,\ \text{where } a_{i,\alpha}^{lb} \le a_i \le a_{i,\alpha}^{ub},\ i = 1, 2, 3,\ 4 \tag{47}$$

$$(DR)_{\alpha}^{lb} = \min DR_S,\ (DR)_{\alpha}^{ub} = \max DR_S,\ \text{where } a_{i,\alpha}^{lb} \le a_i \le a_{i,\alpha}^{ub},\ i = 1, 2, 3, 4 \tag{48}$$

The MFs of $\overline{TH}$ and $\overline{DR}$ are obtained using

$$\varepsilon_{\overline{TH}}(x) = \begin{cases} L_{TH}(x),\ (TH)_{\alpha=0}^{lb} \le x < (TH)_{\alpha=1}^{ub} \\ 1, \quad (TH)_{\alpha=1}^{lb} = x = (TH)_{\alpha=1}^{ub} \\ R_{TH}(x),\ (TH)_{\alpha=1}^{ub} < x \le (TH)_{\alpha=0}^{ub} \end{cases} \tag{49}$$

$$\varepsilon_{\overline{DR}}(x) = \begin{cases} L_{DR}(x),\ (DR)_{\alpha=0}^{lb} \le x < (DR)_{\alpha=1}^{ub} \\ 1, \quad (DR)_{\alpha=1}^{lb} = x = (DR)_{\alpha=1}^{ub} \\ R_{DR}(x),\ (DR)_{\alpha=1}^{ub} < x \le (DR)_{\alpha=0}^{ub} \end{cases} \tag{50}$$

Developing solutions for membership functions analytically is highly challenging due to their intricate nature. To approximate the shape of these membership functions, we utilize a numerical method, treating the system parameters as trapezoidal fuzzy numbers. To exemplify, we use trapezoidal numbers as the fuzzy input parameters.

Example: The input parameters are considered as a trapezoidal fuzzy input parameters, and it is defined as follows:

$$\overline{\lambda} = [x_{11}, x_{12}, x_{13}, x_{14}],\ \text{where } x_{11} < x_{12} < x_{13} < x_{14} \tag{51}$$

$$\overline{\mu} = [x_{21}, x_{22}, x_{23}, x_{24}],\ \text{where } x_{21} < x_{22} < x_{23} < x_{24} \tag{52}$$

$$\overline{\eta} = [x_{31}, x_{32}, x_{33}, x_{34}],\ \text{where } x_{31} < x_{32} < x_{33} < x_{34} \tag{53}$$

$$\overline{\beta} = [x_{41}, x_{42}, x_{43}, x_{44}],\ \text{where } x_{41} < x_{42} < x_{43} < x_{44} \tag{54}$$

The membership function of the trapezoidal fuzzy input numbers $\overline{\lambda}$, $\overline{\mu}$, $\overline{\eta}$, and $\overline{\beta}$ is defined as follows.

$$\varepsilon_{\overline{\lambda}}(x_1) = \begin{cases} \frac{x_1 - x_{11}}{x_{12} - x_{11}},\ x_{11} \le x_1 \le x_{12} \\ 1, \quad x_{12} \le x_1 \le x_{13} \\ \frac{x_{14} - x_1}{x_{14} - x_{13}},\ x_{13} \le x_1 \le x_{14} \\ 0, \quad \textit{Otherwise} \end{cases} \tag{55}$$

$$\varepsilon_{\overline{\mu}}(x_2) = \begin{cases} \frac{x_2 - x_{21}}{x_{22} - x_{21}}, & x_{21} \le x_2 \le x_{22} \\ 1, & x_{22} \le x_2 \le x_{23} \\ \frac{x_{24} - x_2}{x_{24} - x_{23}}, & x_{23} \le x_2 \le x_{24} \\ 0, & Otherwise \end{cases} \tag{56}$$

$$\varepsilon_{\overline{\eta}}(x_3) = \begin{cases} \frac{x_3 - x_{31}}{x_{32} - x_{31}}, & x_{31} \le x_3 \le x_{32} \\ 1, & x_{32} \le x_3 \le x_{33} \\ \frac{x_{34} - x_3}{x_{34} - x_{33}}, & x_{33} \le x_3 \le x_{34} \\ 0, & Otherwise \end{cases} \tag{57}$$

$$\varepsilon_{\overline{\beta}}(x_4) = \begin{cases} \frac{x_4 - x_{41}}{x_{42} - x_{41}}, & x_{41} \le x_4 \le x_{42} \\ 1, & x_{42} \le x_4 \le x_{43} \\ \frac{x_{44} - x_4}{x_{44} - x_{43}}, & x_{43} \le x_4 \le x_{44} \\ 0, & Otherwise \end{cases} \tag{58}$$

The lower bound and upper bound of the trapezoidal number are given as follows: For

$$\overline{\lambda},\ (x_1)_\alpha^L = (x_{12} - x_{11})\alpha + x_{11},\ (x_1)_\alpha^U = x_{14} - (x_{14} - x_{13})\alpha \tag{59}$$

For

$$\overline{\mu},\ (x_2)_\alpha^L = (x_{22} - x_{21})\alpha + x_{21},\ (x_2)_\alpha^U = x_{24} - (x_{24} - x_{23})\alpha \tag{60}$$

For

$$\overline{\eta},\ (x_3)_\alpha^L = (x_{32} - x_{31})\alpha + x_{31},\ (x_3)_\alpha^U = x_{34} - (x_{34} - x_{33})\alpha \tag{61}$$

For

$$\overline{\beta},\ (x_4)_\alpha^L = (x_{42} - x_{41})\alpha + x_{41},\ (x_4)_\alpha^U = x_{44} - (x_{44} - x_{43})\alpha \tag{62}$$

The membership function of the expected number of customer in the system $\overline{EL_S}$ is given as follows:

$$\varepsilon_{\overline{EL_S}}(x) = \sup \min \left\{ \varepsilon_{\overline{\Phi_i}}(x_i) : i = 1,2,3,4 \middle| x = \frac{m_5 \times m_7}{L} + \frac{m_4 L + L m_1}{m_2 L R_1} + \frac{m_3 \{\rho_1 L + \kappa m_4 R_1\}}{\beta L R_1} \right\} \tag{63}$$

where $L = (1-\ell)^2,\ m_1 = (\mu K + \ell\eta), m_2 = (\lambda + \gamma\upsilon), m_3 = (\delta + \beta), m_4 = \mu\rho_1, m_5 = Rl, m_6 = \rho_1,\ m_7 = l.$

Since it is a very difficult task to plot the membership function of the $\overline{EL_S}$, therefore, we use Eq. (43). Thus, parametric nonlinear programming (P-NLP) is employed to find the α-cut of $\overline{EL_S}$ by using Zadeh's extension principle which is

Table 2 Cost set for different cost elements (in $)

Cost set	C_{hold}	C_{iwv}	C_{bwv}	C_{fnb}	C_{bnb}	C_{μ}	C_{η}
I	70	10	35	25	50	25	10
II	100	10	45	25	70	25	10
III	130	20	65	30	90	30	15

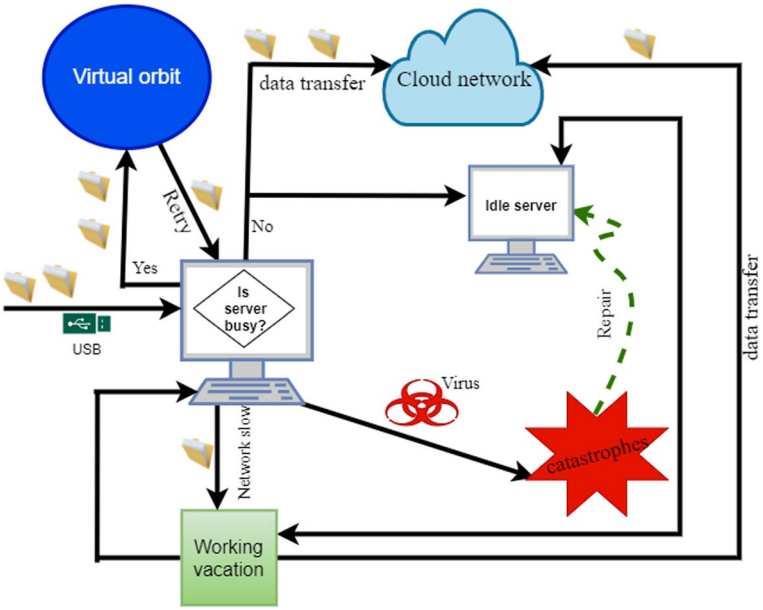


Fig. 3 Scenario for file transfer in cloud storage

given in Sect. 7.1. For example purpose, we set $\overline{\lambda} = [3,\ 3.25,\ 3.5,\ 4.25]$, $\overline{\mu} = [4.5,\ 4.75,\ 5,\ 5.25]$, $\overline{\eta} = [4.25,\ 4.5,\ 4.75,\ 5]$, $\overline{\beta} = [0.5,\ 0.75,\ 1,\ 1.25]$. The system organizer has the ability to assess the fuzzy expected number of customers in the system ($\overline{EL_S}$). For $\alpha = 0\ (0.1)1$, we determine $\left(\overline{EL_S}\right)_{\alpha}^{L}$ and $\left(\overline{EL_S}\right)_{\alpha}^{U}$ using the findings provided in Eqs. (41) and (42).

8 Illustration and numerical results

The tractability of analytical results through numerical results has been validated. We also briefly describe the application of our model in cloud network. For computation purpose, the coding has been done with MATLAB software to measure the performance indices and facilitate the sensitivity analysis. For the performance assessment of the proposed mode, we compute the system metrics established in

Table 3 System performance indices by varying parameters

(μ, η, λ, q)	P_{FWV}	P_{BWV}	P_{FNB}	P_{BNB}	DR	EL_3	EL_1	$TC(\mu, \eta)$
(2,4,4,0.3)	0.15	0.11	0.15	0.49	0.10	3.29	0.71	298.35
(3,4,4,0.3)	0.18	0.13	0.19	0.42	0.08	2.68	0.71	286.10
(4,4,4,0.3)	0.20	0.14	0.22	0.37	0.07	2.29	0.71	286.30
(5,4,4,0.3)	0.22	0.16	0.23	0.32	0.06	2.04	0.71	295.20
(6,2,4,0.3)	0.55	0.20	0.12	0.11	0.02	0.87	0.36	235.53
(6,3,4,0.3)	0.37	0.20	0.19	0.20	0.04	1.34	0.53	268.61
(6,4,4,0.3)	0.24	0.17	0.24	0.28	0.06	1.88	0.71	309.44
(6,5,4,0.3)	0.16	0.14	0.27	0.36	0.07	2.56	0.89	359.67
(6,4,1,0.3)	0.11	0.23	0.27	0.32	0.06	3.30	2.13	384.18
(6,4,2,0.3)	0.16	0.20	0.27	0.31	0.06	2.45	1.28	333.78
(6,4,3,0.3)	0.20	0.19	0.25	0.30	0.06	2.08	0.91	316.07
(6,4,4,0.3)	0.24	0.17	0.24	0.28	0.06	1.88	0.71	309.44
(6,4,5,0.3)	0.28	0.16	0.23	0.27	0.05	1.75	0.58	307.98
(6,4,4,0.5)	0.20	0.14	0.28	0.32	0.06	2.75	0.71	356.72
(6,4,4,0.7)	0.16	0.12	0.30	0.35	0.07	4.17	0.71	437.68
(6,4,4,0.9)	0.14	0.10	0.32	0.37	0.07	6.11	0.71	550.30

the previous sections. The default parameters for the considered classical model developed in Sect. 5 are set as follows:$\lambda = 3$, $\mu = 4.5$, $\eta = 4$, $\beta = 1.5$, $\delta = 0.5$, $\gamma = 4$, $\upsilon = 0.5$, $\theta = 0.5$, $q_v = 0.6$, and $q = 0.8$. To understand units of each parameter, we consider an exponential distribution for service time, retrial time, vacation time, and a Poisson distribution the arrival rate, and set the parameters as follows: $\mu = 4.5$ files/minutes, $\eta = 4$ files/minutes, arrival rate $\lambda = 3$ files/minutes, and $\gamma = 4$ files/minutes. Others parameters $\upsilon = 0.5$, $\theta = 0.5$, $q_v = 0.6$, and $q = 0.8$ are considered as probability values. For cost function computation, we set the cost elements for three cost sets as given in Table 2.

8.1 Application for data backup in cloud network

In high-speed cloud computing network, data backup is a critical problem. When data backup takes place by the clients via internet connection, some issues such as internet connectivity, software bugs, and so on may occur. During low traffic, the data are transferred quickly; however, when the traffic is high and network is in sleep mode, the file transfer becomes slow, i.e., the case of WV; and after completion of data transfer, the server returns to normal busy state (i.e., completion of WV). If the file is not uploaded due to busy network, the file waits in the retrial orbit and the file makes re-attempts until uploading is done successfully. In specific circumstances, when the computer/mobile device hangs (i.e., catastrophes occur), it will again start operation after rebooting/maintenance. The pictorial view of the data backup system is shown in Fig. 3. To understand the more practical situation, we consider exponential distribution for service time & retrial time, arrival rate as Poisson distribution and set the parameters as $\mu = 5$ files/minutes, $\eta = 4$ files/minutes, arrival rate $\lambda = 3$ files/minutes, $\gamma = 4$ files/minutes, $\upsilon = 0.5$, $\theta = 0.5$, $q_v = 0.6$, and $q = 0.8$. For fix

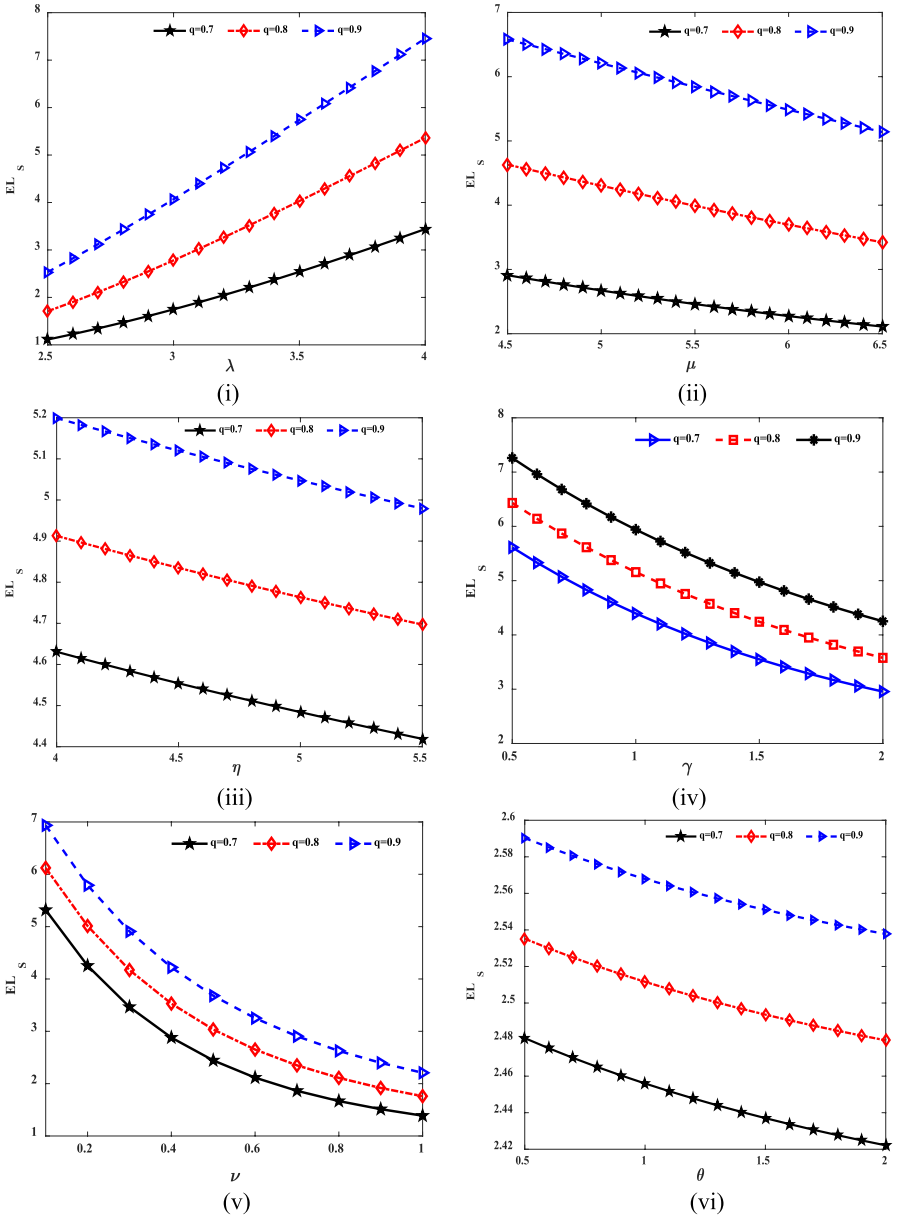


Fig. 4 Trends of EL_S for different values of q and varying parameters (i) λ(ii) μ(iii) η (iv) γ (v) ν (vi) θ

values of parameters, using Eq. (24), we get find the average count of files present in the cloud storage area of the drive as 15.75 ≈22.

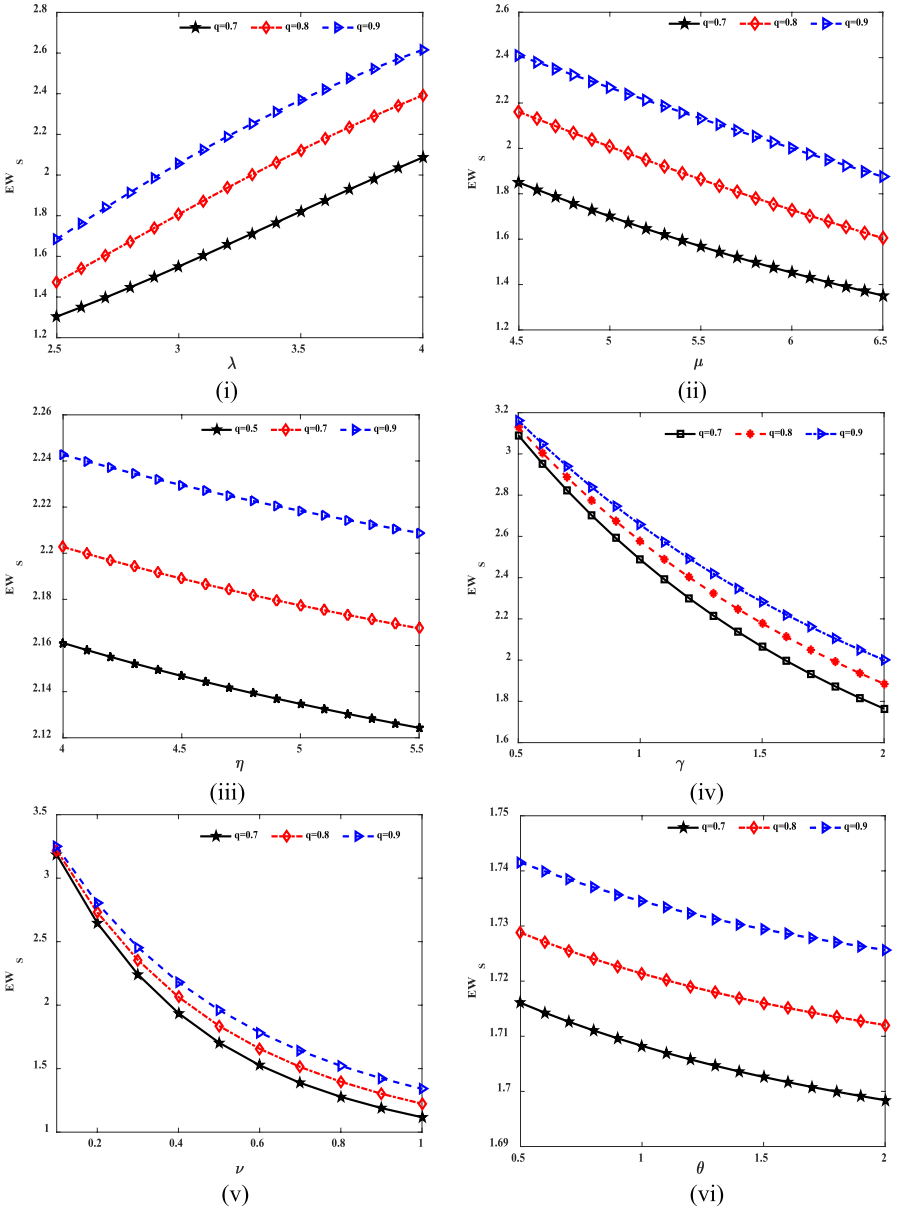


Fig. 5 Trends of EW_S for different values of q and varying parameters (i) λ (ii) μ (iii) η (iv) γ (v) ν (vi) θ

8.2 Sensitivity analysis

The sensitivity analysis by varying parameters has been performed for various performance measures viz. expected system length $\left(EL_S\right)$ in the system, expected spent time $\left(EW_S\right)$ in the system, throughput (TH), and catastrophes (DR), etc.

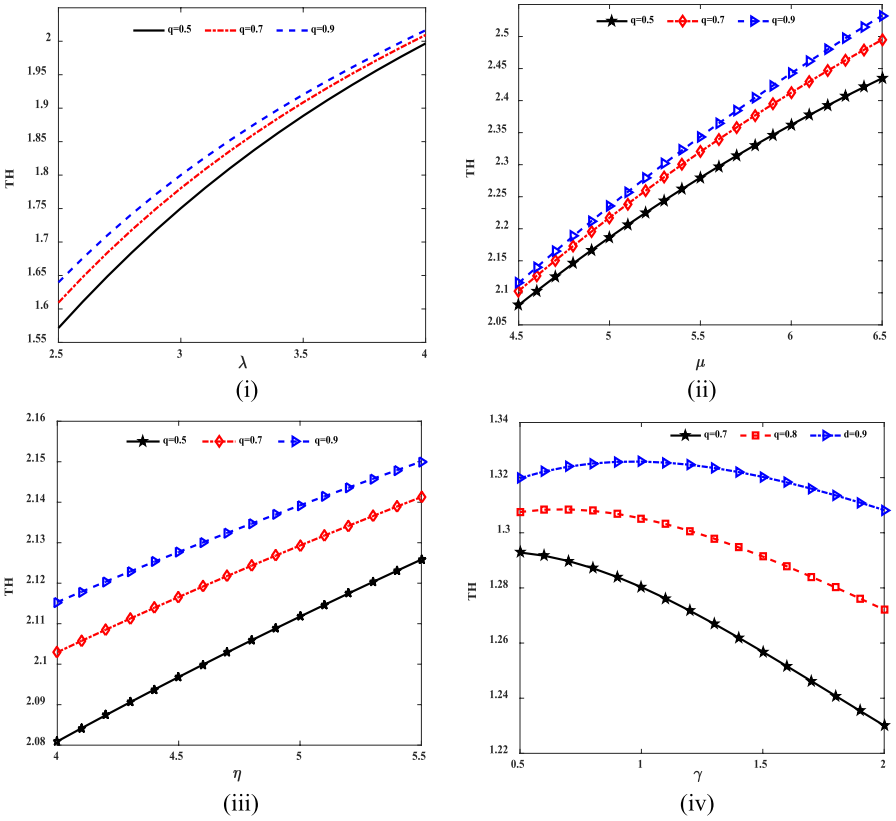


Fig. 6 Trends of TH for different values of q and varying parameters (i) λ (ii) μ(iii) η (iv) γ

The numerical results for the long-run probabilities, average system lengths for WV state and NB state, and total cost for varying parameters (μ, η, λ, q) are reported in Table 3. From Figs. 3, 4, 5, 6 and 7, we have the following observations.

- The average system length (EL_S) and the average time spent in the system (EW_S) both increase with the increasing value of the arrival rate (λ). This is due to the fact that the system becomes more congested if more users join it and results in long queue and delay in service to the users.
- We see that the EL_S and EW_S exhibit the decreasing trend by increasing the service rates (μ, η), retrial rate (γ), reneging rate (ν), and vacation rate (θ) for different values of joining probability (q).
- The throughput (TH) increases by the increment in both arrival rate (λ) and service rates (μ, η); however, TH decreases by increasing the values of the retrial rate (γ).

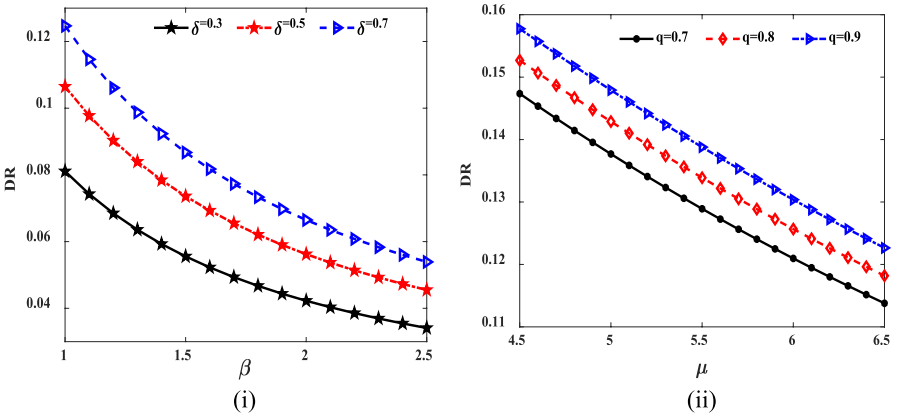


Fig. 7 Trends of *DR* for different values of δ and q and varying parameters (i) β (ii) μ

Table 4 α-cuts of fuzzy parameters

α	$(x_1)_\alpha^{lb}$	$(x_1)_\alpha^{ub}$	$(x_2)_\alpha^{lb}$	$(x_2)_\alpha^{ub}$	$(x_3)_\alpha^{lb}$	$(x_3)_\alpha^{ub}$	$(x_4)_\alpha^{lb}$	$(x_4)_\alpha^{ub}$
0	3.000	4.250	4.500	5.250	4.250	5.000	0.500	1.250
0.1	3.025	4.175	4.525	5.225	4.275	4.975	0.525	1.225
0.2	3.050	4.100	4.550	5.200	4.300	4.950	0.550	1.200
0.3	3.075	4.025	4.575	5.175	4.325	4.925	0.575	1.175
0.4	3.100	3.950	4.600	5.150	4.350	4.900	0.600	1.150
0.5	3.125	3.875	4.625	5.125	4.375	4.875	0.625	1.125
0.6	3.150	3.800	4.650	5.100	4.400	4.850	0.650	1.100
0.7	3.175	3.725	4.675	5.075	4.425	4.825	0.675	1.075
0.8	3.200	3.650	4.700	5.050	4.450	4.800	0.700	1.050
0.9	3.225	3.575	4.725	5.025	4.475	4.775	0.725	1.025
1	3.250	3.500	4.750	5.000	4.500	4.750	0.750	1.000

Table 5 α-cuts of fuzzified performance metrics

$[EL_S]_\alpha^{lb}$	$[EL_S]_\alpha^{ub}$	$[EW_S]_\alpha^{lb}$	$[EW_S]_\alpha^{ub}$	TH_α^{lb}	TH_α^{ub}	DR_α^{lb}	DR_α^{ub}
6.391	10.294	2.971	3.235	1.723	2.205	0.087	0.170
6.465	10.053	2.977	3.221	1.735	2.181	0.088	0.164
6.538	9.812	2.982	3.207	1.747	2.157	0.089	0.158
6.612	9.571	2.988	3.192	1.759	2.133	0.090	0.152
6.686	9.330	2.993	3.177	1.771	2.109	0.091	0.147
6.760	9.088	2.998	3.161	1.783	2.085	0.092	0.143
6.834	8.845	3.003	3.144	1.795	2.060	0.093	0.138
6.908	8.602	3.008	3.126	1.806	2.036	0.094	0.134
6.982	8.359	3.013	3.108	1.818	2.011	0.095	0.130
7.056	8.116	3.018	3.089	1.830	1.985	0.096	0.127
7.130	7.872	3.023	3.068	1.842	1.960	0.097	0.123

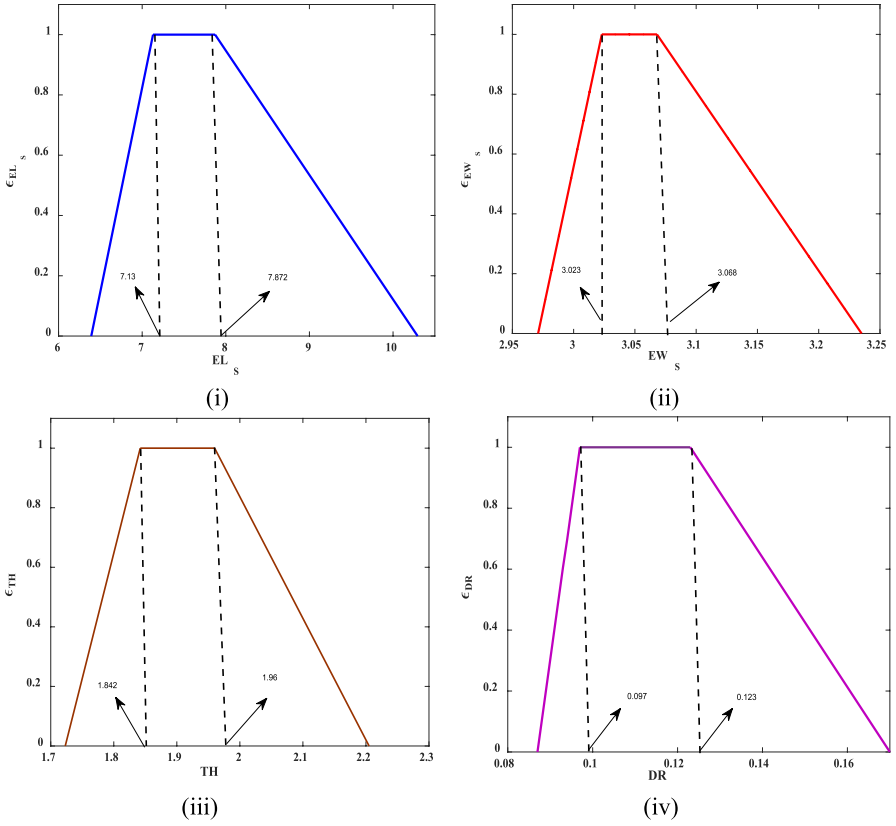


Fig. 8 The fuzzy MF for (i) EL_S (ii) EW_S (iii) TH (iv) DR

- The disaster rate (DR) exhibits the downfalls with respect to the increments in the value of repair rate β and service rate μ for different values of δ and q as shown in Fig. 7(i) and 7(ii), respectively.

8.3 Numerical results for fuzzy model

The fuzzified performance indices such as average system length $\left(\overline{EL_S}\right)$, average waiting time $\left(\overline{EW_S}\right)$, throughput $\left(\overline{TH}\right)$, and catastrophes $\left(\overline{DR}\right)$ are computed by parametric nonlinear programming (P-NLP) approach which is described in the previous section. The arrival rate and service rates during NB and WV states of the server and repair rate are taken as trapezoidal fuzzy numbers (TFN) and are given by.

Table 6 Experimental parameters of QNM and PSO

Parameters	Value	
	QNM	PSO
Lower bound (μ, η)	[4.5 4]	[4.5 4]
Upper bound (μ, η)	[8 7]	[8 7]
Dimensions (D)	2	2
Number of iterations	20	50
Stopping criteria $(IterMax)$	–	50
Tolerance	$\varepsilon = 10^{-7}$	50
Population size (N)	–	40
Acceleration parameters (c_0, c_1)	–	(2,2)
Inertia weight (w)		0.6

$\overline{\lambda} = [3,\ 3.25,\ 3.5,\ 4.25]$, $\overline{\mu} = [4.5,\ 4.75,\ 5,\ 5.25]$, $\overline{\eta} = [4.25,\ 4.5,\ 4.75,\ 5]$, $\overline{\beta} = [0.5,\ 0.75,\ 1,\ 1.25]$.

The classical intervals of fuzzified parameters, respectively, are taken as

(i) $\left[\left(x_1\right)_\alpha^{lb},\ \left(x_1\right)_\alpha^{ub}\right] = [3 + 0.25\alpha,\ 4.25 - 0.75\alpha]$,

(ii) $\left[\left(x_2\right)_\alpha^{lb},\ \left(x_2\right)_\alpha^{ub}\right] = [4.5 + 0.25\alpha,\ 5.25 - 0.25\alpha]$,

(iii) $\left[\left(x_3\right)_\alpha^{lb},\ \left(x_3\right)_\alpha^{ub}\right] = [4.25 + 0.25\alpha,\ 5 - 0.25\alpha]$,

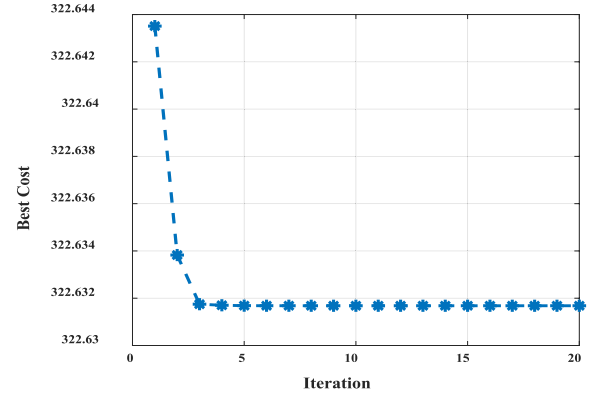
(iv) $\left[\left(x_4\right)_\alpha^{lb},\ \left(x_4\right)_\alpha^{ub}\right] = [0.5 + 0.25\alpha,\ 1.25 - 0.25\alpha]$.

The α-cuts of performance measures for different parameters are given in Table 4. The fuzzified performance indices $\overline{EL_S}$, $\overline{EW_S}$, $\overline{TH}$, and $\overline{DR}$ for $\alpha = 0\ (0.1)\ 1$ are summarized in Table 5.

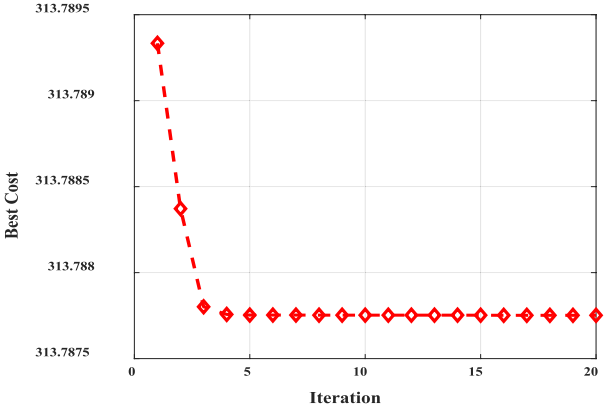
Figure 8i-iv exhibits the MFs of $\varepsilon_{\overline{EL_S}}$, $\varepsilon_{\overline{EW_S}}$, $\varepsilon_{\overline{TH}}$, $\varepsilon_{\overline{DR}}$, respectively, for $\alpha = 0\ (0.1)\ 1$. From Table 5, for $\alpha = 0$, we observe that fuzzified $\overline{EL_S}$, $\overline{EW_S}$, $\overline{TH}$, and $\overline{DR}$ lie in the interval [6.391, 10.294], [2.971, 3.235], [1.723, 2.205], and [0.087, 0.170], respectively. This shows that average system length, average waiting time, throughput, and catastrophes can be never fall outside these intervals. However, for $\alpha = 1$, fuzzified metrics $\overline{EL_S}$, $\overline{EW_S}$, $\overline{TH}$, and $\overline{DR}$ fall in the intervals [7.130, 7.812], [3.025, 3.068], [1.842, 1.960], and [0.097, 0.123], respectively. The similar patterns can also be observed for the MFs of $\overline{EL_S}$, $\overline{EW_S}$, $\overline{TH}$, and $\overline{DR}$ as shown in Fig. 8i-iv, respectively.

8.4 Cost optimization

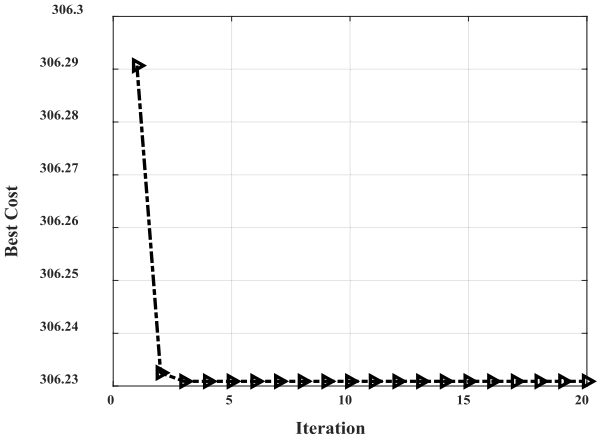
The minimum cost $TC(\mu^*, \eta^*)$ is attained at optimal value (μ^*, η^*) by solving the minimization problem formulated as: -



(i)

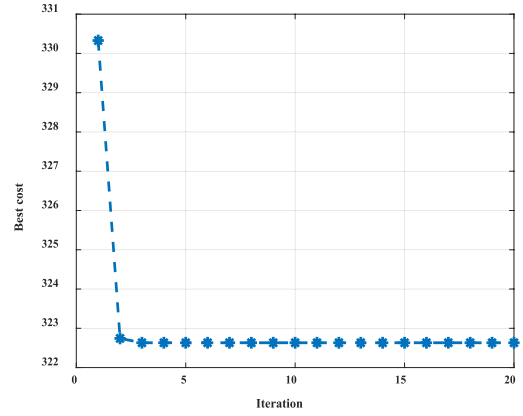


(ii)

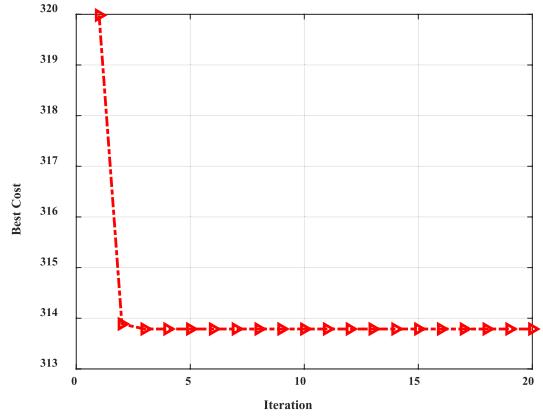


(iii)

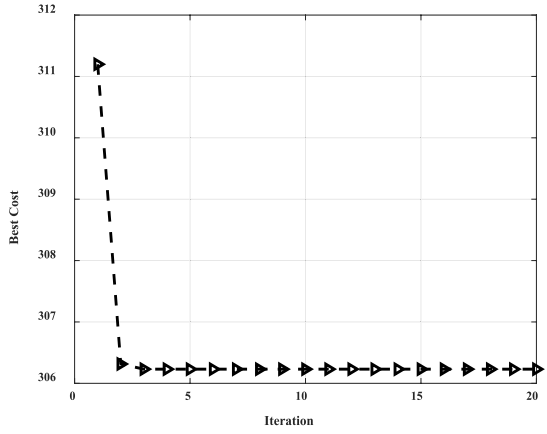
Fig. 9 (i-iii): The convergence graph of PSO for Cost Set I when $q = 0.7$ and $\gamma = 4$



(i)

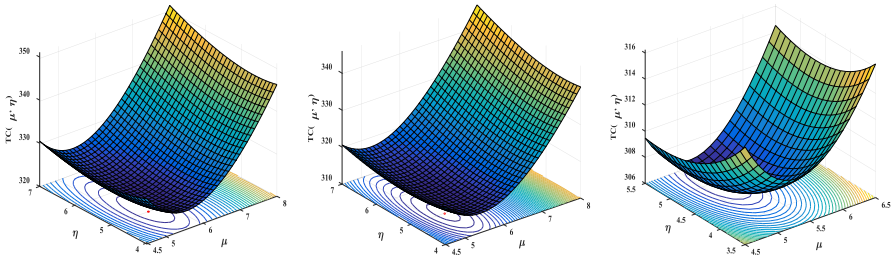


(ii)



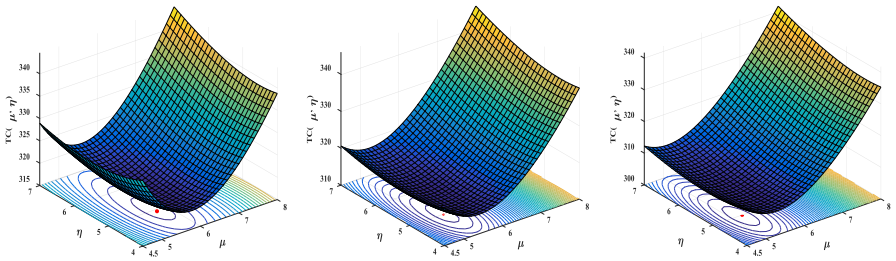
(iii)

Fig. 10 (i-iii): The convergence graph of QNM for Cost Set I when $q = 0.7$ and $\gamma = 4$



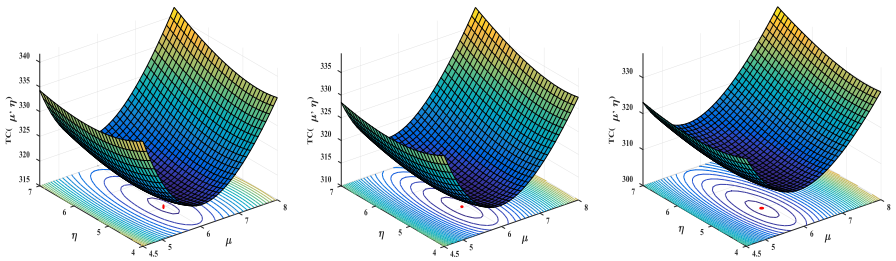
(i) $\gamma = 4$ (ii) $\gamma = 5$ (iii) $\gamma = 6$

Fig. 11 (i-iii): Total cost $TC(\mu, \eta)$ by varying values μ and η for cost set I when $q = 0.7$



(i) $\gamma = 4$ (ii) $\gamma = 5$ (iii) $\gamma = 6$

Fig. 12 (i-iii): Total cost $TC(\mu, \eta)$ by varying values μ and η for cost set II when $q = 0.8$



(i) $\gamma = 4$ (ii) $\gamma = 5$ (ix) $\gamma = 6$

Fig. 13 (i-iii): Total cost $TC(\mu, \eta)$ by varying values μ and η for cost set III when $q = 0.9$

$$(\text{OP}) : \ TC(\mu^*, \ \eta^*) = \underset{(\mu, \ \eta)}{\text{minimize}} \ TC(\mu, \ \eta) \tag{64}$$

subject to $\eta < \mu$; $\eta_{lb} < \eta < \eta_{ub}$; $\mu_{lb} < \mu < \mu_{ub}$.

We shall use the PSO and QNM for solving the optimization problem (OP) given in (64). The default cost element sets for optimization are given in Table 3. To determine the output through QNM and PSO, the numerical values chosen for the input parameters related to optimization techniques are summarized in Table 6.

Table 7 $(\mu^*,\ \eta^*, TC(\mu^*,\ \eta^*))$ for varying values q using QNM and PSO

q	γ	Method	Cost set I	Cost set II	Cost set III
0.7	4	PSO	(5.4980, 5.043, 322.63)	(6.2930, 6.4898, 368.52)	(6.5280, 5.6403, 477.83)
		QNM	(5.4980, 5.0438 322.63)	(6.2930, 6.4898, 368.52)	(6.5280, 5.6402, 477.83)
	5	PSO	(5.3483, 5.0036, 313.78)	(6.1014, 6.4150, 358.05)	(6.3263, 5.5859, 464.71)
		QNM	(5.3483, 5.0036 313.78)	(6.1014, 6.4150, 358.05)	(6.3262, 5.5859, 464.71)
	6	PSO	(5.2188, 4.9667, 306.23)	(5.9379, 6.3479, 349.13)	(6.1543, 5.5368, 453.51)
		QNM	(5.2188, 4.9667, 306.23)	(5.9379, 6.3479, 349.13)	(6.1543, 5.5367, 453.51)
0.8	4	PSO	(5.7102, 5.0309, 317.63)	(6.4232, 6.3995, 359.93)	(6.6473, 5.5770, 466.55)
		QNM	(5.7102, 5.0309, 317.63)	(6.4232, 6.3995, 359.93)	(6.6473, 5.5770, 466.55)
	5	PSO	(5.5969, 4.9977, 311.55)	(6.2860, 6.3424, 352.83)	(6.5037, 5.5346, 457.67)
		QNM	(5.5969, 4.9977, 311.55)	(6.2860, 6.3424, 352.83)	(6.5037, 5.5346, 457.67)
	6	PSO	(5.4966, 4.9672, 306.18)	(6.1651, 6.2906, 346.57)	(6.3773, 5.4958, 449.82)
		QNM	(5.4966, 4.9672, 306.18)	(6.1651, 6.2906, 346.57)	(6.3773, 5.4958, 449.82)
0.9	4	PSO	(5.9618, 5.0260, 317.80)	(6.6321, 6.3420, 357.93)	(6.8526, 5.5354, 463.61)
		QNM	(5.9618, 5.0260, 317.80)	(6.6321, 6.3420, 357.93)	(6.8526, 5.5354, 463.61)
	5	PSO	(5.8724, 4.9986, 313.27)	(6.5268, 6.2971, 352.70)	(6.7429, 5.5016, 457.05)
		QNM	(5.8724, 4.9986, 313.27)	(6.5268, 6.2971, 352.70)	(6.7429, 5.5016, 457.05)
	6	PSO	(5.7916, 4.9732, 309.18)	(6.4320, 6.2558, 347.98)	(6.4400, 5.4707, 451.14)
		QNM	(5.7916, 4.9732, 309.18)	(6.4320, 6.2558, 347.98)	(6.4400, 5.4707, 451.14)

For cost set I and parameters $q = 0.7$ and $\gamma = 4, 5$ and 6, the convergence trends of the TC by implementing PSO and QNM are shown in Figs. 9i-iii–10i-iii, respectively. It is observed that PSO performs better compared to QNM algorithm based on convergence criteria. The surface plots of $T(\mu,\ \eta)$ for cost sets I, II, and III are depicted in Figs. 11i-iii, 12, and 13i-iii.

Based on numerical results obtained via cost optimization, we have the following observations.

- We set the initial value of $(\mu,\ \eta) = (4.5,\ 4)$, and then, initially, we get from Eq. (35), $T(\mu,\ \eta) = \$330.33$. But, after 7th iteration of QNM, the set tolerance 10^{-7} is achieved. We observe that expected cost $T(\mu^*,\ \eta^*) = \$322.63$ is obtained at $(\mu^*,\ \eta^*) = (5.4980, 5.0439)$ which is reported in Table 7.
- Using PSO, the cost function $T(\mu,\ \eta)$ is minimized by varying parameters μ and η in the range of 4.5–8 and 4–7, respectively. For cost set I, by keeping fixed parameter $q = 0.7$ and varying parameters $\gamma = 4,\ 5,\ 6$, we obtain the numerical results of optimal service rates (μ^*, η^*) and corresponding cost $TC(\mu^*, \eta^*)$ which are summarized in Table 7.
- The cost function for cost set I, the same parameter $q = 0.7$, and $\gamma = 4,\ 5,\ 6$, respectively, is also displayed in Fig. 11i-iii.
- It is observed from Figs. 11i-iii, 12i-iii, and 13i-iii that the cost function is of convex nature with respect to service rates $(\mu,\ \eta)$.
- The procedure of QNM and PSO is implemented for both cost set II and cost set III. The optimal values of service rates and corresponding minimum costs are

recorded in Table 7. The numerical results of cost function are identically same while computing by the QNM and PSO.

- Based on numerical results, we conclude that for fixed cost set, the cost optimum $TC(\mu^*, \eta^*)$ decreases with an increase in the retrial rate γ.
- The retrial user can easily access the server more frequently, and the system throughput can be improved by setting the optimal design parameters.

8.5 Managerial insights

The analysis done in this article facilitates the performance metrics of Markovian retrial queueing model with WV policy and mixed joining strategies. The prime aim of the queue management is to facilitate the quality service at optimum cost by reducing the waiting time and improving the service rate. To ensure that every user receives quality service and faces less delay, the quantitative assessments of performance indices and joining strategies are required. The system organizer aims to ensure that their server provides maximum service to the consumers. At the same time, the organizer should minimize their costs and maximize their profits. Our study will be helpful to develop management insights into how to design and improve a service system with the quantitative assessment of performance metrics and cost–benefit trade-offs. The sensitivity analysis and numerical simulations carried out for the concerned model demonstrate the need of the assessment of expected waiting times so that the customers can access the system while keeping in mind the expected length of the queue. The cost optimization can help the system designers to overcome the techno-economic challenges against industry standards in the direction of scalability planning. By employing the queueing analytics, the decision-makers may quickly detect bottlenecks and peak hours just in time, and may check out about the vacation schedule, and other controllable factors such as service rates, etc. The study done may be helpful to take proactive actions ahead of time for effective management of congestion situations.

9 Conclusions

In this research work, we have analyzed the retrial M/M/1 queueing system in both classical and fuzzified environments. Using analytical methods based on difference equations, we have evaluated probabilities of the system states in explicit form. The key performance measures, namely, mean system length, mean waiting time, throughput, and machine catastrophes, have been examined analytically as well as via sensitivity analysis. The noble features of balking/reneging, retrial orbit, and working vacation have been incorporated, which make the proposed model facilitate valuable managerial insights for the concerned service system. The α-cuts and

P-NLP approaches have been used successfully for the performance prediction of the concerned queueing system operating in a fuzzy environment.

The fuzzy FM/FM/1 queue model extends by the classical M/M/1 model by incorporating fuzzy number. In the classical model, performance metrics such as average queue length, average system length, and average waiting time in the queue are clearly predicted through numerical values. However, in a fuzzy environment, the performance metrics encompass uncertainties and variabilities inherent in the queueing system, offering a range of outcomes that can inform more robust and resilient system designs. Moreover, in both cases, the average waiting time decreases when service rates increase. However, in a fuzzy environment, the average waiting time is contained within an interval of range values.

This research contributes significantly to the optimization of cloud storage systems. It offers a comprehensive, realistic, and adaptable model that effectively incorporates the complexities of real-world scenarios. The use of PSO in this context highlights the efficacy of evolutionary algorithms in addressing the complex operational challenges faced by cloud storage systems. This application demonstrates the versatility of these algorithms but also their potential in enhancing the efficiency and cost-effectiveness of cloud storage service delivery. In conclusion, our research provides a robust and practical framework that can be adapted for various cloud storage systems, ensuring improved service quality and operational efficiency.

By including the users' reneging and balking behavior, our model investigates realistic scenarios of congestion problems encountered in many places including cloud computing, telecommunication, call centers, banks, hospitals, etc. The study can be further external to design strategies for restricted admission of the jobs based on F-policy and service control using threshold-based N-policy. The systems organizers may utilize the findings of the sensitivity analysis and cost optimization to anticipate the system performance of the concerned system in advance. In queueing modeling, too much generalizability of the results may be infeasible, but the current study may be further extended to some more realistic domain such as feedback service, working breakdown, differentiated vacation, Bernoulli service, etc. The performance modeling of non-Markovian queueing system having general service can also be done but it will require more cumbersome analysis and computational complexity.

Author contribution Sibasish Dhibar helped in concept & model design, computational results, and manuscript write-up. Dr. Madhu Jain helped in model design, analysis verification, and manuscript write-up.

Funding The author (Sibasish Dhibar) is grateful to the Ministry of Education, India, for supporting the present research work via senior research fellowship (SRF), Grant (MHC01-23-200-428).

Data availability Not applicable, the study does not report any data.

Declarations

Conflict of interest All the authors declare that they have no known conflicts of interest.

Ethical approval This research paper does not contain any studies with human participants or animals performed by any of the authors.

References

1. Xiong Q, Zhang X, Liu W, Ye S, Du Z, Liu D, Zhu D, Liu Z, Yao X (2020) An efficient row key encoding method with ASCII code for storing geospatial big data in HBase. ISPRS Int J Geoinf 9(11):625. https://doi.org/10.3390/ijgi9110625
2. Bjeladinovic S, Marjanovic Z, Babarogic S (2020) A proposal of architecture for integration and uniform use of hybrid SQL/NoSQL database components. J Syst Softw 168:110633. https://doi.org/10.1016/j.jss.2020.110633
3. Krishnamoorthy A, Gopakumar B, Narayanan VC (2012) A retrial queue with server interruptions, resumption and restart of service. Oper Res 12:133–149. https://doi.org/10.1007/s12351-011-0112-8
4. Gao S, Niu X, Li T (2017) Analysis of a constant retrial queue with joining strategy and impatient retrial customers. Math Probl Eng 2017:1–8. https://doi.org/10.1155/2017/9618215
5. Jain M, Dhibar S (2020) Transient analysis of M/M/1 retrial queue with balking, imperfect service and working vacation. Mathematical modeling and computation of real-time problems. CRC Press, Boca Raton, pp 21–32
6. Zhang Y, Wang J (2023) Managing retrial queueing systems with boundedly rational customers. J Oper Res Soc 74(3):748–761. https://doi.org/10.1080/01605682.2022.2053305
7. Shi X, Liu L (2023) Equilibrium joining strategies in the retrial queue with two classes of customers and delayed vacations. Methodol Comput Appl Probab 25:52. https://doi.org/10.1007/s11009-023-10029-y
8. Falin GI (2008) The M/M/1 retrial queue with retrials due to server failures. Queueing Syst 58:155–160. https://doi.org/10.1007/s11134-008-9065-x
9. Sherman NP, Kharoufeh JP (2006) An M/M/1 retrial queue with unreliable server. Oper Res Lett 34:697–705. https://doi.org/10.1016/j.orl.2005.11.003
10. Chang J, Wang J (2018) Unreliable M/M/1/1 retrial queues with set-up time. Qual Technol Quant Manag 15:589–601. https://doi.org/10.1080/16843703.2017.1320459
11. Zhang Y, Wang J (2020) Strategic joining and information disclosing in Markovian queues with an unreliable server and working vacations. Qual Technol Quant Manag 18:298–325. https://doi.org/10.1080/16843703.2020.1809062
12. Jain M, Rani S (2021) Markovian model of unreliable server retrial queue with discouragement. Proc Natl Acad Sci India Sect A Phys Sci 91:217–224. https://doi.org/10.1007/s40010-020-00667-z
13. Bura GS (2019) Transient solution of an M/M/∞ queue with catastrophes. Commun Stat Theory Methods 48:3439–3450. https://doi.org/10.1080/03610926.2018.1477960
14. Jain M, Kaur S, Singh P (2021) Supplementary variable technique (SVT) for non-Markovian single server queue with service interruption (QSI). Oper Res 21:2203–2246. https://doi.org/10.1007/s12351-019-00519-8
15. Li K, Wang J (2021) Equilibrium balking strategies in the single-server retrial queue with constant retrial rate and catastrophes. Qual Technol Quant Manag 18:156–178. https://doi.org/10.1080/16843703.2020.1760464
16. de Souza MO, Rodriguez PM (2021) On a fractional queueing model with catastrophes. Appl Math Comput 410:126468. https://doi.org/10.1016/j.amc.2021.126468
17. Danilyuk E, Plekhanov A, Moiseeva S, Sztrik J (2022) Asymptotic diffusion analysis of retrial queueing system M/M/1 with impatient customers, collisions and unreliable servers. Axioms 11:699. https://doi.org/10.3390/axioms11120699
18. Servi LD, Finn SG (2002) M/M/1 queues with working vacations (M/M/1/WV). Perform Eval 50:41–52. https://doi.org/10.1016/S0166-5316(02)00057-3
19. Li J, Li T (2019) An M/M/1 retrial queue with working vacation, orbit search and balking. Eng Lett 27:97–102
20. Ameur L, Berdjoudj L, Abbas K (2019) Sensitivity analysis of the M/M/1 retrial queue with working vacations and vacation interruption. Int J Manag Sci Eng Manag 14:293–303. https://doi.org/10.1080/17509653.2019.1566034

21. Do NH, Van Do T, Melikov A (2020) Equilibrium customer behavior in the M/M/1 retrial queue with working vacations and a constant retrial rate. Oper Res 20:627–646. https://doi.org/10.1007/s12351-017-0369-7
22. Kumar A, Jain M (2021) M/M/1 queue with bi-level network process and bi-level vacation policy with balking. Commun Stat Theory Methods 52:5502–5526. https://doi.org/10.1080/03610926.2021.2012197
23. Jain M, Dhibar S, Sanga SS (2022) Markovian working vacation queue with imperfect service, balking and retrial. J Ambient Intell Humaniz Comput 13:1907–1923. https://doi.org/10.1007/s12652-021-02954-y
24. Muthusamy S, Devadoss N, Ammar SI (2022) Reliability and optimization measures of retrial queue with different classes of customers under a working vacation schedule. Discret Dyn Nat Soc 2022:1–17. https://doi.org/10.1155/2022/6806104
25. Dhibar S, Jain M (2023) Strategic behaviour for M/M/1 double orbit retrial queue with imperfect service and vacation. Int J Math Oper Res 25:369–385. https://doi.org/10.1504/IJMOR.2022.10048415
26. Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1:3–28
27. Li RJ, Lee ES (1989) Analysis of fuzzy queues. Comput Math with Appl 17:1143–1147. https://doi.org/10.1016/0898-1221(89)90044-8
28. Negi DS, Lee ES (1992) Analysis and simulation of fuzzy queues. Fuzzy Sets Syst 46:321–330. https://doi.org/10.1016/0165-0114(92)90370-J
29. Chen SP (2006) A bulk arrival queueing model with fuzzy parameters and varying batch sizes. Appl Math Model 30:920–929. https://doi.org/10.1016/j.apm.2005.06.002
30. Pardo MJ, de la Fuente D (2007) Optimizing a priority-discipline queueing model using fuzzy set theory. Comput Math Appl 54:267–281. https://doi.org/10.1016/j.camwa.2007.01.019
31. Chen G, Govindan K, Yang ZZ et al (2013) Terminal appointment system design by non-stationary M(t)//c(t) queueing model and genetic algorithm. Int J Prod Econ 146:694–703. https://doi.org/10.1016/j.ijpe.2013.09.001
32. Jain M, Kumar P, Meena RK (2020) Fuzzy metrics and cost optimization of a fault-tolerant system with vacationing and unreliable server. J Ambient Intell Humaniz Comput 11:5755–5770. https://doi.org/10.1007/s12652-020-01951-x
33. Sanga SS, Jain M (2019) FM/FM/1 double orbit retrial queue with customers' joining strategy: a parametric nonlinear programing approach. Appl Math Comput 362:124542. https://doi.org/10.1016/j.amc.2019.06.056
34. Sanga SS, Jain M (2022) Fuzzy modeling of single server double orbit retrial queue. J Ambient Intell Humaniz Comput 13:4223–4234. https://doi.org/10.1007/s12652-022-03705-3
35. Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: Proceedings of the 6th International Symposium Micro Machine and Human Science, pp 39–43. https://doi.org/10.1109/MHS.1995.494215.
36. Wang J, Zhang X, Huang P (2017) Strategic behavior and social optimization in a constant retrial queue with the N-policy. Eur J Oper Res 256(3):1–9. https://doi.org/10.1016/j.ejor.2016.06.034
37. Zhou M, Liu L, Chai X, Wang Z (2020) Wang, Equilibrium strategies in a constant retrial queue with setup time and the N -policy. Commun Stat Theory Methods 49:1695–1711. https://doi.org/10.1080/03610926.2019.1565779
38. Wang Z, Liu L, Zhao YQ (2021) Equilibrium customer and socially optimal balking strategies in a constant retrial queue with multiple vacations and N-policy. J Comb Optim 43(4):870–908. https://doi.org/10.1007/s10878-021-00814-1
39. Sumathi D, Manivannan SS (2021) Stochastic approach for channel selection in cognitive radio networks using optimization techniques. Telecomm Sys 76(2):167–186. https://doi.org/10.1007/s11235-020-00705-6
40. Meena RK, Jain M, Assad A, Sethi R, Garg D (2022) Performance and cost comparative analysis for M/G/1 repairable machining system with N-policy vacation. Math Comp Simul 200:315–328. https://doi.org/10.1016/j.matcom.2022.04.012
41. Jain M, Dhibar S (2023) ANFIS and metaheuristic optimization for strategic joining policy with re-attempt and vacation. Math Comput Simul 211:57–84. https://doi.org/10.1016/j.matcom.2023.03.024
42. Li T, Zhang L, Gao S (2016) Performance of an M/M/1 retrial queue with working vacation interruption and classical retrial policy. Adv Oper Res 2016:1–9. https://doi.org/10.1155/2016/4538031

43. Yang DY, Wu CH (2019) Performance analysis and optimization of a retrial queue with working vacations and starting failures. Math Comput Model Dyn Syst 25:463–481. https://doi.org/10.1080/13873954.2019.1660378
44. Jain M, Sanga SS (2021) Unreliable single server double orbit retrial queue with balking. Proc Natl Acad Sci India Sect A Phys Sci 91:257–268. https://doi.org/10.1007/s40010-020-00725-6
45. Lakaour L, Aissani D, Adel-Aissanou K et al (2022) An unreliable single server retrial queue with collisions and transmission errors. Commun Stat Theory Methods 51:1085–1109. https://doi.org/10.1080/03610926.2020.1758943
46. Elaydi S (2006) An introduction to difference equations, 2nd edn. Springer, New York
47. Jain M, Kumar P, Sanga SS (2020) Fuzzy Markovian modeling of machining system with imperfect coverage, spare provisioning and reboot. J Ambient Intell Humaniz Comput 12:7935–7947. https://doi.org/10.1007/s12652-020-02523-9

Authors and Affiliations

Sibasish Dhibar[1] · Madhu Jain[1]

✉ Sibasish Dhibar
sdhibar@ma.iitr.ac.in

Madhu Jain
madhu.jain@ma.iitr.ac.in

[1] Indian Institute of Technology Roorkee, Roorkee 247667, India