Chapter 2

Multi-Server Service System with Impatient and Bernoulli Scheduled Modified Vacation

2.1 Introduction

In the classical service models, servers are always available. But in many practical queueing scenarios of service systems, we observe that server(s) may become unavailable for a random period because of various reasons such as checking for the maintenance before starting the new job, for reducing the idle time, or simply for taking a rest. The period of unavailability of the server is known as the server's vacation. The vacation queueing theory was developed as a generalization of classical queueing theory in the late 1970s to cover the increasing complexity and inadequacy of the classical queueing models and several stochastic service systems, namely, telecommunication networks, manufacturing system, supply chain systems et cetera. Levy and Yechiali [168] first introduced the concept of server' vacation in the waiting line problem and obtained the Laplace Stieltjes transforms of waiting time of customers, vacation period of server and occupation period. Later, [63] surveyed the vacation queueing models and demonstrated an overview of some general decomposition results with their corresponding methodology. In the context of vacation policies, the reference books by [251], [252] have valuable contents and references therein. Many studies on interesting realtime queueing problems with different types of vacation policies have been done in recent also (*cf.* [10], [138], [274], [188], [244]) for looking merits and restrictions of server' vacation.

In this chapter, we consider that at the completion of the vacation, if the server finds no waiting customer to serve, either the server leaves for a vacation of random period again or waits idly for the next arrival to which the server provides the service immediately, *i.e.* the server follows Bernoulli scheduled modified vacation policy. Number of research articles (*cf.* [43], [45], [184], [300], [294], [186], [324]) have appeared in the literature on the queueing theory in which the server follows the single vacation policy. The classical vacation theory with the Bernoulli schedule scheme was first proposed and studied by [134]. Tadj et al. [247] considered a bulk service queueing model with random set-up time and N-policy under Bernoulli vacation schedule and suggested an algorithm to allocate the optimal management policy. Jain and Agrawal [115] analyzed a batch arrival queueing model using modified Bernoulli schedule vacation under N-policy and determined queue size distribution using the generating function technique. Ke et al. [133] examined $M/M/c$ retrial queue with Bernoulli single vacation policy as a quasi-birth and death process and provided the necessary and sufficient condition of system equilibrium. Having increased applicability in the state-of-the-art strategy, in recent years also several researchers (*cf.* [44], [157], [42], [83]) used the concept of Bernoulli scheduled vacation policy and provided several numerical illustrations after deriving the explicit expressions for various system performance measures.

In recent years, queue-based service systems considering customer impatient behavior have obtained significant attention of several researchers due to their practical applications to administration in service systems and e-commerce. It directly affects goodwill, revenue, satisfaction level in the service system. However, Some works in the queueing survey consider the impatient behavior of customers with various vacation policies. In the present chapter, we consider a service model with balking behavior in which the impatient customer may or may not join the queue depending on the number of waiting customers present in the service system. Haight [95] first introduced the balking behavior of customers for an $M/M/1$ queueing problem. Haight [96] again proposed the reneging behavior of customers for the single server Markovian problem. Later, [9] exhibited a comprehensive analysis of $M/M/1$, $M/G/1$, and $M/M/c$ queue with server's vacation and customer impatience, where customers' impatience is due to an absentee of servers upon arrival. They obtained various closed-form expressions by discussing both single and multiple vacation cases. To show the effect of the change in system parameters on the cost function, [280] performed the optimal and sensitivity analysis and provided various numerical experiments for the illustration purpose. Ward [285] surveyed the results for the $GI/GI/N+GI$ queueing model and studied the reneging behavior of the corresponding queueing problem in the conventional heavy traffic and Halfin-Whitt limit regimes. Yang et al. [304] studied the equilibrium balking behavior of the customers with server breakdown and repair for an $Geo/Geo/1$ queue and obtained the stationary distribution of the system with the mean sojourn time for an arriving customer.

Recently, [219] established the stationary relationship between the number of customers at different characteristic epochs for an $M/G/1$ multiple vacation system with balking. For single server Markovian queues with double adaptive working vacation, [244] studied customers' equilibrium and social optimal balking behavior. More recently, [126] analyzed both single and multiple vacation policies for a multi-server balking retrial queue and provided different numerical results for the comparison of both the vacation policies under optimal operating conditions. For a single server Markovian queue, [160] studied customers' equilibrium balking strategies with a single working vacation and vacation interruption. They derived the Nash equilibrium threshold strategies for fully and partially observable queues and conducted the sensitivity analysis for equilibrium threshold and equilibrium joining probabilities by varying the values of several system parameters. In addition, these researchers also examined the queueing problems with server's vacation and impatient behavior of the customers (*cf.* [106], [271], [175]).

From the economical point of view, the customers' impatience is perceived as a potential loss of customers and thereby results in the loss of total revenue due to their negative effect on the prospective financial circumstances of a service system. Therefore, the consideration of customer' retention policy acknowledges a requirement for a better service. Hence, system analysts employ a number of customer retention schemes to maintain a satisfactory level up to the mark. The concept of retention of reneged customers was first introduced by [154] for a finite capacity single server queueing problem and obtained the closed-form expressions of various performance measures analytically. Kumar and Sharma [150] extended their previous work by considering the balking strategies of customer and carried out a sensitivity analysis of the model. Recently, [303] investigated a finite capacity Markovian queueing model with working breakdown and retention of the impatient customer. They employed the matrix approach for computing the steady-state probabilities explicitly and formulated a cost function to find the minimal expected cost and optimal service rates. Kumar and Sharma [151] dealt with an infinite capacity multi-server system with balking and retention of reneged customers and derived the time-dependent probabilities using Bessel function and generating function technique.

From the literature survey, we note the following research gaps for which no research article is found: (i) the study of the effect of emergency vacation besides modified schedule vacation; (ii) retention of reneged customers in a queue-based service system with different vacation policies; and (iii) determination optimal vacation time using the metaheuristic technique. The main objectives of this chapter are threefold (i) to calculate the steady-state probabilities and various service system performance characteristics in vector form for which we employ an efficient matrix method; (ii) to develop the expected cost function for the service system with vacation to determine the optimal joint values of decision parameters R, μ, μ_1 and θ at the minimum cost value; (iii) an efficient nature-inspired optimization technique: PSO has been taken into consideration to depict the optimal operating values of various system parameters with minimal expected total cost.

The studied model has many applications in commerce sectors like banking, real estate, e-commerce, e-shopping, IPO, insurance, etc. For better services from service providers like schemes in real estate, e-shopping, or IPOs of good company, customers are overwhelming which may lead to over crowed. The imbalance between the availability of the products and demands leads the customer reluctance which indirectly increases the loss of customers, profits, goodwill, etc. To retain the reluctant customers, the service provider(s) may announce attractive and beneficiary schemes like discount, gift vouchers, gifts, fun games, etc. In greed for better services and attractive benefits customers may retain in the service system in spite of over-demand and long waiting. Depending on number of customers, service provider(s) decide the vacation and service strategies. Having more customers, less service provider(s) take vacation for lesser duration and provide faster service to avoid the reluctance among the customers. Our model is also applicable in client-mail service provider having the additional facility of virus scan.

The remainder of this chapter is consolidated as follows. In section 2.2, we present the description of the governing finite capacity multi-server service system with Bernoulli scheduled modified vacation and retaining policy of reneged customer. In subsection 2.2.1, we employ the matrix representation of the studied service system in the closed-form block matrices. Also, we provide the solution algorithm to obtain the steady-state probabilities in vector form in subsection 2.2.2. Various system performance measures are provided in vector form in section 2.3. In section 2.4, a cost function is developed to obtain the optimal values of several decision parameters at minimal expected cost. For the tractability of the studied model, some special cases are presented in section 2.5. The metaheuristic optimization technique, namely particle swarm optimization is implemented to deal with the optimization problem in section 2.6. Some numerical results are provided to illustrate the optimal analysis and simulation of various system performance measures in section 2.7. Lastly, section 2.8 gives conclusions and future perspective.

2.2 Problem Formulation and Notations

We consider a finite capacity multi-server service system with Bernoulli scheduled modified vacation policy and a realistic impatience behavior of customers. There are R homogeneous servers, arranged in parallel, to facilitate the required service to the customers in a service system having a capacity of K customers. For an investigation purpose, we also consider the following notations and assumptions:

Arrival Process

- The prospective customers join the service system according to the Poisson process, with the mean rate λ. The arrived customer forms a single queue if all available servers are busy otherwise customer gets service immediately. The idle server chooses the customer from the queue on the basis of the sequence of their arrivals *i.e.* *First Come First Serve* ($FCFS$) service discipline is followed.

Service Process

- The service times of customers are considered as identically and independently distributed and follow an exponential distribution with state-dependent meantime $\frac{1}{\mu}$ and $\frac{1}{\mu_1}$, $(\mu < \mu_1)$ depending on the number of customers waiting in the service system. On finding more than R customers in the service system, servers switch to the faster service rate to reduce the overload of the service system, which is referred to as a state-dependent queue-based service system.

Impatience Behavior

- If a newly arriving customer finds waiting customer(s) are queueing up in the service system, the customer may either balk without taking service with probability ξ or enter in the service system with the complementary probability $\bar{\xi}$.
- If the customer does not get service till random waiting period then the customer may exhibit an impatience behavior and may renege from the service system without being served with probability p_1 or may retain in the service system for the quality or extra beneficiary service with the complementary probability q_1. This random waiting time before reneging follows an exponential distribution with the meantime $\frac{1}{\zeta}$. This is referred to as the retention of the reneged customer.

Vacation Policy

- At the point of service completion of any customer, if the server has no customer to serve, the server leaves for a vacation of the random time period which also follows an exponential distribution with the state-dependent meantime $\frac{1}{\theta_i}$ where $i(\leq R)$ represents a number of servers on vacation.
- At the end of the vacation, if the server finds a lesser number of customers then available servers, he may opt for another vacation of random duration with probability p_2 or may join and remain idle in the service system with the complementary probability q_2. This is referred to as Bernoulli's scheduled modified vacation.
- The server may also take a vacation of the random time interval in an emergency during the busy period without completing the ongoing service to the

waiting customer. The occurrence of the emergent situation at which servers may opt emergency vacation follows a Poisson process with the mean rate δ and vacation time also follows an exponential distribution with meantime $\frac{1}{\theta_i}$.

- The vacation time of the server is forcefully curtailed shortly if more servers are on vacation using pressure coefficient to balance a load of customers to avoid excessive loss due to their impatience behavior. The effective state-dependent mean vacation time $\frac{1}{\theta_i}$ is defined in term of pressure factor (ψ) and number of servers on vacation (i) where θ_i is expressed as:

$$\theta_i = \left(i\theta + \left\{ \frac{i(R+i+1)}{R+i} \right\}^{\psi} \theta \right); \quad 1 \le i \le R, \psi > 0$$

  The pressure factor represents the degree to which the parameter of vacation is affected by the number of servers on vacation. The parameter θ is reciprocal of the meantime of random vacation duration when there is no pressure of more customers and fewer servers, *i.e.* all servers are available for providing the service.

All processes and events are repeated all over again and independent to the states of the other.

In this chapter, we furnish the steady-state analysis of the multi-server state dependent service system with Bernoulli modified vacation policy and retention of the reneged customer. For this purpose, we characterize various system performance measures in terms of steady-state probabilities. The states of the governing service model are defined by the pair

$$\Theta \equiv \{(j,n); j = 0,1,\cdots,R \text{ and } n = 0,1,\cdots,K\}$$

where the subscript j represents the number of available servers and n denotes the number of customers in the service system. In the steady-state condition, we use the following notation

$P_{j,n} \equiv$ Probability that there are n customers in the service system and $(R-j)$ servers are on vacation

where, $j = 0,1,\cdots,R$ and $n = 0,1,\cdots,K$.

The matrix-analytic method is a widely used approach to calculate the stationary distribution of a Markov chain having repeating structure after some point. It is a useful tool for constructing stochastic models and computing their probability distribution. The matrix-analytic method was first introduced by [200] to explore various features of the embedded Markov chain of many queueing problems.

2.2.1 Matrix Representation

For applying the matrix-analytic method, we define the probability vector $\Pi_n; n = 0, 1, \cdots, K$, a row vector whose j^{th} element is the steady-state probability $P_{j,n}$; $j = 0, 1, \cdots, R$, *i.e.* $\Pi_n = [P_{0,n}, P_{1,n}, \cdots, P_{R,n}]$. The corresponding block-tridiagonal structure of the transition rate matrix $\mathbf{Q}$ of the continuous-time Markov chain (CTMC) is represented as follows

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{B}_0 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{C}_1 & \mathbf{A}_1 & \mathbf{B}_1 & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \mathbf{A}_2 & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_R & \mathbf{B}_R & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{R+1} & \mathbf{A}_{R+1} & \mathbf{B}_R & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{R+2} & \mathbf{A}_{R+2} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{K-1} & \mathbf{B}_R \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_K & \mathbf{A}_K \end{bmatrix}$$

This rate matrix $\mathbf{Q}$ of the Markov process is analogous to the quasi-birth and death process. Each element of the rate matrix $\mathbf{Q}$ is a block square matrix of order $(R+1)$ that is represented as follows

$$\mathbf{A}_n = \begin{bmatrix} x_1^n & y_1^n & 0 & \cdots & 0 & 0 & 0 \\ z_2^n & x_2^n & y_2^n & \cdots & 0 & 0 & 0 \\ 0 & z_3^n & x_3^n & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & x_{R-1}^n & y_{R-1}^n & 0 \\ 0 & 0 & 0 & \cdots & z_R^n & x_R^n & y_R^n \\ 0 & 0 & 0 & \cdots & 0 & z_{R+1}^n & x_{R+1}^n \end{bmatrix}$$

We express each element of the above block matrix $\mathbf{A}_n$ for $0 \leq n \leq K$ in the terms of scalar values x_j^n, y_j^n, and z_j^n for different values of j and the closed form expression of x_j^n, y_j^n and z_j^n are as follows:

$$x_j^n = \begin{cases} -\left(\bar{\xi}\lambda + \theta_R\right); & n = 0 \;\&\; j = 1 \\ -\left\{\lambda + \theta_{R-j+1} + (j-1)\delta\right\}; & n = 0 \;\&\; 2 \leq j \leq R \\ -(\lambda + R\delta); & n = 0 \;\&\; j = R+1 \\ -\left[\bar{\xi}\lambda + \theta_{R-j+1} + (n-j+1)\zeta p_1 \right. & \\ \quad \left. +(j-1)\mu + (j-1)\delta\right]; & 1 \leq n \leq R-2 \;\&\; 1 \leq j \leq R \;\&\; j \leq n+1 \\ -\left[\lambda + \theta_{R-j+1} + n\mu + (j-1)\delta\right]; & 1 \leq n \leq R-2 \;\&\; 1 \leq j \leq R \;\&\; j > n+1 \\ -[\lambda + R\delta + n\mu]; & 1 \leq n \leq R-2 \;\&\; j = R+1 \\ -\left[\bar{\xi}\lambda + \theta_{R-j+1} + (R-j)\zeta p_1 \right. & \\ \quad \left. +(j-1)\mu + (j-1)\delta\right]; & n = R-1 \;\&\; 1 \leq j \leq R \\ -[\lambda + R\delta + (R-1)\mu]; & n = R-1 \;\&\; j = R+1 \\ -\left[\bar{\xi}\lambda + \theta_{R-j+1} + (n-j+1)\zeta p_1 \right. & \\ \quad \left. +(j-1)\mu_1 + (j-1)\delta\right]; & R \leq n \leq K-1 \;\&\; 1 \leq j \leq R \\ -\left[\bar{\xi}\lambda + (n-R)\zeta p_1 + R\mu_1 + R\delta\right]; & R \leq n \leq K-1 \;\&\; j = R+1 \\ -\left[\theta_{R-j+1} + (K-j+1)\zeta p_1 \right. & \\ \quad \left. +(j-1)\mu_1 + (j-1)\delta\right]; & n = K \;\&\; 1 \leq j \leq R \\ -[(K-R)\zeta p_1 + R\mu_1 + R\delta]; & n = K \;\&\; j = R+1 \end{cases}$$

$$y_j^n = \begin{cases} \theta_{R-j+1}; & 0 \leq n \leq K \;\&\; 1 \leq j \leq R \\ 0; & \text{otherwise} \end{cases}$$

and

$$z_j^n = \begin{cases} (j-1)\delta; & 0 \leq n \leq K \;\&\; 2 \leq j \leq R+1 \\ 0; & \text{otherwise} \end{cases}$$

Similarly, we represent the block square matrix $\mathbf{B}_n$ for $0 \leq n \leq R$ in the diagonal form as

$$\mathbf{B}_n = \begin{bmatrix} b_1^n & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & b_2^n & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & b_3^n & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{R-1}^n & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & b_R^n & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & b_{R+1}^n \end{bmatrix}$$

where,

$$b_j^n = \begin{cases} \bar{\xi}\lambda; & 0 \le n \le R \ \& \ 1 \le j \le R+1 \ \& \ j \le n+1 \\ \lambda; & \text{otherwise} \end{cases}$$

Similarly, the structure of the block square matrix $\mathbf{C}_n$ for $0 \le n \le K$ is represented as

$$\mathbf{C}_n = \begin{bmatrix} u_1^n & 0 & 0 & \cdots & 0 & 0 & 0 \\ v_2^n & u_2^n & 0 & \cdots & 0 & 0 & 0 \\ 0 & v_3^n & u_3^n & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & u_{R-1}^n & 0 & 0 \\ 0 & 0 & 0 & \cdots & v_R^n & u_R^n & 0 \\ 0 & 0 & 0 & \cdots & 0 & v_{R+1}^n & u_{R+1}^n \end{bmatrix}$$

where the closed form representation of u_j^n and v_j^n are given as

$$u_j^n = \begin{cases} [(j-1)\mu + (n-j+1)\zeta p_1]; & 1 \le n \le R-1 \ \& \ 1 \le j \le R+1 \ \& \ j \le n \\ nq_2\mu; & 1 \le n \le R-1 \ \& \ 1 \le j \le R+1 \ \& \ j > n \\ [(j-1)\mu_1 + (n-j+1)\zeta p_1]; & n = R \ \& \ 1 \le j \le R+1 \ \& \ j \le n \\ nq_2\mu_1; & n = R \ \& \ 1 \le j \le R+1 \ \& \ j > n \\ [(j-1)\mu_1 + (n-j+1)\zeta p_1]; & R+1 \le n \le K \ \& \ 1 \le j \le R+1 \end{cases}$$

and

$$v_j^n = \begin{cases} np_2\mu; & 1 \le n \le R-1 \ \& \ 2 \le j \le R+1 \ \& \ j > n \\ np_2\mu_1; & n = R \ \& \ 1 \le j \le R+1 \ \& \ j > n \\ 0; & \text{otherwise} \end{cases}$$

2.2.2 Matrix Analytic Solution Algorithm

Let Π represents the steady-state probability vector associated with the rate matrix $\mathbf{Q}$. Assuming the partition of the vector Π as $\Pi = [\Pi_0, \Pi_1, \Pi_2, \ldots, \Pi_{K-1}, \Pi_K]$, where each Π_n is further expressed in the vector form as $\Pi_n = [P_{0,n}, P_{1,n}, \cdots, P_{R-1,n}, P_{R,n}]$ for $n = 0, 1, 2, \cdots, K$, we have following homogeneous system of equations

$$\Pi\mathbf{Q} = \mathbf{0} \tag{2.1}$$

The governing homogenous system of the equation can easily be expressed in terms of pre-defined block matrix in the following steady-state matrix equations as

$$\Pi_0\mathbf{A}_0 + \Pi_1\mathbf{C}_1 = \mathbf{0} \tag{2.2}$$
$$\Pi_{n-1}\mathbf{B}_{n-1} + \Pi_n\mathbf{A}_n + \Pi_{n+1}\mathbf{C}_{n+1} = \mathbf{0};\ 1 \le n \le R \tag{2.3}$$
$$\Pi_{n-1}\mathbf{B}_R + \Pi_n\mathbf{A}_n + \Pi_{n+1}\mathbf{C}_{n+1} = \mathbf{0};\ R+1 \le n \le K-1 \tag{2.4}$$
$$\Pi_{K-1}\mathbf{B}_R + \Pi_K\mathbf{A}_K = \mathbf{0} \tag{2.5}$$

Now, after suitable matrix manipulation and recursive substitution, we obtain

$$\Pi_0 = \Pi_1\mathbf{C}_1\left(-\mathbf{A}_0^{-1}\right) = \Pi_1\mathbf{X}_0$$
$$\Pi_n = \Pi_{n+1}\mathbf{C}_{n+1}\left[-\left\{\mathbf{X}_{n-1}\mathbf{B}_{n-1} + \mathbf{A}_n\right\}^{-1}\right] = \Pi_{n+1}\mathbf{X}_n;\ n = 1, 2, \ldots, R$$
$$\Pi_n = \Pi_{n+1}\mathbf{C}_{n+1}\left[-\left\{\mathbf{X}_{n-1}\mathbf{B}_R + \mathbf{A}_n\right\}^{-1}\right] = \Pi_{n+1}\mathbf{X}_n;\ n = R+1, \ldots, K-1$$

i.e., the matrix notation $\mathbf{X}_n$ for $n = 0, 1, \cdots, K-1$ has the following closed form

$$\mathbf{X}_n = \begin{cases} -\mathbf{C}_1\mathbf{A}_0^{-1}; & n = 0 \\ -\mathbf{C}_{n+1}\left(\mathbf{X}_{n-1}\mathbf{B}_{n-1} + \mathbf{A}_n\right)^{-1}; & 1 \le n \le R \\ -\mathbf{C}_{n+1}\left(\mathbf{X}_{n-1}\mathbf{B}_R + \mathbf{A}_n\right)^{-1}; & R+1 \le n \le K-1 \end{cases}$$

Again by the recursive approach, we relate each steady-state probability vector Π_n in the product form of $\mathbf{X}_n; n = 0, 1, \cdots, K-1$ as

$$\Pi_n = \Pi_K\left\{\mathbf{X}_{K-1}\mathbf{X}_{K-2}\mathbf{X}_{K-3}\ldots\mathbf{X}_{n+2}\mathbf{X}_{n+1}\mathbf{X}_n\right\};\ n = 0, 1, 2, \ldots, K-1$$
$$\Pi_n = \Pi_K\prod_{i=n}^{K-1}\mathbf{X}_i = \Pi_K\Phi_n;\ n = 0, 1, 2, \ldots, K-1 \tag{2.6}$$

The normalizing condition of probability is $\Pi\mathbf{e} = 1$, where $\mathbf{e}$ is column vector of dimension $(K+1)R$ having all elements 1. Let $\mathbf{u}$ be also column vector of dimension

$(R+1)$ having all elements 1. Now, we get

$$\sum_{n=0}^{K} \Pi_n \mathbf{u} = 1$$
$$[\Pi_0 + \Pi_1 + \Pi_2 + \cdots + \Pi_{K-1} + \Pi_K]\,\mathbf{u} = 1$$
$$\Pi_K\,[\Phi_0 + \Phi_1 + \cdots + \Phi_{K-1} + I]\,\mathbf{u} = 1$$

So, the closed form representation of above relation is

$$\Pi_K \left[\sum_{n=0}^{K-1} \Phi_n + \mathbf{I} \right] \mathbf{u} = 1 \tag{2.7}$$

Therefore, Π_K is obtained by solving the eq^n(2.5) and eq^n(2.7). Hence, using the obtained Π_K and eq^n(2.6), all the other steady-state probabilities are also easily determined.

2.3 System Performance Measures

For the performance characterization of the queue-based service system, there are some standard system performance indices. We also employ some performance measures to delineate the modeling and methodology used for finite capacity multi-server service system with Bernoulli scheduled modified vacation and retention of reneged customers and to exhibit the parametric analysis for the decision purpose. These performance measures are quite correlated, and each recognizes as increased importance in a particular environment. In this section, we describe these system performance measures in matrix form by using the steady-state probabilities introduced in the previous section as follows:

- Expected number of the customers in the service system

$$\begin{aligned}
L_S &= \sum_{j=0}^{R} \sum_{n=0}^{K} n P_{j,n} \\
&= 0\Pi_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + 1\Pi_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + 2\Pi_2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \cdots + (K-1)\,\Pi_{K-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + K\,\Pi_K \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \sum_{n=0}^{K} n \Pi_n \mathbf{u}
\end{aligned} \tag{2.8}$$

where, each vector Π_n is a row vector of order $(R+1)$ having the vector form $\Pi_n = [P_{0,n}, P_{1,n}, P_{2,n}, \cdots, P_{R-1,n}, P_{R,n}]; \forall n = 0,1,2,\cdots,K.$

- Expected number of the customers in the queue

$$\begin{aligned}
L_Q &= \sum_{j=0}^{R}\sum_{n=j}^{K}(n-j)P_{j,n} \\
&= \Pi_0 \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \Pi_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \Pi_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \Pi_{K-1} \begin{bmatrix} K-1 \\ K-2 \\ K-3 \\ \vdots \\ K-R-1 \end{bmatrix} + \Pi_K \begin{bmatrix} K \\ K-1 \\ K-2 \\ \vdots \\ K-R \end{bmatrix} \\
&= \sum_{n=0}^{K} \Pi_n \mathbf{u}_n \qquad (2.9)
\end{aligned}$$

- Expected number of servers on vacation in the service system

$$\begin{aligned}
E(V) &= \sum_{j=0}^{R}\sum_{n=0}^{K}(R-j)P_{j,n} \\
&= \Pi_0 \begin{bmatrix} R \\ R-1 \\ R-2 \\ \vdots \\ 0 \end{bmatrix} + \Pi_1 \begin{bmatrix} R \\ R-1 \\ R-2 \\ \vdots \\ 0 \end{bmatrix} + \Pi_2 \begin{bmatrix} R \\ R-1 \\ R-2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \Pi_{K-1} \begin{bmatrix} R \\ R-1 \\ R-2 \\ \vdots \\ 0 \end{bmatrix} + \Pi_K \begin{bmatrix} R \\ R-1 \\ R-2 \\ \vdots \\ 0 \end{bmatrix} \\
&= \sum_{n=0}^{K} \Pi_n \mathbf{v}_1 \qquad (2.10)
\end{aligned}$$

- Expected number of idle servers in the service system

$$\begin{aligned}
E(I) &= \sum_{j=0}^{R}\sum_{n=0}^{j}(j-n)P_{j,n} \\
&= \Pi_0 \begin{bmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ R \end{bmatrix} + \Pi_1 \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ R-1 \end{bmatrix} + \Pi_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ R-2 \end{bmatrix} + \cdots + \Pi_{R-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} + \Pi_R \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\
&= \sum_{n=0}^{R-1} \Pi_n \mathbf{w}_n \qquad (2.11)
\end{aligned}$$

- Average balking rate

$$
\begin{aligned}
ABR &= \sum_{j=0}^{R-1}\sum_{n=j}^{R-1} \xi\lambda P_{j,n} + \sum_{j=0}^{R}\sum_{n=R}^{K-1} \xi\lambda P_{j,n} \\
&= \xi\lambda\,\Pi_0 \begin{bmatrix} 1\\0\\0\\ \vdots \\0\\0 \end{bmatrix} + \xi\lambda\,\Pi_1 \begin{bmatrix} 1\\1\\0\\ \vdots \\0\\0 \end{bmatrix} + \xi\lambda\,\Pi_2 \begin{bmatrix} 1\\1\\1\\ \vdots \\0\\0 \end{bmatrix} + \cdots + \xi\lambda\,\Pi_{R-1} \begin{bmatrix} 1\\1\\1\\ \vdots \\1\\0 \end{bmatrix} + \xi\lambda\,\Pi_R \begin{bmatrix} 1\\1\\1\\ \vdots \\1\\1 \end{bmatrix} + \cdots \\
&\qquad \cdots + \xi\lambda\,\Pi_{R+1} \begin{bmatrix} 1\\1\\1\\ \vdots \\1\\1 \end{bmatrix} + \cdots + \xi\lambda\,\Pi_{K-2} \begin{bmatrix} 1\\1\\1\\ \vdots \\1\\1 \end{bmatrix} + \xi\lambda\,\Pi_{K-1} \begin{bmatrix} 1\\1\\1\\ \vdots \\1\\1 \end{bmatrix} \\
&= \xi\lambda \sum_{n=0}^{R-1} \Pi_n \mathbf{a}_n + \xi\lambda \sum_{n=R}^{K-1} \Pi_n \mathbf{u} \qquad (2.12)
\end{aligned}
$$

- Average reneging rate

$$
\begin{aligned}
ARR &= \sum_{j=0}^{R-1}\sum_{n=j+1}^{R} (n-j)\zeta p_1 P_{j,n} + \sum_{j=0}^{R}\sum_{n=R+1}^{K} (n-j)\zeta p_1 P_{j,n} \\
&= \zeta p_1\,\Pi_0 \begin{bmatrix} 0\\0\\0\\ \vdots \\0\\0 \end{bmatrix} + \zeta p_1\,\Pi_1 \begin{bmatrix} 1\\0\\0\\ \vdots \\0\\0 \end{bmatrix} + \zeta p_1\,\Pi_2 \begin{bmatrix} 2\\1\\0\\ \vdots \\0\\0 \end{bmatrix} + \cdots
\end{aligned}
$$

$$
\begin{aligned}
&\cdots+\zeta p_1\Pi_{R-1}\begin{bmatrix}R-1\\R-2\\R-3\\\vdots\\0\\0\end{bmatrix}+\cdots+\zeta p_1\Pi_R\begin{bmatrix}R\\R-1\\R-2\\\vdots\\1\\0\end{bmatrix}\\
&\cdots+\zeta p_1\Pi_{R+1}\begin{bmatrix}R+1\\R\\R-1\\\vdots\\2\\1\end{bmatrix}+\cdots+\zeta p_1\Pi_K\begin{bmatrix}K\\K-1\\K-2\\\vdots\\K-R-1\\K-R\end{bmatrix}\\
&=\zeta p_1\sum_{n=0}^{K}\Pi_n\mathbf{u}_n
\end{aligned}\tag{2.13}
$$

- Throughput of the service system

$$
\begin{aligned}
\tau_P&=\sum_{j=0}^{R-1}\sum_{n=0}^{j}n\mu P_{j,n}+\sum_{j=0}^{R-2}\sum_{n=j+1}^{R-1}j\mu P_{j,n}+\sum_{n=0}^{R-1}n\mu P_{R,n}+\sum_{j=0}^{R}\sum_{n=R}^{K}j\mu_1 P_{j,n}\\
&=\mu\Pi_0\begin{bmatrix}0\\0\\0\\\vdots\\0\\0\end{bmatrix}+\mu\Pi_1\begin{bmatrix}0\\1\\1\\\vdots\\1\\1\end{bmatrix}+\mu\Pi_2\begin{bmatrix}0\\1\\2\\\vdots\\2\\2\end{bmatrix}+\cdots+\mu\Pi_{R-1}\begin{bmatrix}0\\1\\2\\\vdots\\R-1\\R-1\end{bmatrix}+\mu_1\Pi_R\begin{bmatrix}0\\1\\2\\\vdots\\R-1\\R\end{bmatrix}+\cdots\\
&\cdots+\mu_1\Pi_{R+1}\begin{bmatrix}0\\1\\2\\\vdots\\R-1\\R\end{bmatrix}+\cdots+\mu_1\Pi_{K-1}\begin{bmatrix}0\\1\\2\\\vdots\\R-1\\R\end{bmatrix}+\mu_1\Pi_K\begin{bmatrix}0\\1\\2\\\vdots\\R-1\\R\end{bmatrix}\\
&=\mu\sum_{n=0}^{R-1}\Pi_n\mathbf{b}_n+\mu_1\sum_{n=R}^{K}\Pi_n\mathbf{v}_2
\end{aligned}\tag{2.14}
$$

- Expected waiting time of the customer in the service system

$$W_S = \frac{L_S}{\lambda_{\text{eff}}} \tag{2.15}$$

where,

$$\begin{aligned}
\lambda_{\text{eff}} &= \sum_{j=1}^{R}\sum_{n=0}^{j-1} \lambda P_{j,n} + \sum_{j=0}^{R}\sum_{n=j}^{K-1} \bar{\xi}\lambda P_{j,n} \\
&= \lambda\,\Pi_0 \begin{bmatrix} \bar{\xi} \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} + \lambda\,\Pi_1 \begin{bmatrix} \bar{\xi} \\ \bar{\xi} \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} + \lambda\,\Pi_2 \begin{bmatrix} \bar{\xi} \\ \bar{\xi} \\ \bar{\xi} \\ \vdots \\ 1 \\ 1 \end{bmatrix} + \cdots + \lambda\,\Pi_{R-1} \begin{bmatrix} \bar{\xi} \\ \bar{\xi} \\ \bar{\xi} \\ \vdots \\ \bar{\xi} \\ 1 \end{bmatrix} \\
&\quad + \lambda\,\Pi_R \begin{bmatrix} \bar{\xi} \\ \bar{\xi} \\ \bar{\xi} \\ \vdots \\ \bar{\xi} \\ \bar{\xi} \end{bmatrix} + \cdots + \lambda\,\Pi_{K-1} \begin{bmatrix} \bar{\xi} \\ \bar{\xi} \\ \bar{\xi} \\ \vdots \\ \bar{\xi} \\ \bar{\xi} \end{bmatrix} \\
&= \lambda \sum_{n=0}^{R-1} \Pi_n \mathbf{c}_n + \bar{\xi}\lambda \sum_{n=R}^{K-1} \Pi_n \mathbf{u}, \text{ a effective arrival rate.}
\end{aligned}$$

The vectors $\mathbf{u}$, $\mathbf{u}_n$, $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{w}_n$, $\mathbf{a}_n$, $\mathbf{b}_n$ and $\mathbf{c}_n$ are the column vectors of dimension $(R+1)$. Following are the respective structures of these vectors with corresponding $j^{th}; j = 1,2,\cdots,R+1$ element

$$\begin{aligned}
&\mathbf{u} = (u_j) \text{ s.t.} \, u_j = 1 \\
&\mathbf{u}_n = (u_{nj}) \text{ s.t.} \, u_{nj} = \max\{n-j+1, 0\}; \, 0 \le n \le K \\
&\mathbf{v}_1 = (v_{1j}) \text{ s.t.} \, v_{1j} = R+1-j \\
&\mathbf{v}_2 = (v_{2j}) \text{ s.t.} \, v_{2j} = j-1 \\
&\mathbf{w}_n = (w_{nj}) \text{ s.t.} \, w_{nj} = \max\{j-n-1, 0\}; \, 0 \le n \le R-1 \\
&\mathbf{a}_n = (a_{nj}) \text{ s.t.} \, a_{nj} = \begin{cases} 1; & 0 \le n \le R-1 \,\&\, j \le n+1 \\ 0; & 0 \le n \le R-1 \,\&\, j > n+1 \end{cases}
\end{aligned}$$

$$\mathbf{b}_n = \left(b_{nj}\right) \text{ s.t. } b_{nj} = \begin{cases} 0; & n = 0 \\ (j-1); & 1 \le n \le R-1 \,\&\, j \le n+1 \\ n; & 1 \le n \le R-1 \,\&\, j > n+1 \end{cases}$$

$$\mathbf{c}_n = \left(c_{nj}\right) \text{ s.t. } c_{nj} = \begin{cases} \bar{\xi}; & 0 \le n \le R-1 \,\&\, j \le n+1 \\ 1; & 0 \le n \le R-1 \,\&\, j > n+1 \end{cases}$$

2.4 Cost Analysis

Cost analysis is a systematic approach to estimate the strengths and weaknesses of decision variables. It is used to govern alternatives that provide the best approach to achieve benefits from the service system. Cost analysis is the act of characterizing a cost incurred into its constituents and studying and reporting on each factor. For this purpose, we formulate a cost function for the governing model, in which four decision variables namely, number of servers (R), service rates (μ, μ_1), and mean vacation rate (θ) are used. The decision variable R is intuitively to be a positive integer (Z^+) and the continuous decision variables μ, μ_1, and θ are non-negative real numbers. Our main objective is to find out the optimal number of servers in the service system (R^*), the optimal values of service rates (μ^*, μ_1^*), and optimal vacation rate (θ^*) so as to minimize the expected total cost. Moreover, we use various unit cost elements associated with several states of the service system, having a significant idea about the system performance. The definition of the governing unit cost elements are as follows

$C_h \equiv$ Unit holding cost for each customer present in the service system

$C_v \equiv$ Unit cost associated with each server on vacation

$C_i \equiv$ Unit cost associated with each idle server

$C_m \equiv$ Unit cost for offering service with rate μ for each customer present in the service system

$C_{m_1} \equiv$ Unit cost for offering service with rate μ_1 for each customer present in the service system

$C_t \equiv$ Unit cost associated with parameter θ for the server on vacation

$C_r \equiv$ Unit cost associated with each server

$C_d \equiv$ Unit cost associated with each emergent condition which prompts vacation of the server

$C_1 \equiv$ Unit cost associated with each balking and reneging of customers

Using the above-defined cost elements and derived performance measures in the previous section, we develop the following expected total cost function

$$\begin{aligned} TC(R,\mu,\mu_1,\theta) = C_h L_S + C_i E(I) - C_v E(V) + C_m \mu + C_{m_1}\mu_1 + C_t \theta + C_r R + C_d \delta \\ + C_1(ABR + ARR);\ R \in Z^+ \text{ and } \mu,\mu_1,\theta \in R^+ \cup \{0\} \end{aligned} \tag{2.16}$$

The cost minimization problem of the designed model is formulated mathematically as a constraint problem

$$\begin{aligned} & TC(R^*,\mu^*,\mu_1^*,\theta^*) = \underset{(R,\mu,\mu_1,\theta)}{\text{minimize}}\ TC(R,\mu,\mu_1,\theta) \\ & \text{subject to} \qquad 0 < \mu < \mu_1 \end{aligned} \tag{2.17}$$

The cost elements listed above are considered to be linear in nature. Due to extremely high non-linearity and complexity of the expected total cost function, it would have been an exhausting work to analytically establish the optimal values of R^*, μ^*, μ_1^*, and θ^*. To compute the optimal values of decision variables with a global minimum of the expected total cost (TC), we use metaheuristic optimizing technique particle swarm optimization based on swarm intelligence.

2.5 Special Cases

For the validation and tractability of the investigated model, by relaxing one or more assumptions, our results resemble with the results available in the existing literature.

Case 1: For $\xi = 0$ and $\zeta = 0$, our model and results coincide with the results of the model proposed by [131] . The model reduced to a multi-server queueing model with a modified Bernoulli vacation.

Case 2: By setting the parameters value as $\theta = 0$ and $q_1 = 0$, our model will equivalent to multi-server queueing model with balking and reneging which is proposed by [8]. They have presented the transient solution for queue size distribution.

Case 3: In the case 1 when $p_2 = 1$ and $q_1 = 0$, the model further deduced to multi-server model with single vacation studied by [296] having very vast application.
Case 4: If we set $\xi = 0$, our model exhibits the results which was obtained by [9]. They studied both single and multiple vacations for the server independently with reneging and compared the results for $M/M/1$, $M/M/c$ and $M/G/1$ queueing models.
Case 5: For $R = 1$ and $q_2 = 0$, the present model reduces to single server finite capacity queueing model with balking, reneging and single vacation which was investigate by [323] in past. Their numerical results are identical to the results of the present model.
Case 6: The infinite capacity queueing model with balking, reneging and, the single vacation was investigated by [189], which is a special case of case 5.

2.6 Particle Swarm Optimization

For the optimal analysis of the multi-server service system with the impatient behavior of customers and Bernoulli's scheduled modified vacation, we have implemented the PSO technique in this section. The numerical simulation has been performed for several combinations of system parameters, and results are depicted with the help of different tables and several generations of the PSO algorithm. For more details about the PSO technique, see the section 1.10.3 and its pseudo-code. The discrete and continuous system design parameters R, μ, μ_1, and θ are referred as x_1, x_2, x_3, and x_4 respectively and the expected cost function TC as the objective function f.

2.7 Numerical Results

In this section, we perform the numerical simulation for the state-dependent finite capacity multi-server service system with Bernoulli scheduled modified vacation and retention of the reneged customer. We observe the results of the various system performance indices by calibrating the system parameters, particularly the mean vacation rate (θ) and service rates (μ, μ_1), etc. We validate our formulation of the service model and methodology used by providing numerical illustrations through various graphs and tables, which is helpful in system design, assessment and operation and to the system designers, decision-makers in improving the system performance.

For the analysis purpose, we fix the default values of the system parameters as K=15, R=2, λ=5, μ=7.5, μ_1=10, θ=1.5, δ=0.05, ξ=0.3, ζ=0.7, p_1=0.5, and p_2=0.5. With

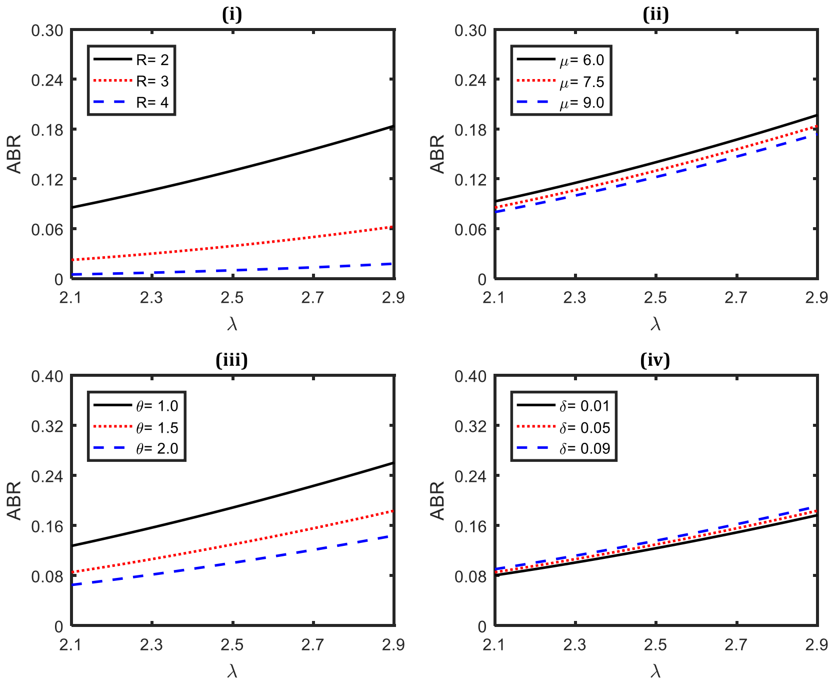

Figure 2.1: Average balking rate of customers in the service system wrt λ for (i) R, (ii) μ, (iii) θ, and (iv) δ.

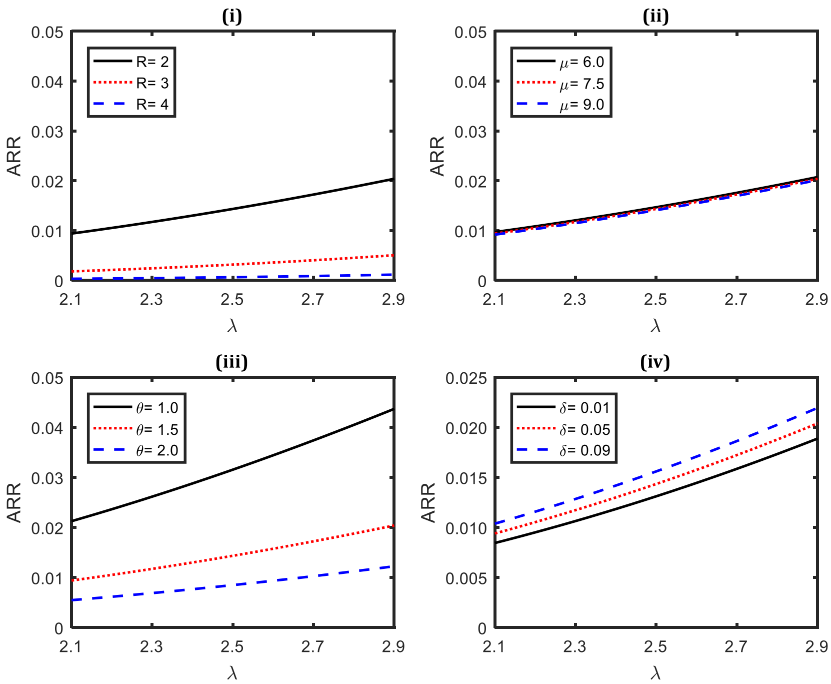

Figure 2.2: Average reneging rate of customers in the service system wrt λ for (i) R, (ii) μ, (iii) θ, and (iv) δ.

the aid of MATLAB software, the steady-state probabilities of various states are exhibited numerically using the matrix-analytic method. In Fig. 2.1, we vary the values of parameters R, μ, θ, and δ for varied λ and review the effect on the average balking rate of customers in the service system. It is noted from each sub-figure of Fig. 2.1

that when we increase the value of λ, the average balking rate of customers increases, which is quite obvious. Also for any test instance, when we fix the value of λ and increase the values of R and μ, the value of average balking rate decreases, which is depicted in the respective Figs. 2.1(i)-(ii). The change is factual because as the value of R and μ increases, the balking probability of the customer decreases. Fig. 2.1(iii) shows that as we increase the value of θ, the value of average balking rate decreases. The result is apparent from the fact that means vacation time is inversely proportional to the vacation mean rate θ, so the incremental change in θ decreases the vacation times, which results in a decrease in the probability of balking as expected. Moreover, the incremental change in δ increases the value of the average balking rate as intuitively expected which is depicted in Fig. 2.1(iv). Similarly, we also observe the effects of these parameters on the average reneging rate of customers in the service system in Fig. 2.2. With these illustrations, system analyst may get complete information about the impatience behavior of the customers and may opt some measures to reduce to some extent.

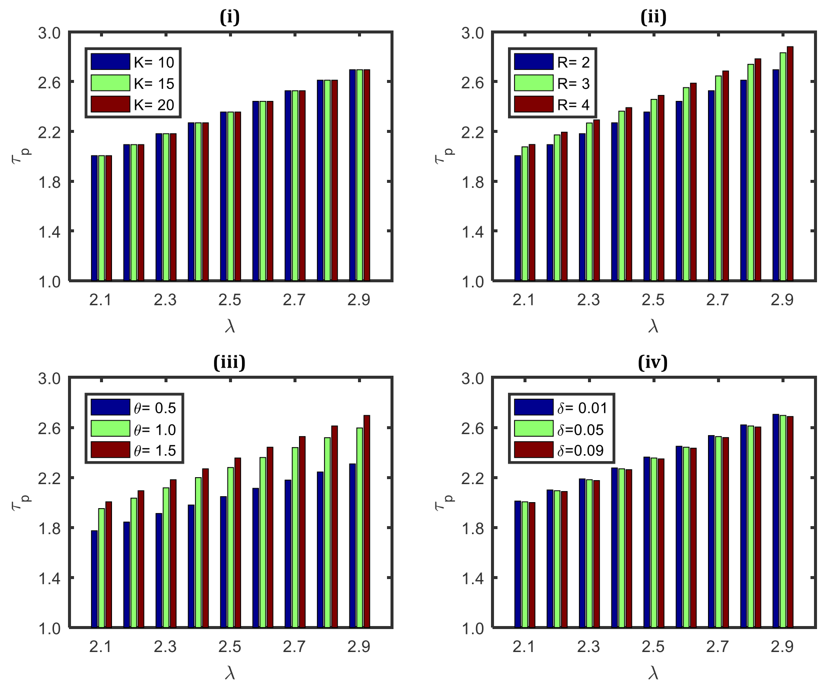

Figure 2.3: Throughput of the service system (τ_p) wrt λ for (i) K, (ii) R, (iii) θ, and (iv) δ.

Figs. 2.3 and 2.4 show the impact of parameters K, R, θ, and δ with varied arrival rate λ and service rate μ respectively on the throughput of the service system (τ_p). We observe from Fig. 2.3 that a larger value of throughput is obtained if a number of customers in the service system is more. From Fig. 2.3(i), it is clear that for a fixed value of λ, the higher value of K is insufficient to affect the throughput of the service system and hence designed model is very much suitable to the system

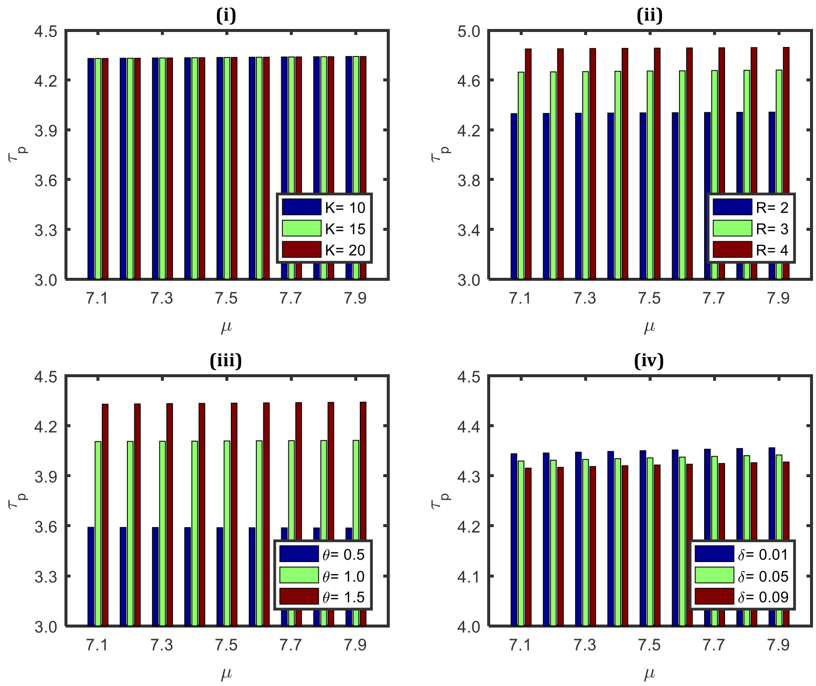

Figure 2.4: Throughput of the service system (τ_p) wrt μ for (i) K, (ii) R, (iii) θ, and (iv) δ.

analyst for decision purpose. From Figs. 2.3(ii)-(iii), it can be concluded that for any fixed value of λ, the increasing values of R and θ respectively lead to the increase in throughput of the service system while for the increasing values of δ, the value of throughput slight decreases as in Fig. 2.3(iv). Similarly, it appears from Fig. 2.4 that as we increase the value of μ, the value of throughput of the service system increases, which is quite obvious and on fixing the value of μ, we observe that the increasing values of R increase the value of throughput too. In the same fashion, the increasing values of θ impact the throughput of the service system in an increasing manner because of inverse proportionality of mean vacation time with the vacation mean rate θ. From Fig. 2.4(iv), it is noticeable that the throughput (τ_p) is unaffected for higher values of μ with some fix values of δ but it decreases for increasing values of δ as we fix the value of μ. So, as the conclusive remark, an extra effort for maintaining the higher service rate is unworthy and hence there is a requirement to incorporate the optimal service rate.

For the different pairs of system parameters, the change in the value of the expected total cost (TC) defined in eq^n(2.16) is demonstrated in Figs. 2.5 and 2.6 respectively. To calculate the expected total cost (TC), first we choose following unit cost elements $C_h = 200$, $C_m = 15$, $C_{m_1} = 25$, $C_t = 10$, $C_r = 15$, $C_v = 80$, $C_i = 20$, $C_d = 5$, and $C_1 = 150$ as default cost elements. Fig. 2.5(i) represents the effect of varied values of λ and K simultaneously on TC. As intuitively anticipated, the expected cost TC increases with the incremental change in these parameters. From

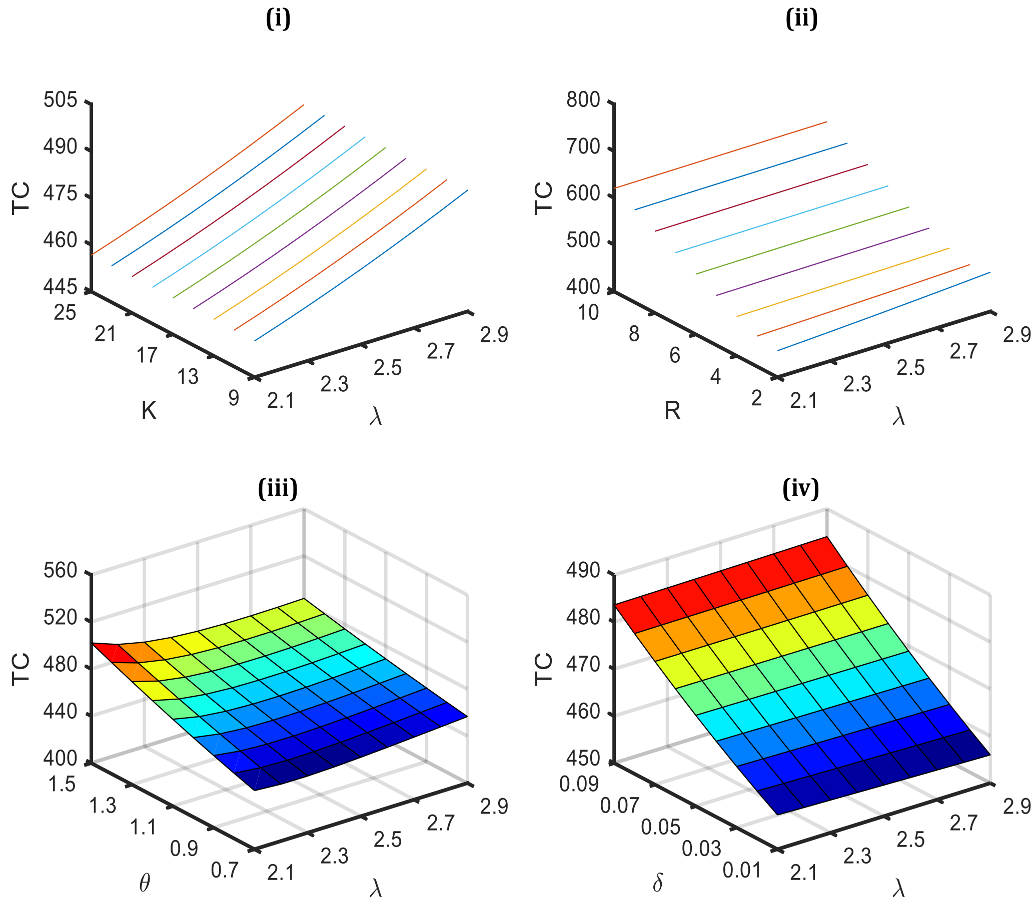

Figure 2.5: Expected total cost of the service system (TC) wrt system parameters (i) (K, λ), (ii) (R, λ), (iii) (θ, λ), and (iv) (δ, λ).

Fig. 2.5(iii), we observe that for smaller values of λ and higher values of θ, the expected total cost TC is calculated higher as compared to others. Similarly, it is consistent with our intuition that the value of TC increases with the increasing values of μ and K as in Fig. 2.6(i) but in Fig. 2.6(ii), we observe that first, the value of TC decreases with the increasing values of R and μ simultaneously and then increases more rapidly. It shows that we are nearby the optimal expected cost along with the optimal value of parameter R. In a similar way, we can define these results for the rest of all the figures also. So, each of these figures incites that all governing system parameters are commendable in system designing and play a key role in the development of studied service model.

We perform varied numerical experiments to observe the effect on various system performance measures of varying values of the several system parameters and results are summarized in Table 2.1 and 2.2. In Table 2.1, we can see that if we increase the value of K, the value of expected number of customers in the service system (L_S), effective arrival rate (λ_{eff}) and expected waiting time in the service system

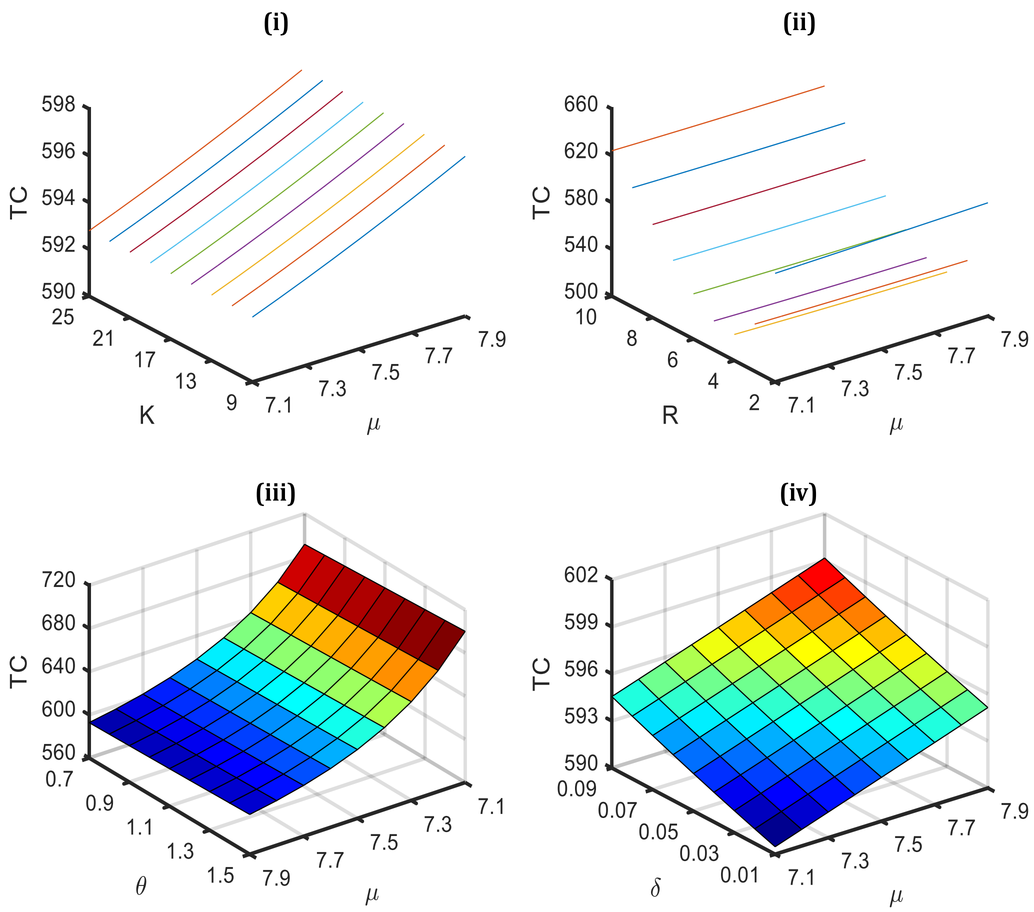

Figure 2.6: Expected total cost of the service system (TC) wrt system parameters (i) (K, μ), (ii) (R, μ), (iii) (θ, μ), and (iv) (δ, μ).

Table 2.1: Numerical simulation of various system characteristics wrt K, λ, μ, and θ.

(K,λ,μ,θ)	L_S	$E(V)$	$E(I)$	τ_p	λ_{eff}	W_S
(10, 5.0, 6.5, 1.5)	0.74802	0.62941	0.81789	4.31948	4.38784	0.17047
(15, 5.0, 6.5, 1.5)	0.74805	0.62940	0.81789	4.31951	4.38788	0.17048
(20, 5.0, 6.5, 1.5)	0.74805	0.62940	0.81789	4.31951	4.38788	0.17048
(10, 5.5, 6.5, 1.5)	0.82837	0.65635	0.75289	4.68014	4.76331	0.17391
(10, 6.0, 6.5, 1.5)	0.91022	0.67873	0.69419	5.03272	5.13182	0.17737
(10, 5.0, 7.0, 1.5)	0.72806	0.63666	0.83115	4.32797	4.39652	0.16560
(10, 5.0, 7.5, 1.5)	0.70963	0.64338	0.84337	4.33581	4.40455	0.16111
(10, 5.0, 6.5, 2.0)	0.69770	0.50147	0.92844	4.44065	4.48531	0.15555
(10, 5.0, 6.5, 2.5)	0.67458	0.41527	1.00424	4.51624	4.54917	0.14829

(W_S) increases while the throughput of the service system (τ_p) and the value of the expected number of servers on vacation $(E(V))$ decreases. Also, if we increase the value of λ, the value of L_S increases and $E(I)$ decreases as obvious. Table 2.2 shows that (i) L_S and W_S decrease while $E(V)$, $E(I)$, and τ_p increase as R increases. (ii)

Table 2.2: Numerical simulation of various system characteristics wrt R, δ, ξ, and ζ.

(R,δ,ξ,ζ)	L_S	$E(V)$	$E(I)$	τ_p	λ_{eff}	W_S
(2, 0.01, 0.2, 0.5)	0.76284	0.62592	0.84089	4.54198	4.59939	0.16586
(3, 0.01, 0.2, 0.5)	0.69546	0.82921	1.55544	4.78064	4.80066	0.14487
(4, 0.01, 0.2, 0.5)	0.67702	0.95033	2.39998	4.90738	4.91421	0.13777
(2, 0.05, 0.2, 0.5)	0.77217	0.64077	0.82757	4.53152	4.59165	0.16817
(2, 0.09, 0.2, 0.5)	0.78161	0.65532	0.81454	4.52117	4.58404	0.17051
(2, 0.01, 0.3, 0.5)	0.70884	0.62825	0.85508	4.36637	4.41442	0.16057
(2, 0.01, 0.4, 0.5)	0.65843	0.62987	0.86916	4.20015	4.23952	0.15531
(2, 0.01, 0.2, 0.7)	0.75509	0.62626	0.84240	4.52217	4.60049	0.16413
(2, 0.01, 0.2, 0.9)	0.74782	0.62657	0.84387	4.50332	4.60154	0.16251

L_S, $E(V)$, W_S increase and $E(I)$, τ_p decrease with increasing values of δ. (iii) L_S, τ_p decrease and $E(V)$, $E(I)$ increase with the increasing values of ζ. So as the concluding remark these research findings would help the system analyst to make a better decision keeping in mind the objectives of the problem and to provide the optimal service strategy based on the desired performance measures.

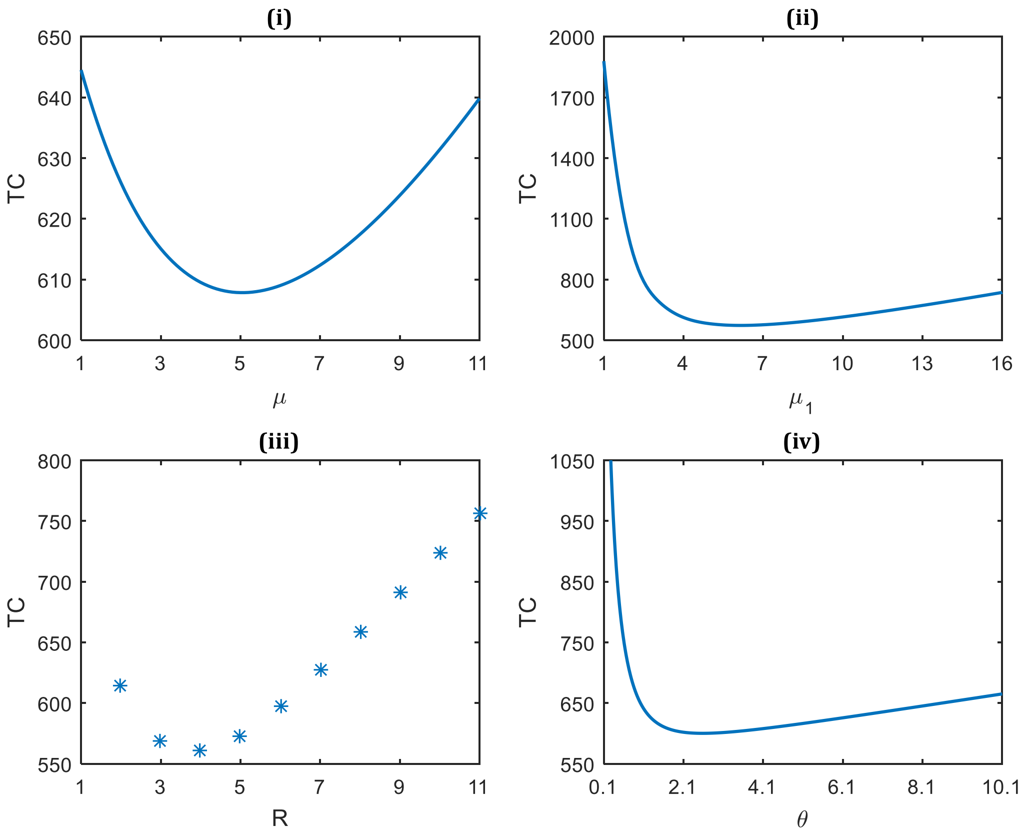

Figure 2.7: Convex expected total cost (TC) wrt decision parameters (i) μ, (ii) μ_1, (iii) R, and (iv) θ.

From above graphs and tables, it is observed that there is need to evaluate optimal service strategy to expect minimum incured expected total cost. Now, to find the optimal values of the decision parameters R, μ, μ_1, and θ, we also provide several numerical experiments to verify the convexity of the developed cost function since it is an arduous task to obtain the analytical expressions of R^*, μ^*, μ_1^*, and θ^* due to the extreme convexity and non-linearity of the cost function. So in the context of this purpose, the values of various system parameters along with several cost elements associated with their respective performance measures, are set as follows: $K = 15$, $R = 2$, $\lambda = 5$, $\mu = 7.5$, $\mu_1 = 10$, $\theta = 1.5$, $\delta = 0.05$, $\xi = 0.3$, $\zeta = 0.7$, $p_1 = 0.5$, $p_2 = 0.5$, $C_h = 200$, $C_r = 15$, $C_v = 80$, $C_i = 20$, $C_d = 5$, $C_m = 15$, $C_{m_1} = 25$, $C_t = 10$, and $C_1 = 150$. The lower and upper limits of decision variables μ, μ_1 are taken as [1 11] and [1 16] respectively. Similarly, the lower and upper limits of R, θ are taken as [1 11] and [0.1 10], respectively. The varied values of decision parameters R, μ, μ_1, and θ are shown graphically in Fig. 2.7, which leads that the desired expected cost function $TC(R, \mu, \mu_1, \theta)$ is convex in nature.

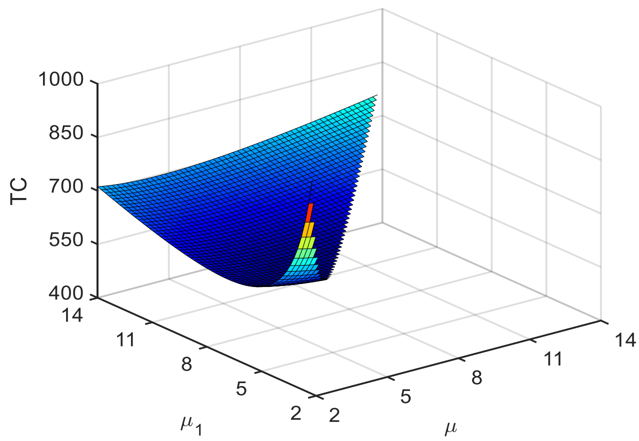

Figure 2.8: Surface plot for TC wrt combination of (μ, μ_1).

From Figs. 2.8 and 2.9, we see that the expected cost function (TC) is very much convex in nature for the joint values of decision parameters (μ, μ_1) and (μ, θ) respectively. Therefore, by combining results of surface plots and line plots in Figs. 2.7–2.9 we can infer that the expected total cost function $TC(R, \mu, \mu_1, \theta)$ is convex wrt to combined values of R and all continuous decision parameters μ, μ_1, and θ.

PSO and many other metaheuristic optimization techniques are generation based, as discussed in the previous section in detail. We can examine the inherent characteristics of PSO that the interactions among particles are performed with the shared knowledge on the best position obtained by neighbors. When a search particle within the neighborhood sets up a position with an optimum local value, which is better than neighbors' value, the other particles will make corresponding adjustments and tend to achieve nearest to that position. So, as the generations pass, all the search particles

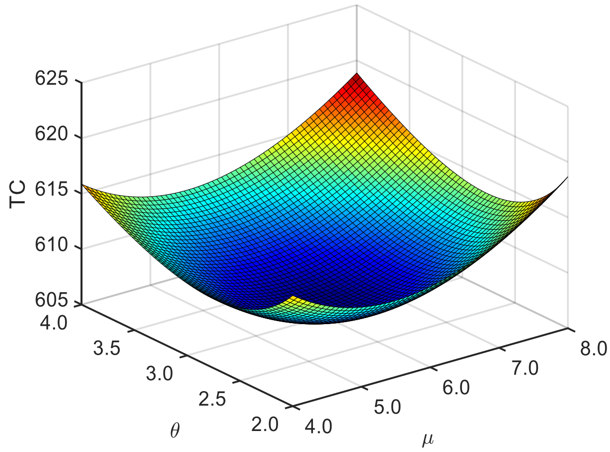

Figure 2.9: Surface plot for TC wrt combination of (μ, θ).

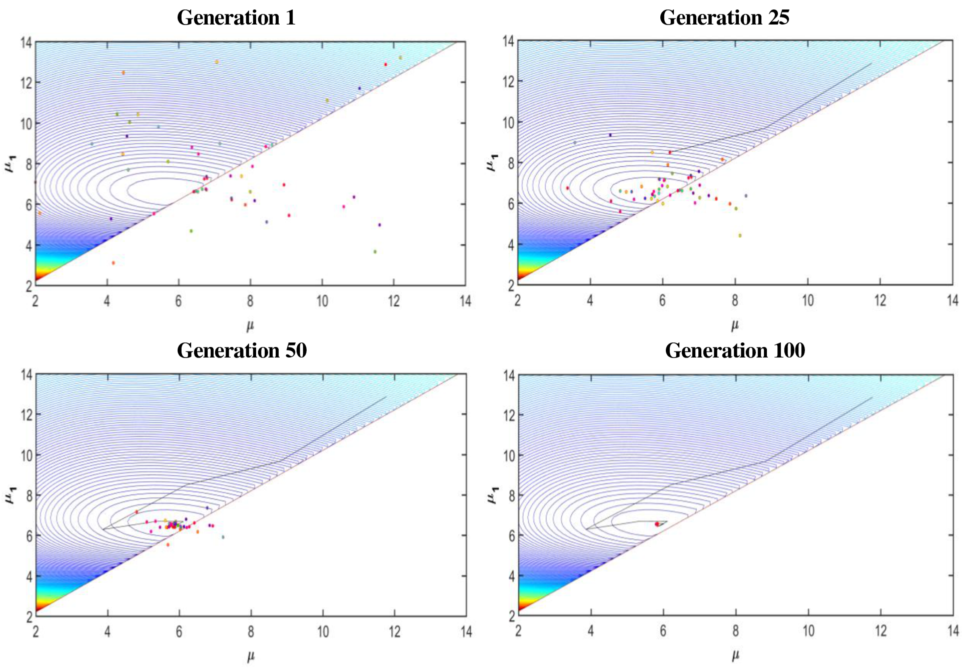

Figure 2.10: Several generations of PSO algorithm on contour of $TC(R, \mu, \mu_1, \theta)$ wrt μ and μ_1.

converge to the optimum and best of which is defined as the position with optimal values of the decision variables for a minimum expected total cost $TC(R, \mu, \mu_1, \theta)$ as shown in Figs. 2.10–2.11.

Because the PSO technique does not incorporate the computation of gradient, it is a convenient and flexible technique for optimization of single/multimodal complex problems with non-differentiable objective function. Furthermore, it can be used to search for the optimal values of discrete decision variables and continuous decision variables at the same instant. Now, to examine the validity and the performance of PSO algorithm for developed cost function, we provide some numerical experiments

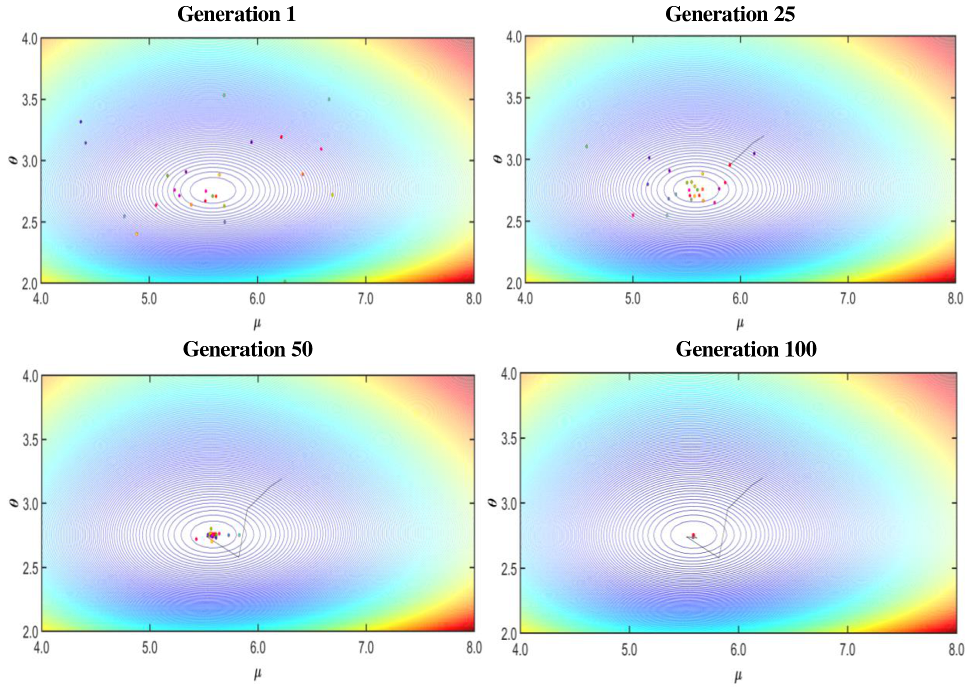

Figure 2.11: Several generations of PSO algorithm on contour of $TC(R,\mu,\mu_1,\theta)$ wrt μ and θ.

in Tables 2.3 and 2.4 which consider the following cost elements $C_h = 200$, $C_m = 15$, $C_{m_1} = 25$, $C_t = 10$, $C_r = 15$, $C_v = 80$, $C_i = 20$, $C_d = 5$, and $C_1 = 150$ with the default parameters of PSO algorithm: $\kappa_1 = 2$, $\kappa_2 = 2$, and $\omega_2 = 0.5$. The value of κ_1, κ_2, and ω_2 are taken from the references cited in previous section for better exploration and exploitation in search space. If we choose lower value, prospective solution point stuck locally and search the local optimal very slowly and if we choose higher value, searching in space proceed very fast and many good solution point overshoot and may miss even global optimal. We set the lower and upper limits for both μ, μ_1 in the interval [2 15] and for θ in the interval [0.5 8.0] respectively. Also, we set the number of servers (R) in the range [2, 10] and obtained the optimal values of the continuous decision variables up to the sixth place of decimal as in both the Tables. These bounds for decision parameters are decided in such way that multi-dimensional search space should be convex in nature, feasible for computation and appropriate for proper exploration and exploitation. Such search space can be determined with a number of trial experiments. We can easily depict from Table 2.3 that as the value of θ increases, (R^*,μ^*,μ_1^*) decreases while the cost of the developed service model increases. The increasing values of (R^*,μ^*,μ_1^*) are obtained as we increase the value of λ. In the same manner, as we increase the values of δ and ξ, R^* keeps the same optimal value but the values of μ^*,μ_1^* increases respectively. Similarly, we can analyze all the other results for Table 2.4.

We depict the numerical results in Table 2.3 and 2.4 by taking 50 random generated particles, 100 independent generations, and 20 independent runs for each experiment by the PSO algorithm. From literature, we decide these values for the number of particles, generations, and runs. The lower value does not do complete exploitation and exploration and has lack of randomness. Higher value makes searching very slow and requires computationally efficient computing machine since it calculates the objective function $\#_{\text{particles}} \times \#_{\text{generations}} \times \#_{\text{runs}}$. We also use the concept of statistical inferences namely mean and maximum of the ratio of optimal TC in all runs of the algorithm, to show the robustness of the PSO algorithm. Moreover, we examine from both of the Tables that the mean and max values of $\left(\frac{TC}{TC^*}\right)$ lie between 1.00000232 and 1.00056970, where TC is the optimal value of the objective function obtained by PSO algorithm and TC^* is the best (minimum) solution among 20 independent runs of PSO. It implies that the PSO technique is potent for all of the test instances and the searching quality of PSO algorithm is outstanding.

2.8 Conclusion and Future Prospective

In this chapter, we have examined a finite capacity multi-server service system with a retention policy of reneged customers and Bernoulli scheduled modified vacation policy. To determine the steady-state probabilities, we have employed the matrix-analytic method, and hence computed various system performance measures in vector form. We have also developed a cost function and formulate the cost minimization constraint problem. To examine the optimal values of decision parameters R, μ, μ_1, and θ with the optimal stability condition and a global minimum of the cost function, we have used a metaheuristic optimization algorithm: PSO with the aid of MATLAB software. Further, we have also presented various generations and surface plots for the pairs of decision parameters (μ, μ_1) and (μ, θ) by the PSO algorithm to show the robustness of PSO algorithm in the context of providing the converging results. The numerical simulation of various system performance measures has been accomplished to study the effects of all the system parameters. Finally, several numerical experiments have been provided to illustrate and achieve optimal solutions.

The cost analysis signifies the validity and profitability of the developed model in a very effective manner and will be helpful to the system designers and decision-makers in minimizing the cost of service, which is a highly desired trait of any organization. This work can be extended by incorporating some more features, such as bulk arrival and synchronous vacation and can also be modeled for machine interference problem.

Table 2.3: Optimal expected total cost for R^*, μ^*, and μ_1^*.

$(K,\theta,\lambda,\delta,\xi,\zeta,p_1,p_2)$	R^*	μ^*	μ_1^*	TC^*	Mean $\left\{\frac{TC}{TC^*}\right\}$	Max $\left\{\frac{TC}{TC^*}\right\}$
(15, 1.5, 5.0, 0.05, 0.3, 0.7, 0.5, 0.7)	5	5.193842	5.693922	460.59987	1.00014523	1.00036772
(20, 1.5, 5.0, 0.05, 0.3, 0.7, 0.5, 0.7)	5	5.171721	5.671732	460.59362	1.00001515	1.00004885
(25, 1.5, 5.0, 0.05, 0.3, 0.7, 0.5, 0.7)	5	5.203232	5.703274	460.60318	1.00003230	1.00010696
(15, 2.0, 5.0, 0.05, 0.3, 0.7, 0.5, 0.7)	4	5.202133	5.702535	485.72671	1.00006074	1.00019731
(15, 2.5, 5.0, 0.05, 0.3, 0.7, 0.5, 0.7)	4	5.166356	5.666484	501.29345	1.00005969	1.00020616
(15, 1.5, 5.5, 0.05, 0.3, 0.7, 0.5, 0.7)	5	5.473370	5.973777	475.29117	1.00014382	1.00044979
(15, 1.5, 6.0, 0.05, 0.3, 0.7, 0.5, 0.7)	6	5.666422	6.166971	488.47843	1.00006459	1.00015754
(15, 1.5, 5.0, 0.10, 0.3, 0.7, 0.5, 0.7)	5	5.201614	5.701768	455.96482	1.00010937	1.00034245
(15, 1.5, 5.0, 0.15, 0.3, 0.7, 0.5, 0.7)	5	5.230113	5.730176	451.16686	1.00004501	1.00007345
(15, 1.5, 5.0, 0.05, 0.5, 0.7, 0.5, 0.7)	5	5.216774	5.717873	468.32566	1.00009441	1.00036129
(15, 1.5, 5.0, 0.05, 0.7, 0.7, 0.5, 0.7)	5	5.225122	5.725173	475.17661	1.00005569	1.00019532
(15, 1.5, 5.0, 0.05, 0.3, 0.3, 0.5, 0.7)	5	5.168750	5.669237	460.15646	1.00005218	1.00018596
(15, 1.5, 5.0, 0.05, 0.3, 0.5, 0.5, 0.7)	5	5.161368	5.661368	460.37196	1.00005904	1.00016985
(15, 1.5, 5.0, 0.05, 0.3, 0.7, 0.3, 0.7)	4	5.323611	5.823639	467.69326	1.00001602	1.00002481
(15, 1.5, 5.0, 0.05, 0.3, 0.7, 0.7, 0.7)	4	5.180638	5.680639	460.89583	1.00005485	1.00011278
(15, 1.5, 5.0, 0.05, 0.3, 0.7, 0.5, 0.3)	4	5.085715	5.585837	505.36024	1.00001217	1.00002842
(15, 1.5, 5.0, 0.05, 0.3, 0.7, 0.5, 0.5)	4	5.263157	5.763355	484.15135	1.00002719	1.00008311

Table 2.4: Optimal expected total cost for μ^*, μ_1^*, and θ^*.

$(K,R,\lambda,\delta,\xi,\zeta,p_1,p_2)$	μ^*	μ_1^*	θ^*	TC^*	Mean $\left\{\frac{TC}{TC^*}\right\}$	Max $\left\{\frac{TC}{TC^*}\right\}$
(15, 2, 5.0, 0.05, 0.3, 0.7, 0.5, 0.5)	5.785978	6.286288	2.666312	554.24364	1.00000287	1.00000421
(20, 2, 5.0, 0.05, 0.3, 0.7, 0.5, 0.5)	5.799533	6.299718	2.680370	554.24031	1.00000395	1.00000897
(25, 2, 5.0, 0.05, 0.3, 0.7, 0.5, 0.5)	5.823453	6.323528	2.672808	554.24011	1.00000271	1.00000986
(15, 3, 5.0, 0.05, 0.3, 0.7, 0.5, 0.5)	5.446347	5.946669	1.441888	502.93985	1.00017624	1.00041796
(15, 4, 5.0, 0.05, 0.3, 0.7, 0.5, 0.5)	5.550451	6.050885	0.820234	461.99804	1.00024538	1.00056970
(15, 2, 5.5, 0.05, 0.3, 0.7, 0.5, 0.5)	6.219710	6.719719	2.973011	589.42423	1.00001086	1.00003033
(15, 2, 6.0, 0.05, 0.3, 0.7, 0.5, 0.5)	6.630894	7.130969	3.279414	624.08982	1.00000781	1.00002461
(15, 2, 5.0, 0.10, 0.3, 0.7, 0.5, 0.5)	5.821786	6.322794	2.725657	556.04741	1.00000928	1.00002315
(15, 2, 5.0, 0.15, 0.3, 0.7, 0.5, 0.5)	5.844926	6.345254	2.803383	557.80824	1.00003461	1.00012736
(15, 2, 5.0, 0.05, 0.5, 0.7, 0.5, 0.5)	5.669801	6.170191	2.797450	575.83652	1.00005854	1.00022254
(15, 2, 5.0, 0.05, 0.7, 0.7, 0.5, 0.5)	5.610778	6.111249	2.973040	595.55291	1.00007196	1.00026642
(15, 2, 5.0, 0.05, 0.3, 0.5, 0.5, 0.5)	5.808808	6.309664	2.669619	553.43042	1.00000232	1.00000493
(15, 2, 5.0, 0.05, 0.3, 0.9, 0.5, 0.5)	5.808909	6.309013	2.675202	555.02691	1.00001097	1.00002311
(15, 2, 5.0, 0.05, 0.3, 0.7, 0.3, 0.5)	5.812826	6.313019	2.642338	553.09873	1.00000329	1.00000753
(15, 2, 5.0, 0.05, 0.3, 0.7, 0.7, 0.5)	5.799808	6.300189	2.685496	555.33646	1.00001368	1.00004020
(15, 2, 5.0, 0.05, 0.3, 0.7, 0.5, 0.3)	5.816812	6.317704	2.195232	548.30094	1.00001442	1.00003097
(15, 2, 5.0, 0.05, 0.3, 0.7, 0.5, 0.7)	5.792581	6.294541	3.033639	559.00744	1.00003704	1.00005364