PERFORMANCE ANALYSIS AND OPTIMAL CONTROL STRATEGIES FOR STATE DEPENDENT QUEUEING SYSTEM

Ph. D. THESIS

by

SUDEEP SINGH SANGA

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE - 247 667 (INDIA)
JULY, 2019

PERFORMANCE ANALYSIS AND OPTIMAL CONTROL STRATEGIES FOR STATE DEPENDENT QUEUEING SYSTEM

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

MATHEMATICS

by

SUDEEP SINGH SANGA

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE - 247 667 (INDIA)
JULY, 2019

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**PERFORMANCE ANALYSIS AND OPTIMAL CONTROL STRATEGIES FOR STATE DEPENDENT QUEUEING SYSTEM**" in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Mathematics of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from December, 2014 to July, 2019 under the supervision of Dr. Madhu Jain, Associate Professor, Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institution.

(SUDEEP SINGH SANGA)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(Madhu Jain)
Supervisor

Date: **July , 2019**

I would like to dedicate this thesis to my loving parents.

Abstract

The performance modeling of Markovian and non-Markovian queueing models plays a vital role in design of real time queueing systems. The application of such models can be seen at many places including telecommunication system, computer networks, industrial and production system, etc. The state-dependent queueing models are of practical use and robust in depicting many real life congestion scenarios. These queues deal with the many realistic situations such as queues with discouragement, time sharing system, machine repair problems, etc. Optimal control of parameters of the queueing system is the key concern as far as the organizer as well customer's point of view. The arriving customers decide before joining the system whether to join or not to join the system based on their prior assessment of the queue length. So far as the controlling of the arrivals in the finite capacity system is concerned, admission control F-policy is quite useful to control the congestion of the customers/jobs and it can be helpful in reducing the lost customers/jobs in particular when the system capacity is full. The admission control F-policy mainly restricts the customers/jobs from an entry in the queueing system when the system capacity is exhausted and further admission of customers/jobs is allowed when enough customers/jobs are served so that the number of customers/jobs in the system drops to a threshold level $'F'$. The admission control F-policy can be employed to resolve the issue of controlling of the arrivals in the queueing system so as to avoid the loss of revenue and inconvenience to the customers.

In the present thesis, we investigate state dependent queueing models applicable to several queueing scenarios. The noble features of the investigation done are design of the control policies for some Markov and non-Markov queueing models by incorporating several features such as admission control F-policy, balking, reneging, feedback, unreliable server, retrial orbit, vacation, etc. In order to study the concerned queueing models, various system metrics such as number of customers in the system and in the queue, throughput, customer's loss, long run probabilities and reliability indices have been obtained using the relevant analytical/numerical techniques. The cost optimization and evaluation of optimal control parameters of the concerned queueing models, have been done using various methods such as quasi-Newton method, genetic algorithm, Harmony search algorithm, etc. The soft computing approaches namely fuzzy logic, neuro fuzzy technique, parametric non-linear programing are also employed for the prediction of performance indices of the concerned queueing models.

The main focus of our investigation in the present thesis is to the state dependent queueing models along with optimal control strategies. The work done on the optimal strategies and evaluation of performance indices of state dependent queueing models is organized into eight chapters. Some state-dependent queueing models have been developed by incorporating customer's joining strategies, admission control policy, double orbit, etc. The numerical results based on sensitivity analysis are also performed to validate the results derived for the concerned queueing model. The investigations done in the thesis are divided into eight chapters which are described as follows.

Chapter 1 is devoted for an overview and the motivation of the relevant topics alongwith preliminary concepts used for the concerned queueing models. The brief description of the methodologies used and the literature survey of relevant topics have been highlighted. Chapter 2 is concerned with an unreliable retrial queueing model under the admission control *F*-policy by incorporating the startup time and threshold policy. The adaptive neuro fuzzy inference system (ANFIS) technique is implemented to compare the numerical results obtained by Gauss-Seidel method. Chapter 3 presents the single server state-dependent model with general retrial attempts under admission control *F*-policy. The minimum cost of the system corresponding to optimal threshold parameter and optimal service rate is also determined using direct search method and quasi-Newton method. Chapter 4 deals with the multi-server finite queueing model with customer's balking behaviour. The concepts of admission control of customers based on *F*-policy and one additional server are incorporated to shorten the queue length formed by the customers in the rush hour. The system cost is minimized using direct search method and quasi-Newton method to obtain the admission control parameter.

Chapter 5 contains various results for the single server finite capacity queueing model with discouragement and general retrial times while the system operates under admission control *F*-policy. The soft computing based artificial neuro fuzzy inference system (ANFIS) method is applied to validate the results obtained by analytical method. Cost analysis is also done using genetic algorithm (GA) and quasi-Newton method by evaluating the optimal control parameters. Chapter 6 deals with the finite population models with general distributed retrial time under admission control *F*-policy. The fuzzy cost analysis for the finite population model is done by considering the cost elements as trapezoidal fuzzy numbers. Furthermore, the signed distance method is used to defuzzify the cost function. To determine the optimal control parameter and minimum cost of the system, genetic algorithm (GA) is also applied. Chapter 7 presents finite crisp and fuzzy population model for the multi-component machining system with general repair, standby support and server vacation. The cost analysis is done using Harmony search algorithm to determine optimal control parameters.

In Chapter 8, three infinite capacity double orbit retrial queueing models are studied. The first model is concerned with the customers' joining strategy in a double orbit retrial queueing system with balking. This model is transformed into fuzzy environment to study the fuzzified indices using the parametric nonlinear programing (P-NLP). The cost optimization is also done to determine optimal service rates using GA. The second model deals with double orbit feedback model. In the third double orbit model, the single server is taken as unreliable. In order to validate the feasibility of use neuro- fuzzy controller, ANFIS technique is also implemented. The cost function is framed and used to obtain the optimal service rates using quasi-Newton method.

At the end of the thesis, the concluding remarks and future scope have been outlined. The relevant references have been listed in the end of the thesis in alphabetical order.

Acknowledgements

I bow my head before ‘The God’, without whose blessings the present thesis would not have been possible in the present form and thank him for giving me patience and strength to overcome the difficulties, which crossed my way in the accomplishment of this endeavor and made me able to make my parents one dream come true. There have been many people with me throughout the whole journey for completing this thesis for doctorate degree. I would like to acknowledge all of them for their support, blessings and cooperation.

I would like to express my sincere thanks to my advisor and supervisor, Dr. Madhu Jain, Department of Mathematics, IIT Roorkee. I shall always remain indebted to her for appropriate comments and suggestions rendered by her at many critical junctures during the entire research work. Her astute guidance, constant advice and encouragement have been the impetus for me to learn a lot in every phase of my doctorate program. In-spite of the heavy pressure of work on her, she spared countless hours of her valuable time and evinced keen interest and personal care at every stage of my research. Further her bold initiative and uncompromising gesture made a highly challenging path of this research journey feasible.

I am thankful to Professor N. Sukavanam, Head, Department of Mathematics, IIT Roorkee and Professor V.K. Katiyar, Ex. Head, Department of Mathematics, IIT Roorkee, for their valuable direction, support and for all necessary facilities for carrying out my research work till its completion. I am also thankful to Professor Kusum Deep, DRC Chairperson, Professor S.P. Yadav, SRC Chairman, Professor R. Balasubramanian, External Expert, Professor S.K. Gupta, Internal Expert, all faculty members and staff members of the department for their valuable support and encouragement at all levels. I am highly indebted to Professor G.C. Sharma. Ex. Pro Vice Chancellor, Dr. B.R. Ambedkar University, Agra who has always encouraged me to do my research work as good as possible. Moreover, his editing suggestions and precise sense of language contributed significantly upto the end of my research work.

I owe my heartfelt thanks to Minupriya, who has always been a constant source of inspiration and helped me. My sincere thanks to Birendra Singh for his help, valuable suggestions and discussions during my research work. I am fortunate to acknowledge the support of some individuals who were always there beside me during happy and hard moments to push me and motivate me; namely, Dr Rakesh Kumar Meena, Pankaj Kumar, Shobha Rani, Mayank, Anshul

Kumar, Sibasish Djhibar, Sandeep Kaur for their cooperation and support for making my stay at IIT Roorkee a memorable one.

I wish to put on records that whatever I have achieved in my life is due to the blessings of my mother Smt. Uma Sanga and father Shri M.S. Sanga. I cannot restrain myself in offering my heartfelt thanks to my elder sisters Nishtha and Shreshtha and my brother-in-laws Amit and Gagan for their cordial support and blessings.

I also wish to express my sincere thanks to Ministry of Human Resource Development, Govt. of India for its financial support in the form of junior research fellow (JRF) and senior research fellow (SRF) via Grant No. MHR-02-23-200-429. I gratefully acknowledge the financial support received from the Dean of Resources and Alumni Affairs, IIT Roorkee to participate in International Conference in Bangkok, Thailand.

Last but not the least, I would like to thank everyone who has directly or indirectly helped me in the successful completion of the thesis.

Sudeep Singh Sanga

List of Publications

Research papers accepted for publication

[1] Madhu Jain & **Sudeep Singh Sanga**, Fuzzy cost optimization and admission control for machine interference problem with general retrial, Journal of Testing and Evaluation, American Society for Testing and Materials (ASTM) International, 48(6), 2020. doi:10.1520/JTE20180882.

[2] Rakesh Kumar Meena, Madhu Jain, **Sudeep Singh Sanga** & Assif Assad, Fuzzy modeling and Harmony search optimization for machining system with general repair, standby support and vacation, Applied Mathematics and Computation, Elsevier, 361, 858-873, 2019.

[3] **Sudeep Singh Sanga** & Madhu Jain, FM/FM/1 double orbit retrial queue with customers' joining strategy: A parametric nonlinear programing approach, Applied Mathematics and Computation, Elsevier, 2019. doi: 10.1016/j.amc.2019.06.056 (*in press*).

[4] Madhu Jain & **Sudeep Singh Sanga**, Admission control for finite capacity queueing model with general retrial times and state dependent rates, Journal of Industrial and Management Optimization, American Institute of Mathematical Sciences, 2019. (*Accepted for publication*).

[5] Madhu Jain & **Sudeep Singh Sanga**, Control F-policy for fault tolerance machining system with general retrial attempts, National Academy Science Letter, Springer, 40(5), 359–364, 2017.

[6] Madhu Jain & **Sudeep Singh Sanga**, F-policy for M/M/1/K retrial queueing model with state-dependent rates, In: Performance Prediction and Analytics of Fuzzy, Reliability and Queuing Models, 1st ed., Springer Singapore, 127–138, 2019.

[7] Madhu Jain & **Sudeep Singh Sanga**, Performance modeling and ANFIS computing for finite buffer retrial queue under F-policy, In: Proceedings of Sixth International Conference on Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing, Springer Singapore, 547, 248–258, 2017.

[8] Madhu Jain & **Sudeep Singh Sanga**, Rakesh Kumar Meena, Control *F*-policy for Markovian retrial queue with server breakdowns, In: 1st International Conference on Power Electronics, Intelligent Control and Energy Systems, IEEE, 1–5, 2016

Research papers communicated for publication

[1] **Sudeep Singh Sanga** & Madhu Jain, Cost optimization and ANFIS computing for admission control of M/M/1/K queue with general retrial times and discouragement, Applied Mathematics and Computation, Elsevier. (Revised and Under Review).

[2] Madhu Jain & **Sudeep Singh Sanga**, Unreliable single server double orbit retrial queue with balking, Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, Springer. (Under Review).

[3] Madhu Jain & **Sudeep Singh Sanga**, Optimal control F-policy for M/M/R/K queue with an additional server and balking, International Journal of Applied and Computational Mathematics, Springer. (Under Review).

[4] Madhu Jain & **Sudeep Singh Sanga**, State dependent queueing models under admission control *F*-policy: An overview, Operational Research, Springer. (Under Review).

[5] Madhu Jain & **Sudeep Singh Sanga**, Fuzzy model of double orbit retrial queue with balking and feedback, National Academy Science Letter, Springer. (Under Review)

Participation in the Conferences

[1] Presented a paper entitled "Cost Optimization for Single Server Double Orbit Retrial Queueing Model with Server Breakdown" in "Annual Conference of Vijñāna Parishad of India on Modeling, Optimization and Computing for Technological and Sustainable Development", SRM Institute of Science and Technology, Delhi-NCR Campus, 26-28 April, 2019.

[2] Presented a paper entitled "Optimal Management for M/M/1 Double Orbit Retrial Queueing Model with Unreliable Server: Quasi-Newton Method" in "International Conference on Emerging Issues in Business, Technology and Applied Sciences", Bangkok, Thailand, 22-23 December, 2018.

[3] Participated in National Conference on "Modeling, Optimization and Computing for Engineering Problems: Use of Technical Hindi Terminology" sponsored by CSTT, MHRD, Government of India, Indian Institute of Technology Roorkee, Roorkee, India, 12-14 October, 2018.

[4] Presented a paper entitled "F-policy for M/M/1/K Retrial Queueing Model with State Dependent Rates" in "International Conference on Recent Trends in Operations Research and Statistics", Indian Institute of Technology Roorkee, Roorkee, India, 28-30 December, 2017.

[5] Presented a paper entitled "Control F-policy for Time Sharing Queueing Model with General Retrial Times" in "National Conference on Mathematical Sciences and Scientific Computing for Industrial Development & International Symposium on Probabilistic Model and Applications of Special Functions" Manipal University, Jaipur, India, 24-26 November, 2017.

[6] Presented a paper entitled "Performance Modeling and ANFIS Computing for Finite Buffer Retrial Queue Under F-policy" in "Sixth International Conference on Soft Computing for Problem Solving" Thapar University, Patiala, India, 23-24 December, 2016.

[7] Presented a research paper entitled "Control F-policy for Markovian Retrial Queue with Server Breakdowns" in "First IEEE International Conference on Power Electronics, Intelligent Control, and Energy Systems", Delhi Technological University, New Delhi, India, 4-6 July, 2016.

Participation in the Workshops

[1] "Applied Stochastic Models and Optimization", Indian Institute of Technology Roorkee, 26-27 May, 2017.

[2] "Probability Meeting", Indian Statistical Institute, Bangalore, 12-14 May, 2017.

[3] "Modeling, Optimization and Simulation of Stochastic Systems", Indian Institute of Technology Roorkee, Roorkee, 26 November, 2016.

[4] "Optimization Techniques for Solving Industrial Problems", Indian Institute of Technology Roorkee, Roorkee, 15 October, 2016.

[5] "Advanced Instructional School on Optimization", Indian Institute of Technology Bombay, 9-21 May, 2016.

[6] Modeling Week and Study Group Meeting on "Industrial Problem", ITM Vadodara, 10-14 March, 2016.

[7] "Static and Dynamic Mechanism Design", Indian Statistical Institute, Delhi, 1-4 August, 2015.

[8] "Computational Techniques for Differential Equations with MATLAB", Indian Institute of Technology Roorkee, Roorkee, 2-6 July, 2015.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. MOTIVATION

The optimal control policy can be implemented to facilitate the quality of service to the customers and to enhance the profit of the service systems operating in many real world queueing scenarios. To encourage the jobs/customers for getting served by the server, the system organizer should focus on the optimal control parameters such as admission/service control threshold parameter, joining threshold probability, service/repair rate, startup rate, vacation rate, etc. Optimal control parameters in queueing systems are important factors for the customers to decide whether to queue up or not for desired service, based on their assessment of queue length or total waiting time. These parameters play key role for the design and management of queueing service systems so that the trade-off between system cost and waiting time of jobs/customers may be done. Optimal control strategies provide valuable insights to the system designers and decision makers to reduce the system cost and delay in congestion problems and can be used for the improvement of the concerned system. Optimal control for the queueing system may be helpful to check the discouragement behavior of the customers in several queue length dependent scenarios including the computer and communication networks, manufacturing and production units, service and distribution systems, and many more systems.

In several queueing systems, the rates may be state-dependent, i.e. the arrival and service may be dependent on the number of customers present in the system. It is also seen that the server may render service with faster rate as soon as the queue size increases. On the contrary, sometimes it may happen that the server becomes slow due to stress. The importance of the state dependent queue may be realized in many day-to-day queueing scenarios including queues with customer’s discouragement behavior, machine repair systems, time-sharing systems, etc. The state-dependent queueing model may also be applicable to the queueing system in which the decision-maker can facilitate the additional servers on observing the long queue based on threshold policy.

The customers' discouragement behavior in the queueing situations, is a common issue and may be noticed at many places in our routine lives including at railways reservation counters, banks, restaurants, post office, VISA service centers, etc. In queueing scenarios, sometimes, it is seen that the arriving customer may decide not to join the queue for receiving the service due to some reason. In such a case, the discouraged customer may depart from the system forever without joining the system; and this phenomenon is known as balking. On the other hand, it can be seen that the arriving customer joins the queue for receiving the service and due to impatience after waiting for some time, he may leave the system without being served; this kind of discouragement behavior is known as reneging. The discouragement behavior of the customers in the queueing system affects the profit as well as the goodwill of the system organizers. Most often, it is experienced that after getting served, the customers do not feel satisfied with the service. Once the customer is served by the server and if he is not satisfied with the service then he may re-join the queue for the service; this phenomenon is termed as *customers' feedback*.

In many organizations/industries, it is noticed that the machines used for rendering the service to jobs/units are prone to failures. The system managers constantly try to facilitate the service to their customers at a fast pace, however, the system may stop functioning due to unpredictable failure of the machines. In order to reduce the congestion of failed machines in the system and to maintain the smooth functioning of the machining system, the managers of the concerned system provide the immediate repair to the failed machines at optimal cost.

The formation of queues and consequently discouragement behavior of the customers as well as delay in the service are the major problems for both customers as well as the system organizers. In order to maintain the smooth functioning of the service systems where long queues are built up, the arrivals should be controlled and this can be done by implementing the optimal admission control F-policy. The intention of the admission control F-policy is to control the admission of joining customers in case when the waiting space is full. In admission control F-policy, the jobs/customers are completely restricted to join the queue when the system waiting space becomes exhausted and during this period only jobs present in the system are served by the server to reduce the waiting line to a certain extent. The further admission of customers can be allowed when 'F' jobs remain after departure of the served customers in the system. The queueing models investigated under admission control F-policy have several applications at various places such as day-to-day service systems, hospitals, call centers, assembly lines, etc.

To illustrate the applicability of optimal *F*-policy, we cite an example of transport service system where loaded trucks arrive at a warehouse for unloading. Due to the limited space available for trucks at parking area, the admission control under *F*-policy for loaded trucks can be employed for the smooth functioning of the unloading service. When the capacity of the parking area becomes full, the newly arrived loaded trucks can be stopped for the time being to enter in the parking area and as soon as the number spaces for the trucks in the parking area ceases to a pre-defined level '*F*'. The loaded trucks are further allowed to park in the available space for unloading.

In queueing situations, sometimes it is experienced that the arriving customer may not like to wait in the queue on finding the busy server and prefers to do some other work in the virtual place called retrial orbit. After a random period of time, the customers waiting in the orbit, may retry for getting the service with the expectation that the service facility is likely to be free; such a special type of queue is termed as retrial queue. The formation of retrial queues may be seen at many places including banks, ATMs, computer communication networks, business, industries, etc. To examine the practical applicability of retrial queues, consider an example of call center wherein the caller may try for a call to the center and if the dialed number is busy, the caller may get a message of busy line. The caller may disconnect the call and would like to remain in the retrial orbit and may try after some time with the hope that the line becomes free to connect the call.

It is often realized that some of the arriving customers may not be willing to join the queue if the facility available in the waiting zone is not up to their expectation. Some customers do not bother about paying more money for getting better comfort during waiting period. To deal with such type of situations, the provision of double orbits in retrial queue namely, ordinary orbit and premium orbit can be provided. If the arriving customers find the server busy, the system organizer directs them to shift to ordinary orbit or premium orbit as per their demand and paying capacity and after a period of time customers retry for the service again from their respective orbits. To understand the single server double orbit retrial queue in a practical way, we cite an example of e-commerce website which allows customers to shop directly from the manufacturing unit through the internet. The e-commerce website may deal with two types of customers namely, ordinary customers and premium customers. During the festive season, the e-commerce website seems too busy due to high access by a large number of customers at the same time. It is experienced that some of arriving customers may not wish to continue shopping due to the heavily loaded website and become discouraged and leave the website forever. On the other hand, many arriving ordinary customers (premium customers) join the queue on finding the website busy. However, the ordinary (premium) customers move to the ordinary orbit (premium orbit) and later re-attempt for the shopping.

The server breakdown is a common phenomenon in queueing systems. Due to over congestion of the customers in the system, sometimes server may not bear the load of the customers and this results in server break down. Such example can be seen at railway reservation counter where operator provides the tickets via computer to the customers. One can experience that the computer may stop working due to several reasons such as software failure, network problem, system overload, hardware issues, power backup, etc. Similar situation can be seen at many places where server may breakdown such as telecommunication system, data center, etc. In queueing problems, server failure is key issue so that the system organizers focus on it when dealing with the queueing models.

In queueing literature, a very few researchers have studied the admission control related issues for queueing situations by developing generic state dependent finite model. There is scarcity of works towards queueing models having double retrial orbit. The joining strategies and optimal control of system parameters have been found interesting and require more attention by noticing that there is significant gap between theoretical and application oriented research works. Our study on admission control joining strategies of customers is motivated by its ample applications in failure prone machining environment and day-to-day congestion scenarios. The main objective of the present study is the performance modeling and design of the control policies for some Markov and non-Markov queueing models for realistic queueing situations. Markovian as well as non-Markovian models under steady state have been analyzed for both crisp and fuzzy descriptors. By developing state dependent queueing models, we have studied the admission control strategies by incorporating many realistic features including fuzzy parameters, retrial attempts, double orbits, discouragement, feedback, etc. The rest of the introductory chapter is organized in the following manner. Section 1.2 is devoted to introductory aspects of queue with optimal control strategies and the admission control *F*-policy. Section 1.3 describes the mathematical formulation of state-dependent queue operating under admission control *F*-policy. The methodologies and techniques used for the state-dependent queueing models have been discussed in Section 1.4. In Section 1.5, some special queueing models under admission control *F*-policy are presented. Literature review of the work done in the thesis is given in Section 1.6. Section 1.7 highlights the overview and outlines of the thesis. Finally, the concluding remarks are given in Section 1.8.

1.2. QUEUE WITH OPTIMAL CONTROL STRATEGIES

In many congestion situations, it is experienced that the arriving customers observe the queue size and decide whether to join the queue for the service or not. However, when the system seems to be congested, it results in the loss of revenue in case arriving customers balk or waiting customers renege. It is most often inconvenient to the customers to wait in the queue. In order to encourage the customers for joining the system for the service, the system organizer may offer a reward to the customer for being served. Also, a waiting cost is imposed when the customers remain in the system. The arriving customers can make such a strategy to join the system so that the reward offered by the system organizer is greater than the waiting cost incurred on them. If the customers observe that the reward is greater (less) than the waiting cost then the customers always prefer to join (balk) the system for the service (without being served) whereas the customers can be indifferent between joining the system and balking when the reward and the cost are equal. In order to make optimal strategies so that the net profit would be maximized, the system organizer can establish a reward-cost structure as follows. Consider a single server queue in which the customers join the system as per Poisson distribution with rate λ and service is provided by the single server as per exponential distribution with rate μ. It is assumed that the arriving customer enters into the system with probability q and balks with probability $(1-q)$. After being served, each customer receives a reward R per unit time and spends waiting cost C per unit time. Based on reward-cost structure, the profit function $f(q)$ can be constructed as

$$f(q) = \lambda(q)\mathrm{R} - \mathrm{C} \times \text{Waiting time}.$$

For more details of optimal control strategies based on cost-reward structure, we refer the book by Hassin and Haviv (2003) and papers by Nobel and Tijms (1999, 2000).

Gupta (1995) introduced the concept of admission control *F*-policy for finite capacity single server Markov queueing system. According to Gupta (1995), *F*-policy problem addresses the issue of controlling arrivals in a finite capacity (K) queueing system. If the system reaches to its capacity K (i.e. the system becomes full), no further customers are allowed to enter the system

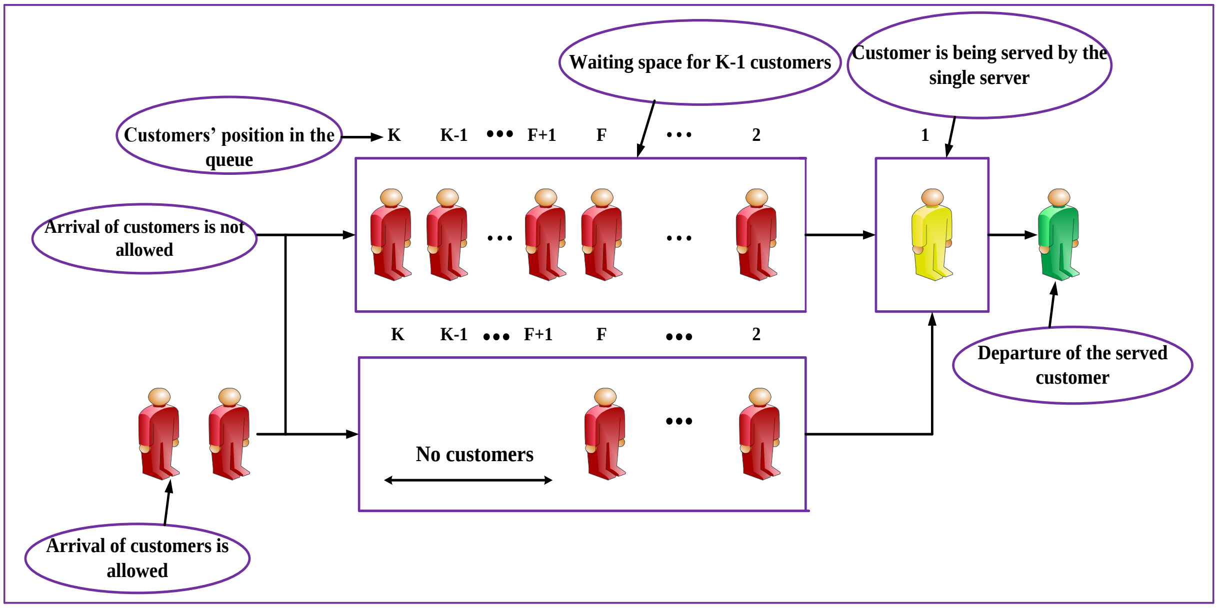


Figure 1.1: Pictorial view of admission control *F*-policy

until enough customers who are already in the system have been served so that the number of customers in the system drops down to a threshold value $F\,(0 \le F \le K-1)$. At that point, a setup time is required to start allowing the customers in the system. Then after the system behaves normally until it reaches its capacity and above process is repeated all over again. To understand the concept of *F*-policy, a pictorial view of admission control *F*-policy is shown in Figure 1.1.

1.3. STATE DEPENDENT QUEUE UNDER ADMISSION CONTROL *F*-POLICY

To construct the mathematical model for the admission control *F*-policy for state dependent single server finite capacity (*K*) Markov model, certain assumptions made are as follows:

- The arriving customers/jobs join the system in Poisson pattern with rate λ.
- The customers/jobs are served by the single server following the exponential distribution with mean $1/\mu$.
- As *K* customers/jobs accumulated in the system, the customers/jobs are restricted till the number of customers/jobs in the system further reduces to '*F*' $(0 \le F \le K-1)$. To start admission of customers/jobs in the system, a startup job is required as per exponential distribution with mean $1/\theta$.

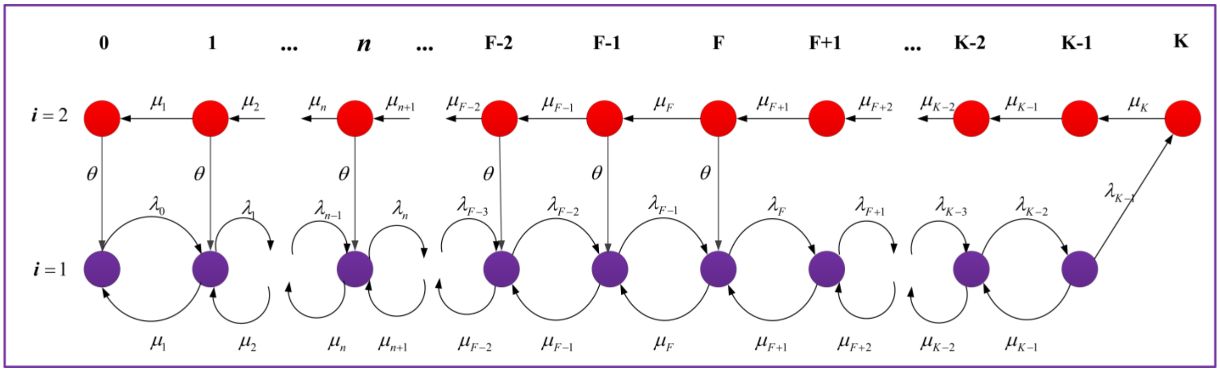


Figure 1.2: Transition state diagram for state dependent model under F-policy

For F-policy model, we denote the status of the server at time τ as follows

$S(\tau)=i=1(2)$, when the server is busy in rendring the service and admission of customers is allowed (not allowed).

When the server is in state i, denoting the probability of $n\,(n=0,1,2,...,K)$ customers at time τ by $P_{i,n}(\tau)$, the transient birth-death equations can be framed as follows (see Figure 1.2):

For $i=1$, $0\le n\le K-1$, *when the admission of customers is allowed.*

$$\frac{d}{d\tau}P_{1,0}(\tau)=-\lambda_0P_{1,0}(\tau)+\mu_1P_{1,1}(\tau)+\theta P_{2,0}(\tau) \tag{1.1}$$

$$\frac{d}{d\tau}P_{1,n}(\tau)=-(\lambda_n+\mu_n)P_{1,n}(\tau)+\lambda_{n-1}P_{1,n-1}(\tau)+\mu_{n+1}P_{1,n+1}(\tau)+\delta_{n-F}\theta P_{2,n}(\tau),\ n=1,2,...,K-2 \tag{1.2}$$

$$\frac{d}{d\tau}P_{1,K-1}(\tau)=-(\lambda_{K-1}+\mu_{K-1})P_{1,K-1}(\tau)+\lambda_{K-2}P_{1,K-2}(\tau) \tag{1.3}$$

For $i=2$, $0\le n\le K$, *when the admission of customers is restricted.*

$$\frac{d}{d\tau}P_{2,n}(\tau)=-(\theta+\delta_{n-F}\mu_n)P_{2,n}(\tau)+\mu_{n+1}P_{2,n+1}(\tau),\ n=0,2,...,F \tag{1.4}$$

$$\frac{d}{d\tau}P_{2,n}(\tau)=-\mu_nP_{2,n}(\tau)+\mu_{n+1}P_{2,n+1}(\tau), n=F+1,...,K-1 \tag{1.5}$$

$$\frac{d}{d\tau}P_{2,K}(\tau)=-\mu_KP_{2,K}(\tau)+\lambda_{K-1}P_{1,K-1}(\tau) \tag{1.6}$$

For the steady state, when $\tau\to\infty$, we denote probabilities by $P_{i,n}=\lim_{\tau\to\infty}P_{i,n}(\tau)$. For the steady state, Equations (1.1)-(1.6) alongwith normalizing condition can be solved using various methods available in queueing literature such as recursive method, matrix analytical method (Wang and Yang, 2009), maximum entropy method (Jain and Bhagat, 2015a), successive over relaxation (SOR) method (Jain and Meena, 2017), etc.

Gupta (1995) developed M/M/1/K model under *F*-policy by denoting the steady state probability of n customers by $p_{i,n}$, where $i = 0(1)$ are used to represent the state when the customers are not allowed (allowed) to enter in the system, respectively. The birth-death equations of Gupta's model can be easily framed by using the appropriate transitions rate as shown in Figure 1.3. This is special case of state dependent model when $\lambda_n = \lambda$ and $\mu_n = \mu$.

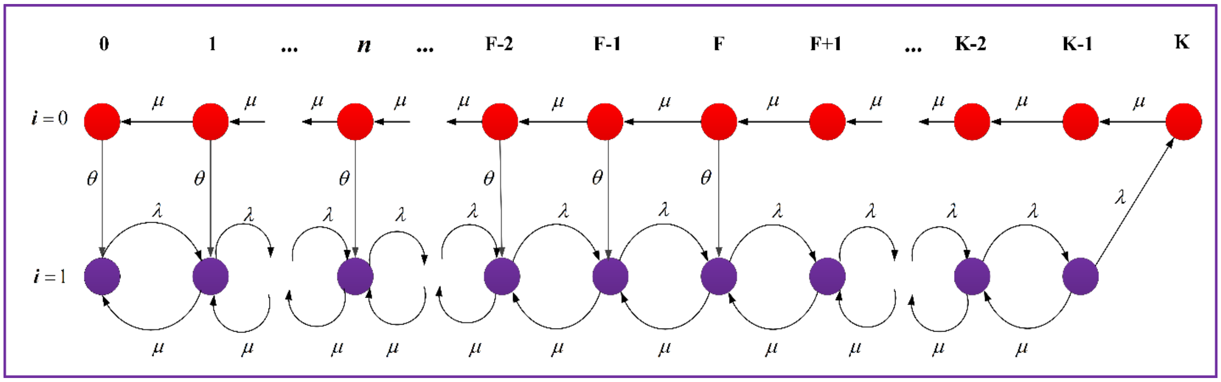


Figure 1.3: Transition state diagram for M/M/1/K model under *F*-policy

1.4. METHODOLOGICAL ASPECTS

In order to analyze the state dependent queueing models, the worth-noting analytical techniques viz. birth death process, Markov process, supplementary variable technique, recursive method, probability generating method, etc. are used. The parametric non-linear programing approach and Gauss-Seidel method have also been employed for the development of model and mathematical analysis of the concerned queueing models. Numerical optimization methods such as direct search method, quasi-Newton method have been implemented to determine the optimal control parameters. The soft computing approaches namely fuzzy logic, neuro fuzzy technique, genetic algorithm (GA) and harmony search method have been used for the prediction of concerned queueing models. To explore the performance of the queueing systems, several system metrics namely, expected system size, expected number of customers in the queue and in retrial orbit, expected waiting time in the system and in the queue, throughput, expected delay time, etc. have been derived. To analyze the status of the server, long run probabilities have also been established. The reliability indices such as system availability, operative efficiency, failure frequency have been derived for the queueing system having failure prone units/servers. In order to increase the revenue to the system designers and to facilitate the customers with better service, the optimal control parameters such as admission threshold parameter, joining probability, service rate, vacation rate, number of servers, etc. have been determined by minimizing the cost function. In order to validate the investigation done for the concerned queueing models, the effects of the system parameters on different performance measures have been explored by taking suitable numerical examples. Now, we discuss some methodologies/techniques which we have used in our study to establish performance metrics.

1.4.1. Stochastic and Markov Process

A *stochastic process* is a family of random variables, $\{X(\tau):\tau \in T\}$, where T denotes a parameter space and at time τ, $X(\tau)$ is called a state of the process. Stochastic process is said to be discrete or continuous depending on whether T is discrete parameter space or continuous parameter space.

Markov processes are memory-less in nature. *Markov process* is defined as "given that the present condition of the process, its future evolution is independent of the past". Mathematically, Markov process is a discrete or continuous parameter stochastic process if, for any set of n point $\tau_1 < \tau_2 < ... < \tau_n$ in parameter space, the conditional distribution of $X(\tau_n)$, given the value of $X(\tau_1), X(\tau_2),..., X(\tau_{i-1})$ depends only on $X(\tau_{n-1})$. For any real numbers $x_1, x_2,..., x_n$, Markov process is defined as

$$\text{Prob}\{X(\tau_n) \le x_n | X(\tau_1)=x_1,...,X(\tau_{n-1})=x_{n-1}\}=\text{Prob}\{X(\tau_n) \le x_n \mid X(\tau_{n-1})=x_{n-1}\}.$$

For the detailed description of Markov process, we refer the book by Medhi (2003).

1.4.2. Birth-Death Process and Markov Queueing Model

A birth-death process is a continuous time or discrete-time Markov chain with the restriction that at each step, the state transitions, if any, can occur only between neighboring states. Here, birth means arrival of a customer in the system whereas death represents the departure of a customer from the system. It is worthwhile to construct the basic mathematical model for the finite capacity single server Markovian model with state dependent rates using birth-death process. The customers arrive in the finite capacity (K) system as per Poisson process with rate λ and served according to an exponential distribution with rare μ.

Consider continuous time Markov chain (CTMC) $\{N(\tau), \tau > 0\}$ and $N(0) = 0$, where $N(\tau)$ represents the number of customers upto time τ. Let $S(\tau) = n$ denotes the state of the system at time τ. For state dependent model, consider the transition between τ and $\tau + \Delta\tau$ as follows:

Prob{state n to state $n+1$ in $\Delta\tau$ time} $= \lambda_n \Delta\tau$,

Prob{state $n+1$ to state n in $\Delta\tau$ time} $= \mu_n \Delta\tau$,

Prob{no change in state n in $\Delta\tau$ time} $= 1 - (\mu_n \Delta\tau + \lambda_n \Delta\tau)$.

$$P_0(\tau + \Delta\tau) = (1 - \lambda_0 \Delta\tau) P_0(\tau) + \mu_1 \Delta\tau P_1 \tag{1.7}$$

Equation (1.7) can be re-written as

$$\left[\frac{P_0(\tau + \Delta\tau) - P_0(\tau)}{\Delta\tau}\right] = -\lambda_0 P_0(\tau) + \mu_1 P_1 \tag{1.8}$$

Taking $\Delta\tau \to 0$, Equation (1.8), yields

$$\frac{d}{d\tau} P_0(\tau) = -\lambda_0 P_0(\tau) + \mu_1 P_1(\tau) \tag{1.9}$$

Similarly, the classical birth death equations for finite state space $1 \le n \le K$ can be formulated as

$$\frac{d}{d\tau} P_n(\tau) = -(\lambda_n + \mu_n) P_n(\tau) + \lambda_{n-1} P_{n-1}(\tau) + \mu_{n+1} P_{n+1}(\tau), 1 \le n \le K-1 \tag{1.10}$$

$$\frac{d}{d\tau} P_K(\tau) = -\mu_K P_K(\tau) + \lambda_{K-1} P_{K-1}(\tau). \tag{1.11}$$

For more details of birth-death process, we refer the book by Cox and Miller (2017).

Let the state dependent arrival and service rates in Markov queueing model be denoted by λ_n and μ_n, respectively. By choosing appropriate values of λ_n and μ_n, the state dependent model can be transformed to finite population model, balking model, reneging model, time sharing model, as follows:

(i) Single server finite population model: $\lambda_n = (K-n)\lambda$ and $\mu_n = \mu.$

(ii) Balking model: $\lambda_n = \begin{cases} q\lambda, & \text{for constant balking,} \\ \lambda e^{-nq}, & \text{for exponential balking.} \end{cases}$

(iii) Single server reneging model: $\mu_n = \mu + r(n)$, where $r(n)$ is a reneging function assumed exponential with $r(0) = r(1) = 0$.

(iv) Time sharing model: $\lambda_n = \dfrac{\lambda}{n+1}$ and $\mu_n = \dfrac{\mu}{n}$.

1.4.3. Non-Markov Queueing Model

A classical queueing model is termed as non-Markovian queueing model if either of inter-arrival times or service times is not exponentially distributed. In order to explain the non-Markov queueing model, it is worthwhile to describe the M/G/1 queueing model. We assume that the customers arrive according to Poisson fashion with rate λ. The service times are assumed to be independent identical distributed (i.i.d.) random variables with density function $\mu(x)\,(x \geq 0)$ with mean $1/\mu$.

At the time τ, let $N(\tau)$ and $U(\tau)$ denote the number of customers in the system and remaining service time, respectively. The system state probability at time epoch τ is defined as follows:

$P_0(\tau) = \text{Prob}\{N(\tau) = 0\}$,

$P_n(\tau) = \text{Prob}\{N(\tau) = n, x < U(\tau) \leq x + dx\}, x \geq 0, n \geq 1.$

Also, $P_n(\tau) = \text{Prob}\{N(\tau) = n\} = \int_0^\infty P_n(u,\tau)dx, n \geq 1.$

At steady state, i.e., when $\tau \to \infty$, we define:

$P_n = \lim_{\tau\to\infty} P_n(\tau), n \geq 0.$

Now by using the probability arguments, Chapman-Kolmogorov equations are formulated as follows:

$$-\lambda P_0 + P_1(0) = 0 \tag{1.12}$$

$$\frac{d}{dx}P_n(x) = \lambda P_n(x) - \lambda P_{n-1}(x) - P_{n+1}(0)\mu(x), n \geq 1. \tag{1.13}$$

Equations (1.12)-(1.13) can be solved using Laplace transformation and then recursive method. The startup, retrial time, vacation time, etc. may also be governed by non-Markovian process.

1.4.4. Recursive Method

This method is commonly used as a tool for solving Chapman-Kolmogorov difference equations. In this method, the governing difference equations can be solved to obtain a sequence of state probabilities $P_1, P_2, ..., P_n$ each in terms of P_0. The normalizing condition to $\sum_{n=0}^{\infty} P_n = 1$ can be used to find P_0 and further all the state probabilities of the queueing system.

1.4.5. Probability Generating Function

This method can be used to find the steady state probabilities, say $P_1, P_2, ..., P_n$, of the concerned queueing model. This method involves with power series expansion of a function. Let us consider a function $\Pi(z) = \sum_{n=0}^{\infty} P_n z^n$ which has a power series expansion

i.e.
$$\Pi(z) = \sum_{n=0}^{\infty} P_n z^n = P_0 + P_1 z + P_2 z^2 + P_3 z^3 + ... \tag{1.14}$$

The power series given in (1.14) is called generating function if the power series converges for some range of z. Probability generating function is appropriate method to solve Chapman-Kolmogorov difference equations related to queueing system. The steady state probabilities can be determined by equating the coefficients of z^n in the expansion of $\Pi(z)$ and normalizing condition $\Pi(1) = 1$ (cf. Gross et al., 2008).

1.4.6. Supplementary Variable Technique

The supplementary variable technique provides an approach to analyze the non-Markovian queueing models such as M/G/1, GI/M/1, M/M/1 with general retrial time, etc. In the queueing literature, there are two types of supplementary variable techniques viz. elapsed time (cf. Cox, 1955) and remaining time (cf. Henderson, 1972). In order to make the system Markovian, the supplementary variable can be added to stochastic process. Now, we describe the use of the supplementary variable as remaining service (retrial) time to non-Markovian queueing models which was first introduced by Henderson (1972). In order to explain the supplementary variable technique, it is worthwhile to describe the M/M/1/K queueing model with general retrial attempts. We consider the customers arrive according to Poisson fashion with rate λ and served according to exponential distribution with rate μ whereas the retrial times are independent identical distributed random variables with density function $\gamma(x)(x \geq 0)$ with mean $1/\gamma$.

At the time τ, let $N(\tau)$, $S(\tau)$ and $U(\tau)$ denote the number of customers in the system, status of the server and remaining service time, respectively. $S(\tau)$ is defined as

$$S(\tau)=\begin{cases}0, & \text{the customers are compelled to join the retrial pool on finiding the server being busy,}\\ 1, & \text{the server is occupied and the jobs are permitted to enter in the system.}\end{cases}$$

The system state probabilities at time epoch τ are defined as follows:

$$P_{0,0}(\tau) = \text{Prob}\{S(\tau)=0, N(\tau)=0\},$$

$$P_{0,n}(\tau) = \text{Prob}\{S(\tau)=0, N(\tau)=n, x < U(\tau) \leq x+dx\}, x \geq 0, n \geq 1,$$

$$P_{1,n}(\tau) = \Pr\{S(\tau)=1, N(\tau)=n\},\ 0 \leq n \leq K-1$$

Also, $P_{0,n}(\tau) = \text{Prob}\{N(\tau)=n\} = \int_0^\infty P_{0,n}(x,\tau)dx, n \geq 1.$

Now by using the probability arguments, Chapman-Kolmogorov equations are formulated for the time τ and $\tau + \Delta\tau$ as follows:

$$P_{0,n}(x-\Delta\tau, \tau+\Delta\tau) = (1-\lambda_n\Delta\tau)P_{0,n}(x,\tau) + \mu\Delta\tau P_{1,n}(x,\tau), 1 \leq n \leq K-1 \tag{1.15}$$

Equation (1.15) yields the following partial differential equation when $\Delta\tau \to 0$

$$\left(\frac{\partial}{\partial\tau} - \frac{\partial}{\partial x}\right)P_{0,n} = -\lambda_n P_{0,n}(x,\tau) + \mu P_{1,n}(x,\tau) 1 \leq n \leq K-1 \tag{1.16}$$

Similarly, by following same arguments, we have

$$\frac{d}{d\tau}P_{0,0}(\tau) = -\lambda_0 P_{0,0}(\tau) + \mu P_{1,0}(\tau) \tag{1.17}$$

$$\frac{d}{d\tau}P_{1,0}(\tau) = -(\lambda_1+\mu)P_{1,0}(\tau) + \lambda_0 P_{0,0}(\tau) + P_{0,1}(0,\tau) \tag{1.18}$$

$$\frac{d}{d\tau}P_{1,n}(\tau) = -(\lambda_{n+1}+\mu)P_{1,n}(\tau) + \lambda_n P_{0,n}(\tau) + \lambda_n P_{1,n-1}(\tau) + P_{0,n+1}(0,\tau), 1 \leq n \leq K-1. \tag{1.19}$$

At steady state, i.e., when $\tau \to \infty$, we define:

$$P_{i,n} = \lim_{\tau \to \infty} P_{i,n}(\tau), 0 \le n \le K-1,\ i = 0,1.$$

At steady state, Equations (1.16)-(1.19) take the form

$$-\frac{d}{dx} P_{0,n} = -\lambda_n P_{0,n}(x) + \mu P_{1,n}(x) 1 \le n \le K-1 \tag{1.20}$$

$$-\lambda_0 P_{0,0} + \mu P_{1,0} = 0 \tag{1.21}$$

$$-(\lambda_1 + \mu) P_{1,0} + \lambda_0 P_{0,0} + P_{0,1}(0) = 0 \tag{1.22}$$

$$-(\lambda_{n+1} + \mu) P_{1,n} + \lambda_n P_{0,n} + \lambda_n P_{1,n-1} + P_{0,n+1}(0) = 0, 1 \le n \le K-1. \tag{1.23}$$

The above set of Equations (1.20)-(1.23) can be solved using recursive method as discussed in Section 1.4.4.

1.4.7. Soft Computing Techniques

(i) Fuzzy Model

In order to incorporate fuzzy descriptors for the concerned queueing model, it is worthwhile to give a brief account of basic fuzzy concepts used to formulate a fuzzy model. The fuzzy model for machining system with general repair can be symbolically represented as FM/FG/1/K/K machining system. If we merge the fuzzy descriptors with Kendall's notations, FM and FG denote the fuzzified Markov input and fuzzy general repair time, respectively. For FM/FG/1/K/K machining system, the system parameters $(\tilde{\Lambda})$ such as failure rate, repair rate, and vacation rate are used as linguistic quantifiers.

Let $\tilde{\Lambda} = [\delta_1, \delta_2, \delta_3]$ be a triangular fuzzy number corresponding to the parameter $\tilde{\Lambda}$ with membership function $\eta_{\tilde{\Lambda}}(y)$. Then

$$\eta_{\tilde{\Lambda}}(y) = \begin{cases} \dfrac{y-\delta_1}{\delta_2-\delta_1}, & \delta_1 \le y < \delta_2, \\ 1, & y = \delta_2, \\ \dfrac{\delta_3 - y}{\delta_3-\delta_2}, & \delta_2 < y \le \delta_3. \end{cases} \tag{1.24}$$

We assume that $g(y)$ is the system characteristics of interest. Here $g(\tilde{\Lambda})$ is also a fuzzy number and is defined by

$$g(\tilde{\Lambda}) = \{(z, \eta_{g(\tilde{\Lambda})}(z)) : z \in Z\} \quad . \tag{1.25}$$

where Z is the universal crisp set of $g(\tilde{\Lambda})$. Now, we formulate the membership function of $g(\tilde{\Lambda})$, following Zadeh's extension principle (cf. Zadeh, 1978) as

$$\eta_{g(\tilde{\Lambda})}(z) = \sup_{y \in Y} \min\{\eta_{\tilde{\Lambda}}(y) \mid z = g(y)\} \tag{1.26}$$

The $\alpha-$cuts of $\tilde{\Lambda}$ are given by

$$\Lambda(\alpha) = \{y : \eta_{\tilde{\Lambda}}(y) \geq \alpha\}. \tag{1.27}$$

The $\alpha-$cuts defined in (1.27) can be expressed in term of crisp intervals as:

$$\Lambda(\alpha) = \left[(y)_\alpha^{LB}, (y)_\alpha^{UB}\right] = \left[\min_{y \in Y}\{y : \eta_{\tilde{\Lambda}}(y) \geq \alpha\}, \max_{y \in Y}\{y : \eta_{\tilde{\Lambda}}(y) \geq \alpha\}\right]. \tag{1.28}$$

According to the convexity of fuzzy number $(\tilde{\Lambda})$, the upper and lower bounds of intervals defined in (1.28) can be obtained in terms of α as

$$\left[(y)_\alpha^{LB}, (y)_\alpha^{UB}\right] = \left[\min \eta_{\tilde{\Lambda}}^{-1}(\alpha), \max \eta_{\tilde{\Lambda}}^{-1}(\alpha)\right]. \tag{1.29}$$

Let $\tilde{U} = (u_1, u_2, u_3, u_4), u_1 < u_2 < u_3 < u_4,$ represent a trapezoidal fuzzy number whose membership function is given by

$$\xi_{\tilde{U}}(u) = \begin{cases} \dfrac{u - u_1}{u_2 - u_1}, & u_1 \leq u \leq u_2 \\ 1, & u_2 \leq u \leq u_3 \\ \dfrac{u_4 - u}{u_4 - u_3}, & u_3 \leq u \leq u_4 \end{cases} \tag{1.30}$$

If $\tilde{U} = (u_1, u_2, u_3, u_4)$ and $\tilde{V} = (v_1, v_2, v_3, v_4)$ be two trapezoidal fuzzy numbers with membership functions $\xi_{\tilde{U}}(u)$ and $\xi_{\tilde{V}}(v)$ respectively, then the fuzzy number $\tilde{U} * \tilde{V}$ is given by the membership function $\xi_{\tilde{U}*\tilde{V}}(w) = \sup_{(u,v): w = u*v} \min\{\xi_{\tilde{U}}(u), \xi_{\tilde{V}}(v)\}$, where "*" represents the algebraic operation between $\tilde{U}$ and $\tilde{V}$ (cf. Verma et al., 2009).

- **Addition:** $\tilde{U} \oplus \tilde{V} = (u_1 + v_1, u_2 + v_2, u_3 + v_3, u_4 + v_4)$.
- **Subtraction:** $\tilde{U} \ominus \tilde{V} = (u_1 - v_1, u_2 - v_2, u_3 - v_3, u_4 - v_4)$.
- **Multiplication:** $\tilde{U} \otimes \tilde{V} = (u_1 v_1, u_2 v_2, u_3 v_3, u_4 v_4)$.

- **Scalar multiplication:** $a \otimes \tilde{U} = \begin{cases} (au_1, au_2, au_3, au_4), \text{ if } a \geq 0 \\ (au_4, au_3, au_2, au_1), \text{ if } a < 0. \end{cases}$

- **Division:** $\tilde{U} \oslash \tilde{V} = \left(\frac{u_1}{v_4}, \frac{u_2}{v_3}, \frac{u_3}{v_2}, \frac{u_4}{v_1} \right)$.

The $\alpha-$cut of a fuzzy number $\tilde{U}$ in terms of crisp interval is given by

$$U(\alpha) = [u_1 + \alpha(u_2 - u_1), u_4 - \alpha(u_4 - u_3)] = [U_\alpha^L, U_\alpha^U] \tag{1.31}$$

Signed distance of fuzzy numbers

Let $W(\mathrm{R})$ denotes the family of fuzzy sets on the real number. For $\tilde{U}, \tilde{V} \in W(\mathrm{R}),$ the signed distance of $\tilde{V}$ to $\tilde{U}$ is denoted by $D(\tilde{U}, \tilde{V})$. We have

$$D(\tilde{U}, \tilde{V}) = \frac{1}{1-0} \int_0^1 [M(U(\alpha)) - M(V(\alpha))] d\alpha \tag{1.32}$$

or
$$D(\tilde{U}, \tilde{V}) = \frac{1}{2} \int_0^1 \left[(U_\alpha^U + U_\alpha^L) - (V_\alpha^U + V_\alpha^L) \right] d\alpha \tag{1.33}$$

where $M(U(\alpha)) = \frac{1}{2}\left[U_\alpha^L + U_\alpha^U \right]$ and $M(V(\alpha)) = \frac{1}{2}\left[V_\alpha^L + V_\alpha^U \right]$ for every α.

In particular, the signed distance of $\bar{0}$ to $\bar{U}$ is defined as

$$D(\tilde{U}, \tilde{0}) = \frac{1}{2} \int_0^1 (U_\alpha^U + U_\alpha^L) \, d\alpha \tag{1.34}$$

(ii) Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS is a hybrid soft computing technique and uses the features of both the neural network and fuzzy inference system. In ANFIS, 1st order Takagi-Sugano (TS) fuzzy system (cf. Zimmermann, 1996) is used for the fuzzification whereas, in order to train the data, it uses two phase algorithms, namely forward pass and backward pass. The basic features of ANFIS are briefly described as follows:

Phase 1: Least square method is applied to measure the consequent parameters of Takagi-Sugano (TS) type rules.

Phase 2: Gradient- descent method is applied to arrange the parameters of the antecedents.

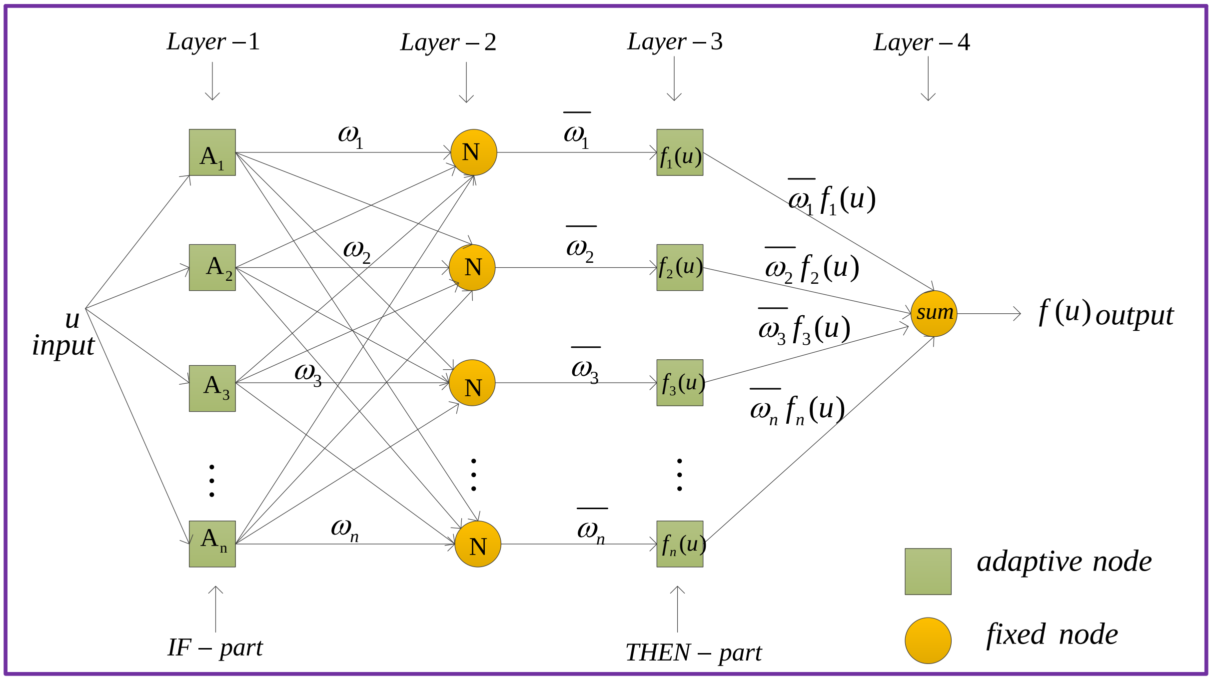


Figure 1.4: ANFIS Architecture

The prediction of ANFIS engages membership functions, fuzzy operators and *IF-THEN* rules. The 1st order TS fuzzy model with one input u and one out $f(u)$ is displayed in Figure 1.4. In general i^{th} rule is defined as:

$$i^{th}\ Rule: IF\ u\ \text{is}\ A_i\ THEN\ \ f_i = a_i u + b_i u + c_i,\ i = 1, 2, \ldots, n.$$

where a_i, b_i, c_i denote the adaptive parameters and A_i is the i^{th} fuzzy set.

Figure 1.4 represents the architecture for the ANFIS that consists of four layers which can be explained as:

Layer 1: This layer consists of adaptive nodes that implement the fuzzification. The membership function $m_i(u)$ of the i^{th} fuzzy set A_i is considered as Gaussian:

$$m_i(u) = exp\left[-\left(\frac{u - x_i}{y_i}\right)^2\right] \tag{1.35}$$

where x_i and y_i are the premise parameters.

Layer 2: In layer 2, all nodes are fixed and this layer gives the normalization value according to the formula:

$$\overline{\omega}_i = \frac{\omega_i}{\sum_{j=1}^{n} \omega_j} \tag{1.36}$$

where ω_i is the matching degree of the rule R_i.

Layer 3: This layer gives the conclusion part of the fuzzy rule and calculates the normalization and an affine function according to $\overline{\omega}_i f_i(u)$.

Layer 4: This single node is a fixed node labeled by sum and gives the overall output as the summation according to the formula:

$$f(u) = \sum_{i=1}^{n} \overline{\omega}_i f_i(u). \tag{1.37}$$

This method can be employed to concerned queueing model to validate the results obtained by analytical method with results obtained with ANFIS technique (cf. Jang, 1993; Jain and Meena, 2017).

1.4.8. Optimization Techniques

(i) Quasi-Newton Method (QNM)

This method is used to evaluate the minimum value of the system cost of the concerned queueing model corresponding to continuous optimal control parameters. This is an iterative method and provides the approximate minimum value of continuous and unimodal function $f(x_1, x_2)$ in the feasible range of x_1 and x_2. Let $f(x_1^*, x_2^*)$ denote the minimum value of $f(x_1, x_2)$ at optimal value (x_1^*, x_2^*).

Algorithmic steps of QNM

Inputs: Tolerance ε_{x_i} of $\left|\frac{\partial TC(x_1, x_2)}{\partial x_1}\right|$, $\left|\frac{\partial TC(x_1, x_2)}{\partial x_2}\right|$.

Output: Approximate solution of $(x_1^*, x_2^*, f(x_1^*, x_2^*))$.

Step 1: Set initial trial solution $\vec{M}_0 = [x_1, x_2]^T$ and evaluate the value of cost function $f(\vec{M}_0)$.

Step 2: Evaluate the cost gradient $\vec{\nabla} f(\vec{M}_0) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right]^T_{\vec{M}_0}$ and the Hessian matrix

$$H(\vec{M}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}.$$

Step 3: Find the new trial solution $M_{j+1} = M_j - \left[H(\vec{M}) \right]^{-1} \vec{\nabla} f(\vec{M}_j)$.

Step 4: Set $j = j+1$, and repeat step until $\max \text{tolerance} \left(\left| \frac{\partial f}{\partial x_1} \right|, \left| \frac{\partial f}{\partial x_2} \right| \right) < \varepsilon_{x_i}$, is the tolerance.

Step 5: Obtain the approximate minimum value $f(\vec{M}_j) = f(x_1^*, x_2^*)$.

Step 6: End.

(ii) Genetic Algorithm (GA)

A genetic algorithm is a search based approach which mimics the natural genetic. In order to evaluate the optimal value of the fitness function (objective function), it works in a multidimensional search space. Genetic algorithm can be proposed for the evaluation of optimal control parameters of the concerned queueing model. GA involves the fitness function (cost function), randomly generated population (all possible solutions to the given problem) and the three operators namely, (i) selection operator, (ii) crossover operator and (iii) mutation operator. We evaluate fitness function at every chromosome (one solution to the given problem). After evaluating fitness value at every chromosomes, we check how good solution we have obtained. This process is done repeatedly.

(i) ***Selection*** is a very important operator in GA to select the fittest chromosomes (chromosomes which have the best fineness value) from the entire population to crossover. Chromosomes with best fitness value have the strongest probability to be selected for the next step. We use a tournament method for the selection of chromosomes (parents). (ii) In the ***crossover***, the fittest chromosomes (parents) are selected from the population to crossover and produce children (offspring) for the next generation. Two point crossover method is adopted for the crossover of chromosomes. Crossover is said to be successful if the new children (offsprings) have the best fitness value as compare to their parents. Crossover probability is considered to be high to produce new children if it is chosen very less then produced new children are the same as parents. (iii) ***Mutation*** is an alteration in produced new children (offspring) from the crossover of chromosomes. In the evaluation of optimal service rate of queueing model, bit inversion mutation is selected (*i.e.* 0 is changed into 1 or 1 is changed into 0) and a small close interval [0, 1] is chosen to generate random values within it if the generated value is less than the mutation rate then a bit at a random position is selected and its value is changed.

After mutation, we check the predefined stopping criteria. If the search process is met to stopping criteria then best chromosomes are picked up and final fitness value (optimal value) is evaluated at that chromosomes, otherwise, the search process is continued until stopping criteria is met. The stopping criteria is taken in terms of a number of generations.

The description of GA, we refer Mitchell (1998), Hourani (2004). The pictorial view of the flowchart of the proposed genetic algorithm is shown in Figure 1.5.

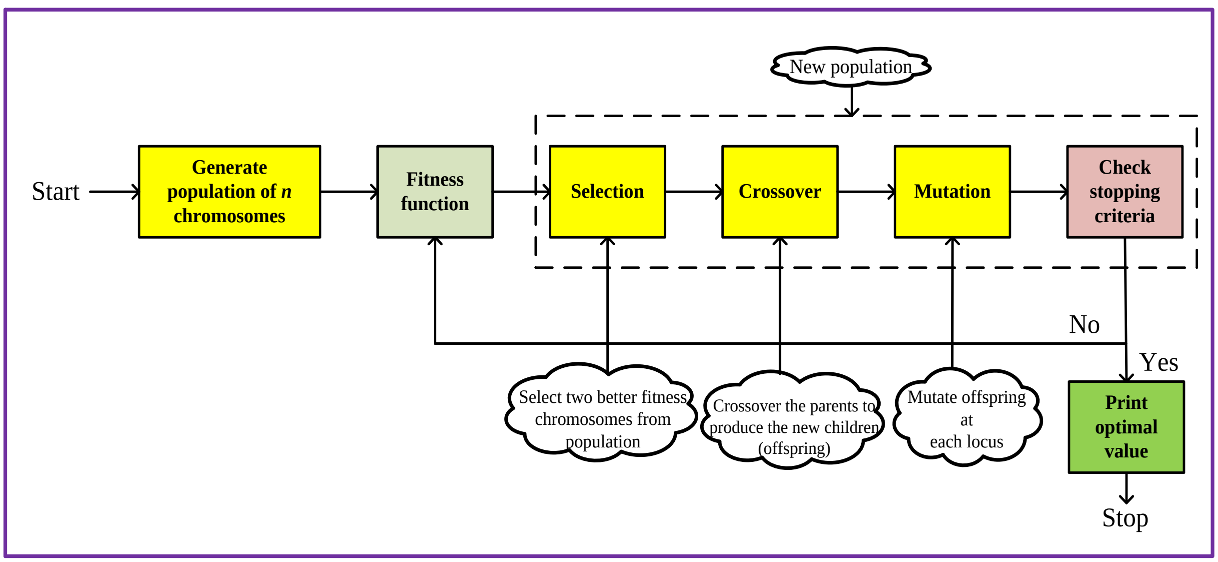


Figure 1.5: Genetic algorithm flow chart

The pseudo code for GA used for getting the optimal control parameter and minimum cost of the concerned queueing model is briefly described in Table 1.1.

Table 1.1: Pseudo code for GA

```
GA( )
initialize population of chromosomes
evaluate fitness of chromosomes
  while (stopping criteria is reached) do
    parent selection
    crossover
    mutation
  decode and fitness calculation
  find best
return best
```

(iii) Harmony Search Algorithm

This algorithm is also used to determine the optimal control parameters and minimum cost of the queueing models. Harmony search (HS) is a musician inspired meta-heuristic algorithm that has found diverse applications since its inception in 2001. Harmony search is a simple yet efficient algorithm and is explained briefly in this section. While playing music, a musician has the following three choices:

a) Play some famous piece of music exactly as it is.

b) Play a known piece of music with few alterations.

c) Compose entirely new nodes.

The three options were formulated by Geem et al. (2001) as an optimization process and the corresponding components became harmony memory (HM), pitch adjusting, and randomization. Table 1.2 provides the pseudo code of harmony search algorithm. Steps 1 through 4 define the control parameters of the algorithm. In Step 5, the HM is initialized with random solutions, keeping each component of the solution within bounds. In order to use harmony memory effectively, a harmony memory parameter corresponding to rate HMCR $\in [0,1]$ is utilized. A high value of HMCR may produce inferior results as the solution space which is not explored properly and on the other hand, a low value of HMCR may result in slow convergence because the best solution remains under utilized. The typical range of HMCR is 0.7 to 0.95.

Table 1.2: Harmony search algorithm

Define objective function $f(H)$.
Define the harmony memory consideration rate (HMCR).
Define pitch adjustment rate (PAR) and bandwidth (BW).
Define harmony memory size (HMS).
Initialize harmony memory (HM).
while (Stopping Criteria Not Reached) ***do***
  Find current **Worst** and **Best** harmony in *HM*.
    for $i = 1$ to d ***do***
      if($rand \leq HMCR$) ***then***
        $H_i = HM_i^j$ where $j = rand\_\mathrm{int}(1, HMS)$
        if(rand $\leq$ PAR) ***then***
          $H_i = H_i \pm rand \times BW$
        end if
      Else
        Generate H_i randomly within the allowed bounds.
      end if
    end for
  if (H is better than worst harmony in HM) ***then***
    Update HM by replacing WORST harmony by ***H***.
  end if
end while
print ***Best*** Harmony as an obtained solution.

The second component pitch adjustment is determined by a pitch bandwidth (BW) or frets width (FW) and pitch adjusting rate (PAR) (cf. Geem and Yoon, 2017)). Pitch adjustment corresponds to altering the current solution for generating a slightly different solution. Pitch can be adjusted linearly or nonlinearly, however, most often linear adjustment is used. Thus, we have

$$H_i^{new} = H_i^{old} + BW \times r_i,\ r_i \in [-1,1] \text{ and } 1 \leq i \leq d \tag{1.38}$$

where H_i^{old} (H_i^{new}) is the i^{th} component of the existing (new) harmony or solution and BW is the bandwidth. The above relation (1.38) yields a new solution around the existing solution by altering it slightly. Here r_i is a random number lies between -1 to 1 and *d* represents the total number of components in the Harmony. The recommended PAR value lies in $[0.1, 0.5]$.

The third component of the HS is the randomization, which is used to enhance the exploration of the search space. In the pseudo code of harmony search algorithm given in Table 1.2. Here H represents a potential solution or Harmony, $\text{rand} \in [0,1]$ is a uniformly distributed random number generator, rand_int (1, HMS) generates a uniformly distributed integer random number between 1 and HMS, size of harmony memory is represented as HMS and *d* is the dimension of the problem.

The parameters HMCR, PAR, and BW of the HS algorithm are tuned depending on the problem at hand.

1.5. SOME QUEUEING MODELS UNDER ADMISSION CONTROL *F*-POLICY (ACF-P)

In this section, we briefly describe the state dependent finite capacity models under admission control *F*-policy dealing with a variants features. In many real life, the queue length can never be infinite due to some physical constraints viz. finite waiting space. The admission control *F*-policy for finite capacity Markov model was introduced in mid 90s to control the congestion of arrivals in the system.

1.5.1. M/M/1/K Queue Model under ACF-P without Startup Time

To formulate the admission control scenario, the basic M/M/1/K queueing model without startup time under *F*-policy is described as follows:

The arriving customers/jobs join the system in Poisson fashion with rate λ and are served by single server following exponential distribution with mean $1/\mu$. When *K* customers/jobs accumulate in the system, a set-up job as per exponential distribution with rate ε is required to restrict the customers/jobs from joining the system.

Define $\delta_{n-F} = \begin{cases} 1, & n = 1, 2, ..., F, \\ 0, & \text{elsewhere.} \end{cases}$ and $\delta'_{n-F} = \begin{cases} 1, & n = F-1, \\ 0, & n = 1, 2, ..., F-2, F, ..., K-2. \end{cases}$

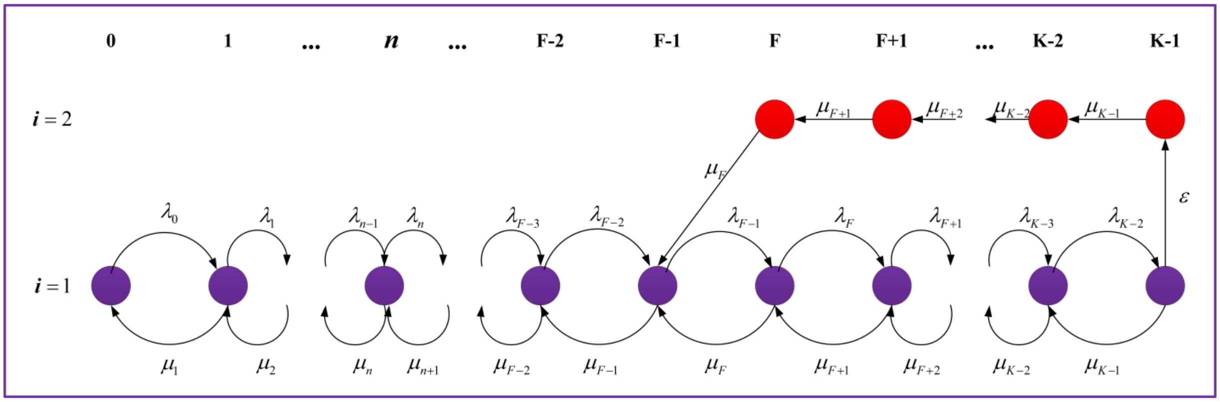


Figure 1.6: Transition state diagram for M/M/1/K model under *F*-policy without startup time

The further entry of jobs are allowed till number of jobs in the system ceases to '*F*'. For the first-come-first-server discipline, Chapman-Kolmogorov (C-K) steady state equations are framed and respective transition state diagram is depicted in Figure 1.6.

$$-\lambda_0 P_{1,0} + \mu_1 P_{1,1} = 0 \tag{1.39}$$

$$-(\lambda_n + \mu_n) P_{1,n} + \lambda_{n-1} P_{1,n-1} + \mu_{n+1} P_{1,n+1} + \delta'_{n-F} \mu_{n+1} P_{2,n+1} = 0,\ n = 1, 2, ..., K-2 \tag{1.40}$$

$$-(\varepsilon + \mu_{K-1}) P_{1,K-1} + \lambda_{K-2} P_{1,K-2} = 0 \tag{1.41}$$

$$\mu_n P_{2,n} = \mu_{n+1} P_{2,n+1} = \varepsilon P_{1,K-1},\ n = F, F+1, ..., K-2 \tag{1.42}$$

Classical queueing methods such as matrix method, recursive method, etc. can be applied to solve the Equations (1.39)-(1.42). Once probability distribution in product form is established, several performance indices including average queue length, carried load, etc. can be obtained.

1.5.2. Finite Retrial Queueing Model under ACF-P

We formulate the state dependent retrial queueing model operating under *F*-policy by considering startup time to return back to normal state from the state where admission of customers are not allowed.

For M/M/1 model having retrial attempts following exponential distribution with rate γ and state dependent arrival rate (λ_n) and service rate (μ_n), the governing Chapman-Kolmogorov (C-K) equations can be formulated by balancing the in-flow with out-flow rates as shown in Figure 1.7.

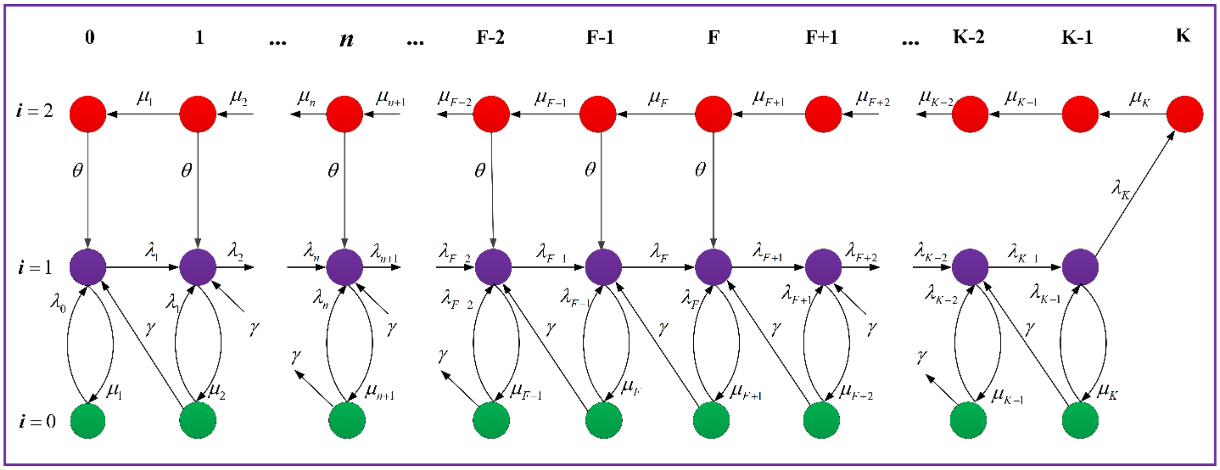


Figure 1.7: Finite capacity retrial model under *F*-policy

The server status $S(\tau)$ is defined as

$$S(\tau)=i=\begin{cases}1(2): & \text{server is busy in rendring the service and admission of customers is allowed} \\ & \text{(not allowed),} \\ 0: & \text{customers are in retrail orbit.}\end{cases}$$

For finite capacity retrial queueing model, the governing equations are given by:

$$-(\lambda_{n+1}+\mu_{n+1})P_{1,n}+\delta_{n-F}\lambda_n P_{1,n-1}+\lambda_n P_{0,n}+\gamma P_{0,n+1}+\theta P_{2,n}=0,\ n=0,2,...,F \tag{1.43}$$

$$-(\lambda_{n+1}+\mu_{n+1})P_{1,n}+\lambda_n P_{1,n-1}+\lambda_n P_{0,n}+\gamma P_{0,n+1}=0, n=F+1,...,K-2 \tag{1.44}$$

$$-(\lambda_K+\mu_K)P_{1,K-1}+\lambda_{K-1}P_{1,K-2}+\lambda_{K-1}P_{0,K-1}=0 \tag{1.45}$$

$$-(\theta+\delta_{n-F}\mu_n)P_{2,n}+\mu_{n+1}P_{2,n+1}=0,\ n=0,1,...,F \tag{1.46}$$

$$-\mu_n P_{2,n}+\mu_{n+1}P_{2,n+1}=0, n=F+1,...,K-1 \tag{1.47}$$

$$-\mu_K P_{2,K}+\lambda_K P_{1,K-1}=0 \tag{1.48}$$

$$-\lambda_0 P_{0,0}+\mu_1 P_{1,K-1}=0 \tag{1.49}$$

$$-(\lambda_n+\gamma)P_{0,n}+\mu_{n+1}P_{1,n}=0, n=1,2,...,K-1 \tag{1.50}$$

1.5.3. Finite Queueing Model with Unreliable Server under ACF-P

In the queueing system; the server's breakdown is a common phenomenon. Due to over congestion of customers in the system, sometimes server may not bear the load as such break down occurs. Now, we describe two unreliable server finite models having state dependent rates and operating under *F*-policy. In the first model the extra feature of startup time on reaching the threshold level 'F' in Level 2 (see Figure 1.8, $i=2$) is included so as to return back to Level 1 (see Figure 1.8, $i=1$) where admission of customers is allowed. The second model is developed by taking same assumptions excluding startup time.

(i) Unreliable Server Model with Startup Time

Now we formulate the state dependent single unreliable server Markov model under admission control F-policy which coincides with model described in Section 1.3 when server is taken to be reliable. The server may fail at any time following Poisson process with rate α and the system organizer immediately sends it to repair so that the service can be resumed without much delay. The repair process follows exponential distribution with rate β. The transition state diagram is depicted in Figure 1.8.

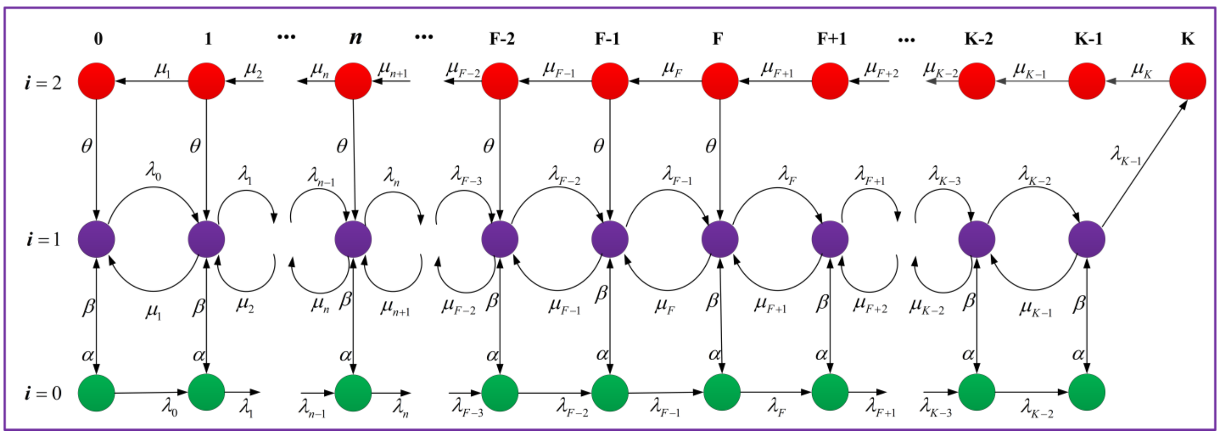


Figure 1.8: Transition diagram for M/M/1/K queueing model with unreliable server

The server status $S(\tau)$ is

$$S(\tau)=i=\begin{cases}1(2): & \text{server is busy in rendring the service and admission of customers is allowed} \\ & \text{(not allowed),} \\ 0: & \text{server is in broken down state.}\end{cases}$$

The steady state equations of finite capacity queueing model with unreliable server under admission control F-policy are formulated as follows:

$$-(\lambda_n+\delta_{n-F}\mu_n+\alpha)P_{1,n}+\delta_{n-F}\lambda_{n-1}P_{1,n-1}+\mu_{1,n+1}P_{1,n+1}+\beta P_{0,n}+\theta P_{2,n}=0, n=0,1,2,...,F \tag{1.51}$$

$$-(\lambda_n+\mu_n+\alpha)P_{1,n}+\mu_{1,n+1}P_{1,n+1}+\beta P_{0,n}=0, n=F+1,F+2,...,K-2 \tag{1.52}$$

$$-(\lambda_{K-1}+\mu_{K-1}+\alpha)P_{1,K-1}+\lambda_{K-2}P_{1,K-2}+\beta P_{0,K-1}=0 \tag{1.53}$$

$$-(\theta+\delta_{n-F}\mu_n)P_{2,n}+\mu_{n+1}P_{2,n+1}=0, n=0,1,...,F \tag{1.54}$$

$$-\mu_n P_{2,n} + \mu_{n+1} P_{2,n+1} = 0, n = F+1, F+2, ..., K-1 \quad (1.55)$$

$$-\mu_K P_{2,K} + \lambda_{K-1} P_{1,K-1} = 0 \quad (1.56)$$

$$-(\lambda_0 + \beta) P_{0,0} + \alpha P_{1,0} = 0 \quad (1.57)$$

$$-(\lambda_n + \beta) P_{0,n} + \lambda_{n-1} P_{0,n-1} + \alpha P_{1,n} = 0, n = 1, 2, ..., K-2 \quad (1.58)$$

$$-\beta P_{0,K-1} + \lambda_{K-2} P_{0,K-2} + \alpha P_{1,K-1} = 0 \quad (1.59)$$

To derive probability distributions, Equations (1.51)-(1.59) can be solved using various methods available in the queueing literature. Furthermore, to determine the several system metrics the queue size distribution can be used.

(ii) Unreliable Server Model without Startup Time

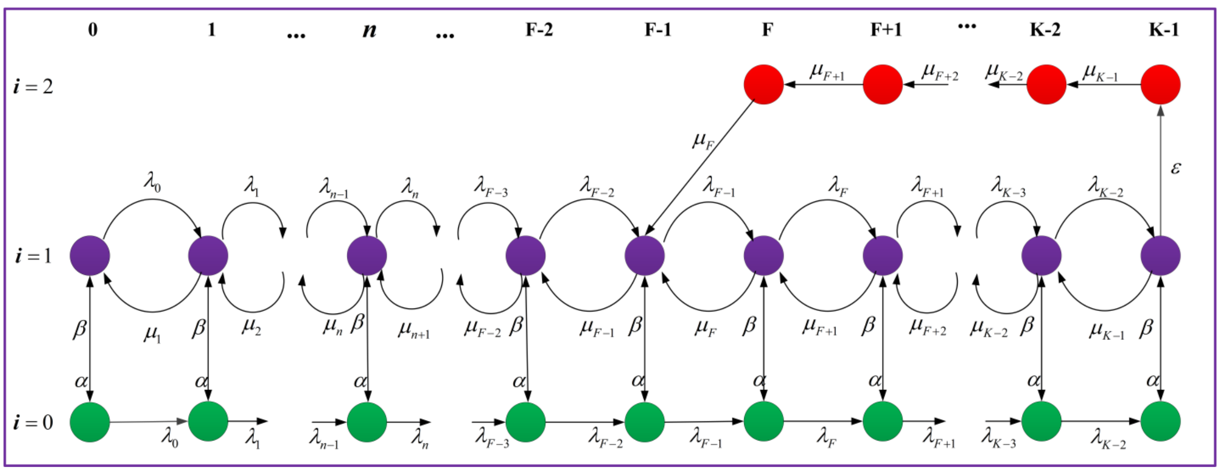


Figure 1.9: M/M/1/K queueing model with unreliable server under F-policy without startup time

F-policy model for unreliable server can also be formulated by taking setup job (ε) and relaxing the startup times (θ). The state transition diagram is shown in Figure 1.9. Then admission F-policy without startup times the single server Markovian queueing model with state dependent rates and unreliable server is formulated by framing C-K equations as follows:

$$-(\lambda_0 + \alpha) P_{1,0} + \mu_1 P_{1,1} + \beta P_{0,0} = 0 \quad (1.60)$$

$$-(\lambda_n + \mu_n + \alpha) P_{1,n} + \mu_{n+1} P_{1,n+1} + \lambda_{n-1} P_{1,n-1} + \beta P_{0,n} + \delta'_{n-F} \mu_{n+1} P_{2,n+1} = 0,\ n = 1, 2, ..., K-2 \quad (1.61)$$

$$-(\varepsilon + \mu_{K-1} + \alpha) P_{1,K-1} + \lambda_{K-2} P_{1,K-2} + \beta P_{0,K-1} = 0 \quad (1.62)$$

$$\mu_n P_{2,n} = \mu_{n+1} P_{2,n+1} = \varepsilon P_{1,K-1},\ n = F, F+1, ..., K-2 \quad (1.63)$$

$$-(\lambda_0 + \beta) P_{0,0} + \alpha P_{1,0} = 0 \quad (1.64)$$

$$-(\lambda_n+\beta)P_{0,n}+\alpha P_{1,n}+\lambda_{n-1}P_{0,n-1}=0,\ n=1,2,...,K-2 \tag{1.65}$$

$$-\beta P_{0,K-1}+\lambda_{K-2}P_{0,K-2}+\alpha P_{1,K-1}=0 \tag{1.66}$$

1.5.4. Finite Population Model under Admission Control *F*-Policy

This section presents the admission control based on *F*-policy for some finite population model. The description of basic finite population models and their mathematical formulation are given (cf. Jayaraman and Matis, 2011). We describe admission control *F*-policy for the finite Markov model with population size K and single repairman who takes care of failed unites/machines. Let the number of failed units/machines present in the system at any time τ is n, then the life time of failed units/machines and repair process of failed units/machines are governed by exponential distribution with parameter λ and μ, respectively. Thus, the effective failure rate is given by $\lambda_n=(K-n)\lambda$. The steady state balance equations are framed by equating the in-flows and out-flows in particular case of state dependent rates when $\tau\to\infty$ i.e. for steady state, by making substitution $\lambda_n=(K-n)\lambda$ and $\mu_n=\mu$ in Equations (1.1)-(1.6).

1.6. LITERATURE SURVEY

In many queueing systems, the important issue is to determine optimal control parameters in order to achieve the goal of minimum cost and pre-specified grade for the service. The optimization problem of a queueing system is mainly concerned with the determination of optimal control parameters, such as optimal input rate, strategic joining probability, optimal service rate, optimal number of servers, optimal threshold parameter, etc. There are wide applications of such queueing models in day-to-day congestion scenarios. It is worthwhile to present the survey of literature related to the work done in the thesis.

The remarkable contributions by several researchers in the field of optimal control strategies for state dependent queueing systems have been described into various sub-sections. The finite queueing models under admission control *F*-policy incorporating different features are given in Sections 1.6.1-1.6.5. In Section 1.6.6, the literature review of the state-dependent queues dealing with finite population model, time sharing model and discouragement models is given. Section 1.6.7 is devoted to the literature review on fuzzy queueing models.

1.6.1. Finite Capacity Model under Admission Control *F*-Policy

The notable works related to finite Markov model under admission control F-policy are as follows. Gupta (1995) was the first who introduced the concept of admission control F-policy in finite capacity queueing model. By incorporating the admission control F-policy in the single server finite capacity queueing model, he used recursive method to derive the steady state results for the probability distributions. He also investigated the finite model under N-policy based on control on service and then developed the relation between F-policy and N-policy.

Ke et al. (2010) dealt with the finite capacity model using the concept of primary service and second optional service. To control the congestion of customers in the system, the admission control F-policy feature was incorporated by Ke et al. (2010). They derived various performance measures and constructed a cost function. Furthermore, numerical illustration to validate the concerned finite capacity model was given.

Huang et al. (2011) investigated admission control policy for two server finite capacity Markov model. In this study, they controlled the arrival of the customers by adding extra server along with the existing server when the capacity of the customers crossed the pre-defined fixed value (say M). The added extra server was removed when the capacity of the system decreased to a pre-specified value (say Q). The analytical result for the queue size distribution provided was further used to establish the various system metrics and cost structure.

Yang and Chang (2015) studied F-policy model of Gupta (1995) by transforming it into fuzzy environment using arrival rate, service rate and start-up rate as triangular fuzzy numbers. They have used $\alpha-$cuts and non-linear programing approaches to construct the membership functions for the system size. Jain and Bhagat (2015b) dealt with admission control for finite retrial queueing model and threshold recovery policy. In their study, the transient behavior of the system was studied. Runga-Kutta fourth order method was used to solve the set of differential equations. The numerical results for queue length and cost function were presented by taking an appropriate illustration.

1.6.2. Finite Working Vacation Model under Admission Control *F*-Policy

In literature on admission control F-policy for finite capacity model, some works have involved the working vacation concept. The worth noting articles in this areas are as follows:

Yang at al. (2010) used the concept of single working vacation in finite queueing model under admission control *F*-policy. They applied matrix analytic method to analyze the steady state model and derived various system metrics. A non-linear cost function was framed and minimized using direct search method and quasi-Newton method in order to determine optimal control parameters.

Jain et al. (2016) investigated the finite population model by controlling the arrivals of failed machines in the system. The single working vacation feature was used to make the model more effective. Further, Chapman-Kolmogorov steady state equations of finite population model were solved using successive over relaxation method. Several performance measures and cost function have derived to explore the utility of model in practical situations.

Jain et al. (2017) investigated the admission control *F*-policy for single server vacation policy for unreliable server queue and developed finite population model. To analyze the finite population model along with working vacation feature, they framed the steady state equations which were solved using SOR method.

Jain and Meena (2017) dealt with fault tolerant multi-component machining system with working vacation and admission control *F*-policy. The steady state analysis of fault tolerant multi-component machining system was done by solving steady state equations. They derived several system metrics by taking an illustrative example to examine the influences of system parameters on system metrics. Further, ANFIS technique, which is one of the hybrid soft computing approach was given to validate the feasibility of computational results.

1.6.3. Finite Unreliable Server Queue under Admission Control *F*-Policy

By surveying the literature, it is found that the some research works have been done towards the admission control *F*-policy for unreliable server queueing model. Wang and Yang (2009) discussed admission control *F*-policy model for unreliable server. The steady state analysis of the system was done using matrix analytical method. They derived various system indices and also constructed a cost function. Quasi-Newton method based numerical technique was used to minimize the system cost.

Chang et al. (2011) considered $M/H_2/1$ model with unreliable server operating under admission control *F*-policy. In this case, the unreliable server provides two type of service with some probability. Steady state analysis was done to establish the system indices and cost function. Jain et al. (2012) investigated M/M/2/K model with combined *N*-policy and *F*-policy. As per their model, the first server renders the service when at-least *N* customers accumulated. The second server is turned on only when the system reaches to its full capacity. The repair of failed server is provided in multi-phase.

Chang et al. (2014a) extended the finite queueing model of Gupta by incorporating the concept of balking and unreliable server. The steady state equations for the finite queueing model was represented in matrix form. Moreover, optimal management of the concerned model was done by minimizing the system cost corresponding to different system parameters using direct search and quasi-Newton method. Jain and Bhagat (2015b) investigated the finite optimal control *F*-policy while developing the finite capacity Markov model with unreliable server. The transient behavior of the concerned system was investigated and several system metrics ware derived using system probabilities.

1.6.4. Finite Population Models under Admission Control *F*-Policy

The relevant works done in queueing literature towards finite population model under admission control *F*-policy are as follows:

Kumar and Jain (2013) dealt with Markov model for multi-component machining system with single repairman by incorporating admission control *F*-policy and service control *N*-policy. They investigated multi-component machining system in steady state. The recursive method was used to derive the analytical results for both *F*-policy and *N*-policy models. Chang et al. (2013) studied the controlling of arrival of failed machines for a warm standby supported multi component machining system with switching failure probability. The steady state solution was obtained using matrix method. Further, the minimum cost was evaluated at optimal system parameters using direct search method and quasi-Newton method.

Jain and Bhagat (2014) have done the transient analysis of finite population model under the provision of admission control *F*-policy and threshold recovery policy. Numerical technique based on Runga-Kutta method was used to solve the transient equations and to derive various performance metrics and cost function. Jain et al. (2016) discussed the admission control for machine repair problem with working vacation for the in-flow of failed machines. The steady state analysis was done by solving Chapman-Kolmogorov equations by employing the SOR method. The system indices such as queue size of failed machines, delay in repair of failed machines, long run probabilities, etc. were derived. A non-linear cost function corresponding to system capacity, threshold parameter F and service rate was constructed which is further minimized using direct search method.

Shekhar et al. (2017) dealt with F-policy and N-policy for redundant machining system with permanent single repairman and one additional repairman as per requirement. In this study, they considered that both repairmen (permanent and additional) provide repair as per time sharing basis. The steady state equations were solved using Cramer's rule to obtain the queue size distributions and other performance metrics. Jain et al. (2017) dealt with machine repair problem with admission control and server working vacation policy. The SOR method was used to determine the solution of the steady state linear equations. A nonlinear cost function was framed with decision variables threshold parameter 'F' and repair rate. Direct search and quasi-Newton method were used to determine the optimal parameter values. Jain and Meena (2017) investigated fault tolerance machining system under the admission control of failed machines in the system for repair and server working vacation. In this study, they analyzed the fault tolerance system by applying the numerical technique based SOR method to find unknown steady state probabilities which are further used in derivation of system metrics.

Kumar et al. (2019) considered two unreliable servers for MRP with warm standbys and F-policy. The transient state analysis of the finite population model was done using matrix method. Several system metrics such as number of failed machines in the system, throughput, long run probabilities, etc. were derived. A cost function was also framed with decision variable repair rate. To minimize it, direct search method was implemented. Jain et al. (2019) investigated F-policy for single repairman fault tolerance machining system with warm standby and working vacation. The transient analysis of the fault tolerance model was done using Runga-Kutta fourth order method. Several system metrics such as long run probabilities, queueing and reliability indices were derived.

1.6.5. Non-Markovian Queueing Model under Admission Control *F*-Policy

In the queueing literature, only a very few papers related to non-Markovian queueing model have appeared; the brief account of the same is as follows:

Karaesmen and Gupta (1997) developed the duality relationship between N-policy and F-policy by considering G/G/1/K queueing model which controlled the service process or arrival process. As per their study, the server turned off when the service process was controlled whereas the arrival process stopped/started depending upon the queue size. Wang et al. (2007)

used supplementary variable approach to analyze finite non-Markov queue by considering the admission control of customers in a finite capacity system. They considered service time for the customers as general distributed and arrival times and startup times as exponentially distributed. The recursive approach was used to establish the probability distributions of queue size. For the computational purpose, exponential distribution and Erlang-3 distribution were used for the service times to examine the optimality of threshold parameter '*F'*. Wang et al. (2008) dealt with a finite model by involving the control of arriving customers using *F*-policy. In their model, the inter-arrival time of the customers was considered general distributed whereas service times and startup times are taken as exponential. The supplementary variable technique (SVT) corresponding to remaining arrival times was used to develop the steady state model. Further, recursive method was used to derive the results for the finite non-Markov model. Moreover, exponential distribution, Erlang-3 distribution and deterministic distribution were considered for inter-arrival time to compute the numerical results and optimal value of '*F'*.

Kuo et al. (2011) established an inter-relation between *N*-policy based on service control policy and *F*-policy based on admission control policy. They presented two models viz. M/G/1/K model under *N*-policy and G/M/1/K model under *F*-policy. Using SVT and recursive method, steady state analytical results for both models were derived. Further, they concluded that the results of M/G/1/K model under *N*-policy can be obtained from the results for G/M/1/K model under *F*-policy. Goswami (2016) investigated two finite queueing models under (i) (*p*, *F*)-policy and (ii) (*q*, *N*)-policy. The author considered discrete distribution viz. geometric distribution for arrival times and service times in place of continuous distribution. The steady state analysis was done for both models to establish the queue size distribution using recursive method. Further, the relation between (*p*, *F*)-policy model and (*q*, *N*)-policy model have established which is further verified results for (*p*, *F*)-policy model using (*q*, *N*)-policy model.

1.6.6. State Dependent Queueing Models

In queueing literature, many researchers have developed the state dependent queueing models under different assumptions (cf. Hadidi, 1974; Gupta and Rao, 1998; Gupta et al., 2017). Massey (2002) considered the time-varying rates to explore the performance of a queueing system and discussed the applications of the developed queueing model in a telecommunication system. A single server queueing system in which the arrival and service rates depend on the system states, was analyzed by Adan and Kulkarni (2003). An important work on M/M/1 retrial queue was done by Parthasarathy and Sudhesh (2007) by considering the state dependent rates. Lee (2011) developed the state-dependent stochastic networks using the birth-death process and established the different moments and stability properties of the system. A single server bulk queueing model using threshold policy was studied by Banerjee and Gupta (2012) by including the features of controlling the arrivals and batch service schedule. Kumara and Dharsana (2015) analyzed the congestion problem by developing a single server Markovian model with queue size dependent arrival rate and impatient customers. Recently, an M/M/1 queueing model with queue size dependent service rate was studied by Rodrigues et al. (2016). Recently, Ernst et al. (2018) investigated single-server multi-class fluid queue by considering state-dependent arrival rate. In order to analyze the fluid model, they obtained stability condition for the fluid queue. Hu et al. (2019) dealt with fluid queue models for traffic circulation systems by taking state-dependent service rate.

Based on arrival and or service rates, the state dependent queueing models can be further classified into different categories including the queueing models with the additional removable server, queue with discouragement, finite population i.e. machine interference models, time-sharing models, etc.

(i) Finite population model

The failure of machines is a major difficulty not only for the users but also the loss of revenue to the organizers/manufactures in the concerned machining system. In such cases, to overcome the problems of machine failures and delay in production, many researchers contributed towards finite population models which also dealt with machine repair problems under different assumption (cf. Haque and Armstrong, 2007; Liou, 2014; Jain et al., 2016; Huang et al., 2016; Chen, 2018). Sometimes, it is seen that due to less workload, the server may remain idle most often which is the wastage of revenue as well as time. The applications of machine repair model with retrial attempts can be found in computer repair shop, telecommunication system, call centers, etc. (Choudhury and Ke, 2014; Ke et al., 2013; Wang et al., 2018). Recently, Yang and Chang (2018) dealt with machine repair problem with general retrial policy to explore the performance of the system. They have used the supplementary variable technique to provide explicit results for the queue size distributions.

(ii) Time sharing model

The concept of time sharing is also used in a few scenarios including computer communication networks, manufacturing and production systems, etc. Many researchers have paid their attention to time-sharing models for computer systems (cf. Adiri and Avi-Itzhak, 1969; Jain et al., 2012; Kim and Kim, 2007). Jain et al. (2005) investigated a time-shared machine repair problem with mixed spares. In this investigation, they have considered that the caretaker of failed machines operates under N-policy. The online optimization issues of machining system used for cloud computing were examined by Chandrasekaran et al. (2013). For manufacturing–remanufacturing systems, the time sharing machining system was studied by Flapper et al. (2014). Telek and Houdt (2018) investigated MAP/GI/LPS-k(m) processor share queue to study the response time distribution using numerical method.

(iii) Queue with customer's discouragement

Most often, the balking behavior of the customers occurs only in a queuing system where customers can actually observe the queues. Some renowned researchers contributed their pioneer works towards the multi-server queue with balking (cf. Abou-El-Ata and Hariri, 1992; Do and Chakka, 2010; Ke and Wang, 1999; Wang and Chang, 2002; Wüchner et al., 2009). A few researchers have contributed significantly towards the finite capacity queueing model with balking and reneging (cf. Choudhury and Medhi, 2011; Kumar and Sharma, 2012; Vijaya Laxmi et al., 2013). Recently, Bouchentouf and Messabihi (2018) and Som and Kumar (2018) developed a heterogeneous queueing model by considering balked and reneged customers. They presented analytical results for queue size distribution and other performance metrics by employing the recursive method. In the queueing literature, a few researchers have presented their extensive works on different aspects related to retrial queues with balking (cf. Artalejo and Lopez-Herrero, 2000; Ke and Chang, 2009; Chang et al., 2018). A single server retrial queue with balking was also investigated by Gao et al. (2017). In this study, they analyzed Nash equilibrium to explore the customers' joining strategies based on the cost reward structure. Most recently, Ke et al. (2019) analyzed a multiple server Markov model for retrial queue with vacation and balking by using the matrix geometric method.

1.6.7. Fuzzy Queueing Models

The queueing literature has a large number of researches towards the crisp queueing models with retrial attempts. In such studies, the queueing parameters corresponding to inter-arrival, service and retrial process in crisp queueing models, follow some specific probability distributions. Several queueing parameters can be defined in linguistic manner and may be both probabilistic and possibilistic. In literature, a very few number of researchers have focused on the fuzzy queueing modeling while its application may be seen in every sphere of life due to the fact that the some of the queueing descriptors may not be crisp and should be defined in terms of linguistic variables. For the practical point of view, fuzzy models (cf. Mohanty and Passi, 2010; Mishra et al., 2017; Jaggi et al., 2018) are much appropriate and applicable in many real time systems as compared to frequently used crisp models. Yang and Chang (2015) used the results of Gupta (1995) and extended their work to a finite fuzzy queue with control *F*-policy. Bagherinejad and Pishkenari (2016) obtained fuzzified expected queue length and the average waiting time of M/M/C queueing model. They constructed the membership functions by using a parametric nonlinear programing approach. Mueen et al. (2017) investigated a single fuzzy queue, by adopting parametric nonlinear programing to frame a hexagonal membership function. Fuzzy bulk queueing model for the communication system was studied by Bhardwaj (2017) and Bhardwaj et al. (2019) in fuzzy environment using Zadeh's extension principle and $\alpha-$cut approach.

1.7. OUTLINE OF THE THESIS

The applicability of the state dependent queueing model along with optimal control strategies can be seen in every sphere of life where queues are formed. The work done on the optimal strategies and evaluation of performance indices of state dependent queueing models is organized into eight chapters of the thesis. The ongoing Chapter 1 is concerned with the introduction of the modeling and methodological aspects of research investigation done in the thesis. Chapters 2-5 contribute towards the finite capacity queueing models under admission control *F*-policy with variant features. Chapters 6-7 present the finite population models with specific features such as retrial, vacation and standbys, etc. Chapter 8 is devoted to the joining and balking strategies of customers for double orbit retrial queueing model. The chapter-wise brief outlines of the contents of the thesis are as follows:

Chapter 1 is devoted to the introductory part of the investigation done in the thesis. The preliminaries of the control strategies and *F*-policy for queueing models are given. The methodologies and techniques used to establish the performance indices and cost optimization for the concerned queueing model are briefly described. The mathematical formulation of some specific models related to our research works have been briefly outlined. The literature survey of relevant topics has been given. The organization of the thesis and research scope of the work done are briefly described.

Chapter 2 is concerned with the control policy for the retrial queueing model with server breakdowns. The startup time and threshold policy are taken into account to develop Markov model. Gauss-Seidel method is applied to present the steady state results. Adaptive neuro fuzzy inference system (ANFIS) technique is also implemented to compare the numerical results obtained with ANFIS results.

Chapter 3 deals with the finite state-dependent queueing model under admission control *F*-policy. The model is investigated by incorporating the general retrial attempts. The supplementary variable corresponding to the remaining retrial time and recursive method are used to derive the performance indices. The optimal control parameters and minimum cost of the system are determined using direct search method and quasi-Newton method.

Chapter 4 presents the multi-server finite queueing model under admission control of customers based upon *F*-policy along with customer's balking behavior. In order to reduce the balking behavior of the customers, there is the provision of one additional server so as to shorten the queue length formed by the customers in rush hour. The recursive technique is applied to establish the steady-state queue size probability distributions. Cost analysis is also done by evaluating the optimal control parameters such as admission control parameter, capacity of the system, number of servers and the service rate.

Chapter 5 deals with the admission control policy for the single server finite capacity queueing model with general distributed retrial times and discouraged customers. The recursive method and soft computing based artificial neuro fuzzy inference system (ANFIS) approaches are applied to establish various performance indices. Genetic algorithm (GA) and quasi-Newton method are used to minimize the expected cost of the system.

Chapter 6 contributes the queueing analysis of the finite population models with retrial orbit under admission control *F*-policy. The machine repair problem of repairable redundant system with finite retrial orbit and general distributed retrial policy is investigated. The supplementary variable corresponding to remaing retrial time is used to develop C-K equations. The recursive method is used to derive the steady state probabilities and queueing and reliability indices. A fuzzy cost function is formulated by considering the cost elements as trapezoidal fuzzy numbers. The defuzzification of the cost function is done using signed distance method. The genetic algorithm (GA) is applied to determine the optimal control parameter and minimum cost of the system.

Chapter 7 is concerned with the multi-component M/G/1/K machine repair system with standby support and vacation. The queue size distributions have been derived using supplementary variable technique followed by recursive method. To deal with a realistic scenario, the machine repair model is transformed from crisp to fuzzy environment by considering the system parameters as fuzzy numbers. The harmony search algorithm is also implemented to determine optimal control parameters and minimum cost of the system.

Chapter 8 contains three double orbit retrial queueing models in which the arriving customers are categorized into two classes namely, ordinary customers and premium customers. The steady state analytical solution for the probability distribution and system performance measures are derived by using probability generating function. The first model investigates the customers' joining strategy in a double orbit retrial queueing system with balking. Double orbit retrial model is then transformed into fuzzy environment by using system parameters expressed as linguistic variables. The cost optimization is also done using GA to determine optimal service rates. The second model is concerned with the double orbit feedback model with balking. The third model develops the unreliable server Markov model with customers' balking behavior. ANFIS technique is also implemented to authenticate the numerical results. Optimal service rates are evaluated using quasi-Newton method by constructing the cost function.

In the end of the thesis, the conclusions and future scope of the investigation done have been outlined to highlight the contributions and significance of the work done. The relevant references have been listed in alphabetical order at the end of this thesis.

1.8. CONCLUDING REMARKS

The optimal policies and strategic management are important means of the queueing models which can be used to control the congestion and improve the quality of the service. The numerous applications of state-dependent queueing models under optimal control strategies encourage the queueing theorists to develop state dependent queueing models which can be well suited to the real life queueing scenarios. The present chapter contributes to the introductory part of the work done in the thesis. The brief account about the preliminary concepts related to the work done, state dependent queues, methodologies, relevant literature, etc. are provided in this chapter. The control policies described may be successfully used in many congestion problems where queues are built up. In view of various applications of admission control policy, some queueing models under strategic joining policy and threshold control policies are investigated. The literature available on optimal state-dependent queues under control strategies exhibits the wide applications in call center, production and manufacturing system, health care center, etc.

Chapter 2

F-Policy for Markovian Retrial Queue with Server Breakdown

2.1. INTRODUCTION

Retrial queue can be seen everywhere around us such as at shopping malls, banks, in front of post office, etc. In several real-life day-to-day as well as in industrial/business queueing scenarios, the jobs arriving in the system may be compelled to leave the service area and move to the retrial pool in the case when the server is busy. The jobs from the retrial pool can try again and again for the service after a random period of time so as to avail the service. The recent works on retrial queue can be found in the articles by Nobel (2016), Phung-Duc et al. (2017) and Chang and Wang (2018). Most of the existing research works referred to retrial queue mainly focus on the reliable server model, however retrial queueing model with server breakdowns depicts more realistic queueing situations. Sherman and Kharoufeh (2006) proposed Markovian retrial queueing model with an unreliable server. They provided the stability conditions and several stochastic decomposability results for the concerned model.

The most common issue involved in queueing system is to control the admission and service of the customers. The research works regarding controllable queues can be divided into two broad categories (i) control the service and (ii) control the arrival of the customers. The concept of admission control *F*-policy is mainly used to maintain the smooth functioning of the queueing system. *F*-policy can be applied to control the congestion by not allowing the customers to join the system when its capacity is full. For example, in a factory, all incoming raw materials are handled by one machine. Due to space limitations (say K), incoming raw materials are not allowed to enter the system when the system capacity is full. The further raw materials join only when the current stock of raw materials falls to a certain threshold value (F). Preparation time must also be accounted before raw materials are allowed to enter the system. *F*-policy can also be implemented in several other queueing scenarios of communication networks, manufacturing and production units and many more systems to prevent the critical problems of blocking and delay in the system.

So far as the controlling of the arrivals in the finite capacity system is concerned, *F*-policy is quite useful to control the congestion of the jobs and helps in reducing the lost jobs in particular when the system capacity is full. In *F*-policy, the admission of the jobs is to be restricted due to the limitation of buffer size and it can be compensated by providing an opportunity to join the system again by pushing the jobs in the retrial orbit.

Jain and Bhagat (2015) dealt with *F*-policy for the finite capacity retrial queueing model with delayed repair and threshold recovery. They obtained the transient probabilities using Runge-Kutta method in order to evaluate various performance indices. Control *F*-policy is recently appeared in the work of Yeh et al. (2017) who analyzed the two phase single server finite capacity queueing model under (p, *F*) policy. They have used a matrix method to establish the steady state solution of the queue size distribution and several other system indices. In recent years, some research works have appeared on queueing model towards *F*-policy (cf. Yang and Yang, 2018; Kuamr et al., 2019).

Adaptive neuro fuzzy inference system (ANFIS) is a combination of neural networks and fuzzy logic and this can be used for the performance prediction of complex systems for which analytical model cannot be developed easily. ANFIS which is a hybrid soft computing technique may be used for the performance prediction of a wide variety of problems including the financial engineering, automobile with automatic transmission, telecommunication system, etc. The architecture and learning process of ANFIS technique was first proposed by Jang (1993) to study the mapping of input-output based human knowledge. In recent years, modeling of electro chromic device by using ANFIS was done by Dounis et al. (2016). They used ANFIS network and performed many experiments by taking training data and testing data. For ANFIS modeling in queueing theory, the contributions of Jain and Upadhyaya (2009) and Jain and Meena (2017) are worth-noting.

In the present chapter, we develop a single server retrial Markov queueing system operating under *F*-policy and supported by an unreliable server. The adaptive neuro-fuzzy inference system (ANFIS) approach is successfully implemented to authenticate the numerical results obtained by analytical method of the concerned retrial queueing system. The rest of the chapter is organized in the following manner. In Section 2.2, the description of the model is given. Chapman-Kolmogorov steady state equations and system indices are given Section 2.3, and Section 2.4 respectively. Section 2.5 is devoted to the sensitivity analysis by taking a numerical example and using ANFIS approach. Finally, the concluding remarks are given in Section 2.6.

2.2. DESCRIPTION OF THE MODEL

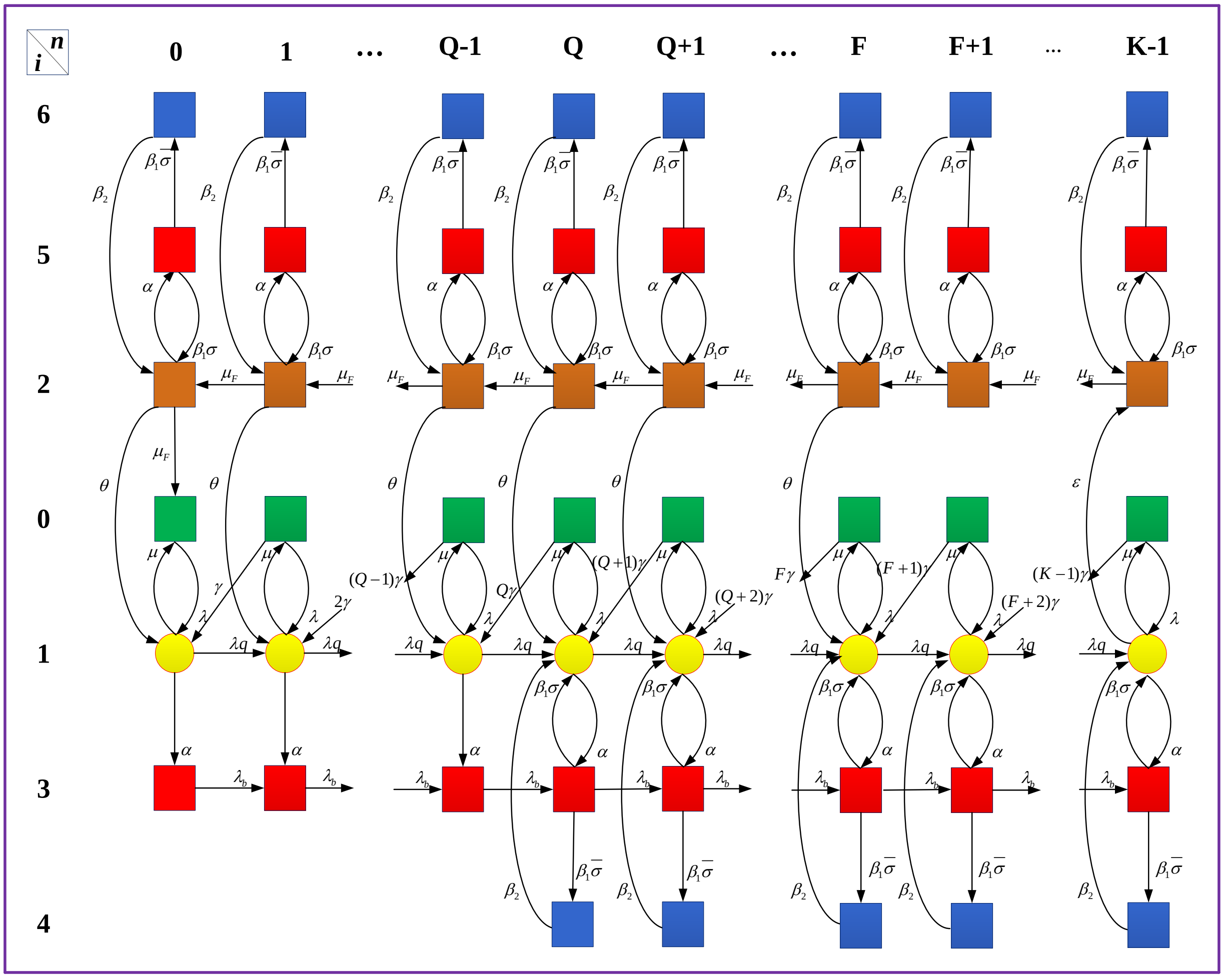


Figure 2.1: Transition state diagram

Consider M/M/1/K retrial queueing model under admission control *F*-policy and startup time. The arriving customers join the system as per first-come-first-served (FCFS) discipline. The formulation of Markov model is done based on the certain assumptions which are as follow.

The arriving customer enters into the system with probability q according to the Poisson process with rate λ. The service to the customers is rendered according to exponential distribution with mean $1/\mu$. If the arriving customer finds the server busy, he enters to the retrial orbit of finite capacity (K). From the retrial orbit, the customers retry for the service following exponential distribution with mean $1/\gamma$. In the case when an arriving customer finds the server free, he immediately gets the service and leaves the system. When the system capacity becomes full, then a setup job is required to stop the customers from joining the system; the time required for the setup is assumed to be exponentially distributed with rate ε. Once the system capacity becomes full, the customers are not allowed to join the system until the number of customers in the system are further ceases to prefixed threshold level 'F'. It is assumed that the server requires a startup time with mean $1/\theta$ before allowing the customers in the system after the number of customers in the system drops to 'F' level. The customers are served during *F*-policy according to exponential distribution with mean $1/\mu_F$. The server may breakdown at any time; the breakdown occurs as per Poisson process with rate α. When the server breaks down, it cannot be repaired immediately until the number of customers in the system accumulated to a threshold value 'Q'. The broken-down server is sent for repair where primary repair job is done according to exponential distribution with mean $1/\beta_1$. After getting primary repair, the server is recovered with probability σ; otherwise broken down server goes to secondary phase of repair with probability $\bar{\sigma}=1-\sigma$, and the repair is done by the same repairman according to exponential distribution with mean $1/\beta_2$.

At the time τ, the random variables $N(\tau), S(\tau)$ denote the number of customers in the system and status of the server, respectively. The server status $S(\tau)$ is defined as follows:

$$S(\tau)=\begin{cases} 0, & \text{Server is busy and the arrivals are forced to join the orbit,} \\ 1, & \text{Server is busy and the arrivals are allowed to enter the system,} \\ 2, & \text{Server is busy and the arrivals are not allowed to enter the system,} \\ 3, & \text{Server is in broken down during normal operation and is primary phase repair,} \\ 4, & \text{Server is broken down during normal operation and is secondary phase repair,} \\ 5, & \text{Server is broken down during F-policy and is primary phase repair,} \\ 6, & \text{Server is broken down during F-policy when customers' arrived is not allowed and} \\ & \text{is secondary phase repair.} \end{cases}$$

We define the system states probabilities at epoch τ by $P_{i,n}(\tau)=\text{Prob}\{S(\tau)=i, N(\tau)=n\}$ for node (i,n) as depicted in Figure 2.1. It is noted that $\{S(\tau), N(\tau): \tau \geq 0\}$ is a bi-variate Markov process which is discrete in state space and continuous in time. We shall analyze the model at steady state, *i.e.*, when $\tau \rightarrow \infty$ denoting probability by $P_{i,n}=\lim_{t\rightarrow\infty} P_{i,n}(\tau)$

2.3. GOVERNING EQUATIONS

Now by using the probability arguments, the steady state Chapman- Kolmogorov equations for the system states for seven levels $S(\tau) = i\,(0 \le i \le 6)$ are constructed as follows:

(i) **For system state** $i = 0, 0 \le n \le K-1$.

$$-\lambda P_{0,0} + \mu_F P_{2,0} + \mu P_{1,0} = 0 \tag{2.1}$$

$$-(\lambda + n\gamma)P_{0,n} + \mu P_{1,n} = 0,\ 1 \le n \le K-1 \tag{2.2}$$

(ii) **For system state** $i = 1, 0 \le n \le K-1$.

$$-(\alpha + \mu + q\lambda)P_{1,0} + \lambda P_{0,0} + \theta P_{2,0} + \gamma P_{0,1} = 0 \tag{2.3}$$

$$-(\alpha + \mu + q\lambda)P_{1,n} + \lambda q P_{1,n-1} + \lambda P_{0,n} + \theta P_{2,n} + (n+1)\gamma P_{0,n+1} = 0, 1 \le n \le Q-1 \tag{2.4}$$

$$-(\alpha + \mu + q\lambda)P_{1,n} + \lambda q P_{1,n-1} + \lambda P_{0,n} + \theta P_{2,n} + (n+1)\gamma P_{0.n+1} + \beta_1 \sigma P_{3,n} + \beta_2 P_{4,n} = 0, \quad Q \le n \le F \tag{2.5}$$

$$-(\alpha + \mu + q\lambda)P_{1,n} + \lambda q P_{1,n-1} + \lambda P_{0,n} + (n+1)\gamma P_{0,n+1} + \beta_1 \sigma P_{3,n} + \beta_2 P_{4,n} = 0, \quad F+1 \le n \le K-2 \tag{2.6}$$

$$-(\alpha + \mu + \varepsilon)P_{1,K-1} + \lambda q P_{1,K-2} + \lambda P_{0,K-1} + \beta_1 \sigma P_{3,K-1} + \beta_2 P_{4,K-1} = 0 \tag{2.7}$$

(iii) **For system state** $i = 2,\ 0 \le n \le K-1$

$$-(\alpha + \theta + \mu_F)P_{2,n} + \beta_2 P_{6,n} + \beta_1 \sigma P_{5,n} + \mu_F P_{2,n+1} = 0, 0 \le n \le F \tag{2.8}$$

$$-(\alpha + \mu_F)P_{2,n} + \beta_2 P_{6,n} + \beta_1 \sigma P_{5,n} + \mu_F P_{2,n+1} = 0, F+1 \le n \le K-2 \tag{2.9}$$

$$-(\alpha + \mu_F)P_{2,K-1} + \beta_2 P_{6,K-1} + \beta_1 \sigma P_{5,K-1} + \varepsilon P_{1,K-1} = 0 \tag{2.10}$$

(iv) **For system state** $i = 3,\ 0 \le n \le K-1$

$$-\lambda_b P_{3,0} + \alpha P_{1,0} = 0 \tag{2.11}$$

$$-\lambda_b P_{3,n} + \alpha P_{1,n} + \lambda_b P_{3,n-1} = 0, 1 \le n \le Q-1 \tag{2.12}$$

$$-(\lambda_b + \beta_1)P_{3,n} + \alpha P_{1,n} + \lambda_b P_{3,n-1} = 0,\ Q \le n \le K-2 \tag{2.13}$$

$$-\beta_1 P_{3,K-1} + \alpha P_{1,K-1} + \lambda_b P_{3,K-2} = 0 \tag{2.14}$$

(v) **For system state** $i=4,\ Q\le n\le K-1$

$$-\beta_2 P_{4,n}+\beta_1\bar{\sigma}P_{3,n}=0,\ Q\le n\le K-1 \tag{2.15}$$

(vi) **For system state** $i=5,\ 0\le n\le K-1$

$$-\beta_1 P_{5,n}+\alpha P_{2,n}=0,\ 0\le n\le K-1 \tag{2.16}$$

(vii) **For system state** $i=6,\ 0\le n\le K-1$

$$-\beta_2 P_{6,n}+\beta_1\bar{\sigma}P_{5,n}=0,\ 0\le n\le K-1 \tag{2.17}$$

By realizing that the analytical approach to solve the system of equations (2.1)-(2.17) is quite tedious, we convert it into matrix equations $\mathbf{AP}=\mathbf{0}$**,** where $\mathbf{A}$ is a square coefficient matrix formed by the coefficients of unknown probabilities $P_{i,n}$ of equations (2.1)-(2.17), $\mathbf{P}$ is a probability vector whose elements are $P_{i,j}, i=0,1,...,6$ and $j=0,1,...,K-1$. To get $P_{i,j}$ we replace the last row of the coefficient matrix $\mathbf{A}$ by a row with all elements 1 and vector **0** by a vector $\mathbf{B}$ whose all elements are zero except last one which is taken as 1. So, we have a matrix equations $\mathbf{A'P}=\mathbf{B}$ which can be easily solved by using numerical technique viz. Gauss-Seidel method**.** The system states probabilities can be further used to evaluate the system performance measures as established in the next section.

2.4. PERFORMANCE MEASURES

To analyze the system characteristics, we establish various performance measures which are as follows:

(i) Expected number of customers in the system

$$E[N_S]=\sum_{n=0}^{K-1} nP_{0,n}+\sum_{n=0}^{K-1}\sum_{i=1,2,3,5,6}(n+1)P_{i,n}+\sum_{n=Q}^{K-1}(n+1)P_{4,n} \tag{2.18}$$

(ii) Expected number of customers in the queue and in the retrial orbit are

$$E[N_q]=\sum_{n=0}^{K-1}\sum_{i=1,2,3,5,6} nP_{i,n}+\sum_{n=Q}^{K-1} nP_{4,n} \tag{2.19a}$$

and $$E[N_R]=\sum_{n=0}^{K-1} nP_{0,n} \tag{2.19b}$$

(iii) System throughput is obtained using

$$TP = \mu \sum_{n=0}^{K-1} P_{1,n} + \mu_0 \sum_{n=0}^{K-1} P_{2,n} \tag{2.20}$$

(iv) The probability of the server being idle and busy respectively, are

$$P_I = \sum_{n=0}^{K-1} P_{0,n} \text{ and } P_{SB} = \sum_{n=0}^{K-1} (P_{1,n} + P_{2,n}) \tag{2.21a-b}$$

(v) The probabilities that server is in broken down state and waiting for the repair

$$P_{BD} = \sum_{n=0}^{Q-1} P_{3,n} \tag{2.22}$$

(vi) The probabilities that server is broken down while the customers are allowed to join the system

$$P_{SR} = \sum_{n=0}^{K-1} (P_{5,n} + P_{6,n}) + \sum_{n=Q}^{K-1} (P_{3,n} + P_{4,n}) \tag{2.23}$$

(vii) Failure frequency of the server

$$F_f = \alpha \sum_{n=0}^{K-1} (P_{1,n} + P_{2,n}) \tag{2.24}$$

(viii) Cost function

We construct a cost function per unit time in the system with service rate as decision variable. For constructing the cost function, various cost elements associated with different activities are used. The cost factors involved in cost function are as follows:

C_I : Cost per unit time when the server is idle,
C_B : Cost per unit time when the server is busy,
C_H : Holding cost per unit time of each customer present in the system,
C_R : Cost per unit time incurred on repairng of the broken down server,
C_F : Cost per unit time for providing service to the customer when the arrivals are not allowed,
C_A : Cost per unit time for providing service to the customer when the arrivals are allowed,
C_S : Cost for startup process when the customers are allowed to enter in the system,
C_O : Cost per unit time incurred on each customer in the retrial orbit.

Now we formulate the cost function as:

$$TC(\mu) = C_I P_I + C_B P_{SB} + C_H E[N_q] + C_R [P_{BD} + P_{SR}] + \mu_F C_F + \mu C_A + \theta C_S + C_O E[N_R] \tag{2.25}$$

2.5. NUMERICAL RESULTS

In this section, we present the numerical illustration to analyze the effects of system parameters on various performance measures. The numerical results are obtained by coding the computer program in MATLAB software. For the computation purpose, we set the default parameters as $K=6,\ F=3,\ \lambda=2,\ \mu=\mu_F=10,\ \gamma=0.2,\ \varepsilon=0.2,\ \theta=3,\ \alpha=0.7,\ \beta_1=0.8,$ $\beta_2=0.9,\ \sigma=0.5,\ \lambda_b=0.3$. To determine the optimal service rate and associated minimum cost, we consider the three cost sets as given in the Table 2.1.

Table 2.1: Cost elements for cost sets (in $)

Cost Set	C_I	C_B	C_H	C_R	C_F	C_A	C_S	C_o
I	200	400	10	50	4	5	10	15
II	100	200	10	50	4	5	10	15
III	200	400	20	25	4	5	10	15

The optimal control parameter 'μ' is determined using a direct search method and corresponding minimum cost which is depicted in Figure 2.2. The optimal service rate and corresponding minimum cost $TC(\mu^*)$ for the three cost sets are recorded in Table 2.2.

Table 2.2: $(\mu^*, TC(\mu^*))$ for different cost sets

Cost Set	$(\mu^*, TC(\mu^*))$		
	$q=0.3$	$q=0.6$	$q=0.9$
I	(5.322, $326.06)	(6.667, $344.40)	(7.595, $359.14)
II	(4.080, $207.50)	(5.180, $224.26)	(5.798, $237.43)
III	(5.544, $326.52)	(7.010, $346.36)	(8.059, $362.26)

To carry out the sensitivity analysis, numerical results are displayed in Tables 2.3-2.6 and Figures 2.4-2.7.

Table 2.3: Performance measures by varying values of λ and *q*

q	λ	$E[N_Q]$	$E[N_R]$	TP	F_f	TC
	2	1.0856	0.8859	1.1157	0.1116	1122.08
0.3	4	2.1820	1.2214	1.7661	0.1766	1131.17
	6	2.8819	1.1525	2.0730	0.2073	1140.20
	2	1.3444	1.2288	1.2424	0.1242	1118.77
0.6	4	2.5744	1.4556	1.8517	0.1852	1132.73
	6	3.1489	1.2591	2.1176	0.2118	1141.64
	2	1.5804	1.5047	1.3322	0.1332	1117.91
0.9	4	2.7461	1.5500	1.8837	0.1884	1133.87
	6	3.2617	1.3036	2.1389	0.2139	1142.18

Table 2.4: Performance measures by varying values of μ and q

q	μ	$E[N_Q]$	$E[N_R]$	TP	F_f	TC
	10	0.7913	0.7929	1.2197	0.0854	1535.65
0.3	25	0.3623	0.5138	1.4857	0.0416	3615.36
	40	0.2417	0.3820	1.6141	0.0282	5704.44
	10	0.9333	1.1031	1.3402	0.0938	1532.01
0.6	25	0.3845	0.6781	1.5464	0.0433	3613.87
	40	0.2506	0.4949	1.6517	0.0289	5703.82
	10	1.0884	1.4045	1.4445	0.1011	1529.99
0.9	25	0.4088	0.8543	1.6089	0.0450	3612.52
	40	0.2596	0.6131	1.6904	0.0296	5703.23

Table 2.5: Performance measures by varying values of θ and q

q	θ	$E[N_Q]$	$E[N_R]$	TP	F_f	TC
	3	1.0856	0.8859	1.1157	0.1116	1122.08
0.3	6	1.0872	0.8889	1.1165	0.1117	1152.02
	9	1.0879	0.8903	1.1169	0.1117	1181.99
	3	1.3444	1.2288	1.2424	0.1242	1118.77
0.6	6	1.3505	1.2389	1.2449	0.1245	1148.61
	9	1.3534	1.2438	1.2462	0.1246	1178.54
	3	1.5804	1.5047	1.3322	0.1332	1117.91
0.9	6	1.5933	1.5248	1.3368	0.1337	1147.66
	9	1.5994	1.5346	1.3389	0.1339	1177.55

Table 2.6: Performance measures by varying values of γ and q

q	γ	$E[N_Q]$	$E[N_R]$	TP	F_f	TC
	0.2	1.0856	0.8859	1.1157	0.1116	1122.08
0.3	1.1	0.7948	0.1987	0.9051	0.0905	1140.89
	2.0	0.7556	0.1119	0.8719	0.0872	1143.95
	0.2	1.3444	1.2288	1.2424	0.1242	1118.77
0.6	1.1	0.8991	0.2747	0.9944	0.0994	1139.56
	2.0	0.8389	0.1531	0.9463	0.0946	1143.69
	0.2	1.5804	1.5047	1.3322	0.1332	1117.91
0.9	1.1	1.0296	0.3622	1.0970	0.1097	1138.15
	2.0	0.9418	0.2004	1.0336	0.1034	1143.28

The ANFIS approach is also implemented using the neuro-fuzzy tool in MATLAB software. The membership functions of the input parameters λ, μ, θ and γ are taken as a trapezoidal function by taking each one as (i) very low (ii) low (iii) average (iv) high and (v) very high values as depicted in Figure 2.3. The ANFIS results (see tick marks of circle, square and diamond) for the expected number of customers in the system have been plotted in Figures 2.4-2.7 along with the numerical results by varying parameters λ, μ, θ and γ obtained. As expected, $E[N_S]$ increases (decreases) as λ and q (μ and γ) increases. These figures show

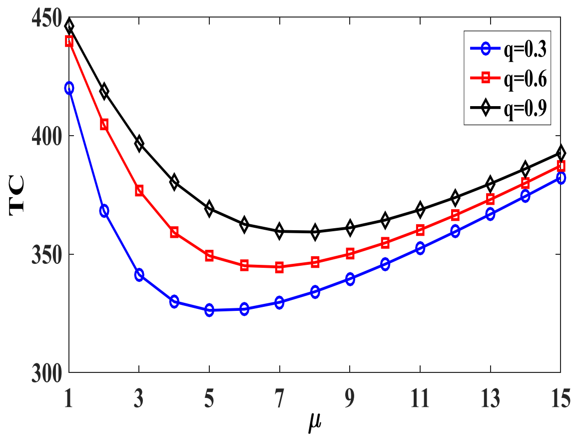


Figure 2.2: TC vs μ for different values of q

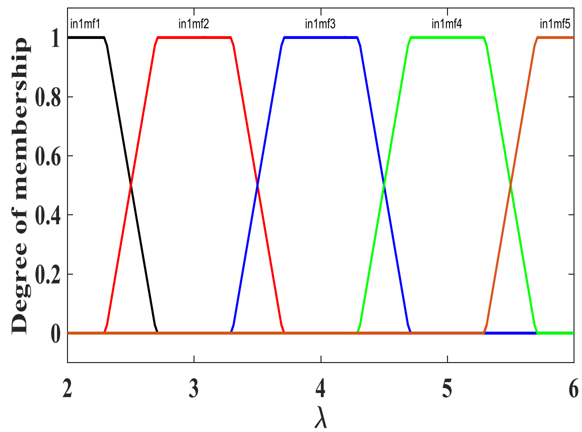


Figure 2.3: Membership function for λ

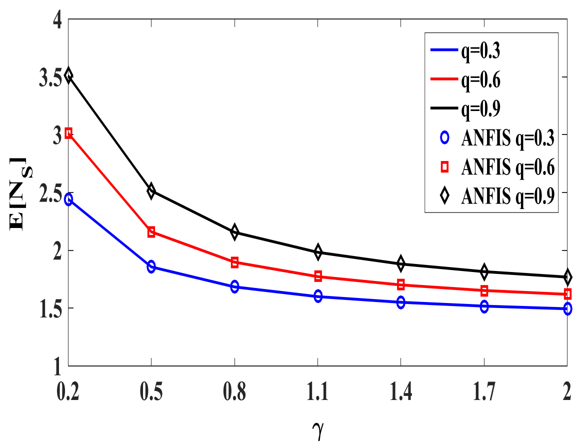


Figure 2.4: $E[N_S]$ vs γ for $q = 0.3, 0.6, 0.9$

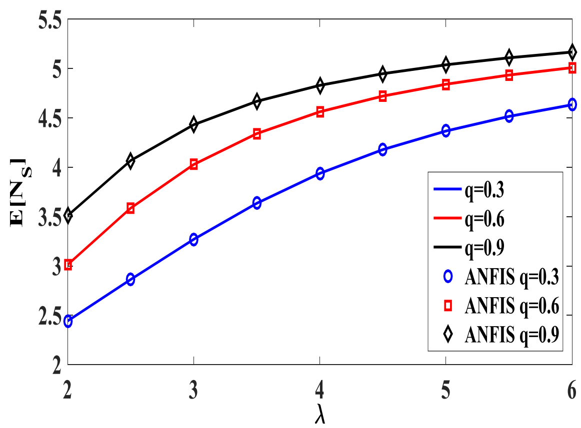


Figure 2.5: $E[N_S]$ vs λ for $q = 0.3, 0.6, 0.9$

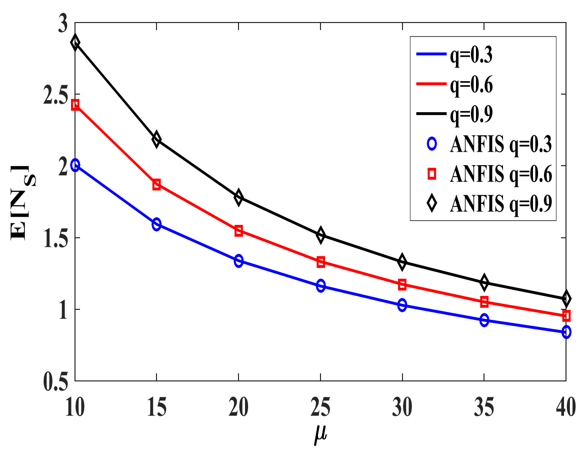


Figure 2.6: $E[N_S]$ vs μ for $q = 0.3, 0.6, 0.9$

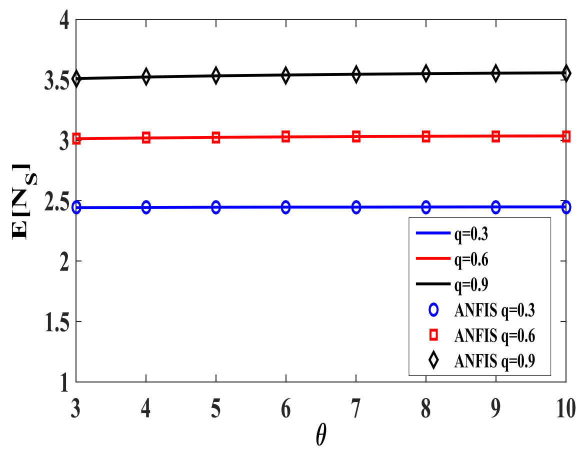


Figure 2.7: $E[N_S]$ vs θ for $q = 0.3, 0.6, 0.9$

almost a collinear trend for both numerical and ANFIS results which imply that the results obtained by ANFIS are at par with the analytical results.

2.6. CONCLUDING REMARKS

In this chapter, we have investigated the performance analysis of M/M/1/K retrial queueing system with server breakdown by incorporating several realistic features such as balking, threshold recovery and admission control policy. Numerical results are compared with the results obtained by using adaptive neuro fuzzy inference system (ANFIS) which demonstrates the future scope and usefulness of neuro-fuzzy tool for the performance prediction of real world queueing systems operating under several techno-economic constraints.

Chapter 3

Finite Capacity Queue with General Retrial and State-Dependent Rates

3.1. INTRODUCTION

Queueing models with state dependent arrival of jobs and state dependent service have useful applications in various congestion situations of day to day life activities. In industrial scenarios, the utilities of queueing models can also be seen in production and manufacturing systems, transportation and service systems, etc. In many congestion situations, the arrival and service rates depend on the present state of the system. The speed of service provided by the server may depend on the present work load. For example, the queue size affects the efficiency of the server; the server may render service with faster rate, in case of long queue. Similarly, some servers act slowly under the pressure of the long queue size which decreases the service rate. There are two types of queueing systems; in some systems, the service terminates at any specified time whereas in the other, service continues till the queue becomes empty. The state dependent service rate may be applicable for the service systems having a human being as server and can be seen in the production system, where the service rate of the server is relatively less when there is much burden of the workload. For the notable works related to state dependent queueing models, we refer the work by Banerjee and Gupta (2012) and Rodrigues et al. (2016).

Based on *F*-policy, non-Markovian M/G/1/K and G/M/1/K queueing models were investigated by Wang et al. (2007, 2008). For the detail description of the admission control *F*-policy, we refer Section 1.5 of Chapter 1. To highlight the practical utility of *F*-policy in finite capacity retrial queueing models, we cite the queue formed at the shopping center wherein the arriving jobs on finding the busy server, may wait in the retrial pool and return back after some time with the hope that the server becomes free. In such queueing scenarios, when the capacity of system becomes full, the admission of customers can be controlled via *F*-policy.

The M/G/1 queueing model has been extensively implemented by several researchers using supplementary variable technique in different frameworks to study the non-Markovian queueing service systems. Some researchers have contributed towards the analysis of retrial queueing models in different structures (cf. Chang and Wang, 2018; Phung-Duc and Kawanishi, 2019). An M/G/1 retrial queueing model was investigated by Moreno (2004) by considering general retrial times. The author also presented the condition of ergodicity for the system and established analytical results for the stationary distribution and other performance measures. Gao et al. (2014) developed an M/G/1 queueing model by considering the general distributed retrial times, working vacations and interruptions due to server breakdown. They have obtained the stationary state probability distribution by using the supplementary variable method. To obtain the queue size distribution and probability generating function (PGF) of the joint distributions of the queue size, M/G/1 queue with general distributed retrials times and Bernoulli vacation was dealt by Choudhury and Ke (2014). A stochastic comparison of Markov chains was proposed by Boualem et al. (2014) for the study of single server queue with retrial times as general distributed. In the work of Yang et al. (2016), the unreliable server retrial queue with general distributed retail attempts was studied by employing the supplementary variable method to establish several performance measures.

In this chapter, we develop a finite queueing model in generic set up by considering many realistic features such as admission control policy, general retrial times and state-dependent arrival and service processes. To analyze the retrial model under *F*-policy, this chapter is arranged in different sections. Section 3.2 presents a model description of the concerned problem. In Section 3.3, equations for the non-Markovian model are framed by using supplementary variable corresponding to remaining of retrial time. Some special models deduced from our study are given in Section 3.4. Various system indices and cost structure are established in Section 3.5. By taking the appropriate illustration, numerical experiment and sensitivity analysis and cost optimization of machine repair problem and time sharing model, are presented in Section 3.6. Finally, Section 3.7 presents the conclusion of the investigation done.

3.2. DESCRIPTION OF THE MODEL

For the queueing scenario with admission control according to *F*-policy, we consider a single server finite capacity (say K) queueing model with general retrial attempts. The service discipline for rendering the service to the jobs follows the first-come-first-served (FCFS) rule. The formation of the model is based on certain assumptions which are outlined as follows:

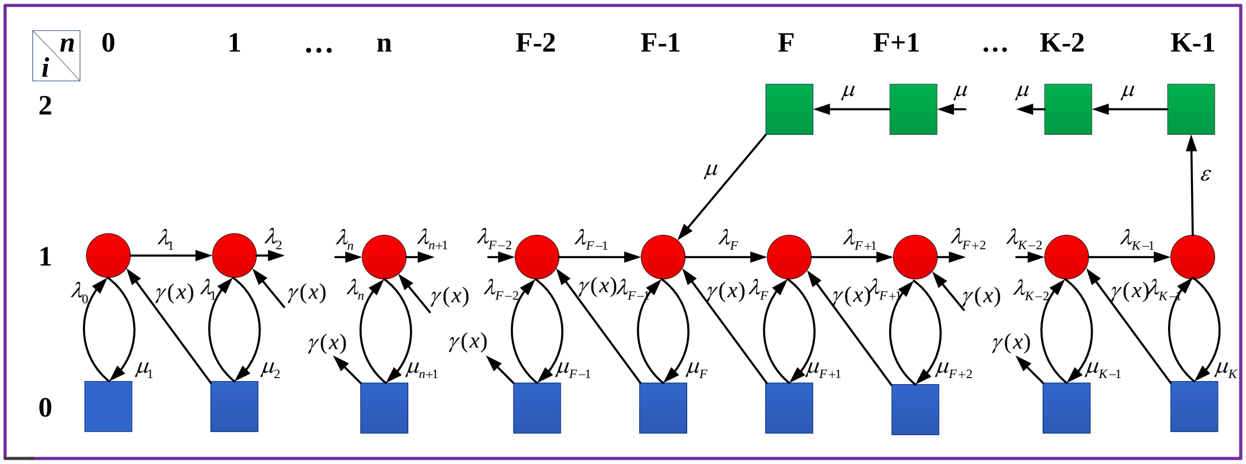


Figure 3.1: Transition state diagram

- The jobs join the system according to Poisson fashion with parameter λ.
- If the arriving job finds the server free, the job gets served according to an exponential distribution with rate μ .
- If the server is occupied then the arriving job joins the retrial pool. From the orbit, the job re-attempts for the service with general distributed retrial time having probability distribution $G(x)(x \geq 0)$ with $G(0)=0$, the probability density function $\gamma(x)$, and mean retrial time $1/\gamma$.
- When the system attains its full capacity, then setup time is required to stop the arriving jobs from joining the queue; the time required for the setup is assumed to exponentially distributed with mean $1/\varepsilon$.
- Once the system becomes full, the arrivals are restricted from joining the system. The further admission of jobs in the system is permitted when the number of jobs in the system ceases to a prefixed threshold value $'F'$ $(0 \leq F < K-1)$.

For developing the state-dependent model for the retrial queueing system, we denote the state-dependent arrival and service rates by λ_n and μ_n , respectively. The supplementary variable (U) is used corresponding to the remaining retrial time of the jobs while residing in the retrial pool. At the time τ , $N(\tau)$ denotes the number of jobs present in the system. The status of the server at a time τ is denoted by $S(\tau)$. To formulate the mathematical model, the random variable $S(\tau)$ is defined as follows:

$$S(\tau)=\begin{cases}0, & \text{the jobs are compelled to join the retrial pool on finiding the server being busy,}\\ 1(2), & \text{the server is occupied and the jobs are permitted (not permitted) to enter in the system.}\end{cases}$$

The system state probabilities at time epoch τ are as follows:

$$\left.\begin{array}{l} P_{0,0}(\tau)=\text{Prob}\{S(\tau)=0, N(\tau)=0\} \\ P_{0,n}(x,\tau)dx=\text{Prob}\{S(\tau)=0, N(\tau)=n, x<U(\tau)\le x+dx\}, x\ge 0, 1\le n\le K-1 \\ P_{1,n}(\tau)=\text{Prob}\{S(\tau)=1, N(\tau)=n\},\ 0\le n\le K-1 \\ P_{2,n}(\tau)=\text{Prob}\{S(\tau)=2, N(\tau)=n\},\ F\le n\le K-1 \end{array}\right\} \tag{3.1}$$

We denote $P_{0,n}(\tau)=\text{Prob}\{S(\tau)=0, N(\tau)=n\}=\int_0^\infty P_{0,n}(x,\tau)dx,\ 1\le n\le K-1$ (3.2)

At steady state, i.e., when $\tau\to\infty$, we define

$P_{0,0}=\lim_{\tau\to\infty} P_{0,0}(\tau)$,

$P_{0,n}(x)=\lim_{\tau\to\infty} P_{0,n}(x,\tau),\ 1\le n\le K-1$,

$P_{1,n}=\lim_{\tau\to\infty} P_{1,n}(\tau),\ 0\le n\le K-1$,

$P_{2,n}=\lim_{\tau\to\infty} P_{2,n}(\tau),\ F\le n\le K-1$.

3.3. GOVERNING EQUATIONS AND QUEUE SIZE DISTRIBUTION

To establish the steady state probabilities of the system state space, the governing equations for three levels (i.e., when $S(\tau)=i=0,1,2$) of the non-Markovian model are constructed by introducing the supplementary variable corresponding to remaining of retrial time. The in-flows and out-flows of system states (n,i) are depicted in the transition diagram shown in Figure 3.1.

Now by using the probability arguments, Chapman-Kolmogorov equations are formulated as follows:

(i) For $i=0$:

$$\lambda_0 P_{0,0}=\mu_1 P_{1,0} \tag{3.3}$$

$$-\frac{d}{dx}P_{0,n}(x)=-\lambda_n P_{0,n}(x)+\mu_{n+1}P_{1,n}\gamma(x),\ 1\le n\le K-1. \tag{3.4}$$

(ii) For $i=1$:

$$(\lambda_1+\mu_1)P_{1,0}=\lambda_0 P_{0,0}+P_{0,1}(0) \tag{3.5}$$

$$(\lambda_{n+1}+\mu_{n+1})P_{1,n}=\lambda_n P_{1,n-1}+\lambda_n P_{0,n}+P_{0,n+1}(0),\ 1\le n\le F-2 \tag{3.6}$$

$$(\lambda_F+\mu_F)P_{1,F-1}=\lambda_{F-1}P_{1,F-2}+\lambda_{F-1}P_{0,F-1}+P_{0,F}(0)+\mu P_{2,F} \tag{3.7}$$

$$(\lambda_{n+1}+\mu_{n+1})P_{1,n}=\lambda_n P_{1,n-1}+\lambda_n P_{0,n}+P_{0,n+1}(0),\ F\le n\le K-2 \tag{3.8}$$

$$(\varepsilon+\mu_K)P_{1,K-1}=\lambda_{K-1}P_{1,K-2}+\lambda_{K-1}P_{0,K-1} \tag{3.9}$$

(iii) For $i=2$:

$$\mu P_{2,n+1}=\mu P_{2,n+2}=\varepsilon P_{1,K-1},\ F\le n\le K-3 \tag{3.10}$$

Define Laplace-Stieltjes transform (LST) of $\gamma(x)$ and $P_{0,n}(x)$ by $\gamma^*(\theta)$ and $P^*_{0,n}(\theta)$, respectively.

Also, $P^*_{0,n}(\theta)=P_{0,n}\gamma^*(\theta)$ (3.11)

The symbols S_n and R_F are used for the brevity of notations and are defined as follows:

$$S_n=\varepsilon\left(\prod_{i=n+2}^{K-1}\lambda_i\right)+\varepsilon\sum_{i=n+3}^{K}\left(\left(\prod_{l=i}^{K-1}\lambda_l\right)\left(\prod_{j=n+2}^{i-1}\mu_j\gamma^*(\lambda_{j-1})\right)\right)+\left(\prod_{j=n+2}^{K}\mu_j\gamma^*(\lambda_{j-1})\right)$$

$$R_F=\mu_1\left(\prod_{j=1}^{F-1}\mu_{j+1}\gamma^*(\lambda_j)\right)\left(\varepsilon\left(\prod_{i=F+1}^{K-1}\lambda_i\right)+\mu_{F+1}\gamma^*(\lambda_F)S_F\right).$$

Theorem 3.1: The steady state queue size distribution for the state dependent retrial model operating under F-policy is given by

$$P_{0,n}=\begin{cases}\dfrac{\lambda_0}{\mu_1}\left(\dfrac{1}{\gamma^*(\lambda_1)}-1\right)P_{0,0}, & n=1,\\ \left(\prod_{i=0}^{n-1}\dfrac{\lambda_i}{\mu_{i+1}}\right)\dfrac{1}{\left(\prod_{j=1}^{n-1}\gamma^*(\lambda_j)\right)}\left(\dfrac{1}{\gamma^*(\lambda_n)}-1\right)P_{0,0}, & 1<n\le F-1,\\ \left(\prod_{i=0}^{n}\lambda_i\right)\dfrac{\mu_{n+1}(1-\gamma^*(\lambda_n))}{\lambda_n}\dfrac{S_n}{R_F}P_{0,0}, & F\le n\le K-2,\\ \left(\prod_{i=0}^{n}\lambda_i\right)\dfrac{\mu_{n+1}(1-\gamma^*(\lambda_n))}{\lambda_n}\dfrac{1}{R_F}P_{0,0}, & n=K-1.\end{cases} \tag{3.12}$$

$$P_{1,n}=\begin{cases}\dfrac{\lambda_0}{\mu_1}P_{0,0}, & n=0,\\ \left(\prod_{i=0}^{n}\dfrac{\lambda_i}{\mu_{i+1}}\right)\dfrac{1}{\left(\prod_{j=1}^{n}\gamma^*(\lambda_j)\right)}P_{0,0}, & 1\le n\le F-1,\\ \left(\prod_{i=0}^{n}\lambda_i\right)\dfrac{S_n}{R_F}P_{0,0}, & F\le n\le K-2,\\ \left(\prod_{i=0}^{n}\lambda_i\right)\dfrac{1}{R_F}P_{0,0}, & n=K-1.\end{cases} \tag{3.13}$$

$$P_{2,n}=\left(\prod_{i=0}^{K-1}\lambda_i\right)\frac{\varepsilon}{\mu R_F}P_{0,0},\quad F\le n\le K-1. \tag{3.14}$$

Proof: Taking LST on both sides of (3.4), we obtain

$$(\lambda_n-\theta)P_{0,n}^*(\theta)=\mu_{n+1}P_{1,n}\gamma^*(\theta)-P_{0,n}(0),\ 1\le n\le K-1 \tag{3.15}$$

Using (3.15), we have

$$P_{1,n}=\frac{1}{\mu_{n+1}\gamma^*(\lambda_n)}P_{0,n}(0)\ \text{and}\ P_{1,n}=\frac{\lambda_n}{\mu_{n+1}(1-\gamma^*(\lambda_n))}P_{0,n} \tag{3.16a-b}$$

Using (3.3),

$$P_{1,0}=\frac{\lambda_0}{\mu_1}P_{0,0} \tag{3.17}$$

Now (3.17) and (3.5) yield

$$P_{0,1}(0)=\frac{\lambda_0\lambda_1}{\mu_1}P_{0,0} \tag{3.18}$$

Using (3.16a) and (3.18), we obtain

$$P_{1,1}=\frac{\lambda_0\lambda_1}{\mu_1\mu_2\gamma^*(\lambda_1)}P_{0,0} \tag{3.19}$$

Further, using (3.16b) and (3.19), we get

$$P_{0,1}=\frac{\lambda_0}{\mu_1}\left(\frac{1}{\gamma^*(\lambda_1)}-1\right)P_{0,0} \tag{3.20}$$

Setting $n=1$ in (3.6) and using (3.17), (3.19) and (3.20), the probability $P_{0,2}(0)$ is obtained as

$$P_{0,2}(0)=\frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\gamma^*(\lambda_1)}P_{0,0} \tag{3.21}$$

From (3.21), (3.16a) and (3.16b), we obtain

$$P_{1,2}=\frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3\gamma^*(\lambda_1)\gamma^*(\lambda_2)}P_{0,0} \text{ and } P_{0,2}=\frac{\lambda_0\lambda_1}{\mu_1\mu_2\gamma^*(\lambda_1)}\left(\frac{1}{\gamma^*(\lambda_2)}-1\right)P_{0,0} \qquad \text{(3.22-3.23)}$$

In general, we obtain

$$P_{1,n}=\left(\prod_{i=0}^{n}\frac{\lambda_i}{\mu_{i+1}}\right)\frac{1}{\left(\prod_{j=1}^{n}\gamma^*(\lambda_j)\right)}P_{0,0},\ 1\le n\le F-1 \tag{3.24}$$

$$P_{0,n}=\left(\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}\right)\frac{1}{\left(\prod_{j=1}^{n-1}\gamma^*(\lambda_j)\right)}\left(\frac{1}{\gamma^*(\lambda_n)}-1\right)P_{0,0},\ 1<n\le F-1 \tag{3.25}$$

From (3.7)-(3.9) and using (3.24) and (3.25), we obtain

$$P_{1,n}=\left(\prod_{i=0}^{n}\lambda_i\right)\frac{S_n}{R_F}P_{0,0},\quad F\le n\le K-2 \tag{3.26}$$

$$P_{0,n}=\left(\prod_{i=0}^{n}\lambda_i\right)\frac{\mu_{n+1}(1-\gamma^*(\lambda_n))}{\lambda_n}\frac{S_n}{R_F}P_{0,0},\quad F\le n\le K-2 \tag{3.27}$$

$$P_{1,K-1}=\left(\prod_{i=0}^{K-1}\lambda_i\right)\frac{1}{R_F}P_{0,0} \tag{3.28}$$

$$P_{0,K-1}=\left(\prod_{i=0}^{K-1}\lambda_i\right)\frac{\mu_K(1-\gamma^*(\lambda_{K-1}))}{\lambda_{K-1}}\frac{1}{R_F}P_{0,0} \tag{3.29}$$

Now, using (3.28) in (3.10), result given in (3.14) is obtained. Also, $P_{0,0}$ can be determined using normalizing condition given by

$$\sum_{n=0}^{K-1}P_{0,n}+\sum_{n=0}^{K-1}P_{1,n}+\sum_{n=F}^{K-1}P_{2,n}=1 \tag{3.30}$$

Remark. It should be noted that when $F = K-1$, we have M/M/1/K model with state dependent rates and general retrial. In this case, F-policy is not taken into account as such $\varepsilon = 0,\ S_F = 1$.

3.4. SPECIAL MODELS

In this section, some special models are deduced by setting suitable parameter values for the state dependent rates. First of all, by setting state dependent arrival rates, we consider a finite population model for machine repair problem (MRP) and its particular case when the control of arrivals is not taken into consideration. Then after, by setting the state dependent service rate time sharing model is discussed.

3.4.1. Finite Population Model (FPM)

In the present scenario of modern lifestyle, machines are needed to perform day-to-day as well as specific jobs. It is noticed that the unexpected failures of machines have an adverse impact on the system efficiency/availability and also increases the production cost and downtime of the system. In this sub-section, the machine repair model which is a finite population model is presented as follows:

By setting $\lambda_n = (K-n)\lambda$ and $\mu_n = \mu$, equations (3.12), (3.13) and (3.14) yield the queue size distribution. Thus, for the finite population model for MRP, we get

$$P_{0,n} = \begin{cases} \dfrac{K\lambda}{\mu}\left(\dfrac{1}{\left(\gamma^*(K-1)\lambda\right)} - 1\right)P_{0,0}, & n = 1, \\ \dfrac{K\lambda}{\mu^n}\left(\displaystyle\prod_{i=1}^{n-1}\dfrac{((K-i)\lambda)}{\gamma^*((K-i)\lambda))}\right)\left(\dfrac{1}{\gamma^*((K-n)\lambda)} - 1\right)P_{0,0}, & 1 < n \le F-1, \\ \left(\displaystyle\prod_{i=0}^{n}((K-i)\lambda)\right)\dfrac{\mu(1-\gamma^*((K-n)\lambda))}{(K-n)\lambda}\dfrac{S_n}{R_F}P_{0,0}, & F \le n \le K-2, \\ \left(\displaystyle\prod_{i=0}^{n}((K-i)\lambda)\right)\dfrac{\mu(1-\gamma^*((K-n)\lambda))}{(K-n)\lambda}\dfrac{1}{R_F}P_{0,0}, & n = K-1. \end{cases} \tag{3.31}$$

$$P_{1,n} = \begin{cases} \dfrac{K\lambda}{\mu} P_{0,0}, & n=0, \\ \dfrac{K\lambda}{\mu^{n+1}} \left(\prod_{i=1}^{n} \dfrac{(K-i)\lambda}{\gamma^*((K-i)\lambda)} \right) P_{0,0}, & 1 \le n \le F-1, \\ \left(\prod_{i=0}^{n} (K-i)\lambda \right) \dfrac{S_n}{R_F} P_{0,0}, & F \le n \le K-2, \\ \left(\prod_{i=0}^{n} (K-i)\lambda \right) \dfrac{1}{R_F} P_{0,0}, & n = K-1, \end{cases} \tag{3.32}$$

$$P_{2,n} = \left(\prod_{i=0}^{K-1} \big((K-i)\lambda\big) \right) \frac{\varepsilon}{\mu R_F} P_{0,0}, \quad F \le n \le K-1 \tag{3.33}$$

In particular, when F-policy is not taken into account, so that $\mu_n = \mu,\ \lambda_n = (K-n)\lambda$, Equations (3.12) and (3.13) yield

$$P_{0,n} = \begin{cases} \dfrac{K\lambda}{\mu} \left(\dfrac{1}{\gamma^*((K-1)\lambda)} - 1 \right) P_{0,0}, & n=1, \\ \dfrac{K\lambda^n}{\mu^n} \left(\prod_{i=1}^{n-1} \dfrac{(K-i)}{\gamma^*((K-i)\lambda)} \right) \left(\dfrac{1}{\gamma^*((K-1)\lambda)} - 1 \right) P_{0,0}, & 2 \le n \le K-1. \end{cases} \tag{3.34}$$

$$P_{1,n} = \begin{cases} \dfrac{K\lambda}{\mu} P_{0,0}, & n=0, \\ \dfrac{K\lambda^{n+1}}{\mu^{n+1}} \left(\prod_{i=1}^{n} \dfrac{(K-i)}{\gamma^*((K-i)\lambda)} \right) P_{0,0}, & 1 \le n \le K-1. \end{cases} \tag{3.35}$$

Equations (3.34) and (3.35) provide the same results as obtained by Yang and Chang (2018).

3.4.2. Time-Sharing Model (TSM)

The time-sharing system involves the sharing of source among many tasks by means of parallel operations or allocating a very small quantum of time to each task in round robin fashion. In case of single server time sharing queueing system, the arriving jobs may wait in the queue for the service for a small pre-specified time duration; if they do not get served within this duration, these jobs have to join the end of the queue. When the jobs again join the server for service, the same rule of time-sharing is again applied until they get served. If some jobs are already present for the service, the arriving job has to join the retrial orbit.

The state-dependent time sharing model is formulated by setting $\mu_n = \frac{\mu}{n}$ and $\lambda_n = \frac{\lambda}{n+1}$ in (3.12)-(3.14). The queue length distribution for the time-sharing system with state-dependent rates is obtained as

$$P_{0,n} = \begin{cases} \frac{\lambda}{\mu}\left(\frac{1}{\gamma^*\left(\lambda/2\right)} - 1\right) P_{0,0}, & n = 1, \\ \left(\frac{\lambda}{\mu}\right)^n \frac{1}{\left(\prod_{j=1}^{n-1} \gamma^*\left(\frac{\lambda}{j+1}\right)\right)} \left(\frac{1}{\gamma^*\left(\frac{\lambda}{n+1}\right)} - 1\right) P_{0,0}, & 1 < n \le F-1, \\ \frac{\lambda^{n+1}}{(n+1)!} \frac{\mu\left(1 - \gamma^*\left(\frac{\lambda}{n+1}\right)\right)}{\lambda} \frac{S_n}{R_F} P_{0,0}, & F \le n \le K-2, \\ \frac{\lambda^{n+1}}{(n+1)!} \frac{\mu\left(1 - \gamma^*\left(\frac{\lambda}{n+1}\right)\right)}{\lambda} \frac{1}{R_F} P_{0,0}, & n = K-1. \end{cases} \tag{3.36}$$

$$P_{1,n} = \begin{cases} \frac{\lambda}{\mu} P_{0,0}, & n = 0, \\ \left(\frac{\lambda}{\mu}\right)^{n+1} \frac{1}{\left(\prod_{j=1}^{n} \gamma^*\left(\frac{\lambda}{j+1}\right)\right)} P_{0,0}, & 1 \le n \le F-1, \\ \frac{\lambda^{n+1}}{(n+1)!} \frac{S_n}{R_F} P_{0,0}, & F \le n \le K-2, \\ \frac{\lambda^{n+1}}{(n+1)!} \frac{1}{R_F} P_{0,0}, & n = K-1. \end{cases} \tag{3.37}$$

$$P_{2,n} = \frac{\lambda^K}{K!} \frac{\varepsilon}{\mu R_F} P_{0,0}, \quad F \le n \le K-1. \tag{3.38}$$

In particular case when the arrival rate is constant, by setting $\lambda_n = \lambda$ in (3.36)-(3.38), we get the results for *F*-policy general retrial model with constant arrival rate.

Special Case: When retrial time follows exponential distribution in TSM

Laplace-Stieltjes transform of $\gamma(x)$ for exponential distribution (Exp) is taken as $\gamma^*(\theta)=\dfrac{\gamma}{\theta+\gamma}$.

Equations (3.36)-(3.38) give the result for TSM when retrial time is taken as exponential as

$$P_{0,n}=\begin{cases}(\delta_n-1)\left(\prod_{i=0}^{n-1}\rho_i\right)\left(\prod_{j=1}^{n-1}\delta_j\right)P_{0,0},\ 1\le n\le F-1\\ \dfrac{\Lambda_{0,n}(1-1/\delta_n)}{\rho_n}\dfrac{S_n}{R_F}P_{0,0}, \qquad F\le n\le K-2\\ \dfrac{\Lambda_{0,K-1}(1-1/\delta_{K-1})}{\rho_{K-1}}\dfrac{1}{R_F}P_{0,0}, \qquad n=K-1,\end{cases} \tag{3.39}$$

$$P_{1,n}=\begin{cases}\left(\prod_{i=0}^{n}\rho_i\right)\left(\prod_{j=1}^{n}\delta_j\right)P_{0,0},\ 1\le n\le F-1,\\ \dfrac{\Lambda_{0,n}S_n}{R_F}P_{0,0}, \qquad F\le n\le K-2,\\ \dfrac{\Lambda_{0,K-1}}{R_F}P_{0,0}, \qquad n=K-1,\end{cases} \tag{3.40}$$

$$P_{2,n}=\frac{\Lambda_{0,K-1}\varepsilon}{\mu R_F}P_{0,0},\ F\le n\le K-1 \tag{3.41}$$

where $\delta_i=\dfrac{\lambda_i+\gamma}{\gamma}$, $\rho_i=\dfrac{\lambda_i}{\mu_{i+1}}$ and $\Lambda_{0,n}=\prod_{j=0}^{n}\lambda_j$.

3.5. PERFORMANCE PREDICTION

To analyze the queueing characteristics of the concerned retrial service system and to make the model applicable to the real-time situation, it is beneficial to establish various system indices and cost analysis.

3.5.1. Performance Indices

The queueing model developed in the previous section for a single server finite model with general retrial attempts under admission control according to *F*-policy is analyzed by deriving some system indices as follows:

(i) The average number of jobs in the system and in the queue respectively, are

$$E[N_S]=\sum_{n=0}^{K-1}nP_{0,n}+\sum_{n=0}^{K-1}(n+1)P_{1,n}+\sum_{n=F}^{K-1}(n+1)P_{2,n} \tag{3.42a}$$

and $E[N_q]=\sum_{n=0}^{K-1}nP_{1,n}+\sum_{n=F}^{K-1}nP_{2,n}$ (3.42b)

(ii) The average number of jobs in the orbit is

$$E[N_R]=\sum_{n=0}^{K-1}nP_{0,n} \tag{3.43}$$

(iii) The system throughput is

$$TP=\sum_{n=0}^{K-1}\mu_{n+1}P_{1,n}+\mu\sum_{n=F}^{K-1}P_{2,n} \tag{3.44}$$

(iv) The status of the server can be represented by the probability of the server being free (P_I) and being engaged (P_{SB}) in rendering service, respectively. Thus we obtain

$$P_I=\sum_{n=0}^{K-1}P_{0,n} \text{ and } P_{SB}=\sum_{n=0}^{K-1}P_{1,n}+\sum_{n=F}^{K-1}P_{2,n} \tag{3.45a-b}$$

3.5.2. Cost Function

For the retrial queueing system operating under F-policy, the organizers may be interested in the optimal service rate that optimizes the total system cost. To formulate the cost function, the cost components associated with different activities are used. To evaluate the threshold parameter (F) and service rate that optimize the cost function $TC(F,\mu)$, we formulate the total cost per unit time for operating the system as follows:

$$TC(F,\mu)=C_IP_I+C_BP_{SB}+C_HE[N_q]+\mu C_F+C_OE[N_R] \tag{3.46}$$

where

C_I : Cost associated with the server per unit time during idle state,
C_B : Cost of the server per unit time when he is busy in rendering the service,
C_H : Holding cost of each job residing in the system,
C_F : Cost involved per unit time in rendering the service to the job,
C_O : Cost spent on each customer while residing in the orbit.

The cost function given in (3.46) is highly non-linear and complex, therefore, its explicit analytical solution may not possible, however numerical methods can be easily employed. To get optimal decision parameters (F,μ), (i) direct search method is used to determine optimal threshold parameter (F^*) which is discrete decision parameter and then (ii) quasi-Newton method is used to evaluate optimal service rate (μ^*) which is continuous decision parameter.

Direct search method to evaluate (F^*): To control the admission of jobs in the system, we find the optimal threshold parameter (F^*) so as to optimize the system cost given in (3.43). It is noticed that the threshold parameter $'F'$ $(0 \le F < K-1)$ has integer values. Thus, direct search method based on a heuristic approach is used by successively substituting $F = 0,1,2,...,$ $K-2$ to compute cost function given in (3.46). The optimal threshold parameter (F^*) is evaluated by using the inequalities. $TC(F^*-1,\mu) \ge TC(F^*,\mu)$ and $TC(F^*+1,\mu) \ge TC(F^*,\mu)$.

Quasi-Newton method to evaluate (μ^*): After determining the optimal threshold parameter (F^*), we use the quasi-Newton method to determine optimal service rate (μ^*) by minimizing the cost function $TC(F^*,\mu)$. The algorithmic steps of the quasi-Newton method are given in Section 1.4.8 of Chapter 1. We set the following input parameters for execution of quasi-Newton method.

Inputs: $K,\ F^*,\ \lambda,\ \mu_0,\ \gamma,\ \varepsilon,\ C_I,\ C_B,\ C_H,\ C_F,\ C_O$, and tolerance ε_0 of $\left|\dfrac{\partial TC(F^*,\mu)}{\partial \mu}\right|$.

Output: Approximate optimal solution of service rate (μ) and total cost per unit time (TC) as

$\mu^*,\ TC(F^*,\mu^*).$

3.6. ILLUSTRATION AND NUMERICAL RESULTS

F-policy state dependent retrial queueing model developed has applications in several real time congestion problems including MRP and time-sharing system. To illustrate MRP, we consider a computer repair shop in which finite number (say K) of computers can be repaired under a maintenance contract. The failed computers arrive for the repair to the shop by following Poisson process with the rate λ. The repair job of a failed computer is done by the repairman following the exponential distribution with mean $1/\mu$. If the caretaker of the failed computer finds the repairman busy, the failed computers are put in the orbit; from the orbit, it can be sent again for the repair job; the retrial time is assumed to be general distributed with mean $1/\gamma$. Due to the limited space of the shop, the arriving failed computers are not permitted to enter the shop as soon as the capacity of the shop becomes full. The failed computers are further permitted to join the shop only when the workload of failed computers reduces to a predefined level $'F'$ in terms of a number of failed computers. Before allowing the failed computers to enter the shop for repair, the setup time is required which is also assumed to be exponentially distributed with the rate ε.

The applicability of *time sharing queueing model* is quite prevalent in the multiplexed information and computing service system, operating systems, computer and communication system, cloud computing centers, etc. To be specific, we cite the illustration of a call center with a single server to serve the queries of arriving calls. The arriving calls contact the agent *i.e.* server of the call center to receive the service. If the agent is free at that time, then the call gets service immediately. When the agent is busy, then the arriving call has to wait in the retrial orbit. After some random time, the call requests for the service to the agent again. In the case when the number of calls accumulated in the call center reaches to the system capacity (i.e., K) of the call center, the arriving calls are not allowed to join until the number of calls in the system reduces to pre-fixed level (i.e., F).

The numerical simulation and cost optimization have been carried out for both machine repair problem (MRP) and the time-sharing model (TSM). The numerical experiment performed may be helpful to examine the effects of parameters on various performance measures and to determine the optimal threshold parameter and optimal service rate. The three distributions for the retrial time, have been considered. The Laplace- Stieltjes transform of $\gamma(x)$ for exponential (Exp), Erlang-3 (E_3), and deterministic (D) distribution are taken as $\gamma^*(\theta)=\dfrac{\gamma}{\theta+\gamma}$, $\gamma^*(\theta)=\left(\dfrac{3\gamma}{\theta+3\gamma}\right)^3$ and $\gamma^*(\theta)=e^{-\theta/\gamma}$, respectively.

The software 'MATLAB' is used to develop the computer program to compute the system indices and cost function. For the machine repair problem (MRP) and the time-sharing model (TSM), we set the state dependent rates as (i) $\lambda_n=\lambda(K-n)$, $\mu_n=\mu$ and (ii) $\lambda_n=\lambda/(n+1)$, $\mu_n=\mu/n$, respectively.

3.6.1. Numerical Results for the Machine Repair Model (MRP)

To validate the practical application of the computer repair shop model, we evaluate the performance indices numerically for exponential distributed retrial time by setting the default parameters as $K=7, F=4, \lambda=1$unit/hour, $\mu=2$ unit/hour, $\gamma=0.5$ unit/hour, $\varepsilon=0.5$unit/hour. For the computer repair example, the average number of failed computers in the shop is obtained as 5.65.

(i) Sensitivity analysis for MRP

To explore the sensitivity of the repair rate (μ), failure rate (λ) and retrial rate (γ) with respect to the indices $E[N_S]$ and TP, the graphs are plotted in Figures 3.2(a-c) and 3.3(a-c), respectively. Some other performance indices have also been summarized in Tables 3.1-3.3 for varying the values of these parameters. For the computation of system indices, the default parameters are fixed as $K=7,\ F=4,\ \lambda=0.5, \mu=8, \gamma=0.5, \varepsilon=1$.

Based on numerical experiments performed, we present the sensitivity of the parameters as follows:

Effect of μ **:** It is observed that as the repairman (server) repairs the failed machines with a faster rate, the average number of failed machines decreases which also demonstrates the validity of analytical results. Also the status of the repairman (idle or busy) completely depends on the number of failed machines. From Table 3.1, it is noticed that as repair rate (μ) increases, the average number of failed machines in the queue $(E[N_q])$ decreases. Also the probability of repairman being busy (idle) decreases (increases) by enhancing the μ. The graph plotted in Figure 3.2(a) depicts that $E[N_S]$ lowers down as the service rate goes up. From Figure 3.3(a), it is clear that the throughput (TP) grows up as the service rate (μ) speeds up which is the same as we expect.

Table 3.1: Various performance measures for varying values of μ for MRP

μ	$E[N_q]$			P_I			P_{SB}			TC		
	Exp	E_3	*D*	*Exp*	E_3	*D*	*Exp*	E_3	*D*	*Exp*	E_3	*D*
1	3.164	3.110	3.060	0.303	0.335	0.354	0.697	0.665	0.646	532.25	543.43	547.66
2	2.043	2.013	1.964	0.492	0.543	0.569	0.508	0.457	0.431	473.61	506.62	518.70
3	1.482	1.477	1.436	0.580	0.643	0.672	0.420	0.357	0.328	431.77	485.08	505.76
4	1.155	1.166	1.135	0.633	0.699	0.730	0.367	0.301	0.270	403.18	469.82	499.15
5	0.940	0.963	0.941	0.671	0.735	0.768	0.329	0.265	0.232	385.15	459.00	496.05

Table 3.2: Various performance measures for varying values of λ for MRP

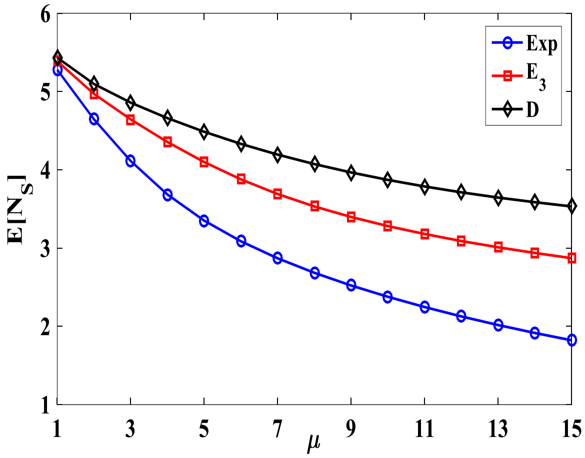
λ	$E[N_q]$			P_I			P_{SB}			TC		
	Exp	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D
1	1.132	1.060	0.998	0.704	0.771	0.793	0.296	0.229	0.207	562.12	599.89	618.34
2	1.742	1.604	1.542	0.644	0.687	0.702	0.356	0.313	0.298	677.66	662.77	667.59
3	2.161	2.026	1.985	0.582	0.617	0.626	0.418	0.383	0.374	711.49	688.00	689.76
4	2.494	2.372	2.347	0.530	0.559	0.564	0.470	0.441	0.436	729.36	704.04	704.74
5	2.767	2.661	2.645	0.486	0.510	0.514	0.514	0.490	0.486	741.01	715.85	716.15

Table 3.3: Various performance measures for varying values of γ for MRP

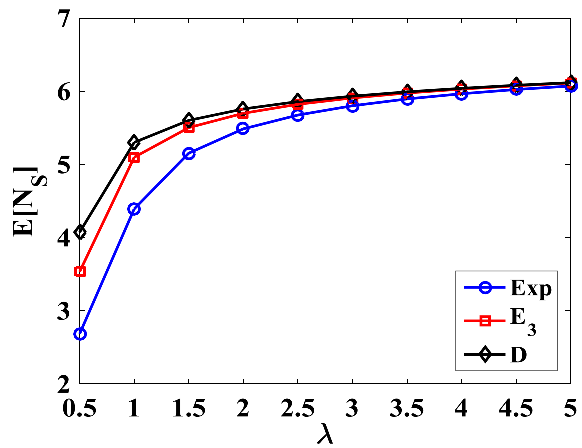
γ	$E[N_q]$			P_I			P_{SB}			TC		
	Exp	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D
0.5	0.581	0.633	0.628	0.745	0.799	0.830	0.255	0.201	0.170	384.56	448.74	500.10
0.6	0.562	0.627	0.637	0.733	0.784	0.817	0.267	0.216	0.183	365.13	423.37	477.68
0.7	0.544	0.617	0.643	0.723	0.771	0.805	0.277	0.229	0.195	348.73	403.30	459.70
0.8	0.527	0.606	0.645	0.714	0.760	0.795	0.286	0.240	0.205	334.57	386.37	444.12
0.9	0.510	0.593	0.644	0.706	0.750	0.784	0.294	0.250	0.216	322.17	371.45	429.80

Effect of λ **:** For the MRP, if the repairman provides repair job to the machines with constant rate and the failure rate (λ) of machines increases, then $E[N_q]$ seems to increase at a faster pace. Table 3.2 displays the numerical results which demonstrate that by keeping the service rate constant, if the rate of the failure of machines increases, then the average queue size $(E[N_q])$ enhances. The probability of the repairman being busy (idle) also seems to increase (decrease) by increasing the failure rate of the machines. From graphs shown in Figures 3.2(b) and 3.3(b), it is evident that $E[N_S]$ and TP grow up as λ increases.

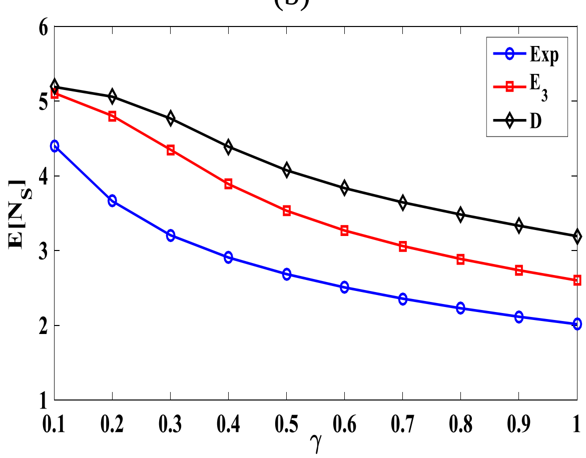
Effect of γ **:** Table 3 presents the negligible effect of the retrial rate on various performance indices. We observe that $E[N_q]$ decreases very slowly as the retrial rate increases. The probability of a repairman being busy (idle) remains almost constant as the retrial rate grows up. Figure 3.2(c) reveals the trends of $E[N_S]$ which decreases at a slow pace as the retrial rate (γ) increases. Figure 3.3(c) shows that the throughput (TP) of the system enhances with a very slow rate as γ enhances.



(a)

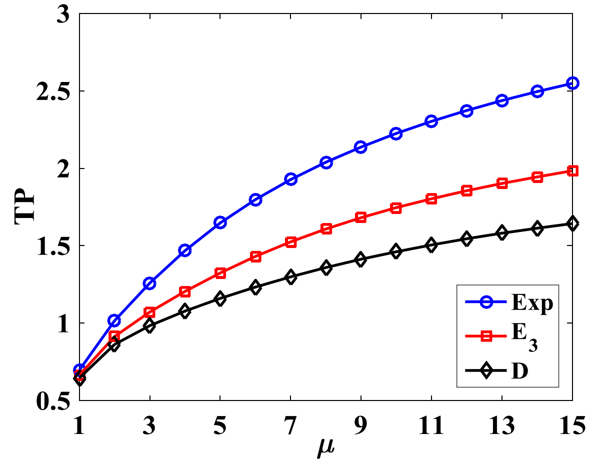


(b)

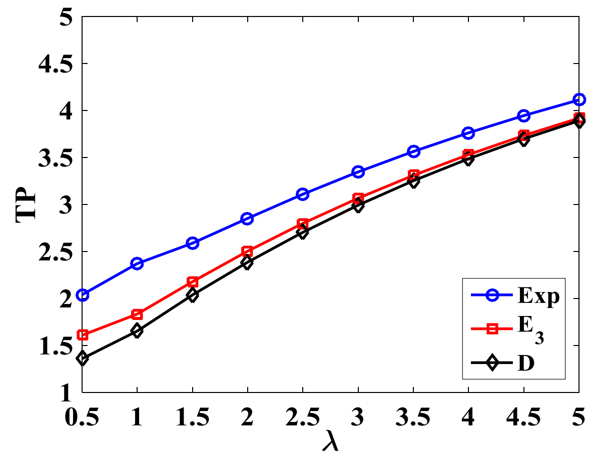


(c)

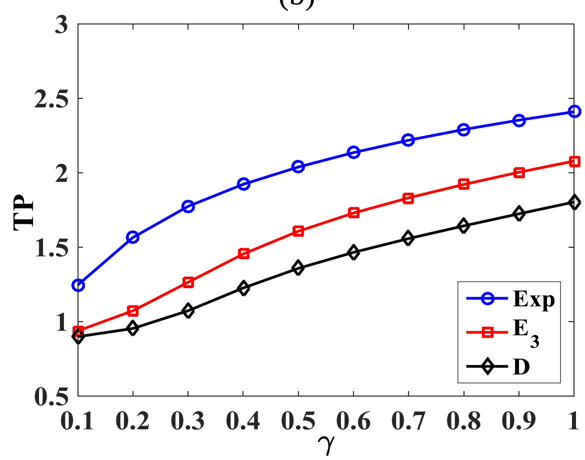
Figure 3.2: $E[N_S]$ vs (a) μ (b) λ (c) γ for different distributions for MRP



(a)



(b)



(c)

Figure 3.3: TP vs (a) μ (b) λ (c) γ for different distributions for MRP

(ii) Cost optimization for MRP

In machine repair problem, the system organizers may be interested to evaluate optimal threshold parameter $'F'$, optimal service rate (μ) and corresponding minimal cost of the system. From the practical point of view, the system capacity K can be treated as an upper bound to determine the threshold parameter $'F'$ which can be searched in the desired feasible region. For evaluating the total cost, the four cost sets given in Table 3.4, have been taken into account.

Table 3.4: Cost sets with different cost elements (in $) for MRP

Cost Set	C_I	C_B	C_H	C_F	C_O
I	30	30	50	70	40
II	10	10	120	15	90
III	15	5	120	15	90
IV	20	20	100	15	90

Optimal threshold parameter (F^*): In order to compute the optimal threshold parameter and minimum cost for the MRP, a direct search method, based on a heuristic approach is applied. To determine F^* and corresponding total cost $TC(F^*,\mu)$ for Cost Set - I, the default parameters are fixed as $K=25$, $\lambda=0.1$, $\mu=1.5$, $\gamma=0.5$, $\varepsilon=1$, and then the total cost is computed by varying F in the feasible region from 0 to 23. The total cost function for different distributions (exponential (Exp), Erlang-3 (E_3) and deterministic (D)) is computed and plotted in Figure 4 for varying values of F. From Figure 3.4, it is seen that the expected cost function is unimodal and convex in the feasible range (0, 23) with respect to the threshold parameter $'F'$. We indicate the minimum cost corresponding to the optimal value of the threshold parameter $'F'$ for Exp, E_3 and D distributions of retrial time. Table 3.5 displays the optimal threshold parameter F^* and corresponding optimal cost $TC(F^*,\mu)$ for different distributions for varying values of retrial rate as $\gamma=0.3$, 0.5 and 0.7.

Table 3.5: Searching the optimal F for MRP for different γ

γ	$(F^*, TC(F^*,\mu))$		
	Exp	E_3	D
0.3	(12, 852.01)	(14, 898.49)	(15, 937.18)
0.5	(10, 821.20)	(12, 848.52)	(13, 874.16)
0.7	(8, 806.92)	(10, 821.04)	(11, 836.66)

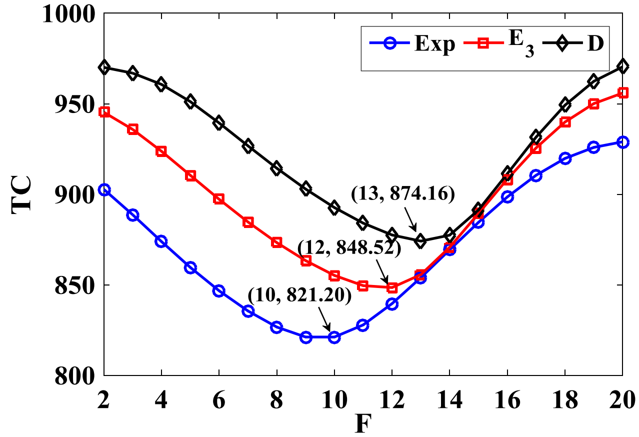
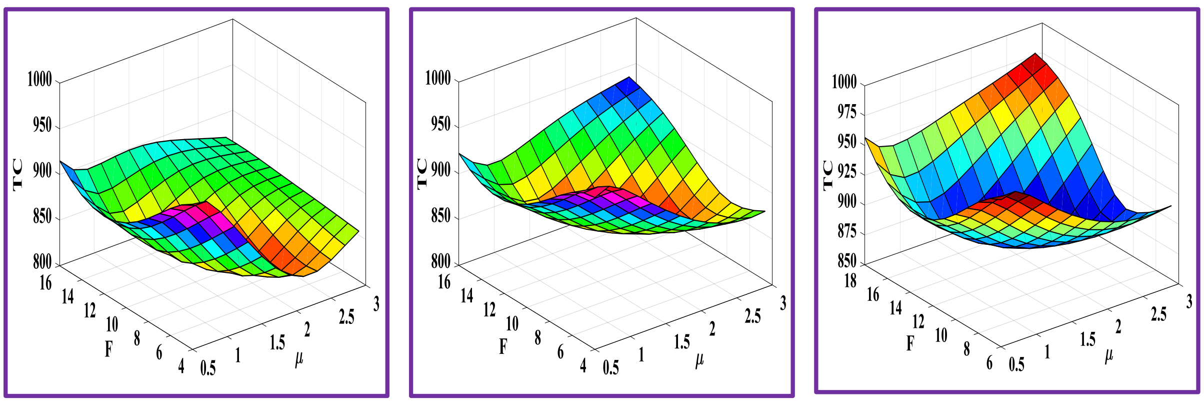


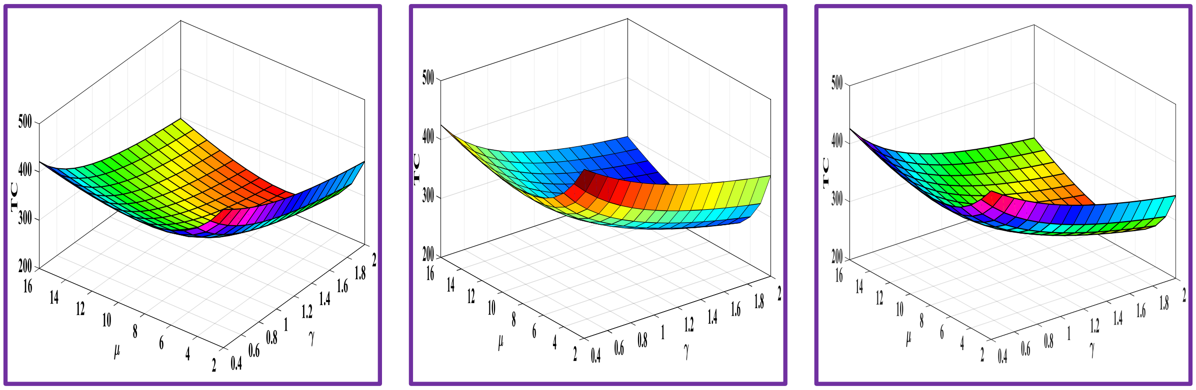
Figure 3.4: TC vs F for different distributions when $\gamma = 0.5$ for MRP



(a) (b) (c)

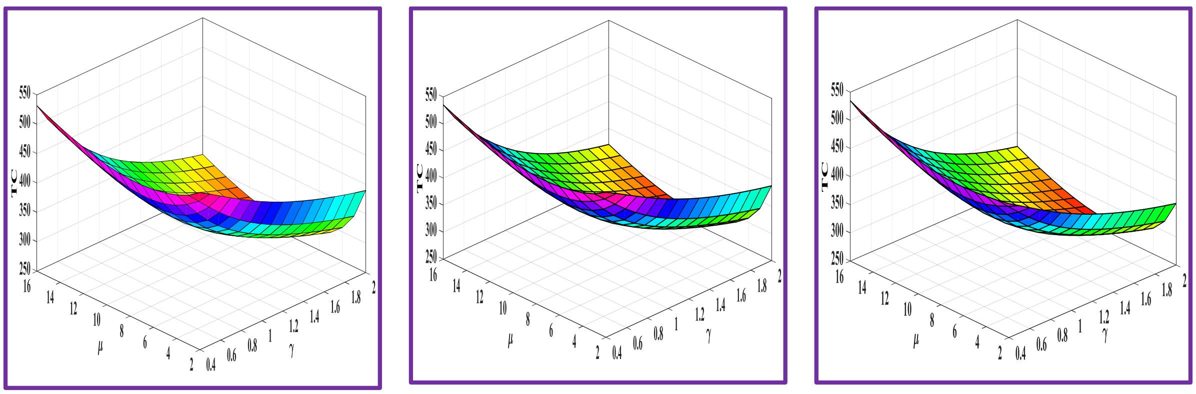
Figure 3.5: TC vs μ and F for (a) Exp (b) E_3 (c) D by taking Cost Sets- I for MRP

Optimal service rate (μ^*)**:** Since μ is a continuous decision parameter, we apply the quasi-Newton method to evaluate the optimal service rate (μ^*). To compute the optimal service rate (μ^*), we choose the threshold optimal parameter (F^*) which is already obtained using direct search method (see Table 3.5). Using algorithmic steps of the quasi-Newton method given in Section 1.4.8 of Chapter 1, for the exponential distribution, we see in Table 3.6 that after performing 7 iterations, the optimal service rate (μ^*) is attained; the corresponding minimum cost $TC(F^*, \mu^*)$ is \$821.20. The total minimum costs corresponding to the optimal threshold parameter (F^*) and optimal service rate (μ^*) for different distributions (Exp, E_3 and D) by considering $\gamma = 0.3$, 0.5 and 0.7 are recorded in Table 3.7.



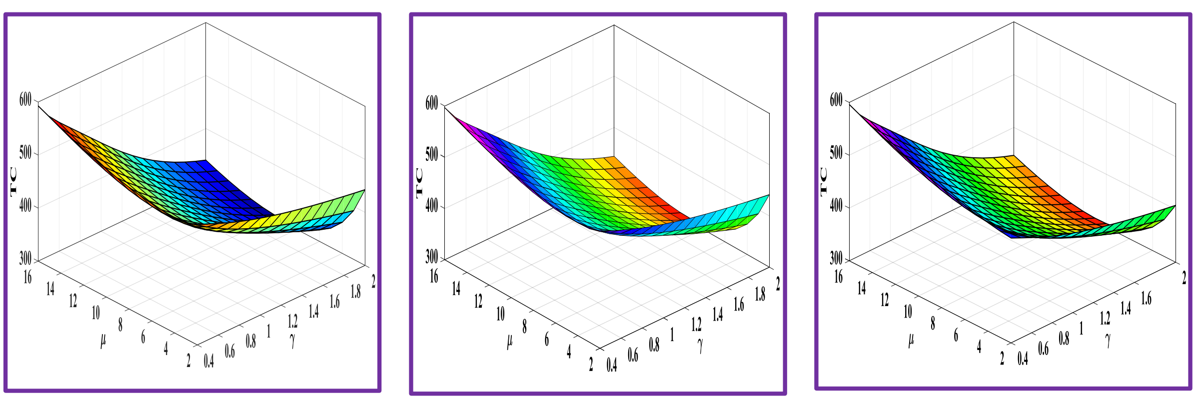
(a) (b) (c)

Figure 3.6: TC vs μ and γ for exponential distribution by taking (a) Cost Set-II (b) Cost Set-III (c) Cost Set-IV for MRP



(a) (b) (c)

Figure 3.7: TC vs μ and γ for Erlang-3 distribution by taking (a) Cost Set-II (b) Cost Set-III (c) Cost Set-IV for MRP



(a) (b) (c)

Figure 3.8: TC vs μ and γ for deterministic distribution by taking (a) Cost Set-II (b) Cost Set-III (c) Cost Set-IV for MRP

From the results given in Table 3.7, we notice that the expected minimum cost corresponding to optimal service rate in case of the exponential distribution is least as compared to other two distributions viz. Erlang-3 and deterministic for the retrial time. For Exp, E_3 and D retrial time distributions, the surface graphs of cost function TC by varying parameters $'F'$ and $'\mu'$ are depicted in Figures 3.5(a)-3.5(c) respectively; these graphs reveal the convexity of TC with respect to F and μ both.

By setting the parameters $K=7, F=4,\ \lambda=0.5, \varepsilon=1$ for the Cost Sets II, III, and IV, the surface graphs for the expected cost function $TC(F,\mu)$ are also shown in Figures 3.6-3.8 for the exponential (Exp), Erlang-3 (E_3) and deterministic (D) distributions, respectively. In these figures, the total cost TC is displayed with varying the values of μ from 2 to 16 and γ from 0.4 to 2. We can see that the cost functions are convex and unimodal in a feasible range of service rate.

Table 3.6: Searching the $\mu*$ by quasi-Newton method for exponential distribution $(F^*=10,\ \gamma=0.5)$

Iterations	F^*	μ	$TC(F^*,\mu)$	Max. tolerance
0	10	1	853.525	9.87E+01
1	10	2	842.70	5.69E+01
2	10	1.6344	823.783	3.5E+01
3	10	1.0491	821.311	8.28E+00
4	10	1.5255	821.207	2.27E+00
5	10	1.4917	821.199	6.35E-02
6	10	1.4990	821.199	4.12E-04
7	10	1.4988	821.199	1.02E-05

Table 3.7: Minimum cost $(F^*,\mu^*,TC(F^*,\mu^*))$ for $\gamma=0.3$, 0.5 and 0.7 for MRP

γ	$(F^*,\mu^*,TC(F^*,\mu^*))$		
	Exp	E_3	D
0.3	(12, 1.601, 851.05)	(14,1.668, 897.73)	(15, 1.323, 936.78)
0.5	(10, 1.499, 821.20)	(12, 1.579, 847.89)	(13, 1.691, 872.05)
0.7	(8, 1.489, 806.90)	(10, 1.543, 820.77)	(11, 1.627, 834.89)

3.6.2. Numerical Results for the Time-Sharing Model (TSM)

The analytical derivation of the system indices for the time-sharing model is done in Section 3.5.1. However, to understand the system behavior, the sensitivity of the parameters for different indices is required.

(i) Sensitivity analysis for TSM

For the TSM, the pictorial representations of $E[N_S]$ and TP are shown in Figures 3.9(a-c) and 3.10(a-c), respectively. The numerical results for the probability of the server being busy or idle and $E[N_q]$ are summarized in Tables 3.8-3.10 for varying values of μ, λ and γ, respectively. For the computation of system indices, the default parameters are fixed as $K=7, F=4, \lambda=2.8, \mu=10, \gamma=1, \varepsilon=1$.

Table 3.8: Various performance measures for varying values of μ for TSM

μ	$E[N_q]$			P_I			P_{SB}			TC		
	Exp	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D
6	0.876	0.995	1.041	0.552	0.557	0.564	0.448	0.443	0.436	274.65	307.11	323.69
8	0.476	0.570	0.623	0.654	0.656	0.659	0.346	0.344	0.341	241.50	270.93	290.14
10	0.286	0.352	0.399	0.721	0.722	0.723	0.279	0.278	0.277	237.26	261.73	280.32
12	0.187	0.235	0.272	0.767	0.767	0.768	0.233	0.233	0.232	247.50	267.75	284.72
14	0.131	0.166	0.196	0.800	0.800	0.801	0.200	0.200	0.199	265.16	282.20	297.46

Table 3.9: Various performance measures for varying values of λ for TSM

λ	$E[N_q]$			P_I			P_{SB}			TC		
	Exp	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D
2	0.107	0.124	0.138	0.800	0.800	0.800	0.200	0.200	0.200	193.36	201.93	208.81
4	0.783	0.948	1.014	0.613	0.619	0.628	0.387	0.381	0.372	341.33	391.91	418.07
6	1.726	1.843	1.829	0.493	0.516	0.535	0.507	0.484	0.465	504.56	547.33	559.06
8	2.340	2.357	2.303	0.426	0.456	0.478	0.574	0.544	0.522	589.23	614.24	619.62
10	2.737	2.699	2.632	0.380	0.412	0.434	0.620	0.588	0.566	634.09	649.61	652.77

Table 3.10: Various performance measures for varying values of γ for TSM

γ	$E[N_q]$			P_I			P_{SB}			TC		
	Exp	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D	*Exp*	E_3	D
1	0.286	0.352	0.399	0.721	0.722	0.723	0.279	0.278	0.277	237.26	261.73	280.32
1.5	0.225	0.256	0.276	0.719	0.721	0.722	0.281	0.279	0.278	214.93	226.13	234.03
2	0.195	0.212	0.223	0.718	0.720	0.721	0.282	0.280	0.279	204.02	210.31	214.43
2.5	0.177	0.188	0.194	0.717	0.719	0.720	0.283	0.281	0.280	197.58	201.59	204.06
3	0.165	0.173	0.177	0.716	0.718	0.719	0.284	0.282	0.281	193.35	196.12	197.75

Effect of μ **:** Numerical results for $E[N_q], P_I, P_{SB}$ and cost function (TC) for varying values of μ are summarized in Table 3.8. It is seen that as the service rate increases, $E[N_q]$ and the probability of the server being busy (idle) decreases (increases). From Figure 3.9(a), it is clear that $E[N_S]$ decreases as μ grows up. Figure 3.10(a) reveals that TP goes up as μ speeds up.

Effect of λ **:** Table 3.9 displays the numerical results of various performance indices for varying values of λ. It is noted that as λ increases, $E[N_q]$ also increases. The probability of the server being idle (busy) seems to decrease (increase) by increasing the value of λ and keeping the μ as constant. The trends shown in Figures 3.9(b) and 3.10(b) indicate that $E[N_S]$ and TP of the system grow up as the arrival rate of the jobs increases.

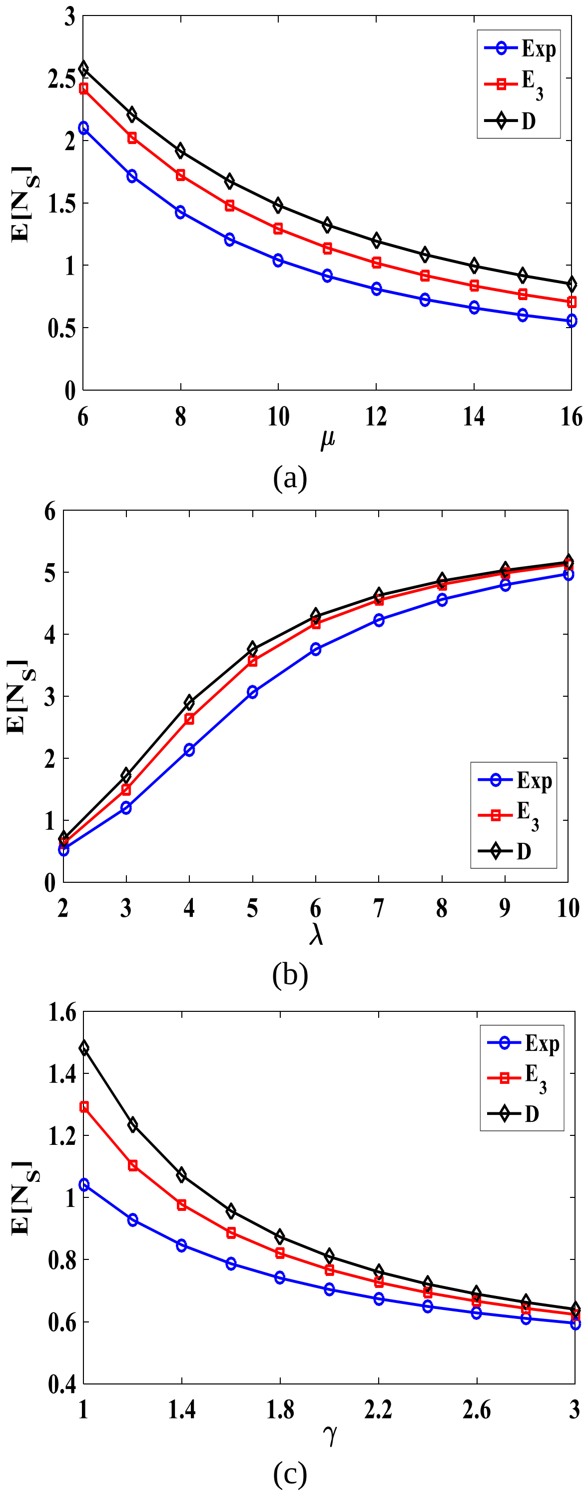


Figure 3.9: $E[N_S]$ vs (a) μ (b) λ (c) γ for different distributions for TSM

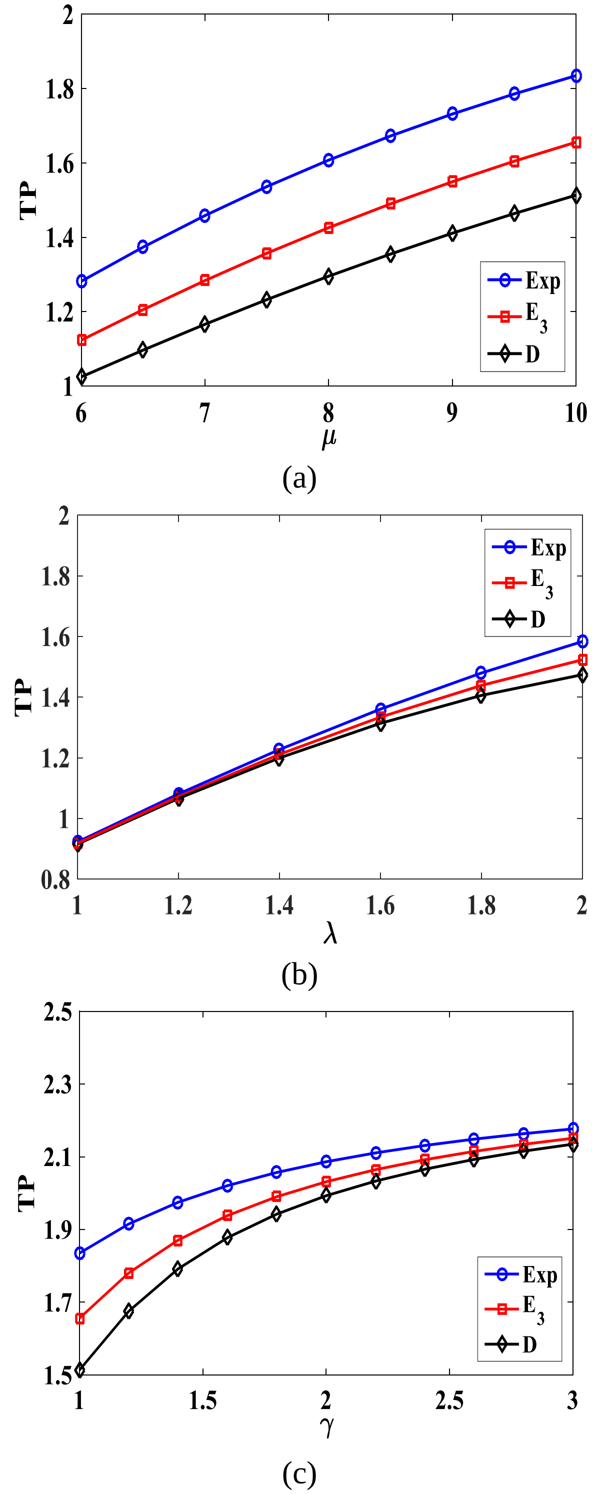


Figure 3.10: TP vs (a) μ (b) λ (c) γ for different distributions for TSM

Effect of γ **:** Table 3.10 reveals that the retrial rate (γ) of the jobs has no significant effects on the system indices. As the retrial rate goes up, $E[N_q]$ seems to decrease with a very slow pace. From Table 3.10, it is noticed that as the retrial rate increases, the probability of the server being busy (idle) increases (decreases) with very slow rate. From Figure 3.9(c), it is clear that $E[N_S]$ goes up very slowly as the retrial rate (γ) increases. Figure 3.10(c) shows that the throughput of the system increases as γ grows.

(ii) Cost optimization for TSM

The time-sharing model is explored to determine the optimal threshold parameter (F^*) and optimal service rate (μ^*) and the corresponding minimum cost $TC(F^*, \mu^*)$ of the system. It is noticed that the system capacity K can be used as an upper bound for the feasible search space of $'F'$. The combined direct search method and the quasi-Newton method are used to evaluate F^* and μ^*, respectively. For determining the total cost, the four cost sets have been taken into consideration as given in Table 3.11.

Table 3.11: Cost sets with different cost elements (in $) for TSM

Cost Set	C_I	C_B	C_H	C_F	C_O
I	30	40	120	60	90
II	10	10	120	15	90
III	15	5	120	15	90
IV	10	10	100	15	110

Optimal threshold parameter (F^*)**:** To evaluate the optimal threshold parameter $'F'$, the default parameters are set as $K=20,\ \lambda=4,\ \gamma=3,\ \varepsilon=1$. For cost set-I, by varying the value of F from 0 to 18, the total cost TC is computed. Table 3.12 provides the minimum cost of the system corresponding to the optimal threshold parameter for the exponential, Erlang-3 and deterministic distributed retrial time by taking $\gamma = 1, 3, 5$. The trend of TC by varying F shown in Figure 3.11 reveals that the optimal threshold parameter (F^*) lies in the feasible range $(0 \le F < K-1)$ of F .

Table 3.12: Searching the optimal F for TSM

γ	$(F^*, TC(F^*, \mu))$		
	Exp	E_3	D
1	(2, 822.28)	(1, 858.94)	(1, 872.35)
3	(5, 665.29)	(5, 678.51)	(5, 687.12)
5	(6, 627.96)	(6, 632.94)	(6, 635.92)

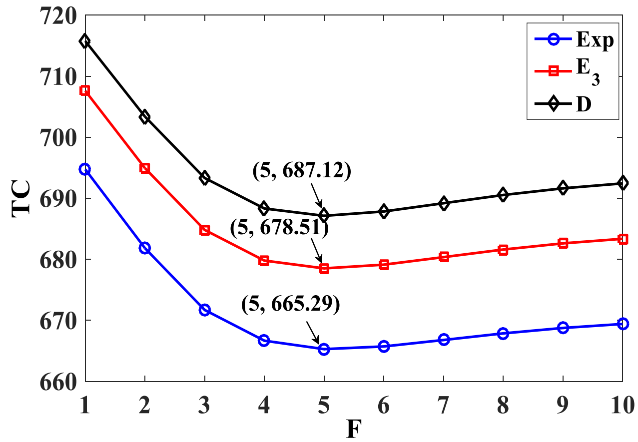
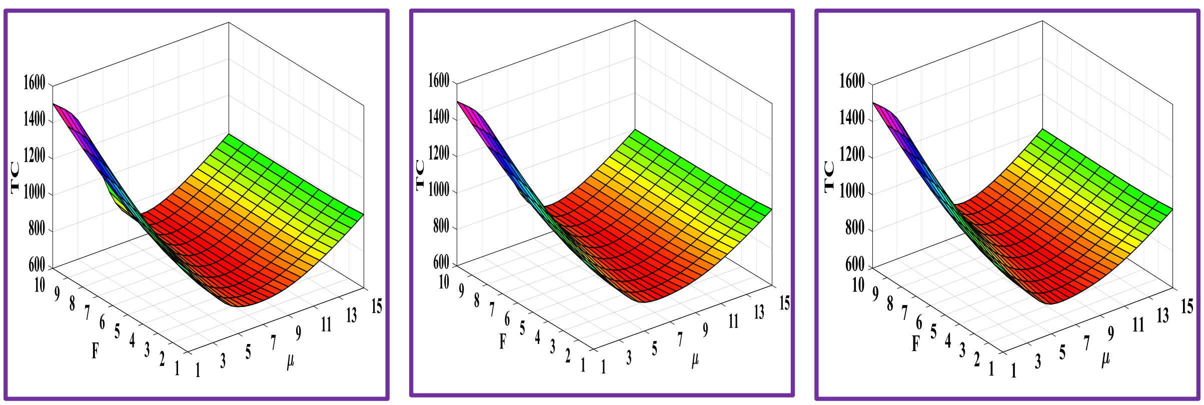


Figure 3.11: TC vs F for different distributions when $\gamma = 3$ for TSM



(a) (b) (c)

Figure 3.12: TC vs μ and F for (a) Exp (b) E_3 (c) D by taking Cost Sets- I for TSM

Optimal service rate (μ^*)**:** For the time-sharing system, the service can be controlled by the system developers so as to provide the better service but at minimum cost. For $\gamma = 3, F^* = 5$, to determine the optimal service rate (μ^*), quasi-Newton method is applied by considering the retrial times as exponential. It is noticed that after 5th iterations, the minimum cost $TC(F^*, \mu^*)$ is achieved at \$659.19 corresponding to $\mu^* = 7.271$ as can be seen from Table 3.13.

Now, we perform a numerical experiment to determine μ^*. First of all, we use the optimal threshold parameter F^* which is given in Table 3.12 and then use the quasi-Newton method for exponential, Erlang-3 and deterministic distributions by taking $\gamma = 1,3,5$. The optimal values of parameters and corresponding minimum cost are shown in Table 3.14.

Table 3.13: Searching the μ^* by quasi-Newton method for exponential distribution $(F^*=5, \gamma=3)$

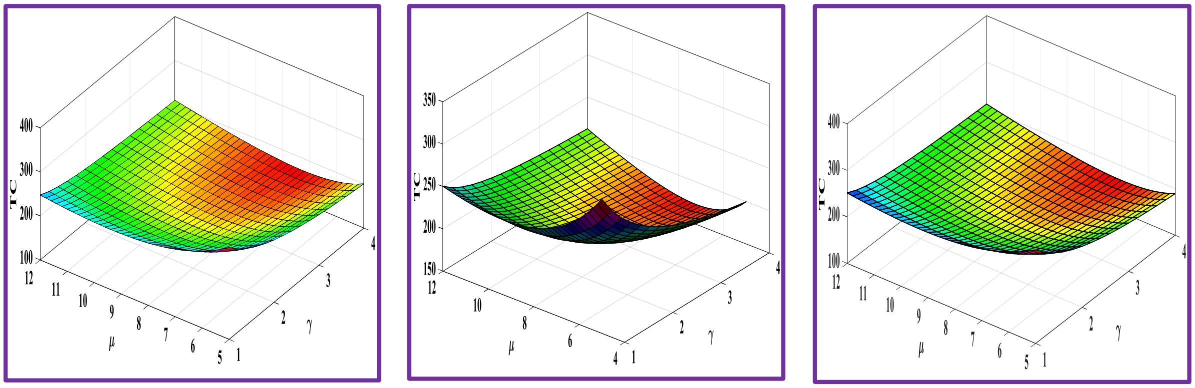
Iterations	F^*	μ	$TC(F^*,\mu)$	Max. tolerance
0	5	8	665.286	1.58E+01
1	5	7	660.172	7.35E+00
2	5	7.3170	669.221	1.16E+00
3	5	7.2739	659.194	6.86E-02
4	5	7.2712	659.194	7.04E-04
5	5	7.2712	659.194	0

Table 3.14: Minimum cost $(F^*, \mu^*, TC(F^*, \mu^*))$ for $\gamma = 1, 3$ and 5 for TSM

γ	$(F^*, \mu^*, TC(F^*, \mu^*))$		
	Exp	E_3	*D*
1	(2, 6.169, 803.45)	(1, 5.854, 809.85)	(1, 6.092, 820.80)
3	(5, 7.271, 659.19)	(5, 7.341, 673.70)	(5, 7.378, 682.92)
5	(6, 7.012, 615.23)	(6, 7.048, 621.22)	(6, 7.068, 624.76)

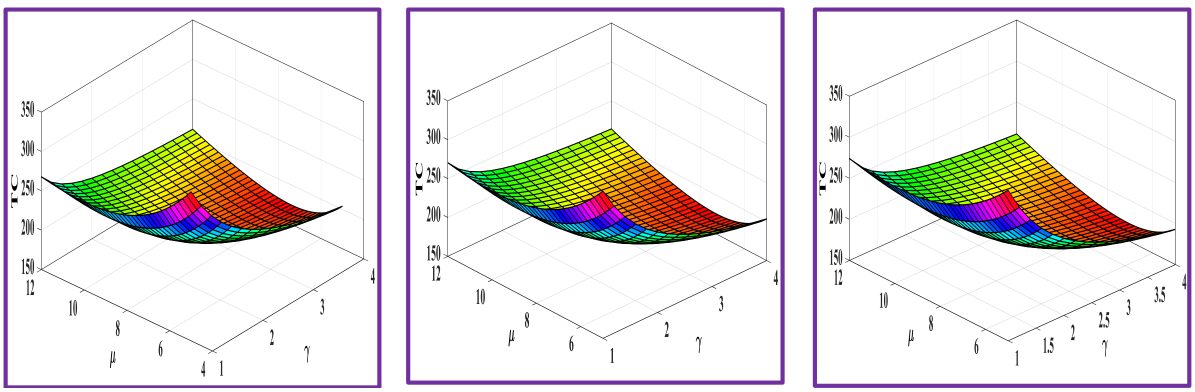
It is seen that the minimum cost for the exponentially distributed retrial time is less as compared to Erlang-3 and deterministic distributions for the retrial time. From Figures 3.12(a-c), it is clear that the cost function is convex and the minimum value of the cost is achieved at the optimal threshold parameter and optimal service rate. From the graphs plotted in Figures 3.12(a-c), it is also observed that the exponential distribution gives the lowest value of the minimum cost in comparison to other distributions for retrial time.

For the cost sets II, III and IV, by fixing $K=7$, $F=4$, $\lambda=2.8, \varepsilon=1$, the surface graphs for the total cost function are shown in Figures 3.13-3.15 by varying μ from 5 to 12 and γ from 1 to 4. It is seen that the cost functions are convex for all the three distributions viz. exponential, Erlang-3 and deterministic distributions and the minimum value of the cost can be attained at an optimal service rate.



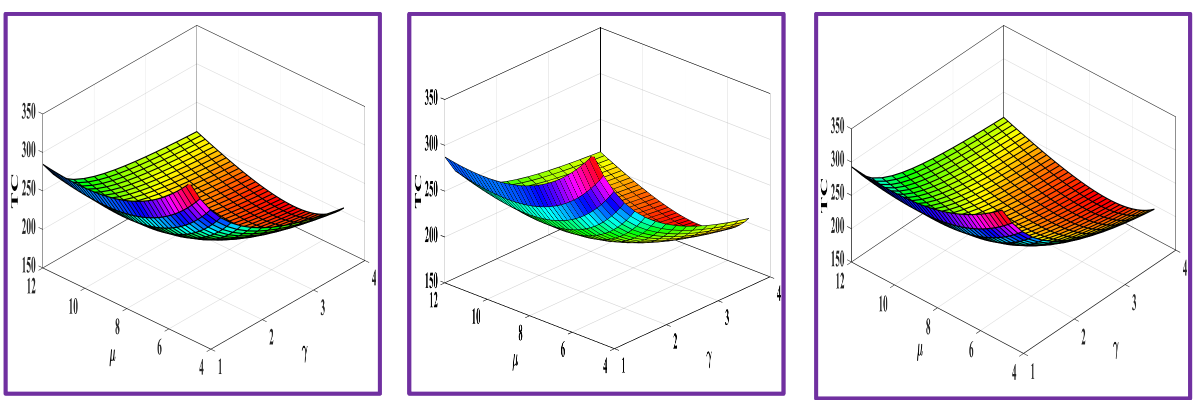
(a) (b) (c)

Figure 3.13: *TC* vs μ and γ for exponential distribution by taking (a) Cost Set-II (b) Cost Set-III (c) Cost Set-IV for TSM



(a) (b) (c)

Figure 3.14: *TC* vs μ and γ for Erlang-3 distribution by taking (a) Cost Set-II (b) Cost Set-III (c) Cost Set-IV for TSM



(a) (b) (c)

Figure 3.15: *TC* vs μ and γ for deterministic distribution by taking (a) Cost Set-II (b) Cost Set-III (c) Cost Set-IV for TSM

3.7. CONCLUDING REMARKS

This chapter investigates the single server finite capacity queueing model with general retrial times by adding the realistic features of admission control *F*-policy and state dependent rates using the supplementary variable corresponding to remaining retrial times. The recursive method is applied to establish the steady state queue size distributions of the system. Some specific distributions viz. exponential, Erlang-3 and deterministic distributions are considered for the general retrial times for providing the computational results for various performance indices. The system indices measured by taking numerical illustration show the validity of the model in real time system. The performance analysis of the concerned model may be applicable to many real-world congestion problems to chalk out the admission control policy by controlling the arrivals of jobs in the system in particular when the capacity of the system is not sufficient for heavy traffic. Further, optimal control parameters and corresponding minimal cost of the system determined by direct search method and quasi-Newton method which may be helpful to the system organizers and decision makers for improving the grade of service of existing system.

Chapter 4

F-Policy for Multi-Server Queue with an Additional Server and Balking

4.1. INTRODUCTION

In the real world, multi-server queues are formed everywhere including various service centers such as at post office, ATMs and many other places. The customers' balking behavior is a common issue in many congestion problems as such a queueing model incorporating the concept of balking may be more appropriate to analyze these problems. In practice, additional server queueing model which also deals with the balking behavior of the customers, has certain advantages by reducing the waiting time of the customers and enhancing the profit to the organization. In recent years, some researchers have also paid their attention towards multi-server Markovian congestion problem. The stationary behavior of M/M/C queueing model in the random environment was analyzed by Liu and Yu (2016). Baumann and Sandmann (2017) dealt with a multi-server tandem queueing model by considering the Markovian arrival process. Laxmi and Kassahun (2017) presented a novel work for the infinite capacity queue with multi-server and discouraged customers. In their investigation, they have analyzed Markovian feedback queue with reneging and balking.

Sometimes in the queueing system, it is noticed that due to the long queue in front of the servers, the customers do not wish to join the system and may balk without receiving the service. The concept of additional server may reduce the queue length and balking behavior of the customers. To be more specific, we illustrate the queueing scenario of customer care department of production organization where calls come regarding many complaints/queries. Sometime due to congestion of the calls, the subscribers may get a busy signal and wait in the queue. To reduce the long queue of the calls, one additional customer care executive may be added to the pool of existing executives. In queueing literature a few researchers have developed such type of queueing models by considering additional servers after a certain workload (cf. Mokaddis and

Jaki, 1983; Abou-El-Ata and Shawky, 1992; Shawky, 1997). Jain and Sharma (2002) dealt with the Markovian multi-server finite queue with additional servers and balking behavior of the customers. In this investigation, they derived various system indices. Markovian multi-server queueing model with balking, provision of additional servers was analyzed by Jain and Singh (2002). By considering the ‘R’ permanent and ‘r’ additional removable repairmen, a multi-component machine repair problem was studied by Jain et al. (2014). They have developed a Markovian model to analyze the queueing situations by including reneging parameter and time-sharing concept. The steady-state analysis of multi-server queue with customer’s discouragement behavior was investigated by Jain et al. (2019). Queue size probability distributions was obtained using recursive method which is further used to derive various performance metrics.

The optimal *F*-policy can be implemented to the queueing system in order to facilitate the quality of the service to the customers by controlling the arrivals and to enhance the profit of the service systems operating in many real-world queueing scenarios. The main objective of the *F*-policy is to deal with the control of the arrival process in case when the waiting area is fully occupied. For the detail description of *F*-policy, we refer Section 1.5 of Chapter 1.

In this chapter, we develop a Markov multi-server queueing model with the provision of additional removable server and balking. By including the realistic features of the balking behavior of the customers and secondary server in rush hour, the present study on the finite multi-server queue under *F*-policy portraits many real-world queueing scenarios. The present chapter is arranged in different sections apart from the ongoing introduction section. In Section 4.2, the description of the concerned model is given. Chapman-Kolmogorov equations are framed and solved by using the recursive method in Section 4.3. The system performance indices and cost function are formulated in Section 4.4. Section 4.5 presents the numerical illustration, ANFIS computing and cost analysis. Finally, Section 4.6 concludes the noble features and the outcomes of the queueing model studied.

4.2. MODEL DESCRIPTION

Consider the finite capacity multi-server Markovian queueing model with two type of customers and the provision of an additional server under admission control *F*-policy by including the balking behavior of the customers. The assumptions for the mathematical model are outlined as follows:

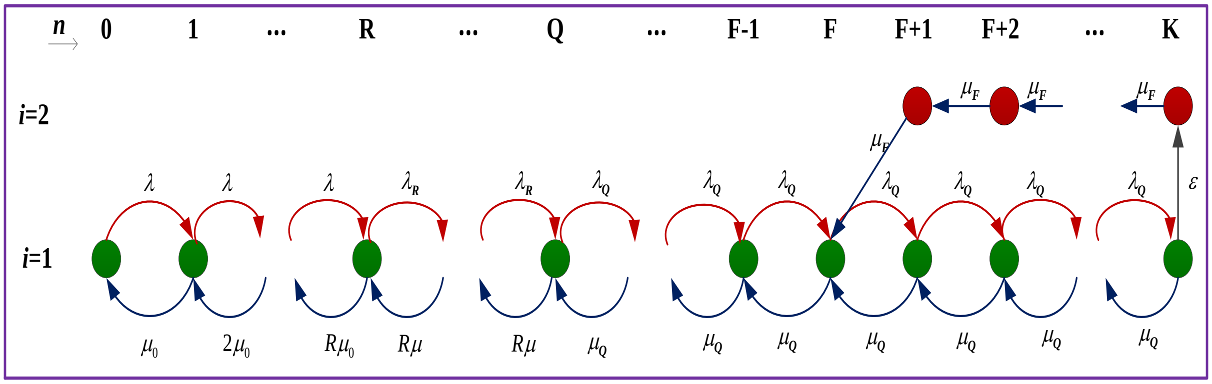


Figure 4.1: Transition state diagram for M/M/R/K model

- Two types of the customers namely, Class-1 and Class-2 join the system in Poisson fashion with rate λ_0 and λ_1, respectively.
- There is the provision of R permanent servers who are always available in the system to render the service to the customers according to FCFS discipline.
- The permanent servers provide service to both classes of customers according to an exponential distribution with mean $1/\mu_0$. When all the permanent servers are engaged, they switch over to faster service mode to reduce the queue size and render the service with rate μ.
- When all permanent servers (R) become busy, the Class-2 customers show the normal behavior in joining the system whereas the Class-1 customers may balk with probability $(1-\beta_0)$.
- At the threshold level Q of queue length, the Class-1 customers are completely lost in the system whereas the Class-2 customers join the system with probability β_1.
- At threshold level Q, one additional server is added to reduce the queue length and balking behavior of the customers. Due to the feasibility of additional server at threshold Q, the integrated service rate at this level 'Q' of queue size, is $\mu_Q = R\mu + \mu_1$ where μ_1 denotes the service rate of an additional server.
- As the system becomes full *i.e.* all waiting space is occupied by the customers, Class-2 customers (Class-1 customers already lost) are not permitted to join the system. To stop the customers from joining the system when the system capacity becomes full, exponentially distributed setup job with rate ε is required. However when the queue size drops to a prefixed threshold value 'F' $(0 \le F < K-1)$, Class-2 customers are further permitted to queue up. When the customers are not allowed due to F-policy, the servers

provide service to the customers with integrated rate $\mu_F = R\mu + \mu_2$, where μ_2 is the service rate of the additional server in the faster mode when new customers are restricted to join the system.

Let at the time τ, the random variables $N(\tau)$ and $S(\tau)$ denote the number of customers and servers' status respectively. The server status $S(\tau)$ is defined as $S(\tau) = 1(2)$ in the case when the servers are engaged and the customers are permitted (not permitted) to join in the system. The state transition diagram of system states is shown in Figure 4.1. To formulate the model, let us consider a bi-variate Markov process $\{S(\tau), N(\tau) : \tau \geq 0\}$, which is discrete in state space and continuous in time. The system states probability at a time τ for the node (i,n) is defined by $P_{i,n}(\tau) = \text{Prob}\{S(\tau) = i, N(\tau) = n\}$. Markov model is analyzed at steady state for which we define the state probabilities by $P_{i,n} = \lim_{\tau\to\infty} P_{i,n}(\tau), i = 1, 2.$

The state-dependent arrival and service rate are respectively, as follows:

$$\lambda(n) = \begin{cases} \lambda, & 0 \leq n \leq R-1, \\ \lambda_R, & R \leq n \leq Q, \\ \lambda_Q, & Q+1 \leq n \leq K, \end{cases} \quad \text{and} \quad \mu(n) = \begin{cases} n\mu_0, & 1 \leq n \leq R-1, \\ R\mu, & R \leq n \leq Q, \\ \mu_Q, & Q+1 \leq n \leq K, \\ \mu_F, & F+1 \leq n \leq K, \end{cases}$$

where $\lambda = \lambda_0 + \lambda_1$, $\lambda_R = \lambda_0\beta_0 + \lambda_1$, $\lambda_Q = \lambda_1\beta_1$, $\mu_Q = R\mu + \mu_1$, $\mu_F = R\mu + \mu_2$.

4.3. GOVERNING EQUATIONS AND SOLUTION

To investigate the queue size of multi-server queueing system, the steady state probabilities of the system state are required. Chapman-Kolmogorov equations for two levels (*i.e.*, when $S(\tau) = i = 1, 2$) of the model can be framed by using the probability arguments and in-flow rates and out-flow rates of each state (see Figure 4.1). Now, we formulate the equations governing the model by balancing the in-flows and out-flows as follows:

(i) ***For*** $i = 1$: ***When the customers are permitted to enter the system.***

$$-\lambda P_{1,0} + \mu_0 P_{1,1} = 0 \tag{4.1}$$

$$-(\lambda + n\mu_0)P_{1,n} + \lambda P_{1,n-1} + (n+1)\mu_0 P_{1,n+1} = 0,\ 1 \leq n \leq R-1 \tag{4.2}$$

$$-(\lambda_R + R\mu_0)P_{1,R} + \lambda P_{1,R-1} + R\mu P_{1,R+1} = 0 \tag{4.3}$$

$$-(\lambda_R + R\mu)P_{1,n} + \lambda_R P_{1,n-1} + R\mu P_{1,n+1} = 0,\ R+1 \leq n \leq Q-1 \tag{4.4}$$

$$-(R\mu+\lambda_Q)P_{1,Q}+\lambda_R P_{1,Q-1}+\mu_Q P_{1,Q+1}=0 \tag{4.5}$$

$$-(\mu_Q+\lambda_Q)P_{1,n}+\lambda_Q P_{1,n-1}+\mu_Q P_{1,n+1}=0,\ Q+1\le n\le F-1 \tag{4.6}$$

$$-(\mu_Q+\lambda_Q)P_{1,F}+\lambda_Q P_{1,F-1}+\mu_Q P_{1,F+1}+\mu_F P_{2,F+1}=0 \tag{4.7}$$

$$-(\mu_Q+\lambda_Q)P_{1,n}+\lambda_Q P_{1,n-1}+\mu_Q P_{1,n+1}=0,\ F+1\le n\le K-1 \tag{4.8}$$

$$-(\mu_Q+\varepsilon)P_{1,K}+\lambda_Q P_{1,K-1}=0 \tag{4.9}$$

(ii) ***For*** $i=2$ ***: When the customers are restricted to enter the system.***

$$\mu_F P_{2,n}=\mu_F P_{2,n+1}=\varepsilon P_{1,K},\ F+1\le n\le K-1 \tag{4.10}$$

Theorem 4.1: The steady-state queue size distribution for the multi-server queueing model with additional server operating under F-policy is given by

$$P_{i,n}=\begin{cases} \dfrac{\lambda^n}{n!\mu_0^n}P_{1,0}, & 1\le n\le R, i=1,\\ \dfrac{\lambda_R^{n-R}\lambda^R}{R!R^{n-R}\mu^{n-R}\mu_0{}^R}P_{1,0}, & R+1\le n\le Q, i=1,\\ \dfrac{\lambda_Q{}^{n-Q}\lambda_R^{Q-R}\lambda^R}{R!\mu_Q{}^{n-Q}R^{Q-R}\mu^{Q-R}\mu_0{}^R}P_{1,0}, & Q+1\le n\le F, i=1,\\ \dfrac{\lambda_Q{}^{n-F}g(n)}{\varepsilon\lambda_Q{}^{K-F-1}+\mu_Q g(F+1)}\dfrac{\lambda_Q{}^{F-Q}\lambda_R^{Q-R}\lambda^R}{R!\mu_Q{}^{F-Q}R^{Q-R}\mu^{Q-R}\mu_0{}^R}P_{1,0}, & F+1\le n\le K-1, i=1,\\ \dfrac{\lambda_Q{}^{K-F}}{\varepsilon\lambda_Q{}^{K-F-1}+\mu_Q g(F+1)}\dfrac{\lambda_Q{}^{F-Q}\lambda_R^{Q-R}\lambda^R}{R!\mu_Q{}^{F-Q}R^{Q-R}\mu^{Q-R}\mu_0{}^R}P_{1,0}, & n=K, i=1,\\ \dfrac{\varepsilon\lambda_Q{}^{K-F}}{\mu_F\left[\varepsilon\lambda_Q{}^{K-F-1}+\mu_Q g(F+1)\right]}\dfrac{\lambda_Q{}^{F-Q}\lambda_R^{Q-R}\lambda^R}{R!\mu_Q{}^{F-Q}R^{Q-R}\mu^{Q-R}\mu_0{}^R}P_{1,0}, & F+1\le n\le K, i=2, \end{cases} \tag{4.11}$$

where $g(n)=\mu_Q{}^{K-n-1}(\mu_Q+\varepsilon)+\varepsilon\sum_{j=1}^{K-n-1}\left(\mu_Q{}^{K-n-1-j}\lambda_Q{}^j\right).$

Proof: From (4.1), it follows

$$P_{1,1}=\frac{\lambda}{\mu_0}P_{1,0} \tag{4.12}$$

Set $n=1$ in (4.2) and using (4.12), we have

$$P_{1,2}=\frac{\lambda^2}{2\mu_0^2}P_{1,0} \tag{4.13}$$

In general, we have

$$P_{1,n}=\frac{\lambda^{n}}{n!\mu_0^{n}}P_{1,0},1\le n\le R \tag{4.14}$$

Now (4.3) and (4.14) yield

$$P_{1,R+1}=\frac{\lambda_R}{R\mu}\frac{\lambda^{R}}{R!\mu_0{}^{R}}P_{1,0} \tag{4.15}$$

Set $n=R+1$ in (4.4) and using (4.14) and (4.15), we obtain

$$P_{1,R+2}=\frac{\lambda_R^{2}}{R^{2}\mu^{2}}\frac{\lambda^{R}}{R!\mu_0{}^{R}}P_{1,0} \tag{4.16}$$

In general, we have

$$P_{1,n}=\frac{\lambda_R{}^{n-R}\lambda^{R}}{R!R^{n-R}\mu^{n-R}\mu_0{}^{R}}P_{1,0},\ R+1\le n\le Q \tag{4.17}$$

Now (4.5) and (4.17) yield

$$P_{1,Q+1}=\frac{\lambda_Q}{\mu_Q}\frac{\lambda_R^{Q-R}}{R^{Q-R}\mu^{Q-R}}\frac{\lambda^{R}}{R!\mu_0{}^{R}}P_{1,0} \tag{4.18}$$

Set $n=Q+1$ in (4.6) and using (4.17) and (4.18), we obtain

$$P_{1,Q+2}=\frac{\lambda_Q^{2}}{\mu_Q^{2}}\frac{\lambda_R^{Q-R}}{R^{Q-R}\mu^{Q-R}}\frac{\lambda^{R}}{R!\mu_0{}^{R}}P_{1,0} \tag{4.19}$$

In general, we have

$$P_{1,n}=\frac{\lambda_Q{}^{n-Q}\lambda_R^{Q-R}\lambda^{R}}{R!\mu_Q{}^{n-Q}R^{Q-R}\mu^{Q-R}\mu_0{}^{R}}P_{1,0},Q+1\le n\le F \tag{4.20}$$

Adding (4.7)-(4.9) and using (4.10) and (4.20),we find

$$P_{1,n}=\frac{\lambda_Q{}^{n-F}g(n)}{\varepsilon\lambda_Q{}^{K-F-1}+\mu_Q g(F+1)}\frac{\lambda_Q{}^{F-Q}\lambda_R^{Q-R}\lambda^{R}}{R!\mu_Q{}^{F-Q}R^{Q-R}\mu^{Q-R}\mu_0{}^{R}}P_{1,0},F+1\le n\le K-1 \tag{4.21}$$

Using (4.21) in (4.9), we obtain

$$P_{1,K}=\frac{\lambda_Q{}^{K-F}}{\varepsilon\lambda_Q{}^{K-F-1}+\mu_Q g(F+1)}\frac{\lambda_Q{}^{F-Q}\lambda_R^{Q-R}\lambda^{R}}{R!\mu_Q{}^{F-Q}R^{Q-R}\mu^{Q-R}\mu_0{}^{R}}P_{1,0} \tag{4.22}$$

Equations (4.10) and (4.22) yield

$$P_{2,n} = \frac{\varepsilon \lambda_Q^{\,K-F}}{\mu_F \left[\varepsilon \lambda_Q^{\,K-F-1} + \mu_Q g(F+1) \right]} \frac{\lambda_Q^{\,F-Q} \lambda_R^{Q-R} \lambda^R}{R!\, \mu_Q^{\,F-Q} R^{Q-R} \mu^{Q-R} \mu_0^{\,R}} P_{1,0},\ F+1 \le n \le K \tag{4.23}$$

where $P_{1,0}$ is determined by using normalizing condition

$$\sum_{n=0}^{K} P_{1,n} + \sum_{n=F+1}^{K} P_{2,n} = 1 \tag{4.24}$$

4.4. PERFORMANCE MEASURES AND COST FUNCTION

To examine the characteristics and behavior of the concerned queueing system, several performance measures namely average system size $E[N_S]$, an average number of idle permanent servers $E[I_S]$ and the average number of busy permanent servers $E[B_S]$ are derived. The long-run probabilities of the server status namely probability of additional server being busy P_{ASB} and the probability of all permanent servers being busy P_{PSB} have also been derived. Several system indices established are as follows:

(i) $$E[N_s] = \sum_{n=1}^{K} nP_{1,n} + \sum_{n=F+1}^{K} nP_{2,n} \tag{4.25}$$

(ii) $$E[I_s] = \sum_{n=0}^{R-1} (R-n) P_{1,n} \tag{4.26}$$

(iii) $$E[B_s] = \sum_{n=1}^{R-1} nP_{1,n} + R\left(\sum_{n=R}^{K} P_{1,n} + \sum_{n=F+1}^{K} P_{2,n} \right) \tag{4.27}$$

(iv) $$P_{ASB} = \sum_{n=Q+1}^{K} P_{1,n} + \sum_{n=F+1}^{K} P_{2,n} \tag{4.28}$$

(v) $$P_{PSB} = \sum_{n=R}^{K} P_{1,n} + \sum_{n=F+1}^{K} P_{2,n} \tag{4.29}$$

(vi) Cost function

For the multi-server queueing model under F-policy and supported by one additional server in rush hour to reduce the balking behavior of the customers, the organization may be interested in the total expected cost incurred in various activities of the system. The cost function per unit time depends on various cost elements. To frame cost function, we consider per unit time cost elements related to (i) holding cost of each customers present in the system

(C_H) (ii) cost for providing service when at least one server idle (C_{μ_0}) (iii) cost incurred during all permanent servers being busy (C_μ), (iv) cost C_{μ_Q} (C_{μ_F}) incurred when one additional server is added and customers are allowed (not allowed) and (v) the cost incurred on each idle permanent server (C_I). The cost function involves the service rate (μ), the number of permanent servers (R), the threshold parameter (F) and the capacity of the system (K) as decision variables. The optimal values of (F, K, R, μ), say (F^*, K^*, R^*, μ^*) can be determined by minimizing the expected total cost per unit time TC. Now we construct the cost function per unit time in terms of cost factors related to different activities as follows:

$$\begin{aligned} TC(F,K,R,\mu) = C_H E[N_S] + \mu_0 C_{\mu_0} \sum_{n=1}^{R-1} nP_{1,n} + R\mu C_\mu \left(\sum_{n=R}^{K} P_{1,n} + \sum_{n=F+1}^{K} P_{2,n} \right) + \mu_1 C_{\mu_Q} \sum_{n=Q+1}^{K} P_{1,n} \\ + \mu_2 C_{\mu_F} \sum_{n=F+1}^{K} P_{2,n} + C_I \sum_{n=0}^{R-1} (R-n) P_{1,n}. \end{aligned} \tag{4.30}$$

4.5. NUMERICAL RESULTS AND COST OPTIMIZATION

In this section, the numerical experiment is done for examining the admission control policy for finite capacity multi-server queue with an additional server and balking. The sensitivity analysis is carried out which may be very useful to analyze the effects of the parameters on several system indices.

4.5.1. Sensitivity Analysis

For the computational purpose, we fix the system parameters as $\lambda_0 = 5,\ \lambda_1 = 3,\ \mu = 0.8,$ $\beta_0 = 0.5,\ \beta_1 = 0.9,\ \varepsilon = 1,\ \mu_0 = 0.9\mu,\ \mu_1 = 1.1\mu,\ \mu_2 = 1.2\mu,\ R = 4,\ Q = 11,\ F = 15,\ K = 20.$

Based on the numerical result obtained, we explain the effect of the parameters on various system indices.

As expected, it is observed that by keeping arrival rate constant as the servers provide service with faster rate, the average number of idle (busy) servers, increases (decreases). The status of the servers depends on the service rate as such the probability of the additional server being busy and the probability of the permanent servers busy become low by facilitating the service with a faster rate. Table 4.1 and Table 4.3 reveal the same facts. From Table 4.2 and Table 4.4, it is noticed that as λ_1 increases, $E[I_S]$ decreases while $E[I_S]$ grows up. On the other hand, we see that the long run probability P_{ASB} increases with faster rate whereas P_{PSB} increases at a slower rate as λ_1 increases. Figure 4.3 and 4.4 depict the effect of λ_1 on $E[N_S]$ for varying values of joining parameters β_1 and β_0, respectively. It is seen that as λ_1 grows, $E[N_S]$ also grows which is quite obvious due to the fact that, by a faster arrival rate, the system becomes crowded.

Table 4.1: System indices by varying values of μ and β_1

β_1	μ	$E[I_s]$	$E[B_s]$	P_{ASB}	P_{PSB}
0.5	1	0.024269	3.975731	0.107796	0.982184
	2	0.649855	3.350145	0.002557	0.623634
	3	1.397539	2.602461	0.000080	0.318095
0.7	1	0.022613	3.977387	0.168659	0.983399
	2	0.649019	3.350981	0.003840	0.624118
	3	1.397487	2.602513	0.000117	0.318120
0.9	1	0.020534	3.979466	0.245094	0.984925
	2	0.648052	3.351948	0.005324	0.624678
	3	1.397430	2.602570	0.000157	0.318148

Table 4.2: System indices by varying values of λ_1 and β_1

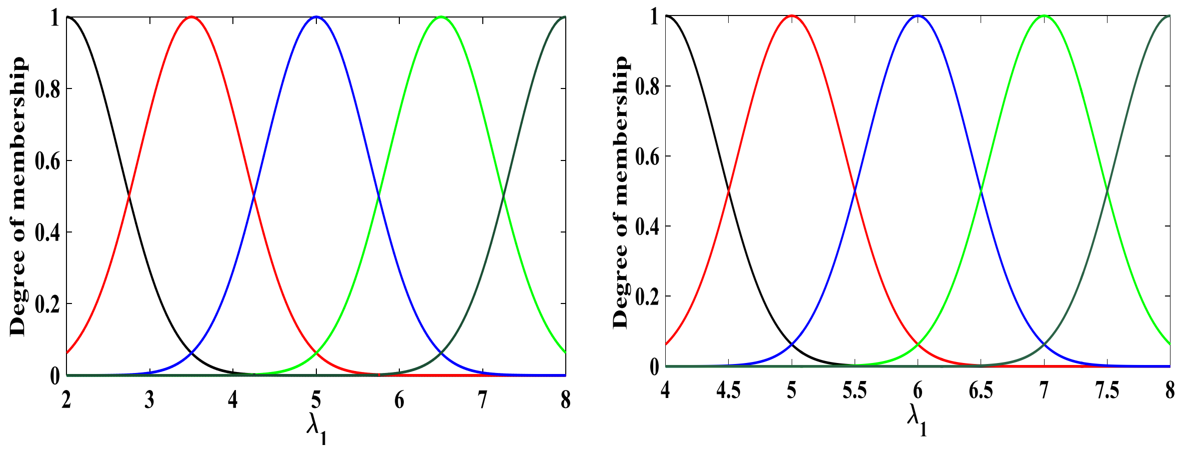
β_1	λ_1	$E[I_s]$	$E[B_s]$	P_{ASB}	P_{PSB}
0.5	2	0.019206	3.980794	0.089887	0.985561
	4	0.001233	3.998767	0.327901	0.999017
	6	0.000100	3.999900	0.613910	0.999917
0.7	2	0.018209	3.981791	0.137106	0.986310
	4	0.000890	3.999110	0.514932	0.999291
	6	0.000039	3.999961	0.847570	0.999967
0.9	2	0.017020	3.982980	0.193488	0.987205
	4	0.000538	3.999462	0.706742	0.999571
	6	0.000013	3.999987	0.948871	0.999989

Table 4.3: System indices by varying values of μ and β_0

β_0	μ	$E[I_s]$	$E[B_s]$	P_{ASB}	P_{PSB}
0.5	1	0.020534	3.979466	0.245094	0.984925
	2	0.648052	3.351948	0.005324	0.624678
	3	1.397430	2.602570	0.000157	0.318148
0.7	1	0.007900	3.992100	0.303625	0.994200
	2	0.509237	3.490763	0.013471	0.705073
	3	1.323189	2.676811	0.000479	0.354373
0.9	1	0.003286	3.996714	0.343834	0.997588
	2	0.373200	3.626800	0.026882	0.783859
	3	1.233683	2.766317	0.001216	0.398046

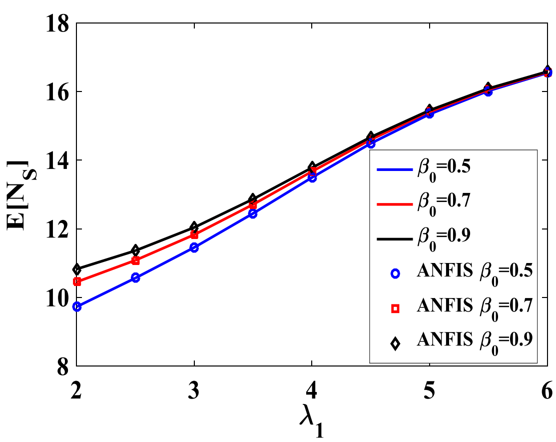
Table 4.4: System indices by varying values of λ_1 and β_0

β_0	λ_1	$E[I_s]$	$E[B_s]$	P_{ASB}	P_{PSB}
0.5	2	0.017020	3.982980	0.193488	0.987205
	4	0.000538	3.999462	0.706742	0.999571
	6	0.000013	3.999987	0.948871	0.999989
0.7	2	0.005386	3.994614	0.249467	0.995951
	4	0.000204	3.999796	0.731001	0.999837
	6	0.000006	3.999994	0.951773	0.999995
0.9	2	0.001919	3.998081	0.286209	0.998557
	4	0.000087	3.999913	0.747112	0.999931
	6	0.000003	3.999997	0.953894	0.999997



(a) (b)

Figure 4.2: Membership function of λ_1 (a) for $\beta_0 = 0.5$ (b) for $\beta_1 = 0.5$



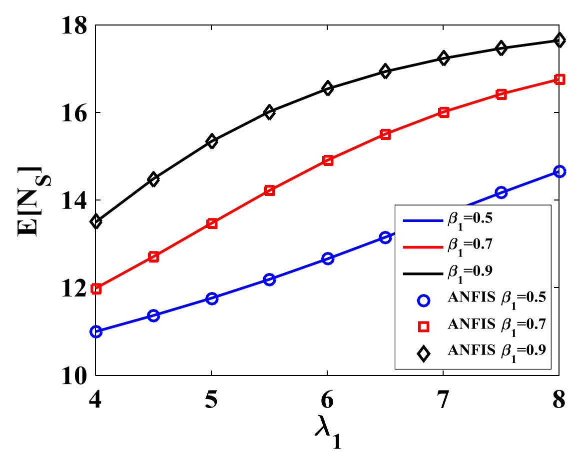


Figure 4.3: $E[N_s]$ vs λ_1 for $\beta_1 = 0.5, 0.7, 0.9$ Figure 4.4: $E[N_s]$ vs λ_1 for $\beta_0 = 0.5, 0.7, 0.9$

The ANFIS approach is also applied to compute the results using the neuro-fuzzy tool in MATLAB software and these results are compared with numerical results obtained by the recursive method. For our model, we consider the one input parameter as λ_1 and one output parameter as $E[N_s]$ (cf. Section 1.4.7 of Chapter 1). We notice that both results almost coincide that validates the usefulness of ANFIS in real time complex queueing systems. By considering the Gaussian membership function and five linguistic values for the input parameter λ_1 as very low, low, average, high, and very high, we proceed to obtain ANFIS results. The membership functions for the input parameter λ_1 are shown in Figures 4.2(a-b). In Figures 4.3 and 4.4, the ANFIS results for $E[N_s]$ are shown by tick marks (circle, square and diamond) while the smooth lines show the trends obtained by an analytical approach.

4.5.2. Cost Optimization

To evaluate the optimal parameters (F^*, K^*, R^*, μ^*) and corresponding minimal cost $TC(F^*, K^*, R^*, \mu^*)$, the cost elements are fixed as $C_H = \$20$, $C_{\mu_0} = \$50$, $C_\mu = \$50$, $C_{\mu_Q} = \$55$, $C_{\mu_F} = \$60$, $C_I = \$10$.

(i) Direct Search Method (DSM)

Since the parameters *'F'*, *'K'* and *'R'* are having the integer values, a direct search method based on a heuristic approach is used.

Optimal threshold parameter 'F': To control the admission of the customers in the system, we compute the optimal threshold parameter F^* so as to minimize the cost function. By choosing the value of $\mu = 0.5$ and using other default system parameters, we compute the F^* and corresponding minimum cost in the desired range of F. F^* and expected minimum cost of the system $(F^*, TC(F^*))$ are obtained for different values of β_0 = 0.3, 0.4 and 0.5 as (12, \$456.05), (13, \$458.54), (13, \$460.30). Figure 4.5 reveals that the cost function is convex and unimodal in the desired range as such minimum cost can be attained.

Optimal system capacity 'K': In order to find the optimal capacity of the system, we set the default system parameters as $\mu = 0.56$, $\beta_0 = 0.5$, $\beta_1 = 0.6$, $F^* = 13$. A unimodal and convex graph is obtained for varying values of K from 18 to 28. From Figure 4.6, it is noticed that when K goes up from 18 to 22 then the expected cost decreases and then after starts increasing up to 28. The optimal capacity of the system is noticed as $K^* = 22$ and the corresponding minimum cost $TC(F^*, K^*)$ is found as \$369.18.

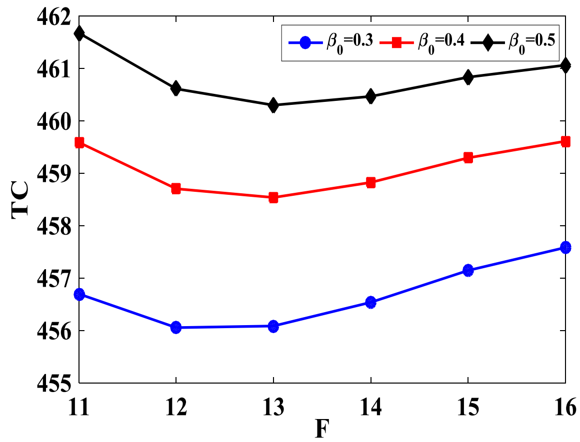


Figure 4.5: TC vs F for $\beta_0 = 0.3, 0.4, 0.5$

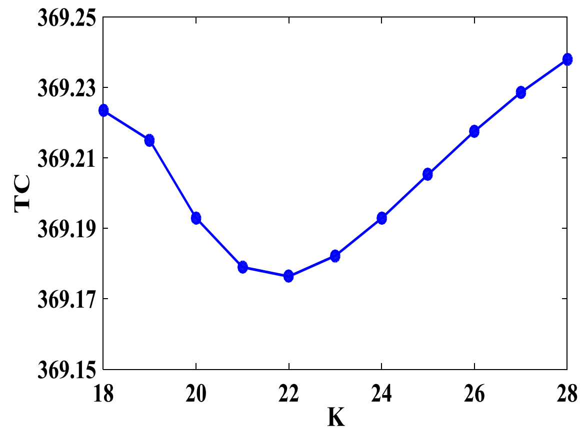


Figure 4.6: Variation in TC for different K

Optimal permanent servers 'R': Now we compute the optimal value of the permanent servers (R^*) by considering the updated values in the default system parameters as $F^* = 13$, $K^* = 22$, $\mu = 0.56$, $\beta_0 = 0.5$ and $\beta_1 = 0.6$. The minimum expected cost of the system corresponding to an optimal number of servers (R^*) is computed for different values of $\lambda_1 = 3, 4, 5, 6$. It is observed that the optimal number of servers (R^*) and corresponding minimum expected cost $(R^*, TC(F^*, K^*, R^*))$ are achieved at (4, \$369.18), (5, \$411.10), (7, \$453.47), (8, \$495.05), respectively. Figure 4.7 shows the surface graph for the expected cost for the varying values of R and β_1.

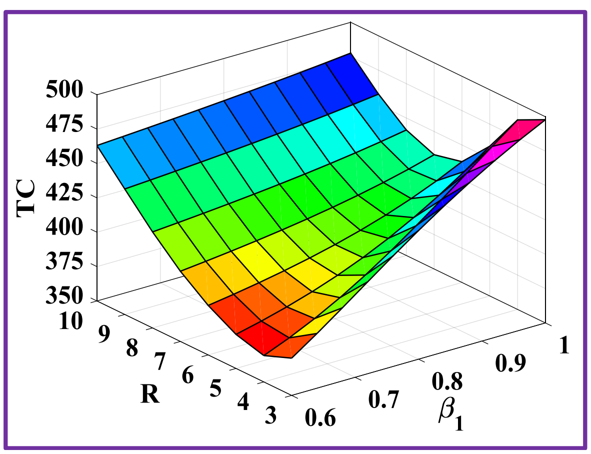


Figure 4.7: TC vs R for varying values of β_1

(ii) Quasi-Newton method (QNM)

For the algorithmic steps of quasi-Newton method, we refer Section 1.4.8 of Chapter 1.

Optimal service rate (μ^*): Now we employ a quasi-Newton approach to compute the optimal service rate (μ^*) which is a continuous decision parameter. For evaluating optimal value μ^* and corresponding minimum cost $(\mu^*, TC(F^*, K^*, R^*, \mu^*))$, we fix the optimal threshold parameters as $F^* = 13$, $K^* = 22$, $R^* = 4$, and $\beta_0 = 0.5$, $\beta_1 = 0.6$.

After performing 5 iterations, as displayed in Table 4.5, it is noticed that maximum tolerance is 1.267×10^{-7} which is less than that of $|\partial TC / \partial \mu| < \varepsilon_0 = 10^{-6}$. Therefore after 5 iterations, we achieve μ^* and minimum system cost $TC(\mu^*, TC(F^*, K^*, R^*, \mu^*))$ as (0.5753, \$369.14). It is also noticed from Figures 4.8 and 4.9, that the expected cost function $TC(\mu)$ is convex with respect to service rate (μ) and β_0, β_1, respectively.

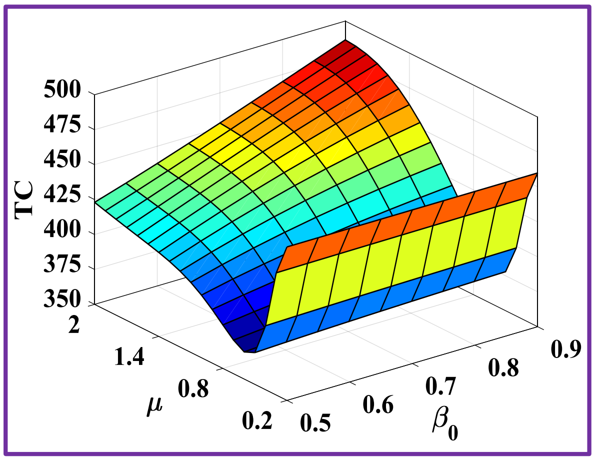


Figure 4.8: TC vs μ for varying values of β_0

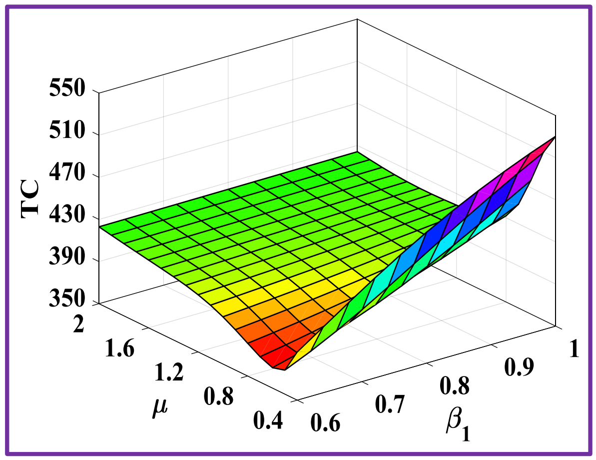


Figure 4.9: TC vs μ for varying values of β_1

Table 4.5: Searching the optimal value of μ by using quasi-Newton method

Iterations	(F^*, K^*, R^*)	μ	$TC(F^*, K^*, R^*, \mu)$	Max. tolerance
0	(13,22,4)	1	394.801	59.2
1	(13,22,4)	0.6111	369.687	28.6
2	(13,22,4)	0.5748	369.142	0.459
3	(13,22,4)	0.5754	369.142	9.59×10^{-2}
4	(13,22,4)	0.5753	369.142	2.48×10^{-4}
5	(13,22,4)	0.5753	369.142	1.267×10^{-7}

4.6. CONCLUDING REMARKS

F-policy for Markovian multi-server finite queueing model with the additional server and balking has investigated by employing the recursive technique to establish the queue size distribution for the customers. The provision of the additional server made in the multi-server queue may be helpful in reducing the balking behavior of the customers. Adaptive neuro fuzzy inference system (ANFIS) approach is successfully applied that validates the feasibility of the use of fuzzy parameters and neural network in our queueing model. The validation of ANFIS in our model gives insight that it may be implemented in several complex queueing problems for which analytical results cannot be derived in closed form. It may also beneficial to earn more profit and reduce waiting time in several queueing scenarios including the hospitals, call centers, computer communication network, ATMs, shopping malls, admission counter in the educational system, and many more. The cost analysis performed may be helpful in setting optimal threshold parameters at the minimum expected cost incurred on the system. Some other system parameters treated as decision variables include the optimal service rate, the optimal number of permanent servers, and the optimal capacity of the system.

Chapter 5

Markovian Queue with General Retrial Times and Discouragement

5.1. INTRODUCTION

In the past, some researchers have focused towards the performance analysis of non-Markovian retrial queueing models and enriched the literature by contributing significantly for both the theoretical and applicable aspects (cf. Phung-Duc, 2017). The non-Markovian queueing model with general retrial attempts and Bernoulli schedule was studied by Gao et al. (2016). In this study, they applied the supplementary variable technique to establish the stationary distribution of the system size. Single server retrial queueing model with an optional vacation for the unreliable server, batch arrivals and M-vacation was studied by Jailaxmi et al. (2017). They have evaluated the probability generating function for the queue size distribution after introducing the supplementary variable (SV). Zirem et al. (2018) studied a single server queue with general retrial times in which customers arrive in a batch and the server is subject to failure. They have used SVT to carry out their study to establish queue size distributions and performance indices.

The cost optimization and evaluation of optimal control parameter for the queueing model using genetic algorithm has rarely appeared in the queueing literature. It is worthwhile to have a look at notable contributions towards optimization problems in different contexts using GA (cf. Raman et al., 2009; Jana and Sharma, 2010; Ke et al., 2010; Huang et al., 2011; Lin and Ke, 2011; Majumdar et al., 2018). In order to balance the cloudlet's loads in a mobile cloud computing system, Rashidi and Sharifian (2017) developed a multi-server queueing model. They have used GA along with ant colony optimization technique for load balancing and determined the optimum mean completion time of offloaded jobs. For the study of queueing characteristics of the allocation problem related to the hub network, Hasanzadeh et al. (2018) employed GA to estimate the optimum waiting cost and queueing capacity of each hub.

In this chapter, we investigate the finite queueing model with general retrial attempts and discouragements under admission control F-policy. The soft computing approaches viz. ANFIS and GA have been used for the performance analysis and cost optimization of the concerned finite model. To investigate the finite queueing model with general repeated attempts, the contents of present chapter are arranged in different sections. Section 5.2 presents the description of the finite capacity retrial model. In Section 5.3, Chapman-Kolmogorov equations are framed by using the SVT corresponding to remaining retrial time. The mathematical analysis of the model is also given in the same section. Several performance metrics and cost optimization algorithms are given in Sections 5.4 and 5.5, respectively. An application example alongwith numerical results and cost optimization are given in section 5.6. Section 5.7 concludes the present investigation by highlighting the noble features.

5.2. MODEL DESCRIPTION

Consider M/M/1//K queueing model with general retrial attempts. The realistic features such as F-policy and discouraged customers are taken into account to develop the finite capacity general retrial model. The customers follow the first-come-first-served (FCFS) rule to join the system. The mathematical construction of the model is based on certain assumptions which are as follows:

- The arriving customers enter into the system in Poisson fashion with parameter λ .
- The arriving tagged customer observes the queue length and decides whether to join the queue or balk. Let β_n be the probability of joining of the customer in the system and $(1-\beta_n)$ represents the balking probability of the customers when there are already n customers present in the system.
- On arrival, if the customer finds the server engaged in the service of other customer, then he is forced to join the retrial orbit. From the retrial orbit, he re-attempts for the service with general distributed retrial time having distribution function $G(x)$, the probability density function $\gamma(x)$ $(x \geq 0)$ with mean retrial time $1/\gamma$. If the arriving customer finds the server free, he immediately gets served; the service time follows the exponential distribution with mean $1/\mu$.
- After admitting in the queue, each customer waits for his turn up to a certain time and then gets impatient and leaves the queue without getting service. The impatience time period

T is governed by an exponential distribution with rate υ. The reneging function for the state $'n'$ which denotes the number of customers present in the system is defined by

$$r_n = \begin{cases} (n-1)\upsilon, & 1 \le n \le K \\ 0, & n > K \end{cases} \tag{5.1}$$

- A setup job is required before restricting the arriving customers from joining the system when the system capacity becomes full. The time required for the setup follows the exponential distribution with mean $1/\varepsilon$.
- As the system reaches its full capacity K i.e. there is no space for waiting, the arriving customers are not permitted to enter into the system until number of customers drops to a threshold level F $(0 \le F < K-1)$.

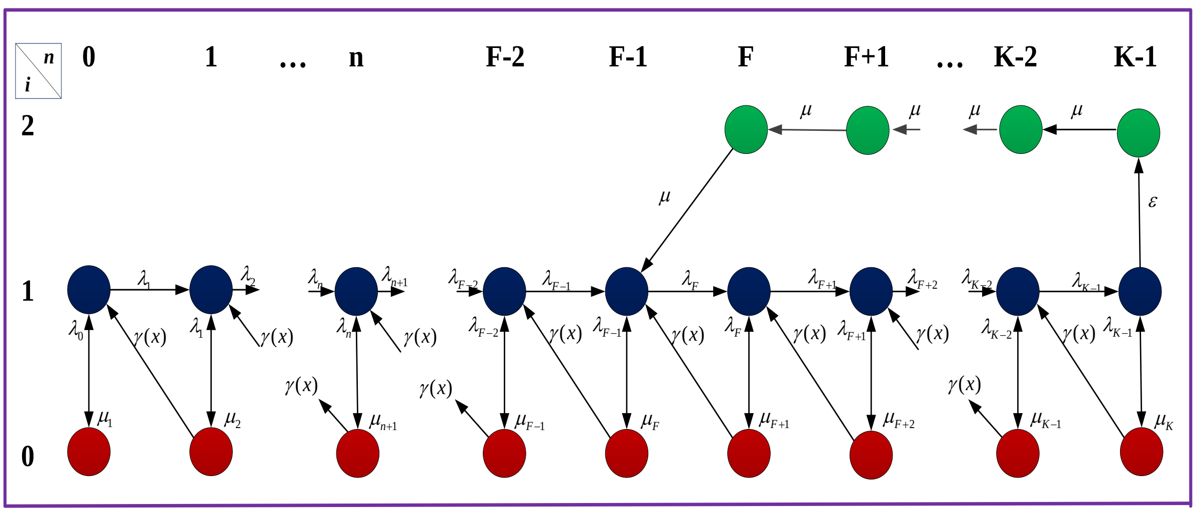


Figure 5.1: State transition diagram

For developing the admission control finite capacity model with general distributed retrial times and discouraged customers, we denote the state dependent arrival rates and service rates by λ_n and μ_n, respectively. Thus

$$\lambda_n = \begin{cases} \lambda\beta_n, & 0 \le n \le K-1 \\ 0, & n \ge K \end{cases} \tag{5.2a}$$

and
$$\mu_n = \mu + r_n, \; 1 \le n \le K. \tag{5.2b}$$

At any time τ, $N(\tau)$ and $S(\tau)$ denote the number of customers present in the system and status of the server, respectively. To formulate the mathematical model, $S(\tau)$ is defined as follows:

$$S(\tau)=i=\begin{cases}0, \text{ the customers are forced to join the retrial pool on finiding the server being busy,}\\ 1(2), \text{ the server is occupied and the customers are permitted (not permitted) to join}\\ \quad \text{the system.}\end{cases}$$

Let the supplementary variable (U) be used corresponding to the remaining retrial time of the customers while waiting in the retrial pool.

The system state probabilities at the time $'\tau'$ are defined as follows:

$P_{0,0}(\tau)=\text{Prob}\{S(\tau)=0, N(\tau)=0\}$

$P_{0,n}(x,\tau)dx=\text{Prob}\{S(\tau)=0, N(\tau)=n, x<U(\tau)\le x+dx\}, x\ge 0, 1\le n\le K-1$

$P_{1,n}(\tau)=\text{Prob}\{S(\tau)=1, N(\tau)=n\},\ 0\le n\le K-1$

$P_{2,n}(\tau)=\text{Prob}\{S(\tau)=2, N(\tau)=n\},\ F\le n\le K-1$

Also $P_{0,n}(\tau)=\text{Prob}\{S(\tau)=0, N(\tau)=n\}=\int_0^\infty P_{0,n}(x,\tau)dx,\ 1\le n\le K-1$.

When $\tau\to\infty$, the steady state probabilities are denoted by

$P_{0,0}=\lim_{\tau\to\infty}P_{0,0}(\tau)$; $P_{0,n}(x)=\lim_{\tau\to\infty}P_{0,n}(x,\tau),\ 1\le n\le K-1$

$P_{1,n}=\lim_{\tau\to\infty}P_{i,n}(\tau),\ 0\le n\le K-1$; $P_{2,n}=\lim_{\tau\to\infty}P_{2,n}(\tau),\ F\le n\le K-1$.

Denote Laplace Stieltjes transform (LST) of $\gamma(x)$ and $P_{0,n}(x)$ by $\gamma^*(\theta)$ and $P_{0,n}^*(\theta)$, respectively. Also, $P_{0,n}^*(\theta)=P_{0,n}\gamma^*(\theta)$.

5.3. GOVERNING EQUATIONS AND ANALYSIS

For the analysis, the supplementary variable technique (SVT) is used by introducing the random variable corresponding to the remaining retrial time. Now, we construct the steady state probability distribution of the non-Markovian model for three levels *i.e.* for $i=0,1,2$. The system state transition diagram is depicted in Figure 5.1. Chapman-Kolmogorov equations for the system states are formulated as follows:

(i) ***For*** $i=0$ **:** ***When the customers are waiting in the retrial orbit.***

$$-\frac{d}{dx}P_{0,n}(x)=-\lambda_n P_{0,n}(x)+\mu_{n+1}P_{1,n}\gamma(x),\ 1\le n\le K-1 \tag{5.3}$$

$$\lambda_0 P_{0,0}=\mu_1 P_{1,0} \tag{5.4}$$

(ii) ***For*** $i=1$: ***When the customers are allowed to enter into the system.***

$$(\lambda_1+\mu_1)P_{1,0}=\lambda_0 P_{0,0}+P_{0,1}(0) \tag{5.5}$$

$$(\lambda_{n+1}+\mu_{n+1})P_{1,n}=\lambda_n P_{1,n-1}+\lambda_n P_{0,n}+P_{0,n+1}(0),\ 1\le n\le F-2 \tag{5.6}$$

$$(\lambda_F+\mu_F)P_{1,F-1}=\lambda_{F-1}P_{1,F-2}+\lambda_{F-1}P_{0,F-1}+P_{0,F}(0)+\mu P_{2,F} \tag{5.7}$$

$$(\lambda_{n+1}+\mu_{n+1})P_{1,n}=\lambda_n P_{1,n-1}+\lambda_n P_{0,n}+P_{0,n+1}(0),\ F\le n\le K-2 \tag{5.8}$$

$$(\varepsilon+\mu_K)P_{1,K-1}=\lambda_{K-1}P_{1,K-2}+\lambda_{K-1}P_{0,K-1} \tag{5.9}$$

(iii) ***For*** $i=2$ **:** ***When the customers are not allowed to enter into the system due to F-policy.***

$$\mu P_{2,n+1}=\mu P_{2,n+2}=\varepsilon P_{1,K-1},\ F\le n\le K-3 \tag{5.10}$$

The queue size distribution is obtained in terms of probability $P_{0,0}$. The normalizing condition is given by

$$\sum_{n=0}^{K-1}P_{0,n}+\sum_{n=0}^{K-1}P_{1,n}+\sum_{n=F}^{K-1}P_{2,n}=1 \tag{5.11}$$

Taking LST of (5.3), we obtain

$$(\lambda_n-\theta)P_{0,n}^{*}\left(\theta\right)=\mu_{n+1}P_{1,n}\gamma^{*}\left(\theta\right)-P_{0,n}\left(0\right),\ 1\le n\le K-1 \tag{5.12}$$

For brevity of notations, we note

$$\prod_{i=n+1}^{n}\delta_i=1,\ \Lambda_{0,n}=\prod_{i=0}^{n}\lambda_i,\ \Lambda_{1,n}^{*}=\left(\prod_{j=1}^{n}\gamma^{*}(\lambda_j)\right)^{-1}\text{ and }\rho_{0,n}=\prod_{i=0}^{n}\frac{\lambda_i}{\mu_{i+1}}.$$

Theorem 6.1: The steady state probability distributions for M/M/1/K retrial queueing model with discouraged customers and operating under *F*-policy, are given by

$$P_{1,n}=\begin{cases}\rho_{0,n}\Lambda_{1,n}^{*}P_{0,0}, & 0\le n\le F-1,\\ \Lambda_{0,n}T_nR_F^{-1}P_{0,0}, & F\le n\le K-2,\\ \Lambda_{0,n}R_F^{-1}P_{0,0}, & n=K-1,\end{cases} \tag{5.13a}$$

$$P_{0,n}=\begin{cases}\rho_{0,n-1}\Lambda_{1,n-1}^{*}\left(\Lambda_{n,n}^{*}-1\right)P_{0,0}, & 1\le n\le F-1,\\ \rho_{n,n}^{-1}\Lambda_{0,n}(1-\gamma^{*}(\lambda_n))T_nR_F^{-1}P_{0,0}, & F\le n\le K-2,\\ \rho_{n,n}^{-1}\Lambda_{0,n}(1-\gamma^{*}(\lambda_n))R_F^{-1}P_{0,0}, & n=K-1,\end{cases} \tag{5.13b}$$

$$P_{2,n} = \frac{\Lambda_{0,K-1}\varepsilon}{\mu R_F} P_{0,0},\ F \le n \le K-1, \tag{5.13c}$$

where $T_n = \varepsilon\Lambda_{n+2,K-1} + \varepsilon \sum_{i=n+3}^{K} \left(\Lambda_{i,K-1} \left(\prod_{j=n+2}^{i-1} \mu_j \gamma^*(\lambda_{j-1}) \right) \right) + \left(\prod_{j=n+2}^{K} \mu_j \gamma^*(\lambda_{j-1}) \right),$

$R_F = \mu_1 \left(\prod_{j=1}^{F-1} \mu_{j+1} \gamma^*(\lambda_j) \right) \left(\varepsilon\Lambda_{F+1,K-1} + \mu_{F+1}\gamma^*(\lambda_F) T_F \right)$ and

$$P_{0,0} = [1 + \rho_{0,n-1}\Lambda^*_{1,n-1}\left(\Lambda^*_{n,n} - 1\right) + \rho^{-1}_{n,n}\Lambda_{0,n}(1-\gamma^*(\lambda_n))T_n R_F^{-1} + \rho^{-1}_{n,n}\Lambda_{0,n}(1-\gamma^*(\lambda_n))R_F^{-1} + \rho_{0,n}\Lambda^*_{1,n} + \Lambda_{0,n}T_n R_F^{-1} + \Lambda_{0,n}R_F^{-1} + \Lambda_{0,K-1}\mu^{-1}R_F^{-1}\varepsilon]^{-1}.$$

Proof: Using (5.12), we get

$$P_{1,n} = \frac{1}{\mu_{n+1}\gamma^*(\lambda_n)} P_{0,n}(0) \tag{5.14a}$$

$$P_{1,n} = \frac{\lambda_n}{\mu_{n+1}(1-\gamma^*(\lambda_n))} P_{0,n} \tag{5.14b}$$

Applying recursive technique, equations (5.4)-(5.10) are solved to obtain expressions for the state probabilities as given in equations (5.13a-c). $P_{0,0}$ is obtained using (5.13a-c) in (5.11).

Remark 6.1: When the customers' discouragement behavior and admission control *F*-policy are not taken into account and $\mu_n = \mu,\ \lambda_n = (K-n)\lambda$, the results given in (5.13a) and (5.13b) coincide with the results obtained by Yang and Chang (2018).

$$P_{0,n} = \begin{cases} \dfrac{K\lambda}{\mu}\left(\dfrac{1}{\gamma^*((K-1)\lambda)} - 1\right) P_{0,0}, & n = 1, \\ \dfrac{K\lambda^n}{\mu^n}\left(\displaystyle\prod_{i=1}^{n-1} \frac{(K-i)}{\gamma^*((K-i)\lambda)}\right)\left(\dfrac{1}{\gamma^*((K-1)\lambda)} - 1\right) P_{0,0}, & 2 \le n \le K-1, \end{cases} \tag{5.15a}$$

$$P_{1,n} = \begin{cases} \dfrac{K\lambda}{\mu} P_{0,0}, & n = 0, \\ \dfrac{K\lambda^{n+1}}{\mu^{n+1}}\left(\displaystyle\prod_{i=1}^{n} \frac{(K-i)}{\gamma^*((K-i)\lambda)}\right) P_{0,0}, & 1 \le n \le K-1. \end{cases} \tag{5.15b}$$

Remark 6.2: For general retrial queueing model, Laplace-Stieltjes transforms $\gamma^*(\theta)$ of some specific probability distributions $\gamma(x)$ for general retrial time are given by

$$\gamma^*(\theta)=\begin{cases}\left(\dfrac{k\gamma}{\theta+k\gamma}\right)^k, & \text{for k-phase Erlang distribution,}\\ \left(\dfrac{\gamma}{\theta+\gamma}\right)^k, & \text{for gamma distribution,}\\ p\dfrac{\gamma}{\theta+\gamma}+(1-p)\dfrac{\gamma^2}{\theta+\gamma^2},\ p\in[0,1], & \text{for hyper}-\text{exponential distribution.}\end{cases}$$

Exponential and deterministic distributions are particular cases of the $k-$phase Erlang distribution (E_k) and are obtained when $k=1$ and when $k\to\infty$, respectively.

5.4. PERFORMANCE MEASURES AND COST FUNCTION

To explore the proposed model, various system indices and cost function have been derived which can be further utilized to resolve the queueing problems of the real time system.

5.4.1. Performance Indices

The following system indices are established in terms of probabilities:

(i) The average number of customers in the system and in the queue respectively, are

(a) $$E[N_S]=\sum_{n=0}^{K-1} nP_{0,n}+\sum_{n=0}^{K-1}(n+1)P_{1,n}+\sum_{n=F}^{K-1}(n+1)P_{2,n} \tag{5.16a}$$

(b) $$E[N_q]=\sum_{n=0}^{K-1} nP_{0,n}+\sum_{n=0}^{K-1} nP_{1,n}+\sum_{n=F}^{K-1} nP_{2,n} \tag{5.16b}$$

(ii) System throughput is

$$TP=\sum_{n=0}^{K-1}\mu_{n+1}P_{1,n}+\mu\sum_{n=F}^{K-1}P_{2,n} \tag{5.17}$$

(iii) The probability of the server being idle and busy respectively, are

(a) $$P(S_I)=\sum_{n=0}^{K-1}P_{0,n} \tag{5.18}$$

(b) $$P(S_B)=\sum_{n=0}^{K-1}P_{1,n}+\sum_{n=F}^{K-1}P_{2,n} \tag{5.19}$$

(iv) Average balking rate, average reneging rate and the average rate of customers loss respectively, are

(a) $B_{avg} = \sum_{n=0}^{K-1} \lambda(1-\beta_n)P_{0,n} + \sum_{n=0}^{K-2} \lambda(1-\beta_{n+1})P_{1,n}$ (5.20)

(b) $R_{avg} = \sum_{n=1}^{K-1} n\upsilon P_{1,n}$ (5.21)

(c) $CL = B_{avg} + R_{avg}$ (5.22)

5.4.2. Cost Function

One of the key concern of analyzing the queueing system operating under the admission control policy is to run the system at minimum cost so that the maximum profit can be earned. In any system, the organizers would like to provide faster service to the customers but it may increase the total system cost. To tradeoff between the excessive cost due to facilitating better service and inconvenience due to longer waiting, we compute the optimal service rate and corresponding expected minimum cost. For this purpose, a cost function per unit of time which is composed of different cost elements per unit of time corresponding to different activities, is framed as:

$$TC(\mu) = C_I P(S_I) + C_B P(S_B) + C_H E[N_q] + \mu C_A \quad (5.23)$$

where different cost components in per unit time used are as follows:

$C_I(C_B)$: Cost incurred on the server during the idle state (busy state),
C_H : Holding cost of each customer residing in the system,
C_A: Cost for rendering the service to the customer.

5.5. COST OPTIMIZATION

To determine the optimal service rate (μ^*) and corresponding minimum total cost $TC(\mu^*)$, we employ quasi-Newton method (QNM). The genetic algorithm (GA) is also used for the cost optimization.

5.5.1. Quasi-Newton Method (QNM)

For the general retrial queueing model with customers' discouragement, we apply QNM to find the optimal service rate (μ^*) and minimum cost $TC(\mu^*)$. The algorithmic steps for the QNM can be seen in Section 1.4.8 of Chapter 1. The following input parameters are used for QNM to determine the minimum cost.

Inputs: K, F, λ, γ, ε, β, α, υ, C_I, C_B, C_H, C_A, tolerance $\varepsilon_0 = 10\text{E}-06$ and the initial trial solution μ_0 for μ.

Output: Approximate optimal service rate (μ^*) and $TC(\mu^*)$.

5.5.2. Genetic Algorithm (GA)

For the queueing model, we apply the GA (cf. Section 1.4.8 (ii), Chapter 1) to evaluate optimal service parameter (μ^*) by using input parameters K, F, λ, γ, ε, β, α, υ, C_I, C_B, C_H and C_A, population size Y_μ, genes X_μ and probabilities of crossover and mutation. The algorithm steps for implementing GA are briefly given below.

Step 1: ***Initialization of population.*** Consider the fixed size Y_μ of the population which is represented by genes (X_μ) of chromosomes. Use binary encoding in which every chromosome is a string of bits 1 or 0.

Step 2: ***Evaluation of fitness function***. Compute the fitness function (cost function) at each chromosome.

Step 3: ***Selection.*** Select the fittest chromosomes (fittest parents) from the Y_μ population. We select $Y_\mu/5$ population from the entire population for the next generation using tournament selection.

Step 4: ***Crossover.*** To produce the next generation, the crossover of the selected chromosomes is done. Two point crossover method is considered and crossover probability is taken as P_C.

Step 5: ***Mutation***. It is done to maintain diversity in the population. The bit inversion mutation is adopted and the mutation rate is selected as M_R.

Step 6: ***Stopping criteria.*** The stopping criteria of GA in terms of generations, stall generation and function tolerance.

5.6. APPLICATION EXAMPLE AND NUMERICAL RESULTS

The finite capacity retrial queueing system with discouraged customers and admission control has varied applications in the real life queueing scenarios including the patients waiting for treatment in the hospitals, failed computers/vehicles arriving in the repair shop, jobs arriving in the production system, etc. In order to illustrate the applicability of the model developed, we cite an example of a queue formed at the reception counter in the hospital. The patients arrive for the treatment to the hospital following Poisson process with rate λ. The arriving patient may be discouraged due to the crowd present at the reception counter of a hospital having finite waiting space and decides not to join the queue for treatment and balks from the hospital with probability $(1-\beta)$. The receptionist at the counter allocates the doctor to the patients as per need of treatment according to the exponential distribution with rate μ. After waiting for a certain period of time in the queue, the patients may get impatient and renege without being getting the appointment according to exponential distribution with rate υ. If the arriving patient finds the receptionist busy with another patient then he is shifted to the waiting hall available in the hospital and re-attempts for the appointment for the doctor after some time by following the exponential distribution with mean $1/\gamma$. Due to capacity constraint, as soon as the capacity (say K) is full, the arriving patients are not allowed to get appointments. Further, the patients are allowed for admission at the hospital when the number of patients for admission ceases to a predefined number 'F'. To understand the more practical situation, we consider exponential distribution for general retrial time and set the parameters as $K = 30$ patients, $F = 20$ patients, $\mu = 8$ patients/hour, $\lambda = 5$ patients/hour, $\gamma = 1$ patient/hour, $\varepsilon = 1$, $\upsilon = 0.5$, $\beta = 0.8$. For these fix parameter values, using Equation (5.16a), we find the average number of patients present in the reception area of the hospital as $21.45 \approx 22$.

5.6.1. Sensitivity Analysis

This section explores the effect of system descriptors μ, λ and γ on various performance measures derived in Section 5.4.1. In order to analyze the impact of the system parameters on different performance measures, we illustrate a numerical example. We consider two distributions, namely exponential distribution (Exp) and 3-stage Erlang distribution (E_3) for the general retrial times.

$$\gamma^*(\theta) = \begin{cases} \dfrac{\gamma}{\gamma+\theta}, & \text{for exponential distribution,} \\ \left(\dfrac{3\gamma}{3\gamma+\theta}\right)^3, & \text{for 3-stage Erlang distribution.} \end{cases}$$

The following two balking functions namely, constant balking (CB) and exponential balking (EB) are taken into account to compute the numerical results.

$$\beta_n = \begin{cases} \beta, & \text{constant balking,} \\ e^{-\alpha n}, & \text{exponential balking.} \end{cases}$$

The MATLAB software is used to perform a numerical experiment by assigning the default values to the system parameters as $K = 25,\ F = 15,\ \mu = 8,\ \lambda = 3,\ \gamma = 1,\ \varepsilon = 1,\ \upsilon = 0.5,$ $\beta = 0.6,\ \alpha = 0.05,\ C_I = 20,\ C_B = 30,\ C_H = 100,\ C_F = 35,\ C_A = 30.$

The numerical results for constant balking and exponential balking are displayed in Table 5.1 and Table 5.2, respectively. The trends of $E[N_S]$ by varying parameters μ, λ and γ for both constant and exponential balking cases and exponential as well as Erlang-3 distributions are depicted in Figures 5.3(a-c).

The following observations are made after performing numerical experiments by varying the different system parameters:

- *Effect of service rate* (μ): In case, when the system organizers provide service to the customers with faster rate *i.e.* when μ is higher, $E[N_S]$, $E[N_q]$ and CL become less. The numerical results for constant balking (CB) and for exponential balking (EB) are summarized in Tables 5.1 and 5.2 which reveal the similar pattern of $E[N_q]$ and CL for varying values of μ for both Exp and E_3 distributions for the retrial time. From Figure 5.3(a), we notice that when the customers follow constant balking, the number of customers decreases rapidly at the initial stage as μ increases upto 7 and then after as μ goes up, the decreasing effect diminishes before reaching to an asymptotically constant value. Furthermore, in case of both constant and exponential balking, $E[N_S]$ decreases with faster rate when retrial time follows Exp distribution in comparison to the model when the retrial time is E_3 distributed.
- *Effect of retrial rate* (γ): From Tables 5.1-5.2, it is noticed that as γ goes up, $E[N_q]$ lower down but CL goes up in all the cases of balking and retrial time distributions. It is observed in Figure 5.3(c) that $E[N_S]$ reduces gradually at the initial stage as γ enhances while later on it becomes asymptotically constant when γ goes beyond 2.5.

Table 5.1: Performance measures for $\beta_n = \beta$ by varying values of (μ, λ, γ)

(μ,λ,γ)	*Exp*			E_3		
	$E[N_q]$	CL	TC	$E[N_q]$	CL	TC
(1,3,1)	8.75	2.65	963.89	13.58	2.75	1445.06
(2,3,1)	6.76	2.30	829.17	11.61	2.52	1313.19
(3,3,1)	4.93	2.04	710.99	9.59	2.28	1176.14
(4,3,1)	3.46	1.90	629.49	7.63	2.06	1044.92
(5,3,1)	2.44	1.87	591.93	5.86	1.89	933.58
(8,1,1)	0.06	0.93	546.70	0.07	0.93	547.91
(8,2,1)	0.31	1.65	572.59	0.48	1.56	589.57
(8,3,1)	1.05	2.06	646.91	2.57	1.80	798.82
(8,4,1)	3.27	2.34	869.42	12.53	2.62	1794.80
(8,5,1)	9.02	3.08	1443.98	18.72	3.51	2413.57
(8,3,1)	1.05	2.06	646.91	2.57	1.80	798.82
(8,3,1.5)	0.62	2.23	603.86	1.00	2.08	642.31
(8,3,2)	0.45	2.32	586.73	0.62	2.23	603.68
(8,3,2.5)	0.35	2.38	577.60	0.45	2.32	587.08
(8,3,3)	0.30	2.42	571.94	0.36	2.38	577.99

Table 5.2: Performance measures for $\beta_n = e^{-\alpha n}$ by varying values of (μ, λ, γ)

(μ,λ,γ)	*Exp*			E_3		
	$E[N_q]$	CL	TC	$E[N_q]$	CL	TC
(1,3,1)	9.19	2.65	1007.63	11.32	2.74	1219.61
(2,3,1)	8.16	2.32	969.13	10.48	2.50	1200.49
(3,3,1)	7.17	2.02	935.56	9.69	2.28	1186.64
(4,3,1)	6.24	1.75	907.31	8.95	2.08	1177.54
(5,3,1)	5.38	1.52	885.58	8.26	1.89	1172.77
(8,1,1)	0.17	0.79	558.52	0.23	0.76	564.50
(8,2,1)	1.03	1.00	645.18	1.96	0.79	738.00
(8,3,1)	3.32	1.13	874.40	6.40	1.43	1181.97
(8,4,1)	6.55	1.95	1197.99	10.62	2.57	1604.24
(8,5,1)	9.55	3.03	1497.54	14.14	3.68	1955.18
(8,3,1)	3.32	1.13	874.40	6.40	1.43	1181.97
(8,3,1.5)	2.06	1.18	748.85	3.63	1.13	905.19
(8,3,2)	1.48	1.29	691.22	2.32	1.14	774.53
(8,3,2.5)	1.16	1.40	659.48	1.66	1.24	708.71
(8,3,3)	0.96	1.48	639.71	1.28	1.35	671.57

- *Effect of arrival rate* (λ): Now we examine the case when the system organizer serves the customers with constant service rate and the arrival rate of the customers increases, *i.e.* when the system becomes more crowded. Tables 5.1 and 5.2 show that $E[N_q]$ and CL increase as the arrival rate of the customers, goes up in all the cases. The trends shown in Figure 5.3(b) depict that in case of constant balking (CB) as λ starts increasing, initially $E[N_S]$ increases with slower rate and later on steep increment is reported before reaching an asymptotically constant value as λ crosses 8. When the customers follow exponential balking (EB) behavior, $E[N_S]$ increases linearly as λ goes up for both retrial time distributions (Exp and E_3); however, $E[N_S]$ is always less for Exp than E_3 distribution.

5.6.2. ANFIS Computing and Results

Now, we compare the numerical results obtained by analytical formulae and ANFIS approach. Using the *evalfis* function in MATLAB software, ANFIS results are generated for default system parameter values given in Section 5.6.1. The membership functions corresponding to input parameters (i) μ (ii) λ (iii) γ are selected as Gaussian function and the linguistic variables for input parameters are defined as given in Table 5.3.

Table 5.3: Input parameters, membership function and linguistic variables for ANFIS

Sr. No.	Input parameter	Number of membership function	Linguistic variables
1.	μ	3	Small; medium; large
2.	λ	3	Small; medium; large
3.	γ	3	Small; medium; large

The absolute percentage error (Δ) is evaluated by the formula

$$\Delta = \frac{|I\{E[N_S]\} - M\{E[N_S]\}|}{M\{E[N_S]\}} \times 100\% \tag{5.24}$$

where $\Delta =$ absolute percentage error, $I\{E[N_S]\} =$ estimated value of $E[N_S]$ by ANFIS, $M\{E[N_S]\} =$ computed value of analytical formula for $E[N_S]$.

The absolute percentage errors (Δ) and accuracy of the estimated value in percentage of $E[N_S]$ are also summarized in Tables 5.4-5.6. The low values of Δ represent how ANFIS approach facilitates good approximate results. The graph for the membership function for the input parameter (i) μ (ii) λ (iii) γ are shown in Figures 5.2(a-c), respectively.

From Tables 5.4-5.6, it is recorded that the computed values of $E[N_S]$ by analytical formula (5.16a) and the estimated values of $E[N_S]$ by ANFIS are very close which authenticate the feasibility of ANFIS in the performance evaluation of real time complex system. The ANFIS results in Figures 5.3(a-c) are shown by tick marks and the analytical results are represented by continuous lines and discrete lines for the prediction of $E[N_S]$. From these Figures, we notice that both results have a good match as tick marks can be seen almost over the curved lines.

Table 5.4: Absolute percentage error (Δ) and $I\{E[N_S]\}$ obtained by ANFIS by varying μ

μ	Constant balking $\beta_n = \beta$						Exponential balking $\beta_n = e^{-\alpha n}$					
	Exp			E_3			*Exp*			E_3		
	M{E[N_S]}	$I\{E[N_S]\}$	Δ	$I\{E[N_S]\}$	$M\{E[N_S]\}$	Δ	M{E[N_S]}	$I\{E[N_S]\}$	Δ	$I\{E[N_S]\}$	M{E[N_S]}	Δ
1	9.110	9.072	0.416	13.822	13.796	0.191	9.541	9.527	0.147	11.583	11.574	0.078
2	7.111	7.150	0.551	11.851	11.882	0.265	8.496	8.510	0.167	10.732	10.740	0.080
3	5.269	5.225	0.847	9.830	9.798	0.323	7.500	7.485	0.203	9.933	9.924	0.096
4	3.784	3.796	0.325	7.866	7.864	0.022	6.557	6.559	0.021	9.183	9.185	0.020
5	2.732	2.766	1.250	6.097	6.140	0.697	5.680	5.696	0.287	8.477	8.486	0.099
6	2.033	1.993	1.967	4.648	4.610	0.815	4.883	4.869	0.300	7.812	7.804	0.103
7	1.571	1.572	0.085	3.553	3.542	0.291	4.182	4.177	0.138	7.186	7.183	0.039
8	1.258	1.290	2.554	2.764	2.808	1.600	3.582	3.601	0.533	6.597	6.607	0.149
9	1.039	1.019	1.914	2.202	2.181	0.954	3.079	3.072	0.234	6.047	6.044	0.052
10	0.879	0.866	1.471	1.800	1.777	1.261	2.663	2.651	0.475	5.536	5.528	0.136
11	0.759	0.782	3.003	1.505	1.537	2.148	2.321	2.337	0.661	5.063	5.072	0.161
12	0.666	0.662	0.590	1.284	1.281	0.241	2.040	2.041	0.012	4.630	4.632	0.027
13	0.592	0.578	2.464	1.115	1.092	2.060	1.808	1.795	0.719	4.236	4.228	0.197
14	0.533	0.548	2.859	0.982	1.005	2.343	1.616	1.628	0.765	3.879	3.887	0.193
15	0.484	0.470	2.865	0.875	0.854	2.421	1.455	1.443	0.808	3.558	3.550	0.211
Average of Δ			1.544		1.042			0.365			0.109	
Accuracy in predicted value (%)			98.456		98.958			99.635			99.891	

Table 5.5: Absolute percentage error (Δ) and $I\{E[N_S]\}$ obtained by ANFIS by varying γ

γ	Constant balking $\beta_n = \beta$						Exponential balking $\beta_n = e^{-\alpha n}$					
	Exp			E_3			*Exp*			E_3		
	M{E[N_S]}	$I\{E[N_S]\}$	Δ	$I\{E[N_S]\}$	M{E[N_S]}	Δ	M{E[N_S]}	$I\{E[N_S]\}$	Δ	$I\{E[N_S]\}$	M{E[N_S]}	Δ
1	1.258	1.258	0.000	2.764	2.764	0.001	3.582	3.582	0.000	6.597	6.597	0.000
1.5	0.832	0.832	0.001	1.213	1.213	0.002	2.356	2.356	0.001	3.883	3.883	0.001
2	0.662	0.662	0.001	0.830	0.830	0.003	1.795	1.795	0.001	2.607	2.607	0.002
2.5	0.572	0.572	0.001	0.666	0.666	0.004	1.486	1.486	0.001	1.965	1.965	0.002
3	0.516	0.516	0.001	0.576	0.576	0.002	1.294	1.294	0.001	1.604	1.604	0.001
Average of Δ			0.001		0.002			0.001			0.001	
Accuracy in predicted value (%)			99.999		99.998			99.999			99.999	

Table 5.6: Absolute percentage error (Δ) and $I\{E[N_S]\}$ obtained by ANFIS by varying λ

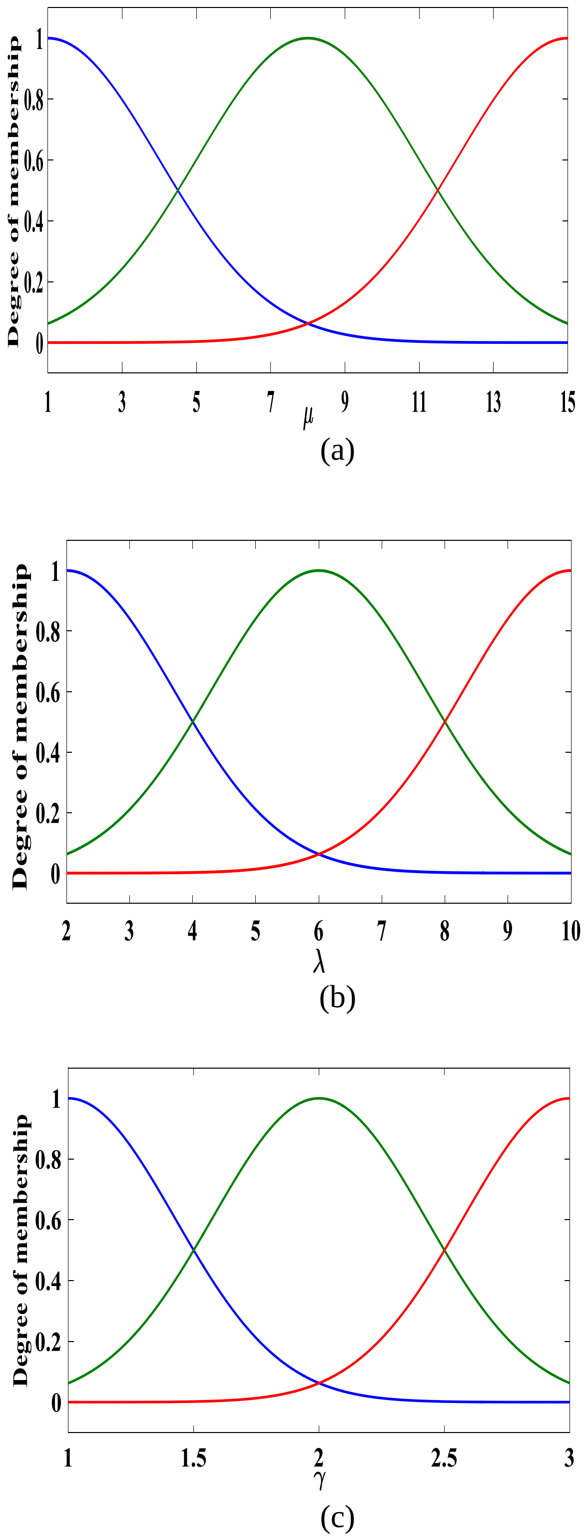
λ	Constant balking $\beta_n = \beta$						Exponential balking $\beta_n = e^{-\alpha n}$					
	Exp			E_3			*Exp*			E_3		
	M{E[N_S]}	$I\{E[N_S]\}$	Δ	$I\{E[N_S]\}$	M{E[N_S]}	Δ	M{E[N_S]}	$I\{E[N_S]\}$	Δ	$I\{E[N_S]\}$	M{E[N_S]}	Δ
2	0.457	0.434	5.069	0.626	0.666	6.354	1.250	1.227	1.847	2.162	2.124	1.739
3	1.258	1.234	1.934	2.764	2.647	4.230	3.582	3.567	0.417	6.597	6.579	0.281
4	3.520	3.539	0.527	12.704	12.350	2.782	6.814	6.814	0.008	10.804	10.791	0.116
5	9.261	9.294	0.354	18.895	19.304	2.162	9.797	9.808	0.111	14.302	14.330	0.202
6	15.982	15.910	0.445	20.011	19.758	1.262	12.391	12.374	0.131	17.135	17.099	0.213
7	18.797	18.860	0.338	20.449	20.607	0.772	14.659	14.678	0.131	19.128	19.166	0.198
8	19.838	19.796	0.212	20.713	20.624	0.427	16.575	16.556	0.110	20.255	20.226	0.146
9	20.356	20.374	0.088	20.902	20.932	0.143	18.093	18.105	0.068	20.818	20.832	0.068
10	20.675	20.734	0.285	21.052	21.212	0.760	19.207	19.239	0.165	21.105	21.148	0.207
Average of Δ			0.826		1.907			0.246			0.249	
Accuracy in predicted value (%)			99.174		98.093			99.754			99.751	

5.6.3. Cost Analysis using QNM and GA

In the finite queueing model, the customers always demand faster service to avoid waiting time in the system but it enhances the system cost also. The objective of the cost analysis in the present section is to evaluate the optimal service rate so as to minimize the total cost. It is seen that the system capacity (K) has an upper bound, which indicates that the expected minimum cost will be computed in the desired feasible region. In order to evaluate the optimal service rate (μ^*) and corresponding minimum cost $TC(\mu^*)$, we set the default values of the system parameters as $K = 25,\ F = 15,\ \mu = 8,\ \lambda = 8,\ \gamma = 1,\ \varepsilon = 1,\ \upsilon = 0.5,\ \beta = 0.6,\ \alpha = 0.05$. The three cost sets for the numerical experiment are taken into consideration as given in Table 5.7.

Table 5.7: Cost sets with different cost elements (in $)

Cost set	C_H	C_A	C_I	C_B
I	100	65	20	30
II	110	55	20	30
III	100	65	50	50



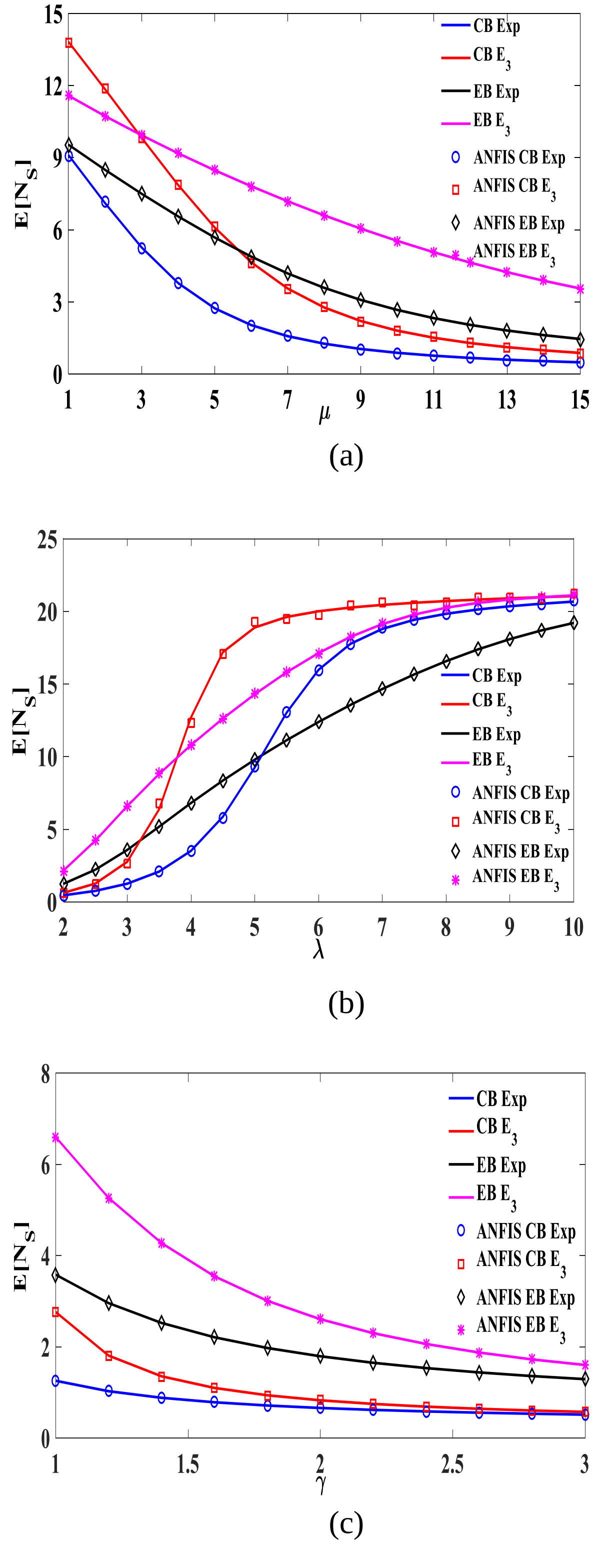


Figure 5.2: Membership function for (a) μ (b) λ (c) γ

Figure 5.3: $E[N_S]$ for varying values of (a) μ (b) λ (c) γ

To achieve the goal of the optimal service rate and the minimum expected cost of the system, it is noticed that the cost function given in (5.23) is highly non-linear and complex and it is not an easy task to minimize it by using the analytical method. However, the search method namely, quasi-Newton method (QNM) and genetic algorithm (GA) can be applied for the evaluation of the optimal values of the service rate (μ^*).

The quasi-Newton method (QNM) is used for three cost sets given in Table 5.7 with initial solution $\mu_0 = 3$ and $\mu_0 = 4$ of μ for Exp retrial time and E_3 retrial time, respectively. The values of μ^* and corresponding minimum cost of the system are gathered in Tables 5.9-5.12 by taking $\beta_n = \beta$ and $\beta_n = e^{-\alpha n}$ for different values of other parameters namely β, α and γ. From Figures 5.4(a-c)-5.7(a-c), it is noted that as the constant (exponential) balking parameter β (α) increases (decreases), the cost of the system enhances. On the other hand, it is also observed that the expected cost function is unimodal and convex in the feasible range of μ; the minimum cost is achieved in the desired search space for different balking parameters.

To achieve the optimal service rate (μ^*) and the corresponding minimum expected cost $TC(\mu^*)$ of the system, metaheuristic algorithm GA is successfully implemented. To obtain the minimum cost of the system for constant balking (CB) and exponential balking (EB), the numerical results of μ^* and $TC(\mu^*)$ are evaluated using the computational steps mentioned in Section 5.5.2 by setting the input parameters as $K = 25,\ F = 15,\ \lambda = 3,\ \gamma = 1,\ \varepsilon = 1,\ \upsilon = 0.5,$ $\beta = 0.6,\ \alpha = 0.05,\ \Delta\mu = 0.01.$ The numerical results are generated for three different cost sets displayed in Table 5.7. GA parameters used to compute μ^* are shown in Table 5.8.

Repeating steps of GA from 2 to 5 given in Section 5.5.2, it is noted that after every 51 iterations, the weighted average change in the fitness function (cost function) value over stall generations is less than function tolerance. Therefore, the algorithm meets the stopping criteria and the approximate values of μ^* and $TC(\mu^*)$ are recorded in Tables 5.9-5.12. We notice that the optimal values obtained by GA almost coincide with the optimal values obtained by the QNM.

Table 5.8: GA parameters for computation of μ^* and $TC(\mu^*)$

Population size (Y_μ)	20
Probability of crossover (P_C)	1
Mutation rate (M_R)	0.08
Stopping criteria	
(i) Generations	100
(ii) Stall generation	50
(iii) Function tolerance	10E-06

Table 5.9: $(\mu^*, TC(\mu^*))$ by varying value of β for $\beta_n = \beta$ using QNM and GA

Distribution for general retrial times	Cost Set	Method	$\beta = 0.5$	$\beta = 0.6$	$\beta = 0.7$
Exp	I	QNM	(4.206, 466.6314)	(5.558, 587.1582)	(7.085, 718.9362)
		GA	(4.204, 466.6315)	(5.558, 587.1582)	(7.085, 718.9363)
	II	QNM	(4.714, 437.6154)	(6.162, 547.1549)	(7.785, 666.1737)
		GA	(4.711, 437.6155)	(6.158, 547.1551)	(7.781, 666.1739)
	III	QNM	(4.193, 493.6604)	(5.547, 614.4117)	(7.076, 746.3861)
		GA	(4.190, 493.6606)	(5.553, 614.4146)	(7.083, 746.3868)
E_3	I	QNM	(5.597, 593.3789)	(7.993, 798.8203)	(11.029, 1048.7280)
		GA	(5.593, 593.3791)	(7.995, 798.8203)	(11.025, 1048.7278)
	II	QNM	(6.207, 553.3419)	(8.749, 738.6806)	(11.936, 962.3292)
		GA	(6.199, 553.3429)	(8.769, 738. 6847)	(11.934, 962.3292)
	III	QNM	(5.588, 621.1035)	(7.987, 826.8685)	(11.024, 1077.0490)
		GA	(5.587, 621.1036)	(7.989, 826.8686)	(11.019, 1077.0490)

Table 5.10: $(\mu^*, TC(\mu^*))$ by varying value of β for $\beta_n = e^{-\alpha n}$ using QNM and GA

Distribution for general retrial times	Cost Set	Method	$\alpha = 0.03$	$\alpha = 0.04$	$\alpha = 0.05$
Exp	I	QNM	(8.739, 969.9668)	(7.755, 916.1820)	(6.886, 868.0854)
		GA	(8.739, 969.9668)	(7.751, 916.1821)	(6.882, 868.0855)
	II	QNM	(10.067, 909.8519)	(9.175, 866.0991)	(8.389, 826.7751)
		GA	(10.070, 909.8519)	(9.178, 866.0991)	(8.386, 826.7751)
	III	QNM	(8.729, 997.4295)	(7.744, 943.5102)	(6.873, 895.2829)
		GA	(8.730, 997.4295)	(7.746, 943.5103)	(6.876, 895.2830)
E_3	I	QNM	(10.681, 1440.5781)	(7.699, 1292.4704)	(5.690, 1171.8401)
		GA	(10.659, 1440.5791)	(7.696, 1292.4704)	(5.689, 1171.8401)
	II	QNM	(14.158, 1376.7692)	(11.638, 1259.1987)	(9.636, 1160.3067)
		GA	(14.166, 1376.7693)	(11.650, 1259.1990)	(9.624, 1160.3070)
	III	QNM	(10.668, 1468.8665)	(7.680, 1320.5287)	(5.668, 1199.6807)
		GA	(10.666, 1468.8665)	(7.675, 1320. 5288)	(5.666, 1199.6807)

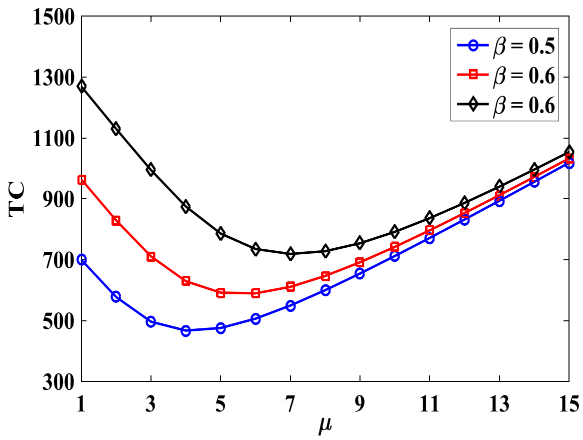
Table 5.11: $(\mu^*, TC(\mu^*))$ by varying value of γ for $\beta_n = \beta$ using QNM and GA

Distribution for general retrial times	Cost Set	Method	$\gamma=1$	$\gamma=2$	$\gamma=3$
Exp	I	QNM	(5.558, 587.1582)	(3.825, 427.3411)	(3.230, 369.5056)
		GA	(5.558, 587.1582)	(3.830, 427.3415)	(3.226, 369.5060)
	II	QNM	(6.162, 547. 1549)	(4.298, 400.8663)	(3.650, 347.4191)
		GA	(6.167, 547.1553)	(4.296, 400.8663)	(3.671, 347.4264)
	III	QNM	(5.547, 614.4117)	(3.809, 453.4553)	(3.210, 394.9680)
		GA	(5.548, 614.4118)	(3.795, 453.4589)	(3.210, 394.9680)
E_3	I	QNM	(7.993, 798.8203)	(4.404, 481.8967)	(3.486, 394.5844)
		GA	(7.992, 798.8203)	(4.405, 481.8967)	(3.486, 394.5844
	II	QNM	(8.749, 738.6806)	(4.923, 451.0108)	(3.929, 370.6347)
		GA	(8.745, 738.6807)	(4.920, 451.0109)	(3.930, 370.6347)
	III	QNM	(7.987, 826.8685)	(4.389, 508.4852)	(3.467, 420.3520)
		GA	(7.990, 826.8686)	(4.387, 508.4852)	(3.458, 420.3535)

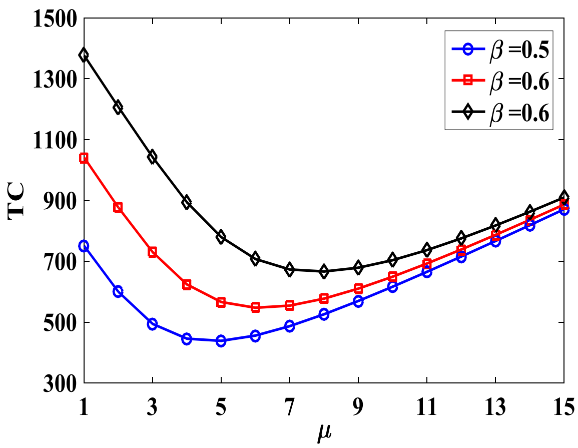
Table 5.12: $(\mu^*, TC(\mu^*))$ by varying value of γ for $\beta_n = e^{-\alpha n}$ using QNM and GA

Distribution for general retrial times	Cost Set	Method	$\gamma=1$	$\gamma=2$	$\gamma=3$
Exp	I	QNM	(6.886, 868.0854)	(5.242, 636.0440)	(4.522, 545.3628)
		GA	(6.895, 868.0858)	(5.239, 636.0441)	(4.525, 545.3629)
	II	QNM	(8.389, 826.7751)	(6.173, 603.3290)	(5.285, 516.7588)
		GA	(8.391, 826.7752)	(6.164, 603.3297)	(5.307, 516.7635)
	III	QNM	(6.873, 895.2929)	(5.225, 662.0901)	(4.502, 570.6971)
		GA	(6.866, 895.2832)	(5.224, 662.0901)	(4.504, 570.6971)
E_3	I	QNM	(5.690, 1171.8401)	(5.849, 747.2241)	(4.887, 600.7302)
		GA	(5.688, 1171.8401)	(5.841, 747.2245)	(4.884, 600.7303)
	II	QNM	(9.636, 1160.3067)	(7.101, 712.9964)	(5.773, 570.6814)
		GA	(9.638, 1160.3067)	(7.110, 712.9969)	(5.771, 570.6815)
	III	QNM	(5.668, 1199.6807)	(5.832, 773.8704)	(4.868, 626.4996)
		GA	(5.665, 1199.6807)	(5.831, 773.8704)	(4.867, 626.4996)

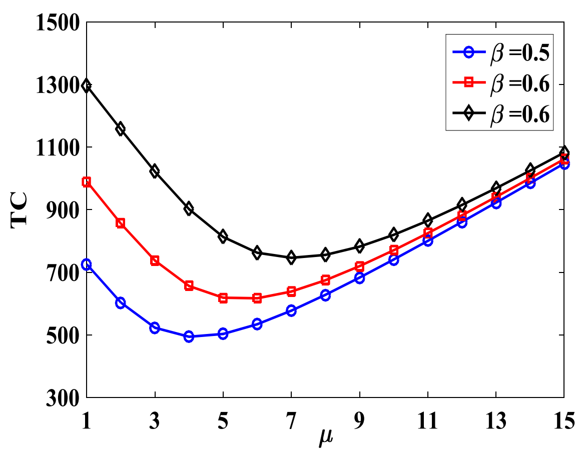
It can also be noticed that when customer re-tries from the orbit for the service as per exponential distribution, the minimum cost is less as compared to when he re-tries according to Erlang-3 distribution. Also, when the customer balks following constant balking (exponential balking) then system cost seems less (more). These facts can be seen in Tables 5.9-5.12. Surface graphs shown in Figures 5.8(a-c)-5.11(a-c) depict that the cost function is convex with respect μ in the feasible interval as such minimum cost of the system is achieved at μ^*.



(a) Cost Set-I

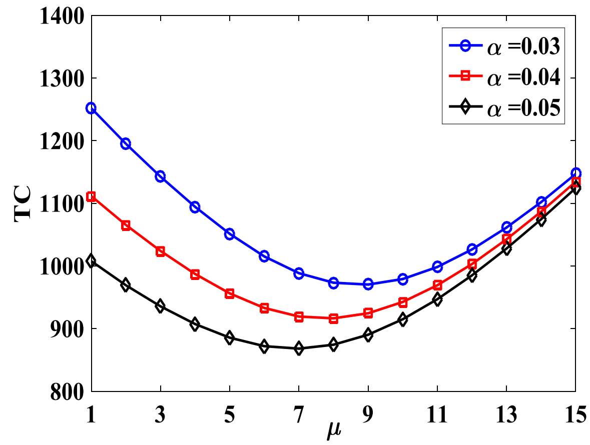


(b) Cost Set-II

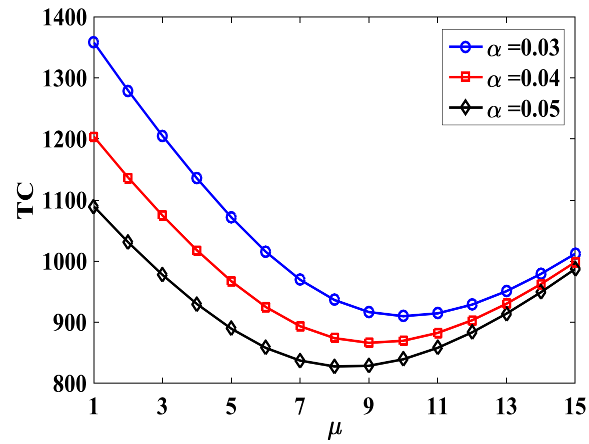


(c) Cost Set-III

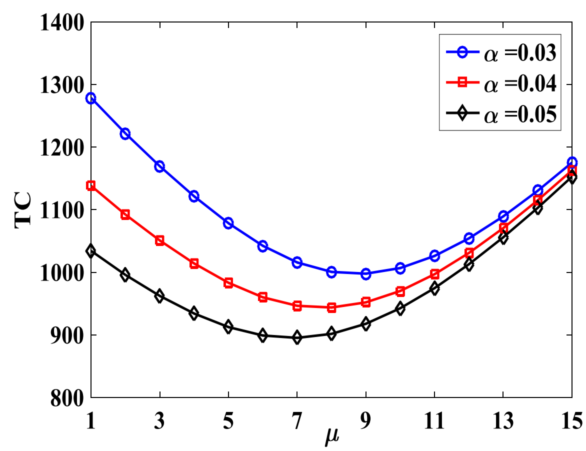
Figure 5.4: TC vs μ for $\beta_n = \beta$ and exponential retrial time



(a) Cost Set-I



(b) Cost Set-II



(c) Cost Set-III

Figure 5.5: TC vs μ for $\beta_n = e^{-\alpha n}$ and exponential retrial time

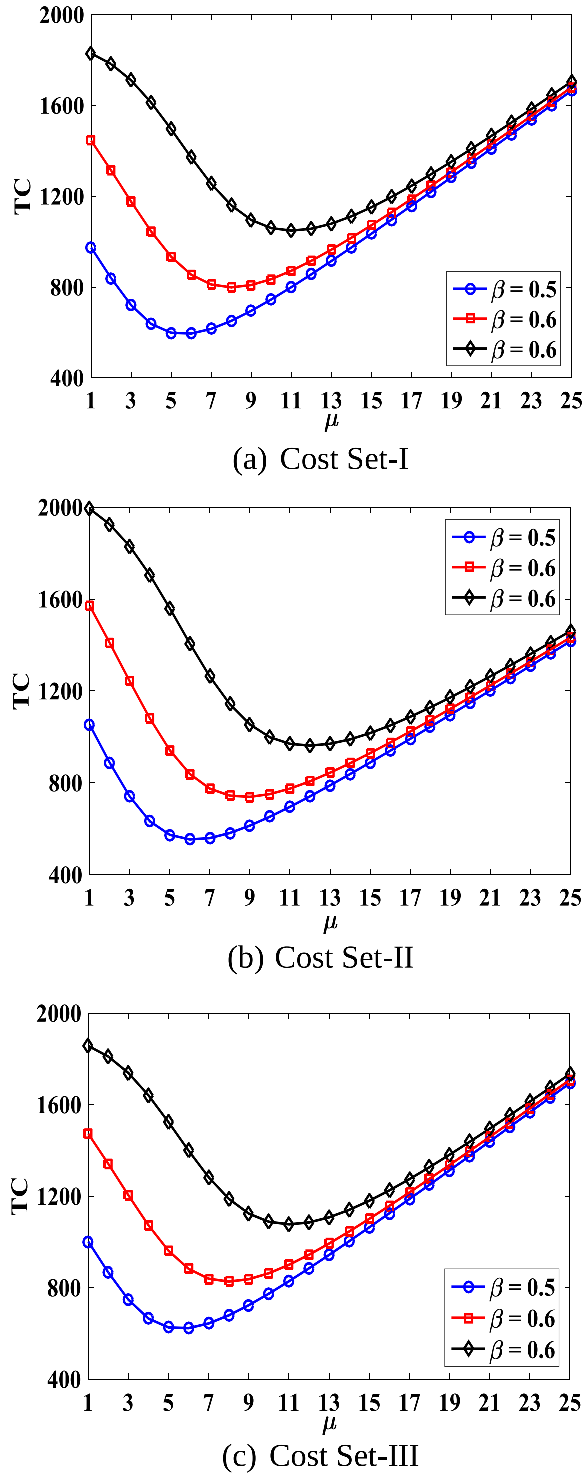


Figure 5.6: TC vs μ for $\beta_n = \beta$ and Erlang-3 retrial time

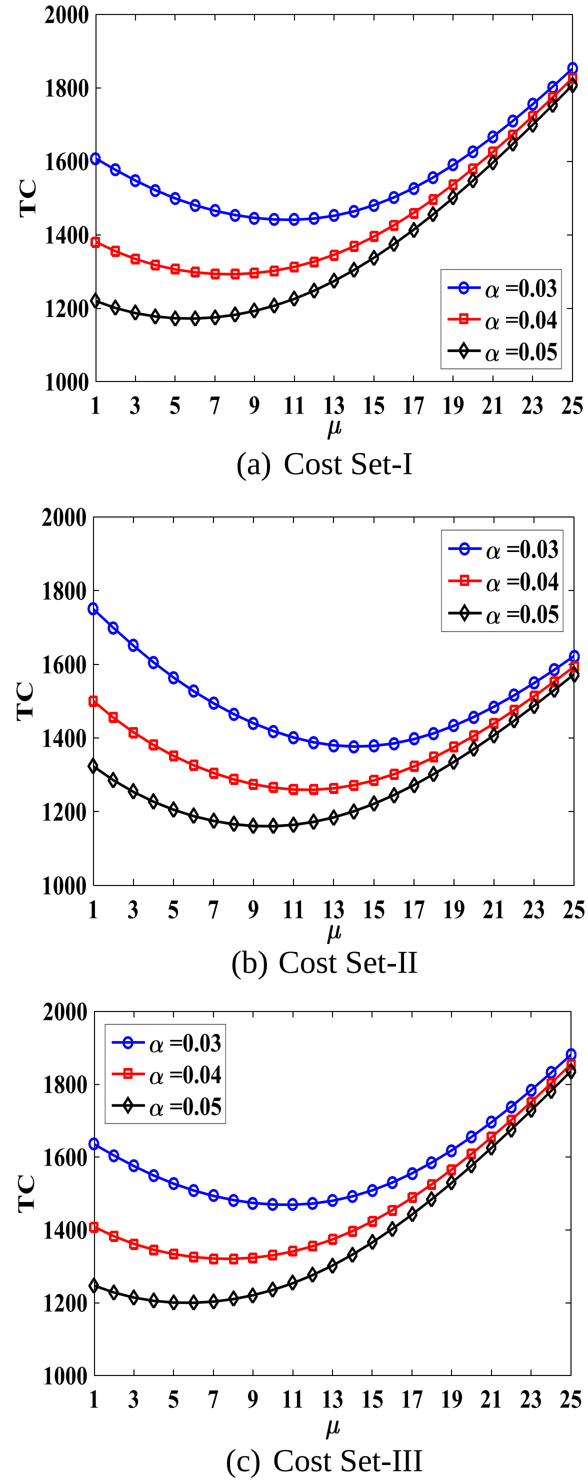
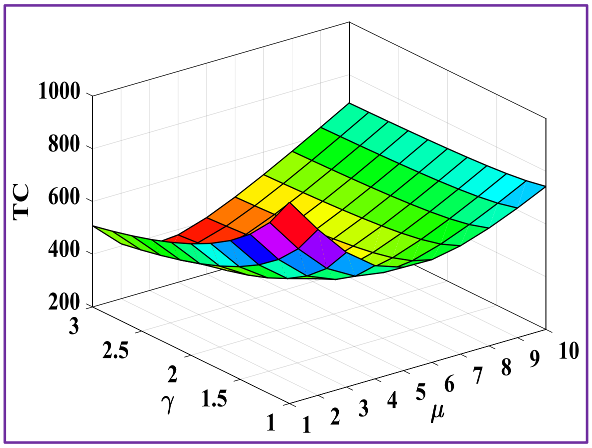
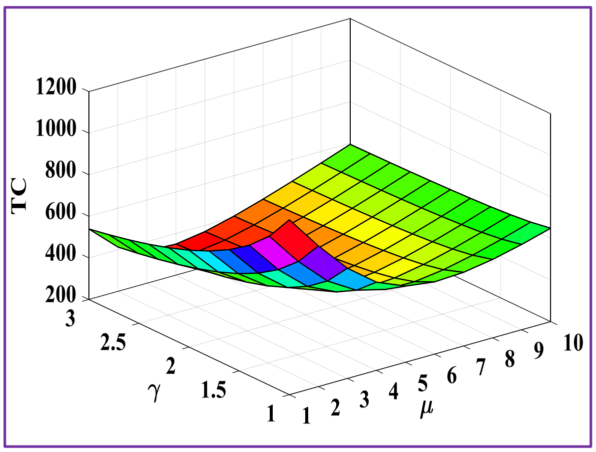


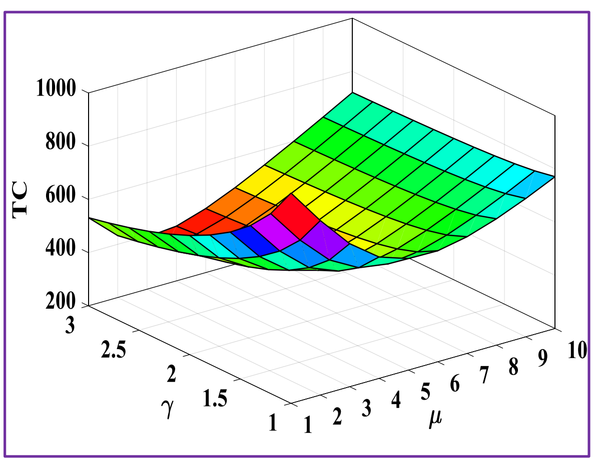
Figure 5.7: TC vs μ for $\beta_n = e^{-\alpha n}$ and Erlang-3 retrial time



(a) Cost Set-I

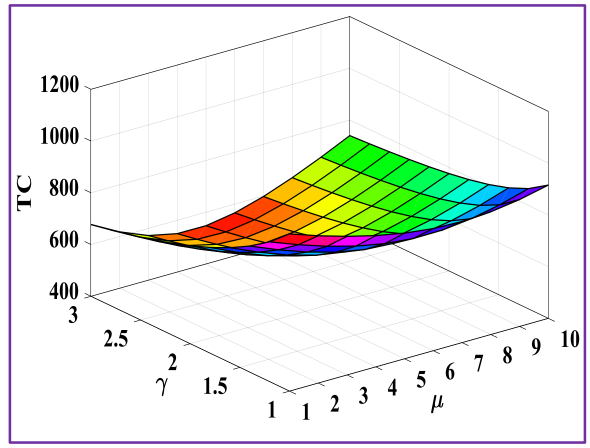


(b) Cost Set-II

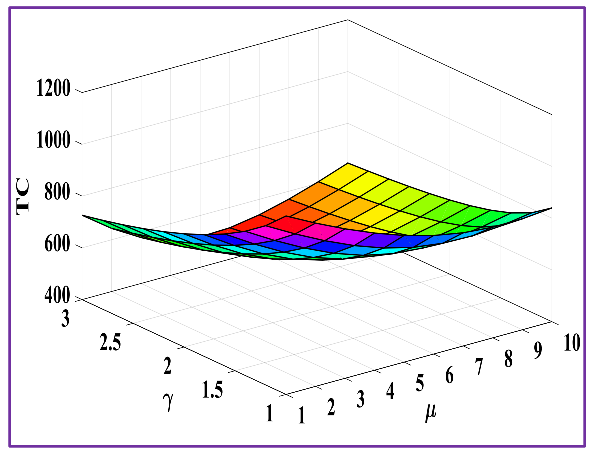


(c) Cost Set-III

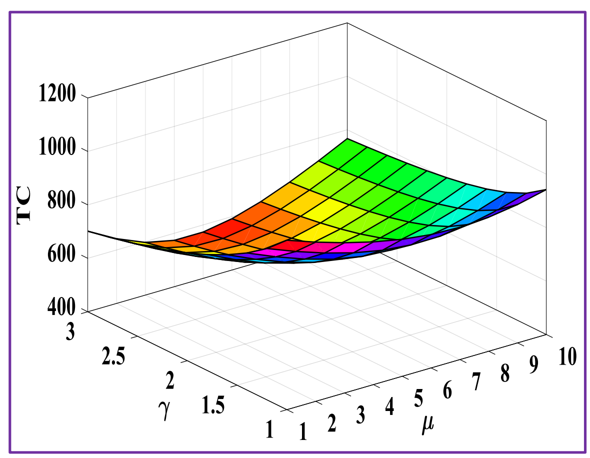
Figure 5.8: Total cost for varying values of μ and γ for $\beta_n = \beta$ and exponential retrial time



(a) Cost Set-I

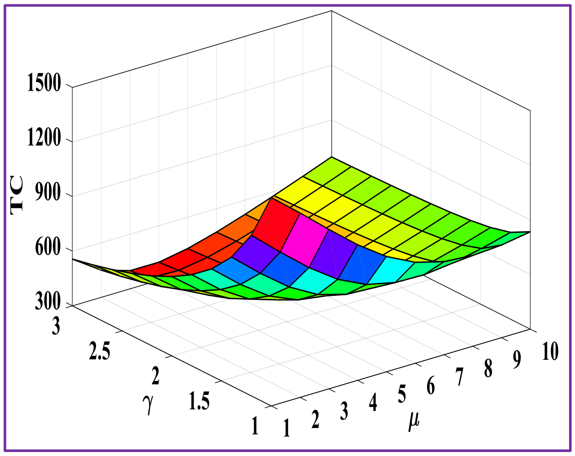


(b) Cost Set-II

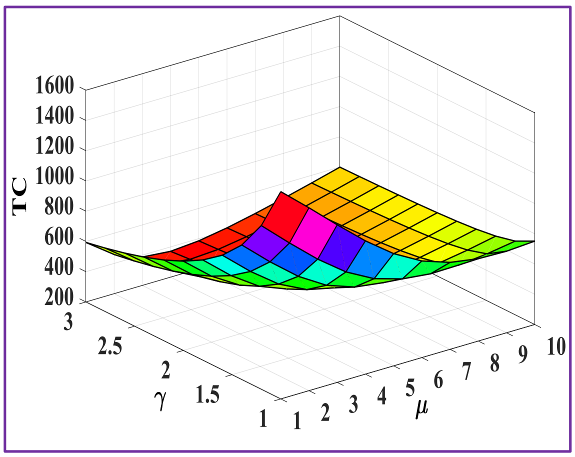


(c) Cost Set-III

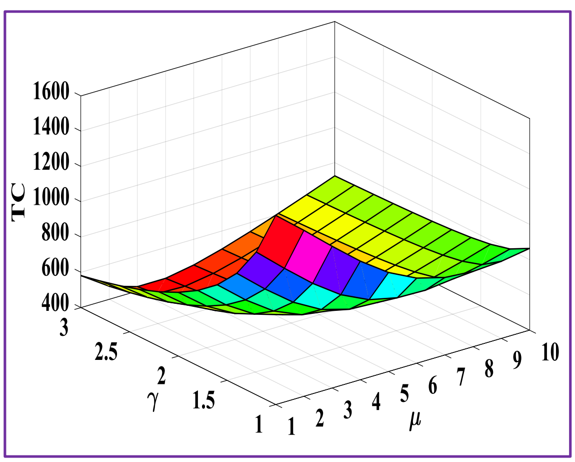
Figure 5.9: Total cost for varying values of μ and γ for $\beta_n = e^{-\alpha n}$ and exponential retrial time



(a) Cost Set-I

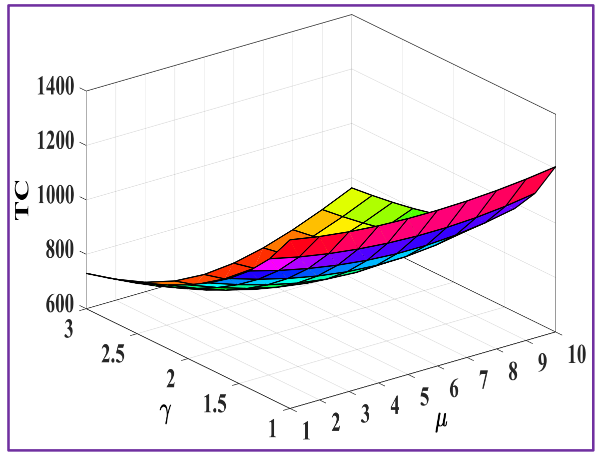


(b) Cost Set-II

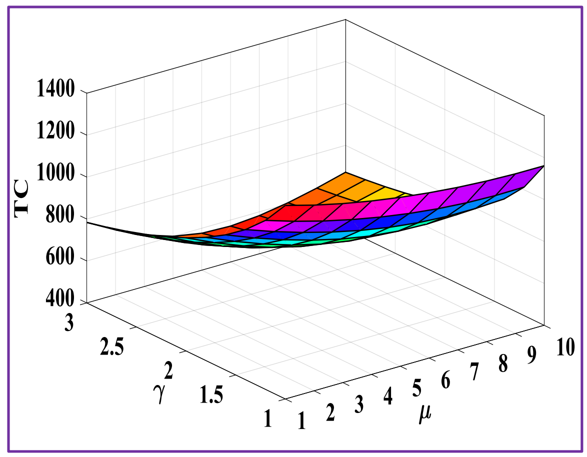


(c) Cost Set-III

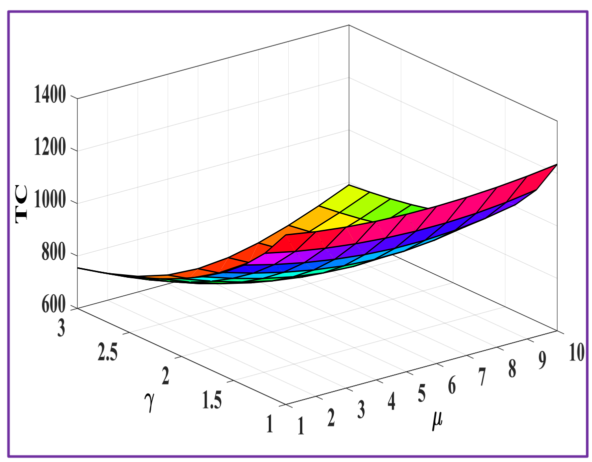
Figure 5.10: Total cost for varying values of μ and γ for $\beta_n = \beta$ and Erlang-3 retrial time



(a) Cost Set-I



(b) Cost Set-II



(c) Cost Set-III

Figure 5.11: Total cost for varying values of μ and γ for $\beta_n = e^{-\alpha n}$ and Erlang-3 retrial time

5.7. CONCLUDING REMARKS

The study done in this chapter on the admission control for M/M/1/K queueing model with general retrial attempt and discouraged customers is supported by implementing soft computing approach. The supplementary variable used corresponding to retrial time helps in providing the exact analytical results by using a recursive method. The exponential distribution and Erlang-3 distribution are considered for the general retrial times for illustration purpose. The incorporation of realistic features *i.e.* balking and reneging alongwith *F*-policy make our model closer to real world queueing problem seen at many day-to-day as well as industrial queueing problems. The sensitivity analysis conducted by taking a numerical illustration shows the effect of variation in parameters on the performance indices. The adaptive neuro fuzzy inference system is successfully implemented and authenticates the scope of the ANFIS technique for the complex queueing system for which analytical results cannot be established in closed form in particular when the network of queues are formed. The applicability of quasi-Newton method and genetic algorithm to optimize the cost function shown by implementing it, validates the scope of soft computing approaches for the future design of complex queueing systems for which analytical results are difficult to derive in explicit form.

Chapter 6

F-Policy for Machine Repair System with General Retrial Attempts

6.1. INTRODUCTION

With the advancement of modern technology, the life of the human being depends on machines to perform various activities. But due to unexpected failures of machines, there may not be only inconvenience to the users but also an undesirable loss of revenue along with goodwill of the concerned systems. Using recursive method, Gupta and Rao (1994) investigated machine interference problem with arbitrary repair to present the steady state results for the performance analysis. To analyze the unexpected breakdowns of the machines, many researchers working in the field of queueing theory have paid their attention towards machine repair problems (cf. Haque and Armstrong, 2007; Huang et al., 2016; Chen, 2018). In many finite population retrial queues, if the caretaker of the failed machine finds the repairman occupied, then he may leave the service area and can wait in the retrial orbit; from retrial orbit, after a random time duration he repeats his attempt for the service. To be more specific, one can notice the computer repair shop where failed computers join the system for the repair job; in the case when a repairman is busy, the failed computers may wait in the pool of blocked computers called retrial orbit and from there those failed computers seek for the repairing job and try again later. Whenever, the repairman becomes free, it takes the next failed computer for the repair, however, other failed computers may also try to get repair from the orbit. Due to enormous applications of retrial queueing problems in computer and communication systems, service and transportation systems, production and manufacturing systems, etc., many researchers contributed towards retrial queueing models in different frameworks.

F-policy can be used to restrict the failed machines from joining the system for repair when the system size is full. Jain et al. (2016) investigated the machine repair problem with working vacation and *F*-policy. The matrix method is used to establish the steady state probability distribution of the system size. Recently, a few researchers have paid their attention towards F-policy and presented a variety of works by considering F-policy (Kumar et al., 2019, Jain et al. 2019).

The system managers/organizers of the service systems may be interested to facilitate the service to their customers at a fast pace; however, the system may stop functioning due to the failure of machines. In order to reduce the congestion of failed machines in the system and to maintain the smooth functioning of the machining system, the managers/organizers of the concerned system should provide the immediate repair to the failed machines at optimal cost. Now-a-days, the industrial engineers design fault tolerant system (FTS) using optimal maintenance policy. In the queueing literature, a few researchers have paid their attention to the maintenance issues of the multi-component machining systems (cf. Chang et al., 2014b; Liou, 2014; Wang et al., 2014; Chen et al., 2016;) by developing cost models having parameters and cost elements as crisp. In recent years, some attempts have also been made to study the fuzzy queue in which arrival and/or service parameters were considered as fuzzy numbers (cf. Bagherinejad and Pishkenari, 2016; Ke et al., 2007; Mueen et al., 2017). In the crisp environment, the cost incurred on various activities are considered to be crisp (cf. Tandra et al., 2004). Kapur and Sachdeva (2016) dealt with the optimization problem to analyze the software faults of different severity to achieve desired system reliability. Kapur et al. (2017) also formulated optimization problem based on Bass diffusion model to depict the amount in cost-saving in manufacturing process. Recently, Bagyam and Chandrika (2019) studied single server retrial queue with admission control and vacation in which the customers join the queue in batch. Using Zadeh's extension principle they analyzed the model in environment.

In real time system, the cost elements do not always have definite values. Therefore, we need to analyze the cost function of the finite population model in a fuzzy environment. In some organizations, the cost factors associated with the total cost are linguistic variables and may be both probabilistic and possibilistic. The fuzzy cost function gives more robust design in comparison to the traditionally used crisp cost function. The same in the case with machine interference system for which finite population model can be used to deal with a variety of applications in real time machining system by considering fuzzy cost factors. The system organizers/managers should have an idea of fuzzified system cost, in particular when the cost function is analyzed for machining system operating in the fuzzy economic environment. The process in which the crisp value extracted from the fuzzy set, is often known as *defuzzification*. The fuzzified cost output is also to be converted into crisp cost output by choosing the suitable method of defuzzification (cf. Leekwijck and Kerre, 1999; Roychowdhury and Pedrycz, 2001). Some authors have used the signed distance method to defuzzify the fuzzy set (cf. Abbasbandy et al., 2013; Jaggi et al., 2016; Sahoo, 2017). The fuzzy cost function can be transformed into crisp equivalent cost function by using signed distance method (cf. Yao and Wu, 2000) of defuzzification of trapezoidal fuzzy numbers.

The objective of this chapter is to study the machine repair problem with the provision of retrial orbit. In the generic setup of general retrial times and admission control *F*-policy, the present study present the cost analysis in the fuzzy environment for machine repair problem by constructing the cost function and associated cost elements to be trapezoidal fuzzy numbers. The contents of this chapter along with ongoing section are arranged in different sections. The machine interference problem is described in Section 6.2. Chapman-Kolmogorov equations and steady-state results of the concerned model are given in Section 6.3. The various system metrics are established in Section 6.4. An application of the concerned study is given in Section 6.5. Section 6.6 is devoted to numerical results of the investigation done. In Section 6.7, the cost function of the concerned model is analyzed in the fuzzy environment. The cost optimization is done in Section 6.8 using a genetic algorithm. Section 6.9 concludes the whole investigation done in this chapter.

6.2. MODEL DESCRIPTION

Consider the finite population single repairman machine interference system with retrial attempts under admission control of failed machines. The system consists of '*K*' identical machines. We assume that the machines are subject to failure. The transition flow diagram of MIP is depicted in Figure 6.1. The following assumptions are made to formulate the machine interference problem (MIP):

- The failed machine joins the repair station by following a Poisson distribution with rate λ.
- The effective arrival rate λ_n for machine inference problem is defined by $\lambda_n = (K-n)\lambda,$ where *K* denotes the population size of the machines.
- If the caretaker of the failed machine finds the repairman free then the failed machine is repaired by following an exponential distribution with mean $1/\mu$.
- On an arriving of the caretaker of failed machines in the system, if the repairman seems busy then the failed machines shifted to the retrial orbit of finite capacity '*F*'. Later on, after some duration, the failed machine can be made to re-attempt for the repair job by following general

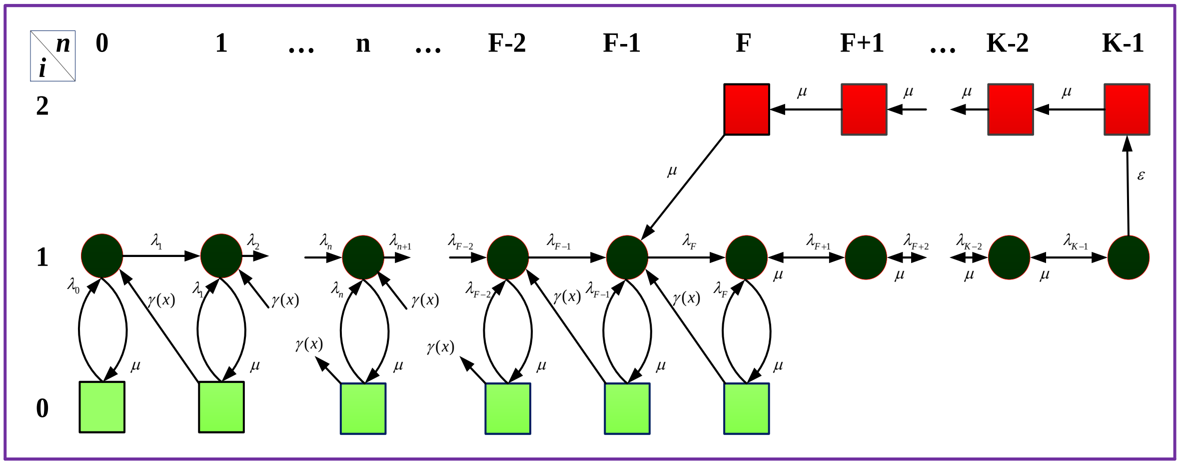


Figure 6.1: Transition state diagram for MRP

distribution of retrial time (G) with cumulative distribution function (CDF) $G(x)\,(x \geq 0)$ and instantaneous rate $\gamma(x)$ and meantime $1/\gamma$.

- As the capacity of the system becomes full, then the failed machines cannot join the system; a setup job following exponential distribution with rate ε is needed to restrict the further entry of failed machines in the system until the system size reduces to prefixed threshold level 'F' $(0 \leq F < K-1)$.
- The life time of the operating machine, repair time of the repairman, and retrial time of the failed machines are mutually independent.

Let at the time τ, $N(\tau)=\{0,1,...,K\}$ and $S(\tau)=i=\{0,1,2\}$ denote the number of failed machines in the system and the status of the repairman, respectively. The random variable $Y(\tau)$ is defined as follows:

$$S(\tau)=\begin{cases} 0 & \text{if the repairman is busy and the failed machines move to the retrial orbit,} \\ 1(2) & \text{if the repairman is busy and the failed machines are allowed (not allowed) to enter} \\ & \text{the system.} \end{cases}$$

Define the probabilities of the system states as follows:

$$P_{0,0}(\tau)=\text{Prob}\{S(\tau)=0, N(\tau)=0\}$$
$$P_{0,n}(x,\tau)dx=\text{Prob}\{S(\tau)=0, N(\tau)=n\}, x \leq U(\tau) \leq x+dx, x \geq 0, 1 \leq n \leq F$$
$$P_{0,n}(\tau)=\text{Prob}\{S(\tau)=0, N(\tau)=n\}=\int_0^\infty P_{0,n}(x,\tau)dx,\ 1 \leq n \leq F$$
$$P_{1,n}(\tau)=\text{Prob}\{S(\tau)=1, N(\tau)=n\}=\int_0^\infty P_{1,n}(x,\tau)dx,\ 1 \leq n \leq K-1$$

Define $\delta_{n-F,0}=\begin{cases}1, & 1\le n\le F,\\ 0, & \text{elsewhere.}\end{cases}$

We use the supplementary variable (U) corresponding to remaining retrial times for the failed machines while put in the retrial pool.

We shall explore the model at steady state, i.e., when $\tau\to\infty$.The steady-state probability $P_{i,n}$ is defined as

$$P_{i,n}=\lim_{\tau\to\infty}\begin{cases}P_{0,n}(\tau), & 0\le n\le F,\ i=0,\\ P_{1,n}(\tau), & 0\le n\le K-1,\ i=1,\\ P_{2,n}(\tau), & F\le n\le K-1,\ i=2.\end{cases}$$

Define the Laplace-Stieltjes transform (LST) of $\gamma(x)$ and $P_{0,n}(x)$ by $\gamma^*(\theta)$ and $P_{0,n}^*(\theta)$, respectively.

Also, $\prod_{i=n}^{n-1} y_i=1, n=1,2,3,...$ (6.1)

6.3. GOVERNING EQUATIONS AND ANALYSIS

To analyze the finite population model with a retrial and under control F-policy, the steady state Chapman-Kolmogorov equations for the three levels when $S(\tau)=\{0,1,2\}$ are framed by introducing the supplementary variable (U) corresponding to remaining retrial times as follows:

(i) ***For level*** $i=0$: ***When failed machines are waiting for repair job in retrial orbit.***

$$\lambda_0 P_{0,0}=\mu P_{1,0} \tag{6.2}$$

$$-\frac{d}{dx}P_{0,n}(x)=-\lambda_n P_{0,n}(x)+\mu P_{1,n}\gamma(x), 1\le n\le F. \tag{6.3}$$

(ii) ***For level*** $i=1$: ***When failed machines are allowed to join the system for a repair job.***

$$(\lambda_{n+1}+\mu)P_{1,n}=\delta_{n-F,0}\lambda_n P_{1,n-1}+\lambda_n P_{0,n}+P_{0,n+1}(0), 0\le n\le F-2 \tag{6.4}$$

$$(\lambda_F+\mu)P_{1,F-1}=\lambda_{F-1}P_{1,F-2}+\lambda_{F-1}P_{0,F-1}+P_{0,F}(0)+\mu P_{2,F} \tag{6.5}$$

$$(\lambda_{n+1}+\mu)P_{1,n}=\lambda_n P_{1,n-1}+\delta_{n-F,0}\lambda_n P_{0,n}+\mu P_{1,n+1}, F\le n\le K-2 \tag{6.6}$$

$$(\varepsilon+\mu)P_{1,K-1}=\lambda_{K-1}P_{1,K-2} \tag{6.7}$$

(iii) ***For level*** $i=2$: ***When failed machines are not allowed to join the system for repair job. during F-policy.***

$$\mu P_{2,n}=\mu P_{2,n+1}=\varepsilon P_{1,K-1}, F\le n\le K-2 \tag{6.8}$$

Define $P^*_{0,n}(\theta) = P_{0,n}\gamma^*(\theta)$ (6.9)

Taking LST on both sides of (6.3), we get

$$(\lambda_n - \theta)P^*_{0,n}(\theta) = \mu P_{1,n}\gamma^*(\theta) - P_{0,n}(0), 1 \le n \le F \tag{6.10}$$

$$P_{1,n} = \frac{\lambda_n}{\mu(1-\gamma^*(\lambda_n))}P_{0,n} = \frac{1}{\mu\gamma^*(\lambda_n)}P_{0,n}(0),\ 1 \le n \le F \tag{6.11}$$

Using (6.2), we get

$$\mathrm{P}_{1,0} = \frac{K\lambda}{\mu}\mathrm{P}_{0,0} \tag{6.12}$$

Using (6.4) and (6.12), we have

$$P_{0,1}(0) = \frac{K(K-1)\lambda^2}{\mu}P_{0,0} \tag{6.13}$$

Using (6.10) and (6.13), we obtain

$$P_{1,1} = \frac{K(K-1)\lambda^2}{\mu^2\gamma^*[(K-1)\lambda]}P_{0,0} \tag{6.14}$$

Using (6.10), (6.13) and (6.14), we have

$$P^*_{0,1}(0) = \frac{K\lambda}{\mu}\left[\frac{1}{\gamma^*(\lambda_1)} - 1\right]P_{0,0} \tag{6.15}$$

Using (6.9), (6.10), (6.13) and (6.14), we obtain

$$P_{0,1} = \frac{K\lambda}{\mu}\left[\frac{1}{\gamma^*(\lambda_1)} - 1\right]P_{0,0} \tag{6.16}$$

Setting $n=1$ in (6.4), and using (6.11) and (6.14), we get

$$P_{0,2}(0) = \frac{K(K-1)(K-2)\lambda^3}{\mu^2\gamma^*(\lambda_1)}P_{0,0} \tag{6.17}$$

Using (6.10) and (6.17), we have

$$P_{1,2} = \frac{P_{0,2}(0)}{\mu\gamma^*(\lambda_2)} = \frac{K(K-1)(K-2)\lambda^3}{\mu^3\gamma^*(\lambda_2)\gamma^*(\lambda_1)}P_{0,0} \tag{6.18}$$

Using (6.10), (6.17) and (6.18), we get

$$P^*_{0,2}(0) = P_{0,2} = \frac{K(K-1)\lambda^2}{\mu^2\gamma^*(\lambda_1)}\left[\frac{1}{\gamma^*(\lambda_2)} - 1\right]P_{0,0} \tag{6.19}$$

In general, we obtain

$$P_{1,n} = \begin{cases} \Phi_{1,n}, & 0 \le n \le F-1, \\ \mu^{n+1-F} f(n)\Phi_{1,n}R^{-1}, & F \le n \le K-2, \\ \mu^{n-F}\Phi_{1,n-1}R^{-1}, & n = K-1. \end{cases} \tag{6.20}$$

$$P_{0,n} = \begin{cases} \Psi_{0,n-1}, & 1 \le n \le F-1, \\ \mu\gamma^*(\lambda_n) f(n)\Psi_{0,n-1}R^{-1}, & n = F. \end{cases} \tag{6.21}$$

$$P_{0,n}(0) = \begin{cases} \Theta_{0,n-1}, & 1 \le n \le F-1, \\ \mu\gamma^*(\lambda_n) f(n)\Theta_{0,n-1}R^{-1}, & n = F. \end{cases} \tag{6.22}$$

$$P_{2,n} = \frac{\varepsilon\mu^{n-F-1}}{R}\Phi_{1,n-1},\ F \le n \le K-1. \tag{6.23}$$

Here $P_{0,0}$ can be determined by following the normalizing condition given by

$$\sum_{n=0}^{F} P_{0,n} + \sum_{n=0}^{K-1} P_{1,n} + \sum_{n=F}^{K-1} P_{2,n} = 1 \tag{6.24}$$

where $\Phi_{1,n} = \dfrac{K\lambda^{n+1}}{\mu^{n+1}} \prod_{i=1}^{n}\left(\dfrac{K-i}{\gamma^*(\lambda_i)}\right)\prod_{i=F}^{n}\left(\gamma^*(\lambda_i)\right)P_{0,0}$,

$$\Psi_{0,n-1} = \frac{K\lambda^n}{\mu^n}\prod_{i=1}^{n-1}\frac{(K-i)}{\gamma^*(\lambda_i)}\left[\frac{1}{\gamma^*(\lambda_n)} - 1\right]P_{0,0},$$

$$\Theta_{0,n-1} = \frac{K(K-n)\lambda^{n+1}}{\mu^n}\prod_{i=1}^{n-1}\frac{K-i}{\gamma^*(\lambda_i)}P_{0,0},$$

$$f(F) = \left(\mu^{K-F-2}(\mu+\varepsilon) + \varepsilon\sum_{j=1}^{K-F-2}\left(\mu^{K-F-2-j}\prod_{i=1}^{j}\lambda_{K-i}\right)\right),$$

$$R = \varepsilon\prod_{i=F+1}^{K-1}(\lambda_i) + \mu\gamma^*(\lambda_F) f(F).$$

Special Case: When retrial time is takes as exponential

We put $\gamma^*(\theta) = \dfrac{\gamma}{\gamma+\lambda}$ for exponential retrial time. The results given in (6.20), (6.21) and (6.23) reduce in the following results:

$$P_{1,n} = \begin{cases} \dfrac{K\lambda^{n+1}}{\mu^{n+1}} \displaystyle\prod_{i=1}^{n} (K-i) \left(\dfrac{\gamma+\lambda_i}{\gamma}\right) P_{0,0}, & 0 \le n \le F-1, \\ \dfrac{K\lambda^{n+1}}{\mu^{n+1}} \dfrac{\mu^{n+1-F} f(n)}{R} \displaystyle\prod_{i=1}^{n} (K-i) \left(\dfrac{\gamma+\lambda_i}{\gamma}\right) \prod_{i=F}^{n} \dfrac{\gamma}{\gamma+\lambda_i} P_{0,0}, & F \le n \le K-2, \\ \dfrac{K\lambda^{n+1}}{\mu^{n+1}} \dfrac{\mu^{n+1-F}}{R} \displaystyle\prod_{i=1}^{n} (n+1-i) \left(\dfrac{\gamma+\lambda_i}{\gamma}\right) \prod_{i=F}^{n} \dfrac{\gamma}{\gamma+\lambda_i} P_{0,0}, & n = K-1 \end{cases} \tag{6.25}$$

$$P_{0,n} = \begin{cases} \dfrac{K\lambda^{n}}{\mu^{n}} \displaystyle\prod_{i=1}^{n-1} (K-i) \left(\dfrac{\lambda_i}{\gamma+\lambda_i}\right) P_{0,0}, & 1 \le n \le F-1 \\ \dfrac{\mu\gamma f(n)}{(\gamma+\lambda_n)R} \dfrac{K\lambda^{n}}{\mu^{n}} \displaystyle\prod_{i=1}^{n-1} (K-i) \left(\dfrac{\lambda_i}{\gamma+\lambda_i}\right) P_{0,0}, & n = F \end{cases} \tag{6.26}$$

$$P_{2,n} = \frac{\varepsilon}{R} \frac{K\lambda^{n}}{\mu^{F+1}} \prod_{i=1}^{n-1} (K-i) \left(\frac{\gamma+\lambda_i}{\gamma}\right) \prod_{i=F}^{n-1} \left(\frac{\gamma}{\gamma+\lambda_i}\right) P_{0,0}, \; F \le n \le K-1 \tag{6.27}$$

6.4. PERFORMANCE MEASURES

To explore the behavior of the system, we establish the various performance indices as follows:

- The expected number of failed units in the system, is

$$E[N_S] = \sum_{n=0}^{F} nP_{0,n} + \sum_{n=0}^{K-1} (n+1) P_{1,n} + \sum_{n=F}^{K-1} (n+1) P_{2,n} \tag{6.28}$$

- Expected number of failed units in the queue and in the retrial orbit respectively, are

$$E[N_q] = \sum_{n=0}^{K-1} nP_{1,n} + \sum_{n=F}^{K-1} nP_{2,n} \text{ and } E[N_R] = \sum_{n=0}^{F} nP_{0,n} \tag{6.29a-b}$$

- The probability of the server being idle and busy respectively, are

$$P_I = \sum_{n=0}^{F} P_{0,n} \text{ and } P_{SB} = \sum_{n=0}^{K-1} P_{1,n} + \sum_{n=F}^{K-1} P_{2,n} \tag{6.30a-b}$$

- The unit availability and operative efficiency respectively, are

$$MA = 1 - \frac{E[N]}{K} \text{ and } O.E. = 1 - \sum_{n=0}^{F} P_{0,n} \tag{6.31a-b}$$

Cost Function: We construct a cost function with service rate as decision variable corresponding to total cost incurred per unit time in the system. The cost function is defined as

$$TC(\mu) = C_I P_I + C_B P_{SB} + C_H E[N_q] + \mu C_F + \mu C_A + C_O E[N_R] \quad (6.32)$$

where

TC : Total cost per unit time

C_I (C_B) : Cost per unit time when the server is idle (busy)

C_H : Holding cost per unit time of each unit present in the system

C_F (C_A): Cost per unit time for providing service to the unit when the arrivals are not allowed (allowed)

C_O : Cost per unit time incurred on each unit in the retrial orbit

6.5. ILLUSTRATION

To illustrate the practical utility of admission control for machine interference model with a retrial orbit, here we cite an example of automobile repair workshop where vehicles can be repaired. In the repair workshop, following Poisson process with rate λ, the motor vehicles of finite capacity (say K) arrive for the repair/maintenance. The repair job of motor vehicles is done by a single repairman as per exponential distribution with rate μ on basis of first-come-first-served rule. If the arriving motor vehicles find the repairman busy with the repair job of other vehicle, then the motor vehicles go to the parking area (i.e. retrial pool) of finite capacity (say F) available in the workshop. Due to the limited parking capacity of the workshop, the admission control policy for the arrived motor vehicles may be applied to maintain the smooth functioning of the repair workshop. As soon as the capacity of the workshop is full, the vehicles arriving for the repair will not be allowed to enter in the workshop until the parking places of the workshop decreases to a predefined threshold level 'F' at which the vehicles are further allowed to join the workshop to get repair/maintenance services.

6.6. NUMERICAL RESULTS

We present the numerical illustration to analyze the effects of system parameters on various performance measures. For the computational purpose, the default parameters are fixed as $K = 7$, $F = 4$, $\lambda = 3$, $\mu = 8$, $\gamma = 1$, $\varepsilon = 1$, $C_I = 10$, $C_B = 10$, $C_H = 120$, $C_F = 10$, $C_A = 5$, $C_o = 90$.

For the computational purpose, the three distributions for the retrial times namely, exponential (Exp), Erlang-3 (E_3) and deterministic (D) are taken. It is noted that

$$\gamma^*(\theta)=\begin{cases}\dfrac{\gamma}{\theta+\gamma}, & \text{for exponential retrial time,}\\ \left(\dfrac{3\gamma}{\theta+3\gamma}\right)^3, & \text{for Erlang-3 retrial time,}\\ e^{-\theta/\gamma}, & \text{for deterministic retrial time.}\end{cases}$$

The numerical results for $E[N_S]$, *TC* and *MA* are displayed in Figures 6.2-6.3 and Table 6.1.

Table 6.1: Machine availability (MA) and Total Cost (TC) by varying λ, μ and γ

(λ,μ,γ)	*MA*			*TC*		
	Exp	E_3	*D*	*Exp*	E_3	*D*
(1,8,1)	0.46	0.41	0.39	477.31	509.48	526.23
(3,8,1)	0.28	0.28	0.27	617.53	622.46	623.36
(5,8,1)	0.22	0.21	0.21	677.31	678.53	678.63
(3,4,1)	0.21	0.21	0.21	629.51	630.61	630.80
(3,8,1)	0.28	0.28	0.27	617.53	622.46	623.36
(3,12,1)	0.33	0.32	0.32	633.95	643.06	644.74
(3,8,1)	0.28	0.27	0.27	617.53	622.46	623.36
(3,8,2)	0.29	0.28	0.28	612.79	619.99	622.73
(3,8,3)	0.30	0.29	0.29	608.77	617.30	620.57

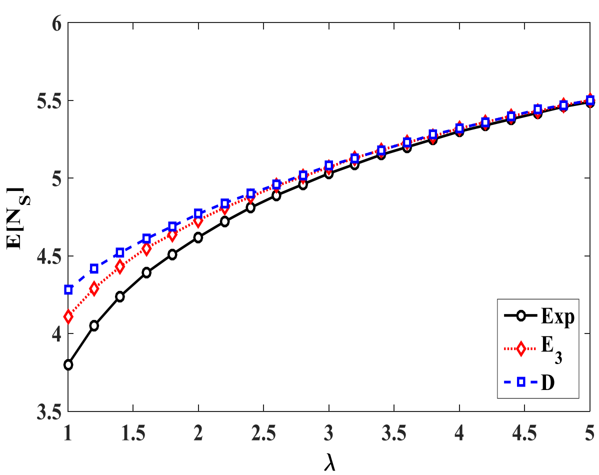


Figure 6.2: $E[N_S]$ vs λ for different distributions

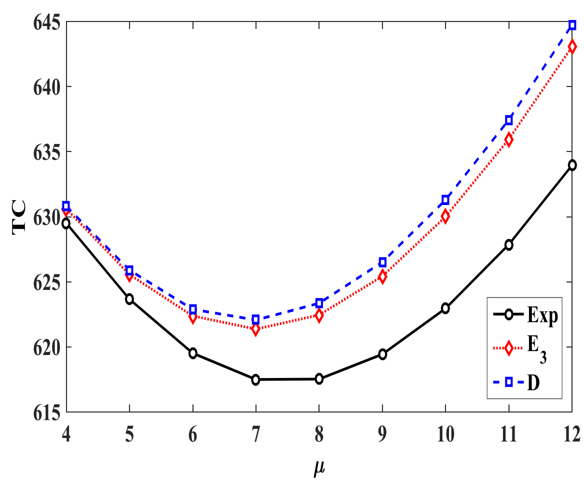


Figure 6.3: TC vs μ for different distributions

From Table 6.1, we notice that the machine availability decreases as λ increases but increases as μ and γ increase. The optimal control parameter 'μ' is determined using heuristic approach based on direct search approach by computing the cost function. From Figure 6.3, we find $(\mu^*, TC(\mu^*)) = \{(7.477,\ \$617.27),(6.966,\ \$621.37),(6.876,\ \$622.06)\}$ corresponding to three distributions {exponential, Erlang-3, deterministic} for retrial times.

6.7. FUZZY COST FUNCTION

The decision makers of the queueing system/organization are always interested to facilitate the repair job at minimum cost by setting the optimal control parameter μ. Cost analysis of the concerned system can point out benefit to the care-taker of the failed machines. The same in the case with present retrial machine interference problem wherein the cost function in the fuzzy environment is taken into account by considering the cost factors as trapezoidal fuzzy numbers. The total fuzzy cost $\overline{TC}(\bar{C}_O,\bar{C}_H,\bar{C}_I,\bar{C}_B,\bar{C}_F,\bar{C}_A)$ per unit of time incurred due to different cost elements treated as trapezoidal fuzzy numbers, is given by

$$\overline{TC}(\bar{C}_O,\bar{C}_H,\bar{C}_I,\bar{C}_B,\bar{C}_F,\bar{C}_A)=\bar{C}_O\otimes E[N_R]+\bar{C}_H\otimes E[N_q]+\bar{C}_I\otimes P_I+\bar{C}_B\otimes P_{SB}+\mu\otimes\bar{C}_F+\mu\otimes\bar{C}_A \quad (6.33)$$

where

$\bar{C}_O(\bar{C}_H)$: fuzzy cost spent on each failed machine in the retrial orbit (queue),

$\bar{C}_I(\bar{C}_B)$: fuzzy cost when the repairman is idle (busy),

$\bar{C}_A(\bar{C}_F)$: fuzzy cost when failed machines are allowed (not allowed) in the system.

The following fuzzy cost factors are represented by trapezoidal fuzzy numbers as:

$\bar{C}_O=(h_{1,\mathrm{O}},h_{2,\mathrm{O}},h_{3,\mathrm{O}},h_{4,\mathrm{O}})$, $\bar{C}_H=(h_{1,\mathrm{H}},h_{2,\mathrm{H}},h_{3,\mathrm{H}},h_{4,\mathrm{H}})$, $\bar{C}_I=(h_{1,\mathrm{I}},h_{2,\mathrm{I}},h_{3,\mathrm{I}},h_{4,\mathrm{I}})$,

$\bar{C}_B=(h_{1,\mathrm{B}},h_{2,\mathrm{B}},h_{3,\mathrm{B}},h_{4,\mathrm{B}})$, $\bar{C}_F=(h_{1,\mathrm{F}},h_{2,\mathrm{F}},h_{3,\mathrm{F}},h_{4,\mathrm{F}})$, $\bar{C}_A=(h_{1,\mathrm{A}},h_{2,\mathrm{A}},h_{3,\mathrm{A}},h_{4,\mathrm{A}})$.

From (6.33), we have

$$\begin{aligned}\overline{TC}(\bar{C}_O,\bar{C}_H,\bar{C}_I,\bar{C}_B,\bar{C}_F,\bar{C}_A)=&(h_{1,\mathrm{O}},h_{2,\mathrm{O}},b_{3,\mathrm{O}},b_{4,\mathrm{O}})\otimes E[N_R]\oplus(h_{1,\mathrm{H}},h_{2,\mathrm{H}},h_{3,\mathrm{H}},h_{4,\mathrm{H}})\otimes E[N_q]\oplus\\&(h_{1,\mathrm{I}},h_{2,\mathrm{I}},h_{3,\mathrm{I}},h_{4,\mathrm{I}})\otimes P_I\oplus(h_{1,\mathrm{B}},h_{2,\mathrm{B}},h_{3,\mathrm{B}},h_{4,\mathrm{B}})\otimes P_{SB}\oplus\\&\mu\otimes(h_{1,\mathrm{F}},h_{2,\mathrm{F}},h_{3,\mathrm{F}},h_{4,\mathrm{F}})\oplus\mu\otimes(h_{1,\mathrm{A}},h_{2,\mathrm{A}},h_{3,\mathrm{A}},h_{4,\mathrm{A}}).\end{aligned} \quad (6.34)$$

Simplifying (6.34) by using arithmetic operations defined in Section 1.4.7 of Chapter 1, we get

$$\begin{aligned}\overline{TC}=(&h_{1,\mathrm{O}}E[N_R]+h_{1,\mathrm{H}}E[N_q]+h_{1,\mathrm{I}}P_I+h_{1,\mathrm{B}}P_{SB}+h_{1,\mathrm{F}}\mu+h_{1,\mathrm{A}}\mu,\ h_{2,\mathrm{O}}E[N_R]+h_{2,\mathrm{H}}E[N_q]+h_{2,\mathrm{I}}P_I+h_{2,\mathrm{B}}P_{SB}+\\&h_{2,\mathrm{F}}\mu+h_{2,\mathrm{A}}\mu,\ h_{3,\mathrm{O}}E[N_R]+h_{3,\mathrm{H}}E[N_q]+h_{3,\mathrm{I}}P_I+h_{3,\mathrm{B}}P_{SB}+h_{3,\mathrm{F}}\mu+h_{3,\mathrm{A}}\mu,\ h_{4,\mathrm{O}}E[N_R]+h_{4,\mathrm{H}}E[N_q]+\\&h_{4,\mathrm{I}}P_I+h_{4,\mathrm{B}}P_{SB}+h_{4,\mathrm{F}}\mu+h_{4,\mathrm{A}}\mu)\end{aligned} \quad (6.35)$$

The total fuzzy cost given is (6.35) is also a trapezoidal number having lower bound $(TC)_\alpha^L$ and upper bound $(TC)_\alpha^U$ of $\alpha-$cuts as

$$(TC)_\alpha^L = [h_{1,\mathrm{O}}E[N_R] + h_{1,\mathrm{H}}E[N_q] + h_{1,\mathrm{I}}P_I + h_{1,\mathrm{B}}P_{SB} + h_{1,\mathrm{F}}\mu + h_{1,\mathrm{A}}\mu] + \alpha[(h_{2,\mathrm{O}} - h_{1,\mathrm{O}})E[N_R] + (h_{2,\mathrm{H}} - h_{1,\mathrm{H}})E[N_q] + (h_{2,\mathrm{I}} - h_{1,\mathrm{I}})P_I + (h_{2,\mathrm{B}} - h_{1,\mathrm{B}})P_{SB} + (h_{2,\mathrm{F}} - h_{1,\mathrm{F}})\mu + (h_{2,\mathrm{A}} - h_{1,\mathrm{A}})\mu] \tag{6.36}$$

$$(TC)_\alpha^U = [h_{4,\mathrm{O}}E[N_R] + h_{4,\mathrm{H}}E[N_q] + h_{4,\mathrm{I}}P_I + h_{4,\mathrm{B}}P_{SB} + h_{4,\mathrm{F}}\mu + h_{4,\mathrm{A}}\mu] - \alpha[(h_{4,\mathrm{O}} - h_{3,\mathrm{O}})E[N_R] + (h_{4,\mathrm{H}} - h_{3,\mathrm{H}})E[N_q] + (h_{4,\mathrm{I}} - h_{3,\mathrm{I}})P_I + (h_{4,\mathrm{B}} - h_{3,\mathrm{B}})P_{SB} + (h_{4,\mathrm{F}} - h_{3,\mathrm{F}})\mu + (h_{4,\mathrm{A}} - h_{3,\mathrm{A}})\mu] \tag{6.37}$$

Defuzzification: A Signed Distance Method

To extract the crisp value from the fuzzy value of the system cost, assigned distance method is adopted by defuzzifying the $\overline{TC}$. Refer to Section 1.4.7 of Chapter 1, the signed distance from $\overline{0}$ to $\overline{TC}(\overline{C}_O, \overline{C}_H, \overline{C}_I, \overline{C}_B, \overline{C}_F, \overline{C}_A)$ is given by

$$D(\overline{TC}(\overline{C}_O, \overline{C}_H, \overline{C}_I, \overline{C}_B, \overline{C}_F, \overline{C}_A), \overline{0}) = \frac{1}{2}\int_0^1 \left((TC)_\alpha^L + (TC)_\alpha^U\right) d\alpha \tag{6.38}$$

Using (6.36) and (6.37) in (6.38), we obtain

$$D(\overline{TC}(\overline{C}_O, \overline{C}_H, \overline{C}_I, \overline{C}_B, \overline{C}_F, \overline{C}_A), \overline{0}) = \frac{1}{4}[(h_{1,\mathrm{O}} + h_{2,\mathrm{O}} + h_{3,\mathrm{O}} + h_{4,\mathrm{O}})E[N_R] + (h_{1,\mathrm{H}} + h_{2,\mathrm{H}} + h_{3,\mathrm{H}} + h_{4,\mathrm{H}}) E[N_q] + (h_{1,\mathrm{I}} + h_{2,\mathrm{I}} + h_{3,\mathrm{I}} + h_{4,\mathrm{I}})P_I + (h_{1,\mathrm{B}} + h_{2,\mathrm{B}} + h_{3,\mathrm{B}} + h_{4,\mathrm{B}})P_{SB} + \mu(h_{1,\mathrm{F}} + h_{2,\mathrm{F}} + h_{3,\mathrm{F}} + h_{4,\mathrm{F}}) + \mu(h_{1,\mathrm{A}} + h_{2,\mathrm{A}} + h_{3,\mathrm{A}} + h_{4,\mathrm{A}})] \equiv H(\mu) \tag{6.39}$$

We consider the right-hand side of (6.39) as a function of μ, i.e., $H(\mu)$ and minimize it with respect to μ using genetic algorithm (GA) in the next section.

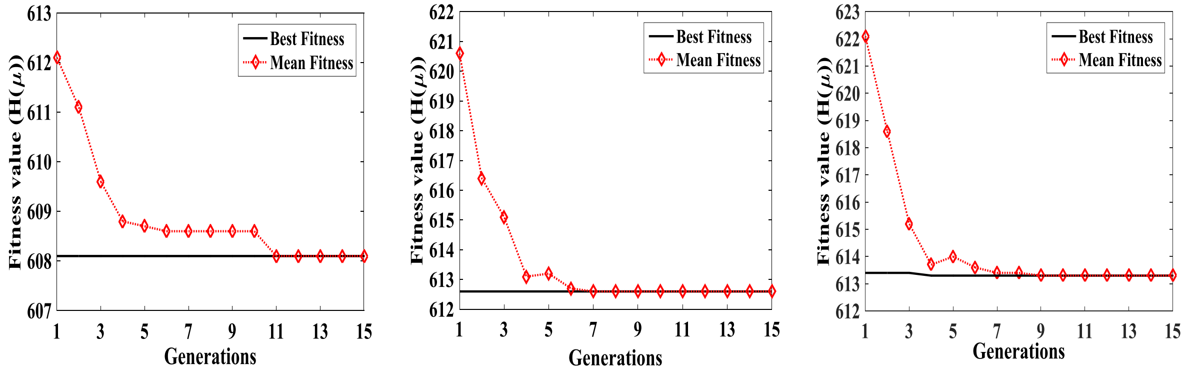
6.8. COST OPTIMIZATION USING A GENETIC ALGORITHM (GA)

It can be observed that the cost function given in (6.39) is non-linear in μ and thus it is not an easy job to find its minimum value by analytical method. Genetic algorithm (GA) is a search based approach which mimics the natural genetic. In order to evaluate the optimal value of the fitness function (objective function), it works in multidimensional search space. In this section, we determine the optimal control parameter μ by implementing a genetic algorithm to minimize $H(\mu)$. Genetic algorithm is associated with the fitness function (i.e. cost function), randomly generated a population of chromosomes, selection, cross over, and mutation. The description of GA can be seen in Section 1.4.8 of Chapter 1. To minimize (6.39), we determine minimum cost of the system and corresponding optimal control parameter $\mu *$; GA program is developed in MATLAB software.

The system input parameters $K, F, \lambda, \gamma, \varepsilon, \sigma,$ trapezoidal fuzzy numbers and GA parameters namely initial population size $(PS(\mu))$, genes (g_μ), selection operator (S), crossover rate (CR) and mutation rate (MR) are summarized in Table 6.2. The outputs $(\mu^*, H(\mu^*))$ of GA for different values of λ and γ are displayed in Table 6.3.

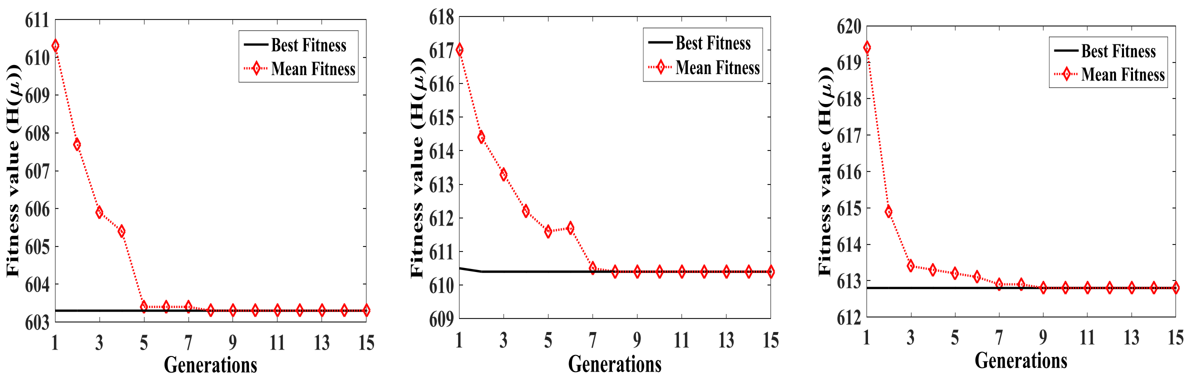
Table 6.2: Input parameters for computation of μ and $H(\mu^*)$ using GA

Parameters	Value assigned	Method
(K, F, ε) λ γ	$(7, 4, 1)$ 1, 2, 3 1, 2, 3	---
$\bar{C}_O = (h_{1,\mathrm{O}}, h_{2,\mathrm{O}}, h_{3,\mathrm{O}}, h_{4,\mathrm{O}})$ $\bar{C}_H = (h_{1,\mathrm{H}}, h_{2,\mathrm{H}}, h_{3,\mathrm{H}}, h_{4,\mathrm{H}})$ $\bar{C}_I = (h_{1,\mathrm{I}}, h_{2,\mathrm{I}}, h_{3,\mathrm{I}}, h_{4,\mathrm{I}})$ $\bar{C}_B = (h_{1,\mathrm{B}}, h_{2,\mathrm{B}}, h_{3,\mathrm{B}}, h_{4,\mathrm{B}})$ $\bar{C}_F = (h_{1,\mathrm{F}}, h_{2,\mathrm{F}}, h_{3,\mathrm{F}}, h_{4,\mathrm{F}})$ $\bar{C}_A = (h_{1,\mathrm{A}}, h_{2,\mathrm{A}}, h_{3,\mathrm{A}}, h_{4,\mathrm{A}})$	$(85, 88, 91, 94)$ $(115, 118, 121, 123)$ $(5, 7, 11, 14)$ $(6, 9, 11, 13)$ $(5, 8, 11, 13)$ $(2, 4, 6, 8)$	--
PS(μ)	50	Binary encoding
S	PS(μ) / 4	Tournament selection
CR	1	2-Point crossover
MR	0.08	Bit inversion mutation
Stopping criteria: (i) Generations (ii) Stall generation (iii)Function tolerance	 100 50 1×10^{-6}	--



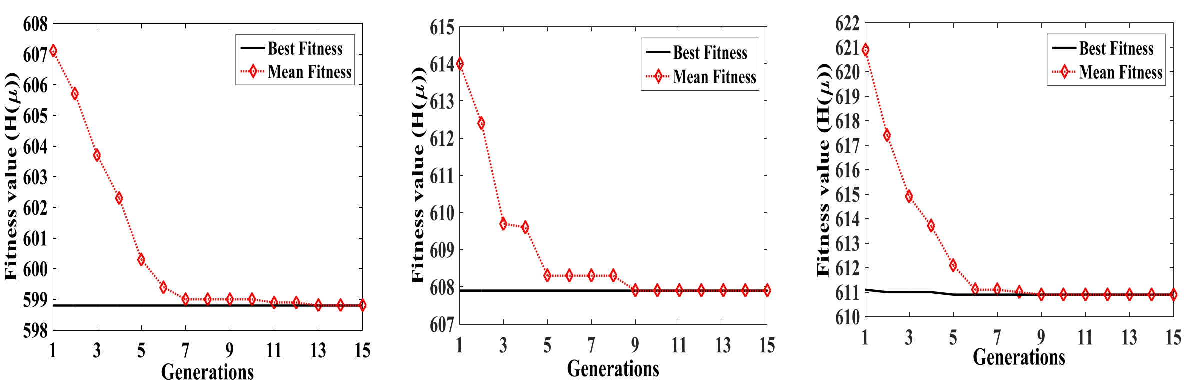
(a) (b) (c)

Figure 6.4: Fitness value vs Generations for (a) *Exp* (b) E_3 (c) D distributions when $(\lambda,\gamma)=(3,1)$



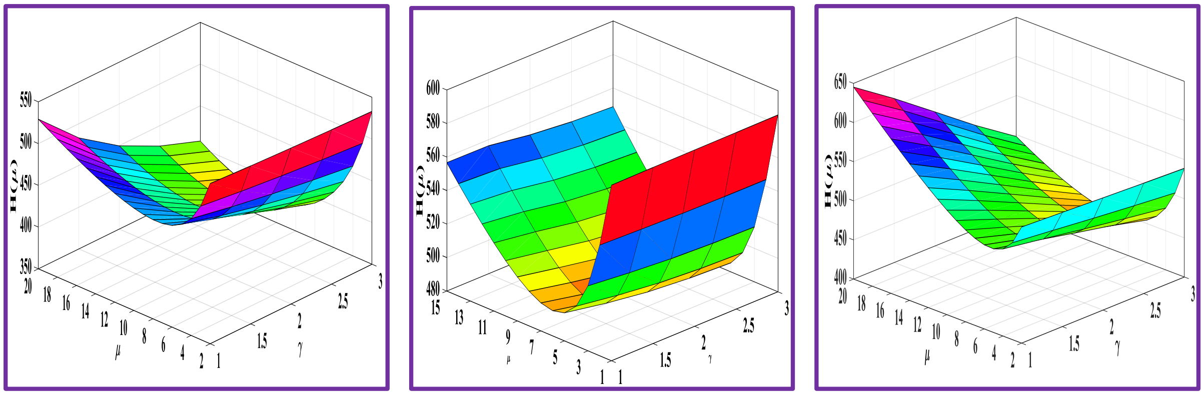
(a) (b) (c)

Figure 6.5: Fitness value vs Generations for (a) *Exp* (b) E_3 (c) D distributions when $(\lambda,\gamma)=(3,2)$



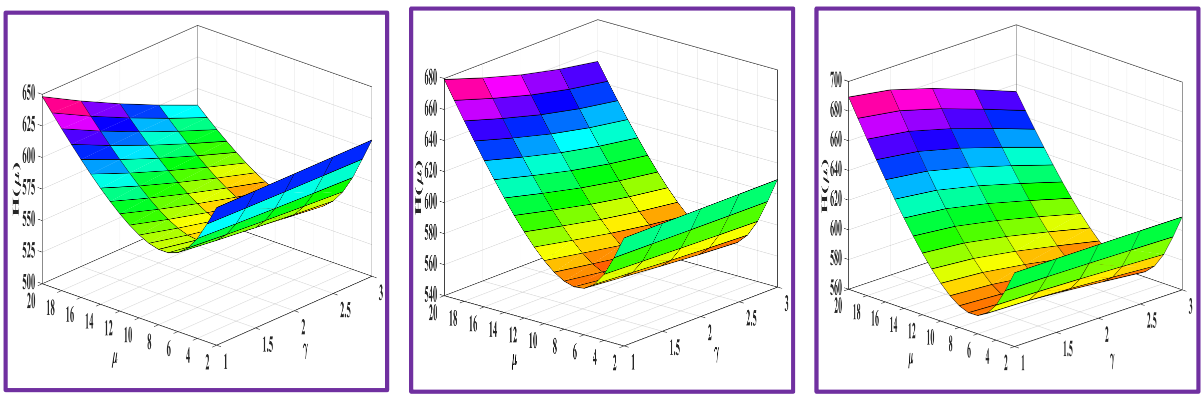
(a) (b) (c)

Figure 6.6: Fitness value vs Generations for (a) *Exp* (b) E_3 (c) D distributions when $(\lambda,\gamma)=(3,3)$



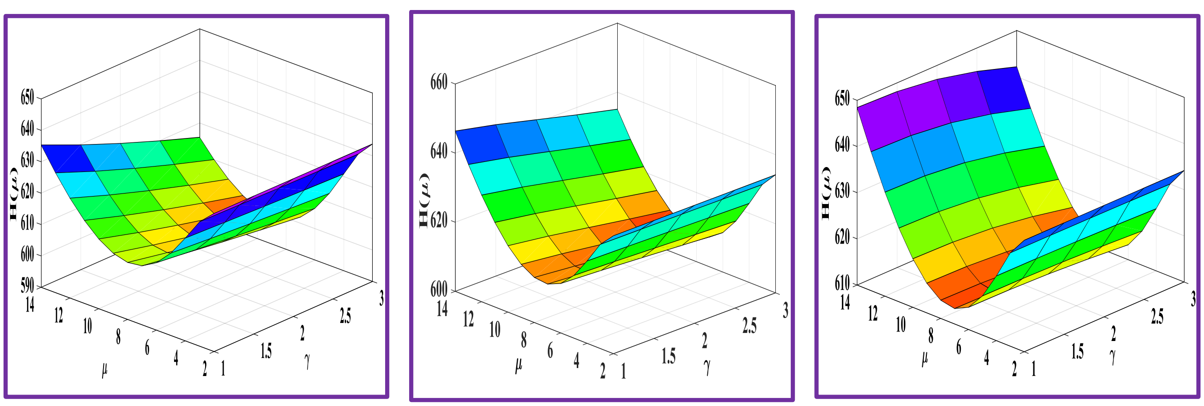
(a) (b) (c)

Figure 6.7: Effect of (μ, γ) on system cost for (a) Exp (b) E_3 (c) D distributions when $\lambda = 1$



(a) (b) (c)

Figure 6.8: Effect of (μ, γ) on system cost for (a) Exp (b) E_3 (c) D distributions when $\lambda = 2$



(a) (b) (c)

Figure 6.9: Effect of (μ, γ) on system cost for (a) Exp (b) E_3 (c) D distributions when $\lambda = 3$

While implementing genetic algorithm, it is noticed that the weighted average variation in the fitness function (cost function) value over stall generations is less than the function tolerance. The graphs are plotted for the fitness values (best fitness value and the mean fitness value) versus a number of iterations (generations) for different distributions for $\lambda = 3$. In case of three distributions namely exponential, Erlang-3 and deterministic of retrial times, Figures 6.4-6.6 display the GA trends of fitness values versus generations for $\gamma = 1, 2, 3$, respectively. The GA trends for $\lambda = 1$ and 2 can be plotted in a similar way.

Table 6.3: Total cost (in \$) for varying values of γ and λ

λ	γ	$(\mu^*, H(\mu^*))$		
		Exp	E_3	D
1	1	(7.663, 468.61)	(5.933, 495.00)	(5.333, 506.56)
	2	(11.446, 418.56)	(6.110, 485.95)	(6.712, 479.82)
	3	(12.475, 375.99)	(6.191, 482.31)	(8.776, 446.45)
2	1	(7.511, 559.39)	(6.771, 569.02)	(6.552, 571.82)
	2	(8.420, 547.83)	(7.210, 563.14)	(6.847, 568.06)
	3	(9.366, 536.61)	(7.532, 558.24)	(7.378, 560.99)
3	1	(7.821, 608.13)	(7.288, 612.55)	(7.194, 613.30)
	2	(8.382, 603.32)	(7.561, 610.42)	(7.261, 612.78)
	3	(8.910, 598.75)	(7.861, 607.91)	(7.491, 610.93)

The surface graphs for the total cost of the system are shown in Figures 6.7-6.9 for $\lambda = 1, 2, 3$ and three distributions namely exponential, Erlang-3 and deterministic for retrial times respectively, by varying values of γ on $x-$axis and μ on $y-$axis. Figures 6.7-6.9 depict that the cost function given in (6.39) is convex and optimal value of μ exists. The output numerical values for minimum cost $H(\mu^*)$ of the system corresponding to the optimal parameter μ^* for varying values of γ are recorded in Table 6.3.

The information listed in Table 6.3 of the minimum cost of the system corresponding to optimal control parameter (μ^*) on the basis of arrival rate and retrial rate of the failed machines can support to the decision maker to make the budget required for the machine interference system. It is quite clear that for the fixed value of arrival rate of failed machines, the system cost is less for exponential retrial time as compared to Erlang-3 and deterministic retrial times.

6.9. CONCLUDING REMARKS

The applicability of investigation done on the machine repair model of machine interference problem with general distributed retrial attempts and operating under F-policy can be realized in many real time systems. The cost optimization for machine interference problem has been studied in the fuzzy environment using a genetic algorithm. Various performance measures derived are further used to formulate the fuzzy cost function. The trapezoidal fuzzy number is considered for the cost elements associated with a cost function. $\alpha-$cut approach is applied on trapezoidal fuzzy cost function to get lower and upper bound of the fuzzy cost. A signed distance method of defuzzification for fuzzy cost is applied to convert the fuzzy cost into crisp cost function. The numerical simulation and sensitivity analysis carried out provide the valuable insights to the decision makers and system designers for controlling the system descriptors to achieve the desired output at optimum cost. The genetic algorithm is successfully implemented to determine optimal repair rate and corresponding minimum cost of the machine interference problem. The application of this investigation can be found in various places including the telecommunication system, production system, assembly lines, call centers, shopping malls, etc.

Chapter 7

Fuzzy Model for Machining System with General Repair and Vacation

7.1. INTRODUCTION

In recent past, some queueing theorists have studied the finite population models to deal with machine repair problems with standby support (cf. Sivazlian and Wang, 1989; Wang, 1995). A few researchers have also developed the machine repair models by incorporating the feature of server vacation to deal with more realistic scenarios (Gupta, 1997; Jain et al., 2004). To derive queueing and reliability measures such as failure frequency, mean time to failure (MTTF), reliability, availability of the *K*-out-of-*M*:G system, Ke and Lin (2005) studied a Markov model for machine repair system with standby support and multiple vacations. Ke and Wang (2007) and Jain and Upadhyaya (2009) analyzed the performance of a repairable multi-component machining system with standby support by incorporating the concepts of single vacation and multiple vacations, respectively. The profit analysis of machine repair problem with standby support and multiple vacations using particle swarm optimization has been carried out by Wang et al. (2014). Recently, Jain and Meena (2017) and Jain et al. (2019) proposed a Markovian model for machine repair problem (MRP) to analyze the performance of redundant fault tolerant machining system by incorporating admission control *F*-policy, imperfect coverage, working vacation and standby support.

To analyze non-Markovian model, the supplementary variable technique can be used (cf. Section 1.4.6 of Chapter 1). Some noticeable works on non-Markovian queueing modeling and performance analysis of machine interference systems can be seen in queueing literature (cf. Gupta and Rao, 1994; Wang et al., 2005). The repairable multi-component machining system with imperfect coverage and multiple vacations was studied by Jain and Gupta (2013). Using supplementary variable technique and recursive method, they derived the steady-state analytical expressions for various system indices. By treating remaining repair time as a supplementary variable, Ke and Liu (2014) proposed a non-Markovian queueing model for machining system with imperfect coverage, reboot and standby support. Ke et al. (2016) used the supplementary variable technique and recursive method to analyze the steady state behavior of machine repair system with switching failure and warm standby support.

The applicability of fuzzy queueing models in the real world scenario seems to be more appropriate than the crisp queueing models (cf. Chen and Chang, 2006). The notable works on fuzzy models can be seen in the literature (cf. Yang and Chang, 2015; Bouazzi et al., 2017; Jana et al., 2016; Gupta and Mohanty, 2016; Gupta and Mohanty, 2017). Verma et al. (2005) formulated a fuzzy optimization problem for conventional dc flow based crisp linear programing (CLP) model. Recently, Mueen et al. (2017) presented a fuzzy Markov single server queue by considering a hexagonal membership functions for determining the waiting time in the queue and mean system size. Bhardwaj et al. (2018) proposed the fuzzy analysis of two servers queue placed in series combination using Zadeh's extension principle. They employed $\alpha-$cut approach for constructing the triangular membership function for the valuation of the system size by taking arrival and service rate as fuzzy parameters. In the queueing literature, most of the research works related to the performance modeling of machining system have been done in the crisp environment. In many studies on MRP, the crisp system parameters were used to develop the performance models of the machining system having failure prone on-line as well as standby components (cf. Wang et al., 2007). To analyze Markov model of multi-components machining system with standby support in a fuzzy environment, the system parameters should be considered in the linguistic form (cf. Buckley, 2004). A very few researches on the performance analysis of machining system have considered the fuzzy queueing descriptors. Chen (2006) used $\alpha-$cut approach to study the fuzzy queueing model for cost optimization of machine repair problem (MRP) by taking cost coefficients and the machine breakdown as the trapezoidal fuzzy numbers. Ke et al. (2008) developed Markov model for a redundant repairable system operating in fuzzy environment by incorporating the imperfect coverage. They have used the parametric nonlinear programming approach via $\alpha-$cut to represent the membership functions for MTTF and machine availability by considering the failure/repair rates of operating and standby units as trapezoidal fuzzy numbers.

Harmony search (HS) which is a musicians behavior inspired evolutionary algorithm for optimization, was introduced by Geem et al. (2001). In literature, it is noticed that there is no work on queueing modeling in which harmony search approach has been used for the cost optimization. For the brief description of HS method is presented Section 1.4.8 of Chapter 1.

In this chapter, we develop finite population M/G/1/K machine repair model with standby support and vacationing server in both crisp and fuzzy environments. Using the supplementary variable technique (SVT) and recursive approach, the stationary queue size distributions and various performance metrics of the concerned model have been established. The prime objective of present study is to develop fuzzy model and facilitate the cost analysis which has been carried out using harmony search algorithm. The remaining part of the chapter is organized in the following manner. In Section 7.2, the model description and governing equations of finite population non-Markovian M/G/1/K model are presented. The various system metrics and cost function are given in Section 7.3. In Section 7.4, the performance metrics of multi-component machining system are analyzed by developing FM/FG/1/K/K model with vacation in the fuzzy environment. Next Section 7.5 is devoted to the numerical results of performance metrics of both M/G/1/K/K and FM/FG/1/K/K models. The cost optimization problem has been studied to determine the optimal control parameters viz. repair rate and vacation rates using harmony search algorithm. Finally, we conclude the findings of the present investigation in Section 7.6.

7.2. MODEL DESCRIPTION

The finite population M/G/1/K/K multi-component machining system comprising of M operating and S warm standby machines is studied. For the repair job of failed machines, there is the provision of a single repairman who may be allowed to take a vacation in case of no pending repair jobs of failed machines.

To develop the non-Markovian finite population M/G/1/K queueing model, the following assumptions are made:

- The operating (standby) machines are prone to failure; the lifetime of operating (standby) machines follow exponential distribution with mean $1/\lambda(1/a)$.
- As soon as the operating machine fails, it is replaced by the standby machine if available. The failure characteristic of replaced standby machine is assumed to be the same as that of the operating machine. The switching time of the failed operating machine to warm standby machine after the repair is assumed to be negligible.
- The repair time of the operating machine is assumed to govern by general distribution with a probability distribution $B(x)$, the probability density function $b(x)\,(x \geq 0)$ and mean $1/\mu$. The repair jobs of failed machines are done according to their failure order i.e. first-come-first-served (FCFS) basis.

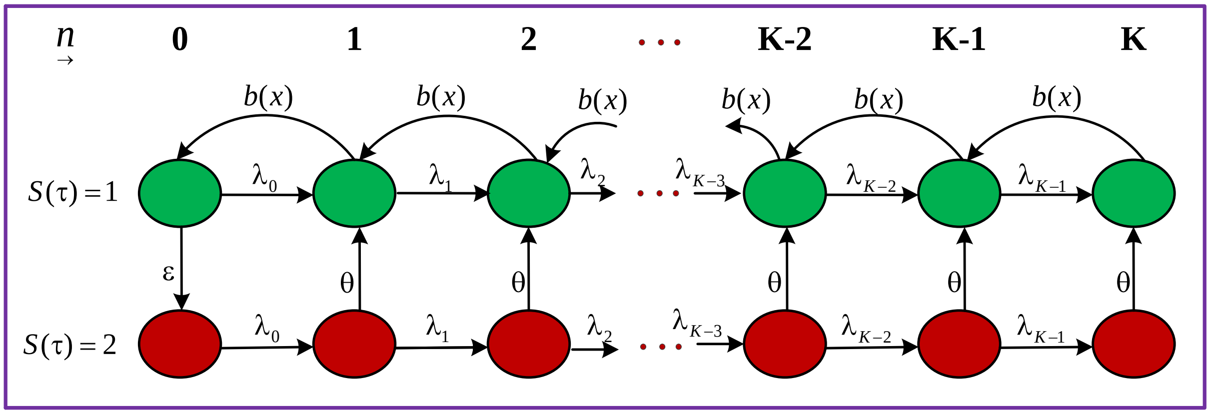


Figure 7.1: Transition state diagram for M/G/1/K/K model

- As the system becomes empty i.e. there is no failed machines in the system, the repairman goes for a vacation with a rate ε and returns back from vacation when any failed machine enters the system. The vacation time is governed by exponential distribution with mean $1/\theta$.

To analyze the non-Markovian finite population M/G/1/K model, we introduce $U(\tau)$ as a supplementary variable corresponding to the remaining repair time at time τ. Let $\{N(\tau), S(\tau); \tau \geq 0\}$ be a continuous time bi-variate stochastic process. At the time τ, $N(\tau)$ and $S(\tau)$ denote the number of failed machines in the system and status of the system, respectively. $S(\tau)$ holds values 1 and 2 for normal operating mode and vacation mode of the repairman, respectively.

We define system states probabilities at time epoch τ as follows:

(i) ***Normal operating state.***

$$P_{1,0}(\tau) = \text{Prob}\{S(\tau) = 1, N(\tau) = 0\}$$

$$P_{1,n}(x,\tau)dx = \text{Prob}\{S(\tau) = 1, N(\tau) = n, x < U(\tau) < x + dx\},\ 1 \leq n \leq K.$$

(ii) ***Vacation state.***

$$P_{2,n}(\tau) = \text{Prob}\{S(\tau) = 2, N(\tau) = n\}, 0 \leq n \leq K.$$

We denote $P_{1,n}(\tau) = \text{Prob}\{S(\tau) = 1, N(\tau) = n\} = \int_0^\infty P_{1,n}(x,\tau)dx,\ 1 \leq n \leq K-1$.

Define Laplace-Stieltjes transform of any function $b(x)$ as $b^*(s)$.

The state-dependent failure rate λ_n of the machines is defined by:

$$\lambda_n = \begin{cases} M\lambda + (S-n)a, & 0 \le n < S, \\ (M-n)\lambda, & S \le n < M+S \equiv K, \\ 0, & \text{otherwise.} \end{cases} \tag{7.1}$$

7.2.1. The Governing Equations and Queue Size Distributions

Chapman-Kolmogorov equations for the finite population model are framed using the state-transition rate (See Figure 7.1) relating to the individual states of the system at the time τ and $\tau + d\tau$ as follows:

$$\frac{d}{d\tau} P_{1,0}(\tau) = -(\lambda_0 + \varepsilon) P_{1,0}(\tau) + P_{1,1}(0,\tau) \tag{7.2}$$

$$\left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}\right) P_{1,n}(x,\tau) = -\lambda_n P_{1,n}(x,\tau) + \lambda_{n-1} P_{1,n-1}(x,\tau) + \theta b(x) P_{2,n-1}(x,\tau) + b(x) P_{1,n+1}(0,\tau), 1 \le n \le K-1 \tag{7.3}$$

$$\left(\frac{\partial}{\partial \tau} - \frac{\partial}{\partial x}\right) P_{1,K}(x,\tau) = \lambda_{K-1} P_{1,K-1}(x,\tau) + \theta b(x) P_{2,K}(x,\tau) \tag{7.4}$$

$$\frac{d}{d\tau} P_{2,0}(\tau) = -\lambda_0 P_{2,0}(\tau) + \varepsilon P_{1,0}(\tau) \tag{7.5}$$

$$\frac{d}{d\tau} P_{2,n}(\tau) = -(\lambda_n + \theta) P_{2,n}(\tau) + \lambda_{n-1} P_{2,n-1}(\tau),\ 1 \le n \le K-1 \tag{7.6}$$

$$\frac{d}{d\tau} P_{2,K}(\tau) = -\theta P_{2,K}(\tau) + \lambda_{K-1} P_{2,K-1}(\tau) \tag{7.7}$$

At the steady state, we have

$$P_{1,0} = \lim_{\tau \to \infty} P_{1,0}(\tau)$$

$$P_{1,n}(x) = \lim_{\tau \to \infty} P_{1,n}(x,\tau), 1 \le n \le K$$

$$P_{2,n}(x) = \lim_{\tau \to \infty} P_{2,n}(x,\tau),\ 0 \le n \le K.$$

Also, we define

$$P_{1,n}(x) = b(x) P_{1,n}, \quad 0 \le n \le K \tag{7.8}$$

The steady state equations corresponding to transient equations (7.2)-(7.7) can be written as

$$P_{1,1}(0) - (\lambda_0 + \varepsilon) P_{1,0} = 0 \tag{7.9}$$

$$\frac{d}{dx} P_{1,n}(x) - \lambda_n P_{1,n}(x) + \lambda_{n-1} P_{1,n-1}(x) + \theta b(x) P_{2,n}(x) + b(x) P_{1,n+1}(0) = 0,\ 1 \le n \le K-1 \tag{7.10}$$

$$\frac{d}{dx}P_{1,K}(x)+\lambda_{K-1}P_{1,K-1}(x)+\theta b(x)P_{2,K}=0 \tag{7.11}$$

$$-\lambda_0 P_{2,0}+\varepsilon P_{1,0}=0 \tag{7.12}$$

$$-(\lambda_n+\theta)P_{2,n}+\lambda_{n-1}P_{2,n-1}=0,\ 1\le n\le K-1 \tag{7.13}$$

$$-\theta P_{2,K}+\lambda_{K-1}P_{2,K-1}=0 \tag{7.14}$$

Define $\zeta_i=\dfrac{\lambda_i}{(\lambda_i+\theta)}$.

Now, from (7.10), (7.12) and (7.13), we obtain

$$\frac{d}{dx}P_{1,n}(x)-\lambda_n P_{1,n}(x)+\lambda_{n-1}P_{1,n-1}(x)+\theta b(x)\frac{\varepsilon}{\lambda_n}\prod_{i=1}^{n}\zeta_i P_{1,0}+b(x)P_{1,n+1}(0)=0,\ \ 1\le n\le K-1 \tag{7.15}$$

Also using (7.11)-(7.14), we obtain

$$-\frac{d}{dx}P_{1,K}(x)+\lambda_{K-1}P_{1,K-1}(x)+\varepsilon\, b(x)\prod_{i=1}^{K-1}\zeta_i P_{1,0}=0 \tag{7.16}$$

Laplace-Stieltjes transforms (LST) of (7.15)-(7.16) yield

$$\left(s\,P_{1,n}^*(s)-P_{1,n}(0)\right)-\lambda_n P_{1,n}^*(s)+\lambda_{n-1}b^*(s)P_{1,n-1}^*+\theta b^*(s)\frac{\varepsilon}{\lambda_n}\prod_{i=1}^{n}\zeta_i P_{1,n}+b^*(s)P_{1,n+1}(0)=0,1\le n\le K-1 \tag{7.17}$$

$$\left(sP_{1,K}^*(s)-P_{1,K}(0)\right)+\lambda_{K-1}P_{1,K-1}^*(s)+\varepsilon b^*(s)\prod_{i=1}^{K-1}\zeta_i P_{1,0}=0 \tag{7.18}$$

Substituting $s=0$ in (7.17), we get

$$P_{1,n+1}(0)=\lambda_n\,P_{1,n}+\varepsilon\prod_{i=1}^{n}\zeta_i P_{1,0},\ \ 1\le n\le K-1 \tag{7.19}$$

Using (7.17) and (7.19) and then substituting $s=\lambda_n$, we get

$$P_{1,n}=\left[\left(\frac{\lambda_0+\varepsilon}{\lambda_n}\right)\prod_{i=1}^{n}\frac{(1-b^*(\lambda_i))}{b^*(\lambda_i)}+\frac{\varepsilon}{\lambda_n}\sum_{i=1}^{n-1}\left(\prod_{j=1}^{i}\zeta_j\prod_{j=i+1}^{n}\frac{1-b^*(\lambda_j)}{b^*(\lambda_j)}\right)\right]P_{1,0},\ 1\le n\le K-1 \tag{7.20}$$

To find $P_{1,K}$, differentiate (7.18) with respect to $'s'$ and then substitute $s=0$. Thus, we obtain

$$P_{1,K}=b_1\left(\lambda_{K-1}P_{1,K-1}+\varepsilon\prod_{i=1}^{K-1}\zeta_i P_{1,0}\right) \tag{7.21}$$

Using (7.20) in (7.21), we get

$$P_{1,K}=b_1\left[(\lambda_0+\varepsilon)\prod_{i=1}^{K-1}\frac{(1-b^*(\lambda_i))}{b^*(\lambda_i)}+\varepsilon\prod_{i=1}^{K-1}\zeta_i+\varepsilon\sum_{i=1}^{K-2}\left(\prod_{j=1}^{i}\zeta_j\prod_{j=i+1}^{K-1}\frac{1-b^*(\lambda_j)}{b^*(\lambda_j)}\right)\right]P_{1,0} \tag{7.22}$$

where $b_1 = -\frac{d}{ds} b^*(0)$.

Solving recursively (7.12)-(7.14), we obtain

$$P_{2,n} = \begin{cases} \frac{\varepsilon}{\lambda_0} P_{1,0}, & n = 0, \\ \prod_{i=0}^{n-1} \zeta_i \frac{\varepsilon}{\lambda_0} P_{1,0}, & 1 \le n \le K-1, \\ \prod_{i=0}^{n-2} \zeta_i \frac{\varepsilon \lambda_{K-1}}{\theta \lambda_0} P_{1,0}, & n = K. \end{cases} \tag{7.23}$$

The unknown probability $P_{1,0}$ is evaluated using normalizing condition

$$\sum_{i=1}^{2} \sum_{n=0}^{K} P_{i,n} = 1 \tag{7.24}$$

Remark 7.1: For specific phase type distributions of general repair time, Laplace-Stieltjes transforms $b^*(\lambda_j)$ is given by

$$b^*(\lambda_j) = \begin{cases} \left(\frac{k\mu}{\lambda_j + k\mu} \right)^k, & \text{for k-phase Erlang distribution,} \\ \prod_{i=1}^{k} \left(\frac{\mu_i}{\lambda_j + \mu_i} \right), & \text{for coxian distribution,} \\ \sum_{i=1}^{L} p_i \left(\frac{k_i \mu_i}{\lambda_j + \mu_i} \right)^{k_i}, & \text{for mixed-Erlang distribution.} \end{cases}$$

Remarks 7.2: The $k-$phase Erlang distribution (E_k) reduces to exponential distribution if $k = 1$ and it becomes deterministic distribution when $k \to \infty$. When $\mu_i = \mu,$ then coxian distribution takes the form of gamma distribution. It is noted that for $k_i = 1$, mixed-Erlang distribution becomes hyper exponential distribution.

7.3. SYSTEM PERFORMANCE MEASURES AND COST FUNCTION

To predict the behavior of multi-component machining system, the performance measures are formulated as follows:

(i) The expected queue length of failed machines in the system (L_S) is

$$L_S = \sum_{n=0}^{K} n (P_{1,n} + P_{2,n}) \tag{7.25}$$

(ii) The machine availability (MA) is

$$MA = 1 - \frac{L_S}{K} \tag{7.26}$$

(iii) The long-run probabilities of the system being in operating state and vacation state respectively, are

$$P_{SB} = \sum_{n=0}^{K} P_{1,n} \text{ and } P_V = \sum_{n=0}^{K} P_{2,n} \tag{7.27a-b}$$

Now, we construct a cost function to analyze the cost associated with different activities of the machining system. The total cost incurred per unit time of the system is determined by considering repair rate (μ) and vacation rate (θ) as decision variables. To make machining system cost-economic, the total cost function $TC(\mu,\theta)$ is minimized corresponding to repair rate (μ) and vacation rate (θ).

The cost elements incurred on different activities of the system are as follows:

C_H : Holding cost per unit time incurred on each failed machine waiting for the repair.

C_B : The cost incurred per unit time on a busy repairman.

C_V : The cost incurred per unit time when the repairman is on vacation.

C_m : The cost incurred per unit time on the repair with rate μ.

Now, the cost function is

$$TC(\mu,\theta) = C_H L_S + C_B P_{SB} + C_V P_V + \mu C_m \tag{7.28}$$

The cost optimization problem (OP) is formulated as follows:

$$(\text{OP}) \qquad TC(\mu^*,\theta^*) = min\, TC(\mu,\theta) \tag{7.29}$$

It is quite tedious task to optimize $TC(\mu,\theta)$ analytically due to non-linear nature of the cost function. We use soft computing approach based on harmony search algorithm to determine the minimum expected total cost $TC(\mu^*,\theta^*)$ and optimal value of decision variables $\mu*$ and $\theta*$.

7.4. FM/FG/1/K MODEL WITH VACATION

Now, we consider a fuzzy model for multi-component machining system which is more realistic than the traditional used crisp parameter based M/G/1/K model for the machining system with vacation and spare part support. The fuzzified parameters $\tilde{\lambda},\ \tilde{\mu},$ and $\tilde{\theta}$ are considered corresponding to crisp parameters with membership functions $\eta_{\tilde{\lambda}}(p),\ \eta_{\tilde{\mu}}(q),$ and $\eta_{\tilde{\theta}}(r)$, respectively. Let $\tilde{\Lambda}_j$ be a fuzzy set with the membership function $\eta_{\tilde{\Lambda}_j}(z_j)$. Thus

$$\tilde{\Lambda}_j = \{(z_j, \eta_{\tilde{\Lambda}_j}(z_j)) : z_j \in Z_j\}, \quad j = 1,2,3. \tag{7.30}$$

For the FM/FG/1/K model, we have

$$\tilde{\Lambda}_j = \begin{cases} \tilde{\lambda}, & for\ j = 1, \\ \tilde{\mu}, & for\ j = 2, \\ \tilde{\theta}, & for\ j = 3. \end{cases} \text{ and } z_j = \begin{cases} p, & for\ j = 1, \\ q, & for\ j = 2, \\ r, & for\ j = 3. \end{cases}$$

Since $\tilde{\Lambda}_j$ represents the fuzzy numbers, we notice that $\tilde{L}_S$ and $M\tilde{A}$ are also fuzzy numbers corresponding to metrics L_S and MA. Now, we determine that the system characteristics of interest such as a expected number of failed machines in the system ($\tilde{L}_S$) and machine availability ($M\tilde{A}$). Thus

$$g(z_1, z_2, z_3) = L_S \tag{7.31}$$

The membership functions of $\tilde{L}_S$ is given by

$$\eta_{\tilde{L}_S}(z) = sup\, min\left\{\eta_{\tilde{\Lambda}_j}(z_i) : j = 1,2,3 \,|\, z = L_S\right\}. \tag{7.32}$$

Likewise, the membership function of $M\tilde{A}$ is

$$\eta_{M\tilde{A}}(z) = sup\, min\left\{\eta_{\tilde{\Lambda}_j}(z_i) : j = 1,2,3 \,|\, z = MA\right\}. \tag{7.33}$$

Now, we employ the parametric nonlinear programming (P-NLP) to find $\alpha-$cuts of $\tilde{L}_S$ and $M\tilde{A}$ using extension principle.

7.4.1. Parametric Nonlinear Programming (P-NLP)

In this subsection, our aim is to formulate the membership functions of $\tilde{L}_S$ and $M\tilde{A}$. From (7.32), we notice that $\eta_{\tilde{L}_S}(z)$ is the minimum of $\eta_{\tilde{\Lambda}_j}(z_i),\ j = 1,2,3$. Therefore, to formulate the membership function of $\tilde{L}_S$, one of the following three cases must hold such that $z = \tilde{L}_S$, which satisfies $\eta_{\tilde{L}_S}(z) = \alpha$:

Case 1: $\eta_{\tilde{\Lambda}_1}(z_1) = \alpha,\ \eta_{\tilde{\Lambda}_j}(z_j) \geq \alpha,\ j = 2,3.$

Case 2: $\eta_{\tilde{\Lambda}_2}(z_2) = \alpha,\ \eta_{\tilde{\Lambda}_j}(z_j) \geq \alpha,\ j = 1,3.$

Case 3: $\eta_{\tilde{\Lambda}_3}(z_3) = \alpha,\ \eta_{\tilde{\Lambda}_j}(z_j) \geq \alpha,\ j = 1,2.$

Now, for Cases 1-3, the lower bound (z_α^{LB}) and upper bound (z_α^{UB}) of $\alpha-$cut of $\tilde{L}_S$ are obtained as*:*

$$z_{j,\alpha}^{LB} = \min L_S \tag{7.34}$$

and $$z_{j,\alpha}^{UB} = \max L_S,\ j = 1,2,3 \tag{7.35}$$

For $0 < \alpha_2 < \alpha_1 \leq 1$, we have $\left[(z_j)_{\alpha_1}^{LB},\ (z_j)_{\alpha_1}^{UB}\right] \subseteq \left[(z_j)_{\alpha_2}^{LB},\ (z_j)_{\alpha_2}^{UB}\right],\ j = 1,2,3$ as $\alpha-$cuts form a nested structure with respect to α. Thus, (7.34) has the same smallest values whereas (7.35) has the same largest value. Now, to determine z_α^{LB} and z_α^{UB}, it is sufficient to evaluate the left and right part of $\eta_{\tilde{L}_S}(z)$ defined in (7.32). Thus, we have

$$(L_S)_\alpha^{LB} = z_\alpha^{LB} = \min L_S \quad \text{where } p_\alpha^{LB} \leq p \leq p_\alpha^{UB},\ q_\alpha^{LB} \leq q \leq q_\alpha^{UB}, r_\alpha^{LB} \leq r \leq r_\alpha^{UB}. \tag{7.36a}$$

$$(L_S)_\alpha^{UB} = z_\alpha^{UB} = \max L_S \quad \text{where } p_\alpha^{LB} \leq p \leq p_\alpha^{UB},\ q_\alpha^{LB} \leq q \leq q_\alpha^{UB}, r_\alpha^{LB} \leq r \leq r_\alpha^{UB}. \tag{7.36b}$$

Here (7.36a-b) are the special cases of P-NLPs (Gal, 1979). In order to hold $\eta_{\tilde{L}_S} = \alpha$, one of z_i must hits the boundary of its $\alpha-$cut. We observe that z_α^{LB} is increasing function whereas z_α^{UB} is decreasing function subject to α. Thus, $z_{\alpha_2}^{LB} \leq z_{\alpha_1}^{LB}$ and $z_{\alpha_2}^{UB} \geq z_{\alpha_1}^{UB};\ 0 < \alpha_2 < \alpha_1 \leq 1$.

Now, the membership function of $\tilde{L}_S$ is framed as

$$\eta_{\tilde{L}_S}(z) = \begin{cases} L_S(z), & (L_S)_{\alpha=0}^{LB} \leq z < (L_S)_{\alpha=1}^{LB} \\ 1, & (L_S)_{\alpha=1}^{LB} = z = (L_S)_{\alpha=1}^{UB} \\ R_S(z), & (L_S)_{\alpha=1}^{UB} < z \leq (L_S)_{\alpha=0}^{UB}. \end{cases} \tag{7.37}$$

For the analytical solutions of $(L_S)_\alpha^{LB}$ and $(L_S)_\alpha^{UB}$, closed-form results cannot be determined. However, the numerical technique can be applied to compute $(L_S)_\alpha^{LB}$ and $(L_S)_\alpha^{UB}$ which is enough to determine the shape of the membership function $(\eta_{\tilde{L}_S}(z))$ of $\tilde{L}_S$. In (7.37), $L_S(z)$ and

$R_S(z)$ denote the left and right parts of membership function of $\tilde{L}_S$ and are obtained as $L_S(z)=[(L_S)_\alpha^L]^{-1}$ and $R_S(z)=[(L_S)_\alpha^U]^{-1}$, respectively.

Now, we construct the lower bound $((MA)_\alpha^{LB})$ and upper bound $((MA)_\alpha^{UB})$ of $\alpha-$cut of $M\tilde{A}$ as

$$(MA)_\alpha^L = min\, MA \text{ where } p_\alpha^{LB} \le p \le p_\alpha^{UB},\ q_\alpha^{LB} \le q \le q_\alpha^{UB}, r_\alpha^{LB} \le r \le r_\alpha^{UB}. \tag{7.38a}$$

$$(MA)_\alpha^U = max\, MA \text{ where } p_\alpha^{LB} \le p \le p_\alpha^{UB},\ q_\alpha^{LB} \le q \le q_\alpha^{UB}, r_\alpha^{LB} \le r \le r_\alpha^{UB}. \tag{7.38b}$$

The membership function $\eta_{M\tilde{A}}$ of $M\tilde{A}$ is

$$\eta_{M\tilde{A}}(z)=\begin{cases} L_A(z), & MA_{\alpha=0}^{LB} \le z < MA_{\alpha=1}^{LB}, \\ 1, & MA_{\alpha=1}^{LB} = z = MA_{\alpha=1}^{UB}, \\ R_A(z), & MA_{\alpha=1}^{UB} < z \le MA_{\alpha=0}^{UB}. \end{cases} \tag{7.39}$$

where $((MA)_\alpha^{LB})^{-1} = L_A(z)$ and $((MA)_\alpha^{UB})^{-1} = R_A(z)$.

For FM/FG/1/K/K model, the implementation of P-NLP is done by considering $\tilde{\Lambda}_j$ as the triangular fuzzy number corresponding to crisp parameters λ, μ, θ respectively. Thus, we have $\tilde{\Lambda}_j = [\delta_{1,j}, \delta_{2,j}, \delta_{3,j}];\ \delta_{1,j} < \delta_{2,j} < \delta_{3,j}$.

The membership functions of $\tilde{\Lambda}_j$ is given by

$$\eta_{\tilde{\Lambda}}(z_j)=\begin{cases} \dfrac{z_j-\delta_{1,j}}{\delta_{2,j}-\delta_{1,j}}, & \delta_{1,j} \le z_j < \delta_{2,j}, \\ 1, & \delta_{2,j} = z_j, \\ \dfrac{\delta_{3,j}-z_j}{\delta_{3,j}-\delta_{2,j}}, & \delta_{2,j} < z_j \le \delta_{3,j}. \end{cases} \tag{7.40}$$

$\alpha-$cut of $\tilde{\Lambda}_j$ in terms of the crisp interval is given by

$$\Lambda(\alpha) = [(z_j)_\alpha^{LB}, (z_j)_\alpha^{UB}] = [\delta_{1,j} + \alpha(\delta_{2,j}-\delta_{1,j}),\ \delta_{3,j} - \alpha(\delta_{3,j}-\delta_{2,j})],\ j=1,2,3. \tag{7.41}$$

Thus, (7.36a) and (7.36b) yield

$$(L_S)_\alpha^{LB} = min\, L_S \tag{7.42}$$

and $$(L_S)_\alpha^{UB} = max\, L_S \tag{7.43}$$

where $\delta_{1,j} + \alpha(\delta_{2,j}-\delta_{1,j}) \le z_j \le \delta_{3,j} - \alpha(\delta_{3,j}-\delta_{2,j}),\ j=1,2,3.$

Similarly, from (7.38a) and (7.38b)

$$(MA)_\alpha^{LB} = min\, MA \tag{7.44}$$

and $(MA)_{\alpha}^{UB} = max\, MA$ (7.45)

where $\delta_{1,j} + \alpha(\delta_{2,j} - \delta_{1,j}) \le z_j \le \delta_{3,j} - \alpha(\delta_{3,j} - \delta_{2,j}),\ j = 1, 2, 3.$

7.5. NUMERICAL RESULTS

The numerical results by taking an illustration of machining system with standby (MSS) for different repair time distributions are presented. For computational purpose, we consider exponential distribution (M), 3-phase Erlang distribution (E_3) and deterministic distribution (D) of repair time and their LSTs can be found by putting $k = 1,\ k = 3$ and $k \to \infty$ in $k-$phase Erlang distribution.

7.5.1. Illustration of a Flexible Manufacturing System (FMS)

To reveal the practical applicability of the model proposed in this investigation, we cite an illustration of a flexible manufacturing system (FMS) where robots are used for packing purpose. The flexible manufacturing system consists of M operating and S standby robots; $K = M + S$. The operating (standby) robots are subjected to failure having lifetimes following an exponential process with parameter λ (a). The switching time of the failed operating robot to standby robot after the repair is assumed to be negligible. A skilled repairman is available in the system for providing repair to failed robots. If there are no failed robots in the system, the repairman goes for vacation by following the exponential distribution with parameter ε. The vacation time of the repairman is assumed to be distributed exponentially with parameter θ. The repair time of operating robot is governed by general distribution with cumulative distribution function $B(x)\,(x \ge 0)$, probability density function $b(x)\,(x \ge 0)$, and mean repair time $1/\mu$.

7.5.2. Sensitivity Analysis of M/G/1/K/K Model

To predict the behavior of the system with respect to different parameters, we carried out the numerical experiments using the Matlab software. The system indices by computing different state probabilities $P_{1,n}$ and $P_{2,n}$ $(0 \le n \le K)$ are obtained to characterize the system behavior by fixing default system parameters as $K = 4, M = 2,\ S = 2,\ \lambda = 0.2,\ k = 3,$ $a = 0.06,\ \mu = 2,\ \varepsilon = 0.5,\ \theta = 0.6.$

Numerical results displayed in Table 7.1 indicate the increasing trend of machine availability (MA) and vacation state probability (P_V) of the system with the increase in repair

Table 7.1: Performance measures vs μ for three models

μ	M/M/1/K			M/E3/1/K			M/D/1/K		
	MA	P_{SB}	P_V	MA	P_{SB}	P_V	MA	P_{SB}	P_V
0.1	0.052	1.000	0.000	0.044	1.000	0.000	0.042	1.000	0.000
0.6	0.377	0.896	0.104	0.260	0.984	0.016	0.227	0.994	0.006
1.1	0.566	0.738	0.262	0.461	0.901	0.099	0.425	0.928	0.072
1.6	0.630	0.662	0.338	0.571	0.813	0.187	0.552	0.836	0.164
2.1	0.657	0.624	0.376	0.625	0.751	0.249	0.615	0.767	0.233
2.6	0.671	0.604	0.396	0.652	0.710	0.290	0.647	0.721	0.279
3.1	0.679	0.591	0.409	0.668	0.681	0.319	0.665	0.689	0.311
3.6	0.684	0.582	0.418	0.678	0.661	0.339	0.676	0.667	0.333

Table 7.2: Performance measures vs θ for three models

θ	M/M/1/K			M/E3/1/K			M/D/1/K		
	MA	P_{SB}	P_V	MA	P_{SB}	P_V	MA	P_{SB}	P_V
0.5	0.579	0.418	0.582	0.587	0.519	0.481	0.585	0.529	0.471
0.7	0.620	0.471	0.529	0.620	0.572	0.428	0.618	0.582	0.418
0.9	0.642	0.507	0.493	0.638	0.607	0.393	0.636	0.616	0.384
1.1	0.655	0.532	0.468	0.649	0.631	0.369	0.646	0.640	0.360
1.3	0.664	0.552	0.448	0.655	0.649	0.351	0.652	0.658	0.342
1.5	0.670	0.567	0.433	0.660	0.663	0.337	0.656	0.671	0.329
1.7	0.674	0.579	0.421	0.662	0.674	0.326	0.659	0.682	0.318
1.9	0.676	0.589	0.411	0.664	0.683	0.317	0.661	0.691	0.309

rate (μ) for all the three models M/M/1/K, M/E$_3$/1/K, and M/D/1/K. The operative state probability (P_{SB}) also increases as the value of repair rate (μ) increases for all the three models.

The machine availability (MA) and operative state probability (P_{SB}) of the system increase with the increase in vacation return rate (θ). On the contrary, the vacation state probability (P_V) shows a decreasing trend as rate (θ) grows up; this fact can be seen in Table 7.2. Figures 7.2 and 7.3 depict the trend of mean queue length of failed machines (L_S) in the system for M/M/1, M/E$_3$/1 and M/D/1 models with respect to failure rate (λ) and repair rate (μ) respectively. Figure 7.2 depicts that L_S increases rapidly initially for the increasing value of failure rate from $\lambda = 0.1$ to $\lambda = 1$; but beyond $\lambda = 1$, the mean queue length gradually increases for the further increment in $\lambda = 1$ to $\lambda = 2$. On the contrary, Figure 7.3 shows a reverse trend for L_S for varying values of repair rate (μ). In Figure 7.4, the machine availability (MA) decreases rapidly initially as the value of λ increases from $\lambda = 0.1$ to $\lambda = 1.2$, but then after the gradually

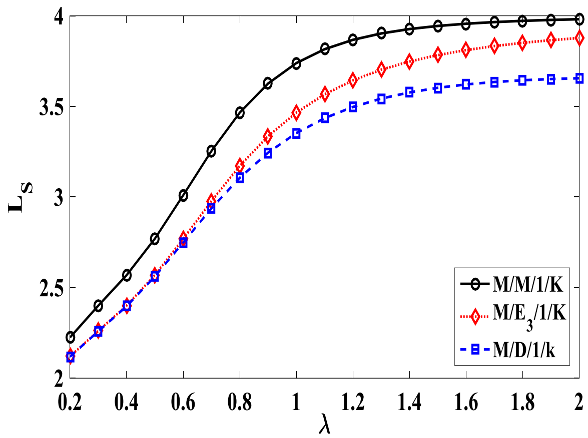


Figure 7.2: Effect of λ on L_S

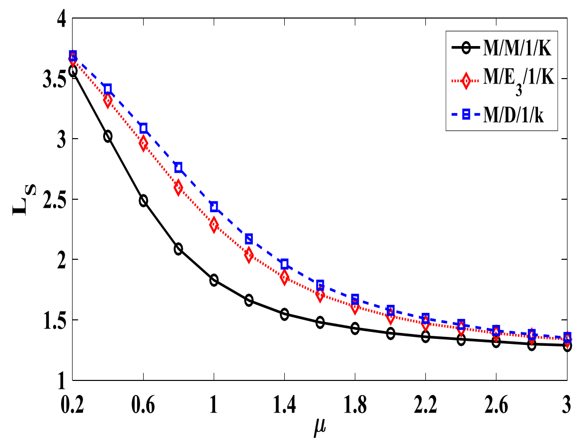


Figure 7.3: Effect of μ on L_S

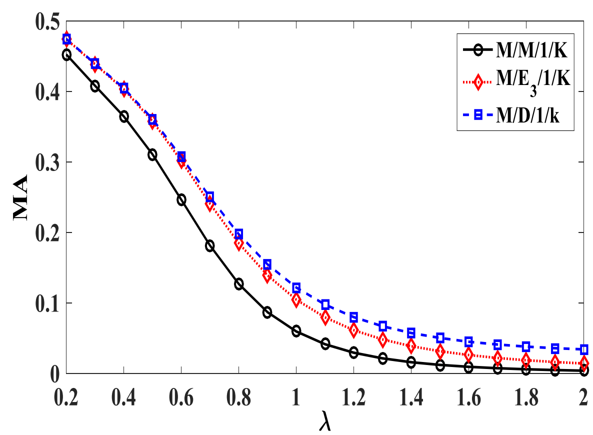


Figure 7.4: Effect of λ on MA

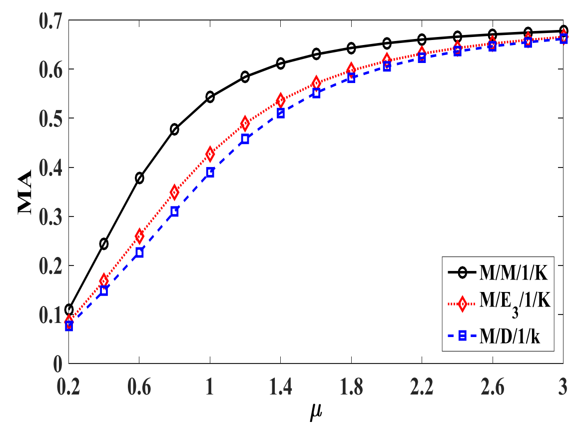


Figure 7.5: Effect of μ on MA

decreasing trend is noticed as λ varies from 1.2 to 2. In Figure 7.5, the trends for machine availability (MA) exhibit the increasing trend as μ increases.

7.5.3. Numerical Results for FM/FG/1/K Model

The fuzzy approach described in Section 7.4 is implemented for the performance analysis of multi-component machining system along with standby support and vacation. A number of real-life applications of the fuzzy model for multi-component machining system are often encountered in our day to day life. To reveal the practical applicability of fuzzy model for the performance modeling of multi-component machining system, we compute numerical results for the fuzzy model which fitted well for the performance analysis of flexible manufacturing system where robots are used for packing purpose.

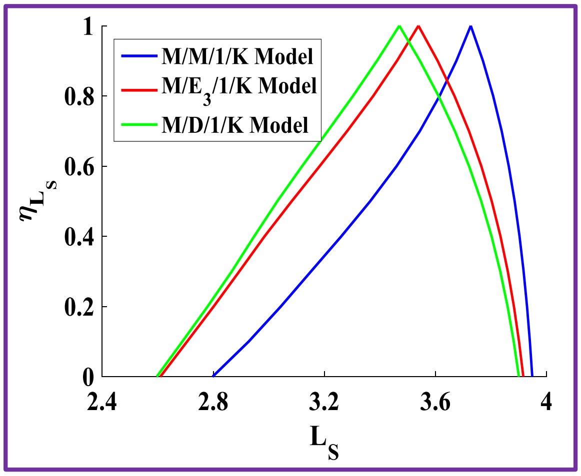


Figure 7.6: The fuzzy membership function for $\tilde{L}_S$

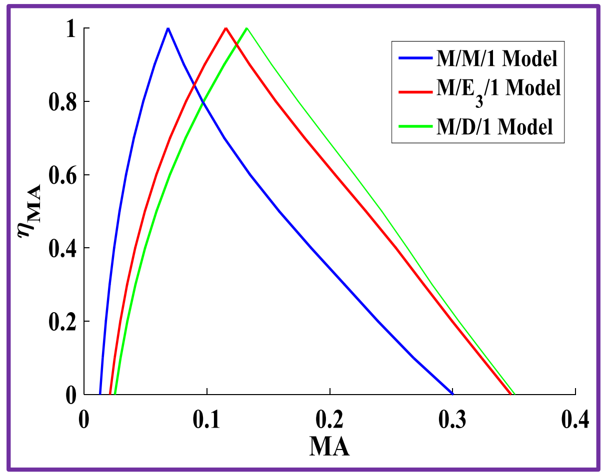


Figure 7.7: The fuzzy membership function for $M\tilde{A}$

For the computational purpose, we consider $K=4$, $M=2$ and $S=2$. The failure rate of robots in standby mode is $a=0.2$. In case when there is no failed robot in the system, the repairman goes for vacation with rate $\varepsilon=0.5$. The failure rate and repair rate of the robots, and vacation time of repairman are considered as triangular fuzzy numbers and are represented in units per day as $\tilde{\lambda}=[0.5,1,1.5]$, $\tilde{\mu}=[1,2,3]$, and $\tilde{\theta}=[0.5,0.7,0.9]$. In this practical example, the crisp interval $[(L_S)_\alpha^{LB},(L_S)_\alpha^{UB}]$ for the mean number of failed robots $(\tilde{L}_S)$ and crisp interval $[(MA)_\alpha^{LB},(MA)_\alpha^{UB}]$ for the machine (robot) availability $(M\tilde{A})$ are determined using Equations (7.42)-(7.43) and (7.44)-(7.45), respectively.

From (7.41), we have $[(p)_\alpha^{LB},(p)_\alpha^{UB}]=[0.5+0.5\alpha,1.5-0.5\alpha]$, $[(q)_\alpha^{LB},(q)_\alpha^{UB}]=[1+\alpha,3-\alpha]$, $[(r)_\alpha^{LB},(r)_\alpha^{UB}]=[0.5+0.2\alpha,0.9-0.2\alpha]$. Now, our aim is to determine the decision variables $(p,q\,\text{and}\,r)$ using Equations (7.42)-(7.45) for $\alpha=0(0.1)1$. The optimization problems given in (7.42)-(7.45), are highly nonlinear and complex as such the analytical solution of (7.42)-(7.45) in terms of α is not easy to found. To achieve the goal of determining $[(L_S)_\alpha^{LB},(L_S)_\alpha^{UB}]$ and $[(MA)_\alpha^{LB},(MA)_\alpha^{UB}]$, the interior point algorithm via '*fmincon*' function of the software MATLAB is implemented. The numerical results of three models viz. M/M/1/K, M/E$_3$/1/K and M/D/1/K for $[(L_S)_\alpha^{LB},(L_S)_\alpha^{UB}]$ and $[(MA)_\alpha^{LB},(MA)_\alpha^{UB}]$ are displayed in Tables 7.3-7.4, respectively. The membership functions for $\alpha=0(0.1)1$ are plotted for $\eta_{\tilde{L}_S}$ and $\eta_{M\tilde{A}}$ as shown in Figures 7.6 and 7.7, respectively.

Table 7.3: The $\alpha-$cuts of $\tilde{\lambda}, \tilde{\mu}, \tilde{\theta}$ and the mean number of failed robots (L_S) in the system

							M/M/1/K		M/E3/1/K		M/D/1/K	
α	p_α^{LB}	p_α^{UB}	q_α^{LB}	q_α^{UB}	r_α^{LB}	r_α^{UB}	$(L_s)_\alpha^{LB}$	$(L_s)_\alpha^{UB}$	$(L_s)_\alpha^{LB}$	$(L_s)_\alpha^{UB}$	$(L_s)_\alpha^{LB}$	$(L_s)_\alpha^{UB}$
0	0.50	1.50	1	3	0.50	0.90	2.80	3.95	2.61	3.92	2.60	3.90
0.1	0.55	1.45	1.1	2.9	0.52	0.88	2.93	3.94	2.71	3.90	2.69	3.88
0.2	0.60	1.40	1.2	2.8	0.54	0.86	3.04	3.93	2.80	3.88	2.78	3.86
0.3	0.65	1.35	1.3	2.7	0.56	0.84	3.15	3.92	2.89	3.86	2.87	3.83
0.4	0.70	1.30	1.4	2.6	0.58	0.82	3.26	3.90	2.98	3.83	2.95	3.80
0.5	0.75	1.25	1.5	2.5	0.60	0.8	3.36	3.88	3.08	3.80	3.03	3.76
0.6	0.80	1.20	1.6	2.4	0.62	0.78	3.46	3.86	3.18	3.76	3.12	3.72
0.7	0.85	1.15	1.7	2.3	0.64	0.76	3.54	3.84	3.28	3.72	3.21	3.67
0.8	0.90	1.10	1.8	2.2	0.66	0.74	3.61	3.81	3.38	3.67	3.30	3.61
0.9	0.95	1.05	1.9	2.1	0.68	0.72	3.67	3.77	3.46	3.61	3.39	3.54
1	1.00	1	2	2	0.7	0.7	3.73	3.73	3.54	3.54	3.47	3.47

Table 7.4: The $\alpha-$cuts of $\tilde{\lambda}, \tilde{\mu}, \tilde{\theta}$ and the machine (robots) availability (MA)

							M/M/1/K		M/E3/1/K		M/D/1/K	
α	p_α^{LB}	p_α^{UB}	q_α^{LB}	q_α^{UB}	r_α^{LB}	r_α^{UB}	$(MA)_\alpha^{LB}$	$(MA)_\alpha^{UB}$	$(MA)_\alpha^{LB}$	$(MA)_\alpha^{UB}$	$(MA)_\alpha^{LB}$	$(MA)_\alpha^{UB}$
0	0.5	1.5	1	3	0.5	0.9	0.01	0.30	0.02	0.35	0.03	0.35
0.1	0.55	1.45	1.1	2.9	0.52	0.88	0.02	0.27	0.02	0.32	0.03	0.33
0.2	0.60	1.40	1.2	2.8	0.54	0.86	0.02	0.24	0.03	0.30	0.04	0.31
0.3	0.65	1.35	1.3	2.7	0.56	0.84	0.02	0.21	0.03	0.28	0.04	0.28
0.4	0.70	1.30	1.4	2.6	0.58	0.82	0.02	0.18	0.04	0.25	0.05	0.26
0.5	0.75	1.25	1.5	2.5	0.60	0.80	0.03	0.16	0.05	0.23	0.06	0.24
0.6	0.80	1.20	1.6	2.4	0.62	0.78	0.03	0.14	0.06	0.20	0.07	0.22
0.7	0.85	1.15	1.7	2.3	0.64	0.76	0.04	0.11	0.07	0.18	0.08	0.20
0.8	0.90	1.10	1.8	2.2	0.66	0.74	0.05	0.10	0.08	0.16	0.10	0.17
0.9	0.95	1.05	1.9	2.1	0.68	0.72	0.06	0.08	0.10	0.13	0.11	0.15
1	1	1	2	2	0.7	0.7	0.07	0.07	0.12	0.12	0.13	0.13

In Table 7.4, it is noticed that at $\alpha = 0$, the mean number of failed robots in the system can never cross the interval [2.80, 3.95] for M/M/1/K model, [2.61, 3.92] for M/E$_3$/1/K model, [2.60, 3.90] for M/D/1/K model. It is also found that at $\alpha = 1$, the mean number of robots in the system are 3.73 for M/M/1/K model, 3.54 for M/E$_3$/1/K model and 3.47 for M/D/1/K model. It is noticed that the mean number of robots in the system for M/M/1/K model is always greater than that of M/E$_3$/1/K, M/D/1/K. Moreover, if the decision maker specifies the mean number of failed robots between 3.36 and 3.88 for M/M/1/K model then α - level possibility should be 0.5 and desired range for arrival rate, repair rate and vacation rate are [0.75, 1.25], [1.5, 2.5] and [0.6, 0.8], respectively.

7.5.4. Cost Analysis

The cost-benefit analysis plays a significant role to improve the future system design. In the competitive world of industrial scenarios where failures/faults of machines cannot be avoided, the repair rate and vacation rate should be optimized by considering it as decision variables so as to make the system operative at economic cost. The cost function $TC(\mu,\theta)$ seems to be non-linear and complex in nature; therefore it is quite a tedious task to optimize such a function analytically. Here, we determine the optimal repair rate (μ^*) and optimal vacation rate (θ^*) by minimizing the cost function $TC(\mu,\theta)$ of the concerned system for M/M/1/K, $M/E_3/1/K$ and M/D/1/K models using harmony search algorithm.

To minimize the cost function $TC(\mu,\theta)$, the variations in the values of μ and θ are considered in the feasible range 0.1 to 2 and 0.5 to 2, respectively. In order to evaluate cost function $TC(\mu,\theta)$ using harmony search algorithm, each harmony is a two-dimensional vector; the first dimension corresponds to μ and the second dimension corresponds to θ. To determine the optimal repair rate (μ^*) and optimal vacation rate (θ^*) using harmonic search (HS) algorithm for the following cost sets:

Cost Set I: $C_H = \$70,\ C_B = \$60,\ C_V = \$30,\ C_m = \$45.$

Cost Set II: $C_H = \$80,\ C_B = \$75,\ C_V = \$30,\ C_m = \$55.$

Cost Set III: $C_H = \$90,\ C_B = \$75,\ C_V = \$35,\ C_m = \$55.$

HS approach for the cost optimization is used by fixing the default parameters as chosen for the sensitivity analysis in Subsection 7.5.2.

Initially, each component of harmony is initialized within the range 0.1 to 2 and 0.5 to 2 for μ and θ respectively, and is evaluated using objective function given in (7.28). The new harmony (H) is generated either using harmony memory or using randomization, and evaluated using objective function and replaces the worst harmony from HM in case it is better (Steps 18 and 19 of Algorithm 1 in Section 1.4.8 of Chapter 1). The above procedure is repeated for 100 generations. The values of parameters HMCR, PAR and HMS have been chosen as 0.9, 0.3 and 5, respectively. Harmony search algorithm has been implemented in Matlab and numerical results for the costs $TC(\mu^*,\theta^*)$ and (μ^*,θ^*) for three models M/M/1/K, $M/E_3/1/K$ and M/D/1/K are depicted in Figures 7.8-7.10, respectively.

Table 7.5: (μ^*,θ^*) and $TC(\mu^*,\theta^*)$ (in $) for different cost sets

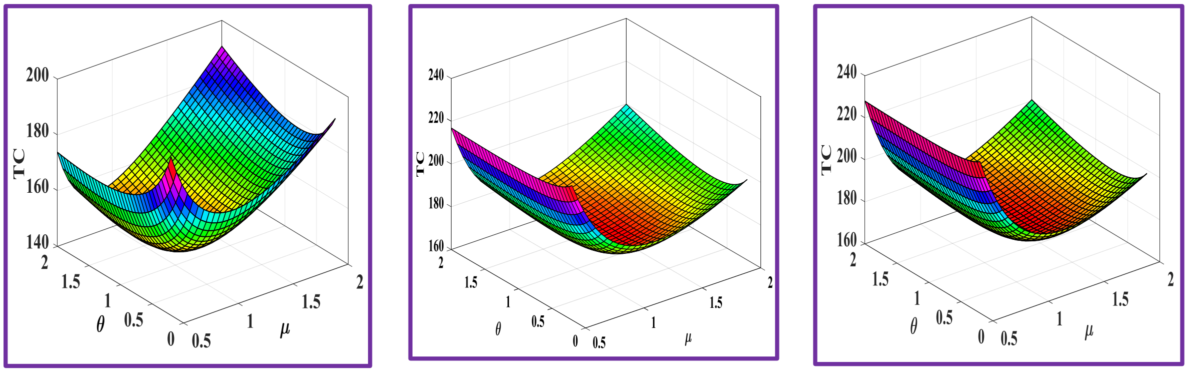
Cost Set	M/M/1/K			M/E3/1/K			M/D/1/K		
	μ^*	θ^*	$TC(\mu^*,\theta^*)$	μ^*	θ^*	$TC(\mu^*,\theta^*)$	μ^*	θ^*	$TC(\mu^*,\theta^*)$
I	0.95	1.05	148.20	1.15	0.90	172.33	1.25	0.85	174.94
II	0.90	1.10	170.64	1.15	0.90	201.29	1.25	0.90	204.51
III	0.95	1.05	184.55	1.20	0.90	215.17	1.25	0.85	218.37

Table 7.6: (μ^*,θ^*) and $TC(\mu^*,\theta^*)$ (in$) for varying values of λ

λ	M/M/1/K			M/E3/1/K			M/D/1/K		
	μ^*	θ^*	$TC(\mu^*,\theta^*)$	μ^*	θ^*	$TC(\mu^*,\theta^*)$	μ^*	θ^*	$TC(\mu^*,\theta^*)$
0.2	1.01	1.27	171.81	1.24	0.84	201.51	1.25	0.84	204.50
0.4	1.04	0.51	277.34	1.26	0.63	291.42	1.29	0.63	296.43
0.6	0.79	0.72	339.43	1.04	0.71	346.33	1.11	0.61	351.53
0.8	0.57	1.67	364.28	0.86	0.87	372.43	1.01	0.70	378.46

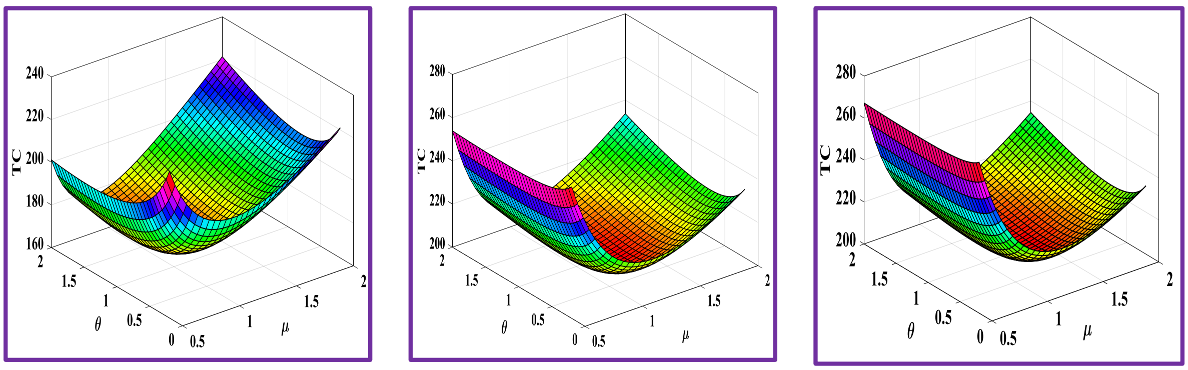
The surface graphs are plotted for the cost function $TC(\mu,\theta)$ to demonstrate the variability as well as computational tractability for real-time systems. The trends of total cost $TC(\mu,\theta)$ by varying parameters μ and θ are depicted in Figures 7.8-7.10 for M/M/1/K, M/E3/1/K, and M/D/1/K models. It is noticed that the cost function has the convex nature with respect to the repair rate (μ) and vacation rate (θ). The optimal system cost $TC(\mu^*,\theta^*)$ and the corresponding optimal parameters (μ^*,θ^*) obtained for three cost sets are displayed in Tables 7.5-7.6.

The comparative study of minimum cost $TC(\mu^*,\theta^*)$ is done using harmony search algorithm among the three models by varying values of λ. Based on the numerical results, the order of total cost is found as $TC_{M/M/1/K} < TC_{M/E_3/1/K} < TC_{M/D/1/K}$, which can be noticed in the Figures 7.8-7.10 and Tables 7.5-7.6 also.



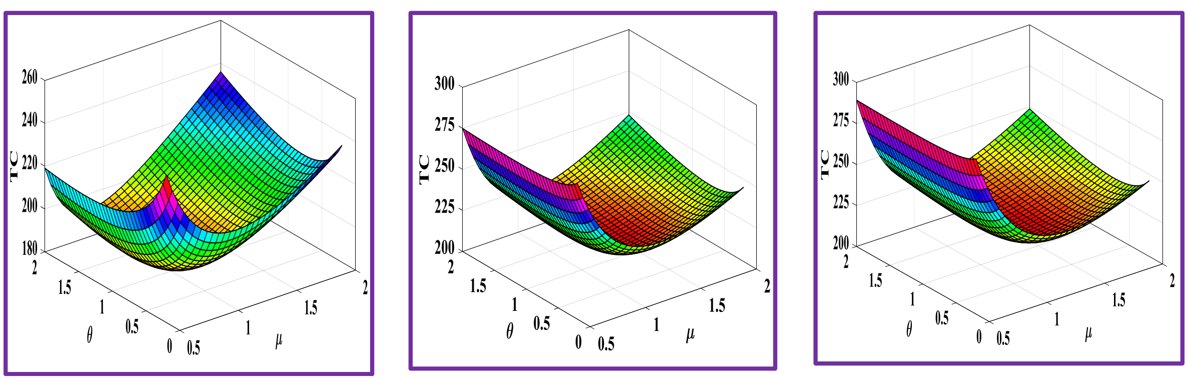
(a) (b) (c)

Figure 7.8: TC vs (μ, θ) (a) M/M/1/K, (b) M/E_3/1/K, and (c) M/D/1/K for Cost Set-I



(a) (b) (c)

Figure 7.9: TC vs (μ, θ) (a) M/M/1/K, (b) M/E_3/1/K, and (c) M/D/1/K for Cost Set-II



(a) (b) (c)

Figure 7.10: TC vs (μ, θ) (a) M/M/1/K, (b) M/E_3/1/K, and (c) M/D/1/K for Cost Set-III

7.6. CONCLUDING REMARKS

The performance analysis of the finite population M/G/1/K multi-component machining system with server vacation and standby support in both crisp, as well as the fuzzy environment, is done. Using the analytical technique based on recursive and supplementary variable approaches, the queue size distribution and system metrics are derived in explicit form. The multi-components machining system with general repair time is then investigated in the fuzzy environment by taking system parameters as the fuzzy numbers. The meta-heuristic, harmony search algorithm is implemented to minimize the cost function which demonstrates the potential applicability and tractability of it in the real-time system. The cost optimization done reveals the scope of the model for the fault-tolerant machining system operating with standbys and repair support and provision of server vacation.

Chapter 8

Customers' Joining Strategies for Double Orbit Retrial Queue

8.1. INTRODUCTION

In many real life congestion scenarios, it is noticed that the customers do not like to wait in front of the unreliable server so that the system organizer provides them facility of waiting hall; from there they can re-attempt their request for the service (cf. Artalejo and Falin, 2002). In the queueing literature, a few researchers have presented their pioneer works on the unreliable server retrial queue (cf. Chang and Wang, 2018; Choudhury et al., 2015; Yang et al., 2016). The balking behavior of the customers in case of unreliable server retrial queue is a common issue in our daily life; this situation may be realized at a number of places. There are scarcity of works on unreliable server retrial queue with balking. Most recently, Ke et al. (2019) analyzed a multiple server Markov model for retrial queue with vacation and balking by using the matrix geometric method. In the queueing literature, some researchers have focused on retrial queue with feedback (cf. Kumar and Sharma, 2014; Tao et al., 2014). Singh et al. (2017) dealt with the non-Markov bulk input queueing model by incorporating both discouraged behavior and feedback. Chang et al. (2018) considered an unreliable server retrial queueing model along with customers' impatient and feedback behaviour. They investigated Markov model by using quasi-birth and death processes. Based on the linear cost-reward structure, Hemachandra and Narahari (2000) considered an MMPP/GI/1 queue and analyzed two optimization problems i.e. profit per unit time (PUT) and profit per accepted customers (PAC).

The single server crisp queueing model with retrial attempts can be further extended to single server double orbit fuzzy queueing model with balking behavior of the customers, server breakdown and customer's feedback. This extended study may have wider applications in a real time systems and would be beneficial to the decision makers is facilitating better service for which the customers are agreed to pay. The work on double retrial orbit in queueing literature is rarely found. The applicability of double orbit retrial queue in our daily life may be noticed including railway station where two types of waiting rooms (ordinary and premium) are available for the arriving passengers. Double orbit retrial queueing model with priority customers was presented by Jain and Bhagat (2013). They presented a transient solution of the queue size distribution by using the Runge-Kutta 4^{th} order method. Jain et al. (2015) dealt with a single unreliable server finite queueing model with double orbit retrial. They categorized customers into two classes namely, priority and non-priority and then steady state analysis is done based on the matrix method.

The fuzzy model for double orbit retrial queue deals with more versatile and realistic queueing scenarios. Many researchers working in the area of queueing theory have contributed towards crisp retrial queueing models but a very few researchers have paid their attention to study the fuzzy retrial queueing model (cf. Ke et al., 2006, 2007; Kalayanaraman et al, 2010; Yang and Chang, 2015). The applicability of fuzzy retrial queueing models is more realistic as compared to crisp retrial queue models from an application viewpoint. Rao et al. (2007) studied a test interval optimization problem using fuzzy-genetic approach. In fuzzy environment, Sharma et al. (2017) considered trapezoidal fuzzy number to frame the cost function of inventory model so as to determine the optimal values of the system descriptors. In the crisp retrial queueing model, the system parameters namely arrival rate, service rate, retrial rate, etc. have a constant value whereas in fuzzy retrial queueing model these parameters are represented in linguistic way. The system parameters used in the fuzzy queue are possibilistic and probabilistic (cf. Buckeley, 2004). The priority-based fuzzy queueing model was studied by Bouazzi et al. (2017). They used the fuzzy logic controller in wireless sensor networks to propose a fuzzy logic algorithm. Bhardwaj et al. (2018) proposed a Markov model with two queues in series and reneged customers in fuzzy environment.

In the queueing literature, the single server double orbit retrial queue with customers' balking behavior has not appeared earlier while it has a variety of applications in our routine life including at repair workshops and communications systems, etc. In this chapter, we deal with the double orbit queueing models with balking behavior of the customers and unreliable server model. In Section 8.2, the single server double orbit queue with customers' joining strategy has been investigated in both crisp environments. Section 8.3 is devoted to the optimal management of double orbit retrial queue with balking. Further, the model discussed in Section 8.2 is transformed into fuzzy environment and analyzed in Section 8.4. In Section 8.5, the double orbit queue is discussed by incorporating the concept of feedback. In Section 8.6, the concept of unreliable server is incorporated to develop double orbit retrial queueing model along with customer's balking behaviour. The soft computing based ANFIS technique is also used to validate the numerical results for unreliable server model. In Section 8.7, the concluding remarks of the investigation done are given.

8.2. DOUBLE ORBIT RETRIAL MODEL WITH BALKING

In the present model, we study the retrial queueing model by incorporating the concepts of double orbit along with balking in both crisp and fuzzy environment. The applications of such queueing model can be seen at many places including automobile industries, shopping malls, communication system, etc.

8.2.1. Model Description of Double Orbit Model

Consider a single server infinite capacity retrial queueing model with balking. To develop Markov model, the concept of double orbit namely, ordinary orbit and premium orbit is incorporated. The following assumptions are made to construct the mathematical model:

(i) The arriving customers join the system in Poisson fashion with rate λ.

(ii) On arrival, the customers decide whether to join the queue or not. It is assumed that the arriving customers join the system with probability q and balk with probability $(1-q)$.

(iii) Reaching at the service center, if the customer finds the server busy then there are two options namely, either he waits in the queue or join the ordinary or premium orbit with probability σ or $(1-\sigma)$. Those whose paying capacity is less, joins the ordinary orbit and those who are in the position to pay high, join the premium orbit.

(iv) After a random period of time, both types of customers re-attempt for the service according to exponential distribution with mean $1/\gamma$.

(v) The server facilitates the service to the ordinary customers and the premium customers according to an exponential distribution with rates μ_1 and μ_2, respectively.

(vi) The customers present in the queue/orbit are served according to first-come-first-served (FCFS) rule.

For the convenience of notations, we define $\mu_O = \sigma\mu_1$ and $\mu_P = (1-\sigma)\mu_2$.

Let $N(\tau)$ and $S(\tau)$ be the random variables denoting the number of customers in the system and status of the server at any time τ, respectively. Now, the three levels of system states are defined as follows:

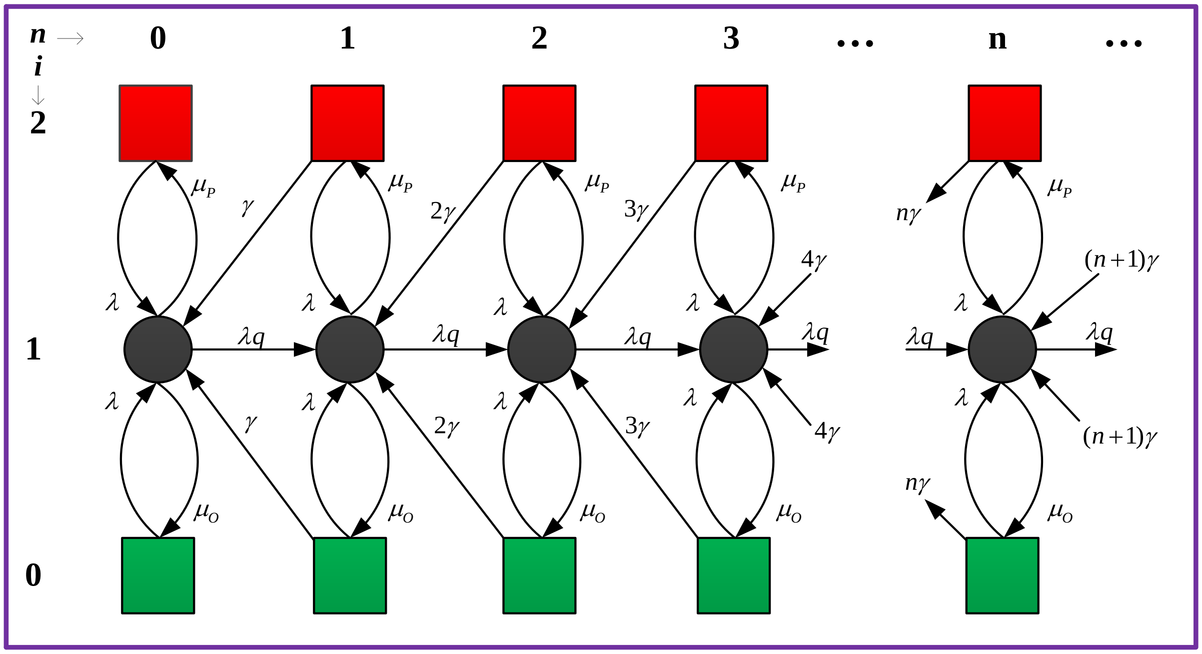


Figure 8.1: Transition state diagram for double orbit model

$$S(\tau) = i = \begin{cases} 0(2), & \text{when the server is free and the customers are residing in the ordinary} \\ & \text{(premium) orbit} \\ 1, & \text{when the server is busy.} \end{cases}$$

We consider a bi-variate stochastic process $\{S(\tau), N(\tau) : \tau \geq 0\}$ which portrays the continuous time Markov chain (CTMC) with the state space $\{0,1,2\} \times \{0,1,2,...\}$. Markov retrial model is developed in terms of steady state probabilities $P_{i,n} = \lim_{\tau \to \infty} P_{i,n}(\tau), i = 0,1,2.$

It can be easily verified that the system is stable if (cf. Gross et al., 2008) $\rho = \dfrac{\lambda q}{\mu_O + \mu_P} < 1.$

8.2.2. The Governing Equations and Analysis

For the mathematical analysis of the model, the steady state Chapman-Kolmogorov (C-K) equations are constructed for the system states at three levels $i = 0,1$ and 2 by balancing the in-flows and out-flows of different states as follows:

$$-(\lambda + n\gamma)P_{0,n} + \mu_o P_{1,n} = 0,\ n = 0,1,2,... \tag{8.1}$$

$$-(\lambda q + \mu_o + \mu_P)P_{1,0} + \lambda P_{0,0} + \lambda P_{2,0} + \gamma P_{0,1} + \gamma P_{2,1} = 0 \tag{8.2}$$

$$-(\lambda q + \mu_o + \mu_P)P_{1,n} + \lambda q P_{1,n-1} + \lambda P_{0,n} + \lambda P_{2,n} + (n+1)\gamma P_{0,n+1} + (n+1)\gamma P_{2,n+1} = 0,\ n = 1,2,... \tag{8.3}$$

$$-(\lambda + n\gamma)P_{2,n} + \mu_P P_{1,n} = 0,\ n = 0,1,2,... \tag{8.4}$$

The probability generating functions (PGFs) $\Pi_i(z)$ where $|z| \le 1$ for the three levels $i = 0,1$ and 2, *i.e.* when the customers are in ordinary orbit, busy state and in the premium orbit respectively, are defined as follows:

$$\Pi_0(z) = \sum_{n=0}^{\infty} P_{0,n} z^n \,, \tag{8.5}$$

$$\Pi_1(z) = \sum_{n=0}^{\infty} P_{1,n} z^n \tag{8.6}$$

and $\Pi_2(z) = \sum_{n=0}^{\infty} P_{2,n} z^n$. (8.7)

Lemma 8.1: The probability generating functions $\Pi_0(z), \Pi_1(z)$ and $\Pi_2(z)$ are given by

(i) $$\Pi_0(z) = \frac{\Lambda \mu_O}{\mu_O + \mu_P}(1-\rho z)^{\frac{-\lambda}{\gamma}} \tag{8.8}$$

(ii) $$\Pi_1(z) = \frac{\Lambda \lambda}{\mu_O + \mu_P}(1-\rho z)^{\frac{-\lambda}{\gamma}-1} \tag{8.9}$$

(iii) $$\Pi_2(z) = \frac{\Lambda \mu_P}{\mu_O + \mu_P}(1-\rho z)^{\frac{-\lambda}{\gamma}} \tag{8.10}$$

where $\Lambda = \dfrac{(1-\rho)^{\frac{\lambda}{\gamma}+1}}{(1-\rho)+\dfrac{\rho}{q}}$.

Proof: Multiplying (8.1) and (8.4) by z^n and summing over n, we get

$$\lambda \Pi_0(\mathrm{z}) + \gamma \mathrm{z} \Pi_0'(\mathrm{z}) = \mu_O \Pi_1(\mathrm{z}) \tag{8.11}$$

$$\lambda \Pi_2(\mathrm{z}) + \gamma \mathrm{z} \Pi_2'(\mathrm{z}) = \mu_P \Pi_1(\mathrm{z}) \tag{8.12}$$

where $\Pi'(z)$ is the first derivative of $\Pi(z)$ with respect to *z*.

Multiplying (8.3) by z^n and summing over n, and then adding (8.2), we have

$$(\lambda q + \mu_O + \mu_P - \lambda q z)\Pi_1(\mathrm{z}) = \lambda \Pi_0(\mathrm{z}) + \gamma \Pi_0'(\mathrm{z}) + \lambda \Pi_2(\mathrm{z}) + \gamma \Pi_2'(\mathrm{z}) \tag{8.13}$$

Denote $\Pi(z) = \Pi_0(z) + \Pi_2(z)$ (8.14)

Adding (8.11) and (8.12) and using (8.14), we find

$$\lambda\Pi(z) + \gamma z\Pi'(z) = (\mu_O + \mu_P)\Pi_1(z) \tag{8.15}$$

Now (8.13) and (8.14) yield,

$$(\lambda q + \mu_O + \mu_P - \lambda q z)\Pi_1(z) = \lambda\Pi(z) + \gamma\Pi'(z) \tag{8.16}$$

Solving (8.15) for $\Pi_1(z)$ and then substituting into (8.16), we get

$$\Pi'(z) = \frac{\rho\lambda}{\gamma(1-\rho z)}\Pi(z) \tag{8.17}$$

Equation (8.17) yields

$$\Pi(z) = k(1-\rho z)^{\frac{-\lambda}{\gamma}}, \ k \text{ being constant.} \tag{8.18}$$

Using (8.15), (8.17) and (8.18), we get

$$\Pi_1(z) = \frac{k\rho}{q}(1-\rho z)^{\frac{-\lambda}{\gamma}-1} \tag{8.19}$$

Adding (8.18) and (8.19) and then using normalizing condition $\sum_{i=0}^{2}\Pi_i(1) = 1$, the constant $k = \Lambda$.

Also, $\Pi_1(z)$ given in (8.9) is obtained by using the value of k into (8.19). Substituting $\Pi_1(z)$ from (8.9) into (8.11) and (8.12) respectively, and solving for $\Pi_0(z)$ and $\Pi_2(z)$, probability generating functions given in (8.8) and (8.10) are obtained.

Theorem 8.1: The steady state probabilities for the double orbit single server queueing model with balking are given by

$$P_{i,n} = \begin{cases} \dfrac{\Lambda\mu_O}{\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n \prod_{j=0}^{n-1}(\lambda + j\gamma)\right], & i=0,\ n=0,1,2,..., \\ \dfrac{\Lambda\lambda}{\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n \prod_{j=0}^{n-1}(\lambda + (j+1)\gamma)\right], & i=1,\ n=0,1,2,..., \\ \dfrac{\Lambda\mu_P}{\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n \prod_{j=0}^{n-1}(\lambda + j\gamma)\right], & i=2,\ n=0,1,2,.... \end{cases} \tag{8.20}$$

where $\prod_{j=n}^{n-1} x_j = 1; n = 0,1,...$.

Proof: By expanding $(1-\rho z)^{\frac{-\lambda}{\gamma}}$ and $(1-\rho z)^{\frac{-\lambda}{\gamma}-1}$ given in (8.8)-(8.10) and then collecting the coefficients of z^n, we obtain results given in (8.20).

Lemma 8.2: The long run probabilities for the idle server and the busy server are respectively, given by

$$P_I = \frac{q(1-\rho)}{q(1-\rho)+\rho} \tag{8.21}$$

$$P_{SB} = \frac{\rho}{q(1-\rho)+\rho} \tag{8.22}$$

Proof: Adding (8.8) and (8.10) and putting $z=1$, we obtain (8.21). The result given in (8.22) can be derived from (8.9) by substituting $z=1$.

Theorem 8.2: When the customer enters into the system with probability q, the expected number of customers in the ordinary orbit, in front of the server, in the premium orbit and in the system are respectively, given by

(i) $$E[N_O] = \frac{\mu_O}{\gamma}\frac{\rho^2}{q(1-\rho)+\rho} \tag{8.23}$$

(ii) $$E[N_q] = \frac{\rho^2}{1-\rho}\frac{1}{q(1-\rho)+\rho}\left(\frac{\lambda+\gamma}{\gamma}\right) \tag{8.24}$$

(iii) $$E[N_P] = \frac{\mu_P}{\gamma}\frac{\rho^2}{q(1-\rho)+\rho} \tag{8.25}$$

(iv) $$E[N_S] = \frac{\rho^2}{q(1-\rho)+\rho}\left[\frac{\mu_O+\mu_P}{\gamma}+\frac{1}{1-\rho}\left(\frac{\lambda+\gamma}{\gamma}\right)+\frac{1}{\rho}\right] \tag{8.26}$$

Proof: Differentiating (8.8), (8.9) and (8.10) with respect to z, and then putting $z=1$, we obtain (8.23), (8.24) and (8.25). Expected number of customers in the system $(E[N_S])$ given in (8.26) is obtained by summing the total number of customers in the queue and average customers in the service.

Remark 8.1: Expected waiting time in the system $(E[W_S])$ is determined using Little's formula given by $\frac{E[N_S]}{\lambda_{eff}}$

$$E[W_S]=\frac{\rho^2}{\lambda q}\left[\frac{\mu_O+\mu_P}{\gamma}+\frac{1}{1-\rho}\left(\frac{\lambda+\gamma}{\gamma}\right)+\frac{1}{\rho}\right] \tag{8.27}$$

where $\lambda_{eff}(q)=\frac{\lambda q}{q(1-\rho)+\rho}$. (8.28)

Remark 8.2: Total number of customers in the queue $(E[N_T])$ is given by the sum of customers waiting in the ordinary orbit $(E[N_O])$, premium orbit $(E[N_P])$ and in front of the server $(E[N_q])$. Thus

$$E[N_T]=\frac{\rho^2}{q(1-\rho)+\rho}\left(\frac{\mu_O+\mu_P}{\gamma}+\left(\frac{\lambda+\gamma}{\gamma}\right)\frac{1}{1-\rho}\right) \tag{8.29}$$

8.2.3. Numerical Results of Double Orbit Retrial Model with Balking

In this section, the sensitivity analysis of system parameters on various performance measures is conducted by taking a numerical illustration. For the computation of numerical results, MATLAB software is used by setting default parameters as $\lambda=22$, $\mu_1=20$, $\mu_2=30$, $\gamma=10$, $q=0.75$, $\sigma=0.5$ and cost elements per unit time as $C_O=60$, $C_P=120$, $C_H=30$, $C_I=30$, $C_B=110$, $C_1=50$, $C_2=50$.

The numerical experiments are conducted by varying the values of $\lambda,\mu_1,\mu_2,\gamma,q$ and the numerical values of the expected total number of customers in the queue $(E[N_T])$, expected waiting time spent by the customers in the system $(E[W_S])$ and long run probabilities namely, server being idle (P_I) and busy (P_{SB}) are recorded in Tables 8.1-8.4. Also, the trends of the expected number of customers in the system $(E[N_S])$ are shown in Figures 8.2-8.5 against $\mu_1,\mu_2,\lambda,\gamma$ respectively, for varying the value of q.

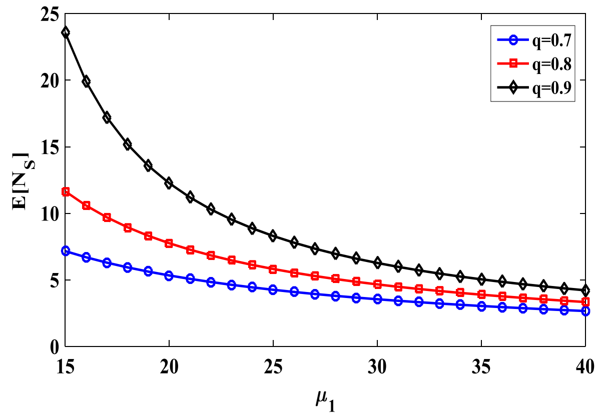


Figure 8.2: $E[N_S]$ vs μ_1 for $q = 0.7, 0.8, 0.9$

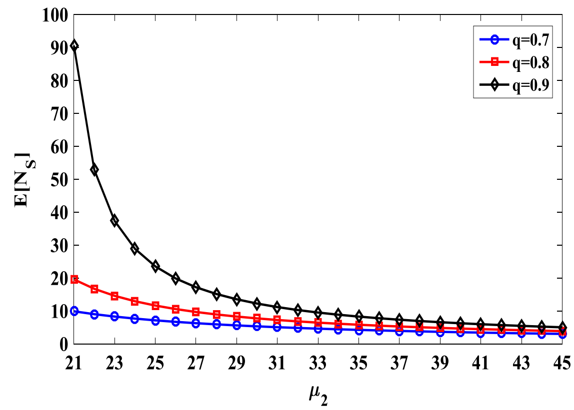


Figure 8.3: $E[N_S]$ vs μ_2 for $q = 0.7, 0.8, 0.9$

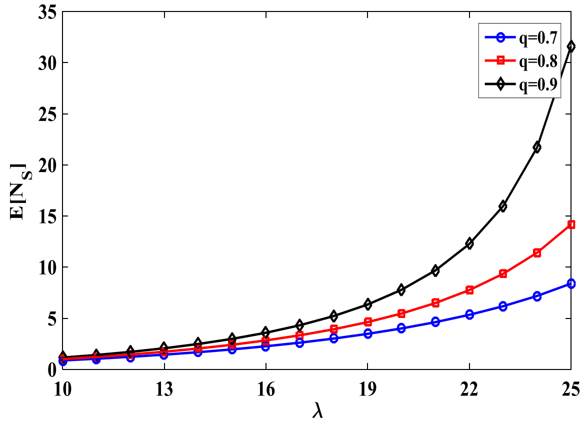


Figure 8.4: $E[N_S]$ vs λ for $q = 0.7, 0.8, 0.9$

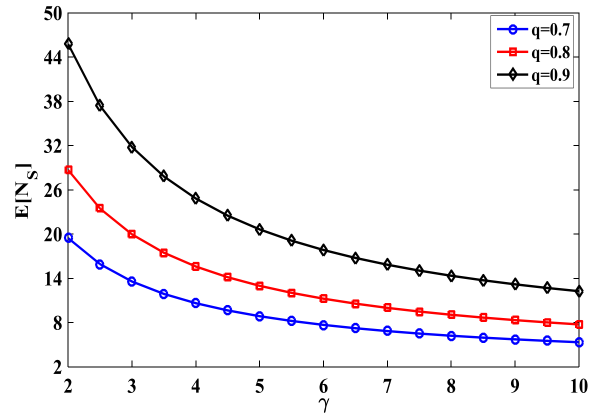


Figure 8.5: $E[N_S]$ vs γ for $q = 0.7, 0.8, 0.9$

Table 8.1: Effect of q and μ_1 on various performance measures

q	μ_1	$E[N_T]$	$E[W_S]$	P_{SB}	P_I	TC
	16	5.97	0.39	0.74	0.26	1497.50
0.7	24	3.79	0.25	0.65	0.35	1609.94
	32	2.76	0.18	0.59	0.41	1762.48
	16	9.78	0.57	0.80	0.20	1635.54
0.8	24	5.44	0.32	0.70	0.30	1677.08
	32	3.7	0.22	0.62	0.38	1805.46
	16	19.02	0.99	0.87	0.13	1939.77
0.9	24	8.12	0.44	0.75	0.25	1778.41
	32	5.06	0.28	0.66	0.34	1862.29

Table 8.2: Effect of q and μ_2 on various performance measures

q	μ_2	$E[N_T]$	$E[W_S]$	P_{SB}	P_I	TC
	25	6.42	0.42	0.76	0.24	1481.57
0.7	30	4.64	0.31	0.70	0.30	1545.84
	35	3.62	0.24	0.65	0.35	1632.99
	25	10.84	0.63	0.82	0.18	1636.78
0.8	30	7.02	0.41	0.75	0.25	1637.15
	35	5.13	0.31	0.69	0.31	1697.05
	25	22.66	1.18	0.89	0.11	2017.90
0.9	30	11.46	0.61	0.81	0.19	1794.10
	35	7.56	0.41	0.74	0.26	1791.12

Table 8.3: Effect of q and λ on various performance measures

q	λ	$E[N_T]$	$E[W_S]$	P_{SB}	P_I	TC
	15	1.45	0.15	0.51	0.49	1399.55
0.7	20	3.36	0.25	0.65	0.35	1492.22
	25	7.63	0.44	0.77	0.23	1659.23
	15	1.89	0.18	0.54	0.46	1421.66
0.8	20	4.78	0.32	0.69	0.31	1551.45
	25	13.34	0.68	0.83	0.17	1856.67
	15	2.43	0.21	0.57	0.43	1448.48
0.9	20	7.04	0.42	0.74	0.26	1638.69
	25	30.69	1.39	0.91	0.09	2408.18

Table 8.4: Effect of q and γ on various performance measures

q	γ	$E[N_T]$	$E[W_S]$	P_{SB}	P_I	TC
	2	18.76	1.12	0.70	0.30	2252.39
0.7	6	6.99	0.44	0.70	0.30	1663.60
	10	4.64	0.31	0.70	0.30	1545.84
	2	27.94	1.53	0.75	0.25	2612.73
0.8	6	10.51	0.60	0.75	0.25	1799.75
	10	7.02	0.41	0.75	0.25	1637.15
	2	44.96	2.26	0.81	0.19	3222.12
0.9	6	17.04	0.88	0.81	0.19	2032.11
	10	11.46	0.61	0.81	0.19	1794.10

It is assumed that both ordinary, as well as premium customers, join the system with constant arrival rate and the system organizer facilitates the waiting spaces as per choice of the customers in ordinary and premium orbits. If the customers are served with faster service rate, then $E[N_T]$, $E[W_S]$ and P_{SB} (P_I) would be decreased (increased) and this fact has been reported in Tables

8.1-8.2. Figures 8.2 and 8.3 reveal that $E[N_S]$ decreases on increasing the service rates μ_1 and μ_2, respectively. But $E[N_S]$ increases with respect to q; however later on, it becomes asymptotically constant when μ_1 and μ_2 become high. Table 8.3 depicts that $E[N_T], E[W_S]$ and P_{SB} rise but P_I downfalls as λ increases by keeping the service rate μ_1 and μ_2 as constants. Also, an increment in $E[N_S]$ is reported (see Figure 8.4) with λ. As the customers re-attempt for the service with increasing γ, there seems to be a downfall in $E[N_T]$ and $E[W_S]$ as can be seen in Table 8.4. Similarly, Figure 8.5 shows that $E[N_S]$ becomes less as γ goes up and at a later stage for higher γ, it becomes asymptotically constant.

8.3. OPTIMAL MANAGEMENT OF DOUBLE ORBIT RETRIAL QUEUE

This section presents the optimal admission control policy for the double orbit retrial queue with balking. We analyze the customers' joining/balking strategy based on profit function. The system cost per unit time is also optimized to determine the optimal service rates.

8.3.1. Optimal Joining Strategy and Profit Function

In the section, we formulate the profit function to determine the optimal joining probability of the customer for the single server double orbit retrial queue. We assume that the arriving customers upon joining the system, are facilitated by the single server and then they receive reward R units for their satisfaction. On the other hand, each customer has to pay a waiting cost C per unit of time when they remain in the system to receive the service. Arriving customers before joining the system, estimate the expected reward related to the service received and the expected waiting cost as per information gathered from the system and then decide whether to queue-up or balk. In particular, customers always prefer to join the system if the reward is greater than the expected waiting cost and are neutral when the rewards are equal to the expected waiting cost. Now, we construct the profit function $f(q)$ for the single server double orbit retrial queue as follows:

$$f(q) = \lambda_{eff}(q)R - C.E[N_T] \tag{8.30}$$

where $\lambda_{eff}(q)$ and $E[N_T]$ are given by (8.28) and (8.29), respectively.

After getting served, the customers are strictly preferred to receive maximum profit from the system organizer. Our aim is to maximize the profit function given in (8.30) corresponding to joining probability q. Thus, the optimization problem is constructed as follows:

$$\max f(q) = f(q^*) \text{ where } 0 \le q \le 1 \tag{8.31}$$

Theorem 8.3: In the single server double orbit retrial queue, when the server is busy, the customers join the system with probability q^* that maximizes the profit per unit time given by

$$q^* = \begin{cases} 0, & 0 < \dfrac{R}{C} < u \\ \dfrac{\sigma\mu_1 + (1-\sigma)\mu_2 - \omega_1}{\lambda}, & u \le \dfrac{R}{C} \le v \\ 1, & v < \dfrac{R}{C}, \end{cases} \tag{8.32}$$

where $u = \dfrac{(\lambda + \sigma\mu_1 + (1-\sigma)\mu_2)(\lambda + \sigma\mu_1 + (1-\sigma)\mu_2 + \gamma)}{\gamma(\sigma\mu_1 + (1-\sigma)\mu_2)^2}$,

$$v = \frac{(\sigma\mu_1 + (1-\sigma)\mu_2)^2(\sigma\mu_1 + (1-\sigma)\mu_2 + \gamma) + \lambda\gamma(\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda)}{\gamma(\sigma\mu_1 + (1-\sigma)\mu_2)(\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda)^2},$$

$$\omega_1 = \frac{C(\sigma\mu_1 + (1-\sigma)\mu_2)(\lambda + \gamma) + \sqrt{X}}{C(\gamma - \sigma\mu_1 - (1-\sigma)\mu_2)) + \gamma RC(\sigma\mu_1 + (1-\sigma)\mu_2)},$$

$$\begin{aligned} X &= (C(\sigma\mu_1 + (1-\sigma)\mu_2)(\lambda + \gamma))^2 + \lambda C(\sigma\mu_1 + (1-\sigma)\mu_2)(\lambda + \gamma)(C(\gamma - \sigma\mu_1 - (1-\sigma)\mu_2)) \\ &\quad + \gamma RC(\sigma\mu_1 + (1-\sigma)\mu_2)). \end{aligned}$$

Proof: Substituting $\lambda_{eff}(q)$ and $E[N_T]$ from (8.28) and (8.29) into (8.30), we have

$$f(q) = \frac{\lambda q R}{q(1-\rho) + \rho} - \frac{C\rho^2}{q(1-\rho) + \rho}\left(\frac{\sigma\mu_1 + (1-\sigma)\mu_2}{\gamma} + \left(\frac{\lambda + \gamma}{\gamma}\right)\frac{1}{1-\rho}\right) \tag{8.33}$$

After some algebra, (8.33) yields

$$f(q) = \frac{\lambda(\sigma\mu_1 + (1-\sigma)\mu_2)R}{\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda q + \lambda} - \frac{C\lambda^2(q\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda q + \lambda + \gamma)}{\gamma(\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda q + \lambda)(\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda q)} \tag{8.34}$$

Now, (8.34) can be rewritten as

$$f(q) = \frac{\lambda C}{\gamma} + \frac{\lambda[((\gamma - \sigma\mu_1 - (1-\sigma)\mu_2)C + (\sigma\mu_1 + (1-\sigma)\mu_2)\gamma R)(\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda q) - C(\lambda + \gamma)(\sigma\mu_1 + (1-\sigma)\mu_2)]}{\gamma(\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda q)(\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda q + \lambda)} \tag{8.35}$$

Fsor the brevity of notation, we assume $\omega = \sigma\mu_1 + (1-\sigma)\mu_2 - \lambda q.$ Now, (8.35) yields

$$f(\omega)=\frac{\lambda C}{\gamma}+\frac{\lambda[((\gamma-\sigma\mu_1-(1-\sigma)\mu_2)C+(\sigma\mu_1+(1-\sigma)\mu_2)\gamma R)\omega-C(\lambda+\gamma)(\sigma\mu_1+(1-\sigma)\mu_2)]}{\gamma\omega(\omega+\lambda)} \tag{8.36}$$

where $\omega\in[\sigma\mu_1+(1-\sigma)\mu_2-\lambda,\sigma\mu_1+(1-\sigma)\mu_2]$.

Profit function $f(q)$ given in (8.30) for $0\le q\le 1$ is equivalent to profit function $f(\omega)$ given in (8.36) for $\sigma\mu_1+(1-\sigma)\mu_2-\lambda\le\omega\le\sigma\mu_1+(1-\sigma)\mu_2$. Maximizing profit function given in (8.30) is now equivalent to maximizing profit function given in (8.36). Differentiating (8.36) with respect to ω, we have

$$\frac{df(\omega)}{d\omega}=\frac{\lambda[-A\omega^2+B\omega+D]}{\gamma\omega^2(\omega+\lambda)^2} \tag{8.37}$$

where $A=((\gamma-\sigma\mu_1-(1-\sigma)\mu_2)C+(\sigma\mu_1+(1-\sigma)\mu_2)\gamma R)$, $B=2C(\lambda+\gamma)(\sigma\mu_1+(1-\sigma)\mu_2)$ and $D=\lambda C(\lambda+\gamma)(\sigma\mu_1+(1-\sigma)\mu_2)$. We notice that $B>0$ and $D>0$ whereas either $A\le 0$ or $A>0$.

Case 1: If $A\le 0\Leftrightarrow$

$$((\gamma-\sigma\mu_1-(1-\sigma)\mu_2)C+(\sigma\mu_1+(1-\sigma)\mu_2)\gamma R)\le 0\Leftrightarrow\frac{R}{C}\le\frac{1}{\gamma}-\frac{1}{\sigma\mu_1+(1-\sigma)\mu_2}.$$

We observe $\frac{df(\omega)}{d\omega}>0$. Therefore, $f(\omega)$ is strictly increasing function in ω and hence its maximum value is attained at $\omega=\sigma\mu_1+(1-\sigma)\mu_2$ or equivalently, $q^*=0$. When $\frac{R}{C}\le\frac{1}{\gamma}-\frac{1}{\sigma\mu_1+(1-\sigma)\mu_2}$, then customers would not prefer to join the system and balk, i.e. $q^*=0$.

Case 2: If $A>0\Leftrightarrow$

$$((\gamma-\sigma\mu_1-(1-\sigma)\mu_2)C+(\sigma\mu_1+(1-\sigma)\mu_2)\gamma R)>0\Leftrightarrow\frac{R}{C}>\frac{1}{\gamma}-\frac{1}{\sigma\mu_1+(1-\sigma)\mu_2}.$$

In this case, we cannot predict whether $f(\omega)$ is increasing or decreasing in ω for $\sigma\mu_1+(1-\sigma)\mu_2-\lambda\le\omega\le\sigma\mu_1+(1-\sigma)\mu_2$. Solving $\frac{df(\omega)}{d\omega}=0$, we have

$$\omega_1 = \frac{C(\sigma\mu_1 + (1-\sigma)\mu_2)(\lambda+\gamma) + \sqrt{X}}{C(\gamma - \sigma\mu_1 - (1-\sigma)\mu_2)) + \gamma RC(\sigma\mu_1 + (1-\sigma)\mu_2)} > 0 \tag{8.38}$$

$$\omega_2 = \frac{C(\sigma\mu_1 + (1-\sigma)\mu_2)(\lambda+\gamma) - \sqrt{X}}{C(\gamma - \sigma\mu_1 - (1-\sigma)\mu_2)) + \gamma RC(\sigma\mu_1 + (1-\sigma)\mu_2)} < 0 \tag{8.39}$$

Thus, we notice that $f(\omega)$ is decreasing in $(-\infty, \omega_2) \cup (\omega_1, \infty)$ and increasing in $[\omega_2, \omega_1]$. Now there are three subcases as follows:

***Case 2.1*:** When $\sigma\mu_1 + (1-\sigma)\mu_2 < \omega_1 \Leftrightarrow \frac{R}{C} < \frac{(\sigma\mu_1 + (1-\sigma)\mu_2 + \lambda)(\lambda+\gamma)}{\gamma(\sigma\mu_1 + (1-\sigma)\mu_2)^2}$. This implies that $f(\omega)$ is an increasing function in $[\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda, \sigma\mu_1 + (1-\sigma)\mu_2]$. Thus $f(\omega)$ has its maximum value at its upper bound i.e., at $\omega = \sigma\mu_1 + (1-\sigma)\mu_2$ or equivalently, when $q^* = 0$. Therefore, the best response of the customers is balking.

***Case 2.2*:** When $\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda \le \omega_1 \le \sigma\mu_1 + (1-\sigma)\mu_2 \Leftrightarrow u < \frac{R}{C} < v$, $f(\omega)$ increases in $[\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda, \omega_1]$ and decreases in $[\omega_1, \infty]$. Thus, $f(\omega)$ attains its maximum value at $\omega = \omega_1$ or equivalently, when $q^* = \frac{\sigma\mu_1 + (1-\sigma)\mu_2 - \omega_1}{\lambda}$. To get the maximum profit after getting served when $u < \frac{R}{C} < v$, customers should join the system with probability $q^* = \frac{\sigma\mu_1 + (1-\sigma)\mu_2 - \omega_1}{\lambda}$.

***Case 2.3*:** When $\omega_1 < \sigma\mu_1 + (1-\sigma)\mu_2 - \lambda \Leftrightarrow \frac{R}{C} > v$. This implies that $f(\omega)$ is a decreasing function in $[\sigma\mu_1 + (1-\sigma)\mu_2 - \lambda, \sigma\mu_1 + (1-\sigma)\mu_2]$. Therefore, $f(\omega)$ attains its maximum value at $\omega = \sigma\mu_1 + (1-\sigma)\mu_2 - \lambda$ or equivalently, $q^* = 1$. This shows that arriving customers necessarily join the system to get maximum profit.

8.3.2. Cost Function

For the single server double orbit retrial queue with balking, the system organizers always prefer to serve the customers as per their demand and earn as much profit as possible. To reduce the system cost and provide better service to the customers, we construct a cost function that can be used to determine the minimum system cost per unit of time corresponding to the optimal service rate. The cost function is framed by considering various cost factors per unit time including $C_O\,(C_P)$ as cost per unit of time incurred on each customer while residing in ordinary orbit (premium orbit); C_H as holding cost of each customer in the system; $C_I\,(C_B)$ as cost spent when the server is idle (busy); $C_1\,(C_2)$ as cost per unit of time incurred on each ordinary customer (premium customer) when he is being served by the server. Thus, total cost function is given by

$$TC(\mu_1,\mu_2)=C_O E[N_O]+C_P E[N_P]+C_H E[N_q]+C_I P_I+C_B P_{SB}+C_1\sigma\mu_1+C_2(1-\sigma)\mu_2 \quad (8.40)$$

The cost function given in (8.40) looks highly non-linear in μ_1 and μ_2. Thus, in such a case we may not be able to minimize it by an analytical method. Thus, we use the genetic algorithm (GA) based on soft computing technique to deal with the cost function.

8.3.3. Numerical Results

(i) Joining Strategy of Customers and Profit Function

In this section, we analyze the impact of system parameters on profit function given in (8.31) and on optimal joining probability. For illustration purpose, a numerical example is considered which will be helpful to validate the tractability of the profit function along with optimal joining probability. For computation, we take default parameters as $\mu_1=6$, $\mu_2=8$, $\sigma=0.5$, $\gamma=1$, $R=10$ and $C=1$. When the system organizer serves the customers with faster rates or customers from retrial orbit re-request for the service with faster rates, the customers are encouraged to join the system with higher probability and this fact can be noticed in Figures 8.6(a-b) and Figure 8.6(d) for $R=8,10,12$. Figure 8.6(c) reveals that when the system seems congested then the probability of joining the system is monotonically decreasing.

Figures 8.7(a-b) and Figure 8.7(d) show that maximum profit enhances rapidly as the service rates and retrials of customers increase. This is due to the fact that the customers spend less waiting time in the system so that their waiting cost reduces and profit increases. On the other hand, we see in Figure 8.7(c) that when the arrival rate goes up then maximum profit increases gradually as the system becomes congested and the customers spend more time in the system. The surface graphs plotted in Figures 8.8(a-d) for the expected profit function versus joining probability by taking system parameters (a) μ_1 (b) μ_2 (c) γ and (d) R are concave and the maximum profit corresponding to optimal joining probability occurs.

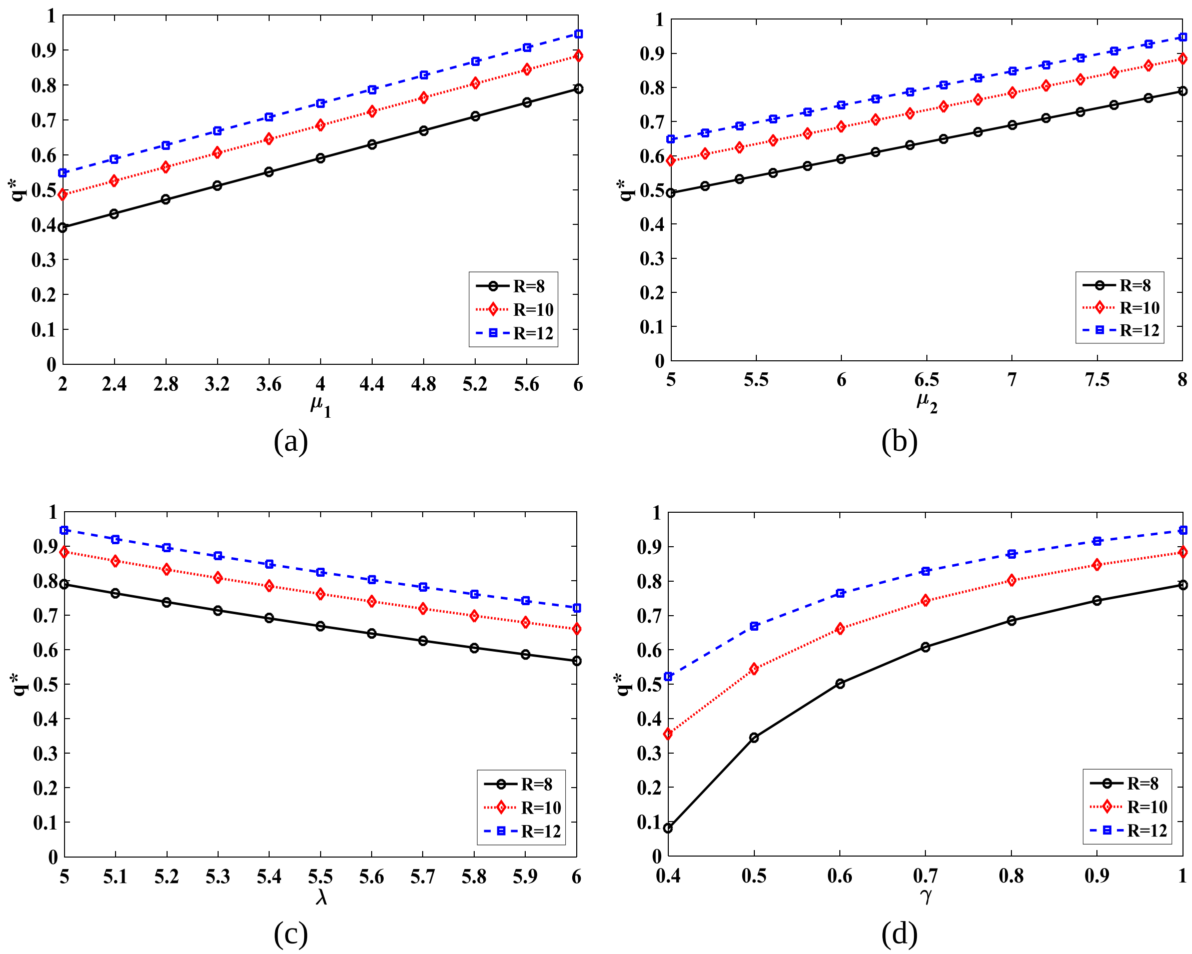
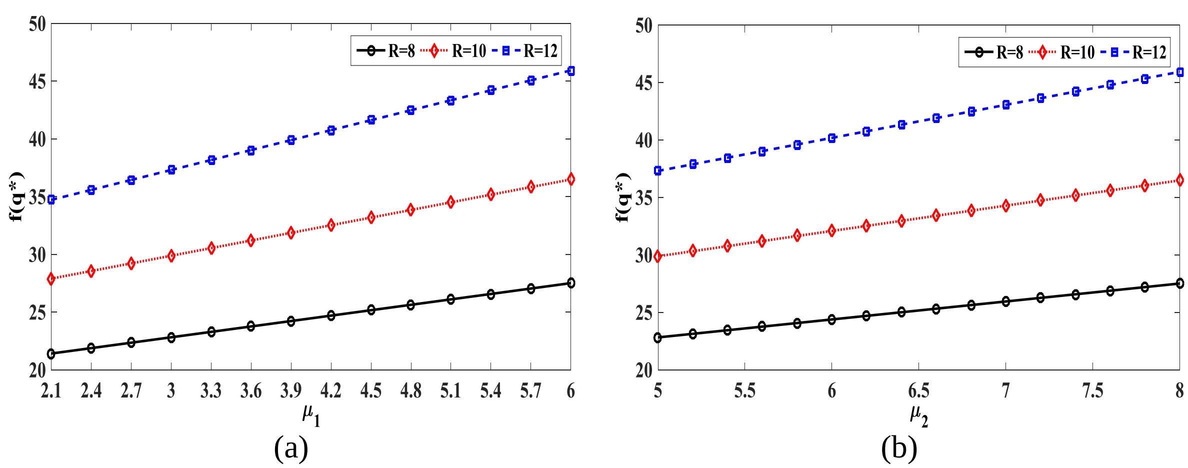


Figure 8.6: q^* vs (a) μ_1 (b) μ_2 (c) λ (d) γ for different values of R



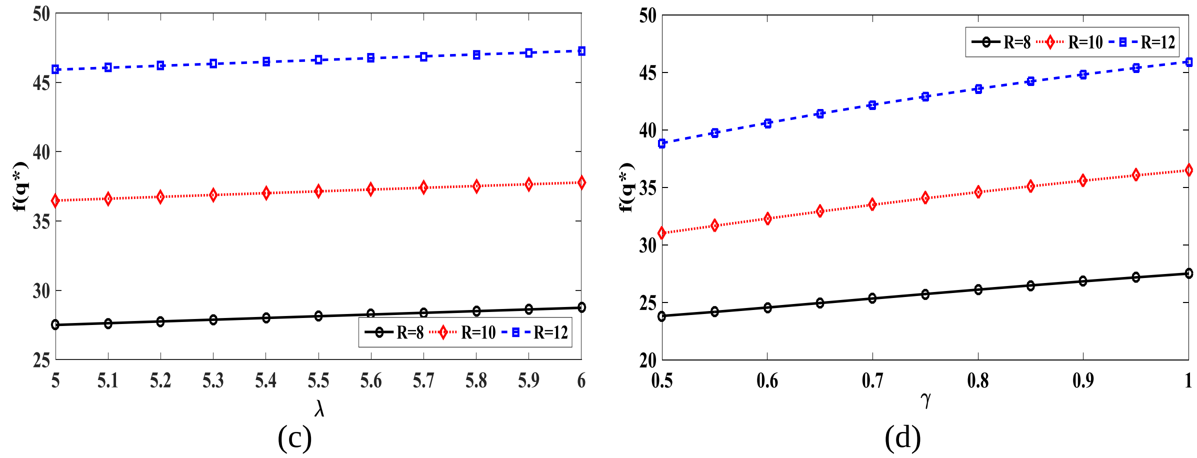


Figure 8.7: $f(q^*)$ vs (a) μ_1 (b) μ_2 (c) λ (d) γ for different values of R

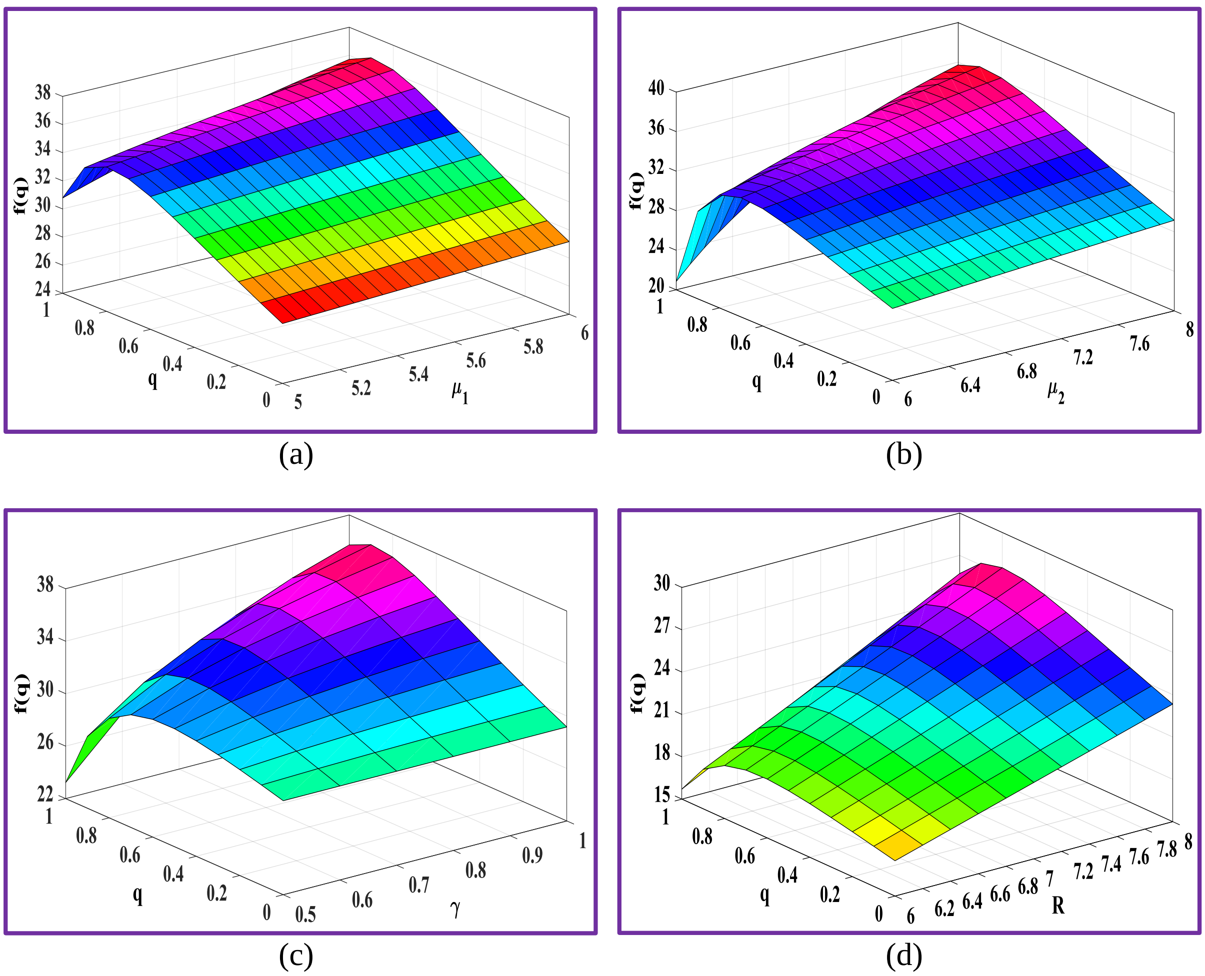


Figure 8.8: $f(q)$ for varying values of q and (a) μ_1 (b) μ_2 (c) γ (d) R

(ii) Cost Optimization using Genetic Algorithm (GA)

To minimize the system cost alongwith the best service to the customers is the main concern of the system organizer. In this section, our main goal is to determine optimal service rates (μ_1^*, μ_2^*) for both ordinary and premium customers and the corresponding minimum costs $(TC(\mu_1^*), TC(\mu_2^*))$ of the system. Since the cost function given in Equation (8.40) of the previous Section 8.3.2, is highly nonlinear and complex so it is not an easy task to analyze its nature analytically. A population search based genetic algorithm is implemented to achieve the goal of optimal service rates by following the algorithmic steps given in Section 1.4.8 of Chapter 1.

For the M/M/1 double orbit retrial queueing model with balking, the input parameters $\lambda, q, \gamma, C_O, C_P, C_H, C_I, C_B, C_1, C_2$ and population size (P_μ), genes (G_μ), crossover rate (C_R) and mutation rate $(\mathrm{M_R})$ are used. The output of the GA are (i) μ_1^* and $TC(\mu_1^*)$ and (ii) μ_2^* and $TC(\mu_2^*)$. To determine the output through GA, the numerical values for the input parameters are summarized in Tables 8.5-8.6.

After following the steps for GA (cf. Mitchell, 1996), Figures 8.9-8.12 indicate the best fitness values and best mean fitness in each generation versus iteration number. Also, it is noticed that the weighted average change in the fitness function (cost function) value over stall generations is less than function tolerance. Therefore, GA is executed successfully after 51 iterations and the output results (i) μ_1^* and $TC(\mu_1^*)$ and (ii) μ_2^* and $TC(\mu_2^*)$ for varying values of q and γ are recorded in Table 8.7. Minimum values of the system cost summarized in Table 8.7 reveal that when the customers join the system with higher probability, then system becomes congested; in such scenario, the system organizer has to render the service at faster rate. On the other hand, when the customers retry from their respective orbits for the service with faster rate, the system gradually becomes less congested; also the system cost decreases. Cost function given in (8.40) is convex as can be seen in the surface graphs shown in Figures 8.13-8.16. The surfaces graphs are made for total cost versus service rates, (μ_1, μ_2) by considering different values of q and γ.

Table 8.5: GA parameters for computation of μ_1^* and $TC(\mu_1^*)$

Parameters	Value assigned	Method
$(\lambda, \mu_2, \sigma, q, \gamma)$ $(C_O, C_P, C_H, C_I, C_B, C_1, C_2)$	$(22, 30, 0.5, 0.75, 10)$ $(60, 120, 30, 30, 110, 50, 50)$	---
Population size (P_{μ_1})	20	Binary encoding
Selection	$P_{\mu_1}/5$	Tournament selection
Crossover rate (C_R)	1	2-Point crossover
Mutation rate (M_R)	0.08	Bit inversion mutation
Stopping criteria: (i) Generations (ii) Stall generation (iii) Function tolerance	 100 50 1×10^{-6}	---

Table 8.6: GA parameters for computation of μ_2^* and $TC(\mu_1^*)$

Parameters	Value assigned	Method
$(\lambda, \mu_1, \sigma, q, \gamma)$ $(C_O, C_P, C_H, C_I, C_B, C_1, C_2)$	$(22, 20, 0.5, 0.75, 10)$ $(60, 120, 30, 30, 110, 50, 50)$	---
Population size (P_{μ_1})	20	Binary encoding
Selection	$P_{\mu_1}/5$	Tournament selection
Crossover rate (C_R)	1	2-Point crossover
Mutation rate (M_R)	0.08	Bit inversion mutation
Stopping criteria: (i) Generations (ii) Stall generation (iii) Function tolerance	 100 50 1×10^{-6}	---

Table 8.7: $(\mu_1^*, TC(\mu_1^*))$ and $(\mu_2^*, TC(\mu_2^*))$ for different q and γ

(q, γ)	$(\mu_1^*, TC(\mu_1^*))$	$(\mu_2^*, TC(\mu_2^*))$
(0.70, 10)	(12.56, $1479.84)	(21.99, $1466.43)
(0.75, 10)	(15.15, $1554.88)	(24.61, $1545.99)
(0.80, 10)	(17.73, $1629.28)	(27.23, $1624.90)
(0.75, 6)	(18.07, $1719.31)	(27.15, $1713.08)
(0.75, 9)	(15.67, $1584.34)	(25.07, $1575.57)
(0.75, 12)	(14.35, $1508.75)	(23.90, $1499.98)

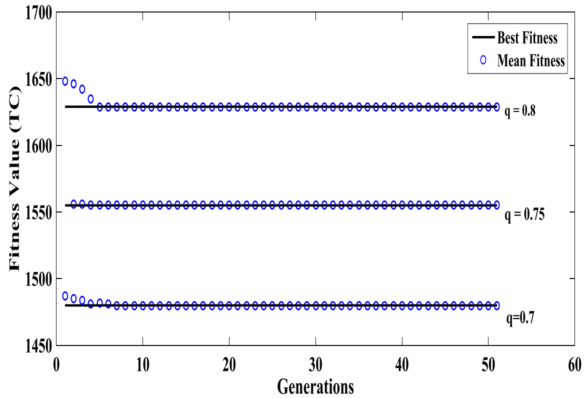


Figure 8.9: TC vs Generation for evaluation of μ_1 by varying value of q

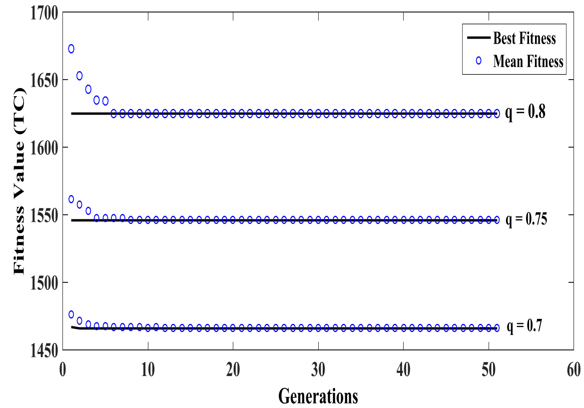


Figure 8.10: TC vs Generation for evaluation of μ_2 by varying value of q

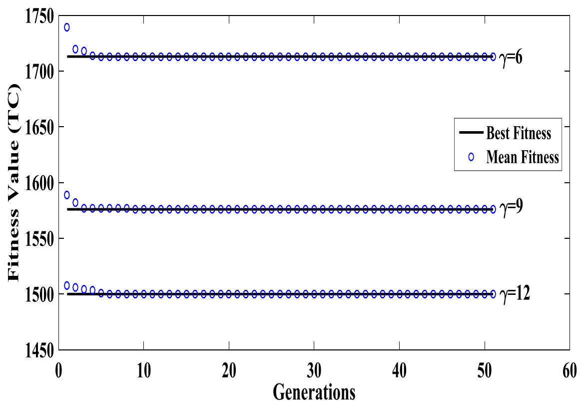


Figure 8.11: TC vs Generation for evaluation of μ_1 by varying value of γ

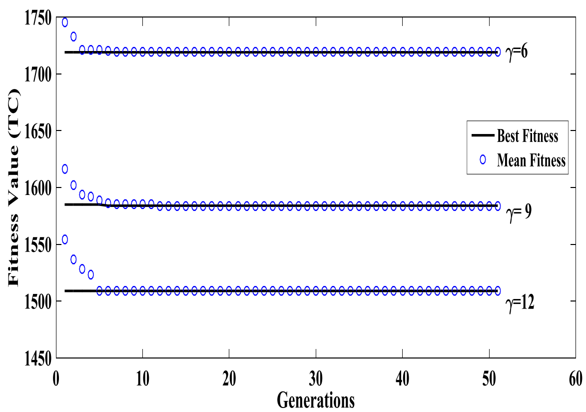


Figure 8.12: TC vs Generation for evaluation of μ_2 by varying value of γ

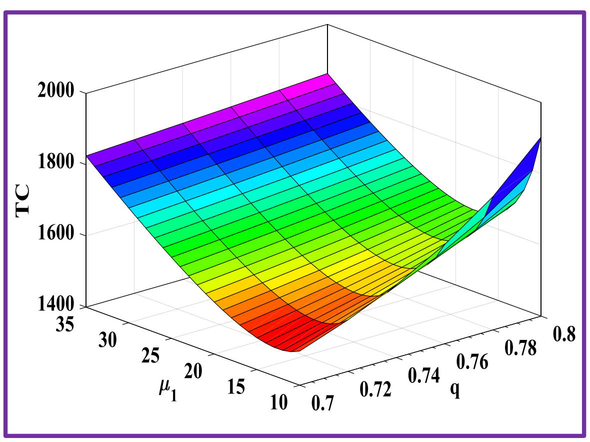


Figure 8.13: TC for varying of μ_1 and q

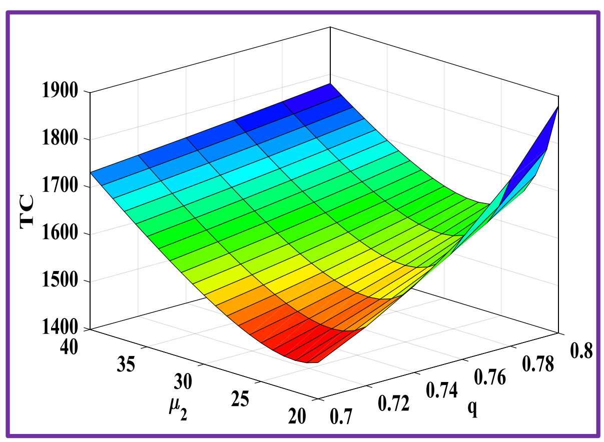


Figure 8.14: TC for varying of μ_2 and q

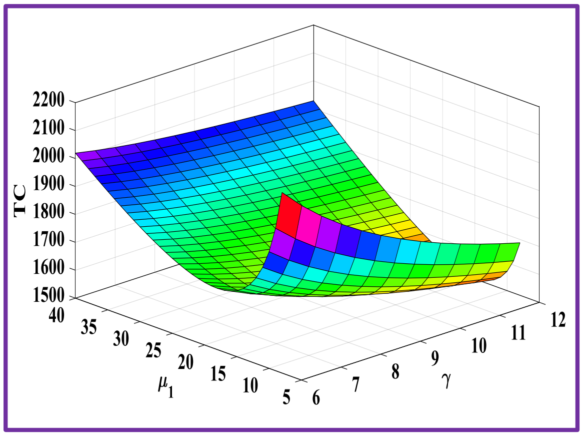


Figure 8.15: TC for varying of μ_1 and γ

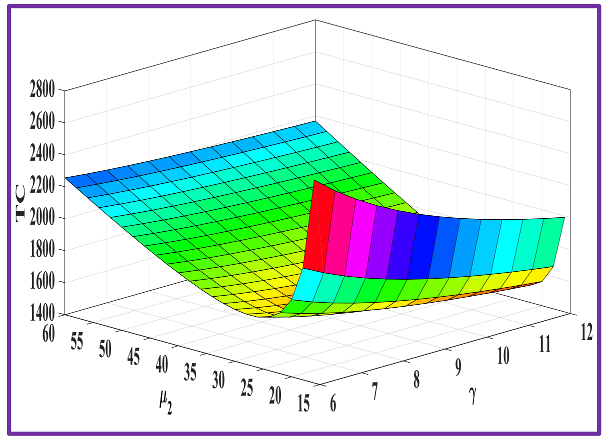


Figure 8.16: TC for varying of μ_2 and γ

8.4. FM/FM/1 DOUBLE ORBIT RETRIAL FUZZY QUEUE

In this section, double orbit retrial queueing model with infinite capacity is developed in fuzzy environment. The input parameters namely arrival rate (λ), the effective service rate for ordinary customers (μ_O), the effective service rate for premium customers (μ_P) and the retrial rate (γ) used in M/M/1 model developed in previous section are to be considered as fuzzified arrival rate $(\tilde{\lambda})$, the effective service rate for the ordinary customers $(\tilde{\mu}_O)$, the effective service rate for the premium customers $(\tilde{\mu}_P)$ and the retrial rate $(\tilde{\gamma})$, respectively. It is noted that $\tilde{\lambda}, \tilde{\mu}_O, \tilde{\mu}_P$ *and* $\tilde{\gamma}$ are the convex fuzzy numbers and their membership functions are $\xi_{\tilde{\lambda}}(l)$, $\xi_{\tilde{\mu}_O}(m_O), \xi_{\tilde{\mu}_P}(m_P)$ and $\xi_{\tilde{\gamma}}(g)$, respectively. Assume $\tilde{\theta}_j$ denotes a convex fuzzy set and its membership function is denoted by $\xi_{\tilde{\theta}_j}(x_j)$. Thus,

$$\tilde{\theta}_j = \{(x_j, \xi_{\tilde{\theta}_j}(x_j)) : x_j \in X_j\}, \quad j = 1,2,3,4 \tag{8.41}$$

where X_j denotes the crisp universal set corresponding to $\tilde{\theta}_j$. Also for FM/FM/1 model, we get $\tilde{\theta}_1 = \tilde{\lambda},\ \tilde{\theta}_2 = \tilde{\mu}_O,\ \tilde{\theta}_3 = \tilde{\mu}_P,\ \tilde{\theta}_4 = \tilde{\gamma}$ and $x_1 = l,\ x_2 = m_O,\ x_3 = m_P,\ x_4 = g$.

Let $f(x_1, x_2, x_3, x_4)$ denote the system performance index. Here $\tilde{\theta}_j$ are fuzzy numbers, so that $f(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4)$ is also a fuzzy number. Now, we define $f(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4)$ by

$$f(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4) = \{(y, \xi_{f(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4)}(y)) : y \in Y\}\ . \tag{8.42}$$

where Y is the universal crisp set of $f(\tilde{\theta}_1,\tilde{\theta}_2,\tilde{\theta}_3,\tilde{\theta}_4)$. Based on Zadeh's extension principle (Zadeh, 1978), the membership function of $f(\tilde{\theta}_1,\tilde{\theta}_2,\tilde{\theta}_3,\tilde{\theta}_4)$ is determined by

$$\xi_{f(\tilde{\theta}_1,\tilde{\theta}_2,\tilde{\theta}_3,\tilde{\theta}_4)}(y)=\begin{cases}\sup\limits_{(x_1,x_2,x_3,x_4)=f^{-1}(y)} \min\{\xi_{\tilde{\theta}_j}(x_j): j=1,2,3,4\}; & if\ f^{-1}(y)\, exists\\ 0 & otherwise\end{cases} \tag{8.43}$$

Thus, $$\xi_{f(\tilde{\theta}_1,\tilde{\theta}_2,\tilde{\theta}_3,\tilde{\theta}_4)}(y)=\sup_{x_j\in X_j} \min\{\xi_{\tilde{\theta}_j}(x_j): j=1,2,3,4 \mid y=f(x_1,x_2,x_3,x_4)\} \tag{8.44}$$

Referring to Theorem 8.2, the membership functions for the expected number of customers $(E[\tilde{N}_S])$ and the expected waiting time in the system $(E[\tilde{W}_S])$ respectively, are given by

$$\xi_{E[\tilde{N}_S]}(y)=\sup_{r<1} \min\left\{\xi_{\tilde{\theta}_j}(x_j),\ j=1,2,3,4 \mid y=\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right]\right\} \tag{8.45}$$

$$\xi_{E[\tilde{W}_S]}(y)=\sup_{r<1} \min\left\{\xi_{\tilde{\theta}_j}(x_j),\ j=1,2,3,4 \mid y=\frac{r^2}{lq}\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right]\right\} \tag{8.46}$$

where $r=\dfrac{lq}{m_O+m_P}$ and $\delta=\dfrac{r^2}{q(1-r)+r}$.

From (8.45) and (8.46), it is not an easy task to plot the membership function of $E[\tilde{N}_S]$ and $E[\tilde{W}_S]$. The parametric nonlinear programing (P-NLP) is used to find $\alpha-$cut of $E[\tilde{N}_S]$ and $E[\tilde{W}_S]$ by using Zadeh's extension principle.

8.4.1. Parametric Nonlinear Programing (P-NLP)

The membership functions are framed by using a parametric non-linear programing approach. Based on Zadeh's extension principle, $\alpha-$cuts of $E[\tilde{N}_S]$ are constructed as

$$E[N_S](\alpha)=\{y:\xi_{E[\tilde{N}_S]}(y)\ge\alpha\},\ \alpha\in[0,1] \tag{8.47}$$

The $\alpha-$cuts of $\tilde{\theta}_j$ are defined by

$$\theta_j(\alpha)=\{x_j:\xi_{\tilde{\theta}_j}\ge\alpha\},\ \ j=1,2,3,4 \tag{8.48}$$

The $\alpha-$cuts of $\tilde{\theta}_j$ in terms of the crisp interval are obtained as

$$\theta_j(\alpha)=\left[(x_j)_\alpha^L,\ (x_j)_\alpha^U\right]=\left[\min_{x_j\in X_j}\{x_j:\xi_{\tilde{\theta}_j}(x_j)\ge\alpha\},\ \max_{x_j\in X_j}\{x_j:\xi_{\tilde{\theta}_j}(x_j)\ge\alpha\}\right],\quad j=1,2,3,4. \tag{8.49}$$

Since $\tilde{\theta}_j$ is a convex fuzzy number, therefore the upper and lower bounds of $\tilde{\theta}_j$ can be represented in terms of α such that

$$\left[(x_j)_\alpha^L,\ (x_j)_\alpha^U\right]=\left[\min \xi_{\tilde{\theta}_j}^{-1}(\alpha),\ \max \xi_{\tilde{\theta}_j}^{-1}(\alpha)\right],\ j=1,2,3,4. \tag{8.50}$$

Based on Zadeh's extension principle, (8.45) holds if at least one of the four cases holds:

Case 1: $\xi_{\tilde{\theta}_1}(x_1)=\alpha,\ \xi_{\tilde{\theta}_j}(x_j)\ge\alpha,\ j=2,3,4,$

Case 2: $\xi_{\tilde{\theta}_2}(x_2)=\alpha,\ \xi_{\tilde{\theta}_j}(x_j)\ge\alpha,\ j=1,3,4,$

Case 3: $\xi_{\tilde{\theta}_3}(x_3)=\alpha,\ \xi_{\tilde{\theta}_j}(x_j)\ge\alpha,\ j=1,2,4,$

Case 4: $\xi_{\tilde{\theta}_4}(x_4)=\alpha,\ \xi_{\tilde{\theta}_j}(x_j)\ge\alpha,\ j=1,2,3.$

Thus

$$y=\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.51}$$

satisfies $\xi_{E[\tilde{N}_S]}(y)=\alpha$.

The parametric nonlinear programing (P-NLP) is used to construct the lower bound (y_α^L) and upper bound (y_α^U) of $\alpha-$cut of $E[\tilde{N}_S]$ through the Cases 1-4. Thus, we obtain

$$y_{1,\alpha}^L=\min_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.52a}$$

$$y_{1,\alpha}^U=\max_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.52b}$$

$$y_{2,\alpha}^L=\min_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.53a}$$

$$y_{2,\alpha}^U=\max_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.53b}$$

$$y_{3,\alpha}^{L}=\min_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.54a}$$

$$y_{3,\alpha}^{U}=\max_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.54b}$$

$$y_{4,\alpha}^{L}=\min_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.55a}$$

$$y_{4,\alpha}^{U}=\max_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.55b}$$

Since $x_j\in\theta_j(\alpha)$, $x_j\in\left[(x_j)_\alpha^L,(x_j)_\alpha^U\right]$. For given $0<\alpha_2<\alpha_1\le 1$, we conclude $\left[(x_j)_{\alpha_1}^L,(x_j)_{\alpha_1}^U\right]\subseteq\left[(x_j)_{\alpha_2}^L,(x_j)_{\alpha_2}^U\right]$ for $j=1,2,3,4$; this is because $\alpha-$cuts form a nested structure with respect to α. Therefore, it is clear that (8.52a), (8.53a), (8.54a) and (8.55a) have the same smallest values. On the other hand (8.52b), (8.53b), (8.54b) and (8.55b) have the same largest value. To plot the membership function $\left(\xi_{E[\tilde{N}_S]}\right)$ of $E[\tilde{N}_S]$, it is sufficient to determine the lower and upper bound of the membership function $\xi_{E[\tilde{N}_S]}$, which is equivalent to determine the lower bound (y_α^L) and upper bound (y_α^U) of $\alpha-$cut of $E[\tilde{N}_S]$ given as follows:

$$E[N_S]_\alpha^L=y_\alpha^L=\min_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.56}$$

where $l_\alpha^L\le l\le l_\alpha^U$, $(m_O)_\alpha^L\le m_O\le(m_O)_\alpha^U$, $(m_P)_\alpha^L\le m_P\le(m_P)_\alpha^U$, $g_\alpha^L\le g\le g_\alpha^U$

and
$$E[N_S]_\alpha^U=y_\alpha^U=\max_{r<1}\delta\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.57}$$

where $l_\alpha^L\le l\le l_\alpha^U$, $(m_O)_\alpha^L\le m_O\le(m_O)_\alpha^U$, $(m_P)_\alpha^L\le m_P\le(m_P)_\alpha^U$, $g_\alpha^L\le g\le g_\alpha^U$.

At least one of $x_j;\, j=1,2,3,4$ must be on the boundary of their $\alpha-$cut that satisfies $\xi_{E[\tilde{N}_S]}=\alpha$. Here, (8.56) and (8.57) are the special cases of P-NLPs (Gal, 1979). y_α^L and y_α^U are respectively increasing and decreasing functions as α goes up because $y_{\alpha_2}^L\le y_{\alpha_1}^L$ and $y_{\alpha_2}^U\ge y_{\alpha_1}^U$ through $0<\alpha_2<\alpha_1\le 1$. Therefore, the membership function $\xi_{E[\tilde{N}_S]}$ of $E[\tilde{N}_S]$ is constructed as

$$\xi_{E[\tilde{N}_S]}(y)=\begin{cases} L_N(y), & E[N_S]_{\alpha=0}^L \le y \le E[N_S]_{\alpha=1}^L \\ 1, & E[N_S]_{\alpha=1}^L \le y \le E[N_S]_{\alpha=1}^U \\ R_N(y), & E[N_S]_{\alpha=1}^U \le y \le E[N_S]_{\alpha=0}^U \\ 0, & otherwise \end{cases} \tag{8.58}$$

where $(E[N_S]_\alpha^L)^{-1}=L_N(y)$ and $(E[N_S]_\alpha^U)^{-1}=R_N(y)$ are the left and the right portion of the portrait of the membership function $\xi_{E[\tilde{N}_S]}$ of $E[\tilde{N}_S]$.

By following the same procedure from (8.47) to (8.58), the lower bound $(E[W_S]_\alpha^L)$ and upper bound $(E[W_S]_\alpha^U)$ of $\alpha-$cut of $E[\tilde{W}_S]$ are obtained as

$$E[W_S]_\alpha^L = \min_{r<1} \frac{r^2}{lq}\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.59}$$

where $l_\alpha^L \le l \le l_\alpha^U$, $(m_O)_\alpha^L \le m_O \le (m_O)_\alpha^U$, $(m_P)_\alpha^L \le m_P \le (m_P)_\alpha^U$, $g_\alpha^L \le g \le g_\alpha^U$

and $$E[W_S]_\alpha^U = y_\alpha^U = \max_{r<1} \frac{r^2}{lq}\left[\frac{m_O+m_P}{g}+\frac{1}{1-r}\left(\frac{l+g}{g}\right)+\frac{1}{r}\right] \tag{8.60}$$

where $l_\alpha^L \le l \le l_\alpha^U$, $(m_O)_\alpha^L \le m_O \le (m_O)_\alpha^U$, $(m_P)_\alpha^L \le m_P \le (m_P)_\alpha^U$, $g_\alpha^L \le g \le g_\alpha^U$.

The membership function $\xi_{E[\tilde{W}_S]}$ of $E[\tilde{W}_S]$ is constructed as

$$\xi_{E[\tilde{W}_S]}(y)=\begin{cases} L_W(y), & E[W_S]_{\alpha=0}^L \le y \le E[W_S]_{\alpha=1}^L \\ 1, & E[W_S]_{\alpha=1}^L \le y \le E[W_S]_{\alpha=1}^U \\ R_W(y), & E[W_S]_{\alpha=1}^U \le y \le E[W_S]_{\alpha=0}^U \\ 0, & otherwise \end{cases} \tag{8.61}$$

where $(E[W_S]_\alpha^L)^{-1}=L_W(y)$ and $(E[W_S]_\alpha^U)^{-1}=R_W(y)$ are the left and the right portion of the portrait of the membership function $\xi_{E[\tilde{W}_S]}$ of $E[\tilde{W}_S]$.

8.4.2. Defuzzification Approach: Yager Ranking Index

In many real time applications, the system organizers/managers prefer to have real (crisp) value instead of fuzzy value. In order to find the crisp values corresponding to fuzzy output values, Yager's ranking index approach is applied. Yager's ranking index is given as (cf. Yager, 1981)

$$R(\tilde{\Theta}) = \frac{1}{2}\int_0^1 (\Theta_\alpha^L + \Theta_\alpha^U)\, d\alpha \tag{8.62}$$

where $\left[\Theta_\alpha^L, \Theta_\alpha^U\right]$ is an $\alpha-$cut of a convex fuzzy number $\tilde{\Theta}$.

8.4.3. Evaluation of Extrema of $E[N_S]_\alpha$ and $E[W_S]_\alpha$

Let $\tilde{\lambda}, \tilde{\mu}_O, \tilde{\mu}_P, \tilde{\gamma}$ be the trapezoidal fuzzy numbers and they are represented by $\tilde{\lambda} = [a_1, a_2\, a_3, a_4]$, $\tilde{\mu}_O = [b_1, b_2\, b_3, b_4]$, $\tilde{\mu}_P = [c_1, c_2\, c_3, c_4]$, $\tilde{\gamma} = [d_1, d_2\, d_3, d_4]$, respectively; and $a_i < a_{i+1}$, $b_i < b_{i+1}, c_i < c_{i+1}, d_i < d_{i+1}$, $i = 1, 2, 3.$ The membership functions for $\tilde{\lambda}, \tilde{\mu}_O, \tilde{\mu}_P$ and $\tilde{\gamma}$ are respectively, given as

$$\xi_{\tilde{\lambda}}(l) = \begin{cases} \dfrac{l - a_1}{a_2 - a_1}, & a_1 \le l \le a_2, \\ 1, & a_2 \le l \le a_3, \\ \dfrac{a_4 - l}{a_4 - a_3}, & a_3 \le l \le a_4. \end{cases} \tag{8.63}$$

$$\xi_{\tilde{\mu}_O}(m_O) = \begin{cases} \dfrac{m_O - b_1}{b_2 - b_1}, & b_1 \le m_O \le b_2, \\ 1, & b_2 \le m_O \le b_3, \\ \dfrac{b_4 - m_O}{b_4 - b_3}, & b_3 \le m_O \le b_4. \end{cases} \tag{8.64}$$

$$\xi_{\tilde{\mu}_P}(m_P) = \begin{cases} \dfrac{m_P - c_1}{c_2 - c_1}, & c_1 \le m_P \le c_2, \\ 1, & c_2 \le m_P \le c_3, \\ \dfrac{c_4 - m_P}{c_4 - c_3}, & c_3 \le m_P \le c_4. \end{cases} \tag{8.65}$$

$$\xi_{\tilde{\gamma}}(g) = \begin{cases} \dfrac{g - d_1}{d_2 - d_1}, & d_1 \le g \le d_2, \\ 1, & d_2 \le g \le d_3, \\ \dfrac{d_4 - g}{d_4 - d_3}, & d_3 \le g \le d_4. \end{cases} \tag{8.66}$$

With the help of (8.49), $\alpha-$cuts of $\tilde{\lambda}, \tilde{\mu}_O, \tilde{\mu}_P$ and $\tilde{\gamma}$ are given by

$$\lambda(\alpha) = [l_\alpha^L, l_\alpha^U] = [a_1 + \alpha(a_2 - a_1),\ a_4 - \alpha(a_4 - a_3)] \tag{8.67}$$

$$\mu_O(\alpha) = [(m_O)_\alpha^L, (m_O)_\alpha^U] = \left[b_1 + \alpha(b_2 - b_1),\ b_4 - \alpha(b_4 - b_3)\right] \tag{8.68}$$

$$\mu_P(\alpha) = [(m_P)_\alpha^L, (m_P)_\alpha^U] = \left[c_1 + \alpha(c_2 - c_1),\ c_4 - \alpha(c_4 - c_3)\right] \tag{8.69}$$

$$\gamma(\alpha) = [g_\alpha^L, g_\alpha^U] = \left[d_1 + \alpha(d_2 - d_1),\ d_4 - \alpha(d_4 - d_3)\right] \tag{8.70}$$

Thus, (8.56) and (8.57) yield

$$E[N_S]_\alpha^L = \min_{r<1} \delta\left[\frac{m_O + m_P}{g} + \frac{1}{1-r}\left(\frac{l+g}{g}\right) + \frac{1}{r}\right] \tag{8.71}$$

$$E[N_S]_\alpha^U = \max_{r<1} \delta\left[\frac{m_O + m_P}{g} + \frac{1}{1-r}\left(\frac{l+g}{g}\right) + \frac{1}{r}\right] \tag{8.72}$$

and satisfy $a_1 + \alpha(a_2 - a_1) \le l \le a_4 - \alpha(a_4 - a_3)$, $b_1 + \alpha(b_2 - b_1) \le m_O \le b_4 - \alpha(b_4 - b_3)$, $c_1 + \alpha(c_2 - c_1) \le m_P \le c_4 - \alpha(c_4 - c_3)$, $d_1 + \alpha(d_2 - d_1) \le g \le d_4 - \alpha(d_4 - d_3)$.

To determine the extremum of the objective functions given in (8.71) and (8.72), we employ the concepts of differential calculus.

Differentiating (8.71) partially with respect to l, m_O, m_P and g respectively, we notice that $\frac{\partial}{\partial l} E[N_S]_\alpha^L > 0$, $\frac{\partial}{\partial m_O} E[N_S]_\alpha^L < 0$, $\frac{\partial}{\partial m_P} E[N_S]_\alpha^L < 0$ and $\frac{\partial}{\partial g} E[N_S]_\alpha^L < 0$ for $a_1 \le l \le a_4$, $b_1 \le m_O \le b_4, c_1 \le m_P \le c_4$, $d_1 \le g \le d_4$. Therefore, the minimum value of $E[N_S]_\alpha^L$ occurs at $l = a_1 + \alpha(a_2 - a_1),\ m_O = b_4 - \alpha(b_4 - b_3),\ m_P = c_4 - \alpha(c_4 - c_3)$ and $g = d_4 - \alpha(d_4 - d_3)$ and the maximum value of $E[N_S]_\alpha^U$ is found at $l = a_4 - \alpha(a_4 - a_3),\ m_O = b_1 + \alpha(b_2 - b_1),$ $m_P = c_1 + \alpha(c_2 - c_1)$ and $g = d_1 + \alpha(d_2 - d_1)$. Now, we have

$$E[N_S]_\alpha^L = \frac{q\left(a_1(\alpha-1) - \alpha a_2\right)^2 \left(\frac{1}{\alpha(d_3 - d_4) + d_4} + \frac{1}{q(a_1 - \alpha a_1 + \alpha a_2)} + \frac{a_1 + \alpha(a_2 - a_1 + d_3 - d_4) + d_4}{(\alpha(d_3 - d_4) + d_4)\left(u_1 + q((\alpha-1)a_1 - \alpha a_2)\right)}\right)}{b_4 + c_4 + a_1(\alpha-1)(q-1) + \alpha(a_2 + b_3 - b_4 + c_3 - c_4 - qa_2)} \tag{8.73}$$

$$E[N_S]_\alpha^U = \frac{q\left(a_4(1-\alpha)+\alpha a_3\right)^2 \left(\frac{1}{\alpha(d_2-d_1)+d_1} + \frac{1}{q(a_4-\alpha a_4+\alpha a_3)} + \frac{a_4+\alpha(a_3-a_4-d_1+d_2)+d_1}{(\alpha(d_2-d_1)+d_1)\left(v_1-q[(1-\alpha)a_4+\alpha a_3]\right)} \right)}{b_1+c_1-a_4(-1+q)-\alpha(a_4+b_1-b_2+c_1-c_2+a_3(-1+q)-qa_4)}$$

(8.74)

By following the same procedure for (8.59) and (8.60), $E[W_S]_\alpha^L$ attains its minimum value at $l=a_1+\alpha(a_2-a_1),\ m_O=b_4-\alpha(b_4-b_3),\ m_P=c_4-\alpha(c_4-c_3)$ and $g=d_4-\alpha(d_4-d_3)$.

Also attains its maximum at $l=a_4-\alpha(a_4-a_3),\ m_O=b_1+\alpha(b_2-b_1),\ m_P=c_1+\alpha(c_2-c_1)$ and $g=d_1+\alpha(d_2-d_1)$. Thus, we obtain

$$E[W_S]_\alpha^L = \frac{q\left(a_1(1-\alpha)+\alpha a_2\right)}{u_1}\left(\frac{1}{\alpha(d_3-d_4)+d_4} + \frac{1}{q(a_1-\alpha a_1+\alpha a_2)} + \frac{a_1+\alpha(a_2-a_1+d_3-d_4)+d_4}{(\alpha(d_3-d_4)+d_4)\left(u+q((\alpha-1)a_1-\alpha a_2)\right)} \right)$$

(8.75)

$$E[W_S]_\alpha^U = \frac{q\left(a_4(1-\alpha)+\alpha a_3\right)}{v_1}\left(\frac{1}{\alpha(d_2-d_1)+d_1} + \frac{1}{q(a_4-\alpha a_4+\alpha a_3)} + \frac{a_4+\alpha(a_3-a_4-d_1+d_2)+d_1}{(\alpha(d_2-d_1)+d_1)\left(v-q[(1-\alpha)a_4+\alpha a_3]\right)} \right)$$

(8.76)

where $u_1=b_4+\alpha(b_3-b_4+c_3-c_4)+c_4$ and $v_1=b_1+\alpha(b_2-b_1+c_2-c_1)+c_1$.

8.4.4. Example of FM/FM/1 Double Orbit Retrial Queue

To explore the proposed fuzzy retrial model in fuzzy environment, we present an application example of movie ticket booking counter which makes the concerned study more beneficial for the system organizers. Let us assume that the movie theatre has single booking counter which provides two types of tickets namely, ordinary ticket and premium ticket. Ordinary customers buy ordinary tickets and the premium customers buy a premium ticket from the booking counter. The arriving customers are assumed to follow Poisson process. The arriving customers may observe the queue length of customers before joining the queue and decide to join the queue with probability q and balk with probability $(1-q)$. The ordinary (premium) customers are served by the booking counter according to exponential distribution with service rates $\mu_1(\mu_2)$. If the arriving ordinary (premium) customer finds the booking counter busy then he is moved to the ordinary (premium) waiting hall and retries for the booking after a random period of time. The arrival rate and the effective service rate of ordinary customers, the effective service rate of premium customers and retrial rate are trapezoidal fuzzy numbers and are given by parameter values for per hour as $\tilde{\lambda} = [28, 30, 32, 34]$, $\tilde{\mu}_O = [18, 19, 20, 21]$, $\tilde{\mu}_P = [20, 21, 22, 23]$, $\tilde{\gamma} = [2, 3, 4, 5]$, respectively. The system organizers may find fuzzy expected number of customers in the system $(E[\tilde{N}_S])$ and fuzzy expected waiting time in the system $(E[\tilde{W}_S])$ for different values of $q = 0.5, 0.7, 0.9$. Using results given by (8.73)-(8.74) and (8.75)-(8.76), the numerical values are obtained and displayed in Tables 8.8-8.9 for $\alpha = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$.

From Table 8.8, it is noticed that at $\alpha = 0$, fuzzy value of $E[\tilde{N}_S]$ lies in the intervals [3.32, 14.88], [5.46, 30.38] and [8.91, 74.51] when $q = 0.5, 0.7, 0.9$, respectively. It means that the fuzzy value of the expected number of customers would not lie outside of the intervals for $\alpha = 0$. On the other hand, when $\alpha = 1$ and $q = 0.5, 0.7, 0.9$, the interval range for $E[\tilde{N}_S]$ are reported as [4.99, 80.6], [8.68, 15.04] and [15.37, 30.07], respectively. The same fact is shown in Figures 8.17(a-c).

Numerical results shown in Table 8.9 depict that fuzzy value of expected waiting time (in hour) for the customers in the system would not be decreased (increased) below (above) 0.16 (0.63) when $q = 0.5$ and $\alpha = 0$. At $\alpha = 1$ and $q = 0.5$, the fuzzy expected time in the system lies in the range [0.23 0.35]. Similarly, when joining the probability of the customers is $q = 0.5, 0.7$ and $\alpha = 1$, then the expected waiting time lies in [0.35 0.58], [0.55, 1.02], respectively. These facts are depicted in Figures 8.18(a-c).

The system organizers may be interested in the crisp values of the expected number of customers in the system and expected waiting time spent by the customers in the system instead of fuzzy values. To defuzzify the fuzzy value, we follow the defuzzification rule stated in Section 8.4.2. Using (8.73) and (8.74) in (8.62) for $q =$ 0.5, 0.7, and 0.9., the crisp values of the expected number of customers in the system is obtained as 7.52, 14.18 and 29.70, respectively. Similarly, crisp values of expected waiting time in the system are obtained as 0.33 hour, 0.55 hour, 1 hour for $q =$ 0.5, 0.7, and 0.9.

Table 8.8: $\alpha-$cuts of $\tilde{\lambda}, \tilde{\mu}_O, \tilde{\mu}_P, \tilde{\gamma}$ and $E[\tilde{N}_S]$

α		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
l_α^L		28	28.2	28.4	28.6	28.8	29	29.2	29.4	29.6	29.8	30
l_α^U		34	33.8	33.6	33.4	33.2	33	32.8	32.6	32.4	32.2	32
$(m_O)_\alpha^L$		18	18.1	18.2	18.3	18.4	18.5	18.6	18.7	18.8	18.9	19
$(m_O)_\alpha^U$		21	20.9	20.8	20.7	20.6	20.5	20.4	20.3	20.2	20.1	20
$(m_P)_\alpha^L$		20	20.1	20.2	20.3	20.4	20.5	20.6	20.7	20.8	20.9	21
$(m_P)_\alpha^U$		23	22.9	22.8	22.7	22.6	22.5	22.4	22.3	22.2	22.1	22
g_α^L		2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3
g_α^U		5	4.9	4.8	4.7	4.6	4.5	4.4	4.3	4.2	4.1	4
$q=0.5$	$E[N_S]_\alpha^L$	3.32	3.45	3.59	3.73	3.88	4.04	4.21	4.39	4.57	4.77	4.99
	$E[N_S]_\alpha^U$	14.88	13.87	12.96	12.14	11.39	10.71	10.09	9.52	8.99	8.51	8.06
$q=0.7$	$E[N_S]_\alpha^L$	5.46	5.7	5.96	6.23	6.52	6.83	7.15	7.5	7.86	8.26	8.68
	$E[N_S]_\alpha^U$	30.38	28.03	25.93	24.05	22.36	20.83	19.45	18.19	17.05	16	15.04
$q=0.9$	$E[N_S]_\alpha^L$	8.91	9.37	9.86	10.39	10.95	11.56	12.21	12.91	13.67	14.48	15.37
	$E[N_S]_\alpha^U$	74.51	66.92	60.39	54.71	49.77	45.43	41.6	38.21	35.19	32.49	30.07

Table 8.9: $\alpha-$cuts of $\tilde{\lambda}, \tilde{\mu}_O, \tilde{\mu}_P, \tilde{\gamma}$ and $E[\tilde{W}_S]$

α		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
l_α^L		28	28.2	28.4	28.6	28.8	29	29.2	29.4	29.6	29.8	30
l_α^U		34	33.8	33.6	33.4	33.2	33	32.8	32.6	32.4	32.2	32
$(m_O)_\alpha^L$		18	18.1	18.2	18.3	18.4	18.5	18.6	18.7	18.8	18.9	19
$(m_O)_\alpha^U$		21	20.9	20.8	20.7	20.6	20.5	20.4	20.3	20.2	20.1	20
$(m_P)_\alpha^L$		20	20.1	20.2	20.3	20.4	20.5	20.6	20.7	20.8	20.9	21
$(m_P)_\alpha^U$		23	22.9	22.8	22.7	22.6	22.5	22.4	22.3	22.2	22.1	22
g_α^L		2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3
g_α^U		5	4.9	4.8	4.7	4.6	4.5	4.4	4.3	4.2	4.1	4
$q=0.5$	$E[W_S]_\alpha^L$	0.16	0.16	0.17	0.17	0.18	0.19	0.19	0.2	0.21	0.22	0.23
	$E[W_S]_\alpha^U$	0.63	0.59	0.55	0.52	0.49	0.46	0.44	0.41	0.39	0.37	0.35
$q=0.7$	$E[W_S]_\alpha^L$	0.23	0.24	0.25	0.26	0.27	0.28	0.3	0.31	0.32	0.34	0.35
	$E[W_S]_\alpha^U$	1.13	1.05	0.97	0.91	0.85	0.79	0.74	0.7	0.66	0.62	0.58
$q=0.9$	$E[W_S]_\alpha^L$	0.34	0.35	0.37	0.39	0.41	0.43	0.45	0.47	0.49	0.52	0.55
	$E[W_S]_\alpha^U$	2.39	2.16	1.95	1.78	1.63	1.49	1.37	1.27	1.17	1.09	1.02

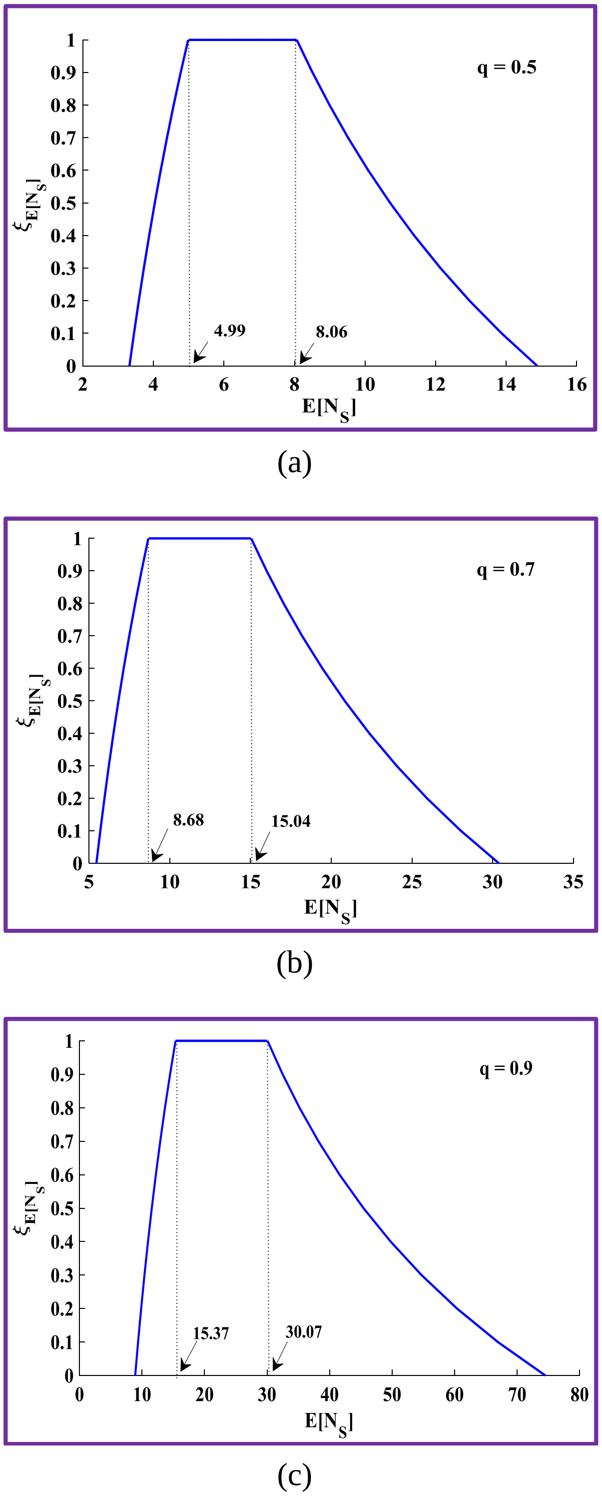


Figure 8.17: The membership function for $E[N_S]$ for (a) $q = 0.5$ (b) $q = 0.7$ (c) $q = 0.9$

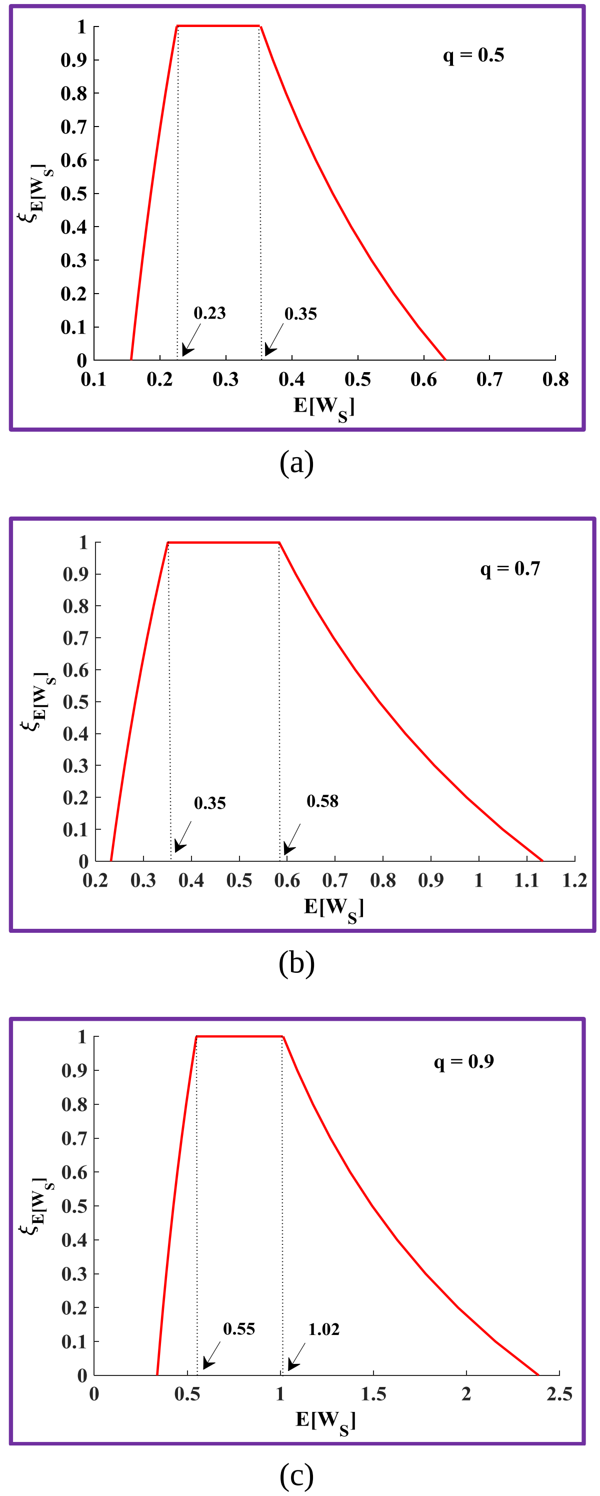


Figure 8.18: The membership function for $E[W_S]$ for (a) $q = 0.5$ (b) $q = 0.7$ (c) $q = 0.9$

8.5. DOUBLE ORBIT RETRIAL MODEL WITH FEEDBACK

If the served ordinary customer is not satisfied with service then he re-joins the system for the re-service. It is assumed that after getting served, β proportion of the ordinary customers leave the system whereas $(1-\beta)$ proportion of customers are feedback to the system again. The transition state diagram for ordinary customer's feedback is depicted in Figure 8.19.

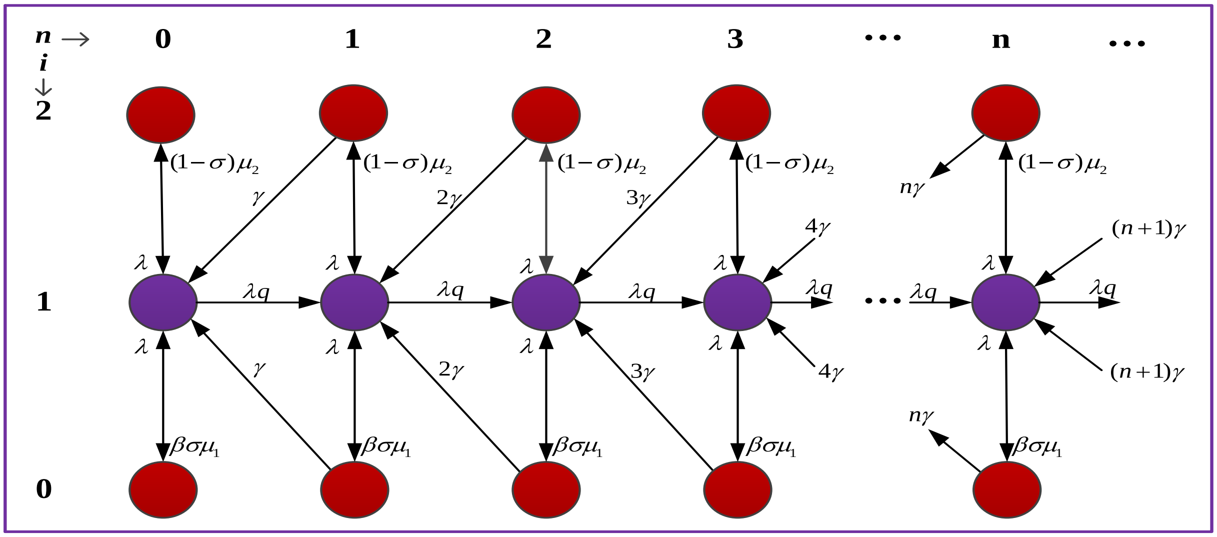


Figure 8.19: Transition state diagram for double orbit feedback model

The probability generating functions and steady state probabilities for the feedback model are given as:

$$\Pi_i(z)=\begin{cases}\dfrac{\Lambda'\beta\mu_O}{\beta\mu_O+\mu_P}(1-\rho' z)^{\frac{-\lambda}{\gamma}}, & for\, i=0,\\[2mm] \dfrac{\Lambda'\lambda}{\beta\mu_O+\mu_P}(1-\rho' z)^{\frac{-\lambda}{\gamma}-1,} & for\, i=1,\\[2mm] \dfrac{\Lambda'\mu_P}{\beta\mu_O+\mu_P}(1-\rho' z)^{\frac{-\lambda}{\gamma}}, & for\, i=2.\end{cases} \tag{8.77}$$

$$P_{i,n}=\begin{cases}\dfrac{\Lambda'\beta\mu_O}{\beta\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n\prod_{j=0}^{n-1}(\lambda+j\gamma)\right], & i=0,\\[2mm] \dfrac{\Lambda'\lambda}{\beta\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n\prod_{j=0}^{n-1}(\lambda+(j+1)\gamma)\right], & i=1,\\[2mm] \dfrac{\Lambda'\mu_P}{\beta\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n\prod_{j=0}^{n-1}(\lambda+j\gamma)\right], & i=2.\end{cases} \tag{8.78}$$

where $\rho' = \dfrac{\lambda q}{\beta\mu_O + \mu_P}$ and $\Lambda' = \dfrac{(1-\rho')^{\frac{\lambda}{\mu}+1}}{(1-\rho') + \dfrac{\rho'}{q}}$.

Special Case: When $\beta = 1$ then double orbit retrial model with feedback reduces to double orbit retrial model without feedback as described in Section 8.2.

8.6. UNRELIABLE SERVER DOUBLE ORBIT RETRIAL MODEL

In this model, we deal with double orbit queue having failure prone server and balking behavior of the customers. ANFIS approach is also proposed to show the feasibility of the implementation of the soft computing approach for the performance evaluation of the present queueing system.

8.6.1. Model Description of Unreliable Server Model

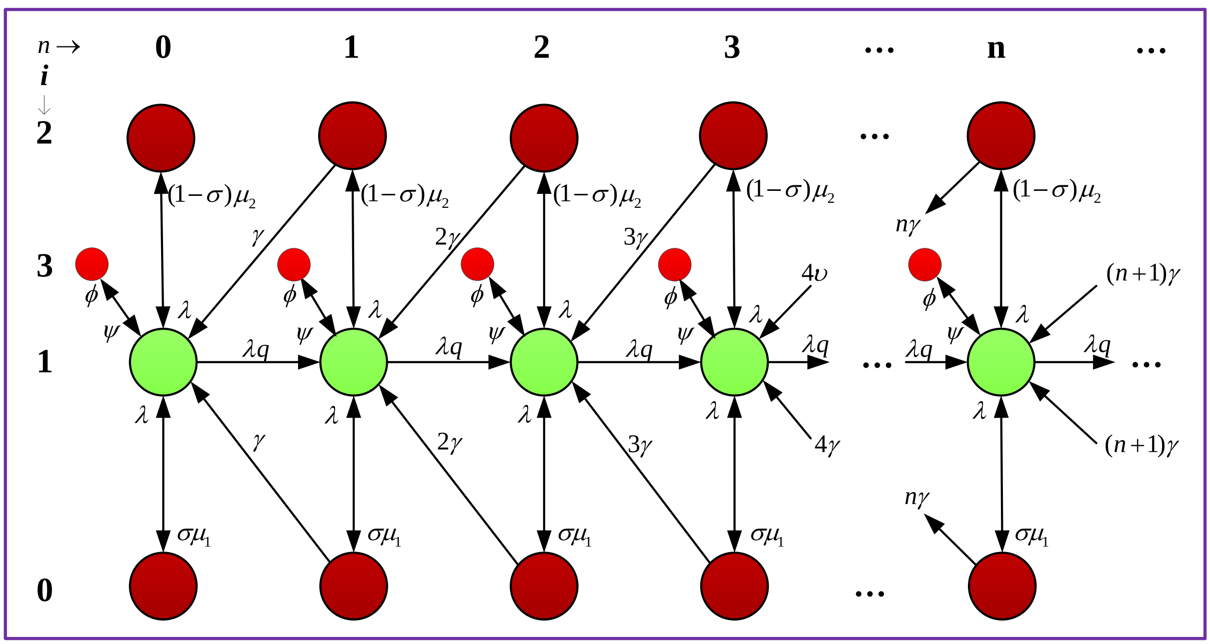


Figure 8.20: Transition state diagram for double orbit retrial model with unreliable server

Now we formulate the unreliable server double orbit model which coincides with the double orbit model described in Section 8.2 when server is taken to be reliable. In this model, the concept of unreliable server is also taken into account. To describe the unreliable server model, the additional assumption alongwith the assumptions (i)-(vi) described in Section 8.2.1 is that the server may fail at any time according to Poisson distribution with rate ϕ. As the server goes to breakdown state, it is immediately sent for repair which is performed according to exponential distribution with rate ψ. The transition state diagram for unreliable server model is depicted in Figure 8.20.

Now, the fourth level of system states $S(\tau)$ is defined as follows:

$S(\tau)=i=3,$ the server is brokendown and under repair.

8.6.2. The Analysis of the Unreliable Server Model

In order to derive the queue size and other performance metrics, the steady state Chapman-Kolmogorov equations by balancing the in-flows and out-flows (see Figure 8.20) are framed for all system states $i=0,1,2,3$ as follows:

$$-(\lambda+n\gamma)P_{0,n}+\mu_O P_{1,n}=0,\ n=0,1,2,... \tag{8.79}$$

$$-(\lambda p+\mu_O+\mu_P+\phi)P_{1,0}+\lambda P_{0,0}+\lambda P_{2,0}+\gamma P_{0,1}+\gamma P_{2,1}+\psi P_{3,0}=0 \tag{8.80}$$

$$-(\lambda q+\mu_O+\mu_P+\phi)P_{1,n}+\lambda q P_{1,n-1}+\lambda P_{0,n}+\lambda P_{2,n}+(n+1)\gamma P_{0,n+1}+(n+1)\gamma P_{2,n+1}+\psi P_{3,n}=0, \quad n=1,2,... \tag{8.81}$$

$$-(\lambda+n\gamma)P_{2,n}+\mu_P P_{1,n}=0,\ n=0,1,2,... \tag{8.82}$$

$$-\psi P_{3,n}+\phi P_{1,n}=0,\ n=0,1,2,... \tag{8.83}$$

Define the probability generating functions (PGFs) for unreliable server model as follows:

$$\Pi_i(z)=\sum_{n=0}^{\infty}P_{i,n}\,z^n, |z|\le 1,\ i=0,1,2,3$$

Lemma 8.3: For the unreliable server model, the probability generating functions $\Pi_i(z)$ is

$$\Pi_i(z)=\begin{cases}\dfrac{\Lambda''\mu_O}{\mu_O+\mu_P}(1-\rho z)^{-\frac{\lambda}{\gamma}}, & i=0,\\[2ex] \dfrac{\Lambda''\lambda}{\mu_O+\mu_P}(1-\rho z)^{-\frac{\lambda}{\gamma}-1}, & i=1,\\[2ex] \dfrac{\Lambda''\mu_P}{\mu_O+\mu_P}(1-\rho z)^{-\frac{\lambda}{\gamma}}, & i=2,\\[2ex] \dfrac{\phi\Lambda''\lambda}{\psi(\mu_O+\mu_P)}(1-\rho z)^{-\frac{\lambda}{\gamma}-1}, & i=3.\end{cases} \tag{8.84}$$

where $\rho = \dfrac{\lambda q}{\mu_O + \mu_P}$ and $\Lambda'' = \dfrac{(1-\rho)^{\frac{\lambda}{\mu}+1}}{(1-\rho)+\left(1+\dfrac{\phi}{\psi}\right)\dfrac{\rho}{q}}$.

Proof: Proof is similar as Lemma 8.1.

Theorem 8.4: For the unreliable server model, the steady state probability distribution, $P_{i,n}$ is

$$P_{i,n} = \begin{cases} \dfrac{\Lambda''\mu_O}{\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n \displaystyle\prod_{j=0}^{n-1}(\lambda + j\gamma)\right], & i=0,\ n=0,1,2,..., \\ \dfrac{\Lambda''\lambda}{\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n \displaystyle\prod_{j=0}^{n-1}(\lambda + (j+1)\gamma)\right], & i=1,\ n=0,1,2,..., \\ \dfrac{\Lambda''\mu_P}{\mu_O+\mu_P}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n \displaystyle\prod_{j=0}^{n-1}(\lambda + j\gamma)\right], & i=2,\ n=0,1,2,..., \\ \dfrac{\phi\Lambda''\lambda}{\psi(\mu_O+\mu_P)}\left[\dfrac{1}{n!}\left(\dfrac{\lambda}{\gamma}\right)^n \displaystyle\prod_{j=0}^{n-1}(\lambda + (j+1)\gamma)\right], & i=3,\ n=0,1,2,... \ . \end{cases} \tag{8.85}$$

Proof: Expanding $(1-\rho z)^{\frac{-\lambda}{\gamma}}$ and $(1-\rho z)^{\frac{-\lambda}{\gamma}-1}$ in (8.84), then after collecting the coefficient of z^n, we get (8.85).

8.6.3. Performance Measures for Unreliable Server Model

In order to explore the system performance, we derive steady state metrics namely, long run probabilities, mean queue size, mean waiting time and failure frequency, as follows:

(i) Queue Size

Let $E[N_T]$ and $E[N_S]$ denote the total mean number of customers in the queue and mean total number of customers in the system, respectively. Then

(a) $E[N_T] = \left[\dfrac{d}{dz}\displaystyle\sum_{i=0}^{3}\Pi_i(z)\right]_{z=1}$

$$E[N_T] = \frac{\rho^2\psi}{\left(\rho(1-q)+q\right)\psi+\rho\phi}\left[\frac{\mu_O+\mu_P}{\gamma}+\frac{1}{1-\rho}\left(1+\frac{\lambda}{\gamma}\right)\left(\frac{\phi+\psi}{\psi}\right)\right] \tag{8.86}$$

(b) $E[N_S] = E[N_T] +$ mean number of customers in the service

$$E[N_S]=\frac{\rho^2\psi}{\left(\rho(1-q)+q\right)\psi+\rho\phi}\left[\frac{\mu_O+\mu_P}{\gamma}+\frac{1}{1-\rho}\left(1+\frac{\lambda}{\gamma}\right)\left(\frac{\phi+\psi}{\psi}\right)+\frac{1}{\rho}\right] \tag{8.87}$$

(ii) Waiting Time

The mean waiting time in the queue and mean waiting time in the system spent by both types of customers are donated by $E[W_T]$, $E[W_S]$, respectively. Then

(a) $$E[W_T]=\frac{\rho^2}{\lambda q}\left[\frac{\mu}{\gamma}+\frac{1}{1-\rho}\left(1+\frac{\lambda}{\gamma}\right)\left(\frac{\phi+\psi}{\psi}\right)\right] \tag{8.88}$$

(b) $$E[W_S]=\frac{\rho^2}{\lambda q}\left[\frac{\mu}{\gamma}+\frac{1}{1-\rho}\left(1+\frac{\lambda}{\gamma}\right)\left(\frac{\phi+\psi}{\psi}\right)+\frac{1}{\rho}\right] \tag{8.89}$$

Results given in (8.88) and (8.89) are obtained by using Little's formulae $E[N_T]=\lambda_{eff}E[W_T]$ and $E[N_S]=\lambda_{eff}E[W_S]$, where λ_{eff} is effective arrival rate and it is given by $\lambda_{eff}=\dfrac{\lambda q\psi}{\left(\rho(1-q)+q\right)\psi+\rho\phi}$.

(iii) Failure Frequency

At any time, the server may be broken down so that the failure frequency (F_f) of the server is obtained using

$$F_f=\phi\Pi_1(1) \tag{8.90}$$

(iv) Long Run Probabilities

The long run probabilities of the server being idle (P_I), busy (P_{SB}) and broken down (P_{BD}) respectively, are obtained as

(a) $$P_I=\lim_{z\to1}\left(\Pi_0(z)+\Pi_2(z)\right)=\frac{q\psi(1-\rho)}{\left(\rho(1-q)+q\right)\psi+\rho\phi} \tag{8.91}$$

(b) $$P_{SB}=\lim_{z\to1}\Pi_1(z)=\frac{\psi\rho}{\left(\rho(1-q)+q\right)\psi+\rho\phi} \tag{8.92}$$

(c) $$P_{BD}=\lim_{z\to1}\Pi_3(z)=\frac{\phi\rho}{\left(\rho(1-q)+q\right)\psi+\rho\phi} \tag{8.93}$$

(v) Cost Function

To facilitate the customer's quality of service (QoS) in economic way, the system organizers must provide service in such a manner that the system cost can be reduced by controlling the service rate. Now, the total cost function $TC(\mu_1, \mu_2)$ per unit of time in terms of decision variables μ_1 and μ_2 is framed by including different cost elements per unit time. Let $C_{\mu_1}(C_{\mu_2})$ be cost/unit time incurred on each ordinary customers (premium customers) when they are getting service, $C_I(C_{SB})$ be the cost/unit time when the server is in idle (busy) state, C_H be the holding cost/unit time when the customers are waiting in the orbits. Thus

$$TC(\mu_1, \mu_2) = C_{\mu_1}\sigma\mu_1 + C_{\mu_2}(1-\sigma)\mu_2 + C_{SB}P_{SB} + C_I P_I + C_H E[N_T] \tag{8.94}$$

8.6.4. Numerical Results of Unreliable Server Model

In this section, we visualize the effects of system parameters namely, λ, γ, μ_1, μ_2 on different performance metrics of the double orbit retrial queueing model with unreliable server. For the cost analysis, quasi-Newton method is applied so as to obtain optimal service rates at minimum cost. For the computational purpose, default values of the system parameters are chosen as given in Table 8.10.

Table 8.10: Default system parameters

Parameter	q	λ	γ	σ	μ_1	μ_2	ϕ	ψ
Default values	0.8	20	10	0.5	20	30	2	3

(i) Sensitivity of the System Parameters

We analyze the effects of arrival rate (λ), retrial rate (γ), and service rates (μ_1, μ_2) on various performance metrics viz. $E[N_T]$, $E[W_T]$, $E[W_S]$, P_{SB}, P_I, F_f, TC and recorded in Tables 8.11-8.14. Also the variations in $E[N_S]$ corresponding to parameters λ, γ, μ_1, μ_2 are shown in Figures 8.22-8.25, respectively. Now we examine the sensitivity of different parameters based on numerical results as follows:

Table 8.11: Effect of λ on performance metrics

q	λ	$E[N_q]$	$E[W_q]$	$E[W_S]$	F_f	P_{SB}	P_I	P_{BD}	TC
	10	0.27	0.04	0.08	0.53	0.265	0.558	0.177	1304.17
0.4	15	0.65	0.08	0.12	0.68	0.341	0.432	0.227	1320.05
	20	1.25	0.13	0.17	0.79	0.397	0.338	0.265	1341.43
	10	0.46	0.07	0.11	0.56	0.280	0.533	0.187	1310.72
0.6	15	1.19	0.13	0.17	0.73	0.366	0.390	0.244	1337.55
	20	2.51	0.23	0.27	0.86	0.432	0.281	0.288	1381.21
	10	0.70	0.09	0.13	0.59	0.297	0.505	0.198	1318.93
0.8	15	1.99	0.20	0.24	0.79	0.395	0.342	0.263	1363.44
	20	4.96	0.42	0.46	0.94	0.472	0.213	0.315	1457.01

Table 8.12: Effect of γ on performance metrics

q	γ	$E[N_T]$	$E[W_T]$	$E[W_S]$	F_f	P_{SB}	P_I	P_{BD}	TC
	4	2.66	0.27	0.31	0.79	0.397	0.338	0.265	1383.78
0.4	8	1.49	0.15	0.19	0.79	0.397	0.338	0.265	1348.48
	12	1.10	0.11	0.15	0.79	0.397	0.338	0.265	1336.72
	4	5.28	0.49	0.53	0.86	0.432	0.281	0.288	1464.28
0.6	8	2.97	0.28	0.32	0.86	0.432	0.281	0.288	1395.05
	12	2.20	0.20	0.24	0.86	0.432	0.281	0.288	1371.98
	4	10.29	0.87	0.91	0.94	0.472	0.213	0.315	1617.01
0.8	8	5.84	0.49	0.53	0.94	0.472	0.213	0.315	1483.67
	12	4.36	0.37	0.41	0.94	0.472	0.213	0.315	1439.23

Table 8.13**:** Effect of μ_1 on performance metrics

q	μ_1	$E[N_T]$	$E[W_T]$	$E[W_S]$	F_f	P_{SB}	P_I	P_{BD}	TC
	10	1.82	0.21	0.26	0.88	0.441	0.265	0.294	1111.18
0.4	15	1.49	0.16	0.20	0.84	0.418	0.303	0.279	1224.72
	20	1.25	0.13	0.17	0.79	0.397	0.338	0.265	1341.43
	10	4.21	0.44	0.49	0.97	0.484	0.194	0.323	1185.32
0.6	15	3.15	0.31	0.35	0.91	0.456	0.240	0.304	1277.02
	20	2.51	0.23	0.27	0.86	0.432	0.281	0.288	1381.21
	10	11.57	1.08	1.13	1.07	0.536	0.107	0.357	1409.29
0.8	15	6.98	0.62	0.66	1.00	0.502	0.163	0.335	1394.61
	20	4.96	0.42	0.46	0.94	0.472	0.213	0.315	1457.01

- *Effect of arrival rate* (λ): From Table 8.11, it is seen that when γ, μ_1, μ_2 are constant and λ goes up, $E[N_T]$, $E[W_T]$, $E[W_S]$, P_{SB}, P_{BD} and F_f increase whereas P_I falls down. The trends for different joining probability (q) shown in Figure 8.22 depict the increasing trend of $E[N_S]$ as λ increases, which is same as per our expectation.

Table 8.14: Effect of μ_2 on performance metrics

q	μ_2	$E[N_T]$	$E[W_T]$	$E[W_S]$	F_f	P_{SB}	P_I	P_{BD}	TC
	16	2.22	0.27	0.32	0.92	0.462	0.231	0.308	1024.15
0.4	20	1.82	0.21	0.26	0.88	0.441	0.265	0.294	1111.18
	24	1.55	0.17	0.21	0.85	0.423	0.296	0.282	1201.71
	16	5.69	0.62	0.68	1.02	0.508	0.153	0.339	1131.36
0.6	20	4.21	0.44	0.49	0.97	0.484	0.194	0.323	1185.32
	24	3.32	0.33	0.37	0.92	0.462	0.231	0.308	1257.38
	16	23.55	2.31	2.37	1.13	0.566	0.057	0.377	1670.38
0.8	20	11.57	1.08	1.13	1.07	0.536	0.107	0.357	1409.29
	24	7.59	0.68	0.72	1.02	0.508	0.153	0.339	1388.31

- *Effect of retrial rate* (γ) *:* The retrial rate (γ) affects significantly $E[N_T]$, $E[W_T]$, $E[W_S]$ as can be seen from Table 8.12. It is observed that P_{SB} , P_I , P_{BD} and F_f do not much change with respect of γ . From Table 8.12, we see that as the customers retry for the service with faster rate, mean number of customers in the orbits (ordinary and executive) becomes less and subsequently their waiting time is also reduced. The same fact can also be seen from Figure 8.23.
- *Effect of service rates* $(\mu_1$ and $\mu_2)$: The mean number of customers in the orbit or in the system and mean waiting time spent by the customers always depend on the service provided by the system organizer. When the system organizer provides faster service and arrival rate of the customers is constant, then $E[N_T]$, $E[W_T]$, $E[W_S]$, F_f , P_{BD} and P_{SB} seem lower down whereas P_I increases as can be seen from the Tables 8.13 and 8.14. Figures 8.24 and 8.25 reveal that $E[N_S]$ becomes less when the server facilitates service to the customers with faster rate.

(ii) ANFIS computing

The artificial neuro fuzzy inference system (ANFIS) is a soft computing tool which works based on neural network and fuzzy inference system. ANFIS has been successfully used for the prediction of many complex systems in the diverse area. The applications of ANFIS may be seen in the fields of traffic modeling apart from financial engineering, industrial engineering, food engineering, telecommunication system, etc. ANFIS technique is implemented to authenticate the steady state results. The procedural steps of ANFIS can be seen in Section 1.4.7 of Chapter 1. For analyzing the unreliable server double orbit retrial queueing model with balking, we consider the fuzzified input parameters (i) λ (ii) γ (iii) μ_1 (iv) μ_2. The Gaussian membership function is taken corresponding to input parameters (i) λ (ii) γ (iii) μ_1 (iv) μ_2 and the linguistic variables for the input parameters are given in Table 8.15.

Table 8.15: Input parameters, membership function and linguistic variables for ANFIS

Sr. No.	Input parameter	Number of membership function	Linguistic variables
1.	λ	5	Very small, small, medium, large, very large
2.	γ	3	Small, medium, large
3.	μ_1	4	Small, medium, large, very large
4.	μ_2	4	Small, medium, large, very large

Matlab software is used to produce the ANFIS results to compare the numerical results obtained by analytical method by employing the *anfisedit* command and setting the default parameters as given in Table 8.10. The accuracy in the ANFIS results for $E[N_S]$ is examined by evaluating the absolute percentage errors (Δ) given by

$$\Delta = \frac{|E[\hat{N}_S] - E[N_S]|}{E[N_S]} \times 100\% \tag{8.95}$$

where Δ is absolute percentage error, $E[\hat{N}_S]$ is estimated value of expected number of customers in the system by ANFIS; $E[N_S]$ is exact value expected number of customers in the system by analytical method.

The exact values of $E[N_S]$ by analytical method using (8.87) and the estimated values $E[\hat{N}_S]$ of $E[N_S]$ by ANFIS technique for varying values of (i) λ (ii) γ (iii) μ_2 (iv) μ_1 for $q = 0.6, 0.7, 0.8$ are recorded in Tables 8.16-8.19. We notice that the values of $E[\hat{N}_S]$ and $E[N_S]$ are almost coincide which shows the feasibility of the ANFIS approach in complex model of real time system. Further, absolute percentage error and accuracy in estimated value in percentage of $E[N_S]$ are evaluated by using (8.95) and summarized in Tables 8.16-8.19. We see that less value of Δ shows that how our ANFIS approach is closer with the results obtained by analytical method.

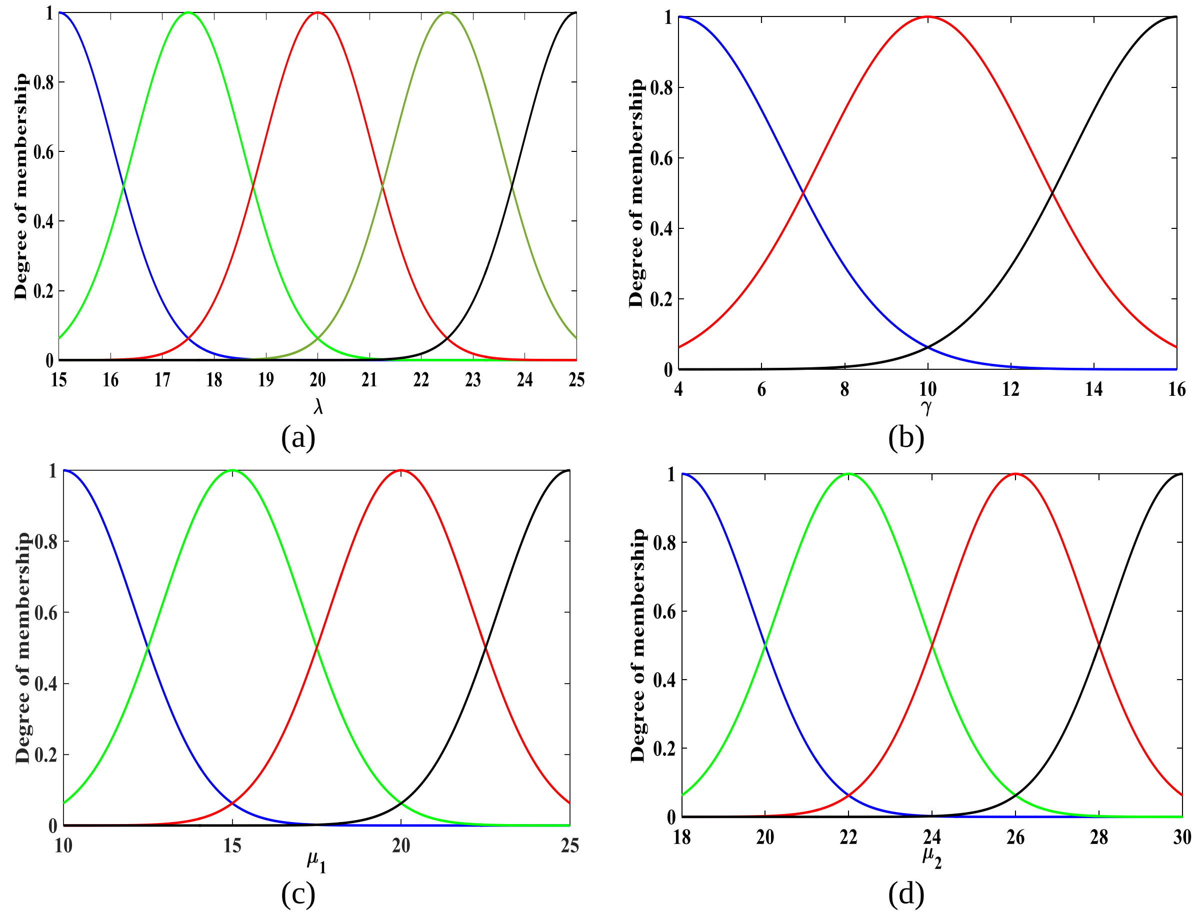


Figure 8.21: Membership function for (a) λ (b) γ (c) μ_1 (d) μ_2

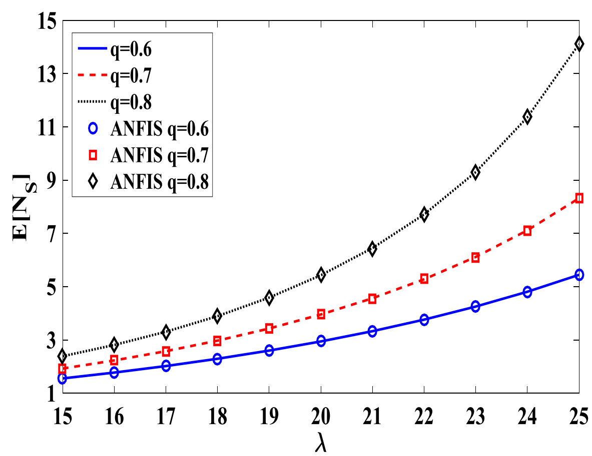


Figure 8.22: $\mathrm{E}[\mathrm{N_S}]$ vs λ for $q = 0.6, 0.7, 0.8$

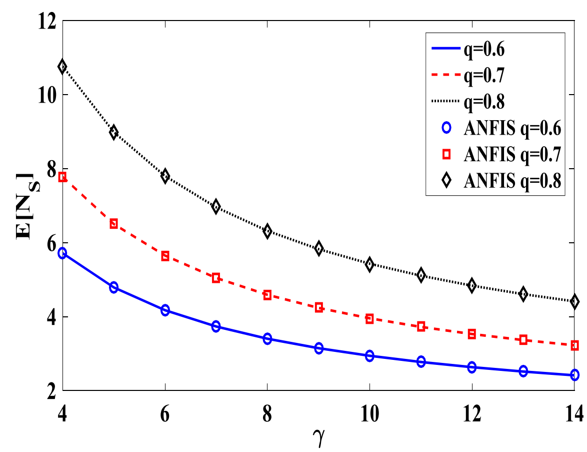


Figure 8.23: $\mathrm{E}[\mathrm{N_S}]$ vs γ for $q = 0.6, 0.7, 0.8$

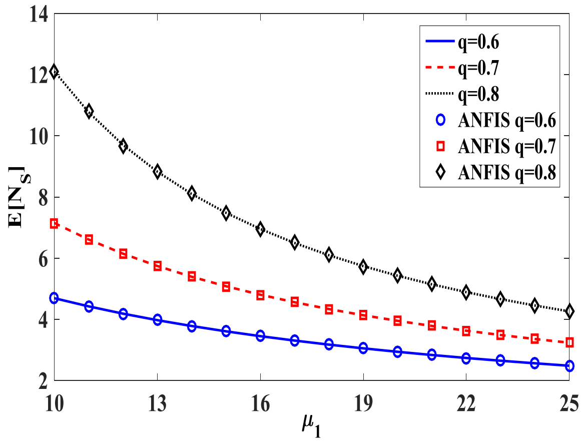


Figure 8.24: $\mathrm{E[N_S]}$ vs μ_1 for $q = 0.6, 0.7, 0.8$

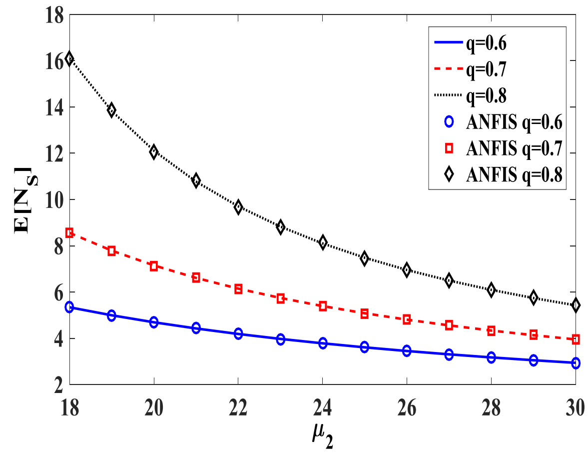


Figure 8.25: $\mathrm{E[N_S]}$ vs μ_2 for $q = 0.6, 0.7, 0.8$

The membership function for input parameter (i) λ (ii) γ (iii) μ_1 (iv) μ_2 respectively, are shown in Figures 8.21(a-d). Figures 8.22-8.25 show best match between exact value of $E[N_S]$ and $E[\hat{N}_S]$. The ANFIS results $E[\hat{N}_S]$ are depicted in Figures 8.22-8.25 by the tick marked whereas the analytical results are indicated by the lines (continuous line, dash line, and dotted line) for the prediction of $E[N_S]$.

Table 8.16**:** Δ, $E[N_S]$ and $E[\hat{N}_S]$ by varying λ for $q = 0.6, 0.7, 0.8$

λ	$q = 0.6$			$q = 0.7$			$q = 0.8$		
	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)
15	1.553	1.552	0.008	1.924	1.924	0.015	2.387	2.386	0.034
16	1.773	1.773	0.027	2.229	2.230	0.049	2.810	2.813	0.104
17	2.018	2.017	0.072	2.576	2.573	0.124	3.307	3.298	0.248
18	2.292	2.295	0.119	2.973	2.978	0.199	3.892	3.907	0.382
19	2.599	2.595	0.154	3.428	3.419	0.249	4.590	4.569	0.459
20	2.942	2.947	0.171	3.954	3.964	0.268	5.428	5.453	0.472
21	3.327	3.322	0.144	4.563	4.553	0.218	6.448	6.425	0.366
22	3.761	3.765	0.108	5.276	5.284	0.158	7.711	7.730	0.253
23	4.252	4.249	0.067	6.116	6.110	0.095	9.302	9.289	0.146
24	4.809	4.810	0.032	7.115	7.118	0.044	11.360	11.367	0.064
25	5.444	5.443	0.011	8.319	8.318	0.014	14.107	14.104	0.020
Average of Δ			0.083		0.130			0.232	
Accuracy in predicted value (%)			99.917		99.870			99.768	

Table 8.17: Δ, $E[N_S]$ and $E[\hat{N}_S]$ by varying γ for $q=0.6,0.7,0.8$

γ	$q=0.6$			$q=0.7$			$q=0.8$		
	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)
4	5.711	5.711	0.007	7.772	7.771	0.007	10.761	10.760	5.711
5	4.788	4.789	0.026	6.499	6.501	0.026	8.983	8.986	4.788
6	4.173	4.169	0.076	5.650	5.646	0.077	7.798	7.792	4.173
7	3.733	3.739	0.145	5.044	5.052	0.148	6.952	6.962	3.733
8	3.403	3.396	0.220	4.590	4.580	0.225	6.317	6.302	3.403
9	3.147	3.156	0.276	4.236	4.248	0.282	5.823	5.840	3.147
10	2.942	2.934	0.252	3.954	3.943	0.259	5.428	5.414	2.942
11	2.774	2.779	0.191	3.722	3.729	0.197	5.105	5.115	2.774
12	2.634	2.631	0.111	3.529	3.525	0.114	4.835	4.830	2.634
13	2.516	2.517	0.039	3.366	3.367	0.041	4.607	4.609	2.516
14	2.414	2.414	0.011	3.226	3.226	0.011	4.412	4.411	2.414
Average of Δ			0.123			0.126			0.128
Accuracy in predicted value (%)			99.877			99.874			99.872

Table 8.18: Δ, $E[N_S]$ and $E[\hat{N}_S]$ by varying μ_2 for $q=0.6,0.7,0.8$

μ_2	$q=0.6$			$q=0.7$			$q=0.8$		
	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)
18	5.341	5.340	0.022	8.557	8.554	0.032	16.110	16.101	0.055
19	4.996	4.999	0.065	7.791	7.799	0.098	13.823	13.848	0.184
20	4.694	4.689	0.103	7.153	7.141	0.168	12.107	12.066	0.344
21	4.427	4.431	0.096	6.612	6.623	0.170	10.772	10.814	0.389
22	4.189	4.188	0.024	6.149	6.145	0.060	9.704	9.686	0.184
23	3.976	3.974	0.063	5.747	5.742	0.090	8.830	8.818	0.144
24	3.785	3.788	0.098	5.395	5.404	0.166	8.102	8.130	0.354
25	3.611	3.609	0.050	5.085	5.079	0.102	7.485	7.466	0.261
26	3.453	3.452	0.030	4.808	4.807	0.033	6.956	6.955	0.023
27	3.309	3.311	0.077	4.561	4.567	0.126	6.498	6.514	0.251
28	3.176	3.174	0.062	4.338	4.333	0.117	6.097	6.080	0.268
29	3.055	3.055	0.024	4.137	4.139	0.056	5.743	5.751	0.148
30	2.942	2.942	0.004	3.954	3.953	0.014	5.428	5.425	0.046
Average of Δ			0.055			0.095			0.204
Accuracy in predicted value (%)			99.945			99.905			99.796

Table 8.19: Δ, $E[N_S]$ and $E[\hat{N}_S]$ by varying μ_1 for $q = 0.6, 0.7, 0.8$

μ_1	$q=0.6$			$q=0.7$			$q=0.8$		
	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)	$E[N_S]$	$E[\hat{N}_S]$	Δ (%)
10	4.694	4.692	0.041	7.153	7.148	0.062	12.107	12.094	0.106
11	4.427	4.431	0.111	6.612	6.624	0.173	10.772	10.806	0.315
12	4.189	4.184	0.110	6.149	6.138	0.181	9.704	9.670	0.354
13	3.976	3.976	0.011	5.747	5.747	0.006	8.830	8.832	0.022
14	3.785	3.789	0.118	5.395	5.405	0.187	8.102	8.130	0.348
15	3.611	3.609	0.047	5.085	5.080	0.087	7.485	7.470	0.198
16	3.453	3.450	0.098	4.808	4.801	0.148	6.956	6.939	0.256
17	3.309	3.312	0.088	4.561	4.568	0.150	6.498	6.518	0.304
18	3.176	3.178	0.063	4.338	4.342	0.089	6.097	6.105	0.139
19	3.055	3.051	0.108	4.137	4.130	0.177	5.743	5.723	0.340
20	2.942	2.941	0.014	3.954	3.953	0.011	5.428	5.428	0.008
21	2.837	2.840	0.102	3.786	3.792	0.161	5.146	5.161	0.296
22	2.740	2.740	0.028	3.632	3.630	0.054	4.892	4.886	0.120
23	2.650	2.648	0.079	3.491	3.487	0.121	4.663	4.653	0.213
24	2.566	2.568	0.085	3.360	3.365	0.140	4.454	4.466	0.266
25	2.486	2.486	0.032	3.239	3.237	0.056	4.263	4.258	0.110
Average of Δ			0.071			0.113			0.212
Accuracy in predicted value (%)			99.929			99.887			99.788

(iii) Cost Optimization

In this subsection, our objective is to minimize the cost function in order to determine the decision variables which are service rates μ_1 and μ_2. From the cost function given in Equation (8.94), it is noticed that it is highly non-linear and complex so it is a tough task to minimize it analytically. Therefore, we use computational technique to achieve our goal of minimum cost at optimal service rates. The decision variables $(\mu_1 \text{ and } \mu_2)$ and minimum cost are evaluated by implementing quasi-Newton method in MATLAB software.

For the implementation of the quasi-Newton method, we consider the following two cost sets:

Cost Set-I: $C_{\mu_1} = \$50, C_{\mu_2} = \$50, C_{SB} = \$110, C_I = \$30, C_H = \$30.$

Cost Set-II: $C_{\mu_1} = \$50, C_{\mu_2} = \$50, C_{SB} = \$90, C_I = \$35, C_H = \$50.$

Quasi-Newton Method (QNM)

The quasi-Newton method is used to find the approximate minimum value of $TC(\mu_1,\mu_2)$ in the feasible range of μ_1 and μ_2. Let $TC(\mu_1^*,\mu_2^*)$ denote the minimum value of $TC(\mu_1,\mu_2)$ at optimal value (μ_1^*,μ_2^*). The minimization problem is mathematically formulated as

(OP) $$\min_{\mu_1>0,\mu_2>0} TC(\mu_1,\mu_2) \text{ subject to } \rho<1 \tag{8.96}$$

Using the algorithmic steps of QNM given in Section 1.4.8 of Chapter-1, we implement QNM.

Inputs: $q,\ \lambda,\gamma,\ \sigma,\ \phi,\ \psi,\ C_{\mu_1},\ C_{\mu_2},\ C_{SB},\ C_I,\ C_H$ and tolerance ε of $\left|\dfrac{\partial TC(\mu_1,\mu_2)}{\partial\mu_1}\right|$, $\left|\dfrac{\partial TC(\mu_1,\mu_2)}{\partial\mu_2}\right|$.

Output: Approximate solution of $(\mu_1^*,\mu_2^*,TC(\mu_1^*,\mu_2^*))$.

Applying algorithmic steps of quasi-Newton method, the minimum cost for Cost Set-I using default parameters given in Table 8.10 and initial trial solution $(\mu_1,\mu_2)=(15,20)$ is $TC(\mu_1,\mu_2)=\$2512.44$. Also, we notice that the expected minimum cost $TC(\mu_1^*,\mu_2^*)=\$1528.76$ is attained at $(\mu_1^*,\mu_2^*)=(20.4435,25.4435)$ after performing 10 iterations (for tolerance $\varepsilon=10^{-6}$) as can be seen from Table 8.20.

Again, we select Cost Set-1 and parameter values $\lambda=22,\ \phi=5,\ \psi=7$, initial trial solution $(\mu_1,\mu_2)=(15,22)$ so that the corresponding cost is $TC(\mu_1,\mu_2)=\$4174.38$. Using the QNM, the minimum value of $TC(\mu_1,\mu_2)$ is achieved after 10 iterations. The procedures of QNM are repeated for Cost Set -II and minimum values of cost at optimal service rates are recorded in Table 8.21. The surface graphs for the cost function for varying values of service rates μ_1 and μ_2 are also shown in Figures 8.26-8.29. The convexity of $TC(\mu_1,\mu_2)$ is quite clear from these figures.

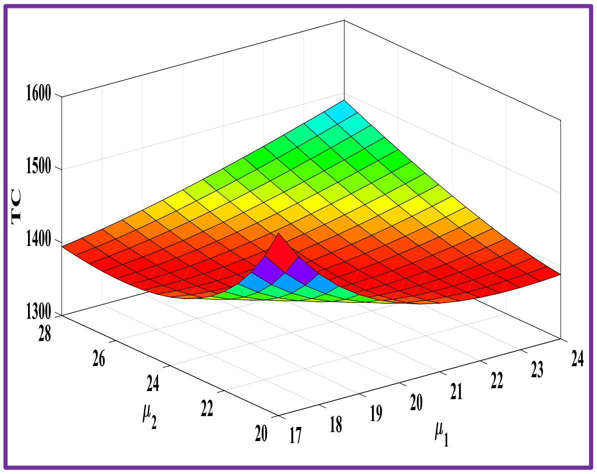


Figure 8.26: TC vs (μ_1,μ_2) for Cost Set-I and $(\lambda,\phi,\psi)=(20,2,3)$

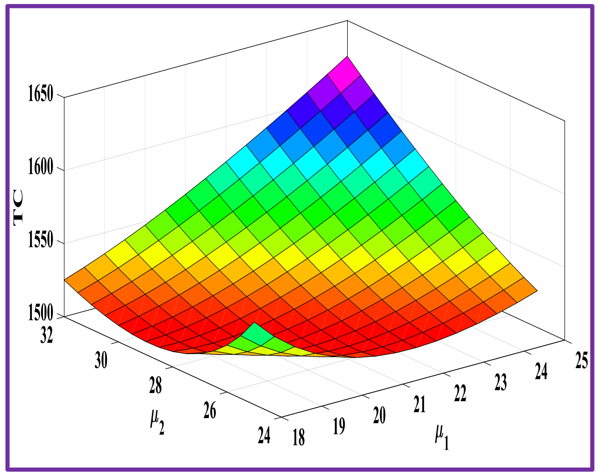


Figure 8.27: TC vs (μ_1,μ_2) for Cost Set-I and $(\lambda,\phi,\psi)=(22,5,7)$

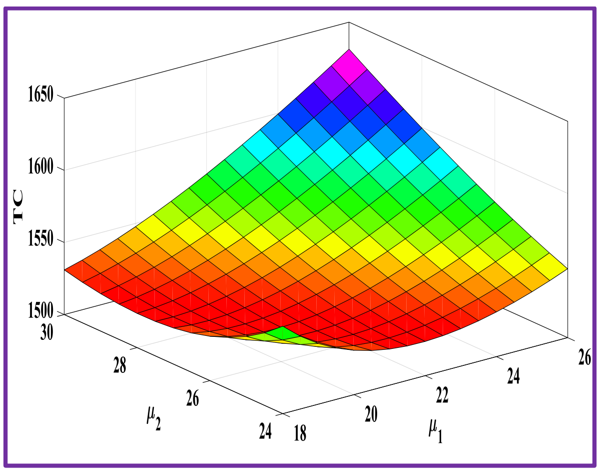


Figure 8.28: TC vs (μ_1,μ_2) for Cost Set-II and $(\lambda,\phi,\psi)=(20,2,3)$

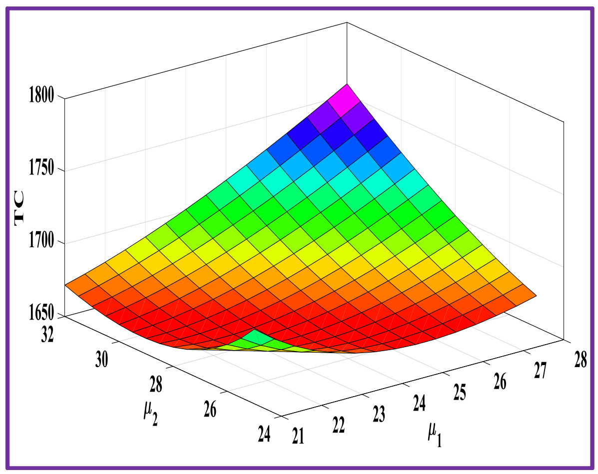


Figure 8.29: TC vs (μ_1,μ_2) for Cost Set-II and $(\lambda,\phi,\psi)=(22,5,7)$

Table 8.20: Quasi-Newton method in searching of (μ_1^*,μ_2^*) and $TC(\mu_1^*,\mu_2^*)$. (Cost Set -1, $\lambda=20,\phi=2,\psi=3$).

Iterations	μ_1	μ_2	$TC(\mu_1,\mu_2)$	Max. Tolerance
0	15	20	1885.67	295
1	16.0000	21.0000	1551.09	90.5
2	16.4417	21.4417	1486.44	58.5
3	17.2487	22.2487	1420.98	26.5
4	17.9174	22.9174	1395.93	12.1
5	18.4837	23.4837	1386.92	4.23
6	18.7865	23.7865	1385.36	1
7	18.8806	23.8806	1385.26	0.109
8	18.8921	23.8921	1385.25	0.00319
9	18.8924	23.8925	1385.25	1.02×10^{-5}

Table 8.21: Optimal table for evaluating $TC(\mu_1^*, \mu_2^*)$.

(λ, ϕ, ψ)	$(\mu_1^*, \mu_2^*, TC(\mu_1^*, \mu_2^*))$	
	Cost Set-I	Cost Set-II
(20, 2, 3)	(18.8924, 23.8925, $1385.25)	(19.9211, 25.9211, $1523.68)
(22, 5, 7)	(19.9365, 26.9356, $1508.78)	(23.0966, 27.0966, $1660.60)

8.7. CONCLUDING RAMARKS

The steady state analysis of the single server double orbit retrial queue with customers' balking is done by considering both reliable server and unreliable server. The probability generating function method is used to obtain explicit formulae for various queueing system indices. The model is also investigated in fuzzy environment by constructing membership functions for the system performance metrics by parametric non-linear programing approach. The customers' behavior of joining the system based on reward-cost structure analyzed would be beneficial to the customers' point of view to decide whether to join the queue or not to join the queue. A cost function framed is minimized using genetic algorithm which can help the decision maker to facilitate service at minimum cost. Furthermore, the applications of our study on single server double orbit queueing system can be noticed in daily routine life including hospitals, malls, online shopping, telecommunication systems, etc. A real life application example of ticket booking counter presented exhibits the realistic and fuzzified metrics for the future system design.

In unreliable server model, a soft computing based ANFIS approach is applied which authenticates the steady state results obtained by analytical method. For illustration and validity of cost function, the minimum cost of the system and optimal service rates determined by employing quasi-Newton method demonstrate the applicability of model for decision making for the concerned system.

Conclusions

The state-dependent queueing models with control strategies investigated in the present thesis are of significant importance in the real life congestion scenarios. Due to various applications of such queues in day to day activities and several other places have motivated us to develop and analyze the new models in this area. In the present thesis, the state dependent queueing models investigated by incorporating various realistic key features such as balking, reneging, feedback, server's breakdown, server's vacation, retrial orbit, control policies, threshold recovery, etc. may be helpful to study the complex systems arising in many congestion situations. The state-dependent queueing situations studied for repairable machining system and time-sharing system can be seen at computers/communication networks, manufacturing/production lines, etc. It is worth-while to highlight the main contributions and noble features of the present thesis as follows:

- The ***finite capacity*** queueing models with various features including admission control *F*-policy are developed in Chapters 2-5. The applications of such models may be noticed at various places including call centers, shopping malls, banks, industrial and manufacturing system, etc. The machine repair problems with specific concepts such as server vacation, standbys, retrial orbit, etc. are studied by developing the finite population model in Chapters 6-7.
- The arriving customers in the system or the customers waiting in the queue for their turn may be discouraged due to formation of long queue in the system or slow service provided by the server. The ***discouragement*** behavior of the customers considered in Chapters 2,3,5 and 8 is a most common and realistic phenomenon in day to day activities and congestion scenarios and this can be noticed where long queues are built.
- In real time systems such as industrial and production system, communication system, computer network, and many others, the facility of retrial orbit for waiting of jobs/customers may be noticed. The retrial queues deal with congestion situations encountered in daily routine life and hold a significant place in the area of cellular networks, telecommunication systems, railway ticket counters, ATMs, etc. The ***retrial orbit*** queueing models are developed in Chapters 2, 3, 5, 6, 8.
- It is worth-noting that the concepts of ***double orbit*** and two types of customers namely, ordinary and premium are taken into account and are incorporated in infinite capacity

retrial model with customer's balking, feedback and unreliable as can be seen in Chapter 8.

- Due to space, cost, delay and other constraints; the admission control of customers/jobs in queueing system is one of the important measures. Optimal control strategies can be employed to control the customer's discouragement behavior (cf. Chapters 4 and 8) and may be helpful in resolving the delay and blocking issues in the queueing system. The ***admission control F-policy*** proposed in Chapters 2-6, may be beneficial in many places where a limited number of customers can queue up as such control of admission of customers/jobs becomes essential.
- In queueing systems with the vacation, server may be unavailable for a certain period of time due to some reasons. The concept of ***vacation*** incorporated for the study of multi-component machining system in Chapter 8 can be realized in the real time industrial systems operating in machining environment. One can also notice that the service facility is not always available. The concept of server's vacation and/or unreliable server can also be seen in many day-to-day as well as industrial queueing situations.
- In the present thesis some queueing models have studied in fuzzy environment (cf. Chapters 7-8). As far as the practical utility is concerned, the fuzzy queueing models portray more versatile real time congestion situations as compared to frequently used crisp queueing models.

From application view points, by incorporating admission control F-policy to the concerned queueing systems, the problem of waiting can be resolved economically to some extent. It is noticed that the congestion of customers/jobs in the concerned queueing system with retrial orbit can also be managed by incorporating the admission control F-policy. It is suggested that the discouragement behavior of the customers may be reduced by many ways such as providing fast service to the customers, adding removable additional server(s) alongwith the existing permanent servers, providing some reward to the customers after getting service for their satisfaction, etc.

The investigations done in the present thesis can be further extended by incorporating the concept of priority, bulk input and/or general input but the analysis and computation will become tedious. The service control based on N-policy may also be incorporated along with admission control F-policy and threshold recovery policy. There is further scope to work on the admission control F-policy for multi-server queueing model for the time sharing systems. Multi-component machine repair model operating under optimal control policy can be extended by using the concepts of working vacation and threshold recovery policy. There is need to study the optimal control of the real time queueing systems in generic frameworks by including the customer' balking behavior, general retrial times, reneged customers, feedback, etc.

There are enormous applications of optimal control strategies in real time congestion problems encountered in routine life congestion scenarios as well as in industrial delay problems such as in data centers, wireless communication systems, computer networks, telecommunication systems, manufacturing and production system, etc. We hope that our investigations on the optimal control strategies for state-dependent queues may facilitate the valuable insights to the system designers and the decision makers to tackle with system cost minimization, customers' profit maximization and traffic management issues involved in the concerned congestion problems encountered in day-to-day as well as industrial scenarios.

References

[1] Abbasbandy, S., Nuraei, R., & Ghanbari, M. (2013). Revision of sign distance method for ranking of fuzzy numbers. Iranian Journal of Fuzzy Systems, 10(4), 101–117.

[2] Abou-El-Ata, M. O., & Shawky, A. I. (1992). The single-server Markovian overflow queue with balking, reneging and an additional server for longer queues. Microelectronics Reliability, 32(10), 1389–1394.

[3] Abou-El-Ata, M., & Hariri, A. (1992). The M/M/c/N queue with balking and reneging. Computers & Operations Research, 19(8), 713–716.

[4] Adan, I. J. B. F., & Kulkarni, V. G. (2003). Single-server queue with Markov-dependent inter-arrival and service times. Queueing Systems, 45(2), 113–134.

[5] Adiri, I and Avi-Itzhak, B. (1969). A time-sharing queue. Management Science, 15(11), 639–657.

[6] Artalejo, J. R., & Lopez-Herrero, M. J. (2000). On the single server retrial queue with balking. INFOR: Information Systems and Operational Research, 38(1), 33–50.

[7] Artalejo, J., & Falin, G. (2002). Standard and retrial queueing systems: A comparative analysis. Revista Matematica Complutense, 15(1), 101–129.

[8] Bagherinejad, J., & Pishkenari, S. B. (2016). Analysis of FM/FM/c queuing system: using fuzzy approach and parametric nonlinear programming. International Journal of Industrial and Systems Engineering, 23(2), 125–140.

[9] Bagyam, J. E. A., & Chandrika, K. U. (2019). Fuzzy analysis of bulk arrival two phase retrial queue with vacation and admission control. The Journal of Analysis, 27(1), 209–232.

[10] Banerjee, A., & Gupta, U. C. (2012). Reducing congestion in bulk-service finite-buffer queueing system using batch-size-dependent service. Performance Evaluation, 69(1), 53–70.

[11] Baumann, H., & Sandmann, W. (2017). Multi-server tandem queue with Markovian arrival process, phase-type service times, and finite buffers. European Journal of Operational Research, 256(1), 187–195.

[12] Bhardwaj, R. (2017). Fuzzy parametric bulk queue model to communication system through α -cut. Journal of Mathematics & Systems Sciences, 13(1), 51–56.

[13] Bhardwaj, R., Singh, T. P., & Kumar, V. (2019). Designing bulk arrival queue model to an interdependent communication system with fuzzy parameters. In Performance Prediction and Analytics of Fuzzy, Reliability and Queuing Models. Asset Analytics (Performance and Safety Management) (pp. 225–234). Springer, Singapore.

[14] Bhardwaj, R., Singh, T., & Kumar, V. (2018). Mathematical study of queue system with impatient customers under fuzzy environment. In Soft Computing: Theories and Applications (pp. 679–688). Springer, Singapore.

[15] Boualem, M., Djellab, N., & Aïssani, D. (2014). Stochastic bounds for a single server queue with general retrial times. Bulletin of the Iranian Mathematical Society, 40(1), 183–198.

[16] Bouazzi, I., Bhar, J., & Atri, M. (2017). Priority-based queuing and transmission rate management using a fuzzy logic controller in WSNs. ICT Express, 3(2), 101–105.

[17] Bouchentouf, A. A., & Messabihi, A. (2018). Heterogeneous two-server queueing system with reverse balking and reneging. OPSEARCH, 55(2), 251–267.

[18] Buckley, J. J. (2004). Fuzzy Probabilities and Fuzzy Sets for Web Planning (1st ed.). Springer-Verlag Berlin Heidelberg.

[19] Chandrasekaran, M., Muralidhar, M., & Dixit, U. S. (2013). Online optimization of multipass machining based on cloud computing. The International Journal of Advanced Manufacturing Technology, 65(1–4), 239–250.

[20] Chang, C.-J., Chang, F.-M., & Ke, J.-C. (2014a). Economic application in a Bernoulli F -policy queueing system with server breakdown. International Journal of Production Research, 52(3), 743–756.

[21] Chang, C.-J., Chang, F.-M., & Ke, J.-C. (2014b). Optimization of machine repair system with controlling arrival and switching failure. Journal of Testing and Evaluation, 42(5), 1278–1287.

[22] Chang, C.-J., Ke, J.-C., & Huang, H.-I. (2011). The optimal management of a queueing system with controlling arrivals. Journal of the Chinese Institute of Industrial Engineers, 28(3), 226–236.

[23] Chang, F. M., Liu, T. H., & Ke, J. C. (2018). On an unreliable-server retrial queue with customer feedback and impatience. Applied Mathematical Modelling, 55, 171–182.

[24] Chang, F.-M., Ke, J.-C., & Liou, C.-H. (2013). Controlling arrival for the machine repair problem with switching failure. Journal of Physics: Conference Series, 410(1), 012116.

[25] Chang, J., & Wang, J. (2018). Unreliable M/M/1/1 retrial queues with set-up time. Quality Technology & Quantitative Management, 15(5), 589–601.

[26] Chen, J.-Y., Yen, T.-C., & Wang, K.-H. (2016). Cost optimization of a single-server queue with working breakdowns under the N Policy. Journal of Testing and Evaluation, 44(5), 2059–2067.

[27] Chen, S. P. (2006). Non-linear programming for the optimization of machine repair problems in fuzzy environments. Engineering Optimization, 38(7), 789–799.

[28] Chen, S. P., & Chang, P. C. (2006). A mathematical programming approach to supply chain models with fuzzy parameters. Engineering Optimization, 38(6), 647–669.

[29] Chen, W.-L. (2018). Reliability and sensitivity analysis of the controllable repair system with warm standbys under the recovery threshold policy. Journal of Testing and Evaluation, 47(2), 1311–1331.

[30] Choudhury, A., & Medhi, P. (2011). Some aspects of balking and reneging in finite buffer queues. RAIRO - Operations Research, 45(3), 223–240.

[31] Choudhury, G., & Ke, J.-C. (2014). An unreliable retrial queue with delaying repair and general retrial times under Bernoulli vacation schedule. Applied Mathematics and Computation, 230, 436–450.

[32] Choudhury, G., Tadj, L., & Deka, M. (2015). An unreliable server retrial queue with two phases of service and general retrial times under Bernoulli vacation schedule. Quality Technology & Quantitative Management, 12(4), 437–464.

[33] Cox, D. R. (1955). The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. Mathematical Proceedings of the Cambridge Philosophical Society, 51(03), 433.

[34] Cox, D.R., & Miller, H.D. (2017). The Theory of Stochastic Processes. Routledge, Taylor and Francis Group.

[35] Dounis, A. I., Leftheriotis, G., Stavrinidis, S., & Syrrokostas, G. (2016). Electrochromic device modeling using an adaptive neuro-fuzzy inference system: A model-free approach. Energy and Buildings, 110, 182–194.

[36] Durga Rao, K., Gopika, V., Kushwaha, H. S., Verma, A. K., & Srividya, A. (2007). Test interval optimization of safety systems of nuclear power plant using fuzzy-genetic approach. Reliability Engineering & System Safety, 92(7), 895–901.

[37] Ernst, P. A., Asmussen, S., & Hasenbein, J. J. (2018). Stability and busy periods in a multiclass queue with state-dependent arrival rates. Queueing Systems, 90(3–4), 207–224.

[38] Flapper, S. D., Gayon, J.-P., & Lim, L. L. (2014). On the optimal control of manufacturing and remanufacturing activities with a single shared server. European Journal of Operational Research, 234(1), 86–98.

[39] Gal, T. (1979). Postoptimal analysis, parametric programming, and related topics (2nd ed.). Berlin, McGraw-Hill New York.

[40] Gao, S., Niu, X., & Li, T. (2017). Analysis of a constant retrial queue with joining strategy and impatient retrial customers. Mathematical Problems in Engineering, 2017, 1–8.

[41] Gao, S., Wang, J., & Li, W. W. (2014). An M/G/1 retrial queue with general retrial times, working vacations and vacation interruption. Asia-Pacific Journal of Operational Research, 31(2), 1440006.

[42] Gao, S., Wang, J., & Van Do, T. (2016). A repairable retrial queue under Bernoulli schedule and general retrial policy. Annals of Operations Research, 247(1), 169–192.

[43] Geem, Z. W., & Yoon, Y. (2017). Harmony search optimization of renewable energy charging with energy storage system. International Journal of Electrical Power & Energy Systems, 86, 120–126.

[44] Geem, Z. W., Kim, J. H. & Loganathan, G. V. (2001). A new heuristic optimization algorithm: Harmony Search. Simulation, 76(2), 60–68.

[45] Goswami, V. (2016). Relationship between randomized F-policy and randomized N-policy in discrete-time queues. OPSEARCH, 53(1), 131–150.

[46] Gross, D., Shortie, J. F., Thompson, J. M., & Harris, C. M. (2008). Fundamentals of Queueing Theory (4th ed.). New Jersey: John Wiley & Sons, Inc.

[47] Gupta, M. K., Hemachandra, N., & Venkateswaran, J. (2017). Optimal revenue management in two class pre-emptive delay dependent Markovian queues. Applied Mathematical Modelling, 45, 31–54.

[48] Gupta, M., & Mohanty, B. K. (2016). Attribute partitioning in multiple attribute decision making problems for a decision with a purpose - a fuzzy approach. Journal of Multi-Criteria Decision Analysis, 23(3–4), 160–170.

[49] Gupta, M., & Mohanty, B. K. (2017). Finding the numerical compensation in multiple criteria decision-making problems under fuzzy environment. International Journal of Systems Science, 48(6), 1301–1310.

[50] Gupta, S. M. (1995). Interrelationship between controlling arrival and service in queueing systems. Computers & Operations Research, 22(10), 1005–1014.

[51] Gupta, S. M. (1997). Machine interference problem with warm spares, server vacations and exhaustive service. Performance Evaluation, 29(3), 195–211.

[52] Gupta, U. C., & Srinivasa Rao, T. S. S. (1994). A recursive method to compute the steady state probabilities of the machine interference model: (M/G/1)/K. Computers & Operations Research, 21(6), 597–605.

[53] Gupta, U. C., & Srinivasa Rao, T. S. S. (1998). On the analysis of single server finite queue with state dependent arrival and service processes: M(n)/G (n)/1/K. OR Spectrum, 20(2), 83–89.

[54] Hadidi, N. (1974). Busy period of queues with state dependent arrival and service rates. Journal of Applied Probability, 11(4), 842–848.

[55] Haque, L., & Armstrong, M. J. (2007). A survey of the machine interference problem. European Journal of Operational Research, 179(2), 469–482.

[56] Hasanzadeh, H., Bashiri, M., & Amiri, A. (2018). A new approach to optimize a hub covering location problem with a queue estimation component using genetic programming. Soft Computing, 22(3), 949–961.

[57] Hassin, R., & Haviv, M. (2003). To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems. Boston, MA: Springer Science & Business Media.

[58] Hemachandra, N., & Narahari, Y. (2000). A mathematical programming approach to optimal Markovian switching of Poisson arrival streams to queueing systems. Queueing Systems, 36(4), 443–461.

[59] Henderson, W. (1972). Alternative approaches to the analysis of the M/G/1 and G/M/1 queues. Journal of the Operations Research Society of Japan, 15, 92–101.

[60] Hourani, M. (2004). Genetic Algorithm Application to Queuing Network and Gene-Clustering Problems. (Doctoral dissertation, Miami University).

[61] Hu, L., Zhao, B., Zhu, J., & Jiang, Y. (2019). Two time-varying and state-dependent fluid queuing models for traffic circulation systems. European Journal of Operational Research, 275(3), 997–1019.

[62] Huang, H.-I., Hsu, P.-C., & Ke, J.-C. (2011). Controlling arrival and service of a two-removable-server system using genetic algorithm. Expert Systems with Applications, 38(8), 10054–10059.

[63] Huang, H.-I., Wang, T.-Y., & Ke, J.-C. (2016). Random policy for an unreliable server system with delaying repair and setup time under Bernoulli vacation schedule. Journal of Testing & Evaluation, 44(3), 1400–1408.

[64] Jaggi, C. K., Mishra, B. K., & Panda, T. C. (2018). A fuzzy model for deteriorating items with allowable shortage and inspection under the trade credit. In Handbook of Research on Promoting Business Process Improvement Through Inventory Control Techniques (pp. 233–249). IGI Global.

[65] Jaggi, C. K., Pareek, S., Khanna, A., & Nidhi, N. (2016). Optimal replenishment policy for fuzzy inventory model with deteriorating items and allowable shortages under inflationary conditions. Yugoslav Journal of Operations Research, 26(4), 507–526.

[66] Jailaxmi, V., Arumuganathan, R., & Senthil Kumar, M. (2017). Performance analysis of an M/G/1 retrial queue with general retrial time, modified M-vacations and collision. Operational Research, 17(2), 649–667.

[67] Jain, M., & Bhagat, A. (2013). Transient analysis of retrial queues with double orbits and priority customers. In 8th International Conference on Queuing Theory and Network Applications (pp. 235–240). Taichung, Taiwan.

[68] Jain, M., & Bhagat, A. (2014). F-policy retrial queue with unreliable server: A computational approach. In International Conference on Emerging Trends in Global Management Practices – An Interdisciplinary Approach (pp. 741–759).

[69] Jain, M., & Bhagat, A. (2015a). Analysis of bulk retrial queue using maximum entropy principle. International Journal of Operational Research, 23(4), 477.

[70] Jain, M., & Bhagat, A. (2015b). Transient analysis of finite F-policy retrial queues with delayed repair and threshold recovery. National Academy Science Letters, 38(3), 257–261.

[71] Jain, M., & Gupta, R. (2013). Optimal replacement policy for a repairable system with multiple vacations and imperfect fault coverage. Computers & Industrial Engineering, 66(4), 710–719.

[72] Jain, M., & Meena, R. K. (2017). Markovian analysis of unreliable multi-components redundant fault tolerant system with working vacation and F-policy. Cogent Mathematics, 4(1), 1–17.

[73] Jain, M., & Sharma, G. (2002). M/M/M/K queue with additional servers and discouragement. International Journal of Engineering, 15(4), 349–354.

[74] Jain, M., & Singh, P. (2002). M/M/m queue with balking, reneging and additional servers. International Journal of Engineering, 15, 169–178.

[75] Jain, M., & Upadhyaya, S. (2009). Threshold N-policy for degraded machining system with multiple type spares and multiple vacations. Quality Technology & Quantitative Management, 6(2), 185–203.

[76] Jain, M., Bhagat, A., & Shekhar, C. (2015). Double orbit finite retrial queues with priority customers and service interruptions. Applied Mathematics and Computation, 253, 324–344.

[77] Jain, M., Kumari, S., Qureshi, R., & Shankaran, R. (2019). Markovian multi-server queue with reneging and provision of additional removable servers. In Performance Prediction and Analytics of Fuzzy, Reliability and Queuing Models (pp. 203–217). Springer Singapore.

[78] Jain, M., Rakhee, & Singh, M. (2004). Bilevel control of degraded machining system with warm standbys, setup and vacation. Applied Mathematical Modelling, 28(12), 1015–1026.

[79] Jain, M., Sharma, G. C., & Rani, V. (2014). M/M/R+r machining system with reneging, spares and interdependent controlled rates. International Journal of Mathematics in Operational Research, 6(6), 655–679.

[80] Jain, M., Sharma, G. C., & Sharma, R. (2012). Optimal control of (N, F) policy for unreliable server queue with multi-optional phase repair and start-up. International Journal of Mathematics in Operational Research, 4(2), 152-174.

[81] Jain, M., Sharma, G. C., & Shekhar, C. (2005). Processor-shared service systems with queue-dependent processors. Computers & Operations Research, 32(3), 629–645.

[82] Jain, M., Shekhar, C., & Meena, R. K. (2017). Admission control policy of maintenance for unreliable server machining system with working vacation. Arabian Journal for Science and Engineering, 42(7), 2993–3005.

[83] Jain, M., Shekhar, C., & Meena, R. K. (2019). Performance analysis and control F-policy for fault-tolerant system with working vacation. OPSEARCH, 56(2), 409–431.

[84] Jain, M., Shekhar, C., & Shukla, S. (2016). Queueing analysis of machine repair problem with controlled rates and working vacation under F-Policy. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 86(1), 21–31.

[85] Jana, R. K., & Sharma, D. K. (2010). Genetic algorithm-based fuzzy goal programming for class of chance-constrained programming problems. International Journal of Computer Mathematics, 87(4), 733–742.

[86] Jana, R. K., Sharma, D. K., & Chakraborty, B. (2016). A hybrid probabilistic fuzzy goal programming approach for agricultural decision-making. International Journal of Production Economics, 173, 134–141.

[87] Jang, J.-S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernetics, 23(3), 665–685.

[88] Jayaraman, R., & Matis, T. I. (2011). Finite population models-single station queues. In Wiley Encyclopedia of Operations Research and Management Science (pp. 1–6). Hoboken, NJ, USA: John Wiley & Sons, Inc.

[89] Kalayanaraman, R., Thillaigovindan, N., & Kannadasan, G. (2010). A single server fuzzy queue with unreliable server. International Journal of Computational Cognition, 8(1), 1–4.

[90] Kapur, P. K., & Sachdeva, N. (2016). Optimization problems for a modular software with faults of different severity. International Journal of Reliability, Quality and Safety Engineering, 23(3), 1640004.

[91] Kapur, P. K., Sachdeva, N., & Singh, O. (2017). Optimal profit for manufacturers in product remanufacturing diffusion dynamics. Journal of Industrial and Production Engineering, 34(8), 568–579.

[92] Karaesmen, F., & Gupta, S. (1997). Duality relations for queues with arrival and service control. Computers & Operations Research, 24(6), 529–538.

[93] Ke, J.-B., Ke, J.-C., & Lin, C. (2010). Cost optimization of an M/M/r queueing system with queue-dependent servers. In Proceedings of the 5th International Conference on Queueing Theory and Network Applications (pp. 82–86). New York, USA.

[94] Ke, J.-C., & Chang, F.-M. (2009). Modified vacation policy for M/G/1 retrial queue with balking and feedback. Computers & Industrial Engineering, 57(1), 433–443.

[95] Ke, J.-C., & Lin, C.-H. (2005). A Markov repairable system involving an imperfect service station with multiple vacations. Asia-Pacific Journal of Operational Research, 22(4), 555–582.

[96] Ke, J.-C., & Liu, T.-H. (2014). A repairable system with imperfect coverage and reboot. Applied Mathematics and Computation, 246, 148–158.

[97] Ke, J.-C., & Wang, K.-H. (1999). Cost analysis of the M/M/R machine repair problem with balking, reneging, and server breakdowns. Journal of the Operational Research Society, 50(3), 275–282.

[98] Ke, J.-C., & Wang, K.-H. (2007). Vacation policies for machine repair problem with two type spares. Applied Mathematical Modelling, 31(5), 880–894.

[99] Ke, J.-C., Chang, F.-M., & Chang, C.-J. (2010). Controlling arrivals for a Markovian queueing system with a second optional service. International Journal of Industrial Engineering: Theory, Applications and Practice, 17(1), 48–57.

[100] Ke, J.-C., Chang, F.-M., & Liu, T.-H. (2019). M/M/c balking retrial queue with vacation. Quality Technology & Quantitative Management, 16(1), 54–66.

[101] Ke, J.-C., Huang, H.-I., & Lin, C.-H. (2006). Parametric programming approach for batch arrival queues with vacation policies and fuzzy parameters. Applied Mathematics and Computation, 180(1), 217–232.

[102] Ke, J.-C., Huang, H.-I., & Lin, C.-H. (2007). On retrial queueing model with fuzzy parameters. Physica A: Statistical Mechanics and Its Applications, 374(1), 272–280.

[103] Ke, J.-C., Huang, H.-I., & Lin, C.-H. (2008). A redundant repairable system with imperfect coverage and fuzzy parameters. Applied Mathematical Modelling, 32(12), 2839–2850.

[104] Ke, J.-C., Liu, T.-H., & Yang, D.-Y. (2016). Machine repairing systems with standby switching failure. Computers & Industrial Engineering, 99, 223–228.

[105] Ke, J.-C., Yang, D., Sheu, S., & Kuo, C. (2013). Availability of a repairable retrial system with warm standby components. International Journal of Computer Mathematics, 90(11), 2279–2297.

[106] Kim, J., & Kim, B. (2007). The processor-sharing queue with bulk arrivals and phase-type services. Performance Evaluation, 64(4), 277–297.

[107] Kumar, K., & Jain, M. (2013). Threshold F-policy and N-policy for multi-component machining system with warm standbys. Journal of Industrial Engineering International, 9(1), 9–28.

[108] Kumar, K., Jain, M., & Shekhar, C. (2019). Machine repair system with F-policy, two unreliable servers, and warm standbys. Journal of Testing and Evaluation, 47(1), 361–383.

[109] Kumar, R., & Sharma, S. K. (2012). M/M/1/N queuing system with retention of reneged customers. American Journal of Operational Research, 2(1), 1–5.

[110] Kumar, R., & Sharma, S. K. (2014). A multi-server Markovian feedback queue with balking reneging and retention of reneged customers. Advanced Modeling and Optimization, 16(2), 395–406.

[111] Kumara, E. R., & Dharsana, S. (2015). Analysis of M/M/1 queueing system with state dependent arrival and detainment of retracted customers. Malaya Journal of Matematik, (Special Issue 1), 89–98.

[112] Kuo, C.-C., Wang, K.-H., & Pearn, W. (2011). The interrelationship between N-policy M/G/1/K and F-policy G/M/1/K queues with startup time. Quality Technology & Quantitative Management, 8(3), 237–251.

[113] Laxmi, P. V., & Kassahun, T. W. (2017). A multi-server infinite capacity Markovian feedback queue with balking, reneing and retension of reneged customers. International Journal of Mathematical Archive, 8(2), 53–59.

[114] Lee, C. (2011). On moment stability properties for a class of state-dependent stochastic networks. Journal of the Korean Statistical Society, 40(3), 325–336.

[115] Lin, C., & Ke, J. (2011). On the multi-server retrial queue with geometric loss and feedback: computational algorithm and parameter optimization. International Journal of Computer Mathematics, 88(5), 1083–1101.

[116] Liou, C.-D. (2014). Optimization analysis of the machine repair problem with multiple vacations and working breakdowns. Journal of Industrial and Management Optimization, 11(1), 83–104.

[117] Liu, Z., & Yu, S. (2016). The M/M/C queueing system in a random environment. Journal of Mathematical Analysis and Applications, 436(1), 556–567.

[118] Majumdar, R., Kapur, P. K., Khatri, S. K., & Shrivastava, A. K. (2018). Evolutionary algorithm based faults optimization of multi-modular software. In S. Satapathy, V. Bhateja, & S. Das (Eds.), Smart Computing and Informatics (pp. 281–291). Springer, Singapore.

[119] Massey, W. A. (2002). The analysis of queues with time-varying rates for telecommunication models. Telecommunication Systems, 21(2–4), 173–204.

[120] Medhi, J. (2003). Stochastic Models in Queueing Theory (2nd ed.). Elsevier.

[121] Mishra, S. S., Sharma, D. K., Yadav, S. K., & Rawat, S. (2017). Computational approach to fuzzified profit optimisation of inventory flow in supply chain with deteriorating items. International Journal of Operational Research, 30(1), 83–98.

[122] Mitchell, M. (1996). An introduction to genetic algorithms. Computers & Mathematics with Applications, 32(6), 133.

[123] Mitchell, M. (1998). An Introduction to Genetic Algorithms (5th ed.). MA, USA: MIT Press Cambridge.

[124] Mohanty, B. K., & Passi, K. (2010). Agent based e-commerce systems that react to buyers' feedbacks – A fuzzy approach. International Journal of Approximate Reasoning, 51(8), 948–963.

[125] Mokaddis, G., & Zaki, S. (1983). The problem of the queuing system M/M/1 with additional servers for a longer queue. Indian Journal of Pure & Applied Mathematics, 14(3), 345–354.

[126] Moreno, P. (2004). An M/G/1 retrial queue with recurrent customers and general retrial times. Applied Mathematics and Computation, 159(3), 651–666.

[127] Mueen, Z., Ramli, R., & Zaibidi, N. (2017). Parametric nonlinear programming approach with fuzzy queues using hexagonal membership functions. Journal of Computational and Theoretical Nanoscience, 14(10), 4979–4985.

[128] Nobel, R.D. (2016). Retrial queueing models in discrete time: a short survey of some late arrival models. Annals of Operations Research, 247(1), 37–63.

[129] Nobel, R.D., & Tijms, H. C. (1999). Optimal control for an MX/G/1 queue with two service modes. European Journal of Operational Research, 113(3), 610–619.

[130] Nobel, R.D., & Tijms, H. C. (2000). Optimal control of a queueing system with heterogeneous servers and setup costs. IEEE Transactions on Automatic Control, 45(4), 780–784.

[131] Parthasarathy, P. R., & Sudhesh, R. (2007). Time-dependent analysis of a single-server retrial queue with state-dependent rates. Operations Research Letters, 35(5), 601–611.

[132] Phung-Duc, T. (2017). Single server retrial queues with setup time. Journal of Industrial & Management Optimization, 13(3), 1329–1345.

[133] Phung-Duc, T., & Kawanishi, K. (2019). Multiserver retrial queue with setup time and its application to data centers. Journal of Industrial & Management Optimization, 15(1), 15–35.

[134] Phung-Duc, T., Dohi, T., Ano, K., & Kasahara, S. (2017). Retrial queueing models: A survey on theory and applications. Stochastic Operations Research in Business and Industry, 1–26.

[135] Raman, D., Nagalingam, S. V, & Gurd, B. W. (2009). A genetic algorithm and queuing theory based methodology for facilities layout problem. International Journal of Production Research, 47(20), 5611–5635.

[136] Rashidi, S., & Sharifian, S. (2017). A hybrid heuristic queue based algorithm for task assignment in mobile cloud. Future Generation Computer Systems, 68, 331–345.

[137] Rodrigues, J., Prado, S. M., Balakrishnan, N., & Louzada, F. (2016). Flexible M/G/1 queueing system with state dependent service rate. Operations Research Letters, 44(3), 383–389.

[138] Roychowdhury, S., & Pedrycz, W. (2001). A survey of defuzzification strategies. International Journal of Intelligent Systems, 16(6), 679–695.

[139] Sahoo, L. (2017). An approach for solving fuzzy matrix games using signed distance method. Journal of Information and Computing Science, 12(1), 73–80.

[140] Sharma, A. K., Jaggi, C. K., & Tiwari, S. (2017). Optimal replenishment policy for price dependent demand in different financial scenario under fuzzy environment. International Journal of Inventory Research, 4(2/3), 103.

[141] Shawky, A. I. (1997). The single-server machine interference model with balking, reneging and an additional server for longer queues. Microelectronics Reliability, 37(2), 355–357.

[142] Shekhar, C., Jain, M., Raina, A. A., & Iqbal, J. (2017). Optimal (N, F) policy for queue-dependent and time-sharing machining redundant system. International Journal of Quality & Reliability Management, 34(6), 798–816.

[143] Sherman, N. P., & Kharoufeh, J. P. (2006). An M/M/1 retrial queue with unreliable server. Operations Research Letters, 34(6), 697–705.

[144] Singh, C. J., Kaur, S., & Jain, M. (2017). Waiting time of bulk arrival unreliable queue with balking and Bernoulli feedback using maximum entropy principle. Journal of Statistical Theory and Practice, 11(1), 41–62.

[145] Sivazlian, B., & Wang, K. H. (1989). Economic analysis of the M/M/R machine repair problem with warm standbys. Microelectronics Reliability, 29(1), 25–35.

[146] Som, B. K., & Kumar, R. (2018). A heterogeneous queuing system with reverse balking and reneging. Journal of Industrial and Production Engineering, 35(1), 1–5.

[147] Tandra, R., Hemachandra, N., & Manjunath, D. (2004). DiffServ node with join minimum cost queue policy and multiclass traffic. Performance Evaluation, 55(1–2), 69–91.

[148] Tao, L., Zhang, L., & Gao, S. (2014). M/M/1 retrial queue with working vacation interruption and feedback under N-policy. Journal of Applied Mathematics, 2014, 1–9.

[149] Telek, M., & Van Houdt, B. (2018). Response time distribution of a class of limited processor sharing queues. ACM SIGMETRICS Performance Evaluation Review, 45(2), 143–155.

[150] Van Do, T. & Chakka, R. (2010). An efficient method to compute the rate matrix for retrial queues with large number of servers. Applied Mathematics Letters, 23(5), 638–643.

[151] Van Leekwijck, W., & Kerre, E. E. (1999). Defuzzification: criteria and classification. Fuzzy Sets and Systems, 108(2), 159–178.

[152] Verma, A. K., Srividya, A., & Gaonkar, R. S. P. (2009). Fuzzy-Reliability Engineering: Concepts and Applications. New Delhi, India: Narosa Publishing.

[153] Verma, A.K., Srividya, A., & Deka, B. C. (2005). Composite system reliability assessment using fuzzy linear programming. Electric Power Systems Research, 73(2), 143–149.

[154] Vijaya Laxmi, P., Goswami, V., & Jyothsna, K. (2013). Optimization of balking and reneging queue with vacation interruption under N-policy. Journal of Optimization, 2013, 1–9.

[155] Wang, K.-H. (1995). An approach to cost analysis of the machine repair problem with two types of spares and service rates. Microelectronics Reliability, 35(11), 1433–1436.

[156] Wang, K.-H., & Chang, Y.-C. (2002). Cost analysis of a finite M/M/R queueing system with balking, reneging, and server breakdowns. Mathematical Methods of Operations Research, 56(2), 169–180.

[157] Wang, K.-H., & Yang, D. Y. (2009). Controlling arrivals for a queueing system with an unreliable server: Newton-Quasi method. Applied Mathematics and Computation, 213(1), 92–101.

[158] Wang, K.-H., Ke, J.-B., & Ke, J.-C. (2007). Profit analysis of the M/M/R machine repair problem with balking, reneging, and standby switching failures. Computers & Operations Research, 34(3), 835–847.

[159] Wang, K.-H., Kuo, C. C., & Pearn, W. L. (2007). Optimal control of an M/G/1/K queueing system with combined F policy and startup time. Journal of Optimization Theory and Applications, 135(2), 285–299.

[160] Wang, K.-H., Kuo, C.-C., & Pearn, W. L. (2008). A recursive method for the F-policy G/M/1/K queueing system with an exponential startup time. Applied Mathematical Modelling, 32(6), 958–970.

[161] Wang, K.-H., Liou, C.-D., & Wang, Y.-L. (2014). Profit optimisation of the multiple-vacation machine repair problem using particle swarm optimisation. International Journal of Systems Science, 45(8), 1769–1780.

[162] Wang, K.-H., Liu, Y.-C., & Pearn, W. L. (2005). Cost benefit analysis of series systems with warm standby components and general repair time. Mathematical Methods of Operations Research, 61(2), 329–343.

[163] Wang, K.-H., Yen, T.-C., & Chen, J.-Y. (2018). Optimization analysis of retrial machine repair problem with server breakdown and threshold recovery policy. Journal of Testing and Evaluation, 46(6), 2630–2640.

[164] Wüchner, P., Sztrik, J., & de Meer, H. (2009). Finite-source retrial queue with search for balking and impatient customers from the orbit. Computer Networks, 53(8), 1264–1273.

[165] Yager, R. R. (1981). A procedure for ordering fuzzy subsets of the unit interval. Information Sciences, 24(2), 143–161.

[166] Yang, D. Y., & Yang, N. C. (2018). Performance and cost analysis of a finite capacity queue with two heterogeneous servers under F-policy. International Journal of Services Operations and Informatics, 9(2), 101–115.

[167] Yang, D.-Y., & Chang, P.-K. (2015). A parametric programming solution to the F-policy queue with fuzzy parameters. International Journal of Systems Science, 46(4), 590–598.

[168] Yang, D.-Y., & Chang, Y.-D. (2018). Sensitivity analysis of the machine repair problem with general repeated attempts. International Journal of Computer Mathematics, 95(9), 1761–1774.

[169] Yang, D.-Y., Chang, F.-M., & Ke, J.-C. (2016). On an unreliable retrial queue with general repeated attempts and J optional vacations. Applied Mathematical Modelling, 40(4), 3275–3288.

[170] Yang, D.-Y., Wang, K.-H., & Wu, C.-H. (2010). Optimization and sensitivity analysis of controlling arrivals in the queueing system with single working vacation. Journal of Computational and Applied Mathematics, 234(2), 545–556.

[171] Yao, J.-S., & Wu, K. (2000). Ranking fuzzy numbers based on decomposition principle and signed distance. Fuzzy Sets and Systems, 116(2), 275–288.

[172] Yeh, C., Lee, Y.-T., Chang, C.-J., & Chang, F.-M. (2017). Analysis of a two-phase queue system with <p, F>- policy. Quality Technology & Quantitative Management, 14(2), 178–194.

[173] Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1(3–28), 61–72.

[174] Zimmermann, H.-J. (1996). Fuzzy Set Theory—and Its Applications (3rd ed.). Springer, Dordrecht

[175] Zirem, D., Boualem, M., Adel-Aissanou, K., & Aïssani, D. (2018). Analysis of a single server batch arrival unreliable queue with balking and general retrial time. Quality Technology & Quantitative Management, 1–24.