PERFORMANCE ANALYSIS OF RETRIAL QUEUEING SYSTEMS

Ph.D. THESIS

by

AMITA BHAGAT

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)
JULY, 2014

PERFORMANCE ANALYSIS OF RETRIAL QUEUEING SYSTEMS

A THESIS

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

DOCTOR OF PHILOSOPHY
in

MATHEMATICS

by

AMITA BHAGAT

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)
JULY, 2014

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 OF RETRIAL QUEUEING SYSTEMS**' 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 July, 2009 to July, 2014 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 Institute.

(AMITA BHAGAT)

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

(Madhu Jain)
Supervisor

Dated:

ABSTRACT

The queueing models with reattempts or returning customers are realistic and robust in formulating many real world congestion situations. Retrial queues which deal with repeated attempts are characterized by the phenomenon that whenever, a customer finds the server busy or blocked, then he is obliged to join another queue or a virtual pool of blocked customers called ‘orbit’. The applications of retrial queues can also be realized in other industrial scenarios including manufacturing and production processes, telecommunication systems, transportation and service systems, etc. The customer deprived of service may make reattempts in order to get served as visible in telephone systems. A telephone subscriber, who finds a busy route usually repeats the call until the connection is made, such subscribers form retrial queues.

Performance modeling plays a vital role in the design, development and analysis of a variety of real time practical systems. Queueing models are often used for the performance and reliability modeling of these systems where retrial queues are often built up. The queueing analysis based on Markovian or non-Markovian processes provides valuable insight to the decision makers for the improvement of retrial queueing systems in different frameworks. Our study on retrial queues is basically motivated by their abundant applications with the advancement of technology in the area of communication and computer networks.

It is significant to study how the phenomenon of making reattempts by the customer affect the performance of various queueing systems. In the present thesis work, we investigate retrial queueing models in different frameworks applicable to various real life congestion scenarios. The noble features of the present investigation are the modeling as well as the analysis of retrial queueing systems by incorporating several practical features like vacation, balking, reneging, bulk, priority, unreliable server, etc. Using different techniques, several performance measures namely queue length, waiting time, server utilization, long run probabilities, etc. have been obtained for the retrial models under consideration. Some cost optimization problems are also framed so as to obtain optimal parameters and optimum cost incurred on the system. The whole thesis devoted to retrial queueing models is structured in ten chapters. Some retrial models have been developed including various features like phase service, phase repair, priority, vacation, control policy, discouragement, etc. and analyzed using suitable techniques. The

sensitivity analysis has been carried out to examine the validity of performance indices evaluated using analytical methods. At the end of the thesis, conclusions and future scope of the present research work has been added to highlight the contributions and significance of present doctoral work. The relevant references have been listed in alphabetical order in the end of this work. The brief outlines of the thesis work are as follows:

Chapter 1 is devoted to an overview of the conceptual aspects along with motivational factors to study the retrial queueing systems in different frameworks. The related literature review has been briefly presented by classifying the retrial queues based on modeling and methodological aspects. The contents of the thesis and concluding remarks are also given.

Chapter 2 is concerned with the analysis of unreliable retrial queue with impatient customers. Using supplementary variable technique (SVT) and probability generating function (PGF); the queue size distributions of the orbit and system size and other performance indices have been obtained. Further, the maximum entropy principle (MEP) has been used to determine the approximate results for the steady state probabilities of the system states, queue length and expected waiting time.

Chapter 3 deals with a batch arrival general retrial queue with multioptional services, vacation and impatient customers. The study extends the work presented in *chapter 2* by incorporating the features of phase repair and Bernoulli vacation schedule. To obtain queue size distribution and various performance measures, SVT and PGF have been used. The neuro fuzzy approach has also been used to provide the computational results for some performance measures.

Bulk arrival M/G/1 retrial queue with impatient customers and modified vacation policy has been analysed in *chapter 4*. The service is provided in k compulsory phases and the repair of broken down server is performed in *d* compulsory phases. As soon as the orbit becomes empty, the server goes for vacation and takes at most *J* vacations until at least one customer is noticed in the system. Using SVT and PGF approach, the queue size distributions of the number of customers in the orbit and system have been obtained. The maximum entropy principle is also employed to obtain the approximate results for the queue length and expected waiting time.

The performance analysis of bulk arrival retrial queue with priority customers, unreliable server, balking, multi essential service, multi phase repair has been presented in *chapter 5*. Using queue theoretic approach based on SVT and PGF, the queue size distributions of priority and non-priority customers and other performance indices have been established.

In *chapter 6,* the bulk arrival retrial queue with negative customers and multi-services subject to server breakdowns has been considered. The system allows the arrival of two types of customers; positive customers and negative customers in the system. Moreover, the customers may renege from the system out of impatience. The server has the provision to initiate the service when there are N customers accumulated in the system. The SVT has been used to analyze the model under consideration.

Chapter 7 is concerned with the performance prediction of a batch arrival retrial queue with multioptional services and phase repair under Bernoulli vacation schedule. The customers arrive in batches and are admitted to join the system following Bernoulli admission control policy. By applying the embedded Markov chain method, the ergodicity condition for the stability and various queueing measures are established.

Chapter 8 deals with two finite capacity retrial queueing models with threshold recovery. The first model deals with Markovian retrial queues with unreliable server and geometric arrivals. The second model is concerned with the finite capacity retrial queueing model with F-policy. The numerical approach based on the Runge Kutta method of fourth order has been employed to study the transient behavior of both the models.

The unreliable server retrial queue with the provision of additional temporary server in the context of application in web faction has been investigated in *chapter 9*. The primary server can serve a maximum of 'K' customers in the system. The additional server is turned on if the number of customers exceeds this limit. The matrix geometric approach is employed to obtain the steady state probabilities of the system states and other performance measures.

In *chapter 10*, we consider the arrival of two types of customers known as priority and non priority customers which have the facility of different waiting spaces i.e. orbits. The double orbit finite capacity retrial queue with unreliable server has been taken into consideration from modeling point of view. Both transient as well as steady state analysis has been done using matrix method. The numerical simulation has been carried out by taking an illustration with an application to cellular radio network.

The modeling and analysis of retrial queueing systems in different frameworks consistent with various real life scenarios have been presented in the present research work. The models developed can be successfully used in abundant congestion problems ranging from day-to-day to telecommunication networks. Keeping in mind the significance of retrial queues a variety of problems have been explored using different methodologies. A variety of prominent features namely additional server, double orbits, finite capacity, phase service, vacation, phase repair etc. have been incorporated to frame versatile retrial queueing models applicable to different real life congestion scenarios. Different cost functions have been structured corresponding to different retrial models and optimal parameters have also been obtained to determine the optimal cost of the concerned queueing systems. The numerical simulation has been done to examine the computational tractability of analytical results using various classical queueing methodologies. It is hoped that the performance and analysis of retrial queueing systems presented in this work may be helpful in improving the grade of the service of many existing systems and may provide valuable insight to the system designers, developers and practitioners to frame more optimal and efficient models which will be more suitable in various real life congestion situations.

ACKNOWLEDGEMENTS

Pursuing Ph.D. is just like climbing a skyscraper peak step by step accompanied with hardships, hope and despair, encouragements and obstacles, frustration and success, and help and cooperation of the people around you. A strong power has always been supporting me during the entire course of the study and I am very much confident that this has been the force of Almighty, "The God", who never left me trapped in difficulties. I express my gratitude to the Almighty for this divine favor.

Through this thanks giving note, at the outset, I express my profound sense of gratitude to my advisor and supervisor, Dr. Madhu Jain, Department of Mathematics, IIT Roorkee. I shall always remain indebted to her for appropriate comments rendered by her at many a critical juncture during the entire course of study. Her physical insight, moral support, guidance and encouragement have been the impetus for me to learn a lot in every phase of my Ph.D programme. At this stage of my career when I paused to look back to my research period, I find that at every stage her perpetual encouragement, her enthusiasm and above all morale boosting inspiration served a vital source to bring the present work to conclusion. Further, her bold initiative and uncompromising gesture made a highly challenging path feasible.

I am also indebted to Prof. R.C. Mittal, Head, Department of Mathematics, IIT Roorkee, for his valuable direction, encouragement, support and for providing me all the necessary facilities for carrying out this work till its completion. I also thank Prof. P.N. Agarwal, DRC Chairman, all the members of Student Research Committee, faculty members and staff members of Department of Mathematics for their valuable support and encouragement at all levels.

I am highly indebted to Prof. G.C. Sharma, Ex. Pro. V.C., 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 to the end of my research work.

I owe my heartfelt thanks and gratitude to Puneet Rana, who has always been a constant source of inspiration and help to me. My sincere thanks are also due to Dr. Richa Sharma, Dr. Shweta Upadhyay, Dr. Kapil Kumar, Dr. K.G. Arora and Dr. Sudhir Pundir for their valuable suggestions and discussions during my research work.

It's my fortune to gratefully acknowledge the support of some special individuals who were always there beside me during the happy and hard moments to push me and motivate me; Stuti, Shikha, Jyotima, Manjula, Khushboo, Harman, Dipasha, Dr. Anupam Yadav, Dr. Harish Garg for their cooperation and support and for making my stay at IIT Roorkee a memorable one. My special thanks go to my buddies Deepika Saini, Dr. Neha Yadav, Rajeev Saxena, Akarsh, Rakesh, and Kamlesh for their wholehearted help and assistance.

I wish to put on records that whatever I have achieved in my life is due to the blessings of my mother Smt. Asha Bhagat and father Shri Sewant Kumar Bhagat. I cannot restrain myself in offering my heartfelt thanks to my elder sister Kavita Wadhwa without whose support this research work would have been incomplete. I am forever grateful to my younger sister Shivani Bhagat, brother Vishal Bhagat and brother-in-law Ravi Wadhwa, without their support this work would not have been possible. They supported me unconditionally in many ways to accomplish this arduous task.

I am also grateful to my grandparents, all uncles and aunts, cousins who drove me morally all the way through this journey of research in numerous ways. It is not possible to list the name of everyone, but I owe my gratitude to each one of them.

I shall be failing in my duty if I do not express my deep sense of gratitude to Ministry of Human Resource and Development for their financial assistance in the form of JRF and SRF. I gratefully acknowledge the financial assistance from CSIR, New Delhi, for their indispensable help in the form of travel grant to participate in International Conference abroad to Taichung, Taiwan.

Finally, I would like to thank everyone who has knowingly and unknowingly helped me in the successful realization of Thesis, as well as expressing my apology that I could not mention personally one by one. Above all, I owe it all to Almighty God for granting me the wisdom, health and strength to undertake this research task and enabling me to its completion.

AMITA BHAGAT

LIST OF PUBLICATIONS

INTERNATIONAL JOURNALS/CONFERENCE PROCEEDINGS

1. **Amita Bhagat** and Madhu Jain, (2013): Unreliable $M^X/G/1$ Retrial Queue with Multi-Optional Services and Impatient Customers, International Journal of Operational Research, Vol. 17, No. 2, pp. 248-273.

2. Madhu Jain and **Amita Bhagat** (2012): Finite Population Retrial Queueing Model with Threshold Recovery, Geometric Arrivals and Impatient Customers, Journal of Information and Operations Management, Vol. 3, No. 1, pp. 162-165.

3. Madhu Jain and **Amita Bhagat** (2014): Unreliable Bulk Retrial Queues with Delayed Repairs and Modified Vacation Policy, Journal of Industrial Engineering International, Springer, Vol. 10, pp. 1-19.

4. Madhu Jain and **Amita Bhagat** (2014): Analysis of Bulk Retrial Queue Using Maximum Entropy Principle, International Journal of Operational Research (In press).

5. Madhu Jain and **Amita Bhagat** (2014): Transient Analysis of Finite F-policy Retrial Queues with Delayed Repair and Threshold Recovery, National Academy of Science Letters, Springer (In Press).

6. Madhu Jain and **Amita Bhagat** (2012): Retrial Queue with Threshold Recovery, Geometric Arrivals and Finite Capacity, Advances in Intelligent and Soft Computing, Vol. 131, pp. 1039-1048, Springer Proceedings of the International conference SocPros-2011, held at IIT Roorkee, India from 20-22 Dec, 2011.

7. Madhu Jain and **Amita Bhagat** (2013): Transient Analysis of Retrial Queues with Double Orbits and Priority Customers, Electronic Proceedings of 8th International Conference on Queueing theory and Network Applications (QTNA-2013) held at Taichung, Taiwan from 30 Jul-2 Aug, 2013, pp. 235-240.

8. Madhu Jain and **Amita Bhagat** (2014): F-policy Finite Retrial Queue with Unreliable Server and Threshold Recovery: A Computational Approach, Electronic proceedings of The International Conference on Emerging Trends in Global Management Practices-An Interdisciplinary Approach, held at Symbiosis Centre for Management Studies, Noida, March 6-7, 2014, pp. 241-260.

9. Madhu Jain and **Amita Bhagat** (2011): Maximum Entropy Analysis of Queue Length of Unreliable $M^x/G/1$ Retrial Queue with Impatient Customers, Electronic proceedings of International Conference AMOC-2011, held at Roorkee, India from 5-7 Dec, 2011, pp. 675-683.

RESEARCH PAPERS REVISED/COMMUNICATED FOR PUBLICATION

1. Madhu Jain, **Amita Bhagat** and Chandra Shekhar, "Double Orbit Finite Retrial Queues with Priority and Unreliable Server", *Applied Mathematics and Computation* (Revised for publication).
2. Madhu Jain and **Amita Bhagat**, "Performance Modeling and Neuro Fuzzy approach for $M^x/G/1$ Retrial Vacation Queue", *Quantity technology and Quantitative Management* (QTQM) (Revised for publication).
3. Madhu Jain and **Amita Bhagat**, "Embedded Markov Chain Approach to Retrial Queue with Vacation, Phase Repair and Multioptional Services", *OPSEARCH* (Revised for publication).
4. Madhu Jain and **Amita Bhagat**, "N-policy for $M^x/G/1$ Unreliable Retrial G-queue with Preemptive Resume and Multi-services", Communicated to *Journal of Egyptian Mathematical Society, Elsevier*.
5. Madhu Jain, **Amita Bhagat** and Richa Sharma, "Matrix Geometric Approach for Unreliable Retrial Queue with Additional Temporary Server and Constant Capacity", Communicated to *Sadhana- Academy Proceedings in Engineering Science (under review).*
6. Madhu Jain and **Amita Bhagat**, "On a Batch Arrival Non-preemptive Priority Retrial Queue with Multiple Phase Service and Breakdowns", Communicated to *Applications and Applied Mathematics: An International Journal (AAM).*

LIST OF CONFERENCES PARTICIPATED

1. Participated and presented a paper entitled "Transient Analysis of Retrial Queues with Double Orbits and Priority Customers" at **8th International Conference on Queueing theory and Network Applications (QTNA-2013)** held at Taichung, Taiwan from 30 Jul-2 Aug, 2013.
2. Presented a paper entitled "Finite Population Retrial Queueing Model with Threshold Recovery, Geometric Arrivals and Impatient Customers" at **International Symposium on recent Advances in IT and IT Management (RAITM'12)** held at Dehradun from 28-29 Feb 2012.
3. Participated and presented a paper entitled "Maximum Entropy Analysis of Queue Length of $M^x/G/1$ Retrial Queue with Impatient Customers" at **International Conference on Advances in Modeling, Optimization and Computing (AMOC-2011)**, held at Roorkee, India from 5-7 Dec, 2011.
4. Presented a paper entitled "Retrial Queue with Threshold Recovery, Geometric Arrivals and Finite Capacity" at the **International conference SocPros-2011**, held at IIT Roorkee, India from 20-22 Dec, 2011.
5. Presented a paper entitled "Maximum Entropy Approach for the Analysis of $M^x/G/1$ Retrial Queue with Impatient Customers" at **Advanced Workshop and Tutorial on Operations Research (AWTOR-2012)**, held at IIM Indore, August 22-25, 2012.
6. Participated in **International Congress of Mathematicians 2010 (ICM-2010)**, held at Hyderabad, India from 19-27 August, 2010.

TABLE OF CONTENTS

ACRONYMS

R.V.	Random variable
pdf	Probability density function
cdf	Cumulative distribution function
PGF	Probability generating function
i.i.d.	Independent and identically distributed
EMC	Embedded Markov chain
MEP	Maximum entropy principle
SVT	Supplementary variable technique
MGM	Matrix geometric method
LST	Laplace–Stieltjes transform
FES	First essential service
SOS	Second optional service

LIST OF NOTATIONS

Arrival Process

λ	Mean arrival rate
X	Random variable denoting the batch size
c_m	Pr{X=m}, (m ≥1)
$C(z)$	PGF of batch size X

Retrial Process

$A(.)$	cdf of retrial time
$\tilde{A}(s)$	LST of $A(.)$
$a(.)$	pdf of retrial time
$\tilde{a}(s)$	LST of $a(.)$
$D_n(t)$	Probability that the server is idle during the retrial of the customers at time t when there are n customers in the system
D_n	$\lim_{t\to\infty} D_n(t)$, n ≥ 0
$D(w,z)$	PGF of retrial time

Service Process

$\varpi(t)$	R.V. corresponding to elapsed service/ vacation time
μ_i	Mean service rate during i^{th} phase of service
$B_i(.)$	cdf for service time during i^{th} phase of service
$\tilde{B}_i(s)$	LST of $B_i(.)$
$b_i(.)$	pdf of service time during i^{th} phase of service
$\tilde{b}_i(s)$	LST of $b_i(.)$
$b_i^{(r)}$	r^{th} moment of i^{th} phase service time
$P_{i,n}(t,x)$	Probability that there are n customers in the system at time t and the server is busy in providing i^{th} phase of service to the customer and the customer is being served with elapsed service time lying between x and x+dx
$P_{i,n}(x)$	$\lim_{t\to\infty} P_{i,n}(t,x)$, n ≥ 1

$P_i(x,z)$	PGF of i^{th} phase service time
$\rho_i = \frac{\lambda E[X]}{\mu_i}$	Traffic intensity of the server

Repair Process

α_i	Mean breakdown rate of the server broken down during i^{th} phase of service
$\sigma(t)$	R.V. corresponding to elapsed repair time/set up time at time t
β_i	Mean repair rate of the server broken down during i^{th} phase service
$\beta_{i,j}$	Mean repair rate of j^{th} phase repair of the server when broken down during i^{th} phase of service
$G_i(.)$	cdf for repair time during i^{th} phase of service
$G_{i,j}(.)$	cdf for repair time during i^{th} phase of service and at the j^{th} phase repair
$\tilde{G}_i(s)$	LST of $G_i(.)$
$\tilde{G}_{i,j}(s)$	LST of $G_{i,j}(.)$
$g_i(.)$	pdf of repair time of the server broken down during i^{th} phase service
$g_{i,j}(.)$	pdf of repair time of j^{th} phase of repair of the server broken down during i^{th} phase of service
$\tilde{g}_i(s)$	LST of $g_i(.)$
$\tilde{g}_{i,j}(s)$	LST of $g_{i,j}(.)$
$R_{i,n}(t,x,y)$	Joint probability that at time and there are n customers in the system, the server is under repair while fails during the i^{th} phase service with elapsed service time x and elapsed repair time lying between y and y+dy
$R_{i,j,n}(t,x,y)$	Joint probability that at time t there are n customers in the system, the server is under j^{th} phase of repair while fails during the i^{th} phase service with elapsed service time x and elapsed repair time

	lying between y and y+dy
$R_{i,n}(x,y)$	$\lim_{t\to\infty} R_{i,n}(t,x,y)$
$R_{i,j,n}(x,y)$	$\lim_{t\to\infty} R_{i,j,n}(t,x,y)$
$R_i(x,y,z)$	PGF of repair time for the server broken down during i^{th} service
$R_{i,j}(x,y,z)$	PGF of j^{th} phase of repair time for the server broken down during i^{th} phase of service
$\tilde{g}_i^{(r)}$	r^{th} moment of the repair time for the server broken during i^{th} phase of service
$\tilde{g}_{i,j}^{(r)}$	r^{th} moment of j^{th} phase repair time for the server broken during i^{th} phase of service

Set up Process

ξ_i	Mean set up rate before repair for the server broken down during i^{th} phase of service
$N_i(.)$	cdf for set up time of the server broken during i^{th} phase of service
$\tilde{N}_i(s)$	LST of $N_i(.)$
$\eta_i(.)$	pdf of set up time of the server broken during i^{th} phase of service
$\tilde{\eta}_i(s)$	LST of $\eta_i(.)$
$S_{i,n}(t,x,y)$	Joint probability that at time t there are n customers in the system and the broken down server during i^{th} service is under setup before repair with elapsed service time x and elapsed setup time lying between y and y+dy
$S_{i,n}(x,y)$	$\lim_{t\to\infty} S_{i,n}(t,x,y)$, $n \geq 1$
$S_i(x,y,z)$	PGF of setup process before repair for the server broken down during i^{th} service
$\tilde{\eta}_i^{(r)}$	r^{th} moment of set up time before the repair process for the server broken during i^{th} phase of service

Vacation Process

ψ	Mean vacation rate of the server
$W(.)$	cdf for vacation time of the server
$\tilde{W}(s)$	LST of $W(.)$
$w(.)$	pdf of vacation time of the server
$\tilde{w}(s)$	LST of $w(.)$
$\tilde{w}^{(r)}$	r^{th} moment of vacation time
$V_n(t,x)$	Probability that at time t there are n customers in the system at and the server is on vacation with elapsed vacation time lying between x and x+dx
$V_n(x)$	$\lim_{t\to\infty} V_n(t,x)$, $n \geq 1$
$V(x,z)$	PGF of vacation time

LIST OF FIGURES

Chapter 5

Chapter 7

Chapter 8

Chapter 9

Chapter 10

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

Queues are essential phenomenon of real life congestion situations and are visible everywhere from reservation counters to admission counters, from supermarkets to doctor's clinic, etc. Queueing theory deals with the waiting line models to predict the behavior of the queueing systems which provide service for the randomly arising demands of the customers. The earliest studied queueing problem was that of telephone traffic congestion. There are many other notable applications of queueing theory, most of which have been well documented in the literature on the probability theory, stochastic processes, operations research, management science, and industrial engineering, etc. Some common examples of queueing scenarios are traffic flow (vehicles, aircraft, people, communications), scheduling (patients in hospitals, jobs on machines, programs on a computer), and service facility (banks, post offices, amusement parks, fast-food restaurants), etc.

In many queueing situations, the customer instead of waiting in the queue may prefer to do some other work and would like to try again for the service with a hope to find the server free on his next attempt. Such situations give rise to special queues known as retrial queues. Retrial queue is a special type of queueing situation, which is characterized by the phenomenon of reattempts. In retrial queues, a customer who is deprived of immediate service due to breakdown or unavailability of the server is obliged to wait in the virtual pool known as retrial orbit so as to try its chance of getting service again after a random interval of time. The classical queueing systems work with the assumption that the arriving customer on finding the server occupied or unavailable either joins the waiting line or leaves the system forever. In real life situations, this is however not true always and thus classical queueing models lack the fact that a customer is not lost forever while it returns back after some random amount of time for the service. All of us have experienced the tone 'call back when free'; 'call waiting' etc. while dialing telephone; which represents the retrial queues. The flow of calls circulating in a telephone network consists of two parts (i) the flow of primary calls and (ii) flow of repeated calls. For the analysis of such queueing systems, queues with returning customers form a new class of queueing systems known as ***retrial queueing systems***. The repeating calls cannot be ignored and thus classical queueing models are not applicable here. Due to the associated retrial phenomenon, there are many such cases where standard queueing models and their structures are not suitable for the performance prediction and thus concept of retrial queues came into existence.

Performance evaluation of retrial queues plays a vital role in the design, development and analysis of real time systems including the computer and telecommunication systems. Queueing models are often used for the performance prediction of such systems, and retrial queues are frequently applied to a special type of queueing situations as discussed above. The concept of reattempts plays a significant role in many queueing systems. It is realized that in many practical applications, queueing models with retrials are more appropriate whereas in other cases, it has a little negative effect on the performance measures. The mathematical study on the retrial queues originated with an idea to study the behavior of retrying customers in the context with their competition to primary customers who join the queue to attain the service. The motivation for studying retrial queueing systems is due to their numerous applications in various telecommunication processes, call centres, computer systems, shopping complexes and many more industrial and management situations wherein retrial phenomenon takes place.

The concept of retrial queues is not only limited to the theoretical investigations and mathematical modeling, but finds a very significant and prestigious place in depicting various real life congestion situations. The application of retrial queues is extended from the applicability in telecommunication systems to manufacturing and production processes. The working of call centres is completely based on the study of retrial queues. It is realized that the mathematical version of retrial queues came into existence after the development of telecommunication systems, where the waiting calls or voice beeps like *ring back when free* forced mathematicians to think of a special type of queueing system which is now known as retrial queueing system. No area of our daily routine activities seems to be untouched with the formation of retrial queues wherein either forcefully or wishfully we return back to get served after some random amount of time. We discuss here two significant applications of retrial queues that can be easily realized in the working of call centres and cellular mobile networks.

The working of ***call centers*** can be considered as an analogue to the application of retrial queues. Call centres are basically set either by private companies or government to deal with the queries or problems related to a particular issue by means of telephone. Most major companies use call centres to communicate with their clients. A large volume of calls are received and transmitted every day, also a call dialed may find the server busy/waiting and may receive the signal ring back later on (formation of retrial queues). In call centres, there is usually call blending of incoming and outgoing calls. Markovian retrial queueing models are commonly developed for the study of blocking and delay situations encountered at the working of call centres but depending on their usage and situations other models have also been developed by many researchers working in the area of queueing theory.

Retrial queues find direct applications in the area of wireless communication system and particularly in ***cellular radio networks***. Actually, a cellular network is a radio network in which the whole geographical area is divided into cells and in each cell there is a base station. Each cell uses a different set of frequencies from neighboring cells, to avoid interference and provide guaranteed bandwidth within each cell. These cells when joined together provide radio coverage over the whole area which further enables our mobiles or other wireless devices to connect and communicate with each other via base stations. A retrial queueing formulation of cellular radio networks provides valuable insight. A number of distinguished researchers have put forward their views and investigations on the performance analysis of cellular radio networks.

The main objective of our investigation in the present thesis is to develop queueing models with retrial attempts by incorporating more realistic features namely unreliable server, vacation policy, bulk arrivals, discouragement behavior of the customers, etc. The present introductory chapter provides an overview of the basic concepts that hold important place in the field of retrial queueing systems with respect to both modeling and methodological view points. The rest of the chapter is organized in the following manner. Section 1.2 describes some aspects of the retrial queueing system. Some specialized queueing models in the context of retrial phenomenon have been described in section 1.3. The techniques used for the analysis of retrial models have been discussed in section 1.4. The literature review of retrial queues has been presented in section 1.5. The objectives of our investigation in the thesis have been discussed in section 1.6. Section 1.7 provides an overview of the work presented in the thesis. Finally, conclusions are drawn in section 1.8.

1.2 RETRIAL QUEUES

Retrial queueing systems are used at a large scale for the stochastic modeling of many real life congestion situations mainly in telecommunication systems including local and wide area networks with random multiple access protocols, call centres, etc. The existence of a retrial queueing model is mainly attributed to the unavailability of the server at the particular moment when the customer demands for the service and the customer's resistance to wait in the queue either due to impatience or any other reason. Such situations give rise to retrial queues and thus customer is not lost forever but repeats its attempt to attain service at its own convenience and availability of the server. For fundamental concepts and mathematical analysis of retrial queueing systems we refer books by Gross and Harris (1985), Falin and Templeton (1997), Artalejo and Gomez-Gomez-Corral (2008).

The main characteristics of a basic retrial queueing system are (see fig. 1.1):

i. The new incoming fresh customers arrive in the system to seek service. An arrival is served if the server is free otherwise it may leave the service area or retry back after some random amount of time.
ii. The customers deprived of the service are assumed to wait in the ***orbit*** which is a virtual pool of customers, from where they can make reattempts for the service.
iii. The customers in the orbit cannot monitor the current status of the server while waiting in the orbit. They have the only option to come and check for the status of the server or make reattempts so as to avail the service. Due to this reason there may be a difference in the time when the server becomes free and the reattempts made by orbit customers.
iv. Moreover, customers from the orbit are served in a random order as it depends on the chance. Hence, orbit can be called as a queue with random service discipline.

A retrial queueing system in its basic form consists of two nodes:

I. **Main (primary) node:** In this node blocking is possible. The customers arrive in the system and are blocked if server is either busy or unavailable for the service. Thus, customers are not served and are forced to wait.
II. **Secondary node:** This node is also known as delay node for repeated trials. The blocked customers waiting in the orbit retry for the service randomly after some interval of time.

A variety of queueing models including various prominent features like bulk arrival, priority customers, negative customers, vacation, and impatient customers have been studied by eminent researchers using different techniques. These models seem to be of keen interest for the practitioners and the system analysts for the performance evaluation of many real life and congestion situations.

The significance of such retrial queues is reflected from the existence of a series of international workshops on retrial queues which began with an aim to explore new ideas and thoughts in the development of retrial queues all over the world. Our study on retrial queues is basically motivated by its abundant applications with the advancement of technology in the area of communication and computer networks.

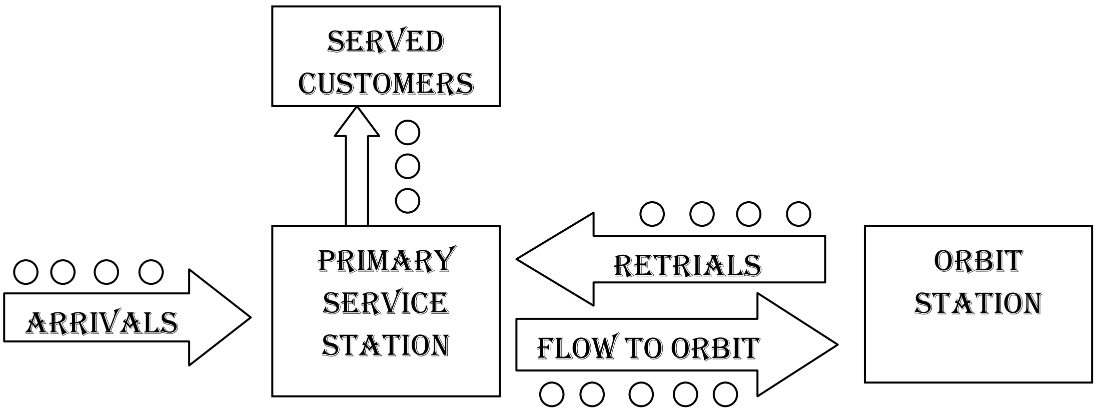


Fig. 1.1: Schematic diagram for the basic retrial queueing system

The modeling of retrial queues can be broadly classified in two main categories as markovian retrial queueing models and non-markovian retrial queueing models.

Markovian retrial models are those models in which all the associated probability distributions follow markovian property. In these cases, we assume that the arrival process, service process, retrial process and any other phenomenon (if present) deals with markovian class of distributions. In probability theory and statistics, the term Markov property refers to the memoryless property of a stochastic process. A stochastic process is said to possess the Markov property if the conditional probability distribution of future states of the process (conditional on both past and present values) depends only upon the present state and is independent of other sequence of previous events.

Basically, markov property is used in reference to the exponential distribution and often with the geometric distribution (cf. Trivedi, 2001). For markovian retrial queueing models, we assume that a customer arrives in the system following Poisson process and is served following exponential discipline. It is worth noting that the inter arrival times follow exponential distribution if arrivals occur in Poisson fashion. Moreover, the retrial attempts are also assumed to follow exponential discipline. Much works have been done on markovian retrial queues in the recent past. For recent research articles on markovian retrial queues, we refer Sherman and Kharoufeh (2006), Artalejo (2010), Wang and Zhang (2013) etc. Various techniques namely probability generating functions, product type solution, and numerical approach based on Runge Kutta method are widely used to obtain the solutions of markovian systems.

In queueing theory, it is not always possible to present every real life situation in the form of markovian model. There exist situations which are complex enough to be sufficiently fit with the distributions holding markov property. Hence, a variety of complex situations are formulated as ***non-markovian retrial queueing models*** wherein at least one or more than one process say service, arrival, vacation, retrial, lifetime, repair etc. follow non-markovian distribution. Such models are categorized as non-markovian models. Enormous work has been done in the direction of solution and modeling of non-markovian retrial queues. The vast literature can be found in the survey articles by Artalejo (1999 a, b) and Artalejo and Falin (2002). The classical techniques namely supplementary variables technique, embedded markov chain etc. are basically used to obtain analytical solutions of such models.

The performance modeling and analysis of retrial queues can be done to determine various significant performance measures which can add meaning to the practical utility of the model. A variety of performance indices are used to judge the efficiency and validity of the model. It is worthwhile to give a brief account of various measures that can be evaluated to study the retrial queueing system. These measures are of significant importance to the system designers and analysts to design more efficient system from the performance point of view. Some of the prominent indices are listed below as:

- **Queue length:** The queue length is the most important key feature associated with any queueing model. The number of customers that can be accommodated in the queue or system under various constraints is of high interest for the engineers to frame more appropriate models in terms of the queue length.
- **Waiting time:** The customer usually needs to wait in the queue so as to get served. The amount of time he spends for waiting in the queue or in the system is termed as average waiting time in the queue or in the system.
- **Long run probabilities:** The stable state of the queueing model can basically be judged after a long run of time. Therefore, we can say that long run probabilities refer

to the probability with which the server remains in a particular state after a long run of time.

- **Reliability:** In many real life congestion situations, the server is unreliable and is prone to breakdowns. This property of unreliable server affects the performance of the system as far as queueing and reliability indices are concerned. Hence, reliability of the server is usually measured to account for the significance of the model.
- **Availability:** It corresponds to the probability of server being available in the system to serve the customers. Hence, the term availability measures the extent to which the formation of retrial queues can be controlled by providing better service.
- **Failure Frequency:** It refers to the probability with which the server fails while servicing the customers.
- **Throughput:** It measures the number of effective services given to the customers.
- **Hazard rate:** It is the instantaneous (conditional) rate of distribution and is defined as the ratio of pdf to the complement of cdf.

The above mentioned indices can be determined or computed to have an idea of the applicability of the modeling of queueing systems. In the present thesis, we have established various performance measures in terms of system state probabilities for the retrial queueing models in different frameworks. These measures are then computed numerically so as to have a better idea of the sensitivity and computational tractability of the concerned models.

1.3 SOME SPECIFIC QUEUEING MODELS

Queueing theory has been a subject of deep interest in recent years because of its theoretical structure as well as its applicability in various real life congestion situations. The pioneer investigator in the field of queueing theory was the Danish mathematician A. K. Erlang, who, in 1909, published "The Theory of Probabilities and Telephone Conversations". Queueing theory originated as a very practical subject, but much of the literature up to the middle 1980s was of little direct practical value. However, queueing theorists became concerned about the applications of the sophisticated theory that has largely arisen since the close of World War II. Numerous models with a variety of distinct features like vacation, impatience, bulk arrival, phase services, repair, control parameters etc. have been investigated by the queue theorists. Here, in this section we present a brief account of some queueing models developed by eminent researchers from time to time which are closely related to our study done in the present thesis.

1.3.1 Bulk Queueing Models

The concept of bulk arrivals and bulk services has gained tremendous significance in present situations. Bulk queueing models are associated with the phenomenon that either service or arrival of the customers or both the processes occurs in batches (i.e. bulk or group). In the present scenario, we may find that some studies are devoted to the ***bulk queues*** in combination to other prominent features like vacation, priority, N-policy, retrial process etc. Bulk queues can be categorized in three main categories (see fig. 1.2).

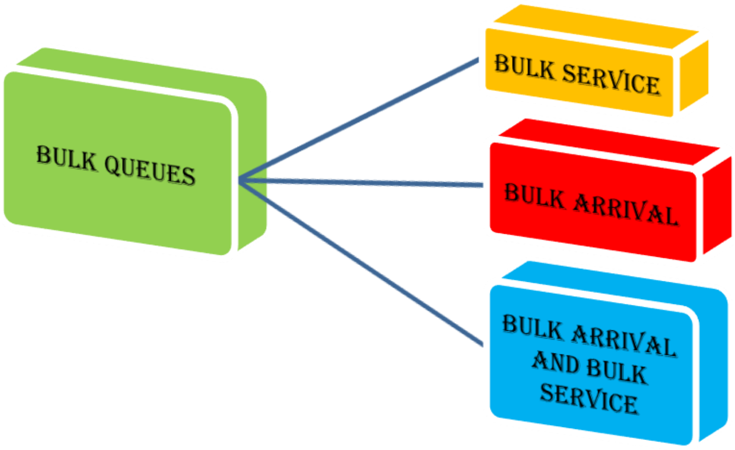


Fig. 1.2: Classification of bulk queueing models

In most of the queueing literature, it is assumed that the customers arrive singly at a service facility. But this assumption is violated in many real-world queueing situations; for example letters arriving at a post office, ships arriving at a port in convoy, people coming to restaurant, election campaigning and so on, are some of the examples of queueing situations in which the customers do not arrive singly, but in bulk or groups which represents ***bulk arrival queues***. Also, the size of an arriving group may be a random variable or a fixed number. Mathematically and also from the practical point of view, the cases when the size of an arriving group is a random variable, are more common, and also more difficult to handle.

Bulk service queueing models can be visualized in traffic signal systems, in loading and unloading of cargo at seaport, in many congestion situations, etc. Bailey (1954) introduced the concept of bulk service and the same was later studied by a number of researchers while developing a variety of queueing models.

1.3.2 Queueing Models with Vacation

Vacation models with the feature of reattempts have been the area of interest for the researchers since recent past. It usually happens that the server may go for some recreation activities or may get involved in some other work and may not be available to serve the next customer. This is usually considered as the vacation of the server of some finite or random length. During this period, the server is not available for servicing and alternates between busy and idle states. During the busy period, the server works at a full speed while during the idle period it does not process any work. In case of working vacation, the server may render service with a lower speed during vacation period.

The applications of vacation models can be realized in almost all congestion scenarios. The study on retrial queues with vacations is motivated due to its numerous applications in real life scenarios from the bank counters to shopping malls, where the concepts of reattempts as well as vacations exist simultaneously. Enormous works has been done on different types of vacation policies keeping in mind the applications of such models to real life scenarios. On the basis of nature of vacation taken by the server in any queueing system, we broadly classify them in the following categories (cf. fig. 1.3)

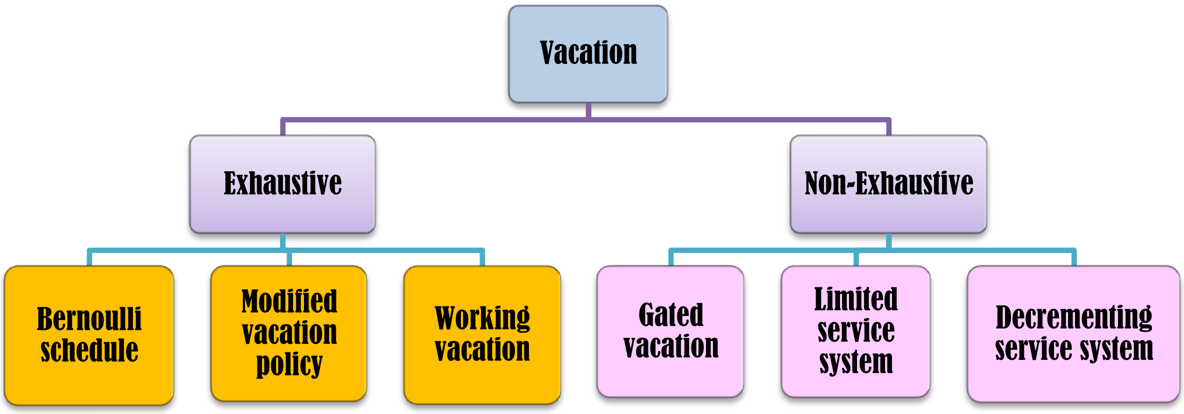


Fig. 1.3: Classification of various types of vacations

Exhaustive Service System: In this type of system, the server takes vacation only if no customers are available in the system for the service. This can be further categorized as: (i) ***single vacation system*** and (ii) ***multiple vacation system***, depending on the number of vacations permitted to the server. If a server can take at most one vacation between two successive busy periods then it is termed as single vacation system. If on coming back from the vacation, the server finds no customers waiting for the system then it may go for another vacation in case of multiple vacation system.

Non-Exhaustive Service system: In this type of service system, the server may go for vacation even if some customers are present in the system.

Different vacation policies namely ***Bernoulli vacation schedule, modified J vacation*** policy have been proposed and studied from time to time to develop more realistic vacation models. Different vacation policies which we have incorporated in our thesis work can be summarized as follows:

- *Bernoulli vacation schedule:* According to this policy, the server after completing service of a customer, has the option either to go for a vacation with some probability say 'p' $(0 \le p \le 1)$ or continue serving the next customer with probability $(1-p)$. As this vacation schedule works like Bernoulli distribution and is hence termed as Bernoulli vacation schedule.
- *Working vacation:* Sometimes, instead of taking completely off from the service, the server prefers to do another job or service at slower rate; such queueing situations are known as ***working vacation***. In this case, the service of the system is not completely switched off and the server is allowed to serve the customers with slower rate.
- *Modified J vacation policy:* Modified vacation policy states that when no customers are recorded in the system, the server may go for at most J vacations repeatedly until at least one customer is recorded in the orbit on returning back from the vacation. In case no customer is found even after J^{th} vacation, then the server will remain with the system in the dormant state.

1.3.3 Unreliable Server Queueing Models

The researches on retrial queueing theory also realized the need of development of more realistic models with respect to the reliability of the server. Earlier classical retrial queueing models were developed under the assumption that the server providing service to the customers is reliable. But this concept of reliable server is not quite true in the context of realistic application of the model as no server is ideally reliable. The server is subject to breakdowns while serving the customers and can be repaired. Some studies appeared in the area of performance analysis of unreliable server retrial queueing system by considering different repair criteria. The repair of the server can either be completed in a single step or in a series of some essential or optional steps. Sometimes, the repair process is also completed following threshold based recovery.

1.3.4 Phase Type Queueing Models

The classical queueing models were developed with the fact that a job/service is completed in only single phase. But this is not true in regard to all the various queueing situations which require more than one phase of service. Queueing models with phase service and phase repair play a vital role in depicting and analyzing many queueing situations. For example, the repair process of the broken down of two-wheeler vehicle can either be completed in a single step if minor problem is detected. But in case when the vehicle is damaged to a greater extent, then it may require more than one step repair. Moreover, the owner of two-wheeler may demand for extra repair like new gears, new body paint of the vehicle which may add some optional repair demands. We can broadly classify phase type service queueing systems as shown in fig. 1.4.

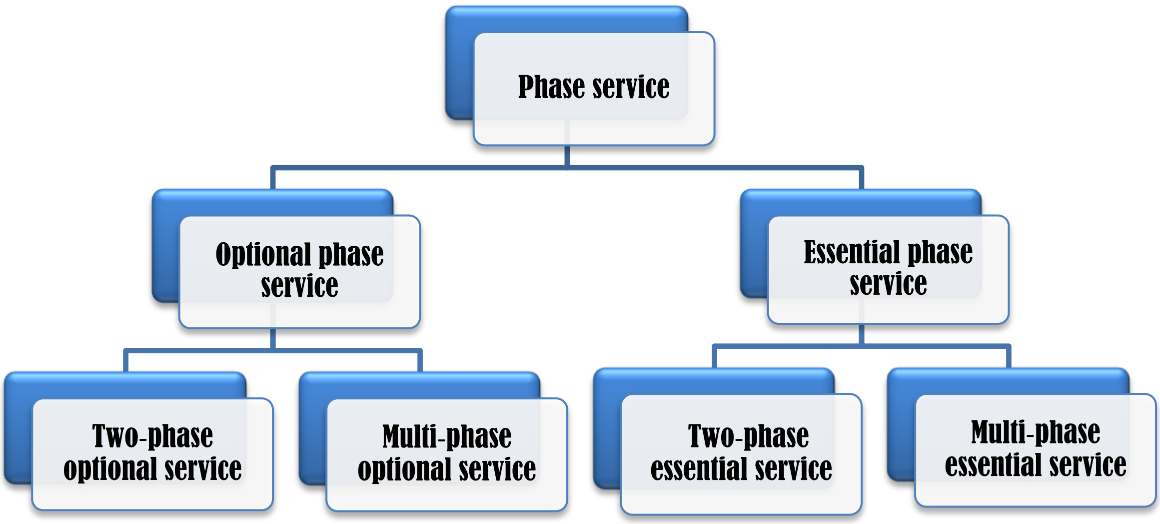


Fig. 1.4: Classification of phase service queueing models

- ***Two-phase essential service:*** In this type of the retrial queueing models, the service process is completed in two compulsory phases. Here, the customer is provided two essential services one-by-one by the server. These types of situations usually arise in manufacturing and industrial process where the work is completed in two compulsory sequential steps.
- ***Two-phase optional service:*** There exist situations in which a customer after completing the first phase of service opts for the second phase of service; for example we cite the case of ATM transactions in which a customer may opt for second preceding services like withdrawal, balance enquiry etc.

- ***Multi-phase essential service:*** A variety of real life congestion situations are associated with the fact that the service is completed in a series of a finite number of compulsory phases rather than in a single phase. For example, medical checkup at doctor's clinic can be completed via a number of compulsory phases like blood pressure checkup, ECG test, X-rays, etc. Such type of situations motivate queue theorists to model retrial queueing systems in which the service is assumed to be completed in more than one essential/optional phases and customers retry for their server again and again on finding the service busy with some other customer.
- ***Multi-optional phase service:*** These models are associated with the fact that after completing first essential phase of the service, the customer has the option to demand for other available services with some probability or can leave the system without taking the optional service. The food orders at restaurants, shopping at malls, etc. are some situations where after getting the first essential service, the probability to avail other optional service depends on customer's choice.

1.3.5 Priority Queueing Models

In the classical queueing models associated with the congestion situations, the customers are assumed to be served on the basis of first come first serve (FCFS) criterion. But there may be queueing situations in which the customers/jobs are assigned or classified according to some priority index.

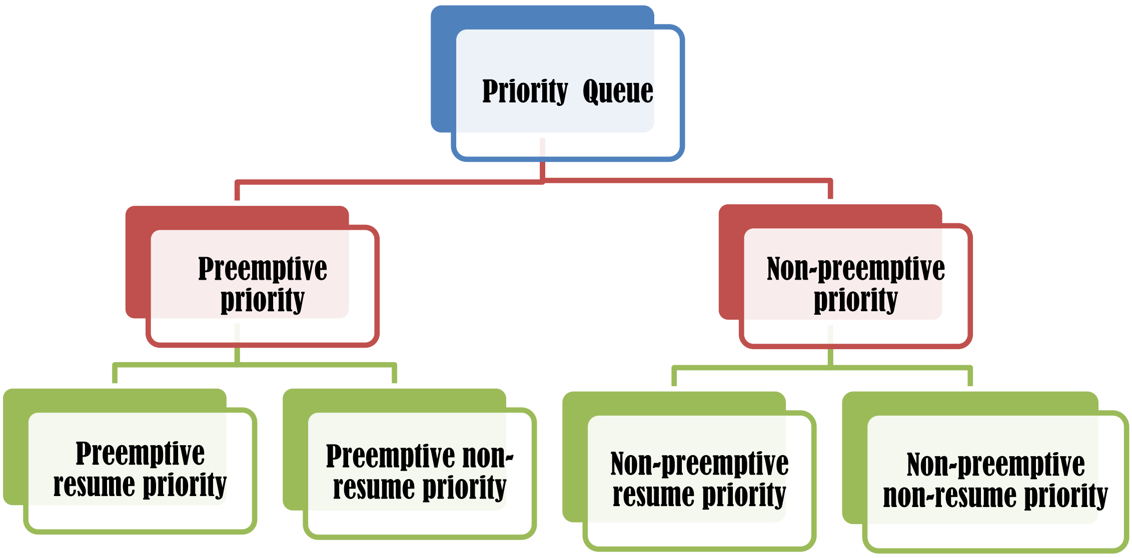


Fig. 1.5: Classification of priority queueing models

The retrial queues with priority also find a variety of applications in many realistic queueing scenarios including the admission at hospitals, data transmission, etc. For example, a critical patient is handled prior to other patients having minor problems or injuries. The related literature on priority queue can be found in the book by Jaiswal (1968). The priority queueing models can be categorized as ***preemptive priority models*** and ***non-preemptive priority models*** (see fig. 1.5).

The preemptive priority models are associated with the fact that when a priority customer enters the system then the service of non-priority customer is immediately stopped and the service of priority customer is started. This fact can be further categorized whether the service of non-priority customer will be resumed or it will start from the initial stage. On the other hand, in case of non-preemptive priority, the priority customer is served after the completion of the service of the customer already with the server (priority or non-priority).

1.3.6 Queueing Models with Negative Customers

It usually happens that the computers and other electronic devices are affected by the attack of virus or malwares which affect the normal functioning of the system. These unwanted arrivals are termed as negative customers and usually enter the system either while accessing internet, pairing our own devices with already affected devices, etc. Negative customers are like unwanted arrivals which affect the normal working of the system either by stopping the service process or by lowering down the rate of service process. The arrivals of negative customers affect the system in a variety of ways, i.e., either they may damage the system completely or may remove the customer at the end/head of the queue waiting for the service. Unlike positive customers, the negative customers are not accumulated in the queue to get served. The concept of negative customers finds various applications in communication systems, computer protocols, neural networks modeling, etc. where the system gets destroyed or failed with the arrival of unwanted customers like virus in the computers or extra order of some inventory items, etc. The queueing models with negative customers are also known as G-queues after the name of Gelenbe (1989) who introduced the concept of negative customers in queueing theory.

1.3.7 Queueing Models with Discouragement

To wait in the queue for the service is the unwanted job for the customers. The customers waiting in the queue for their turn may become impatient and may act in different manners. These situations of impatience behavior of the customers can be realized in almost every real life congestion situations from registration counters at doctor’s clinic where the patients get discouraged due to delay in service. Being motivated from such realistic situations, the queue theorists also developed a variety of retrial queueing models enriched with the phenomenon of discouragement behavior of the customers. A customer waiting in the queue may behave in the following manners due to discouragement:

- ***Balking:*** The customer on seeing a long queue of the customers waiting for the service may decide not to join the queue and leaves the system without joining the queue.
- ***Reneging:*** The customer waiting for the service in the queue may get tired of waiting and gets impatient. In this behavior, the customer leaves the system after waiting sometime in the queue and before the start of the service, and is said to be reneged.
- ***Jockeying:*** This behavior of impatience is visible at those situations where there are multi server systems. Jockeying can be described as the movement of a waiting customer from one queue to another (of shorter length or which appears to be moving faster, etc.) so as to get served at an earlier stage.

1.3.8 Control Policies for Queueing Models

Optimal control is one of the main issues behind the mathematical modeling of queueing systems. The problem of admission/service control is of great importance due to its applicability in inventory, telecommunication process, production processes, etc. In case if the admission in the queues are not controlled, then it may result in the bursting of the system and situation may become out of control of the authorities. The congestion in the queueing system is controlled by various ways either by closing the gate, or by charging an extra fee to some event or so on. Control policies are applied to many real life situations; for example we refer an exhibition of newly launched cars wherein the provision of entrance passes/fees is done for the audience/spectators to control the rush. In case no restriction is imposed on the arrival of the spectators to exhibition then it would have been irresistible to control the crowd. The control policies are not only aimed to control the number of customers in the system but also to make the system more productive and economic. The control policies which are incorporated in the concerned models in our thesis are as follows:

- ***N-policy:*** According to this policy, the server starts servicing only if N customers are accumulated in the system otherwise system is said to be in build up state. Once service is initiated, the server renders service to the customers till the system becomes empty.

- ***Threshold based recovery:*** This kind of control policy is used to frame more economic system by applying control policy on the repair process. As per this policy, the broken down server is repaired after a threshold (pre-decided) customers are accumulated in the system. This is done to optimize the time and money spent on the repair process.
- ***Bernoulli admission mechanism:*** Bernoulli admission mechanism can be used to control the admission or arrival of the customers in the queues with reattempts or retrials. Artalejo and Atencia (2004) and Artalejo *et al.* (2005) proposed this policy for the admission control in continuous and discrete queueing systems respectively. It is reasonable to assume that the arrival of the new customers is controlled in such a way that each individual blocked customer is admitted or allowed to join the system with probability say $\zeta\,(0 \le \zeta \le 1)$. If the arriving customer/batch finds the server in idle state, then one of the admitted customers joins the server whereas rest of the customers join the retrial group; otherwise if the server is busy, the whole batch joins the orbit. This mechanism which is known as Bernoulli admission policy can be considered as an admission control rule to reduce the congestion at the initial stage.
- ***F-policy:*** The arrival in the system can be controlled using F-policy also. According to this policy, no more customers will be allowed to enter the system if the capacity of the system is full but again the arrival process will be initiated at the later stage if a sufficient number of customers served is less than threshold value (say F).

1.4 METHODOLOGICAL ASPECTS

With the advancement of stochastic modeling of retrial queueing systems, the classical techniques required to solve such complicated systems also grew with time. As the retrial queueing systems became more complex due to their applicability to real life congestion scenarios, the methodologies required for their solutions also developed simultaneously. A variety of techniques both analytic and numerical methods have been used to obtain the solutions for different retrial queueing systems. For detailed understanding of different techniques and methodologies used for the analysis of retrial queues, we refer monographs by Artalejo and Gomez-Corral (2008).The use of methodology depends on the nature of the retrial queueing problem under consideration. A large number of techniques have been used for the mathematical analysis of retrial queueing systems. Here, we discuss in brief only those techniques which have been used in our thesis work in order to predict the performance measures of concerned retrial queueing models.

1.4.1 Probability Generating Function (PGF)

The tool of generating functions as developed by Euler is widely used to obtain solution for those stochastic processes which involve non-negative random variables. The importance of using generating functions lies in the fact that a single function can easily be the representative of whole set of values involved. Let $\{a_k\}$ be a sequence of real numbers then using a new variable say 'z', we may define a function

$$A(z) = \sum_{r=0}^{\infty} a_k z^k \tag{1.1}$$

If power series given by (1.1) converges in interval say $-z_0 < z < z_0$, then A(z) is called the generating function of the sequence $\{a_k\}$ (cf. Medhi, 1991). To understand it clearly, we refer mathematical results obtained by Sherman and Kharoufeh (2006) and Krishna Kumar *et al.* (2010) who employed this technique of probability generating function to obtain performance indices of interest for the concerned queueing systems. Brandwajn and Begin (2008) used conditional probability approach to study M/G/1 like queues.

1.4.2 Supplementary Variable Technique (SVT)

The non-markovian retrial queueing models depict more realistic congestion situations as they are not restricted to the statistical distributions having memoryless property. The supplementary variable technique (SVT) is a very elegant and classical technique which is used for the solution of such non-markovian systems. The literature on retrial queueing models has abundance evidences in support of the use of supplementary variables technique for the solution purpose of non-markovian systems. By using this technique, non-markovian process in continuous time is made markovian by introducing one or more supplementary variables (cf. Cox, 1955). This particular technique has been widely used in the literature for the analysis of a variety of non-markovian retrial queueing systems. The mathematical models using supplementary variables technique can be developed in two manners either using elapsed process time or the remaining process time corresponding to non-markovian random variable.

In our investigations on retrial queues with unreliable server, we have used SVT along with generating function (cf. chapters 2-6). It is worthwhile to describe the retrial queueing model with unreliable server (cf. Wang *et al.*, 2001). In this model, the service process and the repair process are general distributed and thus the model is non-markovian. Therefore, supplementary variables are introduced corresponding to service and repair process and markovian model is obtained. Let us consider that the customers arrive in Poisson fashion with rate λ and are served according to general distribution with rate μ. Further, we assume that the retrial duration is exponential distributed with rate θ. The unreliable server may break down exponentially with rate α and is repaired following general distribution with rate β. Let N(t) be the number of customers in the system at time t. Now, we introduce random variables X(t) and Y(t) corresponding to the elapsed service time and elapsed repair time of the server at any time t. The state probabilities are as:

$P_{W,i,1}(t,x)dx$: Joint probability that at time t, there are *i* customers in the retrial group and the elapsed service time of the customer lies between x and x+dx.

$P_{R,i,1}(t,x,y)dy$: Joint probability that at time t, there are *i* customers in the retrial group, a customer is being served with elapsed service time x and elapsed repair time of the server being under repair lies between y and y+dy.

$P_{I,i,0}(t)$: Probability that the server is idle at time t and there are *i* customers in the retrial group.

Using supplementary variable technique, following governing equations can be framed:

$$\left[\frac{d}{dt}+\lambda+i\theta\right]P_{I,i,0}(t)=\int_0^\infty P_{W,i,1}(t,x)\mu(x)dx \tag{1.2}$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda+\alpha+\mu(x)\right]P_{W,i,1}(t,x)=\int_0^\infty P_{R,i,1}(t,x,y)\beta(y)dy+\lambda P_{W,i-1,1}(t,x) \tag{1.3}$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+\lambda+\beta(y)\right]P_{R,i,1}(t,x,y)=\lambda P_{R,i-1,1}(t,x,y) \tag{1.4}$$

The boundary conditions are:

$$P_{R,i,1}(t,x,0)=\alpha P_{W,i,1}(t,x) \tag{1.5}$$

$$P_{W,i,1}(t,0)=\lambda P_{I,i,0}(t)+(i+1)\theta P_{I,i+1,0}(t) \tag{1.6}$$

The normalizing condition is given by

$$\sum_{i=0}^{\infty}\left\{P_{I,i,0}(t)+\int_0^\infty P_{W,i,1}(t,x)dx+\int_0^\infty\int_0^\infty P_{R,i,1}(t,x,y)dxdy\right\}=1 \tag{1.7}$$

The above set of equations (1.2)-(1.7) can be further solved by using probability generating function technique as discussed in section 1.4.1. Various performance measures can be further determined by using required probabilities.

1.4.3 Embedded Markov Chain Technique (EMC)

In queueing literature, embedded markov chain is widely used to obtain the stationary probability distribution of the continuous time markov chain. Embedded markov chain is actually a regular discrete time markov process, and each element in the one-step transition probability represents the conditional probability of transferring from one state to another. This classical technique is widely used to analyze the queueing model where the distributions of the inter arrival time or the service time do not possess the memoryless property, i.e. are not exponential. For more details on embedded markov chain analysis of the retrial queueing models, we refer book by Falin and Templeton (1997).

In chapter 7, we analyze the steady-state behavior of a batch arrival retrial queue with multioptional services and phase repair under Bernoulli vacation by using EMC. To explain this method in brief we describe the model investigated by Artalejo and Atencia (2004). They considered a batch arrival retrial queueing system wherein customers arrive following Poisson process with rate λ and the probability that a batch of k customers arrive is $c_k\,(k\geq 1)$. The customers are admitted in the system following Bernoulli admission policy; therefore let ‘p’ be the probability of admission of each individual customer. Hence, $a_n\,(n\geq 0)$ is the probability that a group of n customers is allowed in the system is as follows:

$$a_0=\sum_{k=1}^{\infty}c_k(1-p)^k \text{ and } a_n=\sum_{k=n}^{\infty}c_k\binom{k}{n}p^n(1-p)^{k-n},\quad n\geq 1. \tag{1.8}$$

If the server is free, then one of the customers from the batch is served and remaining customers join the retrial orbit. However, all the arriving customers are forced to join the orbit if the server is busy. The customers from the retrial group make reattempts following exponential law with rate $\gamma_j=\nu(1-\delta_{n,0})+n\gamma$ where n is the number of customers in the orbit, $\delta_{n,0}$ is the Kronecker’s delta function and γ is the retrial rate, respectively. Let n^th^ service completion or departure occurs at any time instant say τ_n and C_n =C (τ_n-) be the state of the server before the time instant τ_n. Also, let N_n = N (τ_n+) denotes the number of retrying customers present in the system just before τ_n. We have the sequence N_n=N (τ_n+) which is embedded Markov renewal process corresponding to continuous time Markov process Z (t). Then we have the following transition-

$$(N_n / N_{n-1}=J) = \begin{cases} J-1+B_n & \text{with probability } \dfrac{\gamma_J}{\bar{\lambda}+\gamma_J} \\ J+V_n-1+B_n & \text{with probability } \dfrac{\bar{\lambda}}{\bar{\lambda}+\gamma_J} \end{cases} \qquad (1.9)$$

where, B_n is the number of customers that arrive during the n^{th} service time and V_n is the number of customers allowed to join the system if the n^{th} customer proceeds from a batch arrival. Also, $\bar{\lambda} = \lambda(1-a_0)$ due to Bernoulli admission mechanism involved.

1.4.4 Matrix Geometric Method (MGM)

Sometimes, it is not possible to obtain analytical solution of the queueing system under consideration. Here comes the requirement of some numerical procedures to obtain the solution for the system. Matrix geometric method (MGM) is one of the powerful numerical techniques that permit us to deal with the models whose activities are performed in phases. The advantage of this method lies in the fact that it can be used to solve a large number of equations at a time. Basically, MGM can be applied for the analysis of queueing problems for which the system states can be divided into two categories (i) initial portion which acts as boundary condition and (ii) repetitive structure which acts as a base to form rate matrix. The matrix geometric method to determine the probability vector is applicable for the system of equations whose transition matrices have special block structure with repetition of elements of sub matrices. The concerned model can be structured as a square matrix of infinite dimension that converges to finite dimension matrix using the minimal matrix to get recursive relation of the probability vectors.

Neuts (1978, 1981) deserves the credit to develop the matrix geometric method and provided a number of solutions to a variety of queueing problems using matrix geometric approach. Several queue theorists developed repetitive matrix block structured model to obtain solutions of various queueing problems using matrix geometric approach. This method requires tridiagonal block matrix structure as follows:

$$Q = \begin{pmatrix} \mathbf{F_0} & \mathbf{F_1} & & & & \\ \mathbf{F_2} & \mathbf{G_1} & \mathbf{G_2} & & & \\ & \mathbf{G_0} & \mathbf{G_1} & \mathbf{G_2} & & \\ & & \mathbf{G_0} & \mathbf{G_1} & \mathbf{G_2} & \\ & & & \mathbf{G_0} & \mathbf{G_1} & \mathbf{G_2} \\ & & & \ddots & \ddots & \ddots \end{pmatrix} \qquad (1.10)$$

where, F_0, F_1, F_2, G_0, G_1 and G_2 are sub-matrices of appropriate dimension.

Using probability vector $\boldsymbol{\pi} = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3, ...]$, the balance equations can be constructed as follows:

$$\pi_0 \mathbf{F_0} + \pi_1 \mathbf{F}_2 = 0 \tag{1.11}$$

$$\pi_0 \mathbf{F}_1 + \pi_1 \mathbf{G}_1 + \pi_2 \mathbf{G}_0 = 0 \tag{1.12}$$

$$\pi_{i-1} \mathbf{G}_2 + \pi_i \mathbf{G_1} + \pi_{i+1} \mathbf{G}_0 = 0,\ i \geq 2 \tag{1.13}$$

The sub vectors are related to each other as

$$\pi_i = \pi_1 \mathbf{R}^{i-1}, \tag{1.14}$$

where **R** is the constant matrix known as rate matrix or Neut's rate matrix. Using (1.14), the equations (1.11)-(1.13) can be written as:

$$\begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{bmatrix} \mathbf{F_0} & \mathbf{F_1} \\ \mathbf{F_2} & \mathbf{G_1} + \mathbf{RG_0} \end{bmatrix} = (0, 0) \tag{1.15}$$

which can be further used to solve for $\begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix}$ and other sub vectors.

1.4.5 Maximum Entropy Principle (MEP)

Entropy defines the degree of randomness or unevenness of any system or probabilistic distribution. It also measures the expected value of information contained in any message. The maximum entropy principle (MEP) was introduced by Shannon (1948) to study the problems of information theory as the measurement of uncertainty. This principle is applicable to select the appropriate probability distributions for the queueing situation (cf. Kapur, 1989; Karmeshu, 2003).

In many queueing scenarios, sometimes it happens that all the information available is not sufficient to estimate the distribution. For queueing models solved analytically or numerically, the available information may either be present in the form of explicit expressions for the queue length, waiting time, long run probabilities etc. in the form of constraints. But there may be a wide range of distributions which satisfy those constraints. In many realistic problems, the question arises about the best or right distribution which fits the queueing situation. Here, arises the role of maximum entropy principle which helps in providing the best suited distribution based on available information which can be treated as constraints.

To understand the applicability of the maximum entropy approach, we cite the unreliable $M^x/G/1$ model which has been extended by incorporating many realistic features in chapters 2 and 4. Let us define,

P_0 = Probability that there are no customers in the system

$P_1(n)$ = Probability that there are n ($\geq$1) customers in the system when the server is in operation.

$P_2(n)$ = Probability that there are n ($\geq$1) customers in the system when the server is broken down.

The entropy function Y of the queueing system under N-policy is framed as:

$$Y = -P_0 \log P_0 - \sum_{n=0}^{\infty} P_1(n) \log P_1(n) - \sum_{n=1}^{\infty} P_2(n) \log P_2(n) \qquad (1.16)$$

subject to the constraints as-

i. $$P_0 + \sum_{n=0}^{\infty} P_1(n) + \sum_{n=1}^{\infty} P_2(n) = 1, \qquad (1.17)$$

ii. $$\sum_{n=1}^{\infty} P_1(n) = \rho, \qquad (1.18)$$

iii. $$\sum_{n=1}^{\infty} P_2(n) = \rho\alpha / \beta \qquad (1.19)$$

where, ρ , α (β) denote the probability of server being busy, failure rate (repair rate) of the server, respectively.

The entropy function (1.16) subject to constraints (1.17) – (1.19) can be further solved using Lagrange’s function to find the steady state probabilities of various states of the server.

1.5 SURVEY OF LITERATURE

Queueing theory was developed to predict the behaviour of the congestion systems in different frameworks. Due to abundant applications, a plethora of literature is available on the performance analysis of retrial queueing models. Since the pioneering works on retrial queues published in 1950’s, retrial queues have been widely used to provide probabilistic solution to the problems arising in cellular mobile networks, manufacturing and production processes, computer networks and many other real life congestion scenarios. In the present thesis, we have developed many retrial queueing models which are enriched with various prominent features like unreliable server, bulk arrival, vacation, impatient customers, optional service, optimal control policy etc. In this section we give literature review of the prominent researches that took place in the field of retrial queues related to our study. We have reported important contributions of recent years especially in the last decade. The review presented here has been divided into various subsections dealing with retrial models incorporating different features and variety of methodologies used for their solution. The notable contributions by various researchers in the field of retrial queues related to our research works are as follows:

1.5.1 Bulk Retrial Queueing Models

Many real life congestion situations deal with the fact that the arrival or service of the customers may take place in batches. This fact has motivated researchers working in retrial queues to incorporate this practical feature in their investigations so as to frame more general models which are consistent enough with the real life situations. Nobel and Tijms (1999) studied $M^X/G/1$ queue with optimal control. For the review on bulk retrial models we refer survey articles by Artalejo (1999 a, b), Artalejo and Falin (2002), and Artalejo (2010). It is worthwhile to mention the significant contributions in the area of ***bulk arrival retrial queues*** using supplementary variables technique by Choudhury and Deka (2009), Ke and Chang (2009a), Choudhury *et al.* (2010). Choudhury (2007) and Falin (2010) investigated non-markovian bulk queues using embedded markov chain technique. Arrar *et al.* (2012) investigated the asymptotic behavior of M/G/1 retrial queue wherein arrival process takes place in batches using embedded markov chain. Nobel (2013) investigated batch arrival queueing model with retrials and a tolerant server using generating function.

Apart from bulk arrival retrial queueing situations, service may also be rendered in batches. Initially, Borthakur and Medhi (1974) have studied a queueing system with arrival and service in batches of variable size. Laxmi and Gupta (1999) investigated finite buffer bulk service queue. Sikdar and Gupta (2005), Goswami *et al.* (2006), Banik *et al.* (2007), Sikdar *et al.* (2009) and many more have contributed significantly to the study of bulk service queues. A very few papers are available on the bulk service retrial queues. Cordeau and Chaudhry (2009) gave a complete solution to bulk arrival bulk service queue and obtained average queue length of the system. Chaudhry *et al.* (2010) provided results for the number of customers for bulk service queues. Haridass *et al.* (2012) obtained cost analysis of bulk service retrial queue and also obtained probability generating function of queue size distribution. Claeys et al. (2011, 2013) analyzed versatile batch service queueing model with correlation in the arrival process.

1.5.2 Retrial Queueing Models with Vacation

Vacation is a key feature of queueing models in a variety of real life congestion scenarios. The unavailability of the server in the system when no more customers are available for the service can be realized in many queueing situations including the bank counters, manufacturing systems with machine breakdown and processor schedules in computer and switching systems, etc. The first investigation on the queues when the server is unavailable for some times i.e. on vacation was done by Miller (1964). For notable works on vacation queues in recent past we refer Doshi (1986, 1990), Takagi (1986, 1991) and Medhi (1991). Numerous works has been done in the direction of retrial queues with different types of vacation. It is to be worth noting that several researchers have paid their attention to the study on retrial queues with single/ multiple vacation, modified vacation, working vacation and Bernoulli vacation schedule. For notable researches in this area we refer significant works done by Krishna Kumar and Arivudainambi (2002), Wenhui (2005), Choudhury (2007), Boualem *et al.* (2009), Aissani (2009), Banik (2009, 2010) etc. They all studied single server retrial queues with ***Bernoulli vacation schedule*** using embedded markov chain technique. Ke and Chang (2009a) investigated bulk arrival retrial queue with using supplementary variable technique. Recently, Choudhury and Ke (2012, 2014) analyzed batch arrival retrial queue with Bernoulli schedule using embedded markov chain and obtained important performance characteristics of the queueing system.

Li *et al.* (2009) investigated non-markovian queue with working vacations using matrix analytic approach. Goswami and Selvaraju (2010), Arivudainambi *et al.* (2013), and Aissani *et al.* (2014) did the performance analysis of retrial queues with working vacation by considering the general distributed service process. Tao *et al.* (2014) investigated M/M/1 retrial queue with working vacation and feedback under N-policy using matrix analytical methods.

Ke and Chang (2009b) and Jain and Bhargava (2009) studied non-markovian retrial queues with ***modified vacation policy*** using supplementary variable technique and obtained queue size distribution.

1.5.3 Unreliable Retrial Queueing Models

The modeling for unreliable retrial queueing systems has been done by many queue theorists and a considerable amount of literature is also available in this direction. For literature on unreliable queues we refer a recent survey article by Krishnamoorthy *et al.* (2014). A variety of unreliable retrial queueing models incorporating various prominent features like vacation, bulk queues, impatience etc. has been investigated by researchers in recent past. The worth mentioning contributions in the area of unreliable retrial queues are due to Sherman and Kharoufeh (2006), Atencia *et al.* (2006a), Mokaddis *et al.* (2007), Atencia *et al.* (2008), Jain and Agarwal (2009), etc. Krishna Kumar *et al.* (2010), Krishna Kumar *et al.* (2011) and Krishnamoorthy *et al.* (2012) investigated retrial queues in different contexts.

More advanced repairable queueing models have also been studied by queue theorists using a variety of mathematical techniques. Dimitriou and Langaris (2010), Choudhury *et al.* (2010) and Choudhury and Deka (2012), Jain *et al.* (2013) analyzed repairable retrial queueing model and provided performance indices to study the effect of breakdown of the server on the queueing and reliability indices of the system. Zhang and Wang (2013) obtained performance analysis of unreliable retrial queues with finite number of sources. Recently, single server retrial queues with server breakdown is studied by Boualem (2014), and Lakshmi and Ramnath (2014). Choudhury and Ke (2014) investigated the steady state behavior of the unreliable retrial queue with Bernoulli schedule and delaying repair.

1.5.4 Retrial Queueing Models with Priority

Retrial queues with priority finds significant place in the literature of retrial queueing theory. Some notable results on single server priority retrial queues can be found in the articles by Langaris and Moutzoukis (1995), Choi and Chang (1999). Choi and Park (1990) investigated a single server M/G/1 retrial queue with two types of customers and obtained joint distribution of the queue lengths. Krishna Kumar *et al.* (2002) studied retrial queue with preemptive resume priority where two types of customers arrive following markovian arrival process. Two class priority markovian queueing systems were investigated by Tarabia (2007a, 2007b). The other prominent contributions in the field of retrial queues with priority are due to Atencia *et al.* (2006b), Jain and Bhargava (2008) and Wang (2008). Goswami and Selvaraju (2013) investigated working vacation queue with priority and breakdowns. Dimitriou (2013a) studied batch arrival priority retrial queue with admission control and repeated demands. Peng *et al.* (2014) obtained results for the non-markovian retrial queue with break downs and preemptive resume priority using probability generating function technique. Recently, Vadivu *et al.* (2014) investigated non-Markovian loss system with priority and breakdowns. The multi server retrial queueing system with random number of servers and priority has been studied by Vinayak *et al.* (2014).

1.5.5 Retrial Queueing Models with Negative Customers

Some papers on the study of negative customers have appeared in the queueing literature. The remarkable contributions of Gelenbe (1989, 1991, and 2000) opened ways for the future research on the queues with negative customers known as G-queues. Harrison (1993, 1996) investigated queues with negative customers using a new technique and paved a new dimension to the modeling of unreliable queues with negative customers. Shin (2007) analyzed multiserver G-queues with disasters and reattempts. The recent articles on G-queues can be found in the bibliography on negative customers by Do (2011a). Liu *et al.* (2009) and Dimitriou (2013b) studied negative arrival retrial queues with unreliable server and pre emptive resume and gave interesting mathematical results for non-markovian queues. Wu and Yin (2011), Wu and Lian (2013a, 2013b) and Krishna Kumar *et al.* (2013) investigated single server retrial G-queue with priority and breakdowns. The queueing and reliability analysis of non-markovian retrial G-queue has been done by Gao and Wang (2014). Recently, Berdjoudj and Aissani (2014) analysed M/G/1 retrial queue with negative arrivals using martingale technique.

1.5.6 Retrial Queueing Models with Discouragement

The significance of the retrial queues with discouragement can be felt from the research done by a number of researchers in this area. The impatient nature of customers was mathematical structured by Li and Zhao (2005) who studied retrial queue with constant retrial rate and impatient customers. Ke and Chang (2009b) investigated M/G/1 retrial queue with modified vacation policy by incorporating balking and reneging concepts. Wang and Li (2009) developed a queueing model with impatient customers and second phase of service. The transient as well as steady state analysis of M/M/1 queue with impatient customers and failures has been done by Tarabia (2011).

For more recent articles, it is worthwhile to mention the significant contributions by Economou *et al.* (2011), Zhang *et al.* (2013), and Selvaraju and Goswami (2013) who studied equilibrium balking strategies of Markovian queues.

1.5.7 Phase Type Retrial Queueing Models

Phase service queueing models have been studied extensively by many eminent researchers for improving the grade of service in many industrial problems. Different type of phase service, viz. compulsory phases, homogenous optional phases, heterogeneous phase service etc. have been explored by many researchers in the literature. There is a significant literature available on the queue with different types of phase type service. We mention some recent prominent works on various types of phase service retrial queueing models as given in table 1.1.

Table 1.1: Contributions to phase service retrial queueing models

Type of phase service	Research Contributions on retrial queues
Two-phase essential service	Dimitriou and Langaris (2008), Choudhury (2008a), Choudhury and Deka (2008), Wang and Li (2009), Dimitriou and Langaris (2010), Senthil and Arumuganathan (2010), Choudhury and Deka (2013)
Two-phase optional service	Choudhury and Deka (2009), Maraghi *et al.* (2010), Ke *et al.* (2011)
Multi-phase essential service	Langaris and Dimitriou (2010), Kim *et al*. (2012), Dudina (2013)
Multi-optional phase service	Jain and Upadhyaya (2010), Lakshmi and Ramnath (2014)

Table 1.2: Contributions to various control policy for retrial queueing models

Type of control policy	Research Contributions
N-policy	Choudhury *et al.* (2009), Liu *et al.* (2009)
Threshold based recovery	Efrosinin and Winkler (2011), Yang *et al.* (2013)
Bernoulli admission mechanism	Artalejo and Atencia (2004), Artalejo *et al.* (2005), Choudhury (2007), Choudhury and Deka (2013)
F-policy	Wang *et al.* (2007a), Wang *et al.* (2008a), Wang and Yang (2009)

1.5.8 Retrial Queueing Models with Variant Control Policies

Many prominent researchers contributed to the study of retrial queueing models under different control policies. The implementation of one or more control policies acts as the golden rule so as to reduce the congestion in the system. The optimal control models find a special importance in many real life scenarios; therefore several authors have contributed in this area. The optimal control of a queueing system with set up costs has been done by Nobel and Tijms (2000). The pioneering works on various control policies for retrial queueing systems have been summarized in table 1.2.

1.5.9 Finite Retrial Queueing Models

With the growth of queueing literature and its applications, scientists also realized the need of more specific models for finite queueing systems. The systems with ***finite capacity*** and ***finite population*** have been studied by a few researchers in recent past. Such models find enormous applications in machine repair problems, hospitals, institutes etc. where either the calling population or capacity of the system is finite. Ramalhoto and Gomez-Corral (1998), Falin (1999), Alfa and Isotupa (2004), Almasi *et al.* (2005), Zhong *et al.* (2007, 2008), Sharma and Karmeshu (2009) etc. analyzed finite retrial queues. Zhang and Wang (2012) did stochastic analysis of a finite source retrial queue with orbital search. Shin and Moon (2013) investigated M/M/s/K retrial queue with non-persistent customers. Ponomarov and Lebedev (2013) studied the finite source retrial queue with state dependent service rates.

1.5.10 Methodological Perspective

In this sub-section, we present literature review for the analysis of retrial queueing systems based on the methodological aspects which form the basis of our analysis used in this thesis. Various techniques have been used for the modeling and solution of retrial queueing systems. However, some contributions which are important from methodological view point are listed in table 1.3.

1.5.11 Retrial Queueing Models Applicable in Call Centres and Cellular Radio Networks

The concept of retrial queues is not only limited to the theoretical investigations and mathematical modeling, but finds a very significant and prestigious place in various real life congestion situations. The working of call centres can be interpreted as an example of retrial queues. We may also claim that the mathematical version of retrial queues came into existence after the development of telecommunication systems, where the waiting calls or voice beeps like *ring back*. The significant contributions by various prominent researchers in the area of call centres and cellular radio networks in recent years are mentioned in table 1.4.

Table 1.3: Contributions on retrial queueing models using different techniques

Technique used	Research Contributions
Probability Generating Function	Atencia and Moreno (2006), Krishna Kumar *et al.* (2010), Jain *et al.* (2011)
Supplementary Variable Technique	Wang *et al.* (2001), Choudhury and Deka (2008), Ke and Chang (2009a), Wu and Yin (2009), Chang and Ke (2009), Liu *et al.* (2009), Choudhury and Deka (2009), Choudhury *et al.* (2010), Dimitriou and Langaris (2010), Jain and Upadhyaya (2012), Wu and Lian (2013a)
Embedded Markov Chain Technique	Atencia and Moreno (2005), Wenhui (2005), Choudhury (2007), Choudhury (2008b), Boualem *et al.* (2009), Falin (2010a, 2010b), Wu and Lian (2013b), Choudhury and Deka (2013), Choudhury and Ke (2014), Gao and Wang (2014)
Matrix Geometric Method	Neuts (1978, 1981), Zhang and Tian (2003), Li and Tian (2007), Tian *et al.* (2008), Lin and Ke (2009), Jain and Jain (2010) , Jain *et al.* (2010), Luh (2010), Bhargava and Jain (2014), Lakshmi and Ramanath (2013)
Maximum Entropy Approach	(2007b), Wang and Huang (2009), Wang *et al.* (2011), Jain *et al.* (2012a).

Table 1.4: Contributions to applications of retrial queueing models

Application area	Research Contributions
Call centres	Bernett *et al.* (2002), Bhulai and Koole (2003), Deslauriers (2007), Begin *et al.* (2010), Phung-Duc and Kawanishi (2011), Artalejo and Phung-Duc (2013), Phung-Duc and Kawanishi (2014)
Cellular mobile networks	Tran-Gia and Mandjes (1997), Choi *et al.* (1999), Marsan *et al.* (2001), Trivedi *et al.* (2003), Roszik *et al.* (2005), Liu and Fapojuwo (2006), Dharamraja *et al.* (2008), Brandwajn and Begin (2009), Economou and Herrero (2009), Xu *et al.* (2009), Phung-Duc *et al.* (2009), Wang and Luh (2011), Do (2010, 2011b), Karmeshu and Khandelwa (2013), Kajiwara and Phung-Duc (2014)

1.6 OBJECTIVE OF THE THESIS

Modeling and analysis of retrial queueing models is significant not only in mathematical terms but also to study practically the effect of various sensitive parameters on the performance measures of the system. Enormous literature is available on the performance analysis of retrial queueing models. But still much more work can be done in this direction because of change in technology and its ever growing day-to-day as well as industrial applications in teletraffic, computer and communication networks, etc. The primary goal of our research work is the modeling and analysis of more practical and general retrial models which are consistent enough to deal with more realistic congestion situations. Using various queueing techniques, a variety of retrial problems has been studied. All the retrial queueing models are developed under the assumption of unreliable server so as deal with more realistic situations. The main objective of present thesis work is to develop new retrial models which are applicable to real life congestion situations and incorporate a number of features altogether. Some of the retrial models investigated are as follows:

Queues with discouragement: Discouragement is a common phenomenon of any individual who seeks for the service and it affects the queue length and the efficiency of the system. It is realized that the limited literature is available on retrial queues with reneging. Therefore, to analyze retrial queues with impatient customers, it is worthwhile to study how impatient (i.e. reneging) behavior affects the performance of the system in various situations dealing with retrial attempts of the customers. In chapters 2, 3, 4 and 6, the concept of discouragement of the customers along with noble other features like Bernoulli vacation schedule, modified vacation policy, negative customers etc. have been taken into account.

Threshold recovery for unreliable queues: It is noticed that an extensive work has been done in the direction of retrial queues considering the reliable of the server. But every server is not ideally reliable; therefore there is a need to develop more realistic models dealing with the unreliability of the server. The unreliable servers may breakdown during any course of service and hence need to be repaired. Since repair of the server is an essential component of the server, therefore in order to provide service to the customers there is a need to develop models that can provide repair in an optimal manner. The concept of threshold based recovery can be used to provide repair of the broken down server in the optimal sense. In this case, the repairman is usually called upon when a threshold number of customers are already accumulated in the system so as to save time and money. In chapters, 8 and 10, the repair process has been taken into account using threshold based recovery and optimal parameters have been obtained wherever possible.

Control policy queues: It is evident from the literature that a limited work has been done in the direction of retrial queues with F-policy; therefore we have developed the Markovian models operating under F-policy to analyze retrial queues so as to design optimal control policies by constructing the cost function. Chapter 8 deals with the finite retrial queue model using F-policy and threshold recovery.

Phase type queues: There is a limited literature available on the retrial queues with phase type models wherein either service or repair or both are processed in series. Therefore, from modeling point of view there is a need to frame new models which take into consideration the feature of multi-phases service/repair along with other features namely vacation, discouragement, unreliable, bulk, set up and many more simultaneously. In chapters 3-6, the modeling of retrial queues is done by incorporating either multi-phase service or phase repair or both.

To analyze retrial queues in the broader sense, we have also considered the concept of finite *double orbits* instead of single orbit. Moreover, retrial model with additional server has also been developed so as to study the effect of additional server on the performance of the system.

The research work under consideration is not only limited to the modeling of new models incorporated with a number of practical features altogether; it is also aimed to analyze and provide analytical solution for these models. Stochastic processes namely markovian and non-markovian processes are used for the problem formulation. The solutions of mathematical models on retrial queueing system have been obtained by employing one or more of the following methods:

- Generating Functions
- Embedded Markov Chain
- Supplementary variables technique
- Matrix method
- Matrix geometric method
- Maximum Entropy Principle

In our study, we have been interested to derive various performance measures which are important enough to judge the efficiency and validity of the model to practical situations. Some important measures evaluated are queue size, long run probabilities, expected queue length, throughput, carried load, server's utilization, reliability/ availability, failure frequency, total cost incurred, etc. To validate the analytical results, some numerical examples and graphs are facilitated. The numerical simulation and sensitivity analysis are also carried out which will be helpful in examining the effect of various parameters on the system performance.

1.7 OVERVIEW OF THE THESIS

The modeling and analysis of retrial queueing systems in different frameworks consistent with various real life scenarios is the main objective of present research work. Retrial queueing models are successfully used in abundant congestion problems ranging from day-to-day to many industrial scenarios. Therefore, keeping in mind the significance of retrial queues a variety of queueing problems with retrial attempts are explored using suitable approaches. The present thesis can be broadly classified into two categories; non-markovian retrial queues and markovian queues. The whole thesis work has been structured into ten chapters. Chapters 2-7 deal with non-markovian retrial queues while chapters 8-10 are devoted to markovian retrial queues. For solution purpose, various methodologies like supplementary variable technique, embedded markov chain, matrix geometric method, matrix method, generating function, R-K method etc. are used. To validate the analytic results obtained in various chapters, numerical illustrations are also given for the better understanding of real life queueing problems where retrial attempts are common feature of the system. For some models dealing with complex problems the approximate results of various system performance indices have been obtained using matrix entropy principle (MEP). For illustration purpose we have coded the computer programs in MATLAB software so as to provide performance measures of retrial queueing models. The chapter wise brief outlines of the thesis work are as follows:

In this ongoing ***chapter 1,*** entitled Introduction, we present the overview of the works done, methodology and some preliminaries concepts related to retrial queues. The related literature has been briefly discussed by classifying the retrial queues based on modeling and methodological aspects.

In ***chapter 2*** we investigate $M^x/G/1$ retrial queue with unreliable server and general retrial times. The server is subject to breakdowns and takes some setup time before starting the repair. The server renders first essential phase of service (FES) to all the arriving customers whereas second optional phase services (SOS) are provided after

FES to only those customers who opt for it. The impatient customers are allowed to balk depending upon the server's status; they may also renege after waiting sometime in the queue. By incorporating the supplementary variables corresponding to service time, repair time, retrial time and setup time and using generating function method, the queueing analysis has been done to obtain the queue size and orbit size distributions. Using maximum entropy approach, a comparative analysis has been performed between exact analytic results and those obtained by using maximum entropy approach.

In ***chapter 3*** a batch arrival general retrial queue with multioptional services, vacation and impatient customers undergoing reneging has been considered. The study of chapter 2 is extended in chapter 3 by incorporating the feature of phase repair and Bernoulli vacation schedule. The server may go for vacation if he finds no customer waiting for the service and returns back to the service center again as soon as the customer approaches for the service. The server is unreliable and subject to breakdowns during the service; as soon as the server fails, it is immediately sent for the repair so as to restore its functionality as before failure. The repairman assigned for the repair of the server also takes some setup time before commencing the repair process. We employ supplementary variables technique and probability generating function method to obtain the explicit expressions for the queue size distribution and other performance measures. Also, the neuro fuzzy approach has been used to approximate the analytical results.

Bulk arrival M/G/1 retrial queue with impatient customers and modified vacation policy has been analysed in ***chapter 4***. The service is provided in k essential phases to all the customers by the single server which may breakdown while rendering service to the customers. The broken down server is sent to a repair facility wherein the repair is performed in d compulsory phases. As soon as the orbit becomes empty, the server goes for vacation and takes at most J vacations until at least one customer is noticed. The incoming customers are impatient and may renege on seeing a long queue of the customers for the service. The probability generating functions and queue length for the number of customers in the orbit and queue have been obtained using supplementary variable technique. Various system characteristics viz. average number of customers in the queue and orbit, long run probabilities of the system states etc. are obtained. Using maximum entropy approach, a comparative analysis has been performed between exact analytic results and that obtained by using maximum entropy approach. The effects of several parameters on the system performance are examined numerically by taking an illustration.

The steady state analysis of bulk arrival retrial queue with unreliable server and multi essential services has been taken into account in ***chapter 5***. The server renders service to two types of customers; the type 1(2) customers are considered as priority (ordinary) customers. The ordinary customers are forced to join the orbit if they find the server in busy or broken down condition on their arrival whereas priority customers join the queue in front of the server so as to get served. The service is provided in *k* essential phases for both types of customers. The server is unreliable and may break down during any phase of service. As soon as the server fails it is sent for repair to restore it so as to make it as good as before failure. The broken down server is repaired in *d* essential phases. Also, the customers are affected by the traffic and may balk at seeing the long queue. In the present investigation, the supplementary variable technique and the method of generating function have been used to derive the explicit expressions for the average queue length of the system. Moreover, application of the model to healthcare system has also been discussed.

A bulk arrival retrial queue with negative customers and multi-services subject to server breakdowns has been considered in ***chapter 6.*** The system allows the arrival of two types of customers; positive customers and negative customers in the system. The negative customers make the server fail if they find the server in busy state, whereas positive customers are served; otherwise if the server is idle they join the virtual pool of customers called orbit. The customers from the retrial orbit try their chance again for the service. The customers have the option of obtaining more than one service. Moreover, the customers are impatient and may renege from the system with probability ‘r’. The server is sent for the repair as soon as it breakdowns; after repair, the service process starts again. Also, the server has the provision to initiate the service when there are N customers accumulated in the system. Using supplementary variables technique and generating functions, various performance measures like reliability and queueing indices have been obtained.

In ***chapter 7*** the steady-state behavior of a batch arrival retrial queue with multioptional services and phase repair under Bernoulli vacation schedule is studied. The customers arrive in batches and are admitted to join the system following Bernoulli admission control policy. The incoming customers are forced to join the retrial group if they find the server busy, broken down or on vacation. The customers are served in two phases i.e. the first essential service (FES) followed by second optional services (SOS). The server is unreliable and is repaired in *d*- compulsory phases so as to become as good as earlier. After each service completion, the server may go for a vacation following Bernoulli vacation schedule or continue to serve the next customer. By applying the embedded Markov chain method, we first obtain the ergodicity condition for the stability of the system and then obtain steady-state results to examine some queueing measures.

Chapter 8 deals with two finite retrial queueing models with threshold recovery. The first model deals with the finite capacity Markovian retrial queues with unreliable server wherein the customers arrive following geometric distribution while the service pattern follows exponential distribution. The customer occupies the server if it is idle, otherwise he is forced to join the orbit and retry for the service later. The customers are served in two stages i.e. first essential service (FES) and second optional service (SOS) which depends on the customer's demand. The repair process follows threshold recovery according to which the repair starts when a minimum number of customers say q (≥ 1) has been accumulated in the system. The transient state solution of the equations governing the model has been obtained using Runge Kutta method of fourth order.

The second model is concerned with the finite capacity retrial queueing model with F-policy. The server is unreliable and may break down while providing service to the customers. The failed server is sent to the repair facility where after required setup time, the repair is done as per pre-specified rule known as threshold recovery policy for the repair. The arrival of the customers to the system is controlled by using F-policy. The numerical approach based on Runge Kutta method of fourth order has been employed to study the transient behavior of the system. Various performance measures like expected queue length, waiting time, failure frequency, availability, throughput, etc. have been obtained. The cost optimization and sensitivity analysis have been done to explore the effect of different parameters on various performance indices.

The unreliable server retrial queue with the provision of additional temporary server in the context of application in web faction has been investigated in ***chapter 9***. In order to reduce the load consumption and memory usage, a temporary server is usually installed when primary server is overloaded. The secondary server which is temporary one, is turned on when the work load reaches its maximum value i.e., a fixed queue length of 'K' customers including the customer with the primary server, has been build up. The system has the facility of retrial orbit where customers can wait for their service when they find the server in busy state. Using matrix geometric approach, we determine the steady probabilities of the system states. The cost function has also been structured to determine the optimal cost of running the system.

In ***chapter 10***, double orbit finite capacity retrial queue with unreliable server has been taken into consideration. The system facilitates the arrival of two type of customers known as priority and non priority customers and can hold a maximum of L priority customers and K non-priority customers as per its capacity. The priority customers are served prior to the non-priority customers. Moreover, the server is unreliable which may breakdown while servicing either priority or non-priority customer. The failed server is sent for repair following threshold recovery policy to become as good as earlier. Both transient as well as steady state analysis of the model has been done using matrix method. The application of the model to cellular radio network has been discussed by taking an illustration.

Overall conclusions and future scope of the models investigated has been presented at the end of the thesis to highlight the contributions of the carried out research works and its importance to real life congestion situations.

1.8 CONCLUDING REMARKS

Retrial queues due to their numerous applications in manufacturing processes, industries, production systems, telecommunication systems have forced queue theorists to develop new models which can be suited to real life situations. The literature on retrial queues has widely grown since past few decades which clearly exhibit their significance in research due to their abundance applications. In the present chapter, a brief account about the modeling and techniques used for the solution of retrial queues has been presented.

The ongoing chapter provides an overview of the researches that took place in the field of retrial queueing models in different frameworks. Numerous research papers have been cited on the basis of modeling as well as methodological concepts for the analysis of concerned retrial models. The design, development and configuration of retrial queueing systems can be well understood in terms of various performances measures viz., queue size distribution, long run probabilities, average queue length, etc. The queueing analysis of such systems can provide valuable insight to the system designers and decision makers for the improvement and enhancement of retrial queueing models studied in different frameworks by incorporating more realistic features.

CHAPTER 2

UNRELIABLE RETRIAL QUEUE WITH IMPATIENT CUSTOMERS

2.1 INTRODUCTION

Retrial queueing models are significantly used for the performance analysis of many telecommunication processes, including local and wide area networks, switching systems, shared bus local area networks, etc. This is due to the fact that the return of customer is usually a non-neglectable part in many practical situations. An extensive survey on retrial queues can be found in the notable survey articles by Artalejo (1999 a, b), Artalejo and Falin (2002) and Artalejo (2010).

In the real life congestion situations, the server may be unreliable and easily prone to breakdowns. Krishna Kumar and Madheshwari (2003) investigated $M^x/G/1$ retrial queueing model with starting failures. Atencia *et al.* (2006a) studied M/G/1 retrial queue with active breakdowns and Bernoulli schedule. Mokaddis *et al.* (2007) considered the M/G/1 retrial queue with Bernoulli feedback and single vacation where the server is subject to starting failures. In many real time unreliable server queueing situations, the server may take some time called setup time to start the repair. Jain *et al.* (2007) studied the M/G/1 retrial queueing model with set up, server breakdown and repair. Xu *et al.* (2009) obtained distribution for the additional queue length for M/M/1 queue with working vacation and set up times.

In recent past, queues with an optional second phase of service have also attracted the attention of many researchers working in the field of queueing theory. In such queueing scenario, after completing the first essential phase of service (FES), the customer has the option either to go for any of the secondary services of his choice (SOS) provided by the same server or quits the system. Madan (2000) studied an M/G/1 queue with second optional service with service time of FES being governed by general distribution whereas the second optional service is exponentially distributed. The $M^x/G/1$ unreliable retrial queue with two phases of service and Bernoulli admission mechanism was further explored by Choudhury and Deka (2009). Jain and Upadhyaya (2010) investigated $M^x/G/1$ queue with multi-optional services and Bernoulli vacation. Moreover, Senthil and Arumuganathan (2010) analysed bulk arrival retrial queues with optional services.

The retrial queueing models with bulk arrival are always in demand as they depict many practical queueing situations in a better way; to cite we refer a group of people arriving at restaurants. Jain and Bhargava (2009a) studied bulk arrival retrial queue with unreliable server and priority subscribers. Chang and Ke (2009) considered a batch retrial model where the server can take at most J vacations. Maraghi *et al.* (2010) investigated bulk arrival queues with Bernoulli vacation. Other work in this series has been done by Choudhury *et al.* (2010) by considering a batch arrival retrial queueing system with two phases of service and server interruption. Moreover, Balasubramanian and Arumuganathan (2011) discussed steady state analysis of a bulk arrival and general bulk service queueing system. Recently, Choudhury and Ke (2012) analyzed a batch arrival retrial queue with delaying repair and Bernoulli vacation schedule.

Due to numerous applications in computer communication systems, the retrial queues with impatient customers have now-a-days become the point of attraction for the queue theorists. Sometimes, the customers may get discouraged on seeing the long queue ahead, therefore either the customers would not join the queue i.e. would like to balk or leave the queue after waiting for some time i.e. renege. Both reneging and balking can be visualized in day-to-day congestion situations where some impatient customers quit the system without getting the service; such a scenario is prevalent at supermarkets, reservation counters, call centres etc. Ke and Chang (2009b) investigated M/G/1 retrial queue with modified vacation policy by incorporating balking and reneging concepts. Using supplementary variable technique, Arrar *et al.* (2012) investigated asymptotic behaviour of M/G/1 retrial queues with batch arrivals and impatience phenomenon.

The maximum entropy principle (MEP) was introduced by Shannon to study the problems of information theory as the measurement of uncertainty. This principle is applicable to select the appropriate probability distributions for the queueing situation. It was also used by many researchers to obtain the queue size distribution of various queueing systems. Wang *et al.* (2002) used the maximum entropy principle to examine the M/G/1 queueing system in different frameworks. Further, Wang *et al.* (2007b) carried out the maximum entropy analysis of M^x/M/1 queueing system with multiple vacations. Moreover, Wang and Huang (2009) made a comparative study between the exact analytical results and approximate results obtained using MEP. Maximum entropy principle has also been used for discrete time unreliable server queue with working vacation by Jain *et al.* (2012a).

The present chapter aims to investigate the retrial queues with impatient customers and batch arrival. The results of Wang and Li (2009) have been extended by incorporating the concept of bulk arrival and setup time. In this chapter, we apply the MEP to analyze the various system characteristics for a bulk arrival retrial queueing model with second optional services, balking, reneging and setup times. Section 2.2 describes the assumptions required to formulate the model. The generating functions of the queue size distribution are obtained in section 2.3. Various performance indices are computed in section 2.4. Section 2.5 deals with special cases deduced from our analytical results by setting different parameters. In section 2.6, the approximate results for the various performance measures have been obtained by implementing the principle of maximum entropy. A comparative analysis is performed between the exact results and the approximate result obtained by using maximum entropy principle. To validate the analytical results, numerical results are presented in section 2.7. Finally, in section 2.8 we wind up our investigation by highlighting the noble features of the work done.

2.2. MODEL DESCRIPTION

Consider $M^X/G/1$ single server retrial queueing system with impatient customers. The following assumptions have been made to formulate the mathematical model to be investigated:

- **Arrival Process:** The customers arrive in batches according to the Poisson process with rate λ. Let X be the random variable denoting the batch size defined by $\Pr\{X = m\} = c_m;\ m \geq 1$ such that $\sum_{m=1}^{\infty} c_m = 1$.
- **Retrial Process**: The incoming customers are served if they find the server idle otherwise they are forced to join the virtual pool of customers called orbit from where they can try again for the service. The customers waiting in the retrial orbit are known as retrial customers and they retry after a random interval of time with exponential distributed rate γ.
- **Reneging:** In context of the retrial customer, it can be explained as follows. As soon as the server becomes free, both primary as well as orbit customers try for the service. If a primary customer arrives earlier as compared to the retrial customer, then either retrial customer cancels its attempt for the service and returns to its initial position with probability r or quits the system forever with probability (1-r).
- **Service Process:** If an incoming batch of the customers finds the server in idle state, then a customer at the head of the batch joins the server to get served. The customers are served in two stages, the first essential service (FES) with service rate μ_0 which is

compulsory for all the customers and second optional services (SOS) with service rate μ_i $(1 \le i \le k)$, which depends on the customer's choice to avail it or not. After completing FES, the customer may opt for any one of the 'k' different optional services with probability p_i $(1 \le i \le k)$ or quits the system with probability $p_0 = (1 - \sum_{i=1}^{k} p_i)$. The service times of FES and all phases of SOS are i.i.d. and general distributed.

- **Balking:** Sometimes, the customer may get discouraged on seeing a long queue and may decide to leave the queue without joining it i.e. balks with probability b_1, b_2 and b_3 in case for the server being in busy state, breakdown state, setup state, respectively.
- **Breakdown and Setup Process:** The server under consideration is unreliable which can breakdown during any course of service. The server's lifetime is exponentially distributed with mean α_0^{-1} during FES and α_i^{-1} during SOS of i^{th} $(1 \le i \le k)$ type. But before starting the repair process, the server takes some time called setup time with rate ξ_i $(1 \le i \le k)$ to make some preliminary settings i.e. there is delay-in-repair. The setup time and repair time are i.i.d. and general distributed.
- **Repair Process:** The repair process is assumed to be i.i.d. and general distributed and is completed with rate β_i for the server broken during i^{th} $(0 \le i \le k)$ phase of service.

2.3 QUEUE SIZE DISTRIBUTION

To analyze the retrial queueing system, we need to construct the mathematical equations for the system state probabilities. The retrial process, service process and repair process are assumed to be general distributed; therefore the model under consideration is non-markovian. In order to formulate the equations for the present non-markovian system, we use supplementary variable technique to analyze the model.

Let N(t) represents number of customers in the system and $S_1(t) \in \{0,1,2,...,k\}$ denotes the phase of the service at any time t.

The state of the server at any time t is given by

$$Y(t) = \begin{cases} 1, & \text{server is in idle state} \\ 2, & \text{server is busy in providing FES to the customers} \\ 3, & \text{server is busy in providing SOS to the customers} \\ 4, & \text{server is broken down and under setup before repair} \\ 5, & \text{server is broken down and under repair} \end{cases}$$

In the steady state, the joint distributions of the server state and queue size are defined as-

$$D_n = \lim_{t\to\infty} \Pr\{Y(t)=1, N(t)=n\}, n\geq 0$$

$$P_{0,n}(x) = \lim_{t\to\infty} \Pr\{Y(t)=2, x\leq \varpi(t)\leq x+dx, N(t)=n, 0, S_1(t)=0\}, n\geq 0$$

$$P_{i,n}(x) = \lim_{t\to\infty} \Pr\{Y(t)=3, x\leq \varpi(t)\leq x+dx, N(t)=n, S_1(t)=i\}, n\geq 0, (1\leq i\leq k)$$

$$S_{i,n}(x,y) = \lim_{t\to\infty} \Pr\{Y(t)=4, \varpi(t)=x, y\leq \sigma(t)\leq y+dy, N(t)=n, S_1(t)=i\},$$
$$n\geq 0, (0\leq i\leq k)$$

$$R_{i,n}(x,y) = \lim_{t\to\infty} \Pr\{Y(t)=5, \varpi(t)=x, y\leq \sigma(t)\leq y+dy, N(t)=n, S_1(t)=i\},$$
$$n\geq 0, (0\leq i\leq k)$$

2.3.1 Mathematical Formulation

Before constructing the governing equations, we give below the proposition stating the stability condition for the model.

Proposition 2.1: The necessary and sufficient condition for the system to be stable is $\pi < 1 - r(1-\tilde{a}(\lambda))$

where, $\pi = C'(1)\left[\rho_0\left(b_1+\alpha_0\left(\frac{b_3}{\xi_0}+\frac{b_2}{\beta_0}\right)\right)+\sum_{i=1}^{k}\rho_i p_i\left(b_1+\alpha_i\left(\frac{b_3}{\xi_i}+\frac{b_2}{\beta_i}\right)\right)\right]$.

Proof: To study the steady state behaviour of the system, we require the stability condition before formulating the governing equations. Wang *et al.* (2001) discussed the proof for the establishment of stability condition for M/G/1 model. Following the same approach, we derive stability condition for our model as-

$$1-r(1-\tilde{a}(\lambda)) > C'(1)\left[\rho_0\left(b_1+\alpha_0\left(\frac{b_3}{\xi_0}+\frac{b_2}{\beta_0}\right)\right)+\sum_{i=1}^{k}\rho_i p_i\left(b_1+\alpha_i\left(\frac{b_3}{\xi_i}+\frac{b_2}{\beta_i}\right)\right)\right]$$

Governing Equations

$$\left[\frac{d}{dx}+\lambda b_1+\alpha_0+\mu_0(x)\right]P_{0,n}(x) = \lambda b_1\sum_{m=1}^{n} c_m P_{0,n-m}(x) + \int_0^\infty R_{0,n}(x,y)\beta_0(y)dy \tag{2.1}$$

$$\left[\frac{d}{dx}+\lambda b_1+\alpha_i+\mu_i(x)\right]P_{i,n}(x) = \lambda b_1\sum_{m=1}^{n} c_m P_{i,n-m}(x) + \int_0^\infty R_{i,n}(x,y)\beta_i(y)dy \tag{2.2}$$

$$\lambda D_n(0) = p_0\int_0^\infty P_{0,0}(x)\mu_0(x)dx + \sum_{i=1}^{k}\int_0^\infty P_{i,0}(x)\mu_i(x)dx, (1\leq i\leq k) \tag{2.3}$$

$$\left[\frac{d}{dw}+\lambda+\gamma(w)\right]D_n(w) = 0; \qquad n\geq 1 \tag{2.4}$$

$$\left[\frac{\partial}{\partial y}+\lambda b_2+\beta_0(y)\right]R_{0,n}(x,y)=\lambda b_2\sum_{m=1}^{n}c_m R_{0,n-m}(x,y), \quad n\geq 0 \tag{2.5}$$

$$\left[\frac{\partial}{\partial y}+\lambda b_2+\beta_i(y)\right]R_{i,n}(x,y)=\lambda b_2\sum_{m=1}^{n}c_m R_{i,n-m}(x,y), \quad n\geq 0 \tag{2.6}$$

$$\left[\frac{\partial}{\partial y}+\xi_0(y)+\lambda b_3\right]S_{0,n}(x,y)=\lambda b_3\sum_{m=1}^{n}c_m S_{0,n-m}(x,y), \quad n\geq 0 \tag{2.7}$$

$$\left[\frac{\partial}{\partial y}+\xi_i(y)+\lambda b_3\right]S_{i,n}(x,y)=\lambda b_3\sum_{m=1}^{n}c_m S_{i,n-m}(x,y), \quad n\geq 0 \tag{2.8}$$

Boundary Conditions

$$D_n(0)=p_0\int_0^\infty P_{0,n}(x)dx\mu_0(x)dx+\sum_{i=1}^{k}\int_0^\infty P_{i,n}(x)\mu_i(x)dx, \quad n\geq 1 \tag{2.9}$$

$$P_{0,0}(0)=\int_0^\infty D_1(w)\gamma(w)dw+(1-r)\lambda\int_0^\infty D_1(w)dw+\lambda D_0 \tag{2.10}$$

$$P_{0,n}(0)=\int_0^\infty D_{n+1}(w)\gamma(w)dw+(1-r)\lambda\int_0^\infty D_{n+1}(w)dw+r\lambda\int_0^\infty D_n(w)dw, \quad n\geq 1 \tag{2.11}$$

$$P_{i,n}(0)=p_i\int_0^\infty P_{0,n}(x)\mu_0(x)dx \qquad n\geq 1 \tag{2.12}$$

$$R_{0,n}(x,0)=\int_0^\infty S_{0,n}(x,y)\xi_0(y)dy \tag{2.13}$$

$$R_{i,n}(x,0)=\int_0^\infty S_{i,n}(x,y)\xi_i(y)dy, \quad n\geq 1 \quad (1\leq i\leq k) \tag{2.14}$$

$$S_{0,n}(x,0)=\alpha_0 P_{0,n}(x) \tag{2.15}$$

$$S_{i,n}(x,0)=\alpha_i P_{i,n}(x), \qquad n\geq 1,\ (1\leq i\leq k) \tag{2.16}$$

Normalizing Condition

$$D_0+\sum_{n=1}^{\infty}\int_0^\infty D_n(w)dw+\sum_{n=0}^{\infty}\sum_{i=0}^{k}\int_0^\infty P_{i,n}(x)dx+\sum_{n=0}^{\infty}\sum_{i=0}^{k}\int_0^\infty\int_0^\infty R_{i,n}(x,y)dxdy+\sum_{n=0}^{\infty}\sum_{i=0}^{k}\int_0^\infty\int_0^\infty S_{i,n}(x,y)dxdy=1$$

(2.17)

2.3.2 Probability Generating Function

We use probability generating functions (PGF) corresponding to different states of the server to solve the set of differential difference equations so as to obtain the steady state solution of the retrial queueing model. The probability generating functions corresponding to different states are defined as follows:

Retrial state : $D(w,z)=\sum_{n=1}^{\infty} D_n(w)z^n\,;|z|\le 1$

Busy state : $P_i(x,z)=\sum_{n=0}^{\infty} P_{i,n}(x)z^n;\ |z|\le 1$

Repair state : $R_i(x,y,z)=\sum_{n=0}^{\infty} R_{i,n}(x,y)z^n\,;|z|\le 1$

Set up state : $S_i(x,y,z)=\sum_{n=0}^{\infty} S_{i,n}(x,y)z^n;\ |z|\le 1$

Batch size : $C(z)=\sum_{m=1}^{\infty} c_m z^n\ ;|z|\le 1$

The hazard rates corresponding to the different states are as follows:

Retrial state	$\gamma(w)=\dfrac{a(w)}{1-A(w)}$
Busy state for the server being in ith $(1\le i\le k)$ phase of service	$\mu_i(x)=\dfrac{b_i(x)}{1-B_i(x)}$
Repair state for the server broken during ith $(1\le i\le k)$ phase of service	$\beta_i(y)=\dfrac{g_i(x)}{1-G_i(x)}$
Set up state before repair process for the server broken during ith $(1\le i\le k)$ phase of service	$\xi_i(y)=\dfrac{\eta_i(y)}{1-N_i(y)}$

Now, we give our results for the partial generating functions and marginal generating functions for the different states of the system in the form of theorems as follows:

Theorem 2.1: At random epochs, the partial probability generating functions of the joint distribution of the server being in idle state, FES state, i^{th} $(1\le i\le k)$ SOS busy state, under repair while broken down states during FES and i^{th} $(1\le i\le k)$ SOS, during setup while broken down during FES and i^{th} $(1\le i\le k)$ SOS, respectively are given by

$$D(w,z)=D(0,z)\exp\{-\lambda w\}\bar{A}(w) \tag{2.18}$$

$$P_0(x,z)=P_0(0,z)\exp\{-H_0(z)x\}\overline{B_0}(x) \tag{2.19}$$

$$P_i(x,z)=p_i\tilde{b}_0(H_0(z))P_0(0,z)\exp\{-H_i(z)x\}\overline{B_i}(x) \tag{2.20}$$

$$R_0(x,y,z)=\alpha_0 P_0(x,z)\tilde{\eta}_0\{\lambda b_3(1-C(z))\}\exp\{-\lambda b_2 y(1-C(z))\}\bar{G}_0(y) \tag{2.21}$$

$$R_i(x,y,z)=\alpha_i P_i(x,z)\tilde{\eta}_i\{\lambda b_3(1-C(z))\}\exp\{-\lambda b_2 y(1-C(z))\}\bar{G}_i(y) \tag{2.22}$$

$$S_0(x,y,z)=\alpha_0 P_0(x,z)\exp\{-\lambda b_3 y\left(1-C(z)\right)\}\bar{N}_0\left(y\right) \tag{2.23}$$

$$S_i(x,y,z)=\alpha_i P_i(x,z)\exp\{-\lambda b_3 y\left(1-C(z)\right)\}\bar{N}_i\left(y\right) \tag{2.24}$$

where,

$$H_i\left(z\right)=\lambda b_1\left(1-C(z)\right)+\alpha_0(1-M_i(z)),\ (0\le i\le k) \tag{2.25}$$

$$M_i(z)=\tilde{\eta}_i\{\lambda b_3(1-C(z))\}\tilde{g}_i(\lambda b_2(1-C(z))),\quad (0\le i\le k) \tag{2.26}$$

$$\text{D }(0,\text{ z})=\frac{\lambda D_0 z[1-p_0\tilde{b}_0(H_0(z))-\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))]}{[p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))][1+r(z-1)(1-\tilde{a}\ (\lambda))]-z} \tag{2.27}$$

$$P_0(0,z)=\frac{\lambda D_0(1-z)(1-r+r\tilde{a}(\lambda))}{[p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))][1+r(z-1)(1-\tilde{a}\ (\lambda))]-z} \tag{2.28}$$

$$D_0=\frac{1-\pi-r(1-\tilde{a}(\lambda))}{(1-r+r\tilde{a}(\lambda))\left[1+\rho_0\left(1+\alpha_0\left(\frac{1}{\xi_0}+\frac{1}{\beta_0}\right)\right)+\sum_{i=1}^{k}\rho_i p_i\left(1+\alpha_i\left(\frac{1}{\xi_i}+\frac{1}{\beta_i}\right)\right)\right]-\pi\tilde{a}(\lambda)} \tag{2.29}$$

Proof: Multiplying eqs (2.1)-(2.16) by the appropriate powers of z and summing over n=0, 1, 2, 3, 4,... and then solving, we get eqs (2.18)-(2.28). Here, D_0 can be determined by using normalizing condition (2.17).

Theorem 2.2: The marginal probability generating functions at random epochs when the server is in idle state, busy with ith $(0\le i\le k)$ phase service, under repair while breakdown during ith $(0\le i\le k)$ phase service, under set up state while broken down during FES and i^{th} $(1\le i\le k)$ SOS, respectively are given by

$$\text{D(z)}=\frac{D_0 z[1-p_0\tilde{b}_0(H_0(z))-\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))](1-\tilde{a}(\lambda))}{[p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))][1+r(z-1)(1-\tilde{a}(\lambda))]-z} \tag{2.30}$$

$$P_0(z)=\frac{\lambda D_0(1-z)(1-r+r\tilde{a}(\lambda))}{p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))][1+r(z-1)(1-\tilde{a}(\lambda))]-z}\frac{[1-\tilde{b}_0(H_0(z))]}{H_0(z)} \tag{2.31}$$

$$P_i(z)=\frac{\lambda p_i\tilde{b}_0(H_0(z))D_0(1-z)(1-r+r\tilde{a}(\lambda))}{p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(R_i(z))][1+r(z-1)(1-\tilde{a}(\lambda))]-z}\frac{[1-\tilde{b}_i(H_i(z))]}{H_i(z)} \tag{2.32}$$

$$R_0(z)=\frac{\alpha_0\lambda D_0(1-z)(1-r+r\tilde{a}(\lambda))}{p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))][1+r(z-1)(1-\tilde{a}(\lambda))]-z}\times\frac{[1-\tilde{b}_0(H_0(z))]}{H_0(z)}\frac{[1-\tilde{g}_0(\lambda b_2(1-C(z)))]}{\lambda b_2(1-C(z))} \quad (2.33)$$

$$R_i(z)=\frac{\alpha_i\lambda p_i\tilde{b}_0(H_0(z))D_0(1-z)(1-r+r\tilde{a}(\lambda))}{p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))][1+r(z-1)(1-\tilde{a}(\lambda))]-z}\times\frac{[1-\tilde{b}_i(H_i(z))]}{H_i(z)}\frac{[1-\tilde{g}_i(\lambda b_2(1-C(z)))]}{\lambda b_2(1-C(z))} \quad (2.34)$$

$$S_0(z)=\frac{\alpha_0\lambda D_0(1-z)(1-r+r\tilde{a}(\lambda))}{[p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))][1+r(z-1)(1-\tilde{a}(\lambda))]-z}\times\frac{[1-\tilde{b}_0(H_0(z))]}{H_0(z)}\frac{[1-\tilde{\eta}_0(\lambda b_3(1-C(z)))]}{\lambda b_3(1-C(z))} \quad (2.35)$$

$$S_i(z)=\frac{\alpha_i p_i\lambda D_0(1-z)(1-r+r\tilde{a}(\lambda))}{[p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))][1+r(z-1)(1-\tilde{a}(\lambda))]-z}\times\frac{[1-\tilde{b}_0(H_i(z))]}{H_i(z)}\frac{[1-\tilde{\eta}_i(\lambda b_3(1-C(z)))]}{\lambda b_3(1-C(z))} \quad (2.36)$$

Proof: The marginal probability generating functions for different states of the server given in equations (2.30)-(2.36) can be determined by using

$$D(z)=\int_0^\infty D(w,z)dw\ ,\ P_0(z)=\int_0^\infty P_0(x,z)dx\ ,\ P_i(z)=\int_0^\infty P_i(x,z)dx \quad (2.37a)$$

$$R_0(z)=\int_0^\infty\int_0^\infty R_0(x,y,z)dxdy\ ,\ R_i(z)=\int_0^\infty\int_0^\infty R_i(x,y,z)dxdy \quad (2.37b)$$

$$S_0(z)=\int_0^\infty\int_0^\infty S_0(x,y,z)dxdy\ ,\ S_i(z)=\int_0^\infty\int_0^\infty S_i(x,y,z)dxdy \quad (2.37c)$$

Theorem 2.3: The generating function for the number of customers in the retrial queue is given by

$$P(z)=D_0[1+\frac{z[1-\varphi(z)][1-\tilde{a}(\lambda)]}{\varphi(z)\psi(z)-z}+\frac{\lambda(1-z)(1-\tilde{b}_0(H_0(z))(1-r+r\tilde{a}(\lambda))\theta_0(z)}{\{\varphi(z)\psi(z)-z\}H_0(z)\lambda b_3(1-C(z))}+\frac{\sum_{i=1}^{k}p_i\lambda(1-z)\tilde{b}_0(H_0(z))(1-\tilde{b}_i(H_i(z))(1-r+r\tilde{a}(\lambda))\theta_i(z)}{\{\varphi(z)\psi(z)-z\}H_i(z)\lambda b_3(1-C(z))}] \quad (2.38)$$

where,

$$\theta_i(z)=[\lambda b_3(1-C(z))+\alpha_i(1-\tilde{\eta}_i(\lambda b_3(1-C(z))))+\alpha_i M_i(z)\lambda b_3(1-C(z))],\ 0\le i\le k$$

$$\varphi(z)=p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))]$$

$$\psi(z)=[1+r(z-1)(1-\tilde{a}(\lambda))]$$

Proof: We have,

$$P(z)=D_0+D(z)+P_0(z)+\sum_{i=1}^{k}P_i(z)+R_0(z)+\sum_{i=1}^{k}R_i(z)+S_0(z)+\sum_{i=1}^{k}S_i(z) \tag{2.39}$$

where,

$$D(z)=\int_0^\infty D(w,z)dw,\ P_0(z)=\int_0^\infty P_0(x,z)dx, \tag{2.40a}$$

$$P_i(z)=\int_0^\infty P_i(x,z)dx,\ R_0(z)=\int_0^\infty\int_0^\infty R_0(x,y,z)dxdy\ \ R_i(z)=\int_0^\infty\int_0^\infty R_i(x,y,z)dxdy, \tag{2.40b}$$

$$S_0(z)=\int_0^\infty\int_0^\infty S_0(x,y,z)dxdy,\ S_i(z)=\int_0^\infty\int_0^\infty S_i(x,y,z)dxdy \tag{2.40c}$$

Also, using (2.39) and (2.40a)-(2.40c), we obtain generating function for the number of customers in the retrial group as given by (2.38).

Theorem 2.4: The generating function for the number of customers in the system is given by

$$L(z)=D_0[1+\frac{z[1-\phi(z)][1-\tilde{a}(\lambda)]}{\phi(z)\psi(z)-z}+\frac{\lambda z(1-z)(1-\tilde{b}_0(H_0(z))(1-r+r\tilde{a}(\lambda))\theta_0(z)}{\{\phi(z)\psi(z)-z\}H_0(z)\lambda b_3(1-C(z))}$$
$$+\frac{\sum_{i=1}^{k}p_i\lambda z(1-z)\tilde{b}_0(H_0(z))(1-\tilde{b}_i(H_i(z))(1-r+r\tilde{a}(\lambda))\theta_i(z)}{\{\phi(z)\psi(z)-z\}H_i(z)\lambda b_3(1-C(z))}] \tag{2.41}$$

where,

$$\theta_i(z)=[\lambda b_3(1-C(z))+\alpha_i(1-\tilde{\eta}_i(\lambda b_3(1-C(z))))+\alpha_i M_i(z)\lambda b_3(1-C(z))],\ 0\le i\le k$$

$$\varphi(z)=p_0\tilde{b}_0(H_0(z))+\sum_{i=1}^{k}p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))],\ \ \psi(z)=[1+r(z-1)(1-\tilde{a}(\lambda))]$$

Proof: The generating function for the number of customers in the system can be obtained by using results from (2.40a)-(2.40c) in the following equation

$$L(z)=D_0+D(z)+zP_0(z)+z\sum_{i=1}^{k}P_i(z)+zR_0(z)+z\sum_{i=1}^{k}R_i(z)+zS_0(z)+z\sum_{i=1}^{k}S_i(z) \tag{2.42}$$

2.4 PERFORMANCE MEASURES

Various performance measures like long run probabilities, queue length, availability as well as the failure frequency during different states are required for the analysis of any unreliable server queueing model. Some of them are obtained as follows:

(A) Long Run Probabilities

The long run probabilities of the system states are the probabilities with which the server remains at different states after attaining the steady state. These probabilities can be established as follows:

Theorem 2.5: The long run probabilities of the server at different states are obtained as:

(i) The probability that the server is idle and the system is empty, is

$$I_1 = D_0 \tag{2.43a}$$

(ii) The probability that the server is idle but the system is not empty, is

$$I_2 = \pi D_0 (1-\tilde{a}(\lambda)) / (1-\pi - r(1-\tilde{a}(\lambda)) \tag{2.43b}$$

(iii) The probability of the server being in busy state, is

$$P_B = D_0 (1-r+r\tilde{a}(\lambda))(\rho_0 + \sum_{i=1}^{k} \rho_i p_i) / (1-\pi - r(1-\tilde{a}(\lambda))) \tag{2.43c}$$

(iv) The probability of the server being under repair, is

$$P_R = D_0 (1-r+r\tilde{a}(\lambda))((\alpha_0 \rho_0 / \beta_0) + \sum_{i=1}^{k} (\rho_i p_i \alpha_i / \beta_i)) / (1-\pi - r(1-\tilde{a}(\lambda))) \tag{2.43d}$$

(v) The probability that the server is under setup, is

$$P_S = D_0 (1-r+r\tilde{a}(\lambda))((\alpha_0 \rho_0 / \xi_0) + \sum_{i=1}^{k} (\rho_i p_i \alpha_i / \xi_i)) / (1-\pi - r(1-\tilde{a}(\lambda))) \tag{2.43e}$$

Proof: Various long run probabilities given in (2.43a)-(2.43e), respectively are obtained by considering the following limiting behaviour of the server at various levels.

$$I_2 = \lim_{z\to 1} \int_0^\infty D(w,z)dw, \tag{2.44a}$$

$$P_B = \lim_{z\to 1} \left[\int_0^\infty P_0(x,z)dx + \sum_{i=1}^{k} \int_0^\infty P_i(x,z)dx \right], \tag{2.44b}$$

$$P_R = \lim_{z\to 1} \left[\int_0^\infty \int_0^\infty R_0(x,y,z)dxdy + \sum_{i=1}^{k} \int_0^\infty \int_0^\infty R_i(x,y,z)dxdy \right], \tag{2.44c}$$

$$P_S = \lim_{z\to 1} \left[\int_0^\infty \int_0^\infty S_0(x,y,z)dxdy + \sum_{i=1}^{k} \int_0^\infty \int_0^\infty S_i(x,y,z)dxdy \right] \tag{2.44d}$$

(B) Queueing Measures

The applicability of any queueing model depends on its mean performance measures such as mean queue length and mean size of the orbit. Now, we establish mean queue lengths of the orbit and system, and expected waiting time under investigation in the following theorem:

Theorem 2.6: The mean queue length of the retrial orbit (L_R) and mean queue length of the system (L_S) are respectively given by

$$\begin{aligned} L_R &= D_0(1-\tilde{a}(\lambda))(N_2/W_2)+\lambda D_0(1-r+r\tilde{a}(\lambda))(N_2/W_2) \\ &+\lambda D_0(1-r+r\tilde{a}(\lambda))\alpha_0((T_1+T_2)/W_4) \\ &+\lambda D_0(1-r+r\tilde{a}(\lambda))\alpha_0((T_5+T_6)/W_6 \\ &+\lambda D_0(1-r+r\tilde{a}(\lambda))\left\{\sum_{i=1}^{k} p_i(N_3/W_3+\alpha_i((T_3+T_4)/W_5)+\alpha_i((T_7+T_8)/W_7))\right\} \end{aligned} \tag{2.45}$$

$$L_S = L_R + V_1 + V_3 + V_5 + \sum_{i=1}^{k} p_i(V_2+V_4+V_6) \tag{2.46}$$

where, $N_1 = ac + D'b$, $N_2 = E_0'' F_0' a + E_0' F_0' b + E_0' F_0'' a$

$$N_3 = F_i' a\left(2E_i'\bar{E}_0 - E_i''\bar{E}_0\right) + E_i'\bar{E}_0\left(F_i' b + F_i'' a\right)$$

$$T_1 = 25aF_0'U'\{(-12E_0''L_0' - 12E_0'L_0'')(1-K_0) - 8E_0'L_0'(1-K_0)'\}$$

$$T_2 = 20E_0'L_0'\bar{K}_0\{12bF_0'U' + 11aF_0''U' + 12aF_0'U''\}$$

$$T_3 = 25aF_i'U'[[(-12E_i''L_i' - 12E_i'L_i'')(1-K_i) - 8E_i'L_i'(1-K_i)'](1-E_0) - 16E_i'L_i'(1-K_i)(1-E_0)']$$

$$T_4 = 20E_i'L_i'\bar{K}_i(1-E_0)\left\{12bF_i'U' + 11aF_i''U' + 12aF_i'U''\right\}$$

$$T_5 = 25aF_0'Q_0'\{-12E_0''K_0' - 12K_0''E_0'\} , \quad T_6 = 30E_0'K_0'\left\{12bF_0'Q' + 11aF_0''Q' + 12aF_0'Q''\right\}$$

$$T_7 = 25aF_i'Q'\{(-12E_i''K_i' - 12E_i'K_i'')(1-E_0) - 8E_i'K_i'(1-E_0)'\}$$

$$T_8 = 20E_i'K_i'\bar{E}_0\left\{12bF_i'Q' + 11aF_i''Q' + 12aF_i'Q''\right\}$$

$W_1 = 2a^2$, $W_2 = 2F_0'^2 a^2$, $W_3 = 2F_0'^2 a^2$, $W_4 = 432U'^2 F_0'^2 a^2$, $W_5 = 432U'^2 F_i'^2 a^2$

$W_6 = 432Q'^2 F_0'^2 a^2$, $W_7 = 432Q'^2 F_i'^2 a^2$,

$$V_1 = \frac{-2E_0'}{b}, \quad V_2 = \frac{-2E_i''\bar{E}_0}{aF_i'}, \quad V_3 = \frac{-4E_0'L_0'K_0}{5aU'F_0'},$$

$$V_4 = \frac{-4E_i'L_i'\overline{K}_iE_0}{5aU'F_i'}\ ,\ V_5 = \frac{-4E_0'K_0'}{5aQ'F_0'}\ ,\ V_6 = \frac{-4E_i'L_i'K_i'\overline{E}_0}{5aQ'F_i'}$$

$$a = (D' + r(1-\tilde{a}(\lambda)) - 1),\quad b = [D\{1 + r(z-1)(1-\tilde{a}(\lambda))\} - z]'',\ \ c = (-2D' - D'')$$

$$M_i(z) = \tilde{\eta}_i\{\lambda b_3(1-C(z))\}\tilde{g}_i(\lambda b_2(1-C(z)),\ M_i'(1) = -(\xi_i^{-1} + \beta_i^{-1}),$$

$$M_i''(1) = \xi_i^{(2)} + 2\xi_i^{(1)}\beta_i^{(1)} + \beta_i^{(2)},\ H_i(z) = \lambda b_1(1-C(z)) + \alpha_0(1-M_i(z),$$

$$H_i'(1) = -\lambda b_1E[X] + \alpha_i(\xi_i^{-1} + \beta_i^{-1})\ \ D = [p_0\tilde{b}_0(H_0(z)) + \sum_{i=1}^{k} p_i\tilde{b}_0(H_0(z))\tilde{b}_i(H_i(z))],$$

$$D' = \frac{p_0H_0'(1)}{\mu_0} + \frac{\sum_{i=1}^{k} p_iH_0'(1)}{\mu_0} + \frac{\sum_{i=1}^{k} p_iH_i'(1)}{\mu_i}\ \ H_i''(1) = -\lambda b_1E[X^2] - \alpha_iM_i''(1)$$

$$D'' = \frac{p_0H_0''(1)}{\mu_0} + b_0^{(2)}[H_0'(1)]^2 + (\sum_{i=1}^{k} p_iH_0''(1)/\mu_0) + b_0^{(2)}(\sum_{i=1}^{k} p_iH_0'(1)H_i'(1)/\mu_i)\left(\frac{1}{\mu_0} + \frac{1}{\mu_i}\right)$$

$$+ (\sum_{i=1}^{k} p_iH_i''(1)/\mu_0) + \sum_{i=1}^{k} p_ib_i^{(2)}[H_i'(1)]^2$$

$$E_i = [1 - \tilde{b}_i(H_i(z))]\ ,\ \ E_i'(1) = H_i'(1)/\mu_i,\ E_i''(1) = (-H_i''(1)/\mu_i) - b_i^{(2)}[H_i'(1)]^2,(0 \le i \le k)$$

$$F_i = H_0(z),\ \ F_i' = H_0'(z) = -\lambda E[X]\left[b_1 - \alpha_i\left(\frac{b_3}{\xi_i} + \frac{b_2}{\beta_i}\right)\right], F_0'' = (-\lambda b_1E[X^2] - \alpha_iM_i''(1))$$

$$K_i = [1 - \tilde{\eta}_i(\lambda b_3(1-C(z))],\ K_i' = (\lambda b_3E[X]/\xi_0),$$

$$K_i'' = -\xi_i^{(2)}[(\lambda b_3E[X])^2 + (\lambda b_3E[X^2]/\xi_0)]$$

$$L_i = [1 - \tilde{g}_i(\lambda b_2(1-C(z))],\ L_i' = (\lambda b_2E[X]/\beta_0),$$

$$L_i'' = -g_i^{(2)}[(\lambda b_2E[X])^2 + (\lambda b_2E[X^2]/\beta_0)]$$

$$U_i = \lambda b_2(1-C(z)),\ U_i' = -\lambda b_2E[X],\ \ U_i'' = -\lambda b_2E[X^2]$$

$$Q_i = \lambda b_3(1-C(z)),\ Q_i' = -\lambda b_3E[X],\ \ Q_i'' = -\lambda b_3E[X^2]$$

Proof: The mean queue length of the retrial orbit can be obtained by using $L_R = \lim_{z\to 1} P'(z)$, where P(z) is the generating function of retrial orbit. Similarly, the mean queue length of the system can be obtained by using $L_s = \lim_{z\to 1} L'(z)$, where L(z) is the generating function of the system size. Here, L - Hospital rule has been used six times to evaluate the limiting value when z → 1.

Theorem 2.7: The expected waiting time can be determined as:

$$W_s = \frac{L_s}{\lambda_{eff}E[X]} \tag{2.47}$$

where, $\lambda_{eff} = \lambda\left(I_2 + b_1P_B + b_2P_R + b_3P_S\right)$

Proof: The exact expected waiting time W_s is obtained using Little's formula (cf. Gross and Harris, 1985).

(C) Reliability Measures

The reliability indices of unreliable server give an idea of the availability and reliability of the server to perform its job successfully without any breakdowns.

Theorem 2.8: The steady state availability (A_v) and failure frequency (F_f) of the server are obtained using

$$A_v = \frac{D_0}{\pi}\left[r(1-\tilde{a}(\lambda)) - 1 + \pi\tilde{a}(\lambda) + (1-r+r\tilde{a}(\lambda))\left(\rho_0 + \sum_{i=1}^{k} p_i\rho_i\right)\right] \tag{2.48}$$

$$F_f = \frac{D_0}{\pi}\left[-(1-r+r\tilde{a}(\lambda))\left(\alpha_0\rho_0 + \sum_{i=1}^{k}\alpha_i p_i\rho_i\right)\right] \tag{2.49}$$

Proof: To derive above results (2.48) and (2.49), we have used

$$A_v = D_0 + \int_0^\infty D(w,1)dw + \int_0^\infty P_0(x,1)dx + \sum_{i=1}^{k}\int_0^\infty P_i(x,1)dx \tag{2.50}$$

$$F_f = \int_0^\infty \alpha_0 P_0(x,1)dx + \sum_{i=1}^{k}\int_0^\infty \alpha_i P_i(x,1)dx \tag{2.51}$$

2.5 SPECIAL CASES

Now, we consider some special cases of our model. The model under consideration can be reduced to various existing models available in the literature by setting some appropriate parameters as follows:

- **M/G/1 retrial queue with impatient customers and multi optional services**

  By relaxing the assumptions of setup before repair and bulk arrival, we get this particular case. The eq. (2.41) of our model coincides with the eq. (60) of Wang and Li (2009) if

  $b_1=b_2=b_3=b$, C (z) =z, $C'(z)=1$, $\tilde{\eta}_i=1\,(0\le i\le k)$.

- **M/G/1 retrial queue with single phase optional service, no setup and without impatient customers**

  On substituting,

  i=1, $b_1=b_2=b_3=1$, C (z) = z, $C'(z)=1$, r=0, $\tilde{\eta}_i=1\,(0\le i\le 1)$, $\alpha_i=0\,(0\le i\le 1)$

  eq. (2.41) coincides with eq. (4.13) of Wang *et al.* (2001).

- **M/G/1 queue with single phase optional service and server breakdowns**

Using, i=1, $\gamma = 0$, b_1=b_2=b_3=1, C (z) =z, $C'(z) = 1$, r = 0, $\tilde{\eta}_i$ =1 , $\alpha_i = 0$ $(2 \le i \le k)$ and second optional service as exponentially distributed, we get the same results as obtained by Wang (2004).

- **M/G/1 queue with Poisson input and optional service with general service distributions**

  For the specific deduction of our model, we consider both the services being general distributed and other parameters suitably adjusted as i =1, b_1=b_2=b_3=1, C (z) =z, $C'(z) =$ 1, r = 0, $\tilde{\eta}_i$ =1, $\alpha_i = 0$ $(0 \le i \le k)$, $\gamma = 0$. In this case, we get the results compatible with that obtained by Medhi (2002).

- **M/G/1 queue with second optional service and no breakdowns**

  For this case, we set i=1, $\gamma = 0$, b_1=b_2=b_3=1, C (z) =z, $C'(z) = 1$, r=0, $\tilde{\eta}_i$ =1 , $\alpha_i = 0$ $(0 \le i \le k)$ in our model. Furthermore, if the distribution of FES being general distributed and SOS as exponentially distributed, we get the results as obtained by Madan (2000).

2.6 MAXIMUM ENTROPY PRINCIPLE

The principle of maximum entropy (MEP) can be used for estimating the probabilistic information measures which can be further used to obtain queue size distribution of queueing systems in different frameworks. In this section, we employ maximum entropy principle for the $M^{[x]}$/G/1 retrial queueing system with impatient customers in order to determine the steady state probabilities of n customers in the system when the server being in i^{th} busy state $P_{i,n}$ $(0 \le i \le k)$, in repair state $R_{i,n}$ $(0 \le i \le k)$, in setup state $S_{i,n}$ $(0 \le i \le k)$ and being in idle state D_n. For the analysis purpose, we adopt the following procedure (cf. Wang *et al.*, 2007b):

(i) The construction of Lagrange's function H using the method of Lagrange's multipliers subject to a set of constraints in the terms of known indices.

(ii) Partial differentiation of Lagrange's function H w.r.t. $P_{i,n}$, $R_{i,n}$, $S_{i,n}$, and D_n and setting the results to zero.

(iii) Finally, solving the equations obtained in (ii) to derive results for the required probabilities.

The maximum entropy function Y (cf. El-Affendi and Kouvatos, 1983) is formulated in order to evaluate the steady state probabilities by using several known constraints in terms of performance characteristics as follows:

$$Y=-\sum_{n=1}^{\infty}\sum_{i=0}^{k}\mathrm{P}_{i,n}\log\mathrm{P}_{i,n}-\sum_{n=1}^{\infty}\sum_{i=0}^{k}R_{i,n}\log R_{i,n}-\sum_{n=1}^{\infty}\sum_{i=0}^{k}S_{i,n}\log S_{i,n}-\sum_{n=1}^{\infty}D_n\log D_n \tag{2.52}$$

subject to the constraints

(i) $\sum_{n=1}^{\infty}P_{i,n}=P_B$

(ii) $\sum_{n=1}^{\infty}D_n=I_2$

(iii) $\sum_{n=1}^{\infty}R_{i,n}=P_R$

(iv) $\sum_{n=1}^{\infty}S_{i,n}=P_S$

(v) $\sum_{n=1}^{\infty}n\left\{\sum_{i=0}^{k}\mathrm{P}_{i,n}+\sum_{i=0}^{k}R_{i,n}+\sum_{i=0}^{k}S_{i,n}+D_n\right\}=L_S\ ;\ (0\le i\le k)$

Lagrange's Function

To determine the maximum value of entropy function, we construct Lagrange's function $H(P_{i,n},R_{i,n},S_{i,n},D_n)$ by introducing the Lagrange's multipliers θ_i (0≤ i≤ k), δ_i (0 ≤ i≤ k), θ_{k+1} and φ_i (0 ≤ i≤ k+1) corresponding to the known information i.e. constraints available in the form of derived analytical results. Thus, we have

$$\begin{aligned}&H(P_{i,n},R_{i,n},S_{i,n},D_n)=\\&-\sum_{n=1}^{\infty}\sum_{i=0}^{k}P_{i,n}\log P_{i,n}-\sum_{n=1}^{\infty}\sum_{i=0}^{k}R_{i,n}\log R_{i,n}-\sum_{n=1}^{\infty}\sum_{i=0}^{k}S_{i,n}\log S_{i,n}-\sum_{n=1}^{\infty}D_n\log D_n-\sum_{i=0}^{k}\theta_i\left[\sum_{i=0}^{k}P_{i,n}-P_B\right]\\&-\theta_{k+1}\left[\sum_{n=1}^{\infty}D_n-I_2\right]-\sum_{i=0}^{k}\delta_i\left[\sum_{i=0}^{k}R_{i,n}-P_R\right]-\sum_{i=0}^{k}\phi_i\left[\sum_{i=0}^{k}S_{i,n}-P_S\right]\\&-\phi_{k+1}\left[\sum_{n=1}^{\infty}n\left\{\sum_{i=0}^{k}\mathrm{P}_{i,n}+\sum_{i=0}^{k}R_{i,n}+\sum_{i=0}^{k}S_{i,n}+D_n\right\}-L_S\right]\end{aligned} \tag{2.53}$$

Using MEP, we summarize our results for the approximate probabilities of different states in the form of following theorem:

Theorem 2.9: The maximum entropy solutions for approximate values of probabilities $P_{i,n}\,(0\le i\le k)$, $R_{i,n}\,(0\le i\le k)$, $S_{i,n}\,(0\le i\le k)$ and D_n subject to the constraints are

$$\begin{cases} \hat{P}_{i,n} = \dfrac{P_B\sigma\left[L_S-\sigma\right]^{n-1}}{L_S{}^n}, \ \hat{R}_{i,n} = \dfrac{P_R\sigma\left[L_S-\sigma\right]^{n-1}}{L_S{}^n}, \\ \hat{D}_n = \dfrac{I_2\sigma\left[L_S-\sigma\right]^{n-1}}{L_S{}^n}, \ \hat{S}_{i,n} = \dfrac{P_S\sigma\left[L_S-\sigma\right]^{n-1}}{L_S{}^n} \end{cases} \tag{2.54}$$

where, $\sigma = \sum_{i=0}^{k} P(B_i) + \sum_{i=0}^{k} P(R_i) + \sum_{i=0}^{k} P(S_i) + P(I)$

Proof: By following the procedure as stated above, we get

$$D_n = e^{-(1+\theta_{k+1})} e^{-n\phi_{k+1}}, \quad n \geq 0 \tag{2.55}$$

$$P_{i,n} = e^{-(1+\theta_i)} e^{-n\phi_{k+1}}, \ n \geq 0 \tag{2.56}$$

$$S_{i,n} = e^{-(1+\phi_i)} e^{-n\phi_{k+1}}, \quad n \geq 0 \tag{2.57}$$

$$R_{i,n} = e^{-(1+\delta_i)} e^{-n\phi_{k+1}}, \quad n \geq 0, (0 \leq i \leq k) \tag{2.58}$$

For brevity, we use the following notations:

$$e^{-(1+\phi_i)} = \psi_i \tag{2.59}$$

$$e^{-(1+\phi_i)} = \gamma_i, (0 \leq i \leq k+1) \tag{2.60}$$

$$e^{-(1+\delta_i)} = \chi_i, \tag{2.61}$$

$$e^{-\phi_{k+1}} = d_{k+1}, (0 \leq i \leq k) \tag{2.62}$$

Using eqs (2.59)-(2.62), eqs (2.55)-(2.58) reduce to-

$$P_{i,n} = \psi_i d_{k+1}^n, \tag{2.63}$$

$$R_{i,n} = \chi_i d_{k+1}^n, \tag{2.64}$$

$$S_{i,n} = \gamma_i d_{k+1}^n, \tag{2.65}$$

$$D_n = \psi_{k+1} d_{k+1}^n \tag{2.66}$$

Now, using results from eqs (2.63)-(2.66) in constraints, we get results for the long run probabilities of the server being in different states as-

$$P_B = \frac{\psi_i d_{k+1}}{1-d_{k+1}}, \tag{2.67}$$

$$P_R = \frac{\chi_i d_{k+1}}{1-d_{k+1}}, \tag{2.68}$$

$$P_S = \frac{\gamma_i d_{k+1}}{1-d_{k+1}}, \tag{2.69}$$

$$I_2 = \frac{\psi_{k+1} d_{k+1}}{1-d_{k+1}}, (0 \le i \le k) \tag{2.70}$$

Now, in order to determine the approximate queue length of the system, we substitute eqs (2.63)-(2.66) in

$$\sum_{n=1}^{\infty} n \left\{ \sum_{i=0}^{k} \mathrm{P}_{i,n} + \sum_{i=0}^{k} R_{i,n} + \sum_{i=0}^{k} S_{i,n} + D_n \right\} = L_S$$

and get

$$\hat{L}_s = \frac{d_{k+1}\left[\sum_{i=0}^{k+1} \psi_i + \sum_{i=0}^{k} \chi_i + \sum_{i=0}^{k} \gamma_i\right]}{\left(1-d_{k+1}^2\right)} \tag{2.71}$$

We denote $\sigma = P_B + P_R + P_S + I_2$ and using eqs (2.67) - (2.71), we have

$$L_s = \frac{\sigma}{1-d_{k+1}} \tag{2.72}$$

$$d_{k+1} = \frac{L_s - \sigma}{L_s} \tag{2.73}$$

Further, using (2.67)-(2.71) and (2.72)-(2.73), we get

$$\psi_i = \frac{P_B \sigma}{L_s - \sigma} \tag{2.74}$$

$$\chi_i = \frac{P_R \sigma}{L_s - \sigma} \tag{2.75}$$

$$\gamma_i = \frac{P_S \sigma}{L_s - \sigma} \tag{2.76}$$

$$\psi_{k+1} = \frac{I_2 \sigma}{L_s - \sigma}, (0 \le i \le k) \tag{2.77}$$

Finally, substituting results from eqs (2.72)-(2.77) in (2.63)-(2.66), we get expressions given in eq. (2.54).

Theorem 2.10: By using the maximum entropy principle, the approximate expected waiting time in the system is

$$\begin{aligned}
\hat{W}_s = &\sum_{n=1}^{\infty}\left[\sum_{i=0}^{k} \frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right] D_n + \sum_{i=0}^{k}\sum_{n=1}^{\infty}\left[\frac{n}{\mu_i} + \frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right] P_{i,n} \\
&+ \sum_{i=0}^{k}\sum_{n=1}^{\infty}\left[\frac{g_i^{(2)}}{2g_i^{(1)}} + \frac{n}{\mu_i} + \frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right] R_{i,n} \\
&+ \sum_{i=0}^{k}\sum_{n=1}^{\infty}\left[\frac{\eta_i^{(2)}}{2\eta_i^{(1)}} + \frac{n}{\mu_i} + \frac{1}{\beta_i} + \frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right] S_{i,n}
\end{aligned} \tag{2.78}$$

Proof: Following Wang *et al.* (2007b), maximum entropy principle can be used to obtain the approximate expected waiting time in the system. We proceed as follows:

Let us consider that a tagged customer say 'U' when arrives in the system, finds n customers preceding him in the queue. The server can be in any of the states i.e. idle, busy, under repair or under setup when customer 'U' arrives. These following cases may arise-

1. Idle state: If on the arrival, the customer 'U' finds the server in idle state then the incoming batch will be immediately served. The expected waiting time for a customer in this case includes the time taken by the additional customers in the batch preceding him to be served and is given as-

$$W_I = \sum_{n=1}^{\infty}\left[\sum_{i=0}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right] \tag{2.79}$$

2. Busy state: If the server is in busy state, then the incoming batch joins the orbit and the customers in the batch wait for their turn. For this case, the waiting time of the customer 'U' includes the serving time $\sum_{i=0}^{k}\frac{n}{\mu_i}$ of those n customers already present in the queue plus the waiting time $\sum_{i=0}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)$ of those who precedes 'U' in the batch. The total expected waiting time in the busy state is given by-

$$W_B = \sum_{n=1}^{\infty}\left[\sum_{i=0}^{k}\frac{n}{\mu_i}+\sum_{i=0}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right] \tag{2.80}$$

3. Setup State: When the server breaks down, it is sent for repair; but before repair, it is required for the repairman to make some preliminary settings before starting the repair. If the incoming customer finds the server under set up state when broken down during any ith $(0 \le i \le k)$ state of servicing, then it has to wait for the server to complete set up procedure with remaining set up time $\sum_{i=0}^{k}\frac{\eta_i^{(2)}}{2\eta_i^{(1)}}, (0 \le i \le k)$, repair time $\frac{1}{\beta_i}, (0 \le i \le k)$ as well as the service time $\sum_{i=0}^{k}\frac{n}{\mu_i}$ of n customers already present in the system. Moreover, the customers preceding 'U' will also take some time $\sum_{i=0}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)$ to get served. The mean waiting time in the set up state is

$$W_S=\sum_{n=1}^{\infty}\left[\sum_{i=0}^{k}\frac{\eta_i^{(2)}}{2\eta_i^{(1)}}+\sum_{i=0}^{k}\frac{1}{\beta_i}+\sum_{i=0}^{k}\frac{n}{\mu_i}+\sum_{i=0}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right] \tag{2.81}$$

4. Repair state: When the server is in the repair state, the incoming batch will be served after completion of the repair of the server plus the servicing of those n customers already waiting in the queue. The mean remaining repair time is given by $\sum_{i=0}^{k}\frac{g_i^{(2)}}{2g_i^{(1)}}$ when the server breaks down during i[th] $(0\le i\le k)$ state of the servicing, the waiting time for the servicing of n customers is $\sum_{i=0}^{k}\frac{n}{\mu_i}$. Moreover, the customers preceding 'U' will also take some time $\sum_{i=0}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)$. Hence, the total mean waiting when the server is in repair state is

$$W_R=\sum_{n=1}^{\infty}\left[\sum_{i=0}^{k}\frac{g_i^{(2)}}{2g_i^{(1)}}+\sum_{i=0}^{k}\frac{n}{\mu_i}+\sum_{i=0}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right] \tag{2.82}$$

Therefore, approximate expected waiting time in the queue given by (2.78) can be obtained by adding all the above expressions (2.79)-(2.82) for the waiting time obtained for different cases.

2.7 NUMERICAL ILLUSTRATION

The numerical results of the retrial queueing model under consideration are obtained by coding program in 'MATLAB' software. We have divided our numerical illustration in two sub parts; (A) sensitivity analysis and (B) maximum entropy results. All the numerical computations have been done by considering only two optional services i.e. k=2. Moreover, the retrial time, repair time and setup time are also assumed to be exponentially distributed.

(A) Sensitivity Analysis

To study the effect of various parameters on different performance indices, we consider default parameters for computational purpose as-
$\lambda=0.5,\alpha_0=0.1,\alpha_1=0.2,\alpha_2=0.1,\beta_0=0.5=\beta_1=\beta_2,\gamma=0.1,\mu=\mu_0=\mu_1=\mu_2=5,p_0=0.4,$
$p_1=0.2,p_2=0.4,r=0.1,\xi_0=\xi_1=\xi_2=0.7.$
At some stages of simulation, homogenous breakdown rate is also used $(\alpha=\alpha_0=\alpha_1=\alpha_2=0.1)$.

- ***Effect on reliability indices**:* Reliability indices basically give an idea of the availability, failure frequency and reliability of the server. Table 2.1 clearly indicates that the availability of the server decreases with an increase in the value of mean batch size (E[X]) for fixed values of other parameters like r, α and γ. Also, an increase in λ results in the decrement of availability and increment of failure frequency. We notice that an increase in r also decreases the availability of the server. However, failure frequency increases with the increase in breakdown rate (α).

Table 2.2 presents the variation in the reliability indices with varying values of service rate (μ) and other parameters. The increment in service rate with fixed values of r, α and γ results in the increment (decrement) of the availability (failure frequency) of the server due to the speeding up of the servicing of the customers. But for fixed value of μ and on increasing r, availability (failure frequency) decreases (increases).

- ***Effect on Queue length:*** The effect of parameters α, λ, γ, r, μ on the number of customers in the system (L_s) is presented by means of figs 2.1-2.4. The service pattern has been considered as Erlangian, exponential and gamma distributed for computational purpose. Fig. 2.1 exhibits the trend of L_s with respect to the arrival rate for different values of E[X] for Erlangian, exponential and gamma service time distributions. It is observed that the length of the system increases with the an increase in the mean batch size E[X] which is quite obvious as more and more customers in a batch will increase the number of customers in the system. Fig. 2.2 depicts the variation in the L_s with retrial rate γ. As the value of γ increases, L_s seems to increase; this may be due to fact that as more and more customers retry for the service, the accumulation of the customers increases.

On comparing the results for all the concerned distributions, we notice that L_s attains its highest value in the case of gamma distribution. The effect of variation with ‘r’ on L_s has been demonstrated in fig. 2.3. The variation in L_s also follows the increasing pattern with an increase in the values of r. An increase in the breakdown rate (α) interrupts in the servicing procedure, thereby increasing the queue length of the system. Fig. 2.4 depicts the effect of variation of α on L_s. The graphs plotted in fig. 2.4 show that L_s increases with an increase in the breakdown rate. This is because, simultaneous effect of increase in α as well as in λ increases the number of customers at a faster rate and the longer queue might build up in such a situation.

(B) Exact Results vs. Maximum Entropy Results

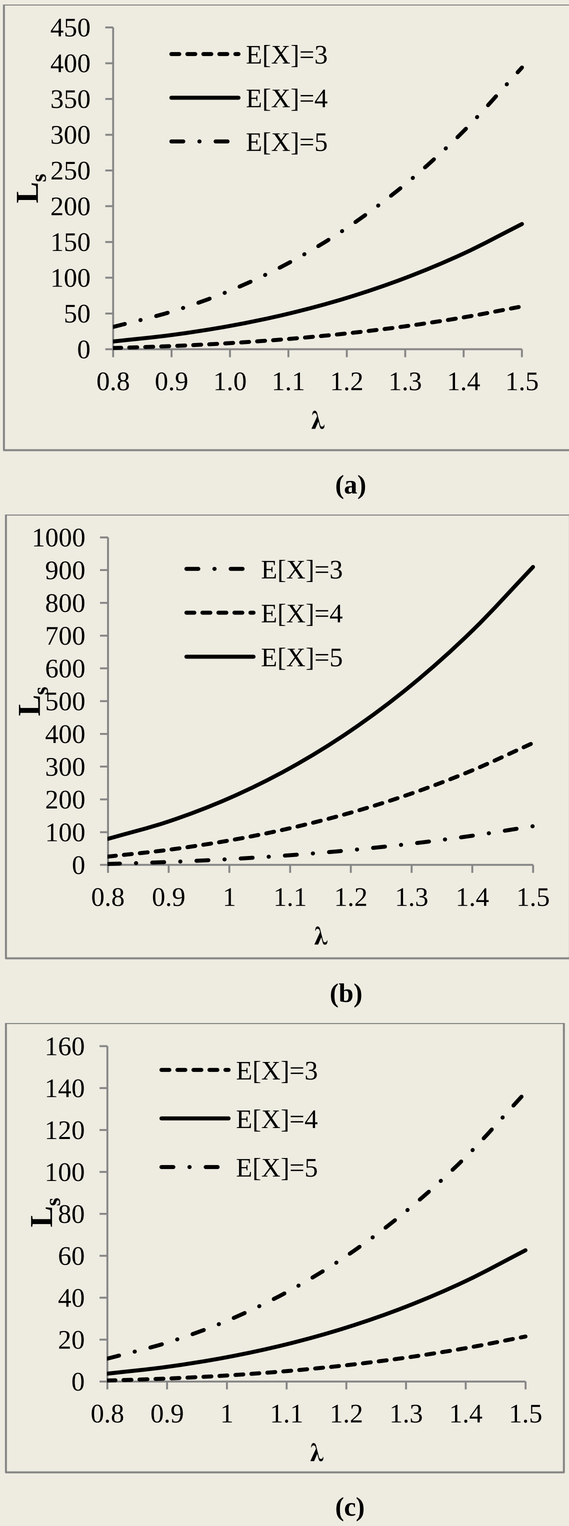
In this section, we perform numerical experiment to facilitate a comparison between the exact average queue length (L_s) and approximate average queue length ($\hat{L}_s$), exact

Table 2.1: Effect of λ on A_V and F_f for $M^{[x]}$/M/1 retrial queueing model

λ	(r, γ, α)	A_V E[X]=3	A_V E[X]=4	A_V E[X]=5	F_f E[X]=3	F_f E[X]=4	F_f E[X]=5
	(0.1,0.2,0.1)	0.6312	0.6056	0.5761	0.0163	0.0177	0.0194
	(0.2,0.2,0.1)	0.6251	0.5962	0.5625	0.0160	0.0172	0.0187
0.5	**(0.3,0.2,0.1)**	0.6177	0.5846	0.5452	0.0157	0.0168	0.0181
	(0.1,0.3,0.1)	0.6083	0.5697	0.5226	0.0167	0.0184	0.0204
	(0.1,0.2,0.2)	0.5790	0.5418	0.4973	0.0313	0.0341	0.0374
	(0.1,0.2,0.3)	0.5458	0.4873	0.4116	0.0338	0.0382	0.0438
	(0.1,0.2,0.1)	0.3728	0.3289	0.2785	0.0268	0.0286	0.0308
	(0.2,0.2,0.1)	0.3602	0.3096	0.2503	0.0273	0.0295	0.0320
1.0	**(0.3,0.2,0.1)**	0.3442	0.2845	0.2127	0.0280	0.0305	0.0336
	(0.1,0.3,0.1)	0.3230	0.2503	0.1601	0.0289	0.0320	0.0359
	(0.1,0.2,0.2)	0.3098	0.2508	0.1807	0.0514	0.0558	0.0610
	(0.1,0.2,0.3)	0.2517	0.1766	0.0849	0.0741	0.0816	0.0907

Table 2.2: Effect of μ on A_V and F_f for $M^{[x]}$/M/1 retrial queueing model

μ	(r, γ, α)	A_V E[X]=3	A_V E[X]=4	A_V E[X]=5	F_f E[X]=3	F_f E[X]=4	F_f E[X]=5
	(0.1,0.2,0.1)	0.3728	0.3289	0.2785	0.0268	0.0286	0.0308
	(0.2,0.2,0.1)	0.3602	0.3096	0.2503	0.0273	0.0295	0.0320
5.0	**(0.3,0.2,0.1)**	0.3442	0.2845	0.2127	0.0280	0.0305	0.0336
	(0.1,0.3,0.1)	0.3230	0.2503	0.1601	0.0289	0.0320	0.0359
	(0.1,0.2,0.2)	0.3098	0.2508	0.1807	0.0514	0.0558	0.0610
	(0.1,0.2,0.3)	0.2517	0.1766	0.0849	0.0741	0.0816	0.0907
	(0.1,0.2,0.1)	0.5301	0.5059	0.4790	0.0201	0.0211	0.0222
	(0.2,0.2,0.1)	0.5230	0.4955	0.4645	0.0204	0.0215	0.0229
7.0	**(0.3,0.2,0.1)**	0.5142	0.4822	0.4456	0.0207	0.0221	0.0237
	(0.1,0.3,0.1)	0.5027	0.4645	0.4200	0.0212	0.0229	0.0248
	(0.1,0.2,0.2)	0.4776	0.4445	0.4069	0.0389	0.0414	0.0442
	(0.1,0.2,0.3)	0.4283	0.3855	0.3358	0.0566	0.0609	0.0658



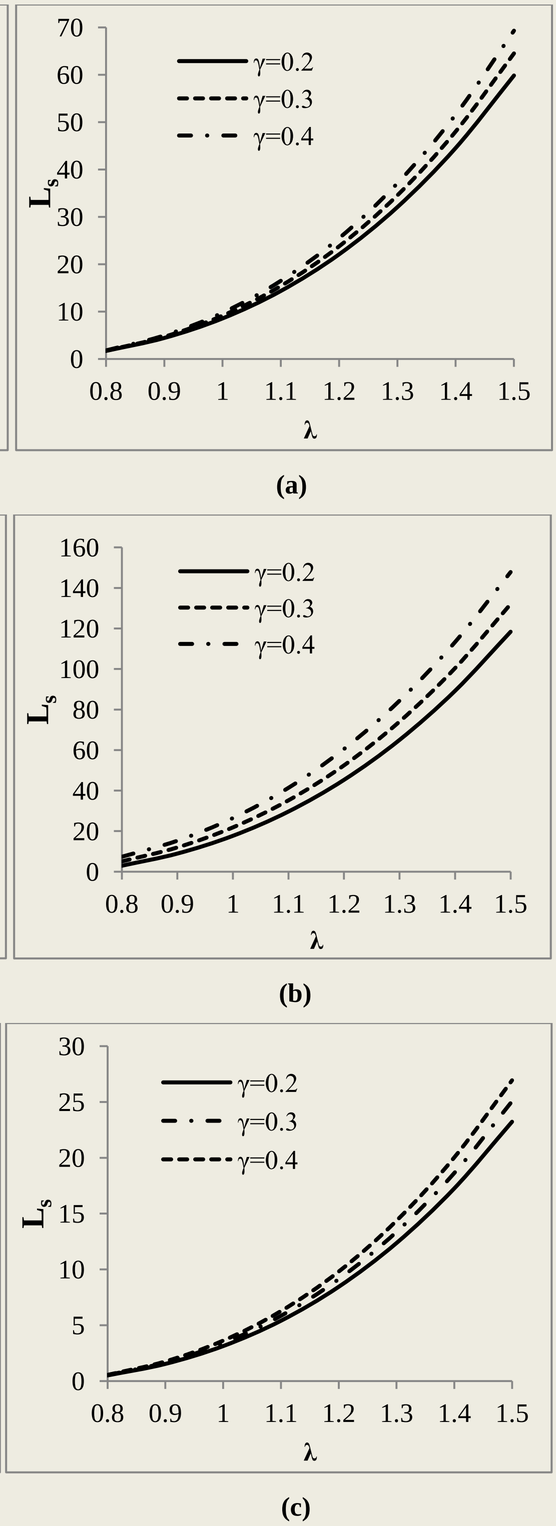
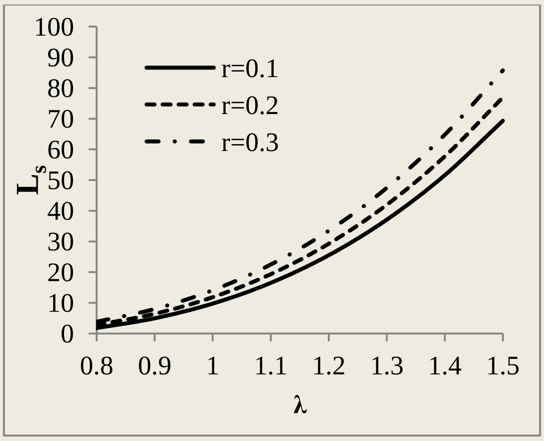
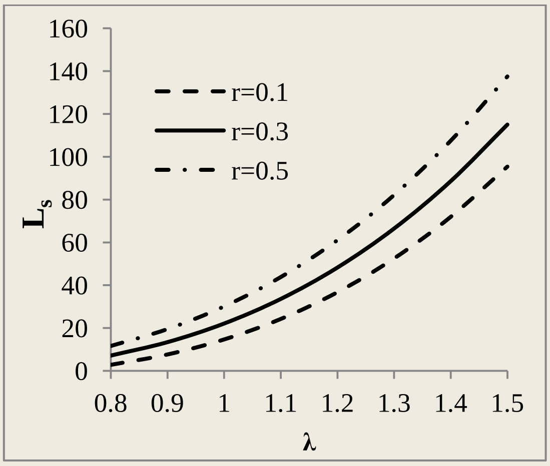


Fig. 2.1: Effect of λ and E[X] on L_s for (a) $M^x/M/1$ (b) $M^x/\gamma/1$ (c) $M^x/E_2/1$

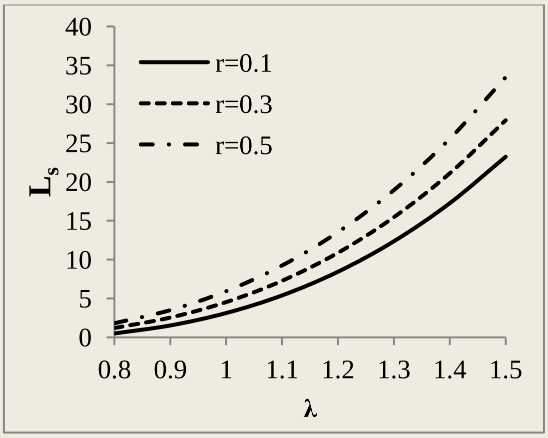
Fig. 2.2: Effect of λ and γ on L_s for (a) $M^x/M/1$ (b) $M^x/\gamma/1$ (c) $M^x/E_2/1$



(a)

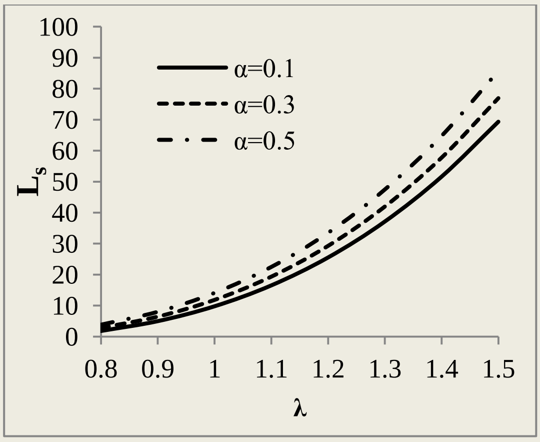


(b)

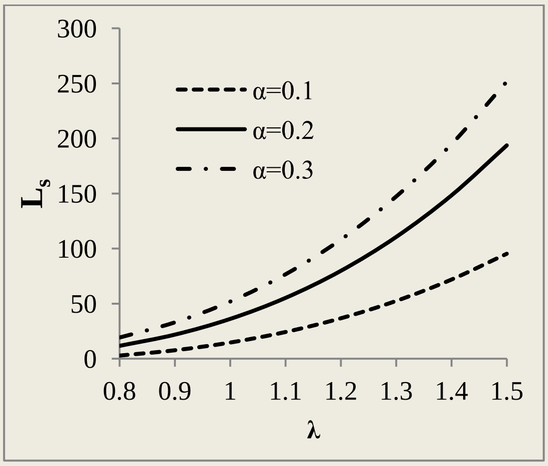


(c)

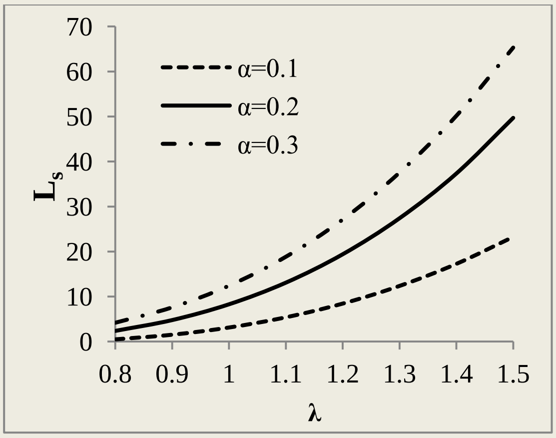
Fig. 2.3: Effect of λ and r on L_s for (a) $M^x/M/1$ (b) $M^x/\gamma/1$ (c) $M^x/E_2/1$



(a)



(b)



(c)

Fig. 2.4: Effect of λ and α on L_s for (a) $M^x/M/1$ (b) $M^x/\gamma/1$ (c) $M^x/E_2/1$

waiting time (W_s) and approximate waiting time ($\hat{W}_S$) obtained by applying maximum entropy principle. To examine how close are the approximate results obtained by MEP to exact results; we calculate absolute percentage error APE1 for the waiting time and APE2 for the queue length for different service time distributions. The default parameters for the computational purpose are fixed as $\lambda = 1, \alpha_0 = \alpha_1 = 0.2, \alpha_2 = 0.02, \beta_0 = 1, \beta_1 = \beta_2 = 0.1, p_0 = 0.4, p_1 = 0.2, p_2 = 0.4, r = 0.1,$ $\xi_0 = \xi_1 = \xi_2 = 1, \gamma = 0.1, \mu_1 = \mu_2 = \mu_3 = 9.$

It is noticed from table 2.3 that when the service time is Erlangian-2 distributed and repair time is exponentially distributed, the absolute percentage error for the average waiting time (APE1) increases with an increase r whereas it decreases with an increase in the retrial rate (γ). On the other hand absolute percentage error (APE2) for the average queue length exhibits opposite behaviour. It decreases with an increase in both r and γ. APE1 increases significantly with an increase in r. The effects of r and γ on APE1 and APE2 for $M^x/M/1$ retrial model have been displayed in table 2.4. The waiting time increases with an increase in r. However, APE2 gives exactly the same value as obtained for Erlangian distributed service time but queue lengths differ significantly.

Table 2.3: Maximum entropy results for $M^x/E_2/1$ retrial model

r	W_s	$\hat{W}_S$	APE1 (%)	L_s	$\hat{L}_s$	APE2 (%)
0.10	269.35	253.67	5.82	538.71	531.73	1.29
0.12	274.89	253.69	7.71	549.78	543.19	1.19
0.14	280.49	253.74	9.53	560.99	554.80	1.10
0.16	286.18	253.83	11.30	572.37	566.57	1.01
0.18	291.96	253.95	13.01	583.92	578.52	0.92
0.20	297.82	254.10	14.68	595.65	590.65	0.83
γ	**W_s**	**$\hat{W}_S$**	**APE1 (%)**	**L_s**	**$\hat{L}_s$**	**APE2 (%)**
0.05	261.93	247.00	5.70	523.87	516.49	1.40
0.10	269.35	253.67	5.82	538.71	531.73	1.29
0.15	276.44	260.09	5.91	552.88	546.29	1.19
0.20	283.21	266.27	5.98	556.42	560.23	1.09
0.25	289.68	272.23	6.02	579.37	573.58	1.00
0.30	295.89	277.97	6.05	591.78	586.37	0.91

Table 2.4: Maximum entropy results for $M^x/M/1$ retrial model

r	W_s	$\hat{W}_S$	APE1 (%)	L_s	$\hat{L}_s$	APE2 (%)
0.10	354.31	292.17	17.53	708.63	699.45	1.29
0.12	361.55	292.69	19.04	723.10	714.44	1.19
0.14	368.88	293.23	20.50	737.76	729.61	1.13
0.16	376.31	293.80	21.92	752.62	745.00	1.01
0.18	383.84	294.39	23.30	767.69	760.60	0.92
0.20	391.50	295.00	24.64	783.00	776.43	0.83
γ	W_s	$\hat{W}_S$	**APE1 (%)**	L_s	$\hat{L}_s$	**APE2 (%)**
0.05	344.44	283.79	17.60	688.89	679.19	1.40
0.10	354.31	292.17	17.53	708.63	699.45	1.29
0.15	363.74	300.24	17.45	727.48	718.81	1.19
0.20	372.74	308.02	17.36	745.49	737.35	1.09
0.25	381.36	315.52	17.26	762.73	755.10	1.00
0.30	389.61	322.76	17.15	779.23	772.11	0.91

Table 2.5: Maximum entropy results for $M^x/\gamma/1$ retrial model

r	W_s	$\hat{W}_S$	APE1 (%)	L_s	$\hat{L}_s$	APE2 (%)
0.10	291.83	263.85	9.58	583.66	576.10	1.29
0.12	297.82	264.01	11.35	595.65	588.52	1.19
0.14	303.90	264.20	13.06	607.81	601.10	1.10
0.16	310.06	264.42	14.71	620.13	613.85	1.01
0.18	316.32	264.67	16.32	632.64	626.79	0.92
0.20	322.67	264.95	17.88	645.35	639.94	0.83
γ	W_s	$\hat{W}_S$	**APE1 (%)**	L_s	$\hat{L}_s$	**APE2 (%)**
0.05	283.79	256.74	9.53	567.59	559.59	1.40
0.10	291.83	263.85	9.58	583.66	576.10	1.25
0.15	299.50	270.70	9.61	599.01	591.88	1.19
0.20	306.84	277.29	9.63	613.68	606.98	1.09
0.25	313.86	283.64	9.62	627.72	621.44	1.00
0.30	320.58	289.77	9.61	641.16	635.30	0.91

The minimum % error for the queue length has been obtained by varying the values of both r and γ. It is found that the results obtained by using maximum entropy are reasonably close with that obtained by the analytic technique at higher reneging and higher retrial rates. The data captured in table 2.5 has been calculated for a queueing model dealing with service time as gamma distributed and repair time as exponentially distributed. Table 2.5 exhibits the effect of r and γ on the waiting time and queue length. The variation in absolute % error for queue length is also reported for all the cases with a chosen set of default parameters.

2.8 DISCUSSION

The performance analysis of unreliable bulk arrival retrial models with impatient customers and optional services is studied. We can summarise our findings in the present study as follows:

- L_s increases significantly with an increase in the values of arrival rate λ, mean batch size E[X] and reneging probability. The system performance is also sensitive to the minor changes in the values of retrial rate γ and server breakdown rate α. Simultaneous increment in both α and λ contributes significantly to the increase in L_s.
- As expected, reliability indices viz. availability (failure frequency) is also affected by the related parameters. The availability of the server can be increased by improving the repair rate which will also help in reducing the failure frequency.
- It is noticed that out of the three distributions namely exponential, gamma and Erlangian for service pattern, gamma distribution gives maximum queue length while Erlangian, which is phase type distribution, gives minimum queue length of the system. The least absolute percentage error for waiting time is obtained in the case when the service time is Erlangian distributed and repair time is exponentially distributed.
- The absolute percentage error for waiting time increases significantly with the increase in r and increases gradually with an increase in γ. Of all the three cases involving repair process as exponential process, higher percentage errors are obtained when both service as well as repair process are exponential distributed while minimum errors are obtained in case of Erlangian-2 service time distribution.

CHAPTER 3

RETRIAL QUEUE WITH BERNOULLI VACATION SCHEDULE

3.1 INTRODUCTION

Vacation retrial queueing models find a significant place in the mathematical modeling of a variety of congestion situations where server may opt for vacation of random length due to the non-availability of the customers to be served. The recent past comprehensive study on queueing systems with server vacations was due to Doshi (1986) and Takagi (1993).

Adaptive Neuro Fuzzy Inference System (ANFIS) can be used along with the traditional classical techniques to analyze the complex queueing systems in a more efficient manner. Jang and Sun (1995) opened new ways to the study of ANFIS by suggesting algorithms of adaptive network based fuzzy inference systems. A detailed description of ANFIS can be found in the articles by Tettamanzi and Tomassini (2001). Lin and Liu (2009) also studied ANFIS technique to predict the CMP manufacturing parameters. Moreover, queueing network modeling of flexible manufacturing system using mean value analysis has also been done by Jain *et al.* (2008). Jain and Upadhyaya (2009) analysed threshold N-policy for degraded machining system using ANFIS approach. Bhargava and Jain (2014) studied unreliable multiserver queueing system with modified vacation policy and compared the results using ANFIS approach.

A bulk arrival retrial queue with vacation and impatient customers is studied in this chapter by using generating function and supplementary variables technique. Various features incorporated in this investigation are (i) retrial queue (ii) bulk arrival (iii) vacation (iv) multioptional services (v) phase repair and (vi) impatient customers. The rest of the chapter is organized in the following manner. Section 3.2 describes the mathematical assumptions and requisite notations to develop the model under consideration. The governing equations and generating functions of the queue size distribution are obtained in sections 3.3 and 3.4 respectively. Some special cases of our model are discussed in section 3.5. The performance measures are derived in section 3.6 to characterize the queueing and reliability characteristics. Section 3.7 is devoted to the sensitivity analysis by taking numerical illustration and using ANFIS approach. Finally, we wind up our investigation with conclusions in section 3.8.

3.2 MODEL DESCRIPTION

We consider a bulk arrival $M^{[x]}/G/1$ retrial queueing system with multioptional services and reneging. The customers arrive in Poisson manner with arrival rate λ in batches. The service is provided to the customer if the server is idle; otherwise he is forced to join the virtual pool (i.e. retrial orbit) of the customers, from where he tries again and again for the service after a random interval of time which is exponential distributed with rate γ. The service process is general distributed and is performed in the same manner as described in chapter 2.2. The server is unreliable and may breakdown in Poisson fashion during any course of service. The other new assumptions in this model are as follows:

- **Repair Process:** The repair process is assumed to be general distributed and completed in d compulsory phases with rate $\beta_j, (1 \le j \le d)$ for the server broken during i^{th} $(0 \le i \le k)$ phase service and at j^{th} phase repair. But before starting the repair process, the server takes some time called setup time with rate ξ to make some preliminary settings i.e. there is delay-in-repair. The setup time and repair time of j^{th} $(1 \le j \le d)$ repair phase are considered to be i.i.d. and general distributed.
- **Bernoulli Vacation Schedule:** After each service completion, the server may also go for a vacation of random length with probability θ or may continue to serve the next customer with complementary probability $(1-\theta)$. The vacation time is assumed to be general distributed with rate ψ .

3.3 QUEUE SIZE DISTRIBUTION

In order to formulate the equations for the present non-markovian system, the supplementary variable technique has been employed by introducing the supplementary variables for elapsed service time, elapsed vacation time, the elapsed repair time and elapsed set up time.

Let N(t) represents the number of customers in the system and $S_1(t)$ and $S_2(t)$ denote the phase of the service and phase of repair respectively, at any time t.

The state of the server at any time t is given by

$$
\text{Y(t)} = \begin{cases} 1, & \text{server is in idle state} \\ 2, & \text{server is busy in providing FES to the customers} \\ 3, & \text{server is busy in providing SOS to the customers} \\ 4, & \text{server is broken down and under setup before repair} \\ 5, & \text{server is broken down and under repair} \\ 6, & \text{server is under vacation} \end{cases}
$$

In the steady state, the joint distributions of the server state and queue size are defined as-

$$D_n = \lim_{t\to\infty} \Pr\{Y(t)=1, N(t)=n\}, n \ge 0$$

$$P_{0,n}(x) = \lim_{t\to\infty} \Pr\{Y(t)=2, x \le \varpi(t) \le x+dx, N(t)=n, S_1(t)=0\}, n \ge 0$$

$$P_{i,n}(x) = \lim_{t\to\infty} \Pr\{Y(t)=3, x \le \varpi(t) \le x+dx, N(t)=n, S_1(t)=i\}, n \ge 0, (1 \le i \le k)$$

$$S_n(x,y) = \lim_{t\to\infty} \Pr\{Y(t)=4, \varpi(t)=x, y \le \sigma(t) \le y+dy, N(t)=n\}, \ n \ge 0$$

$$R_{i,j,n}(x,y) = \lim_{t\to\infty} \Pr\{Y(t)=4, \varpi(t)=x, y \le \sigma(t) \le y+dy, N(t)=n, S_2(t)=j\},$$

$$n \ge 0, \ (0 \le j \le k), (1 \le j \le d)$$

$$V_n(x) = \lim_{t\to\infty} \Pr\{Y(t)=6, x \le \varpi(t) \le x+dx, N(t)=n\}, n \ge 0$$

Below we construct a set of steady state Kolomogorov forward equations after introducing the supplementary variables as follows:

$$\lambda D_n(0) = \theta p_0 \int_0^\infty P_{0,0}(x)\mu_0(x)dx + \theta \sum_{i=1}^{k} \int_0^\infty P_{i,0}(x)\mu_i(x)dx + \int_0^\infty V_n(x)\psi(x)dx; \ (1 \le i \le k) \tag{3.1}$$

$$\left[\frac{d}{dw} + \lambda + \gamma(w)\right] D_n(w) = 0; \ n \ge 1 \tag{3.2}$$

$$\left[\frac{d}{dx} + \lambda + \alpha_0 + \mu_0(x)\right] P_{0,n}(x) = \lambda \sum_{m=1}^{n} c_m P_{0,n-m}(x) + \int_0^\infty R_{0,d,n}(x,y)\beta_d(y)dy, \tag{3.3}$$

$$\left[\frac{d}{dx} + \lambda + \alpha_i + \mu_i(x)\right] P_{i,n}(x) = \lambda \sum_{m=1}^{n} c_m P_{i,n-m}(x) + \int_0^\infty R_{i,d,n}(x,y)\beta_d(y)dy; \ (1 \le i \le k) \tag{3.4}$$

$$\left[\frac{d}{du} + \lambda + \psi(x)\right] V_n(x) = \lambda \sum_{m=1}^{n} c_m V_{n-m}(x); \quad n \ge 1 \tag{3.5}$$

$$\left[\frac{\partial}{\partial y} + \lambda + \beta_j(y)\right] R_{i,j,n}(x,y) = \lambda \sum_{m=1}^{n} c_m R_{i,j,n-m}(x,y); \ (0 \le i \le k), (1 \le j \le d), \ n \ge 0 \tag{3.6}$$

$$\left[\frac{\partial}{\partial y} + \xi(y) + \lambda\right] S_n(x,y) = \lambda \sum_{m=1}^{n} c_m S_{n-m}(x,y); \ n \ge 0 \tag{3.7}$$

$$D_n(0) = \theta p_0 \int_0^\infty P_{0,n}(x)dx\mu_0(x)dx + \theta \sum_{i=1}^{k} \int_0^\infty P_{i,n}(x)\mu_i(x)dx + \int_0^\infty V_n(x)\psi(x)dx; \quad n \ge 1 \tag{3.8}$$

$$P_{0,0}(0)=\int_0^\infty D_1(w)\gamma(w)dw+(1-r)\lambda\int_0^\infty D_1(w)dw+\lambda D_0 \quad (3.9)$$

$$P_{0,n}(0)=\int_0^\infty D_{n+1}(w)\gamma(w)dw+(1-r)\lambda\int_0^\infty D_{n+1}(w)dw+r\lambda\int_0^\infty D_n(w)dw; \quad n\geq 1 \quad (3.10)$$

$$P_{i,n}(0)=p_i\int_0^\infty P_{0,n}(x)\mu_0(x)dx;\ n\geq 1,\ (1\leq i\leq k) \quad (3.11)$$

$$S_n(x,0)=\alpha_i P_{i,n}(x),\ \ (0\leq i\leq k);\ n\geq 1 \quad (3.12)$$

$$R_{i,1,n}(x,0)=\int_0^\infty S_n(x,y)\xi(y)dy;\ (0\leq i\leq k), n\geq 1 \quad (3.13)$$

$$R_{i,j,n}(x,0)=\int_0^\infty R_{i,j-1,n}(x,y)\beta_{j-1}\,dy; n\geq 1,\ \ (0\leq i\leq k),\ (2\leq j\leq d) \quad (3.14)$$

$$V_n(0)=(1-\theta)\ p_0\int_0^\infty P_{0,n}(x)\mu_0(x)dx+(1-\theta)\sum_{i=1}^{k}\int_0^\infty P_{i,n}(x)\mu_i(x)dx \quad (3.15)$$

The normalizing condition is given by

$$D_0+\sum_{n=1}^{\infty}\int_0^\infty D_n(w)dw+\sum_{n=0}^{\infty}\left[\int_0^\infty P_{0,n}(x)dx+\sum_{i=1}^{k}\int_0^\infty P_{i,n}(x)dx\right]$$
$$+\sum_{n=0}^{\infty}\sum_{i=0}^{k}\sum_{j=1}^{d}\int_0^\infty\int_0^\infty R_{i,j,n}(x,y)dxdy+\sum_{n=0}^{\infty}\int_0^\infty\int_0^\infty S_n(x,y)dxdy+\sum_{n=0}^{\infty}\int_0^\infty V_n(x)dx=1 \quad (3.16)$$

3.4 PROBABILITY GENERATING FUNCTION

As considered in chapter 2, here also we use probability generating function technique to obtain the steady state solution of the retrial queueing model. The generating function corresponding to the service time and batch size is same as considered in chapter 2. The generating function corresponding to phase-repair, vacation and setup time, respectively for this analysis are defined as:

$$R_{i,j}(x,y,z)=\sum_{n=0}^{\infty}R_{i,j,n}(x,y)z^n \quad (0\leq i\leq k),(1\leq j\leq d),\ V(x,z)=\sum_{n=0}^{\infty}V_n(x)z^n;$$
$$S(x,y,z)=\sum_{n=0}^{\infty}S_n(x,y)z^n;\ |z|\leq 1. \quad (3.17)$$

The hazard rate corresponding to phase repair, vacation and set up time are respectively, defined as

$$\beta_{i,j}(y)=\frac{g_{i,j}(x)}{1-G_{i,j}(x)},\ \ (0\leq i\leq k),(1\leq j\leq d),\ \xi(y)=\frac{\eta(y)}{1-N(y)}\ \text{and}\ \psi(x)=\frac{w(x)}{1-W(x)}.$$

Now, we establish some results expressed in the form of theorems as stated below:

Theorem 3.1: The partial generating functions for the server being in idle state, FES state, i[th] $(1 \le i \le k)$ SOS busy state, under j[th] $(1 \le j \le d)$ phase repair while broken down in i^{th} $(1 \le i \le k)$ phase service, under setup and on vacation at random epoch respectively, are

$$D(w,z) = D(0,z)\exp\{-\lambda w\}\bar{A}(w) \tag{3.18}$$

$$P_0(x,z) = P_0(0,z)\exp\{-M_0(z)x\}\overline{B_0}(x) \tag{3.19}$$

$$P_i(x,z) = p_i\tilde{b}_0(M_j(z))\exp\{-M_i(z)x\}\overline{B_i}(x)P_0(0,z) \quad ,\left(1 \le i \le k\right) \tag{3.20}$$

$$R_{i,j}(x,y,z) = \alpha_i p_i P_i(x,z)\exp\{-\lambda(1-C(z))y\}H_{i,j}(z)\bar{G}_{i,j}(y);\ \left(0 \le i \le k\right),\left(1 \le j \le d\right) \tag{3.21}$$

$$S(x,y,z) = \alpha_i p_i P_i(x,z)\exp\{-\lambda(1-C(z))y\}\bar{N}(y),\left(0 \le i \le k\right) \tag{3.22}$$

$$V(x,z) = (1-\theta)\sum_{i=0}^{k} p_i P_0(0,z)\tilde{b}_i(M_i(z))\bar{W}(x)\exp\{-\lambda(1-C(z))x\} \tag{3.23}$$

And,

$$\text{D(0,z)} = \frac{\lambda D_0[1-\theta X(z)-(1-\theta)Y(z)]}{[1+r(z-1)(1-\tilde{a}(\lambda))][\theta X(z)+(1-\theta)Y(z)]-z} \tag{3.24}$$

$$P_0(0,z) = \frac{\lambda D_0[1+r(z-1)(1-\tilde{a}(\lambda))-z]}{[1+r(z-1)(1-\tilde{a}(\lambda))][\theta X(z)+(1-\theta)Y(z)]-z} \tag{3.25}$$

where,

$$H_{i,j}(z) = \tilde{\eta}\{\lambda(1-C(z))\}\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}\{\lambda(1-C(z))\}\right),\left(0 \le i \le k\right),\left(1 \le j \le d\right)$$

$$H^{(i)}(z) = \tilde{\eta}\{\lambda(1-C(z))\}\left(\prod_{r=1}^{d}\tilde{g}_{i,r}\{\lambda(1-C(z))\}\right),\left(0 \le i \le k\right) \tag{3.26}$$

$$M_i(z) = \lambda(1-C(z)) + \alpha_i(1-H^{(i)}(z)),\left(0 \le i \le k\right) \tag{3.27}$$

$$X(z) = p_0\tilde{b}_0(M_i(z)) + \sum_{i=1}^{k} p_i\tilde{b}_0(M_i(z))\tilde{b}_i(M_i(z)) \tag{3.28}$$

$$Y(z) = \sum_{i=0}^{k} p_i\tilde{b}_i(M_i(z))\tilde{w}(\lambda(1-C(z))) \tag{3.29}$$

Note: For the brevity, we use the product $\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}(-\lambda(1-C(z)))\right)$ in our results. However, the value of $\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}(-\lambda(1-C(z)))\right)$=1 when j=1.

Proof: The proof follows on the lines of theorem 2.1.

Theorem 3.2: At random epochs, the marginal probability generating functions when the server is in idle state, busy with FES, busy with ith $(0 \le i \le k)$ phase service, under jth $(1 \le j \le d)$ phase repair while broken down during i^{th} $(0 \le i \le k)$ phase service, under set up and on vacation respectively, are

$$D(z) = \frac{D(0,z)(1-\tilde{a}(\lambda))}{\lambda} \tag{3.30}$$

$$P_0(z) = \frac{P_0(0,z)(1-\tilde{b}_0(M_0(z))}{M_0(z)} \tag{3.31}$$

$$P_i(z) = \frac{p_i\tilde{b}_0(M_0(z))P_0(0,z)(1-\tilde{b}_i(M_i(z)))}{M_i(z)}, (1 \le i \le k) \tag{3.32}$$

$$R_{i,j}(z) = \frac{\alpha_i P_i(z) H_{i,j}(z)(1-\tilde{g}_{i,j}(\lambda(1-C(z)))}{(\lambda(1-C(z)))}, (0 \le i \le k), (1 \le j \le d) \tag{3.33}$$

$$S(z) = \frac{\alpha_i P_i(z)(1-\tilde{\eta}(\lambda(1-C(z)))}{(\lambda(1-C(z))}, (1 \le i \le k) \tag{3.34}$$

$$V(z) = (1-\theta)\sum_{i=0}^{k} p_i\tilde{b}_i(M_i(z))P_0(0,z)\frac{[1-\tilde{w}(\lambda(1-C(z)))]}{\lambda(1-C(z))} \tag{3.35}$$

Proof: The proof follows on the lines of theorem 2.2.

Theorem 3.3: The generating function for the number of customers in the retrial queue is

$$\begin{aligned} K(z) = D_0 &+ \frac{D(0,z)(1-\tilde{a}(\lambda))}{\lambda} + \frac{P_0(0,z)(1-\tilde{b}_0(M_0(z)))}{M_0(z)} \\ &+ \frac{\sum_{i=1}^{k} p_i\tilde{b}_0(M_0(z))(1-\tilde{b}_i(M_i(z)))P_0(0,z)}{M_i(z)} + \sum_{i=0}^{k}(1-\theta)p_i\tilde{b}_i(M_i(z))P_0(0,z)\frac{(1-\tilde{w}(\lambda(1-C(z))))}{\lambda(1-C(z))} \\ &+ \sum_{i=0}^{k}\sum_{j=1}^{d}\frac{\alpha_i P_i(z)H_{i,j}(z)(1-\tilde{g}_{i,j}(\lambda(1-C(z)))}{(\lambda(1-C(z)))} + \sum_{i=0}^{k}\frac{\alpha_i P_i(z)(1-\tilde{\eta}(\lambda(1-C(z)))}{(\lambda(1-C(z))}, (1 \le j \le d) \end{aligned} \tag{3.36}$$

Proof: The probability generating function for the number of customers in the retrial orbit is obtained using

$$K(z) = D_0 + D(z) + P_0(z) + \sum_{i=1}^{k} P_i(z) + \sum_{i=0}^{k}\sum_{j=1}^{d} R_{i,j}(z) + S(z) + V(z), \tag{3.37}$$

Theorem 3.4: The generating function for the number of customers present in the system, is

$$L(z)=D_0+\frac{D(0,z)(1-\tilde{a}(\lambda))}{\lambda}+\frac{zP_0(0,z)(1-\tilde{b}_0(M_0(z)))}{M_0(z)}$$

$$+\frac{\sum_{i=1}^{k} zp_i\tilde{b}_0(M_0(z))(1-\tilde{b}_i(M_0(z)))P_0(0,z)}{M_i(z)}+\sum_{i=0}^{k}(1-\theta)p_i\tilde{b}_i(M_i(z))P_0(0,z)\frac{(1-\tilde{w}(\lambda(1-C(z))))}{\lambda(1-C(z))}$$

$$+z\sum_{i=0}^{k}\sum_{j=1}^{d}\frac{\alpha_i P_i(z)H_{i,j}(z)(1-\tilde{g}_{i,j}(\lambda(1-C(z)))}{(\lambda(1-C(z)))}+z\sum_{i=0}^{k}\frac{\alpha_i P_i(z)(1-\tilde{\eta}(\lambda(1-C(z)))}{(\lambda(1-C(z))},(1\le j\le d) \qquad (3.38)$$

Proof: The generating function of the number of customers in the system is obtained by using results of marginal generating functions given by

$$L(z)=D_0+D(z)+zP_0(z)+z\sum_{i=1}^{k}P_i(z)+z\sum_{i=0}^{k}\sum_{j=1}^{d}R_{i,j}(z)+zS(z)+V(z) \qquad (3.39)$$

3.5 SPECIAL CASES

In the present section, we deduce some special cases by setting appropriate parameters. Now some special cases are deduced as:

(i) Bulk arrival M/G/1 queue with unreliable server and single vacation

On setting, γ=0, r =0, α_i=0 (1≤ i≤ k), d=1, p_i=0, $\tilde{\eta}=1$,

our model reduces to that investigated by Haridass and Arumuganathan (2008).

(ii) M/G/1 retrial queue with impatient customers and multi optional services

On setting, C (z) =z, $C'(z)=1$, $\tilde{w}=1$, $\tilde{\eta}=1$, $d=1, \theta=0$, our results coincide with the results obtained by Wang and Li (2009).

(iii) M/G/1 unreliable retrial queueing system with two phase service without vacation and impatient customers

By substituting, C (z) =z, $C'(z)=1, \tilde{w}=1, \tilde{\eta}=1, d=1, \theta=0, r=0, k=1,$ our results correspond to those of Choudhury and Deka (2008).

(iv) M/M/1 reliable queue without vacation and single service

On substituting, $\theta=0, \alpha_i=0, \beta_j=0, p_i=0, r=0, \gamma=0, \tilde{\eta}=1, \tilde{w}=1, C(z)=z, C'(z)=1,$ the explicit results for the queue length coincide with the results for classical M/M/1 model of Gross and Harris (1985).

(v) M/G/1 queue with second optional service and no breakdowns

On setting i=1,γ=0, C (z) =z, $C'(z)=1, \tilde{\eta}=1, \tilde{w}=1, \theta=0$, r=0, $\alpha_i = 0$ (0≤ i≤ k) in our model and considering FES being general distributed and SOS as exponentially distributed, we get the results as obtained by Madan (2000).

3.6 PERFORMANCE MEASURES

Some performance indices characterizing the queue size, long run probabilities and reliability issues are derived below in various categories as follows:

(A) Queueing Measures

The computation of performance measures namely average system size and queue length of the orbit are the key indices which determine the effectiveness and validity of any retrial system.

Theorem 3.5: The mean queue length of the retrial orbit (L_R) and mean queue length of the system (L_S) are

$$L_R = \frac{D_0(1-\tilde{a}(\lambda))(A'-1)B'' - 2A'D'B' - B'A''}{2(A'+D'-1)^2} + \frac{\lambda D_0(a''b''' - a'''b'')}{12(A'+D'-1)^2 Q'^2} + \frac{\sum_{i=1}^{k}\lambda p_i D_0(a''d''' - a'''d'')}{12(A'+D'-1)^2 Q'^2}$$

$$+ \frac{\sum_{i=0}^{k}\alpha_i \lambda p_i D_0(5f^{iv}h''' - 5h^{iv}f''')}{624(A'+D'-1)^2 Q'^2 \lambda^2 (C'(1))^2} + \frac{\sum_{i=0}^{k}(1-\theta)\lambda p_i D_0(j''k''' - j'''k'')}{12(A'+D'-1)^2 N'^2} + \frac{\sum_{i=0}^{k}\alpha_i \lambda p_i D_0(5m'''n^{iv} - 5m^{iv}n''')}{240 N'^2 (A'+D'-1)^2 Q'^2}$$

$$L_S = L_R + \frac{\lambda D_0 (A-E)'\bar{F}'}{(A'+D'-1)Q'} + \frac{\sum_{i=1}^{k}\lambda D_0 p_i (A-E)'F'\bar{G}'}{(A'+D'-1)Q'} + \frac{\sum_{i=0}^{k} 2\alpha_i \lambda D_0 p_i (3d''\bar{M}'H_j(z))}{(a''N')} + \frac{\sum_{i=0}^{k}\alpha_i \lambda D_0 p_i (d''\bar{J}')}{(a''N')}$$

where,

$$X' = \left[\frac{1}{\mu_0} + \sum_{i=1}^{k}\frac{p_i}{\mu_i}\right]\left(\lambda C'(1) + \alpha_i H^{(i)\prime}(1)\right); Y' = \sum_{i=1}^{k}\frac{p_i}{\mu_i}\left(\lambda C'(1) + \alpha_i H^{(i)\prime}(1)\right) + \frac{\sum_{i=1}^{k}\lambda C'(1) p_i}{\xi}$$

$$A = 1 + r(z-1)(1-\tilde{a}(\lambda));\; A' = r(1-\tilde{a}(\lambda));\; B = 1-\theta X(z) - (1-\theta)Y(z);\; B' = -\theta X'(1) - (1-\theta)Y'(1)$$

$$D = 1-B,\; D' = -B';\; E = z, E' = 1;\; F = \tilde{b}_0(M_0(z)), F' = (\lambda C'(1) + \alpha_i H^{(i)\prime}(1))/\mu_0$$

$$G = \tilde{b}_i(M_i(z)), G' = \frac{1}{\mu_i}\left(\lambda C'(1) + \alpha_i H^{(i)\prime}(1)\right);\; I = \tilde{w}(\lambda(1-C(z)), I' = \frac{\lambda C'(1)}{\psi},$$

$$H^{(i)\prime}(1) = \lambda C'(1)\left[\frac{1}{\xi} + \sum_{r=1}^{d}\frac{1}{\beta_r}\right], H_{i,j}(1) = \lambda C'(1)\left[\frac{1}{\xi} + \sum_{r=1}^{j-1}\frac{1}{\beta_r}\right]$$

$$Q = M_i(z), Q' = (-\lambda C'(1) - \alpha_i H^{(i)\prime}(1));\; S = \tilde{w}(\lambda(1-C(z)), S' = (\lambda C'(1)/\theta);$$

$$J=\tilde{\eta}(\lambda(1-C(z)), J'=\frac{\lambda C'(1)}{\xi};$$
$$N=\lambda(1-C(z)), N'=-\lambda C'(z);$$
$$M=\tilde{g}_j(\lambda(1-C(z)), M'=\frac{\lambda C'(1)}{\beta_j}$$
$$a=(AD-E)Q, a''=2(A'+D'-1)Q', a'''=3(AD-E)''Q'+3(AD-E)'Q''$$
$$b=(A-E)\bar{F}, b''=2(A-E)'\bar{F}', b'''=3(A-E)'\bar{F}'',$$
$$d=(A-E)F\bar{G}, d''=2(A-E)'F\bar{G}', d'''=6(A-E)'F'\bar{G}';\ h=aN, h'''=3a''N', h^{iv}=(4a'''N'+6a''N'')$$
$$f=d(1-M)H_{i,j}(z), f'''=3d''\bar{M}'(H_{i,j}(z)), f^{iv}=(12d''\bar{M}'H'_{i,j}(z))$$
$$s=(AD-E)N, s''=2(AD-E)'N', s'''=3(AD-E)''N'+3(AD-E)'N''$$
$$t=(A-E)M\bar{S}, t''=2(A-E)'M\bar{S}', t'''=5(A-E)'M\bar{S}'+3(A-E)'M\bar{S}''+(A-E)'M\bar{S}''$$
$$m=aN, m'''=3a''N', m^{iv}=4a'''N'+6a''N'', n=d\bar{J}, n'''=3d''\bar{J}', n^{iv}=4d'''\bar{J}'+6d''\bar{J}''$$

Proof: The mean queue length of the retrial orbit and mean queue length of the system are derived by using

$$L_R=\lim_{Z\to 1}K'(z),\ L_S=\lim_{Z\to 1}L'(z).$$

Theorem 3.6: The mean waiting time (WT) of the customers in the system is

$$WT=L_s/\lambda$$

Proof: The mean waiting time is evaluated by using Little's formula

Mean queue length = Mean waiting time $\times$ Effective arrival rate

(B) Long Run Probabilities

Here, we derive the explicit expressions for the long run probabilities of the server being in different states. To evaluate the status of the server, we employ the corresponding generating function and results are given in the following theorem.

Theorem 3.7: The long run probabilities of the server being in idle (P_I), busy (P_B), under set up (P_S), repair (P_R), and vacation (P_V) states respectively, are given below as-

$$P_I=\frac{(1-\tilde{a}(\lambda))D_0B'}{(A'+B'-1)} \quad (3.40)$$

$$P_B=\sum_{i=1}^{k}\frac{\lambda D_0(A-E)'}{(A'+D'-1)Q'}\left[F'+\sum_{i=1}^{k}p_iF'\bar{G}'\right] \quad (3.41)$$

$$P_S=\sum_{i=1}^{k}\alpha_i p_i\lambda D_0\frac{(d''\bar{J}')}{(a''N')} \quad (3.42)$$

$$P_R=\sum_{i=1}^{k}\sum_{j=1}^{d}2\alpha_i p_i\lambda D_0\frac{d'\bar{M}'(H_{i,j}{}'(1))}{a''N'} \quad (3.43)$$

$$P_V = 2\sum_{i=0}^{k}(1-\theta)\lambda p_i \frac{(A-E)'M\bar{S}'}{(A'+D'-1)N'} \tag{3.44}$$

Proof: The proof follows on the lines of theorem 2.5.

(C) Reliability Measures

The reliability indices of interest viz. availability and failure frequency are given in the following theorem:

Theorem 3.8: The steady state availability (A_v) and failure frequency (F_f) of the server are

$$A_v = \frac{D_0}{(1-A'-D')}\left[1-A'+B'\tilde{a}(\lambda)+(1-r(1-\tilde{a}(\lambda)))(\rho_0+\sum_{i=1}^{k}p_i\rho_i)\right] \tag{3.45}$$

$$F_f = \frac{D_0}{(1-A'-D')}\left[(1-r(1-\tilde{a}(\lambda)))(\alpha_0\rho_0+\sum_{i=1}^{k}\alpha_i p_i\rho_i)\right] \tag{3.46}$$

Proof: The results given in equations (3.45) and (3.46) are obtained in the similar manner as in theorem 2.8.

3.7 NUMERICAL ILLUSTRATION

The present section deals with the numerical results of our model to examine the effects of various parameters on the performance indices. The batch size, retrial time, repair time and setup time have been assumed to be exponentially distributed. The computational work has been done by considering two optional services i.e. *k*=2 and two phase repair system i.e. *d*=2. The numerical simulation has been performed by assuming service distribution to be Erlangian-5 (E_5) distributed. The numerical results have been summarized in tables 3.1-3.3 and displayed in figures 3.1-3.4. The sensitivity of performance indices with respect to different parameters are explained below.

(A) Effect of Key Parameters on the Performance Indices

The default parameters for evaluating the computational results are taken as follows: $\lambda = 0.5,\ \alpha_0 = \alpha_1 = \alpha_2 = 0.01, \beta = \beta_0 = 0.08 = \beta_1 = \beta_2,\ \gamma = 0.1,\ \mu_0 = \mu_1 = \mu_2 = 5,\ p_0 = 0.4,$ $p_1 = 0.2, p_2 = 0.4,\ r = 0.1, \xi = 0.3,\ \psi = 0.3,\ \theta = 0.8.$

Table 3.1 shows the effect of μ on the performance indices for varying values of set up rate (ξ) and repair rate (β). The data shown in table 3.1 clearly indicates that P_I, P_B,

Table 3.1: Effect of μ with varying values of (ξ, β) on various performance indices

μ	(ξ, β)	P_I	P_B	P_R	P_S	P_V	WT	F_f
5	(0.1,0.04)	0.0780	0.0121	0.0628	0.0024	1.4046	52.41	0.0013
	(0.1,0.06)	0.0808	0.0165	0.0407	0.0024	1.4544	42.63	0.0013
	(0.1,0.08)	0.0820	0.0187	0.0301	0.0024	1.4806	38.14	0.0013
	(0.1,0.10)	0.0827	0.0199	0.0239	0.0024	1.4968	35.57	0.0013
	(0.2,0.10)	0.1022	0.0246	0.0295	0.0029	0.7484	14.51	0.0016
	(0.3,0.10)	0.1093	0.0263	0.0316	0.0032	0.4989	9.68	0.0017
	(0.5,0.10)	0.1152	0.0278	0.0333	0.0033	0.2994	6.61	0.0018
7	(0.1,0.04)	0.0782	0.0090	0.0665	0.0017	1.3663	49.75	0.0010
	(0.1,0.06)	0.0806	0.0124	0.0435	0.0017	1.3996	40.76	0.0010
	(0.1,0.08)	0.0817	0.0140	0.0323	0.0017	1.4168	36.28	0.0010
	(0.1,0.10)	0.0823	0.0150	0.0257	0.0017	1.4274	33.63	0.0010
	(0.2,0.10)	0.1024	0.0187	0.0320	0.0022	0.7137	13.72	0.0012
	(0.3,0.10)	0.1098	0.0200	0.0343	0.0023	0.4758	9.10	0.0013
	(0.5,0.10)	0.1159	0.0212	0.0362	0.0025	0.2855	6.12	0.0014
9	(0.1,0.04)	0.0781	0.0072	0.0687	0.0014	1.3459	48.33	0.0008
	(0.1,0.06)	0.0802	0.0099	0.0452	0.0014	1.3709	39.80	0.0008
	(0.1,0.08)	0.0811	0.0112	0.0336	0.0014	1.3837	35.33	0.0008
	(0.1,0.10)	0.0817	0.0120	0.0268	0.0014	1.3915	32.65	0.0008
	(0.2,0.10)	0.1020	0.0150	0.0334	0.0017	0.6958	13.33	0.0009
	(0.3,0.10)	0.1095	0.0161	0.0359	0.0019	0.4638	8.82	0.0010
	(0.5,0.10)	0.1158	0.0171	0.0380	0.0020	0.2783	5.89	0.0011

Table 3.2: Effect of α with varying values of (r, γ) on various performance indices

α	(r, γ)	P_I	P_B	P_R	P_S	P_V	WT	F_f
0.01	(0.08,0.1)	0.1077	0.0246	0.0401	0.0031	0.4893	10.83	0.0017
	(0.1,0.1)	0.1084	0.0246	0.0398	0.0031	0.4935	10.97	0.0017
	(0.3,0.1)	0.1133	0.0249	0.0359	0.0032	0.5526	12.95	0.0017
	(0.5,0.1)	0.1110	0.0253	0.0296	0.0032	0.6804	17.20	0.0018
	(0.1,0.08)	0.1114	0.0245	0.0395	0.0031	0.4943	10.98	0.0017
	(0.1,0.3)	0.0851	0.0260	0.0424	0.0033	0.4883	10.86	0.0018
	(0.1,0.5)	0.0700	0.0268	0.0440	0.0034	0.4854	10.80	0.0019
0.02	(0.08,0.1)	0.1002	0.0135	0.0840	0.0062	0.4587	10.98	0.0034
	(0.1,0.1)	0.1010	0.0135	0.0835	0.0062	0.4619	11.12	0.0034
	(0.3,0.1)	0.1085	0.0136	0.0770	0.0062	0.5052	13.10	0.0034
	(0.5,0.1)	0.1123	0.0138	0.0665	0.0063	0.5925	16.98	0.0035
	(0.1,0.08)	0.1040	0.0134	0.0829	0.0061	0.4625	11.15	0.0034
	(0.1,0.3)	0.0784	0.0141	0.0880	0.0064	0.4580	10.92	0.0035
	(0.1,0.5)	0.0641	0.0145	0.0908	0.0066	0.4557	10.80	0.0036
0.03	(0.08,0.1)	0.0904	0.0019	0.1315	0.0091	0.4317	14.32	0.0050
	(0.1,0.1)	0.0913	0.0019	0.1309	0.0091	0.4340	14.62	0.0050
	(0.3,0.1)	0.1006	0.0019	0.1229	0.0092	0.4653	18.52	0.0050
	(0.5,0.1)	0.1090	0.0019	0.1102	0.0092	0.5246	25.20	0.0051
	(0.1,0.08)	0.0941	0.0019	0.1301	0.0090	0.4345	14.67	0.0050
	(0.1,0.3)	0.0702	0.0020	0.1366	0.0094	0.4312	14.26	0.0052
	(0.1,0.5)	0.0569	0.0020	0.1401	0.0096	0.4295	14.04	0.0053

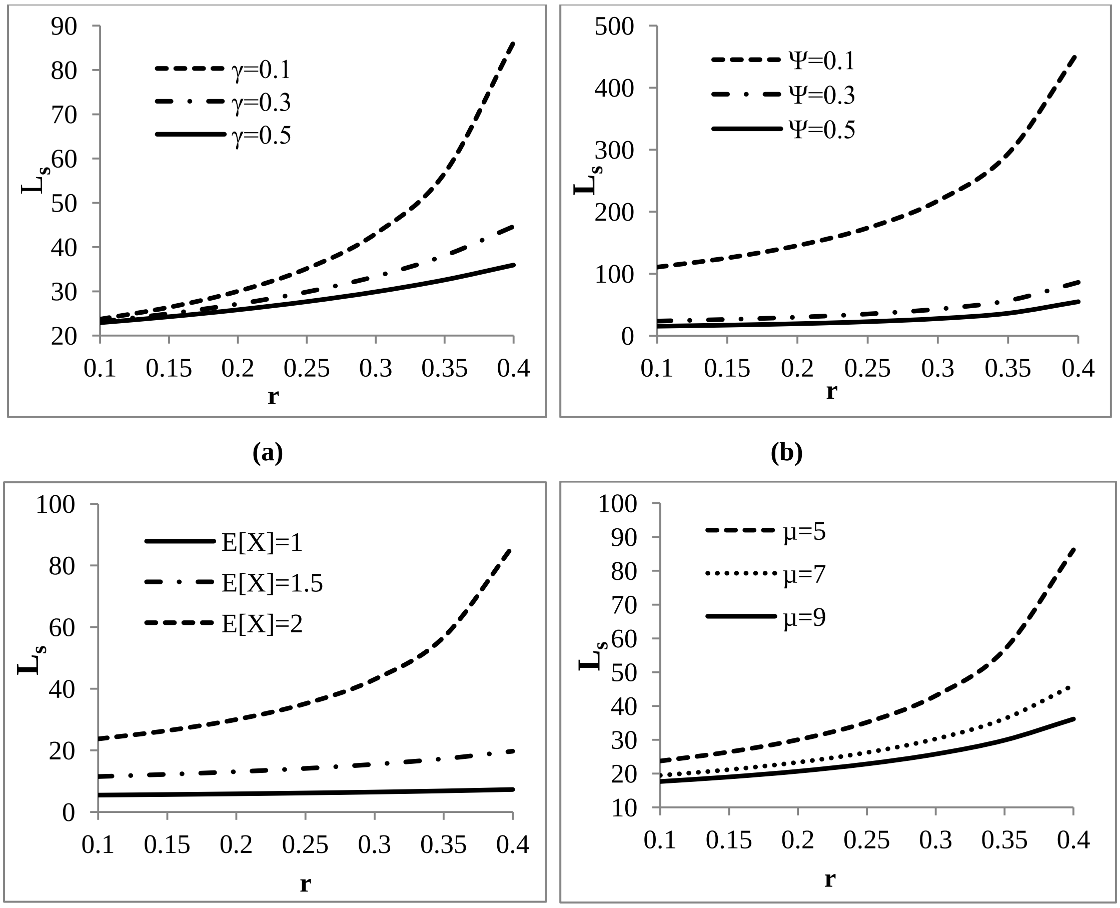
Table 3.3: Effect of λ with varying values of (ψ, β) on various performance indices

λ	(ψ, β)	P_I	P_B	P_R	P_S	P_V	WT	F_f
0.1	(0.1,0.06)	0.0532	0.0009	0.0172	0.0024	0.0776	5.01	0.0004
	(0.1,0.08)	0.0539	0.0011	0.0129	0.0024	0.0778	4.63	0.0004
	(0.1,0.1)	0.0543	0.0012	0.0103	0.0024	0.0779	4.41	0.0004
	(0.2,0.1)	0.0347	0.0013	0.0107	0.0012	0.0728	2.68	0.0004
	(0.3,0.1)	0.0273	0.0014	0.0109	0.0008	0.0712	2.21	0.0004
	(0.5,0.1)	0.0209	0.0014	0.0110	0.0005	0.0700	1.88	0.0004
0.3	(0.1,0.06)	0.1174	0.0074	0.0311	0.0066	0.3531	13.05	0.0012
	(0.1,0.08)	0.1177	0.0087	0.0231	0.0066	0.3577	11.74	0.0012
	(0.1,0.1)	0.1179	0.0095	0.0184	0.0066	0.3605	11.06	0.0012
	(0.2,0.1)	0.1012	0.0106	0.0236	0.0032	0.2700	6.82	0.0012
	(0.3,0.1)	0.0866	0.0109	0.0252	0.0021	0.2492	5.83	0.0012
	(0.5,0.1)	0.0711	0.0112	0.0265	0.0012	0.2347	5.21	0.0011
0.5	(0.1,0.06)	0.0782	0.0181	0.0223	0.0099	1.2314	24.34	0.0018
	(0.1,0.08)	0.0758	0.0212	0.0160	0.0099	1.2894	21.84	0.0018
	(0.1,0.1)	0.0743	0.0231	0.0125	0.0099	1.3270	20.70	0.0018
	(0.2,0.1)	0.1138	0.0256	0.0270	0.0048	0.5912	11.47	0.0018
	(0.3,0.1)	0.1093	0.0263	0.0316	0.0032	0.4989	9.68	0.0017
	(0.5,0.1)	0.0973	0.0269	0.0352	0.0019	0.4436	8.61	0.0017

P_V increase with the increase in repair rate (β) while other performance indices show decreasing trend. The failure frequency F_f remains almost constant for the increased values of β. The waiting time decreases with the increase in service rate μ. We notice that as set up rate (ξ) increases, long run probabilities P_I, P_B, P_R and P_S show an increasing trend while WT and P_V decrease.

Table 3.2 displays the effect of breakdown rate (α) for varying values of r and retrial rate (γ) by keeping fixed values of other key parameters. It can be noticed that an increment in breakdown rate (α) causes an increase in the values of WT, P_R, P_S and F_f whereas P_I, P_B and P_V decrease. We observe that P_I, P_B and P_V increase whereas P_R decreases as we increase r. Even for the varying values of γ, the long run probabilities P_I, P_R and waiting time decrease whereas other performance indices show the opposite behavior.

The effect of arrival rate λ for the varying values of vacation rate (ψ) and repair rate (β) is displayed in table 3.3. As the arrival rate increases from 0.1 to 0.5, the long run probabilities P_B, P_S, F_f, P_V and waiting time increase while P_I and P_R show the fluctuating trend. An increase in the vacation rate (ψ) for constant λ leads to an increase in P_B and P_R whereas P_B, P_V and WT show decreasing pattern. The waiting time increases drastically with an increase in the arrival rate λ. This is due to the fact that an increase in the arrival rate leads to an increase in the number of customers in the system which automatically increases the waiting time.



(c) (d)

Fig. 3.1: Effect of r on the queue length (L_s)

The simultaneous increase in both ψ and λ reveal that P_R, P_B and P_S increase while P_V decreases. Moreover an increase in β leads to an increment in P_B. But simultaneous increase in λ and β increases P_B manifold; the reason for this fact is that an increase in λ keeps the server more occupied due to the arrival of more customers. It is observed that the waiting time decreases with an increase in β which is quite obvious, as an increase in the repair rate makes the server available for the servicing of the customers and hence reduces the waiting time.

(B) Average Queue Length (L_s)

Now we demonstrate the effect of reneging probability (1-r) on the average queue length (L_s) of the system with respect to other key parameters. Figs 3.1(a-d) exhibit the trends for L_s with respect to r to judge the sensitivity. The analytical result of the queue length has been simulated numerically for this purpose. Fig. 3.1 (a) illustrates the effect of retrial rate γ for the increasing values of r. It is quite interesting to note that L_s increases abruptly as r increases and γ decreases. It is very obvious that with an increase in r, probability (1-r) decreases which in turn increases the number of customers.

Fig. 3.1(b) depicts the effect of r with varying values of vacation rate ψ on L_s. The queue length decreases as the vacation rate increases; the reason behind this is that when a server goes for vacation, there is no customer in the system which implies the reduction in the number of customers in the system. Figs 3.1(c) and 3.1(d) depict the behavior of the expected batch size E[X] and service rate μ respectively on the L_s by varying r. L_s increases with an increase (decrease) in batch size E[X] (μ).

(C) ANFIS Results of L_s

For complex queueing models, it is difficult to obtain exact results in explicit form which are computationally manageable. In the present section, the analytical results are compared with the approximate results obtained by developing the Adaptive Network Based Fuzzy Inference System (ANFIS) using Neuro-fuzzy tool in MATLAB. The Gaussian function has been used to depict the membership function of the fuzzy parameters. The ANFIS is trained for 5 epochs and the results are computed by considering 5 linguistic values (say very low, low, average, high, and very high) for the input parameter λ. The Gaussian membership function for λ is shown in fig. 3.2. To compare the numerical results obtained using analytical queue length (L_s), various graphs have been plotted against different key parameters. Fig. 3.3 depicts the sensitivity of L_s towards different parameters. ANFIS results plotted are shown by discrete lines in figs. 3.3 (a-d). To compare these results with the exact results, we have plotted the analytical results by continuous lines in the same figures.

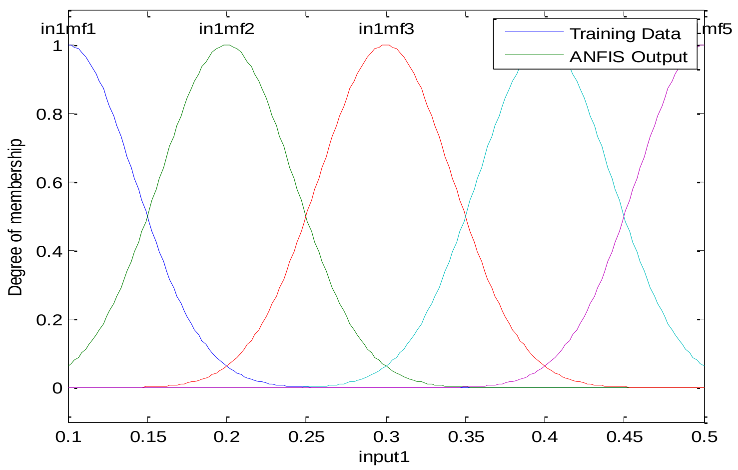


Fig. 3.2: Membership Function

Fig 3.3(a) demonstrates the effect of batch size E[X] on L_s which increases considerably by increasing E[X]. This is quite obvious as the number of customers will increase if the batch size increases. Figs 3.3(b)-3.3(d) illustrate the behavior of L_s with the other significant parameters like set up rate (ξ), repair rate (β) and vacation rate (ψ). It is noticed from fig. 3.3(b) that L_s decreases as the set up rate (ξ) increases; an increase in ξ speeds up the repair process which automatically decreases the number of customers in the system. Also, an increase in the repair rate β reduces L_s as shown by fig. 3.3(c).

Similarly, an increase in the vacation rate (ψ) decreases the number of customers in the system; this may be due to frequent visits of the server for vacation. Along with the exact results, the results obtained by neuro fuzzy technique namely ANFIS are also plotted in figures 3.3(a- d). It is found that the continuous and discrete curves almost overlap which demonstrates that exact analytical results are in good agreement with that of ANFIS results.

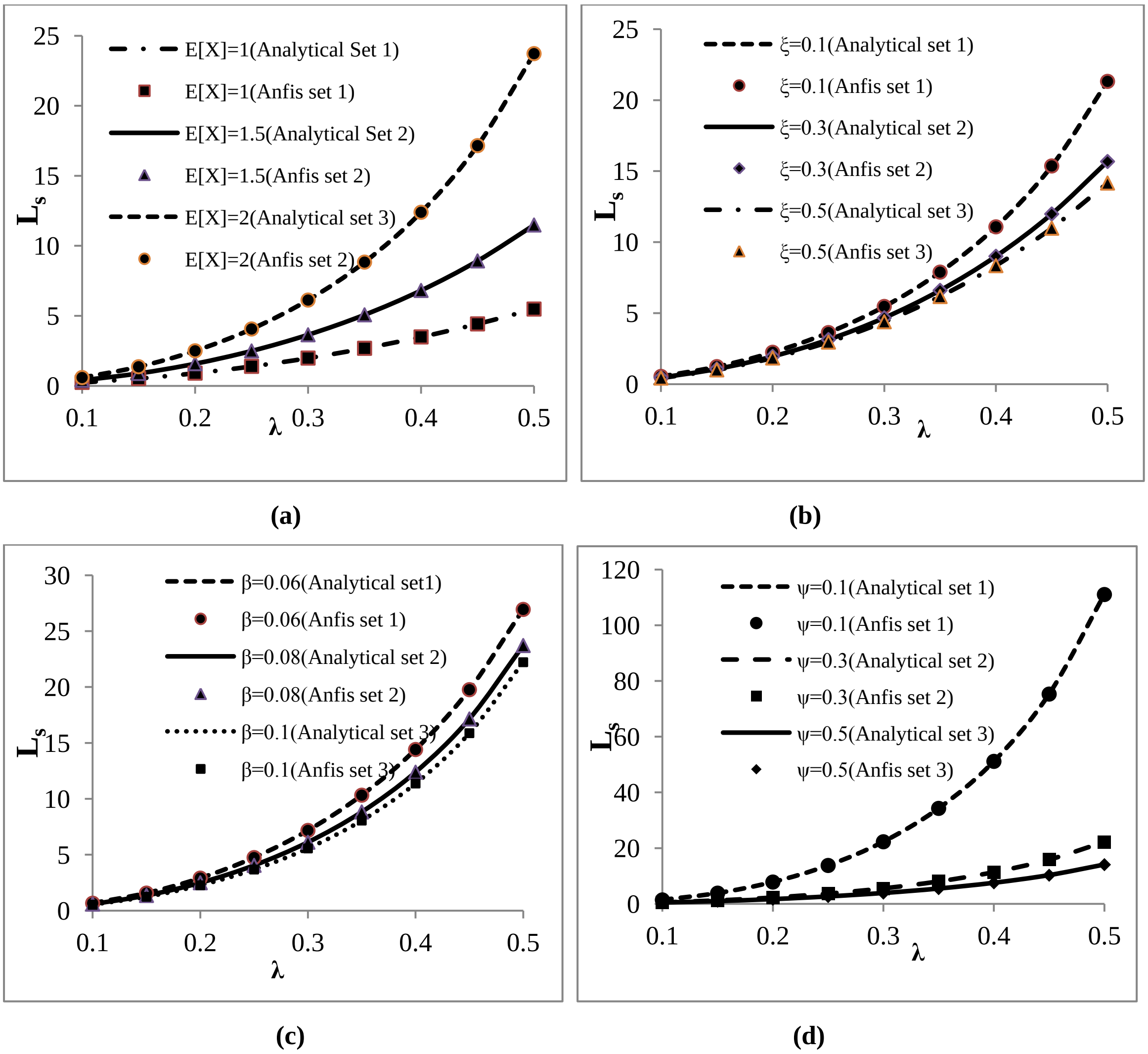


Fig. 3.3: Variation in L_s with λ for various parameters

(D) ANFIS Results of A_V

One of the important reliability measure namely availability (A_V) of the server has been plotted using the ANFIS approach. Various figures have been drawn to display the availability vs. λ for both exact (continuous lines) and neuro fuzzy (discrete lines) approach for different values of parameters viz. α, E[X], γ and r in fig. 3.4(a-d), respectively. The availability of the server decreases as the breakdown rate α increase as shown in fig. 3.4(a). Figs 3.4(b) and 3.4(d) display that an increment in E[X] and r lead to the decrement in the availability of the server. From fig. 3.4(c), it is clear that as the retrial rate (γ) increases, more and more customers try for the service which in turn reduces the availability of the server to all the customers. It can be observed that the plots for L_s as well as A_v obtained by ANFIS are very close (almost overlapping) with that obtained by using analytical explicit expressions.

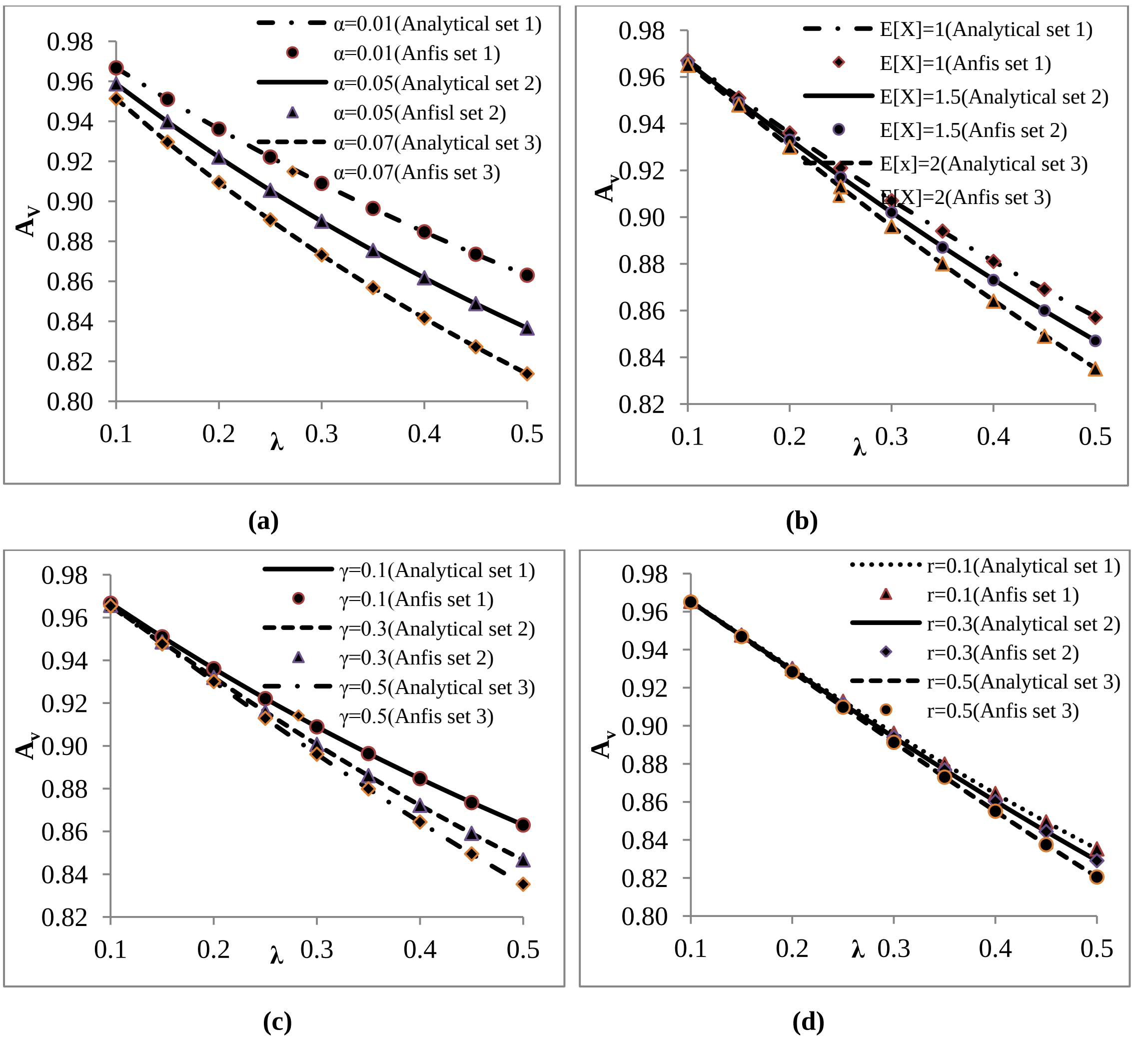


Fig. 3.4: Variation in A_v with λ for various parameters

3.8 DISCUSSION

The retrial queueing model developed in this paper depicts the queueing scenario dealing with the impatient customers, phase repair and multioptional services. Now based on numerical simulation and experience with ANFIS approach, we can conclude that-

- The total number of customers in the system increases (decreases) with an increase in the 'r' and batch size (repair rate, vacation rate and set up rate). The queue length of the concerned queueing model can be controlled by reducing the batch size and increasing the repair rate.
- As expected, the availability of the server decreases by increasing the breakdown rate α, batch size and retrial rate γ. It is found that the control over the parameters like repair rate and γ can make the server more available to serve the incoming customers.
- It is noticed that both analytical and ANFIS results are in strong agreement with each other which demonstrates that computing approach based on ANFIS can be successfully employed for the prediction of the queueing and reliability indices in many real time complex systems. The numerical simulation done by using the classical SVT approach are consistent with those obtained by soft computing approach based on ANFIS approach.

CHAPTER 4

MODIFIED VACATION POLICY FOR BULK RETRIAL QUEUE

4.1 INTRODUCTION

A variety of vacation policies have been developed by the queue theorists during the course of their research since long. This is due to the fact that different real life congestion situations are dealt by the server in different manners. In chapter 3, we dealt with Bernoulli vacation schedule policy of the server. Another vacation policy namely modified vacation policy has been studied in this chapter. The server can avail only a fixed number of vacations of random length in case no customer is present in the system for the service. Chang and Ke (2009) considered a batch retrial model where the server can take at most J vacations; the customer can go for a series of continuous J vacations if no customers/jobs are available in the orbit. Ke and Chang (2009b) investigated modified vacation policy for M/G/1 retrial queue and obtained various performance measures. Bhargava and Jain (2014) investigated unreliable multiserver queueing system with modified vacation policy.

The pattern of servicing also plays a significant role in the modeling of retrial queues. In the queue literature, various kinds of services like single service, optional services, multioptional services, phase services have been studied. Choudhury and Tadj (2011) studied the optimal control of bulk arrival M/G/1 unreliable server with two phases of service and Bernoulli vacation schedule. The servicing in phases clearly relates to many realistic day-to-day situations. The admission in any institute requires a number of formalities and filling of a number of forms. This process is completed in various compulsory phases from getting the form to the submission of completed application form. Various intermediate steps like submission of medical fitness certificates, previous educational qualifications proof, completion of admit card, fees deposition, etc. are also involved in the completion of the process. All these are compulsory phases and admission cannot be completed if any of these compulsory steps is skipped. Jain and Agrawal (2010) analyzed a batch arrival queueing system with N-policy and Bernoulli vacation schedule wherein the customer undergoes l-essential stage service procedure to avail the service.

In this chapter, bulk arrival retrial vacation queue with unreliable server has been studied. Both service process and repair process are done in fixed compulsory phases in succession. Moreover, the server takes some time to start the repair so as to make some preliminary settings, known as setup time. Moreover, the concept of modified vacation has been incorporated along with the discouragement behavior of the customers. Section 4.2 describes the requisite assumptions to formulate the model. The governing equations along with the boundary conditions and generating functions of the queue size distribution are obtained in section 4.3. The performances measures are derived in section 4.4. Section 4.5 deals with the maximum entropy analysis of the retrial model. Section 4.6 is devoted to the sensitivity analysis which is carried out by taking numerical illustration. Finally, the conclusions have been drawn in section 4.7.

4.2 MODEL DESCRIPTION

Consider bulk arrival M/G/1 retrial queue model with unreliable server, reneging and vacation policy. The server is unreliable and may breakdown in an exponential manner while providing service to the customers. The broken down server is repaired following general distribution in *d*-compulsory phases as considered in chapter 3. The other assumptions underlying the model are as follows:

- **Arrival Process**

The customers arrive in batches in the system following Poisson distribution with state dependent arrival rate λ_I depending on the server's status; 'I' takes value 1, 2, 3, 4 and 5 when the server is in retrial state, busy state, setup state, repair state and in vacation state, respectively. The incoming customers are served if they find the server idle otherwise they are forced to join the virtual pool of the customers called orbit from where they can try again for the service with exponential retrial rate γ.

- **Service Process**

If an incoming batch of the customers finds the server in idle state, then a customer at the head of the batch joins the server to get served. All the customers are served in k essential phases with service rate μ_i $(1 \le i \le k)$ for a customer availing i^{th} phase of service.

- **Vacation Policy**

If no more customers are present in the system, then the server takes at most J vacations repeatedly with rate ψ_l $(1 \le l \le J)$ for l^{th} vacation and returns back if at least one job is found in the orbit after returning from the vacation. This process repeats again if no more jobs are available in the system i.e. the server may reactivate at the end of l^{th} ($1\le l \le J$) vacation if any customer/job is available in the system. But the server remains dormant in the system if no job is present in the system at the end of J^{th} vacation.

4.3 QUEUE SIZE DISTRIBUTION

Let N(t) represents the number of customers in the system and $S_1(t)$, $S_2(t)$ and $S_3(t)$ denote the phase of the service, phase of repair and state of vacation respectively, at any time t.

The state of the server at any time t is given by

$$Y(t)=\begin{cases}1, & \text{server is in idle state}\\ 2, & \text{server is busy in providing service to the customers}\\ 3, & \text{server is broken down and under setup before repair}\\ 4, & \text{server is broken down and under repair}\\ 5, & \text{server is in vacation state}\end{cases}$$

In the steady state, the joint distributions of the server state and the queue size are defined as-

$$D_n = \lim_{t\to\infty}\Pr\{Y(t)=1, N(t)=n\}, n\ge 0$$

$$P_{i,n}(x) = \lim_{t\to\infty}\Pr\{Y(t)=2, x\le \varpi(t)\le x+dx, N(t)=n, S_1(t)=i\}, n\ge 0, (1\le i\le k)$$

$$S_{i,n}(x,y) = \lim_{t\to\infty}\Pr\{Y(t)=3, \varpi(t)=x, y\le\sigma(t)\le y+dy, N(t)=n, S_1(t)=i, S_2(t)=j\},\ n\ge 0,\ (1\le i\le k),\ (1\le j\le d)$$

$$R_{i,j,n}(x,y) = \lim_{t\to\infty}\Pr\{Y(t)=4, \varpi(t)=x, y\le\sigma(t)\le y+dy, N(t)=n, S_1(t)=i, S_2(t)=j\},\ n\ge 0, (1\le i\le k), (1\le j\le d)$$

$$V_{l,n}(x) = \lim_{t\to\infty}\Pr\{Y(t)=5, x\le\varpi(t)\le x+dx, N(t)=n, S_1(t)=i, S_2(t)=j, S_3(t)=l\},\ n\ge 0, (1\le i\le k), (1\le j\le d), (1\le l\le J)$$

4.3.1 Mathematical Formulation

Before framing the governing equations for the model, we give the proposition stating the stability condition for the model as follows:

Proposition 4.1: The necessary and sufficient condition for the system to be stable is

$$r(1-\tilde{a}(\lambda_1))+Y'<1$$

where, $Y'=\prod_{q=1}^{k} b_q^{(1)} H_q'(1),\ H_i'(z)=-\lambda_2 C'(1)-\alpha_i M_i'(1),$

$$M_i'(z) = \eta_i^{(1)}(-\lambda_3 C'(1)) + \prod_{r=1}^{j} g_{i,r}^{(1)}(-\lambda_4 C'(1))$$

Proof: The above result is proved in the same manner as in proposition 2.1.

Now, we formulate the set of equations along with the boundary conditions governing the model by using the supplementary variables technique as follows:

Governing Equations

$$\lambda_1 D_0 = \int_0^\infty V_{J,0}(x)\psi_J(x)dx \tag{4.1}$$

$$\left[\frac{d}{dw} + \lambda_1 + \gamma(w)\right] D_n(w) = 0;\ n \ge 1 \tag{4.2}$$

$$\left[\frac{d}{dx} + \lambda_2 + \alpha_i + \mu_i(x)\right] P_{i,n}(x) = \lambda_2 \sum_{m=1}^{n} c_m P_{i,n-m}(x) + \int_0^\infty R_{i,d,n}(x,y)\beta_{i,d}(y)dy; \quad (1 \le i \le k) \tag{4.3}$$

$$\left[\frac{\partial}{\partial y} + \xi_i(y) + \lambda_3\right] S_{i,n}(x,y) = \lambda_3 \sum_{m=1}^{n} c_m S_{i,n-m}(x,y);\quad (1 \le i \le k), n \ge 0 \tag{4.4}$$

$$\left[\frac{\partial}{\partial y} + \lambda_4 + \beta_{i,j}(y)\right] R_{i,j,n}(x,y) = \lambda_4 \sum_{m=1}^{n} c_m R_{i,j,n-m}(x,y);\ (1 \le i \le k), (1 \le j \le d),\ n \ge 0 \tag{4.5}$$

$$\left[\frac{d}{dx} + \lambda_5 + \psi_l(x)\right] V_{l,0}(x) = 0;\quad (1 \le l \le J) \tag{4.6}$$

$$\left[\frac{d}{dx} + \lambda_5 + \psi_l(x)\right] V_{l,n}(x) = \lambda_5 \sum_{m=1}^{n} c_m V_{l,n-m}(x);\ (1 \le l \le J),\ n \ge 1 \tag{4.7}$$

Boundary Conditions

$$D_n(0) = \sum_{l=1}^{J} \int_0^\infty V_{l,n}(x)\psi_l(x)dx + \int_0^\infty P_{k,n}(x)\mu_k(x)dx \tag{4.8}$$

$$P_{i,n}(0) = \int_0^\infty P_{i-1,n}(x)\mu_{i-1}(x)dx;\ (1 \le i \le k), n \ge 1 \tag{4.9}$$

$$S_{i,n}(x,0) = \alpha_i P_{i,n}(x),\ (1 \le i \le k);\ n \ge 1 \tag{4.10}$$

$$R_{i,1,n}(x,0) = \int_0^\infty S_{i,n}(x,y)\xi_i(y)dy;\quad (1 \le i \le k), n \ge 1 \tag{4.11}$$

$$R_{i,j,n}(x,0) = \int_0^\infty R_{i,j,n-1}(x,y)\beta_{i,j-1}\,dy;\ (1 \le j \le d), (1 \le i \le k)\ n \ge 1 \tag{4.12}$$

$$P_{0,0}(0)=\int_0^\infty D_1(w)\gamma(w)dw+(1-r)\lambda_1\int_0^\infty D_1(w)dw+\lambda_1 D_0 \tag{4.13}$$

$$P_{1,n}(0)=\int_0^\infty D_{n+1}(w)\gamma(w)dw+(1-r)\lambda_1\int_0^\infty D_{n+1}(w)dw+r\lambda_1\int_0^\infty D_n(w)dw; \quad n\geq 1 \tag{4.14}$$

$$V_{1,0}(0)=\begin{cases}\int_0^\infty P_{1,0}(x)\mu_1(x)dx, & n=0\\ 0, & n\geq 1\end{cases} \tag{4.15}$$

$$V_{l,0}(0)=\begin{cases}\int_0^\infty V_{l-1,0}(x)\psi_{l-1}(x)d, & n=0 \quad l=2,3,..,J\\ 0, & n\geq 1 \quad l=2,3,..,J\end{cases} \tag{4.16}$$

Also, the normalizing condition is given as follows

$$D_0+\sum_{n=1}^{\infty}\int_0^\infty D_n(w)dw+\sum_{n=0}^{\infty}\sum_{i=1}^{k}\int_0^k P_{i,n}(x)dx+\sum_{n=0}^{\infty}\sum_{i=1}^{k}\sum_{j=1}^{d}\int_0^\infty\int_0^\infty R_{i,j,n}(x,y)dxdy$$

$$+\sum_{n=0}^{\infty}\sum_{i=1}^{k}\int_0^\infty\int_0^\infty S_{i,n}(x,y)dxdy+\sum_{n=0}^{\infty}\sum_{l=1}^{J}\int_0^\infty V_{l,n}(x)dx=1$$

(4.17)

4.3.2 Probability Generating Function

As used in chapters 2 and 3, here also we use probability generating functions (pgf) corresponding to different states of the server to solve the set of differential difference equations so as to obtain the steady state solution of the retrial queueing model. The generating functions and hazard rates corresponding to service process and setup process are same as considered in chapter 2; the generating function and hazard rate for repair process are same as taken in chapter 3. The generating function for the vacation state is expressed as $V_l(x,z)=\sum_{n=0}^{\infty}V_{l,n}(x)z^n;\ |z|\leq 1,\ (1\leq l\leq J)$. Also, the hazard function corresponding to vacation state is given by $\psi_l(x)=\frac{w_l(x)}{1-W_l(x)},\ (1\leq l\leq J)$.

Now, we establish some theorems to present queue size distributions as follows:

Theorem 4.1: The partial generating functions for the server being in idle state, busy with i^{th} $(1\leq i\leq k)$ phase of service, under j^{th} $(1\leq j\leq d)$ repair state while broken down during i^{th} $(1\leq i\leq k)$ phase of service, during set up state, in l^{th} vacation $(1\leq l\leq J)$ at random epoch respectively, are

$$D(w,z) = D(0,z)\exp\{-\lambda_1 w\}\bar{A}(w) \tag{4.18}$$

$$P_i(x,z) = P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)\exp\{-H_i(z)x\}\bar{B}_i(x) \tag{4.19}$$

$$\begin{aligned} R_{i,j}(x,y,z) &= \alpha_i P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)\exp\{-H_i(z)x\}\bar{B}_i(x)\tilde{\eta}_i(-\lambda_3(\bar{C}(z))) \\ &\times\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}(-\lambda_4(\bar{C}(z)))\right)\exp\{-\lambda_4(\bar{C}(z))y\}\bar{G}_{i,j}(y), (1\le i\le k), (1\le j\le d) \end{aligned} \tag{4.20}$$

$$S_i(x,y,z) = \alpha_i P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)\exp\{-H_i(z)x\}\bar{B}_i(x)\exp\{-\lambda_3(\bar{C}(z))y\}\bar{N}_i(y), \quad (1\le i\le k) \tag{4.21}$$

$$V_l(x,z) = V_l(0,z)\exp\{-\lambda_5(\bar{C}(z))x\}\bar{U}_l(x) \tag{4.22}$$

where,

$$V_l(0,z) = \frac{\lambda_1 D_0}{\left[\tilde{w}(\lambda_5)\right]^{J-l+1}}, \quad \text{for } l=1, 2,.., J \tag{4.23}$$

$$\text{D(0,z)} = \frac{\lambda_1 D_0\left[z\left(\left(\frac{[\tilde{w}_l(\lambda_5(\bar{C}(z))-1)][1-(\tilde{w}(\lambda_5))^J]}{\lambda_5[1-\tilde{w}(\lambda_5)](\tilde{w}(\lambda_5))^J}\right)+\left[\tilde{b}_k(H_k(z))\prod_{q=1}^{k-1}\tilde{b}_q(H_q(z))\right]-1\right)\right]}{z-\left[\tilde{b}_k(H_k(z))\prod_{q=1}^{k-1}\tilde{b}_q(H_q(z))\right]\left[1+r(z-1)(1-\tilde{a}(\lambda_1))\right]} \tag{4.24}$$

$$P_1(0,z) = \frac{\lambda_1 D_0\left[\left[1+r(z-1)(1-\tilde{a}(\lambda))\right]\left(\left(\frac{[\tilde{w}_l(\lambda_5(\bar{C}(z))-1)][1-(\tilde{w}(\lambda_5))^J]}{\lambda_5[1-\tilde{w}(\lambda_5)](\tilde{w}(\lambda_5))^J}\right)-1\right)+z\right]}{z-\left[\tilde{b}_k(H_k(z))\prod_{q=1}^{k-1}\tilde{b}_q(H_q(z))\right]\left[1+r(z-1)(1-\tilde{a}(\lambda_1))\right]} \tag{4.25}$$

$$M_i(z) = \tilde{\eta}_i(\lambda_3(\bar{C}(z)))\prod_{r=1}^{d}\tilde{g}_{i,r}(\lambda_4(\bar{C}(z))) \tag{4.26}$$

$$H_i(z) = \left[\lambda_2(\bar{C}(z))+\alpha_i(1-M_i(z))\right] \tag{4.27}$$

Note: For the brevity, we use the product $\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}(-\lambda_4(\bar{C}(z)))\right)$ and $\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)$ in our results. However, the value of $\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}(-\lambda_4(\bar{C}(z)))\right)$=1 when j=1 and $\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)$=1 when i=1.

Proof: The proof of the theorem is obtained in the same manner as done in theorem 2.1.

Theorem 4.2: The marginal probability generating functions at random epochs, when the server is in idle state, busy with i^{th} $(1 \le i \le k)$ phase service, under jth $(1 \le j \le d)$ phase repair while breakdown during i^{th} $(1 \le i \le k)$ phase service, under set up before repair and under l^{th} $(1 \le l \le J)$ vacation are respectively, given by

$$D(z) = \frac{D(0,z)(1-\tilde{a}(\lambda_1))}{\lambda_1} \tag{4.28}$$

$$P_i(z) = \frac{P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z)\right)(1-\tilde{b}_i(H_i(z))}{H_i(z)}; \quad (1 \le i \le k) \tag{4.29}$$

$$R_{i,j}(z) = \frac{\alpha_i P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}(\lambda_4(\bar{C}(z))\right)\tilde{\eta}_i(\lambda_3(\bar{C}(z))(1-\tilde{b}_i(H_i(z))(1-\tilde{g}_{i,j}(\lambda_4(\bar{C}(z)))}{H_i(z)(\lambda_4(\bar{C}(z)))},$$

$$(1 \le i \le k), (1 \le j \le d) \tag{4.30}$$

$$S_i(z) = \frac{\alpha_i P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)(1-\tilde{b}_i(H_i(z)))(1-\tilde{\eta}_i(\lambda_3(\bar{C}(z)))}{H_i(z)(\lambda_3(\bar{C}(z))}; \quad (1 \le i \le k) \tag{4.31}$$

$$V_l(z) = \frac{\lambda_1 D_0[\tilde{w}_l(\lambda_5(1-C(z)))]}{\left[\tilde{w}_l(\lambda_5)\right]^{J-l+1}}; \qquad (1 \le l \le J) \tag{4.32}$$

Proof: The proof is done on the lines of theorem 2.2.

Theorem 4.3: The generating function for the number of customers in the retrial queue is

$$\begin{aligned}K(z) = D_0 &+ \frac{D(0,z)(1-\tilde{a}(\lambda_1))}{\lambda_1} + \sum_{i=1}^{k} P_1(0,z)\frac{\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z)\right)(1-\tilde{b}_i(H_i(z))}{H_i(z)} \\ &+ \sum_{i=1}^{k}\sum_{j=1}^{d}\frac{\alpha_i P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}(\lambda_4(\bar{C}(z))\right)\tilde{\eta}_i(\lambda_3(\bar{C}(z))(1-\tilde{b}_i(H_i(z))(1-\tilde{g}_{i,j}(\lambda_4(\bar{C}(z)))}{H_i(z)(\lambda_4(\bar{C}(z)))} \\ &+ \sum_{i=1}^{k}\frac{\alpha_i P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)(1-\tilde{b}_i(H_i(z)))(1-\tilde{\eta}_i(\lambda_3(\bar{C}(z)))}{H_i(z)(\lambda_3(\bar{C}(z))} + V_0[\tilde{w}_l(\lambda_5(\bar{C}(z)))]\end{aligned} \tag{4.33}$$

Proof: We can obtain the probability generating function for the number of customers in the retrial queue by using

$$K(z) = D_0 + D(z) + \sum_{i=1}^{k} P_i(z) + \sum_{i=1}^{k}\sum_{j=1}^{d} R_{i,j}(z) + \sum_{i=1}^{k} S_i(z) + \sum_{l=1}^{J} V_l(z) \tag{4.34}$$

Theorem 4.4: The generating function for the number of customers present in the system is

$$L(z) = D_0 + \frac{D(0,z)(1-\tilde{a}(\lambda_1))}{\lambda_1} + z\sum_{i=1}^{k} P_1(0,z)\frac{\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z)\right)(1-\tilde{b}_i(H_i(z))}{H_i(z)}$$
$$+ z\sum_{i=1}^{k}\sum_{j=1}^{d}\frac{\alpha_i P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)\left(\prod_{r=1}^{j-1}\tilde{g}_{i,r}(\lambda_4(1-C(z))\right)\tilde{\eta}_i(\lambda_3(1-C(z))(1-\tilde{b}_i(H_i(z))(1-\tilde{g}_{i,j}(\lambda_4(1-C(z)))}{H_i(z)(\lambda_4(1-C(z)))}$$
$$+ z\sum_{i=1}^{k}\frac{\alpha_i P_1(0,z)\left(\prod_{q=1}^{i-1}\tilde{b}_q(H_q(z))\right)(1-\tilde{b}_i(H_i(z)))(1-\tilde{\eta}_i(\lambda_3(1-C(z)))}{H_i(z)(\lambda_3(1-C(z))} + V_0[\tilde{w}_l(\lambda_5(1-C(z)))] \tag{4.35}$$

Proof: The generating function of the number of customers in the system is obtained by using the results of marginal generating functions and is obtained by

$$L(z) = D_0 + D(z) + z\sum_{i=1}^{k} P_i(z) + z\sum_{i=1}^{k}\sum_{j=1}^{d} R_{i,j}(z) + z\sum_{i=1}^{k} S_i(z) + \sum_{l=1}^{J} V_l(z) \tag{4.36}$$

4.4 PERFORMANCE INDICES

Some of the important performance indices are derived by using generating functions in various categories as follows:

(A) Long Run Probabilities

The long run probabilities of the server being in idle (P_I), busy (P_B), repair (P_R), set up (P_S) and vacation (P_V) states respectively, are given below in the form of following theorem.

Theorem 4.5:

(i) Long run probability of the server being in the idle state, is

$$P_I = (1-\tilde{a}(\lambda_1))D_0\left(\frac{(\chi_3' + \chi_1')}{(1-\chi_1' - \chi_2')}\right) \tag{4.37}$$

(ii) Long run probability of the server being in the busy state, is

$$P_B = \sum_{i=1}^{k}\lambda_1 D_0\left(\frac{\varsigma_9''}{\varsigma_8''}\right) \tag{4.38}$$

(iii) Long run probability of the server being in the repair state, is

$$P_R = \sum_{i=1}^{k} \alpha_i \lambda_1 D_0 \left(\frac{\varsigma_{12}'''}{\varsigma_{11}'''} \right) \tag{4.39}$$

(iv) Long run probability of the server being under set up state, is

$$P_S = \sum_{i=1}^{k} \alpha_i \lambda_1 D_0 \frac{(\varsigma_{13}''')}{(\varsigma_{14}''')} \tag{4.40}$$

(v) Long run probability of the server being in the vacation state, is

$$P_V = \frac{\lambda_1 D_0}{\lambda_5 \left[\tilde{w}(\lambda_5)\right]^{J-l+1}} \left(\frac{\varsigma_7'}{\chi_8'} \right) \tag{4.41}$$

where,

$$\varsigma_1 = \prod_{q=1}^{i-1} \tilde{b}_q(H_q(z)), \varsigma_1' = \prod_{q=1}^{i-1} b_q^{(1)}(H_q'(1)), \varsigma_1'' = \prod_{q=1}^{i-1} b_q^{(2)}(H_q'(1)) + \prod_{q=1}^{i-1} b_q^{(1)}(H_q''(1))$$

$$\varsigma_2 = 1 - \tilde{b}_i(H_i(z)), \varsigma_2' = -b_i^{(1)} H_i'(1),\ \varsigma_2'' = -b_i^{(2)} H_i''(1)^2 - b_i^{(1)} H_i''(1)$$

$$\varsigma_3 = 1 - \tilde{g}_{i,j}(\lambda_4(1-C(z))),\ \varsigma_3' = -g_{i,j}^{(1)}(-\lambda_4 E[X]), \varsigma_3'' = -g_{i,j}^{(2)}(\lambda_4 E[X])^2 - g_{i,j}^{(1)}(-\lambda_4 E[X^2])$$

$$\varsigma_4 = \tilde{\eta}_i(\lambda_3(1-C(z)), \varsigma_4' = -\eta_i^{(1)} \lambda_3 E[X], \varsigma_4'' = \eta_i^{(2)} \left(-\lambda_3 E[X]\right)^2 - \eta_i^{(1)} \lambda_3 E[X^2]$$

$$\varsigma_5 = \prod_{r=1}^{j-1} \tilde{g}_{i,r}\left(\lambda_4(1-C(z))\right), \varsigma_5' = \prod_{r=1}^{j-1} g_{i,r}^{(1)}\left(-\lambda_4 E[X]\right),\ \varsigma_5'' = \prod_{r=1}^{j-1} g_{i,r}^{(1)}\left(-\lambda_4 E[X^2]\right) + \prod_{r=1}^{j-1} g_{i,r}^{(2)}\left(-\lambda_4 E[X]\right)^2$$

$$\varsigma_6 = 1 - \tilde{\eta}_i(\lambda_3(1-C(z))),\ \varsigma_6' = -\eta_i^{(1)}(-\lambda_3 E[X]),\ \varsigma_6'' = -\eta_i^{(1)}(-\lambda_3 E[X^2]) - \eta_i^{(2)}(\lambda_3 E[X])^2$$

$$\varsigma_7 = 1 - \tilde{u}_l(\lambda_5(1-C(z))), \varsigma_7' = -u_l^{(1)}\left(-\lambda_5 E[X]\right), \varsigma_7'' = -u_l^{(1)}\left(-\lambda_5 E[X^2]\right) - u_l^{(2)}\left(-\lambda_5 E[X]\right)^2$$

$$\varsigma_8 = (z - \chi_1\chi_2)\chi_5, \varsigma_8'' = 2(1 - \chi_1' - \chi_2')\chi_5',\ \varsigma_8''' = 3(z - \chi_1\chi_2)''\chi' + 3(z - \chi_1\chi_2)'\chi_5''$$

$$\varsigma_9 = \chi_2\varsigma_1\varsigma_2,\ \varsigma_9'' = 2\chi_2'\varsigma_2',\ \varsigma_9''' = 3\chi_2''\varsigma_2' + 6\chi_2'\varsigma_1'\varsigma_2' + 3\chi_2'\varsigma_2'',$$

$$\varsigma_{11} = \varsigma_8\chi_6,\ \varsigma_{11}''' = 3\varsigma_8''\chi_6',\ \varsigma_{11}^{iv} = 4\varsigma_8'''\chi_6' + 6\varsigma_8''\chi_6'',$$

$$\varsigma_{10} = \varsigma_4\varsigma_5\varsigma_3,\ \varsigma_{10}' = \varsigma_3',\ \varsigma_{10}'' = 2\varsigma_4'\varsigma_5'\varsigma_3' + 2\varsigma_4\varsigma_5'\varsigma_3' + \varsigma_4\varsigma_5\varsigma_3'',$$

$$\varsigma_{12} = \varsigma_9\varsigma_{10},\ \varsigma_{12}''' = 3\varsigma_9''\varsigma_{10}',\ \varsigma_{12}^{iv} = 4\varsigma_9'''\varsigma_{10}' + 6\varsigma_9''\varsigma_{10}''\ \varsigma_{14} = \varsigma_8\chi_7,\ \varsigma_{14}''' = 3\varsigma_8''\chi_7',\ \varsigma_{14}^{iv} = 4\varsigma_8'''\chi_7' + 6\varsigma_8''\chi_7'',$$

$$\varsigma_{13} = \varsigma_9\varsigma_6,\ \varsigma_{13}''' = 3\varsigma_9''\varsigma_6',\ \varsigma_{13}^{iv} = 4\varsigma_9'''\varsigma_6' + 6\varsigma_9''\varsigma_6''$$

$$\chi_1 = 1 + r(z-1)(1 - \tilde{a}(\lambda_1));\ \chi_1' = r(1 - \tilde{a}(\lambda_1))$$

$$\chi_5 = H_i(z) = \lambda_2(1 - C(z)) + \alpha_i(1 - M_i(z)),\ \chi_5' = H_i'(z) = -\lambda_2 E[X] - \alpha_i M_i'(1),$$
$$\chi_5'' = H_i''(z) = -\lambda_2 E[X^2] - \alpha_i M_i''(1)$$

$$\chi_4 = \prod_{q=1}^{k-1} \tilde{b}_q(H_q(z))\tilde{b}_k(H_k(z)), \chi_4' = \prod_{q=1}^{k} b_q^{(1)} H_q'(1), \chi_4'' = \prod_{q=1}^{k} b_q^{(2)} H_q'(1) + \prod_{q=1}^{k} b_q^{(1)} H_q''(1)$$

$$\chi_6 = \lambda_4(1 - C(z)),\ \chi_6' = -\lambda_4 E[X],\ \chi_6'' = -\lambda_4 E[X^2]$$

$$\chi_3 = \frac{\left[1-\tilde{w}(\lambda_5)^J\right]}{\lambda_5\left[1-\tilde{w}(\lambda_5)\right]\left[\tilde{u}(\lambda_5)\right]^J}\left[\tilde{w}_s(\lambda_5(1-C(z)))\right],\ \chi_3' = \frac{\left[1-\tilde{w}(\lambda_5)^J\right]}{\lambda_5\left[1-\tilde{w}(\lambda_5)\right]\left[\tilde{w}(\lambda_5)\right]^J}\tilde{w}_l^{(1)}(-\lambda_5 E[X])$$

$$\chi_7 = \lambda_3(1-C(z)), \chi_7' = -\lambda_3 E[X],\ \chi_7'' = -\lambda_3 E[X^2]$$

$$\chi_3'' = \frac{\left[1-\tilde{w}(\lambda_5)^J\right]}{\lambda_5\left[1-\tilde{w}(\lambda_5)\right]\left[\tilde{w}(\lambda_5)\right]^l}\left(\tilde{w}_l^{(2)}(-\lambda_5 E[X])^2 + \tilde{w}_l^{(1)}(-\lambda_5 E[X^2])\right)$$

$$\chi_{10} = \lambda_5(1-C(z)), \chi_{10}' = -\lambda_5 E[X],\ \chi_{10}'' = -\lambda_5 E[X^2]$$

$$M_i(z) = \tilde{\eta}_i(\lambda_3(1-C(z)))\tilde{g}_{i,j}(\lambda_4(1-C(z)))\prod_{r=1}^{j-1} g_{i,r}(\lambda_4(1-C(z)))$$

$$M_i'(z) = \eta_i^{(1)}(-\lambda_3 E[X]) + \prod_{r=1}^{j} g_{i,r}^{(1)}(-\lambda_4 E[X]),$$

$$M_i''(z) = \eta_i^{(2)}(-\lambda_3 E[X])^2 + \eta_i^{(1)}(-\lambda_3 E[X^2]) + \eta_i^{(1)}(-\lambda_3 E[X])\prod_{r=1}^{j} g_{i,r}^{(1)}(-\lambda_4 E[X])$$
$$+\prod_{r=1}^{j} g_{i,r}^{(2)}(-\lambda_4 E[X])^2 + \prod_{r=1}^{j} g_{i,r}^{(1)}(-\lambda_4 E[X^2]),$$

Proof: For proof, we follow the procedure as used for the proof of theorem 2.5.

(B) Queueing Measures

The analytic expressions for the queue lengths of both system and retrial orbit is obtained in the following theorem:

Theorem 4.6: The mean queue length of the retrial orbit (L_R) and that of the system (L_S) are

$$L_R = \frac{D_0(1-\tilde{a}(\lambda_1))(1-\chi_1'-\chi_2')\left(\left(\chi_3'+\chi_1'\right)+\left(\chi_3''+\chi_1''\right)\right)+\left(\chi_4''+\chi_2''\right)\left(\chi_3'+\chi_1'\right)}{2(1-\chi_1'-\chi_2')^2}$$
$$+\sum_{i=1}^{k}\frac{\lambda_1 D_0\left(a''b'''-a'''b''\right)}{12(1-\chi_1'-\chi_2')^2\chi_5'^2}+\sum_{i=1}^{k}\frac{\alpha_i\lambda_1 D_0(5e^{iv}d'''-5d^{iv}e''')}{624(1-\chi_1'-\chi_2')^2\chi_5'^2\chi_6'^2} \quad (4.42)$$
$$+\sum_{i=1}^{k}\frac{\alpha_i\lambda_1 D_0(5f^{iv}h'''-5h^{iv}f''')}{624(1-\chi_1'-\chi_2')^2\chi_5'^2\chi_7'^2}+\sum_{l=1}^{J}\frac{\lambda_1 D_0}{\lambda_5\left[\tilde{w}(\lambda_5)\right]^{J-l+1}}\left(\frac{\chi_8'\chi_9''-\chi_8''\chi_9'}{2\chi_8'^2}\right)$$

$$L_S = L_R + \frac{\lambda_1 D_0\varsigma_9''}{\varsigma_8''}+\sum_{i=1}^{k}\frac{\alpha_i\lambda_1 D_0\varsigma_{12}'''}{\varsigma_{11}'''}+\sum_{i=1}^{k}\frac{\alpha_i\lambda_1 D_0\varsigma_{13}'''}{\varsigma_{14}'''} \quad (4.43)$$

Proof: The mean queue length of the retrial orbit and mean queue length of the system are obtained in eqs (4.42) and (4.43) respectively, by using

$$L_R = \lim_{z\to 1} K'(z) \text{ and } L_s = \lim_{z\to 1} L'(z)$$

Here, L - Hospital rule has been used six times to evaluate the limiting value when z → 1.

Theorem 4.7: The exact expected waiting time for a customer in the system is given by

$$W_s = \frac{L_s}{\lambda_{eff}} \tag{4.44}$$

where, $\lambda_{eff} = \left[\lambda_1 P_I + \lambda_2 P_B + \lambda_3 P_S + \lambda_4 P_R + \lambda_5 P_V\right] E[X]$.

Proof: The exact expected waiting time W_s is obtained by using Little's formula (cf. Gross and Harris, 1985) as given by eq. (4.44).

(C) Reliability Measures

In this section, we derive some important reliability measures namely availability and failure frequency.

Theorem 4.8: The steady state availability (A_v) and failure frequency (F_f) of the server are

$$A_v = D_0\left[\tilde{a}(\lambda_1) + \sum_{i=1}^{k}\lambda_1 \frac{\left[1+\chi_3' - r(1-\tilde{a}(\lambda_1))\right]}{(1-\chi_2'-\chi_1')\mu_i}\right] \tag{4.45}$$

$$F_f = D_0 \sum_{i=1}^{k}\lambda_1\alpha_i \frac{\left[1+\chi_3' - r(1-\tilde{a}(\lambda_1))\right]}{(1-\chi_2'-\chi_1')\mu_i} \tag{4.46}$$

Proof: The availability and failure frequency for the system given in eqs (4.45) and (4.46) are obtained using

$$A_v = D_0 + \int_0^\infty D(w,1)dw + \sum_{i=1}^{k}\int_0^\infty P_i(x,1)dx \text{ and } F_f = \sum_{i=1}^{k}\int_0^\infty \alpha_i P_i(x,1)dx$$

4.5 MAXIMUM ENTROPY PRINCIPLE

In this section, we employ maximum entropy approach in order to determine the steady state probabilities $P_{i,n}(1 \le i \le k)$, $R_{i,j,n}(1 \le i \le k),(1 \le j \le d)$, $S_{i,n}(1 \le i \le k)$, $V_{l,n}(1 \le l \le J)$ and D_n for the concerned $M^{[x]}/G/1$ retrial queueing system with modified vacation policy. For the analysis purpose, we follow the following procedure (cf. Wang *et al.*, 2007b) as used in section 2.6 and frame maximum entropy function as:

$$Y = -\sum_{n=1}^{\infty}\sum_{i=1}^{k} \mathrm{P}_{i,n}\log \mathrm{P}_{i,n} - \sum_{n=1}^{\infty}\sum_{i=1}^{k}\sum_{j=1}^{d} R_{i,j,n}\log R_{i,j,n} - \sum_{n=1}^{\infty}\sum_{i=1}^{k} S_{i,n}\log S_{i,n}$$
$$-\sum_{n=1}^{\infty} D_n \log D_n - \sum_{n=1}^{\infty}\sum_{l=1}^{J} V_{l,n}\log V_{l,n} \tag{4.47}$$

subject to the constraints

(i) $$\sum_{n=1}^{\infty}\sum_{i=1}^{k} P_{i,n} = P_B,\ \sum_{n=1}^{\infty}\sum_{i=1}^{k}\sum_{j=1}^{d} R_{i,j,n} = P_R,\ \sum_{n=1}^{\infty}\sum_{i=1}^{k} S_{i,n} = P_S \tag{4.48}$$

(ii) $\sum_{n=1}^{\infty} D_n = P_I, \sum_{n=1}^{\infty}\sum_{l=1}^{J} V_{l,n} = P_V$ (4.49)

(iii) $\sum_{n=1}^{\infty} n\left\{\sum_{i=1}^{k} P_{i,n} + \sum_{i=1}^{k}\sum_{j=1}^{d} R_{i,j,n} + \sum_{i=1}^{k} S_{i,n} + \sum_{l=1}^{J} V_{l,n} + D_n\right\} = L_S$ (4.50)

Lagrange's Function

To determine the maximum value of entropy function, we construct Lagrange's function $H(P_{i,n}, R_{i,n}, S_{i,n}, D_n, V_{l,n})$ by introducing the Lagrange's multipliers θ_i (1 ≤ i≤ k), δ_{ij} (1 ≤ i≤ k), φ_i (1 ≤ i≤ k), θ_{k+1}, ε_l (1 ≤ l≤ J) and ϕ_{k+1} corresponding to the information as constraints (4.48)-(4.50) available in the form of derived analytical results. Thus, we have

$$H(P_{i,n}, R_{i,n}, S_{i,n}, D_n, V_{l,n}) =$$
$$\sum_{n=1}^{\infty}\sum_{i=1}^{k} \mathrm{P}_{i,n}\log \mathrm{P}_{i,n} - \sum_{n=1}^{\infty}\sum_{i=1}^{k}\sum_{j=1}^{d} R_{i,j,n}\log R_{i,j,n} - \sum_{n=1}^{\infty}\sum_{i=1}^{k} S_{i,n}\log S_{i,n} - \sum_{n=1}^{\infty} D_n \log D_n - \sum_{n=1}^{\infty}\sum_{l=1}^{J} V_{l,n}\log V_{l,n}$$
$$-\sum_{i=1}^{k}\theta_i\left[\sum_{n=1}^{\infty}\sum_{i=1}^{k} P_{i,n} - P_B\right] - \sum_{i=1}^{k}\sum_{j=1}^{d}\delta_{ij}\left[\sum_{n=1}^{\infty}\sum_{i=1}^{k}\sum_{j=1}^{d} R_{i,j,n} - P_R\right] - \sum_{i=1}^{k}\phi_i\left[\sum_{n=1}^{\infty}\sum_{i=1}^{k} S_{i,n} - P_S\right] - \theta_{k+1}\left[\sum_{n=1}^{\infty} D_n - P_I\right]$$
$$-\sum_{l=1}^{J}\varepsilon_l\left[\sum_{n=1}^{\infty}\sum_{l=1}^{J} V_{l,n} - P_V\right] - \phi_{k+1}\left[\sum_{n=1}^{\infty} n\left\{\sum_{i=1}^{k} P_{i,n} + \sum_{i=1}^{k}\sum_{j=1}^{d} R_{i,j,n} + \sum_{i=1}^{k} S_{i,n} + \sum_{i=1}^{k} V_{l,n} + D_n\right\} - L_S\right]$$

(4.51)

The results obtained for approximate probabilities of different states of the server are presented in the form of following theorem.

Theorem 4.9: The maximum entropy solutions for the approximate values of probabilities $\mathrm{P}_{i,n}, \mathrm{R}_{i,j,n}, \mathrm{D}_n, S_{i,n}$ and $\mathrm{V}_{l,n}$, $(1 \le i \le k), (1 \le j \le d), (1 \le l \le J), n \ge 1$ subject to the constraints are

$$\hat{P}_{i,n} = \frac{P_B\sigma\left[L_S - \sigma\right]^{n-1}}{L_S{}^n} \quad \hat{R}_{i,j,n} = \frac{P_R\sigma\left[L_S - \sigma\right]^{n-1}}{L_S{}^n} \quad \hat{S}_{i,n} = \frac{P_S\sigma\left[L_S - \sigma\right]^{n-1}}{L_S{}^n}$$

$$\hat{D}_n = \frac{P_I\sigma\left[L_S - \sigma\right]^{n-1}}{L_S{}^n} \quad \hat{V}_{l,n} = \frac{P_V\sigma\left[L_S - \sigma\right]^{n-1}}{L_S{}^n} \tag{4.52}$$

where,

$$\sigma = P_B + P_R + P_V + P_s + P_I \tag{4.53}$$

Proof: The proof of this theorem is done on the lines of theorem 2.9.

Theorem 4.10: Using the principle of maximum entropy, the approximate expected waiting time in the system is

$$
\begin{aligned}
\hat{W}_s = & \sum_{n=1}^{\infty}\left[\sum_{i=1}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right]\hat{D}_n+\sum_{n=1}^{\infty}\left[\sum_{i=1}^{k}\frac{n}{\mu_i}+\sum_{i=1}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right]\hat{P}_{i,n} \\
& +\sum_{n=1}^{\infty}\left[\sum_{i=1}^{k}\sum_{j=1}^{d}\frac{g_{ij}^{(2)}}{2g_{ij}^{(1)}}+\sum_{i=1}^{k}\frac{n}{\mu_i}+\sum_{i=1}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right]\hat{R}_{i,j,n} \\
& +\sum_{n=1}^{\infty}\left[\sum_{i=1}^{k}\sum_{j=1}^{d}\frac{\eta_i^{(2)}}{2\eta_i^{(1)}}+\sum_{i=1}^{k}\frac{n}{\mu_i}+\frac{1}{\beta_{ij}}+\sum_{i=1}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right]\hat{S}_{i,n} \\
& +\sum_{n=1}^{\infty}\left[\sum_{l=1}^{J}\frac{w_l^{(2)}}{2w_l^{(1)}}+\sum_{i=1}^{k}\frac{n}{\mu_i}+\sum_{i=1}^{k}\frac{1}{2\mu_i}\left(\frac{E[X^2]}{E[X]}-1\right)\right]\hat{V}_{l,n}
\end{aligned}
\qquad (4.54)
$$

Proof: The approximate expected waiting time is obtained by following the same procedure as used for the proof of theorem 2.10.

4.6. NUMERICAL ILLUSTRATION

The present section deals with the sensitivity analysis of the performance indices of queueing model with respect to the various parameters.

(A) Queue Length (L_S)

To study the sensitivity of queue length towards various parameters, figures 4.1-4.6 have been plotted corresponding to three service time distributions namely Erlangian-2, exponential and gamma distributions. The vacation time, retrial process, set up process as well as repair process is assumed to follow exponential distribution. The set of default parameters assumed for the numerical illustration are as follows-

$\lambda=\lambda_1=\lambda_2=\lambda_3=\lambda_4=\lambda_5=0.5, \xi=1, \mu=\mu_1=\mu_2=5, \alpha=\alpha_1=\alpha_2=0.01, \gamma=0.1, \psi=2,$
$r=0.1, \beta=\beta_{11}=\beta_{12}=\beta_{21}=\beta_{22}=0.9.$

The effect of parameters namely breakdown rate (α) and repair rate (β) on the queue length of the system have been demonstrated in figs 4.1 and 4.2. It is clear from the figs 4.1(a-c) that the queue length increases as the breakdown rate increases from 0.008 units to 0.01 units for all the service time distributions. The maximum number of customers or queue length is observed in the fig. 4.1(c), when the service time is supposed to be gamma distributed. On the other hand in fig. 4.2 wherein graphs are plotted for different values of repair rate β (0.6, 0.8 and 1), L_S decreases with an increase in the repair rate. This is due to the fact that an increase in the breakdown rate forces the customers to accumulate in the system due to the non-working condition of the server, and hence increases the queue length.

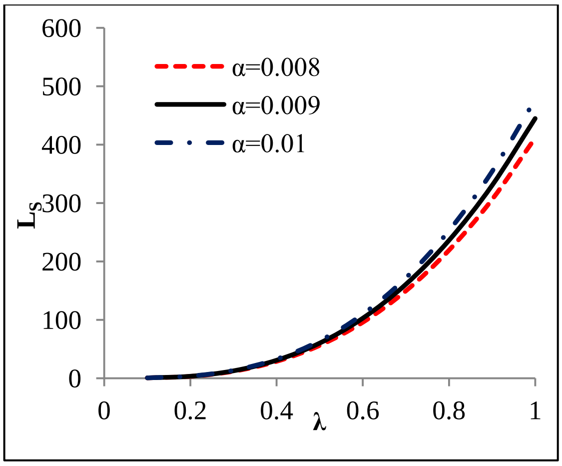
In figs 4.3 (a-c) queue lengths have been plotted for various values of service rate μ for (a) Erlangian-2 (b) exponential and (c) gamma service time distributions. The graphs plotted in figures 4.3 (a-c) clearly demonstrate that the queue length of the system decreases with an increase in the service rate for all the three service time distributions. Figs 4.4 (a-c) depict the sensitivity of 'r' on the queue length of the system. The number of customers in the system increases as r increases from 0.1 to 0.3 units. The variations of the queue length with the vacation rate ψ and setup rate ξ are explored through figures 4.5 and 4.6.

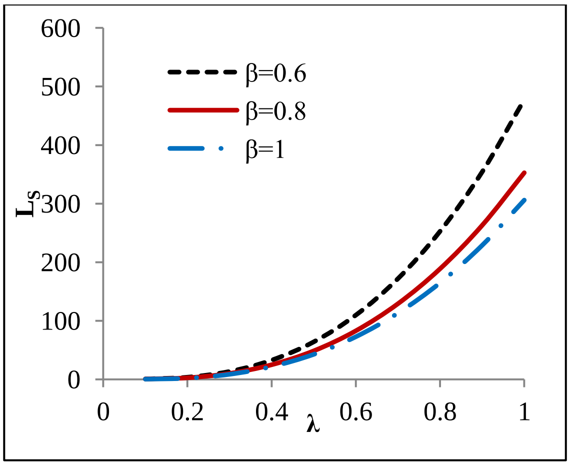
Moreover, the set up rate also affects the system size; an increase in the setup rate decreases the number of customers in the system as demonstrated by figs 4.6 (a-c). This pattern is due to the fact that an increase in the setup rate improves the repair process of the server which in turn increases the availability of the server and thus a reduction in the number of customers in the system is observed.

(B) Comparison of Exact and Approximate Average Waiting Time

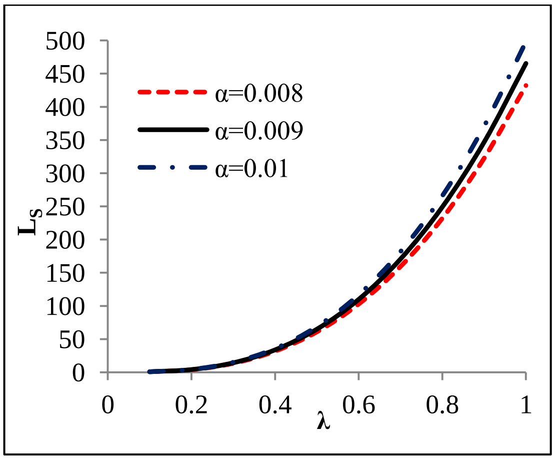
In the present subsection, a comparison between exact average waiting time (W_q) and approximate average waiting time has been presented in tables 4.1- 4.2. Table 4.1 shows the comparison between exact and approximate average waiting time for two types of service time distributions namely exponential and gamma distributions. The absolute percentage error (APE %) has been obtained for variation in different parameters namely (i) setup rate ξ, (ii) retrial rate γ and (iii) breakdown rate $\alpha=\alpha_1=\alpha_2$. An increase in the set up rate ξ from 1.0 unit to 2.5 units affects the waiting time of the customer in the queue.

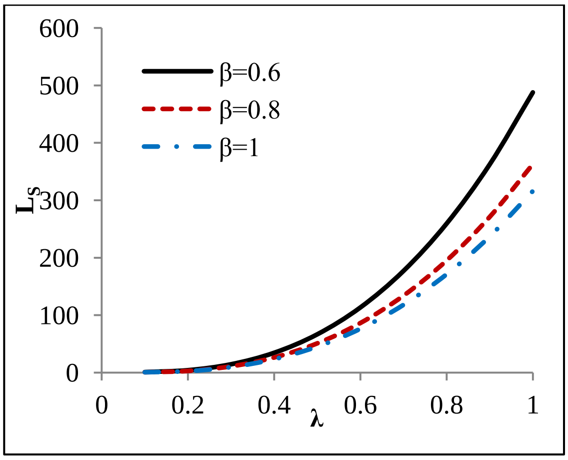
APE decreases with the increase in retrial rate for both exponential as well as gamma distributions with maximum % error as 16.20 % for exponential distribution and 15.92 % for gamma distribution in Case 2. The data captured in Case 3 depicts the effect of breakdown rate α on the waiting time of the customer in the system. Both exact and approximate waiting time increases with the increase in break down rate from 0.006 units to 0.01 units. Table 4.2 displays the data for the average waiting time for a queueing model with Erlangian-2 and deterministic distributed service time. With an increase in retrial rate γ and breakdown rate α, APE as well as average waiting time decreases.



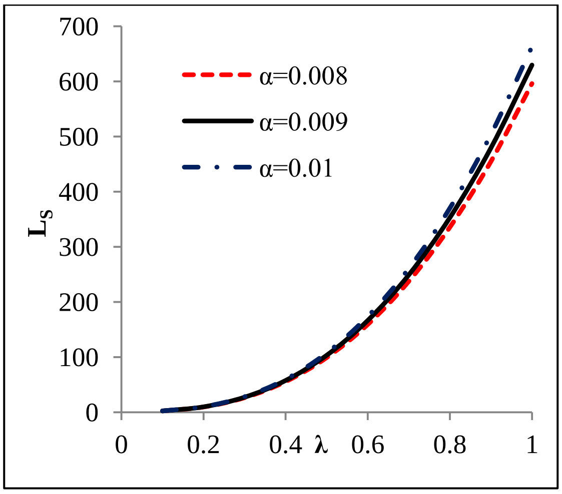


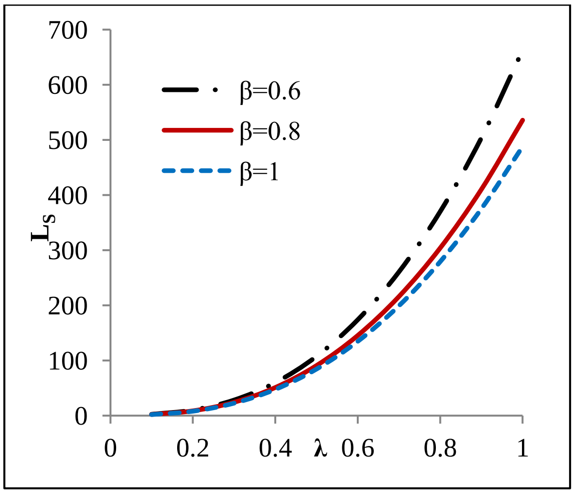
(a) **(a)**





(b) (b)

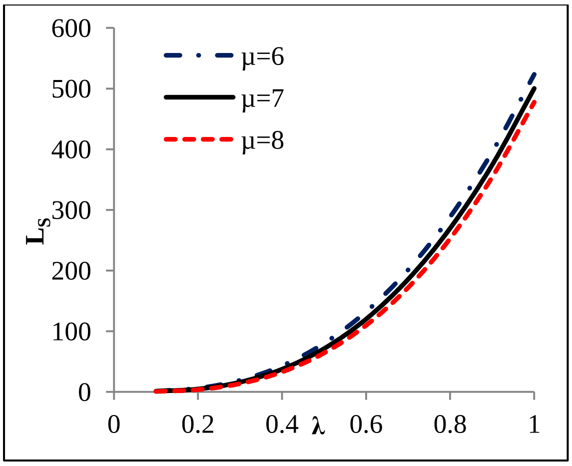




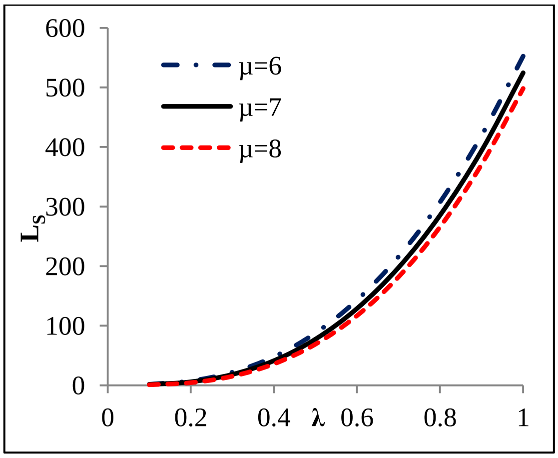
(c) (c)

Fig. 4.1: Effect of α on L_s for
(a) Erlangian-2
(b) Exponential
(c) Gamma distributed service time

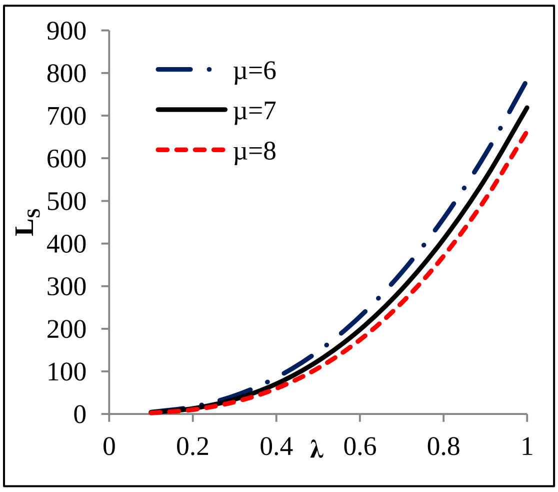
Fig. 4.2: Effect of β on L_s for
(a) Erlangian-2
(b) Exponential
(c) Gamma distributed service time



(a)

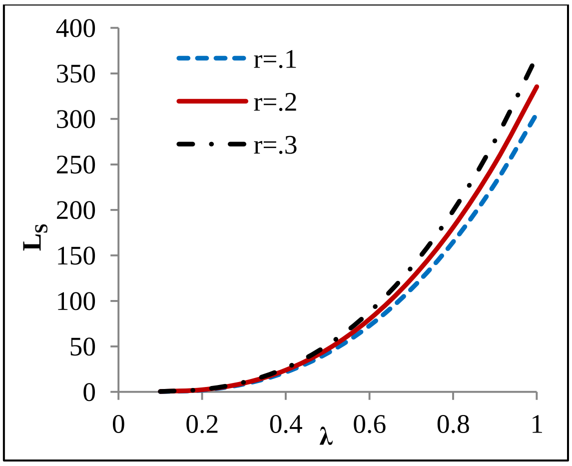


(b)

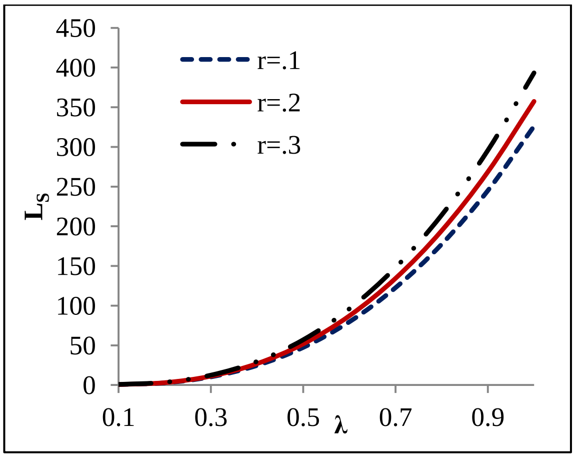


(c)

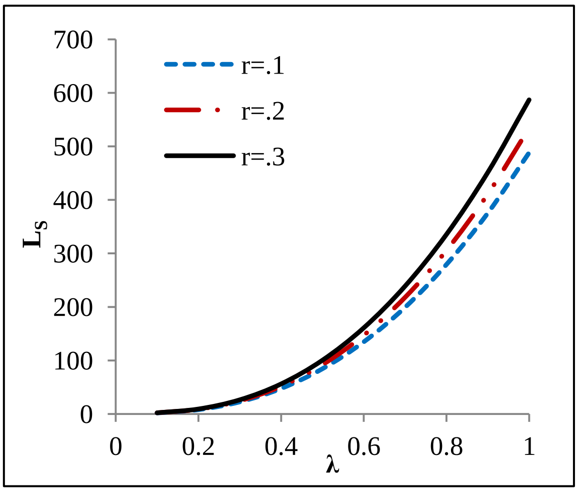
Fig. 4.3: Effect of μ on L_s for
(a) Erlangian-2
(b) Exponential
(c) Gamma distributed service time



(a)

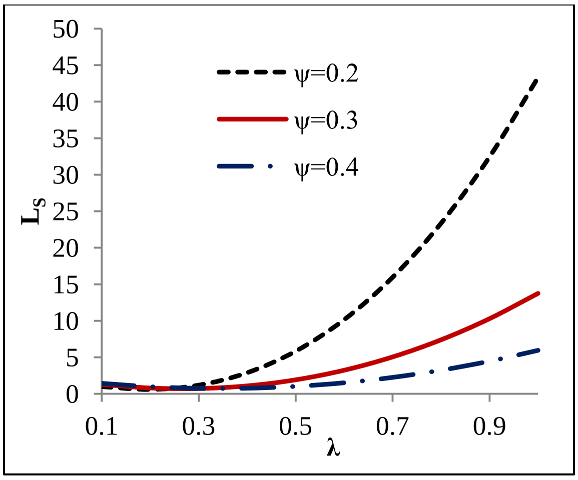


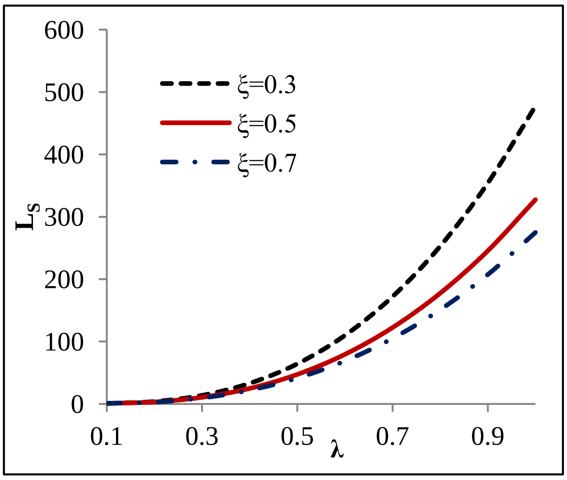
(b)



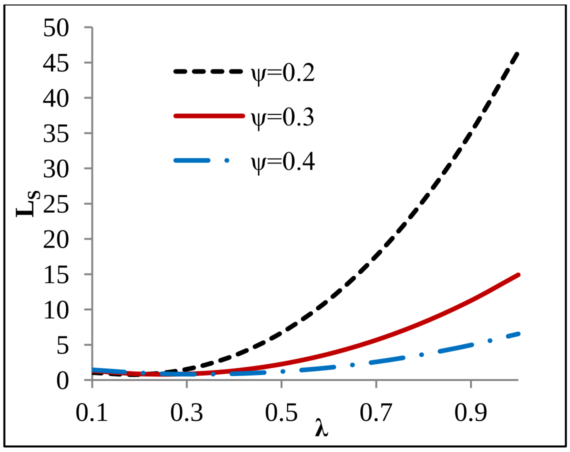
(c)

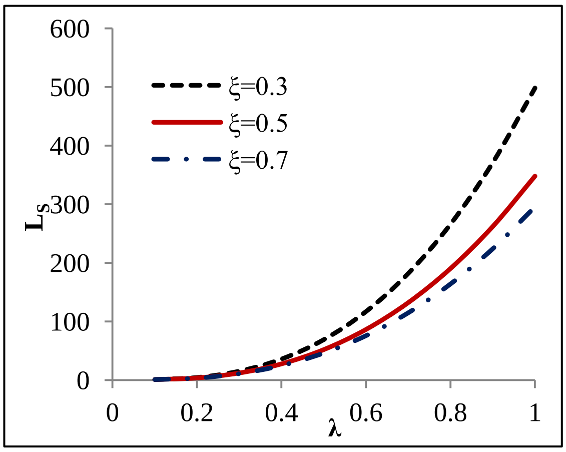
Fig. 4.4: Effect of r on L_s for
(a) Erlangian-2
(b) Exponential
(c) Gamma distributed service time



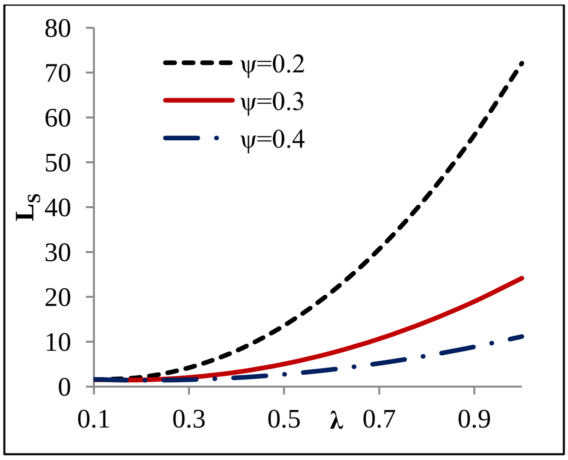


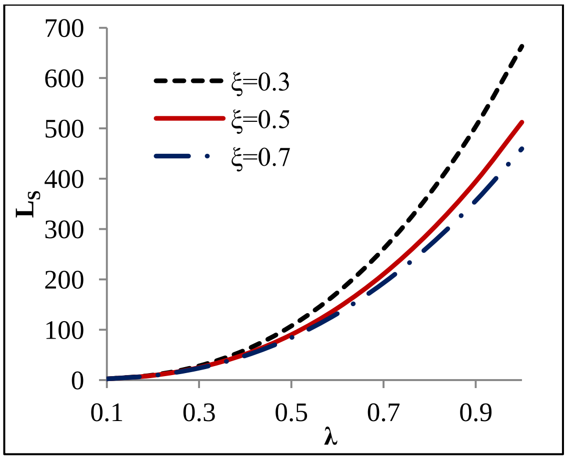
(a) **(a)**





(b) **(b)**





(c) **(c)**

Fig. 4.5: Effect of ψ on L_s for
(a) Erlangian-2
(b) Exponential
(c) Gamma distributed service time

Fig. 4.6: Effect of ξ on L_s for
(a) Erlangian-2
(b) Exponential
(c) Gamma distributed service time

Table 4.1: Comparison of exact and approximate average waiting time for Exponential and Gamma distributed service time

	Service time as Exponential			Service time as Gamma distributed		
	Case 1			Case 1		
ξ	Wq	$\hat{W}_q$	APE (%)	Wq	$\hat{W}_q$	APE (%)
1.0	4.1838	4.1969	0.3143	3.8552	3.8575	0.0592
1.8	4.1718	4.2041	0.7755	3.8460	3.8577	0.3066
2.0	4.1685	4.2098	0.9900	3.8430	3.8624	0.5062
	Case 2			Case 2		
γ	Wq	$\hat{W}_q$	APE (%)	Wq	$\hat{W}_q$	APE (%)
0.080	4.2578	4.9479	16.2072	3.9054	4.5272	15.9205
0.090	4.2195	4.5384	7.5560	3.8795	4.1622	7.2858
0.10	4.1838	4.1969	0.3143	3.8552	3.8575	0.0592
	Case 3			Case 3		
α	Wq	$\hat{W}_q$	APE (%)	Wq	$\hat{W}_q$	APE (%)
0.006	4.1146	3.6017	12.4659	3.7856	3.3029	12.7513
0.008	4.1498	3.8788	6.5305	3.8210	3.5611	6.8004
0.01	4.1838	4.1969	0.3143	3.8552	3.8575	0.0592

4.7 COST ANALYSIS

In the present section, we construct the expected total cost function (ETC) for the retrial queueing model with modified vacation policy under consideration. The cost function is formulated as:

$$ETC = C_h L_s + C_b P_B + C_s P_s + C_R P_R + C_V P_V + C_I P_I$$

where,

C_h = Holding cost per unit customer

C_b = Cost per unit time while servicing the customers

C_s = Cost per unit time for making pre repair settings

C_R = Cost per unit time for providing repair to the broken down server

C_V = Cost per unit time in the system when the server is on vacation

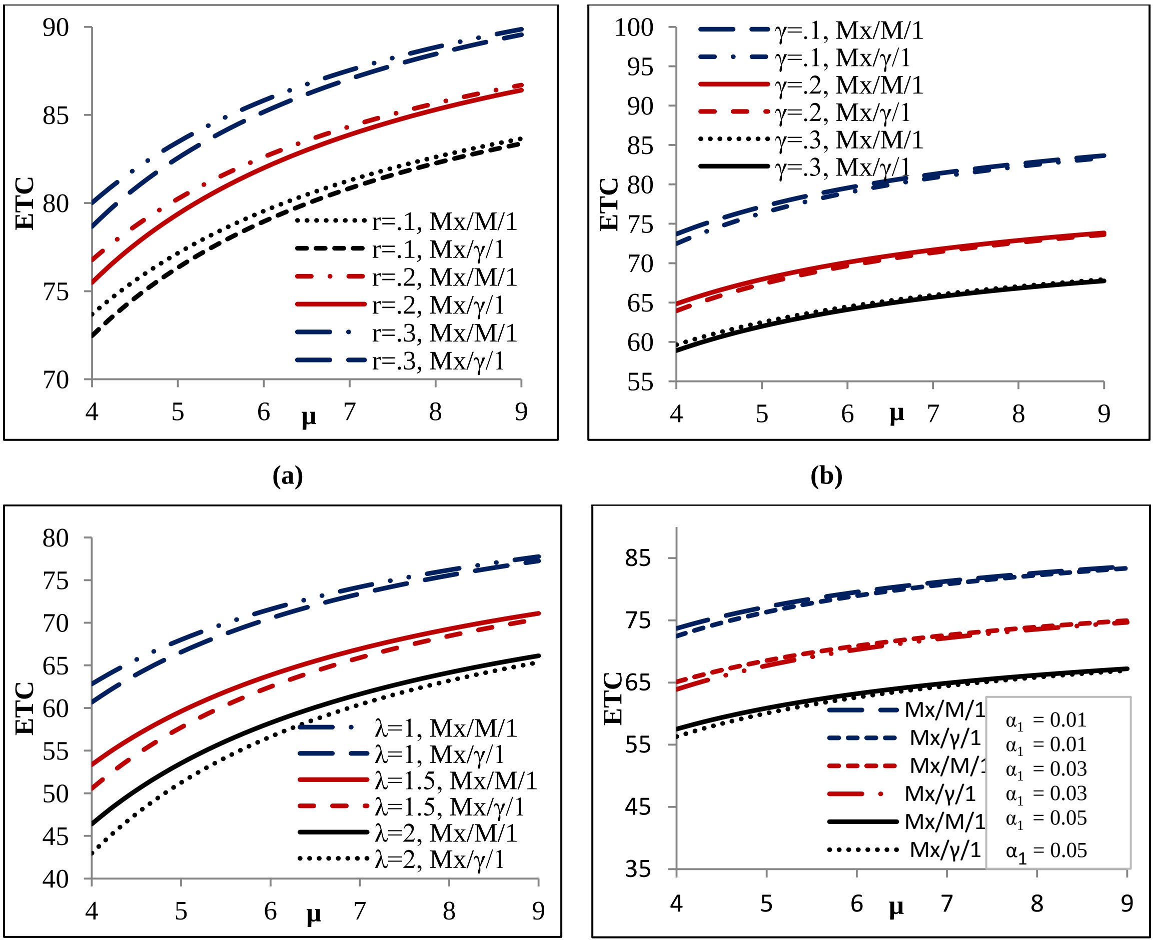
C_I = Cost per unit time when the customer retry for the service

Table 4.2: Comparison of exact and approximate average waiting time for Erlangian-2 and Deterministic distributed service time

	Service time as Erlangian-2 distributed			Service time as Deterministic distributed		
	Case 1			Case 1		
ξ	**Wq**	$\hat{W}_q$	**APE (%)**	**Wq**	$\hat{W}_q$	**APE (%)**
1.0	4.2248	4.2393	0.3434	4.2659	4.2818	0.3720
1.8	4.2125	4.2474	0.8290	4.2533	4.2907	0.8815
2.0	4.2092	4.2532	1.0452	4.2499	4.2966	1.0994
	Case 2			Case 2		
γ	**Wq**	$\hat{W}_q$	**APE (%)**	**Wq**	$\hat{W}_q$	**APE (%)**
0.080	4.3019	5.0005	16.2398	4.3459	5.0531	16.2716
0.090	4.2621	4.5854	7.5868	4.3046	4.6324	7.6169
0.10	4.2248	4.2393	0.3434	4.2659	4.2818	0.3720
	Case 3			Case 3		
α	**Wq**	$\hat{W}_q$	**APE (%)**	**Wq**	$\hat{W}_q$	**APE (%)**
0.006	4.1969	3.6764	12.4015	4.1558	3.6391	12.4333
0.008	4.2320	3.9582	6.4696	4.1909	3.9185	6.4998
0.01	4.2659	4.2818	0.3720	4.2248	4.2393	0.3434

The effect of various parameters on the total cost of the system has been examined so as to visualize the nature of the cost function towards various parameters. The set of default cost elements are taken as C_I =10, C_R =50, C_h =5, C_b =50, C_V=20, C_s =20. Figures 4.7 (a-d) display the effect of various parameters namely reneging probability (r), retrial rate (γ), arrival rate (λ) and breakdown rate (α_1) respectively, on the total cost (ETC) of the system. The graphs are plotted with ETC on the y-axis and service rate μ (=μ_1= μ_2) on the x-axis. The sensitivity of cost with respect to service rate (μ) for different values of r is displayed in fig. 4.7 (a).

The total cost of the system increases with an increase in the reneging probability for both exponential and gamma service time distributions. It is seen that the total cost of the system decreases with the increase in retrial rate (γ) from 0.1 units to 0.3 units for the fixed values of other parameters as displayed in fig. 4.7(b) for both types of distributions. The sensitivity of total cost with varying values of arrival rate (λ) and breakdown rate (α_1) has been depicted in figs 4.7(c) and 4.7 (d), respectively. Fig. 4.7(d) displays the effect of



(c) (d)

Fig. 4.7: Effect of (a) reneging probability r (b) retrial rate γ (c) arrival rate λ and (d) break down rate α_1 on ETC of the system

breakdown rate and service rate on ETC; it is quite interesting to observe that the total cost values for different service time distributions namely exponential and gamma service time are quite close enough.

4.8 DISCUSSION

The bulk arrival retrial queueing model with discouragement and modified vacation has been analyzed. The concepts of phase service and phase repair incorporated along with the delaying repair makes the present model close to many real life queueing scenarios. Overall, we may conclude that-

- The maximum queue length is observed in case of a system with gamma service time distribution as compared to the system following exponential and Erlangian-2 service time distributions.

- The queue length decreases as the vacation rate increases; this is due to the fact that a server goes for vacation only when there is no customer in the system which implies reduction in the number of customers in the system.
- APE increases with an increase in the setup rate (ξ) for both Erlangian-2 and deterministic distributed service process. Hence, the choice of appropriate service distribution may help in reducing the waiting time of the customers in the system.
- As expected, the total cost increases with an increase in r, retrial rate (γ), arrival rate (λ), breakdown rate (α_1) and service rate (μ).

CHAPTER 5

PRIORITY RETRIAL QUEUE WITH MULTIPLE PHASE SERVICE

5.1 INTRODUCTION

The formation of retrial queues can also be visualized in many systems which allow the arrivals of different class of customers. One such example of real life congestion situations is the working of hospitals where variety of patients are admitted and are treated depending on the injury/problem with which they are suffering. The emergency patients with major injuries are obviously treated earlier than the patients with minor problems and thus later are forced to wait for their turn in the orbit. Such retrial queues are said to be retrial queues with priority customers.

In the queues associated with the real life congestion situations, the customers are basically served on the basis of first come first serve (FCFS) criterion. But there may be queueing situations in which jobs are assigned or classified according to some priority index. Some notable results on single server priority retrial queues can be found in the article by Choi and Chang (1993). Gomez-Coral (2002) analysed single server retrial queue with non-preemptive priority. Wang (2008) analyzed the single server retrial queue with priority customers where service times for both the customers are assumed to be general distributed. Dimitriou (2013) studied priority retrial queue with negative arrivals, unreliable server and multiple vacations.

The objective of our study in this chapter is to investigate the batch arrival retrial queues with two types of customers viz. non-preemptive priority and ordinary customers. The service of the customers is completed in multi essential phases. Further, we assume that the repair of the failed server is also done in multi essential phases. The present chapter is organized in the following manner. Section 5.2 deals with the description of the model. Section 5.3 presents the queue size distribution of the model and the methodology used to solve the model. The probability generating functions associated with the various states of the server are obtained in section 5.4 while several performance measures for the analysis of the system are established in section 5.5. The numerical simulation and sensitivity analysis of the model have been facilitated in section 5.6. Finally, we wind up our investigation with the discussion in the section 5.7 at the end.

5.2 MODEL DESCRIPTION

Let us consider a single server unreliable bulk retrial queue with multi essential services and two types of customers. The basic assumptions for the formulation of the model are as follows:

- **Arrival Process:** The single server renders service to two types of customers, (i) non-preemptive priority customers and (ii) ordinary customers. Both types of customers arrive in batches according to a Poisson process with arrival rate λ_1 (λ_2) for the type 1 (2) customers. The priority is assigned to type 1 customers whereas type 2 customers are known as ordinary customers. Let X_1 and X_2 be the random variables denoting the batch sizes defined by Pr $\{X_1=m\} = c_m$; m $\geq$1 such that $\sum_{m=1}^{\infty} c_m$ =1 and Pr $\{X_2=n\} = c_n$; n $\geq$1 such that $\sum_{n=1}^{\infty} c_n$ =1 for priority and ordinary customers, respectively.
- **Non Preemptive Priority and Retrial Process:** The type 1 customers are served on arrival if they find the server idle or free otherwise they join the queue of priority customers. Moreover, type 2 customers are forced to join the retrial orbit if they find the server unavailable or busy with type 1 customers. Also, if an incoming priority customer finds that the server is busy with the ordinary customer, then he waits for the completion of service of that particular ordinary customer and cannot interrupt in between. This is known as non-preemptive priority scheme. But an incoming fresh ordinary customer can try for the service only when all the priority customers are served. The ordinary customers retry for their service with retrial rate γ to get served. The retrial process is assumed to be exponentially distributed.
- **Balking:** The customers may balk on seeing a long queue waiting for the service. The priority and ordinary incoming customers may balk with balking probability ($1-b_1$) and ($1-b_2$), respectively.
- **Service Process:** The priority customers are served earlier than the ordinary customers. For both priority as well as ordinary customers, the service is provided in k ($k \geq 1$) essential phases. The priority (ordinary) customer after completing first essential phase of a service moves to second phase with rate $\mu_{1,1}$ ($\mu_{2,1}$). Similarly, a priority and ordinary customer after being served in the i^{th} ($1 \leq i \leq k-1$) phase move to $(i+1)^{th}$ phase with rate $\mu_{1,i}$ and $\mu_{2,i}$, respectively. The service for all the customers are assumed to be

general distributed. The pdf, cdf , LST of pdf of the service pattern are $b_{v,i}$(t) , $B_{v,i}$(t) and $\tilde{b}_{v,i}(s)$ respectively (v =1 for priority customers and v =2 for ordinary customers).

- **Repair Process:** The server may break down exponentially with rate α_i ($1\leq i\leq k$) while providing any phase of service to the either type of customer (priority or ordinary). The broken down server is sent for repair immediately and the repair is completed in '*d*' compulsory phases for a server brokendown during any phase of service. The repair process is general distributed with repair rate $\beta_{1,i,j}$ ($\beta_{2,i,j}$), ($1\leq i\leq k$), ($1\leq j \leq d$) for the priority (ordinary customers). The pdf, cdf , LST of pdf for repair time are given as $g_{v,i,j}$(t), $G_{v,i,j}$(t) and $\tilde{g}_{v,i,j}(s)$ respectively where ($1\leq i \leq k$), ($1\leq j \leq d$) (v=1 for priority customers and v =2 for ordinary customers).

Application to Healthcare Departments

The application of the present queueing model can be realized in healthcare centers. The working of healthcare centers involves a number of medical and non-medical steps starting from the admission of the patient to the discharge of the patient after required treatment. There are several departments in the health care units where queues are built up; as such their performance and working can be assessed quantitatively via queue theoretic approach by developing suitable model based on the realistic assumptions. It may happen that a variety of patients arrive at the hospitals/healthcare departments for the treatment; some of them have ordinary health problems (like minor injuries, viral fever, routine checkup, etc.) whereas others may have severe problems (like major accidents, dengue fever, cancer, etc.).

The emergency patients with severe problems need treatment early as compared to those patients who are suffering with mild problems. There might be a group of patients waiting in the hospitals to seek doctors to have their test reports or to have some other type of consultations. The patients arriving in the group can be categorized in two broad categories on the basis of priority of treatments required. When there is no emergency patient in the system, the treatment to the ordinary patient is provided.

In our model we consider two types of customers arriving in batches, (i) critical patients (i.e. priority customers) and (ii) ordinary patients who can be treated only when there are no emergency/critical patients. In case an emergency patient arrives and the doctor is free, then he immediately starts the medical treatment to the priority patient; however if the concerned doctor is occupied, then the priority customer has to wait for his turn in the queue as the life of the patient under treatment can't be risked over the arrival of a new patient. In case when the doctor is busy and ordinary patients arrive, they have to wait in the waiting room (i.e. retrial orbit) for their turn; the patients from the orbit may try again and again for the treatment by checking whether the doctor is free or not.

The treatment of a patient is usually completed in a series of compulsory steps from the admission of patient, filling of forms, various tests, observation and many more and end with the discharge of patient. Various phases of compulsory services vary from patient to patient depending upon the treatment required. On arrival, a patient may balk on seeing many patients waiting for the treatment and can move to other hospitals with a hope to get faster treatment without waiting for a longer time.

It may happen that the appliances used to examine the patients are unavailable/ not in working state due to technical faults. But to continue the treatment smoothly, the broken down appliance/machine is to be repaired at once so as to continue the treatment. From the description of application in health care organization as mentioned above and shown by block diagram in Figure 5.1, it is clear that the working of hospitals is a very close real life application of the retrial queueing model with priority.

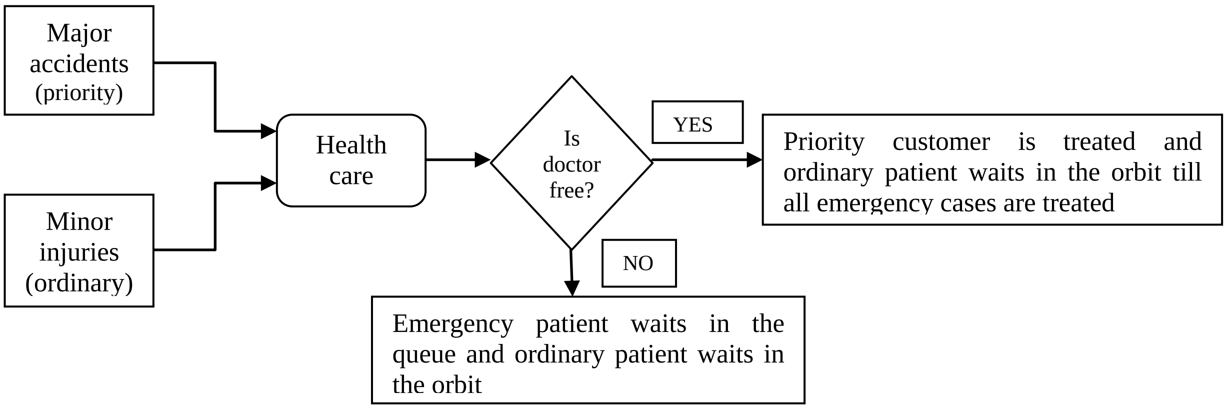


Fig. 5.1: Block diagram of the application of model in health care organization

5.3 QUEUE SIZE DISTRIBUTION

Let $N_1(t)$ and $N_2(t)$ represent the number of priority and ordinary customers respectively, in the system and $S_1(t)$ and $S_2(t)$ denote the phase of service and repair respectively, at time t.

The state of the server at any time t is given by

$$
Y(t)=\begin{cases} 0, & \text{server is in idle state} \\ 1, & \text{server is busy in providing service to the priority customers} \\ 2, & \text{server is busy in providing service to the ordinary customers} \\ 3, & \text{server is brokendown and under repair while servicing priority customers} \\ 4, & \text{server is brokendown and under repair while servicing ordinary customers} \end{cases}
$$

In the steady state, the joint distributions of the server state and queue size are defined as-

$$P_{0,m,n} = \lim_{t\to\infty} \Pr\{Y(t)=0, N_1(t)=m, N_2(t)=n\}, m\geq 0, n\geq 0$$

$$P_{1,m,n,i}(x) = \lim_{t\to\infty} \Pr\{Y(t)=1, x\leq \varpi(t)\leq x+dx, N_1(t)=m, N_2(t)=n, S_1(t)=i\},$$
$$m\geq 0, n\geq 0, (1\leq i\leq k)$$

$$P_{2,m,n,i}(x) = \lim_{t\to\infty} \Pr\{Y(t)=2, x\leq \varpi(t)\leq x+dx, N_1(t)=m, N_2(t)=n, S_1(t)=i\},$$
$$m\geq 0, n\geq 0, (1\leq i\leq k)$$

$$R_{1,m,n,i,j}(x,y) = \lim_{t\to\infty} \Pr\{Y(t)=3, \varpi(t)=x, y\leq \sigma(t)\leq y+dy, N_1(t)=m, N_2(t)=n, S_1(t)=i, S_2(t)=j\},$$
$$m\geq 0, n\geq 0, 1\leq i\leq k, 1\leq j\leq d$$

$$R_{1,m,n,i,j}(x,y) = \lim_{t\to\infty} \Pr\{Y(t)=4, \varpi(t)=x, y\leq \sigma(t)\leq y+dy, N_1(t)=m, N_2(t)=n, S_1(t)=i, S_2(t)=j\},$$
$$m\geq 0, n\geq 0, 1\leq i\leq k, 1\leq j\leq d$$

Before framing the equations, the stability condition for the retrial model under consideration is established in the form of proposition as follows:

Proposition 5.1: The necessary and sufficient condition for the system to be stable is

$$\rho_{11}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)+\rho_{21}\left(1+\sum_{j=1}^{d}\frac{\alpha_2}{\beta_{2,1,j}}\right)<1$$

where, $\rho_{11}=\dfrac{\lambda_1 E[X_1]b_1}{\mu_{1,1}}$ and $\rho_{21}=\dfrac{\lambda_2 E[X_2]b_2}{\mu_{2,1}}$

Proof: The proof is done on the pattern similar to that of proposition 2.1.

Now, using the approach based on supplementary variable technique, we formulate the equations for the model under consideration as follows:

Governing Equations

The set of governing equations and boundary conditions for different states of the server after introducing the supplementary variables are constructed as follows:

$$\left[\frac{d}{dx}+\lambda_1 b_1+\lambda_2 b_2+\mu_{1,i}(x)+\alpha_i\right]P_{1,m,n,i}(x)=\int_0^\infty R_{1,m,n,i,d}(x,y)\beta_{1,i,d}(y)dy+\sum_{m_1=1}^{s}\lambda_1 b_1 c_{1m_1}P_{1,m-m_1,n,i}(x) + \sum_{m_1=1}^{s}\lambda_2 b_2 c_{2m_1}P_{1,m,n-m_1,i}(x) \tag{5.1}$$

$$\left[\frac{d}{dx}+\lambda_1 b_1+\lambda_2 b_2+\mu_{2,i}(x)+\alpha_i\right]P_{2,m,n,i}(x)=\int_0^\infty R_{2,m,n,i,d}(x,y)\beta_{2,i,d}(y)dy+\sum_{m_1=1}^{s}\lambda_1 b_1 c_{1m_1}P_{2,m-m_1,n,i}(x) + \sum_{m_1=1}^{s}\lambda_2 b_2 c_{2m_1}P_{2,m,n-m_1,i}(x) \tag{5.2}$$

$$(\lambda_1 b_1+\lambda_2 b_2+n\gamma)P_{0,0,n}=\int_0^\infty P_{1,0,n,1}(x)\mu_{1,1}(x)dx+\int_0^\infty P_{2,0,n,1}(x)\mu_{2,1}(x)dx \tag{5.3}$$

$$\left[\frac{\partial}{\partial y}+\lambda_1 b_1+\lambda_2 b_2+\beta_{1,i,j}(y)\right]R_{1,m,n,i,j}(x,y)=\sum_{m_1=1}^{s}\lambda_1 b_1 c_{1m_1}R_{1,m-m_1,n,i,j}(x,y) + \sum_{m_1=1}^{s}\lambda_2 b_2 c_{2m_1}R_{1,m,n-m_1,i,j}(x,y) \tag{5.4}$$

$$\left[\frac{\partial}{\partial y}+\lambda_1 b_1+\lambda_2 b_2+\beta_{2,i,j}(y)\right]R_{2,m,n,i,j}(x,y)=\sum_{m_1=1}^{s}\lambda_1 b_1 c_{1m_1}R_{2,m-m_1,n,i,j}(x,y) + \sum_{m_1=1}^{s}\lambda_2 b_2 c_{2m_1}R_{2,m,n-m_1,i,j}(x,y) \tag{5.5}$$

Boundary Conditions

$$P_{1,m,n,1}(0)=\sum_{m_1=1}^{m+1}\lambda_1 b_1 c_{1m_1}P_{0,0,m-m_1+1}\delta_{i,0}+\int_0^\infty P_{1,m+1,n,1}(x)\mu_{1,1}(x)dx+\int_0^\infty P_{2,m+1,n,1}(x)\mu_{2,1}(x)dx \tag{5.6}$$

$$P_{2,m,n,1}(0)=\sum_{m_1=1}^{n+1}\lambda_2 b_2 c_{2m_1}P_{0,0,n-m_1+1}+(n+1)\gamma P_{0,0,n+1}, \quad \text{if } m=0 \tag{5.7}$$

$$P_{2,m,n,1}(0)=0,\ if\ m\geq 1 \tag{5.8}$$

$$P_{1,m,n,i}(0)=\int_0^\infty P_{1,m,n,i-1}(x)\mu_{1,i-1}(x)dx, \quad (2\leq i\leq k) \tag{5.9}$$

$$P_{2,m,n,i}(0)=\int_0^\infty P_{2,m,n,i-1}(x)\mu_{2,i-1}(x)dx, \quad (2\leq i\leq k) \tag{5.10}$$

$$R_{1,m,n,i,1}(0)=\alpha_i P_{1,m,n,i}(x), \qquad (1\leq i\leq k) \tag{5.11}$$

$$R_{2,m,n,i,1}(0)=\alpha_i P_{2,m,n,i}(x), \qquad (1\leq i\leq k) \tag{5.12}$$

$$R_{1,m,n,i,j}(0)=\int_0^\infty R_{1,m,n,i,j-1}(x,y)\,\beta_{1,i,j-1}(y)\,dy,\ (1\le i\le k),(2\le j\le d) \quad (5.13)$$

$$R_{2,m,n,i,j}(0)=\int_0^\infty R_{2,m,n,i,j-1}(x,y)\,\beta_{2,i,j-1}(y)\,dy,\ (1\le i\le k),(2\le j\le d) \quad (5.14)$$

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\left[P_{0,0,n}+\sum_{i=1}^{k}\left[\int_0^\infty P_{1,m,n,i}(x)dx+\int_0^\infty P_{2,m,n,i}(x)dx\right]\right] + \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{i=1}^{k}\sum_{j=1}^{d}\left[\int_0^\infty\int_0^\infty R_{1,m,n,i,j}(x,y)dxdy+\int_0^\infty\int_0^\infty R_{2,m,n,i,j}(x,y)dxdy\right]=1 \quad (5.15)$$

Following the procedure as used in chapters 2 and 3, we define the probability generating functions corresponding to the various states as:

$$P_0(z_1,z_2)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}P_{0,0,n}z_1{}^m z_2{}^n,\qquad P_{v,i}(z_1,z_2,x)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}P_{v,m,n,i}(x)z_1{}^m z_2{}^n,$$

$$P_{v,d}(z_1,z_2,x)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}P_{v,m,n,d}(x)z_1{}^m z_2{}^n,\qquad C_1(z_1)=\sum_{m_1=1}^{\infty}C_{1m_1}z_1{}^{m_1},\ C_2(z_2)=\sum_{m_1=1}^{\infty}C_{2m_1}z_2{}^{m_1}$$

$$R_{v,i,j}(z_1,z_2,x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}R_{v,m,n,i,j}(x,y)z_1{}^m z_2{}^n$$

Here $v=1$ for priorty and v=2 for non-priorty customers; $1\le i\le k, 1\le j\le d$.

The hazard rates corresponding to the the service state and repair state are as follows:

$$\mu_{1,i}(x)=\frac{b_{1,i}(x)}{1-B_{1,i}(x)}\quad and\quad \mu_{2,i}(x)=\frac{b_{2,i}(x)}{1-B_{2,i}(x)}\ (1\le i\le k)$$

$$\beta_{1,i,j}(y)=\frac{g_{1,i,j}(y)}{1-G_{1,i,j}(y)}\quad and\quad \beta_{2,i,j}(y)=\frac{g_{2,i,j}(y)}{1-G_{2,i,j}(y)}\ (1\le j\le d)$$

Multiplying equations (5.1)-(5.15) by $z_1{}^m$. $z_2{}^n$ and summing over all values of m and n; and then using generating functions, the above set of equations (5.1)-(5.15) converts to

$$\left[\frac{d}{dx}+\lambda_1 b_1(1-C_1(z_1)+\lambda_2 b_2(1-C_2(z_2)+\mu_{1,i}(x)+\alpha_i\right]P_{1,i}(z_1,z_2,x)=\int_0^\infty R_{1,i,d}(z_1,z_2,x,y)g_{1,i,d}(y)dy \quad (5.16)$$

$$\left[\frac{d}{dx}+\lambda_1 b_1(1-C_1(z_1)+\lambda_2 b_2(1-C_2(z_2)+\mu_{2,i}(x)+\alpha_i\right]P_{2,i}(z_1,z_2,x)=\int_0^\infty R_{2,i,d}(z_1,z_2,x,y)g_{2,i,d}(y)dy \quad (5.17)$$

$$(\lambda_1 b_1+\lambda_2 b_2)P_{0,0}(z_1,z_2)+z_2\,\gamma P'_{0,0}(z_1,z_2)=\int_0^\infty P_{1,1}(0,z_2,x)\mu_{1,1}(x)dx+\int_0^\infty P_{2,1}(0,z_2,x)\mu_{2,1}(x)dx \quad (5.18)$$

$$\left[\frac{\partial}{\partial y}+\lambda_1 b_1(1-C_1(z_1)+\lambda_2 b_2(1-C_2(z_2)+\beta_{1,i,j}(y)\right]R_{1,i,j}(z_1,z_2,x,y)=0 \quad (5.19)$$

$$\left[\frac{\partial}{\partial y}+\lambda_1 b_1(1-C_1(z_1)+\lambda_2 b_2(1-C_2(z_2)+\beta_{2,i,j}(y)\right]R_{2,i,j}(z_1,z_2,x,y)=0 \quad (5.20)$$

$$z_1P_{1,1}(z_1,z_2,0)=z_1\lambda_1b_1C_1(z_1)P_0(z_2)+\int_0^\infty\left[P_{1,1}(z_1,z_2,x)-P_{1,1}(0,z_2,x)\right]\mu_{1,1}(x)dx$$
$$+\int_0^\infty\left[P_{2,1}(z_1,z_2,x)-P_{2,1}(0,z_2,x)\right]\mu_{2,1}(x)dx \tag{5.21}$$

$$P_{1,i}(z_1,z_2,0)=\int_0^\infty P_{1,i-1}(z_1,z_2,x)\mu_{1,i-1}(x)dx,\quad (2\le i\le k) \tag{5.22}$$

$$P_{2,1}(z_1,z_2,0)=\frac{\lambda_2b_2C_2(z_2)P_{0,0}(z_1,z_2)}{z_2}+\gamma\frac{dP_{0,0}(z_1,z_2)}{z_2} \tag{5.23}$$

$$P_{2,i}(z_1,z_2,0)=\int_0^\infty P_{2,i-1}(z_1,z_2,x)\mu_{2,i-1}(x)dx,\quad (2\le i\le k) \tag{5.24}$$

$$R_{1,i,1}(z_1,z_2,x,0)=\alpha_iP_{1,i}(z_1,z_2,x),\quad (1\le i\le k) \tag{5.25}$$

$$R_{2,i,1}(z_1,z_2,x,0)=\alpha_iP_{2,i}(z_1,z_2,x),\quad (1\le i\le k) \tag{5.26}$$

$$R_{1,i,j}(z_1,z_2,x,0)=\int_0^\infty R_{1,i,j-1}(z_1,z_2,x,y)\,\beta_{1,i,j-1}(y)\,dy,\ (1\le i\le k),(2\le j\le d) \tag{5.27}$$

$$R_{2,i,j}(z_1,z_2,x,0)=\int_0^\infty R_{2,i,j-1}(z_1,z_2,x,y)\,\beta_{2,i,j-1}(y)\,dy,\ (1\le i\le k),(2\le j\le d) \tag{5.28}$$

$$\lim_{z_1,z_2\to1}\left[P_0(z_1,z_2)+\sum_{i=1}^{k}\left(\int_0^\infty P_{1,i}(z_1,z_2,x)dx+\int_0^\infty P_{2,i}(z_1,z_2,x)dx\right)\right]$$
$$+\lim_{z_1,z_2\to1}\sum_{i=1}^{k}\sum_{j=1}^{d}\left[\int_0^\infty\int_0^\infty R_{1,i,j}(z_1,z_2,x,y)dxdy+\int_0^\infty\int_0^\infty R_{2,i,j}(z_1,z_2,x,y)dxdy\right]=1$$

Now, we give our results for the partial generating functions and marginal generating functions for the different states of the system in the form of theorems as follows:

Theorem 5.1: The partial probability generating functions for the server being in idle state, busy in servicing priority customers, busy in servicing ordinary customers, under repair when failed while servicing the priority and ordinary respectively, are:

$$P_0(z_2)=\frac{\left[1-\rho_{11}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)-\rho_{21}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{2,1,j}}\right)\right]}{\left[1-\rho_{11}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)-\rho_{21}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{2,1,j}}\right)+\sum_{v=1}^{2}\left(\sum_{i=1}^{k}\sum_{j=1}^{d}\left(\rho_{v1}\frac{1}{\mu_{v,1}}\left(\prod_{s=1}^{i}\frac{1}{\mu_{v,s}}\right)\left(1+\prod_{r=1}^{j}\frac{\alpha_v}{\beta_{v,1,j}}\right)\right)\right)\right]}$$
$$\times\exp\left[\frac{1}{\gamma}\int_1^{z_2}\frac{\lambda_1b_1+\lambda_2b_2-\lambda_1b_1C_1(h(v))-\dfrac{\lambda_2b_2C_2(v)k_{2,1}(h(v),v)}{v}}{\left[k_{2,1}(h(v),v)-v\right]}dv\right] \tag{5.29}$$

$$P_{1,i}(z_1,z_{2,}x)=\left\{\begin{array}{l}\left[\left[\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(h(z_2))-\dfrac{k_{2,1}(h(z_2),z_2)\lambda_2 b_2 C_2(z_2)}{z_2}\right)\left(k_{2,1}(z_1,z_2)-z_2\right)\right]\right] \\ -\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(z_1)-\dfrac{k_{2,1}(z_1,z_2)\lambda_2 b_2 C_2(z_2)}{z_2}\right)\left[k_{2,1}(h(z_2),z_2)-z_2\right]\end{array}\right\}$$

$$\times\prod_{s=1}^{i-1}\tilde{b}_{1,s}\left[N_{1,s}(z_1,z_2)\right]\times\exp\left[-\left\{N_{1,i}(z_1,z_2)\right\}x\right]\bar{B}_{1,i}(x)$$

$$\times\left[z_1-k_{1,1}(h(z_2),z_2)\right]^{-1}\left[k_{2,1}(h(z_2),z_2)-z_2\right]^{-1}P_0(z_2) \qquad (1\le i\le k) \tag{5.30}$$

$$P_{2,i}(z_1,z_{2,}x)=\left[\frac{\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(h(z_2))-\lambda_2 b_2 C_2(z_2)\right)}{\left[k_{2,1}(h(z_2),z_2)-z_2\right]}\right]\exp\left[-\left\{N_{2,i}(z_1,z_2)\right\}x\right]\bar{B}_{2,i}(x)$$

$$\times\prod_{s=1}^{i-1}\tilde{b}_{2,s}\left[N_{2,r}(z_1,z_2)\right]P_0(z_2) \qquad (1\le i\le k) \tag{5.31}$$

$$R_{1,i,j}(z_1,z_2,x,y)=\alpha_i P_{1,i}(z_1,z_2,x)\prod_{r=1}^{j-1}\tilde{g}_{1,i,r}(M(z_1,z_2))\exp\left[-\left\{M(z_1,z_2)\right\}y\right]\bar{G}_{1,i,j}(y),$$
$$(1\le i\le k),(1\le j\le d) \tag{5.32}$$

$$R_{2,i,j}(z_1,z_2,x,y)=\alpha_i P_{2,i}(z_1,z_2,x)\prod_{r=1}^{j-1}\tilde{g}_{2,i,r}(M(z_1,z_2))\exp\left[-\left\{M(z_1,z_2)\right\}y\right]\bar{G}_{2,i,j}(y)$$
$$(1\le i\le k),(1\le j\le d) \tag{5.33}$$

where,

$$M(z_1,z_2)=\lambda_1 b_1(1-C_1(z_1))+\lambda_2 b_2(1-C_2(z_2)) \tag{5.34}$$

$$N_{1,i-1}(z_1,z_2)=M(z_1,z_2)+\alpha_{i-1}\left[1-\prod_{r=1}^{d}\tilde{g}_{1,i-1,r}\left(M(z_1,z_2)\right)\right] \tag{5.35}$$

$$N_{2,i-1}(z_1,z_2)=M(z_1,z_2)+\alpha_{i-1}\left[1-\prod_{r=1}^{d}\tilde{g}_{2,i-1,r}\left(M(z_1,z_2)\right)\right] \tag{5.36}$$

$$\tilde{b}_{1,1}\left(N_{1,1}(z_1,z_2)\right)=k_{1,1}(z_1,z_2)\ and\ \tilde{b}_{2,1}\left(N_{2,1}(z_1,z_2)\right)=k_{2,1}(z_1,z_2)$$

Note: For the brevity, we use the product $\left(\prod_{r=1}^{j-1}\tilde{g}_{1,i,r}(M(z_1,z_2))\right)$ and $\left(\prod_{r=1}^{j-1}\tilde{g}_{2,i,r}(M(z_1,z_2))\right)$ in our results. However, the value of these products equals 1 when j=1.

Proof: To obtain the probability generating functions we proceed as:

On solving equations (5.19) and (5.20), we get

$$R_{1,i,j}(z_1,z_2,x,y)=R_{1,i,j}(z_1,z_2,x,0)\exp\left[-\{M(z_1,z_2)\}y\right]\bar{G}_{1,i,j}(y) \tag{5.37}$$

$$R_{2,i,j}(z_1,z_2,x,y)=R_{2,i,j}(z_1,z_2,x,0)\exp\left[-\{M(z_1,z_2)\}y\right]\bar{G}_{2,i,j}(y) \tag{5.38}$$

Using eqs (5.25)-(5.26) in (5.37)-(5.38), we get results given in (5.32) and (5.33). Further, using eqs (5.32)-(5.33) and (5.16)-(5.17), we get

$$P_{1,i}(z_1,z_2,x)=P_{1,i}(z_1,z_2,0)\exp\left[-\{N_{1,i-1}(z_1,z_2)\}x\right]\bar{B}_{1,i}(x)P_0(z_2) \tag{5.39}$$

$$P_{2,i}(z_1,z_2,x)=P_{2,i}(z_1,z_2,0)\exp\left[-\{N_{2,i-1}(z_1,z_2)\}x\right]\bar{B}_{2,i}(x)P_0(z_2) \tag{5.40}$$

substituting i= (i-1) in eq.(5.39) and using it in eq. (5.22), we obtain

$$P_{1,i}(z_1,z_2,0)=P_{1,i-1}(z_1,z_2,0)\tilde{b}_{1,i-1}\left[N_{1,i-1}(z_1,z_2)\right],(2\le i\le k) \tag{5.41}$$

using eqs (5.39)-(5.40) for i=1 and (5.18) and (5.21), we obtain

$$\begin{aligned} z_1P_{1,1}(z_1,z_2,0)&=z_1\lambda_1b_1C_1(z_1)P_0(z_2)+P_{1,1}(z_1,z_2,0)\tilde{b}_{1,1}\left[N_{1,1}(z_1,z_2)\right]\\ &+P_{2,1}(z_1,z_2,0)\tilde{b}_{2,1}\left[N_{2,1}(z_1,z_2)\right]-(\lambda_1b_1+\lambda_2b_2)P_0(z_2)-z_2\,\gamma P_0'(z_2)\end{aligned} \tag{5.42}$$

Further using eqs (5.23) and (5.42), we get

$$P_{1,1}(z_1,z_2,0)=\frac{\gamma P_0'(z_2)\left[\tilde{b}_{2,1}\left[N_{2,1}(z_1,z_2)\right]-z_2\right]+\left[\dfrac{\lambda_2b_2C_2(z_2)\left(\tilde{b}_{2,1}\left[N_{2,1}(z_1,z_2)\right]\right)}{z_2}-\lambda_1b_1-\lambda_2b_2+\lambda_1b_1C_1(z_1)\right]P_0(z_2)}{\left[z_1-\tilde{b}_{1,1}\left(N_{1,1}(z_1,z_2)\right)\right]}$$

Let $\tilde{b}_{1,1}\left(N_{1,1}(z_1,z_2)\right)=k_{1,1}(z_1,z_2)$ and $\tilde{b}_{2,1}\left(N_{2,1}(z_1,z_2)\right)=k_{2,1}(z_1,z_2)$

Thus,

$$P_{1,1}(z_1,z_2,0)=\frac{\gamma P_0'(z_2)\left[k_{2,1}(z_1,z_2)-z_2\right]-\left[\dfrac{-\lambda_2b_2C_2(z_2)k_{2,1}(z_1,z_2)}{z_2}+\lambda_1b_1+\lambda_2b_2-\lambda_1b_1C_1(z_1)\right]P_0(z_2)}{\left[z_1-k_{1,1}(z_1,z_2)\right]}$$

Also, $P_0(z_2)=P_0(z_1,z_2)$.

If $\left[z_1-k_{1,1}(z_1,z_2)\right]=0$, then

$$\gamma P_0'(z_2)\left[k_{2,1}(z_1,z_2)-z_2\right]-\left[\frac{-\lambda_2b_2C_2(z_2)k_{2,1}(z_1,z_2)}{z_2}+\lambda_1b_1+\lambda_2b_2-\lambda_1b_1C_1(z_1)\right]P_0(z_2)=0$$

which is a first order linear differential equation and on solving gives eq. (5.29). Mathematically, the above differential equation has a unique root $z_1=h(z_2)$ in the $|z|\le 1$ [cf. Falin and Templeton, 1997]. Hence, we obtain partial probability generating functions in the above manner. Further,we denote $z_1=h(z_2)$ so as to obtain probability generating function. Here, $z_1 = h(z_2)$ is defined as the generating function of the number of ordinary jobs that arrive during the busy period formed by the priority customers.

Theorem 5.2: The marginal probability generating functions at random epochs when the server is busy with i^{th} ($1\le i\le k$) phase service of priority and ordinary customers and under j^{th} ($1\le j\le d$) phase repair while brokendown either during servicing of priority or ordinary customers respectively, are

$$P_{1,i}(z_1,z_2) = P_0(z_2)\left\{\begin{array}{l}\left[\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(h(z_2))-k_{2,1}(h(z_2),z_2)\lambda_2 b_2 C_2(z_2)\right)\left(k_{2,1}(z_1,z_2)-z_2\right)\right]\\ -\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(z_1)-\dfrac{k_{2,1}(z_1,z_2)\lambda_2 b_2 C_2(z_2)}{z_2}\right)\left[k_{2,1}(h(z_2),z_2)-z_2\right]\end{array}\right\}$$
$$\times\left[z_1-k_{1,1}(h(z_2),z_2)\right]^{-1}\left[k_{2,1}(h(z_2),z_2)-z_2\right]^{-1} \tag{5.43}$$
$$\times\left[\frac{\left[1-\tilde{b}_{1,i}\left(N_{1,i}(z_1,z_2)\right)\right]}{N_{1,i}(z_1,z_2)}\prod_{s=1}^{i-1}\tilde{b}_{1,s}\left(N_{1,s}(z_1,z_2)\right)\right],(1\le i\le k)$$

$$P_{2,i}(z_1,z_2) = P_0(z_2)\left[\frac{\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(h(z_2))-\lambda_2 b_2 C_2(z_2)\right)}{\left[k_{2,1}(h(z_2),z_2)-z_2\right]}\right]$$
$$\times\left[\frac{\left[1-\tilde{b}_{2,i}\left(N_{2,1}(z_1,z_2)\right)\right]}{N_{2,i}(z_1,z_2)}\prod_{s=1}^{i-1}\tilde{b}_{2,s}\left(N_{2,s}(z_1,z_2)\right)\right],(2\le i\le k) \tag{5.44}$$

$$R_{1,i,j}(z_1,z_2)=\alpha_i P_{1,i}(z_1,z_2)\prod_{r=1}^{j-1}\tilde{g}_{1,i,r}(M(z_1,z_2))\frac{\left[1-\tilde{g}_{1,i,j}(M(z_1,z_2))\right]}{M(z_1,z_2)}, \tag{5.45}$$
$$(1\le i\le k)\text{ and }(1\le j\le d)$$

$$R_{2,i,j}(z_1,z_2)=\alpha_i P_{2,i}(z_1,z_2)\prod_{r=1}^{j-1}\tilde{g}_{2,i,r}(M(z_1,z_2))\frac{\left[1-\tilde{g}_{2,i,j}(M(z_1,z_2))\right]}{M(z_1,z_2)}, \tag{5.46}$$
$$(1\le i\le k)\text{ and }(1\le j\le d)$$

Note: For the brevity, we use the product $\prod_{s=1}^{i-1}\tilde{b}_{1,s}\left(N_{1,s}(z_1,z_2)\right)$ and $\prod_{s=1}^{i-1}\tilde{b}_{2,s}\left(N_{2,s}(z_1,z_2)\right)$ as notations in our results. However, the value of these products equals 1 when i=1.

Proof: The marginal generating functions for various states of the server are obtained as-

$$P_{1,i}(z_1,z_2)=\int_0^\infty P_{1,i}(z_1,z_2,x)dx,\ P_{2,i}(z_1,z_2)=\int_0^\infty P_{2,i}(z_1,z_2,x)dx,\ (1\le i\le k)$$

$$R_{1,i,j}(z_1,z_2)=\int_0^\infty\int_0^\infty R_{1,i,j}(z_1,z_2,x,y)dxdy,$$

$$R_{2,i,j}(z_1,z_2)=\int_0^\infty\int_0^\infty R_{2,i,j}(z_1,z_2,x,y)dxdy,(1\le i\le k)\,\text{and}\,(1\le j\le d)$$

Theorem 4.3: The generating function for the queue size distribution under steady state is

$$\begin{aligned}
L(z_1,z_2)=P_0(z_2)+\sum_{i=1}^{k}\sum_{j=1}^{d}P_0(z_2)&\left(1+\prod_{r=1}^{j-1}\tilde{g}_{1,i,r}(M(z_1,z_2))\frac{\alpha_i\left[1-\tilde{g}_{1,i,j}(M(z_1,z_2))\right]}{M(z_1,z_2)}\right)\\
&\times\left[z_1-k_{1,1}(h(z_2),z_2)\right]^{-1}\left[k_{2,1}(h(z_2),z_2)-z_2\right]^{-1}\\
&\times\left\{\begin{array}{l}\left[\left(\lambda_1b_1+\lambda_2b_2-\lambda_1b_1C_1(h(z_2))-k_{2,1}(h(z_2),z_2)\lambda_2b_2C_2(z_2)\right)\left(k_{2,1}(z_1,z_2)-z_2\right)\right]\\-\left(\lambda_1b_1+\lambda_2b_2-\lambda_1b_1C_1(z_1)-\dfrac{k_{2,1}(z_1,z_2)\lambda_2b_2C_2(z_2)}{z_2}\right)\left[k_{2,1}(h(z_2),z_2)-z_2\right]\end{array}\right\}\\
&\times\left[\frac{\left[1-\tilde{b}_{1,i}\left(N_{1,1}(z_1,z_2)\right)\right]}{N_{1,i}(z_1,z_2)}\prod_{s=1}^{i-1}\tilde{b}_{1,s}\left(N_{1,s}(z_1,z_2)\right)\right]\\
&+\sum_{i=1}^{k}\sum_{j=1}^{d}P_0(z_2)\left(1+\prod_{r=1}^{j-1}\tilde{g}_{2,i,r}(M(z_1,z_2))\frac{\alpha_i\left[1-\tilde{g}_{2,i,j}(M(z_1,z_2))\right]}{M(z_1,z_2)}\right)\\
&\times\left[\frac{\left(\lambda_1b_1+\lambda_2b_2-\lambda_1b_1C_1(h(z_2))-\lambda_2b_2C_2(z_2)\right)}{\left[k_{2,1}(h(z_2),z_2)-z_2\right]}\right]\left[\frac{\left[1-\tilde{b}_{2,i}\left(N_{2,1}(z_1,z_2)\right)\right]}{N_{2,i}(z_1,z_2)}\prod_{r=1}^{i-1}\tilde{b}_{2,r}\left(N_{2,r}(z_1,z_2)\right)\right]
\end{aligned}\tag{5.47}$$

Proof: The queue size distribution is obtained using

$$L(z_1,z_2)=P_0(z_2)+\sum_{i=1}^{k}P_{1,i}(z_1,z_2)+\sum_{i=1}^{k}P_{2,i}(z_1,z_2)+\sum_{i=1}^{k}\sum_{j=1}^{d}R_{1,i,j}(z_1,z_2)+\sum_{i=1}^{k}\sum_{j=1}^{d}R_{2,i,j}(z_1,z_2),$$
$$(1\le i\le k)\,\text{and}\,(1\le j\le d)$$

5.4 PERFORMANCE MEASURES

The key performance measures which are of interest for the analysis of unreliable server queueing system are long run probabilities, queue length, availability as well as failure frequency which are determined as follows:

(A) Long Run Probabilities

The probability measures characterizing the server status over a long run of time are obtained in the following theorem:

Theorem 4.4: The various long run probabilities of the server are as follows:

- Long run probability of the server being in idle state is

$$P(I) = P_0(1) \tag{5.48}$$

- Long run probability of the server being busy in rendering service to the priority customer

$$P(B_1) = \sum_{i=1}^{k} \frac{\rho_{11}\mu_{11}}{\left[1-\rho_{11}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)-\rho_{21}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{2,1,j}}\right)\right]}\left(\prod_{s=1}^{i}\frac{1}{\mu_{1,s}}\right)P_0(1), (1 \le i \le k) \tag{5.49}$$

- Long run probability of the server being busy in rendering service to the ordinary customer

$$P(B_2) = \sum_{i=1}^{k} \frac{\rho_{21}\mu_{21}}{\left[1-\rho_{11}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)-\rho_{21}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{2,1,j}}\right)\right]}\left(\prod_{s=1}^{i}\frac{1}{\mu_{2,s}}\right)P_0(1), (1 \le i \le k) \tag{5.50}$$

- Long run probability that the server is broken down while rendering service to the priority customer and under repair is

$$P(R_1) = \sum_{i=1}^{k}\sum_{j=1}^{d} \frac{\rho_{11}\mu_{11}}{\left[1-\rho_{11}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)-\rho_{21}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{2,1,j}}\right)\right]}\left(\prod_{s=1}^{i}\frac{1}{\mu_{1,s}}\right)\left(\alpha_i\prod_{r=1}^{j}\frac{1}{\beta_{1,1,j}}\right)P_0(1) \tag{5.51}$$

- Long run probability that the server is broken down while rendering service to the ordinary customer and under repair is

$$P(R_2) = \sum_{i=1}^{k}\sum_{j=1}^{d} \frac{\rho_{21}\mu_{21}}{\left[1-\rho_{11}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)-\rho_{21}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{2,1,j}}\right)\right]}\left(\prod_{s=1}^{i}\frac{1}{\mu_{2,s}}\right)\left(\alpha_i\prod_{r=1}^{j}\frac{1}{\beta_{2,1,j}}\right)P_0(1) \tag{5.52}$$

Proof: The above expressions (5.29)-(5.33) for the long run probabilities are obtained by using

$$P(I) = \lim_{z_2\to 1} P_0(z_2),$$

$$P(B_1) = \lim_{z_1,z_2\to 1}\left[P_{1,1}(z_1,z_2)+\sum_{i=2}^{k}P_{1,i}(z_1,z_2)\right],\; P(B_2) = \lim_{z_1,z_2\to 1}\left[P_{2,1}(z_1,z_2)+\sum_{i=2}^{k}P_{2,i}(z_1,z_2)\right]$$

$$P(R_1) = \lim_{z_1,z_2\to 1}\left[\sum_{j=1}^{d}R_{1,1,j}(z_1,z_2)+\sum_{j=1}^{d}\sum_{i=2}^{k}R_{1,i,j}(z_1,z_2)\right]$$

$$P(R_2) = \lim_{z_1,z_2\to 1}\left[\sum_{j=1}^{d}R_{2,1,j}(z_1,z_2)+\sum_{j=1}^{d}\sum_{i=2}^{k}R_{2,i,j}(z_1,z_2)\right]$$

(B) Reliability Measures

The present section deals with the reliability indices of the retrial model under consideration. The reliability measures are established in the following theorem.

Theorem 5.5: The steady state availability (A_v) and failure frequency (F_f) of the server are obtained using

$$A_v = \sum_{i=1}^{k}\left(\frac{\left(\rho_{11}\mu_{11}\left(\left(\prod_{s=1}^{i}\frac{1}{\mu_{1,s}}\right)\right)+\rho_{21}\mu_{21}\left(\prod_{s=1}^{i}\frac{1}{\mu_{2,s}}\right)\right)}{\left[1-\rho_{11}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)-\rho_{21}\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{2,1,j}}\right)\right]}+1\right)P_0(1) \tag{5.53}$$

$$F_f = \sum_{i=1}^{k}\left(\frac{\alpha_i\left(\rho_{11}\mu_{11}\left(\left(\prod_{s=1}^{i}\frac{1}{\mu_{1,s}}\right)\right)+\rho_{21}\mu_{21}\left(\prod_{s=1}^{i}\frac{1}{\mu_{2,s}}\right)\right)}{\left[1-\rho_{11}b_1\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{1,1,j}}\right)-\rho_{21}b_2\left(1+\sum_{j=1}^{d}\frac{\alpha_1}{\beta_{2,1,j}}\right)\right]}\right)P_0(1) \tag{5.54}$$

Proof: The above results (5.34) and (5.35) for the steady state availability and failure frequency respectively, are derived using

$$A_v = P(I)+P(B_1)+P(B_2)$$

$$F_f = \alpha_i\left(P(B_1)+P(B_2)\right),(1\le i\le k)$$

(C) Queue Length

The results for the average queue length of the system for both priority and ordinary customers are presented below in the form of the theorem. For brevity, the various notations used in the explicit expressions for the queue length are described here as-

$$A=\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(h(z_2))-\frac{\lambda_2 b_2 C_2(z_2)k_{2,1}(h(z_2),z_2)}{z_2}\right)$$

$$B=\left(k_{2,1}(z_1,z_2)-z_2\right),\ \frac{\partial B(1)}{\partial z_1}=\frac{\lambda_1 b_1 C_1'(1)}{\mu_{2,1}}\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right),\ \frac{\partial B(1)}{\partial z_2}=\frac{\lambda_2 b_2 C_2'(1)}{\mu_{2,1}}\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right)-1$$

$$\frac{\partial^2 B}{\partial z_1^{\,2}}=b_{2,1}^{(2)}\left(\frac{\partial g}{\partial z_1}\right)^2+b_{2,1}^{(1)}\left(\frac{\partial^2 g}{\partial z_1^2}\right),\ \frac{\partial^2 B}{\partial z_2^{\,2}}=b_{2,1}^{(2)}\left(\frac{\partial g}{\partial z_2}\right)^2+b_{2,1}^{(1)}\left(\frac{\partial^2 g}{\partial z_2^2}\right)$$

$$C=\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(z_1)-\frac{k_{2,1}(z_1,z_2)\lambda_2 b_2 C_2(z_2)}{z_2}\right),\frac{\partial C(1)}{\partial z_1}=-\lambda_1 b_1 C_1'(1)-\rho_{2,1}\lambda_1 b_1 C_1'(1)\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right)$$

$$\frac{\partial C(1)}{\partial z_2}=\lambda_2 b_2-\lambda_2 b_2 C_2'(1)-\rho_{2,1}\lambda_2 b_2 C_2'(1)\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right),\frac{\partial^2 C}{\partial z_1{}^2}=-\lambda_2 b_2\left(\frac{\partial^2 B}{\partial z_1{}^2}\right)-\lambda_1 b_1 C_1''(1)$$

$$\frac{\partial^2 C}{\partial z_2{}^2}=-\lambda_2 b_2\left(\frac{\partial^2 B}{\partial z_2{}^2}\right)-2\lambda_2 b_2 C_2''(1)\left(1+\frac{\partial B}{\partial z_2}\right)+2\lambda_2 b_2\left(1+\frac{\partial B}{\partial z_2}\right)-\lambda_2 b_2 C_2''(1)+2\lambda_2 b_2 C_2'(1)$$

$$D=\left[k_{2,1}(h(z_2),z_2)-z_2\right],\ \frac{\partial D}{\partial z_2}=\frac{\lambda_2 b_2 C_2'(1)}{\mu_{2,1}}\left[\frac{\lambda_1 b_1 C_1'(1)\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right)}{1-\rho_{1,1,j}\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right)}+1\right]\left[\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right)-1\right]$$

$$E=z_1-k_{1,1}(z_1,z_2),\ \frac{\partial E(1)}{\partial z_1}=1-\rho_{1,1}b_1\left(1+\frac{\alpha_1}{\beta_{1,1,j}}\right),\frac{\partial E(1)}{\partial z_2}=-\frac{\lambda_2 b_2 C_2'(1)}{\mu_{1,1}}\left(1+\frac{\alpha_1}{\beta_{1,1,j}}\right)$$

$$\frac{\partial^2 E}{\partial z_1{}^2}=-b_{1,1}^{(2)}\left(\frac{\partial G_1}{\partial z_1}\right)^2-b_{1,1}^{(1)}\left(\frac{\partial^2 G_1}{\partial z_1^2}\right),\frac{\partial^2 E}{\partial z_2^2}=-b_{1,1}^{(2)}\left(\frac{\partial G_1}{\partial z_2}\right)^2-b_{1,1}^{(1)}\left(\frac{\partial^2 G_1}{\partial z_2^2}\right)$$

$$F_1=\left[1-\tilde{q}_{1,1}\left(N_{1,1}(z_1,z_2)\right)\right],\ \frac{\partial F_1}{\partial z_1}=-b_1\rho_{1,1}\left(1+\frac{\alpha_1}{\beta_{1,1,j}}\right),\frac{\partial F_1}{\partial z_2}=-\frac{\lambda_2 b_2 C_2'(1)}{\mu_{1,1}}\left(1+\frac{\alpha_1}{\beta_{1,1,j}}\right)$$

$$\frac{\partial^2 F_1}{\partial z_1{}^2}=\frac{\partial^2 E}{\partial z_1{}^2},\frac{\partial^2 F_1}{\partial z_2^2}=\frac{\partial^2 E}{\partial z_2^2},F_i=\left[1-\tilde{q}_{1,i}\left(N_{1,i}(z_1,z_2)\right)\right]\ (2\le i\le k),$$

$$\frac{\partial F_i}{\partial z_1}=-b_1\rho_{1,i}\left(1+\frac{\alpha_i}{\beta_{1,i,j}}\right),\frac{\partial F_i}{\partial z_2}=-\frac{\lambda_2 b_2 C_2'(1)}{\mu_{1,i}}\left(1+\frac{\alpha_i}{\beta_{1,i,j}}\right),\frac{\partial^2 F_i}{\partial z_1^2}=-b_{1,i}^{(2)}\left(\frac{\partial G_i}{\partial z_1}\right)^2-b_{1,i}^{(1)}\left(\frac{\partial^2 G_i}{\partial z_1^2}\right)$$

$$\frac{\partial^2 F_i}{\partial z_2^2}=-b_{1,i}^{(2)}\left(\frac{\partial G_i}{\partial z_2}\right)^2-b_{1,i}^{(1)}\left(\frac{\partial^2 G_i}{\partial z_2^2}\right),$$

$$\frac{\partial^2 G_i}{\partial z_1{}^2}=-\lambda_1 b_1 C_1''(1)+\alpha_i\left(\frac{\partial^2 m}{\partial z_1{}^2}\right),\ \frac{\partial^2 G_i}{\partial z_2{}^2}=-\lambda_2 b_2 C_2''(1)+\alpha_i\left(\frac{\partial^2 m}{\partial z_2{}^2}\right)$$

$$H=\prod_{s=1}^{i-1}\tilde{q}_{1,s}\left(N_{1,s}(z_1,z_2)\right),\ \frac{\partial H}{\partial z_1}=\sum_{s=1}^{i-1}\frac{1}{\mu_{1,s}}\lambda_1 b_1 C_1'(1)\left(1+\frac{\alpha_1}{\beta_{1,s,j}}\right),\ \frac{\partial H}{\partial z_2}=\sum_{s=1}^{i-1}\frac{1}{\mu_{1,s}}\lambda_2 b_2 C_2'(1)\left(1+\frac{\alpha_1}{\beta_{1,s,j}}\right)$$

$$\frac{\partial^2 H}{\partial z_1^2}=\sum_{s=1}^{i-1}b_{1,s}^{(2)}\left(-\lambda_1 b_1 C_1'(1)\left(1+\frac{\alpha_s}{\beta_{1,s,j}}\right)\right)^2+\sum_{s=1}^{i-1}b_{1,s}^{(1)}\left(-\lambda_1 b_1 C_1''(1)+\alpha_s(-y_{1,s,j}^{(2)}\left(\lambda_1 b_1 C_1'(1)\right)^2+y_{1,s,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right))\right)$$

$$\frac{\partial^2 H}{\partial z_2^2}=\sum_{s=1}^{i-1}b_{1,s}^{(2)}\left(-\lambda_2 b_2 C_2'(1)\left(1+\frac{\alpha_s}{\beta_{1,s,j}}\right)\right)^2+\sum_{s=1}^{i-1}b_{1,s}^{(1)}\left(-\lambda_2 b_2 C_2''(1)+\alpha_s(-y_{1,s,j}^{(2)}\left(\lambda_2 b_2 C_2'(1)\right)^2+y_{1,s,j}^{(1)}\left(\lambda_2 b_2 C_2''(1)\right))\right)$$

$$P=\left(\lambda_1 b_1+\lambda_2 b_2-\lambda_1 b_1 C_1(h(z_2))-\lambda_2 b_2 C_2(z_2)\right),\ \frac{\partial P}{\partial z_2}=-\lambda_1 b_1 C_1'(1)h'(1)-\lambda_2 b_2 C_2'(1)$$

$$R=\left[1-\tilde{q}_{2,1}\left(N_{1,1}(z_1,z_2)\right)\right],\frac{\partial R}{\partial z_1}=-\frac{\lambda_1 b_1 C_1'(1)}{\mu_{2,1}}\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right),\ \frac{\partial R}{\partial z_2}=-b_2\rho_{2,1}\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right)$$

$$\frac{\partial^2 R}{\partial z_1{}^2} = -b_{2,1}^{(2)}\left(\frac{\partial G_1}{\partial z_1}\right)^2 + b_{2,1}^{(1)}\left(\frac{\partial^2 G_1}{\partial z_1{}^2}\right), \frac{\partial^2 R}{\partial z_2{}^2} = -b_{2,1}^{(2)}\left(\frac{\partial G_1}{\partial z_2}\right)^2 + b_{2,1}^{(1)}\left(\frac{\partial^2 G_1}{\partial z_2{}^2}\right)$$

$$g = N_{2,1}(z_1, z_2),\ \frac{\partial g}{\partial z_1} = -\lambda_1 b_1 C_1'(1)\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right), \frac{\partial^2 g}{\partial z_1{}^2} = -\lambda_1 b_1 C_1''(1) + y_{2,1,j}^{(2)}\left(\lambda_1 b_1 C_1'(1)\right)^2 + y_{2,1,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right),$$

$$\frac{\partial g}{\partial z_2} = -\lambda_2 b_2 C_2'(1)\left(1+\frac{\alpha_1}{\beta_{2,1,j}}\right), \frac{\partial^2 g}{\partial z_2{}^2} = -\lambda_2 b_2 C_2''(1) + y_{2,1,j}^{(2)}\left(\lambda_2 b_2 C_2'(1)\right)^2 + y_{2,1,j}^{(1)}\left(\lambda_2 b_2 C_2''(1)\right)$$

$$J = N_{2,i}(z_1, z_2),\ \frac{\partial J}{\partial z_1} = -\lambda_1 b_1 C_1'(1)\left(1+\frac{\alpha_1}{\beta_{2,i,j}}\right), \frac{\partial J}{\partial z_2} = -\lambda_2 b_2 C_2'(1)\left(1+\frac{\alpha_1}{\beta_{2,i,j}}\right)(2 \le i \le k)$$

$$\frac{\partial^2 J}{\partial z_1{}^2} = -\lambda_1 b_1 C_1''(1) + \alpha_i\left(\frac{\partial^2 c}{\partial z_1{}^2}\right),\ \frac{\partial^2 J}{\partial z_2{}^2} = -\lambda_2 b_2 C_2''(1) + \alpha_i\left(\frac{\partial^2 c}{\partial z_2{}^2}\right)$$

$$s = \prod_{r=1}^{i-1}\tilde{q}_{2,r}\left(N_{2,r}(z_1, z_2)\right),\ \frac{\partial s}{\partial z_1} = \sum_{r=1}^{i-1}\frac{1}{\mu_{2,r}}\lambda_1 b_1 C_1'(1)\left(1+\frac{\alpha_1}{\beta_{2,r,j}}\right),\ \frac{\partial s}{\partial z_2} = \sum_{r=1}^{i-1}\frac{1}{\mu_{2,r}}\lambda_2 b_2 C_2'(1)\left(1+\frac{\alpha_1}{\beta_{2,r,j}}\right)$$

$$\frac{\partial^2 s}{\partial z_1^2} = \sum_{r=1}^{i-1} b_{1,r}^{(2)}\left(-\lambda_1 b_1 C_1'(1)\left(1+\frac{\alpha_r}{\beta_{2,r,j}}\right)\right)^2 + \sum_{r=1}^{i-1} b_{1,r}^{(1)}\left(-\lambda_1 b_1 C_1''(1) + \alpha_r(-y_{2,r,j}^{(2)}\left(\lambda_1 b_1 C_1'(1)\right)^2 + y_{2,r,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right))\right)$$

$$\frac{\partial^2 s}{\partial z_2^2} = \sum_{r=1}^{i-1} b_{1,r}^{(2)}\left(-\lambda_2 b_2 C_2'(1)\left(1+\frac{\alpha_r}{\beta_{1,r,j}}\right)\right)^2 + \sum_{r=1}^{i-1} b_{1,r}^{(1)}\left(-\lambda_2 b_2 C_2''(1) + \alpha_r(-y_{2,r,j}^{(2)}\left(\lambda_2 b_2 C_2'(1)\right)^2 + y_{2,r,j}^{(1)}\left(\lambda_2 b_2 C_2''(1)\right))\right)$$

$$a = \left[1-\tilde{g}_{1,1,j}(M(z_1, z_2))\right],\ \frac{\partial a}{\partial z_1} = \frac{-\lambda_1 b_1 C_1'(1)}{\beta_{1,1,j}},\ \frac{\partial a}{\partial z_2} = \frac{-\lambda_2 b_2 C_2'(1)}{\beta_{1,1,j}}$$

$$\frac{\partial^2 a}{\partial z_1{}^2} = -y_{1,1,j}^{(2)}\left(\lambda_1 b_1 C_1'(1)\right)^2 + y_{1,1,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right),\ \frac{\partial^2 a}{\partial z_2{}^2} = -y_{1,1,j}^{(2)}\left(\lambda_2 b_2 C_2'(1)\right)^2 + y_{1,1,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right)$$

$$b = M(z_1, z_2),\ \frac{\partial b}{\partial z_1} = -\lambda_1 b_1 C_1'(1), \frac{\partial b}{\partial z_2} = -\lambda_2 b_2 C_2'(1)\ \ \frac{\partial^2 b}{\partial z_1{}^2} = -\lambda_1 b_1 C_1''(1), \frac{\partial^2 b}{\partial z_2{}^2} = -\lambda_2 b_2 C_2''(1)$$

$$Y = \left[1-\tilde{g}_{2,1,j}(M(z_1, z_2))\right],\ \frac{\partial Y}{\partial z_1} = \frac{-\lambda_1 b_1 C_1'(1)}{\beta_{2,1,j}},\ \frac{\partial Y}{\partial z_2} = \frac{-\lambda_2 b_2 C_2'(1)}{\beta_{2,1,j}}$$

$$\frac{\partial^2 Y}{\partial z_1{}^2} = -y_{2,1,j}^{(2)}\left(\lambda_1 b_1 C_1'(1)\right)^2 + y_{2,1,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right),\ \frac{\partial^2 Y}{\partial z_2{}^2} = -y_{2,1,j}^{(2)}\left(\lambda_2 b_2 C_2'(1)\right)^2 + y_{2,1,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right)$$

$$c = \left[1-\tilde{g}_{2,i,j}(M(z_1, z_2))\right],\ \frac{\partial c}{\partial z_1} = \frac{-\lambda_1 b_1 C_1'(1)}{\beta_{2,i,j}},\ \frac{\partial c}{\partial z_2} = \frac{-\lambda_2 b_2 C_2'(1)}{\beta_{2,i,j}}$$

$$\frac{\partial^2 c}{\partial z_1{}^2} = -y_{2,i,j}^{(2)}\left(\lambda_1 b_1 C_1'(1)\right)^2 + y_{2,i,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right),\ \frac{\partial^2 c}{\partial z_2{}^2} = -y_{2,i,j}^{(2)}\left(\lambda_2 b_2 C_2'(1)\right)^2 + y_{2,i,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right)$$

$$T = \left[1-\tilde{q}_{2,i}\left(N_{2,i}(z_1, z_2)\right)\right], \frac{\partial T}{\partial z_1} = -\frac{\lambda_1 b_1 C_1'(1)}{\mu_{2,i}}\left(1+\frac{\alpha_1}{\beta_{2,i,j}}\right),\ \frac{\partial T}{\partial z_2} = -b_2\rho_{2,i}\left(1+\frac{\alpha_1}{\beta_{2,i,j}}\right)$$

$$\frac{\partial^2 T}{\partial z_1{}^2} = -b_{2,1}^{(2)}\left(\frac{\partial g}{\partial z_1}\right)^2 + b_{2,1}^{(1)}\left(\frac{\partial^2 g}{\partial z_1{}^2}\right), \frac{\partial^2 T}{\partial z_2{}^2} = -b_{2,1}^{(2)}\left(\frac{\partial g}{\partial z_2}\right)^2 + b_{2,1}^{(1)}\left(\frac{\partial^2 g}{\partial z_2{}^2}\right)$$

$$m=\left[1-\tilde{g}_{1,i,j}(M(z_1,z_2))\right],\ \frac{\partial m}{\partial z_1}=\frac{-\lambda_1 b_1 C_1'(1)}{\beta_{1,i,j}},\ \frac{\partial m}{\partial z_2}=\frac{-\lambda_2 b_2 C_2'(1)}{\beta_{1,i,j}}$$

$$\frac{\partial^2 m}{\partial z_1{}^2}=-y_{1,i,j}^{(2)}\left(\lambda_1 b_1 C_1'(1)\right)^2+y_{1,i,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right),\ \frac{\partial^2 m}{\partial z_2{}^2}=-y_{1,i,j}^{(2)}\left(\lambda_2 b_2 C_2'(1)\right)^2+y_{1,i,j}^{(1)}\left(\lambda_1 b_1 C_1''(1)\right)$$

$$f=Ts,\ f'=T's+Ts',\ f''=T''s+2T's'+Ts''$$

Theorem 5.6(a): The expected number of priority customers in the system is given by

$$L_P=P_0(1)\left\{T_1+T_2+T_3+T_4+T_5+T_6+T_7\right\} \tag{5.55}$$

where

$$T_1=-\left[\frac{E'G_1'\left(C''F_1'+C'F_1''\right)-C'F_1'\left(E''G_1'+E'G_1''\right)}{2\left(E'G'\right)^2}\right]$$

$$T_2=-\left[\frac{3\left(E'G_i'\right)\left(2C'F_i'+C''F_i'+C'F_i''\right)-2\left(E''G_i'+E'G_i''\right)\left(C'F'\right)}{8\left(E'G_i'\right)^2}\right]$$

$$T_3=\left(\frac{P'}{D'}\right)\left[\left(\frac{R''g'-R'g''}{2g'^2}\right)-\left(\frac{f''J'-f'J''}{2J'^2}\right)\right],\ T_4=-\alpha_1\left[\frac{\left(E'G_i'b'\right)\left(C'F_i'a'\right)'-\left(E'G_i'b'\right)'\left(C'F_i'a'\right)}{2\left(E'G_i'b'\right)^2}\right]$$

$$T_5=-\alpha_i\left[\frac{2\left(E'G_i'b'\right)\left(C'F_i'H'\right)'-2\left(E'G_i'b'\right)'\left(2C'F_i'H'+C''F_i'H+C'F_i''H\right)}{4\left(E'G_i'b'\right)^2}\right]$$

$$T_6=\alpha_1\left[\left(\frac{P'}{D'}\right)\left(\frac{\left(g'b'\right)\left(R'Y'\right)'-\left(R'Y'\right)\left(g'b'\right)'}{2\left(g'b'\right)^2}\right)\right],\ T_7=\alpha_i\left[\left(\frac{P'}{D'}\right)\left(\frac{\left(J'b'\right)\left(f'c'\right)'-\left(f'c'\right)\left(J'b'\right)'}{2\left(J'b'\right)^2}\right)\right]$$

where dash ($'$) denotes the partial differentiation of the terms w.r.t 'z_1'

Proof: The expected number of customers in the priority queue is obtained by using

$$L_P=\lim_{z_1,z_2\to 1}\frac{\partial L(z_1,z_2)}{\partial z_1}$$

where, $L(z_1,z_2)$ is the generating function of the queue size distribution.

Theorem 5.6(b): The expected number of customers in the ordinary queue is given by

$$L_{NP}=P_0(1)\left\{\frac{1}{\gamma}.\frac{A'}{D'}+T_1+V_1+V_2+V_3+V_4+V_5+V_6+V_7+V_8\right\} \tag{5.56}$$

where,

$$V_1=\left(\frac{\left(D'E'G_1'\right)\left(A'B'F_1'\right)'-\left(A'B'F_1'\right)\left(D'E'G_1'\right)'}{2\left(D'E'G_1'\right)^2}\right)+\left(\frac{\left(DEG_i\right)''\left(ABF_iH\right)^{iv}-\left(ABF_iH\right)'''\left(DEG_i\right)^{iv}}{4\left(6D'E'G_i'\right)^2}\right)$$

$$V_2 = -\left[\frac{3(E'G')(2C'F'H + C''F'H + C'F''H) - 2(E''G' + E'G'')(C'F'H)}{8(E'G')^2}\right]$$

$$V_3 = \left(\frac{(D'g')(P''R' + P'R'') - (P'R')(D''g' + D'g'')}{2(D'g')^2}\right) + \left(\frac{(DJ)''(Pf)''' - (Pf)''(DJ)'''}{6(D'J')^2}\right)$$

$$V_4 = 288\alpha_1\left(\frac{(D'E'G_1'b')(A'B'F'a')' - (A'B'F'a')(D'E'G_1'b')'}{(D'E'G_1'b')^2}\right)$$

$$V_5 = -\alpha_1\left(\frac{(E'G_1'b')(C'F'a')' - (C'F'a')(E'G_1'b')'}{2(E'G_1'b')^2}\right)$$

$$V_7 = \alpha_i\left(\frac{(E'G_i'b')\left(2C'F_i'm'H + C''F_i'm'H + C'F_i''m'H\right) - (C'F_i'm'H)(E''G_i'b' + E'G_i''b' + E'G_i'b'')}{2(E'G_i'b')^2}\right)$$

$$V_8 = \alpha_1\left(\frac{(D'g'b')(P'R'Y')' - (P'R'Y')(D'g'b')'}{(D'g'b')^2}\right) + \alpha_i\left(\frac{(D'J'b')(P'f'c')' - (P'f'c')(D'J'b')'}{2(D'J'b')^2}\right)$$

where dash ($'$) denotes the partial differentiation of the terms w.r.t 'z_2'

Proof: The expected number of customers in the priority queue is obtained by using

$$L_{NP} = \lim_{z_1,z_2\to 1}\frac{\partial L(z_1,z_2)}{\partial z_2}$$

where, $L(z_1,z_2)$ is the generating function of the queue size distribution.

5.5 NUMERICAL ILLUSTRATION

The present section deals with the numerical simulation of our model to examine the effects of various parameters on the performance indices. 'MATLAB' software has been used to develop the code for computer program to study the sensitivity analysis under consideration. For the computational purposes, the batch size, retrial time as well as repair time have been taken to be exponentially distributed. Further, numerical results have been obtained by assuming two phase service system i.e.k=2 and two phase repair system i.e. d=2. The different distributions namely exponential, Erlangian-5 and gamma have been considered for the service time. The numerical results have been displayed by means of tables and figures by assuming the values of default parameters as λ_1=1.5, λ_2=0.5, γ=0.1, b_1=b_2=0.5, β_{111} = β_{112} =β_{121}= β_{122} =β_{11}=1, β_{211} = β_{212} =β_{221}= β_{222} =β_{21}=0.5, μ_{11}= μ_{12}= μ_1=4, μ_{21}= μ_{22}= μ_2=3, α_1=0.02, α_2=0.1, E[X_1]= E[X_2]= E[X]=2.

Tables 5.1 and 5.2 display the effect of service rates μ_2 and μ_1 on the long run probabilities of the server at different states. It is noted from the tables 5.1 and 5.2 that the long run probabilities for the idle state P(I) decreases with the increase in breakdown rate α_1 (α_2) for a fixed value of service rate. The P(B_1) and P(B_2) show almost constant values with small variations in α_1 while it increases with the increase in α_2. The long run probabilities P(R_1) and P(R_2) increase with the increase in α_1 (α_2) for fixed service rates for both types of customers. With an increase in the service rate the long run probabilities, P(I), P(B_1) and P(R_1) increase whereas P(B_2) and P(R_2) decrease for fixed breakdown rates.

Tables 5.3 and 5.4 demonstrate the effect of service rates on the long run probabilities of the system states for the varying values of joining probabilities b_1 and b_2. With an increase in joining probability b_1, the long run probabilities P(I), P(B_2) and P(R_2) decrease whereas P(B_1) and P(R_1) increase. On the other hand, P(I), P(B_1) and P(R_1) show a decreasing trend whereas P(B_2) and P(R_2) exhibit increasing behavior for the increasing values of b_2.The effects of arrival rates λ_1 and λ_2 with varying values of the repair rates (β_{11}, β_{21}) on the long run probabilities are shown in tables 5.5 and 5.6. With an increase in arrival rate λ_1 for priority customers, the long run probabilities P(B_1) and P(R_1) increase whereas P(I), P(B_2) and P(R_2) decrease which is quite obvious. This is due to the fact that an increase in arrival rate for priority customers, has an adverse effect on the idleness of the server and as such increases the probability of the server being in a busy state.

Similarly, with an increase in arrival rate λ_2, P(B_2) and P(R_2) increase whereas P(I), P(B_1) and P(R_1) show a decreasing trend for the fixed values of both λ_1 and λ_2. It is observed that, P(R_1) decreases with an increase in breakdown rate β_{11} for constant value of β_{21}, whereas other long run probabilities increase. It is also seen from the tables 5.5 and 5.6 that P(R_2) decreases with an increase in β_{21} while other long run indices increase with an increase in β_{21}. Tables 5.7-5.8 depict the effect of breakdown rate (α_1) and repair rate (β_{11}) with arrival rate (λ_1) and service rate (μ_1) on the availability (A_v) and failure frequency (F_f) of the server. We note that as λ_1 increases, the availability (A_v) decreases whereas failure frequency (F_f) increases. Table 5.7 demonstrates that an increase in breakdown rate (α_1) reduces the availability of the server whereas a slight increment is observed in the values of F_f with decrease in α_1. The variation in repair rate (β_{11}) on the reliability indices is shown in table 5.8. A slight increment is observed in the availability as well as failure frequency of the server with an increase in β_1 for fixed values of other parameters.

Table 5.1: Effect of μ_2 and (α_1,α_2) on the long run probabilities of the server states

μ_2	(α_1,α_2)	P(I)	P(B_1)	P(B_2)	P(R_1)	P(R_2)
	(0.02,0.1)	0.2130	0.3522	0.3993	0.0101	0.0256
	(0.03,0.1)	0.2057	0.3522	0.3993	0.0126	0.0303
3	(0.04,0.1)	0.1984	0.3522	0.3993	0.0151	0.0351
	(0.02,0.2)	0.2086	0.3449	0.3911	0.0148	0.0407
	(0.02,0.3)	0.2044	0.3380	0.3832	0.0193	0.0552
	(0.02,0.1)	0.2982	0.3815	0.2919	0.0109	0.0175
	(0.03,0.1)	0.2916	0.3815	0.2919	0.0136	0.0214
4	(0.04,0.1)	0.2849	0.3815	0.2919	0.0163	0.0253
	(0.02,0.2)	0.2937	0.3758	0.2876	0.0161	0.0268
	(0.02,0.3)	0.2894	0.3702	0.2833	0.0212	0.0359
	(0.02,0.1)	0.3522	0.3968	0.2267	0.0113	0.0130
	(0.03,0.1)	0.3462	0.3968	0.2267	0.0142	0.0162
5	(0.04,0.1)	0.3401	0.3968	0.2267	0.0170	0.0194
	(0.02,0.2)	0.3480	0.3920	0.2240	0.0168	0.0192
	(0.02,0.3)	0.3439	0.3874	0.2213	0.0221	0.0253

Table 5.2: Effect of μ_1 and (α_1,α_2) on the long run probabilities of the server states

μ_1	(α_1,α_2)	P(I)	P(B_1)	P(B_2)	P(R_1)	P(R_2)
	(0.02,0.1)	0.2130	0.3522	0.3993	0.0101	0.0256
	(0.03,0.1)	0.2057	0.3522	0.3993	0.0126	0.0303
5	(0.04,0.1)	0.1984	0.3522	0.3993	0.0151	0.0351
	(0.02,0.2)	0.2086	0.3449	0.3911	0.0148	0.0407
	(0.02,0.3)	0.2044	0.3380	0.3832	0.0193	0.0552
	(0.02,0.1)	0.2643	0.2888	0.4126	0.0079	0.0264
	(0.03,0.1)	0.2571	0.2888	0.4126	0.0101	0.0314
6	(0.04,0.1)	0.2643	0.2888	0.4126	0.0079	0.0264
	(0.02,0.2)	0.2590	0.2831	0.4044	0.0113	0.0421
	(0.02,0.3)	0.2540	0.2776	0.3966	0.0146	0.0571
	(0.02,0.1)	0.3019	0.2436	0.4211	0.0065	0.0269
	(0.03,0.1)	0.2949	0.2436	0.4211	0.0084	0.0320
7	(0.04,0.1)	0.2880	0.2436	0.4211	0.0103	0.0371
	(0.02,0.2)	0.2961	0.2389	0.4130	0.0090	0.0429
	(0.02,0.3)	0.2905	0.2344	0.4052	0.0115	0.0584

Table 5.3: Effect of μ_1 and (b_1, b_2) on the long run probabilities of the server states

μ_1	(b_1, b_2)	P(I)	P(B_1)	P(B_2)	P(R_1)	P(R_2)
5	(0.4,0.5)	0.2700	0.2878	0.4079	0.0082	0.0261
	(0.5,0.5)	0.2130	0.3522	0.3993	0.0101	0.0256
	(0.6,0.5)	0.1583	0.4138	0.3910	0.0118	0.0250
	(0.5,0.4)	0.2724	0.3650	0.3310	0.0104	0.0212
	(0.5,0.6)	0.1576	0.3402	0.4629	0.0097	0.0296
6	(0.4,0.5)	0.3132	0.2346	0.4189	0.0065	0.0268
	(0.5,0.5)	0.2643	0.2888	0.4126	0.0079	0.0264
	(0.6,0.5)	0.2168	0.3414	0.4064	0.0094	0.0260
	(0.5,0.4)	0.3276	0.2997	0.3425	0.0082	0.0219
	(0.5,0.6)	0.2053	0.2787	0.4778	0.0077	0.0306
7	(0.4,0.5)	0.3445	0.1971	0.4259	0.0053	0.0273
	(0.5,0.5)	0.3019	0.2436	0.4211	0.0065	0.0269
	(0.6,0.5)	0.2603	0.2890	0.4163	0.0077	0.0266
	(0.5,0.4)	0.3681	0.2530	0.3498	0.0067	0.0224
	(0.5,0.6)	0.2404	0.2349	0.4872	0.0063	0.0312

Table 5.4: Effect of μ_2 and (b_1, b_2) on the long run probabilities of the server states

μ_2	(b_1, b_2)	P(I)	P(B_1)	P(B_2)	P(R_1)	P(R_2)
3	(0.4,0.5)	0.2700	0.2878	0.4079	0.0082	0.0261
	(0.5,0.5)	0.2130	0.3522	0.3993	0.0101	0.0256
	(0.6,0.5)	0.1583	0.4138	0.3910	0.0118	0.0250
	(0.5,0.4)	0.2724	0.3650	0.3310	0.0104	0.0212
	(0.5,0.6)	0.1576	0.3402	0.4629	0.0097	0.0296
4	(0.4,0.5)	0.3620	0.3123	0.2988	0.0089	0.0179
	(0.5,0.5)	0.2982	0.3815	0.2919	0.0109	0.0175
	(0.6,0.5)	0.2372	0.4475	0.2854	0.0128	0.0171
	(0.5,0.4)	0.3461	0.3898	0.2387	0.0111	0.0143
	(0.5,0.6)	0.2523	0.3735	0.3430	0.0107	0.0206
5	(0.4,0.5)	0.4200	0.3251	0.2322	0.0093	0.0133
	(0.5,0.5)	0.3522	0.3968	0.2267	0.0113	0.0130
	(0.6,0.5)	0.2876	0.4650	0.2214	0.0133	0.0127
	(0.5,0.4)	0.3915	0.4025	0.1840	0.0115	0.0105
	(0.5,0.6)	0.3141	0.3912	0.2682	0.0112	0.0153

Table 5.5: Effect of λ_1 and (β_{11}, β_{21}) on the long run probabilities of the server states

λ_1	(β_{11}, β_{21})	P(I)	P(B_1)	P(B_2)	P(R_1)	P(R_2)
3.0	(1.0,0.5)	0.1961	0.5795	0.1947	0.0185	0.0111
	(1.5,0.5)	0.1992	0.5818	0.1955	0.0124	0.0112
	(2.0,0.5)	0.2007	0.5829	0.1958	0.0093	0.0112
	(1.0,1.0)	0.1994	0.5811	0.1953	0.0186	0.0056
	(1.0,1.5)	0.2006	0.5817	0.1954	0.0186	0.0037
3.5	(1.0,0.5)	0.1317	0.6497	0.1871	0.0208	0.0107
	(1.5,0.5)	0.1348	0.6526	0.1879	0.0139	0.0107
	(2.0,0.5)	0.1364	0.6540	0.1883	0.0105	0.0108
	(1.0,1.0)	0.1347	0.6515	0.1876	0.0208	0.0054
	(1.0,1.5)	0.1357	0.6521	0.1878	0.0209	0.0036
4.0	(1.0,0.5)	0.0720	0.7147	0.1801	0.0229	0.0103
	(1.5,0.5)	0.0753	0.7181	0.1810	0.0153	0.0103
	(2.0,0.5)	0.0769	0.7198	0.1814	0.0115	0.0104
	(1.0,1.0)	0.0748	0.7165	0.1806	0.0229	0.0052
	(1.0,1.5)	0.0757	0.7172	0.1807	0.0229	0.0034

Table 5.6: Effect of λ_2 and (β_{11}, β_{21}) on the long run probabilities of the server states

λ_2	(β_{11},β_{21})	P(I)	P(B_1)	P(B_2)	P(R_1)	P(R_2)
1.0	(1.0,0.5)	0.2769	0.5978	0.1004	0.0191	0.0057
	(1.5,0.5)	0.2804	0.6002	0.1008	0.0128	0.0058
	(2.0,0.5)	0.2822	0.6014	0.1010	0.0096	0.0058
	(1.0,1.0)	0.2787	0.5987	0.1006	0.0192	0.0029
	(1.0,1.5)	0.2793	0.5989	0.1006	0.0192	0.0019
2.0	(1.0,0.5)	0.1961	0.5795	0.1947	0.0185	0.0111
	(1.5,0.5)	0.1992	0.5818	0.1955	0.0124	0.0112
	(2.0,0.5)	0.2007	0.5829	0.1958	0.0093	0.0112
	(1.0,1.0)	0.1994	0.5811	0.1953	0.0186	0.0056
	(1.0,1.5)	0.2006	0.5817	0.1954	0.0186	0.0037
3.0	(1.0,0.5)	0.1201	0.5623	0.2834	0.0180	0.0162
	(1.5,0.5)	0.1228	0.5644	0.2845	0.0120	0.0163
	(2.0,0.5)	0.1242	0.5655	0.2850	0.0090	0.0163
	(1.0,1.0)	0.1247	0.5646	0.2846	0.0181	0.0081
	(1.0,1.5)	0.1262	0.5654	0.2849	0.0181	0.0054

Table 5.7: Effect of α_1 on Availability and Failure frequency

α_1	λ_1	A_v			F_f		
		μ_1=4	μ_1=5	μ_1=6	μ_1=4	μ_1=5	μ_1=6
0.02	**1.5**	0.9851	0.9868	0.9878	0.1159	0.0993	0.0885
	2.0	0.9831	0.9853	0.9866	0.1379	0.1171	0.1032
	2.5	0.9813	0.9838	0.9854	0.1587	0.1342	0.1176
	3.0	0.9795	0.9824	0.9843	0.1784	0.1507	0.1316
0.04	**1.5**	0.9789	0.9811	0.9825	0.1232	0.1055	0.0940
	2.0	0.9761	0.9788	0.9806	0.1465	0.1244	0.1097
	2.5	0.9734	0.9767	0.9788	0.1686	0.1425	0.1249
	3.0	0.9709	0.9745	0.9770	0.1896	0.1601	0.1398
0.06	**1.5**	0.9728	0.9755	0.9772	0.1304	0.1117	0.0996
	2.0	0.9691	0.9724	0.9747	0.1552	0.1317	0.1161
	2.5	0.9656	0.9695	0.9722	0.1785	0.1509	0.1323
	3.0	0.9623	0.9667	0.9698	0.2007	0.1695	0.1480

Table 5.8: Effect of β_{11} on Availability and Failure frequency

β_{11}	λ_1	A_v			F_f		
		μ_1=4	μ_1=5	μ_1=6	μ_1=4	μ_1=5	μ_1=6
1.0	**1.5**	0.9851	0.9868	0.9878	0.1159	0.0993	0.0885
	2.0	0.9831	0.9853	0.9866	0.1379	0.1171	0.1032
	2.5	0.9813	0.9838	0.9854	0.1587	0.1342	0.1176
	3.0	0.9795	0.9824	0.9843	0.1784	0.1507	0.1316
1.5	**1.5**	0.9875	0.9885	0.9892	0.1161	0.0994	0.0886
	2.0	0.9862	0.9876	0.9884	0.1382	0.1172	0.1033
	2.5	0.9851	0.9866	0.9877	0.1590	0.1344	0.1177
	3.0	0.9840	0.9857	0.9869	0.1789	0.1509	0.1317
2.0	**1.5**	0.9887	0.9894	0.9899	0.1162	0.0995	0.0886
	2.0	0.9878	0.9887	0.9893	0.1383	0.1173	0.1034
	2.5	0.9870	0.9881	0.9888	0.1592	0.1344	0.1178
	3.0	0.9862	0.9874	0.9882	0.1791	0.1510	0.1318

Figures 5.2- 5.5 are plotted to display the effect of various varying parameters on the queue length of the customers for both priority and ordinary customers. The service time distribution has been considered as Erlangian-5, exponential and gamma respectively for all the figures. Figures 5.2 and 5.3 exhibit the effect of arrival rate for both types of customers on their respective queue lengths for varying values of mean batch sizes. It is quite clear from the graphs plotted in figs 5.2(5.3) that the queue length for priority (ordinary) customers increases with the increase in the arrival rate and batch size, which is quite obvious. The average queue length obtained for the priority customers is higher than that of ordinary customers for all the three service time distributions under

consideration. However, Erlangian-5 distribution reveals the maximum queue length for both types of customers.

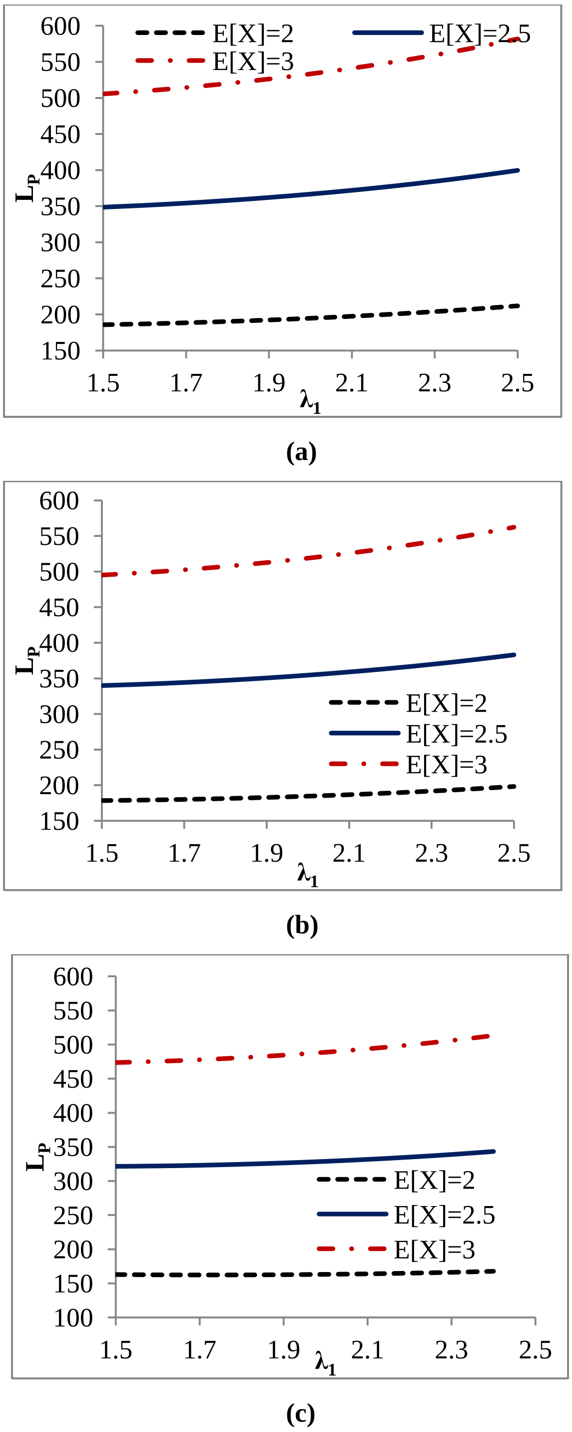


Fig. 5.2: Effect of λ_1 and E[X] on L_P for (a) $M^x/E_5/1$ (b) $M^x/M/1$ (c) $M^x/\gamma/1$

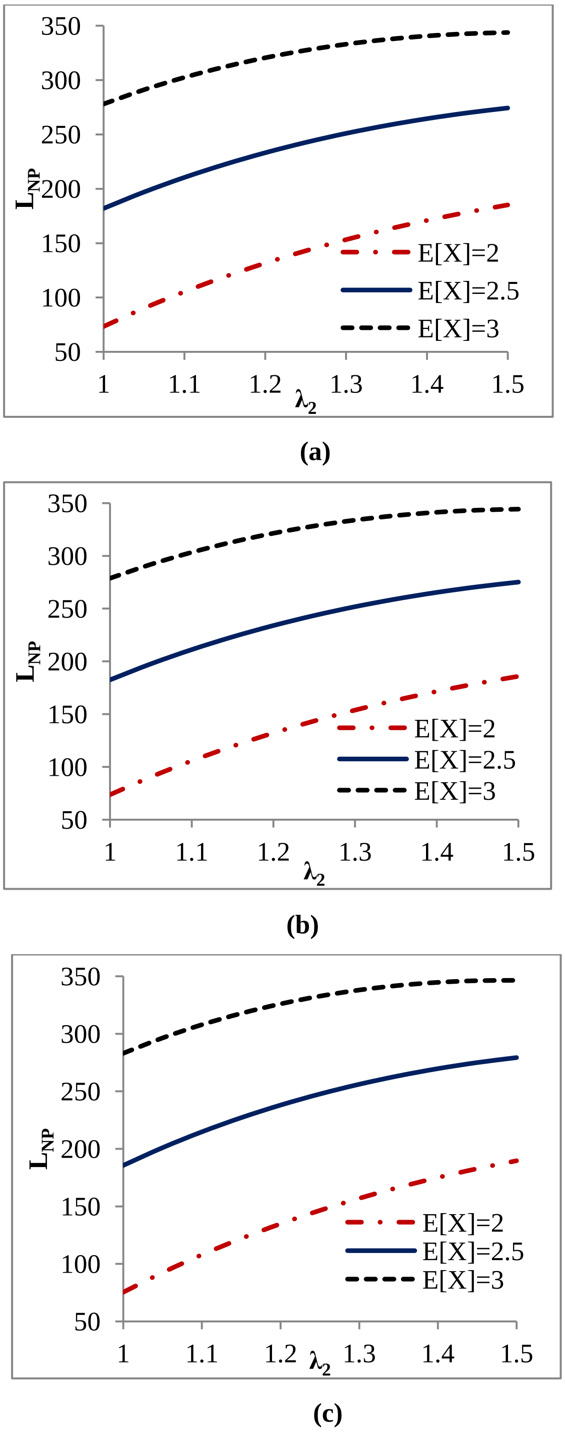
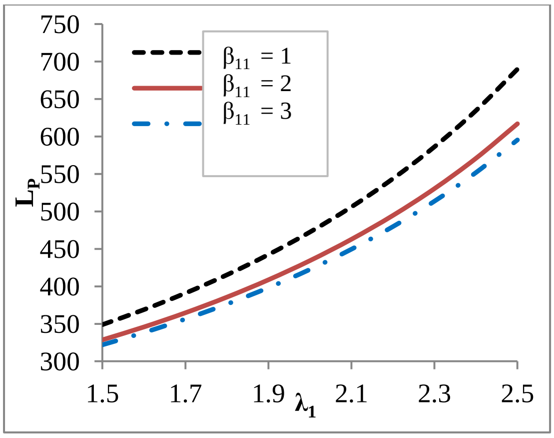
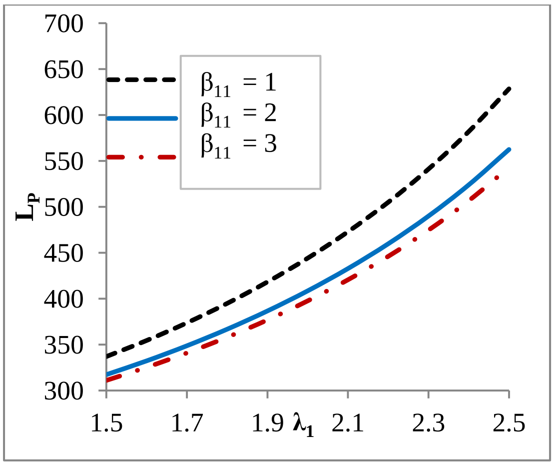


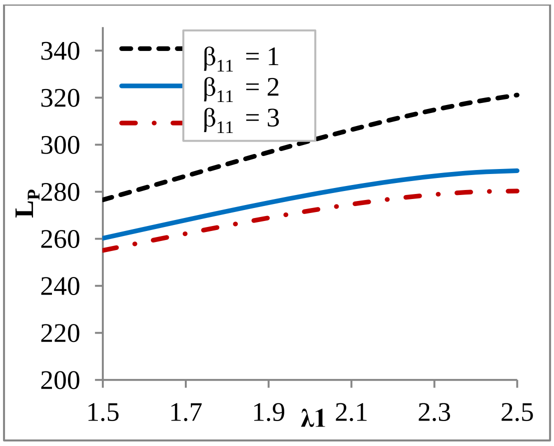
Fig. 5.3: Effect of λ_2 and E[X] on L_{NP} (a) $M^x/E_5/1$ (b) $M^x/M/1$ (c) $M^x/\gamma/1$



(a)

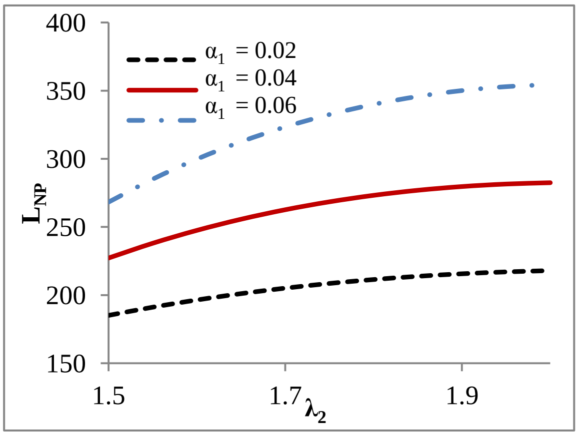


(b)

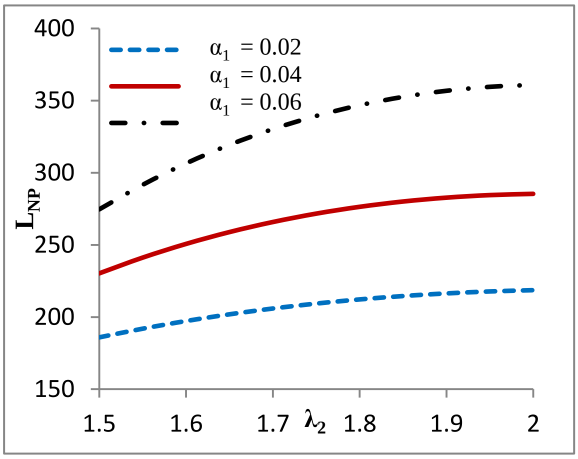


(c)

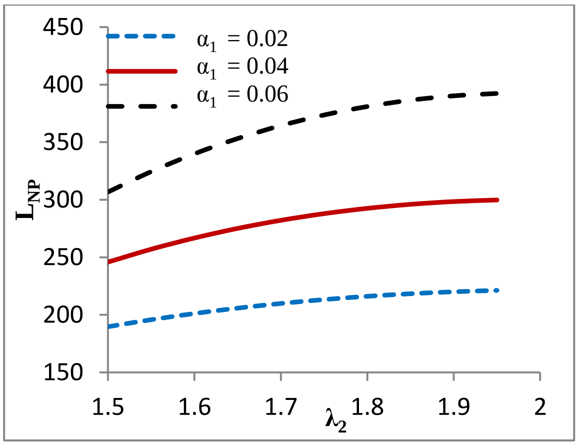
Fig. 5.4: Effect of λ_1 and β_{11} on L_P for (a) $M^x/E_5/1$ (b) $M^x/M/1$ (c) $M^x/\gamma/1$



(a)



(b)



(c)

Fig. 5.5: Effect of λ_2 and α_1 on L_{NP} for (a) $M^x/E_5/1$ (b) $M^x/M/1$ (c) $M^x/\gamma/1$

Fig. 5.4 exhibits the effect of the repair rate as well as arrival rate on the average queue length LP of the priority customers. As the arrival rate increases from 1.5 units to

2.5 units, the LP also builds up which can be realized in realistic queueing situations. Moreover, it is clear from the figure that as the repair rate increases for fixed (λ_1), the number of customers decreases; this is due to the fact that an increase in repair rate, speeds up the servicing of the customers thus reducing the queue length. The figure 5.5 demonstrates the effect of arrival rate (λ_2) along with breakdown rate (α_1) of the priority customers on L_{NP}. This particular case has been examined because the breakdowns of the server for priority customers also affects the queue length (L_{NP}) of ordinary customers. The number of customers increases with an increase in the breakdown rate α_1 for all the three distributions under consideration as depicted by figs 5.5 (a-c). It is quite obvious that an increase in the breakdown rate for priority customers keeps the priority customers in the waiting queue for a long period and thereby increases the queue length of ordinary customers which cannot be served prior to priority customers.

5.6 DISCUSSION

The bulk arrival non-Markovian system with two types of customers have investigated by assuming different distributions for the service time. The queueing analysis of the retrial model with priority doneto establish explicit expressions for the queue length and other performance measures. Overall, we can conclude that-

- The queue lengths for both types of customers increase with an increase in the arrival rate and mean batch size.
- L_{NP} increases with an increase in breakdown rate α_1, as an increase in the breakdown rate of priority customers indirectly increases the queue length of ordinary customers by delaying the service. Hence, a control over repair of the server can help in reducing the queue length.
- The long run probabilities of server states are affected to a great extent with the service rate and arrival rates.

CHAPTER 6

RETRIAL G-QUEUE WITH PREEMPTIVE RESUME

6.1 INTRODUCTION

The attack of malware or virus on the computer system is a very common phenomenon. Almost every computer system is affected by the attack of malwares/cookies/virus arrived independently either by using internet services or unprotected web sites or affected external drives. These virus/malwares in queueing theory are termed as negative customers, which not only destroy the normal functioning of the system but also make the system under repair. Mathematical queues with the arrival of negative customers are known as G-queues. Performance modeling of such queues holds a significant place in the queueing theory. G-queues or queues with negative customers were first analyzed by Gelenbe (1989). The arrival of negative customers usually affects the normal working of the system either by stopping the service process or by lowering down the working process. The concept of negative customers finds various applications in communication systems, computer protocols, neural networks modeling, etc. where the system get destroyed or failed with the arrival of unwanted customers like virus in the computers.

The concept of negative customers can also be considered as a step towards study of new control policy as negative customers usually reduce the congestion in the system by removing the customers. The remarkable contributions of Gelenbe (1991, 2000) opened ways for the future research on G-queues. Harrison (1993, 1996) investigated queues with negative customers using a new technique and paved a new dimension to the modeling of unreliable queues with negative customers. Shin (2007) analyzed multiserver G-queues with disasters and reattempts. The recent articles on G-queues can be found in the bibliography on negative customers by Do (2011a). Liu *et al.* (2009) and Dimitriou (2013a, b) studied negative arrival retrial queues with pre emptive resume with breakdowns and gave interesting mathematical results for the non-markovian queues.

In the present chapter, we develop the semi Markov model for the unreliable retrial G-queue wherein the customers can renege due to impatience. The server renders service in multi-phases and may breakdown while providing service to the customers.

The concept of bulk arrival and impatience customers seeking for more optional services has been incorporated to study the effect of negative customers on the performance indices of the concerned retrial queueing system. The rest of the chapter has been organized in the following manner. Section 6.2 presents the mathematical description and various assumptions underlying the model. The generating functions corresponding to various server states have been obtained in section 6.3. By setting appropriate parameters, some special cases have been deduced in section 6.4. Section 6.5 describes various performance measures for the model. Finally, the conclusions are drawn in section 6.6.

6.2 MODEL DESCRIPTION

The semi-Markov model is developed to analyze a bulk arrival retrial queue with negative customers and reneging. The server allows the service of every customer in essential as well as in optional phases. The basic assumptions describing the model are as follows:

Assumptions

We consider $M^x/G/1$ retrial queue wherein the service is interrupted due to the arrival of negative customers. The server is affected by the negative customer only when it is in idle state or busy state. The assumptions underlying the model are as follows:

- **Arrival Process** The system allows the arrival of two types of customers; positive customers and negative customers according to the Poisson process with arrival rates λ^+ and λ^- respectively s.t. $\lambda = \lambda^+ + \lambda^-$.
- **Retrial Process:** If an arriving customer finds the server busy or in non-working condition then he joins the retrial orbit so as to wait in order to retry for the service. The customers retry with constant retrial rate γ so as to get served. Moreover, there is a competition between retrying customers and new arrivals to obtain the service. In case a retrial customer fails to get service while competing with the new primary customer then either it leaves the system forever with probability (1-r) (i.e. reneges) or returns back to the orbit with probability 'r' to get the service.
- **Service Process:** The incoming customers are served in two phases with general distributed service time; first phase corresponds to the compulsory phase known as first phase of service (FPS) with mean service rate μ_0. After completing FPS, the interested customers may go for second phase of service (SPS) with probability p_o or may enter the orbit with probability 1- p_o which is further completed in 'k'

compulsory phases with service rate μ_i (1 ≤i≤ k), to make retrials until successful services are achieved.

- **Pre-emptive resume:** The server follows the policy of pre-emptive resume of service; according to which the customer under FES may be forced to move to the orbit with probability *p* so as to make retrials after some time to new incoming positive customer or continue the service with probability (1-p) of the customer already in service.
- **N-policy***:* In case no more customers are available in the system then server moves to the dormant state and comes back to the working state only after the accumulation of N customers in the system. This can be considered as N-policy for the optimal control of service rendered by the server.
- **Repair Process:** The server under consideration is unreliable and may fail due to the arrival of negative customers in the system. The server is affected by the arrival of negative customer when it is in idle state or busy state. The system remains unaffected if it is in dormant or repair state. The server breakdown may occur in Poisson fashion with failure rate α_i (0≤ i≤ k, i=0 for FES).The repair process is completed in *d*-compulsory phases with repair rate β_{ij} (0 ≤i≤ k), (1≤ j≤ *d*) to bring the system back to its original state.

To highlight the novelty of the model, we cite its application in computer system which is prone to the attack of malware or virus. The negative customers (viruses etc.) and positive customers (files to be worked on) are available in the system independent of each other in batches. The work files (positive customers) can be served in various phases of service; there may be a series of work procedure which have to be performed depending on the requirement of the file currently under work. The arrival of negative customer destroys the normal functioning of the system and makes the system under repair. The repair process is usually completed in a series of compulsory phases starting from the booting or restarting of the system, to launch the start up of the computer (in case it is damaged) and finally moving through other mandatory steps to the scanning of the system so as to stop the attack of further malwares in the system.

6.3 THE ANALYSIS

In this section, we present mathematical formulation and analysis of the concerned retrial model based on certain notations and assumptions as stated in section 6.2. The system state probabilities, construction of governing Chapman-Kolmogorov equations and methodology based on generating function and supplementary variables technique has been used to solve the model.

State Probabilities

The model under consideration is non-markovian as service process and repair process are assumed to follow general distribution. We define the state of the server at any time t as follows:

$$Y(t)=\begin{cases} V, & \text{server is in dormant period at time t} \\ 0, & \text{server is busy in providing FPS at time t} \\ 1, & \text{server is busy in providing } i^{th}\ (1 \le i \le k) \text{ SPS at time t} \\ 2, & \text{server is in idle state at time t} \\ 3, & \text{server is brokendown and under repair at time t} \end{cases}$$

Let N(t) denotes the number of customers in the system at any time t and $S_1(t) \in \{0,1,2,...,k\}$, $S_2(t) \in \{1,2,...,d\}$ represent the phase of service and repair process respectively, at any time t.

In the steady state, the joint distributions of the system states and corresponding probabilities are defined as-

$$V_n = \lim_{t\to\infty} \Pr\{Y(t)=V, N(t)=n\}, 1 \le n \le N-1$$

$$P_{0,n}(x) = \lim_{t\to\infty} \Pr\{Y(t)=0, x \le \varpi(t) \le x+dx, N(t)=n, S_1(t)=0\}, n \ge 0$$

$$P_{i,n}(x) = \lim_{t\to\infty} \Pr\{Y(t)=1, x \le \varpi(t) \le x+dx, N(t)=n, S_1(t)=i\}, n \ge 0, (1 \le i \le k)$$

$$D_n = \lim_{t\to\infty} \Pr\{Y(t)=2, N(t)=n\}, n \ge 1$$

$$R_{i,j,n}(x,y) = \lim_{t\to\infty} \Pr\{Y(t)=3, \varpi(t)=x, y \le \sigma(t) \le y+dy, S_1(t)=0=i, S_2(t)=0=j, N(t)=n\},$$
$$n \ge 0, (0 \le i \le k), (1 \le j \le d)$$

Governing Equations

Now, we construct Chapman-Kolmogorov equations using supplementary variables technique which are as follows:

$$\lambda^+ V_n = \lambda^+ V_{n-1}, \ \ 1 \le n \le N \tag{6.1}$$

$$\lambda^+ V_0 = \int_0^{+\infty} P_{k,0}(x)\mu_k(x)dx + \sum_{i=0}^{k} \int_0^{+\infty} R_{i,d,0}(x,y)\beta_{i,d}(y)dy \tag{6.2}$$

$$\frac{dP_{0,n}(x)}{dx} = -\left(\lambda + \mu_0(x)\right)P_{0,n}(x) + \sum_{m=1}^{n} c_m \lambda^+ \bar{p} P_{0,n-m}(x)(1-\delta_{n,0}),\ n \ge 0 \tag{6.3}$$

$$\frac{dP_{i,n}(x)}{dx}=-\left(\lambda+\mu_i(x)\right)P_{i,n}(x)+\sum_{m=1}^{n}c_m\lambda^+P_{i,n-m}(x)(1-\delta_{n,0}),\ n\geq 0,(1\leq i\leq k) \tag{6.4}$$

$$\begin{aligned}\left(\lambda+n\gamma\right)D_n&=\int_0^{+\infty}\bar{p}_oP_{0,n-1}(x)\mu_0(x)dx+\int_0^{+\infty}P_{k,n}(x)\mu_k(x)dx\\&+\sum_{i=0}^{k}\int_0^{+\infty}R_{i,d,n}(x,y)\beta_{i,d}(y)dy+\lambda^+V_{n-1}\delta_{n,N},\ n\geq 1\end{aligned} \tag{6.5}$$

$$\begin{aligned}\frac{\partial R_{i,j,n}(x,y)}{\partial y}&=-\left(\lambda^++\beta_{i,j}(y)\right)R_{i,j,n}(x,y)+\sum_{m=1}^{n}c_m\lambda^+R_{i,j,n-m}(x,y)(1-\delta_{n,0}),\ n\geq 0,\\&(1\leq i\leq k),(1\leq j\leq d)\end{aligned} \tag{6.6}$$

Boundary Conditions

$$\begin{aligned}P_{0,n}(0)&=\lambda^+D_n(1-\delta_{n,0})+(n+1)\gamma D_{n+1}+(1-\delta_{n,0})\int_0^{+\infty}\lambda^+pP_{0,n-1}(x)dx+r\lambda^+D_n(1-\delta_{n,0})\\&+(1-r)\lambda^+D_{n+1},n\geq 0\end{aligned} \tag{6.7}$$

$$P_{1,n}(0)=\int_0^{+\infty}p_oP_{0,n}(x)\mu_0(x)dx,\ n\geq 0 \tag{6.8}$$

$$P_{i,n}(0)=\int_0^{+\infty}P_{i-1,n}(x)\mu_{i-1}(x)dx,\ n\geq 0,\ (2\leq i\leq k) \tag{6.9}$$

$$R_{i,1,n}(x,0)=\lambda^-\left[\int_0^{+\infty}P_{0,n}(x)dx+\int_0^{+\infty}P_{i,n}(x)dx+D_n(1-\delta_{n,0})\right],\ n\geq 0,(0\leq i\leq k) \tag{6.10}$$

$$R_{i,j,n}(x,0)=\int_0^{\infty}R_{i,j-1,n}(x,y)\beta_{j-1}(y)dy,\ n\geq 0,\ (2\leq j\leq d)\ ,\ (0\leq i\leq k) \tag{6.11}$$

Normalizing Condition

$$\sum_{n=1}^{\infty}D_n+\sum_{n=0}^{N-1}V_n+\sum_{n=0}^{\infty}\int_0^{+\infty}P_{0,n}(x)dx+\sum_{n=0}^{\infty}\sum_{i=1}^{k}\int_0^{\infty}P_{i,n}(x)dx+\sum_{n=0}^{\infty}\sum_{i=0}^{k}\sum_{j=1}^{d}\int_0^{\infty}\int_0^{\infty}R_{i,j,n}(x,y)dxdy=1 \tag{6.12}$$

To obtain the steady state solution of the system of differential equations (6.1)-(6.12), we use probability generating function technique. The probability generating functions and hazard rates for the batch size, idle state, busy state and repair state are same as defined in chapter 3. The generating function corresponding to the dormant period is considered as

$V(z)=\sum_{n=0}^{N-1}V_nz^n;\quad |z|\leq 1$.

Partial Generating Functions (PGFs) corresponding to dormant state, idle state, busy with first service (i=0), busy with ith ($1\leq$ i$\leq$ k) service, under jth ($1\leq$ j$\leq$ *d*) repair state respectively, are given in the following equations (6.13)-(6.20).

$$D(z)=D'(0)\left(\exp\left\{\int_0^z T_4(u)du\right\}\right)\int_0^z A_1(t)e^{-\int T_4(u)du}dt \tag{6.13}$$

$$V(z)=\frac{(\lambda^+(1+r)+\gamma)}{\lambda^+}D'(0)\left(\frac{1-z^N}{1-z}\right)$$
$$\times\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left\{1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)-p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)\tilde{b}_i(\lambda)\right\}\tilde{g}_{i,d}(\lambda^+)\right] \tag{6.14}$$

$$P_0(x,z)=P_0(0,z)\exp((\lambda^+\bar{p}C(z)-\lambda)x)\bar{B}_0(x) \tag{6.15}$$

$$P_i(x,z)=p_oP_0(0,z)\left[\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z))\right]\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^+C(z))\right]$$
$$\times\exp((\lambda^+C(z)-\lambda)x)\bar{B}_i(x),\ (1\le i\le k) \tag{6.16}$$

$$R_{i,j}(x,y,z)=R_{i,j}(x,0,z)\exp((\lambda^+C(z)-\lambda^+)y)\bar{G}_{ij}(y),\ (0\le i\le k),\ (1\le j\le d) \tag{6.17}$$

where,

$$P_0(0,z)=\frac{D'(0)(\lambda-\lambda^+\bar{p}C(z))\left[(z\lambda^++rz\lambda^++(1-r)\lambda^+)A_3(z)+z\gamma(A_1(z)+T_4(z)A_3(z))\right]}{z\left[\lambda-\lambda^+\bar{p}C(z)-z\lambda^+p+z\lambda^+p\tilde{B}_0(\lambda-\lambda^+\bar{p}C(z))\right]} \tag{6.18}$$

$$R_{i,1}(x,0,z)=\lambda^-P_0(0,z)\tilde{\tilde{B}}_0(\lambda-\lambda^+\bar{p}C(z))+\lambda^-D(z)$$
$$+\lambda^-\left[p_oP_0(0,z)\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^+C(z))\right]\left[\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z))\right]\tilde{\tilde{B}}_i(\lambda-\lambda^+C(z))\right],(0\le i\le k) \tag{6.19}$$

$$R_{i,j}(x,0,z)=\prod_{s=1}^{j-1}\tilde{g}_{i,s}(\lambda^+-\lambda^+C(z))R_{i,j}(x,0,z),\ (2\le j\le d),(0\le i\le k) \tag{6.20}$$

$$D'(0)=\frac{1}{\left(A_3(1)+A_2(1)+T_5(1)(1-\tilde{b}_0(\lambda-\lambda^+\bar{p}))+T_7(1)T_5(1)+\lambda^-\sum_{i=0}^{k}\sum_{j=1}^{d}\left(\beta_{i,j}^{-1}\right)T_6(1)T_5(1)\right)}$$

To derive expressions (6.13)-(6.20), we proceed as follows:

Multiplying eq. (6.1) by the appropriate powers of z and summing over n=0, 1, 2, 3, 4…, we get

$$V(z)=\frac{V_0(1-z^N)}{(1-z)} \tag{6.21}$$

Now, multiplying eqs (6.3), (6.4) and (6.6) by appropriate powers of z and summing over n=0, 1, 2, 3, 4…, we get

$$P_0(x,z)=P_0(0,z)\exp\left[(\lambda^+\bar{p}C(z)-\lambda)x\right]\bar{B}_0(x) \tag{6.22}$$

$$P_i(x,z)=P_i(0,z)\exp\left[(\lambda^+C(z)-\lambda)x\right]\bar{B}_i(x);\quad (1\le i\le k) \tag{6.23}$$

$$R_{ij}(x,y,z)=R_{ij}(x,0,z)\exp\left[(\lambda^{+}C(z)-\lambda^{+})y\right]\bar{G}_{ij}(y);\ \ (0\le i\le k),(1\le j\le d) \tag{6.24}$$

Using generating functions for the boundary conditions (6.7)-(6.11) respectively, we get

$$P_0(0,z)=\left[\lambda^{+}+r\lambda^{+}+\frac{(1-r)\lambda^{+}}{z}\right]D(z)+\gamma D'(z)+z\lambda^{+}p\int_0^{\infty}P_0(x,z)dx \tag{6.25}$$

$$P_1(0,z)=\int_0^{\infty}p_o\mu_0(x)P_0(x,z)dx \tag{6.26}$$

$$P_i(0,z)=\int_0^{+\infty}\mu_{i-1}(x)P_{i-1}(x,z)dx;\ \ (2\le i\le k) \tag{6.27}$$

$$R_{i,1}(x,0,z)=\lambda^{-}\left[\int_0^{\infty}P_0(x,z)dx+\int_0^{\infty}P_i(x,z)dx+D(z)\right];\quad (0\le i\le k) \tag{6.28}$$

$$R_{i,j}(x,0,z)=\int_0^{\infty}\beta_{j-1}(y)R_{i,j-1}(x,y,z)dy;\quad (0\le i\le k),(2\le j\le d) \tag{6.29}$$

Using (6.22) in (6.26) and (6.23) in (6.27), we obtain

$$P_1(0,z)=p_oP_0(0,z)\tilde{b}_0(\lambda-\lambda^{+}\bar{p}C(z)) \tag{6.30}$$

$$P_i(0,z)=P_{i-1}(0,z)\tilde{b}_{i-1}(\lambda-\lambda^{+}C(z)) \tag{6.31}$$

Solving (6.30) and (6.31) for different values of i, in general we get

$$P_i(0,z)=p_oP_0(0,z)\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^{+}C(z))\right]\tilde{b}_0(\lambda-\lambda^{+}\bar{p}C(z)) \tag{6.32}$$

Using generating function, from (6.10) and (6.11), we get

$$\begin{aligned}R_{i,1}(x,0,z)&=\lambda^{-}\left[P_0(0,z)\bar{\tilde{B}}_0(\lambda-\lambda^{+}\bar{p}C(z))+D(z)\right]\\&\quad+\lambda^{-}p_oP_0(0,z)\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^{+}C(z))\right]\tilde{b}_0(\lambda-\lambda^{+}\bar{p}C(z))\bar{\tilde{B}}_i(\lambda-\lambda^{+}C(z))\end{aligned} \tag{6.33}$$

$$R_{i,j}(x,0,z)=\prod_{s=1}^{j-1}\tilde{g}_{i,s}(\lambda^{+}-\lambda^{+}C(z))R_{i,1}(x,0,z),\ (0\le i\le k),(2\le j\le d) \tag{6.34}$$

Using (6.22) and (6.26), we have

$$P_0(0,z)=\frac{(\lambda^{+}-\lambda^{+}\bar{p}C(z))\left[(z\lambda^{+}+r\lambda^{+}z+(1-r)\lambda^{+})D(z)+z\gamma D'(z)\right]}{z\left[\lambda-\lambda^{+}\bar{p}C(z)-z\lambda^{+}p+z\lambda^{+}p\bar{\tilde{B}}_0(\lambda^{+}-\lambda^{+}\bar{p}C(z))\right]} \tag{6.35}$$

For n=0, we get

$$R_{i,j,0}(x,y)=R_{i,j,0}(x,0,0)\exp(-\lambda^{+}y)\bar{G}_{ij}(y),\ (0\le i\le k),(1\le j\le d)$$

Therefore, $\int_0^\infty R_{i,d,0}(x,y)\beta_d(y)dy = \int_0^\infty R_{i,d,0}(x,0,0)\exp(-\lambda^+ y)\bar{G}_{i,d}(y)\beta_{i,d}(y)dy$ (6.36)

$$R_{i,d,0}(x,0,0) = \frac{\lambda^-}{\lambda} P_0(0,0)\left[1-\tilde{b}_0(\lambda) + p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)(1-\tilde{b}_i(\lambda))\right] \tag{6.37}$$

where, $P_0(0,0) = \left[\lambda^+(1+r)+\gamma\right]D'(0)$ (6.38)

Therefore, $\int_0^\infty R_{i,d,0}(x,y)\beta_{i,d}(y)dy = R_{i,d,0}(x,0,0)\tilde{g}_{i,d}(\lambda^+)$ (6.39)

Similarly, $\int_0^\infty P_{k,0}(x)\mu_k(x)dx = p_o(\lambda^+(1+r)+\gamma)D'(0)\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)$ (6.40)

Using (6.39) and (6.40) in (6.2), we get

$$V_0 = \frac{(\lambda^+(1+r)+\gamma)D'(0)}{\lambda^+} \times\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda) + \sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left[1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)(1-\tilde{b}_i(\lambda))\right]\tilde{g}_{i,d}(\lambda^+)\right] \tag{6.41}$$

Further, using (6.41) in (6.21), we get the expression for the partial generating function for the server in dormant state V(z) as given by (6.14).

Now, solving (6.1), (6.2) and (6.5), we get

$$(\lambda^+ - \lambda^+ z)V(z) + \lambda D(z) + \gamma z D'(z) = z\bar{p}_o P_0(0,z)\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z)) + P_k(0,z)\tilde{b}_k(\lambda-\lambda^+C(z)) + \sum_{i=0}^{k} R_{i,d}(x,0,z)\tilde{g}_{i,d}(\lambda-\lambda^+C(z)) \tag{6.42}$$

which on further algebraic manipulation yields

$$D'(z) = T_3(z) + T_4(z)D(z) \tag{6.43}$$

where,

$$T_1(z) = \frac{(\lambda-\lambda^+\bar{p}C(z))\left[z\bar{p}_o + p_o\prod_{r=1}^{k}\tilde{b}_r(\lambda-\lambda^+C(z))\right]\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z))}{z\left(\lambda-\lambda^+\bar{p}C(z) - z\lambda^+ p + z\lambda^+ p\tilde{B}_0(\lambda-\lambda^+\bar{p}C(z))\right)} \tag{6.44}$$

$$T_2(z) = \frac{\left[1-\tilde{B}_0(\lambda-\lambda^+C(z)) + \sum_{i=0}^{k} p_o\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^+C(z))\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z))\tilde{\bar{B}}_i(\lambda-\lambda^+C(z))\right]}{z\left(\lambda-\lambda^+\bar{p}C(z) - z\lambda^+ p + z\lambda^+ p\tilde{B}_0(\lambda-\lambda^+\bar{p}C(z))\right)} \tag{6.45}$$

$$T_3(z)=(\lambda^+(1+r)+\gamma)D'(0)(z^N-1)$$
$$\times\frac{\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left[1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)(1-\tilde{b}_i(\lambda))\right]\tilde{g}_{i,d}(\lambda^+)\right]}{\left[z\gamma-z\gamma(T_1(z)+\lambda^-T_2(z)\sum_{i=0}^{k}\left(\prod_{s=1}^{d}\tilde{g}_{i,s}(\lambda^+-\lambda^+C(z))\right)\right]}\tag{6.46}$$

$$T_4(z)=\frac{(\lambda^+z(1+r)+(1-r)\lambda^+)T_1(z)+\left\{(\lambda^+z(1+r)+(1-r)\lambda^+)\lambda^-T_2(z)-\lambda^-\right\}\sum_{i=0}^{k}\left(\prod_{s=1}^{d}\tilde{g}_{i,s}(\lambda^+-\lambda^+C(z))\right)-\lambda}{\left[z\gamma-z\gamma(T_1(z)+\lambda^-T_2(z)\sum_{i=0}^{k}\left(\prod_{s=1}^{d}\tilde{g}_{i,s}(\lambda^+-\lambda^+C(z))\right)\right]}\tag{6.47}$$

Now (6.43) yields

$$D(z)=e^{\int_0^z T_4(u)du}\left[\int_0^z\left(T_3(t)e^{-\int_0^z T_4(u)du}\right)dt\right]\tag{6.48}$$

***Computation of* $\mathbf{D'(0)}$**

Now, we proceed to compute the value of $D'(0)$. Here,

$T_3(z)=A_1(z)D'(0)\ \Rightarrow T_3(1)=A_1(1)D'(0);$
$V(z)=A_2(z)D'(0)\ \Rightarrow V(1)=A_2(1)D'(0);$
$D(z)=A_3(z)D'(0)\Rightarrow D(1)=A_3(1)D'(0);$

where

$$A_1(z)=(\lambda^+(1+r)+\gamma)(z^N-1)$$
$$\times\frac{\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left[1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)(1-\tilde{b}_i(\lambda))\right]\tilde{g}_{i,d}(\lambda^+)\right]}{\left[z\gamma-z\gamma(T_1(z)+\lambda^-T_2(z)\sum_{i=0}^{k}\left(\prod_{s=1}^{d}\tilde{g}_{i,s}(\lambda^+-\lambda^+C(z))\right)\right]}\tag{6.49}$$
$$(0\le i\le k),$$

$$A_2(z)=\frac{(\lambda^+(1+r)+\gamma)}{\lambda^+}$$
$$\times\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left[1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)(1-\tilde{b}_i(\lambda))\right]\tilde{g}_{i,d}(\lambda^+)\right]\left(\frac{1-z^N}{1-z}\right)\tag{6.50}$$

$$A_3(z)=e^{\int_0^z T_4(u)du}\left[\int_0^z\left(A_1(t)e^{-\int_0^z T_4(u)du}\right)dt\right]\tag{6.51}$$

$$A_1(1)=\frac{(\lambda^+(1+r)+\gamma)N\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left[1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)(1-\tilde{b}_i(\lambda))\right]\tilde{g}_{i,d}(\lambda^+)\right]}{\gamma\left[1-(T_1'(1)+\lambda^-T_2'(1)-\lambda^-\lambda^+T_2(1)\sum_{i=0}^{k}\sum_{s=1}^{d}g_{i,s}^{(1)}\right]}$$

(6.52)

$$A_2(1)=\frac{N(\lambda^+(1+r)+\gamma)}{\lambda^+}\times\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left[1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)(1-\tilde{b}_i(\lambda))\right]\tilde{g}_{i,d}(\lambda^+)\right] \quad (6.53)$$

Hence, $A_3(1)=e^{\int_0^1 T_4(u)du}\left[\int_0^1\left(A_1(t)e^{-\int_0^1 T_4(u)du}\right)dt\right]$ (6.54)

Now, $\lim_{z\to1}T_1(z)=\lim_{z\to1}\frac{(\lambda-\lambda^+\overline{p}C(z))\left[z\overline{p}_o+p_o\prod_{r=1}^{k}\tilde{b}_r(\lambda-\lambda^+C(z))\right]\tilde{b}_0(\lambda-\lambda^+\overline{p}C(z))}{\lambda-\lambda^+\overline{p}C(z)-z\lambda^+p+z\lambda^+p\tilde{B}_0(\lambda-\lambda^+\overline{p}C(z))}$

$$\Rightarrow T_1(1)=\frac{(\lambda-\lambda^+\overline{p})\left[(1-p_o)+p_o\prod_{r=1}^{k}\tilde{b}_r(\lambda^-)\right]\tilde{b}_0(\lambda-\lambda^+\overline{p})}{\lambda^-+\lambda^+p\tilde{B}_0(\lambda-\lambda^+\overline{p})}, \quad (6.55)$$

$$T_2(1)=\frac{\left[1-\tilde{B}_0(\lambda^-)+\sum_{i=0}^{k}p_o\prod_{r=1}^{i-1}\tilde{b}_r(\lambda^-)\tilde{b}_0(\lambda-\lambda^+\overline{p})\tilde{\tilde{B}}_i(\lambda^-)\right]}{\lambda^-+\lambda^+p\tilde{B}_0(\lambda-\lambda^+\overline{p})} \quad (6.56)$$

$$T_4(1)=\frac{(\lambda^+(1+r))(T_1(1)+T_2(1))+(2\lambda^+)(T_1'(1)+T_2'(1))+(2\lambda^+T_2(1)-\lambda^-)\sum_{i=0}^{k}\sum_{s=1}^{d}g_{i,s}^{(1)}}{\gamma\left[1-(T_1(1)+\lambda^-T_2(1))-(T_1'(1)+\lambda^-T_2'(1)+\lambda^-T_2(1)\sum_{i=0}^{k}\sum_{s=1}^{d}g_{i,s}^{(1)}\right]}$$

Also, $D'(1)=D'(0)\left[A_1(1)+T_4(1)A_3(1)\right]$ (6.57)

To obtain $D'(0)$, we use generating function approach on the normalizing condition as given by eq. (6.12). Thus,

$$\lim_{z\to1}\left[D(z)+V(z)+\int_0^\infty P_0(x,z)dx+\sum_{i=1}^{k}\int_0^\infty P_i(x,z)dx+\sum_{i=1}^{k}\sum_{j=1}^{d}\int_0^\infty\int_0^\infty R_{i,j}(x,y,z)dxdy\right]=1$$

$$D'(0)=\frac{1}{\left(A_3(1)+A_2(1)+T_5(1)(1-\tilde{b}_0(\lambda-\lambda^+\overline{p}))+T_7(1)T_5(1)+\lambda^-\sum_{i=0}^{k}\sum_{j=1}^{d}\left(\beta_{i,j}^{-1}\right)T_6(1)T_5(1)\right)} \quad (6.58)$$

where,

$$T_5(1)=\frac{(2\lambda^+)A_3(1)+\gamma(A_1(1)+T_4(1)A_3(1))}{\lambda^-+\lambda^+ p\tilde{b}_0(\lambda-\lambda^+\overline{p})}$$

$$T_6(1)=\left[\frac{(1-\tilde{b}_0(\lambda-\lambda^+\overline{p}))}{(\lambda-\lambda^+\overline{p})}+\sum_{i=0}^{k}p_o\prod_{r=1}^{i-1}\tilde{b}_r(\lambda^-)\tilde{b}_0(\lambda-\lambda^+\overline{p})\frac{(1-\tilde{b}_i(\lambda^-))}{(\lambda^-)}\right]$$

$$T_7(1)=\sum_{i=1}^{k}\left(p_o(\lambda-\lambda^+\overline{p})\frac{(1-\tilde{b}_i(\lambda^-))}{(\lambda^-)}\prod_{r=1}^{i-1}\tilde{b}_r(\lambda^-)\tilde{b}_0(\lambda-\lambda^+\overline{p})\right)$$

Hence, partial generating functions for different states of the server obtained in the above manner and are given by expressions (6.13) - (6.20).

Corollary: The necessary and sufficient condition for the system to be stable is

$$\left(A_3(1)+A_2(1)+T_5(1)(1-\tilde{b}_0(\lambda-\lambda^+\overline{p}))+T_7(1)T_5(1)+\lambda^-\sum_{i=0}^{k}\sum_{j=1}^{d}\left(\beta_{i,j}^{-1}\right)T_6(1)T_5(1)\right)>1 \qquad (6.59)$$

Proof: In order to find the stability condition, we follow the approach of Wang *et al.* (2001).

Theorem 6.1: The marginal generating functions corresponding to the idle state, dormant state, busy with first service, busy with ith (1 ≤ i≤ k) service, under jth (1 ≤ j≤ *d*) repair state are

$$D(z)=D'(0)\left(\exp\left\{\int_0^z T_4(u)du\right\}\right)\int_0^z A_1(t)e^{-\int T_4(u)du}dt \qquad (6.60)$$

$$V(z)=\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left\{1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)-p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)\tilde{b}_i(\lambda)\right\}\right]$$
$$\times\frac{(\lambda^+(1+r)+\gamma)}{\lambda^+}D'(0)\left(\frac{1-z^N}{1-z}\right) \qquad (6.61)$$

$$P_0(z)=\frac{D'(0)\left[1-\tilde{b}_0(\lambda-\lambda^+\overline{p}C(z))\right]\left[(z\lambda^++rz\lambda^++(1-r)\lambda^+)A_3(z)+z\gamma(A_1(z)+T_4(z)A_3(z))\right]}{z\left[\lambda-\lambda^+\overline{p}C(z)-z\lambda^+p+z\lambda^+p\tilde{B}_0(\lambda-\lambda^+\overline{p}C(z))\right]} \qquad (6.62)$$

$$P_i(z)=p_oP_0(0,z)\left[\tilde{b}_0(\lambda-\lambda^+\overline{p}C(z))\right]\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^+C(z))\right]\frac{\left[1-\tilde{b}_i(\lambda-\lambda^+C(z))\right]}{(\lambda-\lambda^+C(z))}, \qquad (6.63)$$
$$(1\le i\le k)$$

$$R_{i,j}(z)=R_{i,j}(x,0,z)\frac{\left[1-\tilde{g}_{i,j}(\lambda^+-\lambda^+C(z))\right]}{(\lambda^+-\lambda^+C(z))},\ (1\le j\le d),(0\le i\le k) \qquad (6.64)$$

Proof: The marginal probability generating functions for the different states of the server given in equations (6.60)-(6.64) can be determined by using

$$P_0(z)=\int_0^\infty P_0(x,z)dx, P_i(z)=\int_0^\infty P_i(x,z)dx, R_{i,j}(z)=\int_0^\infty\int_0^\infty R_{i,j}(x,y,z)dxdy,(0\le i\le k),(1\le j\le d)$$

Here, D (z) and V(z) are same as obtained earlier in theorem 6.1.

Theorem 6.2: The generating function for the number of customers in the orbit is

$$K_1(z)=\frac{(\lambda^+(1+r)+\gamma)}{\lambda^+}D'(0)\left(\frac{1-z^N}{1-z}\right)$$
$$\times\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left\{1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)-p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)\tilde{b}_i(\lambda)\right\}\tilde{g}_{i,d}(\lambda^+)\right]$$
$$+D'(0)F(z)\left[\frac{(1-\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z)))}{(\lambda-\lambda^+\bar{p}C(z)))}\right]+D'(0)A_3(z)$$
$$+D'(0)F(z)\sum_{i=1}^{k}p_o\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^+C(z))\right]\frac{\left[1-\tilde{b}_i(\lambda-\lambda^+C(z))\right]}{(\lambda-\lambda^+C(z))}\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z))$$
$$+\sum_{i=0}^{k}\sum_{j=1}^{d}\lambda^-D'(0)\left[1+\prod_{s=1}^{j-1}\tilde{g}_{is}(\lambda^+-\lambda^+C(z))\right]\frac{\left[1-\tilde{g}_{ij}(\lambda^+-\lambda^+C(z))\right]}{(\lambda^+-\lambda^+C(z))}$$
$$\times\left\{F(z)\left\{\tilde{\tilde{B}}_0(\lambda-\lambda^+\bar{p}C(z))+p_o\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^+C(z))\right]\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z))\tilde{\tilde{B}}_i(\lambda-\lambda^+C(z))\right\}+A_3(z)\right\}$$

(6.65)

$$\text{where, } F(z)=\frac{(\lambda-\lambda^+\bar{p}C(z))\left[(z\lambda^++rz\lambda^++(1-r)\lambda^+)A_3(z)+z\gamma(A_1(z)+T_4(z)A_3(z))\right]}{z\left[\lambda-\lambda^+\bar{p}C(z)-z\lambda^+p+z\lambda^+p\tilde{B}_0(\lambda-\lambda^+\bar{p}C(z))\right]} \quad (6.66)$$

Proof: The generating function for the number of customers in the orbit is obtained as

$$K_1(z)=V(z)+D(z)+P_0(z)+\sum_{i=1}^{k}P_i(z)+\sum_{i=0}^{k}\sum_{j=1}^{d}R_{i,j}(z) \quad (6.67)$$

Theorem 6.3: The generating function for the number of customers in the system is

$$K_2(z)=\frac{(\lambda^+(1+r)+\gamma)}{\lambda^+}D'(0)\left(\frac{1-z^N}{1-z}\right)$$
$$\times\left[p_o\prod_{r=0}^{k-1}\tilde{b}_r(\lambda)\tilde{b}_k(\lambda)+\sum_{i=0}^{k}\frac{\lambda^-}{\lambda}\left\{1-\tilde{b}_0(\lambda)+p_o\tilde{b}_0(\lambda)\prod_{r=1}^{i-1}\tilde{b}_r(\lambda)(1-\tilde{b}_i(\lambda))\right\}\tilde{g}_{i,d}(\lambda^+)\right]$$
$$+zD'(0)F(z)\left[\frac{(1-\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z)))}{(\lambda-\lambda^+\bar{p}C(z)))}\right]+D'(0)A_3(z)$$
$$+zD'(0)F(z)\sum_{i=1}^{k}p_o\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^+C(z))\right]\frac{\left[1-\tilde{b}_i(\lambda-\lambda^+C(z))\right]}{(\lambda-\lambda^+C(z))}\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z))$$
$$+z\sum_{i=0}^{k}\sum_{j=1}^{d}\lambda^-D'(0)\left[1+\prod_{s=1}^{j-1}\tilde{g}_{is}(\lambda^+-\lambda^+C(z))\right]\frac{\left[1-\tilde{g}_{ij}(\lambda^+-\lambda^+C(z))\right]}{(\lambda^+-\lambda^+C(z))}$$
$$\times\left\{F(z)\left\{\tilde{\tilde{B}}_0(\lambda-\lambda^+\bar{p}C(z))+p_o\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda-\lambda^+C(z))\right]\tilde{b}_0(\lambda-\lambda^+\bar{p}C(z))\tilde{\tilde{B}}_i(\lambda-\lambda^+C(z))\right\}+A_3(z)\right\}$$

(6.68)

Proof: The generating function for the number of customers in the system is obtained by using

$$K_2(z) = V(z) + D(z) + zP_0(z) + z\sum_{i=1}^{k} P_i(z) + z\sum_{i=0}^{k}\sum_{j=1}^{d} R_{i,j}(z) \quad (6.69)$$

6.4. SPECIAL CASES

In order to justify the general framework of our model, we discuss here some special cases by setting appropriate parameters as discussed below:

(a) If $E(X)=1, E(X^2)=0, k=1, d=1, r=0,$ then the present model reduces to M/G/1 retrial queue with single repair without discouragement. In this case our model matches with the model studied by Liu *et al.* (2009).

(b) On substituting, $E(X)=1, E(X^2)=0, \lambda^-=0, r=0, N=0, p=0, i(1\le i\le k)=0, \quad d=1,$ $p_o=1,$ our model deals with the non-markovian retrial queue with breakdowns; the same model was studied by Falin (2010a).

(c) If $E(X)=1, E(X^2)=0, \lambda^-=0, r=0, N=0, p=0, d=1, i(1\le i\le k)=0,$ $b^{(1)}=\frac{1}{\mu_0},$ $b^{(2)}=\frac{1}{2\mu_0^2},$ then the present model reduces to M/M/1 retrial queue with breakdowns studied by Sherman and Kharoufeh (2006).

(d) If $\lambda^-=0, r=0, p=0, d=1, i(2\le i\le k)=0, \alpha_i=0, \beta_{ij}=0, (0\le i\le k), (1\le j\le d),$ then we come across the model dealing with bulk retrial queue with additional service under N-policy which was also studied by Choudhury and Paul (2004).

(e) If $E(X)=1, E(X^2)=0, r=0, N=0, p=0, d=1, i(1\le i\le k)=0,$ then our model provides results for M/G/1 retrial queue with disasters and failures. Here, if we consider the arrival of negative customers as the arrival or attack rate of catastrophes, then our model portrays the same queueing problem as explored by Wang *et al.* (2008b).

6.5. PERFORMANCE MEASURES

The significant performance indices are derived as follows:

(A) Long Run Probabilities

The long run probabilities of the system states are determined by using the probability generating functions as follows:

Theorem 6.4: The long run probabilities corresponding to various states of the server are:

- The probability of the server being idle, is

$$P_I = D'(0)A_3(1) \tag{6.70}$$

- The probability that the server being in dormant state, is

$$P_V = D'(0)A_2(1) \tag{6.71}$$

- The probability that the server being busy with first essential service, is

$$P_{B_0} = \frac{D'(0)\left[1-\tilde{b}_0(\lambda-\lambda^+\overline{p})\right]\left[(2\lambda^+)A_3(1)+\gamma(A_1(1)+T_4(1)A_3(1))\right]}{\left[\lambda^- + \lambda^+ p\tilde{B}_0(\lambda-\lambda^+\overline{p})\right]} \tag{6.72}$$

- The probability that the server being busy with i^th^ (1 ≤i≤ k) phase optional service, is

$$P_{B_i} = D'(0)\left[1-\tilde{b}_0(\lambda-\lambda^+\overline{p})\right]\left[(2\lambda^+)A_3(1)+\gamma(A_1(1)+T_4(1)A_3(1))\right]$$
$$\times \frac{\left[\prod_{r=1}^{i-1}\tilde{b}_r(\lambda^-)\right]\left[\tilde{b}_0(\lambda-\lambda^+\overline{p})\right]\left[1-\tilde{b}_i(\lambda^-)\right]}{\lambda^-\left[\lambda^- + \lambda^+ p\tilde{B}_0(\lambda-\lambda^+\overline{p})\right]} \tag{6.73}$$

- The probability that the broken down server being under first phase of repair, is

$$P_{R_1} = \lambda^- D(1) + \sum_{i=1}^{k}\lambda^-\left\{\frac{(\lambda-\lambda^+\overline{p})\left[(2\lambda^+)A_3(1)+\gamma(A_1(1)+T_4(1)A_3(1))\right]}{\left[\lambda^- + \lambda^+ p\tilde{B}_0(\lambda-\lambda^+\overline{p})\right]}\right\}$$
$$\times\left[\left(\frac{(1-\tilde{b}_0(\lambda-\lambda^+\overline{p}))}{(\lambda-\lambda^+\overline{p})}\right)+p_o\prod_{r=1}^{i-1}\tilde{b}_r(\lambda^-)\tilde{b}_0(\lambda-\lambda^+\overline{p}))\left(\frac{(1-\tilde{b}_i(\lambda^-))}{(\lambda^-)}\right)\right],(0\le i\le k) \tag{6.74}$$

- The probability that the broken down server is under j^th^ $(2 \le j \le d)$ phase of repair, is

$$P_{R_j} = \sum_{i=1}^{k}\sum_{j=1}^{d}\lambda^- D(1)(\beta_{i,j}{}^{-1}) + \sum_{i=1}^{k}\sum_{j=1}^{d}\lambda^-(\beta_{i,j}{}^{-1})\left\{\frac{(\lambda-\lambda^+\overline{p})\left[(2\lambda^+)A_3(1)+\gamma(A_1(1)+T_4(1)A_3(1))\right]}{\left[\lambda^- + \lambda^+ p\tilde{B}_0(\lambda-\lambda^+\overline{p})\right]}\right\}$$
$$\times\left[\left(\frac{(1-\tilde{b}_0(\lambda-\lambda^+\overline{p}))}{(\lambda-\lambda^+\overline{p})}\right)+p_o\prod_{r=1}^{i-1}\tilde{b}_r(\lambda^-)\tilde{b}_0(\lambda-\lambda^+\overline{p}))\left(\frac{(1-\tilde{b}_i(\lambda^-))}{(\lambda^-)}\right)\right],\ (2\le j\le d),(0\le i\le k)$$

(6.75)

Proof: The proof follows on the lines of theorem 2.5.

(B) Queue Length

The average number of customers waiting for the service in the queue and in the orbit is the key metrics which are obtained as follows:

Theorem 6.5: The average queue length of the retrial orbit (L_R) and average queue length of the system (L_S) are given by:

$$L_R = E_1 + E_2 + E_3 + E_4 + E_5(F_2 + F_3 + F_4) + \sum_{i=0}^{k} F_1(E_6 + E_7 + E_8) \tag{6.76}$$

and, $L_S = L_R + P_0(1) + \sum_{i=1}^{k} P_i(1) + \sum_{i=0}^{k}\sum_{j=1}^{d} R_{i,j}(1)$ (6.77)

where,

$$H_6' = \sum_{s=1}^{j-1} g_{ij}^{(1)}\lambda^+ E[X], H_7' = g_{ij}^{(1)}(-\lambda^+ E[X]), H_7'' = g_{ij}^{(2)}(-\lambda^+ E[X])^2 + g_{ij}^{(1)}(-\lambda^+ C''(1)),$$

$$H_8' = (-\lambda^+ E[X]), H_8'' = (-\lambda^+ C''(1))$$

$$F_1 = \lambda^- D'(0) g_{ij}^{(1)}, F_2 = (\lambda - \lambda^+ \overline{p}) F(1) \tilde{\tilde{B}}_0(\lambda - \lambda^+ \overline{p}),$$

$$F_3 = (\lambda - \lambda^+ \overline{p}) F(1) p_o \prod_{r=1}^{i-1} \tilde{b}_r(\lambda^-) \tilde{b}_0(\lambda - \lambda^+ \overline{p}) \tilde{\tilde{B}}_i(\lambda^-), F_4 = (\lambda - \lambda^+ \overline{p}) F(1) A_3(1)$$

$$E_1 = \frac{(\lambda^+(1+r)+\gamma)}{\lambda^+} D'(0) \left(\frac{N(N-1)}{2} \right)$$

$$\times \left[p_o \prod_{r=0}^{k-1} \tilde{b}_r(\lambda) \tilde{b}_k(\lambda) + \sum_{i=0}^{k} \frac{\lambda^-}{\lambda} \left\{ 1 - \tilde{b}_0(\lambda) + p_o \tilde{b}_0(\lambda) \prod_{r=1}^{i-1} \tilde{b}_r(\lambda) - p_o \tilde{b}_0(\lambda) \prod_{r=1}^{i-1} \tilde{b}_r(\lambda) \tilde{b}_i(\lambda) \right\} \right]$$

$$E_2 = D'(0) A_3'(1), E_3 = D'(0) \left[F'(1)(1 - \tilde{b}_0(\lambda - \lambda^+ \overline{p})) + F(1)(-\tilde{b}_0'(\lambda - \lambda^+ \overline{p}))(-\lambda^+ \overline{p} E[X]) \right]$$

$$E_4 = \sum_{i=1}^{k} p_o D'(0) \left[\frac{\lambda^-(M'(1)) - M(1)(-\lambda^+ E[X])}{(\lambda^-)^2} \right],$$

$$E_5 = \sum_{i=0}^{k}\sum_{j=1}^{d} \lambda^- D'(0) \left[\frac{2H_6' H_7' H_8' + H_8' H_7'' - H_7' H_8''}{2(H_8')^2} \right]$$

$$E_7 = (-\lambda^+ \overline{p} C'(1)) F(1) p_o \prod_{r=1}^{i-1} \tilde{b}_r(\lambda^-) \tilde{b}_0(\lambda - \lambda^+ \overline{p}) \tilde{\tilde{B}}_i(\lambda^-)$$

$$+ (\lambda - \lambda^+ \overline{p}) F'(1) p_o \prod_{r=1}^{i-1} \tilde{b}_r(\lambda^-) \tilde{b}_0(\lambda - \lambda^+ \overline{p}) \tilde{\tilde{B}}_i(\lambda^-)$$

$$+ (\lambda - \lambda^+ \overline{p}) F(1) p_o \prod_{r=1}^{i-1} \tilde{b}_r'(\lambda^-)(-\lambda^+ E[X]) \tilde{b}_0(\lambda - \lambda^+ \overline{p}) \tilde{\tilde{B}}_i(\lambda^-)$$

$$+ (\lambda - \lambda^+ \overline{p}) F(1) p_o \prod_{r=1}^{i-1} \tilde{b}_r(\lambda^-) \tilde{b}_0'(\lambda - \lambda^+ \overline{p})(-\lambda^+ \overline{p} E[X]) \tilde{\tilde{B}}_i(\lambda^-)$$

$$+ (\lambda - \lambda^+ \overline{p}) F(1) p_o \prod_{r=1}^{i-1} \tilde{b}_r(\lambda^-) \tilde{b}_0(\lambda - \lambda^+ \overline{p}) \tilde{\tilde{B}}_i'(\lambda^-)(-\lambda^+ E[X])$$

$$E_6 = F'(1)(\lambda - \lambda^+ \overline{p}) \tilde{\tilde{B}}_0(\lambda - \lambda^+ \overline{p}) + F(1)(-\lambda^+ \overline{p} E[X]) \tilde{\tilde{B}}_0(\lambda - \lambda^+ \overline{p})$$

$$+ F(1)(\lambda - \lambda^+ \overline{p}) \tilde{\tilde{B}}_0(\lambda - \lambda^+ \overline{p})(-\lambda^+ \overline{p} E[X])$$

$$E_8 = (\lambda - \lambda^+ \overline{p}) F'(1) A_3(1) + (-\lambda^+ \overline{p} E[X]) F(1) A_3(1) + (\lambda - \lambda^+ \overline{p}) F(1) A_3'(1)$$

Proof: The average queue length of the retrial orbit and the system are obtained by using $L_R = \lim_{z \to 1} K_1'(z)$ and $L_S = \lim_{z \to 1} K_2'(z)$, respectively.

(C) Reliability Measures

Here, we derive explicit expressions for the reliability indices namely availability and failure frequency of the system as given in the following theorem:

Theorem 6.6: The steady state availability (A_v) and failure frequency (F_f) of the server are given by

$$A_v = 1 - \left[\sum_{i=0}^{k} \sum_{j=1}^{d} \lambda^- D(1) (\beta_{ij})^{-1} + \sum_{i=0}^{k} \sum_{j=1}^{d} \lambda^- (\beta_{i,j}{}^{-1}) \left\{ (\lambda - \lambda^+ \overline{p}) F_2(1) \right\} T_6(1) \right], \ (1 \le j \le d) \qquad (6.78)$$

$$F_f = \lambda^- D'(0) \left(A_3(1) + F_2(1)(\lambda^-)^{-1} \left[1 - \tilde{b}_0(\lambda - \lambda^+ \overline{p}) \right] \left(\lambda^- + \sum_{i=1}^{k} \left[\prod_{r=1}^{i-1} \tilde{b}_r(\lambda^-) \right] \left[\tilde{b}_0(\lambda - \lambda^+ \overline{p}) \right] \left[1 - \tilde{b}_i(\lambda^-) \right] \right) \right) \qquad (6.79)$$

where, $$F_2(1) = \frac{\left[(2\lambda^+) A_3(1) + \gamma (A_1(1) + T_4(1) A_3(1)) \right]}{\left[\lambda^- + \lambda^+ p \tilde{B}_0(\lambda - \lambda^+ \overline{p}) \right]} \qquad (6.80)$$

Proof: The availability of the server can be determined by using

$$A_v = 1 - P_{R_j}, \ (1 \le j \le d) \qquad (6.81)$$

$$F_f = \lambda^- \left(D(1) + \int_0^{\infty} P_0(x,1) dx + \sum_{i=1}^{k} \int_0^{\infty} P_i(x,1) dx \right) \qquad (6.82)$$

6.6 DISCUSSION

The negative arrival retrial queue with preemptive resume investigated in this chapter incorporates many realistic features viz. bulk arrival and impatient behavior of the customers. The explicit expressions derived for the generating functions and various performance indices including queue length, reliability indices etc. make our investigation applicable for the quantitative prediction of delay metrics which can be further used for the improvement of many real time systems wherein the working of such system is not only degraded but also stopped due to the attack of negative customers. The model under consideration depicts many real life congestion situations as it involves the feature of availing extra services as per choice of the customers. Our model can be easily implemented for present day computer communication congestion situations to obtain cost optimal solution for specified grade of service for various practical situations under technical constraints.

CHAPTER 7

EMBEDDED MARKOV CHAIN ANALYSIS OF RETRIAL QUEUE WITH VACATION

7.1 INTRODUCTION

In this chapter, we incorporate the concept of Bernoulli admission mechanism to vacation retrial queue. Bernoulli admission mechanism can be used to control the admission or arrival of the customers in the queues with reattempts or retrials. Artalejo and Atencia (2004) and Artalejo *et al.* (2005) proposed this policy for the admission control in continuous and discrete queueing systems, respectively. It is reasonable to assume that the arrival of the new customers is controlled in such a way that each individual blocked customer is admitted or allowed to join the system with probability $\zeta\,(0 \le \zeta \le 1)$. If the arriving customer/batch finds the server in idle state, then one of the admitted customers joins the server whereas rest of the customers join the retrial group; otherwise if the server is busy, the whole batch joins the orbit. This mechanism which is known as Bernoulli admission policy can be considered as a control device to reduce the congestion at the initial stage.

Three important retrial policies namely classical retrial policy, constant retrial policy and linear retrial policy have been proposed in the literature by the researchers from time to time. In classical retrial policy, the customer make reattempts with exponentially distributed rate $n\gamma$ where n is the number of customers present in the retrial group. The reattempts are allowed with retrial rate $\nu(1-\delta_{n,0})$ during constant retrial policy with the intervals between successive retrial attempts following exponential distribution; here $\delta_{n,0}$ denotes the Kronecker's delta. The linear retrial policy deals with the fact that the customers retry from the orbit with retrial rate $\nu(1-\delta_{n,0})+n\gamma$, where n is the number of customers in the orbit, $\delta_{n,0}$ is the Kronecker's delta function and γ is the retrial rate, respectively.

The objective of our investigation in the present chapter is to analyze M/G/1 bulk arrival retrial queue with admission control using embedded markov chain technique. In comparison to earlier existing models, the queueing model under consideration is developed by keeping more versatile congestion scenarios in the sense that it includes various noble features altogether such as (i) multi-optional services, (ii) phase repair, (iii) Bernoulli vacation schedule, (iv) Bernoulli admission policy, (v) unreliable server and (vi) bulk arrival. The rest of the investigation has been organized as follows. Section 7.2 is devoted to the description of the model with various assumptions and notations describing the model. The limiting distribution of the queue size using embedded markov technique has been obtained in section 7.3. Section 7.4 presents the distribution of the number of customers in the orbit and queue at equilibrium. Section 7.5 explains the stochastic decomposition property for the model. Various performance indices are obtained in section 7.6. Some special cases of our model are deduced by setting some appropriate parameters in section 7.7. Further, cost function for the model has been framed in section 7.8. The numerical simulation of obtained analytical results has been done in section 7.9. Finally, we wind up our investigation by highlighting the noble features of the work done in section 7.10.

7.2. DESCRIPTION OF THE MODEL

A single server retrial queueing system with bulk arrival and Bernoulli vacation schedule has been considered. The incoming customers are admitted in the system following Bernoulli admission mechanism. The various assumptions and notations describing the model are stated as follows:

- **Arrival Process and Bernoulli Admission Mechanism**

The customers arrive in batches of m customers with probability c_m ; m ≥1. Let a_n be the probability with which a batch of n customers joins the system following Poisson process with arrival rate λ. Thus,

$$a_0 = \sum_{m=1}^{\infty} c_m (1-\zeta)^m \text{ and } a_n = \sum_{m=n}^{\infty} c_m \binom{m}{n} \zeta^m (1-\zeta)^{m-n}, \quad n \geq 1$$

The probability generating function of the sequences $(a_n; n \geq 0)$ is $a(z) = \sum_{n=0}^{\infty} a_n z^n$. The two sequences are related to each other as $a(z) = C((1-\zeta)+\zeta z)$. If $\zeta = 1$, then $a(z) = C(z) = \sum_{n=0}^{\infty} c_n z^n$. Moreover, first moment is $a = \zeta E[X]$, and $a^{(r)} = \zeta^r c^{(r)}$ where $a^{(r)}$ is the rth (r ≥2) factorial moment of $\{a_n\}_{n=1}^{\infty}$.

- **Retrial Process and Bernoulli Vacation Schedule**

On finding the server busy or in non-working state, the customers admitted to the system wait in the orbit from where they can retry for the service.

The server may go for a vacation of random interval after each service completion with probability $\sigma_1(\sigma_i)$ during FES (i[th] SOS, 2≤ i≤ k) or it may continue to serve new incoming customers with complementary probability $\bar{\sigma}_1(\bar{\sigma}_i)$.

- **Service and Repair Processes**

The customers admitted to the system are served in multi-optional phases where first phase being compulsory phase for all the customers and is known as first essential service (FES). The customer after completing FES may avail 'i' (2≤ i≤ k) number of second optional services (SOS) in succession out of total 'k-1' optional services available. After completing first optional service, the customer may opt for the next second optional service and so on. In general after completing $(i\text{-}1)^{th}$ optional service, the customer may go for the i^{th} (2≤ i≤ k) optional service with probability p_{i-1} (2≤ i≤ k) or may move out of the system with complementary probability (1- p_{i-1}). After completion of k^{th} service, it departs from the system.

The server is unreliable and may breakdown in Poisson fashion with breakdown rate $\alpha_i, (1 \le i \le k)$ during any state of service. The repair process is completed in 'd' compulsory phases with repair rate $\beta_{ij}, (1 \le i \le k)(1 \le j \le d)$. The cdf, pdf, LST of pdf, mean and r^{th} (≥ 2) moments for the i[th] $(1 \le i \le k)$ phase service process are respectively given by $B_i(t)$, $b_i(t)$, $\tilde{b}_i(s)$, b_i and $b_i^{(r)}$. The cdf, pdf, LST of pdf, mean and r^{th} (≥ 2) moments for the repair process are given by $G_{ij}(t)$, $g_{ij}(t)$, $\tilde{g}_{ij}(s)$, g_{ij} and $g_{ij}^{(r)}$, respectively.

- **Generalized Service Time**

The generalized service time H_i (t) $(1 \le i \le k)$ of i^{th} $(1 \le i \le k)$ phase service can be defined as the total time taken to complete i^{th} phase service including both service time and repair time. Now, LST for the distribution function of the generalized service time of H_i (t) is

$$\begin{aligned}\tilde{H}_i(s) &= \sum_{n=0}^{\infty}\int_0^{\infty} e^{-st} e^{-\alpha_i t}\left[\frac{(\alpha_i t)^n}{n!}\right]\left(\prod_{j=1}^{d}\tilde{g}_{ij}(s)\right)^n dB_i(t) \\ &= \tilde{b}_i(s+\alpha_i(1-\prod_{j=1}^{d}\tilde{g}_{ij}(s))\end{aligned} \tag{7.1}$$

The first and second moment of i^{th} phase generalized service time respectively, are given as:

$$h_i = b_i(1+\alpha_i \sum_{j=1}^{d} g_{ij}) \text{ and } h_i^{(2)} = b_i^{(2)}(1+\alpha_i \sum_{j=1}^{d} g_{ij})^2 + b_i(\alpha_i \sum_{j=1}^{d} g_{ij}^{(2)})$$

Let N(t) represents the number of customers in the orbit at any time t and the phase of service and phase of repair, respectively at any time t are given as $S_1(t) \in \{1,2,...,k\}$, $S_2(t) \in \{1,2,...,d\}$. The state of the server at any time t is given by-

$$\mathrm{Y(t)} = \begin{cases} 0, & \text{when the server is in idle state} \\ 1, & \text{when the server is busy with essential phase of service} \\ \mathrm{k}', & \text{when the server is busy with k}'\text{optional service, } (2 \le \mathrm{k}' \le \mathrm{k}) \\ \mathrm{k}+1, & \text{when the server is in vacation state} \end{cases}$$

If, $Y(t) \in (2,3,...,k')$, then we define $\delta(t)$ as the elapsed service time. So, the state of the system at any time t can be stated as $Z(t) = (Y(t), S_1(t), S_2(t), N(t), \delta(t))$.

- **Modified Service Time**

The modified service time, as introduced by Keilson and Servi (1986) can be defined as the time required for completing a service cycle. It is defined as the time taken by a customer from the start of its service until the moment he exits from the system. Thus, the modified service time includes the service time, repair time and vacation time for the model under consideration. For our model, we define the modified service time B as:

$$B = \begin{cases} H_1 & \text{with probability } \bar{p}_1\bar{\sigma}_1 \\ H_1 + V_1 & \text{with probability } \bar{p}_1\sigma_1 \\ H_1 + \sum_{i=2}^{k'} H_i & \text{with probability } \prod_{i=1}^{k'-1} p_i\bar{\sigma}_{k'} \quad (2 \le k' \le k) \\ H_1 + \sum_{i=2}^{k'} H_i + V_{k'} & \text{with probability } \prod_{i=1}^{k'-1} p_i\sigma_{k'} \quad (2 \le k' \le k) \end{cases}$$

Here k' is the last optional service availed by the customer starting continuously from 2 to k' $(2 \le k' \le k)$. Here, H_i $(1 \le i \le k)$ is the cdf for the i^{th} generalized service time, $V_{k'}$ $(2 \le k' \le k)$ corresponds to the cdf for vacation after availing k'^{th} optional service.

7.3 EMBEDDED MARKOV CHAIN

Let n^{th} service completion or departure occurs at any time instant say τ_n and $C(\tau_n-) = C_n$ be the state of the server before the time instant τ_n. Also, let $N(\tau_n+) = N_n$

denotes the number of retrying customers present in the system just after τ_n. Also, we have the sequence $N_n=N(\tau_n+)$ which is embedded Markov renewal process corresponding to continuous time Markov process Z(t). Then we have the following transition-

$$(N_n / N_{n-1}=J) = \begin{cases} J-1+B_n, & \text{with probability } \dfrac{\gamma_J}{\bar{\lambda}+\gamma_J} \\ \\ J+W_n\text{-}1\text{+}B_n, & \text{with probability } \dfrac{\bar{\lambda}}{\bar{\lambda}+\gamma_J} \end{cases} \tag{7.2}$$

where, B_n is the number of customers that arrive during the n[th] modified service time and W_n is the number of customers allowed to join the system if the n[th] customer proceeds from a batch arrival. Also, $\bar{\lambda}=\lambda(1-a_0)$ due to Bernoulli admission mechanism involved in our model.

7.3.1 Ergodicity of the Embedded Markov Chain

Now, we proceed to determine the condition that lead to a stationary process being ergodic. We know that a markov chain is said to be ergodic if it is possible to go from every state to every other state, may be in one move or more than one moves or if a single realization of the process can make all the conclusions about the probability law that generate the process. To check the ergodicity of the markov chain N_n, we have the following theorem:

Theorem 7.1: The embedded markov chain is ergodic iff

$$\rho < \frac{\bar{\lambda}+\gamma_J}{\left[\bar{\lambda}+\gamma_J+\left[h_1\bar{p}_{1\bar{\sigma}_1}+v_1\bar{p}_{1\sigma_1}+\sum_{k'=2}^{k}\left(h_{k'}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}\right)+\sum_{k'=2}^{k}\left(h_{k'}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}\right)\right]^{-1}\right]}$$

Proof: In order to investigate the ergodicity of the sequence N_n, we use classical theory of Lyapunov function (see Artalejo and Gomez Corral, 1997; Sennott *et al.*, 1983) which is based on mean drift criteria. Further, we employ Foster's criterion (cf. Pakes, 1969), according to which an irreducible and aperiodic Markov chain N_n with state space Z^+ is ergodic if there exists a non negative function f(n), $n \in Z^+$, ε>0 and if the mean drift

$$\phi_n = E[N_{n+1}-N_n \mid N_n = n],$$

is finite for all $n \in Z^+$ except perhaps a finite number.

Let us choose f (n) = n as test function or Lyapunov function. Then we have,

$$\phi_n = E[-1+B_{n+1}]\frac{\gamma_J}{\overline{\lambda}+\gamma_J} + E[-1+B_{n+1}+V_{n+1}]\frac{\overline{\lambda}}{\overline{\lambda}+\gamma_J}$$

$$= (-1+\rho) + \frac{\lambda a}{\overline{\lambda}+\theta_J}$$

For $\phi_n \le -\varepsilon$, we have the following stability conditions-

(i) If $\nu > 0$ and $\gamma = 0$, then

$$\rho < \frac{\overline{\lambda}+\gamma_J}{\left[\overline{\lambda}+\gamma_J + \left[h_1\overline{p}_{1\overline{\sigma}_1} + v_1\overline{p}_{1\sigma_1} + \sum_{k'=2}^{k}\left(h_{k'}\left(\prod_{i=1}^{k'-1} p_i\right)\overline{\sigma}_{k'}\right) + \sum_{k'=2}^{k}\left(h_{k'}\left(\prod_{i=1}^{k'-1} p_i\right)\sigma_{k'}\right)\right]^{-1}\right]}$$

(ii) If $\nu \ge 0$ *and* $\gamma > 0$, then $\rho < 1$.

Therefore, it can be said that the embedded markov chain $\{N_n : n \in Z^+\}$ is ergodic. Also, sequence $\{N_n\}_{n=1}^{\infty}$ is positive recurrent so as to guarantee that the limiting probabilities $\pi_J = \lim_{n\to\infty} P\{N_n = J\}$, J ≥0 exists and are positive.

7.3.2. Limiting Distribution

Let us assume that e_{i,n_i}, r_{i,n_i} $(1 \le i \le k)$ are the probabilities of admission of n_i customers in the system during i^{th} $(1 \le i \le k)$ phase service time and during the vacation after the completion of i^{th} $(1 \le i \le k)$ phase service, respectively. Therefore, we have

$$e_{i,n_i} = \begin{cases} \int_0^\infty \exp(-\overline{\lambda}t)dH_i(t), & \text{if } n_i = 0 \\ \sum_{n_i=1}^{\infty}\int_0^\infty \frac{\exp(-\lambda t)(\lambda t)^{n_i} a_m^{(n_i)} dH_i(t)}{n_i!}, & \text{if } n_i \ge 1 \end{cases} \quad (7.3)$$

$$\text{and, } r_{i,n_i} = \begin{cases} \int_0^\infty \exp(-\overline{\lambda}t)dV_i(t), & \text{if } n_i = 0 \\ \sum_{n_i=1}^{\infty}\int_0^\infty \frac{\exp(-\lambda t)(\lambda t)^{n_i} a_m^{(n_i)} dV_i(t)}{n_i!}, & \text{if } n_i \ge 1 \end{cases} \quad (7.4)$$

Here $\{a_m^{(n)}\}_{m=0}^{\infty}$ denotes the *n*-fold convolution of the sequence $\{a_m\}_{m=0}^{\infty}$. Let us assume that the 'n' customers arrive with probabilities $e_n^{(1)} = e_{1,n}$, $e_n^{(k')}$, $l_n^{(1)}$ and $l_n^{(k')}$ during the time intervals (H_1), ($H_1+\sum_{i=2}^{k'} H_i$), (H_1+V_1), and ($H_1+\sum_{i=2}^{k'} H_i + V_{k'}$), respectively s.t. $(2 \le k' \le k)$, then $\sum_{i=1}^{k} n_i = n$. Now, the required probabilities are determined using $e_n^{(k')} = \sum_{0 \le n_1+n_2+...+n_{k'} \le n} e_{1,n_1} ....e_{k',n_{k'}}, (1 \le k' \le k)$,

$l_n^{(1)} = \sum_{m=0}^{n} e_m^{(1)} r_{1,n-m}$; $l_n^{(k')} = \sum_{m=0}^{n} e_m^{(k')} r_{k',n-m}$ $\left(2 \le k' \le k\right)$.

7.3.3 Transition matrix

The one step transition matrix is P= (p_{uv}), where p_{uv} =Pr (N_{n+1}=v/ N_n=u) associated with Markov chain $\{N_n\}_{n=1}^{\infty}$ is obtained as

$$p_{uv} = \begin{cases} \dfrac{\gamma_J}{\lambda+\gamma_J}\left[\bar{p}_1\bar{\sigma}_1 e_0^{(1)} + \bar{p}_1\sigma_1 l_0^{(1)} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\bar{\sigma}_{k'} e_0^{(k')} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\sigma_{k'} l_0^{(k')}\right], \\ \qquad\qquad \text{if } u \ge 1,\ v = u\text{-}1 \\ \dfrac{\bar{\lambda}}{\bar{\lambda}+\gamma_J}\sum_{n=1}^{v-u+1} a_n \left[\bar{p}_1\bar{\sigma}_1 e_{v-u+1-n}^{(1)} + \bar{p}_1\sigma_1 l_{u-v+1-n}^{(1)} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\bar{\sigma}_{k'} e_{v-u+1-n}^{(k')} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\sigma_{k'} l_{v-u+1-n}^{(k')}\right] \\ \quad + \dfrac{\gamma_J}{\bar{\lambda}+\gamma_J}\left[\bar{p}_1\bar{\sigma}_1 e_{v-u+1}^{(1)} + \bar{p}_1\sigma_1 l_{v-u+1}^{(1)} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\bar{\sigma}_{k'} e_{v-u+1}^{(k')} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\sigma_{k'} l_{v-u+1}^{(k')}\right], \\ \qquad\qquad \text{if } 0 \le u \le v \end{cases} \tag{7.5}$$

Now, using π= π P, the Kolmogorov equation associated with Markov chain can be written as-

$$\pi_v = \sum_{n=0}^{v} \frac{\bar{\lambda}\pi_n}{\bar{\lambda}+\gamma_J} \sum_{u=1}^{v-n+1} a_u \left[\bar{p}_1\bar{\sigma}_1 e_{v-n+1-u}^{(1)} + \bar{p}_1\sigma_1 l_{v-n+1-u}^{(1)} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\bar{\sigma}_{k'} e_{v-u+1-n}^{(k')} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\sigma_{k'} l_{v-u+1-n}^{(k')}\right]$$
$$+ \sum_{u=1}^{v+1} \frac{\gamma_u\pi_u}{\bar{\lambda}+\gamma_u}\left[\bar{p}_1\bar{\sigma}_1 e_{v-u+1}^{(1)} + \bar{p}_1\sigma_1 l_{v-u+1}^{(1)} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\bar{\sigma}_{k'} e_{v-u+1}^{(k')} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\sigma_{k'} l_{v-u+1}^{(k')}\right], \quad v \ge 0 \tag{7.6}$$

Now, we multiply equation (7.6) by z^v and then summing it over v ≥ 0, we get

$$\sum_{v=0}^{\infty} \pi_v z^v = \sum_{v=0}^{\infty}\sum_{n=0}^{v} \frac{\bar{\lambda}\pi_n z^n}{\bar{\lambda}+\gamma_J} \sum_{u=1}^{v-n+1} z^u a_u$$
$$\times\left[\bar{p}_1\bar{\sigma}_1 e_{v-n+1-u}^{(1)} + \bar{p}_1\sigma_1 l_{v-n+1-u}^{(1)} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\bar{\sigma}_{k'} e_{v-u+1-n}^{(k')} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\sigma_{k'} l_{v-u+1-n}^{(k')}\right] z^{v-n+1-u} z^{-1}$$
$$+ \sum_{n=0}^{\infty}\sum_{u=1}^{n+1} \frac{\gamma_u\pi_u z^{u-1}}{\bar{\lambda}+\gamma_u}\left[\bar{p}_1\bar{\sigma}_1 e_{v-u+1}^{(1)} + \bar{p}_1\sigma_1 l_{v-u+1}^{(1)} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\bar{\sigma}_{k'} e_{v-u+1}^{(k')} + \sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1} p_i\right)\sigma_{k'} l_{v-u+1}^{(k')}\right] z^{v-u+1} \quad v \ge 0 \tag{7.7}$$

7.3.4 Generating Function

In this sub-section, we define generating functions as follows:

$$\Pi(z)=\sum_{n=0}^{\infty}\pi_n z^n\ ,\psi(z)=\sum_{n=0}^{\infty}\frac{\pi_n z^n}{(\bar{\lambda}+\gamma_J)}\ ,\ L_1(z)=\sum_{n=0}^{\infty}l_n^{(1)}z^n\ ,L_{k'}(z)=\sum_{n=0}^{\infty}l_n^{(k')}z^n\ ,$$

$$E_1(z)=\sum_{n=0}^{\infty}e_n^{(1)}z^n\ ,\ E_i(z)=\sum_{n=0}^{\infty}(e_{i,n_i})z^n\ ,\ E_{(k')}(z)=\sum_{n=0}^{\infty}e_n^{(k')}z^n=\prod_{i=1}^{k'}E_i(z)\ ,\ \left(1\le k'\le k\right)$$

$$P_i(z)=\sum_{n=0}^{\infty}P_{i,n}z^n\ ,R_i(z)=\sum_{n=0}^{\infty}r_n^i z^n\quad (1\le i\le k)\ ,\ L_1(z)=E_1(z)R_1(z),L_{k'}(z)=E_{(k')}(z)R_{k'}(z)$$

Using (7.3), we have $E_i(z)=\tilde{H}_i\left(\lambda\left(1-a(z)\right)\right),\ \left(1\le k'\le k\right)$. Also, using (7.1), we have

$$E_i(z)=\tilde{b}_i(A_i(z))\ \text{ s.t. }\ A_i(z)=(\lambda(1-a(z))+\alpha_i(1-\prod_{j=1}^{d}\tilde{g}_{ij}(\bar{\lambda}(1-a(z))))),\ \left(1\le i\le k\right)$$

Similarly, $R_i(z)=\tilde{V}_i(\lambda(1-a(z)),$

$$L_1(z)=\tilde{b}_1(A_1(z))\tilde{V}_1(\lambda(1-a(z))\ \text{ and }\ L_{k'}(z)=\left(\prod_{i=1}^{k'}\tilde{b}_i(A_i(z))\right)\tilde{V}_{k'}(\lambda(1-a(z))\ ,\ \left(2\le k'\le k\right)$$

Now, consider

$$\Pi(z)=\sum_{n=0}^{\infty}\pi_n z^n=\sum_{n=0}^{\infty}\left(\frac{\bar{\lambda}+\gamma_J}{\bar{\lambda}+\gamma_J}\right)\pi_n z^n \tag{7.8}$$

We consider the case of linear retrial policy, as for ν=0 the obtained results reduces to that for classical retrial policy and for γ =0, we can get results for constant retrial policy. Therefore, we have

$$\gamma_J=\nu(1-\delta_{n,0})+J\gamma\ \text{ and }\ \psi(z)=\sum_{n=0}^{\infty}\frac{\pi_n z^n}{(\bar{\lambda}+\gamma_J)} \tag{7.9}$$

Following Choudhury (2009) equation (7.8) reduces to

$$\Pi(z)=\bar{\lambda}\sum_{J=0}^{\infty}\frac{\pi_J z^J}{(\bar{\lambda}+\gamma_J)}+\sum_{J=0}^{\infty}\frac{\pi_J z^J}{(\bar{\lambda}+\gamma_J)}(\nu(1-\delta_{J,0})+J\gamma) \tag{7.10}$$

$$=(\bar{\lambda}+\nu)\psi(z)+\gamma z\psi'(z)-\frac{\nu\pi_0}{\bar{\lambda}} \tag{7.11}$$

This can be written as

$$\Pi(z)=\left[\frac{\lambda\psi(z)\left(a(z)-a_0\right)}{z}+\frac{\nu}{z}\psi(z)-\frac{\nu\pi_0}{z\bar{\lambda}}+\gamma\psi'(z)\right]Q(z)$$

$$=\frac{Q(z)}{z}\left(\gamma z\psi'(z)+\psi(z)\{\bar{\lambda}+\nu-\lambda(1-a(z)\}-\nu\pi_0\bar{\lambda}^{(-1)}\right) \tag{7.12}$$

where,

$$Q(z)=\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\bar{\sigma}_{k'}\right)E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right] \tag{7.13}$$

Solving (7.8) and (7.12), we get

$$\gamma z\psi'(z)+\psi(z)\left[(\bar{\lambda}+\nu)-\frac{\lambda(1-a(z))}{Q(z)-z}Q(z)\right]=\nu\pi_0\bar{\lambda}^{(-1)} \tag{7.14}$$

$$\prod(z)=\frac{\lambda(1-a(z))\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]}{\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_0(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]-z}\psi(z) \tag{7.15}$$

Computation of $\prod(1)$

Using the arguments of normalizing condition i.e. the sum of probabilities as unity, we have $\prod(1)=1$. Hence, using normalizing condition in equation (7.15) and applying L-Hospital's rule, we find

$$\prod(1)=\frac{(-\lambda a'(1)Q(1)\psi(1))}{Q'(1)-1}=1 \tag{7.16}$$

$$\psi(1)=\frac{Q'(1)-1}{-\lambda a'(1)}=\frac{1-\rho}{\lambda a} \tag{7.17}$$

where, $a=\zeta E[X]$,

$$\rho=\lambda\zeta E(X)\left[h_1\bar{p}_1\bar{\sigma}_1+v_1\bar{p}_1\sigma_1+\sum_{k'=2}^{k}\left(h_{k'}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}\right)+\sum_{k'=2}^{k}\left(h_{k'}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}\right)\right] \tag{7.18}$$

Theorem 7.2: The probability generating function $\psi(z)$ for the stationary queue size distribution for (i) Linear retrial policy (ν >0, γ >0) (ii) Classical retrial policy (ν=0, γ >0) and (iii) Constant retrial policy (ν> 0 and γ =0) are respectively, given by

$$\text{(i)}\,\psi(z)=z^{-\tau_1}e^{\left\{\lambda/\gamma\int_z^1(\Omega(y)/y)dy\right\}}\left[\tau_2+\nu\pi_0\bar{\lambda}\gamma^{-1}\int_z^1 x^{(\tau_1-1)}e^{\left\{(-\lambda/\gamma)\int_x^1(\Omega(y)/y)dy\right\}}dx\right] \tag{7.19}$$

$$\pi_0=\tau_2\gamma\bar{\lambda}\nu^{-1}\left[\int_0^1 x^{(\tau_2-1)}e^{\left\{(-\lambda/\gamma)\int_x^1(\Omega(y)/y)dy\right\}}dx\right]^{-1} \tag{7.20}$$

$$\text{(ii)}\ \psi(z)=\tau_2\exp\left\{(\gamma^{-1})\int_z^1\left(\bar{\lambda}-\lambda\Omega(y)\right)\frac{dy}{y}\right\},\tag{7.21a}$$

$$\pi_0=\tau_2\lambda\left[\exp\left\{(-\lambda/\gamma)\int_0^1\left(\bar{\lambda}-\lambda\Omega(y)\right)\frac{dy}{y}\right\}\right]\tag{7.21b}$$

$$\text{(iii)}\ \psi(z)=\frac{\nu\pi_0}{\bar{\lambda}\left[\left(\bar{\lambda}+\nu\right)-\dfrac{\lambda(1-a(z)Q(z))}{(Q(z)-z)}\right]},\tag{7.22a}$$

$$\pi_0=(1-a_0)\lambda\tau_2\left(1+\bar{\lambda}\nu^{-1}\right)-\bar{\lambda}\nu^{-1}\tag{7.22b}$$

Proof: The probability generating function can be obtained by solving the linear differential equation (7.14) under different retrial schemes. We initially consider the case for linear retrial policy. On solving differential equation (7.14), we get the generating function of the stationary queue size distribution given by (7.19). Further, by putting z=0 in equation (7.19), we get the value of π_0.

For classical retrial policy, the differential equation (7.14) reduces to homogenous equation and directly gives the result as given by (7.21a). The value of π_0 as specified in equation (7.21b) for classical retrial policy is obtained by using the relation $\pi_0=\lambda\psi(0)$. Now we turn our attention for the constant retrial policy. For this case we set γ =0 in equation (7.14) and get

$$\psi(z)\left[(\bar{\lambda}+\nu)-\frac{\lambda(1-a(z)}{Q(z)-z}Q(z)\right]=\nu\pi_0\bar{\lambda}^{(-1)}\tag{7.22c}$$

On further simplifying (7.22c), we get result (7.22a). The value of π_0 as given in equation (7.22b) is obtained by taking the limit z$\to$1 in equation (7.22a) and then using L-hospital rule.

7.4. JOINT DISTRIBUTION AND EMBEDDED MARKOV RENEWAL PROCESS

In order to investigate the joint distribution of the state of the server and the number of customers in the retrial group, we employ Markov regenerative process (MRGP). The similar approach is used by Jain *et al.* (2012b); Choudhury (2013), and many more to analyze queueing models in different frameworks. MRGP is a continuous-time stochastic process associated with more general conditions of regeneration than regenerative processes and can be used for determining the limiting behavior of some queues with general distributions. Let us assume that the limiting distribution of the system and the number in the retrial group is

$P_{l,s} = \lim_{t\to\infty} \Pr\{Z(t)=(l,s)\}, (l,s) \in \{Y(t), N(t)\}$.

It is evident from Burke's theorem (cf. Cooper, 1981) that stationary probabilities $\left(P_{l,s} : l \in Y(t), s \in N(t)\right)$ exist and are positive for embedded Markov chain. For brevity, we consider the following notations:

$$\Omega(y) = \frac{(1-a(y))Q(y))}{(Q(y)-y)}, \tau_1 = (\bar{\lambda}+\nu)/\theta \text{ and } \tau_2 = \frac{(1-\rho)}{\lambda\zeta E(X)} \tag{7.23}$$

7.4.1 Embedded Markov Renewal Process

In order to determine the limiting distribution we use some existing results established by Cinlar (1975). Let us assume that,

$\phi_n(l,s)$ = Expected time spent by the process $\{Z(t), t \ge 0\}$ in the state *(l, s)* during a service cycle.

ϕ_n = Expected length of the service cycle where service cycle can be referred as the span of time between two consecutive completion epochs by considering the fact that n customers are present in the retrial orbit at the beginning of this interval.

Therefore, $$P_{l,s} = \frac{\sum_{n=0}^{\infty} \pi_n \phi_n(l,s)}{\sum_{n=0}^{\infty} \pi_n \phi_n} \tag{7.24}$$

Further, for model under consideration, ϕ_n is obtained as:

$$\phi_n = \frac{1}{\bar{\lambda}+\gamma_n} + \frac{\rho}{\lambda a}, n \ge 0$$

and the mean service cycle is given by

$$\sum_{n=0}^{\infty} \pi_n \phi_n = (\lambda a)^{-1} \tag{7.25}$$

Theorem 7.3: The limiting probabilities for the queue size distribution are obtained as follows:

$$P_{0,s} = [\lambda\zeta E(X)]\sum_{n=0}^{\infty}\frac{1}{\bar{\lambda}+\gamma_n}\delta_{s,n}\pi_n \tag{7.26}$$

$$P_{1,s} = [\lambda\zeta E(X)]\left[\sum_{n=0}^{s}\left(\frac{\lambda\pi_n}{\bar{\lambda}+\gamma_n}(1-\delta_{s,n-1})\sum_{m=1}^{s-n+1}a_m f_{1,s-n+1-m} + \sum_{n=1}^{s+1}\frac{\gamma_n\pi_n}{\bar{\lambda}+\gamma_n} f_{1,s-n+1}\right)\right]\pi_n \tag{7.27}$$

$$P_{k',s} = \left[\lambda\zeta E(X)\right](\prod_{i=1}^{k'-1} p_i)\left(\frac{\lambda\pi_n}{\bar{\lambda}+\gamma_n}(1-\delta_{s,n-1})\sum_{m=1}^{s-n+1}a_m\sum_{i=1}^{s-n+1-m}e_i^{(k'-1)}f_{k',s-n+1-i-m}\right)$$
$$+\left[\lambda\zeta E(X)\right](\prod_{i=1}^{k'-1} p_i)\left(\sum_{n=1}^{s+1}\frac{\gamma_n\pi_n}{\bar{\lambda}+\gamma_n}(\prod_{i=1}^{k'-1} p_i)\sum_{i=1}^{s-n+1}e_i^{(k'-1)}f_{k',s-n+1-i}\right), n\ge 0, s\ge 0, (2\le k'\le k) \tag{7.28}$$

$$P_{k+1,s} = (\lambda\zeta E(X))[\frac{\lambda}{\bar{\lambda}+\gamma_n}\bar{p}_1\sigma_1(1-\delta_{s,n-1})\sum_{m=1}^{s-n+1}a_m\sum_{i=1}^{s-n+1-m}e_{1,i}u_{1,s-n+1-i-m}$$
$$+\bar{p}_1\sigma_1\frac{\gamma_n}{\bar{\lambda}+\gamma_n}\sum_{i=1}^{s-n+1}e_{1,i}u_{1,s-n+1-i} + \frac{\lambda}{\bar{\lambda}+\gamma_n}(\prod_{i=1}^{k'-1}p_i\sigma_{k'})(1-\delta_{s,n-1})\sum_{m=1}^{s-n+1}a_m\sum_{i=1}^{s-n+1-m}e_i^{(k'-1)}u_{k+1,s-n+1-i-m} \tag{7.29}$$
$$+\sum_{n=1}^{s+1}\frac{\gamma_n}{\bar{\lambda}+\gamma_n}(\prod_{i=1}^{k'-1}p_i\sigma_{k'})\sum_{i=1}^{s-n+1}e_i^{(k'-1)}u_{k+1,s-n+1-i}]\pi_n, n\ge 0, s\ge 0, (2\le k'\le k)$$

where, $f_{i,n} = \begin{cases}\int_0^{\infty} e^{-\bar{\lambda}t}(1-H_i(t))dt, & n=0\\ \sum_{n=1}^{\infty}\int_0^{\infty}\frac{\exp(-\lambda t)(\lambda t)^n a_m^{(n)}(1-H_i(t))dt}{n!}, n\ge 1, & (1\le i\le k)\end{cases}$

and,

$$u_{i,n} = \begin{cases}\int_0^{\infty} e^{-\bar{\lambda}t}(1-V_i(t))dt, & n=0\\ \sum_{n=1}^{\infty}\int_0^{\infty}\frac{\exp(-\lambda t)(\lambda t)^n a_m^{(n)}(1-V_i(t))dt}{n!}, n\ge 1, & (1\le i\le k)\end{cases}$$

Proof: The limiting probabilities for the various states of the server can be obtained in the following manner:

Case 1: When Y(t) = 0, and N(t)=s.

In this case the server is in idle state with *s* customers in the retrial orbit at any time t and n retrial customers at the beginning of this interval. Then, we have

$$\phi_n(0,s) = \frac{1}{\bar{\lambda}+\gamma_n}\delta_{s,n}, n\ge 0, s\ge 0 \tag{7.30}$$

where, $\delta_{s,n} = \begin{cases}1, & \text{if } s=n\\ 0, & \text{if } s\ne n.\end{cases}$

Using, (7.30) in (7.24), we have

$$P_{0,s} = [\lambda\zeta E(X)]\sum_{n=0}^{\infty}\frac{1}{\bar{\lambda}+\gamma_n}\delta_{s,n}\pi_n$$

Case 2: When Y(t)=1.

In this case, the server is busy with the essential generalized service of the customer. Thus time spent by the process in the state (1, s) is given by

$$\phi_n(1,s) = \frac{\lambda}{\bar{\lambda}+\gamma_n}(1-\delta_{s,n-1})\sum_{m=1}^{s-n+1}a_m f_{1,s-n+1-m} + \sum_{n=1}^{s+1}\frac{\gamma_n}{\bar{\lambda}+\gamma_n}f_{1,s-n+1}, n\geq 0, s\geq 0 \quad (7.31)$$

Therefore, using (7.31) in (7.24), we get (7.27).

Case 3: When $\mathrm{Y(t)} = k'$; $(2\leq k'\leq k)$

During this state, the server provides optional services to the customer. A customer undergoing service can utilize k' $(2\leq k'\leq k)$ successive optional services. Then, the expected amount of time spent by the Markov process is:

$$\begin{aligned}\phi_n(k',s) &= \frac{\lambda}{\bar{\lambda}+\gamma_n}(\prod_{i=1}^{k'-1}p_i)(1-\delta_{s,n-1})\sum_{m=1}^{s-n+1}a_m\sum_{i=1}^{s-n+1-m}e_i^{(k'-1)}f_{k',s-n+1-i-m}\\ &+\sum_{n=1}^{s+1}\frac{\gamma_n}{\bar{\lambda}+\gamma_n}(\prod_{i=1}^{k'-1}p_i)\sum_{i=1}^{s-n+1}e_i^{(k'-1)}f_{k',s-n+1-i}, n\geq 0, s\geq 0\end{aligned} \quad (7.32)$$

and, thus using (7.32) in (7.24), we have (7.28).

Case 4: When $\mathrm{Y(t)} = k+1$.

The server undergoes vacation during this state and returns back after a random interval of time to serve the next customer if any. Here, we have

$$\begin{aligned}\phi_n(k+1,s) &= \frac{\lambda}{\bar{\lambda}+\gamma_n}\bar{p}_1\sigma_1(1-\delta_{s,n-1})\sum_{m=1}^{s-n+1}a_m\sum_{i=1}^{s-n+1-m}e_{1,i}u_{1,s-n+1-i-m} + \bar{p}_1\sigma_1\frac{\gamma_n}{\bar{\lambda}+\gamma_n}\sum_{i=1}^{s-n+1}e_{1,i}u_{1,s-n+1-i}\\ &+\frac{\lambda}{\bar{\lambda}+\gamma_n}(\prod_{i=1}^{k'-1}p_i\sigma_{k'})(1-\delta_{s,n-1})\sum_{m=1}^{s-n+1}a_m\sum_{i=1}^{s-n+1-m}e_i^{(k'-1)}u_{k+1,s-n+1-i-m}\\ &+\sum_{n=1}^{s+1}\frac{\gamma_n}{\bar{\lambda}+\gamma_n}(\prod_{i=1}^{k'-1}p_i\sigma_{k'})\sum_{i=1}^{s-n+1}e_i^{(k'-1)}u_{k+1,s-n+1-i}, n\geq 0, s\geq 0\end{aligned} \quad (7.33)$$

Therefore, limiting distribution (7.29) for vacation state can be obtained by solving (7.24) and (7.33).

Theorem 7.4: The generating functions for the various states of the server are expressed as:

$$P_0(z) = [\lambda\zeta E(X)]\psi(z) \quad (7.34)$$

$$P_1(z)=\frac{\left[1-\tilde{b}_1(A_1(z)\right]P_0(z)}{Q(z)-z} \tag{7.35}$$

$$P_{k'}(z)=\frac{(\prod_{i=1}^{k'-1}p_i)(\prod_{i=1}^{k'-1}E_i(z))(1-\tilde{b}_{k'}(A_{k'}(z)))P_0(z)}{Q(z)-z},\ (2\le k'\le k) \tag{7.36}$$

$$P_{k+1,b}(z)=\frac{P_0(z)}{Q(z)-z}$$
$$\times\left[\tilde{b}_1(A_1(z))(1-\tilde{V}_1(\lambda(1-a(z)))\bar{p}_1\sigma_1+\left(\prod_{i=1}^{k'-1}p_i\sigma_{k'}\right)\left(\prod_{i=1}^{k'}\tilde{b}_i\left(A_i(z)\right)\right)\left(1-\tilde{V}_{k'}(\lambda(1-a(z)))\right)\right] \tag{7.37}$$

Proof: We define $P_l(z)=\sum_{s=0}^{\infty}P_{l,s}z^s \quad \forall l\in 1,2,...,k+1.$

Also, $F_i(z)=\frac{1-\tilde{b}_i(A_i(z))}{\lambda(1-a(z))}$ and $U_i(z)=\frac{1-\tilde{V}_i(\lambda(1-a(z)))}{\lambda(1-a(z))}$.

Now, the partial generating functions for various states can be obtained by multiplying equations (7.26)-(7.29) by required powers of z followed by summation of s for s ≥0. Thus, we get (7.34)-(7.37).

Theorem 7.5: The generating function defining probability of the number of customers in the orbit is given by

$$K_o(z)=\frac{(1-z)\left[\lambda\zeta E(X)\right]\psi(z)}{\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\bar{\sigma}_{k'}\right)E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]-z}$$

Proof: The generating function for the number of customers in the orbit is obtained by using

$$K_o(z)=P_0(z)+P_1(z)+\sum_{k'=2}^{k}P_{k'}(z)+P_{k+1}(z),\ (2\le k'\le k).$$

Theorem 7.6: The generating function defining probability of the number of customers in the system is given by

$$K_s(z)=\frac{(1-z)\left[\lambda\zeta E(X)\right]\psi(z)\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]}{\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]-z}$$

Proof: It can be obtained by using $K_s(z)=P_0(z)+z\left[P_1(z)+\sum_{k'=2}^{k}P_{k'}(z)+P_{k+1}(z)\right],\ (2\le k'\le k).$

7.5 STOCHASTIC DECOMPOSITION

The stochastic decomposition property of the queue with vacation has been studied by a number of researchers (Yang and Templeton, 1987; Fuhramann, 1985). Stochastic decomposition for retrial queues with vacation can be established easily. According to this decomposition, the generating function for the queue length of the system for the retrial queues with vacations can be stochastically decomposed as the product of the stationary distribution of retrial queues without vacations and additional term due to vacations. In this section, we present the stochastic decomposition of retrial queues with multi-optional services with Bernoulli vacations and multi-essential repair. From theorem 7.6, we have

$$K_s(z) = \frac{(1-z)\left[\lambda\zeta E(X)\right]\psi(z)\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]}{\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]-z} \tag{7.38}$$

The above expression for the generating function of the system size can also be written as (cf. Choudhury and Deka, 2013):

$$K_s(z) = T_1(z)\times T_2(z) \tag{7.39}$$

where,

$$T_1(z) = \left(\frac{(1-z)(1-\rho)\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]}{\left[\bar{p}_1\bar{\sigma}_1E_1(z)+\bar{p}_1\sigma_1L_1(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\bar{\sigma}_{k'}E_{(k')}(z)+\sum_{k'=2}^{k}\left(\prod_{i=1}^{k'-1}p_i\right)\sigma_{k'}L_{k'}(z)\right]-z}\right)$$

and $T_2(z) = \left(\frac{\left[\lambda\zeta E(X)\right]\psi(z)}{(1-\rho)}\right)$

Here, $T_1(z)$ represents the system size distribution of bulk arrival non-markovian retrial queue equipped with phase service and Bernoulli admission mechanism under Bernoulli vacation schedule. The second term, $T_2(z)$ gives the distribution of the average number of customers in the retrial group under the condition that the server is in idle state.

7.6 PERFORMANCE INDICES

In the present section, we derive some performance measures of interest in theorem form as follows:

(a) Long Run Probabilities

The long run probabilities are given in the following theorem:

Theorem 7.7: The long run probabilities of the server in various states are

(i) When the server is in idle state:

$$P_I = (1-\rho) \tag{7.40}$$

(ii) When the server is busy with first essential phase of generalized service:

$$P_E = \lambda\zeta E(X) b_1 (1+\alpha_1 \sum_{j=1}^{d} g_{ij}) \tag{7.41}$$

(iii) When the server is busy with k', (2≤ k'≤ k) optional phase of generalized service:

$$P_{O_{k'}} = \left(\lambda\zeta E(X) \left(\prod_{i=1}^{k'-1} p_i \right) b_{k'} (1+\alpha_{k'} \sum_{j=1}^{d} g_{k'j}) \right); \quad (2 \le k' \le k) \tag{7.42}$$

(iv) When the server is under vacation state:

$$P_v = \lambda\zeta E(X) \left(\bar{p}\sigma_1 v_1 + (\prod_{i=1}^{k'-1} p_i)\sigma_{k'} v_{k'} \right); \quad (2 \le i \le k) \tag{7.43}$$

Proof: The long run probabilities can be obtained by using

$$P_I = \lim_{z\to 1} P_0(z), P_E = \lim_{z\to 1} P_1(z),$$
$$P_{O_{k'}} = \lim_{z\to 1} P_{k'}(z) \ (2 \le k' \le k), P_V = \lim_{z\to 1} P_{k+1}(z)$$

(b) Average Queue Length and Average Waiting Time

The average queue lengths of the number of customers in the retrial orbit and in the system are established in the following theorems:

Theorem 7.8: The average queue length of the customers in the retrial orbit E [N_o] and average queue length of the system E [N_s] are given by

$$\text{E}\,[\text{N}_\text{o}] = \left[\frac{\zeta E(X^2)}{E(X)} + \frac{\lambda\zeta E(X)\psi'(1)}{(1-\rho)} + \frac{Q''(1)}{2(1-\rho)} \right] \tag{7.44}$$

$$\text{E}\,[\text{N}_\text{s}] = \text{E}[\text{N}_o] + \rho \tag{7.45}$$

where,

(i) If $\nu > 0, \gamma > 0$, then

$$\psi'(1) = \left(\lambda\zeta E(X)\gamma\right)^{-1} \left[\nu\left(\lambda\pi_0 E(X) + \rho - 1\right) + \lambda(1-a_0)(\rho-1) + \zeta E(X) \right]$$

(ii) If $\nu = 0, \gamma > 0$, then

$$\psi'(1) = \left(\lambda\zeta E(X)\gamma\right)^{-1} \left[\lambda(1-a_0)(\rho-1) + \zeta E(X) \right]$$

(iii) If $\nu > 0, \gamma = 0$, then

$$\psi'(1) = \frac{\nu\lambda\pi_0}{\bar{\lambda}}\left[\frac{K_3'(1)}{\left[(\bar{\lambda}+\nu)-\lambda K_3(1)\right]^2}\right]$$

$$K_3(1) = \frac{-a'(1)}{\rho-1}$$

$$K_3'(1) = \frac{a''(1)}{2(1-\rho)} + \frac{a'(1)Q''(1)}{2(\rho-1)^2}$$

$$Q''(1) = \bar{p}_1\bar{\sigma}_1E_1''(1) + \bar{p}_1\sigma_1L_1''(1) + \prod_{i=1}^{k'-1} p_i\bar{\sigma}_{k'}E_{(k')}''(1) + \prod_{i=1}^{k'-1} p_i\sigma_{k'}L_{k'}''(1),\ (2 \le k' \le k)$$

$$E_1''(1) = b_1^{(2)}\left[\lambda\zeta E(X)(1+\alpha_1 g_{1j})\right]^2 + b_1 A_1''(1),\ (1 \le j \le d)$$

$$E_{(k')}''(1) = \sum_{i=1}^{k'}\left[b_i^{(2)}\left[\lambda\zeta E(X)(1+\alpha_i\sum_{j=1}^{d} g_{ij})\right]^2 + b_i A_i''(1)\right],\ (2 \le k' \le k)$$

$$L_{k'}''(1) = \sum_{i=1}^{k'}\left(b_i^{(2)}\left[\lambda\zeta E(X)(1+\alpha_i\sum_{j=1}^{d} g_{ij})\right]^2 + b_iA_i''(1)\right) + 2\left(\lambda\zeta E(X)(1+\alpha_i\sum_{j=1}^{d} g_{ij})b_i\right)\left(\lambda\zeta E(X)v_{k'}\right)$$
$$+\left(v_{k'}^{(2)}\left(\lambda\zeta E(X)\right)^2 + v_{k'}\left(-\lambda\zeta^2 E(X^2)\right)\right),(2 \le k' \le k),(1 \le j \le d)$$

$$A_i''(1) = -\lambda\zeta^2 E(X^2) - \alpha_i\left(\lambda^2 E(X)^2 g_{ij}^{(2)} + \zeta^2\lambda E(X^2)g_{ij}^{(1)}\right),\ (1 \le i \le k),\ (1 \le j \le d)$$

Proof: The average queue lengths for the retrial orbit and system are obtained (cf. Choudhury, 2008) as:

$$\mathrm{E}\,[\mathrm{N_o}] = \lim_{z\to 1}\frac{d\mathrm{K}_o(z)}{dz} \ \text{ and }\ \mathrm{E}\,[\mathrm{N}_s] = \lim_{z\to 1}\frac{d\mathrm{K}_s(z)}{dz} = \mathrm{E}\,[\mathrm{N}_o] + \rho$$

Theorem 7.9: The average waiting time (W_s) spend by a customer in the system is

$$\frac{1}{\lambda}\left[\frac{\zeta E(X^2)}{E(X)} + \frac{\lambda\zeta E(X)\psi'(1)}{(1-\rho)} + \frac{Q''(1)}{2(1-\rho)} + \rho\right] \tag{7.46}$$

Proof: The mean time spend by a customer for the service is determined using Little's formula (Gross and Harris, 1985) as:

$$\mathrm{W_s} = E[N_s]/\bar{\lambda}\,.$$

7.6 SPECIAL CASES

In order to validate our model, we deduce some special cases of our study by assigning some specific values to various parameters. Some such cases are listed below:

a) If $\zeta = 1, k = 2, d = 1,$ then our model reduces to model analyzed by Jain *et al.* (2012b).

b) If $\alpha_i = 0, \beta_i = 0, k = 2, d = 1, p_i(1 \le i \le k) = 1, \sigma_i(1 \le i \le k-1) = 0,$ then the present model reduces to model analyzed by Choudhury and Deka (2013).

c) If $E(X) = 1, E(X^2) = 0, k = 2, d = 1, p_i(1 \le i \le k) = 1, \sigma_i(1 \le i \le k-1) = 0, \zeta = 1,$ then the present model coincides with the model studied by Choudhury (2009).

d) If $k = 2, d = 1, p_i(1 \le i \le k) = 1, \sigma_i(1 \le i \le k-1) = 0, \zeta = 1, \upsilon = 0, \alpha_i = 0,$ then our model is same as studied by Choudhury (2007).

e) If $E(X) = 1, E(X^2) = 0, k = 1, d = 1, \zeta = 1, \sigma_i = 0,$ then the model developed provides results obtained by Falin (2010a).

f) If $\alpha_i = 0, \beta_i = 0, k = 2, d = 1, p_i(1 \le i \le k) = 1, \sigma_i(1 \le i \le k-1) = 0, \zeta = 1, E(X) = 1,$ then our model coincides with queueing model studied by Choudhury (2008b).

g) If $k = 1, d = 1, \sigma_i(1 \le i \le k) = 0, p_i(1 \le i \le k) = 0, \alpha_i = 0,$ then our model deduces to the queueing model analyzed by Artalejo and Atencia (2004).

h) If $E(X) = 1, \alpha_i = 0, \beta_i = 0, k = 2, d = 1, p_i(1 \le i \le k) = 1, \sigma_i(1 \le i \le k-1) = 0, \zeta = 1,$ then the model under consideration yields performance measures obtained by Artalejo and Choudhury (2004).

i) If $\alpha_i = 0, \beta_i = 0, k = 1, d = 1, p_i(1 \le i \le k) = 0, \sigma_i(1 \le i \le k-1) = 0, \zeta = 1,$ present model coincides with that studied by Falin (2010b).

7.8 COST FUNCTION

In the present section, we frame the function which provides the analytical formula for the expected total cost (ETC) for the concerned retrial queueing model. The cost function is formulated by considering service rate (μ_1) and admission probability (ζ) as decision variables.

The cost function has been formulated in the following manner as:

$$ETC(\mu_1, \zeta) = C_h L_s + \zeta C_1 + C_{b_1} P_E + C_{b_{k'}} P_{O_{k'}} + C_V P_V + C_I P_I; \;\; (2 \le k' \le k) \qquad (7.48)$$

where,

C_h = Holding cost per unit customer;

C_1 = Fixed cost incurred when a customer is admitted to join the system according to admission policy;

C_{b_1} = Cost per unit time while providing first essential service;

$C_{b_{k'}}$ = Cost per unit time while providing k'^{th} $(2 \le k' \le k)$ optional service;

C_V = Cost per unit time in the system when the server is on vacation;

C_I = Cost per unit time when the customer retry for the service.

Here, our aim is to find out the optimal expected total cost of the system by determining optimal decision variables μ_1* and ζ*. Since the cost function framed is non-linear in nature, so we first check the convexity of the function and then find out corresponding optimal expected cost.

The following procedure has been adopted to find optimal decision variables μ_1* and ζ*:

Step 1: Initialize $\zeta = \zeta_{0.}$

Step 2: Vary μ_1 in predefined range say ($a \le \mu_1 \le b$) and obtain total cost corresponding to different values of μ_1 and fixed $\zeta_{0.}$

Step 3: Search for the optimal value of μ_1 (say μ_1*), such that ETC ($\mu_1{*}^{-}$, ζ) $\le$ ETC (μ_1*, ζ) $\le$ ETC ($\mu_1{*}^{+}$, ζ) is satisfied.

Step 4: Now, fix $\mu_1 = \mu_1$* and vary ζ within the prescribed limits and similar to step 2 find total costs corresponding to different values of ζ and fixed μ_1*.

Step 5: Search for the optimal value of ζ (say ζ*) such that such that ETC (μ_1*, $\zeta{*}^{-}$) $\le$ ETC (μ_1*, $\zeta{*}^{-}$) $\le$ ETC (μ_1*, $\zeta{*}^{-}$) is satisfied.

Step 6: Record (μ_1*, ζ*) and ETC (μ_1*, ζ*), which provide optimal parameters and corresponding optimal cost.

7.9 NUMERICAL RESULTS AND SENSITIVITY ANALYSIS

The efficiency of any retrial queueing model is best deciphered by means of the numerical analysis of derived analytic results. In this section, we provide the sensitivity analysis to examine various performance indices towards variation in different parameters. The numerical tractability of the present generalized queueing model will provide insight for the validation of the real time queueing system in a much better way. The numerical results are obtained by coding computer program in 'MATLAB' software. The batch size has been assumed to follow geometric distribution while retrial process, vacation process and repair process are assumed to be exponentially distributed for numerical purposes. Different distributions namely exponential and gamma are considered for the service time. For illustration, the computational results have been obtained for a total 3 phase services (1 essential and 2 optional) by setting the set of default parameters as:

$$\lambda = 0.5, \alpha_1 = 0.2, \alpha_2 = 0.2, \alpha_3 = 0.2, \beta_1 = 0.8, \beta_2 = 1.2, \beta_3 = 1, \mu_2 = 5, \mu_3 = 4, p_1 = 0.4,$$
$$p_2 = 0.2, p_3 = 0.4, E[X] = 1, \gamma = 0.2, \upsilon = 0.4, \sigma_1 = 0.7, \sigma_1 = 0.4, \sigma_1 = 0.2, C_h = \$5,$$
$$C_1 = \$5, C_{b_1} = \$20, C_{b_2} = \$35, C_{b_3} = \$52, C_V = \$10, C_I = \$50.$$

Figs 7.1-7.2 are plotted to compute the expected total cost and optimal values of μ_1 and ξ for the assumed set of default parameters. Figs 7.1(a-b) depict the model $M^x/M/1$ when service pattern is exponentially distributed service while figs 7.2 (a-b) are plotted for gamma distributed service time. For $M^x/M/1$ model, we proceed as follows:

(i) Fix ζ as 0.4 units and vary μ_1 from 1 to 5 units to search for the corresponding optimal cost as shown in fig. 7.1(a).

(ii) It is noticed that min. ETC (μ_1, ξ) is \$40 at μ_1=2.3 units and ζ =0.4 units so that μ_1*=2.3 and ETC (μ_1*, ξ) = \$40.

(iii) Now, we fix μ_1=2.3 units and vary ζ as 0.01: 0.01: 0.4 units and check the corresponding minimum expected total cost.

(iv) The optimal minimum cost ETC (μ_1, ζ) = \$37.65 which is attained at ζ *=0.2 units when μ_1*=2.3 units as shown in fig. 7.1(b).

(v) Now, (μ_1*, ζ*) = (0.2, 2.3) and ETC (μ_1*, ζ*) = \$ 37.65.

Similarly, we proceed to find out (μ_1*, ζ *) when service time distribution is assumed to follow gamma distribution. In this case minimum expected total cost ETC (μ_1*, ζ*) = \$37.39 is achieved at (μ_1*, ζ*) = (0.2, 0.21). On comparing the cost for two different service time distributions, we claim that the minimum optimal cost is obtained when service time follows exponential distribution.

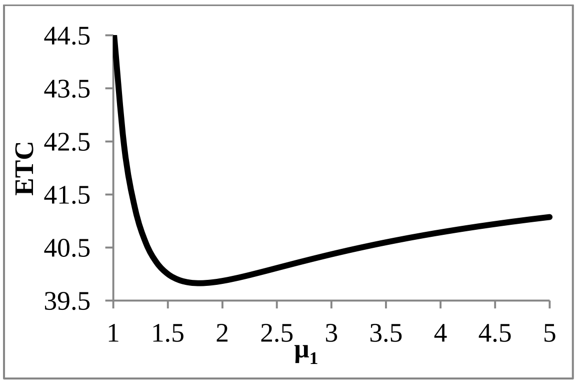


Fig. 7.1(a): ETC (μ_1, ζ) vs. μ_1 for $M^x/M/1$ retrial queue

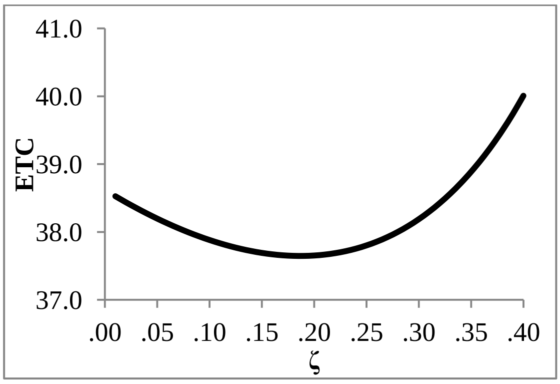


Fig. 7.1(b): ETC (μ_1, ζ) vs. ζ for $M^x/M/1$ retrial queue

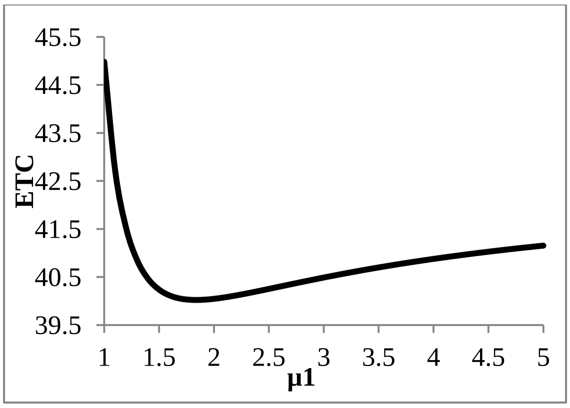


Fig. 7.2(a): ETC (µ$_1$, ζ) vs. µ$_1$ for $M^x/\gamma/1$ retrial queue

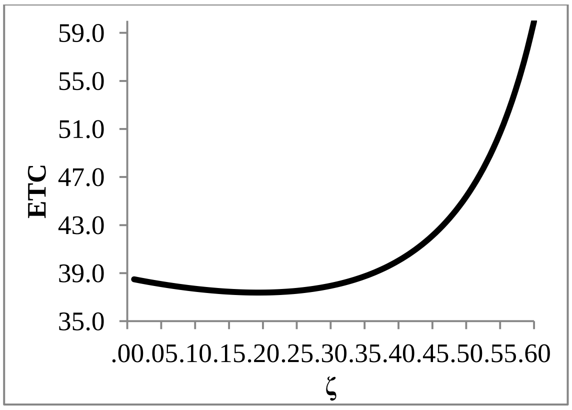


Fig. 7.2(b): ETC (µ$_1$, ζ) vs. ζ for $M^x/\gamma/1$ retrial queue

However, ζ* is higher for gamma distributed service time and on the contrary μ_1* is higher for exponentially distributed service pattern.

The sensitivity of various parameters towards performance indices is detailed in the following subsections. The default parameters for computational purpose are fixed as:

$$\lambda = 0.5, \alpha_1 = 0.2, \alpha_2 = 0.2, \alpha_3 = 0.2, \beta_1 = 0.8, \beta_2 = 1.2, \beta_3 = 1, \mu_1 = 2, \mu_2 = 5, \mu_3 = 4,$$
$$p_1 = 0.4, p_2 = 0.2, p_3 = 0.4, E[X] = 1, \gamma = 0.2, \zeta = 0.6, \upsilon = 0.4, \sigma_1 = 0.7, \sigma_1 = 0.4, \sigma_1 = 0.2.$$

(A) Long Run Probabilities: The sensitivity of long run probabilities and waiting time of the system towards various parameters viz. λ, γ, μ_1, μ_2, σ_1, β_1 for different values of ζ have been displayed by means of tables 7.1-7.3. It is clear from the data given in tables 7.1-7.3 that P_I decreases with an increase in arrival rate (λ), vacation probability (σ_1) while it increases with an increase in service rates (μ_1, μ_2) and repair rate (β_1). On the other hand, the probabilities of the server being busy with essential service (P_E) and being busy with optional services ($P_{o_{k'}}$) increase with an increase in λ, ζ, and their corresponding service rates μ_1 and μ_2 but decreases with an increase in repair rate β_1.

(B) Queue Length: The effects of various parameters on the number of customers in both system and the orbit are displayed by means of figures 7.3-7.4. In figs 7.1(a-c), the queue length of the system $E[N_s]$ and orbit $E[N_o]$ have been plotted against varying values of λ, μ_1, μ_2, σ_1, ζ for $M^x/M/1$ model. The solid lines (—) in the graphs 7.3(a-c) correspond to $E[N_s]$ where as dashed lines (---) are plotted for $E[N_o]$. It is quite clear from the figs 7.3(a-c) that both $E[N_s]$ and E $[N_o]$ increase with an increase in the arrival rate (λ), vacation probability (σ1), admission probability (ζ) as well as with breakdown rate (α_1) which is quite obvious. The effects of various parameters like ζ, β_1 and γ with λ

Table 7.1: Effect of λ, γ and ζ on the long run probabilities

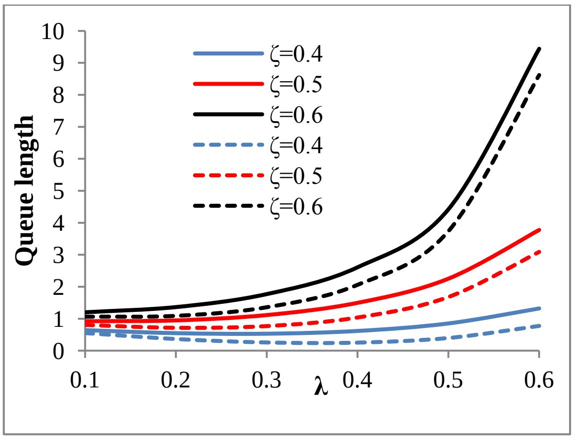
		ζ = 0.6					ζ = 0.4				
λ	γ	P_I	P_E	P_{O3}	P_V	Ws	P_I	P_E	P_{O3}	P_V	Ws
0.5	0.2	0.4944	0.2371	0.0249	0.1992	4.98	0.6630	0.1580	0.0166	0.1328	0.77
0.7	0.2	0.2922	0.3319	0.0349	0.2789	8.54	0.5281	0.2213	0.0233	0.1859	1.13
0.9	0.2	0.0900	0.4267	0.0449	0.3586	31.21	0.3933	0.2845	0.0299	0.2390	2.20
0.5	0.3	0.4944	0.2371	0.0249	0.1992	4.46	0.6630	0.1580	0.0166	0.1328	1.27
0.7	0.3	0.2922	0.3319	0.0349	0.2789	6.62	0.5281	0.2213	0.0233	0.1859	1.37
0.9	0.3	0.0900	0.4267	0.0449	0.3586	21.60	0.3933	0.2845	0.0299	0.2390	2.00
0.5	0.4	0.4944	0.2371	0.0249	0.1992	4.19	0.6630	0.1580	0.0166	0.1328	1.52
0.7	0.4	0.2922	0.3319	0.0349	0.2789	5.66	0.5281	0.2213	0.0233	0.1859	1.49
0.9	0.4	0.0900	0.4267	0.0449	0.3586	16.79	0.3933	0.2845	0.0299	0.2390	1.90

Table 7.2: Effect of μ_1, μ_2 and ζ on the long run probabilities

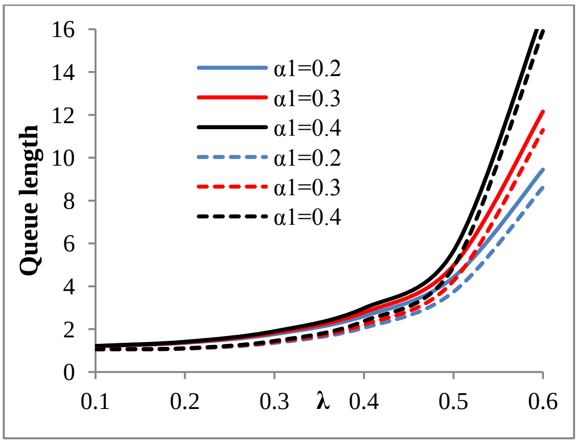
		ζ = 0.6					ζ = 0.4				
μ_1	μ_2	P_I	P_E	P_{O3}	P_V	Ws	P_I	P_E	P_{O3}	P_V	Ws
2	5	0.3759	0.3556	0.0249	0.1992	7.13	0.5839	0.2371	0.0166	0.1328	1.34
3	5	0.4944	0.2371	0.0249	0.1992	4.98	0.6630	0.1580	0.0166	0.1328	0.77
4	5	0.5537	0.1778	0.0249	0.1992	4.21	0.7025	0.1185	0.0166	0.1328	0.53
2	5.5	0.3809	0.3556	0.0239	0.1992	7.02	0.5873	0.2371	0.0160	0.1328	1.31
3	5.5	0.4995	0.2371	0.0239	0.1992	4.91	0.6663	0.1580	0.0160	0.1328	0.75
4	5.5	0.5587	0.1778	0.0239	0.1992	4.15	0.7058	0.1185	0.0160	0.1328	0.51
2	6	0.3851	0.3556	0.0231	0.1992	6.92	0.5901	0.2371	0.0154	0.1328	1.29
3	6	0.5037	0.2371	0.0231	0.1992	4.85	0.6691	0.1580	0.0154	0.1328	0.73
4	6	0.5037	0.2371	0.0231	0.1992	4.85	0.7086	0.1185	0.0154	0.1328	0.49

Table 7.3: Effect of σ_1, β_1 and ζ on the long run probabilities

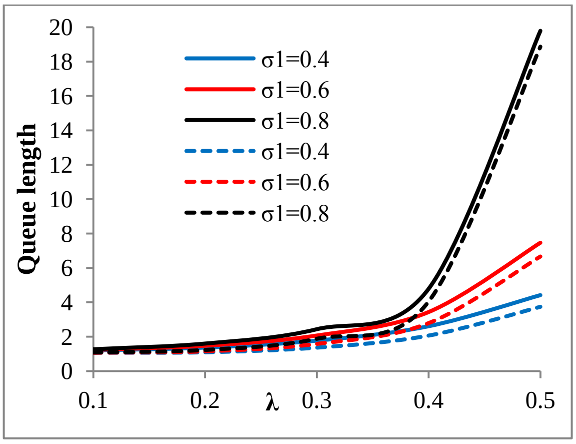
		ζ = 0.6					ζ = 0.4				
σ_1	β_1	P_I	P_E	P_{O3}	P_V	Ws	P_I	P_E	P_{O3}	P_V	Ws
0.3	1.4	0.3855	0.3556	0.0249	0.1896	6.91	0.5903	0.2371	0.0166	0.1264	1.28
0.5	1.4	0.2655	0.3556	0.0249	0.3096	10.70	0.5103	0.2371	0.0166	0.2064	2.01
0.7	1.4	0.1455	0.3556	0.0249	0.4296	20.26	0.4303	0.2371	0.0166	0.2864	2.92
0.3	1.0	0.3544	0.3850	0.0256	0.1896	7.64	0.5696	0.2567	0.0170	0.1264	1.43
0.5	1.0	0.2344	0.3850	0.0256	0.3096	12.25	0.4896	0.2567	0.0170	0.2064	2.20
0.7	1.0	0.1144	0.3850	0.0256	0.4296	25.92	0.4096	0.2567	0.0170	0.2864	3.19
0.3	0.8	0.3192	0.4188	0.0261	0.1896	8.63	0.5461	0.2792	0.0174	0.1264	1.61
0.5	0.8	0.1992	0.4188	0.0261	0.3096	14.58	0.4661	0.2792	0.0174	0.2064	2.45
0.7	0.8	0.0792	0.4188	0.0261	0.4296	37.66	0.3861	0.2792	0.0174	0.2864	3.53



(a)

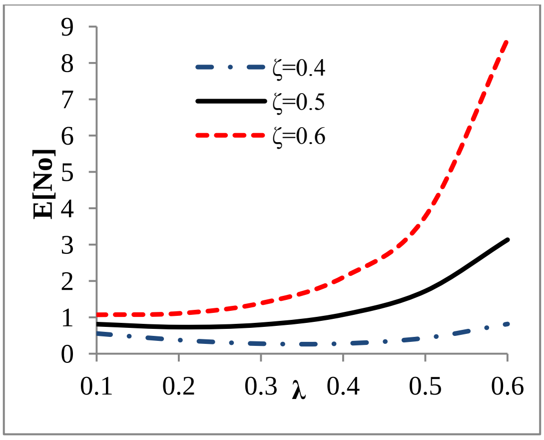


(b)

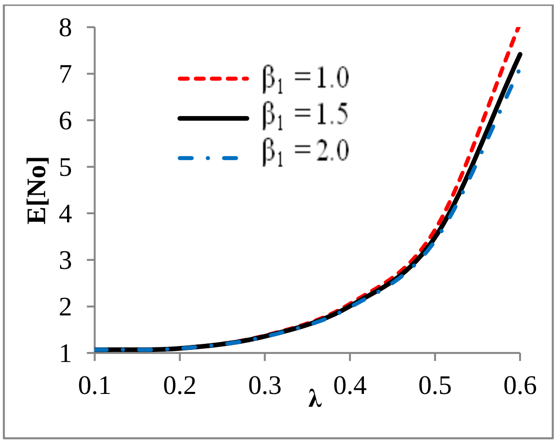


(c)

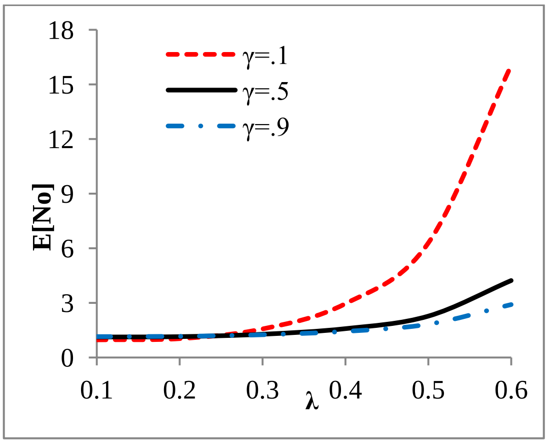
Fig.7.3: Queue length vs. λ for $M^x/M/1$ retrial model for (a) ξ, (b) α_1 (c) σ_1



(a)



(b)



(c)

Fig.7.4: $E[N_o]$ vs. λ for $M^x/\gamma/1$ retrial model for (a) ζ, (b) β_1 (c) γ

on E $[N_o]$ by varying arrival rate for gamma distributed service pattern, have been shown in figs 7.4(a-c). It is clear from the graphs plotted in fig. 7.4(a) that the number of customers in the orbit increases as ζ increases, which is same as we expect. On the other hand, $E[N_o]$ decreases with an increase in repair rate (β_1) and retrial rate (γ), as an increase in repair rate helps in fast servicing of the customers and thus reduces the queue length. If we increase γ, then obviously the number of customers retrying for the service from orbit increases which in turn reduces the queue length of the orbit.

7.10 DISCUSSION

The unreliable retrial queue with bulk input, multi-optional services and phase repair has been investigated using embedded Markov chain technique. Overall based on numerical simulation, we can conclude that-

- An increase in the service rate speeds up the servicing of the customers thereby waiting time reduces while an increase in the vacation probability builds up the queue length as such results in the increment in the waiting time of the customers.
- An increase in the vacation probability and admission probability builds up the queue length of the customers. A control over the customers in both orbit and the system can be maintained by increasing the repair rate and service rate. Thus, an increase in repair rate and service rate can help in reducing the customers in both orbit and the system.
- The average time spend by a customer while waiting (W_s) increases with more arrivals in the system and vacation probability (σ_1) but decreases with the increase in service rates (μ_1, μ_2) and repair rate (β_1).
- Optimal Bernoulli admission control parameter ζ^* is higher for gamma distributed service time and on the contrary optimal service rate μ_1^* is higher for exponentially distributed service pattern.
- The developed cost function is convex w.r.t. service rate μ_1 and admission control parameter ζ for retrial queueing model under consideration. The optimal cost computed can be useful for the system designers and decision makers to have a better idea to trade off between cost and delay of the system.

CHAPTER 8

RETRIAL QUEUE WITH THRESHOLD RECOVERY

8.1 INTRODUCTION

Enormous real life congestion situations deal with the systems where either capacity or population is finite. The systems with finite capacity as well as finite population have been studied in recent past. Such models find numerous applications in machine repair problems, hospitals, educational institutes, telecommunication systems, inventory etc. where either the calling population or capacity of the system is finite.

In previous chapters 2-7, the steady state behavior of non-markovian retrial queues has been analyzed using classical techniques namely supplementary variable technique and embedded markov chain technique, etc. This chapter is devoted to study the transient behavior of the finite capacity retrial queues with threshold based recovery using numerical approach based on Runge-Kutta method. Two different finite models have been explored in this chapter. Section 8.2 deals with the finite capacity retrial queueing model with geometric arrivals. Section 8.3 is concerned with the F-policy finite retrial queueing model. Finally, conclusions are drawn in section 8.4.

8.2 FINITE CAPACITY RETRIAL QUEUEING MODEL WITH THRESHOLD RECOVERY

In this model, we consider unreliable Geo/M/1 retrial queueing model with finite capacity K. Let N(t) be the number of customers in the system at time t and X(t) denote the status of the server s.t. X (t) = {0, 1, 2, 3, 4} for idle, FES, SOS, failure during FES, broken down during SOS, respectively. We define our state space as X= {(x, n): n € N, x € (0, 1, 2, 3, 4)}. To formulate the model, the assumptions made are as follows:

- **Retrial Process**: The customers in the orbit retry for the service with a constant retrial rate γ and compete with the external customers for the service.
- **Arrival Process**: The customers arrive in the system following geometric distribution with state dependent arrival rate Λ. The arrival rate follows geometric distribution with

probability defined by $\Pr\{N(t)=n;\ n\ge 1\}=\Lambda\bar{\Lambda}^{i-1}$, $i\ge 1$. The arrival rates are described here as:

$$\Lambda=\begin{cases}\lambda, \text{when the server is idle}\\ \lambda_e, \text{when the server is busy during FES}\\ \lambda_o, \text{when the server is busy during SOS}\\ \lambda_{re}, \text{when the server is under repair while breakdown during FES}\\ \lambda_{ro}, \text{when the server is under repair while breakdown during SOS}\end{cases}$$

- **Service Process**: The servicing of a customer is basically completed in two stages; first essential service (FES) and second optional service (SOS). The service process follows exponential distribution with parameter $\{\mu_e\}$ for FES and $\{\mu_o\}$ for SOS. All the incoming customers at the first stage are served with rate $p\mu_e$. After completing the FES, the customer may either go to SOS with rate μ_o or leave the system with rate $\bar{p}\mu_e$.
- **Breakdown Process**: The server is unreliable and subject to unpredictable breakdowns. It may breakdown during FES (SOS) according to Poisson process with rate α_e (α_o).
- **Repair Process**: As soon as the server breakdowns, it is immediately sent for repair and after repair it becomes as good as before failure. The repair process follows threshold recovery; the repairing starts only when a sufficient number say q (threshold) customers have been accumulated in the system. The repair rates are defined as-

$$\beta=\begin{cases}\beta_e, \text{when the server is brokendown during FES}\\ \beta_o, \text{when the server is brokendown during SOS}\end{cases}$$

8.2.1 The Governing Equations

Chapman Kolmogorov equations governing the model have been constructed for the finite capacity model which can accommodate K number of customers in the system.

(i) Idle state

$$P'_{0,0}(t)=p\mu_e P_{1,1}(t)+\mu_o P_{2,1}(t)-\lambda P_{0,0}(t) \tag{8.1}$$

$$P'_{0,i}(t)=p\mu_e P_{1,i+1}(t)+\mu_o P_{2,i+1}(t)-(\lambda\bar{\lambda}^{i-1}+\gamma)P_{0,i}(t);\ \ 1\le i\le K-1 \tag{8.2}$$

$$P'_{0,K}(t)=-\gamma P_{0,K}(t) \tag{8.3}$$

(ii) First essential service

$$P'_{1,1}(t)=\gamma P_{0,1}(t)-(\bar{p}\mu_e+p\mu_e+\alpha_e+\lambda_e)P_{1,1}(t)+\lambda P_{0,0}(t) \tag{8.4}$$

$$\begin{aligned}P'_{1,i}(t)=\gamma P_{0,i}(t)+\lambda\bar{\lambda}^{i-2}P_{0,i-1}(t)-(\bar{p}\mu_e+p\mu_e+\alpha_e+\lambda_e\bar{\lambda}_e^{i-1})P_{1,i}(t)\\ +\lambda_e\bar{\lambda}_e^{i-2}P_{1,i-1}(t);\ \ 2\le i\le q-1\end{aligned} \tag{8.5}$$

$$P'_{1,i}(t)=\gamma P_{0,i}(t)+\lambda\bar{\lambda}^{i-2}P_{0,i-1}(t)-(\bar{p}\mu_e+p\mu_e+\alpha_e+\lambda_e\bar{\lambda}_e^{i-1})P_{1,i}(t)$$
$$+\beta_e P_{3,1}(t)+\lambda_e\bar{\lambda}_e^{i-2}P_{1,i-1}(t);\; q\le i\le K-1 \quad (8.6)$$

$$P'_{1,K}(t)=\gamma P_{0,K}(t)+\lambda\bar{\lambda}^{K-2}P_{0,K-1}(t)-(\bar{p}\mu_e+p\mu_e+\alpha_e)P_{1,K}(t)$$
$$+\beta_e P_{3,K}(t)+\lambda_e\bar{\lambda}_e^{K-2}P_{1,K-1}(t) \quad (8.7)$$

(iii) Second optional service

$$P'_{2,1}(t)=\bar{p}\mu_e P_{1,1}(t)-(\lambda_o+\alpha_o+\mu_o)P_{2,1}(t) \quad (8.8)$$

$$P'_{2,i}(t)=\bar{p}\mu_e P_{1,i}(t)+\lambda_o\bar{\lambda}_o^{i-2}P_{2,i-1}(t)-(\lambda_o\bar{\lambda}_o^{i-1}+\alpha_o+\mu_o)P_{2,i-1}(t);\; 2\le i\le q-1 \quad (8.9)$$

$$P'_{2,i}(t)=\bar{p}\mu_e P_{1,i}(t)+\lambda_o\bar{\lambda}_o^{i-2}P_{2,i-1}(t)-(\lambda_o\bar{\lambda}_o^{i-1}+\alpha_o+\mu_o)P_{2,i-1}(t)$$
$$+\beta_o P_{4,i}(t);\; q\le i\le K-1 \quad (8.10)$$

$$P'_{2,K}(t)=\bar{p}\mu_e P_{1,K}(t)+\lambda_o\bar{\lambda}_o^{K-2}P_{2,K-1}(t)-(\alpha_o+\mu_o)P_{2,K}(t)+\beta_o P_{4,K}(t) \quad (8.11)$$

(iv) Repair state for the failure during FES

$$P'_{3,1}(t)=\alpha_e P_{1,1}(t)-\lambda_{re}P_{3,1}(t) \quad (8.12)$$

$$P'_{3,i}(t)=\alpha_e P_{1,i}(t)+\lambda_{re}\bar{\lambda}_{re}^{i-2}P_{3,i-1}(t)-\lambda_{re}\bar{\lambda}_{re}^{i-1}P_{3,i}(t),\; 2\le i\le q-1 \quad (8.13)$$

$$P'_{3,i}(t)=\alpha_e P_{1,i}(t)+\lambda_{re}\bar{\lambda}_{re}^{i-2}P_{3,i-1}(t)-(\lambda_{re}\bar{\lambda}_{re}^{i-1}+\beta_e)P_{3,i}(t),\; q\le i\le K-1 \quad (8.14)$$

$$P'_{3,K}(t)=\alpha_e P_{1,K}(t)+\lambda_{re}\bar{\lambda}_{re}^{K-2}P_{3,K-1}(t)-\beta_e P_{3,K}(t) \quad (8.15)$$

(v) Repair state for the failure during SOS

$$P'_{4,1}(t)=\alpha_o P_{2,1}(t)-\lambda_{ro}P_{4,1}(t) \quad (8.16)$$

$$P'_{4,i}(t)=\alpha_o P_{2,i}(t)+\lambda_{ro}\bar{\lambda}_{ro}^{i-2}P_{4,i-1}(t)-\lambda_{ro}\bar{\lambda}_{ro}^{i-1}P_{4,i}(t),\; 2\le i\le q-1 \quad (8.17)$$

$$P'_{4,i}(t)=\alpha_o P_{2,i}(t)+\lambda_{ro}\bar{\lambda}_{ro}^{i-2}P_{4,i-1}(t)-(\lambda_{ro}\bar{\lambda}_{ro}^{i-1}+\beta_o)P_{4,i}(t),\; q\le i\le K-1 \quad (8.18)$$

$$P'_{4,K}(t)=\alpha_o P_{2,K}(t)+\lambda_{ro}\bar{\lambda}_{ro}^{K-2}P_{4,K-1}(t)-\beta_o P_{4,K}(t) \quad (8.19)$$

8.2.2 Performance Indices

Various performance measures such as the expected number of customers during idle state, busy state and repair state and other indices have been formulated in terms of transient state probabilities as follows:.

- Expected number of customers in the system at any time t is

$$E[N(t)]=\sum_{n=1}^{K}\sum_{i=0}^{4}nP_{i,n}(t) \quad (8.20)$$

- Expected number of customers in the queue at any time t is

$$E[N_q(t)] = \sum_{n=2}^{K}(n-1)\{P_{1,n}(t)+P_{2,n}(t)\} + \sum_{n=1}^{K} n\{P_{3,n}(t)+P_{4,n}(t)\} \tag{8.21}$$

- Throughput at any time t is

$$TP(t) = \mu_e \sum_{n=1}^{K} P_{1,n}(t) + \mu_o \sum_{n=1}^{K} P_{2,n}(t) \tag{8.22}$$

- Reliability of the server at any time t is

$$R(t) = 1 - \left(\sum_{n=1}^{K} P_{3,n}(t) + \sum_{n=1}^{K} P_{4,n}(t) \right) \tag{8.23}$$

- Failure frequency of the system is

$$F_f(t) = \alpha_e \sum_{n=1}^{K} P_{1,n}(t) + \alpha_o \sum_{n=1}^{K} P_{2,n}(t) \tag{8.24}$$

- The expected delay time is

$$E\left[D(t)\right] = \frac{E[N(t)]}{TP(t)} \tag{8.25}$$

The long run probabilities of the system depicting the status of the server at different levels namely idle, busy, repair, etc. are established as

- The long run probability of the server being idle, is

$$P_I = \sum_{n=0}^{K} P_{0,n}(t) \tag{8.26}$$

- The long run probability of the server being busy, is

$$P_B = \sum_{n=1}^{K} P_{1,n}(t) + \sum_{n=1}^{K} P_{2,n}(t) \tag{8.27}$$

- The long run probability of the server being in broken down state and waiting for the repair due to threshold policy, is

$$P_{R1} = \sum_{n=1}^{L-1} P_{3,n}(t) + \sum_{n=1}^{L-1} P_{4,n}(t) \tag{8.28}$$

- The long run probability of the server being broken-down and under repair, is

$$P_{R2} = \sum_{n=L}^{K} P_{3,n}(t) + \sum_{n=L}^{K} P_{4,n}(t) \tag{8.29}$$

8.2.3 Cost Function

To study the effect of sensitiveness of different parameters on the total cost, we construct the cost function involving costs incurred in different activities of the system (cf. Jaggi and Arneja (2010, 2011 a, 2011b). Firstly, we describe here various cost elements associated with different stages of the system as:

C_I : Cost per unit time when the server is idle

C_B : Cost per unit time when the server is busy with either during FES or SOS

C_H : Holding cost per unit time of each customer present in the system

C_R : Repair cost incurred per unit time for a broken down server

The cost function can be constructed in terms of above defined cost elements. The expected total cost per unit time is

$$E[TC(t)] = C_I P_I(t) + C_B P_B(t) + C_H E[N(t)] + C_R [P_{R1}(t) + P_{R2}(t)]$$

8.2.4 Numerical Illustration

The present section provides numerical simulation for the finite retrial model. Here we employ R-K method of fourth order using "ode45" function of MATLAB software to solve the set of differential equations (8.1)- (8.19).

The effect of various parameters on the performance measures has been displayed by means of tables as well as graphs. For tables, the time span of [5-15] with an interval of 5 units has been taken so as to know the sensitivity of performance measures with respect to different parameters. The figures 8.1-8.4 depict the effect of various parameters on the queue length, throughput and reliability on a vast time span of 0 to 200 units.

For computational purpose, we have assign default values as p=0.6, γ=0.1, $\lambda = \lambda_o = \lambda_e = \lambda_{re} = \lambda_{ro} = 0.5, \mu_e = 0.8, \mu_o = 0.8, \alpha_e = 0.001, \alpha_o = 0.002, \beta_e = 0.01, \beta_o = 0.01.$

Tables 8.1-8.3 have been constructed to explore the transient results using heterogeneous arrival rate on various performance measures. The heterogeneous arrival rates are taken as $\lambda_e = 0.7\lambda, \lambda_o = 0.8\lambda, \lambda_{re} = 0.5\lambda, \lambda_{ro} = 0.4\lambda$.

The default cost elements are fixed as

$$C_I = 50, C_B = 10, C_H = 10, C_D = 10$$

It is noticed from table 8.1 that the failure frequency $F_f(t)$ is affected to a very less extent with the variation in the arrival rate. However, E[D(t)], $P_I(t)$, $P_R(t)$ and $P_B(t)$ show significant increment in their values with the growth of time and arrival rate λ. The cost E[TC(t)] is also affected with the growth of time and arrival rate. As the arrival rate increases, the system cost increase which is quite obvious. The effect of retrial rate γ has been displayed in table 8.2. As the retrial rate γ increases, the number of customers in the queue increases. This is so because with the increase in retrial rate, the customers from the orbit try for the service rapidly, which further builds up larger queue in the system. The cost is affected to a lesser extent with the increase in arrival rate.

Table 8.1: Effect of heterogeneous λ on various performance indices

λ	t	E[TC(t)	ED(t)	$P_I(t)$	$P_B(t)$	$P_R(t)$	$F_f(t)$
0.3	**5**	24.78	1.98	0.6562	0.3426	0.0012	0.0004
	10	29.13	2.93	0.6390	0.3575	0.0035	0.0005
	15	32.54	3.69	0.6336	0.3607	0.0057	0.0005
0.5	**5**	29.08	2.44	0.5301	0.4682	0.0018	0.0006
	10	36.85	3.85	0.5353	0.4600	0.0047	0.0006
	15	42.10	4.98	0.5530	0.4396	0.0075	0.0006
0.7	**5**	32.95	2.79	0.4517	0.5462	0.0022	0.0007
	10	41.83	4.48	0.5004	0.4942	0.0055	0.0006
	15	46.54	5.76	0.5471	0.4447	0.0083	0.0006

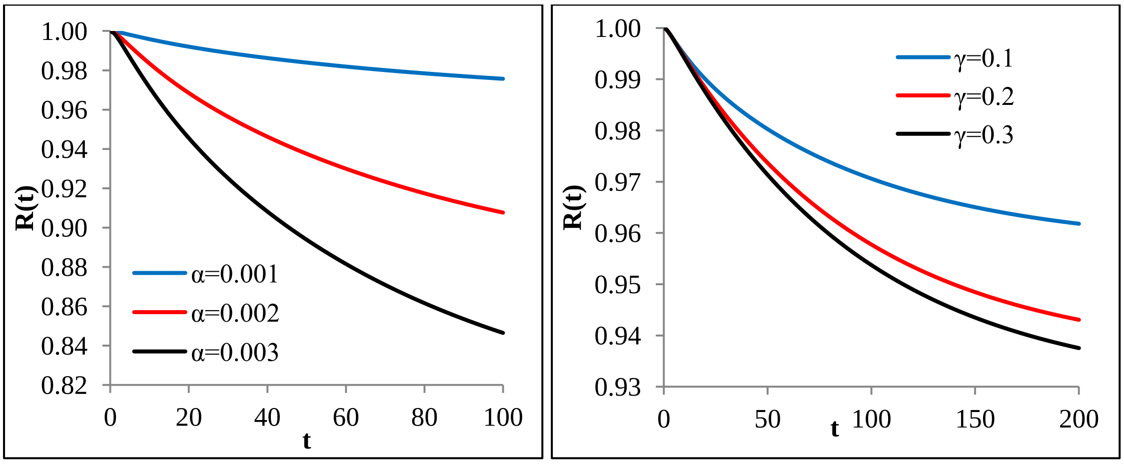
Table 8.2: Effect of retrial rate γ on various performance indices

γ	t	E[TC(t)]	ED(t)	$P_I(t)$	$P_B(t)$	$P_R(t)$	$F_f(t)$
0.2	**5**	30.57	2.66	0.5190	0.4792	0.0018	0.0006
	10	38.20	3.92	0.5080	0.4873	0.0048	0.0006
	15	42.34	4.68	0.5140	0.4784	0.0077	0.0006
0.3	**5**	30.39	2.59	0.5100	0.4882	0.0018	0.0006
	10	37.37	3.64	0.4846	0.5105	0.0050	0.0007
	15	40.86	4.20	0.4820	0.5101	0.0080	0.0007
0.4	**5**	30.24	2.54	0.5022	0.4960	0.0018	0.0006
	10	36.70	3.44	0.4663	0.5287	0.0050	0.0007
	15	39.72	3.87	0.4584	0.5335	0.0082	0.0007

Table 8.3: Effect of breakdown rate α on various performance indices

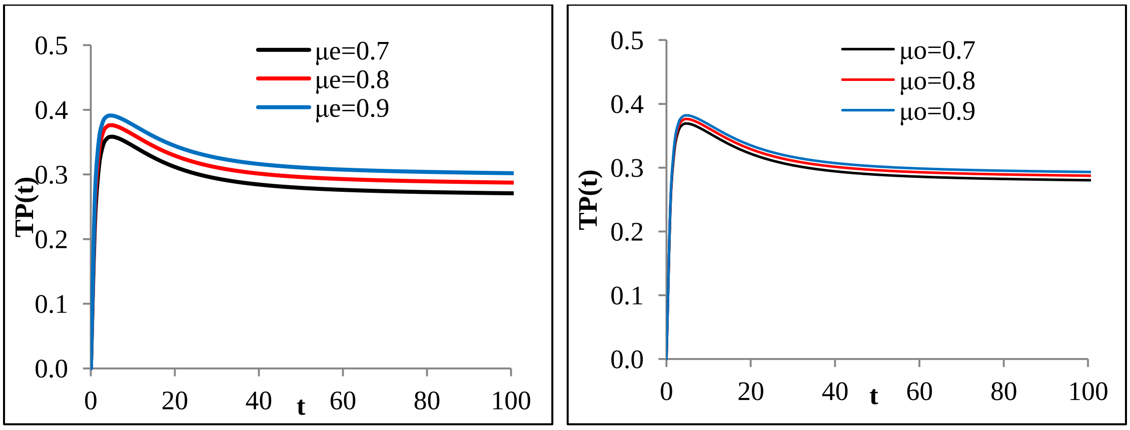
α	t	E[TC(t)]	ED(t)	$P_I(t)$	$P_B(t)$	$P_R(t)$	$F_f(t)$
0.001	**5**	30.76	2.73	0.5295	0.4690	0.0015	0.0005
	10	39.20	4.31	0.5390	0.4574	0.0037	0.0005
	15	44.25	5.45	0.5599	0.4344	0.0058	0.0004
0.004	**5**	30.95	2.75	0.5278	0.4663	0.0059	0.0019
	10	39.75	4.39	0.5336	0.4518	0.0149	0.0018
	15	45.18	5.61	0.5509	0.4265	0.0230	0.0017
0.007	**5**	31.15	2.78	0.5262	0.4637	0.0102	0.0032
	10	40.30	4.47	0.5283	0.4462	0.0258	0.0031
	15	46.08	5.78	0.5421	0.4187	0.0398	0.0029

The probability for the server being in idle state decreases adversely whereas delay time increases with an increase in the values of γ and t. The variation in various indices with breakdown rate (α) is displayed in table 8.3. It is observed that $F_f(t)$ and $P_R(t)$ show a tremendous growth as the system stops working due to the server failure.



(a) (b)

Fig. 8.1: Variation in R(t) with (a) α and (b) γ

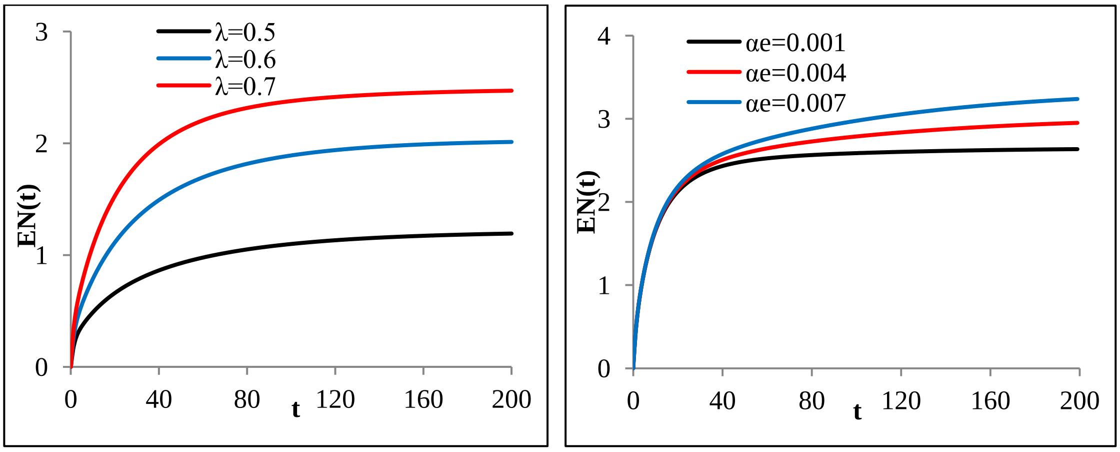


(a) (b)

Fig. 8.2: Variation in TP (t) with (a) μ_e (b) μ_o

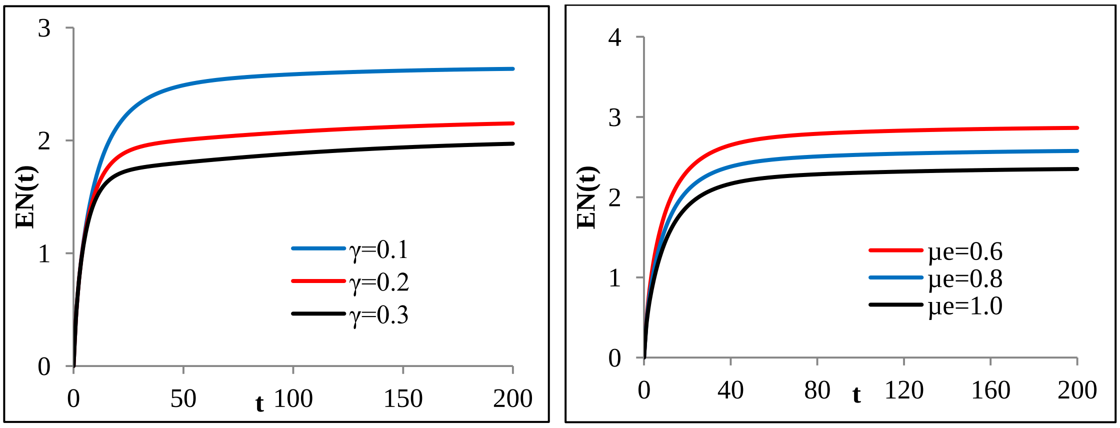
The sensitivity of various affecting parameters has also been demonstrated by means of graphs as shown in figs 8.1-8.4. To exhibit the long run effect on the performance indices, the time span of 0-200 units has been taken into consideration. Fig. 8.1 reveals the variation in R(t) for varying values of breakdown rate (α) and retrial rate (γ). It is clear that the reliability decreases as the value of α increases. With an increase in γ, we notice that R(t) decreases exponentially. Fig. 8.2 graphically displays the effect of service rates (μ_e and μ_o) on the throughput TP(t) with time.

It is clear from the figures that TP(t) increases sharply for initial values of t with varying values of service rates and then decreases till asymptotic constant value is achieved. Fig. 8.3 demonstrates the effect of heterogeneous arrival rate and breakdown rate (α_e) on the expected number of customers in the system E[N(t)]. The number of customers in the system increases with the increase in both arrival rate and breakdown rate. The effect of other parameters namely retrial rate (γ) and service rate (μ_e) are shown by means of graphs plotted



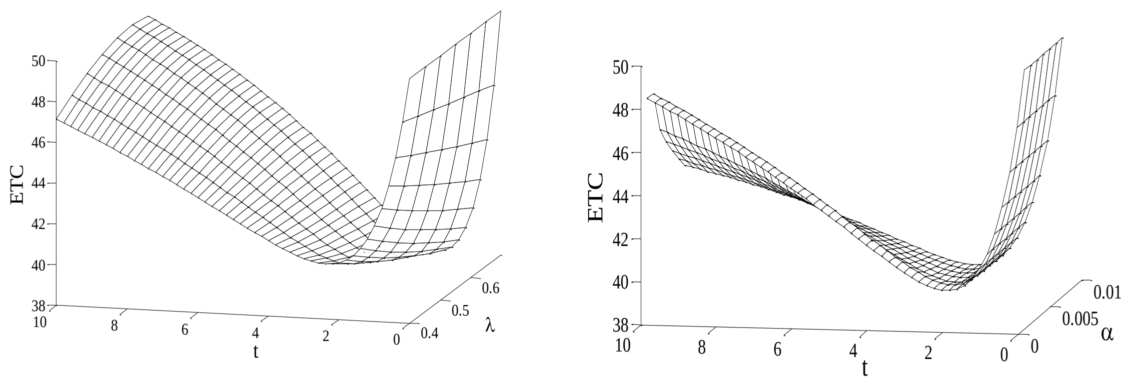
(a) **(b)**

Fig. 8.3: Variation in EN(t) with (a) λ (b) α_e



(a) **(b)**

Fig. 8.4: Variation in EN(t) with (a) γ (b) μ_e



(a) **(b)**

Fig 8.5: Variation in E[TC(t)] with varying parameters

in fig. 8.4. As the service rate increases, the number of customers decreases due to fast servicing to the customers, thereby reducing the length of the system with an increased rate.

To explore the effect of various parameters on the cost of the system the surface graphs have been plotted for varying values of t to exhibit E[TC(t)] as shown in fig. 8.5. Fig. 8.5(a) displays the variation in E[TC(t)] with homogenous arrival rate λ with increasing values of t. It can be easily interpreted that the cost decreases with the increase in λ and t. Similarly, fig. 8.5 (b) shows variation in the E[TC(t)] with the increasing values of α. The curve follows initial decrease and further increase in the cost with time.

8.3 F- POLICY FINITE CAPACITY RETRIAL QUEUE

In previous section 8.2, we dealt with the finite retrial queues with threshold recovery and unreliable server. Besides threshold based recovery, other control policies can also be used to reduce congestion in the system. The present investigation incorporates F-policy to control the arrivals in the system. According to F-policy, no more customers will be allowed to enter the system if the capacity of the system is full but again the arrival process will be initiated at the later stage if a sufficient number of customers are served so as the number of customers in the system ceases to a threshold value ‘F’. In the present section, an unreliable server retrial queueing model with control policy namely F-policy to control the arrivals of the customers in the system has been studied.

8.3.1 Model Description

Consider a finite M/M/1 retrial queue with unreliable server. As soon as K customers are accumulated in the system, further incoming customers may wait in the retrial orbit, from where they retry later on for the service with retrial rate γ. Let N(t) be the number of customers in the system at any time t. Further, $\Theta(t)$ denotes the status of the server such that

$$\Theta(t) = \begin{cases} 1, \text{ if the server is busy in rendering service when the arrivals are allowed} \\ 2, \text{ if the server is busy but the arrivals are not allowed} \\ 3, \text{ retrial state} \\ 4, \text{ setup state before repair of the brokendown server when the arrivals are allowed} \\ 5, \text{ repair state of the server when the arrivals are allowed} \\ 6, \text{ setup state before repair of the brokendown server when the arrivals are not allowed} \\ 7, \text{ repair state of the server when the arrivals are not allowed} \end{cases}$$

Therefore, the state space for the system at time t can be defined as $\Omega(t)=\left(\Theta(t),N(t)\right)$. Moreover $P_{i,n}(t)$ denotes the transient state probability of n customers in the system at time t when the server being in i^{th} state and $i\in\Theta(t)$.

The basic assumptions underlying the model are:

- **Arrival Process:** The arrival pattern of the customers follows a Poisson distribution with the state dependent arrival rate $\Lambda(t)$. The arrival of the customers is continued in the system until K customers have been accumulated and the arrival is controlled using F-policy. At this stage, the arrival of the customers is allowed after a start up time following an exponential distribution with rate ξ. The arrival rate is defined by

$$\Lambda(t)=\begin{cases}\lambda, & \text{if } \Theta(t)=1 \text{ and } 3\\ \lambda b_1, & \text{if } \Theta(t)=4\\ \lambda b_2, & \text{if } \Theta(t)=5\end{cases}$$

  where b_1 (b_2) denotes the joining probabilities of the customers.

- **Service Process:** The customer who finds the server in the idle state is served immediately with service rate $\mu_1\left(\mu_2\right)$ if $\Theta(t)=1(2)$.
- **Breakdown Process:** The server is unreliable and is prone to breakdowns while servicing; the server may fail according to Poisson distribution with rate $\alpha_1(\alpha_2)$ while servicing when the maximum strength of the system is not full (full).
- **Repair Process and Set up before Repair:** Before starting the repair process some preparation time known as setup time is required before starting the repair; and the set up time is exponentially distributed with rate υ. The process of repair is assumed to be exponentially distributed and starts immediately for the server failed while servicing when the arrivals are allowed i.e. $\Theta(t)=2$. On the other hand, the repair process follows the concept of threshold recovery for the server failed when $\Theta(t)=1$. The repair starts only when a minimum number of customers (threshold value) say, q (≥ 1) has been accumulated in the system. The failed system gets repaired with repair rate β_1 (or β_2) while it is failed in case when the arrivals are allowed (or when the maximum strength has been achieved).

8.3.2 Governing Equations

We describe the system by constructing the differential equations for all system state probabilities by using the rates of inflow and outflow. The transient equations for different states of the server are constructed below as:

(i) The server is busy and the arrivals are allowed.

$$P'_{1,0}(t) = \xi P_{2,0}(t) + \gamma P_{3,1}(t) + \mu_1 P_{1,1}(t) - (\lambda + \alpha_1) P_{1,0}(t) \tag{8.30}$$

$$P'_{1,n}(t) = \xi P_{2,n}(t) + \gamma P_{3,n+1}(t) + \mu_1 P_{1,n+1}(t) + \lambda P_{1,n-1}(t) - (\lambda + \alpha_1 + \mu_1) P_{1,n}(t), \quad n = 1, 2, ..., q-1 \tag{8.31}$$

$$P'_{1,n}(t) = \xi P_{2,n}(t) + \gamma P_{3,n+1}(t) + \mu_1 P_{1,n+1}(t) + \lambda P_{1,n-1}(t) + \beta_1 P_{5,n}(t) - (\lambda + \alpha_1 + \mu_1) P_{1,n}(t), \quad n = q, q+1, .., F \tag{8.32}$$

$$P'_{1,n}(t) = \gamma P_{3,n+1}(t) + \mu_1 P_{1,n+1}(t) + \lambda P_{1,n-1}(t) + \beta_1 P_{5,n}(t) - (\lambda + \alpha_1 + \mu_1) P_{1,n}(t), \quad n = F, F+1, ..., K-2 \tag{8.33}$$

$$P'_{1,K-1}(t) = \gamma P_{3,K}(t) + \lambda P_{1,K-2}(t) + \beta_1 P_{5,K-1}(t) - (\lambda + \alpha_1 + \mu_1) P_{1,K-1}(t) \tag{8.34}$$

(ii) The server is busy and the arrivals are not allowed.

$$P'_{2,0}(t) = \mu_2 P_{2,1}(t) + \beta_2 P_{7,0}(t) - (\alpha_2 + \xi) P_{2,0}(t) \tag{8.35}$$

$$P'_{2,n}(t) = \mu_2 P_{2,n+1}(t) + \beta_2 P_{7,n}(t) - (\alpha_2 + \xi + \mu_2) P_{2,n}(t), \quad n = 1, 2...., q-1, q, ...F \tag{8.36}$$

$$P'_{2,n}(t) = \mu_2 P_{2,n+1}(t) + \beta_2 P_{7,n}(t) - (\alpha_2 + \mu_2) P_{2,n}(t), \quad n = F+1, ..., K-1 \tag{8.37}$$

$$P'_{2,n}(t) = \mu_2 P_{2,n+1}(t) + \beta_2 P_{7,n}(t) - (\alpha_2 + \mu_2) P_{2,n}(t), \quad n = F+1, ..., K-1 \tag{8.38}$$

(iii) The server is under retrial state.

$$P'_{3,0}(t) = -\lambda P_{3,0}(t) \tag{8.39}$$

$$P'_{3,n}(t) = \lambda P_{3,n-1}(t) - (\lambda + \gamma) P_{3,n}(t), \quad n = 1, 2, ..., (K-1) \tag{8.40}$$

$$P'_{3,K}(t) = \lambda P_{3,K-1}(t) - \gamma P_{3,K}(t) \tag{8.41}$$

(iv) The server is under set up before starting the repair process of the broken down server that failed while servicing when new customers were allowed in the system.

$$P'_{4,1}(t) = \alpha_1 P_{1,1}(t) - (\lambda b_1) P_{4,1}(t) \tag{8.42}$$

$$P'_{4,n}(t) = \alpha_1 P_{1,n}(t) + (\lambda b_1) P_{4,n-1}(t) - (\lambda b_1) P_{4,n}(t), \quad n = 2, ..., (q-1) \tag{8.43}$$

$$P'_{4,n}(t) = \alpha_1 P_{1,n}(t) + (\lambda b_1) P_{4,n-1}(t) - (\lambda b_1 + \upsilon) P_{4,n}(t), \quad n = q, ..., (K-1) \tag{8.44}$$

(v) The repair state of the failed server, broken down while servicing when new customers were allowed in the system.

$$P'_{5,1}(t) = -\lambda b_2 P_{5,1}(t)$$

$$P'_{5,n}(t) = \lambda b_2 P_{5,n-1}(t) - (\lambda b_2) P_{5,n}(t), \quad n = 2, ..., (q-1) \tag{8.46}$$

$$P'_{5,n}(t) = \lambda b_2 P_{5,n-1}(t) + \upsilon P_{4,n}(t) - (\lambda b_2 + \beta_1) P_{5,n}(t), n = q, q+1, ..., (K-1) \tag{8.47}$$

(vi) The server is under set up before starting the repair process of the broken down server that failed while servicing when no more customers were allowed in the system.

$$P'_{6,n}(t) = \alpha_2 P_{2,n}(t) - \upsilon P_{6,n}(t), \quad n = 1, 2, ..., K \tag{8.48}$$

(vii) The repair state of the failed server, which break down while servicing when no more customers were allowed to enter the system.

$$P'_{7,n}(t) = \upsilon P_{6,n}(t) - \beta_2 P_{7,n}(t), \quad n = 1, 2, ..., K \tag{8.49}$$

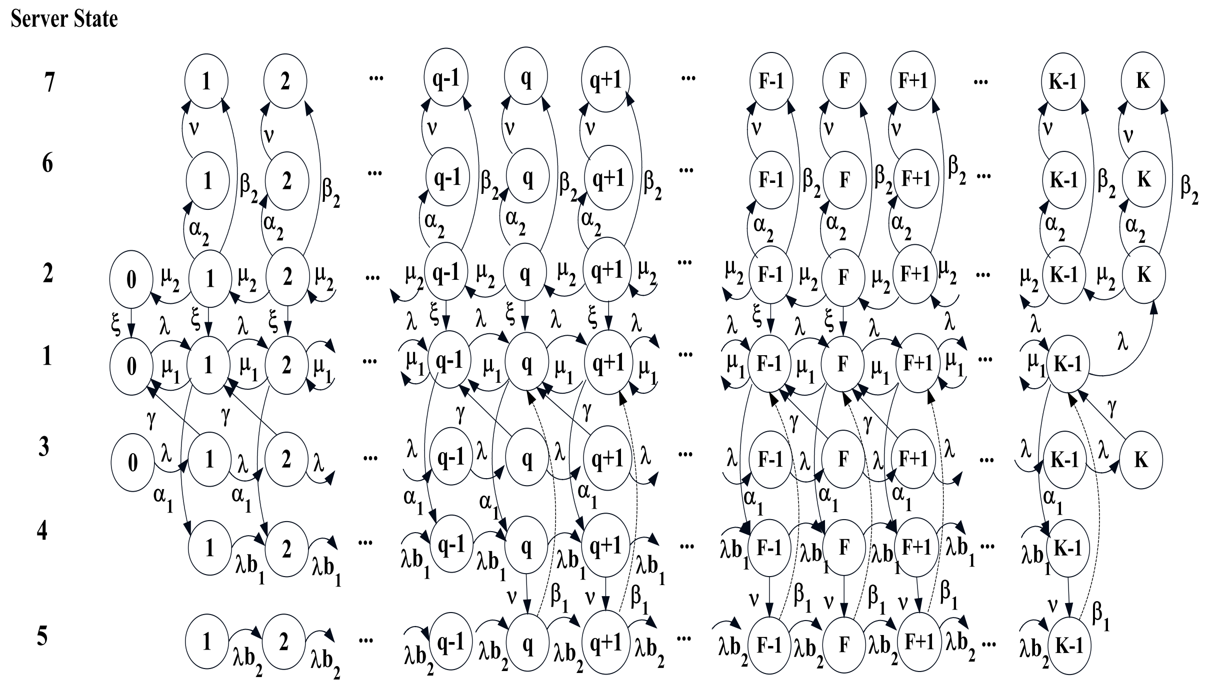


Fig. 8.6: State transition rate diagram for retrial queueing system

8.3.4 Performance Measures

In the present sub-section, we derive analytic expressions for the various performance measures namely queue length, reliability, throughput, failure frequency, waiting time etc., in terms of the transient probabilities.

(A) Server State Probabilities

Various probabilities for the different states of the server at time t are established as follows:

- The probability that the server being busy in providing service to the customers, is

$$P_B(t) = \sum_{n=1}^{K-1} P_{1,n}(t) + \sum_{n=1}^{K} P_{2,n}(t) \tag{8.50}$$

- The probability that the server starts to allow the customers to enter the system, is

$$P_s(t)=\sum_{n=0}^{F}\mathrm{P}_{2,\mathrm{n}}(t) \quad (8.51)$$

- The probability that the system is blocked, is

$$P_L(t)=\sum_{n=0}^{K}P_{2,n}(t)+\sum_{n=0}^{K}P_{6,n}(t)+\sum_{n=0}^{K}P_{7,n}(t) \quad (8.52)$$

- The probability that the broken down server is under setup before repair, is

$$P_T(t)=\sum_{n=0}^{K-1}P_{4,n}(t)+\sum_{n=0}^{K}P_{6,n}(t) \quad (8.53)$$

- The probability that the server starts the repair, is

$$P_R(t)=\sum_{n=q+1}^{K-1}P_{5,n}(t)+\sum_{n=q+1}^{K}P_{7,n}(t) \quad (8.54)$$

- The probability that the customer retry for the service, is

$$P_{RE}(t)=\sum_{n=1}^{K}P_{3,n}(t) \quad (8.55)$$

(B) Queueing Measures

Below we describe some of the transient queueing measures related to our model.

- Expected queue length at time t, is

$$L_s(t)=\sum_{i=1}^{7}\sum_{n=1}^{K-1}nP_{i,n}(t)+K\sum_{i=2,3,6,7}P_{i,K}(t) \quad (8.56)$$

- Expected waiting time in the system at time t, is

$$W_s(t)=L_s(t)/\lambda_{eff}(t) \quad (8.57)$$

where, the effective arrival rate $\lambda_{eff}(t)$ at time t is obtained by using

$$\lambda_{eff}(t)=\sum_{n=0}^{K-1}\lambda\left[P_{1,n}(t)+P_{3,n}(t)+b_1P_{4,n}(t)+b_2P_{5,n}(t)\right]$$

- Throughput at time t is obtained by using

$$TP(t)=\mu_1\sum_{n=1}^{K-1}P_{1,n}(t)+\mu_2\sum_{n=1}^{K}P_{2,n}(t) \quad (8.58)$$

(C) Reliability Measures

Some of the reliability indices are as follows:

- Availability of the server at time t, is

$$A_v(t)=\sum_{n=1}^{K-1}P_{1,n}(t)+\sum_{i=2}^{3}\sum_{n=1}^{K}P_{i,n}(t) \quad (8.59)$$

- Failure frequency, is

$$F_f(t) = \alpha_1 \sum_{n=0}^{K-1} P_{1,n}(t) + \alpha_2 \sum_{n=0}^{K} P_{2,n}(t) \tag{8.60}$$

(D) Cost Function

We construct the cost function for the finite retrial model under consideration. The cost function is framed as:

$$TC(F,q) = C_B P_B(t) + C_h L_s(t) + C_R P_R(t) + C_L \lambda P_L(t) + C_1 \mu_1(t) + C_2 \mu_2(t) + \xi C_S + \gamma C_{RE} + \nu C_{SET} \tag{8.61}$$

where,

C_L : Cost for every lost customer when the arrivals are not allowed;

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 : Repair cost incurred per unit time for a broken down server;

C_1 : Cost for providing the service to the customer when the arrivals are allowed;

C_2 : Cost for providing the service to the customer when the arrivals are not allowed;

C_s : Fixed cost for startup process per unit customer when the customers are allowed to enter;

C_{RE} : Fixed cost incurred for each retrial customer at each time;

C_{SET} : Fixed cost for setup process before starting the repair process.

8.3.5 Numerical Results and Cost Analysis

To evaluate the optimal values of decision parameters 'F' and 'q', we consider the cost function given by (8.61). We find the optimal values of decision variables F and q by numerical computation based on discrete allocation. Three cost sets are taken into consideration to find out the best value for 'F' and 'q' with pre-assigned values of other parameters. We mathematically formulate the optimization problem as-

$$TC(F^*, q^*) = \text{Minimize } TC(F, q)$$

subject to $1 \le F \le K-1$ and $1 \le q \le F-1$

For discrete optimization, we use the inequalities

$$TC(F-1, q+1) \ge TC(F^*, q^*),\ TC(F+1, q-1) \ge TC(F^*, q^*) \text{ and } TC(F+1, q+1) \ge TC(F^*, q^*)$$

The various cost sets under consideration are:

Set 1: C_h=5, C_b= 200, C_L=300, C_1=70, C_2=70, C_s= 250, C_{SET} = 100, C_R=20,C_{RE} = 200

Set 2: C_h=50, C_b= 30, C_L=150, C_1=40, C_2=70, C_s= 300, C_{SET} = 150, C_R=100,C_{RE} = 15

Set 3: C_h=100, C_b= 200, C_L=200, C_1=50,C_2=50, C_s= 400, C_{SET} = 50, C_R=200,C_{RE} = 200

In the present cost optimization problem as structured in (8.61), the cost parameters are assumed to be linear in nature. It is difficult to obtain the analytic solution, therefore we use numerical approach based on Runge Kutta method to find out the system state probabilities

Table 8.4: Effect of (λ, μ_2) on optimal (F^*, q^*) for different cost sets

	(λ, μ_2)								
	(1,3)	**(2,3)**	**(2.5,3)**	**(1,5)**	**(2,5)**	**(2.5,5)**	**(1,7)**	**(2,7)**	**(2.5,7)**
Cost Set 1									
F^*	9	9	3	9	5	3	9	5	3
q^*	8	8	2	8	4	2	8	4	2
TC(F^*, q*)	1061.00	1129.30	1333.50	1201.70	1266.90	1439.70	1341.70	1405.10	1563.70
Cost Set 2									
F^*	9	4	2	9	4	2	9	3	2
q^*	8	3	1	8	3	1	8	2	1
TC(F^*, q*)	1033.90	1088.90	1275.00	1173.90	1225.20	1387.80	1313.90	1363.50	1515.10
Cost Set 3									
F^*	9	3	2	9	3	2	9	3	2
q^*	8	2	1	8	2	1	8	2	1
TC(F^*, q*)	1308.30	1453.70	1863.80	1408.30	1546.70	1918.20	1508.30	1643.50	1996.50

Table 8.5: Effect of (λ, β_1) on optimal (F^*, q^*) for different cost sets

	(λ, β_1)								
	(1,0.1)	**(2,0.1)**	**(3,0.1)**	**(1,0.3)**	**(2,0.3)**	**(3,0.3)**	**(1,0.5)**	**(2,0.5)**	**(3,0.5)**
Cost Set 1									
F^*	8	5	2	9	5	2	9	5	2
q^*	7	4	1	8	4	1	8	4	1
TC(F^*, q^*)	1202.20	1266.90	2276.90	1201.90	1268.30	2282.30	1202.10	1269.20	2286.70
Cost Set 2									
F^*	9	4	2	2	4	2	9	4	2
q^*	8	3	1	1	3	1	1	3	1
TC(F^*, q^*)	1173.90	1225.20	2136.10	1173.70	1226.70	2141.20	1173.00	1227.60	2145.30
Cost Set 3									
F^*	9	3	2	9	3	2	9	3	2
q^*	8	2	1	8	2	1	1	2	1
TC(F^*, q^*)	1408.30	1546.70	3524.80	1408.90	1550.80	3538	1405.70	1553.10	3547.70

and further employ direct search approach based on discrete allocation to find out the optimal threshold values of ‘F’ and ‘q’.

The optimal control parameters ‘F’ and ‘q’ are determined using direct search approach by computing the cost. For this purpose, time t is fixed as 200 units and all the performance measures involved in the determination of total cost have been calculated at that particular time. Tables 8.4 and 8.5 display the values of optimal parameters F^* and q^* for three different costs sets and varying values of other parameters namely arrival rate (λ), service rate (μ_2) and repair rate (β_1). The capacity of the system is fixed as K=10 and we vary the values of F and q in their permissible range to obtain the minimum expected cost. It is observed from table 8.4 that the total minimum cost increases as arrival rate λ increases. In table 8.5, various sets of optimal (F^*, q^*) are summarized corresponding to different sets of (λ, β_1).

Now, using R-K method we examine various transient state performance measures under various conditions from sensitivity analysis view point. The default parameters for this purpose are fixed as:

$$\lambda = 1,\ \gamma = 0.1,\ \mu_1 = 4,\ \mu_2 = 5,\ \beta = \beta_1 = \beta_2 = 0.1,\ \alpha = \alpha_1 = \alpha_2 = 0.01,\ \nu = 0.06,\ \xi = 2,\ b_1 = b_2 = 1.$$

The effect of sensitive parameters on various performance indices viz. server state probabilities, reliability indices and queue length of the system are examined by varying different parameters as given below.

(A) Server State Probabilities

These probabilities of the server can be interpreted as the proportion of time for which the system/server remains at a particular state. Table 8.6 demonstrates the sensitivity of these measures towards arrival rate (λ). With an increase in time t, probabilities $P_B(t)$, $P_T(t)$ and $P_R(t)$ decrease while $P_L(t)$ and $P_S(t)$ increase. It is quite clear from the table 8.6 that as the arrival rate λ increases from 1 unit to 2 units, the system probabilities of the server being in busy state ($P_B(t)$), the probability that the server starts to allow the customers to enter in the system ($P_S(t)$) and the probability of blocked customer ($P_L(t)$), increase whereas probability of the server under setup state before repair ($P_T(t)$) decreases.

(B) Reliability Indices

Table 8.7 displays the effect of arrival rate (λ) and repair rate (β) on the reliability indices namely availability of the server ($A_v(t)$) and failure frequency ($F_f(t)$) of the system at time t. The failure frequency decreases with the increase in λ from 1 unit to 1.5 units but increases with the increase in λ from 1.5 units to 2 units. The effects of other parameters on the $A_v(t)$ are demonstrated by means of graphs as shown in figs 8.7(a-d). Fig. 8.7 (a) shows the effect of λ on the availability of the server. The effect of breakdown rate (α) on $A_v(t)$ at

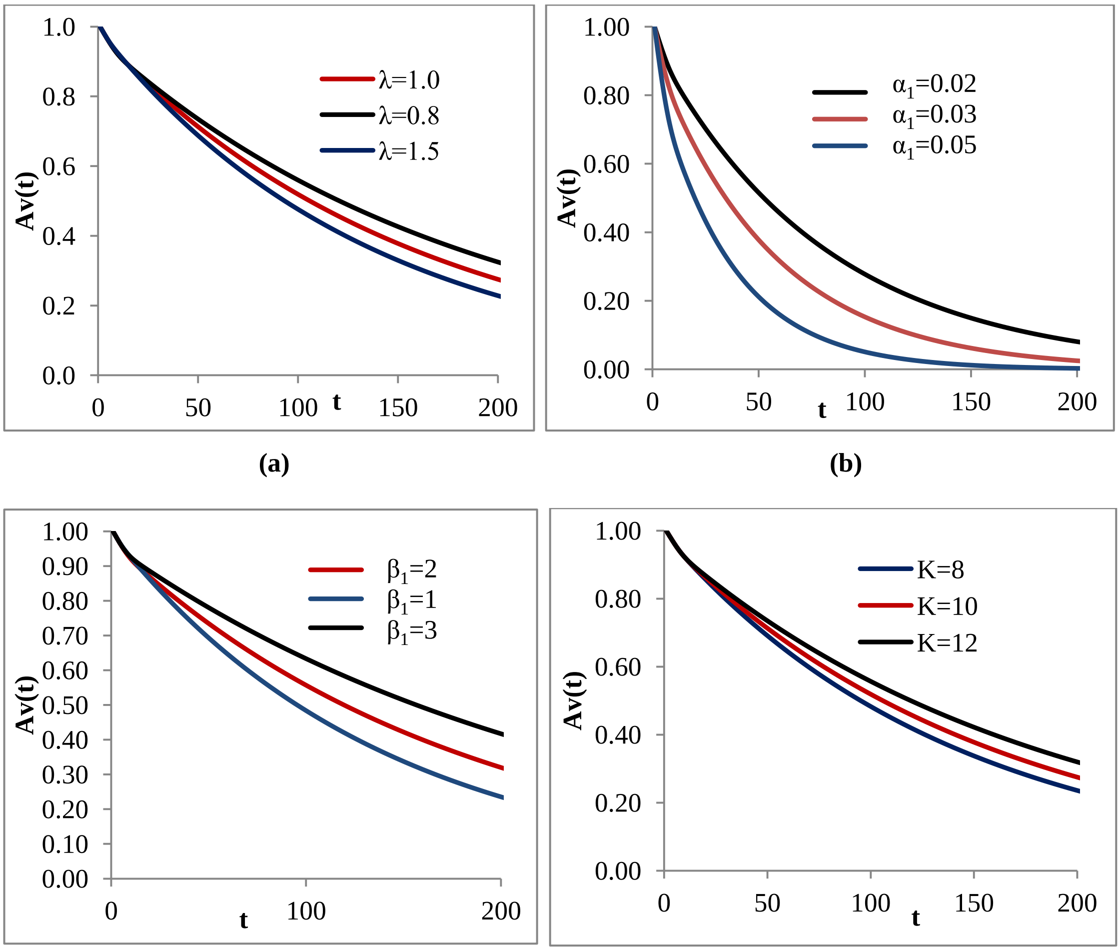
Table 8.6: Effect of λ on the server state probabilities

λ	t	$P_B(t)$	$P_S(t)$	$P_L(t)$	$P_T(t)$	$P_R(t)$	$W_s(t)$
1	10	0.22577	0.00004	0.00015	0.06919	0.05506	0.98
	20	0.20725	0.00008	0.00027	0.06749	0.05329	1.04
	30	0.18977	0.00008	0.00027	0.06179	0.05041	1.04
	40	0.17375	0.00007	0.00026	0.05658	0.04617	1.04
	50	0.15909	0.00006	0.00024	0.05181	0.04228	1.04
1.5	10	0.34091	0.00030	0.00097	0.04968	0.05655	0.84
	20	0.31306	0.00029	0.00100	0.04572	0.05223	0.85
	30	0.28738	0.00027	0.00097	0.04200	0.04800	0.85
	40	0.26382	0.00025	0.00093	0.03857	0.04410	0.85
	50	0.24219	0.00023	0.00087	0.03541	0.04050	0.85
2	10	0.48513	0.00280	0.00903	0.03794	0.10002	0.89
	20	0.47835	0.00287	0.00985	0.03762	0.09972	0.90
	30	0.47137	0.00284	0.01020	0.03730	0.09875	0.90
	40	0.46459	0.00281	0.01035	0.03689	0.09762	0.90
	50	0.45795	0.00278	0.01038	0.03644	0.09641	0.90

Table 8.7: Effect of λ on the availability and failure frequency

λ	t	Av(t)			$F_f(t)$		
		β_1=0.1	β_1 =0.5	β_1 =0.9	β_1=0.1	β_1 =0.5	β_1 =0.9
1.0	10	0.9069	0.9041	0.9061	0.0091	0.0090	0.0091
	20	0.8291	0.8480	0.8557	0.0083	0.0085	0.0086
	30	0.7587	0.7899	0.8022	0.0076	0.0079	0.0080
	40	0.6943	0.7358	0.7520	0.0069	0.0074	0.0075
	50	0.6855	0.6855	0.7049	0.0069	0.0069	0.0070
1.5	10	0.9081	0.9134	0.9161	0.0091	0.0091	0.0092
	20	0.8279	0.8434	0.8502	0.0083	0.0084	0.0085
	30	0.7549	0.7788	0.7892	0.0075	0.0078	0.0079
	40	0.6883	0.7192	0.7325	0.0069	0.0072	0.0073
	50	0.6276	0.6641	0.6799	0.0063	0.0066	0.0068
2.0	10	0.9176	0.9228	0.9256	0.0092	0.0092	0.0093
	20	0.8473	0.8607	0.8673	0.0085	0.0086	0.0087
	30	0.7823	0.8023	0.8128	0.0078	0.0078	0.0081
	40	0.7224	0.7488	0.7617	0.0072	0.0075	0.0076
	50	0.6671	0.6985	0.7139	0.0067	0.0070	0.0071

time t is displayed in fig. 8.7(b). The effects of repair rate (β_1) and capacity (K) of the system on $A_v(t)$ are displayed by means of figs 8.7(c) and 8.7(d), respectively. As the repair rate (β_1) increases, the availability of the server also increases because faster repair rate makes the server more available for the customers.



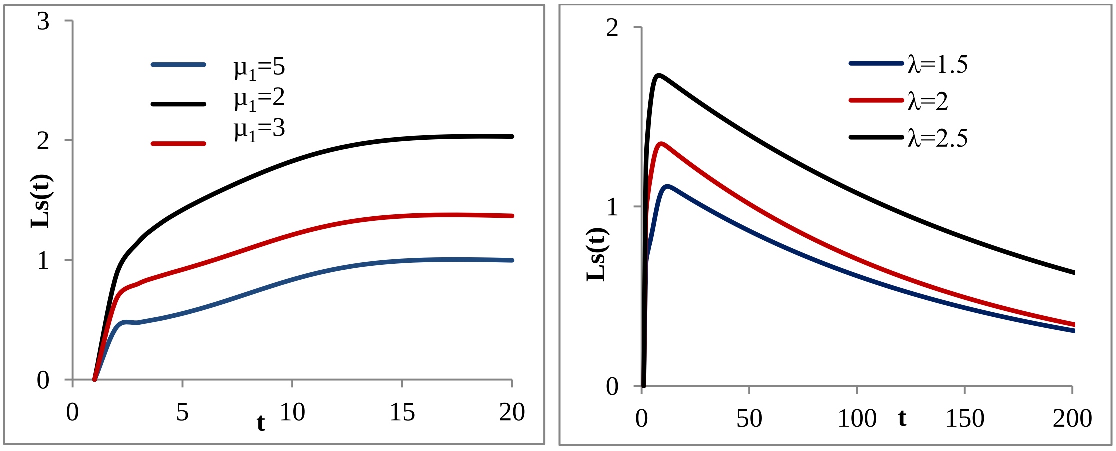
(c) (d)

Fig. 8.7: Effect of various parameters on $A_v(t)$ with time t

(C) Queueing Indices

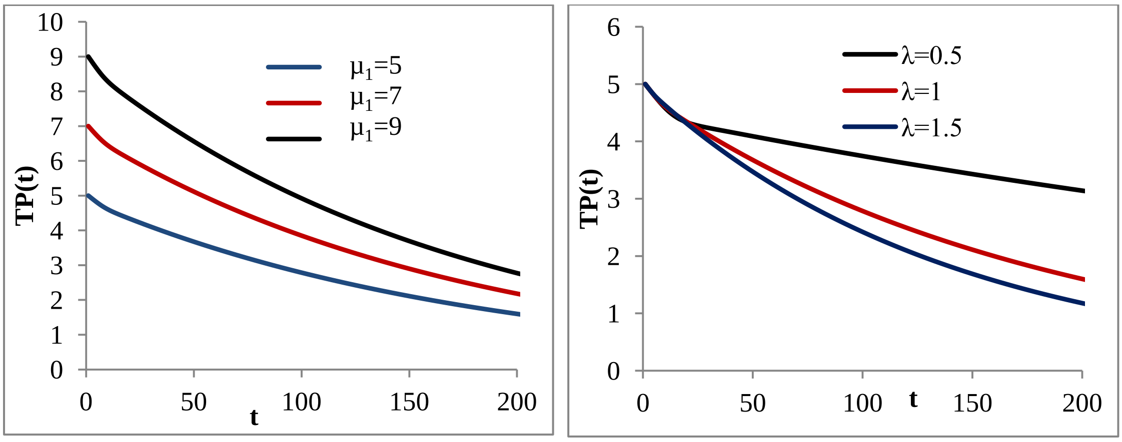
The effects of various parameters on the waiting time $W_s(t)$ have been displayed in table 8.6. The waiting time $W_s(t)$ increases with an increase in time while it decreases with the growth of arrival rate from 1 unit to 1.5 units but increases as λ grows upto 2 units. Figs 8.8 (a-b) are plotted so as to examine the sensitiveness of the queue length towards other parameters. The effect of service rate μ_1 on the queue length is shown in fig. 8.8(a). The number of customers in the system reduces as the service rate increases which is obvious. On the other hand, reverse effects are observed on the queue length ($L_s(t)$) with respect to λ.

Figs 8.9(a-b) display the effect of service rate (μ_1) and arrival rate (λ) on the throughput TP(t) of the system. Fig. 8.9(a) reveals the effect of service rate (μ_1) on TP(t) by varying time t. As μ_1 increases, TP(t) increases which corresponds to higher number of successful services at any instant of time. The reverse trends are seen when TP(t) is plotted against t as demonstrated by figure 8.9 (b).



(a) **(b)**

Fig. 8.8: Effect of μ_1 and λ on the queue length $L_s(t)$ of the system

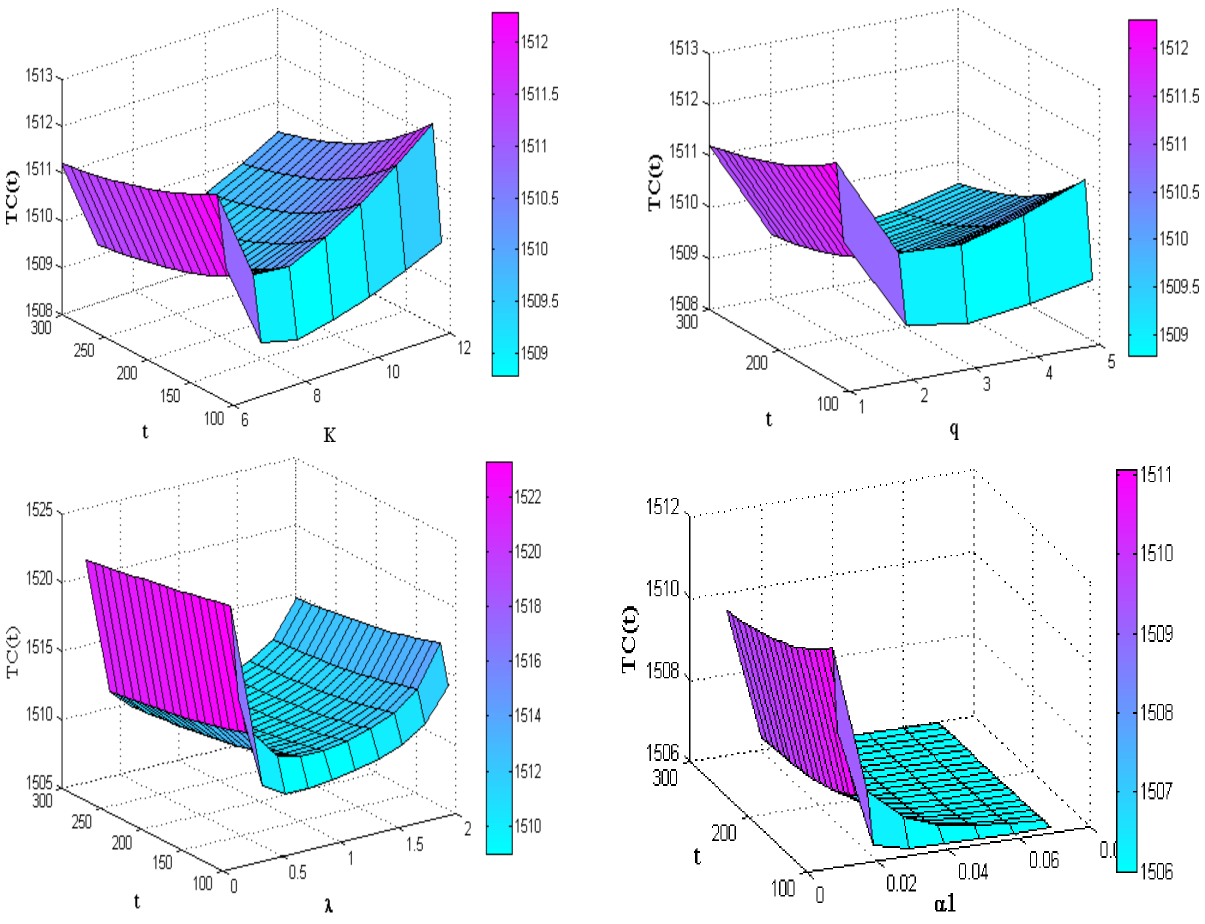


(a) **(b)**

Fig. 8.9: Effect of μ_1 and λ on the Throughput TP(t) of the system

Here, we deal with the sensitivity analysis of total expected cost TC(t) towards various parameters for cost set 1 for finite capacity system. The surface graphs are plotted as shown in figs 8.10(a-d) to explore the effect of parameters on the cost TC(t). A range of 200 units i.e. 100-300 time span is taken on the y-axis with total cost TC(t) taken on z-axis. Fig. 8.10(a) is plotted for varying values of capacity K of the system with maximum cost of 1513 units. A smooth convex curve is obtained for varying values of K. As K increases from K=6 to around K=9, TC(t) decreases and then increases upto K=12.

Similarly, fig. 8.10 (b) shows the variation in the cost of the system with threshold parameter q and time t. A smooth convex surface graph is plotted with minimum cost at q*=4 for this case. Hence, a model with fixed capacity K=10 and F=5 gives optimal threshold parameter as q*=4. Figs 8.10(c) and 8.10(d) demonstrate the effect of arrival rate (λ) and breakdown rate (α_1) on the cost function. In fig. 8.10(c), TC(t) decreases with an increase in λ upto λ=1 (approx.) and then increases hereby proving the convexity of the



(c) **(d)**

Fig. 8.10: Effect of (a) Capacity K, (b) threshold parameter q, (c) arrival rate λ and (d) breakdown rate α_1 on the total cost

function. Furthermore, in fig. 8.10(d) a continuous decreasing function is noticed with the increasing values of α_1, implying that the total cost decreases as the breakdown rate increases for these parameters.

8.4 DISCUSSION

Two finite capacity retrial queueing models with threshold recovery have been investigated using numerical approach based on Runge-Kutta method. Overall, we can conclude that-

- The threshold based recovery parameter q* determined by using heuristic search approach plays an important role in various real life congestion situations like telecommunication systems, traffic systems wherein the server needs to control the arrivals of messages/jobs so as to maintain the efficiency of the system and enhance the capability and reliability.
- The optimized cost function can be utilized in order to determine the optimal threshold parameters and minimum cost to design optimal and more efficient systems.
- The total number of customers in the system increases (decreases) with an increase in the arrival rate and breakdown rate (service rate and retrial rate).
- Reliability decreases with the increase in breakdown rate which is consistent with the realistic situations. As the arrival rate increases, the availability of the server in the working state $A_v(t)$ decreases at any instant.
- The state probabilities of the server states, queueing measures including the queue length and reliability measures of a system can also be maintained by controlling various parameters.
- It is noticed that the efficiency of the system can be improved by increasing the service rate so as to reduce the accumulation of the customers in the system and hence reduction in the waiting time of the customers in the system.

CHAPTER 9

RETRIAL QUEUE WITH ADDITIONAL SERVER

9.1 INTRODUCTION

In the retrial queueing systems, the server load can sometimes be lowered by the provision of additional server to serve the customers. It is usually observed in day to day routine activities that the provision of temporary servers in the case when the server load increases, can play a significant role to improve the efficiency of the system. The provision of additional temporary server is basically done to reduce the load on a single server; this may also be helpful in reducing the waiting time of the customers. The concept of installing temporary server finds several applications in real life congestion problems such as telecommunication systems, web servers, computer protocols, message transmission, admission counters, dispensaries and many other situations. Web faction is one of the key areas where temporary server can be installed in the case when the server load crosses a threshold value. The secondary server is usually installed with an aim to reduce the waiting time of the customers and to increase the efficiency of the system in terms of faster service rendered.

In this chapter, we study a retrial queueing system in which the primary server is prone to breakdowns and can serve only a limited number of customers. There is provision of temporary server which is switched on only when the load on the first server crosses the pre specified threshold load in terms of the number of customers in the system. The main objective of the investigation presented in this chapter is to obtain the server state probabilities and various performance measures using matrix geometric approach. The rest of the chapter is organized in the following manner. Section 9.2 deals with the detailed description of the model including various assumptions, applications and equations governing the model. Section 9.3 contains the methodology used and provides the analysis of the queueing model under consideration. Sections 9.4 and 9.5 respectively, present the various performance measures and cost function of the system under consideration. The numerical illustration and sensitivity analysis has been carried out in section 9.6 and 9.7, respectively. Finally conclusions are drawn in section 9.8.

9.2 THE MATHEMATICAL MODEL

Consider a retrial queue with unreliable primary server in which the system has the provision of installing a second temporary server which is turned on when the number of customers with the first server exceeds a pre-specified level. The various features of the model are discussed in the following subsections.

9.2.1 Model Description

The retrial model under consideration has the provision of two servers, out of which second temporary server is activated only when the work load with the primary server crosses a threshold level. The various assumptions and notations underlying the model are as follows:

- **Arrival Process:** The arrivals of the customers in the system follow Poisson pattern with arrival rate λ. There is a provision of two servers; the first primary server and second temporary server. The second temporary server is installed only if 'K' customers are already queued up before the primary server including the one in the service. If an arriving customer finds less than 'K' customers with the primary server, then either he waits for his turn in the queue with the primary server or may join the waiting space i.e. orbit. But if on arrival, the primary server's buffer is fully occupied with 'K' customers, then the new arrival has no other option rather than to join the buffer of the secondary server.
- **Retrial Process:** The customers accumulated in the orbit retry with exponentially distributed retrial rate γ and compete for the service with the primary customers as soon as they find the server idle.
- **Service Process:** The customers are served following exponential distribution with rate μ_i, if queued before i^{th} server (i=1 for primary server and i=2 for secondary server). The number of customers joining the secondary server is unlimited. Both the servers have their own independent queues but the formation of second queue takes place when the buffer of primary server is full. No queue shifting is permitted to the customers once they join it.
- **Breakdown and Repair Process:** The primary server is unreliable and may breakdown while serving the customers; the broken down server is sent for the repair immediately and after repair, it becomes as good as before failure. However, the temporary second server is considered as reliable server. The life time and repair time

of the primary server follow the exponential distribution with rates α_1 and β_1, respectively.

9.2.2 State of the Server

To describe the state of the server at any instant, we consider the following three random variables that describe the system completely:

(i) $\Theta(t)$, represents the state of the server and takes values 0,1 or 2 when the server is in retrial, busy or broken down state/under repair states, respectively.

(ii) $N_1(t)$, denotes the number of customers with the first server, such that $N_1(t) = i$, $(0 \le i \le K)$

(iii) $N_2(t)$, denotes the number of customers with the second server, such that $N_2(t) = j$, $j \ge 0$.

Now, the state space of stochastic process of concerned model is completely specified as

$$\Omega = \{(\Theta(t), \mathrm{N}_1(\mathrm{t}) \ , \mathrm{N}_2(\mathrm{t}) \ ) : \Theta(t) \in (0,1,2), \mathrm{N}_1(\mathrm{t}) \in (0 \le i \le K), \mathrm{N}_2(\mathrm{t}) = j(\ge 0)\}.$$

Also, the probability of the server at any instant of time is denoted as $P_{k,i,j}$ for $k \in (0,1,2); (0 \le i \le K); j \ge 0$.

9.2.3 Application of the Model to Web Faction

There are enormous applications of the present retrial queueing model. Here, we cite a useful real life application in case of web faction which can be explained as follows. Web faction is a system which provides a complete web hosting service, with everything required to set up and run web services. It provides rights and powers to run basic blogs to advanced web applications in which heavy load are shared by multiple servers. Web faction conducts daily backup of all home directories, email accounts and this backup is usually retained for a maximum of ten days. It has a key tool known as control panel which helps us in linking to all the portions of clients account together. The account and data can also be accessed in two other ways either by using File Transfer Protocol (FTP) or by working with each machine remotely with Secure Shell (SSH). While working with web faction, all the data base and customer activities are stored under the umbrella of an account which runs by unique username and password which is provided by the web server.

This unique combination of username and password is identified by the control panel while logging in any account. An individual while accessing internet, is able to access websites in its own domain only. A domain name is a unique and human-readable label and all browsers can use that label to locate a particular web device on internet. Most of the time, the web faction accounts use a single server, where all the applications, databases and files are safely stored and served from a single assigned server. However, in case of heavy load, the new temporary server is also installed so as to distribute the load consumption and memory usage. This installation of new reliable temporary server helps in balancing the load in particular when the traffic to a particular server exceeds a threshold level. Moreover, some applications on server also demand more memory usage which can be completed by the installation of the new server.

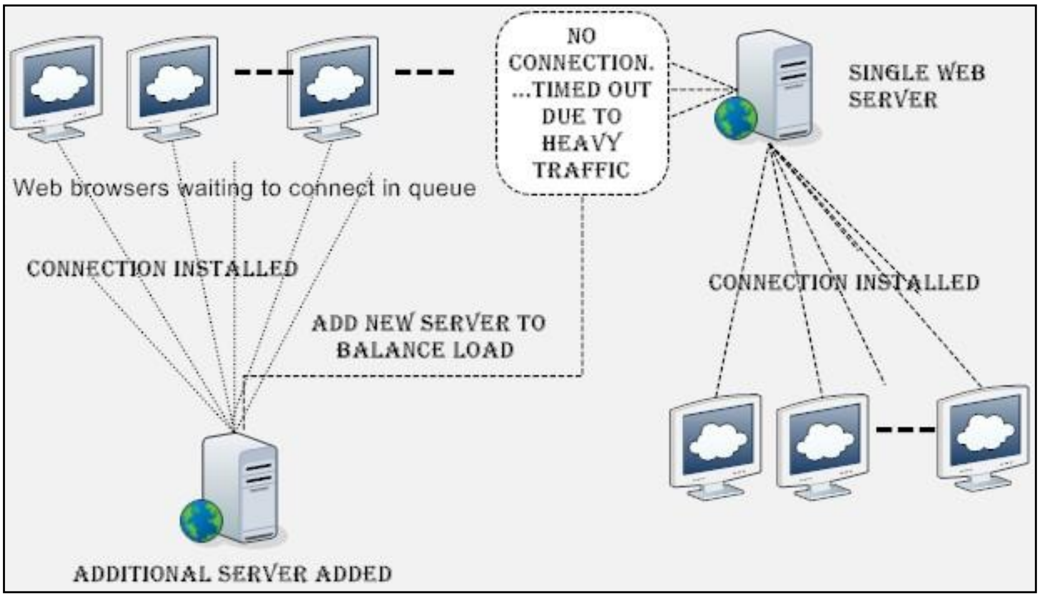


Fig. 9.1: Pictorial presentation of addition of temporary servers in the web faction system

Each incoming request is directed to the new server, if buffer of first server is already full with the maximum number of requests. Also, some requests are directed to buffer (orbit) so as to retry if first server is in busy state. The pictorial representation of web server supported by additional temporary server is shown in fig. 9.1.

9.2.4 Governing Equations

We frame steady state equations governing the model by using appropriate transition rates of birth-death process. Chapman-Kolmogorov equations corresponding to different system states are formulated as:

Retrial state

$$(\lambda+\gamma)P_{0,1,0}=0 \tag{9.1}$$

$$(\lambda+\gamma)P_{0,i,0}=\lambda P_{0,i-1,0},\ (1\le i\le K-1) \tag{9.2}$$

$$\gamma P_{0,K,0}=\lambda P_{0,K-1,0} \tag{9.3}$$

Busy State

$$(\lambda+\alpha_1)P_{1,0,0}=\mu_1 P_{1,1,0}+\beta_1 P_{2,0,0}+\mu_2 P_{1,0,1} \tag{9.4}$$

$$(\lambda+\alpha_1+\mu_1)P_{1,i,0}=\mu_1 P_{1,i+1,0}+\beta_1 P_{2,i,0}+\mu_2 P_{1,i,1}+\lambda P_{1,i-1,0}+\gamma P_{0,i,0},\ (1\le i\le K-1) \tag{9.5}$$

$$(\lambda+\alpha_1+\mu_1)P_{1,K,0}=\beta_1 P_{2,K,0}+\mu_2 P_{1,K,1}+\lambda P_{1,K-1,0}+\gamma P_{0,K,0} \tag{9.6}$$

$$(\lambda+\alpha_1+\mu_2)P_{1,0,j}=\mu_1 P_{1,1,j}+\beta_1 P_{2,0,j}+\mu_2 P_{1,1,j+1},\ j\ge 1 \tag{9.7}$$

$$(\lambda+\alpha_1+\mu_1+\mu_2)P_{1,i,j}=\mu_1 P_{1,i+1,j}+\beta_1 P_{2,i,j}+\mu_2 P_{1,i,j+1}+\lambda P_{1,i-1,j},\ (1\le i\le K-1),\ j\ge 1 \tag{9.8}$$

$$(\lambda+\alpha_1+\mu_1+\mu_2)P_{1,K,j}=\beta_1 P_{2,K,j}+\lambda p P_{1,K,j-1}+\mu_1 P_{1,K-1,j},\ \ j\ge 1 \tag{9.9}$$

Repair State

$$(\lambda+\beta_1)P_{2,0,j}=\alpha_1 P_{1,0,j},\ j\ge 0 \tag{9.10}$$

$$(\lambda+\beta_1)P_{2,i,j}=\alpha_1 P_{1,i,j}+\lambda P_{2,i-1,j},(1\le i\le K-1),\ j\ge 0 \tag{9.11}$$

$$(\lambda+\beta_1)P_{2,K,j}=\alpha_1 P_{1,K,j}+\lambda P_{2,K-1,j},\ j\ge 0 \tag{9.12}$$

In order to determine the solution of eqs (9.1) - (9.12), we employ matrix geometric method as explained in the next section 9.3.

9.3 THE ANALYSIS

The matrix geometric method (cf. Neuts, 1981) can be used to solve the stationary state probabilities for the vector space Markov process with repetitive structure. Therefore, in order to find the solution for the system of equations constructed in section 9.2.4, we employ this technique to determine the associated state probability vector.

Matrix Geometric Method

The matrix geometric method to determine the probability vector is applicable for the system of equations whose transition matrices have special block structure with repetition of elements of sub matrices. The concerned model can be structured as a square matrix of infinite dimension that converges to finite dimension matrix using the minimal matrix to get recursive relation of probability vectors. The above set of eqs (9.1)-(9.12) can be written in matrix form as $\boldsymbol{\pi}\mathbf{Q}=\mathbf{0}$, where $\mathbf{Q}$ is the infinitesimal generator of the continuous time Markov chain and '$\mathbf{0}$' is a zero column vector of suitable dimension. Also, let $\boldsymbol{\pi}=(\boldsymbol{\pi_0},\boldsymbol{\pi_1},\boldsymbol{\pi_2},\boldsymbol{\pi_3},\ldots)$ be the vector defining the steady state probabilities of all the governing states of the retrial queueing system under consideration. The matrix $\mathbf{Q}$ can be given in partition form as

$$\mathbf{Q}=\begin{bmatrix} \mathbf{F_0} & \mathbf{F_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\ldots \\ \mathbf{F_2} & \mathbf{F_3} & \mathbf{F_4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\ldots \\ \mathbf{0} & \mathbf{F}_5 & \mathbf{F_3} & \mathbf{F_4} & \mathbf{0} & \mathbf{0} & \mathbf{0}\ldots \\ \mathbf{0} & \mathbf{0} & \mathbf{F}_5 & \mathbf{F_3} & \mathbf{F_4} & \mathbf{0} & \mathbf{0}\ldots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{F}_5 & \mathbf{F_3} & \mathbf{F_4} & \mathbf{0}\ldots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{9.13}$$

$$\mathbf{F_0}=\begin{bmatrix} \mathbf{A_0} & \mathbf{B_0} \\ \mathbf{0} & \mathbf{A_1} \end{bmatrix}_{(2n+1)\times(2n+1)} \quad \mathbf{F_1}=\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B_1} & \mathbf{C_1} \end{bmatrix}_{(2n+2)\times(2n+2)} \quad \mathbf{F_2}=\begin{bmatrix} \mathbf{0} & \mathbf{D_1} \\ \mathbf{0} & \mathbf{G_1} \end{bmatrix}_{(2n+2)\times(2n+1)}$$

$$\mathbf{F_3}=\begin{bmatrix} \mathbf{E_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H_1} \end{bmatrix}_{(2n+2)\times(2n+2)}, \mathbf{F_4}=\begin{bmatrix} \mathbf{C_1} & \mathbf{0} \\ \mathbf{B_1} & \mathbf{0} \end{bmatrix}_{(2n+2)\times(2n+2)}, \mathbf{F}_5=\begin{bmatrix} \mathbf{0} & \mathbf{D_1} \\ \mathbf{0} & \mathbf{G_1} \end{bmatrix}_{(2n+2)\times(2n+2)}$$

$$\mathbf{A_1}=\begin{bmatrix} -(\lambda+\alpha_1) & \lambda & 0 & 0\ldots.. & 0 & \\ \mu_1 & -(\lambda+\alpha_1+\mu_1) & \lambda & 0\ldots.. & 0 & \\ 0 & \mu_1 & -(\lambda+\alpha_1+\mu_1) & \lambda & 0 & \\ \vdots & \ddots & & \ddots & \ddots & \\ & & \mu_1 & -(\lambda+\alpha_1+\mu_1) & \lambda & \\ & & & \mu_1 & -(\lambda+\alpha_1+\mu_1) & \end{bmatrix}_{(n+1)\times(n+1)}$$

$$\mathbf{A_0}=\begin{bmatrix} -(\lambda+\gamma) & \lambda & 0 & 0\ldots.. & 0 \\ 0 & -(\lambda+\gamma) & \lambda & 0\ldots.. & 0 \\ 0 & 0 & -(\lambda+\gamma) & \lambda & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & & & -(\lambda+\gamma) & \lambda \\ 0 & 0 & 0 & & -\gamma \end{bmatrix}_{(n)\times(n)} \quad \mathbf{B_0}=\begin{bmatrix} 0 & \gamma & 0 & 0 & 0\ldots.. & 0 \\ 0 & 0 & \gamma & 0 & 0\ldots.. & 0 \\ 0 & 0 & 0 & \gamma & 0\ldots.. & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & & \ddots & \gamma & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma \end{bmatrix}_{(n)\times(n+1)}$$

$$\mathbf{E_1}=\begin{bmatrix} -(\lambda+\beta_1) & \lambda & 0 & 0\ldots.. & 0 \\ 0 & -(\lambda+\beta_1) & \lambda & 0\ldots.. & 0 \\ 0 & 0 & -(\lambda+\beta_1) & \lambda & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & & & -(\lambda+\beta_1) & \lambda \\ 0 & 0 & 0 & 0 & -(\lambda+\beta_1) \end{bmatrix}_{(n+1)\times(n+1)} \quad \mathbf{C_1}=\begin{bmatrix} 0 & 0 & 0\ldots.0 & 0 \\ \vdots & \ddots & \vdots & 0 \\ \vdots & & \ddots\ \vdots & 0 \\ 0 & 0\ldots. & 0 & \lambda \end{bmatrix}_{(n+1)\times(n+1)}$$

$$\mathbf{B_1}=\alpha_1\mathbf{I}_{(n+1)}, \mathbf{D_1}=\beta_1\mathbf{I}_{(n+1)}, \mathbf{G_1}=\mu_2\mathbf{I}_{(n+1)}$$

where, $\mathbf{I}_{(n+1)}$ is the identity matrix of order (n+1). The normalizing condition is represented by $\boldsymbol{\pi}\mathbf{e}=1$, where '**e**' is a column vector of suitable dimension with all its entries as 1. In order to determine the probability vector, we partition vector $\boldsymbol{\pi}$ conformably with the blocks of matrix **Q** as

$$\boldsymbol{\pi}_0 = (P_{0,0,0}, P_{1,0,0}; P_{0,1,0}, P_{1,1,0}; ...; P_{0,K,0}, P_{1,K,0}); \; P_{0,0,0} = 0 \tag{9.14}$$

$$\boldsymbol{\pi}_j = (P_{2,0,j-1}, P_{1,0,j}; P_{2,1,j-1}, P_{1,1,j}; ...; P_{2,K,j-1}, P_{1,K,j}); \; j \geq 1 \tag{9.15}$$

Using matrix geometric approach (cf. Neuts, 1981), we have

$$\boldsymbol{\pi}_\mathbf{j} = \boldsymbol{\pi}_\mathbf{1} \mathbf{R}^{j-1}, (j \geq 2) \tag{9.16}$$

where, **R** is the minimal non-negative matrix known as rate matrix.
The balance equation for the repeating states is

$$\boldsymbol{\pi}_\mathbf{j-1}\mathbf{F_4} + \boldsymbol{\pi}_\mathbf{j}\mathbf{F_3} + \boldsymbol{\pi}_\mathbf{j+1}\mathbf{F_5} = \mathbf{0}; \quad j = 2, 3, 4, ... \tag{9.17}$$

The balance equations for the boundary states are

$$\boldsymbol{\pi}_\mathbf{0}\mathbf{F_0} + \boldsymbol{\pi}_\mathbf{1}\mathbf{F_2} = \mathbf{0} \tag{9.18}$$

$$\boldsymbol{\pi}_\mathbf{0}\mathbf{F_1} + \boldsymbol{\pi}_\mathbf{1}\mathbf{F_3} + \boldsymbol{\pi}_\mathbf{2}\mathbf{F_5} = \mathbf{0} \tag{9.19}$$

The value of $\boldsymbol{\pi}_j, (j \geq 2)$ is a probability function of the transition between the states with *j*-1 queued customers and states with *j* queued customers. Using (9.16) and (9.17), we have

$$\begin{aligned} &\boldsymbol{\pi}_\mathbf{1}\mathbf{R}^{\mathbf{j-2}}\mathbf{F_4} + \boldsymbol{\pi}_\mathbf{1}\mathbf{R}^{\mathbf{j-1}}\mathbf{F_3} + \boldsymbol{\pi}_\mathbf{1}\mathbf{R}^{\mathbf{j+1}}\mathbf{F_5} = \mathbf{0}; \quad j = 2, 3, 4, ... \\ &\Rightarrow \mathbf{F_4} + \mathbf{RF_3} + \mathbf{R^2F_5} = \mathbf{0} \end{aligned} \tag{9.20}$$

On solving (9.20), we get the rate matrix **R**, which can be further used to compute steady state probabilities for the repeating states. Now, using (9.16) for j=2 in (9.19), we get

$$\boldsymbol{\pi}_\mathbf{0}\mathbf{F_1} + \boldsymbol{\pi}_\mathbf{1}(\mathbf{F_3} + R\mathbf{F_5}) = \mathbf{0} \tag{9.21}$$

Eqs (9.18) and (9.21) can be further written in matrix form as $\begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{bmatrix} \mathbf{F_0} & \mathbf{F_2} \\ \mathbf{F_1} & \mathbf{F_3} + \mathbf{RF_5} \end{bmatrix} = 0$

In order to find $\boldsymbol{\pi}_\mathbf{0}$, we use normalizing condition

$$\boldsymbol{\pi}_\mathbf{0} e + \boldsymbol{\pi}_\mathbf{1} \left(\sum_{j=0}^{\infty} \mathbf{R}^j \right) \mathbf{e} = 1 \; . \tag{9.22}$$

The eigenvalues of **R** lie inside the unit circle which means that (I-**R**) is non-singular and hence we have

$$\left(\sum_{j=0}^{\infty} \mathbf{R}^j \right) = (I - \mathbf{R})^{-1}. \tag{9.23}$$

Since, it is not an easy job to obtain rate matrix and steady state probabilities using huge algebraic manipulations. Therefore, in order to obtain steady state probabilities for various states of the server, we use numerical approach based on MGM. For this purpose, we first obtain the rate matrix **R** by iterative procedure as **R** (0) =0, and move further with successive approximation using

$$\mathbf{R}(n+1) = -[\mathbf{F}_4 + \mathbf{R}^2\mathbf{F}_5]\mathbf{F}_3^{-1}, \quad n \geq 0 \tag{9.24}$$

Since, $-\mathbf{F}_3^{-1}$ is a non-negative matrix, therefore it can be concluded that the sequence $\{\mathbf{R}(n)\}_n$ is a non-decreasing sequence which converges monotonically to a non-negative matrix **R**. The stage when $\|\mathbf{R}(n+1) - \mathbf{R}(n)\| < \varepsilon$ (ε is a constant) is satisfied, we terminate the solution process and obtain **R** which helps further in determining the steady state probabilities numerically.

9.4 PERFORMANCE MEASURES

In this section, we derive various performance measures in terms of steady state probabilities as follows:

(A) Server State Probabilities

The probabilities of the server being present in different states are expressed as:

- Probability of the primary server being in retrial state is framed as:

$$P_r = \sum_{n=0}^{K} P_{0,n,0} \tag{9.25}$$

- Probability of the primary server being busy is:

$$P_{B_1} = \sum_{n=1}^{K} P_{1,n,0} \tag{9.26}$$

- Probability that both primary and temporary servers are busy in servicing is:

$$P_{B_2} = \sum_{j=1}^{\infty} \sum_{n=1}^{K} P_{1,n,j} \tag{9.27}$$

- Probability of the primary server being in broken down state is:

$$P_D = \sum_{j=0}^{\infty} \sum_{n=0}^{q} P_{2,n,j} \tag{9.28}$$

(B) Queue Length

The expected number of customers at various states of the server can be obtained in terms of steady state probabilities as follows:

- The expected number of customers in the retrial orbit, is:

$$E[N_r] = \sum_{n=1}^{K} nP_{0,n,0} \tag{9.29}$$

- The expected number of customers in the busy state when primary server is on, is:

$$E[N_1] = \sum_{n=1}^{K} nP_{1,n,0} \tag{9.30}$$

- The expected number of customers in the busy state when both the servers are busy in rendering the service to the customers, is:

$$E[N_2] = \sum_{j=1}^{\infty}\sum_{n=1}^{K} nP_{1,n,j} + \sum_{j=1}^{\infty} jP_{1,K,j},\ j \geq 1 \tag{9.31}$$

- The expected number of customers when primary server is in the broken state, is:

$$E[N_d] = \sum_{n=0}^{K} nP_{2,n,j},\ j \geq 0 \tag{9.32}$$

- The expected number of customers in the system, is:

$$E[N] = E[N_r] + E[N_1] + E[N_2] + E[N_d] \tag{9.33}$$

(C) Throughput

In terms of steady state probabilities, throughput is given by

$$TP = \mu_1 \sum_{n=0}^{K} P_{1,n,0} + (\mu_1 + \mu_2)\sum_{j=1}^{\infty}\sum_{n=0}^{K} P_{1,n,j} \tag{9.34}$$

(D) Expected Delay

The expected delay experienced by the customers in the system is

$$E[D] = E[N] / TP \tag{9.35}$$

(E) Waiting Time

The average waiting time of the customers in the system is expressed as

$$E[W] = \frac{E[N]}{\lambda} \tag{9.36}$$

9.5 COST FUNCTION

We evaluate here the cost function in terms of various performance measures and associated cost elements so as to study the system in monetary terms. The cost function is constructed as:

$$TC(\mu_1, \mu_2) = C_{\text{h}} E[N] + C_1 \mu_1 P_{B_1} + C_2 (\mu_1 + \mu_2) P_{B_2} + C_{\text{d}} P_D + C_{\text{r}} \gamma \tag{9.37}$$

where,

C_1 : Fixed cost per unit time when the first server is busy;

C_2 : Fixed cost per unit time when the both the servers are busy;

C_h : Holding cost per unit time for each customer present in the system;

C_d : Repair cost incurred per unit time when the primary server is broken down;

C_r : Fixed cost incurred per unit time when a customer from the orbit retries for the service.

We further aim to find total optimal cost by determining the optimal service rates for both the servers using 'Direct search approach'. The optimization problem is mathematically formulated as:

$$\text{(OP): } TC(\mu_1, \mu_2) = \text{Minimize}\, TC(\mu_1, \mu_2) \quad (9.38)$$

In order to analyze the nature and sensitivity of the cost function towards various parameters, we give illustrations as given further in sections 9.6 and 9.7.

9.6 NUMERICAL ILLUSTRATION

In this section, we perform numerical simulation so as to compute the rate matrix and other performance indices. The computer program for this purpose is executed in MATLAB software using the set of default parameters as $\lambda = 0.5, \mu_1 = 4, \mu_2 = 5, \gamma = 0.5, \alpha_1 = 0.1, \beta_1 = 1, K = 5.$ Various sub matrices and rate matrix **R** are computed as:

$$\mathbf{F_0} = \begin{bmatrix} -1 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & -1 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & -1 & 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & -0.5 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & -0.6 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -4.6 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & -4.6 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & -4.6 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0.1 \end{bmatrix}, \quad \mathbf{F_1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}$$

$$\mathbf{F_2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix} \mathbf{F_3} = \begin{bmatrix} -1.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -5.6 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & -9.6 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & -9.6 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & -9.6 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & -9.6 \end{bmatrix},$$

$$
\mathbf{F}_4 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\
0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix},
\quad
\mathbf{F}_5 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5
\end{bmatrix}
$$

$$
\mathbf{R} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.33 & 0 & 0 & 0 & 0 & 0 \\
0.33 & 0.11 & 0.037 & 0.012 & 0.004 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.33 & 0.11 & 0.037 & 0.012 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.33 & 0.11 & 0.037 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.33 & 0.11 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.33 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 4 & -10 & 0.5 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & -10 & 0.5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & -10
\end{bmatrix}
$$

The various performance indices are obtained as:

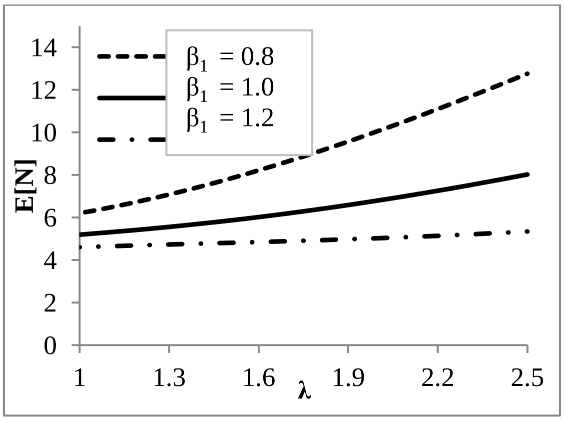
E[N] = 4.60, E[W]=9.21, TP=10.12, P_r= 0.2759, P_D=0.4568, E[D]=0.44, TC=270.82 units.

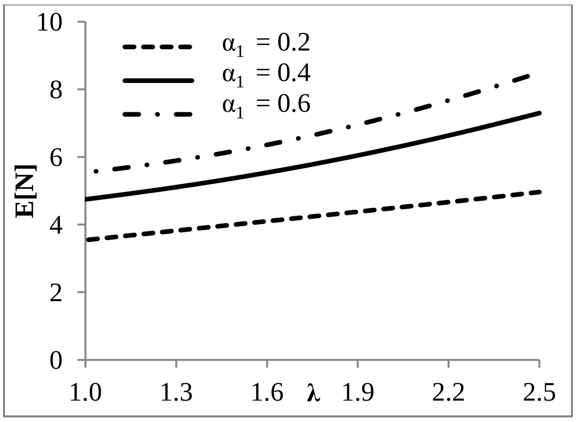
9.7 SENSITIVITY ANALYSIS

In the present section, the sensitivity of various performance indices towards different parameters has been analyzed. To study the system performance measures, we set the default parameters as: $\lambda = 0.5, \mu_1 = 4, \mu_2 = 5, \gamma = 0.5, \alpha_1 = 0.1, \beta_1 = 1, K = 5.$ Based on the computational results obtained, the effects of parameters on various measures are interpreted as follows:

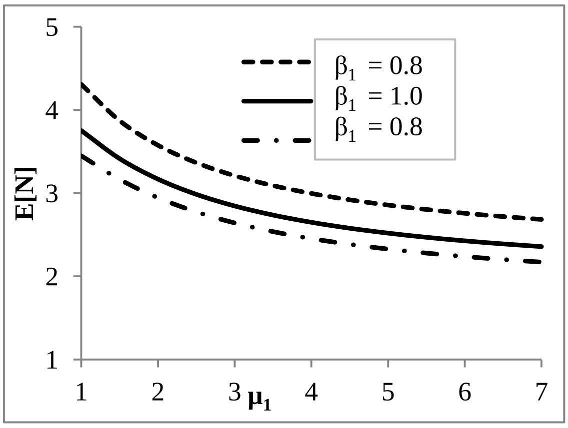
(A) Queue Length of the System

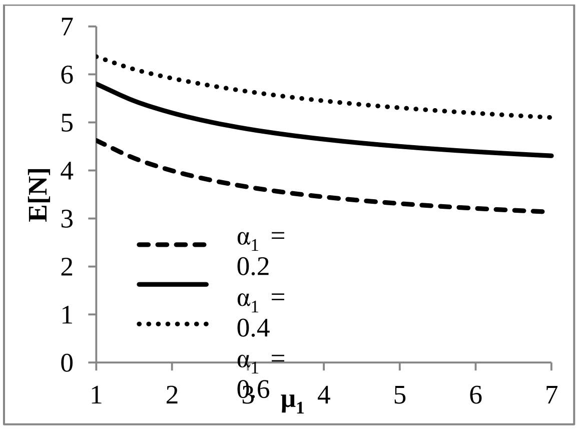
The first server can serve 'K' number of customers and the rest of the customers who arrive in the system have to accept services from the second server. The sensitivity of E [N] towards arrival rate λ and service rate μ_1 is shown in figs 9.2(a-b). We notice that the queue length of the system increases (decreases) with an increase in λ (μ_1). This is due to the fact that an increase in the arrival rate automatically increases the customers in the system and thus the need of installing second server increases proportionally.





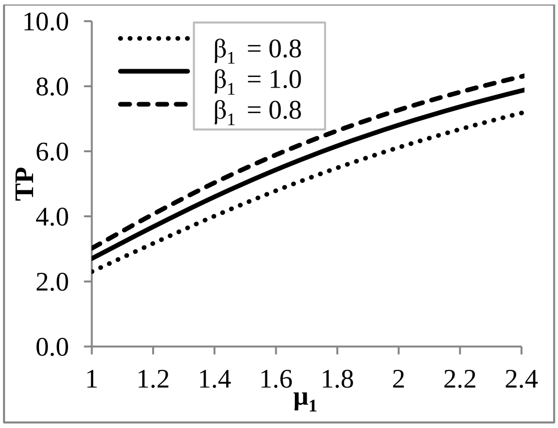
(a) **(b)**

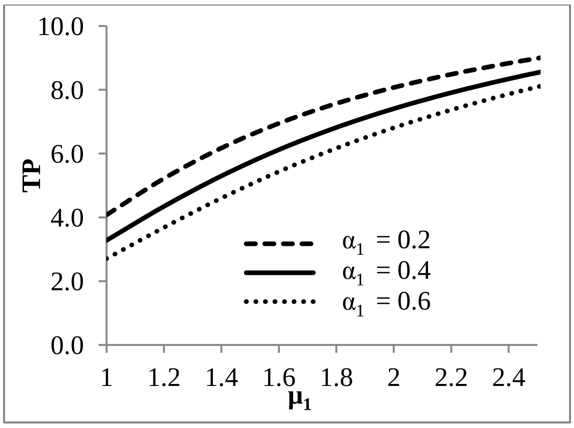




(c) **(d)**

Fig. 9.2: Effect of various parameters on the expected number of customers in the system E[N]





(a) **(b)**

Fig. 9.3: Effect of various parameters on the throughput TP of the system

Table 9.1: Effect of μ_2 and λ on the system performance indices

			μ_2=4				μ_2=6			
λ	**α_1**	**β_1**	**E[D]**	**E[N]**	**E[W]**	**TP**	**E[D]**	**E[N]**	**E[W]**	**TP**
0.5	0.2	**0.8**	0.40	3.41	6.82	8.46	0.34	3.44	6.89	10.09
		1.0	0.36	3.03	6.07	8.33	0.31	3.05	6.11	9.90
		1.2	0.34	2.78	5.56	8.21	0.29	2.80	5.59	9.73
	0.4	**0.8**	0.53	4.73	9.46	8.99	0.44	4.79	9.59	10.89
		1.0	0.47	4.18	8.37	8.84	0.40	4.23	8.46	10.66
		1.2	0.44	3.80	7.60	8.68	0.37	3.83	7.67	10.42
0.7	0.2	**0.8**	0.40	3.68	5.26	9.32	0.33	3.75	5.36	11.23
		1.0	0.35	3.25	4.64	9.33	0.29	3.30	4.71	11.22
		1.2	0.32	2.97	4.24	9.30	0.27	3.01	4.31	11.16
	0.4	**0.8**	0.53	5.06	7.23	9.55	0.45	5.21	7.44	11.63
		1.0	0.46	4.42	6.31	9.64	0.39	4.53	6.48	11.71
		1.2	0.42	4.00	5.71	9.63	0.35	4.09	5.84	11.67
0.9	0.2	**0.8**	0.41	3.98	4.42	9.80	0.35	4.11	4.57	11.89
		1.0	0.35	3.44	3.83	9.96	0.29	3.55	3.95	12.07
		1.2	0.31	3.13	3.48	10.02	0.27	3.22	3.58	12.13
	0.4	**0.8**	0.56	5.44	6.04	9.65	0.48	5.71	6.35	11.79
		1.0	0.46	4.64	5.15	9.98	0.40	4.85	5.39	12.18
		1.2	0.41	4.15	4.61	10.13	0.35	4.33	4.82	12.34

However, on the other hand as per our expectation, the number of customers in the system reduces on speeding up the service rate. In fig. 9.2(c), we see that an increase in the breakdown rate of first server (α_1) highly affects the number of customers in the system.

An increase in the breakdown rate is highly responsible for the installation of second server due to increased congestion in the system. But, on the other hand an increase in the repair rate β_1 (see fig. 9.2(d)) helps in reducing the congestion in the system as seen from decrement in the queue length.

(B) Throughput

Throughput is a direct measure to study the efficiency of any queueing model. Figs 9.3(a-b) exhibit the effect of service rate (μ_1) on the throughput (TP) of the system. An increase in μ_1 results in an increase in the throughput. Moreover, an increase in the repair rate also increases the number of served customers while the TP decreases with an increase in the breakdown rate of the server. The effect of μ_2 on TP is tabulated in table 9.1.

(C) Waiting Time and Expected Delay

The total time that a customer spends in the system and expected delay in the service are the key factors that affect the performance and efficiency of any waiting system. Table 9.1 depicts the effect of parameters λ, α_1, β_1 and μ_2 on the waiting time E[W] and expected delay E[D], expected number of customers E[N] and throughput TP of the system. It is noticed that E[D], E[N] and E[W] decrease with the increase in repair rate (β_1) but increase with an increase in the breakdown rate α_1. However, E[D] and E[W] reveal lower values for the higher values of service rates.

(D) Effect of Service Rates on the Cost of the System

In our model, as the customers load in the system increases beyond a pre-specified limit (K), the second server is installed so as to serve the customers. The service rate of second server significantly influences the performance of the system as well as the total cost spent on the system for the servicing of the customers. To visualize the nature of cost function towards the service rate of both the servers, we consider the following illustrations:

Illustration 9.1: Consider the additional server retrial queueing system that can accommodate a queue length of maximum K = 5 customers with the first server. The system works under the set of default parameters assumed as $\lambda = 0.5,\ \mu_1 = 1.5, \gamma = 0.5, \alpha_1 = 0.1, \beta_1 = 1,\ C_h = 10,\ C_r = 40,\ C_1 = 10,\ C_2 = 25,\ C_d = 20.$

To obtain the optimal service rate ($\mu_2{*}$), we vary μ_2 for feasible range say (0.5: 0.05: 2.0) and search for the optimal point.

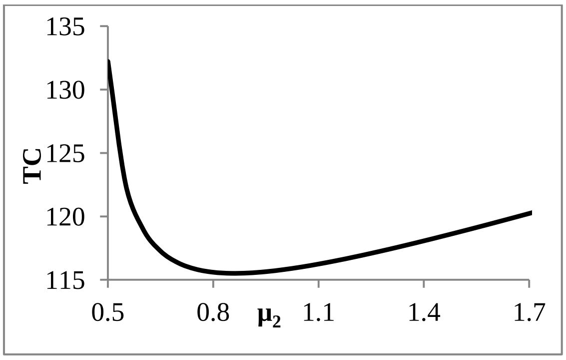


Fig. 9.4: Convexity of TC with μ_2

Table 9.2(a): Effect of arrival rates (λ) on μ_2* and TC (μ_1, μ_2*)

	λ= 0.5			**λ= 0.7**			**λ= 0.9**		
α_1	**0.1**	**0.3**	**0.5**	**0.1**	**0.3**	**0.5**	**0.1**	**0.3**	**0.5**
μ_2*	0.85	1.17	1.5	1.77	1.49	1.88	1.68	2.21	2.75
TC(μ_1, μ_2*)	115.51	133.53	149.96	132.47	151.48	168.09	129.70	153.57	173.54
TP	3.44	3.24	3.10	4.85	3.55	3.29	4.53	4.64	4.70
E[N]	2.50	3.68	4.56	2.52	3.91	4.77	1.87	3.32	4.40

Table 9.2(b): Effect of breakdown rate α_1 on μ_2* and TC (μ_1, μ_2*)

	μ_1= 1			μ_1= 2			μ_1= 3		
α_1	**0.1**	**0.3**	**0.5**	**0.1**	**0.3**	**0.5**	**0.1**	**0.3**	**0.5**
μ_2*	0.76	1.13	1.54	0.89	1.20	1.50	0.93	1.25	1.54
TC(μ_1,μ_2*)	110.29	128.40	146.01	118.40	137.45	154.04	122.12	144.01	162.16
TP	2.65	2.53	2.44	4.00	3.73	3.53	4.95	4.54	4.24
E[N]	2.70	3.91	4.80	2.34	3.51	4.39	2.14	3.31	4.19

It is clear from the fig. 9.4 that the cost function is convex in nature as well as unimodular within the pre-specified range of μ_2. The minimum cost is obtained at μ_2*=0.85 units with TC (μ_1, μ_2*) = 115.51 units. Tables 9.2(a-b), display the optimal service rate (μ_2*) with the variation in various parameters. The effects of arrival rate (λ), service rate (μ_1) and breakdown rate (α_1) have been presented in tables 9.2(a-b) on the optimal service rate (μ_2*) of the second server and other corresponding metrics namely TC (μ_1, μ_2*), TP and E[N]. It is noticed that the μ_2* is sensitive to the variation in λ with α_1; an increase in λ affects the optimal μ_2* as well as increases the total cost of the server. An increment in μ_1 also significantly affects the optimal μ_2* and increases the total cost of the system which is quite obvious.

Illustration 9.2: Consider the retrial queueing system with two servers in the system with $\lambda = 0.5$, $\gamma = 0.5$, $\alpha_1 = 0.5$, $\beta_1 = 1$, C_h=10, C_r=40, C_1=10, C_2=25, C_d=20, K=5. Now, we need to find optimal service rates (μ_1*, μ_2*) and the corresponding optimal cost TC (μ_1*, μ_2*). As shown in fig. 9.5, we vary μ_1 for interval (1.8≤ μ_1≤ 2.6) and μ_2 for interval (1≤ μ_2 ≤ 2.5) and search for the optimal pair (μ_1*, μ_2*).

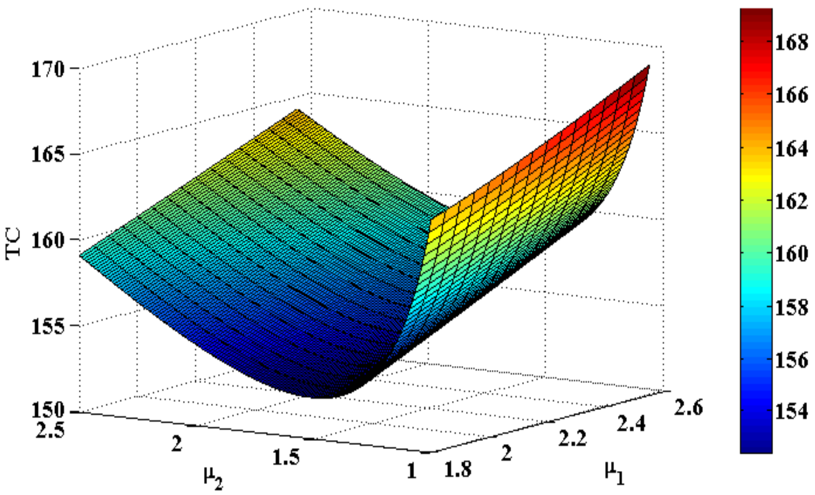


Fig. 9.5: Determination of optimal pair (μ_1*, μ_2*) and TC (μ_1*, μ_2*)

The optimal cost is obtained as **TC (μ_1*, μ_2*) = 152.40** units at **(μ_1*, μ_2*) = (1.85, 1.51**).

(E) Effect of Retrial Rate γ on the Total Cost

The retrial rate is one of the key system descriptor that significantly affects the system performance indices and total cost of the system. The number of customers in the retrial orbit and the rate with which they retry also affects the system expenses. We study the sensitivity of total cost TC (μ_1*, μ_2*) towards retrial rate by considering the

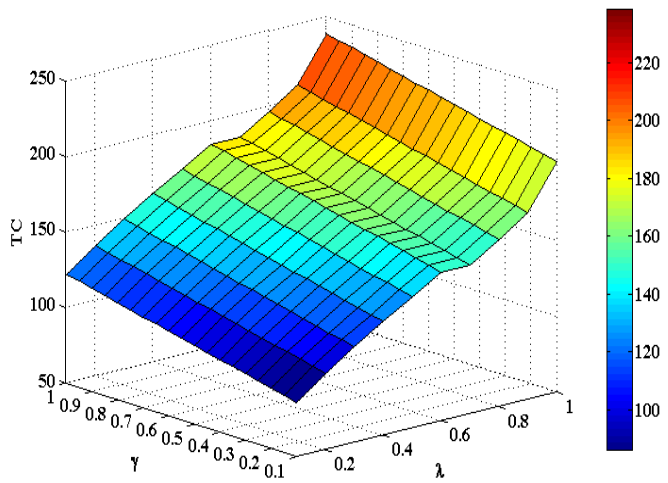


Fig. 9.6 (a): Variation in TC with α_1 and γ

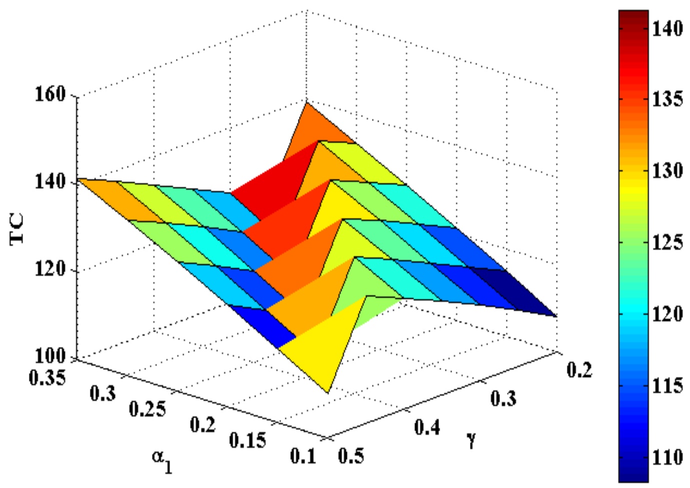


Fig. 9.6 (b): Variation in TC with λ and γ

optimal service rates as obtained in illustration 9.2. For the set of default parameters as taken in illustration 9.2 with $\mu_1 = 1.85, \mu_2 = 1.51$, we plot total cost of the system in figures 9.6(a-b) to explore the combined effect of retrial rate γ along with α_1 and β_1. The total cost TC of the system increases continuously with an increment in both retrial rate and repair rate as displayed in fig. 9.6(a). This is due to the fact that with every reattempt performed by the customer, a cost is incurred. It is noticed that the TC also increases with an increase in the breakdown rate.

9.8 DISCUSSION

In this chapter, markovian retrial queueing system is analyzed using matrix geometric approach. The cost function formulated has been further used to compute the optimal values of service rates for both the servers by taking numerical illustrations. The model under consideration seems to be applicable to a variety of real life congestion situations where usually second server is installed depending on the increase in the load of primary server. The application of the model to web faction has also been discussed with the view point of requirement of the additional temporary server due to load increment. Overall, we can conclude that

- An increase in the breakdown rate and the arrival rate is responsible for the increase in the congestion in the system as such single server is not sufficient to serve the customers.
- TP increases with the high values of μ_2 and λ. Hence, the higher values of service rates of both the servers may be helpful in maintaining a highly efficient system.
- A speedy service and higher repair rate results in the reduction in the waiting and delay experienced by the customers in the system.
- The total cost TC of the system increases continuously with an increment in both retrial rate and repair rate.
- This study may be useful for the system designers and decision makers to have a better idea to trade off between cost and delay of the system in particular when traffic load in the system is sufficiently high.

CHAPTER 10

DOUBLE ORBIT RETRIAL QUEUES WITH PRIORITY

10.1 INTRODUCTION

The majority of research in retrial queues deals with the systems with homogenous customers. There exist systems which allow the arrival of heterogeneous customers with variable rates and are even served with different service rates. In this chapter, we focus on the formation of double orbits by different class of customers. In previous chapters, we have studied single class of customers in the retrial queue. Here, we consider broader class of customers with different priorities and hence different waiting spaces (i.e. retrial orbits). Domenech-Benlloch *et al.* (2009) investigated retrial queueing model with two types of orbits for different class of customers using extrapolation. They considered both the orbits of infinite capacity. Following this work, Avreachenkov et al. (2010) considered single server retrial queueing model where two types of customers join different class of orbits if server is not available. They considered two orbits in which one orbit was of infinite capacity and other one of finite capacity. Since all the algorithmic schemes assume truncations in deriving the approximate results, therefore we consider a system in which different class of customers arrive with different arrival rates and are kept in separate buffers of fixed capacity.

The present chapter deals with finite double orbits retrial queue along with unreliable server and priority customers. Moreover, the broken down server is repaired following threshold recovery for both priority as well as non priority customers. Non priority customers are served only if no priority customers are present in the system. The transient solution of the model has been explored using matrix method. The cost function has been optimized to determine the optimal parameters. The rest of the chapter is organized in the following manner. Section 10.2 presents the description of the model along with underlying assumptions and notations. The governing equations are framed in section 10.3. Various performance measures have been obtained in section 10.4. An application on cellular mobile network has been developed in section 10.5. The numerical simulation has been carried out in section 10.6. Section 10.7 is devoted to the cost optimization of the concerned retrial queueing system. Finally, the results and findings are discussed in section 10.8.

10.2 MODEL DESCRIPTION

We consider an unreliable single server finite retrial queueing model with two types of customers; priority and non-priority customers with double orbits. The basic operation of the model can be described as:

(i) Arrival and Retrial Process: Two class of customers namely priority and non-priority customers arrive in the system. The priority (non-priority) customers follow Poisson process with mean arrival rate λ_1 (λ_2). On finding the server idle or unavailable for the service, the customers on their arrival may either join the queue in the front of the server or they can wait for their turn in their respective retrial orbits. The queue in the front of server can hold a maximum of L priority and K non-priority customers. Further, we assume that the capacity of retrial orbit 1(orbit 2) is of L-1 (K-1) priority (non-priority) customers.

The priority (non priority) customers retry for their service from their respective orbits; the retrial time is exponentially distributed with rate γ_1 (γ_2). The retrial process occurs only when single type of customers (either priority or non-priority) is present in the system. However, if both types of customers are present in the system then the priority customers are served like a classical queue and no retrial phenomenon takes place in such case.

(ii) Service Process: All the customers are served following the first come first serve (FCFS) service discipline. The priority and non priority customers are served according to exponential distribution with rates $\mu_1 \text{ and } \mu_2$, respectively. The priority customers are served prior to the non-priority customers. The server after serving the last priority customer present in the queue automatically starts the servicing of non-priority customers waiting for the service. There is no retrial mechanism in this case.

(iii) Breakdown state: The server is unreliable and may break down while serving the customers. The server failures occur in Poisson fashion with rates α_1 (α_2) while servicing priority (non priority) customers. The server failures occur while the server is busy in serving either type of the customer. No break down occurs during retrial and idle states. In case when the server breakdown occurs, the service of the customer already in the service is resumed and continued as soon as the repair process of the server is completed.

Moreover, the system transition from repair state is allowed only into the busy state and not in the retrial state.

(iv) Repair process: The server may breakdown while serving either priority or non-priority customers. The broken down server is sent for the repair so as to regain its functionality. The repair process is completed according to the following threshold recovery policy:

a) If the server breakdown occurs while serving the priority customers, then the repair process starts if a sufficient number of priority customers say q_1 ($1 \leq q_1 \leq L-1$) are already present in the system.

b) If the server breakdown occurs while serving the non- priority customers, then the repair process starts if a threshold number of non- priority customers say q_2 ($1 \leq q_2 \leq K-1$) are available in the system.

The various notations used to formulate the model are summarized below:

λ_1 (λ_2): Arrival rate for priority (non priority) customers

μ_1 (μ_2): Service rate for priority (non priority) customers

γ_1 (γ_2): Retrial rate for priority (non priority) customers

α_1 (α_2): Breakdown rate for the server while servicing priority (non priority) customers

q_1 (q_2): Threshold limit on the number of priority (non-priority) customers for the repair

β_1 (β_2): Repair rate for the server broken down during the service of priority (non priority) customers.

10.3 MATHEMATICAL FORMULATION OF THE MODEL

The retrial queueing model under consideration is Markovian. In order to obtain the solution of the system we first develop mathematical model for the system using notations and assumptions discussed in previous section 10.2. Chapman Kolmogorov equations are established to obtain the transient solution of the system. These are further explained in subsections 10.3.1 and 10.3.2.

10.3.1 Transient State Probabilities

Let ‘n’ and ‘m’ represent the number of priority and non priority customers present in the system at any time t, respectively. The transient state probabilities of the system states are denoted as:

$P_{0,0}(t)$ = Probability of the server being in idle or inactive state with no customers in the system at time t.

$P_{n,m}(t)$ = Probability that the server is in busy state with n (0≤ n ≤ L) priority and m (0≤ m ≤ K) non- priority customers in the system at time t (except n=m=0).

$Q_n^{(1)}(t)$ = Probability that there are n (1≤ n ≤ L-1) priority customers in the retrial orbit 1 at time t. In this state, there is no service however, the server is ready to work and the corresponding orbit is not empty.

$Q_m^{(2)}(t)$ = Probability that there are m (1≤ m≤ K-1) non priority customers in the retrial orbit 2 at time t, there is no service and the server is ready to work and the corresponding orbit is not empty.

$R_{n,m}(t)$ = Probability that the server is under repair state with n (n≥0) priority and m (m≥0) non- priority customers in the system at time t (except n=m=0).

The steady state probabilities are given as:

$$P_{n,m} = \lim_{t\to\infty} P_{n,m}(t);\ Q_n^{(1)}(t) = \lim_{t\to\infty} Q_n^{(1)}(t)\ ;\ Q_m^{(2)}(t) = \lim_{t\to\infty} Q_m^{(2)}(t);\ R_{n,m} = \lim_{t\to\infty} R_{n,m}(t)$$

Also, we denote $P'_{n,m}(t) = \dfrac{dP_{n,m}(t)}{dt}$.

10.3.2 Governing Equations

Chapman-Kolmogorov equations for the different states of the model are constructed as written below. The following indicator functions are used for the formulation of the equations governing the model:

$$I_A = \begin{cases} 1,\ if\ (0 \le n \le q_1 - 1), (0 \le m \le K-1) \\ 0,\ if\ (q_1 \le n \le L-1),\ (0 \le m \le K-1) \end{cases}$$

$$I_B = \begin{cases} 1,\ if\ (0 \le m \le q_2 - 1), (0 \le n \le L-1) \\ 0,\ if\ (q_2 \le m \le K-1), (0 \le n \le L-1) \end{cases}$$

(i) Inactive state

This state corresponds to the idle state of the server when neither priority nor non-priority customer is present in the system. The equation in this case is:

$$P'_{0,0}(t) = -(\lambda_1 + \lambda_2)P_{0,0}(t) + \mu_1 P_{1,0}(t) + \mu_2 P_{0,1}(t) \qquad (10.1)$$

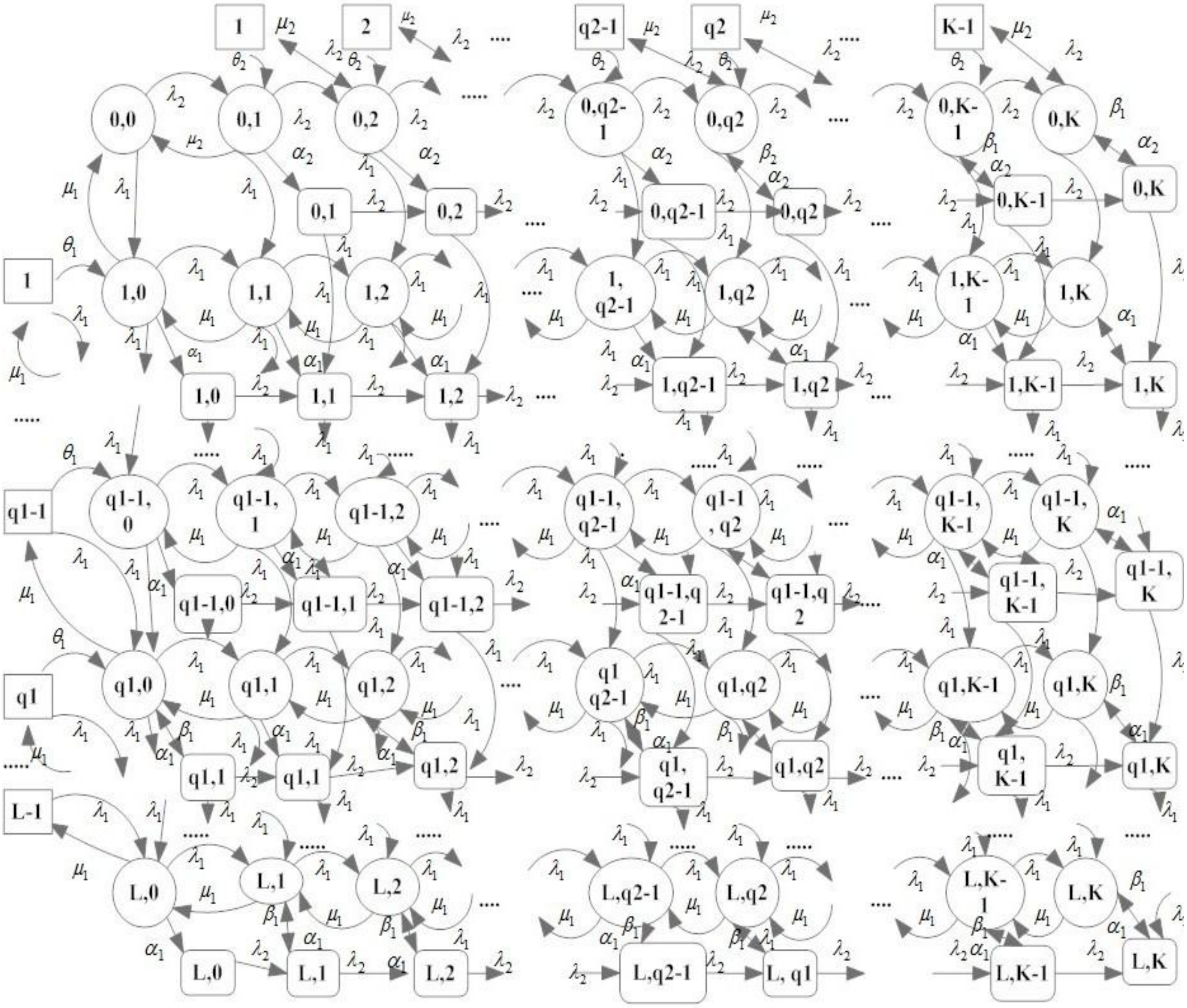


Fig. 10.1: State Transition Diagram

(ii) Busy state

The present state corresponds to that state of the server when it is busy in rendering service to the customers. Depending on the number of customers present in the system, various cases can be framed as:

(a) When there is no non-priority customer in the system.

$$P'_{1,0}(t) = -(\lambda_1 + \lambda_2 + \alpha_1 + \mu_1)P_{1,0}(t) + \lambda_1 P_{0,0}(t) + \gamma_1 Q_1^{(1)}(t) + (1 - I_A)\beta_1 R_{1,0}(t) \qquad (10.2)$$

$$P'_{n,0}(t) = -(\lambda_1 + \lambda_2 + \alpha_1 + \mu_1)P_{n,0}(t) + \lambda_1 P_{n-1,0}(t) + \gamma_1 Q_n^{(1)}(t) + \lambda_1 Q_{n-1}^{(1)}(t) + (1 - I_A)\beta_1 R_{n,0}(t), \quad (2 \le n \le L-1) \qquad (10.3)$$

$$P'_{L,0}(t) = -(\lambda_2 + \alpha_1 + \mu_1)P_{L,0}(t) + \lambda_1 P_{L-1,0}(t) + \lambda_1 Q_{L-1}^{(1)}(t) + \beta_1 R_{L,0}(t) \qquad (10.4)$$

(b) When both priority and non-priority customers are present in the system.

$$P'_{n,m}(t) = -(\lambda_1 + \lambda_2 + \alpha_1 + \mu_1)P_{n,m}(t) + \lambda_1 P_{n-1,m}(t) + \lambda_2 P_{n,m-1}(t)$$
$$+ \mu_1 P_{n+1,m}(t) + (1 - I_A)\beta_1 R_{n,m}(t),\ (1 \le n \le L-1),\ (1 \le m \le K-1) \quad (10.5)$$

$$P'_{L,m}(t) = -(\lambda_2 + \alpha_1 + \mu_1)P_{L,m}(t) + \lambda_1 P_{L-1,m}(t) + \lambda_2 P_{L,m-1}(t) + \beta_1 R_{L,m}(t),$$
$$(1 \le m \le K-1) \quad (10.6)$$

$$P'_{L,K}(t) = -(\alpha_1 + \mu_1)P_{L,K}(t) + \lambda_1 P_{L-1,K}(t) + \lambda_2 P_{L,K-1}(t) + \beta_1 R_{L,K}(t) \quad (10.7)$$

$$P'_{n,K}(t) = -(\lambda_1 + \alpha_1 + \mu_1)P_{n,K}(t) + \lambda_1 P_{n-1,K}(t) + \lambda_2 P_{n,K-1}(t)$$
$$+ (1 - I_A)\beta_1 R_{L,K}(t) + \mu_1 P_{n+1,K}(t),\ (1 \le n \le L-1) \quad (10.8)$$

(c) When there is no priority customer in the system.

$$P'_{0,1}(t) = -(\lambda_1 + \lambda_2 + \alpha_2 + \mu_2)P_{0,1}(t) + \lambda_2 P_{0,0}(t) + \gamma_2 Q_1^{(2)}(t) + \mu_1 P_{1,1}(t)$$
$$+ (1 - I_B)\beta_2 R_{0,1}(t) \quad (10.9)$$

$$P'_{0,m}(t) = -(\lambda_1 + \lambda_2 + \alpha_2 + \mu_2)P_{0,m}(t) + \lambda_2 P_{0,m-1}(t) + \lambda_2 Q_{m-1}^{(2)}(t) + \gamma_2 Q_m^{(2)}(t)$$
$$+ \mu_1 P_{1,m}(t) + (1 - I_B)\beta_2 R_{0,m}(t),\ (2 \le m \le K-1) \quad (10.10)$$

$$P'_{0,K}(t) = -(\lambda_1 + \alpha_2 + \mu_2)P_{0,K}(t) + \lambda_2 P_{0,K-1}(t) + \lambda_2 Q_{K-1}^{(2)}(t)$$
$$+ \mu_1 P_{1,K}(t) + \beta_2 R_{0,K}(t) \quad (10.11)$$

(iii) Repair State

The server is unreliable and may break down while serving either the priority or non-priority customers. Therefore, the governing equations of the states corresponding to the repair process of the broken down server in various situations are given below:

(a) When there is no non priority customers i.e. m=0 and priority customers n ≥ 1.

$$R'_{1,0}(t) = -(\lambda_1 + \lambda_2)R_{1,0}(t) + \alpha_1 P_{1,0}(t) \quad (10.12)$$

$$R'_{n,0}(t) = -(\lambda_1 + \lambda_2 + (1 - I_A)\beta_1)R_{n,0}(t) + \alpha_1 P_{n,0}(t) + \lambda_1 R_{n-1,0}(t), (2 \le n \le L-1) \quad (10.13)$$

$$R'_{L,0}(t) = -(\lambda_2 + \beta_1)R_{L,0}(t) + \alpha_1 P_{L,0}(t) + \lambda_1 R_{L-1,0}(t) \quad (10.14)$$

(b) When the number of non priority customer m≥ 1 and priority customers n≥ 1.

$$R'_{n,m}(t) = -(\lambda_1 + \lambda_2 + (1 - I_A)\beta_1)R_{n,m}(t) + \alpha_1 P_{n,m}(t) + \lambda_1 R_{n-1,m}(t) + \lambda_2 R_{n,m-1}(t),$$
$$(1 \le n \le L-1), (1 \le m \le K-1) \quad (10.15)$$

$$R'_{n,K}(t) = -(\lambda_1 + (1 - I_A)\beta_1)R_{n,K}(t) + \alpha_1 P_{n,K}(t) + \lambda_1 R_{n-1,K}(t) + \lambda_2 R_{n,K-1}(t),$$
$$(1 \le n \le L-1) \quad (10.16)$$

$$R'_{L,m}(t) = -(\lambda_2 + (1 - I_A)\beta_1)R_{L,m}(t) + \alpha_1 P_{L,m}(t) + \lambda_1 R_{L-1,m}(t) + \lambda_2 R_{L,m-1}(t),$$
$$(1 \le m \le K-1) \quad (10.17)$$

$$R'_{L,K}(t) = -\beta_1 R_{L,K}(t) + \alpha_1 P_{L,K}(t) + \lambda_2 R_{L,K-1}(t) + \lambda_1 R_{L-1,K}(t) \quad (10.18)$$

(c) When the number of priority customers n=0 and non priority customers m≥ 1.

$$R'_{0,1}(t) = -(\lambda_1 + \lambda_2)R_{0,1}(t) + \alpha_2 P_{0,1}(t) \tag{10.19}$$

$$R'_{0,m}(t) = -(\lambda_1 + \lambda_2 + (1 - I_A)\beta_2)R_{0,m}(t) + \alpha_2 P_{0,m}(t) + \lambda_2 R_{0,m-1}(t), (2 \le m \le K-1) \tag{10.20}$$

(d) When the number of non priority customers m= K and priority customers n= 0.

$$R'_{0,K}(t) = -(\lambda_1 + \beta_2)R_{0,K}(t) + \alpha_2 P_{0,K}(t) + \lambda_2 R_{0,K-1}(t) \tag{10.21}$$

(iv) For Retrial Orbit 1

$$Q_n^{(1)\prime}(t) = -(\gamma_1 + \lambda_1)Q_n^{(1)}(t) + \mu_1 P_{n+1,0}(t) \qquad \text{for } n = 1, 2, 3, ..., L-1 \tag{10.22}$$

(v) For Retrial Orbit 2

$$Q_m^{(2)\prime}(t) = -(\gamma_2 + \lambda_2)Q_m^{(2)}(t) + \mu_2 P_{0,m+1}(t) \qquad \text{for } m = 1, 2, 3, ..., K-1 \tag{10.23}$$

Normalization Condition is

$$\sum_{n=1}^{L-1} Q_n^{(1)}(t) + \sum_{m=1}^{K-1} Q_m^{(2)}(t) + \sum_{n=0}^{L}\sum_{m=0}^{K} P_{n,m}(t) + \sum_{n=0}^{L}\sum_{m=0}^{K} R_{n,m}(t) = 1 \tag{10.24}$$

10.4 MATRIX METHOD

The present section deals with the transient solution of the system of equations by using the 'Matrix method'. At the initial stage, Laplace transforms of the equations are taken to convert them in differential free form. The obtained new set of equations is then arranged in the form of block matrix so as to obtain transient state probabilities in terms of eigenvalues of the determinant of coefficient matrix. The block matrix developed can be solved by a number of numerical techniques to obtain the eigen values of the coefficient matrix. The transient state probabilities are obtained in terms of the eigenvalues which can be further used to find the performance indices. Before proceeding further, we define transient state probabilities as:

$$\Pi(t) = [\Pi_1(t), \Pi_2(t), ..., \Pi_K(t), \Pi_{K+1}(t), \Pi_{K+2}(t), \Pi_{K+3}(t)]^T;$$

where,

$$\mathbf{\Pi_1}(t) = \mathbf{[}P_{0,0}(t), Q_1^{(2)}(t), ..., Q_{K\text{-}1}^{(2)}(t), P_{0,1}(t), ..., P_{0,K}(t)\mathbf{]}^T{}_{(2K)\times(1)};$$

$$\mathbf{\Pi_2}(t) = \mathbf{[}R_{0,1}(t), ..., R_{0,K}(t), Q_1^{(1)}(t), ..., Q_{L\text{-}1}^{(1)}(t)\mathbf{]}^T{}_{(K+L\text{-}1)\times(1)};$$

$$\Pi_{l+\mathbf{3}}(t) = [P_{1,l}(t), ..., P_{L,l}(t), R_{1,l}(t), ..., R_{1,l}(t)]^T{}_{(2L)\times(1)} \qquad (0 \le l \le K)$$

For the sake of convenience, we define transient state probabilities in terms of $\pi(t)$ as-

$$\mathbf{\Pi_1}(t) = \mathbf{[}\pi_1(t), \pi_2(t), ..., \pi_{2K}(t)\mathbf{]}^T$$

$$\mathbf{\Pi_2}(t) = [\pi_{2K+1}(t), \pi_{2K+2}(t), \ldots, \pi_{3K+L-1}(t)]^T$$

$$\Pi_{l+\mathbf{3}}(t) = [\pi_{3K+L+(l\times 2L)}(t), \ldots, \pi_{3K+L-1+(l+1)\times 2L)}(t)]^T \qquad (0 \le l \le K) \tag{10.25}$$

The Laplace transforms of $\pi_i(t), \Pi_i(t), \Pi(t)$ are denoted by $\tilde{\pi}_i(s), \tilde{\Pi}_i(s), \tilde{\Pi}(s)$ respectively.

In terms of Laplace transform, the above transient state probabilities reduces to

$$\tilde{\Pi}(s) = [\tilde{\Pi}_1(s), \tilde{\Pi}_2(s), \tilde{\Pi}_3(s), \ldots, \tilde{\Pi}_{K+1}(s), \tilde{\Pi}_{K+2}(s), \tilde{\Pi}_{K+3}(s)]^T$$

$$\tilde{\mathbf{\Pi}}_{\mathbf{1}}(t) = [\tilde{\pi}_1(t), \tilde{\pi}_2(t), \ldots, \tilde{\pi}_{2K}(t)]^T$$

$$\tilde{\mathbf{\Pi}}_{\mathbf{2}}(t) = [\tilde{\pi}_{2K+1}(t), \tilde{\pi}_{2K+2}(t), \ldots, \tilde{\pi}_{3K+L-1}(t)]^T \tag{10.26}$$

$$\tilde{\Pi}_{l+\mathbf{3}}(t) = [\tilde{\pi}_{3K+L+(l\times 2L)}(t), \ldots, \tilde{\pi}_{3K+L-1+(l+1)\times 2L)}(t)]^T \qquad (0 \le l \le K)$$

Noting that initially at time t=0, the system is empty i.e.

$$P_{0,0}(0) = 1, P_{n,m}(0) = 0 \ \forall\, n \ne 0, m \ne 0. \tag{10.27}$$

The initial vector can be defined as,

$$\Pi(0) = [1, 0, 0, \ldots, 0]_{(2KL+3K+3L-1)\times 1} \tag{10.28}$$

After taking Laplace transforms, the set of differential equations (10.1)-(10.23) can be written in matrix form as:

$$\mathbf{A(s)}\tilde{\Pi}\mathbf{(s)} = \Pi\mathbf{(0)} \tag{10.29}$$

where,

$$\mathbf{A(s)} = \begin{bmatrix} \mathbf{A_0} & \mathbf{B_0} & \mathbf{B_1} & \mathbf{C_1} & \mathbf{C_2} & \mathbf{C_3} & \cdots & \mathbf{C_{K-1}} & \mathbf{C_K} \\ \mathbf{D_0} & \mathbf{A_1} & \mathbf{G_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{G} & \mathbf{D_1} & \mathbf{A_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{E_1} & \mathbf{F_1} & \mathbf{D_2} & \mathbf{A_3} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{E_2} & \mathbf{F_2} & \mathbf{0} & \mathbf{D_2} & \mathbf{A_3} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{E_3} & \mathbf{F_3} & \mathbf{0} & \mathbf{0} & \mathbf{D_2} & \mathbf{A_3} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & & \mathbf{0} & \mathbf{0} \\ \mathbf{E_{K-1}} & \mathbf{F_{K-1}} & \vdots & \vdots & \cdots & & \mathbf{D_2} & \mathbf{A_3} & \mathbf{0} \\ \mathbf{E_K} & \mathbf{F_K} & \mathbf{0} & \mathbf{0} & \cdots & & \mathbf{0} & \mathbf{D_2} & \mathbf{A_4} \end{bmatrix}_{(2LK+3K\text{-}1)\times(2LK+3K\text{-}1)}$$

$$\mathbf{A_0} = \begin{bmatrix} \mathbf{C} & \mathbf{B} \\ \mathbf{D} & \mathbf{E} \end{bmatrix}_{(2K)}, \quad \mathbf{A_1} = \begin{bmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_1 \end{bmatrix}_{(K+L\text{-}1)}, \mathbf{A_2} = \begin{bmatrix} \mathbf{M_0} & \mathbf{M} \\ \mathbf{M}_1 & \mathbf{U} \end{bmatrix}_{(2L)}, \mathbf{A_3} = \begin{bmatrix} \mathbf{S} & \mathbf{M} \\ \mathbf{M}_1 & \mathbf{U} \end{bmatrix}_{(2L)}, \mathbf{A_4} = \begin{bmatrix} \mathbf{V} & \mathbf{M} \\ \mathbf{M}_1 & \mathbf{W} \end{bmatrix}_{(2L)}$$

$$\mathbf{B_0} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} \end{bmatrix}_{(K+L\text{-}1)\times(2K)}, \mathbf{B_1} = \begin{bmatrix} \mathbf{J_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{(2K)\times(2L)}, \mathbf{G}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{Z} & \mathbf{0} \end{bmatrix}_{(K+L\text{-}1)\times(2K)}, \mathbf{G} = \begin{bmatrix} \mathbf{T_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{(2L)\times(2K)}$$

$$\mathbf{D_0}=\begin{bmatrix}\mathbf{0} & \mathbf{M}_2\\ \mathbf{0} & \mathbf{0}\end{bmatrix}_{(K+L\text{-}1)\times(2K)},\mathbf{D_1}=\begin{bmatrix}\mathbf{0} & \mathbf{K_1}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}_{(2L)\times(K+L\text{-}1)}$$

$$\mathbf{C}_l=\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{J}_l & \mathbf{0}\end{bmatrix}_{(2K)\times(2L)},\mathbf{E}_l=\begin{bmatrix}\mathbf{0} & \mathbf{T}_l\\ \mathbf{0} & \mathbf{0}\end{bmatrix}_{(2L)\times(2K)},\mathbf{F}_l=\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{T}_l & \mathbf{0}\end{bmatrix}_{(2L)\times(K+L-1)} \quad l=1,2,...K$$

The various sub matrices can be further defined as:

$\mathbf{B}=\mu_2\mathbf{I}_{(K)}$, $\mathbf{R_0}=\alpha_2\mathbf{I}_{(K)}$, $\mathbf{I}_1=[-(s+\theta_1+\lambda_1)]\mathbf{I}_{(L-1)}$, $\mathbf{M}_1=\alpha_1\mathbf{I}_{(L)}$, $\mathbf{D}_2=(\lambda_2)\mathbf{I}_{(2L)}$

where, $\mathbf{I}_{(m)}$ is the identity matrix of order m.

$\mathbf{T}_l=[t_{ij}]_{(L\times K)}$

$$t_{ij}=\begin{cases}\lambda_1; & for\, i=1,\, j=l\\ 0; & otherwise\end{cases}$$

$\mathbf{F}=[f_{ij}]_{(K\times K)}$

$$f_{ij}=\begin{cases}\beta_2; & for\, i=j\, and\ q_2\le j\le K\\ 0; & otherwise\end{cases}$$

$\mathbf{K_1}=[k_{ij}]_{(L\times(L-1))}$

$$k_{ij}=\begin{cases}\gamma_1; & for\, i=j\\ \lambda_1; & for\ i=j+1\ and\, 1\le j\le L-1\end{cases}$$

$\mathbf{J}_l=[j_{ij}]_{(K\times L)}$

$$j_{ij}=\begin{cases}\mu_1; & for\, i=l\ and\ j=1\\ 0; & otherwise\end{cases}$$

$\mathbf{Z}=[z_{ij}]_{(L-1)\times L}$

$$z_{ij}=\begin{cases}\mu_1; & for\ 1\le i\le L-1\ and\ j=i+1\\ 0; & otherwise\end{cases}$$

$\mathbf{M}=[m_{ij}]_{(L\times L)}$

$$m_{ij}=\begin{cases}\beta_1; & for\, i=j\, and\ q_1\le j\le L\\ 0; & otherwise\end{cases}$$

$\mathbf{C}=[a_{ij}]_{(K\times K)}$

$$a_{ij}=\begin{cases}-(s+\lambda_1+\lambda_2); & for\, i=j=1\\ -(s+\gamma_2+\lambda_2); & for\, i=j\ and\ 2\le i\le K\end{cases}$$

$\mathbf{D}=[d_{ij}]_{(K\times K)}$

$$d_{ij}=\begin{cases}\lambda_2; & for\, i=j, 1\le i\le K\\ \gamma_2; & for\ \ j=i+1\ and\ 1\le i\le K-1\end{cases}$$

$\mathbf{E}=[e_{ij}]_{(K\times K)}$

$$e_{ij}=\begin{cases}-(s+\lambda_1+\lambda_2+\alpha_2+\mu_2); & for\, i=j\ and\ 1\le i\le K-1\\ -(s+\lambda_2+\alpha_2+\mu_2); & for\, i=j\ and\ i=K\\ \lambda_2; & for\ i=j+1 and\, 1\le j\le K-1\end{cases}$$

$\mathbf{M}_0=[\partial_{ij}]_{(L\times L)}$

$$\partial_{ij}=\begin{cases}-(s+\lambda_1+\lambda_2+\alpha_1+\mu_1); & for\, i=j\ and\ 1\le i\le L-1\\ -(s+\lambda_2+\alpha_1+\mu_1); & for\, i=j\ and\ i=L\\ \lambda_1; & for\ i=j+1 and\, 1\le j\le L-1\end{cases}$$

$\mathbf{U}=[u_{ij}]_{(L\times L)}$

$$u_{ij}=\begin{cases}-(s+\lambda_1+\lambda_2);\ for\, i=j\ and\ 1\le i\le q_1-1\\ -(s+\lambda_1+\lambda_2+\beta_1);\ for\, i=j\ and\ q_1\le i\le L-1\\ -(s+\lambda_2+\beta_1);\ for\, i=j=L\\ \lambda_1;\ for\ i=j+1 and 1\le j\le L-1\end{cases}$$

$\mathbf{S}=[s_{ij}]_{(L\times L)}$

$$s_{ij}=\begin{cases}-(s+\lambda_1+\lambda_2+\alpha_1+\mu_1);\ for\, i=j\ and\ 1\le i\le L-1\\ -(s+\lambda_2+\alpha_1+\mu_1);\ for\, i=j\ and\ i=L\\ \lambda_1;\ for\ i=j+1 and 1\le j\le L-1\\ \mu_1;\ for\ j=i+1 and 1\le i\le L-1\end{cases}$$

$\mathbf{H}=[h_{ij}]_{(K\times K)}$

$$h_{ij}=\begin{cases}-(s+\lambda_1+\lambda_2);\ for\, i=j\ and\ 1\le i\le q_2-1\\ -(s+\lambda_1+\lambda_2+\beta_2);\ for\, i=j\ and\ q_2\le i\le K-1\\ -(s+\lambda_1+\lambda_2+\beta_2); for\, i=j=K\\ \lambda_2;\ for\ i=j+1 and 1\le j\le K-1\end{cases}$$

$\mathbf{W}=[w_{ij}]_{(L\times L)}$

$$w_{ij}=\begin{cases}-(s+\lambda_1);\ for\, i=j\ and\ 1\le i\le q_1-1\\ -(s+\lambda_1+\beta_1);\ for\, i=j\ and\ q_1\le i\le L-1\\ -(s+\beta_1);\ for\, i=j=L\\ \lambda_1;\ for\ i=j+1 and 1\le j\le L-1\end{cases}$$

$\mathbf{V}=[v_{ij}]_{(L\times L)}$

$$v_{ij}=\begin{cases}-(s+\lambda_1+\alpha_1+\mu_1);\ for\, i=j\ and\ 1\le i\le L-1\\ -(s+\alpha_1+\mu_1)\quad ; for\, i=j=L\\ \lambda_1;\ for\ i=j+1 and 1\le j\le L-1\\ \mu_1;\ for\ j=i+1 and 1\le i\le L-1\end{cases}$$

Now, we aim to find out the probability of the system at any time t. At the initial stage, we use Cramer's rule so as to determine the transient probabilities of the server at different states on matrix $\mathbf{A}(s)$. From equation (10.29), we obtain

$$\tilde{\pi}_i(s)=\frac{\det[\mathrm{A}_i(s)]}{\det\left[\mathrm{A}(s)\right]},\quad (i=1,\ 2,3,4,.........,(2LK+3L+3K-1)) \tag{10.30}$$

where, $\det\left[\mathrm{A}(s)\right]$ is the determinant of the matrix $\boldsymbol{A}$(s) and $\det[\mathrm{A}_i(s)]$ is the determinant of matrix which has been obtained by replacing the respective ith column vector, $(i=1,\ 2,3,4,.........,(2LK+3L+3K-1))$ of $\mathbf{A}(s)$ with initial vector $\Pi(0)$. Now, in order to obtain the explicit expression for the equation (10.30), we proceed as follows:

It is clear that s=0 is a root of $\det\left[\mathrm{A}(s)\right]$=0. Now substituting s= -δ, we obtain

$$\mathbf{A}(-\delta)=(\mathbf{B}-\delta\mathbf{I}) \tag{10.31}$$

where, $\mathrm{B}=\mathrm{A}(0)$ is a square matrix of order $(2LK+3L+3K-1)$ and I is the identity matrix of order $(2LK+3L+3K-1)$. Using equations (10.29) and (10.31), we obtain

$$\mathbf{A}(-\delta)\tilde{\Pi}(s)=(\mathbf{B}-\delta\mathbf{I})\tilde{\Pi}(s)=\Pi(0) \tag{10.32}$$

Now, we find other distinct eigenvalues $\delta_x\left(\delta_x\neq 0, x=1,2,......,(2LK+3L+3K-2)\right)$ of the matrix $\mathbf{B}-\delta\mathbf{I}$. For this purpose, we equate its determinant equals to zero. The eigenvalues are either, real (excluding zero) or complex. We assume that there are x real and y pairs of distinct conjugate complex eigenvalues which we denote by $\delta_1,\delta_2,.............,\delta_x$ *and* $\left(\delta_{x+1},\bar{\delta}_{x+1}\right),\left(\delta_{x+2},\bar{\delta}_{x+2}\right),.................,\left(\delta_{x+y},\bar{\delta}_{x+y}\right)$ respectively. Moreover, $x+2y=(2LK+3L+3K-2)$ and thus, we have

$$\det\left[\mathrm{A}(s)\right]=s\left[\prod_{k=1}^{x}\left(s+\delta_k\right)\right]\left[\prod_{k=1}^{y}\left\{s^2+\left(\delta_{x+k}+\bar{\delta}_{x+k}\right)s+\delta_{x+k}\bar{\delta}_{x+k}\right\}\right] \tag{10.33}$$

Using equations (10.30) and (10.33), we get

$$\tilde{\pi}_i(s)=\frac{\det[\mathrm{A}_i(s)]}{s\left[\prod_{k=1}^{x}\left(s+\delta_k\right)\right]\left[\prod_{k=1}^{y}\left\{s^2+\left(\delta_{x+k}+\bar{\delta}_{x+k}\right)s+\delta_{x+k}\bar{\delta}_{x+k}\right\}\right]},$$

$$(i=1,\ 2,3,4,.........,(2LK+3L+3K-1)) \tag{10.34}$$

On expanding by partial fractions, we get

$$\tilde{\pi}_i(s)=\frac{a_0}{s}+\sum_{m=1}^{x}\frac{a_m}{s+\delta_m}+\sum_{m=1}^{y}\frac{b_m s+c_m}{s^2+\left(\delta_{x+m}+\bar{\delta}_{x+m}\right)s+\delta_{x+m}\bar{\delta}_{x+m}}\ ,$$

$$(i=1,\ 2,3,4,.........,(2LK+3L+3K-1)) \tag{10.35}$$

where a_0 and $a_m\left(m=1,2,............,x\right)$ are real numbers and are obtained as

$$a_0=\frac{\det[\mathrm{A}_i(0)]}{\left(\prod_{k=1}^{x}\delta_k\right)\left(\prod_{k=1}^{y}\delta_{x+k}\bar{\delta}_{x+k}\right)} \tag{10.36}$$

$$a_m=\frac{\det[\mathrm{A}_i(-\delta_m)]}{(-\delta_m)\left[\prod_{\substack{k=1\\k\neq m}}^{x}\left(\delta_k-\delta_m\right)\right]\left[\prod_{k=1}^{y}\left\{(-\delta_m)^2+\left(\delta_{x+k}+\bar{\delta}_{x+k}\right)(-\delta_m)+\delta_{x+k}\bar{\delta}_{x+k}\right\}\right]},\mathrm{m}=1,\ 2,\ldots,\ \mathrm{x} \tag{10.37}$$

$$b_m(-\delta_{x+m})+c_m=\frac{\det\left[A_i(-\delta_{x+m})\right]}{(-\delta_{x+m})\left[\prod_{k=1}^{x}(\delta_k-\delta_{x+m})\right]\left[\prod_{\substack{k=1\\k\neq m}}^{y}\left\{(-\delta_{x+m})^2+\left(\delta_{x+k}+\bar{\delta}_{x+k}\right)(-\delta_{x+m})+\delta_{x+k}\bar{\delta}_{x+k}\right\}\right]}$$

m=1,2,…,y (10.38)

On taking inverse Laplace transformation of equation (10.35), the probability of the system state at any time t, are given by

$$\pi_i(t)=a_0+\sum_{m=1}^{x}a_m e^{-\delta_m t}+\sum_{m=1}^{y}\left[b_m e^{-v_m t}\cos(w_m t)+\frac{c_m-b_m v_m}{w_m}e^{-v_m t}\sin(w_m t)\right],$$

$$(1\le i\le(2LK+3L+3K-1))$$

(10.39)

where, $a_0, a_m, b_m, c_m, v_m, w_m$ are real numbers.

10.5 MATRIX RECURSIVE APPROACH

In the previous section, we have used matrix method to obtain transient state probabilities of the system in terms of eigenvalues of the determinants. Now, to obtain the steady state probabilities of the system we use 'Matrix Recursive Approach'. This method is widely used to deal with various queueing models for exact steady state solutions.

The steady state probabilities are denoted by:

$$P_{n,m}=\lim_{t\to\infty}P_{n,m}(t);\ Q_n^{(1)}(t)=\lim_{t\to\infty}Q_n^{(1)}(t)\ ;\ Q_m^{(2)}(t)=\lim_{t\to\infty}Q_m^{(2)}(t);$$

$$R_{n,m}=\lim_{t\to\infty}R_{n,m}(t);\prod_n=\lim_{t\to\infty}\prod_n(t)$$

Eq. (10.29) can be written as:

$$\begin{bmatrix}
\mathbf{A_0} & \mathbf{B_0} & \mathbf{B_1} & \mathbf{C_1} & \mathbf{C_2} & \mathbf{C_3} & \cdots & \mathbf{C_{K-1}} & \mathbf{C_K}\\
\mathbf{D_0} & \mathbf{A_1} & \mathbf{G_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\
\mathbf{G} & \mathbf{D_1} & \mathbf{A_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\
\mathbf{E_1} & \mathbf{F_1} & \mathbf{D_2} & \mathbf{A_3} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\
\mathbf{E_2} & \mathbf{F_2} & \mathbf{0} & \mathbf{D_2} & \mathbf{A_3} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\
\mathbf{E_3} & \mathbf{F_3} & \mathbf{0} & \mathbf{0} & \mathbf{D_2} & \mathbf{A_3} & \cdots & \mathbf{0} & \mathbf{0}\\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & & \mathbf{0} & \mathbf{0}\\
\mathbf{E_{K-1}} & \mathbf{F_{K-1}} & \vdots & \vdots & \cdots & & \mathbf{D_2} & \mathbf{A_3} & \mathbf{0}\\
\mathbf{E_K} & \mathbf{F_K} & \mathbf{0} & \mathbf{0} & \cdots & & \mathbf{0} & \mathbf{D_2} & \mathbf{A_4}
\end{bmatrix}
\begin{bmatrix}\mathbf{\Pi_1}\\ \mathbf{\Pi_2}\\ \mathbf{\Pi_3}\\ \mathbf{\Pi_4}\\ \mathbf{\Pi_5}\\ \vdots\\ \mathbf{\Pi_{K+1}}\\ \mathbf{\Pi_{K+2}}\\ \mathbf{\Pi_{K+3}}\end{bmatrix}
=
\begin{bmatrix}\mathbf{\Pi_1}(0)\\ \mathbf{\Pi_2}(0)\\ \mathbf{\Pi_3}(0)\\ \mathbf{\Pi_4}(0)\\ \mathbf{\Pi_5}(0)\\ \vdots\\ \mathbf{\Pi_{K+1}}(0)\\ \mathbf{\Pi_{K+2}}(0)\\ \mathbf{\Pi_{K+3}}(0)\end{bmatrix}$$

Therefore we have the following set of equations:

$$\mathbf{A_0}\Pi_1+\mathbf{B_0}\Pi_2+\mathbf{B_1}\Pi_3+\sum_{i=1}^{K}\mathbf{C}_i\Pi_{i+3}=1, \qquad (10.40)$$

$$\mathbf{D_0}\Pi_1+\mathbf{A_1}\Pi_2+\mathbf{G_1}\Pi_3=0 \qquad (10.41)$$

$$\mathbf{G}\Pi_1+\mathbf{D_1}\Pi_2+\mathbf{A_2}\Pi_3=0 \qquad (10.42)$$

$$\mathbf{E}_i\Pi_1+\mathbf{F}_i\Pi_2+\mathbf{D_2}\Pi_{i+2}+\mathbf{A_3}\Pi_{i+3}=0,\ (1\le i\le K-1) \qquad (10.43)$$

$$\mathbf{E_K}\Pi_1+\mathbf{F_K}\Pi_2+\mathbf{D_2}\Pi_{\mathbf{K+2}}+\mathbf{A_4}\Pi_{\mathbf{K+3}}=0 \qquad (10.44)$$

On solving (10.41) and (10.42), we get

$$\Pi_3=-\mathbf{G_1^{-1}}\{\mathbf{D_0}\Pi_1+\mathbf{A_1}\Pi_2\}=-\mathbf{A_2^{-1}}\{\mathbf{G}_1\Pi_1+\mathbf{D_1}\Pi_2\} \qquad (10.45a)$$

$$\text{and, }\Pi_2=\left\{\mathbf{G_1^{-1}A_1}-\mathbf{A_2^{-1}D_1}\right\}^{-1}\left\{\mathbf{A_2^{-1}G}-\mathbf{G_1^{-1}D_0}\right\}\Pi_1 \qquad (10.45b)$$

Using the value of Π_2 in $\Pi_3=-\mathbf{G_1^{-1}}\{\mathbf{D_0}\Pi_1+\mathbf{A_1}\Pi_2\}$, we get

$$\Pi_3=-\mathbf{G_1^{-1}}\left\{\mathbf{D_0}+\left\{\mathbf{G_1^{-1}A_1}-\mathbf{A_2^{-1}D_1}\right\}^{-1}\left\{\mathbf{A_2^{-1}G}-\mathbf{G_1^{-1}D_0}\right\}\right\}\Pi_1 \qquad (10.46)$$

Proceeding in a similar pattern, we can obtain a general result for equations (10.43) in the following form:

$$\Pi_i=-\mathbf{A}_3^{-1}\left\{\sum_{r=1}^{i-3}\left(\mathbf{E}_{i-2-r}\left(-\mathbf{D}_2\mathbf{A}_3^{-1}\right)^{r-1}\right)\Pi_1+\sum_{r=1}^{i-3}\left(\mathbf{F}_{i-2-r}\left(-\mathbf{D}_2\mathbf{A}_3^{-1}\right)^{r-1}\right)\Pi_2+\left(\mathbf{D}_2\right)^{i-3}\left(-\mathbf{A}_3^{-1}\right)^{i-4}\Pi_3\right\} \qquad (10.47)$$

$$4\le i\le K+2$$

Equation (10.48) yields, $\Pi_{\mathbf{K+3}}=-\mathbf{A}_4^{-1}\{\mathbf{E_K}\Pi_1+\mathbf{F_K}\Pi_2+\mathbf{D_2}\Pi_{\mathbf{K+2}}\}$ (10.48)

By using the values of unknowns from Π_2 to Π_{K+2} in equation (10.40), i.e. we can get Π_1.

10.5 PERFORMANCE MEASURES

The validity of any retrial queueing model can be best deciphered in terms of its performance indices. Various indices namely average queue length, system state probabilities, throughput etc. can be determined so as to judge the efficiency of the system.

10.5.1 Server State Probabilities

The probabilities of the different states of the server are important in deciding the efficiency and other metrics of the system. The various server state probabilities for the different states of the server at time t are established as follows:

(a) Busy state:

The probability that the server is busy in providing service to the customers at time t, is

$$P_B(t) = \sum_{n=1}^{L}\sum_{m=1}^{K} P_{n,m}(t) + \sum_{n=1}^{L} P_{n,0}(t) + \sum_{m=1}^{K} P_{0,m}(t) \tag{10.49}$$

(b) Broken down state:

The probability that the server being in broken down state at time t, is

$$P_f(t) = \sum_{n=1}^{L}\sum_{m=1}^{K} R_{n,m}(t) \tag{10.50}$$

(c) Repair State:

(i) The probability that the server is under repair of the server when failed while servicing the non- priority customers at time t, is

$$P_{R1}(t) = \sum_{m=q_2}^{K} R_{0,m}(t) \tag{10.51}$$

(ii) The probability that the server is under repair when failed while servicing the priority customers at time t, is

$$P_{R2}(t) = \sum_{n=q_1}^{L}\sum_{m=0}^{K} R_{n,m}(t) \tag{10.52}$$

(iii) The probability that the server is under repair at time t, is

$$P_R(t) = P_{R1}(t) + P_{R2}(t) \tag{10.53}$$

(iv) The probability that the server is in broken down state but repair is not started at time t

$$P_{R3}(t) = \sum_{n=1}^{q_1-1}\sum_{m=0}^{K} R_{n,m}(t) + \sum_{m=1}^{q_2-1} R_{0,m}(t) \tag{10.54}$$

(d) ***Retrial State:***

(i) The probability that the priority customer retry for the service at time t, is

$$P_{RE1}(t) = \sum_{n=1}^{L-1} Q_n^{(1)}(t) \tag{10.55}$$

(ii) The probability that the non-priority customer retry for the service at time t, is

$$P_{RE2}(t) = \sum_{m=1}^{K-1} Q_m^{(2)}(t) \tag{10.56}$$

10.5.2 Queueing Indices

The computation of queueing indices of any system is the most significant and promising measure to upgrade any system. It really helps the system designers for better management of delay situations and efficient functioning of the queueing systems.

(a) Queue Length

The assessment of queue length is the primary objective of any queueing model. Here we give some indices which are related to the queue length of the concerned model.

(i) Expected number of priority customers in the system at time t, is

$$E\left\{N_1(t)\right\} = \sum_{n=1}^{L-1} nQ_n^{(1)}(t) + \sum_{n=1}^{L}\sum_{m=0}^{K} n\left(P_{n,m}(t) + R_{n,m}(t)\right) \tag{10.57}$$

(ii) Expected number of non-priority customers in the system at time t, is

$$E\left\{N_2(t)\right\} = \sum_{m=1}^{K-1} mQ_m^{(2)}(t) + \sum_{n=0}^{L}\sum_{m=1}^{K} m\left(P_{n,m}(t) + R_{n,m}(t)\right) \tag{10.58}$$

(iii) Expected number of customers in the system at time t, is

$$E\left\{N(t)\right\} = E\left\{N_1(t)\right\} + E\left\{N_2(t)\right\} \tag{10.59}$$

(iv) Expected number of customers waiting in the retrial orbits at any time t, is

$$E\left\{N_r(t)\right\} = \sum_{n=1}^{L-1} nQ_n^{(1)}(t) + \sum_{m=1}^{K-1} mQ_m^{(2)}(t) \tag{10.60}$$

(v) Expected number of customers in the breakdown state at any time t, is

$$E\left\{N_f(t)\right\} = \sum_{n=1}^{L}\sum_{m=0}^{K} nR_{n,m}(t) + \sum_{n=0}^{L}\sum_{m=1}^{K} mR_{n,m}(t) \tag{10.61}$$

(b) Throughput

Throughput can be considered as the average rate of successful services rendered to the customers by the server in a queueing system and can be expressed as:

$$TP(t) = \mu_1 \sum_{n=1}^{L}\sum_{m=0}^{K} P_{n,m}(t) + \mu_2 \sum_{m=1}^{K} P_{0,m}(t) \tag{10.62}$$

(c) Carried load

The carried load at time t is given by:

$$\begin{aligned} C_L(t) = {} & \lambda_1\left[\sum_{n=1}^{L-1} Q_n^{(1)}(t) + \sum_{n=1}^{L}\sum_{m=0}^{K} P_{n,m}(t) + \sum_{n=1}^{L}\sum_{m=1}^{K} R_{n,m}(t)\right] \\ & + \lambda_2\left[\sum_{m=1}^{K-1} Q_m^{(2)}(t) + \sum_{n=0}^{L}\sum_{m=1}^{K} P_{n,m}(t) + \sum_{n=1}^{L}\sum_{m=1}^{K} R_{n,m}(t)\right] \end{aligned} \tag{10.63}$$

10.5.3 Reliability measures

The reliability indices also play a major role in improving the availability and efficiency of the concerned unreliable server queueing system. These indices can be further used to improve the system during design and development phases at time t.

(i) Availability of the server at any time t.

It measures the probability of the server being available in the system at time t and is given by

$$A_V(t)=\sum_{n=1}^{L-1}Q_n^{(1)}(t)+\sum_{m=1}^{K-1}Q_m^{(2)}(t)+\sum_{n=0}^{L}\sum_{m=0}^{K}P_{n,m}(t) \tag{10.64}$$

(ii) Failure Frequency

The rate of failure of the server at different states is used to obtain the failure frequency of the system. Thus, the failure frequency at any time t is expressed as:

$$F_f(t)=\alpha_1\sum_{n=1}^{L}\sum_{m=0}^{K}P_{n,m}(t)+\alpha_2\sum_{m=1}^{K}P_{0,m}(t) \tag{10.65}$$

10.5 APPLICATION TO CELLULAR RADIO NETWORK

We illustrate the real life application of queueing model under consideration in cellular radio network wherein the whole geographical area is divided into cells and in each cell there is a base station. We consider the radio transmission via base station by considering a single channel in the microcell to serve the incoming calls which are generated in the Poisson fashion.

It is assumed that before arrival of any call in the system, ‘n’ handoff calls and ‘m’ new calls are already present in the system. The arriving calls are of two types i.e. (i) new calls which are assumed to be originated in the coverage area of the cell and (ii) handoff calls which are transferred from the neighboring cell due to mobility of the subscribers. The traffic flow in cellular radio network is depicted in figure 10.2. A cutoff priority is given to the handoff calls. Once the connection is established, the call should be continued till completion by reserving some channels called guard channels for them but not at the cost of new calls. New call requests are allowed to be queued in the buffer whenever no channel is available at the arrival instant; in such a case the handover calls are treated as the blocked calls. The interruption in the transmission due to unavailability of the channel is managed by providing the buffer of capacity L (K) for handover (new) attempts. If any handover call is present in the cell, the new call is not served. If the channel is busy, the new as well as handover calls wait in the orbits from where they retry for the service. The channel is subject to breakdown and repair. The repairing of the channel is permissible only when a pre-specified number of calls are accumulated in the buffer of coverage area. Here, we consider that handover calls arrive with rate λ_1 and new calls arrive with arrival rate λ_2. The handover (new) calls are served with rate $\mu_1(\mu_2)$. Excluding the call being served at the moment by the channel, extra handover and new calls present in the queue retry for the service with retrial rate $\gamma_1(\gamma_2)$. A handoff call can retry for the service only if no new calls are present in the system, i.e. the retrials cannot be made by handoff calls at the loss or cost of new (non-priority) calls. Based upon above discussion, it be realized that the congestion problem in cellular radio network can be considered as the direct implication of our retrial queueing model.

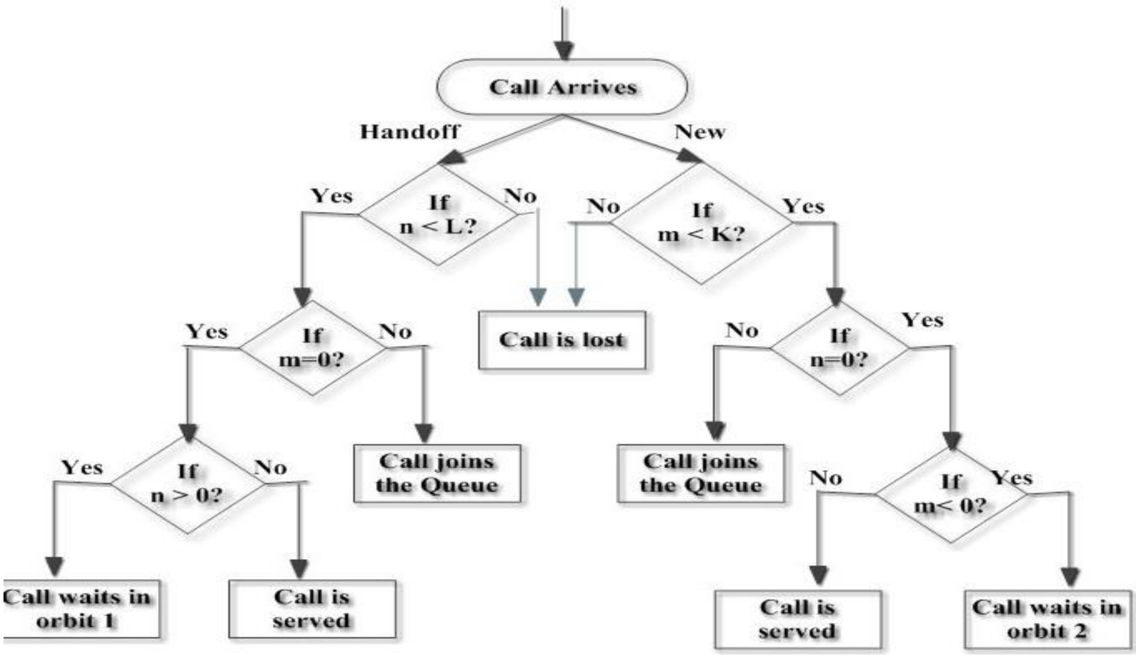


Fig. 10.2: Flow chart of traffic in cellular radio network

10.6 NUMERICAL RESULTS

In this section, we perform numerical experiment by taking the illustration of the cellular mobile network as explained in above section 10.5. To study the effect of various parameters on the sensitivity of the system performance, the values for default parameters are taken as $\lambda_1 = 5,\ \lambda_2 = 6,\ \gamma_1 = 0.5, \gamma_2 = 1,\ \mu_1 = 8,\ \mu_2 = 7,\ \beta_1 = 1, \beta_2 = 2,\ \alpha_1 = 0.5,\ \alpha_2 = 0.5.$

(A) Queue Length

The single channel (server) can serve only one call at a time, therefore rest of the incoming calls either handoff or incoming form queue in the system. The queue length of either type of calls is affected by various parameters namely service rate, arrival rate of calls and many other factors. To examine the effect of system parameters we consider following numerical example.

Illustration 10.1: We consider the set of default parameters and vary q_1 from 1 to 5 at a time range of 1 to 10 units and find the corresponding queue length of priority customers.

Table 10.1: Variation in E [N_1(t)] with q_1 and t

t	q_1=1	q_1=2	q_1=3	q_1=4	q_1=5
1	1.1776	**0.9441**	0.9587	0.9564	0.6970
2	1.2986	**1.2763**	1.3742	1.4676	1.5244
3	1.4899	**1.4380**	1.5810	1.7348	1.8711
4	1.8354	**1.5015**	1.6631	1.8437	2.0183
5	2.2537	**1.5256**	1.6941	1.8842	2.0727
6	2.7402	**1.5353**	1.7061	1.8999	2.0942
7	3.3142	**1.5393**	1.7111	1.9065	2.1038
8	3.9983	**1.5411**	1.7133	1.9095	2.1085
9	4.8169	**1.5420**	1.7143	1.9108	2.1109
10	5.7985	**1.5424**	1.7148	1.9115	2.1121

To minimize the congestion in the system, we intend to compute the optimal threshold recovery parameter q_1 which minimizes the queue length of priority customers (handover calls). Table 10.1 displays the value of queue length of priority customers (handover calls) i.e. E[N_1(t)] corresponding to different threshold parameters q_1 and time t. It is noticed from the table that at every time t, the minimum queue length is obtained for q_1=2. In fig. 10.3, it is observed that the number of calls increases with an increase in time t. However, E[N_1(t)] decreases from q_1=1 to q_1=2 and then increases up to q_1=5. It is very clear from the data given in table 10.1 that q_1=2 seem to be the optimal threshold recovery point where minimum queue length is observed.

Figs 10.4(a)-10.4(b) depict the trends of queue length of handoff calls i.e. priority customers w.r.t. arrival rate (λ_1) and breakdown rate (α_1) on E[N_1(t)]. An increase in the arrival rate (breakdown rate) increases the accumulation of handoff calls in the system (make the system more prone to failures). Figs 10.5(a)-10.5(c) exhibit the variation in the queue length of both handoff calls (priority customers) and new calls (non-priority customers) with time t and other parameters. The variation in the queue length of new incoming calls (non-priority customers) with time t and parameters q_2 and K has been demonstrated in Figs 10.5(b)-10.5(c). The queue length $E[N_2(t)]$ of the incoming calls (non-priority customers) increases with an increase in time t as well as with q_2 and K.

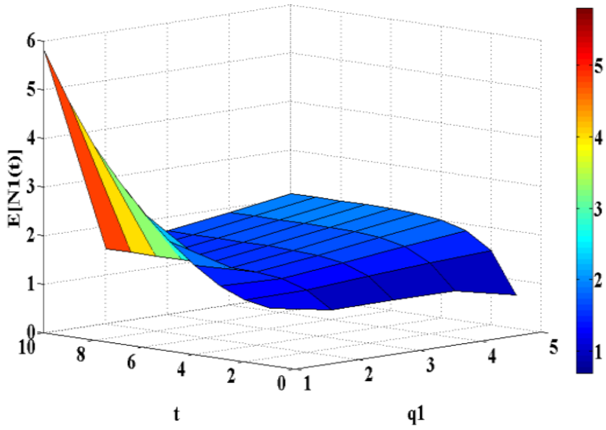


Fig. 10.3: Effect of q_1 on $E[N_1(t)]$ with time t

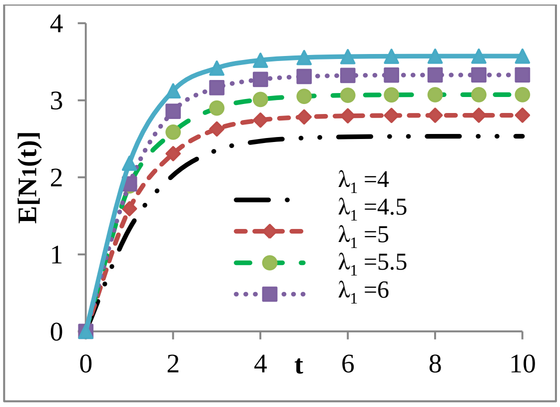


Fig. 10.4(a): Effect of λ_1 on E $[N_1(t)]$

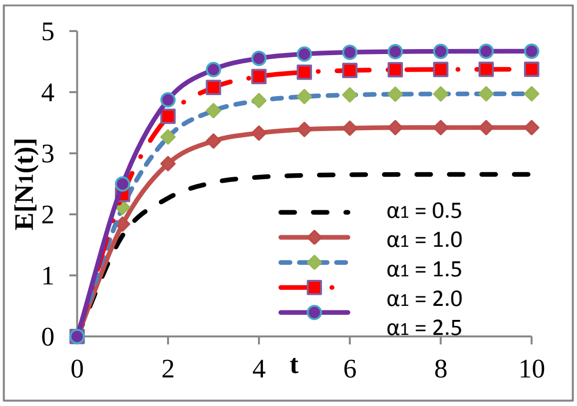


Fig. 10.4(b): Effect of α_1 on $E[N_1(t)]$

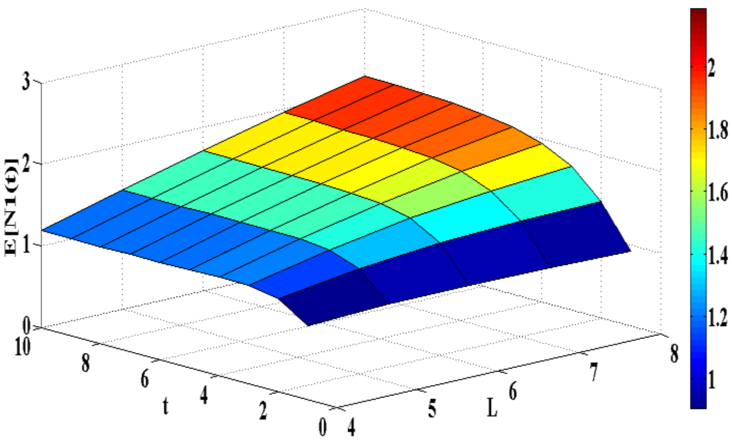


Fig. 10.5(a): Effect of L on $E[N_1(t)]$ with t

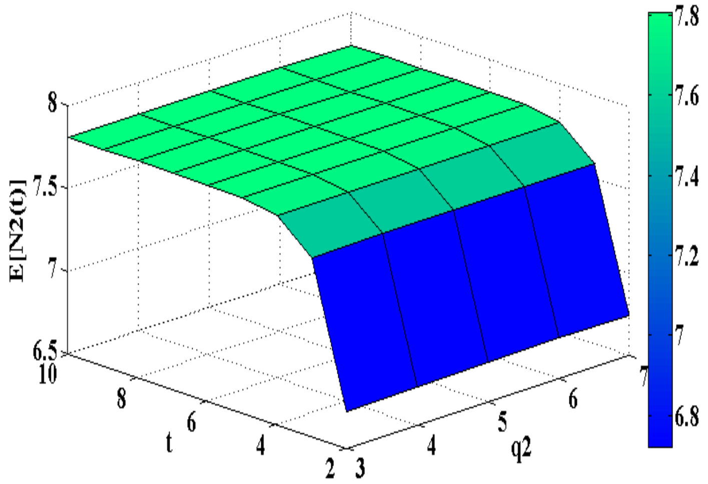


Fig. 10.5(b): Effect of q_2 on $E[N_2(t)]$ with t

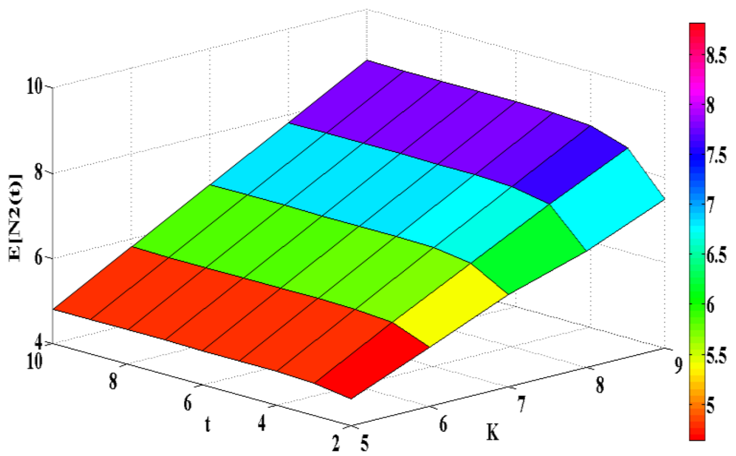


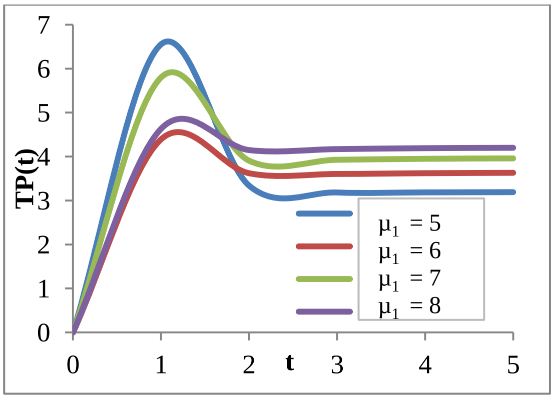
Fig. 10.5(c): Effect of K on $E[N_2(t)]$ with time t

(B) Throughput

Figs 10.6(a-b) display the effect of service rate (μ_1) and arrival rate (λ_1) respectively, on the throughput TP(t). At a particular instant, TP(t) is maximum in both the graphs 10.6(a-b) when $\mu_1 = \lambda_1$. However, in fig. 10.6(a) we can see that at t=2 units and onwards, TP(t) shows a steady state behavior even on increasing the service rate μ_1. In fig. 10.6(b), TP(t) increases with the growth of λ_1; this is due to the fact that growth in the number of new calls automatically increases TP (t).

(C) Reliability Indices

Tables 10.2 and 10.3 present the effect of various parameters on the reliability indices; namely availability and failure frequency on the system.



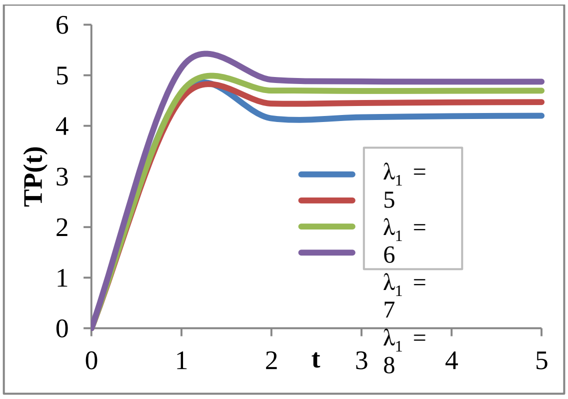


Fig. 10.6(a): Effect of μ_1 on TP(t) with t **Fig. 10.6(b): Effect of λ_1 on TP(t) with t**

Both availability $A_v(t)$ and failure frequency ($F_f(t)$) of the system decrease as time t grows. From data depicted in table 10.2, it is clear that $A_v(t)$ increases but $F_f(t)$ decreases with the growth of service rate μ_1 from 6 units to 8 units. This is due to the fact that an increase in the service rate of the server makes the server more available for the service. Further, $A_v(t)$ ($F_f(t)$) exhibits decreasing (increasing) trend with the growth of arrival rates λ_1 and λ_2. An increase in the arrival rate makes the server busier with the customers and thus it becomes less available for new incoming calls.

(D) Server state probabilities

Tables 10.3 and 10.4, display the variation in the server state probabilities with varying values of service rates (μ_1,μ_2) and retrial rates (γ_1, γ_2), respectively. The long run

Table 10.2: Effect of λ_1 on $A_v(t)$ and $F_f(t)$

λ_1	t	$A_v(t)$			$F_f(t)$		
		μ_1=6	μ_1 =7	μ_1 =8	μ_1=6	μ_1 =7	μ_1 =8
3	1	0.7837	0.7901	0.7966	0.2718	0.2680	0.2628
	3	0.6894	0.6974	0.7041	0.2490	0.2404	0.2338
	5	0.6852	0.6936	0.7009	0.2545	0.2445	0.2369
3.5	1	0.7802	0.7860	0.7915	0.2796	0.2780	0.2723
	3	0.6884	0.6962	0.7031	0.2613	0.2514	0.2436
	5	0.6832	0.6915	0.6990	0.2663	0.2550	0.2462
4	1	0.7743	0.7830	0.1059	0.2762	0.2858	0.2219
	3	0.6869	0.6946	0.6872	0.2727	0.2618	0.2314
	5	0.6810	0.6891	0.6963	0.2771	0.2649	0.2547

Table 10.3: Effect of λ_2 on $A_v(t)$ and $F_f(t)$

λ_2	t	$A_v(t)$			$F_f(t)$		
		β_1=0.8	β_1=1.2	β_1=1.4	β_1=0.8	β_1=1.2	β_1=1.4
4	1	0.7913	0.8163	0.8415	0.1194	0.2766	0.3044
	3	0.6781	0.7369	0.7575	0.2427	0.2666	0.2744
	5	0.6685	0.7361	0.7585	0.2431	0.2674	0.2751
5	1	0.7869	0.7987	0.8046	0.2790	0.2762	0.2770
	3	0.6710	0.7299	0.7509	0.2505	0.2746	0.2829
	5	0.6593	0.7278	0.7505	0.2505	0.2763	0.2845
6	1	0.7765	0.7909	0.7968	0.2951	0.3059	0.3100
	3	0.6640	0.7236	0.7449	0.2564	0.2813	0.2900
	5	0.6516	0.7210	0.7440	0.2563	0.2833	0.2920

probabilities $P_B(t)$, $P_f(t)$, $P_R(t)$ and $P_{RE2}(t)$ increase with the growth of time t whereas $P_{R3}(t)$ and $P_{RE1}(t)$ exhibit decreasing pattern. The long run probability of the server being in busy state i.e. $P_B(t)$ increases with the rise in service rate and retrial rates. This is so because an increase in service rates makes the server more available to serve and increases its busy behavior. An increase in retrial rates for both types of calls also increases $P_B(t)$.

10.7 COST OPTIMIZATION

The present section is devoted to the cost optimization. The cost function is

$$TC(L,K,q_1,q_2) = C_B P_B(t) + C_h E[N(t)] + C_a P_R(t) + C_b P_{R3}(t) + \gamma_1 C_c + \gamma_2 C_d \qquad (10.66)$$

where,

C_B : Cost per unit time when the channel is busy;

C_h : Holding cost per unit time of each call present in the system;

C_a : Repair cost incurred per unit time for a broken down channel;

C_b : Cost incurred per unit time for the channel being in broken down state but the repair is not yet started;

C_c : Fixed cost incurred when handoff calls (priority customer) retry for the service each time;

C_d : Fixed cost incurred when new calls (non-priority customer) retry for the service each time.

Table 10.4: Effect of (μ_1, μ_2) on the server state probabilities

(μ_1,μ_2)	t	$P_B(t)$	$P_f(t)$	$P_R(t)$	$P_{R3}(t)$	$P_{RE1}(t)$	$P_{RE2}(t)$
(6,7)	**1**	0.5182	0.1872	0.1243	0.0721	0.0182	0.0541
	3	0.5844	0.3095	0.2768	0.0398	0.0003	0.0986
	5	0.5906	0.3171	0.2882	0.0351	0.0000	0.0861
(6.5,7)	**1**	0.5323	0.2135	0.1469	0.0863	0.0209	0.2003
	3	0.5723	0.3049	0.2698	0.0433	0.0004	0.1140
	5	0.5778	0.3124	0.2806	0.0391	0.0000	0.1025
(8,7)	**1**	0.5021	0.2000	0.1392	0.0767	0.0253	0.1397
	3	0.5408	0.2915	0.2516	0.0512	0.0008	0.1555
	5	0.5436	0.2982	0.2605	0.0483	0.0000	0.1476
(6,8)	**1**	0.4854	0.2266	0.1665	0.0800	0.0203	0.2339
	3	0.5761	0.3047	0.2725	0.0393	0.0004	0.1116
	5	0.5834	0.3130	0.2844	0.0347	0.0000	0.0975
(6,9)	**1**	0.5522	0.2086	0.1469	0.0775	0.0214	0.1857
	3	0.5686	0.3002	0.2684	0.0388	0.0004	0.1237
	5	0.5763	0.3091	0.2808	0.0343	0.0000	0.1085

Table 10.5: Effect of (γ_1, γ_2) on the server state probabilities

(γ_1, γ_2)	t	$P_B(t)$	$P_f(t)$	$P_R(t)$	$P_{R3}(t)$	$P_{RE1}(t)$	$P_{RE2}(t)$
(1,0.5)	**1**	0.5162	0.2016	0.1365	0.0851	0.0216	0.2250
	3	0.5341	0.2878	0.2484	0.0506	0.0005	0.1664
	5	0.5375	0.2947	0.2574	0.0478	0.0000	0.1574
(1.5,0.5)	**1**	0.5188	0.2021	0.1369	0.0852	0.0188	0.2243
	3	0.5339	0.2880	0.2484	0.0507	0.0003	0.1666
	5	0.5374	0.2947	0.2574	0.0478	0.0000	0.1574
(2,0.5)	**1**	0.5203	0.2025	0.1371	0.0854	0.0164	0.2243
	3	0.5338	0.2880	0.2485	0.0507	0.0002	0.1667
	5	0.5374	0.2947	0.2574	0.0478	0.0000	0.1574
(2,2)	**1**	0.5897	0.2066	0.1421	0.0827	0.0170	0.1473
	3	0.5509	0.2984	0.2574	0.0525	0.0004	0.1385
	5	0.5538	0.3040	0.2656	0.0492	0.0000	0.1314
(2,2.5)	**1**	0.5929	0.2082	0.1429	0.0839	0.0174	0.1411
	3	0.5553	0.3011	0.2598	0.0529	0.0004	0.1311
	5	0.5581	0.3065	0.2678	0.0495	0.0000	0.1246

In order to obtain the optimal values of capacity and threshold recovery parameters (L, K, q_1, q_2), we minimize the total cost of the system. The non-linear optimization problem is solved by using direct search approach based on discrete allocation.

The optimization problem (OP) is formulated mathematically as:

$$\text{(OP):}\quad TC(L^*, q_1^*, K^*, q_2^*) = \text{Minimize}\, TC(L, q_1, K, q_2)$$

$$\text{subject to:}\quad 1 \le q_1 \le L-1$$

$$\text{and,}\quad 1 \le q_2 \le K-1$$

For numerical computations, the default values of different cost parameters are taken as C_B=20, C_h=15, C_a=25, C_b=10, C_c=40, C_d=40. The default values for other parameters are considered as $\lambda_1 = 5, \lambda_2 = 6, \gamma_1 = 0.5, \gamma_2 = 1,\ \mu_1 = 8,\ \mu_2 = 7,\ \beta_1 = 1, \beta_2 = 2, \alpha_1 = 0.5, \alpha_2 = 0.5$.

Also, we fix t as 2 units and determine the optimal values of (L, K, q_1, q_2) using direct search approach.

10.7.1 Determination of optimal set (q_1*, q_2*)

Now, we proceed to find out the optimal set of threshold parameters (q_1*, q_2*) that produces minimum cost for the concerned queueing system.

Illustration 10.2: Let us consider the double orbit retrial queueing model with the following features:

(a) The system can accommodate a maximum of 6 (i.e. L=6) handoff calls and 8 (i.e. K=8) new calls.

(b) The threshold parameters are q_1 and q_2 such that $1 \le q_1 \le L-1;\ 1 \le q_2 \le K-1$.

(c) The values of other default parameters are fixed as: $\lambda_1 = 5, \lambda_2 = 6, \gamma_1 = 0.5, \gamma_2 = 1,\ \mu_1 = 6,\ \mu_2 = 7,\ \beta_1 = 1, \beta_2 = 2, \alpha_1 = 0.5, \alpha_2 = 0.5$

In order to find the optimal values, we use "*Direct Search Approach*". In table 10.6(a), we vary q_1 and q_2 within permissible limits and search for that optimal set of threshold parameters (q_1*, q_2*) that produces minimum cost for the queueing system. Table 10.6(a) displays the numerical result for illustration 10.2. The optimal values for (q_1*, q_2*) are obtained as **(2, 2)** as shown by bold digits with their corresponding optimal cost, average queue length and throughput. Tables 10.6(b)-10.6(d) display the optimal parameters obtained for different values of arrival rates (λ_1, λ_2), retrial rates (γ_1, γ_2) and repair rates (β_1, β_2), respectively.

Table 10.6(a): Effect of arrival rates (λ_1, λ_2) on the optimal parameters (q_1*, q_2*)

q_1	q_2	TC(t)	E[N(t)]	TP(t)	q_1	q_2	TC(t)	E[N(t)]	TP(t)
1	1	269.78	11.935	5.2523	3	4	226.88	9.8972	3.6198
1	2	253.55	11.121	5.0683	3	5	227.36	9.9351	3.5720
1	3	253.95	11.141	5.0843	3	6	227.33	9.9319	3.5622
1	4	254.28	11.158	5.0996	3	7	227.94	9.969	3.5730
1	5	254.48	11.167	5.1081	4	1	242.07	10.735	3.7173
1	6	254.56	11.17	5.1078	4	2	227.7	9.9826	3.4739
1	7	254.85	11.185	5.1123	4	3	227.62	9.9765	3.4661
2	1	239.71	10.503	3.9798	4	4	228.31	10.02	3.4594
2	**2**	**224.94**	**9.7637**	**3.7011**	4	5	228.68	10.045	3.4382
2	3	225.22	9.7811	3.6982	4	6	228.66	10.044	3.4248
2	4	225.52	9.7996	3.6912	4	7	229.46	10.087	3.4775
2	5	225.78	9.8159	3.6814	5	1	242.6	10.812	3.5632
2	6	225.69	9.8087	3.6725	5	2	228.62	10.076	3.3167
2	7	226.32	9.8476	3.6756	5	3	225.54	9.8734	3.4303
3	1	257.31	11.782	2.6954	5	4	227.66	10.009	3.4261
3	2	226.45	9.8784	3.5984	5	5	229.62	10.14	3.2812
3	3	225.7	9.823	3.6143	5	6	229.57	10.136	3.2689
					5	7	230.2	10.18	3.2716

Table 10.6(b): Effect of arrival rates (λ_1, λ_2) on the optimal parameters (q_1*, q_2*)

(λ_1, λ_2)	(4,5)	(4,6)	(5, 5)	(5,6)	(5,7)	(6,5)	(6,6)	(6,7)
q_1*	4	5	4	2	5	2	3	3
q_2*	4	2	6	2	2	2	7	3
TC(q_1*,q_2*)	184.86	180.02	153.25	224.94	220.95	231.49	171.75	222.79
E[N(t)]	7.4692	7.231	5.46	9.76	9.65	10.15	6.20	9.55
TP(t)	1.1814	5	0.96	3.70	2.78	3.79	3.56	3.79

Table 10.6(c): Effect of retrial rates (γ_1, γ_2) on the optimal parameters (q_1*, q_2*)

(γ_1, γ_2)	(0.5,0.5)	(0.5,1.0)	(0.5,1.5)	(1.0,0.5)	(1.0,1.0)	(1.0,1.5)	(1.5,0.5)	(1.5,1.0)
q_1*	2	2	5	2	2	5	3	2
q_2*	1	2	4	1	2	4	1	2
TC(q_1*,q_2*)	218.93	224.94	175.49	218.69	244.92	196.66	228.94	264.85
E[N(t)]	10.46	9.76	5.74	9.13	9.76	5.8192	8.51	9.7574
TP(t)	3.93	3.70	0.15	3.71	3.70	0.17534	3.60	3.7029

Table 10.6(d): Effect of repair rates (β_1, β_2) on the optimal parameters (q_1*, q_2*)

(β_1, β_2)	(1,2)	(2,2)	(3,2)	(4,2)	(1,3)	(1,4)	(2,3)	(2,4)
q_1*	2	2	2	2	4	4	3	5
q_2*	2	2	1	1	6	3	1	1
TC(q_1*,q_2*)	224.94	220.21	185.76	196.1	214.62	223.95	88.737	172.11
E[N(t)]	9.76	9.4885	6.9324	7.6528	9.1293	9.6608	0.34043	5.9801
TP(t)	3.70	4.1921	5.1593	5.1652	3.3621	4.5583	6.3451	5.4462

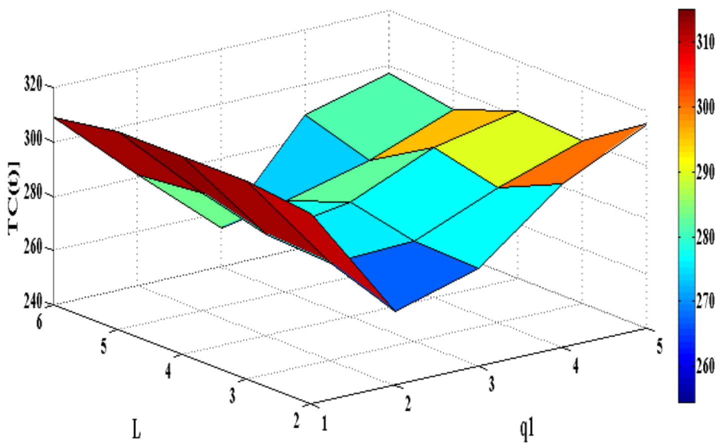


Fig. 10.7: Determination of optimal parameter (q_1*, L*)

Illustration 10.3: Consider the default parameters $\lambda_1 = 5, \lambda_2 = 6, \gamma_1 = 0.5, \gamma_2 = 1,$ $\mu_1 = 6,\ \mu_2 = 7,\ \beta_1 = 1, \beta_2 = 2, \alpha_1 = 0.5, \alpha_2 = 0.5$ for the retrial queueing system with double orbits having fixed capacity of new calls as 8 units and threshold parameter for new calls as 4 units with the set of default parameters. Moreover, we vary the capacity of handover calls (L) from one base station to another base station from 2 to 6 and the corresponding threshold recovery parameter (q_1) from 1 to 5 (i.e. L-1). Figure 10.7 displays the variation in the optimal cost of the system with different values of 'L' by 'q_1'at constant t=2 units. It is clear from the figure that a change point i.e. dip is visible corresponding to q_1=2. However, the cost increases with the increment in L.

10.7.2 Determination of optimal parameter set (L*, q_1*, K*, q_2*)

In the previous subsection, we have computed the optimal parameter set (q_1*, q_2*) and further corresponding optimal cost of the queueing model by keeping L and K as constant. Now, we determine optimal parameter set (L*, q_1*, K*, q_2*) corresponding to the minimum cost using direct search approach by varying L, q_1, K and q_2 within the assumed bounds. The various bounds are as:

(a) Capacity of the handover calls (L): $2 \le L \le 6$

(b) Capacity of the new calls (K): $2 \le K \le 8$

(c) Threshold to start the repair of channel for handover calls (q_1): $1 \le q_1 \le L-1$

(d) Threshold to start the repair of channel for new calls (q_2): $1 \le q_2 \le K-1$

Different sets of optimal parameters (L*, q_1*, K*, q_2*) are obtained corresponding to various sets of default parameters.

We fix $\mu_1 = 6$ and 8 for tables 10.7(a) and 10.7(b), respectively. The set of other parameters taken are: $\lambda_1 = 5, \lambda_2 = 6, \gamma_1 = 0.5, \gamma_2 = 1, \mu_2 = 7, \beta_1 = 1, \beta_2 = 2, \alpha_1 = 0.5, \alpha_2 = 0.5.$ Tables 10.7(a-b) summarizes the various optimal set for (L*, q_1*, K*, q_2*) for different set of (λ_1, λ_2) and (γ_1, γ_2).

Illustration 10.4: From table 10.6(c), we consider the case when (γ_1, γ_2) = (1.0, 1.5) and corresponding optimal threshold parameters (q_1*, q_2*) as (5, 4). We vary the capacity size of handover calls (L) from 6 to 12 and capacity parameter of new calls i.e. K from 5 to 8 (minimum value of L and K must be greater than their corresponding threshold repair parameters i.e. q_1 =5 and q_2=4). Figure 10.8 shows the variation in the total cost TC (t) of the system with varying values of L and K. It is observed from the figure as well from the data depicted in table 10.8 that the minimum cost (shown by bold letters) 92 units approximately is obtained for (L*, K*) = (8, 7).

Table 10.7(a): Effect of (λ_1, λ_2) on optimal parameters (L*, q_1*, K*, q_2*)

Optimal parameters	λ_1=2, λ_2=3	λ_1=4, λ_2=6	λ_1=5, λ_2=5	λ_1=3, λ_2=4
L*	5	6	4	6
q_1*	3	5	2	1
K*	6	3	8	8
q_2*	4	2	1	2
TC (L*,q_1*,K*,q_2*)	21.50	135.29	81.53	48.14

Table 10.7(b): Effect of (γ_1, γ_2) on optimal parameters (L*, q_1*, K*, q_2*)

Optimal parameters	γ_1=1.0, γ_2=1.0	γ_1=1.5, γ_2=1.0	γ_1=2,γ_2=3	γ_1=0.5,γ_2=0.5
L*	2	3	3	4
q_1*	1	2	2	2
K*	2	3	3	3
q_2*	1	2	2	2
TC (L*,q_1*,K*,q_2*)	85.46	175.47	276.36	120.38

Table 10.8: Total cost of the system corresponding to different values of L and K

L\K	5	6	7	8
6	220	233	243	253
7	219	231	242	252
8	190	166	**92**	262
9	231	244	255	264
10	233	245	257	266
11	235	249	258	268
12	236	247	260	275

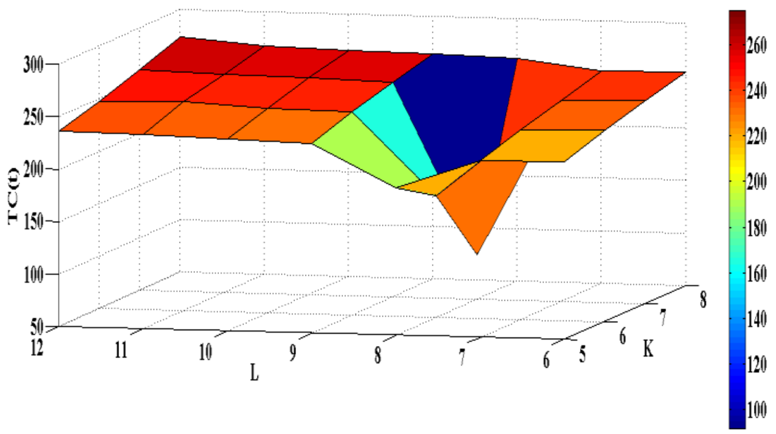


Fig. 10.8: Determination of optimal parameter (L*, K*)

The optimal parameters (L*, q_1*, K*, q_2*) determined can be used for the design of optimal systems by setting the buffer capacity of orbit size. A cellular mobile network with optimal capacity for both types of calls and optimal threshold parameters can be designed which may prove economical to deal with the dropping of calls in real time system.

10.7.3 Determination of optimal service rates

In sub sections 10.7.1 and 10.7.2 we have computed optimal cost TC (L^*, K^*, q_1^*, q_2^*) by using the optimal parameters. Now, we intend to determine optimal service rates (μ_1^*, μ_2^*) so as to know the optimal service rate at which calls must be served at the minimum cost.

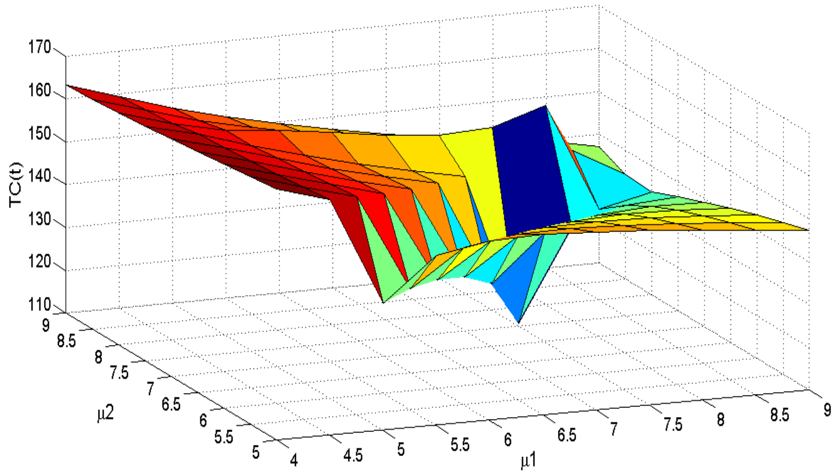


Fig. 10.9: Determination of optimal parameter (μ_1^*, μ_2^*)

Illustration 10.5: From table 10.7(a), we consider optimal parameters (L^*,K^*,q_1^*,q_2^*) = (6,5,3,2) with default parameters as $\lambda_1 = 4, \lambda_2 = 6, \gamma_1 = 0.5, \gamma_2 = 1, \beta_1 = 1, \beta_2 = 2,$ $\alpha_1 = 0.5, \alpha_2 = 0.5$ and vary service rates μ_1 from 4 to 9 units and μ_2 from 5 to 9 units. Now, proceed to find out the optimal values of both service rates which provide minimum cost. In figure 10.9, we display the TC(t) corresponding to different service rates and cost values.

We can easily observe from the table 10.7(a) that corresponding to this particular set of optimal parameters (L^*, K^*, q_1^*, q_2^*) = (6, 5, 3, 2) we have obtained optimal service rates μ_1=8 and μ_2=7 and the corresponding cost was TC (t) = 135.29 units. But it is remarkable to observe from the fig. 10.9 that with the variation in service rates minimum cost is obtained as TC (t) = 110.52 units at **μ_1= μ_2=7.5 units.** By taking illustration, we have demonstrated that the system can be made more economical by serving the calls or customers at optimal service rates. This gives optimal decision parameters as $L^*=6$, $K^*=5$, $q_1^*=3$, $q_2^*=2$, $\mu_1= \mu_2=7.5$.

10.8 DISCUSSION

In this chapter, the double orbit finite retrial queue with two types of customers had been investigated. The cost optimization of the system had been proposed to determine the optimal parameters. The work presented in this chapter seems to be useful for the construction of optimal system designs wherein priority to one kind of traffic is given in comparison to other type of traffic. The model can be applicable to various congestion situations encountered in telecommunication systems, hospitals, banks, manufacturing systems which involve servicing of two types of customers under certain priority rule. The sensitivity analysis demonstrates the tractability of proposed model in context of its applications to the cellular mobile network. Overall, we can conclude that:

- We infer that the availability of the server can be increased by controlling the arrival rate of the new incoming calls. Moreover, the system can be made more efficient by enhancing the service rate, which in turn reduces the failure frequency of the server.
- The determination of optimal threshold recovery parameters (q_1^*, q_2^*) can be of interest in order to initiate the start of the repair of broken down server. On this basis, cellular mobile network can handle the breakdown of the channel and its repair by choosing optimal threshold parameters which may prove economical in terms of both time and money.

CONCLUSIONS

Stochastic modeling of congestion problems with reattempts holds a significant place in the area of queueing theory in the form of retrial queues. Due to the abundant applications of such queues in day-to-day activities as well as in various industrial scenarios including manufacturing and production systems, computer and telecommunication systems, etc. have forced queue theorists to develop new models which can be well suited to real life congestion situations. In the present doctoral work, an attempt is made to develop and analyze retrial queueing models enriched with various prominent features like unreliable server, vacation, bulk arrival, impatience behavior, threshold recovery, priority, etc. so as to study the complex queueing systems arising out of a variant of congestion phenomenon. The highlights of the noble features of the work done in the present doctoral thesis are as follows:

- The reliability of the server greatly affects the performance and efficiency of any queueing model. Therefore, unreliability of the server cannot be neglected and thus retrial queueing models with unreliable server are more consistent enough with the real life situations. We have developed retrial queueing models by incorporating the assumption of ***unreliable server*** which has wide applicability in many areas such as in computer and communication systems. It is worth noting that the concept of unreliability of the server is taken into account in all the models studied in chapters from 2-10.
- ***Vacation retrial queueing models*** are examined in chapters 2 and 3, which may be useful for the queueing scenarios with re-attempts. The concept of ***Bernoulli vacation schedule*** is incorporated in chapter 3 where the server may go for vacation after each service or may continue to serve the customers with complementary probability. ***Modified vacation policy*** discussed in chapter 4 allows the server to go for a maximum of J vacations in case no customers are present in the system.
- ***Priority retrial queueing models*** considered in chapter 4, 5 and 10 have numerous applications in many real life situations encountered in computer networks, communication systems, transportation and many others.
- The repair of the broken down server is a key issue in order to continue the service of the customers waiting in the queue. ***Threshold based recovery*** can be used by

the technicians and maintenance engineers to repair the broken down server. The repair of the server should be done in optimized manner so as to save both time and money. This policy is incorporated in chapters 7 and 8 in order to determine the optimal threshold parameter and corresponding optimal cost.

- ***Bulk arrivals*** have significant impact on the behavior of the queueing systems with reattempts and have a variety of applications in queueing situations encountered in computer sharing systems, communication traffic, manufacturing processes, etc. The bulk arrival retrial queueing models investigated in chapters 2-7 also include other prominent features like vacation, N-policy, optional services, etc.
- The service pattern usually differs from one system to another; it may be single phase service pattern, multi-phase service or a series of some optional services. Due to its wide applicability to numerous systems, the concept of ***multi services/repairs*** is incorporated in chapters 3-5.
- The incorporation of ***discouragement behavior*** (i.e. balking and reneging) in the modeling of most of the retrial queueing systems is done in chapters 2- 6. The balking or reneging behavior of the customers is common and realistic, which arises due to impatience of the customers on seeing a long queue or server being busy. The customer's satisfaction is the main goal of any service sector
- Sometimes, the system designer is more interested in knowing the behavior of the server at a particular instant of time instead of judging the long run or steady state behavior of the system. The ***transient state solution*** is thus required to have a idea of the server's status and provides more realistic characteristics of the queueing systems (cf. chapters 8, 9 and 10).
- ***Optimal control*** of the queue is also a major key concern of the queue theorists and system designers. This is basically useful in optimizing the cost or to prevent the queueing system from bursting due to excessive crowd in the system. Various control policies namely N-policy (cf. chapter 6), F-policy (cf. chapter 8), threshold recovery (cf. chapters 8 and 10) have incorporated in our study to frame retrial queueing systems so as to determine the optimal system parameters.

The retrial queueing models investigated in this doctoral work are not limited to the concepts considered in this research work; they can be further enhanced by adding more realistic assumptions so as to model more complex real life situations. The stochastic modeling and performance analysis of retrial queueing models can also be done by using real time data for some industry rather than using heuristic data. The future extensions of present work can be done in the following directions:

- There is a wide scope of bulk service retrial models. Various retrial queueing models developed can be further extended by incorporating bulk service concept but analytical results will become more cumbersome to derive.
- Various control policies namely N-policy, F-policy, threshold recovery can be established by combining together with fuzzy parameters. The genetic algorithm based optimization and fuzzy logic may be of great advantage to the system organizers and industrial engineers to design the concerned queueing systems in an optimized manner.
- Retrial queueing model with finite double orbits and priority investigated in our research work (chapter-10) can be further extended by adding more orbits for multi-class customers.

The retrial queues find applications in almost every sphere of life from daily routine activities to various complex real life situations. Most of the queueing situations can be modeled as retrial queueing problem and our study can be utilized to suggest the means and ways for improving the grade of service in terms of trade-off between delay and cost. Keeping this in view, cost optimization Markov models to study the performance analysis of retrial queues are developed. It is hoped that the investigations presented in the present thesis may be helpful in upgrading the many existing queueing systems with retrials. The research work done provides valuable insight to the system managers and decision makers for the quantitative assessment of the performance of the concerned systems. The performance measures obtained may be helpful to the system designers and decision makers in improving the efficiency of the systems. The study done will be of great importance not only from theoretical point of view but will strongly reflect the practical and managerial implementation.

REFERENCES

[1] Aissani, A. (2009): An M^X/G/1 energetic retrial queue with vacations and it's control, Electronic Notes in Theoretical Computer Science, Vol. 253, No. 3, pp. 33-44.

[2] Aissani, A., Taleb, S. Kernane, T. Saidi, G. and Hamadouche, D. (2014): An M/G/1 retrial queue with working vacation, Advances in Intelligent Systems and Computing, Vol. 240, pp. 443-452.

[3] Alfa, A. S. and Isotupa, K. P. S. (2004): An M/PH/k retrial queue with finite number of sources, Computers and Operations Research, Vol. 31, No. 9, pp. 1455-1464.

[4] Almasi, B., Roszik, J. and Sztrik, J. (2005): Homogeneous finite-source retrial queues with server subject to breakdowns and repairs, Mathematical and Computer Modeling, Vol. 42, No. 5-6, pp. 673-682.

[5] Arivudainambi, D., Godhandaraman, P. and Rajadurai, P. (2013): Performance analysis of a single server retrial queue with working vacation, OPSEARCH, doi: 10.1007/s12597-013-0154-1.

[6] Arrar, N. K., Djellab, N. V. and Baillon, J. B. (2012): On the asymptotic behavior of retrial queues with batch arrivals and impatience phenomenon, Mathematical and Computer Modelling, Vol. 55, No. 3-4, pp. 654-665.

[7] Artalejo, J. R. (1999a): A classified bibliography of research on retrial queues: Progress in1990-1999, Top, Vol. 7, pp. 187-211.

[8] Artalejo, J. R. (1999b): Accessible bibliography on retrial queues, Mathematical Computer Modeling, Vol. 30, No. 3-4, pp. 1-6.

[9] Artalejo, J. R. (2010): Accessible bibliography on retrial queues, Mathematical Computer Modeling, Vol. 51, No. 9-10, pp. 1071-1081.

[10] Artalejo, J. R. and Atencia, I. (2004): On the single server retrial queue with batch arrivals, Sankhyá, Vol. 66, No. 1, pp. 140–158.

[11] Artalejo, J. R., Atencia, I. and Moreno, P. (2005): A discrete-time $Geo^{[x]}$/G/1 retrial queue with control of admission, Applied Mathematical Modelling, Vol. 29, No. 11, pp. 1100–1120.

[12] Atencia, I., Bouza, G. and Moreno, P. (2008): An $M^{[X]}/G/1$ retrial queue with server breakdowns and constant rate of repeated attempts, Annals of Operations Research, Vol. 157, No. 1, pp. 225-243.

[13] Artalejo, J. R. and Choudhury, G. (2004): Steady state analysis of an M/G/1 queue with repeated attempts and two-phase service, Quality Technology and Quantitative Management, Vol. 1, No. 2, pp. 189-199.

[14] Artalejo, J. R. and Falin, G. I. (2002): Standard and retrial queueing systems: A comparative analysis, Revista Matemática Complutense, Vol. 15, No. 1, pp. 101-129.

[15] Artalejo, J. R. and Gomez-Corral, A. (1997): Steady state solution of a single-server with linear repeated requests, Journal of Applied Probability, Vol. 34, No. 1, pp. 223-233.

[16] Artalejo, J. R. and Gomez-Corral, A. (2008): Retrial Queueing Systems: A Computational Approach, Springer-Verlag, Berlin.

[17] Artalejo, J. R. and Phung-Duc, T. (2013): Single server retrial queues with two way communication, Applied Mathematical Modelling, Vol. 37, No. 4, pp. 1811–1822.

[18] Atencia, I., Fortes, I., Moreno, P. and Sanchez, S. (2006a): An M/G/1 retrial queue with active breakdowns and Bernoulli schedule in the server, Information and Management Sciences, Vol. 17, No. 1, pp. 1-17.

[19] Atencia, I., Moreno, P. and Bouza, G. (2006b): An $M_2/G_2/1$ retrial queue with priority customers, 2^{nd} optional service and linear retrial policy, Revista de Investigación Operacional, Vol. 27, No. 3, pp. 229-248.

[20] Avrachenkov, K., Dudin, A. and Klimenok, V. (2010): Retrial queueing model $MMAP/M_2/1$ with two orbits, Multiple Access Communications Lecture Notes in Computer Science, Vol. 6235, pp. 107-118.

[21] Bailey, N.T. J. (1954): On queueing processes with bulk service, Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 16, pp. 80-87.

[22] Balasubramanian, M. and Arumuganathan, R. (2011): Steady state analysis of a bulk arrival general bulk service queueing system with modified M-vacation policy and variant arrival rate, International Journal of Operational Research, Vol. 11, No. 4, pp. 383-407.

[23] Banik, A. D. (2009): The infinite-buffer single server queue with a variant of multiple vacation policy and batch Markovian arrival process, Applied Mathematical Modelling, Vol. 33, No.7, pp. 3025-3039.

[24] Banik, A. D. (2010): Analysis of single working vacation in GI/M/1/N and GI/M/1/∞ queueing systems, International Journal of Operational Research, Vol. 7, No. 3, pp. 314-333.

[25] Banik, A. D., Gupta, U. C. and Chaudhry, M. L. (2007): Finite-buffer bulk service queue under Markovian service process, Valuetools'07, Proceedings of the 2nd international conference on Performance evaluation methodologies and tools, October 23-25, 2007, Nantes, France, Article 55.

[26] Begin, T., Baynat, B., Sourd, F. and Brandwajn, A. (2010): A DFO technique to calibrate queueing models, Computers and Operations Research, Vol. 37, No. 2, pp. 273-281

[27] Berdjoudj, L. and Aissani, D. (2014): An analysis of the M/G/1 retrial queue with negative arrivals using martingale technique, International Journal of Mathematical Sciences, Vol. 196, No. 1, pp. 11-14.

[28] Bernett, H. G., Fischer, M. J., and Masi, D. M. B. (2002): Blended call center performance analysis, IT Professional, Vol. 4, No. 2, pp. 33–38.

[29] Bhargava, C. and Jain, M. (2014): Unreliable multiserver queueing system with modified vacation policy, OPSEARCH, Vol. 51, No. 2, pp. 159-182.

[30] Bhulai, S., and Koole, G. (2003): A queueing model for call blending in call centers. IEEE Transactions on Automatic Control, Vol. 48, No. 8, pp. 1434–1438.

[31] Borthakur, A. and Medhi, J. (1974): A queueing system with arrivals and service in batches of variable size, Cah. Du. Centred'eludes de Rech. Oper., Vol. 16, pp. 117-126.

[32] Boualem, M. (2014): Insensitive bounds for the stationary distribution of a single server retrial queue with server subject to active breakdowns, Advances in Operations Research, Article ID 985453(In Press).

[33] Boualem, M., Djellab, N. and Aissani, D. (2009): Stochastic inequalities for M/G/1 retrial queues with vacations and constant retrial policy, Mathematical and Computer Modelling, Vol. 50, No. 1-2, pp. 207-212.

[34] Brandwajn, A. and Begin, T. (2009): Higher-order distributional properties in closed queueing networks, Performance Evaluation, Vol. 66, pp. 660-659.

[35] Brandwajn, A. and Wang, H. (2008): A Conditional Probability Approach to M/G/1-like Queues, Performance Evaluation, Vol. 65, pp. 366-381.

[36] Chang, F. M. and Ke, J. C. (2009): On a batch retrial model with J vacations, Journal of Computational and Applied Mathematics, Vol. 232, No. 2, pp. 402-414.

[37] Chaudhry, M. L., Yoon, B. K. and Kim, N. K. (2010): On the distribution of the number of customers in the D-BMAP/$G^{a,\ b}/1/M$ queue-A simple approach to a complex problem, INFOR: Information Systems and Operational Research, Vol. 48, No. 2, pp. 121-132.

[38] Choi, B. D. and Chang, Y. (1999): Single server retrial queues with priority calls, Mathematical and Computer Modelling, Vol. 30, No. 3-4, pp. 7-32.

[39] Choi, B. D., Chang, Y. and Kim, B. (1999): MAP_1, MAP_2/M/c retrial queue with guard channels and its applications to cellular networks, Top, Vol. 7, No. 2, pp. 231-248.

[40] Choi, B. D. and Park, K. K. (1990): The M/G/1 retrial queue with Bernoulli schedule, Queueing Systems, Vol.7, No. 2, pp. 219-228.

[41] Choudhury, G. (2007): A two phase batch arrival retrial queueing system with Bernoulli vacation schedule, Applied Mathematics and Computation, Vol. 188, No. 2, pp. 1455-1466.

[42] Choudhury, G. (2008a): A single server queueing system with two phases of service and vacations, Quality Technology and Quantitative Management, Vol. 5, No. 1, pp. 33-49.

[43] Choudhury, G. (2008b): Steady state analysis of an M/G/1 queue with linear retrial policy and two phase service under Bernoulli vacation schedule, Applied Mathematical Modelling, Vol. 32, No. 12, pp. 2480-2489.

[44] Choudhury, G. (2009): An M/G/1 retrial queue with an additional phase of second service and general retrial times, International Journal of Information and Management Sciences, Vol. 20, pp. 1-14.

[45] Choudhury, G. and Deka, K. (2008): An M/G/1 retrial queueing system with two phases of service subject to server breakdown and repair, Performance Evaluation, Vol. 65, No. 10, pp. 714-724.

[46] Choudhury, G. and Deka, K. (2009): An M^X/G/1 unreliable retrial queue with two phases of service and Bernoulli admission mechanism, Applied Mathematics and Computation, Vol. 215, No. 3, pp. 936-949.

[47] Choudhury, G. and Deka, M. (2012): A single server queueing system with two phases of service subject to server breakdown and Bernoulli vacation, Applied Mathematical and Modelling, Vol. 36, No. 12, pp. 6050-6060.

[48] Choudhury, G. and Deka, K. (2013): A batch arrival retrial queue with two phases of service and Bernoulli vacation schedule, Acta Mathematicae Applicatae Sinica, English Series, Vol. 29, pp. 15–34.

[49] Choudhury, G. and Ke, J. C. (2012): A batch arrival retrial queue with general retrial times under Bernoulli vacation schedule for unreliable server and delaying repair, Applied Mathematical Modelling, Vol. 36, No. 1, pp. 255–269.

[50] Choudhury, G. and Ke, J. C. (2014): An unreliable retrial queue with delaying repair and general retrial times under Bernoulli vacation schedule, Applied Mathematics and Computation, Vol. 230, pp. 436–450.

[51] Choudhury, G., Ke, J. C. and Tadj, L. (2009): The N-policy for an un-reliable server with delaying repair and two phases of service, Journal of Computational and Applied Mathematics, Vol. 231, No. 1, pp. 349-364.

[52] Choudhury, G. and Paul, M. (2004): A batch arrival queue with an additional service channel under N-policy, Applied Mathematics and Computation, Vol. 156, No. 1, pp. 115-130.

[53] Choudhury, G. and Tadj, L. (2011): The optimal control of an $M^x/G/1$ unreliable server queue with two phases of service and Bernoulli vacation schedule, Mathematical and Computer Modelling, Vol. 54, No. 1-2, pp. 673-688. (chapter-4)

[54] Choudhury, G., Tadj, L. and Deka, K. (2010): A batch arrival retrial queueing system with two phases of service and service interruptions, Computers and Mathematics with Applications, Vol. 59, No. 1, pp. 437-450.

[55] Cinlar, E. (1975): Introduction to Stochastic Processes. Prentice-Hall, Englewood Cliffs, New Jersey.

[56] Claeys, D., Steyaert, B., Walraevens, J., Laevens, K. and Bruneel, H. (2013): Analysis of a versatile batch-service queueing model with correlation in the arrival process, Performance Evaluation, Vol. 70, No. 4, pp. 300-316.

[57] Claeys, D., Walraevens, J., Laevens, K. and Bruneel, H. (2011): Analysis of threshold-based batch-service queueing systems with batch arrivals and general service times, Performance Evaluation, Vol. 68, No. 6, pp. 528-549.

[58] Cooper, R. B. (1981): Introduction to Queueing theory, North-Holland, New York.

[59] Cordeau, J. L. and Chaudhry, M. L. (2009): A simple and complete solution to the stationary queue-length probabilities of a bulk-arrival bulk-service queue, INFOR: Information Systems and Operational Research, Vol. 47, No. 4, pp. 283-288.

[60] Cox, D. R. (1955): The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables, Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 51, No. 3, pp. 433-441.

[61] Deslauriers, A., L'Ecuyer, P., Pichitlamken, J., Ingolfsson, A., Avramidis, and A.N. (2007): Markov chain models of a telephone call center with call blending, Computers and Operations Research, Vol. 34, No. 6, pp. 1616–1645.

[62] Dharmaraja, S., Jindala, V. and Alfa, A. S. (2008): Phase-type models for cellular networks supporting voice, video and data traffic, Mathematical and Computer Modelling, Vol. 47, No. 11-12, pp. 1167–1180.

[63] Dimitriou, I. (2013a): A mixed priority retrial queue with negative arrivals, unreliable server and multiple vacations, Applied Mathematical Modeling, Vol. 37, No. 3, pp. 1295–1309.

[64] Dimitriou, I. (2013b): A preemptive resume priority retrial queue with state dependent arrivals, unreliable server and negative customers, Top, Vol. 21, No. 3, pp. 542–571.

[65] Dimitriou, I. and Langaris, C. (2008): Analysis of a retrial queue with two-phase service and server vacations, Queueing Systems, Vol. 60, No. 1-2, No. 1-2, pp. 111-129.

[66] Dimitriou, I. and Langaris, C. (2010): A repairable queueing model with two-phase service, start-up times and retrial customers, Computers and Operations Research, Vol. 37, No. 7, pp. 1181-1190.

[67] Do, T. V. (2010): A new computational algorithm for retrial queues to cellular mobile systems with guard channels, Computers and Industrial Engineering, Vol. 59, No. 4, pp. 865–872.

[68] Do, T. V. (2011a): Bibliography on G-networks, negative customers and applications, Mathematical and Computer Modelling, Vol. 53, No. 1-2, pp. 205–212.

[69] Do, T. V. (2011b): Solution for a retrial queueing problem in cellular networks with the fractional guard channel policy, Mathematical and Computer Modelling, Vol. 53, No. 11-12, pp. 2059-2066.

[70] Domenech-Benlloch, M. J., Giménez-Guzmán, J. M., Pla, V., Giner, V. C. and Martínez-Bauset, J. (2009): Solving multiserver systems with two retrial orbits using value extrapolation: A comparative perspective, Vol. 5513, pp. 56-70.

[71] Doshi, B. T. (1986): Queueing systems with vacation-A survey, Queueing Systems, Vol. 1, No. 1, pp. 29-66.

[72] Doshi, B. T. (1990): Single server queues with vacations, In H. Takagi (ed.): Stochastic Analysis of Computer and Communication Systems, North-Holland, Amsterdam, pp. 217-265.

[73] Dudina, O., Kim, C. and Dudin, S. (2013): Retrial queuing system with Markovian arrival flow and phase-type service time distribution, Computers and Industrial Engineering, Vol. 66, No. 2, pp. 360–373.

[74] Economou, A., Gomez-Corral, A. and Kanta, S. (2011): Optimal balking strategies in single-server queues with general service and vacation times, Performance Evaluation, Vol. 68, No. 10, pp. 967-982.

[75] Economou, A. and Lopez-Herrero, M. J. (2009): Performance analysis of cellular mobile network with retrials and guard channels using waiting and first passage time measures, European Transactions on Telecommunications, Vol. 20, No. 4, pp. 389-401.

[76] Efrosinin, D. and Winkler, A. (2011): Queueing system with a constant retrial rate, non-reliable server and threshold-based recovery, European Journal of Operational Research, Vol. 210, No. 3, pp. 594-605.

[77] El-Affendi, M. A. and Kouvatsos, D. D. (1983): A maximum entropy analysis of the M/G/1 and G/M/1 queueing systems in equilibrium, Acta Information, Vol.19, No. 4, pp. 339-355.

[78] Erlang, A. K. (1909): Probability and telephone calls, Nyt. Tidsskr Krarup Mat. Ser. B., Vol. 20, pp. 33-39.

[79] Falin, G. (2010a): An M/G/1 retrial queue with an unreliable server and general retrial times, Performance Evaluation, Vol. 67, No. 7, pp. 569-582.

[80] Falin, G. (2010b): A single-server batch arrival retrial queue with returning customers, European Journal of Operational Research, Vol. 201, No. 3, pp. 786-790.

[81] Falin, G. I. (1999): A multiserver retrial queue with a primary finite number of sources of primary calls, Mathematical and Computer Modeling, Vol. 30, No. 3-4, pp. 33-49.

[82] Falin, G. I. and Templeton, J. G. C. (1997): Retrial Queues, Chapman and Hall, London.

[83] Fuhramann, S. W. and Cooper, R. B. (1985): Stochastic decompositions in the M/G/1 queue with generalized vacations, Operations Research, Vol. 33, No. 5, pp. 1117-1129.

[84] Gao, S. and Wang J. (2014): Performance and reliability analysis of an M/G/1-G retrial queue with orbital search and non-persistent customers, European Journal of Operational Research, Vol. 236, No. 2, pp. 561-572.

[85] Gelenbe, E. (1989): Random neural networks with negative and positive signals and product form solution, Neural Computation, Vol. 1, No. 4, pp. 502-510.

[86] Gelenbe, E. (1991): Product-form queueing networks with negative and positive customers, Journal of Applied Probability, Vol. 28, pp. 656-663.

[87] Gelenbe, E. (2000): The first decade of G-networks, European Journal of Operations Research, Vol. 126, No. 2, pp. 231-232.

[88] Gomez-Corral, A. (2002): Analysis of a single-server retrial queue with quasi-random input and non preemptive priority, Computers and Mathematics with Applications, Vol. 43, No. 6-7, pp. 767-782.

[89] Goswami, V., Gupta, U. C. and Samanta, S. K. (2006): Analyzing discrete-time bulk-service $Geo/Geo^b/m$ queue, RAIRO - Operations Research, Vol. 40, No. 3, pp. 267-284.

[90] Goswami, C. and Selvaraju, N. (2010): The discrete-time MAP/PH/1 queue with multiple working vacations, Applied Mathematical Modelling, Vol. 34, pp. 931-946.

[91] Goswami, C. and Selvaraju, N. (2013): A working vacation queue with priority customers and vacation interruptions, International Journal of Operational Research, Vol. 17, No. 3, pp. 311-332.

[92] Gross, D. and Harris, C. M. (1985): Fundamentals of Queueing Theory, 2nd ed., John Wiley and Sons, New York.

[93] Haridass, M. and Arumuganathan, R. (2008): Analysis of a bulk queue with unreliable server and single vacation, International Journal of Open Problems in Computer Science and Mathematics, Vol. 1, No. 2, pp. 130-148.

[94] Haridass, M., Arumuganathan, R. and M. Senthilkumar (2012): Cost analysis of a bulk service retrial queue, Vol. 14, No. 1, pp. 94-119.

[95] Harrison, P. G. and Pitel, E. (1993): Sojourn times in single server queues with negative customers, Journal of Applied Probability, Vol. 30, pp. 943–963.

[96] Harrison, P. G. and Pitel, E. (1996): The M/G/1 queue with negative customers, Advances in Applied Probability, Vol. 28, pp. 540–566.

[97] Jaggi, C. K. and Arneja, N. (2010): Periodic inventory model with unstable lead-time and setup cost with backorder discount, International Journal of Applied Decision Sciences, Vol. 3, No. 1, pp. 53-73.

[98] Jaggi, C. K. and Arneja, N. (2011a): Stochastic integrated vendor-buyer model with unstable lead time and setup cost, International Journal of Industrial Engineering Computations, Vol. 2, No. 1, pp. 123-140.

[99] Jaggi, C. K. and Arneja, N. (2011b): Periodic inventory model with reduced setup cost under service level constraint, Electronic Journal of Applied Statistical Analysis, Vol. 4, No. 2, pp. 111-123.

[100] Jain, M. and Agrawal, P. K. (2009): Optimal policy for bulk queue with multiple types of server breakdown, International Journal Operational Research, Vol. 4, No. 1, pp. 35-54.

[101] Jain, M. and Agrawal, P. K. (2010): N-policy for state dependent batch arrival queueing system with *l*-stage service and modified Bernoulli schedule vacation, Quality Technology and Quantitative Management, Vol. 7, No. 3, pp. 215-230.

[102] Jain, M. and Bhargava, C. (2008): Bulk arrival retrial queue with unreliable server and priority subscribers, International Journal of Operations Research, Vol. 2, No. 3, pp. 260-278.

[103] Jain, M. and Bhargava, C. (2009): Unreliable server M/G/1 queueing system with Bernoulli feedback, repeated attempts, modified vacation, phase repair and discouragement, Journal of King Abdulaziz University: Engineering Science, Vol. 20, No. 2, pp. 45-77.

[104] Jain, M. and Jain, A. (2010): Working vacations queueing models with multiple types of server breakdowns, Applied Mathematical Modelling, Vol. 34, No. 1, pp. 1-13.

[105] Jain, M., Maheshwari, S. and Baghel, K. P. S. (2008): Queueing network modeling of flexible manufacturing system using mean value analysis, Applied Mathematical Modelling, Vol. 32, No. 5, pp.700–711.

[106] Jain, M., Sharma, G. C. and Pundhir, R. S. (2007): The M/G/1 retrial queueing system with set up, server breakdown and repair, Ganita, Vol. 58, No. 2, pp. 137-155.

[107] Jain, M., Sharma, G. C. and Sharma, R. (2010): Matrix-geometric method for working vacation queueing system with second optional service and vacation interruptions, Mathematics Today, Vol. 26, pp. 14-27.

[108] Jain, M., Sharma, G. C. and Sharma, R. (2011): Working vacation queue with service interruption and multi optional repair, International Journal of Information and Management Sciences, Vol. 22, pp. 157-175.

[109] Jain, M., Sharma, G. C. and Sharma, R. (2012a): Maximum entropy approach for discrete-time unreliable server $Geo^X/Geo/1$ queue with working vacation, International Journal of Mathematics in Operational Research, Vol. 4, No. 1, pp. 56-77.

[110] Jain, M., Sharma, G. C. and Sharma, R. (2012b): A batch arrival retrial queueing system for essential and optional services with server breakdown and Bernoulli vacation, International Journal of Internet and Enterprise Management, Vol. 8, No. 1, pp. 16-45.

[111] Jain, M., Sharma, G. C. and Sharma R. (2013): Unreliable server M/G/1 queue with multi-optional services and multi-optional vacations, International Journal of Mathematics of Operational Research, Vol. 5, No. 2, pp. 145-169.

[112] Jain, M. and Upadhyaya, S. (2009): Threshold N-Policy for degraded machining system with multiple type spares and multiple vacations, Quality Technology and Quantitative Management, Vol. 6, No. 2, pp. 185-203.

[113] Jain, M. and Upadhyaya, S. (2010): Optimal repairable $M^x/G/1$ queue with multi-optional services and Bernoulli vacation, International Journal of Operational Research, Vol. 7, No. 1, pp. 109-132.

[114] Jain, M. and Upadhyaya, S. (2012): Optimal repairable $M^x/G/1$ queue with Bernoulli feedback and setup, International Journal of Mathematics in Operational Research, Vol. 4, No. 6, pp. 679-702.

[115] Jaiswal, N. K. (1968): Priority Queues, Academic Press, New York.

[116] Jang, J. S. R. and Sun, C. T. (1995): Neuro-fuzzy modeling and control, Proceedings of the IEEE, Vol. 83, No. 3, pp. 378-406.

[117] Kajiwara, K. and Phung-Duc, T. (2014): Asymptotic and numerical analysis of multiserver retrial queue with guard channel for cellular networks, Proceedings of

the Eight International Conference on Matrix-Analytic Methods in Stochastic Models (MAM8), NIT Calicut, Kerala, India, pp. 85-102, January 06-10, http://arxiv.org/abs/1405.1916.

[118] Kapur, J. N. (1989): Maximum-Entropy Models in Science and Engineering, John Wiley and Sons.

[119] Karmeshu (Ed.) (2003): Entropy Measures, Maximum Entropy Principle and Emerging Applications, Studies in Fuzziness and Soft Computing, Vol. 119, Springer-Verlag, Berlin.

[120] Karmeshu and Khandelwal, V. (2013): On the applicability of average channel capacity in log-normal fading environment, Wireless Personal Communications, Vol. 68, No. 4, pp. 1393–1402.

[121] Ke, J. C. and Chang, F. M. (2009a): M^X/ (G_1, G_2)/1 retrial queue under Bernoulli vacation schedules with general repeated attempts and starting failures, Applied Mathematics and Computation, Vol. 32, No. 4, pp. 443-458.

[122] Ke, J. C. and Chang, F. M. (2009b): Modified vacation policy for M/G/1 retrial queue with balking and feedback, Computers and Industrial Engineering, Vol. 57, No. 1, pp. 433-443.

[123] Ke, J. C., and Wu, C. H. and Pearn, W. L. (2011): Multi-server retrial queue with second optional service: Algorithmic computation and optimization, International Journal of Systems Science, Vol. 42, No. 10, pp. 1755-1769.

[124] Keilson, J. and Servi, L. D. (1986): Oscillating random walk models for GI/G/1 vacation system with Bernoulli schedules, Journal of Applied Probability, Vol. 23, pp. 790-802.

[125] Kim, C.S., Vilena, V., Alexander, M. and Dudin, N. (2012): Computation of the steady state distribution for multi-server retrial queues with phase type service process, Annals of Operations Research, Vol. 201, pp. 307–323.

[126] Krishna Kumar, B. and Arivudainambi, D. (2002): An M/G/1 retrial queue with Bernoulli schedule and general retrial times, Computers and Mathematics with Applications, Vol. 43, pp. 15-30.

[127] Krishna Kumar, B. and Madheshwari, S. P. (2003): M^X/G/1 retrial queue with multiple vacations and starting failures, OPSEARCH, Vol. 40, No. 1, pp. 115-137.

[128] Krishna Kumar, B., Pavai Madheswari, S. and Anantha Lakshmi, S. R. (2011): Queuing system with state-dependent controlled batch arrivals and server under maintenance, Top, Vol. 19, No. 2, pp. 351-379.

[129] Krishna Kumar, B., Pavai Madheswari, S. and Anantha Lakshmi, S. R. (2013): An M/G/1 Bernoulli feedback retrial queueing system with negative customers, Operational Research, Vol. 13, No. 2, pp. 187-210.

[130] Krishna Kumar, B., Vijaykumar, A. and Arivudainambi, D. (2002): An M/G/1 retrial queueing system with two-phase service and preemptive resume, Annals of Operations Research, Vol. 113, pp. 61-79.

[131] Krishna Kumar, B., Vijayalakshmi, G., Krishnamoorthy, A. and Sadiq Basha, S. (2010): A single server feedback retrial queue with collisions, Computers and Operations Research, Vol. 37, No.7, pp. 1247-1255.

[132] Krishnamoorthy, A., Gopakumar, B. and Narayanan, V. C. (2012): A retrial queue with server interruptions, resumption and restart of service, Operational Research, Vol. 12, No. 2, pp. 133-149.

[133] Krishnamoorthy, A., Pramod, P. K. and Chakravarthy, S. R. (2014): Queues with interruptions: A survey, Top, Vol. 22, No. 1, pp. 290-320.

[134] Lakshmi, K. and Ramanath, K. (2013): A matrix geometric approach to the M/M/1 two-phase multi optional retrial queue with Bernoulli feedback, impatient customers and a server subject to breakdown and repair, OPSEARCH, Vol. 51, No. 1, pp. 3649.

[135] Lakshmi, K. and Ramanath, K. (2014): An M/G/1 two phase multi-optional retrial queue with Bernoulli feedback, non-persistent customers and breakdown and repair, International Journal of Operational Research, Vol. 19, No. 1, pp. 78-95.

[136] Langaris, C. and Dimitriou, I. (2010): A queueing system with *n*-phases of service and (n-1) types of retrial customers, European Journal of Operational Research, Vol. 205, No. 3, pp. 638-649.

[137] Langaris, C. and Moutzoukis, E. (1995): A retrial queue with structured batch arrivals, priorities and server vacations, Queueing Systems, Vol. 20, No. 3-4, pp. 341-368.

[138] Laxmi, P. V. and Gupta, U. C. (1999): On the finite-buffer bulk-service queue with general independent arrivals: $GI/M^{[b]}/1/N$, Operations Research Letters, Vol. 25, No. 5, pp. 241-245.

[139] Li, J. H. and Tian N. S. (2007): The discrete-time GI/Geo/1 queue with working vacations and vacation interruption, Applied Mathematics and Computation, Vol. 185, No. 1, pp. 1-10.

[140] Li, J. H., Tian, N. S., Zhang, Z. G. and Luh, H. P. (2009): Analysis of the M/G/1 queue with exponentially working vacations - A matrix analytic approach, Queueing Systems, Vol. 61, No. 2-3, pp.139-166.

[141] Li, H. and Zhao, Y. Q. (2005): A retrial queue with a constant retrial rate, server downs and impatient customers, Stochastic Models, Vol. 21, No. 2-3, pp. 531-550.

[142] Lin, C. H. and Ke, J. C. (2009): Multi-server system with single working vacation, Applied Mathematical Modelling, Vol. 33, No. 7, pp. 2967-2977.

[143] Lin, Z. C. and Liu, C. Y. (2003): Analysis and application of the adaptive neuro-fuzzy inference system in prediction of CMP machining parameters, International Journal of Computer Applications in Technology, Vol. 17, No. 2, pp. 80-89.

[144] Liu, X. and Fapojuwo, A. O. (2006): Performance analysis of hierarchical cellular networks with queueing and user retrials, International Journal of Communication Systems, Vol. 19, No. 6, pp. 699-721.

[145] Liu, Z., Wu, J. and Yang, G. (2009): An M/G/1 retrial G-queue with preemptive resume and feedback under N-policy subject to the server breakdowns and repairs, Computers and Mathematics with Applications, Vol. 58, No. 9, pp. 1792–1807.

[146] Luh, H. (2010): Matrix geometric analysis of departure processes of queues with applications to a pull serial line, International Journal of Operations Research, Vol. 7, No. 2, pp. 1-18.

[147] Madan, K. C. (2000): An M/G/1 queue with second optional service, Queueing Systems, Vol. 34, No. 1-4, pp. 37-46.

[148] Maraghi A. F., Madan, K. C. and Dowman, D. K. (2010): Bernoulli schedule vacation queue with batch arrivals and random system breakdowns having general repair time distribution, International Journal of Operational Research, Vol. 7, No. 2, pp. 240-256.

[149] Marsan, M. A., Carolis, G. D., Leonardi, Cigno, R. L. and Meo, M. (2001): Efficient estimation of call blocking probabilities in cellular mobile telephony networks with customer retrials, IEEE Journal on Selected Areas in Communications, Vol. 19, No. 2, pp. 332-346.

[150] Medhi, J. (1991): Stochastic Models in Queueing Theory, Academic Press, San Diego.

[151] Medhi, J. (2002): A single server Poisson input queue with a second optional channel, Queueing Systems, Vol. 42, No. 3, pp. 239-242.

[152] Miller, L. W. (1964): Alternating Priorities in Multi-Class Queue, Ph.D. Dissertation, Cornell University, Ithaca, New York.

[153] Mokaddis, G. S., Metwally, S. A. and Zaki, B. M. (2007): Optimal control of $M^X/H_2/1$ queue with a removable and non-reliable server, Information and Management Sciences, Vol. 18, No. 4, pp. 387-402.

[154] Neuts, M. F. (1978): Markov chains with applications in queueing theory, which have a matrix geometric invariant probability vector, Advances in Applied Probability, Vol. 10, pp. 185-212.

[155] Neuts, M. F. (1981): Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach, The Johns Hopkins University Press, Baltimore and London.

[156] Nobel, R. D. (2013): A queueing model in discrete time with retrials and a tolerant server, Proceedings of the 8th International Conference on Queueing Theory and Network Applications (QTNA2013), Taichung, pp. 181-186, June 30-July 1, 2013.

[157] Nobel, R. D. and Tijms, H. C. (1999): Optimal control for an $M^X/G/1$ queue with two service modes, European Journal of Operational Research, Vol. 113, No. 3, pp. 610-619.

[158] Nobel, R. D. and Tijms, H. C. (2000): Optimal control of a queueing system with heterogeneous servers and setup costs, IEEE Transactions on Automatic Control, Vol. 45, No. 4, pp. 780-784.

[159] Pakes, A.G. (1969): Some conditions for ergodicity and recurrence of Markov chains, Operations Research, Vol. 17, No. 6, pp. 1058-1061.

[160] Peng, Y., Liu, Z. and Wu, J. (2014): An M/G/1 retrial G-queue with preemptive resume priority and collisions subject to the server breakdowns and delayed repairs, Journal of Applied Mathematics and Computation, Vol. 44, No. 1-2, pp. 187–213.

[161] Phung-Duc, T. and Kawanishi, K. (2011): Multiserver retrial queues with after-call work, Numerical Algebra, Control and Optimization, Vol. 1, No. 4, pp. 639-656.

[162] Phung-Duc, T. and Kawanishi, K. (2014): Performance analysis of call centers with abandonment, retrial and after-call work, Performance Evaluation doi: 10.1016/j.peva.2014.03.001 (In press).

[163] Phung-Duc, T., Masuyama, H., Kasahara, S. and Takahashi, Y. (2009): Performance analysis of optical burst switched networks with limited-range wavelength conversion, retransmission and burst segmentation, Journal of the Operations Research Society of Japan, Vol. 52, No. 1, pp. 58-74.

[164] Ponomarov, V. and Lebedev, E. (2013): finite source retrial queues with state-dependent service rate, Modern Probabilistic Methods for Analysis of Telecommunication Networks Communications in Computer and Information Science, Vol. 356, pp. 140-146.

[165] Ramalhoto, M. F. and Gomez-Corral, A. (1998): Some decomposition formulae for the M/M/r/r+d queue with constant retrial rate, Communications in Statistics-Stochastic Models, Vol. 14, pp. 123-146.

[166] Roszik, J., Sztrik, J. and Kim, C. S. (2005): Retrial queues in the performance modeling of cellular mobile networks using MOSEL, International Journal of Simulation: Systems, Science and Technology, Vol. 6, pp. 38-47.

[167] Sharma, S. and Karmeshu (2009): Power law characteristics and loss probability: Finite buffer queueing systems, IEEE Communications Letters, Vol. 13, No. 12, pp. 971-973.

[168] Selvaraju, N. and Goswami, C. (2013): Impatient customers in an M/M/1 queue with single and multiple working vacations, Computers and Industrial Engineering, Vol. 65, pp. 207–215.

[169] Sennott, L. I., Humblet, P. A. and Tweedie, R. L. (1983): Mean drifts and the non-ergodicity of Markov chain, Operations Research, Vol. 15, No. 4, pp. 783-789.

[170] Senthil Kumar, M. and Arumuganathan, R. (2010): An M^X/G/1 retrial queue with two-phase service subject to active server breakdowns and two types of repair, International Journal of Operational Research, Vol. 8, No. 3, pp. 261 - 291.

[171] Shannon, C. E. (1948): A mathematical theory of communication, Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656.

[172] Sherman, N. P. and Kharoufeh, J. P. (2006): An M/M/1 retrial queue with unreliable server, Operations Research Letters, Vol. 34, No. 6, pp. 697-705.

[173] Shin, Y. W. (2007): Multi-server retrial queue with negative customers and disasters, Queueing Systems, Vol. 55, No. 4, pp. 223–237.

[174] Shin, Y. W. and Moon, D. H. (2013): Approximation of M/M/s/K retrial queue with non persistent customers, Applied Mathematical Modeling, Vol. 37, No. 3, pp.753-761.

[175] Sikdar, K. and Gupta U. C. (2005): Analytic and numerical aspects of batch service queues with single vacation, Computers and Operations Research, Vol. 32, No. 4, pp. 943-966.

[176] Sikdar, K., Gupta, U. C. and Banik, A. D. (2009): Performance analysis of a finite-buffer batch service queue with general input and exponential multiple vacations, Journal of Probability and Statistical Science, Vol. 7, No. 2, pp. 191-210.

[177] Takagi, H. (1986): Analysis of Polling Systems, MIT Press, Cambridge, M. A.

[178] Takagi, H. (1991): Queueing Analysis: A Foundation of Performance Evaluation, Vacation and Priority Systems, Part I, New York: North-Holland.

[179] Takagi, H. (1993): Questing Analysis: A Foundation of Performance Evaluation, Vol. 3, Discrete-Time Systems, North-Holland, Amsterdam.

[180] Tao, L., Liu, Z. and Wang Z. (2014): M/M/1 retrial queue with collisions and working vacation interruption under N-policy, RAIRO - Operations Research, Vol. 46, No. 3, pp. 55-371.

[181] Tarabia, A. M. K. (2007a): Analysis of M/M/1 queueing system with two priority classes, OPSEARCH, Vol. 44, No. 4, pp. 346-365.

[182] Tarabia, A. M. K. (2007b): Two-class priority queueing system with restricted number of priority customers, AEU-International Journal of Electronics and Communications, Vol. 61, No. 8, pp. 534-539.

[183] Tarabia, A. M. K. (2011): Transient and steady state analysis of an M/M/1 queue with balking, catastrophes, server failures and repairs, Journal of Industrial and Management Optimization, Vol. 7, No. 4, pp. 811-823.

[184] Tettamanzi, A. and Tomassini, M. (2001): Soft Computing, Springer-Verlag.

[185] Tian, M., Ma, Z. and Liu, M. (2008): The discrete time Geom/Geom/1 queue with multiple working vacations, Applied Mathematical Modelling, Vol. 32, No. 12, pp. 2941-2953.

[186] Tran-Gia, P. and Mandjes, M. (1997): Modeling of customer retrial phenomenon in cellular mobile networks, IEEE journal on selected areas in communications, Vol. 15, No. 8, pp. 1406-1414.

[187] Trivedi, K.S. (2001): Probability and Statistics with Reliability, Queuing and Computer Science Application, 2nd ed., John Wiley and Sons.

[188] Trivedi, K.S., Ma, X. and Dharmaraja, S. (2003): Performability model on wireless communication systems, International Journal of Communication Systems, Vol. 16, pp. 561-577.

[189] Vadivu, A. S., Vinayak, R., Dharmaraja, S. and Arumuganathan, R. (2014): Performance analysis of voice over internet protocol via non Markovian loss system with preemptive priority and server break down, OPSEARCH, Vol. 5, No. 1, pp. 50–75.

[190] Vinayak, R., Dharmaraja, S. and Arunachalam, V. (2014): On the study of simultaneous service by random number of servers with retrial and preemptive priority, International Journal of Operational Research, Vol. 20, No. 1, pp. 68-90.

[191] Wang, J. (2004): An M/G/1 queue with second optional service and server breakdowns, Computers and Mathematics with Applications, Vol. 47, No. 10-11, pp. 1713-1723.

[192] Wang, J. (2008): On the single server retrial queue with priority subscribers and server breakdowns, Journal of Systems Science and Complexity, Vol. 21, No. 2, pp. 304-315.

[193] Wang, J., Cao, J. and Li, J. (2001): Reliability analysis of the retrial queue with server breakdowns and repairs, Queueing Systems, Vol. 38, No. 4, pp. 363-380.

[194] Wang, K. H., Chan, M. C. and Ke, J. C. (2007b): Maximum entropy analysis of $M^{[x]}/M/1$ queueing system with multiple vacations and server breakdowns, Computers and Industrial Engineering, Vol. 52, No. 2, pp. 192-202.

[195] Wang, K. H., Chuang, S. L. and Pearn, W. L. (2002): Maximum entropy analysis to the N-policy M/G/1 queueing system with a removable server, Applied Mathematical Modelling, Vol. 26, No. 12, pp. 1151-1162.

[196] Wang, K. H. and Huang, K. B. (2009): A maximum entropy approach for the (p, N)-policy M/G/1 queue with removable and unreliable server, Applied Mathematical Modelling , Vol. 33, No. 4, pp. 2024-2034.

[197] Wang, K. H., Kuo, C. C. and Pearn, W. L. (2007a): Optimal control of an M/G/1/K queueing system with combined F policy and startup time, Journal of Optimization Theory and Applications, Vol. 135, No. 2, pp. 285-299.

[198] Wang, K. H., Kuo, C. C. and Pearn, W. L. (2008a): A recursive method for the F-policy G/M/1/K queueing system with an exponential startup time, Applied Mathematical Modelling, Vol. 32, No. 6, pp. 958-970.

[199] Wang, J. and Li, J. (2009): A single server retrial queue with general retrial times and two phase service, Journal of Systems Science and Complexity, Vol. 22, No. 2, pp. 291-302.

[200] Wang, J., Liu, B. and Li, J. (2008b): Transient analysis of an M/G/1 retrial queue subject to disasters and server failures, European Journal of Operational Research, Vol.189, No. 3, pp. 1118–1132.

[201] Wang, C. H. and Luh, H. (2011): Analysis of bandwidth allocation on end-to-end qos networks under budget control, Computers and Mathematics with Applications, Vol. 62, No. 1, pp. 419-439.

[202] Wang, K. H. and Yang, D. Y. (2009): Controlling arrivals for a queueing system with an unreliable server: Newton-Quasi method, Applied Mathematics and Computation, Vol. 213, No. 1, pp. 92-101.

[203] Wang, K. H., Yang, D. Y. and Pearn, W. L. (2011): Comparative analysis of a randomized N-policy queue: An improved maximum entropy method, Expert Systems with Applications, Vol. 38, No. 8, pp. 9461-9471.

[204] Wang, J. and Zhang, F. (2013): Strategic joining in M/M/1 retrial queues, European Journal of Operational Research, Vol. 230, No. 1, pp. 76-87.

[205] Wenhui, Z. (2005): Analysis of a single-server retrial queue with FCFS orbit and Bernoulli vacation, Applied Mathematics and Computation, Vol. 161, pp. 353-364.

[206] Wu, J. and Lian, Z. (2013a): A single-server retrial G-queue with priority and unreliable server under Bernoulli vacation schedule, Computers and Industrial Engineering, Vol. 64, No. 1, pp. 84-93.

[207] Wu, J. and Lian, Z. (2013b): Analysis of the M_1, M_2/G/1 G-queueing system with retrial customers, Nonlinear Analysis: Real World Applications, Vol. 14, No. 1, pp. 365-382.

[208] Wu, J. and Yin, X. (2011): An M/G/1 retrial G-queue with non-exhaustive random vacations and an unreliable server, Computers and Mathematics with Applications, Vol. 62, No. 5, pp. 2314–2329.

[209] Xu, K., Hassanein, H. S., Takahara, G. and Wang, Q. (2010): Relay node deployment strategies in heterogeneous wireless sensor networks, IEEE Transactions on Mobile Computing, Vol. 9, No. 2, pp. 145- 159.

[210] Xu, X., Zhang, Z. and Tian, N. (2009): The M/M/1 queue with single working vacation and set-up times, International Journal of Operational Research, Vol. 6, No. 3, pp. 420 - 434.

[211] Yang, D. Y., Chiang, Y. C. and Tsou, C. S. (2013): Cost analysis of a finite capacity queue with server breakdowns and threshold-based recovery policy, Journal of Manufacturing Systems, Vol. 32, No. 1, pp. 174–179.

[212] Yang, T. and Templeton, J. G. C. (1987): A survey on retrial queues, Queueing Systems, Vol. 2, No. 3, pp. 201–233.

[213] Zhang, Z. G. and Tian, N. (2003): Analysis on queueing systems with synchronous of partial servers, Performance Evaluation, Vol. 52, No. 4, pp. 269-282.

[214] Zhang, F. and Wang, J. (2012): Stochastic analysis of a finite source retrial queue with spares and orbit search, Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance Lecture Notes in Computer Science, Vol. 7201, pp. 16-30.

[215] Zhang, F. and Wang, J. (2013): Performance analysis of the retrial queues with finite number of sources and service interruptions, Journal of the Korean Statistical Society, Vol. 42, No. 1, pp. 117-131.

[216] Zhang, F., Wang, J. and Liu, B. (2013): Equilibrium balking strategies in Markovian queues with working vacations, Applied Mathematical Modelling, Vol. 37, No. 16-17, pp. 8264–8282.

[217] Zhong, L., Alajaji, F. and Takahara, G. (2007): A binary communication channel with memory based on a finite queue, IEEE Transactions on Information Theory, Vol. 53, No. 8, pp. 2815-2840.

[218] Zhong, L., Alajaji, F. and Takahara, G. (2008): A model for correlated rician fading channels based on a finite queue, IEEE Transactions on Vehicular Technology, Vol. 57, No. 1, pp. 79-89.