STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

DECLARATION

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma of any other university or institution and, to the best of my knowledge and belief, it contains no material previously published by any other person, except where due reference made in the text of the thesis.

NARAYANAN, V

Calicut University Campus
12th October 2006

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

Certificate

This is to certify the work reported in this thesis entitled **STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS** that is being submitted by Sri Narayanan, V for the award of Doctor of Philosophy, to the University of Calicut, is based on the bonafide research work carried out by him under my supervision and guidance in the Department of Statistics, University of Calicut. The results embodied in this thesis have not been included in any other thesis submitted previously for the award of any degree or diploma of any other University or institution.

Dr M. Manoharan
Reader and Head
Department of Statistics
University of Calicut
Kerala – 673 635
India

Calicut University Campus
12th October 2006

HEAD
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

ACKNOWLEDGEMENT

I place on record my indebtedness to Dr. M. Manoharan, Reader and Head, Department of Statistics, University of Calicut, for the inspiring guidance, constructive suggestions and constant encouragement during the entire period of study. I am thankful for his intellectual contributions and methodological guidance, as well as his understanding of my various interests.

I would like to express my deep sense of gratitude to the authorities of National Institute of Technology Calicut, for permitting me to do part time research under the guidance of Dr. M. Manoharan, University of Calicut.

I would like to express my deep sense of gratitude to Dr. M.J. Jacob, Professor and Head, Department of Mathematics, National Institute of Technology Calicut, for his encouragement and valuable discussions in writing the thesis.

I am thankful to Dr. M.J. Jacob, Dr. M. Chandramohan, Dr. O.T. George, Dr. Sunil Jacob John and other faculty members of the Department of Mathematics, National Institute of Technology Calicut, for their help and all sorts of co-operation.

I am thankful to Dr. N. Raju, Dr. C. Chandran, Dr. K. Jayakumar for their help and encouragement.

Thanks are due to research scholars, librarian, students and office staff of Department of Statistics, University of Calicut.

I am thankful to Mr. Beljith for his excellent typesetting.

Finally my thanks goes to wife Sobha and daughters Sujina and Nijisha whom offered encouragement, understanding and patience during this research period.

NARAYANAN, V

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

CONTENTS

CHAPTER 1 INTRODUCTION **1**

1.1 Introduction 1

1.2 Queueing Systems 5

1.3 Problems in Queueing System 14

1.4 Queueing Parameters 17

1.5 Some Preliminaries 18

1.6 Waiting Time Distribution 24

1.7 Review of State-dependent Markovian Queueing Models 29

1.8 Main Contributions of the Thesis 34

CHAPTER 2 GENERAL STATE-DEPENDENT MARKOVIAN QUEUEING MODELS **40**

2.1 Introduction 40

2.2 Model I: $\lambda_n = \lambda,\ n \geq 0;\ \mu_n = n\mu,\ n \geq 1\ (\lambda > 0,\ \mu > 0)$ 41

2.3 Model II: $\lambda_n = \frac{\lambda}{n+1},\ n \geq 0;\ \mu_n = \mu,\ n \geq 1\ (\lambda > 0,\ \mu > 0)$ 47

2.4 Model III: $\lambda_n = \frac{\lambda}{n+1},\ n \geq 0;\ \mu_n = n\mu,\ n \geq 1 (\lambda > 0,\ \mu > 0)$ 53

2.5 Model IV: $\lambda_n = (n+1)\lambda,\ n \geq 0;\ \mu_n = n\mu,\ n \geq 1 (\lambda > 0,\ \mu > 0)$ 57

2.6 Transient/Steady-state Analysis of State-dependent Markovian Queueing Models 67

2.7 Summary of Results and Concluding Remarks 70

CHAPTER 3 PERFORMANCE MEASURES OF STATE DEPENDENT MARKOVIAN QUEUEING MODELS **72**

3.1 Introduction 72

3.2 Linear Model characterized by $\lambda_n = \lambda,\ n \geq 0,\ \mu_n = 2n\mu,\ n \geq 1$ $(\lambda = 0, \mu > 0)$ 73

3.3 State-dependent Model characterized by $\lambda_n = \frac{\lambda}{2(n+1)},\ n \geq 0,$ $\mu_n = \mu,\ n \geq 1,\ (\lambda = 0, \mu > 0)$ 78

3.4 State-dependent Model characterized by $\lambda_n = \frac{\lambda}{2(n+1)}, n \geq 0,$ 81

$\mu_n = 2n\mu, n \geq 1, \ (\lambda = 0, \mu > 0)$

3.5 Summary of Results and Concluding Remarks 84

CHAPTER 4 MODIFIED STATE-DEPENDENT MARKOVIAN QUEUEING MODELS 86

4.1 Introduction 86
4.2 Model AI 86
4.3 Model AII 93
4.4 Model AIII 98
4.5 Model AIV 102
4.6 Model AV 109
4.7 Model AVI 113
4.8 Model AVII 116
4.9 Summary of Results and Concluding Remarks 119

CHAPTER 5 NON-LINEAR STATE-DEPENDENT MARKOVIAN QUEUEING MODELS 122

5.1 Introduction 122
5.2 Model BI 122
5.3 Model BII 129
5.4 Model BIII 132
5.5 Model BIV 137
5.6 Model BV 142
5.7 Model BVI 144
5.8 Summary of Results and Concluding Remarks 147

CHAPTER 6 THE WAITING TIME DISTRIBUTION OF MARKOVIAN QUEUEING MODELS WITH STATE-DEPENDENT ARRIVAL RATES – A COMPARISON 150

6.1 Introduction 150

6.2 Model I: $\lambda_n = \frac{\lambda}{2(2n+1)}, n \geq 0; \quad \mu_n = \mu, n \geq 1$ 152

$(\lambda > 0, \mu > 0)$

6.3 Model II: $\lambda_n = \frac{\lambda}{(n+1)^2}, n \geq 0; \mu_n = \mu, (\lambda > 0, \mu > 0)$ 155

6.4 Comparison of expected waiting time and coefficient of variation for different models 160
6.5 Summary of Results and Concluding Remarks 161

CHAPTER 7 COMPARISON BETWEEN STATE-DEPENDENT QUEUEING MODELS 162

7.1 Introduction 162
7.2 Numerical comparison of measures of service capacity for the different state-dependent queueing models 169
7.3 Numerical comparison of $\lambda\hat{W}, \lambda\hat{B}$, and $\lambda\hat{S}$ for different state-dependent queueing models 171
7.4 Comparison of system size 175
7.5 Comparison between linear and non-linear state-dependent queueing models 177
7.6 Summary of Results and Concluding Remarks. 184

REFERENCES 186

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

Chapter 1

INTRODUCTION

1.1 INTRODUCTION

The interplay of supply and demand has always been a central feature of many realm of human activity. We have a range of obvious and well known examples from Economics. But in this thesis we are rather be concerned with systems typified by units which demand service and facilities which attempt to provide service. We shall not consider the cognate questions of the price to be paid for service, nor the penalty for not providing it. It will be a fundamental principle of our investigation that as much service as can reasonably be given shall be conceded to demand, while on the other hand, demand will take account of the capacity of the service facility and regulate its requirements accordingly. That is, we look at ways of optimizing the performance of demand-service facility by the use of adaptive mechanisms.

We shall be concerned with the family of single server queueing models. When a customer arrives at the service point and demand service, it is generally regarded as in the best interest of the management to provide that service. Without overworking the system, the management would like to maximize the facilities offered. The environment under which the system operates does not remains the same and so we cannot expect the inherent parameters of the system as constant. The changes often develop as a result of extraneous uncontrollable factors. To the conscious management, the question always being confronted with is: "Does there exist a better system whose application would bring better results in the same situation?" Even though the system currently in use gives some what satisfactory results, there may be in some sense a more efficient system whose implementation would be more rewarding or useful.

The present study is concerned with some state-dependent Markovian queueing models. The underlying process in each model is examined at the outset to give more information on its performance measures. The central model is the generalization of M/M/N system in which the arrival rate λ_n and the service rate μ_n are both functions of n, the number of customers already present in the system. In many real life situations λ_n and μ_n change whenever n changes, so that both demand (arrival) and service (departure) correspond to the system state. We are interested in queueing models restricted to those favouring statistical equilibrium.

There is a large literature concerned with different methods for the reduction of congestion at high traffic intensity. For example, consider customers being lost when the system reaches a certain level or customers leaving the system before being served due to long waiting time (balking). The method described by Harris (1967) is to alter the service interval when the system reaches a certain preassigned level. Also one may think of switching to a completely different system when the traffic intensity reaches a certain pre determined level. These are only examples of state dependent queueing system which is our main concern in the thesis. The simplest generalization which gives a state-dependent queuing system consists in supposing the parameters λ and μ of M/M/1 queueing system to be state dependent with λ_n and μ_n as the arrival rate and service rate respectively, when there are n customers/units in the system. The model described by Conolly and Hadidi (1968) has the feature of a direct relationship between arrivals and service intervals but not state-dependence. However, it is an adaptive system. The so called 'correlated' queue described there has the speciality that the arrival pattern influences the service behaviour such that the service time of the n^{th} customer directly depends on his inter arrival interval.

State-dependent queueing models has been discussed far and widely in the literature since the work of Cox and Smith (1961). In the more general context of birth and death processes they have been reviewed and analyzed, for example, by Jackson (1963), Keilson (1964), Harris (1967), Hadidi and Conolly, B.W (1969), Natvig (1974), Conolly (1975), Vandoorn (1981) and many other researchers. Earlier work by Hadidi and Conolly point to the operational advantage of state-dependent queueing models. Models of this kind are of special interest in Operations Research.

Our study of state-dependent Markovian queueing models is motivated by queueing scenarios where the arrival rate and/or service rate depends on the amount of work present, like production system and internet. Queueing systems where the service speed is work load-dependent are well-known, specifically in the studies of dams and storage processes, for example see Harrison and Resnick (1976), Brockwell et. al (1982) and Kaspi et. al (1996). Also in production systems the speed of the server often depends on the amount of work. This is particularly true if the server is not represented by a machine, but rather by a human being, where the speed of the server is relatively high or low when there is much work or there is little work. In addition to the general service speeds, the rate at which jobs arrive at the system may also depend on the amount of work present. In the human server production system, we may try to control the arrival rate of jobs to optimize server performance.

A queueing system can be described as customers arriving for service, waiting for service if it is not immediate, and if having waited for service, leaving the system after being served. One would need to know answers to such questions as "How long must a customer wait"? and "How many people will form the line"?. Queueing theory attempts to answer these questions through detailed mathematical analysis. Queueing system would require a detailed characterization of the underlying processes. We feel it unnecessary to conduct a general survey and appreciation of the realm of queueing theory, but we shall first identify that aspect of queueing theory to which this contribution belongs. It will be a fundamental principle of our investigation to consider ways of optimizing or regulating the performance of a demand-service facility by the use of adaptive mechanisms, more specifically with state dependence.

A major difficulty in the practical application of queueing system is congestion. Congestion occurs when demand is at the limit of what service can provide. The consequences is that new demands for service are increasingly likely to be obliged to wait or sometimes be denied as times goes by. Hence it is clearly of importance to have a measure of the load imposed by demand on a service system. A frequently employed measure is the mean number of arrivals during an average service time, often called congestion index or traffic intensity.

1.2 QUEUEING SYSTEMS

Generally a queueing system is any facility at which customers/units arrive and then stay for a certain duration before departing/servicing; the unit demanding service is called the 'customer' and the device at which or the person by whom it get served is known as the 'server'. This terminology is used in a wide context. Here are a few realistic examples of this customer – server mechanism.

(i) Vehicles demanding service arrive in a garage and depending on the number of employees, one or more vehicles may be repaired at a time.

(ii) Patients arrive at a doctor's clinic for treatment. Even if some appointment system exists, due to the emergency service rendered, there is a random element present in the arrival scheme and, there is a possibility of a waiting line building up.

(iii) Passengers demanding tickets queue up in front of a ticket counter

(iv) In a telephone exchange, the incoming calls are the customers who demands service in the form of telephone conversations

It is possible to give numerous examples of this type, where a queue situation exists in one form or the other. As suggested by the above problems, some of these situations differ from each other in several details. However, it is not difficult to see that all these situations have certain common basic characteristics, which we call the basic elements of a queueing system.

Basic Elements of a Queueing System

The arrival (input) process, service mechanism, queue discipline and cost structure are the four basic elements of a queueing system.

1. *The input process*:

If the arrivals and service are strictly according to schedule a queue can be avoided. But they are not, especially the arrivals. In most situations arrivals are controlled by external factors and these factors contribute to the uncertain nature of arrivals. For instance, the arrival could be in groups of random or constant size or in the simplest case, one at a time. The time intervals between successive arrivals can be considered as random variables, having certain distributions. Further, the arrivals could be emanating from a finite or an infinite source. For instance, if the customers are the machines needing repair in an industrial concern the number of machines is a finite number; whereas customers at an ordinary store window could be considered as coming from an infinite source. Therefore, the source of arrivals, the type of arrivals and the inter arrival times should be specified in a complete specification of an input process.

2. *Service mechanism*:

The number of service is an integral part of the service mechanism. So also are the duration of service and whether service is given to groups of customers or to one at a time. Because of the uncertainty involved in the length of service – a telephone conversation is a good example – we can consider the service time as a random variable having a certain distribution.

3. *Queue discipline*:

We can consider all other factors regarding the rules of the conduct of the queue under this heading, the simplest of these is known as the first-come, first served (FCFS) discipline. This specifies that customers arriving when the server is busy will be taken for service in the order of their arrival. Other disciplines which are also common in use include (i) Last-come, first served (LCFS) – according to it the last arrival in the system is served first, (ii) Service-in Random order (SIRO), where the arrivals are served randomly irrespective of their arrivals in the system. It is more likely for the customer to experience both short and long waiting times under SIRO than under FCFS, a result which can be backed by intuition. Clearly when customers arrive in groups, it is assumed that those in a group are ordered for the sake of service. In addition to these one may introduce some sort of variations in customers getting impatient after waiting, may leave the system without getting served. Some may jocket for positions in the waiting line or some customers may be considered as having higher priorities in service than others. The system may not allow more than a certain number to be waiting at a certain time; that is, the size of the waiting room could be finite, and those arriving when the waiting room is full are allowed to be lost to the system. These are only a few examples of various types that can be derived from the simple system described. It is essential, therefore, for the complete description of a queueing system, the rule to be followed has to be specified.

4. *The cost structure*

The cost structure of a queueing system specifies the payment made by the customers for the service they receive (the reward), and the various operating costs of the system. These costs are expressed in terms of the amount of service provided by the system, measured, for example, by the number of servers. The system also incurs a holding cost, expressed as a function of the number of customers waiting or the delay suffered by them.

Based on these descriptions, the problems arising in Queueing theory can be classified into three.

(i) *Behavioral problems of the system*:

The study of behavioral problems aims at understanding a particular situation as thoroughly as possible. This is done by using mathematical models. Naturally, these are idealized models to varied degrees of realisms. As done in many other branches of Science, these models are studied analytically in isolation, hoping that the information obtained form such a study would be useful in the decision making process regarding such situations. Some of the characteristics considered in this connection are the distributions, expected values and other moments connected with the queue length, waiting time and the length of the busy period in a queueing system.

(ii) *Statistical problems of the system*:

By statistical problems we mean the problem of the study of empirical data, estimating and tests of hypotheses regarding queue situations. For an insight into the correct mathematical model, which could be studied analytically to derive its properties, a statistical study is essential. Otherwise, the analytical study would be divorced from the practical situation, thus rendering it less useful for the applied researcher.

(iii) *Optimization problems of the system*:

Looking at the variety of queueing systems possible, it is only natural to assume that some models are more appropriate in certain contexts than the rest. Further, as queue situations can be controlled according to specifications, the knowledge of the right model for the right situation becomes essential. The suitability of a model is decided by a comparison among several models of the return in benefit to the individual concerned. Naturally, some cost factors have to be considered. Because of the external factors in solved, the internal characteristics have to be changed according to these. Thus some decision rules should be spelled out.

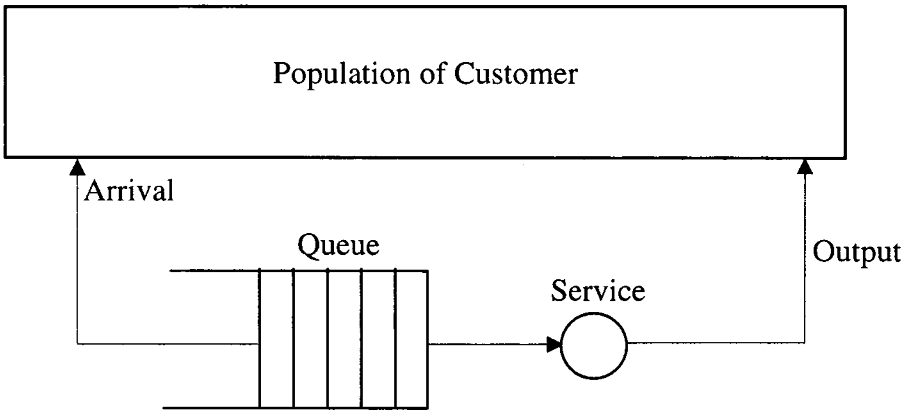
Our aims in studying a queueing system are:

(i) To understand the queueing phenomenon and

(ii) To control it- that is, exercise some influence on this phenomenon with a view to operating the system most economically.

Queue sizes are important considerations in the design of fixed facilities and in management decisions. Usually some compromise between queue size, waiting time and the costs of service have to be made. The success or failure to a queueing system cannot be judged as easily as the failure resistance of a beam. Whether or not the queueing system can be considered a failure can only be judged on a competitive basis.

The following figure shows the elements of a single server queueing system.



Most quantitative parameters (like average queue length, average time spent in the system) do not depend on the queue discipline. This is why most models either do not take the queue discipline into account at all or assume the normal FCFS queue. In fact the only parameter that depends on the queue discipline is the variance (or standard deviation) of the waiting time. In the steady-state, a change of queue discipline does not affect the corresponding output process, unless it interferes with the servicing system. The study on the effect of queue discipline on output process seems to have been started with the work of Muntz (1973). It has long been known, see for example Mirasol (1963) or

Kendall (1964), that output process from an $M/G/\infty$ queue is a Poisson process (insensitivity property). Shanbhag and Tambouratzis (1973) have established that for an M/G/s loss model, the steady-state output process is Poisson with the same parameter as the input process. Kelly (1976) has studied an M/G/1 system with the pre-emption last come first served (LCFS) queue discipline. Kelly proved that the departures in the system in equilibrium is also Poisson and that distribution of the number of customers in the system is Geometric. For a concise survey of departures and related characteristics of queueing models, one may refer to Manoharan et. al. (2003).

The two extreme values of the waiting time variance are for the FCFS queue (minimum) and the LCFS queue (maximum). The study of "waiting in line" is a fundamental aspect of Engineering design and analysis across a range of application domain. Modern telecommunication networks and packet-switched networks (internet), as well as computer architecture and distributed system design are all based on queueing systems. Hence, they benefit from the knowledge gained through the theory of queues. Some of the questions that engineers seek to answer about a queueing system include:

(i) What is the steady-state (long-term average) mean time that a customer is in the system - including both waiting and service time?

(ii) What is the steady-state mean delay or waiting time experienced by a customer in the system?

(iii) What is the steady-state mean number of customers in the system or in the queue?

(iv) What is the steady state utilization (percentage of time spent servicing customers) of the system? (Note: this is unity minus the probability the system is empty)

(v) What are the state probabilities for a given number of customers in the system?

(vi) What is the complete distribution of time that customers spent in the system (sojourn time)?

(vii) What is the distribution of time before the system reaches a certain state, e.g. the time until the delay exceeds some threshold (first passage time)?

The Kendall – Lee notation for Queueing Systems

Kendall introduced a A/B/C queueing notation in 1953. It has since been extended to 1/2/3/(4/5/6) where the numbers are replaced with

1. A code describing the arrival process. The codes used are:

   M stands for "Markovian" implying exponential distribution for inter arrival times or service times.

   D stands for "deterministic" service times.

   E_K stands for an Erlang distribution with K as shape parameter.

   G stands for a "general distribution"

2. A similar code, as for arrived process, representing the service process. The same symbols are used.

3. The number of service channels

4. The priority order that jobs in the line are served

   (i) First Come Fist Served (FCFS)

(ii) Last Come First Served (LCFS)

(iii) Service in Random Order (SIRO) and

(iv) Processor Sharing (PS).

5. The maximum size of the system. The maximum number of customers allowed in the system including those in service when the number at this maximum, further arrivals are turned away.

6. The size of calling source. The size of the population from which the customers come. This limit is the arrival rate as more jobs queued up there are fewer available to arrive into the system.

The literature on queueing theory and the diverse areas of its applications has grown tremendously (exponentially, as claimed by some writers) over the years. The following list of books will provide a foundation on the topic of study.

1. Thomas L. Saaty (1961) - Elements of Queueing theory; with Applications. Mc Graw-Hill Book Company; New York
2. Cohen, J.W (1969) - The single server queue - North-Holland, Amsterdam
3. Leonard Kleinrock (1975) - Queueing systems; Vol I: Theory. Wiley New York
4. Brian Conolly (1975), Lecture Notes on Queueing System. Ellis Horwood Publisher Chichester.
5. Cooper, R.B (1981). Introduction to Queueing theory, 2nd ed. New York: North Holland.
6. B.D Bunday (1986) - Basic queueing theory; Edward Arnold, Australia.

7. Baccelli, F and Bremand, P (1994) - Elements of Queueing theory. Springer, New York.
8. N.U. Prabhu (1997) - Foundations of queueing theory; Kluwer Academic Publications, New York.
9. Donald Gross & Carl M. Harris (1998) - Fundamentals of queueing theory; Third edition. John Wiley & Sons (Asia) Pvt. Ltd; Singapore.
10. D.L Minh (2002) - Applied Probability models - Thomson Asia Pvt. Ltd. Singapore.
11. Asmussen, S (2003), Applied Probability and Queues, Second edn., Springer Verlag, New York.
12. Wayne L. Winston (2004) - Introduction to probability models IV edition. Thomson Asia Pvt. Ltd; Singapore.

1.3 PROBLEMS IN A QUEUEING SYSTEM

The ultimate objective of the analysis of queueing systems is to understand the behaviour of their underlying processes so that informed and intelligent decisions can be made in their management. Three types of problems can be identified in this process.

1. *Behavior problems*: The study of behavior problems of queueing systems is intended to understand how it behaves under various conditions. The bulk of results in queueing theory is based on research in behavioral problems. Mathematical models for the probability relationships among the various elements of the underlying process is used in the analysis. To make the ideas

concrete let us define a few terms which is defined formally later. In the context of queueing system the number of customers with time as a parameter is a stochastic process. Let $Q(t)$ be the number of customers in the system at time t. This number is the difference the number of arrivals and departures during (0, t). Let A(t) and D(t) respectively, be these numbers, then $Q(t) = A(t) - D(t)$. In order to manage the system efficiently one has to understand how the process $Q(t)$ behaves over time. Since $Q(t)$ is dependent on A(t) and D(t), both of which are also stochastic processes, their properties and dependence characteristics between the two should also be understood. In addition to the number of customers in the system, which we call the queue length, the time a new arrival has to wait till its service begins, which we call the waiting time, and the length of time the server is continuously busy, which we call busy period, are major characteristics of interest. It should be noted that the queue length and waiting time are stochastic processes and the busy period is a random variable. Distribution characteristics of the stochastic processes and random variables are needed to understand their behaviour. Since time is a factor, the analysis has to make a distinction between the time dependent, also known as transient, and the limiting, also known as the long term, behavior. Under certain conditions a stochastic process may settled down to what is commonly called a steady state, in which its distribution properties are independent of time.

2. *Statistical Problem*: Under statistical problems we include the analysis of empirical data in order to identify the correct mathematical model, and validation methods to determine whether the proposed model is appropriate.

Chronologically, the statistical study proceeds the behavioral study as could be seen from the early papers in this area.

3. *Operational Problem*: Under this heading we include all problems that are inherent in the operation of queueing systems. Some such problems are statistical in nature. Others are related to the design, control, and the measurement of effectiveness of the systems.

A queueing system can be studied under two different assumptions. One would aim at the short term behaviour of the system, in which the results turn out to be time dependent (transient behaviour). Instead, one may assume that the system has been in progress sufficiently long, so that it has settled into exhibiting a stable behaviour. This can happen only under certain conditions (such as traffic intensity less than one). The results obtained in this study are independent of time and hence are much more simple in form. Because of this simplicity, investigators in need of queueing results, have tended to use steady state results, in most of the situations. When the system has not settled into an equilibrium state, at best, this approach would give some approximations to the actual behaviour.

The importance of the study of queueing systems in finite time was noted as early as 1934 by Pollaczek. The difficulty in such a study is that the processes involved are not simple and more sophisticated mathematical procedures are necessary. For instance, the birth and death process equations are simple enough in the case of a simple queue with Poisson arrivals and exponential service times.

But for time dependent solution the use of transforms is necessary. This solution was given for the first time by Bailey (1956) and Lederman and Reuter (1956). While Bailey used the method of Generating functions for the differential equations, Lederman and Reuter used Spectral theory in its solution. Laplace transforms have also been used for the same problem, and it has been realized that generating functions and/or Laplace transforms form a useful technique in the solution of such differential - difference equations.

1.4 QUEUEING PARAMETERS

For an understanding, and efficient management, of queueing systems, the following three queue parameters are generally investigated.

1. The queue length:

The queue length is the number of customers/units waiting in the queue, or present in the system. In the latter case it is sometimes called the system size. The determination of the probability distribution of this discrete random variable is important for the design of the system. For example, for providing an extra server or a large waiting space in the case of large queues.

2. The waiting time:

The waiting time is defined to be the amount of time spent by the entering customer up to the beginning of service. Sometimes, the service time is also in the waiting time. Some authors, like Cox and Smith (1961), use the term queueing time and waiting time according as the service time is excluded or included. This is a continuous random variable and its probability distribution is important from the customer's point of view.

3. The busy period:

In a single server system suppose that the server is free initially and a customer arrives. He will be served immediately. During his service time some more customers will arrive and will be served in their turn. The process will continue in this way until no customer is left and the server becomes free again. When this happen, we say that a busy period has just ended. In contrast we have idle periods during which no customers are present in the system. A busy period and the idle period following it together constitute a busy cycle.

In a system with c servers ($2 \le c < \infty$) the natural definition of busy period is the time during which at least one of the servers is busy. However, we may also be interested in the period of time during which all c servers are busy, may call this the system busy period.

1.5 SOME PRELIMINARIES

Birth-Death Methodology

A large class of queueing systems are modeled by Markov processes of the so - called "Birth-and-Death" type. The defining characteristics of these processes is that the only non-zero instantaneous transition rates out of state n ($n = 0, 1, 2, 3, \ldots$) are to the neighbouring states (if any), $n + 1$ and $n - 1$. In the queueing frame work, the process, N_t, represents the number of customers present in some system at time t. Transitions from one state to another occur when customer arrives and when they depart. Then the above condition amounts to a requirement that customer arrive and depart singly (rather than in batches). Denote the arrival rate and departure rate when the process is in state n, by λ_n and μ_n respectively.

Denote the limiting probability (steady state probability) of state n by p_n where n is 0, 1, 2, ... These probabilities satisfy the following global balance equations (Total rate of transitions out of state n equals total rate of transitions into state n).

$$\lambda_0 \, p_0 = \mu_1 \, p_1$$

$$(\lambda_n + \mu_n) \, p_n = \lambda_{n-1} \; p_{n-1} + \mu_{n+1} \; p_{n+1}, n \geq 1$$

By solving these systems of differential difference equations recursively, we get

$$p_1 = \frac{\lambda_0}{\mu_1} p_0$$

$$p_2 = \frac{\lambda_0 \, \lambda_1}{\mu_1 \, \mu_2} p_0$$

$$p_n = \frac{\lambda_0 \, \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \, \mu_2 \ldots \mu_n} p_0$$

Steady state exists for the Birth-and-Death process if and only if the solution p_n can be made to satisfy $\sum_{n=0}^{\infty} p_n = 1$

$$p_0 \left[1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 . \lambda_1}{\mu_1 . \mu_2} + ... \right] = 1$$

Hence, $p_0 = \left[1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0.\lambda_1}{\mu_1.\mu_2} + \frac{\lambda_0.\lambda_1.\lambda_2}{\mu_1.\mu_2.\mu_3} + ...\right]^{-1}$, which is the steady state probability of an empty system.

Concept of Effective Arrival/Service Rate

When we consider Markovian state-dependent queueing models or adaptive models it may be desirable to look into the queueing characteristics in terms of the so called effective arrival rate and service rate.

Effective arrival rate, λ^*, is defined as the mean number of arrivals per unit time who enters and remains in the system or mean number of arrivals who actually enters in a queueing system and is defined by $\lambda^* = \sum_{n=0}^{\infty} \lambda_n p_n$.

Effective service rate, μ^* is defined on similar lines and is given by

$$\mu^*(1 - p_o) = \sum_{1}^{\infty} \mu_n \, p_n$$

(The factor $1 - p_o$ appears because a service can be completed only when the system is non empty) where p_n is the steady-state probability that there are n customers in the system.

For the standard M/M/1 model, effective service rate, μ^* is given by

$$\mu^*(1 - p_o) = \sum_{1}^{\infty} \mu_n \, p_n = \sum_{1}^{\infty} \mu \, p_n = \mu(1 - p_o)$$

The global balance equation $\lambda_n\, p_n = \mu_{n+1}\, p_{n+1}$ for M/M/1 queue model implies that $\sum_{n=0}^{\infty} \lambda_n\, p_n = \sum_{n=0}^{\infty} \mu_{n+1}\, p_{n+1} = \sum_{n=1}^{\infty} \mu_n\, p_n$, which may be deduced directly from the fact that the long run average arrival rate must equal the long run average service rate. Hence by definition it follows that $\lambda^* = \mu^*\,(1 - p_o)$. Hence from the knowledge of effective arrival rate λ^*, effective service rate can be obtained as $\mu^* = \lambda^*/(1\text{-}p_o)$.

Modified Bessel Function, $I_m(x)$:

The modified Bessel function $I_m\,(x)$ are solutions of Bessel's modified equation, $x^2 y'' + xy' - (x^2 + m^2)\, y = 0$, $m > 0$, $x > 0$.

The series representation for the modified Bessel function of the first kind of order m is

$I_m(x) = i^{-m} J_m\,(i\,x) = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2k+m}}{k!(k+m)!}$, $m = 0, 1, 2, \ldots$ whose values are available in tables, see for example Friedrich Losch (1960).

When $m = n$, $n = 0, 1, 2, \ldots$ we have that

$$I_{-n}\,(x) = I_n\,(x).$$

That is, the modified Bessel function with negative integer are the same as those with positive integer index. The modified Bessel functions are commonly associated with probability density functions in probability theory.

Many of the identities associated with modified Bessel function of the first kind are similar to those of the standard Bessel functions, some of which are displayed below:

1. $I_o\,(0) = 1$; $I_m\,(0) = 0,\ m > 0$

2. $e^{\pm x\cos\theta} = \sum_{n=-\infty}^{\infty} I_n(x).\cos n\theta$

3. $I_n(x) = \frac{1}{\pi}\int_{\theta}^{\pi} e^{x\cos\theta} \cos n\theta\ d\theta\ ,\ n = 0, 1, 2, \ldots$

4. $I_{m-1}\,(x) - I_{m+1}\,(\mathrm{x}) = \frac{2m}{x}\, I_m\,(x).$

5. $I_m(x) \approx \frac{\left(\frac{x}{2}\right)^m}{\Gamma(1+m)}\ ,\ m \neq -1, -2, -3, \ldots\ x \to 0^+$

6. $I_{1/2}(x) = \sqrt{\frac{2}{\pi x}}\ \sinh x\ ;\ I_{-1/2}(x) = \sqrt{\frac{2}{\pi x}}\ \cos hx$

7. $I_m(x) \approx \frac{e^x}{\sqrt{2\pi x}}$, $x \to \infty$, which is independent of m.

Modified Bessel function, $I_m(x)$ is completely a real function, identical with $J_n(x)$ (Bessel function of the first kind of order n), except its terms, instead of alternating in sign, are all positive. This new function, which is related to $J_n(x)$ in the same way that Coshx and sinhx are related to cos x and sin x, is denoted by $I_n(x)$.

Probability Generating Function (PGF)

It is defined as a function $G(t)$ of the form $G(t)=\sum_{k=0}^{\infty} p_k t^k$. The coefficients p_i are the probabilities of some random variable N taking the value i. One can determine the mean and variance of N using PGF as $G'(1)=E(N)$ and $Var(N)=G''(1)+G'(1)-[G'(1)]^2$.

We use the Birth - death methodology to analyse the various transient and steady state Markovian state-dependent queueing models. We obtain the partial differential equation (PDE) satisfied by the probability generating function (PGF), $G(z, t) = \sum_{0}^{\infty} p_n(t) z^n$, where $p_n(t)$ is the probability that there are n customers in the system at time t. Now from $G(z, t)$, $p_n(t)$ can be derived using the formula,

$$p_n(t)=\frac{1}{n!}\frac{\partial^n}{\partial z^n} G(z,t)\Big]_{z=0}$$

The PDE satisfied by $G(z, t)$ is of the form $P\frac{\partial}{\partial t}G(z,t)+Q\frac{\partial}{\partial z}G(z,t)=R$ where P, Q and R are functions of G, z and t, called Lagranges first order linear PDE. For obtaining the general solution of, we form the related equations, $\frac{dt}{P}=\frac{dz}{Q}=\frac{dG}{R}$.

If $u = u(t, z, G) = C_1$, a constant and $V = V(t, z, G) = C_2$, a constant are two particular integrals of these related equations, then the general solutions is given by $u = f(v)$ where $f(.)$ is an arbitrary function, from which we can write down the PGF, $G(z, t)$.

We can make use of $G(z, t)$ to derive $p_n(t)$ and its moments.

$$p_n(t) = \frac{1}{n!} \frac{\partial^m}{\partial z^n} G(z,t)\Big]_{z=0}$$

$p_n(t)$ is also the Coefficient of z^n in the series expansion of $G(z, t)$.

We may use $G(z, t)$ to find the mean and variance of $N(t)$.

System size at time $t = E\{N(t)\}$

$$= \frac{\partial}{\partial z} G(z,t)\Big]_{z=1} \quad (= G'(1))$$

$$Var\{N(t)\} = G''(1) + G'(1) - [G'(1)]^2$$

Little's Formula

This formula gives a very important relation between L_S, the mean number of customers in the system, W_S, the mean sojourn time and λ, the average number of customers entering the system per unit time. It states that $L_S = \lambda\, W_S$. Here it is assumed that the capacity of the system is sufficient to deal with the customers (i.e., the number of customers does not grow to ∞)

1.6 WAITING TIME DISTRIBUTION

Often we are interested in the amount of time that a typical customer spends in a queueing system. We define W_S as the expected time a customer spends in the queueing system, including time in queue plus time in service, and

W_q as the expected time a customer spend waiting in queue. Both W_S and W_q are computed under the assumption that steady state has been reached. By using the powerful result/formula known as Little's formula, $L = \lambda W$, W_S and W_q may easily computed from L_S and L_q (λ = average number of customer arriving in the queueing system; L_S = average number of customers present in the system and L_q = average number of customers waiting in the queue for service). These waiting time is applicable for all queue disciplines. It is to be noted that the probability distribution of waiting times, $W_S(t)$ and $W_q(t)$ are not needed to obtain W_S and W_q respectively. Questions concerning probabilities of waits greater than or less than specified seconds or minutes can be answered using the pdf of waiting time. It is to be noted that the pdf of waiting time depends on the queue discipline. Customers queueing time or waiting time are of direct interest when there is an economic loss if a customer is kept queueing. If the loss per unit delay is constant, only the mean queueing time, W_q or waiting time, W_S need be considered. Instances where the distribution and not just the mean waiting time are of interest arise when a customer may leave the system if delayed time, or when there is a penalty if a customer is delayed longer than some critical time.

For example, if the customers are items of material that have passed through one stage of processing and are waiting for a second stage, it may happen that small delays are of no consequence, but long delays will spoil the final product. In such a situation we shall be interested in the distribution of queueing time. On the other hand suppose one is interested to know the proportion of time the customer has spend in the system, then for computing this probability he/she needs the pdf of the waiting time in the system.

The objective of any queueing model is to determine how to provide service to customers so as to minimize the total cost of service and waiting time of customers by manipulating certain factors such as the number of servers, the rate of service, and order of service.

There are many situations in which customers are not served on a FCFS basis. We consider the service in Random order (SIRO) and LCFS queue discipline. Let W_{FCFS}, W_{SIRO} and W_{LCFS} be the random variables representing a customers waiting time in queueing systems under the disciplines FCFS, SIRO, and E (W_{FCFS}) = E (W_{SIRO}) = E (W_{LCFS}). It can also be shown that:

$$Var(W_{FCFS}) < Var(W_{SIRO}) < Var(W_{LCFS})$$

Since a large variance is usually associated with a random variable that has a relatively large chance of assuming extreme values, the above relation indicate that relatively large waiting times are most likely to occur with an LCFS discipline and least likely to occur with an FCFS discipline. This is reasonable, because in an LCFS system a customer can get lucky and immediately enter service but can also be bumped to the end of a long queue. In FCFS, however, the customer cannot be bumped to the end of a long line, so a very long wait is relatively unlikely. It is more likely for the customer to experience both short and long waiting times under SIRO discipline than under FCFS, a result which can be backed by intuition, see for example, L. Flatto (1997).

Components of a Waiting Time Study

The events defining waiting period and the design to register them are the two most important components of a waiting time study. The study design can be classified as either prospective or retrospective to refer to the approach of subject identification. For example, in health care problems, while retrospective design is used to study how long the patients who received treatment were required to wait from treatment design, prospective studies examine how long the patients accepted for treatment had to wait for it.

Methods to Characterize Waiting Time Variations

We will examine how the statistical variation of waiting times can be characterized. The variation of waiting times is fully described by the cumulative distribution function, $F(t)$, which is the probability that the waiting time T is less than some stated value t, $F(t) = P(T < t)$. For example, in health services research, where interest usually lies in describing how quickly patients receive service, we will call it the access function. Simply stated, the access function shows the probability of access by time in a queue. Thus higher rates of one access function relative to another mean quicker access. Largely employed in time-to-event analysis, the survival function shows chances of staying in a queue beyond a certain time. It is the complement of he access function $S(t) = 1 - F(t)$. As time to admission is the number of service planning cycles, we assume that waiting time is a discrete random variable that can take on positive integer values $j = 1, 2, \ldots$ with probabilities p_j, and $\Sigma\, p_j = 1$. The access function for a waiting time i is then defined as follows: $F_i = \sum_{j \le i} p_j$. If there is a limit M for waiting time values, and all probabilities p_j are equal, then the access function follows the discrete uniform distribution $F_i = \frac{i}{M}$, where $I = 1, 2, \dots M$.

The access function provides a useful descriptive measure of the overall pattern of waiting times. First, it may be used to export access probabilities at some meaningful moments of time on the waiting list, such as maximum time that patient can be safely waiting for surgery. Second, it may be used to estimate quantiles of the waiting times distribution (i.e., the moments of time when access reaches certain rates). The access function quantiles in one group of patients then could be compared with the quantiles in another groups.

Significance of Waiting Time in the Queue

In the design of queueing systems, it may be more important to focus on waiting time in the queue than on the total time in the system. If this is the case, it may be proved that the optimal number of servers (when the service rate is also a variable) is markedly different from well-known results for minimizing the total system time, one may refer Xiuli, Chaao and Carlten Scott (2000).

In a wide variety of manufacturing and service applications the waiting time in the queue is more significant than total time in the system (sojourn time). This is the case, for instance, in semiconductor fabrication and in the steel industry, where longer waiting time (for wafers) means more contamination and hence low yield. Examples in the service sector include amusement parks, where customers generally prefer a shorter wait and longer service; and good class restaurants, where customers prefer to be seated as soon as possible rather than to queue at the door, but many prefer a more leisurely service rate once seated. This is consistent with Maister (1985), who states that "pre-process waiting feels longer than in-process waiting", as a principle of queueing. In all these situations, customer satisfaction can be enhanced by having more (even though slower) servers rather than fewer (but faster) servers.

1.7 REVIEW OF STATE-DEPENDENT MARKOVIAN QUEUEING MODELS

In the study of queueing systems, the arrival rate λ_n and service rate μ_n are generally assumed to be constant regardless of how many customers are already in the system. Unfortunately, this is not the case in many practical queueing systems. When there is a large back-log of work (i.e., a long queue), it is quite likely that a server will tend to work faster than when the back-log is small or non-existent. That is, the service rate μ_n depends on n, the number of customers already in the system. Similarly situations may occur where customers reluctant to join the queue because of long waiting by seeing a large number of customers in the system (i.e., here the arrival rate λ_n depends on n, the number of customers already in the system). These kind of queueing systems where the arrival rate and/or service rate depends on the number of customers already in the system are called state-dependent or adaptive queueing systems. The state dependent queueing Models are of interest for customers/servers interaction. Their analysis, in particular, the specification of measures of effectiveness belong to the realm of applied probability.

Our study of state dependent queueing models is motivated by queueing scenarios where the arrival rate and/or the speed of the server depends on the amount of work present like production systems and internet. Queueing systems were the service speed is work load dependent are well-known, specifically in the studies of dams and storage processes, see for example Brockwell et. al. (1982); Bertrand, J.W.M., Van Ooijen, H.P.G. (2002); Harrison, J.M., Resnick, S.L. (1976) and Kaspi et. al. (1996).

In production system the speed of the server often depends on the amount of work. This is particularly true if the server is not being represented by a machine, but rather by a human-being, see for example Bertrand, J.W.M., and Van Ooijen, H.P.G., (2002), where the speed of the server is relatively high or low when there is much work or little work. In addition to general service fields, the rate at which jobs arrive at the system may also depend on the amount of work present. In the human-server production systems we may try to control the arrival rate of jobs to optimize server performance. In packet switched connection systems, the transmission rate of data connections may dynamically adapted based on buffer content, see for example Mitra, D. and Elwalid, A.I. (1994). These examples illustrate the state-dependent queue systems which will be the main concern in this thesis work. The simplest generalization giving a state-dependent system is the M/M/1 in which the parameters λ and μ assumed to be state-dependent.

The central model is the generalization of M/M/N system in which the mean arrival rate λ_n and the mean service rate μ_n are both functions of n, the number of customers are already present in the system. In many real situations λ_n and μ_n changes whenever n changes, both demand and service correspond to the system state. We are interested in models restricted to those favouring statistical equilibrium. There is a large literature concerned with different methods for the reduction of congestion at high traffic intensity. For example, consider customers turned away (lost) when the system state reaches a certain level or customers leaving the system before being served due to long waiting line. The second one is called "balking" and is well described in Haight (1957, 1959) and also in Ancker and Gafarin (1963).

State-dependent/adaptive queueing models have been discussed (at least) in the literature since Cox and Smith (1961). In the more general context of birth and death processes they have been reviewed and analysed, for example, by Keilson (1964, 1965); Hadidi and Conolly B.W. (1969 a, b); Hadidi (1969, 1972, 1973); Natvig, B. (1973); Conolly (1974) and many other researchers. Models of this kind are special interest in Operations Research because they can be interpreted as representing queueing systems in which the parameters of demand (arrival) and service (departure) are dependent on the actual number of customers already in the system at a given moment.

The state-dependent Queueing Models in the more general context of birth and death processes have been discussed far and widely in the literature by Conolly (1975); Conolly B.W., (1974, 1975); Hadidi and Conolly B.W. (1964); Parthasarathy, P.R., and Selvaraju, N., (2001) and many other researchers. In "Lecture notes on queueing systems", Conolly (1975) studied three sate-dependent Queueing Models, viz. Model A, Model B and Model C respectively.

Model A, which is a single server system with linearly state-dependent service is characterized by $\lambda_n = \lambda$, $n \geq 0$; $\mu_n = n\mu$, $n \geq 1$ where λ_n is the arrival rate and μ_n is service rate when there are n customers in the system and $\lambda > 0$, $\mu > 0$. In recent years there have been a considerable interest in computing state probabilities of various queueing Models which is used in performance analysis of computer/communication system. Since performance measures are functions of state probabilities their exact evaluation is important. Also many times a system designer is interested in probability distribution of number of customers in the system at arbitrary, departure and pre-arrival epochs which provide more information than few moments. For Model A, Conolly obtained $p_n(t)$, which is the probability distribution of the queue length at time t. From $p_n(t)$, he obtained the study state probability distribution p_n and using Little's formula, he also obtained the expected waiting time of an arbitrary arrival in the steady-state.

Model B is characterized by $\lambda_n = \frac{\lambda}{n+1}$, $n \geq 0$; $\mu_n = \mu$, $n \geq 1$ where $\lambda > 0$, $\mu > 0$. This is a state-dependent queue Model where potential customers are discouraged by queue length. For this model he obtained p_n using the birth-death methodology and using the result, the pdf of the waiting time distribution. He also obtained mean and variance of waiting time distribution under FCFS queue discipline.

Model C, which is a combination of Model A and Model B is characterized by $\lambda_n = \frac{\lambda}{n+1}$, $n \geq 0$; $\mu_n = n\mu$ $n \geq 1$ where $\lambda > 0$, $\mu > 0$.

He obtained the steady-state probability that there are n customers in the system as

$$\frac{\rho^n}{n!\, n!\, I_0(2\sqrt{\rho})}$$

where $I_m(x)$ is the Modified Bessel function of the first kind of order m defined by

$$I_m(x) = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2k+m}}{k!(k+m)!}, \qquad m = 0, 1, 3, \ldots$$

The expected system L_S and other performance measures are obtained in this study. He also evaluated numerically the various steady-state performance measures of these Models.

1.8 MAIN CONTRIBUTIONS OF THE THESIS

The present study is concerned with some state-dependent Markovian queueing models. The thesis consists of seven chapters starting with an introduction which brings together the relevant materials on state-dependent Markovian queueing models. Through the remaining six chapters various models are investigated to gain information on their performance measures.

Our study of state-dependent queueing models is motivated by queueing scenarios where the arrival rate and/or server speeds depends on the amount of work present, like production systems and internet. The study is mainly concentrated on state-dependent queueing models and their performance evaluations.

We study different linear and non-linear state-dependent Markovian queueing models and their modified forms. We obtain transient/steady-state probability distribution of queue length using probability generating function (PGF)/birth-death methodology. Using the probability distribution of queue length we obtained the pdf of waiting time and expected waiting time in the steady-state. We also obtain expected waiting time using Little's formula. Some statistics of the waiting time are also investigated. We obtain various other performance measures in the steady-state. Numerical evaluation of the various performance measures of the state-dependent Markovian queueing models are performed and compared numerically thereby suggesting the relative supremacy of the models. We also compare numerically these performance measures with that of the standard M/M/1 model. Numerical comparison of some statistics of waiting time distribution has also investigated and thus determine the model having less waiting time. Finally, numerical comparison between state dependent queueing models is done in terms of (i) mean busy period, (ii) mean inter arrival time, (iii) measures of queue density, and (iv) mean service rate and system state. We also discuss general linear and non linear state dependent queueing models. Analyzing these state dependent queueing models, we establish that there are linear and non linear state dependent queueing models having the same probability distribution of queue length and thus having the same queueing characteristics. The number of customers served during the waiting time of an arbitrary arrival in these models can be different.

In section 1.1 of Chapter 1, we introduce queueing systems with the basic elements of the system. We describe briefly the problem arising in queueing theory which are classified into three viz., (i) Behavioural problems of the system, (ii) Statistical problems of the system and (iii) Optimization problems of the system. Aims in studying a queueing system are presented in 1.3. The queueing parameters needed for an understanding and efficient management of a system are mentioned in section 1.4. Some preliminaries needed for the study is described in section 1.5. In section 1.6, various aspects of waiting time distribution are briefly discussed. The significance of waiting time in queue are also highlighted. A review of state-dependent Markovian queueing models are elaborated with a literature survey in section 1.7. The concept of effective arrival/service rate is also discussed. Finally the main contribution of the thesis are elaborated in section 1.8.

Chapter 2, deals with some general state-dependent queueing Models. Specifically we study four state-dependent Markovian queueing Models, viz. Model I characterized by the arrival rate $\lambda_n = \lambda$, $n \geq 0$; service rate $\mu_n = n\mu$, $n \geq 1$; Model II characterized by $\lambda_n = \dfrac{\lambda}{n+1}$, $n \geq 0$; $\mu_n = \mu$, $n \geq 1$; Model III characterized by $\lambda_n = \dfrac{\lambda}{n+1}$, $n \geq 0$; $\mu_n = n\mu$, $n \geq 1$; and Model IV characterized by $\lambda_n = (n + 1)\lambda$, $n \geq 0$; $\mu_n = n\mu$, $n \geq 1$; corresponding to different types of arrival/service rates where $\lambda > 0$, $\mu > 0$. We obtain transient/steady-state probability distribution of queue length for these models using the probability generating function/birth-death methodology. Using $p_n(t)/p_n$, we obtain various performance measures of the models. We also obtain the pdf of the waiting time distribution and expected waiting time in the steady state. We evaluate numerically the various steady-state performance measures of the models and compare numerically thereby suggesting the relative supremacy of the models. We also compare numerically these performance measures with that of non-adaptive M/M/1 model. By means of this comparison, it is proved that the transient probability of queue length of the state-dependent queueing model characterized by the arrival rate $\lambda_n = (n + 1)\lambda$, $n \geq 0$; service rate $\mu_n = n\mu$ $n \geq 1$, where $\lambda > 0$, $\mu > 0$ and M/M/1 model are vastly different even though the steady-state probabilities are same.

In Chapter 3, performance measures of state-dependent queueing models are discussed. We obtain the queue characteristics of various state-dependent queueing models. We evaluate numerically the performance measures of these models and compare them. We also compare these performance measures with that of the standard M/M/1 model. Finally a brief summary of results and concluding remarks are presented in section 3.5.

In Chapter 4, we introduce Modified state-dependent Markovian queueing models. We discuss several Modified versions of the Models considered in Chapter 2 and obtain the transient/steady state probability distribution of queue length using probability generating function/birth-death Methodology. Using this probability distribution, we obtain various performance measures. We also obtain waiting time distribution and expected waiting time in the system. For the model AVI, where $\lambda_n = \lambda$, n ≥ 0; $\mu_n = (n + 2)\mu$, n ≥ 1 ($\lambda,\mu > 0$), the number of customers served during the waiting time of an arbitrary arrival is equal to the number of customers in the system when $\lambda = \mu$. It may be proved that the steady-state probability distribution of the number of customers for the linearly state-dependent Markovian queueing models characterized by the arrival rate $\lambda_n = (n + k)\lambda$, $n \geq 0$; service rate $\mu_n = (n + r)\mu$, $n \geq 1$ where k and r are non-negative integers satisfying $k \geq 1$ and $r \geq 0$ follows Geometric law given by:

$p_n = (1 - \rho)\rho^n$, $n \geq 0$ and $\rho = \dfrac{\lambda}{\mu} < 1$ for pairs of $\{(k,r)\} = \{(1,0),(2,1),(3,2),...\}$,

which is the same as that of the standard M/M/1 model. It is verified that when ρ increases, μW_S, the number of customers served during the waiting time of an arbitrary arrival, decreases. Finally, a brief summary of results and concluding remarks are presented in section 4.9.

In Chapter 5, we introduce some non-linear state-dependent Markovian queueing models. We obtain the probability distribution of queue length and the various performance measures for these models. We obtain the pdf of waiting time distribution for the queueing models with state-dependent arrival rates and constant service rate. We also obtain the expected waiting time and the number of customers served during the waiting time of an arbitrary arrival. We also evaluate numerically the performance measures of these models and make a comparison between them. It is proved that the steady-state probability distribution of queue length of the non-linear state-dependent queueing models characterized by the arrival rate $\lambda_n = (n + 1)^m, n \geq 0$; service rate $\mu_n = n^m\mu, n \geq 1$ where $m \geq 2$ follows Geometric law given by $p_n = (1 - \rho)\rho^n, n \geq 0$ and $\rho = \frac{\lambda}{\mu} < 1$ which is the same as that of the non-adaptive M/M/1 model. It is verified that there are linear and non-linear state-dependent queueing Models having the same steady-state probability distribution of queue length given by $p_n = (1 - \rho)\rho^n, n \geq 0$ and $\rho = \frac{\lambda}{\mu} < 1$. With reference to the performance measure p_o, the state probability of an empty system, one can do better than the model characterized by $\lambda_n = \frac{\lambda}{2n+1}, n \geq 0$; $\mu_n = 2n\mu$, $n \geq 1$. An example is provided by $p_0 = \frac{1}{I_0(\sqrt{\rho})}$, which can be achieved by the state-dependent model characterized by the arrival rate $\lambda_n = \frac{\lambda}{2n+2}, n \geq 0$; $\mu_n =$

$2n\mu$, $n \geq 1$, where $I_0(\sqrt{\rho})$ is the modified Bessel function of the first kind of order zero.

In Chapter 6, we discuss the pdf of waiting time distribution of queueing models with the state-dependent arrival rate and constant service rate. We also compute numerically some statistics of the waiting time distribution and make a comparison and thus decide the state-dependent queueing model having less waiting time. We also compare these measures with that of the standard M/M/1 model and verified that the expected waiting time and coefficient of variation of waiting time of the two models considered in this chapter is less than that of the standard M/M/1 model. Finally a brief summary of results and concluding remarks are given in section 6.8.

Finally in Chapter 7, a comparison between state-dependent queueing models is made in terms of (i) $\frac{\hat{B}}{\hat{T}}$, a measure of queue density, (ii) $\lambda\hat{B}$ (which measure the number of arrivals during the busy period), $\lambda\hat{W}$, $\lambda\hat{S}$ and (iii) Mean system size, where $\hat{B}$ = Mean busy period, $\hat{T}$ = Mean inter arrival time and $\hat{S}$ = Mean service time. We also discuss general linear and non-linear state-dependent queueing models. Analyzing these state-dependent queueing models, we show that there are linear and non-linear state-dependent queueing models having the same probability distribution of queue length and thus having the same queueing characteristics. The number of customers served during the waiting time of an arbitrary arrival in the system for these models can be different.

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

Chapter 2

GENERAL STATE-DEPENDENT MARKOVIAN QUEUEING MODELS

2.1 INTRODUCTION

In this Chapter, we shall study various partial/complete state-dependent Markovian queueing models. In analyzing these models, one would like to compute certain performance measures, such as the steady state queue length distribution $\{p_n\}$, transient queue length distribution $\{p_n(t)\}$, average number of customers in the queue or in the system, expected waiting time of an arrival, sojourn time distribution, time for the queue to reach some specified number, etc. For a specified model, these quantities may all be characterized as solutions to certain equations. Thus, computing the performance measures amounts to solving these equations (Differential-difference equations) together with appropriate initial conditions. Such measures of performance measures for the purpose of making recommendations about the design of the system.

We shall consider the following particular state dependent Markovian queueing models and obtain transient and/or steady state results and make a numerical comparison of performance measure of these models with the results of non-adaptive standard queueing model, *M/M/*1. The method of probability generating function (PGF) is used in the analysis.

Model I : $\lambda_n = \lambda, n \geq 0, \mu_n = n\mu, n \geq 1$

Model II : $\lambda_n = \dfrac{\lambda}{n+1}, n \geq 0; \mu_n = \mu,\ n \geq 1$

Model III : $\lambda_n = \dfrac{\lambda}{n+1}, n \geq 0; \mu_n = n\mu,\ n \geq 1$

Model IV : $\lambda_n = (n+1)\lambda, n \geq 0; \mu_n = n\mu,\ n \geq 1$

where λ_n is the arrival rate and μ_n is the service rate when there are n customers already in the system ($\lambda > 0$, $\mu > 0$).

2.2 MODEL I

Consider a single server partially state dependent Markovian queueing model where the rate of arrivals, λ_n, does not depend on the number of customers, n, already present in the system but the service rate, μ_n, does depends on the number, n, present in the system. For this model, a long queue causes the server to serve more quickly in order to avoid congestion and obviously no queue forms in this case, since an arrival immediately enters service. This model seems realistic in system where the server reacts to the number of customers waiting.

Let $N(t)$ denote the number of customers present in the system at time t. We shall find $p_{in}(t)$ or $p_n(t) = P\ \{N(t) = n/N(0) = i\}$ which is the probability that $N(t)$ takes on the value n = 0, 1, 2, 3 ... at time t given that initially there are i customers in the system, in a general Birth-and-death single server queueing process with parameters λ_n and μ_n defined above and obtain the queue characteristics as under:

The Differential - difference equations for $p_n(t)$ are:

$$p_0'(t) = -\lambda\, p_o(t) + \mu\, p_1(t) \tag{2.2.1}$$

$$p_n'(t) = -(\lambda + n\mu)\, p_n(t) + \lambda\, p_{n-1}(t) + (n+1)\mu\, p_{n+1}(t) \tag{2.2.2}$$

Let,

$$G(z,t) = \sum_{o}^{\infty} p_n(t)\, z^n \tag{2.2.3}$$

be the probability generating function (PGF) of $p_n(t)$

$$\frac{\partial}{\partial z} G(z,t) = \Sigma\, p_n(t)\, n\, z^{n-1} \;\&\; \frac{\partial}{\partial t} G(z,t) = \sum_{0}^{\infty} p_n'(t)\, z^n$$

$$\frac{\partial}{\partial t} G(z,t) - p_0'(t) = \sum_{1}^{\infty} p_n'(t)\, z^n \tag{2.2.4}$$

$$\frac{\partial}{\partial t} G(z,t) - [\lambda\, p_o(t) + \mu\, p_1(t) = -\lambda \sum_{1}^{\infty} p_n(t)\, z^n - \mu \sum_{1}^{\infty} n\, p_n(t)\, z^n + \lambda \sum_{1}^{\infty} p_{n-1}(t)\, z^n + \mu \sum_{1}^{\infty} (n+1) p_{n+1}(t)\, z^n$$

(using (2.2.1) and (2.2.2) suitably in (2.2.4))

i.e.,
$$\frac{\partial}{\partial t} G(z,t) + \mu(z-1) \frac{\partial}{\partial z} G(z,t) = \lambda(z-1)\, G(z,t) \tag{2.2.5}$$

(2.2.5) is the partial differential equation (PDE) satisfied by $G(z, t)$ which is to be solved using the initial condition $G(z, 0) = z^i$ (initially there are i customers in the system).

For (2.2.5), the related equations are

$$\frac{dt}{1} = \frac{dz}{\mu(z-1)} = \frac{dG}{\lambda(z-1)G}$$

$\frac{dt}{1}=\frac{dz}{\mu(z-1)}$ gives $\frac{dz}{dt}=\mu(z-1)$ which yields $u(t,z,G)=(1-z)e^{-\mu t}=C_1$ (a constant).

$\frac{dz}{\mu(z-1)}=\frac{dG}{\lambda(z-1)G}$ gives $\frac{dG}{dz}=\frac{\lambda}{\mu}G$ which yields $U(t,z,G)=Ge^{-\frac{\lambda}{\mu}z}=C_2$ (a constant).

Thus the general solution of (2.2.5) is given by $U=g(u)$

i.e., $$G(z,t)=e^{\frac{\lambda}{\mu}z}\, g\{(1-z)e^{-\mu t}\} \tag{2.2.6}$$

where g(.) is an arbitrary function.

$$G(z,0)=e^{\frac{\lambda}{\mu}z}\; g\{(1-z)\}$$

i.e., $z^i=e^{\rho z}\; g(1-z)$, where $\rho=\frac{\lambda}{\mu}$.

With $1-z=y$, $z=1-y$

$$g(y)=e^{-\rho(1-y)},[1-y]^i$$

Hence from (2.2.6),

$$G(z,t)=e^{\rho z}\,e^{-p[1-(1-z)e^{-\mu t}]}\,[1-(1-z)e^{-\mu t}]^i$$

i.e., $G(z,t)=\; e^{-p(1-z)(1-e^{-\mu t})}\,[1-(1-z)\,e^{-\mu t}]^i$ (2.2.7)

Now we shall use (2.2.7) to obtain $p_n(t)$, $E\{N(t)]$ and $\mathrm{Var}\{N(t)\}$

$$p_n(t) = \frac{1}{n!}\frac{\partial^n}{\partial z^n} G(z,t)]_{z=0}$$

$$= \frac{1}{n!}\left\{\begin{array}{l} \left[1-(1-z)e^{-\mu t}\right]^i e^{-\rho(1-z)(1-e^{-\mu t})}\rho^n(1-e^{-\mu t})^n \\ + nC_1\left(i\left[1-(1-z)e^{-\mu t}\right]^{i-1} e^{-\mu t}\; e^{-\rho(1-z)(1-e^{-\mu t})}\rho^{n-1}(1-e^{-\mu t})^{n-1}\right) \\ + nC_2\left(i(i-1)\left[1-(1-z)\,e^{-\mu t}\right]^{i-2} e^{-2\mu t}\; e^{-\rho(1-z)(1-e^{-\mu t})}\rho^{n-2}(1-e^{-\mu t})^{n-2}\right) + \ldots \\ + (iC_n)\, n!\, e^{-n\mu t}\left[1-(1-z)e^{-\mu t}\right]^{i-n} e^{-\rho(1-z)(1-e^{-\mu t})} \end{array}\right\}_{z=0}$$

(using Leibnitz's result for the n^{th} derivative of a product)

i.e.,
$$p_n(t) = \frac{1}{n!}\left\{\begin{array}{l} \rho^n\left(1-e^{-\mu t}\right)^n e^{-\rho(1-e^{-\mu t})}(1-e^{-\mu t})^i \\ + n\left[i\left(1-e^{-\mu t}\right)^{i-1} e^{-\mu t}\; e^{-\rho(1-e^{-\mu t})}\rho^{n-1}(1-e^{-\mu t})^{n-1}\right] + \ldots \\ \ldots + iC_n\; n!\, e^{-n\mu t}\left(1-e^{-\mu t}\right)^{i-n} e^{-\rho(1-e^{-\mu t})} \end{array}\right\}$$

Then for steady state solution

$$p_n = \lim_{t\to\infty} p_n(t) = \frac{e^{-\rho}\rho^n}{n!}\;, n \geq 0 \qquad (2.2.8)$$

which shows that after a long time the number in the system has a Poisson distribution with parameter $\rho = \dfrac{\lambda}{\mu}$

Now $E\{N(t)\}$ = Average number of customers in the system at time t.

$$= \frac{\partial}{\partial z}G(z,t)_{z=1}\,(=G'(1)) \qquad (2.2.9)$$
$$= \rho\left(1-e^{-\mu t}\right) + i\,e^{-\mu t}$$

$$\text{Var}\left[N(t)\right] = G''(1) + G'(1) - \left[G'(1)\right]^2$$

$$\begin{aligned} &= 2i\rho e^{-\mu t}\left(1 - e^{-\mu t}\right) - ie^{-2\mu t} \\ &+ ie^{-\mu t} + \rho\left(1 - e^{-\mu t}\right) \end{aligned} \tag{2.2.10}$$

When $t \to \infty$ $E\{N(t)\} \to E(N) = L_S$ (system size at any time) and $\text{Var}\{N(t) \to \text{Var}(N)\} = \rho$

From the knowledge of L_S and λ one can obtain the expected waiting time in the system, W_S, of an arrival in the steady state as

$$W_S = \frac{L_S}{\lambda} \qquad \text{(Little's formula)}$$

$= \frac{\rho}{\lambda} = \frac{1}{\mu}$, independent of arrival rate λ. Hence for all ρ, the number of customers served during the waiting time of an arbitrary arrival, is given by

$$\mu W_S = 1 \tag{2.2.11}$$

We shall also obtain p_n, using the Birth-death methodology as under:

$$p_n = \frac{\lambda_{0,}\lambda_1 \ldots \lambda_{n-1}}{\mu_1, \mu_2 \ldots \mu_n} p_0 = \frac{\lambda_{,}\lambda_1 \ldots \lambda}{\mu(2\mu)\ldots(n\mu)}$$

$$= \frac{\rho^n}{n!} p_0, \ \rho = \frac{\lambda}{\mu}$$

using $\Sigma p_n = 1$ gives $p_o = e^{-\rho}$

Hence $$p_n = \frac{e^{-\rho}\rho^n}{n!}, n \geq 0 \tag{2.2.12}$$

which agrees with the result obtained in (2.2.8).

One can also obtain L_S and Var(N) using p_n as under:

System size at any time, $L_S = E(N) = \Sigma n \dfrac{e^{-\rho}\rho^n}{n!} = \rho$

$$E(N^2) = \Sigma n^2 p_n = \Sigma[n(n-1)+n]\,\frac{e^{-\rho}\rho^n}{n!} = \rho^2 + \rho$$

Hence,

$$Var(N) = E(N^2) - [E(N)]^2 = \rho^2 + \rho - \rho^2 = \rho$$

L_S and Var(N) agrees with the result obtained in (2.2.9) and (2.2.10) respectively while $t \to \infty$.

λ^* = Mean number of customers who actually enters in the system (effective arrival rate).

$$= \Sigma \lambda_n\, p_n = \Sigma \lambda \frac{e^{-\rho}\,\rho^n}{n!} = \lambda \tag{2.2.13}$$

which is obviously true.

Effective service rate μ^* is given by

$$\mu^*(1 - p_0) = \sum_{n=1}^{\infty} \mu_n p_n = \lambda$$

Hence,

$$\mu^* = \frac{\lambda}{1 - p_o} = \frac{\lambda}{1 - e^{-\rho}} \tag{2.2.14}$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = 1 - e^{-\rho} = 1 - p_o \tag{2.2.15}$$

which is the number of customers at the server. It is also the fraction of time the server is working which increases with increasing value of ρ.

2.3 MODEL II

This is a single service partially dependent Markovian queueing model where customers will be discouraged from joining the queue if on arrival they find a large number of customers in the system already. Arrivals are geared (or could be controlled) in accordance with availability of service.

The discouraged arrivals queue has been studied in the past by Cox and Smith (1961), B.W. Conolly (1974), Natvig. B (1975) and E.A. Van Doorn (1981). It is a model where demand is regulated by what a new customer sees waiting when he arrives: It expresses the deterrent effect of a long queue and has so been denominated by Cox and Smith.

The discouraged arrivals single service queueing model is useful to model a computing facility that is solely dedicated to batches - jobs processing. Job submissions are discouraged when the facility is heavily used and can be modeled as a Poisson process with the state dependent arrival rate. The time taken to process each job is exponentially distributed with a constant service rate regardless of the number of jobs in the system.

Model II is characterized by $\lambda_n = \dfrac{\lambda}{n+1}$, $n \geq 0$; $\mu_n = \mu, n \geq 1$ $(\lambda, \mu > 0)$

$N(t)$ denotes the number of customers in the system at time t. We shall find $p_n(t) = p_{on}(t) = P\{N(t) = n / N(0) = 0\}$ which is the probability $N(t)$ takes on the value $n = 0,1,2,3$ ... and obtain the queue characteristics as under:

The differential - difference equations for $p_n(t)$ are:

$$p_o'(t) = -\lambda\, p_o(t) + \mu\, p_1(t) \tag{2.3.1}$$

$$p_n'(t) = -\left(\frac{\lambda}{n+1} + \mu\right) p_n(t) + \frac{\lambda}{n}\; p_{n+1}(t),\; n \geq 1 \tag{2.3.2}$$

It is difficult to solve these system of equations to obtain $p_n(t)$, we shall find p_n, which is the steady state probability of n customers in the system.

By using the Birth-Death methodology we have,

$$p_n = \frac{\lambda_0\, \lambda_1 \ldots \lambda_{n-1}}{\mu_1\, \mu_1 \ldots \mu_n}\, p_0 = \frac{\frac{\lambda}{1}\frac{\lambda}{2}\cdots\frac{\lambda}{n}}{\mu(2\mu)\ldots(n\mu)}\, p_0$$

$$= \frac{\rho^n}{n!}\, p_0,\; \rho = \frac{\lambda}{\mu}$$

$\Sigma\, p_n = 1$ gives $p_o = e^{-\rho}$

Hence

$$p_n = \frac{e^{-\rho}\rho^n}{n!},\; n \geq 0 \tag{2.3.3}$$

which is a Poisson distribution with parameter ρ.

Note that Models I and II have the same steady state distribution,

$$p_n = \frac{e^{-\rho}\rho^n}{n!},\; n \geq 0$$

Average number of customers in the system in the steady state,

$$L_S = \Sigma n\, p_n = \rho \tag{2.3.4}$$

and Var(N) = ρ

λ^* = Average number of customers who actually enters in the system (Effective arrival rate)

$$= \Sigma \lambda_n\, p_n = \Sigma \frac{\lambda}{n+1} \frac{e^{-\rho} \rho^n}{n!} = \mu e^{-\rho} \left(e^{\rho} - 1\right) = \mu \left(1 - e^{-\rho}\right)$$

Effective service time, μ^*

$$\mu^* (1 - p_o) = \sum_{n=1}^{\infty} \mu_n\, p_n = \mu (1 - p_o)$$

Hence $\mu^* = \mu$, which is obviously true.

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = 1 - e^{-\rho} = 1 - p_o \tag{2.3.5}$$

which is the average number of customers at the server. It is also the proportion of the time the server is working which is exactly as for Model I.

Expected waiting time of an arrival in the steady state is given by

$$W_S = \frac{L_S}{\lambda^*} \quad \text{(using Little's formula)}$$

$$W_S = \frac{\rho}{\mu (1 - e^{-\rho})}$$

$$\mu W_S = \frac{\rho}{1-e^{-\rho}} \qquad (2.3.6)$$

which is the number of customers served during the waiting time of an arbitrary arrival. We prove below that (2.3.6) is a number greater than 1 and increases with increasing values of ρ.

Proof: $\mu W_s = \frac{\rho}{1-e^{-\rho}}$

$$= \left\{1-\left(\frac{\rho}{2}-\frac{\rho^2}{6}+...\right)\right\}^{-1}$$

$$= 1-\left(\frac{\rho}{2}-\frac{\rho^2}{6}+...\right)+\left(\frac{\rho}{2}-\frac{\rho^2}{6}+...\right)^2...$$

$$= 1+\frac{\rho}{2}+\frac{\rho^2}{12} \text{ (approximately)}$$

Clearly this value is always greater than 1 and increases with the increasing values of ρ. This result can also be verified by means of numerical evaluation.

Now we shall derive the pdf of the waiting time distribution of this model

Waiting Time Distribution

It is to be noted that the probability distribution of waiting time, T is not needed to find W_S. Questions concerning probabilities of waits greater than or less than specified minutes can be answered using the pdf of waiting time. Naturally this probability distribution will change with queue - discipline and is important from the customers point of view. Assume that customers are served in the order of arrival (ie; assume FCFS queue discipline). The duration of services are independent and identically distributed (i.i.d) random variables, independent of the arrival process. The service time distribution is exponential with parameter μ. If we write $f_n(t)$ for the pdf of a wait, duration t, including service for the n^{th} arrival and suppose that $f_n(t) \to f(t)$ as $t \to \infty$ then,

$$f(t) = \sum_{n=0}^{\infty} \frac{\lambda_n p_n}{\lambda} \mu \frac{e^{-\mu t}(\mu t)^n}{n!} \quad , \text{ so that } \int_0^{\infty} f(t)\,dt = 1$$

$$= \frac{e^{-\mu t} \lambda e^{-\rho}}{1-e^{-\rho}} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!(n+1)!}$$

$$= \frac{e^{-\mu t} \lambda e^{-\rho}}{1-e^{-\rho}} \frac{1}{\sqrt{\lambda t}} \sum_{n=0}^{\infty} \frac{\left(\sqrt{\lambda t}\right)^{2n+1}}{n!(n+1)!}$$

$$= \frac{e^{-\mu t}}{e^{\rho}-1} \sqrt{\frac{\lambda}{t}}\, I_1\left(2\sqrt{\lambda t}\right) \tag{2.3.7}$$

where $I_m(x) = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2k+m}}{k!(k+m)!}$ is the modified Bessel function of the first kind of order in $m = 0, 1, 2, \ldots.$

We shall obtain the expected waiting time, W_S, of an arrival using (2.3.7).

$W_S = \int_0^{\infty} t\, f(t)\,dt \; (= E(T))$, T denotes waiting time

$$= \int_0^{\infty} \frac{t\, e^{-\mu t}}{e^{\rho}-1} \sqrt{\frac{\lambda}{t}}\, I_1\left(2\sqrt{\lambda t}\right) dt$$

$$= \frac{1}{e^{\rho}-1} \int_0^{\infty} e^{-ut} \sqrt{\lambda t} \left[\sqrt{\lambda t} + \frac{\left(\sqrt{\lambda t}\right)^3}{1!2!} + \frac{\left(\sqrt{\lambda t}\right)^5}{2!3!} + \ldots \right] dt$$

(Using the series expansion of $I_1\left(2\sqrt{\lambda t}\right)$

$$= \frac{1}{e^{\rho}-1} \int_0^{\infty} e^{-ut} \left[\lambda t + \frac{(\lambda t)^2}{1!2!} + \frac{(\lambda t)^3}{2!3!} + \ldots \right] dt$$

$$= \frac{1}{e^{\rho}-1} \left\{ \lambda \left(\frac{1}{\mu^2}\right) + \frac{\lambda^2}{2!} \left(\frac{2!}{\mu^3}\right) + \frac{\lambda^3}{12} \left(\frac{3!}{\mu^4}\right) + \ldots \right\}$$

$$= \frac{1}{e^{\rho}-1} \frac{\lambda}{\mu^2} e^{\rho} = \frac{\rho}{\mu\left(1-e^{-\rho}\right)} \quad (2.3.8)$$

which agrees with the result obtained in (2.3.6) by Little's formula.

Now, $$E(T^2) = \int_0^{\infty} t^2 f(t)\, dt$$

$$= \int_0^{\infty} t^2 \frac{e^{-\mu t}}{e^{\rho}-1} \sqrt{\frac{\lambda}{t}}\, I_1\left(2\sqrt{\lambda t}\right) dt$$

$$= \frac{1}{\mu^2 (e^{\rho}-1)} \left\{\rho^2 e^{\rho} + 2\rho\, e^{\rho}\right\}$$

Hence variance of waiting time distribution is given by,

$$Var(T) = E(T^2) - [E(T)]^2$$

$$= \frac{\rho e^{\rho} (2e^{\rho} - \rho - 2)}{\mu^2 \left(e^{\rho}-1\right)^2}$$

Hence $$S.D \text{ of } T = \frac{\sqrt{\rho e^{\rho} (2e^{\rho} - \rho - 2)}}{\mu\, (e^{\rho}-1)} \quad (2.3.9)$$

Effective service rate (weighted average service rate) μ^* is given by

$$\mu^* (1-p_o) = \sum_1^{\infty} \mu_n\, p_n$$

$$= \mu (1-p_o)$$

Hence $\mu^* = \mu$, which is obviously true (2.3.10)

Effective traffic intensity, $\rho^* = \frac{\lambda^*}{\mu^*} = 1 - e^{-\rho} = 1 - p_o$, which is the mean number of customers at the server. It is also the fraction of time the server is working which increases with increasing value of ρ as in Model I.

2.4 MODEL III

This model embodies both mechanisms of Models I and II, considered in sections (2.2) & (2.3). Thus the model is characterized by

$$\lambda_n = \frac{\lambda}{n+1}, n \geq 0, \mu_n = n\mu, n \geq 1 \; (\lambda, \mu > 0)$$

See Conolly, B.W., (1974), Brian Conolly (1975) and Van Doorn (1981).

Using the Birth - Death methodology, we shall find p_n, which is the steady state probability that there are n customers in the system and obtain the queue characteristics as under:

$$p_n = \frac{\lambda_0, \lambda_1, \ldots \lambda_{n-1}}{\mu_1, \mu_2 \ldots \mu_n} p_o$$

$$= \frac{\frac{\lambda}{1}\frac{\lambda}{2}\cdots\frac{\lambda}{n}}{\mu \; 2\mu \ldots n\mu} p_o = \frac{\rho^n}{n!n!} p_o, \rho = \frac{\lambda}{\mu}$$

Now $\Sigma \, p_n = 1$ gives $p_o = \dfrac{1}{I_o(2\sqrt{\rho})}$

Hence,

$$p_n = \frac{\rho^n}{n!n! \, I_o(2\sqrt{\rho})}, \; n \geq 0 \tag{2.4.1}$$

Probability of an empty system/server is idle is given by $p_o = \dfrac{1}{I_o(2\sqrt{\rho})}$

Average number of customers in the system in the steady state, $L_S = \Sigma \; np_n$ (system size at any time) is

$$L_S = \frac{\rho}{I_o(2\sqrt{\rho})} \sum_{n=1}^{\infty} \frac{\rho^n}{(n-1)!\,n!}$$

$$= \frac{\rho}{I_o(2\sqrt{\rho})} \sum_{m=0}^{\infty} \frac{\rho^{m+1}}{m!\,(m+1)!}$$

$$= \frac{\rho}{I_o(2\sqrt{\rho})} \sum_{m=0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2m+1}}{m!\,(m+1)!}$$

$$= \frac{\sqrt{\rho}\quad I_1\left(2\sqrt{\rho}\right)}{I_o(2\sqrt{\rho})} \tag{2.4.2}$$

Now, $$E(N^2) = \frac{1}{I_o(2\sqrt{\rho})} \Sigma \frac{\rho^n}{(n-1)!(n-1)!}$$

$$= \frac{1}{I_o(2\sqrt{\rho})} \sum_{m=0}^{\infty} \frac{\rho^{m+1}}{m!m!}$$

$$= \frac{\rho}{I_o(2\sqrt{\rho})} \sum_{m=0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2m+0}}{m!m!}$$

$$= \frac{\rho}{I_o\left(2\sqrt{\rho}\right)} I_o\left(2\sqrt{\rho}\right) = \rho$$

Hence,

$$Var(N) = E(N^2) - [E(N)]^2 = \rho \left\{1 - \left[\frac{I_1\left(2\sqrt{\rho}\right)}{I_o\left(2\sqrt{\rho}\right)}\right]^2\right\} \tag{2.4.3}$$

which is the variance of the number of customers in the system in the steady state. Using $L_S = \lambda^* W_S$, the expected waiting time of an arrival in the steady state is given by,

$$W_S = \frac{L_S}{\lambda^*} \tag{2.4.4}$$

λ^* = Mean number of customers who actually enters in the system

$$= \sum_{0}^{\infty} \lambda_n \; p_n$$

$$= \sum_{n=0}^{\infty} \frac{\lambda}{n+1} \frac{\rho^n}{n!n!} I_o\left(2\sqrt{\rho}\right)$$

$$= \frac{\lambda}{I_o\left(2\sqrt{\rho}\right)} \frac{1}{\sqrt{\rho}} \sum_{n=0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2n+1}}{n!\,(n+1)!}$$

$$= \frac{\lambda}{\sqrt{\rho}} \frac{I_1\left(2\sqrt{\rho}\right)}{I_o\left(2\sqrt{\rho}\right)} \tag{2.4.5}$$

Using (2.4.3) in (2.4.4),

$$\mu W_S = 1 \tag{2.4.6}$$

which is a result shared with Model I. Hence for all ρ, the number of customers served during the waiting time of an arbitrary customer equals one.

Using the expansion of $I_o\left(2\sqrt{\rho}\right)$ and $I_1\left(2\sqrt{\rho}\right)$, for "small enough" ρ, it can been seen that, L_S for Model III $= \sqrt{\rho} \; \dfrac{\mathrm{I}_1\left(2\sqrt{\rho}\right)}{\mathrm{I}_o\left(2\sqrt{\rho}\right)}$

$$= \sqrt{\rho} \left\{ \sqrt{\rho} - \frac{\rho^{3/2}}{2} + ... \right\}$$

$$= \rho - \frac{\rho^2}{2} \text{ (approximately)}$$

For $M/M/1$ model,

$$L_S = \frac{\rho}{1-\rho} = \rho(1+\rho+\rho^2+...) = \rho+\rho^2 \text{ (approximately).}$$

Thus in the steady state, for "small enough" ρ, one sees that $L_S = \rho - \frac{\rho^2}{2}$ as compared with ρ for Models I and II and $\rho + \rho^2$ for $M/M/1$, indicating that a reduction of mean queue length is to be anticipated. By means of numerical comparison of performance measure it may be verified that Model III considered in section (2.4) is more accurate than both Models I and II alone.

Effective service rate, μ^* is given by

$$\mu^*(1-p_o) = \sum_1^{\infty} \mu_n p_n$$

$$= \Sigma n\mu \frac{\rho^n}{n!n!I_o(2\sqrt{\rho})}$$

$$= \frac{\mu}{I_o(2\sqrt{\rho})}\left(\sqrt{\rho}\right) I_1\left(2\sqrt{\rho}\right)$$

Hence,

$$\mu^* = \frac{\mu\left(\sqrt{\rho}\right) I_1\left(2\sqrt{\rho}\right)}{I_o(2\sqrt{\rho})-1}$$

or

$$\mu^* = \frac{\lambda^*}{1-p_0} = \frac{\mu\left(\sqrt{\rho}\right) I_1\left(2\sqrt{\rho}\right)}{I_o(2\sqrt{\rho})-1} \qquad (2.4.7)$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \frac{I_o\left(2\sqrt{\rho}\right)-1}{I_o\left(2\sqrt{\rho}\right)} = 1 - p_o \qquad (2.4.8)$$

which is the mean number of customers at the server. It is also the fraction of time the server is working.

2.5 MODEL IV

This queueing model is characterized by $\lambda_n = (n+1)\lambda,\ n \geq 0,\ \mu_n = n\mu,$ $n \geq 1\ (\lambda, \mu > 0)$. See Conolly, B.W., (1974) and Brian Conolly (1975).

We shall obtain, $p_{in}(t) = p_n(t) = P\{N(t) = n\ /\ N(0) = i\}$ which is the probability that there are n customers in the system at time t given there are i customers in the system initially, in a general birth-and-death single server queueing process with parameters. λ_n and μ_n defined above and obtain the queue characteristics.

The differential - difference equations of $p_n(t)$ are:

$$p_o'(t) = -\lambda\, p_o(t) + \mu\, p_1(t) \tag{2.5.1}$$

$$p_n'(t) = -[(n+1)\lambda + n\mu]\, p_n(t) + n\lambda\, p_{n-1}(t) + (n+1)\mu\, p_{n+1}(t) \tag{2.5.2}$$

$$G(z,t) = \sum_0^\infty p_n(t)\, z^n \text{ be PGF of } p_n(t). \tag{2.5.3}$$

Now $$\frac{\partial}{\partial z} G(z,t) = \sum_1^\infty n\, p_n(t)\ z^{n-1}\ \&\ \frac{\partial}{\partial t} G(z,t) = \sum_{n=0}^\infty p_n{}'(t)\ z^n \tag{2.5.4}$$

Hence

$$\frac{\partial}{\partial t} G(z,t) - [-\lambda p_0(t) + \mu p_1(t)] = \sum_1^\infty p_n'(t) z^n$$

$$= -\lambda \sum_1^\infty (n+1)\ p_n(t) z^n - \mu \Sigma\ w\, p_n(t) z^n + \lambda \sum_1^\infty n\, p_{n-1}(t) z^n$$

$$+ \mu \sum_1^\infty (n+1)\ p_{n+1}(t) z^n$$

$$\frac{\partial}{\partial t}G(z,t) = -\lambda \sum_{0}^{\infty} (n+1)\, p_n(t)\, z^n - \mu z \Sigma n\, p_n(t)\, z^{n-1}$$

$$+ \lambda z^2 \sum_{1}^{\infty} n\, p_{n-1}(t)\, z^{n-2} + \mu \sum_{0}^{\infty} (n+1)\, p_{n+1}(t)\, z^n$$

$$= -\lambda z \Sigma\, p_n(t).n\, z^{n-1} - \lambda \sum_{0}^{\infty} p_n(t)\, z^n$$

$$- \mu z \sum_{1}^{\infty} n\, p_n(t)\, z^{n-1} + \lambda z^2 \sum_{m=1}^{\infty} p_m(t)\, m z^{m-1}$$

$$+ \lambda z \sum_{m=0}^{\infty} p_m(t)\, z^m + \mu \sum_{1}^{\infty} p_n(t)\, n\, z^{n-1}$$

$$= -\lambda z \frac{\partial}{\partial z}G(z,t) - \lambda\, G(z,t) - \mu z \frac{\partial}{\partial z}G(z,t) + \lambda z^2 \frac{\partial}{\partial z}G(z,t)$$

$$+ \lambda z^2 \frac{\partial}{\partial z}G(z,t) + \lambda z\, G(z,t) + \mu \frac{\partial}{\partial z}G(z,t)$$

i.e., $$\frac{\partial}{\partial t}G(z,t) + (1-z)\,(\lambda z - \mu)\frac{\partial}{\partial z}G(z,t) = \lambda(z-1)G(z,t) \qquad (2.5.5)$$

which is the partial differential equation satisfied by $G(z, t)$ that is to be solved using the condition that $G(z,0) = z^i$ (There are i customers in the system at time $t = 0$).

The related equations for (2.5.5) are

$$\frac{dt}{1} = \frac{dz}{(1-z)\,(\lambda z - \mu)} = \frac{dG}{\lambda(z-1)G}$$

$$\frac{dt}{1} = \frac{dz}{(1-z)(\lambda z - \mu)} \text{ gives } \frac{dz}{dt} = (1-z)(\lambda z - \mu)$$

which yields $u(t,z,G) = \dfrac{1-z}{\lambda z - \mu} e^{(\lambda - \mu)} = C_1$ (a constant).

$$\frac{dz}{(1-z)\,(\lambda z-\mu)}=\frac{dG}{\lambda(z-1)G} \text{ gives } \frac{dG}{dz}=\frac{-\lambda G}{\lambda z-\mu} \text{ which yields } V\ (t,\ z,\ G) =$$

$(\lambda z-\mu)G = C_2$ (a constant).

The general solution of (2.5.5) is given by

$$V= g\ (u)$$

ie,
$$G(z,t)=\frac{1}{\lambda z-\mu}\ g\left\{\left(\frac{1-z}{\lambda z-\mu}\right)e^{(\lambda-\mu)t}\right\} \qquad (2.5.6)$$

$$G(z,o)=\frac{1}{\lambda z-\mu}\ g\left\{\frac{1-z}{\lambda z-\mu}\right\}$$

$$z^{i}=\frac{1}{\lambda z-\mu}\ g\left\{\frac{1-z}{\lambda z-\mu}\right\}$$

with
$$\frac{1-z}{\lambda z-\mu}=y,\ z=\left(\frac{1+\mu y}{1+\lambda y}\right)$$

Hence,

$$g(y)=\frac{\lambda-\mu}{1+\lambda y}\left(\frac{1+\mu y}{1+\lambda y}\right)^{i}$$

We need $g\left\{\left(\frac{1-z}{\lambda z-\mu}\right)e^{(\lambda-\mu)t}\right\}$

Hence from (2.5.6),

$$G(z,\ t)=\frac{1}{\lambda z-\mu}\left[\frac{\lambda-\mu}{1+\lambda\left(\frac{1-z}{\lambda z-\mu}\right)e^{(\lambda-\mu)t}}\right]\left[\frac{1+\mu\left(\frac{1-z}{\lambda z-\mu}\right)e^{(\lambda-\mu)t}}{1+\lambda\left(\frac{1-z}{\lambda z-\mu}\right)e^{(\lambda-\mu)t}}\right]$$

$$=\left[\frac{\lambda-\mu}{\lambda z-\mu+\lambda(1-z)\,e^{(\lambda-\mu)t}}\right]\left[\frac{\lambda z-\mu+\mu(1-z)\,e^{(\lambda-\mu)t}}{\lambda z-\mu+\lambda(1-z)\,e^{(\lambda-\mu)t}}\right]^{i} \qquad (2.5.7)$$

Using (2.5.7), we may determine the mean (expected value) and variance of the number of customers in the system at time t as under:

$$E\{N(t)\} = \frac{\partial}{\partial z} G(z,t)\Big]_{z=1} (= G'(1))$$

$$= i\, e^{(\lambda-\mu)t} - \frac{(\lambda - \lambda e^{-(\mu-\lambda)t})}{\lambda-\mu} \tag{2.5.8}$$

which is the average number of customers in the system at time t.

$$G''(1) = \frac{-3i\, e^{-(\mu-\lambda)t}(\lambda - \lambda e^{-(\mu-\lambda)t})}{\lambda-\mu} + i(i-1)\, e^{-2(\mu-\lambda)t}$$

$$+ \frac{2(\lambda - \lambda e^{-(\mu-\lambda)t})}{(\lambda-\mu)^2} + \frac{i\,[\lambda - \lambda e^{-(\mu-\lambda)t}]}{\lambda-\mu}\, e^{-(\mu-\lambda)t}$$

Hence,

$$\mathrm{Var}\,(N(t)) = G''(1) + G'(1) - [G'(1)]^2$$

$$= \frac{-3ie^{-(\mu-\lambda)t}(\lambda - \lambda e^{-(\mu-\lambda)t})}{\lambda-\mu} + i(i-1)e^{-2(\mu-\lambda)t} + \frac{2(\lambda - \lambda e^{-(\mu-\lambda)t}}{(\lambda-\mu)^2}$$

$$+ \frac{i[\lambda - \lambda e^{-(\mu-\lambda)t}]}{\lambda-\mu} e^{-(\mu-\lambda)t} - \left\{ \frac{ie^{-(\mu-\lambda)t} - (\lambda - \lambda e^{-(\mu-\lambda)t})}{\lambda-\mu} \right\} \tag{2.5.9}$$

Using (2.5.7), we shall obtain $p_n(t)$ as under:

$$p_n(t) = \frac{1}{n!} \frac{\partial^n G(z,t)}{\partial z^n}\Bigg]_{z=0} \tag{2.5.10}$$

$$= \frac{1}{n!} \frac{\partial^n}{\partial z^n} \left\{ \left[\frac{(\lambda-\mu)}{\lambda z - \mu + \lambda(1-z)e^{-(\mu-\lambda)t}} \right] \left[\frac{\lambda z - \mu + \mu(1-z)e^{-(\mu-\lambda)t}}{\lambda z - \mu + \lambda(1-z)e^{-(\mu-\lambda)t}} \right] \right\}_{z=0}$$

$$
= \frac{(\lambda-\mu)}{n!}\left\{\begin{array}{l}
\left[\dfrac{\lambda z-\mu+\mu(1-z)e^{-(\mu-\lambda)t}}{\lambda z-\mu+\lambda(1-z)e^{-(\mu-\lambda)t}}\right]^{i}\dfrac{(-1)^{n}\,n!\,[\lambda-\lambda e^{-(\mu-\lambda)t}]^{n}}{[\lambda z-\mu+\lambda(1-z)e^{-(\mu-\lambda)t}]^{n+1}} \\
+nC_1\left[\begin{array}{l} i\left(\dfrac{\lambda z-\mu+\mu(1-z)e^{-(\mu-\lambda)t}}{\lambda z-\mu+\lambda(1-z)e^{-(\mu-\lambda)t}}\right)^{i-1}(\lambda-\mu)^2 e^{-(\mu-\lambda)t} \\ (-1)^{n-1}\dfrac{(n-1)!\left(\lambda-\lambda e^{-(\mu-\lambda)t}\right)^{n-1}}{\left(\lambda z-\mu+\lambda(1-z)e^{-(\mu-\lambda)t}\right)^{n}} \end{array}\right]+\ldots \\
+\dfrac{1}{\left(\lambda z-\mu+\lambda(1-z)e^{-(\mu-\lambda)t}\right)}(-1)^{n+1}(iC_n)(n!)^2(\lambda-\mu)^{2n}e^{-(\mu-\lambda)t} \\
\left[\dfrac{[\lambda z-\mu+\mu(1-z)e^{-(\mu-\lambda)t}]}{[\lambda z-\mu+\lambda(1-z)e^{-(\mu-\lambda)t}]}\right]^{i-n}
\end{array}\right\}_{z=0}
$$

$$
= \frac{(\lambda-\mu)}{n!}\left\{\left(\frac{-\mu+\mu e^{-(\mu-\lambda)t}}{-\mu+\lambda e^{-(\mu-\lambda)t}}\right)^{i}(-1)^{n}n!\frac{[\lambda-\lambda e^{-(\mu-\lambda)t}]^{n}}{[-\mu+\lambda e^{-(\mu-\lambda)t}]^{n+1}}\right.
$$

$$
+n\left[i(\lambda-\mu)^2 e^{-(\mu-\lambda)t}\left(\frac{-\mu+\mu e^{-(\mu-\lambda)t}}{-\mu+\lambda e^{-(\mu-\lambda)t}}\right)^{i-1}(-1)^{n-1}(n-1)!\frac{[\lambda-\lambda e^{-(\mu-\lambda)t}]^{n-1}}{[-\mu+\lambda e^{-(\mu-\lambda)t}]^{n}}\right]+\ldots
$$

$$
\left.+\frac{1}{[-\mu+\lambda e^{-(\mu-\lambda)t}]}(-1)^{n+1}(iC_n)(n!)^2(\lambda-\mu)^{2n}e^{-(\mu-\lambda)t}\left(\frac{-\mu+\mu e^{-(\mu-\lambda)t}}{-\mu+\lambda e^{-(\mu-\lambda)t}}\right)^{i-n}\right\}
$$

i.e.,
$$
p_n(t)=\frac{(\lambda-\mu)}{n!}\left\{(-1)n!\rho^{n}\left(\frac{1-e^{-(\mu-\lambda)t}}{1-\rho e^{-(\mu-\lambda)t}}\right)^{i}\frac{(1-e^{-(\mu-\lambda)t})^{n}}{\mu[1-\rho e^{-(\mu-\lambda)t}]^{n+1}}+\right.
$$

$$
n\left[(i)(-1)(n-1)!(\lambda-\mu)^2 e^{-(\mu-\lambda)t}\left(\frac{1-e^{-(\mu-\lambda)t}}{1-\rho e^{-(\mu-\lambda)t}}\right)^{i-1}\left(\frac{\rho^{n}}{\lambda}\right)\right]+\ldots
$$

$$
\left.+(iC_n)(n!)^2(-1)^{n-1}(\lambda-\mu)^{2n}e^{-n(\mu-\lambda)t}\left(\frac{1-e^{-(\mu-\lambda)t}}{1-\rho e^{-(\mu-\lambda)t}}\right)^{i-n}\right\} \qquad (2.5.11)
$$

Now $p_n = \lim_{t\to\infty} p_n(t)$

$$= \frac{(\lambda-\mu)^n}{n!}(-1)^n n! \frac{\lambda^n}{(-\mu)^{n+1}}$$

$$= (1-\rho)\,\rho^n,\ n \geq 0 \text{ where } \rho = \frac{\lambda}{\mu} \tag{2.5.12}$$

which is also steady state probability distribution of queue length for M/M/1 model where arrival rate λ_n and service rate μ_n are independent of n, the number present in the system already.

It may be proved that system size, $L_s = E(N) = \frac{\rho}{1-\rho}$ and $\text{var}(N) = \frac{\rho}{(1-\rho)^2}$

From (2.5.8), as $t \to \infty$, $E(N(t)) \to E(N) = \frac{1}{1-\rho}$ and from (2.5.9), as $t \to \infty$, $\text{Var}\{N(t)\} \to \text{Var}\{N\} = \frac{1}{(1-\rho)^2}$

We may also obtain p_n using the Birth-death methodology as under:

$$p_n = \frac{\lambda_0\lambda_1\ldots\lambda_{n-1}}{\mu_1\mu_2\ldots\mu_n}p_0 = \frac{\lambda 2\lambda\ldots n\lambda}{\mu 2\mu\ldots n\mu} = \rho^n p_0,\ \rho = \frac{\lambda}{\mu}$$

$\sum p_n = 1$ gives $p_0 = (1-\rho)$

Hence,

$$p_n = (1-\rho)\rho^n,\ n \geq 0 \text{ and } 0 < \rho < 1) \tag{2.5.13}$$

which agrees with the result obtained in (2.5.11) while $t \to \infty$.

λ^* = average number of arrivals who actually enters in the system (effective arrival rate).

$$= \sum \lambda_n p_n = \sum (n+1)\lambda(1-\rho)\rho^n = \frac{\lambda}{1-\rho} \qquad (2.5.14)$$

Effective service rate, μ^* (weighted average service time) is given by

$$\mu^*(1-p_0) = \sum_1^{\infty} \mu_n p_n = \sum n\mu(1-\rho)\rho^n = \mu\rho$$

$$\mu^* = \frac{\mu}{(1-\rho)}$$

or

$$\mu^* = \frac{\lambda^*}{(1-p_0)} = \frac{\mu}{1-\rho}$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \rho = 1 - p_0, \qquad (2.5.15)$$

which is the mean number of customers at the server. It is also the fraction of time the server is working.

Expected waiting time, W_S, of an arrival in the steady state is given by

$$W_S = \frac{L_s}{\lambda^*} = \frac{1}{\mu}$$

$$\mu W_S = 1 \qquad (2.5.16)$$

i.e., the number of customers served during the waiting time of an arbitrary customer equals one for $\rho < 1$ and is independent of λ.

Now we evaluate numerically the queue characteristics of the above four state-dependence Markovian queueing Models for various $\rho = \frac{\lambda}{\mu}$ $(0 < \rho < 1)$ and the results tabulated in Tables I, II, III & IV. The performance measures are compared numerically thereby suggesting the relative supremacy of the Models. It is to be noted that the stationary distribution for the queueing Models exists only for ρ which lies in $0 < \rho < 1$.

Table I Numerical Evaluation of the queue characteristics of the Model I in (2.2) for various $\rho = \frac{\lambda}{\mu}$ where $0 < \rho < 1$

ρ	p_0	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9048	0.1	1	λ	1.0504 μ	0.0952
0.2	0.8187	0.2	1	λ	1.1031 μ	0.1813
0.5	0.6065	0.5	1	λ	1.2706 μ	0.3935
0.9	0.4066	0.9	1	λ	1.5167 μ	0.5934
0.95	0.3867	0.95	1	λ	1.5490 μ	0.6133
0.99	0.3716	0.99	1	λ	1.5754 μ	0.6284

Table II Numerical Evaluation of the queue characteristics of the Model II in (2.3) for various $\rho = \frac{\lambda}{\mu}$ where $0 < \rho < 1$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9048	0.1	1.0504	0.0952λ	μ	0.0952
0.2	0.8187	0.2	1.1031	0.1813λ	μ	0.1813
0.5	0.6065	0.5	1.2706	0.3935λ	μ	0.3935
0.9	0.4066	0.9	1.5167	0.5934λ	μ	0.5934
0.95	0.3867	0.95	1.5490	0.6133λ	μ	0.6133
0.99	0.3716	0.99	1.5754	0.6284λ	μ	0.6284

Table III Numerical Evaluation of the queue characteristics of the Model III in (2.4) for various $\rho = \frac{\lambda}{\mu}$ where $0 < \rho < 1$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9076	0.0950	1	0.9498 λ	1.0278μ	0.0924
0.2	0.8277	0.1816	1	0.9080 λ	1.0543μ	0.1723
0.5	0.6400	0.4052	1	0.8104 λ	1.1254μ	0.3600
0.9	0.4699	0.6456	1	0.7174 λ	1.2179μ	0.5301
0.95	0.4541	0.6720	1	0.7073λ	1.2310μ	0.5458
0.99	0.4417	0.6926	1	0.6996 λ	1.2405μ	0.5583

Table IV Numerical Evaluation of the queue characteristics of the Model IV in (2.5) for various $\rho = \frac{\lambda}{\mu}$ where $0 < \rho < 1$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9	0.1111	1	1.1111λ	1.1111μ	0.1
0.2	0.8	0.25	1	1.250λ	1.250μ	0.2
0.5	0.5	1.00	1	2.00λ	2.00μ	0.5
0.9	0.1	9.00	1	10.00λ	10.00μ	0.9
0.95	0.05	19.00	1	20.00λ	20.00μ	0.95
0.99	0.01	99.00	1	100.00λ	100.00μ	0.99

Now we compare numerically the performance measures of the state dependent models, Model I, II, III & IV with the standard M/M/1 model and for comparison purpose we compute the corresponding performance measures of the standard M/M/1 model for various $\rho = \frac{\lambda}{\mu}$ $(0 < \rho < 1)$ and the results tabulated in Table V.

Table V Numerical Evaluation of the queue characteristics of the M/M/1 Model for various ρ ($0 < \rho < 1$)

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9	0.1111	1.111	λ	μ	0.1
0.2	0.8	0.25	1.250	λ	μ	0.2
0.5	0.5	1.00	2.000	λ	μ	0.5
0.9	0.1	9.00	10.000	λ	μ	0.9
0.95	0.05	19.00	20.000	λ	μ	0.95
0.99	0.01	99.00	100.000	λ	μ	0.99

2.6 TRANSIENT/STEADY-STATE ANALYSIS OF STATE-DEPENDENT QUEUEING MODELS

In this section we shall compare the transient probability distribution of queue length of the adaptive queueing models considered in (2.5) and the standard trivially state-dependent M/M/1 Model. We prove that the transient probability of queue length are vastly different even though the steady-state probability of queue length are same.

The transient probability distribution of queue length for the Model (2.5) characterized by the arrival rate $\lambda_n = (n + 1)\lambda$ $n \geq 0$; service rate $\mu_n = n\,\mu$, $n \geq 1$, where $\lambda > 0$, $\mu > 0$ is given by:

$$p_n(t) = (1-\rho)\left\{\frac{\rho^n(1-e^{-(\mu-\lambda)t})^n}{[1-\rho e^{-(\mu-\lambda)t}]^{n+1}}\right\} \tag{2.6.1}$$

(See equation (2.5.11) where initially there are no customers in the system).

The transient probability distribution of queue length for the standard M/M/1 Model characterized by $\lambda_n = \lambda \;\; n \geq 0$; $\mu_n = \mu$, $n \geq 1$, where $\lambda > 0$, $\mu > 0$ is given by:

$$p_n(t) = e^{-(\lambda+\mu)t}\left\{\rho^{n/2} I_n 2\sqrt{\lambda\mu}\, t + \rho^{\frac{n-1}{2}} I_{n+1} 2\sqrt{\lambda\mu}\, t + (1-\rho)\rho^n \sum_{j=n+2}^{\infty} \rho^{-\frac{j}{2}} I_j (2\sqrt{\lambda\mu}\, t)\right\} \quad (2.6.2)$$

for all $n \geq 0$ and initially there are no customers in the system. (See for example Gross and Harris (1998)). Here $I_n(y) = \sum_{k=0}^{\infty} \frac{\left(\frac{y}{2}\right)^{n+2k}}{k!(n+k)!}$ is the infinite series for the Modified Bessel function of the first kind of order $n = 0, 1, 2, \ldots$

When $t \to \infty$ and $\rho = \frac{\lambda}{\mu} < 1$, the steady-state probability distribution of queue length for the Model considered in (2.5) is given by

$$p_n = \lim_{t\to\infty} p_n(t) = (1-\rho)\rho^n, \; n \geq 0 \text{ and } \rho < 1 \quad (2.6.3)$$

which is a Geometric distribution with parameter $\rho < 1$.

From (2.6.2), the steady-state probability distribution of queue length for the standard M/M/1 Model is given by:

$$p_n = \lim_{t\to\infty} p_n(t) = (1-\rho)\rho^n, \; n \geq 0 \text{ and } \rho < 1$$

which is also a Geometric distribution as in (2.6.3).

We now evaluate numerically the transient probabilities of queue length for the two Models for $\lambda = 2$ and $\mu = 4$ so that $\rho = \frac{\lambda}{\mu} < 1$ for various times. We present below a typical contrasting situation.

For the state-dependent Model considered in section 2.5, the transient probabilities $p_n(t)$ for $t = 1$ are, $p_0(1) = 0.5363$; $p_1(1) = 0.2487$; $p_2(1) = 0.1168$; $p_3(1) = 0.0535$ and so on...

Transient probabilities for $t = 2$ are, $p_0(2) = 0.5046$; $p_1(2) = 0.2500$; $p_2(2) = 0.1239$ and so on.. The steady-state probabilities, p_n for the Model in (2.5) are $p_0 = 0.5$; $p_1 = 0.25$; $p_2 = 0.125$; $p_3 = 0.0625$ and so on...

For the M/M/1 Model, the transient probabilities $p_n(t)$ for $t = 1$ are, $p_0(1) = 0.1408$; $p_1(1) = 0.0435$; $p_2(1) = 0.0011$; and so on...

The steady-state probabilities, p_n for the M/M/1 Model are, $p_0 = 0.5$; $p_1 = 0.25$; $p_2 = 0.125$; $p_3 = 0.0625$ and so on...

By the above transient and steady-state analysis for the Models in (2.5) and the standard M/M/1 Model, it is verified that the transient probabilities of queue length are vastly different even though the steady-state probabilities of queue length are same. This clearly indicates the importance of transient measures in adaptive queueing models.

2.7 SUMMARY OF RESULTS AND CONCLUDING REMARKS

1. Model II considered in (2.3) is exactly same as Model I consider in (2.2) with respect to p_n, L_s & ρ^* for a given traffic intensity ρ.
2. μW_S, the number of customers served during the waiting time of an arbitrary arrival for Model I is less than that of Model II, even though both state-dependent Queueing Models have the same probability distribution of queue length.
3. With reference to the performance measure p_o, the state probability of an empty system, Model III is superior to Model I or Model II.
4. With reference to the performance measure p_o, the state probability of an empty system, it is observed that the state-dependant Queueing Models I, II and III are always superior to the standard M/M/1 Model or equivalently Model IV.
5. ρ^*, the effective traffic intensity (i.e., the proportion of time the server is working) of 0.99, extremely heavy, for M/M/1, becomes moderate at 0.5583 for Model III when $\rho \to 1$.
6. By comparing with the standard M/M/1 Model, Models I & II are superior to an increasing extent as $\rho \to 1$, and Model III is better even than these, when considering the equilibrium system state.
7. The steady-state probability distribution of queue length of the state-dependent queue Model IV is exactly same as that of the standard M/M/1 Model and is given by the Geometric law, $p_n = (1-\rho)\rho^n, n \geq 0$ and $\rho = \frac{\lambda}{\mu} < 1$.

8. From the study of transient and steady-state analysis, it is proved that the transient probability distribution of queue length, $p_n(t)$, are vastly different even though the steady-state probabilities p_n are the same for the state-dependent queue Model considered in (2.5) and the standard M/M/1 Model.

9. For Model III, it is observed that $\rho^* = 1 - \frac{1}{I_0(2\sqrt{\rho})}$ is nearly ρ when ρ is small.

10. For Model IV, ρ and ρ^* are equal which distinguishes it from the other state-dependent queueing Models considered in this chapter.

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

Chapter 3

PERFORMANCE MEASURES OF STATE-DEPENDENT MARKOVIAN QUEUEING MODELS

3.1. INTRODUCTION

State-dependent queueing models has been discussed far and widely in the literature since the work of Cox and Smith (1961). In the more general context of birth and death processes they have been reviewed and analyzed, for example, by Jackson (1963), Keilson (1964), Harris (1967), Hadidi and B, W. Conolly (1969b), Natvig (1974) and Conolly (1975).

In analyzing various state-dependent queueing models, one would like to compute certain performance measures, such as transient queue length distribution, the steady-state queue length distribution, mean number of customers, average number of customers in the system and sojourn time distribution. For a specified model, these quantities may all be characterized as solutions to certain equations, called differential-difference equations. Thus, computing the performance measures amounts to solving these equations together with appropriate initial conditions, see for example Conolly, B.W., (1975); Donald Gross and Carl. M. Harris (1998) and Keilson, J. (1964, 1965).

In this chapter we obtain probability distribution of queue length (transient or steady state) for some state dependent queueing models using PGF or a combination of PGF and Laplace Transformation. Using the probability distribution of queue length some performance measures of the Models are investigated and compared numerically thereby suggesting the relative supremacy of the Models.

3.2 LINEAR STATE-DEPENDENT MODEL CHARACTERIZED BY THE ARRIVAL RATE $\lambda_n = \lambda$, $n \geq 0$; SERVICE RATE $\mu_n = 2n\mu$, $n \geq 1$ $(\lambda > 0, \mu > 0)$

This is a model with constant arrival rate and abundant service time. $N(t)$ denote the number of customers present at time t and initially there are i customers in the system.

Let $p_n(t)$ or $p_{in}(t) = P\{N(t) = n/N(0) = i\}$ denote the probability that there are n customers in the system at time t where initially there are i customers in the system.

The differential–difference equations of $p_n(t)$ are:

$$P_o'(t) = -\lambda p_0(t) + 2\mu p_1(t) \tag{3.2.1}$$

$$p_n'(t) = -[\lambda + 2n\mu]p_n(t) + \lambda p_{n-1}(t) + (2n+2)\mu p_{n+1}(t) \quad , n \geq 1 \tag{3.2.2}$$

Define the PGF of $p_n(t)$, $G(z, t) = \sum_0^\infty p_n(t) z^n$ (3.2.3)

Now,

$$\frac{\partial}{\partial z}G(z,t) = \sum_{1}^{\infty} p_n(t)nz^{n-1} \quad \& \quad \frac{\partial}{\partial t}G(z,t) = \sum_{0}^{\infty} p_n{}'(t)z^n \qquad (3.2.4)$$

Hence,

$$\frac{\partial}{\partial t}G(z,t) - [-\lambda p_o(t) + 2\mu p_1(t)] = -\lambda \sum_{1}^{\infty} p_n(t)z^n - 2\mu \sum_{1}^{\infty} np_n(t)z^n$$

$$+ \lambda \sum_{1}^{\infty} p_{n-1}(t)z^n + 2\mu \sum_{1}^{\infty}(n+1)p_{n+1}(t)z^n$$

Using (3.2.1), (3.2.2) & (3.2.3) suitably in (3.2.4).

$$\frac{\partial}{\partial t}G(z,t) = -\lambda G(z,t) - 2\mu z\frac{\partial}{\partial z}G(z,t) + \lambda zG(z,t) + 2\mu \frac{\partial}{\partial z}G(z,t)$$

i.e.

$$\frac{\partial}{\partial t}G(z,t) + 2\mu(z-1)\frac{\partial}{\partial z}G(z,\mathrm{t}) = \lambda(z-1)G(z,t) \qquad (3.2.5)$$

which is the PDE satisfied by $G(z, t)$ that is to be solved using the initial condition $G(z, 0) = z^i$.

The related equations of (3.2.5) are

$$\frac{dt}{1} = \frac{dz}{2\mu(z-1)} = \frac{dG}{\lambda(z-1)G} \qquad (3.2.6)$$

From $\frac{dt}{1} = \frac{dz}{2\mu(z-1)}$ yield the solution

$u(t, z, G) = (z-1)e^{-2\mu t} = C_1$ (a constant)

and $\frac{dz}{2\mu(z-1)} = \frac{dG}{\lambda(z-1)G}$ yield the solutions

$$V(t, z, G) = G(z, t)\, e^{\frac{-\rho}{2}z} = C_2 \text{ (a constant)}, \quad \rho = \frac{\lambda}{\mu}$$

Then the general solution of (3.2.5) is given by

$$G(z, t)\, e^{\frac{-\rho}{2}z} = g\{(z-1)e^{-2\mu t}\} \tag{3.2.7}$$

$$G(z, 0) = e^{\frac{\rho}{2}z}\, g\{(z-1)\}$$

$$z^i = e^{\frac{\rho}{2}z}\, g(z-1)$$

with $z - 1 = y$, $z = 1 + y$

$$g(y) = e^{-\rho/2(1+y)}\,(1+y)^i \tag{3.2.8}$$

Hence from (3.2.7),

$$G(z, t) = e^{\frac{\rho}{2}z} \,.\, e^{\frac{\rho}{2}[1+(z-1)e^{-2\mu t}} \,.\, [1 + (z-1)e^{-2\mu t}]^i$$

$$= e^{\frac{\rho}{2}z(1-\bar{e}^{2\mu t})} \,.\, e^{\frac{\rho}{2}(1-\bar{e}^{2\mu t})} \,.\, [1 + (z-1)\,\bar{e}^{2\mu t}]^i \tag{3.2.9}$$

Now

$$p_n(t) = \frac{1}{n!}\frac{\partial^n}{\partial z^n} G(z,t)\ ]_{z=0} \tag{3.2.10}$$

$$= \frac{e^{\frac{-\rho}{2}}(1-e^{-2\mu})}{n!}\left\{[1+(z-1)e^{-2\mu t}\left(\frac{\rho}{2}\right)^n (1-e^{-2\mu t})^n e^{\frac{\rho}{2}z(1-e^{-2\mu t})}\right\}$$

$$+ nC_1\left[\ i(1+(z-1)\ \bar{e}^{2\mu t})^{i-1}\ \bar{e}^{2\mu t} \,.\, \left(\frac{\rho}{2}\right)^{n-1} (1-\bar{e}^{2\mu t})^{n-1} e^{\frac{\rho}{2}z(1-\bar{e}^{2\mu t})}\ \right]$$

$$+ \ldots + (iC_n)n!\,[1 + (z-1)\,e^{-2\mu t}]^{i-n} \cdot (e^{-2\mu t})^n\, e^{\frac{\rho}{2}z(1-\bar{e}^{2\mu t})} \Big\}_{z=0}$$

$$= \frac{e^{\frac{-\rho}{2}(1-e^{-2\mu t})}}{n!} \left\{ \left(\frac{\rho}{2}\right)^n (1-\bar{e}^{2\mu t})^n (1-e^{-2\mu t})^i + n\left[ie^{-2\mu t}(1-e^{-2\mu t})^{i-1} \right.\right.$$

$$\left.\left. \left(\frac{\rho}{2}\right)^{n-1} (1-e^{-2\mu t})^{n-1} \right] + \ldots\ldots + (iC_n)\, n!\ \bar{e}^{2\mu t})^n\ (1-\bar{e}^{2\mu t})^{i-n} \right\} \tag{3.2.11}$$

(using Leibnitz's theorem for the n^{th} derivative of a product).

Hence

$$p_n = \lim_{t\to\infty} p_n(t) = \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n!}, \quad n \geq 0 \tag{3.2.12}$$

which is a Poisson distribution with parameter $\frac{\rho}{2}$.

System size at any time,

$$L_S = E(N) = \frac{\rho}{2} \tag{3.2.13}$$

Effective arrival rate,

$$\lambda^* = \sum \lambda_n p_n = \sum \lambda p_n = \lambda \tag{3.2.14}$$

By Little's formula, expected waiting time of an arrival in the steady state is given by

$$W_S = \frac{L_s}{\lambda^*} = \frac{1}{2\mu} \tag{3.2.15}$$

Hence, μW_S, which is the number of customers served during the waiting time of an arbitrary arrivals equals ½ for any ρ

Effective service rate, μ^* is given by

$$\mu^*(1-p_o) = \sum_{1}^{\infty} n\mu p_n = \lambda \text{, yielding } \mu^* = \frac{\lambda}{1-p_o} \qquad (3.2.16)$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = 1 - \bar{e}^{\rho/2} = 1-p_o \qquad (3.2.17)$$

which is the number of customers at the server during his service time. It is also the proportion of time the server is working.

Table I: The Numerical evaluation of the performance measures of the Model in (3.2) for various $\rho = \dfrac{\lambda}{\mu}$

ρ	p_0	L_s	μW_s	λ^*	μ^*	ρ^*
0.1	0.9512	0.05	0.5	λ	(20.4918) λ	0.0488
0.2	0.9048	0.10	0.5	λ	(10.5040) λ	0.0952
0.5	0.7788	0.25	0.5	λ	(4.5208) λ	0.2212
0.9	0.6376	0.45	0.5	λ	(2.7594) λ	0.3624
0.99	0.6096	0.495	0.5	λ	(2.5615) λ	0.3904
1.00	0.6065	0.500	0.5	λ	2.5413 λ	0.3935

3.3 STATE-DEPENDENT MODEL CHARACTERIZED BY THE ARRIVAL RATE $\lambda_n = \dfrac{\lambda}{2(n+1)}$, $n \geq 0$; SERVICE RATE $\mu_n = \mu$, $n \geq 1$ $(\lambda > 0 , \mu > 0)$

This is a queue model where potential customers are discouraged by queue length. In this model the sight of long queue discourages fresh customers join it. For the model we shall find p_n, the steady state probability distribution of queue length using the Birth-death methodology and obtain the queue characteristics as under. The differential-difference equations for p_n are,

$$0 = -\frac{\lambda}{2} p_o + \mu p_1 \tag{3.3.1}$$

$$0 = -\left[\frac{\lambda}{2(n+1)} + \mu\right] p_n + \frac{\lambda}{2n} p_{n-1} + \mu p_{n+1}, \quad n \geq 1 \tag{3.3.2}$$

Solving these equations recursively, we obtain the probability distribution queue length in the steady-state as

$$p_n = \frac{\left(\frac{\rho}{2}\right)^n p_0}{n!}, \quad \rho = \frac{\lambda}{\mu} \tag{3.3.3}$$

Using $\qquad \sum p_n = 1$ gives $p_o = e^{-\rho/2}$ (3.3.4)

Hence

$$p_n = \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n!}, \quad n \geq 0 \tag{3.3.5}$$

which is again a Poisson distribution with parameter $\frac{\rho}{2}$.

System size at any time, $L_s = \sum np_n = \frac{\rho}{2}$ (3.3.6)

Effective arrival rate, $\lambda^* = \sum \lambda_n p_n$

$$= \sum \frac{\lambda}{2n+2} \cdot \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n!} = \mu\left(1-e^{-\rho/2}\right) \qquad (3.3.7)$$

Using Little's formula, expected waiting time of an arrival in the steady state is given by

$$Ws = \frac{L_s}{\lambda^*} = \frac{\rho}{2\mu(1-e^{-\rho/2})}$$

Hence

$$\mu\, W_S = \frac{\rho}{2(1-e^{-\rho/2})} \qquad (3.3.8)$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system which is a value greater than one and increases with the increasing values of ρ.

Proof:

$$\mu W_s = \frac{\rho}{2(1-e^{-\rho/2})}$$

$$= \left\{1-\left(\frac{\rho}{4}-\frac{\rho^2}{24}+....\right)\right\}^{-1}$$

$$= 1+\left(\frac{\rho}{4}-\frac{\rho^2}{24}+....\right)+\left(\frac{\rho}{4}-\frac{\rho^2}{24}+....\right)^2+...$$

$$= 1+\frac{\rho}{4}+\frac{\rho^2}{48} \text{ (approximately)}$$

Clearly this is a value greater than one and increases with increasing values of ρ.

The balance equation $\lambda_n p_n = \mu_{n+1} p_{n+1}$ of the M/M/1 model implies that $\sum_{n=0}^{\infty} \lambda_n p_n = \sum_{n=0}^{\infty} \mu_{n+1} p_{n+1} = \sum_{n=1}^{\infty} \mu_n p_n$, which may be deduced directly from the fact that the long-run average arrival rate must equal the long run average service rate. Hence by definition it follows that

$$\lambda^* = \mu^*(1 - p_0)$$

Hence effective service rate, $\mu^* = \dfrac{\lambda^*}{(1 - p_o)}$

$$= \mu \qquad (3.3.9)$$

Effective traffic intensity

$$\rho^* = \frac{\lambda^*}{\mu^*} = 1 - e^{-\rho/2} = 1 - p_0 \qquad (3.3.10)$$

which is the number of customers at the server during his service time. It is also the proportion of time the server is working.

Table II: The Numerical evaluation of the performance measures of the Model in (3.3) for various ρ

ρ	p_0	L_S	$\mu\ W_S$	λ^*	μ^*	ρ^*
0.1	0.9512	0.05	1.0246	$0.0488\,\mu$	μ	0.0488
0.2	0.9048	0.10	1.0569	$0.0952\,\mu$	μ	0.0952
0.5	0.7788	0.25	1.1302	$0.2212\,\mu$	μ	0.2212
0.9	0.6376	0.45	1.2417	$0.3624\,\mu$	μ	0.3624
0.99	0.6096	0.495	1.2679	$0.3904\,\mu$	μ	0.3904
1.0	0.6065	0.50	1.2706	$0.3935\,\mu$	μ	0.3935

3.4 STATE-DEPENDENT MODEL CHARACTERIZED BY ARRIVAL RATE $\lambda_n = \dfrac{\lambda}{2(n+1)}$, $n \geq 0$; SERVICE RATE $\mu_n = 2n\mu$, $n \geq 1$ $(\lambda > 0\ ,\ \mu > 0)$

This is a queue model which embodies both mechanisms of Models 3.2 and 3.3. For this model we shall obtain p_n, steady stare probability distribution of queue length, using the Birth-death methodology and obtain the queue characteristics as under

$$p_n = \frac{\left(\frac{\lambda}{2}\right)\cdot\left(\frac{\lambda}{z}\right)\cdots\left(\frac{\lambda}{2n}\right)p_0}{(2\mu)(4\mu)\ldots(2n\mu)} = \frac{\rho_n p_0}{2^{2n}(n!)^2}\ ,\quad \rho = \frac{\lambda}{\mu}$$

Using $$\sum p_n = 1 \text{ gives } p_0 = \frac{1}{I_0(\sqrt{\rho})} \qquad (3.4.1)$$

where $I_o(\sqrt{\rho})$ is the modified Bessel function of the first kind of order zero.

Hence

$$p_n = \frac{\left(\frac{\sqrt{\rho}}{2}\right)^{2n}}{(n!)^2 I_0(\sqrt{\rho})}, \quad n \geq 0 \tag{3.4.2}$$

System size at any time,

$$L_s = \sum n.p_n = \frac{\sqrt{\rho}/2}{I_0(\sqrt{\rho})} \sum_{k=0}^{\infty} \frac{(\sqrt{\rho}/2)^{2k+1}}{k!(k+1)!} = \frac{\sqrt{\rho}}{2} \frac{I_1(\sqrt{\rho})}{I_0(\sqrt{\rho})} \tag{3.4.3}$$

Effective arrival rate, $\lambda^* = \sum \lambda_n.p_n$

$$= \frac{1}{I_0(\sqrt{\rho})}\left(\frac{\lambda}{\sqrt{\rho}}\right) \sum_{k=0}^{\infty} \frac{(\rho/2)^{2k+1}}{k!(k+1)!}$$

$$= \frac{\lambda I_1(\sqrt{\rho})}{\sqrt{\rho} I_0(\sqrt{\rho})} \tag{3.4.4}$$

Expected waiting time of an arbitrary arrival in the queue is given by

$W_S = \frac{L_s}{\lambda^*} = \frac{1}{2\mu}$ (By Little's formula)

Hence,

$$\mu W_s = ½ \quad \text{for any } \rho \text{ in } 0 < \rho < 1 \tag{3.4.5}$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system.

Effective service rate, μ^* is given by

$$\mu^*(1 - p_0) = \sum_{1}^{\infty} 2n\mu . \frac{(\sqrt{\rho}/2)^{2n}}{I_0(\sqrt{\rho})(n!)^2} = \mu\sqrt{\rho} \frac{I_1(\sqrt{\rho})}{I_0(\sqrt{\rho})}$$

Hence,

$$\mu^* = \frac{\mu(\sqrt{\rho}).I_1(\sqrt{\rho})}{(1-p_0).I_0(\sqrt{\rho})} \tag{3.4.6}$$

Effective traffic intensity

$$\rho^* = \frac{\lambda^*}{\mu^*} = \frac{I_0\sqrt{\rho}-1}{I_0\sqrt{\rho}} = 1 - p_o \tag{3.4.7}$$

which is the number of customers at the server during his service time. It is also the proportion of time the server is working.

Table III: The Numerical evaluation of the performance measures of the Model in (3.4) for various ρ

ρ	p_O	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9748	0.0252	0.5	0.49987 λ	1.9868 μ	0.0252
0.2	0.9512	0.0491	0.5	0.4909 λ	2.0120 μ	0.0488
0.5	0.8849	0.1182	0.5	0.4729 λ	2.0539 μ	0.1151
0.9	0.8073	0.2035	0.5	0.4516 λ	2.1092 μ	0.1927
0.99	0.7933	0.2223	0.5	0.4469 λ	2.1226 μ	0.2084
1.0	0.7898	0.2222	0.5	0.4446 λ	2.1333 μ	0.2084

Remark: It is to be noted that we can evaluate numerically the two characteristics for the Models (3.2), (3.3) and (3.4) for any $\rho > 0$. But stationary condition exists only for ρ where $0 < \rho < 1$.

Now for comparison purpose we also compute the corresponding performance measures of the standard *M/M/*1 Model for various $\rho = \frac{\lambda}{\mu}$ $(0 < \rho < 1)$ and the results tabulated in Table IV.

Table IV : The Numerical evaluation of the performance measures of the standard *M/M/*1 model for various $\rho = \frac{\lambda}{\mu}$ in $0 < \rho < 1$

ρ	P_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9	0.1111	1.1111	λ	μ	0.1
0.2	0.8	0.2500	1.2500	λ	μ	0.2
0.5	0.5	1.0000	2.0000	λ	μ	0.5
0.9	0.1	9.0000	10.0000	λ	μ	0.9
0.99	0.01	99.0000	100.0000	λ	μ	0.99

3.5 SUMMARY OF RESULTS AND CONCLUDING REMARKS

1. The three state dependent models considered in this paper are better than M/M/1 model with reference to p_o, which is the state probability of an empty system.
2. For the standard M/M/1 model, which is a trivially state dependent Markovian queue model, $\mu W_S = L_S + 1$. i.e. the number of customers served per unit time during the waiting time of an arbitrary customer in the system equals one more than the system size.

3. For the state dependent Models in sections (3.2) & (3.3), the system size L_S is near to ρ^*, only for $\rho \leq 0.5$ whereas for model in section (3.4), L_S is almost equal to ρ^* uniformly for all ρ (see table III). So in this sense Model (3.4) is better than the other state dependent models considered.

4. In model (3.2), we have $\mu w_S = L_S$ when $\rho = 1$, which distinguishes the model from other state-dependent Queueing models considered. Also $\mu W_S = 2L_s$ when $\rho = 0.5$.

5. By comparing with the standard M/M/1 model, Models (3.2) & (3.3) are superior to an increasing extent as $\rho \to 1$, and Model (3.4) is even better than these, when considering the equilibrium system state.

6. In terms of the performance measures p_o, the state probability of an empty system one can do better than the Models (3.2) and (3.3). An example is provided by $p_0 = \dfrac{1}{I_0(\sqrt{\rho})}$, which can be achieved by means of the Model in (3.4) (see Table III).

7. The probability distribution of queue length in the steady-state of the models in (3.2) and (3.3) follows a Poisson distribution with parameter $\rho/2$. Thus there can be state-dependent Markovian queueing models having the same steady-state probability distribution of queue length.

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

Chapter 4

MODIFIED STATE DEPENDANT MARKOVIAN QUEUEING MODELS

4.1 INTRODUCTION

In this chapter we shall study several state-dependent Markovian queueing models, called Modified state-dependent Markovian queueing models, which are modified version of the Models considered in chapter 2. We shall obtain the transient and/or steady state probability distribution of queue length and obtain the queue characteristics. We use the Probability Generating Function (PGF) and Birth-Death methodology in the analysis. We also evaluate numerically the performance measures of these models and carryout a comparative study.

4.2 MODEL AI

A partially linear state dependent Markovian queueing Model characterized by $\lambda_n = \lambda$, $n \geq 0$, $\mu_n = 2\, n\mu$, $n \geq 1$ (λ & $\mu > 0$). This model is a self-service system like Model I of Chapter 2 and may be regarded as modified form of it.

$N(t)$ denote the number of customers present in the system at time t. We shall find $p_n(t)$ or $p_{in}(t) = P\{N(t) = n/N(0) = i\}$, which is the probability that there are n customers in the system at time t given that initially there are i customers, using the probability generating function (PGF) of $p_n(t)$ and obtain the queue characteristics as under:

The Differential - difference equations of $p_n(t)$ are:

$$p_0'(t) = -\lambda\, p_o(t) + 2\mu\, p_1(t) \tag{4.2.1}$$

$$p_n'(t) = -[\lambda + 2n\mu]\, p_n(t) + \lambda\, p_{n-1}(t) + (2n+2)\,\mu\, p_{n+1}(t)\ ,\ n \geq 1 \tag{4.2.2}$$

$$G(z,t) = \sum_0^\infty p_n(t)\, z^n \text{ be the PGF of } p_n(t) \tag{4.2.3}$$

Now
$$\frac{\partial}{\partial z} G(z,t) = \Sigma\, p_n(t)\, n\, z^{n-1} \ \&\ \frac{\partial}{\partial t} G(z,t) = \Sigma\, p_n'(t)\, z^n \tag{4.2.4}$$

$$\frac{\partial}{\partial t} G(z,t) - p_0'(t) = \sum_{n=1}^\infty p_n'(t)\, z^n$$

$$\frac{\partial}{\partial t} G(z,t) - [-\lambda\, p_o(t) + 2\mu\, p_1(t)] = -\lambda \sum_1^\infty p_n(t)\, z^n - 2\mu \sum_1^\infty n\, p_n(t)\, z^n + \lambda \sum_1^\infty p_{n-1}(t)\, z^n$$

$$+ 2\mu \sum_1^\infty (n+1)\, p_{n+1}(t)\, z^n$$

(Using (4.2.1), (4.2.2) & (4.2.3) suitably in (4.2.4))

$$\frac{\partial}{\partial t} G(z,t) = -\lambda G(z,t) - 2\mu z \frac{\partial}{\partial z} G(z,t) + z\lambda G(z,t) + 2\mu \frac{\partial}{\partial z} G(z,t)$$

i.e.,
$$\frac{\partial}{\partial t} G(z,t) + 2\mu(z-1) \frac{\partial}{\partial z} G(z,t) = \lambda\,(z-1)\, G(z,t) \tag{4.2.5}$$

which is the PDE satisfied by $G(z, t)$ that is to be solved subject to the initial condition $G(z,0) = z^i$ (initially there are i customers in the system).

The related equations of (4.2.5) are

$$\frac{dt}{1}=\frac{dz}{2\mu(z-1)}=\frac{dG}{\lambda(z-1)G}$$

$\frac{dt}{1}=\frac{dz}{2\mu(z-1)}$, yield the solution

$$u(t,z,G)=(z-1)e^{-2\mu t}=C_1 \text{ (a constant)}$$

$\frac{dz}{2\mu(z-1)}=\frac{dG}{\lambda(z-1)G}$, yield the solution

$$V(t,z,G)=G(z,t)e^{-\frac{\lambda}{2\mu}t}=C_2 \text{ (a constant)}$$

Thus the general solution of (4.2.5) is given by

$$G(z,t)e^{-\rho/2z}=g\{(z-1)e^{-2\mu t}\} \tag{4.2.6}$$

$$G(z,0)=e^{\frac{\lambda}{2\mu}z}\ g\{z-1\}$$

$$z^i=e^{\frac{\lambda}{2\mu}z}\ g\{z-1\}$$

With $z-1=y$, $z=1+y$

$$g(y)=e^{-\rho/2(1+y)}\{1+y\}^i$$

Hence from (4.2.6),

$$G(z,t)=e^{\frac{\lambda}{2\mu}z}\ e^{-\frac{\lambda}{2\mu}[1+(z-1)e^{-2\mu t}]}[1+(z-1)e^{-2\mu t}]^i$$

$$= e^{-\rho/2(1-e^{-2\mu t})} \; e^{-\rho/2 z(1-e^{-2\mu t})} \; [1+(z-1)\,\bar{e}^{2\mu t}]^{i} \tag{4.2.7}$$

Now, $$p_n(t) = \frac{1}{n!}\frac{\partial^n G(z,t)}{\partial z^n}\Bigg]_{z=0}$$

$$= \frac{e^{-\rho/2\left(1-e^{-2\mu t}\right)}}{n!} \left\{ \begin{array}{l} [1+(z-1)\bar{e}^{2\mu t}]^{i} \left(\frac{\rho}{2}\right)^{n} \left(1-\bar{e}^{2\mu t}\right)^{n} e^{-\rho/2(1-e^{-2\mu t})} \\ + n c_1 \left[i\left(1+(z-1)\bar{e}^{2\mu t}\right)^{i-1} \left(\bar{e}^{2\mu t}\right)\left(\frac{\rho}{2}\right)^{n-1} \left(1-\bar{e}^{2\mu t}\right)^{n-1} e^{\frac{\rho}{2} z\left(1-\bar{e}^{2\mu t}\right)} \right] \\ + \ldots + (i c_n) n! \left[1+(z-1)\bar{e}^{2\mu t}\right]^{i-n} \left(\bar{e}^{2\mu t}\right)^{n} e^{\frac{\rho}{2} z\left(1-e^{-2\mu t}\right)} \end{array} \right\}_{z=0}$$

where $\rho = \frac{\lambda}{\mu}$

$$= \frac{e^{-\rho/2\left(1-e^{-2\mu t}\right)}}{n!} \left\{ \begin{array}{l} \left(\frac{\rho}{2}\right)^{n} \left(1-e^{-2\mu t}\right)^{n} \left(1-e^{-2\mu t}\right)^{i} \\ + n\left[i e^{-2\mu t} \left(1-e^{-2\mu t}\right)^{i-1} \left(\frac{\rho}{2}\right)^{n-1} \left(1-e^{-2\mu t}\right)^{n-1} \right] + \\ \ldots + (i c_n) n! \left(e^{-2\mu t}\right)^{n} \left(1-e^{-2\mu t}\right)^{i-n} \end{array} \right\} \tag{4.2.8}$$

(using Leibnitz's rule for the nth derivative of a product).

Hence

$$p_n = \lim_{t\to\infty} p_n(t) = \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^{n}}{n!}, n \geq 0 \tag{4.2.9}$$

which is a Poisson distribution with mean $\frac{\rho}{2}$.

We may also obtain p_n, using the Birth-Death methodology as under:

$$p_n = \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1} \; p_o}{\mu_1 \mu_2 \ldots \mu_n} = \frac{\lambda \lambda \ldots \lambda \; p_o}{2\mu(4\mu)\ldots(2n\mu)} = \frac{\left(\frac{\lambda}{2\mu}\right)^n p_o}{n!}$$

Using $\sum p_n = 1$ gives $p_o = e^{-\rho/2}$, where $\rho = \frac{\lambda}{\mu}$

Hence,

$$p_n = \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n!}, \quad n \geq 0 \qquad (4.2.10)$$

which agrees with the result obtained in (4.2.8) while $t \to \infty$

p_0 = probability that the server is idle/system is empty $= e^{-\rho/2}$ (4.2.11)

System size at any time,

$$L_S = \Sigma \, n \, p_n = \frac{\rho}{2} \qquad (4.2.12)$$

Effective arrival rate λ^* is given by

$$\lambda^* = \Sigma \lambda_n \, p_n = \lambda \qquad (4.2.13)$$

Expected waiting time of an arrival in the steady state,

$$W_S = \frac{L_S}{\lambda} = \frac{1}{2\mu} \text{ (using Little's rule)}$$

Hence for any ρ, the number of customers served during the waiting time of an arbitrary customer,

$$\mu W_S = \frac{1}{2} \tag{4.2.14}$$

There are many customers present in the system when service begins. The effective service rate μ^* is given by

$$\mu^*(1-p_o) = \sum_{n=1}^{\infty} \mu_n \, p_n = \Sigma \, 2n\mu \, \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n!} = \lambda \text{, yielding}$$

$$\mu^* = \frac{\lambda}{1-e^{-\rho/2}} \tag{4.2.15}$$

We may also obtain μ^* by the result $\mu^* = \dfrac{\lambda^*}{1-p_0} = \dfrac{\lambda}{1-e^{-\rho/2}}$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = 1-e^{-\rho/2} = 1-p_0 \tag{4.2.16}$$

which is the average number of customers at the server. It is also the fraction of time the server is working.

Now

$$E(N^2) = \Sigma[n(n-1)+n] \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n!}$$

$$= e^{-\rho/2}\left\{\frac{\Sigma\left(\frac{\rho}{2}\right)^n}{(n-2)!} + \frac{\Sigma\left(\frac{\rho}{2}\right)^n}{(n-1)!}\right\}$$

$$= e^{-\rho/2}\left\{\left(\frac{\rho}{2}\right)^2 e^{\rho/2} + \left(\frac{\rho}{2}\right) e^{\rho/2}\right\} = \left(\frac{\rho}{2}\right)^2 + \left(\frac{\rho}{2}\right)$$

Hence,

$$Var(N) = E(N^2) - [E(N)]^2 = \frac{\rho}{2} \tag{4.2.17}$$

which is the variance of the number of customers in the system in the steady state.

Table I: The numerical evaluation of the queue characteristics of the Model AI in (4.2) for various $\rho = \frac{\lambda}{\mu}$

ρ	p_o	$L_S = E(N)$	μW_S	$Var(N)$	μ^*	ρ^*
0.1	0.9512	0.05	0.50	0.05	λ(20.4918)	0.0488
0.2	0.9048	0.10	0.50	0.10	λ(10.5040)	0.0952
0.5	0.7788	0.25	0.50	0.25	λ(4.5208)	0.2212
0.9	0.6376	0.45	0.50	0.45	λ(2.7594)	0.3624
0.99	0.6096	0.495	0.50	0.495	λ(2.5615)	0.3904

4.3 MODEL AII

A partially state-dependent Markovian queueing model characterized by:

$$\lambda_n = \frac{\lambda}{2n+2},\ n \geq 0,\ \mu_n = \mu,\ n \geq 1\ (\lambda, \mu > 0)$$

(Queue with discouraged arrivals) and may be regarded as a modified form of Model II, considered in chapter 2, see for example Natvig, B. (1975).

We shall find p_n, the steady state probability that there are n customers in the system, using the Birth-death methodology and obtain the queue characteristics as under:

$$p_n = \frac{\frac{\lambda}{2}\cdot\frac{\lambda}{4}\cdot\frac{\lambda}{6}\cdots\left(\frac{\lambda}{2n}\right)p_o}{\mu\cdot\mu\cdot\mu\ldots\mu} = \frac{\left(\frac{\lambda}{2\mu}\right)^n \cdot p_o}{n!}$$

Using $\Sigma\, p_n = 1$ gives

$$p_o = e^{-\rho/2}, \tag{4.3.1}$$

which is the state probability of an empty system.

Hence

$$p_n = \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n!},\ n \geq 0,\ \rho = \frac{\lambda}{\mu} \tag{4.3.2}$$

which is again a Poisson distribution with parameter $\frac{\rho}{2}$.

$$L_S = \text{Expected system size} = \frac{\rho}{2} \tag{4.3.3}$$

Effective arrival rate λ^* is given by

$$\lambda^* = \Sigma \lambda_n \, p_n = \Sigma \frac{\lambda}{2n+2} \; \frac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n!} = \mu\left(1 - e^{-\rho/2}\right) \tag{4.3.4}$$

Expected waiting time of an arrival in the steady state,

$$W_S = \frac{L_S}{\lambda^*}$$

$$= \frac{\frac{\rho}{2}}{\mu\left(1 - e^{-\rho/2}\right)}$$

$$\mu W_S = \frac{\rho}{2\left(1 - e^{-\rho/2}\right)}$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system. Clearly,

$$\frac{\rho}{2(1 - e^{-\rho/2})} > 1 \text{ for any } \rho \tag{4.3.5}$$

There are many customers in the system when service begins. The effective service rate μ^* is given by

$$\mu^*(1 - p_0) = \sum_{1}^{\infty} \mu_n \, p_n = \mu(1 - p_0)$$

Hence $\mu^* = \mu$, which is obviously true

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = 1 - e^{-\rho/2} = 1 - p_o \tag{4.3.6}$$

which is the mean number of customers at the server. It is also the fraction of time the server is working. Thus p_n and the traffic intensity, ρ^*, induced by Model AII is exactly same as for Model AI, and the same technique for comparison with other systems at a given level of traffic intensity is called for.

We shall obtain a closed form pdf of the waiting time distribution under FCFS queue discipline for this Model AII.

Waiting Time Distribution

Write $f_n(t)$ for the pdf of a waiting time for the n^{th} arrival, and suppose that $f_n(t) \to f(t)$ as $t \to \infty$. Assume that customers are served by FCFS queue discipline. The duration of services are independent and identically distributed (i.i.d) random variables, independent of arrival process. The service time distribution is exponential with parameter μ.

Then, $$f(t) = \sum_{n=0}^{\infty} \frac{\lambda_n}{\lambda^*} p_n \mu \frac{\bar{e}^{\mu t} (\mu t)^n}{n!} \text{, so that } \int_0^{\infty} f(t)\,dt = 1$$

$$= \sum_{n=0}^{\infty} \frac{\lambda \mu \bar{e}^{\rho/2} \left(\frac{\rho}{2}\right)^n \bar{e}^{\mu t} (\mu t)^n}{(2n+2)\mu(1-\bar{e}^{\rho/2}) n! n!}$$

$$= \frac{\lambda \bar{e}^{\mu t} \bar{e}^{\rho/2}}{2(1-\bar{e}^{\rho/2})} \sqrt{\frac{2}{\lambda t}} \sum_{n=0}^{\infty} \frac{\left(\sqrt{\frac{\lambda t}{2}}\right)^{2n+1}}{(n+1)! n!}$$

$$= \frac{\bar{e}^{\mu t} \bar{e}^{\rho/2}}{\sqrt{2}(1-\bar{e}^{\rho/2})} \sqrt{\frac{\lambda}{t}} I_1\left(\sqrt{\lambda t}\right) \tag{4.3.7}$$

where $I_1(x) = \sum_{K=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2k+1}}{k!\,(k+1)!}$ is the modified Bessel function of the first kind of order 1.

Using (4.3.7) the expected waiting time in the system,

$$W_S = \int_0^\infty t\, f(t)\, dt$$

$$= \frac{e^{-\rho/2}}{\sqrt{2}\left(1-e^{-\rho/2}\right)} \int_0^\infty t\, e^{-\mu t} \sqrt{\frac{\lambda}{t}}\, I_1\left(\sqrt{\lambda t}\right) dt$$

$$= \frac{e^{-\rho/2}}{\left(1-e^{-\rho/2}\right)} \int_0^\infty e^{-\mu t} \sqrt{\frac{\lambda t}{2}} \left[\sqrt{\frac{\lambda t}{2}} + \frac{\left(\sqrt{\frac{\lambda t}{2}}\right)^3}{1!\,2!} + \frac{\left(\sqrt{\frac{\lambda t}{2}}\right)^5}{2!\,3!} + \ldots \right] dt$$

$$= \frac{e^{-\rho/2}}{\left(1-e^{-\rho/2}\right)} \int_0^\infty e^{-\mu t} \left[\frac{\lambda t}{2} + \frac{\left(\frac{\lambda t}{2}\right)^2}{1!\,2!} + \frac{\left(\frac{\lambda t}{2}\right)^3}{2!\,3!} + \ldots \right] dt$$

$$= \frac{e^{-\rho/2}}{\left(1-e^{-\rho/2}\right)} \left\{ \left(\frac{\lambda}{2}\right) \frac{1}{\mu^2} + \frac{\lambda^2}{2\times 4}\left(\frac{2!}{\mu^3}\right) + \frac{\lambda^3}{8\times 12}\left(\frac{3!}{\mu^4}\right) + \ldots \right\}$$

$$= \frac{e^{-\rho/2}}{\left(1-e^{-\rho/2}\right)} \left(\frac{\lambda}{2\mu^2}\right) \left\{ 1 + \left(\frac{\lambda}{2\mu}\right) + \frac{\left(\frac{\lambda}{2\mu}\right)^2}{2!} + \frac{\left(\frac{\lambda}{2\mu}\right)^3}{3!} + \ldots \right\}$$

$$= \frac{\rho\, e^{-\rho/2}}{2\mu\left(1-e^{-\rho/2}\right)} e^{\rho/2} = \frac{\rho}{2\mu\left(1-e^{-\rho/2}\right)} \qquad (4.3.8)$$

which agrees with the result obtained in (4.3.5) by using Little's formula.

Now, $E(N^2) = \Sigma[n(n-1)+n] \cdot \dfrac{e^{-\rho/2}\left(\frac{\rho}{2}\right)^n}{n\,!}$

$$= e^{-\rho/2}\left\{\Sigma \frac{\left(\frac{\rho}{2}\right)^n}{(n-2)\,!} + \Sigma \frac{\left(\frac{\rho}{2}\right)^n}{(n-1)\,!}\right\}$$

$$= e^{-\rho/2}\left\{\left(\frac{\rho}{2}\right)^2 e^{\rho/2} + \left(\frac{\rho}{2}\right) e^{\rho/2}\right\} = \left(\frac{\rho}{2}\right)^2 + \left(\frac{\rho}{2}\right)$$

Hence,

$$Var(N) = E(N^2) - [E(N)]^2 = \frac{\rho}{2} \tag{4.3.9}$$

which is the variance of the number of customers in the system in the steady state

Table II: The numerical evaluation of the queue characteristics of

Model AII in (4.3) for various $\rho = \dfrac{\lambda}{\mu}$

ρ	p_o	$E(N) = L_S$	μW_S	λ^*	ρ^*
0.1	0.9512	0.05	1.0246	(0.0488)μ	0.0488
0.2	0.9048	0.10	1.0509	(0.0952)μ	0.0952
0.5	0.7788	0.25	1.1302	(0.2212)μ	0.2212
0.9	0.6376	0.45	1.2417	(0.3624)μ	0.3624
0.99	0.6096	0.495	1.2679	(0.3904)μ	0.3904

4.4 Model AIII

A completely state dependent Markovian queueing model characterized by $\lambda_n = \frac{\lambda}{2n+2}, n \geq 0, \mu_n = 2n\mu, n \geq 1 (\lambda, \mu > 0)$. This model embodies both mechanisms of Model AI & AII and may be regarded as another modified form of Model III considered in chapter 2.

We shall find p_n, the steady state probability that there are n customers in the system using the Birth-Death methodology and obtain the queue characteristics as under:

$$p_n = \frac{\frac{\lambda}{2}\frac{\lambda}{4}\cdots\frac{\lambda}{2n}}{2\mu 4\mu \ldots 2n\mu} p_0$$

$$= \frac{\rho^n p_o}{2^n 2^n n! n!}, \quad \rho = \frac{\lambda}{\mu} \tag{4.4.1}$$

Using $\Sigma p_n = 1$ gives

$$p_o \sum_{n=0}^{\infty} \frac{\left(\frac{\sqrt{\rho}}{2}\right)^{2n}}{n!\,n!} = 1$$

$$p_0 = \frac{1}{I_0(\sqrt{\rho})}$$

where $I_m(x) = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2k+m}}{k!(k+m)!}$ is the modified Bessel function of the first kind of order $m = 0, 1, 2, \ldots$

Hence,

$$p_n = \frac{\rho^n}{2^{2n}\; n\,!\, n\,!\, I_o(\sqrt{\rho})}, \quad n \geq 0 \tag{4.4.2}$$

Now

$$L_S = E(N) = \Sigma n\; p_n = \Sigma n \cdot \frac{\rho^n}{2^{2n} \cdot n\,!\, n\,!\, I_o\left(\sqrt{\rho}\right)}$$

$$= \frac{\frac{\sqrt{\rho}}{2}}{I_o\left(\sqrt{\rho}\right)} \sum_{k=0}^{\infty} \frac{\left(\frac{\sqrt{\rho}}{2}\right)^{2k+1}}{k\,!\,(k+1)\,!}$$

$$= \frac{\frac{\sqrt{\rho}}{2} \cdot I_1\left(\sqrt{\rho}\right)}{I_o\left(\sqrt{\rho}\right)} \tag{4.4.3}$$

Effective arrival rate λ^* is given by

$$\lambda^* = \Sigma\lambda_n\, p_n = \Sigma \frac{\lambda}{2n+2} \; \frac{\left(\frac{\sqrt{\rho}}{2}\right)^{2n}}{n!\; n!\, I_0\left(\sqrt{\rho}\right)}$$

$$= \frac{\lambda}{\sqrt{\rho}} \sum_{n=0}^{\infty} \frac{\left(\frac{\sqrt{\rho}}{2}\right)^{2n+1}}{(n+1)!\, n!\, I_0\left(\sqrt{\rho}\right)}$$

$$= \left(\frac{\lambda}{\sqrt{\rho}}\right) \frac{I_1(\sqrt{\rho})}{I_o(\sqrt{\rho})} \tag{4.4.4}$$

Effective service rate, μ^* is given by

$$\mu^*(1-p_o)=\sum_{n=1}^{\infty}(2n\mu)\frac{\left(\frac{\sqrt{\rho}}{2}\right)^{2n}}{n!\,n!\,I_o\left(\sqrt{\rho}\right)}$$

$$=\frac{2\mu}{I_o\left(\sqrt{\rho}\right)}\left(\frac{\sqrt{\rho}}{2}\right)\sum_{K=0}^{\infty}\frac{\left(\frac{\sqrt{\rho}}{2}\right)^{2K+1}}{K!\,(K+1)\,!}$$

$$=\frac{\mu\sqrt{\rho}}{I_o\left(\sqrt{\rho}\right)}I_1\left(\sqrt{\rho}\right)$$

Hence,

$$\mu^*=\frac{\mu\sqrt{\rho}\;I_1\left(\sqrt{\rho}\right)}{\left[I_o\left(\sqrt{\rho}\right)-1\right]}$$

or

$$\mu^*=\frac{\lambda^*}{1-p_0}=\left(\frac{\lambda}{\sqrt{\rho}}\right)\frac{I_1(\sqrt{\rho})}{[I_0(\sqrt{\rho}-1)]} \tag{4.4.5}$$

Effective traffic intensity,

$$\rho^*=\frac{\lambda^*}{\mu^*}$$

$$=\frac{I_0\left(\sqrt{\rho}\right)-1}{I_0\left(\sqrt{\rho}\right)}=1-p_0 \tag{4.4.6}$$

which is the mean number of customers at the server. It is also the fraction of time the server is working.

Using Little's formula, the expected waiting time of an arrival in the steady state is given by

$$W_S=\frac{L_S}{\lambda^*}=\frac{1}{2\mu}$$

Hence,

$$\mu W_S = \frac{1}{2} \tag{4.4.7}$$

i.e., the number of customers served during the waiting time of an arbitrary arrival equals $\frac{1}{2}$ for any ρ.

Remark: We have already discussed the Models AI, AII, and AIII in Chapter 3. However, as modified models, which are modified version of the models considered in Chapter 2, and for comparing with more modified state-dependent Markovian queueing models, we have considered the Models AI, AII and AIII in this chapter also.

Table III: The numerical evaluation of the queue characteristics of Model AIII in (4.4) for various $\rho = \frac{\lambda}{\mu}$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9748	0.0252	0.5	0.4997 λ	1.9868 μ	0.0252
0.2	0.9512	0.0491	0.5	0.4909 λ	2.0120 μ	0.0488
0.5	0.8849	0.1182	0.5	0.4729 λ	2.0539 μ	0.1151
0.9	0.8073	0.2035	0.5	0.4516 λ	2.1092 μ	0.1927
0.99	0.7816	0.2223	0.5	0.4469 λ	2.1226 μ	0.2084
1	0.7898	0.2232	0.5	0.4446 λ	2.1333 μ	0.2084

4.5 MODEL AIV

A partially linearly state dependent Markovian queueing model characterized by $\lambda_n = \dfrac{\lambda}{n+2}, n \geq 0, \mu_n = \mu, n \geq 1 \ (\lambda, \mu > 0)$, another modified model of Model II considered in chapter 2.

We shall first find p_n, the steady state probability distribution of the number, n = 0, 1, 2 ..., of customers in the system using the Birth-death methodology and then find the queue characteristics as under:

$$p_n = \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} p_o$$

$$= \frac{\frac{\lambda}{2} \cdot \frac{\lambda}{3} \cdots \frac{\lambda}{n+1}}{\mu \mu \ldots \mu} p_o$$

$$= \frac{\rho^n p_o}{(n+1)!}, \qquad \rho = \frac{\lambda}{\mu}$$

Using $\Sigma p_n = 1$ gives $p_o = \frac{\rho}{e^\rho - 1}$

Hence,

$$p_n = \frac{\rho^{n+1}}{(e^\rho - 1)\ (n+1)\ !}, n \geq 0, \rho > 0 \tag{4.5.1}$$

Average system size in the steady state,

$$L_S = \sum_1^\infty n\ p_n = \sum_1^\infty \left[(n+1) - 1\right] \frac{\rho^{n+1}}{(e^\rho - 1)\ (n+1)\ !}$$

$$= \frac{1}{(e^\rho - 1)} \left\{\rho\ (e^\rho - 1) - (e^\rho - \rho - 1)\right\}$$

$$= \frac{1}{(e^\rho - 1)} \left\{\rho\ e^\rho + 1 - e^\rho\right\} \tag{4.5.2}$$

Effective arrival rate, λ^* (Average number of customers who actually enters the system), is given by

$$\lambda^* = \sum_0^\infty \lambda_n\ p_n$$

$$= \frac{\lambda}{(e^\rho - 1)}\ \Sigma\ \frac{\rho^{n+1}}{(n+2)\ !} \tag{4.5.3}$$

$$= \frac{\lambda}{\rho\ (e^\rho - 1)} \left\{e^\rho - \rho - 1\right\}$$

Expected waiting time of an arrival is given by

$$W_S = \frac{L_S}{\lambda^*} = \frac{\rho\left[e^{\rho}(\rho-1)+1\right]}{\lambda(e^{\rho}-\rho-1)} = \frac{1}{\mu}\frac{\left[e^{\rho}(\rho-1)+1\right]}{(e^{\rho}-\rho-1)}$$

Hence,

$$\mu W_S = \frac{\left[e^{\rho}(\rho-1)+1\right]}{(e^{\rho}-\rho-1)} \qquad (4.5.4)$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system. It can be proved that μW_S is always greater than one for any ρ.

Proof

$$\mu W_S = \frac{\left[e^{\rho}(\rho-1)+1\right]}{(e^{\rho}-\rho-1)}$$

$$= \frac{\left[\left(1+\frac{\rho}{1}+\frac{\rho^2}{2!}+\frac{\rho^3}{3!}+\ldots\right)(\rho-1)+1\right]}{\frac{\rho^2}{2!}+\frac{\rho^3}{3!}+\frac{\rho^4}{4!}+\ldots}$$

$$= \frac{\left(\frac{\rho^2}{2!}+\frac{\rho^3}{3!}+\ldots\right)}{\frac{\rho^2}{2!}\left[1+\frac{\rho}{3}+\frac{\rho^2}{12}+\ldots\right]}$$

$$= \frac{2}{\rho^2}\left[\frac{\rho^2}{2!}+\frac{\rho^3}{3!}+\ldots\right]\left[1-\frac{\rho}{3}-\frac{\rho^2}{12}-\ldots\right]$$

$$= \left[1+\frac{\rho}{3}+\frac{\rho^2}{12}+\ldots\right]\left[1-\frac{\rho}{3}-\frac{\rho^2}{12}-\ldots\right]$$

which is clearly a value greater than one for any ρ.

Effective service rate, μ^* (weighted average service rate), is given by

$$\mu^*(1-p_o)=\sum_1^\infty \mu_n \cdot p_n$$

$$=\sum_1^\infty \mu \cdot p_n = \mu\,\{1-p_o\},\text{ yielding}$$

Hence,

$$\mu^* = \mu\text{, which is obviously true.} \tag{4.5.5}$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*}$$

$$=\frac{e^\rho-\rho-1}{e^\rho-1} = 1-p_o \tag{4.5.6}$$

which is the average number of customers at the server. It is also the proportion of time the server is working.

Now, $E(N^2)=\sum_1^\infty n^2 \cdot p_n = \sum_1^\infty \left[(n+1)^2-2(n+1)+1\right]\dfrac{\rho^{n+1}}{(e^\rho-1)\,(n+1)\,!}$

$$=\frac{1}{(e^\rho-1)}\left\{\sum_1^\infty \frac{\rho^{n+1}}{n\,!}(n+1)-2\sum_1^\infty\frac{\rho^{n+1}}{n\,!}+\sum_1^\infty\frac{\rho^{n+1}}{(n+1)\,!}\right\}$$

$$=\frac{1}{(e^\rho-1)}\left\{\sum_1^\infty \frac{\rho^{n+1}}{(n-1)\,!}+\sum_1^\infty\frac{\rho^{n+1}}{n\,!}-2\sum_1^\infty\frac{\rho^{n+1}}{n\,!}+\sum_1^\infty\frac{\rho^{n+1}}{(n+1)\,!}\right\}$$

$$=\frac{1}{(e^\rho-1)}\left\{\rho^2\cdot e^\rho-\rho(e^\rho-1)+(e^\rho-\rho-1)\right\}$$

$$=\frac{1}{e^\rho-1}\left\{e^\rho(\rho^2-\rho+1)-1\right\}$$

$$Var(N)=E(N^2)-\left[E(N)\right]^2$$

$$=\frac{[e^{\rho}(\rho^2-\rho+1)-1]}{e^{\rho}-1}-\left\{\frac{\rho\, e^{\rho}+1-e^{\rho}}{(e^{\rho}-1)}\right\}^2 \tag{4.5.7}$$

which is the variance of the number of customers in the system in the steady state

Waiting Time Distribution

We shall derive an explicit expression for the pdf of the Waiting time distribution for the given queue with discouraged arrivals.

Write $f_n(t)$ for the pdf of waiting time for the nth arrival, and suppose that $f_n(t) \to f(t)$ as $t \to \infty$. Assume that customers are served according to FCFS queue discipline. The duration of services are independent and identically distributed (i.i.d) random variables, independent of the arrival process. The service time distribution is exponential with parameter μ.

Then,

$$f(t)=K\sum_{n=0}^{\infty}\lambda_n\cdot p_n\cdot\mu\,\frac{e^{-\mu t}(\mu t)^n}{n\,!}$$

$$=K\sum_{n=0}^{\infty}\frac{\lambda}{n+2}\cdot\frac{\rho^{n+1}}{(e^{\rho}-1)\,(n+1)\,!}\,\frac{e^{-\mu t}(\mu t)^n}{n\,!}$$

$$=\frac{K(\lambda\rho\cdot\mu)\,e^{-\mu t}}{(e^{\rho}-1)}\sum_{n=0}^{\infty}\frac{(\rho\,\mu t)^n}{(n+2)\,!\,n\,!},$$

where K is a constant so that $\int_0^{\infty} f(t)\,dt=1$

$$=\frac{K\,e^{-\mu t}}{(e^{\rho}-1)}(\lambda\rho\mu)\frac{1}{\lambda t}\sum_{n=0}^{\infty}\frac{\left(\sqrt{\lambda t}\right)^{2n+2}}{n\,!\,(n+2)\,!}$$

$$= \frac{K\, e^{-\mu t}}{(e^{\rho}-1)}\ \frac{\lambda}{t}\ I_2\left(2\sqrt{\lambda t}\right)$$

where $I_2\left(2\sqrt{\lambda t}\right)$ is the modified Bessel function of the first kind of order 2.

To find K

$$K\int_0^{\infty}\frac{e^{-\mu t}}{(e^{\rho}-1)}\ \frac{\lambda}{t}\left[\frac{\left(\sqrt{\lambda t}\right)^2}{2!}+\frac{\left(\sqrt{\lambda t}\right)^4}{1!\,3!}+\frac{\left(\sqrt{\lambda t}\right)^6}{2!\,4!}+\ldots\right]dt=1$$

$$\frac{K\lambda}{e^{\rho}-1}\int_0^{\infty}e^{-\mu t}\left[\frac{\lambda}{2}+\frac{\lambda^2 t}{6}+\frac{\lambda^3}{2\times 24}t^2+\ldots\right]dt=1$$

$$\frac{K\lambda}{(e^{\rho}-1)}\left\{\left(\frac{\lambda}{2}\right)\left(\frac{1}{\mu}\right)+\frac{\lambda^2}{6}\left(\frac{1}{\mu^2}\right)+\frac{\lambda^3}{2\times 48}\left(\frac{2}{\mu^3}\right)+\ldots\right\}=1$$

$$\frac{K\lambda}{(e^{\rho}-1)\,\rho}\left\{\frac{\rho^2}{2\,!}+\frac{\rho^3}{3\,!}+\frac{\rho^4}{4\,!}+\ldots\right\}=1$$

$$\frac{K\lambda}{\rho(e^{\rho}-1)}\,(e^{\rho}-\rho-1)=1$$

$$K=\frac{(e^{\rho}-1)}{\mu(e^{\rho}-\rho-1)}$$

Hence

$$f(t)=\frac{\rho\, e^{-\mu t}}{(e^{\rho}-\rho-1)\,t}\ I_2\left(2\sqrt{\lambda t}\right),\qquad t\geq 0 \tag{4.5.8}$$

which is the pdf of the waiting time distribution under FCFS queue discipline.

Using (4.5.8), the expected waiting time, W_S, is given by

$$W_S = \int_0^\infty t \frac{\rho\, e^{-\mu t}}{t(e^\rho - \rho - 1)} I_2\left(2\sqrt{\lambda t}\right) dt$$

$$= \frac{\rho}{(e^\rho - \rho - 1)} \int_0^\infty e^{-\mu t} \left[\frac{\left(\sqrt{\lambda t}\right)^2}{2!} + \frac{\left(\sqrt{\lambda t}\right)^4}{1!3!} + \frac{\left(\sqrt{\lambda t}\right)^6}{2!4!} + ... \right] dt$$

$$= \frac{\rho}{(e^\rho - \rho - 1)} \left[\frac{\lambda}{2}\left(\frac{1}{\mu^2}\right) + \frac{\lambda^2}{6}\left(\frac{2}{\mu^3}\right) + \frac{\lambda^3}{2!4!}\left(\frac{6}{\mu^4}\right) + ... \right]$$

$$= \frac{\rho}{\lambda\,(e^\rho - \rho - 1)} \left[\frac{\rho^2}{2} + \frac{\rho^3}{3} + \frac{\rho^4}{8} + ... \right], \rho = \frac{\lambda}{\mu}$$

$$= \frac{1}{\mu\,(e^\rho - \rho - 1)} \left\{ \rho\left(1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \frac{\rho^3}{3!} + ...\right) - \left(1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \frac{\rho^3}{3!} + ...\right) + 1 \right\}$$

$$= \frac{1}{\mu\,(e^\rho - \rho - 1)} \left\{ \rho\, e^\rho - e^\rho + 1 \right\} \qquad (4.5.9)$$

which agrees with the result obtained in (4.5.4) by Little's formula.

Table IV: The Numerical evaluation of the queue characteristics of the Model AIV in (4.5) for various ρ

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9505	0.0505	1.0192	0.4916 λ	μ	0.0492
0.2	0.9033	0.1034	1.0692	0.4833 λ	μ	0.0967
0.5	0.7708	0.2707	1.1813	0.4584 λ	μ	0.2292
0.9	0.6166	0.5166	1.3475	0.4260 λ	μ	0.3834
0.95	0.5854	0.5505	1.3878	0.4188 λ	μ	0.4146
1.0	0.5820	0.5820	1.3922	0.4180 λ	μ	0.4180

4.6 MODEL AV

A partially linearly state dependent Markovian queueing model characterized by $\lambda_n = \dfrac{\lambda}{n+2}, n \geq 0,\ \mu_n = n\mu, n \geq 1\ \ (\lambda, \mu > 0)$

This model may be regarded as a Modified form of Model III considered in Chapter 2.

Using the Birth-death methodology, we shall obtain p_n, the steady state probability that there are n customers in the system and the queue characteristics as under:

$$p_n = \frac{\frac{\lambda}{2} \cdot \frac{\lambda}{3} \cdots \frac{\lambda}{n+1}}{\mu(2\mu)\ldots n\mu}\, p_o$$

$$= \frac{\rho^n\, p_o}{(n+1)!\, n!},\quad \rho = \frac{\lambda}{\mu}$$

Using $\Sigma p_n = 1$ gives $p_o = \dfrac{\sqrt{\rho}}{I_1\left(2\sqrt{\rho}\right)}$, where $I_1\left(2\sqrt{\rho}\right)$ is the modified Bessel function of the first kind of order 1 defined by

$$I_m(x) = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2k+m}}{k\,!(k+m)\,!},\qquad m = 0, 1, 2, 3\ldots$$

Hence

$$p_n = \frac{\sqrt{\rho}}{I_1(2\sqrt{\rho})} \frac{\rho^n}{n\,!\,(n+1)!},\quad n \geq 0;\ \rho = \frac{\lambda}{\mu} \tag{4.6.1}$$

Average number of customers in the system in the steady state,

$$
\begin{aligned}
L_S &= \sum_{1}^{\infty} n\, p_n \\
&= \sum_{1}^{\infty} n \frac{\sqrt{\rho}}{I_1(2\sqrt{\rho})} \frac{\rho^n}{n\,!\,(n+1)\,!} \\
&= \frac{\sqrt{\rho}}{I_1(2\sqrt{\rho})} \sum \frac{\rho^n}{(n-1)\,!\,(n+1)\,!} \\
&= \frac{\sqrt{\rho}}{I_1(2\sqrt{\rho})} \sum_{m=0}^{\infty} \frac{\rho^{m+1}}{m\,!\,(m+2)\,!} \\
&= \frac{\sqrt{\rho}}{I_1(2\sqrt{\rho})} \sum_{0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2m+2}}{m\,!\,(m+2)\,!} \\
&= \frac{\sqrt{\rho}}{I_1(2\sqrt{\rho})} I_2\left(2\sqrt{\rho}\right) \qquad (4.6.2)
\end{aligned}
$$

Effective arrival rate, λ^* (Average number of customers who actually enters the system), is given by

$$
\begin{aligned}
\lambda^* = \sum_{n=0}^{\infty} \lambda_n\, p_n &= \sum_{0}^{\infty} \frac{\lambda}{n+2} \cdot \frac{\sqrt{\rho}}{I_1\left(2\sqrt{\rho}\right)} \frac{\rho^n}{n\,!\,(n+1)\,!} \\
&= \frac{\sqrt{\rho}\cdot\lambda}{I_1\left(2\sqrt{\rho}\right)} \cdot \frac{1}{\rho} \sum_{n=0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2n+2}}{n\,!\,(n+2)\,!} \\
&= \frac{\lambda}{\sqrt{\rho}} \frac{I_2\left(2\sqrt{\rho}\right)}{I_1\left(2\sqrt{\rho}\right)} \qquad (4.6.3)
\end{aligned}
$$

Expected waiting time, W_S, of an arrival in the steady state is given by

$$W_S = \frac{L_s}{\lambda^*} = \frac{1}{\mu}$$

Hence,

$$\mu W_S = 1, \tag{4.6.4}$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system.

Effective service rate, μ^* (weighted average service rate), is given by

$$\begin{aligned}\mu^*(1-p_o) &= \sum_{n=1}^{\infty} n\mu \cdot \frac{\sqrt{\rho}}{I_1(2\sqrt{\rho})} \frac{\rho^n}{n!(n+1)!} \\ &= \frac{\sqrt{\rho}\cdot\mu}{I_1(2\sqrt{\rho})} \cdot \sum_{m=0}^{\infty} \frac{\rho^{m+1}}{m!(m+2)!} \\ &= \frac{\mu\sqrt{\rho}}{I_1(2\sqrt{\rho})} \cdot \sum_{0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2m+2}}{m!(m+2)!} \\ &= \frac{\mu\sqrt{\rho}}{I_1(2\sqrt{\rho})} \cdot I_2\left(2\sqrt{\rho}\right)\end{aligned}$$

Hence,

$$\mu^* = \frac{\mu\sqrt{\rho}\cdot I_2\left(2\sqrt{\rho}\right)}{I_1(2\sqrt{\rho}) - \sqrt{\rho}}$$

or

$$\mu^* = \frac{\lambda^*}{1-p_0} = \frac{\mu\sqrt{\rho}\cdot I_2\left(2\sqrt{\rho}\right)}{I_1(2\sqrt{\rho}) - \sqrt{\rho}} \tag{4.6.5}$$

Effective traffic intensity,

$$\begin{aligned}\rho^* &= \frac{\lambda^*}{\mu^*} \\ &= \frac{I_1\left(2\sqrt{\rho}\right) - \sqrt{\rho}}{I_1(2\sqrt{\rho})} = 1 - p_o\end{aligned} \tag{4.6.6}$$

which is the average number of customers at the server. It is also the proportion of time the server is working.

Now, $E(N^2) = \Sigma\, n^2 \cdot p_n$

$$= \sum_1^{\infty}\left[(n+1)^2 - 2(n+1) + 1\right] \frac{\sqrt{\rho}}{I_1\left(2\sqrt{\rho}\right)} \cdot \frac{\rho^n}{n\,!\,(n+1)\,!}$$

$$= \frac{\sqrt{\rho}}{I_1\left(2\sqrt{\rho}\right)} \left\{ \sum_1^{\infty}\left(\frac{\rho^n}{(n-1)\,!\,n\,!} + \frac{\rho^n}{n\,!\,(n+1)\,!}\right) - 2\sum_1^{\infty}\frac{\rho^n}{n\,!\,n\,!} + \Sigma\frac{\rho^n}{n\,!\,(n+1)\,!}\right\}$$

$$= \frac{\sqrt{\rho}}{I_1\left(2\sqrt{\rho}\right)} \left\{ \sqrt{\rho} \cdot I_1\left(2\sqrt{\rho}\right) - 2\left[I_o\left(2\sqrt{\rho}\right) - 1\right] + \frac{2}{\sqrt{\rho}}\left(I_1\left(2\sqrt{\rho}\right) - \sqrt{\rho}\right)\right\} \quad (4.6.\ 7)$$

$$Var(N) = E(N^2) - \left[E(N)\right]^2$$

$$= \frac{\sqrt{\rho}}{I_1\left(2\sqrt{\rho}\right)} \left\{ \sqrt{\rho} \cdot I_1\left(2\sqrt{\rho}\right) + \frac{2}{\sqrt{\rho}} I_1\left(2\sqrt{\rho}\right) - 2 I_o\left(2\sqrt{\rho}\right)\right\}$$

$$-\left(\frac{\sqrt{\rho}}{I_1\left(2\sqrt{\rho}\right)} \cdot I_2\left(2\sqrt{\rho}\right)\right)^2 \quad (4.6.8)$$

which is the variance of the number of customers in the system in the steady state

Table V: Numerical evaluation of the queue characteristics of Model AV in (4.6) for various $\rho = \frac{\lambda}{\mu}$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9556	0.0497	1	0.5860 λ	1.3277 μ	0.0441
0.2	0.9116	0.0979	1	0.4804 λ	1.0882 μ	0.0883
0.5	0.7898	0.2310	1	0.4620 λ	1.0990 μ	0.2102
0.9	0.6551	0.3979	1	0.4393 λ	1.1464 μ	0.3449
0.99	0.6255	0.4258	1	0.4356 λ	1.1517 μ	0.3745

4.7 MODEL AVI

A partially linearly state dependent Markovian queueing model characterized by $\lambda_n = \lambda, n \geq 0,\ \mu_n = (n+2)\mu, n \geq 1\ (\lambda, \mu > 0)$, which may be considered as a Modified form of Model II.

We shall find p_n, the steady state probability that there are n customers in the system using the Birth-death methodology and then obtain the queue characteristics as under:

$$p_n = \frac{\lambda \cdot \lambda \ldots \lambda \cdot}{3\mu \cdot 4\mu \ldots (n+2)\mu} p_o = \frac{2\rho^n p_o}{(n+2)\,!},\ \rho = \frac{\lambda}{\mu}$$

Using $\Sigma p_n = 1$ gives

$$\frac{2}{\rho^2} p_o \sum_0^\infty \frac{\rho^{n+2}}{(n+2)\,!} = 1$$

$$\frac{2}{\rho^2} (e^\rho - \rho - 1)\ p_o = 1$$

$$p_o = \frac{\rho^2}{2(e^\rho - \rho - 1)} \tag{4.7.1}$$

Hence,

$$p_n = \frac{\rho^{n+2}}{(e^\rho - \rho - 1)\,(n+2)\,!}, \quad n \geq 0;\ \rho = \frac{\lambda}{\mu} \tag{4.7.2}$$

Expected number of customers in the system in the steady state,

$$L_S = \sum_{1}^{\infty} n\ p_n$$

$$= \sum_{1}^{\infty} \left[(n+2)-2\right] \frac{\rho^{n+2}}{(e^{\rho}-\rho-1)\ (n+2)\ !}$$

$$= \frac{1}{(e^{\rho}-\rho-1)} \left\{\rho(e^{\rho}-\rho-1) - 2\left(e^{\rho} - \frac{\rho}{2} - \rho - 1\right)\right\}$$

$$= \frac{1}{(e^{\rho}-\rho-1)} \left\{\rho e^{\rho} + \rho - 2e^{\rho} + 2)\right\} \qquad (4.7.3)$$

Effective arrival rate, λ^* (Average number of customers who actually enters the system), is given by

$$\lambda^* = \sum_{0}^{\infty} \lambda_n\ p_n = \lambda \text{, which is obviously true} \qquad (4.7.4)$$

Expected waiting time, W_S, of an arrival in the steady state is given by

$$W_S = \frac{L_S}{\lambda^*} \text{ (By Little's formula)}$$

$$= \frac{\rho e^{\rho} + \rho - 2e^{\rho} + 2}{\lambda(e^{\rho}-\rho-1)}$$

Hence,

$$\mu W_S = \frac{\rho\ e^{\rho} + \rho - 2e^{\rho} + 2}{\rho\ (e^{\rho}-\rho-1)} \qquad (4.7.5)$$

which is the number of customers served during the waiting time of an arrival in the system.

Effective service rate, μ^* (weighted average service time), is given by

$$\mu^* (1 - p_o) = \sum_1^{\infty} (n+2)\mu \frac{\rho^{n+2}}{(e^{\rho} - \rho - 1)\ (n+2)\ !}$$

$$= \frac{\mu}{(e^{\rho} - \rho - 1)} \sum_1^{\infty} \frac{\rho^{n+2}}{(n+1)\ !}$$

$$= \frac{\mu\,\rho}{(e^{\rho} - \rho - 1)} \left\{(e^{\rho} - \rho - 1)\right\} = \mu\,\rho \text{, yielding}$$

$$\mu^* = \frac{\mu\rho}{1 - \dfrac{\rho^2}{2(e^{\rho} - \rho - 1)}} = \frac{2\mu\,\rho(e^{\rho} - \rho - 1)}{2(e^{\rho} - \rho - 1) - \rho^2}$$

or

$$\mu^* = \frac{\lambda^*}{1 - p_0} = \frac{2\mu\,\rho\ (e^{\rho} - \rho - 1)}{2\ (e^{\rho} - \rho - 1) - \rho^2} \tag{4.7.6}$$

Effective traffic intensity,

$$\rho^* = \frac{2\ (e^{\rho} - \rho - 1) - \rho^2}{2\ (e^{\rho} - \rho - 1)} = 1 - p_o \tag{4.7.7}$$

which is the number of customers at the server. It is also the proportion of time the server is working.

Table VI: Numerical evaluation of the queue characteristics of the Model AVI in (4.7) for various $\rho = \frac{\lambda}{\mu}$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9615	0.0231	0.2308	λ	2.6μ	0.0385
0.2	0.9346	0.0692	0.3458	λ	3.0571μ	0.0654
0.5	0.8406	0.1812	0.3625	λ	3.1371μ	0.1594
0.9	0.7237	0.3475	0.3861	λ	3.2577μ	0.2763
0.99	0.6989	0.3878	0.3917	λ	3.2877μ	0.3011
1	0.6960	0.3930	0.3930	λ	3.2904μ	0.3039

Remark: It is possible to evaluate Numerically the queue characteristics of the Models AI, AII, AIII, AIV & AV for any $\rho > 0$. But stationary condition exist only for ρ where $0 < \rho < 1$.

4.8 MODEL AVII

A completely linearly state dependent Markovian queueing model characterized by $\lambda_n = (n+2)\lambda, n \geq 0$ $\mu_n = (n+1)\mu, n \geq 1$ $(\lambda > 0, \mu > 0)$, which is a Modified form of Model IV considered in Chapter 2.

We shall find p_n, which is the steady state probability of n customers in the system using the Birth-death methodology. Using p_n, we obtain the steady state queue characteristics.

$$p_n = \frac{2\lambda\, 3\lambda ... (n+1)\lambda}{2\mu\, 3\mu ... (n+1)\mu} p_o$$
$$= \rho^n \cdot p_o,\ \rho = \frac{\lambda}{\mu}$$

Using $\Sigma\, p_n = 1$ gives $p_o = (1-\rho),\ \rho < 1$

Hence,

$$p_n = (1-\rho)\rho^n, \qquad 0 < \rho < 1 \ \text{ and } n \geq 0 \tag{4.8.1}$$

which is the same as the steady state probability of queue length of the standard model, *M/M/*1 queue

$$L_S = \text{system size at any time}$$
$$= \Sigma n \cdot p_n = (1-\rho)\ (\rho + 2\rho^2 + 3\rho^3 + ...)$$
$$= \frac{\rho}{1-\rho} \tag{4.8.2}$$

Effective arrival rate, λ^* (Average number of customers who actually enters the system), is given by

$$\lambda^* = \sum_0^{\infty} (n+2)\lambda \cdot p_n = \lambda \left\{ \Sigma n\, p_n + 2\Sigma\, p_n \right\}$$
$$= \lambda \left\{ \frac{\rho}{1-\rho} + 2 \right\} = \frac{\lambda\,(2-\rho)}{(1-\rho)} \tag{4.8.3}$$

Effective service rate, μ^* (weighted average service time), is given by

$$\mu^*(1-p_o) = \sum_1^{\infty}(n+1)\,\mu\,(1-\rho)\,\rho^n$$

$$= (1-\rho)\,\mu\left\{\sum_1 n\,\rho^n + \sum_1^{\infty}\rho^n\right\}$$

$$= (1-\rho)\,\mu\left\{\frac{\rho}{(1-\rho)^2} + \frac{\rho}{(1-\rho)}\right\}$$

$$= \frac{\mu\,\rho(2-\rho)}{(1-\rho)}\text{, yielding}$$

$$\mu^* = \frac{\mu(2-\rho)}{(1-\rho)}. \tag{4.8.4}$$

Expected weighting time of an arrival in the steady state is given by,

$$W_S = \frac{L_S}{\lambda^*} \quad \text{(By Little's formula)}$$

$$= \frac{1}{\mu(2-\rho)}$$

Hence,

$$\mu W_S = \frac{1}{2-\rho} \tag{4.8.5}$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system for any ρ in $0 < \rho < 1$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \frac{\lambda}{\mu} = \rho = 1 - p_o \tag{4.8.6}$$

which is the average number of customers at the server. It is also the proportion of time the server is working.

Table VII: Numerical evaluation of the queue characteristics of the Model AVII in (4.8) for various $\rho = \frac{\lambda}{\mu}$ (0< ρ < 1)

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9	0.1111	0.5263	2.1111 λ	2.111 μ	0.1
0.2	0.8	0.25	0.5556	2.25 λ	2.25 μ	0.2
0.5	0.5	1.00	0.6667	3.00 λ	3.00 μ	0.5
0.9	0.1	9.00	0.9091	11.00 λ	11.00 μ	0.9
0.99	0.01	99.00	0.9901	101.00 λ	101.00 μ	0.99

4.9 SUMMARY OF RESULTS AND CONCLUDING REMARKS

1. For the Model AI:

   (i) For fixed arrival rate λ, say, an increase in ρ calls for a deduction in mean effective service time.

   (ii) An increase in ρ calls an increase in effective traffic intensity. That is, mean number of customers at the server or the fraction of time the service is working increases with increasing value of ρ

   (iii) In terms of the performance measure, p_o, which is the state probability of an empty system, it is verified that Model AI, which is the modification of Model I, is better than Model I (A self service system)

2. For the Model AII:

   (i) For a fixed service rate μ, an increase in value of ρ calls an increase in effective arrival rate.

(ii) The number of customers served during the waiting time of an arbitrary arrival increases slightly with increase in ρ which is only 0.5 for any ρ.

(iii) Model AII, a Modification of Model II, is better than Model II (A queueing model in which customers are discouraged) by comparing p_o.

(iv) An increase in ρ calls for an increase in ρ^*.

3. For the Model AIII:

(i) With reference to the performance measure p_o, which is also state probability of an empty system, it is proved that Model AIII is better than the Models AI & AII.

(ii) When ρ increases, ρ^* increases, which is lesser than ρ^* of Models AI & AII.

(iii) The number of customers in the system in the steady-state for Model AIII is lesser than that of Model AI and AII.

4. For the Model AIV: $L_s = p_o$ when $\lambda = \mu$. which distinguishes from the other Models considered in this chapter.

5. For the Model AV:

(i) When ρ increases, the system size, L_S increases.

(ii) For fixed service rate μ, an increase in the value of ρ calls for an increase in effective service rate.

(iii) For any $\rho > 0$, the number of customers served during the waiting time of an arbitrary customer equal one.

(iv) For fixed arrival rate λ, an increase in the value of ρ calls for a decrease in effective arrival rate as expected.

6. For the Model AVI:

(i) For fixed service rate μ, an increase in the value of ρ calls an increase in effective service rate.

(ii) The number of customers served during the waiting time of an arbitrary arrival is equal to the number of customers in the system when arrival rate equals service rate.

7. For the Model AVII: for fixed arrival rate λ, and service rate μ, effective arrival rate equals effective service rate for a given $\rho = \frac{\lambda}{\mu}$.

8. With reference to the performance measure p_o, the state probability of an empty system, Model A III is better than the other state-dependent queueing models considered in this chapter.

It may be proved that the steady-state probability distribution of the number of customers in the system for the general linearly state-dependent Markovian queueing models characterized by $\lambda_n = (n + k)\lambda, \quad n \geq 0$; $\mu_n = (n + \gamma)\mu$, $n \geq 1$ respectively follows the Geometric law given by $p_n = (1 - \rho)\, \rho^n$, $n \geq 0$ and $0 < \rho < 1$ which is the same as that of the standard M/M/1 Model for the pairs of $\{(k, \gamma)\} = \{(1, 0), (2, 1), (3, 2), ...\}$. It may also be verified that when k increases, the number of customers served during the waiting time of an arbitrary arrival decreases.

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

Chapter 5

NON-LINEAR STATE DEPENDENT MARKOVIAN QUEUEING MODELS

5.1 INTRODUCTION

In the earlier chapters, we have discussed various state-dependent Markovian queueing models where the arrival rate and service rate are linear functions. The model II of Chapter 2 where potential customers are discouraged by queue length is a non-linear state-dependent model. Here in this chapter, we shall discuss some non-linear state-dependent queueing models with different types of arrival/service rate. In recent years there has been a considerable interest in computing state probabilities of various queueing models. Since performance measures are functions of state probabilities their exact evaluation is important. We obtain the steady state probability distribution of queue length using the Birth-death methodology. We obtain the queue characteristics of the Models in the steady state. We also evaluate numerically the performance measures of the models and make a comparison between them.

5.2 MODEL BI

The non-linear state dependent (NLS) queueing Model characterized by

$\lambda_n = \dfrac{\lambda}{(n+1)^2}, n \geq 0; \mu_n = \mu, n \geq 1 \; (\lambda, \mu > 0)$.

This model may be regarded as modified form of *M*/*M*/1 model with balking in which an arriving customer might be discouraged by the number of customers already in the system and thus do not enter.

This non-linear state-dependant simple server queueing model like Model II discussed in chapter 2 is useful to model a computing facility that is solely dedicated to batch-job processing. Job submissions are discouraged when the facility is heavily used and can be modeled as a Poisson process with state department arrival rate. The time taken to process each job is exponentially distributed with a constant service rate regardless of the number of jobs in the system.

The state- dependant discouraged arrivals queue has been studied in the past by Natvig (1974, 1975), Vandoorn E.A (1981) and here the arrivals are geared (or could be controlled) in accordance with the availability of service.

We shall find p_n, the steady state probability that there are n customers in the system at any time using the Birth-Death methodology and obtain the queue characteristics as under:

$$p_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \; p_o = \frac{\frac{\lambda}{1^2} \frac{\lambda}{2^2} \cdots \frac{\lambda}{n^2} \, p_o}{\mu \, \mu \ldots \mu}$$

$$= \frac{\rho^n}{(n\,!)^2} \; p_o \, , \; \rho = \frac{\lambda}{\mu}$$

Using $\Sigma p_n = 1$ gives $p_o = \frac{1}{I_o(2\sqrt{\rho})}$ where $I_o(2\sqrt{\rho})$ is the modified Bessel function of the first kind of order zero.

Hence

$$p_n = \frac{\rho^n}{(n!)^2 \, I_o(2\sqrt{\rho})}, \; n \geq 0 \tag{5.2.1}$$

Expected number of customers in the system,

$$\begin{aligned} L_S = E(N) &= \sum_{n=0}^{\infty} n \cdot \frac{\rho^n}{(n!)^2 \, I_o(2\sqrt{\rho})} \\ &= \frac{1}{I_o(2\sqrt{\rho})} \sum_{m=0}^{\infty} \frac{\rho^{m+1}}{(m!)(m+1)!} \\ &= \frac{\sqrt{\rho}}{I_o(2\sqrt{\rho})} \Sigma \frac{\left(\sqrt{\rho}\right)^{2m+1}}{m!(m+1)!} \\ &= \frac{\sqrt{\rho}}{I_o(2\sqrt{\rho})} I_1(2\sqrt{\rho}) \end{aligned} \tag{5.2.2}$$

where $I_m(x)$ is the modified Bessel function of the first kind of order $m = 0, 1, 2, \ldots$

Effective arrival rate, (Average number of customers who actually entering the system) λ^*, is given by

$$\begin{aligned}\lambda^* &= \sum_0^\infty \lambda_n\, p_n \\ &= \Sigma \frac{\lambda}{(n+1)^2}\frac{\rho^n}{(n!)^2\, I_o(2\sqrt{\rho})} \\ &= \frac{\lambda}{I_o(2\sqrt{\rho})}\Sigma\frac{\rho^n}{(n+1)!\,(n+1)!} \\ &= \frac{\lambda}{I_o(2\sqrt{\rho})}\frac{\left[I_o(2\sqrt{\rho})-1\right]}{\rho}\end{aligned} \qquad (5.2.3)$$

Using Little's formula, expected waiting time of an arrival in the system is given by,

$$W_S = \frac{L_S}{\lambda^*} = \frac{\sqrt{\rho}\cdot I_1\left(2\sqrt{\rho}\right)}{\mu\left[I_o\left(2\sqrt{\rho}\right)-1\right]}$$

$$\mu W_S = \frac{\sqrt{\rho}\cdot I_1\left(2\sqrt{\rho}\right)}{I_o\left(2\sqrt{\rho}\right)-1} \qquad (5.2.4),$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system. It may be proved that μW_s is a value greater than and increases with the increasing values of ρ.

Proof:

$$\mu W_s = \frac{\sqrt{\rho} I_1(2\sqrt{\rho})}{I_0(2\sqrt{\rho})-1}$$

$$= \frac{\sqrt{\rho}\sum\limits_{k=0}^{\infty}(\sqrt{\rho})^{2k+1}}{k!(k+1)!} \Bigg/ \frac{\sum\limits_{k=0}^{\infty}(\sqrt{\rho})^{2k}}{k!k!}$$

$$= \left(1+\frac{\rho}{1!2!}+\frac{\rho^2}{2!3!}+\ldots\right)\left\{1-\left(1+\frac{\rho}{2!2!}+\frac{\rho^2}{3!3!}+\ldots\right)+\left(1+\frac{\rho}{2!2!}+\frac{\rho^2}{3!3!}+\ldots\right)^2+\ldots\right\}$$

Clearly this is a value greater than one for any ρ which increases with increasing values of ρ. This result is also verified by means of numerical evaluation.

Effective service rate, μ^*, (weighted averaged service rate) is given by

$$\mu^* (1 - p_o) = \sum_{n=1}^{\infty} \mu_n \, p_n = \mu\left(\sum_{n=1}^{\infty} p_n \right) = \mu(1 - p_o) \text{ , yielding} \tag{5.2.5}$$

$$\mu^* = \mu$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \frac{I_o\left(2\sqrt{\rho}\right) - 1}{I_o\left(2\sqrt{\rho}\right)} = 1 - p_o \tag{5.2.6}$$

which is the average number of customers at the server. It is also the proportion of time the server is working.

We shall now obtain an expression for the pdf of waiting time under FCFS queue discipline. Since service rate is a constant exponential rate μ with the pdf $\mu e^{-\mu t}$, the pdf of waiting time may easily be derived. Assume that the customers are served in the order of their arrival. Let T be the time spent in a system by an arrival.

Then the pdf of T is given by $f(t) = \sum_{n=0}^{\infty} \frac{\lambda_n \, p_n}{\lambda^*} \frac{\mu e^{-\mu t} (\mu t)^n}{n!}$, $t > 0$, so that $\int_0^{\infty} f(t) dt = 1$.

i.e., $$f(t) = \frac{\lambda e^{-\mu t}}{(1 - p_0) I_0 (2\sqrt{\rho})} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{(n+1)!(n+1)! \; n!} \tag{5.2.7}$$

Using $f(t)$, expected waiting time,

$$W_S = E(T) = \int_0^\infty t.f(t)dt$$

$$= \frac{\lambda}{(1-p_0)I_0(2\sqrt{\rho})} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n+1)!(n+1)!\ n!} \int_0^\infty e^{-\mu t} t^{n+1} dt$$

$$= \frac{\lambda}{(1-p_0)I_0(2\sqrt{\rho})} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n+1)!(n+1)!\ n!} \left(\frac{1}{\mu^{n+2}}\right) \Gamma(n+2)$$

where $\Gamma(n+2)$ is the gamma function defined by $\Gamma(n+2) = \int_0^\infty \bar{e}^s s^{n+1}\, ds$

i.e., $W_S = \frac{1}{(1-p_0)I_0(2\sqrt{\rho})\mu} \sum_{n=0}^{\infty} \frac{\rho^{n+1}}{(n+1)!\ n!}$, since $\Gamma n + 2 = (n+1)!$

$$= \frac{\sqrt{\rho}}{\mu(1-p_0)I_0(2\sqrt{\rho})} \sum_{n=0}^{\infty} \frac{(\sqrt{\rho})^{2n+1}}{(n+1)!\ n!}$$

$$= \frac{\sqrt{\rho} I_1(2\sqrt{\rho})}{\mu(1-p_0)I_0(2\sqrt{\rho})}$$

Hence

$$\mu W_S = \frac{\sqrt{\rho} I_1(2\sqrt{\rho})}{[I_0(2\sqrt{\rho}) - 1]} \tag{5.2.8}$$

which agrees the result obtain in (5.2.4) by Little's formula

Effective service rate

$$\mu^* = \frac{\lambda^*}{(1-p_0)} = \mu \tag{5.2.9}$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \frac{I_o\left(2\sqrt{\rho}\right)-1}{I_o\left(2\sqrt{\rho}\right)} = 1 - p_o \qquad (5.2.10)$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system. It may be proved that μW_S is greater than one for any ρ and increases with increasing value of ρ.

Now,

$$E(N^2) = \Sigma[n(n-1)+n]\,p_n$$

$$= \sum_{2}^{\infty} n(n-1)\cdot\frac{\rho^n}{(n!)^2\, I_o\left(2\sqrt{\rho}\right)} + L_S$$

$$= \frac{1}{I_o\left(2\sqrt{\rho}\right)} \sum_{m=0}^{\infty} \frac{\rho^{m+2}}{m!\,(m+2)!} + L_S$$

$$= \frac{\rho}{I_o\left(2\sqrt{\rho}\right)} \Sigma \frac{\left(\sqrt{\rho}\right)^{2m+2}}{m!\,(m+2)!} + L_S$$

$$= \frac{\rho}{I_o\left(2\sqrt{\rho}\right)}\cdot I_2\left(2\sqrt{\rho}\right) + \sqrt{\rho}\cdot\frac{I_1\left(2\sqrt{\rho}\right)}{I_o\left(2\sqrt{\rho}\right)}$$

where $I_o\left(2\sqrt{\rho}\right)$ is the modified Bessel function of the first kind of order 2.

$Var(N) = E(N^2) - [E(N)]^2$

$$= \frac{\rho}{I_o\left(2\sqrt{\rho}\right)} I_2\left(2\sqrt{\rho}\right) + \frac{\sqrt{\rho}\cdot I_1\left(2\sqrt{\rho}\right)}{I_o\left(2\sqrt{\rho}\right)} - \left\{\frac{\sqrt{\rho}\; I_1\left(2\sqrt{\rho}\right)}{I_o\left(2\sqrt{\rho}\right)}\right\}^2 \qquad (5.2.10)$$

which is the variance of the number of customers in the system in the steady state

Table I: Numerical evaluation of the queue characteristics of the Model BI in (5.2) for various $\rho = \frac{\lambda}{\mu}$

ρ	p_o	L_S	σ_N	μW_S	λ^*	μ^*	ρ^*
0.1	0.8994	0.0941	0.3012	1.0279	1.0056 λ	μ	0.1006
0.2	0.8277	0.1816	0.4076	1.0543	0.8613 λ	μ	0.1723
0.5	0.6400	0.4052	0.5784	1.1257	0.7199 λ	μ	0.3599
0.9	0.4699	0.6456	0.6958	1.2180	0.5890 λ	μ	0.5301
0.99	0.4386	0.6941	0.7153	1.2364	0.5670 λ	μ	0.5614
1	0.4386	0.6976	0.7162	1.2427	0.5614 λ	μ	0.5614

5.3 MODEL BII

The non-linear state dependent Markovian queueing model characterized by: $\lambda_n = (n+1)^2 \lambda,\ n \geq 0, \mu_n = n^2 \mu,\ n \geq 1\ (\lambda, \mu > 0)$

We shall find, p_n, the steady state probability that there are n customers in the system using the Birth-Death methodology and obtain the queue characteristics as under:

$$p_n = \frac{\lambda \cdot 2^2 \cdot \lambda \cdot 3^2 \lambda \ldots . n^2 \lambda}{1^2 \cdot \mu \cdot (2^2 \mu) \ldots . (n^2 \mu)}\ p_o = \rho^n\ p_o,\ \rho = \frac{\lambda}{\mu}$$

$\Sigma\ p_n = 1$ gives $p_o = (1 - \rho)$

Hence,

$$p_n = (1-\rho)\,\rho^n, \quad n \geq 0 \tag{5.3.1}$$

which is a Geometric distribution with parameter $\rho < 1$.

System size at any time,

$$L_S = \Sigma n\, p_n = \Sigma (1-\rho)\, n\, \rho^n$$

$$= (1-\rho)\left\{1\cdot\rho + 2\cdot\rho^2 + 3\rho^3 + ...\right\}$$

$$= (1-\rho)\cdot\rho\,(1+2\rho+3\rho^2 ....) = \frac{\rho}{(1-\rho)} \tag{5.3.2}$$

Effective arrival rate (Average number of customers actually entering the system)

$$\lambda^* = \Sigma\,\lambda_n\,p_n = \Sigma (n+1)^2 \lambda\cdot(1-\rho)\,\rho^n$$
$$= \lambda\Sigma\,[n\,(n-1)+3n+1]\,(1-\rho)\,\rho^n$$
$$= (1-\rho)\lambda\,\{[2\times 1\rho^2 + 3.2\rho^3 + ...] + 3\,[1\cdot\rho^1 + 2\rho^2 + 3\rho^3 + ...] + [1+\rho+\rho^2+...]\}$$
$$= \lambda(1-\rho)\left\{\frac{2\rho^2}{(1-\rho)^3} + \frac{3\rho}{(1-\rho)^2} + \frac{1}{1-\rho}\right\}$$

$$= \lambda\left\{\frac{2\rho^2 + 3\rho(1-\rho) + (1-\rho)^2}{(1-\rho)^2}\right\} = \lambda\frac{(1+\rho)}{(1-\rho)^2} \tag{5.3.3}$$

Expected waiting time, W_S, of an arrival in the steady state is given by

$$W_S = \frac{L_S}{\lambda^*} \quad \text{(By Little's formula)}$$

$$= \frac{1-\rho}{(1+\rho)\,\mu}$$

$$\mu W_S = \frac{1-\rho}{1+\rho} = 1 - 2\rho + 2\rho^2 \quad \text{(approximately)} \qquad (5.3.4),$$

which is the number of customers served during the waiting time of an arbitrary customer and this number is always less than one for any ρ and decreases with increasing values of ρ.

Effective service rate, μ^*, is given by

$$\mu^* (1-p_o) = \sum_1^{\infty} \mu_n \, p_n$$
$$= \Sigma \, n^2 \cdot \mu \cdot (1-\rho) \, \rho^n$$
$$= \mu(1-\rho) \sum_1^{\infty} \left\{ \frac{2\rho^2}{(1-\rho)^3} + \frac{\rho}{(1-\rho)^2} \right\} = \frac{\mu \rho (1+\rho)}{(1-\rho)^2} \quad \text{, yielding}$$

$$\mu^* = \frac{\mu \rho (1+\rho)}{(1-\rho)^2 \cdot (1-p_o)} = \frac{\mu (1+\rho)}{(1-\rho)^2}$$

or
$$\mu^* = \frac{\lambda^*}{1-p_0} = \frac{\mu(1+\rho)}{(1-\rho)^2} \tag{5.3.5}$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \rho = 1 - p_o \qquad (5.3.6),$$

which is the average number of customers at the server. It is also the proportion of time the server is working.

Now,
$$E(N^2) = \Sigma [n(n-1) + n] \, p_n$$
$$= \Sigma \, n(n-1) \cdot (1-\rho) \rho^n \cdot + \Sigma n \cdot p_n$$
$$= (1-\rho) \left\{ 2 \cdot \rho^2 + (3) 2 \rho^3 + (4) 3 \rho^4 + \ldots \right\} + L_S$$
$$= (1-\rho) \, 2\rho^2 \left\{ 1 + 3\rho + 6\rho^2 + \ldots \right\} + L_S$$
$$= (1-\rho) \, 2\rho^2 \cdot \frac{1}{(1-\rho)^3} + \frac{\rho}{1-\rho} = \frac{\rho^2 + \rho}{(1-\rho)^2}$$

Hence

$$Var(N) = E(N^2) - E[(N)]^2 = \frac{\rho}{(1-\rho)^2} \tag{5.3.7}$$

which is the variance of number of customers in the system in the steady state .

Table II: Numerical evaluation of the queue characteristics of the Model BII in (5.3) for various $\rho = \frac{\lambda}{\mu}$ (0< ρ< 1)

ρ	p_o	L_S	σ_N	μW_S	λ^*	μ^*	ρ^*
0.1	0.9	0.1111	0.3514	0.8182	1.3580 λ	1.3580 μ	0.1
0.2	0.8	0.25	0.5590	0.6667	1.8750 λ	1.8750 μ	0.2
0.5	0.5	1.00	1.4142	0.3333	6.000 λ	6 μ	0.5
0.9	0.1	9.00	9.4868	0.0526	190 λ	190 μ	0.9
0.99	0.01	99.00	99.4987	0.0050	19900 λ	19900 μ	0.99

5.4 MODEL BIII

The non-linear state dependent Markovian queueing model characterized by:

$$\lambda_n = \lambda,\ n \geq 0, \mu_n = n^2\mu,\ n \geq 1\ (\lambda, \mu > 0)$$

We shall find, p_n, the steady state probability that there are n customers in the system, using the Birth-Death methodology and obtain the queue characteristics as under:

$$p_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} p_o = \frac{\rho^n}{(n!)^2} p_o, \ \rho = \frac{\lambda}{\mu}$$

Using $\Sigma\, p_n = 1$ gives $p_o = \frac{1}{I_o\left(2\sqrt{\rho}\right)}$ where $I_o\left(2\sqrt{\rho}\right)$ is the modified Bessel function of the first kind of order zero

Hence,

$$p_n = \frac{\rho^n}{(n!)^2\, I_o\left(2\sqrt{\rho}\right)}, \ n \geq 0 \tag{5.4.1}$$

Expected number of customers in the system,

$$\begin{aligned} L_S = \Sigma\, n\, p_n &= \sum_{n=0}^{\infty} \frac{n \cdot \rho^n}{(n!)^2\, I_o(2\sqrt{\rho})} \\ &= \frac{\sqrt{\rho}}{I_o\,(2\sqrt{\rho})}\, I_1(2\sqrt{\rho}) \end{aligned} \tag{5.4.2}$$

It may be proved that L_S increases with the increasing values of ρ.

Proof:

$$L_S = \frac{\sqrt{\rho}}{I_o\left(2\sqrt{\rho}\right)}\, I_1\left(2\sqrt{\rho}\right)$$

$$= \frac{\sqrt{\rho}\displaystyle\sum_{m=0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2m+1}}{m!(m+1)!}}{\displaystyle\sum_{k=0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2k}}{k!k!}}$$

$$= \frac{\sqrt{\rho}\left[\frac{\sqrt{\rho}}{1!} + \frac{\left(\sqrt{\rho}\right)^5}{2!3!} + \ldots\right]}{[1 + \frac{\left(\sqrt{\rho}\right)^2}{1!} + \frac{\left(\sqrt{\rho}\right)^4}{2!2!} + \ldots]}$$

$= \rho(1-\rho-\frac{\rho^2}{6})$, which is a value less than one and increases with the increase in value of ρ.

Effective arrival rate, λ^*, (Average number of customers who actually entering the system) is given by

$$\lambda^* = \Sigma\lambda_n\, p_n = \Sigma\lambda \cdot p_n = \lambda \tag{5.4.3}$$

Effective service rate, μ^*, (weighted average service rate) is given by

$$\mu^*(1-p_o) = \sum_1^\infty \cdot \mu_n\, \rho_n$$

$$= \sum_1^\infty \cdot n^2 \mu \cdot \frac{\rho^n}{(n!)^2 \cdot I_o(2\sqrt{\rho})}$$

$$= \frac{\mu}{I_o(2\sqrt{\rho})} \sum_{m=0}^{\infty} \frac{\rho^{m+1}}{m!\,m!}$$

$$= \frac{\mu\rho}{I_o(2\sqrt{\rho})} \sum_{m=0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2m}}{m!m!}$$

$$= \frac{\lambda}{I_o(2\sqrt{\rho})} I_o\left(2\sqrt{\rho}\right), \text{ yielding}$$

$$\mu^* = \frac{\lambda}{1 - \frac{1}{I_0(2\sqrt{\rho})}} = \frac{\lambda \cdot I_0(2\sqrt{\rho})}{I_o(2\sqrt{\rho}) - 1}$$

or

$$\mu^* = \frac{\lambda^*}{1-p_0} = \frac{\lambda \cdot I_o\left(2\sqrt{\rho}\right)}{I_o(2\sqrt{\rho})-1} \tag{5.4.4}$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \frac{I_0\left(2\sqrt{\rho}\right)-1}{I_0(2\sqrt{\rho})} = 1 - p_o \qquad (5.4.5),$$

which is the average number of customers at the server. It is also the proportion of time the server is working.

Expected waiting time of an arrival in the steady state is given by,

$$W_S = \frac{L_S}{\lambda^*} = \frac{\sqrt{\rho}\, I_1\left(2\sqrt{\rho}\right)}{I_o(2\sqrt{\rho})\cdot\lambda}$$

Hence

$$\mu W_S = \frac{I_1\left(2\sqrt{\rho}\right)}{I_o(2\sqrt{\rho})\cdot\sqrt{\rho}} \qquad (5.4.6),$$

which is the number of customers served during the waiting time of an arbitrary customer and this number is always less than 1 which decreases with the increasing values of ρ.

Proof

$$\mu W_S = \frac{I_1\left(2\sqrt{\rho}\right)}{I_o((2\sqrt{\rho})\cdot\sqrt{\rho}}$$

$$= \frac{\dfrac{\sum_{m=0}^{\infty}(\sqrt{\rho})^{2m+1}}{m!\,(m+1)!}}{\sqrt{\rho}\sum_{m=0}^{\infty}\dfrac{(\sqrt{\rho})^{2m}}{m!\,m!}}$$

$$= \frac{1 + \frac{(\sqrt{\rho})^2}{1!\,2!} + \frac{(\sqrt{\rho})^4}{2!\,3!} + \dots}{\left\{1 + \frac{(\sqrt{\rho})^2}{1!\,1!} + \frac{(\sqrt{\rho})^4}{2!\,2!} + \dots\right\}}$$

$$= \left(1 + \frac{(\sqrt{\rho})^2}{1!\,2!} + \frac{(\sqrt{\rho})^4}{2!\,3!} + \dots\right)\left\{1 + \left(\frac{(\sqrt{\rho})^2}{1!\,1!} + \frac{(\sqrt{\rho})^4}{2!\,2!} + \dots\right)\right\}^{-1}$$

$$= \left(1 + \frac{\rho}{2} + \frac{\rho^2}{12} + \dots\right)\left(1 - \rho - \frac{\rho^2}{4} - \dots\right)$$

$$= 1 - \frac{\rho}{2} - \frac{2}{3}\rho^2 \text{ (approximately)},$$

which is clearly less than one and decreases with increasing values of ρ.

Now

$$E(N^2) = \Sigma[n(n-1)+n]\, p_n$$

$$= \Sigma[n(n-1)+n]\, \frac{\rho^n}{(n!)^2\, I_o(2\sqrt{\rho})}$$

$$= \frac{\rho}{I_o(2\sqrt{\rho})}\, \frac{\Sigma\left(\sqrt{\rho}\right)^{2m+2}}{m!\,(m+2)!} + L_S$$

$$= \frac{\rho}{I_o(2\sqrt{\rho})}\, I_2(2\sqrt{\rho}) + \sqrt{\rho} \cdot \frac{I_1\left(2\sqrt{\rho}\right)}{I_o(2\sqrt{\rho})}$$

Hence,

$$Var(N) = E(N^2) - [E(N)]^2$$

$$= \frac{\rho \cdot I_2(2\sqrt{\rho})}{I_o(2\sqrt{\rho})} + \sqrt{\rho}\,\frac{I_1(2\sqrt{\rho})}{I_o(2\sqrt{\rho})} - \left\{\sqrt{\rho}\,\frac{I_1(2\sqrt{\rho})}{I_o(2\sqrt{\rho})}\right\}^2 \qquad (5.4.7)$$

which is the variance of number of customers in the system in the steady state

Table III: Numerical evaluation of the queue characteristics of the Model BIII in (5.4) for various $\rho = \frac{\lambda}{\mu}$

ρ	p_o	L_S	μW_S	σ_N	λ^*	μ^*	ρ^*
0.1	0.8994	0.0941	0.9412	0.3012	λ	9.9445 λ	0.1006
0.2	0.8277	0.1816	0.9082	0.4076	λ	5.8054 λ	0.1723
0.5	0.6400	0.4052	0.8104	0.5784	λ	2.7781 λ	0.3600
0.9	0.4699	0.6456	0.7174	0.6958	λ	1.8865 λ	0.5301
0.99	0.4386	0.6941	0.7011	0.7153	λ	1.7813 λ	0.5614
1	0.4386	0.6976	0.6976	0.7162	λ	1.7813 λ	0.5614

5.5 MODEL BIV

The non-linear state dependent Markovian queueing model characterized by:

$$\lambda_n = (n+1)^2\,\lambda,\ n \ge 0, \mu_n = 2n^2\mu,\ n \ge 1\ (\lambda, \mu > 0)$$

We shall find, p_n, the steady state probability that there are n customers in the system using the Birth-Death methodology and obtain the queue characteristics as under:

$$p_n = \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} p_o = \frac{1^2 \lambda 2^2 \lambda \ldots n^2 \lambda \;\; p_o}{2(1^2 \mu) 2(2^2 \mu) \ldots 2(n^2 \mu)}$$

$$= \left(\frac{\lambda}{2\mu}\right)^n p_o = \left(\frac{\rho}{2}\right)^n p_o, \rho = \frac{\lambda}{\mu}$$

$\Sigma p_n = 1$ gives $$p_o = \left(1 - \frac{\rho}{2}\right), \quad \frac{\rho}{2} < 1$$

Hence,

$$p_n = \left(1 - \frac{\rho}{2}\right)\left(\frac{\rho}{2}\right)^n, n \geq 0 \; and \; 0 < \rho < 2 \tag{5.5.1}$$

Expected number of customers in the system,

$$L_S = E(N) = \Sigma n \cdot p_n = \Sigma\, n\left(1 - \frac{\rho}{2}\right)\left(\frac{\rho}{2}\right)^n$$

$$= \left(1 - \frac{\rho}{2}\right)\left\{\left(\frac{\rho}{2}\right) + 2\left(\frac{\rho}{2}\right)^2 + \ldots\right\}$$

$$= \left(1 - \frac{\rho}{2}\right) \cdot \frac{\frac{\rho}{2}}{\left(1 - \frac{\rho}{2}\right)^2} = \frac{\frac{\rho}{2}}{1 - \frac{\rho}{2}} = \frac{\rho}{2 - \rho}$$

$$= \frac{\rho}{2} + \frac{\rho^2}{4} \text{ (approximately)} \tag{5.5.2}$$

which is a value less than one and decreases with increasing values of ρ.

Effective arrival rate λ^* is given by

$$\lambda^* = \Sigma \lambda_n \cdot p_n$$

$$= \sum_{n=0}^{\infty} (n+1)^2 \lambda \cdot \left(1-\frac{\rho}{2}\right)\left(\frac{\rho}{2}\right)^w$$

$$= \lambda\left(1-\frac{\rho}{2}\right)\left\{[n(n-1)+3n+1]\left(\frac{\rho}{2}\right)^n\right\}$$

$$= \lambda\left(1-\frac{\rho}{2}\right)\left\{\frac{2\left(\frac{\rho}{2}\right)^2}{\left(1-\frac{\rho}{2}\right)^3}+\frac{3\left(\frac{\rho}{2}\right)}{\left(1-\frac{\rho}{2}\right)^2}+\frac{1}{\left(1-\frac{\rho}{2}\right)}\right\}$$

$$= \frac{\lambda\left(1+\frac{\rho}{2}\right)}{\left(1-\frac{\rho}{2}\right)^2} = \frac{2\lambda(2+\rho)}{(2-\rho)^2} \tag{5.5.3}$$

The expected waiting time, W_S, of an arrival in the steady state is given by

$$W_S = \frac{L_S}{\lambda^*} = \frac{1}{2\mu}\left(\frac{2-\rho}{2+\rho}\right)$$

$$\mu W_S = \frac{1}{2}\left(\frac{2-\rho}{2+\rho}\right) \tag{5.5.4}$$

which is the number of customers served during the waiting time of an arbitrary arrival and this number is always less than one which decrease with increasing values of ρ.

Proof:

$$\mu W_s = \frac{1}{2}\left(1-\frac{\rho}{2}\right)\left(1+\frac{\rho}{2}\right)^{-1}$$

$$= \frac{1}{2}\left(1-\frac{\rho}{2}\right)\left(1-\frac{\rho}{2}+\frac{\rho^2}{4}-\dots\right)$$

$$= \frac{1}{2} - \frac{\rho}{2} + \frac{\rho^2}{4} \quad \text{(approximately)}$$

which is clearly a value less than one which decreases with increasing values of ρ.

Effective service rate, μ^*, (Weighted average service time) is given by

$$\mu^* (1 - p_o) = \sum_{n=1}^{\infty} 2n^2 \mu \cdot \left(1 - \frac{\rho}{2}\right)\left(\frac{\rho}{2}\right)^n$$

$$= 2\mu\left(1 - \frac{\rho}{2}\right) \sum_{1}^{\infty} [n(n-1) + n] \left(\frac{\rho}{2}\right)^n$$

$$= 2\mu\left(1 - \frac{\rho}{2}\right) \left\{ \begin{array}{l} 2(1)\left(\frac{\rho}{2}\right)^2 + 3(2)\left(\frac{\rho}{2}\right)^3 + \ldots\ldots \\ + \left[\frac{\rho}{2} + 2\left(\frac{\rho}{2}\right)^2 + 3\left(\frac{\rho}{2}\right)^3 + \ldots\ldots\right] \end{array} \right\}$$

$$= 2\mu\left(1 - \frac{\rho}{2}\right) \left\{ \frac{2\left(\frac{\rho}{2}\right)^2}{\left(1 - \frac{\rho}{2}\right)^3} + \frac{\frac{\rho}{2}}{\left(1 - \frac{\rho}{2}\right)^2} \right\}$$

$$= \frac{2\lambda(\rho + 2)}{(2 - \rho)^2}$$

Hence,

$$\mu^* = \frac{2\lambda(\rho + 2)}{(2 - \rho)^2 (1 - p_0)}$$

Or

$$\mu^* = \frac{\lambda^*}{(1 - p_0)} = \frac{2\lambda(\rho + 2)}{(2 - \rho)^2 (1 - p_0)} \qquad (5.5.5)$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \frac{\rho}{2} = 1 - p_o \qquad (5.5.6)$$

which is the mean number of customers at the server. It is also the proportion of the time the server is working.

Now, $E(N^2) = \Sigma[n(n-1)+n]\,p_n$

$$= \Sigma[n(n-1)\cdot\left(1-\frac{\rho}{2}\right)\left(\frac{\rho}{2}\right)^n + L_S$$

$$= \left(1-\frac{\rho}{2}\right)\left\{2(1)\left(\frac{\rho}{2}\right)^2 + 3(2)\left(\frac{\rho}{2}\right)^3 + \ldots\right\} + L_S$$

$$= \left(1-\frac{\rho}{2}\right)2\left(\frac{\rho}{2}\right)^2\left\{1+3\left(\frac{\rho}{2}\right)+6\left(\frac{\rho}{2}\right)^2 + \ldots\right\} + L_S$$

$$= \frac{2\rho^2}{(2-\rho)^2} + \frac{\rho}{2-\rho} = \frac{2\rho+\rho^2}{(2-\rho)^2}$$

Hence,

$$Var(N) = E(N^2) - [E(N)]^2$$

i.e.,
$$Var(N) = \frac{2\rho}{(2-\rho)^2} \qquad (5.5.7)$$

which is the variance of number of customers in the system in the steady state.

Table IV: Numerical evaluation of the queue characteristics of the Model BIV in (5.5) for various $\rho = \frac{\lambda}{\mu}$ $(0 < \rho < 2)$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*	σ_N
0.1	0.95	0.0526	0.4524	1.1634 λ	2.3269 μ	0.05	0.2354
0.2	0.90	0.1111	0.4091	1.3580 λ	2.7160 μ	0.1	0.3514
0.5	0.75	0.3333	0.3000	2.2222 λ	4.4444 μ	0.25	0.6667
0.9	0.55	0.8182	0.1897	4.7934 λ	9.5868 μ	0.45	1.2197
0.99	0.505	0.9802	0.1689	5.8622 λ	11.7243 μ	0.495	1.3932
1	0.5	1	0.1667	6 λ	12 μ	0.5	1.4142
1.5	0.25	3	0.0714	28λ	56μ	0.75	3.4641
1.9	0.05	19	0.0128	780λ	1560μ	0.95	19.4936
1.99	0.005	199	0.0013	79800λ	159600μ	0.995	199.4994

5.6 MODEL BV:

Non linear state-dependent (NLSD) model characterised by the arrival rate

$\lambda_n = \dfrac{\lambda}{(n+1)}$, $n \geq 0$; service rate $\mu_n = \mu$, $n \geq 1$ ($\lambda > 0$, $\mu > 0$)

We have discussed this model in chapter 2. But as a non-linear state-dependent queueing model, we study the model in this chapter also in order to compare this model with that of the other non linear state-dependent queueing models considered.

This model may be considered as *M/M/*1 model with balking in which an arriving customer might be discouraged by the number of customers already in the system and thus do not enter.

As in (2.3.3), the steady state probability distribution of queue length,

$$p_n = \frac{e^{-\rho}\,\rho^n}{n\,!}, \ \rho = \frac{\lambda}{\mu}; \quad where \ n \geq 0 \tag{5.6.1}$$

which is again a Poisson distribution with parameter $\rho > 0$.

System size at any time, $L_S = \Sigma n \ p_n = \rho$ (5.6.2)

Effective arrival rate,

$$\lambda^* = \Sigma \lambda_n \cdot p_n = \mu\,(1 - e^{-\rho}) \tag{5.6.3}$$

Expected waiting time of an arbitrary arrival in the system is given by

$$W_S = \frac{L_S}{\lambda^*} = \frac{\rho}{(1-e^{-\rho})\,\mu} \qquad \text{(by Little's formula)}$$

Hence

$$\mu W_S = \frac{\rho}{1-e^{-\rho}} = 1+\frac{\rho}{2}+\frac{\rho^2}{12} \text{ (approximately)} \tag{5.6.4}$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system. Clearly this is a value greater than one and increases with the increasing value of ρ. This result is also verified by numerical evaluation.

Proof:

$$\mu W_s = \frac{\rho}{1-e^{-\rho}}$$

$$= \left[1-\left(\frac{\rho}{2!}-\frac{\rho^2}{3!}+\ldots\right)\right]^{-1}$$

$$= 1+\left(\frac{\rho}{2!}-\frac{\rho^2}{3!}+\ldots\right)+\left(\frac{\rho}{2!}-\frac{\rho^2}{3!}+\ldots\right)^2+\ldots.$$

$$= 1+\frac{\rho}{2}+\frac{\rho^2}{12} \text{ (approximately)}$$

$$> 1 \text{ for any } \rho$$

which increases with increasing values of ρ.

Effective service rate,

$$\mu^* = \frac{\lambda^*}{(1-p_0)} = \mu \tag{5.6.5}$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = 1 - e^{-\rho} = 1 - p_o$$

$$= \rho - \frac{\rho^2}{2} + \frac{\rho^3}{6} \text{ (approximately)} \qquad (5.6.6)$$

which is the number of customers at the server during his service time. It is also the proportion of time the server is working. This is a value less than one and increases with the increasing values of ρ. This result is also verified by numerical evaluation.

Table V: The performance measures of the Model BV in (5.6)

for various $\rho = \frac{\lambda}{\mu}$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9048	0.1	1.0504	0.0952μ	μ	0.0952
0.2	0.8187	0.2	1.1031	$0.1813\ \mu$	μ	0.1813
0.5	0.6065	0.5	1.2707	$0.3935\ \mu$	μ	0.3935
0.9	0.4066	0.9	1.5167	$0.5934\ \mu$	μ	0.5934
0.99	0.3716	0.99	1.5754	$0.6284\ \mu$	μ	0.6284
1.00	0.3679	1.00	1.5820	$0.6321\ \mu$	μ	0.6321

5.7 MODEL BVI:

Nonlinear state-dependent (NLSD) model characterized by the arrival rate $\lambda_n = (n+1)\lambda$, $n \geq 0$; service rate $\mu_n = n^2\mu$, $n \geq 1$ ($\lambda > 0$, $\mu > 0$).

This model is a completely state-dependent Markovian queueing model with the abundant service rate and linear arrival rate.

We shall derive p_n, the steady state probability distribution of queue length and obtain the queue characteristics as under:

By using Birth-Death methodology,

$$p_n = \frac{\lambda(2\lambda)\ldots(n\lambda)\ p_o}{(1^2\mu)\ (2^2\mu)\ldots(n^2\mu)} = \frac{\rho^n\ p_o}{n\,!}\ ,\quad \rho = \frac{\lambda}{\mu}$$

Using $\Sigma\, p_n = 1$ gives $\quad p_o = e^{-\rho}$ (5.7.1)

Hence,

$$p_n = \frac{e^{-\rho}\ \rho^n}{n\,!}\ ,\quad n \geq 0 \tag{5.7.2}$$

which is Poisson distribution with parameter $\rho > 0$.

Equation (5.7.2) is also probability distribution of queue length in the steady state for $M/M/\infty$ model (self service model) which is a model often used to analyse manufacturing process and to model phenomena in telecommunication networks.

System size at any time,

$$L_S = \Sigma n\ p_n\ = \rho \tag{5.7.3}$$

Effective arrival rate,

$$\lambda^* = \sum_0^\infty \lambda_n \; p_n \; = \sum_0^\infty (n+1) \; \lambda \cdot p_n = \lambda(\rho+1) \qquad (5.7.4)$$

Expected waiting time of an arrival is given by

$$W_S = \frac{L_S}{\lambda^*} = \frac{\rho}{\lambda \; (\rho+1)} \quad \text{(by Little's formula)}$$

Hence,

$$\mu W_S = \frac{1}{1+\rho} = 1-\rho+\rho^2-\rho^3 \text{ (approximately)} \qquad (5.7.5)$$

which is clearly less than one and decreases with increasing values of ρ. This result is also verified by numerical evaluation. Equation (5.7.5) gives the number of customers served during the waiting time of an arbitrary arrival.

Effective service rate

$$\mu^* = \frac{\lambda^*}{(1-p_0)} = \frac{\lambda(1+\rho)}{(1-p_0)} \qquad (5.7.6)$$

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = 1- \; e^{-\rho} = 1-p_o \qquad (5.7.7)$$

which is the number of customers at the server during his service time. It is also the proportion of time the server is working.

Table VI: The performance measures of the Model BVI in (5.7)

for various $\rho = \frac{\lambda}{\mu}$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9048	0.1	0.9091	1.1 λ	1.1555 μ	0.0952
0.2	0.8187	0.2	0.8333	1.2 λ	1.3238 μ	0.1813
0.5	0.6065	0.5	0.6667	1.5 λ	1.9060 μ	0.3935
0.9	0.4066	0.9	0.5263	1.9 λ	2.8817 μ	0.5934
0.99	0.3716	0.99	0.5025	1.99 λ	3.1351 μ	0.6284
1.00	0.3679	1.00	0.5000	2.00 λ	3.1641 μ	0.6321

5.8 SUMMARY OF RESULTS AND CONCLUDING REMARKS

1. With reference to the measure p_o, the state probability of an empty system, it is verified that Model (5.2) is better than the other non linear state-dependent model considered.

2. The probability distribution of queue length for the NLSD Models (5.6) and (5.7) are same which follows Poisson distribution with parameter ρ. i.e., there are non linear state-dependent Markovian queueing models having the same probability distribution of queue length.

3. From the model (5.3), it may be verified that the steady state probability distribution of queue length of the NLSD queueing models characterized by $\lambda_n = (n + 1)^m \lambda, \quad n \geq 0; \mu_n = n^m \mu, \quad n \geq 1,$

where $m \geq 2$, $\lambda > 0$, $\mu > 0$ follows Geometric distribution given by $p_n = (1-\rho)\rho^n$, $n \geq 0$ and $0 < \rho < 1$, which is the same as the steady-state probability distribution of queue length for the M/M/1 Model.

4. With reference to the measure p_0, one can do better than the Models (5.2) (5.3), (5.4), (5.6) and (5.7). An example is provided by $p_0 = 1 - \frac{\rho}{2}$, $0 < \rho < 2$ which can be achieved by the Model (5.5) characterized by $\lambda_n = (n+1)^2 \lambda$, $n \geq 0$; $\mu_n = 2\, n^2 \mu$, $n \geq 1$, which is a model very close to the non linear state-dependent queueing model considered in this chapter.

5. Considering the model (5.5), for fixed service rate μ, an increase in ρ calls for a reduction in the proportion of the time actually the server is working than the models (5.1), (5.2), (5.3), (5.4), (5.6) and (5.7).

6. It is verified that the effective traffic intensity ρ^* of the better model is exactly half of ρ^* of the better state-dependent queueing model considered in (5.3).

7. With reference to the measure p_0, the state probability of an empty system, one can do better than the model (5.7). An example is provided by $p_0 = \frac{1}{I_0(\sqrt{\rho})}$ where $I_0(\sqrt{\rho})$ is Modified Bessel function of order zero, which can be achieved by the model characterized by the arrival rate $\lambda_n = \frac{\lambda}{2n+2}$, $n \geq 0$, service rate

$\mu_n = 2\, n\mu, \quad n \geq 1$ where $\lambda > 0, \; \mu > 0$. It is to be noted that both models have the same service rate.

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

Chapter 6

THE WAITING TIME DISTRIBUTION OF MARKOVIAN QUEUEING MODELS WITH STATE-DEPENDENT ARRIVAL RATES – A COMPARISON

6.1 INTRODUCTION

In this chapter, we study some single server Markovian queueing models with state dependent arrival rates and constant service rate. We obtain the steady state probability distribution of queue length and waiting time distribution of these models. Some statistics of the waiting time distribution are also investigated and compared numerically thereby suggesting the relative supremacy of the models.

Queues with state dependent arrival rates have wide application in computers, communication systems, queueing networks etc. In recent years there has been great interest in obtaining queue length distribution of these models which are used in performance analysis of computer/communication systems.

We study some Markovian single service queueing models with state dependent arrival rates λ_n and constant service rate μ_n when the number of customers in the system is n. We obtain the steady state probability distribution of queue length. We shall also carry out a steady state analysis of the waiting time. Narayanan and Manoharan (2006) have considered state dependent Markovian Queueing Models with state dependent arrival rates and constant service time. Here the steady state behaviour of the model is discussed and obtained explicit expression for the steady state probability distribution of queue length and waiting time distribution.

The waiting time of an arrival is queueing time plus service time. The waiting time is merely the service time if the system state upon arrival is zero. Brain Conolly (1975) and Hadidi (1969) gave an expression for the waiting time distribution using direct analysis. However, the method used is more tedious and complicated. The moments of waiting time distribution can be found only by contour integration. Natvig (1975) derives the pdf of waiting time in the steady state of a queue model where potential customers are discouraged by queue length. He also compute the mean and variance of the waiting time.

In this section, we consider some Markovian single server queueing Models with state dependent arrival rates λ_n and service rate $\mu_n = \mu$, a constant when the number of customers in the system is n. We obtain the steady state probability distribution of queue length and waiting time distribution. Some statistics of the waiting time distribution are also evaluated.

6.2 MODEL I

The state dependent single server Markovian Queueing Models characterized by the arrival rate $\lambda_n = \frac{\lambda}{2(n+1)}$, $n \geq 0$, service rate $\mu_n = \mu$, $n \geq 1$ $(\lambda > 0, \mu > 0)$.

We have already discussed this model as Model AII in chapter 4. However, as a Markovian queueing model with the state-dependent arrival rates and for comparing some statistics of the waiting time distribution for the state-dependent queueing models, we consider this model in this chapter also.

As in (4.3.2), the steady state probability distribution of queue length is given by,

$$p_n = \frac{\bar{e}^{\rho/2}\left(\frac{\rho}{2}\right)^n}{n!}, \; n \geq 0 \text{ and } \rho = \frac{\lambda}{\mu} \tag{6.2.1}$$

Assume that customers are served in the order of their arrival (i.e., assume FCFS queue discipline). The duration of services are independent and identically distributed (i.i.d) random variables and independent of the arrival process. The service time distribution is exponential with mean $\frac{1}{\mu}$. Let T be the time spent by an arrival in the steady state. Then the pdf of the Waiting time T is given by

$$f(t) = \sum_{n=0}^{\infty} \frac{\lambda_n}{\lambda^*} p_n \frac{\mu\, e^{-\mu t} (\mu t)^n}{n!}, \text{ so that } \int_0^{\infty} f(t)\, dt = 1,$$

where effective arrival rate $\lambda^* = \sum_{n=0}^{\infty} \lambda_n\, p_n = \mu(1 - e^{-\rho/2})$ (6.2.2)

i.e., as in (4.3.7), $f(t) = \dfrac{\lambda e^{-\rho/2}\, e^{-\mu t}}{2(1-e^{-\rho/2})} \sqrt{\dfrac{2}{\lambda t}} \displaystyle\sum_{n=0}^{\infty} \dfrac{\left(\sqrt{\dfrac{\lambda t}{2}}\right)^{2n+1}}{n!(n+1)!}$

$$= \frac{\lambda e^{-\mu t}}{\sqrt{2\lambda t}\,(e^{\rho/2} - 1)} I_1\left(\sqrt{\lambda t}\right) \quad (6.2.3)$$

where $I_1\left(\sqrt{\lambda t}\right)$ is the Modified Bessel function of the first kind of order one.

The expected waiting time in the system,

$$W_S = E(T) = \int_0^{\infty} t.f(t)\, dt$$

$$= \frac{\lambda}{(e^{\rho/2} - 1)} \int_0^{\infty} t \left[\frac{1}{2} + \frac{\lambda t}{2^2(2)} + \frac{(\lambda t)^2}{2^3(12)} + \ldots\right] e^{-\mu t} dt$$

$$= \frac{\rho}{2\mu(1 - e^{-\rho/2})} \quad (6.2.4)$$

Hence μW_S, the expected number of customers served during the waiting time of an arbitrary arrival equals $\dfrac{\rho}{2\,(1 - e^{-\rho/2})}$, which is clearly a value greater than one and increases with increasing values of ρ.

Proof:

$$\mu W_S = \frac{\rho}{2(1-e^{-\rho/2})}$$

$$= \left\{1-\left(\frac{\rho^2}{4}-\frac{\rho^3}{24}+\dots\right)\right\}^{-1}$$

$$= 1+\frac{\rho^2}{4}-\frac{\rho^3}{24} \text{ nearly}$$

Clearly this is a value greater than one and increases with increasing values of ρ.

This can also verify by numerical evaluation.

Now, $E(T^2) = \int_0^\infty t^2 f(t)\,dt$

$$= \frac{\lambda}{(e^{\rho/2}-1)} \left\{ \begin{matrix} \frac{1}{2}\left(\frac{2}{\mu^3}\right)+\frac{\lambda}{2\times 4}\left(\frac{6}{\mu^4}\right)+ \\ \frac{\lambda^2}{8\times 12}\left(\frac{24}{\mu^5}\right)+\frac{\lambda^3}{16(6\times 24)}\left(\frac{120}{\mu^6}\right)+\dots \end{matrix} \right\}$$

$$= \frac{\rho}{\mu^2(e^{\rho/2}-1)} \left\{ 1+\left(\frac{\rho}{2}\right)+\frac{\left(\frac{\rho}{2}\right)^2}{2!}+\frac{\left(\frac{\rho}{2}\right)^3}{3!}+\dots+\frac{\rho}{4}\left[1+\frac{\rho}{2}+\frac{\left(\frac{\rho}{2}\right)^2}{2!}+\frac{\left(\frac{\rho}{2}\right)^3}{3!}+\dots\right] \right\}$$

$$= \frac{\rho}{\mu^2(e^{\rho/2}-1)} \left\{e^{\rho/2}+\rho/4\ e^{\rho/2}\right\} \qquad (6.2.5)$$

Hence,

$$Var(T),\ \sigma_T^2 = E(T^2)-[E(T)]^2$$

$$= \frac{\rho}{\mu^2\left(1-\bar{e}^{\rho/2}\right)^2} \left\{(1-e^{\rho/2})\ (1+\rho/4) - \rho/4\right\}$$

Hence,

$$\mu\sigma_T = \frac{\sqrt{\rho\left\{[1-e^{\rho/2}]\,(1+\rho/4)-\rho/4\right\}}}{\left(1-\bar{e}^{\rho/2}\right)} \qquad (6.2.6)$$

Now, the Coefficient of Variation (C.V) of T is given by

$$\text{C.V. of } T = \frac{\sigma_T}{W_S} = 2\sqrt{\frac{\left(1-\bar{e}^{\rho/2}\right)(1+\rho/4)-\rho/4}{\rho}} \qquad (6.2.7)$$

Table I: The Numerical evaluation of p_o and statistics of the Waiting time distribution of the Model I in (6.2) for different values of λ, $\mu\,(\lambda = \rho,\ \mu = 1)$

$\lambda\,(=\rho)$	p_o	μW_S	$\mu^2\sigma_T^2$	$\mu\sigma_T$	C.V. of T
0.1	0.9512	1.0256	1.0521	1.0257	1.0001
0.2	0.9048	1.0508	1.0980	1.0479	0.9971
0.5	0.7788	1.1302	1.2664	1.1253	0.9957
0.9	0.6376	1.2418	1.4976	1.2238	0.9856
0.95	0.6219	1.2563	1.5306	1.2372	0.9848
0.99	0.6096	1.2678	1.5559	1.2474	0.9837
1.00	0.6065	1.2708	1.5625	1.2500	0.9836
2	0.3679	1.5820	2.2434	1.4978	0.9468
5	0.0821	2.7236	4.8382	2.1996	0.8076
9	0.0111	4.5505	8.8715	2.9785	0.6545

6.3 MODEL II

The state dependent single server Markovian queue model characterized by the arrival rate $\lambda_n = \frac{\lambda}{(n+1)^2}$, $n \geq 0$, service rate $\mu_n = \mu$, $n \geq 1$ $(\lambda > 0, \mu > 0)$.

We have already discussed this model as Model BI in chapter 5. However, as a Markovian queueing model with state-dependent arrival rates, and for comparing some statistics of the waiting time distribution for the state-dependent models, we consider the model in this chapter also.

As in (5.2.1), the steady state probability distribution of queue length,

$$p_n = \frac{\rho^n}{(n!)^2\, I_o\left(2\sqrt{\rho}\right)},\ n \geq 0 \text{ and } \rho = \frac{\lambda}{\mu} \tag{6.3.1}$$

where $I_o\left(2\sqrt{\rho}\right)$ is the Modified Bessel function of the first kind of order zero.

System size at any time, $L_S = \Sigma n . p_n$

$$= \frac{\sqrt{\rho}\, I_1\left(2\sqrt{\rho}\right)}{I_o\left(2\sqrt{\rho}\right)} \tag{6.3.2}$$

Effective arrival rate, $\lambda^* = \Sigma \lambda_n . p_n$

$$= \frac{\mu\left[I_o\left(2\sqrt{\rho}\right)-1\right]}{I_o\left(2\sqrt{\rho}\right)} \tag{6.3.3}$$

Using Little's formula, the expected waiting time of an arrival in the system is given by,

$$W_S = \frac{L_S}{\lambda^*} = \frac{\sqrt{\rho}.\, I_1\left(2\sqrt{\rho}\right)}{\mu\left[I_o\left(2\sqrt{\rho}\right)-1\right]} \tag{6.3.4}$$

Hence,

$$\mu W_S = \frac{\sqrt{\rho}.\, I_1\left(2\sqrt{\rho}\right)}{\left[I_o\left(2\sqrt{\rho}\right)-1\right]} \tag{6.3.5}$$

which is the expected number of customers served during the waiting time of an arbitrary arrival.

Assume that customers are served in the order of their arrival. Service time distribution is exponential with mean $\frac{1}{\mu}$ and is independent of the arrival process. Let T be the waiting time of an arrival in the system.

As in (5.2.7), the pdf of T is given by

$$f(t) = \sum_{n=0}^{\infty} \frac{\lambda_n}{\lambda^*} p_n \frac{\mu e^{-\mu t} (\mu t)^n}{n!}, \text{ so that } \int_0^{\infty} f(t)\, dt = 1$$

i.e.,
$$f(t) = \frac{\lambda e^{-\mu t}}{[I_o(2\sqrt{\rho}) - 1]} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{(n+1)!(n+1)!\, n!}, \quad t > 0 \qquad (6.3.6)$$

Now,
$$E(T)\ (= W_S) = \int_0^{\infty} t \,.\, f(t)\, dt$$

$$= \sum_{n=0}^{\infty} \frac{\lambda^{n+1} \int_0^{\infty} e^{-\mu t}\, t^{n+1}\, dt}{\left[I_o\left(2\sqrt{\rho}\right) - 1\right](n+1)!(n+1)!\, n!}$$

$$= \frac{1}{\left[I_o\left(2\sqrt{\rho}\right) - 1\right]} \sum_{n=0}^{\infty} \frac{\rho^{n+1}}{\mu\, (n+1)!(n+1)!\ n!} \Gamma(n+2)$$

where $\Gamma(n+2)$ is the Gamma function and $\Gamma(n+2) = (n+1)\,!$, since n is a + ve integer.

Thus
$$W_S = \frac{1}{\left[I_o(2\sqrt{\rho})-1\right]\mu} \sum_{n=0}^{\infty} \frac{\rho^{n+1}}{(n+1)!\ n!}$$

$$= \frac{\sqrt{\rho}}{[I_o(2\sqrt{\rho})-1]\mu} \sum_{n=0}^{\infty} \frac{\left(\sqrt{\rho}\right)^{2n+1}}{n!\ (n+1)!}$$

$$= \frac{\sqrt{\rho}[I_1(2\sqrt{\rho}]}{\mu[I_o(2\sqrt{\rho})-1]} \tag{6.3.7}$$

Hence,

$$\mu W_S = \frac{\sqrt{\rho}\ \left[I_1\left(2\sqrt{\rho}\right)\right]}{\left[I_o\left(2\sqrt{\rho}\right)-1\right]} \tag{6.3.8}$$

which agrees with the result obtained by Little's formula.

Now,
$$E(T^2) = \int_0^{\infty} \frac{t^2 \lambda e^{-\mu t}}{[I_o(2\sqrt{\rho})-1]} \left(\sum_{n=0}^{\infty} \frac{(\lambda t)^n}{(n+1)!(n+1)!\ n!}\right) dt$$

$$= \frac{1}{[I_o(2\sqrt{\rho})-1]} \sum_{n=0}^{\infty} \frac{\lambda^{n+1}}{[(n+1)!]^2 n!} \int_0^{\infty} e^{-\mu t}\ t^{n+2}\ dt$$

$$= \frac{1}{\mu^2 [I_o(2\sqrt{\rho})-1]} \sum_{n=0}^{\infty} \frac{\rho^{n+1}}{[(n+1)!]^2\ n!}\ \Gamma(n+3)$$

where $\Gamma(n+3) = (n+2)! = (n+2)(n+1)!$

i.e., $$E(T^2)=\frac{1}{\mu^2[I_o(2\sqrt{\rho})-1]}\sum_{n=0}^{\infty}\frac{(n+2)\,\rho^{n+1}}{(n+1)!\;n!}$$

$$=\frac{1}{\mu^2\,[I_o\,(2\sqrt{\rho})-1]}\left\{\rho\;I_2\left(2\sqrt{\rho}\right)+2\sqrt{\rho}.I_1\left(2\sqrt{\rho}\right)\right\}$$

Hence,

$Var(T),\;\sigma_T^2=E(T^2)-[E(T)]^2$

$$=\frac{1}{\mu^2\,[I_o\,(2\sqrt{\rho})-1]}\left\{\rho\;I_2\left(2\sqrt{\rho}\right)+2\sqrt{\rho}.I_1\left(2\sqrt{\rho}\right)\right\}-\left\{\frac{\sqrt{\rho}\;I_1\left(2\sqrt{\rho}\right)}{\mu[I_o\,(2\sqrt{\rho})-1]}\right\}^2$$

i.e., $$\mu^2\,\sigma_T^2=\frac{\rho\;I_2\left(2\sqrt{\rho}\right)+2\sqrt{\rho}\;\;I_1\left(2\sqrt{\rho}\right)}{[I_o\,(2\sqrt{\rho})-1]}-\left\{\frac{\sqrt{\rho}\;\;I_1\left(2\sqrt{\rho}\right)}{I_o\left(2\sqrt{\rho}\right)-1}\right\}^2 \qquad (6.3.9)$$

Coefficient of variation (C.V) of T is given by C.V. of $T=\frac{\sigma_T}{W_S}$, which is a dimensionless quantity and is a measure of consistency.

Table II: The Numerical evaluation of p_o and statistics of the waiting time distribution of Model II in (6.3) for different values of λ, μ ($\lambda=\rho$; $\mu=1$)

$\lambda\,(=\rho)$	p_o	μW_S	$\mu^2\,\sigma_T^2$	$\mu\,\sigma_T$	C.V of T
0.1	0.9076	1.0279	1.0496	1.0245	0.9967
0.2	0.8277	1.0543	1.0985	1.0481	0.9941
0.5	0.6400	1.1258	1.2440	1.1153	0.9907
0.9	0.4699	1.2180	1.4341	1.1975	0.9832
0.95	0.4541	1.2310	1.4564	1.2068	0.9804
0.99	0.4417	1.2405	1.4750	1.2145	0.9790
1.00	0.4386	1.2427	1.4796	1.2164	0.9788
2	0.2344	1.4689	1.9242	1.3872	0.9443
5	0.0585	2.3663	2.3569	1.5352	0.6488
9	0.0149	3.0459	3.1751	1.7819	0.5852

6.4 COMPARISON OF EXPECTED WAITING TIME AND COEFFICIENT OF VARIATION FOR DIFFERENT MODELS

In the following table, we provide the values of expected waiting time T and corresponding coefficient of variation for different queueing models for $\lambda = \rho$, $\mu = 1$

ρ	*M/M/1 Model*		*Model I in section (6.2)*		*Model II in section (6.3)*	
	μW_S	$C.V\ of\ T$	μW_S	$C.V\ of\ T$	μW_S	$C.V\ of\ T$
0.1	1.1111	1	1.0256	1.0000	1.0279	0.9967
0.2	1.250	1	1.0508	0.9971	1.0543	0.9941
0.5	2.000	1	1.1302	0.9957	1.1258	0.9907
0.9	10.000	1	1.2418	0.9856	1.2180	0.9832
0.95	20.000	1	1.2563	0.9848	1.2310	0.9804
0.99	100.00	1	1.2678	0.9837	1.2405	0.9790
1.00	–	–	1.2707	0.9836	1.2427	0.9788
2	–	–	1.5820	0.9468	1.4689	0.9443
5	–	–	2.7236	0.8076	2.3663	0.6488
9	–	–	4.5505	0.6545	3.0451	0.5852

6.5 SUMMARY OF RESULTS AND CONCLUDING REMARKS

(1) The expected waiting time and coefficient of variation of Models I & II in sections (6.2) and (6.3) respectively are less than that of the standard M/M/1 Model.

(2) Among the models considered, Model II is better than the Model I. Hence customer can always choose Model II to have less waiting time in the system.

(3) For the Model II, $L_S = \dfrac{\sqrt{\rho}\ I_1\left(2\sqrt{\rho}\right)}{I_o(2\sqrt{\rho})}$; $W_S = \dfrac{I_1\left(2\sqrt{\rho}\right).\sqrt{\rho}}{\mu\,[I_o(2\sqrt{\rho})-1]}$ &

$$\lambda^* = \frac{\mu\,[I_o(2\sqrt{\rho})-1]}{I_o\left(2\sqrt{\rho}\right)}$$

Thus $L_S = \lambda^* W_s$, a verification of Little's formula. Similarly Little's formula may be verified for Model I.

(4) With reference to the performance measure p_o, the state probability of an empty system, Model I is better than Model II. However Model II is better than the standard M/M/1 model.

(5) The comparisons in terms of expected waiting time and coefficient of variation of waiting time indicates the reduction in these quantities for Models I and II. This suggests the effect of queue discouraged by potential customers. The effect is more significant for higher values of the traffic intensity parameter ρ.

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

Chapter 7

COMPARISON BETWEEN STATE-DEPENDENT MODELS

7.1 INTRODUCTION

This chapter is devoted to the comparison between the state dependent models developed in the previous chapters viz.

(i) Model I: The state-dependent queue model characterized by the arrival rate $\lambda_n = \lambda, n \geq 0$; service rate $\mu_n = n\mu, n \geq 1$ where $\lambda > 0, \mu > 0$.

(ii) Model II: The state-dependent queue model characterized by the arrival rate $\lambda_n = \dfrac{\lambda}{n+1}$, $n \geq 0$; service rate $\mu_n = \mu, n \geq 1$ where $\lambda > 0, \mu > 0$.

(iii) Model III: The state-dependent queue model characterized by the arrival rate $\lambda_n = \dfrac{\lambda}{n+1}$, $n \geq 0$; service rate $\mu_n = n\mu, n \geq 1$ where $\lambda > 0, \mu > 0$.

(iv) Model IV: The state-dependent queue model characterized by the arrival rate $\lambda_n = (n + 1)\,\lambda$, $n \geq 0$; service rate $\mu_n = n\mu, n \geq 1$ where $\lambda > 0$, $\mu > 0$.

(v) Model V: The state-dependent queue model characterized by the arrival rate $\lambda_n = \dfrac{\lambda}{2n+1}$, $n \geq 0$; service rate $\mu_n = 2n\mu, n \geq 1$ where $\lambda > 0$, $\mu > 0$.

The numerical evaluation of system characteristics permit us to have a meaningful comparison between these models. In order to illustrate the advantages of specific adaptive mechanism we also compare these models with the worst possible non-adaptive system – M/M/1.

Suppose we are operating an M/M/1 queueing system with parameters λ and μ. The traffic intensity γ is measured by the ratio

$\gamma_{M/M/1} = \frac{\hat{S}}{\hat{T}}$ where $\hat{S}$ = mean service time, $\hat{T}$ = mean inter arrival time

$= \frac{\lambda}{\mu} \; (=\rho) \quad \text{(say)}$

Suppose now, keeping the same λ and μ , that we consider the adaptive mechanism of Model I, characterized by the arrival rate $\lambda_n = \lambda$, $n \geq 0$; service rate $\mu_n = n\mu$, $n \geq 1$ where $\lambda > 0$, $\mu > 0$.

This generates a new mean effective service time $\hat{S}$ given by $\hat{S} = \frac{1}{\mu^*} = \frac{1}{\lambda}(1-\bar{e}^{\rho})$ where μ^* = effective service rate and consequently a new traffic intensity is generated,

$$\gamma_{Model\,I} = \lambda\,\hat{S} = 1-\bar{e}^{\rho} \qquad (7.1.1)$$

which is clearly less than $\rho = \gamma_{M/M/1}$ for $\rho > 0$

i.e., the compensatory adaptive mechanism of Model I causes the system to operate at a lower traffic intensity.

When ρ is very small, there is not much difference between $\gamma_{Model\ I}$ and $\gamma_{M/M/1}$ (= ρ). But when ρ has the largest possible value for the operation of M/M/1 in steady state, i.e.; $\rho = 1$, $\gamma_{Model\ I} = 1 - e^{-1} = 0.63$. Therefore Model I offers better operating characteristics than M/M/1, see for example, Brian Conolly, (1975).

We have

$$\mu \sum_{1}^{\infty} p_n = \mu(1 - p_o)$$

Hence,

$$\frac{\mu \sum_{1}^{\infty} p_n}{1 - p_o} = \mu = \frac{1}{E\ (service\ time)}$$

For Model I, $\mu_n = n\mu$

$$\frac{\sum_{1}^{\infty} n\mu\, p_n}{1 - p_o} = \frac{\mu \sum_{1}^{\infty} n\, p_n}{1 - p_o} = \frac{\lambda}{1 - e^{-\rho}} (= \mu) = \frac{1}{E\ (service\ time)}$$

Now consider the state dependent mechanism of Model II, characterized by $\lambda_n = \dfrac{\lambda}{n+1}$, $n \geq 0$; service rate $\mu_n = \mu$, $n \geq 1$ where $\lambda > 0$, $\mu > 0$. This model is the counter part to effective service time for Model I. For this model, effective arrival rate,

$$\lambda^* = \frac{1}{\hat{T}} = \sum_{0}^{\infty} \lambda_n . p_n = \sum_{0}^{\infty} \frac{\lambda}{n+1} \frac{e^{-\rho} \rho^n}{n!} = \mu(1 - e^{-\rho})$$

The explicit results of Hadidi and Natvig (1969) for effective inter arrival interval (*EII*) are too complicated to quote but they lead to the result.

$$E\,(EII) = \bar{\mu}^{1}\,(1-e^{-\rho})^{-1}\,(=\hat{T})$$

The new traffic intensity generated by Model II is given by

$$\gamma_{Model\ II} = \frac{\hat{S}}{\hat{T}} = \frac{\frac{1}{\mu^*}}{\frac{1}{\lambda^*}} = 1 - e^{-\rho} \qquad (7.1.2)$$

which is also the traffic intensity induced by Model I. i.e., as in Model I, the compensatory adaptive mechanism of Model II causes the system to operate at lower traffic intensity.

Thus when ρ is very small, there is not much difference between $\gamma_{Model\ I}$ $(= \gamma_{Model\ II})$ and $\gamma_{M/M/1}$ $(= \rho)$. But when ρ has the largest possible value for the operation of M/M/1 in steady state, i.e., $\rho = 1$, $\gamma_{Model\ II} = 1 - e^{-1} = 0.63$.

Therefore, Model II like Model I offers better operating characteristics than M/M/1.

Now we consider Model III characterized by the arrival rate $\lambda_n = \frac{\lambda}{n+1}$, $n \geq 0$; service rate $\mu_n = n\mu$, $n \geq 1$ where $\lambda > 0$, $\mu > 0$. Keeping the same λ and μ, we consider Model III which is a combination of Model I and Model II and for this model, effective arrival rate $\lambda^* = \frac{\lambda\, I_1\left(2\sqrt{\rho}\right)}{\sqrt{\rho}\;\, I_o(2\sqrt{\rho})}$ and effective service rate,

$$\mu^* = \frac{\mu\,\sqrt{\rho}\; I_1\left(2\sqrt{\rho}\right)}{\left[I_o(2\sqrt{\rho}) - 1\right]}$$

This generates a new mean effective service time $\hat{S} = \frac{1}{\mu^*}$ given by

$$\hat{S} = \frac{\left[I_0\left(2\sqrt{\rho}\right)-1\right]}{\mu.\sqrt{\rho}\; I_1(2\sqrt{\rho})} \text{ and mean inter arrival time, } \hat{T} = \frac{1}{\lambda^*} = \frac{\sqrt{\rho}\; I_o\left(2\sqrt{\rho}\right)}{\lambda\, I_1(2\sqrt{\rho})}$$

Consequently the new traffic intensity induced by the Model III is given by

$$\gamma_{Model\ III} = \frac{\hat{S}}{\hat{T}} = 1 - \frac{1}{I_o\left(2\sqrt{\rho}\right)} < \rho = \gamma_{M/M/1} \quad \text{for } \rho > 0 \qquad (7.1.3)$$

i.e., Model III causes the system to operate at a lower traffic intensity. When ρ is very small, there is not much difference between $\gamma_{Model\ III}$ and $\gamma_{M/M/1}$ $(= \rho)$. But when ρ has the largest possible value for the operation of M/M/1 in steady state, i.e., $\rho = 1$,

$$\gamma_{Model\ III} = 1 - \frac{1}{I_o(2)} = 0.56 \qquad (7.1.4)$$

Therefore, Model III offers better operating characteristics than the state dependent Models I & II where Models I & II are better than the standard M/M/1 Model.

The same question may be put in a number of different ways. For example, suppose the system is operating under M/M/1. For the Model I, $\gamma_{Model\ I} = 1 - e^{-1} = 0.63$ when ρ has the largest possible value in the steady state, i.e., when $\rho = 1$. Since traffic intensity (i.e., effective traffic intensity) for Model I is less than that of the standard M/M/1 model, Model I offers a better operating characteristics than M/M/1.

It is now obvious that Model III offers a better operating characteristics than Model I or Model II.

Now we consider the state dependent queue model characterized by $\lambda_n = (n + 1)\,\lambda$, , $n \geq 0$; service rate $\mu_n = n\mu$, $n \geq 1$ where $\lambda > 0$, $\mu > 0$. For this model called Model IV, effective arrival rate, $\lambda^* = \dfrac{\lambda}{1-\rho}$ and effective service rate $\mu^* = \dfrac{\mu}{1-\rho}$

This generates a new mean effective service time, $\hat{S} = \dfrac{(1-\rho)}{\mu}$ and a mean inter arrival time $\hat{T} = \dfrac{(1-\rho)}{\lambda}$. Consequently the new traffic intensity induced by the Model IV is given by

$$\gamma_{Model\ IV} = \frac{\hat{S}}{\hat{T}} = \rho = \gamma_{M/M/1} \qquad (7.1.5)$$

i.e., Model IV causes the system to operate at the same traffic intensity that of M/M/1.

Now we consider a single server generalization of M/M/1 queueing model in which the parameters λ_n and μ_n of arrival and service rate, respectively, are caused to depend on n, the number of customers present in the system, which is the state of the system at time t after initiation. This model is defined by

$\lambda_n = \dfrac{\lambda}{2n+1}$, $n \geq 0$; $\mu_n = 2n\mu$, $n \geq 1$ where $\lambda, \mu > 0$. We call this model as Model V.

Suppose we are operating an M/M/1 queueing system with parameters λ and μ. The traffic intensity γ is measured by the ratio

$\gamma_{M/M/1} = \dfrac{\hat{S}}{\hat{T}}$ where $\hat{S}$ = mean service time and $\hat{T}$ = mean inter arrival time

Suppose now, keeping the same λ and μ, that we consider adaptive mechanism, the state dependent model V. This generates a new mean effective service time $\hat{S}$ given by

$$\frac{e^{\sqrt{\rho}} + e^{-\sqrt{\rho}} - 2}{\mu\sqrt{\rho}(e^{\sqrt{\rho}} - e^{-\sqrt{\rho}})} \tag{7.1.6}$$

The mean effective inter arrival time $\hat{T}$ is given by

$$\hat{T} = \frac{\sqrt{\rho}\left(e^{\sqrt{\rho}} + e^{-\sqrt{\rho}}\right)}{(e^{\sqrt{\rho}} - e^{-\sqrt{\rho}})} \tag{7.1.7}$$

Consequently the new traffic intensity induced by the above Model V is given by:

$$\gamma_{Model\ V} = \frac{\hat{S}}{\hat{T}} = \frac{\left(e^{\sqrt{\rho}} + e^{-\sqrt{\rho}} - 2\right)\rho}{(e^{\sqrt{\rho}} + e^{-\sqrt{\rho}})} < \rho = \gamma_{M/M/1} \quad \text{for } \rho > 0 \tag{7.1.8}$$

When ρ is very small, there is not much difference between $\gamma_{\text{Model V}}$ and $\gamma_{M/M/1} = \rho$. But when ρ has the largest value for the operation of M/M/1 in steady state, there is much difference between $\gamma_{\text{Model V}}$ and $\gamma_{M/M/1}$ $(= \rho)$.

When $\rho = 0.99$, $\gamma_{Model\ V} = 0.3495$ and

When $\rho = 1$, $\gamma_{Model\ V} = 0.52$

Therefore, Model V offers better operating characteristics than Model III. Thus Model V offers better operating characteristics than Models I, II, III and IV (or M/M/1).

7.2 NUMERICAL COMPARISON OF MEASURE OF SERVICE CAPACITY FOR THE DIFFERENT STATE-DEPENDENT QUEUEING MODELS

Since $\hat{B}$ is the mean busy period and $\hat{T}$ is the mean inter arrival time, the ratio $\frac{\hat{B}}{\hat{T}}$ gives the average number of customers queueing in the busy period. This is a measure of service capacity of the system. In this section we compare this measure for different state-dependent queueing models.

For Model I, $\frac{\hat{B}}{\hat{T}} = e^{\rho} - 1$;

For Model II, $\frac{\hat{B}}{\hat{T}} = \frac{(e^{\rho} - 1)^2}{\rho e^{\rho}}$;

For Model III, $\frac{\hat{B}}{\hat{T}} = \frac{\left[I_o\left(2\sqrt{\rho}\right) - 1\right]}{I_o\left(2\sqrt{\rho}\right)}$ &

For M/M/1, $\frac{\hat{B}}{\hat{T}} = \frac{1}{1-\rho}$

where $\hat{B} = E$ (BP) and $\hat{T} = E$ (inter arrival time)

Now, let us compare Numerically $\frac{\hat{B}}{\hat{T}}$ for Models I, II & III in (7.1) with M/M/1 Model for various $\rho = \frac{\lambda}{\mu}$. The results are tabulated and given in the Table I.

Table I

ρ	$\frac{\hat{B}}{\hat{T}}$ for M/M/1 where $\rho < 1$	$\frac{\hat{B}}{\hat{T}}$ for Model I	$\frac{\hat{B}}{\hat{T}}$ for Model II	$\frac{\hat{B}}{\hat{T}}$ for Model III
0.1	1.1111	0.1052	0.1002	0.0924
0.2	1.250	0.2214	0.2007	0.1723
0.5	2.000	0.6487	0.5105	0.3599
0.9	10.000	1.4596	0.9624	0.5300
0.95	20.000	1.5857	1.0236	0.5459
0.99	100.000	1.6912	1.0736	0.5583
1.00	–	1.7183	1.0862	0.5614
2	–	6.3891	2.7622	0.7656
5	–	147.4132	29.6840	0.9415
9	–	8102.0839	900.1204	0.9851
10	–	22025.4658	2202.4932	0.9889

From the Table (I) it is observed that,

$$\left(\frac{\hat{B}}{\hat{T}}\right)_{Model\ III} < \left(\frac{\hat{B}}{\hat{T}}\right)_{Model\ II} < \left(\frac{\hat{B}}{\hat{T}}\right)_{Model\ I} < \left(\frac{\hat{B}}{\hat{T}}\right)_{M/M/1} \qquad (7.2.1)$$

7.3 NUMERICAL COMPARISON OF $\lambda \hat{W}, \lambda \hat{B}$ & $\lambda \hat{S}$ FOR DIFFERENT STATE-DEPENDENT QUEUEING MODELS

In this section we evaluate numerically $\lambda \hat{W}, \lambda \hat{B}$ (which measures the number of arrivals during the busy period) & $\lambda \hat{S}$ for the state dependent Models I, II & III under various values of ρ. The values are tabulated and given in Tables I, II, III respectively where $\hat{B} = E(BP)$; $\hat{W} = E(W)$ and $\hat{S} = E$ (service time) and λ is the arrival rate. Then we compare these models with that of the standard M/M/1 model.

Table II: Numerical evaluation of $\lambda\hat{W}$, $\lambda\hat{B}$ and $\lambda\hat{S}$ for Model I in (7.1) for various ρ

ρ	$\lambda\hat{W}$	$\lambda\hat{B}$	$\lambda\hat{S}$
0.1	0.1	0.1052	0.0952
0.2	0.2	0.2214	0.1813
0.5	0.5	0.6487	0.3935
0.9	0.9	1.4596	0.5934
0.95	0.95	1.5857	0.6133
0.99	0.99	1.6912	0.6284
1.00	1.00	1.7183	0.6321

Table III: Numerical evaluation of $\lambda\hat{W}$, $\lambda\hat{B}$ and $\lambda\hat{S}$ for the Model II in (7.1) for various ρ

ρ	$\lambda\hat{W}$	$\lambda\hat{B}$	$\lambda\hat{S}$
0.1	0.1051	0.1052	0.1
0.2	0.2207	0.2214	0.2
0.5	0.6354	0.6487	0.5
0.9	1.3649	1.4596	0.9
0.95	1.4717	1.5857	0.95
0.99	1.5596	1.6912	0.99
1.00	1.5820	1.7183	1.00

Table IV: Numerical evaluation of $\lambda\hat{W}$, $\lambda\hat{B}$ and $\lambda\hat{S}$ for Model III in (7.1) for various ρ

ρ	$\lambda\hat{W}$	$\lambda\hat{B}$	$\lambda\hat{S}$
0.1	0.1	0.1018	0.0973
0.2	0.2	0.2081	0.1897
0.5	0.5	0.5624	0.4441
0.9	0.9	1.1280	0.7389
0.95	0.95	1.2020	0.7717
0.99	0.99	1.2640	0.7981
1.00	1.00	1.280	0.8047

For comparison purposes we also evaluate numerically $\lambda\hat{W}$, $\lambda\hat{B}$ and $\lambda\hat{S}$ for the standard M/M/1 model and the results tabulated in the following Table V.

Table V: Numerical evaluation of $\lambda\hat{W}$, $\lambda\hat{B}$ and $\lambda\hat{S}$ for the M/M/1 model for various ρ

ρ	$\lambda\hat{W}$	$\lambda\hat{B}$	$\lambda\hat{S}$
0.1	0.1111	0.1111	0.1
0.2	0.250	0.250	0.2
0.5	1.000	1.000	0.5
0.9	9.000	9.000	0.9
0.95	19.000	19.000	0.95
0.99	99.000	99.000	0.99

From Table IV, it is observed that $\lambda\hat{S}_{\text{Model I}} < \lambda\hat{S}_{\text{Model III}} < \lambda\hat{S}_{\text{Model II}}$ and for $\rho < 1$, we have,

$\lambda\hat{S}_{\text{Model I}} < \lambda\hat{S}_{\text{Model III}} < \lambda\hat{S}_{\text{Model II}} = \lambda\hat{S}_{\text{M/M/1}}$

Also $\lambda\hat{W}_{\text{Model I}} < \lambda\hat{W}_{\text{Model II}} < \lambda\hat{W}_{\text{M/M/1}}$

and $\lambda\hat{B}_{\text{Model III}} < \lambda\hat{B}_{\text{Model I}} = \lambda\hat{B}_{\text{Model II}} < \lambda\hat{B}_{\text{M/M/1}}$

Now we compute $\lambda\hat{W}$ and $\lambda\hat{S}$ for the Model IV, characterized by $\lambda_n = (n+1)\lambda$, $n \geq 0$, $\mu_n = n\mu$, $n \geq 1$ ($\lambda > 0$, $\mu > 0$). The measures are evaluated and tabulated in tables V and VI respectively.

Table VI: Numerical evaluation of $\lambda\hat{W}$, and $\lambda\hat{S}$ for the Model IV in (7.1) for various ρ in $0 < \rho < 1$

ρ	$\lambda\hat{W}$	$\lambda\hat{S}$
0.1	0.1	0.09
0.2	0.2	0.16
0.5	0.5	0.25
0.9	0.9	0.09
0.95	0.95	0.05
0.99	0.99	0.01

Table VII: Numerical evaluation of $\lambda\hat{W}$, and $\lambda\hat{S}$ for the Model V in (7.1) for various ρ in $0 < \rho < 1$

ρ	$\lambda\hat{W}$	$\lambda\hat{S}$
0.1	0.0508	0.0492
0.2	0.1033	0.0967
0.5	0.2708	0.2293
0.9	0.5166	0.3834
0.95	0.5491	0.4009
0.99	0.5754	0.4166

From Tables VI and VII, it is observed that:

$$\lambda\hat{W}_{\text{Model V}} < \lambda\hat{W}_{\text{Model IV}}$$

7.4 COMPARISON OF SYSTEM SIZE

In this section we evaluate numerically the system size in steady state for the state dependent models I, II & III where each model is operating under the same traffic load. We also compare these state dependent Queueing Models with that of the standard M/M/1 model.

For example, for M/M/1 model, $L_s = \hat{N} = \dfrac{\gamma}{1-\gamma}$ and for Models I & II, $L_S = \rho$.

Now $1 - p_o = \gamma \Rightarrow 1 - e^{-\rho} = \gamma$

Hence $\rho = -\log(1 - \gamma)$, system size for Models I & II. Therefore mean system size for Models I & II are given by

$$\hat{N} = -\log(1-\gamma) \tag{7.4.1}$$

For Model III, effective traffic intensity induced by model III is

$$\gamma_{\text{Model III}} = 1 - p_o = 1 - \frac{1}{I_o(2\sqrt{\rho})}.$$

Now when $\gamma = 0.1$, $\dfrac{1}{I_o(2\sqrt{\rho})} = 0.9$ which implies $\rho = 0.3086$

Hence,

$$\hat{N} = \sqrt{\rho}.\frac{I_1(2\sqrt{\rho})}{I_1(2\sqrt{\rho})} = 0.1741 \tag{7.4.2}$$

Also when $\gamma = 0.5$, $I_o(2\sqrt{\rho}) = 2$ which implies $\rho = 0.8190$

Hence

$$\hat{N} = 0.6017 \qquad (7.4.3)$$

Continuing in this manner, when $\gamma = 0.99$, $I_o(2\sqrt{\rho}) = 100$ which implies

$\rho = 10.3362$

Hence

$$\hat{N} = 2.9115 \qquad (7.4.4)$$

System size $\hat{N}$ in steady state for M/M/1 and Models I – III are evaluated numerically and tabulated in Table VIII. This table compares mean system size of M/M/1 model with that of models I – III at common levels of traffic intensity for each.

Table VIII: System size in steady state for Models I – III compared with M/M/1 for fixed traffic intensities

ρ	*Mean system size* $\hat{N}$			
	M/M/1	Model I	Model II	Model III
0.1	0.1111	0.1054	0.1054	0.1741
0.2	0.25	0.223	0.223	0.211
0.5	1.00	0.6931	0.6931	0.6017
0.9	9.00	2.3026	2.3026	1.6397
0.95	19.00	2.9957	2.9957	2.0749
0.99	99.00	4.6052	4.6052	2.9115

By comparing the mean system size, the advantages of Models I – III are maintained. It is verified that at low traffic intensities there is not a lot to be gained by comparison with the standard M/M/1 model but the advantage increases with traffic intensity. It may be observed that the mean system size increases with traffic intensity as expected and the state dependent Model III is superior to the models I & II, which are still superior to the standard M/M/1 model with reference to the mean system size $\hat{N}$.

7.5 COMPARISON BETWEEN LINEAR AND NON-LINEAR STATE-DEPENDENT QUEUEING MODELS

In this section, we study some general linear (Model C_1) and non-linear state-dependent Markovian queueing models (Model C_2). We obtain the steady-state probability distribution of queue length and important performance measures for these models. We prove that the steady-state probability distribution of queue length for the linear and non-linear state dependent queueing models considered follows the Geometric distribution, $p_n = (1 - \rho)\rho^n$, $n \geq 0$, $0 < \rho \left(= \dfrac{\lambda}{\mu}\right) < 1$, which is also the steady state probability distribution of queue length for the standard M/M/1 model. We also evaluate numerically the steady-state performance measures for the models considered and make a comparison between them. With reference to the performance measure p_0, the state probability of an empty system, we obtain a better non-linear state-dependent model than a particular case of state-dependent queueing model of the general non-linear state-dependent queueing model considered.

Model C₁

General linear state-dependent queueing model characterized by the arrival rate $\lambda_n = (n + k)\lambda$, $n \geq 0$; service rate $\mu_n = (n + r)\mu$, $n \geq 1$ where r & k are non-negative integers satisfying $k \geq 1$ and $r \geq 0$ and $\lambda > 0$, $\mu > 0$.

Using the Birth-death methodology, we obtain the steady state probability distribution of queue length and performance measures as under:

$$p_n = \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} p_o$$
$$= \frac{k\lambda\,(k+1)\lambda \ldots (k+n-1)\lambda\, p_o}{(r+1)\,\mu(r+2)\mu \ldots (r+n)\mu}$$

$= \rho^n\, p_o$ for the pairs of $\{(k,r)\} = \{(1,0),(2,1),(3,2),\ldots\}$ where $\rho = \dfrac{\lambda}{\mu}$.

Using $\sum p_n = 1$ gives $p_o = (1-\rho)$

Hence,

$$p_n = (1-\rho)\,\rho^n,\ n \geq 0,\ 0 < \rho < 1. \qquad (7.5.1)$$

which is a Geometric distribution with parameter $\rho < 1$.

System size at any time t, $L_S = \sum n\, p_n$

$$= \frac{\rho}{1-\rho} = \rho + \rho^2, \text{(approximately)} \qquad (7.5.2)$$

which is clearly less than one and increases with increasing values of ρ in $0 < \rho < 1$.

Effective arrival rate,

$$\lambda^* = \sum_{n=0}^{\infty} \lambda_n . p_n$$

$$= \sum \lambda (n+k)\, p_n = \lambda \{L_S + k\}$$

$$= \frac{\lambda [k + \rho(1-k)]}{1-\rho} \qquad (7.5.3)$$

For k = 1, 2, 3, ..., the effective arrival rate λ^* (approximately) are respectively given by $\lambda\,(1 + \rho + \rho^2)$, $\lambda\,(2 + \rho + \rho^2)$, $\lambda\,(3 + \rho + \rho^2)$, ... and for a given arrival rate λ, λ^* increases with increasing values of k as expected.

For a particular value of k, λ^* increases with increasing values of ρ.

Using Little's formula, the expected waiting time of an arbitrary arrival is given by

$$W_S = \frac{L_S}{\lambda^*} = \frac{1}{\mu[k + \rho(1-k)]}$$

Hence $$\mu W_S = \frac{1}{k + \rho(1-k)} \qquad (7.5.4)$$

Which is the number of customers served during the waiting time of an arbitrary arrival in the system. For k = 1, 2, 3, ... the number of customers served during the waiting time of an arbitrary arrival are respectively given by (approximately), 1, $\frac{1}{2} + \frac{\rho}{4} + \frac{\rho^2}{8}$, $\frac{1}{3} + \frac{2}{9}\rho + \frac{4}{27}\rho^2$,... which decreases with increasing values of k.

i.e.; when the rate of arrivals increases, the number of customers served during the waiting time of an arbitrary arrival decreases.

Effective service rate μ^* is given by

$$\mu^* = \frac{\lambda^*}{1-p_o} = \frac{\mu\,[k+\rho(1-k)]}{1-\rho} \qquad (7.5.5)$$

For $k = 1, 2, 3, \ldots, \mu^*$ are respectively given by

$\mu(1+\rho+\rho^2)$, $\mu(2+\rho+\rho^2)$, $\mu(3+\rho+\rho^2)$, ... approximately.

For a given service rate μ, μ^* increases with increasing values of ρ as expected for a specified values of k.

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \rho = 1 - p_o \qquad (7.5.6)$$

which is the number of costumers at the server. It is also the proportion of time the server is working and $\rho^* = \rho$ means that the number of customers at the server or the proportion of time the server is working equals the minimum number of servers to cope with the arriving traffic.

Remark: A particular case of this general linear model when $k = 2$, $r = 1$ gives the model A VII discussed in Chapter 4.

Model C₂:

General non-linear state-dependent queueing model characterized by the arrival rate $\lambda_n = (n+1)^m \lambda$, $n \geq 0$; service rate $\mu_n = n^m\mu$, $n \geq 1$ where $m \geq 2$ and $\lambda > 0, \mu > 0$.

As in the Model C_1,

$$p_n = \frac{(1^m\lambda)\,(2^m\lambda)\,...((n^m\lambda)\;p_o}{(1^m\mu)\,(2^m\mu)...(n^m\mu)} = \rho^n p_o$$

using $\sum p_n = 1$ gives $p_o = 1 - \rho$

Hence,

$$p_n = (1-\rho)\,\rho^n,\ n \geq 0,\ 0 < \rho < 1 \qquad (7.5.7)$$

which is a Geometric distribution as in Model C_1.

System size at any time, $L_S = \sum n\,p_n = \dfrac{\rho}{1-\rho}$

$$= \rho + \rho^2 + \rho^3 \text{ approximately} \qquad (7.5.8)$$

which is a value less than one and increases with increasing values of ρ in $0 < \rho < 1$. This result may also be verified by numerical evaluation.

Effective arrival rate, $\lambda^* = \sum_{n=0}^{\infty} \lambda_n . p_n$

$$= \sum (n+1)^m . \lambda . p_n$$

$$= \frac{\lambda(1+\rho)}{(1-\rho)^2}\ \text{, where } m = 2$$

$$= \lambda(1+3\rho+5\rho^2)\ \text{, approximately} \qquad (7.5.9)$$

For a given arrival rate λ, λ^* increases with increasing values of ρ in $0 < \rho < 1$.

Using Little's formula, the expected waiting time of an arbitrary arrival,

$$W_S = \frac{L_S}{\lambda^*} = \frac{(1-\rho)}{\mu(1+\rho)}$$

Hence,

$$\mu W_S = \frac{1-\rho}{1+\rho} = 1 - 2\rho + 2\rho^2 - 2\rho^3, \text{ nearly} \quad (7.5.10)$$

which is the number of customers served during the waiting time of an arbitrary arrival in the system, which decreases with increasing values of ρ in $0 < \rho < 1$.

Effective service rate, $\mu^* = \dfrac{\lambda^*}{1-p_o}$

$$= \frac{\mu(1+\rho)}{(1-\rho)^2}$$

$$= \mu(1+3\rho+5\rho^2+7\rho^3) \text{ , nearly} \quad (7.5.11)$$

Hence for a given service rate μ, μ^* increases with increasing values of ρ.

Effective traffic intensity,

$$\rho^* = \frac{\lambda^*}{\mu^*} = \rho = 1 - p_o \quad (7.5.12)$$

This result is also true for the general linear state-dependent queueing Model C_1.

From the Model C_1, when $k = 1$ and $r = 0$, we have the queueing model characterized by $\lambda_n = (n+1)\lambda, n \geq 0$; $\mu_n = n\mu, n \geq 1$ where $\lambda > 0$, $\mu > 0$, which is a model introduced in B.W, Conolly (1974) and Brain Conolly (1975).

As in the general state-dependent Model C_1, the steady-state probability distribution of queue length for this particular model is given by, $p_n = (1-\rho)\rho^n$, $n \geq 0$ and $0 < \rho < 1$.

$$L_S = \frac{\rho}{1-\rho} = \rho + \rho^2 + \rho^3 \text{ approximately}$$

By putting $k = 1$ in (7.5.3), $\lambda^* = \frac{\lambda}{1-\rho}$, $\mu^* = \frac{\mu}{1-\rho}$, $\mu W_S = 1$ and $\rho^* = \rho$, a result shared with Model C_1. Now we evaluate numerically the performance measures of certain specific models of section (7.5) viz., (i) the Model C_1 where $k = 1$, $r = 0$ and (ii) the Model C_2 where $m = 2$ for various ρ in $0 < \rho < 1$ and the results tabulated in Table I and Table II respectively.

Table I: Performance measures for the Model C_1 for various ρ in $0 < \rho < 1$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9	0.1111	0.5263	2.1111 λ	2.1111 μ	0.1
0.2	0.8	0.25	0.5556	2.25 λ	2.25 μ	0.2
0.5	0.5	1.00	0.6667	3.00 λ	3.00 μ	0.5
0.9	0.1	9.000	0.9091	11.00 λ	11.00 μ	0.9
0.99	0.01	99.000	0.9901	101.00 λ	101.00 μ	0.99

Table II: Performance measures for the Model C_2 for various ρ in $0 < \rho < 1$

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9	0.1111	0.8182	1.3580 λ	1.3580 μ	0.1
0.2	0.8	0.25	0.6667	1.8750 λ	1.8750 μ	0.2
0.5	0.5	1.00	0.3333	6.00 λ	6.00 μ	0.5
0.9	0.1	9.000	0.0526	190.00 λ	190.00μ	0.9
0.99	0.01	99.000	0.0050	19900.00 λ	19900.00μ	0.99

Now for comparison purposes, we also evaluate numerically the performance measures with that of the standard M/M/1 model for various ρ in $0 < \rho < 1$ and the results tabulated in Table III

Table III

ρ	p_o	L_S	μW_S	λ^*	μ^*	ρ^*
0.1	0.9	0.1111	1.1111	λ	μ	0.1
0.2	0.8	0.250	2.250	λ	μ	0.2
0.5	0.5	1.000	2.000	λ	μ	0.5
0.9	0.1	9.000	10.000	λ	μ	0.9
0.99	0.01	99.000	100.000	λ	μ	0.99

7.6 SUMMARY OF RESULTS AND CONCLUDING REMARKS

1. If $\hat{S}$ is the mean service time and λ, the arrival rate, then

$$\lambda\hat{S}_{\text{Model I}} < \lambda\hat{S}_{\text{Model II}} = \lambda\hat{S}_{\text{M/M/1}}$$

Also

$$\lambda\hat{S}_{\text{Model I}} < \lambda\hat{S}_{\text{Model III}}$$

2. $\lambda\hat{S}_{\text{Model IV}} < \lambda\hat{S}_{\text{Model I}} < \lambda\hat{S}_{\text{Model II}} = < \lambda\hat{S}_{\text{M/M/1}}$

3. $\lambda\hat{S}_{\text{Model V}} < \lambda\hat{S}_{\text{Model I}} < \lambda\hat{S}_{\text{Model II}}$

4. On comparing the measure of service capacity of the models we have,

$$\left(\frac{\hat{B}}{\hat{T}}\right)_{\text{Model III}} < \left(\frac{\hat{B}}{\hat{T}}\right)_{\text{Model II}} < \left(\frac{\hat{B}}{\hat{T}}\right)_{\text{Model I}}$$

5. If $\hat{W}$ denotes the expected waiting time and λ, the arrival rate, then

$$(\lambda\hat{W})_{\text{Model III}} = (\lambda\hat{W})_{\text{Model I}} < (\lambda\hat{W})_{\text{Model II}}$$

6. If $\hat{B} = E(\text{BP})$, then

$(\lambda \hat{B})_{\text{Model III}} < (\lambda \hat{B})_{\text{Model I}} = (\lambda \hat{B})_{\text{Model II}}$

7. One can find a better state-dependent model which offers better operating characteristics than Models I, II, III & IV (or M/M/1). An example is provided by the Model V characterized by the arrival rate $\lambda_n = \dfrac{\lambda}{2n+1}$, $n \geq 0$, service rate $\mu_n = 2n\mu$, $n \geq 1$ ($\lambda > 0$, $\mu > 0$) where $\gamma_{\text{Model V}} = 0.352$ when $\rho = 1$.

8. With reference to the system size, $\hat{N}$ in steady state, it is proved that Model III is superior to the Model I or Model II. However, Model I or Model II is superior to the standard M/M/1 Model, under the same criterion.

9. By analyzing the general linear and non-linear state-dependent queueing models in section 7.5, we establish that there are linear and non-linear state-dependent queueing models having the same probability distribution of queue length. The number of customers served during the waiting time of an arbitrary arrival in the system is less for the non-linear state-dependent queueing Models than the linear state-dependent queueing Model. But the number of customers at the server or the proportion of time the server is working is equal to the minimum number of servers to cope with the arrival traffic (i.e., $\rho^* = \rho$).

STATE-DEPENDENT MARKOVIAN QUEUEING MODELS AND THEIR PERFORMANCE EVALUATIONS

Thesis submitted to the

University of Calicut

For the award of the Degree of

DOCTOR OF PHILOSOPHY

Under the Faculty of Science

by

Narayanan, V.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

October 2006

REFERENCES

1. Ancker, C.J., Jr. and Gafarin, A.V (1963), Some queueing problems with the balking and reneging 2, Operations Research 11, pp. 928-937.

2. Asmussen, S (2003), *Applied Probability and Queues* second edition, Springer, New York.

3. Bailey, N (1952), A study of queues and appointment systems in hospital outpatients departments with special reference to waiting times. *Journal of Royal Statistical Society*, 14, 2, pp. 185-199.

4. Bailey, N.T. J (1954), A continuous time treatment of a simple queue using generating functions, *Journal of Royal Statistical Society*, B, 16, pp. 288-291.

5. Bhat, U.N (1964), On a single server bulk queueing processes with binomial input, *Operations Research*, 12, pp. 527-533.

6. Brian Conolly (1975), *Lecture Notes on Queueing Systems*, Ellis Horwood Limited Chichester.

7. Brockwell, P.J., Resnick, S.L., and Tweedie, R.L (1982), Storage processes with general release rule and additive inputs, *Advances in Applied Probability*, 14, pp.392-433.

8. Brian D. Bunday (1986), *Basic Queueing Theory*, Edward Arnold Publishers Limited, Australia.

9. Baccelli, F. and Bermand P (1994), *Elements of Queueing Theory*, Springer, New York.

10. Bertrand, J.W.M., and Van Ooijen, H.P.G (2002), Workload based order release and productivity: a missing link, *Production Planning and Control*, 13, pp. 665-678.

11. Cox, D.R., and Smith, W.L (1961), *Queues Methuen*, London.

12. Conway, R.W, and Maxwell, W.L (1962), A queueing model with state dependent service rates, *Journal of Industrial Engineering*, 12, pp. 132-136.

13. Conolly, B.W., and Hadidi, N (1968), A correlated queue, *Journal of Applied Probability*, 6, pp. 122-136.

14. Cohen, J.W (1969), *The Single Server Queue*, North Holland, Amsterdam.

15. Conolly, B.W., and Hadidi, N (1969), A comparison of the operational features of a self regulating system, *Applied Statistics*, 18, pp. 41-53.

16. Conolly, B.W (1971), Some applications of theory of infinite capacity service systems to a single server system with linearly state-dependent service, *Journal of Applied Probability*, 8, pp. 202-207.

17. Conolly, B.W (1975), The generalized state-dependent Erlangian queues: Speculations about calculating measures of effectiveness, *Journal of Applied Probability*, 12, pp. 358-363.

18. Cooper, R.B (1981), *Introduction of Queueing Theory*, second edition, New York, North Holland.

19. Donald Gross and Carl, M. Harris, *Fundamentals of Queueing Theory*, third edition, John Wiley & Sons, New York.

20. Friedrich Losch (1960), *Tables of Higher Functions,* sixth edition, McGraw Hill Book Company, New York.

21. Flatto, L. (1997), The waiting time distribution of random order service M/M/1 queues, *The Annals of Applied Probability*, vol. 7, No. 2, pp. 382-409.

22. Gupta, U.C., and Srinivasa Rao, T.S.S (1998), On the analysis of single server finite queue with state-dependent arrival and service process, *OR. Spectrum*, 20, pp. 83-89.

23. Haight, F.A (1957), Queueing with balking, *Biometrika*, 1, 46, pp. 360-369.

24. Haight, F.A (1959), Queueing with reneging, *Biometrika*, 2, pp. 186-197.

25. Harris, C.M (1967), Queues with the state-dependent stochastic service rates, *Operations Research,* 15, pp. 117-130.

26. Hadidi, N (1969), On the service time distribution and waiting process of a potentially infinite capacity queueing systems, *Journal of Applied Probability*, 6, pp. 594-603.

27. Hadidi, N., and Conolly, B.W (1969), On the improvement of the operational characteristics of a single server queue mechanism, *Applied Statistics*, 18, pp. 229-240.

28. Harrison, J.M., Resnick, S.L (1976), The stationary distribution and first exists probability of a storage process with general release rule, *Mathematics of OR*, 1, pp. 347-358.

29. Jackson, J.R (1963), Job shop-like queueing systems, *Management Science*, 10, pp. 131-142.

30. Jenyenz Cheah and J, Macgregor Smith (1994), M/G/C/C state dependent queueing models and pedestrian traffic flows, *Queueing Systems*, pp. 365-386.

31. Kendall, D.G (1994), Some recent work and further problems in the theory of queues, *Theory Probability Application*, 9, pp. 1-13.

32. Keilson, J (1964), A review of transient behaviour in the regular diffusion and birth-death process, I, *Journal of Applied Probability*, 1, pp. 247-266.

33. Keilson, J (1965), A review of transient behaviour in regular diffusion and birth-death process, II, *Journal of Applied Probability*, 2, pp. 405-428.

34. Kleinrock, L (1975), *Queueing Systems Vol. 1 Theory*, Wiley New York.

35. Kelly, F.P (1976), The departure process from a queueing system, *Math. Proc. Camb. Phil. Soc.*, 80, pp. 283-285.

36. Kaspi, H., Kella, O., and Perry, D (1996), Dam process with the state-dependent batch sizes and intermittent production processes with state-dependent rates, *Queueing Systems,* 24, pp. 37-57.

37. Ledermann, W., and Reuter, G.E.H (1954), Spectral theory for the differential equations of simple birth and death processes, *Phil. Trans. Roy. Soc.*, A - 246, pp. 321-369.

38. Mirasol, N.M (1963), The output of $M|G|\infty$ queueing system is Poisson, *Operations Research,* 11, pp.282-284.

39. Muntz, R.R (1973), Poisson departure process and queueing networks, *IBM Research Report RC* 4145, IBM Thomas, J. Watson Research Center, Yorktown, Heights, New York.

40. Mosner, M (1973), Simple server queues with service time dependent on waiting time, *Operations Research* 21, pp. 610-616,.

41. Maister, D.A (1985), The psychology of waiting lines, J.A. Czepiel, M.R. Solomon, C.F. Surprenent, eds, The Service Encounter Health and Company, Lexington, M.A., pp. 113-123.

42. Mitra, D., and Elwalid, A.I (1994), Statistical multiplexing with loss priorities in rate based conjunction control of high speed networks, *IEEE Trans. Commun.* 42, pp. 2989-3002.

43. Minh, D.L (2002), *Applied Probability Models*, Thomson Asia Pvt. Ltd., Singapore.

44. Manoharan, M., Alamatsaz, M.H., and Shanbag, D.N (2003), Departures and Related Characteristics in queueing models, *Handbook of Statistics, Vol.21, Chapter 15*, pp. 557-572.

45. Manoharan, M., Narayanan, V. (2006), Performance measures of state-dependent Markovian queueing models (Presented in the National Seminar on Stochastic process modeling, Distribution theory and Order statistics & Annual Conference of Kerala Statistical Association, Kerala University, March 16-18, 2006.

46. Natvig, B (1973), On the busy period for a queueing model where potential customers are discouraged by queue length, *Department of Mathematics, University of Trondheim - NTH.*

47. Natvig, B (1974), On the transient state probabilities where potential customers are discouraged by queue length, *Journal of Applied Probability,* 11, pp. 345-354.

48. Natvig, B (1975), On a queueing model where potential customers are discouraged by queue length, *Scand. J. Statist.* 2, pp. 34-42.

49. Manoharan, M., Alamatsaz, M.H., and Shanbag, D.N (2003), Departures and related characteristics in queueing models, *Handbook of Statistics*, vol. 21, Chapter 15, pp. 557-572.

50. Narayanan, V., and Manoharan, M (2006), On some state-dependent Markovian queueing models (under revision with 4 OR).

51. Narayanan, V., and Manoharan, M (2006), On some state-dependent Markovian queueing models (under revision).

52. Narayanan, V., and Manoharan, M (2006), Some results on the state-dependent Markovian queueing models (communicated).

53. Narayanan, V., and Manoharan, M (2006), On the waiting time distribution of Markovian queue models with state-dependent arrival rates (communicated).

54. Narayanan, V. Manoharan, M., (2006), Transient analysis of an adaptive queueing system (Presented in the National Seminar on Stochastic process modeling, Distribution theory and Order statistics & Annual Conference of Kerala Statistical Association, Kerala University, March 16-18, 2006.

55. Prabhu, N.U (1997), *Foundations of Queueing Theory*, Kluwer Academic Publishers, USA.

56. Parthasarathy, P.R., and Selvaraju N (2001), Transient analysis of a queue where potential customers are discouraged by queue length, *Mathematical Problems in Engineering*, vol. 7, pp. 433-454.

57. Rosenshine, M (1967), Queues with state-dependent service times, *Transportation Research I*, pp. 97-104.

58. Shanbag, D.N., and Tambouratzis, D.G (1973), Erlang's formula and some results on the departure process for a loss system, *Journal of Applied Probability*, 10, pp. 233-240.

59. Susana, V., Mondschein and Gabrial, Y., Weintraub (2003), Appointment policies in service operations: A critical analysis of the economic frame work, *Production and Operations Management*, vol. 2, No. 2, pp. 266-286.

60. Tackac's, L (1955), Investigating waiting time problems while reducing to Markove process, *Acta, Math. Acad. Sci.* Hung 6, pp. 101-129.

61. Thomas, L. Satty (1961), *Elements of Queueing Theory*, McGraw Hill Book Company, New York.

62. Taylor, S (1994), Waiting for service: The relationship between delays and evaluation of service, *Journal of Marketing*, 58, pp. 56-69.

63. Vandoorn, E.A. (1981), The transient state probabilities for a queueing model where potential customers are discouraged by queue length, *Journal of Applied Probability*, 18, pp. 499-506.

64. Van Ooijen, H.P.G., and Bertrand, J.W.M (2003), The effects of a single arrival rate control policy on thoroughput and work-in-process in production systems with work load dependent processing rates, *International Journal of Production Economics*, 85, pp. 61-68.

65. Weber, R.R (1983), A note on waiting time in a single server queues, *Operations Research*, 31, pp. 950-957.

66. Wayne, L. Winston (2004), *Introduction to Probability Models*, fourth edition, Thomson Asia Pvt. Ltd., Singapore.

67. Xiuli Chao and Carlton Scott (2000), Several results in the design of queueing system, *Operations Research*, vol. 48, No. 6, pp. 965-970.