2006

Single server retrial queueing models.

Xiaoyong Wu
University of Windsor

SINGLE SERVER RETRIAL QUEUEING MODELS

by

XIAOYONG WU

A Dissertation
Submitted to the Faculty of Graduate Studies and Research
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

2005

Your file Votre référence
ISBN: 978-0-494-17128-8
Our file Notre référence
ISBN: 978-0-494-17128-8

Abstract

Most retrial queueing research assumes that each retrial customer has its own orbit, and the retrial customers retry to enter service independently of each other. A small selection of papers assume that the retrial customers themselves form a queue, and only one customer from the retrial queue can attempt to enter at any given time. Retrial queues with exponential retrial times have been extensively studied, but little attention has been paid to retrial queues with general retrial times.

In this thesis, we consider four retrial queueing models of the type in which the retrial customers form their own queue. Model I is a type of $M/G/1$ retrial queue with general retrial times and server subject to breakdowns and repairs. In addition, we allow the customer in service to leave the service position and keep retrying for service until the server has been repaired. After repair, the server is not allowed to begin service on other customers until the current customer (in service) returns from its temporary absence. We say that the server is in "reserved" mode, when the current customer is absent and the server has already been repaired. We define the server to be blocked if the server is busy, under repair or in reserved mode.

In Model II, we consider a single unreliable server retrial queue with general retrial times and balking customers. If an arriving primary customer finds the server blocked, the customer either enters a retrial queue with probability p or leaves the system with probability $1-p$. An unsuccessful arriving customer from the retrial queue either returns to its position at the head of the retrial queue with probability q or leaves the system with the probability $1-q$. If the server fails, the customer in service either

remains in service with probability r or enters a retrial service orbit with probability $1 - r$ and keeps returning until the server is repaired.

We give a formal description for these two retrial queueing models, with examples. The stability of the system is analyzed by using an embedded Markov chain. We get a necessary and sufficient condition for the ergodicity of the embedded Markov chain.

By employing the method of supplementary variables, we describe the state of the system at each point in time. A system of partial differential equations related to the models is derived from a stochastic analysis of the model. The steady state distribution of the system is obtained by means of probability generating functions. In steady state, some performance measures of the system are reported, the distribution of some important performance characteristics in the waiting process are investigated, and the busy period is discussed. In addition, some numerical results are given.

Model III consists of a single-server retrial queue with two primary sources and both a retrial queue and retrial orbits. Some results are obtained using matrix analytic methods. Also simulation results are obtained.

Model IV consists of a single server system in which the retrial customers form a queue. The service times are discrete. A stability condition and performance measures are presented.

I dedicate my efforts to:

my parents for their love and support;

my wife, Yin Zhou, for her love, support and encouragement; and

my children, David (Yeda) and Katherine (Yedi), for giving meaning to it all.

Acknowledgements

I wish to express my deep appreciation to my co-supervisors Dr. M. Hlynka and Dr. P. Brill for their guidance and encouragement in every stage of this dissertation. Thanks are also due to committee members Dr. M. Hlynka, Dr. P. Brill, Dr. S. R. Chakravarthy, Dr. T. Traynor, Dr. A. Hussein and Dr. S. Goodwin for their useful comments and suggestions which greatly improved this dissertation.

I would like to thank the Department of Mathematics and Statistics, and the University of Windsor, for giving me financial support in terms of graduate teaching assistantships and tuition scholarships. The financial support from Dr. Myron Hlynka and Dr. Percy Brill through their NSERC research grants during my graduate studies is gratefully acknowledged.

I would like to thank Dr. M. Hlynka, Dr. P. Brill, Dr. S. R. Paul, Dr. S. E. Ahmed, Dr. C. S. Wong, Dr. K. Fung, Dr. A. Hussein, Dr. T. Traynor for their readiness to help and for stimulating discussions during my graduate course work in Statistics.

Finally, I would like to thank my parents, my wife, my daughter and my son for their love, support and understanding.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The earliest queueing paper with "retrial" in the title appears to be by Fredericks and Reisner (1979), although earlier papers on the subject of retrial queues exist, e.g. (Cohen, 1957). Retrial queues have applications to performance evaluation of telephone switching systems, telecommunication networks, computer networks and computer systems (see Choi and Chang, 2001; Krishna Kumar et al, 2002; Wang et al, 2001).

Retrial queues have the following basic and important characteristic. When all servers are busy and the waiting room accessible for service is full, an arriving customer will leave the service area but after some random time repeat its demand for service over and over until the customer obtains service. That is, an arriving customer becomes a potential source of repeated customers. This implies that the flow of customers consists of two parts: the flow of primary customers, each of which would like immediate service, and the flow of repeated customers due to failure of the previous attempt to access the server. Therefore, retrial queues are different from ordinary queueing systems, in which a customer arriving when the server is busy and the buffer is full, must leave the system. Yang and Templeton (1997), Falin (1990) and Kulkarni and Liang (1997) provide extensive surveys on retrial queues.

Historically, retrial queues with exponential retrial times have received considerable attention. "The first results on the $M/G/1$ retrial queues are due to Keilson, Cozzolino and Young (1968) who used the method of supplementary variables to investigate the joint distribution of the channel state and the number of customers in orbit in steady state." (Falin (1975)). Aleksandrov (1974) and Falin (1975) considered the case of arbitrarily distributed service times and obtained the joint distribution of the server state and the queue length in steady state by methods different from those used in Keilson, Cozzolino and Young (1968). Variations have been studied by a number of authors including, Falin (1986), Kulkarni (1983), Choi and Park (1990), Neuts and Ramalhoto (1984), Farahmand (1990), Yang and Li (1994) and Wang, Cao and Li (2001), Chakravarthy and Dudin (2003, 2004, who allowed for MAP and BMAP arrivals).

However, less attention has been paid to the $M/G/1$ retrial queues with non-exponential or general retrial times. The first results on this model are due to Kapyrin (1977), who assumed that each customer in the retrial orbit generates a stream of repeated attempts independent of other customers in the retrial orbit and the server state. Later, a method convenient for practical applications, based on the stochastic decomposition property, was proposed in Yang, Posner, Templeton and Li (1994). They developed an approximation method for the calculation of steady state performance measures of the queue described by Kapyrin (1977).

A model where only one customer at a time from orbit can retry for service was suggested by Fayolle (1986). Farahmand (1990) calls such a system a retrial queue with a FCFS orbit. We will refer to such queueing systems as retrial queues with FCFS discipline. Under Markovian assumptions, Fayolle obtained a necessary and sufficient condition for ergodicity. The analysis of the waiting time process is a much more difficult problem. Fayolle gave a solution of the problem in terms of a meromorphic function, the poles and residues of which are easily computed recursively. He also investigated the tail behavior of the waiting time distribution in steady state. Fayolle investigated a telephone exchange model as an $M/M/1$ retrial queue in which an orbiting customer who finds the server busy joins the tail of a retrial queue and only the customer at the head of the retrial queue is allowed to attempt to receive service after an exponentially distributed retrial time. Farahmand (1990) considered a model in which the channel holding times are generally distributed and retrial times have an exponential distribution.

Recently, $M/G/1$ retrial queues with general retrial times and with a FCFS discipline received some attention. An $M/G/1$ retrial queue with general retrial times was proposed by Gómez-Corral (1999), who assumed that customers who find the server busy are queued in the retrial queue according to a FCFS discipline and only the customer at the head of the retrial queue is allowed access to the server. A retrial attempt begins only when the server completes a service rather than at a service attempt failure. According to Kumar and Arivudainambi (2002), this kind of retrial queue is important in the control policy for the ALOHA protocol in communication systems. A retrial queue with two-phase service and preemptive resume was studied by Kumar, Vijayakumar and Arivudainambi (2002), who discussed system stability and some performance measures.

A model allowing non-persistence of customers was first considered by Cohen (1957) for the $M/M/c$ type retrial queue, with impatient customers. The model allows some repeated customers to leave the system without service. Keilson and Servi (1993) considered a wide class of Markovian single-server retrial queues with non-persistence of customers.

The study of single-server queues with an unreliable server goes back to Avi-Itzhak and Naor (1962), continued by Thiruvengadam (1962), Mitrany and Avi-Itzhak (1968), Neuts and Lucantoni (1979), Cao and Cheng (1982). A single-server retrial queue with unreliable server was considered by Kulkarni (1990), Aissani (1993), Yang and Li (1994), Artalejo (1994) and Aissani (1995) and Wang, Cao and Li (2001).

In this thesis, we extend Gómez-Corral's model to allow for service breakdown and add the new concept of a retrial (service) orbit for the customer in service, during which the server is "reserved" for that customer. A condition for the stability of the system is presented, with a variety of justifications. In steady state, some important performance characteristics are analyzed, such as, the distribution of the queue length, the distribution of the waiting time, probabilities for the system to be empty and busy. Also, an $M/G/1$ retrial queue with general retrial times, balking or non-persistence of customers and unreliable server is discussed. A mixed model with a FCFS discipline retrial queue and retrial orbits is considered and analyzed using matrix analytic methods.

Retrial queues with continuous service times have been extensively studied. However, little attention has been paid to retrial queues with discrete service times. One paper that allows for deterministic service is Kobe (2000) in which the author studied system stability for an $M/D/1$ retrial queue with a limited waiting time. In this thesis, we consider a retrial queue with FCFS discipline in which each customer has a discrete service taking value D_j with probability p_j, $j = 1, 2, \ldots$. We call this retrial queue an $M/\{D_n\}/1$ retrial queue. The $M/D/1$ retrial queue is a special case. Classical queueing systems with discrete service times have been considered by the authors such as Erlang (translation in 1948), Crommelin (1932) , Iversen and Staalhagen (1999), Franx (2001), Brun and Garcia (2000), Brill (2002), Shortle and Brill (2005).

The rest of the thesis is organized as follows. In Chapter 2, we discuss Model I, a specific $M/G/1$ retrial queue with general retrial times and server subject to breakdown, repairs. A new feature introduced in this thesis is the concept of a retrial of a customer in service during server breakdown. While the server is being repaired, the customer in service can leave temporarily and return repeatedly to see if the repair is complete. After repair, the server must wait for the customer in service to return. The time between when the server has been repaired and the return of the customer in service is called the "reserved" time. In Chapter 2, we introduce some basic background and preliminaries. A complete description of the model with definitions and conventions is given. The evolution of the queueing process is exhibited. The states of the system are defined through elapsed times. A necessary and sufficient condition for the stability of the system is obtained by analyzing the ergodicity of an embedded Markov chain. Several examples of systems which can be modelled by our retrial queue are provided.

In Chapter 3, we follow the procedure of the method of supplementary variables and derive a system of forward equations for the model of Chapter 2. The steady state solution for the model is obtained with the help of generating functions. In steady state, we are interested in some main performance measures, such as the probabilities that the system is in the different states and the expected value of the number of customers in the retrial queue. An explicit expression for these performance measures is derived. In the waiting process, we also discuss the waiting time that a primary customer spends in the retrial queue and its expected value in steady state.

Chapter 4 discusses Model II, an $M/G/1$ retrial queue which has general retrial times, balking customers and unreliable server, and a reserved server for the customer in service orbit. We fully describe the model and the evolution of the process. The stability of the system is investigated. Some comments about the stability condition are made.

In Chapter 5, the system of forward equations for Model II is derived by considering transitions between states of the system. Given that the system is stable, we obtain the steady state distribution of the system. For the waiting process and the busy period, we discuss the distribution of some important performance measures.

In Chapter 6, we introduce Model III, which consists of a single server retrial queue with two primary sources and two different types of retrials. The description of the model and a sufficient condition for the stability of the system are given.

In Chapter 7, we look at special cases of Model III, and obtain some numerical results. In Chapter 8, we propose Model IV, a retrial queue with discrete service times. In Chapter 9, we obtain some results on the waiting time and the busy period of Model IV.

In Chapter 10, we give some graphical results for Models I and II. In Chapter 11, we give some simulation results for Model III. In Chapter 12, we present possibilities for future work.

The unifying theme for all four models in this thesis is the use of a queue for retrial customers rather than having the retrial customers each behave independently. In Models I and II, there is also a retrial within the service area. In model III, the retrial queue type model is combined with the more popular retrial model which treats retrial customers as independent entities. Because of the complexity already inherent in Model III, the condition of allowing the customer in service to leave and return is removed. Model IV looks at general discrete service time retrial queues, again with the retrial customers joining a retrial queue. Since so little work has been done on discrete service retrial queues, we did not allow the retrial of customer in service during breakdown.

Chapter 2

Model I Description and Stability

2.1 Description of Model I

The model described in this chapter is referred to as Model I. We consider a single-server retrial queue in which primary customers arrive according to a Poisson process with rate λ. We assume that there is only room for the customer being served and there is no waiting room. If an arriving primary customer finds the server idle, the customer begins service immediately and leaves the system after service completion. The server is said to be blocked if the server is busy, under repair or reserved. If an arriving primary customer finds the server blocked, the customer enters a retrial queue according to a FCFS discipline and becomes a repeated customer. (We give examples of this type of model later in this chapter.) When the server becomes idle, the customers from the retrial queue compete with new primary customers to reach the server first. For customers in the retrial queue, only the customer at the head of the retrial queue is allowed to attempt to reach the server, in a time generally distributed with distribution function $A(x)$, density function $a(x)$ and Laplace transform $L_A(s)$, measured from the instant the server becomes idle. The time generated by $A(x)$ is called a retrial time. The retrial customer is required to cancel the attempt for service if a primary customer arrives first while the server is idle. In that case, the retrial customer returns to its position in the retrial queue. The time in which a customer is being served by the server is called the service time, generally distributed with distribution function $B(x)$, density function $b(x)$, Laplace transform $L_B(s)$ and first two moments β_1 and β_2 which are equal to $-\frac{d}{ds}L_B(s)|_{s=0}$ and $\frac{d^2}{ds^2}L_B(s)|_{s=0}$. This time does not include any repair time which may be needed.

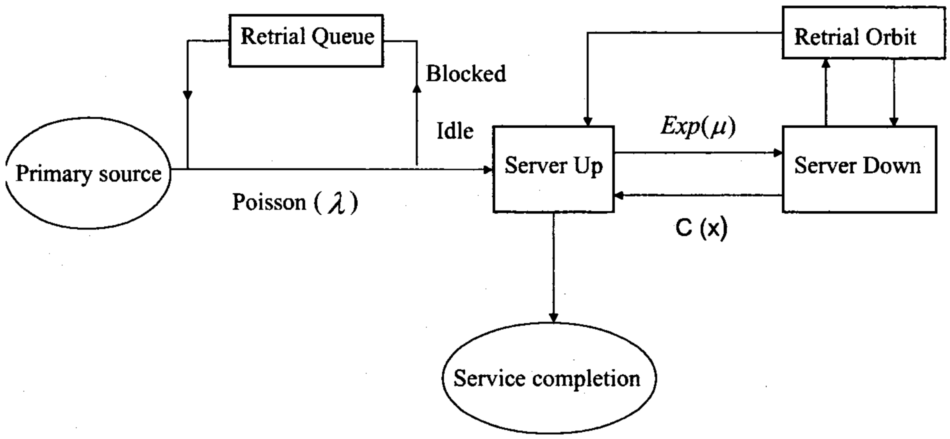
We assume that the server may fail in a time exponentially distributed with mean $\frac{1}{\mu}$, but failure can occur only when a customer is being served.

When the server fails, repair begins immediately. The repair time has distribution function $C(x)$, density function $c(x)$, Laplace transform $L_C(s)$ and first two moments γ_1 and γ_2 which are equal to $-\frac{d}{ds}L_C(s)|_{s=0}$ and $\frac{d^2}{ds^2}L_C(s)|_{s=0}$.

During the repair time, the customer in service enters a retrial orbit, different from the retrial queue, and keeps repeating its request for service continually at times exponentially distributed with mean $\frac{1}{\theta}$, until the server is repaired. The time length of each retrial orbit, is called an orbit time.

After repair, the server must wait for the customer from the retrial orbit to return. During this time, we say the server is "reserved" and refer to the time as the reserved time. The reserved time is exponentially distributed with mean $\frac{1}{\theta}$. The description of the model is illustrated in Figure 2.1.

Figure 2.1: An M /G /1 retrial queue with retrials of the customer in service



The service time for a customer is cumulative (recalling that the server may breakdown) and the server is as good as new after repair. The service time for a customer is resumed after repair time and reserved time. Note that the service time is not the length of time measured from when a customer begins to be served until its service is completed because of possible breakdowns. We define the generalized service time as the length of time from when a customer begins to be served until the service is completed. We assume that retrial times, service times, repair times and reserved times are mutually independent. The time until failure measured from the start of service of a customer, is independent of the other times except for the service time, which must be an upper bound on the time until failure, given that a failure occurs. However we assume that the time until the next failure is generated independently of the service times from an exponential distribution and no failure occurs if the generated value exceeds the remaining service time.

Such random variables can be expressed in terms of the conditional completion rate (or hazard rate or the specific failure rate) $\alpha(x), \beta(x)$ and $\gamma(x)$ as

$$\alpha(x) = \frac{a(x)}{1 - A(x)}, \quad \beta(x) = \frac{b(x)}{1 - B(x)}, \quad \gamma(x) = \frac{c(x)}{1 - C(x)}.$$

The Laplace-Stieltjes transform of the retrial time is defined as

$$L_A(s) = \int_0^\infty exp\,(-sx)\, dA(x),$$

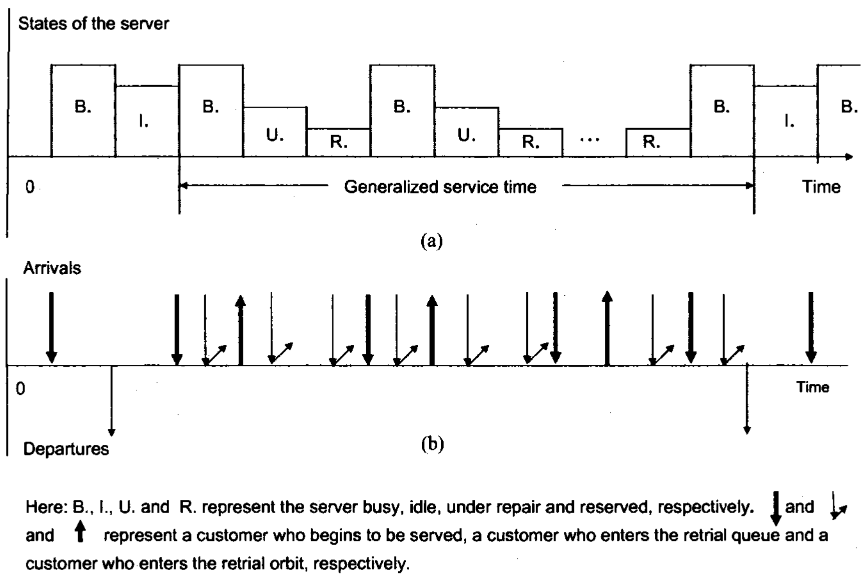
and similarly for $L_B(s)$, $L_C(s)$, $s \geq 0$.

2.2 Evolution of the queueing system

Suppose that the customers in the system are numbered by their order of service. Let d_{i-1}, $i \geq 2$, be the successive instants of completion of service, i.e., the successive departure epochs, when the $(i-1)$st customer completes its service and leaves the system. At each d_{i-1}, $i \geq 2$, the server becomes idle. Since there was no waiting room in front of the server, an idle time κ_i occurs, after which the server begins serving a primary customer or a customer from the retrial queue. At d_{i-1}, if the retrial queue is empty, κ_i is exponentially distributed with rate λ. If the retrial queue is nonempty, then there is a competition between an arriving primary customer and an arriving retrial customer to determine who will be the ith customer in service. That is, whether a primary customer or a retrial customer will be the ith customer in service depends upon who arrives first. If the retrial queue is nonempty, then κ_i

is equal to the minimum of two independent times, one of which is exponentially distributed with rate λ and the other is generally distributed with Laplace transform $L_A(s)$. The probability that the ith customer is a primary customer is equal to $1 - L_A(\lambda)$ and the probability that the ith customer is a retrial customer is equal to $L_A(\lambda)$. This is because the probability that a primary customer wins the competition is $\int_0^\infty \int_0^y \lambda e^{-\lambda x} dx dA(y) = 1 - L_A(\lambda)$, while the probability that a retrial customer wins the competition is $\int_0^\infty \int_y^\infty \lambda e^{-\lambda x} dx dA(y) = L_A(\lambda)$. At time $d_{i-1} + \kappa_i$, the ith customer begins to be served. Suppose next that the server fails after a time $S_i^{(1)}$. Thus, at time $d_{i-1} + \kappa_i + S_i^{(1)}$, the server begins to be repaired and the ith customer enters the retrial orbit and retries in exponential times with mean $\frac{1}{\theta}$ until the server is repaired. Once the server is up after a repair time $w_i^{(1)}$, generally distributed with distribution function $C(x)$, the server will resume service for the customer in the retrial orbit only after a reserved time $\varsigma_i^{(1)}$ which by the memoryless property, is also exponentially distributed with rate θ. Otherwise, if the customer in the retrial orbit arrives before the server is up, the customer must reenter the retrial orbit again. Possibly after several retrials, the customer reenters service again after a repair time plus a reserved time $w_i^{(1)} + \varsigma_i^{(1)}$. This implies that at time $d_{i-1} + \kappa_i + S_i^{(1)} + w_i^{(1)} + \varsigma_i^{(1)}$, the ith customer resumes service. The remaining components of the ith generalized service time occurs similarly. Suppose that k represents the number of failures during the ith generalized service time, then the ith customer completes service at time $d_i \equiv d_{i-1} + \kappa_i + S_i^{(1)} + w_i^{(1)} + \varsigma_i^{(1)} + S_i^{(2)} + w_i^{(2)} + \varsigma_i^{(2)} + ... + S_i^{(k)} + w_i^{(k)} + \varsigma_i^{(k)} + S_i^{(k+1)}$, where $S_i \equiv S_i^{(1)} + S_i^{(2)} + ... + S_i^{(k+1)}$ represents the real service time, $w_i \equiv w_i^{(1)} + w_i^{(2)} + ... + w_i^{(k)}$ represents the total repair time and $\varsigma_i \equiv \varsigma_i^{(1)} + \varsigma_i^{(2)} + ... + \varsigma_i^{(k)}$ represents the total reserved time. The evolution of the queueing process during the ith generalized service time is illustrated in Figure 2.2. For $d_i, d_{i+1}, \ldots$, the queueing process goes on in a similar way.

Figure 2.2: Queueing process during the ith generalized service time. (a) States of the server versus time. (b) Arrivals and departures versus time.



2.3 Examples related to the model

There exist a number of practical examples which are modelled by an $M/G/1$ retrial queue with general retrial times and server subject to breakdowns, repairs and reservations. Some examples follow.

Example 2.1 Consider a call center having a single server with a telephone and an answering machine. If a customer calls and finds the server blocked, the customer leaves a message on the answering machine according to a FCFS discipline. Thus, all calls come from a primary source, while all messages in the answering machine form a retrial queue. When the server becomes idle, the messages on the answering machine are checked and the first caller on the answering machine is contacted unless a primary customer phones the center before the contact is made. The server may fail (or be interrupted) during service. During the repair time, the telephone is reserved by the customer in service. The customer may become impatient and begin other tasks, checking occasionally to see if the server is up. After repair, the server has to wait for the customer to return to the line before continuing service. (This is precisely how the University of Windsor Information Technology Help Center operates although there are limits on the amount of time that the server will continue to wait for return of the customer in service, during the temporary absence.)

Example 2.2 (This is a slight variant of Example 2.1) A person phones a call center asking for service and finds the server busy. The message on the phone invites the customer to e-mail a message detailing the type of service needed. If the customer sends an e-mail, that e-mail becomes a retrial customer with FCFS discipline. When the server is ready to serve another customer, the server will approach the oldest e-mail message first. If the server is in the middle of reading the message, the server will not allow a phone call to interrupt his reading. If the server is just about to start reading, a new phone call will represent a primary customer entering ahead of the retrial customers. If the server is dealing with a retrial customer message, the server may phone the customer with information and be put on hold (so the system is in reserved mode). At the end of the holding period the message is delivered and service is completed.

Example 2.3 A manufacturing company has a machine shared by all units of the company. If the machine is busy, a new arrival who needs the machine signs up on a waiting list. When a service is completed, the machine manager contacts the next person on the list unless another arrival enters before the contact is made. The next person accepted for the machine is the current user. The waiting list corresponds to the retrial queue. Their server is the machine. While in service, the machine may break down or operate poorly. During repair, the user leaves to perform other tasks, returning at various times to see if the machine is fixed, so that service can continue. The machine cannot accept new jobs while in reserved mode.

Example 2.4 (This example is only intended to illustrate the nature of a system being in reserved mode with the server up and the customer absent. It is not an example of the whole model). Consider a grocery store with a line of customers. One customer in line is in the middle of service, with about half of the grocery items being registered in the cash register. The server (cashier) needs to change the receipt paper roll. Since there is no activity, the customer takes advantage of the down time to rush off to pick up another item. Then the cashier completes the paper roll change. The server cannot continue beyond the items present however, since the customer is temporarily absent. No other customer can use the server, who is in the middle of a transaction. Here the server is in "reserved" mode.

2.4 States of the system

We present a table which lists the states of the system and supplementary variables. The full description appears below.

$J(t)$	State of Server
0	idle
1	busy
2	under repair
3	reserved
$\xi_0(t)$	elapsed retrial time
$\xi_1(t)$	elapsed service time
$\xi_2(t)$	elapsed repair time
$\xi_3(t)$	elapsed reserved time

Since the distribution of the service time is not exponential, the stochastic process $\{(J(t), Q(t)); t \geq 0\}$, as in ordinary $M/G/1$ queues, is not necessarily Markovian, where at time t, $J(t)$ represents the server state (0, 1, 2 or 3 stand for the server idle or busy or under repair or reserved, respectively) and $Q(t)$ denotes the number of customers in the retrial queue. In order to analyze the stochastic process, we introduce supplementary variables, as referred to by Cox (1955), Kleinrock (1975), Riordan (1962), Keilson, Cozzolino and Young (1968). If $J(t) = 0$ and $Q(t) > 0$, we define $\xi_0(t)$ as the elapsed retrial time of a customer in the retrial queue. If $J(t) = 1$ or 2 or 3, we define $\xi_1(t)$ as the elapsed service time; if $J(t) = 2$, we define $\xi_2(t)$ as the elapsed repair time. If $J(t) = 3$, we define $\xi_3(t)$ as the elapsed reserved time of a customer in the retrial orbit. Thus, the stochastic process $\{X(t); t \geq 0\} = \{(J(t),\ Q(t),\ \xi_0(t),\ \xi_1(t),\ \xi_2(t),\ \xi_3(t)); t \geq 0\}$ is a Markovian process with the state space $\Omega = \{(0,\ 0)\} \cup \{(0,\ i,\ w); i = 1, 2, ...,\ 0 \leq w < \infty\} \cup \{((1,\ i,\ x),\ (2,\ i,\ x,\ y),\ (3,\ i,\ x,\ \tau)); i = 1, 2, ...,\ 0 \leq x,\ y,\ \tau < \infty\}$, where $(0,\ 0)$ means that the server is idle and the retrial queue is empty; $(0,\ i,\ w)$ means that the server is idle and there are i customers in the retrial queue with elapsed retrial time w; $(1,\ i,\ x)$ means that the server is busy with elapsed service time x and i customers are in the retrial queue; $(2,\ i,\ x,\ y)$ means that the server is under repair with elapsed service time x and with elapsed repair time y and i customers are in the retrial queue; $(3,\ i,\ x,\ \tau)$ means that the server is reserved with elapsed service time x and with elapsed reserved time τ and i customers are in the retrial queue.

2.5 An embedded Markov chain

Let us consider the system immediately after a customer completes service and leaves the system. The next service will begin after an idle period. Let Q_n be the number of customers in the retrial queue just after the nth departure point in the retrial queue. Let $S^{(n)}$ be the nth generalized service time. Let D_{n+1} be the number of primary customers arriving during the $(n+1)$st generalized service time $S^{(n+1)}$ of the $(n+1)$st customer. As stated in section 3, if the retrial queue is nonempty. then there is a competition between an arriving primary customer and an arriving retrial customer. For convenience, let "$t_p < t_r$" denote the event that the primary customer wins the competition and "$t_r < t_p$" denote the event that the retrial customer wins the competition.

Therefore, we have

$$Q_{n+1} = \begin{cases} D_{n+1}, & \text{if } Q_n = 0, \\ Q_n + D_{n+1}, & \text{if } Q_n > 0 \text{ and } t_p < t_r, \\ Q_n + D_{n+1} - 1, & \text{if } Q_n > 0 \text{ and } t_r < t_p. \end{cases} \tag{2.1}$$

Let S_n, $n = 1, 2, ...$, denote the nth service time. By our model assumption, S_n, $n = 1, 2, ...$, are independent and identically distributed. The time until the server fails, measured from the start of service, is exponentially distributed at rate μ. Recall that S^n is the generalized service time. For convenience, we suppress the superscript and denote a generalized service time by S^*. The number of failures during the time S^* is Poisson distributed with mean μS where S is a service time. Suppose the server fails exactly k times in time S. Then S^* is the sum of S, plus k independent repair times, plus k independent reserved times. Recall that the repair time has cdf $C(x)$. Recall that the service time has cdf $B(x)$. Let $B^*(x)$ be the cdf of the generalized service time S^*. Then

$$\begin{aligned} B^*(x) = P(S^* \le x) = \int_0^x \frac{(\mu y)^0}{0!} e^{-\mu y}\, dB(y) \\ + \int_0^x \sum_{k=1}^{\infty} \frac{(\mu y)^k}{k!} e^{-\mu y}\, dB(y) \int_0^{x-y} \frac{\theta(\theta z)^{k-1}}{(k-1)!} e^{-\theta z} C^{(k)}(x-y-z)\, dz, \end{aligned} \tag{2.2}$$

where $C^{(k)}(\cdot)$ denotes the k-fold convolution of $C(\cdot)$. It can be seen that

$$\begin{aligned} L_{B^*}(s) &= \int_0^{\infty} e^{-sx} \sum_{k=0}^{\infty} \frac{(\mu y)^k}{k!} e^{-\mu y} \left(\int_0^{\infty} e^{-s(y+z)} \theta e^{-\theta z}\, dC(y)\, dz \right)^k dB(x) \\ &= L_B\left(s + \mu - \frac{\mu\theta}{s+\theta} L_C(s)\right), \end{aligned} \tag{2.3}$$

Since the interarrival times of primary customers is exponentially distributed, the random variable D_{n+1} depends only on S^* and not on the retrial queue or on the type of customer initiating service. Let D denote the number of primary customers arriving during a generalized service time. Thus, one-step transition probabilities can be derived as follows.

Consider the following cases.

(i) Suppose that the retrial queue is empty at time d_n. Then only a primary customer can arrive, that is, the $(n+1)$st customer in service must be a primary customer. Therefore, for $i = 0$ and $j \geq 0$, we have

$$P(Q_{n+1} = j \mid Q_n = 0) = \int_0^\infty P(D = j \mid S_* = x)\, dB^*(x)$$
$$= \int_0^\infty \frac{(\lambda x)^j}{j!} e^{-\lambda x}\, dB^*(x). \tag{2.4}$$

(ii) Suppose there are $i\,(i > 0)$ customers in the retrial queue at time d_n and $i-1$ customers in the retrial queue at time d_{n+1}. Then a retrial customer won the competition between a primary arrival and a retrial customer. Also no primary customers arrived during the $(n+1)$st generalized service time. Therefore, for $i > 0$ and $j = i - 1$, we have

$$P(Q_{n+1} = j \mid Q_n = i) = L_A(\lambda) P(Q_{n+1} = j \mid Q_n = i,\, t_r < t_p)$$
$$= L_A(\lambda) \int_0^\infty P(D = 0 \mid Q_n = i,\, S_* = x,\, t_r < t_p)\, dB^*(x)$$
$$= L_A(\lambda) \int_0^\infty e^{-\lambda x}\, dB^*(x). \tag{2.5}$$

(iii) Suppose there are $i\,(i > 0)$ customers in the retrial queue at time d_n and $j\,(> i-1)$ customers in the retrial queue at time d_{n+1}. Then there was a competition between a primary customer and a retrial customer to enter service at time d_n. If a primary customer won this competition, then $(j - i)$ primary customers arrived during the $(n+1)$st generalized service time. Otherwise, if a retrial customer won this competition, then $(j-i+1)$ primary customers arrived during the $(n+1)$st generalized service time. Therefore, for $i > 0$ and $j > i-1$, we have

$$P(Q_{n+1} = j \mid Q_n = i) = (1 - L_A(\lambda))P(Q_{n+1} = j \mid Q_n = i,\, t_p < t_r)$$
$$+L_A(\lambda)P(Q_{n+1} = j \mid Q_n = i,\, t_r < t_p)$$
$$= (1 - L_A(\lambda)) \int_0^\infty P(D_{n+1} = j - i \mid Q_n = i,\, S^{(n+1)} = x,\, t_p < t_r)\, dB^*(x)$$
$$+L_A(\lambda) \int_0^\infty P(D_{n+1} = j - i + 1 \mid Q_n = i,\, S^{(n+1)} = x,\, t_r < t_p)\, dB^*(x)$$
$$= (1 - L_A(\lambda)) \int_0^\infty \frac{(\lambda x)^{j-i}}{(j-i)!} e^{-\lambda x}\, dB^*(x)$$
$$+L_A(\lambda) \int_0^\infty \frac{(\lambda x)^{j-i+1}}{(j-i+1)!} e^{-\lambda x}\, dB^*(x). \tag{2.6}$$

(iv) For all other cases, $P(Q_{n+1} = j \mid Q_n = i) = 0$.

Hence, the sequence of random variables $\{Q_n,\, n \geq 1\}$ constitutes a Markov chain, which is an embedded Markov chain for the queueing system.

2.6 Stability of the system

Theorem 2.1 The inequality $\lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1)) < L_A(\lambda)$ is a necessary and sufficient condition for the system to be stable.

Proof. Let $S^{(n)}$ be the nth generalized service, $n = 1, 2, \ldots$. Then according to (2.3), $\{S^{(n)}\}$ are independent and identically distributed with expected value

$$ES^{(n)} = \beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)).$$

Since primary customers arrive in a Poisson process, we use Burke's condition (given in Cooper, 1981, p. 187) which essentially states that the steady state probabilities of $\{(C(t), Q(t)); t \geq 0\}$ exist and are positive if and only if the Markov chain $\{Q_n, n = 1, 2, ...\}$ is ergodic. Therefore, it is sufficient to prove that $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$ is a necessary and sufficient condition for $\{Q_n, n = 1, 2, ...\}$ to be ergodic.

Suppose that $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$. Note that $\{Q_n; , n = 1, 2, ...\}$ is an irreducible and aperiodic Markov chain. In order to show that $\{Q_n; n = 1, 2, ...\}$ is ergodic, it remains to prove that it is positive recurrent. We use Theorem 2 from Pakes (1969), which states that an irreducible and aperiodic Markov chain $\{Q_n; n = 1, 2, ...\}$ is positive recurrent if $|\chi_i| < \infty$ for all i and $\lim_{i\to\infty} \sup \chi_i < 0$, where the mean drift $\chi_i \equiv E(Q_{n+1} - Q_n | Q_n = i)$. Thus, we have

$$\chi_0 = \lambda ES^{(n)} = \lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)),$$

and for $i > 0$,

$$\begin{aligned}\chi_i &= (1 - L_A(\lambda))\lambda ES^{(n)} + L_A(\lambda)(\lambda ES^{(n)} - 1) \\ &= \lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) - L_A(\lambda).\end{aligned}$$

Since $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$, then it is clear that $\lim_{i\to\infty} \sup \chi_i < 0$ so $\{Q_n, n = 1, 2, ...\}$ is positive recurrent and therefore is ergodic.

Next assume

$$\{Q_n, n = 0, 1, ...\} \text{ is ergodic.} \tag{2.7}$$

We will show that $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$.

To do this, we look at the contrapositive.

Suppose $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) \geq L_A(\lambda)$.

Define the mean down drift to be $\delta_i = \sum_{j<i} P(Q_{n+1} = j \mid Q_n = i)$.

But

$$P(Q_{n+1} = j \mid Q_n = i) = 0 \text{ for } j < i \text{ unless } j = i-1, \text{ for } i > 0.$$

Thus, by (2.5),

$$\delta_i = -P(Q_{n+1} = i-1 \mid Q_n = i) = -L_A(\lambda) \int_0^\infty e^{-\lambda x}\, dB^*(x).$$

Also, $\delta_0 = 0$.

We already showed that the mean drift $\chi_0 = \lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1))$ and for $i > 0$,

$$\chi_i = \lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) - L_A(\lambda) \tag{2.8}$$

Kaplan's condition, as presented in Sennott et al. (1983), states that (1) χ_i is finite for all i, (2) there exists N such that $\chi_i \geq 0$ for $i = N, N+1, ...$, (3) $\delta_i \geq -\epsilon$ for some $\epsilon > 0$. In Sennott et al. (1983), Theorem 1, we find that if Kaplan's condition holds, then the Markov chain $\{Q_n, n = 1, 2, ...\}$ is not ergodic. The contrapositive of this statement reads " If the Markov chain is ergodic, then Kaplan's condition does not hold."

By (2.7), $\{Q_n, n = 1, 2, ...\}$ is ergodic. Thus, Kaplan's condition does not hold. However if we check Kaplan's condition, we find that (1) and (3) clearly hold. Thus (2) must be false. So there does not exist N such that $\chi_i \geq 0$ for $n = N, N+1, ...$ Thus $\chi_i \leq 0$ for some $i = N+1, N+2, ...$ But χ_i is constant $\forall i = N+1, N+2, ...$ and by (2.8), $\chi_i < 0$, i.e., $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$. Thus, we have shown the necessity.

Remark 2.1 In this model, there exist three blocked states of the server, that is, the server is busy or under repair or reserved. It can be easily shown that the joint Laplace transform of W_1, W_2 and W_3 is given by

$$L(s_1, s_2, s_3) \equiv E(e^{-s_1W_1-s_2W_2-s_3W_3}) = L_B(s_1 + \mu - \frac{\mu\theta}{s_3+\theta}L_C(s_2)), \tag{2.9}$$

where W_1, W_2 and W_3 are, respectively, the service time, the total repair time and the total reserved time during a generalized service time. Also, their corresponding mean times are, respectively, β_1, $\beta_1\mu\gamma_1$ and $\frac{\beta_1\mu}{\theta}$.

Remark 2.2 It is well-known that for a classical $M/G/1$ queue, $\lambda\beta < 1$ is a necessary and sufficient condition for the system to be stable, where λ and β are, respectively, the average arrival rate and the expected value of the generalized service time. We may view $\lambda\beta$ as the average number of customers added to the queue during a generalized service time β. However if the queue is nonempty, the completion of a generalized service implies that the queue can be reduced by only one customer and then a service begins immediately. Therefore, for the stability of the system, $\lambda\beta < 1$ must be satisfied. On the other hand, for the $M/G/1$ retrial queue, the queue of interest has changed into the retrial queue. In this case, $\lambda\beta$ is the average number of customers added to the retrial queue during the mean service time β. Given that the retrial queue is nonempty, only when a repeated customer wins the competition between a primary customer and a retrial customer, can the retrial queue be decreased by one. Note that the success of a primary customer in the competition does not influence the average length of the retrial queue. Therefore, for stability of the retrial queueing system, $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$ must be satisfied. Here $L_A(\lambda)$ is just the probability that the retrial customer wins the competition over a primary customer.

Remark 2.3 The $M/G/1$ retrial queue with general retrial times and server subject to breakdowns and repairs and reserved customers may be viewed as an extension of some other retrial queues.

(i) If $\mu = 0$, our retrial queue reduces to the $M/G/1$ retrial queue with server without breakdowns.

(ii) If $\mu \neq 0$, $\theta_1 = \infty$, our retrial queue reduces to the classical $M/G/1$ retrial queue with repairable server but no recycles.

(iii) If the retrial time distribution is $A(x) = 1 - e^{-\alpha x}$ and if $\alpha \to \infty$, then the $M/G/1$ retrial queue becomes the ordinary $M/G/1$ queue.

Remark 2.4 In analyzing the stability of the $M/G/1$ retrial queue with general retrial times and server subject to breakdowns and repairs and recycles, we may view it as an ordinary $M/G/1$ queueing system. It can be seen that the average inter-arrival time is equal to $\frac{1}{\lambda}$, while the "service time" may be viewed as the generalized service time plus the arrival time if a primary customer wins the competition or plus the retrial time if a repeated customer wins the competition. That is, the "service time" is equal to the generalized service time plus the minimum of the arrival time and the retrial time.

The average "service time" is

$$\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) + \int_0^\infty dA(y) \int_0^y x\lambda e^{-\lambda x}\, dx$$

$$+\int_0^\infty y\,dA(y)\int_y^\infty \lambda e^{-\lambda x}\,dx$$
$$=\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))+\frac{1}{\lambda}(1-L_A(\lambda)). \tag{2.10}$$

Therefore, the system is stable if and only if

$$\frac{1}{\lambda}>\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))+\frac{1}{\lambda}(1-L_A(\lambda)), \tag{2.11}$$

or equivalently,

$$\lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))<L_A(\lambda). \tag{2.12}$$

Remark 2.5 Suppose the retrial queue has a large number of customers in the following discussion. Let $P(\text{block})$ be the probability that the system is blocked. We have an alternating renewal process between the blocked and idle states for the server. Let $E(S^*)$ be the expected blocked time. Let $E(T)$ be the expected idle time. Then

$$P(\text{block})=\frac{E(S^*)}{E(S^*)+E(T)},$$

and

$$P(\text{not block})=\frac{E(T)}{E(S^*)+E(T)}.$$

The arrival rate to the retrial queue is $\lambda P(\text{block})$. If the server becomes idle, then $E(T)$ is the expected idle time. The probability that the next customer in service comes from the retrial queue is $L_A(\lambda)$.

The entrance rate to the retrial queue is $\lambda P(\text{block})$.

The exit rate from the retrial queue is

$$\frac{L_A(\lambda)P(\text{not block})}{E(T)}.$$

For stability, we need the arrival rate to be less than the exit rate

i.e.,

$$\lambda P(\text{block}) < \frac{L_A(\lambda)P(\text{not block})}{E(T)}.$$

This is equivalent to

$$\lambda E(S^{(k)}) < L_A(\lambda),$$

i.e.

$$\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)) < L_A(\lambda).$$

Thus we have presented four different arguments to indicate why our stability condition is correct.

Chapter 3

Model I Steady State Distribution

3.1 A system of forward equations

In this chapter, we carry out a stochastic analysis of the queueing system of Model I. Our objective is to obtain the joint distribution of the state of the server and the length of the retrial queue in steady state. In order of do so, we first define the state probabilities, state probability densities and joint state probability densities as follows (some of the following notation came from section 2.3).

(i) $P_{(0,0)}(t) \equiv P(J(t) = 0,\, Q(t) = 0),\ 0 \le t < \infty,$

(ii) $P_{(0,i)}(t,\, w)\, dw \equiv P(J(t) = 0,\, Q(t) = i,\, w < \xi_0(t) < w + dw)$, for $0 \le t, w < \infty,\ i = 1, 2, ...,$

(iii) $P_{(1,i)}(t,\, x)\, dx \equiv P(J(t) = 1,\, Q(t) = i,\, x < \xi_1(t) < x + dx)$, for $0 \le t,\, x < \infty,\ i = 0, 1, ...,$

(iv) $P_{(2,i)}(t,\, x,\, y)\, dx\, dy \equiv P(J(t) = 2,\, Q(t) = i,\, x < \xi_1(t) < x + dx,\, y < \xi_2(t) < y + dy)$, for $0 \le t,\, x,\, y < \infty\, i = 0, 1, ...,$

(v) $P_{(3,i)}(t, x, \tau)\, dx\, d\tau \equiv P(J(t) = 3, Q(t) = i, x < \xi_1(t) < x + dx, \tau < \xi_3(t) < \tau + d\tau)$, for $0 \leq t, x, \tau < \infty, i = 0, 1, \ldots$.

To find $P_{(0,0)}(t)$, $P_{(0,i)}(t, w)$, $P_{(1,i)}(t, x)$, $P_{(2,i)}(t, x, y)$ and $P_{(3,i)}(t, x, \tau)$, we must know how the system moves from one state at time t to other states at time $t + \Delta t$. The investigation is carried out as follows.

(i) The event $(J(t + \Delta t) = 0, Q(t + \Delta t) = 0)$ occurs either when the event $(J(t) = 0, Q(t) = 0)$ occurs and there are no primary customers arriving during Δt, or when $(J(t) = 1, Q(t) = 1, \xi_1(t) = x)$ occurs and the service is completed during Δt. This leads to the difference equation:

$$P_{(0,0)}(t + \Delta t) = (1 - \lambda \Delta t)P_{(0,0)}(t) + \Delta t \int_0^\infty P_{(1,0)}(t, x)\beta(x)\, dx + o(\Delta t). \quad (3.1)$$

(ii) The event $(J(t + \Delta t) = 0, Q(t + \Delta t) = i, \xi_0(t + \Delta t) = w + \Delta t)$ occurs only when the event $(J(t) = 0, Q(t) = i, \xi_0(t) = w)$ occurs and no primary customers or repeated customers arrive during Δt. Hence, for $i = 1, 2, \ldots, 0 < w < \infty$,

$$P_{(0,i)}(t + \Delta t, w + \Delta t) = (1 - (\lambda + \alpha(w)) \Delta t)P_{(0,i)}(t, w) + o(\Delta t). \quad (3.2)$$

(iii) The event $(J(t+\Delta t) = 1, Q(t+\Delta t) = i, \xi_1(t+\Delta t) = x+\Delta t)$ occurs either when the event $(J(t) = 1, Q(t) = i, \xi_1(t) = x)$ occurs and it is impossible that primary customers arrive or the service is completed or the server is under repair during Δt, or when the event $(J(t) = 1, Q(t) = i - 1, \xi_1(t) = x)$ occurs and a primary customer arrives during Δt, or when the event $(J(t) = 3, Q(t) = i, \xi_1(t) = x, \xi_3(t) = \tau)$ occurs and the reserved customer begins service again during Δt. Thus, for $i = 0, 1, \ldots, 0 < w < \infty, 0 < x < \infty$,

$$P_{(1,i)}(t + \Delta t, x + \Delta t) = (1 - (\lambda + \mu + \beta(x)) \Delta t)P_{(1,i)}(t, x)$$

$$+\lambda \bigtriangleup t P_{(1,i-1)}(t,\,x) + \theta \bigtriangleup t \int_0^\infty P_{(3,i)}(t,\,x,\,\tau)\,d\tau + o(\Delta t), \tag{3.3}$$

where $P_{(1,-1)}(t,\,x) \equiv 0$.

(iv) The event $(J(t+\Delta t)=2,\, Q(t+\Delta t)=i,\, \xi_1(t+\Delta t)=x,\, \xi_2(t+\Delta t)=y+\Delta t)$ occurs either when the event $(J(t)=2,\, Q(t)=i,\, \xi_1(t)=x,\, \xi_2(t)=y)$ occurs and there are no primary customers arriving and the server is still under repair during Δt, or when the event $(J(t)=2,\, Q(t)=i-1,\, \xi_1(t)=x,\, \xi_2(t)=y)$ occurs and a primary customer arrives during Δt. Then, for $i=0,1,...,\ 0<x<\infty,\ 0<y<\infty$,

$$P_{(2,i)}(t+\Delta t,\,x,\,y+\Delta t) = (1-(\lambda+\gamma(y)) \bigtriangleup t)P_{(2,i)}(t,\,x,\,y)$$

$$+\lambda \bigtriangleup t P_{(2,i-1)}(t,\,x,\,y) + o(\Delta t), \tag{3.4}$$

where $P_{(2,-1)}(t,\,x,\,y) \equiv 0$.

(v) The event $(J(t+\Delta t)=3,\, Q(t+\Delta t)=i,\, \xi_1(t+\Delta t)=x,\, \xi_3(t+\Delta t)=\tau+\Delta t)$ occurs either when the event $(J(t)=3,\, Q(t)=i,\, \xi_1(t)=x,\, \xi_3(t)=\tau)$ occurs and no primary customers or reserved customers arrive during Δt, or when the event $(J(t)=3,\, Q(t)=i-1,\, \xi_1(t)=x,\, \xi_3(t)=\tau)$ occurs and a primary customer arrives during Δt. Therefore, for $i=0,1,...,\ 0<x<\infty,\ 0<\tau<\infty$,

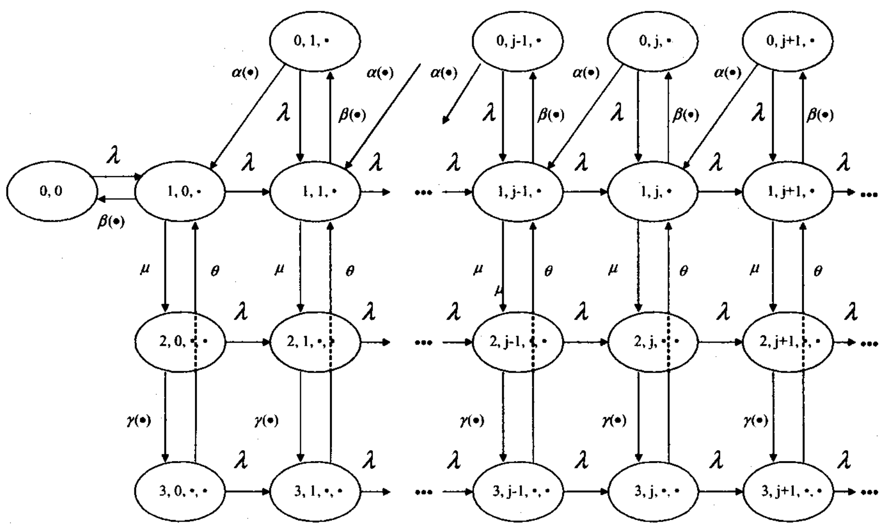
$$P_{(3,i)}(t+\Delta t,\,x,\,\tau+\Delta t) = (1-(\lambda+\theta) \bigtriangleup t)P_{(3,i)}(t,\,x,\,y)$$

$$+\lambda \bigtriangleup t P_{(3,i-1)}(t,\,x,\,\tau) + o(\Delta t), \tag{3.5}$$

where $P_{(3,-1)}(t,\,x,\,\tau) \equiv 0$.

The rate transition diagram for the model is illustrated in Figure 3.1.

Figure 3.1: Rate transition diagram-M/G/1 retrial queue with general retrial times and server subject to breakdowns, repairs and reserved server.



Similarly, we have following boundary conditions for $P_{(0,\,i)}(t,\,w)$, $P_{(1,\,i)}(t,\,x)$, $P_{(2,\,i)}(t,\,x,\,y)$ and $P_{(3,\,i)}(t,\,x,\,\tau)$ as follows.

$$P_{(0,\,i)}(t,\,0) = \int_0^\infty P_{(1,\,i)}(t,\,x)\beta(x)\,dx, \tag{3.6}$$

$$P_{(1,\,i)}(t,\,0) = \lambda\delta_{i0}P_{(0,\,0)}(t) + \lambda(1-\delta_{i0})\int_0^\infty P_{(0,\,i)}(t,\,w)\,dw$$
$$+\int_0^\infty P_{(0,\,i+1)}(t,\,w)\alpha(w)\,dw, \tag{3.7}$$

$$P_{(2,\,i)}(t,\,x,\,0) = \mu P_{(1,\,i)}(t,\,x), \tag{3.8}$$

$$P_{(3,\,i)}(t,\,x,\,0) = \int_0^\infty P_{(2,\,i)}(t,\,x,\,y)\gamma(y)\,dy. \tag{3.9}$$

where δ_{i0} is the Kronecker function.

Also, the normalization equation is given by

$$P_{(0,0)}(t) + \sum_{i=1}^{\infty} \int_0^{\infty} P_{(0,i)}(t, w)\, dw + \sum_{i=0}^{\infty} (\int_0^{\infty} P_{(1,i)}(t, x)\, dx$$
$$+ \int_0^{\infty} \int_0^{\infty} P_{(2,i)}(t, x, y)\, dx\, dy + \int_0^{\infty} \int_0^{\infty} P_{(3,i)}(t, x, \tau)\, dx\, d\tau) = 1. \quad (3.10)$$

From (3.1)-(3.5), the corresponding differential-difference equations are found by subtracting $P_{(0,0)}(t)$, $P_{(0,i)}(t, w)$, $P_{(1,i)}(t, x)$, $P_{(2,i)}(t, x, y)$ or $P_{(3,i)}(t, x, \tau)$ from both sides, dividing through by $\triangle t$ and taking the limit as $\triangle t \to 0$. Thus we get

$$(\frac{d}{dt} + \lambda) P_{(0,0)}(t) = \int_0^{\infty} P_{(1,0)}(t, x) \beta(x)\, dx, \quad (3.11)$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial w} + \lambda + \alpha(w)) P_{(0,i)}(t, w) = 0, \quad (3.12)$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial x} + \lambda + \mu + \beta(x)) P_{(1,i)}(t, x) = \lambda P_{(1,i-1)}(t, x)$$
$$+\theta \int_0^{\infty} P_{(3,i)}(t, x, \tau)\, d\tau, \quad (3.13)$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial y} + \lambda + \gamma(y)) P_{(2,i)}(t, x, y) = \lambda P_{(2,i-1)}(t, x, y), \quad (3.14)$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial \tau} + \lambda + \theta) P_{(3,i)}(t, x, \tau) = \lambda P_{(3,i-1)}(t, x, \tau). \quad (3.15)$$

3.2 The steady state distribution

It has been shown that $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$ is a necessary and sufficient condition for the system to be stable. Thus, if $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$, from the discussion above, we can obtain the joint distribution of the server state and the queue length, in steady state. Also, some main performance measures in steady state, such as: P(the system is empty), P(the system is nonempty and the server is idle), P(the server is busy), P(the server is under repair) and P(the server is reserved), can be explicitly given.

Theorem 3.1 For the retrial queue under consideration, if $\lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1)) < L_A(\lambda)$, then in steady state there exists the joint distribution of the state of the server and the length of the retrial queue in terms of the probability generating functions given by

$$P_{(0,0)} = \frac{L_A(\lambda) - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))}{L_A(\lambda)}, \tag{3.16}$$

$$P_0(z,\, w) = \frac{\lambda(L_A(\lambda) - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))z(1-L_B(K(\lambda(1-z))))}{L_A(\lambda)(L_A(\lambda)(1-z)L_B(K(\lambda(1-z))) - z(1-L_B(K(\lambda(1-z)))))}$$
$$\times \exp(-\lambda w)(1-A(w)), \tag{3.17}$$

$$P_1(z,\, x) = \frac{\lambda(L_A(\lambda) - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))(1-z)}{L_A(\lambda)(1-z)L_B(K(\lambda(1-z))) - z(1-L_B(K(\lambda(1-z)))))}$$
$$\times \exp(-K(\lambda(1-z))x)(1-B(x)), \tag{3.18}$$

$$P_2(z,\, x,\, y) = \frac{\lambda\mu(L_A(\lambda) - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))(1-z)}{L_A(\lambda)(1-z)L_B(K(\lambda(1-z))) - z(1-L_B(K(\lambda(1-z)))))}$$
$$\times \exp(-K(\lambda(1-z))x - \lambda(1-z)y)(1-B(x))(1-C(y)), \tag{3.19}$$

$$P_3(z,\, x,\, \tau) = \frac{\lambda\mu(L_A(\lambda - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))(1-z)L_C(\lambda(1-z))}{L_A(\lambda)(1-z)L_B(K(\lambda(1-z))) - z(1-L_B(K(\lambda(1-z)))))}$$
$$\times \exp(-K(\lambda(1-z))x - (\theta+\lambda(1-z))\tau)(1-B(x)), \tag{3.20}$$

where

$$K(x) \equiv x + \mu - \frac{\mu\theta}{x+\theta}L_C(x). \tag{3.21}$$

Proof. If $\lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1)) < L_A(\lambda)$, then $\lim_{t\to\infty} P_{(0,0)}(t)$, $\lim_{t\to\infty} P_{(0,i)}(t,\, w)$, $\lim_{t\to\infty} P_{(1,i)}(t,\, x)$, $\lim_{t\to\infty} P_{(2,i)}(t,\, x,\, y)$ and $\lim_{t\to\infty} P_{(3,i)}(t, x, \tau)$ exist and are denoted by $P_{(0,0)}$, $P_{(0,i)}(w)$, $P_{(1,i)}(x)$, $P_{(2,i)}(x,\, y)$ and $P_{(3,i)}(x,\, \tau)$, respectively.

Hence, from the equations (3.6)-(3.15) we can derive that

$$\lambda P_{(0,\,0)} = \int_0^\infty P_{(1,\,0)}(x)\beta(x)\,dx, \tag{3.22}$$

$$(\frac{d}{dw}+\lambda+\alpha(w))P_{(0,\,i)}(w) = 0, \tag{3.23}$$

$$(\frac{d}{dx}+\lambda+\mu+\beta(x))P_{(1,\,i)}(x) = \lambda P_{(1,i-1)}(x)+\theta\int_0^\infty P_{(3,\,i)}(x,\,\tau)\,d\tau, \tag{3.24}$$

$$(\frac{\partial}{\partial y}+\lambda+\gamma(y))P_{(2,\,i)}(x,\,y) = \lambda P_{(2,i-1)}(x,\,y), \tag{3.25}$$

$$(\frac{\partial}{\partial \tau}+\lambda+\theta)P_{(3,\,i)}(x,\,\tau) = \lambda P_{(3,\,i-1)}(x,\,\tau), \tag{3.26}$$

$$P_{(0,\,i)}(0) = \int_0^\infty P_{(1,\,i)}(x)\beta(x)\,dx, \tag{3.27}$$

$$\begin{aligned} P_{(1,\,i)}(0) &= \lambda\delta_{i0}P_{(0,\,0)} + \lambda(1-\delta_{i0})\int_0^\infty P_{(0,\,i)}(w)\,dw \\ &+ \int_0^\infty P_{(0,\,i+1)}(w)\alpha(w)\,dw, \end{aligned} \tag{3.28}$$

$$P_{(2,\,i)}(x,\,0) = \mu P_{(1,\,i)}(x), \tag{3.29}$$

$$P_{(3,\,i)}(x,\,0) = \int_0^\infty P_{(2,\,i)}(x,\,y)\gamma(y)\,dy, \tag{3.30}$$

$$\begin{aligned} &P_{(0,\,0)} + \sum_{i=1}^\infty \int_0^\infty P_{(0,\,i)}(w)\,dw + \sum_{i=0}^\infty(\int_0^\infty P_{(1,\,i)}(x)\,dx \\ &+ \int_0^\infty\int_0^\infty P_{(2,\,i)}(x,\,y)\,dx\,dy + \int_0^\infty\int_0^\infty P_{(3,\,i)}(x,\,\tau)\,dx\,d\tau) = 1. \end{aligned} \tag{3.31}$$

In order to solve the system of equations (3.22)-(3.31), we introduce the following probability generating functions:

$$P_0(z,\,w)=\sum_{i=1}^{\infty}P_{(0,\,i)}(w)z^i,\quad P_1(z,\,x)=\sum_{i=0}^{\infty}P_{(1,\,i)}(x)z^i,$$

$$P_2(z,\,x,\,y)=\sum_{i=0}^{\infty}P_{(2,\,i)}(x,\,y)z^i,\quad P_3(z,\,x,\,\tau)=\sum_{i=0}^{\infty}P_{(3,\,i)}(x,\,\tau)z^i.$$

Thus, when both sides of the equations (3.23)–(3.31) are multiplied by z^i and summed over i, we obtain the following equations:

$$(\frac{\partial}{\partial w}+\lambda+\alpha(w))P_0(z,\,w)=0, \tag{3.32}$$

$$(\frac{\partial}{\partial x}+\lambda+\mu+\beta(x))P_1(z,\,x)=\lambda zP_1(z,\,x)+\theta\int_0^{\infty}P_3(z,\,x,\,\tau)\,d\tau, \tag{3.33}$$

$$(\frac{\partial}{\partial y}+\lambda+\gamma(y))P_2(z,\,x,\,y)=\lambda zP_2(z,\,x,\,y), \tag{3.34}$$

$$(\frac{\partial}{\partial \tau}+\lambda+\theta)P_3(z,\,x,\,\tau)=\lambda zP_3(z,\,x,\,\tau), \tag{3.35}$$

$$P_0(z,\,0)=-\lambda P_{(0,\,0)}+\int_0^{\infty}P_1(z,\,x)\beta(x)\,dx, \tag{3.36}$$

$$P_1(z,\,0)=\lambda P_{(0,0)}+\lambda\int_0^{\infty}P_0(z,\,w)\,dw+\frac{1}{z}\int_0^{\infty}P_0(z,\,w)\alpha(w)\,dw, \tag{3.37}$$

$$P_2(z,\,x,\,0)=\mu P_1(z,\,x), \tag{3.38}$$

$$P_3(z,\,x,\,0)=\int_0^{\infty}P_2(z,\,x,\,y)\gamma(y)\,dy, \tag{3.39}$$

$$\begin{aligned}&P_{(0,\,0)}+\lim_{z\to 1-}(\int_0^{\infty}P_0(z,\,w)\,dw+\int_0^{\infty}P_1(z,\,x)\,dx\\&+\int_0^{\infty}\int_0^{\infty}P_2(z,\,x,\,y)\,dx\,dy+\int_0^{\infty}\int_0^{\infty}P_3(z,\,x,\,\tau)\,dx\,d\tau)=1.\end{aligned} \tag{3.40}$$

These equations are solved as follows.

It is clear that the solutions to (3.32), (3.34) and (3.35) are given by

$$P_0(z,\,w)=P_0(z,\,0)e^{-\lambda w}(1-A(w)), \tag{3.41}$$

$$P_2(z,\, x,\, y) = P_2(z,\, x,\, 0)e^{-\lambda(1-z)y}(1 - C(y)), \tag{3.42}$$

$$P_3(z,\, x,\, \tau) = P_3(z,\, x,\, 0)e^{-(\lambda(1-z)+\theta)\tau}. \tag{3.43}$$

The substitution of (3.41) into (3.37) yields

$$P_1(z,\, 0) = \lambda P_{(0,\,0)} + \frac{1}{z}P_0(z,\, 0)(L_A(\lambda) + (1 - L_A(\lambda))z). \tag{3.44}$$

However, with the help of (3.38) and (3.39), (3.42) and (3.43) can be rewritten as

$$P_2(z,\, x,\, y) = P_1(z,\, x)\mu e^{-\lambda(1-z)y}(1 - C(y)), \tag{3.45}$$

$$P_3(z,\, x,\, \tau) = P_1(z,\, x)\mu L_C(\lambda(1 - z))e^{-(\lambda(1-z)+\theta)\tau}. \tag{3.46}$$

Substituting (3.46) into (3.33), we get

$$P_1(z,\, x) = P_1(z,\, 0)e^{-K(\lambda(1-z))x}(1 - B(x)), \tag{3.47}$$

which is used in (3.36) to yield

$$P_0(z,\, 0) = -\lambda P_{(0,\,0)} + L_B(K(\lambda(1 - z)))P_1(z,\, 0). \tag{3.48}$$

Thus, from (3.44) and (3.48), we have

$$P_0(z,\, 0) = \frac{\lambda z(1 - L_B(K(\lambda(1 - z))))P_{(0,\,0)}}{L_A(\lambda)(1 - z)L_B(K(\lambda(1 - z))) - z(1 - L_B(K(\lambda(1 - z))))}, \tag{3.49}$$

$$P_1(z,\, 0) = \frac{\lambda L_A(\lambda)(1 - z)P_{(0,\,0)}}{L_A(\lambda)(1 - z)L_B(K(\lambda(1 - z))) - z(1 - L_B(K(\lambda(1 - z))))}, \tag{3.50}$$

which implies that $P_0(z,\, w)$, $P_1(z,\, x)$, $P_2(z,\, x,\, y)$ and $P_3(z,\, x,\, \tau)$ depend only on $P_{(0,\,0)}$.

However, $P_{(0,0)}$ can be found by using the normalization condition

$$\begin{aligned} &P_{(0,0)} + \int_0^\infty P_0(1,\, w)\, dw + \int_0^\infty P_1(1,\, x)\, dx \\ &+ \int_0^\infty \int_0^\infty P_2(1,\, x,\, y)\, dx\, dy + \int_0^\infty \int_0^\infty P_3(1,\, x,\, \tau)\, dx\, d\tau = 1. \end{aligned} \tag{3.51}$$

This completes the proof.

3.3 Performance Measures

In steady state, let J represent the state of the server and Q represent the number of customers in the retrial queue, namely, $J = \lim_{t\to\infty} J(t)$ and $Q = \lim_{t\to\infty} Q(t)$. The steady state solution of the model allows us to calculate some important performance measures. Of the performance measures, the probabilities that the system is in different states are very important. All of the following definitions in this paragraph refer to probabilities in steady state. Let $E \equiv P$(the system is empty), $I \equiv P$ (the system is nonempty and the server is idle), $B \equiv P$(the server is busy), $F \equiv P$(the server is under repair), $R \equiv P$(the server is reserved). We obtain the following Corollary.

Corollary 3.1 If $\lambda\beta_1(1 + \mu(\frac{1}{\theta} + \gamma_1)) < L_A(\lambda)$, then

(i) $E = P_{(0,0)} = \frac{L_A(\lambda) - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))}{L_A(\lambda)}$,

(ii) $I = \frac{(1-L_A(\lambda)\lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))}{L_A(\lambda)}$,

(iii) $B = \lambda\beta_1$,

(iv) $F = \lambda\beta_1\mu\gamma_1$,

(v) $R = \frac{\lambda\beta_1\mu}{\theta}$.

Proof. It is evident that $E = P(J=0,\, Q=0) = P_{(0,0)}$. Also, we can write

$$
\begin{aligned}
I &= P(J=0,\, Q=1,2,\ldots)\\
&= \sum_{i=1}^{\infty}\int_0^{\infty} P_{(0,i)}(w)\,dw\\
&= \lim_{z\to 1-}\int_0^{\infty} P_0(z,\, w)\,dw\\
&= \lim_{z\to 1-}\frac{(1-L_A(\lambda))(L_A(\lambda)-\lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))\lambda z(1-L_*(z))}{\lambda L_A(\lambda)(L_A(\lambda)(1-z)L_*(z)-z(1-L_*(z))}, \qquad (3.52)
\end{aligned}
$$

where $L_*(z) \equiv L_B(K(\lambda(1-z)))$,

$$
\begin{aligned}
B &= P(J=1,\, Q=0,1,\ldots)\\
&= \sum_{i=0}^{\infty}\int_0^{\infty} P_{(1,i)}(x)\,dx\\
&= \lim_{z\to 1-}\int_0^{\infty} P_1(z,\, x)\,dx\\
&= \lim_{z\to 1-}\frac{\lambda(L_A(\lambda)-\lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))(1-z)(1-L_*(z))}{(L_A(\lambda)(1-z)L_*(z)-z(1-L_*(z))K(\lambda(1-z))}, \qquad (3.53)
\end{aligned}
$$

$$
\begin{aligned}
F &= P(J=2,\, Q=0,1,\ldots)\\
&= \sum_{i=0}^{\infty}\int_0^{\infty}\int_0^{\infty} P_{(2,i)}(x\,y)\,dx\,dy\\
&= \lim_{z\to 1-}\int_0^{\infty} P_2(z,\, x\,y)\,dx\,dy\\
&= \lim_{z\to 1-}\frac{\lambda\mu(L_A(\lambda)-\lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))(1-z)}{L_A(\lambda)(1-z)L_*(z)-z(1-L_*(z))}\\
&\times\frac{(1-L_*(z))(1-L_C((\lambda(1-z))}{\lambda(1-z)K(\lambda(1-z))}, \qquad (3.54)
\end{aligned}
$$

$$
\begin{aligned}
R &= P(J=3,\, Q=0,1,\ldots)\\
&= \sum_{i=0}^{\infty}\int_0^{\infty}\int_0^{\infty} P_{(3,i)}(x\,\tau)\,dx\,d\tau\\
&= \lim_{z\to 1-}\int_0^{\infty} P_3(z,\, x\,\tau)\,dx\,d\tau
\end{aligned}
$$

$$= \lim_{z\to 1-} \frac{\lambda\mu(L_A(\lambda - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1))(1-z)L_C(\lambda(1-z))}{L_A(\lambda)(1-z)L_*(z) - z(1-L_*(z))}$$
$$\times \frac{1-L_*(z))}{(\theta+\lambda(1-z))K(\lambda(1-z))}. \tag{3.55}$$

Substituting the steady state solution of Theorem 3.1 into (3.52)-(3.55), we obtain the results of Corollary 3.1.

Now

$$\frac{d}{dz}(L_B(K(\lambda(1-z))) \mid_{z=1} = \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1)). \tag{3.56}$$

Therefore, we employ L'Hôpital's rule in (3.52)-(3.55) and then obtain these main performance measures. This completes the proof.

Remark 3.1 Corollary 3.1 gives us many standard measures of performance. We immediately also have that proportion of time that the server is in idle mode is $E+I$. In steady state, in addition to the probabilities of being in different states of the system, it is desirable to obtain the distribution of the number of customers in the retrial queue and the distribution of the number of customers in the system, from which the expected number of customers in the retrial queue and the expected value of the number of customers in the system are easily derived. Therefore we next turn our attention to the derivation of the distribution of these two random variables and their expected values.

Corollary 3.2 In steady state, let N_q represent the number of customers in the retrial queue and N represent the number of customers in the system. Then the probability generating functions of N_q and N are respectively given by

$$p_q(z) = \frac{(L_A(\lambda) - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1)))(1-z)}{L_A(\lambda)(1-z)L_B(K(\lambda(1-z))) - z(1-L_B(K(\lambda(1-z))))}, \tag{3.57}$$

$$p(z) = \frac{(L_A(\lambda - \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1)))(1-z)L_B(K(\lambda(1-z)))}{L_A(\lambda)(1-z)L_B(K(\lambda(1-z))) - z(1-L_B(K(\lambda(1-z))))}. \tag{3.58}$$

The expected values of N_q and N are respectively given by

$$L_q = \frac{\lambda^2(\theta\mu\gamma_2 - 2)\beta_1^3 + 2\rho_*(\lambda+\theta-L_A(\lambda)\theta)\beta_1^2 + \rho_*^2\theta\beta_2}{2\theta(L_A(\lambda)-\rho_*)\beta_1^2}, \tag{3.59}$$

$$L = \frac{\lambda^2(\theta\mu\gamma_2 - 2)\beta_1^3 + 2\rho_*(\lambda+\theta-\rho_*\theta)\beta_1^2 + \rho_*^2\theta\beta_2}{2\theta(L_A(\lambda)-\rho_*)\beta_1^2}, \tag{3.60}$$

where

$$\rho_* \equiv \lambda\beta_1(1+\mu(\frac{1}{\theta}+\gamma_1)). \tag{3.61}$$

Proof. It can be easily seen that

$$\begin{aligned}
p_q(z) &= E(z^{N_q}) \\
&= P_{(0,0)} + \sum_{i=1}^{\infty} z^i P(J=0,\, Q=i) + \sum_{i=1}^{\infty} z^i P(J=0,\, Q=i) \\
&+ \sum_{i=1}^{\infty} z^i P(J=0,\, Q=i) + \sum_{i=1}^{\infty} z^i P(J=0,\, Q=i) \\
&= P_{(0,0)} + \sum_{i=1}^{\infty}\int_0^{\infty} P_{(0,i)}(w)\,dw + \sum_{i=0}^{\infty}\int_0^{\infty} P_{(1,i)}(x)\,dx \\
&+ \sum_{i=0}^{\infty}\int_0^{\infty}\int_0^{\infty} P_{(2,i)}(x,\,y)\,dx\,dy + \sum_{i=0}^{\infty}\int_0^{\infty}\int_0^{\infty} P_{(3,i)}(x,\,\tau)\,dx\,d\tau \\
&= P_{(0,0)} + \int_0^{\infty} P_0(z,\,w)\,dw + \int_0^{\infty} P_1(z,\,x)\,dx \\
&+ \int_0^{\infty} P_2(z,\,x,\,y)\,dx\,dy + \int_0^{\infty} P_3(z,\,x,\,\tau)\,dx\,d\tau,
\end{aligned} \tag{3.62}$$

$$p(z) = Ez^{N_q} = P_{(0,0)} + \int_0^{\infty} P_0(z,\,w)\,dw + z(\int_0^{\infty} P_1(z,\,x)\,dx$$

$$+\int_0^\infty P_2(z,\, x\, y)\, dx\, dy + \int_0^\infty P_3(z,\, x,\, \tau)\, dx\, d\tau), \tag{3.63}$$

which yields the expressions of (3.57) and (3.58) by substituting the steady state solution of Theorem 3.1 into (3.62) and (3.63).

Since

$$L_q = \lim_{z \to 1-} p_q'(z), \tag{3.64}$$

and

$$L = \lim_{z \to 1^-} p'(z), \tag{3.65}$$

these expected values are easily obtained.

Remark 3.2 The mean number of repairs during a service period is $\mu\beta_1$. Also, for this retrial queue, we have a relationship between $p_q(z)$ and $p(z)$ and between L_q and L:

$$p(z) = L_B(K(\lambda(1-z)))p_q(z), \tag{3.66}$$

and

$$L = L_q + E + I. \tag{3.67}$$

This follows from (3.57) and (3.58).

3.4 The waiting process

In this section we proceed to study the steady-state distribution of the waiting time that an arriving primary customer spends in the retrial queue and the distribution of the waiting time that an arriving primary customer spends in the system. According to the characteristics of the retrial queue, it can be seen that the waiting process for the retrial queue is more complicated than the waiting process for an ordinary queueing system. The waiting time that an arriving primary customer spends in the system for the retrial model is equal to the waiting time an arriving primary customer spends in the retrial queue plus its generalized service time, while the waiting time that an arriving primary customer spends in the system for an ordinary queueing system is equal to the waiting time that an arriving primary customer spends in the queue plus its regular service time.

Theorem 3.2 Let W and W_q denote, respectively, the time that an arriving primary customer spends in the system and in the retrial queue, in steady state. If the system is stable, then the Laplace transforms of W and W_q are respectively given by

$$L_W(s) = (1-\rho_*)L_B(K(s)) + \frac{\lambda(L_A(\lambda)-\rho_*)}{\lambda - s - \lambda\pi(s)} \times \frac{\pi(s)(1-\pi(s))(L_B(K(s)) - L_B(K(\lambda(1-\pi(s)))))}{L_A(\lambda)(1-\pi(s))L_B(K(\lambda(1-\pi(s))) - \pi(s)(1-L_B(K(\lambda(1-\pi(s)))))}, \tag{3.68}$$

$$L_{W_q}(s) = 1-\rho_* + \frac{\lambda(L_A(\lambda)-\rho_*)}{(\lambda - s - \lambda\pi(s))L_B(K(s))} \times \frac{\pi(s)(1-\pi(s))(L_B(K(s)) - L_B(K(\lambda(1-\pi(s)))))}{L_A(\lambda)(1-\pi(s))L_B(K(\lambda(1-\pi(s))) - \pi(s)(1-L_B(K(\lambda(1-\pi(s)))))}. \tag{3.69}$$

Their corresponding expected values are obtained as

$$EW = \frac{\lambda^2(\theta\mu\gamma_2 - 2)\beta_1^3 + 2\rho_*(\lambda+\theta-\rho_*\theta)\beta_1^2 + \rho_*^2\theta\beta_2}{2\lambda\theta(L_A(\lambda)-\rho_*)\beta_1^2}, \tag{3.70}$$

$$EW_q = \frac{\lambda^2(\theta\mu\gamma_2 - 2)\beta_1^3 + 2\rho_*(\lambda + \theta - L_A(\lambda)\theta)\beta_1^2 + \rho_*^2\theta\beta_2}{2\lambda\theta(L_A(\lambda) - \rho_*)\beta_1^2}, \tag{3.71}$$

where

$$\pi(s) \equiv \frac{(s+\lambda)L_A(s+\lambda)L_B(K(s))}{s+\lambda-\lambda(1-L_A(s+\lambda))L_B(K(s))}. \tag{3.72}$$

Proof. Note that the system is empty or the server is idle, busy, under repair or reserved, when a primary customer arrives. Thus, we have

$$\begin{aligned} L_W(s) &= P_{(0,0)}E(e^{-sW}|J=0,\, Q=0) \\ &+\sum_{i=1}^{\infty}\int_0^{\infty} P_{(0,i)}(w)E(e^{-sW}\,|\,J=0,\, Q=i,\, \xi_0^{(e)}=w)\,dw \\ &+\sum_{i=0}^{\infty}(\int_0^{\infty} P_{(1,i)}(x)E(e^{-sW}\,|\,J=1,\, Q=i,\, \xi_1^{(e)}=x)\,dx \\ &+\int_0^{\infty}\int_0^{\infty} P_{(2,i)}(x,\, y)E(e^{-sW}\,|\,J=2,\, Q=i,\, \xi_1^{(e)}=x,\, \xi_2^{(e)}=y)\,dx\,dy \\ &+\int_0^{\infty}\int_0^{\infty} P_{(3,i)}(x,\, \tau)E(e^{-sW}|J=3,\, Q=i,\, \xi_1^{(e)}=x,\, \xi_3^{(e)}=\tau)\,dx\,d\tau), \end{aligned} \tag{3.73}$$

where J, Q, $\xi_0^{(e)}$, $\xi_1^{(e)}$, $\xi_2^{(e)}$ and $\xi_3^{(e)}$ denote, respectively, the server state, the number of customers in the retrial queue, the elapsed retrial time, the elapsed service time, the elapsed repair time and the elapsed reserved time when the primary customer arrives.

It can be seen that W coincides with a generalized service time if the system is empty or the server is idle when the primary customer arrives. In this case, $E(e^{-sW}\,|\,J=0, Q=0) = E(e^{-sW}\,|\,J=0, Q=i, \xi_0^{(e)}=w) = L_B(K(s))$. Let $W^{(*)}$ represent the waiting time that the customer being served spends in the system from the instant the primary customer arrives and $W^{(i+1)}$ represent the total waiting time the $(i+1)$ customers in the retrial queue spend in the system from the instant the customer originally being served leaves the system. Then W is equal to $W^{(*)}+W^{(i+1)}$ if i customers are already in the retrial queue and the server is busy, under repair or reserved when the primary customer arrives. Since $W^{(*)}$ and $W^{(i+1)}$ are independent, we obtain

$$E(e^{-sW}|J=1,\, Q=i,\, \xi_1^{(e)}=x) = E(e^{-sW^{(i+1)}})$$
$$\times E(e^{-sW^{(*)}}|\, J=1, Q=i, \xi_1^{(e)}=x), \tag{3.74}$$

$$E(e^{-sW}|J=2,\, Q=i,\, \xi_1^{(e)}=x,\, \xi_2^{(e)}=y) = E(e^{-sW^{(i+1)}})$$
$$\times E(e^{-sW^{(*)}}|J=2,\, Q=i,\, \xi_1^{(e)}=x,\, \xi_2^{(e)}=y), \tag{3.75}$$

$$E(e^{-sW}|J=3,\, Q=i,\, \xi_1^{(e)}=x,\, \xi_3^{(e)}=\tau) = E(e^{-sW^{(i+1)}})$$
$$\times E(e^{-sW^{(*)}}|J=3,\, Q=i,\, \xi_1^{(e)}=x,\, \xi_3^{(e)}=\tau). \tag{3.76}$$

In order to get $E(e^{-sW^{(*)}}|\, J=1,\, Q=i,\, \xi_1^{(e)}=x)\, E(e^{-sW^{(*)}}|J=2,\, Q=i,\, \xi_1^{(e)}= x,\, \xi_2^{(e)}=y)$ and $E(e^{-sW^{(*)}}|J=3,\, Q=i,\, \xi_1^{(e)}=x,\, \xi_3^{(e)}=\tau)$, we employ the formulas

$$P(y<\xi_1^{(r)}<y+dy \mid \xi_1>x) = \frac{b(x+y)\,dy}{1-B(x)}, \tag{3.77}$$

$$P(y<\xi_2^{(r)}<y+dy \mid \xi_2>x) = \frac{c(x+y)\,dy}{1-C(x)}, \tag{3.78}$$

$$P(y<\xi_3^{(r)}<y+dy \mid \xi_3>x) = \theta e^{-\theta y}\,dy. \tag{3.79}$$

where ξ_1, ξ_2 and ξ_3 represent the service time, repair time and reserved time, respectively, and $\xi_1^{(r)}$, $\xi_2^{(r)}$ and $\xi_3^{(r)}$ represent the remaining service time, remaining repair time and remaining reserved time at the instant the primary customer arrives. Thus, we have

$$E(e^{-sW^{(*)}}|J=1,\, Q=i,\, \xi_1=x)$$

$$= \sum_{n=0}^{\infty} \frac{1}{1-B(x)} \int_0^{\infty} b(x+u) e^{-su} \frac{(\mu u)^n}{n!} e^{-\mu u} (\frac{\theta}{s+\theta} L_C(s))^n \, du$$
$$= \frac{1}{1-B(x)} \int_x^{\infty} b(u) e^{-K(s)(u-x)} \, du, \tag{3.80}$$

$$E(e^{-sW^{(*)}} | J=2, \, Q=i, \, \xi_1 = x, \, \xi_2 = y)$$
$$= \sum_{n=0}^{\infty} \frac{1}{(1-B(x))(1-C(y))} \int_0^{\infty} \int_0^{\infty} b(x+u) c(y+v) e^{-s(u+v)}$$
$$\times \frac{(\mu u)^n}{n!} e^{-\mu u} \frac{\theta}{s+\theta} (\frac{\theta}{s+\theta} L_C(s))^n \, du \, dv$$
$$= \frac{\theta}{(s+\theta)(1-B(x))(1-C(y))} \int_x^{\infty} \int_y^{\infty} b(u) c(v)$$
$$\times e^{-K(s)(u-x)-s(v-y)} \, du \, dv, \tag{3.81}$$

$$E(e^{-sW^{(*)}} | J=3, \, Q=i, \, \xi_1 = x, \, \xi_3 = \tau)$$
$$= \sum_{n=0}^{\infty} \frac{1}{1-B(x)} \int_0^{\infty} \int_0^{\infty} b(x+u) \theta e^{-\theta v} e^{-s(u+v)} \frac{(\mu u)^n}{n!} e^{-\mu u}$$
$$\times \frac{\theta}{s+\theta} (\frac{\theta}{s+\theta} L_C(s))^n \, du \, dv$$
$$= \frac{\theta}{(s+\theta)(1-B(x))} \int_x^{\infty} b(u) e^{-K(s)(u-x)} \, du. \tag{3.82}$$

In order to obtain $L_W(s)$, it remains to find $E(e^{-sW^{(i+1)}})$. We first get $E(e^{-sW^{(1)}})$. After a service completion there exists a competition for service between a primary customer and a customer in the retrial queue. Therefore

$$E(e^{-sW^{(1)}}) = \sum_{n=0}^{\infty} L_B(K(s)) \int_0^{\infty} \int_y^{\infty} e^{-sy} \lambda e^{-\lambda x} \, dA(y) \, dx$$
$$\times (L_B(K(s)) \int_0^{\infty} \int_0^{y} e^{-sx} \lambda e^{-\lambda x} \, dA(y) \, dx)^n$$
$$= \sum_{n=0}^{\infty} L_A(s+\lambda) L_B(K(s)) (\frac{\lambda}{s+\lambda} (1 - L_A(s+\lambda)) L_B(K(s)))^n$$
$$= \frac{(s+\lambda) L_A(s+\lambda) L_B(K(s))}{s+\lambda-\lambda(1-L_A(s+\lambda)) L_B(K(s))} \equiv \pi(s). \tag{3.83}$$

However, according to the characteristics of the model, it can be seen that

$$E(e^{-sW^{(i+1)}}) = (\pi(s))^{i+1},\ i = 0, 1, ... \tag{3.84}$$

We substitute (3.84) into (3.74)-(3.76), (3.80) into (3.74), (3.81) into (3.75), (3.82) into (3.76) and then put the new expressions for (3.74)-(3.76) into (3.73). This yields $L_W(s)$ with the help of the steady state solution of the model. Note that

$$L_W(s) = L_B(K(\lambda(1-s))L_{W_q}(s). \tag{3.85}$$

This leads to our expression for $L_{W_q}(s)$.

In order to calculate EW and EW_q, we use the formulas $EW = -\lim_{s\to 0^+} L'_W(s)$ and $EW_q = -\lim_{s\to 0^+} L'_{W_q}(s)$ and then get (3.70) and (3.71).

Chapter 4

Model II Description and Stability

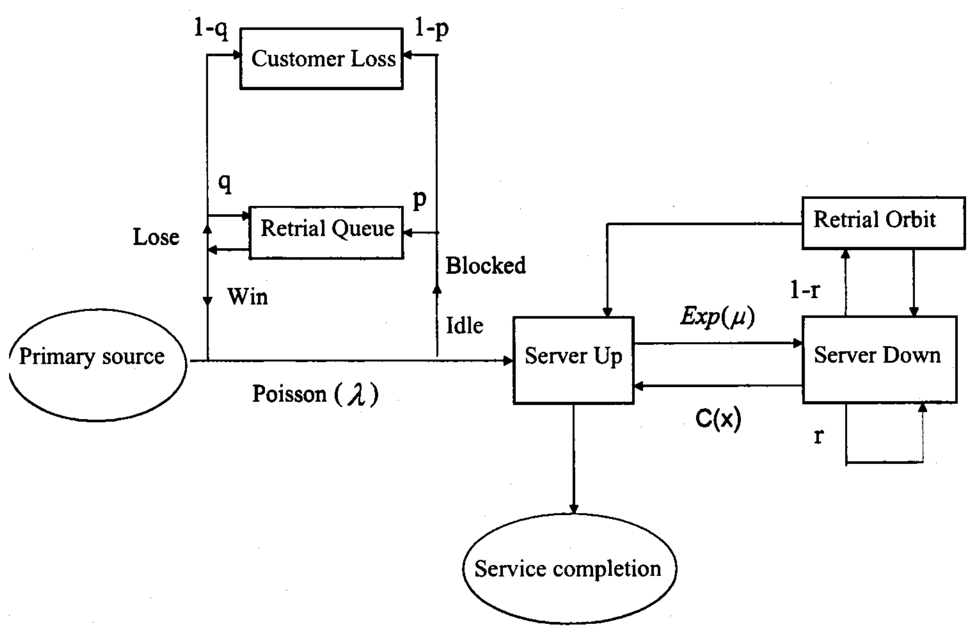
4.1 Description of model II

A single unreliable server retrial queue with general retrial times and balking customers is considered. Primary customers arrive in a Poisson process with rate λ. There is no waiting room in front of the server. If an arriving primary customer finds the server idle, the customer begins service immediately and leaves the system after service completion. If an arriving primary customer finds the server busy, the primary customer either enters a retrial queue with probability p according to a FCFS discipline and becomes a repeated customer or leaves the system with probability $1-p$. For customers in the retrial queue, only the customer at the head of the retrial queue is allowed to attempt to reach the server at a time generally distributed and measured from the instant that the server becomes idle. If a primary customer arrives first while the server is idle, the retrial customer is required to cancel the attempt for service. At that moment, the repeated customer either returns to the position at the head of the retrial queue with probability q or leaves the system with probability $1-q$. The server fails in an exponential time with rate μ but failures can only occur when a customer is being served. When the server fails, repair begins immediately. Meanwhile, the customer in service either stays in the service position with probability r until the server is up again or enters a retrial orbit with probability $1-r$ and keeps making a retrial at times exponentially distributed with rate θ until the customer finds the server up again.

Example 4.1 We again consider an information center with a single server using a telephone and an answering machine. If a customer calls and finds the server busy, the customer either leaves a message on the answering machine according to a FCFS discipline (with probability p) or gives up (with probability 1-p). All messages on the answering machine form the retrial queue. When the server becomes idle, the messages on the answering machine are checked and the first caller on the answering machine is contacted unless a new primary customer phones the center before the contact with the retrial customer is made. The customer whose message is on the answering machine may choose to remove that message after certain time intervals (with probability 1-q), if the time is too long or the service is no longer useful. The server may fail (or be interrupted) during service. During the repair time (interruption time), the telephone is engaged by the customer in service. The customer may become impatient and begin other tasks (with probability 1-r), checking occasionally to see if the server is back. The customer in service remains waiting for the server (with probability r) After the server returns, and the customer returns from its other tasks (if any), the customer resumes service. The retrial times are independently and identically distributed with distribution function $A(x)$, density function $a(x)$ and Laplace transform $L_A(s)$. The service times are independent and identically distributed with distribution function $B(x)$, density function $b(x)$, Laplace transform $L_B(s)$ and first two moments β_1 and β_2. The repair times are independent and identically distributed with distribution function $C(x)$, density function $c(x)$, Laplace transform $L_C(s)$ and first two moments γ_1 and γ_2.

The service time for a customer is cumulative and after repair the server is as good as new. Retrial times, service times, failure times, repair times and the reserved times are also assumed to be mutually independent. The model is described in Figure 4.1.

Figure 4.1: Outline of the $M/G/1$ retrial queue with general retrial times, balking, retrials of customer in service and repairable server.



4.2 Evolution of the queueing system and its states

Let d_i represent the departure time for ith customer in service, $i = 1, 2 \ldots,$. At time d_{i-1}, the $(i-1)$st customer completes service and leaves the system so the server is then idle. According to the assumptions of the model, the ith customer begins to be served after an idle time κ_i. At time d_{i-1}, if the retrial queue is empty, κ_i is exponentially distributed with rate λ; otherwise, if the retrial queue is nonempty, there is a competition between primary customers and a repeated customer to decide who will be the next customer in service. Then κ_i is the minimum of two random times, one exponentially distributed with rate λ and the other generally distributed with Laplace transform $L_A(s)$. If the retrial queue is nonempty, it can be proved that the probability the ith customer is a primary customer after this competition is equal to $1-L_A(\lambda)$. The probability that the repeated customer returns to its position at the head of the retrial queue and does not win this competition is equal to $q(1-L_A(\lambda))$. The probability that the repeated customer does not win this competition and leaves the system is equal to $(1-q)(1-L_A(\lambda))$. The probability that the ith customer is a retrial customer is equal to $L_A(\lambda)$. At time $d_{i-1}+\kappa_i$, the ith customer begins to be served. Later, the server may fail after a busy time $S_i^{(1)}$, which implies that at time $d_{i-1}+\kappa_i+S_i^{(1)}$, the server begins to be repaired.

Meanwhile the ith customer either stays in the service position with probability r or enters the retrial orbit with probability $1-r$ and keeps making a retrial until the service resumes. Suppose that the time needed until the ith customer resumes its service is $\psi_i^{(1)}$. This is equal to a repair time $v_i^{(1)}$ distributed with distribution function $C(x)$ if the ith customer stays in the service position when the server is being repaired, or equal to a repair time $v_i^{(1)}$ distributed with distribution function $C(x)$ plus a exponential reserved time $\varsigma_i^{(1)}$ with rate θ if the ith customer enters the retrial orbit when the server is being repaired. At time $d_{i-1}+\kappa_i+S_i^{(1)}+\psi_i^{(1)}$, the

ith customer resumes service. The queueing process develops similarly during the ith generalized service time until the customer leaves the system at time d_i.

Suppose that the ith customer completes the service and leaves the system at time $d_i \equiv d_{i-1}+\kappa_i+S_i^{(1)}+\psi_i^{(1)}+S_i^{(1)}+\psi_i^{(2)}+...+S_i^{(k)}+\psi_i^{(k)}+S_i^{(k+1)}$, where $S_i^{(1)}+S_i^{(2)}+...+S_i^{(k)}$ and $\psi_i^{(1)}+\psi_i^{(2)}+...+\psi_i^{(k)}+S_i^{(k+1)}$ represent the total busy time and the total repair or reserved time for the ith customer. From the time $d_{i-1}+\kappa_i$ to the time d_i, an arriving primary customer either enters a retrial queue with probability p or leaves the system with probability $1-p$. After time d_i, the queueing process continues in a similar way as time goes on.

The state of the system at time t can be described by the Markov process $\{X(t), t \geq 0\} = \{(J(t), J^*(t), Q(t), \xi_0(t), \xi_1(t), \xi_2(t), \xi_3(t)); t \geq 0\}$ where $J(t)$ represents the server state $(0, 1, 2, 3$ denote the server idle or busy or under repair or reserved, respectively) and $J^*(t)$ represents the customer state in the service area (0 or 1 denote the retrial customer in service position when the server is being repaired or in the retrial orbit after the server breakdown, respectively). $Q(t)$ denotes the number of customers in the retrial queue at time t. If $J(t) = 0$ and $Q(t) > 0$, then $\xi_0(t)$ denotes the elapsed retrial time for a customer in the retrial queue. If $J(t) = 1$, $J(t) = 2$ or $J(t) = 3$, then $\xi_1(t)$ denotes the elapsed service time . If $J(t) = 2$ and $J^*(t) = 0$ or 1, then $\xi_2(t)$ denotes the elapsed repair time. If $J(t) = 3$, then $\xi_3(t)$ denotes the elapsed reserved time. The functions $\alpha(x), \beta(x)$ and $\gamma(x)$ are the conditional completion rates for the retrial attempt, the service and the repair at time x, respectively, i.e., $\alpha(x) = a(x)(1-A(x))^{-1}$, $\beta(x) = b(x)(1-B(x))^{-1}$, and $\gamma(x) = c(x)(1-C(x))^{-1}$. Therefore, $\{X(t),\, t \geq 0\} = \{(J(t),\, J^*(t),\, Q(t),\, \xi_0(t),\, \xi_1(t),\, \xi_2(t),\, \xi_3(t));\, t \geq 0\}$ is a

Markovian process with the state space $\Omega = \{((0,0)\} \cup \{(0,i,w)); 1 \leq i < \infty\} \cup \{((1,i,x),\ (2,0,i,x,y),\ (2,1,i,x,y),\ (3,i,x,\tau);\ 0 \leq i < \infty,\ 0 \leq x, y, \tau < \infty\}$, where $(0,0)$ means that the server is idle and the retrial queue is empty, $(0,i,w)$ means that the server is idle and there are i customers in the retrial queue with elapsed retrial time w, $(1,i,x)$ means that the server is busy with elapsed service time x and i customers are in the retrial queue, $(2,0,i,x,y)$ means that the server is under repair and the customer stays in the service position with elapsed service time x and with elapsed repair time y and i customers are in the retrial queue after breakdown, $(2,1,i,x,\tau)$ means that the customer is in the retrial orbit and the server is reserved with elapsed service time x and with elapsed reserved time τ and i customers are in the retrial queue after breakdown until the server is up, $(3,i,x,\tau)$ means that the customer is in the retrial orbit and the server is reserved with elapsed service time x and with elapsed reserved time τ and i customers are in the retrial queue after the server is up.

4.3 An embedded Markov chain

The stochastic process $\{(J(t),\ J^*(t),\ Q(t)); t \geq 0\}$ is not a Markov process, however, the embedded stochastic process $\{Q_n; t \geq 1\}$ is a Markov chain, where Q_n represents the number of customers in the retrial queue at the nth departure time d_n, $n = 1, 2, \ldots$. In order to see that $\{Q_n; n = 1, 2, \ldots\}$ is a Markov chain, let S_n, $S^{(n)}$ represent the nth service time and generalized service time and D_{n+1} represent the number of primary customers arriving to the system during $S^{(n)}$. Let $N_1(D_{n+1})$ represent the number of primary customers who enter the retrial queue among the D_{n+1} primary customers arriving during $S^{(n+1)}$. Note that if the retrial queue is nonempty at any departure point, then there exists a competition between a primary customer and a repeated customer to determine who is the next customer served. For convenience, let "$t_p < t_r$" denote the event that the primary customer wins in this competition, let "$t_p < t_r \wedge q$" denote the event that the primary customer wins in this competition and the repeated customer returns to the retrial queue, let "$t_p < t_r \vee q$" denote the event that the primary customer wins in this competition and the repeated customer leaves the system, and let "$t_r < t_p$" denote the event that the repeated customer wins in this competition. Therefore,

$$Q_{n+1} = \begin{cases} N_1(D_{n+1}), & \text{if } Q_n = 0, \\ Q_n + N_1(D_{n+1}), & \text{if } Q_n > 0 \text{ and } t_p < t_r \wedge q, \\ Q_n + N_1(D_{n+1}) - 1, & \text{if } Q_n > 0 \text{ and } t_p < t_r \vee q \text{ or } t_r < t_p. \end{cases} \tag{4.1}$$

To find $P(Q_{n+1} = j \mid Q_n = i)$, we must obtain the distribution of $S^{(n)}$. Let $N_2(S_n)$ represent the number of failures during S_n and $N_3(N_2(S_n))$ represent the number of times that the customer in service stays in the service position for the $N_2(S_n)$ failures. By the assumption of the model, it is easily seen that $N_1(k)$ is binomially distributed with the parameters k and p, that $N_2(t)$ is a Poisson random variable with rate μt, and that $N_3(i))$ is binomially distributed with the parameters i and r. Therefore, $S^{(n)}$ is distributed with distribution function

$$\begin{aligned} B^*(x) &= P(S^{(n)} \leq x) \\ &= \sum_{k=0}^{\infty} \int_0^x P(S^{(n)} \leq x \mid N_2(x),\, S_n = x) \frac{(\mu y)^k}{k!} e^{-\mu y}\, dB(y) \end{aligned}$$

$$= \sum_{k=0}^{\infty} \sum_{i=0}^{k} \int_0^x P(S^{(n)} \le x \mid N_2(x),\, N_3(k) = i,\, S_n = x)$$

$$\times \binom{k}{i} r^i (1-r)^{k-i} \frac{(\mu y)^k}{k!} e^{-\mu y}\, dB(y)$$

$$= \sum_{k=0}^{\infty} \sum_{i=0}^{k} \int_0^x \binom{k}{i} r^i (1-r)^{k-i} \frac{(\mu y)^k}{k!} e^{-\mu y} C^{(2)}_{k,i}(x-y)\, dB(y), \tag{4.2}$$

and Laplace transform

$$L_{B^*}(s) = \sum_{i=0}^{\infty} \int_0^x E(e^{-sS^{(k)}} \mid N_1(y) = i,\, S_k = y) \frac{(\mu y)^i}{i!} e^{-\mu y}\, dB(y)$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{i} \int_0^x E(e^{-sS^{(k)}} \mid N_1(y) = i,\, N_2(i) = j,\, S_k = y)$$

$$\times \binom{i}{j} r^j (1-r)^{i-j} \frac{(\mu y)^i}{i!} e^{-\mu y}\, dB(y)$$

$$= L_B(s + \mu - \mu \frac{rs+\theta}{s+\theta} L_C(s)), \tag{4.3}$$

where $C^{(2)}_{k,i}(x)$ represents the convolution of $C^{(k)}(x)$ (which is the k-fold convolution of $C(x)$) and the gamma distribution with the parameters $k-i$ and $\frac{1}{\theta}$.

(1) For $i = 0$ and $j \ge 0$, we have

$$P(Q_{n+1} = j \mid Q_n = 0) = \int_0^\infty P(Q_{n+1} = j \mid S^{(n)} = x)\, dB^*(x)$$

$$= \sum_{k=j}^{\infty} \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x} P(N_1(D_{n+1}) = j \mid S^{(n)} = x,\, D_{n+1} = k)\, dB^*(x)$$

$$= \sum_{k=j}^{\infty} \int_0^\infty \frac{(\lambda x)^k}{k!} e^{-\lambda x} C^j_k p^j (1-p)^{k-j}\, dB^*(x)$$

$$= \int_0^\infty \frac{(p\lambda x)^j}{j!} e^{-p\lambda x}\, dB^*(x). \tag{4.4}$$

(2) For $i = 1, 2, \ldots$ and $j = i-1$, we have

$$P(Q_{n+1} = j \mid Q_n = i) = (1 - L_A(\lambda)) P(Q_{n+1} = i-1 \mid Q_n = i,\, t_p < t_r)$$

$$+ L_A(\lambda) P(Q_{n+1} = i-1 \mid Q_n = i,\, t_r < t_p)$$

$$= (1 - L_A(\lambda))(1-q)\sum_{k=0}^{\infty}\int_0^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}(1-p)^k\,dB^*(x)$$
$$+L_A(\lambda)\sum_{k=0}^{\infty}\int_0^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}(1-p)^k\,dB^*(x)$$
$$= (1 - L_A(\lambda))(1-q)\int_0^{\infty}e^{-p\lambda x}\,dB^*(x) + L_A(\lambda)\int_0^{\infty}e^{-\lambda x}\,dB^*(x)$$
$$= (1-q+qL_A(\lambda))\int_0^{\infty}e^{-\lambda x}\,dB^*(x). \tag{4.5}$$

(3) For $i = 1, 2, \ldots$ and $j = i, i+1, \ldots$, we have

$$P(Q_{n+1} = j \mid Q_n = i)$$
$$= (1-L_A(\lambda))q\sum_{k=j-i}^{\infty}\int_0^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}P(N_1(D_{n+1}) = j-i \mid$$
$$S^{(n)} = x,\ D_{n+1} = k)\,dB^*(x)$$
$$+(1-L_A(\lambda))(1-q)\sum_{k=j-i+1}^{\infty}\int_0^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}P(N_1(D_{n+1}) = j-i+1 \mid$$
$$S^{(n)} = x,\ D_{n+1} = k)\,dB^*(x)$$
$$+L_A(\lambda)\sum_{k=j-i+1}^{\infty}\int_0^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}P(N_1(D_{n+1}) = j-i+1 \mid$$
$$S^{(n)} = x,\ D_{n+1} = k)\,dB^*(x)$$
$$= (1-L_A(\lambda))q\sum_{k=j-i}^{\infty}\int_0^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}C_k^{j-i}p^{j-i}(1-p)^{k-j+i}\,dB^*(x)$$
$$+(1-L_A(\lambda))(1-q)\sum_{k=j-i+1}^{\infty}\int_0^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}C_k^{j-i+1}p^{j-i+1}$$
$$\times(1-p)^{k-j+i-1}\,dB^*(x) + L_A(\lambda)\sum_{k=j-i+1}^{\infty}\int_0^{\infty}\frac{(\lambda x)^k}{k!}e^{-\lambda x}C_k^{j-i+1}$$
$$\times p^{j-i+1}(1-p)^{k-j+i-1}\,dB^*(x)$$
$$= (1-L_A(\lambda))q\int_0^{\infty}\frac{(p\lambda x)^{j-i}}{(j-i)!}e^{-p\lambda x}\,dB^*(x)$$
$$+(1-q+qL_A(\lambda))\int_0^{\infty}\frac{(p\lambda x)^{j-i+1}}{(j-i+1)!}e^{-p\lambda x}\,dB^*(x). \tag{4.6}$$

(4) Otherwise, $P(Q_{n+1} = j \mid Q_n = i) = 0$.

Summarizing the results above, we obtain the one-step transition probabilities of the embedded stochastic process $\{Q_n,\, n \geq 1\}$

$$p_{ij} = \begin{cases} \int_0^\infty \frac{(p\lambda x)^j}{j!} e^{-p\lambda x}\, dB^*(x), & \text{if} i = 0, j = 0, 1, ..., \\ (1 - q + qL_A(\lambda)) \int_0^\infty e^{-p\lambda x} dB^*(x), & \text{if} i = 1, 2, ..., j = i - 1, \\ q(1 - L_A(\lambda)) \int_0^\infty \frac{(p\lambda x)^{j-i+1}}{(j-i+1)!} e^{-p\lambda x}\, dB^*(x) & \\ +(1 - q + qL_A(\lambda)) \int_0^\infty \frac{(p\lambda x)^{j-i}}{(j-i)!} e^{-p\lambda x}\, dB^*(x), & \text{if} i = 1, 2, ..., j = i, i + 1, ..., \\ 0, & \text{otherwise.} \end{cases} \tag{4.7}$$

Thus, the sequence of random variables $\{Q_n, n \geq 1\}$ constitutes a Markov chain, which is an embedded Markov chain for the queueing system.

4.4 The stability of the system

Theorem 4.1 The system is stable if and only if $p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)) < 1-q+qL_A(\lambda)$.

Proof. Let Q_n be the system size, i.e., the retrial queue length at the nth departure. Since primary customers arrive in a Poisson process, we use Burke's result (see Cooper (1981)) which essentially states that the steady state probabilities of $\{(J(t), J^*(t), Q(t)); t \geq 0\}$ exist and are positive if and only if $\{Q_n, n \geq 1\}$ is ergodic. Therefore,it is sufficient to prove that $p\lambda\beta_1(1 + \mu(\frac{1-r}{\theta} + \gamma_1)) < 1 - q + qL_A(\lambda)$ is a necessary and sufficient condition for an embedded Markov chain $\{Q_n;\, n \geq 1\}$ to be ergodic.

Assume $p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)) < 1-q+qL_A(\lambda)$. To show ergodicity, it is easily seen that $\{Q_n;\ n \geq 1\}$ is irreducible and aperiodic. It remains to prove that it is positive recurrent. We employ Theorem 2 of Pakes (1969) which states that for an irreducible and aperiodic Markov chain $Q_n;\ n \geq 1\}$ with the state space Ξ, a sufficient condition for ergodicity for $\{Q_n;\ n \geq 1\}$ to be positive recurrent is the existence of a nonnegative function $f(x), x \in S$ and $\epsilon > 0$ such that the mean drift χ_j is finite for all $j \in \Xi$ and $\chi_j \leq -\epsilon$ for all j except perhaps a finite number, where $\chi_j \equiv E(Q_{n+1} - Q_n | Q_n = j)$. However, according to equation (4.3), it can be shown that the Laplace transform of $S^{(n)}$ is

$$L_{B^*}(s) = L_B(s + \mu - \mu\frac{rs+\theta}{s+\theta}L_C(s)), \tag{4.8}$$

and its expected value is given by

$$E(S_n^{(*)}) = \beta_1(1 + \mu(\frac{1-r}{\theta} + \gamma_1)). \tag{4.9}$$

Therefore,

$$\chi_j = p\lambda\beta_1(1 + \mu(\frac{1-r}{\theta} + \gamma_1)) - (1 - \delta_{j0})(1 - q + qL_A(\lambda)), \tag{4.10}$$

for any j where δ_{jk} is the Kronecker delta. This implies that $\{Q_n,\ n = 1, 2, ...\}$ is positive recurrent and therefore is ergodic.

Conversely, if the system is stable and we assume that $p\lambda\beta_1(1 + \mu(\frac{1-r}{\theta} + \gamma_1)) \geq 1-q+qL_A(\lambda)$ holds, we employ Theorem 1 of Sennott et al. (1983) which states that the Markov chain $\{Q_n;\ n \geq 1\}$ is not ergodic if it satisfies Kaplan's condition and the mean drift χ_j is finite for all j and there exists N such that $\chi_j \geq 0$ for $j \geq N$. In our case, $P_{ij} = 0$ $(j < i,\ i > 0)$ which implies that Kaplan's condition is satisfied by using

Theorem 1 of Sennott et al. (1983). Thus, $\{Q_n;\ n \geq 1\}$ is not ergodic. Therefore, we have shown that $p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)) < 1-q+qL_A(\lambda)$ is a necessary and sufficient condition for $\{Q_n;\ n \geq 1\}$ to be ergodic.

This completes the proof of the theorem. '

Remark 4.1 There exist four states of a non-idle server, that is, the server is busy or under repair in two cases (customer stays in service or orbits) or reserved. It can be shown that the joint Laplace transform of T_1, T_2, T_3 and T_4 is given by

$$\begin{aligned} &L(s_1, s_2, s_3, s_4) \equiv E(e^{-s_1T_1-s_2T_2-s_3T_3-s_4T_4)}) \\ &= L_B(s_1+\mu-\mu(rL_C(s_2)+\frac{(1-r)\theta}{s_4+\theta}L_C(s_3))), \end{aligned} \tag{4.11}$$

where T_1, T_2, T_3 and T_4 are, respectively, the service time, the total repair time for the customer in the service position, the total retrial time for the customer in the retrial orbit and the total reserved times. Also, their corresponding mean times are $\beta_1, r\beta_1\mu\gamma_1, (1-r)\beta_1\mu\gamma_1$ and $\frac{(1-r)\beta_1\mu}{\theta}$.

Remark 4.2 The $M/G/1$ retrial queue with retrials of the customer in service and balking may be viewed as an extension of other queueing models.

(i) If $A(x) = 1 - e^{-\alpha x}$ and $\alpha = \infty$, then this retrial queue reduces to the classical $M/G/1$ queue with repairable server and balking.

(ii) If $\mu = 0$, this retrial queue reduces to the $M/G/1$ retrial queue without repairable server.

(iii) If $p = q = r = 1$, $\mu \neq 0$, $\theta = \infty$, this retrial queue reduces to the $M/G/1$ retrial queue with repairable server but no retrials of the customer in service and no customer loss.

Remark 4.3 Assume the retrial queue has a large number of customers. Note that $p\lambda$ is the arrival rate of primary customers to the retrial queue and that $\frac{L_A(\lambda)}{\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}$ is the exit rate of retrial customers to the service position. Also $\frac{1-L_A(\lambda)}{\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}$ is the exit rate of retrial customers who are unsuccessful in accessing the server and must return to the retrial queue or leave. This implies that $\frac{(1-q)(1-L_A(\lambda))}{\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}$ is the loss rate of the retrial customers. Therefore, $\frac{L_A(\lambda)+(1-q)(1-L_A(\lambda))}{\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))} = \frac{1-q+qL_A(\lambda)}{\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}$ is the total exit rate of the retrial customers from the retrial queue. For the system stability, the arriving rate of primary customers to the retrial queue must be less than the total exit rate for retrial customers from the retrial queue. That is, the inequality $p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)) < 1-q+qL_A(\lambda)$ must hold.

Remark 4.4 Suppose the retrial queue has a large number of customers in the following discussion. Let $P(\text{block})$ be the probability that the system is blocked. We have an alternating renewal process between the blocked and idle states for the server. As in Section 2.5, let $E(S^*)$ be the expected blocked time. Let $E(T)$ be the expected idle time. Then

$$P(\text{block}) = \frac{E(S^*)}{E(S^*)+E(T)},$$

and

$$P(\text{not block}) = \frac{E(T)}{E(S^*)+E(T)}.$$

The arrival rate to the retrial queue is $\lambda p P(\text{block})$. If the server becomes idle, then

$E(T)$ is the expected idle time. The probability that the next customer in service comes from the retrial queue is $L_A(\lambda)$. If the next customer is a primary customer then with probability $(1-q)$, the retrial customer will exit the system.

The exit rate from the retrial queue by entering service is

$$P(\text{block})L_A(\lambda)\frac{1}{E(T)}.$$

The exit rate from the retrial queue by leaving the system when a primary customer arrives first to the server is

$$P(\text{not block})(1-L_A(\lambda))\frac{1}{E(T)}(1-q).$$

The total exit rate from the retrial queue is

$$P(\text{not block})\frac{1}{E(T)}(1-q+qL_A(\lambda)).$$

For stability, we need the arrival rate to be less than the total exit rate

i.e.

$$\lambda p P(\text{block}) < P(\text{not block})\frac{1}{E(T)}(1-q+qL_A(\lambda)).$$

This is equivalent to

$$\lambda p E(S^*) < 1-q+qL_A(\lambda),$$

i.e.

$$p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)) < 1-q+qL_A(\lambda).$$

Chapter 5

Model II Steady State Equations and Their Solution

5.1 The forward equations of the system

The probability densities of the system states are defined as follows:

(i) $P_{(0,0)}(t) \equiv P(J(t) = 0,\ Q(t) = 0)$, for $0 \leq t < \infty$,

(ii) $P_{(0,i)}(t,\ w)\, dw \equiv P(J(t) = 0,\ Q(t) = i,\ w < \xi_0(t) < w + dw)$, for $0 \leq t,\ w < \infty,\ i = 1, 2, ...,$

(iii) $P_{(1,i)}(t,\ x)\, dx \equiv P(J(t) = 1,\ Q(t) = i,\ x < \xi_1(t) < x + dx)$, for $0 \leq t,\ x < \infty,\ i = 0, 1, ...,$

(iv) $P_{(2,0,i)}(t,\ x,\ y)\, dx\, dy \equiv P(J(t) = 2,\ J^*(t) = 0,\ Q(t) = i,\ x < \xi_1(t) < x + dx, y < \xi_2(t) < y + dy)$, for $0 \leq t,\ x,\ y < \infty,\ i = 0, 1, ...,$

(v) $P_{(2,1,i)}(t,\ x,\ y)\, dx\, dy \equiv P(J(t) = 2,\ J^*(t) = 1,\ Q(t) = i,\ x < \xi_1(t) < x + dx,\ y < \xi_2(t) < y + dy)$, for $0 \leq t,\ x,\ y < \infty,\ i = 0, 1, ...,$

(vi) $P_{(3,i)}(t,\ x,\ \tau)\, dx\, d\tau \equiv P(J(t) = 3,\ Q(t) = i,\ x < \xi_1(t) < x + dx,\ \tau < \xi_3(t) <$

$\tau + d\tau)$, for $0 \leq t,\, x,\, \tau < \infty$, $i = 0, 1, \ldots$.

Next, we analyze $P_{(0,1)}(t), P_{(0,i)}(t,\, w)$, $P_{(1,i)}(t,\, x)$, $P_{(2,0,i)}(t,\, x,\, y)$, $P_{(2,1,i)}(t,\, x,\, y)$ and $P_{(3,i)}(t,\, x,\, \tau)$, respectively.

By using a manner similar to (3.1.1) and (3.1.2), we obtain the following difference equations for $P_{(0,0)}(t)$ and $P_{(0,i)}(t,\, w)$.

$$P_{(0,0)}(t + \Delta t) = (1 - \lambda \triangle t)P_{(0,0)}(t) + \Delta t \int_0^\infty P_{(1,0)}(t,\, x)\beta(x)\, dx + o(\Delta t), \quad (5.1)$$

$$P_{(0,i)}(t + \Delta t, w + \Delta t) = (1 - (\lambda + \alpha(w)) \triangle t)P_{(0,i)}(t,\, w) + o(\Delta t). \quad (5.2)$$

Next,we derive other equations related to $P_{(1,i)}(t,\, x)$, $P_{(2,0,i)}(t,\, x,\, y)$, $p_{(2,1,i)}(t,\, x,\, y)$, $P_{(3,i)}(t,\, x,\, \tau)$ as follows.

(1) The event $(C(t + \Delta t) = 1,\, Q(t + \Delta t) = i,\, \xi_1(t + \Delta t) = x + \Delta t)$ occurs in one of four cases. In the first case, the event $(J(t) = 1,\, Q(t) = i,\, \xi_1(t) = x)$ occurs and there is an arriving primary customer during Δt and the arriving primary customer leaves the system or no arriving primary customers or service completions or breakdowns during Δt. In the second case, n the event $(J(t) = 1,\, Q(t) = i - 1,\, \xi_1(t) = x)$ occurs and there is an arriving primary customer during Δt and the customer enters the retrial queue. In the third case the event $(J(t) = 2,\, J^*(t) = 0,\, Q(t) = i,\, \xi_1(t) = x,\, \xi_2(t) = y)$ occurs and the server is up again during Δt. In the fourth case, the event $(J(t) = 3,\, Q(t) = i,\, \xi_1(t) = x,\, \xi_3(t) = \tau)$ occurs and the customer in the retrial queue obtains the service in Δt. This leads to the difference equation

$$P_{(1,i)}(t + \Delta t, x + \Delta t) = (1 - (p\lambda + \mu + \beta(x)) \triangle t)P_{(1,i)}(t,\, x)$$
$$+ p\lambda \triangle t P_{(1,i-1)}(t,\, x) + \Delta t \int_0^\infty P_{(2,0,i)}(t,\, x,\, y)\gamma(y)\, dy$$

$$+\theta \triangle t \int_0^{\infty} P_{(3,i)}(t, x, \tau)\, d\tau + o(\triangle t), \tag{5.3}$$

where $P_{(1,-1)}(t, x) \equiv 0$.

(2) The event $(C(t+\triangle t) = 2,\ C^*(t+\triangle t) = 0,\ Q(t+\triangle t) = i,\ \xi_1(t+\triangle t) = x,\ \xi_2(t+\triangle t) = y + \triangle t)$ occurs in one of two case. In the first case, the event $(J(t) = 2,\ J^*(t) = 0,\ Q(t) = i,\ \xi_1(t) = x,\ \xi_2(t) = y)$ occurs and there is an arriving primary customer during $\triangle$t and the arriving primary customer leaves the system or there are no arriving primary customers and the server is still under repair during $\triangle$t. In the second case, the event $(J(t) = 2,\ J^*(t) = 0,\ Q(t) = i-1,\ \xi_1(t) = x,\ \xi_2(t) = y)$ occurs and there is an arriving primary customer and the customer enters the retrial queue during $\triangle$t. This leads to the following difference equation

$$P_{(2,0,i)}(t+\triangle t, x, y+\triangle t) = (1-(p\lambda+\gamma(y))\triangle t)P_{(2,0,i)}(t, x, y)$$

$$+p\lambda \triangle t P_{(2,0,i-1)}(t, x, y) + o(\triangle t), \tag{5.4}$$

where $P_{(2,0,-1)}(t, x, y) \equiv 0$.

(3) The event $(C(t+\triangle t) = 2,\ C^*(t+\triangle t) = 1,\ Q(t+\triangle t) = i,\ \xi_1(t+\triangle t) = x,\ \xi_2(t+\triangle t) = y + \triangle t)$ occurs in one of two cases. In the first case, the event $(J(t) = 2,\ J^*(t) = 1,\ Q(t) = i,\ \xi_1(t) = x,\ \xi_2(t) = y)$ occurs and there is an arriving primary customer during $\triangle$t and the arriving primary customer leaves the system or there are no arriving primary customers and the server is still under repair. In the second case, the event $(J(t) = 2,\ J^*(t) = 1,\ Q(t) = i-1,\ \xi_1(t) = x,\ \xi_2(t) = y)$ occurs and there is an arriving primary customer during $\triangle$t and the customer enters the retrial queue during $\triangle$t. This leads to the following difference equation

$$P_{(2,1,i)}(t+\triangle t, x, y+\triangle t) = (1-(p\lambda+\gamma(y))\triangle t)P_{(2,1,i)}(t, x, y)$$

$$+p\lambda \triangle t P_{(2,1,i-1)}(t,\, x,\, y) + o(\Delta t), \tag{5.5}$$

where $P_{(2,1,-1)}(t,\, x,\, y) \equiv 0$.

(4) The event $(C(t+\Delta t) = 3,\, Q(t+\Delta t) = i,\, \xi_1(t+\Delta t) = x,\, \xi_3(t+\Delta t) = \tau + \Delta t)$ occurs in one of two cases. In the first case, the event $(J(t) = 3,\, Q(t) = i,\, \xi_1(t) = x,\, \xi_3(t) = \tau)$ occurs and there is an arriving primary customer during Δt and the arriving primary customer leaves the system or there are no arriving primary customers or arriving repeated customers in the retrial queue during Δt. In the second case, the event $(J(t) = 3,\, Q(t) = i-1,\, \xi_1(t) = x,\, \xi_3(t) = \tau)$ occurs and there is an arriving primary customer during Δt and the customer enters the retrial queue. This leads to the following difference equation

$$\begin{aligned} &P_{(3,i)}(t+\Delta t,\, x,\, \tau+\Delta t) = (1-(p\lambda+\theta)\triangle t)P_{(3,i)}(t,\, x,\, y) \\ &+p\lambda \triangle t P_{(3,i-1)}(t,\, x,\, \tau) + o(\Delta t), \end{aligned} \tag{5.6}$$

where $P_{(3,-1)}(t,\, x,\, y) \equiv 0$.

Similarly, we have the following boundary conditions:

$$P_{(0,i)}(t,\, 0) = \int_0^\infty P_{(1,i)}(t,\, x)\beta(x)\, dx, \tag{5.7}$$

$$\begin{aligned} &P_{(1,i)}(t,\, 0) = \lambda\delta_{i0}P_{(0,0)}(t) + q\lambda(1-\delta_{i0})\int_0^\infty P_{(0,i)}(t,\, w)\, dw \\ &+(1-q)\lambda\int_0^\infty P_{(0,i+1)}(t,\, w)\, dw + \int_0^\infty P_{(0,i+1)}(t,\, w)\alpha(w)\, dw, \end{aligned} \tag{5.8}$$

$$P_{(2,0,i)}(t,\, x,\, 0) = r\mu P_{(1,i)}(t,\, x), \tag{5.9}$$

$$P_{(2,1,i)}(t,\, x,\, 0) = (1-r)\mu P_{(1,i)}(t,\, x), \tag{5.10}$$

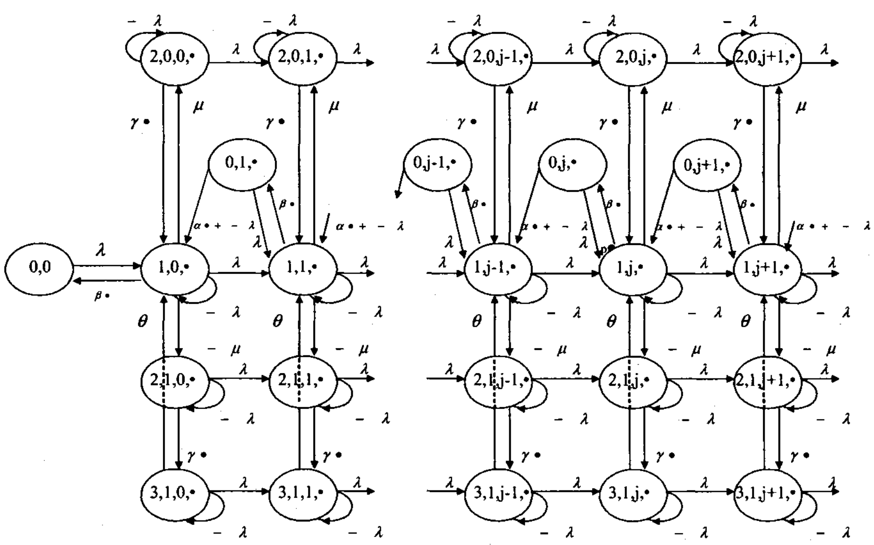
$$P_{(3,i)}(t,\, x,\, 0) = \int_0^\infty P_{(2,1,i)}(t,\, x,\, y)\gamma(y)\, dy. \tag{5.11}$$

The normalization equation is given by

$$
\begin{aligned}
&P_{(0,0)}(t) + \sum_{i=1}^{\infty} \int_0^{\infty} P_{(0,i)}(t,\, w)\, dw + \sum_{i=0}^{\infty} \int_0^{\infty} P_{(1,i)}(t,\, x)\, dx \\
&+ \sum_{i=0}^{\infty} \int_0^{\infty} \int_0^{\infty} P_{(2,0,i)}(t,\, x,\, y)\, dx\, dy + \sum_{i=0}^{\infty} \int_0^{\infty} \int_0^{\infty} P_{(2,1,i)}(t,\, x,\, y)\, dx\, dy \\
&+ \sum_{i=0}^{\infty} \int_0^{\infty} \int_0^{\infty} P_{(3,i)}(t,\, x,\, \tau)\, dx\, d\tau = 1. \qquad (5.12)
\end{aligned}
$$

The rate transition diagram in the model is illustrated in Figure 5.1.

Figure 5.1: Rate transition diagram-M/G/1 retrial queue with general retrial times, unreliable server and balking.



These difference equations yield the following differential equations:

$$(\frac{d}{dt} + \lambda)P_{(0,0)}(t) = \int_0^\infty P_{(1,0)}(t,\, x)\beta(x)\, dx, \tag{5.13}$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial w} + \lambda + \alpha(w))P_{(0,\, i)}(t,\, w) = 0, \tag{5.14}$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial x} + p\lambda + \mu + \beta(x))P_{(1,\, i)}(t,\, x) = \int_0^\infty P_{(2,\, 0,\, i)}(t,\, x,\, y)\gamma(y)\, dy$$
$$+\theta \int_0^\infty P_{(3,\, i)}(t,\, x,\, \tau)\, d\tau + p\lambda P_{(1,\, i-1)}(t,\, x), \tag{5.15}$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial y} + p\lambda + \gamma(y))P_{(2,\, 0,\, i)}(t,\, x,\, y) = p\lambda P_{(2,\, 0,\, i-1)}(t,\, x,\, y), \tag{5.16}$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial y} + p\lambda + \gamma(y))P_{(2,\, 1,\, i)}(t,\, x,\, y) = p\lambda P_{(2,\, 1,\, i-1)}(t,\, x,\, y), \tag{5.17}$$

$$(\frac{\partial}{\partial t} + \frac{\partial}{\partial \tau} + p\lambda + \theta)P_{(3,\, i)}(t,\, x,\, \tau) = p\lambda P_{(3,\, i-1)}(t,\, x,\, \tau). \tag{5.18}$$

Assume that $p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)) < 1-q+qL_A(\lambda)$, which implies that the system is stable. Therefore, $\lim_{t\to+\infty} P_{(0,0)}(t)$, $\lim_{t\to+\infty} P_{(0,i)}(t,\, w)$, $\lim_{t\to+\infty} P_{(1,i)}(t,\, x)$, $\lim_{t\to+\infty} P_{(2,0,i)}(t,\, x,\, y)$, $\lim_{t\to+\infty} P_{(2,1,i)}(t,\, x,\, y)$ and $\lim_{t\to+\infty} P_{(3,i)}(t,\, x,\, \tau)$ exist and are denoted by $P_{(0,0)}$, $P_{(0,i)}(w)$, $P_{(1,i)}(x)$, $P_{(2,i)}(x,\, y)$, $P_{(2,1,i)}(x,\, y)$ and $P_{(3,i)}(x,\, \tau)$, respectively.

Thus, (5.7)-(5.18) lead to the following results:

$$\lambda P_{(0,0)} = \int_0^\infty P_{(1,0)}(x)\,\beta(x)\,dx, \tag{5.19}$$

$$(\frac{d}{dw}+\lambda+\alpha(w))P_{(0,i)}(w) = 0, \tag{5.20}$$

$$\begin{aligned}&(\frac{d}{dx}+p\lambda+\mu+\beta(x))P_{(1,i)}(x) = \int_0^\infty P_{(2,0,i)}(x,\, y)\gamma(y)\,dy\\&+\theta\int_0^\infty P_{(3,i)}(x,\, \tau)\,d\tau + p\lambda P_{(1,i-1)}(x),\end{aligned} \tag{5.21}$$

$$(\frac{\partial}{\partial y}+p\lambda+\gamma(y))P_{(2,0,i)}(x,\, y) = p\lambda P_{(2,0,i-1)}(x,\, y), \tag{5.22}$$

$$(\frac{\partial}{\partial y}+p\lambda+\gamma(y))P_{(2,1,i)}(x,\, y) = p\lambda P_{(2,1,i-1)}(x,\, y), \tag{5.23}$$

$$(\frac{\partial}{\partial \tau}+p\lambda+\theta)P_{(3,i)}(x,\, \tau) = p\lambda P_{(3,i-1)}(x,\, \tau), \tag{5.24}$$

$$P_{(0,i)}(0) = \int_0^\infty P_{(1,i)}(x)\beta(x)\,dx, \tag{5.25}$$

$$\begin{aligned}P_{(1,i)}(0) &= \int_0^\infty P_{(0,i+1)}(w)\alpha(w)\,dw + (1-q)\lambda\int_0^\infty P_{(0,i+1)}(w)\,dw\\&+(1-\delta_{i0})q\lambda\int_0^\infty P_{(0,i)}(w)\,dw + \delta_{i0}\lambda P_{(0,0)},\end{aligned} \tag{5.26}$$

$$P_{(2,0,i)}(x,\, 0) = r\mu P_{(1,i)}(x), \tag{5.27}$$

$$P_{(2,1,i)}(x,\, 0) = (1-r)\mu P_{(1,i)}(x), \tag{5.28}$$

$$P_{(3,i)}(x,\, 0) = \int_0^\infty P_{(2,1,i)}(x,\, y)\gamma(y)\,dy, \tag{5.29}$$

$$P_{(0,0)} + \sum_{i=1}^{\infty} \int_0^{\infty} P_{(0,i)}(w)\,dw + \sum_{i=0}^{\infty} (\int_0^{\infty} P_{(1,i)}(x)\,dx + \int_0^{\infty}\int_0^{\infty} P_{(2,0,i)}(x,\,y)\,dx\,dy$$
$$+ \int_0^{\infty}\int_0^{\infty} P_{(2,1,i)}(x,\,y)\,dx\,dy + \int_0^{\infty}\int_0^{\infty} P_{(3,i)}(x,\,\tau)\,dx\,d\tau) = 1. \tag{5.30}$$

where

$$G(x) \equiv x + \mu - \mu\frac{rx+\theta}{x+\theta}L_C(x). \tag{5.31}$$

In order to solve the system of equations (5.19)-(5.30), we introduce the following probability generating functions:

$$P_0(z,\,w) = \sum_{i=1}^{\infty} P_{(0,i)}(w)z^i, \quad P_1(z,\,x) = \sum_{i=0}^{\infty} P_{(1,i)}(x)z^i,$$

$$P_{20}(z,\,x,\,y) = \sum_{i=0}^{\infty} P_{(2,0,i)}(x,\,y)z^i, \quad P_{21}(z,\,x,\,y) = \sum_{i=0}^{\infty} P_{(2,1,i)}(x,\,y)z^i,$$

$$P_3(z,\,x,\,\tau) = \sum_{i=0}^{\infty} P_{(3,i)}(x,\,\tau)z^i.$$

Theorem 5.1. If $p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)) < 1-q+qL_A(\lambda)$, then there exist the following steady state solutions of the model

$$(i)P_{(0,0)} = \frac{1-q+qL_A(\lambda)-p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}{(1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda)}, \tag{5.32}$$

$$(ii)P_0(z,\,w) = \frac{\lambda z(1-K(z))\exp(-\lambda w)(1-A(w))P_{(0,0)}}{(1-q(1-L_A(\lambda))(1-z))K(z)-z(1-K(z))}, \tag{5.33}$$

$$(iii)P_1(z,\,x) = \frac{\lambda(1-q+qL_A(\lambda))(1-z)}{(1-q(1-L_A(\lambda))(1-z))K(z)-z(1-K(z))},$$
$$\exp(-G(p\lambda(1-z))x)(1-B(x))P_{(0,0)} \tag{5.34}$$

$$(iv)P_{20}(z,\,x,\,y) = \frac{r\lambda\mu(1-q+qL_A(\lambda))(1-z)}{(1-q(1-L_A(\lambda))(1-z))K(z)-z(1-K(z))}$$

$$\times \exp(-G(p\lambda(1-z))x - p\lambda(1-z)y)(1-B(x))(1-C(y))P_{(0,0)}, \qquad (5.35)$$

$$(v) P_{21}(z, x, y) = \frac{(1-r)\lambda\mu(1-q+qL_A(\lambda))(1-z)}{(1-q(1-L_A(\lambda))(1-z))K(z) - z(1-K(z))}$$
$$\times \exp(-G(p\lambda(1-z))x - p\lambda(1-z)y)(1-B(x))(1-C(y))P_{(0,0)}, \qquad (5.36)$$

$$(vi), P_3(z, x, \tau) = \frac{(1-r)\lambda\mu(1-q+qL_A(\lambda))(1-z)L_C(p\lambda(1-z))}{(1-q(1-L_A(\lambda))(1-z))K(z) - z(1-K(z))}$$
$$\times \exp(-G(p\lambda(1-z))x - (p\lambda(1-z)+\theta)\tau)(1-B(x))P_{(0,0)}, \qquad (5.37)$$

where

$$G(x) \equiv x + \mu - \mu\frac{rx+\theta}{x+\theta}L_C(x), \qquad (5.38)$$

$$K(x) \equiv L_B(G(p\lambda(1-x))).$$

Proof. By multiplying both sides of equations (5.20-(5.30) by z^i and summing over i, then we have

$$(\frac{\partial}{\partial w} + \lambda + \alpha(w))P_0(z, w) = 0, \qquad (5.39)$$

$$(\frac{\partial}{\partial x} + p\lambda + \mu + \beta(x))P_1(z, x) = \int_0^\infty P_{20}(z, x, y)\gamma(y)\, dy$$
$$+\theta\int_0^\infty P_3(z, x, \tau)\, d\tau + p\lambda z P_1(z, x), \qquad (5.40)$$

$$(\frac{\partial}{\partial y} + p\lambda + \gamma(y))P_{20}(z, x, y) = p\lambda z P_{20}(z, x, y), \qquad (5.41)$$

$$(\frac{\partial}{\partial y} + p\lambda + \gamma(y))P_{21}(z, x, y) = p\lambda z P_{21}(z, x, y), \qquad (5.42)$$

$$(\frac{\partial}{\partial \tau} + p\lambda + \theta)P_3(z, x, \tau) = p\lambda z P_3(z, x, \tau), \qquad (5.43)$$

$$P_0(z, 0) = \int_0^\infty P_1(z, x)\beta(x)\, dx - \lambda P_{(0,0)}, \qquad (5.44)$$

$$P_1(z,\,0) = \frac{\lambda(1-q+qz)}{z}\int_0^\infty P_0(z,\,w)\,dw$$
$$+\frac{1}{z}\int_0^\infty P_0(z,\,w)\alpha(w)\,dw + \lambda P_{(0,0)}, \tag{5.45}$$

$$P_{20}(z,\,x,\,0) = r\mu P_1(z,\,x), \tag{5.46}$$

$$P_{21}(z,\,x,\,0) = (1-r)\mu P_1(z,\,x), \tag{5.47}$$

$$P_3(z,\,x,\,0) = \int_0^\infty P_{21}(z,\,x,\,y)\gamma(y)\,dy, \tag{5.48}$$

$$P_{(0,0)} + \lim_{z\to 1-}\Big(\int_0^\infty P_0(z,\,w)\,dw + \int_0^\infty P_1(z,\,x)\,dx + \int_0^\infty\int_0^\infty P_{20}(z,\,x,\,y)\,dx\,dy$$
$$+\int_0^\infty\int_0^\infty P_{21}(z,\,x,\,y)\,dx\,dy + \int_0^\infty\int_0^\infty P_3(z,\,x,\,\tau)\,dx\,d\tau\Big) = 1. \tag{5.49}$$

From Equations (5.41)-(5.43), we get

$$P_{20}(z,\,x,\,y) = P_{20}(z,\,x,\,0)e^{-p\lambda(1-z)y}(1-C(y)), \tag{5.50}$$

$$P_{21}(z,\,x,\,y) = P_{21}(z,\,x,\,0)e^{-p\lambda(1-z)y}(1-C(y)), \tag{5.51}$$

$$P_3(z,\,x,\,\tau) = P_3(z,\,x,\,0)e^{-(p\lambda(1-z)+\theta)\tau}, \tag{5.52}$$

Substituting (5.46) into (5.50) and substituting (5.47) into (5.51), we have

$$P_{20}(z,\,x,\,y) = P_1(z,\,x)r\mu e^{-p\lambda(1-z)y}(1-C(y)), \tag{5.53}$$

$$P_{21}(z,\,x,\,y) = P_1(z,\,x)(1-r)\mu e^{-p\lambda(1-z)y}(1-C(y)). \tag{5.54}$$

Thus,

$$P_3(z,\,x,\,0) = (1-r)\mu L_C(p\lambda(1-z))P_1(z,\,x), \tag{5.55}$$

and therefore

$$P_3(z,\,x,\,\tau) = (1-r)\mu L_C(p\lambda(1-z))e^{-(p\lambda(1-z)+\theta)\tau}P_1(z,\,x). \tag{5.56}$$

As a result,

$$
\begin{aligned}
&(\frac{\partial}{\partial x} + p\lambda + \mu + \beta(x))P_1(z, x) = P_1(z, x) \\
&\times(\frac{\mu(rp\lambda(1-z)+\theta)L_C(p\lambda(1-z))}{p\lambda(1-z)+\theta} + p\lambda z),
\end{aligned} \tag{5.57}
$$

which yields

$$P_1(z, x) = e^{-G(p\lambda(1-z))x}(1 - B(x))P_1(z, 0). \tag{5.58}$$

and then

$$P_0(z, 0) = P_1(z, 0)L_B(G(p\lambda(1-z))) - \lambda P_{(0,0)}. \tag{5.59}$$

Since

$$P_0(z, w) = e^{-\lambda w}(1 - A(w))P_0(z, 0). \tag{5.60}$$

Then

$$P_1(z, 0) = \frac{1 - q(1-z)(1 - L_A(\lambda))P_0(z, 0)}{z} + \lambda P_{(0,0)}. \tag{5.61}$$

Therefore,

$$P_0(z, 0) = \frac{\lambda z(1 - K(z))P_{(0,0)}}{(1 - q(1 - L_A(\lambda))(1-z))K(z) - z(1 - K(z))}, \tag{5.62}$$

$$P_1(z, 0) = \frac{\lambda(1 - q + qL_A(\lambda))(1-z)P_{(0,0)}}{(1 - q(1 - L_A(\lambda))(1-z))K(z) - z(1 - K(z))}. \tag{5.63}$$

Then, we get (5.33)-(5.37). However, $P_{(0,0)}$ can be obtained by the use of the normalization equation.

This completes the proof.

5.2 Performance Measures

The previous theorem allows us to find some important performance measures of the model in steady state. The following corollary gives use results about the proportion of time the system is empty, the proportion of time that the system is nonempty and the server is idle. The sum of these two pieces gives the proportion of time that the server is idle. The corollary also gives the proportion of time that the server is busy, plus several other useful performance measures.

Corollary 5.1 If the system is stable, then

$$(i)P_{(0,0)} = \frac{1-q+qL_A(\lambda)-p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}{(1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda)}, \tag{5.64}$$

$$(ii)P_0 = \frac{(1-L_A(\lambda))p\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}{(1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda)}, \tag{5.65}$$

$$(iii)P_1 = \frac{(1-q+qL_A(\lambda))\lambda\beta_1}{(1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda)}, \tag{5.66}$$

$$(iv)P_{20} = \frac{r(1-q+qL_A(\lambda))\lambda\beta_1\mu\gamma_1}{(1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda)}, \tag{5.67}$$

$$(v)P_{21} = \frac{(1-r)(1-q+qL_A(\lambda))\lambda\beta_1\mu\gamma_1}{(1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda)}, \tag{5.68}$$

$$(vi)P_3 = \frac{(1-r)(1-q+qL_A(\lambda))\lambda\beta_1\mu}{\theta((1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda))}, \tag{5.69}$$

where the steady state $P_{(0,0)} \equiv P($ the system is empty$)$, $P_0 \equiv P($ the system is nonempty and the server is idle$)$, $P_1 \equiv P($ the server is busy$)$, $P_{20} \equiv P($ the server is under repair and the customer in service after server breakdown remains the service position$)$, $P_{21} \equiv P($ the server is under repair and the customer in service after server breakdown is in the retrial orbit$)$, $P_3 \equiv P($ the server is reserved$)$.

Proof. Note that

$$P_0 = \lim_{z\to 1-} \int_0^\infty P_0(z,\, w)\, dw, \quad P_1 = \lim_{z\to 1-} \int_0^\infty P_1(z,\, x)\, dx,$$

$$P_{20} = \lim_{z\to 1-} \int_0^\infty \int_0^\infty P_3(z,\, x,\, y)\, dx\, dy, \quad P_{21} = \lim_{z\to 1-} \int_0^\infty \int_0^\infty P_{21}(z,\, x,\, y)\, dx\, dy,$$

$$P_3 = \lim_{z\to 1-} \int_0^\infty \int_0^\infty P_3(z,\, x,\, \tau)\, dx\, d\tau.$$

By substituting the system steady state solution into the above formulas, we obtain the performance measures (i)-(vi).

Corollary 5.2 The long run proportion of customers who leave the system without service is:

$$\frac{(1-q+(q-p)L_A(\lambda))\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}{(1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda)}.$$

Proof. In steady state, let G denote the event that an arriving customer eventually leaves the system without service, let G_1 denote the event that the server is busy, under repair or reserved when the customer arrives at the system, and let G_2 denote the event that the customer enters the retrial queue. Recall that p is the probability that an arriving customer who encounters a busy server will enter the retrial queue. Then

$$P(G) = (1-p)P(G_1) + P(G_1)P(G_2 \mid G_1)P(G \mid G_2)$$

$$= (1-p)P(G_1) + pP(G_1)P(G \mid G_2)$$

However, the probability that the customer at the head of the retrial queue loses this competition is $1 - L_A(\lambda)$. Therefore,

$$P(G \mid G_2) = (1-q)(1-L_A(\lambda)) + q(1-L_A(\lambda))P(G \mid G_2)$$

or equivalently,

$$P(G \mid G_2) = \frac{(1-q)(1-L_A(\lambda))}{1-q(1-L_A(\lambda))}$$

This and the results of Corollary 5.1 directly yield $P(G)$, which is the long run proportion of customers who leave the system without service.

Corollary 5.3 The throughput to the system i(i.e. the rate of customers completing service) is the arrival rate times the long run proportion of customers who eventually receive service. So the throughput is

$$\lambda(1 - \frac{(1-q+(q-p)L_A(\lambda))\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1))}{(1-q+(q-p)L_A(\lambda))(1+\lambda\beta_1(1+\mu(\frac{1-r}{\theta}+\gamma_1)))+pL_A(\lambda)}).$$

Corollary 5.4 Let N_q and N be the number of customers in the retrial queue and in the system in steady state, respectively. Then the probability generating function of N_q and N are, respectively, given by

$$\begin{aligned} p_q(z) &\equiv E(z^{N_q}) \\ &= \frac{(1-q+qL_A(\lambda))(1-(1-p+pz)K(z))-pL_A(\lambda)z(1-K(z))}{(1-q(1-L_A(\lambda))(1-z))K(z)-z(1-K(z))}, \end{aligned} \tag{5.70}$$

$$\begin{aligned} p(z) &\equiv E(z^{N}) \\ &= \frac{(1-q+qL_A(\lambda))(z-(z-p+pz)K(z))-pL_A(\lambda)z(1-K(z))}{(1-q(1-L_A(\lambda))(1-z))K(z)-z(1-K(z))}. \end{aligned} \tag{5.71}$$

Proof. It can be shown that

$$p_q(z) = P_{(0,0)} + \int_0^\infty P_0(z, w)\, dw + \int_0^\infty P_1(z, x)\, dx$$

$$+\int_0^\infty\int_0^\infty P_{20}(z,\,x,\,y)\,dx\,dy+\int_0^\infty\int_0^\infty P_{21}(z,\,x,\,y)\,dx\,dy$$
$$+\int_0^\infty\int_0^\infty P_3(z,\,x,\,\tau)\,dx\,d\tau \tag{5.72}$$

and

$$p(z)=P_{(0,0)}+\int_0^\infty P_0(z,\,w)\,dw+z(\int_0^\infty P_1(z,\,x)\,dx$$
$$+\int_0^\infty\int_0^\infty P_{20}(z,\,x,\,y)\,dx\,dy+\int_0^\infty\int_0^\infty P_{21}(z,\,x,\,y)\,dx\,dy$$
$$+\int_0^\infty\int_0^\infty P_3(z,\,x,\,\tau)\,dxd\tau). \tag{5.73}$$

These are solved using the steady state solutions of Theorem 5.1.

Corollary 5.5 Let N_q be the number of customers in the Model II retrial queue in steady state. Then

$$E(N_q)=p\lambda\beta_1(1-\mu\frac{r-1}{\theta}+\mu\gamma_1)((p-q)L_A(\lambda)-1+q)-p(1-q)-pqL_A(\lambda).$$

5.3 The waiting process

The following theorem gives the joint distribution of the waiting time that a primary customer spends in the retrial queue and the number of customers served during the waiting time.

Theorem 5.2 In steady state, let W represent the waiting time that a primary customer spends in the retrial queue and N represent the number of customers served during this waiting time. Then we obtain the joint distribution of W and N in terms of the Laplace transform

$$\Phi(s,\, z) \equiv E(e^{-sW} z^{N}) = 1 - p + p(P_{(0,0)} + P_0)$$
$$+\frac{p\lambda(1-q+qL_A(\lambda))z(1-Q^*(s,\, z))Q(s,\, z)P_{(0,0)}}{(1-q(1-L_A(\lambda))(1-Q^*(s,\, z))K(Q^*(s,\, z)) - Q^*(s,\, z)(1-K(Q^*(s,\, z)))}$$
$$\times\frac{L_B(G(s)) - K(Q^*(s,\, z))}{p\lambda(1-Q^*(s,\, z))) - s}, \tag{5.74}$$

where

$$Q(s,\, z) = \frac{(1-q)\lambda + (s+q\lambda)L_A(s+\lambda)}{s+\lambda - q\lambda(1-L_A(s+\lambda))zL_B(G(s))}, \tag{5.75}$$

and $Q^*(s,\, z) \equiv zL_B(G(s))Q(s,\, z)$.

Proof. Note that when a primary customer PC arrives, the system is empty, or the system is nonempty and the server is idle, or the server is busy, or the server is under repair and the customer in service remains in the service position during server repair, or the server is under repair and the customer in service enters the retrial queue during server repair. Therefore,

$$\Phi(s,\, z) \equiv E(e^{-sW} z^{N})$$
$$= P_{(0,0)} + \sum_{i=1}^{\infty} \int_0^{\infty} P_{(0,\,i)}(w)\, dw$$
$$+ \sum_{i=0}^{\infty} \int_0^{\infty} E(e^{-sW} z^{N} \mid C = 1,\, Q = i,\, \xi_1 = x) P_{(1,\,i)}(x)\, dx$$
$$+ \sum_{i=0}^{\infty} \int_0^{\infty}\int_0^{\infty} E(e^{-sW} z^{N} \mid C = 2,\, C^* = 0,\, Q = i,\, \xi_1 = x,$$
$$\xi_2 = y) P_{(2,0,\,i)}(x,\, y)\, dx\, dy + \sum_{i=0}^{\infty} \int_0^{\infty}\int_0^{\infty} E(e^{-sW} z^{N} \mid C = 2,$$
$$C^* = 1,\, Q = i,\, \xi_1 = x,\, \xi_2 = y) P_{(2,1,\,i)}(x,\, y)\, dx\, dy$$
$$+ \sum_{i=0}^{\infty} \int_0^{\infty}\int_0^{\infty} E(e^{-sW} z^{N} \mid C = 3,\, C^* = 1,\, Q = i,\, \xi_1 = x,$$

$$\xi_3 = \tau)P_{(3,i)}(x, \tau)\, dx\, d\tau, \tag{5.76}$$

where C, C^*, Q, ξ_1, ξ_2 and ξ_3 denote, respectively, the server state, the state of the customer in service after the server breakdown, the number of customers in the retrial queue, the elapsed service time, the elapsed repair time and the elapsed reserved time when PC arrives to the system in steady state.

Now let us consider the state of the system when PC arrives. Suppose that there are i customers in the retrial queue. If the server is busy, under repair and the customer in service remains in the service position or enters the retrial queue or is reserved, then PC enters the retrial queue and becomes the $(i+1)$st retrial customer or leaves the system. Let $\zeta_1 = 1$ if PC enters the retrial queue and $\zeta_1 = 0$ if PC leaves the system. Thus we have

$$\begin{aligned} &E(e^{-sW}z^N \mid C = 1,\, Q = i,\, \xi_1 = x) \\ &= 1 - p + pE(e^{-sW}z^N \mid C = 1,\, Q = i,\, \xi_1 = x,\, \zeta_1 = 1), \end{aligned} \tag{5.77}$$

$$\begin{aligned} &E(e^{-sW}z^N \mid C = 2,\, C^* = 0,\, Q = i,\, \xi_1 = x,\, \xi_2 = y) \\ &= 1 - p + pE(e^{-sW}z^N \mid C = 2,\, C^* = 0,\, Q = i,\, \xi_1 = x,\, \xi_2 = y,\, \zeta_1 = 1), \end{aligned} \tag{5.78}$$

$$\begin{aligned} &E(e^{-sW}z^N \mid C = 2,\, C^* = 1,\, Q = i,\, \xi_1 = x,\, \xi_2 = y) \\ &= 1 - p + pE(e^{-sW}z^N \mid C = 2,\, C^* = 1,\, Q = i,\, \xi_1 = x,\, \xi_2 = y,\, \zeta_1 = 1), \end{aligned} \tag{5.79}$$

$$E(e^{-sW}z^N \mid C = 3,\, C^* = 1,\, Q = i,\, \xi_1 = x,\, \xi_3 = \tau)$$

$$= 1 - p + pE(e^{-sW}z^N \mid C = 3,\ C^* = 1,\ Q = i,\ \xi_1 = x,\ \xi_3 = \tau,\ \zeta_1 = 1). \tag{5.80}$$

In order to calculate $E(e^{-sW}z^N \mid C = 1,\ Q = i,\ \xi_1 = x,\ \zeta_1 = 1)$, $E(e^{-sW}z^N \mid C = 2,\ C^* = 0,\ Q = i,\ \xi_1 = x,\ \xi_2 = y,\ \zeta_1 = 1)$, $E(e^{-sW}z^N \mid C = 2,\ C^* = 1,\ Q = i,\ \xi_1 = x,\ \xi_2 = y,\ \zeta_1 = 1)$ and $E(e^{-sW}z^N \mid C = 3,\ C^* = 1,\ Q = i,\ \xi_1 = x,\ \xi_3 = \tau,\ \zeta_1 = 1)$, we employ the remaining service time when the server is busy, the remaining service time and the remaining repair time when the customer in service remains in the service position and the remaining service time and the remaining repair time or the remaining reserved time when the customer in service enters the retrial queue. In addition, we consider the number of failures during the remaining service time. Thus, recalling the definition of S^i from Chapter 4, we get

$$\begin{aligned}
&E(e^{-sW}z^N \mid C = 1,\ Q = i,\ \xi_1 = x,\ \zeta_1 = 1)\\
&= \sum_{k=0}^{\infty} Q^{(i+1)}(s,\ z)E(e^{-s(S^{(1)}+S^{(2)}+\dots+S^{(i)})}z^i)\frac{1}{1-B(x)}\\
&\times \int_x^\infty e^{-su}z(\frac{rs+\theta}{s+\theta}L_C(s))^k\frac{(\mu u)^k}{k!}e^{-\mu u}b(u+x)\,du,
\end{aligned} \tag{5.81}$$

$$\begin{aligned}
&E(e^{-sW}z^N \mid C = 2,\ C^* = 0,\ Q = i,\ \xi_1 = x,\ \xi_2 = y,\ \zeta_1 = 1)\\
&= \sum_{k=0}^{\infty} Q^{(i+1)}(s,\ z)E(e^{-s(S^{(1)}+S^{(2)}+\dots+S^{(i)})}z^i)\frac{1}{(1-B(x))(1-C(x))}\\
&\times \int_x^\infty\int_y^\infty e^{-s(u+v)}(\frac{rs+\theta}{s+\theta}L_C(s))^k\frac{(\mu u)^k}{k!}e^{-\mu u}b(u+x)b(u)c(v)\,du\,dv,
\end{aligned} \tag{5.82}$$

$$\begin{aligned}
&E(e^{-sW}z^N \mid C = 2,\ C^* = 1,\ Q = i,\ \xi_1 = x,\ \xi_2 = y,\ \zeta_1 = 1)\\
&= \sum_{k=0}^{\infty} Q^{(i+1)}(s,\ z)E(e^{-s(S^{(1)}+S^{(2)}+\dots+S^{(i)})}z^i)\frac{1}{(1-B(x))(1-C(x))}
\end{aligned}$$

$$\times \int_x^\infty \int_y^\infty e^{-s(u+v)} z \frac{\theta}{s+\theta} (\frac{rs+\theta}{s+\theta} L_C(s))^k \frac{(\mu u)^k}{k!} e^{-\mu u} b(u+x) c(y+v)\, du\, dv,$$

(5.83)

$$E(e^{-sW} z^N \mid C=3,\, C^*=1,\, Q=i,\, \xi_1=x,\, \xi_3=\tau,\, \zeta_1=1)$$
$$= \sum_{k=0}^\infty Q^{(i+1)}(s,\, z) E(e^{-s(S^{(1)}+S^{(2)}+\dots+S^{(i)})} z^i) \frac{1}{1-B(x)} \int_x^\infty e^{-su} z$$
$$\times \frac{\theta}{s+\theta} (\frac{rs+\theta}{s+\theta} L_C(s))^k \frac{(\mu u)^k}{k!} e^{-\mu u} b(u+x) b(u)\, du, \qquad (5.84)$$

where $Q^{(i+1)}(s,\, z)$ is defined below. Recall that if primary customer PC enters the retrial queue, it sees i customers in the retrial queue. Consider the customer at the head of this retrial queue. Let $W^{(1)}$ represent the waiting time that the customer at the head of the retrial queue spends in the retrial queue measured from the next service completion time point. Let $W^{(j)}$, $j = 2, 3, \dots, i+1$, represent the waiting time that the jth retrial customer spends at the head of the retrial queue. Let $N^{(j)}$ represent the number of customers served during $W^{(j)}$. Define $Q^{(i+1)}(s,\, z) = E(e^{s(W^{(1)}+W^{(2)}+S^{(2)}+\dots+W^{(i)}+W^{(i+1)}} z^{N^{(1)}+N^{(2)}+\dots+N^{(i)}+N^{(i+1)}})$. According to the model, we see that $(W^{(1)},\, N^{(1)}), (W^{(2)},\, N^{(2)}), \dots, (W^{(i+1)},\, N^{(i+1)})$ are independent and identically distributed with distribution, in terms of the Laplace transform,

$$Q(s,\, z) \equiv E(e^{-sW_1^{(q)}} z^{N_1^{(q)}})$$
$$= \int_0^\infty \int_y^\infty \lambda e^{-\lambda x - sy}\, dA(y)\, dx$$
$$+ \int_0^\infty \int_0^y (1-q+qzL_B(G(s))Q(s,z)) \lambda e^{-\lambda x - sx}\, dA(y)\, dx$$
$$= L_A(s+\lambda) + \frac{(1-q)\lambda}{s+\lambda}(1-L_A(s+\lambda))$$
$$+ \frac{q\lambda z}{s+\lambda}(1-L_A(s+\lambda)) L_B(G(s)) Q(s,\, z), \qquad (5.85)$$

which yields

$$Q(s,\, z) = \frac{(1-q)\lambda + (s+q\lambda)L_A(s+\lambda)}{s+\lambda - q\lambda(1-L_A(s+\lambda))zL_B(G(s))}. \tag{5.86}$$

Thus, (5.77)-(5.80) can be rewritten as

$$\begin{aligned} &E(e^{-sW}z^N \mid C=1,\, Q=i,\, \xi_1=x) \\ &= 1-p+\frac{pzQ(s,z)Q^*(s,z)^i}{1-B(x)}\int_x^\infty e^{-G(s)(u-x)}b(u)\,du, \end{aligned} \tag{5.87}$$

$$\begin{aligned} &E(e^{-sW}z^N \mid C=2,\, C*=0,\, Q=i,\, \xi_1=x,\, \xi_2=y) \\ &= 1-p+\frac{pzQ(s,\, z)Q^*(s,\, z)^i}{(1-B(x))(1-C(y))}\int_x^\infty\int_y^\infty e^{-G(s)(u-x)-s(v-y)}b(u)c(v)\,du\,dv, \end{aligned} \tag{5.88}$$

$$\begin{aligned} &E(e^{-sW}z^N \mid C=2,\, C^*=1,\, Q=i,\, \xi_1=x,\, \xi_2=y) \\ &= 1-p+\frac{p\theta zQ(s,\, z)Q^*(s,\, z)^i}{(s+\theta)(1-B(x))(1-C(y))}\int_x^\infty\int_y^\infty e^{-G(s)(u-x)-s(v-y)}b(u)c(v)\,du\,dv, \end{aligned} \tag{5.89}$$

$$\begin{aligned} &E(e^{-sW}z^N \mid C=3,\, C^*=1,\, Q=i,\, \xi_1=x,\, \xi_3=\tau) \\ &= 1-p+\frac{p\theta zQ(s,\, z)Q^*(s,\, z)^i}{(s+\theta)(1-B(x))}\int_x^\infty e^{-G(s)(u-x)}b(u)\,du, \end{aligned} \tag{5.90}$$

where $Q^*(s,\, z) \equiv zL_B(G(s))Q(s,\, z)$.

Therefore,

$$\begin{aligned} &\Phi(s,\, z) = P_{(0,0)} + \lim_{z\to 1-}\int_0^\infty P_0(z,\, w)\,dw \\ &+(1-p)\lim_{z\to 1-}\Big(\int_0^\infty P_1(z,\, x)\,dx + \int_0^\infty\int_0^\infty P_{20}(z,\, x,\, y)\,dx\,dy \\ &+\int_0^\infty\int_0^\infty P_{21}(z,\, x,\, y)\,dx\,dy + \int_0^\infty\int_0^\infty P_3(z,\, x,\, \tau)\,dx\,d\tau\Big) \end{aligned}$$

$$+\lim_{z\to 1-} pzQ(s,z)\Big(\int_0^\infty \frac{P_1(Q^*(s,\,z)\,,x)}{1-B(x)}\int_x^\infty e^{-G(s)(u-x)}b(u)\,du\,dx$$
$$+\int_0^\infty\int_0^\infty \frac{P_{20}(Q^*(s,\,z),\,x,\,y)}{(1-B(x))(1-C(y))}\int_x^\infty\int_y^\infty e^{-G(s)(u-x)-s(v-y)}$$
$$\times b(u)c(v)\,du\,dv\,dx\,dy$$
$$+\frac{\theta}{s+\theta}\int_0^\infty\int_0^\infty \frac{P_{21}(Q^*(s,\,z),\,x,\,y)}{(1-B(x))(1-C(y))}\int_x^\infty\int_y^\infty e^{-G(s)(u-x)-s(v-y)}$$
$$\times b(u)c(v)\,du\,dv\,dx\,dy$$
$$+\frac{\theta}{s+\theta}\int_0^\infty\int_0^\infty \frac{P_3(Q^*(s,z),x,\tau)}{(1-B(x))}\int_x^\infty e^{-G(s)(u-x)}b(u)\,du\,dx\,d\tau\Big), \tag{5.91}$$

which yields the desired result by substituting the steady state results of Theorem 5 into (5.91).

5.4 The busy period

A (primary) busy period is defined as the time period from when an primary customer arrives and finds the system empty until the system is empty again. Such a primary customer is called IPC (an initial customer). Any customer who arrives during the generalized service time of IPC and chose to enter retrial queue is called an IRC (an initial retrial customer).

Suppose that there are k IRCs, say Λ_i, $i = 1, 2, ..., k$, in the system. Let t_0 denote the departure point of IPC and t_i denote the departure point of Λ_i. We call $c_i \equiv t_i - t_{i-1}$ the effective waiting time of Λ_i. This effective waiting time can consist of at least one idle time, and at least one service time depending on whether or not primary customers also enter the system. Suppose that there are n_i primary customers, say Λ_{ij}, $j = 1, 2, ..., n_i$, arriving during the effective waiting time of Λ_i and let t_{ij} denote the departure point of Λ_{ij}. We call $c_{ij} \equiv t_{ij} - t_{ij-1}$ the effective waiting time of Λ_{ij}, where $\Lambda_{10} \equiv t_k$. We continue similarly. We call Λ_{ij} the first generation offspring of Λ_i. We define primary customers arriving during the effective waiting time of Λ_i as the first generation offspring of Λ_i. Similarly we can define second generation offspring of Λ_i, third generation offspring of Λ_i and so on. We define Λ_i and all its generations of offspring as a family from Λ_i and define the effective waiting time of all members of a family from Λ_i as the ith retrial busy period.

It can be seen that a busy period in the retrial queue is different from that in an ordinary queueing system. The server is always busy in a busy period of the ordinary queueing system, while the server alternates between a busy state and an idle state (with the retrial queue nonempty) in a busy period of the retrial queue.

Theorem 5.3 Let T represent the length of a busy period and N represent the number of customers served during the busy period in steady state. Then the joint distribution of T and N, in terms of the generating function Laplace transform, is

$$\Psi(s,z) \equiv E(e^{-sT}z^N) = zL_B(G(s+\lambda-\lambda(1-p+p\Psi^{(q)}(s,z)))), \tag{5.92}$$

where

$$\Psi^{(q)}(s,z) \equiv \frac{(1-q)\lambda+(s+q\lambda)L_A(s+\lambda))\Psi(s,z)}{s+\lambda-q\lambda(1-L_A(s+\lambda))\Psi(s,z)}. \tag{5.93}$$

Proof. Let $S^{(*)}$ denote the generalized service time of IPC. If no primary customers arrive during $(0, S^{(*)})$, then $T = S^{(*)}$ and $N = 1$. Otherwise, if n primary customers arrive but only k customers become IRC, then $T = S^{(*)} + T_1^{(q)} + T_2^{(q)} + ... + T_k^{(q)}$ and $N = 1 + N_1^{(q)} + N_2^{(q)} + ... + N_k^{(q)}$, where $T_j^{(q)}$ and $N_j^{(q)}$, represent the length of the ith retrial busy period and the number of customers served during it, respectively, for $j = 1, 2, ..., k$.

Therefore,

$$\Psi(s,\, z) = \int_0^\infty E(e^{-sT} z^N \mid S^{(*)} = x)\, dB^*(x)$$
$$= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \int_0^\infty \frac{(\lambda x)^n}{n!} e^{-\lambda x} \tbinom{n}{k} p^k (1-p)^{n-k} E(e^{-sT} z^N \mid$$
$$S^{(*)} = x,\, N(x) = n,\, N^{(q)}(x) = k)\, dB^*(x)$$
$$= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \int_0^\infty \frac{(\lambda x)^n}{n!} e^{-\lambda x} \tbinom{n}{k} p^k (1-p)^{n-k} e^{-sx} z$$
$$\times E(e^{-s(T_1^{(q)}+T_2^{(q)}+...+T_k^{(q)})} z^{N_1^{(q)}+N_2^{(q)}+...+N_k^{(q)}})\, dB^*(x), \tag{5.94}$$

where $T_0^{(q)} = 0$ and $N_0^{(q)} = 0$.

According to the characteristics of the model, it can be seen that the primary arrival process is independent of the retrials and service and that $T_i^{(q)}$ consists of the sum of disjoint time intervals. Suppose that only one retrial customer is in the retrial queue during the generalized service time of the initial customer and let $T_*^{(q)}$ correspond to the length of the retrial busy period and $N_*^{(q)}$ correspond to the number of customers served during it. It can be seen that $(T_1^{(q)},\, N_1^{(q)})$, $(T_2^{(q)},\, N_2^{(q)})$, ..., $(T_k^{(q)},\, N_k^{(q)})$ are independent and have the same distribution as $(T_*^{(q)},\, N_*^{(q)})$ so

$$E(e^{-s(T_1^{(q)}+T_2^{(q)}+...+T_k^{(q)})} z^{N_1^{(q)}+N_2^{(q)}+...+N_k^{(q)}}) = \left(E(e^{-sT_*^{(q)}} z^{N_*^{(q)}})\right)^k. \tag{5.95}$$

Thus, equation (5.94) can be rewritten as

$$\Psi(s,\, z) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} \int_0^\infty \frac{(\lambda x)^n}{n!} e^{-\lambda x} \tbinom{n}{k} p^k (1-p)^{n-k} e^{-sx} z (\Psi^{(q)}(s,\, z))^k\, dB^*(x), \tag{5.96}$$

where $\Psi^{(q)}(s,\, z) = E(e^{-sT_*^{(q)}} z^{N_*^{(q)}})$.

In order to get the joint distribution of $T_*^{(q)}$ and $N_*^{(q)}$ in terms of the generating function Laplace transform $\Psi^{(q)}(s,z)$, we must analyze the evolution of the process from the time that IPC leaves the system. There exists a competition between an arriving primary customer and the first IRC from the time that IPC leaves the system. If the first IRC wins this competition, then $T_*^{(q)}$ corresponds to the retrial time plus a random variable which has the same distribution as T, and $N_*^{(q)}$ corresponds to a random variable which has the same distribution as N. Otherwise, if an arriving primary customer wins and the first IRC chooses to return to the retrial queue, $T_*^{(q)}$ corresponds to the arriving time of the primary customer plus the sum of two independent random variables, of which one has the same distribution as T and the other has the same distribution as $T_*^{(q)}$. Here $N_*^{(q)}$ is equal to the sum of two random variables, one of which has the same distribution as N and the other has the same distribution as $N_*^{(q)}$. If an arriving primary customer wins and the first IRC chooses to leave the system, $T_*^{(q)}$ corresponds to the arriving time of the primary customer plus one random variable, which has the same distribution as T, and $N_*^{(q)}$ corresponds to a random variable which has the same distribution as N. Thus we have

$$\Psi^{(q)}(s,\, z) = \int_0^\infty \int_y^\infty \lambda e^{-\lambda x - sy}\Psi(s,\, z)\, dA(y)\, dx$$
$$+ \int_0^\infty \int_0^y \lambda e^{-\lambda x - sx}((1-q)\Psi(s,\, z) + q\Psi(s,\, z)\Psi^{(q)}(s,z))\, dA(y)\, dx. \tag{5.97}$$

However,

$$\int_0^\infty \int_y^\infty \lambda e^{-\lambda x - sy}\Psi(s,\, z)\, dA(y)\, dx = L_A(s+\lambda). \tag{5.98}$$

Accordingly, (5.97) can be rewritten as

$$\Psi^{(q)}(s,\, z) = (L_A(s+\lambda) + \frac{(1-q)\lambda}{s+\lambda}(1 - L_A(s+\lambda))\Psi(s,\, z)$$

$$+\frac{q\lambda}{s+\lambda}(1-L_A(s+\lambda))\Psi(s,z)\Psi^{(q)}(s,\ z), \tag{5.99}$$

or equivalently,

$$\Psi^{(q)}(s,\ z)=\frac{(1-q)\lambda+(s+q\lambda)L_A(s+\lambda))\Psi(s,\ z)}{s+\lambda-q\lambda(1-L_A(s+\lambda))\Psi(s,\ z)},$$

which is substituted into equation (5.96) and yields the desired result.

This completes the proof.

Chapter 6

Model III Description and Stability

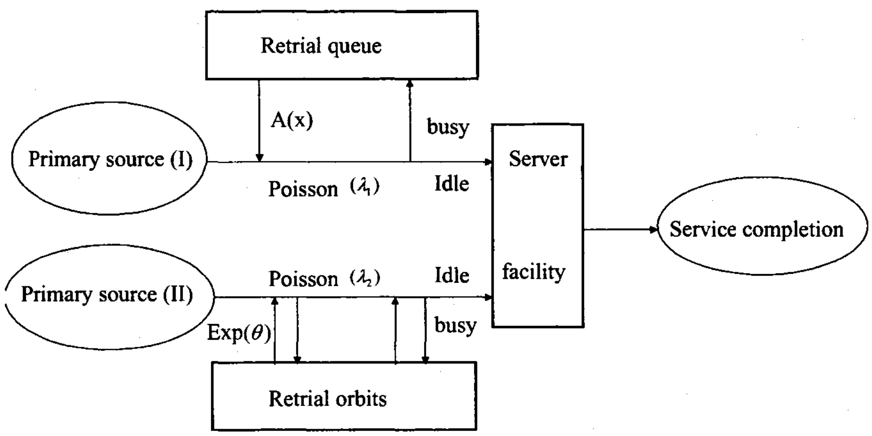
6.1 Description of model III

An $M/G/1$ queue with general retrial times, two primary sources and one retrial queue and one collection of retrial orbits, with different disciplines is investigated. The description of model III is as follows. We consider a single-server queueing system with two primary sources (called primary source I and primary source II) and no waiting room in front of the server. We assume that primary customers from primary source I arrive in a Poisson process with rate λ_1 and primary customers from primary source II arrive in a Poisson process with rate λ_2. If an arriving primary customer finds the server idle, that customer begins service immediately and leaves the system after the service completion. If an arriving primary customer from primary source I finds the server busy, the customer enters a retrial queue according to a FCFS discipline. For customers in the retrial queue, only the customer at the head of the retrial queue is allowed to try for service in a random time measured from the moment that the server becomes idle. It is required to cancel the demand for service if a primary customer (or a source II customer from the retrial orbits) arrives first and the retrial customer then returns to the position at the head of the retrial queue. The retrial times in the retrial queue are distributed with distribution function $A(x)$, density function $a(x)$, Laplace transform $L_A(s)$ and first two moments α_1 and α_2.

If an arriving primary customer from primary source II finds the server busy, that customer enters its own retrial orbit (not a retrial queue) and keeps making retrials at random times, measured from when the server becomes idle, until the customer eventually obtains service. All such source II retrial customers form independent retrial sources each with its own retrial orbit. The retrial times in the retrial orbits are each exponentially distributed with rate θ.

The service time for a customer from primary source I or from the retrial queue have distribution function $B_1(x)$, density function $b_1(x)$, Laplace transform $L_{B_1}(s)$ and the first two moments $\beta_1^{(1)}$, $\beta_2^{(1)}$. The service times for customers from primary source (II) or from the retrial orbits have distribution function $B_2(x)$, density function $b_2(x)$, Laplace transform $L_{B_2}(s)$ and first two moments $\beta_1^{(2)}$, $\beta_2^{(2)}$. All times involved are mutually independent. The model is illustrated in Figure 6.1.

Figure 6.1: Outline of the M/G/1 queue with two primary sources a retrial queue and retrial orbits.



Remark 6.1 Model III generalizes Models I and II subject to certain restrictions. If we assume that $\lambda_2 = 0$ and $\lambda_1 \neq 0$, then we get the earlier models with no retrials of customers in service, no server breakdown, no opportunity for early exit from the system. We could include these extra conditions but a discussion of a model containing both a retrial queue and retrial orbits does not seem to exist in queueing literature, so the simple model is the best model to consider. Furthermore, the "simple" model con-

sidered here is not simple at all. Rather, it is very complex. Note that if $\lambda_1 = 0$ and $\lambda_2 \neq 0$, then we get the most common type of retrial queueing system as a special case.

Example 6.1 To illustrate Model III, we continue with our example of a call center with an answering machine, as is used at a university computer help system. A customer phones the help system for assistance. If the server is busy, then the customer leaves a message on the answering machine and waits for the server (consultant) to phone back. These customers are waiting in a retrial queue. This makes sense for customers working in an office on campus where the computers are connected without using the phone lines. However, many customers work at home and connect to the university computer through their phone lines. It does not make sense to leave a phone message and wait for a phone call because such customers would not longer be able to use the phone lines for their computer connection. As a result, these customers occasionally disconnect their computer from the phone line and phone for service, hoping to access an idle server. These types of customers form retrial orbits. Thus the computer call center is dealing with both types of customer. It is also reasonable to expect that the arrival rates and service rates of the two type of customer could be different.

6.2 Evolution of the Model III and its states

From the description of the model, it is readily seen that the queueing process develops as follows. Suppose that customers are numbered by their order of service and that the $(i-1)$th customer completes service at time d_{i-1} and leaves the system so the server then is idle. Therefore, the ith customer begins service after a idle time κ_i by the assumptions of the model. At time d_{i-1}, if the retrial queue is empty and j customers are in the retrial orbits, there exists a competition between a primary customer from primary source I, a primary customer from primary source II and all retrial customers from retrial orbits before the beginning of the ith customer's service. That is, κ_i is equal to the minimum of three times, one of which is exponentially distributed with rate λ_1, one of which is exponentially distributed with rate λ_2 and the other is exponentially distributed with rate $j\theta$. It can be shown that the probability that the ith customer is a primary customer from primary source I is equal to $\frac{\lambda_1}{\lambda_1+\lambda_2+j\theta}$ and the probability the ith customer is a primary customer from primary source II is equal to $\frac{\lambda_2}{\lambda_1+\lambda_2+j\theta}$ and the probability the ith customer is a retrial customer II from a retrial orbit is equal to $\frac{j\theta}{\lambda_1+\lambda_2+j\theta}$. Otherwise, if the retrial queue is nonempty and j customers are in the retrial orbit, there is a competition between a primary customer from primary source I, a primary customer from primary source II, a retrial customer II from the retrial orbits, and a retrial customer I from the retrial queue, before the ith customer begins to be served. Therefore, κ_i is equal to the minimum of four times, one of which is exponentially distributed with rate λ_1, one of which is exponentially distributed with rate λ_2, one of which is generally distributed with Laplace transform $L_A(s)$ and the other is exponentially distributed with rate $j\theta$. The probability that the ith customer is a primary customer from primary source I is equal to $\frac{\lambda_1(1-L_A(\lambda_1+\lambda_2+j\theta))}{\lambda_1+\lambda_2+j\theta}$, the probability the ith customer is a primary customer from primary source II is $\frac{\lambda_2(1-L_A(\lambda_1+\lambda_2+j\theta))}{\lambda_1+\lambda_2+j\theta}$, the probability the ith customer is a repeated customer I from the retrial queue is equal to $L_A(\lambda_1+\lambda_2+j\theta)$, and the probability the ith customer is a retrial customer II from a retrial orbit is equal to $\frac{j\theta(1-L_A(\lambda_1+\lambda_2+j\theta))}{\lambda_1+\lambda_2+j\theta}$. At time $d_{i-1}+\kappa_i$, the ith customer begins to be served and continues for a busy time S_i, distributed with distribution function $B_1(x)$ or $B_2(x)$. At time $d_i = d_{i-1}+\kappa_i+S_i$, the server is idle again. After that, the queueing process goes on similarly.

The state of the system at time t can be described by the Markov process $\{X(t);\, t \geq 0\} = \{(J(t),\, Q_I(t),\, Q_{II}(t),\, \xi(t),\, \eta_1(t),\, \eta_2(t)); t \geq 0\}$ where at time t, (i). $J(t) = 0$ if the server is idle. (ii). $J(t) = 1$ if the server is busy with a customer from primary source (I) or from the retrial queue of type (I) customers. (iii). $J(t) = 2$ if the server is busy with a customer from primary source (II) or from the retrial orbits of type II customers. $Q_I(t)$ denotes the number of customers in the retrial queue and $Q_{II}(t)$ denotes the number of customers in the retrial orbits. If $J(t) = 0$ and $Q_I(t) > 0$, then $\xi(t)$ represents the elapsed retrial time for a customer in the retrial queue. If $J(t) = 1$, then $\eta_1(t)$ represents the elapsed service time for a type (I) customer. If $J(t) = 2$, then $\eta_2(t)$ represents the elapsed service time for a type (II) customer. The functions $\alpha(x), \beta_1(x)$ and $\beta_2(x)$ are the conditional completion rates for the retrial customers from the retrial queue, the service time for a type (I) customer and the service time for a type (II) customer, respectively, i.e., $\alpha(x) = a(x)(1-A(x))^{-1}$, $\beta_1(x) = b_1(x)(1-B_1(x))^{-1}$, and $\beta_2(x) = b_2(x)(1-B_2(x))^{-1}$.

6.3 Sufficient Condition for System Stability

Theorem 6.1 If $\lambda_1\beta_1^{(1)} + \lambda_2\beta_1^{(2)} < L_A(\lambda_1 + \lambda_2)$, then the system is stable.

Proof. Let d_n be the nth departure point and let $J^{(n)} = J(d_n - 0)$, $Q_I^{(n)} = Q_I(d_n - 0)$, and $Q_{II}^{(n)} = Q_{II}(d_n - 0)$. Then $\{(J^{(n)}, Q_I^{(n)}, Q_{II}^{(n)}); n \geq 0\}$ is a Markov chain with state space $S = \{1, 2\} \times Z_+^2$. It is readily seen that the Markov chain is irreducible and aperiodic. We first define a function $f(k, i, j) = c_1 i + c_2 j$ where c_1 and c_2 are two constants, $(k, i, j) \in S = \{1, 2\} \times Z_+^2$.

Next, we calculate the mean drift Δ_{kij}, for $k = 1$, 2, $i \geq 0$, $j \geq 0$, as follows.

(1) For $k = 1$ or 2, $i = 0$ and $j \geq 0$, define

$$\begin{aligned}
&\Delta_{k0j} \equiv E(f(J^{(n+1)},\, Q_I^{(n+1)},\, Q_{II}^{(n+1)}) - f(J^{(n)},\, Q_I^{(n)},\, Q_{II}^{(n)}) \\
&\mid (J^{(n+1)},\, Q_I^{(n)},\, Q_{II}^{(n)}) = (k,\, 0,\, j)) \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2 + j\theta}(\lambda_1\beta_1^{(1)} c_1 + \lambda_2\beta_1^{(1)} c_2) + \frac{\lambda_2}{\lambda_1 + \lambda_2 + j\theta}(\lambda_1\beta_1^{(2)} c_1 + \lambda_2\beta_1^{(2)} c_2) \\
&+ \frac{j\theta}{\lambda_1 + \lambda_2 + j\theta}(\lambda_1\beta_1^{(2)} c_1 + (\lambda_2\beta_1^{(2)} - 1)c_2) \\
&= (1 - \lambda_1\beta_1^{(1)} + \lambda_1\beta_1^{(2)})c_1 - (1 - \lambda_2\beta_1^{(2)} + \lambda_2\beta_1^{(1)})c_2 - (1 - \lambda_1\beta_1^{(1)})c_1 + \lambda_2\beta_1^{(1)} c_2 \\
&- \frac{1}{\lambda_1 + \lambda_2 + j\theta}(\lambda_1((1 - \lambda_1\beta_1^{(1)} + \lambda_1\beta_1^{(2)})c_1 - (1 - \lambda_2\beta_1^{(2)} + \lambda_2\beta_1^{(1)})c_2) - \lambda_1 c_1 - \lambda_2 c_2).
\end{aligned} \tag{6.1}$$

(2) For $k = 1$ or 2, $i \geq 1$ and $j \geq 0$, we have

$$\begin{aligned}
&\Delta_{kij} \equiv E(f(J^{(n+1)},\, Q_I^{(n+1)},\, Q_{II}^{(n+1)}) - f(J^{(n)},\, Q_I^{(n)},\, Q_{II}^{(n)}) \\
&\mid (J^{(n+1)},\, Q_I^{(n)},\, Q_{II}^{(n)}) = (k,\, i,\, j)) \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2 + j\theta}(1 - L_A(\lambda_1 + \lambda_2 + j\theta))(\lambda_1\beta_1^{(1)} c_1 + \lambda_2\beta_1^{(1)} c_2) \\
&+ \frac{\lambda_2}{\lambda_1 + \lambda_2 + j\theta}(1 - L_A(\lambda_1 + \lambda_2 + j\theta))(\lambda_1\beta_1^{(2)} c_1 + \lambda_2\beta_1^{(2)} c_2)
\end{aligned}$$

$$+\frac{j\theta}{\lambda_1+\lambda_2+j\theta}(1-L_A(\lambda_1+\lambda_2+j\theta))(\lambda_1\beta_1^{(2)}c_1+(\lambda_2\beta_1^{(2)}-1)c_2)$$
$$+L_A(\lambda_1+\lambda_2+j\theta)((\lambda_1\beta_1^{(1)}-1)c_1+\lambda_2\beta_1^{(1)}c_2)$$
$$=(1-\lambda_1\beta_1^{(1)}+\lambda_1\beta_1^{(2)})c_1-(1-\lambda_2\beta_1^{(2)}+\lambda_2\beta_1^{(1)})c_2-(1-\lambda_1\beta_1^{(1)})c_1+\lambda_2\beta_1^{(1)}c_2$$
$$-\frac{1-L_A(\lambda_1+\lambda_2+j\theta)}{\lambda_1+\lambda_2+j\theta}(\lambda_1((1-\lambda_1\beta_1^{(1)}+\lambda_1\beta_1^{(2)})c_1$$
$$-(1-\lambda_2\beta_1^{(2)}+\lambda_2\beta_1^{(1)})c_2)-\lambda_1c_1-\lambda_2c_2)$$
$$+L_A(\lambda_1+\lambda_2+j\theta)((1-\lambda_1\beta_1^{(1)}+\lambda_1\beta_1^{(2)})c_1-(1-\lambda_2\beta_1^{(2)}+\lambda_2\beta_1^{(1)})c_2). \qquad (6.2)$$

We employ Theorem 2 from Pakes (1981) which for our model, states that for our irreducible and aperiodic Markov chain $\{(J^{(n)}, Q_I^{(n)}, Q_{II}^{(n)}); n \geq 1\}$ with the state space $S = \{1, 2\} \times Z_+^2$, a sufficient condition for ergodicity is the existence of nonnegative function $f(k, i, j) = c_1 i + c_2 j$ and $\epsilon > 0$ such that the mean drift $\Delta_{kij} \leq -\epsilon$ is finite for all $(k, i, j) \in S = \{1, 2\} \times Z_+^2$ except a finite number of triples.

By examining (6.1) and (6.2), we construct $c_1 = 1 - \lambda_2\beta_1^{(2)} + \lambda_2\beta_1^{(1)}$ and $c_2 = 1 - \lambda_1\beta_1^{(1)} + \lambda_1\beta_1^{(2)}$.

We next show that for this choice of c_1 and c_2, that the condition for ergodicity is satisfied. Observer that

(1) for $k = 1$ or 2, $i = 0$ and $j \geq 0, \Delta_{k0j}$ can be rewritten as

$$\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+j\theta}+\lambda_1\beta_1^{(1)}+\lambda_2\beta_1^{(2)}-1$$

(2) for $k = 1$ or 2, $i \geq 1$ and $j \geq 0, \Delta_{kij}$ can be rewritten as

$$\frac{(\lambda_1+\lambda_2)(1-L_A(\lambda_1+\lambda_2+j\theta))}{\lambda_1+\lambda_2+j\theta}+\lambda_1\beta_1^{(1)}+\lambda_2\beta_1^{(2)}-1.$$

Note that

$$\lim_{j\to\infty}\Delta_{k0j}=\lambda_1\beta_1^{(1)}+\lambda_2\beta_1^{(2)}-1$$

and the function $\frac{1-L_A(\lambda_1+\lambda_2+j\theta)}{\lambda_1+\lambda_2+j\theta}$ is decreasing in j and

$$\Delta_{kij} \le \lambda_1\beta_1^{(1)} + \lambda_2\beta_1^{(2)} - L_A(\lambda_1+\lambda_2+j\theta).$$

Therefore, $\lambda_1\beta_1^{(1)} + \lambda_2\beta_1^{(2)} < L_A(\lambda_1+\lambda_2)$ is a sufficient condition for ergodicity for the Markov chain $\{(J^{(n)}, Q_I^{(n)}, Q_{II}^{(n)}); n \ge 1\}$, which implies that the system is stable since the arrival process is a Poisson process.

6.4 States of the system

The probabilities of different states of the system are defined as

(i) $P_{(0,0,j)}(t) \equiv P(J(t)=0, Q_I(t)=i, Q_{II}(t)=j)$, for $0 \le t < \infty$, $0 \le j < \infty$,

(ii) $P_{(0,i,j)}(t, x)\,dx \equiv P(J(t)=0, Q_I(t)=i, Q_{II}(t)=j, x < \xi(t) < x+dx)$, for $0 \le t, x < \infty$, $1 \le i < \infty$, $0 \le j < \infty$,

(iii) $P_{(1,i,j)}(t, y)\,dy \equiv P(J(t)=1, Q_I(t)=i, Q_{II}(t)=j, y < \eta_1(t) < y+dy)$, for $0 \le t, y < \infty$, $0 \le i, j < \infty$,

iv) $P_{(2,i,j)}(t, y)\,dy \equiv P(J(t)=2, Q_I(t)=i, Q_{II}(t)=j, y < \eta_2(t) < y+dy)$, for $0 \le t, y < \infty$, $0 \le i, j < \infty$.

Considering the transition rates between t and $t+\Delta t$, we obtain the following difference equations:

$$P_{(0,0,j)}(t+\Delta t) = (1-(\lambda_1+\lambda_2+j\theta)\Delta t)P_{(0,0,j)}(t)$$
$$+\Delta t\int_0^\infty P_{(1,0,j)}\beta_1(y)(t, y)\,dy + \Delta t\int_0^\infty P_{(2,0,j)}\beta_2(y)(t,y)\,dy + o(\Delta t), \quad (6.3)$$

$$P_{(0,i,j)}(t+\Delta t, x+\Delta t) = (1-(\lambda_1+\lambda_2+j\theta\alpha(x))\Delta t)P_{(0,i,j)}(t, x) + o(\Delta t), \quad (6.4)$$

$$P_{(1,i,j)}(t+\Delta t, y+\Delta t) = (1-(\lambda_1+\lambda_2+\beta_1(y))\Delta t)P_{(1,i,j)}(t, y)$$

$$+\lambda_1\triangle tP_{(1,i-1,j)}(t,\, y)+\lambda_2\triangle tP_{(1,i,j-1)}(t,\, y)+o(\triangle t), \tag{6.5}$$

$$P_{(2,i,j)}(t+\triangle t,\, y+\triangle t)=(1-(\lambda_1+\lambda_2+\beta_1(y))\triangle t)P_{(1,i,j)}(t,\, y)$$

$$+\lambda_1\triangle tP_{(2,i-1,j)}(t,\, y)+\lambda_2\triangle tP_{(2,i,j-1)}(t,y)+o(\triangle t), \tag{6.6}$$

$$P_{(0,i,j)}(t,0)=\int_0^\infty P_{(1,i,j)}\beta_1(y)(t,y)\,dy+\int_0^\infty P_{(2,i,j)}\beta_2(y)(t,\, y)\,dy, \tag{6.7}$$

$$P_{(1,i,j}(t,\, 0)=(1-\delta_{0i})\lambda_1\int_0^\infty P_{(0,i,j)}(t,\, x)\,dx$$

$$+\int_0^\infty P_{(0,i+1,j)}(t,\, x)\alpha(x)\,dx+\delta_{0i}\lambda_1P_{(0,0,j)}(t), \tag{6.8}$$

$$P_{(2,i,j}(t,\, 0)=\delta_{0i}(\lambda_2P_{(0,0,j)}(t)+(j+1)\theta P_{(0,0,j+1)}(t))$$

$$+(1-\delta_{0i})(\lambda_1\int_0^\infty P_{(0,i,j)}(t,\, x)\,dx+\int_0^\infty P_{(0,i,j+1)}(t,\, x)(j+1)\theta\,dx), \tag{6.9}$$

$$\sum_{i=0}^\infty P_{(0,0,j)}(t)+\sum_{i=0}^\infty\sum_{i=0}^\infty\int_0^\infty P_{(0,i,j)}(t,\, x)\,dx$$

$$+\sum_{i=0}^\infty\sum_{i=0}^\infty\int_0^\infty P_{(1,i,j)}(t,\, y)\,dy+\sum_{i=0}^\infty\sum_{i=0}^\infty\int_0^\infty P_{(2,i,j)}(t,\, y)\,dy=1. \tag{6.10}$$

(6.3)-(6.6) lead to the following differential equations:

$$(\frac{d}{dt}+\lambda_1+\lambda_2+j\theta)P_{(0,0,j)}(t)=\int_0^\infty P_{(1,0,j)}\beta_!(y)(t,\, y)\,dy$$

$$+\int_0^\infty P_{(2,0,j)}\beta_1(y)(t,\, y)\,dy, \tag{6.11}$$

$$(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_1+\lambda_2+j\theta\alpha(x))P_{(0,i,j)}(t,\, x)=0, \tag{6.12}$$

$$(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_1+\lambda_2+\beta_1(y))P_{(1,i,j)}(t,\, y)=\lambda_1P_{(1,i-1,j)}(t,\, y)$$

$$+\lambda_2P_{(1,i,j-1)}(t,\, y), \tag{6.13}$$

$$(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_1+\lambda_2+\beta_!(y))P_{(1,i,j)}(t,\,y)=\lambda_1 P_{(1,i-1,j)}(t,\,y)$$
$$+\lambda_2 P_{(1,i,j-1)}(t,\,y). \tag{6.14}$$

Henceforth, we assume that $\lambda_1\beta_1^{(1)}+\lambda_2\beta_1^{(2)}<L_A(\lambda_1+\lambda_2)$.

Let $P_{(0,0,j)}=\lim_{t\to+\infty}P_{(0,0,j)}(t)$, $P_{(0,i,j)}(x)=\lim_{t\to+\infty}P_{(0,i,j)}(t,\,x)$,

$P_{(1,i,j)}(y)=\lim_{t\to+\infty}P_{(1,i,j)}(t,\,y)$ and $P_{(1,i,j)}(y)=\lim_{t\to+\infty}P_{(1,i,j)}(t,\,y)$.

Therefore, equations(6.7)-(6.14) can be converted into

$$(\lambda_1+\lambda_2+j\theta)P_{(0,0,j)}(t)=\int_0^\infty \beta_!(y)P_{(1,0,j)}\,dy+\int_0^\infty P_{(2,0,j)}\beta_1(y)(y)\,dy, \tag{6.15}$$

$$(\frac{d}{dx}+\lambda_1+\lambda_2+j\theta+\alpha(x))P_{(0,i,j)}(x)=0, \tag{6.16}$$

$$(\frac{d}{dy}+\lambda_1+\lambda_2+\beta_1(y))P_{(1,i,j)}(y)=\lambda_1 P_{(1,i-1,j)}(y)+\lambda_2 P_{(1,i,j-1)}(y) \tag{6.17}$$

$$(\frac{d}{dy}+\lambda_1+\lambda_2+\beta_1(y))P_{(1,i,j)}(y)=\lambda_1 P_{(1,i-1,j)}(y)+\lambda_2 P_{(1,i,j-1)}(y) \tag{6.18}$$

$$P_{(0,i,j)}(0)=\int_0^\infty \beta_1(y)P_{(1,i,j)}\,dy+\int_0^\infty \beta_2(y)P_{(2,i,j)}\,dy, \tag{6.19}$$

$$P_{(1,i,j}(0)=(1-\delta_{0i})\lambda_1\int_0^\infty P_{(0,i,j)}(x)\,dx+\int_0^\infty \alpha(x)P_{(0,i+1,j)}(x)\,dx+\delta_{0i}\lambda_1 P_{(0,0,j)}, \tag{6.20}$$

$$P_{(2,i,j}(0)=\delta_{0i}(\lambda_2 P_{(0,0,j)}+(j+1)\theta P_{(0,0,j+1)})$$
$$+(1-\delta_{0i})(\lambda_2\int_0^\infty P_{(0,i,j)}(x)\,dx+\int_0^\infty P_{(0,i,j+1)}(x)(j+1)\theta\,dx), \tag{6.21}$$

$$\sum_{j=0}^\infty P_{(0,0,j)}+\sum_{i=1}^\infty\sum_{i=0}^\infty\int_0^\infty P_{(0,i,j)}(x)\,dx$$
$$+\sum_{i=0}^\infty\sum_{i=0}^\infty(\int_0^\infty P_{(1,i,j)}(y)\,dy+\int_0^\infty P_{(2,i,j)}(y)\,dy)=1. \tag{6.22}$$

In order to solve these equations, we introduce generating functions:

$$P_{(0,0)}(v)=\sum_{j=0}^\infty P_{(0,0,j)}v^j,\quad P_{(0,1)}(u,\,v,\,x)=\sum_{i=1}^\infty\sum_{j=0}^\infty P_{(0,i,j)}u^i v^j,$$

$$P_{(1)}(u,\,v,\,y)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P_{(1,i,j)}u^iv^j,\quad P_{(2)}(u,\,v,\,y)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}P_{(2,i,j)}u^iv^j.$$

Multiplying both sides of (6.15) by v^j and of (6.16)-(6.22) by u^iv^j summing over all possible i and j, we have

$$(\frac{\partial}{\partial x}+\lambda_1+\lambda_2+\alpha(x)+\theta v\frac{\partial}{\partial v})P_{(0,1)}(u,\,v,\,x)=0, \tag{6.23}$$

$$(\frac{\partial}{\partial x}+\lambda_1+\lambda_2+\beta_1(y))P_{(1)}(u,\,v,\,y)=(\lambda_1u+\lambda_2v)P_{(1)}(u,\,v,\,y), \tag{6.24}$$

$$(\frac{\partial}{\partial x}+\lambda_1+\lambda_2+\beta_1(y))P_{(2)}(u,\,v,\,y)=(\lambda_1u+\lambda_2v)P_{(2)}(u,\,v,\,y), \tag{6.25}$$

$$\begin{aligned}P_{(0,1)}(u,\,v,\,0)&=\int_0^{\infty}\beta_2(y)P_{(2)}(u,\,v,\,y)\,dy+\int_0^{\infty}\beta_1(y)P_{(1)}(u,\,v,\,y)\,dy\\&-(\lambda_1+\lambda_2+\theta v\frac{d}{dv}P_{(0,0)}(v)),\end{aligned} \tag{6.26}$$

$$P_{(1)}(u,\,v,\,0)=\frac{1}{u}\int_0^{\infty}P_{(0,1)}(u,\,v,\,x)\alpha(x)\,dx+\lambda_1\int_0^{\infty}P_{(0,1)}(u,\,v,\,x)\,dx+\lambda_1P_{(0,0)}(v), \tag{6.27}$$

$$\begin{aligned}P_{(2)}(u,\,v,\,0)&=\lambda_2P_{(0,0)}(v)+\theta\frac{d}{dv}P_{(0,0)}(v)\\&+\lambda_2\int_0^{\infty}P_{(0,1)}(u,\,v,\,x)\,dx+\theta\int_0^{\infty}\frac{\partial}{\partial v}P_{(0,1)}(u,\,v,\,x)\,dx,\end{aligned} \tag{6.28}$$

$$\begin{aligned}&\lim_{v\to1^-}P_{(0,0)}(v)+\lim_{v\to1^-}\lim_{u\to1^-}(\int_0^{\infty}P_{(0,1)}(u,\,v,\,x)\,dx\\&+\int_0^{\infty}P_{(1)}(u,\,v,\,y)\,dy+\int_0^{\infty}P_{(2)}(u,\,v,\,y)\,dy)=1.\end{aligned} \tag{6.29}$$

The solution of the system of equations above is still unfinished work. We will make a further investigation in the future.

Chapter 7

Numerical Results Related to Model III

7.1 Quasi-Birth-and-Death Process

In this chapter, we consider a special case of the $M/G/1$ model III of Chapter 6. In order to overcome the difficulties of the $M/G/1$ model III, we consider an $M/M/1$ model III. We also place an upper bound on the number of customers allowed into the retrial orbits. Much of the first part of Chapter 6 is repeated (except for a change in the service description and the upper bound change on the source II customers) so that the current chapter can stand alone.

An $M/M/1$ queue with general retrial times, two primary sources, a retrial queue and retrial orbits, with different disciplines is investigated. The description of model III is as follows. We consider a single-server queueing system with two primary sources (called primary source I and primary source II and no waiting room in front of the server. We assume that primary customers from primary source I arrive in a Poisson process with rate λ_1 and primary customers from primary source II arrive in a Poisson process with rate λ_2. If an arriving primary or retrial customer finds the server idle, that customer begins service immediately and leaves the system after the service completion. If an arriving primary customer from primary source I finds the server busy, the customer enters a retrial queue according to a FCFS discipline. For customers in the retrial queue, only the customer at the head of the retrial queue is allowed to demand service in a time measured from the time the server is idle. It is required to cancel the demand for service if a primary customer arrives first and the retrial customer then returns to the position at the head of the retrial queue. If an arriving primary source II customer encounters a busy server, and the number of source II customers in orbit equals c, then the customer is lost to the system. If an arriving primary source II customer finds the server busy, and the number of source II customers in orbit is less than or equal to some value c, that customer enters its own retrial orbit (not a retrial queue) and keeps making retrials until the customer eventually obtains service. All such retrial customers form independent retrial sources each with its own retrial orbit. The retrial times in the retrial queue are distributed with distribution function $A(x)$, density function $a(x)$, Laplace transform $L_A(s)$ and first two moments α_1 and α_2. The retrial times in the retrial orbits are each exponentially distributed with rate θ. The service times for any customer (regardless of whether the customer is a primary customer from (I) or is a retrial customer from the retrial queue or a retrial orbit) have an exponential distribution with mean $\frac{1}{\mu}$. All times considered are mutually independent.

At time t, let $C(t)$ denote the state of the server. That is, $C(t) = 0$ means that the server is idle and $C(t) = 1$ means that the server is busy. Let $N_1(t)$ and $N_2(t)$

denote the number of customers in the retrial orbits and the retrial queue, respectively. It can seen that the process $\{(J(t), N_1(t), N_2(t)); t \geq 0\}$ may be viewed as a QBD process (quasi-birth-and-death process) on the state space $\Theta = \{(i, j, k); 0 \leq i \leq 1, 0 \leq j \leq c, 0 \leq k \leq \infty\}$, where the states are ordered in groups as follows, $(0,0,0), (0,1,0), ...(0,c,0); (1,0,0), (1,1,0), ..., (1,c,0); (0,0,1), (0,1,1), ..., (0,c,1);$ $(1,0,1), (1,1,1), ..., (1,c,1); ....$ Thus, it is easily seen that the infinitesimal generator Q is given by

$$Q = \begin{pmatrix} B_0 & A_0 & 0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & 0 & A_2 & A_1 & A_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where

$$B_0 = \begin{pmatrix} B_{11}^{(0)} & B_{12}^{(0)} \\ B_{21}^{(0)} & B_{22}^{(0)} \end{pmatrix},$$

$$B_{11}^{(0)} = \begin{pmatrix} -(\lambda_1+\lambda_2) & 0 & 0 & \cdots & 0 \\ 0 & -(\lambda_1+\lambda_2+\theta) & 0 & \cdots & 0 \\ 0 & 0 & -(\lambda_1+\lambda_2+2\theta) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda_1+\lambda_2+c\theta) \end{pmatrix},$$

$$B_{12}^{(0)} = \begin{pmatrix} \lambda_1+\lambda_2 & 0 & 0 & \cdots & 0 \\ \theta & \lambda_1+\lambda_2 & 0 & \cdots & 0 \\ 0 & 2\theta & \lambda_1+\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_1+\lambda_2 \end{pmatrix}, B_{21}^{(0)} = \begin{pmatrix} \mu & 0 & 0 & \cdots & 0 \\ 0 & \mu & 0 & \cdots & 0 \\ 0 & 0 & \mu & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mu \end{pmatrix},$$

$$B_{22}^{(0)} = \begin{pmatrix} -(\lambda_1+\lambda_2+\mu) & \lambda_2 & 0 & \cdots & 0 \\ 0 & -(\lambda_1+\lambda_2+\mu) & \lambda_2 & \cdots & 0 \\ 0 & 0 & -(\lambda_1+\lambda_2+\mu) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda_1+\lambda_2+\mu) \end{pmatrix},$$

$$A_0 = \begin{pmatrix} A_{11}^{(0)} & A_{12}^{(0)} \\ A_{21}^{(0)} & A_{22}^{(0)} \end{pmatrix},$$

$$A_{11}^{(0)} = A_{12}^{(0)} = A_{21}^{(0)} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, A_{22}^{(0)} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_1 \end{pmatrix},$$

$$A_1 = \begin{pmatrix} A_{11}^{(1)} & A_{12}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{pmatrix},$$

$$A_{11}^{(1)} = \begin{pmatrix} -(\lambda_1+\lambda_2+\alpha) & 0 & \cdots & 0 \\ 0 & -(\lambda_1+\lambda_2+\alpha+\theta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -(\lambda_1+\lambda_2+\alpha+c\theta) \end{pmatrix},$$

$$A_{12}^{(1)} = \begin{pmatrix} \lambda_1+\lambda_2 & 0 & 0 & \cdots & 0 \\ \theta & \lambda_1+\lambda_2 & 0 & \cdots & 0 \\ 0 & 2\theta & \lambda_1+\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_1+\lambda_2 \end{pmatrix},$$

$$A_{21}^{(1)} = \begin{pmatrix} \mu & 0 & 0 & \cdots & 0 \\ 0 & \mu & 0 & \cdots & 0 \\ 0 & 0 & \mu & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mu \end{pmatrix},$$

$$A_{22}^{(1)} = \begin{pmatrix} -(\lambda_1+\lambda_2+\mu) & \lambda_2 & 0 & \cdots & 0 \\ 0 & -(\lambda_1+\lambda_2+\mu) & \lambda_2 & \cdots & 0 \\ 0 & 0 & -(\lambda_1+\lambda_2+\mu) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda_1+\mu) \end{pmatrix},$$

$$A_2 = \begin{pmatrix} A_{11}^{(2)} & A_{12}^{(2)} \\ A_{21}^{(2)} & A_{22}^{(2)} \end{pmatrix}, \; A_{11}^{(2)} = A_{21}^{(2)} = A_{22}^{(2)} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$A_{12}^{(2)} = \begin{pmatrix} \alpha & 0 & 0 & \cdots & 0 \\ 0 & \alpha & 0 & \cdots & 0 \\ 0 & 0 & \alpha & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha \end{pmatrix}.$$

7.2 Approximate Stationary Distribution

According to Neuts (1981, p. 82), for the QBD process, the matrix R, which is needed to solve for the steady state probabilities, is the minimal nonnegative solution to the quadratic matrix equation

$$R^2 A_2 + RA_1 + A_0 = 0. \tag{7.1}$$

We first consider the case when $c = 1$. In this case A_2, A_1, and A_0 are all 4×4 matrices. Thus R must also be 4×4. In order to find R, let $R = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{pmatrix}$.

Thus,

$$-(\lambda_1 + \lambda_2 + \alpha)x_{11} + \mu x_{13} = 0, \tag{7.2}$$

$$-(\lambda_1+\lambda_2+\alpha+\theta)x_{12}+\mu x_{14}=0, \tag{7.3}$$

$$-(\lambda_1+\lambda_2+\alpha)x_{21}+\mu x_{23}=0, \tag{7.4}$$

$$-(\lambda_1+\lambda_2+\alpha+\theta)x_{22}+\mu x_{24}=0, \tag{7.5}$$

$$-(\lambda_1+\lambda_2+\alpha)x_{31}+\mu x_{33}=0, \tag{7.6}$$

$$-(\lambda_1+\lambda_2+\alpha+\theta)x_{32}+\mu x_{34}=0, \tag{7.7}$$

$$-(\lambda_1+\lambda_2+\alpha)x_{41}+\mu x_{43}=0, \tag{7.8}$$

$$-(\lambda_1+\lambda_2+\alpha+\theta)x_{42}+\mu x_{44}=0, \tag{7.9}$$

$$\alpha({x_{11}}^2+x_{12}x_{21})+x_{13}x_{31}+x_{14}x_{41}+(\lambda_1+\lambda_2)x_{11}+\theta x_{12}-(\lambda_1+\lambda_2+\mu)x_{13}=0, \tag{7.10}$$

$$\alpha(x_{11}x_{12}+x_{12}x_{22})+x_{13}x_{32}+x_{14}x_{42}+(\lambda_1+\lambda_2)x_{12}+\lambda_2 x_{13}-(\lambda_1+\mu)x_{14}=0, \tag{7.11}$$

$$\alpha(x_{11}x_{21}+x_{21}x_{22})+x_{23}x_{31}+x_{24}x_{41}+(\lambda_1+\lambda_2)x_{21}+\theta x_{22}-(\lambda_1+\lambda_2+\mu)x_{23}=0, \tag{7.12}$$

$$\alpha(x_{12}x_{21}+{x_{22}}^2)+x_{23}x_{32}+x_{24}x_{42}+(\lambda_1+\lambda_2)x_{22}+\lambda_2 x_{23}-(\lambda_1+\mu)x_{24}=0, \tag{7.13}$$

$$\alpha(x_{31}x_{11}+x_{32}x_{21})+x_{33}x_{31}+x_{34}x_{41}+(\lambda_1+\lambda_2)x_{31}+\theta x_{32}-(\lambda_1+\lambda_2+\mu)x_{33}=-\lambda_1, \tag{7.14}$$

$$\alpha(x_{31}x_{12}+x_{32}x_{22})+x_{33}x_{32}+x_{34}x_{42}+(\lambda_1+\lambda_2)x_{32}+\lambda_2 x_{33}-(\lambda_1+\mu)x_{34}=0, \tag{7.15}$$

$$\alpha(x_{41}x_{11}+x_{42}x_{21})+x_{43}x_{31}+x_{44}x_{41}+(\lambda_1+\lambda_2)x_{41}+\theta x_{42}-(\lambda_1+\lambda_2+\mu)x_{43}=0, \tag{7.16}$$

$$\alpha(x_{41}x_{12}+x_{42}x_{22})+x_{43}x_{32}+x_{44}x_{42}+(\lambda_1+\lambda_2)x_{42}+\lambda_2 x_{43}-(\lambda_1+\mu)x_{44}=-\lambda_1. \tag{7.17}$$

Because some of the equations are not linear in x_{ij}, we are not able obtain the matrix R exactly, even when $c=1$. Thus, we will use an iterative scheme to obtain a numerical result for R. Our numerical method will work for any value of c.

Let $\lambda_1=\lambda_2=\alpha=\theta=1$ and $\mu=7$ and $c=2$, as a typical example. These values satisfy the condition of the previous chapter to guarantee that the limiting probabilities exist. We use a numerical method from Grassmann (2000, p.170) to get the matrix R. Here

$$Q = \begin{pmatrix} B_0 & A_0 & 0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & 0 & A_2 & A_1 & A_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where

$$B_0 = \begin{pmatrix} B_{11}^{(0)} & B_{12}^{(0)} \\ B_{21}^{(0)} & B_{22}^{(0)} \end{pmatrix}, \; B_{11}^{(0)} = \begin{pmatrix} -(\lambda_1+\lambda_2) & 0 & 0 \\ 0 & -(\lambda_1+\lambda_2+\theta) & 0 \\ 0 & 0 & -(\lambda_1+\lambda_2+2\theta) \end{pmatrix},$$

$$B_{12}^{(0)} = \begin{pmatrix} \lambda_1+\lambda_2 & 0 & 0 \\ \theta & \lambda_1+\lambda_2 & 0 \\ 0 & 2\theta & \lambda_1+\lambda_2 \end{pmatrix}, \; B_{21}^{(0)} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \mu \end{pmatrix},$$

$$B_{22}^{(0)} = \begin{pmatrix} -(\lambda_1+\lambda_2+\mu) & \lambda_2 & 0 \\ 0 & -(\lambda_1+\lambda_2+\mu) & \lambda_2 \\ 0 & 0 & -(\lambda_1+\mu) \end{pmatrix},$$

$$A_0 = \begin{pmatrix} A_{11}^{(0)} & A_{12}^{(0)} \\ A_{21}^{(0)} & A_{22}^{(0)} \end{pmatrix}, \; A_{11}^{(0)} = A_{12}^{(0)} = A_{21}^{(0)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; A_{22}^{(0)} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix},$$

$$A_1 = \begin{pmatrix} A_{11}^{(1)} & A_{12}^{(1)} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{pmatrix},$$

$$A_{11}^{(1)} = \begin{pmatrix} -(\lambda_1+\lambda_2+\alpha) & 0 & 0 \\ 0 & -(\lambda_1+\lambda_2+\alpha+\theta) & 0 \\ 0 & 0 & -(\lambda_1+\lambda_2+\alpha+2\theta) \end{pmatrix},$$

$$A_{12}^{(1)} = \begin{pmatrix} \lambda_1+\lambda_2 & 0 & 0 \\ \theta & \lambda_1+\lambda_2 & 0 \\ 0 & 2\theta & \lambda_1+\lambda_2 \end{pmatrix}, A_{21}^{(1)} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \mu \end{pmatrix},$$

$$A_{22}^{(1)} = \begin{pmatrix} -(\lambda_1+\lambda_2+\mu) & \lambda_2 & 0 \\ 0 & -(\lambda_1+\lambda_2+\mu) & \lambda_2 \\ 0 & 0 & -(\lambda_1+\mu) \end{pmatrix},$$

$$A_2 = \begin{pmatrix} A_{11}^{(2)} & A_{12}^{(2)} \\ A_{21}^{(2)} & A_{22}^{(2)} \end{pmatrix}, A_{11}^{(2)} = A_{21}^{(2)} = A_{22}^{(2)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, A_{12}^{(2)} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix}.$$

We use the following iterative scheme to obtain an approximation for R. Let

$$Y_{(i)} = \sum_{j=0}^{\infty} R_{(i)}^{j} A_{j+1}, \tag{7.18}$$

and

$$R_{(i+1)} = A_0(I - Y_{(i)})^{-1}. \tag{7.19}$$

First, we set

$$R_{(0)} = \begin{pmatrix} 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \end{pmatrix}.$$

It follows from (7.18) and (7.19) that

$$R_{(6)} = \begin{pmatrix} 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \\ 0.292851 & 0.036898 & 0.050391 & 0.167343 & 0.026356 & 0.004319 \\ 0.065238 & 0.220483 & 0.029237 & 0.037279 & 0.157488 & 0.025060 \\ 0.025490 & 0.082397 & 0.191091 & 0.014565 & 0.058855 & 0.163792 \end{pmatrix},$$

$$R_{(7)} = \begin{pmatrix} 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. & 0. & 0. \\ 0.292851 & 0.036898 & 0.005039 & 0.167343 & 0.026356 & 0.004319 \\ 0.065238 & 0.220484 & 0.029237 & 0.037279 & 0.157488 & 0.025060 \\ 0.025490 & 0.082397 & 0.191091 & 0.014565 & 0.058855 & 0.163792 \end{pmatrix}.$$

Since $R_{(7)} = R_{(6)}$, at least to the number of decimal places shown, we use $R_{(7)}$ as our approximation for R. Define π to be the stationary probability vector for the QBD process. Write $\pi = (\pi_0, \pi_1, \ldots)$. By using theorem 3.1.1 of Neuts (1981, p. 82)

$$\pi_0(B_0 + RB_1) = 0 \tag{7.20}$$

$$\pi_0(I-R)^{-1}e = 1 \tag{7.21}$$

$$\pi_n = \pi_0 R^n, \ n = 0, 1, 2, \ldots \tag{7.22}$$

so using MAPLE, we obtain

$\pi_0 = (0.357068, 0.168106, 0.093073, 0.102019, 0.072045, 0.053184)$

and therefore $\pi_1 = (0.035932, 0.024031, 0.012783, 0.020532, 0.017165, 0.010957)$,

and $\pi_2 = (0.007412, 0.005445, 0.002699, 0.004235, 0.003889, 0.002313)$, ..., and so on.

7.3 Performance Measures

Now we discuss some performance measures. In steady state, let C denote the state of the server, N_1 denote the number of customers in the retrial orbits, N_2 denote the number of customers in the retrial queue and N denote the number of customers in the system. Let $\pi_{ijk} = P(C = i, N_1 = j, N_2 = k), i = 0, 1; j = 0, 1, \ldots, c; k = 0, 1, 2, \ldots$. Then we easily obtain the following performance measures.

Property 7.1 Under the assumptions and notation above,

(i) $P(C = 0, N_1 = 0, N_2 = 0) = \pi_{000}$ =Prob(empty system),

(ii) $P(N_1 = 0) = \sum_{i=0}^{1}\sum_{k=0}^{\infty}\pi_{i0k}$ =Prob(no customer in retrial orbit),

(iii) $P(N_2 = 0) = \sum_{i=0}^{1}\sum_{j=0}^{c}\pi_{ij0}$ =Prob(empty retrial queue),

(iv) $P(C = 0, N_1 + N_2 \neq 0) = \sum_{j+k\neq 0}\pi_{0jk}$ =Prob(idle server but nonempty system),

(v) $P(C = 1, N_1 = 2) = \sum_{k=0}^{\infty}\pi_{12k}$ =Prob(busy server, retrial orbits full),

(vi) $EN_1 = \sum_{i=0}^{1}\sum_{j=0}^{c}\sum_{k=0}^{\infty} j\pi_{ijk}$ =expected retrial orbit number,

(vii) $EN_2 = \sum_{i=0}^{1}\sum_{j=0}^{c}\sum_{k=0}^{\infty} k\pi_{ijk}$ =expected retrial queue size,

(viii) $EN = \sum_{i=0}^{1}\sum_{j=0}^{c}\sum_{k=0}^{\infty}(i + j + k)\pi_{ijk}$ =expected system size, As an example, let $\lambda_1 = \lambda_2 = \alpha = \theta = 1$ and $\mu = 7$ and $c = 2$. We will obtain approximate values of the performance measures above. According to the discussion in the previous section, we obtain

$$\begin{pmatrix} \pi_{000} & \pi_{010} & \pi_{020} & \pi_{100} & \pi_{110} & \pi_{120} \\ \pi_{001} & \pi_{011} & \pi_{021} & \pi_{101} & \pi_{111} & \pi_{121} \\ \pi_{002} & \pi_{012} & \pi_{022} & \pi_{102} & \pi_{112} & \pi_{122} \\ \pi_{003} & \pi_{013} & \pi_{023} & \pi_{103} & \pi_{113} & \pi_{123} \\ \pi_{004} & \pi_{014} & \pi_{024} & \pi_{104} & \pi_{114} & \pi_{124} \\ \pi_{005} & \pi_{015} & \pi_{025} & \pi_{105} & \pi_{115} & \pi_{125} \\ \pi_{006} & \pi_{016} & \pi_{026} & \pi_{106} & \pi_{116} & \pi_{126} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$= \begin{pmatrix} 0.357068 & 0.168106 & 0.093073 & 0.102019 & 0.072045 & 0.053184 \\ 0.035932 & 0.024031 & 0.012783 & 0.020532 & 0.017165 & 0.010957 \\ 0.007412 & 0.005445 & 0.002699 & 0.004235 & 0.003889 & 0.002313 \\ 0.001553 & 0.001204 & 0.000577 & 0.000887 & 0.000860 & 0.000494 \\ 0.000328 & 0.000263 & 0.000124 & 0.000187 & 0.000187 & 0.000106 \\ 0.000069 & 0.000057 & 0.000026 & 0.000040 & 0.000040 & 0.000023 \\ 0.000015 & 0.000012 & 0.000005 & 0.000008 & 0.000008 & 0.000005 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Therefore, $P(C = 0, N_1 = 0, N_2 = 0) = 0.357068, P(N_1 = 0) = 0.530285, P(N_2 = 0) = 0.845495, P(C = 0, N_1 + N_2 \neq 0) = 0.710782, P(C = 1, N_1 = 2) = 0.067082, EN_1 = 0.646050, EN_2 = 0.196484, EN = 1.070110$.

We next discuss the waiting time that a customer from primary source I and

Source II spend in the system, respectively. Let $W_{(i,j,k,1)}$ denote the expected waiting time that a customer from primary source I spends in the system when it arrives at the state (i,j,k), and let $W_{(i,j,k,2)}$ denote the expected waiting time that a customer from primary source II spends in the system when it arrives at the state $(i,j,k), i = 0,1; j = 0,1,\ldots,c; k = 0,1,\ldots$. Let $\lambda = \lambda_1 + \lambda_2$. Then $W_{(0,j,k,l)} = \frac{1}{\mu}$ for $l = 1,2$ and

$$\begin{aligned}
W_{(1,j,k,1)} &= \frac{\mu}{\lambda+\mu}[\frac{\lambda}{\lambda+\alpha+j\theta}(\frac{1}{\lambda+\mu}+\frac{1}{\lambda+\alpha+j\theta}+W_{(1,j,k,1)}) \\
&+\frac{j\theta}{\lambda+\alpha+j\theta}(\frac{1}{\lambda+\mu}+\frac{1}{\lambda+\alpha+j\theta}+W_{(1,j-1,k,1)}) \\
&+\frac{\alpha}{\lambda+\alpha+j\theta}(\frac{1}{\lambda+\mu}+\frac{1}{\lambda+\alpha+j\theta}+\frac{1}{\mu}+W_{(1,j,k-1,1)})] \\
&+\frac{\lambda_1}{\lambda+\mu}(\frac{1}{\lambda+\mu}+W_{(1,j,k,1)})+\frac{\lambda_2}{\lambda+\mu}(\frac{1}{\lambda+\mu}+W_{(1,j+1,k,1)}) \\
&= \frac{1}{\lambda+\mu}+\frac{\mu+\alpha}{(\lambda+\mu)(\lambda+\alpha+j\theta)}+\frac{j\mu\theta}{(\lambda+\mu)(\lambda+\alpha+j\theta)}\times W_{(1,j-1,k,1)} \\
&+\frac{j\mu\theta}{(\lambda+\mu)(\lambda+\alpha+j\theta)}\times W_{(1,j,k-1,1)}+(\frac{\lambda\mu}{(\lambda+\mu)(\lambda+\alpha+j\theta)}+\frac{\lambda_1}{\lambda+\mu})\times W_{(1,j,k,1)} \\
&+\frac{\lambda_2}{\lambda+\mu}\times W_{(1,j+1,k,1)}
\end{aligned} \tag{7.23}$$

and

$$\begin{aligned}
W_{(1,j,k,2)} &= \frac{\mu}{\lambda+\mu}[\frac{\lambda}{\lambda+\alpha+(j+1)\theta}(\frac{1}{\lambda+\mu}+\frac{1}{\lambda+\alpha+(j+1)\theta}+W_{(1,j+1,k,2)}) \\
&+\frac{\theta}{\lambda+\alpha+(j+1)\theta}(\frac{1}{\lambda+\mu}+\frac{1}{\lambda+\alpha+(j+1)\theta}+\frac{1}{\mu}) \\
&+\frac{j\theta}{\lambda+\alpha+(j+1)\theta}(\frac{1}{\lambda+\mu}+\frac{1}{\lambda+\alpha+(j+1)\theta}+W_{(1,j,k,2)}) \\
&+\frac{\alpha}{\lambda+\alpha+(j+1)\theta}(\frac{1}{\lambda+\mu}+\frac{1}{\lambda+\alpha+(j+1)\theta}+W_{(1,j+1,k-1,2)})] \\
&+\frac{\lambda_1}{\lambda+\mu}(\frac{1}{\lambda+\mu}+W_{(1,j+1,k+1,2)})+\frac{\lambda_2}{\lambda+\mu}(\frac{1}{\lambda+\mu}+W_{(1,j+2,k,2)}) \\
&= \frac{1}{\lambda+\mu}+\frac{\mu+\theta}{(\lambda+\mu)(\lambda+\alpha+(j+1)\theta)}+\frac{j\mu\theta}{(\lambda+\mu)(\lambda+\alpha+(j+1)\theta)}\times W_{(1,j,k,2)} \\
&+(\frac{\mu\alpha}{(\mu\alpha)(\lambda+\alpha+(j+1)\theta)}\times W_{(1,j+1,k-1,2)}
\end{aligned}$$

$$+\frac{\lambda\mu}{(\lambda+\mu)(\lambda+\alpha+(j+1)\theta)}\times W_{(1,j+1,k,2)}$$
$$+\frac{\lambda_1}{\lambda+\mu}\times W_{(1,j+1,k+1,2)}+\frac{\lambda_2}{\lambda+\mu}\times W_{(1,j+2,k,2)}, \tag{7.24}$$

where $W_{(1,-1,k,1)} \equiv 0, W_{(1,j,-1,1)} \equiv 0, W_{(1,j,-1,2)} \equiv 0, W_{(1,c,k,l)} \equiv 0, W_{(1,c+1,k,l)} \equiv 0.$

It can be seen these equations are obtained. For example, when $c = 2$, we obtain $W_{(1,0,1,1)}, W_{(1,1,1,1)}, W_{(1,2,1,1)}$ which can be solved, and so on.

Chapter 8

Model IV Description and Stability

8.1 Model description and stability condition

Model IV is described as follows. We consider a single server retrial queueing system with a discrete service time taking on value D_j with probability p_j, $j = 1, 2, ...$, for each customer (primary or repeated customer). Primary customers arrive according to a Poisson process with rate λ. There is no waiting room in front of the server. If an arriving primary customer finds the server idle, the customer begins service immediately and leaves the system after service completion. If the server is found to be busy, the arriving primary customer enters a retrial queue according to a FCFS discipline. When the server becomes idle, if the retrial queue is empty, a primary customer is served next; otherwise, if the retrial queue is nonempty, the customer at the head of the retrial queue (called an active retrial customer) competes with the next arriving primary customer to reach the server first. The time for the active customer to reach the server is generally distributed with cdf $A(x)$, pdf $a(x)$ and Laplace-Stieltjes transform $A^*(s)$. The active retrial customer is required to cancel its attempt for service if a primary customer arrives first. In that case, the active retrial customer remains at the head of the retrial queue. We call this retrial queue an $M/\{D_n\}/1$ retrial queue. Inter-arrival times, retrial times and service times are mutually independent.

Example 8.1 Again we consider the call center with an answering machine. The answering machine holds the information of the FCFS retrial customers. The nature of the service may be such that a computer program is to be run based on the characteristics of the customer. The nature of the service required by the customer indicates which level of program should be run: minimal, standard, complete, etc. The various levels of the program take different (essentially) discrete amounts of time.

At time t, let $J(t)$ denote the server state. $J(t) = 0$ means that the server is idle and $J(t) = j$ means that the server is busy with service time D_j, $j = 1, 2, \ldots$. Let $Q(t)$ denote the number of customers in the retrial queue. Since the service times are discrete, the stochastic process $\{(J(t), Q(t)); t \geq 0\}$ is not necessarily Markovian. In order to analyze the stochastic evolution of our retrial queueing model, we introduce the following two supplementary variables. At time t, if $J(t) = 0$ and $Q(t) > 0$, we define $\xi(t)$ as the elapsed retrial time; if $J(t) = j(\neq 0)$, we define $\eta(t)$ as the elapsed service time. Thus, the stochastic process $\{X(t), t \geq 0\} = \{(J(t), Q(t), \xi(t), \eta(t)); t \geq 0\}$ is a Markov process. The conditional completion rate of the retrial time is defined as $\alpha(x) = a(x)(1 - A(x))^{-1}$.

The following theorem gives a necessary and sufficient condition for the system to be stable.

Theorem 8.1 An $M/\{D_n\}/1$ retrial queueing system is stable if and only if $\lambda \sum_{j=1}^{\infty} p_j D_j < A^*(\lambda)$.

Proof. Let Q_n be the number of customers in the retrial queue just before the nth departure point from the system. Since primary customers arrive in a Poisson process, we use Burke's theorem (see [7], p.187) to conclude that the steady state probabilities of $\{(J(t), Q(t)); t \geq 0\}$ exist and are positive if and only if $\{Q_n, n = 1, 2, ...\}$ is ergodic. Note that the mean drift $\chi_i \equiv E(Q_{n+1} - Q_n \,|\, Q_n = i) = \lambda \sum_{j=1}^{\infty} p_j D_j - \delta_{i0} L_A(\lambda)$, $i = 0, 1, ...$, where δ_{i0} is the Kronecker delta. By using results of Pakes (1969) and Sennott et al. (1983), we can show that $\lambda \sum_{j=1}^{\infty} p_j D_j < L_A(\lambda)$ is a necessary and sufficient condition for $\{Q_n, n = 1, 2, ...\}$ to be ergodic. The theorem follows.

Corollary 8.1 For an $M/\{D_n\}/1$ regular(non retrial) queueing system, the system is stable if and only if $\lambda \sum_{j=1}^{\infty} p_j D_j < 1$.

Proof. Let $A(x) = 1 - e^{-\alpha x}$ and let $\alpha \to \infty$. By using the result of Theorem 1, we get our conclusion.

Now we give two less formal remarks about the condition of the system stability.

Remark 8.1 Let $P(\text{busy})$ be the probability that the server is busy. Let S be a service time. Then $ES = \sum_{j=1}^{\infty} p_j D_j$. However, the rate out of the retrial queue is

$\frac{1}{ES}P(\text{busy})L_A(\lambda)$ and the rate into the retrial queue is $\lambda P(\text{busy})$. Hence, for stability,

$$\lambda P(\text{busy}) < \frac{1}{ES}P(\text{busy})L_A(\lambda),$$

i.e.,

$$\lambda \sum_{j=1}^{\infty} D_j p_j < L_A(\lambda).$$

This gives us a second method to obtain the stability condition.

Remark 8.2 In the analysis of the system stability of our model, we may view it as a regular $M/\{D_n\}/1$ queue. We define the generalized service time as a time interval from the instant that a customer completes service and the retrial queue is nonempty until the next customer completes the service. Let S^* and S denote generalized service time and service time, respectively. Let X_P and X_R denote inter-arrival time and retrial time, respectively. Then $S^* = S + min(X_P, X_R)$. Note that $P(min(X_P, X_R) > x) = e^{-\lambda x}(1 - A(x))$. Therefore,

$$\begin{aligned} E(min(X_P, X_R)) &= \int_0^{\infty} P(min(X_P, X_R) > x)\, dx \\ &= \int_0^{\infty} e^{-\lambda x}\, dx - \int_0^{\infty} e^{-\lambda x} A(x)\, dx \\ &= \frac{1 - L_A(\lambda)}{\lambda}. \end{aligned}$$

Hence, the system is stable if and only if $\frac{1}{\lambda} > \sum_{j=1}^{\infty} p_j D_j + \frac{1-L_A(\lambda)}{\lambda}$, i.e., $\lambda \sum_{j=1}^{\infty} p_j D_j < L_A(\lambda)$. This method of analysis of the system stability adds descriptive insight.

8.2 Steady state equations and their solution

In this section, our objective is to obtain the joint distribution of the server state and the queue length in steady state. We first introduce the new concept "memory of service level". The memory of service level $J^*(t)$ is defined as an indicator for the most recent service completed up to time t. That is, "$J^*(t) = j$" means the most recent service completed prior to time t has length D_j, $j = 1, 2, \ldots$. Next we define the probabilities and the probability densities as follows.

(i) $P_E(t) = P(J(t) = 0,\ Q(t) = 0)$,

(ii) $P_{Ej}(t) = P(J(t) = 0,\ J^*(t) = j,\ Q(t) = 0)$, $j = 1, 2, \ldots,$

(iii) $P_{Ii}(t, x)\,dx = P(J(t) = 0,\ Q(t) = i,\ x < \xi(t) < x + dx)$, $i = 1, 2, \ldots,$

(iv) $P_{Iij}(t, x)\,dx = P(J(t) = 0,\ J^*(t) = j, Q(t) = i,\ x < \xi(t) < x + dx)$, $i = 1, 2, \ldots$, $j = 1, 2, \ldots,$

(v) $P_{Bij}(t, x)\,dx = P(J(t) = j,\ Q(t) = i,\ x < \eta(t) < x + dx)$, $i = 0, 1, \ldots$, $j = 1, 2, \ldots$.

It is obvious that $P_E(t) = \sum_{j=1}^{\infty} P_{Ej}(t)$ and $P_{Ii}(t, x) = \sum_{j=1}^{\infty} P_{Iij}(t, x)$, $i = 1, 2, \ldots$.

To get the steady state equations, we need the following results.

Lemma 8.1 (i) For $i = 0, 1, \ldots$, $j = 1, 2, \ldots$, $P(J(t + D_j) = 0,\ J^*(t + D_j) = j,\ Q(t + D_j) = 0 \mid J(t) = 0,\ J^*(t) = i,\ Q(t) = 0) = e^{-\lambda D_j}$ if $i = j$ and 0 otherwise.

(ii) For $i = 0, 1, \ldots$, $j = 1, 2, \ldots$, $0 < x < D_j$, $P(J(t + D_j) = 0,\ J^*(t + D_j) = j,\ Q(t + D_j) = 0 \mid J(t) = i,\ Q(t) = 0,\ \eta(t) > x) = e^{-\lambda D_j}$ if $i = j$ and 0 otherwise.

Proof. (i) If the event $(J(t) = 0,\ J^*(t) = i,\ Q(t) = 0)$ and the event $(J(t + D_j) = 0,\ J^*(t + D_j) = j,\ Q(t + D_j) = 0)$ occur, then no primary customers arrive between time t and time $t+D_j$. Therefore, $P(J(t+D_j)=0,\ J^*(t+D_j)=j,\ Q(t+D_j)=0 \mid J(t)=0,\ J^*(t)=i,\ Q(t)=0)=e^{-\lambda D_j}$. If $i \neq j$, the event $(J(t)=0,\ J^*(t)=i,\ Q(t)=0)$ and the event $(J(t+D_j)=0,\ J^*(t+D_j)=j,\ Q(t+D_j)=0)$ occur only when a service with length of D_j is completed between time t and time $t+D_j$. However, this is impossible and therefore $P(J(t+D_j)=0,\ J^*(t+D_j)=j \mid J(t)=0,\ J^*(t)=i)=0$.

(ii) Given that the event $(J(t)=i,\ Q(t)=0,\ \eta(t)>x)$ occurs, the event $(J(t+D_j)=0,\ J^*(t+D_j)=j,\ Q(t+D_j)=0)$ occurs if no primary customers arrive between time t and time $t+D_j$. Therefore, $P(J(t+D_j)=0,\ J^*(t+D_j)=j,\ Q(t+D_j)=0 \mid J(t)=i,\ Q(t)=0,\ \eta(t)>x)=e^{-\lambda D_j}$. If $i \neq j$, since the event $(J(t)=i,\ J(t+D_j)=0,\ J^*(t+D_j)=j)$ does not occur, $P(J(t+D_j)=0,\ J^*(t+D_j)=j,\ Q(t+D_j)=0 \mid J(t)=i,\ Q(t)=0,\ \eta(t)>x)=0$.

Corollary 8.2 For $j=1,2,...,$

$$P_{Ej}(t+D_j)=e^{-\lambda D_j}(P_{Ej}(t)+\int_0^{D_j} P_{B0j}(t,\,x)\,dx). \tag{8.1}$$

Proof. This is easily shown by Lemma 8.1.

Lemma 8.2 For $i=1,2,...,\ j=1,2,...,$

$$\begin{aligned} P_{Iij}(t+D_j,\,0) &= \lambda p_j\frac{(\lambda D_j)^i}{i!}e^{-\lambda D_j}P_E(t) \\ &+\sum_{k=1}^{i}\int_0^\infty(\lambda p_j\frac{(\lambda D_j)^{i-k}}{(i-k)!}e^{-\lambda D_j}+\alpha(x)p_j\frac{(\lambda D_j)^{i-k+1}}{(i-k+1)!}e^{-\lambda D_j})P_{Ik}(t,\,x)\,dx \\ &+\int_0^\infty \alpha(x)p_je^{-\lambda D_j}P_{Ii+1}(t,\,x)\,dx. \end{aligned} \tag{8.2}$$

Proof. Note that $P_{Iij}(t+D_j,0) = \lim_{h\to 0}\frac{g(h)}{h}$, where $g(h) = P(J(t+D_j) = 0,\ J^*(t+D_j) = j,\ Q(t+D_j) = i,\ 0 < \xi(t+D_j) < h)$. We will consider the contribution to $g(h)$ for the following cases (i) through (iii).

(i) The probability that the system is empty at time $t-h$ is $P_E(t-h)$. A primary customer arrives between time $t-h$ and time t with probability $\lambda h + o(h)$. It obtains service time D_j with probability p_j. Then i primary customers arrive during its service time with probability $\frac{(\lambda D_j)^i}{i!}e^{-\lambda D_j}$. Therefore, the contribution to $g(h)$ is given by

$$g_1(h) = \lambda h p_j \frac{(\lambda D_j)^i}{i!} e^{-\lambda D_j} P_E(t-h) + o(h).$$

(ii) The probability that the server is idle at time $t-h$ and there are k $(k = 1, 2, \ldots, i)$ customers in the retrial queue is $\int_0^\infty P_{Ik}(t-h,x)dx$. Given that the elapsed retrial time at time $t-h$ equals x, a primary customer arrives between time $t-h$ and time t with probability $\lambda h + o(h)$, while a repeated customer arrives between time $t-h$ and time t with probability $\alpha(x)h + o(h)$. Each of them obtains service time D_j with probability p_j. If a primary customer obtains the service, then $i-k$ primary customers arrive during its service time with probability $\lambda h p_j \frac{(\lambda D_j)^{i-k}}{(i-k)!}e^{-\lambda D_j}$. Otherwise, if a repeated customer obtains the service, $i-k+1$ primary customers arrive during its service time with probability $\lambda h p_j \frac{(\lambda D_j)^{i-k+1}}{(i-k+1)!}e^{-\lambda D_j}$. Therefore, the contribution to $g(h)$ is given by

$$g_k(h) = \int_0^\infty (\lambda h p_j \frac{(\lambda D_j)^{i-k}}{(i-k)!} e^{-\lambda D_j} + \alpha(x) h p_j \frac{(\lambda D_j)^{i-k+1}}{(i-k+1)!} e^{-\lambda D_j}) P_{Ik}(t-h,x)\,dx + o(h).$$

(iii) The probability that the server is idle at time $t-h$ and there are $i+1$ customers in the retrial queue is $\int_0^\infty P_{Ii+1}(t-h,\ x)\,dx$. Given that the elapsed retrial time at time $t-h$ equals x, a repeated customer arrives between time $t-h$ and time t with probability $\alpha(x)h + o(h)$. It obtains service time D_j with probability p_j. During its service time, no primary customers arrive with probability $e^{-\lambda D_j}$. Therefore, the contribution to $g(h)$ is given by

$$g_{i+1}(h) = \int_0^\infty \alpha(x) h p_j e^{-\lambda D_j} P_{Ii+1}(t-h,\, x)\, dx + o(h).$$

Summarizing the results of (i), (ii) and (iii), we obtain

$$g(h) = g_1(h) + \sum_{k=1}^{i} g_k(h) + g_{i+1}(h),$$

which leads to (8.2).

We are interested in the limiting behavior of the queueing process. Assume that $\lambda \sum_{j=1}^{\infty} p_j D_j < L_A(\lambda)$. Then $\lim_{t\to\infty} P_E(t), \lim_{t\to\infty} P_{Ej}(t), \lim_{t\to\infty} P_{Ii}(t,\, x)$, $\lim_{t\to\infty} P_{Iij}(t,\, x)$ and $\lim_{t\to\infty} P_{Bij}(t,\, x)$ exist and we denote them by $P_E, P_{Ej}, P_{Ii}(x)$, $P_{Iij}(x)$ and $P_{Bij}(x)$. By using the results above and the supplementary variable technique, we obtain the following system of steady state equations related to the model.

$$(e^{\lambda D_j} - 1) P_{Ej} = \int_0^{D_j} P_{B0j}(t,\, x)\, dx, \tag{8.3}$$

$$(\frac{d}{dx} + \lambda + \alpha(x)) P_{Iij}(x) = 0, \tag{8.4}$$

$$(\frac{d}{dx} + \lambda) P_{B0j}(x) = 0, \tag{8.5}$$

$$(\frac{d}{dx} + \lambda) P_{Bij}(x) = \lambda P_{Bi-1j}(x),\ i = 1, 2, ..., \tag{8.6}$$

$$\begin{aligned} P_{Iij}(0) = {} & \lambda p_j \frac{(\lambda D_j)^i}{i!} e^{-\lambda D_j} P_E + \sum_{k=1}^{i} \int_0^\infty (\lambda p_j \frac{(\lambda D_j)^{i-k}}{(i-k)!} e^{-\lambda D_j} \\ & + \alpha(x) p_j \frac{(\lambda D_j)^{i-k+1}}{(i-k+1)!} e^{-\lambda D_j}) P_{Ik}(x)\, dx + \int_0^\infty \alpha(x) p_j e^{-\lambda D_j} P_{Ii+1}(x)\, dx, \end{aligned} \tag{8.7}$$

$$P_{B0j}(0) = \lambda p_j P_E + \int_0^\infty \alpha(x) p_j P_{I1}(x)\, dx, \tag{8.8}$$

$$P_{Bij}(0) = \int_0^\infty \lambda p_j P_{Ii}(x)\, dx + \int_0^\infty \alpha(x) p_j P_{Ii+1}(x)\, dx,\ i = 1, 2, ..., \tag{8.9}$$

$$P_E + \sum_{i=1}^\infty \sum_{j=1}^\infty P_{Iij}(x)\, dx + \sum_{i=0}^\infty \sum_{j=1}^\infty \int_0^{D_j} P_{Bij}(x)\, dx = 1. \tag{8.10}$$

In order to solve the system of equations above, we introduce the following probability generating functions:

$$P_I(z,\, x) = \sum_{i=1}^\infty \sum_{j=1}^\infty P_{Iij}(x) z^i, \quad P_{Bj}(z,\, x) = \sum_{i=0}^\infty P_{Bij}(x) z^i,\ j = 1, 2, ...$$

When both sides of the equations (8.4), (8.6), (8.7) and (8.9) are multiplied by z^i and both sides of the equations (8.4) and (8.7) are summed over i and j and both sides of the equations (8.6) and (8.9) are summed over i, we obtain the following equations:

$$(\frac{\partial}{\partial x} + \lambda + \alpha(x)) P_I(z,\, x) = 0, \tag{8.11}$$

$$(\frac{\partial}{\partial x} + \lambda) P_{Bj}(z, x) = \lambda z P_{Bj}(z,\, x), \tag{8.12}$$

$$\begin{aligned} P_I(z,\, 0) &= \sum_{j=1}^\infty (\lambda p_j e^{-\lambda D_j}(e^{\lambda D_j z} - 1) P_E + p_j e^{-\lambda D_j (1-z)} \\ &\times \int_0^\infty (\lambda + \frac{\alpha(x)}{z}) P_I(z,\, x)\, dx - \int_0^\infty \alpha(x) p_j e^{-\lambda D_j} P_{I1}(x)\, dx), \end{aligned} \tag{8.13}$$

$$P_{Bj}(z,\, 0) = \lambda p_j P_E + \lambda p_j \int_0^\infty P_I(z,\, x)\, dx + \frac{1}{z}\int_0^\infty \alpha(x) p_j P_I(z,\, x)\, dx. \tag{8.14}$$

It is easily seen that (3.10) becomes

$$P_E + \lim_{z\to 1^-} (\int_0^\infty P_I(z, x)\, dx + \sum_{j=1}^\infty \int_0^{D_j} P_{Bj}(z,\, x)\, dx) = 1. \tag{8.15}$$

From (8.5), we get

$$P_{B0j}(x) = P_{B0j}(0) e^{-\lambda x}. \tag{8.16}$$

This and (3.3) lead to

$$P_{B0j}(0) = \lambda e^{\lambda D_j} P_{Ej}, \tag{8.17}$$

which yields

$$\int_0^\infty \alpha(x) p_j P_{I1}(x)\, dx = \lambda(e^{\lambda D_j} P_{Ej} - p_j P_E). \tag{8.18}$$

Also

$$P_I(z,\, x) = P_I(z,\, 0) e^{-\lambda x}(1 - A(x)), \tag{8.19}$$

and

$$P_{Bj}(z,\, x) = P_{Bj}(z,\, 0) e^{-\lambda(1-z)x}. \tag{8.20}$$

Therefore, (8.13) becomes

$$P_I(z,\, 0) = \sum_{j=1}^{\infty} (\lambda(p_j e^{-\lambda D_j(1-z)} - 1) P_E + (1 - (1 - \frac{1}{z}) L_A(\lambda)) p_j e^{-\lambda D_j(1-z)} P_I(z,\, 0)), \tag{8.21}$$

or equivalently,

$$P_I(z,\, 0) = \frac{\lambda z (1 - \sum_{j=1}^\infty p_j e^{-\lambda D_j(1-z)}) P_E}{L_A(\lambda)(1-z) \sum_{j=1}^\infty p_j e^{-\lambda D_j(1-z)} - z(1 - \sum_{j=1}^\infty p_j e^{-\lambda D_j(1-z)})}. \tag{8.22}$$

Similarly, we get

$$P_{Bj}(z,\, 0) = \frac{\lambda p_j L_A(\lambda)(1-z) P_E}{L_A(\lambda)(1-z) \sum_{j=1}^\infty p_j e^{-\lambda D_j(1-z)} - z(1 - \sum_{j=1}^\infty p_j e^{-\lambda D_j(1-z)})}. \tag{8.23}$$

Therefore, by substituting (8.22) and (8.23) into (8.19) and (8.20), respectively, and substituting their results into (8.15), we obtain

$$P_E = 1 - \frac{\lambda \sum_{j=1}^\infty p_j D_j}{L_A(\lambda)}.$$

Thus, we have proved the following theorem.

Theorem 8.2. If $\lambda\sum_{j=1}^{\infty}p_jD_j < L_A(\lambda)$, then the joint steady-state distribution of the state of the server and the length of the retrial queue is, in terms of the probability generating functions Laplace transform given by

$$P_E = 1 - \frac{\lambda\sum_{j=1}^{\infty}p_jD_j}{L_A(\lambda)}, \tag{8.24}$$

$$P_I(z,\, x) = \frac{\lambda z(1-\sum_{j=1}^{\infty}p_je^{-\lambda D_j(1-z)})e^{-\lambda x}(1-A(x))P_E}{L_A(\lambda)(1-z)\sum_{j=1}^{\infty}p_je^{-\lambda D_j(1-z)} - z(1-\sum_{j=1}^{\infty}p_je^{-\lambda D_j(1-z)})}, \tag{8.25}$$

$$P_{Bj}(z,\, x) = \frac{\lambda p_jL_A(\lambda)(1-z)e^{-\lambda(1-z)x}P_E}{L_A(\lambda)(1-z)\sum_{j=1}^{\infty}p_je^{-\lambda D_j(1-z)} - z(1-\sum_{j=1}^{\infty}p_je^{-\lambda D_j(1-z)})},\ j=1,2,.... \tag{8.26}$$

By using the results of Theorem 8.2, we will derive some important performance measures.

Corollary 8.2 If $\lambda\sum_{j=1}^{\infty}p_jD_j < L_A(\lambda)$, then

(i) $P_E = 1 - \frac{\lambda\sum_{j=1}^{\infty}p_jD_j}{L_A(\lambda)}$,

(ii) $P_I = \frac{(1-L_A(\lambda))\lambda\sum_{j=1}^{\infty}p_jD_j}{L_A(\lambda)}$,

(iii) $P_B = \lambda\sum_{j=1}^{\infty}p_jD_j$,

where in steady state, $P_E = P($ the system is empty$)$, $P_I = P$ (the system is non-empty and the server is idle) and $P_B = P($the server is busy$)$.

Proof. It is easily seen that $P_I = \lim_{z\to 1-}\int_0^{\infty}P_I(z,\, x)\,dx$ and $P_B = \lim_{z\to 1-}\sum_{j=1}^{\infty}\int_0^{D_j}P_{Bj}(z,\, x)\,dx$. By substituting the results of Theorem 8.2 into the formulas, we can obtain the explicit expressions for P_I and P_B through direct calculation.

Corollary 8.3 For an $M/\{D_n\}/1$ regular(non retrial) queueing system, if $\lambda \sum_{j=1}^{\infty} p_j D_j < 1$, then

(i) $P_E^{(c)} = 1 - \lambda \sum_{j=1}^{\infty} p_j D_j$,

(ii) $P_B^{(c)} = \lambda \sum_{j=1}^{\infty} p_j D_j$,

where in steady state, $P_E^{(c)} = P($ the system is empty) and $P_B^{(c)} = P($the server is busy).

Proof. Let $A(x) = 1 - e^{-\alpha x}$ and let $\alpha \to \infty$. By using the results of Corollary 8.2, we get our results.

Corollary 8.4 Let N_q and N be the number of the customers in the retrial queue and in the system, respectively, in steady state. Then N_q and N have distributions, in terms of generating functions given by

$$p_q(z) \equiv Ez^{N_q} = \frac{(L_A(\lambda) - \lambda \sum_{j=1}^{\infty} p_j D_j)(1-z)}{L_A(\lambda)(1-z)\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)} - z(1 - \sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)})},$$

$$p(z) \equiv Ez^{N} = \frac{(L_A(\lambda) - \lambda \sum_{j=1}^{\infty} D_j p_j)\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)}(1-z)}{L_A(\lambda)(1-z)\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)} - z(1 - \sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)})},$$

respectively.

Also, their expected values are, respectively, given by

$$EN_q = \frac{\lambda^2 \sum_{j=1}^{\infty} p_j {D_j}^2 + 2\lambda(1 - L_A(\lambda))\sum_{j=1}^{\infty} p_j D_j}{2(L_A(\lambda) - \lambda \sum_{j=1}^{\infty} p_j D_j)},$$

$$EN = \frac{\lambda^2 \sum_{j=1}^{\infty} p_j {D_j}^2 + 2\lambda \sum_{j=1}^{\infty} p_j D_j(1 - \lambda \sum_{j=1}^{\infty} p_j D_j)}{2(L_A(\lambda) - \lambda \sum_{j=1}^{\infty} p_j D_j)}.$$

Proof. Note that

$$p_q(z) = P_E + \int_0^\infty P_I(z,x)\,dx + \sum_{j=1}^{\infty}\int_0^{D_j} P_{Bj}(z,\,x)\,dx,$$

$$p(z) = P_E + \int_0^\infty P_I(z,x)\,dx + \sum_{j=1}^{\infty}\int_0^{D_j} zP_{Bj}(z,\,x)\,dx.$$

The substitution of the steady state distribution of Theorem 8.2 into the formulas above gives our results.

From the results of Corollary 8.3, we have $p(z) = p_q(z)\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)}$.

Corollary 8.5 For an $M/\{D_n\}/1$ regular (non retrial) queueing model, let $N_q^{(c)}$ and $N^{(c)}$ be the number of the customers in the queue and in the system, respectively. If $\lambda\sum_{j=1}^{\infty} p_j D_j < 1$, then the distributions of $N_q^{(c)}$ and $N^{(c)}$ have the distributions, in terms of generating functions given by

$$p_q^{(c)}(z) = \frac{(1-\lambda\sum_{j=1}^{\infty} p_j D_j)(1-z)}{\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)} - z},$$

$$p^{(c)}(z) = \frac{(1-\lambda\sum_{j=1}^{\infty} D_j p_j)\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)}(1-z)}{\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-z)} - z},$$

respectively.

Also, their expected values are, respectively, given by

$$EN_q^{(c)} = \frac{\lambda^2\sum_{j=1}^{\infty} p_j{}^2}{D_j} 2(1-\lambda\sum_{j=1}^{\infty} p_j D_j),$$

$$EN^{(c)} = \frac{\lambda^2\sum_{j=1}^{\infty} p_j D_j{}^2 + 2\lambda\sum_{j=1}^{\infty} p_j D_j(1-\lambda\sum_{j=1}^{\infty} p_j D_j)}{2(1-\lambda\sum_{j=1}^{\infty} p_j D_j)},$$

Proof. Let $A(x) = 1 - e^{-\alpha x}$ and let $\alpha \to \infty$. By using the results of Corollary 8.4, we get our results.

The result of Corollary 8.5 is interesting. First, the result can be (and has been) checked using the Pollaczek Khinchin formula. The results can also be compared to the results obtained for the waiting time distribution as obtained by Shortle and Brill (2005) for the $M/\{iD\}/1$ queue. The derivation of waiting time distribution is much more complex than its expected value. Shortle and Brill used sums of partitions for their formula. This implies that counting over the partitions simplifies greatly when computing expected values. The reasons for this may be of interest in number theory.

Chapter 9

Model IV Waiting Time and Busy Period

9.1 Steady-state distribution of the waiting time

The following theorem gives the joint steady-state distribution of the waiting time that a primary customer spends in the retrial queue and the number of customers served during the waiting time.

Theorem 9.1. In steady state, let W represent the waiting time that a primary customer spends in the retrial queue and N represent the number of customers served during this waiting time. Then the joint distribution of W and N in terms of generating function Laplace transform is

$$L(s,z) \equiv E(e^{-sW}z^N) = P_E + P_I + \frac{\lambda L_1(s,z) z L_A(\lambda)(1 - L_1^*(s,z))P_E}{\lambda(1 - L_1^*(s,z)) - s}$$
$$\times \frac{\sum_{j=1}^{\infty} p_j e^{-sD_j}(1 - e^{-(\lambda(1-L_1^*(s,z))-s)D_j})}{L_A(\lambda)(1 - L_1^*(s,z))\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-L_1^*(s,z))} - L_1^*(s,z)(1 - \sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-L_1^*(s,z))})}, \tag{9.1}$$

where

$$L_1^*(s,z) = \sum_{j=1}^{\infty} p_j e^{-sD_j} z L_1(s,z), \tag{9.2}$$

$$L_1(s,z) = \frac{(s+\lambda)L_A(s+\lambda)}{s+\lambda-\lambda\sum_{j=1}^{\infty} p_j e^{-sD_j} z(1-L_A(s+\lambda))}. \tag{9.3}$$

Proof. Note that when a primary customer (called PC) arrives, the system is empty, or the system is nonempty and the server is idle, or the server is busy. Therefore,

$$\begin{aligned} L(s,\, z) &= P_E E(e^{-sW} z^N \mid J=0,\, Q=0) \\ &+ \sum_{i=1}^{\infty} \int_0^{\infty} P_{Ii}(x) E(e^{-sW} z^N \mid J=0,\, Q=i,\, \xi=x)\, dx \\ &+ \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \int_0^{D_j} P_{Bij}(x) E(e^{-sW} z^N \mid J=j,\, Q=i,\, \eta=x)\, dx, \end{aligned} \tag{9.4}$$

where J, Q, ξ and η denote, respectively, the server state, the number of customers in the retrial queue, the elapsed retrial time and the elapsed service time when PC arrives to the system in steady state. It is obvious that $E(e^{-sW} z^N \mid J=0,\, Q=0) = E(e^{-sW} z^N \mid J=0,\, Q=i,\, \xi=x) = 1$. Now let us consider the state of the server busy when PC arrives. Suppose that a customer (called SC) is being served with service time D_j and the elapsed service time is x and there are i ($i = 0, 1, ...$) customers in the retrial orbit when PC arrives. Then PC enters the retrial orbit and becomes the $(i+1)$th retrial customer. For convenience, let these $(i+1)$ retrial customers be numbered in order of their arrival. Let S_k denote the service time of the kth retrial customer, $k = 1, 2, ...i$. Let W_1 denote the waiting time that the 1*st* retrial customer spends in the retrial queue measured from the instant that SC completes its service and W_k denote the waiting time that the kth retrial customer spends in the retrial queue measured from the instant that the $(k-1)$th retrial customer completes its service, $k = 2, 3, ..., i+1$. Let N_k denote the number of customers served during W_k, $k = 1, 2, ..., i+1$. Then $W = D_j - x + \sum_{k=1}^{i} S_k + \sum_{k=1}^{i+1} W_k$ and $N = 1 + i + \sum_{k=1}^{i+1} N_k$. Hence,

$$E(e^{-sW} z^N \mid J = j,\, Q = i,\, \eta = x) = e^{-s(D_j - x)} z (\sum_{k=1}^{\infty} p_k e^{-sD_k} z)^i \, E(e^{-s\sum_{k=1}^{i+1} W_k} z^{\sum_{k=1}^{i+1} N_k}). \tag{9.5}$$

From the model assumptions, we see that $(W_1, N_1), (W_2, N_2), ..., (W_{i+1}, N_{i+1})$ are independent and identically distributed with generating function Laplace transform

$$\begin{aligned} L_1(s,z) &\equiv E(e^{-sW_1} z^{N_1}) = \sum_{n=0}^{\infty} \int_0^{\infty} \int_y^{\infty} e^{-sy} \lambda e^{-\lambda x} \, dx \, dA(y) \\ &\times (\int_0^{\infty} \int_0^{y} e^{-sx} \lambda e^{-\lambda x} \sum_{k=1}^{\infty} p_k e^{-sD_k} z \, dx \, dA(y))^n \\ &= \sum_{n=0}^{\infty} L_A(s+\lambda) (\frac{\lambda \sum_{k=1}^{\infty} p_k e^{-sD_k} z}{s+\lambda} (1 - L_A(s+\lambda)))^n , \end{aligned} \tag{9.6}$$

Then (9.3) follows from (9.6). Also $E(e^{-s\sum_{k=1}^{i+1} W_k} z^{\sum_{k=1}^{i+1} N_k}) = (L_1(s,\, z))^{i+1}$. Using (9.4) and the results above, we obtain

$$\begin{aligned} L(s,z) &= P_E + \sum_{i=1}^{\infty} \int_0^{\infty} P_{Ii}(x) \, dx \\ &+ \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} e^{-sD_j} z L_1(s,z) \int_0^{D_j} P_{Bij}(x) e^{sx} (L_1^*(s,z))^i \, dx \\ &= P_E + P_I + \sum_{j=1}^{\infty} e^{-sD_j} z L_1(s,z) \int_0^{D_j} P_{Bj}(L_1^*(s,z), x) e^{sx} \, dx. \end{aligned} \tag{9.7}$$

Thus, with the help of the steady-state solution in Theorem 8.2, we obtain (9.1) .

Shortle and Brill [109] gave an analytical solution and numerical algorithms to calculate the distribution of the waiting time in the $M/\{iD\}/1$ queue and $M/D_N/1$ queue. However, the formula may be difficult to compute because it involves partitions of integers, and numerical stability issues. It would be interesting later to compare the results in our model with the results the authors obtained.

Corollary 9.1 For an $M/\{D_n\}/1$ regular (non retrial) queueing model, if $\lambda\sum_{j=1}^{\infty} p_j D_j < 1$, then in steady state, the joint distribution of the waiting time and the number of customers served during it has generating function Laplace transform

$$L^{(c)}(s,z) = P_E^{(c)} + \frac{(\lambda z(1-L^{(c*)}(s,z))\sum_{j=1}^{\infty} p_j e^{-sD_j}(1-e^{-(\lambda(1-L^{(c*)}(s,z))-s)D_j})P_E^{(c)}}{(\lambda(1-L^{(c*)}(s,z))-s)(\sum_{j=1}^{\infty} p_j e^{-\lambda D_j(1-L^{(c*)}(s,z))} - L^{(c*)}(s,z))}, \tag{9.8}$$

where

$$L^{(c*)}(s,z) = \sum_{j=1}^{\infty} p_j e^{-sD_j} z. \tag{9.9}$$

Proof. Let $A(x) = 1 - e^{-\alpha x}$ and let $\alpha \to \infty$. By using the results of Theorem 9.1, we get our results.

9.2 The busy period

A (primary) busy period is defined as the time period from when a primary customer arrives and finds the system empty until the system is first empty again. Such a primary customer is called IPC (an initial primary customer). Any customer arriving during the service time of IPC is called an IRC (an initial repeated customer).

Suppose that there are k IRCs, say Λ_i, $i = 1, 2, ..., k$, in the system. Let t_0 denote the departure point of IPC and t_i denote the departure point of Λ_i. Define $c_i \equiv t_i - t_{i-1}$ to be the effective waiting time of Λ_i. This effective waiting time will consist of at least one idle time and at least one service time. There may be more depending on whether or not primary customers also enter the system. Suppose that there are n_i primary customers, say Λ_{ij}, $j = 1, 2, ..., n_i$, arriving during the effective waiting time of Λ_i and let t_{ij} denote the departure point of Λ_{ij}. We call $c_{ij} \equiv t_{ij} - t_{ij-1}$ the effective waiting time of Λ_{ij}, where $t_{10} = t_k$, $t_{i0} = t_{i-1n_{i-1}}$, $i = 2, 3, ...$ We continue similarly. We call Λ_{ij} the first generation offspring of Λ_i. Similarly we can define second generation offspring of Λ_i, third generation offspring of Λ_i and so on. We define Λ_i and all its generations of offspring as a family from Λ_i and define the total effective waiting time of all members of a family from Λ_i as the ith retrial busy period.

It can be seen that a busy period in the retrial queue is different from that in an ordinary queueing system. The server is always busy in a busy period of the ordinary queueing system, while the server alternates between a busy state and an idle state (with the retrial queue nonempty) where the first state and the last state are busy states in a busy period of the retrial queue. The following result gives a functional equation which is useful for describing the system.

Theorem 9.2 Let T represent the length of a busy period and K represent the number of customers served during the busy period in steady state. Then the joint distribution of T and K, in terms of the probability generating function Laplace transform, is

$$\Psi(s, z) \equiv E(e^{-sT} z^K) = \sum_{j=1}^{\infty} p_j e^{-(s+\lambda-\lambda\Psi^*(s,z))D_j}, \tag{9.10}$$

where

$$\Psi^*(s, z) = \frac{(s+\lambda)L_A(s+\lambda))\Psi(s, z)}{s+\lambda-\lambda(1-L_A(s+\lambda))\Psi(s, z)}. \tag{9.11}$$

Proof. Suppose that IPC takes a service time D_j. If no primary customers arrive during its service time, then $T = D_j$ and $K = 1$. Otherwise, $n\,(n = 1, 2, ...)$ primary customers arrive and each one becomes an IRC. Then $T = D_j + T_1 + T_2 + ... + T_n$ and $K = 1 + K_1 + K_2 + ... + K_n$, where T_j and K_j represent the length of the jth retrial busy period and the number of customers served during it, respectively, $j = 1, 2, ..., n$. Therefore,

$$\Psi(s, z) = \sum_{j=1}^{\infty}\sum_{n=0}^{\infty} p_j \frac{(\lambda D_j)^n}{n!} z e^{-\lambda D_j} e^{-sD_j} E(e^{-s(T_1+T_2+...+T_n)} z^{K_1+K_2+...+K_n}), \tag{9.12}$$

where $T_0 = 0$ and $K_0 = 0$.

By the assumptions of the model, it can be seen that the primary arrival process is independent of the retrials and service and that T_j consists of the sum of disjoint time intervals. Suppose that only one repeated customer is in the retrial queue during the service time of IPC and let T_* be the length of the retrial busy period and N_* be the number of customers served during it. It can be seen that $(T_1, K_1), (T_2, K_2), ..., (T_n, K_n)$ are independent and have the same distribution as (T_*, K_*). Therefore,

$$E(e^{-s(T_1+T_2+...+T_n)} z^{K_1+K_2+...+K_n}) = (E(e^{-sT_*} z^{K_*}))^n. \tag{9.13}$$

Thus, (5.3) becomes

$$\Psi(s, z) = \sum_{j=1}^{\infty}\sum_{n=0}^{\infty} p_j \frac{(\lambda D_j)^n}{n!} e^{-(\lambda+s)D_j} z(\Psi^*(s, z))^n, \tag{9.14}$$

where $\Psi^*(s,\, z) = E(e^{-sT_*} z^{K_*})$.

In order to get $\Psi^*(s, z)$, we must analyze the evolution of the process from the time that IPC leaves the system. There exists a competition between an arriving primary customer and this IRC from the time that IPC completes its service. If this IRC wins, then T_* is its retrial time plus a random variable which has the same distribution as T, and K_* is a random variable which has the same distribution as K. Otherwise, this IRC returns to the first position in the retrial queue, T_* is the inter-arrival time plus the sum of two independent random variables, of which one has the same distribution as T and the other has the same distribution as T_*. Here K_* is equal to the sum of two random variables, one of which has the same distribution as K and the other has the same distribution as K_*. Thus,

$$\Psi^*(s,\, z) = \int_0^\infty \int_y^\infty \lambda e^{-\lambda x - sy} \Psi(s,\, z)\, dA(y)\, dx$$
$$+ \int_0^\infty \int_0^y \lambda e^{-\lambda x - sx} \Psi(s,\, z) \Psi^*(s,\, z)\, dA(y)\, dx. \tag{9.15}$$

Since

$$\int_0^\infty \int_y^\infty \lambda e^{-\lambda x - sy} \Psi(s,\, z)\, dA(y)\, dx = L_A(s + \lambda), \tag{9.16}$$

and

$$\int_0^\infty \int_0^y \lambda e^{-\lambda x - sx} \Psi(s, z)\, dA(y)\, dx = 1 - L_A(s + \lambda), \tag{9.17}$$

it follows that

$$\Psi^*(s,\, z) = L_A(s + \lambda)\Psi(s,\, z) + \frac{\lambda}{s + \lambda}(1 - L_A(s + \lambda))\Psi(s,\, z)\Psi^*(s,\, z), \tag{9.18}$$

which yields (9.11) and therefore the substitution of (9.11) into (9.14) yields (9.10).

Corollary 9.2 For an $M/\{D_n\}/1$ regular (non retrial) queueing model, if $\lambda \sum_{j=1}^{\infty} p_j D_j < 1$, then in steady state, the joint distribution of the length of a busy period and the number of customers served during it has probability generating function Laplace transform

$$\Psi^{(c)}(s,\, z) = \sum_{j=1}^{\infty} p_j e^{-(s+\lambda-\lambda\Psi^{(c)}(s,\, z))D_j}. \tag{9.19}$$

Proof. Let $A(x) = 1 - e^{-\alpha x}$ and let $\alpha \to \infty$. By using the results of Theorem 9.2, we get our result.

Chapter 10

Graphical Results for Models I and II

10.1 $\mathrm{E(W_q)}$ versus different combinations of parameters for retrials and service

In this chapter, several examples illustrate the influence of different parameters on W_q in different combinations of retrial times and service times, for the models described in chapter 3 and 5. Here we display graphs of $E(W_q)$ for the following cases.

(i) The retrial time is exponentially distributed with density function $a_1(x) = \alpha e^{-\alpha x}$.

(ii) The retrial time is Erlang distributed with density function $a_2(x) = 4\alpha^2 x e^{-2\alpha x}$.

(iii) The retrial time is hyperexponentially distributed with density function $a_3(x) = \frac{2}{9}\alpha(e^{-\frac{2\alpha x}{3}} + 4e^{-\frac{4\alpha x}{3}})$.

We begin with Model I from Chapter 3. Figure 10.1 displays $E(W_q)$ versus (α, λ) for the set of parameters $(\beta_1, \beta_2, \mu, \theta, \gamma_1, \gamma_2) = (0.001, 0.1, 2, 1, 0.001, 0.1)$. From the expression for $E(W_q)$ in Chapter 3, we do not need to know the full distribution of the service time. We only need the first two moments of the service time β_1, and β_2. Here Figure 10.1 (A_1) considers the case of an exponential retrial distribution. We treat μ as fixed in this case. It can be seen that $E(W_q)$ shows a downward trend as α increases (since the retrial is more likely to win a competition against a primary customer) and an upward trend as λ increases (since a primary customer is more likely to win a competition against a retrial customer). Figure 10.1 (A_2) and Figure 10.1 (A_3) consider the cases of an Erlang retrial distribution and a hyperexponential retrial distribution. In each of these two cases, there is a downward trend for $E(W_q)$ as α increases and an upward trend for $E(W_q)$ as λ increases. In our examples, the parameters have been chosen so that the expected value of the retrial time is $\frac{1}{\alpha}$ in all three cases (exponential, Erlang, hyperexponential). It turns out that the lowest expected time in the retrial queue occurs for the hyperexponential case and the highest expected time in the retrial queue occurs for Erlang retrial times. Note that the expected time in the retrial queue is quite close for all three distributions as can be seen by the scale given on the vertical axis of the three diagrams.

Figure 10.2 displays $E(W_q)$ versus (μ, λ) for the fixed set of parameters $(\beta_1, \beta_2, \alpha, \theta, \gamma_1, \gamma_2) = (0.001, 0.1, 5, 1, 0.001, 0.1)$. Here we treat α as fixed and let μ vary. In each of the three cases, there exists an upward trend for $E(W_q)$ as μ increases (since more breakdowns slow the whole system) and an upward trend for $E(W_q)$ as λ increases, just as we would expect. Again there is very little difference in the results. However, the Erlang distribution again appears to be slightly higher than the other two, at least near $\mu = 1$ and $\lambda = 1.2$.

Figure 10.3 displays $E(W_q)$ versus $(\theta,\ \lambda)$ for the fixed set of parameters $(\beta_1,\ \beta_2,\ \alpha,\ \mu,\ \gamma_1,\ \gamma_2) = (0.225,\ 0.1,\ 3,\ 1,\ 0.01,\ 0.1)$. In this figure, we allow θ to vary. Since θ is the retrial orbit rate, we would guess that the expected wait $E(W_q)$ should decrease as θ increases. However, the amount of change of $E(W_q)$ for a small change in θ is quite large (and interesting). In each of the three cases, there exists a downward trend for $E(W_q)$ as θ increases (since this will speed up the service retrial orbit) and an upward trend for $E(W_q)$ as λ increases.

Figure 10.4 displays $E(W_q)$ versus $(\alpha,\ \mu)$ for the fixed set of parameters $(\beta_1,\ \beta_2,\ \lambda,\ \theta,\ \gamma_1,\ \gamma_2) = (0.01,\ 0.01,\ 0.21,\ 1.2,\ 0.001,\ 0.1)$. Here we treat α and μ as variables and the other parameters as fixed. In each of the three cases, there exists a downward trend for $E(W_q)$ as α increases and an upward trend for $E(W_q)$ as μ increases. From the scale on the vertical axis, we see that the change in the time spent in the retrial queue is very small compared to the relatively larger change in the rates α and μ. The same comment holds for Figure 10.1.

Figure 10.5 displays $E(W_q)$ versus $(\alpha,\ \theta)$ for the fixed set of parameters $(\beta_1,\ \beta_2,\ \lambda,\ \mu,\ \gamma_1,\ \gamma_2) = (0.001,\ 0.001,\ 2,\ 5,\ 0.001,\ 0.1)$. The three graphs again are quite similar in shape. This time, both the inner retrial rate and the outer retrial rate are shown. Note that the scale for α is quite different than in the other graphs. In each of the three cases, there exists a downward trend for $E(W_q)$ as α increases and a downward trend for $E(W_q)$ as θ increases. Near $\alpha = 2$ and $\theta = 2$ the function shows a large increase. In fact, $E(W_q)$ converges to ∞ as (α) tends to zero and θ tends to 0.

Figure 10.6 displays $E(W_q)$ versus $(\mu,\ \theta)$ for the fixed set of parameters $(\beta_1,\ \beta_2,\ \lambda,\ \alpha, \gamma_1,\ \gamma_2) = (0.01,\ 0.1,\ 2.6,\ 2,\ 0.01,\ 0.1)$. In each of these two cases, there is an upward trend for $E(W_q)$ as μ increases and a downward trend for $E(W_q)$ as θ increases. The graphs look close to being planes which suggests a linear approximation could be useful in this case. The middle graph (Erlang) shows the highest values for $E(W_q)$.

Figure 10.1: $E(W_q)$ versus (α, λ). (A_1) Exponential retrial distribution. (A_2) Erlang retrial distribution. (A_3) Hyper-exponential retrial distribution.

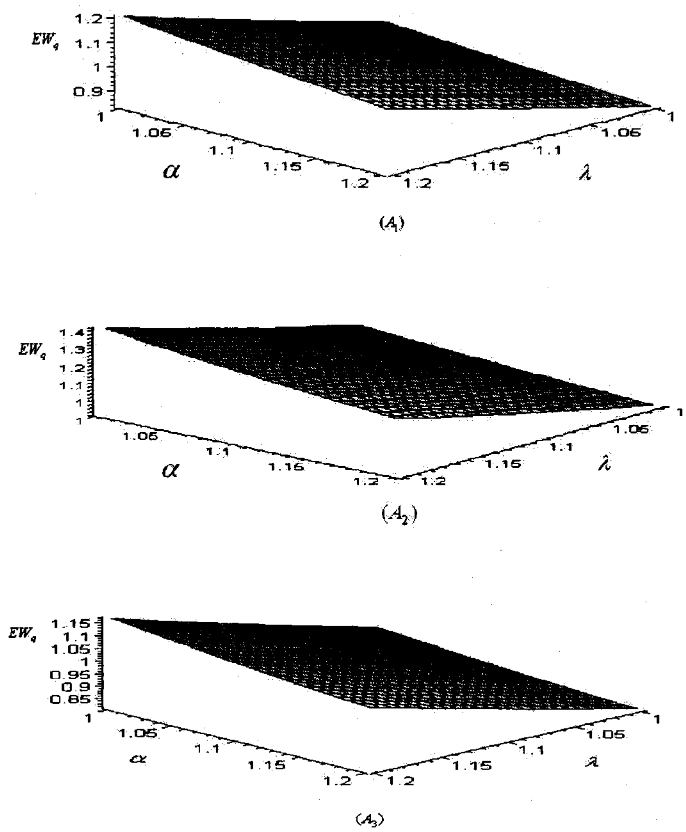


Figure 10.2: $E(W_q)$ versus (μ, λ). (B_1) Exponential retrial distribution. (B_2) Erlang retrial distribution. (B_3) Hyper-exponential retrial distribution.

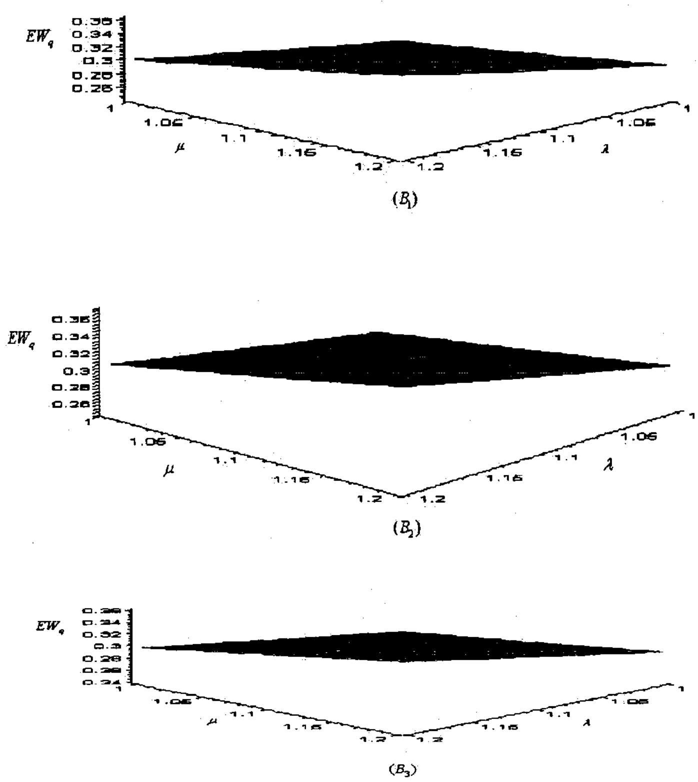
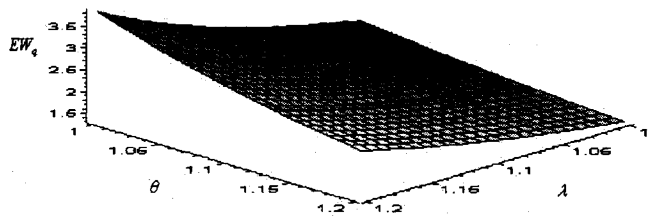
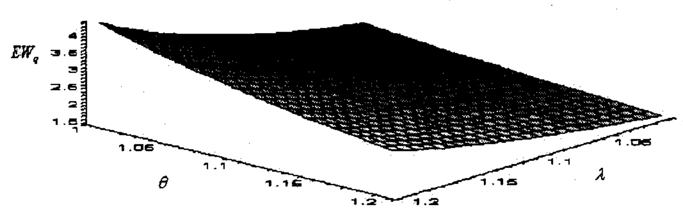


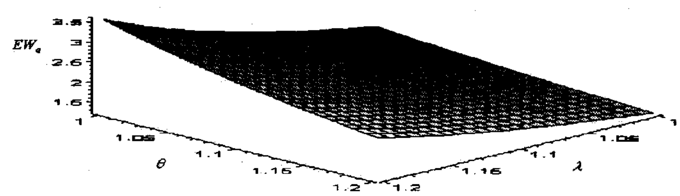
Figure 10.3: $E(W_q)$ versus (θ, λ). (C_1) Exponential retrial distribution. (C_2) Erlang retrial distribution. (C_3) Hyper-exponential retrial distribution.



(C_1)



(C_2)



(C_3)

Figure 10.4: $E(W_q)$ versus (α, μ). (D_1) Exponential retrial distribution. (D_2) Erlang retrial distribution. (D_3) Hyper-exponential retrial distribution.

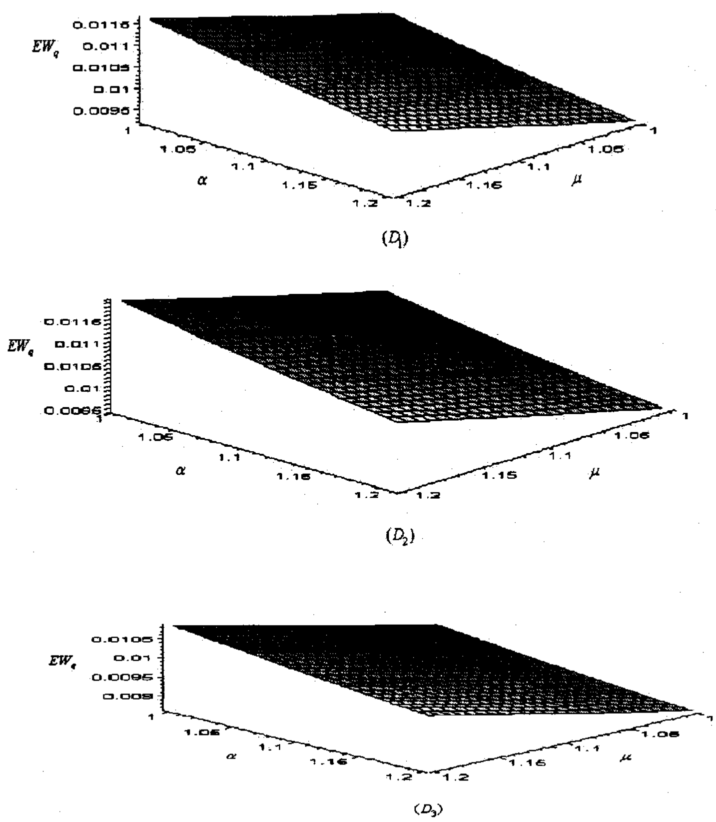


Figure 10.5: $E(W_q)$ versus (α, θ). (E_1) Exponential retrial distribution. (E_2) Erlang retrial distribution. (E_3) Hyper-exponential retrial distribution.

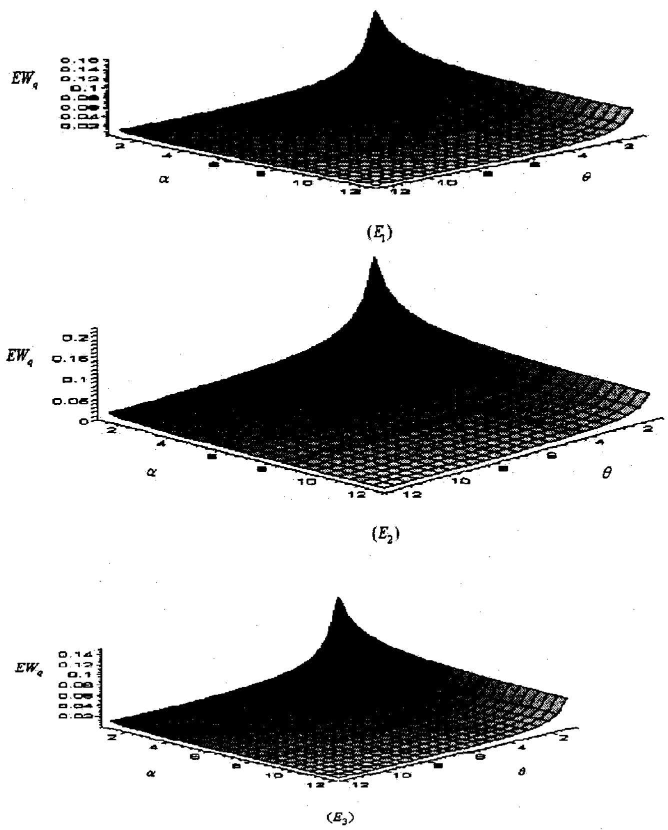
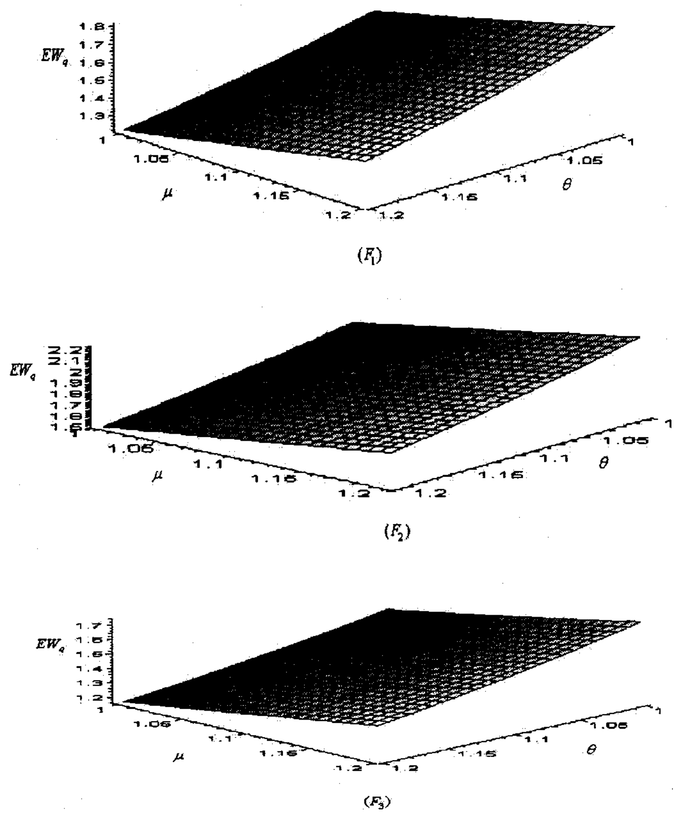


Figure 10.6: $E(W_q)$ versus (μ, θ). (F_1) Exponential retrial distribution. (F_2) Erlang retrial distribution. (F_3) Hyper-exponential retrial distribution.



10.2 $\mathbf{E(W_q)}$ versus different parameters: Two dimensional graphs, Model I.

In the previous section, the three dimensional graphs made it difficult to examine the changes caused by changing a single parameter because the three dimensions do not lend themselves to careful examination when printed on a two dimensional page. Therefore, in this section, several examples illustrate the influence of different single parameters on $E(W_q)$ for different combinations of retrial times and service times, for the model I of chapter 3. Here we display the graphs for the following cases.

(i) The retrial time is exponentially distributed with density function $a_1(x) = \alpha e^{-\alpha x}$.

(ii) The retrial time is Erlang distributed with density function $a_2(x) = 4\alpha^2 x e^{-2\alpha x}$.

(iii) The retrial time is hyperexponentially distributed with density function $a_3(x) = \frac{2}{9}\alpha(e^{-\frac{2\alpha x}{3}} + 4e^{-\frac{4\alpha x}{3}})$.

(iv) The service time is exponentially distributed with density function $b_1(x) = \frac{1}{\beta_1}e^{-\frac{x}{\beta_1}}$.

(v) The service time is Erlang distributed with density function $b_2(x) = \frac{4x}{\beta_1^2}e^{-\frac{2x}{\beta_1}}$.

(vi) The service time is hyperexponentially distributed with density function $b_3(x) = \frac{2}{9\beta_1}(e^{-\frac{2x}{3\beta_1}} + 4e^{-\frac{4x}{3\beta_1}})$.

Here we give three service time distributions. We have selected these so that all three have the same expected service time β_1. From the service time density functions, we can compute the second moment about 0, namely β_2. The three dimensional graphs given earlier are only partially successful in getting an understanding of the effects of changing parameters. Two dimensional graphs can help our understanding. We have 3 retrial time distributions. We have 3 service time distributions. We plot $E(W_q)$ against 6 parameters, namely α, β_1, γ_1, θ, λ, and μ. Since $3\times3\times6 = 54$, we need 54 graphs. But each diagram has 3 different graphs, so 54/3=27 diagrams will suffice. Each Figure has 3 such diagrams, so $27/3 = 9$ figures appear in the following pages.

For the set of parameters $(\alpha = 0.8, \beta_1 = 0.1, \gamma_1 = 0.4, \gamma_2 = 2, \mu = 0.01, \theta = 10)$. Figure 10.7 (A_1) displays $E(W_q)$ versus λ for $a_1(x)$ and $b_1(x)$, $a_2(x)$ and $b_1(x)$ and $a_3(x)$ and $b_1(x)$. Figure 10.7 (B_1) displays $E(W_q)$ versus λ for $a_1(x)$ and $b_2(x)$, $a_2(x)$ and $b_2(x)$ and $a_3(x)$ and $b_2(x)$. Figure 10.7 (C_1) displays $E(W_q)$ versus λ for $a_1(x)$ and $b_3(x)$, $a_2(x)$ and $b_3(x)$ and $a_3(x)$ and $b_3(x)$. It can be seen that in each case above, $E(W_q)$ shows an upward trend as λ increases, which is expected intuitively. We note that the three graphs of $E(W_q)$ vs λ for combinations of hyperexponential retrial time a_3 and service time b_i, $i = 1, 2, 3$ are almost linearly increasing in λ. The exponential retrial case shows slightly more curvature and the Erlang case shows the most curvature.

For the set of parameters $(\lambda = 0.1, \beta_1 = 0.5, \gamma_1 = 0.4, \gamma_2 = 0.2, \mu = 0.2, \theta = 0.5)$. Figure 10.8 (A_2) displays $E(W_q)$ versus α for $a_1(x)$ and $b_1(x)$, $a_2(x)$ and $b_1(x)$ and $a_3(x)$ and $b_1(x)$. Figure 10.8 (B_2) displays $E(W_q)$ versus α for $a_1(x)$ and $b_2(x)$, $a_2(x)$ and $b_2(x)$ and $a_3(x)$ and $b_2(x)$. Figure 10.8 (C_2) displays $E(W_q)$ versus α for $a_1(x)$

and $b_3(x)$, $a_2(x)$ and $b_3(x)$ and $a_3(x)$ and $b_3(x)$. It can be seen that in each case above, $E(W_q)$ shows a downward trend as α increases, as expected. Again the graph of the exponential case lies between the Erlang and the hyperexponential cases.

For the set of parameters ($\lambda = 0.1, \alpha = 0.2, \gamma_1 = 0.4, \gamma_2 = 0.1, \mu = 0.4, \theta = 0.3$). Figure 10.9 ($A_3$) displays $E(W_q)$ versus β_1 for $a_1(x)$ and $b_1(x)$, $a_2(x)$ and $b_1(x)$ and $a_3(x)$ and $b_1(x)$. Figure 10.9 (B_3) displays $E(W_q)$ versus α for $a_1(x)$ and $b_2(x)$, $a_2(x)$ and $b_2(x)$ and $a_3(x)$ and $b_2(x)$. Figure 10.9 (C_3) displays $E(W_q)$ versus α for $a_1(x)$ and $b_3(x)$, $a_2(x)$ and $b_3(x)$ and $a_3(x)$ and $b_3(x)$. It can be seen that in each case above, $E(W_q)$ is increasing in β_1. Note the units on the horizontal and vertical axis. We see that a small increase in the service time can cause a large change in the total time spent in the retrial queue. From a planning point of view, we should pay close attention to the service time if we want to improve our performance.

For the set of parameters ($\lambda = 0.1, \alpha = 0.3, \beta_1 = 1.5, \gamma_1 = 0.4, \gamma_2 = 2, \theta = 0.4$). Figure 10.10 ($A_4$) displays $E(W_q)$ versus μ for $a_1(x)$ and $b_1(x)$, $a_2(x)$ and $b_1(x)$ and $a_3(x)$ and $b_1(x)$. Figure 10.10 (B_4) displays $E(W_q)$ versus α for $a_1(x)$ and $b_2(x)$, $a_2(x)$ and $b_2(x)$ and $a_3(x)$ and $b_2(x)$. Figure 10.10 (C_4) displays $E(W_q)$ versus α for $a_1(x)$ and $b_3(x)$, $a_2(x)$ and $b_3(x)$ and $a_3(x)$ and $b_3(x)$. It can be seen that in each case above, $E(W_q)$ shows an upward trend as μ increases. Again look at the vertical and horizontal units. It can be seen that a small increase in the breakdown rate will cause a large increase in the total time spent in the retrial queue. We can use this information to help with planning. Of course, this result will depend strongly on the parameters chosen for γ_1 and θ.

For the set of parameters ($\lambda = 0.2, \alpha = 0.3, \beta_1 = 0.8, \mu = 0.2, \gamma_2 = 1.2, \theta = 0.1$).

Figure 10.11 (A_5) displays $E(W_q)$ versus γ_1 for $a_1(x)$ and $b_1(x)$, $a_2(x)$ and $b_1(x)$ and $a_3(x)$ and $b_1(x)$. Figure 10.11 (B_5) displays $E(W_q)$ versus α for $a_1(x)$ and $b_2(x)$, $a_2(x)$ and $b_2(x)$ and $a_3(x)$ and $b_2(x)$. Figure 10.11 (C_5) displays $E(W_q)$ versus α for $a_1(x)$ and $b_3(x)$, $a_2(x)$ and $b_3(x)$ and $a_3(x)$ and $b_3(x)$. It can be seen that in each case above, $E(W_q)$ shows an upward trend as γ_1 increases. Note that we see here a very large difference in expected time spent in the retrial queue. The Erlang case is much larger. Note that the retrial time (i.e. the time for the customer at the head of the retrial queue to reach the server in competition with a primary customer) has three different distributions in each of A5, B5, C5. The expression for $E(W_q)$ depends on the Laplace transform of the retrial time and that accounts for the difference.

For the set of parameters ($\lambda = 0.1, \alpha = 0.3, \beta_1 = 1, \mu = 0.2, \gamma_1 = 1, \gamma_2 = 2$). Figure 10.12 ($A_6$) displays $E(W_q)$ versus θ for $a_1(x)$ and $b_1(x)$, $a_2(x)$ and $b_1(x)$ and $a_3(x)$ and $b_1(x)$. Figure 10.12 (B_6) displays $E(W_q)$ versus α for $a_1(x)$ and $b_2(x)$, $a_2(x)$ and $b_2(x)$ and $a_3(x)$ and $b_2(x)$. Figure 10.12 (C_6) displays $E(W_q)$ versus α for $a_1(x)$ and $b_3(x)$, $a_2(x)$ and $b_3(x)$ and $a_3(x)$ and $b_3(x)$. It can be seen that in each case above, $E(W_q)$ shows a downward trend as θ increases. In this figure, for each section , the three graphs are almost identical. This shows that the influence of θ is very slight for the particular parameters chosen.

Figure 10.7: (A_1) Exponential service distribution. (B_1) Erlang service distribution. (C_1) Hyper-exponential service distribution.

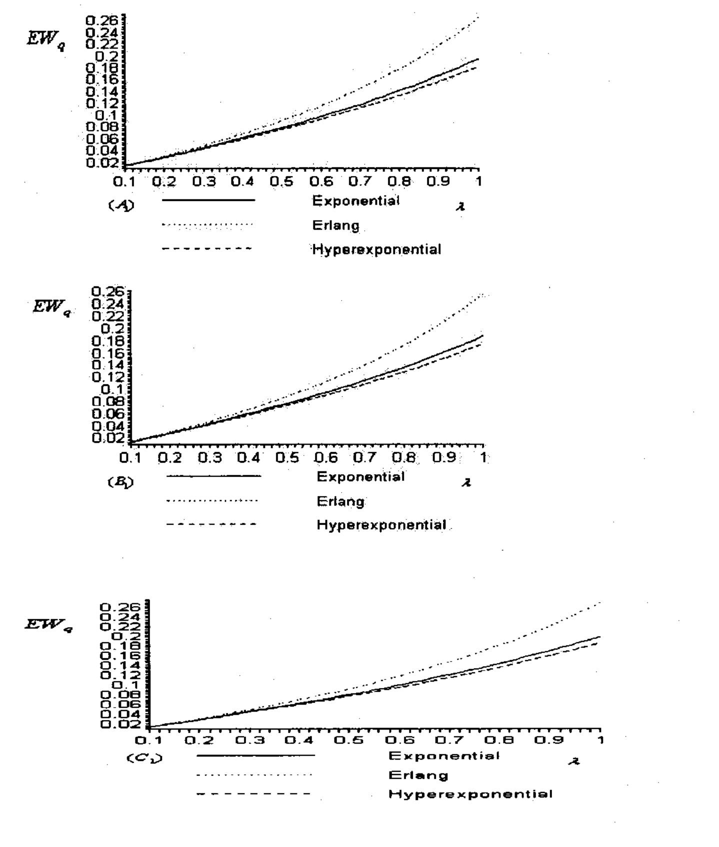


Figure 10.8: (A_2) Exponential service distribution. (B_2) Erlang service distribution. (C_2) Hyper-exponential service distribution.

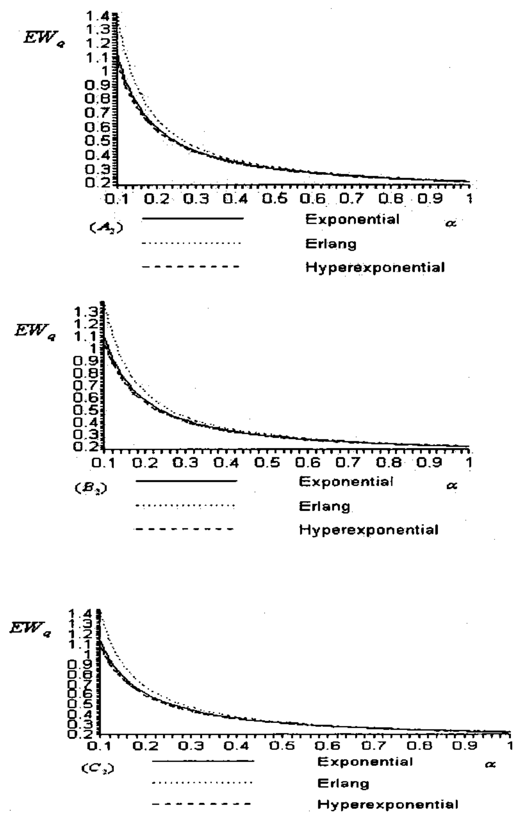


Figure 10.9: (A_3) Exponential service distribution. (B_3) Erlang service distribution. (C_3) Hyper-exponential service distribution.

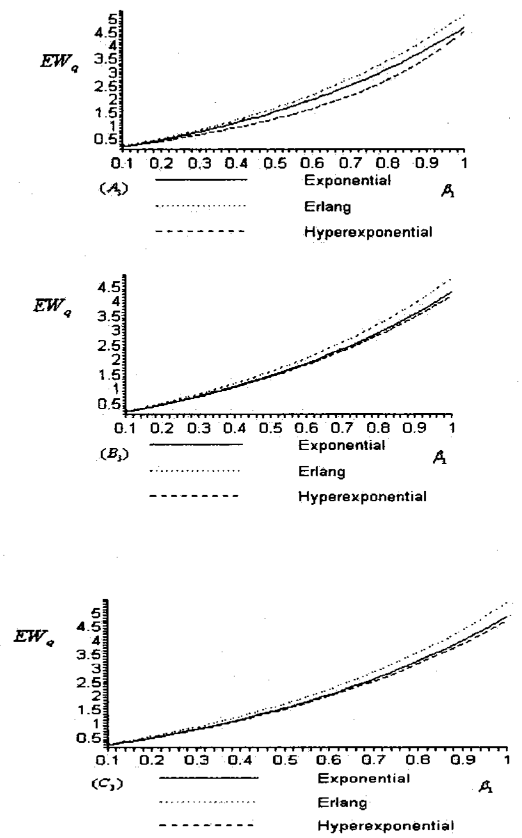


Figure 10.10: (A_4) Exponential service distribution. (B_4) Erlang service distribution. (C_4) Hyper-exponential service distribution.

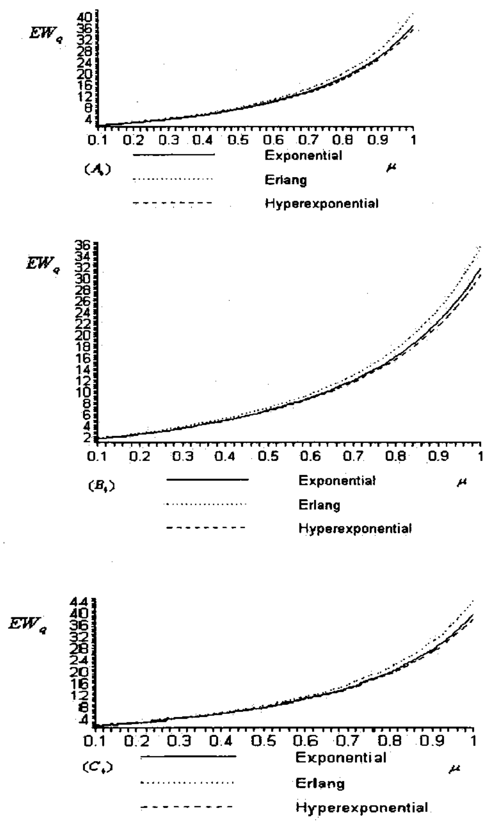


Figure 10.11: (A_5) Exponential service distribution. (B_5) Erlang service distribution. (C_5) Hyper-exponential service distribution.

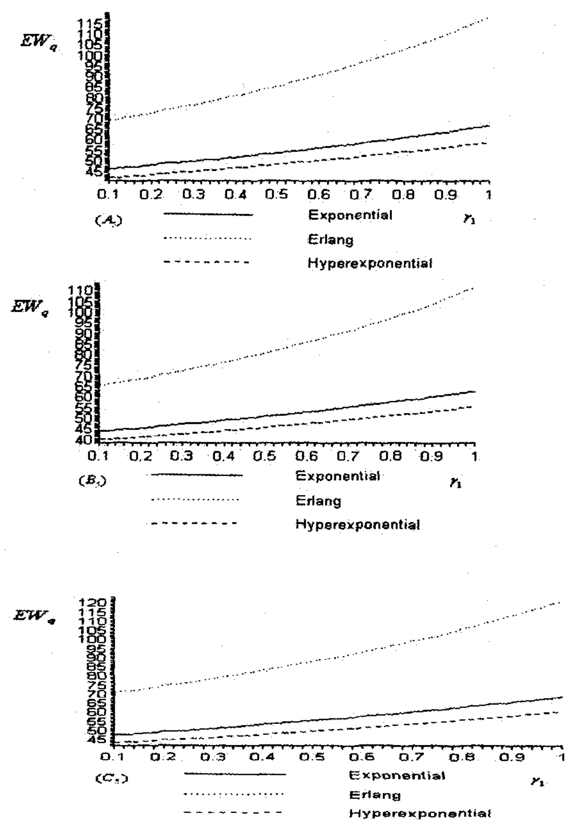
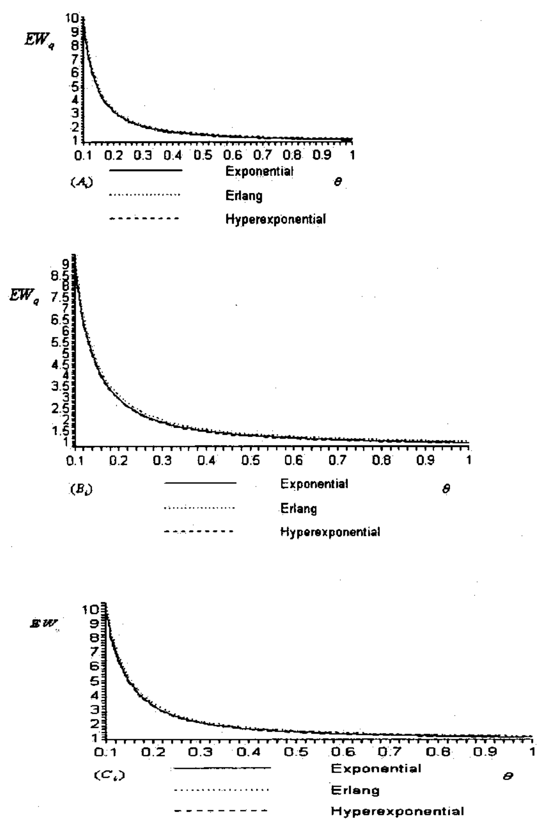


Figure 10.12: (A_6) Exponential service distribution. (B_6) Erlang service distribution. (C_6) Hyper-exponential service distribution.



10.3 $\mathbf{P_{(0,0)}, P_0, P_1, P_{20}, P_{21}}$ and $\mathbf{P_3}$ versus $\mathbf{p, q}$ and $\mathbf{r}$

This section gives graphical results for Model II of Chapter 4. Recall that p is the probability a primary customer who cannot enter service upon arrival will choose to join the retrial queue. that Recall that $1-q$ is the probability that a retrial customer would be lost from the system at each retrial attempt from the head of the retrial queue, if unsuccessful. Recall that r is the probability that the customer in service remains in the service position when the server breaks down. For all of the figures in this section, the probabilities on the vertical axes depend only on the mean service time rather than on the complete service time distribution. The$beta_1$ value is always given.

For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, q = 0.5, r = 0.5$), Figure 10.13 ($A_7$) displays $P_{(0,0)}$ (probability of an empty system) versus p for $a_1(x)$, $a_2(x)$ and $a_3(x)$, as described in the previous section. For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, p = 0.5, r = 0.5$), Figure 10.13 ($B_7$) displays $P_{(0,0)}$ versus q for $a_1(x)$, $a_2(x)$ and $a_3(x)$. For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, p = 0.5, q = 0.5$), Figure 10.13 ($C_7$) displays $P_{(0,0)}$ versus r for $a_1(x)$, $a_2(x)$ and $a_3(x)$. Note that $P_{(0,0)}$ shows an downward trend as p (as expected). Also $P_{(0,0)}$ shows a downward trend as q increases, and $P_{(0,0)}$ shows an upward trend as r increases, since this should decrease the time that a customer takes to leave the system with completed service. Figure 10.13 (A_7) and 10.13 (C_7) give very similar results for all three retrial time distributions. For Figure 9.13 (B_7), the Erlang results in a lower value for $P_{(0,0)}$.

For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, q =$

$0.5, r = 0.5$), Figure 10.14 (A_8) displays P_0 (probability of idle server, nonempty retrial queue) versus p for $a_1(x)$, $a_2(x)$ and $a_3(x)$. We would expect that as p increases, so should P_0 and indeed this is the case. For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, p = 0.5, r = 0.5$), Figure 10.14 ($B_8$) displays P_0 versus q for $a_1(x)$, $a_2(x)$ and $a_3(x)$. Again we expect an increasing function. For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, p = 0.5, q = 0.5$), Figure 10.14 ($C_8$) displays P_0 versus r for $a_1(x)$, $a_2(x)$ and $a_3(x)$. As r increases, the server finishes customers in service faster. This allows customers in the retrial queue to exit so P_0 shows a downward trend as r increases. Note also that the curves are almost linear so that linear approximations could be quite useful.

For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, q = 0.5, r = 0.5$), Figure 10.15 ($A_9$) displays P_1 (probability that the server is busy, not including repair time) versus p for $a_1(x)$, $a_2(x)$ and $a_3(x)$. The more likely it is that customers stay in the system, the server will be busy for a higher proportion of time. So we expect and see an increasing function. For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, p = 0.5, r = 0.5$), Figure 10.15 ($B_9$) displays P_1 versus q for $a_1(x)$, $a_2(x)$ and $a_3(x)$. We expect an increasing function and that is what occurs. For Figures 10.13 and 10.14, the graph of the probability versus r displayed a different trend than the graphs versus p and q. However, in this case, we would expect that as r increases, the server will become free faster and thus there will be fewer customers lost to the system. Hence the server will be busy more often. For the set of parameters ($\lambda = 3, \alpha = 5, \beta_1 = 0.06, \gamma_1 = 0.4, \mu = 2, \theta = 0.5, p = 0.5, q = 0.5$), Figure 10.15 ($C_9$) displays P_1 versus r for $a_1(x)$, $a_2(x)$ and $a_3(x)$. P_1 shows an upward trend as p increases or q increases or r increases.

For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 = 6, \mu = 1, \theta = 0.08, q = 0.5, r = 0.5$), Figure 10.16 ($A_{10}$) displays P_{20} (the steady state probability that the server is under repair and the customer in service after server breakdown remains in the service position) versus p for $a_1(x)$, $a_2(x)$ and $a_3(x)$. For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 = 6, \mu = 1, \theta = 0.08, p = 0.5, r = 0.5$), Figure 10.16 ($B_{10}$) displays P_{20} versus q for $a_1(x)$, $a_2(x)$ and $a_3(x)$. For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 = 6, \mu = 1, \theta = 0.08, p = 0.5, q = 0.5$), Figure 10.16 ($C_{10}$) displays P_{20} versus r for $a_1(x)$, $a_2(x)$ and $a_3(x)$. P_{20} shows an upward trend as p increases or q increases or r increases. As p increases, fewer customers are lost to the system so we would expect the system to have customers in the service area more often, and hence the situation described by P_{20} should occur more often. Thus we expect the function to be increasing and it is. For the same reason, we expect P_{20} to be increasing in q. It is even more clear that P_{20} should be increasing in r since r is the probability that a customer in service stays in service area during breakdown. Note that vertical axis of the third graph covers a much larger region reflecting the strong dependence of P_{20} on p.

For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 = 6, \mu = 1, \theta = 0.05, q = 0.5, r = 0.5$), Figure 10.17 ($A_{11}$) displays P_{21} (limiting probability that the customer in service is in the service retrial orbit and the server is under repair) versus p for $a_1(x)$, $a_2(x)$ and $a_3(x)$. For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 = 6, \mu = 1, \theta = 0.05, p = 0.5, r = 0.5$), Figure 10.17 ($B_{11}$) displays P_{21} versus q for $a_1(x)$, $a_2(x)$ and $a_3(x)$. For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 =$

$6, \mu = 1, \theta = 0.05, p = 0.5, q = 0.5$), Figure 10.17 ($C_{11}$) displays P_{21} versus r for $a_1(x)$, $a_2(x)$ and $a_3(x)$. We note that P_{21} and P_{20} are partial opposites. So we would expect that all three plots will move in the opposite direction from those in Figure 10.16. However, we would be wrong. Only the third graph is decreasing in Figure 10.17. The reason that the other two graphs are increasing is that P_{21} is also affected by the probability that the server is busy, which is increasing in p and q, (see Figure 10.15) and the amount of increases affects P_{21}.

In Figure 10.18, we deliberately choose parameters for which the stability condition does not hold. This results in some unusual graphs, with probabilities taking on values greater than 1. This emphasizes the fact that we should always check the stability condition for the parameters selected. Otherwise, our results will be meaningless. For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 = 6, \mu = 1, \theta = 0.01, q = 0.5, r = 0.5$), Figure 10.18 ($A_{12}$) displays P_3 (limiting probability that the server is up but and a customer is in the retrial service orbit) versus p for $a_1(x)$, $a_2(x)$ and $a_3(x)$. For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 = 6, \mu = 1, \theta = 0.01, p = 0.5, r = 0.5$), Figure 10.18 ($B_{12}$) displays P_3 versus q for $a_1(x)$, $a_2(x)$ and $a_3(x)$. For the set of parameters ($\lambda = 5, \alpha = 0.8, \beta_1 = 0.04, \gamma_1 = 6, \mu = 1, \theta = 0.01, p = 0.5, q = 0.5$), Figure 10.18 ($C_{12}$) displays P_3 versus r for $a_1(x)$, $a_2(x)$ and $a_3(x)$.

Figure 10.13: (A_7) $P_{(0,0)}$ versus p. (B_7) $P_{(0,0)}$ versus q. (C_7) $P_{(0,0)}$ versus r.

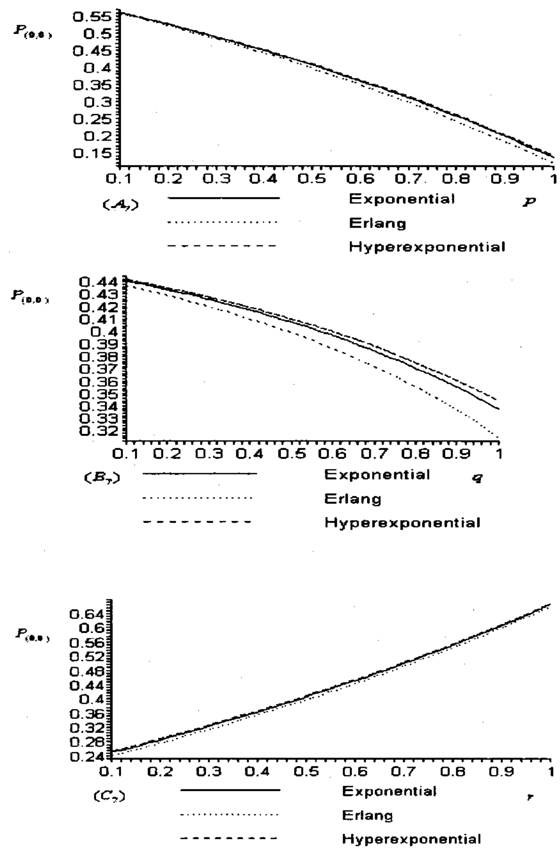


Figure 10.14: (A_8) P_0 versus p. (B_8) P_0 versus q. (C_8) P_0 versus r.

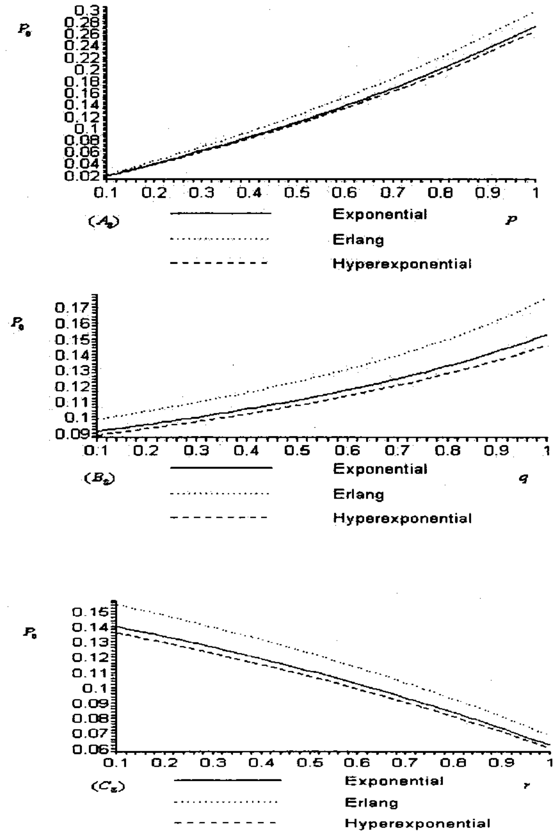


Figure 10.15: (A_9) P_1 versus p. (B_9) P_1 versus q. (C_9) P_1 versus r.

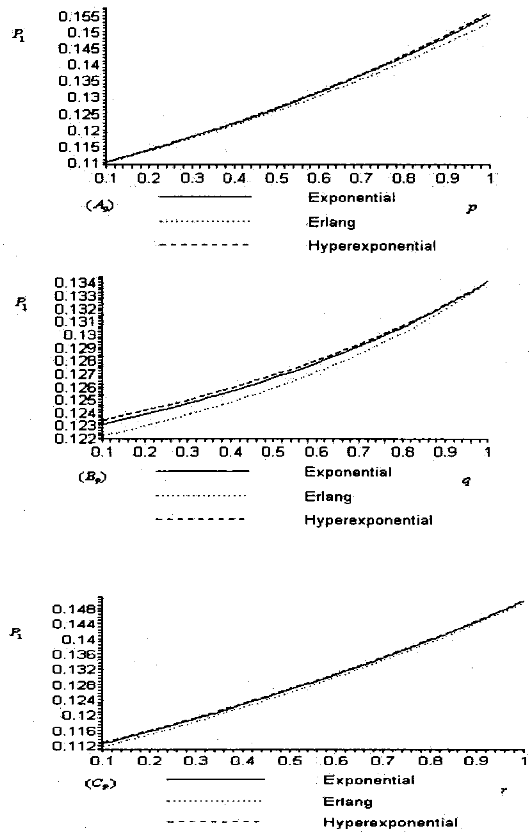


Figure 10.16: (A_{10}) P_{20} versus p. (B_{10}) P_{20} versus q. (C_{10}) P_{20} versus r.

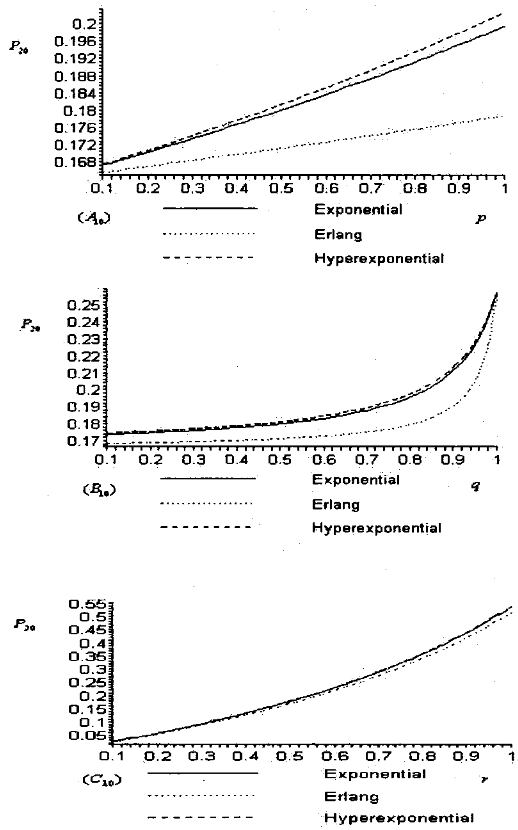


Figure 10.17: (A_{11}) P_{21} versus p. (B_{11}) P_{21} versus q. (C_{11}) P_{21} versus r.

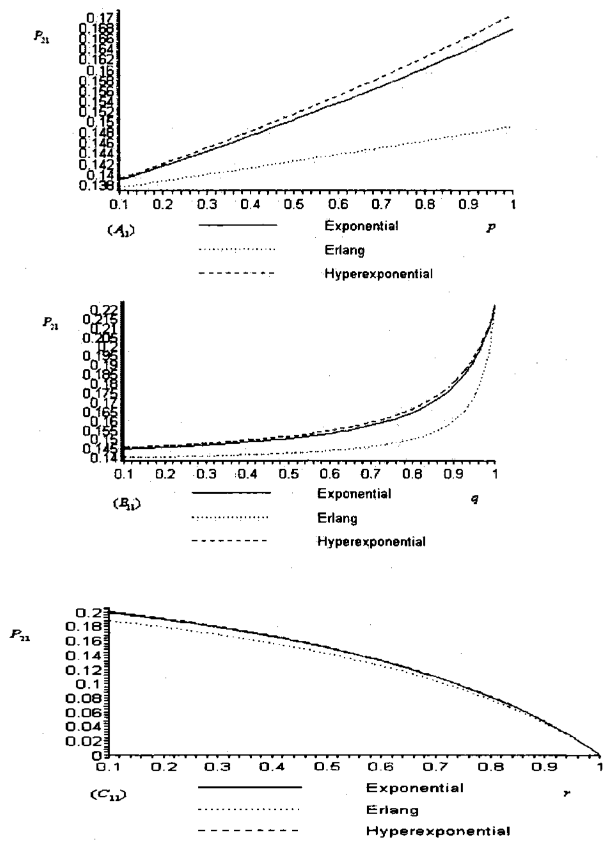
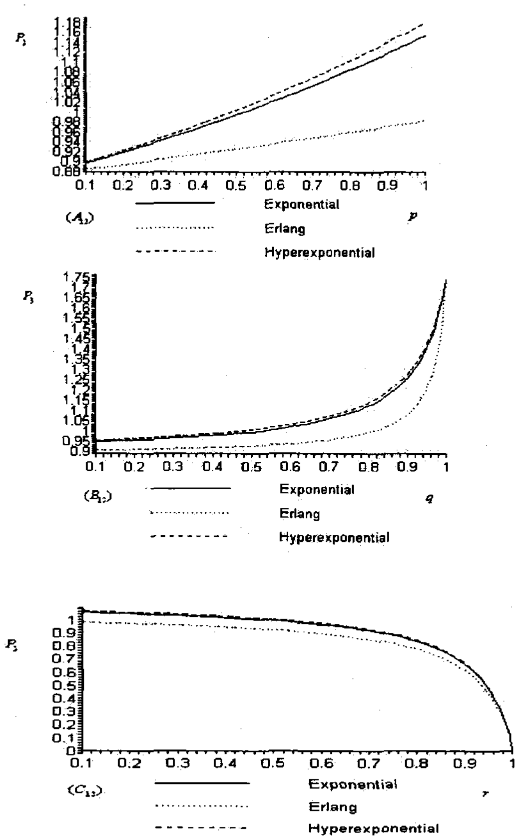


Figure 10.18: (A_{12}) P_3 versus p. (B_{12}) P_3 versus q. (C_{12}) P_3 versus r.



Chapter 11

Simulation

In this chapter, a simulation for model III of Chapter 6 is considered. We make the following assumptions for our simulation. Primary customers from primary source I arrive according to a Poisson process with rate λ_1 and primary customers from the primary source II arrive according to a Poisson process with rate λ_2. The retrial time of the customer at the head of the retrial queue, measured from the last departure time, is exponentially distributed with rate α. A retrial time of any retrial customer from the retrial (service) orbit is exponentially distributed with rate θ. The service time for a customer from primary source I or from the retrial queue has an Erlang distribution with parameters n_1 and μ_1 and the service time for a customer from primary source II or from the retrial orbits has an Erlang distribution with parameters n_2 and μ_2.

Here we are interested in simulating model III to get the following quantities:

(i). The average time W_1 that a customer, out of the first 10000 arriving customers, from the primary source (I) spends in the system.

(ii). The average time W_2 that a customer, out of the first 10000 arriving customers, from the primary source (II) spends in the system.

(iii). The number of repeated customers left in the retrial queue after the first 10000 arriving customers obtain service.

(iv). The number of repeated customers left in the retrial orbits after the first 10000 arriving customers obtain service.

In order to do simulation, we set $n_1 = n_2 = 2$, $\mu_1 = 30$ and $\mu_1 = 20$, as an example. The same method will work for any reasonable set of parameters. The results of our simulation are displayed in the following table.

We analyze the effects of the parameters λ_1, λ_2, α and θ on the waiting time W_1 and W_2 from the numerical results of the tables.

(i) Consider one factor case. We fix three parameters and allow one to vary.

(a) For any given λ_1, λ_2 and θ, if α increases (or decreases), then W_1 decreases (or increases) and W_2 increases (or decreases). Of course, our intuition expects this.

(b) For any given λ_1, λ_2 and α, if θ increases (or decreases), then W_1 increases (or decreases) and W_2 decreases (or increases).

(c) For any given λ_2, α and θ, if λ_1 increases (or decreases), then W_1 increases (or decreases) and W_2 increases (or decreases).

(d) For any given λ_1, α and θ, if λ_2 increases (or decreases), then W_1 increases (or decreases) and W_2 increases (or decreases).

(ii) Consider the two factors case. We fix two parameters and allow two to vary.

(a) For any given λ_1 and λ_2, if α increases (or decreases) and θ increases (or decreases), then W_1 decreases (or increases) and W_2 decreases (or increases).

(b) For any given λ_1 and α, if λ_2 increases (or decreases) and θ decreases (or increases), then W_1 increases (or decreases) and W_2 increases (or decreases).

(c) For any given λ_1 and θ, if λ_2 increases (or decreases) and α decreases (or increases), then W_1 decreases (or increases) and W_2 decreases (or increases).

(d) For any given λ_2 and α, if λ_1 increases (or decreases) and θ decreases (or increases), then W_1 decreases (or increases) and W_2 decreases (or increases).

(e) For any given λ_2 and θ, if λ_1 increases (or decreases) and α decreases (or increases), then W_1 decreases (or increases) and W_2 decreases (or increases).

(f) For any given α and θ , if λ_1 increases (or decreases) and $\lambda_2 increases (ordecreases), then$ W_1 decreases (or increases) and W_2 decreases (or increases).

(iii). Consider three factors case. We fix one parameter and allow three to vary.

(a) For any given λ_1, if λ_2 increases (or decreases) and α decreases (or increases) and θ decreases (or increases), then W_1 increases (or decreases) and W_2 increases (or decreases).

(b) For any given λ_2, if λ_1 increases (or decreases) and α decreases (or increases) and θ decreases (or increases), then W_1 increases (or decreases) and W_2 increases (or decreases).

(c) For any given α, if λ_1 increases (or decreases) and λ_2 increases (or decreases) and θ decreases (or increases), then W_1 increases (or decreases) and W_2 increases (or decreases).

(d) For any given θ, ifλ_1 increases (or decreases) and λ_2 increases (or decreases) and α decreases (or increases), then W_1 increases (or decreases) and W_2 increases (or decreases).

(iv) Consider the four factors case.

(a) If λ_1 increases (or decreases) and λ_2 increases (or decreases) and α decreases (or increases) and θ decreases (or increases), then W_1 increases (or decreases) and W_2 increases (or decreases).

Evidently, extremely large values of N_1 and N_2 may reflect the non-stability of the system. From the tables, it can be seen that λ_2 is an important factor that influence the size of the system and the stability of the system.

Tables 10.1, 10.2, 10.4 indicate only a small number of customers in the retrial queue and in the retrial orbits after the 10000 customers have completed service. These small numbers indicate that the system is stable. However, in Table 10.3, the number of customers remaining in the retrial queue and the retrial orbits are both large. This indicates that the system is not stable. (Recall that we only have an analytic expression for the sufficiency condition for stability so that numerical results can be useful to evaluate stability.) In Table 10.5, the number of customers in the retrial queue is large, but the number of customers in the retrial orbits is small. Again this is an indicator of the nonstability of the system, although as α and θ both increase, the system seems to be approaching stability, with the value of N_1 decreasing.

Note also in Tables 10.3 and 10.5 that the number of customers in the retrial queue is much larger than the number of customers in the retrial orbits. This results because of the competition for service. As the number of customers in the retrial orbits increases, the probability that the next customer will come from the retrial orbits (rather than from the retrial queue) increases.

Tables 10.6, 10.7, 10.8, 10.9 all show a large number of customers in the retrial queue. Tables 10.7 and 10.8 show only a small number customers in the retrial orbits. All four of these tables indicate a lack of stability of the system. The results in Table 10.7 and 10.8 show that the customers in the retrial queue rarely win in competition against the retrial orbit customers. Primary customers of both types will enter the retrial queue and retrial orbits with high probability. Almost all of the customers entering service are coming from the retrial orbits.

All of the tables indicate that simulation is a valuable tool in identifying the stability of a system. For those systems which are stable, simulation provides a useful method of obtaining measures of the expected system time for both types of customer and of find the expected number of customers for both types in the system.

Table 11.1: Waiting Time and Number of Remaining Customers for $\lambda_1 = 2$ and $\lambda_2 = 2$.

α	θ	W_1	W_2	N_1	N_2
25	5	0.12584	0.253396	0	0
25	15	0.1309	0.182806	0	0
25	25	0.136851	0.169701	0	1
75	5	0.10993	0.255817	1	3
75	15	0.115198	0.186225	0	0
75	25	0.109987	0.166118	0	1
125	5	0.106525	0.262208	0	1
125	15	0.106692	0.184814	0	0
125	25	0.110408	0.168261	0	0

Table 11.2: Waiting Time and Number of Remaining Customers for $\lambda_1 = 2$ and $\lambda_2 = 6$.

α	θ	W_1	W_2	N_1	N_2
25	5	0.380624	0.866875	0	1
25	15	0.519893	0.485171	0	3
25	25	0.616633	0.390787	3	3
75	5	0.22409	0.983544	0	10
75	15	0.243619	0.545938	1	6
75	25	0.272702	0.405862	0	1
125	5	0.172242	0.90444	0	0
125	15	0.204812	0.548501	1	2
125	25	0.223155	0.453981	1	2

Table 11.3: Waiting Time and Number of Remaining Customers for $\lambda_1 = 2$ and $\lambda_2 = 10$.

α	θ	W_1	W_2	N_1	N_2
25	5	3531.39	18.5566	12020	136
25	15	3701.2	7.95697	12032	11
25	25	2493.4	4.13147	9667	44
75	5	1451.49	29.1288	6836	563
75	15	2943.47	19.574	10738	227
75	25	2215.47	9.80123	7693	63
125	5	641.981	33.7943	3247	492
125	15	1603.56	20.0344	6021	129
125	25	1535.71	14.2988	5866	108

Table 11.4: Waiting Time and Number of Remaining Customers for $\lambda_1 = 6$ and $\lambda_2 = 2$.

α	θ	W_1	W_2	N_1	N_2
25	5	0.307344	0.513719	1	1
25	15	0.397468	0.302041	0	0
25	25	0.435626	0.244354	1	0
75	5	0.198339	0.609883	0	0
75	15	0.209593	0.351517	0	0
75	25	0.22101	0.289333	2	1
125	5	0.159505	0.587283	1	0
125	15	0.177931	0.36003	0	0
125	25	0.179514	0.278399	1	0

Table 11.5: Waiting Time and Number of Remaining Customers for $\lambda_1 = 6$ and $\lambda_2 = 6$.

α	θ	W_1	W_2	N_1	N_2
25	5	211.678	1.6488	2179	6
25	15	202.194	0.705627	2125	2
25	25	214.649	0.576809	2373	0
75	5	58.9881	3.20914	744	13
75	15	51.9127	1.28611	555	2
75	25	55.3215	0.826194	543	1
125	5	18.3619	5.00401	297	21
125	15	51.2605	1.81334	553	7
125	25	32.7024	1.22532	335	5

Table 11.6: Waiting Time and Number of Remaining Customers for $\lambda_1 = 6$ and $\lambda_2 = 10$.

α	θ	W_1	W_2	N_1	N_2
25	5	9058.77	15.0327	102227	616
25	15	9784.75	9.5851	114692	100
25	25	8218.19	4.07659	93347	52
75	5	6011.67	21.3132	81390	744
75	15	6719.15	11.9893	74706	148
75	25	6426.72	7.60958	76261	330
125	5	3286.09	31.3541	44138	670
125	15	5186.84	13.2773	64137	267
125	25	5715.6	17.5037	62122	211

Table 11.7: Waiting Time and Number of Remaining Customers for $\lambda_1 = 10$ and $\lambda_2 = 2$.

α	θ	W_1	W_2	N_1	N_2
25	5	99.256	0.914037	1687	1
25	15	76.2671	0.406674	1444	0
25	25	77.9401	0.298497	1397	0
75	5	1.60998	1.42349	22	6
75	15	1.74936	0.653023	35	3
75	25	4.08878	0.461258	15	0
125	5	1.13622	2.18495	12	6
125	15	0.810793	0.83446	19	1
125	25	0.749861	0.58142	2	1

Table 11.8: Waiting Time and Number of Remaining Customers for $\lambda_1 = 10$ and $\lambda_2 = 6$.

α	θ	W_1	W_2	N_1	N_2
25	5	548.8	1.78203	9559	12
25	15	557.922	0.7586	10138	1
25	25	544.494	0.548596	9631	3
75	5	301.269	3.62512	5427	15
75	15	307.275	1.29491	5521	5
75	25	349.701	1.00964	5898	6
125	5	271.048	5.38533	4764	13
125	15	278.995	1.93139	4905	12
125	25	272.347	1.26056	4899	5

Table 11.9: Waiting Time and Number of Remaining Customers for $\lambda_1 = 10$ and $\lambda_2 = 10$.

α	θ	W_1	W_2	N_1	N_2
25	5	14782.4	19.5192	259007	316
25	15	13871.7	10.7021	239900	50
25	25	14713.2	4.87113	263737	25
75	5	7234.2	27.3107	155560	587
75	15	9643.13	9.80852	176890	200
75	25	10702	8.62262	184179	136
125	5	5755.44	33.8066	120691	726
125	15	6813.14	16.8038	135891	348
125	25	7379.47	10.3777	148788	244

Chapter 12

Conclusion and Future Research

In this thesis, we studied four models (models I,II,III and discrete service model) related to retrial queues. All of the models have the common characteristic that some of the customers will enter a retrial queue of the type where the customers attempt to re-enter according to a FCFS discipline. The first two models included a customers in service who also had a service retrial orbit. The latter characteristic was dropped for the third (combined retrial queue and retrial orbits model) and fourth models (discrete service time model) because these models have not been analyzed anywhere else before and the added complexity of allowing the customer in service to go in to orbit would detract from the primary focus.

The models discussed in this thesis include some important aspects of queueing systems, for example, breakdowns and repairs and reservations of customers in service and balking and non-persistence of customers. Now we give some further directions for future research.

(i) Models can be extended to allow retrial times to be measured from the last retrial, rather than measured from the time at which the server becomes idle. Although there may be a signal that the server has become free and this signal could prompt a retrial, other models may not have such a signal. As before we would assume that retrial times are independent and distributed with a general distribution.

(ii) Model I can be extended to a model in which a customer is allowed to enter when the server is up but the unfinished customer is in the service area (but not in service). When the original customer returns, it has preemptive priority over the new customer. Such a model would improve the efficiency of the system.

(iii) The single-server retrial queues can be extended to the multi-server retrial queues, for our models I, II and III.

(iv) The single-server retrial queues can allow a larger buffer in Models I, II and III. We might begin by just considering the capacity of the waiting room to be 1.

(v) We may consider a series of retrial queues without waiting rooms, in which an arriving customer must get a series of service before the customer leaves the system. Here, an arriving customer enters the first retrial queue if the first server is found to be busy. The customer requests service from the second server after the customer completes the first service and enters the second retrial queue if the second server is found to be busy and so on.

Bibliography

[1] Alfa, A. S. and Isotupa, K. P. S. (2004). An M/PH/k retrial queue with finite number of sources, *Computers & Operations Research* 31, 1455-1464.

[2] Aissani, A. (1993). Unreliable queueing with repeated orders. *Microelectronics and Reliability*, 33, 2093-2106.

[3] Aissani, A. (1995). Retrial queue with redundancy and unreliable server. *Queueing Systems*, 17, 431-449.

[4] Aleksandrov, A. M. (1974). A queueing system with repeated orders. *Engineering Cybernetics Review* 12, 1-4.

[5] Altman, E. and Borovkov, A. A. (1997). On the stability of retrial queues. *Queueing Systems*, 26, 343-363.

[6] Anisimov, V. V. and Artalejo, J. R. (2001). Analysis of Markov multiserver retrial queues with negative arrivals. *Queueing Systems*, 39, 157-182.

[7] J.R. Artalejo and M.J. Lopez-Herrero, *On the Single Server Retrial Queue with Balking,* INFOR 38 (2000) 33-50.

[8] Artalejo, J. R. (1994). New results in retrial queueing systems with breakdown of the servers. *Statistica Neerlandica*, 48, 23-36.

[9] Artalejo, J. R. (1999). Accessible bibliography on retrial queues. *Mathematical and Computer Modelling*, 30, 1-6.

[10] Artalejo, J. R. and Falin, G. I. (1994). Stochastic decomposition for retrial queues. *Top*, 2, 329-342.

[11] Artalejo J. R. and Falin, G. I. (1996). On the orbit characteristics of the M/G/1 retrial queue. *Naval Research Logistics*, 43, 1147-1161.

[12] Artalejo, J. R. and Falin, G. I. (2002). Standard and retrial queueing systems: a comparative analysis. *Revista Matematica Complutense*, 15, 101-129.

[13] Artalejo J. R. and Gómez-Corral, A. (1997). Steady state solution of a single-server queue with linear requests repeated. *Journal of Applied Probability*, 34, 223-233.

[14] Artalejo, J. R., Rajagopalan, V. and Sivasamy, R. (2000). On finite Markovian queues with repeated attempts. *Investigacion Operativa*, 9, 83-94.

[15] Asmussen, S. (2003). *Applied Probability and Queues,* Springer-Verlag, New York.

[16] Atencia, I. and Moreno, P. (2004). A Discrete Time Geo/G/1 Retrial Queue with General Retrial Times, *Queueing Systems*, 48, 5-21.

[17] Avi-Itzhak, B. and Naor, P. (1962). Some queueing problems with the service station subject to breakdown. *Operations Research*, 10, 303-320.

[18] Bertsekas, D. and Gallager, R. (1992). *Data Networks (Second Edition)*, Prentice Hall, Englewood Cliffs, New Jersey.

[19] Borovkov, A. A. (1976). *Stochastic Processes in Queueing Theory.* Springer-Verlag, New York.

[20] Borovkov, A. A. (1989). Ergodicity and stability of multidimensional Markov chains. *Theory of Probability and its Applications*, 35, 542-546.

[21] Brahimi, M. and Worthington, D. J. (1991). The finite capacity multi-server queue with inhomogeneous arrival rate and discrete service time distribution and its application to continuous service time problems, *European Journal of Operational Research*, 50 310-324.

[22] Brill, P. H. (2002). Properties of the waiting time in $M/D_N/1$ queues, Technical report, *Department of Mathematics and Statistics, University of Windsor*, Report WMSR, 02-01.

[23] Brun, O. and Garcia, J. (2000). Analytical solutions of finite capacity $M/D/1$ queues, *Journal of Applied Probability*, 37 1092-1098.

[24] Chakravarthy, S. R. and Dudin, A. (2002). A multiserver retrial queue with BMAP arrivals and group services. *Queueing Systems.* 42, 5-31.

[25] Chakravarthy, S. R. and Dudin, A. (2003). Analysis of a retrial queueing model with MAP arrivals and two types of customers. *Mathematical and Computer Modelling Systems.* 37, 343-363.

[26] Cao, J. and Cheng, K. (1982). Analysis of an M/G/1 queueing system with repairable service station,. *Acta Mathematicae Applicatae Sinica*, 5, 113-127.

[27] Choi, B. D. and Chang, Y. (1999). Single server retrial queues with priority calls. *Mathematical and Computer Modelling*, 30, 7-32.

[28] Choi, B. D. and Kulkarni, V. G. (1992). Feedback retrial queueing systems, in Queueing and Related Models, (editors Bhat and Basawa) *Oxford Science Publications, Oxford.* 93-105.

[29] Choi, B. D. and Park, K. K. (1990). The $M/G/1$ retrial queue with Bernoulli schedule. *Queueing Systems*, 7, 219-227.

[30] Choi, B. D. and Rhee, K. H. (1996). An $M/G/1$ retrial queue with a threshold in the retrial group. *Kyungpook Mathematical Journal*, 35, 469-479.

[31] Choi, B. D., Choi, K. B. and Lee, Y. W. (1995). $M/G/1$ retrial queueing systems with two types of calls and finite capacity. *Queueing Systems*, 19, 215-229.

[32] Choi, B.D., Chung, Y.H. and Dudin, A.N. (2001). The $BMAP/SM/1$ retrial queue with controllable operation modes, *European Journal of Operational Research* 131, 16-30.

[33] Choi, B. D., Han, D. H. and Falin, G. I. (1993). On the virtual waiting time for an $M/G/1$ retrial queue with two types of calls. *Journal of Applied Mathematics and Statistic Analysis*, 6, 11-24.

[34] Choi, B. D., Kim, B. and Choi, S. H. (2003). An $M/G/1$ queue with multiple types of feedback, gated vacations and FCFS policy. *Computers & Operations Research*, 30, 1289-1309.

[35] Choi, B. D., Kim, Y. C. and Lee, Y. W. (1998). The $M/M/c$ retrial queue with geometric loss and feedback. *Computers & Mathematics with Applications*, 36(6), 41-52.

[36] Choi, B. D., Park, K. K. and Pearce, C. E. M. (1993). An $M/M/1$ retrial queue with control policy and general retrial times. *Queueing Systems*, 14, 275-292.

[37] Choi, B. D., Rhee, K. H. and Park, K. K. (1993). The $M/G/1$ retrial queue with retrial rate control policy. *Probability in the Engineering and Information Sciences*, 7, 29-46.

[38] Choo, Q. H. and Conolly, B. (1979). New results in the theory of repeated orders queueing systems. Journal of Applied Probability, 16, 631-640.

[39] Choudhury, G. (2003). Some aspects of an $M/G/1$ queueing system with optional second service. *Top*, 11, 141-150.

[40] Cinlar, E. (1975). *Introduction to Stochastic Processes,* Prentice-Hall Englewood Cliffs, New Jersey.

[41] Cohen, J. W. (1957). Basic problems of telephone traffic theory and the influence of calls. *Philips Telecommunication Review*, 18, 49-100.

[42] Cohen, J. W. (1982). *The Single Server Queue,* North-Holland, Amsterdam.

[43] Cooper, R. B. (1970). Queues served in cyclic order: Waiting times. *The Bell Systems Technical Journal*, 49, 339-413.

[44] Cooper, R. B. (1981). *Introduction to Queueing Theory.* North-Holland, New York.

[45] Cox, D. R. (1955). The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Proceedings of the Cambridge Philosophical Society*, 51, 433-441.

[46] Crommelin, C. D. (1932). Delay probability formulae when the holding times are constant. *Post Office Electrical Engineering Journal*, 25, 41-50.

[47] Deul, N. (1980). Stationary conditions for multi-server queueing systems with repeated calls. *Elektronische Informationsverarbeitung und Kybernetik*, 16, 607-613.

[48] Diamond, J. E. and Alfa, A. S. (1999). Approximation methods for $M/PH/1$ retrial queues with phase type inter-retrial times. *European Journal of Operational Research*, 113, 620-631.

[49] Doshi, B. T. (1991). Analysis of a two phase queueing system with general service times. *Operations Research Letters*, 10, 265-272.

[50] Dudin, A. and Klimenok, V. (2000). A retrial $BMAP/SM/1$ system with linear requests. *Queueing Systems*, 34, 47-66.

[51] Elcan, A. (1999). Asymptotic bounds for an optimal state-dependent retrial rate of the $M/M/1$ queue with returning customers. *Mathematical and Computer Modelling*, 30, 129-140.

[52] Erlang, A. K. (1909). The theory of probabilities and telephone conversations, *Nyt Tidsskrift for Matematik B*, 20, 33-39, English translation in *The Life and Works of A. K. Erlang*, by E. Brockmeyer, H.L Halstrom, and Arne Jensen, The Copenhagen Telephone Company, Copenhagen, 1948.

[53] Falin, G. I. (1984). On sufficient conditions for ergodicity of multichannel queueing systems with repeated calls. *Advances in Applied Probability*, 16, 447-448.

[54] Falin, G. I. (1986a). On the waiting time process in single-line queue with repeated calls. *Journal of Applied Probability*, 23, 185-192.

[55] Falin, G. I. (1986b). Single-line repeated orders queueing systems. *Optimization*, 17, 649-667.

[56] Falin, G. I. (1990). A survey of retrial queues. *Queueing systems*, 7, 127-167.

[57] Falin, G. I. and Fricker, C. (1991). On the virtual waiting time in an $M/G/1$ retrial queue. *Journal of Applied Probability*, 28, 446-460.

[58] Falin, G. I. and Templeton, J. G. C. (1997). *Retrial queues.* Chapman & Hall, New York.

[59] Farahmand, K. (1990). Single line queue with recurrent repeated demands. *Queueing Systems*, 6, 223-228.

[60] Fayolle, G. (1986). *"A simple telephone exchange with delayed feedbacks,"* in *Teletraffic Analysis and Computer Performance Evaluation.* O. J. Boxma, J. W. Cohen, and H. C. Tijms (Editors), Elsevier, Amsterdam.

[61] Fayolle, G. (1995). *Topics in the Constructive Theory of Countable Markov Chains.* Cambridge University Press.

[62] Federgruen, A. and Green, L. (1986). Queueing systems with service interruptions. *Operations Research*, 34, 752-768.

[63] Franx, G. J. (2001). A simple solution for the $M/D/c$ waiting time distribution, *Operations Research Letters*, 29, 221-229.

[64] Fredericks, A. A. and Reisner, G. A. (1979). Approximation to stochastic service systems with an application to a retrial model. *The Bell System Technical Journal*, 58, 557-576.

[65] Gaver, D. P. (1962). A waiting line with nterrupted service, including priorities. *Journal of the Royal Statistical Society, Series B*, 24, 73-90.

[66] Gómez-Corral, A. (1999). Stochastic analysis of a single server retrial queue with general retrial times. *Naval Research Logistics*, 46, 561-581.

[67] Grassmann, W. K. (2000). *Computer Probability.* Kluwer Academic Publishers.

[68] Gross, D. and Harris, C. M. (1998). *Fundamentals of Queueing Theory (third Edition).* John Wiley and Sons.

[69] Greenberg, B. S. (1989). $M/G/1$ queueing systems with returning customers. *Journal of Applied Probability,* 26, 152-163.

[70] Hanschke, T. (1984). A Model for planning switching networks. *Operations Research Proceedings Springer, Berlin,* 555-562.

[71] He, Q. M., Li, H. and Zhao, Y. Q. (2000). Ergodicity of the $BMAP/PH/s/s+K$ retrial queue with PH-retrial times. *Queueing Systems,* 35, 323-347.

[72] Iversen, V. B. and Staalhagen, L. (1999). Waiting time distribution in $M/D/1$ queueing systems, *Electronics Letters,* 35, 2184-2185.

[73] Lopez-Herrero, M. J. (2002). Distribution of the number of customers served in an $M/G/1$ retrial queue. *Journal of Applied Probability,* 39, 407-412.

[74] Martin, M. and Artalejo, J. R. (1995). Analysis of an $M/G/1$ queue with two types of impatient units. *Advances in Applied Probability,* 27, 840-861.

[75] Moreno, P. (2004). An M/G/1 retrial queue with recurrent customers and general retrial times. *Applied Mathematics and Computation,* 159, 651-666.

[76] Kapyrin, V. A. (1977). A study of the stationary characteristics of a queueing system with recurring demands. *Cybernetics,* 13, 584-590.

[77] Karlin, S. and McGregor, J. L. (1957). The differential equations of birth-and-death processes and the Stieltjes moment problem. *Transactions of American Mathematical Society*, 85, 489-546.

[78] Keilson. J. and Servi, L. D. (1993). The matrix $M/M/\infty$ system: Retrial models and Markov modulated sources. *Advances in Applied Probability*, 25, 453-471.

[79] Keilson, J., Cozzolino, J. and Young, H. (1968). A service system with unfilled requests repeated. *Operations Research*, 16, 1126-1137.

[80] Khalil, Z. and Falin, G. I. (1994). Stochastic inequalities for $M/G/1$ retrial queues. *Operations Research Letters*, 16, 285-290.

[81] Kleinrock, L. (1975). *Queueing Systems, Volume 1: Theory.* Wiley-Interscience, New York.

[82] Kobe, E. (2000). Stability condition for $M/D/1$ retrial queueing system with a limited waiting time. *Cybernetics and System Analysis*, 36, 312-314.

[83] Krishna, C. M. and Lee, Y. H. (1990). A study of two phase service. *Operations Research Letters*. 9, 91-97.

[84] Krishna Kumar, B., Vijayakumar, A. and Arivudainambi, D. (2002). An $M/G/1$ retrial queueing system with two-phase service and preemptive resume. *Annals of Operations Research*, 113, 61-79.

[85] Kulkarni, V. G. (1983). On queueing systems with retrials. *Journal of Applied Probability*, 20, 380-389.

[86] Kulkarni, V. G. and Choi, B. D. (1990). Retrial queues with server subject to breakdowns and repairs. *Queueing Systems*, 7, 191-208.

[87] Kulkarni, V. G. and Liang, H. M. (1997). *"Retrial queues revisited."* in *Frontiers in queueing models and applications in science and engineering.* J. H. Dshalalow (editor), CRC Press, Boca Raton.

[88] Kumar, B. K. and Arivudainambi, D. (2002). The $M/G/1$ retrial queue with Bernoulli schedules and general retrial times. *Computers and Mathematics with Applications*, 43, 15-30.

[89] Kumar, B. K. and Madheswari, S. P. (2003). $M^X/G/1$ retrial queue with multiple vacations and starting failures, *Opsearch*, 40, 115-137.

[90] Kumar, B. K., Vijayakumar, A. and Arivudainambi, D. (2002). An $M/G/1$ retrial queueing system with two-phase service and preemptive resume. *Annals of Operation Research*, 113, 61-79.

[91] Levy, Y. and Yechiali, U. (1975). Utilization of idle time in an $M/G/1$ queueing system. *Management Science*, 22, 202-211.

[92] Li, H. and Yang, T. (1995). A single-server retrial queue with server vacations and a finite number of input sources. *European Journal of Operational Research*, 85, 149-160.

[93] Liang, H. M. and Kulkarni, V. G. (1993). Stability condition for a single server retrial queue. *Advances in Applied Probability*, 25, 690-701.

[94] Lopez-Herero, M. J. (2002). Distribution of the number of customers served in an $M/G/1$ retrial queue. *Journal of Applied Probability*, 39, 407-412.

[95] Lucantoni, D. M. (1991). New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7, 1-46.

[96] Madan, K. C. (2000). An $M/G/1$ queue with second optional service. *Queueing Systems*, 34, 37-46.

[97] Marcellán, F. and Pérez, G. (2003). The moments of the $M/M/s$ queue length process. *Queueing Systems*, 44, 281-304.

[98] Martin, M. and Artalejo, J. R. (1995). Analysis of an $M/G/1$ queue with two types of impatient units. *Advances in Applied Probability*, 27, 840-861.

[99] Medhi, J. (2002). A single server Poisson input queue with a second optional channel. *Queueing Systems*, 42, 239-242.

[100] Medhi, J. (2003).*Stochastic Models in Queueing Theory*. Second Edition, Academic Press, Amsterdam.

[101] Mitrany, I. L. and Avi-Itzhak, B. (1968). A many-server queue with service interruptions. *Operations Research*, 16, 628-638.

[102] Neuts, M. F. and Lucantoni, D. M. (1979). A Markovian queue with N servers subject to breakdowns and repairs. *Management Science*, 25, 849-861.

[103] Neuts, M. F. and Ramalhoto, M. F. (1984). A service model in which the server is required to search for customers. *Journal of Applied Probability*, 21, 157-166.

[104] Pakes, A. G. (1969). Some conditions for ergodicity and recurrence of Markov chains. *Operations Research*, 17, 1058-1061.

[105] Ramaswami, V. (1980). The $N/G/1$ queue and its detailed analysis. *Advances in Applied Probability*, 12, 222-261.

[106] Riordan, J. (1961). Delays for last-come first-served service and the busy period. *The Bell System Technical Journal*, 40, 785-793.

[107] Ross, S. M. (1995). *Stochastic processes (second edition).* Wiley, New York.

[108] Sennott, L. I., Humblet, P. A. and Tweedie, R. L. (1983). Mean drifts and the non-ergodicity of Markov chains. *Operations Research*, 31, 783-789.

[109] Shortle, J. F. and Brill, P. H., Analytic distribution of waiting time in an $M/\{iD\}/1$ Queue, *Queueing Systems*, 50 (2005) 185-197.

[110] Spieksma, F. M. and Tweedie, R. L. (1994). Strengthening ergodicity to geometric ergodicity for Markov chains. *Stochastic Models*, 10, 45-74.

[111] Takács, L. (1962). *Introduction to The Theory of Queues.* Oxford University Press, Oxford.

[112] Takács, L. (1963). A single-server queue with feedback. *Bell System Technical Journal*, 42, 509-519.

[113] Takagi, H. (1991). *Queueing Analysis: A Foundation of Performance Evaluation, Volume 1: Vacation and Priority Systems.* Elsevier Science Publishers.

[114] Takagi, H. (1993). *Queueing Analysis: A Foundation of Performance Evaluation, Volume 2: Finite Systems.* Elsevier Science Publishers.

[115] Takagi, H. (1993). *Queueing Analysis: A Foundation of Performance Evaluation, Volume 3: Discrete Time Systems.* Elsevier Science Publishers.

[116] Teghem, J. (1986). Control of the service process in a queueing system. *European Journal of Operations Research*, 23, 141-158.

[117] Thiruvengadam, K. (1963). Queueing with breakdown. *Operations Research*, 11, 62-71.

[118] Tweedie, R. L. (1976). Criteria for classifying general Markov chains. *Advances in Applied Probability*, 8, 737-771.

[119] Wang, J. T., Cao, J. H. and Li, Q. (2001). Reliability analysis of the retrial queue with server breakdowns and repairs. *Queueing Systems*, 38, 363-380.

[120] Wolff, R. (1982). Poisson arrivals see time averages. *Operations Research*, 30, 223-231.

[121] Wolff, R. (1989). *Stochastic Modeling and the Theory of Queues.* Prentice-Hall, Englewood Cliffs, New Jersey.

[122] Wu, X., Brill, P. H., Hlynka, M. and Wang, J. (2005). An $M/G/1$ retrial queue with balking and retrials during service, *International Journal of Operational Research*, 1, 30-51.

[123] Yang, T. and Templeton, J. G. C. (1987). A survey on retrial queues. *Queueing Systems*, 2, 201-233.

[124] Yang, T. and Li, H. (1994). The $M/G/1$ retrial queue with the server subject to starting failures. *Queueing Systems*, 16, 83-96.

[125] Yang, T. and Li, H. (1995). On the steady-state queue size distribution of the discrete-time $Geo/G/1$ queue with repeated customers. *Queueing Systems*, 21, 199-215.

[126] Yang, T., Posner, M. J. M. and Templeton, J. G. C. (1990). The $M/G/1$ retrial queue with nonpersistent customers. *Queueing Systems*, 7, 209-218.

[127] Yang, T., Posner, M. J. M., Templeton, J. G. C. and Li, H. (1994). An approximation method for the $M/G/1$ retrial queue with general retrial times. *European Journal of Operational Research*, 76, 552-562.

Vita Auctoris

Name: Xiaoyong Wu

Place of Birth: Jiangxi, China

Education: B.S. in Mathematics, 1987, Nanchang University of China

M.S. in Mathematics, 1995, Beijing Normal University of China

Ph.D. in Statistics, 2005, University of Windsor, Canada