Stochastic Modelling: Analysis and Applications

On Queues with Interruption and Protection

Thesis submitted to the
Cochin University of Science and Technology
for the award of the degree of

DOCTOR OF PHILOSOPHY

under the Faculty of Science by

GOPAKUMAR B

Department of Mathematics
Cochin University of Science and Technology
Cochin - 682 022

September 2014

TO

MY PARENTS

Certificate

This is to certify that the thesis entitled ***On Queues with Interruption and Protection*** *submitted to the Cochin University of Science and Technology by* Mr. GOPAKUMAR B *for the award of the degree of Doctor of Philosophy under the Faculty of Science is a bona fide record of studies carried out by him under my supervision in the Department of Mathematics, Cochin University of Science and Technology. This report has not been submitted previously for considering the award of any degree, fellowship or similar titles elsewhere.*

Dr. A. KRISHNAMOORTHY

September 17, 2014.

(Supervising Guide)

Emeritus Professor

Department of Mathematics

Cochin University of Science and Technology

Cochin- 682022, Kerala.

Declaration

I, GOPAKUMAR B, hereby declare that this thesis entitled **On Queues with Interruption and Protection** contains no material which had been accepted for any other Degree, Diploma or similar titles in any University or institution and that to the best of my knowledge and belief, it contains no material previously published by any person except where due references are made in the text of the thesis.

Gopakumar B

September 17, 2014.

Research Scholar

(Part Time, Reg. No.3364)

Department of Mathematics

Cochin University of Science and Technology

Cochin-682022, Kerala.

Acknowledgement

A few words of gratitude. I express my gratitude to all those who made this thesis possible.

I would like to express my heartfelt gratitude to my supervisor, Dr. A. KRISHNAMOORTHY, Emeritus Professor, Department of Mathematics, Cochin University of Science and Technology for his guidance during my research. He has been a constant source of inspiration and a tremendous mentor for me. I owe him for the encouragement and support he offered during my research. It is his guidance that has given my research the proper direction. Without his guidance, persistent help and parental affection this thesis would not have been possible.

I should mention the motivation and support from Dr. Viswa nath C Narayanan, Govt. Engineering College, Thrissur. It is the constant interactions with him that made me focussed in my work. He is the co-author of all my papers. I thank him for the immense help provided.

I thank Prof.P.G.Romeo, Head of the Department of Mathematics, for the encouragement and help he had given during my research. I am also grateful to Prof.M.N.Nampoothiri, Prof.B.Lakshmy and Prof.A.Vijayakumar for their encouraging words. I also thank the office staff and librarian of the Department of Mathematics for their support and help of various kinds. My gratitude also goes to the authorities of Cochin University of Science and Technology for the facilities they provided.

I thank Dr K Vijayakumar, The Principal, Govt. Engineering College, Thrissur for his support and good wishes.

I express my feeling of gratitude to my colleagues Dr Sheeba M B, Mr. M P Rajan, Mrs Sindu Mathew, Dr. Seema Varghese for the interest they showed in my work.

My fellow researchers Dr Sajeev S Nair, Mr Sathian M K, Dr.Varghese Jacob, Dr.Sreenivasan, Dr.Pramod, Dr.Manikandan, Dr Ajayan and Mr Manjunath were always ready to discuss many research problems and to share new ideas. I thank all of them for their help.

I pay my homage to late Prof M D Paul, St Thomas College, Thrissur, commemorating the succour he had provided.

I also thank my friend Mr. A K Prakas, for the help he had given me during my difficult days.

I am also extremely thankful to my family for allowing me to divert myself towards the research leaving them behind for a while.

Gopakumar B

On Queues with
Interruption and Protection

Contents

Chapter 1

INTRODUCTION

1.1 Queueing Theory

The study of the congestion in telephone network by A K Erlang lead to the origin of Queueing Theory. Over the years, the subject found its applications in diverse areas like Telecommunications, Traffic flow, Computer networking, Computing etc.. This is a branch of science that deals with the study of waiting lines. When customers/units requiring some kind of service gather at a service centre, a queue is formed. In Queueing Theory, we model such systems mathematically and predict some characteristics like average waiting time of a customer, average queue length etc..

The basic features that characterises a queueing system are the following:

a) **Arrival Pattern:** This describes the manner in which the units arrive and join the system. The customers may arrive in single or in batches. Time interval between any two consecutive arrivals is called the inter-arrival time. The arrival pattern is usually represented by the probability distribution of the inter-arrival time.

   On arrival, if a customer sees a long queue, he may decide not to join the queue and may leave the station. This customer behaviour is called *balking*. Some customers join the queue, wait for a while but loosing their patience, may leave the system without waiting further for service. This situation is referred to as *reneging*.

   If there are more than one queue, customers have a tendency to switch from one queue to another. This is called *jockeying*.

b) **Service Pattern:** This indicates the manner in which the service is rendered. Like the arrivals, the service also is provided in single or in batches. The probability distribution of the service time describes the service pattern.

c) **Queue Discipline:** Queue Discipline tells us the rule by which the customers are taken for service. Some of the commonly used disciplines include *first in first out* (FIFO), *last in first out* (LIFO), *service in random order* (SIRO) and *server sharing*. In some systems customers may be given priorities so that the service is rendered in the order of their priorities.

d) **Number of service channels:** This refers to the number of servers providing service to the customers in the system.

e) **Capacity of the system:** The capacity of the system is the maximum number of customers it can accommodate. It may be finite or infinite.

A queueing system is often analysed by modelling it as a Markov chain. Some basic concepts employed in this direction are given briefly in the following sections.

1.2 Some basic concepts

1.2.1 Stochastic Process

A *stochastic process* or a *random process* is a collection

$$\{X(t)/t \in T\}$$

of random variables where T is some index set. The index t is usually referred to as time. If the index set T is countable, then the process is called a *discrete time process*. Otherwise it is called a *continuous time process*. The set of all possible values of $X(t)$ for each $t \in T$ is the *state space* of the process.

1.2.2 Counting Process

A counting process $\{N(t)/t \geq 0\}$ is a stochastic process if $N(t)$ represents the total number of events occurred by time t such that

i) $N(t) \geq 0$.

ii) $N(t)$ is integral valued.

iii) If $s < t$, then $N(s) \leq N(t)$.

iv) For $s < t, N(t) - N(s)$ is the number of events occurred in the interval $(s, t]$.

A counting process $\{N(t)/t \geq 0\}$ is said to be have *independent increments* if for all $t_1, t_2, ..., t_n, \quad t_1 < t_2 < ... < t_n$, the random variables $N(t_2) - N(t_1), N(t_3) - N(t_2), ..., N(t_n) - N(t_{n-1})$ are independent. It has *stationary increments* if the distribution of $N(t) - N(s)$ depends only on $t - s$.

1.2.3 Markov Process

Definition 1.2.1. A stochastic process $\{X(t)/t \in T\}$ is called a Markov Process if

$$P\left[X(t_n) = x_n / X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, ..., X(t_0) = x_0\right]$$
$$= P\left[X(t_n) = x_n / X(t_{n-1}) = x_{n-1}\right]$$

whenever $t_0 < t_1 < ... < t_{n-1} < t_n$ for every n.

A discrete time Markov Process is called a Markov chain. Thus a *Markov chain* is a stochastic process $\{X_n/n = 0, 1, 2, ...\}$

for which

$$P\left(X_n = x_n / X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, ..., X_0 = x_0\right)$$
$$= P\left(X_n = x_n / X_{n-1} = x_{n-1}\right)$$

for $x_0, x_1, ..., x_n$ in the state space.

The probability $p_{i,j}^{n,n+1} = P\left(X_{n+1} = j / X_n = i\right)$ is called one step transition probability of the Markov chain. If the transition probabilities are independent of time, the chain is said to be *homogeneous*. If it depends on time, the chain is called *non-homogeneous*. For a homogeneous Markov chain, the one step transition probabilities are denoted by $p_{i,j}$. The matrix $P = [p_{i,j}]$ is called the one step transition matrix of the chain.

More generally the probability that starting from a state i, the chain reaches the state j in exactly m transitions is called the $m -$ *step transition probability*. For a homogeneous chain this probability is denoted by $p_{ij}^{(m)}$. Thus

$$p_{ij}^{(m)} = P\left(X_{n+m} = j / X_n = i\right).$$

A subset C of the state space of a chain is said to be *closed* if no state outside C can be reached from any state in C. If the chain has no proper closed subset other than the state space itself, it is called an *irreducible chain*.

A state i is *recurrent* if and only if, starting from state i, the probability of returning to state i after some finite time is certain. A non-recurrent state is said to be *transient*. For a recurrent state if the mean recurrence time is finite, it is called *positive recurrent*. The greatest common divisor of the recurrence times of a state is called its *period*. If the period is one, the state is said to be *aperiodic*. A positive recurrent aperiodic state of a Markov chain is said to be *ergodic*. A Markov chain is ergodic if all its states are ergodic.

For a homogeneous Markov chain, the vector

$$\pi = (\pi_1, \pi_2, \pi_3, ...)$$

is called a *stationary probability vector* if for every j in the state space,

$$\pi_j = \sum_i \pi_i p_{ij} \text{ such that } 0 \leq \pi_j \leq 1 \text{ and } \sum_j \pi_j = 1.$$

An irreducible chain has a stationary distribution if and only if all of its states are positive recurrent.

The probability vector $\varpi = (\varpi_1, \varpi_2, ...)$ is called the limiting distribution of the chain if $p_{ij}^{(n)} \to \varpi_j$ as $n \to \infty$.

If a positive recurrent chain is both irreducible and aperiodic, it has a limiting distribution.

Theorem 1.2.1. *For an irreducible ergodic Markov chain, the limiting distribution exists and is same as its stationary distribution.*

1.3 Modelling Tools

In this section we describe some tools we used in analysing the models introduced in this thesis.

1.3.1 Exponential Distribution

A random variable X is said to have an exponential distribution with parameter $\lambda > 0$ if it has the probability density function

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x}, 0 \leq x < \infty \\ &= 0, x < 0. \end{aligned}$$

Its distribution function is given by $F(x) = 1 - e^{-\lambda x}, x \geq 0$.

The mean of this distribution is $\frac{1}{\lambda}$ and variance is $\frac{1}{\lambda^2}$. The moment generating function is $M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}$.

The important properties that make exponential distribution much useful in modelling queueing systems are the following:-

a) **Memoryless property (Non-ageing property):** This property implies that if X denotes the duration of some activity, and if the the activity is still going on, the distribution of the duration of the remaining part of the activity is same as that of X, no matter when the activity has begun. In other words the remaining part of the activity can be treated as a new activity.

$$\text{ie; } P\left(X \geq x+y | X \geq x\right) = P\left(X \geq y\right).$$

Exponential distribution is the only continuous distribution having this property.

b) **Minimum of two exponential variables is exponential:** Let X_1 and X_2 be two exponential random variables with parameters λ_1 and λ_2 then $min\left(X_1, X_2\right)$ is exponential with parameter $\lambda_1 + \lambda_2$. Also $P\left(X_i < X_j\right) = \frac{\lambda_i}{\lambda_i + \lambda_j}$.

1.3.2 Poisson Process

A counting process $\{N(t)/t \geq 0\}$ is called a Poisson Process with rate $\lambda > 0$ if

i) $N\left(0\right) = 0$.

ii) It has stationary and independent increments.

iii) The distribution of $N(t)$ is Poisson with mean λt.

$$\text{ie;} \quad P\left(N(t) = n\right) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, n = 0, 1, 2, ... \quad .$$

.

A detailed description on Poisson process and related distributions is given in Medhi [53]. We state two important theorems.

Theorem 1.3.1. *For a homogeneous Poisson process with mean λt, the inter-occurrence times are independently and identically distributed exponential random variables with mean $\frac{1}{\lambda}$.*

Theorem 1.3.2. *If the interval between successive occurrences of an event E are independently and exponentially distributed with mean $\frac{1}{\lambda}$, then the events E will form a Poisson process with mean λt.*

1.3.3 Phase Type Distribution

Consider a Markov Process on the states $\{1, 2, 3, ..., m, m+1\}$ with the infinitesimal generator

$$Q = \begin{bmatrix} T & T^0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{1.1}$$

where the $m \times m$ matrix T satisfies $T_{ii} < 0$ for $1 \leq i \leq m$, and $T_{ij} \geq 0$, for $i \neq j$. Also $Te + T^0 = 0$. Let initial probability vector of this process be (α, α_{m+1}) with $\alpha e + \alpha_{m+1} = 1$. Also assume that the states $1, 2, .., m$ are transient so that absorption into the state $m+1$ is certain.

Definition 1.3.1. A probability distribution $F(.)$ on $[0, \infty)$ is said be a phase type distribution (PH-distribution) of order m with representation (α, T) if and only if it is the distribution of the time until absorption of a finite Markov process defined in (1.1).

If $F(.)$ is a phase type distribution described by the Markov process defined in (1.1), then

$$F(x) = 1 - \alpha.exp(Tx)\boldsymbol{e}, \text{ for } x \geq 0.$$

For a PH distribution $F(.)$ with representation (α, T),

i) The distribution $F(.)$ has a jump at $x = 0$ of magnitude α_{m+1}.

ii) The corresponding probability density function $f(.)$ is given by

$$f(x) = \alpha.exp(Tx)T^0, x \geq 0.$$

iii) The Laplace-Stieltjes transform $f(s)$ of $F(.)$ is given by

$$f(s) = \alpha_{m+1} + \alpha(sI - T)^{-1}T^0, \text{ for } Re(s) \geq 0.$$

iv) The i^{th} raw moment ${\mu'}_i$ is given by

$${\mu'}_i = (-1)^i i! \left(\alpha T^{-i}\boldsymbol{e}\right), i = 1, 2, 3, \ldots .$$

Example 1.3.1. Erlang distribution

A random variable X is said to follow an Erlang-k distribution, $k = 1, 2, 3, ..$ if it has the probability density function

$$f(x) = \frac{(\mu x)^{k-1}}{(k-1)!}\mu e^{-\mu x}.$$

The mean of this distribution is $\frac{k}{\mu}$ and variance is $\frac{k}{\mu^2}$. Its moment generating function is $M_X(t) = (1 - \frac{t}{\mu})^{-k}$.

From the moment generating function it follows that the sum of k mutually independent exponential random variables, each with common population mean $\frac{1}{\mu}$ is an Erlang-k distribution with mean $\frac{k}{\mu}$.

Now consider a random variable X with phase type probability distribution F(.) represented by (α, T) where

$$\alpha = (1, 0, 0, ..., 0)_{1\times m} \quad \text{and}$$

$$T = \begin{bmatrix} -\mu & \mu & 0 & . & . & . & . & 0 \\ 0 & -\mu & \mu & 0 & . & . & . & 0 \\ 0 & 0 & -\mu & \mu & 0 & . & . & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & . & . & 0 & . & -\mu & \mu \\ 0 & 0 & . & . & 0 & . & -\mu & \mu \end{bmatrix}_m$$

Since the corresponding Markov process always start from the first phase, the time until absorption, X is the sum of time spent in each of the m phases. Hence X is the sum of m exponentially distributed random variables with mean $\frac{1}{\mu}$. That is the distribution of X is Erlang-m. Thus Erlang distribution is a phase type distribution.

Example 1.3.2. Exponential distribution

When $k = 1$ the Erlang distribution reduces to Exponential distribution. Hence Exponential distribution can be considered as an Erlang-1 distribution. Therefore exponential distribution is a phase type distribution with a single phase.

1.4 Matrix Analytic Methods

When Queueing theory found its applications in several new areas like computer networking, mobile phone communications etc., the usual methods like Method of generating functions, Methods using Transforms etc. failed to provide much tractability in the analysis of many queueing models especially when the distribution of inter-arrival time or service time is not exponential. But the introduction of Matrix analytic methods gave us the ability to analyse much complicated Stochastic models in an algorithmic way and to numerically explore the problems more deeply. In this thesis, the Matrix analytic methods are used to analyse quasi-birth-and-death processes.

1.4.1 Level independent quasi-birth-and-death processes

Consider a Markov process with state space

$$E = \{(i,j), i \geq 0, 1 \leq j \leq m\}.$$

We partition the state space as

$$E = \bigcup_i E_i \text{ where } E_i = \{(i,j), 1 \leq j \leq m\}.$$

The states in E_i are said to be in level i. Such a Markov process is called a level independent quasi-birth-and-death process

(LIQBD) if its infinitesimal generator is the irreducible tridiagonal matrix Q given by

$$Q = \begin{bmatrix} B_0 & A_0 & & & & & \\ B_1 & A_1 & A_0 & & & & \\ & A_2 & A_1 & A_0 & & & \\ & & A_2 & A_1 & A_0 & & \\ & & & \cdots & \cdots & \cdots & \end{bmatrix}.$$

Then we have the following theorem (Neuts [49]).

Theorem 1.4.1. *The process Q is positive recurrent if and only if, the minimal non-negative solution R to the matrix quadratic equation*

$$R^2 A_2 + RA_1 + A_0 = 0 \tag{1.1}$$

has spectral radius less than one and the finite system of equations

$$\begin{aligned} x_0 \left(B_0 + RB_1\right) &= 0 \\ x_0 \left(I - R\right)^{-1} e &= 1 \end{aligned} \tag{1.2}$$

has a unique positive solution for x_0. If the matrix $A = A_0 + A_l + A_2$ is irreducible, then $sp(R) < 1$ if and only if, $\pi A_0 e < \pi A_2 e$, where π is the stationary probability vector of the matrix A. The stationary probability vector $x = (x_0, x_l, x_2, ...)$ of Q is given by

$$x_i = x_0 R^i, i \geq 0. \tag{1.3}$$

To find the solution R of equation(1.1), we use the iterative formula

$$R_n = -A_0 \left(A_1 + R_{n-1} A_2 \right)^{-1}, n = 1, 2, 3, ... \tag{1.4}$$

with an initial approximation R_0. If $sp(R) < 1$ then R_n converges to R. More powerful iterative methods can be found in Latouche and Ramaswami [50].

1.4.2 Level dependent quasi-birth-and-death processes

A Level dependent quasi-birth-and-death process (LDQBD) is a Markov process with state space

$$E = \{(i, j)/i \geq 0, 1 \leq j \leq n_i\}$$

whose infinitesimal generator Q is given by

$$Q = \begin{bmatrix} A_{1,0} & A_{0,0} & & & & & \\ A_{2,1} & A_{1,1} & A_{0,1} & & & & \\ & A_{2,2} & A_{1,2} & A_{0,2} & & & \\ & & A_{2,3} & A_{1,3} & A_{0,3} & & \\ & & & \cdots & \cdots & \cdots & \\ & & & & \cdots & \cdots & \cdots \end{bmatrix}.$$

The state space is partitioned into different levels where level i is given by $E_i = \{(i, j)/1 \leq j \leq n_i\}, i = 0, 1, 2, ...$. The transitions are to the adjacent levels alone. But the transition rate will depend on the level in which the process is then. Assuming that the process is irreducible, we have the following theorems (Latouche and Ramaswami [50]).

Theorem 1.4.2. *If an LDQBD is aperiodic and positive recurrent, its limiting probability vector* $\pi = \{\pi_1, \pi_2, \pi_3, ...\}$ *satisfies the relation*

$$\pi_n = \pi_{n-1} R_n, n \geq 1$$

where the matrices R_n *are the minimal non-negative solutions of the system of equations*

$$R_n R_{n+1} A_{2,n+1} + R_n A_{1,n} + A_{0,n} = 0.$$

Theorem 1.4.3. *The LDQBD is positive recurrent if and only if there exists a strictly positive solution of the system*

$$\pi_0 = \pi_0 \left(A_{1,0} + R_1 A_{2,1} \right)$$

normalized by

$$\pi_0 \sum_{n \geq 0} \left(\prod_{1 \leq k \leq n} R_k \right) e = 1.$$

To calculate R_n we use Nuets - Rao Truncation method (Neuts and Rao [51]) for retrial queues. In this method an upper level N is selected such that the transitions between the levels higher than N are independent of the level using the approximation that

$$A_{2,i} = A_{2,N}, A_{1,i} = A_{1,N} \text{ and } A_{0,i} = A_{0,N} \text{ for } i > N.$$

For retrial queues this makes sense since if the number of retrying customers are very large, most of the retrials fail. So the retrials exceeding a large number N will have no effect on the system. Then R_N is the the minimal non-negative solution of the equation

$$R_N^2 A_{2,N} + R_N A_{1,N} + A_{0,N} = 0.$$

Using this R_N we can find the steady state vector π_N which converges to π as $N \to \infty$.

In more general cases we choose the method proposed by Bright and Taylor [52].

1.5 Summary of the thesis

In our day to day life, in processes like banking, internet, business, agriculture, scientific experiments etc., we are faced with different kinds of interruptions like a power failure affecting a banking procedure or the working of certain machinery. Though the facilities are improving/increasing each day, which reduce the severity of interruptions, increasing needs bring them back.

For an example consider the following situations. Suppose we have a computer with a backup of one hour. This backup is sufficient for many purposes like saving a work that got interrupted due to power failure. However, suppose that we have to run certain program, which requires more than one hour for execution on this computer. While running such a program, we are worried about a power failure. This example applies to many other situations, as the world is run by computers.

Another example from day to day life: Consider a work person who uses a power tool that runs on electricity, who has more than one work sites to attend. If a power failure lasts for long or repeats randomly on one work site, he may choose to shift to another site with constant power supply.

An emergency call for an ambulance service getting interrupted due to network issues is yet another example for interruptions in day to day life.

An ideal world, as one would expect, be that which is free from all types of interruptions. However, as we discussed earlier, we do not belong there yet. Facing the reality of interruption, we hope to minimize its severity by introducing some protection mechanism. Analysis of queueing models with service interruption and protection is therefore important.

Naturally, researchers got involved in modelling these scenarios in a queueing theory perspective. A brief review on the related works is added in the beginning of each chapter. The abundance of works in this direction tells the practical importance. This is the reason for the selection of a few systems in which the service process is susceptible for interruption and having some measures for immunizing the server from it in this thesis. The thesis is arranged in 6 chapters including the present introductory chapter.

In **Chapter 2**, we consider a single server queueing system where the service time distribution is phase type. The service process may face some interruptions during the service. The interruption occurs according to a Poisson process. Interruptions are assumed to occur only when a service is in progress and not when the server is idle. The interrupted service is either resumed or repeated based on which of the two renewal processes, started simultaneously with the interruption, renews first. Customers arrive according to a Poisson process with different means depending on whether the server is interrupted or not. The customers waiting in the queue for service may leave the system without waiting further for service while the server is interrupted. Stability of the above system is analysed and steady state vector is calculated using Neuts-Rao truncation. A thorough numerical study of various performance measures such as mean and variance of waiting time of a customer are carried out.

Chapter 3 analyses a single server retrial queueing model with service interruptions, resumption/repeat of interrupted service. On arrival if a customer finds an idle server, he is immediately taken for service. If the server is busy when a customer arrives, this customer goes to an orbit of infinite capacity from where he makes repeated attempts for service according to a Poisson process. After an unsuccessful retrial he rejoins the orbit with probability p or leaves the system without waiting for service with probability $q = 1 - p$. The service time durations follow PH distribution when there is no interruption. The service process is subject to interruptions, which occur according to a Poisson Process. The interrupted service is either resumed or repeated as in the model described in chapter 2. The system is found to be always stable if $q > 0$. The case $q = 0$ is also analysed. Using Matrix analytic method, expressions for important system characteristics such as expected service time, expected number of interruptions etc. are obtained. System performance measures are numerically explored and the effect of service interruptions in a retrial set up is studied.

Chapter 4 is devoted to a model in which service time distribution is Erlang of order m. The server is subjected to interruption. The arrival of customers as well as the occurrence of interruptions is according to a Poisson Process. The interrupted server is taken for repair immediately. The repair time follows exponential distribution. The interrupted service is either resumed or restarted after repair according to the time taken for repair to be done. As a means to reduce the impact of interruption, a protective mechanism is employed. To reduce the chance of a service reaching completion being restarted all over again, the final n phases of service are immunized from interruption. Thus a service that completed the first $m - n$ phases will no longer face any interruption. The condition for stability is determined and the service process is thoroughly analysed. The steady state probabilities are evaluated using Matrix Analytic method. Many system performance measures like expected number of customers in the system, expected waiting time, expected number of interruptions during a service, expected interruption duration etc. are also investigated. A cost analysis is done numerically to find the number of phases to be protected at an optimum cost.

In **Chapter 5**, we consider a single server queueing system where customers arrive according to a Poisson process. Service time distribution is exponential. The service process is subject to interruptions, which occurs according to a Poisson process. We assume that during interruption, the customer being served waits there until his service is completed. The interrupted service is restarted after repair. Repair time is exponentially distributed. To minimize the loss due to the interruptions, some protection is given to the server. There will be no interruption if the server is in protected mode. But the way in which the server is protected differs from the method adopted in the previous chapter. Here the server is brought to the protected mode after a random time from the start of the service. Stability of the above system is analysed and steady state vector is calculated. Explicit formulas for system performance measures such as expected number of customers in the system, expected interruption rate, waiting time of a customer in the system etc. are also obtained. A cost analysis is also done numerically to find the time after which the service has to be protected at an optimum cost.

In **Chapter 6**, a system similar to one discussed in the previous chapter is analysed, but in this model the service time has Erlang-m distribution. The strategy used to protect the service is the same as the one used in chapter 5. The condition for stability of the system is obtained. The service process is well studied and important system performance measures are evaluated. A cost analysis is made to determine the optimum time at which the protection is to be started in a cost effective manner. A comparison between the two strategies of protection is also done.

It may be noted that the protection mechanism introduced in chapter 4 looks similar to the N-policy in queueing system. In contrast those introduced in chapters 5 and 6 are similar to the T-Policy.

Chapter 2

A QUEUEING MODEL WITH INTERRUPTION, RESUMPTION/REPEAT AND RENEGING

2.1 Introduction

In the fast growing field of communication networks like Internet, Queueing models with interruption have an important role to play. Service interruption models studied in the literature include different types of service unavailability that may be due to server taking vacations, server breakdowns, customer induced interruptions, arrival of a priority customer etc. Queues

---

1. Presented in the International Symposium on Probability Theory and Stochastic Process held in honour of Prof.S R S Varadhan FRS, Feb 6-9,2009, at Cochin University of Science and Technology, Kochi.
2. Published in Bulletin of Kerala Mathematical Association, Special Issue, Guest Editor S R S Varadhan 29-45, October 2009.

with service interruption was first studied by White and Christie [1] which is an $M/M/1$ queueing model with exponentially distributed service interruption durations. They assumed that at the end of each interruption the service of the interrupted customer is resumed (from where it got interrupted). Some of the earlier papers, which analyse queueing models with service interruptions, assuming general distribution for the service and interruption times, are by Jaiswal [2] ,[3], Gaver, Jr. [4], Keilson [5], Avi-Itzhak and Naor [6] and Thiruvengadam [7]. In all these papers it is assumed that the arrival of a higher priority customer interrupts the service of a low priority customer. Some other papers on service interruption models include Federgruen and Green [8], Van Dijk [9], Takine and Sengupta [10], Masuyama and Takine [11]. Among these, Gaver, Jr. [4] considers a pre-emptive repeat or repeat and re-sampling, whereas Keilson [5] considers repeat or resumption of the interrupted service. The rest of the above mentioned works consider pre-emptive resume discipline.

The queueing model analysed by Krishnamoorthy and Usha kumari [12] where disaster can occur to the unit undergoing service and the one by Wang, Liu and Li [13] with disaster and unreliable server can be considered as models with service interruption. Vacation to server can also be considered as a particular type of service interruption. We refer to Doshi [14] for some general decomposition results for vacation models. Takagi [15] gives a detailed analysis on vacation queueing models. Classical vacation models assume that the vacation (either single or multiple) period starts at a service completion epoch, either due to absence of customers or when a preassigned number of customers are already served since the last vacation. Deviating from this Takagi and Leung [16] analysed a queueing system where the server takes vacation when the service period exceeds certain specified duration. The interrupted service is resumed after vacation. In a discrete $MAP/PH/1$ queue analysed by Alfa [17] a similar situation is considered. During vacation the server may attend other work. In both the models, vacation has the nature of a service interruption. Li and Tian [18] introduces a vacation model where the vacation can be interrupted by assuming that the server can come back to the normal working level, without completing the vacation period. Boxma, Mandjes and Kella [19] study a single server vacation model where the length of vacation depends on the length of the previous active period. Gray, Wang and Scot [20] analyse a vacation queueing model where service breakdowns can occur during a service and the service is resumed after the repair; nevertheless they assume that the vacation period starts after a busy period.

In almost all papers on queues with service interruptions, the service is either resumed or repeated on removal of interruption. Fiems, Maertens and Bruneel [21], consider a queueing system with different types of server interruptions, namely destructive and non-destructive. They assume that the interrupted service is repeated after a destructive interruption and resumed from where it stopped in the other case. Following this there has been an extensive study on such models by numerous authors.

Recently Krishnamoorthy and co-authors studied a few queueing models with interruptions, where a special stress was given in deciding whether to repeat or resume an interrupted service. For a detailed report on queueing models with interruption, we refer to the papers by Krishnamoorthy and Pramod [23]. A recent paper by Krishnamoorthy, Pramod and Deepak [22] considers a queueing model with service interruption and repair where the decision on whether to repeat or resume the interrupted service is taken after completion of interruption according to the realization of a phase type distributed random variable.

Customer impatience due to waiting long in the queue being a common phenomena in many queueing systems have attracted various researchers to consider this while modelling. Pioneering studies in this direction are by Haight [24],[25] and Barrer [26]. Among these, [24] was on balking of customers and the other two were on reneging of customers. These studies then followed by the one due to Ancker Jr.and Gafarian [27], who assumed negative exponential distribution to model the time to renege as well as the service time and obtained several important system performance measures. Haghighi, Medhi and Mohanty [28] studied a multi-server queueing system with balking and reneging. Wang and Chang [29] study a queueing system with balking reneging and server breakdowns. The studies by Zhang, Yue D and Yue W [30], Yue D, Yue W and Sun Y [31] and Altman and Yechiali [32],[33] combines the notions of customer impatience and server vacations in a queueing model. Baruah, Madan and Eldabi [34] consider a queueing system, where the customers may chose to renege during system break downs and server vacations.

The queueing model considered in the present chapter differs mainly from that discussed in [22] in three aspects: first, here two random variables compete on the onset of interruption to decide whether to repeat or resume the interrupted service (as in [22]); second, no repair time is assumed if the decision is to repeat the service after the interruption and finally we assume reneging of customers during a service interruption. Here one may wonder why two random variables compete to determine the nature of service to be provided, namely, resumption or repetition. After the interruption if the repair takes too much time to damage the service rendered till the interruption or to make the customer impatient, the service is repeated without waiting for the repair to be completed with another identical server if necessary. On the other hand if the repair is done quickly, the service is resumed from where it was interrupted. This policy differs from that in Fiems, Maertens and Bruneel [21] because in the latter the nature of the recommenced service on completion of interruption is determined at the epoch of onset of interruption. Another feature of the model discussed in this chapter, unlike many others on service interruption, is that the interruption may not need repair. This is the case when the interrupted server is replaced by an identical one instantaneously or the interruption is not due to a server failure. Here we wish to point out that such a case may arise if we combine the models in Alfa [17], Takagi and Leung [16], where the service interruption is due to vacation, with the model in Li and Tian [18] where a server on vacation has to come back without completing the vacation period.

While modelling systems with service interruptions, repetition of service on completion of an interruption can be one of the following:

i) Repeat identical: In this case, the service on completion of interruption has the same distribution as the one offered prior to the onset of interruption.

ii) repeat same: This is much more complex than the first case.repeated service has to follow the same pattern as that prior to the interruption. This means one has to keep all relevant information about the earlier service.

In the model described here as well as in the other chapters we consider identical repetition of interrupted service.

One real life situation where the model in this chapter is appropriate is the following. Consider a person downloading a software from some site. The downloading may be interrupted for some reason; may be the server site becoming jammed by too many users or it may be some ISP problems or virus attack. Now at this point, the downloading is disrupted and the patience of the person who is browsing starts to decay and he/she may decide to repeat the whole downloading process; or it may happen that before his/her patience reaches the threshold, the download may resume from where it stopped. This is a usual phenomenon for sites which have more visitors than its capacity.

2.2 The mathematical model

We consider a single server queueing system in which the service time follows PH distribution with representation $(\boldsymbol{\alpha}, S)$ of order m. The service is interrupted at an exponentially distributed duration with parameter θ. At the epoch when an interruption occurs, two renewal processes, namely, resume clock and repeat clock are started, realization times of which follow exponential distribution with parameters γ and δ, respectively. If the realization of the resume clock occurs first, the interrupted service is resumed whereas if the repeat clock realizes first then the interrupted service has to be repeated. The customers arrive to the system according to a Poisson process with rate λ_1 while the service is stopped due to an interruption and with rate λ_0 otherwise. At the stoppage of a service due to interruption the customers, except the one being served, may leave the system without waiting for service. Such reneging of customers is assumed to follow Poisson distribution with rate $k\beta$ when there are k customers waiting for service. Interruptions are assumed to occur only when a service is in progress and not when the server is idle. We consider only the case in which, when a service is interrupted, no further interruption befalls on that until the present interruption is cleared. This situation resembles the type I counter (see for example Karlin and Taylor [35]).

A diagrammatic representation of the model is given in Figure (2.1).

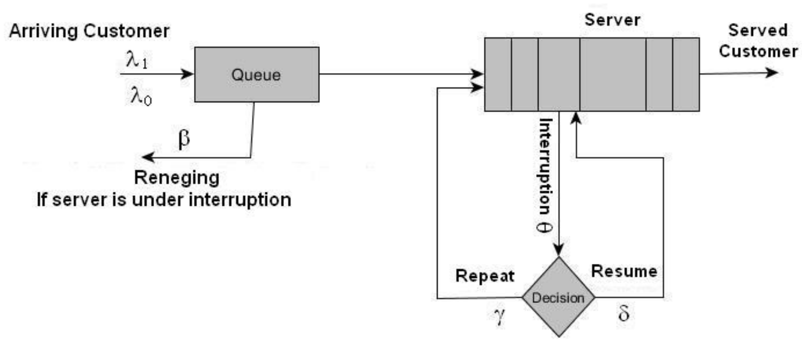


Figure 2.1: *An $M/Ph/1$ Queue with service interruptions.*

Let $N(t)$ be the number of customers in the orbit, $S(t)$ the status of the server which is 0 or 1 according as the server is uninterrupted or interrupted and $J(t)$ the phase of the service process at time t. Then the above model can be represented by the Markov process $\boldsymbol{\mathcal{X}} = \{X(t)/t \geq 0\} = \{(N(t), S(t), J(t))\,/t \geq 0\}$. The state space is $\{0\} \cup (\{1, 2, 3, ...\} \times \{0, 1\} \times \{1, 2, 3, ..., m\})$.

The one step transitions of the above process from a state are restricted to the states in the same level or to states in one level up or one level down. The level decreases by one when a service completion occurs or reneging occurs while the ongoing service is facing an interruption. The rate at which a service completion occurs at level $k, k \geq 0$ in the phase j is s_j^0 and the probability that of the next service starts in the phase i is α_i. Hence the rate at which transitions happens from $(k, 0, j)$ to $(k-1, 0, i)$ is $\alpha_i s_j^0$. The level goes down by one while a reneging occurs consequent to the the server being interrupted. Such an event will not alter the phase of the service. For such transitions the rate is $(k-1)\beta$. These transitions are depicted in Figure 2.2(a)

The only way the level is increased by one is the arrival of a customer. As this happens at rate λ_1 or λ_0 according as the server is facing an interruption or not and leaves the phase undisturbed, we see that the transition rate from (k, i, j) to $(k+1, i, j)$ is $\lambda_1$while there is an interruption $(i = 1)$ and λ_0 otherwise $(i = 0)$. Figure 2.2(b) illustrates such transitions.

Now the transitions that will not change the level are those among the phases, given by S, occurrences of interruptions at the rate θ, realizations of resume clock at the rate γ and that of repeat clock at the rate δ as shown in Figure (2.2(c))

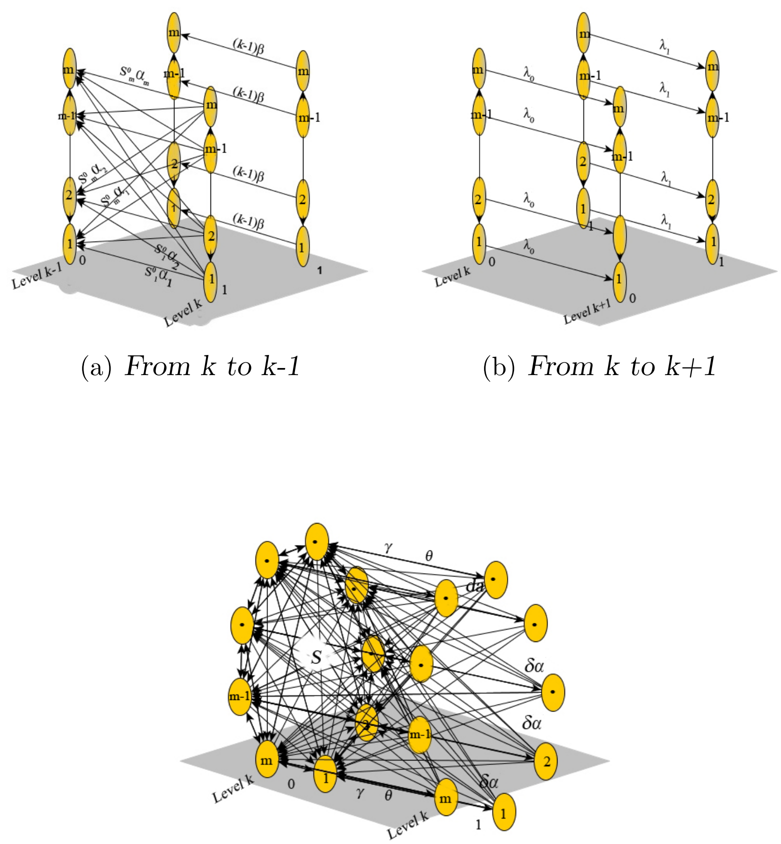
These transitions can be described by the following matrices

$$A_0 = \begin{bmatrix} \lambda_0 I & \mathbf{0} \\ \mathbf{0} & \lambda_1 I \end{bmatrix}$$

$$A_{1,k} = \begin{bmatrix} S - (\theta + \lambda_0) I & \theta I \\ \gamma I + \delta \boldsymbol{e\alpha} & -(\gamma + \delta + (k-1)\beta + \lambda_1) I \end{bmatrix}$$

and

$$A_{2,k+1} = \begin{bmatrix} S^0\boldsymbol{\alpha} & \mathbf{0} \\ \mathbf{0} & k\beta I \end{bmatrix}, k = 1, 2, 3, \dots \ .$$



(a) *From k to k-1* (b) *From k to k+1*

(c) *From k to k*

Figure 2.2: *State transition diagrams*

Hence the infinitesimal generator matrix $\boldsymbol{Q}$ is given by

$$\boldsymbol{Q} = \begin{bmatrix} A_{1,0} & A_{0,0} & & & & \\ A_{2,1} & A_{1,1} & A_0 & & & \\ & A_{2,2} & A_{1,2} & A_0 & & \\ & & \cdots & \cdots & \cdots & \\ & & & \cdots & \cdots & \cdots \end{bmatrix}$$

where

$$A_{1,0} = \begin{bmatrix} -\lambda_0 \end{bmatrix} \ A_{0,0} = \begin{bmatrix} \lambda_0 \boldsymbol{\alpha} & 0\boldsymbol{\alpha} \end{bmatrix} \ A_{2,1} = \begin{bmatrix} S^0 \\ \mathbf{0} \end{bmatrix}.$$

2.3 Steady State Analysis

2.3.1 Neuts Rao Truncation

Since the model described in the previous section is a level dependent QBD, We use an algorithmic solution based on Neuts-Rao Truncation method (Neuts and Rao [51]) for further analysis. Application of this method modifies the process $\boldsymbol{\mathcal{X}}$ into the process $\tilde{\boldsymbol{\mathcal{X}}}$ with infinitesimal generator

$$\tilde{\mathbf{Q}} = \begin{bmatrix} A_{1,0} & A_{0,0} & & & & & & & \\ A_{2,1} & A_{1,1} & A_0 & & & & & & \\ & A_{2,2} & A_{1,2} & A_0 & & & & & \\ & & & \cdots & \cdots & \cdots & & & \\ & & & & A_{2,N-1} & A_{1,N-1} & A_0 & & \\ & & & & & A_2 & A_1 & A_0 & \\ & & & & & & A_2 & A_1 & A_0 \\ & & & & & & & \cdots & \cdots & \cdots \end{bmatrix}$$

where $A_1 = A_{1,N}$ and $A_2 = A_{2,N}$.

2.3.2 Stability condition for the truncated system

If $\boldsymbol{\pi} = (\boldsymbol{\pi_0}, \boldsymbol{\pi_1})$ is the stationary probability vector of the generator matrix $A = A_0 + A_1 + A_2$, then we know that the system is stable if and only if $\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}$. For the truncated model described above, we have

$$A = \begin{bmatrix} S - \theta I + S^0 \boldsymbol{\alpha} & \theta I \\ \gamma I + \delta \boldsymbol{e} \boldsymbol{\alpha} & -(\gamma + \delta) I \end{bmatrix}$$

A simple arithmetic using the relations

$$\boldsymbol{\pi_0}\left(S - \theta I + S^0 \boldsymbol{\alpha}\right) + \boldsymbol{\pi_1}\left(\gamma I + \delta \boldsymbol{e} \boldsymbol{\alpha}\right) = \mathbf{0}$$

$$\theta \boldsymbol{\pi_0} - (\gamma + \delta)\, \boldsymbol{\pi_1} = \mathbf{0}$$

$$\boldsymbol{\pi_0} \boldsymbol{e} + \boldsymbol{\pi_1} \boldsymbol{e} = 1$$

gives us

$$\boldsymbol{\pi} A_0 \boldsymbol{e} = \frac{(\gamma + \delta)\lambda_0 + \theta\lambda_1}{\gamma + \delta + \theta}$$

and

$$\boldsymbol{\pi} A_2 \boldsymbol{e} = \frac{(\gamma + \delta)\tilde{\mu} - \delta\theta + (N-1)\beta\theta}{\gamma + \delta + \theta}$$

where

$$\tilde{\mu} = \frac{1}{-\boldsymbol{\alpha}\left(S - \frac{\delta\theta}{\gamma+\delta} I\right)^{-1} \boldsymbol{e}}$$

Hence we have the following theorem.

Theorem 2.3.1. *The system $\tilde{\boldsymbol{\mathcal{X}}}$ is stable if and only if*

$$\frac{(\gamma + \delta)\lambda_0 + \theta\lambda_1}{(\gamma + \delta)\tilde{\mu} - \delta\theta + (N-1)\beta\theta} < 1$$

The above theorem shows that the mean service rate is $\frac{\delta\theta}{\gamma+\delta}$ less than that of a service process having phase type representation $\left(\boldsymbol{\alpha}, S - \frac{\delta\theta}{\gamma+\delta} I\right)$ and the proportion of time it is available is $\frac{\gamma+\delta}{\gamma+\delta+\theta}$. The system remains interrupted for $\frac{\theta}{\gamma+\delta+\theta}$ of the time. The expected number of customers renege during this time is $\frac{(N-1)\beta\theta}{\gamma+\delta+\theta}$, which contributes to the outflow from the system.

Thus for large N the system $\tilde{\boldsymbol{\mathcal{X}}}$ is stable, irrespective of the parameters involved. However, as N tends to ∞, the truncated system $\tilde{\boldsymbol{\mathcal{X}}}$ becomes identical with the original system $\boldsymbol{\mathcal{X}}$. This leads to

Theorem 2.3.2. *The system $\boldsymbol{\mathcal{X}}$ is always stable, irrespective of the parameters involved*

2.3.3 Steady state vector

Let $\boldsymbol{x} = (x_0, x_1, x_2, ...)$ be the steady state probability vector of the Markov process $\boldsymbol{\mathcal{X}}$. We assume that

$$x_{N+i} = x_{N-1} R_N^{i+1}, \quad i = 0, 1, 2, 3, ... \tag{2.1}$$

where R_N is the minimal solution of the matrix quadratic equation

$$R^2 A_{2,N} + R A_{1,N} + A_0 = 0.$$

Let η_N be the spectral radius of R_N, $N > 0$.
The truncation level N is so chosen that the stability condition stated in Theorem 2.3.1 is satisfied and such that $|\eta_N - \eta_{N+1}| < \epsilon$, for some preassigned number ϵ.
Again $\boldsymbol{x}\tilde{\mathbf{Q}} = \mathbf{0}$ leads us to

$$x_{N-i} = x_{N-i-1} R_{N-i} \tag{2.2}$$

where

$$R_{N-i} = -A_0 (A_{1,N-i} + R_{N-i+1} A_{2,N-i+1})^{-1}, i = 1, 2, ..., N-2$$

and

$$x_1 = x_0 R_1 \tag{2.3}$$

where

$$R_1 = -A_{0,0}(A_{1,1} + R_2 A_{2,2})^{-1}$$

Finally $x_0 A_{1,0} + x_1 A_{2,1} = 0$. We find x_0 as the steady state distribution of the finite state Markov chain with the generator $A_{1,0} + R_1 A_{2,1}$. Then using equations (2.1), (2.2) and (2.3), we find x_i for $i \geq 1$. Now $\boldsymbol{x}$ can be calculated by dividing each x_i with the normalizing constant $\sum_{i=0}^{\infty} x_i \boldsymbol{e}$.

2.4 Analysis of the service process

2.4.1 Expected service time

The service process with interruption can be viewed as a Markov process $\Psi = \psi(t) = \{(S(t), J(t)) / t \geq 0\}$ where $S(t)$ is the status of the server which is 0 if the server is uninterrupted and 1 otherwise and $J(t)$ is the phase of the service process at time t. This process has $2m$ transient states given by $\{0,1\} \times \{1,2,3,...,m\}$ and one absorbing state Δ. The absorbing state Δ denotes the service completion. Let T be the time until absorption of the process Ψ. The infinitesimal generator

$\tilde{Q}$ of this process is given by

$$\tilde{Q} = \begin{bmatrix} B & B^0 \\ 0 & 0 \end{bmatrix}$$

where

$$B = \begin{bmatrix} S - \theta I & \theta I \\ \gamma I + \delta \boldsymbol{e\alpha} & -(\gamma + \delta) I \end{bmatrix} \quad \text{and} \quad B^0 = \begin{bmatrix} S^0 \\ \mathbf{0} \end{bmatrix}$$

The $2m \times 2m$ matrix B satisfies $B_{ii} < 0$ for $1 \le i \le 2m$ and $B_{ij} \ge 0$ for $i \ne j$. Also $B\boldsymbol{e} + B^0 = \mathbf{0}$ and the initial probability vector of the process is $(\boldsymbol{\xi}, 0)$, where $\boldsymbol{\xi} = (\boldsymbol{\alpha}, 0\boldsymbol{\alpha})$.

The probability distribution function $F(.)$ of T is given by

$$F(x) = 1 - \boldsymbol{\xi}.exp\,(Bx)\,\boldsymbol{e}, x \ge 0$$

Its density function $F'(x)$ in $(0, \infty)$ is given by

$$F'(x) = \boldsymbol{\xi}.exp\,(Bx)\,B^0.$$

The Laplace-Stieltjes transform $f(s)$ of $F(.)$ is

$$f(s) = \boldsymbol{\xi}(sI - B)^{-1}B^0, \text{ for } Re(s) \ge 0.$$

The non-central moments $\acute{\mu_i}$ of X are given by

$$\acute{\mu_i} = (-1)^i i!(\boldsymbol{\xi} B^{-i} \boldsymbol{e}), i = 1, 2, 3, ...$$

In particular we have the following lemma

Lemma 2.4.1. *The expected service time is given by*

$$E(T) = \xi(-B)^{-1}\boldsymbol{e} = -\left(1 + \frac{\theta}{\gamma+\delta}\right)\boldsymbol{\alpha}\left(S + \frac{\theta\delta}{\gamma+\delta}\left(\boldsymbol{e}\boldsymbol{\alpha} - I\right)\right)^{-1}\boldsymbol{e}$$

2.4.2 Expected number of interruptions during a single service

We have the following lemma, about the expected number of interruptions during a single service.

Lemma 2.4.2. *Expected number of interruptions during a single service $E(i)$ is given by,*

$$E\left(i\right) =$$
$$-\tfrac{\theta}{\gamma+\delta}\boldsymbol{\alpha}(S-\theta I)^{-1}(\gamma I + \delta\boldsymbol{e}\boldsymbol{\alpha})\left[I + \tfrac{\theta}{\gamma+\delta}(S-\theta I)^{-1}(\gamma I + \delta\boldsymbol{e}\alpha)\right]^{-1}\boldsymbol{e}$$

Proof. In any particular service, there are no more interruptions before an interrupted service is repaired. So the number of interruptions faced is independent of the time spend under interruptions. Therefore in evaluating the number of interruptions the information that whether the server is interrupted or not at time t is irrelevant.

Hence to get the distribution of the number of interruptions during a single service, we consider the Markov process χ =

$\{(N(t), J(t))/t \geq 0\}$ where $N(t)$ is the number of interruptions occurred during the service process till time t and $J(t)$ is the phase of the service process at time t. This process has the state space $\{\hat{0}\} \cup \{0, 1\} \times \{1, 2, 3, ..., m\}$ where $\hat{0}$ is the absorbing state denoting the service completion. The infinitesimal generator matrix of this process is given by

$$K = \begin{bmatrix} 0 & 0 & 0 & 0 & .... \\ S^0 & S - \theta I & C & 0 & ... \\ S^0 & 0 & S - \theta I & C & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \end{bmatrix}$$

where

$$C = \frac{\theta}{\gamma + \delta} (\gamma I + \delta \boldsymbol{e\alpha})$$

Let y_k be the probability that the number of interruptions during a single service is k. Then y_k is the probability that the absorption occurs from the level k for the process χ. Hence y_k are given by

$$y_0 = -\boldsymbol{\alpha} (S - \theta I)^{-1} S^0$$

and for $k = 1, 2, 3, ...,$

$$y_k = -\left(\frac{-\theta}{\gamma + \delta}\right)^k \boldsymbol{\alpha} \left[(S - \theta I)^{-1} (\gamma I + \delta \boldsymbol{e\alpha})\right]^k (S - \theta I)^{-1} S^0$$

Therefore, the expected number of interruptions during any particular service,

$$E(i) = \sum_{k=0}^{\infty} k y_k$$
$$= -\tfrac{\theta}{(\gamma+\delta)}\boldsymbol{\alpha}(S-\theta I)^{-1}(\gamma I + \delta \boldsymbol{e}\boldsymbol{\alpha})\left[I + \tfrac{\theta}{\gamma+\delta}(S-\theta I)^{-1}(\gamma I + \delta \boldsymbol{e}\boldsymbol{\alpha})\right]^{-1}\boldsymbol{e}.$$

□

Since the mean duration of an interruption is $\frac{1}{\gamma+\delta}$, obviously we have,

Corollary 2.4.3. The expected time spent under interruption during each service is

$$-\tfrac{\theta}{(\gamma+\delta)^2}\boldsymbol{\alpha}(S-\theta I)^{-1}(\gamma I + \delta \boldsymbol{e}\boldsymbol{\alpha})\left[I + \tfrac{\theta}{\gamma+\delta}(S-\theta I)^{-1}(\gamma I + \delta \boldsymbol{e}\boldsymbol{\alpha})\right]^{-1}\boldsymbol{e}.$$

2.5 Performance measures

2.5.1 Expected waiting time

For computing expected waiting time of a particular customer who joins as the r^{th} customer, $r > 0$, in the queue, we consider the Markov process

$$W(t) = \{(N(t); S(t); J(t))/t \geq 0\}$$

where $N(t)$ is the rank of the customer, $S(t) = 1$ or 0 according as the service is under interruption or not and $J(t)$ is the phase of the service process at time t. The rank $N(t)$ of the customer is assumed to be i if he is the i^{th} customer in the queue at time t. His rank may decrease to 1 as the customers ahead of him leave the system either after completing the service or due to reneging. Mean while it may happen that the tagged customer himself may renege from the system. Since the customers who arrive after the tagged customer cannot change his rank, level-changing transitions in $W(t)$ can only take place to one side of the diagonal. We arrange the state space of $W(t)$ as $\{r, r-1, ..., 2, 1\} \times \{0, 1\} \times \{1, 2, ..., m\} \cup \left\{\tilde{0}\right\}$, where $\tilde{0}$ is the absorbing state denoting that the tagged customer is either selected for service or he leaves the system without waiting for service. Thus the infinitesimal generator $\mathbf{W}$ of the process $W(t)$ takes the form

$$\mathbf{W} = \begin{bmatrix} \tilde{T} & \tilde{T}^0 \\ 0 & 0 \end{bmatrix}$$

where

$$\tilde{T} = \begin{bmatrix} \tilde{A}_{1,r} & \tilde{A}_{0,r} & & & & & \\ & \tilde{A}_{1,r-1} & \tilde{A}_{0,r-1} & & & & \\ & & \tilde{A}_{1,r-2} & \tilde{A}_{0,r-2} & & & \\ & & & ... & ... & ... & \\ & & & & \tilde{A}_{1,2} & \tilde{A}_{0,2} & \\ & & & & & & \tilde{A}_{1,1} \end{bmatrix} \quad \tilde{T}^0 = \begin{bmatrix} \tilde{B} \\ \tilde{B} \\ \tilde{B} \\ ... \\ \tilde{B} \\ \tilde{B}^0 \end{bmatrix}$$

with

$$\tilde{A}_{1,i} = \begin{bmatrix} S - \theta I & \theta I \\ \gamma I + \delta \boldsymbol{e\alpha} & -\left(\gamma + \delta + (i-1)\beta\right) I \end{bmatrix} i = 1, 2, 3, .., r.$$

$$\tilde{A}_{0,i} = \begin{bmatrix} S^0\boldsymbol{\alpha} & \mathbf{0} \\ \mathbf{0} & (i-1)\beta I \end{bmatrix} i = 2, 3, .., r. \quad \tilde{B} = \begin{bmatrix} 0 \\ \beta\boldsymbol{e} \end{bmatrix} \quad \tilde{B}^0 = \begin{bmatrix} S^0 \\ \beta\boldsymbol{e} \end{bmatrix}.$$

Now, the waiting time W of a customer, who joins the queue as the r^{th} customer is the time until absorption of the Markov chain $W(t)$. Thus the expected waiting times of this particular customer according to the phase of the service process at the time of his arrival are given by the column vector,

$$E_W^{(r)} = \left[-A_{1,r}^{-1} \left(I + \sum_{i=1}^{r-1} (-1)^i \prod_{j=1}^{i} A_{0,r+1-j} A_{1,r-j}^{-1} \right) \right] \boldsymbol{e}.$$

The second moments of waiting times of the tagged customer are given by the column vector $E_{W^2}^r$ which is the first block of the matrix $2(-\tilde{T})^{-2}\boldsymbol{e}$.

Hence, the expected waiting time of a general customer in the queue is,

$$W_L = \sum_{r=1}^{\infty} x(r) E_W^{(r)}.$$

The second moment of W is

$$W_L^{(2)} = \sum_{r=1}^{\infty} x(r) E_{W^2}^r.$$

2.5.2 Expected waiting time of customer who was served

The expected waiting time of a customer who waited till he gets the service, is obtained as in the previous section. In this case the $\mathbf{W}$ can be obtained by replacing $\tilde{B}$ and $\tilde{B}^0$ with zero vector and $\begin{bmatrix} S^0 \\ \mathbf{0} \end{bmatrix}$ respectively and adjusting the diagonals so as to form a generator. Here again we can calculate the first two moments $E_{W_s}^{(r)}$ and $E_{W^2}^{r}$ of waiting time as done earlier. Using these the expected waiting time W_L^s and the variance V_L^s are computed as

$$W_L^s = \sum_{r=1}^{\infty} x(r) E_{W_s}^{(r)}$$

$$V_L^s = W_L^{s(2)} - (W_L^s)^2$$

2.5.3 Other performance measures

The steady state probability vector $\boldsymbol{x} = (x_0, x_1, x_2, ...)$ calculated in section (2.3) can be partitioned by writing x_i as

$$x_i = (x_i', x_i''), i = 1, 2, ...$$

where

$$x_i' = (x_i'(1), x_i'(2), ..., x_i'(m))$$

and

$$x_i'' = (x_i''(1), x_i(2), ..., x_i''(m))$$

are steady state probability vectors corresponding to the states in which the system is not interrupted and is under interruption respectively with i customers in the system.

- Probability that there is no customer in the system,
$$P_C(0) = x_0.$$
- Probability that there are i customers in the system,
$$P_C(i) = x_i\boldsymbol{e}.$$
- Probability that there are $i > 1$ customers in the system when system is uninterrupted,
$$P'_C(i) = x'_i\boldsymbol{e}.$$
- Probability that there are $i > 1$ customers in the system when system is under interruption,
$$P''_C(i) = x''_i\boldsymbol{e}.$$
- Expected number of customers in the system,
$$E(C) = \sum_{i=0}^{\infty} iP_C(i).$$
- Expected number of customers in the queue
$$E(Q) = \sum_{i=0}^{\infty}(i-1)P_C(i).$$
- Expected number of customers in the system when the server is uninterrupted,
$$E'(C) = \sum_{i=1}^{\infty} iP'_C(i).$$

- Expected number of customers in the system when the system is under interruption,

$$E''(C) = \sum_{i=1}^{\infty} i P''_C(i).$$

- Variance of the number of customers in the system,

$$V(C) = \sum_{i=0}^{\infty} i^2 P_C(i) - \left(\sum_{i=0}^{\infty} i P_C(i) \right)^2.$$

- Variance of the number of customers in the system when the server is not interrupted

$$V'(C) = \sum_{i=0}^{\infty} i^2 P'_C(i) - \left(\sum_{i=0}^{\infty} i P'_C(i) \right)^2.$$

- Variance of the number of customers in the system when the system is under interruption,

$$V''(C) = \sum_{i=0}^{\infty} i^2 P''_C(i) - \left(\sum_{i=0}^{\infty} i P''_C(i) \right)^2.$$

- Probability that the system is under interruption,

$$P_S(I) = \sum_{i=0}^{\infty} x''_i \boldsymbol{e}.$$

- Effective interruption rate,

$$EI = \sum_{i=0}^{\infty} \theta x'_i \boldsymbol{e}.$$

- Effective reneging rate,

$$ER = \sum_{i=0}^{\infty} (i-1) \beta x''_i \boldsymbol{e}.$$

- Effective rate of repetition of service,

$$ERST = \sum_{i=0}^{\infty} \delta x_i'' \boldsymbol{e}.$$

- Effective service resumption rate,

$$ERSM = \sum_{i=0}^{\infty} \gamma x_i'' \boldsymbol{e}.$$

2.6 Numerical illustration

For numerical study, we have taken $\lambda_0 = \lambda_1 = \lambda$. Also we used the following values: Number of phases of the service process = 3

$$S = \begin{bmatrix} -17 & 4 & 7 \\ 1 & -17 & 6 \\ 4 & 5 & -17 \end{bmatrix}, \quad S^0 = \begin{bmatrix} 6 \\ 10 \\ 8 \end{bmatrix}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} 0.3 & 0.5 & 0.2 \end{bmatrix}, \quad -\alpha S^{-1} \boldsymbol{e} = 8.2186.$$

Table 2.1 indicates the variation in the system performance measures with arrival rate λ. The increase in the values of the performance measures like expected number of customers in the system, expected reneging rate and expected waiting time are on expected lines. Increase in effective interruption rate, effective resumption rate and effective repeat rate can be due to increase in number of services. Reneging may be the reason behind the increase in variance of number of customers. Same reasoning can be attributed to the smaller variance in the number of customers when the server is busy than when the service is interrupted. Also note that $E''(C) > E'(C)$, as seen from the table, which can be attributed to the increase in system breakdown probability.

From Table 2.2 we can make the following observations. As the interruption rate increases, the breakdown probability also increases and so is the expected resumption and repeat rates. Because further breakdowns can cause a delay, more customers accumulate in the queue and so an increased reneging rate can be expected. Large reneging from the system can cause an increased variance in the number of customers. The more interrupted the service, the more the mean and the variance of waiting time.

Increase in the rate δ of repetition of service decreases the expected number of customers, their waiting time and the corresponding variances. It makes the system busy for a longer duration so that $E'(C)$ increases and $E''(C)$decreases. Also as δ increases the relation $E''(C) > E'(C)$, gets reversed due to the decrease in $PS(I)$. As the system becomes busier, the effective interruption rate increases. The decrease in number of customers in the queue decreases the effective reneging rate. Effective repeat rate increases whereas effective resume rate decreases. All these can be observed from Table 2.3.

All performance measures except ERST and ERSM show the same behaviour with increase in resumption rate γ as with increase in repeat rate δ. While effective resumption rate increases, effective repeat rate decreases with increase in γ. This can be seen from Table 2.4.

Table 2.5. shows that, as interruption rate θ tends to zero, our results agree with corresponding results in the $M/PH/1$ queue without reneging.

Table 2.1: **Variation of the arrival rate λ.**

γ =2, δ =1.5, $\theta = 5$, $\beta = 1.0$, Expected service time = 0.29554963, Effective service rate = 3.38352656, Expected number of interruptions during a particular service = 0.608484685.

λ	1	3	5	7	10
$E(C)$	0.3779903	1.74073625	4.06366062	7.16970205	12.40080
$E''(C)$	0.2329257	1.08578670	2.47443676	4.25922298	7.207237
$E'(C)$	0.1450645	0.654949486	1.58922386	2.91047907	5.193565
$V(C)$	0.5036651	3.09771872	7.56597757	12.23357771	8.93583
$V''(C)$	0.3645396	2.74650860	8.60955811	19.38838964	6.59236
$V'(C)$	0.2067040	1.77348113	6.82128763	17.63794524	7.20600
$P_S(I)$	0.1614215	0.402823180	0.531474292	0.576763630	0.587703
ER	0.0715042	0.682963550	1.94296241	3.68245935	6.619534
$ERST$	0.2421323	0.604234755	0.797211468	0.865145445	0.881555
$ERSM$	0.3228431	0.805646360	1.06294858	1.15352726	1.175406
EI	0.5649754	1.40988111	1.86016011	2.01867270	2.056962
W_L	0.1035736	0.351978898	0.632030845	0.884171963	1.140170
V_L	0.0689032	0.208551750	0.334332943	0.449822873	0.644813
W_L^s	0.1493769	0.530649960	1.00692415	1.49856973	2.114290
V_L^s	0.1440628	0.442976147	0.674007535	0.773701608	0.801752

Table 2.2: **Variation of the interruption rate θ.**

$$\gamma = 2, \delta = 1.5, \beta = 1.0, \lambda = 3$$

θ	3	5	7	9
$E(C)$	1.33159864	1.74073625	2.06954503	2.33114314
$E''(C)$	0.67691880	1.08578670	1.43486571	1.72605133
$E'(C)$	0.65467989	0.65494948	0.63467919	0.60509198
$V(C)$	2.50664020	3.09771872	3.42105627	3.58161974
$V''(C)$	1.83268714	2.74650860	3.36910987	3.76818156
$V'(C)$	1.56028330	1.77348113	1.87330532	1.90227795
$P_S(I)$	0.27052372	0.40282318	0.50353527	0.58035075
ER	0.40639510	0.68296355	0.93133044	1.14570057
$ERST$	0.40578559	0.60423475	0.75530290	0.87052619
$ERSM$	0.54104745	0.80564636	1.00707054	1.16070151
EI	0.94683295	1.40988111	1.76237345	2.03122783
W_L	0.24848799	0.35197889	0.43808069	0.50836688
V_L	0.13973821	0.20855175	0.26554176	0.31274968
W_L^s	0.33858171	0.53064996	0.71729731	0.89381492
V_L^s	0.25643634	0.44297614	0.63161248	0.81720197
EI	0.36506450	0.60848468	0.85193550	1.0954150
ES	0.22599230	0.29554963	0.36511531	0.43468853
μ_S^I	4.42492962	3.38352656	2.7388608	2.3004977

Table 2.3: **Variation of the repeat rate δ.**

$\gamma = 2, \lambda = 3, \theta = 5, \beta = 1.0$

δ	0	1	2	3	4
$E(C)$	2.29792905	1.88868237	1.61907673	1.43267190	1.29816723
$E''(C)$	1.71670914	1.24744129	0.956216455	0.763889909	0.630068719
$E'(C)$	0.581219852	0.641241074	0.662860274	0.668782055	0.668098509
$V(C)$	3.70396972	3.29580617	2.91606927	2.60717702	2.36347437
$V''(C)$	3.88564682	3.08722901	2.45171666	1.98258507	1.63756526
$V'(C)$	1.81389380	1.80839837	1.73202848	1.64634371	1.56780493
$P_S(I)$	0.56100690	0.445889860	0.366695255	0.309857279	0.267497152
ER	1.15570211	0.801551402	0.589521229	0.454032630	0.362571567
$ERST$	0.00000000	0.445889860	0.733390510	0.929571807	1.06998861
$ERSM$	1.12201381	0.891779721	0.733390510	0.619714558	0.534994304
EI	1.12201381	1.33766949	1.46678102	1.54928625	1.60498285
W_L	0.504173100	0.391752839	0.319675118	0.270985782	0.236557841
V_L	0.337761730	0.240672797	0.183276370	0.146600634	0.121784367
W_L^s	0.960977197	0.625915349	0.46016511	0.364068389	0.302659810
V_L^s	1.08261943	0.565738022	0.359015137	0.254279107	0.193604380
$E(i)$	0.60836923	0.60846072	0.608502030	0.608525693	0.608540833
$E(T)$	0.42585837	0.32451230	0.273825884	0.243410215	0.223131612
μ_S^I	2.3481984	3.08154726	3.65195560	4.10829115	4.48165989

Table 2.4: **Variation of the Resume rate γ.**

$\delta = 2$, $\lambda = 3$, $\theta = 5$, $\beta = 1.0$

γ	0	1	2	3	4
$E(C)$	2.29874682	1.88897383	1.61907673	1.43250191	1.29789412
$E''(C)$	1.71726036	1.24761605	0.95621645	0.76380670	0.62994641
$E'(C)$	0.58148640	0.64135772	0.66286027	0.66869527	0.66794765
$V(C)$	3.70502734	3.29632473	2.91606927	2.60679483	2.36282945
$V''(C)$	3.88699865	3.08776593	2.45171666	1.98228478	1.63710999
$V'(C)$	1.81515551	1.80889547	1.73202848	1.64601791	1.56726182
$P_S(I)$	0.56108725	0.44591891	0.36669525	0.30984091	0.26747170
ER	1.15617311	0.80169719	0.58952122	0.45396575	0.36247471
$ERST$	1.12217450	0.89183783	0.73339051	0.61968183	0.53494340
$ERSM$	0.00000000	0.44591891	0.73339051	0.92952281	1.06988680
EI	1.12217450	1.33775675	1.46678102	1.54920459	1.60483015
W_L	0.50440818	0.39183455	0.31967511	0.27094003	0.23648545
V_L	0.33798757	0.24074511	0.18327637	0.14656364	0.12172775
W_L^s	0.96164107	0.62608504	0.46016511	0.36399617	0.30255371
V_L^s	1.08377659	0.56596720	0.35901513	0.25420251	0.19349990
$E(i)$	0.60861176	0.60854083	0.60850203	0.60847747	0.60846072
$E(T)$	0.42602813	0.32455506	0.27382588	0.24339097	0.22310221
μ_S^I	2.34726286	3.08114123	3.65195560	4.10861588	4.48225069

Table 2.5: **Variation of the Interruption rate θ.**

$\gamma = 2$, δ=2, $\lambda = 3$, $\beta = 1.0$.
For an $M/PH/1$ queue with above λ, S, S^0, the following results were obtained:
P(1 customer in the system)=0.238256142, P(No customers in the system)=0.391630888, Exp queue length = 0.945055604, Exp.No. of customers in the system = 1.55342472, Expected waiting time in the queue = 0.189011127.

θ	0.1	0.05	0.005	0.0005	0.00005	0
$P_C(1)$	0.23405	0.236195	0.2381349	0.2383294	0.2383489	0.23835110
$P_C(0)$	0.37935	0.385440	0.3910071	0.3915683	0.3916247	0.39163094
$E(Q)$	1.01428	0.979415	0.9478278	0.9446583	0.9443398	0.94430494
$E(C)$	1.63493	1.593975	1.5568206	1.553089	1.5527150	1.55267394
$E''(C)$	0.05802	0.028930	0.00288492	.0002884	.000028839	0.00000000
$E'(C)$	1.57690	1.565044	1.55393565	1.5528014	1.5526862	1.55267394
$V(C)$	4.22801	4.096664	3.97532463	3.9630260	3.9617877	3.96165204
$V''(C)$	0.26689	0.132962	.013247648	.00132425	.000132419	0.00000000
$V'(C)$	4.14414	4.054256 3	.97104311	3.9625973	3.9617447	3.96165204
$P_S(I)$	.017240	.00865576	.000868748	.00008690	.00000869	0.00000000
ER	0.40789	.02027464	.002016178	.00020150	.000020149	0.00000000
$ERST$	.025860	.01298364	.001303123	.00013035	.000013036	0.00000000
$ERSM$	.034480	.01731153	.001737497	.00017381	.000017381	0.00000000
EI	.060340	.03029518	.003040620	.00030417	.000030418	0.00000000
W_L	0.20285	0.1958831	0.18956555	0.1889316	0.18886804	0.18886101
V_L	.090564	.08609868	.082032784	.08162374	.081582680	.0 81578128
W_L^s	0.20747	0.1981575	0.18978993	0.1889540	0.18887026	0.18886101
V_L^s	.095844	.08869975	.082289248	.08164934	.081585235	.0 81578128
$E(i)$	.012167	.00608370	.0006083693	.00006083	.000006083	0.00000000
$E(T)$	0.12515	0.1234122	0.12184768	0.1216912	0.12167555	0.12167382
μ_S^I	7.99036	8.1029205	8.20696735	8.2175188	8.21857738	8.21869469

Chapter 3

A RETRIAL QUEUE WITH SERVER INTERRUPTION RESUMPTION AND RESTART OF SERVICE

3.1 Introduction

In many real life situations, where a Queue of customers may develop, often one can see customers who neither like to be queued up nor want to leave without getting service. Such customers may temporarily quit the system for accomplishing some other goal and may retry for service after some time. Though we can consider such retrying customers as new arrivals while modelling, it will be more realistic if we assume that retrials are happening from a separate pool of customers. This is be-

1. Presented at the 8^{th} International Workshop on Retrial Queues, July 27 - 29, 2010, Beijing, China.
2. Published in Operational Research Int. Journal (2012) 12 133-149.

cause the inter-occurrence time distribution between two fresh arrivals and that between two retrials need not be the same; also the inter-occurrence time between two retrials generally diminishes as the number of retrying customers increases. Queueing models, which allow retrial of customers, are now-a-days very popular and the literature on such models is vast. We refer to books by Falin and Templeton [36] and Artalejo and Gmez-Corral [42], for an extensive analysis of theory, applications and numerical procedure on retrial queues.

A common situation, which one faces while modelling a queueing system is the possibility of interruptions to an ongoing service that may be due to server breakdowns, server going on vacations, arrival of priority customers and even customer induced interruptions. Since interruptions causes a natural increase in the length of a service, we cannot expect the other customers to patiently wait in the queue. This is why we included the possibility of reneging of customers from the queue in the model described in the previous chapter. But in a system where service is subject to interruptions, there is a high possibility that the customers may choose to leave the system temporarily and retry for service after some time.

Two retrial systems in which the server is subject to interruptions, one with a finite number of homogeneous and the other with heterogeneous customers is presented by Almasi, Roszik and Sztrik [43],[44].Wang, Zhao and Zhang [45] analysed a retrial queue with a finite number of sources in which the server is subject to breakdowns and repairs. Kulkarni and Choi [38] study two models of single server retrial queue with server breakdowns. In the first model, the customer whose service is interrupted either leaves the system or joins the orbit; whereas in the second model the interrupted service is repeated after the repair is completed. Some other papers which study retrial queues with an unreliable server include Artalejo [37] Aissani and Artalejo [39], Artalejo and Gomez-Corral [40], Wang, Cao and Li [41], Artalejo and Gmez-Corral [42] and Chen, Zhu and Zhang [46].

In this chapter a retrial queueing model with service interruptions is analysed. More precisely, here we assume that a customer, encountering a busy or interrupted server, proceeds to an orbit of infinite capacity and from there retries for service. We also assume that unsuccessful retrials may lead to customer leaving the system without receiving the service.

As a motivating example for the present model, consider a person browsing internet for some purpose like booking a train ticket. The process of booking the ticket consists of a few steps, during which it may get interrupted for different reasons like too many customers login to the same website. After some random time, it may happen that the interrupted service get resumed from the same step at which interruption occurred or sometimes one sees the message like the web page has been expired, forcing the person to repeat the whole process from the very beginning. For considering such problems, the model described in this chapter is motivated by the fact that the more apt modelling could be made by an M/PH/1 retrial queueing system rather than by the classical M/PH/1 queueing system.

In this chapter, we extend the model described in chapter 2 to the retrial set up.

3.2 Model description

We consider a single server queueing model in which customers arrive according to a Poisson process with parameter λ. On arrival if a customer finds an idle server, he is immediately taken for service. If the server is busy when the customer arrives, the customer goes to an orbit of infinite capacity from where he makes repeated attempts for service according to a Poisson process with parameter β. After an unsuccessful retrial he rejoins the orbit with probability p or leaves the system without waiting for service with the probability $q = 1 - p$. The service time follows PH distribution with representation $(\boldsymbol{\alpha}, S)$ of order m. During service, interruptions to the service process may occur. The interruptions occur according to a Poisson process with parameter θ. At the epoch an interruption strikes, two clocks, namely resume clock and repeat clock are started exactly as in the model studied in the previous chapter to determine whether to resume or repeat the interrupted service. Their realization times follow distinct exponential distributions with parameters γ and δ respectively. If the resume clock realizes first, the interrupted service is resumed whereas the realization of the repeat clock before the resume clock prompts the system to repeat the current service.Figure (3.1) gives a picture of the model

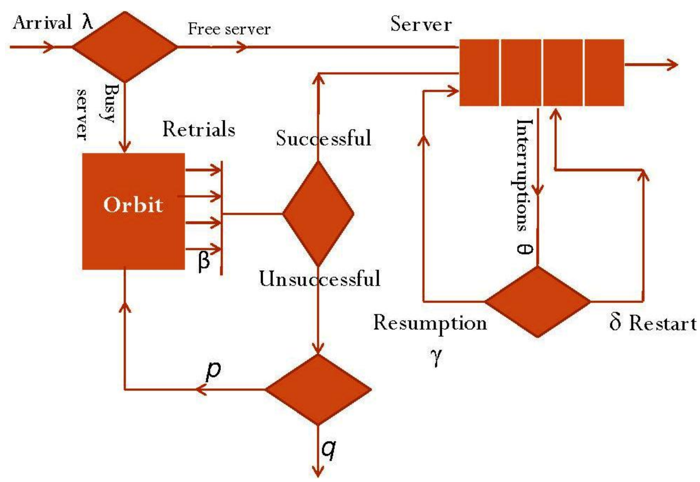


Figure 3.1: *An $M/Ph/1$ Retrial Queue with service interruptions.*

Let $N(t)$ be the number of customers in the orbit, $S(t)$ the status of the server which is 0, 1 or 2 according as the server is idle, uninterrupted or interrupted and $J(t)$ the phase of the service process at time t. Then the above model can be represented by the Markov process

$$\boldsymbol{X} = \{X(t)/t \geq 0\} = \{(N(t), S(t), J(t))\, /t \geq 0\}\, .$$

Its state space is given by

$$(\{0, 1, 2, 3, ...\} \times \{0\}) \cup (\{0, 1, 2, 3, ...\} \times \{1, 2\} \times \{1, 2, 3, ..., m\})\, .$$

Like the model described in Chapter 2, this process $\boldsymbol{X}$ is also a QBD. The number of customers in the orbit, $N(t)$ is the level in which the system is in at time t. In a single step, the process can go either one level up or one level down or remain in the same level. The forward transitions occurs when a new customer joins the orbit. This will happen only when the server is busy or is interrupted. The arrival of a customer to the orbit have no effect on the phase of the ongoing service, the rates at which the forward transitions occur are given by the entries in the matrix

$$A_0 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \lambda I \end{bmatrix}. \tag{3.1}$$

These transitions are depicted in Figure (3.2(a)).

A back transition occurs when a customer is selected for service after a successful retrial or he leaves the system on failure of a retrial. After a successful retrial, a new service is begun according to the initial probability vector $\boldsymbol{\alpha}$ whereas after a failed retrial the service process remains in the same phase where it was. Figure (3.2(b)) illustrates these transitions. The corresponding matrix is

$$A_{2,k} = \begin{bmatrix} \mathbf{0} & k\beta\boldsymbol{\alpha} & \mathbf{0} \\ \mathbf{0} & kq\beta I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & kq\beta I \end{bmatrix}, k = 1, 2, 3, ... \quad . \tag{3.2}$$

The transitions within the same level are the phase changes during a service, occurrence of interruptions, realizations of resume/repeat clocks and completion of a service resulting the server to be in idle state as pictorially represented in Figure

(3.2(c)). The matrix whose entries are these rates is given by

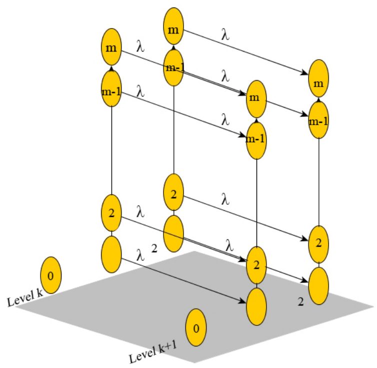
$$A_{1,k} = \begin{bmatrix} -(\lambda + k\beta) & \lambda\boldsymbol{\alpha} & \mathbf{0} \\ S^0 & S - (\theta + \lambda + kq\beta)I & \theta I \\ \mathbf{0} & \gamma I + \delta \boldsymbol{e}\boldsymbol{\alpha} & -(\lambda + \gamma + \delta + kq\beta)I \end{bmatrix}, k = 0, 1, 2, ... \tag{3.3}$$

Hence the infinitesimal generator matrix of the process $\boldsymbol{X}$ is

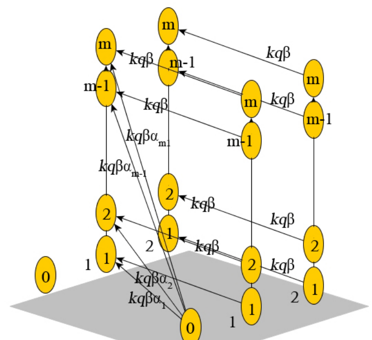
$$\boldsymbol{Q} = \begin{bmatrix} A_{1,0} & A_0 & & & & \\ A_{2,1} & A_{1,1} & A_0 & & & \\ & A_{2,2} & A_{1,2} & A_0 & & \\ & & A_{2,3} & A_{1,3} & A_0 & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \end{bmatrix}$$

3.3 Analysis of the service process

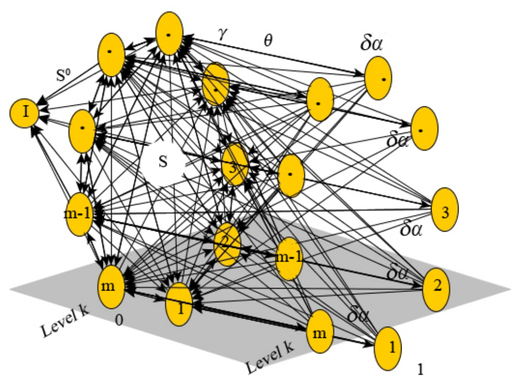
Since the service process for this model is the same as the one described in the previous chapter, the expressions for the expected service time, expected number of interruptions and the expected total duration of interruptions are as those obtained in Chapter 2 section 2.4.



(a) *From k to k+1*



(b) *From k to k-1*



(c) *From k to k*

Figure 3.2: *State transition diagrams*

3.4 Stability and steady state analysis

3.4.1 Stability analysis

From expressions (3.1) (3.2) and (3.3) we have

$$A = A_{2,k} + A_{1,k} + A_0 = \begin{bmatrix} -(\lambda + k\beta) & (\lambda + k\beta)\boldsymbol{\alpha} & \mathbf{0} \\ S^0 & S - \theta I & \theta I \\ \mathbf{0} & \gamma I + \delta \boldsymbol{e}\boldsymbol{\alpha} & -(\gamma + \delta)I \end{bmatrix}$$

Let $\boldsymbol{\pi} = (\boldsymbol{\pi_0}, \boldsymbol{\pi_1}, \boldsymbol{\pi_2})$, where $\boldsymbol{\pi_i} = (\pi_{i1}, \pi_{i2}, ..., \pi_{im}), i = 0, 1, 2$ be the invariant probability vector of A such that $\boldsymbol{\pi}\boldsymbol{e} = 1$.

Thus π_0, π_1 and π_2 satisfy the equations

$$(\lambda + k\beta)\boldsymbol{\pi_0} = \boldsymbol{\pi_1} S^0 \tag{3.1}$$

$$\theta \boldsymbol{\pi_1} = (\gamma + \delta)\boldsymbol{\pi_2} \tag{3.2}$$

and

$$\boldsymbol{\pi_1}\left[S^0\alpha + S - \frac{\delta\theta}{\gamma + \delta}(I - \boldsymbol{e}\boldsymbol{\alpha})\right] = \mathbf{0} \tag{3.3}$$

Therefore

$$\boldsymbol{\pi} A_0 \boldsymbol{e} = \frac{\lambda(\gamma + \delta + \theta)}{\gamma + \delta}\pi_1 \boldsymbol{e}$$

and

$$\boldsymbol{\pi} A_{2,k} \boldsymbol{e} = kq\beta + \frac{kp\beta}{\lambda + k\beta}\pi_1 S^0$$

Hence if $q \neq 0$, then

$$\frac{\boldsymbol{\pi} A_0 \boldsymbol{e}}{\boldsymbol{\pi} A_{2,k} \boldsymbol{e}} \to 0 \quad \text{as } k \to \infty.$$

Hence the system is always stable.

For $q = 0$ we proceed as follows.

From (3.3)

$$\boldsymbol{\pi_1} \left[S - S\boldsymbol{e}\alpha - \frac{\delta\theta}{\gamma+\delta}\left(I - \boldsymbol{e\alpha}\right) \right] = \mathbf{0}$$

$$\boldsymbol{\pi_1} \left[S - \frac{\delta\theta}{\gamma+\delta} I \right] \left[I - \boldsymbol{e\alpha} \right] = \mathbf{0}$$

Hence

$$\boldsymbol{\pi_1} \left[S - \frac{\delta\theta}{\gamma+\delta} I \right] = \varpi\alpha \tag{3.4}$$

for some scalar ϖ.

Therefore

$$\boldsymbol{\pi_1} = \varpi\alpha \left[S - \frac{\delta\theta}{\gamma+\delta} I \right]^{-1}$$

Hence

$$-\tilde{\mu}\boldsymbol{\pi_1}\boldsymbol{e} = \varpi. \tag{3.5}$$

where

$$\tilde{\mu} = \frac{1}{-\alpha \left[S - \frac{\delta\theta}{\gamma+\delta} I \right]^{-1} \boldsymbol{e}}$$

Now post multiplying (3.4) by $\boldsymbol{e}$, we get

$$-\boldsymbol{\pi_1} S^0 - \frac{\delta\theta}{\gamma+\delta}\boldsymbol{\pi_1}\boldsymbol{e} = \varpi \tag{3.6}$$

From (3.5) and (3.6), it follows that

$$\boldsymbol{\pi_1} S^0 = \left(\tilde{\mu} - \frac{\delta\theta}{\gamma+\delta}\right)\boldsymbol{\pi_1 e} \tag{3.7}$$

But in this case,
$\boldsymbol{\pi} A_{2,k}\boldsymbol{e} = \frac{k\beta}{\lambda+k\beta}\pi_1 S^0 \to \pi_1 S^0 \quad \text{as} \quad k \to \infty.$

Hence the condition for stability is

$$\frac{\lambda(\gamma+\delta+\theta)}{\gamma+\delta}\boldsymbol{\pi_1 e} < \pi_1 S^0.$$

Using (3.7), this condition becomes

$$\frac{\lambda(\gamma+\delta+\theta)}{\gamma+\delta} < \left(\tilde{\mu} - \frac{\delta\theta}{\gamma+\delta}\right)$$

Combining all these together we have the following theorem.

Theorem 3.4.1. *The Markov chain* $\boldsymbol{X}$ *is always stable irrespective of the system parameters if* $q > 0$*. For* $q = 0$ *the chain is stable if and only if*

$$\frac{\lambda(\gamma+\delta+\theta)}{\gamma+\delta} < \frac{1}{-\boldsymbol{\alpha}\left(S - \frac{\theta\delta}{\gamma+\delta}I\right)^{-1}\boldsymbol{e}} - \frac{\theta\delta}{\gamma+\delta}$$

From the stability condition it follows that when $q = 0$ the mean service time is $\frac{\theta\delta}{\gamma+\delta}$ less than that of a phase type service process with representation $\left(\boldsymbol{\alpha}, S - \frac{\theta\delta}{\gamma+\delta}I\right)$. Since the proportion of time an uninterrupted service is available is $\frac{\gamma+\delta}{\gamma+\delta+\theta}$, the effective service rate becomes

$$\frac{\gamma+\delta}{(\gamma+\delta+\theta)}\left[\frac{1}{-\boldsymbol{\alpha}\left(S-\frac{\theta\delta}{\gamma+\delta}I\right)^{-1}\boldsymbol{e}}-\frac{\theta\delta}{\gamma+\delta}\right]$$

3.4.2 Steady state analysis

The steady state vector $\boldsymbol{\xi}=(\boldsymbol{\xi_0},\boldsymbol{\xi_1},\boldsymbol{\xi_2},...)$ is given by

$$\boldsymbol{\xi}Q=0 \tag{3.8}$$

Writing $\boldsymbol{\xi_i}=(x_i,y_i,z_i)$ where x_i is the probability that the server is idle and $y_i=(y_i(1),y_i(2),...,y_i(m))$ and $z_i=(z_i(1),z_i(2),...,z_i(m))$ are probability vectors denoting the probabilities that the server is functioning or is under interruption respectively, with i customers in the orbit, we have from (3.8)

$$x_i=\frac{1}{\lambda+i\beta}y_iS^0, i=0,1,2,3,... \tag{3.9}$$

Eliminating x_is from equation (3.8) using equation (3.9) and rearranging we see that the vector $\boldsymbol{\zeta}=(\boldsymbol{\zeta_0},\boldsymbol{\zeta_1},\boldsymbol{\zeta_2},...)$ where $\boldsymbol{\zeta_i}=(y_i,z_i)$satisfies the equation

$$\boldsymbol{\zeta}B=0 \tag{3.10}$$

where

$$B = \begin{bmatrix} B_{1,0} & B_0 & & & & \\ B_{2,1} & B_{1,1} & B_0 & & & \\ & B_{2,2} & B_{1,2} & B_0 & & \\ & & B_{2,3} & B_{1,3} & B_0 & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \end{bmatrix}$$

in which

$$B_{1,k} = \begin{bmatrix} \frac{\lambda}{\lambda+k\beta} S^0 \boldsymbol{\alpha} + S - (\theta + \lambda + qk\beta) I & \theta I \\ \gamma I + \delta \boldsymbol{e} \boldsymbol{\alpha} & -(\gamma + \delta + \lambda + qk\beta) I \end{bmatrix}$$

$$B_{2,k} = \begin{bmatrix} \frac{k\beta}{\lambda+k\beta} S^0 \boldsymbol{\alpha} + qk\beta I & \mathbf{0} \\ \mathbf{0} & qk\beta I \end{bmatrix}, k = 0, 1, 2, 3, ... \text{ and}$$

$B_0 = \lambda I_{2m}$.

For finding the vector $\boldsymbol{\zeta}$, we consider two cases.

Case 1. $\boldsymbol{q > 0}$.

Here we use Neuts-Rao Truncation (Neuts and Rao [51]) to determine $\boldsymbol{\zeta}$. In this method, we assume that $B_{1,i} = B_{1,N}$ and $B_{2,i} = B_{2,N}$ for all $i \geq N$. With this approximation, the vector $\boldsymbol{\zeta}$ can be then given by

$$\boldsymbol{\zeta_{N+i}} = \boldsymbol{\zeta_{N-1}} R_N^{i+1}, i = 0, 1, 2, 3, ...$$

where R_N is the minimal non-negative solution of the matrix quadratic equation $R^2 B_{2,N} + R B_{1,N} + B_0 = 0$ and $\boldsymbol{\zeta_{N-i}} = \boldsymbol{\zeta_{N-i-1}} R_{N-i}$ where $R_{N-i} = -B_0(B_{1,N-i} + R_{N-i+1} B_{2,N-i+1})^{-1}, i =$

$1, 2, 3, ..N - 1$.(Neuts [49]).

Finally, we have $\boldsymbol{\zeta_0} B_{10} + \boldsymbol{\zeta_1} B_{21} = 0$. Substituting $\boldsymbol{\zeta_1} = \boldsymbol{\zeta_0} R_1$, we get, $\boldsymbol{\zeta_0} \left[B_{10} + R_1 B_{21}\right] = 0$. We note that the matrix $B_{10} + R_1 B_{21}$ is an infinitesimal generator and hence we take $\boldsymbol{\zeta_0}$ as the steady state vector of this matrix. Using $\boldsymbol{\zeta_0}$, we can find all the $\boldsymbol{\zeta_i}$ s. Once $\boldsymbol{\zeta_i}$ s are obtained, x_i s are given by

$$x_i = \frac{1}{\lambda + i\beta} \boldsymbol{\zeta_i} \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix} S^0, i = 0, 1, 2, 3, ... \quad .$$

It remains that the vectors x_is and $\boldsymbol{\zeta_i}$s are to be normalized for finally arriving at the steady state vector $\boldsymbol{\xi}$ and the normalizing constant is given by $\sum x_i + \sum \boldsymbol{\zeta_i} \boldsymbol{e}$. Since we do not have the explicit expressions for the R_i matrices, such an expression for the normalizing constant seems impossible and therefore it is obtained numerically only.

Case 2. $\boldsymbol{q = 0}$

The case $q = 0$ deserves a special note. In this case, from equation (3.10), we have the equations,

$$\boldsymbol{\zeta_0} B_{1,0} + \boldsymbol{\zeta_1} B_{2,1} = 0$$

$$\boldsymbol{\zeta_{i-1}} B_0 + \boldsymbol{\zeta_i} B_{1,i} + \boldsymbol{\zeta_{i+1}} B_{2,i+1} = 0, i = 1, 2, 3, ...$$

A further manipulation of the above equations by post multiplying with the column vector of 1s gives:

$$\lambda \boldsymbol{\zeta_{i-1}} \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e} \end{bmatrix} = \zeta_i \begin{bmatrix} \frac{i\beta}{\lambda+i\beta} S^0 \\ \mathbf{0} \end{bmatrix}, i = 1, 2, 3, ...$$

These relations help us to write,

$$\boldsymbol{\zeta_i} = \boldsymbol{\zeta_{i-1}} R_i, \ i = 1, 2, 3, ...$$

where

$$R_i = -\lambda \begin{bmatrix} S - (\theta+\lambda)I + \lambda \boldsymbol{e}\boldsymbol{\alpha} + \frac{\lambda}{\lambda+i\beta} S^0 \boldsymbol{\alpha} & \theta I \\ \gamma I + (\delta+\lambda)\boldsymbol{e}\boldsymbol{\alpha} & -(\gamma+\delta+\lambda)I \end{bmatrix}^{-1},$$

$$i = 1, 2, 3, ... \quad .$$

As in case 1, the substitution $\boldsymbol{\zeta_1} = \boldsymbol{\zeta_0} R_1$ in the equation $\zeta_0 B_{1,0} + \zeta_1 B_{2,1} = 0$ implies that $\boldsymbol{\zeta_0}$ satisfies the equation

$$\boldsymbol{\zeta_0} \begin{bmatrix} S - (\theta+\lambda)I + \lambda \boldsymbol{e}\boldsymbol{\alpha} + S^0 \boldsymbol{\alpha} & \theta I \\ \gamma I + (\delta+\lambda)\boldsymbol{e}\boldsymbol{\alpha} & -(\gamma+\delta+\lambda)I \end{bmatrix} = 0$$

Here also we note that the square matrix in the above equation is an infinitesimal generator and the same procedure as in case (1) can be applied for finding the steady state vector.

3.5 Performance Measures

3.5.1 Expected waiting time of a customer in the orbit

To find the expected waiting time $E(W_L)$ with $q \geq 0$, we consider a system similar to the system we are considering, but with finite orbital capacity N. If $E_N(W_L)$ denotes the waiting time of a customer in the orbit in this finite orbit system, then $E(W_L)$ is obtained as the limit of $E_N(W_L)$ as $N \to \infty$.

The waiting time of a customer, who joins the system with finite orbit capacity N, is the time until absorption of the Markov process $W(t) = (N(t), S(t), J(t))$ where $N(t)$ is the number of customers in the orbit, $S(t)$ is the status of the server which is 0, 1 or 2 according as the server is idle, uninterrupted and busy or interrupted respectively during a service and $J(t)$ is the phase of the service process at time t. This process has the state space $\{1, 2, 3, ..., N\} \times \{0, 1, 2\} \times \{1, 2, 3, ..., m\} \cup \{\Delta\}$ where Δ is the absorbing state which denotes the event that the tagged customer is selected for service or leaves the system after an unsuccessful retrial. The infinitesimal generator of this process is

$$\boldsymbol{W} = \begin{bmatrix} \tilde{\boldsymbol{T}}_{\boldsymbol{N}} & \tilde{\boldsymbol{T}}^{\boldsymbol{0}}_{\boldsymbol{N}} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

where

$$\tilde{\boldsymbol{T}}_{\boldsymbol{N}} = \begin{bmatrix} A_{11} & A_0 & & & & \\ A_{21} & A_{12} & A_0 & & & \\ & A_{22} & A_{13} & A_0 & & \\ & & . & . & . & \\ & & & . & . & . \\ & & & & A_{2N-1} & \tilde{A}_{1N} \end{bmatrix}, \tilde{\boldsymbol{T}}^{\boldsymbol{0}}_{\boldsymbol{N}} = \begin{bmatrix} \tau \\ \tau \\ . \\ . \\ . \\ \tau \end{bmatrix}$$

with

$$\tau = \begin{bmatrix} \beta \\ q\beta\boldsymbol{e} \\ q\beta\boldsymbol{e} \end{bmatrix} \text{ and } \tilde{A}_{1N} = \begin{bmatrix} -(\lambda + N\beta) & \lambda\alpha & 0 \\ S^0 & S - (\theta + Nq\beta)I & \theta I \\ 0 & \gamma I + \delta\boldsymbol{e\alpha} & -(\gamma + \delta + Nq\beta)I \end{bmatrix}$$

The expected waiting time of a customer in the system with finite orbital capacity N is then given by

$$E_N(W_L) = -\xi^{(N)}\tilde{\boldsymbol{T}}_{\boldsymbol{N}}^{-1}\boldsymbol{e}$$

where

$$\xi^{(N)} = (\tilde{\xi}_0, \tilde{\xi}_1, ..., \tilde{\xi}_{N-1})$$

and

$$\tilde{\xi}_i = (0, y_i, z_i).$$

The variance in waiting time $V_N(W_L)$ is given by

$$V_N(W_L) = E_N(W_L^2) - [E_N(W_L)]^2$$

where

$$E_N(W_L^2) = 2\xi^{(N)} \left(-\tilde{\boldsymbol{T}}_{\boldsymbol{N}}{}^{-1}\right)^2 \boldsymbol{e}$$

The probability that a customer leaves the system without completing the service given that the server is busy on his arrival,

$$P_L^N = -\xi^{(N)}\tilde{T}_N^{-1}\hat{T}^0$$

where

$$\hat{T}^0 = \begin{bmatrix} \hat{\tau} \\ \hat{\tau} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\tau} \end{bmatrix} \quad \text{with} \quad \hat{\tau} = \begin{bmatrix} 0 \\ q\beta\boldsymbol{e} \\ q\beta\boldsymbol{e} \end{bmatrix}.$$

The variance in waiting time of a customer, $V(W_L)$ and the probability P_L that a customer may leave the system without waiting for service given the system was busy at the time of his arrival, P_L are given by

$$V(W_L) = \lim_{N\to\infty} V_N(W_L).$$

$$P_L = \lim_{N\to\infty} P_L^N$$

3.5.2 Other performance measures

The following performance measures are found numerically.

- Probability that the server is idle,

$$P_{idle} = \sum_{i=1}^{\infty} x_i$$

- Probability that the server is interrupted,

$$P_{int} = \sum_{i=1}^{\infty} z_i \boldsymbol{e}$$

- Probability that the server is uninterrupted,

$$P_{unin} = \sum_{i=1}^{\infty} y_i \boldsymbol{e}$$

- Probability that the server is idle with customers in the orbit,

$$P_{idle}(C) = P_{idle} - x_0$$

- Expected number of customers in the orbit,

$$E_O(C) = \sum_{i=1}^{\infty} i\xi_i \boldsymbol{e}$$

- Expected number of customers in the orbit when the server is under interruption,

$$E_O^I(C) = \sum_{i=1}^{\infty} iz_i \boldsymbol{e}$$

- Expected number of customers in the orbit when the server is uninterrupted,

$$E_O^U(C) = \sum_{i=1}^{\infty} iy_i \boldsymbol{e}$$

- Expected number of customers in the orbit when the server is idle,
$$E_O^{Id}(C) = \sum_{i=1}^{\infty} i x_i$$
- Effective rate at which the customers leave the system after unsuccessful retrials,
$$ELR = \beta(1 - P_{idle})$$

3.6 Numerical illustration

The results of numerical experiments conducted to study the effect of various parameters on the system performance are analysed below. For numerical calculations we used the following values:

Number of phases of the service process = 3

$$S = \begin{bmatrix} -9 & 4 & 2 \\ 6 & -15 & 4 \\ 3 & 7 & -14 \end{bmatrix}, \boldsymbol{\alpha} = \begin{bmatrix} 0.3 & 0.5 & 0.2 \end{bmatrix}.$$

The effect of the retrial rate β on the system performance is computed in Table 3.1. As the retrial rate β increases, there is a decrease in the expected number of customers in the orbit. At the same time there is an increase in the server idle probability. Together, they point towards the loss of retrying customers from the system. At this point, it is to be noted that the arrival rate (= 6.0) is greater than the service rate (= 3.058) and so a portion of the arrivals must be lost (after retrials) for the stability of the system. Also notice the increase in the loss rate ELR in the table. As β tends to infinity, the expected number of customers in the orbit is tending to 0, while the idle probability is converging to some value (0.3376 in the table) and the loss rate ELR is converging to $\lambda - \lambda P_{idle}$. That is as β tends to infinity, the system under study becomes a queueing system where only those who find an idle server get service. In the table, the fraction of time the server is found interrupted P_{int} is slowly decreasing, with increasing β, before finally converging to some value. This is due to a decrease in the number of services rendered as a result of the increased loss of customers from the orbit. Hence the decrease in P_{int} is in fact indicating the loss of customers to the system due to increased reneging rate. Hence a decrease in the interruption rate is not always a positive sign for the system. The same reasoning can be made for the decreasing behaviour in the probabilities P_{unin} and $P_{idle}(C)$.

As the loss probability q increases, the expected number of customers in the orbit decreases as in the case of increase in retrial rate β; but it is to be noted that in the case of increase in β, the expected number is converging to 0, whereas even if the loss probability q is equal to 1, the expected number of customers in the orbit is converging to that in the corresponding $M/PH/1$ retrial queue where an orbital customer leaves the system after one retrial attempt. Because the number of customers decreases, number of services also decreases and the idle probability of the server increases. For similar reason, the server interruption probability P_{int} as well as the probability of finding an uninterrupted server P_{unint} decreases. The increase in the server idle probability with customers in the system $P_{idle}(C)$ is brought by the increased loss rate while the server is busy or uninterrupted. Also notice the increase in the difference $P_{idle} - P_{idle}(C)$ with increase in q, which also indicates the effect of loss rate of customers on the server idle probability. Table 3.2 gives the corresponding numerical values.

Table 3.3 contains the values of various performance measures for different values of θ. As the interruption rate θ increases, first note the increase in the expected service time $E(T)$; which leads to an increase in the expected number of customers $E_O(C)$ and $E_O^I(C)$. Quite naturally, as the number of interruptions increases, the probability P_{int} increases and the probability P_{unin} decreases. Since the fraction of time the system may be found uninterrupted decreases, so does the expected number $E_O^U(C)$. The decrease in the sever idle probability is brought by the increase in the expected service time.

As the realization rate γ of the resume clock increases, first notice the decrease in the expected service time $E(T)$, which follows from the analytical expression

$$E(T) = -\left(1 + \frac{\theta}{\gamma+\delta}\right) \boldsymbol{\alpha} \left(S + \frac{\theta\delta}{\gamma+\delta}(\boldsymbol{e\alpha} - I)\right)^{-1} \boldsymbol{e}. \quad (3.1)$$

In the above expression, as γ increases, the quantity $1 + \frac{\theta}{\gamma+\delta}$ is clearly decreasing. The remaining part is also decreasing as that being the absorption time in an m-dimensional Markov chain with generator matrix $\left(S + \frac{\theta\delta}{\gamma+\delta}(\boldsymbol{e\alpha} - I)\right)$, where the transition rates in the generator matrix is decreasing with increase in γ. Thus the decrease in the expected service time with increase in γ is brought by the decrease in the interruption duration. From the above expression, we can also infer that the expected service time is converging to $-\boldsymbol{\alpha}S^{-1}\boldsymbol{e}$ as $\gamma \to \infty$. The decrease in the interruption duration leads to a small decrease in the loss rate. The increase in the service rate results in a decrease in the expected number of customers in the orbit $E_O(C)$. Here note that the expected number of customers in the orbit when the server is either uninterrupted or is idle is increasing due to the increase in the corresponding probabilities P_{unin} and P_{idle}. The importance of an increased realization rate γ is that it diminishes the severity of the interruption. These findings are supported by the numerical results arranged in Table (3.4).

The effect of the realization rate of the repeat clock δ on the system performance is similar to that of the resume clock γ, this is so as δ increases, interruption duration decreases and hence the expected service time $E(T)$ decreases. Here again consider the expression for the expected service time $E(T)$ given by equation (3.1). As in the case of γ, here also an increase in the rate δ causes a decrease to $E(T)$; but it is to be noted that here the decrease rate also depends on the nature of the initial probability vector $\boldsymbol{\alpha}$ and on the matrix S. All the performance measures studied except the expected number of interruptions $E(i)$ show the same behaviour with increase in δ as in the case of increase in γ. This is because in the case of δ, the behaviour of the expected number of interruptions $E(i)$ depends on the nature of the phase type distribution. For example here one can see that $E(i)$ is decreasing slightly but if we consider an Erlang distribution for the service time, $E(i)$ shows an increase. Numerical data supporting these findings are tabulated in Table (3.5).

Table (3.6) considers the case where the probability that the customer may quit the system after an unsuccessful retrial, $q = 0$. Note that in this case we have obtained a different stability condition. For the numerical study for Table (3.6) we have taken the arrival rate $\lambda = 2.5$ and the rest of the parameters are the same as in other tables. In table 6, we compare the case $q = 0$ with the case $q = 0.3$. Note the high values for the expected number of customers when $q = 0$. The service rate being the same in two cases the low values for the expected number of customers when $q = 0.3$ reflects the loss of customers from the orbit. The data emphasize the fact that the increase in idle probability with increase in retrial rate when customer loss is allowed is due to the loss of customers and does not indicate an improvement in the efficiency of the system. Another important observation is about the waiting time of an orbital customer, which is very high in the case $q = 0$ compared to $q = 0.3$. Figures (3.3(a)), (3.3(b)), (3.3(c)) and (3.3(d)) illustrate these points.

Table 3.1: **Effect of the retrial rate** β

arrival rate $\lambda = 6$, Interruption rate $\theta = 2.0$, Realization rate of resume clock $\gamma = 3.0$, Realization rate of repeat clock $\delta = 4.0$, Probability of customer leaving the system on unsuccessful retrial $q = 0.3$

β	1.0	3.0	5.0	7.0	∞
$E_O(C)$	13.0287	4.3474	2.6112	1.8669	0.0000
$E_O^I(C)$	2.5476	0.8812	0.5445	0.3980	0.0000
$E_O^U(C)$	8.7281	2.9180	1.7594	1.2639	0.0000
$E_O^{id}(C)$	1.7530	0.5482	0.3073	0.2050	0.0000
P_{idle}	0.1440	0.1560	0.1680	0.1792	0.3376
P_{int}	0.1902	0.1876	0.1849	0.1824	0.1472
P_{unint}	0.6657	0.6564	0.6471	0.6384	0.5152
$P_{idle}(C)$	0.1440	0.1457	0.1278	0.1070	0.0000
ELR	3.3827	3.4193	3.4559	3.4901	3.9745
W_L	2.1715	0.7246	0.4352	0.3112	0.0000
P_L	0.5638	0.5699	0.5760	0.5817	0.6624

Table 3.2: **Effect of q on various performance measures**

arrival rate $\lambda = 6$, Interruption rate $\theta = 2.0$, Realization rate of resume clock $\gamma = 3.0$, Realization rate of repeat clock $\delta = 4.0$, Retrial rate $\beta = 2$

q	0.1	0.3	0.5	0.7	1.0
$E_O(C)$	17.0574	6.5177	4.1938	3.1308	2.2884
$E_O^I(C)$	3.5845	1.2989	0.8120	0.5945	0.4259
$E_O^U(C)$	12.288	4.3690	2.7068	1.9725	1.4079
$E_O^{id}(C)$	1.1849	0.8498	0.6750	0.5638	0.4546
P_{idle}	0.0760	0.1499	0.1885	0.2131	0.2372
P_{int}	0.2053	0.1889	0.1803	0.1749	0.1695
P_{unint}	0.7187	0.6612	0.6311	0.6120	0.5933
$P_{idle}(C)$	0.0759	0.1480	0.1793	0.1924	0.1961
ELR	3.1745	3.4008	3.5188	3.5938	3.6676
W_L	2.8429	1.0863	0.6990	0.5218	0.3814
P_L	0.5291	0.5668	0.5865	0.5990	0.6113

Table 3.3: **Effect of the Interruption rate θ**

arrival rate $\lambda = 6$, Realization rate of resume clock $\gamma = 3.0$, Realization rate of repeat clock $\delta = 4.0$, Retrial rate $\beta = 2$, Probability of customer leaving the system on unsuccessful retrial $q = 0.3$

θ	1.0	2.0	4.0	6.0	8.0
$E_O(C)$	6.1668	6.5177	7.0610	7.4598	7.7641
$E_O^I(C)$	0.6833	1.2989	2.3441	3.1842	3.8671
$E_O^U(C)$	4.5705	4.3690	3.9741	3.6182	3.3082
$E_O^{id}(C)$	0.9130	0.8498	0.7428	0.6575	0.5889
P_{idle}	0.1703	0.1499	0.1206	0.1006	0.0862
P_{int}	0.1037	0.1889	0.3198	0.4151	0.4874
P_{unint}	0.7260	0.6612	0.5596	0.4843	0.4264
$P_{idle}(C)$	0.1675	0.1480	0.1196	0.1000	0.0858
$E(T)$	0.2914	0.3270	0.3981	0.4688	0.5392
ELR	3.1522	3.4008	3.7909	4.0814	4.3052
$E(i)$	0.2549	0.5087	1.0133	1.5145	2.0129
W_L	1.0278	1.0863	1.1768	1.2433	1.2940
P_L	0.5254	0.5668	0.6318	0.6802	0.7175

Table 3.4: **Effect of the Realization rate of resume clock** γ

arrival rate $\lambda = 6$, Interruption rate $\theta = 2.0$, Probability of customer leaving the system on unsuccessful retrial $q = 0.3$, Realization rate of repeat clock $\delta = 4.0$, retrial rate $\beta = 2$

γ	3	5	7	9	∞
$E_O(C)$	6.5177	6.3695	6.2688	6.1958	5.7422
$E_O^I(C)$	1.2989	1.026	0.8468	0.7204	0
$E_O^U(C)$	4.369	4.4657	4.5257	4.5661	4.7591
$E_O^{id}(C)$	0.8498	0.8778	0.8963	0.9093	0.9831
P_{idle}	0.1499	0.1582	0.164	0.1682	0.1964
P_{int}	0.1889	0.1531	0.1286	0.1109	0
P_{unint}	0.6612	0.6887	0.7074	0.7209	0.8036
$P_{idle}(C)$	0.148	0.156	0.1616	0.1657	0.1923
$E(T)$	0.327	0.3112	0.3011	0.2941	0.2556
ELR	3.4008	3.295	3.2235	3.1719	2.8555
$E(i)$	0.5087	0.5092	0.5096	0.5098	0.5111
W_L	1.0863	1.0616	1.0448	1.0326	0.957
P_L	0.5668	0.5492	0.5372	0.5287	0.4759

Table 3.5: **Effect of the Realization rate of repeat clock δ**

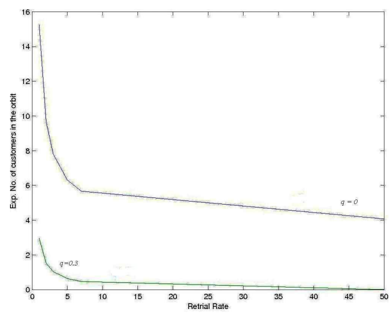
arrival rate $\lambda = 6$, Interruption rate $\theta = 2.0$, Probability of customer leaving the system on unsuccessful retrial $q = 0.3$, Realization rate of resume clock $\gamma = 4.0$, retrial rate $\beta = 2$

δ	4	6	8	10	∞
$E_O(C)$	6.5177	6.3644	6.2602	6.1848	5.7162
$E_O^I(C)$	1.2989	1.0249	0.8453	0.7187	0
$E_O^U(C)$	4.369	4.4606	4.517	4.5548	4.7291
$E_O^{id}(C)$	0.8498	0.8788	0.8979	0.9113	0.9871
P_{idle}	0.1499	0.1585	0.1645	0.1689	0.1981
P_{int}	0.1889	0.153	0.1285	0.1108	0
P_{unint}	0.6612	0.6885	0.707	0.7203	0.8019
$P_{idle}(C)$	0.148	0.1563	0.1621	0.1663	0.1939
$E(T)$	0.327	0.3107	0.3003	0.2931	0.2536
ELR	3.4008	3.2913	3.2174	3.1641	2.8375
$E(i)$	0.5087	0.5084	0.5081	0.508	0.5071
W_L	1.0863	1.0607	1.0434	1.0308	0.9527
P_L	0.5668	0.5486	0.5362	0.5273	0.4729

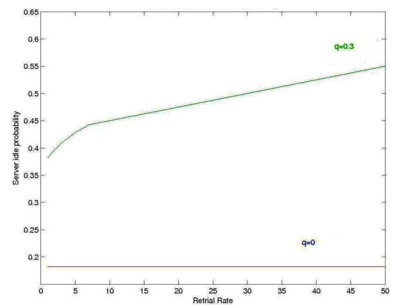
Table 3.6: **Effect of the Realization rate of repeat clock δ**

Arrival rate $\lambda = 6$, Interruption rate $\theta = 2.0$, Probability of customer leaving the system on unsuccessful retrial $q = 0.3$, Realization rate of resume clock $\gamma = 4.0$, Retrial rate $\beta = 2$

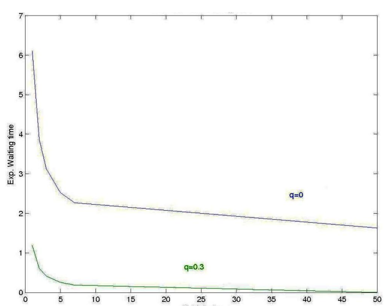
β	1		5		7		∞	
	$q=0$	$q=0.3$	$q=0$	$q=0.3$	$q=0$	$q=0.3$	$q=0$	$q=0.3$
$E_O(C)$	15.2746	2.9661	7.8039	1.0316	5.6694	0.4644	4.0685	0
$E_O^I(C)$	2.9906	0.4731	1.6333	0.1888	1.2454	0.0962	0.9546	0
$E_O^U(C)$	10.24	1.5548	5.4893	0.5815	4.1319	0.2829	3.1139	0
$E_O^{id}(C)$	2.044	0.9382	0.6813	0.2613	0.292	0.0853	0	0
P_{idle}	0.1824	0.3814	0.1824	0.4091	0.1824	0.4427	0.1824	0.5502
P_{int}	0.1817	0.1375	0.1817	0.1313	0.1817	0.1238	0.1817	0.1
P_{unint}	0.6359	0.4812	0.6359	0.4596	0.6359	0.4334	0.6359	0.3499
$P_{idle}(C)$	0.1792	0.3245	0.1351	0.1662	0.0801	0.0681	0	0
W_L	6.1098	1.1864	3.1215	0.4126	2.2677	0.1858	1.6274	0
ELR	0	0.6084	0	0.6932	0	0.7961	0	1.1246



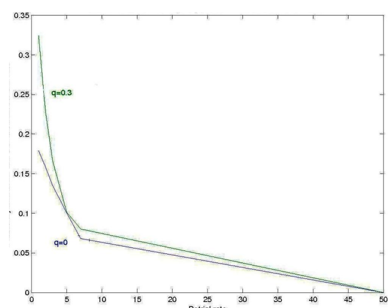
(a) *Retrial Rate V/S Expected number of customers in the orbit*



(b) *Retrial Rate V/S Server Idle Probability*



(c) *Retrial Rate V/S Waiting time in the orbit*



(d) *Retrial Rate V/S Waiting time in the orbit*

Figure 3.3: *Comparison of variations in Expected number of customers in the orbit, system idle probability, probability that system is idle with customers in the orbit and expected waiting time in the orbit for the cases $q = 0$ and $q = 0.3$ as retrial rate changes*

Chapter 4

AN $M/E_m/1$ QUEUE WITH PROTECTED AND UNPROTECTED PHASES FROM INTERRUPTION

4.1 Introduction

As revealed by the models discussed in the previous chapters, interruption is a common phenomenon in many fields like industry, health care, education and many more; this makes scientists to include the possibility of different kinds of interruptions while modelling a practical situation. In many of these situations the effect of interruptions may prove to be fatal. An interruption not only slows down the service rate and makes the system less efficient but also damages the ongoing service and makes heavy

---

Presented at the 5^{th} International conference on Queueing Theory and Network Applications (QTNA 2010), July 24 - 26, 2010, Beijing, China.

loss to the system. This raised the need of protecting the service from interruption using some more resources so that the damage caused by the interruption may be minimized. Providing battery back up to our desk top computers is a good example for this situation from our daily life.

Klimenok, Kim and Kuznetsov [47] analyses a model with two types of customers ordinary customers and negative customers where a negative customer can eliminate an ordinary customer in service and a partial protection is provided for ordinary customers. A similar system with BMAP input process is considered in Klimenok and Dudin [48]

In this chapter, we consider a single server queueing system where the server is subject to interruption. The interrupted server is taken for repair immediately; where both the interruption and repair time follow independent exponential distributions. The service time follows an Erlang distribution with m phases. As soon as an interruption occurs, a random clock is started, which runs along with the repair clock. The purpose of this random clock is to decide whether to repeat or resume the interrupted service after the completion of repair. If the random clock realizes before the completion of repair, the interrupted service needs to be repeated from the first phase, otherwise it is resumed in the phase at which the interruption occurred. The highlight of this model is that, here we divide the m phases of the Erlang service process in to two groups, namely protected and unprotected, in the sense that no interruption can affect the server while the service process is in the protected phases. More precisely we assume that the final n phases are protected and the service process will not be interrupted while being in these phases. The first $m-n$ phases are unprotected and service in these phases are vulnerable to interruptions.

Interruption being a common phenomenon, the present model is motivated by the fact that certain stages of services (for example in health care) are so important that they cannot afford an interruption.

4.2 Model Description

Consider a single server queuing system in which the arrival process is Poisson with rate λ. The service time distribution is Erlang of order m having density

$$f(t) = \frac{m\mu(m\mu t)^{m-1}e^{-m\mu t}}{(m-1)!}, t \geq 0$$

with mean $\frac{1}{\mu}$.The service time may be assumed to be consisting of m independent exponential stages, each with mean $\frac{1}{m\mu}$. A customer taken for service has to complete these m stages of service. Until he has completed all the stages of service, the other customers in the system have to wait in a queue. The first $m-n$ stages of the service process are subject to interruptions and the final n phases are protected in the sense that the service will not be interrupted while being in these phases. The interruptions occur according to a Poisson process with mean $\frac{1}{\theta}$. The interrupted server is taken for repair immediately with repair duration following an exponential distribution with mean $\frac{1}{\delta}$. A random clock is started at the beginning of each repair to decide whether to restart or resume the service after repair. If the random clock realizes before a repair, the service needs to be restarted, otherwise the service is resumed in the phase from where interruption occurred. The realization time of the random clock also follows an exponential distribution, with mean $\frac{1}{\gamma}$. This model is pictured in Figure 4.1. This queueing model can be defined by the

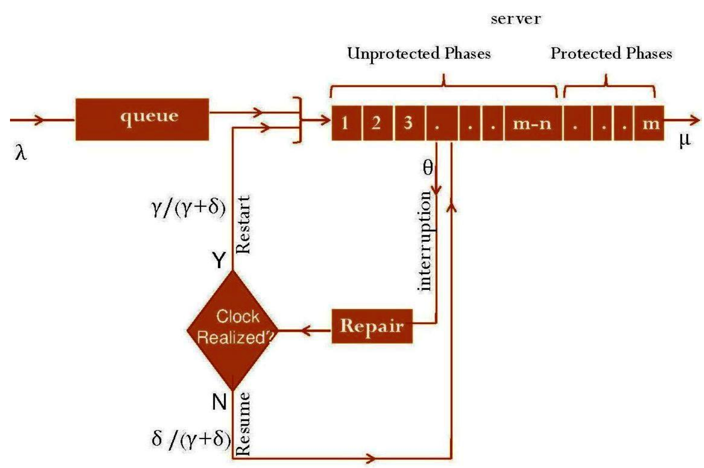
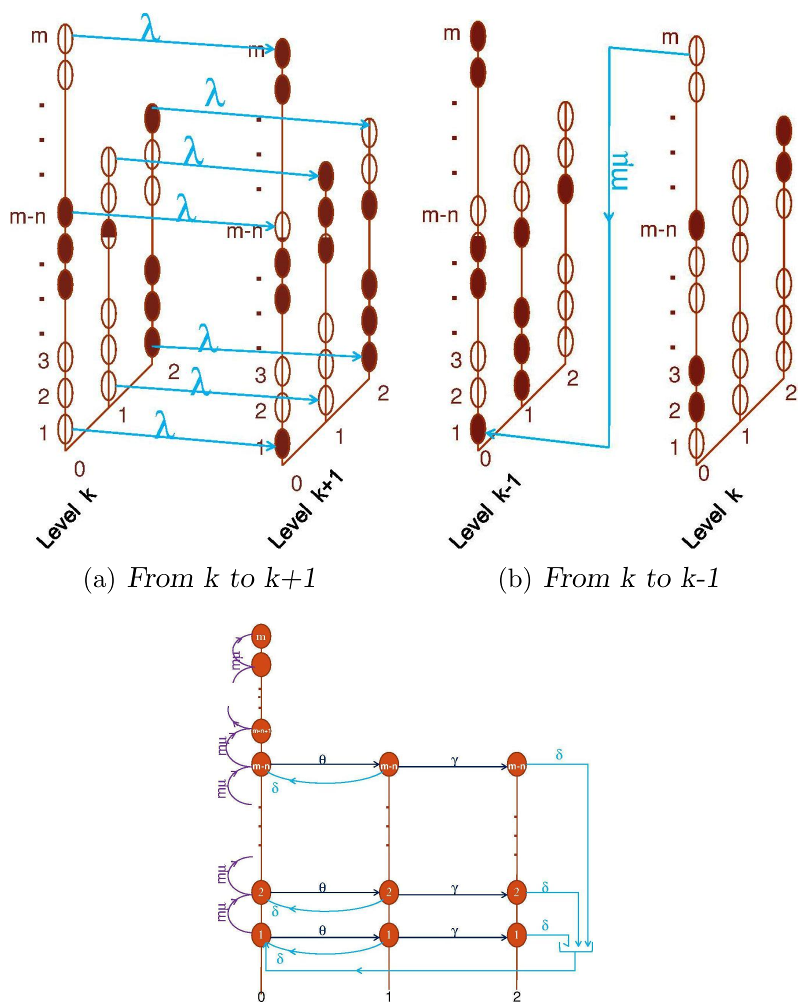


Figure 4.1: *An $M/E_m/1$ Queue with Protected and Unprotected Phases from Interruption*

Markov process $\boldsymbol{X} = \{X(t)/t \geq 0\} = \{(N(t), S(t), J(t)), t \geq 0\}$ where $N(t)$ is the number of customers in the system, $S(t)$ is the status of the server which is 0, 1 or 2 according as the service is uninterrupted, interrupted with a running clock or interrupted with a realized clock and $J(t)$ is the phase of the service process at time t. The state space is given by $\{0, 1, 2, 3, ...\} \times \{0\} \times$

$\{1, 2, ..., m\} \cup \{0, 1, 2, 3, ...\} \times \{1, 2\} \times \{1, 2, ..., m-n\}$. the state transitions are described in the Figure 4.2.



(a) *From k to k+1*

(b) *From k to k-1*

(c) *From k to k*

Figure 4.2: *State transition diagrams*

The infinitesimal generator matrix given by

$$Q = \begin{bmatrix} A_{10} & A_{00} & & & & \\ A_{21} & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \cdot & \cdot & \cdot \\ & & & & & \\ & & & & & \end{bmatrix}$$

where

$$A_{10} = \begin{bmatrix} -\lambda \end{bmatrix} \quad A_{00} = \begin{bmatrix} \lambda\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 & .... & 0 \end{bmatrix}$$

$$A_0 = \begin{bmatrix} \lambda I_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \lambda I \end{bmatrix} \quad A_{21} = \begin{bmatrix} S^0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad A_2 = \begin{bmatrix} S^0\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$A_1 = \begin{bmatrix} S - \begin{bmatrix} \theta I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} - \lambda I_m & \begin{bmatrix} \theta I \\ \mathbf{0} \end{bmatrix} & \mathbf{0} \\ \begin{bmatrix} \delta I & \mathbf{0} \end{bmatrix} & -(\gamma+\delta+\lambda)I & \gamma I \\ \delta\boldsymbol{e}\boldsymbol{\alpha} & \mathbf{0} & -(\delta+\lambda)I \end{bmatrix}$$

$$S^0 = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ m\mu \end{bmatrix} \quad S = m\mu \begin{bmatrix} -1 & 1 & 0 & & & \\ & -1 & 1 & 0 & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \\ & & & & & & -1 \end{bmatrix}$$

and I is the identity matrix of order $m-n$.

4.3 Steady state Analysis

4.3.1 Stability analysis

$$\text{Let } A = A_0+A_1+A_2 = \begin{bmatrix} S - \begin{bmatrix} \theta I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + S^0\boldsymbol{\alpha} & \begin{bmatrix} \theta I \\ \mathbf{0} \end{bmatrix} & \mathbf{0} \\ \begin{bmatrix} \delta I & \mathbf{0} \end{bmatrix} & -(\gamma+\delta)I & \gamma I \\ \delta \boldsymbol{e}\boldsymbol{\alpha} & \mathbf{0} & -\delta I \end{bmatrix}$$

and $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$ be the vector such that $\boldsymbol{\pi}A = 0$ and $\boldsymbol{\pi}\boldsymbol{e} = 1$. Then

$$\pi_0 \left(S - \begin{bmatrix} \theta I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + S^0\boldsymbol{\alpha} \right) + \pi_1 \begin{bmatrix} \delta I & \mathbf{0} \end{bmatrix} + \pi_2 \delta \boldsymbol{e}\boldsymbol{\alpha} = 0 \tag{4.1}$$

$$\pi_0 \begin{bmatrix} \theta I \\ \mathbf{0} \end{bmatrix} - (\gamma+\delta)\,\pi_1 = 0 \tag{4.2}$$

$$\gamma\pi_1 - \delta\pi_2 = 0 \tag{4.3}$$

Solving we get

$$\pi_1 = \frac{\theta}{\gamma+\delta}\pi_{0u}$$

and

$$\pi_2 = \frac{\gamma\theta}{\delta(\gamma+\delta)}\pi_{0u}$$

where $\pi_0 = (\pi_{0u}, \pi_{0p})$. Therefore equation (4.1) reduces to

$$\pi_0 \left(S - \begin{bmatrix} \theta I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + S^0 \boldsymbol{\alpha} \right) + \frac{\delta\theta}{\gamma+\delta} \pi_{0u} \begin{bmatrix} \delta I & \mathbf{0} \end{bmatrix} + \frac{\gamma\theta}{\gamma+\delta} \pi_{0u} \boldsymbol{e}\boldsymbol{\alpha} = 0 \tag{4.4}$$

This leads to $\pi_{0p} = \pi_{0,m-n}\,(1, 1, 1, ..., 1)$ and

$$\pi_{0,m-n-i} = \left[1 + \frac{\gamma\theta}{m\mu(\gamma+\delta)} \right]^i \pi_{0,m-n},\ i = 1, 2, ..., m-n-1.$$

Thus

$$\pi_{0,m-n} = \tfrac{\gamma\theta}{m\mu(\gamma+\delta)} \left[\tfrac{\delta+\theta}{\delta} \left(\left(1 + \tfrac{\gamma\theta}{m\mu(\gamma+\delta)} \right)^{m-n} - 1 \right) + \tfrac{n\gamma\theta}{m\mu(\gamma+\delta)} \right]^{-1}$$

It follows that

$$\begin{aligned} \boldsymbol{\pi} A_2 \boldsymbol{e} &= m\mu\pi_{0,m-n} \\ &= \tfrac{\gamma\theta}{(\gamma+\delta)} \left[\tfrac{\delta+\theta}{\delta} \left(\left(1 + \tfrac{\gamma\theta}{m\mu(\gamma+\delta)} \right)^{m-n} - 1 \right) + \tfrac{n\gamma\theta}{m\mu(\gamma+\delta)} \right]^{-1} \end{aligned}$$

and $\boldsymbol{\pi} A_0 \boldsymbol{e} = \lambda$ Hence the condition for stability is

$$\lambda \left[\frac{\delta+\theta}{\delta} \left(\left(1 + \frac{\gamma\theta}{m\mu(\gamma+\delta)} \right)^{m-n} - 1 \right) + \frac{n\gamma\theta}{m\mu(\gamma+\delta)} \right] < \frac{\gamma\theta}{(\gamma+\delta)}$$

That is $\rho < 1$ where,

$$\rho = \left[\frac{\delta+\theta}{\delta} \frac{\gamma+\delta}{\gamma\theta} \left(\left(1 + \frac{\gamma\theta}{m\mu(\gamma+\delta)} \right)^{m-n} - 1 \right) \right] + \frac{n}{m\mu}$$

If all the phases are protected, then $n = m$ and the stability condition becomes $\lambda < \mu$. The reason is trivial since if all the phases are protected then there will be no interruption throughout the service and the mean service time is $\frac{1}{\mu}$.

4.3.2 Stationary probabilities

The stationary probability vector $\boldsymbol{x}$ is given by $\boldsymbol{x}Q = 0$, together with the normalizing condition $\boldsymbol{xe} = 1$. On partitioning the steady state vector as $\boldsymbol{x} = (x_0, x_1, x_2, ...)$, the equation $\boldsymbol{x}Q = 0$ reduces to the following equations.

$$\begin{aligned} x_0 A_{10} + x_1 A_{21} &= 0 \\ x_0 A_{00} + x_1 A_1 + x_2 A_2 &= 0 \\ x_i A_0 + x_{i+1} A_1 + x_{i+2} A_2 &= 0, \;\; i = 1, 2, 3, ... \end{aligned}$$

Post multiplying these equations by $\boldsymbol{e}$, we get

$$\lambda x_i \boldsymbol{e} = x_{i+1} \begin{bmatrix} S^0 \\ 0 \\ 0 \end{bmatrix}, i = 0, 1, 2, ...$$

Also note that $A_2 = \begin{bmatrix} S^0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} & 0 & 0 \end{bmatrix}$.

Solving the above system of equations togethor with the last two expressions we get

$$x_0 = 1-\rho,\ x_i = (1-\rho)\beta R^i \text{ for } i = 1,2,3,...$$

where $\beta = (\alpha, 0, 0)$ and $R = -\lambda[A_1 + \lambda \boldsymbol{e}\beta]^{-1}$.

4.4 Analysis of the service process

4.4.1 Expected service time

The service process with the introduction of interruption, becomes a Markov Process Ψ with $3m-2n$ transient states given by

$$\{1,2,...,m\} \times \{0\} \cup \{1,2,...,m-n\} \times \{1,2\}$$

and one absorbing state $\widetilde{0}$. The absorbing state denotes the service completion. The process Ψ can be represented by $\psi(t) = (i,j)$ where i is the phase of service, $j=0$ if the service is uninterrupted, $j=1$ if the service is interrupted with repeat clock running and $j=2$ if the service is interrupted with the realized repeat clock. Let τ be the time until absorption of the process Ψ with the initial probability vector $\boldsymbol{\beta} = (\boldsymbol{\alpha}, \mathbf{0}, \mathbf{0})$. The infinitesimal generator of this process is given by $\widetilde{Q} = \begin{bmatrix} T & T^0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where

$$T = \begin{bmatrix} S - \begin{bmatrix} \theta I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} & \begin{bmatrix} \theta I \\ \mathbf{0} \end{bmatrix} & \mathbf{0} \\ \begin{bmatrix} \delta I & \mathbf{0} \end{bmatrix} & -(\gamma+\delta)I & \gamma I \\ \delta \boldsymbol{e}\boldsymbol{\alpha} & \mathbf{0} & -\delta I \end{bmatrix} \quad T^0 = \begin{bmatrix} S^0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

The expected service time is the time until absorption of the above process which is given by,

$$\begin{aligned}
E(\tau) &= \boldsymbol{\beta}(-T)^{-1}\boldsymbol{e} \\
&= \frac{\frac{\delta+\theta}{\delta}\left[\left(1+\frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{m-n}-1\right]+\frac{n\gamma\theta}{m\mu(\gamma+\delta)}}{\frac{\gamma\theta}{\gamma+\delta}} \\
&= \frac{(\delta+\theta)}{\delta}\frac{(\gamma+\delta)}{\gamma\theta}\left[\left(1+\frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{m-n}-1\right]+\frac{n}{m\mu}.
\end{aligned}$$

Let y_i be the expected time spent in the state i during a single service, $i = 1, 2, ..., 3m - 2n$.
Then $\boldsymbol{\beta}(-T)^{-1} = (y_1, y_2, ..., y_{3m-2n})$. So we have,

$$y_i = \begin{cases}
\frac{1}{m\mu}\left(1+\frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{m-n-i}, & i = 1, 2, ..., m-n \\
\frac{1}{m\mu}, & i = m-n+1, ..., m \\
\frac{\theta}{(\gamma+\delta)m\mu}\left(1+\frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{2m-n-i} & i = m+1, ..., 2m-n \\
\frac{\gamma\theta}{(\gamma+\delta)m\mu\delta}\left(1+\frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{3m-2n-i}, & i = 2m-n+1, ..., 3m-2n
\end{cases}$$

Therefore, during a single service

- The expected time spent in each of the n protected phases is
$$\frac{1}{m\mu}$$
.

- The expected time spent in the i^{th} phase is

$$y_i + y_{m+i} + y_{2m-n+i} = \frac{(\delta+\theta)}{m\mu\delta}\left(1+\frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{m-n-i}, \quad i = 1, 2, ..., m-n$$

- Expected time spent in phase i under interruption is

$$\frac{\theta}{m\mu\delta}\left(1+\frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{m-n-i}, i = 1, 2, 3, ..., m-n$$

- Expected duration of interruption is

$$\frac{\gamma+\delta}{\gamma\delta}\left[\left(1+\frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{m-n-i} - 1\right].$$

It can be seen that expected time spent in the i^{th} phase increases with increase in γ. Hence expected service time also increases as γ increases. This justifies our intuition that a higher clock realization rate results in more repeat of service on completion of interruption and results in a longer service time.

A higher value of repair rate δ causes the repair to get finished before the realization of the clock and hence an interrupted service can be resumed. So the service time gets lowered. This follows from the expression for expected time spent in a phase under interruption.

As expected, increase in the interruption rate θ results in an increase in the expected interruption duration and so the service time also increases.

4.4.2 Expected number of interruptions during a single service

To compute the expected number of interruptions faced by a customer during his service, we consider the Markov process $\boldsymbol{\chi} = \{(N(t), J(t))\,/t \geq 0\}$ where $N(t)$ is the number of interruptions that occurred up to time t and $J(t)$ is the phase of the service process at time t. $\boldsymbol{\chi}$ has the state space $\{0, 1, 2, 3, ...\} \times \{1, 2, 3, ..., mn\} \cup \{\Delta\}$ where Δ is the absorbing state which denotes that the service process reached the first protected phase and there will be no more interruption. The infinitesimal generator $\tilde{Q}$ of this process is

$$Q = \begin{bmatrix} D_2 & D_1 & D_0 & . & & \\ D_2 & 0 & D_1 & D_0 & . & \\ D_2 & 0 & 0 & D_1 & D_0 & . \\ . & . & . & . & . & \\ . & . & . & . & . & . \\ & & & & & \end{bmatrix}$$

where

$$D_1 = \begin{bmatrix} -m\mu-\theta & m\mu & 0 & & & \\ & -m\mu-\theta & m\mu & 0 & & \\ & & -m\mu-\theta & m\mu & . & \\ & & & . & . \; . & \\ & & & & . \; . & . \\ & & & & & -m\mu-\theta \end{bmatrix}_{m-n}$$

$$D_0 = \begin{bmatrix} \theta & 0 & 0 & & & \\ \frac{\gamma\theta}{\gamma+\delta} & \frac{\delta\theta}{\gamma+\delta} & & & & \\ \frac{\gamma\theta}{\gamma+\delta} & 0 & \frac{\delta\theta}{\gamma+\delta} & & . & \\ \frac{\gamma\theta}{\gamma+\delta} & 0 & 0 & \frac{\delta\theta}{\gamma+\delta} & . \; . & \\ . & . & . & . & . \; . \; . & \\ & & & & & \end{bmatrix}_{m-n} \quad \text{and } D_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ m\mu \end{bmatrix}_{(m-n)\times 1}$$

If z_j is the probability that there are exactly j interruptions during a single service, then

$$z_j = \zeta(-D_1^{-1}D_0)^j(-D_1^{-1}D_2), j = 0, 1, 2, ...$$

where

$$\zeta = (1, 0.0..., 0).$$

Expected number of interruptions during a single service,

$$
\begin{aligned}
E(I) &= \sum_{j=0}^{\infty} j z_j \\
&= \zeta(-D_1^{-1}D_0)(I - D_1^{-1}D_0)^{-1}\boldsymbol{e} \\
&= \left(1 + \frac{\delta}{\gamma}\right)\left[\left(1 + \frac{\gamma\theta}{m\mu(\gamma+\delta)}\right)^{m-n} - 1\right]
\end{aligned}
$$

4.5 Performance Analysis

4.5.1 Expected waiting time

For computing expected waiting time of a particular customer who joins as the r^{th} customer in the queue, we consider the Markov process $\boldsymbol{W} = \{W(t)/t \geq 0\} = \{(N(t), S(t), J(t))\,/t \geq 0\}$ where $N(t)$ is the rank of the customer, $S(t) = 0$, 1 or 2 according as the service is not under interruption or under interruption with clock running or finished and $J(t)$ is the phase of the service process at time t. The rank $N(t)$ of the customer is assumed to be r if he joins as the r^{th} customer in the queue. His rank decrease by 1 as one customer ahead of him leaves the system after completing the service. Since the customers who arrive after the tagged customer cannot change his rank, level-changing transitions in $W(t)$ can only take place to one side of the diagonal.

We arrange the state space of $W(t)$ as $\{r, r-1, ..., 2, 1\} \times \{0\} \times \{1, 2, ..., m\} \cup \{r, r-1, ..., 2, 1\} \times \{1, 2\} \times \{1, 2, ..., m-n\} \cup \{\Delta\}$ where Δ is an absorbing state in the sense that the tagged customer is selected for service.

Thus the infinitesimal generator $\tilde{\boldsymbol{Q}}$ of the process $\boldsymbol{W}$ takes the form

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} B & B^0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \text{ where } B = \begin{bmatrix} T & A_2 & & & & \\ & T & A_2 & & & \\ & & T & A_2 & & \\ & & & \cdot & \cdot & \\ & & & & \cdot & \cdot \\ & & & & & T \end{bmatrix} \text{ and } B^0 = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ S^0 \end{bmatrix}$$

Now, the waiting time W of a customer, who joins the queue as the r^{th} customer is the time until absorption in the Markov chain $W(t)$. Thus the expected waiting time of this particular customer is given by the column vector

$$\begin{aligned} E_W^r &= (-B)^{-1}\boldsymbol{e} \\ &= (-T)^{-1}\left[I - (A_2T^{-1}) + (A_2T^{-1})^2 - ... + (-1)^{r-1}(A_2T^{-1})^{r-1}\right]\boldsymbol{e} \\ &= (-T)^{-1}\boldsymbol{e} + (r-1)E(\tau)\boldsymbol{e} \end{aligned}$$

Hence, the expected waiting time of a general customer,

$$\begin{aligned} W_L &= \sum_{r=1}^{\infty} x(r) E_W^r \\ &= \sum_{r=1}^{\infty} \left[x(r)(-T)^{-1}\boldsymbol{e} + (r-1)E(\tau)x(r)\boldsymbol{e} \right] \\ &= \rho\sigma(-T)^{-1}\boldsymbol{e} + E(\tau)E_Q(C) \end{aligned}$$

where $\sigma = \frac{\beta}{E(\tau)}(-T)^{-1}$ and $E_Q(C) = \sum_{r=1}^{\infty}(r-1)x(r)\boldsymbol{e}$ is the expected number of customers in the queue.

4.5.2 Other performance measures

The steady state probability vector $\boldsymbol{x} = (x_0, x_1, ...)$ can be partitioned by writing x_i as $x_i = \left(x_i^{(0)}, x_i^{(1)}, x_i^{(2)}\right)$ where

$$\begin{aligned} x_i^{(0)} &= \left(x_i^{(0)}(1), x_i^{(0)}(2), ..., x_i^{(0)}(m)\right) \\ x_i^{(1)} &= \left(x_i^{(1)}(1), x_i^{(1)}(2), ..., x_i^{(1)}(m-n)\right) \\ x_i^{(2)} &= \left(x_i^{(2)}(1), x_i^{(2)}(2), ..., x_i^{(2)}(m-n)\right), \text{ for } i = 1, 2, 3, ... \end{aligned}$$

are steady state probability vectors corresponding to the states in which the system is working and under interruption with clock running and finished respectively

- Probability that there is no customer in the system

$$P_C(0) = x_0$$

- . Probability that there are i customers in the system,

$$P_C(i) = x_i \boldsymbol{e}$$

- . Probability that there are i customers in the system when system is under Interruption,

$$P_C^I(i) = \left(x_i^{(1)} + x_i^{(2)} \right) \boldsymbol{e}$$

- . Probability that there are i customers in the system when system is uninterrupted,

$$P_C^B(i) = x_i^{(0)} \boldsymbol{e}$$

- . Expected number of customers in the system,

$$E_S(C) = \sum_{i=0}^{\infty} i P_C(i)$$

- Expected number of customers in the queue,

$$E_Q(C) = \sum_{i=0}^{\infty} (i-1) P_C(i)$$

- Expected number of customers in the system when the system is under interruption,

$$E^I(C) = \sum_{i=0}^{\infty} i P_C^I(i)$$

- Expected number of customers in the system when the server is uninterrupted,

$$E^B(C) = \sum_{i=0}^{\infty} i P_C^B(i)$$

- Variance of the number of customers in the system,

$$V_S(C) = \sum_{i=1}^{\infty} i^2 P_C(i) - \left(\sum_{i=1}^{\infty} i P_C(i) \right)^2$$

- Variance of the number of customers in the system when the system is under interruption,

$$V^I(C) = \sum_{i=1}^{\infty} i^2 P_C^I(i) - \left(\sum_{i=1}^{\infty} i P_C^I(i) \right)^2$$

- Variance of the number of customers in the system when the server is busy

$$V^B(C) = \sum_{i=1}^{\infty} i^2 P_C^B(i) - \left(\sum_{i=1}^{\infty} i P_C^B(i) \right)^2$$

- Probability that the system is under interruption,

$$P_S(I) = \sum_{i=1}^{\infty} \left(x_i^1 + x_i^1 \right) \boldsymbol{e} = \frac{\lambda}{\delta} E(I)$$

- Effective interruption rate,

$$EI = \theta \sum_{i=1}^{\infty} x_i^{(0)} \boldsymbol{e}$$

- Effective service resumption rate,

$$ERSM = \delta \sum_{i=1}^{\infty} x_i^{(1)} \boldsymbol{e}$$

- Effective rate of repetition of service,

$$ERPT = \delta \sum_{i=1}^{\infty} x_i^{(1)} \boldsymbol{e}$$

- Probability that the system is in protected phase,

$$P_{pr} = \sum_{i=1}^{\infty} \sum_{j=m-n+1}^{m} x_i^{(0)}(j) = \frac{n\lambda}{m\mu}$$

- Probability that the system is in protected phase given system is busy,

$$P_B(Pr) = \frac{n\lambda}{\rho m\mu}$$

4.6 Numerical Illustration

We use the following notations in this section

$E_S(C)$	:	Expected number of customers in the system
$V_S(C)$	:	Variance of number of customers in the system
$E_Q(C)$	:	Expected number of customers in the queue
$E^B(C)$	:	Expected number of customers in the system when the system is uninterrupted
$E^I(C)$	:	Expected number of customers in the system when the system is interrupted
$P_S(I)$	:	Probability that the system is interrupted
P_{pr}	:	Probability that the system is in a protected state
$P_S(idle)$	:	Probability that the system is idle
$E(I)$	:	Expected number of interruptions during a service
$E^I(D)$	:	Expected time spent in interruption during a service
$E(\tau)$	:	Expected service time
W_L	:	Expected waiting time of a customer in the queue

For numerical study we have taken the number of phases of the Erlang process, $m = 5$ and arrival rate $\lambda = 2$. Table 4.1 depicts the harm that interruptions can cause to the system and justifies the need for protected phases. From the table it can be seen that as the interruption rate θ increases, the expected service time $E(\tau)$ increases. This leads to the increase in the various expected numbers of customers $E_S(C), E_Q(C), E^B(C)$ and $E^I(C)$. Note that the $E_Q(C)$ comes closer to $E_S(C)$ as θ decreases. This indicates how much severe can be the effect of interruptions. Also note the high increase in the variance $V_S(C)$ of the number of customers in the system with increase in θ. A similar high rate of increase can be seen in the waiting time of a customer. The decrease in the server idle probability should be read together with the increase in the queue length which indicates that the server may be working in vain instead of a constant arrival rate.

The destructive effect of interruptions to the system performance has been revealed from the above discussion. To overcome this we came up with the idea of protection. In Table 4.2, the effect of the introduction of protected phases is studied. The table shows that by giving protection to sufficient number of phases, the expected service time can be brought down to a reasonable level. This decrease in $E(\tau)$in turn causes a decrease in the expected numbers $E_S(C), E_Q(C), E^B(C)$ and $E^I(C)$. Also note the variance $V_S(C)$ of the number of customers in the system and the ratio $\frac{E_S(C)}{E_Q(C)}$ can be lowered to desired levels in the presence of required number of protected phases. The high rate of decrease in certain performance measures like $W_L, E^I(D), P_S(I)$ etc. show the importance of providing protection to enough number of phases of a system that is subject to interruptions.

Table 4.3 shows the effect of service rate μ on various performance measures. The changes are as expected. When μ increases, expected number of customers in the system and in the queue will decrease. Here note the difference between these two quantities that points out the effect of interruption. This difference decreases as μ increases. $E^I(C)$ is another quantity that reflects the effect of interruption.

Table 4.4 describes the variations in the performance measures with γ which is narrow. The increase in $E_S(C), E_Q(C), E^B(C)$ and $E^I(C)$ are due to an increase in $E(\tau)$. This leads to a natural decrease in the server idle probability and the probability that the system is in an interrupted state. The increase in W_L also is due to the increase in $E(\tau)$.

Finally Table 4.5 gives the variation in performance measures with variation in repair rate δ. The effect of δ is just opposite to the effect of γ on $E(\tau)$: as δ increases, $E(\tau)$ decreases. Naturally the effect of δ on other measures is also opposite to that of γ.

Table 4.1: **Effect of the rate of Interruption θ on various performance measures**

Number of protected phases $n = 3$, Service rate $\mu = 8.0$,
Realizing rate of random clock $\gamma = 3.0$, Repair rate $\delta = 2.0$.

θ	1	2	3	4	6
$E_S(C)$	0.454	0.6346	0.8503	1.1114	1.8437
$V_S(C)$	0.83	1.4809	2.3409	3.5063	7.4997
$E_Q(C)$	0.1529	0.2819	0.4447	0.6524	1.2757
$E^I(C)$	0.1147	0.2493	0.4095	0.6034	1.1469
$E^B(C)$	0.3393	0.3855	0.4408	0.5079	0.6969
$P_S(I)$	0.0504	0.1015	0.1534	0.206	0.3135
$P_S(idle)$	0.6989	0.647	0.5944	0.541	0.432
$E(I)$	0.0504	0.1015	0.1534	0.206	0.3135
$E^I(D)$	0.0252	0.0507	0.0767	0.103	0.1568
$E(\tau)$	0.1506	0.1765	0.2028	0.2295	0.284
W_L	0.0764	0.1409	0.2223	0.3262	0.6379

Table 4.2: **Effect of the number of protected phases n on various performance measures**

Interruption Rate $\theta = 1$, Service rate $\mu = 5.0$, Realizing rate of random clock $\gamma = 3.0$, Repair rate $\delta = 2.0$.

n	1	2	3	4	5
$E_S(C)$	1.5136	1.1768	0.9204	0.72	0.56
$V_S(C)$	4.7508	3.0963	1.994	1.2279	0.6773
$E_Q(C)$	0.936	0.6481	0.4376	0.28	0.16
$E^I(C)$	0.5481	0.3602	0.2136	0.096	0
$E^B(C)$	0.9655	0.8166	0.7069	0.624	0.56
$P_S(I)$	0.1659	0.1229	0.081	0.04	0
P_{pr}	0.08	0.16	0.24	0.32	0.4
$P_S(idle)$	0.4224	0.4713	0.5171	0.56	0.6
$E(I)$	0.1659	0.1229	0.081	0.04	0
$E^I(D)$	0.0829	0.0615	0.0405	0.02	0
$E(\tau)$	0.2888	0.2644	0.2414	0.22	0.2
W_L	0.468	0.324	0.2188	0.14	0.08

4.7 Analysis of a Cost Function

From our discussion, it follows that to nullify the adverse effect of interruptions to the system there are three alternatives: Increase the repair rate, choose a higher service rate or protect the critical phases of service from interruptions. The first option is expensive and less efficient as there will always be some repair work to do during a service process. The second one also is expensive. The server has to work more for nothing. The efficiency is reduced in the sense that the number of customers

Table 4.3: **Effect of the rate of service μ on various performance measures**

Number of protected phases $n = 3$, Interruption Rate $\theta = 1.0$, Realizing rate of random clock $\gamma = 3.0$, Repair rate $\delta = 2.0$.

μ	5	8	10	12
$E_S(C)$	0.9204	0.454	0.3412	0.2737
$V_S(C)$	1.994	0.83	0.5993	0.4696
$E_Q(C)$	0.4376	0.1529	0.1005	0.0732
$E^I(C)$	0.2136	0.1147	0.0885	0.0722
$E^B(C)$	0.7069	0.3393	0.2527	0.2015
$P_S(I)$	0.081	0.0504	0.0402	0.0335
P_{pr}	0.24	0.15	0.12	0.1
$P_S(idle)$	0.5171	0.6989	0.7593	0.7995
$E(I)$	0.081	0.0504	0.0402	0.0335
$E^I(D)$	0.0405	0.0252	0.0201	0.0168
$E(\tau)$	0.2414	0.1506	0.1204	0.1003
W_L	0.2188	0.0764	0.0503	0.0366

served is the same in spite of the high service rate. Here is the importance of the third alternative

When we choose this option, the questions naturally arise are which are the phases to be protected and how many of them should be protected. For the first question the answer is to protect the critical phases. For an Erlang process these are the final phases. Now about the number of phases. Of course protecting every phase is a nice idea but it may be too expensive. The cost of protection depends on various system costs. Based on these costs we have to find an optimum number of phases

Table 4.4: **Effect of the rate of Realization of the random clock γ on various performance measures**

Number of protected phases $n = 3$, Service rate $\mu = 5.0$, Interruption Rate $\theta = 1.0$, Repair rate $\delta = 2.0$.

γ	3	4	6	10
$E_S(C)$	0.9204	0.9221	0.9242	0.926
$V_S(C)$	1.994	2.001	2.0088	2.0159
$E_Q(C)$	0.4376	0.4389	0.4406	0.442
$E^I(C)$	0.2136	0.2141	0.2147	0.2152
$E^B(C)$	0.7069	0.7081	0.7095	0.7108
$P_S(I)$	0.081	0.0811	0.0812	0.0813
$P_S(idle)$	0.5171	0.5168	0.5164	0.516
$E(I)$	0.081	0.0811	0.0812	0.0813
$E^I(D)$	0.0405	0.0405	0.0406	0.0407
$E(\tau)$	0.2414	0.2416	0.2418	0.242
W_L	0.2188	0.2195	0.2203	0.221

to be protected. This is explained by the following numerical experiment

For checking the optimality of the number of unprotected phases, we first study, how this parameter affects different performance measures. Table 4.6 below shows that as the number of unprotected phases increases, expected interruption rate increases. As a consequence there is an increase in the expected number of customers in the system, in the expected resume rate, and in the expected repeat rate. This was expected as the number of unprotected phases increases, the possibility of interruption increases and therefore there will be an increase in

Table 4.5: **Effect of the rate of Repair δ on various performance measures**

Number of protected phases $n = 3$, Service rate $\mu = 5.0$, Interruption Rate $\theta = 1.0$, Realizing rate of random clock $\gamma = 3.0$.

δ	3	4	6	10
$E_S(C)$	0.7533	0.6892	0.6362	0.6011
$V_S(C)$	1.2315	1.0006	0.8413	0.7549
$E_Q(C)$	0.2983	0.2479	0.2086	0.1845
$E^I(C)$	0.1148	0.0767	0.0452	0.0244
$E^B(C)$	0.6385	0.6125	0.591	0.5767
$P_S(I)$	0.0538	0.0403	0.0268	0.0161
$P_S(idle)$	0.545	0.5587	0.5724	0.5834
$E(I)$	0.0806	0.0805	0.0804	0.0803
$E^I(D)$	0.0269	0.0201	0.0134	0.008
$E(\tau)$	0.2275	0.2207	0.2138	0.2083
W_L	0.1491	0.124	0.1043	0.0923

the expected repeat/resume rates and also in the queue size. In contrast to this, the fraction of time, the service is in protected phases decreases with an increase in the number of unprotected phases. Hence to find an optimal value for the number of unprotected phases, we construct the following cost function

$$
\begin{aligned}
CF = CRPT \times ERPT + CRSM \times ERSM \\
+ CHOLD \times E_S(C) + CINT \times EI \\
+ CP \times P_{pr}
\end{aligned}
$$

where CRPT and CRSM are the unit time costs for repeating or resuming an interrupted service respectively. CHOLD is the holding cost per unit time per customer and CINT is the cost per unit time per interruption. Finally CP is the unit time cost for giving service in the protected phases.

In table 4.7 the behaviour of cost function is shown. It follows from the table that, when the protection cost CP is 600, the optimal number of unprotected phases is 1 whereas when this cost is increased to 650, the optimal value is 4; further if we increase this cost to 700, the optimal value becomes 7. Though these optimal values depend on the system parameters and costs, table 4.7 shows that we can have a control over the number of unprotected phases in favour of the system.

Table 4.6: **Effect of the number of protected phases n on various performance measures involved in the cost function**

No. of phases of the Erlang process $m = 10$, Arrival Rate $\lambda = 1.5$, Service Rate $\mu = 2.1$, Interruption Rate $\theta = 1.0$, Realizing rate of random clock $\gamma = 1.0$, Repair rate $\delta = 10.0$.

n	9	8	7	3	2	1
$ERPT$	0.0065	0.013	0.02	0.046	0.053	0.059
$ERSM$	0.065	0.13	0.196	0.46	0.527	0.595
$E_S(C)$	1.753	1.815	1.885	2.253	2.376	2.517
EI	0.071	0.143	0.215	0.506	0.58	0.654
P_{pr}	0.643	0.571	0.5	0.214	0.143	0.071

Table 4.7: **Effect of the number of protected phases n on the cost function**

$CRSM = 100, CRPT = 400, CINT = 400, CHOLD = 100$, No. of phases of the Erlang process $m = 10$, Arrival Rate $\lambda = 1.5$, Service Rate μ =2.1, Interruption Rate $\theta = 1.0$, Realizing rate of random clock $\gamma = 1.0$, Repair rate $\delta = 10.0$.

No of Protected Phases	Cost		
	CP= 600	CP=650	CP=700
1	598.7	630.8	663
2	599.9	628.4	657
3	601.9	626.9	652
4	604.9	626.4	647.8
5	609	626.9	644.7
6	614.3	628.6	642.9
7	621	631.7	642.4
8	629.3	636.4	643.5
9	639.4	643	646.6

Chapter 5

AN $M/M/1$ QUEUE WITH INTERRUPTION AND PROTECTION

5.1 Introduction

In chapter 4 we described a model in which the final few phases of the service is protected from interruptions so as to improve the system performance. This runs parallel to the N-policy. However, if the chance of interruption is considerably large, this method may not be fruitful enough. In such cases it may take much time to finish the unprotected phases. Here protecting more phases may not be a good idea. Moreover, in many real life situations, the instant at which protection is needed may vary from customer to customer. For example, in the treatment of chronic diseases, patients undergo a series of phases and during some of these phases the risk is higher. So patients in the risky phases should be given additional attention. The time to reach a serious phase depends on the physical condition of the patients. So one cannot insist that special attention (Protection) will begin only at a particular phase. Some patients may require protection soon after they develop the disease while some others can withstand for a long duration. This procedure of protecting a service is similar to the T-policy in Queueing theory.

Motivated by this chronic care model, we introduce a system in which the server is protected from interruptions after a random time from the start of each service.

5.2 Model Description

We consider a single server queueing model in which customers arrive according to a Poisson process with parameter λ. The service time follows exponential distribution with mean $\frac{1}{\mu}$. While rendering service the server may face some interruptions. The interruptions occur according to a Poisson process with parameter θ . The interrupted service restarts after a repair and the repair time follows exponential distribution with mean $\frac{1}{\delta}$. To diminish the effect of interruptions a protection mechanism is arranged. Once the protection mechanism is on, the service will continue without any further interruptions. This mechanism is provided after getting an uninterrupted service for a random time. This is done with the help of a random clock whose realization time is exponentially distributed with mean

$\frac{1}{\varphi}$. The clock is started simultaneously with the service process. If there were no interruptions until the realization of the clock, the protection for the service is provided at the epoch of the realization of the clock. If there is any interruption before the realization of the clock, it is reset and started again with the restart of the service. This model is described in Figure 5.1

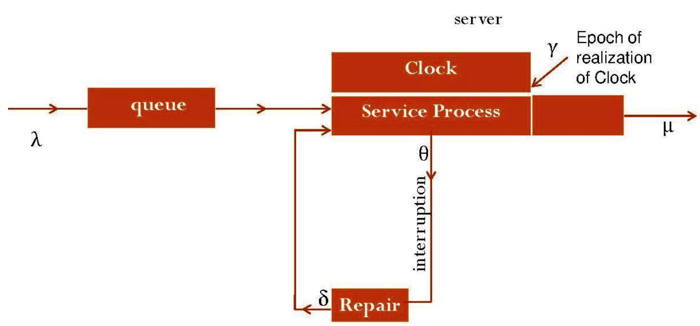


Figure 5.1: *An $M/M/1$ Queue with Interruption and Protection*

5.3 Mathematical Model

The system described in the previous section can be mathematically modelled as a Markov process

$$\boldsymbol{X} = \{X(t)/t \geq 0\} = \{(N(t), C(t), J(t))\,/t \geq 0\}$$

where $N(t)$ is the number of customers in the system , $C(t)$ is the status of the clock; it is 0 if clock is running and 1 if the clock is realized and $J(t)$ is the state of the server which is 0 if server is not interrupted and 1 if it is interrupted. The state space of the process is given by $\{0\}\cup(\{1,2,3,...\}\times\{0,1\}\times\{0,1\})$. The infinitesimal generator matrix of the process is given by

$$\mathbf{Q}=\begin{bmatrix} A_{10} & A_{00} & & & & & \\ A_{21} & A_1 & A_0 & & & & \\ & A_2 & A_1 & A_0 & & & \\ & & A_2 & A_1 & A_0 & & \\ & & & & \cdots & \cdots & \cdots & \\ & & & & & \cdots & \cdots & \cdots \end{bmatrix}$$

where

$$A_{10}=\begin{bmatrix}-\lambda\end{bmatrix}\quad A_{00}=\begin{bmatrix}\lambda & 0 & 0\end{bmatrix}\quad A_{21}=\begin{bmatrix}\mu\\ 0\\ \mu\end{bmatrix}$$

$$A_2=\begin{bmatrix}\mu & 0 & 0\\ 0 & 0 & 0\\ \mu & 0 & 0\end{bmatrix}, A_1=\begin{bmatrix}-(\mu+\theta+\varphi+\lambda) & \theta & \varphi\\ \delta & -(\delta+\lambda) & 0\\ 0 & 0 & -(\mu+\lambda)\end{bmatrix}$$

and $A_0=\lambda I$.

5.4 Stability Analysis

Let $A=A_0+A_1+A_2=\begin{bmatrix}-(\theta+\varphi) & \theta & \varphi\\ \delta & -\delta & 0\\ \mu & 0 & -\mu\end{bmatrix}$ and $\pi=(\pi_1,\pi_2,\pi_3)$ be the invariant vector of A such that $\pi\boldsymbol{e}=1$. Then the system is stable if and only if $\pi A_0\boldsymbol{e}<\pi A_2\boldsymbol{e}$ Hence we have the following theorem.

Theorem 5.4.1. *The system* $\boldsymbol{X}$ *is stable if and only if*

$$\frac{\lambda}{\mu} < \left(1+\frac{\varphi}{\mu}\right)\left(1+\frac{\varphi}{\mu}+\frac{\theta}{\delta}\right)^{-1}$$

5.5 Steady State Analysis

The steady state probability vector is obtained from the equation $\boldsymbol{xQ} = 0$. Writing $\boldsymbol{x} = (x_0, x_1, x_2, ...)$, a matrix geometric solution of the above equation is given by

$$x_0 = 1 - \rho$$

$$x_i = (1-\rho)\alpha R^i, i = 1, 2, 3, ...$$

where

$$\rho = \frac{\lambda}{\mu+\varphi}\left(1+\frac{\varphi}{\mu}+\frac{\theta}{\delta}\right)$$

$$R = \frac{\lambda}{\mu\left(\mu+\lambda+\varphi\right)}\begin{bmatrix} \mu+\lambda & \frac{\theta(\mu+\lambda)}{\delta+\lambda} & \varphi \\ \mu+\lambda & \frac{\theta(\mu+\lambda)}{\delta+\lambda}+\frac{\mu(\mu+\lambda+\varphi)}{\delta+\lambda} & \varphi \\ \lambda & \frac{\theta\lambda}{\delta+\lambda} & \mu+\lambda \end{bmatrix}$$

and $\alpha = (1, 0, 0)$.

5.6 Analysis of the Service Process

Obviously we can see that the service time follows phase type distribution with representation (α, S) where $\alpha = (1, 0, 0)$ and

$$S = \begin{bmatrix} -(\mu+\theta+\varphi) & \theta & \varphi \\ \delta & -\delta & 0 \\ 0 & 0 & -\mu \end{bmatrix}$$

The following results can be easily derived.

- The expected service time, $E(\tau) = \frac{1}{\mu} + \frac{\theta}{\mu+\varphi}\frac{1}{\delta}$
- Increase in service time due to interruption $= \frac{\theta}{\mu+\varphi}\frac{1}{\delta}$
- During a service, the time spent in the unprotected uninterrupted state $= \frac{1}{\mu+\varphi}$
- Time spent in the protected state $= \frac{\varphi}{\mu+\varphi}\frac{1}{\delta}$
- Probability that the service is completed before protection starts $= \frac{\mu}{\mu+\varphi}$

5.6.1 Expected number of interruptions during any particular service

Let $N(t)$be the number of interruptions during a particular service at time t. Let $J(t)$ be the status of the server at time

t. Then $Y = \{(N(t), J(t))\,/t \geq 0\}$ is a Markov Process whose infinitesimal generator matrix is given by

$$\tilde{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & & & & & \\ S_2 & S_1 & S_0 & & & & & \\ S_2 & \mathbf{0} & S_1 & S_0 & & & & \\ S_2 & \mathbf{0} & \mathbf{0} & S_1 & S_0 & & & \\ \vdots & & & & \cdots & \cdots & & \\ \vdots & & & & & \cdots & \cdots & \end{bmatrix}$$

where

$$S_0 = \begin{bmatrix} \mu+\varphi \\ 0 \end{bmatrix}, S_1 = \begin{bmatrix} -(\mu+\theta+\varphi) & 0 \\ 0 & -\delta \end{bmatrix} \text{ and } S_2 = \begin{bmatrix} 0 & \theta \\ \delta & 0 \end{bmatrix}.$$

Let y_k be the probability that there are exactly k interruptions during a service. Then

$$y_0 = \alpha(-S_1)^{-1}S_2$$

and

$$y_k = \alpha\left[(-S_1)^{-1}S_0\right]^k(-S_1)^{-1}S_2, \;\; k = 1, 2, 3, \ldots \quad .$$

Simplifying we get,

$$y_k = \frac{\mu+\varphi}{\mu+\varphi+\theta}\left(\frac{\theta}{\mu+\varphi+\theta}\right)^k, \;\; k = 0, 1, 2, \ldots \quad .$$

Theorem 5.6.1. *The mean number of interruptions occur dur-*

ing a single service is

$$E(I) = \frac{\theta}{\mu + \varphi}.$$

Proof. The mean number of interruptions,

$$E(I) = \sum_{k-0}^{\infty} k y_k = \frac{\theta}{\mu + \varphi}$$

$\square$

5.7 Expected Waiting Time

Consider a customer who joins as the r^{th} customer in the queue, $r > 0$. The waiting time of this customer may be described as the time until absorption of a Markov Chain $\mathbf{W} = \{W(t)/t \geq 0\} = \{(N(t), C(t), J(t))\,/t \geq 0\}$ where $N(t)$ is the rank of the tagged customer, $C(t)$ is 0 if the service is unprotected and 1 otherwise and $J(t)$ is 1 if the service is interrupted and 0 otherwise at time t. Thus the waiting time of the tagged customer has a phase type distribution with representation (β, B) where

$$B = \begin{bmatrix} S & S^0\boldsymbol{\alpha} & & & \\ & S & S^0\boldsymbol{\alpha} & & \\ & & \cdots & \cdots & \\ & & & & S \end{bmatrix}$$

and β is the initial probability vector which ensures that the chain always starts from level r.

Therefore the expected waiting time of the tagged customer according to the state of the server at the time of joining the queue,

$$\begin{aligned}
E_W^r &= -B^{-1}\boldsymbol{e} \\
&= -S^{-1}\left[I - S^0\boldsymbol{\alpha}S^{-1} + ... + (-1)^{r-1}\left(S^0\boldsymbol{\alpha}S^{-1}\right)^{r-1}\right]\boldsymbol{e} \\
&= -S^{-1}\boldsymbol{e} + \frac{r-1}{\mu+\varphi}\left(1 + \frac{\theta}{\delta}\frac{\varphi}{\mu}\right)\boldsymbol{e} \\
&= -S^{-1}\boldsymbol{e} + (r-1)E(\tau)\boldsymbol{e}.
\end{aligned}$$

Hence the expected waiting time of a customer who has to wait is

$$E(W) = \sum_{r=1}^{\infty} x_r E_W^r.$$

5.8 Some Performance Measures

- Probability that the system is idle,

$$P_s(idle) = 1 - \rho$$

- Probability that the system is busy and uninterrupted without protection on

$$\frac{\lambda}{\mu+\varphi}$$

- Probability that the system is in interruption,

$$\frac{\theta}{\mu+\varphi}\frac{\lambda}{\delta}$$

- Probability that the system is under protection,

$$\frac{\varphi}{\mu+\varphi}\frac{\lambda}{\mu}$$

- Expected number of customers in the system

$$\rho+\frac{\lambda}{1-\rho}\left[\frac{\rho}{\mu}+\frac{\lambda\theta\left(\mu+\varphi+\theta+\delta\right)}{(\mu+\varphi)^2\delta^2}\right]$$

5.9 Numerical Illustration

A numerical study of the effect of various parameters involved on various performance measures is carried out in this section.

Table 5.1 shows the effect of φ, the rate of realization of the protection clock. As φ increases the service is protected from interruption that much quicker and hence reduces the chance of being interrupted. This improves the performance of the system. As the expected number of interruptions is decreased, the service time is reduced. This results in a reduction of expected number of customers in the system and expected waiting time. When $\varphi \to \infty$, the system reduces to an ordinary $M/M/1$ queue.

In Table 5.2, the effect of interruption rate θ on various performance measures is illustrated. As expected, an increase in θ increases the probability that the system is interrupted and hence the service time increases. As a consequence, the expected number of customers in the system and expected waiting time are also increased.

Table 5.3 explains the effect of the repair rate δ on the performance of the system.As the repair rate increases, the time spent in the interrupted state decreases and results in a higher service rate. This explains the changes in different measures. If δ is very large, the repair is done instantaneously and the time spent in the interrupted state becomes zero.

Table 5.1: **Variation in different performance measures with the rate of realization of protection clock**

Arrival Rate $\lambda = 3$, Service Rate $\mu = 10$, Interruption Rate $\theta = 3$, repair rate $\delta = 5$.

φ	$P_s(Idle)$	$P_s(Prot)$	$P_s(I)$	$E_s(C)$	$E(W)$	$E(Ser)$	$E(I)$
1	0.4818	0.0273	0.2182	0.8166	0.1827	0.1727	0.3636
2	0.5000	0.0500	0.2000	0.7200	0.1567	0.1667	0.3333
3	0.5154	0.0692	0.1846	0.6458	0.1370	0.1615	0.3077
4	0.5286	0.0857	0.1714	0.5873	0.1217	0.1571	0.2857
5	0.5400	0.1000	0.1600	0.5400	0.1095	0.1533	0.2667
6	0.5500	0.1125	0.1500	0.5011	0.0995	0.1500	0.2500
7	0.5588	0.1235	0.1412	0.4687	0.0913	0.1471	0.2353
8	0.5667	0.1333	0.1333	0.4412	0.0845	0.1444	0.2222
9	0.5737	0.1421	0.1263	0.4176	0.0786	0.1421	0.2105
10	0.5800	0.1500	0.1200	0.3972	0.0736	0.1400	0.2000
large	0.7000	0.3000	0.0000	0.1286	0.0129	0.1000	0.4000

Table 5.2: **Variation in different performance measures with the interruption rate**

Arrival Rate $\lambda = 3$, Service Rate $\mu = 10$, Protection clock realization rate $\varphi = 4$ Rate $\theta = 3$, repair rate $\delta = 5$.

θ	$P_s(Idle)$	$P_s(Prot)$	$P_s(I)$	$E_s(C)$	$E(W)$	$E(Ser)$	$E(I)$
1	0.6571	0.0857	0.0429	0.2124	0.0316	0.1143	0.0714
2	0.6143	0.0857	0.0857	0.3140	0.0551	0.1286	0.1429
3	0.5714	0.0857	0.1286	0.4371	0.0845	0.1429	0.2143
4	0.5286	0.0857	0.1714	0.5873	0.1217	0.1571	0.2857
5	0.4857	0.0857	0.2143	0.7714	0.1690	0.1714	0.3571
6	0.4429	0.0857	0.2571	0.9995	0.2297	0.1857	0.4286
7	0.4000	0.0857	0.3000	1.2857	0.3086	0.2000	0.5000
8	0.3571	0.0857	0.3429	1.6509	0.4125	0.2143	0.5714
9	0.3143	0.0857	0.3857	2.1273	0.5524	0.2286	0.6429
10	0.2714	0.0857	0.4286	2.7677	0.7456	0.2429	0.7143

5.10 Analysis of Cost Function

The results in the previous sections show the effect of interruptions on the system performance and the extend to which it can be minimized through giving protection. The cost of service will increase due to the introduction of protection. Longer the time the service is protected the more will be the cost. So a question naturally arises about the epoch at which the protection starts so that the service cost is optimum. For checking this optimality we consider the following factors:- The expected number of interruptions $E(I)$,The fraction of time in which the system is protected $T(Pr)$, The fraction of time in which the system is unprotected $T(unpr)$, The number of customers in

Table 5.3: **Variation in different performance measures with the repair rate**

Arrival Rate $\lambda = 3$, Service Rate $\mu = 10$, Protection clock realization rate $\varphi = 5$ Interruption rate $\theta = 3$.

δ	$P_s(Idle)$	$P_s(Prot)$	$P_s(I)$	$E_s(C)$	$E(W)$	$E(Ser)$	$E(I)$
1	0.1000	0.1	0.6000	25.5000	8.2300	0.3000	**0.2**
2	0.4000	0.1	0.3000	1.9500	0.5300	0.2000	0.2
3	0.5000	0.1	0.2000	0.8600	0.2033	0.1667	0.2
4	0.5500	0.1	0.1500	0.5455	0.1143	0.1500	0.2
5	0.5800	0.1	0.1200	0.4076	0.0771	0.1400	0.2
6	0.6000	0.1	0.1000	0.3333	0.0578	0.1333	0.2
7	0.6143	0.1	0.0857	0.2880	0.0464	0.1286	0.2
8	0.6250	0.1	0.0750	0.2580	0.0391	0.1250	0.2
9	0.6333	0.1	0.0667	0.0341	0.2368	0.1222	0.2
10	0.6400	0.1	0.0600	0.2213	0.0306	0.1200	0.2
large	0.7000	0.1	0.0000	0.1286	0.0129	0.1000	0.2

the system $E_s(C)$.

Each time an interruption occurs, the repair has to be done and let RC be the repair cost. Let PC be the unit time cost of running the server with protection and UPC be that without protection. Let HC be the holding cost for retaining a customer for unit time. Then the service cost per unit time is given by

$$C = E(I) \times RC + E_s(C) \times HC + T(pr) \times PC + T(upr) \times UPC.$$

The variation in cost with the time to start protection when $\mu = 10$,$\lambda = 3$, $\theta = 4$ and δ =5 with costs $RC = 4$,$HC = 0.5$, $PC = 20$ and UPC= 2.5 is given in Table 5.4. In this case the optimum value for φ is 6

Table 5.4: **The variation in cost with the time to start protection**

Arrival Rate $\lambda = 3$, Service Rate $\mu = 10$, Interruption rate $\theta = 4$, repair rate $\delta = 5$.

φ	2	3	4	5	**6**	7	8	9
C	5.985	5.9768	5.9722	5.97	**5.9693**	5.9696	5.9706	5.972

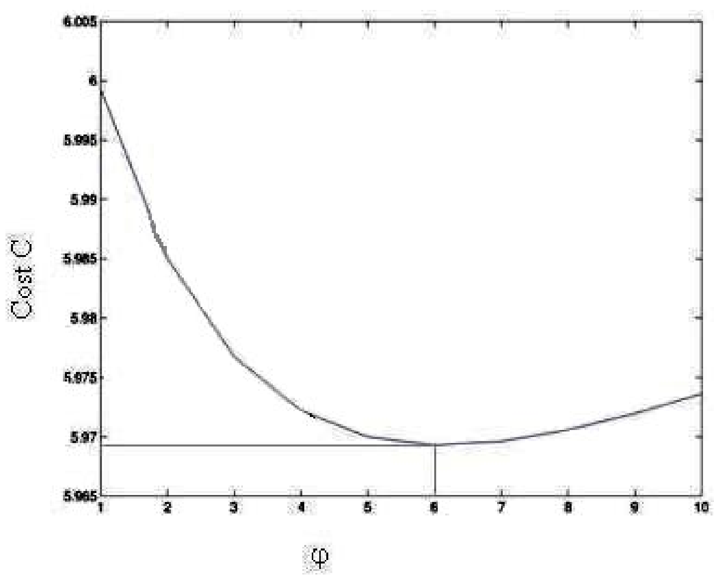


Figure 5.2: *Variation in Cost with* φ

Chapter 6

AN $M/E_m/1$ QUEUE WITH PROTECTION BASED ON T-POLICY

6.1 Introduction

In the last chapter we analysed a queueing model having a service process with interruption and a time based method of protection, some thing like a T-policy. But the service time, if there were no interruption was assumed to be exponential. In many practical situations as pointed out in chapter 4, the service process may consist of more than one phase. In such systems, the repeat and resumption of service have distinct roles to play. The chance of repeating an interrupted service makes this T-policy protection very interesting.

Moreover, we have already introduced a method of protecting some phases of the service process from interruption, say an N-policy. It would be meaningful to compare the two types of protection. This is one of the motive of this chapter

6.2 Model Description

Consider a single server queuing system in which the arrival process is Poisson with rate λ. The service time distribution is Erlang of order m with mean $\frac{1}{\mu}$. Interruptions occurs to the service process at an exponentially distributed time durations with mean $\frac{1}{\theta}$. As in the previous models, the interrupted server is taken for repair immediately with repair duration follows an exponential distribution with mean $\frac{1}{\delta}$. A random clock is started at the beginning of each repair to decide whether to restart or resume the service after repair. If the random clock realizes before a repair, the service needs to be repeated, otherwise the service is resumed in the phase at which the interruption occurred. The realization time of the random clock also follows an exponential distribution, with mean $\frac{1}{\gamma}$. To avoid the situation where the system is interrupted while the service nearing completion, and had to repeat the service from the beginning, a protective mechanism is provided. To minimize the cost of running the system, this mechanism will be used only while the service is continued for sufficiently long time. Thus a clock is started simultaneously with the beginning of a new or an interrupted service. The protection is provided for the part of the service that remains after the realization of this random clock. The realization time of this clock is assumed to be an exponential variable with mean $\frac{1}{\varphi}$. In effect we extend the service protection mechanism introduced in Chapter 5 to Erlang distributed service time.

This queueing model can be defined by the Markov process $\boldsymbol{X} = \{X(t)/t \geq 0\} = \{(N(t), P(t), S(t), J(t)), t \geq 0\}$ where $N(t)$ is the number of customers in the system, $P(t)$ is the status of the protective mechanism which is 0 or 1 according as the mechanism is off or on, $S(t)$ is the status of the server which is 0, 1 or 2 according as the service is uninterrupted, interrupted with a running clock or interrupted with a realized clock and $J(t)$ is the phase of the service process at time t. The state space is given by $\{0, 1, 2, 3, ...\} \times \{0, 1\} \times \{0\} \times \{1, 2, ..., m\} \cup \{0, 1, 2, 3, ...\} \times \{0\} \times \{1, 2\} \times \{1, 2, ..., m\}$ and the infinitesimal generator matrix given by

$$Q = \begin{bmatrix} A_{10} & A_{00} & & & & \\ A_{21} & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & . & . & . \\ & & & & & \\ & & & & & \end{bmatrix}$$

where

$$A_{10} = \begin{bmatrix} -\lambda \end{bmatrix} \quad A_{00} = \begin{bmatrix} \lambda\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 & .... & 0 \end{bmatrix}$$

$A_0 = \lambda I_{4m}$

$$A_{21} = \begin{bmatrix} S^0 \\ \mathbf{0} \\ \mathbf{0} \\ S^0 \end{bmatrix} \quad A_2 = \begin{bmatrix} S^0\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ S^0\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$A_1 = \begin{bmatrix} S-(\theta+\varphi+\lambda)I & \theta I & \mathbf{0} & \varphi I \\ \delta I & -(\gamma+\delta+\lambda)I & \gamma I & \mathbf{0} \\ \delta \boldsymbol{e\alpha} & \mathbf{0} & -(\delta+\lambda)I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & S-\lambda I \end{bmatrix}$$

$$S^0 = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ m\mu \end{bmatrix} \quad S = m\mu \begin{bmatrix} -1 & 1 & 0 & & & \\ & -1 & 1 & 0 & & \\ & & . & . & . & \\ & & & . & . & . \\ & & & & . & . & . \\ & & & & & & -1 \end{bmatrix}$$

and I is the identity matrix of order m.

6.3 Analysis of the service process

6.3.1 Expected service time

The service process with the introduction of interruption, becomes a Markov Process Ψ with $4m$ transient states given by

$$\{0,1,2,3\} \times \{1,2,...,m\}$$

and one absorbing state $\widetilde{0}$. The absorbing state denotes the service completion. The process Ψ can be represented by $\psi(t) = (i, j)$ where $i = 0$ if the service is uninterrupted, $i = 1$ if the service is interrupted with restart clock running $i = 2$ if the service is interrupted with a finished restart clock and $i = 3$ if the service is protected and j is the phase of service. The initial probability vector is $\beta = (\alpha, 0, 0, 0)$ Let τ be the time until absorption of the process Ψ. The infinitesimal generator of this process is given by $\widetilde{Q} = \begin{bmatrix} T & T^0 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where

$$T = \begin{bmatrix} S-(\theta+\varphi)\,I & \theta I & \mathbf{0} & \varphi I \\ \delta I & -(\gamma+\delta)I & \gamma I & \mathbf{0} \\ \delta \boldsymbol{e\alpha} & \mathbf{0} & -\delta I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & S \end{bmatrix} \quad T^0 = \begin{bmatrix} S^0 \\ \mathbf{0} \\ \mathbf{0} \\ S^0 \end{bmatrix}$$

$\beta(-T)^{-1} = (y_1, y_2, ..., y_{4m})$ gives us the time spend in each state in a single service and can be calculated as

$$y_i = \begin{cases} \left[\frac{1}{m\mu}\left(m\mu+\varphi+\frac{\gamma\theta}{\gamma+\delta}\right)\right]^{m-i}\omega, & i = 1, 2, ..., m \\ \frac{\theta}{\gamma+\delta}\left[\frac{1}{m\mu}\left(m\mu+\varphi+\frac{\gamma\theta}{\gamma+\delta}\right)\right]^{m-i}\omega, & i = m+1, ..., 2m \\ \frac{\gamma\theta}{\delta(\gamma+\delta)}\left[\frac{1}{m\mu}\left(m\mu+\varphi+\frac{\gamma\theta}{\gamma+\delta}\right)\right]^{m-i}\omega, & i = 2m, ..., 3m \\ \frac{\varphi}{m\mu}\left(\sum_{j=4m-i}^{m-1}\left[\frac{1}{m\mu}\left(m\mu+\varphi+\frac{\gamma\theta}{\gamma+\delta}\right)\right]^{j}\right)\omega, & i = 3m+1, ..., 4m. \end{cases}$$

where

$$\frac{1}{\omega} = m\mu + \varphi \sum_{j=0}^{m-1} \left[\frac{1}{m\mu} \left(m\mu + \varphi + \frac{\gamma\theta}{\gamma+\delta} \right) \right]^{j}$$

Therefore, during a single service

- The expected time spent in the protected phases is

$$\frac{\varphi}{m\mu} \left(\sum_{j=1}^{m} j \left[\frac{1}{m\mu} \left(m\mu + \varphi + \frac{\gamma\theta}{\gamma+\delta} \right) \right]^{j-1} \right) \omega.$$

- The expected time spent in the uninterrupted unprotected phases is

$$\left(\sum_{j=1}^{m} \left[\frac{1}{m\mu} \left(m\mu + \varphi + \frac{\gamma\theta}{\gamma+\delta} \right) \right]^{j-1} \right) \omega.$$

- Expected interruption duration is

$$\frac{\theta}{\delta} \left(\sum_{j=1}^{m} \left[\frac{1}{m\mu} \left(m\mu + \varphi + \frac{\gamma\theta}{\gamma+\delta} \right) \right]^{j-1} \right) \omega.$$

The expected service time is the expect time until absorption of the above process which is given by,

$$\begin{aligned} E(\tau) &= \beta(-T)^{-1}\boldsymbol{e} \\ &= \left(1 + \frac{\theta}{\delta}\right) \left(\sum_{j=1}^{m} \left[\frac{1}{m\mu} \left(m\mu + \varphi + \frac{\gamma\theta}{\gamma+\delta} \right) \right]^{j-1} \right) \omega \end{aligned}$$

$$+\frac{\varphi}{m\mu}\left(\sum_{j=1}^{m} j\left[\frac{1}{m\mu}\left(m\mu+\varphi+\frac{\gamma\theta}{\gamma+\delta}\right)\right]^{j-1}\right)\omega.$$

6.3.2 Expected number of interruptions during a single service

To compute the expected number of interruptions faced by a customer during his service, we consider the Markov process $\boldsymbol{\chi} = \{(N(t), J(t))\, /t \geq 0\}$ where $N(t)$ is the number of interruptions that occurred up to time t and $J(t)$ is the phase of the service process at time t. $\boldsymbol{\chi}$ has the state space $\{0, 1, 2, 3, ...\} \times \{1, 2, 3, ..., m\} \cup \{\Delta\}$ where Δ is the absorbing state which denotes that the service process reached the first protected phase and there will be no more interruptions. The infinitesimal generator $\tilde{Q}$ of this process is

$$Q = \begin{bmatrix} D_2 & D_1 & D_0 & . & & & \\ D_2 & 0 & D_1 & D_0 & . & & \\ D_2 & 0 & 0 & D_1 & D_0 & . & \\ . & . & . & . & . & & \\ . & . & . & . & . & . & \\ & & & & & & \end{bmatrix}$$

where

$$D_1 = -m\mu \begin{bmatrix} -1 & 1 & \mathbf{0} & & & \\ & -1 & 1 & & & \\ & & -1 & & & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & 1 \\ & & & & & & -1 \end{bmatrix}_m - (\theta + \varphi) I_m$$

$$D_0 = \begin{bmatrix} \theta & \mathbf{0} & \mathbf{0} & \cdot & & & \\ \frac{\gamma\theta}{\gamma+\delta} & \frac{\delta\theta}{\gamma+\delta} & 0 & \cdot & & & \\ \frac{\gamma\theta}{\gamma+\delta} & \mathbf{0} & \frac{\delta\theta}{\gamma+\delta} & & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\gamma\theta}{\gamma+\delta} & \mathbf{0} & \mathbf{0} & \cdot & \cdot & \cdot & \frac{\delta\theta}{\gamma+\delta} \end{bmatrix}_m \text{ and } D_2 = \begin{bmatrix} \varphi \\ \varphi \\ \varphi \\ \cdot \\ \cdot \\ m\mu + \varphi \end{bmatrix}_{(m-n)\times 1}$$

If z_j is the probability that there are exactly j interruptions during a single service, then

$$z_j = \zeta(-D_1^{-1}D_0)^j(-D_1^{-1}D_2), j = 0, 1, 2, ...$$

Expected number of interruptions during a single service,

$$\begin{aligned} E(I) &= \sum_{j=0}^{\infty} j z_j \\ &= \zeta(-D_1^{-1}D_0)\left[I - \left(-D_1^{-1}\right) D_0\right]^{-1} \boldsymbol{e} \\ &= \theta\left(\kappa^m - 1\right)\left(\varphi\kappa^m + \frac{\gamma\theta}{\gamma+\delta}\right)^{-1} \end{aligned}$$

where

$$\kappa = 1 + \frac{\varphi}{m\mu} + \frac{\gamma\theta}{m\mu\,(\gamma+\delta)}$$

6.4 Steady state Analysis

6.4.1 Stability analysis

Let $A = A_0 + A_1 + A_2 = \begin{bmatrix} S-(\theta+\varphi)I+S^0\boldsymbol{\alpha} & \theta I & \mathbf{0} & \varphi I \\ \delta I & -(\gamma+\delta)I & \gamma I & \mathbf{0} \\ \delta\boldsymbol{e}\boldsymbol{\alpha} & \mathbf{0} & -\delta I & \mathbf{0} \\ S^0\boldsymbol{\alpha} & \mathbf{0} & \mathbf{0} & S \end{bmatrix}$
and $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \pi_3)$ be the vector such that $\boldsymbol{\pi}A = 0$ and $\boldsymbol{\pi}\boldsymbol{e} = 1$. Then

$$\pi_0\left(S-(\theta+\varphi)I+S^0\boldsymbol{\alpha}\right)+\delta\pi_1+\pi_2\delta\boldsymbol{e}\boldsymbol{\alpha}+\pi_3S^0\boldsymbol{\alpha}=0 \quad (6.1)$$

$$\theta\pi_0-(\gamma+\delta)\,\pi_1=0 \quad (6.2)$$

$$\gamma\pi_1-\delta\pi_2=0 \quad (6.3)$$

$$\varphi\pi_0-\pi_3S=0 \quad (6.4)$$

From the above equations, we get

$$\pi_0\left[\left(S+S^0\boldsymbol{\alpha}\right)-\left(\frac{\gamma\theta}{\gamma+\delta}+\varphi\right)(I-\boldsymbol{e}\boldsymbol{\alpha})\right]=0$$

Solving, we get

$$\pi_0 = \sigma(1, k, k^2, ..., k^{m-1})$$

where σ is a non-zero constant and $k = \frac{1}{m\mu}\left(m\mu + \varphi + \frac{\gamma\theta}{\gamma+\delta}\right)$. Therefore using equations (6.1), (6.2) and (6.3) we have,

$$\pi_0 \boldsymbol{e} = (1 + k + k^2 + ... + k^{m-1})\sigma$$

$$\pi_1 \boldsymbol{e} = \frac{\theta}{\gamma+\delta}(1 + k + k^2 + ... + k^{m-1})\sigma$$

$$\pi_2 \boldsymbol{e} = \frac{\gamma\theta}{\delta(\gamma+\delta)}(1 + k + k^2 + ... + k^{m-1})\sigma$$

$$\pi_3 \boldsymbol{e} = \frac{\varphi}{m\mu}(1 + 2k + 3k^2 + ... + mk^{m-1})\sigma$$

and

$$\pi \boldsymbol{e} = \left[\left(1 + \frac{\theta}{\gamma+\delta} + \frac{\gamma\theta}{\delta(\gamma+\delta)}\right)(1 + k + k^2 + ... + k^{m-1}) + \frac{\varphi}{m\mu}(1 + 2k + 3k^2 + ... + mk^{m-1})\right]\sigma$$

Therefore

$$\sigma = \left[\left(1 + \frac{\theta}{\gamma+\delta} + \frac{\gamma\theta}{\delta(\gamma+\delta)}\right)(1 + k + k^2 + ... + k^{m-1}) + \frac{\varphi}{m\mu}(1 + 2k + 3k^2 + ... + mk^{m-1})\right]^{-1}$$

Also it follows that $\boldsymbol{\pi} A_2 \boldsymbol{e} = \left[\varphi\left(1 + k + k^2 + ... + k^{m-1}\right) + m\mu\right]\sigma$ and $\boldsymbol{\pi} A_0 \boldsymbol{e} = \lambda$ Hence the condition for stability is

$$\lambda\left[\left(1 + \frac{\theta}{\gamma+\delta} + \frac{\gamma\theta}{\delta(\gamma+\delta)}\right)\left(1 + k + k^2 + ... + k^{m-1}\right) + \frac{\varphi}{m\mu}(1 + 2k + 3k^2 + ... + mk^{m-1})\right] < \varphi\left(1 + k + k^2 + ... + k^{m-1}\right) + m\mu \quad (6.5)$$

6.4.2 Stationary probabilities

The stationary probability vector $\boldsymbol{x}$ is given by $\boldsymbol{x}Q = 0$, together with the normalizing condition $\boldsymbol{xe} = 1$. On partitioning the steady state vector as $\boldsymbol{x} = (x_0, x_1, x_2, ...)$, the equation $\boldsymbol{x}Q = 0$ reduces to the following equations.

$$\begin{aligned} x_0A_{10} + x_1A_{21} &= 0 \\ x_0A_{00} + x_1A_1 + x_2A_2 &= 0 \\ x_iA_0 + x_{i+1}A_1 + x_{i+2}A_2 &= 0, \;\; i = 1, 2, 3, ... \quad . \end{aligned}$$

Also $\lambda x_i\boldsymbol{e} = x_{i+1}A_{21}, i = 0, 1, 2, ...$ and $A_2 = A_{21}\begin{bmatrix}\boldsymbol{\alpha} & 0 & 0 & 0\end{bmatrix}$ Solving we get $x_0 = 1-\rho, x_i = (1-\rho)\beta R^i$ for $i = 1, 2, 3, ...$where $\beta = (\alpha, 0, 0, 0)$, $R = -\lambda[A_1 + \lambda\boldsymbol{e}\beta]^{-1}$. and $\rho = -\lambda\boldsymbol{\alpha}T^{-1}\boldsymbol{e}$.

6.5 Performance Analysis

6.5.1 Expected waiting time

The expected waiting time W_L of a customer in the queue is found by the tagged customer method as in section (4.5.1).Thus W_L is given by

$$W_L = \lambda\beta(-T)^{-2}\boldsymbol{e} + E(\tau)E_Q(C)$$

where $E_Q(C) = \sum_{r=1}^{\infty}(r-1)x(r)\boldsymbol{e}$ is the expected number of customers in the queue and $E(\tau)$ is the expected service time.

6.5.2 Other performance measures

Each probability vector x_i except x_0 in the steady state probability vector

$$\boldsymbol{x} = (x_0, x_1, x_2, ...)$$

can be partitioned as

$$x_i = \left(x_i^{(0)}, x_i^{(1)}, x_i^{(2)}, x_i^{(3)}\right)$$

where

$$x_i^{(j)} = \left(x_i^{(j)}(1), x_i^{(j)}(2), ..., x_i^{(j)}(m)\right), j = 0, 1, 2, 3 \text{ and } i = 1, 2, 3, ...$$

With this notation, we have the following results.

- Probability of no customer in the system,

$$P_C(0) = x_0.$$

- Probability of i customers in the system,

$$P_C(i) = x_i\boldsymbol{e}.$$

- Expected number of customers in the system,

$$E_S(C) = \sum_{i=0}^{\infty} i P_C(i).$$

- Expected queue length,

$$E_Q(C) = \sum_{i=0}^{\infty} (i-1) P_C(i).$$

- Probability that there are i customers in the system when the server is interrupted,

$$P_C^I(i) = \left(x_i^{(1)} + x_i^{(2)}\right) \boldsymbol{e}.$$

- Expected number of customers in the system when the system is under interruption,

$$E^I(C) = \sum_{i=0}^{\infty} i P_C^I(i).$$

- Probability that there are i customers in the system when the server is uninterrupted,

$$P_C^I(i) = \left(x_i^{(0)} + x_i^{(3)}\right) \boldsymbol{e}.$$

- Expected number of customers in the system when the server is uninterrupted,

$$E^B(C) = \sum_{i=0}^{\infty} i P_C^B(i).$$

- Variance of the number of customers in the system,

$$V_S(C) = \sum_{i=0}^{\infty} i^2 P_C(i) - \left(\sum_{i=0}^{\infty} i P_C(i) \right)^2.$$

- Probability that the system is under interruption,

$$P_S(I) = \sum_{i=0}^{\infty} P_C^I(i).$$

- Probability that the system is protected state,

$$P_{pr} = \sum_{i=1}^{\infty} x_i^{(3)} \boldsymbol{e}.$$

- Effective interruption rate,

$$EI = \theta \sum_{i=1}^{\infty} x_i^{(0)} \boldsymbol{e}.$$

- Effective rate of service resumption,

$$ERSM = \delta \sum_{i=1}^{\infty} x_i^{(1)} \boldsymbol{e}.$$

- Effective rate of repetition of service,

$$ERSM = \delta \sum_{i=1}^{\infty} x_i^{(2)} \boldsymbol{e}.$$

6.6 Numerical Study of the Model

This section a numerical analysis of the model is done. This will be helpful in understanding the effect of various parameters in the performance of the system. In this examination, we consider a service process in which the service time distribution is Erlang-5. The arrival rate is assumed to be 2. For the easiness of comparison, we use the same notations used in section 4.6.

The results of the study with different values of θ is given in Table 6.1. The value of θ is increased from 0 to 6. As θ increases, the interruption occurs more frequently. Thus the number of interruptions per service as well as the duration of interruption during a service are increased. This results in an increase in the service time. This explains the changes in the other measures like expected number of customers, the waiting time, probability that the system is interrupted and the probability that the system is idle.

One would be interested to see what happens to the system when rate of realization of protection clock φ is increased. Table 6.2 gives us an idea about this. As φ increases, the system gets protected sooner. So the probability of being interrupted decreases, resulting in a faster service. Hence the service time is lowered and so is the number of customers in the system and the waiting time. Also note the increase in idle probability with increase in φ.

Next we will see how the rate of realization of repeat clock, γ acts on the system performance measures. Table 6.3 comprises of the numerical results obtained. As the possibility of repeating the service increases with increase in γ, the service rate is slightly increased. So there is more chance that the system will be protected. This is the reason why the change in service rate is small. Because of the same reason the number of interruption during a service remains almost unchanged as the interruption duration. Increase in service time results in increased number of customers waiting and longer waiting times.

Our Numerical experiment shows heavy dependency of the system performance measures on the rate of repair γ as shown in Table 6.4. As γ increases, the interruptions are repaired quickly so the duration of each interruption decreases. Also the chance that an interrupted service is resumed from where it was interrupted increases. Due to these reasons the service time decreases. The service may be completed even before it is protected. So the probability that the system is protected is decreased slightly. Due to the increased service rate, the number of customers in the system, average queue length and the waiting time are decreased. Changes in the number of interruptions during a service and probability that system is idle are to be noted.

Variations in the service rate μ have a direct impact on the performance of the system. As the expected service time decreases with increase in μ, the system behaves in a nice way with high values for μ as illustrated in Table 6.5

6.7 Comparison of the models described in chapter 4 and chapter 6

The difference in the models described in chapter 4 and chapter 6 is in where and when the protection is applied. In the first model the final n phases are protected in a deterministic way, irrespective of the circumstances. But in the latter, protection is given only if the service could not be completed in a reasonable time. Thus in this case, we wait for a while after the start of a new service and if it seems that the service may go long due to interruption, the protection is given to the rest of the service. Hence any number of phases may be protected. The protection can start even from the middle of a phase. In the first model we have no control over the time to complete the unprotected phases whereas the second model has a handicap of protecting unnecessary phases. The two types of protection methods have some similarity to the N-policy and T-Policy in queueing systems.

Table 4.1 and Table 6.1 can be compared to get a glimpse of the behaviour of the two systems with respect to the interruption rate θ. It may be noted that, for the 'N-Policy' model the probability that the system is protected remains constant with respect to θ. This is because we protect only the last n phases of the service process, whatever may be the value of θ. Hence the protection cost remains the same. But in the 'T-Policy' model as θ increases the chance that the system is protected increases. This means that the protection is started at an earlier phase for high values of θ compared to the N-Policy model. This increases the cost of protection. But the values of the expected service times from these tables shows that even though more protection is is provided in the T-Policy model, the N-Policy model has the lowest service time for large values of θ. In the T-Policy model to start the protection, we wait until we get an uninterrupted service for sufficiently long duration. For large values of θ it takes long to fulfil this condition as follows from the high values of $E(I)$ and $E^I(D)$. So much time is elapsed before switching the protection on. This results in a high service time. This is reflected in the values of the number of customers in the system and their waiting time. Thus the comparison of these tables reveals that when the chance of interruption is small, the T-Policy would be beneficial, but for large values of θ, the N-Policy model dominates the other.

A comparison of Table 4.2 and Table 6.2 gives us a relation between the number of protected phases in the N-Policy model and the rate of realization of the protection clock in the corresponding T-Policy model. For the choice of other parameters, it can be seen that an N-Policy model with 1 protected phase performs almost similarly to a T-Policy model with $\varphi = 2$.

The role of the service rate μ in the performance of the two systems is revealed in the comparison of Table 4.3 and Table 6.5. It can be seen that the probability P_{pr} is decreases considerably as μ increases for both the models. For the N-Policy model this decrement is due to the increase in the speed of service. But for the T-Policy model there is one more reason other than this. As the service is rendered quickly, most part of service might have completed before giving protection. So for high values of μ we are not getting the benefit of protection. This is why the expected number and duration of interruptions is more for the T-Policy model. This is reflected in the number of customers and their waiting time also. So we conclude that for large values of μ, the N-Policy is better.

The role of the repeat rate γ on the two models can be unveiled on the comparison of Table 4.4 and Table 6.3. With high repeat rate, most of the interrupted services have to be repeated. So the time to reach the protected phases is more for the N-Policy model. There fore the service time increases with increase in repeat rate. The effect of γ on the T-Policy model is the same. With a proper choice of φ both the models give almost the same performance.

On the other hand, a high repair rate δ forces most of the interrupted services to be resumed. Hence in an N-Policy model, the service process reaches the protected phases quickly lowering the service time. So the model performs better with high repair rate. See Table 4.5for numerical illustration.

But for the T-Policy model, with a high repair rate, though the repair is done very fast, the probability further interruptions denies a customer uninterrupted service for sufficient time to switch on the protection. Hence it may take long to get the service protected. As a result the service time will not be decreased in spite of high repair rate. Thus we conclude that with a high repair rate, N-Policy model is superior to the N-Policy model.

6.8 Cost analysis

In systems with interruptions, the interruptions is a cause of loss. We introduce protection to minimize the loss due to interruption. To protect the server from interruptions, additional resources might be used. This adds to the cost of service. So one must me interested to know the time to start the protection so that the cost of running the system is minimum.

we conducted a numerical experiment with the same cost function used in chapter 4, given by

$$\begin{aligned} CF = CRPT \times ERPT &+ CRSM \times ERSM \\ &+ CHOLD \times E_S(C) + CINT \times EI \\ &+ CP \times P_{pr}. \end{aligned}$$

where CRPT and CRSM are the unit time costs for repeating or resuming an interrupted service respectively. CHOLD is the holding cost per unit time per customer and CINT is the cost per unit time per interruption. Finally CP is the unit time cost for giving service in the protected phases.

The results obtained for two different costs of protection are given in Table 6.6 and Table 6.7 with the values of the parameters involved. For this analysis we took

$$\lambda = 1.5, m = 10, \mu = 5.0, \theta = 1.0, \gamma = 1.0\delta = 10.0$$

The results shows that for $CP = 600$, the optimum duration until protection is 42.5 and for$CP = 650$, the optimal value is 6.505. This assures that we can control the value of φ in favour of the system.

Table 6.1: **Effect of interruption rate.**

Service rate, $\mu = 8$, Repair rate, $\delta = 2$ Rate of realization of the repeat clock, $\gamma = 3$, Rate of realization of the protection clock, $\varphi = 15$.

θ	0	1	2	3	4	5	6
$E_S(C)$	0.3	0.4669	0.6648	0.9030	1.1953	1.5625	2.0377
$V_S(C)$	0.3267	0.8717	1.5890	2.5557	3.8962	5.8216	8.7121
$E_Q(C)$	0.05	0.1613	0.3030	0.4845	0.7196	1.0291	1.4460
$E^I(C)$	0	0.1229	0.2682	0.4426	0.6563	0.9241	1.2701
$E^B(C)$	0.3	0.3440	0.3966	0.4603	0.5390	0.6384	0.7675
$P_S(I)$	0	0.0534	0.1074	0.1620	0.2171	0.2728	0.3290
$P_S(idle)$	0.75	0.6944	0.6382	0.5815	0.5243	0.4666	0.4083
$E(I)$	0	0.0534	0.1074	0.1620	0.2171	0.2728	0.3290
$E^I(D)$	0	0.0267	0.0537	0.0810	0.1086	0.1364	0.1645
$E(\tau)$	0.1250	0.1528	0.1809	0.2092	0.2379	0.2667	0.2958
P_{pr}	0.1438	0.1454	0.1469	0.1485	0.1500	0.1515	0.1530
W_L	0.0250	0.0807	0.1515	0.2422	0.3598	0.5145	0.7230

Table 6.2: **Effect of Realization rate of protection clock.**

Service rate, $\mu = 5$, Repair rate, $\delta = 2$ Rate of realization of the repeat clock, $\gamma = 3$, Interruption rate, $\theta = 1$.

φ	0	2	5	10	15	20	10000
$E_S(C)$	1.9719	1.5227	1.1869	0.9372	0.8218	0.7575	0.5604
$V_S(C)$	7.3757	4.7936	3.1326	2.0508	1.5960	1.3557	0.6785
$E_Q(C)$	1.3424	0.9435	0.6556	0.4499	0.3580	0.3079	0.1603
$E^I(C)$	0.7972	0.5482	0.3602	0.2189	0.1529	0.1159	0.0002
$E^B(C)$	1.1747	0.9746	0.8267	0.7183	0.6688	0.6416	0.5601
$P_S(I)$	0.2098	0.1649	0.1219	0.0821	0.0606	0.0475	0.0001
$P_S(idle)$	0.3705	0.4208	0.4687	0.5127	0.5362	0.5504	0.5999
$E(I)$	0.2098	0.1649	0.1219	0.0821	0.0606	0.0475	0
$E^I(D)$	0.1049	0.0825	0.0609	0.0410	0.0303	0.0237	0
$E(\tau)$	0.3147	0.2896	0.2656	0.2437	0.2319	0.2248	0.2001
Ppr	0	0.0844	0.1656	0.2411	0.2821	0.3073	0.3998
W_L	0.6712	0.4718	0.3278	0.2249	0.1790	0.1539	0.0801

Table 6.3: **Effect of rate of realization of repeat clock.**

Service rate, $\mu = 5$, Repair rate, $\delta = 2$ Rate of realization of the protection clock, $\varphi = 15$, Interruption rate, $\theta = 1$.

γ	1	3	4	6	10	15	20
$E_S(C)$	0.8142	0.8218	0.8234	0.8253	0.827	0.828	0.8285
$V_S(C)$	1.5672	1.596	1.6014	1.6074	1.6126	1.6152	1.6166
$E_Q(C)$	0.3519	0.358	0.3592	0.3606	0.3619	0.3626	0.363
$E^I(C)$	0.1521	0.1529	0.1531	0.1533	0.1535	0.1536	0.1536
$E^B(C)$	0.6621	0.6688	0.6703	0.672	0.6736	0.6744	0.6749
$P_S(I)$	0.0604	0.0606	0.0606	0.0606	0.0606	0.0607	0.0607
$P_S(idle)$	0.5377	0.5362	0.5358	0.5353	0.5349	0.5346	0.5345
$E(I)$	0.0604	0.0606	0.0606	0.0606	0.0606	0.0607	0.0607
$E^I(D)$	0.0302	0.0303	0.0303	0.0303	0.0303	0.0303	0.0303
$E(\tau)$	0.2311	0.2319	0.2321	0.2323	0.2326	0.2327	0.2328
P_{pr}	0.2809	0.2821	0.2824	0.2828	0.2832	0.2834	0.2835
W_L	0.176	0.179	0.1796	0.1803	0.181	0.1813	0.1815

Table 6.4: **Effect of rate of repair.**

Service rate, $\mu = 5$, Rate of realization of the repeat clock, $\gamma = 3$, Rate of realization of the protection clock, $\varphi = 15$, Interruption rate, $\theta = 1$.

δ	1	3	4	6	10	15	20
$E_S(C)$	1.4027	0.7044	0.6577	0.6183	0.5918	0.5802	0.5747
$V_S(C)$	5.7409	1.0776	0.9137	0.7985	0.735	0.7112	0.7011
$E_Q(C)$	0.8775	0.2614	0.2251	0.1964	0.1785	0.1712	0.168
$E^I(C)$	0.4978	0.0837	0.0563	0.0334	0.0181	0.0114	0.0083
$E^B(C)$	0.905	0.6207	0.6014	0.5849	0.5737	0.5687	0.5664
$P_S(I)$	0.1212	0.0403	0.0302	0.0201	0.0121	0.0081	0.006
$P_S(idle)$	0.4747	0.5569	0.5674	0.578	0.5867	0.591	0.5932
$E(I)$	0.0606	0.0605	0.0605	0.0604	0.0604	0.0604	0.0604
$E^I(D)$	0.0606	0.0202	0.0151	0.0101	0.006	0.004	0.003
$E(\tau)$	0.2626	0.2215	0.2163	0.211	0.2067	0.2045	0.2034
P_{pr}	0.2828	0.2817	0.2814	0.2809	0.2805	0.2802	0.28
W_L	0.4387	0.1307	0.1126	0.0982	0.0892	0.0856	0.084

Table 6.5: **Effect of service rate.**

Repair rate, $\delta = 2$, Rate of realization of the repeat clock, $\gamma = 3$, Rate of realization of the protection clock, $\varphi = 15$, Interruption rate, $\theta = 1$.

μ	3	5	8	10	12	15	20
$E_S(C)$	2.2076	0.8218	0.4669	0.3696	0.3077	0.2472	0.1872
$V_S(C)$	6.3922	1.596	0.8717	0.6926	0.5802	0.4702	0.36
$E_Q(C)$	1.472	0.358	0.1613	0.1189	0.0946	0.0727	0.0528
$E^I(C)$	0.2368	0.1529	0.1229	0.11	0.0996	0.0871	0.0719
$E^B(C)$	1.9708	0.6688	0.344	0.2596	0.2082	0.1601	0.1153
$P_S(I)$	0.0647	0.0606	0.0534	0.049	0.0451	0.0401	0.0337
$P_S(idle)$	0.2644	0.5362	0.6944	0.7493	0.7868	0.8255	0.8656
$E(I)$	0.0647	0.0606	0.0534	0.049	0.0451	0.0401	0.0337
$E^I(D)$	0.0324	0.0303	0.0267	0.0245	0.0226	0.0201	0.0169
$E(\tau)$	0.3678	0.2319	0.1528	0.1254	0.1066	0.0873	0.0672
P_{pr}	0.5415	0.2821	0.1454	0.1037	0.0779	0.0541	0.0333
W_L	0.736	0.179	0.0807	0.0594	0.0473	0.0364	0.0264

Table 6.6: **Time to Protection versus Cost**

Cost of protection per unit time = 600

φ	$ERPT$	$ERSM$	$E_S(C)$	EI	P_{pr}	$Cost$
35.0	0.0004	0.0390	2.0303	0.0429	0.7073	648.5794
40.0	0.0003	0.0341	2.0248	0.0375	0.7126	648.5721
42.0	0.0003	0.0325	2.0229	0.0357	0.7144	648.5716
42.5	0.0003	0.0321	2.0225	0.0353	0.7148	**648.5715**
43.0	0.0003	0.0317	2.0221	0.0349	0.7152	648.5716
45.0	0.0003	0.0303	2.0205	0.0333	0.7167	648.5722
50.0	0.0003	0.0273	2.0171	0.0300	0.7201	648.5758

Table 6.7: **Time to Protection versus Cost**
Cost of protection per unit time = 600

φ	$ERPT$	$ERSM$	$E_S(C)$	EI	P_{pr}	$Cost$
6.000	0.0021	0.2111	2.2686	0.2322	0.5204	679.9899
6.300	0.0020	0.2027	2.2551	0.2230	0.5295	679.9660
6.490	0.0020	0.1977	2.2472	0.2175	0.5349	679.9617
6.500	0.0020	0.1975	2.2468	0.2172	0.5352	679.9617
6.505	0.0020	0.1973	2.2466	0.2171	0.5353	**679.9616**
6.510	0.0020	0.1972	2.2464	0.2169	0.5354	679.9617
6.520	0.0020	0.1970	2.2460	0.2167	0.5357	679.9617
7.000	0.0019	0.1853	2.2278	0.2039	0.5483	679.9832

CONCLUSION

In this thesis we discussed a few queueing models involving interruption of service and protection against interruption. Second and third chapters where on interruption without protection of service . As a consequence service of a customer has to be resumed or repeated depending on factors deciding which one to opt. Chapters 4, 5 and 6 introduced protection mechanism against interruption. The protection mechanism of chapter 4 has the flavour of N-policy where as those in chapters 5 and 6 have the flavour of T-policy.

The applications of the models discussed in this thesis are numerous, some of which are indicated in the introduction and in the relevant chapters. The results of chapters 4 and 6 are compared for efficiency.

The models discussed in this thesis can be extended to Markovian arrival process and arbitrarily distributed service time with rational Laplace Stieltjes transform. Several other variations and generalizations are on the anvil.

Bibliography

[1] H White, L Christie, Queuing with preemptive priorities or with breakdown. Operations Research, vol. 6 (1), 7995, 1958.

[2] Jaiswal N K, Preemptive resume priority queue, Operations Res., Vol. 9, No. 5, 732-742, 1961.

[3] Jaiswal N K, Time dependant solution of the head of the line priority queue, J. Roy. Stat. Soc. B24, 91-101, 1962.

[4] D Gaver Jr., A waiting line with interrupted service, including priorities. Journal of the Royal Statistical Society, vol. B24:7390, 1962.

[5] Keilson J, Queues subject to service interruptions, Ann. Math. Statistics, 33, 1314-1322, 1962.

[6] Avi-Itzhak B, Naor P, Some Queueing Problems with the Service Station subject to breakdown, Operations Research, Vol. 11, Issue 3, 303-320, 1963.

[7] K Thiruvengadam. Queuing with breakdowns. Operations Research, vol. 11(1):6271, 1963.

[8] A Federgruen, L Green, Queueing systems with service interruptions, Operations Research,vol. 34(5):752768, 1986.

[9] N Van Dijk, Simple bounds for queueing systems with breakdowns, Performance Evaluation, vol. 8(2):117128, 1988.

[10] T Takine, B Sengupta, A single server queue with service interruptions, Queueing Systems, vol. 26:285300, 1997.

[11] H Masuyama, T Takine. Stationary queue length in a FIFO single server queue with service interruptions and multiple batch Markovian arrival streams, Journal of the Operations Research Society of Japan vol. 46(3), 319341, 2003.

[12] A Krishnamoorthy, P V Ushakumari, On an M/G/1 retrial queue with disaster to the customer in service, In International Workshop on Retrial Queues, Madrid, September 1998.

[13] J Wang, B Liu and J Li, Transient analysis of an M/G/1 retrial queue subject to disasters and server failures, European Journal of Operational Research, Vol. 189, Issue 3, 1118-1132, 2008.

[14] B T Doshi, Queueing systems with vacations-a survey, Queueing Systems: Theory and Applications, Vol 1, Issue 1, pp. 29-66, 1986.

[15] H Takagi, Queueing Analysis Vol. 1. Vacation and Priority Systems, North-Holland, ISBN: 0444889108, 1991.

[16] H Takagi, K K Leung, Analysis of a discrete-time queueing system with time-limited service, Queueing Systems, Vol. 18, No. 1-2, pp. 183-197, 1994.

[17] A S Alfa, A discrete MAP/PH/1 queue with vacations and exhaustive time-limited service, Operations Research Letters, Vol. 18, Issue 1, pp. 31-40, 1995.

[18] J Li, N Tian, The M/M/1 queue with working vacations and vacation interruptions, Journal of Systems Science and Systems Engineering, Vol. 16, No.1, pp. 121-127, 2007.

[19] Boxma, M Mandjes and O Kella, On a queueing model with service interruptions, Probability in the Engineering and Informational Sciences, Vol. 22, pp 537-555, doi: 10.1017/S0269964808000326, 2008.

[20] W J Gray, P P Wang and M Scot, A vacation queueing model with service breakdowns, Applied Mathematical Modelling, Vol. 24, Issues 5-6, pp. 391-400, 2000.

[21] D Fiems, T Maertens and H Bruneel, Queueing systems with different types of server interruptions, European Journal of Opernational Research, Vol. 188, No. 3, pp. 838-845, 2008.

[22] A Krishnamoorthy, P K Pramod and T G Deepak, On a queue with interruptions and repeat or resumption of service, Nonlinear Analysis: Theory, Methods & Applications, Volume 71, Issue 12, e1673-e1683, 15 December 2009.

[23] A Krishnamoorthy and P K Pramod, Queues With Interruption and Repeat/Resumption of Service- A Survey and Further Results, Fourth International Conference on Neural, Parallel & Scientific Computations, August 11-14, MorehouseCollege, Atlanta, Georgia, USA, 2010.

[24] Haight F A , Queueing with Balking, Biometrika, Vol. 44, No. 3-4, pp.360-369, 1957.

[25] Haight F A , Queueing with Reneging, Metrika, Vol. 2, No. 1, pp.186-197, 1959.

[26] Barrer D Y, Queueing with Impatient Customers and Ordered Service, Operations Research, Vol.5, Issue 5, pp. 650-656,1957.

[27] Ancker C J Jr. and Gafarian A V, Some Queueing Problems with Balking and reneging, Operations Research, Vol. 11, No. 1, pp.88-100, 1963.

[28] Haghighi A M, Medhi J and Mohanty S G, On a Multiserver Markovian Queueing System with Balking and Reneging, Computers & Operations Research, Vol. 13, Issue 4, pp. 421-425, 1986.

[29] Wang H and Chang Y C, Cost Analysis of a Finite M/M/R Queueing System with Balking Reneging and Server Breakdowns, Mathematical Methods of Operations Research, Vol. 56, pp.169-180, 2002.

[30] Zhang Y, Yue D and Yue W, Analysis of an M/M/1/N Queue with Balking, Reneging and Server Vacations, International Symposium on OR and Its Applications, www.aporc.org/LNOR/5/ISORA2005F04.pdf, 2005.

[31] Yue D , Yue W and Sun Y, Analysis of an M/M/c/N Queue with Balking Reneging and Server Vacations, Proc. Sixth International Symposium on Operations Research and Its Applications, pp.128-143, 2006.

[32] Altman E and Yechiali U, Analysis of Customers Impatience in Queues with Server Vacations, Queueing Systems, Vol. 52, Issue 4, pp. 261-279, 2006.

[33] Altman E and Yechiali U, Infinite Server Queues with Systems Additional Tasks and Impatient Customers, Probability in Engineering and Informational Sciences, Vol. 22, Issue 4, pp.477-493, 2008.

[34] Baruah M, Madan K C and Eldabi T, A Two Stage Batch Arrival Queue with Reneging During Vacation and Breakdown Periods, American Journal of Operations Research, Vol. 3, No. 6, DOI:10.4236/ajor.2013.36054, 2013.

[35] S. Karlin and H. J. Taylor, A First Course in Stochastic Processes. Academic Press, 1975.

[36] G I Falin and J G C Templeton , Retrial Queues, Chapman and Hall, London (1997).

[37] Artalejo J R, New Results in Retrial Queueing Systems with Breakdown of the Servers, Statistica Neerlandica, Vol. 48, Issue 1, 1994, pp.23-36

[38] V G Kulkarni and B D Choi, Retrial queues with server subject to breakdowns and repairs, Queueing Systems, Vol. 7, No.2, 191-208, 1990.

[39] Aissani A and Artalejo J R, On the single server retrial queue subject to breakdowns, Queueing Systems Vol. 30 (3-4), 309-321, 1998.

[40] Artalejo J R and Gomez-Corral A, Unreliable retrial queues due to service interruptions arising from facsimile networks. Belg. J. Oper. Res. Statist. Comput. Sci. Vol. 38 (1), 31-41, 1998.

[41] Wang J, Cao J and Li Q L, Reliability analysis of the retrial queue with server breakdowns and repairs. Queueing Systems, 38 (4), 363-380, 2001.

[42] Artalejo J R and Gmez-Corral A , Retrial Queueing Systems: A Computational Approach, Springer-Verlag, Berlin (2008).

[43] B Almasi, J Roszik and J Sztrik, Homogeneous Finite-Source Retrial Queues with Server Subject to Breakdowns and Repairs, Mathematical and Computer Modelling, Vol. 42, Issues 5-6, pp. 673-682 , 2006.

[44] B Almasi and J Roszik and J Sztrik, Heterogeneous Finite-Source Retrial Queues with Server Subject to Breakdowns

and Repairs, Journal of Mathematical Sciences, Vol. 132, No. 5, 2006, pp. 677-685.

[45] Wang J, Zhao L and Zhang F, Performance Analysis of the Finite Source Retrial Queue with Server Breakdowns and Repairs, Proceedings of the Fifth International Conference on Queueing Theory and Network Applications, pp. 169-176, 2010.

[46] Chen P, Zhu Y and Zhang Y, A Retrial Queue with Modified Vacations and Server Breakdowns, Computer Science and Information Technology (ICCSIT), 3rd International Conference on, pp. 26-30, DOI: 10.1109/ICCSIT.2010.5563560, 2010

[47] Klimenok V, Kim C S and Kuznetsov A, A Multi-Server Queue with Negative Customers and Partial Protection of Service, Proceedings of the 13th International Conference on Analytical and Stochastic Modelling Techniques and Applications (ASMTA06), Bonn, Germany, Eds. K Al-Begain, pp 143-148, 2006.

[48] Klimenok V and Dudin A N, A BMAP/PH/N Queue with Negative Customers and Partial Protection of Service, Communications in Statistics-Simulation and Computation, Vol. 41, Issue 7, pp. 1062-1082, 2012.

[49] M F Neuts, Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach, Second Edition, Dover Publications, Inc., New York, 1994.

[50] Latouche G and Ramaswami V, Introduction to Matrix Analytic Methods in Stochastic Modeling, SIAM, 1999.

[51] M F Neuts and B M Rao, Numerical investigation of a multiserver retrial model. Queueing systems, 7:169-190,1990.

[52] L Bright and P G Taylor, Equilibrium distribution for level-dependent quasi-birth-and-death processes, Comm. Stat. Stochastic Models, 11:497-525, 1995.

[53] J Medhi, Stochastic Processes, Second Edition, New Age International Publishers, New Delhi, 1982.