STOCHASTIC MODELLING AND APPLICATIONS

ON QUEUES WITH INTERRUPTIONS AND REPEAT OR RESUMPTION OF SERVICE

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

PRAMOD P.K

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

January 2010

Certificate

This is to certify that the thesis entitled '**On Queues with Interruptions and Repeat or Resumption of Service**' submitted to the Cochin University of Science and Technology by Mr. Pramod P.K for the award of the degree of Doctor of Philosophy under the Faculty of Science, is a bonafide 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
(Supervisor)
Professor(Retd.)
Department of Mathematics
Cochin University of Science and Technology
Cochin- 682022, Kerala.

Cochin-22
04/01/2010.

Declaration

I, PRAMOD.P.K hereby declare that this thesis entitled '**On Queues with Interruptions and Repeat or Resumption of Service**' 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.

Pramod.P.K
Research Scholar
Registration No.2953
Department of Mathematics
Cochin University of Science and Technology
Cochin-682022, Kerala.

Cochin-22
04/01/2010.

To

My

Parents and Teachers

Acknowledgement

I would like to express my gratitude to all those who helped me to complete this thesis. My special thanks to parents, friends , relatives and teachers.

The greatest thanks I owe is to my supervisor and guide Dr. A. Krishnamoorthy, Professor (Retd), Department of Mathematics, Cochin University of Science and Technology for giving opportunity to do research under his guidance. His excellent guidance and timely help made me this thesis a great success.

I wish to express my sincere thanks to Dr. M.N. Naryanan Namboothiri, Department of Mathematics, CUSAT, Dr. Rammohan, Dean, Marine Sciences, CUSAT, Dr. A.V. Alex, Department of Physics, St.Paul's College, Kalamassery, Mr. Sasi Gopalan, School of Engineering, CUSAT for their love and affection which guide me to start my doctoral work. I am also grateful to Dr. R. S. Chakravarti, Head, Department of Mathematics, CUSAT and other faculty members of the Department: Dr.B. Lakshmi, Dr. M. Jathavedan, Dr. M.N. Narayanan Namboothiri, Dr. A. Vijayakumar, Ms. Meena.K for their helps during my research work. I am also grateful to office staff: Ms.Indukumari.S, Mr. Devassy V.K, Ms.Vinitha Rani C.S, Ms.Sree Vidhya.S, Mr.Ramesh.M.N, Ms. Ratna Bai.S, Ms.Priya.A.S, Mr.Jaya shankar, Ms. Shefy, Ms.Bindu.S, Ms.Omana P.P, Mr.Sakti C.S, librarian Shajitha. C and former librarian Sailaja.V.K of the Department of Mathematics for their help during my research period. My gratitude also goes to the authorities of Cochin University of Science and Technology for the facilities provided during my research work.

I acknowledge with gratitude Dr. S.R. Chakravarthi, Department of Industrial and Manufacturing Engineering, Kettering University, Flint, USA and Dr. T.G. Deepak, Department of Mathematics, Indian Institute of Space Science and Technology, Trivandrum for permitting to include our joint work in this thesis.

I express my feeling of gratitude to friends in my research area Dr. Viswanath C Narayanan, Dr. Babu.S, Dr. K.P. Jose, Mr. Ajayakumar C.B, Ms. Deepthi.C.P, Ms. Lalitha .K, Mr. Manikandan, Ms. Resmi.T, Mr. Satyan.M.K, Mr. Sreenivasan.C, Mr. Varghese Jacob, Mr. Sajeev.S.Nair, Mr. Gopakumar.G, Ms. Treesa Mary Chacko, for their interest in my research work and they were always ready to share their bright ideas with me. I want to express my warmest thanks to my fellow research scholars Dr. G. Indulal, Dr. Aparna Lakshman, Mr.Pravas.K , Ms. Chithra M. R., Ms. Manju K Menon, , Ms. Lexy Alexander, Dr. Rajeswaridevi M. B., Mr. Santhosh Kumar Pandey, Mr. Shinoj K. M., Ms.Shiny Philip, Mr. Kirankumar, Mr. Didimos.K.V, Ms. Seema Varghese, Mr. Tonny K. B., Mr.Gireesan.K.K and Ms. Viji M. for their interest in my work. I also express gratitude to all my friends who gave moral support during this Doctoral work.

ON
QUEUES WITH INTERRUPTIONS AND REPEAT OR RESUMPTION OF SERVICE

Contents

List of Figures

List of Acronyme

$CTMC$	-	Continuous time Markov Chain.
$Diag$	-	Diagonal matrix.
ER	-	Erlang distribution.
EX	-	Exponential distribution.
$Exp(.)$	-	Exponential distribution with parameter (.)
$FIFO$	-	First in First out.
IP	-	Interruption Clock.
$LIQBD$	-	Level independent Quasi death Process.
MAP	-	Markovian Arrival Process.
MC	-	Markov Chain
$MMAP$	-	Marked Markovian Arrival Process.
PH	-	Phase type distribution.
SC	-	Super Clock.
TC	-	Threshold Clock.

List of symbols

$\bigotimes$	-	Kronecker product.
$A \bigoplus B$	-	Kronecker sum of matrices A and B. ie, $A \bigotimes I + I \bigotimes B$.
e	-	The column vector of dimension r consisting of 1's.
$\underline{e}$	-	column vector of 1's of appropriate order.
$e_j(r)$	-	Column vector of dimension r with 1 in the j^{th} place and zero elsewhere.
I_r	-	Identity matrix of dimension r.
$m \times n$	-	m by n.
E_s	-	Expected time of service completion.
$\mu'_{\tilde{T}}$	-	Mean of $PH(\alpha, \tilde{T})$.
$\sigma'_{\tilde{T}}$	-	Variance of $PH(\alpha, \tilde{T})$.
$'$	-	Denote transpose of a matrix.
R	-	Denote rate matrix.
G	-	Stochastic matrix.
$*$	-	Convolution .
$diag[A_1, A_2]$	-	Denote a diagonal matrix with diagonal elements A_1, A_2.
$diag[F^{(i,i-1)}$	-	A diagonal matrix whose i^{th} diagonal element is $F^{(i,i-1)}$.

Chapter 1

Introduction

Since the last century there have been marked changes in the approach to scientific enquires. There has been greater realization that probability (or non-deterministic) models are more realistic than deterministic models in many situations. Observations taken at different time points rather than those taken at a fixed period of time began to engage the attention of probabilists. This led to new concept of indeterminism in dynamic studies. The period of dynamic indeterminism began roughly with the work of Mendel (1822-1884). The physicists like Chandrasekhar (1943) played a leading role in the development of dynamic indeterminism. Many such phenomenon occurring in physical and life sciences are studied now not only as a random phenomenon but also as one changing with time or space. Similar considerations are also made in other areas, such as, social sciences, engineering and management and so on. The scope of applications of random variables which are functions of time or space or both has been on the increase. This thesis is concerned with stochastic modelling and analysis of real life problems. We confine to problems in queues, though the motivation comes from diverse areas such as Engineering, medicine. Most of the time our process turns out to be a continuous time Markov chain and in one case a discrete time Markov chain. Hence we start with some basic results in Markov chains.

1.1. Markov Chain

Definition 1.1.1. A Markov process is a stochastic process with the property that, given the value of X_t, the values of $X_s, s > t$, do not depend on the values of $X_u, u < t$. If the time is discrete the Markov process is called discrete time Markov chain; otherwise continuous time Markov chain. In this thesis we encounter both.

discrete and continuous time Markov chains.

Theorem 1.1.1. *If a Markov Chain is irreducible and positive recurrent, there exists a unique solution to the linear system $\pi P = \pi$, $\pi e = 1$. If, moreover, the chain is aperiodic, the probabilities $P[X_t = i]$ will converge to π_i as $n \to \infty$*

In this thesis we will demonstrate how certain queueing problems can be modelled as Markov Chain. Markov chain have wide range of applications in different areas of science.

1.2. Queueing Theory

The first work on waiting line (queue) was 'The theory of probabilities and telephone conversations ' by A.K. Erlang [9] who published this paper in 1909 . This was devoted to the study of telephone traffic congestion. Random fluctuations in customer arrival and service processes play a pivotal role here. Queueing Theory is mainly seen as a branch of applied probability theory. Applications of queueing theory in different fields include communication networks, computer systems, machine plants and so forth. These are concerned with the design and planning of service facilities to meet randomly fluctuating demands for service so that congestion is minimized and the economic balance between the cost of service and the cost associated with waiting for that service is maintained. A queueing system consists of customers arriving at random time to some facility where they receive service and then depart. When the service system is available, a certain service discipline decides which customer will be served next. This customer then moves to service facility and depart the queueing system after getting the service.

Queueing systems are classified according to the input process, the service time distribution, the size of buffers, the number of servers and the scheduling discipline. Kendall notation to describe the queueing system and its $a/b/c/d/e/f$ here a-denotes the arrival process, b-service process, c-the number of servers, d-size of queue, e-service discipline, f-size of the source of arrival. Following are some of the distributions used in this thesis.

Definition 1.2.1. Continuous Time Phase Type Distribution
Consider a finite Markov chain(MC) with 'a' transient states and one absorbing state with the transition matrix P partitioned as $P = \begin{bmatrix} T & T^0 \\ \bar{0} & 0 \end{bmatrix}$ where T is a matrix of order a and T^0 is a column vector such that $Te + T^0 = 0$ where e is a column vector of 1's. For eventual absorbtion in to the absorbing state, starting from

any initial state, it is necessary and sufficient that T is non singular. Suppose that the initial state of the MC is chosen according to the probability vector (α, α_{a+1}), with $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_a\}$. Let X denote the time until absorbtion. Then X is a continuous time random variable taking non negative real values with probability distribution function $F(x)$ given by $F(x) = 1 - \alpha e^{Tx}e$, for $x \geqslant 0$.The probability function so constructed is a continuous PH-distribution and (α, T) of order a is a representation. The k^{th} factorial moment is given by $\mu_k' = k!\alpha T^{k-1}(I - T)^{-1}e$, for $k \geqslant 0$

Definition 1.2.2. Discrete Time Phase Type Distribution
Consider a finite Markov chain(MC) with 'a' transient states and one absorbing state with the transition probability matrix P partitioned as $P = \begin{bmatrix} T & T^0 \\ \bar{0} & 1 \end{bmatrix}$ where T is a matrix of order a and T^0 is a column vector such that $Te + T^0 = e$. It is necessary and sufficient that $(I - T)$ is non-singular for eventual absorption in to the absorbing state, staring from any initial state. Suppose that the initial state of the MC is chosen according to the probability vector (α, α_{a+1}). The absorption time X is then a random variable taking non negative integer with probability function (a_k) given by $a_0 = \alpha_{a+1}$ and $a_k = \alpha T^{k-1}T^0$, for $k \geqslant 1$, $k = 1, 2, ..., a$. The probability function so constructed is a discrete PH-distribution represented by (α, T) of order a is a representation; the order of T is called the order of the representation. The k^{th} factorial moment of (a_k) is given by $\mu_k' = k!\alpha T^{k-1}(I - T)^{-1}e$, for $k \geqslant 0$

Definition 1.2.3. Poisson Process
A stochastic process $\{N(t), t \geqslant 0\}$ is said to be a counting process if $N(t)$ represents the total number of 'events' that occur by time t. The counting process $\{N(t), t \geqslant 0\}$ is said to be a Poisson process having rate λ, $\lambda > 0$, if (i) N(0)=0, (ii) the process has independent increments (iii) the number of events in any interval of length t is poisson distributed with mean λt. The Poisson process and distribution arising out of it play a pivotal role in queueing theory.

Definition 1.2.4. Exponential Distribution
Suppose we have a poisson distribution with rate of change λ, the distribution of waiting between successive changes is

$$\begin{aligned} F(x) &= P(X \leqslant x) \\ &= 1 - P(X > x) \\ &= 1 - e^{-\lambda x} \end{aligned}$$

and its density function is $f(x) = \lambda e^{-\lambda x}$. The exponential distribution is the only continuous distribution having memoryless property.

Definition 1.2.5. Erlang Distribution
An Erlang distribution E_n^{λ} with n stages and parameter λ is the distribution of the sum of n independent exponential random variables with parameter λ. It has density function given by $f(t) = \frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}$ for all $t \geqslant 0$; $\lambda > 0$. Its interpretation

as a succession of n exponentially distributions with rate λ. An Erlang distribution can be represented as the holding time in the transient state set $\{1,2,....,n\}$ of a Markov chain with absorbing state n+1 where the only possible transitions occur from a state k to the next state k+1 (for k=1,2,...,n), with rate λ each. Thus an Erlang distribution is PH-distribution with representation (α, T) where

$$\alpha = (1,0,0,....0)\ ,\ T = \begin{bmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & . & & \\ & & & -\lambda & \lambda \\ & & & & -\lambda \end{bmatrix} \text{ and } T^0 = \begin{pmatrix} 0 & 0 & & . & \lambda \end{pmatrix}'$$

Definition 1.2.6. Hyper Exponential Distribution
A hyper-exponential distribution is a finite mixture of $n(\in N)$ exponential distributions with different parameters $\lambda_k(k = 1,2,...,n)$. Its density function is given as $f(t) = \sum_{k=1}^{n} q_k\lambda_k e^{-\lambda_k t}$ with proportions $q_k > 0$ satisfying $\sum_{k=1}^{n} q_k = 1$. This leads to a PH-representation as $\alpha = (\pi_1, \pi_2, ., ..., \pi_n)$, $T = \begin{bmatrix} -\lambda_1 & & & \\ & -\lambda_2 & & \\ & & & \\ & & & -\lambda_n \end{bmatrix}$

and $T^0 = \begin{pmatrix} \lambda_1 & \lambda_2 & . & . & \lambda_n \end{pmatrix}'$

Definition 1.2.7. Geometric Distribution
The geometric distribution is a discrete distribution for x=0,1,2.....having probability density function $f(x) = p(1-p)^x = pq^x$ where $0 < p < 1$ and $q = 1-p$ and distribution function is $F(x) = \sum_{k=0}^{x} f(k) == 1 - q^{n+1}$. The geometric distribution is the only discrete distribution having memoryless property and it is the discrete analog of the exponential distribution.

Definition 1.2.8. Batch Markovian Arrival Process
Consider a two dimensional Markov Process $X(t) = \{N(t), J(t) : t \geqslant 0\}$ on the state space $\{(i,j) : i \geqslant 0, 1 \leqslant j \leqslant m\}$ with infinitesimal generator given by

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

where $D_k, k \geqslant 0$, are $m \times m$ matrices.; D_0 has diagonal elements negative and nonnegative off-diagonal elements; D_k for $k \geqslant 1$ are nonnegative and the matrix D given by $D = \sum_{k=0}^{\infty} D_k$ is an irreducible infinitesimal generator of a continuous time Markov chain. The variable $N(t)$ denotes the number of arrivals in $(0,t]$, and the variable $J(t)$ denotes phase of the arrival process. The transition from a state (i,j) to a state $(i+k,l)$ where $k \geqslant 1$, $1 \leqslant j,l \leqslant m$

in with transition rates governed by the matrix D_k, correspond to the arrival of a batch of size k, while a transition from a state (i,j) to a state $(i,1)$, $1 \leqslant j$, $l \leqslant m$; $j \neq 1$, with transition rates governed by the matrix Do, correspond to no arrival. Thus the matrix D_0 governs transitions that correspond to no arrival and the matrix D_k governs transitions corresponding to a batch arrival of size k, $k \geqslant 1$. We assume that the matrix D_0 is a stable matrix (see Bellman [8]) which makes it non-singular which in turn ensures that the sojourn time in the set of states $\{(i,j) : 1 \leqslant j \leqslant m\}$ is finite with probability 1 for all i. This ensures that the arrival process $X(t)$ never terminates. Let π be the stationary probability vector of the Markov process with generator D. The fundamental arrival rate is then given by $\delta = \pi \left(\sum_{k=0}^{\infty} kD_k \right) e$. For more details on BMAPs we refer to Lucantoni [23]. An excellent survey of BMAP is available in Chakravarthy(2006)[6].

Definition 1.2.9. Markovian Arrival Process

A Markovian Arrival Process (MAP) is a particular case of BMAP where maximum possible batch size is 1, that is, we take $D_k = 0$, for $k \geqslant 2$, so that in this case $D = D_0 + D_1$. This is not the construction of MAP. A construction of MAP with representation matrices (D_0, D_1) of order m is as follows: Consider a Markov process with state space $\{1, 2, ..., m, m+1\}$ with infinitesimal generator $D = \begin{bmatrix} D_0 & d \\ 0 & 0 \end{bmatrix}$ where D_0 is an $m \times m$ matrix, $D_0 e + d = 0$ and $m+1$ is an absorbing state. Since by assumption D_0 is a stable nonsingular matrix, absorption occurs with probability 1 from any initial state. As in the construction of PH-renewal process, when absorption occurs we assume that an arrival has occurred and we immediately restart the process using an initial probability vector. But different from PH-renewal process here this initial probability vector depends also on the state from which absorption occurred and this brings dependence between inter arrival times. Let $\alpha_i \neq 0$, where α_i is an m-dimensional row vector with $\alpha_i e = 1$, be the probability vector which we use to restart the process after absorption has occurred from the state i and define the $m \times m$ matrix D_1 by $(D_1)_{ij}$, $(d_i)(\alpha_i)_j, 1 \leqslant i, j \leqslant m$. Now the matrix $D = D_0 + D_1$ will be the generator matrix of a Markov process $\{Y(t) : t \geqslant 0\}$ on the state space $\{1, 2, ..., m\}$. Let $N(t)$ denotes the number of arrivals in $(0, t]$. Then the 2-dimensional Markov Process $\{(N(t), Y(t)) : t \geqslant 0\}$ with state space $\{(i,j) : i \leqslant 0, 1 \leqslant j \leqslant m\}$ is the arrival process which we constructed above and is called Markovian Arrival Process. The infinitesimal generator of the process is given by $Q = \begin{bmatrix} D_0 & D_1 & & \\ & D_0 & D_1 & \\ & & D_0 & D_1 \\ & & & \ddots \end{bmatrix}$ For more details on MAPs refer to Lucantoni [23], Chakravarthy [6].

1.3. Matrix analytic methods

Queueing systems such as $M/M/1$, $M/M/\infty$, $G/G/1$ etc. are well studied and are well tractable, using the methods of generating functions and Laplace transform methods. However there are increasingly many queueing problems that turn out to be analytically intractable. At best one may get the Laplace transforms of some quantities of interest. Nevertheless, most often these turn out to be difficult to invert if not impossible, thereby rendering the evaluation of system performance measures inaccessible. It is to overcome this difficulty that Neuts (1979)[25] devised efficient algorithms through the introduction of matrix analytic methods. PH distributions have the advantage that an arbitrary distribution with rational Laplace Steiltjes transform can be approximated by the former. Since PH distribution is numerically tractable, one can do quite a bite of manipulations for the distribution having this properly mentioned above. The two books : Matrix Geometric Metholds in Stochastic Process. An Algorithmic Approach (John Hopkins, 1988) and Matrices of M/G/1 type (1991) by Neuts [26] and also by Latouche & Ramaswami [21] make excellent reading. The modelling tools such as Phase type distributions, Markovian Arrival Processes, Batch Markovian Arrival Processes, Markovian Service Processes etc. are well suited for Matrix Analytic Methods. Below we give a brief description of Matrix Analytic Methods applied for solving quasi-birth-and-death processes.

Definition 1.3.1. Level independent quasi-birth-and-death processes
A level independent quasi-birth and death process is a Markov process with state space $\Delta = \{(0,j) : 1 \leqslant j \leqslant n\} \cup \{(i,j); i \geqslant 1, 1 \leqslant j \leqslant m\}$ and with infinitesimal generator Q given by $Q = \begin{bmatrix} C_0 & C_1 & & & \\ C_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \cdot & \cdot & \cdot \\ & & & & \cdot \end{bmatrix}$

The generator Q is obtained in the above form by partitioning the state space E into the set of levels $\underline{0}, \underline{1}, \underline{2}, \ldots$ where $\underline{0} = (0,j) : 1 \leqslant j \leqslant n, \underline{i} = (i,j) : 1 \leqslant j \leqslant n$ for $i \geqslant 1$. The vector $\underline{i}$ is called i^{th} level. C_0 is a square matrix of order $n \times n$ and denotes transition rates from states of level 0 to the states of level 0 itself. C_1 is a matrix of order $n \times m$ and denotes transition rates from level 0 to level 1. The $m \times n$ matrix C_2 denotes transition rates from level 1 to level 0 . A_2, A_1, A_0 are square matrices of order m and denotes transition rates from level i to levels i - 1, i, i + 1 respectively. Assuming that Q is irreducible , we have the following theorem (see Neuts [25]).

Theorem 1.3.1. *The process Q is positive recurrent if and only if, the minimal* *non negative solution R to the matrix quadratic equation*

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

has spectral radius less than 1 and the finite system of equations

$$\begin{aligned} x_0C_0 + x_1C_2 &= 0, \\ x_0C_1 + x_1(A_1 + RA_2) &= 0, \\ x_0e + x_1(I - R)^{-1}e &= 1 \end{aligned} \tag{1.2}$$

has a unique positive solution for x_0, and, x_1. If the matrix $A = A_0 + A_1 + A_2$ 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 generator matrix A.

The stationary probability vector $\mathbf{x} = (x_0, x_1, x_2, ...)$ of Q is given by

$$x_i = x_1R^i, for\ i \geqslant 1. \tag{1.3}$$

To find the minimal solution of 1.2 one can use the iterative formulas (see Neuts [25]):

$$R = -Ao(A_1 + R_{n-1}A_2)^{-1} for\ n \geqslant 1 \tag{1.4}$$

with an initial value R_0, which converges to R if $sp(R) < 1$. An accuracy check for R is given by the equation $RA_2e = A_0e$. Also the above relation 1.4 shows that if any row of A_0 is a row consisting of zeroes only, then the corresponding row of R, also consists of zeros only. So if our A_0 matrix has a special structure, it can be exploited in the evaluation of the R matrix. Another method to find R is to use the relation

$$R = A_0(-A_1 - A_0G)^{-1} \tag{1.5}$$

where the matrix G is the minimal nonnegative solution of the matrix quadratic equation

$$A_2 + A_1G + A_0G^2 = 0. \tag{1.6}$$

The matrix G will be stochastic if $sp(R) < 1$. When $sp(R) < 1$, the Logarithmic Reduction Algorithm due to Ramaswamy [21] (see Latouche and Ramaswamy [21]), which is quadratically convergent, can be used to calculate the G matrix and hence the R matrix using relation 1.5. When G is stochastic, from 1.6 we obtain the relation

$$G = (-A_1 - A_0G)^{-1}A_2 \tag{1.7}$$

which shows that if any column of the A_2 matrix is zero then the corresponding column of the G matrix is also zero. Therefore if the A_2 matrix has a special structure, it can be exploited in the calculation of the G matrix. Also one can efficiently use (Block) Gauss- Seidel iteration method to evaluate the G matrix, particularly, if the matrix A_2 has a special structure. For further details on Matrix Analytic Methods for Level independent QBD's we refer to Neuts [25], Latouche and Ramaswami [21].

1.4. Literature Survey

It may be noted that vacation is a sort of interruption, there is a distinction that usually vacation is associated with the advent of an event whenever server becomes free after serving a few customers/ a certain customers are served continuously. Thus the server goes on vacation at the end of a service. However interruption to service takes place while a service is going on. This can be due to server breakdown or due to a high priority customer arrival and consequent pre-emption of the customer in service. It is the better form of interruption that we dwelve on this thesis. We give below a brief survey of the work reported in queues with interruption. This accuracy is in no way exhausting.

(1) White and Christie (1958)[36] considered two queues (priority I & II) served by a single server, with the lower priority (II) customer being preempted on arrival of a high priority customer. Service times of both type of customers are independent exponentially distributed with distinct parameters. Nevertheless the assumption that service times are exponentially distributed does not help us in distinguishing whether an interrupted service is to be repeated or resumed. They contrast this model with head of line priority. They had shown how service facility breakdown could be considered equivalent to arrivals of items with preemptive priority.

(2) Gaver (1962) [14] discussed a queueing problem with interruption as detailed under: On completion of an interruption either the service is repeated or resumed. There is no specific rule that determines whether the service is to be resumed or repeated. The completion time distribution function for postponable interruption (the interruption starts not at the epoch of its onset; rather only on completion of the present service the interruption takes effect. The duration of the interruption is the cumulated effect of all interruptions that got postponed). He computed the distributions of the completion time of the job in the three cases of repetition, resumption and postponable interruption, in terms of Laplace transforms. With the arrival process assumed to be poisson and service times arbitrarily distributed, with duration of interruption also arbitrarily distributed, Gaver obtained the Laplace transform of the service completion time random variable.

(3) Keilson (1962)[17] considered M/G/1 queue with interruptions of Poisson incidence occasioned either by server break down or the arrival of customers with higher priority. Interruption times and priority service times have arbitrary distribution. After preemptive interruption, ordinary service is either repeated or resumed. The time dependent behavior of the system was discussed in a complete state space and the join density in all system variables of this space is constructed systematically from the densities associated with

a set of simpler first-passage problems. He also obtained equilibrium distributions as limiting forms and server busy period distribution computed. Nevertheless Keilson did not devise a method to distinguish between repeat and resumption of an interrupted service.

(4) Ibe and Trivedi (1990) [35] considered a queue with two stations, that are served by a single server in a cyclic manner. They assumed that at most one customer can be served at a station when the server arrives at the station. The server is subject to breakdown and hence a repair time is associated with such events. They obtained appropriate mean delay of customers in the system. Numerical results were obtained to get a closer view of the performance measures.

(5) Nunez-Queija (2000)[27] considered the sojourn times of customers in an $M/M/1$ queue with the processor sharing service discipline and a server which is subject to breakdown. The duration of the breakdown have a general distribution, whereas the on-periods are exponentially distributed. A branching process approach leads to a decomposition of the sojourn time, in which the components are independent of each other and could be investigated separately. He derived the LaplaceStieltjes transform of the sojourn-time distribution in steady state, and showed that the expected sojourn time is not proportional to the service requirement. In the heavy-traffic limit, the sojourn time, conditioned on the service requirement and scaled by the traffic load, was shown to be exponentially distributed. These results could be used for the performance analysis of elastic traffic in communication networks, in particular, the ABR service class in ATM networks, and best-effort services in IP networks.

(6) Fiems et.al, EJOR (2008)[11] provides specific probability for repeat/resumption of an interrupted service. Specifically, they assumed that an interruption would be destructive (the authors call it disruptive, which is a wrong terminology) with probability p, and so the interrupted service has to be repeated, or with probability $1-p$ it is non destructive and so has to be resumed on removal of interruption. With arrival process forming a poisson process and service times arbitrarily distributed they set up the equation to determine the effective service time of a customer. Closed form expressions for various performance measures were obtained. First the stability of the system was investigated. Using a transform approach, they obtained various performance measures such as the moments of the queue content and waiting times. They illustrate their approach by means of some numerical examples.

(7) Tewfik Kernane(2009) [34] extended the work of Fiems et.al.(2008)[11] to queues with repeated trial (retrial)of customers. He proved that all the results obtained in the latter could be translated to the retrial set up.

(8) Takine and Sengupta (1997)[33] considered a single server, multi-class service system. Arrival of customers according to MAP. Service times arbitrarily distributed. At times server would not be available for service to customers of priority below a given level. They characterize the queue length distribution as well as the waiting time distribution. The computational feasibility of the model is highlighted through numerical procedures.

(9) Atencia and Moreno (2006)[2] discuss a discrete time Geo/G/1 retrial queue with server subject to starting interruption. That is, at the instance of commencement of processing a new job, the server may breakdown. Associated with this there is a repair time, after which the service commences. They obtain the stationary distribution of the system state and then compute a few useful performance measures. They also obtain two stochastic decomposition laws and find a measure of the proximity between the system size distribution of the model and corresponding model without retrials. Further they showed that M/G/1 retrial queue with starting failures can be approximated by its discrete time counterpart.

(10) Atencia and Moreno (2008)[3] consider an $M^{[X]}/G/1$ retrial queue in which customer arrival constitutes compound poisson process. Service times are arbitrarily distributed. There is no waiting space for customers and so, if the server is busy at an arrival epoch, such customers are directed to an orbit of infinite capacity from where they retry to access the server according to an exponentially distributed time. In case the server is idle at an arrival epoch, then one in the arriving group proceeds for service and the rest, if any, to the orbit. Server is subject to failure during service. The customer in service then stays back. The repair time is arbitrarily distributed. The server 'on time'is exponentially distributed. The service that got interrupted get resumed, on repair of server. They obtain long run behavior of the system .

(11) Lin Li, Ying and Zhao (2006)[29] consider a queue in a more general set up than that of Atencia and Moreno (2008)[3], namely the $BMAP/G/1$ Retrial queue with server breakdown and repairs. Here again service times have arbitrary distribution; repair time of server is also arbitrarily distributed. The server 'on time' has exponential distribution (which does not change when the server is idle). Here again resumption of service, on repair of the server, is assumed. Using supplementary variable technique the authors analyze the system. The R-G factorization of the level dependent CTMC of the $M/G/1$ type is used to provide the stationary probability measures.

(12) Gursoy and Xiao (2004)[15] discuss an infinite server queue with Poisson arrival and exponentially distributed service times. Interruption occurs to the system according to a poisson process. When interrupted, the service rate of each server is less than the normal service rate. The time to get back

to normal state is exponentially distributed. Under these assumptions the authors obtain a stochastic decomposition for the number of customers in the system; they prove that one component in the decomposition is precisely the number of customers present in the classical $M/M/\infty$ queue.

- In most of the work reported on interruption, either the service of the interrupted customer is repeated on removal of interruption or it is resumed. An exact rule to determine whether to resume or repeat service on completion of interruption is missing. In most of these Krishnamoorthy et.al.(2009)[19] is also provides specific rules for repetition / resumption of service. However Fiems.et.al (2008)[11] and its extension to retrial set up by Tewfik Kernane (2009)[34] specifically identify the rule to decide whether service is to be repeated or resumed. Nevertheless, this is done at the onset of interruption which may not be the correct decision rule in most situations. To rectify this we have, in this thesis, brought in the rules concerning repetitions/resumption of service to apply immediately after the completion of an interruption. Specifically a random clock (threshold clock) starts ticking the moment interruption starts. At this point a competition between the interruption clock and random clock begins. Whichever stops ticking first determines whether to resume or repeat the service of the interrupted customer. To be more specific, we assume that the interrupted service is resumed if interruption clock realizes before threshold clock and it is repeated otherwise. This is the rule that we follow throughout this thesis.

Another important aspect to be mentioned at this point is that, while on interruption the server does not get affected by the arrival of further interruptions which means that the interruption behaves like a Type I counter (see for example Karlin and Taylor (1975)[20]). All the work reported on interruptions essentially follow this rule except, perhaps that of Gaver [14] where postponable interruption is treated separately. However postponable interruption has the defect that the effective service time of a customer is the actual service time. This so because the interruptions that occurred during a service could be all pooled together and passed on the server at the service completion epoch. Type II counter like interruption is being investigated.

In addition to the above mentioned work there are a few others that are reported on queues with service interruptions. These include Guodong Pang and Ward Whitt(2009)[28], Haridass and Arumuganathan (2008)[16], Boxma et.al.(2008)[5], Chan et.al.(1993)[7], Gursoy and Xiao (2004)[15], Rembowski (1985)[30], Li and Zhao(2004)[29], in continuous time, and in discrete time, Fiems et.al.(2002)[12], Alfa (2002)[1].

Li and Tian (2007)[22] examined a queue with working vacations and vacation interruptions. Precisely, he assumed that a vacation could be interrupted by the arrival of a high priority customer. In the steady state they proved that the stationary queue length can be decomposed into the sum of two independent random variables.

1.5. About the thesis

The theory of stochastic service systems is mainly concerned with services which are subject to interruptions. The most natural cause of such interruptions is server breakdowns. However interruption can also occur to the system when the server preempts the current customer to serve a higher priority customer. Since 1958 several researchers have analyzed different types of interruptions. Most of these consider either repetition or resumption of the interrupted service on removal of interruption. Also there are some work dealing with both repeat and resumption of service, but predetermined with some probability. The work reported in this thesis is a generalization in different directions of the work done by most of these authors.

An example related to the model discussed in this thesis can be described as follows: In a typical client-server interaction over the internet, clients issue requests for applications such as real-time voice or video download, large file transfer via file transfer protocol or interactive sessions with a server. For example, a request from a client can be made for a video streaming from a streaming server. Such video transfers require sustained availability of service from the server. Client requests line up at the server at peak hours of operation. Insistent requests keep the server fully loaded and on the verge of breakdown. The server can go to a repair state and can come back after a stipulated amount of time. The server resumes serving clients requests from the point where it was dropped or repeat the service of the client depending upon the amount of time taken for repair.

Another example is administration of antibiotics. In general, antibiotics are prescribed for a specified duration of time (in days). Interruptions of short durations are permitted. However if the medicine is not taken continuously for a few days, the whole process has to be repeated. A third example comes from Ayurveda (the Indian system of medicine): Certain types of ailments need a type of massaging, twice or thrice a day, for 21/41/..., days. Interruption to this, due to patients health condition, is permissible only for 3 to 4 days at a stretch. Beyond that the whole process has to start from the very beginning; that is the treatment has to be repeated.

Other contexts of applications could be found in civil engineering (concreting of roofs), buffer sharing and so on. However a very significant example comes from the functioning of artificial heart valves. In a typical situation, a patient, surviving on an artificial heart valve was prescribed a particular tablet to be take once a day, at a specified time, to keep the level of a certain component in the blood between 2.5 and 3.5 units. The patient was advised to take the tablet daily at the specified time without fail (interruption). Nevertheless even when interruptions occur, a tolerance limit is prescribed. Beyond this tolerance level, the heart valve is to be replaced.

More examples can be cited, for example from priority queues, buffer sharing and so on. The above examples are all indicative of a specific characteristic. If within tolerance limit interruption is completed, the service which got interrupted gets resumed ie, start from where it got interrupted; else it gets repeated. In the first case the part of service that was provided is remembered whereas in the latter case the entire past is forgotten. It may be noted that the models under discussion are not restricted to the above applications alone and so the discussions to follow are in a much more general framework.

Throughout the thesis we assume that whenever the server is under interruption, no further interruption befalls the server. ie, the effect of interruption process is governed by a type-I counter (see [36]). The case where duration of interruption getting extended due to further interruptions befalling the server, who is already under interruption, is the subject matter of future work. This thesis entitled '*On Queues with Interruptions and Repeat or Resumption of Service*' introduces several new concepts into queues with service interruption. It is divided into Seven chapters including an introductory chapter. The following are keywords that we use in this thesis: Phase type (PH) distribution, Markovian Arrival Process (MAP), Geometric Distribution, Service Interruption, First in First out (FIFO), threshold random variable and Super threshold random variable. In the second chapter we introduce a new concept called the 'threshold random variable' which competes with interruption time to decide whether to repeat or resume the interrupted service after removal of interruptions. This notion generalizes the work reported so far in queues with service interruptions. In chapter 3 we introduce the concept of what is called 'Super threshold clock' (a random variable) which keeps track of the total interruption time of a customer during his service except when it is realized before completion of interruption in some cases to be discussed in this thesis and in other cases it exactly measures the duration of all interruptions put together. The Super threshold clock is on whenever the service is interrupted and is deactivated when service is rendered. Throughout this thesis the first in first out service discipline is followed except for priority queues.

1.6. Summary of the thesis

The first chapter is an introduction and contains a survey on queues with interruptions. It also contains some basic definitions and terminologies used in this thesis.

Second chapter presents an infinite capacity queueing system with a single server. A customer, on arrival to an idle server, is immediately taken for service; else he joins the tail of the queue. Arrival of customers constitute MAP with representation (D_0, D_1). During the service of a customer, none, one or more interruptions may occur. It takes a random amount of time to clear the interruption. Interruption process occurs according to a Poisson process of rate γ and duration of interruption is PH distributed. When the duration of an interruption exceeds a threshold random variable (also Phase type distributed), the customer whose service got interrupted has to undergo the service process right from the beginning on completion of interruption; else his service is resumed. We investigate this queueing system. Long run system distribution is obtained under stable regime. Several performance measures are evaluated. Numerical illustrations of the system behavior is also provided. As a variation we also consider in this chapter an infinite capacity queueing system with a single server, here with a bound on the number of interruptions that a customer is normally willing to accept, beyond which the system will have to pay a heavy penalty. All other assumptions remain the same as in model I. We investigate the behavior of this queueing system. Long run system distribution is obtained under stable regime. We analyze this problem as a Markov Decision process (MDP) to investigate the optimal N value. Several performance measures are evaluated. Numerical illustrations of the system behavior is also provided.

The third chapter describes an infinite capacity queueing system with a single server. A customer on arrival to an idle system immediately joins service. Arrival of customers constitute MAP. Here we introduce an upper bound N to the number of interruptions that a customer is subjected, beyond which no further interruption is permitted. Thus the probability p in the previous chapter turns out to be 1 for this model on realization of N interruptions. Interruptions occur according to a Poisson process of rate γ and duration of each interruption is PH distributed. Repeat / resumption of interrupted customers service is decided by a comparison between duration of interruption and a threshold random variable which is also PH distributed. When the duration of interruption exceeds the threshold random variable the interrupted customer gets its service repeated on completion of interruption; otherwise it is resumed after the interruption is removed. In addition to the bound on the number of interruptions we bring in another check on the interruption through a 'Super threshold clock'. This measures duration of interruption, except perhaps that of the last interruption. Once the super threshold clock is realized, the customer whose service got interrupted, will face no further interruption (note that the present interruption is allowed to be continued). So the super threshold clock and maximum number of interruptions permitted, act as checks on the actual number of interruptions/ total duration of interruptions a customer encounters. Long run system state distribution is obtained under stable regime. Several performance measures are evaluated. Numerical illustrations of the system behaviour is also provide. An optimization problem is numerically analyzed.

Chapter 4 studies a single server queueing model consisting of two queues—an infinite capacity queue of ordinary customers and a finite capacity (K) of priority customers. Customers join the system according to a Marked Markovian Arrival process. If the server is free, an arriving customer (ordinary/priority) can immediately join for service. During the service of an ordinary customer preemption can take place by the arrival of a priority customer. Then the preempted customer waits at the head of the ordinary queue till he is allowed to continue his service. An $(N+1)$ faced solid figure whose possible out comes are $0,1,\ldots\ldots,N$ with probabilities q_0, $q_1,\ldots\ldots,q_N$, respectively, is tossed at the beginning of the service of an ordinary customer which decide the maximum number of priority customer(s) allowed to be served during the service of the specified ordinary customer. Thus if the face i of the solid figure turns up, then at most i interruptions are permitted to the service of this ordinary customer. The restart/ resumption of preempted service takes place only when the priority queue become empty or the maximum number of priority customers permitted to be served during his service is realized, whichever occurs first. If the threshold random variable, which competes with the duration of preemption, is realized before completion of preemption then the preempted customer has to get its service repeated ; otherwise it is resumed. Here the random variable corresponding to ordinary customers service, priority customers service and threshold random variable are all PH distributed. This system is analyzed under stable regime. A few useful measures for system performance are obtained. These help in designing an efficient system. Numerical results are provided to illustrate the system performance . We also examine the optimal value of N by introducing a suitable cost function. Note that this chapter generalizes the results in the previous one, since, in the former the maximum number of interruptions during a service was systems choice and is fixed at $i, 0 \leqslant i \leqslant N$ whereas in this chapter it is left with customer to decide the maximum number of interruptions he/she is willing to undergo. However the system may encourage a low priority customer to opt for a large number of interruptions by way of providing incentives.

In chapter five we study an infinite capacity single server discrete time queueing system. Customers join the system according to PH distribution. If the system is idle, an arriving customer is immediately taken for service. During service several interruptions may occur where interruption process is geometrically distributed with probability γ for an interruption to occur. When interruption occurs a threshold random clock starts ticking. When the duration of an interruption exceeds a threshold random variable, the interrupted customer has to undergo the service right from the beginning, on completion of interruption; else his service is resumed. Several performance measures are evaluated. Numerical illustrations of the system behavior is also provided. It may be noted thát unlike in the continuous time models discussed in the previous chapters, here several events can occur at a time point. In that way more complexity is involved here. Further we work with transition probability matrices in place of the infinitesimal generator.

The last chapter is a comparison of the performance of all the models discussed through chapters 2 through 5. Some further possible investigations are also suggested in this chapter.

Chapter 2

On a Queue With Interruptions and Repeat or Resumption of service

This chapter presents an infinite capacity queueing system with a single server where service rule is FIFO . A customer, on arrival to an idle server, is immediately taken for service if the server is idle at that instant; else he joins the tail of the queue. Arrival of customers constitute *MAP.* During the service of a customer, one or more interruptions may occur. It takes a random amount of time to clear the interruption. Interruption process occurs according to a Poisson process of rate γ and duration of interruption is PH distributed. When the duration of an interruption exceeds a threshold random variable (also PH distributed), the customer whose service got interrupted has to undergo the service process right from the beginning on completion of interruption; else his service is resumed. We investigate this queueing system. Long run system distribution is obtained under stable regime. The several performance measures are evaluated. Numerical illustrations of the system behavior is also provided.

Part of this chapter appeared in Journal of Non-Linear Analysis .
A. Krishnamoorthy, Pramod.P.K, T.G. Deepak, On a Queue With Interruptions and Repeat or Resumption of Service, Non-Linear Analysis, 2009, (Elsevier). It was also given as invited talk to the Fourth World Conference of Non-Linear Analysts, Orlando, July 2008.

2.1. Model I

Here we consider a service system with a single server, to which customers arrive according to a Markovian arrival process with representation (D_0, D_1), the order of these matrices is r. An arriving customer enters service immediately if the server is free. When the system is in busy state, an interruption occurs to the service, resulting in preemption of the customer in service. When the interruption time exceeds a threshold random variable, the interrupted customer gets its service repeated on completion of interruption. Else service is resumed, that is it starts at the point where it got interrupted. The service time, interruption time and the threshold random variables are all mutually independent PH-distributed with representations (α,S), (β,T), (δ,U) and of orders a, b, c, respectively. Write $S^0 = -S\underline{e}$, $T^0 = -T\underline{e}$ and $U^0 = -U\underline{e}$ where $\underline{e}$ is a column vector of 1's of appropriate order. Let $N(t), S(t), S_1(t), S_2(t), S_3(t)$ denote respectively the number of customers in the system, state of the system, service phase, interruption phase and threshold phase.

If $N(t) \geqslant 1$ and service is going on at t , write $S(t) = 1$; if server is interrupted write $S(t) = 0$. $S_1(t)$ denotes the phase of service where it got interrupted if $S(t) = 0$ and phase of service if $S(t) = 1$. The variable $S_2(t)$ denotes phase of interruption at time t. The random variable $S_3(t)$ equal to 0 if duration of the interruption exceeds threshold random variable and phase of threshold random variable otherwise. The arrival process is represented by the variable $M(t)$.

Write $X(t) = (N(t), S(t), S_1(t), S_2(t), S_3(t), M(t))$; then $\{X(t) : t \geqslant 0\}$ forms a continuous time Markov chain (CTMC) which is a quasi death process (LIQBD) with state space whose n^{th} level is given by $\ell(n)=\bigcup_l \Psi(n,l), l = 0, 1$. The subsets of $\Psi(n,l)$ are defined as $\{(n, 0, i_1, i_2, i_3, i_4); 1 \leqslant i_1 \leqslant a, 1 \leqslant i_2 \leqslant b, 0 \leqslant i_3 \leqslant c, 1 \leqslant i_4 \leqslant r\}$, and $\{(n, 1, i_1, i_4); 1 \leqslant i_1 \leqslant a, 1 \leqslant i_4 \leqslant r\}$. The states in Ψ are listed in lexicographical order, then transitions among subsets $\Psi(n,l); l = 0, 1$ are as follows: Note that when $N(t) = 0$, the only other component in the state vector is $M(t)$. Also, when $S(t) = 1$, the last two components in the state vector stand for the phases of service and arrival process respectively. We now describe the infinitesimal generator matrix Q of this CTMC. Note that by the assumptions made above the CTMC $\{X(t), t \geqslant 0\}$ is a level independent quasi-birth and death process (LIQBD). We have

$$Q = \begin{bmatrix} D_0 & C_0 & 0 & 0 & 0 & 0 & 0 \\ C_1 & A_1 & A_0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 \\ & & & \cdot & \cdot & \cdot & \end{bmatrix} \tag{2.1}$$

In the above boundary matrices D_0, C_0 and C_1 have different dimensions which are further different from those of A_0, A_1 and A_2. First we shall describe the matrices A_0, A_1 and A_2 appearing in the repeating part of Q. A_0 corresponds to arrival of a customer to the system . The transitions in $I_a \otimes I_b \otimes I_{c+1} \otimes D_1$ and $I_a \otimes D_1$ record rates of jumping to $\Psi(n+1,0)$ and $\Psi(n+1,1)$ respectively from $\Psi(n,0)$ and $\Psi(n,1)$, which are of orders $ab(c+1)r$, ar respectively. All other transitions in A_0 are filled by zero matrices and A_0 is of order $ab(c+1)r + ar$. Similarly, A_2 corresponds to departure of a customer after completing the service. During the entire service a customer may encounter one or more interruptions. The only non zero component in A_2 corresponds to the transition to $\Psi(n-1,1)$ from $\Psi(n,1)$ which is given by $S^0\,\alpha \otimes I_r$. Since there cannot be a departure while remaining in interruption state such positions are represented by zero matrices. Also there cannot be departure along with transitions from busy state to interruption state. Thus this position is also occupied by zero matrix. A departure occurs only when the transition occurs from busy state to itself; that position is occupied by the matrix $S^0\,\alpha \otimes I_r$ and is of order ar. The block matrix A_2 is square matrix of order $ab(c+1)r + ar$.

Now consider the matrix A_1, which describes all transitions in which level does not change (that is transitions within levels). Let A_{11}, A_{12}, A_{13} and A_{14} record transitions from $\Psi(n,0)$ to $\Psi(n,0)$, $\Psi(n,0)$ to $\Psi(n,1)$, $\Psi(n,1)$ to $\Psi(n,0)$ and $\Psi(n,1)$ to $\Psi(n,1)$, respectively. A_{11} is described as follows: $I_a \otimes H$ records transitions to $\Psi(n,0)$ from $\Psi(n,0)$ and is of order $ab(c+1)r$, where only transitions within interruption phases and/or phase change of threshold clock occur. We have $H = G_1 \oplus G_2 = G_1 \otimes I_r + I_b \otimes G_2$, $G_1 = T \otimes I_{c+1}$, $G_2 = F \oplus D_0$ and $F = \begin{bmatrix} 0 & \bar{0} \\ U^0 & U \end{bmatrix}$ where F is square matrix of order $(c+1)$ having 0's in the first row and entries of U^0 in the first column, starting with the second element and the remaining part occupied by matrix U. A_{12} in A_1 is described as follows: B_0 records transitions to $\Psi(n,1)$ from $\Psi(n,0)$ and is of order $ab(c+1)r \times ar$ where $\begin{bmatrix} B_{01} & B_{02} & . & . & B_{0,a} \end{bmatrix}'$ and $B_{0j} = T^{(0)} \otimes \begin{bmatrix} \alpha & e_j & e_j & . & . & e_j \end{bmatrix}' \otimes I_r$; e_j is a row vector of appropriate order with 1 in the j^{th} place and zero elsewhere. B_0 represents transitions from interruption to busy state. In this transition, when interruption duration exceeds threshold, the interrupted customer gets its service repeated. Thus events in these transitions correspond to removal of interruption followed by resumption/repeat of service. A_{13} in A_1 records transition to $\Psi(n,0)$ from $\Psi(n,1)$ and is as follows: $\gamma I_a \otimes \beta \otimes \hat{\delta} \otimes I_r$ lists rates corresponding to the transitions from busy state to interruption state and $\hat{\delta} = (0, \delta)$. A_{14} in A_1 records transitions to $\Psi(n,1)$ from $\Psi(n,1)$ and is as follows: $S \oplus D_0 - \gamma I_{ar}$ records transitions in $\Psi(n,1)$ from $\Psi(n,1)$ represents no interruption to service . Now we consider the boundary blocks D_0, C_0 and C_1 in the infinitesimal generator Q. Here $C_0 = \alpha \otimes D_1$ and $C_1 = S^0 \otimes I_r$. In D_0 entries correspond to server remaining idle and transitions occur only in the arrival phase without triggering an arrival. C_0 corresponds to transitions with arrival and C_1 those resulting in departure of customer leaving behind none in the system .

2.2. Stability Condition.

Next we examine the system stability. We can anticipate that a very strong condition is needed here for the same since a service can get interrupted several times. What is needed is that the rate of drift to a lower level from a given one should be higher than that to the next higher level. This means that the Markov chain is stable if and only if $\Pi A_0 e < \Pi A_2 e$ where Π is the unique solution to $\Pi A = 0$, $\Pi e = 1$ and $A = A_0 + A_1 + A_2$. Nevertheless, it is extremely difficult to simplify the above condition because of the complexities involved . However validity of this condition can be checked numerically. Therefore we give two results: one a necessary condition and the other is a sufficient condition for stability. These are of independent interest as well. We need then while discussing the waiting time distribution and bounds for the expected waiting time. However these do not turn out to be necessary and sufficient is easy to check. To start with we formally note the necessary and sufficient condition. Its validity is well known (see Neuts(1981)[25]).

Theorem 2.2.1. *In order that the system is stable it is necessary and sufficient that* $\Pi A_0 e < \Pi A_2 e$.

2.2.1 Necessary Condition

In order to derive a necessary condition we proceed as follows. We assume here that each interruption results in resumption of service. Thus to compute the expected value of the time spent by a customer from the time he is taken for service till he leaves the system on being served completely, we look at random variables representing mutually exclusive and exhaustive events: $S^{(1)}, (S^{(1)} + X_1 + S^{(2)}), (S^{(1)} + X_1 + S^{(2)} + X_2 + S^{(3)}), \ldots\ldots\ldots$. In this the $S^{(1)}$ remaining separate is the full service time; in the second $(S^{(1)} + S^{(2)})$ give the full service time whereas X_1 is the duration of the single interruption; in the third, $(S^{(1)} + S^{(2)} + S^{(3)})$ provides the total duration of service and X_1, X_2 are the durations of the intervening interruptions, and, so on. The probabilities of occurrence of events represented by these random variables are P(Service time < Strike time of the first interruption), P(Service time > Strike time of first interruption, duration of interruption < threshold random variable, length of service time < strike time of the interruption random variable),

.... Thus the distribution of the required time =

$$PH(\alpha,S).P(S^{(1)} < Y) + (PH(\alpha,S) * PH(\beta,T)).P(Y < S).P(X_1 < T_1).$$
$$P(S^{(2)} < Y) + (PH(\alpha,S) * (PH(\beta,T))^{*2}).(P(Y < S))^2.P(X_1 < T_1).P(X_2 < T_1).$$
$$P(S^{(3)} < Y) + \ldots\ldots\ldots \tag{2.2}$$

where Y is an exponentially distributed random variable with parameter γ and T_1 is the threshold random variable which is $PH(\delta, U)$. We notice that the above is convergent by comparison with the renewal function. Further it has finite expected value since service time and interruption time have finite expectations. Now the necessary condition for system stability can be given as follows.

Theorem 2.2.2. *The necessary condition for stability of the system is that the number of arrivals during the expected value(duration) of distribution in (2.2) is less than one.*

2.2.2 Sufficient Condition:

The total time spent by a customer in the system, from the moment he is taken for service till he leaves the system on completion of service, can be represented by the random variables $S, (S_p^{(1)} + X_1 + S), (S_p^{(1)} + X_1 + S_p^{(2)} + X_2 + S), \ldots\ldots\ldots$ The distribution of the total time spent is the sum of the probabilities of the mutually exclusive events represented by these random variables multiplied by the probability of occurrence of these events. The probabilities of occurrence of these mutually exclusive events are respectively, $P(S < Y), P(S > Y).P(X_1 > T_1)P(S < Y), P(S > Y).P(X_1 > T_1).P(S > Y).P(X_2 > T_1).P(S < Y), \ldots\ldots$ where Y is an exponentially distributed random variable with parameter γ and T_1 is $PH(\delta, U)$. Thus the distribution of the time spent by a customer from the epoch at which he is taken for service, until he leaves the system, on completion of service is given by

$$PH(\alpha,S).P(S < Y) + ((PH(\alpha,S))^{*2} * PH(\beta,T)).P(S > Y).P(X_1 > T_1).$$
$$P(S < Y) + ((PH(\alpha,S))^{*3} * (PH(\beta,T))^{*2}).(P(S > Y))^2.(P(X_1 > T_1))^2.$$
$$(P(S < Y)), \ldots\ldots\ldots\ldots \tag{2.3}$$

Note that in the above even, for partial service we have given the full service time distribution. This argument leads to the following sufficient condition for system stability.

Theorem 2.2.3. *A sufficient condition for system stability is that the expected number of customers arriving to the system during the expected duration of (2.3) should be less than one.*

2.3. First passage time analysis

In this section, we investigate the expected length of a busy period and a busy cycle. As a prerequisite to that, we compute the expected length of time needed to reach level i, $i \geqslant 1$ from the level $i+1$. In particular when $i = 0$, we obtain the expected length of a busy period. Let $G_{j,j'}(k,x)$ be the conditional probability that the CTMC, starting in the state (i,j) at time $t = 0$, reaches the level $i-1$ for the first time at or prior to time x, after exactly k transitions to the left and does so by entering the state$(i-1,j')$ for $j' \geqslant 1$. Here by (i,j) we mean the j^{th} state in level i, when they are arranged in the lexicographic order. The matrix with elements $G_{j,j'}(k,x)$ is denoted by $G(k,x)$. Let $G^*(z,s) = \sum\limits_{k=1}^{\infty} z^k \int\limits_0^{\infty} e^{-sx} dG(k,x)$. Then, for $0 < z < 1$, $s > 0$, the matrix $G^*(z,s)$ is the minimal non-negative solution to the equation $zA_2 - (sI - A_1)G^*(z,s) + A_0 G^{*^2}(z,s) = 0$. We know that $\lim_{z\to 1, s\to 0} G^*(z,s)$ =G $= (G_{jj'})$ where $G_{jj'} = P\{\tau < \infty\}$ and $\{(X(\tau), Y(\tau)) = (i-1, j') \mid (X(0), Y(0)) = (i,j)\}$ and τ is the first passage time from the level i to the level $i-1$. Let m_{1j} be the mean first passage time from the level i $(i > 1)$ to the level $i-1$, given that the first passage time started in the state (i,j), and $\tilde{m}_1$ be a row vector with elements m_{1j}. Let m_{2j} be the mean number of transitions to the left during the first passage time from the level i to $i-1$, given that the first passage time started in the state (i,j). Let $\tilde{m}_2$ be a row vector with elements m_{2j}. Then

$$\begin{aligned} \tilde{m}_1 &= -\frac{\partial}{\partial s} G^*(z,s) e \,|_{s=0,z=1} \\ &= -(A_1 + A_0(I+G))^{-1} e. \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} \tilde{m}_2 &= \frac{\partial}{\partial z} G^*(z,s) e \,|_{s=0,z=1} \\ &= -(A_1 + A_0(I+G))^{-1} A_2 e. \end{aligned} \tag{2.5}$$

For computing the length of a busy period we define the matrix $G^{*(1,0)}(z,s)$ and vectors $\tilde{m}_1^{(1,0)}$ and $\tilde{m}_2^{(1,0)}$ which correspond to the first passage time from level 1 to level 0 and $G^{*(0,0)}(z,s)$, $\tilde{m}_1^{(0,0)}$ and $\tilde{m}_2^{(0,0)}$ for the first return time from level zero to the level zero. Then $G^{*(1,0)}(z,s) = z(sI - A_1)^{-1}C_1 + (sI - A_1)^{-1} A_0 G^*(z,s) G^{*(1,0)}(z,s)$ and $G^{*(0,0)}(z,s) = (sI - D_0)^{-1} C_0 G^{*(1,0)}(z,s)$.
Hence

$$\begin{aligned} \tilde{m}_1^{(1,0)} &= -\frac{\partial}{\partial s} G^{*(1,0)}(z,s) e \,|_{s=0,z=1} \\ &= -(A_1 + A_0 G)^{-1}(A_0 \tilde{m}_1 + e). \end{aligned} \tag{2.6}$$

$$\begin{aligned}\tilde{m}_2^{(1,0)} &= \frac{\partial}{\partial z} G^{*(1,0)}(z,s)e\,|_{s=0,z=1} \\ &= -(A_1 + A_0 G)^{-1}(A_0 \tilde{m}_2 + C_1 e).\end{aligned} \tag{2.7}$$

$$\begin{aligned}\tilde{m}_1^{(0,0)} &= -\frac{\partial}{\partial s} G^{*(0,0)}(z,s)e\,|_{s=0,z=1} \\ &= -\, D_0^{-1}(C_0 \tilde{m}_1^{(1,0)} + e).\end{aligned} \tag{2.8}$$

and

$$\begin{aligned}\tilde{m}_2^{(0,0)} &= \frac{\partial}{\partial z} G^{*(0,0)}(z,s)e\,|_{s=0,z=1} \\ &= -\, D_0^{-1}\, C_0 \tilde{m}_2^{(1,0)}.\end{aligned} \tag{2.9}$$

Note that $\tilde{m}_1^{(1,0)}$ and $\tilde{m}_1^{(0,0)}$ represent the mean lengths of busy period and busy cycle respectively. Thus, if the matrix G is known, (which can be computed by Logarithmic Reduction algorithm, see Latouche and Ramaswami (1999)[21]) all these measures can be calculated.
Next we pass on to a discussion of results related to the waiting time of a customer. Since during the service of a customer there can be a number of interruptions, some of which result in repetition of service whereas the rest in resumption, to derive an expression for the waiting time distribution could be quite challenging. Yet we make an attempt to compute the distribution function and derive bounds for the expected waiting time.

2.4. Service Process with Interruption

To describe service process with interruption, we assume that interruption duration and threshold random variable are exponentially distributed while the service time is phase type distributed. The service process with interruption can be expressed as a Markov process Ω with $4 \times a$ states given by $\{0,1\} \times \{0,1\} \times \{1,2,...,a\}$ plus one absorbing state. The absorbing state denotes a service completion. Let $\tilde{T}$ be the time until absorption of the process Ω. The process Ω can be represented by $X(t) = \{(i,j,k); i = 0,1; j = 0,1; k = 1,2,....,a\}$. If $i = 0$ then the server is in interrupted state, $i = 1$ means server busy or idle; if otherwise. If $j = 0$, then the threshold exceeded its saturation level and on completion of interruption, service can repeated and resumed if $j = 1$. The infinitesimal generator of this process is given by $Q_1 = \begin{bmatrix} \tilde{\Delta} & \tilde{\Delta}_0 \\ [0] & 0 \end{bmatrix}$ where $\tilde{\Delta} = \begin{bmatrix} \begin{bmatrix} -\tilde{\delta} I_a & [0] \\ \eta I_a & -(\eta+\tilde{\delta}) I_a \end{bmatrix} & \begin{bmatrix} \tilde{\delta}(e \otimes \beta) \\ \tilde{\delta} I_a \end{bmatrix} \\ \begin{bmatrix} [0] & \gamma I_a \end{bmatrix} & S - \gamma I_a \end{bmatrix}$

and $\tilde{\Delta}_0 = \begin{bmatrix} [0] \\ S^0 \end{bmatrix}$. The probability distribution $F(.)$ of T corresponding to the initial probability vector ξ is given by $F(x) = 1 - \xi.\exp(\tilde{\Delta}x)e, x \geqslant 0$. It's density function $F^{'}(x)$ in $(0, \infty)$ is given by $F^{'}(x) = \xi.\exp(\tilde{\Delta}x)\Delta^0$.
The Laplace-Stieltjes transform $f(s)$ of $F(.)$ is given by $f(s) = \xi.(sI - \tilde{\Delta})\tilde{\Delta}^0$.
The expected time for service completion E_s is given by $E_s = \xi(-\tilde{\Delta})^{-1}e$ and $\mu_s = \frac{1}{E_s}$.
The noncentral moments μ_i^1 of X are given by $\mu_i^{'} = (-1)^1 i!(-\tilde{\Delta})^{-1}e$.

2.5. Waiting time distribution

At the time of arrival of a customer we have the following mutually exclusive and exhaustive cases:
(i) Server is idle.
(ii) Server busy and the phase of service is i, $1 \leqslant i \leqslant a$; there are $(n \geqslant 0)$ customers waiting.
(iii) Server interrupted. The phase of interruption removal process is j, $1 \leqslant j \leqslant b$; the phase of the threshold (random clock) is k, $0 \leqslant k \leqslant c$; there are $(n \geqslant 0)$ customers waiting.
(I): In this case the customer is immediately taken for service. The probability for this event is x_0e. In this case the distribution of the amount of time spent by the customer in the system is his service time; the distribution of this is provided in theorems2.1 and2.2 respectively.
(II): Server busy and there are n customers waiting. Given the phase of service, the remaining service time is again phase-type distributed. This part may encounter interruption, in which case the distributions provided in Theorems2.1 and2.2 can be used to get the distribution of the remaining time to be spent by this customer. This convoluted with a phase type distribution corresponding to the service that was going on at the arrival epoch of the customer under consideration, convoluted further with distribution function of the time to be spent by the n customers for their services and intervening interruptions provides the distribution of the time spent by a customers before taken for service.
(III): Here first we have the phase-type distribution corresponding to the time until completion of interruption. Then depending on whether the interruption is completed before the random clock stopped ticking or the other way round, the service of the customer gets resumed/repeated. Nevertheless, we may need the number of interruptions encountered by the customer in service. The phase type distribution arising this way, convoluted with the distribution of time required for the service completion of the n customers provides the distribution of the time the customer will have to spend before taken for service.

In cases (II) and (III) distribution of the minimum time is n-fold convolution of the distribution given in (2.2) convoluted with the distribution of the time until the present customer (interrupted/undergoing service) completes service and that of the maximum time is the n-fold convolution of the distribution given by (2.3) convoluted with the distribution of the time until the present customer (in service/interrupted) completes service. Their expected values give respectively, the lower and upper bounds for the expected waiting time. In the section providing numerical illustrations, we throw some light on these also. It may be instructive to look at the problem in the following perspective. Assume that all distributions involved are exponential with parameters as described below. Inter arrival times, service times, inter arrival times of interruption, interruption duration and random (threshold) clock are exponentially distributed with rates λ, μ, γ, α, and β respectively. Of course, in this case there is no distinction between resumption and repetition of service and so each interruption results in repetition of service. Thus a complete service time has the distribution

$$\begin{aligned}((Exp(\mu) \leqslant t)).P(Exp(\mu) < Exp(\gamma)) + ((Exp(\mu) \leqslant t)^{*2} * (Exp(\mu) \leqslant t))). \\ P(Exp(\mu) > Exp(\gamma)).P(Exp(\mu) < Exp(\gamma)) + ((Exp(\mu) \leqslant t))^{*3} * \\ (Exp(\alpha) \leqslant t))^{*2}).P(Exp(\mu) > Exp(\gamma))^2. \\ P(Exp(\mu) < Exp(\gamma)) + \dots\dots\dots\end{aligned} \tag{2.10}$$

where $Exp(x)$ stands for exponential random variable with parameters x and $*n$ denotes n-fold convolution of the function with itself. The above series converges as it is dominated by a converging geometric series. Note that the above expression does not involve the exponential random variable with parameter β corresponding to the threshold clock for reasons indicated earlier.

Theorem 2.5.1. *A necessary and sufficient condition for this queueing model, with all underlying distributions exponential with parameters as indicated above, to be stable is that λ.(Expectation of distribution function given by (2.10))*< 1.

In order to have distinction between repeat and resumption of service, henceforth we assume that service time is not exponentially distributed. Any distribution function for service time other than exponential will serve the purpose.

2.6. Stationary distribution

Since the queueing model under study is an LIQBD process, its stationary distribution (if it exists) has a matrix-geometric solution. Assume that condition for system stability is satisfied. Let the stationary vector $\mathbf{x}$ of Q be partitioned by the levels into sub-vectors x_i for $i \geqslant 0$. Then x_i has the matrix-geometric form

$$x_i = x_1 R^{i-1} for \;\; i \geqslant 2 \tag{2.11}$$

where R is the minimal non-negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0, \tag{2.12}$$

and the vectors x_0, x_1 are obtained by solving the equations

$$\begin{aligned} x_0 D_0 + x_1 C_1 &= 0, \\ x_0 C_0 + x_1 (A_1 + RA_2) &= 0 \end{aligned} \tag{2.13}$$

subject to the normalizing condition

$$x_0 e + x_1 (I - R)^{-1} e = 1 \tag{2.14}$$

From the above equation, it is clear that to determine $\mathbf{x}$, a key step is the computation of the rate matrix R.

2.7. Performance Characteristics

Some useful descriptors of the model are listed below.

1. Mean number of customers in the system=$\sum_{n=1}^{\infty} n x_n e = x_1 (I - R)^{-2} e$
2. Fraction of time the server is busy=$\sum_{n=1}^{\infty} x_{n1} e$
3. Fraction of time the server remains interrupted=$\sum_{n=1}^{\infty} x_{n0} e$
4. Thus the fraction of time the server is idle=$x_0 e$
5. Fraction of time service is in interrupted state +Fraction of time service is going on=$\sum_{n=1}^{\infty} x_{n0} e + \sum_{n=1}^{\infty} x_{n1} e$
6. The rate at which server break down occurs=$\gamma \sum_{n=1}^{\infty} x_{n1} e$

7. The rate at which repair completion (removal of interruption) takes place before the threshold is reached) $R^c_{NT} = \sum_{n=1}^{\infty}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{r} x_{n,0,i,j,k,l}\, T^0_j$ where T^0_j is the j^{th} component of T^0

8. Rate at which repair completion takes place after the threshold is reached $R^c_T = \sum_{n=1}^{\infty}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{r} x_{n,0,i,j,0,l}\, S^0_{2j}$

9. Effective service rate $R^c_T = \sum_{n=1}^{\infty}\sum_{i=1}^{a}\sum_{l=1}^{r} x_{n,1,i,l}\, S^0_i$

10. The probability of a customer completing service without any interruption= P(service time< an exponentially distributed random variable with parameter γ) and is given by $\int_0^{\infty} (\alpha e^{Tu} T^0) e^{-\gamma u} du = \alpha(\gamma I - S)^{-1} S^0$

11. The probability that a customer encounters at least one interruption during his service= P(service time> an exponentially distributed random variable with parameter γ) =1-P(Service time <an exponential random variable with parameter γ) $= 1 - \alpha(\gamma I - S)^{-1} S^0$.

12. Probability of m consecutive services with at least one interruption each= (Probability that a customer encounters at least one interruption during his service)^m $= (1 - \alpha(\gamma I - S)^{-1} S^0)^m$

13. Probability that an interruption completion takes place before the threshold is reached= P(Interruption random variable>threshold random variable) =Phase type distribution with representation $(\tilde{\beta}, L)$ of order $bc + b + c$ where $\tilde{\beta} = [\beta \otimes \delta, \delta_{c+1}\beta, \beta_{b+1}\delta]$ and $L = \begin{bmatrix} T \otimes I + I \otimes U & I \otimes U^0 & T^0 \otimes I \\ 0 & T & 0 \\ 0 & 0 & U \end{bmatrix}$

14. Probability that an interruption completion takes place after the threshold is reached= P(interruption random variable<threshold random variable)=Phase type distribution with representation $[\beta \otimes \delta, T \otimes I + I \otimes U]$

2.8. Numerical Examples

In order to illustrate the performance of the system, we present some numerical results. Let $D_0 = \begin{bmatrix} -6.5 & 0.25 \\ 0.25 & -0.75 \end{bmatrix}$, $D_1 = \begin{bmatrix} 6.0 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$, $S = \begin{bmatrix} -12.0 & 6.0 \\ 6.0 & -12.0 \end{bmatrix}$,

$$T = \begin{bmatrix} -12.0 & 3.0 \\ 3.0 & -12.0 \end{bmatrix}, U = \begin{bmatrix} -12.0 & 8.0 \\ 8.0 & -12.0 \end{bmatrix} S^0 = \begin{bmatrix} 6.0 \\ 6.0 \end{bmatrix}, T^0 = \begin{bmatrix} 9.0 \\ 9.0 \end{bmatrix},$$
$$U^0 = \begin{bmatrix} 4.0 \\ 4.0 \end{bmatrix}.$$

When γ is progressively decreased and comes closer and closer to zero, our model converges to classical queueing problem without interruption. Thus the ratio $\frac{\Pi A_0 e}{\Pi A_2 e}$ converges to the traffic intensity ρ of the classical case when γ tend to 0. This is illustrated in Table 1. In this Π is the stationary vector of the infinitesimal generator $A = A_0 + A_1 + A_2$.

Table 1:

γ	6.5	4.5	2.5	0.5	0.4	0.3	0.2	0.1	0.05	0
$\frac{\Pi A_0 e}{\Pi A_2 e}$	0.96875	0.84375	0.71875	0.59375	0.5875	0.58125	0.5750	0.56875	0.56313	0.5625

Let N_{mean}= Mean number of customers in the system.

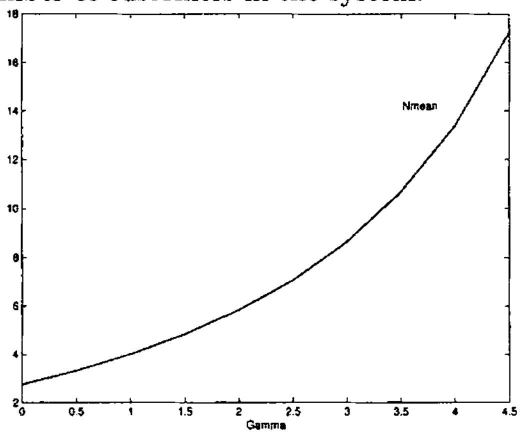

Fig. 2.1: Gamma versus Mean Number of Customers in the system

To investigate the effect of γ on the idle and busy time of the server we introduce the following notations, F_{idle}= Fraction of time the server is idle, F_{int}= Fraction of time server is interrupted and $R_{idle+busy}$= Fraction of time server idle+ Fraction of time server busy. Thus we have results given in

Also introduce the notations R_{BD}= Rate at which server breakdown occurs and E_{SR}= Effective service rate.

We notice from figure 2.1 that the average number of customers increases with increase in the interruption rate; of course this is as expected. Also when interruption duration exceeds threshold, the customer need to get its service repeated. Since the arrival rate is constant, the number of customers increases with increase in the

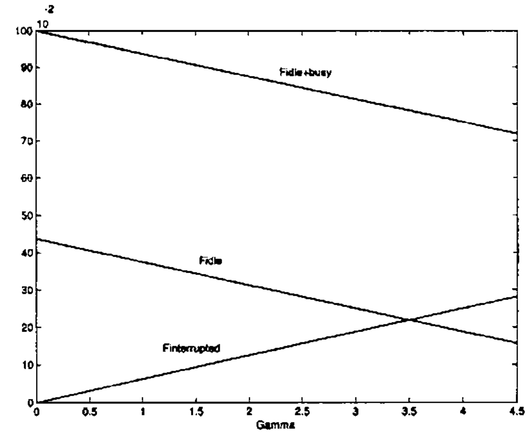

Fig. 2.2: Gamma versus F_{int}, F_{idle}, $F_{idle+busy}$

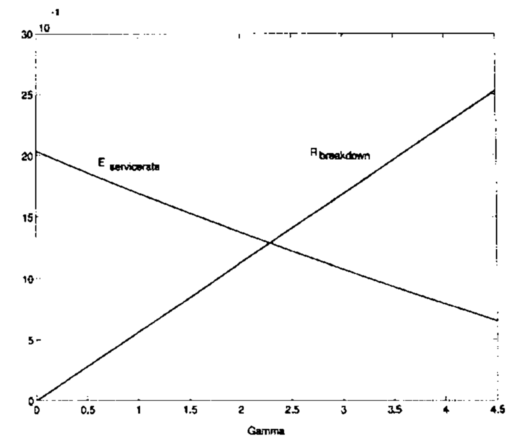

Fig. 2.3: Gamma versus Rate at which server breakdown occurs and Effective service rate

interruption rate. Since no reneging takes place, each customers has to wait for his turn of service. In fig.2.2, we see that fraction of time server is idle decreases with increase in the interruption rate and fraction of time server is interrupted increases with increase in the interruption rate. fig.2.3 shows that rate of server breakdown and effective service rate increase with increase in the interruption rate.

Next we define a busy period as the length of time starting with an arrival to an ideal system, until the server becomes free for the first time (no customer left). Then with γ increasing the expected length of busy period should increase. This is demonstrated in fig. 2.3. The calculations are based on the expressions for first passage time as obtained in section 5.

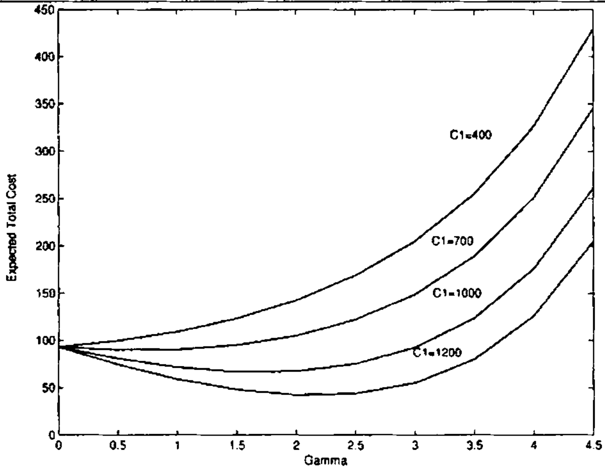

Fig. 2.4: Gamma versus expected total cost

2.9. Cost Function

To construct a cost function we assume that interruption produces higher revenue to the system. For example server may be assigned to serve a high priority customer leading to increased income to the system. Let the per unit time revenue on this be C_1. However there is associated holding cost of the customer whose service got interrupted and also that of the remaining customers waiting in line. The holding cost (C_3) of customer whose service got interrupted can be taken to be higher than that for those waiting in the queue (C_2). Hence here we assume that the holding cost of interrupted customer is greater than those waiting in the queue. Idle time also involves an expenditure to the system. We denote by C_4 the cost per unit time of server remaining idle. Thus we introduce the per unit time cost as follows:
C_1=Revenue per unit time interruption, C_2 =Holding cost of the customer waiting in the queue, C_3=Holding cost of the customer interrupted and C_4=Cost(expense) when the server is idle. Total Expected Cost=-(Fraction of time interrupted)C_1+ (Mean number of customers in the system)C_2+ (Fraction of time interrupted)C_3+ (Fraction of time server idle)C_4. We fix C_2 = \$ 50, C_3 = \$ 75 and C_4 = \$ 25. It is seen from fig.2.4 that the total expected cost decreases first and then increases with increasing γ for sufficiently large values of C_1. In the case of moderately large values of C_1, the expected total cost shows an increasing trend with γ .

Model II: Queue with Finite Number of Interruptions

So far we have permitted any number of interruptions to befall during the service of a customer in this section we restrict this to a finite number . The arrival process, interruption process, phase of service, phase of interruption time, threshold phase are all following the same distributions as in model *I*. Let N be the maximum number of interruptions encountered by a customer in service.

The service time, interruption time and the threshold random variables are all mutually independent PH-distributed with representations (α,S), (β,T), (δ,U) and of orders a, b, c respectively. Write $S^0 = -S\underline{e}$, $T^0 = -T\underline{e}$ and $U^0 = -U\underline{e}$ where $\underline{e}$ is a column vector of 1's of appropriate order and $B_1(t)$, $B_2(t)$ are the random variable corresponding to number of interruptions and number of occasions in which duration of interruption exceeds threshold random variable. All other variables have same meaning as described in model I. Let $X(t) = (N(t), S(t), B_1(t), B_2(t), S_1(t), S_2(t), S_3(t), M(t))$; then $\{X(t) : t \geqslant 0\}$ forms a continuous time Markov chain (CTMC) with state space whose n^{th} level is given by $\ell(n)=\bigcup_l \Psi(n,l), l = 0, 1$. The subsets of $\Psi(n,l)$ are defined as $\{(n, 0, j_1, j_2, i_1, i_2, i_3, i_4)$ $; 0 \leqslant j_1 \leqslant N-1, 0 \leqslant j_2 \leqslant J_1, 1 \leqslant i_1 \leqslant a, 1 \leqslant i_2 \leqslant b, 0 \leqslant i_3 \leqslant c, 1 \leqslant i_4 \leqslant r\}$, and $\{(n, 1, j_1, j_2, i_1, i_4); 0 \leqslant j_1 \leqslant N, 0 \leqslant j_2 \leqslant J_1, 1 \leqslant i_1 \leqslant a, 1 \leqslant i_4 \leqslant r\}$.

Note that when $N(t) = 0$, the only other component in the state vector is $M(t)$. Also when $S(t) = 1$, the last two components in the state vector stand for the phases of service time and arrival process respectively. Note that by the assumptions made above the CTMC $\{X(t), t \geqslant 0\}$ is a level independent quasi-birth (LIQBD) and death process with infinitesimal generator matrix Q

$$Q = \begin{bmatrix} D_0 & C_0 & 0 & 0 & . & . \\ C_1 & A_1 & A_0 & 0 & 0 & . \\ 0 & A_2 & A_1 & A_0 & 0 & . \\ 0 & 0 & A_2 & A_1 & A_0 & . \\ 0 & 0 & 0 & & . & . \\ & & & & & . \end{bmatrix} \tag{2.15}$$

In the above boundary matrices D_0, C_0 and C_1 have different dimensions and these are further different from those of A_0, A_1 and A_2. First we shall describe the matrices A_0, A_1 and A_2 appearing in the repeating part. $I_a \otimes I_b \otimes I_{(c+1)} \otimes I_{N(N+1)/2} \otimes D_1$ and $I_a \otimes I_{(N+1)(N+2)/2} \otimes D_1$ records transitions to $\Psi(n+1, 0)$, $\Psi(n+1, l)$ from $\Psi(n, 0)$ and $\Psi(n, l)$. All other components are zero matrices. Here $I_a \otimes I_b \otimes I_{(c+1)} \otimes I_{N(N+1)/2} \otimes D_1$ and $I_a \otimes I_{(N+1)(N+2)/2} \otimes D_1$ are square matrices corresponding to arrival of customers to the system when the system is in interrupted and busy state of the server, respectively. Obviously A_0 is a square matrix of order $ab(c+1)rN(N+1)/2+ar(N+1)(N+2)/2$.

Similarly A_2 corresponds to departure of a customer after completing the service. Before completion of service there may be a number of interruptions. The only non zero element in A_2 is the transition to $\Psi(n-1, 1)$ starting from $\Psi(n, 1)$ and is the matrix $\left[e_{(N+1)(N+2)/2} \otimes S^0\alpha \otimes I_r, [0]\right]$. Since there cannot be a departure while remaining in interruption state and with transition from interruption state to busy state, such positions are represented by zero block matrices. The block matrix A_2 is square matrix of order $ab(c+1)rN(N+1)/2+ar \times (N+1) \times (N+2)/2$.

Now we discuss the matrix A_1 which records transitions in $\Psi(n,.)$ starting from $\Psi(n,.)$. The components in A_1 are F_1, F_2, F_3, F_4 corresponds to transition from $\Psi(n,0)$ to $\Psi(n,0)$, $\Psi(n,0)$ to $\Psi(n,1)$, $\Psi(n,1)$ to $\Psi(n,0)$ and $\Psi(n,1)$ to $\Psi(n,1)$ respectively.

Here continuing in the interruption state is represented by the matrix F_1 and is of order $ab(c+1)rN \times (N+1)/2$, where only transitions due to interruption and threshold phase changes occur. We have $F_1 = I_{N(N+1)/2} \otimes I_a \otimes H$, $H = G_1 \oplus G_2$, $G_1 = T \otimes I_{c+1}$, $G_2 = F \oplus D_0$ and $F = \begin{bmatrix} 0 & \bar{0} \\ U^0 & U \end{bmatrix}$ where F is square matrix of order $(c+1)$ having $0's$ in the first row and entries of U^0 in the first column, starting with the second element and the remaining part occupied by matrix U. F_2 in A_1 is of order $ab(c+1)rK(K+1)/2 \times ar(N+1)(N+2)/2$ and is $F_2 = \begin{bmatrix} [0] & diag\left(F_2^{(i-1,i)}\right) \end{bmatrix}$; $1 \leqslant i \leqslant N$; where $diag\left(F_2^{(i-1,i)}\right)$ is a diagonal matrix whose i^{th} diagonal element is $F_2^{(i-1,i)}$, which lists the transition to $\Psi(n,1)$ starting from $\Psi(n,0)$ on completion of interruption. Here the matrix $F_2^{(0,1)} = [F_{21}^{(0,1)}, F_{22}^{(0,1)}]$ represents transition from interruption state to busy state on completion of the first interruption. In that, matrices $F_{21}^{(0,1)}$, $F_{22}^{(0,1)}$ corresponding to first interruption completion takes place before threshold is reached and after threshold is reached respectively. Here $F_{21}^{(0,1)} = [B_{01}, B_{02}, \ldots\ldots\ldots, B_{0a}]'$ where $B_{oj} = T^0 \otimes [\bar{0}, \tilde{e}_j, \tilde{e}_j, \ldots\ldots, \tilde{e}_j]' \otimes I_a$ and $\tilde{e}_j$ is a row vector of appropriate order with 1 in the j^{th} place and zero elsewhere. $F_{22}^{(0,1)} = [B'_{01}, B'_{02}, \ldots\ldots\ldots, B'_{0a}]'$ where $B'_{oj} = T^0 \otimes [\bar{\alpha}, \bar{0}, \bar{0}, \ldots\ldots, \bar{0}]' \otimes I_a$. Here the matrix $F_2^{((i-1),i)} = diag \begin{bmatrix} F_{21}^{(0,1)} & F_{22}^{(0,1)} \end{bmatrix}$ is a diagonal matrix of order i whose diagonal element is $\begin{bmatrix} F_{21}^{(0,1)} & F_{22}^{(0,1)} \end{bmatrix}$. F_3 appearing in A_1 records the transition to $\Psi(n,0)$ starting from $\Psi(n,1)$ is the matrix $I_{(N+1)(N+2)/2} \otimes \gamma I_b \otimes \beta \otimes \hat{\delta} \otimes I_a$ where $\hat{\delta} = (0, \delta)$.

F_4 lists transitions in $\Psi(n,1)$ starting from $\Psi(n,1)$. Since there is bound on number of interruptions (N), no interruption during the service of the customer who has already N interruptions. The matrix F_4 is given by $F_4 = I_{N(N+1)/2} \otimes [S \oplus D_0 - \gamma I_{ar},\ S \oplus D_0]'$ Now we consider the boundary blocks D_0, C_0 and C_1 in the infinitesimal generator Q. Here $C_0 = [[0]_{01}, [\alpha \otimes D_1], [0]_{02}]$ and $C_1 = \begin{bmatrix} [0] \\ e_{(N+1)(N+2)/2} \otimes S^0 \otimes I_a \end{bmatrix}$ The matrices $[0]_{01}$ and $[0]_{02}$ are matrices of dimensions $r \times (N(N+1)abcr)/2$ and $r \times ((N+1)(N+2)abcr)/2$, $r \times ((N^2+3N+1)ar)/2$ respectively. In D_0, entries correspond to server remaining idle and transitions occur only in the arrival phase without triggering an arrival. C_0 corresponds to transitions with arrival and C_1

those resulting in departure of customer leading behind none in the system .

2.10. Stationary distribution

Since the model is studied as an LIQBD process, its stationary distribution (if it exists) has a matrix-geometric solution. Assume that condition (*) is satisfied. Let the stationary vector π of Q be partitioned by the levels into sub-vectors π_i for $i \geqslant 0$. Then π_i has the matrix-geometric form.

$$\pi_i = \pi_1 R^{i-1} for \;\; i \geqslant 2 \tag{2.16}$$

where R is the minimal non-negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0, \tag{2.17}$$

and the vectors π_0, π_1 are obtained by solving the equations

$$\begin{aligned} \pi_0 D_0 + \pi_1 C_1 &= 0, \\ \pi_0 C_0 + \pi_1(A_1 + RA_2) &= 0 \end{aligned} \tag{2.18}$$

subject to the normalizing condition

$$\pi_0 e + \pi_1 (I - R)^{-1} e = 1 \tag{2.19}$$

From the above equation, it is clear that to determine π, a key step is the computation of the rate matrix R.

2.11. Analysis of Service Process

Expected Service Time

The Markov Process is $X(t) = \{B(t), S(t), S_3(t), S_1(t)\}$. Here $B(t)$ denotes the number of interruptions already encountered by the customer in service, S(t) corresponds to state of the server, $S_3(t)$ is the phase of threshold random variable and $S_1(t)$ corresponds to phase of service at time t. The X be the time until absorption of the process $X(t)$. The state of the process can be represented by $X(t) = \{(n, i, j, k) : i = 0, 1; j = 0, 1; 1 \leqslant k \leqslant a\}$. (a) If $i = 0$ the system is interrupted and 1 otherwise. (b) If $i = 0$ then the number of interruptions $n = 0, 1, ..., N - 1$, including the present one. (c) If $i = 0$ then the value of $j = 0$

or 1 which corresponds to threshold random variable. The saturation of threshold is indicated by $j = 0$ and non saturation by $j = 1$. (d) The service time is phase type distributed with representation (α, S) of order a. The service phases are 1,2,...,a and one absorbing state.
The infinitesimal generator of this process is given by

$$\Delta = \begin{bmatrix} I_{N-1} \otimes \begin{bmatrix} -\beta I_a & [0] \\ \delta I_a & -(\beta+\delta) I_a \end{bmatrix} & \begin{bmatrix} [0] & I_{N-1} \otimes \begin{bmatrix} \beta(e \otimes \alpha) \\ \beta I_a \end{bmatrix} \end{bmatrix} \\ \begin{bmatrix} I_{N-1} \otimes \begin{bmatrix} [0] & \gamma I_a \end{bmatrix} \\ [0] \end{bmatrix} & \begin{bmatrix} I_{N-1} \otimes (S - \gamma I_a) & [0] \\ [0] & S \end{bmatrix} \end{bmatrix}$$ and $\Delta^0 = \begin{bmatrix} [0] \\ e \otimes S^0 \end{bmatrix}$ where e is a column vector of 1's of order $N+1$. Here $Q = \begin{bmatrix} \Delta & \Delta^0 \\ \bar{0} & 0 \end{bmatrix}$ and the initial probability vector of Q is given by $\xi = (0.\alpha, \alpha, 0)$. The probability distribution $F(.)$ of X corresponding to the initial probability vector ξ is given by $F(x) = 1 - \xi.\exp(\Delta x)e, x \geqslant 0$. Its density function $F'(x)$ in $(0, \infty)$ is given by $F'(x) = \xi.\exp(\Delta x)\Delta^0$. The Laplace-Stieltjes transform $f(s)$ of $F(.)$ is given by $f(s) = \xi.(sI - \Delta)\Delta^0$.
The expected time for service completion E_s is given by $E_s = \xi(-\Delta)^{-1}e$ and $\mu_s = \frac{1}{E_s}$. The noncentral moments μ_i' of X are given by $\mu_i' = (-1)^i i!(-\Delta)^{-1}e$. In order that the system is stable the number of arrivals during the effective service time of a customer should be less than one. The we have

Theorem 2.11.1. *The system is stable iff* $\lambda < \mu_s$.

2.12. Expected number of Interruptions

To calculate the expected number of interruptions during a single service, we consider the Markov process $X(t) = (N(t), S(t), S_3(t), S_1(t))$ where $N(t)$ is the number of completed interruptions, $S(t) = 0$ or 1, corresponding to interrupted service or not, and $S_3(t)$ is the phase of threshold random variable and $S_1(t)$, phase of service time. Here interruption time, threshold random variable are exponentially distributed with rates δ, η and service time is phase type distributed and of representation (α, S). $X(t)$ has the state space $\{\Delta\} \cup \{0, 1, 2, ...\} \times \{0, 1\} \times \{0, 1\} \times \{1, 2, ..., a\}$ where Δ is an absorbing state which denotes the service completion. The infinitesimal generator matrix is

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & & \cdot & \cdot \\ B_0 & C_0 & C_1 & 0 & 0 & & \cdot & \cdot \\ A_2 & 0 & A_1 & A_0 & 0 & & & \cdot \\ A_2 & 0 & 0 & A_1 & A_0 & \cdot & \cdot & \cdot \\ A_2 & 0 & 0 & 0 & A_1 & A_0 & \cdot & \cdot \\ 0 & & \cdot & & \cdot & \cdot & & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

where $B_0 = S^0$, $C_0 = S - \gamma I_a$, $C_1 = [\gamma I_a\, [0]]$.

The matrix $A_2 = \begin{bmatrix} [0] \\ S^0 \end{bmatrix}$, $A_1 = \begin{bmatrix} \begin{pmatrix} -\delta(e\otimes\alpha) & [0] \\ \eta I_a & -(\eta+\delta)I_a \end{pmatrix} & \begin{pmatrix} \delta(e\otimes\alpha) \\ \delta I_a \end{pmatrix} \\ [0] & S-\theta I_a \end{bmatrix}$

and $A_0 = \begin{bmatrix} [0] & [0] \\ \gamma I_a & [0] \end{bmatrix}$.

Let p_i be the probability that an absorbtion occurs with exactly i interruptions, the $p_0 = \alpha(-C_0^{-1}B_0)$ and $p_i = \alpha(-C_0^{-1}B_0)(A_1^{-1}A_0)^{i-1}(-A_1^{-1}A_2), i = 1,2,3,.......$

Expected number of interruption before absorption is

$E = \big((I - (-A_1^{-1}A_0))(1-N)(-A_1^{-1}A_0)^N + (-A_1^{-1}A_0)(1-(-A_1^{-1}A_0))^{N-1}\big)$
$(I - (-A_1^{-1}A_0))^{-1}e.$

2.13. Expected Waiting Time

To find the expected waiting time of a customer who joined in the queue, we tag the customer by giving his token number as r. Now we consider the Markov process $X(t) = (N(t), S(t), B(t), S_3(t), S_1(t)$ where $N(t)$ denotes the rank of the customer, $S(t)$ the state of the system, $B(t)$ corresponds to number of customers, S_3 that of threshold random variable and $S_1(t)$ the service phase. The value of $N(t)$ is r if he joins as the k^{th} customer in the queue. $N(t)$ reduces to 1 if the customers ahead of him leave the system after completing their service. If $S(t) = 0$, state space of $X(t)$ as$X(t) = \{k, k-1, k-2, ...3, 2, 1\} \times \{0, 1, 2, ..., N-1\} \times \{0, 1\} \times \{1, 2, ...a\}$ and if $S(t) = 1$, then it is $\{k, k-1, k-2, ...3, 2, 1\} \times \{0, 1, 2, ..., N\} \times \{1, 2, ..., a\}$. Also there is one absorbing state which denotes the tagged customer is selected for service. Thus the infinitesimal generator $\bar{Q}$ of the process $X(t)$ takes the form

$$\bar{Q} = \begin{bmatrix} U & U^0 \\ \bar{0} & 0 \end{bmatrix} \text{ where } U = \begin{bmatrix} A_1 & A_2 & 0 & 0 & & & 0 & 0 \\ 0 & A_1 & A_2 & 0 & & & 0 & 0 \\ 0 & 0 & A_1 & A_2 & & & 0 & 0 \\ & & & . & & & & \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & A_1 & A_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A_1 & A_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_1 \end{bmatrix} \text{ and } U^0 =$$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ e_{N+1}\otimes S^0 \end{bmatrix} \text{ where } A_1 = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} -\beta I_a & [0] \\ \delta I_a & -(\beta+\delta)I_a \end{bmatrix} \\ I_N \otimes \begin{bmatrix} [0] & \gamma I_a \end{bmatrix} \\ [0] \end{bmatrix} & \begin{bmatrix} \begin{bmatrix} I_N \otimes \beta \begin{bmatrix} e\otimes\alpha \\ I_a \end{bmatrix} \end{bmatrix} \\ \begin{bmatrix} I_N \otimes (S-\gamma I_a) & [0] \\ [0] & S \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

and $A_2 = \begin{bmatrix} [0] & [0] \\ [0] & I_{N+1} \otimes S^0 \otimes \alpha \end{bmatrix}$.

Expected waiting time of a customer who joins the queue as r^{th} customer = $-A_1^{-1}(I - (A_2 A_1^{-1})^r)(I - A_2 A_1)^{-1}$

2.14. Performance Characteristics

Some useful general descriptors of our model are listed below.

1. Mean number of customers in the system$=\sum_{n=1}^{\infty} n\pi_n e = \pi_1(I - R)^{-2}e$
2. Fraction of time the server is busy$=\sum_{n=1}^{\infty} \pi_{n1} e$
3. Fraction of time the server remains interrupted$=\sum_{n=1}^{\infty} \pi_{n0} e$
4. Thus the fraction of time the server is idle$=\pi_0 e$
5. Fraction of time service is in interrupted state +Fraction of time service is going on$=\sum_{n=1}^{\infty} \pi_{n0} e + \sum_{n=1}^{\infty} \pi_{n1} e$
6. The rate at which server break down occurs$=\gamma \sum_{n=1}^{\infty} \pi_{n1} e$
7. The rate at which repair completion (removal of interruption) takes place before the threshold is reached) $R_{NT}^c = \sum_{n=1}^{\infty} \sum_{x_1=1}^{N} \sum_{x_2=1}^{x_1} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{r} \pi_{n,0,x_1,x_2,i,j,k,l}\, T_j^0$ where S_{2j}^0 is the j^{th} component of T^0
8. Rate at which repair completion takes place after the threshold is reached $R_T^c = \sum_{n=1}^{\infty} \sum_{x_1=1}^{N} \sum_{x_2=1}^{x_1} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{l=1}^{r} \pi_{n,0,x_1,x_2,i,j,0,l}\, T_j^0$
9. Effective service rate $R_T^c = \sum_{n=1}^{\infty} \sum_{x_1=0}^{N} \sum_{x_2=0}^{x_1} \sum_{i=1}^{a} \sum_{l=1}^{r} \pi_{n,1,x_1,x_2,i,l}\, S_i^0$
10. The probability of a customer completing service without any interruption= P(service time < an exponentially distributed random variable with parameter γ) is given by $\int_0^{\infty} (\alpha e^{Tu} T^0) e^{-\gamma u} du = \alpha(\gamma I - S)^{-1} S^0$

2.15. Numerical Results

To illustrate numerically we fix the following values:

N=3, a=2,b=2,c=2,r=2;$D_0 = \begin{bmatrix} -6.5 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$,$D_1 = \begin{bmatrix} 6.0 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$,

$S = \begin{bmatrix} -12.0 & 6.0 \\ 6.0 & -12.0 \end{bmatrix}, T = \begin{bmatrix} -12.0 & 3.0 \\ 3.0 & -12.0 \end{bmatrix}, U = \begin{bmatrix} -12.0 & 8.0 \\ 8.0 & -12.0 \end{bmatrix}$,

$S^0 = \begin{bmatrix} 6.0 & 6.0 \end{bmatrix}', T^0 = \begin{bmatrix} 9.0 & 9.0 \end{bmatrix}', U^0 = \begin{bmatrix} 4.0 & 4.0 \end{bmatrix}', \alpha = \begin{bmatrix} 0.4 & 0.6 \end{bmatrix}$,
$\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}, \delta = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$.

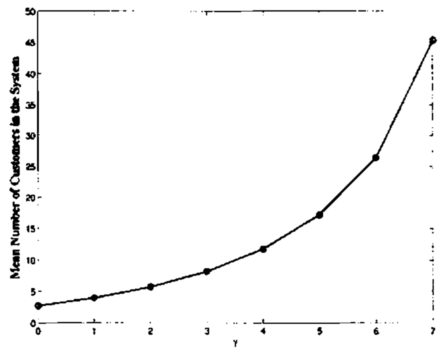

Fig. 2.5: Gamma versus Mean Number of Customers in the System. This shows that the mean number of customers in the system increases with increasing interruption rate, at a nonlinear rate.

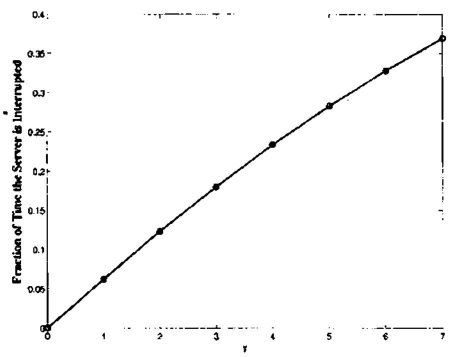

Fig. 2.6: Gamma versus Fraction of Time the Server is Interrupted.

From this figure we learn that the fraction of time the server is interrupted increases with increasing interruption rate. From the figure 2.7 we get a clear picture

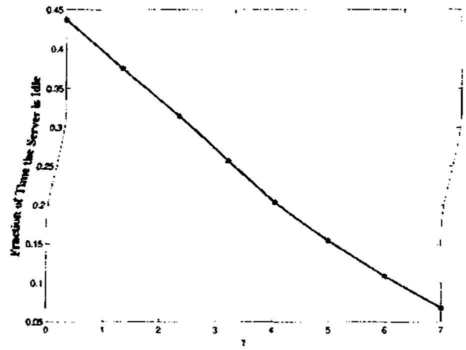

Fig. 2.7: Gamma versus Fraction of Time the Server is Idle.

about the idle time. The idle time decreases with increasing rate of interruption This figure 2.8 points to the fact that effective service rate decreases with increasing

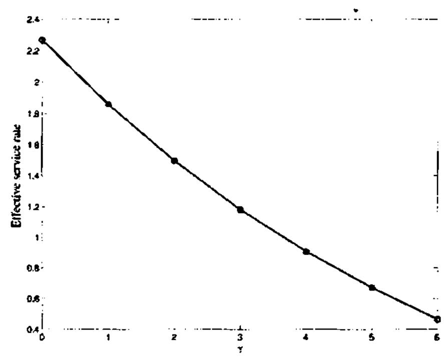

Fig. 2.8: Gamma versus effective service rate.

interruption rate.

Increase in the values of γ results in increase in breakdown. This is verified in the figure 2.9. Note that, though the relationship between the two is linear, the constant of proportionality is less than one service we are at type I counter like situation.

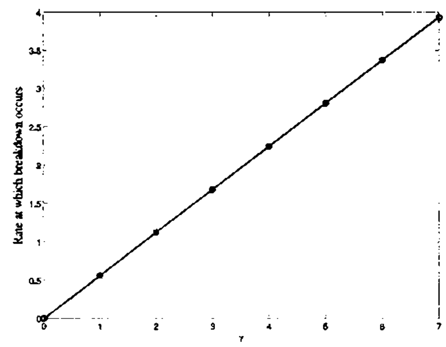

Fig. 2.9: Gamma versus Fraction of time the Server breakdown occurs.

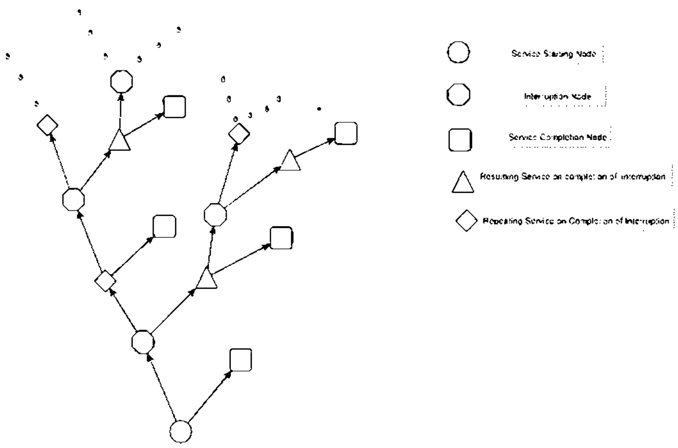

Fig. 2.10: Tree structure of the model

2.16. Cost Function

2.16.1 Cost function I: Here we take into account cost involved with events such as Interruption, Non-interruption, Repeat and Resumption

In this model interruption can occur at any time during a service. Here we can consider interruption as the server disconnecting current service temporarily and to take up better assignment which gives higher income. Every interruption results in repeat/resumption of service. Tree structure of the model is shown in 2.10.

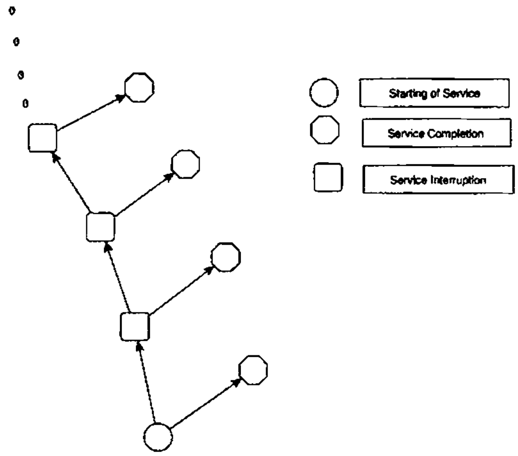

Fig. 2.11: Tree structure of the model

Repeat/resumption of service just after interruption results in an additional cost to the server. We label the variables in the following way: p_i be the probability of the occurrence of i^{th} interruption . C_i be the additional benefit due to i^{th} interruption and C_i' that of non interruption. r_i be the probability of repetition of service just after i^{th} interruption and $1 - r_i$ that of resumption of service. R_i be the cost incurred to the system due to repeat of service just after i^{th} interruption and g_i that of resumption.

If a customer completes his service without any interruption then Expected revenue at that stage=$(1-p_1)C_1'$. If customer completes service after 1^{st} interruption, then Expected revenue at that stage=$2p_1C_1 + (1-r_1)g_1 + r_1R_1 + 2(1-p_2)C_2'$ If customer completing service after 2^{nd} interruption, then Expected revenue at that stage=$2^2\sum_{i=1}^{2} p_iC_i + 2\sum_{i=1}^{2}(1-r_i)g_i + \sum_{i=1}^{2} r_iR_i + 2^2(1-p_3)C_3'$. Continuing like this if a customer completing service after N^{th} interruption,
then Expected revenue at that stage=$2^N\sum_{i=1}^{N} p_iC_i + 2^{N-1}\sum_{i=1}^{N}(1-r_i)g_i + \sum_{i=1}^{N} r_iR_i + 2^N(1-p_{N+1})C_{N+1}'$.

2.16.2 Cost Function II: A Decision process, namely, decide to interrupt or not to interrupt a service at the epoch at which interruption occurs.

The model is explained using the tree structure in 2.11.

We find the minimum cost/maximum revenue function by dynamic programming approach. Consider a dynamic system which is reviewed at the end of each interruption i, i=0,1,......,N. At each review the system is classified into one of the possible number of stages and subsequently a decision has to be made. The set of possible stages is denoted by $\triangle$. For each stage $j \in \Omega$, a set $\Omega(j)$ of decisions or actions is given. The set $\triangle$ and the action sets $\Omega(j)$ are assumed to be finite. The economic consequences of the decisions taken at the review times (decision epochs) are reflected in costs. This controlled dynamic system is called a discrete-time Markov decision model when the following Markovian property is satisfied. If at a decision epoch the action action $\Omega(j) = \{interruptions, noninterruptios\}$ is chosen in state i, then regardless of the past history of the system, the following happens. Here stage space is $1, 2, \ldots\ldots\ldots\ldots, N \in \triangle$ and action sets $\Omega(j) = \{interruptions, noninterruptios\}$.

Let p^i be the probability that an interruption occur after $(i-1)^{th}$ interruption and $1 - p^i$ that of no interruption. $i.c$ be the cost due to i^{th} interruption and zero that of no interruption. l be the additional benefit due to an interruption (interruption here considered as leaving the current customer in service temporarily and serving another customer for higher income).

Total return up to state $1 = p(l - c)$.
Total return up to state $2 = p^2(l - 2c) + p(l - c)$
Continuing like this total return up to state N is
$F(N) = lp\left(\frac{1-p^N}{1-p}\right) - cp\left(\frac{1-p^N}{1-p} + \frac{p(1-Np^{N-1}+(N-1)p^N)}{(1-p)^2}\right)$
Optimal N is given by second derivative test in calculus and optimal value of N is $\frac{l+\sqrt{l^2-4c^2p}}{2c}$.

Chapter 3

On a Queue with Interruptions Controlled by a Super Clock and Maximum number of Interruptions

In this chapter we study an infinite capacity queueing system with single server. Two models are discussed. The general features are first indicated. A customer on arrival to an idle system immediately joins service. Arrival of customers constitute MAP. Each service may be subjected to interruption involving a maximum of N interruptions. Interruptions occur according to a Poisson process of rate γ and duration of each interruption is PH distributed. Repeat / resumption of interrupted customers service is decided by a comparison between duration of interruption and a threshold random variable which is also PH distributed. When the duration of interruption exceeds the threshold random variable the interrupted customer gets its service repeated on completion of interruption; otherwise it is resumed after the interruption is removed. During service period of a customer if the total duration of interruption exceeds a super threshold (also PH distributed) or total number of interruptions encountered by that customer reaches its peak (N), whichever occurs first, no further interruption is permitted . ie from that point on, the customer completes its service without any further interruption. Now we turn to the two specific models; In model I, it is assumed that with the realization of the super-

Some results of this chapter are included in the following papers: A. Krishnamoorthy, P.K. Pramod, and S.R. Chakravarthy, A note on characterizing interruptions with phase type distribution, European Journal of Operations Research (communicated),2009
It was also given as Invited talk by Krishamoothy, A, to the special session on Matrix Analytic Methods of INFORMS Applied Probability Conference, Cornel University, Ithaca, USA, July 2009.

clock, the interruption is removed. Instead the present interruption is allowed to continue; no further interruption is permitted. However in model II, immediately on realization of the super-clock, the current interruption is removed. Long run system state distribution is obtained under stable regime. Several performance measures are evaluated. Numerical illustrations of the system behavior in both cases are also provided. An optimization problem is numerically analyzed.

3.1. Model Description

As in the previous chapter, we assume here also that the process of interruption follows the type-I counter and bound on maximum number of interruptions. In addition to this upper bound on the number of interruptions, we bring in super-clock. The random/threshold clock and the duration of interruption of chapter 2 remain intact here also. The super-clock has a great role to play here. The two models considered here reveal the role of the super-clock. The process of interruption that befalls the customer in service is controlled as following. No further interruption is permitted to befall on the customer in service in case either the super-clock is realized or the maximum number of interruptions permitted is arrived at. In model I, we allow the current interruption to continue even if the super-clock is realized. In model 2, however, we insist that the interruption is removed the moment the super-clock is realized. The effects of these distinct assumptions on the performance of the system are quite revealing. More on super-clock and the models will be stated in the sequel.

In this section, we describe the queue with a single server subject to interruptions, has in addition to all components a classical queue, two random clocks. We call the first one the threshold clock as in the previous chapter, and the second, a super clock . The competition between interruption time and threshold clock decides whether to repeat/ resume the interrupted service. In case the interruption time exceeding threshold random variable then, on completion of interruption the interrupted service is repeated; otherwise it is resumed. The super threshold random variable acts like a control on the duration of interruptions. The super clock is set at position zero at the time a new service starts. The moment the first interruption strikes this starts to tick until the interruption is removed or the random variable describing the super clock is realized, whichever occurs first. The same customer is not subject to any further interruption in case the super clock is realized first. Else it can be interrupted, in which case the super clock starts from where it stopped ticking at the conclusion of the previous interruption of the present customer in service. The competition between interruption time and threshold random variable decides the repeat/ resumption of the interrupted ser-

vice. When interruption time exceeds threshold random variable, on completion of interruption, interrupted service is repeated; otherwise it is resumed. No more interruption takes place once the super clock reaches saturation point. Also the number of interruption reaches its maximum level, no more interruption is allowed. The detailed discussion of the two Queueing Model follows:
The Arrival and service processes are described as in chapter II. The arrival of customers to the queue is modelled by MAP with r phases represented by (D_0, D_1), D_0 and D_1 are square matrices of order r. The matrix D_0 has strictly negative diagonal entries and non negative off diagonal entries, and is invertible. Also $(D_0+D_1)e = 0$ where e is a column vector of appropriate order. The Service Process the absence of interruption service times of customers are independent phase type distributed with representation (α, S). Then it has probability density function $\alpha e^{xS}S^0$ and mean service time is $\alpha(-S)^{-1}e$. When the server is interrupted the service phase , where interruption occurred, is recorded. Since the convolution of two phase type distribution is again phase type, the total service time of a particular customer is also phase type distributed.

The status of the server is designated by $S(t)$ which takes 0 or 1 depending on whether server is in interrupted state or in busy mode , respectively . Here we specify once again the rules governing the interruption and the service after interruption. The components of interruption are: the interruption process, a threshold clock, a super clock, the maximum number of interruptions a customer may encounter during its service and the duration of an interruption.(a) The threshold clock and the service, on completion of interruption, are as described in chapter 2. The threshold random variable is PH distributed with representation (δ, U) .(b) The super clock acts like another threshold which is also called super threshold random variable, decides, in model II and partially in model I, the maximum time a customer can be interrupted . If super clock reaches its saturation point, then no more interruption takes place to the customer in service and interrupted customer completes its service on completion of present interruption (model I). The duration of the super-clock is a random variable which is PH distributed and having representation (η, L). (c) The onset of interruption forms a Poisson process with rate γ. (d) Duration of interruption is governed by a phase type distributed random variable.

Remarks:
At the strike of the first interruption to a customers service, threshold clock and super threshold starts ticking. The repeat or resumption of service on completion of current interruption is decided by the current status of interruption time and threshold clock. When interruption time exceeds threshold random variable the customer repeats its service else it is resumed. The super-clock ticking stopped at the instant interruption is removed. Nevertheless, at the next interruption epoch,

to the same customer, the super-clock starts ticking from where it stopped on completion of the previous interruption. Only either on its realization or the present customer leaving the system, the super-clock is set to zero. When super threshold random variable reaches its saturation level no more interruption allowed and the customer who is serving complete its service with out any interruption. If the number of interruption reaches its maximum level N, then no more interruption is allowed during the remaining service of the present customer. The minimum of these two, whichever occurs first gives the red signal for the interruption.

3.2. Model I

Let $N(t), S(t), B_1(t), S_1(t), S_2(t), S_3(t), S_4(t)$ and $M(t)$ denote the number of customers in the system, state of the server, the number of interruptions, the phase of the super clock, the service phase, the interruption phase, the phase of the threshold and the arrival phase, respectively. The process $\Omega = \{(N_1(t), S(t), S_1(t), S_2(t), S_3(t)$ $S_4(t), M(t)), t \geqslant 0\}$; is a continuous time Markov chain (CTMC) which turns out to be LIQBD with n^{th} level given by $\ell(n)=\bigcup_l \Psi(n,l)$, $l = 0, 1$. The subsets of $\Psi(n,l)$ are defined as $\{(n, 0, j_1, i_1, i_2, i_3, i_4, i_5); 0 \leqslant j_1 \leqslant N-1, 0 \leqslant i_1 \leqslant d, 1 \leqslant i_2 \leqslant a, 1 \leqslant i_3 \leqslant b, 0 \leqslant i_4 \leqslant c, 1 \leqslant i_5 \leqslant r\}$, $\{(n, 1, 0, i_2, i_4); 1 \leqslant i_2 \leqslant a, 1 \leqslant i_4 \leqslant r\}$ and $\{(n, 1, j_1, i_1, i_2, i_4); 0 \leqslant j_1 \leqslant N, 0 \leqslant i_1 \leqslant d, 1 \leqslant i_2 \leqslant a, 1 \leqslant i_4 \leqslant r\}$. The random variable $B(t)$ counts the number of interruptions excluding the current one if S(t)=0 and the completed number of interruption if S(t)=1. The threshold clock is set to zero position on completion of each interruption and start ticking at the beginning of a new interruption. The super clock counts the total interruption time (not completely,since the last interruption continues even the super clock is saturated) of a particular customer. The states in Ψ are listed in lexicographical order. We now describe the infinitesimal generator Q of this CTMC. Note that by the assumptions made above the CTMC $\{X(t), t \geqslant 0\}$ is a level independent quasi-birth and death process (LIQBD).

We have

$$Q = \begin{bmatrix} D_0 & C_0 & 0 & 0 & 0 & 0 & 0 \\ C_1 & A_1 & A_0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 \\ & & & & \cdot & \cdot & \cdot \end{bmatrix} \tag{3.1}$$

The matrix A_0 describes the arrival of customers to the system. The only non-zero elements in A_0 are $I_{N(d+1)ab(c+1)} \otimes D_1$ and $I_{(N(d+1)+1)a} \otimes D_1$ which record the transition rates to $\Psi(n+1,0)$, $\Psi(n+1,1)$ starting from $\Psi(n,0)$ and $\Psi(n,1)$, respectively. The matrix A_2 records transition rates corresponds to departure of customer after completing the service. The only non zero block matrix in A_2 is $[A_2^{(1)} \quad [0]]$ which represents the transition from busy state to itself . The matrix $A_2^{(1)} = e_{(N(d+1)+1)} \otimes (S^0 \otimes \alpha \otimes I_r)$ records transition rates in $\Psi(n-1,1)$ starting from $\Psi(n,1)$. Since there is no departure in other transitions, they are listed as zero matrices.

Now we describe the matrix A_1 which records transaction within $\Psi(n,.)$ from $\Psi(n,.)$. The components in A_1 are A_{11}, A_{12}, A_{13}, A_{14} which record transitions from $\Psi(n,0)$ to $\Psi(n,0)$; $\Psi(n,0)$ to $\Psi(n,1)$; $\Psi(n,1)$ to $\Psi(n,0)$; $\Psi(n,1)$ to $\Psi(n,1)$, respectively. The matrix $I_N \otimes H_3$ records transactions in $\Psi(n,0)$ starting from $\Psi(n,0)$ where $H_3 = G_4 \oplus H_1$, and $H_1 = G_2 \oplus G_1$. Here the matrix $G_1 = F_1 \oplus D_0$, $G_2 = T \otimes I_{(c+1)}$, and $G_4 = F_2 \otimes I_{ab(c+1)}$ where $F_1 = \begin{bmatrix} 0 & \bar{0} \\ U^0 & U \end{bmatrix}$ and $F_2 = \begin{bmatrix} 0 & \bar{0} \\ L^0 & L \end{bmatrix}$. The matrix $A_{12} = \begin{bmatrix} [0] & I_N \otimes I_a \otimes B \end{bmatrix}$ lists the rate correspond to repeat/ resumption of service on completion of interruption with $B = [B_{11}, B_{12}, \ldots\ldots, B_{1b}]$ and $B_{1j} = T^0 \otimes [\bar{\alpha}, \bar{e}, \bar{e}, \ldots\ldots \bar{e}]' \otimes I_r$. $A_{13} = \left[[0] \;\; diag[A_{13}^{(1)}, A_{13}^{(2)}\right]$ where $diag[A_{13}^{(1)}, A_{13}^{(2)}]$ is a diagonal matrix with diagonal elements $A_{13}^{(1)}$ and $A_{13}^{(2)}$ which represent transitions corresponding to busy state to interruption state; $A_{13}^{(1)} = \gamma I_a \otimes \bar{\eta} \otimes \beta \otimes \widehat{\delta} \otimes I_r$ and $A_{13}^{(2)} = I_{N-1} \otimes \gamma \begin{bmatrix} [0] & I_d \end{bmatrix}' \otimes I_a \otimes \beta \otimes \widehat{\delta} \otimes I_r$. Since the saturation of super threshold clock just after interruption is impossible, it is represented by the zero block matrix.

Now turning to A_{14}: we have $A_{14} = diag\begin{bmatrix} A_{14}^{(1)} & A_{14}^{(2)} & A_{14}^{(3)} \end{bmatrix}$ the elements $A_{14}^{(1)}$, $A_{14}^{(2)}$ and $A_{14}^{(3)}$ which record transitions from $\Psi(n,1)$ to itself. The matrices $A_{14}^{(1)}, A_{14}^{(2)}, A_{14}^{(3)}$ records transitions in $\Psi(n,1)$ which correspond to service of a customer whose is not interrupted so far; i times interrupted, $1 \leqslant i \leqslant N-1$ and and N times interrupted , respectively with $A_{14}^{(1)} = S \oplus D_0 - \gamma I_{ar}$, $A_{14}^{(2)} = diag\begin{bmatrix} S \oplus D_0, I_d \otimes S \oplus D_0 - \gamma I_{ar} \end{bmatrix}$. The matrix $A_{14}^{(3)} = I_{d+1} \otimes [S \oplus D_0]$. The matrix A_{14} corresponds to transitions when the server is busy. $A_{14}^{(1)}$ denotes transition in busy state when super threshold reaches its saturation level. When super threshold is saturated we make sure that no further interruption occurs. The matrix $A_{14}^{(2)}$ corresponds to transitions when super threshold is not saturated. $A_{14}^{(3)}$ describes transition rates in busy state when the number of interruption has reached maximum level.

The matrices D_0, C_0 and C_1 are, respectively the transitions from level 0 to 0; level 0 to 1 and level 1 to 0 where $C_0 = \begin{bmatrix} [0] & \alpha \otimes D_1 & [0] \end{bmatrix}$ and $C_1 = \begin{bmatrix} [0] \\ e_{(N(d+1)+1)} \otimes (S^{(0)} \otimes I_r \end{bmatrix}$.

Now we turn to the analysis of Model I. Here a customer in service without be subject to further interruptions on realization on the super-clock/maximum number of interruptions which one occurs first. However the present interruption is allowed to be continued even on realization of the super-clock. First we analyze the service time of a customer . By this we mean, the time duration starting with admission to service counter up to leaving the system on completion of service. Thus this includes the intervening interruptions and consequent repeat/resumption of service.

3.3. Description of the phase type distribution for the services

The focus of this section is to describe the time it takes to process a job once it enters into the service facility. We assume that the service times are of phase type with representation given by $(\boldsymbol{\alpha}, S)$ of order a. The services are subject to interruptions and the interruption process is assumed to follow a Poisson process with rate γ. When the current service is interrupted for the first time three clocks, referred to as (a) super clock, (b) interruption clock, and (c) threshold clock, will simultaneously be started. The durations of these clocks are of phase type with representations given by, respectively, $(\boldsymbol{\eta}, L)$ of order d, $(\boldsymbol{\beta}, T)$ of order b, and $(\boldsymbol{\delta}, U)$ of order c. Once the interruption clock expires the service of the interrupted job will begin again. The service will resume (from the phase where the service got interrupted) or repeat (like a new service) depending on whether the interrupted clock expired before the threshold clock or not. This is irrespective of whether the super-clock expired or not. However, if the super-clock expired before the interruption clock, then the service of the current job will not be interrupted anymore once the service begins again for this job. On the other hand, if the interruption clock expires before the super clock, the phase of the super-clock will be frozen and will resume from this phase should there be an another interruption for the existing job. For the job under service, the number of interruptions will be tracked and when this number attains a pre-specified threshold value, $N < \infty$, no further interruptions are allowed. Thus, the super clock will play no role (assuming it is not yet expired) from the moment when the service begins for the current job from the $(N-1)^{th}$ interruption.

Note that a phase type distribution is defined as the time until absorption in a finite state irreducible Markov chain with one absorbing state. For details on PH distributions and their properties, we refer the reader to [25].

Let X denote the duration of the effective service for a job. That is, X is the time between the arrival of a job to the service facility until it leaves the facility with a service. Note that this service time may possibly have i, $0 \leqslant i \leqslant N$

interruptions, and each interruption may cause the service to either resume from where it got interrupted or repeat from the beginning. Thus, we refer to X as the effective service time to distinguish this from the service time given by $(\boldsymbol{\alpha}, S)$. Note that when $\gamma = 0$ (that is, when there are no interruptions) then the effective service time is same as the service time. We will show that X follows a PH-distribution. Towards this end, we first define $J_1(t), J_2(t), J_3(t), J_4(t)$, and $J_5(t)$, respectively, to be the number of interruptions seen by the current job in service, the phase of the super clock, the phase of the current service, the phase of the interruption clock, and the phase of the threshold clock, at time t. Note that some of these phases will be frozen or not defined. For example, when the service is going on, the phases of interruption clock and the threshold clock will not be defined as they are not turned on, and the super clock may not be defined or when defined it will be frozen. Let $*$ denote the absorbing state that corresponds to the completion of the current service.

The states and their description are given in Table 1 below. For use in sequel, we now define sets of states as follows.

- $\boldsymbol{i^*} = \{(i^*, j_2) : 1 \leqslant j_2 \leqslant a\}$, for $1 \leqslant i^* \leqslant N-1$
- $\boldsymbol{0} = \{(0, j_2) : 1 \leqslant j_2 \leqslant a\}$
- $\boldsymbol{i} = \{(i, j_1, j_2) : 1 \leqslant j_1 \leqslant d, 1 \leqslant j_2 \leqslant a\}$, for $1 \leqslant i \leqslant N-1$
- $\hat{\imath} = \{(\hat{\imath}, j_2, j_3) : 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b\}$, for $1 \leqslant \hat{\imath} \leqslant N-1$
- $\hat{\hat{\imath}} = \{(\hat{\hat{\imath}}, j_2, j_3, j_4) : 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b, 1 \leqslant j_4 \leqslant c\}$, for $1 \leqslant \hat{\hat{\imath}} \leqslant N-1$
- $\tilde{\imath} = \{(\tilde{\imath}, j_1, j_2, j_3) : 1 \leqslant j_1 \leqslant d, 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant c\}$, for $1 \leqslant \tilde{\imath} \leqslant N-1$
- $\tilde{\tilde{\imath}} = \{(\tilde{\tilde{\imath}}, j_1, j_2, j_3, j_4) : 1 \leqslant j_1 \leqslant d, 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b, 1 \leqslant j_4 \leqslant c\}$, for $1 \leqslant \tilde{\tilde{\imath}} \leqslant N-1$
- $\bar{N} = \{(\bar{N}, j_2) : 1 \leqslant j_2 \leqslant a\}$
- $\bar{\bar{N}} = \{(\bar{\bar{N}}, j_2, j_3, j_4) : 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b, 1 \leqslant j_4 \leqslant c, \}$
- $\bar{\bar{\bar{N}}} = \{(\bar{\bar{\bar{N}}}, j_2, j_3) : 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b\}$

Table 1: The states and their description

States	Description
$\{(i^*, j_2) :$ $1 \leqslant j_2 \leqslant a,$ $1 \leqslant i^* \leqslant N-1\}$	The super clock expired and the current service is in phase j; no more interruptions from these states possible. The number of interruptions for the current service is i^*.
$\{(\bar{N}, j_2) : 1 \leqslant j_2 \leqslant a\}$	The number of interruptions is at its maximum with the current service in phase j_2.
$\{0, j) : 1 \leqslant j_2 \leqslant a\}$	The current service has seen no interruptions so far and the service is in phase j_2.
$\{(i, j_1, j_2) : 1 \leqslant j_1 \leqslant d,$ $1 \leqslant j_2 \leqslant a,$ $1 \leqslant i \leqslant N-1\}$	The phase of the service is in state j_2 with the super clock frozen in state j_1 and the number of interruptions so far is i.
$\{(\hat{i}, j_2, j_3) : 1 \leqslant j_2 \leqslant a,$ $1 \leqslant j_3 \leqslant b,$ $1 \leqslant \hat{i} \leqslant N-1\}$	The super as well as threshold clocks have expired with the interruption clock in state j_3; the service phase is frozen in j_2 and the number of interruptions including the current one is $\hat{i}$. The service will be repeated from these states.
$\{(\hat{\hat{i}}, j_2, j_3, j_4) : 1 \leqslant j_2 \leqslant a,$ $1 \leqslant j_3 \leqslant b, 1 \leqslant j_4 \leqslant c,$ $1 \leqslant \hat{\hat{i}} \leqslant N-1\}$	The super clock has expired with the interruption clock in state j_3 and the threshold clock in j_4; the service phase is frozen in state j_2, and the number of interruptions including the current one is $\hat{\hat{i}}$.
$\{(\tilde{i}, j_1, j_2, j_3) : 1 \leqslant j_1 \leqslant d,$ $1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b$ $1 \leqslant i \leqslant N-1\}$	The threshold clock has expired with the super in phase j_1, the interruption clock is in state j_3; the service phase is frozen in j_2 and the number of interruptions including the current one is $\tilde{i}$..
$\{(\tilde{\tilde{i}}, j_1, j_2, j_3, j_4) : 1 \leqslant j_1 \leqslant d,$ $1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b,$ $1 \leqslant j_4 \leqslant c, 1 \leqslant \tilde{\tilde{i}} \leqslant N-1\}$	The super clock is in state j_1, the interruption clock in state j_3,the threshold clock in j_4; the service phase is frozen in state j_2, and the number of interruptions including the current one is $\tilde{\tilde{i}}$.
$\{(\bar{\bar{N}}, j_2, j_3, j_4) : 1 \leqslant j_2 \leqslant a,$ $1 \leqslant j_3 \leqslant b, 1 \leqslant j_4 \leqslant c\}$	The service phase is frozen in state j_2 with the number of interruptions at its maximum limit of N; both interruption and threshold clocks are in phase j_3, and j_4, respectively.
$\{(\bar{\bar{\bar{N}}}, j_2, j_3) : 1 \leqslant j_2 \leqslant a,$ $1 \leqslant j_3 \leqslant b\}$	The service phase is frozen in state j_2 with the number of interruptions at its maximum limit of N, and the interruption clock is in phase j_3.

For use in sequel, let $e(r)$, $e_j(r)$ and I_r denote, respectively, the (column) vector of dimension r consisting of 1's, column vector of dimension r with 1 in the j^{th} position and 0 elsewhere, and an identity matrix of dimension r. When there is no need to emphasize the dimension of these vectors we will suppress the suffix. Thus, e will denote a column vector of 1's of appropriate dimension. The notation "$\prime$" appearing in a matrix will stand for the matrix transpose. The notation $\otimes$ will stand for the Kronecker product of two matrices, and $\oplus$ stands for the Kronecker sum of two matrices. Thus, if A is a matrix of order $m \times n$ and if B is a matrix of order $p \times q$, then $A \otimes B$ will denote a matrix of order $mp \times nq$ whose $(i,j)^{th}$ block matrix is given by $a_{ij}B$. The Kronecker sum, $A \oplus B$ of dimension mp is defined by $A \otimes I + I \otimes B$. For more details on Kronecker products and sums, we refer the reader to [4], [24].

The Markov process $\{J_1(t), J_2(t), J_3(t), J_4(t), J_5(t)) : t \geqslant 0\}$ with the absorbing state $*$ is defined on the state space

$$\Omega = \{i^* : 1 \leqslant i^* \leqslant N\} \bigcup \{\hat{N}\} \bigcup \{0\} \bigcup \{i : 1 \leqslant i \leqslant N-1\} \bigcup \{\hat{i} : 1 \leqslant \hat{i} \leqslant N\}$$
$$\bigcup \{\hat{\hat{i}} : 1 \leqslant \hat{\hat{i}} \leqslant N\} \bigcup \{\tilde{i} : 1 \leqslant \tilde{i} \leqslant N\} \bigcup \{\tilde{\tilde{i}} : 1 \leqslant \tilde{\tilde{i}} \leqslant N\} \bigcup \{*\}$$

and its infinitesimal generator is given by

$$Q = \begin{pmatrix} \tilde{T} & \tilde{\boldsymbol{T}}^{\boldsymbol{0}} \\ 0 & 0 \end{pmatrix}, \tag{1}$$

where

$$\tilde{T} = \begin{pmatrix}
T_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & T_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & T_{3,3} & 0 & 0 & 0 & 0 & T_{3,8} & 0 & 0 \\
0 & 0 & 0 & T_{4,4} & 0 & 0 & 0 & T_{4,8} & T_{4,9} & 0 \\
T_{5,1} & 0 & 0 & 0 & T_{5,5} & 0 & 0 & 0 & 0 & 0 \\
T_{6,1} & 0 & 0 & 0 & T_{6,5} & T_{6,6} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & T_{7,4} & T_{7,5} & 0 & T_{7,7} & 0 & 0 & 0 \\
0 & 0 & 0 & T_{8,4} & 0 & T_{8,6} & T_{8,7} & T_{8,8} & 0 & 0 \\
0 & T_{9,2} & 0 & 0 & 0 & 0 & 0 & 0 & T_{9,9} & T_{9,10} \\
0 & T_{10,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & T_{10,10}
\end{pmatrix}, \tag{2}$$

with

$$T_{1,1} = I \otimes S,\ T_{2,2} = S,\ T_{3,3} = S - \gamma I,\ T_{3,8} = \gamma(\boldsymbol{e}_1' \otimes \boldsymbol{\eta} \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta}),$$

$$T_{4,4} = I \otimes (S - \gamma I),\ T_{4,8} = \gamma(\hat{I}_{N-1} \otimes I \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta}),\ T_{4,9} = \gamma(\boldsymbol{e} \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta}),$$

$$T_{5,1} = I \otimes \boldsymbol{e} \otimes \boldsymbol{T}^{\boldsymbol{0}} \boldsymbol{\alpha},\ T_{5,5} = I \otimes I \otimes T),\ T_{6,1} = I \otimes I \otimes \boldsymbol{T}^{\boldsymbol{0}} \otimes \boldsymbol{e},\ T_{6,5} = I \otimes I \otimes I \otimes \boldsymbol{U}^{\boldsymbol{0}},$$

$$T_{6,6} = I \otimes I \otimes (T \oplus U,\ T_{7,4} = I \otimes I \otimes \boldsymbol{e} \boldsymbol{\alpha} \otimes \boldsymbol{T}^{\boldsymbol{0}},\ T_{7,5} = I \otimes \boldsymbol{L}^{\boldsymbol{0}} \otimes I \otimes I,$$

$$T_{7,7} = I \otimes [L \oplus (I \otimes T)],\ T_{8,4} = I \otimes I \otimes I \otimes \boldsymbol{T}^{\boldsymbol{0}} \otimes \boldsymbol{e},\ T_{8,6} = I \otimes \boldsymbol{L}^{\boldsymbol{0}} \otimes I \otimes I \otimes I,$$

$$T_{8,7} = I \otimes I \otimes I \otimes I \otimes \boldsymbol{U}^{\boldsymbol{0}},\ T_{8,8} = I \otimes [L \oplus (I \otimes (T \oplus U))],\ T_{9,2} = I \otimes \boldsymbol{T}^{\boldsymbol{0}} \otimes \boldsymbol{e},$$

$$T_{9,9} = I \otimes (T \oplus U),\ T_{9,10} = I \otimes I \otimes I \otimes \boldsymbol{U}^{\boldsymbol{0}},\ T_{10,2} = \boldsymbol{e} \boldsymbol{\alpha} \otimes \boldsymbol{T}^{\boldsymbol{0}},\ T_{10,10} = I \otimes T, \tag{3}$$

where

$$\tilde{T}^0 = \begin{pmatrix} e\otimes S^0 \\ S^0 \\ S^0 \\ e\otimes S^0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \hat{I}_{N-1} = \begin{pmatrix} 0 & I_{N-2} \\ 0 & 0 \end{pmatrix},$$

Theorem 3.3.1. *The effective service time, X, is of phase type with representation $(\boldsymbol{\zeta}, \tilde{T})$ of order $(N+1)a + Nab(c+1) + (N-1)da[1 + b(c+1)]$, where*

$$\boldsymbol{\beta} = (0, 0, \boldsymbol{\alpha}, 0, 0, 0, 0, 0, 0, 0)\ , \tag{4}$$

and $\tilde{T}$ is as given in (2).

PROOF: First note that a new service will begin in level **0** in state $(0, j_2)$ with probability given by $\alpha_{j_2}(2)$. Once the service begins it can end with or without interruptions and looking through all possible transitions, one will see that the transition matrix is given by $\tilde{T}$ as given in (2). Thus, the service time is nothing but the time until absorption into state $*$ starting from the level **0**. This results in the form of the initial probability vector as given in (4).

Now will show how the mean, $\mu_{\tilde{T}'}$, and the standard deviation, $\sigma_{\tilde{T}}$, of $(\boldsymbol{\zeta}, \tilde{T})$ can be computed recursively (and explicitly). Recall ([25]) that the mean and standard deviation of X is given by

$$\mu_{\tilde{T}'} = \boldsymbol{\zeta}(-\tilde{T})^{-1}\boldsymbol{e}, \qquad \sigma_{\tilde{T}} = \sqrt{[2\boldsymbol{\zeta}(-\tilde{T})^{-2}\boldsymbol{e} - (\mu_{\tilde{T}'})^2]}. \tag{6}$$

Due to the special structure of the matrix $\tilde{T}$ given in (2), we can compute the mean as well as the standard deviation of $\tilde{T}$ explicitly and recursively. First, we define

$$\boldsymbol{\zeta}(-\tilde{T})^{-1} = (\boldsymbol{u}, \hat{\boldsymbol{u}}, \boldsymbol{v}_0, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{a}_N, \boldsymbol{b}_N). \tag{7}$$

We further split the vectors on the right side of (7) as

$$\begin{array}{l} \boldsymbol{u} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N),\ \boldsymbol{v} = (\boldsymbol{v}_1, \cdots, \boldsymbol{u}_{N-1}),\ \boldsymbol{w} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{N-1}) \\ \boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{N-1}),\ \boldsymbol{y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{N-1}),\ \boldsymbol{z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_{N-1}). \end{array} \tag{8}$$

Note that the vectors $\boldsymbol{u}_i, 1 \leqslant i \leqslant N$, $\hat{\boldsymbol{u}}$, and $\boldsymbol{v}_0$ are of dimension a; $\boldsymbol{v}_i, 1 \leqslant i \leqslant N-1$ are of order da; $\boldsymbol{w}_i, 1 \leqslant i \leqslant N-1$ are of order ab; $\boldsymbol{x}_i, 1 \leqslant i \leqslant N-1$ are of order

abc; $\boldsymbol{y}_i, 1 \leqslant i \leqslant N-1$ are of order dbc; $\boldsymbol{z}_i, 1 \leqslant i \leqslant N-1$ are of order $dabc$; $\boldsymbol{a}_N$ is of order abc and $\boldsymbol{b}_N$ is of order ab.

Exploiting the special structure of $\tilde{T}$ and using the notations in (7) and (8), it is easy to verify the following equations.

$$\begin{aligned}
\boldsymbol{v}_0 &= \boldsymbol{\alpha}[\gamma I - S]^{-1}, \\
\boldsymbol{z}_1 &= \gamma \boldsymbol{v}_0[\boldsymbol{\eta} \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta}][-(L \oplus (I \otimes (T \oplus U)))]^{-1}, \\
\boldsymbol{y}_1 &= \boldsymbol{z}_1[I \otimes I \otimes I \otimes \boldsymbol{U^0}][-(L \oplus (I \otimes T))]^{-1}, \\
\boldsymbol{x}_1 &= \boldsymbol{z}_1[\boldsymbol{L^0} \otimes I \otimes I \otimes I][-(I \otimes (T \oplus U))]^{-1}, \\
\boldsymbol{w}_1 &= [\boldsymbol{x}_1(I \otimes I \otimes \boldsymbol{U^0}) + \boldsymbol{y}_1(\boldsymbol{L^0} \otimes I \otimes I)][-(I \otimes T)]^{-1}, \\
\boldsymbol{u}_1 &= [\boldsymbol{w}_1(\boldsymbol{e} \otimes \boldsymbol{T^0}\boldsymbol{\alpha}) + \boldsymbol{x}_1(I \otimes \boldsymbol{T^0} \otimes \boldsymbol{e})][-S]^{-1}, \\
\boldsymbol{v}_i &= [\boldsymbol{y}_1(I \otimes \boldsymbol{e}\boldsymbol{\alpha} \otimes \boldsymbol{T^0}) + \boldsymbol{z}_i(I \otimes I \otimes \boldsymbol{T^0} \otimes \boldsymbol{e})][I \otimes (\gamma I - S)]^{-1}, \ 1 \leqslant i \leqslant N-1, \\
\boldsymbol{z}_i &= \gamma \boldsymbol{v}_{i-1}[I \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta}][-(L \oplus (I \otimes (T \oplus U)))]^{-1}, \ 2 \leqslant i \leqslant N-1, \\
\boldsymbol{y}_i &= \boldsymbol{z}_i[I \otimes I \otimes I \otimes \boldsymbol{U^0}][-(L \oplus (I \otimes T))]^{-1}, \ 2 \leqslant i \leqslant N-1, \\
\boldsymbol{x}_i &= \boldsymbol{z}_i[\boldsymbol{L^0} \otimes I \otimes I \otimes I][-(I \otimes (T \oplus U))]^{-1}, \ 2 \leqslant i \leqslant N-1, \\
\boldsymbol{w}_i &= [\boldsymbol{x}_i(I \otimes I \otimes \boldsymbol{U^0}) + \boldsymbol{y}_i(\boldsymbol{L^0} \otimes I \otimes I)][-(I \otimes T)]^{-1}, \ 2 \leqslant i \leqslant N-1, \\
\boldsymbol{u}_i &= [\boldsymbol{w}_i(\boldsymbol{e} \otimes \boldsymbol{T^0}\boldsymbol{\alpha}) + \boldsymbol{x}_i(I \otimes \boldsymbol{T^0} \otimes \boldsymbol{e})][-S]^{-1}, \ 2 \leqslant i \leqslant N-1, \\
\boldsymbol{a}_N &= \gamma[\boldsymbol{v}_{N-1}(\boldsymbol{e} \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta})][-(I \otimes (T \oplus U))]^{-1}, \\
\boldsymbol{b}_N &= [\boldsymbol{a}_N(I \otimes I \otimes \boldsymbol{U^0})[-(I \otimes T]^{-1}, \\
\boldsymbol{u}_N &= \boldsymbol{a}_N(I \otimes \boldsymbol{T^0} \otimes \boldsymbol{e}) + \boldsymbol{b}_N(\boldsymbol{e}\boldsymbol{\alpha} \otimes \boldsymbol{T^0}).
\end{aligned} \tag{9}$$

By looking at the order in which the equations are displayed in (9), one can see the explicit evaluation of the vectors needed in the computation of μ'_T, which is obtained as

$$\mu'_T = \boldsymbol{u}\boldsymbol{e} + \hat{\boldsymbol{u}}\boldsymbol{e} + \boldsymbol{v}_0\boldsymbol{e} + \boldsymbol{v}\boldsymbol{e} + \boldsymbol{w}\boldsymbol{e} + \boldsymbol{x}\boldsymbol{e} + \boldsymbol{y}\boldsymbol{e} + \boldsymbol{z}\boldsymbol{e} + \boldsymbol{a}_N\boldsymbol{e} + \boldsymbol{b}_N\boldsymbol{e}. \tag{10}$$

Similar to getting an explicit expression for $\boldsymbol{\zeta}(-\tilde{T})^{-1}$ one can derive an explicit expression for $\boldsymbol{\zeta}(-\tilde{T})^{-2}$ by replacing the role played by β in (9) with $(\boldsymbol{u}, \hat{\boldsymbol{u}}, \boldsymbol{v}_0, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{a}_N, \boldsymbol{b}_N$

The details are omitted.

Note that some of the explicit expressions in (9) involve matrices whose orders may become very large as the values of d, a, b, and c become large. In that case one can exploit the structure of these matrices. Some examples to this extent can be seen in [25]. In any computational aspect, it is very important to have some internal accuracy checks. In the case of the computation of $\mu_{\tilde{T}'}$, one can use the fact that $\boldsymbol{\zeta}(-\tilde{T})^{-1}\tilde{\boldsymbol{T}}^{\mathbf{0}} = 1$, which reduces to

$$\left[\hat{\boldsymbol{u}} + \boldsymbol{v}_0 + \sum_{i=1}^{N} \boldsymbol{u}_i\right] \boldsymbol{S}^{\mathbf{0}} + \sum_{i=1}^{N-1} \boldsymbol{v}_i(\boldsymbol{e} \otimes \boldsymbol{S}^{\mathbf{0}}) = 1. \tag{11}$$

In the case of the computation of $\sigma_{\tilde{T}}$, one can use the fact that $\boldsymbol{\zeta}(-\tilde{T})^{-2}\tilde{\boldsymbol{T}}^{\mathbf{0}} = \mu_{\tilde{T}'}$.

3.4. Numerical Examples

In this section we discuss some interesting numerical examples that qualitatively describe the phase type distribution modeling the interrupted services. The correctness and the accuracy of the implementation of the recursive and explicit schemes are verified by a number of accuracy checks such as the one listed in(11). As an additional accuracy check, we obtained the numerical solution for the exponential clocks' case in their simple forms. Next, we implemented the general algorithm, but using the following PH representation: Let R be an irreducible, stable matrix with eigenvalue of maximum real part -$\theta < 0$. Let $\boldsymbol{\alpha}$ denote the corresponding left eigenvector, normalized by $\boldsymbol{\alpha e}$=1. The PH representation $(\boldsymbol{\alpha}, R)$ reduces to the exponential distribution with rate θ. The general algorithm does not utilize this fact in any manner, but the numerical results agreed very much.

In addition to the measures, $\mu_{\tilde{T}'}$ and $\sigma_{\tilde{T}}$, given in (6) there are other measures that can be constructed. A few are listed below along with their formulas.

1. Probability of i interruptions without the super clock expiring: The probability, P_i^{SCNX} that exactly i interruptions occur during a service without the super clock expiring is given by

$$P_i^{SCNX} = \begin{cases} \boldsymbol{v}_0\boldsymbol{S}^{\mathbf{0}}, & i = 0, \\ \boldsymbol{v}_i(\boldsymbol{e} \otimes \boldsymbol{S}^{\mathbf{0}}), & 1 \leqslant i \leqslant N-1, \\ \hat{\boldsymbol{u}}\boldsymbol{S}^{\mathbf{0}}, & i = N. \end{cases}$$

2. Probability of i interruptions with the super clock expiring: The probability, P_i^{SCEX} that exactly i interruptions occur during a service with the super clock expiring is given by

$$P_i^{SCEX} = \boldsymbol{u}_i \boldsymbol{S^0}, \quad 1 \leqslant i \leqslant N-1.$$

3. The mean number of interruptions during a service completion: The mean, μ_{IPS}, number of interruptions that occur during a service completion is given by

$$\mu_{IPS} = \sum_{i=1}^{N-1} i[P_i^{SCNX} + P_i^{SCEX}] + NP_N^{SCNX}.$$

4. Probability of a service completion in which neither super clock expires nor the maximum interruptions occur: The probability that a service completion occurs without the super clock expiring as well as the maximum number of interruptions allowed (which is N) is not attained is given by

$$P^{SCNXNI} = 1 - [\sum_{i=1}^{N} \boldsymbol{u}_i \boldsymbol{S^0} + \hat{\boldsymbol{u}} \boldsymbol{S^0}].$$

5. Probability of a service completion with no interruption: The probability that a service is completed without any interruptions is given by

$$P^{SCNI} = \boldsymbol{\alpha}[\gamma I - S]^{-1} \boldsymbol{S^0}.$$

Note that the above probability is nothing but $P(X < Y)$ where X follows a phase type distribution with representation $(\boldsymbol{\alpha}, S)$ and Y is exponential with parameter γ.

We consider the following set of PH-distributions for our numerical examples. For specific example under consideration, we will identify the choice for each of the four input PH-distributions as follows. By $ER - SC$ we will denote that the super clock (SC) has a PH-distribution given by ER (Erlang of order 5). We use ST for the service time; IP for the interruption clock and TC for the threshold clock.
Erlang of order 5(ER):

$$\boldsymbol{\alpha} = (1,0,0,0,0), S = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Exponential(EX):

$$\boldsymbol{\alpha} = 1, S = \begin{pmatrix} -1 \end{pmatrix}$$

Hyperexponential(HE):

$$\boldsymbol{\alpha} = (0.6, 0.3, 0.1), S = \begin{pmatrix} -100 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

All these three $PH-$distributions will be normalized so as to have a specific (given) mean. However, these are qualitatively different. For example, the coefficient of variation of these three distributions ER, EX, and HE are, respectively,0.44721, 1, and 3.18497.

EXAMPLE 1: The purpose of this example is to see the effect of the variability in the four PH-distributions on the four measures: $\mu_{\tilde{T}'}, \sigma_{tildeT}, \mu_{IPS}$, and σ_{IPS}. Here we fix $N = 20, \gamma = 1, \mu_1' = 4.0, \mu_2' = 1, \mu_3' = 2$, and $\mu_4' = 2$. These measures are displayed for various combinations of the distributions for the super clock, the service time, the interruption time, and the threshold clock in Table 2 below.
An examination of this table reveals the following.

- When service times have less variability (like Erlang), the four measures: the effective service mean, the standard deviation of the effective service time, the mean and the standard deviation of the number of interruptions appear to increase with increasing variability in the threshold times. This observation appears to be true for all the distributions considered for the interruption and for the super-clock. However, for service times having more variability (like hyperexponential), these two measures appear to decrease with increasing variability in the threshold times. This observation appears to be true for all the distributions considered for the interruption and for the super-clock.

- In the case of exponential service time, the four measures under consideration are insensitive to the distribution for the threshold clock. However, this measure depends on the type of distributions assumed for the super-clock and the interruption clock. This is to be expected due to the memoryless property of the service times.

Table 2: Measures for various combinations

			ER-IP			EX-IP			HE-IP		
			ER-TC	EX-TC	HE-TC	ER-TC	EX-TC	HE-TC	ER-TC	EX-TC	HE-TC
ER-SC	ER-ST	$\mu_{\bar{T}'}$	3.05337	3.12978	3.28708	2.98057	3.06744	3.26830	2.97196	3.05405	3.31578
		$\sigma_{\bar{T}}$	2.47212	2.53038	2.66605	2.87395	2.93116	3.07304	6.49081	6.58266	6.86211
		μ_{IPS}	0.94294	0.96430	1.00453	0.92254	0.94878	1.00591	0.95325	0.97947	1.06198
		σ_{IPS}	0.97088	0.99923	1.04790	0.97220	1.01295	1.09577	1.03613	1.08319	1.22449
	EX-ST	$\mu_{\bar{T}'}$	2.57423	2.57423	2.57423	2.59812	2.59812	2.59812	2.77782	2.77782	2.77782
		$\sigma_{\bar{T}}$	2.70037	2.70037	2.70037	3.06655	3.06655	3.06655	6.46701	6.46701	6.46701
		μ_{IPS}	0.78711	0.78711	0.78711	0.79906	0.79906	0.79906	0.88891	0.88891	0.88891
		σ_{IPS}	0.95976	0.95976	0.95976	1.01090	1.01090	1.01090	1.20281	1.20281	1.20281
	HE-ST	$\mu_{\bar{T}'}$	1.30908	1.23914	1.10992	1.38909	1.30249	1.14428	1.77407	1.60960	1.30768
		$\sigma_{\bar{T}}$	2.81346	2.63022	2.22823	3.07857	2.89023	2.49919	5.80647	5.42487	4.73198
		μ_{IPS}	0.37569	0.36102	0.33703	0.40831	0.38581	0.34847	0.55250	0.50482	0.41578
		σ_{IPS}	0.7433	0.70526	0.64523	0.85719	0.79074	0.68502	1.37912	1.20564	0.89689
EX-SC	ER-ST	$\mu_{\bar{T}'}$	2.8924	2.95931	3.10183	2.86836	2.94350	3.12238	2.89777	2.96940	3.20166
		$\sigma_{\bar{T}}$	2.42314	2.48900	2.65168	2.83986	2.89663	3.04900	6.37279	6.45101	6.69785
		μ_{IPS}	0.86983	0.88782	0.92355	0.87098	0.89255	0.94159	0.91767	0.93944	1.00995
		σ_{IPS}	0.94463	0.97608	1.03644	0.94029	0.97917	1.06393	0.99477	1.03568	1.16277
	EX-ST	$\mu_{\bar{T}'}$	2.45024	2.45024	2.45024	2.50000	2.50000	2.50000	2.69414	2.69414	2.69414
		$\sigma_{\bar{T}}$	2.6182	2.61820	2.61820	3.00000	3.00000	3.00000	6.31385	6.31385	6.31385
		μ_{IPS}	0.72512	0.72512	0.72512	0.75000	0.75000	0.75000	0.84707	0.84707	0.84707
		σ_{IPS}	0.92319	0.92319	0.92319	0.96825	0.96825	0.96825	1.14259	1.14259	1.14259
	HE-ST	$\mu_{\bar{T}'}$	1.29838	1.23368	1.11010	1.37718	1.29683	1.14427	1.73074	1.58556	1.30352
		$\sigma_{\bar{T}}$	2.86132	2.69388	2.31756	3.11226	2.93506	2.55405	5.65774	5.32648	4.69555
		μ_{IPS}	0.34814	0.33631	0.31701	0.38123	0.36327	0.33267	0.51298	0.47507	0.40079
		σ_{IPS}	0.70075	0.66379	0.60654	0.80665	0.74659	0.65134	1.27432	1.12518	0.85756
HE-SC	ER-ST	$\mu_{\bar{T}'}$	2.56094	2.60246	2.69460	2.57985	2.62636	2.74136	2.68449	2.73126	2.88733
		$\sigma_{\bar{T}}$	2.03962	2.08850	2.22463	2.53204	2.56863	2.68227	6.01089	6.05561	6.20928
		μ_{IPS}	0.71925	0.72772	0.74551	0.73822	0.74856	0.77361	0.81541	0.82688	0.86682
		σ_{IPS}	0.76376	0.78754	0.83871	0.77553	0.80200	0.86540	0.85985	0.88514	0.96963
	EX-ST	$\mu_{\bar{T}'}$	2.19963	2.19963	2.19963	2.25624	2.25624	2.25624	2.46063	2.46063	2.46063
		$\sigma_{\bar{T}}$	2.26447	2.26447	2.26447	2.67709	2.67709	2.67709	5.85890	5.85890	5.85890
		μ_{IPS}	0.59982	0.59982	0.59982	0.62812	0.62812	0.62812	0.73031	0.73031	0.73031
		σ_{IPS}	0.74525	0.74525	0.74525	0.78993	0.78993	0.78993	0.95773	0.95773	0.95773
	HE-ST	$\mu_{\bar{T}'}$	1.28125	1.22610	1.11105	1.34817	1.28312	1.14487	1.61912	1.51912	1.29230
		$\sigma_{\bar{T}}$	2.95707	2.81630	2.46797	3.17185	3.02778	2.67505	5.26812	5.05473	4.58809
		μ_{IPS}	0.29629	0.29096	0.28229	0.31985	0.31154	0.29658	0.41473	0.39649	0.35665
		σ_{IPS}	0.5751	0.55250	0.51902	0.65213	0.61616	0.55960	0.99015	0.90095	0.73816

As mentioned earlier the probability that a service completion occurs before any interruptions given by $\boldsymbol{\alpha}[\gamma I - S]^{-1}\boldsymbol{S}^0$ is independent of the other PH-distributions and the values for this probability for the set of PH-distributions under study are given in Table 3. From this table we see that hyper-exponential times that have the largest variability yields the highest probability of service completion without any interruption.

Table 3: P(service completion with no interruption)

Service distribution	ER-ST	EX-ST	HE-ST
Probability	0.40188	0.50000	0.74376

Suppose we now look at what happens to the probability of a service completion with exactly one interruption with super clock expiring. That is, we look at the probability P_1^{SCEX}. Table 4 gives this probability for the different scenarios under consideration. It is interesting to note that this measure appears to be independent of the type of distribution used for the threshold clock.

Table 4: Values of P_1^{SCEX} for different scenarios

	ER-ST			EX-ST			HE-ST		
	ER-IP	EX-IP	HE-IP	ER-IP	EX-IP	HE-IP	ER-IP	EX-IP	HE-IP
ER-SC	0.08664	0.11121	0.06638	0.07242	0.09297	0.05549	0.03712	0.04764	0.02844
EX-SC	0.22674	0.19937	0.10798	0.18954	0.16667	0.09027	0.09714	0.08541	0.04626
HE-SC	0.44247	0.38930	0.23773	0.36988	0.32544	0.19873	0.18956	0.16678	0.10185

EXAMPLE 2: The purpose of this example is to see the effect of γ on $\mu_{\bar{T}'}$. Towards this, we fix $N = 20, \mu_1' = 4.0, \mu_2' = 1, \mu_3' = 2$, and $\mu_4' = 2.$. We fix all but one of the four PH-distributions to be either Erlang (ER) or hyperexponential HE, and vary the other one to be one of ER, EX, and HE. Note that these representations are normalized so as to have the specified mean. In Figures 1 and 2, respectively, we display $\mu_{\bar{T}'}$ as functions of γ for these two sets of scenarios.
From Figures 1 and 2, we note that a larger variability for the interruption times (when other times have significantly less variability) yields high values for the effective mean service time for almost all values of γ. However, having a higher or lower variability among the threshold times appears to have very small significant effect on this measure for a fixed value of γ. However, as γ is increased, the effective mean service increases and approaches a constant (which depends on the type of distributions used) and the rate of approaching the constant value is high for lower variability distributions (Figure 1) as compared to higher ones (Figure 2). In both the figures we notice that only in the case of varying interruption times from Erlang to hyper-exponential we see the effective mean service time appears to increase with increasing variability. In other cases this measure appears to decrease with increasing variability.

3.5. Stationary distribution

Since the model is studies as a LIQBD process, its stationary distribution (if it exists) has a matrix-geometric solution. Assume that condition (2) is satisfied. Let the stationary vector x of Q be partitioned by the levels into sub-vectors x_i for $i \geqslant 0$. Then x_i has the matrix-geometric form

$$x_i = x_1 R^{i-1} for \quad i \geqslant 2 \tag{3.2}$$

where R is the minimal non-negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0, \tag{3.3}$$

and the vectors x_0, x_1 are obtained by solving the equations

$$\begin{aligned} x_0 D_0 + x_1 C_1 = 0, \\ x_0 C_0 + x_1(A_1 + RA_2) = 0 \end{aligned} \tag{3.4}$$

subject to the normalizing condition

$$x_0 e + x_1 (I - R)^{-1} e = 1 \tag{3.5}$$

From the above equation, it is clear that to determine x, a key step is the computation of the rate matrix R.

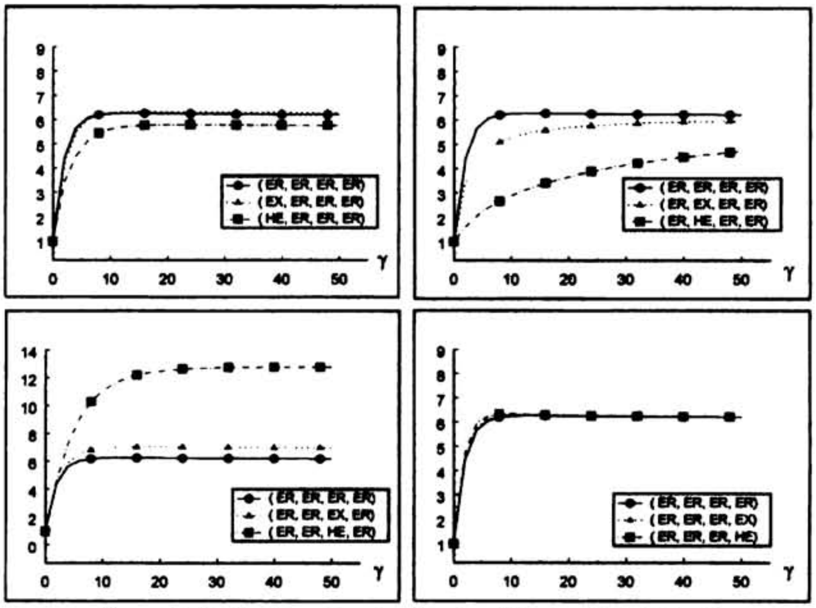

Figure 1: Effective Mean Service Time (fix all but one to be Erlang)

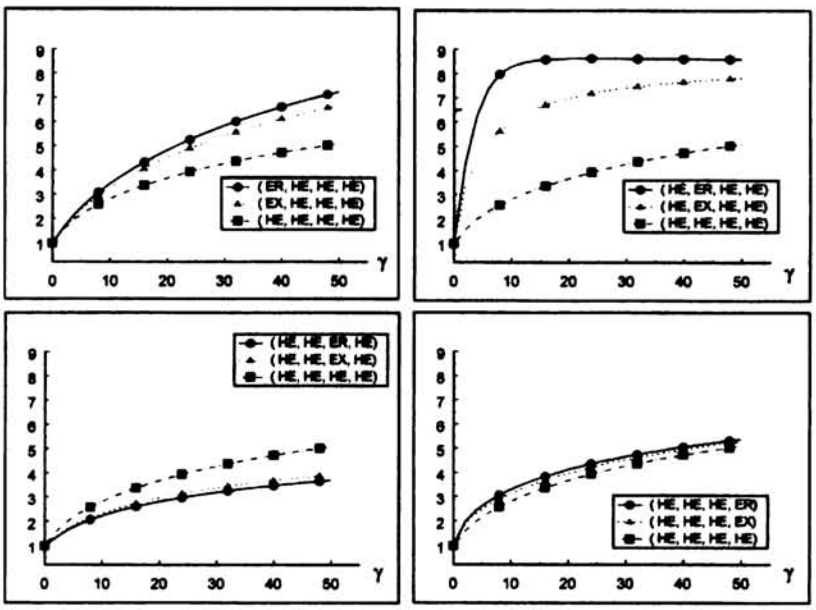

Figure 2: Effective Mean Service Time (fix all but one to be HE)

Theorem 3.5.1. *The process* $\{X(t), t \geqslant 0\}$ *is stable if and only if* $\pi D_1 e < \frac{1}{\mu'_{\tilde{T}}}$ *where* π *is the stationary vector of* $D = D_0 + D_1$ *and* $\mu'_{\tilde{T}}$ *is the mean effective service time.*

3.6. Performance Characteristics

Some useful general descriptors of our model are listed below.

1. Mean number of customers in the system= $\sum_{n=1}^{\infty} n x_n e = x_1(I-R)^{-2}e$

2. Fraction of time the server is busy= $\sum_{n=1}^{\infty} x_{n1} e$

3. Fraction of time the server remains interrupted= $\sum_{n=1}^{\infty} x_{n0} e$

4. Thus the fraction of time the server is idle= $x_0 e$

5. Fraction of time service is in interrupted state +Fraction of time service is going on= $\sum_{n=1}^{\infty} x_{n0} e + \sum_{n=1}^{\infty} x_{n1} e$

6. The rate at which server break down occurs= $\gamma \sum_{n=1}^{\infty} x_{n1} e$

7. Rate at which Interruption completion takes place before threshold is reached $R_I^c b$
$= \sum_{n=1}^{\infty} \sum_{i=1}^{N} \sum_{j=0}^{d} \sum_{k=1}^{a} \sum_{l=1}^{b} \sum_{j'=1}^{c} \sum_{u=1}^{r} x_{n,0,i,j,k,l,j',u} T_l^0$

8. Rate at which interruption completion takes place after the threshold is reached $R_I^c a$
$= \sum_{n=1}^{\infty} \sum_{i=1}^{N} \sum_{j=0}^{d} \sum_{k=1}^{a} \sum_{l=1}^{b} \sum_{u=1}^{r} x_{n,0,i,j,k,l,0,u} T_l^0$

9. Rate at which service completion(with atleast one interruption) takes place before super threshold is reached $R_S^c b$
$= \sum_{n=1}^{\infty} \sum_{i=0}^{N} \sum_{j=0}^{d} \sum_{k=1}^{a} \sum_{u=1}^{r} x_{n,1,i,j,k,u} S_k^0$

10. Rate at which service completion (with atleast one interruption) takes place after super threshold is reached $R_S^c a$
$= \sum_{n=1}^{\infty}\sum_{i=0}^{N}\sum_{k=1}^{a}\sum_{u=1}^{r} x_{n,1,i,0,k,u}\, S_k^0$

11. Effective service rate $E_S^R = \sum_{n=1}^{\infty}\sum_{k=1}^{a}\sum_{u=1}^{r} x_{n,1,0,k,u}\, S_k^0 + \sum_{n=1}^{\infty}\sum_{i=0}^{N}\sum_{j=0}^{d}\sum_{k=1}^{a}\sum_{u=1}^{r} x_{n,1,i,j,k,u}\, S_k^0$

3.7. Numerical Results

We illustrate the behavior of performance measures with variation in the incidence rate γ, of interruption. For this we fix the following values:

N=3, a=2,b=2,c=2,d=2, t=2;$D_0 = \begin{bmatrix} -6.5 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$,$D_1 = \begin{bmatrix} 6.0 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$, $L = \begin{bmatrix} -12.0 & 5.0 \\ 5.0 & -12.0 \end{bmatrix}$, $S = \begin{bmatrix} -12.0 & 6.0 \\ 6.0 & -12.0 \end{bmatrix}$, $T = \begin{bmatrix} -12.0 & 3.0 \\ 3.0 & -12.0 \end{bmatrix}$, $U = \begin{bmatrix} -12.0 & 8.0 \\ 8.0 & -12.0 \end{bmatrix}$
$L^0 = \begin{bmatrix} 7.0 & 7.0 \end{bmatrix}'$, $S^0 = \begin{bmatrix} 6.0 & 6.0 \end{bmatrix}'$, $T^0 = \begin{bmatrix} 9.0 & 9.0 \end{bmatrix}'$, $U^0 = \begin{bmatrix} 4.0 & 4.0 \end{bmatrix}'$,
$\eta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}$, $\alpha = \begin{bmatrix} 0.4 & 0.6 \end{bmatrix}$, $\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}$, $\delta = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$.

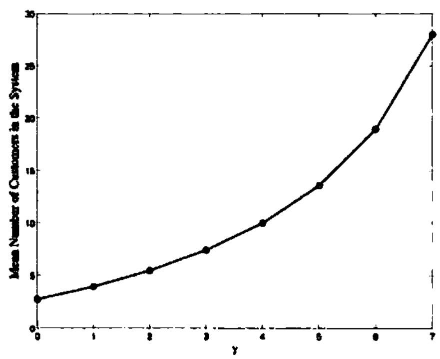

Fig. 3.1: Gamma versus Mean Number of Customers in the System

As expected, mean number of customers in the system increases with increasing value of γ.

The Figure 3.2 indicates that with increase in the value of γ, the fraction of time the server remains interrupted increases.

One expects that the fraction of time the server remains idle (no customer in the system) decreases with increase in value of γ. Experimentally this is validated

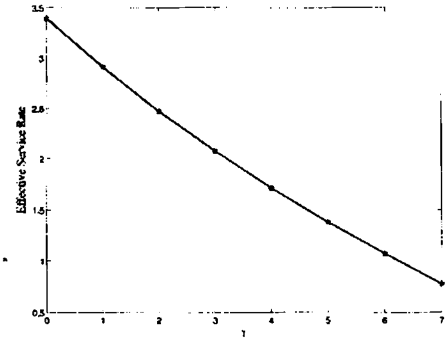

Fig. 3.4: Gamma versus Effective Service Rate

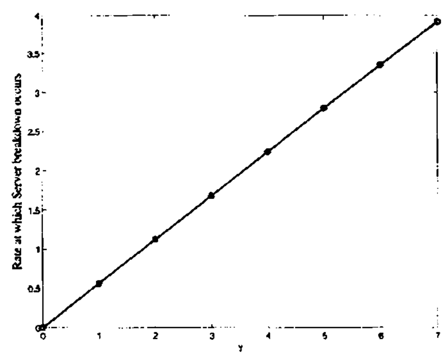

Fig. 3.5: Gamma versus and Rate at which server break down occurs

3.8. Model II

In Model I, the current interruption was allowed to continue even on saturation of the super-clock. However, no further interruption was permitted. In the present model, it is assumed that the moment the super-clock is realized, the interruption is removed. The super-clock saturation did not have any pronounced impact on the effective service time of a customer in Model I. However, the additional assumption brought in here shows that the impact of the super-clock is indeed remarkable. This will be revealed in the numerical illustration. Thus we study the case of a customer whose interruption completion takes place at the moment when the super threshold is saturated. A physical realization of this is the following: On getting interrupted the service of the present customer is suspended. During this interrupted state if the super-clock realized, then immediately another server (hired from within the system) is assigned to complete the service of the interrupted customer. This way the system is made more efficient in the sense that queue length will decrease and service will increase. We proceed to the mathematical model:

Let $N(t)$ denotes number of customers in the system at time t and state of the server be $S(t)$: $S(t) = 1$ the server is busy and $S(t) = 0$ indicates that it is interrupted. The variable $B(t)$ counts the number of interruptions the customer in service/interrupted has already undergone. The super threshold random variable is described by $S_1(t)$. Since current interruption is removed the moment the super threshold is saturated, the super threshold saturation point 0 is not in interruption state, but it keeps its position in busy state. The service phase, interruption phase and threshold random variables are expressed by the variables $S_2(t), S_3(t)$ and $S_4(t)$, respectively. Here as in model I, super threshold, service phase, interruption phase and threshold random variable are all PH distributed with representation $(\eta, V), (\alpha, S), (\beta, T)$ and (δ, U). The process $X_2(t) = \{(N(t), S(t), B(t), S_1(t), S_2(t), S_3(t), S_4(t), M(t)), t \geqslant 0\}$; is a continuous time Markov chain (CTMC) which turns out to be LIQBD with n^{th} level given by $\ell(n)=\bigcup_l \Psi(n,l)$, $l = 0, 1$. The subsets of $\Psi(n,l)$ are defined as $\{(n, 0, j_1, i_1, i_2, i_3, i_4, i_5); 0 \leqslant j_1 \leqslant N-1; 0 \leqslant i_1 \leqslant d; 1 \leqslant i_2 \leqslant a; 1 \leqslant i_3 \leqslant b; 0 \leqslant i_4 \leqslant c; 1 \leqslant i_5 \leqslant r\}$, $\{(n, 1, j_1, i_1, i_2, i_5); 0 \leqslant j_1 \leqslant N; 0 \leqslant i_1 \leqslant d; 1 \leqslant i_2 \leqslant a; 1 \leqslant i_5 \leqslant r\}$ and $\{(n, 1, 0, i_2, i_5); 0 \leqslant i_2 \leqslant a, 1 \leqslant i_5 \leqslant r\}$. Then $\{X_2(t), t \geqslant 0\}$ is a level independent quasi birth and death process.

$$Q = \begin{bmatrix} D_0 & C_0 & 0 & 0 & 0 & 0 & 0 \\ C_1 & A_1 & A_0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 \\ & & & & & & \end{bmatrix} \tag{3.6}$$

The matrix A_0, A_1 and A_2 correspond to transition from level $\Psi(n,.)$ to $\Psi(n+1,.)$; $\Psi(n,.)$ to $\Psi(n,.)$and $\Psi(n,.)$ to $\Psi(n-1,.)$, respectively. The matrix $I_{N(d+1)ab(c+1)} \otimes D_1$ records transition rates to $\Psi(n+1,0)$ from $\Psi(n,0)$. The matrix $I_{(N(d+1)+1)a} \otimes D_1$ records transitions to $\Psi(n+1,1)$ starting from $\Psi(n,1)$. All other components in A_0 are zero matrices. While service completion takes place only during the service period, A_2 is of the form $A_2 = \begin{bmatrix} [0] & [0] & \\ [0] & [A_2^{(1)} & [0]] \end{bmatrix}$, where $A_2^{(1)} = e_{(N(d+1)+1)} \otimes S^0 \otimes \alpha \otimes I_r$.

The components in A_1 are A_{11}, A_{12}, A_{13} and A_{14} which record transitions to $\Psi(n,.)$ from $\Psi(n,.)$. The matrix A_{11} records transition within $\Psi(n,0)$ and is given by: $A_{11} = I_N \otimes H_3$ where $H_3 = G_4 \oplus H_1$, and $H_1 = G_2 \oplus G_1$. Here the matrix $G_1 = F_1 \oplus D_0$, $G_2 = T \otimes I_{c+1}$, and $G_4 = V \otimes I_a \otimes I_b \otimes I_{(c+1)}$. The interruption states to busy state transition is described by $A_{12} = \begin{bmatrix} [0] & A_{12}^{(1)} \end{bmatrix}$ with $A_{12}^{(1)} = \begin{bmatrix} I_{N-1} \otimes V^{0)} \otimes E & I_{N-1} \otimes I_d \otimes B \end{bmatrix}$ where $B = [B_{11}, B_{12}, \ldots\ldots, B_{1a}]$ and $E = [E_{11}, E_{12}, \ldots\ldots, E_{1a}]$. In this $B_{1j} = T^0 \otimes [\bar{\alpha}, \bar{e}_j, \bar{e}_j, \ldots\ldots \bar{e}_j]' \otimes I_r$ and $E_{1j} = e_b \otimes [\bar{\alpha}, \bar{e}_j, \bar{e}_j, \ldots\ldots \bar{e}_j]' \otimes I_r$.

When a service is going on interruption the server can occur at rate γ; this expressed by the matrix $A_{13} = \begin{bmatrix} [0] & \begin{bmatrix} A_{13}^{(1)} & [0] \\ [0] & A_{13}^{(2)} \end{bmatrix} \end{bmatrix}$ where $A_{13}^{(1)} = \gamma I_a \otimes \bar{\eta} \otimes \beta \otimes \widehat{\delta} \otimes I_a$ and $A_{13}^{(2)} = I_{N-1} \otimes \gamma I_d \otimes I_a \otimes \beta \otimes \widehat{\delta} \otimes I_r$.

Service phase change during busy period is described by $A_{14} = \begin{bmatrix} A_{14}^{(1)} & [0] & [0] \\ [0] & A_{14}^{(2)} & [0] \\ [0] & [0] & A_{14}^{(3)} \end{bmatrix}$ where $A_{14}^{(1)} = T \oplus D_0 - \gamma I_{ar}$, $A_{14}^{(2)} = I_{N-1} \otimes \begin{bmatrix} T \oplus D_0 & [0] \\ [0] & I_d \otimes [S \oplus D_0 - \gamma I_{ar}] \end{bmatrix}$ and $A_{14}^{(3)} = I_{a+1} \otimes [S \oplus D_0]$. The matrices D_0, C_0 and C_1 are, respectively the transitions from level 0 to 0; level 0 to 1 and level 1 to 0 where $C_0 = \begin{bmatrix} [0] & \alpha \otimes D_1 & [0] \end{bmatrix}$ and $C_1 = \begin{bmatrix} [0] \\ e_{(N(d+1)+1)} \otimes (S^{(0)} \otimes I_r) \end{bmatrix}$.

3.9. Description of the phase type distribution for the services

The focus of this section is to describe the time it takes to process a job once it enters into the service facility. We assume that the service times are of phase type with representation given by $(\boldsymbol{\alpha}, S)$ of order a. The services are subject to interruptions and the interruption process is assumed to follow a Poisson process with rate γ. When the current service is interrupted for the first time, three clocks, referred to as super clock, interruption clock, and threshold clock, respectively, will simultaneously be started. The durations of these clocks are of phase type with representations given by, respectively, $(\boldsymbol{\eta}, L)$ of order d, $(\boldsymbol{\beta}, T)$ of order b, and $(\boldsymbol{\delta}, U)$ of order c. Once the interruption clock expires/super-clock realizes, whichever occurs first, the service of the interrupted job will begin again. The service will resume (from the phase where the service got interrupted) or repeat (like a new service) depending on whether the interruption clock expired before the threshold clock or not. In addition, if the super-clock expired before the interruption clock, then the service of the current job will not be interrupted anymore once the service begins again for this job. On the other hand, if the interruption clock expires before the super clock, the phase of the super-clock will be frozen and will resume from this phase should there be an another interruption for the existing job. For the job under service, the number of interruptions will be tracked and when this number attains a pre-specified threshold value, $N < \infty$, no further interruptions are allowed.

Let $X_2(t)$ denote the duration of the effective service for a job. That is, $X_2(t)$ is the time between the arrival of a job to the service facility until it leaves the facility with a service. Note that this service time may have i, $0 \leqslant i \leqslant N$ interruptions, and each interruption may cause the service to either resume from where it got interrupted or repeat from the beginning. Thus, we refer to $X_2(t)$ as the effective service time to distinguish this from the service time given by $(\boldsymbol{\alpha}, S)$. Note that when $\gamma = 0$ (that is, when there are no interruptions) then the effective service time is same as the service time. We will show that $X_2(t)$ follows a PH-distribution. Towards this end, we first define $J_1(t), J_2(t), J_3(t), J_4(t)$, and $J_5(t)$, respectively, to be the number of interruptions seen by the current job in service, the phase of the super clock, the phase of the current service, the phase of the interruption clock, and the phase of the threshold clock, at time t. Note that some of these phases will be frozen or not defined. For example, when the service is going on, the phases of interruption clock and the threshold clock will not be defined as they are not turned on, and the super clock may not be defined or when defined it will be frozen indicated that it is realized. Let $*$ denote the absorbing state that corresponds to the completion of the current service.

The states and their description are given in Table 1 below. For use in sequel, we now define sets of states as follows.

- $\boldsymbol{i^*} = \{(i^*, j_2) : 1 \leqslant j_2 \leqslant a\}$, for $1 \leqslant i^* \leqslant N-1$
- $\bar{N} = \{(\bar{N}, j_2) : 1 \leqslant j_2 \leqslant a\}$
- $\boldsymbol{0} = \{(0, j_2) : 1 \leqslant j_2 \leqslant a\}$
- $\boldsymbol{i} = \{(i, j_1, j_2) : 1 \leqslant j_1 \leqslant d, 1 \leqslant j_2 \leqslant a\}$, for $1 \leqslant i \leqslant N-1$
- $\tilde{i} = \{(\tilde{i}, j_1, j_2, j_3) : 1 \leqslant j_1 \leqslant d, 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b\}$, for $1 \leqslant \tilde{i} \leqslant N-1$
- $\tilde{\tilde{i}} = \{(\tilde{\tilde{i}}, j_1, j_2, j_3, j_4) : 1 \leqslant j_1 \leqslant d, 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b, 1 \leqslant j_4 \leqslant c\}$, for $1 \leqslant \tilde{\tilde{i}} \leqslant N-1$
- $\bar{\bar{N}} = \{(\bar{\bar{N}}, j_2, j_3, j_4) : 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b, 1 \leqslant j_4 \leqslant c, \}$
- $\bar{\bar{\bar{N}}} = \{(\bar{\bar{\bar{N}}}, j_2, J_3) : 1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b\}$

Table 1: The states and their description

States	Description
$\{(i^*, j_2) :$ $1 \leqslant j_2 \leqslant a,$ $1 \leqslant i^* \leqslant N-1\}$	The super clock expired and the current service is in phase j; no more interruptions from these states possible. The number of interruptions for the current service is i^*.
$\{(\bar{N}, j_2) : 1 \leqslant j_2 \leqslant a\}$	The number of interruptions is at its maximum with the current service in phase j_2.
$\{0, j) : 1 \leqslant j_2 \leqslant a\}$	The current service has seen no interruptions so far and the service is in phase j_2.
$\{(i, j_1, j_2) : 1 \leqslant j_1 \leqslant d,$ $1 \leqslant j_2 \leqslant a,$ $1 \leqslant i \leqslant N-1\}$	The phase of the service is in state j_2 with the super clock frozen in state j_1 and the number of interruptions so far is i.
$\{(\tilde{i}, j_1, j_2, j_3) : 1 \leqslant j_1 \leqslant d,$ $1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b$ $1 \leqslant i \leqslant N-1\}$	The threshold clock has expired with the super in phase j_1, the interruption clock is in state j_3; the service phase is frozen in j_2 and the number of interruptions including the current one is $\tilde{i}$..
$\{(\tilde{\tilde{i}}, j_1, j_2, j_3, j_4) : 1 \leqslant j_1 \leqslant d,$ $1 \leqslant j_2 \leqslant a, 1 \leqslant j_3 \leqslant b,$ $1 \leqslant j_4 \leqslant c, 1 \leqslant \tilde{\tilde{i}} \leqslant N-1\}$	The super clock is in state j_1, the interruption clock in state j_3,the threshold clock in j_4; the service phase is frozen in state j_2, and the number of interruptions including the current one is $\tilde{\tilde{i}}$.
$\{(\bar{\bar{N}}, j_2, j_3, j_4) : 1 \leqslant j_2 \leqslant a,$ $1 \leqslant j_3 \leqslant b, 1 \leqslant j_4 \leqslant c\}$	The service phase is frozen in state j_2 with the number of interruptions at its maximum limit of N; both interruption and threshold clocks are in phase j_3, and j_4, respectively.
$\{(\bar{\bar{\bar{N}}}, j_2, j_3) : 1 \leqslant j_2 \leqslant a,$ $1 \leqslant j_3 \leqslant b\}$	The service phase is frozen in state j_2 with the number of interruptions at its maximum limit of N, and the interruption clock is in phase j_3.

The Markov process $\{J_1(t), J_2(t), J_3(t), J_4(t), J_5(t)) : t \geqslant 0\}$ with the absorbing state $*$ is defined on the state space

$$\Omega = \{i^* : 1 \leqslant i^* \leqslant N\} \bigcup \{\hat{N}\} \bigcup \{0\} \bigcup \{i : 1 \leqslant i \leqslant N-1\} \bigcup \{\hat{i} : 1 \leqslant \hat{i} \leqslant N-1\}$$

$$\bigcup \{\tilde{\tilde{i}} :. 1 \leqslant \tilde{\tilde{i}} \leqslant N-1\} \bigcup \{\tilde{\hat{N}}\} \bigcup \{\tilde{\hat{\tilde{N}}}\} \bigcup \{*\}$$

and its infinitesimal generator is given by

$$Q = \begin{pmatrix} \tilde{T} & \tilde{\boldsymbol{T}}^0 \\ 0 & 0 \end{pmatrix}, \tag{1}$$

where

$$\tilde{T} = \begin{pmatrix} T_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & T_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & T_{3,3} & 0 & 0 & T_{3,6} & 0 & 0 \\ 0 & 0 & 0 & T_{4,4} & 0 & T_{4,6} & T_{4,7} & 0 \\ T_{5,1} & 0 & 0 & T_{5,4} & T_{5,5} & 0 & 0 & 0 \\ T_{6,1} & 0 & 0 & T_{6,4} & T_{6,5} & T_{6,6} & 0 & 0 \\ 0 & T_{7,2} & 0 & 0 & 0 & 0 & T_{7,7} & T_{7,8} \\ 0 & T_{8,2} & 0 & 0 & 0 & 0 & 0 & T_{8,8} \end{pmatrix}, \tag{2}$$

with

$$T_{1,1} = I \otimes S,\ T_{2,2} = S,\ T_{3,3} = S - \gamma I,\ T_{3,6} = \gamma(e_1' \otimes \eta \otimes I \otimes \beta \otimes \delta),$$

$$T_{4,4} = I \otimes (S - \gamma I),\ T_{4,6} = \gamma(\hat{I}_{N-1} \otimes I \otimes I \otimes \beta \otimes \delta),\ T_{4,7} = \gamma(e \otimes I \otimes \beta \otimes \delta),$$

$$T_{5,1} = I_N \otimes L^0 \otimes e \otimes \alpha,\ T_{5,4} = I \otimes I \otimes e\alpha \otimes T^0,$$
$$T_{5,5} = I \otimes [L \oplus (I \otimes T)],\ T_{6,1} = I_N \otimes e \otimes I_a \otimes L^0 \otimes e$$
$$T_{6,4} = I \otimes I \otimes I \otimes T^0 \otimes e,$$

$$T_{6,5} = I \otimes I \otimes I \otimes I \otimes U^0,\ T_{6,6} = I \otimes [L \oplus (I \otimes (T \oplus U))],\ T_{7,2} = I \otimes T^0 \otimes e,$$

$$T_{7,7} = I \otimes (T \oplus U),\ T_{7,8} = I \otimes I \otimes I \otimes U^0,\ T_{8,2} = e\alpha \otimes T^0,\ T_{8,8} = I \otimes T, \tag{3}$$

where

$$\tilde{T}^0 = \begin{pmatrix} e \otimes S^0 \\ S^0 \\ S^0 \\ e \otimes S^0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \hat{I}_{N-1} = \begin{pmatrix} 0 & I_{N-2} \\ 0 & 0 \end{pmatrix},$$

Theorem 3.9.1. *The effective service time, $X_2(t)$, is of phase type with representation $(\zeta, \tilde{T})$ of order $(N+1)a + Nab(c+1) + (N-1)da[1 + b(c+1)]$, where*

$$\zeta = (0, 0, \alpha, 0, 0, 0, 0, 0)\ , \tag{4}$$

and $\tilde{T}$ is as given in (2).

PROOF: First note that a new service will begin in level 0 in state $(0, j_2)$ with probability given by α_{j_2}. Once the service begins it can end with or without interruptions and looking through all possible transitions, one will see that the transition matrix is given by $\tilde{T}$ as given in (2). Thus, the service time is nothing but the time until absorption into state $*$ starting from the level 0. This results in the form of the initial probability vector as given in (4).
Now will show how the mean, $\mu_{\tilde{T}'}$, and the standard deviation, σ_T, of $(\zeta, \tilde{T})$ can be computed recursively (and explicitly). First recall ([25]) that the mean, μ_1', of (α, S), is given by

$$\mu_1' = \alpha\ (-S)^{-1} e. \tag{5}$$

Recall ([25]) that the mean and standard deviation of X is given by

$$\mu_{\tilde{T}'} = \zeta(-\tilde{T})^{-1} e, \qquad \sigma_{\tilde{T}} = \sqrt{[2\zeta(-\tilde{T})^{-2} e - (\mu_{\tilde{T}'})^2]}. \tag{6}$$

Due to the special structure of the matrix $\tilde{T}$ given in (2), we can compute the mean as well as the standard deviation of $\tilde{T}$ explicitly and recursively. First, we define

$$\zeta(-\tilde{T})^{-1} = (\boldsymbol{u}, \hat{\boldsymbol{u}}, \boldsymbol{v}_0, \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{a}_N, \boldsymbol{b}_N). \tag{7}$$

We further split the vectors on the right side of (7) as

$$\boldsymbol{u} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N),\ \boldsymbol{v} = (\boldsymbol{v}_1, \cdots, \boldsymbol{u}_{N-1}),\ \boldsymbol{y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{N-1}),\ \boldsymbol{z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_{N-1}). \tag{8}$$

Note that the vectors $\boldsymbol{u}_i, 1 \leqslant i \leqslant N$, $\hat{\boldsymbol{u}}$, and $\boldsymbol{v}_0$ are of dimension a; $\boldsymbol{v}_i, 1 \leqslant i \leqslant N-1$ are of order da; $\boldsymbol{y}_i, 1 \leqslant i \leqslant N-1$ are of order dab; $\boldsymbol{z}_i, 1 \leqslant i \leqslant N-1$ are of order $dabc$; $\boldsymbol{a}_N$ is of order abc and $\boldsymbol{b}_N$ is of order ab.

Exploiting the special structure of $\tilde{T}$ and using the notations in (7) and (8), it is easy to verify the following equations.

$$\begin{aligned}
\boldsymbol{v}_0 &= \boldsymbol{\alpha}[\gamma I - S]^{-1}, \\
\boldsymbol{z}_1 &= \gamma \boldsymbol{v}_0[\boldsymbol{\eta} \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta}][-(L \oplus (I \otimes (T \oplus U)))]^{-1}, \\
\boldsymbol{y}_1 &= \boldsymbol{z}_1[I \otimes I \otimes I \otimes \boldsymbol{U}^0][-(L \oplus (I \otimes T))]^{-1}, \\
\boldsymbol{u}_1 &= [\boldsymbol{y}_1(L^0 \otimes e \otimes e\alpha) + \boldsymbol{z}_1(e \otimes I_a \otimes L^0 \otimes e)][-S]^{-1}, \\
\boldsymbol{v}_i &= [\boldsymbol{y}_1(I \otimes \boldsymbol{e\alpha} \otimes \boldsymbol{T}^0) + \boldsymbol{z}_i(I \otimes I \otimes \boldsymbol{T}^0 \otimes \boldsymbol{e})][I \otimes (\gamma I - S)]^{-1},\ 1 \leqslant i \leqslant N-1, \\
\boldsymbol{z}_i &= \gamma \boldsymbol{v}_{i-1}[I \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta}][-(L \oplus (I \otimes (T \oplus U)))]^{-1},\ 2 \leqslant i \leqslant N-1, \\
\boldsymbol{y}_i &= \boldsymbol{z}_i[I \otimes I \otimes I \otimes \boldsymbol{U}^0][-(L \oplus (I \otimes T))]^{-1},\ 2 \leqslant i \leqslant N-1, \\
\boldsymbol{u}_i &= [\boldsymbol{y}_i(I_N \otimes L^0 \otimes e \otimes e\alpha) + \boldsymbol{z}_i(I \otimes e \otimes I_a \otimes L^0 \otimes e)][-S]^{-1},\ 2 \leqslant i \leqslant N-1, \\
\boldsymbol{a}_N &= \gamma[\boldsymbol{v}_{N-1}(\boldsymbol{e} \otimes I \otimes \boldsymbol{\beta} \otimes \boldsymbol{\delta})][-(I \otimes (T \oplus U))]^{-1}, \\
\boldsymbol{b}_N &= [\boldsymbol{a}_N(I \otimes I \otimes \boldsymbol{U}^0)[-(I \otimes T]^{-1}, \\
\boldsymbol{u}_N &= \boldsymbol{a}_N(I \otimes \boldsymbol{T}^0 \otimes \boldsymbol{e}) + \boldsymbol{b}_N(\boldsymbol{e\alpha} \otimes \boldsymbol{T}^0).
\end{aligned} \tag{9}$$

By looking at the order in which the equations are displayed in (9), one can see the explicit evaluation of the vectors needed in the computation of $\mu_{\tilde{T}'}$, which is obtained as

$$\mu_{\tilde{T}'} = \boldsymbol{ue} + \hat{\boldsymbol{u}}\boldsymbol{e} + \boldsymbol{v}_0\boldsymbol{e} + \boldsymbol{ve} + \boldsymbol{ye} + \boldsymbol{ze} + \boldsymbol{a}_N\boldsymbol{e} + \boldsymbol{b}_N\boldsymbol{e}. \tag{10}$$

Similar to getting an explicit expression for $\zeta(-\tilde{T})^{-1}$ one can derive an explicit expression for $\zeta(-\tilde{T})^{-2}$ by replacing the role played by ζ in (9) with $(\boldsymbol{u}, \hat{\boldsymbol{u}}, \boldsymbol{v}_0, \boldsymbol{v}, \boldsymbol{y}, \boldsymbol{z}, a_N, b_N)$

The details are omitted.

Note that some of the explicit expressions in (9) involve matrices whose orders may become very large as the values of $d, a, b,$ and c become large. In that case one can exploit the structure of these matrices. Some examples to this extent can be seen in [25]. In any computational aspect, it is very important to have some internal accuracy checks. In the case of the computation of $\mu_{\tilde{T}'}$, one can use the fact that $\boldsymbol{\zeta}(-\tilde{T})^{-1}\tilde{\boldsymbol{T}}^{\mathbf{0}} = 1$, which reduces to

$$\left[\hat{\boldsymbol{u}} + \boldsymbol{v}_0 + \sum_{i=1}^{N} \boldsymbol{u}_i\right] \boldsymbol{S}^{\mathbf{0}} + \sum_{i=1}^{N-1} \boldsymbol{v}_i(\boldsymbol{e} \otimes \boldsymbol{S}^{\mathbf{0}}) = 1. \tag{11}$$

In the case of the computation of $\sigma_{\tilde{T}}$, one can use the fact that $\boldsymbol{\zeta}(-\tilde{T})^{-2}\tilde{\boldsymbol{T}}^{\mathbf{0}} = \mu_{\tilde{T}'}$.

3.10. Numerical Examples

In this section we discuss some interesting numerical examples that qualitatively describe the phase type distribution modeling the interrupted services. The correctness and the accuracy of the implementation of the recursive and explicit schemes are verified by a number of accuracy checks such as the one listed in(11). As an additional accuracy check, we obtained the numerical solution for the exponential clocks' case in their simple forms. Next, we implemented the general algorithm, but using the following $PH-$ representation: Let R be an irreducible, stable matrix with eigenvalue of maximum real part -$\theta < 0$. Let $\boldsymbol{\alpha}$ denote the corresponding left eigenvector, normalized by $\boldsymbol{\alpha e}$=1. The $PH-$ representation $(\boldsymbol{\alpha}, R)$ reduces to the exponential distribution with rate θ. The general algorithm does not utilize this fact in any manner, but the numerical results agreed very much.

In addition to the measures, $\mu_{\tilde{T}'}$ and $\sigma_{\tilde{T}}$, given in (6) there are other measures that can be constructed. A few are listed below along with their formulas.

1. Probability of i interruptions without the super clock expiring: The probability, P_i^{SCNX} that exactly i interruptions occur during a service without the super clock expiring is given by

$$P_i^{SCNX} = \begin{cases} \boldsymbol{v}_0 \boldsymbol{S}^{\mathbf{0}}, & i = 0, \\ \boldsymbol{v}_i(\boldsymbol{e} \otimes \boldsymbol{S}^{\mathbf{0}}), & 1 \leqslant i \leqslant N-1, \\ \hat{\boldsymbol{u}} \boldsymbol{S}_{\mathbf{2}}^{\mathbf{0}}, & i = N. \end{cases}$$

2. Probability of i interruptions with the super clock expiring: The probability, P_i^{SCEX} that exactly i interruptions occur during a service with the super clock expiring is given by

$$P_i^{SCEX} = \boldsymbol{u}_i \boldsymbol{S}^{\mathbf{0}}, \quad 1 \leqslant i \leqslant N-1.$$

3. The mean number of interruptions during a service completion: The mean, μ_{IPS}, number of interruptions that occur during a service completion is given by

$$\mu_{IPS} = \sum_{i=1}^{N-1} i[P_i^{SCNX} + P_i^{SCEX}] + NP_N^{SCNX}.$$

4. Probability of a service completion in which neither super clock expires nor the maximum interruptions occur: The probability that a service completion occurs without the super clock expiring as well as the maximum number of interruptions allowed (which is N) is not attained is given by

$$P^{SCNXNI} = 1 - [\sum_{i=1}^{N} \boldsymbol{u}_i \boldsymbol{S}^{\mathbf{0}} + \hat{\boldsymbol{u}} \boldsymbol{S}^{\mathbf{0}}].$$

5. Probability of a service completion with no interruption: The probability that a service is completed without any interruptions is given by

$$P^{SCNI} = \boldsymbol{\alpha}[\gamma I - S]^{-1} \boldsymbol{S}^{\mathbf{0}}.$$

Note that the above probability is nothing but $P(X < Y)$ where X follows a phase type distribution with representation $(\boldsymbol{\alpha}, S)$ and Y is exponential with parameter γ.

3.11. Stationary distribution

Since the model is studies as a LIQBD process, its stationary distribution (if it exists) has a matrix-geometric solution. Let the stationary vector x of Q be partitioned by the levels into sub-vectors x_i for $i \geqslant 0$. Then x_i has the matrix-geometric form

$$x_i = x_1 R^{i-1} for \ \ i \geqslant 2 \tag{3.7}$$

where R is the minimal non-negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0, \tag{3.8}$$

and the vectors x_0, x_1 are obtained by solving the equations

$$\begin{aligned} x_0 D_0 + x_1 C_1 &= 0, \\ x_0 C_0 + x_1(A_1 + RA_2) &= 0 \end{aligned} \tag{3.9}$$

subject to the normalizing condition

$$x_0 e + x_1 (I - R)^{-1} e = 1 \tag{3.10}$$

From the above equation, it is clear that to determine x, a key step is the computation of the rate matrix R.

Theorem 3.11.1. *The process $\{X(t), t \geqslant 0\}$ is stable if and only if $\pi D_1 e < \frac{1}{\mu'_{\hat{T}}}$ where π is the stationary vector of $D = D_0 + D_1$ and $\mu'_{\hat{T}}$ is the mean effective service time.*

3.12. Performance Characteristics

Some useful general descriptors of our model are listed below.

1. Mean number of customers in the system$=\sum_{n=1}^{\infty} n x_n e = x_1 (I - R)^{-2} e$
2. Fraction of time the server is busy$=\sum_{n=1}^{\infty} x_{n1} e$
3. Fraction of time the server remains interrupted$=\sum_{n=1}^{\infty} x_{n0} e$
4. Thus the fraction of time the server is idle$=x_0 e$
5. Fraction of time service is in interrupted state
+Fraction of time service is going on$=\sum_{n=1}^{\infty} x_{n0} e + \sum_{n=1}^{\infty} x_{n1} e$
6. The rate at which server break down occurs$=\gamma \sum_{n=1}^{\infty} x_{n1} e$
7. Rate at which Interruption completion takes place before threshold is reached $R_I^c b$
$= \sum_{n=1}^{\infty} \sum_{i=1}^{N} \sum_{j=0}^{d} \sum_{k=1}^{a} \sum_{l=1}^{b} \sum_{j'=1}^{c} \sum_{u=1}^{r} x_{n,0,i,j,k,l,j',u} T_l^0$

8. Rate at which interruption completion takes place after the threshold is reached $R_I^c\, a$
$= \sum_{n=1}^{\infty}\sum_{i=1}^{N}\sum_{j=0}^{d}\sum_{k=1}^{a}\sum_{l=1}^{b}\sum_{u=1}^{r} x_{n,0,i,j,k,l,0,u}\, T_l^0$

9. Rate at which service completion(with atleast one interruption) takes place before super threshold is reached $R_S^c\, b$
$= \sum_{n=1}^{\infty}\sum_{i=0}^{N}\sum_{j=0}^{d}\sum_{k=1}^{a}\sum_{u=1}^{r} x_{n,1,i,j,k,u}\, S_k^0$

10. Rate at which service completion (with atleast one interruption) takes place after super threshold is reached $R_S^c\, a$
$= \sum_{n=1}^{\infty}\sum_{i=0}^{N}\sum_{k=1}^{a}\sum_{u=1}^{r} x_{n,1,i,0,k,u}\, S_k^0$

11. Effective service rate $E_S^R = \sum_{n=1}^{\infty}\sum_{k=1}^{a}\sum_{u=1}^{r} x_{n,1,0,k,u}\, S_k^0 + \sum_{n=1}^{\infty}\sum_{i=0}^{N}\sum_{j=0}^{d}\sum_{k=1}^{a}\sum_{u=1}^{r} x_{n,1,i,j,k,u}\, S_k^0$

3.13. Numerical Results

The numerical experiments in this sections show the marked deviation between models I and II discussed in this chapter. The effect of terminating interruption on realization of the super clock improves the system performance remarkably. For example fix the input parameters: N=3, a=2,b=2,c=2,d=2, r=2;$D_0 = \begin{bmatrix} -6.5 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$,$D_1 = \begin{bmatrix} 6.0 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$, $L = \begin{bmatrix} -12.0 & 5.0 \\ 5.0 & -12.0 \end{bmatrix}$, $S = \begin{bmatrix} -12.0 & 6.0 \\ 6.0 & -12.0 \end{bmatrix}$,
$T = \begin{bmatrix} -12.0 & 3.0 \\ 3.0 & -12.0 \end{bmatrix}$, $U = \begin{bmatrix} -12.0 & 8.0 \\ 8.0 & -12.0 \end{bmatrix}$,
$L^0 = \begin{bmatrix} 7.0 & 7.0 \end{bmatrix}'$, $S^0 = \begin{bmatrix} 6.0 & 6.0 \end{bmatrix}'$, $T^0 = \begin{bmatrix} 9.0 & 9.0 \end{bmatrix}'$, $U^0 = \begin{bmatrix} 4.0 & 4.0 \end{bmatrix}'$,
$\eta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}$, $\alpha = \begin{bmatrix} 0.4 & 0.6 \end{bmatrix}$, $\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}$, $\delta = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$.

Fig. 3.6 shows that increase in mean number of customers in the system is very much reduced, compared to the same situation discussed in Fig.3.1.

Fraction of time server is interrupted increases again; nevertheless, it is not that rapid unlike its counter part in Fig. 3.7

Idle time decrease with increase in γ (see Fig. 3.8); however, this not that much performed unlike that given in Fig. 3.3.

Here again effective service rate should decrease with increase in value of γ; nevertheless it is not that sharp compared to model I (see Fig. 3.4 and 3.9).

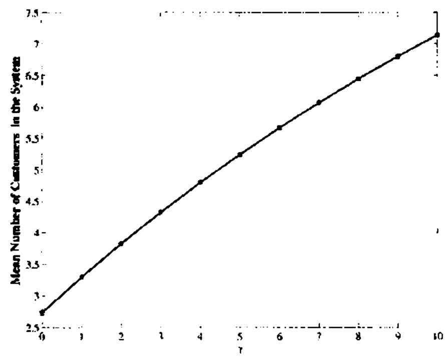

Fig. 3.6: Gamma versus Mean Number of Custŏmers in the System.

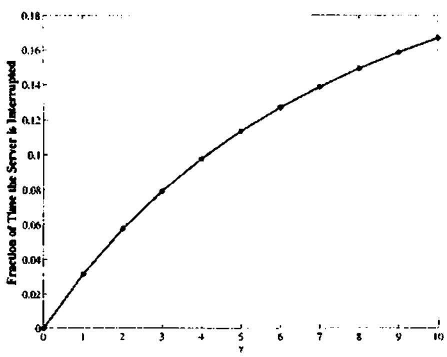

Fig. 3.7: Gamma versus Fraction of Time the Server is Interrupted.

The behavior of rate of breakdown with increase in value of γ is insensitive to the way the server interruption is removed. The two figures 3.5 and 3.10 are therefore identical.

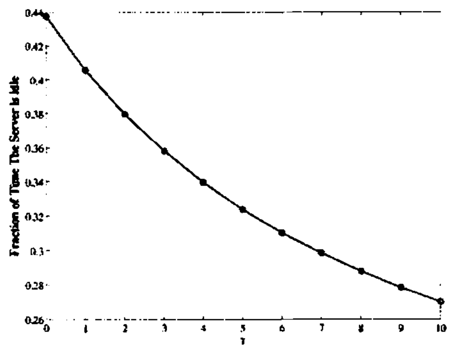

Fig. 3.8: Gamma versus Fraction of Time the Server is Idle

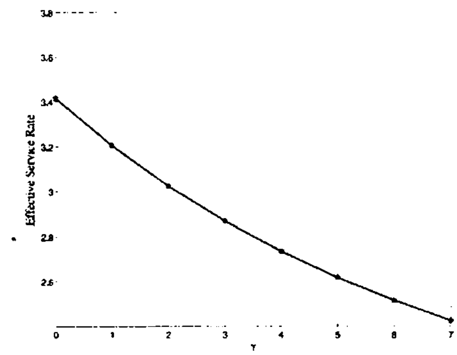

Fig. 3.9: Gamma versus Effective Service Rate.

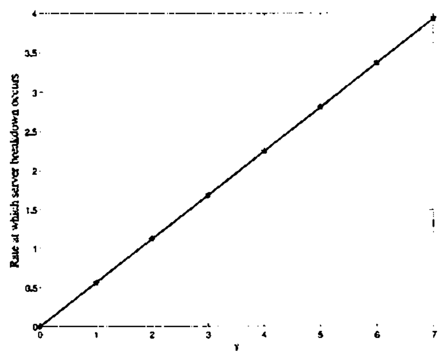

Fig. 3.10: Gamma versus Effective Service Rate and Rate at which server break down occurs

Chapter 4

Queue with Preemptions and Repeat or Resumption of Preempted Service

One of the objective of this chapter is to generalize results in Chapters 2 and 3 were concerned with customers of the same priority. Though we would very well interpret an interruption as the server processing a high priority customer, the models were basically confined to single priority. Otherwise questions like arrival process of priority customers, description of waiting space of such customers would arise. The present chapter is concerned with a two-priority service system. In contrast to fixing N as the upper bound for number of interruptions of a low priority customers service, the customer is given the option to choose the maximum number of interruptions he is willing to undergo, subject to a maximum of N. Nevertheless the customer who chooses to undergo interruptions closer to N (say $> N/2$) will be given incentives, which will not be available to those who do not opt for such length interruptions. Specifically we assume that q_i, $0 \leqslant i \leqslant N$, is the probability of a low priority customer opting for maximum of i interruptions. Thus the results here generalizes those of chapter 3. Specifically we concentrate on a single server queueing model consisting of two queues-an infinite capacity queue of low priority customers and a finite capacity N of high priority customers. Customers join the system according to a MMAP. If the server is free, at the epoch of an arrival of a customer (low priority/ high priority) can immediately join for service. An $(N+1)$ faces solid figure with the face marked 0 to N, is tossed at the beginning of the service of an ordinary customer. i^{th} face turns up with probability q_i $(0 \leqslant i \leqslant N)$. This decides the maximum number of priority customer(s) allowed to be served during the service of the specified ordinary customer. During the service of a low priority customer preemption can take place by the arrival of a high priority customer. Then the preempted customer waits at the head of the low priority queue till either the high priority queue becomes empty or the maximum number of high priority customers permitted to be served, as per the outcome of the toss of solid object, whichever occurs first. The restart/ resumption of preempted customer's service takes place when the high priority queue becomes empty or the maximum number of high priority customer's service permitted during his effective service is realized. We introduce a threshold random variable which competes with the duration of each preemption; if this realizes before completion of preemption then the preempted customer has to get its service repeated ; otherwise the service is resume . Here the random variable corresponding to low priority customers service, high priority customers service and threshold random variable are all distinct and independent PH distributed. The system is analyzed under stable regime. A few useful measures for system performance are obtained. These help in designing an efficient system. Numerical results are provided to illustrate the system performance . We also examine the optimal value of N.

4.1. Mathematical Model

We consider a queueing model in which arrival of low priority and high priority customers occur according to MMAP with representation (D_0, D_1) of order r. The arriving customer is of low (high) priority with probability $p_1(p_2)$. If the server is idle, an arriving customer (low priority or high priority) can immediately join for service. During low priority customers service the arrival of high priority customer preempts him, provided the number of high priority customers served during his preemptions has not reached the maximum allowed by the outcome of the solid figure by his own initial choice, and the preempted low priority customer waits as the head of the infinity capacity queue of low priority customers. Subsequent high priority customers arriving during that period wait in the finite capacity (K) queue. An $N+1$ faced solid figure with markings $0, 1, \ldots\ldots, N$, respectively is tossed at the beginning of a low priority customers service; let q_i be the probability that the tossing results in i, $(0 \leqslant i \leqslant N)$, then i is the maximum number of high priority customers allowed to be served during his service period. It may happen there is no priority customer present to be served during the effective service time of a low priority customer, even when the experimental outcome is $i(\geqslant 1)$. The moment preemption takes place the threshold random clock starts ticking. The preempted customer gets its service repeated /resumed when the high priority queue becomes either empty or the number of high priority customers served during his service period reaches its maximum, whichever occurs first. When the preemption time exceeds a threshold random variable, the interrupted customer gets its service repeated on completion of preemption; else the service is resumed, that is it starts at the point where it got preempted. Duration of services of low and high priority customers are PH distributed random variables with representations (α, S) and (β, T), respectively; the threshold r.v is PH distributed with representation (δ, U). All these random variables are mutually independent.Write $S^0 = -S\underline{e}$, $T^0 = -T\underline{e}$ and $U^0 = -U\underline{e}$ where $\underline{e}$ is a column vector of 1's of appropriate order. Let $N_1(t)$, $N_2(t)$, $S(t)$, $B_1(t)$ and $B_2(t)$ denote, respectively, the number of low priority customers , high priority customers, status of server, maximum number of high priority customers permitted to be served during a low priority customers service and the number of high priority customers so far served , including the present one , if a high priority service is going on by preemption. When $S(t) = 0$, the server is busy with high priority service and a preempted low priority customer is waiting as the head of the queue; when $S(t) = 1$, the server is busy with high priority customer with no preempted customer waiting and $S(t) = 2$ stand for the server busy with low priority service. The process $X(t) = \{(N_1(t), N_2(t), S(t), B_1(t), B_2(t), S_1(t), S_2(t), S_3(t), M(t)), t \geqslant 0\}$; is a continuous time Markov chain (CTMC) which turns out to be LIQBD with n^{th} level given by $\ell(n){=}\bigcup_l \Psi(n, m, l), 0 \leqslant m \leqslant K, l = 0, 1, 2$. The subsets of $\Psi(n, m, l)$ are defined as $\{(n, m, 0, j_1, j_2, i_1, i_2, i_3, i_4); 1 \leqslant j_1 \leqslant N; 1 \leqslant j_2 \leqslant j_1; 1 \leqslant i_1 \leqslant a; 1 \leqslant i_2 \leqslant b; 0 \leqslant i_3 \leqslant c; 1 \leqslant i_4 \leqslant r\}$, for $1 \leqslant m \leqslant K$, $\{(n, m, 1, i_2, i_4); 1 \leqslant i_2 \leqslant b; 1 \leqslant i_4 \leqslant r\}$ and for $0 \leqslant m \leqslant K$, $\{(n, m, 2, j_1, j_2, i_1, i_4); 1 \leqslant j_1 \leqslant N; 1 \leqslant j_2 \leqslant j_1; 1 \leqslant i_1 \leqslant a; 1 \leqslant i_4 \leqslant r\}$. The states in Ψ are listed in lexicographical order.The transitions among subsets $\Psi(n, m, l); l = 0, 1, 2$ are as follows:

- For $1 \leqslant m \leqslant K$, $I_K \otimes I_{N(N+1)/2} \otimes I_a \otimes I_b \otimes I_{(c+1)} \otimes p_1 D_1$, $I_K \otimes I_b \otimes p_1 D_1$ and $I_{(N(N+1)/2)+KN} \otimes I_a \otimes p_1 D_1$ records transition rates to states in $\Psi(n+1, m, 0)$, $\Psi(n+1, m, 1)$ and $\Psi(n+1, m, 2)$ respectively, starting from states in $\Psi(n, m, 0)$, $\Psi(n, m, 1) and \Psi(n, m, 2)$.

- The matrix $D = (D_{1i})_{1\times(N+1)}, i = 1, 2, \ldots\ldots, N+1$ is a column vector with components $D_{11}, D_{12}, \ldots\ldots\ldots, D_{1,N+1}$; $D_{1i} = T^0 \otimes q_{i-1} e_i \otimes \alpha \otimes I_r$, $1 \leqslant i \leqslant N+1$; records transition rates at the beginning of low priority service on completion of a high priority priority customers service where e_i is a column vector of order i with 1 in the 1^{st} place and zero elsewhere.

- The matrix $\tilde{Q} = (\tilde{Q}_{1i})_{1\times(N+1)}, i = 1, 2, \ldots\ldots, N+1$ is a column vector with components $\tilde{Q}_{11}, \tilde{Q}_{12} \ldots\ldots\ldots, \tilde{Q}_{1,N+1}$; $\tilde{Q}_{1i} = S^0 \otimes q_{i-1} e_i \otimes \alpha \otimes I_t$, $1 \leqslant i \leqslant N+1$; records transition rates at the beginning of low priority service on completion of a low priority customers service in state 2 where e_i is a column vector of order i with 1 in the 1^{st} place and zero elsewhere.

- The matrix $B = (B_{1i})_{1\times a}, i = 1, 2, \ldots\ldots, a$ is a column vector with components $B_{11}, B_{12} \ldots\ldots\ldots, B_{1,a}$; $B_{1i} = T^0 \otimes \begin{bmatrix} \alpha & e_j & e_j & . & e_j \end{bmatrix}^T \otimes I_r$, $1 \leqslant i \leqslant a$;

records transition rates corresponding to repeat/resumption of preempted low priority customer's service on completion of a high priority customer's service, where e_i is a column vector of order 1 in the i^{th} place and zero elsewhere.

- For $1 \leqslant m \leqslant K$, the matrices $T^0 \otimes \beta \otimes I_r$, $I_b \otimes p_2 D_1$ records transition rates to states in $\Psi(n, m-1, .)$ and $\Psi(n, m+1, .)$ respectively, from states in $\Psi(n, m, .)$;

The infinitesimal generator of the Markov chain governing the system is given by

$$Q = \begin{bmatrix} C_0 & C_1 & 0 & 0 & 0 & 0 & 0 \\ C_2 & A_1 & A_0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 \\ & & & & . & & \end{bmatrix} \tag{4.1}$$

- The matrix $C_0 = \left[C_0^{(i,j)}\right]$ represents a square matrix of order $r(Kb+1)$ which corresponds to transition from i to j, when the system is free with low priority customer; $0 \leqslant i, j \leqslant K$. The matrices D_0, $\beta \otimes p_2 D_1$, $T^0 \otimes I_r$ specifies elements of $C_0^{(0,0)}$, $C_0^{(0,1)}$ and $C_0^{(1,0)}$,respectively. $I_b \otimes p_2 D_1$ provides the elements of $C_0^{(i,i+1)}$, $1 \leqslant i \leqslant K-1$ and $T^0 \otimes \beta \otimes I_r$ lists the elements of $C_0^{(i,i-1)}$, $2 \leqslant i \leqslant K$, while $S_2 \bigoplus D_0$ records the transitions in $C_0^{(i,i)}$, $1 \leqslant i \leqslant K-1$, $T \bigoplus D_0 \bigoplus p_2 D_1$ corresponds to the transition rates in $C_0^{(i,i)}$, $i = K$.

- The only non zero block in C_1 are the transition from $\Psi(n, m, .)$, n=m=0 to $\Psi(1, m, 2)$; $\Psi(0, m, 1)$ to $\Psi(1, m, 1)$ and are denoted by $C_1^{(1)}$, $C_1^{(2)}$ respectively. Here $C_1^{(1)} = \left[\begin{array}{cc} C_1^{(11)} & [0] \end{array}\right]$ where $C_1^{(11)} = \left[\begin{array}{cccc} E_1 & E_2 & . & . & E_{N+1} \end{array}\right]$ with $E_i = q_{i-1} e_i \otimes \alpha \otimes p_1 D_1$. e_i is a $1 \times i$ row vector having 1 in the 1^{st} place and zero elsewhere. The matrix $C_1^{(2)} = \left[\begin{array}{cc} I_K \otimes I_b \otimes p_1 D_1 & [0] \end{array}\right]$ records arrival of a low priority customer when server's state is 2.

- The matrix $C_2 = \left[\begin{array}{cc} [0] & C_2^{(1)} \end{array}\right]'$ where block [0] indicates no service completion of low priority customer during state 0 and 1 of the server. The matrix $e_{(N+1)(N+2)/2} \otimes S^0 \otimes I_r$ in $C_2^{(1)}$ records transition from $\Psi(n, 0, 2)$ to $\Psi(n, 0, 2)$, if $n = 1$ and $I_K \otimes e_N \otimes (S^0 \otimes \beta \otimes I_r)$ lists the transition rates to $\Psi(0, m, 1)$ from $\Psi(1, m, 2)$, $1 \leqslant m \leqslant K$.

- The matrix A_0 records arrival of low priority customers in to the system where the only non zero elements are diagonal ones. The matrices $I_K \otimes I_{N(N+1)/2} \otimes I_a \otimes I_b \otimes I_{(c+1)} \otimes p_1 D_1$, $I_K \otimes I_b \otimes p_1 D_1$ and $I_{((N+1)(N+2)/2)+KN} \otimes I_a \otimes p_1 D_1$ in A_1 lists the arrival of low priority customer within states $\Psi(n,m,0)$, $\Psi(n,m,1)$ and $\Psi(n,m,2)$.

- The matrix A_2 in Q lists the service completion of low priority service. The matrices $A_2^{(1)} = \begin{bmatrix} [0] & I_K \otimes e_N \otimes S^0 \otimes \beta \otimes I_r \end{bmatrix}'$ and $A_2^{(2)} = \begin{bmatrix} e_{(N+1)(N+2)/2} \otimes Z & [0] \end{bmatrix}$ where $Z = (Z_{1i})'_{1 \times N+1}, i = 1, 2, \ldots\ldots, N+1$ is a column vector of $Z_{11}, Z_{12}, \ldots\ldots\ldots$ $, Z_{1N+1}$; $Z_{1i} = S^0 \otimes q_{i-1} e_i \otimes \alpha \otimes I_r$, $1 \leqslant i \leqslant N+1$ records transition rates in $\Psi(n-1,0,2), \Psi(n-1,m,2)$ of A_2, starting from states in $\Psi(n,0,2)$ and $\Psi(n,m,2)$ respectively.

- The matrix A_1 in Q records transition from $\Psi(n,m,l)$ to itself. The components in A_1 are A_{11}, A_{12}, A_{13}, A_{14}, A_{15} and A_{16} each of which records transition rates from states in $\Psi(n,m,0)$ to $\Psi(n,m,0)$; $\Psi(n,m,0)$ to $\Psi(n,m,1)$; $\Psi(n,m,1)$ to $\Psi(n,m,1)$; $\Psi(n,m,1)$ to $\Psi(n,m,2)$; $\Psi(n,m,2)$ to $\Psi(n,m,0)$ and $\Psi(n,m,2)$ to $\Psi(n,m,2)$.

(a) The matrix A_{11} is as follows: $\mathrm{I}_{\mathrm{N(N+1)/2}} \otimes I_a \otimes H$, where $H = G_1 \oplus G_2$, $G_1 = F \oplus D_0$, $F = \begin{bmatrix} \bar{0} & 0 \\ U & U^0 \end{bmatrix}$, $G_2 = T \otimes I_{c+1}$ records transitions to $\Psi(n,m,0)$ from $\Psi(n,m,0)$, $1 \leqslant m \leqslant K-1$, $I_{N(N+1)/2} \otimes I_a \otimes I_b \otimes I_{(c+1)} \otimes p_2 D_1$ records transition rates from $\Psi(n,m,0)$ to $\Psi(n,m+1,0)$, $1 \leqslant m \leqslant K-1$ and $\mathrm{I}_{\mathrm{N(N+1)/2}} \otimes I_a \otimes (H \oplus p_2 D_1)$ If $m = K$. $I_{N(N+1)/2} \otimes I_a \otimes T^{(0)} \otimes \beta \otimes I_{c+1} \otimes I_r$ records transition rates from $\Psi(n,m,0)$ to $\Psi(n,m-1,0)$, $2 \leqslant m \leqslant K$.

(b) The matrix A_{12} records transitions in $\Psi(n,m,2)$ from $\Psi(n,m,0)$ and is as follows: The matrix $\begin{bmatrix} [0] & diag\left[\begin{pmatrix} [0] & diag(I_j B) \end{pmatrix}\right] \end{bmatrix}, 1 \leqslant j \leqslant N$, $diag\left[\begin{pmatrix} [0] & diag(I_j B) \end{pmatrix}\right]$ denote a diagonal matrix whose i^{th} diagonal element is $\begin{pmatrix} [0] & diag(I_j B) \end{pmatrix}$ and $diag(I_j B)$ is a diagonal matrix whose j diagonal element is $I_j B$, I_j is the identity matrix of order j, records transition from $\Psi(n,1,0)$ to $\Psi(n,0,2)$. The matrix $diag\left[e_i B\right], 2 \leqslant \mathrm{i} \leqslant K$ is a diagonal matrix with i^{th} diagonal element $e_i B$ where e_i is a column vector of order i with 1 in the i^{th} place and 0 elsewhere, records transition from $\Psi(n,m,2)$ to $\Psi(n,m-1,2)$, $2 \leqslant m \leqslant K$.

(c) The matrix A_{13} lists transition rates in $\Psi(n,i,1)$ to $\Psi(n,j,1)$, $1 \leqslant i,j \leqslant N$. $Diag[T \oplus D_0]$ records transition rates in $\Psi(n,m,1)$ from $\Psi(n,m,1)$, $1 \leqslant \mathrm{m} \leqslant K-1$ and $T \oplus D_0 \oplus p_2 D_1$ if $m = K$. $diag[I_b \otimes p_2 D_1], 1 \leqslant \mathrm{i} \leqslant N-1$ lists transition rates in $\Psi(n,m+1,1)$ from $\Psi(n,m,1)$ and $diag[T^0 \otimes \beta \otimes I_r]$

corresponds to transition rates to $\Psi(n,m-1,1)$ from $\Psi(n,m,1)$, $2 \leqslant i \leqslant K$.

(d) The matrix A_{14} records transition rates in $\Psi(n,i,2)$ from $\Psi(n,j,1)$, $1 \leqslant i,j \leqslant K$ where the transition to $\Psi(n,0,2)$ from $\Psi(n,1,1)$ is $A_{14}^{(1)} = \begin{bmatrix} W_{11} & W_{12} & . & . & W_{1(N+1)} \end{bmatrix}$ and $W_{1i} = T^0 \otimes q_{i-1}e_i \otimes \alpha \otimes I_r$ where e_i is a row vector of order i with 1 in the 1^{st} place and zero elsewhere and other transition in A_{14} are [0] block matrices.

(e) The matrix A_{15} in A_1 records transition rates in $\Psi(n,m,0)$ from $\Psi(n,m,2)$ and such transition arises only when the system is busy with low priority service and m=0. Thus in A_{15}, transition to $\Psi(n,1,0)$ from $\Psi(n,0,2)$ is described by $A_{15}(1)$ and is described as follows: $A_{15}(1) = \begin{bmatrix} [0] & A_{15}^{(11)} \end{bmatrix}'$ where $A_{15}^{(11)} = diag\left(I_i(I_a \otimes \beta \otimes \bar{\delta} \otimes p_2 D_1)\right), 1 \leqslant i \leqslant N$.

(f) The matrix A_{16} records transition rates to $\Psi(n,j,2)$ from $\Psi(n,i,2)$, $0 \leqslant i,j \leqslant K$, where the matrix $I_{(N+1)(N+2)/2} \otimes (S \oplus D_0)$ lists transition rates in $\Psi(n,0,2)$ from $\Psi(n,0,2)$. $I_N \otimes (S \oplus D_0)$ lists transition rates within $\Psi(n,m,2)$ for $1 \leqslant m \leqslant K-1$ and $I_N \otimes (S \oplus D_0)$ if m=K while $I_N \otimes I_a \otimes p_2 D_1$ records transition in $\Psi(n,m+1,2)$ from $\Psi(n,m,2)$, $1 \leqslant m \leqslant K-1$.

4.2. Description of the phase type distribution for the services

The focus of this section is to describe the time it takes to process a job once it enters into the service facility. We assume that the service times are of phase type with representation given by $(\boldsymbol{\alpha}, S)$ of order a. The services are subject to preemptions. When the current service is preempted for the first time, counting clocks, which counts the number of priority customers served during his service period, preemption time (service time of current high priority customer), and threshold clock, respectively, will simultaneously be started. The preemption clock and threshold clock are of phase type with representations given by, respectively, $(\boldsymbol{\beta}, T)$ of order b, and $(\boldsymbol{\delta}, U)$ of order c. Once the high priority queue becomes empty/the number of high priority during the customer's service period reaches its maximum allowed level, whichever occurs first, the service of the preempted job will begin again. The service will resume (from the phase where the service got interrupted) or repeat (like a new service) depending on whether the interruption clock expired before the threshold clock or not. In addition, if the number of preemptions during the customers service reaches its maximum allowed level, then the service of the current job will not be preempted anymore once the service begins again for this job. On the other hand, if the preemption clock expires before the number of preemption reaches its maximum, determined by that customer at the beginning of his service, the counting end temporarily and will resume from this should there be another preemption for the existing job.

Following the procedure indicated in the previous chapter we will be able to compute the phase type distribution governing the effective service time. Mean of this phase type distribution can be computed in the usual manner. Thus the system is stable iff the effective service rate is larger than the arrival rate.

4.3. Stationary Distribution

Denote by $\mathbf{x}$ the stationary vector of $X(t)$, and partition $\mathbf{x}$ in to sub vectors $x(n, m, l), 0 \leqslant n \leqslant \infty, 0 \leqslant m \leqslant K, l = 0, 1, 2.$ and satisfying the condition $\mathbf{xQ=0}$ and $\mathbf{xe=1}$. The vectors $\mathbf{x(0)}$ and $\mathbf{x(1)}$ are obtained by solving the equations

$$\begin{aligned} \mathbf{x}(0)C_0 + \mathbf{x}(1)C_2 &= 0, \\ \mathbf{x}(0)C_1 + \mathbf{x}(1)(A_1 + RA_2) &= 0 \end{aligned} \tag{4.2}$$

subject to the normalizing condition

$$\mathbf{x}(0)e + \mathbf{x}(1)(I - R)^{-1}e = 1 \tag{4.3}$$

where $\mathbf{R}$ is the minimal non-negative solution to the matrix equation

$$A_0 + RA_1 + R^2A_2 = 0. \tag{4.4}$$

From these results, we obtain some interesting measures which helps in design of the system. Some of them are as follows:

4.4. Performance Measures

(a) The mean number of low units in the system, $E_lps = \sum_{n=0}^{\infty} n\ x(n)$.

(b) The mean number of high priority units in the system, $E_{hps} = \sum_{n=0}^{\infty} \sum_{m=0}^{K} \mathrm{m}\ x(n, m)$

(c) The fraction of time the server is idle=$\sum_{l=1}^{r} x(0,0,l)$.

(d) The fraction of time the low priority customer is preempted=$\sum_{n=1}^{\infty}\sum_{m=1}^{K} x(n,m,0)$

(e) Fraction of time the server is busy with high priority customer (with no preempted customer in the system)=$\sum_{n=1}^{\infty}\sum_{m=1}^{K} x(n,m,1)$.

(f) Fraction of time the server is busy with high priority customer (with preempted customer in the system)=
$\sum_{n=1}^{\infty}\sum_{m=1}^{K} x(n,m,2)$.

(g) Fraction of time the server is busy with low priority customer =
$\sum_{n=1}^{\infty}\sum_{j_1=0}^{N}\sum_{j_2=0}^{j_1}\sum_{i=1}^{a}\sum_{l=1}^{r} x_{n,0,2,j_1,j_2,i,l} + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\sum_{i_1=1}^{N}\sum_{i=1}^{a}\sum_{l=1}^{r} x_{n,m,2,i_1,i_1,i,l}$.

(h) Effective service rate of low priority customer= $\sum_{n=1}^{\infty}\sum_{j_1=0}^{N}\sum_{j_2=0}^{j_1}\sum_{i=1}^{a}\sum_{l=1}^{r} x_{n,0,2,j_1,j_2,i,l}\, S_i^0$
$+\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\sum_{i_1=1}^{N}\sum_{i=1}^{a}\sum_{l=1}^{r} x_{n,m,2,i_1,i_1,i,l}\, S_i^0$

4.5. Stability Condition.

We examine the system stability. What is needed is that the rate of drift to any lower level from a given level should be higher than that to a higher level. This means that the Markov chain is stable iff

$$\Pi A_0 e < \Pi A_2 e \tag{4.5}$$

where Π is the unique solution to $\Pi A = 0, \Pi e = 1$ where $A = A_0 + A_1 + A_2$. The above condition implies that the arrival rate should be less than the effective service time (reciprocal of the expected time to completely serve a customer).

4.6. Numerical Results

For the above input parameters we have plotted 4 graphs in Fig. 4.1. These represent variations in the mean number of low priority customers against increasing value of the probability of customers of that priority.

K=3, N=2, a=b=c=r=2; $D_0 = \begin{bmatrix} -6.5 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$,$D_1 = \begin{bmatrix} 6.0 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$, $S = \begin{bmatrix} -12.0 & 6.0 \\ 6.0 & -12.0 \end{bmatrix}$, $T = \begin{bmatrix} -12.0 & 3.0 \\ 3.0 & -12.0 \end{bmatrix}$, $U = \begin{bmatrix} -12.0 & 8.0 \\ 8.0 & -12.0 \end{bmatrix}$,
$S^0 = \begin{bmatrix} 6.0 & 6.0 \end{bmatrix}'$, $T^0 = \begin{bmatrix} 9.0 & 9.0 \end{bmatrix}'$, $U^0 = \begin{bmatrix} 4.0 & 4.0 \end{bmatrix}'$,
$\alpha = \begin{bmatrix} 0.4 & 0.6 \end{bmatrix}$, $\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}$, $\delta = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$.

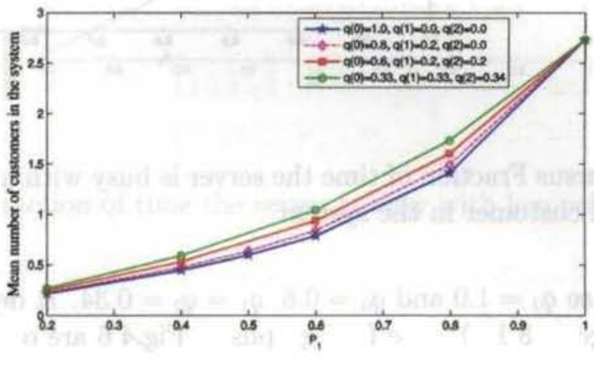

Fig. 4.1: p_1 versus Mean Number of Customers in the System

Depending on the outcome of the toss (the decision of the low priority customers to allow none, one or two priority customers to be served during its effective service time), we have the four graphs given in Fig. 4.1. (i) $q_0 = 1; q_1 = q_2 = 0$ (ii) $q_0 = 0.8, q_1 = 0.2, q_2 = 0$ (iii) $q_0 = 0.6; q_1 = q_2 = 0.2$ (iv) $q_0 = q_1 = 0.33; q_2 = 0.34$. At $p_1 = 0.2$ all the above result in almost the same mean number of customers; at $p_1 = 1.0$ (no high priority customer turns up) all these have the same value, which is not surprising. Fig. 4.2 provides the behavior of fraction of time server is busy with high priority customer for increasing value of p_1 in the four cases discussed in Fig. 4.1. Note that at $p_1 = 1$ this fraction turns out to be zero.

Fig. 4.3 provides the fraction of time the sever is busy with low priority customers with increasing value of p_1 in the four cases indicated in Fig. 4.1 and 4.3. As expected when $p_1 = 1$ (when all arrivals are of low priority), the fraction of time the server is busy serving low priority, turns out to be the maximum. Similarly the fraction of time the server is busy serving high priority customers decrease with increase in p_1 value (see Fig. 4.4)

An unexpected behavior of effective service time versus p_1, when $p_1 > 0.8$ is

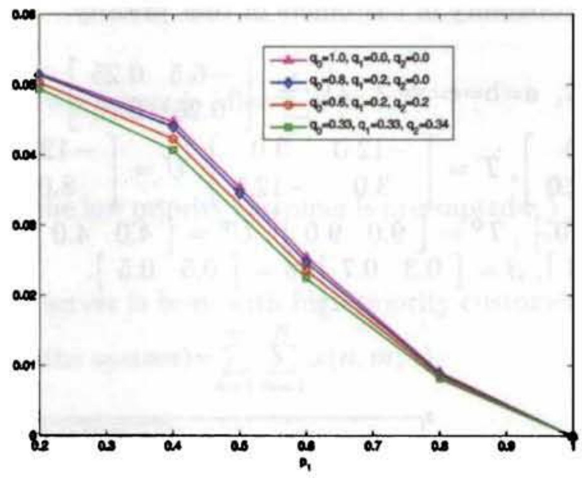

Fig. 4.2: p_1 versus Fraction of time the server is busy with high priority service but no preempted customer in the system

seen in the case $q_0 = 1.0$ and $q_0 = 0.6$, $q_1 = q_2 = 0.34$. It decreasing for increasing p_1 in the range (0.8,1.0). The four graphs in Fig.4.6 are on expected lines.

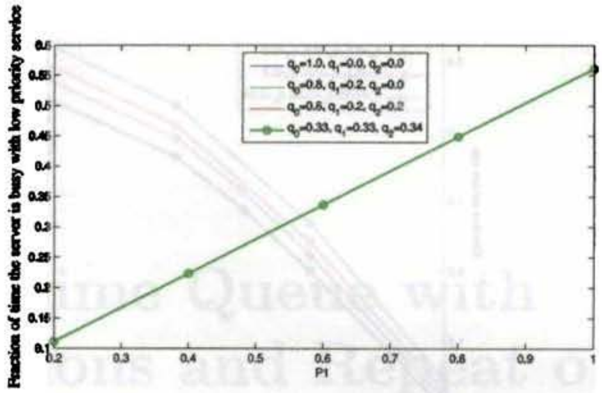

Fig. 4.3: p_1 versus Fraction of time the server is busy with low priority service

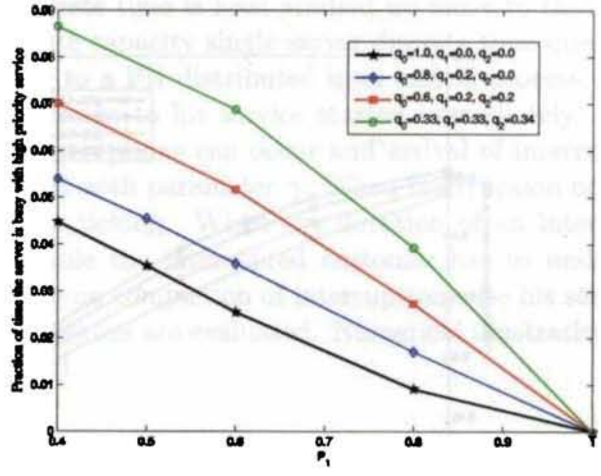

Fig. 4.4: p_1 versus Fraction of time the server is busy with high priority service

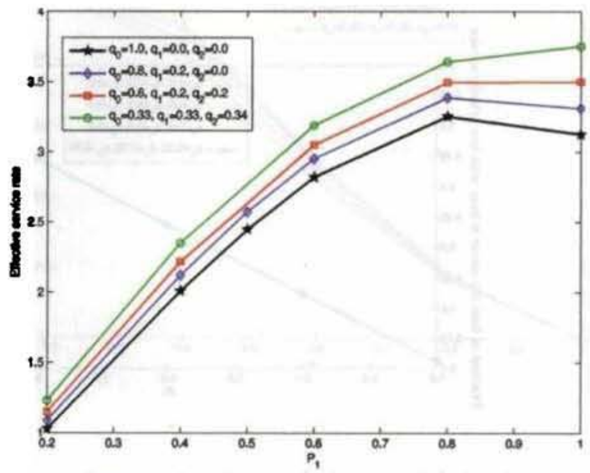

Fig. 4.5: p_1 versus Effective service rate of low priority customers

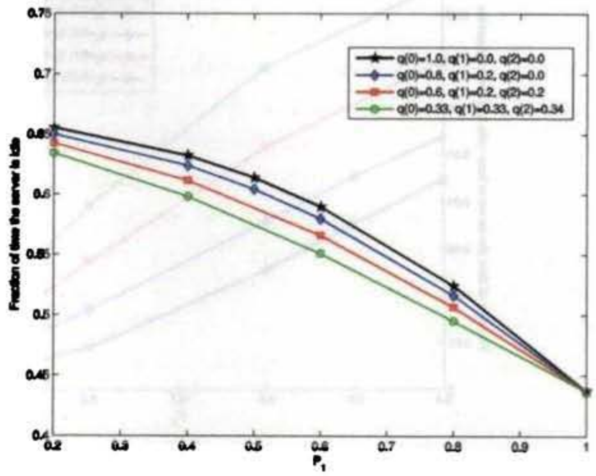

Fig. 4.6: p_1 versus fraction of time the server is idle

Chapter 5

Discrete time Queue with Interruptions and Repeat or Resumption of Service.

In chapter 2 through 4 we considered continuous time LIQBD which represented some of the lively problem in queues with interruption. We have progressively moved from simple models to more complex to more complex situations. Since the counterparts in discrete time is least studied we move to that in this chapter. In this chapter an infinite capacity single server discrete time queue to which customers arrive according to a PH distributed inter arrival process. If the system is idle an arriving customer go to his service started immediately. During the service more/one/none interruptions can occur and arrival of interruption process is geometrically distributed with parameter γ. When interruption occurs the threshold random clock starts ticking. When the duration of an interruption exceeds threshold random variable the interrupted customer has to undergo the service right from the beginning on completion of interruption; else his service is resumed. Several performance measures are evaluated. Numerical illustrations of the system behavior is also provided.

5.1. Model Description

Here we consider a discrete-time queueing system where the time axis is divided into intervals of equal length, called slots. In continuous-time queues, the probability of more than one event such as an arrival and a departure or two departures or more event taking place during a very short interval of time is zero, whereas it is not so in discrete-time queues. Let the time axis be marked by 0,1,2,3,........,m,.......Suppose the departures and interruptions occur in (m-,m),and the arrival occur in (m,m+). In this queueing model arrival process follows phase type distribution represented by (α, L) of dimension . Here the mean inter arrival time $l = 1 - L1$. The service time,interruption time and threshold random variable are also phase type distributed. The phase type representation of service time, interruption time and threshold random variable are (β, S),(η, T),(δ, U) and mean service times are $s = 1-S1$, $t' = 1-T1$, $u = 1-U1$ respectively. During service an interruption can occur where the interruption process is geometrically distributed with parameter γ. When the service is interrupted the server goes for interruption or vacation and threshold random clock starts ticking. Then a competition between interruption time and threshold random variable starts. On completion of interruption the interrupted customer may repeat /resume its service based on the following rule: If the interruption time exceeds threshold the interrupted customer gets its service repeated from the very beginning ; else its is resumed.

The state of the system at time n be ρ_n and its value is 0 if it is in interrupted state and 1 otherwise. If $\rho_n = 0$, let the number of customers in the system including the one in service be H_n, the service phase be B_n, the phase of arrival be J_n, the phase of interruption time be D_n and the threshold random variable be E_n. If $\rho_n = 1$ there are only three r.vs; the number of customers, service phase and arrival process. The process $\Omega = (H_n, \rho_n, B_n, D_n, E_n, J_n)$ is a discrete time markov chain whose n^{th} level is given by $l(n) = \bigcup_l \psi(n,l); n \geqslant 1, l = 0, 1$. The subsets of $\psi(n,l)$ are defined as $\{(n, 0, i_1, i_2, i_3, i_4); 1 \leqslant i_1 \leqslant a; 1 \leqslant i_2 \leqslant b; 0 \leqslant i_3 \leqslant c; 1 \leqslant i_4 \leqslant r\}$, ,$\{(n, 1, i_1, i_4); 1 \leqslant i_1 \leqslant a, 1 \leqslant i_4 \leqslant r\}$. Consider the Markov chain described by $\Delta = \{(0, J_n) \cup (H_n, 0, B_n, D_n, E_n, J_n) \cup (H_n, 1, B_n, J_n)\}, n \geqslant 0$.

The LIQBD has transition probability matrix $\Psi = \begin{bmatrix} C_0 & C_1 & & & \\ C_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & A_0 \\ & & & . & . \end{bmatrix}$

Transition from level 0 to 0 is represented by the matrix $C_0 = L$. When the system is idle an arriving customer is immediately taken for service. This part of ψ is given by C_1. We have C_1 given by $C_1 = \left[[0], \beta \otimes l'\alpha\right]$. The matrix C_2 corresponds to transitions on completion of the service of a customer the server is in level 1 and the matrix is as follows: $C_2 = \begin{bmatrix} [0] \\ S^0 \otimes L \end{bmatrix}$.

Here A_0 records transitions to $\psi(n+1, l)$ from $\psi(n, l)$ given by $A_0 = \begin{bmatrix} A_0^{(1)} & A_0^{(2)} \\ A_0^{(3)} & A_0^{(4)} \end{bmatrix}$. $A_0^{(1)}$ corresponds to transition during interruption period. The matrix $A_0^{(1)} = I_a \otimes$

$T \otimes \bar{U} \otimes l'\alpha$ where $\bar{U} = \begin{bmatrix} 1 & 0 \\ U & U^0 \end{bmatrix}$. Write the matrix $A_0^{(2)} = \begin{bmatrix} F_1 & F_2 & . & F_a \end{bmatrix}'$ where $F_j = T^0 \otimes \begin{bmatrix} \beta & \bar{e}_j & \bar{e}_j & . & . & \bar{e}_j \end{bmatrix}' \otimes l'\alpha$ corresponds to level change from n to $n+1$ when the interruption state changes to busy state. In a discrete time queueing system more than one event can takes place at an epoch. Here the interruption mode first changes to busy and then an arrival occur. Consider the matrix $A_0^{(3)} = \gamma S \otimes \eta \otimes \delta \otimes l'\alpha$. This matrix is related to arrival of customers when the system state changes from busy to interruption. The matrix$A_0^{(4)} = S \otimes l^1\alpha$ corresponds to arrival of customers during busy period.

Now we describe A_2. The matrix A_2 can be represented as $\begin{bmatrix} [0] & [0] \\ A_2^{(1)} & A_2^{(2)} \end{bmatrix}$ where $A_2^{(2)} = S^0\beta \otimes L$ and $A_2^{(1)} = \gamma S^0 e \otimes \eta \otimes \bar{\delta} \otimes L$ which designate level change from n to $n-1$. Here $A_2^{(1)}$ refers to departure of a customer along with the server changing from busy to interruption state and $A_2^{(2)} = S^0\beta \otimes L$ corresponds to departure and server continuing to be busy with the next customer in line.

The matrix A_1 which corresponds to transitions within level, and is as follows: $A_1 = \begin{bmatrix} A_1^{(1)} & A_1^{(2)} \\ A_1^{(3)} & A_1^{(4)} \end{bmatrix}$. Here continuing in the interruption state is represented by the matrix $A_1^{(1)} = I_a \otimes T \otimes \bar{U} \otimes L$ and is order , where only transition due to interruption, threshold and arrival phase change occur. Here $\bar{U} = \begin{bmatrix} 1 & \bar{0} \\ U^0 & U \end{bmatrix}$. The matrix $A_1^{(2)}$ in A_1 is of order $ab(c+1)r \times ar$ and $A_1^{(2)} = \begin{bmatrix} K_1 & K_2 & . & . & . & K_a \end{bmatrix}'$ where $K_j = T^0 \otimes \begin{bmatrix} \beta & \bar{e}_j & \bar{e}_j & . & \bar{e}_j \end{bmatrix}^T \otimes l'\alpha$ and e_j is a row vector of appropriate order with 1 in the j^{th} place and 0 elsewhere. The matrix $A_1^{(2)}$ correspond to transition from interruption to busy state. In this section when interruption time exceeds threshold, the interrupted customer gets its service repeated. Thus events in these transition corresponds to removal of interruption followed by repeat/resume of service. Occurrence of interruption during service is indicated by the matrix $A_1^{(3)}$ and is given by $A_1^{(3)} = \gamma S \otimes \eta \otimes \hat{\delta} \otimes L + \gamma S^0 e \otimes \eta \otimes \delta \otimes \boldsymbol{l}'\alpha$. The first term in $A_1^{(3)}$ stands for no service completion and arrival prior to interruption. The second term stands for a service completion and an arrival take place before it skips to interruption. The matrix $A_1^{(4)} = (1-\gamma)(S \otimes L + S^0\beta \otimes l'\alpha)$ where $\hat{\delta} = (0, \delta)$, shows that the system maintain its status when it is busy with no arrival and service completion in the first term and an arrival and departure in the second term.

5.2. Stationary Distribution

Let $A = A_0 + A_1 + A_2$ and $\pi = \pi A$, $\pi \mathbf{1} = 1$. The LIQBD is positive recurrent if $\pi A_0 1 < \pi A_2 1$. Let x be the invariant vector of P with $x = xP$, $x\mathbf{1} = 1$ where $x = [x_0, x_1, ...]$. By the matrix-geometric theorem (Neuts (1981)[25]) we have $x_{i+1} = x_i R$, where R is the minimal non-negative solution to $R = A_0 + RA_1 + R^2 A_2$ and the vectors x_0, x_1 are obtained by solving the equations

$$\begin{aligned} x_0 C_0 + x_1 C_2 &= x_0 \\ x_0 C_1 + x_1(A_1 + RA_2) &= x_1 \end{aligned} \tag{5.1}$$

subject to the normalizing condition

$$x_0 e + x_1 (I - R)^{-1} e = 1, \tag{5.2}$$

where $[x_0, x_1]$ is the invariant vector of the stochastic matrix $\begin{bmatrix} C_0 & C_1 \\ C_2 & A_1 + RA_2 \end{bmatrix}$. Also $x_{i+1} = x_i R, i \geqslant 1$.

5.3. Description of the service process in Discrete time queue

In this section we describe the time it takes to process a job once it enters into the service facility. We assume that service time, interruption time, threshold random variables are all independent phase type distributed random variables, with representations (α, S), (β, T) and (δ, U) respectively. Let X denote the duration of the effective service for a job. ie, X is the time between the arrival of a job to the service facility until it leaves the facility. There is no restriction on number of interruptions during the service of a customer. The interruption occurs to the service of a customer with rate γ which is distributed geometrically. We define $J_1(t), J_2(t), J_3(t)$ respectively, to the phase of service, phase of interruption and phase of threshold random variable. The states and their description are given in the following table:

$\{J_1,\}$ $1 \leqslant J_1 \leqslant a$	The service is in Phase J_1
$\{(J_1, J_2)\}$,	The threshold clock is expired, the interruption clock is in state J_2 the service phase is frozen in state J_1 .
$\{J_1, J_2, J_3\}$	The interruption clock is in state J_2, the threshold clock is in state J_3 and the service phase is frozen in state J_1.

The Markov process $\{J_1(t), J_2(t), J_3(t)\} : t \geqslant 0$ with absorbing state $*$ is defined on the state space
$\Omega = \{(0, J_2); 1 \leqslant J_2 \leqslant a\} \cup \{(J_1, J_2); 1 \leqslant J_1 \leqslant a, 1 \leqslant J_2 \leqslant b\} \cup$
$\{(J_1, J_2, J_3); 1 \leqslant J_1 \leqslant a, 1 \leqslant J_2 \leqslant b, 1 \leqslant J_1 \leqslant c\}$

and its infinitesimal generator matrix is given by

$$Q = \begin{bmatrix} \Delta & \Delta^0 \\ \bar{0} & \mathrm{I} \end{bmatrix} \tag{5.3}$$

where

$$\Delta = \begin{bmatrix} \Delta_{01} & 0 & \Delta_{03} \\ \Delta_{11} & \Delta_{12} & 0 \\ \Delta_{21} & \Delta_{22} & \Delta_{23} \end{bmatrix} \tag{5.4}$$

. Where

$$\begin{gathered} \Delta_{01} = S - \gamma I, \Delta_{03} = \gamma\beta \otimes \delta \otimes I \\ \Delta_{11} = T^0 \otimes \alpha, \Delta_{12} = I \otimes T \\ \Delta_{21} = I \otimes T^0 \otimes e, \Delta_{22} = I \otimes T \otimes U^0 \\ \Delta_{23} = I \otimes (T \oplus U) \\ \text{and} \\ \Delta^0 = \begin{pmatrix} S^0 & 0 & 0 \end{pmatrix}'. \end{gathered}$$

Theorem 5.3.1. *The effective service time, X, has phase type distribution with representation (ζ, Δ) of order a+bc+abc, where $\zeta = (\alpha, 0, 0)$ and Δ is given in 5.4.*

Proof. First note that a new service will begin in level 0 in state J_1 with probability α_{j_1}. Once the service begins it can end with or without interruptions and looking through all possible transitions, one will see that the transition matrix is given in Δ. Thus, the service time is nothing but the time until absorption into state * starting from level 0.
Mean $\mu'_\Delta = \zeta(-\Delta)^{-1}e$ and standard deviation of X is $\sigma\Delta = \sqrt{2\zeta(-\Delta)^{-2}e - \mu_{\Delta'}}$. Due to the special structure of the matrix Δgiven in (2), we can compute the mean as well as the standard deviation of X explicitly and recursively. First, we define

$$\beta(-\Delta)^{-1} = (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}),.$$

Using the above equation and exploiting the special structure of Δ, we get
u, v and w.

. Also we get the expression for $\mu_\Delta{}' = ue+ve+we$. □

5.4. Stability Condition.

Theorem 5.4.1. *The Markov chain $\{X_t, t \geqslant 0\}$ is stable if only if $\frac{1}{\alpha(-L)^{-1}e} < \frac{1}{\mu'_\Delta}$.*

5.5. Expected waiting time

We compute the expected waiting time of a tagged customer and is positioned r in the queue at the arrival epoch. We consider the Markov process $\{N(t), S(t), J(t)\}, t = 0, 1, 2, ...$, where $N(t)$ is the rank of the customer, $S(t) = \{0, 1\}$ as the state of the server, J(t) as the phase of the service process at time t. The r^{th} rank of the customer may decrease to $r - 1$ when the present customer in service leave the system after completing his service.
We arrange the state space of $X(t)$ as:$\{r, r-1, ..., 3, 2, 1\} \times \{0, 1\} \times \{1, 2, ..., a\} \times \{*\}$ where $*$ is an absorbing state which denote the tagged customer is selected for service.

The infinitesimal generator $\bar{Q} = \begin{bmatrix} \tilde{\Delta} & \tilde{\Delta}^0 \\ \bar{0} & 0 \end{bmatrix}$ where $\tilde{\Delta} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 & & & \\ & \tilde{A}_1 & \tilde{A}_2 & & \\ & & \tilde{A}_1 & \tilde{A}_2 & \\ & & & & \\ & & & & \dot{\tilde{A}}_1 \end{bmatrix}$

and $\tilde{\Delta}^0 = \begin{bmatrix} 0 & 0 & 0 & . & \tilde{B} \end{bmatrix}'$ where $\tilde{B} = [0, S^0]'$. The matrix $\tilde{A}_1 = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{13} & \tilde{A}_{14} \end{bmatrix}$ and $\tilde{A}_{11} = I_a \otimes T \otimes \bar{U}$,$\tilde{A}_{12} = \begin{bmatrix} B_{11} & B_{12} & & B_{1a} \end{bmatrix}'$ where $B_{1j} = T^0 \otimes \begin{bmatrix} \beta & \bar{e}_j & & . & \bar{e}_j \end{bmatrix}$,$\tilde{A}_{13} = \gamma S \otimes \eta \otimes \hat{\delta}$ and $\tilde{A}_{14} = S$. The matrix $\tilde{A}_2 = \begin{bmatrix} 0 & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}$ where $\tilde{A}_{21} = \gamma S^0 \beta \otimes \eta \otimes \hat{\delta}$ and $\tilde{A}_{22} = (S^0 \beta \otimes L)(1 - \gamma)$.

Expected waiting time of a customer who joins the queue as r^{th} customer $= -A_1^{-1}(I - (A_2 A_1^{-1})^r (I - A_2 A_1)^{-1}$

5.6. Performance Characteristics

Some useful descriptors of the model are listed below.

1. Mean number of customers in the system$= \sum_{n=1}^{\infty} n x_n e = x_1 (I - R)^{-2} e$

2. Fraction of time the server is busy$= \sum_{n=1}^{\infty} x_{n1} e$

3. Fraction of time the server remains interrupted$= \sum_{n=1}^{\infty} x_{n0} e$

4. Thus the fraction of time the server is idle=$x_0 e$

5. Fraction of time service is in interrupted state +Fraction of time service is going on=$\sum_{n=1}^{\infty} x_{n0}e + \sum_{n=1}^{\infty} x_{n1}e$

6. The rate at which server break down occurs=$\gamma \sum_{n=1}^{\infty} x_{n1}e$

7. The rate at which repair completion (removal of interruption) takes place before the threshold is reached) $R_{NT}^{c} = \sum_{n=1}^{\infty} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{r} x_{n,0,i,j,k,l}\, T_j^0$ where T_j^0 is the j^{th} component of T^0

8. Rate at which repair completion takes place after the threshold is reached $R_T^c = \sum_{n=1}^{\infty} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{l=1}^{r} x_{n,0,i,j,0,l}\, T_j^0$

9. Effective service rate $R_T^c = \sum_{n=1}^{\infty} \sum_{i=1}^{a} \sum_{l=1}^{t} x_{n,1,i,l}\, S_i^0$

10. The probability of a customer completing service without any interruption=P(service time< geometrically distributed random variable with parameter γ).

5.7. Numerical Results

In order to illustrate the performance of the system, we fix the following values:

$$L = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix}, S = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}, T = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix},$$

$$U = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{bmatrix} L^0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} S^0 = \begin{bmatrix} 0.7 \\ 0.7 \end{bmatrix}, T^0 = \begin{bmatrix} 0.7 \\ 0.7 \end{bmatrix}, U^0 = \begin{bmatrix} 0.3 \\ 0.3 \end{bmatrix}.$$

Fig. 5.1 indicates that mean number of customers in the system increases with increasing γ.

In this model, fraction of time the server is interrupted increases with increasing interruption rate. It is seen from Fig. 5.2 that the graph is almost linear in shape.

Like other models fraction of time the server is idle decreases with increasing interruption rate γ.

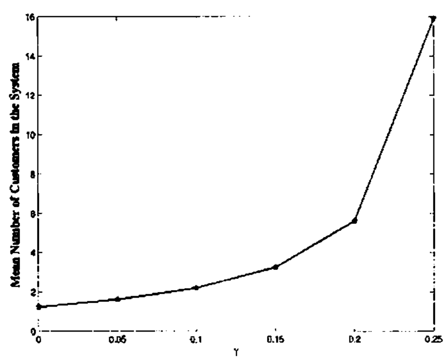

Fig. 5.1: Gamma versus Mean number of customers in the system

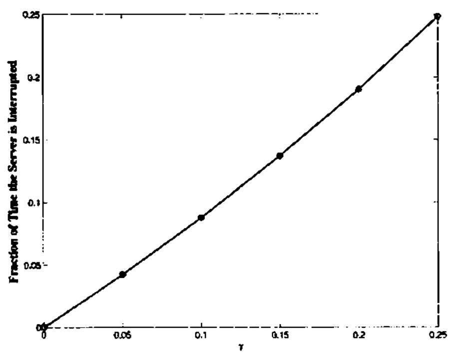

Fig. 5.2: Gamma versus Fraction of time the server is interrupted

With increasing interruption rate from 0 to 0.25, the effective service rate decreases gradually, which is seen in Fig. 5.4.

Like other continuous time models in this thesis, the rate at which sever break down also increases with increasing interruption rate and the curve is linear (see Fig. 5.5).

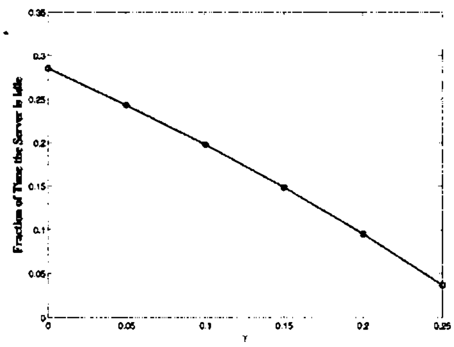

Fig. 5.3: Gamma versus Fraction of time the server is Idle

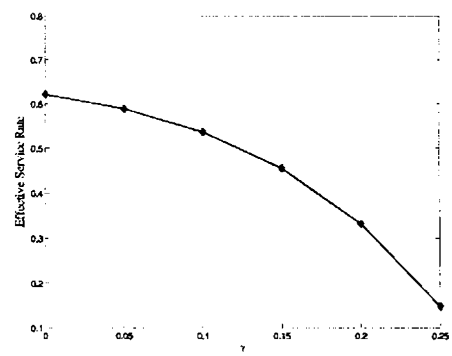

Fig. 5.4: Gamma versus Effective service Rate

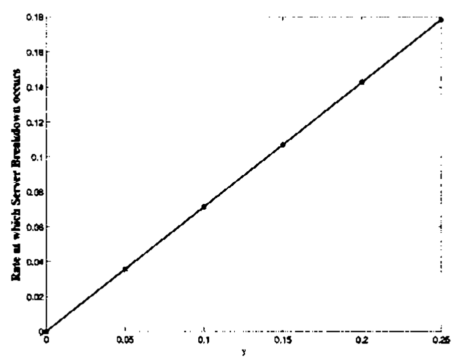

Fig. 5.5: Gamma versus Rate at which server breakdown occurs

Chapter 6

A Comparison Study and Conclusion

In this chapter we compare, wherever possible, the models that were discussed in chapters 2 through 4. Since chapter 5 is on discrete time queues, we do not go for a comparison of that with the rest of the models discussed. However a brief mention of what has been done there will be made.

We recall that in the second chapter two problems were discussed. In the first one we did not specify any upper bound on the number of interruptions a customer can encounter whereas in the second we imposed an upper bound. Then Fig. 6.1 shows that the mean number of customers in the one without a bound for interruption turns up to has much larger number of customers even moderately values of γ than that for the second problem. For example, when $\gamma = 6$ the first one has more than 50 customers on the average, whereas the second has only half that number. When a super clock to control the interruptions is introduced (chapter 3) there turned out to be further drop in the mean number of customer with increasing values of γ in comparison with the two models of chapter 2. However, on realization of the super clcok, the present interruption is not terminated to the customer in service; instead it is allowed to continue until its natural end and thereafter no interruption is permitted to that customer. This is model I of chapter 3. In contrast, in model 2 of that chapter the present interruption is instantly terminated and no further interruption is permitted to the customer in service. This has substantial effect on the mean number of customers in the system- a drastic reduction in that number is the outcome. So this model is an excellent design where customer impatience is involved.

In terms of several other performance measures also model 2 of chapter 3 is the best and worst performance is presented by model I of chapter 2. This is not

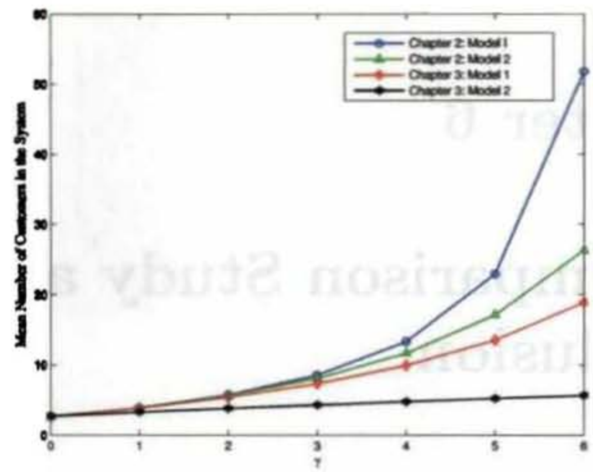

Fig. 6.1: Gamma versus Mean Number of Customers in the System

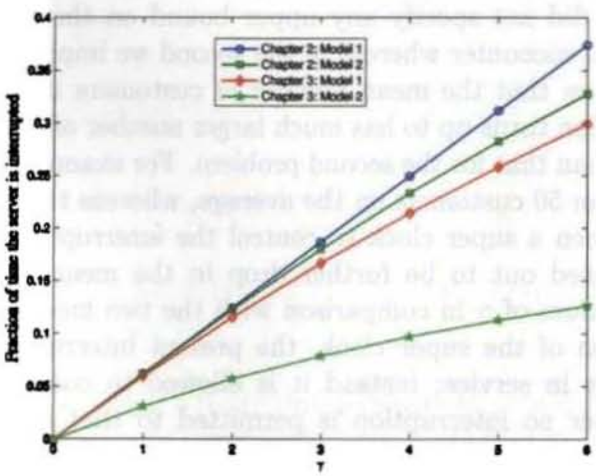

Fig. 6.2: Gamma versus Fraction of time the server is interrupted

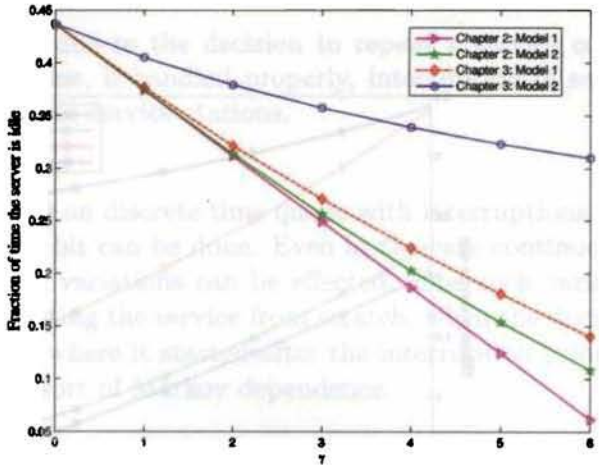

Fig. 6.3: Gamma versus Fraction of time the server is idle

surprising since model I of chapter 2 permits unlimited number of interruptions to a customer, whereas the remaining three models bring in restrictions on the number of interruptions. Fig.6.2, 6.3 and 6.4 provide comparison of the performance measures; fraction of time server is interrupted, fraction of time sever is idle and effective service rate against incidence of interruption, between the four models. In all cases model 2 of chapter 3 is superior to all other models and, as mentioned earlier, model I of chapter 2 is the worst performing one.

Finally, coming to the traffic intensity: note that model 2 of chapter 3 has least traffic intensity whereas model I of chapter 2 has the highest. This leads to the conclusion that the super clock has a telling effect on reduction in congestion; this is further enhanced by the termination of interruption of the present customer at the epoch of realization of the super clock itself. This scenario is presented in Fig. 6.5. Here the graph presents the growth of traffic intensity with increase in value of γ.

In Chapter 5 we concentrated only on the discrete time version of model I of Chapter 2. A few performance measures are computed.

In conclusion, this thesis is a study of queues with interruption of service. Second interpretations of interruption are considered depending on the context. In contrast to earlier work on queues with interruption, we have succeeded in designing a rule to decide whether to resume from where the service was interrupted or repeat the whole thing from the very beginning. The congestion that occurs in

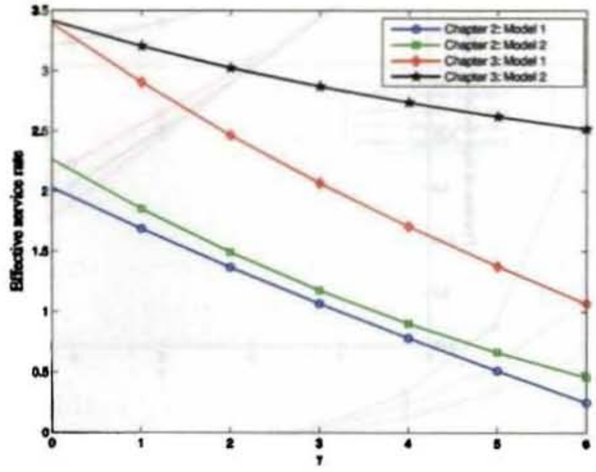

Fig. 6.4: Gamma versus Effective service rate

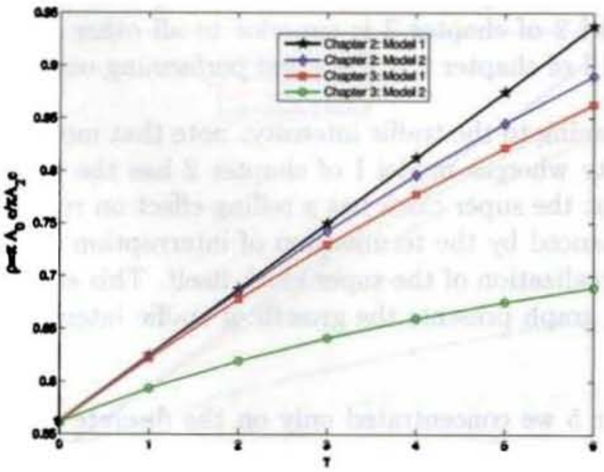

Fig. 6.5: Gamma versus rho=$\pi A_0 e/\pi A_2 e$

these models is mainly due to the decision to repeat a service consequent to an interruption. Nevertheless, if handled properly, interruption of service can be an added source of income to service stations.

We did not elaborate on discrete time queue with interruptions. This is a probable area where quite a bit can be done. Even in the case continuous time models discussed in this thesis, variations can be effected. One such variation is the following: instead of repeating the service from scratch, when the decision is to go for repetition, repeat from where it started after the interruption prior to the present one. This results in as sort of Markov dependence.

It may be noted that the models discussed in this thesis has a wide range of applications, a few of which have been indicated already.

References

[1] Alfa,A.S:Discrete time queues and matrix-analytic methods,TOP,Springer publication, 147-185, 2007.

[2] Atentia, I and Moreno, P: A Discrete-Time Geo/G/1 retrial queue with the server subject to starting failures, Annals of OR, 141:85-107, 2006.

[3] Atencia,I., Bouza,G and Moreno, P: An $M^{[X]}/G/1$ retrial queue with server breakdowns and constant rate of repeated attempts, Ann Oper Res , 225243,2008,

[4] Bellman, R : Introduction to Matrix Analysis . McGraw Hill Book Co., New York, 1960.

[5] Boxma,O, Mandjes, M and Kella, O: On a Queueing model with service interruptions, Probability in the Engineering and Informational Sciences, 537555, 2008.

[6] Chakravarthy, S.R:The batch Markovian arrival process: A review and future work. In Advances in Probability and Stochastic Processes, Ed. A. Krishnamoorthy, N. Raju and V. Ramaswami, Notable Publications, Inc., New Jersey, USA, 21-49, 2001.

[7] Chan,W, Bartoszynski, R and Pearl, D: Queues with beakdowns and customer discouragement, Probability and Mathematical Statistics, Vol.14, 77-87, 1993.

[8] Doshi, B.T; Queueing systems with vacations-a survey, Queueing Systems, Theory and Applications, V.1 n.1 P 29-66, 1986.

[9] Erlang, A.K : The theory of probabilities and telephone conversations, Nyt Tidsskrift Matematik, B.20, pp.139153, 33-39, 1909.

[10] Fiems, D, Steyaert, B and Bruneel: Analysis of a discrete-time GI-G-1 queueing model subjected to bursty interruptions, Computers and Operations Research, vol. 30/1, 2003.

[11] Fiems,D, Maertens,T and Brunee,H: Queueing systems with different types of interruptions, European Journal of Operations Research, vol. 188/3, pp. 838-845,2008.

[12] Fiems, D, Steyaert, B and Brunnel, H: Randomly interrupted GI-G-1 Queues, Service Strategies and Stability Issues, Annals of Operations Research, vol. 112, pp. 171183, 2002.

[13] Gross, D. and Harris, C.M.: Fundamentals of Queueing Theory, 3rd Ed.; JohnWiley and Sons,Inc. : New York, 1988.

[14] Gaver, D: A waiting line with interrupted service including priority, Journal of Royal Statistical Society, B24, -90, 1962.

[15] Gursoy, M.B and Xiao,W: Stochastic Decomposition in M/M/8 Queues with Markov Modulated Service Rates, Queueing Systems, 75-88, 2004.

[16] Haridass, M and Arumuganathan, R:Analysis of a Bulk Queue with Unreliable Server and Single Vacation, Int. J. Open Problems Compt. Math., Vol. 1, 2008.

[17] Julian Keilson: Queues Subject to Service Interruptions,The Annals of Mathematical Statistics, 1962.

[18] Karlin, S and Taylor, H.E: A first course in Stochastic Processes, 2^{nd} edition, Elsevier, 1975.

[19] Krishnamoorthy, A ,Pramod, P.K and Deepak, T.G: On a queue with interruptons and repeat/resumption of service, Non linear Analysis, Theory, Methods and Applications, Elsevier,2009.

[20] Karlin, S, Taylor, H.E: A first course in Stochastic Processes, 2^{nd} edition, Elsevier, 1975.

[21] Latouche, G. and Ramaswami, V.: Introduction to Matrix Analytic Methodsin Stochastic Modellings; ASA-SIAM.: Philadelphia, 1999.

[22] Li,J, Tian, N, Zhang, Z, G and Luh, H,P: Analysis of the M/G/1 queue with exponentially working vacations, Volune 61, 139-166, 2007.

[23] Lucatoni, D: New results on the single server queue with a Batch Markovian Arrival Process, Stochastic Models,7, 1-46, 1991.

[24] Marcus, M., and Minc, H.: A Survey of matrix theory and matrix inequalities, Allyn and Bacon, Boston, MA, 1964.

[25] Neuts, M.F.: Matrix Geometric Solutionsin Stochastic Models:An Algorithmic Approach, 2nd Ed.; Johns Hopkins University Press, Baltimore, 1981.

[26] Neuts, M.F.: Structured Stochastic Matrices of M/G/1 type and their applications, Marcel Dekkar, 1989.

[27] Nunez-Queija, R, Sojourn times in a processor sharing queue with service interruptions, Queueing Systems, 351-386, 2000.

[28] Pang, G and Whitt, W: Heavy-traffic limits for many-server queues with service interruptions, Queueing Syst , 167202, 2009.

[29] Quan-Lin Li, Yu Ying and Yiqiang Q. Zhao: A BMAP/G/1 Retrial Queue with a Server Subject to Breakdowns and Repairs, Ann Oper Res , 141: 233270, 2006

[30] Rembowski, B, Apriority queue with interruptions of service permitted after a time quantum, Acta Informatica, 241251, 1985.

[31] Tadj, L and Choudhary, G: Optimal design and Control of Queues, TOP, 13(2), 359-412, 2005.

[32] Takagi, H: Queueing Analysis, A Foundation of Performance Evaluation, Vol.1, Vacation and Priority Systems, Part1, Elsevier Science Publications B.V, Amsterdam, 1991.

[33] Takine, T and Sengupta, B: A single server queue with service interruptions, Queueing Systems 26, 285-300,1997.

[34] Tewfik Kernane:A Single Server Retrial Queue with Different Types of Server Interruptions(submitted Feb 2009).

[35] Ibe,O,C and Trivedi, K.S: Two Queues with alternating service and server breakdown, Queueing Systems, 253-268, 1990.

[36] White, H and Christie, L.S : Queueing with preemptive priorities or with breakdown, Operations Research, Vol.6, 79-95,1958.

List of Publications

Papers presented

- Krishnamoorthy,A, Pramod, P, K; On a Discrete time Queues with Service Interruptions; Joint Statistical Meeting and International Conference on Statistics, Probability and Related areas.

- Krishnamoorthy, A, Pramod, P, K, Deepak, T, G; On a Queue with Interruptions and Repeat / Resumption of Service. Invited talk by A. Krishnamoorthy to the session on queues in the Fourth International Conference on Non-Linear Analysis, Orlando, Florida, USA, July 2008.

- Krishnamoorthy, A, Pramod, P, K;On a Queue with Interruptions Controlled by a Bound on Maximum Number of Interruptions; National Research Scholars meet of Mathematics and Statistics, IIT Kanpur, December 2008.

- Krishnamoorthy, A, Pramod, P, K; On a Queue with Interruptions Controlled by a Super Clock and a Bound on Maximum Number of Interruptions. Invited talk by Krishamoothy, A, to the special session on Matrix Analytic Methods of INFORMS Applied Probability Conference, Cornel University, Ithaca, USA, July 2009.

Papers published / communicated

- A.Krishnamoorthy, P.K. Pramod, T..G. Deepak, On a Queue with Interruptions and Repeat/Resumption of service, Non-Linear Analysis, Elsevier Journal, 2009.

- Krishnamoorthy, A, Pramod, P, K; On a Queue with Interruption Controlled by a Super Clock and a Bound on Maximum Number of Interruptions, Communicated.

- Krishnamoorthy,A, Pramod, P, K, Chakravarthy, S, R;On a Note on Characterizing Interruptions with Phase Type Distribution, Communicated.