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 $\lambda, \lambda > 0$, 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].