Chapter 1

INTRODUCTION

The objective of this thesis is to study the time dependent behaviour of some complex queueing and inventory models. It contains a detailed analysis of the basic stochastic processes underlying these models. In the theory of queues, analysis of time dependent behaviour is an area very little developed compared to steady state theory. Time dependence seems certainly worth studying from an application point of view, but unfortunately, the analytic difficulties are considerable. Closed form solutions are complicated even for such simple models as M/M/1. Outside M/M/1, time dependent solutions have been found only in special cases and involve most often double transforms which provide very little insight into the behaviour of the queueing systems themselves. In inventory theory also there is not much results available giving the time dependent solution of the system size probabilities. Our emphasis is on explicit results free from all types of transforms and the method used may be of special interest to a wide variety of problems having regenerative structure.
In this thesis we consider different queueing models with Poisson arrivals, general service, with and without vacations to the server and derive transient solutions for all the models. Also we study some \((s,S)\) inventory systems under the assumptions of random lead times, random quantity replenished, bulk demands and vacations to the server. In the following chapters we analyse each of these models. Each chapter contains a self introduction and some important references. In that sense, each chapter is self contained. In this chapter we give a brief general introduction to the subject and some related topics.

1.1 Queuing Theory

The question may arise whether after about 80 years of research in queueing theory, it is still possible to make a substantial contribution to the theory and to come up with some new results. To account for our positive answer to this question, and to place the present work in its proper context, we first give a short historical survey of the origin and development of queueing theory and subsequently discuss the current state of affairs.

Historically, Johannisn's "Waiting Times and Number of Calls" (an article published in 1907 and
reprinted in Post Office Electrical Engineers Journal, London, October 1910) seems to be the first paper in the subject. But the method used in this paper is not mathematically exact and therefore, from the point of view of exact treatment, the paper that has historic importance is A.K. Erlang's "The theory of Probabilities and Telephone Conversations", published in 1909. In this paper, he lays the foundation for the place of Poisson (and hence exponential) distribution in congestion theory. His papers written in the next 20 years contain some of the most important concepts and techniques. That include the notion of statistical equilibrium, the method of writing down the balance of state equations (later called Chapman-Kolmogorov equation). Erlang delay probability and the phase method of Erlang.

Until 1940, the majority of the contribution to queuing theory was made by people active in the field of telephone traffic problems. After the Second World War, the field of operations research rapidly developed and queuing applications were also found in production planning, inventory control and maintenance problems. In this period, much theoretically oriented research on queuing problems were done.
In the fifties and sixties, the theory reached a very high mathematical level (see Cohen (1969) and Takacs (1962)). Advanced mathematical techniques like transform methods, Weiner Hopf decomposition and function theoretic tools were developed and refined. This research resulted in a number of elegant mathematical solutions.

In particular, noting the inadequacy of the equilibrium theory in many queue situations, Pollaczek in 1934 began investigations of the behaviour of the system during a finite interval. Since then he has done considerable work in the analytical behavioural study of queueing systems. The trend towards the analytical study of the basic stochastic processes of the system has continued, and queueing theory has proved to be a fertile field for researchers who wanted to do fundamental research on stochastic processes involving mathematical models.

The processes involved are not simple and for the time dependent analysis, more sophisticated mathematical procedures are necessary. For instance, for the queue with Poisson arrivals and exponential service times, under statistical equilibrium, the balance of state equations are simple and the limiting distribution of the queue size is obtained by recursive arguments and induction. But for the time dependent solution, the use of transforms is necessary.
The time dependent solution was first given by Bailey (1954b) and Ledermann and Reuter (1954). While Bailey used the method of generating functions for the differential equations, Ledermann and Reuter used spectral theory for their solution.

To analyse the case of M/G/1 queues, Kendall (1953) has used the method of regeneration points due to Palm. The method of supplementary variables investigated by Cox (1955) was already used in W. Weston's thesis in 1942. It is extensively discussed in the book by Gnedenko and Kovalenko (1968).

The study of bulk queues is considered to be originated with the pioneering work of Bailey (1954a). In a way the study of bulk queues may be said to have begun with Erlang's investigation of M/M_k/1; for its solution contains implicitly the solution of the model M^k/M/1. Bailey studied the stationary behaviour of a single server queue having simple Poisson input, intermittently available server and service in batches of fixed maximum size. The results of this study are given in terms of probability generating functions, the evaluation of which requires determining the zeroes of a polynomial. This study was followed by a series of papers involving the treatment of queueing processes with group arrivals or batch service. Gaver (1959) seems to
be the first to take up specifically queues involving group arrivals. For more details on bulk queues, one may refer to Medhi (1984). For a detailed treatment of queueing systems and for further references, one can refer any one of the standard books on the subject like, Saaty (1961), Takacs (1962), Cohen (1969), Frawh (1965, 1980), Gnendenko and Kovalenko (1968), Cooper (1972), Gross and Harris (1974), Kleinrock (1975) and Asmussen (1987).

One important name worth mentioning in the study of time dependent behaviour of queueing systems is that of Takacs. He has studied the transient behaviour of \( M/G/1 \) and \( E_k/G/1 \) models. Also the influence of his study of virtual waiting time process has been tremendous in the development of queueing theory. For more details, one may refer to Takacs (1962).

Queueing systems in which the service process is subject to interruptions resulting from unscheduled breakdowns of servers, scheduled off periods, arrival of customers with pre-emptive or non-preemptive priorities or the server working on primary and secondary customers arise naturally. The impact of these service interruptions on the performance of a queueing system will depend on the specific interaction between the interruption process and the service process.
Queueing models with interruptions and their connection to priority models were first studied by White and Christie (1958), who considered the case with exponential service, on-time and off-time distributions. Their results were extended by Gaver (1962), Keilson (1962), Avi-Itzhak and Naor (1962) and Thiruvengadam (1963) to models with general service time and off-time distributions but exponential on-times. When the on periods are not exponential, the problem became very difficult and one such model is studied by Federgruen and Green (1986). A detailed analysis of single server queueing system with server failure is given in Gnedenko and Kovalenko (1968).

Another variation of the interruption model is the vacation model. In this the queueing system incurs a start-up delay whenever an idle period ends or the server takes vacation periods. The vacation model includes server working on primary and secondary customers also. Analysis of queueing systems with vacations to the server is motivated by the study of cyclic queues and Miller (1964) was the first to study such a system. Miller analysed a system in which the server goes for a vacation (a 'rest period') of random length whenever it becomes idle. He also considered a system in which the server behaves normally but the first customer arriving to an empty system has a special service time. These types of systems and similar ones were also
examined by Welsch (1954), Avi-Itzhak, Maxwell and
Miller (1965), Cooper (1970), Pakes (1973), Lemoine (1975),
Levy and Yechiali (1975), Heyman (1977), Van der Duyn
Schouten (1978), Shantikumar (1980, 1982) and Scholl and

All the above models (having rest periods, set-up
time, starter, interruptions etc.) can be jointly called
as vacation models. While the queue with interruption has
preemptive priority for vacation, other types of vacations
have least priority among all work with vacation taken when
the system is empty. Variations of vacation models are
available with single and multiple vacations and exhaustive
and non-exhaustive service disciplines.

When the system becomes empty, server starts a
vacation and the server keeps on taking vacations until,
on return from a vacation, at least one customer is present.
This is called a multiple vacation system. The server taking
exactly one vacation at the end of each busy period, is
called a single vacation system. We say that a vacation
model has the property of exhaustive service in case each
time the server becomes available, he works in a continuous
manner until the system becomes empty. Systems with a
vacation period beginning after every service completion,
(or after any vacation period if the queue is empty) is
known as the single service discipline. There is another non-exhaustive service discipline which is a generalization of both exhaustive and single service disciplines known as the Bernoulli schedule discipline defined as follows. After each service completion, the server takes a vacation with probability $p$ and starts a new service with probability $1-p$. If the system is empty, after a service completion or vacation completion, server always takes a vacation and after any vacation if customers are present server resumes service.


The main results in the vacation systems is the delay analysis by decomposition. The 'stochastic decomposition property' of $M/G/1$ queueing system with vacation says
that the (stationary) number of customers present in the system at a random point in time is distributed as the sum of two independent random variables. One is the (stationary) number of customers present in the corresponding standard $M/G/1$ queue (i.e. without vacation) at a random point in time and the other is the number of arrivals in the forward recurrence time of vacation period. For more details on queueing systems with vacations one may refer to Doshi (1986).

All the above models assume the existence of a stationary distribution and studied some aspects of the queue length and waiting time distribution. Some aspects of the dynamic behaviour of $M/G/1$ queues with vacations is studied by Keilson and Servi (1986c). The time dependent solution for a finite capacity $M/G/1$ queueing system with vacations to the server is given by Jacob and Krishnamoorthy (1987). They have introduced a new method namely the convolution product of matrices, whose elements are the transition probability density functions, to arrive at the solution. Using renewal theoretic arguments, they have given explicit expressions for the time dependent system size probabilities at arbitrary epochs and also the probability distribution of the virtual waiting time in the queue at time $t$. Time dependent solution for a finite capacity $M/G^{b}/1$ queueing system with vacations to the
server is given by Jacob and Madhusoodanan (1988), using the theory of regenerative processes. In this thesis we extend these results to a number of variations of the M/G/1 queue.

1.2. INVENTORY THEORY

By inventory, we mean the measured amount of some items which varies in quantity over time in response to a 'demand' process, which operates to diminish the stock, and a 'replenishment' process, which operates to increase it. Usually the demand is not subject to control, but the timing and magnitude of the replenishment can be regulated.

The real need for analysis of an inventory system was first recognized in industries that had a combination of production scheduling problems and inventory problems. That is, situations where items were produced in lots and then stored at a factory warehouse. The earliest derivation of what is often called the "simple lot size formula" was obtained by Ford Harris in 1915. The same formula have been developed, independently by many researchers since then. It is often referred to as "Wilson's formula", since it was derived by R.H. Wilson as an integral part of an inventory control scheme.
During World War II, a useful stochastic model was developed as the Christmas tree model. Shortly thereafter, a stochastic version of the simple lot size model was developed by Whitin, whose book published in 1953 was the first book in English which dealt in any detail with stochastic inventory models. The paper by the economists, Arrow, Harris and Marschak (1951) was one of the first to provide a rigorous mathematical analysis of a simple type of inventory model. It was followed by the often quoted and rather abstract papers by the mathematicians Dvoretzky, Kiefer and Wolfowitz (1952, 1953). Since then a number of papers by mathematicians have appeared.

A valuable review of the problems in probability theory of storage systems is given by Gani (1957). A systematic account of probabilistic treatment in the study of inventory systems using renewal theoretic arguments is given by Arrow, Karlin and Scarf (1958). Hadley and Whitin (1963) deals with the application of mathematical models to practical situations. The cost analysis of different inventory systems is given by Haddor (1966). Tijms (1972) gives a detailed analysis of inventory systems under (s,S) policy. A practical treatment of the (s,S) inventory systems can be found in the recent books by Silver and Peterson (1985) and Tijms (1986).
Gaver (1959) analyses the case of an (s, S) inventory system with compound Poisson demand and random lead times. Some aspects of (s, S) inventory system with arbitrary interarrival time of demands and random lead times is discussed by Finch (1961). Veinott (1966) gives a detailed review of the status of mathematical inventory theory upto 1965. Gross and Harris (1971) and Gross, Harris and Lechner (1971) deal with one for one ordering inventory policies with state dependent lead times. Sivazlian (1974) considers an (s, S) inventory system with arbitrary interarrival time of demands and zero lead time and Srinivasan (1979) extended these results to allow the lead time to follow arbitrary distribution. The case of an (s, S) inventory system with bulk demands and constant lead time is analysed by Sahin (1979). Also Sahin (1983) discussed the problem of an (s, S) inventory system with compound renewal demand and random lead times and he obtained the binomial moments of the inventory deficit.

A continuous review (s, S) inventory system in random environment is analysed by Feldman (1978). Richards (1979) analyses an (s, S) inventory system with compound Poisson demand. Algorithms for a continuous review (s, S) inventory system in which the demand is according to a versatile Markovian point process is given
by Ramaswami (1981). An inventory system with two ordering levels and random lead times is analysed by Thangaraj and Ramanarayanan (1985). Approximation for the single-product production-inventory problem with compound Poisson demand and two possible production rates where the product is continuously added to inventory is given by De Kok, Tijms and Van der Duyn Schouten (1986). Using Markov decision drift processes, Hordijk and Van der Duyn Schouten (1986) examines the optimality of \((s,S)\) policy in a continuous review inventory model with constant lead time when the demand process is a superposition of a compound Poisson process and a continuous deterministic process.

An important variation of the inventory problem is the perishable commodity inventory system. For details of the work in this area, one may refer to the excellent survey given by Nahmias (1982). The continuous review perishable inventory system can be identified with queueing systems with impatient customers as viewed by Kaspi and Perry (1983, 1984).

In the case of random lead times, the concept of vacations to the server during dry period is introduced in inventory systems by Daniel and Ramanarayanan (1987 a,b).
Usha, Ramanarayanan and Jacob (1987) analyses the case of finite backlog of demands and vacations to the server. Several other models with vacations to the server, bulk demands, varying ordering levels etc. can be found in Jacob (1987).

1.3. NOTATIONS

Here we introduce the following notations that are frequently used in this thesis.

$a$ denotes the convolution operator.

$f^{*n}(x)$ denotes the n-fold convolution of $f(x)$ with itself.

For a distribution function $F(x)$, $\bar{F}(x) = 1 - F(x)$, the survival probability.

Now we define the convolution of two matrices as follows.

If $A(t) = [a_{ij}(t)]$ is a matrix of order $m \times p$, and $B(t) = [b_{ij}(t)]$ is a matrix of order $p \times n$, then $(A*B)(t) = [c_{ij}(t)]$ is a matrix of order $m \times n$ whose elements are given by

$$c_{ij}(t) = \int_0^t \sum_{k=1}^p a_{ik}(u) b_{kj}(t-u) du$$
For a square matrix $A(x)$ of order $m$, let $A^{\times 0}(x)$ be the identity matrix of order $m$ and for $n \geq 1$, let $A^{\times n}(x)$ be the $n$-fold convolution of $A(x)$ with itself.

1.4 RENEWAL THEORY

Renewal processes are the simplest regenerative stochastic processes. To define a renewal process, let 
\[ \{X_n, n=1,2, \ldots \} \]
be a sequence of non-negative independent random variables. Assume that $\Pr \{X_n=0\} < 1$, and that the random variables are identically distributed with distribution function $F(.)$. Since $X_n$ is non-negative, it follows that $E X_n$ exists.

Let 
\[ S_0 = 0, \quad S_n = X_1 + X_2 + \ldots + X_n \quad \text{for} \quad n \geq 1, \]
and let 
\[ F_n(x) = \Pr \{ S_n \leq x \} \]
be the distribution function of $S_n$.

Since $X_i$'s are i.i.d., $F_n(x) = F^{\times n}(x)$.

Define the random variable
\[ N(t) = \sup \{ n \mid S_n \leq t \} \]
The process \( \{N(t), t \geq 0\} \) is called a renewal process.
If for some \( n \), \( S_n = t \), then the \( n \)th renewal is said to occur at \( t \); \( S_n \) gives the time of the \( n \)th renewal and is called the \( n \)th renewal epoch. The random variable \( N(t) \) gives the number of renewals in the interval \((0, t]\).

The function \( M(t) = E[N(t)] \) is called the renewal function of the process with distribution function \( F \). It is easy to see that

\[
N(t) \geq n \iff S_n \leq t
\]

Thus the distribution of \( N(t) \) is given by

\[
Pr \{ N(t) = n \} = F^n(t) - F^{n+1}(t)
\]

and the expected number of renewals is given by

\[
M(t) = \sum_{n=1}^{\infty} F^n(t)
\]

Its derivative

\[
m(t) = \frac{dM(t)}{dt} = \sum_{n=1}^{\infty} f^n(t)
\]

is the renewal density function, assuming the density function \( f(t) \) exists. To better understand the meaning of \( m(t) \), let us consider the increment of \( M(t) \).
\( \delta M(t) = M(t+\delta t) - M(t) \)

\[
= \sum_{n=1}^{\infty} \left[ F^*n(t+\delta t) - F^*n(t) \right]
\]

\[
= \sum_{n=1}^{\infty} \Pr \{ t < S_n \leq t+\delta t \}
\]

On the other hand, we have for \( \delta t \to 0 \),

\( \Pr \{ \text{more than one renewal point in } (t, t+\delta t) \} \to O(\delta t) \)

For \( \delta t \to 0 \), we get

\[
\delta M(t) = \Pr \{ S_1 \text{ or } S_2 \text{ or } S_3 \text{ or } \ldots \text{ lies in } (t, t+\delta t) \}
\]

\[
\to m(t)\delta t
\]

This interpretation of renewal density is important in practical applications.

Now, suppose that the first interoccurrence time \( X_1 \) has a distribution \( G \) which is different from the common distribution \( F \) of the remaining interoccurrence times \( X_2, X_3, X_4, \ldots \).

As before let us define

\[
S_0 = 0, \quad S_n = \sum_{i=1}^{n} X_i
\]
and \[ N_D(t) = \text{Sup} \{ n \mid S_n \leq t \} \]

The stochastic process \( \{N_D(t), t \geq 0\} \) is called a Delayed or Modified renewal process.

Here we have

\[
\text{Pr} \{ N_D(t) = n \} = G*F^{n-1}(t) - G*F^n(t)
\]

so that the modified renewal function is

\[
N_D(t) = \mathbb{E} [ N_D(t) ] = \sum_{n=0}^{\infty} G*F^n(t)
\]

The modified renewal density function is given by

\[
\nu_D(t) = N_D'(t) = \sum_{n=0}^{\infty} G*F^n(t)
\]

under the additional assumptions that the density functions \( g(x) = G'(x) \) and \( f(x) = F'(x) \) exists.

Now, consider a stochastic process \( \{ X(t), t \geq 0 \} \) with state space \( \{ 0,1,2, \ldots \} \), having the property that there
exist time points at which the process (probabilistically) restarts itself. That is, suppose that with probability one, there exists a time $T_1$, such that the continuation of the process beyond $T_1$ is a probabilistic replica of the whole process starting at 0. Note that this property implies the existence of further times $T_2$, $T_3$, ..., having the same property as $T_1$. Such a stochastic process is known as a regenerative process.

From the above, it follows that $\{T_1, T_2, \ldots\}$ forms a renewal process; and we shall say that a cycle is completed every time a renewal occurs. It is easy to see that a renewal process is regenerative, and $T_1$ represents the time of the first renewal.

For details on renewal theory, one may refer to Cox (1962), Feller (1983), Ross (1970) or Cinlar (1975).

1.5. SUMMARY OF THE WORK INCLUDED IN THE THESIS

In this thesis we study the time dependent behaviour of some complex queueing and inventory models. In our analysis, renewal theory plays an important role. In each model, identifying the regeneration points and using matrix convolutions, we obtain the required transition probability density functions.
The thesis is divided into seven chapters.

In the second chapter, we consider a finite capacity general bulk service queueing system. The arrival of customers is according to a homogeneous Poisson process and the service times are generally distributed independent random variables whose distribution depends on the size of the batch being served. Customers are served in batches according to the general bulk service rule (See Neuts (1967)). Using renewal theoretic arguments we derive the probability density function of the busy period and the time dependent system size probabilities at arbitrary time points. Also we derive the probability distribution of the virtual waiting time in the queue at any time $t$.

The next chapter is devoted to the study of an infinite capacity $M^X/G/1$ queueing system with vacations to the server. The arrival of customers is according to a compound Poisson process and service is done one by one with service time following a general distribution. Under exhaustive service discipline, server takes vacations for a random period having a general distribution. The server keeps on taking vacations until on return from a vacation, at least one customer is present. Here also we derive the probability density function of the busy period, time dependent system size probabilities at arbitrary epochs.
and the probability distribution for the virtual waiting time in the queue, using renewal theoretic arguments.

A finite capacity M/G^b/l queueing system with vacations to the server is analysed in chapter four. The general bulk service rule is modified to allow the arriving customers to enter for service, if the maximum service capacity is not attained, without altering the service time. Even if service of a batch is in progress with less than 'b' customers, all the arrivals may not be interested to join for partial service. So we assume that an arriving customer enter for partial service with probability p and wait for full service with probability 1-p, till the service capacity is attained. Server goes for vacation whenever he finds less than 'a' customers in the system and this is a multiple vacation. The vacation periods are independent and identically distributed random variables following a general distribution. Using the theory of regenerative processes, explicit expressions for the busy period, the time dependent system size probabilities and the probability distribution of the virtual waiting time are derived.

In chapter five, we consider a finite capacity M/G^B/l vacation system with Bernoulli schedules. The arrival of customers is according to a homogeneous Poisson process and the service is in batches of maximum capacity B.
The service times are generally distributed random variables which depend upon the size of the batch being served. The Bernoulli vacation model is defined as follows. After each service completion the server starts a new service, if a customer is present, with probability $p$ and takes a vacation with probability $1-p$. If the system is empty after a service completion or a vacation completion, the server always takes a vacation. The decision about taking a vacation after each service completion or vacation completion are independent. If the server finds at least one customer upon return from a vacation, he starts service of the batch. Also the vacations are independent and identically distributed random variables following a general distribution. Here also we derive explicit expressions for the probability density function of a busy period, time dependent system size probabilities and the probability distribution of the virtual waiting time in the queue, using the theory of regenerative processes.

In the next two chapters, we analyse some continuous review $(s, S)$ inventory systems with random lead times. Chapter six is devoted to the study of a single item $(s, S)$ inventory system in which the quantity replenished is random. The demand is for one unit at a time and the
interarrival time of demands and lead times are independent sequences of independent and identically distributed random variables following general distributions. Whenever the inventory level drops to $s$, an order is placed for $S-s$ units. However, the quantity replenished is a random variable that can assume values $s+1, \ldots, S-s$. Explicit expressions for the probability mass function of the stock level at arbitrary epochs are derived, using renewal theoretic arguments. An expression for the total cost over a period of time of length $t$ is obtained. Then we consider the special case of zero lead time and derive the time dependent as well as the stationary distribution of the inventory level. Using this, the associated optimization problem is discussed in detail. Finally, some numerical examples are given.

In the last chapter, we consider an $(s,S)$ inventory system with random lead times depending on the size of the order. The time between successive demands and quantities demanded at these points are independent sequences of independent and identically distributed random variables following general distributions. Whenever the inventory level falls to or below $s$, an order is placed to fill the inventory. The lead times are also independent random variables following general distributions. Whenever the inventory becomes dry, the server goes
for a vacation for a random length of time which follows a general distribution. Using the theory of regenerative process, we derive explicit expressions for the inventory level probabilities at arbitrary epochs.