CHAPTER 1

Introduction

The purpose of this thesis is to combine several concepts from queuing theory and inventory and use them in modelling and analysis. Until 1947 it was assumed, while analyzing problems in queues with finite capacity, when the buffer is full any further arrival is lost. However this is not the case in reality. A customer who could not get admission into the system may keep trying until he succeeds or quits because a time reaches when he does not derive any benefit out of the service, whichever occurs first. This type of queueing problem was first analyzed by Kosten [27] in 1947 and such type of queues are referred to as retrial queues. Retrial queues arise in a natural way in communication systems, at enquiry counters attached with offices, in hospitals and so on. Multiserver retrial queues are complex compared to single server queue. Still more complex is the retrial multiserver queues where the servers are separated, which arises as follows. Suppose there are $c$ servers who are separated so that neither a server nor an arriving customer knows the status of the rest of the $c - 1$ servers. Thus if the present arrival to a particular server finds that server busy then he has to retry to access even other servers. This type of situation arises in, for example, at reception counters where there are a few telephones with distinct numbers. This problem is analyzed in Mushkov, Jacob, Ramakrishnan, Krishnamoorthy and Dudin [50] in 2006.

Inventory system was formally investigated in the most simple situation by Harris in 1915 which was subsequently analysed independently by Wilson in 1918 and the famous Harris-Wilson EOQ formula was realized. Most of the initial work in inventory theory were on deterministic models. Realizing the importance of uncertainty of the demand process and of the lead time, probabilistic models started getting investigated. Nevertheless the basic assumption in all these was that the time required to serve the item(s) was negligible. So in case item is available at demand epoch it is instantly
served. Else a queue gets formed, provided backlog is permitted. Krishnamoorthy and Raju in a series of papers [39, 41], analyzed inventory with local purchase during stock out period, whenever a demand occurs, to earn customer good-will. However these were also restricted to the case of negligible service time. In practice a positive duration of service, deterministic or random, is needed to serve the item(s). Thus Berman, Kim and Shimshack in 1993, came up with the notion of inventory with positive service time. Since then there are several developments in the analysis of such inventory models.

In this thesis we combine models in classical/retrial queues with inventory involving positive service time. In some cases we introduce local purchase during stock out period, to improve the reliability of the system. This local purchase is assumed to be instantly done so that customers are not lost on account of lack of availability of the item. We also introduce disaster that removes all inventoried items instantly.

Next we provide a brief account of queues and inventory. In the sequel we also provide a brief account of the matrix geometric solution. Then we proceed to provide a brief review of the work that were done in the direction of the problems discussed in this thesis.

1.1. Classical and Retrial Queues

Lining up for some form of service is a common phenomenon, be it visible or invisible, by human beings or by inanimate objects. It is more organized or, sometimes, is made to be so in the modern world and therefore a systematic study of a line up or equivalently a queueing process is instinctively more rewarding academically. A classical queueing system can be described as customers arriving for service, waiting for service if service is not immediate and if having waited for service, leaving the system after being served.

A queue is formed when either there is positive service time or there are no sufficient servers for the arriving customers. Some examples of a queue are customers arriving at a bank and aeroplanes waiting for their turn to land in busy airports.

Queueing systems in which arriving customers find all servers and waiting positions (if any) occupied, may retry for service after a period of time. Such queues are called
retrial queues or queues with repeated attempts. One of the most obvious example is provided by a person who desires to make a phone call. If the line is busy, then he cannot queue up, but can try sometime later.

Retrial queues are a type of networking with reserving after blocking. The classical queueing models do not take into account the phenomenon of retrials and therefore cannot be applied in solving a number of practically important problems. Retrial queues have been introduced to solve this deficiency.

1.2. Inventory Systems

In all business firms the system must keep a minimum amount of inventory at the time of order placing of inventory for the smooth and efficient running of the firm. The importance of inventory management for the quality of service of today’s service systems is generally accepted and optimization of systems in order to maximize quality of service is therefore an important topic.

There are several factors affecting the inventory. They are demand, life time of items stored, damage due to external disaster, production rate, the time lag between order and supply, availability of space in the store etc. If all these parameters are known beforehand, then the inventory model is called deterministic inventory model. If some or all of these parameters are not known with certainty then we consider them as random variables with some probability distribution and the resulting inventory model is then called stochastic inventory model.

Efficient management of inventory system is done by finding out optimal values of the decision variables. The important decision variables in inventory system are maximum capacity of the inventory, reordering point and order quantity. Several policies may be used to control an inventory system. Of these, the most important policy is the \((s, S)\) policy. An inventory system may be based on periodic review (e.g., ordering every week or every month), in which new orders are placed at the start of each period. Alternatively the system may be based on continuous review where a new order is placed when the inventory level drops to a certain level, called the reorder point. An example of periodic review occurs in gas stations where new deliveries arrive at the start
of each week. Continuous review occurs in retail stores where items (such as cosmetics) are replenished only when their level on the shelf drops to the reorder point.

The time elapsed between an order and its physical materialization is termed as lead time. If the replenishment is instantaneous then lead time is zero, otherwise the system is said to have positive lead time.

Inventory models have a wide range of applications in the decision making of government military organization, industries, hospitals, banks, educational institutions etc. Study and research in this fast growing field of applied mathematics, taking models from practical situations, contributes significantly to the progress and development of human society.

In most of the analysis of inventory systems the decay and disaster factors are ignored. But in several practical situations these factors play an important role in decision making. Examples are electronic equipments stored and exhibited on a sales counter, perishable goods like food stuffs, chemicals, crops vulnerable to insects and natural calamities like earth quake, rains, storms etc.

1.2.1. Inventory with positive service time. In all works reported in inventory prior to 1993 it was assumed that the time required to serve the item to the customer is negligible. As a consequence if the item is available at a demand epoch, the customer need not have to wait; a queue can be formed only when the inventory level becomes zero and lead time is positive.

We come across several real life situations where the service time is not negligible. In this case a queue will be formed even when the item is available. Thus the problem in inventory with service time may appear as a problem in queue. Nevertheless, this is not the case. The server stays idle even when there are customers in the system in the absence of inventoried items for processing.

Shortages of inventory occur in systems with positive lead time, in systems with negative reordering points or in multi commodity inventory system in which an order is placed only when the inventory level of at least two commodities fall to or below than the reorder level. Shortage cost is the penalty incurred when we run out of stock. It includes potential loss of income and moreover subjective cost of loss in customer’s
goodwill. There are different methods to tackle the stock out periods of the inventory. One of the method is to consider the demands during dry periods as ‘lost sales’. The other is partial or full backlogging of the demands. Lost sale causes a loss in the profit and back logging results in the increase in the waiting time of the customer. In order to avoid these two possibilities in this thesis we adopt the notion of local purchase. If a customer enters for service when the inventory level is zero we make a local purchase of the item at a higher cost. Thus we can decrease the waiting time of the customer and thereby holding cost of the customer. Local purchases are made to improve the goodwill of the customers with the system especially in a newly opened shop or where there is a competition between near by shops.

1.2.2. Quasi-Birth and Death process (QBD). Consider a continuous time Markov chain on the two-dimensional state space \( \{(0, j), 1 \leq j \leq m'\} \cup \{(n, j), n \geq 1, 1 \leq j \leq m\} \). The first co-ordinate \( n \) is called the level and the second co-ordinate \( j \) is called the phase of the state \( (n, j) \). The Markov process is called a QBD if one-step transition from a state are restricted to states in the same level or in the two adjacent levels: it is possible to move in one step from \( (n, j) \) to \( (n', j') \) only if \( n' = n, n + 1 \) or \( n - 1 \) (in the last case \( n \geq 1 \)). If the transition rate from \( (n, j) \) to \( (n', j') \) does not depend on \( n \) and \( n' \), but only on \( n' - n \) then the Markov process is called a Level Independent Quasi-Birth Death (LIQBD) process and the infinitesimal generator \( Q \) is given by

\[
Q = \begin{bmatrix}
B_1 & B_0 & 0 & 0 & \cdots & \cdots \\
B_2 & A_1 & A_0 & 0 & \cdots & \cdots \\
0 & A_2 & A_1 & A_0 & \cdots & \cdots \\
0 & 0 & A_2 & A_1 & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots \\
\end{bmatrix}
\]

where \( B_1 \) is a square matrix of order \( m' \), \( B_0 \) is an \( m' \times m \), \( B_2 \) is an \( m \times m' \) and \( A_0, A_1 \) and \( A_2 \) are square matrices of order \( m \).

If the transition rates depend on the level then the Markov Process is called a Level Dependent Quasi Birth Death (LDQBD) Process and the infinitesimal generator \( Q \) is
then given by
\[
Q = \begin{bmatrix}
A_{10} & A_{00} & 0 & 0 & 0 & 0 & \cdots \\
A_{21} & A_{11} & A_{01} & 0 & 0 & 0 & \cdots \\
0 & A_{22} & A_{12} & A_{02} & 0 & 0 & \cdots \\
0 & 0 & A_{23} & A_{13} & A_{03} & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}.
\]

(1.2.1)

All models discussed in this thesis are either LIQBD or LDQBD

1.2.3. Matrix analytic method. A matrix analytic approach to stochastic models was introduced by Neuts [53] to provide an algorithmic analysis for $M|G|1$ and $GI|M|1$ type of queueing models. Matrix analytic methods constitute a success story, illustrating the enrichment of science, applied probability by a technology, that of digital computers.

The following theorem gives a brief description of Matrix Analytic Method applied for solving Quasi-Birth Death Process (QBD).

THEOREM 1.2.1. A continuous time QBD with infinitesimal generator $Q$ of the form (1.2.1) is positive recurrent if and only if the minimal non-negative solution $R$ to the matrix quadratic equation

\[
R^2A_2 + RA_1 + A_0 = 0
\]

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

\[
x_0A_{10} + x_1A_{21} = 0 \\
x_{i-1}A_{0,i-1} + x_iA_{1i} + x_{i+1}A_{2,i+1} = 0 \quad (1 \leq i \leq N - 2) \\
x_{N-2}A_{0,N-2} + x_{N-1}(A_{1,N-1} + RA_2) = 0
\]

has a unique solution for $x_0, \ldots, x_{N-1}$. If the matrix $A = A_0 + A_1 + A_2$ where $A_{0i} = A_0$, $A_{1i} = A_1$ for $i \geq N$ is irreducible, then $sp(R) < 1$ if and only if $\pi A_0e < \pi A_2e$ where $\pi$ is the stationary probability vector of the generator matrix $A$ and satisfies the equation $\pi A = 0$ and $\pi e = 1$ where $e = (1, \ldots, 1)'$. 
If \( x = (x_0, x_1, \ldots) \) is the stationary probability vector of \( Q \) then \( x_i \)'s \( (i \geq N) \) are given by

\[
x_{N+r-1} = x_{N-1} R^r \text{ for } r \geq 1.
\]

To find the minimal solution of (1.2.2) we can use the iterative formula given by

\[
R_{n+1} = -(R_n^2 A_2 + A_0) A_1^{-1}, \quad n = 0, 1, 2, \ldots \text{ with } R_0 = 0
\]

1.3. Review of Related Work

1.3.1. Works on inventory. In 1915 Harris [24] started the mathematical modelling of inventory problems and derived the famous EOQ formula that was popularized by Wilson. A systematic analysis of the \((s, S)\) inventory system using renewal theoretic arguments is provided in Arrow, Karlin and Scarf [2]. Hadley and Whitin [23] gave several applications of different inventory models. Gross and Harris [21] considered the inventory systems with state dependent lead times. Sivazlian [63] analyzed the continuous review \((s, S)\) inventory system with general inter arrival times and unit demand in which he shows that the limiting distribution of the position inventory is uniform and independent of the inter arrival time distribution. Sahin [60] analyzed continuous review \((s, S)\) inventory with continuous state space and constant lead time. Srinivasan [64] discussed an \((s, S)\) inventory problem with arbitrarily distributed interarrival times and lead times.

Manoharan et.al. [47] discussed the case of non-identically distributed interarrival times. Krishnamoorthy and Lakshmi [35] analyzed problems with Markov dependent re-ordering levels and Markov dependent replenishment quantities. Krishnamoorthy and Manoharan [46] modelled an inventory system with varying reorder levels and random lead time. Krishnamoorthy and Varghese [44] considered a two commodity inventory problem with Markov shift in demand for the commodity. Krishnamoorthy and Raju [39] introduced \(N\)-policy to the \((s, S)\) inventory system with positive lead time and local purchase when the inventory level is zero.

Berman, Kim and Shimshack [13] introduced positive service time in inventory in which the service time is assumed to be constant. They determined optimal order quantity \( Q \) that minimizes the total cost rate using dynamic programing technique.

Viswanath et al. [66] studied an \((s, S)\) inventory policy with service time by considering vacation to server and correlated lead time. They considered quite general distribution for interarrival time, duration of service time and duration of a vacation. Server goes on vacation whenever there is either no customer left behind in the system at departure epoch or when the inventory level drops to zero or both occur simultaneously. Schwarz et al. [61] discussed \(M|M|1\) queueing systems with inventory where the lead times are exponentially distributed. They analyzed the problem for both \((r, Q)\) and \((r, S)\) inventory policies and derived stationary distribution of joint queue length and inventory level in explicit product form. Also they discussed the problem of order placements any where on the set \(\{0, 1, \ldots, s\}\) according to a given probability distribution. Krishnamoorthy et al. [38] introduced the \(N\)-policy for commencement of service, once the server is switched off in the absence of customers in the system. Here the service time is positive and lead time is zero. They obtained analytical solution to this model. They establish a product form solution to the system state and thus produce a decomposition of the state space. Murthy and Ramanarayan [49] discussed \((s, S)\) inventory system with defective items in the replenished items, where the lead time is positive with arbitrary distribution. Krishnamoorthy and Varghese [43] analyzed an inventory model where the items are damaged due to decay and disaster. They assumed that the lead time is zero and the service time is negligible. A detailed survey on inventory with positive service time is given in Krishnamoorthy et al. [36].

1.3.2. Works on retrial queue and retrial inventory. Retrial queues or queues with repeated attempts have been extensively investigated (See the survey papers by Yang and Templeton [67], Falin [18] and the book by Falin and Templeton [19]). Subsequent development on retrial queues can be found in Artalejo [3]. The latest addition to books on retrial queues is authored by Artalejo and Gomez-Correl [6]. In this they discussed the algorithmic approach. Artalejo, Krishnamoorthy and Lopez-Herrero [9]
were the first to study inventory policies with positive lead time coupled with retrial of unsatisfied customers and their approach turns out to be algorithmic. Ushakumari [65] obtained analytical solution to the above problem in 2006. Krishnamoorthy and Mohammad Ekramol Islam [31] analyzed an \( (s, S) \) inventory system with retrial of customers. Here the lead time and inter-retrial times are assumed to be exponentially distributed.

Krishnamoorthy and Jose [33] compared three \( (s, S) \) inventory system with retrial of customers where the service time and lead time are positive. They investigated these systems to obtain performance measures and construct suitable cost functions for the three cases. In 2002 Artelajo et.al [8] discussed an \( M|G|1 \) retrial queue where the server goes for an orbital search, when he is free. Thus the system can decrease the idle time of the server as well as the waiting time of the customer. Neuts and Rao [55] discussed an \( M|M|c \) retrial queue in which the model is LDQBD process and they suggested a truncation procedure, the idea is to make retrial rate to be constant when the number of customers in the orbit exceeds some level.

1.4. An Outline of the Work in this Thesis

This thesis is divided into six chapters including this introductory chapter. Second chapter contains investigation of two models. In the first model we consider a single item, continuous review \( (s, S) \) inventory model with one server. Arrival of customers form a Poisson process with rate \( \lambda \) and service times of customers are exponentially distributed random variables with parameter \( \mu \), one unit of item is needed for each customer. Lead time is assumed to be zero. An arriving customer, who finds the server busy, proceed to an orbit of infinite capacity and makes successive repeated attempts until it finds the server free. The inter retrial times have an exponential distribution with parameter \( i\theta \) when there are \( i \) customers in the orbit. Here we get an analytical solution to the model. We construct a cost function and numerical examples are given. In the second model we consider a more general set up involving arbitrarily distributed service time. All other assumptions are same as that in the first model. We consider the number
of customers in the orbit and the inventory level at the departure epoch of a customer. Thus we have an embedded Markov chain. Here also we analyze a cost function.

In chapter 3, we consider five distinct inventory models with positive service time and positive lead time. In all these it is assumed that customers arrive to a single server system according to a Poisson process with rate $\lambda$ and service times are exponentially distributed random variables with parameter $\mu$. Each customer requires one unit of inventory. We follow an $(s, S)$ inventory policy. When the inventory level depletes to $s$ we place an order for $Q = S - s$ quantity of inventory. The distribution of lead time is exponential with parameter $\beta$. In model 1 customers do not join the system when the inventory level is zero. In model 2 customers join the system even when the inventory level is zero. In model 3 and 4 we make a local purchase of one and $s$ units of items respectively, whenever a customer arrives to find the inventory level zero, at an extra cost. In model 5 under the same situation we make a local purchase of $S$ units, thus cancelling the existing order for procurement of inventory as the maximum capacity of inventory is $S$. Numerical examples are given to compare performance of these models in terms of appropriate cost functions.

In chapter 4 we introduce retrial of unsatisfied customers into the models discussed in chapter 3, with the assumption that there is no waiting space for the customers at the service station other than to the one who is being served. An arriving customer who finds the server busy, proceeds to an orbit of infinite capacity and makes successive repeated attempts until it finds the server free. The inter retrial times can be modelled according to different disciplines depending on each particular application. In telephone systems the repeated attempts are made individually by each blocked customer following an exponential law of rate $\theta$. This is the classical retrial policy where the rate is $i\theta$ when there are $i \geq 0$ customers in the orbit. Another retrial policy is the constant retrial policy in which the probability of repeated attempts is independent of the number of customers in the orbit. Here we assume that the inter retrial times have an exponential distribution with constant rate $\theta$. Here also we compare the cost functions through numerical investigations.
In chapter 5 we consider \((s, S)\) inventory systems with the possibility of destruction of inventoried items due to disasters. Here we discuss two models. Customers arrive to a single server system according to a Poisson process with parameter \(\lambda\) where service times are exponentially distributed random variables with parameter \(\mu\). We assume that disaster destroys all the inventoried items but not the customers. For example, in godowns food items are destroyed by natural calamities. Here we assume the inter disaster times to be exponentially distributed with parameter \(\delta\). It is assumed that lead time is also exponentially distributed random variables with parameter \(\beta\). In Model I we assume further that customers do not join the system when the inventory level is zero. However in Model II it is assumed that customers join even when the inventory level is zero. Thus stability in Model II is affected by the lead time parameter. We compare the two models through numerical examples by constructing suitable cost functions.

In chapter 6 we consider a multi server queue coupled with an inventory following \((s, S)\) policy and retrial of customers. Customers arrive to the system with \(c\)-servers according to Poisson process with rate \(\lambda\). The service times are exponentially distributed with parameter \(\mu\). One item is needed for each customer. An arriving primary customer, who finds all servers busy, will go to an orbit of infinite capacity and tries again for the service. Inter retrial time follows exponential distribution with parameter \(\theta\). The lead time follows exponential distribution with rate \(\beta\). We assume that customers do not join the system when the inventory level is zero. A cost function is constructed and numerically investigated.