Chapter 1

Introduction

Most of us experience queueing systems directly or indirectly; directly through waiting in line ourselves or indirectly through some of our items waiting in line, such as a print job waiting in the printer buffer queue, or a packet waiting at a router node for processing. In all the cases, we want the delay to be minimum and also not to be turned away by the system due to the buffer space being not available. These possibilities of delay and denial are the major issues in a queuing system and how to minimize them is the concern. A trade off between the cost to the system due to customers denied admission as a consequence of overflow and profit due to large number of customers in the system is what is needed. Stochastic modelling of the system along with construction of a suitable cost function provides answers to most of the questions. Thus the performance of a queuing system can be evaluated and information can be generated for making decisions as to when and how to upgrade the system to improve its future performance.

In queueing systems, and all systems that operate over time with uncertainty being model characteristic, we need a sequence or a family of random variables to represent such a phenomenon over time. A stochastic process is a family or a sequence of random variables indexed by a parameter, usually time. A continuous time Markov chain is a continuous time stochastic process that enjoy
Chapter 1: Introduction

memoryless property which means that no matter what the past was, the current state is all that is needed to predict the future. This memoryless property allows flexibility in modelling and produces tractable models. This thesis analyzes a few queueing models by means of continuous time Markov chains. Modelling tools such as Markovian Arrival Process (MAP) and Phase-type service distributions (PH-distributions) are used in this regard.

Queueing systems with phase-type arrival or service mechanisms give rise to transition matrices that are block tridiagonal and are referred to as quasi-birth-death(QBD) processes. The matrix geometric method developed by Neuts is employed in solving such queueing models. In this thesis we have extensively made use of Phase type distributions for service time and Markovian arrival process for arrival of customers. Hence it is apt to give a brief description of these.

1.1 Phase Type distribution
(Continuous time)

Phase type(PH) distributions and related point processes provide a versatile set of tractable models in applied probability. They are based on the method of stages, a technique introduced by A. K. Erlang and generalized by M. F. Neuts. The key idea is to model random time intervals as being made up of a (possibly random) number of exponentially distributed segments and to exploit the resulting Markovian structure without losing computational tractability.

The continuous PH distributions are introduced as a natural generalization of the exponential and Erlang distributions. A PH—distribution is obtained as the distribution of the time until absorption in a finite state space Markov chain with an absorbing state. Phase-type distributions have matrix representations that are not unique. Furthermore, phase-type distributions constitute a versatile class of distributions that can approximate arbitrarily closely any probability
Phase Type distribution

distribution defined on the nonnegative real line.

A non-negative random variable $X$ has a Phase-Type ($PH$) distribution if its distribution function is given by

$$F(t) = P(X \leq t) = 1 - \alpha \exp(Tt)e \equiv 1 - \alpha \left( \sum_{r=0}^{\infty} \frac{t^r T^r}{r!} \right) e, \quad t \geq 0$$

where,

- $\alpha$ is row vector of non-negative elements of order $m(>0)$ satisfying $\alpha e \leq 1$.
- $T$ is an $m \times m$ matrix such that i) all off-diagonal elements are nonnegative
  ii) all main diagonal elements are negative
  iii) all row sums are non-positive
  iv) $T$ is invertible

The 2- tuple $(\alpha, T)$ is called a phase-type representation of order $m$ for the PH distribution and $T$ is called a generator of the PH distribution.

Let $X = \{X(t) : t \geq 0\}$ be a homogeneous Markov chain with finite state space $\{1, \ldots, m, m+1\}$ and generator

$$Q = \begin{pmatrix} T_{m \times m} & T^0 \\ 0 & 0 \end{pmatrix}$$

where the elements of the matrices $T$ and $T^0$ satisfy $T_{ii} < 0$ for $1 \leq i \leq m$, $T_{ij} \geq 0$ for $i \neq j$; $T_i^0 \geq 0$ and $T_i^0 > 0$ for at least one $i$, $1 \leq i \leq m$ and $Te + T^0 = 0$.

Let the initial distribution of $X$ be the row vector $(\alpha, \alpha_{m+1})$, $\alpha$ being a row vector of dimension $m$ with the property that $\alpha e + \alpha_{m+1} = 1$. The states $1, 2, \ldots, m$ shall be transient, while the state $m+1$ is absorbing.

Let $Z = \inf\{t \geq 0 : X(t) = m+1\}$ be the random variable representing the time until absorption in state $m+1$. Then the distribution of $Z$ is Phase type distribution (or shortly PH distribution) with representation $(\alpha, T)$. The dimension $m$ of $T$ is called the order of the distribution. The states $1, 2, \ldots, m$ are also called phases.
• The density function is
\[
f(t) = \alpha \exp(T \cdot t) \cdot T^0 \quad \text{for every } t > 0
\]

• \(E[X^n] = (-1)^n n! \alpha T^{-n} e, \ n \geq 1\).

• The Laplace-Stieltjes transform of \(F(.)\) is
\[
\phi(s) = \alpha_{m+1} + \alpha (sI - T)^{-1} \cdot T^0 \quad \text{for } Re(s) \geq 0.
\]

**Theorem 1.1.1** (see, Latouche and Ramaswami [43]). Consider a PH distribution \((\alpha, T)\). Absorption into state \(m + 1\) occurs with probability 1 from any phase \(i\) in \(\{1, 2, \ldots, m\}\) if and only if the matrix \(T\) is nonsingular.

More over, \((-T^{-1})_{i,j}\) is the expected total time spent in phase \(j\) during the time until absorption, given that the initial phase is \(i\).

For further information about the PH distribution, see, Neuts, [52], Breuer and Baum, [9], Latouche and Ramaswami, [44], and Qi-Ming He, [55]. Usefulness of PH distribution as service time distribution in telecommunication networks is elaborated, e.g., in Pattavina and Parini [53] and Riska, Diev and Smirni [54].

### 1.2 Markovian Arrival Process

Markovian Arrival Processes (MAP) are introduced in Neuts [50]. It is a rich class of point processes that includes many well-known processes such as Poisson, PH-renewal processes and Markov-modulated Poisson process. A salient feature of the MAP is the underlying Markovian structure that fits ideally in the context of matrix-analytic solutions to stochastic models. MAP significantly generalizes the Poisson processes and still keep the tractability for modelling purposes. Currently, the MAP is the most popular mathematical model for the telecommunication networks traffic because it catches the typical features of this traffic such
Markovian Arrival Process

as correlation and burstiness. Furthermore, in many practical applications, notably in communication engineering, production and manufacturing engineering, the arrivals do not usually form a renewal process. So, MAP is a convenient tool to model both renewal and non-renewal arrivals. In [10], Chakravarthy provides an extensive survey of the Batch Markovian Arrival Process (BMAP) in which arrivals are in batches where as it is in singles in MAP.

A continuous time Markovian arrival process is a counting process that is defined on a finite state continuous time Markov chain. However, unlike PH-distributions an underlying Markov chain for a Markovian arrival process has no absorption state (phase). A Markovian arrival process counts the number of arrivals, which can be associated with changes of state in the underlying Markov chain. The arrivals can also occur during the stay in each state of the underlying Markov chain. For a MAP, the transitions of state with arrival, transitions of state without arrival, and arrivals without a transition of state, are all referred to as events. Arrival rates of events can be customized for different states, demonstrating the versatility inherent to MAPs.

In a MAP, the customers arrival is directed by an irreducible continuous time Markov chain \( \{\phi_t, t \geq 0\} \) with the state space \( \{1, 2, \ldots, m\} \). Let \( D \) be the generator of this Markov chain. At the end of a sojourn time in state \( i \), that is exponentially distributed with parameter \( \lambda_i \), one of the following two events could occur: with probability \( p_{ij}(1) \) the transition corresponds to an arrival and the underlying Markov chain is in state \( j \) with \( 1 \leq i, j \leq m \); with probability \( p_{ij}(0) \) the transition corresponds to no arrival and the state of the Markov chain is \( j, j \neq i \). The Markov chain can go from state \( i \) to state \( i \) only through an arrival. Also we have

\[
\sum_{j=1}^{m} p_{ij}(1) + \sum_{j=1, j \neq i}^{m} p_{ij}(0) = 1, \quad 1 \leq i \leq m.
\]

The transition intensities of the Markov chain \( \{\phi_t, t \geq 0\} \) which are accompanied by arrival of \( k \) customers are described by the matrices \( D_k, k = 0, 1 \). Define \( D_0 = \)

\[
\]
(d^{(0)}_{ij}) and D_1 = (d^{(1)}_{ij}) such that d^{(0)}_{ii} = -\lambda_i, 1 \leq i \leq m, d^{(0)}_{ij} = \lambda_i p_{ij}(0), for j \neq i and d^{(1)}_{ij} = \lambda_i p_{ij}(1), 1 \leq i, j \leq m.

By assuming D_0 to be a nonsingular matrix, the inter-arrival time is finite with probability one and the arrival process does not terminate. Hence, we see that D_0 is a stable matrix. The generator D is then given by D = D_0 + D_1. Thus D_0 governs the transitions corresponding to no arrival and D_1 governs those corresponding to an arrival. Vector $\eta$ of the stationary distribution of the process $\{\phi_t, t \geq 0\}$ is the unique solution to the system

$$\eta(D_0 + D_1) = \eta D = 0 \text{ and } \eta e = 1.$$  \hfill (1.1)

Fundamental rate $\lambda$ of the MAP is given by $\lambda = \eta D_1 e$ which gives the expected number of arrivals per unit time in the stationary version of the MAP.

1.3 Quasi-birth-death processes

Quasi-birth-death processes (QBDs) are matrix generalizations of simple birth-and-death processes on the nonnegative integers in the same way as PH distributions are matrix generalization of the exponential distribution. Consider a Markov Chain $\{X_t, t \in \mathbb{R}^+\}$ on the two dimensional state space

$\Omega = \bigcup_{n \geq 0} \{(n, j) : 1 \leq j \leq m\}$. The first coordinate $n$ is called the level, and the second coordinate $j$ is called a phase of the $n^{th}$ level. The number of phases in each level may be either finite or infinite. The Markov chain is called a QBD process if one-step transitions from a state are restricted to phases in the same level or to the two adjacent levels. In other words,

$$(n - 1, j') \Rightarrow (n, j) \Rightarrow (n + 1, j'') \text{ for } n \geq 1.$$ 

If the transition rates are level independent, the resulting QBD process is called level independent quasi-birth-death process (LIQBD); else it is called level dependent quasi-birth-death process (LDQBD). Arranging the elements of $\Omega$ in
lexicographic order, the infinitesimal generator of a LIQBD process is block tridiagonal and has the following form:

\[
Q = \begin{pmatrix}
B_1 & A_0 & & \\
B_2 & A_1 & A_0 & \\
 & A_2 & A_1 & A_0 \\
& & \ddots & \ddots & \ddots
\end{pmatrix}
\] (1.2)

where the sub matrices \(A_0, A_1, A_2\) are square and have the same dimension; matrix \(B_1\) is also square and need not have the same size as \(A_1\). Also, the matrices \(B_2, A_2\) and \(A_0\) are nonnegative and the matrices \(B_1\) and \(A_1\) have nonnegative off-diagonal elements and strictly negative diagonals. The row sums of \(Q\) are equal to zero, so that we have \(B_1 \mathbf{e} + A_0 \mathbf{e} = B_2 \mathbf{e} + A_1 \mathbf{e} + A_0 \mathbf{e} = (A_0 + A_1 + A_2) \mathbf{e} = \mathbf{0}\).

Among the several tools that we employed in this thesis Matrix geometric method plays a key role. A brief description of this is given below.

## 1.4 Matrix Geometric Method

Matrix Geometric Method introduced by M. F. Neuts is a tool to construct and analyze a wide class of stochastic models, particularly queueing systems, using a matrix formalism to develop algorithmically tractable solution. The transform techniques employed in solving QBD processes are replaced largely by the matrix geometric approach with the advent of high speed computers and efficient algorithms. In the matrix geometric method the distribution of a random variable is defined through a matrix; its density function, moments, etc. are expressed with this matrix. The modelling tools such as Phase type distributions, Markovian Arrival Processes, Batch Markovian Arrival Processes, Markovian Service Processes etc. are well suited for Matrix Geometric Methods.

**Theorem 1.4.1** (see Theorem 3.1.1. of Neuts [52]). The process \(Q\) in (1.2) is positive recurrent if and only if the minimal non-negative solution \(R\) to the
Chapter 1: Introduction

matrix-quadratic equation

$$R^2 A_2 + RA_1 + A_0 = O$$  \hspace{1cm} (1.3)

has all its eigenvalues inside the unit disk and the finite system of equations

$$x_0 \ (B_1 + RB_2) = 0$$
$$x_0 \ (I - R)^{-1} e = 1$$  \hspace{1cm} (1.4)

has a unique positive solution $x_0$.

If the matrix $A = A_0 + A_1 + A_2$ is irreducible, then $sp(R) < 1$ if and only if

$$\pi A_0 e < \pi A_2 e$$  \hspace{1cm} (1.5)

where $\pi$ is the stationary probability vector of $A$.

The stationary probability vector $x = (x_0, x_1, \ldots)$ of $Q$ is given by

$$x_i = x_0 R^i \text{ for } i \geq 1.$$  \hspace{1cm} (1.6)

Once $R$, the rate matrix, is obtained, the vector $x$ can be computed. We can use an iterative procedure or logarithmic reduction algorithm (see Latouche and Ramaswami [43]) or the cyclic reduction algorithm (see Bini and Meini [4]) for computing $R$.

1.5 Computation of $R$ matrix

There are several algorithms for computing rate matrix $R$. Here we list two of them.

Iterative algorithm

From (1.3), we can evaluate $R$ in a recursive procedure as follows.
Matrix Geometric Method

Step 0: $R(0) = O$.

Step 1: 

$$R(n + 1) = A_0(-A_1)^{-1} + R^2(n)A_2(-A_1)^{-1}, \quad n = 0, 1, \ldots$$

Continue Step 1 until $R(n + 1)$ is close to $R(n)$.

That is, $||R(n + 1) - R(n)||_\infty < \epsilon$.

Logarithmic reduction algorithm

Logarithmic reduction algorithm is developed by Latouche and Ramaswami [43] which has extremely fast quadratic convergence. This algorithm is considered to be the most efficient one. The main steps involved in the logarithmic reduction algorithm are listed below. For further details on the logarithmic reduction algorithm refer Latouche and Ramaswami [43].

Step 0: $H \leftarrow (-A_1)^{-1}A_0$, $L \leftarrow (-A_1)^{-1}A_2$, $G = L$, and $T = H$.

Step 1:

$$U = HL + LH$$

$$M = H^2$$

$$H \leftarrow (I - U)^{-1}M$$

$$M \leftarrow L^2$$

$$L \leftarrow (I - U)^{-1}M$$

$$G \leftarrow G + TL$$

$$T \leftarrow TH$$
Continue **Step 1**: until $\|e - Ge\|_\infty < \epsilon$.

**Step 2:** $R = -A_0(A_1 + A_0G)^{-1}$.

### 1.6 Supplementary variable technique

In most practical queueing systems inclusion of one or more supplementary variable could make the system Markovian. The use of the supplementary variable technique in queueing dates back to 1942 when it was introduced by Kosten [36]. In this method to get a Markov Process, we keep track of some additional information together with the underlying random variable. Consider an $M/G/1$ queue, where $G$ denotes the distribution of service time. The process $\{N_t, t \in R^+\}$, where $N_t$ gives the state of the system or the system size at an arbitrary time $t$ is then non-Markovian. This process is not Markov. However such a process could be made Markovian by the inclusion of variable $x_t$ defined as the amount of time spent/remaining for the customer in service at time $t$, if any. In other words the collection $\{(N_t, x_t) : t \geq 0, x_t \geq 0\}$ is a Markov process. For the $GI/M/1$ the supplementary information on time elapsed until $t$ since last arrival, in addition to the number of customers at the pre-arrival epoch prior to time $t$, provides a two dimensional Markov chain. For details of the supplementary variable techniques applied to $M/G/1$ queue see Cox [19], Keilson and Kooharain [34] and Cohen [18].

This thesis provides analysis of priority queues. These priorities need not be on the basis of source of external (primary) customers. We have deviated from the classical priority queue by bringing in ‘internal’ priority generation. This root of internal priority generation is considered in Krishnamoorthy et al.,[38] and Gomez-Corral et al.[21]. Krishnamoorthy et al. [38] also analyze multi priority queues with ‘internal priority generation’. Very often internal priority generation may be t higher priorities (like patients waiting in a queue to consult
a physician). In contrast the internal priority generation discussed in this thesis takes the customer to lower priority queue; such queues get generated internally. An example of such situation is a customer interrupting his service to attend a phone call.

1.7 Review of related work

In queueing literature, priority queues stand for customers belonging to different classes joining distinct waiting lines (one for each class) to receive service. The highest classes of customers have priority (preemptive or non-preemptive) over the rest; the next in the order gets priority over all lower class customers and so on.

Priority queues are first considered by White and Christie [64] as a queue with interruption of service of low priority customers to provide service to higher priority customers. A priority queue with preemptive service can be regarded as a queue with service interruption for e.g. a doctor renders his service to a causality patient urgently by interrupting his other consultation. Jaiswal [27] is on preemptive priority queue with resumption of service of the low priority customer and Jaiswal [29] discusses time dependent solution in priority queues. Cobham [17] considers a non-preemptive priority queue and derived equilibrium expected waiting time. A detailed discussion of development in priority queues until 1968 is given in Jaiswal [30]. More recent developments on priority queues could be found in Takagi [61] and in Brodal [8].

Concept of interruption in service is introduced in the context of the failure of service system (see the recent survey paper by A. Krishnamoorthy et al. [39]). Customer induced service interruption as coined by Jacob et al. [26] is a contrast situation to that of interruption due to server failure. This is done for the single server case, where service interrupted customers are given priority over primary customers. Here self-interrupted customer takes an exponentially distributed
Chapter 1: Introduction

time to get out of interruption. This is extended to the multi-server case in Krishnamoorthy and Jacob [37]. All underlying distributions (inter-arrival time, service time, inter-interruption time, interruption fixation time) are assumed to be independent exponential random variables. Dudin et al. [20] extend the above case to Markovian Arrival Process and Phase type service with $c$ servers and negative customers with a few protected service phases.

The priority queuing system considered in the second chapter differs from those discussed above as follows: We assume that the interrupted customers are allotted low priority. As an e.g. a person who applies for credit card with bad credit history. Also in the previous models discussed, the interrupted customers are entered in a buffer space of finite capacity where as in this model the interrupted customers join a waiting line with an infinite capacity. The interruption for a customer can occur a finite number of times, say $N$, resulting in $N + 1$ queues. Each waiting line is generated by the customers in the immediately preceding queue, except the highest priority customers who form the primary queue (external source). Thus the low priority queues are dependent even in its evolution. Both preemptive and non-preemptive service disciplines are analyzed.

Compared to the second chapter, the third chapter analyzes a priority queuing model where low priority lines consists of customers who come back for service getting repeated. Thus the third chapter focuses on priority queues with feedback customers. Various feedback policies on different queuing models are studied in detail in literature. Some of the works are reported in [7], [13], [14], [15], [32], [59] and [60]. Krishnakumar et al. [40] consider M/G/1 retrial queue with feedback, the feedback customer goes to the tail end of the queue. Krishnakumar et al. [41] analyze a multiserver feedback system in which also feedback is to the tail end of the queue. A single server retrial queue with collisions and feedback is analyzed in Krishnakumar et al. [42]. The feedback considered in literature fall mainly in two categories. Either the customer joins the tail end of queue on completion of service to get his service repeated or he occupies the server immediately on
completion of service without joining the queue. In the latter case, service at the head of the queue is paused for a while to provide service to the immediately fed back customer. In both cases there is no separate queue for feedback customers and there is no way of identifying a feedback customer in the first case.

In the third chapter, we introduce feedback queue in a different setting. Even though the customer feedback is instantaneous, it is assigned a lower priority in our system, added to that we assume there is external entry also. For instance, a company providing annual maintenance contract with certain number of free services. Here the waiting lines are not as dependent as discussed in the second chapter. Yet, if we block the external entry to the low priority lines, then the queues will be formed only by the feedback customers (so that analysis of feedback customers will be made easy). We restrict our attention to the case of a single feedback.

From here the work proceeds to a different priority queueing model, which contains a virtual queue of infinite capacity and a finite queue of physically arriving customers. For example, a store may have two types deliveries-one direct and other over phone. Crowdsourcing happens when the store decides to serve indirect customers through willing direct customers, the store being main server and willing customers being servers for the store. Crowdsourcing coined from ‘crowd’ and ‘outsourcing’ according to Howe [25] is the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. For a discussion on the crowdsourcing queueing system one may refer to Chakravarthy and Dudin [11]. They discuss the problem as a priority queue with non-preemption. Motivated from this we analyze a preemptive priority crowdsourcing model.

In all the three models discussed above, it is assumed that the server is completely aware of the service requirement of a customer (see Gross and Harris [22], though there is no mention about exact requirement). In fact this is the case with all models discussed so far in queueing theory. Quite often only one type of ser-
Chapter 1: Introduction

service is offered by the system and so conflict does not occur. It is also true that the customers arriving to such system know the type of service needed. Thus there is no conflict on the service provided to the customer. However, there are several real life situations where the customer (or server or even both) is not knowledgeable about the exact service requirement. This is especially the case when several types of services are available at a service station. As a concrete example we have vehicles for repair at service stations, patients consulting physicians for diagnosis and medication and a specific call for a service at a customer care center. If the right service required is not identified and instead the diagnosis turned out to be wrong the result could be disastrous. A wrong diagnosis and consequent service provided may sometimes turn out to be even fatal or may result in the equipment being rendered unusable or the loss of a customer altogether. These types of diagnostic problems are analyzed in the last two chapters of this thesis.

This thesis analyzes models providing explicit solution for system state distribution and also those that need algorithmic analysis. The matrix-geometric structure of the steady-state distributions introduced by Neuts and an extended version by Miller [49] for doubly infinite queues are used in the models for obtaining solutions.

1.8 Summary of the thesis

This thesis is basically analysis of some priority queues and a problem on diagnostics which we face in many real life situations. In this thesis a few queueing models are studied by means of continuous time Markov chains. The modelling tools such as Markovian Arrival Process (MAP) and Phase type distributions (PH-distributions) are used. We analyze the resulting systems as quasi-birth-death processes, mainly using matrix geometric method.

Now we turn to the content of the thesis. The thesis entitled “Analysis of some priority queues and a problem on diagnostics” is divided into 6 chapters
Summary of the thesis

including the introductory chapter. The chapters 2 and 3 discuss doubly infinite queues, chapter 4 discuss a crowdsourcing model and chapters 5 and 6 are on diagnostic problems. All models discussed in this thesis involve interruption in service in some form or other.

In chapter 2 we consider a priority queueing system where low priority customers are generated by self-interruption while at service. Customers arrive to a single service station from a Poisson stream and form a queue ($P_1$ line) of infinite capacity, if the server is found busy. They are served one at a time according to FIFO discipline. Customers may have a tendency to interrupt their own service while availing the same due to various reasons. Self interrupted customers are pushed to an infinite capacity low priority ($P_2$) queue. If the customer at $P_2$ line interrupts his service again, he is sent to a further lower priority queue ($P_3$ -line) and this may go on a finite number (say $N$) of times. When at a service completion epoch of a $P_i$ customer, if there is none left behind in $P_1$ line, then the server goes to serve customers in $P_{i+1}$ line. The service time for each category is assumed to follow exponential distribution, but at different service rates. The interruptions that happen are also according to exponential distributions with different parameters. We consider both preemptive and non-preemptive service discipline. We analyze a two priority system in detail where we assume that $P_2$ customers are not allowed to interrupt their service. The joint system state distribution is obtained from which the marginals are computed. Waiting time distribution of both type of customers are derived. We extend the results to three priority non-preemptive case and the case of $N + 1$ priorities is briefly discussed.

Chapter 3 is a modification of the hitherto notion of feedback in queueing theory (see page no.3). Here we analyze a two priority queueing system where high priority ($P_1$) customers may feedback according to a Bernoulli process if they are not satisfied with the service provided. but they will have to join the tail end of the low priority($P_2$) line. Arrival of both type of customers are according to independent Poisson processes. Both waiting rooms have infinite capacity. Cus-
customers are served one at a time according to FIFO discipline on priority basis: those in \( P_1 \) are given priority over the ones in the waiting line \( P_2 \). The service time is class dependent phase type. \( P_2 \) line customers will be serviced only when, if there is none left behind in \( P_1 \) line, at the service completion epoch of a high priority customer. Being a two priority system we assume that \( P_2 \) customers are not allowed an additional feedback. Thus the system consists of a primary waiting line and a second waiting line which is generated from the first as well as by customers from outside. We consider both preemptive and non-preemptive service discipline. The joint steady state probability distribution is derived and the corresponding marginal probabilities are computed. The distribution of waiting time of each type of customers is derived. We also point out a situation where there is no external entry to the \( P_2 \) line which makes the \( P_2 \) line exclusively for feedback customers. Even this special case does not boil down to the main problem discussed in chapter 2, since there, the self-interruption is during service.

Going on, we analyze a crowdsourcing queueing model in chapter 4. We consider a \( c \)-server queueing system providing service to two types of customers, \( P_1 \) and \( P_2 \). Customers arrive according to two distinct Poisson processes. A \( P_1 \) customer has to receive service by one of the \( c \) servers while a Type 2 customer may be served by a \( P_1 \) customer who is available to act as a server soon after getting own service or by one of \( c \) servers. A \( P_1 \) customer will be available for serving a \( P_2 \) customer with certain probability provided there is at least one \( P_2 \) customer waiting in the queue at the time of the service completion of that \( P_1 \) customer. With complementary probability, a \( P_1 \) customer will opt out of serving a \( P_2 \) customer, if any, waiting in the system. A free server offers service to a \( P_1 \) customer on a FCFS basis. However, if there is no \( P_1 \) customer waiting in the system, that server will serve a \( P_2 \) customer if one of that type is present in the queue. The service time is exponentially distributed for each category. \( P_1 \) customers have priority over those of \( P_2 \). We consider preemptive service discipline. Condition for system stability is established. Important system
characteristics including the average number of busy servers, the loss probability and the expected waiting time of each type in the system are computed. Some examples are numerically illustrated. Finally the characteristics of this model are compared with that of Chakravarthy and Dudin [11].

Now we turn to the diagnostic problem. In real life, there are several service providing systems offering a multitude of service. Neither the server nor the customer may be fully aware of the exact service requirement. Very often this results in irreparable damage to the customer being served. It is this type of problems that we analyze in chapters 5 and 6.

Chapter 5 discusses a queueing model with a single server offering many services to which arrival is according to a $MAP$ forming a single line. The time taken for completing service is phase type distributed. A service could be appropriate or inappropriate for each customer. If the service starts in an inappropriate state with a positive probability, we assume a clock to start ticking simultaneously. In case the service time exceeds the realization of the clock, then that customer is compelled to leave the system forever without being eligible for the service that he actually requires. Otherwise the customer gets the required service and then leaves the system. Several system performance measures including the rate of loss of customers, rate of customers leaving with correct service, even if started in incorrect service, are computed. Numerical illustrations of the system behavior are also provided. Then this is compared with that of Madan [46] and Medhi [48]. Also we employ arbitrarily distributed service time in certain special cases of the model discussed here and analyze the system using supplementary variable technique [19].

In Chapter 6, we extend the discussion in previous chapter to a single server system offering $n$ distinct services. Arrival of customers is according to $MAP$ and service time has phase type distribution. For a customer any one among the $n$ services is required and the remaining $n - 1$ are damaging (undesirable/ inessential) for him. For different customers the exact service requirement may
differ. When the service starts, a timer starts simultaneously whose realization determines the success of service. Several performance measures, including the expected service time of a customer, are evaluated. Effects of various parameters on the performance measures are numerically investigated.

Finally a section of “concluding remarks and suggestions for future study” is included.