STATISTICAL DISTRIBUTIONS FOR SERVICE TIMES

A Thesis Submitted to the College of
Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon

by

ADEBOLANLE ADEDIGBA

Dedicated to

My Parents;
Late Alhaji S.K Adedigba and Mrs S.F Adedigba

ABSTRACT

Queueing models have been used extensively in the design of call centres. In particular, a queueing model will be used to describe a help desk which is a form of a call centre. The design of the queueing model involves modelling the arrival and service processes of the system.

Conventionally, the arrival process is assumed to be Poisson and service times are assumed to be exponentially distributed. But it has been proposed that practically these are seldom the case. Past research reveals that the log-normal distribution can be used to model the service times in call centres. Also, services may involve stages/tasks before completion. This motivates the use of a phase-type distribution to model the underlying stages of service.

This research work focuses on developing statistical models for the overall service times and the service times by job types in a particular help desk. The assumption of exponential service times was investigated and a log-normal distribution was fitted to service times of this help desk. Each stage of the service in this help desk was modelled as a phase in the phase-type distribution.

Results from the analysis carried out in this work confirmed the irrelevance of the assumption of exponential service times to this help desk and it was apparent that log-normal distributions provided a reasonable fit to the service times. A phase-type distribution with three phases fitted the overall service times and the service times of administrative and miscellaneous jobs very well. For the service times of e-mail and network jobs, a phase-type distribution with two phases served as a good model.
Finally, log-normal models of service times in this help desk were approximated using an order three phase-type distribution.

PERMISSION TO USE

ACKNOWLEDGMENTS

All glory be to GOD Almighty for the love He has for me. I owe much gratitude to GOD Almighty for His Amazing grace leading me through in life.

My profound gratitude goes to my late father, **ALHAJI ADEDIGBA** who despite his demise, his love and care for me is still being observed. I will also like to thank my loving and caring mother **MRS S.F. ADEDIGBA** for her immeasurable love, support in all ways; you are a mother in a million.

When kindness cannot immediately be returned, it should be immediately appreciated. This has given me the opportunity to forward my words of appreciation to who had been of great help to me somehow, somewhere along the line. My sincere appreciation to my able supervisor, **PROFESSOR RAJ SRINIVASAN** under whose direction and supervision this thesis was prepared. I will be ungrateful if I fail to express my sincere gratitude to my husband **ADEKUNLE ADELAKUN SEYI** for his full support throughout the period of my course. You are the best man I have ever known and will ever know. I will forever love you. My appreciation also goes to members of my committee; Professor Chris Soteros and Professor Mik Bickis.

My special regards to my God sent friend; **OLANREWAJU AKINDIPE**, thank you so much for your understanding, advice, efforts, encouragements and total support towards achieving this success. I am highly grateful. I am also grateful to my sisters; Ramat Adedigba, Modinat Adedigba, Ganiyat Adedigba, Mosun Adedigba, Mayowa Adedigba and Abimbola Salami for their support. Special thanks to **ALHAJI OLABUSARI** for his fatherly support throughout the period of my course. My special regards to **PASTOR AND MRS OYETUGA**, **MR AND MRS ADIGUN** and all members of RCCG, Saskatoon, for their encouragements and support.

Contents

List of Figures

List of Tables

Chapter 1
INTRODUCTION

1.1 INTRODUCTION

Teletraffic theory can be described as the application of probability theory to the solution of problems concerning planning, performance evaluation on operation and maintenance of telecommunication systems. Teletraffic study has been of significance to business and industry, especially customer contact centres where services are being provided via telephone, e-mail, fax, internet etc.

Before the existence of customer contact centres, there were what were called call centres where services are being provided by telephone. A call centre provides tele-services, in which customers and service agents are remote from each other. The call centre has now evolved to become a customer contact centre, reflecting that other forms of contact (e-mail, fax, text chat, internet, regular mail etc.) are handled in addition to phone calls. This feature allows them to handle a large volume of transactions in an efficient and effective manner. A typical call centre is equipped with trunk lines, sales representatives(agents) and technology to handle customer calls which involve routing calls to an agent with a particular skill.

A customer contact centre is designed for departments such as help desks, admission office, reservation desk, customer service or ticket sales. The help desk is a special type of call centre for technical support for hardware and software. It is staffed by people who can either solve the problem directly or forward the problem to someone else. The line of communication in a call centre is the trunk line. Trunking of lines allows a group of inlet switches to receive several calls on the same inlets group and route them to a limited number of outlets. Trunking is based on the statistical assumption that not all callers will wish to make telephone calls at the same time.

Figure 1.1 (www.semanainformatica.com)Representatives Handling Customers' Requests

So the call centre management can provide a lesser number of circuits than might otherwise be required, allowing many users to share a smaller number of connections and achieve capacity savings.

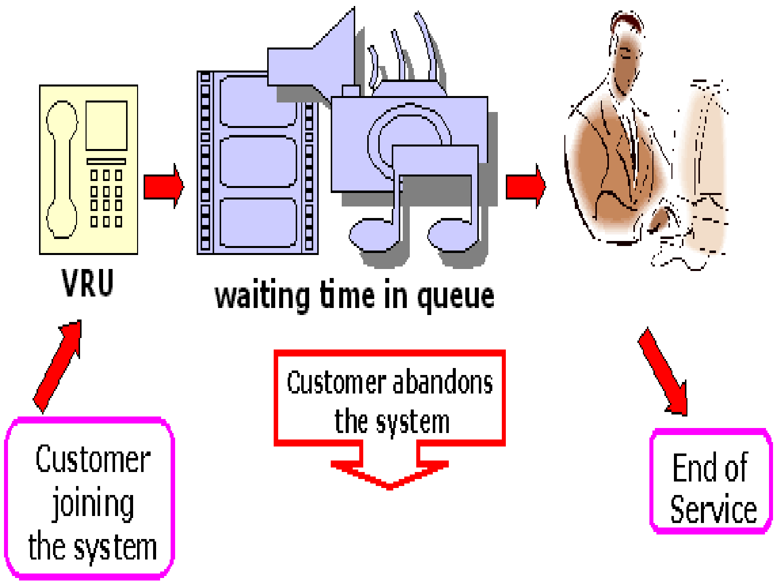

Figure 1.2 Customer Interaction with Call Centre

Figure 1.2 illustrates the process of a transaction between a customer and the call centre. Calls arriving at the call centre are connected to the voice response unit (VRU) or interactive voice response unit (IVRU), where the caller can be self served by following a set of instructions. A good illustration is when customers dial a 1-800 number and they are greeted by the IVRU/VRU after which they are asked to follow a set of instructions, for instance, for service in English, press one etc. Customers can either be served at the VRU or join the invisible queue in order to speak with agents. While waiting in the queue, the customer might abandon the system due to a long waiting time, impatience or other reasons.

Generally, the role of every customer contact centre is to maximize customer satisfaction as well as realize efficiency. This involves delivering customer satisfaction through prompt transaction handling, minimizing costs by using resources and technology efficiently and increasing profitability.

Over time, queueing models have been developed to design call centres and these models were analyzed in order to derive some performance measures. The models developed involve statistical description of the arrival process, service process, number of agents/sales representatives and system population. This research work is basically on the statistical description of the service process in a help desk.

1.2 MOTIVATION

In recent decades, there has been an explosive growth in the number of companies providing services via telephone, e-mail, fax, internet etc. Customer contact centres now comprise a large growing part of the economy. Moreover, 70% of all customer-business interactions occur in call centres and 2-3% of the American workforce is currently employed in call centres [23].

In the call centre industry, the business of real-time matching of the right customer to the right agent or resource in the right time is paramount to a successful customer experience! If done right, customers are more satisfied. The success of a customer contact centre is more than just delivering excellent service. To succeed, the management has to be equally focused on keeping costs in check. Therefore, managing contact centres is complicated because it is essential to serve many different call types requiring different service skills in a timely fashion at minimum cost.

The overall challenge in the design and management of a contact centre is to achieve a balance between operational efficiency and service quality. This involves monitoring and improving service level and measures of performance associated with service. In industrial terms, service level is defined as X percent of the customers answered in Y seconds. For instance, 80% of customers answered in 20 seconds is a common benchmark. Performance measures include probability of blocked calls, probability of abandonment, longest delay in queue, agent's occupancy, schedule efficiency, operational cost etc.

The derivation of the performance measures involve developing models to describe the contact centre and analyzing these models. A typical queueing model of a call centre involves statistical descriptions of the arrival process and the service process. This research work contributes to the analysis of this queueing model by exploring the statistical nature of the service process.

1.3 OBJECTIVE

The goal of system modelling is to provide quantitative forecasts of various system performance measures such as service level, expected waiting time, agent's occupancy, schedule efficiency, cost etc. In the design of a call centre, evaluation of these performance measures is important to making optimal decisions about overall cost, system performance, which has to be within the allowable budget and other performance based constraints.

Designing a call centre involves developing statistical models, which capture the arrival process and the service process in the call centre. At the simplest level, a call centre can be modelled as a $G|G|N|S$ queue where the first G gives a description of the random arrival process to the call centre, the second G provides a statistical description of call holding time/service time, N is the number of sales agents and $S(\geq N)$ is the system capacity i.e. the maximum number of jobs that can be handled by the system at any given time. In this research work, this queueing model will be used to analyze a help desk with a focus on the service process.

Conventionally, in queueing theory, call handling times/ service times are assumed to be independent identically distributed exponential random variables. But it has been suggested that this assumption may not hold for call centre data [24]. This present work confirms this view. It was also reported that, at least for one call centre dataset, the assumption of log-normal service time works well [23],[19]. Going by this assumption, service times in the help desk will be modelled using the log-normal distribution.

However, with the assumption of log-normal service times, the queueing model $G|G|N|S$ describing the help desk is analytically intractable. Also, some services may involve stages or many tasks before completion. This motivates the use of phase-type distributions to model service times. In this present work, service times in the help desk will be modelled using phase-type distributions, where each phase of the phase-type distributions corresponds to each stage/task of service. Statistical goodness of fit tests will be used to assess the models developed.

1.4 OUTLINE

A detailed description of a call centre as well as basic queueing models used in call centres are discussed in Chapter 2 of this work. Chapter 3 entails the description of service time models. In particular, exponential, log-normal and phase-type distributions are discussed. This chapter also gives the estimation techniques employed for estimating the parameters of the distributions as well as the EMpht program and probability plots used in implementing these techniques. The two famous goodness of fit tests, Kolmogorov Smirnov and Anderson-Darling statistical goodness of fit tests implemented in this research work are also discussed. Chapter 4 gives the descriptive analysis of the help desk data being analyzed in this research work. The preliminary analysis of this data are also presented. Chapter 5 entails the contribution of this present work. It presents the models developed for service times of the help desk. The summary of this work, recommendations for future work and conclusions of the present work are given in Chapter 6.

Chapter 2
CALL CENTRE DESCRIPTION

2.1 INTRODUCTION

This chapter gives a detailed description of a call centre. It also describes basic queueing models that have been used in the design and analysis of call centres.

2.2 CALL CENTRE DESCRIPTION

An important type of customer contact centre which provides service via telephone lines is the call centre. Call centres are classified as either inbound or outbound call centres. Inbound call centres handle incoming calls that are instigated by outside callers calling into a centre, while outbound call centres handle outgoing calls, which are calls initiated from within a centre. Usually, outbound call operations are associated with tele-marketing and survey businesses.

As the name suggests, one of the major facilities in a call centre is the trunk line. The trunk is a line of communication connecting the central exchange with specified sub-exchanges. Trunking of lines allows a group of inlet switches to receive several calls on the same inlet group and route them to a limited number of outlets. Trunking is based on the statistical assumption that not all callers will want to make telephone calls at the same time therefore, allowing the call centre management to provide a lesser number of circuits than might otherwise be required. This enables many users to share a smaller number of connections so that the management can achieve capacity savings. Typically, the number of outlets is finite and larger than the number of agents so that while all the agents are busy attending to the customers, there are available telephone lines for the incoming calls.

A representation of a telephone call centre is given in Figure 2.1. Incoming calls arrive at the Interactive Voice Response Unit(IVRU) or Voice Response Unit(VRU)

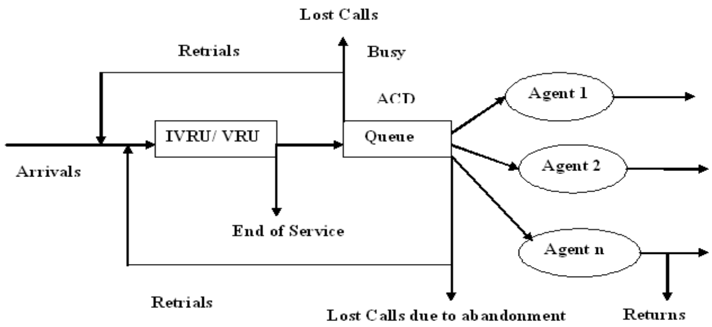

Figure 2.1 Representation Of A Telephone Call centre

where they can communicate their needs to be self served. Some customers are served at the IVRU/VRU, and then release the telephone lines and leave the system. Other customers who want to speak with an agent stays on the line and gets connected to the Automatic Call Distributor (ACD), where they can be connected to available agents. The ACD is a device designed to manage the invisible queue by routing calls to relevant available agents and achieve operational data in the call centre. If all the agents are busy, calls are put on hold and a queue starts to build up at the ACD, thereby reducing the number of trunk lines available for incoming calls and increasing the probability of customers getting busy signal.

A customer calling when all trunk lines are engaged, receives a busy signal. Such a customer might try again (retrial) or give up (lost call). In either case, the probability of a call getting lost is of interest. Customers who succeed in getting through to the ACD at a time when all agents are busy, are delayed in a queue until an agent becomes available. While waiting in the queue, if customers run out of patience before their service begins, they hang up (abandon). After abandoning, customers might try calling later, while others are lost.

The design of a call centre involves establishing the right number of agents (not too few and not too many) and allocating the right number of trunk lines, thereby, balancing the desired level of service against the availability and operational cost of these resources. Given call load data and call handling time data (time spent by agent to provide service (talk time) plus wrap-up time which is the additional activities to complete a call and prepare for the next call), a statistical model can be developed from which some performance measures can be derived.

Performance measures which can be monitored in call centres include the average speed of answer (ASA), which is the average time it takes to answer a telephone call. Another performance measure commonly used is the grade of service (GOS). The grade of service is the percentage of calls that will be answered within a targeted threshold. For instance, a target grade of service may be for 80% of calls to be answered within 20 seconds.

In the literature, queueing models have been used to study call centres [13], [34]. A queueing model has a general notation $A|B|C|D|E$ introduced by Kendall in 1953 [22], where the symbol A describes the arrival process, symbol B denotes service process, symbol C gives the number of agents, symbol D is the waiting room capacity which is the maximal number of calls that can arrive in the system (D is greater or equal to C). Symbol E is the service discipline which is usually, First Come First Served (FCFS) unless in some cases where it can be Last Come First Served (LCFS) or priority. The case of service by priority is well described in a hospital environment where priority has to be given to emergency cases in the presence of other waiting patients.

The questions that arise are about the statistical description of the arrival process (symbol A) and service process (symbol B), the number of agents to employ (symbol C) and number of trunk lines (symbol D). Many of the theories still in use today in call centre modelling were developed between 1909 and 1917 by the Danish Mathematician Agner Krarup Erlang [9]. The simplest model he used was the $M|M|N|N$ model where arrival process is Poisson with exponential interarrival time, service process is exponential, N is the number of agents with the assumption of no waiting space. This model and other models used in the design of call centres are discussed in the next section.

2.3 BASIC QUEUEING MODELS IN CALL CENTRES

In this section, some of the basic queueing models that are used in call centre studies are discussed.

2.3.1 ERLANG-B MODEL

The Erlang B model which was designed and analyzed by A.K. Erlang [9] is represented by $M|M|N|N$ where the first M stands for Poisson arrival process, the second M represents independent and identical exponential service times. The number of agents N is the same as the number of trunk lines N which implies that waiting is not allowed. Erlang B is a "blocked calls lost" model, in which when servers are unavailable, the service requestor is denied service and must retry the request. When all telephone trunks are exhausted, the caller receives a busy signal and must hang up and redial repeatedly until a server becomes available.

Erlang B calculates the blocked call probability (loss probability for a given traffic load and a number of servers).
Let $P_B(N, A)$ be the steady state probability that a call will be blocked when the offered load is A Erlangs,

$$A = \lambda/\mu$$

where λ is the arrival rate and μ is the service rate and the number of trunk lines is N.
Then $P_B(N, A)$ is the steady state probability that a finite state birth-death process (with birth rate λ and state dependent death rates $i\mu$ $(1 \leq i \leq N)$) is in state N can

be derived using results from [20],[9] as follows.

$$P_B(N,A) = \frac{A^N/N!}{\sum_{i=0}^{N} A^i/i!} \quad . \tag{2.1}$$

Using this formula, given A and N, the blocking probability can be determined. The reverse problem is to determine N for a given call load A and a prescribed blocking probability which is done using tables.

2.3.2 ERLANG-C MODEL

Unlike the Erlang B model, in which blocked service requests are considered lost, in the Erlang C model, calls that cannot be answered immediately are delayed until a server is available. The representation of the model is given as , $M|M|N|\infty$, that is, there are N number of servers with infinite waiting room capacity. This model assumes that callers will be willing to wait forever to be answered by agents but in practice, some callers will hang up the phone as soon as they are put on hold, while some will abandon the call after waiting in the queue for some time. Erlang-C formula provides the probability that a call will have to wait for service.

Let $P_c(N,A)$ be the probability that a call will have to wait for service if N agents are assigned to handle traffic of A Erlangs ($A = \lambda/\mu$),

Then $P_c(N,A)$ can be derived as [20] ,[9]:

$$P_c(N,A) = \frac{\frac{A^N N}{N!(N-A)}}{\sum_{i=0}^{N-1} \frac{A^i}{i!} + \frac{A^N N}{N!(N-A)}} \quad . \tag{2.2}$$

2.3.3 ERLANG-A MODEL

As an improvement to the classical $M|M|N|\infty$ queueing model , Erlang-A accommodates the significance of abandonment in modelling and practice.

Palm [29] suggested to enrich Erlang-C by associating with each arriving caller an exponentially distributed patience time with mean θ^{-1}. An arriving customer encounters an offered waiting time, which is defined as the time that this customer would have to wait given that his or her patience is infinite. If the offered wait exceeds the customer's patience time, the call is then abandoned, otherwise the customer awaits service. The patience parameter θ is referred to as the individual abandonment rate. The Erlang-A model is denoted by $M|M|N+M$ and the steady state distribution as well as performance measures were derived by Baccelli [6].

2.4 GENERAL MULTISERVER QUEUEING MODEL

The help desk which is a form of call centre can be modelled as a multiserver queue $G|G|N|S$. In the queueing models described so far $M|M|N|S$, $S \geq N$, the arrival process is assumed to be Poisson and service times are assumed to be exponentially distributed. But recent research suggests that for call centre data these are seldom the cases. The arrival process can be generalized by replacing the Poisson arrival process M with a renewal process $GI|M|N|S$ or by an inhomogenous Poisson process, $M(t)|M|N|S$. The assumed exponential service times can be replaced with hyperexponential, erlang, coxian etc. i.e. $G|G|N|S$. As suggested in [23], another distribution that can be considered for modelling the service times in a call centre is the log-normal distribution but the draw back with this is that, with log-normal service time, the queueing model $G|G|N|S$ is difficult to analyze. Also, in order to model each stage of service, phase-type distribution will be used to model service times of the help desk.

Chapter 3
EXPONENTIAL, LOG-NORMAL AND PHASE-TYPE DISTRIBUTION

3.1 INTRODUCTION

In this chapter, different models of service times are described. This involves giving the description of various statistical distributions that can be used in modelling service times. Techniques for estimating the parameters of the distributions are also provided. EMpht program and probability plots used in implementing these techniques are also described. Specifically, this chapter examines exponential, log-normal and phase-type distributions as models of service times. The two famous goodness of fit tests, Kolmogorov-Smirnov and Anderson-Darling tests implemented in this research work will also be discussed.

3.2 EXPONENTIAL MODEL

A random variable T is said to be exponentially distributed with parameter λ if the probability density function is given as:

$$f(t) = \lambda \exp(-\lambda t),\ \lambda > 0,\ t \geq 0.$$

The cumulative distribution function (CDF), *F(t)* $= P(T \leq t) = 1 - \exp(-\lambda t)$. The reliability function *R(t)*= 1- *F(t)* $= \exp(-\lambda t)$. The failure rate function gives the probability that customers will finish receiving their service during a very small time interval, assuming they were in service at the beginning of the interval. The exponential distribution is the only distribution with a constant failure rate and it is given by $h(t) = \frac{f(t)}{R(t)} = \lambda$. The moment generating function $M_T(x)$ is given as: $E[e^{xT}] = \frac{\lambda}{\lambda - x},\ x < \lambda$. In particular, the random variable T has mean $1/\lambda$ and variance $1/\lambda^2$.

A special property of the exponential distribution is the memoryless property. A random variable T is said to be without memory, or memoryless, if
$P(T > t + s|T > t) = P(T > s)$.
This memoryless property makes the exponential distribution to be widely used in reliability modelling.

3.3 LOG-NORMAL MODEL

A random variable T is log-normal with parameters (μ, σ^2) if $\ln(T)$ is normally distributed with mean μ and variance σ^2. The log-normal distribution has been used widely in reliability applications to model failure times. The probability density function (PDF), *f(t)* is defined as $f(t) = \frac{1}{\sigma t\sqrt{2\pi}} \exp\left(\frac{-(\ln t-\mu)^2}{2\sigma^2}\right)$. It's cumulative distribution function (CDF) *F(t)* which is $P(T \leq t)$ is given as $\Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)$ $t \geq 0,\ \sigma > 0$, where Φ is the cumulative distribution function of the standard normal distribution.

The reliability function *R(t)* which is $P(T > t)$ is given as $\int_t^\infty f(t)dt$. There is no closed-form solution for the log-normal reliability function. The failure rate function *h(t)* is defined as $\frac{f(t)}{R(t)}$. Standard normal tables will be required to solve this failure rate function. The moment generating function (MGF) of log-normal distribution can be computed from the moment generating function of the normal distribution using the transformation $t = \ln(T)$. The moment generating function $M_T(x)$ of normal distribution is given as: $E[e^{xT}] = \exp(\mathrm{x}\mu + \frac{\sigma^2\mathrm{x}^2}{2})$. In particular, the mean of the log-normal distribution is $\exp((\mu + \frac{\sigma^2}{2}))$ and the variance is $(\exp(2\mu + 2\sigma^2)) - (\exp(2\mu + \sigma^2))$.

The Log-normal distribution has played major roles in diverse areas of science. Royston [31] modelled survival time in cancer with an emphasis on prognostic factors using the log-normal distribution. Log-normal distribution also played a major role in modelling failure rate distribution for worst case bound reliability prediction as described by Bishop and Bloomfield [8] .

3.4 PHASE-TYPE DISTRIBUTION

A nonnegative random variable T is said to be of phase-type if T is the time a finite-state Markov process takes to reach an absorbing state from its initial distribution. Phase-type distribution is an extension of the exponential distribution, erlang distribution, sum of exponentials, mixture of exponentials and hyperexponential distributions. More generally, phase-type distribution is any series|parallel|loop arrangement of exponential distributions [21]. If the parameter of the Markov chain is continuous, then the phase-type distribution is a continuous phase-type distribution (CPH) and if the parameter of the Markov chain is discrete it is called discrete phase-type distribution (DPH).

The phase-type distribution was introduced by Neuts in 1975 [26] and since then, this distribution has proven its utility in stochastic modelling such as in queueing theory, reliability theory, renewal theory, lifetime analysis, survival models and many others. The simplest examples of the phase-type distributions are mixtures and convolutions of exponential distributions, in particular, erlang distributions.

This research work focuses on continuous phase-type distributions, in which the time to absorption is being described by the service time, which is the time a customer spends from the beginning of service till end of service before leaving the system. When service times consist of several stages, the phase-type distribution models each stage of service as phases in the phase-type distribution.

3.4.1 REPRESENTATION OF CONTINUOUS PHASE-TYPE DISTRIBUTION

Before giving the representation of phase-type distribution, some basic concepts are given.

Edward [21] defined a stochastic process as a collection of random variables, $X = \{X(t), t \epsilon T\}$, in which the set T is called the index set. When the index set T is countable, X is called a discrete time process and when the index set is an interval of the real line, the stochastic process is called a continuous time process. $X(t)$ takes values in a set S called the state space of the process for every $t \epsilon T$. When S is countable, then the process has a discrete state space and when S is an interval of the real line, the process has a continuous state space.

A stochastic process $X = \{X(t),\ t \geq 0\}$ where $X(t)$ denotes the state of the system at continouus time t, on a discrete state space $S = \{0, 1, 2, ...\}$ is called a continuous time Markov chain (CTMC) if

$$P\{X(t+s) = j | X(s) = i,\ X(u) = x(u),\ 0 \leq u < s\} = P\{X(t+s) = j | X(s) = i\}$$

for all $s \geq 0,\ t \geq 0, x(u),\ 0 \leq u \leq s$.

In a CTMC, the process starts in state i, stays there for an amount of time (sojourn time) which is exponentially distributed before making a change to another state j. CTMC with an absorbing state is of particular interest. In a CTMC with an absorbing state, the process starts in state i(transient state), visits other transient states and eventually gets absorbed in the absorbing state where it stays forever.

The transition rates of moving from one state to another constitute the elements of the matrix called the infinitesimal generator of a CTMC with each row sum equal to zero. Ignoring the information about the exponential sojourn time in each state, the resulting sequence of states visited by the process is a discrete time Markov chain referred to as the embedded discrete time chain. A CTMC is completely characterized by the infinitesimal generator and the starting state probability vector.

Let $X = \{X_t, t \geq 0\}$ be a continuous time Markov chain (CTMC) on discrete states $\{1, 2, 3, ..., k, \triangle\}$, where $\triangle$ is the absorbing state with infinitesimal generator given in the blocked form:

$$Q = \begin{pmatrix} R & r \\ 0 & 0 \end{pmatrix},$$

where R is a $k \times k$ dimensional matrix.

R is the generator matrix (non- singular) block of the Markov process, infinitesimal generator corresponding to the transient states $[1, 2, ..., k]$.

Since the sum of each row in the infinitesimal generator is zero, it follows that, $r = -R1$ where **1** is a column vector of ones.

The initial distribution, which is the probability that service will start from each of the states $\{1, 2, 3, ..., k, \triangle\}$ is given as:

$\mathbf{q} = (q_1, q_2, ..., q_k, 0)$ where $q_i \geq 0, \ \sum_i q_i = 1$.

Therefore the phase-type distribution can be represented as $[q, R]$.

- The transition probability matrix for the embedded Markov chain can be computed as,

$$P_{ij} = \begin{cases} \frac{-R_{ij}}{R_{jj}} & \text{if } j \neq i \ 1 \leq i \leq k; \\ 0 & \text{if } j = i \ 1 \leq i \leq k \end{cases}.$$

- The absorption probability vector (i.e. the probability that the process gets to the absorption state from each of the transient states) is given as,

  $P_{j\triangle} = \frac{-r_j}{R_{jj}} \ 1 \leq j \leq k$

  $(\sum_j P_{ij}) + P_{j\triangle} = 1.$

- The expected length of time the Markov chain spends in each of the transient states $[1, 2, ..., k]$ in seconds can be computed as:

  $M_j = \frac{-1}{R_{jj}} \ 1 \leq j \leq k.$

It should be noted that the parameters of the phase-type distribution are not identifiable. That is, doing different fits for a specific order of the phase-type distribution using different initial distributions produces different sets of parameter estimates describing the same phase-type distribution. For example, fitting an order three phase-type distribution PH(3) to a data set using two different initial distributions will produce two sets of parameter estimates, $[q, R]$ and $[q', R']$ respectively.

The likelihood functions of these two sets of estimates are almost the same and the density functions of the two, when plotted also coincide. Therefore, the two sets of PH(3) model describes the same data.

3.4.2 PROPERTIES OF PHASE-TYPE DISTRIBUTION

Let T be a nonnegative random variable representing time to absorption. Then, given the infinitesimal generator matrix Q, the transition matices P_t can be derived by solving

$$\frac{d}{dt}P_t = QP_t \ .$$

subject to initial condition,

$$P(I) = 0 \ .$$

The unique solution to $\frac{d}{dt}P_t = QP_t$ is

$$P_t = \exp\{tQ\} \ .$$

The transition matrix

$$P_t = \exp\{Qt\} = \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!} = I + \sum_{n=1}^{\infty} \frac{t^n Q^n}{n!}$$

can be block partitioned as:

$$P_t = \begin{pmatrix} e - \exp\{Rt\}e & \exp\{Rt\} \\ 1 & 0, ..., 0 \end{pmatrix}$$

(where e is a column vector of ones) which gives an expression for the cumulative probability distribution function CDF ((F(t))which is $P(T \leq t)$. $F(t) = 1 - q\exp\{tR\}e$. Differentiating *F(t)* with respect to t gives the probability density function PDF of T

$$f(t) = q\exp\{tR\}r, \ r = -Re.$$

Laplace transform

$$LT = \int_0^\infty e^{-st} F_t dt = q(sI - R)^{-1} r = q \left[\frac{det(sI - R)_{ji}}{det(sI - R)} \right] r.$$

and

$$r^{th}\ moment\ E(T^r) = \int_0^\infty F(t)(dt) = (-1)^r r! q R^{-r} e.$$

In particular, phase-type distribution has mean $-qR^{-1}e$ and variance

$$(2q(RXR)^{-1}e) - (-qR^{-1}e)^2.$$

3.4.3 CHARACTERISTICS OF PHASE-TYPE DISTRIBUTIONS

The phase-type distributions are dense since it can be used to approximate all probability distributions on $[0, \infty)$ [4]. Phase-type distribution is structurally informative because it can be used to reflect the qualitative features of the model and also provide useful information on its physical behavior through the interpretation of numerical results. The phase-type distribution can be used for Markov modelling and has a simple probabilistic interpretation. It also preserves the underlying Markov structure of stochastic models.

In the last decade, Phase-type distributions have been used extensively in modelling insurance risk. Phase-type distributions have also played major roles in areas such as telecommunications, teletraffic modelling, queueing theory, reliability theory and biostatistics.

Aruna et al.[3] modelled the service time distribution in cellular networks using phase-type distributions. Marshall and McLean [25] used conditional phase-type distributions to model patient length of stay in hospital, Ishay [19] also fitted phase-type distributions to data from a telephone call centre.

3.5 ESTIMATION TECHNIQUE

A commonly used estimation technique, is the method of maximum likelihood parameter estimation. A likelihood function $L(\alpha)$ is the probability density for the occurrence of a sample configuration $x_1, ..., x_n$ given that the probability density $f(x;\alpha)$ with parameter α is known. The general maximum likelihood parameter estimation technique is described below.

3.5.1 MAXIMUM LIKELIHOOD PARAMETER ESTIMATION

Let T be a continuous random variable with probability density function:

$$f(t;\theta_1, \theta_2, ..., \theta_k)$$

(where $\theta_1, \theta_2, ..., \theta_k$ are k unknown parameters which are to be estimated). With N independent observations $t_1, t_2, ..., t_N$ of T, the likelihood function is given by the following product:

$$L(\theta_1, \theta_2, ..., \theta_k | t_1, t_2, ..., t_N) = L = \prod_{i=1}^{N} f(t_i; \theta_1, \theta_2, ..., \theta_k). \tag{3.1}$$

The logarithmic likelihood function is given by,

$$\Lambda = \ln L = \sum_{i=1}^{N} \ln f(t_i; \theta_1, \theta_2, ..., \theta_k). \tag{3.2}$$

The maximum likelihood estimators of $\theta_1, \theta_2, ..., \theta_k$ are obtained by maximizing the likelihood or log-likelihood given in equation 3.1 or 3.2 respectively. This involves solving the following k simultaneous equations,:

$$\frac{\partial(\Lambda)}{\partial\theta_j} = 0, \ j = 1, 2, 3, ..., k. \tag{3.3}$$

and checking the negativity of the second derivatives at the critical points.

Using the maximum likelihood estimation technique, the maximum likelihood parameter estimate for the exponential distribution is given as:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i} \ . \tag{3.4}$$

The maximum likelihood estimates for the scale parameter μ, and the shape parameter σ, of the log-normal distribution are,

$$\mu = e^m \ , \tag{3.5}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(\ln(t_i) - m)^2}{N}} \ , \tag{3.6}$$

where

$$m = \frac{\sum_{i=1}^{N}\ln(t_i)}{N} \ . \tag{3.7}$$

The maximum likelihood parameter estimates for phase-type distribution are difficult to derive. However, the well known iterative technique, expectation-maximization (EM) algorithm will be used to estimate the parameters of the phase-type distribution. The EM algorithm will be discussed in the next section.

3.5.2 GENERAL EXPECTATION-MAXIMIZATION (EM) ALGORITHM

The EM-algorithm is an iterative method for maximum likelihood estimation. Its area of applications concerns incomplete data, that is, data which can be thought as partial observations of a larger experiments, where a more specified course of events can be observed than in the experiment actually performed. The EM-algorithm can also be used to fit a phase-type distribution to a theoretically given distribution. Each iteration of the algorithm has two parts. The first part is the computation of the expected value of the log-likelihood function conditioned by observed data as a function of an unknown parameter (E-step). The second part is the maximization of the likelihood function over some parametric space (M-step).

3.5.3 EM-ALGORITHM FOR PHASE-TYPE DISTRIBUTION

An observation y of the time to absorption (service time) can be regarded as an incomplete observation of the Markov process X_t. It is incomplete in the sense that it only gives the time when the process hits $\triangle$ (the absorbing state) and does not provide any information about how it got there, where it started, which states it visited and for how long. Consequently, given an observation y of the phase-type distribution, a complete observation of the process X_t on the interval $(0, y]$ can be represented by

$$X = (i_0, i_1, ..., i_{m-1}, s_0, s_1, ..., s_{m-1})$$

where $i_0, i_1, ..., i_{m-1}(i_m = \triangle)$ are states of the embedded Markov chain and $s_0, s_1, ..., s_{m-1}(s_m = \infty)$ are the sojourn times, m is the number of jumps until X_t hits the absorbing state $\triangle$. The sojourn time must satisfy $y = s_0 + s_1 + ... + s_{m-1}$,

Suppose that there are n independent replications of the process $X_t^{[1]}, ..., X_t^{[n]}$ and let $I_0^{[v]}, ..., I_{M^{[v]}-1}^{[v]}$ denote the embedded Markov chain and $S_0^{[v]}, ..., S_{M^{[v]}-1}^{[v]}$ the holding times for the v^{th} process. Hence a sample of size n is represented by

$$X = \left(i_0^{[1]}, ..., i_{m^{[1]}-1}^{[1]}, S_0^{[1]}, ..., S_{m^{[1]}-1}^{[1]} ..., i_0^{[n]}, ..., i_{m^{[n]}-1}^{[n]}, S_0^{[n]}, ..., S_{m^{[n]}-1}^{[n]}\right).$$

This is the complete data set which will be used in the EM- algorithm to find the maximum likelihood estimate (q,R) from the observed sample,

$$\mathbf{y} = (y_1, ..., y_n) = \left(S_0^{[1]} + ... + S_{m^{[1]}-1}^{[1]}, ..., S_0^{[n]} + ... + S_{m^{[n]}-1}^{[n]}\right).$$

The density of the complete sample X can be written in the form,

$$f(x; q; R) = \prod_{i=1}^{k} q_i^{B_i} \prod_{i=1}^{k} exp\{R_{ii} z_i\} \prod_{i=1}^{p} \prod_{j=0, j\neq i}^{k} R_{ij}^{N_{ij}}, \tag{3.8}$$

where

$B_i = \sum_{v=1}^{n} 1_{\{I_0^v = i\}}$ = the number of Markov process starting in state i, $i = 1, ..., k$,

$z_i = \sum_{v=1}^{n} \prod_{u=0}^{m(v)-1} 1_{\{I_u^{[v]} = i\}} S_u^{[v]}$ = the total time spent in state i, $i = 1, ..., k$, and $N_{ij} = \sum_{v=1}^{n} \sum_{u=0}^{m^{[v]}-1} 1_{\{I_u^{[v]}=i,\, I_{u+1}^{[v]}=j\}}$ = the total number of jumps from state i to state j for $i \neq j,\ i = 1, ..., k$, and $j = 0, 1, ..., k$.

Let

$$S = ((B_i)_{i=1,...,k}, (z_i)_{i=1,...,k}, (N_{ij})_{i=1,...,k,j=0,1,...,k,\ i\neq j}) \tag{3.9}$$

represent the sufficient statistics above, it follows either by general theory for exponential families or by explicit calculations (using $-(r_i + \sum_j r_{ij}) = r_{ii}$) [7] that the maximum likelihood estimates based on sample X, are:

$$\begin{aligned} q_i &= \frac{B_i}{n} \\ R_{ij} &= \frac{N_{ij}}{z_i} \\ r_i &= \frac{N_{ij}}{z_i} \\ R_{ii} &= \left(r_i + \sum_{j=1, j\neq i}^{k} R_{ij} \right)\ i, j = 1, ..., k. \end{aligned} \tag{3.10}$$

The detailed derivation can be found in [7].

Let $B_i^{[v]},\ Z_i^{[v]},\ N_{ij}^{[v]}$ be the contributions to S in equation 3.9 from the v^{th} observed process, then the $u + 1_{st}$ iteration of the EM-algorithm becomes:

E-STEP: Calculate

$$\begin{aligned} B_i^{(u+1)} &= \sum_{v=1}^{n} E_{(q,R)(u)}[B_i^{[v]}|y_v]\ i = 1, ..., k \\ Z_i^{(u+1)} &= \sum_{v=1}^{n} E_{(q,R)(u)}[Z_i^{[v]}|y_v]\ i = 1, ..., k \\ N_{ij}^{(u+1)} &= \sum_{v=1}^{n} E_{(q,R)(u)}[N_{ij}^{[v]}|y_v]\ j \neq i, i = 1, ..., k\ j = 0, 1, .., k. \end{aligned} \tag{3.11}$$

M-STEP: The new estimates are given by:

$$q_i^{(u+1)} = \frac{B_i^{(u+1)}}{n}$$

$$r_{ij}^{(u+1)} = \frac{N_{ij}^{(u+1)}}{Z_i^{(u+1)}}$$
$$r_i^{(u+1)} = \frac{N_{i0}^{(u+1)}}{Z_i^{(u+1)}}$$
$$r_{ii}^{(u+1)} = -(r_i^{(u+1)} + \sum_{j=1, j\neq i}^{p} r_{ij}^{(u+1)}). \quad (3.12)$$

The likelihood function,

$$L(q, R; y) = \prod_{i=1}^{n} q \exp\{\mathrm{Ry_i}\}\, \mathrm{r} \quad (3.13)$$

increases in every iteration and the sequence of estimates converges towards a stationary point (q, R) of the likelihood.

Any continuous distribution can be approximated using phase-type distribution by fitting a density (f) of a phase-type distribution to the density (g) of the given distribution. This is minimizing the information divergence (relative entropy or Kullback-Leibler information). The information divergence or relative entropy of f with respect to g is

$$\int log \left(\frac{g(t)}{f(t; q, R)} \right) g(t)dt.$$

The smaller the relative entropy, the more similar the phase-type distribution and the given distribution. This concept is a natural analogue to maximizing the log-likelihood function in fitting phase-type distribution to a sample [4].

3.6 EMPHT PROGRAM

EMpht is a program for fitting phase-type distributions. It can be used either to fit a phase-type distribution to a sample (which may contain censored observations) or to make a phase-type approximation of another continuous distribution. The EMpht program given in Olsson (1996) [27] implements the E-M algorithm developed by Asmussen et al (1996) [4] for the phase-type distributions. It is written in C and is complemented by a Matlab program, PHplot, for graphical display of the fitted phase-type distribution.

EMpht calculates the estimates of the elements in (q,R), the parameters of the phase-type distribution, for a fixed order/phases given by the user. Starting with initial values (q^0, R^0), which are either provided by the user or randomly generated in EMpht, the program produces a sequence of parameter estimates

$$(q^1, R^1), (q^2, R^2), ..., (q^N, R^N).$$

Each of these estimates corresponds to one iteration of an EM-algorithm, which implies that the likelihood function increases in every iteration. By using EMpht to fit a phase-type distribution to a sample $\mathbf{t} = (t_1, ..., t_n)$, it is certain that each new estimate is better than the previous one in the sense that

$$L(q^k, R^k; t) \leq L(q^{k+1}, R^{k+1}; t),$$

where

$$L(q^k, R^k; t) = \prod_{i=1}^{n} q \exp\{Rt_i\} r$$

is the likelihood function of the phase-type distribution.

Approximation of another non-negative continuous distribution by a phase-type distribution is done by minimizing the information divergence (the relative entropy or Kullback- Leibler information). If the distribution to be approximated has density h(.), and the phase-type density is denoted by f(. ; q,R), then the information divergence of f with respect to h is

$$\int log\left(\frac{h(t)}{f(t; q, R)}\right) h(t)dt.$$

For a fixed density h, the information divergence is minimized by maximizing

$$\int log[f(t; q, R)]h(t)dt.$$

The minimization of the information divergence can be regarded as an infinitesimal analogue of maximization of the log-likelihood function. In EMpht program, the same EM-algorithm is used to perform the minimization of the information divergence, as is used in maximization of the log-likelihood function of a sample. The sequence of estimates converges towards a stationary point $(\hat{q}, \hat{R})$ of the likelihood, and if the number of iterations, N, is large enough then, $(q^N, R^N) \approx (\hat{q}, \hat{R})$.

Due to the non-identifiability of the parametrization of the phase-type distribution, there can be different set-ups of estimates $(\hat{q}, \hat{R})$ corresponding to the same phase-type distribution. Typically, when different fits of a specific order of phase-type distribution to a sample or a distribution are performed using EMpht, they will result in different parameter values. However the value of the likelihood function are mostly the same for the different fits, and the corresponding densities are, when plotted, seldom possible to distinguish from each other.

3.7 PROBABILITY PLOT (P-P Plot)

The probability plot is a graphical technique for assessing whether or not a data set follows a given distribution. Given an ordered sample of independent observations $x_{(1)} \leq x_{(2)} \leq ..... \leq x_{(n)}$ from a population with estimated distribution function $\hat{F}$, a probability plot consists of the points

$$\left\{ \left(\hat{F}(x_{(i)}) \, , \frac{i}{n+1} \right) \ : i = 1, ..., n \right\} \ .$$

If $\hat{F}$ is a reasonable model for the population distribution, the points of the probability plot should lie close to the unit diagonal. Significant departures from linearity provide evidence that $\hat{F}$ is not a suitable model for the data [14].

In this research work, the probability plot will be used to check if service times in the help desk come from exponential and log-normal distributions.

3.8 KOLMOGOROV-SMIRNOV STATISTICAL TEST

The Kolmogorov Smirnov (K-S) test is used to test whether a sample comes from a population with a specified distribution. A hypothesis test involves calculation of a test statistic from the data and the probability of obtaining a value at least as large as a tail area if the correct distribution is chosen. The K-S test is based on the empirical cumulative distribution function (ECDF), it measures the difference between the empirical cumulative distribution function and the hypothesized cumulative distribution function. The K-S test is distribution free since its critical values do not depend on the specific distribution being tested. The K-S test is relatively insensitive to differences in the tails but more sensitive to points near the median of the distribution.

The Kolmogorov Smirnov (K-S) test is defined as follows:

H_0 : The data follows a specified distribution

H_1 : The data do not follow the specified distribution.

Test Statistic: A two-tailed test statistics associated with Kolmogorov-Smirnov statistics are,

$$\begin{aligned} D_n^+ &= \max_{i=1,2,..,n} \left(\frac{i}{n} - \hat{F}(x_i) \right) \\ D_n^- &= \max_{i=1,2,...,n} \left(\hat{F}(x_i) - \frac{i-1}{n} \right) \\ D &= \max \left\{ D_n^+, D_n^- \right\} \end{aligned} \tag{3.14}$$

n is the sample size, $x_{(1)}, x_{(2)}, ..., x_{(n)}$ are the order statistics and $\hat{F}(x_i)$ is the cumulative distribution function of the fitted distribution.

The maximum positive difference, D_n^+, detects the largest vertical deviation between the two cumulative distribution functions (CDF's) where the fitted CDF is below the empirical CDF. Likewise the maximum negative difference, D_n^-, detects the largest vertical deviation between the two CDF's where the fitted CDF is above the empirical CDF. The hypothesis regarding the distributional form is rejected if the test statistic $D\sqrt{n}$ is greater than the critical value (CV) obtained from a table. Smaller values of D indicate better fit.

According to Stephens [33], when the hypothesized distribution $\hat{F}(x)$ is continuous and completely specified, then the modified test statistic should be used. The modified test statistic is,

$$D_n = D\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right). \tag{3.15}$$

The outcome of a statistical test depends on the number of observations n available. The larger n is, the better chances of rejecting an inadequate distribution.

3.9 ANDERSON-DARLING (A-D) STATISTICAL TEST

The A-D test is an adjustment of the K-S test. It gives more weight to the tails than the K-S test and is known to be more powerful than the K-S test [11]. Unlike the K-S test, the A-D test makes use of the specific distribution being tested in calculating the critical values. The A-D test is a one-sided test and the null hypothesis is rejected for test-statistic greater than the critical value.

Formally the test is defined as follows:

H_0 : The data follows a specified distribution $F(x_i)$

H_1 : The data do not follow the specified distribution.

Test Statistic is given as,

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln z_i + \ln(1 - z_{n+1-i})]. \tag{3.16}$$

where $z_i = F(x_i)$.

A table of critical values for the K-S test and A-D test for continuous and completely specified distributions as extracted from Stephens [33] is presented below.

Table 3.1 Critical Values (CV) For K-S and A-D Tests

Level of Significance	0.15	0.10	0.05	0.025	0.01
D_n	1.138	1.224	1.358	1.480	1.628
A^2 For all $n \geq 5$	1.610	1.933	2.492	3.070	3.857

The K-S and A-D tests are implemented in this research work.

Chapter 4
HELP DESK DATA AND DESCRIPTIVE ANALYSIS

4.1 INTRODUCTION

This chapter gives a description of the help desk data analyzed in this research work. It also gives some preliminary descriptive information about the data, mostly about transaction load. Further analysis will not be on transaction load but on handling time/service time in the help desk.

4.2 DATA DESCRIPTION

The data analyzed in this research work was drawn from a help desk operation at the University of Saskatchewan. This help desk is a special type of call centre for technical support for hardware or software and are staffed by people who can either solve the problem directly or forward the job to someone else. A representation of the help desk is given in Figure 4.1. Transactions arriving at the help desk which

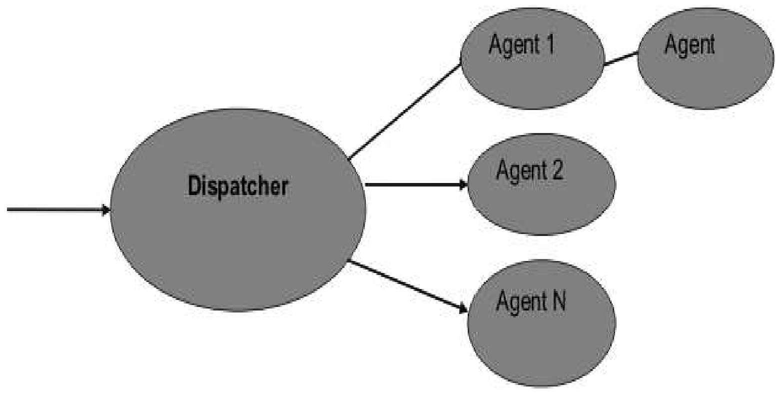

Figure 4.1 Representation of customer interaction with help desk

can be in the form of telephone call or e-mail are received by a main server which then distributes the jobs to relevant available agents depending on the type of job. If the job transferred to particular agent(s) could not be handled, it will be transferred to another agent(s) who can handle the job. For the help desk being analyzed, transactions come by E-mail, telephone, in-person (walk in) and campus mail (postal). These jobs are categorized into thirty-four job types which include, Accounts-Admin, Accounts-Dialup, Administration, Applications, E-mail etc.

The jobs are ticketed as, Administration, Mac, E-mail, etc and all tickets defaults to the main server, dispatcher, before being transferred to agents who will handle the job. A software developed by the help desk operation records the following list of information for each transaction.

Creation Date: This describes the date at which the job was received.

Creation Time: This is the time of the day at which the job was received.

Creation Day: 1=Monday, 2=Tuesday, etc.

Creation Month:1=Jan, 2=Feb, etc.

Creation Year: This is the year the job was created.

Completed Date: This is the date the job was completed.

Completed Time: This was suppose to be the time at which the job was completed, but in some cases, the recorded completed time include the time used in doing other work after the job has been completed.

Completed Day: This is the day the job was completed.

Completed Month: This indicates the month the job was completed.

Completed Year: This is the year the job was completed

Duration: This gives the actual time difference between the time the job was received and when it was completed. It is the handling time used for analysis in this research work.

The help desk operates between 8:30a.m. and 4:30p.m. Monday through Friday. The data obtained consists of 43,689 records between September 2001 and March 2004. A portion of the raw data from the help desk for a particular day i.e. 3rd, September 2001 is given below.

Table 4.1 Sample Data

Creation Date	Creation Time	Creation Day	Creation Month	Creation Year
9/3/2001	8:22:42	2	9	2001
9/3/2001	8:26:08	2	9	2001
9/3/2001	8:36:40	2	9	2001
9/3/2001	8:38:47	2	9	2001

Table 4.2 Sample Data Continued

Completed Date	Completed Time	Completed Day	Completed Month	Completed Year
9/3/2001	13:50:05	2	9	2001
9/3/2001	15:53:04	2	9	2001
9/3/2001	8:38:35	2	9	2001
9/3/2001	8:39:41	2	9	2001

Table 4.3 Sample Data Continued

Duration	Problem Source	Ticket Category	Queue
0:36:00	E-mail	Administration	Win/VMS
0:00:09	Phone	Mac	Mac/Unix
0:01:55	In Person	Administration	Dispatcher
0:00:54	In Person	Administration	Dispatcher

From the second row of the sample data, a job came in by e-mail and was ticketed as an administrative type of job. The job was queued to the Win/VMS server where the job can be done or transferred to other agent(s) who can handle the job. From the creation time(8:22:42) and completed time(13:50:05), the difference in time was supposed to be the duration (0:36:00) but was not because the completed time was posted long after the job was completed. For the transaction in the $4th$ row, the completed time(8:38:35) was posted once the job was completed, therefore, the difference between creation time(8:36:40) and completed time gave the actual duration(0:01:55) i.e. the job was completed in 1 minute and 55 seconds. Therefore the raw data cannot be used directly for the analysis and cleaning of the data becomes necessary.

Cleaning of the data includes, extracting service duration in seconds from the variable "Duration" which was in the form 0:00:00, ignoring service duration of zero seconds and sorting of data to different categories, for instance service times by job types etc. Cleaning of the data reduced the dataset from $43,689$ to $35,269$ transactions.

One of the main objectives of design and operation of a help desk is to determine the right number of support people to employ during a shift, that is, how many different technical support people should be employed in order to provide reasonable service. In order to accomplish this goal, a statistical model describing the help desk has to be developed. The help desk can be modelled as a multiserver queue $G|G|N|S$, where the first G describes the arrival process of jobs which will not be analyzed in this research work, the second G describes the service process, N is the number of agents handling these jobs and S can be referred to as the system capacity. Analysis of the service process is the focus in this research work.

4.3 PRELIMINARY ANALYSIS OF DATA

YEARLY TRANSACTIONS

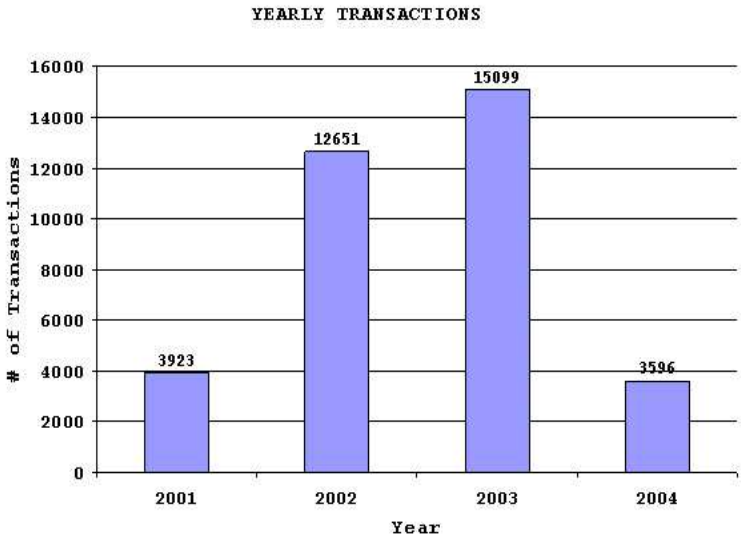

Figure 4.2 Yearly Transactions

Figure 4.2 is the plot of yearly transactions volume at the help desk. Transactions for years 2002 and 2003 are for complete months while 2001 data is between September and December and 2004 is from January to March. In 2002 and 2003, the help desk handled approximately between 12,000 and 15,000 transactions.

MONTHLY TRANSACTIONS

Having examined yearly transactions, it is important to describe the monthly transactions in the help desk. Figure 4.3 gives the plot comparing monthly transactions for the four years 2001 to 2004, while Figure 4.4 is the plot of monthly transactions for the year 2002. From Figure 4.3, it is apparent that monthly transactions have a similar pattern over the years and from the graphs, it can be observed that the peak is in September, meaning that the help desk handles a higher number of

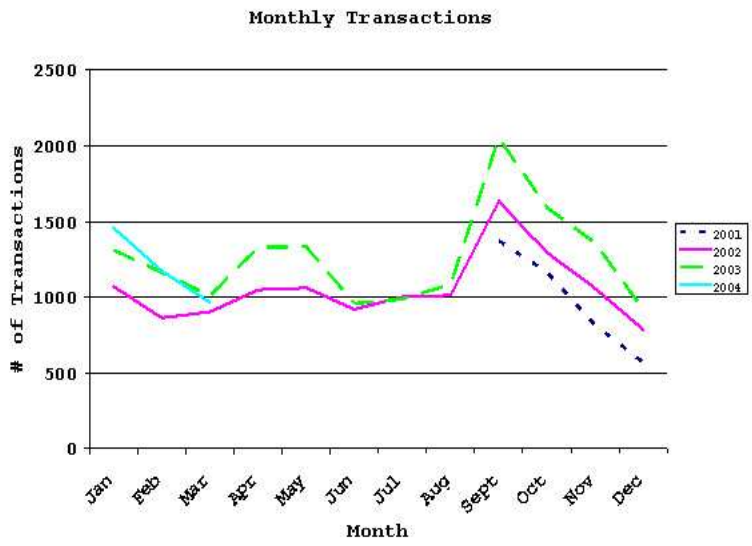

Figure 4.3 Comparison of Monthly Transactions for years 2001 - 2004

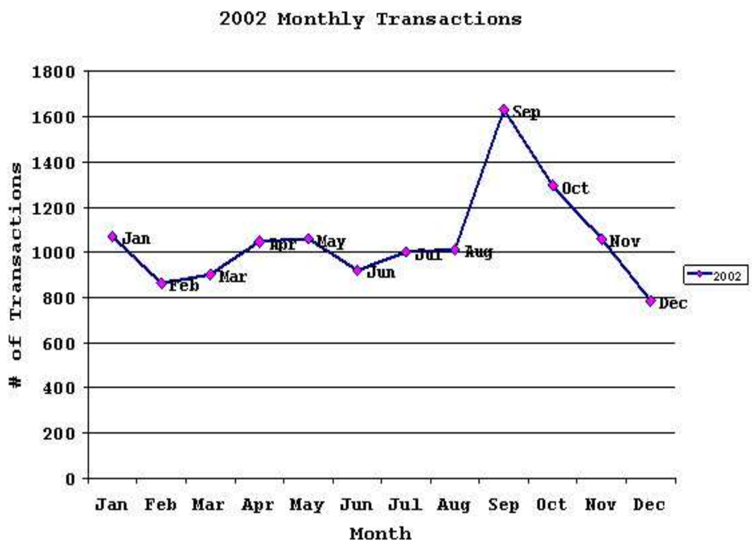

Figure 4.4 Monthly Transactions - Year 2002

transactions in September. This can be explained in terms of academic calender. For the first term, the peak is in September when students arrive for the regular academic session, while in second term, the peak is in January. Transaction volume is higher in

September compared to January because most new students resume at the beginning of the regular session while few students enroll in January.

DAILY TRANSACTIONS

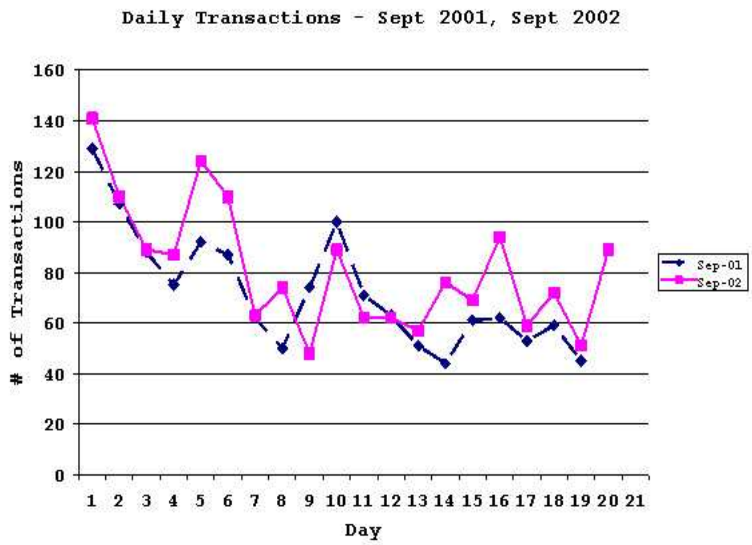

Figure 4.5 Comparison of Daily Transactions - Sept 2001 and Sept 2002

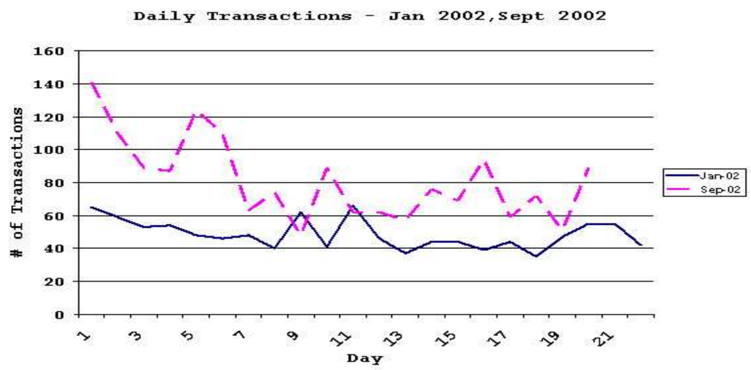

Figure 4.6 Comparison of Daily Transactions - Jan 2002 and Sept 2002

Observing the day to day activities in the help desk, the plot given in Figure 4.5 showing the daily transactions during September over the two years (2001 and 2002) reveals that there is a similar pattern. Therefore, the daily transactions at the help desk appear to be consistent over the months. Also, Figure 4.6 is the plot comparing daily transactions for the peak months, January and September of 2002. Again there is similar pattern, affirming the claim of consistent daily transactions at the help desk.

DAILY TRANSACTIONS ACCORDING TO DAY OF THE WEEK

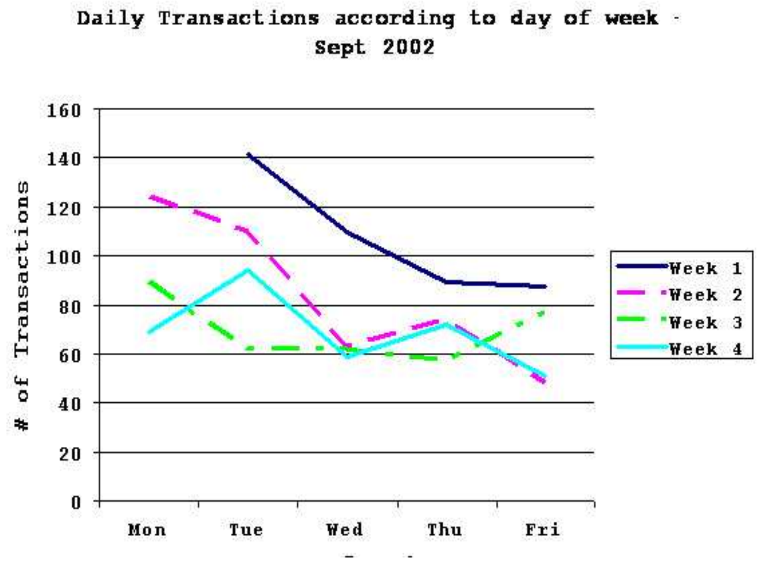

Figure 4.7 Daily Transactions According to Day of The Week - Sept 2002

It is important to be able to describe transaction load per day of the week. The plots comparing the transaction volume for days of the week (Monday, Tuesday,...) for the two peak months, September and January of 2002 are given in Figure 4.7 and Figure 4.8 respectively. From the plots, it can be observed that Monday/Tuesday is the busiest day of the week and the first week of the month is the busiest week. With these findings, it is obvious that the help desk handles higher number of transactions on Mondays and Tuesdays compared to other days of the week and experiences higher traffic during the first week of the month compared to other weeks.

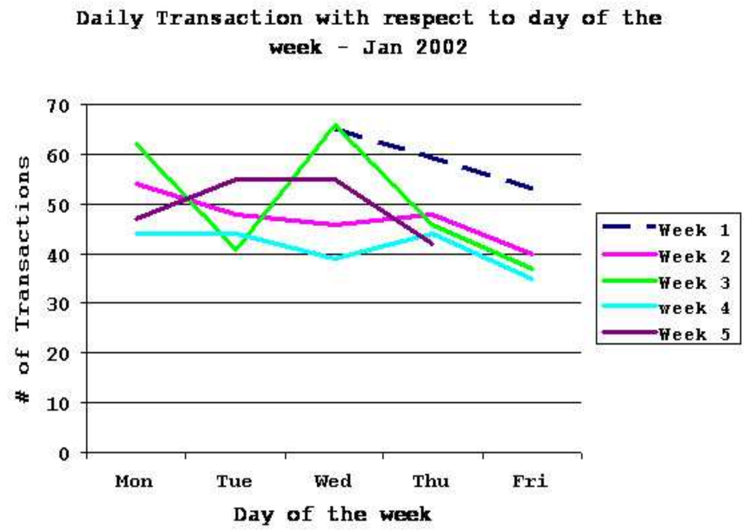

Figure 4.8 Daily Transactions According to Day of The Week - Jan 2002

HOURLY TRANSACTIONS

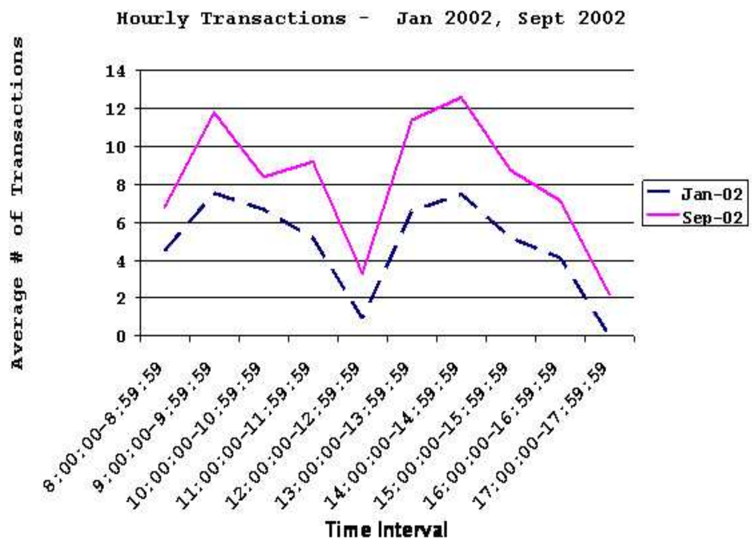

Figure 4.9 Comparison of Hourly Transactions - Jan 2002 and Sept 2002

Having considered yearly, monthly, weekly, and daily transactions of the help desk, it is desirable to look at transactions during the day. Hourly transactions handled at the help desk will be examined. To increase the sample size, the hourly transactions plotted were averaged over the working days of the months in consideration.

Figure 4.9 is the plot comparing the hourly transactions in January and September 2002. Categorizing the hourly transactions into morning and afternoon, it is interesting to know from the plot that the busiest time in the morning is between 9:00a.m and 10:00a.m while in the afternoon, the help desk handles higher volume of transactions between 2:00p.m and 3:00p.m. It is apparent from the plot that between 12:00noon and 1:00p.m there is a drastic drop in the number of transactions handled, this is due to lunch break which is usually during that time. Another drastic drop is obvious between 5:00p.m and 6:00p.m, this can be attributed to the normal closing hour which is 4:30p.m, calls/jobs arriving after 4:30p.m are to be handled next business day.

DESIGN OF HELP DESK SERVICE TIME

- Overall Service Time : The overall service time entails the service times of all the thirty-four job types.
- Service Times By Job Types : These are service times for four out of the thirty-four job types. They are jobs with leading number of transactions and they include the following.
  - Administrative Jobs
  - E-mail Jobs
  - Miscellaneous Jobs
  - Network Jobs

TRANSACTIONS BY JOB TYPE (Year 2002)

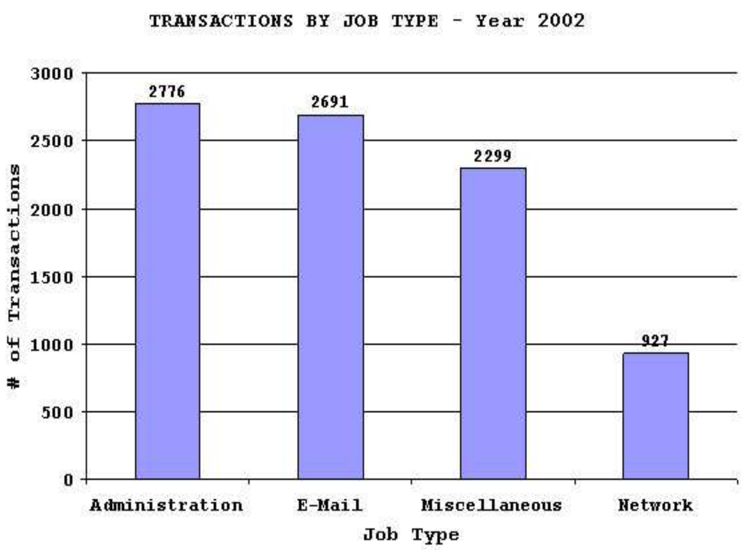

Figure 4.10 Transactions by Job Type - Year 2002

TRANSACTIONS BY JOB TYPE (Sept 2002)

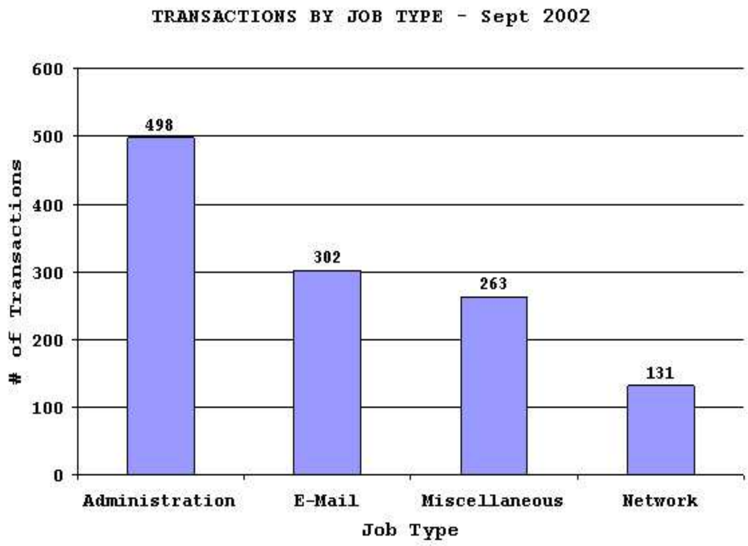

Figure 4.11 Transactions by Job Type - Sept 2002

Since this research work is basically on modelling service times in the help desk, transactions by job types will be described. Figure 4.10 provides the plot of transactions by job types for the year 2002 while Figure 4.11 gives the plot of transactions by job types for the month September of 2002. The plots reveal that the help desk handled 2,776 administrative jobs in the year 2002 out of which 498 transactions were handled in September 2002. 2,691 e-mail transactions were handled in year 2002 out of which 302 occurred in September 2002. Also 2,299 miscellaneous transactions were handled in the year 2002 of which 263 were recorded in September 2002. The help desk handled 927 network transactions in same year out of which 131 were handled in September 2002.

It is important to know how the servers handle each of the job types in the help desk. In order to increase the sample size, the service times were averaged over September 2001, September 2002 and September 2003.

SERVICE TIME - ADMINISTRATIVE JOB

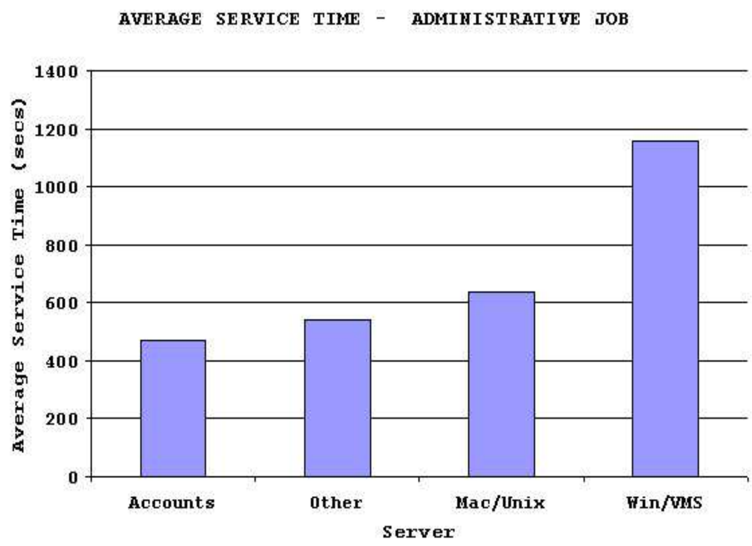

Figure 4.12 Service Time - Administrative Job

There are three major servers handling administrative jobs and they include, accounts, mac/unix and win/vms. Account servers spend lesser time in handling administrative job while win/vms servers spend more time in getting the job done. The server "other" consist of u-connect manager, technical services, SLA, quota, portal, operations, Mgmnt/CSCO, mailserv and housecalls servers respectively (Figure 4.12).

SERVICE TIME - E-MAIL JOB

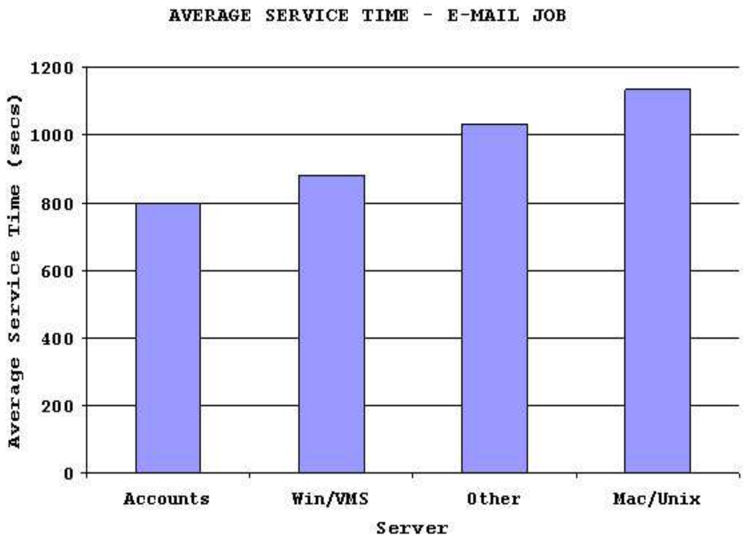

Figure 4.13 Service Time - E-mail Job

E-mail jobs are basically handled by accounts, win/vms and mac/unix servers respectively. Account servers are more efficient while mac/unix servers are slow in handling e-mail jobs. Other servers which occasionally handle e-mail jobs include SLA, quota, portal, operations, Mgmnt/CSCO and housecalls. The graph showing the average service times for these servers is displayed in Figure 4.13.

Miscellaneous jobs are mostly handled by accounts, mac/unix and win/vms servers. Other servers include technical services, SLA, quota, operations, Mgmnt/CSCO, IP REG and housecalls. Win/Vms servers should be least considered in handling miscellaneous jobs because they take more time to handle the jobs. Account servers are better in handling miscellaneous jobs (Figure 4.14).

SERVICE TIME - MISCELLANEOUS JOB

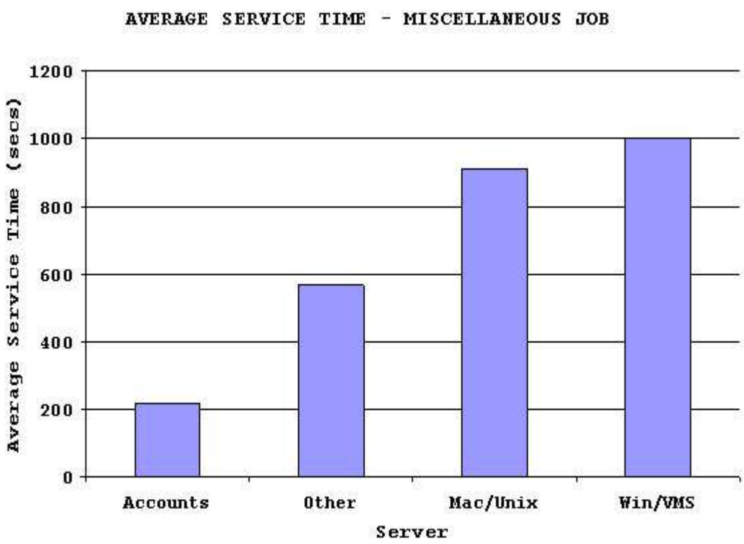

Figure 4.14 Service Time - Miscellaneous Job

Figure 4.15 is the graph of average service times for servers handling network jobs. IP REG is the most efficient of the servers, while win/vms servers take more time in handling network jobs. When all IP REG servers are busy, the next servers to be considered should be network services.

SERVICE TIME - NETWORK JOB

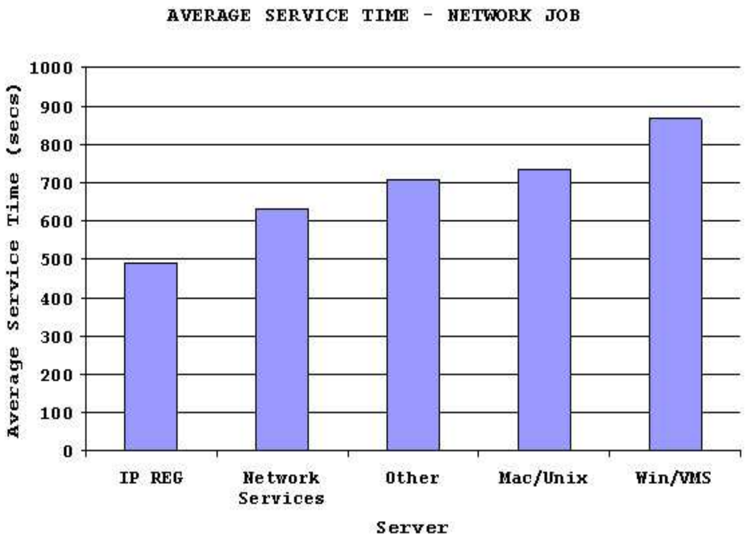

Figure 4.15 Service Time - Network Job

The descriptive information obtained so far clearly identifies yearly and monthly patterns. It is also clear that the month of September is the busiest when compared to other months. There is consistency in the daily transaction in the help desk, with Monday/Tuesday being the busiest day and the first week the busiest week of every month. In terms of hourly transactions, the help desk experiences higher traffic of transactions between the hours of 9:00a.m and 10:00a.m as well as between 2:00p.m and 3:00p.m daily.

Out of about thirty-four job types handled in the help desk, administrative, e-mail, miscellaneous and network jobs had leading number of transactions. Further analysis will be based on overall service time as well as service times for administrative, e-mail, miscellaneous and network jobs recorded for the month of September 2002.

Chapter 5
STATISTICAL MODELS FOR HELP DESK DATA

5.1 INTRODUCTION

Statistical models developed in this research work for various service times are presented in this chapter. In particular, models of exponential, log-normal and phase-type distributions for service times are investigated. In addition, phase-type distributions are used to approximate the fitted log-normal models.

5.2 EXPONENTIAL MODEL OF SERVICE TIMES

Conventionally, in queueing models, service times are assumed to be exponentially distributed, this is due to the memoryless property of the exponential distribution, which makes the analysis of the model, $M|M|N|S$, easier. The relevance of this assumption was examined by fitting exponential distributions to the overall service times as well as to service times by job types namely, administration, e-mail, miscellaneous and network jobs of the help desk.

SERVICE TIME - OVERALL

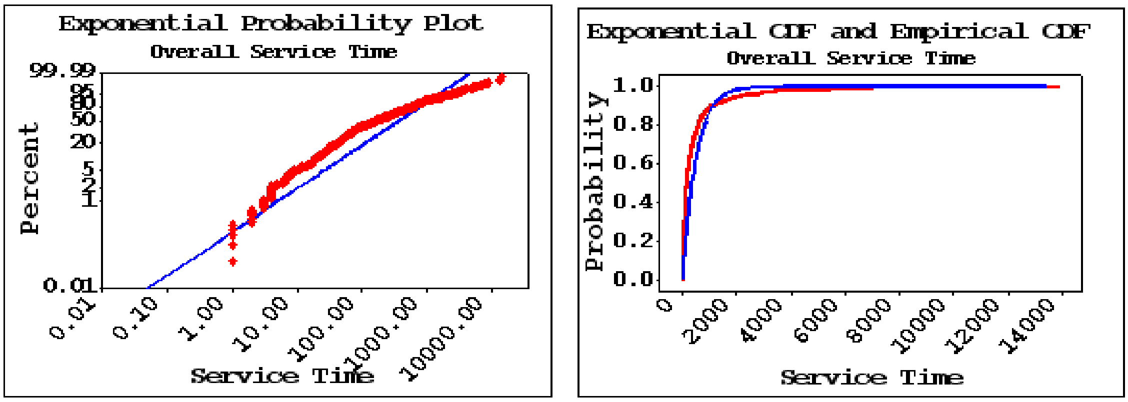

Figure 5.1 Comparison of Exponential Model and Empirical Data - Overall Service Time

SERVICE TIME - ADMINISTRATIVE JOB

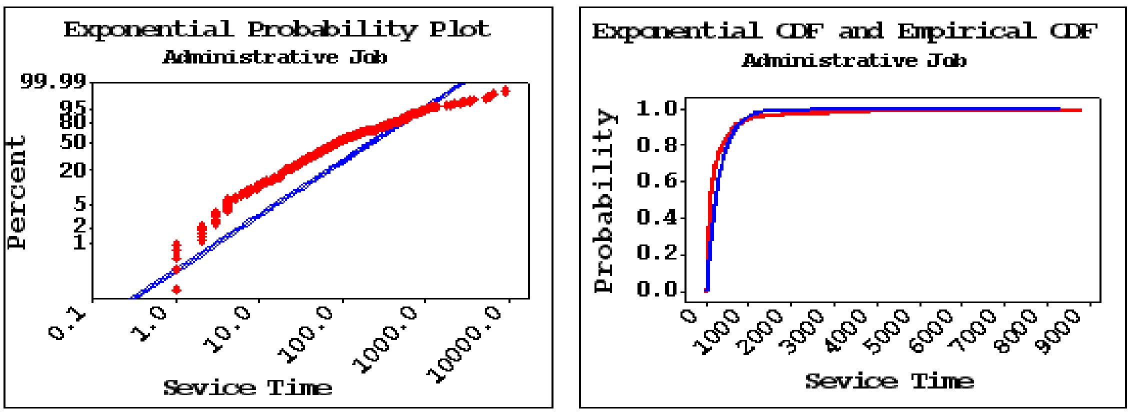

Figure 5.2 Comparison of Exponential Model and Empirical Data - Administrative Jobs

SERVICE TIME - EMAIL JOB

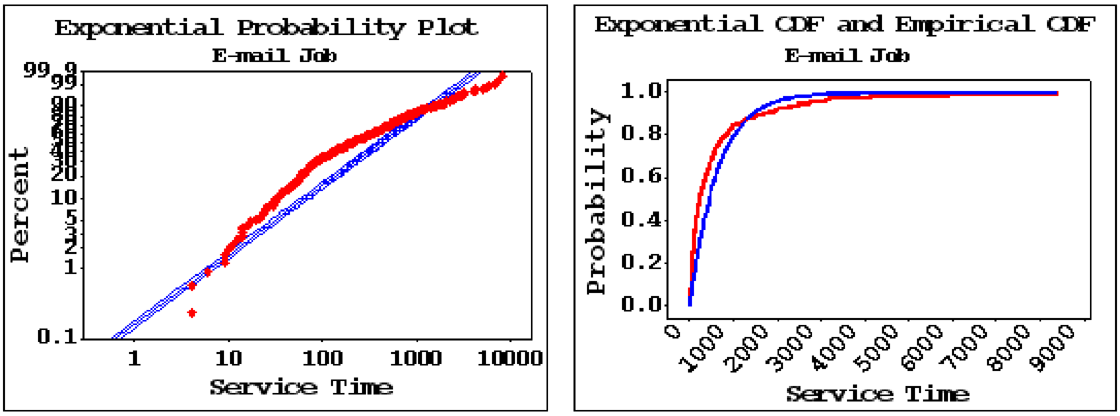

Figure 5.3 Comparison of Exponential Model and Empirical Data - E-mail Jobs

SERVICE TIME - MISCELLANEOUS JOB

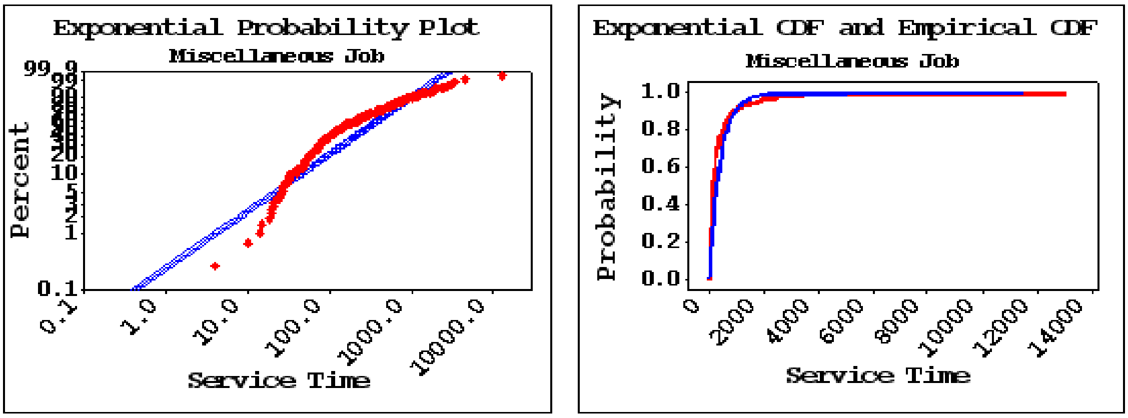

Figure 5.4 Comparison of Exponential Model and Empirical Data - Miscellaneous Jobs

SERVICE TIME - NETWORK JOB

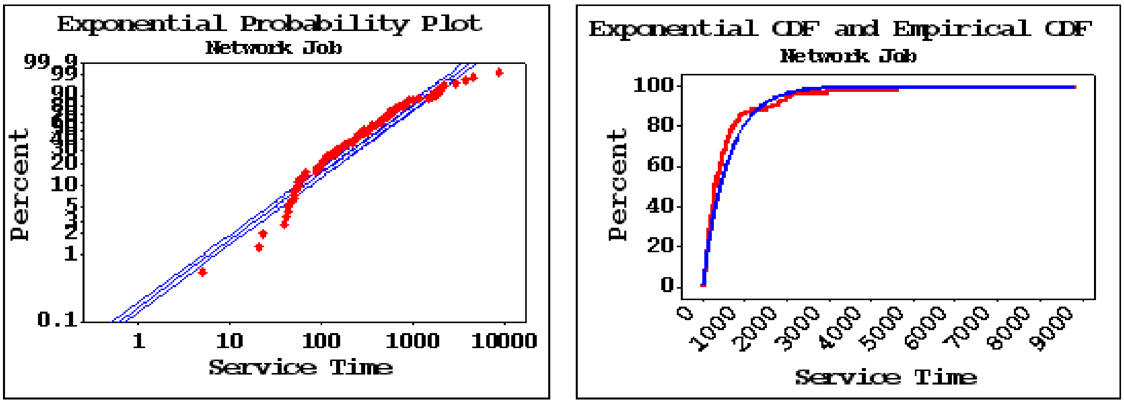

Figure 5.5 Comparison of Exponential Model and Empirical Data - Network Jobs

A probability plot was used to determine if the overall service times and the service times by job types are from exponential distributions. In figures 5.1 - 5.5, the blue line on each of the probability plots is the cumulative distribution function for the chosen theoretical distribution (exponential distribution) with the indicated parameters. The y-values and in some cases the x-values are transformed so that the fitted line is linear. The curved blue lines about the blue straight line is the 95% confidence interval for the fitted distribution. In some of the plots, the confidence bounds are very close to the fitted distribution and they are almost indistinguishable. The red points are created by plotting the value of each observation against its estimated cumulative probability. The y-scale can be percent, probability or score but the y-scale used is the percent. If the service times matches the exponential distribution, the points (red points) should cluster around the blue straight line. The further the points (red points) deviate from this line, the greater the indication of departure from the exponential distribution.

The cumulative distribution function of the empirical data and the fitted exponential model are also compared. If the service time is from an exponential distribution, then the plot of the cumulative distribution function of the empirical data should coincide or almost coincide with the cumulative distribution function of the fitted exponential distribution. Figure 5.1 gives the p-p plot and the comparison of the CDFS for overall service time. Looking at the p-p plot, there is a significant departure from the blue reference line indicating that the overall service time is not exponentially distributed.

Figures 5.2 to 5.5 provide the probability plots and the cumulative distribution functions for the service times of administrative jobs, e-mail jobs, miscellaneous jobs and network jobs respectively. For each of these probability plots, there is a high deviation of the points from the blue reference line. This is an indication that service times cannot be modelled as exponential distribution. However, it is important to have a statistical confirmation, this is done by implementing the Kolmogorov-Smirnov goodness of fit test. The result of the test is presented in Table 5.1.

RESULT OF KOLMOGOROV SMIRNOV (K-S) STATISTICAL TEST - EXPONENTIAL MODEL

H_0 = Service time is exponential.

H_1 = Service time is not exponential

Table 5.1 Result of K-S test - Exponential Model

	Overall	Administration	E-mail	Miscellaneous	Network
Level of Sig	0.05	0.05	0.05	0.05	0.05
KS Statistic(D)	0.2476	0.2876	0.2163	0.2342	0.1274
CV	0.0335	0.0605	0.0776	0.0831	0.1173
Decision	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

The K-S statistics are greater than the critical values (CV) for the overall service times and service times of the four job types. Therefore the null hypothesis H_0 is rejected in all cases. This result confirms the findings by the graphical approach that the service times are not exponentially distributed for the help desk data used in this study.

5.3 LOG-NORMAL DISTRIBUTION OF SERVICE TIME

With the confirmation of the irrelevance of the assumption of exponential service time to this help desk data, as suggested in [23], the log-normal distribution was fitted to the overall service times and service times of the four major job types (administration, e-mail, miscellaneous and network). This was done using the probability plot (p-p) and comparing the cumulative distribution functions of the empirical data and fitted log-normal model. The Kolmogorov-Smirnov goodness of fit test was used to assess the log-normal models.

SERVICE TIME - OVERALL

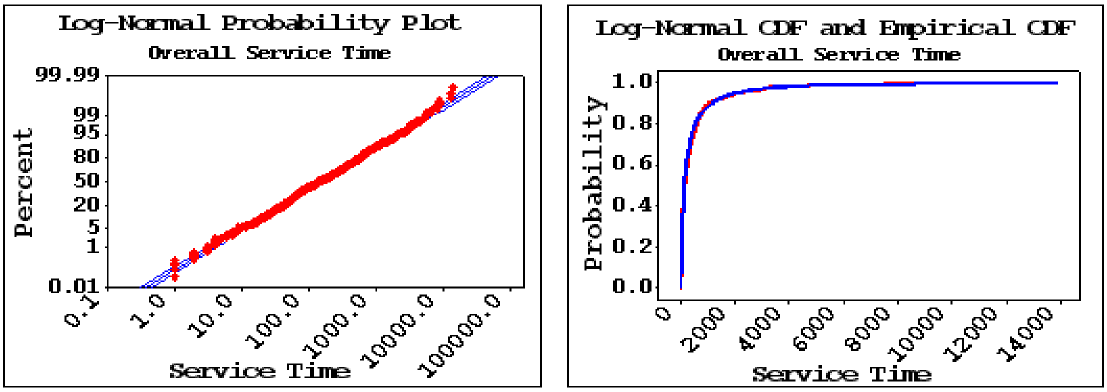

Figure 5.6 Comparison of Log-Normal Model and Empirical Data - Overall Service Time

SERVICE TIME - ADMINISTRATIVE JOB

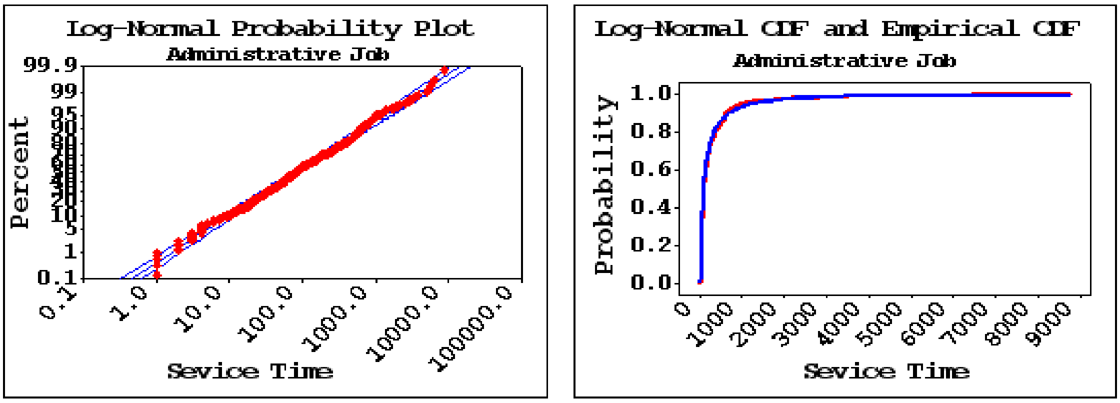

Figure 5.7 Comparison of Log-Normal Model and Empirical Data - Administrative Jobs

SERVICE TIME - EMAIL JOB

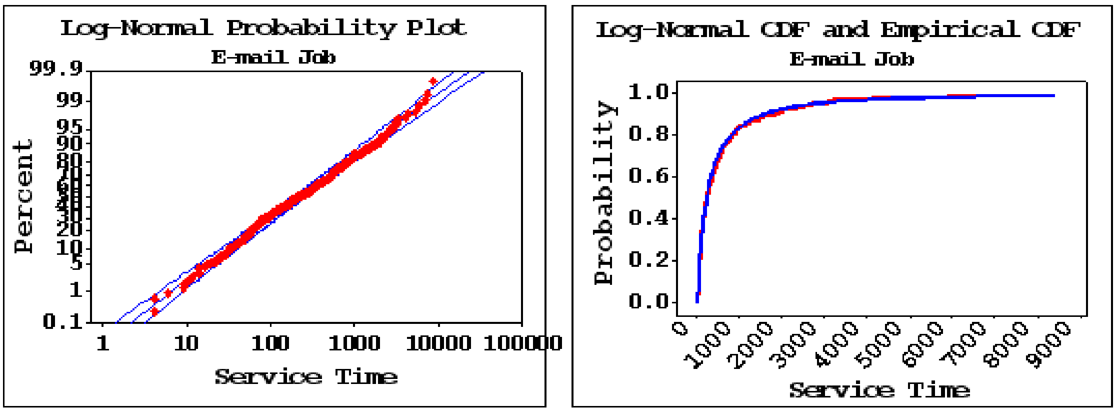

Figure 5.8 Comparison of Log-Normal Model and Empirical Data - E-mail Jobs

SERVICE TIME - MISCELLANEOUS JOB

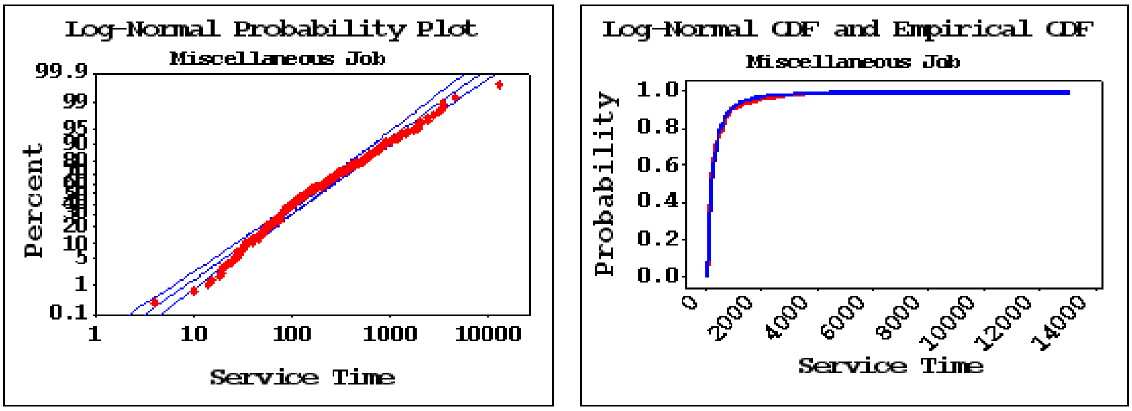

Figure 5.9 Comparison of Log-Normal Model and Empirical Data - Miscellaneous Jobs

SERVICE TIME - NETWORK JOB

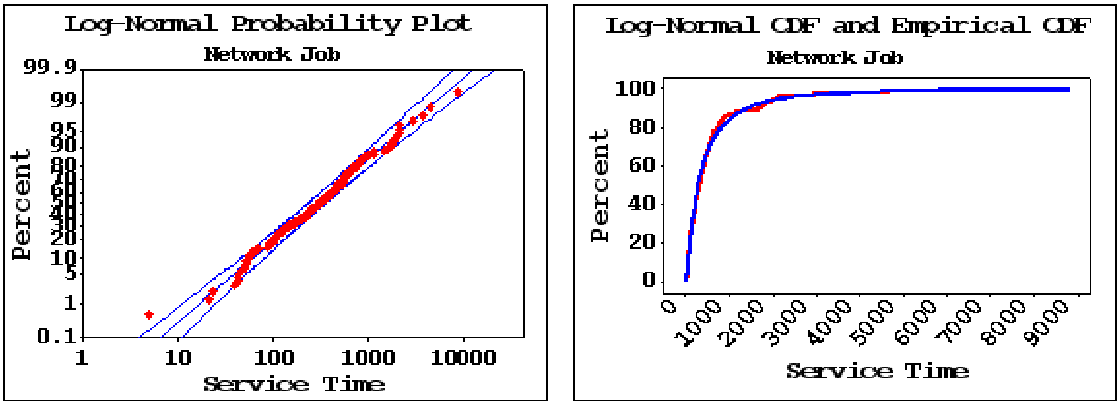

Figure 5.10 Comparison of Log-Normal Model and Empirical Data - Network Jobs

From figures 5.6 - 5.10, unlike the case of exponential models for service times, the points (red) are clustered around the blue reference line on the probability plots for the overall service times and service times of each of the job types. The p-p plots suggest that the service time data are close to what would be expected from the log-normal distribution, so the service times can be modelled using the log-normal distribution. Also, the cumulative distribution functions in Figures 5.6 - 5.10 for the fitted log-normal models and empirical data strongly coincide in all the plots. The log-normal models are further assessed by implementing the Kolmogorov-Smirnov goodness of fit test.

RESULT OF KOLMOGOROV SMIRNOV (K-S) STATISTICAL TEST - LOG-NORMAL MODEL

H_0 = Service time is log-normal.

H_1 = Service time is not log-normal

Table 5.2 Result of K-S test - log-normal model

	Overall	Administration	E-mail	Miscellaneous	Network
Level of Sig	0.05	0.05	0.05	0.05	0.05
KS Statistic(D)	0.0217	0.0275	0.0387	0.0679	0.0430
CV	0.0335	0.0605	0.0776	0.0831	0.1173

Decision: Do not reject H_0.

At 0.05 level of significance, the null hypothesis H_0 was not rejected in all cases, indicating that the log-normal distribution provides a reasonable fit for the overall service times and service times by job types of the help desk. This result affirms the conclusion, based on a visual approach, that log-normal distribution gives a reasonable fit. With this finding, in the $G|G|N|S$ model, the service process can be modelled as log-normal. Using the maximum likelihood technique in Chapter 3, the maximum likelihood estimates of parameters of log-normal models for the service times in the help desk are as follows:

- Overall Service Time - Logn(4.973,1.604)
- Administrative Job - Logn(4.392,1.681)

- E-mail Job - Logn(5.403,1.508)
- Miscellaneous Job - Logn(5.104,1.275)
- Network Job - Logn(5.660,1.228)

The drawback of the log-normal service time is that the model $G|G|N|S$ becomes analytically intractable and it is difficult to derive the desired performance measures of this queueing model. Also some services may involve many tasks or stages before completion and it may be interesting to model each stage of the service. For instance, the opening of an account in a bank may involve one person collecting the information, another entering the information into the database and another creating the account. Services involving stages motivate the use of phase-type distribution to model each of the stage/tasks as phases in the phase-type distribution. In the next section, phase-type distributions will be fitted to service times of the help desk.

5.4 PHASE-TYPE DISTRIBUTION OF SERVICE TIME

In this section, phase-type distributions of general structure fitted to overall service times and service times by job types are presented. Generally, in fitting phase-type distribution to empirical data, the order (phases) can be an arbitrary number but it has to be noted that as the number of phases increases, the state space becomes larger and hence the the model becomes complicated. For the data set from the help desk, phase-type distributions with phases $p = 2, 3$ or 4 provide reasonable fits, therefore general phase-type fits with order(phases) larger than 4 will not be considered. Phase-type distribution with phases $p = 2, 3$ and 4 will be denoted by PH(2), PH(3) and PH(4) respectively. The Kolmogorov-Smirnov and Anderson-Darling statistical tests were used in assessing the models and in model selection. Model selection, (i.e., choosing the appropriate number of phases) will be based on the A-D test rather than the K-S test since the former is more powerful than the later. The smallest phase for which A-D test does not reject the null hypothesis will be used.

SERVICE TIME - OVERALL

The survival functions and the distribution functions of the empirical data(service times) and the phase-type fit of order $p = 2, 3, 4$ were compared. The survival function measures the probability that a customer will be in service up to time t. Figure 5.11 presents the comparison of the survival functions and the distribution functions of the fitted PH(2), PH(3), PH(4) and the empirical data.

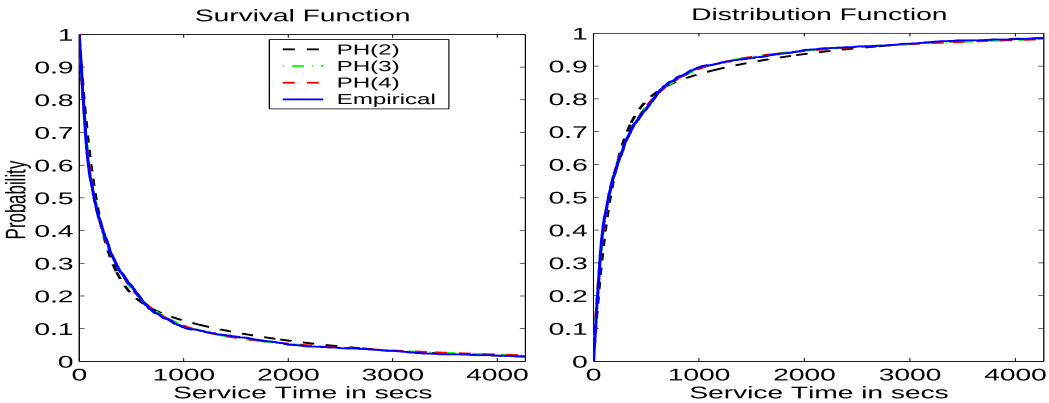

Figure 5.11 Comparison of fitted PH(2), PH(3) and PH(4) model - overall service time.

The survival functions of the empirical data and phase-type fit of order $p = 3$ and $p = 4$ almost coincide when compared to phase-type fit of order $p = 2$. The cumulative distribution functions of the empirical data and phase-type fit of order $p = 3$ and $p = 4$ are very close when compared to phase-type fit of order $p = 2$. From this graphical approach, phase-type distributions of order $p = 3$ and $p = 4$ provide a better description of overall service times of the help desk. For model selection and assessment of models, the K-S and A-D tests were implemented. (Table 5.3).

H_0 = Service time follows Phase-type distribution of order p

H_1 = Service time do not follow Phase-type distribution of order p.

Table 5.3 Result of K-S and A-D tests for fitted Phase-type distribution - Overall Service Times

	PH(2)	PH(3)	PH(4)
D^+	2.5230e-005	2.3330e-004	2.1125e-004
D^-	5.8752e-004	3.7945e-004	4.0149e-004
D	5.8752e-004	3.7945e-004	4.0149e-004
K-S Statistic(D_n)	0.0238	0.0154	0.0163
A-D Statistic(A^2)	8.6865	1.4881	1.3218

The K-S and A-D statistics were compared with their corresponding critical values at various levels of significance in Table 5.4.

Table 5.4 Critical Values (CV) For K-S and A-D Tests

Level of Significance	0.15	0.10	0.05	0.025	0.01
K-S (CV)	1.138	1.224	1.358	1.480	1.628
A-D (CV)	1.610	1.933	2.492	3.070	3.857

According to results of K-S test, at 15%, 10%, 5%, 2.5% and 1% levels of significance, the critical values (CV)(Table 5.4) are greater than the test statistics for PH(2), PH(3) and PH(4)(Table 5.3). Therefore PH(2), PH(3) or PH(4) can be used to model the overall service time. From A-D test, at 15% level of significance, PH(2) is rejected while PH(3) and PH(4) are accepted at all selected levels of significance. Using the criteria for model selection, since PH(3) is the fit with the smallest number of phases that was accepted at all selected levels of significance with preference given to A-D test then, **PH(3)** will be selected as the most appropriate model that describes the overall service times.

In Table 5.5, descriptive statistics such as the mean, standard deviation(Stdev), coefficient of variation(C.V) and the log-likelihood functions (Log-L) are presented for the phase-type distributions with order $2, 3$, and 4.

Table 5.5 Descriptive Statistics For Fitted Models - Overall Service Times

	PH(2)	PH(3)	PH(4)
Mean	484.7292	484.7294	484.7292
Stdev	946.6724	1064.6	1059.6
C.V	1.95	2.20	2.19
Log-L	-11248.709346	-11207.098476	-11205.111469

It can be noted from Table 5.5, that the log-likelihood increases as more phases are added and there is a high probability of getting a better fit as the number of phases increases. But as the number of phases increases, the model becomes more complex.

Having observed that **PH(3)** fits best for the overall service times, its parameters are given as follows:

- The probability of starting in states $[1, 2, 3]$:

$$q = \begin{pmatrix} 0.001 & 0.007 & 0.992 \end{pmatrix}$$

- The Phase Generator Matrix:

$$R = \begin{pmatrix} -0.000583 & 0.000542 & 0.000012 \\ 0.000651 & -0.005392 & 0.004728 \\ 0.000033 & 0.004958 & -0.011752 \end{pmatrix}$$

- The Transition Probability Matrix of the Embedded Markov chain can be calculated as

$$\hat{P} = \begin{pmatrix} 0 & 0.93 & 0.02 \\ 0.12 & 0 & 0.88 \\ 0.003 & 0.42 & 0 \end{pmatrix}$$

- The Absorption probability Vector is:

$$\hat{P}_{k\triangle} = \begin{pmatrix} 0.05 & 0.003 & 0.58 \end{pmatrix}$$

- The expected length of time spent in states [1,2,3] in seconds:

$$\hat{m}_k = \begin{pmatrix} 1715.266 & 185.460 & 85.092 \end{pmatrix}$$

Figure 5.12 provides a pictoral view of the structure behind the phase-type distributions of order $2, 3$ and 4 fitted to the overall service times.

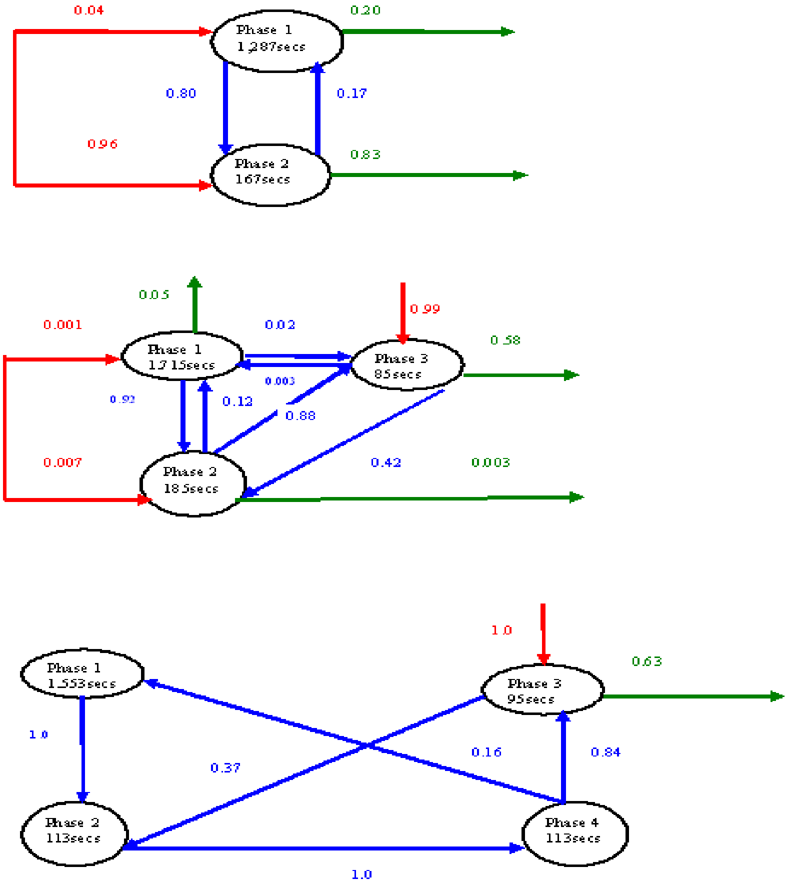

Figure 5.12 PH Structures of Order $p = 2, 3, 4$ - Overall Service Times

Looking at the structure of the selected PH(3) model, less time(85secs) is being spent in phase 3 of the service, while it takes more time(1715secs) in phase 1. It is almost certain that a job will start at phase 3 (prob= 0.99), while the chance of service starting at phase 1 or phase 2 is negligible. As expected, the probability of a job being completed and exiting the system from phase 3 is high(0.58) compared to the negligible possibility of exiting from phase 1(0.05) or phase 2(0.003).

PH MODEL FOR SERVICE TIME BY JOB TYPES

Phase-type distributions of order $p = 2, 3$ and 4 were fitted to service times by job types (administration, e-mail, miscellaneous and network). The survival functions and the cumulative distribution functions were plotted for empirical data(blue), PH(2)(black) , PH(3)(green) and PH(4)(red) for service times of each of the job type. Figure 5.13 provides a comparison of survival functions and cumulative distribution functions of PH(2), PH(3), PH(4) and empirical data for service times of administrative jobs. Similarly, Figure 5.14 gives the comparison of the survival and distribution functions of PH(2), PH(3), PH(4) and empirical of miscellaneous jobs' service times.

For service times of administrative jobs, the survival functions plots for PH(3) and PH(4) almost coincide with the survival function of the empirical data. The cumulative distribution functions curves of PH(3) and PH(4) almost coincide and are closer to the empirical data when compared to PH(2). This suggests that PH(3) or PH(4) gives a better description of service times of administrative jobs when compared to PH(2). In the case of service times of miscellaneous jobs, the plots of the survival functions of PH(2) and PH(4) almost coincide. The plots of the cumulative distribution functions of PH(2), PH(3) and PH(4) are close to the cumulative distribution function of the empirical data. Therefore, PH(2), PH(3) or PH(4) can model service times of miscellaneous jobs. Figure 5.14 presents the plots of survival functions and distribution functions of PH(2), PH(3), PH(4) and empirical data of service times of e-mail jobs. Figure 5.16 provides the same comparison for service times of network jobs. The survival functions of PH(2), PH(3) and PH(4) are almost indistinguishable and are close to the survival function of the empirical data for service times of e-mail and network jobs respectively. In terms of cumulative distribution functions, PH(2), PH(3) and PH(4) are very close to the empirical data of e-mail and network jobs respectively. Using this graphical approach, it is difficult to say which of the models best describe service times of e-mail jobs and network jobs. Therefore, the K-S test and the A-D test are implemented.

SERVICE TIME - ADMINISTRATIVE JOB

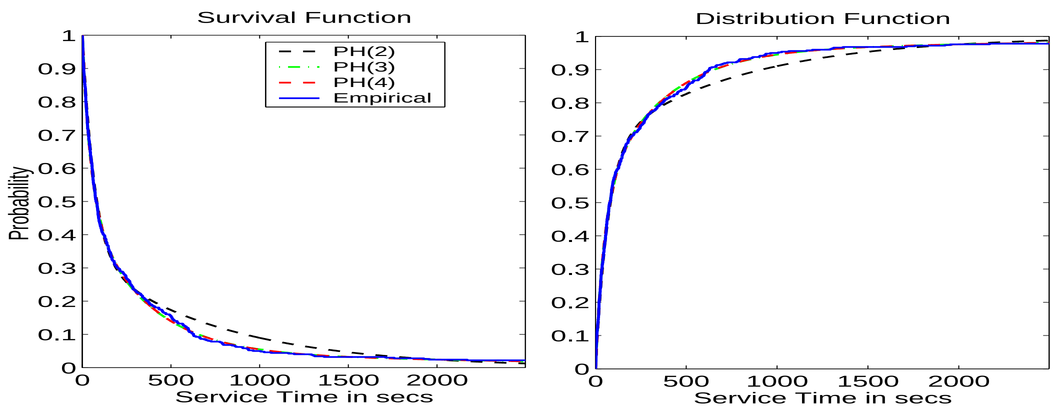

Figure 5.13 Comparison of fitted PH(2), PH(3) and PH(4) model - Administrative Jobs.

SERVICE TIME - MISCELLANEOUS JOB

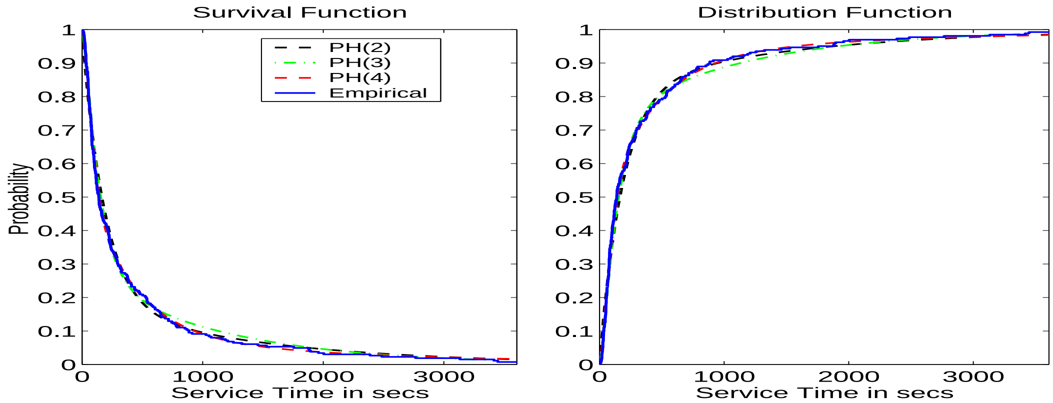

Figure 5.14 Comparison of fitted PH(2), PH(3) and PH(4) model - Miscellaneous Jobs.

SERVICE TIME - E-MAIL JOB

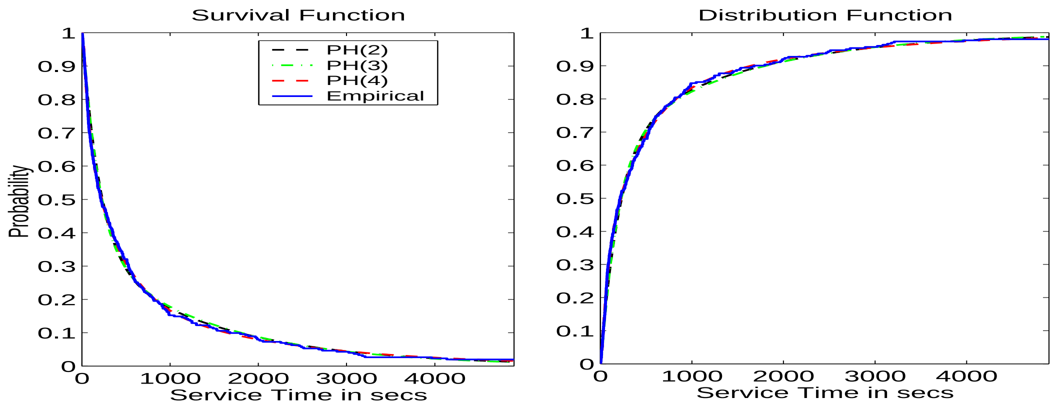

Figure 5.15 Comparison of fitted PH(2), PH(3) and PH(4) model - E-mail Jobs.

SERVICE TIME - NETWORK JOB

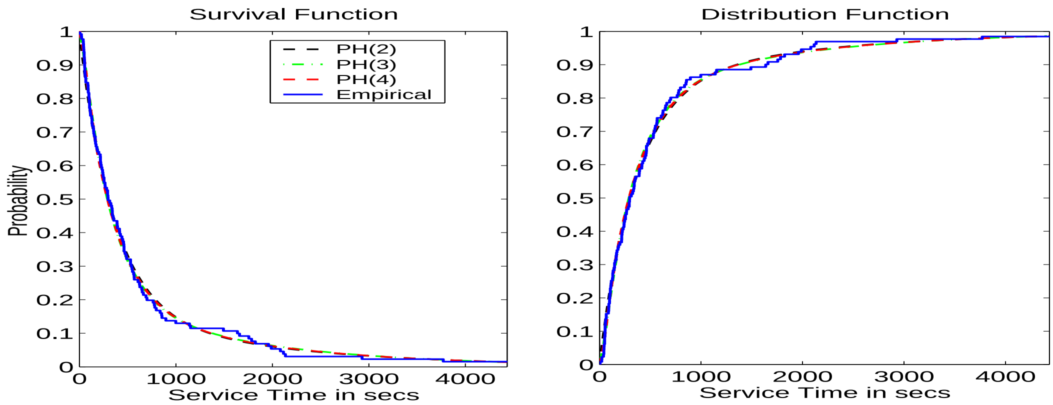

Figure 5.16 Comparison of fitted PH(2), PH(3) and PH(4) model - Network Jobs.

MODEL SELECTION

In order to select phase-type model for service times by job types, the K-S and A-D tests will be implemented. For each of the job types, the K-S and A-D test statistics were compared with the corresponding critical values at various levels of significance. First, phase-type distributions for service times of administrative and miscellaneous jobs will be discussed.

Table 5.6 Critical Values (CV) For K-S and A-D Tests

Level of Significance	0.15	0.10	0.05	0.025	0.01
K-S (CV)	1.138	1.224	1.358	1.480	1.628
A-D (CV)	1.610	1.933	2.492	3.070	3.857

SERVICE TIME - ADMINISTRATIVE JOB

Table 5.7 K-S and A-D Test Statistics - Administrative Jobs

	PH(2)	PH(3)	PH(4)
D^+	3.4667e-006	0.0015	0.0015
D^-	0.0020	4.7680e-004	4.9135e-004
D	0.0020	0.0015	0.0015
KS Statistic(D_n)	0.0450	0.0344	0.0340
A-D Statistic(A^2)	1.6171	0.2905	0.2800

SERVICE TIME - MISCELLANEOUS JOB

Table 5.8 K-S and A-D Test Statistics - Miscellaneous Jobs

	PH(2)	PH(3)	PH(4)
D^+	2.0800e-005	2.5950e-006	2.1403E-004
D^-	0.0038	0.0038	0.0036
D	0.0038	0.0038	0.0036
KS Statistic(D_n)	0.0618	0.0621	0.0586
A-D Statistic(A^2)	2.2499	0.7317	0.2235

For service times of administrative jobs, comparing the test statistics in Table 5.7 with the corresponding critical values in Table 5.6, K-S test accepts PH(2), PH(3) and PH(4) at all selected levels of significance while A-D test rejects PH(2) at 15% level of significance but was accepted at other levels of significance. **PH(3)** and **PH(4)** are accepted at all selected levels of significance.

For service times of miscellaneous jobs,comparison of test statistics in Table 5.8 with critical values in Table 5.6, reveals that K-S test accepts PH(2), PH(3) and PH(4) at all selected levels of significance. Whereas A-D test accepts PH(2) at 2.5% and 1% levels of significance and rejects it at other levels of significance, it accepts PH(3) and PH(4) at all selected levels of significance. Since **PH(3)** is the model with least number of phases that was accepted by A-D test at all selected levels of significance, therefore, **PH(3)** is selected as the appropriate phase-type model for service times of administrative and miscellaneous jobs. Table 5.9 and Table 5.10 give the descriptive statistics for the phase-type models considered.

SERVICE TIME- ADMINISTRATIVE JOB

Table 5.9 Descriptive Statistics For Fitted Models - Administrative Jobs

	PH(2)	PH(3)	PH(4)
Mean	299.5402	299.5403	299.5399
Stdev	550.6081	802.3306	801.1823
C.V	1.84	2.68	2.67
Log-L	-3172.166257	-3147.651258	-3147.279667

SERVICE TIME- MISCELLANEOUS JOB

Table 5.10 Descriptive Statistics For Fitted Models - Miscellaneous Jobs

	PH(2)	PH(3)	PH(4)
Mean	415.3041	415.3043	415.3042
Stdev	801.1881	740.1659	885.9140
C.V	1.93	1.78	2.13
Log-L	-1791.688325	-1781.764748	-1774.733724

From Table 5.9 and Table 5.10 it can be seen that the log-likelihood function increases as the number of phases increases. The estimated parameters for the selected models are given as follows:

SERVICE TIME- ADMINISTRATIVE JOB

For the selected PH(3) model for service times of administrative jobs, the estimated parameters are given as follows:

- The probability of starting in states $[1, 2, 3]$:

$$q = \begin{pmatrix} 0.964 & 0.033 & 0.003 \end{pmatrix}$$

- Phase Generator Matrix:

$$R = \begin{pmatrix} -0.018882 & 0.006775 & 0.000039 \\ 0.004350 & -0.004896 & 0.000287 \\ 0.000063 & 0.000359 & -0.000441 \end{pmatrix}$$

- The Transition Probability Matrix is computed as:

$$\hat{P} = \begin{pmatrix} 0 & 0.36 & 0.002 \\ 0.89 & 0 & 0.06 \\ 0.14 & 0.81 & 0 \end{pmatrix}$$

- The Absorption probability Vector:

$$\hat{P}_{k\triangle} = \begin{pmatrix} 0.64 & 0.05 & 0.04 \end{pmatrix}$$

- Expected time spent in states $[1, 2, 3]$ in seconds:

$$\hat{m}_{k\triangle} = \begin{pmatrix} 53 & 204 & 2268 \end{pmatrix}$$

SERVICE TIME- MISCELLANEOUS JOB

Having observed that PH(3) gives the most appropriate model for service times of miscellaneous jobs, the estimated parameters are given as follows:

- The probability of starting in states $[1, 2, 3]$:

$$q = \begin{pmatrix} 0.0006 & 0.0 & 0.999 \end{pmatrix}$$

- Phase Generator Matrix:

$$R = \begin{pmatrix} -0.001164 & 0.001162 & 0.000002 \\ 0.011789 & -0.050636 & 0.000013 \\ 0.000000 & 0.007805 & -0.007806 \end{pmatrix}$$

- The Transition Probability Matrix is computed as:

$$\hat{P} = \begin{pmatrix} 0 & 0.998 & 0.002 \\ 0.23 & 0 & 0 \\ 0 & 0.999 & 0 \end{pmatrix}$$

- Exit Probability is calculated as;

$$\hat{P}_{k\triangle} = \begin{pmatrix} 0 & 0.77 & 0 \end{pmatrix}$$

- Expected time spent in states $[1, 2, 3]$ in seconds:

$$\hat{m}_{k\triangle} = \begin{pmatrix} 859 & 208 & 128 \end{pmatrix}$$

Figure 5.17 and figure 5.18 are the structural representations of PH(3) models for service times of administrative and miscellaneous jobs respectively.

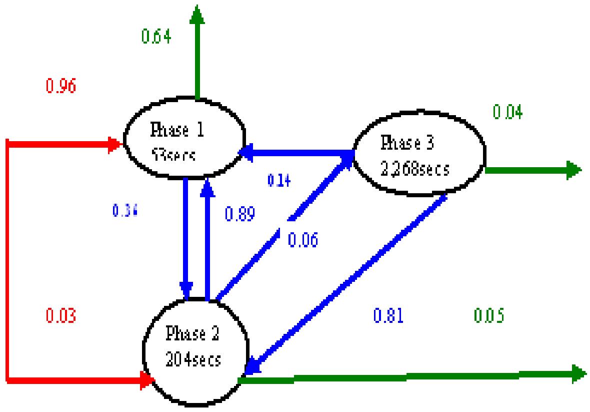

Figure 5.17 Structure of PH(3) model - Administrative Jobs.

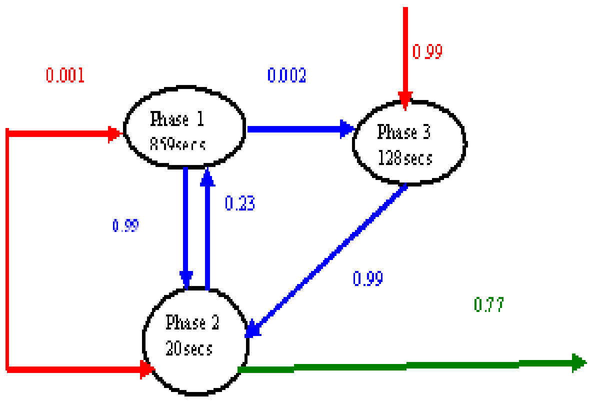

Figure 5.18 Structure of PH(3) model - Miscellaneous Jobs

Phase-type models for service times of e-mail jobs and network jobs are discussed as follows.

SERVICE TIME - E-MAIL JOB

Table 5.11 K-S and A-D Test Statistics - E-mail Jobs

	PH(2)	PH(3)	PH(4)
D^+	0.0013	0.0011	0.0022
D^-	0.0020	0.0022	0.0011
D	0.0020	0.0022	0.0022
KS Statistic(D_n)	0.0356	0.0392	0.0392
A-D Statistic(A^2)	0.9530	0.9099	0.1122

SERVICE TIME - NETWORK JOB

Table 5.12 K-S and A-D Test Statistics - Network Jobs

	PH(2)	PH(3)	PH(4)
D^+	0.0012	9.1921e-004	0.0010
D^-	0.0065	0.0067	0.0066
D	0.0065	0.0067	0.0066
KS Statistic(D_n)	0.0748	0.0777	0.0767
A-D Statistic(A^2)	0.5960	0.2761	0.2541

Considering service times of e-mail jobs, the test statistics in Table 5.11 are compared with the corresponding critical values in Table 5.6. Both the K-S test and the A-D test accept PH(2), PH(3) and PH(4) at all selected levels of significance. Similarly, for service times of network jobs, comparison of the test statistics in Table 5.12 with critical values in Table 5.6 reveals that at all selected levels of significance, K-S test and A-D test accept PH(2), PH(3) and PH(4). Since **PH(2)** is the phase-type model with least number of phases that was accepted by A-D test, therefore **PH(2)** is selected as the appropriate model for service times of e-mail jobs and network jobs. Table 5.13 and Table 5.14 give the descriptive statistics for the phase-type models considered for service times of e-mail jobs and network jobs respectively.

SERVICE TIME- E-MAIL JOB

Table 5.13 Descriptive Statistics For Fitted Models - E-mail Jobs

	PH(2)	PH(3)	PH(4)
Mean	631.0365	631.0376	631.0363
Stdev	1070.6	1053.8	1126.2
C.V	1.70	1.67	1.78
Log-L	-2188.901152	-2186.245633	-2180.360539

SERVICE TIME- NETWORK JOB

Table 5.14 Descriptive Statistics For Fitted Models - Network Jobs

	PH(2)	PH(3)	PH(4)
Mean	599.1754	599.1756	599.1752
Stdev	966.7234	952.2709	957.6118
C.V	1.61	1.59	1.60
Log-L	-956.274215	-953.425937	-953.219774

The estimated parameters for the selected PH(2) models for service times of e-mail jobs and network jobs are presented as follows.

SERVICE TIME- E-MAIL JOB

The estimated parameters for the selected PH(2) model for service times of e-mail jobs are given as follows.

- The probability of starting in states $[1, 2, 3]$:

$$q = \begin{pmatrix} 0.005 & 0.995 \end{pmatrix}$$

- Phase Generator Matrix:

$$R = \begin{pmatrix} -0.000908 & 0.000892 \\ 0.001085 & -0.004579 \end{pmatrix}$$

- The Transition Probability Matrix is computed as:

$$\hat{P} = \begin{pmatrix} 0 & 0.98 \\ 0.24 & 0 \end{pmatrix}$$

- Exit Probability is calculated as;

$$\hat{P}_{k\triangle} = \begin{pmatrix} 0.02 & 0.76 \end{pmatrix}$$

- Expected time spent in states $[1, 2, 3]$ in seconds:

$$\hat{m}_{k\triangle} = \begin{pmatrix} 1101 & 218 \end{pmatrix}$$

SERVICE TIME- NETWORK JOB

The estimated parameters for the selected PH(2) model for service times of network jobs are given as follows.

- The probability of starting in states $[1, 2, 3]$:

$$q = \begin{pmatrix} 0.055 & 0.945 \end{pmatrix}$$

- Phase Generator Matrix:

$$R = \begin{pmatrix} -0.000609 & 0.000310 \\ 0.000257 & -0.002844 \end{pmatrix}$$

- The Transition Probability Matrix:

$$\hat{P} = \begin{pmatrix} 0 & 0.51 \\ 0.09 & 0 \end{pmatrix}$$

- Exit Probability is calculated as;

$$\hat{P}_{k\triangle} = \begin{pmatrix} 0.49 & 0.91 \end{pmatrix}$$

- Expected time spent in states [1, 2, 3] in seconds:

$$\hat{m}_{k\triangle} = \begin{pmatrix} 1642 & 352 \end{pmatrix}$$

Structural representations of PH(2) models for service times of e-mail jobs and network jobs are given in figures 5.19 and 5.20 respectively.

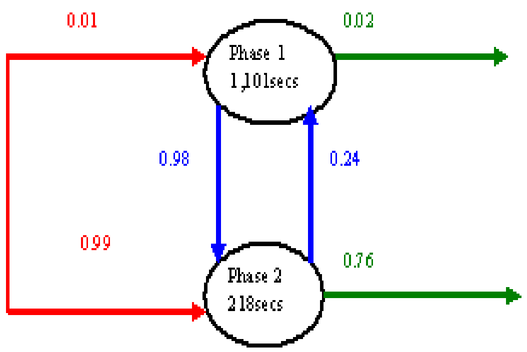

Figure 5.19 Structure of PH(2) Model - E-mail Jobs.

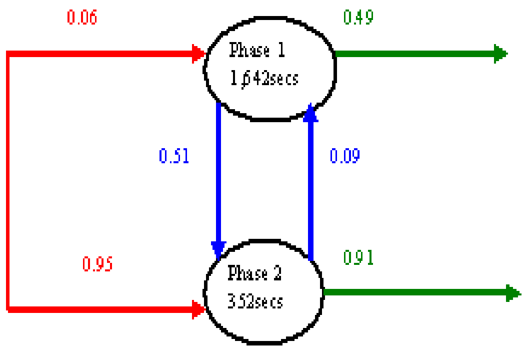

Figure 5.20 Structure of PH(2) Model - Network Jobs.

5.5 APPROXIMATING LOG-NORMAL MODEL USING PHASE-TYPE DISTRIBUTION

Since every continuous distribution can be approximated using phase-type distribution, the log-normal models of service times in this particular help desk were approximated using the phase-type distribution.

Approximating the log-normal distribution minimizes the information divergence, this is equivalent to maximizing the log-likelihood function when fitted to sample [27], [4] . The EMpht program was used to approximate the log-normal models of service times. Phase-type distribution with three phases (PH(3)) gave the best approximation to log-normal models of overall service times and service times by job types.

Figures 5.21- 5.25 provide the comparison of survival functions and distribution functions of log-normal models and PH(3) models for overall service times and service times of administrative , e-mail , miscellaneous and network jobs respectively. Approximations of log-normal models for service times of miscellaneous jobs and network jobs are better than overall service times, service times of administrative and e-mail jobs. Also, as the shape parameter(σ) of the log-normal model becomes smaller, the approximation becomes better.

SERVICE TIME - OVERALL SERVICE TIME

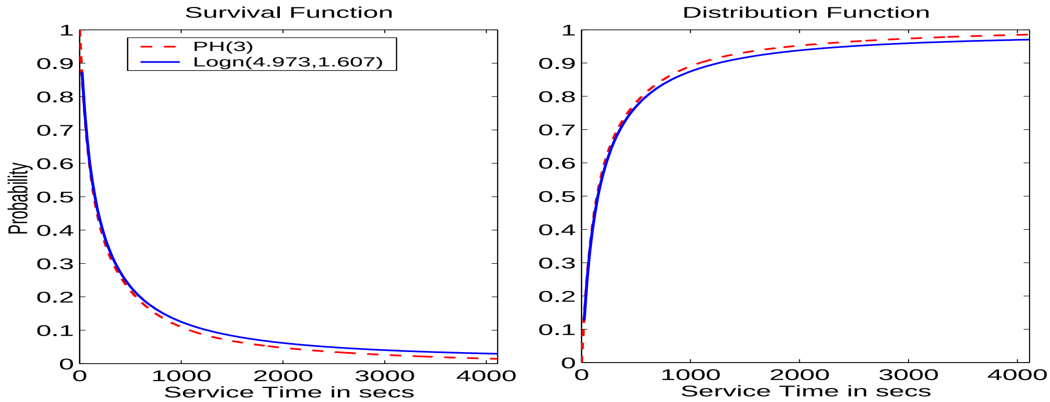

Figure 5.21 Approximated Log-Normal Model - Overall Service Time.

SERVICE TIME - ADMINISTRATIVE JOB

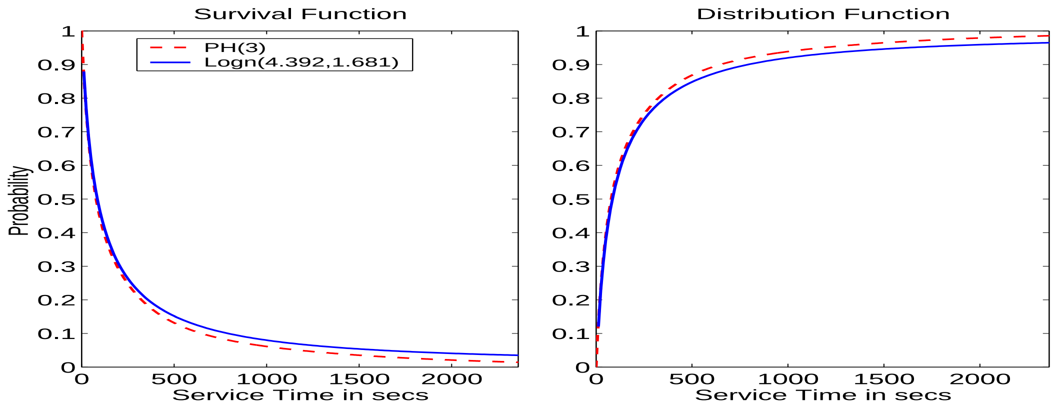

Figure 5.22 Approximated Log-Normal Model - Administrative Jobs.

SERVICE TIME - E-MAIL JOB

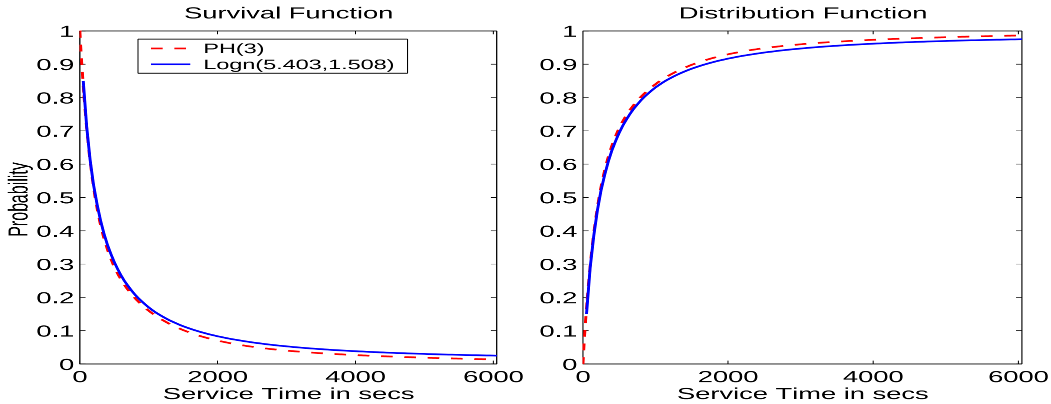

Figure 5.23 Approximated Log-Normal Model - E-mail Jobs.

SERVICE TIME - MISCELLANEOUS JOB

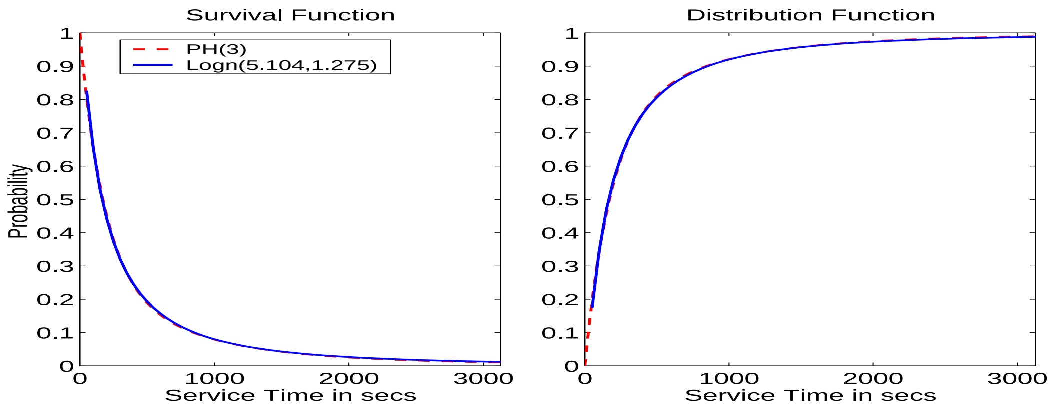

Figure 5.24 Approximated Log-Normal Model - Miscellaneous Jobs

SERVICE TIME - NETWORK JOB

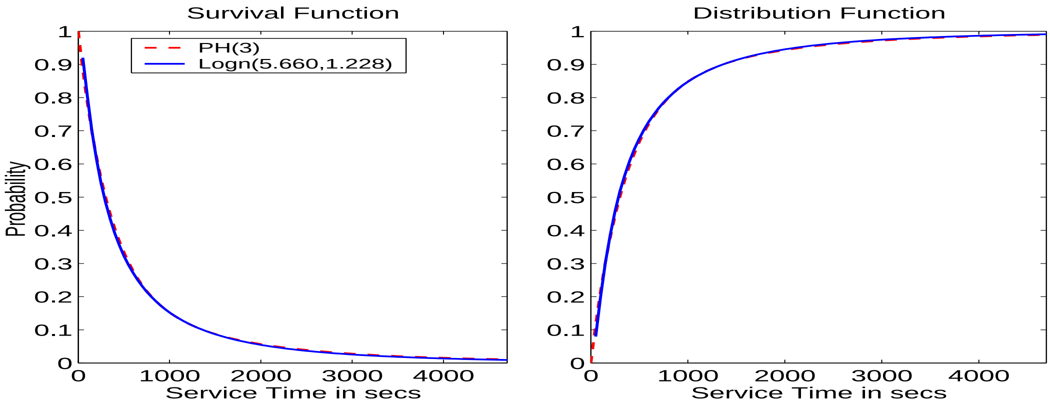

Figure 5.25 Approximated Log-Normal Model - Network Jobs.

The descriptive statistics of approximated log-normal models and phase-type models are presented in Tables 5.15 to 5.19.

SERVICE TIME - OVERALL SERVICE TIME

Table 5.15 Approximated Model - Overall Service Time.

	Logn(4.973,1.607)	PH(3)
Mean	525.434	447.6769
S.d	1837.479	916.3222
C.V	3.50	2.05

SERVICE TIME- ADMINISTRATIVE JOB

Table 5.16 Approximated Model - Administrative Jobs.

	Logn(4.392,1.681)	PH(3)
Mean	331.916	260.8167
S.d	1322.415	525.9529
C.V	3.98	2.02

SERVICE TIME- E-MAIL JOB

Table 5.17 Approximated Model - E-mail Jobs.

	Logn(5.403,1.508)	PH(3)
Mean	692.309	644.8207
S.d	2044.225	1355.8
C.V	2.95	2.10

SERVICE TIME- MISCELLANEOUS JOB

Table 5.18 Approximated Model - Miscellaneous Jobs.

	Logn(5.104,1.275)	PH(3)
Mean	371.227	368.8817
S.d	749.987	692.3616
C.V	2.02	1.88

SERVICE TIME- NETWORK JOB

Table 5.19 Approximated Model - Network Jobs.

	Logn(5.660,1.228)	PH(3)
Mean	610.325	602.5850
S.d	1144.684	1032.3
C.V	1.88	1.71

The log-normal models for overall service times, and service times of administrative, e-mail, miscellaneous and network jobs respectively were approximated by phase-type distribution of order $p = 3$.

Chapter 6
SUMMARY, CONCLUSION AND RECOMMENDATIONS FOR FUTURE WORK

6.1 INTRODUCTION

This chapter gives the summary of this present work, conclusion based on the analysis and findings of this work as well as recommendations for future work.

6.2 SUMMARY

The help desk analyzed in this work is a form of a contact centre for technical support for hardware or software. It is staffed by people who can either solve the problem directly or forward the job to someone else who can handle it. One of the goals of the design of this help desk is to establish the right number of staff to be employed as well as number of trunk lines to be provided in order to maintain a target service level. In order to achieve this goal, statistical model has to be developed to design this help desk.

Queueing models have been used extensively in designing call centres. In this research work, a simple queueing model $G|G|N|S$ was used in designing and analyzing this help desk. Major inputs to the building of this model are the description of the arrival process and the service process. This research work focused on describing service process in this help desk.

From the literature, service times are assumed to be exponentially distributed. Mandelbaum [23] proposed that at least for one call centre, exponential distribution does not describe the service times and suggested a log-normal distribution as an option to model the service times in call centres. Since queues with log-normal service times are difficult to analyze, phase-type distributions that can model several stages of the service process were investigated as possible models of the service times.

This research work confirms the irrelevance of assumption of exponential distribution to this help desk. It also modelled service times in this help desk with log-normal distribution as suggested by Mandelbaum. Phase-type distributions with phases $p = 2, 3, 4$ were fitted to service times of this help desk. Finally, the log-normal models of this help desk were approximated using phase-type distribution of order $p = 3$. Parameters of log-normal models were estimated using the maximum likelihood parameter estimation. Due to the complexity in the derivation of maximum likelihood parameter estimates of the phase-type distribution, the expectation - maximization (EM) algorithm was used in estimating the parameters of the phase-type distribution. This was implemented using the EMpht program. The two famous goodness of fit tests, Kolmogorov-Smirnov and Anderson-Darling tests were implemented to assess the models and for model selection.

6.3 CONCLUSION

Based on the analysis of the data from this particular help desk, it was found that service times in this help desk are not exponentially distributed. Log-normal distributions gave appropriate description of the overall service times and the service times of administrative, e-mail, miscellaneous and network jobs. A phase-type distribution with three phases (PH(3)) provided a reasonable fit to the overall service times and the service times of administrative jobs and miscellaneous jobs. Whereas a phase-type distribution with two phases (PH(2)) provided a suitable model for the service times of e-mail and network jobs. Finally, a phase-type distribution with three phases (PH(3)) was used to approximate the log-normal model for the overall service times and the service times by job types.

6.4 RECOMMENDATION FOR FUTURE RESEARCH

The arrival process in this help desk should be investigated, so as to be able to analyze the model $G|G|N|S$ used in modelling this help desk. Further work is necessary on practical interpretation of the phases in the phase-type model. Analyzing other call centre data should also provide more insights into modelling service times by other non-exponential distributions.

References

1. O.O. Aalen, *On Phase-Type Distribution In Survival Analysis*. Scandinavian Journal Of Statistics, Vol. 22 , p.p.447-463 : 1995.

2. R.B D'Agostino, and M.A. Stephens, *Goodness Of Fit Techniques*. IMarcel Dekker :1988.

3. J. Aruna, G. David, A. John, *Modelling Service Time Distributions In Cellular Networks Using Phase-Type Distributions*. IEEE Trans. Communication, Vol. 2 , p.p.440-444 : 2001 .

4. S. Asmussen, O. Nerman, and M. Olsson, *Fitting Phase-Type Distributions Via The EM Algorithm* Scandinavian Journal of Statistics, Vol. 23, p-p 419-441 : 1996.

5. S. Asmussen, *Applied Probability and Queues*. John Wiley and Sons, : 1987.

6. F. Baccelli, and G. Hebuterne, *On Queues With Impatient Customers.*(In Performance), :1981.

7. I.V Basawa, P. Rao, *Statistical Inference For Stochastic Processes*. Academic Press London, : 1980.

8. P. Bishop, and E. Bloomfield, *Using A Log-Normal Failure Rate Distribution For Worst Case Bound Reliability Prediction* Adelard and Centre For Software Reliability.

9. E. Brockmeyer, H.L. Halstrom, and J. Arne , *The Life and Works Of A.K. Erlang.*Transactions of The Darnish Academy Of Technical Sciences, No. 2, pp 277 : Copenhagen 1948 .

10. J. Chambers, C. Williams, K. Beat, and T. Paul, *Graphical Methods For Data Analysis*. Wadsworth :1983.

11. E. Chlebus, *Empirical Validation of Call Holding Time Distribution in Cellular Communication Systems.*: 1997.

12. C.A. O'Cinneide, *Characterization Of Phase-Type Distributions*. Commun. Statist.- Stochastic Models, Vol. 6 , p-p.1-57 : 1990.

13. B. Cleveland, and J. Mayben , *Call Center Management On Fast Forward: Succeeding In Today's Dynamic Inbound Enviroment.* : Copenhagen 1997 .

14. S. Coles , *An Introduction To Statistical Modelling Of Extreme Values*. Springer-Verlag London Limited :2001 .

15. D.R. Cox, and D. Oakes, *Analysis Of Survival Data*. Chapman Hall

16. E.L. Crow, and K. Shimizu, *Log-Normal Distributions*. Marcel Dekker: 1988.

17. Y. Fang, and I. Chlamtac *Teletraffic Analysis and Mobility Modelling of PCS Networks* IEEE Trans. Communication, Vol. 47, p-p 1062-1072 : July 1999

18. A. Horvath, M. Telek , *Approximating Heavy Tailed Behaviour With Phase-Type Distributions.*Advance in Algorithmic Methods For Stochastic Models G.Latoche and P.Taylor (eds.) p-p 191-214 : 2000 .

19. E. Ishay, *Fitting Phase-Type Distributions to Data From a Telephone Call Center* Master Thesis, Technion: March 2003.

20. ITU-D Study Group2 , *Teletraffic Engineering*.

21. E. Kao, *An Introduction To Stochastic Processes*. ITP: 1997.

22. D.G. Kendall, *The Study of Queues With Multiple Servers*. Seminar Work of Kiefer and Wolfowitz: 1953.

23. A. Mandelbaum, A. Sakov, and S. Zeltyn, *Empirical Analysis of A Call Center*. Technical Report, Technion Isreal Institute of Technology : 2000.

24. A. Mandelbaum, and S. Zeltyn, *The Palm/Erlang A Queue, with Applications To Call Centers* Faculty of Industrial Engineering and Management Technion, Isreal: 2005

25. A.H Marshall, and S.I McLean, *Conditional Phase-Type Distributions For Modelling Patient Length Of Stay In Hospital.*(IFORS), Edinburgh :2002.

26. M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithm Approach.*(The Johns Hopkins University Press), Baltimore :1981.

27. M. Olsson, *The EMpht-programme*. Department of Mathematics, Chalmers University of Technology, and Goteborg University : 1998.

28. M. Olsson, *Estimation Of Phase-Type Distributions From Censored Data*. Scandinavian Journal, Vol. 23 , p-p.443-460 : 1996.

29. C. Palm, *Research on Telephone Traffic Carried By Full Availability Groups.* Tele, Vol. 1, p.p 107 : 1957.

30. A. Riska, V. Diev, E. Smirni, *Efficient Fitting Of Long-Tailed Data Sets Into Hyperexponential Distributions*. Dept. Of Computer Science, College Of Williams and Mary, Williamburg USA.

31. P. Royston, *The Log-Normal Distributions As A Model For Survival Time In Cancer, With Emphasis On Prognostic Factors*.Statistica Neerlandica : 2001 .

32. S.M. Ross , *Stochastic Processes*.Wiley Series in Probability and Mathematical Statistics, Second Edition : 1996 .

33. M.A Stephens, *EDF Statistics For Goodness Of Fit And Some Comparisons*. American Statistical Association Journal, Vol. 69 , p-p.730-737 : 1974.

34. R. Wallace, W. Whitt, *A Staffing Algorithm For Call Centers With Skill-Based Routing*. IBM, The George Washington University and Columbia University : 2005.