# MODELLING WAITING TIME IN QUEUES 

## KARONEY KENNETH KIBET (BSC APPLIED STATISTICS WITH COMPUTING)

A THESIS SUBMITTED FOR EXAMINATION IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS IN THE SCHOOL OF BIOLOGICAL AND PHYSICAL SCIENCES OF MOI UNIVERSITY

## DECLARATION

This thesis is my original work and has never been presented for a degree in any other university. No part of this thesis may be reproduced without the prior permission of the author and/or Moi University.

Signature Date

## KARONEY KENNETH KIBET - MSC/BS/06/15

## Declaration by the supervisors

This thesis has been submitted for examination with our approval as university supervisors.

## Signature

Date. $\qquad$

## Dr. Mathew Kosgei

Department of Statistics and Computer Science
Moi University
P. O. Box 3900-30100,

ELDORET- KENYA.

Signature
Date.
Prof. Kennedy L. Nyongesa
Department of Mathematics
Masinde Muliro University of Science and Technology,
P. O. Box 190-05600,

KAKAMEGA- KENYA.

## DEDICATION

I dedicate this thesis to my family, friends and all my classmates.

## ACKNOWLEDGEMENT

My heartfelt gratitude goes to my supervisors, Prof Kennedy L. Nyongesa and Dr Mathew Kosgei for their guidance, tireless patience, positive criticism and concern during supervision of this study. With profound humility they have provided me with professional and moral support during the entire course of my graduate program.

I am grateful for the love and support from my parents, brothers and sisters to this far. I am greatly indebted to Dr Gregory Kerich for his assistance and encouragement. I would also like to express a lot of appreciation to all members of statistics and computer science department, Moi University, for their inspiration and support. Many thanks to my postgraduate colleagues for their moral support during write up of this thesis. So much was worthwhile through their humility, encouragement and understanding.


#### Abstract

The Mathematical study of queue waiting time is mainly concerned with queue performance measures where several applications have been drawn in past studies. Among the vast uses and applications of the theory are the various queues experienced in day to day lives where the theory has been used to solve the problem of long queues which leads to resource waste. In recent developments, the electronic quality management systems have been adopted to manage queues which can be used in assessment, analysis and even more consequentially to estimate the average queuing times in waiting lines as well as to reveal the steady state properties of queues. Ultimately, these queue performance measures are used to make decisions on improving services. Single server stochastic queuing models have been studied and used extensively to model the mean waiting time in queues. There is need to utilize a joint model comprising of both stochastic and deterministic queuing models in order to solve the inadequacies of the non-randomness property of the stochastic models. This study is intended to model waiting time in queues by use of stochastic and deterministic queuing systems where the main aim of this study was to model waiting time in queues where specifically, models for estimating the mean waiting time were obtained. Consequently, real time queue data was applied on these models and finally, validation of the model through simulation was also performed. To accomplish these, the waiting time in queues was modeled on the basis of both stochastic model represented as M/M/C denoting Markovian arrivals, Markovian service times and multiple (C) servers and a deterministic model represented as D/D/1 denoting Deterministic arrivals, Deterministic service times and one (1) service channel (server). Bank queue data was applied on the models obtained to ascertain whether the various queue performance measures explain the behavior of a queue and to validate the models simulation was performed using MATLAB software. The results reveal that both deterministic and the stochastic delay components are compatible in modeling waiting time. The models also are applicable to real time bank queue data where upon simulation, both models depict a fairly equal waiting times for server utilization factors below 1.For instance when $\rho<1$, and an infinitely increasing delay at $\rho>1$. In conclusion, the models need to be implemented in queue systems so that queues are more efficient in terms of service delivery.


## TABLE OF CONTENTS

Table of Contents
DECLARATION ..... ii
DEDICATION ..... iii
ACKNOWLEDGEMENT ..... iv
ABSTRACT ..... v
TABLE OF CONTENTS ..... vi
LIST OF TABLES ..... viii
LIST OF FIGURES ..... ix
LIST OF SYMBOLS AND NOTATIONS ..... x
CHAPTER ONE ..... 1
INTRODUCTION ..... 1
1.0 Introduction ..... 1
1.1 Background Information .....  1
1.2Basic Concepts ..... 3
1.2.1: Queuing Model ..... 3
1.2.2: Markov Process ..... 3
1.2.3: Markov Property ..... 3
1.2.4: Birth Death Queuing Framework ..... 3
1.2.5: Equilibrium ..... 4
1.3 Statement of the Problem ..... 4
1.4 General objective ..... 5
1.4.1 Specific objectives ..... 5
1.4.2Significance of the study ..... 5
CHAPTER TWO ..... 6
LITERATURE REVIEW ..... 6
2.1 Introduction ..... 6
2.2 Applications of Queuing Theory ..... 6
2.3 Deterministic Delay Models ..... 8
2.4 Steady State Models ..... 9
2.4.1 Exact Expressions ..... 9
2.5 Time Dependent Models ..... 10
CHAPTER THREE ..... 12
METHODOLOGY ..... 12
3.1 Introduction ..... 12
3.2 Model Formulation ..... 12
3.2.1 Deterministic Queue System ..... 12
3.2.2 Stochastic Queuing System ..... 13
3.2.3 Overall Queue Model ..... 14
3.3Data Utilization in the Model ..... 16
3.3.1 Sampling method ..... 16
3.3.2 Calculation of Parameters ..... 16
3.4Model Validation ..... 16
CHAPTER FOUR ..... 17
RESULTS AND DISCUSION ..... 17
4.1 Introduction ..... 17
4.2 Queue Model ..... 17
4.2.1 The Mean of Deterministic Delay Model ..... 17
4.2.2 Mean of Stochastic Delay Component ..... 18
4.2.3 Overall Queue Model ..... 18
4.3 Application of data to the Model ..... 19
4.3.1Computation of Parameters ..... 19
4.4Validation of the model ..... 22
4.4.1 Simulation of E(Wt1) ..... 23
4.4.2 Simulation of $E(W t 2)$ with two and three servers ..... 24
4.4.4 Simulation of $E(W t)$ With Three Servers ..... 25
4.4.5 Simulation of $\mathrm{E}(\mathrm{Wt})$ with Two Servers ..... 26
CHAPTER FIVE ..... 28
CONCLUSION AND RECOMMENDATION ..... 28
5.1 Introduction ..... 28
5.2 Conclusion ..... 28
5.3 Recommendation ..... 28
REFERENCES ..... 29
Appendix A: MATLAB Iteration codes ..... 31
MATLAB Iteration code for simulating $\mathrm{E}[\mathrm{Wt} 1]$ versus $\rho$. ..... 31
MATLAB Iteration code for simulating $\mathrm{E}[\mathrm{Wt2} 2]$ versus $\rho$ with Three and Two Servers. ..... 32
MATLAB Iteration code for simulating $\mathrm{E}[\mathrm{Wt}]$ versus $\rho$ with three servers. ..... 33
MATLAB Iteration code for simulating E[Wt] versus $\rho$ with Two servers. ..... 34
Appendix B: Data Tables ..... 36

## LIST OF TABLES

Table 1Data collected on $16^{\text {th }}$ July 2018 to $20^{\text {th }}$ July 2018 ..... 36
Table 2 Data collected on $23^{\text {rd }}$ July 2018 to $27^{\text {th }}$ July 2018 ..... 37
Table 3 Data collected on $30^{\text {th }}$ July to $3^{\text {rd }}$ August 2018 ..... 38
Table 4 Data collected on $6^{\text {th }}$ August to $10^{\text {th }}$ August 2018 ..... 39

## LIST OF FIGURES

Figure 1 Diagram representing bank electronic quality management queue system layout...... 1


#### Abstract

Figure 2 Diagram representing the deterministic queue delay13


Figure 3 Diagram representing simulation of deterministic component ..... 23
Figure 4 Diagram representing simulation of stochastic component with two and three servers24
Figure 5 Diagram representing simulation of the overall model with three servers ..... 25
Figure 6 diagram representing simulation of overall model with three servers ..... 26

## LIST OF SYMBOLS AND NOTATIONS

$\mathrm{A}(\mathrm{t})$ - Cumulative arrivals.
C - Number of servers.
$\mathrm{c}_{\mathrm{y}}$ - Cycle time (min).
$\mathrm{D}(\mathrm{t})$ - Cumulative departures.
$\mathrm{E}\left[\mathrm{W}_{\mathrm{t}}\right]$ - Expected overall waiting time.
$\mathrm{E}\left[\mathrm{W}_{\mathrm{t}_{1}}\right]$ - Expected deterministic delay.
$\mathrm{E}\left[\mathrm{W}_{\mathrm{t}_{2}}\right]$ - Expected stochastic delay.
$\mathrm{F}_{\mathrm{q}}(\mathrm{t})$ - Waiting time distribution function.
$\mathrm{g}_{0}$ - Time necessary for the queue to dissipate.
$g_{e}$-Effective service time.
iid - Independent and identically distributed.
$L o S$ - Levels of service
$\mathrm{L}_{\mathrm{q}}$ - Queue size on the line.
$\mathrm{L}_{\mathrm{s}}$ - Queue size in the system.
$N(t)$ - Queue size at time ' $t$ '.
$P_{0}$ - Steady state probability of having no customers in the system.
$P_{n}(t)$-Probability that the queue is of size ' $n$ ' at time ' $t$ '.
$r$ - Effective waiting time on the queue before service.
$\mu$ - Departure rate.
W - Overall waiting time.
$\mathrm{W}_{\mathrm{q}}$ - Waiting time on the queue.
$\mathrm{W}_{\mathrm{s}}$ - Waiting time on the system.
$\mathrm{W}_{\mathrm{t}}-$ Overall queue delay.
$\mathrm{W}_{\mathrm{t}_{1}}$ - Deterministic queue delay component.
$\mathrm{W}_{\mathrm{t}_{2}}-$ Stochastic queue delay component.
$\lambda$ - Arrival rate.
$\pi_{\mathrm{w}}$ - Probability of waiting on the queue.
$\rho$ - Utilization factor.

## CHAPTER ONE

## INTRODUCTION

### 1.0 Introduction

This chapter contains the background information, basic concepts, statement of the problem and the objectives of the study.

### 1.1 Background Information

Waiting is one of the most unpleasant experiences of life. Queuing theory deals with delays and queues which are essentials in determining the levels of service in queues. It also evaluates the adequacy of service channels and the economic losses that come about as a result of long waiting lines. Quantifying these delays accurately and appropriately in queues is critical for planning design and analysis of service channels. In modern times, queuing has been automated such that customers arrive and pick ticket numbers from a ticket dispensing machine. Electronic quality management systems were implemented for purposes of instilling order and eliminating or easing/reducing congestion in queues. In figure (1) below, a diagrammatic description of a queue system is presented.


The queue setup assumes the customers do not possess the following characteristics;
i) Balking -This is the refusal of a customer to join the queue if the queue is too long.
ii) Reneging - This is the withdrawal of a customer from the queue because of the length of the waiting line.
iii) Jokeying-When a customer withdraws from a queue to join another one because the new queue is either shorter or moving faster.

From Figure (1) above, it can be intuitively seen that customer arrivals to the queue are a stochastic Poisson process and the ticket dispenser gives ticket numbers to customers. By a customer getting a ticket number, they join in a queue and the customer at the front of the queue will be served by the next idle server who will complete service first among the ' C ' busy severs. That is each server can serve customers of different classes of service, class here to mean the various types of needs customers may seek. Assigning of customers to the queue is stochastic since any customer can be served by any available server. Models that incorporate stochastic random variations of the bank queue systems where the Electronic Quality Management Systems have been recently adopted by organizations to improve queue performance can be very appealing in waiting lines. They were intended to simplify theoretical models that are numerically inconsequential. Of the various queuing models, M/M/C and D/D/1 are intended to be used in the study, where M/M/C implies Markovian arrivals and independent identically distributed service times following an exponential distribution. The ' C ', where $\mathrm{C} \geq 1$ in the system represents the number of service channels. For the $\mathrm{D} / \mathrm{D} / 1$ queuing system, the first D represents uniform arrivals with parameter $\lambda$, the second D represents constant departures with parameter $\mu$ and 1 represents one service channel.

### 1.2Basic Concepts

This section discusses the various concepts and terminologies that are relevant to queuing theory modelling.

### 1.2.1: Queuing Model

A queuing model is a Mathematical representation of a waiting line or queue. In queuing, a model is constructed so that queue lengths and waiting time can be predicted.

### 1.2.2: Markov Process

A stochastic model is a Markov process when it is used to model a random system that changes states according to a transition rule that only depends on the current state.

### 1.2.3: Markov Property

A stochastic process has the Markov property if the conditional probability distribution of future states of the process (conditional on both past and present values) depends only upon the present state; that is, given the present, the future does not depend on the past.

### 1.2.4: Birth Death Queuing Framework

The birth death process is a special case of continuous time Markov process where the state transitions are of only two types: "births" which increases the state variable by one and "death" which decreases the state transition by one. It is a specific type of continuous-time Markov chain. It consists of a set of states $\{0,1,2, \ldots\}$, typically denoting the "population" of some system. State transitions occur as unit jumps up or down from the current state. More specifically, when the system is in state $\mathrm{n} \geq 0$, the time until the next arrival (or "birth") is an exponential random variable with rate $\lambda_{\mathrm{n}}$. At an arrival, the system moves from state n to state $\mathrm{n}+1$. When the system is in staten $\geq 1$, the time until the next departure (or "death") is an exponential random variable with rate $\mu_{\mathrm{n}}$. At a departure, the system moves from state n to state $\mathrm{n}-1$.

### 1.2.5: Equilibrium

If a Markov process is irreducible (all states communicate) then the limiting distribution $\lim _{t \rightarrow \infty} P_{n}(t)=P_{n}$ exists and is independent of the initial conditions of the process. The limits $\left\{P_{n}, n \in N\right\}$ are such that they either vanish identically (i.e. $P_{n}=0 \forall n \in N$ ) or are all positive and form a probability distribution(i.e. $P_{n}>0 \forall n \in N, \sum_{n \in N} P_{n}=1$ ). The limiting distribution $\left\{\mathrm{P}_{\mathrm{n}}, \mathrm{n} \in \mathrm{N}\right\}$ is an irreducible recurrent markov process given by the unique solution of the equation $\mathrm{PA}=0$ and $\sum_{\mathrm{n} \in \mathrm{N}} \mathrm{P}_{\mathrm{n}}=1$ where $\mathrm{P}=\left(\mathrm{P}_{0}, \mathrm{P}_{1}, \mathrm{P}_{2}, \ldots\right)$. One can think of a state of equilibrium in a stochastic process and how that affects the difference differential equations. In a state of equilibrium also known as steady state, the behavior of the process is independent of the time parameter.

### 1.3 Statement of the Problem

Waiting lines are becoming a nuisance in most of our encounters of day to day activities. Single server stochastic queuing models have been studied and used extensively to model the mean waiting time in queues. There is need to utilize a joint model comprising of both stochastic and deterministic queuing models in order to solve the inadequacies of the nonrandomness property of the stochastic models. This study is intended to address the problem by modeling waiting time in queues by use of $\mathrm{M} / \mathrm{M} / \mathrm{C}$ and $\mathrm{D} / \mathrm{D} / 1$ queuing systems.

### 1.4 General objective

To model waiting time in queues by obtaining the mean waiting time models.

### 1.4.1 Specific objectives

The objectives were to:
i) Formulate a joint model that estimates the mean waiting time in queues.
ii) Apply real time queue data to the model.
iii) Validate the model.

### 1.4.2Significance of the study

Modeling waiting times is very fundamental as it reveals queue parameters and performance measures that explain the behavior of a queue which can be used by organizations in making informed decisions during planning and optimization of server resource which may result in satisfaction and improvement on efficiency and effectiveness during service delivery.

## CHAPTER TWO

## LITERATURE REVIEW

### 2.1 Introduction

This chapter provides literature review on the application of queuing theory, deterministic delay models, steady state models and time dependent models.

### 2.2 Applications of Queuing Theory

Queuing theory has become one of the most important and arguably one of the most universally used tool that has applications in diverse fields including telecommunications traffic engineering, computing and design of factories, hospitals and banks. (Erlang, 1909) was the first to study queuing and it resulted in a vast acclaimed Erlang telephone exchange model all over the world. Queuing theory is the mathematical study of waiting lines or queues. The theory enables mathematical analysis of a wide range of processes including the arrival (back of the) queue, waiting in the queue (simply a storage process) and being served at the front of the queue. Several performance measures can be derived from queuing probability of having exactly $n$ customers, average waiting time, steady state probability of encountering the system in certain states such as empty, full, having an available server or having to wait a certain time to be served. Queuing theory has been applied in many areas as outlined by (Wayne, 2003). (Erlang, 1909) did the first application of queuing theory to the problem of the telephone exchange. He examined the telephone network system and tried to determine the effect of fluctuating service demands on calls on the utilization of the automatic dial equipment for one telephone operator. Erlang proceeded to extend the results to the activities of several telephone operators in 1917. Development in the field of telephone traffic continued largely along the lines initiated by Erlang and the main publications were
among them (Molina, 1927). A probabilistic approach to the analysis of queuing was initiated by (Kendall, 1953) when he demonstrated that embedded Markov chains can be identified in the queue length process in systems $\mathrm{M} / \mathrm{G} / 1$ and $\mathrm{GI} / \mathrm{M} / \mathrm{C}$ where $\mathrm{G}, \mathrm{GI}$ and ' $\mathrm{C}^{\prime}$ ' are provided in the list of symbols. (Lindley, 1952) derived integral equations for waiting time distributions defined by embedded Markov points in the general queue, GI/M/C. This investigation led to the use of renewal theory in queuing system analysis by (Çinlar, 1969). Therefore, queuing theory has become one of the most universally used tool by statisticians in diverse fields including telecommunications traffic engineering, computing, banking etc. A queuing model of a system is an abstract representation whose purpose is to isolate the factors that relate to the system's ability to meet service demands whose occurrences and durations are random (Janos, 2010). Queuing models provide the analyst with a powerful tool for designing and evaluating the performance of queuing systems (Banks,J.,Carson,J.,Nelson,B.L,and Nicol,D., 2001). Queuing-based teller models were introduced into the financial industry in the late 1960's and early 1970's primarily to control increasing labor expenses (Brewton, 1989). (Wenny \& Whitney, 2004) determined server scheduling using simulation with arrival rates. (Emuoyibofarhe, Omotoso, Ozichi, \& and Popoola, 2005) studied the queuing problem of banks and applied the theory to solve multiple server problem M/M/C/ $\infty / \infty / \infty$ Where first ' $M$ ' implies Markovian arrivals, second ' $M$ ' implies Markovian service times, the ' $C$ ' represents multiple servers with; capacity of the system, population size and space to queue being infinite. (Bakari, 2014) studied bank Automated Teller Machine queues, analyzed and found out that to contain queue length, utilization (i.e. traffic intensity) must be less than one, the server must have unused capacity. (Bajpai, 2013) determined behavior in queues in four different situations in India in a comparative study. The theory has been used in evaluating levels of service in queuing systems. (Toshiba, 2013) and (Yusuff, 2010) applied the use of queuing theory to establish the queue optimization model. Other queuing theory scenarios
arise in road traffic signalized intersection queues, telephone call centers where there has been a large amount of work in recent years done on flexibility in queuing systems and its potential benefits for performance (You-Tong He, 2012).
(Agbola, A. \& Salawu,R.O., 2008) identified various information and communication technologies in the use and determined how they could be utilized for optimal performance on business transactions in the banking industry. Efforts in this study are directed towards developing an overall queuing model for estimating the mean delay in a queue. From this literature, it is necessary to advance the researches into deterministic and stochastic models which will be used in formulating a joint model that estimates the mean waiting time in queues.

### 2.3 Deterministic Delay Models

(Zukerman, 2012) considered a case where inter-arrival and service times are deterministic. To avoid ambiguity, he assumed that if an arrival and a departure occur at the same time, the departure occurs first. Such an assumption is not required for Markovian queues where the queue size process follows a continuous-time Markov-chain because the probability of two events occurring at the same time is zero, but it is required for deterministic queues. Unlike many of the Markovian queues, steady state queue size distribution for the deterministic queues does not exist because the queue size deterministically fluctuates according to a certain order. According to his work, all the entities that enter the system are served before the next one arrives, the mean queue size of $\mathrm{D} / \mathrm{D} / 1$ must be equal to the mean queue size in the queue system and therefore it is equal to the traffic intensity. In other words, the queue size alternates between the values 1 and 0 , spending a time period of $\frac{1}{\mu}$ at state 1 , then a time period $\frac{1}{\lambda}-\frac{1}{\mu}$ at state 0 , then again $\frac{1}{\mu}$ at state 1 . Without very strict assumptions that a queue follows a regular arrival and departure process, Zukerman could not accurately predict queue
behavior based on the deterministic queue model therefore rendering the model as insufficient to measure and describe queue parameters and hence called for the need to incorporate a stochastic model.

### 2.4 Steady State Models

These models characterize delays based on statistical distributions of the arrival and departure processes. Due to the purely theoretical foundation of the models, they require very strong assumptions to be considered valid.

### 2.4.1 Exact Expressions

(Beckmann,M.J., McGuire,C.B., \& WinstenC.B., 1956) derived the mean delay at fixed-time signals with the assumption of the binomial arrival process and deterministic service. The expected overflow queue used and the restrictive assumption of the binomial arrival process reduce the practical usefulness of Beckman's work. (Little, 1961) analyzed the expected delay at or near traffic signals to a vehicle crossing a Poisson traffic lane. (McNeil, 1968) derived a formula for the expected delay with the assumption of a general arrival process, and a constant departure time. From this work, (Tarko,A., Rouphail,N., \& Akcelik,R.B, 1993) expressed the total customer delay during one cycle as a sum of two delay components. (Beckmann,M.J., McGuire,C.B., \& WinstenC.B., 1956) obtained an identical expression to that obtained by (Darroch, 1964)when the arrival process follows a binomial distribution. With departure process being deterministic, (Darroch, 1964) derived the mean delay expectation when the service is random. (Gazis, 1974) considered the case of the compound Poisson arrival process and the general departure process obtaining a model for estimating the mean delay.

### 2.5 Time Dependent Models

For stochastic equilibrium to be achieved it requires the assumption that the steady state models undergo an infinite time period of stable queue flow conditions. Traffic flows during peak hours are hardly stationary, thus violating the important assumption of steady state models. (Liping,F. \& Bruce,H., 1999) developed a model for estimating arrival that is time dependent and is subjected to large variation because of the randomness of arrivals. The model was constructed on the basis of the delay evolution patterns under two extreme conditions: highly under saturated and highly over saturated conditions. They used the traditional uniform delay model and Canadian capacity guide (Teply,S., Allingham,D.I., Richardson,D.B., \& Stephens,B.W, 1995) to estimate the mean arrival time dependent delay. The model for estimating this delay was established through coordinate transformation based on steady state model and the deterministic model for arrival time dependent overflow delay (Kimber,R. \& Hollis,E., 1979). (Richa, 2014) performed the mathematical analysis of queue with phase service to estimate the average waiting time. (John, 2015) developed a queue model with server walking time dependent average delay model. The literature provided describes studies done on estimating the mean waiting time in queues. The literature was used in this study where the contents were advanced to develop an overall mean waiting time model comprising of a stochastic and a deterministic model with the application of embedding. Stochastic models are more reliable in describing queue behaviors. However they do not completely capture all the aspects of non- randomness in a queue setup thus there was need to combine both aspects of randomness and non-randomness. Ideally a queue setup is largely random but there are non-random occurrences in terms if arrivals and departures which cannot be ignored in modeling a queue system. Both stochastic and deterministic queue data was applied to check if the data fits well with the model was covered in this study. The past difficulty of arriving at numerical solutions for queuing models is no longer a
disadvantage, as simulation can be run to arrive at approximate answers. Simulation of queuing theory models allowed for alteration of the values of variables involved and analyze the results of the change of the variables.

## CHAPTER THREE

## METHODOLOGY

### 3.1 Introduction

This chapter presents the methods that were used in the study. Section 3.2 discusses the formulation of the model and the assumptions of the queuing models. Section 3.2.1 illustrates the use of $\mathrm{D} / \mathrm{D} / 1$ deterministic queuing system mean waiting time while the $\mathrm{M} / \mathrm{M} / \mathrm{C}$ stochastic queuing system mean waiting time is described in section 3.2.2. Section 3.3 discusses the application of real time queue data on the model, Section 3.3.1 describes the sampling method used to collect the sample queue data collected and MATLAB Statistical software for simulation in Section 3.4.

### 3.2 Model Formulation

The model is formulated on the basis of deterministic and stochastic queue systems to obtain a joint mean delay model.

### 3.2.1 Deterministic Queue System

Deterministic delay is denoted by $\mathrm{W}_{\mathrm{t}_{1}}$. In this sectionthe mean of $\mathrm{W}_{\mathrm{t}_{1}}$ was obtained. The mean hence estimated by use of a deterministic queuing system D/D/1.InFigure2below a diagrammatic description of the deterministic delay process is presented.


Figure 2 Diagram representing the deterministic queue delay
The figure displays the deterministic delay experienced in queues.
From the figure, $\mathrm{D}(\mathrm{t})$ and $\mathrm{A}(\mathrm{t})$ represent the cumulative departure and arrivals respectively. The area covered by triangle ABC represents the total deterministic delay in the queue. The Mean delay model was computed from the illustration in figure 2 above.

### 3.2.2 Stochastic Queuing System

In this section, the mean waiting time statistical measure of the stochastic queue system is computed denoted by $\mathrm{W}_{\mathrm{t}_{2}}$. The model is established on the basis of $\mathrm{M} / \mathrm{M} / \mathrm{C}$ queue system. Under this system, arrivals to the queue are in a Poisson process stream where the interarrival times as well as the service times follow a shifted exponential distribution given as

$$
F(t)=1-\lambda e^{-\lambda t}
$$

At steady state, $\mathrm{P}_{\mathrm{n}}=\operatorname{Pr}(\mathrm{N}=\mathrm{n}) ; \mathrm{n}=0,1,2, \ldots$ where steady state is the probability of having a queue length of size n which leads to the computation of the probability of waiting in the queue given as

$$
\begin{equation*}
\pi_{w}=\frac{(C \rho)^{C}}{C!(1-\rho)} P_{0} \quad \text { where } \pi_{w}=P(n \geq c) \tag{3.1}
\end{equation*}
$$

The delay probability model $\pi_{\mathrm{w}}$, was used to compute the length of the queue $\mathrm{L}_{q}$, length of the queue at service $\mathrm{L}_{s}$ and overall length of the queue, L and thereafter employ Little's theorem to compute the corresponding $\mathrm{W}_{q}, \mathrm{~W}_{s}$, and W provided for in the list of symbols.

To obtain the queue length $\mathrm{L}_{\mathrm{q}}$, which is the expectation of the number waiting in the line (i.e. $(\mathrm{n}-\mathrm{c})$ customers waiting in the queue) which yields $\mathrm{L}_{\mathrm{q}}$ as

$$
\begin{equation*}
L_{q}=\frac{\rho(\lambda / \mu)^{C}}{C!(1-\rho)^{2}} P_{0} \tag{3.2}
\end{equation*}
$$

To obtain L , which is the expectation of n number of customers in the system given as

$$
\begin{equation*}
\mathrm{L}=\frac{\rho(\lambda / \mu)^{\mathrm{C}} \mathrm{P}_{0}}{\mathrm{C}!(1-\rho)^{2}}+(\lambda / \mu) \tag{3.3}
\end{equation*}
$$

but from 3.2 above, $L_{q}$ is given above as $L_{q}=\frac{\rho(\lambda / \mu)^{C} P_{0}}{C!(1-\rho)^{2}}$

$$
\begin{gather*}
\text { Thus from } 3.3 \text { above } \mathrm{L}=\mathrm{L}_{\mathrm{q}}+\lambda / \mu \\
\text { implying that } \mathrm{L}_{\mathrm{s}}=\mathrm{L}-\mathrm{L}_{\mathrm{q}} \tag{3.4}
\end{gather*}
$$

therefore giving $L_{s}$ as $L_{s}=\lambda / \mu$
Little's theorem is applied to $L_{s}$ to obtain the corresponding $W_{S}$
Assuming F.I.F.O as the queue discipline, with exponential inter-departure times with the rate $\mathrm{C} \mu$ when $\mathrm{n} \geq \mathrm{C}$ then the distribution waiting time is given as

$$
\begin{equation*}
\mathrm{F}_{q}(\mathrm{t})=\pi_{\mathrm{w}} \mathrm{e}^{-\mathrm{C} \mu(1-\rho) \mathrm{t}} ; \mathrm{t} \geq 0 \tag{3.5}
\end{equation*}
$$

From the distribution waiting time, the expected waiting time on the queue can be obtained.

### 3.2.3 Overall Queue Model

Consider a cumulative arrival and departure of entities in a queue for the time interval $[0, \mathrm{~T}]$. The time taken in the queue herein referred to as overall queue delay is denoted by $\mathrm{W}_{\mathrm{t}}$. The
study seeks to formulate $\mathrm{W}_{\mathrm{t}}$ thatcomprises of deterministic and stochastic delay components as follows,

$$
\begin{equation*}
\mathrm{W}_{\mathrm{t}}=\mathrm{W}_{\mathrm{t}_{1}}+\mathrm{W}_{\mathrm{t}_{2}} \tag{3.6}
\end{equation*}
$$

Where $\mathrm{W}_{\mathrm{t}_{1}}$ is the deterministic delay component representing a delay that is experienced by customers with uniform arrival times and departures within the time interval $\left[t, t+c_{y}\right]$ while $\mathrm{W}_{\mathrm{t}_{2}}$ is the stochastic delay component representing the delay caused by random queues resulting from the random nature of arrivals.

Equation (3.6) is solved under the following assumptions.
i) A single service channel is considered for the deterministic queue model while multiple service channels were considered for the Markovian queue model.
ii) Customer arrivals in the queue are either uniform or random variable following a Poisson process and no queue is present at the time when the prediction is performed.
iii) There is adequate space for queuing.
iv) The queue discipline is F.I.F.O.
v) Customers are patient (no balking, reneging, or jockeying).
vi) The queue system is in steady state.

Queues can be effectively analyzed when the length of the time between two consecutive transitions of the two queue systems is considered as negligible and therefore embedding was applied to $W_{t}$ which is the joint model obtained that represents the overall queue delay where the behavior of the process was considered as an embedded renewal process and therefore the probability that a continuous time markov chain will be at a specific state at a certain time converges to a limiting value which is independent of the initial state.

### 3.3Data Utilization in the Model

### 3.3.1 Sampling method

The data was collected from randomly selected banks in Eldoret town for a period of four weeks between $16^{\text {th }}$ July 2018 and $10^{\text {th }}$ August 2018 on a daily basis between Monday and Friday of every week. Randomization in the choice of banks allows for an unbiased data sample. Data was taken for five days of the week to sufficiently capture queue data in the busiest days of the week. The queue data comprising of the number of customer arrivals and departures on an hourly basis from 9:00 AM to 3:00 PM was recorded. $g_{e}$ and $c_{y}$ represented effective deterministic service time and fixed cycle time of service respectively. Calculation of parameters was performed and the data collected yielded all the parameters of the model.

### 3.3.2 Calculation of Parameters

From the data collected, model parameters were computed.

### 3.4Model Validation

Simulation was performed in this study using statistical software MATLAB (Matrix Laboratory) where the parameters of the model were utilized to obtain data that was used to validate the model.

## CHAPTER FOUR

## RESULTS AND DISCUSION

### 4.1 Introduction

This chapter presents the results obtained in the study. Section 4.2 discusses the queue model, section 4.3 contains the application of data to the model and section 4.4 describes the validation of the model.

### 4.2 Queue Model

### 4.2.1 The Mean of Deterministic Delay Model

To compute the mean, it is assumed that customer arrivals and departures are uniformly distributed with rates $\lambda$ and $\mu$ respectively. The mean delay of customers for this case can be easily determined from figure 2 . The figure shows classical cumulative/departure graph against time for uniform arrival rate approach to the queue. The slope of the cumulative arrival line is the uniform arrival rate of customers per unit time, denoted by $\lambda$. The slope of the cumulative departure line is sometimes zero (when the server is occupied/busy) and sometimes $\rho$ (when a customer is called upon); where $\rho$ is the utilization factor obtained as $\rho=\frac{\lambda}{\mu}$.

Upon utilizing D/D/1 queuing system and the theory behind it, the mean waiting time is given as.

$$
\begin{equation*}
\mathrm{E}\left(\mathrm{~W}_{\mathrm{t}_{1}}\right)=\frac{\mathrm{C}_{\mathrm{y}}\left(1-\frac{\mathrm{g}_{\mathrm{e}}}{\mathrm{C}_{\mathrm{y}}}\right)^{2}}{2\left(1-\frac{\mathrm{g}_{\mathrm{e}}}{\mathrm{C}_{\mathrm{y}}} \rho\right)} \tag{4.1}
\end{equation*}
$$

### 4.2.2 Mean of Stochastic Delay Component

### 4.2.2.1 Expected Waiting Time while at Service

Applying little's theorem to Equation (3.5);

$$
\begin{align*}
& \mathrm{W}_{\mathrm{s}}=\frac{\mathrm{L}_{\mathrm{s}}}{\lambda} \\
& \equiv(\lambda / \mu) \div \lambda \\
& \mathrm{W}_{\mathrm{s}}=1 / \mu \tag{4.2}
\end{align*}
$$

### 4.2.2.2Expected Waiting Time on the Queue

Given the waiting time distribution from Equation 3.6 as;

$$
\mathrm{F}_{q}(\mathrm{t})=\pi_{\mathrm{w}} \mathrm{e}^{-\mathrm{c} \mu(1-\rho) \mathrm{t}} ; \mathrm{t} \geq 0
$$

The expected waiting time on the queue is given as

$$
E(t)=\frac{\pi_{w}}{c \mu(1-\rho)}=W_{q}
$$

Hence the expected waiting time for the stochastic queue is given as

$$
\begin{equation*}
\mathrm{E}\left(\mathrm{~W}_{\mathrm{t}_{2}}\right)=1 / \mu+\frac{\pi_{\mathrm{w}}}{\mathrm{c} \mu(1-\rho)} \tag{4.3}
\end{equation*}
$$

### 4.2.3 Overall Queue Model

Notice that from Equation (3.7), $\mathrm{W}_{\mathrm{t}}$ is split into two independent queue systems, that is $\mathrm{W}_{\mathrm{t}_{1}}$ and $W_{t_{2}}$. In the previous sections, computation of the mean waiting times of $W_{t_{1}}$ and $W_{t_{2}}$ was given. In this section both the deterministic and the stochastic models are combined to obtainE $\left(\mathrm{W}_{\mathrm{t}}\right)$ which is an additive joint model.

To obtain $\mathrm{E}\left(\mathrm{W}_{\mathrm{t}}\right)$

$$
\mathrm{E}\left(\mathrm{~W}_{\mathrm{t}}\right)=\left\{\mathrm{E}\left(\mathrm{~W}_{\mathrm{t}_{1}}\right)+\mathrm{E}\left(\mathrm{~W}_{\mathrm{t}_{2}}\right)\right\}
$$

The length of time between two consecutive transitions of the two queue systems was considered as negligible and therefore embedding $(\Delta)$ was applied to the joint model where $\Delta$ is the embedding factor combining the deterministic and the stochastic model which is given as

$$
\Delta=\left(\lambda+\frac{\lambda}{\mathrm{g}_{\mathrm{e}}-\lambda}\right)^{-1}
$$

Upon utilizing the embedding, the overall model becomes

$$
\begin{align*}
E\left(W_{t}\right)= & \left(\lambda+\frac{\lambda}{g_{e}-\lambda}\right)^{-1}\left(\frac{C_{y}\left(1-\frac{g_{e}}{C_{y}}\right)^{2}}{2\left(1-\frac{g_{e}}{C_{y}} \rho\right)}\right)+\frac{1}{\mu\left(\lambda+\frac{\lambda}{g_{e}-\lambda}\right)} \\
& +\frac{\pi_{w}}{c \mu(1-\rho)\left(\lambda+\frac{\lambda}{g_{e}-\lambda}\right)} \tag{4.4}
\end{align*}
$$

This model is the overall delay model for estimating the overall mean waiting time on a queue on the basis of a deterministic and a stochastic delay model with the application of embedding.

### 4.3 Application of data to the Model

The application has been done to ascertain the representation of the parameters on what the model depicts

### 4.3.1Computation of Parameters

For convenience of the sampling measurement, it was assumed that the data collected on a daily basis represented the queue data on a regular basis in the long run. Considering a single server controlled by the Electronic Queue Management System, the duration of the service times that allows customers to get their anticipated services and the number of customers that are able to acquire service after every cycle time of 240 seconds ( 4 minutes). To simplify, four service channels, where one was deterministic and three as random service channels were considered to represent the several service channels in the model.

Data collected is provided in appendix C.
The parameters are computed as follows:

For three servers (c=3)
From tables 1, 2, 3 and 4 we compute the average effective deterministic service as

$$
\begin{aligned}
\mathrm{g}_{e} & =\frac{1}{4}\left(\frac{2742}{30}+\frac{2661}{30}+\frac{2695}{30}+\frac{3221}{30}\right) \\
& =\frac{1}{4}(378.296) \\
& =94.57 \mathrm{sec}
\end{aligned}
$$

The average arrival rate as

$$
\lambda=\frac{\text { Total arrivals }}{\text { Total number of hours observed in tables } 1,2,3 \text { and } 4}
$$

Total number of arrivals was obtained by adding customer arrivals in tables 1,2,3 and 4

$$
=\frac{11385}{120}
$$

$$
=94.875 \text { Customers per hour } \equiv 1.58125 \text { customers per minute } .
$$

The average service rate is

$$
\mu=\frac{\text { Total Departures }}{\text { Total number of hours observed in tables } 1,2,3 \text { and } 4}
$$

Total number of departures was obtained by adding customers served in tables 1, 2, 3, 4 and 5

$$
=\frac{10283}{120}
$$

$=85.6916$ Customers per hour $\equiv 1.4281$ customers per minute
The probability that a server is idle as

$$
P_{0}=\left\{1+\frac{(\lambda / \mu)^{1}}{1!}+\frac{(\lambda / \mu)^{2}}{2!}+\cdots+\frac{(\lambda / \mu)^{c-1}}{(c-1)!}+\frac{(\lambda / \mu)^{c}}{c!}\left[1+(\lambda / c \mu)+(\lambda / c \mu)^{2}+\cdots\right]\right\}^{-1}
$$

$$
\begin{gathered}
=\left\{1+1.1072+\frac{(1.1072)^{2}}{2!}+\frac{(1.1072)^{3}}{3!(1-0.369080)}\right\}^{-1} \\
=(3.1454075)^{-1} \\
=0.317923
\end{gathered}
$$

For two servers, $(\mathrm{C}=2)$ the probability that a server is idle is given as The average service rate where the first and the second server were considered,

$$
\mu=\frac{\text { Total Departures }}{\text { Total number of hours observed in tables } 1,2,3 \text { and } 4}
$$

Total number of departures was obtained by adding customers served by teller 1 and teller 2 in tables 1, 2, 3, 4 and 5

$$
=\frac{6770}{120}
$$

$$
=56.416667 \text { Customers per hour } \equiv 0.9402 \text { customers per minute }
$$

The probability that the server is idle as

$$
\begin{gathered}
P_{0}=\left\{1+\frac{(\lambda / \mu)^{1}}{1!}+\frac{(\lambda / \mu)^{2}}{2!}+\cdots+\frac{(\lambda / \mu)^{c-1}}{(c-1)!}+\frac{(\lambda / \mu)^{c}}{c!}\left[1+(\lambda / c \mu)+(\lambda / c \mu)^{2}+\cdots\right]\right\}^{-1} \\
=\left\{1+1.681823+\frac{(1.681823)^{2}}{2!(1-0.8409115)}\right\}^{-1} \\
=(11.573276)^{-1} \\
=0.0864
\end{gathered}
$$

The parameters of the model calculated indicate the queue behavior. It can be seen that the effective average deterministic service time is 94.57 despite having fixed the cycle time to be 4 minutes as the maximum time a customer should take to be served deterministically. The average arrival rate to the queue is calculated as well. This indicates approximately two
arrivals to the queue per minute while the average rate of departure is one customer per minute. The figures show that arrivals overwhelm departures. The probability that a server is idle was computed and obtained as 0.317923 for three servers and 0.0864 for two servers. It can be intuitively seen that due to more number of servers then the possibility of having an idle server at a given time is higher due to a higher work rate. If the servers are reduced to one then it is impossible to find the server idle given that the queue exists meaning that the probability that the server is idle reduces to zero.

### 4.4Validation of the model

To validate the model, simulation of the modeled waiting times in queues was provided. MATLAB software simulink functions for $\mathrm{E}\left(\mathrm{W}_{\mathrm{t}_{1}}\right), \mathrm{E}\left(\mathrm{W}_{\mathrm{t}_{2}}\right)$ and $\mathrm{E}\left(\mathrm{W}_{\mathrm{t}}\right)$ were carried out where three and two servers were considered. All the graphs depicted a direct proportionality where the average delays increase with an increase in the utilization factor where the utilization factor is the proportion of time when the servers are busy.

### 4.4.1 Simulation of $E\left(W_{t_{1}}\right)$

Using equation 4.1, the deterministic model $\mathrm{EW}_{\mathrm{t}_{1}}$ is described in figure 3 by MATLAB software.


Figure 3 Diagram representing simulation of deterministic component From figure 3 the deterministic delay model estimates a continuous delay but does not accommodate the aspect of randomness. The model reveals a steady increase in mean delay with a more increase in waiting when the flows approach capacity $\rho>1$ which consequently implies infinite delays in the long run queuing.

### 4.4.2 Simulation of $E\left(W_{t_{2}}\right)$ with two and three servers

Using equation 4.3the stochastic model $\mathrm{EW}_{\mathrm{t}_{2}}$ with three and two servers is described in figure 4 by MATLAB software.


Figure 4 Diagram representing simulation of stochastic component with two and three servers From figure 4, the stochastic delay models with three and two servers are represented as red and green respectively on the above graph. Delays tending to infinity when the arrival flows approach capacity $\rho>1$ are estimated. When arrival flows exceed capacity then oversaturated queues exist and continuous delays occur. However comparing the three server model and the two server model, it depicts an increased delay on the two server model which is as a result of decreased service channels.

### 4.4.4 Simulation of $E\left(W_{t}\right)$ With Three Servers

Using equation 4.4, $\mathrm{E}\left(\mathrm{W}_{\mathrm{t}}\right)$ was split into $\mathrm{EW}_{\mathrm{t}_{1}}$ and $\mathrm{EW}_{\mathrm{t}_{2}}$ as described in figure 5 by MATLLAB software when service times and arrival times follow exponential distributions with parameters $\frac{1}{\mu}$ and $\frac{1}{\lambda}$ respectively.


Figure 5 Diagram representing simulation of the overall model with three servers From figure 5 the stochastic delay model appropriately estimates delays before the flows approach capacity and the delay tends to infinity beyond $\rho>1$ making it an inappropriate model to estimate delays when used in isolation. When arrival flows exceed capacity, oversaturated queues exist and continuous delays occur. The deterministic delay model also depicts estimates of continuous uniform delay but it does not completely deal with the effect of randomness. The figure shows that both components of the overall delay model are compatible as it depicts a steady increase in waiting time accommodating both aspects of randomness as well as uniformity. Therefore the overall delay model is used to bridge the gap between the two models and also the overall model provides a reduced waiting time as
compared to the stochastic and deterministic models in isolation therefore it is a better model to model waiting time in queues.

### 4.4.5 Simulation of $E\left(W_{t}\right)$ with Two Servers

Similarly using equation 4.4 for two servers, $E\left(W_{t}\right)$ was split into $E W_{t_{1}}$ and $E W_{t_{2}}$ as described in figure 6 by MATLLAB software when service times and inter arrival times follow exponential distributions with parameters $\frac{1}{\mu}$ and $\frac{1}{\lambda}$ respectively.


Figure 6 diagram representing simulation of overall model with three servers From figure 6 the stochastic delay model appropriately estimates delays before the flows approach capacity and the delay tends to infinity beyond $\rho>1$ making it an inappropriate
model to estimate delays when used in isolation. It also shows that when arrival flows exceed capacity, oversaturated queues exist and continuous delays occur. The deterministic delay model also depicts that it estimates a continuous uniform delay but it does not completely deal with the effect of randomness. The figure shows that both components of the overall delay model are compatible as it depicts a steady increase in waiting time accommodating both aspects of randomness as well as uniformity. Therefore the overall delay model is used to bridge the gap between the two models and also the overall model provides a reduced waiting time as compared to the stochastic and deterministic models in isolation therefore it is a better model to estimate delays in queues. It is important to also note that Figure 6 indicates a generally increased waiting time as compared to Figure 5 which is explained by the reduced number of servers.

## CHAPTER FIVE

## CONCLUSION AND RECOMMENDATION

### 5.1 Introduction

This chapter presents the conclusions and recommendations from the study. Section 5.2 gives a conclusion derived from the results. Section 5.3 puts forward a recommendation from the study.

### 5.2 Conclusion

Considering the uniform and random properties of queues, the models for estimating deterministic and stochastic delay components in queue delays modeled waiting times in queues on the basis of a joint model incorporating stochastic and deterministic queue models. From the joint delay model obtained, real life data was applied to the model and it fitted well with the model. The model was validated by use of simulation for utilization factors ranging from 0.1 to 1.9 using MATLAB software simulink functions. The simulation results confirm that using the stochastic and the deterministic models in isolation when oversaturated conditions exist renders the models unrealistic, therefore modeling queues by combining the two models is more appropriate. The results also showed that delays increase when the servers are reduced indicating that the model is applicable in real life situations where queues develop.

### 5.3 Recommendation

The study recommends the model obtained that it would be useful if it would be utilized into practical situations when designing queue setups in organizations and/or institutions. Further studies should be conducted for customer flexibility as well as server flexibility type of queue models

## REFERENCES

Agbola, A., \& Salawu,R.O. (2008). Optimizing the use of information and communication Technology (ICT) in Nigerian Banks. Journal of internet babking and commerce, 415.

Bajpai, P. (2013). Measure of Performance of queing models and behavior of customers in rel life applications. International journal of applied physics and mathematics, 5, 3-5.

Bakari, H. (2014). Queing process and its applications to customer service delivery. International Journal of Mathematics and Statistics Invention, 188.

Banks,J.,Carson,J.,Nelson,B.L,and Nicol,D. (2001). Discrete Event system simulation. Prentice Hall International, 3, 1-30.

Beckmann,M.J., McGuire,C.B., \& WinstenC.B. (1956). Studies in the Economics in Transportation. Yale: Yale University Press.

Brewton, J. (1989). Customer experience as the strategic differntiator in retail banking. ABA Banking Journal, 3-7.

Çinlar, E. (1969). Markov Renewal Theory. Advances in Applied Probability Trust. JSTOR 1426216, 1 ( 2), 123-187.

Darroch, J. (1964). On theTraffi Light Queue. Statistics Journal, 380-388.
Emuoyibofarhe, O., Omotoso, I., Ozichi, A., \& and Popoola, O. (2005). Application of queing theory to the optimum control of the waiting times, idle times andqueue in banks. Ladoke Akintola University of Technology, 1-2.

Erlang. (1909). The theory of Probabilities and Telephone conversations. Copenhagen: Nyt Tiddskrift Matematika.

Gazis. (1974). Traffic Science. Intersection Publication Journal, 148-151.
Janos, S. (2010). Queing theory and its application. 8th International Conference on Applied Informatics. 1, pp. 9-30. Eger: University of Debrecen, Faculty of Informatics.

John, S. (2015). A Server Walking-Time Dependent Model. Theory of Queuing Models.
Kendall, D. (1953). Stochastic Processses occurring in the theory of queues and the annalysis method of the embeded markov chain. JSTOR, 8(4), 1-3.

Kimber,R., \& Hollis,E. (1979). Traffic Queues and Delays at Road Junctions. London: Britain Press.

Lindley, D. (1952). The theory of Queues with a single server. Mathematical Proceedings of the cambridge philosophical society(48(2)), 277-289.

Liping,F., \& Bruce,H. (1999). Delay Variability at Signalized Intersection. TransportationResearchRecord.

Little, J. (1961). Approximate Expected Delays fo rSeveral Maneuvers in Poisson Traffic. Journal of Operations Research, 39-52.

McNeil, D. (1968). A Solution to the Fixed Cycle Traffic Light Problem for Compound poisson arrivals. Journal of Applied Probability, 624-635.

Molina, C. E. (1927). Application of the theory of probability to telephone trunking problems. Bell system technical journal(4), 461-462.

Richa, S. (2014). Mathematical Analysis Of Queue With Phase Service. Journal of Statistical operation research, 1-6.

Tarko,A., Rouphail,N., \& Akcelik,R.B. (1993). Overflow Delay at a Signalized. Transportation Research Record, 82-89.

Teply,S., Allingham,D.I., Richardson,D.B., \& Stephens,B.W. (1995). Canadian Capacity Guide for Signalized Intersections. Institute Of Transportation Engineering.

Toshiba. (2013). International Journal of Advance Computational Engineering and Networking. Application of queueing theory for improvement of bank services, 1-3.

Wayne, L. W. (2003). Operations Research Applications and Algorithms (4th ed.). Brooks: Curt Hinrichs.

Wenny, C., \& Whitney, C. (2004). Determining bank tellerschedulling using simulation with changing arrival rates. Journal of Mathematics, 1-8.

You-Tong He. (2012). Model Flexibility In Queuing Systems. Chicago: Adventure works press.

Yusuff, S. (2010). Analysis of expected actual waiting time and service delivery. International Journal of Humanities and Social Studies, 398-401.

Zukerman, M. (2012). Introduction to Queueing Theory and Stochastic Teletraffic Models. City University of Hong Kong.

## Appendix A: MATLAB Iteration codes.

MATLAB Iteration code for simulating $E\left[W_{t_{1}}\right]$ versus $\rho$.
c_y=240;
g_e $=68.23$;
rho=0.1:0.2:2.0;
delta $=0.3556$;
\% \% mean of deterministic delay
\% \%
Ew_t_1=c_y*((1-g_e/c_y).^2)./(2.*(1-(g_e/c_y).*rho));
figure (5.1)
plot(rho,Ew_t_1,'g');
xlabel('utilization factor');
ylabel('mean of deterministic delay');

MATLAB Iteration code for simulating $E\left[W_{t_{2}}\right]$ versus $\rho$ with Three and Two Servers. p_o=0.317923;
rho=0.1:0.2:2.0;
mu_3=1.4281;
mu_2=0.9402;
c_3=3;
c_2=2;
\%\%mean of three server stochastic delay
\%\%
Ew_t_3=1/mu_3+(((c_3*rho).^c_3)/factorial(c_3).*c_3.*mu_3.*(1-rho).^2).*p_o;
$\% \%$ mean of two server stochastic delay
\%\%
Ew_t_2=1/mu_2+(((c_2*rho).^c_2)/factorial(c_2).*c_2.*mu_2.*(1-rho).^2).*p_o;
Figure (4)
plot(rho,Ew_t_3,'r');
hold on
plot(rho,Ew_t_2,'g');
xlabel('utilization factor')
ylabel('mean of stochastic delay')

## MATLAB Iteration code for simulating $E\left[W_{t}\right]$ versus $\rho$ with three servers.

c_y=4;
g _ $\mathrm{e}=1.576167$;
rho=0.1:0.2:2.0;
$\mathrm{mu}=1.4281$;
$\mathrm{c}=3$;
p_o=0.317923;
delta $=0.447427$;
$\% \%$ mean of deterministic delay
\%\%
Ew_t_1=c_y*((1-g_e/c_y).^2)./(2.*(1-(g_e/c_y).*rho));
\%\%mean of stochastic delay
$\% \%$
Ew_t_2=1/mu+(((c*rho).^c)/factorial(c).*c.*mu.*(1-rho).^2).*p_o;
\%\%mean of overall delay
\%\%
Ew_t=delta.*(c_y* $\left(\left(1-g_{-} e^{2} / c_{-} y\right) . \wedge 2\right) . /(2 . *(1-$
$\left.\left.\left(\mathrm{g} \_\mathrm{e} / \mathrm{c} \_\mathrm{y}\right) .{ }^{*} \mathrm{rho}\right)\right)+1 / \mathrm{mu}+(((\mathrm{c} * \mathrm{rho}) . \wedge \mathrm{c}) /$ factorial(c).*c.*mu.*(1-rho).^2).*p_o);
figure(5)
plot(rho,Ew_t_1,'g');
hold on
plot(rho,Ew_t_2,'b');
hold on
plot(rho,Ew_t,'r');
xlabel('utilization factor')
ylabel('mean of stochastic delay')
legend('E[w_t_1]','E[w_t_2]','E[w_t]');

## MATLAB Iteration code for simulating $E\left[W_{t}\right]$ versus $\rho$ with Two servers.

c_y=4;
g_e $=1.576167$;
rho=0.1:0.2:2.0;
$\mathrm{mu}=1.9482$;
$\mathrm{c}=2$;
p_o=0.0864;
delta $=0.447427$;
$\% \%$ mean of deterministic delay
\%\%
Ew_t_1=c_y*((1-g_e/c_y).^2)./(2.*(1-(g_e/c_y).*rho));
$\% \%$ mean of stochastic delay
\%\%
Ew_t_2=1/mu+(((c*rho).^c)/factorial(c).*c.*mu.*(1-rho).^2).*p_o;
\%\%mean of overall delay
\%\%
Ew_t=delta.*(c_y* $\left(\left(1-\mathrm{g} \_\mathrm{e} / \mathrm{c} \_\mathrm{y}\right) . \wedge 2\right) . /(2 . *(1-$
$\left.\left.\left(\mathrm{g} \_\mathrm{e} / \mathrm{c} \_\mathrm{y}\right) . * \mathrm{rho}\right)\right)+1 / \mathrm{mu}+(((\mathrm{c} *$ rho $) . \wedge \mathrm{c}) /$ factorial(c).*c.*mu.*(1-rho).^2).*p_o);
figure(6)
plot(rho,Ew_t_1,'g');
hold on
plot(rho,Ew_t_2,'b');
hold on
plot(rho,Ew_t,'r');
xlabel('utilization factor')
ylabel('mean of stochastic delay')
legend('E[w_t_1]','E[w_t_2]','E[w_t]');

## Appendix B: Data Tables

Table 1Data collected on $16^{\text {th }}$ July 2018 to $20^{\text {th }}$ July 2018

|  |  |  | Customers Served |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Day | Time Interval | Customer Arrivals | $\mathrm{g}_{\mathrm{e}}$ | Teller <br> 1 | Teller 2 | Teller 3 |
| Monday | 9:00-10:00 AM | 68 | 68 | 11 | 33 | 14 |
|  | 10:00-11:00 AM | 152 | 91 | 21 | 39 | 51 |
|  | 11:00-12:00 AM | 77 | 87 | 24 | 17 | 28 |
|  | 12:00-1:00 PM | 92 | 120 | 17 | 38 | 30 |
|  | 1:00-2:00 PM | 87 | 68 | 30 | 20 | 23 |
|  | 2:00-3:00 PM | 119 | 62 | 47 | 25 | 34 |
| Tuesday | 9:00-10:00 AM | 110 | 170 | 43 | 27 | 25 |
|  | 10:00-11:00 AM | 112 | 111 | 23 | 31 | 39 |
|  | 11:00-12:00 AM | 53 | 97 | 1 | 24 | 23 |
|  | 12:00-1:00 PM | 103 | 59 | 11 | 36 | 28 |
|  | 1:00-2:00 PM | 82 | 77 | 19 | 31 | 36 |
|  | 2:00-3:00 PM | 109 | 205 | 10 | 27 | 35 |
| Wednesday | 9:00-10:00 AM | 46 | 63 | 17 | 19 | 18 |
|  | 10:00-11:00 AM | 82 | 79 | 23 | 21 | 35 |
|  | 11:00-12:00 AM | 52 | 120 | 22 | 23 | 14 |
|  | 12:00-1:00 PM | 53 | 230 | 17 | 33 | 11 |
|  | 1:00-2:00 PM | 98 | 197 | 17 | 39 | 33 |
|  | 2:00-3:00 PM | 109 | 80 | 10 | 35 | 27 |
| Thursday | 9:00-10:00 AM | 71 | 73 | 21 | 32 | 21 |
|  | 10:00-11:00 AM | 64 | 131 | 20 | 5 | 21 |
|  | 11:00-12:00 AM | 76 | 71 | 28 | 27 | 15 |
|  | 12:00-1:00 PM | 54 | 68 | 8 | 31 | 19 |
|  | 1:00-2:00 PM | 76 | 206 | 24 | 17 | 28 |
|  | 2:00-3:00 PM | 132 | 70 | 21 | 45 | 54 |
|  |  |  |  |  |  |  |
| Friday | 9:00-10:00 AM | 112 | 72 | 30 | 36 | 48 |
|  | 10:00-11:00 AM | 90 | 151 | 30 | 28 | 26 |
|  | 11:00-12:00 AM | 162 | 68 | 31 | 59 | 53 |
|  | 12:00-1:00 PM | 90 | 86 | 35 | 29 | 19 |
|  | 1:00-2:00 PM | 98 | 68 | 38 | 22 | 38 |
|  | 2:00-3:00 PM | 113 | 93 | 30 | 31 | 22 |
| Total |  | 2742 | $\begin{aligned} & \hline 314 \\ & 1 \end{aligned}$ | 679 | 880 | 868 |

Table 2 Data collected on $23^{\text {rd }}$ July 2018 to $27^{\text {th }}$ July 2018

|  |  |  | Customers Served |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Day | Time Interval | Customer Arrivals | $\mathrm{g}_{\mathrm{e}}$ | Teller 1 | Teller 2 | Teller 3 |
| Monday | 9:00-10:00 AM | 104 | 107 | 36 | 42 | 17 |
|  | 10:00-11:00 AM | 113 | 61 | 27 | 38 | 28 |
|  | 11:00-12:00 AM | 125 | 78 | 61 | 53 | 41 |
|  | 12:00-1:00 PM | 144 | 102 | 10 | 35 | 27 |
|  | 1:00-2:00 PM | 77 | 64 | 38 | 26 | 17 |
|  | 2:00-3:00 PM | 74 | 62 | 26 | 23 | 31 |
| Tuesday | 9:00-10:00 AM | 92 | 107 | 39 | 23 | 31 |
|  | 10:00-11:00 AM | 81 | 111 | 29 | 28 | 15 |
|  | 11:00-12:00 AM | 87 | 70 | 35 | 29 | 19 |
|  | 12:00-1:00 PM | 121 | 64 | 27 | 16 | 24 |
|  | 1:00-2:00 PM | 119 | 68 | 36 | 26 | 35 |
|  | 2:00-3:00 PM | 101 | 65 | 38 | 22 | 38 |
| Wednesday | 9:00-10:00 AM | 110 | 61 | 10 | 35 | 27 |
|  | 10:00-11:00 AM | 104 | 151 | 20 | 17 | 24 |
|  | 11:00-12:00 AM | 112 | 62 | 29 | 32 | 24 |
|  | 12:00-1:00 PM | 82 | 64 | 23 | 21 | 35 |
|  | 1:00-2:00 PM | 53 | 157 | 22 | 23 | 14 |
|  | 2:00-3:00 PM | 68 | 68 | 39 | 20 | 19 |
| Thursday | 9:00-10:00 AM | 103 | 65 | 11 | 36 | 41 |
|  | 10:00-11:00 AM | 74 | 70 | 21 | 32 | 21 |
|  | 11:00-12:00 AM | 82 | 113 | 30 | 37 | 25 |
|  | 12:00-1:00 PM | 55 | 62 | 6 | 13 | 21 |
|  | 1:00-2:00 PM | 109 | 70 | 38 | 47 | 19 |
|  | 2:00-3:00 PM | 80 | 61 | 13 | 23 | 21 |
| Friday | 9:00-10:00 AM | 46 | 73 | 17 | 19 | 18 |
|  | 10:00-11:00 AM | 27 | 179 | 6 | 21 | 17 |
|  | 11:00-12:00 AM | 82 | 64 | 23 | 21 | 35 |
|  | 12:00-1:00 PM | 94 | 203 | 38 | 23 | 35 |
|  | 1:00-2:00 PM | 52 | 70 | 22 | 23 | 14 |
|  | 2:00-3:00 PM | 90 | 60 | 30 | 28 | 26 |
| Total |  | 2661 | 2612 | 800 | 832 | 759 |

Table 3 Data collected on $30{ }^{\text {th }}$ July to $3^{\text {rd }}$ August 2018

|  |  |  | Customers Served |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Day | Time Interval | Customer Arrivals | $\mathrm{g}_{\mathrm{e}}$ | $\begin{array}{\|l\|} \hline \text { Telle } \\ \text { r } 1 \end{array}$ | Teller 2 | Teller 3 |
| Monday | 9:00-10:00 AM | 98 | 71 | 37 | 20 | 35 |
|  | 10:00-11:00 AM | 113 | 93 | 28 | 34 | 32 |
|  | 11:00-12:00 AM | 132 | 63 | 31 | 20 | 33 |
|  | 12:00-1:00 PM | 98 | 209 | 22 | 36 | 39 |
|  | 1:00-2:00 PM | 115 | 72 | 49 | 27 | 36 |
|  | 2:00-3:00 PM | 90 | 56 | 30 | 28 | 26 |
| Tuesday | 9:00-10:00 AM | 98 | 118 | 38 | 22 | 38 |
|  | 10:00-11:00 AM | 62 | 162 | 32 | 61 | 50 |
|  | 11:00-12:00 AM | 76 | 76 | 28 | 27 | 15 |
|  | 12:00-1:00 PM | 90 | 90 | 17 | 31 | 47 |
|  | 1:00-2:00 PM | 40 | 40 | 19 | 10 | 14 |
|  | 2:00-3:00 PM | 112 | 112 | 30 | 31 | 23 |
| Wednesday | 9:00-10:00 AM | 111 | 67 | 36 | 27 | 34 |
|  | 10:00-11:00 AM | 132 | 111 | 27 | 33 | 35 |
|  | 11:00-12:00 AM | 110 | 73 | 26 | 29 | 33 |
|  | 12:00-1:00 PM | 76 | 140 | 29 | 17 | 10 |
|  | 1:00-2:00 PM | 98 | 56 | 32 | 27 | 26 |
|  | 2:00-3:00 PM | 54 | 157 | 6 | 13 | 21 |
| Thursday | 9:00-10:00 AM | 88 | 68 | 16 | 39 | 33 |
|  | 10:00-11:00 AM | 76 | 221 | 23 | 19 | 28 |
|  | 11:00-12:00 AM | 75 | 61 | 28 | 26 | 17 |
|  | 12:00-1:00 PM | 64 | 189 | 17 | 19 | 21 |
|  | 1:00-2:00 PM | 82 | 68 | 23 | 21 | 35 |
|  | 2:00-3:00 PM | 71 | 181 | 21 | 30 | 23 |
|  |  |  |  |  |  |  |
| Friday | 9:00-10:00 AM | 64 | 63 | 13 | 23 | 20 |
|  | 10:00-11:00 AM | 109 | 17 | 28 | 34 | 37 |
|  | 11:00-12:00 AM | 80 | 67 | 33 | 31 | 21 |
|  | 12:00-1:00 PM | 98 | 115 | 39 | 23 | 39 |
|  | 1:00-2:00 PM | 130 | 64 | 21 | 45 | 54 |
|  | 2:00-3:00 PM | 53 | 107 | 10 | 27 | 21 |
| Total |  | 2695 | 2987 | 789 | 830 | 896 |

Table 4 Data collected on $6{ }^{\text {th }}$ August to $10{ }^{\text {th }}$ August 2018

|  |  |  | Customers Served |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Day | Time Interval | Customer Arrivals | ge | Teller $1$ | Teller 2 | Teller 3 |
| Monday | 9:00-10:00 AM | 110 | 51 | 29 | 34 | 26 |
|  | 10:00-11:00 AM | 153 | 170 | 27 | 36 | 30 |
|  | 11:00-12:00 AM | 120 | 52 | 64 | 59 | 72 |
|  | 12:00-1:00 PM | 110 | 59 | 45 | 26 | 25 |
|  | 1:00-2:00 PM | 117 | 107 | 27 | 39 | 31 |
|  | 2:00-3:00 PM | 132 | 53 | 46 | 59 | 19 |
| Tuesday | 9:00-10:00 AM | 90 | 64 | 31 | 40 | 18 |
|  | 10:00-11:00 AM | 111 | 191 | 37 | 26 | 31 |
|  | 11:00-12:00 AM | 124 | 70 | 49 | 29 | 35 |
|  | 12:00-1:00 PM | 112 | 117 | 30 | 36 | 41 |
|  | 1:00-2:00 PM | 74 | 73 | 10 | 31 | 35 |
|  | 2:00-3:00 PM | 40 | 105 | 17 | 19 | 18 |
| Wednesday | 9:00-10:00 AM | 99 | 93 | 27 | 29 | 22 |
|  | 10:00-11:00 AM | 90 | 123 | 13 | 19 | 21 |
|  | 11:00-12:00 AM | 87 | 62 | 27 | 11 | 27 |
|  | 12:00-1:00 PM | 76 | 113 | 29 | 17 | 10 |
|  | 1:00-2:00 PM | 130 | 60 | 19 | 46 | 59 |
|  | 2:00-3:00 PM | 162 | 187 | 33 | 48 | 61 |
| Thursday | 9:00-10:00 AM | 113 | 202 | 36 | 29 | 31 |
|  | 10:00-11:00 AM | 98 | 231 | 31 | 25 | 28 |
|  | 11:00-12:00 AM | 120 | 79 | 27 | 30 | 39 |
|  | 12:00-1:00 PM | 90 | 73 | 47 | 37 | 31 |
|  | 1:00-2:00 PM | 120 | 111 | 23 | 47 | 58 |
|  | 2:00-3:00 PM | 115 | 72 | 31 | 37 | 49 |
|  |  |  |  |  |  |  |
| Friday | 9:00-10:00 AM | 113 | 67 | 31 | 36 | 29 |
|  | 10:00-11:00 AM | 98 | 115 | 28 | 29 | 26 |
|  | 11:00-12:00 AM | 131 | 70 | 31 | 37 | 43 |
|  | 12:00-1:00 PM | 132 | 198 | 23 | 29 | 31 |
|  | 1:00-2:00 PM | 107 | 73 | 38 | 47 | 19 |
|  | 2:00-3:00 PM | 113 | 170 | 30 | 37 | 25 |
| Totals |  | 3287 | $\begin{aligned} & 321 \\ & 1 \end{aligned}$ | 936 | 1024 | 990 |

