APPLICATION OF QUEUING THOERY TO ENHANCE THE OPERATIONAL EFFICIENCY OF THE BANK

By

Aruna T Hettiarachchi

108656E

This dissertation submitted in partial fulfillment of the requirement for the

Master of Science Degree in Operational Research

Faculty of Engineering

Department of Mathematics

University of Moratuwa

NOVEMBER 2015

Declaration of the Candidate and Supervisor

The work submitted in this dissertation is the result of my own investigation, except where otherwise stated.

It has not already been accepted for any degree, and is also not been concurrently submitted for any submitted for any other degree

Aruna T Hettiarachchi Date

I endorse the declaration by the candidate,

Mr. T M J A Cooray - Project Supervisor Date

B.Sc (sp) Mathematics (Moratuwa), PG Dip Maths (Peradeniya), M.Sc (Colombo),

M.Phil (Moratuwa)

Senior Lecturer

Faculty of Engineering

Department of Mathematics

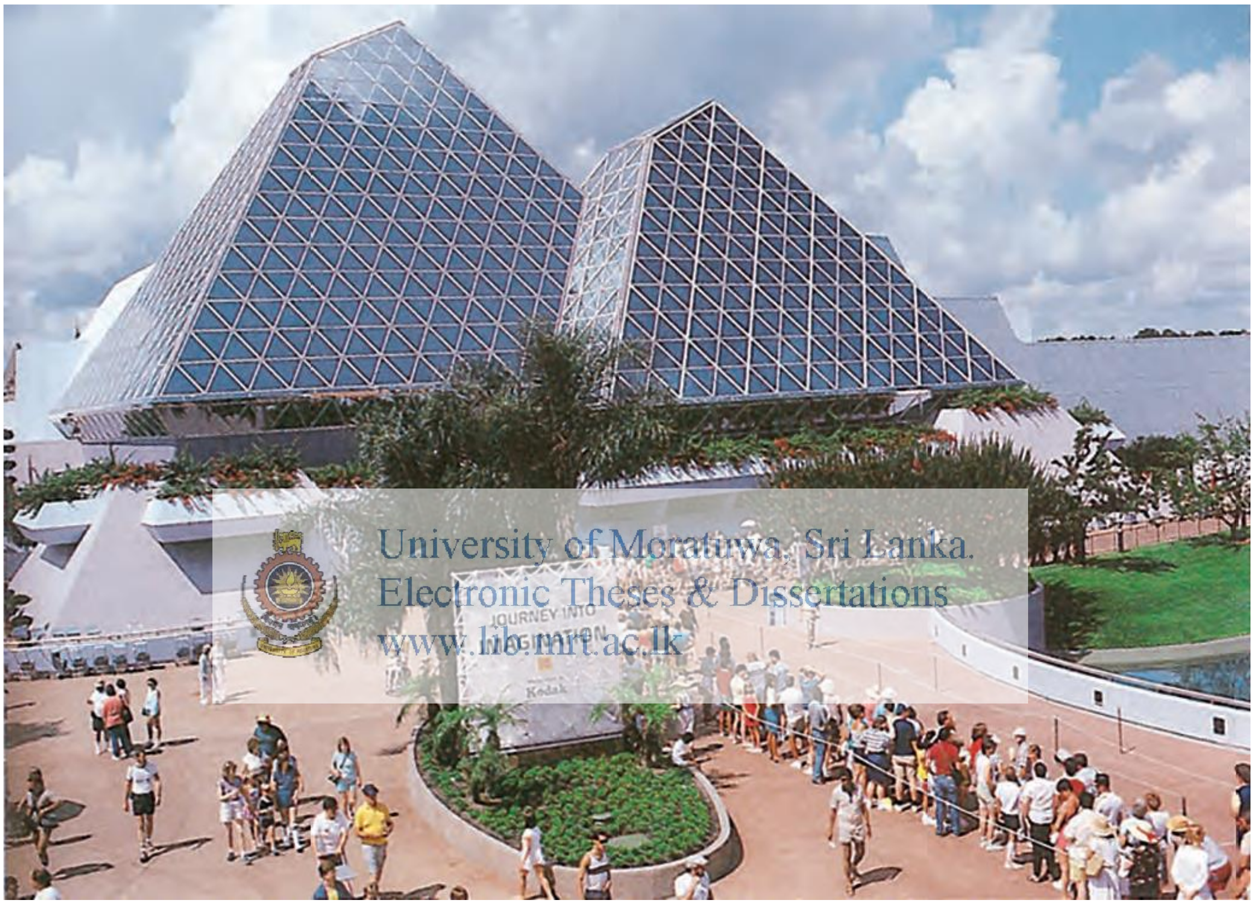
University of Moratuwa.

DEDICATION

My parents, my brother and two sisters

&

My wife Sonali and my daughter Ranudi


JOURNEY INTO
Kodak

ABSTRACT

This study reviews the applications of queuing theory to the field of banking queue management problems. This review proposes a system of classification of queues in the banking sectors, which examined with the assistance of queuing models. The areas described in the literature are the common problems encountered in the queue management strategies in the banking industry. The goal is to identify the best effective method to reduce customer-waiting time at the bank to the maximum possible standard while improving the efficiency of the bank.

Customer satisfaction is a concern to service industries as customers expect to get their service promptly. For a service industry like a bank, there is a need for efficient bank Teller scheduling system that takes into account recognizing various customers' expectations.

This study concentered on the single channel waiting line systems with poison arrivals and exponential service times in Bank of Ceylon, City office, Bank of Ceylon Kuliyapitiya and Bank of Ceylon Bingiriya.

All above branches have separate queues for the separate counters. (Many lines)They use rule of thumb to open/close counters at the branches based on their experience.Consequent to our findings and calculations we have proved that having one line and many counters (One line) is effective than having many lines. Further, with respect to open/close of counters we have suggested queue probability tool. Queue probability is one of significant factor to determine to set up number of counters effectively.

Key words: Queuing Theory, Mathematical modeling, Bank, customers, inter arrival time, service time, queuing system, M/M/1/∞ Model, M/M/Z/∞ Model

CONTENTS

Chapter 1 Introduction Page Number

1.1 Banking industry in Sri Lanka : An overview…………………………………………...01

1.2 Introduction to the queues at the banks………………………………………………02

1.3 Background of the study ……………………………………………………………..06

1.4 Objective of the study…………………………………………………………………07

1.5 Significance of the study……………………………………………………………..07

1.6 Importance of the study………………………………………………………………07

1.7 Design of the study……………………………………………………………………08

1.8 Data Collection ………………………………………………………………………08

1.9 Contents of the study…………………………………………………………………08

Chapter 2 Literature Review ……………………………………………………………09

2.1 Introduction ……………………………………………………………………………09

2.2 Similar studies undertaken and past findings…………………………………………09

2.3 Evolution of Queuing Theory…………………………………………………………11

2.4 Conceptual Framework…………………………………………………………………11

Page Number

3. Chapter Methodology ……………………………………………………………….12

3.1 Introduction ……………………………………………………………………….12

3.2Key terms in the study 12

3.3 Queue discipline…………………………………………………………………14

3.4Behavior of arrivals ……………………………………………………………14

3.5 Inter arrival time…………………………………………………………………15

3.6The exponential distribution for inter-arrival times ………………………………16

3.7Characteristics of a waiting line in a bank ……………………………………………16

3.8Elements of waiting lines ………………………………………………………16

3.9Arrival Characteristics ……………………………………………………………17

3.10 Labels for queuing models ………………………………………………………19

3.11 Service……………………………………………………………………………20

3.12Service time distributions ………………………………………………………20

3.13Service pattern …………………………………………………………………20

3.14Capacity of the system …………………………………………………………21

3.15The number of waiting lines ………………………………………………….21

3.16Waiting line performance measures………………………………………………22

3.17An Elementary Queuing Process……………………………………………….24

3.18Steady-state condition …………………………………………………………24

3.19Queuing Models formulation ……………………………………………….25

Page Number

3.20M/M/Z/∞ Model ……………………………………………………………27

3.21The number of servers ………………………………………………………30

3.22 Some insights about designing queuing system in a bank …………………………32

3.23Chapter summary ……………………………………………………………32

Chapter 4 Analysis ……………………………………………………………33

4.1 Introduction 33

4.2 Analysis of collected data……………………………………………………33

4.3 Rush hours at the bank ………………………………………………………39

4.4 Chi-squared test for goodness of fit……………………………………………45

4.5 Optimal service station…………………………………………………………47

4.6 Why use this model …………………………………………………………48

4.7 Chapter summary ……………………………………………………………48

Chapter 5 Results conclusion and Recommendations ……………………………49

5.1 Limitations……………………………………………………………………53

References ……………………………………………………………………54

LIST OF TABLES

Page Number

Table 4.1:Bingiriya branch figures on a first day of a week ……………………………33

Table 4.2:Bingiriya branch figures on a second day of a week …………………………34

Table 4.3:Bingiriya branch figures on a third day of a week ……………………………34

Table 4.4:Bingiriya branch figures on a fourth day of a week …………………………35

Table 4.5:Bingiriya branch figures on a fifth day of a week ……………………………35

Table 4.6:Summarized figures of table 4.1 to 4.5 ………………………………………36

Table 4.7: Final figures of three branches ………………………………………………37

Table 4.8: Calculation of queue performance at City office 37

Table 4.9: Calculations of queue performance at Kuliyapitiya branch …………………37

Table 4.10: Calculations of queue performance at Bingiriya branch……………………38

Table 4.11:Bingiriya branch figures at rush time on the first day of a week……………39

Table 4.12:Bingiriya branch figures at rush time on the second day of a week …………39

Table 4.13:Bingiriya branch figures at rush time on the third day of a week……………40

Table 4.14:Bingiriya branch figures at rush time on the fourth day of a week…………..40

Table 4.15:Bingiriya branch figures at rush time on the fifth day of a week……………41

Table 4.16:Bingiriya branch summarize figures at rush time……………………………41

Table 4.17: Calculation of queue performance at Bingiriya branch at rush time 41

Table 4.18: Calculation of Chi-squared test for goodness of Bingiriya branch…………45

Table 4.19: Calculation of Chi-squared test for goodness of Bingiriya branch…………46

at rush time

Table 4.20: Calculations of queue probability at City office ……………………………47

Table 4.21: Calculations of queue probability at kuliyapitiya branch……………………47

Table 4.22: Calculations of queue probability at Bingiriya branch………………………48

Table 5.1: Calculations of optimal number of servers at City office…………………….51

Table 5.2: Calculations of optimal number of servers at Kuliyapitiya ………………….51

Table 5.3: Calculations of optimal number of servers at Bingiriya………………………51

LIST OF FIGURES

Page Number

Figure 3.1: Probability distribution Poisson with λ = 2 18

Figure 3.2: Probability distribution Poisson with λ = 4 …………………………………19

Figure 4.1: Total arrivals of the Monday at Bingiriya branch - hourly wise …………….42

Figure 4.2: Total arrivals of the Tuesday at Bingiriya branch - hourly wise…………….42

Figure 4.3: Total arrivals of the Wednesday at Bingiriya branch - hourly wise…………43

Figure 4.4: Total arrivals of the Thursday at Bingiriya branch - hourly wise……………43

Figure 4.5: Total arrivals of the Friday at Bingiriya branch - hourly wise ………………44

Figure 4.6: Total arrivals of the Bingiriya branch- day wise of the week……………….44

ABBREVIATIONS

Abbreviations	Description
ATM	Automated Teller Machine
PLC	Public Limited Company

CHAPTER 01

INTRODUCTION

1.1 Banking industry in Sri Lanka: An overview

The banking sector in Sri Lanka is one of the most dynamic and vibrant sectors of the economy with developments taking place during the last decade in terms of institutions, instruments, range of services, and the geographical coverage. Financial sector reforms have introduced to improve efficiency and stability of the sector and further reforms are underway. Accordingly, the financial sector in Sri Lanka is the backbone of the economy that ensures the smooth functioning of the economy. As in many countries, banks are dominant players within the financial sector since banks are the only organizations that mobilize the large amount of deposits and provide funds for the future developments. Therefore, having sound banking system is one of the significant requirements to a country.

In the early 1950's the banking industry was limited to 16 commercial banks, out of which nine were foreign banks and held more than 60% of banking assets. None of these commercial banks was interested in providing value added customer service, since they had a monopoly at that time. At the same time, those banks were not interested in extending their service to rural areas in Sri Lanka, which comprise of 80% of the population. They were essentially graces to serve the plantation industry and the import export trade. Even though commercial banks were existed at that time, remarkable development appeared only after financial liberalization happened as the result of introducing open economic policies in 1977.

Bank of Ceylon and Peoples bank are the largest government owned banks in Sri Lanka. According to the Central Bank annual report in the year 2013, the banking industry in Sri Lanka comprised with 24 Commercial banks, 09 licensed commercial banks and 47-licensed financial companies. This has been functioning through 1676 branches and 3898 other out lets of Commercial banks. There were 527 branches and 273 other outlets of licensed commercial banks. Total number of financial companies was 212. Furthermore, there were 4622 automated teller machines, and 27689 point of sales machines.

1.2 Introduction to the queues at the Bank

In generally when we consider our day-to-day activities, one of our most inconvenience activities may be "waiting in line" (WAITING IN A QUEUE) until we get our chance. Queue means two or more people or things waiting in a line to purchase some goods or to receive a service. Queues are part of our everyday life. People will wait in queues to buy a ticket to watch a cricket match at the grounds, to pay what they bought from the supermarket, to obtain food at restaurant and **at a bank, deposit as well as withdraw money and other related work**. Lining-up for one service or the other is an inevitable phenomenon of our life. Though it look common, but demands much attention because excessive waiting is very costly in some sense, social cost, loss of customers to the organization, the cost of idle employees or some other important cost.

Waiting in lines is part of everyday life. To get a service people have to wait for some time until the service counter gets free if the counter is busy. Whether it is waiting in line at a grocery store to buy daily items (by taking a number) or checking out at the cash registers (finding the quickest line) waiting in line at the bank for a teller, or waiting at an amusement park to go on the newest ride, we spend a lot of time waiting. We wait in lines at the movies, campus dining room, and university library, the motor vehicle register's office for car registration at the division of motor vehicles, and even at the end of the school term to return books back. This describes waiting lines are how much related to our daily activities.

The queue is where customers wait before they served. The maximum permissible number of customers that it can contain characterizes a queue. Queues may infinite or finite, according to whether this number is infinite or finite. The assumption of an infinite queue is the standard one for most queuing models, even for situations where there actually is a finite upper bound on the permissible number of customers, because dealing with such an upper bound would be a complicating factor in the analysis. However, for queuing systems where this upper bound is small enough that it actually would reach with some frequency; it becomes necessary to assume a finite queue. Abandonment in queues for a longer period hasrecognized as having a significant impact on system performance.

Common queuing situations are as follows.

SITUATION	ARRIVALS IN QUEUE	SERVICE PROCESS
Supermarket	shoppers	Cashiers
Highway toll booth	Vehicles	Payment of tolls at booth
Doctor's office	Patients	Treatment
Computer system	Programs to run	Computer processes jobs
Bank	**Customers**	**Transactions handled by tellers**

Some banks are taking every endeavor to minimize the customer waiting time and some banks do not pay more attention over this task. Especially in the service industry, an organization like a bank, it is one of the significant requirement is having efficient and effective queue management system which takes into account varying customer service demand levels.

Many commercial banks have done great effort to increase the service efficiency and customer satisfaction but the most of them are facing a serious problem of waiting line of customers. In bank, the waiting line of customers appears due to low efficiency of the queuing system, it reflects the lacking of the business philosophy of customer centric, low service rate of the system. The waiting queues of the customer develop because the service to a customer may not deliver immediately as the customer reaches the service facility. Lack of satisfactory service facility would cause the waiting line of customers to form. The only technique is that the service demand can be met with ease is to increase the service capacity and increasing the efficiency of the existing capacity to a higher level.

As service speeds up, time spent waiting on queue decreases. Service cost however, increases as the number of service stations increase. The goal of managers is to schedule as few employees as possible while maintaining a maximum customer service level. Managers should make every endeavor to make the queues short enough so that customers are not dissatisfy and either leaves without transacting their business or transact at once and never returning the future. However, some waiting can allowed if the waiting cost is balance with significant saving in service cost.

In most of thebanks, customer and service information are identify generally based on manual observation and personal judgment (Obtained through information from visits to some local banks). This gives inaccurate results and wastes time. It also requires continuous observation by management personnel and thus results in additional cost. These results have greater possibility to make some customers being dissatisfied as customers who came first may be serving last.

Profit maximization objective may not be easy to achieve in banking, without providing a good level of customer service exceeding of their expectations. In other words, the faster they are attending to, the more the customer would be encouraged to keep their money with a bank. The only technique is that service demand can meet with ease to increase the service level.

This research and study has completely based on theories laid down in the Queuing Theory. Queuing theory is a mathematical approach applied to the analysis of waiting lines. It uses models to represent the various types of queuing systems. Formula for each model indicates how the related queuing system should perform, under a variety of conditions. The queuing model are very powerful tool for determining that how to manage a queuing system in the most effective manner. The queuing theory is known as the random system theory, which studies the content of the behavior problems, the optimization problem, and the statistical inference of queuing system.

Queuing is a challenge for all the branches in the Banking industry. In the developed world, considerable research has effected on how to improve queuing systems in various Banks. Unfortunately, above situation has not occurred in the case of developing countries like Sri Lanka. This study seeks to contribute to this subject by analyzing the queuing situation in public banks in Sri Lanka and to bring its practical value to how decision-making can enhanced in banking industry. Queuing theory is a potent mathematical approach to the analysis of waiting lines performance parameters in the queue management systems. It has increasingly become a common management tool for decision making in the developed world. This vital tool is unfortunately minimally used in most banks in Sri Lanka.It has also been used in reducing costs relating to various aspects of banking industry.

Long waiting queues are symptomatic of inefficiency in the banking industry or any other service industry. Unfortunately; this is the case in many public banks in Sri Lanka. Capacity management decisions in banking industry are arrived on experience and rule of the thumb rather than with the help of strategic research model-based analysis coupled with good mathematical approach.

All the larger banks in Sri Lanka receive a large number of customers in every day and this generally results in long customer waiting times. In response to this challenge, this research study analyses the queuing system of the banking industry in order to develop a model that can help reduce the waiting time of their customers. Specifically, this study seeks to construct a structural model of customer flow within the bank and to model a queuing system using the queuing theory to minimize customers waiting times in the bank.

The objective of queuing analysis is to offer a reasonably satisfactory service to waiting customers. In general, the approach of queuing theory is not an optimization technique but instead it determines the measure of performances of waiting lines, such as average waiting time in the queue and the productivity of the service facility, can used to design the service installation.

The applications of queuing theory extend well beyond waiting in line at a bank. It may take some creative thinking, but if there is any sort of scenario where time passes before a particular event occurs, there is probably some way to develop it into a queuing model. Queues are so commonplace in society that it is highly worthwhile to study them, even if only to reduce a few seconds off one’s wait in the queue

1.3 Background of the study

In service industries, such as banks demand for service is variable and often depends on the day of the month, day of the week, or even on the time of the day. However, service need to deliver promptly when it is needed. Daily work scheduling is required in many service companies such as hospitals and banks. Bad queue management process will result in long customer waiting times, long queues, and consequently, waiting cost. Bad scheduling can also result in loss of productively of Tellers due to idle times. On the other hand, good scheduling results in low waiting cost, good Teller utilization, customer satisfaction, and more profit. Managers are faced with the problem of recognizing the trade-off that must be taken between providing good service and the cost of waiting for such service.

As service speeds up, time spent waiting on queue decreases. Service cost however increases as the level of service increases. The goal of managers is to deploy as few employees as possible while maintaining a maximum customer service level. Manager's greater expectation is to ensure the queues to make shorter as much as possible with the view that customers not dissatisfied and either leaves without transacting their business or transact at once and never return in the future. However, some waiting can allowed if the waiting cost is not significantly impact with service cost.

A typical Teller-Customer scheduling system has two goals:

(i) To determine the minimum number of personnel to satisfy a set of service level requirements, and

(ii) To build a schedule that specifies when a staff should start shift so that, the staffing level requirements of each period of the day are met.

It is therefore important to recognize peak and non-peak periods to decide the staffing needs. However, very often, identifying the changes in the demand level from past or real- time data is not a straightforward matter because of the high variability. Using queuing theory with a control system that monitors the state of the system at all times, one can model and analyze a real time queuing situation in a bank, compare scheduling alternatives and the corresponding service levels, and provide the best scheduling rule based on the desired service level.

Waiting line models require an **arrival rate** and a **service rate.**

1.4 Objectives of the study

Objective of the study are as follows:

1. Apply queuing theory to determine the key measures of performance for a queuing system.
2. Check whether having one line and several counters is more efficient than having several lines.
3. Statistical model to decide how many servers should have to balance the queue.

1.5 Significance of the study

Through the study with the help of queuing models, we can make some suggestions to improve the service quality at the bank counters. At this junction minimize customer waiting time is the ultimate goal.

1.6 Importance of the study

Following are the significant outcomes of this study to enable the bank manager to improve the quality of the customer transaction at the bank.

1. The finding of the study will enable the researcher as well as the manager to develop a statistical model to use for the bank to enhance the efficient & effectiveness by providing speedy customer service.
2. This model will highlight the success factors of the reducing queues at the bank counters.
3. Bank management will be able to recommend the number of staff that should allocate to a branch/department and at what average rate that any individual could perform the service.
4. Offer a reasonably satisfactory service to waiting customers in the queues
5. Bank manager is able to operate minimum number of servers while providing efficient service.
6. Enhance the employee efficiency.

1.7 Design of the study

This study has designed to introduce a common method to minimize the customer waiting time at the bank counters. It is realized that no managers have taken steps to enhance the efficiency at the cash counters at the banks. Although they are very keen on reducing customer, waiting time but they are not applying scientific methods such as queuing theory. Every bank is suffering with this problem. With the view to enhance the efficiency at the cash counters at the banks this study is designed.

1.8 Data collection

Data was collected selecting three branches at Bank of Ceylon. Initially at City office branch at Colombo 1 and then Kuliyapitiya branch and at last Bingiriya branch. City office branch is very busy place in the entire week. Reason for the selecting of above branch is to ensure that the findings of the study will be remaining unchanged despite the location. There will be three counters at all the branches. Data in respect of customer arrival time and service time has extracted from above three branches during the year 2014. Two months data has taken from all above branches for the study.

1.9 Contents of study

Chapter1 describes the introduction to the banking industry in Sri Lanka. Further discuss background of the study, objective of the study, significance and importance of the study and basically cover the areas pertaining to the whole study. Similar studies taken on the same subject area have discussed in the chapter 2 under literature review. Methodology of the study have discussed in detail manner in the chapter 3. Chapter 4 describes the analysis of the collected data from the three designated branches. Through the analysis, it has proved that single line is more effective than having number of lines. Further based on the queue probability the number of counters to open in order to get best efficiency is discussing in this chapter. At last, results and conclusions are discussing in the chapter 5 for any user to implement the proposed model.

CHAPTER 02

LITERATURE REVIEW

2.1 Introduction

This chapter describes with previous researches undertaken in the queuing theory by various people. Many models have been proposed to explain queuing theory and its impact by different individuals. Most of these are about how to reduce waiting time in the counters. Some researchers have extended the research activities to different kind of working places to ensure what is the present queue management method and what are the steps that should take in order to enhance the effectiveness.

2.2 Similar studies undertaken and past findings

S.K.DharandTanzinaRahman (2013) have concluded that a research in similar manner by discussing the benefits of performing queuing analysis to a busy ATM in a Bank at Bangladesh. This research has discussed the application of queuing theory to the Bank ATM. From the result they have obtained that, the rate at which customers arrive in the queuing system is one customer per minute and the service rate is 1.50 customers per minute. The probability of buffer flow if there are three or more customers in the queue is 10 out of 100 customers. The probability of buffer overflow is the probability that customers will run away, because may be they are impatient to wait in the queue. This theory is also applicable for the bank, if they want to calculate all the data daily and this can applied to all branches ATM.

Chuka Emmanuel (2014) has engaged in analysis of the queuing system at First Bank PLC in Nigeria, and to show that the number of their servers was not adequate for the better customer's service. It observed that they need five servers instead of the three at present. This is the appropriate number of servers that can serve the customers as and at when due without waiting for long before customers are been served at the actual time necessary for the service. This increase in servers reduces the waiting time, and the probability that an arrival will have to wait for service is 0.056. However, the system utilization was observed to be 0.235 for an hour. Furthermore, the system capacity of the five servers was observed to be 92.7 for an hour.

It suggests a need to increase the number of servers in order to serve the customers better despite the staff allocation.

MathiasDharmawirya (2011) has concluded a research at restaurant to ensure losing their customers due to a long wait on the line. Some restaurants initially provide more waiting chairs than they actually need to put them in the safe side, and reducing the chairs as the time goes on safe space.

However, waiting chairs alone would not solve a problem when customers withdraw and go to the competitor's door and the service time may need to be improved. This shows a need of a numerical model for the restaurant management to understand the situation better. The researcher aimed to show that queuing theory satisfies the model when tested with a real-case scenario. He obtained the data from a restaurant in Jakarta. He has then derive the arrival rate, service rate, utilization rate, waiting time in queue and the probability of potential customers to balk based on the data using Little's Theorem and M/M/1 queuing model. The arrival rate at the restaurant its busiest period of the day is 2.22 customers per minute (cpm) while the service rate is 2.24 cpm. The average number of customers in the restaurant is 122 and the utilization period is 0.991.

EhsanBolandifar (2014) present an empirical study of an emergency department in which patients may leave the waiting area without being seen by a physician (LWBS patients). Using operational data from a hospital emergency department, they show that both time and number of patients in the waiting area significantly increase a patient's LWBS probability. These factors interact with each other in a non-linear fashion. In addition to these two factors, they showed that observed service rate affects LWBS probability, where the magnitude of its effect depends on waiting time. As waiting time increases, higher observed service rates may encourage patients to wait. They also examined the shape of the hazard rate curves for LWBS behavior. They use these findings to draw insights into modeling LWBS behavior. Further, they have discussed the state-of-the art for existing queuing and simulation models of abandonment and translate how their findings affect the utility of these models.

S.Vijay Prasad& others (2015) have engaged in a study and concluded that the total cost with assumption of certain Waiting cost in single queue and multiple queues and proved that the expected total cost is less for single queue and multi-server model as comparing with multi-queue multi-server model. The problem of customers waiting for the shortest time has studied by means of queuing models, the measure to reduce the time of the customer's queues to achieve the goal of the people oriented and the greatest effectiveness of the checkout stands of supermarkets.

The expected number for customers waiting in the queue is less in the case of single queue as compare with case of mulitiple queues. The expected waiting time of the customer in the queue is less in the case of single queue as compare with case of multiple queues. The expected waiting time of the customer in the system is less in the case of single queue as compare with case of multiple queues

2.3 Evolution of Queuing Theory

In 1908, Copenhagen Telephone Company requested Agner K. Erlang to work on the holding times in a telephone switch. He identified that the number of telephone conversations and telephone holding time fit into Poisson distribution and exponentially distributed. This was thebeginning of the study of queuing theory. Eight years later, he published a report addressing the delays in automatic dialing equipment. At the end of World War II, his early work extended to more problems that are general and to business applications of waiting lines.

The study of waiting lines, called queuing theory, is one of the oldest and most widely used quantitative analysis techniques. Subsequently, researchers, and industry experts have taken the facts incorporated in the theory and expanded the outcomes for different fields with the view to enhance service quality.

2.4 Conceptual Framework

A queuing system can simply described as customers arriving for service, waiting for service, if service not provided immediately and then leave the system after they served. The term "customer" used to denote a unit comes to obtain the service and this unit does not necessarily be a human. Some instances this unit may not be human and could be any item.

.

CHAPTER 03

METHODOLOGY

3.1 Introduction

The principal actors in a queuing theory are the customers and the servers. Customers will be generating from the environment. On arrival from at a service facility; they can start service immediately or wait in a queue if the facility is busy. When a facility completes a service, it automatically pulls a waiting customer, if another from the queue. If the queue is empty, the facility becomes idle until a new customer arrives.

From the standpoint of analyzing queues, the arrival of customers denoted by the inter-arrival time between successive customers, and the service described by the service time per customer. Generally, the interarrival and service time can be probabilistic, as in the case of bank counter and it is deterministic, as in the arrival of applicants for job interviews.

Queue size plays a vital role in the analysis of queues, and it may have a finite size, as in the buffer area between two successive machines.

Queue discipline, which represents the order in which customers selected from a queue, as an important factor in the analysis of queuing models.

3.2 Key terms in the study

Bank : An organization, which accepts money and pay money on behalf of their customers and providing other services based on money

Commercial service system: Providing commercial service to the general public

Constant service time: Every customer has the same service time

Customers: A generic terms that refers to whichever kind of entity is coming to the queuing system to receive service

Interarrival time: The elapsed time between consecutive to a queuing system

Finite queue: A queue that can hold only limited number of customers

Lack of memory property: When referring to arrivals, this property is that the time of the next arrival is completely uninfluenced by when the last arrival occurred

Number of customers in the queue: Number of customers who are waiting to receive the service

Number of customers in the system: The total number of customers in the queuing system, either waiting for the service to begin or currently served

Queue capacity: The maximum number of customers that can accommodate in the queue.

Queue discipline: The rule for determining the order in which members of the queue select to begin service.

Queuing system: Customers gathered in a line to receive a service.

Server: Entity/Individual who provides the service to customers in the queue.

Service time: Time taken by the server to complete a single transaction.

3.3 Queue discipline

This means that how customers should selected from the waiting lines. This can be done on one of the four types given below.

i. First in first out (FIFO) or First come first served (FCFS) basis

   That is the customer who has been queuing the longest time is served first. This is the most common and fairest queue discipline everywhere occurs.

ii. Last in first out (LIFO) basis

   That is the customer who come last or the item that was stocked last is taken first.

iii. Service in random order (SIRO) basis

   For example, in telephone exchangers, the operator has no means of telling how long any caller has been trying to make the call and callers are selected at random.

iv. Service in priority order (PRI) basis

   For example, in the telegraph system urgent messages are sent before the ordinary messages.

The queuing behavior of customer plays a role in waiting line analysis.

3.4Behavior of arrivals

The arrival pattern of the customers to a bank could be determined by the inter-arrival times of successive customer's arrivals.

The following characters are also among the characteristics of interest in the arrival pattern:

i. Whether the customers arrive in singly or in batches,
ii. Whether an arriving customer, for some reason decides not to join the queue.

   That is, whether the **balking** occurs in the arrival pattern.

iii. Whether a customer already in the system decides to leave the system without obtaining the service. That is whether the **reneging** occurs in the arrival pattern.

iv. In the case of multiple queues in parallel it may be possible for a customer to change from the queue he/she first joined to another one if he/she thinks that this would improve his/her change of being through the system earlier.
This is known as **jockeying**.

Unless stated to the contrary it assumed that customers arrive singly and no balking, no reneging, and no jockeying occur in the arrival pattern.

Most queuing models assume that an arriving customer is a patient customer. Patient customers are people or machines that wait in the queue until they are served and do not switch between lines. Unfortunately, life is complicated by the fact that people have been known to balk or to renege. Customers who balkrefuses to join the waiting line because it is too long to suit their needs or interests. Reneging customers are those who enter the queue but then become impatient and leave without completing their transaction. Actually, both of these situations just serve to highlight the need for queuing theory and waiting-line analysis.

3.5 Inter-arrival time

The time between consecutive arrivals is referred to as the inter arrival time. High variability in the arrival times is common for the queuing system. Therefore it is usually impossible to predict just how long until the next customer will arrive. However, after gathering a lot more data, it does possible to do estimate the followings.

1. Estimate the expected number of arrivals per unit time. This quantity is normally referred as to the mean arrival rate.
2. Estimate the form of the probability distribution of inter-arrival time.

The mean of this distribution actually comes directly, since

λ = Mean arrival rate for customers coming to the queuing system

The mean of the probability distribution of inter-arrival time is

Expected arrival time = $1/\lambda$

3.6The exponential distribution for inter-arrival times

In most queuing situations, the arrival customers occur in a very random fashion. Randomness means that the occurrence of an event (arrival of a customer or complete of a service) is not influenced by the length of time that has elapsed since the occurrence of the last event.

Inter arrival and service times are described quantitatively in queuing models by the exponential distribution defined as follows.

$$f(t) = \mu e^{-\mu t}, \quad t > 0$$

For the exponential distribution $E(t) = \frac{1}{\mu}$ and $p\{t \le T\} = \int_0^T \mu e^{-\mu t} dt$

3.7 Characteristics of a waiting line in a bank

At the bank counters, waiting time is affected by the design of the waiting line system in that bank. A waiting line system (waiting lines) has defined by the following elements

1. Arrivals or inputs to the system. These have characteristics such as population size, behavior, and a statistical distribution.

2. Queue discipline, or the waiting line itself. Characteristics of the queue include whether it is limited or unlimited in length and the discipline of people or items in it.

3. Theservice facility at the bank. Its characteristics include its design and the statistical distribution of service times

3.8Elements of waiting lines

At any time there is more customer demand for a service then a waiting line occurs. Customers can be either humans or inanimate objects. Examples of objects that must wait in lines include a machine waiting for repair, a customer order waiting to be processed, subassemblies in a manufacturing plant (that is working process inventory), and electronic messages on the internet, and ships or railcars waiting for unloading.

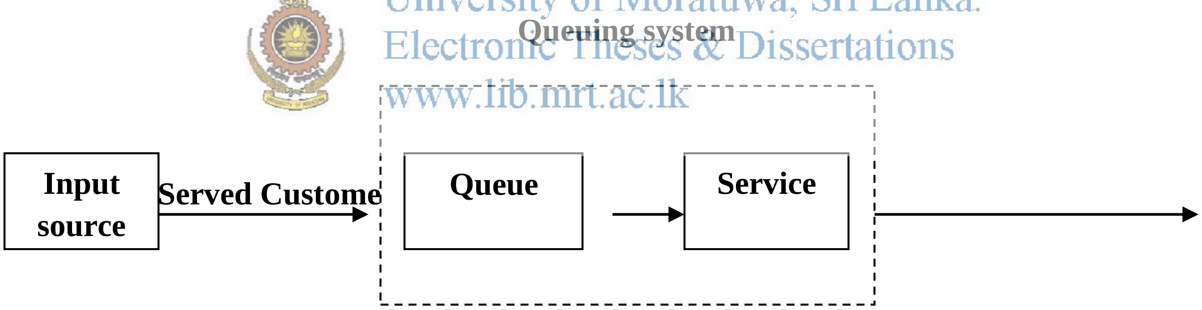
3.9Arrival Characteristics

The input source that generates arrivals or customers for a service system has three major characteristics.

1. Size of the arrival population.
2. Behavior of arrivals.
3. Pattern of arrivals (statistical distribution).

Size of the arrival (Source) population

Population sizes are considered either unlimited (essentially infinite) or limited (finite). When the number of customers or arrivals on hand at any given moment is just a small portion of all potential arrivals, the arrival population is considered **unlimited**, or **infinite**. Most queuing models assume such an infinite arrival population. An example of a **limited**, or **finite**, population is found in a copying shop that has, say, eight copying machines. Each of the copiers is a potential “customer” that may break down and require service



Pattern of Arrivals at the System

Customers arrive at a service facility either according to some known schedule (for example, one patient every 15 minutes or one student every half-hour) or else they arrive randomly. Arrivals are considered random when they are independent of one another and their occurrence cannot be predicted exactly. Frequently in queuing problems, the number of arrivals per unit time can estimate by a probability distribution known as the **Poisson distribution**. For any given arrival time a discrete Poisson distribution can be established by using the formula

$$p_{(x)} = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad for\ x = 0,1,2,3,4,$$

Where $P(x)$ = probability of x arrivals

x = number of arrivals per unit of time

λ = average arrival rate

e = 2.7183 (which is the base of the natural logarithms) with the help of the table which gives the value of $e^{-\lambda}$ for use in the Poisson distribution these values are easy to compute.

Following graphs illustrates the Poisson distribution for λ = 2 and λ = 4. This means that if the average arrival rate is λ = 2 customers per hour, the probability of 0 customers arriving in any random hour is about 13%, probability of 1 customer is about 27%, 2 customers about 27%, 3 customers about 18%, 4 customers about 9%, and so on. The chances that 9 or more will arrive are virtually zero. Arrivals, of course, are not always Poisson distributed (they may follow some other distribution). Patterns, therefore, should examine to make certain that they are well approximated by Poisson before that distribution is applied.

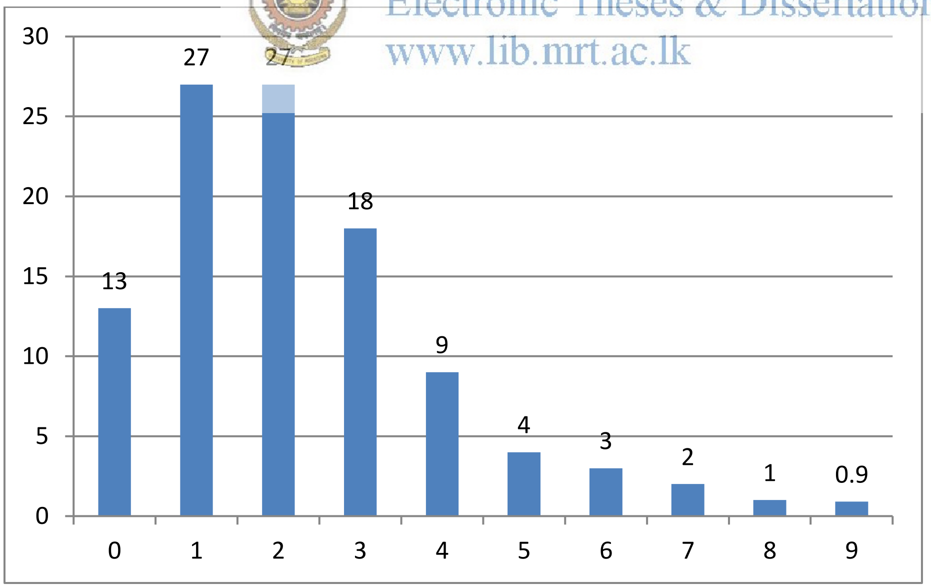


Figure: 3.1: Probability distribution Poisson with $\lambda = 2$

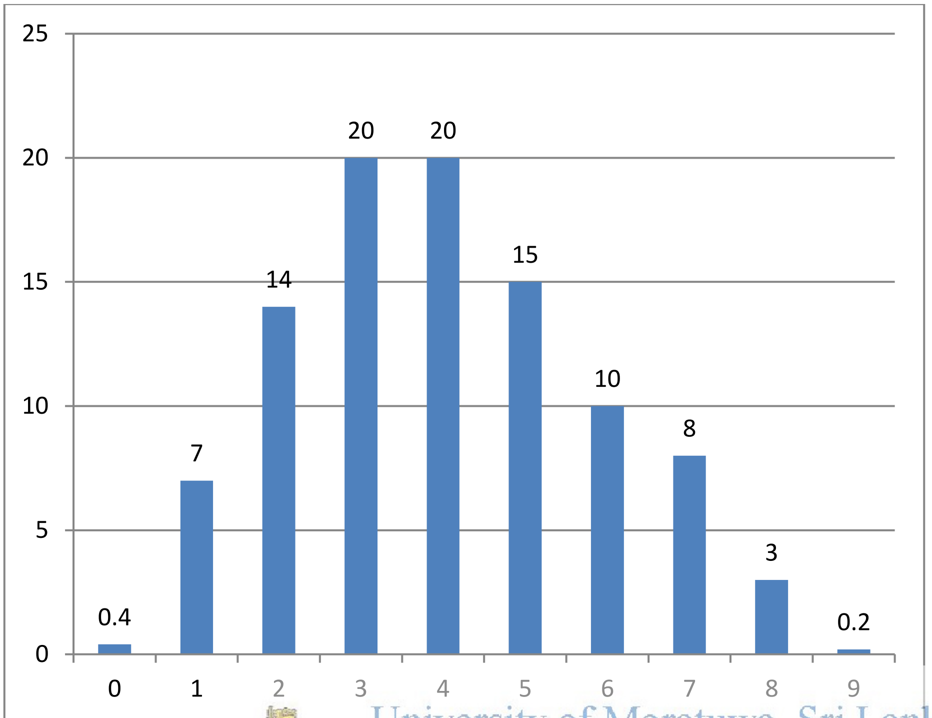
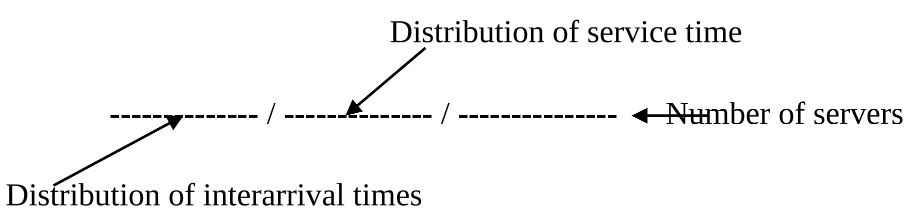


Fig: 3.2: Probability distribution Poisson with λ = 4

3.10 Labels of queuing models

To identify which probability distribution to use for service times and for interarrival times, a queuing model has labeled is follows.



The symbols used for the possible distributions (for the service times or interarrival times) are

M = Exponential distribution (Markovian)

D = Degenerate distribution (Constant times)

The M/M/1 is the single server model that assumes both interarrival times and service times have an exponential distribution. The M/M/2 model is the correspondence model with two servers. Letting “s” be the symbol that represents the number of servers, M/M/s is the correspondent model that permits any number of servers. Similarly, the M/D/s model hasexponential interarrival times, constant service times, and any desired number of servers.

Interarrival times also can have a degenerate distribution instead of an exponential distribution. The D/M/s model has constant interarrival times, exponential service times, and any desired number of servers.

3.11 Service

For a basic queuing system, one of the servers serves each customer individually. A system with one server is called as single server system and a system has servers more than one is called as multiple servers system. When a customer enters service, the elapsed time from the beginning to the end of the service is referred to as the service time. Service times generally vary from one customer to other. However basic queuing models assumes that the service time has a particular probability distribution, independent of which server is providing the service.

The symbol used for the mean of the service time distribution is $\frac{1}{\mu}$

μ = Expected number of service completed per unit time for a single continuously busy server

3.12 Service time distributions

The most popular choice for the probability distribution of service time is the exponential distribution, which has the same as above for the interarrival times. The main reason for this choice is that this distribution is much easier to analyze than any other distributions. Although this distribution provides an excellent for interarrival times for most situations, this is much less true for service times. Depending on the nature of the queuing system, the exponential distribution can provide a somehow similar distribution to the arrivals.

3.13 Service pattern

The service pattern could obtain by watching the system as to how the bank is providing the counter service to its customers. The service time is the time taken by a server to serve one customer. Also of interest is whether one server or more than one server attends a customer completely. Unless stated to the contrary it is assumed that a customer attended completely by one server. That is servers are not connected in series.

3.14 Capacity of the system

The system capacity is the maximum number of customers allowed in the queue/s as well as in service. It could be finite or infinite. If a customer arrives when the system is full, he/she is not allowed to enter the system or to wait the outside the system, but it is forced to leave without service.

The variability in arrival and service patterns causes waiting lines form when several customers request service at approximately the same time. This surge of customers temporarily overloads the service system and a line develops. Waiting line models that assess the performance of service systems usually assume that customers arrive according to a poison probability distribution and service time described by an exponential distribution. The poison probability distribution specifies the probability that a certain number of customers will arrive in a given time period (such as per hour). The exponential distribution describes the service times as the probability that a particular service time will be less than or equal to a given amount of time.

3.15 The number of waiting lines

Waiting line system can have single or multiple lines. Banks often have single lines for customers. Customers wait in line until a teller is free and then proceed to the teller's position. Other examples of single line systems include airline counters, rental car counters, restaurants, and amusement park attractions and call centers.

The advantage of using single line when multiple servers are available is the customer's perception of fairness in terms of equitable waits. That is the customer is not penalized by picking the line but is served in a true first come first served fashion. The single line approach eliminates jockeying behavior. Finally, a single line, multiple systems has better performance in terms of waiting times than the same system with the line for each server.

The multiple-line configuration is appropriate when specialized servers are used for when space considerations make a single line inconvenient. For example in a grocery store, some registers are expressed lanes for customers with a small number of items. Using express lines reduce the waiting time for customers making smaller purchases.

3.16 Waiting line performance measures

Performance measures used to gain useful information about waiting line systems. These measures include.

1. **The average number of customers waiting in line and the system.**

   The number of customers waiting in a line can interpreted in several ways. Short waiting lines can result from relatively constant customer arrivals (no major surges in demand) or by the organization having excess capacity (many cashiers open). On the other hand, long waiting lines can result from poor service efficiency, inadequate system capacity, and/or significant surges in demand.

2. **The average time customers spend for waiting, and the average time a customer spend in the system.**

   A customer often waits long waiting line to poor quality service. When long waiting times occur, one option may be to change demand pattern. That is, the company discounts or better service at less busy times of the day or week. For example, a restaurant offers early bird diners a discount so that demand is more level. The discount moves some demand from prime-time dining hours to the less desired dining hours.

If customers spend too much time in the system, might perceive the competency of the service provider as poor. For example, the amount of time customers spent in the line and in the system at a retail checkout counter can be a result of a new employee not yet proficient at handling the transactions.

3. **The system utilization rate**

Measuring capacity utilization shows the percentage of time the servers are busy. Management's goal is to have enough servers to assure that waiting is within allowable limits but not too many servers as to cost inefficient.

Changing operational characteristics

After calculating the operating characteristics for a waiting line system, sometimes you need to change the system to alter its performance. You can make changes to the different elements of the waiting line system are as follows.

- Arrival rates

You can try to change arrival rates in a number of ways. For example, you can provide discounts or run special promotions during the non-peak hours to attract customers.

- Number and type of service facilities

You can either increase or decrease the number of server facilities. For example, a grocery store can easily change the number of cashier open for business (up to the number of registers available). The grocery increase the number of cashiers open when lines are too long.

Another approach is to dedicate specific servers for specific transactions. One example would be to limit the number of items that can process at a particular cashier (ten items or less) or to limit a cashier to cash-only transactions. Still another possible is to install self-service checkout systems.

- Changing the number of phases

You can use the multi-phase system where servers specialize in a portion of the total service rather than needing to know the entire service provided. Since a server has fever tasks to learn, the individual server proficiency should improve .This goes back to the concept of division of labor.

3.17An Elementary Queuing Process

As already suggested, queuing theory has been applied to many different types of waiting-line situations. However, the most prevalent type of situation is the following. A single waiting line form in the front of a single service facility, within which are stationed one or more servers. One of the servers services each customer generated by an input source, perhaps after some waiting in the queue.

3.18Steady-state condition

After sufficient time has elapsed, the state of the queuing system becomes essentially independent of the initial state and the elapsed time. Then the system is said to be in steady state condition.

Notations for the steady state condition

P_n = probability of exactly n customers in the queuing system.

L_S = expected number of customers in queuing system

$$= \sum_{n=0}^{\infty} nP_n = \frac{\rho}{1-\rho}$$

L_q = expected queue length (excludes customers served)

$$L_q = L_s - \frac{\lambda}{\mu}$$

W_S = waiting time in system (includes service time) for each individual customer.

$$W_S = L_S / \lambda$$

W_q = waiting time in queue (excludes service time) for each individual customer.

$$W_q = L_q / \lambda$$

P_0 = Probability of zero units in the system (Idle time) = $1 - \frac{\lambda}{\mu}$

$P_{n>k}$ = Probability of having more than k units in the system, where n is the number of units in the system = $\left(\frac{\lambda}{\mu}\right)^{k+1}$

The relationships between L , W , L_q and W_q

$L_S = \lambda W_S$

This equation sometimes referred to as Little's formula.

$L_q = \lambda W_q$

Now assume that the mean service time is a constant, $1/\mu$ .Itthen follows that $W_S = W_q + \frac{1}{\mu}$

3.19 Queuing Models formulation

Model 1: (M/M/1)

This is the simplest queuing system to analyze. The system consists of only one server. The arrivals follow poison distribution with a mean arrival rate of λ and the service time has exponential distribution with the average service rate of μ .

$p_n = p(N = n),\ (n = 0, 1, 2,.....)$is the probability distribution of the queue length.

Utilization factor (The fraction of time that servers are busy) will be

$$\rho = \frac{\text{mean arrival rate}}{\text{mean service rate}} = \frac{\lambda}{\mu}$$

Derivation of the steady state equations

The probability that there will be n units (n>0) in the system at time $(t + \Delta t)$ may be expressed as the sum of three independent compound probabilities by using the fundamental properties of probability, poison arrivals and exponential service times.

The following are the three cases.

Time t No. of units	Arrival No. of units	ServiceNo.of units	Time(t+dt) (No. of units)
n	0	0	n
n-1	1	0	n
n+1	0	1	n

By adding above three independent compound probabilities, we obtain the probability of n units in the system at time $(t+\Delta t)$

$$p_n(t+\Delta t) = p_n(t)(1-(\lambda+\mu)\Delta t) + p_{n-1}(t)\mu\Delta dt$$

$$\frac{p_n(t+\Delta t) - p_n(t)}{\Delta t} = -p_n(t)(\lambda+\mu) + p_{n-1}(t)\lambda + p_{n+1}(t)\mu$$

In the steady state $\frac{p_n(t+\Delta t) - p_n(t)}{\Delta t} = 0$

$$p_{n+1}(t)\mu = p_n(t)(\lambda+\mu) - p_{n-1}(t)\lambda$$

$$p_{n+1}(t) = p_n(t)\left(\frac{\lambda}{\mu}+\mu\right) - p_{n-1}(t)\frac{\lambda}{\mu}$$

Or $p_{n+1}(t) = p_n(t)(\rho+1) - p_{n-1}(t)\rho$

In the steady state probabilities for queuing systems are

$$p(t) = t \xrightarrow{\lim} \infty \; p(n)\,(t) \quad (n = 0,1,2,\ldots..)$$

Consider empty system

Time t No. of units	Arrival No. of units	Service	Time(t+dt)
0	0	-	0
1	0	1	0

$$p_0(t+\Delta t) = p_0(t)(1-(\lambda)\Delta t) + p_1(t)\mu\Delta tdt$$

$$\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = -p_0(t)(\lambda) + p_1(t)\mu$$

In the steady state $\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = 0$

Under steady state we have $0 = \lambda p_0 - \mu p_1$ for M/M/1 model, we define the utilization factor as $\rho = \frac{\lambda}{\mu}$ where ρ is the expected number of arrival per mean service time. If $\rho < 1$, then steady state probabilities exist and are given by $p_n = \rho^n(1-\rho)$

Measures of this model are as follows.

Expected (average) number of the customers in the system

$$L_s = \frac{\rho}{1-\rho}$$

Expected (average) queue length

$$L_q = L_s - \frac{\lambda}{\mu} = \frac{\rho^2}{1-\rho}$$

Expected waiting time in the queue

$$W_q = \frac{\rho}{\mu - \lambda}$$

Expected waiting time in the system

$$W_s = W_q + \frac{1}{\mu} = \frac{1}{\mu - \lambda}$$

3.20M/M/Z/∞ Model

This model comprise with the condition in which there are several service stations in parallel and each customers in the waiting queue can be served by more than one server. Consider an M/M/z queue with arrival rate λ, service rate μ, and z number of servers.
The traffic intensity will be,

$$\rho_z = \frac{\rho}{z} = \frac{\lambda}{z\mu}$$

The steady distribution of queuing system is studies as following,
$\rho_n = \rho(N = n)$ $(n = 0,1,2,3....)$ Is the probability distribution of the queue length N, as the system is in steady state, when the number of system servers is z, then we have $\lambda_n = \lambda$, (n= 0, 1, 2...)

If there are n customers in the queuing system at any given time, then the following two cases may arise.

- If n< Z (number of customers in the system is less than the number of servers), then there will be no queue. However, (Z-n) number of servers will not be busy. The combined service rate will be $\mu_n = n\mu : \; n < z$
- If n ≥ Z, (number of customers in the system is more than or equal to the number of servers) then all servers will be busy and the maximum number of customers in the queue will be (n-Z). The combined service rate will be $\mu_n = n\mu : n > z$

The steady state difference equations are

$$p_0 = (t + \Delta t) = p_0(t)(1 - \lambda\Delta t) + p_t(t)\mu\Delta t + 0(\Delta t) \text{ for n=0}$$

$$p_n(t + \Delta t) = p_n(t)(1 - (\lambda + n\mu)\Delta t) + p_{n-1}(t)\lambda\Delta t + p_{n+1}(t)(n+1)\mu\Delta t + 0(\Delta t) \quad \text{n= 1,2,3,....z-1}$$

$$p_n(t + \Delta t) = p_n(t)(1 - (\lambda + z\mu)\Delta t) + p_{n-1}(t)\lambda\Delta t + p_{n+1}(t)z\mu\Delta t + 0(\Delta t) \quad \text{for n=z,z+1,z+2...}$$

Now dividing these equations by Δt and taking limit as $\Delta t \to 0$ the difference equations are

$$p_0(t) = -\lambda\, p_0(t) + \mu\, p_1(t) \text{ for n=0}$$

$$p_n(t) = -(\lambda + n\mu)\, p_n(t) + \lambda\, p_{n-1}(t) + (n+1)\mu\, p_{n+1}(t) \text{ For n=1, 2, 3, s-1}$$

$$p_n(t) = -(\lambda + z\mu)\, p_n(t) + \lambda\, p_{n-1}(t) + z\mu\, p_{n+1}(t) \text{ For n= z, z+1,z+2.....}$$

Considering the case of steady state, when $t \to \infty$, $p_n(t) \to p_n$ and hence $p_n(t) \to 0$ for all n, above equation becomes

$$0 = -\lambda\, p_0 + \mu\, p_1 \text{ for n=0} \qquad (1)$$

$$0 = -(\lambda + n\mu)\, p_n + \lambda\, p_{n-1} + (n+1)\mu\, p_{n+1} \text{ for } 1 \le n \le z - 1 \qquad (2)$$

$$0 = -(\lambda + z\mu)\, p_n + \lambda\, p_{n-1} + z\mu\, p_{n+1} \text{ for } n \ge z \qquad (3)$$

Hence,

From (1) $0 = -\lambda\, p_0 + \mu\, p_1$

From (2) $p_2 = \frac{\lambda}{2\mu} p_1 = \frac{\lambda^2}{2!\mu^2} \; p_0$

From (2) $p_3 = \frac{\lambda}{3\mu} p_2 = \frac{\lambda^3}{3!\mu^3} \quad p_0$

In general $p_n = \frac{\lambda}{n\mu} p_{n-1} = \frac{\lambda^n}{n!\,\mu^n} p_0$ for $1 \le n \le z-1$

From (3) $p_z = \frac{\lambda}{z\mu} p_{z-1} = \frac{1}{z!}\left(\frac{\lambda}{\mu}\right)^z p_0$

$$p_{z+1} = \frac{\lambda}{z\mu} p_z = \frac{1}{zxz!}\left(\frac{\lambda}{\mu}\right)^{z+1} p_0$$

$$p_{z+2} = \frac{\lambda}{z^2\mu} p_{z+1} = \frac{1}{z^2 xz!}\left(\frac{\lambda}{\mu}\right)^{z+2} p_0$$

In general $\quad p_n = \frac{1}{z^{n-z} xz!}\left(\frac{\lambda}{\mu}\right)^{(n)} p_0$ for $n \ge z$

Performance measure of above model is as follows

$\sum_{n=0}^{z-1} p_n + \sum_{n=z}^{\infty} p_n = 1$ This gives the steady state distribution of arrivals (n) as

$$p_n = \begin{cases} \frac{\lambda}{n\mu} p_{n-1} = \frac{\lambda^n}{n!\,\mu^n} p_0 & for\, 1 \le n \le z-1 \\ pn \; = \frac{1}{z^{n-z} xz!}\left(\frac{\lambda}{\mu}\right)^n p_0 & for \;\; n \ge z \end{cases}$$

$$p_0 = \frac{1}{\left[\sum_{n=0}^{z-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^{(n)} + \frac{1}{z!}\left(\frac{\lambda}{\mu}\right)^{(z)} \frac{z\mu}{z\mu - \lambda}\right]} \qquad for\; z\mu > \lambda$$

Length of the queue $L_q = p_z \frac{\rho}{(1-\rho)^2}$ where $p_z = \frac{\left(\frac{\lambda}{\mu}\right)^{(z)} p_0}{z!}$

Length of the system $L_s = L_q + \frac{\lambda}{\mu}$

Waiting time in the queue $W_q = L_q / \lambda$

Waiting time in the system $W_s = L_s / \lambda$

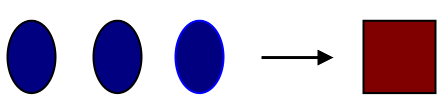
Probability of getting queue $\Pr o(W > 0) = \frac{p_s}{(1-\rho)}$ where as $p_s = \frac{\left(\frac{\lambda}{\mu}\right)^{z} p_0}{z!}$

3.21 The number of servers

System service capacity is the function of the number of service facilities and server proficiency. It assumes that a server or channel can serve one customer at a time. Waiting line systems are either single server (single channel) or multi-server.

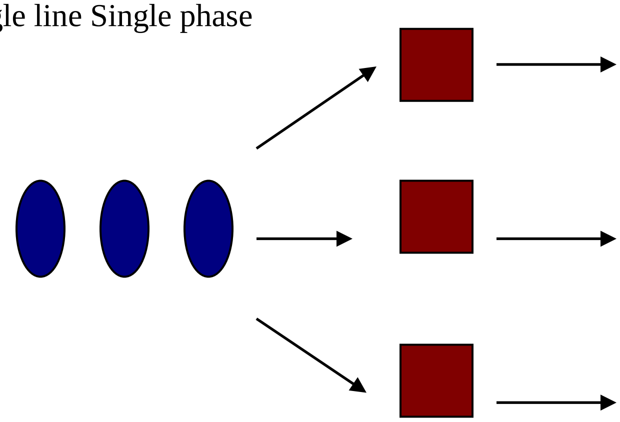
Single server examples include small retail stores with a single checkout counter, atheater with a single person selling tickets and controlling admission in to the show. Multi server systems have parallel service providers offering the same service.

Single -server, single-phase

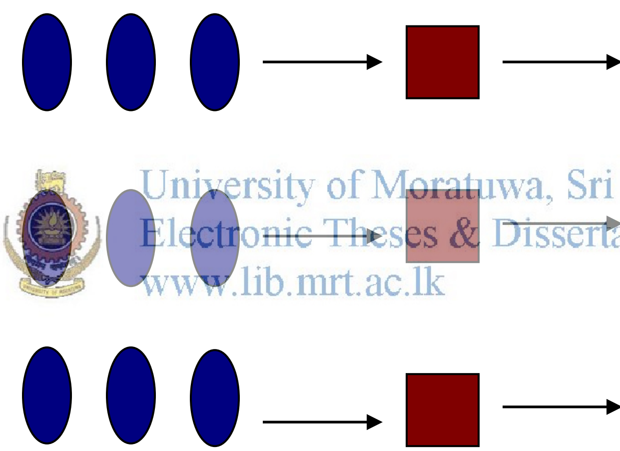


Single –Server, multiphase

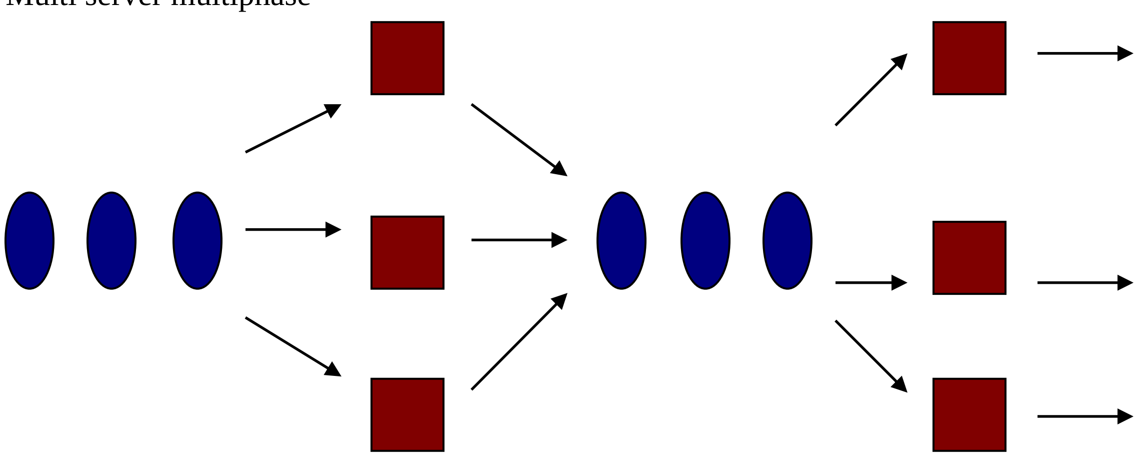
Multi – server, single line Single phase



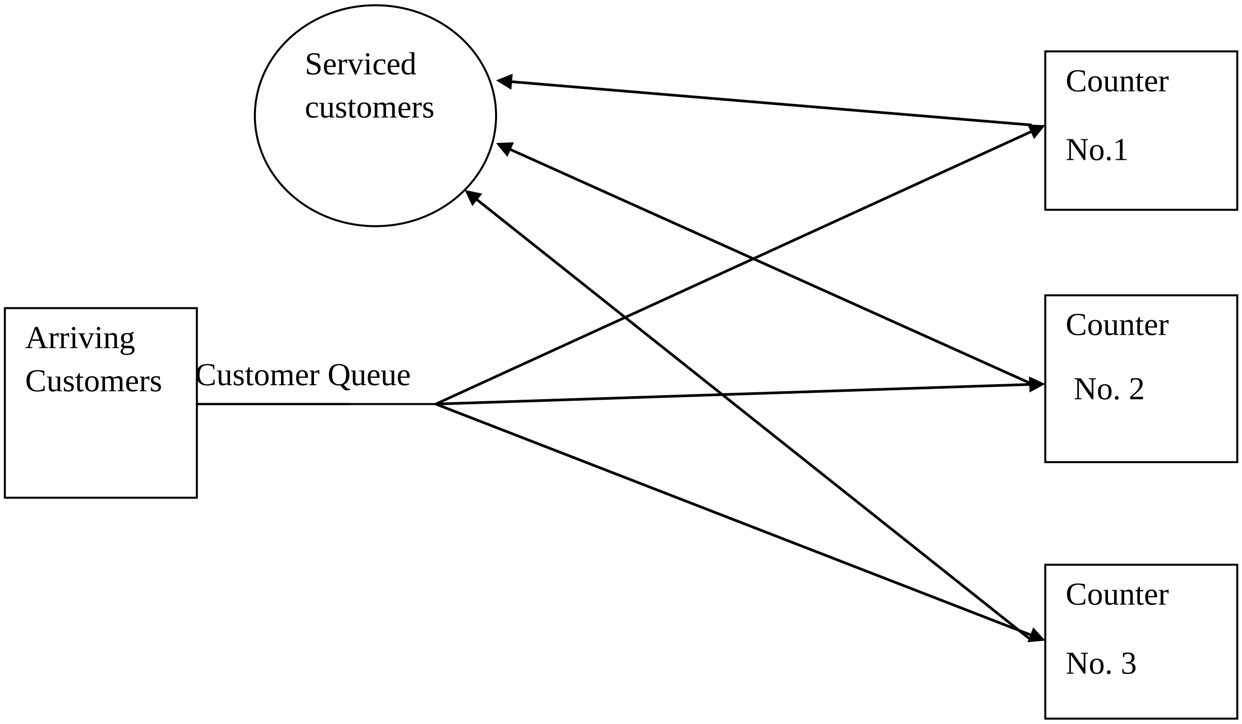
Multi server, multi-line Single-phase



Multi server multiphase



Single line and multiple channels (Three service counters) customer queue at a bank



3.22 Some insights about designing queuing system in a bank

When designing a single server queuing system, by allocating a relatively higher utilization factor to a server will lead surprisingly poor measures of performance for the system.

Managers normally strive for a higher utilization factor for their employees. This is an important part of running an efficient business. A utilization factor of 0.9 or higher would be considered desirable. However all this should change when the employee is the server in a single- server queuing system that has considerable variability in interarrival times and service times.

3.23Chapter summary

This chapter has provided the attention of methodology in which we have perform the study. At the commencement, methodology has highlighted. Definitions of key term that have used in the study have then discussed. Queue discipline and role of exponential distribution with regard to the study have discussed thereafter. Labels for queuing method, service behavior have dealt with subsequently. Queuing models have formulated and performances have measured in different set up and have determined which method is the best to apply for the bank.

CHAPTER 04

ANALYSIS

4.1 Introduction

Analysis means with the use of collected date put in to the above-discussed queuing models and ensure the queue performances.

4.2 Analysis of collected data

As a preliminary analysis, the following are the figures extracted from Bingiriya branch for a period of one week. In order to demonstrate the important terms of the queuing model; we calculate L_s, L_q, W_s and $W_q$respectively and compared to know which one in more efficient, we will analysis it from a technical point as following.

Monday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
8.30-09.30	17	8	15	13	22	21
09.30-10.30	22	9	22	20	26	25
10.30-11.30	23	17	26	24	26	25
11.30-12.30	22	13	22	21	34	33
12.30-01.30	19	10	17	17	23	22
01.30-02.30	11	6	12	12	17	16
02.30-03.30	10	5	12	12	18	17

Table 4.1:Bingiriya branch figures on a first day of a week

Tuesday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
8.30-09.30	17	12	23	21	28	25
09.30-10.30	27	20	35	32	27	34
10.30-11.30	31	27	36	32	42	39
11.30-12.30	38	24	29	27	34	29
12.30-01.30	36	35	31	28	36	33
01.30-02.30	21	18	25	18	24	20
02.30-03.30	18	16	20	17	22	20

Table 4.2:Bingiriya branch figures on a second day of a week

Wednesday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
8.30-09.30	12	11	12	11	24	22
09.30-10.30	19	17	24	23	20	18
10.30-11.30	27	25	25	19	35	33
11.30-12.30	17	14	26	19	32	30
12.30-01.30	14	13	14	10	14	13
01.30-02.30	8	7	7	6	7	5
02.30-03.30	6	5	10	7	5	3

Table 4.3:Bingiriya branch figures on a third day of a week

Thursday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
8.30-09.30	13	12	23	19	20	18
09.30-10.30	19	18	20	18	29	27
10.30-11.30	27	25	27	24	43	39
11.30-12.30	19	17	39	34	40	35
12.30-01.30	24	22	25	18	32	22
01.30-02.30	11	10	11	9	20	16
02.30-03.30	18	15	20	18	24	22

Table 4.4:Bingiriya branch figures on a fourth day of a week

Friday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
8.30-09.30	13	13	22	18	18	13
09.30-10.30	20	19	17	16	28	25
10.30-11.30	23	22	32	23	28	24
11.30-12.30	20	19	29	22	23	19
12.30-01.30	11	11	20	17	22	21
01.30-02.30	10	9	15	7	13	11
02.30-03.30	8	7	12	6	10	7

Table 4.5:Bingiriya branch figures on a fifth day of a week

Above figures have summarized as follows.

		Server 1		Server 2		Server 3	
		Arrival rate	Service rate	Arrival rate	Service rate	Arrival rate	Service rate
Day 1	Total arrival & service rate (per hour)	124	68	126	119	166	159
	Average arrival & service rate (per hour)	17.71	9.71	18	17	23.71	22.71
Day 2	Total arrival & service rate (per hour)	188	152	199	175	213	190
	Average arrival & service rate (per hour)	26.86	21.71	28.43	25	30.43	27.14
Day 3	Total arrival & service rate (per hour)	103	92	118	95	137	124
	Average arrival & service rate (per hour)	14.71	13.14	16.86	13.57	19.57	17.71
Day4	Total arrival & service rate (per hour)	131	119	165	140	208	179
	Average arrival & service rate (per hour)	18.71	17	23.57	20	29.71	25.57
Day 5	Total arrival & service rate (per hour)	105	100	147	109	142	120
	Average arrival & service rate (per hour)	15	14.29	21	15.57	20.29	17.14
Total	Total arrival & service rate	651	531	755	638	866	772
	Average	18.6	15.17	21.57	18.22	24.74	22.05

Table 4.6: Summarized figures of Table 4.1 to 4.5 at Bingiriya branch

Customers arrival rate for server 1(λ1) = 18.6/ hour

Customers arrival rate for server 2 (λ2) = 21.57/hour

Customers arrival rate for server 3 (λ3) = 24.74/hour

Average Customers arrival rate for the servers (λ) = 21.63/hour

Service Rate for Server 1(μ1) = 15.17/hour

Service Rate for Server 2(μ2) = 18.22/hour

Service Rate for Server 3(μ3) = 22.05/hour

Average Service Rate for the Servers (μ) = 18.48/hour

Similar calculations were carried out for other two branches and found the following information

	City Office	Kuliyapitiya	Bingiriya
Average Customers arrival rate (λ)	50	40	21
Average Service Rate (μ)	40	30	18

Table 4.7: Final figures of three branches

City office calculations are as follows

	λ	μ	L_s(customers)	L_q(customers)	W_s(hours)	Wq(hours)
One line & two counters	50	40	2.051	0.801	0.0410 (2.46 minutes)	0.016 (0.96 minutes)
Two lines & two counters	25	40	1.666	1.041	0.066 (3.96 minutes)	0.041 (2.46 minutes)

Table 4.8: Calculation of queue performance at City office

Kuliyapitiya calculations are as follows

	λ	μ	L_s(customers)	L_q(customers)	W_s(hours)	Wq(hours)
One line & two counters	40	30	2.4	1.066	0.06 (3.6 minutes)	0.026 (1.56 minutes)
Two lines & two counters	20	30	2.0	1.333	0.1 (6.0 minutes)	0.066 (3.96 minutes)

Table 4.9: Calculations of queue performance at Kuliyapitiya branch

Bingiriya calculations are as follows

	λ	μ	L_s(customers)	L_q(customers)	W_s(hours)	Wq(hours)
One line & two counters	21	18	1.768	0.601	0.084(5.04 minutes)	0.028 (1.68 minutes)
Two lines & two counters	10.5	18	1.4	0.816	0.133 (7.98 minutes)	0.077 (4.62 minutes)

Table 4.10: Calculations of queue performance at Bingiriya branch

In the case of one line and two counters and two lines and two counters, $L_{q,}$ W and Wq are **smaller in one line and two counters.** In the real world, we have seen there are several lines and several service stations are available at the banks. Each service station has a separate queue. If each service station has a queue according their schedule, then the arrival customers join in each queue as they wish.

Accordingly, the probability of joining to a queue will be 1/ number of available queues.

For example if there are two queues, the system can be considered as two isolated M/M/1 system, and then the arrival rate of each service station will be λ/2. If there is a line, the system will be M/M/2.When there are n lines, it should have n number of service stations. Each service station has a queue based on their schedule and each arrival customer joins in each queue at the probability at 1/n. The mean arrival rate is λ/n and means service rate is μ.

.

4.3 Rush hours at the bank

We observed that rush hours of Bingiriya branch will be from 9.30 a.m. to 1.30 p.m. We have extracted figures at rush time and considered the calculation for the above same method.

Monday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
09.30-10.30	22	9	22	20	26	25
10.30-11.30	23	17	26	24	26	25
11.30-12.30	22	13	22	21	34	33
12.30-01.30	19	10	17	17	23	22

Table 4.11:Bingiriya branch figures at rush time on the first day of a week

Tuesday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
09.30-10.30	27	20	35	32	27	34
10.30-11.30	31	27	36	32	42	39
11.30-12.30	38	24	29	27	34	29
12.30-01.30	36	35	31	28	36	33

Table 4.12:Bingiriya branch figures at rush time on the second day of a week

Wednesday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
09.30-10.30	19	17	24	23	20	18
10.30-11.30	27	25	25	19	35	33
11.30-12.30	17	14	26	19	32	30
12.30-01.30	14	13	14	10	14	13

Table 4.13:Bingiriya branch figures at rush time on the third day of a week

Thursday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
09.30-10.30	19	18	20	18	29	27
10.30-11.30	27	25	27	24	43	39
11.30-12.30	19	17	39	34	40	35
12.30-01.30	24	22	25	18	32	22

Table 4.14:Bingiriya branch figures at rush time on the fourth day of a week

Friday						
	Server 1		Server 2		Server 3	
Time	Arrival rate(per hour)	Service rate(per hour)	Arrival rate (per hour)	Service rate (per hour)	Arrival rate (per hour)	Service rate (per hour)
09.30-10.30	20	19	17	16	28	25
10.30-11.30	23	22	32	23	28	24
11.30-12.30	20	19	29	22	23	19
12.30-01.30	11	11	20	17	22	21

Table 4.15:Bingiriya branch figures at rush time on the fifth day of a week

After summarizing above figures we have arrived to the following information.

Average Customers arrival rate (λ)	23
Average Service Rate (μ)	20

Table 4.16:Bingiriya branch summarize figures at rush time

	λ	μ	L_s(customers)	L_q(customers)	W_s(hours)	Wq(hours)
One line & two counters	23	20	1.718	0.568	0.074 (4.44 minutes)	0.024 (1.44 minutes)
Two lines & two counters	11.5	20	1.352	0.777	0.117 (7.02 minutes)	0.067 (4.02 minutes)

Table 4.17: Calculation of queue performance at Bingiriya branch at rush time

Conclusion

Even at the rush time also “one line and two counters is effective than two lines and two counters”.

Figure 4.1: Total arrivals of Mondays at Bingiriya branch - hourly wise

Arrivals have gradually increased from 8.30 to 12.30 and decreased thereafter.

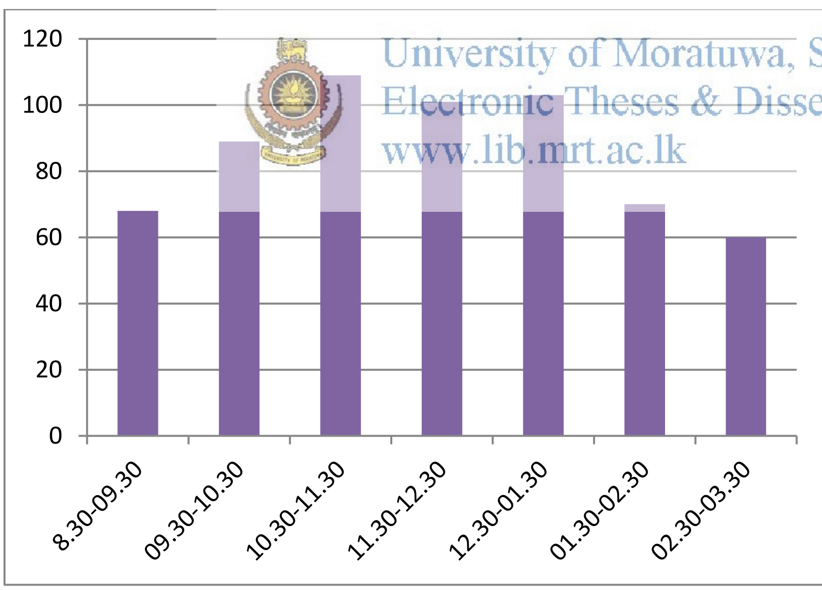


Figure 4.2: Total arrivals of Tuesdays at Bingiriya branch - hourly wise

On Tuesdays arrivals have increased from 8.30 to 11.30, decreased in 11.30 to 12.30, increased in 12.30 to 1.30 and decreased thereafter. Tuesday is the highest arrival recorded day. The reason for this is during the week end customers will collect maximum number of cheques from their customers and deposit on Monday with the bank. Those cheques will be realizing on Tuesday. Therefore customers will gather to withdraw money on Tuesdays than other days in the week.

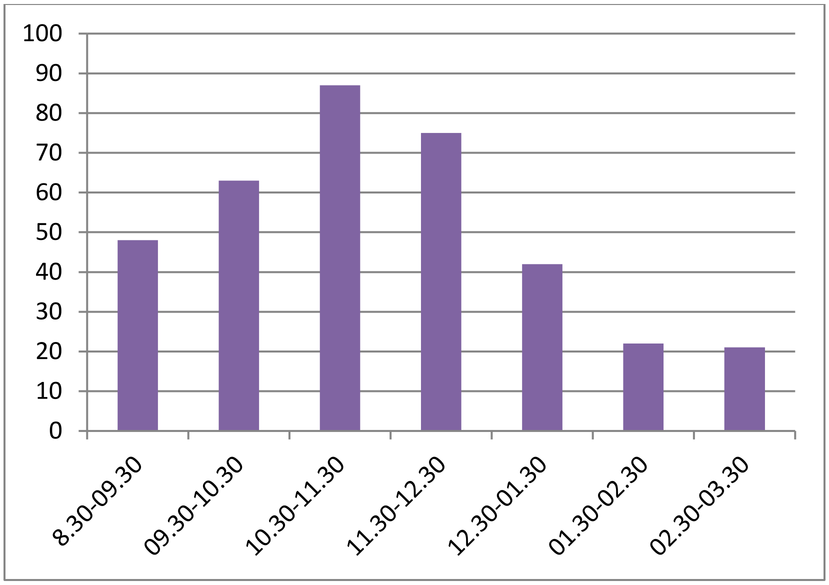


Figure 4.3: Total arrivals of Wednesdays at Bingiriya branch - hourly wise

On Wednesdays also arrivals have increased from 8.30 to 11.30 and decreased thereafter.

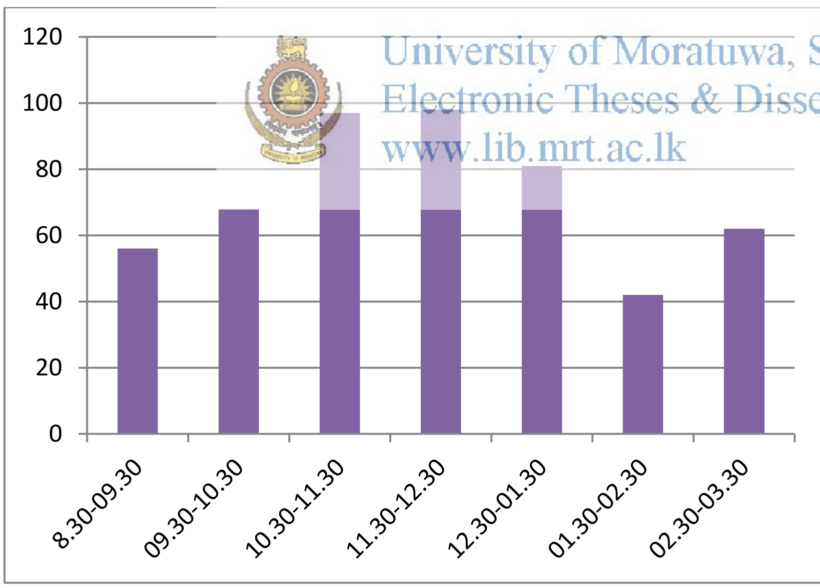


Figure 4.4: Total arrivals of Thursdays at Bingiriya branch - hourly wise

On Thursdays arrivals have increased from 8.30 to 11.30 and decreased up to 2.30 and increased in the last hour

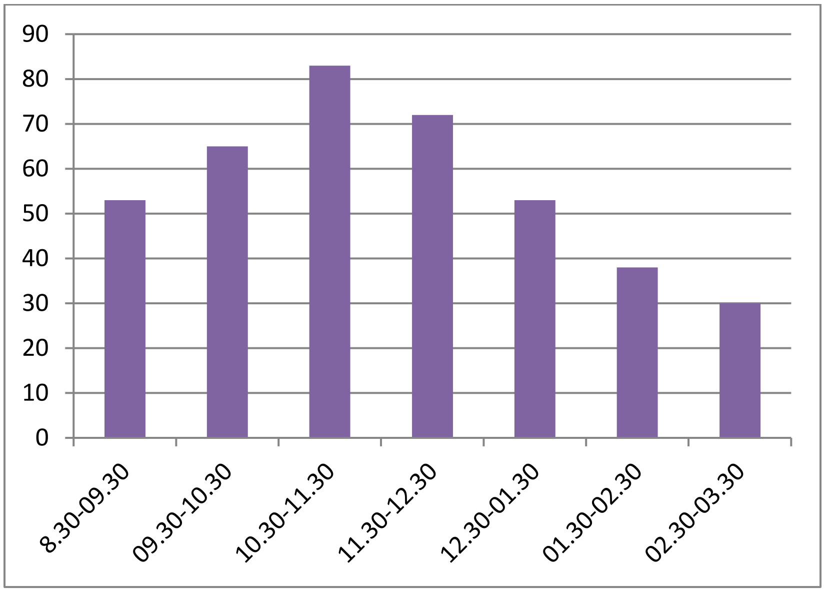


Figure 4.5: Total arrivals of Fridays at Bingiriya branch - hourly wise

On Fridays arrivals have increased from 8.30 to 11.30 and decreased thereafter

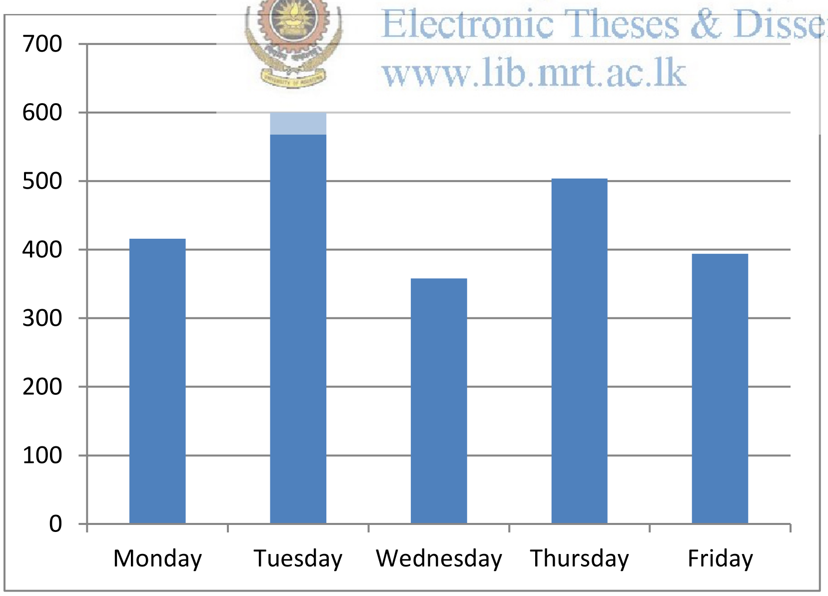


Figure 4.6: Total arrivals of the Bingiriya branch–During the week

During the week arrivals have increased from Monday to Tuesday decreased on Wednesday increased on Thursday and decreased again on Friday.

4.4 Chi-squared test for goodness of fit

Arrivals during the day

Ho: Number of arrivals is independent of the day.

Alternative Hypothesis

H_t:Number of arrivals is not independent of the day.

Under Ho, the expected frequencies are E_i

	Monday	Tuesday	Wednesday	Thursday	Friday	Total
Total arrivals(O_i)	416	600	358	504	394	2272
Mean arrivals (E_i)	454.4	454.4	454.4	454.4	454.4	
$\frac{(O_i - E_i)^2}{E_i}$	3.245	46.653	20.451	5.414	8.028	83.792

Table 4.18: Calculation of Chi-squared test for goodness of Bingiriya branch

The observed value of the test statistic $\sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 83.792$

$\chi_2^2(5\%) = 5.991 X^2$

Since 83.792 >5.991, reject Ho at 5% level and conclude that the number of arrivals is not independent of the day of theweek.

Arrivals during the rush time

Ho: Number of arrived customers is independent of the day.

Alternative Hypothesis

H_t:Number of arrived customers is not independent of the day.

Under Ho, the expected frequencies are E;.

	Monday	Tuesday	Wednesday	Thursday	Friday	Total
Total service time (O_i)	279	402	267	344	273	1565
Mean service time (E_i)	313	313	313	313	313	
$\frac{(O_i - E_i)^2}{E_i}$	3.693	25.306	6.760	3.070	5.111	43.942

Table 4.19: Calculation of Chi-squared test for goodness of Bingiriya branch at rush time

The observed value of the test statistic $\sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 43.942$

$\chi_2^2(5\%) = 5.991X^2$

Since 43.942 >5.991, reject Ho at 5% level and conclude that the number of arrivals is not independent of the day of theweek.

4.5 Optimal service station

Considering the queue length always there will be an accepted waiting time at the bank queues. It is very clear that more the number of servers the bank manager could allocate then greater efficiency will be there. However, there will be some sort of justification as to how many number of servers introduced to ensure accepted efficiency at the bank. Bank manager should come to know by enhancing or decreasing one server the change of queuing probability. So that we must find out how many number of servers should be there.

The probability of having to wait for service

$$\text{Pro (W>0)} = \frac{\left(\frac{\lambda}{\mu}\right)^{s}}{z!\left(1-\rho\right)} p_0$$

Whereas s= number of servers, $\rho = \frac{\lambda}{z\mu}$

When calculate City office figures to ensure the optimal number of servers

Number of servers	2	3	4	5
Probability of queuing	48.06%	12.44%	2.70%	0.50%

Table 4.20: Calculations of queue probability at City office

It is proposed to have three counters.

When calculate Kuliyapitiya figures to ensure the optimal number of servers

Number of servers	2	3	4	5
Probability of queuing	53.33%	13.56%	2.91%	0.53%

Table 4.21: Calculations of queue probability at kuliyapitiya branch

It is proposed to have three counters

When calculate Bingiriya figures to ensure the optimal number of servers

Number of servers	2	3	4	5
Probability of queuing	42.97%	11.31%	2.49%	0.46%

Table 4.22: Calculations of queue probability at Bingiriya branch

It is proposed to have three counters

4.6Why use this model

1. Common queuing system is having number of counters and having number of queues for those counters. (Many lines)
2. We have proved that having one line is more effective than having many lines and quantitative method for having number of counters through the queue probability.
3. When use this one line system definitely efficiency of the queue will upgrade and reduce customer waiting time at the queue.
4. When practice this suggested queue probability method, the manager of the branch will be able to open exact number of counters. This effort will result the manager to obtain optimum service from the other employees at the branch.

4.7 Chapter summary

In this chapter, we have discussed about analysis of collected data to determine the effectiveness of having number of lines or having one line and number of servers. Further, we have discussed how to decide to create number of counters.

The implementation of the model has decided in the chapter. Proved that having one line and several counters are more effective for the queue management.

CHAPTER 05

RESULTS CONCLUSION AND RECOMMENDATIONS

In this, study preliminary information gathered from the branches in Bank of Ceylon located in three different towns. I have visited several bank branches and found that none of these branches has taken scientific approach to determine number of servers as well as queue management strategies.

Upon analyzing, the collected date we found that M/M/s model is more efficient than M/M/1 model.

Results were as follows.

At city office under the condition of M/M/1(Two lines and two counters)

L_q= 1.041 customers W_s= 0.066 (3.96 minutes) W_q= 0.041 (2.46 minutes)

Under the condition of M/M/2(One line and two counters)

L_q= 0.801 customers W_s= 0.041 (2.46 minutes) W_q= 0.016 (0.96 minutes)

It is easily find that L_q, Ws and W_q is lessor in the case one line and two counters than two lines and two counters. That means efficiency is higher in the case one line and two counters than two lines and two counters

L_s =0.801 customers $<$1.041 customers

W_s =0.041 (2.46 minutes) $<$0.066 (3.96 minutes)

W_q =0.016 (0.96 minutes) $<$0.041 2.46 minutes)

At Kuliyapitiya branch under the condition of M/M/1(Two lines and two counters)

L_q=1.333 customers W_s=0.1 (6.0 minutes) W_q= 0.066 (3.96 minutes)

Under the condition of M/M/2(One line and two counters)

L_q=1.066 customers W_s=0.06 (3.6 minutes) W_q= 0.026 (1.56 minutes)

It is easily find that L_q, Ws and W_q is lesser in the case one line and two counters than two lines and two counters. That means efficiency is higher in the case one line and two counters than two lines and two counters

L_s =1.066 customers <1.333 customers

W_s =0.060 (3.6 minutes) <0.100 (6 minutes)

Wq =0.026 (1.56 minutes) <0.066 (3.96 minutes)

At Bingiriya branch under the condition of M/M/1(Two lines and two counters)

L_q=0.816 customers W_s=0.133 (7.98 minutes) W_q= 0.077 (4.62 minutes)

Under the condition of M/M/2(One line and two counters)

L_q=0.601 customers W_s=0.084 (5.04 minutes) W_q= 0.028 (1.68 minutes)

It is easily find that L_q, Ws and W_q is lesser in the case one line and two counters than two lines and two counters. That means efficiency is higher in the case one line and two counters than two lines and two counters

L_s =0.601 customers <0.816 customers

W_s =0.084 (5.04 minutes) <0.133 (7.98 minutes)

W_q =0.028 (1.68 minutes) <0.077 (4.62 minutes)

It is noted and proved precisely that every occasion single line and two counters is very effective than having two lines and two counters. Of course in the case of M/M/1 model the entire efficiency of that queue will be depend on particular person who is in the counter. Counter people’s mentality of having larger queue only in front of his/her counter will further make the delay to the customers. However, in the case of M/M/z model the efficiency of all the counter people will consolidate. Joint effort in combine manner will enhance aggressive working environment. This combination will result of providing more speedy service to their customers. Therefore it is proved that M/M/z is effective than M/M/1 model at the bank.

Further we found that best way to determine number of servers. We are well aware that highest number of servers of course reduces the customer waiting time to the maximum and lower number servers will do the opposite. However, we cannot allocate all the counters opened since bank has to run other functions as well. Therefore, the bank manager should determine in order to guarantee the quality of service, how many service stations should be set up.

Findings of optimal service station will help in this regard.

The probability of having to wait for service

$$\text{Pro (W>0)} = \frac{\left(\frac{\lambda}{\mu}\right)^{s}}{z!\left(1-\rho\right)} p_{0}$$

Whereas s= number of servers, $\rho = \frac{\lambda}{z\mu}$

When calculate City office figures to ensure the optimal number of servers

Number of servers	2	3	4	5
Probability of queuing	48.06%	12.44%	2.70%	0.50%

Table 5.1: Calculations of optimal number of servers at City office

When calculateKuliyapitiya figures to ensure the optimal number of servers

Number of servers	2	3	4	5
Probability of queuing	53.33%	13.56%	2.91%	0.53%

Table 5.2: Calculations of optimal number of servers at Kuliyapitiya

When calculateBingiriya figures to ensure the optimal number of servers

Number of servers	2	3	4	5
Probability of queuing	42.97%	11.31%	2.49%	0.46%

Table 5.3: Calculations of optimal number of servers at Bingiriya

Based on the above figures the manager can decide whether to open how many counters. By looking at the above figures it is reasonably to open three counters at those branches.

Based on the study it is proved that having one line and number of servers is more efficient than having more lines.

Based on the above queue probability the manager will be able to adjust the servers at the bank.

The efficiency of the counter transaction of the bank can improve by adopting above strategies. The queuing number, the service station number has investigated by means of queuing theory. By the example, the results are effective and practical. The time of customer queuing is reduced, the customer satisfaction is increased. It is proved that this optimal model of the queuing is feasible.

Trying to put this knowledge of queuing theory to some use, we took a trip to a foreign bank in Colombo to watch their lines and see if their system aligned toacceptable standard.We found that the system is simple one to keep track of particularly because all of the customers waited in a single queue instead of several separate ones. That made it possible to identify how long each customer had been waiting in line. However, the number of servers on duty during the period of our visit kept changing, and to keep the model from becoming overcomplicated. There were 77 customers kept track of with a chosen value of 5 servers over the course of one hour of observation. On average, a customer arrived every 46 seconds, or every 0.77 minutes. Thus, the sample interarrival rate was 1.30 customers per minute. In addition, service completions took an average of 146 seconds, or 2.44 minutes. This means that the sample service rate was 0.41 service completions per minute. The average waiting time in the queue was 43.6 seconds, or 0.73 minutes. The average total time a customer spent in the system was 3.17 minutes. Ideally, this system could be modeled as $M/M/5/FCFS/\infty/\infty$ system

5.1 Limitation and further recommendations

Although this study was carefully prepared in line with the queuing theory, we all know that there are limitations and shortcomings. The main limitation is the human behavior. Of course, with the findings of the study it is recommended to have one line instead of having number of lines. However there are some customers who prefer to wait at the bank for longer period since they need human responsiveness in the society. Mainly the senior citizen customers are preferred to talk with others and wait in queues at the bank for a longer period.

Nevertheless, it is recommended to apply this model with some value addition features such as

1. Issue Debit cards and arrange to divert the customers from the cash counters to the automated teller machine at the bank.
2. Introduce internet-banking facility to the customers and make them aware to pay online.
3. Install cash deposit machine at the bank premises and make familiar to deposit cash and ensure the accuracy.
4. Deploy an employee at the entrance to inquire the exact requirement of the arrival customer and to guide accordingly.

REFERENCES

1. Central Bank Annual Report - year 2013

2. Dr.Chuka Emmanuel& others (2014), "Analysis of a queuing system in First Bank PLC in Nigeria"

3. EhsanBolandifar (2014) – Modeling the behavior of the patients who leave the hospital without being seen

4. Introduction to Management Science- Fredrick S Hillier &Mrak S Hillier

5. MathiasDharmawirya (2011) – Case study for restaurant queuing model

6. Operations Research – An Introduction – Eighth Edition – Hamdy A Taha, 2007.

7. Quantitative Methods for Management with application in planning & Decision Making Mr. K R M T Karunarathna

8. S.K.DharandTanzinaRahman (2013)- University of Information Technology & Science (UITS) Bangladesh- Case study for bank ATM queuing model

9. Statistics for Management – Seventh Edition – Richard I Levin & David S Rubin

10. S.Vijay Prasad& others (2015) - Mathematical analysis of single queue multi server and multi servers multi queues