

BINDURA UNIVERSITY OF SCIENCE EDUCATION



FACULTY OF SCIENCE AND ENGINEERING DEPARTMENT OF STATISTICS AND MATHEMATICS

*Optimizing Efficiency in the Revenue office of Masvingo Rural District Council. A
Queuing Theory and Simulation Approach*

NAME	: ANOTIDA E. CHIVHOKO
REG NUMBER	: B213526B
DATE OF SUBMISSION	: 20 JUNE 2025
SUPERVISOR	: DR T.W. MAPUWEI

AUTHORSHIP DECLARATION STATEMENT

Title of the Thesis: Optimizing Efficiency in the Revenue office of Masvingo Rural District council. A Queuing Theory and Simulation Approach

Author: Anotida Everjoy Chivhoko

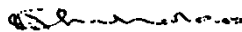
Program: BACHELOR OF SCIENCE HONOURS DEGREE IN STATISTICS AND FINANCIAL MATHEMATICS (HBScSFM)

I, the undersigned author of the above-mentioned thesis, hereby declare that:

1. This thesis is my original work and has been prepared by me in accordance with the institution's requirements.
2. All sources, data, and references used in this thesis have been acknowledged and cited appropriately.
3. This thesis has not been submitted elsewhere for any degree or diploma.
4. I have obtained all necessary permissions for the inclusion of third-party content where applicable.

I affirm that this declaration is made with full integrity and in compliance with the institution's policies and academic practice.

Author's: Anotida Everjoy Chivhoko



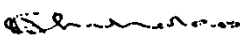
Signature:

Date: 19 June 2025

APPROVAL FORM

This is to certify, that this research project is the result of my own research work and has not been copied or extracted from past sources without acknowledgement. I hereby declare that no part of it has been presented for another degree in this University or elsewhere.

CHIVHOKO ANOTIDA E


.....

19/06/2025

B213526B

Signature

Date

Certified by:



19/06/2025

DR. T.W MAPUWEI

.....

.....

Supervisor

Signature

Date

DR.M. MAGODORA



19/08/2025

Chairperson

Signature

Date

DEDICATION

I dedicate this dissertation to my parents Mr and Mrs Chivhoko who have made sacrifices towards my personal and professional endeavours and my siblings for their encouragement. I thank you for believing in my dreams.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude in writing this acknowledgement as I could not have completed my thesis without the invaluable assistance of my supervisor, family and friends. I extend my deepest appreciation to my supervisor, Dr T. W. Mapuwei for consistently providing me with support throughout the development of my dissertation. Their patience, guidance, motivation, and extensive knowledge proved invaluable during the analysis and writing process. I also thank God for providing me with the care, strength, knowledge and opportunity to pursue my education to this level. Additionally, I would like to thank the lecturers and staff of the Department of Statistics and Mathematics at Bindura University of Science Education for their academic support and knowledge sharing. To my family, I express my sincere gratitude for their unwavering encouragement, advice and prayers all through my studies. I pray that the Lord bless them abundantly. Finally, to my colleagues, I consider you all as family, and I ask for God's guidance and protection to be with you always.

ABSTRACT

This study focuses on enhancing operational efficiency at the Masvingo Rural District Council (MRDC) revenue office through the application of queuing theory and simulation modelling. The research addresses inefficiencies including long customer wait times, low satisfaction levels and suboptimal resource utilization, which hinder revenue collection and service delivery. The research problem centres on the systemic delays caused by a single-server queuing system, leading to customer frustration and suboptimal resource allocation. To tackle this, the study employs queuing theory and simulation modelling to analyse and optimize the revenue office's operations. The primary research objectives are to evaluate the current efficiency of the revenue office in terms of waiting times and staff utilization, develop queuing and simulation models to represent customer flow, reduce average customer waiting times and assess the cost-effectiveness of proposed improvements. The study adopts a descriptive research design, combining observational data collection with analytical tools such as TORA and Arena simulation software. Key metrics including arrival rates, service times and queue lengths, were analysed under three staffing scenarios (1, 2, and 3 servers) to identify the most efficient configuration. Key findings revealed that the existing single-server system results in excessive delays, with an average wait time of 34.24 minutes. Introducing a second server significantly improved performance, reducing wait times to just 1.14 minutes, while a third server offered only marginal benefits, indicating diminishing returns. The results also demonstrate the effectiveness of data-driven approaches in enhancing public sector service delivery. The study concludes that a two-server configuration is the most cost-effective and efficient solution. The research recommends deploying a second cashier to alleviate congestion and implementing process improvements to streamline operations, introducing dynamic staffing during peak hours based on real-time demand and also separating complex and routine transactions into distinct service streams to reduce bottlenecks. These measures are expected to enhance customer satisfaction, optimize staff utilization, and increase revenue collection. Future studies should examine the long-term financial benefits and return on investment (ROI) of applying a two-server system including the potential increase in revenue collection and a decrease in operational costs. Conduct surveys and interviews to quantitatively assess the effect of queuing system improvements on customer satisfaction and perceived service quality. Investigate the feasibility and effectiveness of real-time queue management systems that can dynamically adjust staffing based on predictive analytics and live data.

Contents

AUTHORSHIP DECLARATION STATEMENT	i
APPROVAL FORM	ii
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	v
LIST OF FIGURES	ix
LIST OF TABLES	ix
CHAPTER 1: INTRODUCTION	- 1 -
1.0 INTRODUCTION	- 1 -
1.1 BACKGROUND OF STUDY	- 1 -
1.2 STATEMENT OF THE PROBLEM	- 2 -
1.3 RESEARCH OBJECTIVES	- 3 -
1.4 RESEARCH QUESTION	- 3 -
1.5 ASSUMPTIONS	- 3 -
1.6 SIGNIFICANCE OF THE STUDY	- 3 -
1.7 LIMITATIONS	- 4 -
1.8 DELIMITATIONS	- 4 -
1.9 DEFINATION OF KEY TERMS	- 4 -
1.10 CHAPTER SUMMURY	- 5 -
CHAPTER 2: LITERATURE REVIEW	- 6 -
2.0 INTRODUCTION	- 6 -
2.1 CONCEPTUAL FRAMEWORK	- 6 -
2.2 THEORETICAL LITERATURE	- 6 -
2.2.1 An overview of Queuing Models	- 6 -
2.3 QUEUING THEORIES	- 10 -
2.3.1 Erlang's Queuing Theory	- 10 -
2.3.2 Little's Theorem	- 11 -
2.3.3 Queuing Network Theory	- 11 -
2.3.4 Birth-Death process theory	- 12 -
2.3.5 Markovian queuing theory	- 12 -

2.3.6 Kendall Notations Theory.....	- 13 -
2.4 AN OVERVIEW OF QUEUING CONCEPTS.....	- 13 -
2.4.1 Customer arrival patterns	- 13 -
2.4.2 Queuing discipline.....	- 14 -
2.4.3 Queue Management	- 15 -
2.4.4 Service quality	- 16 -
2.5 SIMULATION APPROACH.....	- 17 -
2.6 COMBINATION OF QUEUING AND SIMULATION APPROACH	- 18 -
2.7 EMPIRICAL EVIDENCE.....	- 18 -
2.8 THE GAP	- 20 -
2.9 CHAPTER SUMMARY	- 20 -
CHAPTER 3: METHODOLOGY	- 22 -
3.0 INTRODUCTION.....	- 22 -
3.1 RESEARCH DESIGN	- 22 -
3.2 STUDY AREA AND STUDY POPULATION.....	- 22 -
3.3 DATA COLLECTION	- 23 -
3.4 DESCRIPTION OF VARIABLES AND EXPECTED RELATIONSHIP	- 23 -
3.5 PRE-TEST	- 24 -
3.6 QUEUING MODEL.....	- 24 -
3.6.1 Advantages of Queuing Model	- 24 -
3.6.2 Kendall's Notation.....	- 25 -
3.6.3 Parameters.....	- 25 -
3.7 MODEL ASSUMPTIONS	- 26 -
3.7.1 Multi-server Queue Model (M/M/k)	- 26 -
3.8 MODEL SELECTION CRITERIA	- 28 -
3.8.1 TORA Software	- 28 -
3.8.2 Arena Software	- 29 -
3.8.3 Simulation Model.....	- 29 -
3.9 MODEL VALIDATION	- 29 -
3.9.1 Input-Output Validation.....	- 29 -
3.9.2 Sensitivity Analysis	- 30 -
3.10 CHAPTER SUMMARY	- 30 -
CHAPTER 4: DATA PRESENTATION, ANALYSIS AND DISCUSSION	- 31 -
4.0 INTRODUCTION.....	- 31 -
4.1 DESCRIPTIVE STATISTICS	- 31 -
4.2 PRE-TEST	- 32 -

4.4.1 Chi-Square Test	- 32 -
4.2.2 Kolmogorov-Smirnov and Chi-Square test	- 33 -
4.3 DATA ANALYSIS AND OBSERVATION.....	- 33 -
4.3.2 Queuing Analysis Using TORA Software and Arena Models	- 35 -
4.3.3 Deeper Analysis and Connection of Arena and Tora	- 37 -
4.4 MODEL VALIDATION TESTS	- 40 -
4.4.1 Input-Output Validation.....	- 40 -
4.4.2 Sensitivity Analysis	- 40 -
4.5 DISCUSSION OF FINDINGS.....	- 41 -
4.6 CHAPTER SUMMARY	- 43 -
CHAPTER 5: SUMMARY CONCLUSION AND RECOMMENDATIONS	- 44 -
5.0: INTRODUCTION.....	- 44 -
5.1 SUMMARY OF THE STUDY AND FINDINGS.....	- 44 -
5.3: RECOMMENDATIONS	- 46 -
5.4 AREAS FOR FURTHER RESEARCH.....	- 46 -
REFERENCES	a
APPENDICES	f

LIST OF FIGURES

Figure 2:1 Conceptual Framework	- 6 -
Figure 2:2: M/M/1 queuing model.....	- 8 -
Figure 2:3 M/M/c queuing model	- 9 -
Figure 1:4 M/G/1 queuing model.....	- 9 -
Figure 1:5 G/G/1 queuing model	- 10 -
Figure 1:6 Queue Discipline	- 14 -
Figure 4:1 Historical of arrival time	- 32 -
Figure 4:2: Service rate data plot	- 33 -
Figure 4:3 Simulation model of the revenue office	- 34 -
Figure 4:4 M/M/2 model.....	- 36 -
Figure 4:5 M/M/3 model.....	- 37 -

LIST OF TABLES

Table 1: Descriptive Statistics	- 31 -
Table 2: Performance measures of multi-server queuing model.....	- 35 -
Table 3 : Tora vs. Arena Queuing Performance Comparison.....	- 38 -

CHAPTER 1: INTRODUCTION

1.0 INTRODUCTION

In our fast-changing world, maintaining an efficient organization is important especially for public offices such as the Masvingo Rural District Council (MRDC) revenue office. This office significantly adds to the council's financial health by collecting payments and taxes from local citizens. However long wait times and delays can lead to frustration for visitors and delay operations. Therefore, this study explores how queuing theory and simulation can improve the efficiency in the office. Queuing theory helps in understanding the reasons behind formation of the line and how to effectively manage the lines and simulation allows us to form a practical model of the office to experiment with many approaches such as changing the number of cashiers or changing shift schedules before applying them in reality. By applying actual data on visitor capacity, wait times and service speed we can learn more effective methods to minimize waiting, ensure smooth operations and optimize the use of existing staff.

1.1 BACKGROUND OF STUDY

The revenue offices worldwide are facing pressure due to the increasing difficulty of tax structures and the growing demands for transparency and responsibility. Most revenue offices depend on the old procedures and handbook systems which often result in slow service and a increases possible errors. Queuing theory and simulation approach aims to improve the overall performance in this study of optimizing efficiency in a revenue office so as to bring more income and productivity to the office. The collection of taxes and utility bills are some of the responsibilities of the revenue office. Increased workload for the staff and less income are due to long time waiting of customers, congestion and inefficient procedures and this lead to disappointment of customers. To confirm level operations and to deliver pleasing service to people in the revenue office of Masvingo rural District Council it is important to optimize efficiency. Effectively handling customer wait times and optimizing server performance is important for guaranteeing efficiency and customer satisfaction. These settings deals with issues related to congestion, delays and uneven resource utilization. Queuing theory offers a mathematical structure for analysing these situations permitting decision makers to estimate queue lengths, waiting times and server workloads under different circumstances.

The queuing theory provides a framework for understanding waiting in line and how to manage efficiently (Brahma, 2013). The queuing theory was initiated by Erlang in 1909 to help in defining the volume of the Danish telephone structure. A range of service businesses have involved and activated it, including health system management and computer networking. Erlang also evaluated a mathematical model called the Erlang C model for a single facility M/M/s queue where arrivals of customers are founded on a Poisson process. Kendall D.G characterises one more important turning point in history of theory queuing. Kendall developed and represents the features of a queuing structure which is the A/B/C code known as the “Kendall’s notation”. It has now developed the known method for categorizing queuing theories.

Using computer software, simulation is a method to model real world sensations. Without interfering with the working of the system, it enables researchers to test out various settings and factors. In the revenue office, simulation can be used to test different scenarios and spot possible obstacles or places for improvement. Queuing theory and simulation can be used to review many parts of a revenue office, including customer arrival patterns, wait times, staffing numbers, queue management tactics, and overall process flow. Reducing wait times for residents while keeping true transaction processing is the aim of revenue efficiency optimization. As an outcome, the rural council may see lower operating expenses, higher levels of residence satisfaction, and better overall service delivery.

1.2 STATEMENT OF THE PROBLEM

Masvingo Rural District Council is failing to maximize on revenue collection regardless of being compelled to do so by many legislations. The revenue office faces inefficiencies in its operations due to limited resources including staff, equipment and infrastructure. This is resulting in the long waiting queues of customers who are paying their rates, taxes, licenses and levy to be served. According to the observations some customers are leaving the office without being served because they will be upset with the services which takes hours of their time for service that could take minutes, therefore there is need to optimize efficiency in the revenue office.

1.3 RESEARCH OBJECTIVES

1. To determine bottlenecks in server utilization and queue handling.
2. To develop optimum server and queuing process.

1.4 RESEARCH QUESTION

1. What queuing and simulation models can be adopted to model customer queues paying bills at MRDC?
2. How many customers can be expected to be present in the revenue collecting office at any given time?
3. How long do clients typically wait in line before being attended to in the revenue collecting office?
4. How many staff members are needed to provide optimal service to customers in the revenue collecting office?

1.5 ASSUMPTIONS

1. Sum of customers visiting the revenue office is relatively constant over time.
2. The time it takes to assist each customer is consistent and does not vary significantly.
3. The revenue office has single server attending to the customers.
4. Customers are assisted in the same order they arrive.

1.6 SIGNIFICANCE OF THE STUDY

The study is important because it can help make the revenue office more efficient by cutting down customer wait times, improving overall satisfaction, and making better use of staff and resources. By changing how the office operates, residents could enjoy faster service with fewer difficulties when dealing with payments. At the same time, staff, infrastructure, and technology could be used more effectively, reducing waste and increasing productivity. Eventually these changes would not only improve the day-to-day experience for customers but also support the council in delivering better, more responsive public services.

1.7 LIMITATIONS

- Stakeholder resistance: resistance from stakeholders like staff and residence may limit the ability to apply changes.
- Budget constraints: optimizing efficiency may be limited by the budget constraints that may avoid MRDC from investing in new technologies and adding staff.
- Resource constraints: unavailability of resources limit the optimizing efficiency of MRDC revenue office.

1.8 DELIMITATIONS

The study will take place in one of Zimbabwe's rural districts which is Masvingo rural district. This study will focus on optimizing local revenue collection processes within the revenue office of MRDC, specifically examining its revenue collection processes and systems over a period of one month, February 2025.

1.9 DEFINATION OF KEY TERMS

1. **Queuing theory** – is a subdivision of calculations that educates exactly how lines form, their purpose and why they breakdown, Erlang, A.K. (1909)
2. **Simulation**- is a computer tests that include generating data by random sampling from known likelihood distributions, Feiveson AH. (2002)
3. **Efficiency**- Efficiency measures how effectively resources are used ndandavapo now
4. by matching the useful output shaped to the total effort needed to reach it. Slack, N., Brandon-Jones, A., & Johnston, R. (2013)
5. **Revenue**- is the income received by a firm from the sales of goods and services to its customers, Samuelson, P.A. , Nordhaus , W.D (1995)
6. **Optimization**- is the process of choosing the most effective decision from a set of alternatives to achieve a desired objective, Hillier, F.S., and Lieberman, G.J. (2020)
7. **Satisfaction**- is achieved when a human need is fulfilled , resulting in a state of contentment and well-being, Maslow, A.H (1943)
8. **Utilization**- refers to the degree to which available resources, such as labour or capital, are used in production process, Keynes, J.M. (1936)

- 9. *Residence***- refers to the physical dwelling or location where individuals or families establish their living arrangements, which often reflects social and economic patterns, Giddens, A.(1984)
- 10. *Stakeholders***- any collection or individuals who can be moved or is affected by the success of the organisation objectives, Freeman, R.E (1984)
- 11. *Operations*** – the supervision of the processes that produce or deliver the organisation's goods and services, Slacks et al, (2013)

1.10 CHAPTER SUMMURY

This chapter outlines the problem faced by the residence in paying their rates in the revenue office of Masvingo Rural District Council. The background of study highlights the importance of efficient revenue collection for local government. The problem is that the residence are facing long waiting queues, delays in service delivery, inefficient use of resources low revenue collection and dissatisfaction among customers in the revenue office whilst they want to pay their rates. This research apply queuing theory and simulation approach to solve the problem.

CHAPTER 2: LITERATURE REVIEW

2.0 INTRODUCTION

The reason of this study is to showcase the efficiency of queuing theory in a revenue office of rural councils. The queuing theory is a study of waiting in line and is a full concept in the area of operations supervision (Brahma, 2013). Improving efficiency at a rural council revenue office is important to keep up even operations and offering people reasonable services. The use of simulation and queuing theory tools can assist in archiving this goal. This literature review will give a brief summary of related research that has looked into the use of simulation and queuing theory to revenue office efficiency optimization.

2.1 CONCEPTUAL FRAMEWORK

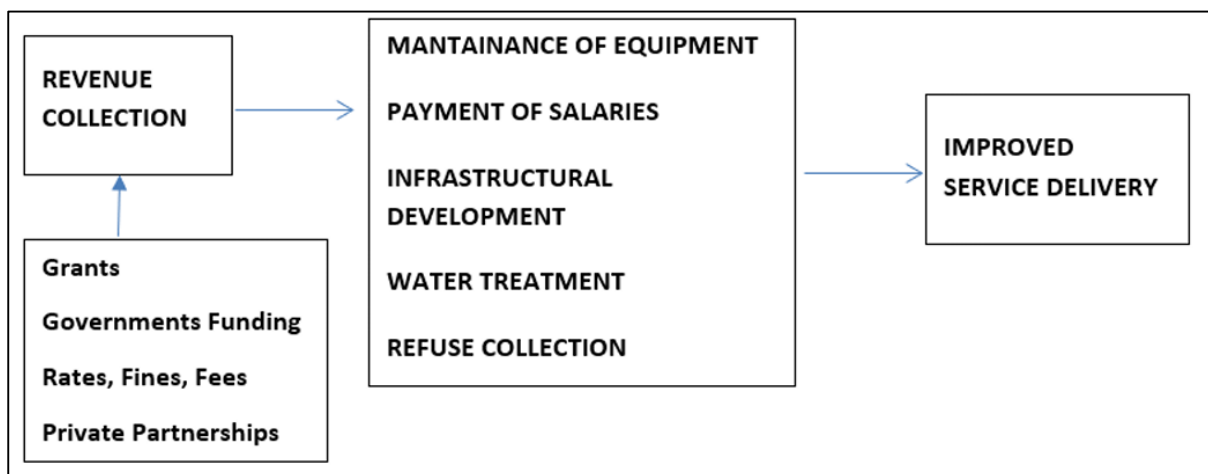


Figure 2:1 Conceptual Framework

2.2 THEORETICAL LITERATURE

The research study is going to use the Queuing Theory and the Simulation approach which are explained below;

2.2.1 An overview of Queuing Models

Queuing models are mathematical structures used to analyse and optimize systems where customers arrive, wait in a queue and then receive service. The main components of a queuing model are the arrival process, service process, queue discipline, number of servers, system capacity and queue length. The arrival process describes how customers arrive at the system, normally modelled using poison processes or any other distribution type. Service process represents how customers are served, usually modelled using exponential or deterministic service times. The method by which the customers are served for example on the FIFO or LIFO basis is what is referred to as the queue discipline. Queues may have multiple servers for example parallel service channels, which influences system performance. The structure volume is the maximum sum of customers acceptable in the system. Number of customers waiting for service is the queue length.

Municipal solid waste collection using the vehicle moving problem to serve number of customers with a fleet of vehicles was a problem solved by Jeroen & Jonas, (2019). They used a 3D-GIS model to allocate trucks to different clients while minimizing costs. They came to the conclusion that waste in the environment and society will be minimal or absent if vehicles are allocated properly. Hasugian, I.A. and Dewi, E., 2020, solved a problem on fast foods restaurant XYZ aims to identify the queuing model that operates in the XYZ cashier area and evaluate its correctness and competence by contrasting the findings of a previous model with those of a simulation run using Any logic software. Three counters and three queue lines make up the restaurant queue system. They concluded that replacing 3 queue lines into 1 queue line which increased the number of customers served for every counter.

A call centre with one agent where calls arrive randomly (poison procedure) and the time consumed with each call is exponential distributed can use a M/M/1 queue model. This model is a single server queue with Poison arrival and exponentially distributed service time, (kleinrock L 1975). The M/M/1 queue model is defined as M (Makovian Arrival Process) is the inter arrival time of customers is exponentially distributed meaning the time between arrivals is memoryless with mean arrival rate denoted as λ and this is also known as the poison process. M (Makovian service time) the service time follows a distribution called exponential with mean service rate denoted as μ . This is also known as the memoryless service procedure where time to serve a customer is independent of the previous service time. 1 (single server) the system consists of single server that deals with customer one at a given time. The M/M/1 queue model has a state for constancy which says for the system to be constant, the rate of

arrival λ need be below the rate of service μ that is, $\lambda < \mu$. If $\lambda \geq \mu$, the structure will become unbalanced and the queue will increase indeterminately, (Gross, D et al 1998).

THE M/M/1 queuing model

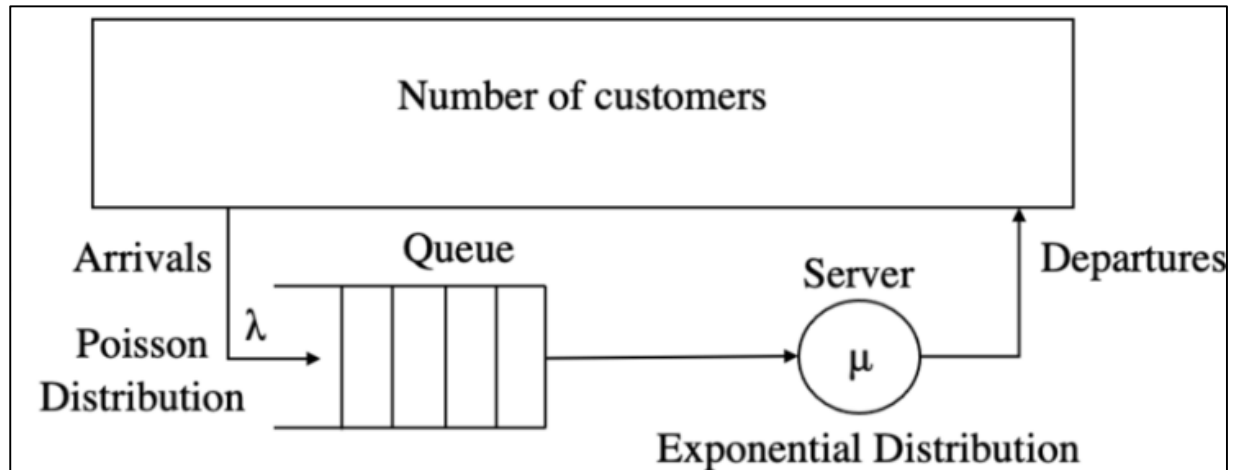


Figure 2:2: M/M/1 queuing model

A queue using multiple servers (c servers), where both arrival and service time are exponentially distributed uses an M/M/ c queue. This model is used in situations where more than one server is needed for instance a bank with multiple tellers serving customers who arrive randomly, (Gross, D et al 1998). The M/M/ c model has been extensively studied and its performance metrics, such as waiting time and queue length have been derived (Erlang 1917; Palm 1938). These matrices are needed in evaluating the efficiency of the system. The general formulation of the M/M/ c structure depend on calculating the system's state probabilities and developing key performance matrices like the structure utilization, the sum of customers in the scheme and the waiting time. M/M/ c queue model takes into account the possibility that customers may leave the queue (reneging) or decide not to join if the system is too busy (balking), (cooper, 1981).

The M/M/c queuing model

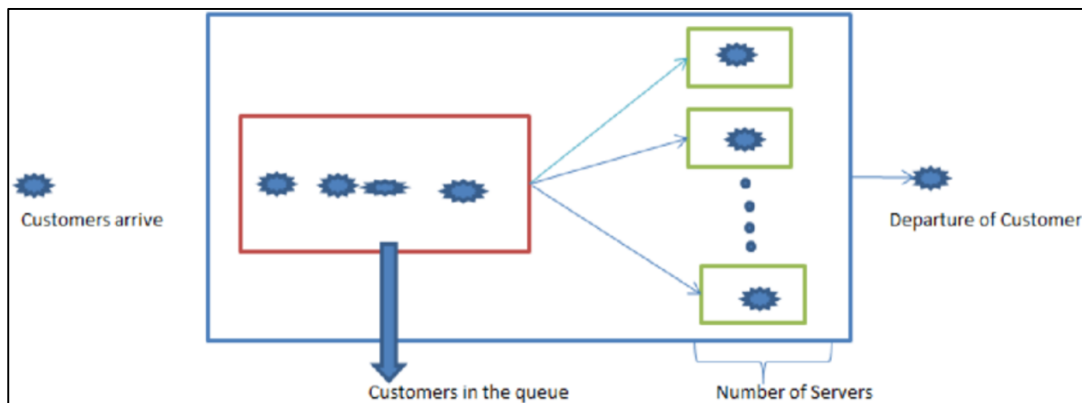


Figure 2:3 M/M/c queuing model

The M/G/1 queuing model

Queue model M/G/1 is a single server where arrival rate is a poisson process but the service rate is commonly distributed. This queuing model extends the M/M/1 queue model by allowing for a general service time distributions (Kendall, 1951). A hospital emergency room where patients arrive at a poisson rate but service times are generally distributed, Wolff, R.W 1989

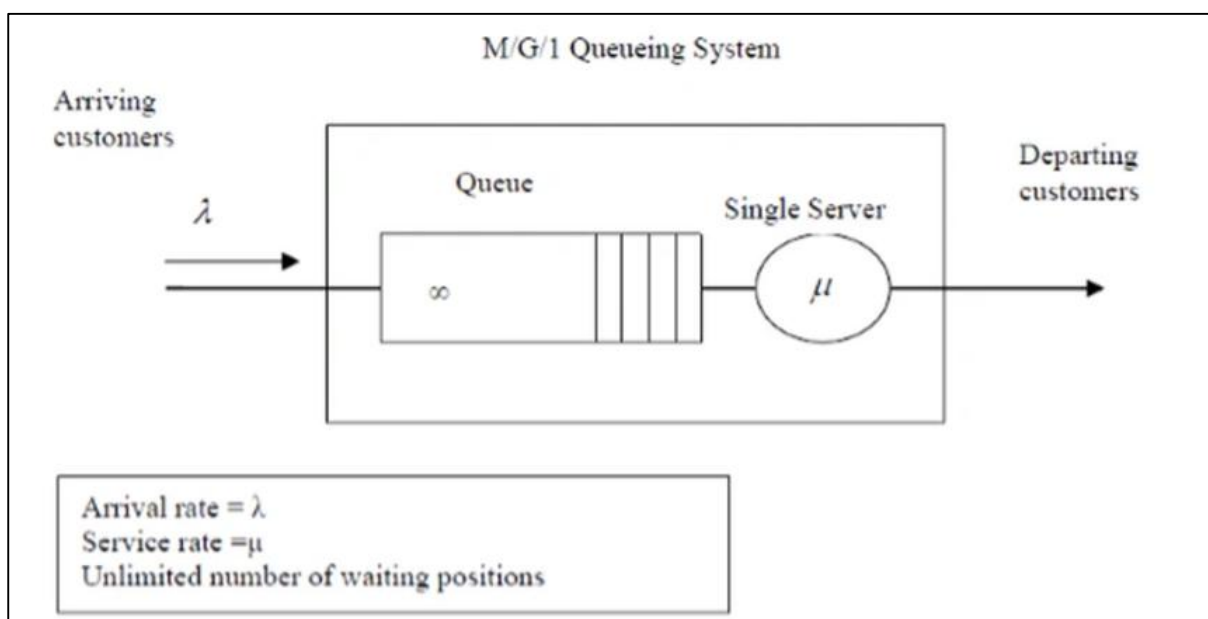


Figure 1:4 M/G/1 queuing model

The G/G/1 queuing model

A general queue model where both service time and the arrival rate are arbitrary distribution is what is called a G/G/1 queue model. The time it takes to assist customers in a supermarket vary on the basis of how many and the kind of items they're purchasing, Takagi, H. (1991). The

G/G/1 queue model have been widely studied using many methods which including the Lindley equation (Lindley, 1952) and the kingman's inequality (kingman, 1962). These techniques gives bounds and calculations for the performance metrics of the structure. The G/G/1 queuing model is an extremely reliable and general model that can be useful to a wide range of real-world structure. The G/G/1 queuing model is the most suitable model for complex system because of its capability to handle arbitrary arrival and service distribution.

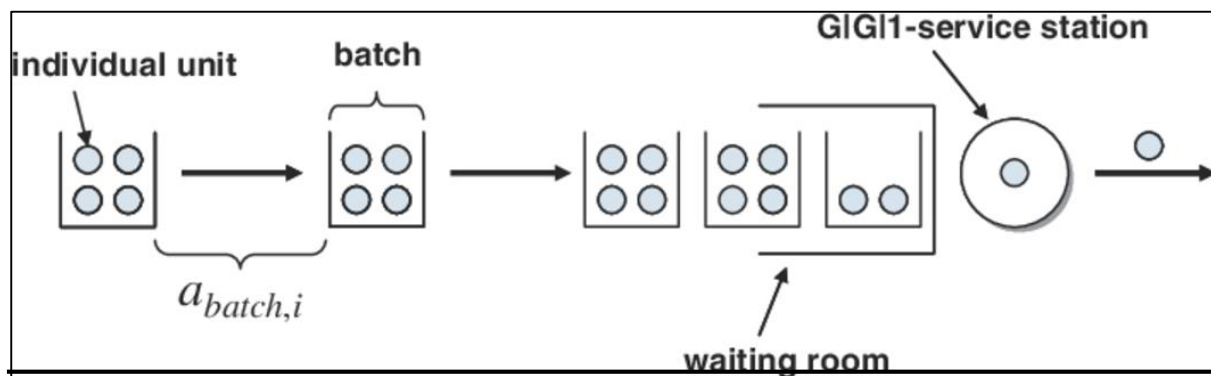


Figure 1:5 G/G/1 queuing model

2.3 QUEUING THEORIES

Queuing theory is a branch of operations research that entails examination of waiting lines or queues with an aim of improving service systems. Queuing theory provides mathematical models for analysing parameters such as arrival rates, service rates, queue lengths, waiting times, and system capacity. Essentially, queuing theory helps organizations make fact-based decisions about resource usage, customer service policy, and process improvement in environments where demand for services changes over time.

2.3.1 Erlang's Queuing Theory

The foundations of modern queuing theory were laid by Agner Krarup Erlang in 1909 a Danish mathematician who was working for the Copenhagen Telephone Company. Erlang developed the first mathematical models to describe the behaviour of telephone call traffic and to determine the number of telephone lines or operators needed in order to minimize lost calls. His pioneering work introduced basic concepts such as Poisson arrival processes, exponential service times, and the Erlang formulas (Erlang B and Erlang C) that are still widely used in telecommunications, call centers, and service industry modelling today. Erlang's work laid the

basis for the analysis and improvement of service systems through quantitative methods, forming the basis of queuing theory as known and practiced today.

2.3.2 Little's Theorem

One of the most significant contributions to queuing theory is the Little's Law introduced by John D.C. Little in 1961. This law creates a major and very simple relationship between three key performance measures in any stable queuing system, $L = \lambda W$ where L represents the average number of customers in the system, λ is the average rate of arrival and W is the average time that a customer will spend in the system. Little's Law is applied broadly to a range of industries and service environments regardless of the specific arrival or service distributions. This flexibility makes it a basis of performance analysis in operations management, manufacturing, logistics, healthcare and any other fields where flow and capacity are essential to understanding. Little's Law links theoretical modelling and practical application allowing managers to assess and optimize system performance using empirical operational measurements.

2.3.3 Queuing Network Theory

J.R. Jackson et.al (1960-1980) extended the basis of queuing theory by developing the concept of queuing networks, a powerful technique for analysing complex service systems composed of multiple interconnected queues. These queuing networks are systems in which entities (customers, data packets, or jobs) move through a number of service stations each with its own arrival and service characteristics. Jackson's understanding was that under a certain set of assumptions such as the Poisson arrivals, service times is exponential, independent routing and first-come-first served discipline the steady state distribution of the entire network could be expressed as the product of the distributions of the individual nodes. This result is known as the product form solution, significantly simplified the analysis of large systems and made it possible to model complex environments like telecommunication systems, computer networks, logistics hubs and multi department service centres. Queuing network theory developed by Jackson and extended by others, including Gordon and Newell, placed the theoretical foundation for much simulation and performance modelling done today offering both analytical insight and practical tools for optimizing multi stage service systems.

2.3.4 Birth-Death process theory

The Birth-Death process developed through the early work of Andrey Kolmogorov et.al is an important mathematical modelling technique for queuing systems. A Birth-Death process is a Markov process modelling systems where changes in state occur due to births (arrivals) and deaths (departures or service completions). The number of entities in the system, and the model assumes that changes occur only to the neighbouring states. Kolmogorov's contribution in the form of his forward and backward equations provides us with the mathematical framework to analyse these processes over time. In queuing theory, the model is used to describe the dynamic process of arrival and servicing of customers that forms the basis of systems such as M/M/1 or M/M/c queues. The Birth-Death process allows analysts to calculate key presentation events such as average number of customers in the system, average waiting time and the server utilization. Its elegance lies in its probabilistic and time dependent nature, making it well suited to real time service systems where both arrivals and service events are random. The theory remains important to the study of queues, underpinning much of modern queuing analysis and simulation modelling.

2.3.5 Markovian queuing theory

The origins of Markovian queuing theory can be traced back to the early 1900s with the foundational work of Andrey Markov who introduced the concept of Markov chains stochastic processes where the upcoming event is determined by the current state and not on the order of events that preceded it that is the memory-less property. This property was essential in queuing theory where Markovian models describe systems with Poisson arrivals and exponentially distributed service times which form so called M/M/1, M/M/c, etc., queues. These models assume that arrivals and service completions are random but at constant average rates and that the chances of a transition from one event to another is a function only of the present state. Markovian queuing models are mathematically tractable in the sense that important performance measures such as average queue length, waiting time, and server utilization can be obtained by analysts through the use of steady state probabilities. Markov's theory provided the mathematical basis for the probabilistic modelling of queuing systems and remains a cornerstone of operations research especially in the modelling of dynamic, time dependent service systems like call centres, healthcare and computer networks.

2.3.6 Kendall Notations Theory

In the development of queuing theory, an innovation in the standardization and classification of queuing models was realized by D.G Kendall in 1953, through what is presently identified as Kendall's notation. This system provides the shorthand notation for defining the features of a queue model as $A/S/c$, where A is the process of arrival, S the service time distribution and c the number of servers. This was later extended to $A/S/c/K/N/D$, including system capacity represented by K , population size which is our N and queue discipline represented by D that is FIFO. For example, $M/M/1$ specifies a model with Poisson arrivals, exponential service times and a single-server. Kendall's notation is necessary because it allows researchers and practitioners to instantly understand and communicate the structure of a queuing system. It also allows different systems to be compared and analysed within the same framework, both for theoretical study and for applied modelling in many areas like telecommunications, healthcare, retail and public services. This notation have become an important factor for queuing theory, facilitating the study and application of the theory to complex service systems.

2.4 AN OVERVIEW OF QUEUING CONCEPTS

The queuing concept is the basic idea of how customers wait in line and receive service in a situation where demand occasionally exceeds capacity. It includes the study and management of waiting lines, with attention to queue action and performance whether in terms of how items arrive, how they get serviced, how long they wait and in what order.

2.4.1 Customer arrival patterns

The common scenario in various service systems such as banks, fast foods restaurant, super markets, etc. is when a customer have a choice of when to arrive to a facility that offers service that the customer wants. Arrivals may happen at organised times or random times. Customer's arrival patterns are influenced by several factors such as time of the day, week day and seasonality. Customers arrival often follows a daily pattern with peak arrivals during certain hours (Larson,1988). Arrival of customers can vary by the day of the week, with higher arrivals on certain days (kimes, 1989). Customer arrivals can exhibit seasonal patterns, with higher arrivals during certain times of the year (Gupta, 1999). A fundamental aspect of studying customer arrival patterns is understanding how they impact service systems, particularly in queuing models (Gross & Harris, 1998).

Brown et al. (2005) proposed a queuing model that captures historical variability in customer arrivals, demonstrating the impact of peak periods on system performance. These results are consistent with Gans et al. (2003), who noted that arrivals often cluster around specific times of the day, requiring dynamic resource allocation. Queuing theory has been widely used to model and analyse customer arrival patterns. Erlang (1909) first introduced the concept of poison arrivals in telecommunication system, which has been since extended to various service industries. More studies such as those by Whitt (2002), have incorporated non-poison arrival processes to account for variability and clustering in real world applications.

2.4.2 Queuing discipline

The queue discipline discusses the rules that determine the order in which customers are assisted in a queuing system. It answers the question of who will be served next. Common queue discipline involves first-in-first-out (FIFO), Last-in-first-out (LIFO), serve in random order (SIRO). Queues can be intangible, such as when a phone call is placed on hold or they can be made up of people or objects standing in a line (Tamrat Yifter et al, 2023). The following figure shows an example of a queuing discipline

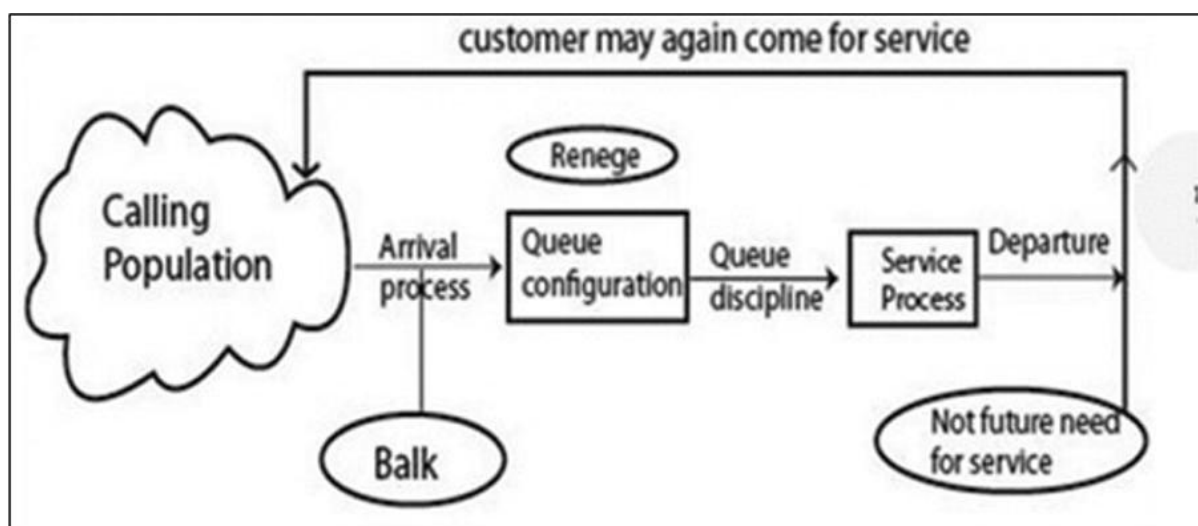


Figure 1:6 Queue Discipline

First In-First Out, or FIFO, suggests that services is offered in the order of arrival but the customers leave in different order due to the time it takes in. Last In-First Out, or LIFO, is the process wherein the item at the back of the line or the one that enters the system first is served (Ailobhio et al, 2021).

Suppose there are three customers X, Y, Z waiting at a counter for service and their service time is 20 minutes, 30 minutes and 40 minutes respectively. It takes $20 + 30 + 40 = 90$ minutes to finish all the service. However, the average waiting time before service for the three customers can be quite different for different service disciplines.

Case 1 (FIFO): The waiting time for the first customer is zero, for the next customer is 10 minutes and the waiting time for the last customer is $20+30=50$ minutes. Therefore, the average waiting time before service is $(0+20+30) = 50/3$

Case 2 (LIFO): the waiting time for first customer is zero, the waiting time for the second customer is 50 minutes and for the third customer is $50+20=70$ minutes. Therefore, the average waiting time before service is $(30+70) = 100/3$ minutes which is twice case 1 resulting in many systems preferring the case 1 scenario.

2.4.3 Queue Management

Queue management is a critical aspect of service system, aiming to minimize customer waiting time and optimize service efficiency. According to Lee (2019) queuing management is a group of guidelines for optimizing customer flow and advancing the queuing process. There are three techniques for queue management which are queue length control, service rate control and queue discipline. The queue length control technique involves controlling the number of customers in the queue to minimize waiting time (Larson,1988). The technique of the service rate control involves adjusting the service rate to match the arrival rate of customers, (Hopp & Spearman, 2000). The technique of the queue discipline involves managing the order in which customers receive service, such as using a FIFO or a priority queue discipline (Kendall, 1951). Strategies for queue management includes the FIFO, priority queuing and dynamic queuing. The most common queue management strategy is the First-in-First-Out (FIFO) where customers served in order they arrive, (Kendall, 1951). Priority queuing strategy assigns priorities to the customers based on their characteristics, such as urgency or importance, (Cobham 1954) for instance in a hospital a patient in an emergency state will be treated first, dynamic queuing strategy adjusts the queue management policy based on real time system conditions, (Maglaras & Meissner, 2006).

Strategies of Queue management includes queue length management, queue balancing and priority queuing. Managing the length of the queue is the process of watching and directing the size of the queues in many areas and the main aim is to make sure that queues do not become long leading to long waiting time and customer disappointment. Queue balancing is the process

of allocating customers evenly among multiple service points to ensure that queues do not become overly overcrowded while other service points remain underutilized and this is necessary to optimize efficiency and reduce wait times. Priority queuing is for certain people such as elderly or disabled. Queue management can help the rural council to effectively manage staff in accordance with current demand. The organization can understand the number of customers and their needs then adjust their staff accordingly so as to provide faster service and smarter use of staff time. Queue management can also make use of physical barriers such as ropes, chairs for customers to seat so as to guide customers in forming orderly lines (Wole 2024).

2.4.4 Service quality

The difference between customer prospects and their observations of the actual service delivered is what is called service quality (Parasuraman et al. 1985). Tangibles, reliability, responsiveness, assurance and empathy can be used to identify the extents of service quality. Tangibles are referred to as the physical evidence of the service which involves facilities and equipment, the capacity of the server to provide service as guaranteed is what is called reliability, the willingness of the server to contribute customers is the responsiveness, the confidence that customers have in the service provider is what is referred to as assurance and the degree to which the server appreciates and cares about customer wants is what is called empathy (Parasuraman et al 1985).

The SERVQUAL model is the most widely adopted agenda for assessing service quality which was developed by Parasuraman, Zaithaml and Berry (1988). This model measures the service quality based on the gap between customers' hopes and insights. Some studies suggested that the SERVQUAL model may not fit all service contexts equally and that it may need to be adjusted to suit specific industries such as healthcare, retail and hospitality (Carman, 1990). Brown et al. (1993) also highlighted the importance of introducing the SERVQUAL model in line with the service industry's nature. The SERPERF model was created and it only focuses on performance rather than the gap between expectations and observations (Cronin & Taylor, 1992). The SERPERF model simplifies the SERVQUAL model by removing the expectation factor that has been criticized on grounds of subjectivity.

There are factors that influence service quality such as employee competence and training, service environment and expectations of customers. The skills and training of employees can significantly impact the service quality (Schneider & Bowen, 1995). The physical environment

in which the service is delivered can impact the customer perceptions of service quality (Bitner, 1992). Customer expectations can significantly impact the customer perceptions of service quality (Parasuraman et al. 1985). Service quality directly impacts customer happiness, devotion and overall business achievement. Understanding service quality is essential for servers to provide maximum quality services that meet or go above customer expectations.

2.5 SIMULATION APPROACH

Simulation is a great instrument used to analyse and optimize difficult systems by simulating their performance over time (Law & Kelton, 2000). The simulation approach has been widely applied in various fields, including operation research, management science, engineering and economics. Building a false past structure and observing the past structure history to infer operating features of the real system is involved in simulation. Simulation can be done by hand, but usually, by computer with the appropriate software. A simulation model is usually developed by a set of expectations concerning the behaviour of a system. After validation and verification of the model, we can use it to investigate “what-if” scenario of the physical world structure and is thus useful in predicting potential areas of changes in the system.

Various types of simulation approach include system dynamics (SD), Monte carlo simulation, agent-based modelling (ABM), discrete event simulation (DES) and continuous simulation. Modelling systems where proceedings happens at a specific point in time is what DES focuses on. According to Law & Kelton (2007), this approach is particularly effective in analysing queuing system, such as customer service lines or production process. Vazquez-Serrano (2021), dynamic and difficult systems are addressed using a stochastic modelling approach called discrete-event simulation. In this research, the researcher found out that the most popular simulation software used is Arena and concluded that there is an up going trend in using discrete-event simulation to address matters at an operational level, yet few of discrete-event simulation applications existing actual implementations of the models in real life situation. ABM models the behaviour of individual agents their interaction within the system. Macal and North (2010) emphasized that agent-based simulation is ideal for studying complex adaptive systems such as social networks or ecological systems. System dynamics as described by Sterman (2000) involves modelling feedback loops and time delays in systems. It is commonly used in policy analysis and organisational decision making.

Advances in computational power and the availability of specialized software such as Any logic, simul8, and Arena have made simulation more accessible to researchers and practitioners. In order to have a successful simulation project, there are different phases that must be understood and followed. Realistic goals with careful planning and expectations are some of the requirements of simulation. Before taking a simulation project, adequate background information should be gathered with respect to the nature of the problem to decide whether simulation is a fitting solution.

2.6 COMBINATION OF QUEUING AND SIMULATION APPROACH

Simulation and queuing theory are two powerful approaches used to study and optimize complex systems. Queuing theory provides a mathematical model for examining waiting lines, simulation allows the analysis of complex systems and the measurement of how the systems work in the long run. Applying the two methods allows for better modelling, particularly in complex or non-standard systems where analytical solution may be challenging or impossible to reach.

Kalwar et al, (2021) decided that the study was conducted with both the queuing theory and simulation method with the aim of improving health care. The results showed that the patients are facing problems of delayed service, prolonged waiting times and the number of doctors is also below the number required. Queuing theory and simulation were used to resolve these problems in order to improve the whole health care system. Osorio and Bierlaire (2013) also conducted a study suggesting a mix of approaches that pairs queuing theory and microscopic traffic simulation to optimize urban traffic network. The queuing component modelled traffic situations at a small scale and simulation accounted for detailed driver behaviour and contact. Bolch et al. (2006) presented how queuing models can be used to forecast system performance under changing traffic levels and simulation can be used in order to study network reliability and scalability in real world settings.

2.7 EMPIRICAL EVIDENCE

A study by Kolesar (1984) provided the empirical validity of queuing theory models for performance prediction in service operations. With data collected from a number of service

settings, the paper evaluates the predictive accuracy of theoretical models for estimating key operational measures such as waiting times and server utilization. Results indicate that, with appropriate parameter estimation and model selection, queuing theory provides accurate predictions that closely follow observed system performance. The results highlight the usefulness of queuing models as practical instruments for enhancing resource allocation and customer service in intricate operational settings.

Green.L. et al (2007) analysed a class of non-stationary queuing models suitable for modelling services with time varying demand and delay sensitive customers. They proposed and assess an approach to calculating staffing requirements when both arrival rates and time of service vary over a period. The approach is an extension of the stationary M/M/s and employs approximations and corrections to address the non-stationary environment. Their numerical experiments and case studies show that the approach performs well in practice, giving good approximations to the time varying system dynamics and useful guidance on staffing levels. Green et al. emphasized the importance of understanding service variability and model calibration, highlighting queuing theory as an important tool in performance optimization in general complicated service system.

A study by Ferreira, R.B & Seneviratne P.N (1992) provides an empirical exploration of applications of queuing theory in urban traffic systems is providing examples of signalized intersections. By collecting and analysing traffic data from several intersections, the research evaluates the reliability and validity of queuing models for estimating vehicle delays and queue lengths. The findings confirm that queuing theory is a valid foundation for traffic flow analysis and control, with significant implications for signal timing optimization and intersection presentation. The study underlines the significance of merging empirical data with theoretical models in an effort to inform successful traffic management policy. Their research provided transportation planners and engineers with a foundation for applying queuing models as reliable tools for the analysis of traffic flow and the design of infrastructure facilities.

A study by Law and Kelton (2000) presented a complete, modern and technically accurate explanation of all major aspects of a simulation study. Their book seeks to make the content available through the use of intuition and a wide range of figures, examples and exercises. It is well suited for use in university coursework, simulation practice and self-study. Some of the topics covered by the book are introductory simulation modelling, complicated system modelling, simulation software, input probability distribution selection, random number

generation, output data analysis and comparing alternative system configurations. With its problem-solving focus and extensive coverage the book is an essential resource for practitioners involved in the simulation modelling and analysis of complex systems. Their work emphasized the value of simulation as a useful means of system behaviour understanding and operational effectiveness improvement.

A study by Aksin et al. (2007) combined queuing theory and simulation approach to analyse call centre performance. Call centres have been a successful area for operations supervision researchers in many areas, including predicting, volume planning and queuing personnel development. Furthermore, as information technology and telecommunication have progressive in recent years, call centre managers operation problems have increased in complexity. Human resource management, marketing and sales-related problems have also become increasingly important to call centre operations and to academia related research. They provided a review of the recent literature on call centre operations management. Besides historical research themes, they gave distinctive attention to new management problems that was addressed. Their work emphasized how essential it is to integrate theory models and empirical findings in order to push decision processes in call centres towards higher efficiency and customer focus.

2.8 THE GAP

Most of the industries such as healthcare, banking and transport have used queuing theory and simulation methods. Few researches focuses on simulating and optimizing revenue office efficiency. The revenue office uses a custom-made queuing model to understand the customer movement, reduce the waiting time and increase working efficiency. These models pays attention to elements such as the arrival rate, service and the number of cashiers to ensure maximum operation efficiency in the revenue office. Simulation is the only system able deal with real world difficult system. Simulation assumption-free permitting for flexible arrival and service times without requiring fixed distributions, making it a more accurate and flexible modelling approach. It can also replicate the behaviour of an existing system by joining suitable distributions and entering relevant data. It is capable of producing more accurate results.

2.9 CHAPTER SUMMARY

This chapter of the literature review gave literature associated with the subject under study. These studies discuss the effectiveness of queuing theory and simulation models in several

industries such as banking, healthcare, transport and fast-food restaurants. The next chapter focus on the research methodology and data collection techniques.

CHAPTER 3: METHODOLOGY

3.0 INTRODUCTION

In this chapter the researcher is going to talk about the methods used to achieve the research purposes and answer the research questions. The researcher presents the research plan, study area, aim population, data collection instrument, data presentation and analysis in this chapter. This chapter specifies the models expected by the study and offers the methods of assessing and examining the model.

3.1 RESEARCH DESIGN

Research design refers to the strategy of responding research questions by collecting, analysing and interpreting data. Research designs are generally categorized into five types based on the nature of the study and these are descriptive, exploratory experimental and research design correlation. The descriptive research design was used adopted was the researcher. This research design explains the characteristics and behaviour as they exist. The methods for collecting data on descriptive research design includes surveys and questionnaires, observational studies and case studies. The researcher used the observational studies on the current operations of the revenue office. The researcher collected data on customer arrival, service time and queue length. A queuing model for the revenue office was developed from this data and the model will be used to simulate different scenarios and identify the most suitable one.

3.2 STUDY AREA AND STUDY POPULATION

The study will take place in one of Zimbabwe's rural districts which is Masvingo rural district. The population under study was the customers and employees in the revenue office. A wide range of clients interact with the revenue office and these includes companies, private citizens, auditors, financial consultants. Council officials in the revenue office are accountable for providing customer service, handling payments, and managing the overall operation of the revenue office.

3.3 DATA COLLECTION

In this study the primary data which was collected from Masvingo rural district council through observations was used by the researcher. The data have information about the average rate of arrival for customers, the average time service for the customers and how many servers in the current system. A queuing model that can stimulate the flow of customers in the revenue office was developed and the areas for improvement will be identified. The researcher used a space of one month to collect the data.

3.4 DESCRIPTION OF VARIABLES AND EXPECTED RELATIONSHIP

Optimizing efficiency in the revenue office using queuing theory and simulation involves analysing different variables related to customer arrivals, service processes and system performance. Independent variables represent the factors that influence the performance and efficiency of the queuing system and these are rate of arrival (λ), rate of service (μ), number of cashiers (s), customer behaviour and service discipline. Rate of arrival (λ) refers to the average number of customer arriving at the revenue office at a time. As the arrival rate increases, the queuing becomes more crowded, leading to longer waiting times and lower efficiency if service capacity is not adjusted. Service rate (μ) is the average number of customers the cashier can serve at a time. Customer wait time is reduced by a higher service rate and it improves efficiency by minimizing the queue length. The number of service counters available to serve customers. Increasing the total number of cashiers reduces the load on each cashier, resulting in faster service and reduced waiting times. Patterns in customer arrival such as peak times or customer impatience that may lead to balking or reneging is what is referred to as the customer behaviour. Higher rates of balking or reneging reduce the effective arrival rate but may indicate inefficiencies in the system. Service discipline refers to the way customers are served. Priority rules can impact the perceived fairness and efficiency of the system, affecting customer satisfaction.

Dependent variables measures the performance and efficiency in the queuing system and these are average waiting time (W_q), average time in the system (W), queue length (L_q), system utilization (ρ), probability of balking or reneging and customer satisfaction. Increased arrival rate or decreased service rate leads to longer wait times. Optimizing service processes reduces W_q . Higher service efficiency reduces the time customers spend in the system. Queue length increases with higher arrival rates and decreases with faster service rates or more service. The

fraction of time at which servers are busy is called system utilization. Higher utilization indicates the system is operating near capacity, which can lead to delays. An optimal utilization rate balances efficiency and customer satisfaction. The likelihood that customers leave without being served is the probability of balking or reneging. High waiting times increase the probability of balking or reneging, reducing customer satisfaction and revenue. Lower wait times, fair service discipline, and reduced balking/reneging improve customer satisfaction.

3.5 PRE-TEST

Pre-tests played a vital role in this study on optimizing operational efficiency in the revenue office by ensuring the reliability and validity of data. The researcher specifically focused on assess data quality, check statistical assumptions.

3.6 QUEUING MODEL

The theory of waiting in line called Queuing theory is the arithmetic study of holding queues. A queue with multiple servers (c servers), where the rate of arrival and service time are exponentially distributed and uses an $M/M/c$ queue model. Arrival of customers is randomly and if they find the server busy they generally join the queue in front of the server and this is what is referred to as the system of the queue. The three components which are arrival process, service mechanism and queue discipline describes the system of the queue. The way customers arrive at the facility describes the process of arrival. Customers arrive to the system. The number of cashiers describes the service mechanism of a queuing system. There are rules that are used by the server to choose the next customer from the line when the server is done with serving of the current customer, this is referred to as the discipline of the queuing system. Queue disciplines normally used are LIFO whereby customers receive service using the last in first out manner, Priority where customers receive their service in mandate of their importance on the basis of service they requirements and on FIFO service is received on the basis of first in first out basis and one of the most popular queuing models is the $M/M/c$ model. The model makes it possible to calculate performance metrics based on arrival rates, service rates, and the number of servers.

3.6.1 Advantages of Queuing Model

Queuing models offers multiple advantages such as it identifies bottlenecks in service delivery by quantifies where and why delays occur and also measures peak vs. off-peak demand, helping

allocate staff and resources effectively. It also optimizes staff and resource allocation through its ability to forecast the perfect number of service counters required to minimize wait times and by balancing cost of staffing vs. customer wait times

They also reduces operational costs by **avoiding** overstaffing or understaffing. They supports data-driven decision making and also tests "what-if" scenarios without real world trial and error.

3.6.2 Kendall's Notation

A/B/m/K/n/D

- A: Distribution of the arrival time
- B: service rate distribution
- m: total number of servers
- K: System capacity
- n: Population size,
- D: Service discipline

3.6.3 Parameters

λ = Average arrival rate.

μ = Server's average service rate

L_q = Average number of customers in the queue

L_s = Average number of customers in the system

W_q = Average time customers spent in the queue

W_s = Average time customers spends in the system

k = the number of servers in the system

P_0 = The likelihood of the system being idle

ρ = traffic intensity that is the chance that the servers are busy at random time

P_s = The likelihood that exactly n customers are in the system

P_w = The chance that a customer has to wait in a queue

Arena software, Tora software and Microsoft Excel will be used for data examination and explanation of results. The data that was collected using observation will be entered in Excel spread sheet and will be imported to Arena for analysis.

3.7 MODEL ASSUMPTIONS

1. First come first served (FCFS) is the queue discipline
2. Arrival of customers per unit time follows a Poison probability distribution.
3. Time of service follows an exponential distribution
4. The queue is endless
5. The average rate of arrival is greater than the rate of service.
6. There is no balking that is no customer leaves without being served regardless of the size of the queue.

3.7.1 Multi-server Queue Model (M/M/k)

The researcher uses the multiple channel queuing structure model, which have more than two servers attending new customers. It assumes that the inter arrival time is exponentially distributed (poison process with rate λ), service time is also exponential with mean μ and c servers.

Key Performance Metrics

Equations of the model

Traffic intensity (Server Utilization):

- The utilization of the system is given by :

$$\rho = \frac{\lambda}{k\mu} \quad (3.1)$$

- This represents the average fraction of time the servers are busy. For stability $\rho < 1$

Probability that a system is idle:

- The possibility that no customers are present in the system:

$$P_0 = \left[\sum_{n=0}^{k-1} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n + \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k \frac{k\mu}{(k\mu - \lambda)} \right]^{-1} \quad (3.2)$$

- This involves summing up two parts:
 1. The sum of probabilities for having k customers (from 0 to k-1).
 2. The chance for taking k or added customers.

Probability that exactly n customers are present in the system:

$$P_n = \begin{cases} P_0 \left(\frac{\lambda}{\mu} \right)^n, & n < k \\ P_0 \frac{\rho^n}{k! k^{n-k}}, & n \geq k \end{cases} \quad (3.3)$$

Number of customers expected in the queue:

$$L_q = \frac{\lambda \mu \left(\frac{\lambda}{\mu} \right)^k}{(k-1)!(k\mu - \lambda)^2} P_0 \quad (3.4)$$

OR

$$L_q = L_s - \rho \quad (3.5)$$

Number of customers likely to be in the system

$$L_s = \frac{\lambda \mu \left(\frac{\lambda}{\mu} \right)^k}{(k-1)!(k\mu - \lambda)^2} + \frac{\lambda}{\mu} P_0 \quad (3.6)$$

Predictable time customers spent in the queue:

$$W_q = \frac{\mu \left(\frac{\lambda}{\mu}\right)^k}{(k-1)!(k\mu - \lambda)^2} P_0 \quad (3.7)$$

OR

$$W_q = W_s - \frac{1}{\mu} = \frac{L_q}{\lambda} \quad (3.8)$$

Expected time customers spent in the system:

$$W_s = \frac{\mu \left(\frac{\lambda}{\mu}\right)^k}{(k-1)!(k\mu - \lambda)^2} P_0 + \frac{1}{\mu} = \frac{L_s}{\lambda} \quad (3.9)$$

Probability that a customer has to wait in a queue:

$$P_w = \frac{\lambda^k \mu k}{k! \mu (k\mu - \lambda)} P_0 \quad (3.10)$$

3.8 MODEL SELECTION CRITERIA

Optimizing efficiency in the revenue office can be achieved by employing queuing theory and simulation approach. In combining the queuing theory and simulation approach, the choice of the model depends how difficulty the system is. The modest model include the M/M/1, M/G/1 and M/M/1, these are the best for simple, small-scale systems. The researcher used the multi-server queue model.

3.8.1 TORA Software

TORA is a software tool designed for solving a variety of operation research problems. It is widely used in academics and industry to analyse optimization problems such as linear programming, queuing theory, inventory management, project management and simulation. It makes use of rate of arrival of customer's λ and the rate of service time μ . Tora software could

suggest optimal staffing levels based on predicted customer traffic, reducing idle time for employees and wait time for customers. With real time data, Tora could adjust workflows dynamically based on observed demand, improving adaptability and performance. It will be used to make different scenarios in order for the researcher to make data driven decisions to optimize efficiency of the revenue office.

3.8.2 Arena Software

Arena is a software for discrete event simulation. It is commonly used for modelling, simulation and analysing complex systems and processes in industries such as manufacturing, logistics healthcare and service systems. Arena allows users to build simulation model visually using flowcharts and modules, making it easier to analyse system behaviour and optimize performance. The researcher is going to use Arena software to analyse queuing system in the revenue office. It helps identifies bottlenecks and inefficiencies in processes.

3.8.3 Simulation Model

A simulation model will be developed for the current system of the revenue office and to test different scenarios and strategies for optimizing efficiency in the revenue office. The model will be used evaluate the “what if” scenarios, testing different scenarios. The model will simulate the arrival pattern of citizens, their service time in the revenue office. By simulating various scenarios, the researcher can recognize areas that need developments and make data-driven decisions to optimize efficiency.

3.9 MODEL VALIDATION

Model validation includes verifying that the mathematical, statistical, and simulation models used in the study exactly reflect real world conditions. Model validation connects the gap between theory and practice. Several techniques and tests are used to validate the model, with each method offering different insights into different aspects of the model's performance.

One of the most common methods is the input-output validation. It ensures that the model correctly represents behaviour of the system. Also sensitivity analysis has a compulsory role to find how the model works with respect to varying input parameters. Together these methods form strong, actionable insights for optimization.

3.9.1 Input-Output Validation.

Input-output validation is a crucial process in revenue office efficiency that assures if our analytical models correctly represents real life processes. For example, if the model predicts that the adding two cashiers will reduce processing times by 30%, verification is to check for this scenario in a test branch and measure the actual impact. This validation process ensures that our suggestions on efficiency are evidence based and minimizes the risk of putting in place inefficient solutions, which leads to more reliable improvements in service delivery and usage of resources. Sensitivity analysis in inspects how variations in important input parameters like customer arrival rates, service times staffing levels affect the operational outputs of the revenue office, such as wait times and resource utilization

3.9.2 Sensitivity Analysis

Sensitivity analysis in reviews how differences in input parameters like customer arrival rates, service times staffing levels affect the operational outputs of the revenue office, such as wait times and resource utilization. By systematically testing different scenarios, we identify which factors that are most significantly influenced by efficiency metrics, allowing us to select high-impact interventions. This method confirms our optimization strategies remain strong under real world uncertainties and changing operational conditions.

3.10 CHAPTER SUMMARY

This chapter revealed the methods through which data was collected. It also shows the research methodology, research strategy, case study and data collection instruments used in the whole process. The next chapter 4 focuses on how the data was presented, interpreted, and analysed.

CHAPTER 4: DATA PRESENTATION, ANALYSIS AND DISCUSSION

4.0 INTRODUCTION

Data interpretation and the process of data analysis for the study are the main topics of this chapter. The main aim is to present a complete perception of the findings as a result of the application of Queuing models in optimization of efficiency. This chapter forms the basis for determining valuable visions, justifying the value of the structure proposed, and addressing the research goals outlined in Chapter 1.

4.1 DESCRIPTIVE STATISTICS

Table 1: Descriptive Statistics

VARIABLE	Service Time	Arrival rate
Statistic	64.0000	32.0000
Mean	6.7655	21.0000
Standard Deviation	7.8465	6.461653
Minimum data value	0.03	13.0000
Maximum data value	33.3	29.0000
Chi-square test	4.1	1.65
Kolmogrov-Smirnov test	0.109	
Standard Deviation	7.85	4.62
p-value	0.141	0.656
Degrees of freedom	2	3
Squared error	0.019576	0.038465

The descriptive statistics provide us with critical information regarding the customer service process in the revenue office. From the service times, we can observe that there were 64 customers registered, with the average service being about 6.77 minutes. Yet the time taken per customer to serve is between 0.03 and 33.3 minutes. The 7.85 minutes standard deviation shows us this huge range, which most likely reflects the range of services involved from simple payments up to more extensive enquiries.

Given the interarrival times, which indicate the frequency at which customers arrive, the data includes 32 readings for 4 weeks of hour-by-hour periods (eight hours of the day per 5 days of working days a week) and has an average of 21 minutes of time between arrivals. The arrival pattern is relatively regular, even though there is some variation as indicated by a standard deviation of 6.46 minutes. This between 13 and 29 minutes suggests a normal but slightly erratic customer flow. These are characteristic of a fairly evenly spaced-out customer flow, which is preferable for modelling a Poisson distribution, where arrivals are random but have a steady average rate.

Statistical tests also validate the choice of theoretical distributions. For inter arrival rates, the Chi-Square test statistic is 1.65 with degrees of freedom =3 with 0.656 as the p value. This significant p value indicates that the data fit extremely well to the Poisson model and there is no statistical ground to reject this model. Similarly, service times were tested for goodness of fit to the exponential model and the resulted in a Chi-Square of 4.1, 2 degree of freedom and a p value of 0.141 again showing good fit. The Kolmogorov-Smirnov value of 0.109 for service times, and low squared errors (0.0196 for service time and 0.0385 for inter arrival rate) asserts that these distributions are appropriate for simulating the queuing system precisely employing M/M/c models.

4.2 PRE-TEST

Prior to simulation and analytical modelling, we conducted principal statistical pre-tests on raw inter arrival rate and service time data to confirm that modelling assumptions were accurate and inform direct methodological decision. Preliminary tests provided an insight into the inherent traits and distributional properties of the data.

4.4.1 Chi-Square Test

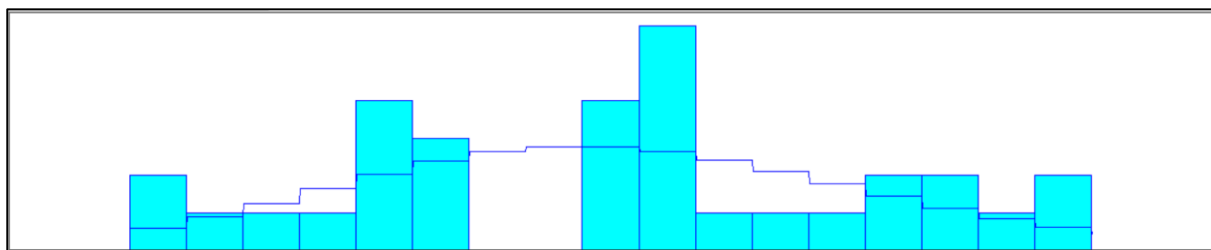


Figure 4:1 Historical of arrival time

The hypothesis test is to assess if the data follows a Poisson distribution with an average rate (λ) of 21. Using the Chi-Square goodness of fit test, the value of the test statistic is 1.65 with 9 degrees of freedom and a p-value of 0.656 well past the standard 0.05 significance level. This shows that there is no good reason to doubt that the data is adhering to a Poisson distribution. In order to confirm this test the sample mean is exactly 21, and the 4.62 standard deviation is actually very close to the predicted $\sqrt{21} \approx 4.58$ for Poisson data. The model also gave a very low squared error (0.038), which again indicates a good fit. Overall, both statistical evidence and simple measures indicate that the data is actually exhibiting a Poisson trend with $\lambda = 21$ quite reasonably.

4.2.2 Kolmogorov-Smirnov and Chi-Square test

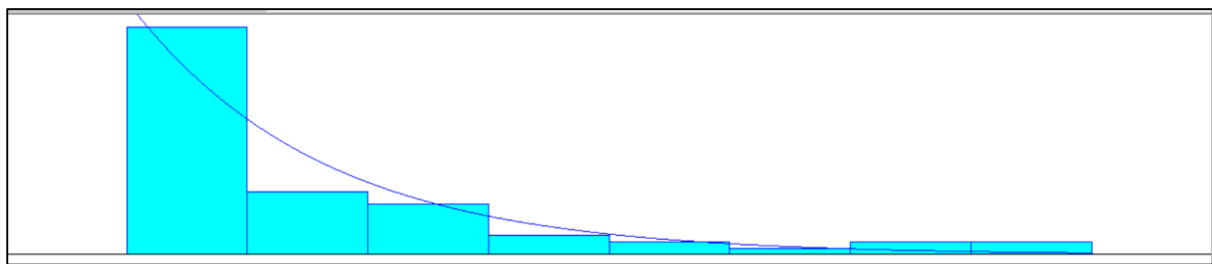


Figure 4:2: Service rate data plot

Statistical tests strictly validate that the data follows an exponential distribution. The Chi-Square test the p-value of 0.141 is larger than the 0.05 endpoint for no unusual lack of fit from an exponential model. The Kolmogorov-Smirnov test p-value is more than 0.15, validating the goodness of the exponential fit. With a squared error of 0.0186 and 64 data points, the exponential distribution EXPO (0) in particular provides a good fit of the data. The findings confirm that the data illustrates the memoryless property and thus is a good fit for models in applications for example, queuing systems or reliability analysis.

4.3 DATA ANALYSIS AND OBSERVATION

Arena simulation was used to model the flow of customers in the MRDC revenue office system. The existing configuration of the revenue office is shown in the figure below, and the observation after this simulation is explained below.

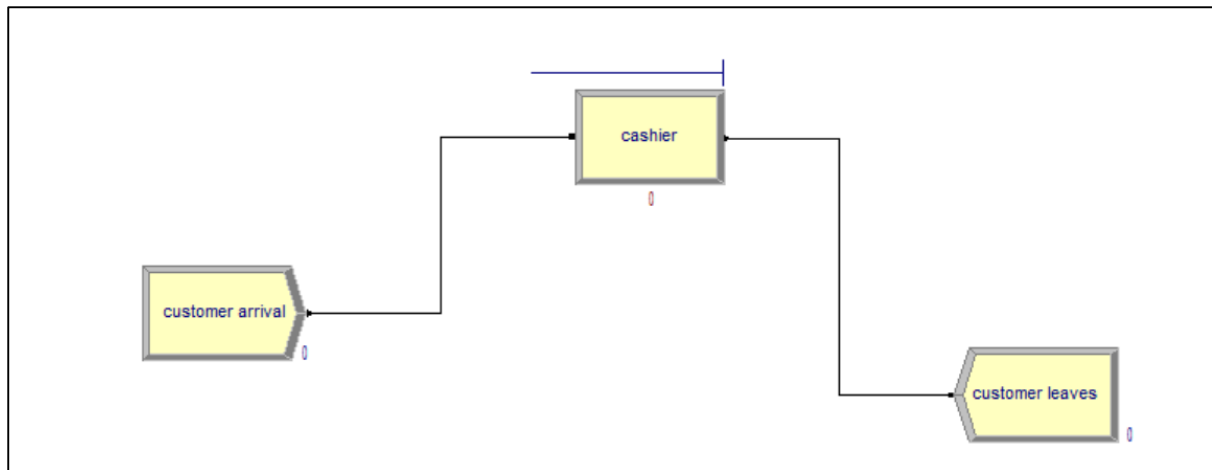


Figure 4:3 Simulation model of the revenue office

Arena simulation findings show very crucial information about the efficiency of the revenue office's operations that reveals enormous issues in managing the flow of customers. One of the most significant results are the averages of time in line to the cashier, which has the aggregate measurement of 34.24 minutes per lag well above acceptable levels of service for most public sector environments. Even worse is the variability of these wait times, with a standard deviation of 24.33 minutes and top delays of 87.44 minutes in certain replications. Taking these and adding them to the mean total system time of 39.73 minutes from arrival to departure, these numbers clearly paint the picture of a system weighed down by inefficiency and congestion. These findings are consistent with the fundamental principle of queuing theory that poor service capacity or poor queue discipline inevitably leads to long waits and systemic congestion (Gross et al., 2008). The magnitude of these delays means that the revenue office frequently is running at or even over its peak capacity, particularly during peak hours.

This is in contrast to the behaviour of a multi-server queuing system (e.g., M/M/c), where not enough cashiers for customer arrivals causes unmanageable congestion (Kleinrock, 1975). The reasons may be unequal difficulty of transactions, unpredictable bursts of arrivals, or inadequate deployment of staff. Also, the high variation between replications (as shown by the wide ranges of wait times) indicates the system's instability, which will undermine public trust in government services. As Naoum (2001) suggests, consistency is critical to service delivery in order to promote citizen confidence, especially in revenue collection, where delay will directly damage compliance and satisfaction. From the optimization viewpoint, such findings map out a number of opportunities for operational improvement. First, increasing the number of service counters at peak-demand periods could decrease congestion and average wait time.

Second implementing a triage system under which simple transactions are processed swiftly and complex cases are prioritized would render the whole process more efficient. Third putting routine processes such as online payments, form submission on auto-pilot could divert much of the demand away from queues, a strategy suggested by Al-Mudimigh et al. (2011) in their study on digital transformation for public services.

Alongside these in-service changes, the simulation provides a robust analytical platform for evidence-based decision making. Through models of different staffing, shift patterns, or queue control policies, policymakers are able to identify the most valuable interventions before they are implemented. A simulation of Arena with one additional cashier during peak store times, for instance, would be able to estimate potential wait time decreases, and priority queue experiments would establish their impact on speed and fairness of service.

These findings validate the worth of consolidating queuing theory and simulation modelling in public administration. Inefficiencies being exposed by this study are not theoretical they are actual life choke points that if addressed, might render service delivery much improved. As the revenue office adopts an evidence-based approach to optimization, it can move towards a more efficient, predictable, and citizen-friendly operation realizing the basic aim of this study

4.3.2 Queuing Analysis Using TORA Software and Arena Models

These findings validate the worth of consolidating queuing theory and simulation modelling in public administration. Inefficiencies being exposed by this study are not theoretical they are actual life choke points that if addressed, might render service delivery much improved. As the revenue office adopts an evidence-based approach to optimization, it can move towards a more efficient, predictable, and citizen-friendly operation realizing the basic aim of this study

Table 2: Performance measures of multi-server queuing model

Scenario	c	Lambda	Mu	L_{da}^{eff}	p_0	L_s	L_q	W_s	W_q
1	1	8.000	10.000	8.000	0.200	4.000	3.200	0.500	0.400
2	2	8.000	10.000	8.000	0.429	0.952	0.152	0.119	0.019
3	3	8.000	10.000	8.000	0.447	0.819	0.019	0.102	0.002

The single-server scenario (Scenario 1) mirrors the chronic understaffing which afflicts most public agencies, yielding shocking results of 80% server utilization, average queues of 3.2 customers, and waiting times of up to 24 minutes.

These findings are a confirmation of the backbiting cycle described by Gross et al. (2008), since system utilization approaches full capacity, delays grow geometrically. The application in real life of this mathematical principle strikes home painfully in the long lines of waiting and outraged citizens that characterize many revenue offices at peak hours.

Adding a second server (Scenario 2) has stunning consequences, shortening wait times to slightly more than one minute (1.14 minutes) while reducing waiting line lengths to a mere 0.15 customers. This 43% usage level is the ideal point of public service peak identified by Naoum (2001) as the right combination of service efficiency and fiscal responsibility. The system provides sufficient capacity for ordinary fluctuations in demand without too much idle, enabling both cost-effective operations and satisfactory citizen experiences. The one that follows is the arena model depicting a two-server environment and the result shows a reduction considerably with more servers.

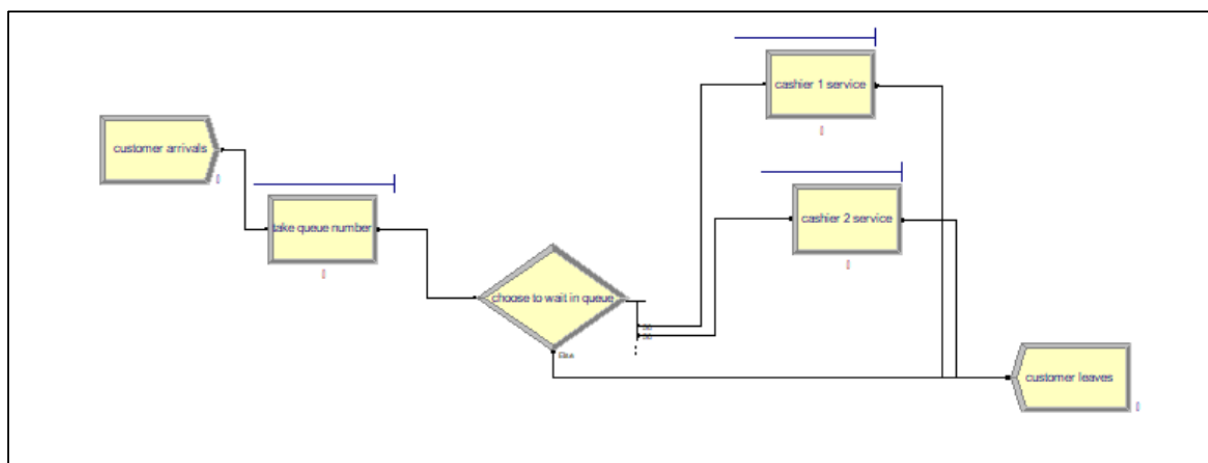


Figure 4:4 M/M/2 model

Despite the three-server system (Scenario 3) bringing down the change of service to rapid (7-second waits), its utilization rate of 45% is a sign of decreasing returns. As Brill (2008) cautions in his analysis of service systems, opportunity cost vs. marginal benefit for incremental capacity needs to be weighed. Again, here, the negligible difference with Scenario 2 suggests that the expense of the third server might be better allocated elsewhere into other service improvements or digital transformation initiatives. And the Arena M/M/3 model is illustrated below and it the model output reflects a slight deviation from the M/M/2 model indicating that the third server is not significantly contributing.

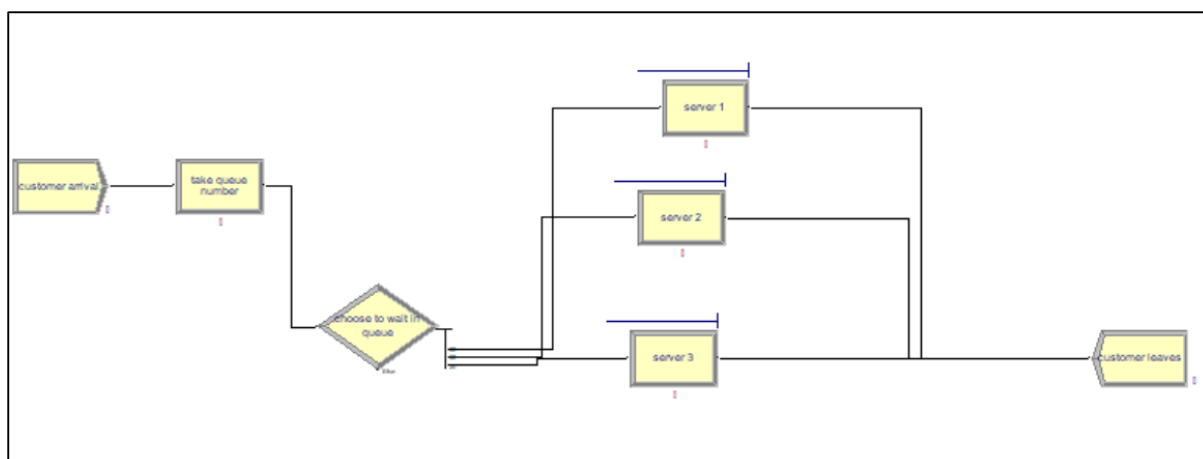


Figure 4:5 M/M/3 model

Implications for policy are clear and compelling. In the modelled demand conditions, two servers are the optimal configuration eliminating unwarranted waits while still maintaining coherent utilization of resources. This finding is even more relevant if it is considered in the context of the Arena simulation results, which measured actual waits over 30 minutes in current operations

Taken together, these analytical approaches offer supporting evidence to justify prompt changes in operations. Beyond direct immediate staffing adjustments, the study demonstrates the way quantitative modelling can inform more general public sector modernization. The TORA Arena combination offers local authorities a useful method for modelling the workings of process change from digital service delivery through to dynamic staffing models prior to implementation. In the Zimbabwean environment, where resources are scarce and demands are increasing for effective services, such evidence-driven approaches to administrative reform could be revolutionary.

This evaluation ultimately supports the core hypothesis of the study that operations research techniques, if used suitably to public service problems, can bridge the gap between mathematical theory and real governance solutions. The TORA results do not simply show better queue management they provide a prescribe-able framework for data driven decision making in resource deprived public institutions across the developing world.

4.3.3 Deeper Analysis and Connection of Arena and Tora

The queuing analysis executed using both Arena simulation and TORA Optimization System gives a complete assessment of service efficiency in Zimbabwe's council revenue offices

TORA vs. Arena: Queuing Performance Comparison

Table 3 : Tora vs. Arena Queuing Performance Comparison

Scenarios number of servers		ARENA RESULTS	TORA RESULTS
1 SERVER	Average in Queue (Lq)	3.77	3.2000
	Average in system (Ls)	5.3	4.000
	Average Wait Time in Queue, hrs (Wq)	0.5707 (34.24 minutes)	0.400 (24 minutes)
	Average Wait Time in system (Ws)	0.6622 (39.73 minutes)	0.50000 (30.0 minutes)
	Server utilization	75%	80 %
2 SERVERS	Average in Queue (Lq)	0.48	0.15238
	Average in system (Ls)	1.12	0.95238
	Average Wait Time in Queue, hrs (Wq)	2.67	0.01905 (1.14 minutes)
	Average Wait Time in system (Ws)	0.1400 (8.4 minutes)	0.11905 (7.14 minutes)
	Server utilization	40 %	43%
3 SERVERS	Average in Queue (Lq)	0.04	0.01892
	Average in system (Ls)	0.93	0.81892
	Average Wait Time in Queue, hour (Wq)	0.0067 (0.4 minutes)	0.00237 (0.14 minutes)
	Average Wait Time in system (Ws)	0.116 (6.96 minutes)	0.10237 (6.14 minutes)

	Server utilization	43 %	45 %
--	--------------------	------	------

The queuing analysis from Arena simulation and TORA Optimization System provides an overall evaluation of service efficiency in Zimbabwe's council revenue offices. Arena simulation captures the dynamic characteristic of real-world conditions of customer flow and reveals some of the severe operational problems. At 34.24-minute average wait and close-to-40-minute total system times, the simulation captures the unacceptable delays that citizens routinely experience. The findings are also validated by large standard deviations across replications, indicating unstable performance typical of common public complaints about erratic delivery of service.

The true worth of the simulation is its ability to capture the confused, variable nature of actual operations, such as unexpected arrival patterns and variable service times. These findings are corroborated by the TORA analysis with theoretical targets given for performance under controlled conditions. The single server test, 80% utilization and 24-minute average wait, is an exact replica of the Arena data, validating both approaches. This test clearly demonstrates how the current understaffed system creates chronic crowding. The dramatic reduction shown in the two-server mode to nearly one minute waits with reasonable 43% utilization makes a compelling case for restructuring for operations. At the same time, the small added benefit of the three-server option illustrates the law of diminishing returns in provision of public services.

The combination of these analytical approaches presents new contributions to public sector optimization. While Arena estimates the material impacts of inefficiencies in the real world, TORA provides the mathematical model for improvement strategies. Both of them constitute an evidence base that is theoretically difficult but practically feasible. The twin model approach is particularly valuable for developing countries, where the limitation of resources necessitates right, data driven decision making. The consistent counsel emerging from both analyses of changing technology to a two-server configuration would greatly improve service delivery while being cost-effective.

Along with short-term staff adjustments, this study creates the model for how operational research will contribute to transforming public administration. The methodology is an achievable framework for inefficient diagnosis and testing of solutions in whatever quantity of government services are wanted. For the local governments of Zimbabwe, adopting such evidence-based techniques could be a huge step toward reformed, citizen-centered-governance.

The findings uncover how quantitative analysis, if applied appropriately in solving public service issues, has the ability to link the gap between actual improvement in people's lives and the theoretical models.

4.4 MODEL VALIDATION TESTS

Model validation is a crucial process in confirming the reliability and precision of financial analytical models. In this chapter, the researcher explains and presents two primary validation tests used to ascertain the performance of the model, namely the Input-Output Validation and sensitivity analysis. The researcher can ascertain the strength, reliability, and analytical ability of the model using the application of these validation tests.

4.4.1 Input-Output Validation

To ensure integrity in our simulation environment, we conducted rigorous validation with two complementary approaches. First was input-output validation, where we made comparisons of key performance metrics between the Arena simulation and TORA's queuing analytical model under the same inputs. Using an arrival rate of 8 customers per hour and a service rate of 10 customers per hour in a M/M/1 environment, the steady state M/M/1 model developed by TORA calculations showed the average waiting time in the queue as 24 minutes and total system time as 30 minutes. The Arena simulation that accounted for real world variability provided values marginally higher at 34.24 minutes for average waiting time and 39.73 minutes for total system time. Though these values are 30-40% higher than TORA's predictions, such deviation was expected and indeed lends support to the simulation in the ability to detect dynamic system behaviour beyond the capability of theoretical models. The direction consistency between the two models again confirms Arena's usage in scenario testing while maintaining its more elaborate portrayal of operational conditions.

4.4.2 Sensitivity Analysis

Sensitivity analysis was the second validation where we systematically changed the number of servers to test model responsiveness. Results indicated an important efficiency benchmark when transitioning from one to two servers. In the TORA model, the improvement cut average waits by 96.3% from 24 min to a mere 1.14 min and reduced queue lengths from 3.2 customers to 0.15. Arena. simulations also showed dramatic improvement with waits. going down by over 85%.

. Adding a third server was, however, found to produce only marginal gains in the others: waits reduced only to 7 seconds in TORA and only improved marginally in Arena while rising in idle.

Both tests discovered this trend of diminishing returns, albeit with methodological difference and thus lent support to the conclusion that two servers is the optimal number. The test experiments in general show that both the Arena and TORA models have a lot of ability to offer good insight into decision-making within operations. The two models support one another well: TORA sets theoretical benchmarks in a simplified environment, whereas Arena simulates the complexities of real operations. Their consistent reporting of two servers as the efficiency frontier provides firm evidence-based advice on staffing levels in the revenue office. The minor inconsistencies between the models actually enhance their collective value, as they reflect the projected differences between theoretical prediction and real implementation. This rigorous validation process ensures that the simulation framework can be trusted to generate significant operational improvements in public service delivery.

4.5 DISCUSSION OF FINDINGS

Findings in this chapter present a detailed report of the MRDC revenue office queuing system through empirical insight and simulation modelling. Descriptive statistics are used to initiate the analysis with the revelation of predominant patterns in customer arrival behaviour as well as customer service. With a mean of 6.77 minutes and a high standard deviation of 7.85 minutes, the service process is found to be extremely variable ranging from very short jobs to lengthy transactions. Similarly, the inter arrival times are also very regular with a mean of 21 minutes and a standard deviation of 6.46 minutes. These assumptions prove that the arrival of customers can be well approximated by a Poisson distribution and service time would be best represented with an exponential distribution. Assumption tests like Chi-Square and Kolmogorov-Smirnov prove that these assumptions are true with high p-values (0.656 for interarrival and 0.141 for service) and lowest squared errors. This verification of response is crucial as it ensures accuracy of M/M/c queuing models to response for analysis.

Before venturing into simulation, pre-test simulation was conducted to confirm appropriateness of data for use in queuing theory. The response of inter arrival times Chi-Square test confirmed Poisson distribution fit by demonstrating a very good agreement between empirical and theoretical standard deviations. Equally, exponentially of service time was confirmed using Kolmogorov Smirnov and Chi-Square tests. This statistical rigour foundation provided a clear

platform from which Arena simulation and TORA optimisation could be conducted confidently since they input parameters are true representations of actual conditions. Arena simulation yielded extremely high levels of inefficiency in the current single-server system.

Average wait time of 34.24 minutes, maximum delay of 87.44 minutes, reflect a grossly overloaded system. This congestion translates to a total system time of nearly 40 minutes per customer, far above acceptable levels for a public service environment. The very high variability in waiting times also implies poor consistency in the provision of services a weak area for organizations that enjoy public trust. From queuing theory and literature, such excessive delays are typically the consequence of operating at or near capacity levels without adequate buffers. The findings strongly show that the single-server model is no longer suitable for the demand level at the MRDC office. The TORA model confirms the findings by analysis standards.

With simulation of different service situations having a constant rate of arrival of 8 customers/hr and rate of service of 10, the single-server model again resulted in disastrous outcomes: 80% server utilization and 24-minute average wait. While a third server delivered marginal increases only, a second server dramatically raised performance metrics that cut down wait times by more than 95% and queues to a negligibly small 0.15 customers. The sensitivity of a third server delivering marginal returns only would reflect diminishing returns. This equilibrium of productivity with cost-effectiveness in the form of two servers is based on tested operations research principles and adheres to the postulate that maximization is not in having excess capacity, but in intelligent utilization of resources.

It is complementarity between models which makes the combination of Arena and TORA stronger. Arena is a more detailed representation of the real-world dynamics like variability and randomness and shows a better view of customers' experience. TORA provides theoretical guidance and real-time what-if analysis at steady-state, however. Outcome similarity between each model most notably the identical recommendation for a two-server system makes study conclusions stronger. Arena's greater response times than TORA are as expected since real systems include inefficiency and randomness not found in analytical models.

Validation experiments were also employed to enhance the validity of the simulation model. Input-output validation showed agreement between Arena and TORA with identical parameters, while sensitivity analysis confirmed the optimality of the two-server system. On both platforms, scaling from one to two servers witnessed the most increase in performance,

while a third server provided the minimum incremental improvement. This conformity between modelling paradigms and sensitivity analyses confirms the validity of the simulation as well as analytic findings. More widely the study validates simulation modelling and queuing theory as effective tools for public administration, enabling evidence based decision making which can raise service delivery significantly while maximizing limited resources. In the case of revenue offices and similar other such public service contexts. This approach offers a realistic avenue to improving working efficiency as well as citizen satisfaction with evidence based improvements.

4.6 CHAPTER SUMMARY

This chapter, based on statistical analysis and advanced modeling techniques, has rigorously examined the service system of the revenue office in terms of efficiency of operations. It graphically portrays the image of a service system still wallowing in inefficiencies as a result of very high variability in processing times and inefficient resource allocation. Analyses provide a diagnostic summary of existing inefficiencies and suggest a clear path to improvement. The outcome suggests that strategic evidence-based interventions will improve service delivery in the revenue office, possibly extending to similar public service settings. The findings now form the foundation for the policy advice and plans for implementation in later chapters.

CHAPTER 5: SUMMARY CONCLUSION AND RECOMMENDATIONS

5.0: INTRODUCTION

This chapter presents the summary of the research findings and conclusions and recommendations the researcher arrived at upon consideration of the objective of the study. The chapter also identifies areas for further research that were not emphasized by the researcher in the study.

5.1 SUMMARY OF THE STUDY AND FINDINGS

This study was focused on assessing and refining the efficiency of service delivery at the Masvingo Rural District Council (MRDC) revenue department, where frequent delays for customers, long queues, and unnecessary underutilization of personnel were noted. The most important research concern was to minimize customer waiting time and optimize personnel resource usage since inefficiency in the current queuing system was impacting operation performance as well as public satisfaction. This the research applied queuing theory and simulation modelling to study system performance and suggest useful improvements.

This study was directed by four major research objectives which were to investigate the efficiency of waiting time and utilization of staff in the office, to develop queuing and simulation models of customers in MRDC revenue office, to reduce the average waiting time of the customer and to reduce the operating costs of the revenue collecting office at Masvingo Rural District Council.

Study delimitation restricted the study to MRDC revenue collection activities alone, focusing on customer arrival rates, service times, staffing (servers) and internal queuing efficiency. External financial or policy procedures were excluded. The study was also restricted to weekday operation (8:00 Am to 4:00 PM) over four weeks.

The research applied quantitative methods based on operational research methodology, and combined simulation and analytical modelling techniques. Raw data for inter arrival and service times were collected and fit tested with assumptions of queuing theory. Times of arrival were Poisson and the service times exponential, hence using M/M/c queuing models was appropriate. Arena simulation software was used to model real queuing behaviour under varied

server conditions and record dynamic variability. At the same time, TORA software was used for theory purposes, calculating steady-state measures to be used for comparative purposes.

In achieving the first objective, the research found that MRDC's current one-server system is highly inefficient. Arena predicted an average wait time of 34.24 minutes in queue and peak wait time of 87.44 minutes, with overall system time at almost 40 minutes. TORA replicated these findings, showing 80% utilization, high queues (3.2 customers), and high delay. These results showed high bottlenecks in service provision, particularly at times of heavy loads when the system tended to run at or even above full capacity.

To achieve the second objective, Arena and TORA were utilized to create satisfactory queuing models representing MRDC's service system for different scenarios (1, 2, and 3 servers). The models allowed intensive analysis of arrival rates, capacity for service and system performance and were a starting point for experimental evaluation of alternative optimization techniques. They confirmed the feasibility of M/M/c modelling in simulating and analysing public service queuing systems.

To the third objective of reducing average customer waiting time, the models positively demonstrated that a growth in servers from one to two immensely reduced waiting times. For TORA, wait time fell from 24 minutes to 1.14 minutes; in Arena, by over 85%. Queue length was markedly reduced, and service consistency increased. A third server offered marginal improvement, indicating that the two-server design offers the best balance between performance and realism.

Finally, regarding the fourth objective of reducing operating costs the study found that while a third server had a negligible impact of reducing wait time, it led to higher idle time and wasting of resources. The two-server system operated at a utilization level of around 43% without wastage. This is an economical running scenario that maximizes service delivery without excessive increase in resources.

In summary, the study validated that the MRDC revenue office can achieve significant gains in efficiency by implementing a two-server queuing system. The conclusion verifies the use of simulation and analytical techniques in locating inefficiencies and creating evidence-based operational reforms for public agencies.

5.3: RECOMMENDATIONS

Based on what the present study has shown, it's clear that the revenue office would benefit more from adding an additional service counter. Just adding one more staff member would cut down on long wait times considerably and get the whole process going a whole lot smoother. This change is an extremely cost-friendly solution with huge dividends without requiring unnecessary expansion of resources. Furthermore, using a separate service process for complex or long cases would maintain the process of normal operations and reduce conflict. To ensure that it runs smoothly in the long term, the office should regularly monitor its performance with the use of simple tools such as queuing models and simulation tools to aid data driven decision making and guarantee continuous development. Staff training investment in order to minimize service time variability could additionally maximize consistency and satisfaction of customers.

5.4 AREAS FOR FURTHER RESEARCH

This work has left some doors open for further research, basing its conclusions and findings on the previous chapters. One of them would be to conduct a rigorous economic analysis of recruiting more employees against the benefits accrued through reduced waiting times and increased customer satisfaction. Somewhere else to check is how technology like online booking systems, queue tracking applications or self-service kiosks can potentially reduce waiting times and crowd levels. Also, a comparison between similar offices in other locations may help identify whether the same solutions work somewhere else. These research possibilities can continue to inform and allow for more informed, effective change in public service delivery.

REFERENCES

- Aksin, O. Z., Armony, M., & Mehrotra, V. (2007). *The modern call centre: A multidisciplinary perspective on operations management*. Springer.
- Ailobhio, D., Owolabi, T., & Ayoo, P. (2021). Application of queuing theory in antenatal clinics. *IOSR Journal of Mathematics (IOSR-JM)*, *16*(6), 42–47.
- Basak, A., & Choudhury, A. (2021). Bayesian inference and prediction in single server M/M/1 queuing model based on queue length. *Communications in Statistics-Simulation and Computation*, *50*(6), 1576–1588. <https://doi.org/10.1080/03610918.2019.1586938>
- Bitner, M. J., Booms, B. H., & Tetreault, M. S. (1990). The service encounter: Diagnosing favorable and unfavorable incidents. *Journal of Marketing*, *54*(1), 71–84. <https://doi.org/10.1177/002224299005400105>
- Bolch, G., Greiner, S., De Meer, H., & Trivedi, K. S. (2006). *Queuing networks and Markov chains: Modeling and performance evaluation with computer science applications* (2nd ed.). Wiley.
- Brahma, S. (2013). Queuing theory. *International Journal of Scientific and Research Publications*, *3*(6), 1–8.
- Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., & Zhao, L. (2005). Statistical analysis of a telephone call center: A queuing-science perspective. *Journal of the American Statistical Association*, *100*(469), 36–50. <https://doi.org/10.1198/016214504000001808>
- Carman, J. M. (1990). Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions. *Journal of Retailing*, *66*(1), 33–55.
- Cobham, A. (1954). Priority assignment in waiting line problems. *Journal of the Operations Research Society of America*, *2*(2), 175–182. <https://doi.org/10.1287/opre.2.2.175>

Cooper, R. B. (1981). *Introduction to queuing theory* (2nd ed.). Elsevier North-Holland.

Cronin, J. J., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, *56*(3), 55–68. <https://doi.org/10.1177/002224299205600304>

Erlang, A. K. (1909). The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B*, *20*, 33–39.

Erlang, A. K. (1917). Solutions to some problems in the theory of probabilities of significance in automatic telephone exchanges. *Elektroteknikerne*, *13*, 5–13.

Ferreira, R. B., & Seneviratne, P. N. (1992). Simulation modeling of passenger queues at airport terminals. *Journal of Transportation Engineering*, *118*(4), 525–539. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1992\)118:4\(525\)](https://doi.org/10.1061/(ASCE)0733-947X(1992)118:4(525))

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman Publishing.

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, *5*(2), 79–141. <https://doi.org/10.1287/msom.5.2.79.16071>

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Polity Press.

Green, L. V., Savin, S., & Murray, M. (2007). Providing timely access to care: What is the right patient panel size? *The Joint Commission Journal on Quality and Patient Safety*, *33*(4), 211–218. [https://doi.org/10.1016/S1553-7250\(07\)33024-2](https://doi.org/10.1016/S1553-7250(07)33024-2)

Gross, D., & Harris, C. M. (1998). *Fundamentals of queuing theory* (3rd ed.). Wiley-Interscience.

Gupta, S. (1999). Impact of seasonality on service systems. *Journal of Service Research*, *1*(2), 151–165. <https://doi.org/10.1177/109467059912006>

Hasugian, I. A., & Dewi, E. (2020). Analysis of queuing models of fast-food restaurant with simulation approach. *IOP Conference Series: Materials Science and Engineering*, *851*(1), 012028. <https://doi.org/10.1088/1757-899X/851/1/012028>

Hillier, F. S., & Lieberman, G. J. (2020). *Introduction to operations research* (11th ed.). McGraw-Hill Education.

Hopp, W. J., & Spearman, M. L. (2000). *Factory physics: Foundations of manufacturing management* (2nd ed.). McGraw-Hill.

Kalwar, M. A., Marri, H. B., Khan, M. A., & Khaskheli, S. A. (2021). Applications of queuing theory and discrete event simulation in health care units of Pakistan. *International Journal of Science and Engineering Investigations*, *10*(109), 6–18.

Kendall, D. G. (1951). Some problems in the theory of queues. *Journal of the Royal Statistical Society: Series B (Methodological)*, *13*(2), 151–185. <https://doi.org/10.1111/j.2517-6161.1951.tb00082.x>

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *Annals of Mathematical Statistics*, *24*(3), 338–354. <https://doi.org/10.1214/aoms/1177728975>

Keynes, J. M. (1936). *The general theory of employment, interest, and money*. Macmillan.

Kimes, S. E. (1989). A simulation model for restaurant table management. *Journal of Foodservice Management*, *2*(1), 31–34.

Kleinrock, L. (1975). *Queuing systems, Volume 1: Theory*. Wiley-Interscience.

Kolesar, P. (1984). Stalking the elusive goal of production scheduling. *Journal of Operations Management*, *4*(2), 109–123. [https://doi.org/10.1016/0272-6963\(84\)90003-2](https://doi.org/10.1016/0272-6963(84)90003-2)

Larson, R. C. (1988). A queuing model for police patrol operations. *Journal of Operations Management*, *7*(1), 1–5. [https://doi.org/10.1016/0272-6963\(88\)90002-0](https://doi.org/10.1016/0272-6963(88)90002-0)

Law, A. M., & Kelton, W. D. (2000). *Simulation modeling and analysis* (3rd ed.). McGraw-Hill.

Law, A. M., & Kelton, W. D. (2007). *Simulation modeling and analysis* (4th ed.). McGraw-Hill Education.

Lindley, D. V. (1952). The theory of queues with a single server. *Proceedings of the Cambridge Philosophical Society*, *48*(2), 277–289. <https://doi.org/10.1017/S0305004100027638>

Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modeling and simulation. *Journal of Simulation*, *4*(3), 151–162. <https://doi.org/10.1057/jos.2010.3>

Maglaras, C., & Meissner, J. (2006). Dynamic pricing strategies for managing queuing systems. *Journal of Service Research*, *9*(1), 3–17. <https://doi.org/10.1177/1094670506289525>

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396. <https://doi.org/10.1037/h0054346>

Osorio, C., & Bierlaire, M. (2013). A queuing network model with a microscopic traffic simulator for traffic congestion management. *Transportation Research Part C: Emerging Technologies*, *38*, 171–185. <https://doi.org/10.1016/j.trc.2013.11.008>

Palm, C. (1938). Analysis of the Erlang traffic formula for busy-signal arrangements. *Ericsson Technics*, *4*, 1–34.

Pareto, V. (1906). *Manuale di economia politica*. Società Editrice Libreria.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, *49*(4), 41–50. <https://doi.org/10.1177/002224298504900403>

Samuelson, P. A., & Nordhaus, W. D. (1995). *Economics* (15th ed.). McGraw-Hill Education.

Schneider, B., & Bowen, D. E. (1995). *Winning the service game*. Harvard Business School Press.

Slack, N., Chambers, S., & Johnston, R. (2013). *Operations management* (7th ed.). Pearson Education.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. McGraw-Hill.

Takagi, H. (1991). *Queuing analysis: A foundation of performance evaluation* (Vol. 1). Wiley-Interscience.

Vázquez-Serrano, J. I., Peimbert-García, R. E., & Cárdenas-Barrón, L. E. (2021). Discrete-event simulation modeling in healthcare: A comprehensive review. *International Journal of Environmental Research and Public Health*, *18*(22), 12262. <https://doi.org/10.3390/ijerph182212262>

Whitt, W. (2002). *Stochastic-process limits: An introduction to stochastic-process limits and their application to queues*. Springer Science & Business Media.

Wolff, R. W. (1989). *Stochastic modeling and the theory of queues*. Prentice Hall.

APPENDICES

TURNIT IN REPORT

Optimizing Efficiency in the Revenue office of Masvingo Rural District Council. A Queuing Theory and Simulation Approach

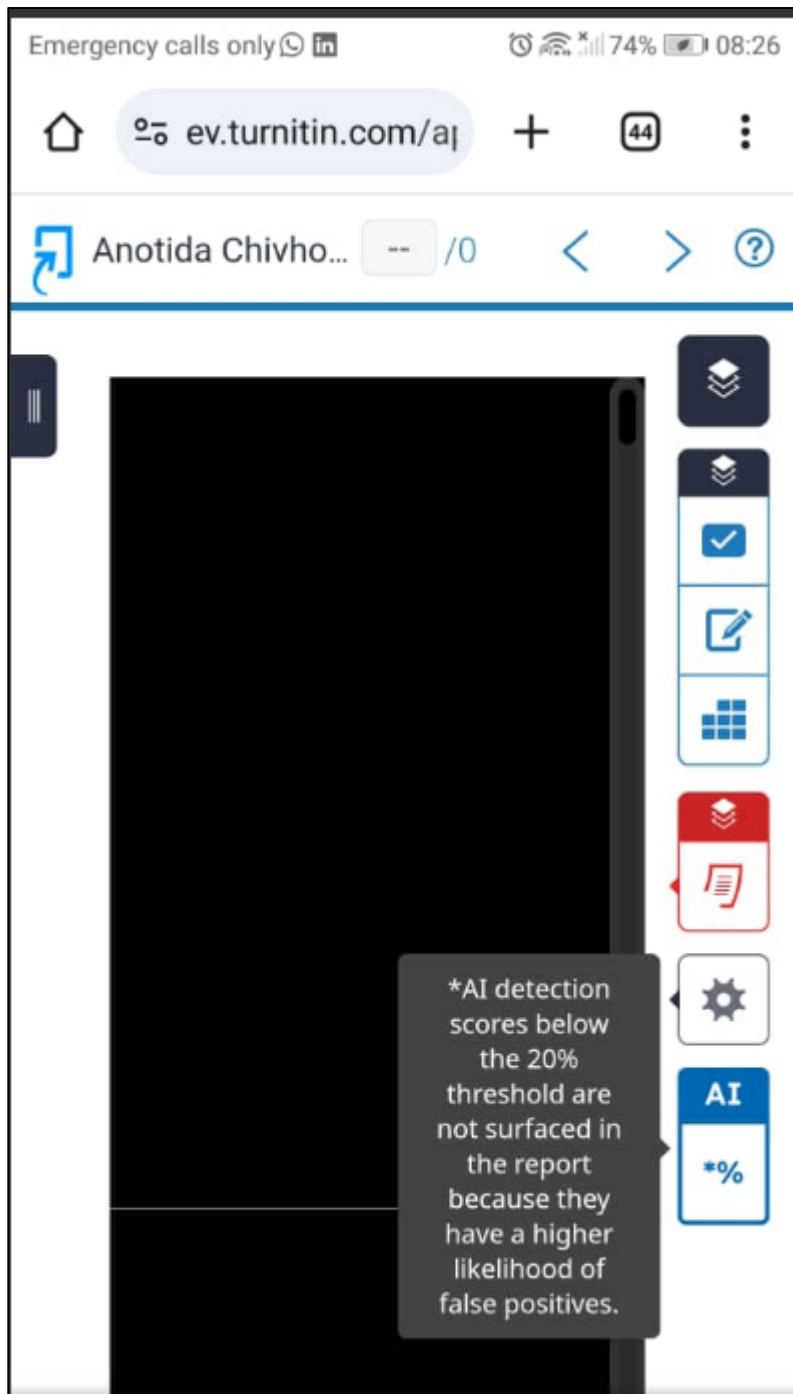
ORIGINALITY REPORT

9%	5%	5%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hkumath.hku.hk Internet Source	1%
2	Manuel Laguna, Johan Marklund. "Business Process Analytics - Modeling, Simulation, and Design", CRC Press, 2025 Publication	1%
3	www.isarsoft.com Internet Source	<1%
4	brightideas.houstontx.gov Internet Source	<1%
5	Statistics for Industry and Technology, 2015. Publication	<1%
6	dreamsupport.us Internet Source	<1%
7	Joti Lal Jain, Sri Gopal Mohanty, Walter Böhm. "A Course on Queueing Models", Chapman and Hall/CRC, 2019 Publication	<1%
8	repository.dinus.ac.id Internet Source	<1%
9	Submitted to Federal University of Technology Student Paper	<1%
10	Submitted to University of Technology Student Paper	<1%

AI Report

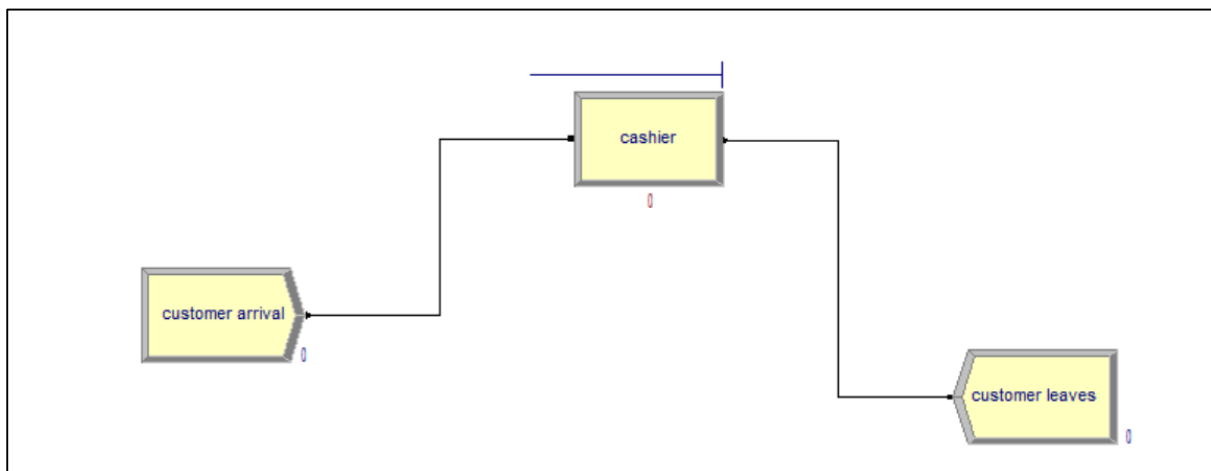


Queuing Theory

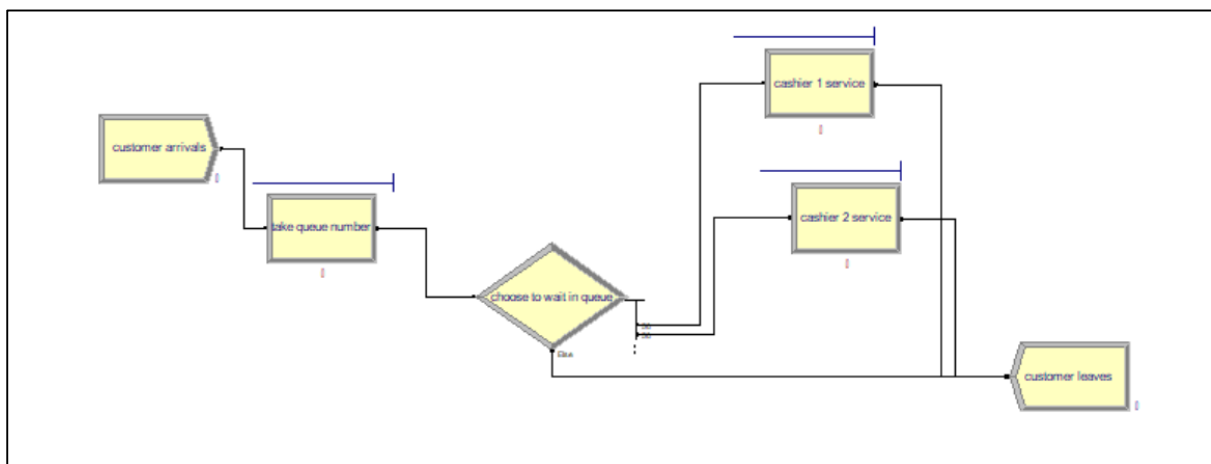
Scenario	c	Lambda	Mu	<u>L'da eff</u>	p0	<u>Is</u>	<u>Lq</u>	<u>Ws</u>	<u>Wq</u>
1	1	8.000	10.000	8.000	0.200	4.000	3.200	0.500	0.400
2	2	8.000	10.000	8.000	0.429	0.952	0.152	0.119	0.019
3	3	8.000	10.000	8.000	0.447	0.819	0.019	0.102	0.002

Simulation models

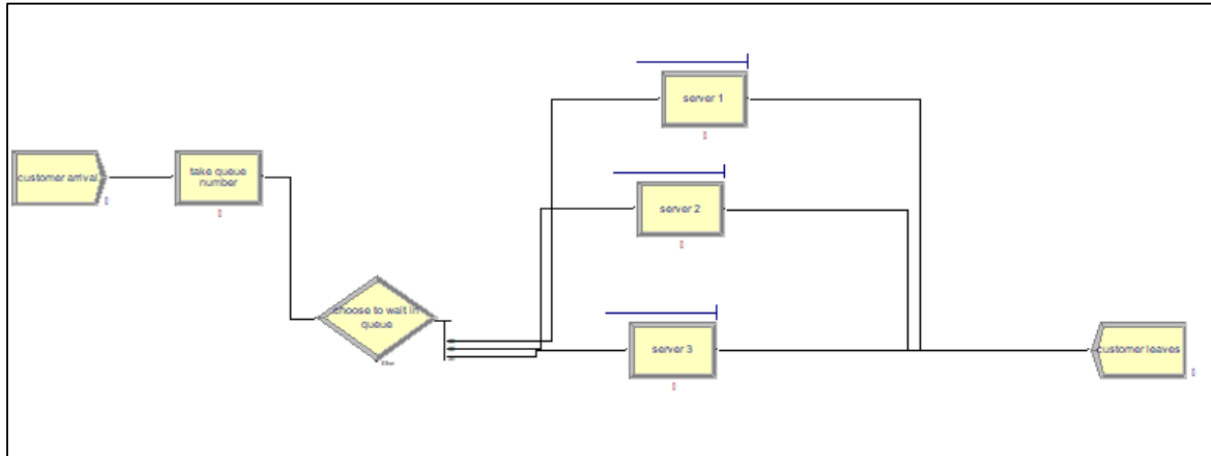
M/M/1 Model



M/M/2 Model



M/M/3 Model



All correspondences to be addressed to:
THE CHIEF EXECUTIVE OFFICER



**MASVINGO RURAL DISTRICT
COUNCIL**

P.O. Box 724 Masvingo
Tel: 039-262079/ 0773907417/ 0712648515
Website: www.masvingordc.co.zw



28 January 2025

To Whom it May Concern

RE: PERMISSION TO CONDUCT DATA COLLECTION

This letter serves to confirm that Masvingo Rural District Council has granted permission to Anotida E Chivhoko a student from Bindura University of Science Education, to conduct data collection at our organisation in partial fulfilment of their academic research requirements.

The purpose of the study, as communicated to us, is to Optimize Efficiency in the revenue office of Masvingo Rural District Council. The student is permitted to access relevant non-confidential information. All activities must be conducted in accordance with our operational protocols and ethical guidelines.

We are pleased to support this academic effort and wish the Anotida every success in her study.

Yours faithfully

Mr C. Shaimarai

Finance Director

