

**PERFORMANCE EVALUATION AND OPTIMAL TELLER
CAPACITY
DETERMINATION USING AN $M/M/c$ QUEUEING MODEL**

Abstract. Efficient queue management plays an important role in enhancing customer service delivery and operational efficiency at GCB Bank PLC, Tamale branch. Long customer waiting times arising from inadequate teller capacity can adversely affect customer satisfaction and the overall performance of banking operations. This study evaluated the performance of an $M/M/c$ queueing model and determined the optimal teller capacity for improving customer service delivery at GCB Bank PLC, Tamale branch. Specifically, the study sought to determine the steady-state performance measures of the banking queueing system, examine the effect of increasing teller capacity through sensitivity analysis, and determine the optimal teller capacity that minimises customer waiting while maintaining efficient utilisation of banking resources. The study employed the classical $M/M/c$ queueing model under the assumptions of Poisson customer arrivals, exponential service times, First-Come, First-Served (FCFS) service discipline, infinite calling population, unlimited queue capacity and steady-state operating conditions. An arrival rate of 45 customers per hour and a service rate of 12 customers per hour per teller were used in the analysis. The steady-state performance measures comprising traffic intensity, probability of waiting, expected queue length, average waiting time in the queue, average time spent in the system and expected number of customers in the system were computed for teller capacities ranging from eight to twelve. In addition, sensitivity analysis was conducted to evaluate the effect of increasing teller capacity on the performance of the banking queueing system. The results revealed that increasing teller capacity from eight to twelve progressively reduced traffic intensity from 0.46875 to 0.31250, the probability of waiting from 0.01473 to 0.00025, and the expected queue length from 0.01296 to 0.00011 customers. Similarly, the average waiting time in the queue decreased from 0.00029 hours to 0.000024 hours, while the average time spent in the system reduced slightly from 0.08362 hours to 0.08334 hours. Sensitivity analysis further showed that increasing teller capacity substantially improved queue performance, although the magnitude of improvement decreased as more tellers were added. Additionally, the data showed that teller idle capacity increased continuously, indicating diminishing gains in service enhancement above a particular staffing level. Ten tellers were found to be the ideal staffing level for the banking system under the projected arrival and service rates based on the trade-off between queue performance and teller utilisation. According to the study's findings, the $M/M/c$ queueing model offers a useful framework for making decisions on how best to staff tellers and assess the performance of banking services. The results show that while maintaining the effective use of banking resources, optimal teller allocation can greatly minimise customer waiting times. The study recommends the adoption of queueing models in banking operations to support evidence-based staffing decisions and improve customer service delivery.

2020 Mathematics Subject Classification. 60K25, 90B22, 60J2760K25, 90B22, 60J27.

Key words and phrases. queueing theory, $M/M/c$ model, teller capacity, customer waiting time, traffic intensity, sensitivity analysis.

1. Introduction

Queueing theory is a branch of operations research and applied probability that studies congestion and waiting lines in service systems. It provides mathematical models for analysing the interaction between customers arriving for service and the service facilities available to attend to them. Since its introduction by Agner Krarup Erlang in the early twentieth century to analyse telephone traffic congestion, queueing theory has evolved into one of the most important analytical tools for solving service-related problems in banking, healthcare, transportation, manufacturing, telecommunication, computer networks and public administration. The banking industry is one of the major sectors where queueing systems are frequently encountered. Customers continuously arrive at banking halls seeking different services including cash withdrawals, deposits, account opening, loan enquiries, cheque processing and customer support. Since the number of service personnel (tellers) is limited, customers often experience delays before receiving service. Long waiting times reduce customer satisfaction, lower service quality, increase operational costs and may eventually result in customer attrition. Modern banking institutions therefore seek efficient mechanisms for balancing customer waiting time with operational cost. Increasing the number of tellers indiscriminately reduces waiting time but significantly increases labour and operational costs. Conversely, reducing the number of servers minimises operating expenses but leads to longer queues and customer dissatisfaction. Queueing theory provides a scientific approach for determining the optimal balance between these competing objectives. Among the various queueing models developed in the literature, the multi-server Markovian queue, denoted by the $M/M/c$ model, has gained considerable attention because it closely represents many banking environments where multiple tellers simultaneously serve customers arriving randomly. The model assumes that customer arrivals follow a Poisson process, service times follow an exponential distribution, there are c identical servers operating in parallel, the calling population is infinite and customers are served according

to the First-Come-First-Served (FCFS) discipline.

A recent banking application also reported that increasing service capacity within a queueing framework was associated with lower customer waiting times, reinforcing the relevance of server-capacity analysis in bank service systems (Iwa & Audu, 2023).

Several empirical studies have demonstrated the usefulness of queueing models in improving banking operations. Odirichukwu et al. (2014) developed a queue management system for Nigerian banks using the First-In-First-Out (FIFO) principle and reported improvements in service **organisation** and reductions in customer congestion. Similarly, Afolalu et al. (2019) reviewed the application of queueing theory in banking and concluded that models such as $M/M/1$, $M/G/1$, $M/M/c$, Artificial Neural Networks and Erlang formulations have significantly improved productivity and customer service, although additional improvements remain necessary due to changing banking environments. Recent empirical studies have increasingly focused on multi-server systems because they better represent modern commercial banks. Akande et al. (2022) applied the $M/M/c$ queueing model to Automated Teller Machine (ATM) services and demonstrated that the model effectively evaluates server **utilisation**, average waiting time, queue length and service efficiency. Their findings further showed that queueing analysis assists management in determining whether additional servers are economically justified. In addition to mathematical modelling, technological advancements have introduced automated queue management systems that improve service delivery by electronically managing customer flow, allocating service tokens and monitoring service performance. Automated systems provide useful operational data that support queueing analysis and informed managerial decision-making.

Simulation-based analysis at the Commercial Bank of Ethiopia likewise showed that queueing models can be used to compare alternative service configurations and evaluate waiting-time performance (Yifter et al., 2023).

Despite these developments, long customer waiting times remain common in many commercial banks, particularly during peak operating hours. Consequently, there is a continuing need to develop mathematical models capable of analysing banking queue performance and recommending optimal service configurations. This study therefore applies the $M/M/c$ queueing model to analyse customer waiting times in a banking system with the aim of improving service efficiency while **minimising** customer waiting time and operational cost. Kendall (1953) introduced a **standardised** notation for describing queueing systems. Kendall's notation simplified the classification of queueing models and remains the universally accepted framework for describing arrival processes, service distributions, number of servers, queue capacity and service discipline. Over the past several decades, queueing theory has experienced remarkable growth due to advances in stochastic processes, probability theory, Markov chains and computer simulation. According to Bolch et al. (2006), advances in computational approaches have enabled researchers to analyse complex multi-server systems that were previously difficult to solve analytically. Computer-based queueing analysis now supports decision-making in banks, cloud computing, telecommunication networks, manufacturing industries, transportation systems and healthcare institutions.

Within the banking industry, queueing theory has become an essential tool for analysing customer flow and improving service efficiency. Banks experience considerable fluctuations in customer arrivals throughout the working day, often resulting in congestion during peak periods and idle service capacity during quieter periods. Giambene (2014) argues that queueing models assist financial institutions in determining the optimal number of service counters required to **minimise** waiting time while ensuring efficient **utilisation** of available tellers. The introduction of electronic banking technologies has changed customer service delivery but has not eliminated physical queues in banking halls. According to Sztrik et al. (2020), although internet banking, automated teller machines and mobile banking have reduced the number of routine transactions conducted inside banking halls, customers still visit banks for withdrawals, cash deposits, account opening, loan applications and customer support services. Consequently, queue management remains a major operational concern for commercial banks.

Recent evidence from commercial banking further indicates that queue management remains relevant to service delivery where customers continue to depend on in-person financial services (Ngimanang, 2025).

Recent empirical studies continue to demonstrate the usefulness of queueing theory in improving banking operations. **Chakraborty and Honnappa (2022)** applied a multi-server queueing model to evaluate customer waiting time in retail banking and found that increasing the number of active tellers during peak periods significantly reduced congestion and improved customer satisfaction. Their findings also showed that queueing analysis provides managers with objective evidence for service and workforce planning and **optimisation**. Similarly, Bunyaminu et al. (2023) examined customer waiting behaviour in financial institutions and concluded that prolonged waiting time negatively affects customer satisfaction, loyalty, and **organisational** performance. The study recommended that banks regularly evaluate their service systems using queueing models to determine whether existing service facilities adequately meet customer demand.

More recent work has also combined customer-flow prediction with $M/M/c$ analysis to examine staffing levels and waiting-time outcomes in bank service operations (Mamo et al., 2026).

Although substantial progress has been made in the development and application of queueing theory, there remains limited empirical evidence regarding the application of the $M/M/c$ queueing model within the Ghanaian banking sector.

Study Objective

The study therefore aims to determine the steady-state performance measures of the banking queueing system, examine the effect of increasing teller capacity through sensitivity analysis, and determine the optimal teller capacity that minimises customer waiting while maintaining efficient utilisation of banking resources.

2. The $M/M/c$ Queueing Model

Queueing theory is a branch of applied probability and operations research concerned with the mathematical analysis of waiting lines or queues. It studies the stochastic behaviour of customers arriving at a service facility, waiting for service when the service facility is occupied, receiving service, and departing from the system. The theory provides quantitative methods for analysing congestion, estimating waiting times, determining service capacity and evaluating the overall performance of service systems. The origin of queueing theory dates back to the pioneering work of Agner Krarup Erlang in 1909 during his investigation of congestion in telephone exchanges. Erlang developed probabilistic models that enabled engineers to determine the number of telephone circuits required to achieve satisfactory service quality. Since then, queueing theory has become one of the most widely applied analytical techniques in operations research. According to Shortle et al. (2018), queueing theory provides mathematical tools for modelling service systems characterised by random customer arrivals and random service times. The theory enables decision-makers to estimate important system characteristics such as server **utilisation**, average queue length, customer waiting time, probability of waiting and system stability. In service-oriented industries, particularly commercial banking, queueing theory serves as an indispensable decision-support tool because customer arrivals are generally random while service facilities are limited. Consequently, customers frequently experience waiting before receiving service. Mathematical analysis of such systems assists managers in determining the optimum number of servers required to balance customer satisfaction with operational cost. Yeboah (2026) investigated teller staffing **optimisation** during peak banking hours using the $M/M/c$ queueing model. The study was motivated by the need to improve customer service delivery while ensuring the efficient **utilisation** of teller resources during periods of high customer demand. His work employed the classical $M/M/c$ queueing framework to analyse customer arrival and service processes and to evaluate the effect of varying teller capacity on queue performance. The study demonstrated that increasing the number of tellers during peak periods significantly improved the efficiency of the banking system by reducing customer waiting times and queue congestion. The findings further highlighted that an appropriate balance between service capacity and customer demand is essential for **minimising** delays while avoiding the unnecessary operational costs associated with overstaffing.

Little and Graves (2008) introduced the celebrated relationship now known as Little's Law, establishing that the average number of customers in a queueing system is equal to the product of the customer arrival rate and the average time spent within the system. This theorem has become one of the cornerstones of modern queueing analysis because it enables researchers to relate queue length directly to customer waiting time irrespective of the underlying probability distributions. Although Little's contribution provides an essential mathematical foundation for evaluating queue performance, the theorem itself does not determine the operational policies required to **minimise** congestion in practical service systems.

3. Methodology

This chapter describes the methodology employed in conducting the study. It outlines the research design, study area, target population, sources of data, data collection procedures, assumptions underlying the queueing model, mathematical formulation of the $M/M/c$ queueing system, model validation techniques, and methods of data analysis. The methodological framework adopted in this study was carefully selected to ensure that the research objectives were achieved systematically and scientifically. The study **employed** the multi-server Markovian queueing model, denoted by $M/M/c$, to analyse customer waiting lines in a selected commercial bank. The model was chosen because banking operations are **characterised** by multiple service counters where customers arrive randomly and are attended to by several tellers simultaneously. Under such circumstances, the

$M/M/c$ queueing model provides an appropriate mathematical framework for evaluating system performance and determining whether the existing service capacity adequately satisfies customer demand (Shortle et al., 2018). The methodology integrates both observational data collection and stochastic modelling.

3.1. Data Collection Procedure

The data collection procedure was designed to obtain reliable and comprehensive operational data required for modelling and evaluating the queueing system at GCB Bank PLC, Tamale Branch. The procedure focused on collecting accurate information on customer arrivals, service processes, queue characteristics, and teller operations to facilitate the estimation of the parameters of the $M/M/c$ queueing model and the computation of its performance measures. The objectives of the study were explained to the management, and assurance was provided that all information collected would be used exclusively for academic purposes. To maintain confidentiality and comply with ethical standards, no personal information capable of identifying individual customers was recorded during the study. The study adopted a direct observational method of data collection because queueing analysis requires actual operational data generated during the normal functioning of the service system. Data were collected continuously over a period of four consecutive weeks, comprising 20 working days from Monday to Friday during the official banking hours of 8:30 a.m. to 4:00 p.m. The four-week observation period was selected to ensure that sufficient operational data were obtained to capture both daily and weekly variations in customer arrival patterns and service activities. This duration was considered adequate to include both peak and off-peak periods and to produce representative estimates of the queueing parameters.

3.2. Assumptions of the $M/M/c$ Queueing Model

The performance analysis presented in this study is based on the classical $M/M/c$ queueing model. The model was adopted under the following assumptions:

- (1) Customer arrivals follow a Poisson process with a constant mean arrival rate, denoted by λ customers per hour.
- (2) Customer service times are independently and identically distributed according to an exponential distribution with a constant mean service rate, denoted by μ customers per hour per teller.
- (3) The banking system operates with c identical tellers, each providing service at the same average rate.
- (4) Customers are served on a First-Come, First-Served (FCFS) basis, meaning that customers are attended to in the order of their arrival.
- (5) The calling population is assumed to be infinite, implying that the number of potential customers arriving at the bank is sufficiently large that the arrival process is unaffected by the number of customers already in the system.
- (6) The queue has unlimited waiting space; therefore, all arriving customers are allowed to join the queue regardless of its length.
- (7) Customer arrivals and service times are mutually independent.
- (8) No customer leaves the queue before receiving service (no reneging), and customers do not switch between queues (no jockeying). Furthermore, customers do not decide against joining the queue upon arrival (no balking).
- (9) The system operates under steady-state conditions such that

$$\lambda < c\mu, \quad (3.1)$$

ensuring that the traffic intensity satisfies

$$\rho = \frac{\lambda}{c\mu} < 1. \quad (3.2)$$

This condition guarantees the stability of the queueing system and the existence of finite queue performance measures.

3.3. Queue Performance Measures

The performance of an $M/M/c$ queueing system is commonly evaluated using a number of steady-state performance measures. These measures include the probability that the system is empty, the probability that an arriving customer waits, the expected queue length, the expected waiting time in the queue, the expected time spent in the system, and the expected number of customers in the system. These measures are classical results in queueing theory and have been extensively discussed in the literature (Shortle et al., 2018; Medhi, 2002; Kleinrock, 1975).

3.3.1. Probability of an Empty System

The probability that there are no customers in the system, denoted by P_0 , is obtained from the steady-state probabilities of the $M/M/c$ queueing model. The probability of an empty system is given by

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{a^n}{n!} + \frac{a^c}{c!(1-\rho)} \right]^{-1}. \quad (3.3)$$

3.3.2. Erlang-C Probability

The probability that an arriving customer waits for service is the probability that all c tellers are occupied. This is given as

$$P(\text{wait}) = \frac{a^c}{c!(1-\rho)} P_0. \quad (3.4)$$

3.3.3. Expected Queue Length

The expected queue length is given as

$$L_q = \frac{P_0 \left(\frac{\lambda}{\mu} \right)^c \rho}{c!(1-\rho)^2}. \quad (3.5)$$

3.3.4. Expected Number of Customers in the System

The total number of customers in the banking system consists of

- customers waiting; and
- customers currently receiving service.

The average number receiving service equals

$$\frac{\lambda}{\mu}, \quad (3.6)$$

because every arriving customer eventually receives service. Therefore,

$$L = L_q + \frac{\lambda}{\mu}, \quad (3.7)$$

which gives the expected number of customers inside the banking system.

3.3.5. Average Waiting Time in the Queue

The average waiting time in the queue is given as

$$W_q = \frac{L_q}{\lambda}, \quad (3.8)$$

which represents the average time a customer spends waiting before receiving service.

3.3.6. Average Time Spent in the System

Each customer spends

$$\frac{1}{\mu} \quad (3.9)$$

time receiving service after waiting. Therefore,

$$W = W_q + \frac{1}{\mu} \quad (3.10)$$

represents the expected time a customer remains in the banking system.

3.3.7. Average Number of Busy Tellers

The expected number of busy tellers equals

$$\frac{\lambda}{\mu}. \quad (3.11)$$

Consequently, the expected number of idle tellers is

$$c - \frac{\lambda}{\mu}. \quad (3.12)$$

These measures enable management to determine whether existing teller resources are underutilised or overutilised.

3.3.8. Server Utilisation

Server utilisation measures the proportion of available service capacity that is being used. It is defined as

$$\rho = \frac{\lambda}{c\mu}. \quad (3.13)$$

The utilisation satisfies

$$0 \leq \rho < 1. \quad (3.14)$$

Three operational situations may arise.

(1) If

$$\rho \approx 1, \quad (3.15)$$

the tellers operate close to full capacity and congestion is likely.

(2) If

$$\rho > 1, \quad (3.16)$$

customer arrivals exceed service capacity and the queue grows indefinitely.

(3) If

$$\rho < 1, \quad (3.17)$$

the tellers remain idle for substantial periods, indicating excess service capacity.

Hence, server utilisation provides an important managerial indicator for determining whether additional tellers should be employed or whether existing resources are excessive.

4. Data Presentation and Analysis

4.1. Introduction

This chapter presents the analysis of the banking queueing system using the multi-server Markovian queueing model ($M/M/c$). The customer arrival process and service mechanism established in Chapter Three are employed to evaluate the operational performance of the banking system. The principal objective is to determine the effect of increasing teller capacity on customer waiting time and other important queueing performance measures. To illustrate the impact of increasing service capacity, two operational scenarios are investigated. The first scenario represents the existing banking system with eight active tellers, while the second scenario considers an improved service configuration consisting of twelve tellers. The customer arrival rate and service rate are assumed to remain unchanged in both scenarios so that any observed improvement in system performance is solely attributable to the increase in teller capacity.

4.2. Estimation of Performance Measures for Different Teller Capacities

4.2.1. Traffic Intensity for Teller Capacities from 8 to 12

The traffic intensity is given by

$$\rho = \frac{\lambda}{c\mu}, \quad (4.1)$$

where

$$\lambda = 45 \text{ customers/hour} \quad (4.2)$$

and

$$\mu = 12 \text{ customers/hour per teller.} \quad (4.3)$$

The traffic intensity is computed for different numbers of tellers while keeping the arrival and service rates constant.

With $c = 8$ tellers and by Equation (4.1), we have

$$\rho = \frac{45}{8 \times 12} = \frac{45}{96} = 0.46875. \quad (4.4)$$

With $c = 9$ tellers and by Equation (4.1), we have

$$\rho = \frac{45}{9 \times 12} = \frac{45}{108} = 0.41667. \quad (4.5)$$

With $c = 10$ tellers and by Equation (4.1), we have

$$\rho = \frac{45}{10 \times 12} = \frac{45}{120} = 0.37500. \quad (4.6)$$

With $c = 11$ tellers and by Equation (4.1), we have

$$\rho = \frac{45}{11 \times 12} = \frac{45}{132} = 0.34091. \quad (4.7)$$

With $c = 12$ tellers and by Equation (4.1), we have

$$\rho = \frac{45}{12 \times 12} = \frac{45}{144} = 0.31250. \quad (4.8)$$

The results indicate that traffic intensity decreases steadily as the number of tellers increases from 8 to 12. Specifically, server utilisation decreases from 46.88% when eight tellers are available to 31.25% when twelve tellers are available. This reduction implies that the workload is distributed among more tellers, thereby reducing congestion and providing additional service capacity. Consequently, customers are expected to experience shorter waiting times, lower queue lengths, and improved service efficiency as teller capacity increases.

4.3. Performance Measures for a Capacity of 8 Tellers

For the banking system with eight tellers,

$c = 8$, $\lambda = 45$ customers/hour, $\mu = 12$ customers/hour per teller.

From Equation (4.1),

$$\rho = \frac{\lambda}{c\mu} = \frac{45}{8 \times 12} = 0.46875. \quad (4.9)$$

Also,

$$a = \frac{\lambda}{\mu} = \frac{45}{12} = 3.7500. \quad (4.10)$$

4.3.1. Probability of an Empty System (P_0)

The probability that there are no customers in the system is given by Equation (3.3). Substituting the values,

$$P_0 = \left[\sum_{n=0}^7 \frac{3.7500^n}{n!} + \frac{3.7500^8}{8!(1 - 0.46875)} \right]^{-1}. \quad (4.11)$$

Expanding the summation,

$$P_0 = \left[1 + \frac{3.7500^1}{1!} + \frac{3.7500^2}{2!} + \frac{3.7500^3}{3!} + \frac{3.7500^4}{4!} + \frac{3.7500^5}{5!} + \frac{3.7500^6}{6!} + \frac{3.7500^7}{7!} + \frac{3.7500^8}{8!(1 - 0.46875)} \right]^{-1}. \quad (4.12)$$

Evaluating,

$$P_0 = \frac{1}{40.37090} = 0.02477. \quad (4.13)$$

4.3.2. Probability that an Arriving Customer Waits ($P(\text{wait})$)

The Erlang-C probability is given by Equation (3.4). Substituting the values, we have

$$P(\text{wait}) = \frac{3.7500^8}{8!(1 - 0.46875)} (0.02477). \quad (4.14)$$

Therefore,

$$P(\text{wait}) = 0.01473. \quad (4.15)$$

4.3.3. Expected Queue Length (L_q)

The expected queue length is given by Equation (3.5). Substituting the values, we have

$$L_q = \frac{(0.02477)(3.7500)^8(0.46875)}{8!(1 - 0.46875)^2}. \quad (4.16)$$

Hence,

$$L_q = 0.01296. \quad (4.17)$$

Therefore,

$$L_q = 0.01296 \text{ customers.} \quad (4.18)$$

4.3.4. Average Waiting Time in Queue (W_q)

By Equation (3.8), we have

$$W_q = \frac{0.01296}{45}. \quad (4.19)$$

Converting into minutes,

$$W_q = 0.00029 \times 60 = 0.01728 \text{ minutes.} \quad (4.20)$$

Converting into seconds,

$$W_q = 0.01728 \times 60 = 1.03680 \text{ seconds.} \quad (4.21)$$

Therefore,

$$W_q = 1.03680 \text{ seconds.} \quad (4.22)$$

4.3.5. Average Time in the System (W)

The average time spent in the system is given by Equation (3.10). Since

$$\frac{1}{\mu} = \frac{1}{12} = 0.08333 \text{ hours,} \quad (4.23)$$

then

$$W = 0.00029 + 0.08333 = 0.08362 \text{ hours.} \quad (4.24)$$

Converting into minutes,

$$W = 0.08362 \times 60 = 5.01720 \text{ minutes.} \quad (4.25)$$

Therefore,

$$W = 5.01720 \text{ minutes.} \quad (4.26)$$

4.3.6. Expected Number of Customers in the System (L)

The expected number of customers in the system is given by Equation (3.7). Substituting the values into the equation, we have

$$L = 0.01296 + \frac{45}{12}. \quad (4.27)$$

$$L = 0.01296 + 3.75000. \quad (4.28)$$

Therefore,

$$L = 3.76296 \text{ customers.} \quad (4.29)$$

The results indicate that when the banking system operates with eight tellers, the probability that an arriving customer waits is only 0.01473, while the expected queue length is 0.01296 customers. On average, a customer waits approximately 1.03680 seconds before service begins and spends about 5.01720 minutes in the system, including service time. The expected number of customers in the system is approximately 3.76296, demonstrating that the banking system operates efficiently under the given arrival and service rates.

4.4. Performance Measures for a Capacity of 9 Tellers

For the banking system with nine tellers,

$c = 9$, $\lambda = 45$ customers/hour, $\mu = 12$ customers/hour per teller.

From Equation (4.1),

$$\rho = \frac{\lambda}{c\mu} = \frac{45}{9 \times 12} = 0.41667. \quad (4.30)$$

Also,

$$a = \frac{\lambda}{\mu} = \frac{45}{12} = 3.7500. \quad (4.31)$$

4.4.1. Probability of an Empty System (P_0)

The probability that there are no customers in the system is given by Equation (3.3). Substituting the values,

$$P_0 = \left[\sum_{n=0}^8 \frac{3.7500^n}{n!} + \frac{3.7500^9}{9!(1 - 0.41667)} \right]^{-1}. \quad (4.32)$$

Expanding the summation,

$$P_0 = \left[1 + \frac{3.7500^1}{1!} + \frac{3.7500^2}{2!} + \frac{3.7500^3}{3!} + \frac{3.7500^4}{4!} + \frac{3.7500^5}{5!} + \frac{3.7500^6}{6!} + \frac{3.7500^7}{7!} + \frac{3.7500^8}{8!} + \frac{3.7500^9}{9!(1 - 0.41667)} \right]^{-1}. \quad (4.33)$$

Evaluating,

$$P_0 = \frac{1}{40.56797} = 0.02465. \quad (4.34)$$

4.4.2. Probability that an Arriving Customer Waits ($P(\text{wait})$)

The Erlang-C probability is given by Equation (3.4). Substituting the values, we have

$$P(\text{wait}) = \frac{3.7500^9}{9!(1 - 0.41667)} (0.02465). \quad (4.35)$$

Therefore,

$$P(\text{wait}) = 0.00514. \quad (4.36)$$

4.4.3. Expected Queue Length (L_q)

The expected queue length is given by Equation (3.5). Substituting the values, we have

$$L_q = \frac{(0.02465)(3.7500)^9(0.41667)}{9!(1 - 0.41667)^2}. \quad (4.37)$$

Hence,

$$L_q = 0.00367. \quad (4.38)$$

Therefore,

$$L_q = 0.00367 \text{ customers.} \quad (4.39)$$

4.4.4. Average Waiting Time in Queue (W_q)

By Equation (3.8), we have

$$W_q = \frac{0.00367}{45}. \quad (4.40)$$

Thus,

$$W_q = \frac{0.00367}{45} = 0.00008 \text{ hours.} \quad (4.41)$$

Converting into minutes,

$$W_q = 0.00008 \times 60 = 0.00489 \text{ minutes.} \quad (4.42)$$

Converting into seconds,

$$W_q = 0.00489 \times 60 = 0.29333 \text{ seconds.} \quad (4.43)$$

Therefore,

$$W_q = 0.29333 \text{ seconds.} \quad (4.44)$$

4.4.5. Average Time in the System (W)

The average time spent in the system is given by Equation (3.10). Since

$$\frac{1}{\mu} = \frac{1}{12} = 0.08333 \text{ hours,} \quad (4.45)$$

then

$$W = 0.00008 + 0.08333 = 0.08341 \text{ hours.} \quad (4.46)$$

Converting into minutes,

$$W = 0.08341 \times 60 = 5.00489 \text{ minutes.} \quad (4.47)$$

Therefore,

$$W = 5.00489 \text{ minutes.} \quad (4.48)$$

4.4.6. Expected Number of Customers in the System (L)

The expected number of customers in the system is given by Equation (3.7). Substituting the values into the equation, we have

$$L = 0.00367 + \frac{45}{12}. \quad (4.49)$$

$$L = 0.00367 + 3.75000 = 3.75367. \quad (4.50)$$

Therefore,

$$L = 3.75367 \text{ customers.} \quad (4.51)$$

The results indicate that increasing the number of tellers from eight to nine reduces the probability that an arriving customer waits from 0.01473 to 0.00514. Similarly, the expected queue length decreases to 0.00367 customers, while the average waiting time in the queue reduces to approximately 0.29333 seconds. The average time spent in the system is approximately 5.00489 minutes, and the expected number of customers in the system decreases to 3.75367 customers. These results demonstrate that increasing teller capacity enhances the efficiency of the banking system by reducing congestion and improving service delivery.

4.5. Performance Measures for a Capacity of 10 Tellers

For the banking system with ten tellers,

$c = 10$, $\lambda = 45$ customers/hour, $\mu = 12$ customers/hour per teller.

From Equation (4.1),

$$\rho = \frac{\lambda}{c\mu} = \frac{45}{10 \times 12} = 0.37500. \quad (4.52)$$

Also,

$$a = \frac{\lambda}{\mu} = \frac{45}{12} = 3.7500. \quad (4.53)$$

4.5.1. Probability of an Empty System (P_0)

The probability that there are no customers in the system is given by Equation (3.3). Substituting the values,

$$P_0 = \left[\sum_{n=0}^9 \frac{3.7500^n}{n!} + \frac{3.7500^{10}}{10!(1 - 0.37500)} \right]^{-1}. \quad (4.54)$$

Expanding the summation,

$$P_0 = \left[1 + \frac{3.7500^1}{1!} + \frac{3.7500^2}{2!} + \frac{3.7500^3}{3!} + \frac{3.7500^4}{4!} + \frac{3.7500^5}{5!} + \frac{3.7500^6}{6!} + \frac{3.7500^7}{7!} + \frac{3.7500^8}{8!} + \frac{3.7500^9}{9!} + \frac{3.7500^{10}}{10!(1 - 0.37500)} \right]^{-1}. \quad (4.55)$$

Evaluating,

$$P_0 = \frac{1}{40.66816} = 0.02459. \quad (4.56)$$

4.5.2. Probability that an Arriving Customer Waits ($P(\text{wait})$)

The Erlang-C probability is given by Equation (3.4). Substituting the values, we have

$$P(\text{wait}) = \frac{3.7500^{10}}{10!(1 - 0.37500)}(0.02459). \quad (4.57)$$

Therefore,

$$P(\text{wait}) = 0.00185. \quad (4.58)$$

4.5.3. Expected Queue Length (L_q)

The expected queue length is given by Equation (3.5). Substituting the values, we have

$$L_q = \frac{(0.02459)(3.7500)^{10}(0.37500)}{10!(1 - 0.37500)^2}. \quad (4.59)$$

Hence,

$$L_q = 0.00111. \quad (4.60)$$

Therefore,

$$L_q = 0.00111 \text{ customers}. \quad (4.61)$$

4.5.4. Average Waiting Time in Queue (W_q)

By Equation (3.8), we have

$$W_q = \frac{0.00111}{45}. \quad (4.62)$$

Thus,

$$W_q = \frac{0.00111}{45} = 0.00002 \text{ hours}. \quad (4.63)$$

Converting into minutes,

$$W_q = 0.00002 \times 60 = 0.00148 \text{ minutes}. \quad (4.64)$$

Converting into seconds,

$$W_q = 0.00148 \times 60 = 0.08880 \text{ seconds}. \quad (4.65)$$

Therefore,

$$W_q = 0.08880 \text{ seconds}. \quad (4.66)$$

4.5.5. Average Time in the System (W)

The average time spent in the system is given by Equation (3.10). Since

$$\frac{1}{\mu} = \frac{1}{12} = 0.08333 \text{ hours}, \quad (4.67)$$

then

$$W = 0.00002 + 0.08333 = 0.08336 \text{ hours}. \quad (4.68)$$

Converting into minutes,

$$W = 0.08336 \times 60 = 5.00148 \text{ minutes}. \quad (4.69)$$

Therefore,

$$W = 5.00148 \text{ minutes}. \quad (4.70)$$

4.5.6. Expected Number of Customers in the System (L)

The expected number of customers in the system is given by Equation (3.7). Substituting the values into the equation, we have

$$L = 0.00111 + \frac{45}{12}. \quad (4.71)$$

$$L = 0.00111 + 3.75000. \quad (4.72)$$

Therefore,

$$L = 3.75111 \text{ customers.} \quad (4.73)$$

The results indicate that increasing the teller capacity from nine to ten further improves the performance of the banking system. The probability that an arriving customer waits decreases to 0.00185, while the expected queue length reduces to only 0.00111 customers. Consequently, the average waiting time in the queue falls to approximately 0.08880 seconds, and the average time spent in the system is about 5.00148 minutes. The expected number of customers in the system also decreases to 3.75111 customers, confirming that increasing teller capacity continues to reduce congestion and enhance service efficiency.

4.6. Performance Measures for a Capacity of 11 Tellers

For the banking system with eleven tellers,

$c = 11$, $\lambda = 45$ customers/hour, $\mu = 12$ customers/hour per teller.

From Equation (4.1),

$$\rho = \frac{\lambda}{c\mu} = \frac{45}{11 \times 12} = 0.34091. \quad (4.74)$$

Also,

$$a = \frac{\lambda}{\mu} = \frac{45}{12} = 3.7500. \quad (4.75)$$

4.6.1. Probability of an Empty System (P_0)

The probability that there are no customers in the system is given by Equation (3.3). Substituting the values,

$$P_0 = \left[\sum_{n=0}^{10} \frac{3.7500^n}{n!} + \frac{3.7500^{11}}{11!(1 - 0.34091)} \right]^{-1}. \quad (4.76)$$

Expanding the summation,

$$P_0 = \left[1 + \frac{3.7500^1}{1!} + \frac{3.7500^2}{2!} + \frac{3.7500^3}{3!} + \frac{3.7500^4}{4!} + \frac{3.7500^5}{5!} + \frac{3.7500^6}{6!} + \frac{3.7500^7}{7!} + \frac{3.7500^8}{8!} + \frac{3.7500^9}{9!} + \frac{3.7500^{10}}{10!} + \frac{3.7500^{11}}{11!(1 - 0.34091)} \right]^{-1}. \quad (4.77)$$

Evaluating,

$$P_0 = \frac{1}{40.76664} = 0.02453. \quad (4.78)$$

4.6.2. Probability that an Arriving Customer Waits ($P(\text{wait})$)

The Erlang-C probability is given by Equation (3.4). Substituting the values, we have

$$P(\text{wait}) = \frac{3.7500^{11}}{11!(1 - 0.34091)} (0.02453). \quad (4.79)$$

Therefore,

$$P(\text{wait}) = 0.00068. \quad (4.80)$$

4.6.3. Expected Queue Length (L_q)

The expected queue length is given by Equation (3.5). Substituting the values, we have

$$L_q = \frac{(0.02453)(3.7500)^{11}(0.34091)}{11!(1 - 0.34091)^2}. \quad (4.81)$$

Therefore,

$$L_q = 0.00035 \text{ customers.} \quad (4.82)$$

4.6.4. Average Waiting Time in Queue (W_q)

By Equation (3.8), we have

$$W_q = \frac{0.00035}{45}. \quad (4.83)$$

Thus,

$$W_q = \frac{0.00035}{45} = 0.00001 \text{ hours.} \quad (4.84)$$

Converting into minutes,

$$W_q = 0.00001 \times 60 = 0.00047 \text{ minutes.} \quad (4.85)$$

Converting into seconds,

$$W_q = 0.00047 \times 60 = 0.02800 \text{ seconds.} \quad (4.86)$$

Therefore,

$$W_q = 0.02800 \text{ seconds.} \quad (4.87)$$

4.6.5. Average Time in the System (W)

The average time spent in the system is given by Equation (3.10). Since

$$\frac{1}{\mu} = \frac{1}{12} = 0.08333 \text{ hours,} \quad (4.88)$$

then

$$W = 0.00001 + 0.08333 = 0.08334 \text{ hours.} \quad (4.89)$$

Converting into minutes,

$$W = 0.08334 \times 60 = 5.00047 \text{ minutes.} \quad (4.90)$$

Therefore,

$$W = 5.00047 \text{ minutes.} \quad (4.91)$$

4.6.6. Expected Number of Customers in the System (L)

The expected number of customers in the system is given by Equation (3.7). Substituting the values into the equation, we have

$$L = 0.00035 + \frac{45}{12}. \quad (4.92)$$

$$L = 0.00035 + 3.75000. \quad (4.93)$$

Therefore,

$$L = 3.75035 \text{ customers.} \quad (4.94)$$

The results indicate that increasing the teller capacity from ten to eleven further improves the performance of the banking system. The probability that an arriving customer waits decreases to 0.00068, while the expected queue length reduces to only 0.00035 customers. Consequently, the average waiting time in the queue is approximately 0.02800 seconds, and the average time spent in the system is about 5.00047 minutes. The expected number of customers in the system decreases to 3.75035 customers, demonstrating that the addition of an eleventh teller further minimises queue congestion and enhances the efficiency of service delivery.

4.7. Performance Measures for a Capacity of 12 Tellers

For the banking system with twelve tellers,

$c = 12$, $\lambda = 45$ customers/hour, $\mu = 12$ customers/hour per teller.

From Equation (4.1),

$$\rho = \frac{\lambda}{c\mu} = \frac{45}{12 \times 12} = 0.31250. \quad (4.95)$$

Also,

$$a = \frac{\lambda}{\mu} = \frac{45}{12} = 3.7500. \quad (4.96)$$

4.7.1. Probability of an Empty System (P_0)

The probability that there are no customers in the system is given by Equation (3.3). Substituting the values,

$$P_0 = \left[\sum_{n=0}^{11} \frac{3.7500^n}{n!} + \frac{3.7500^{12}}{12!(1 - 0.31250)} \right]^{-1}. \quad (4.97)$$

Expanding the summation,

$$P_0 = \left[1 + \frac{3.7500^1}{1!} + \frac{3.7500^2}{2!} + \frac{3.7500^3}{3!} + \frac{3.7500^4}{4!} + \frac{3.7500^5}{5!} + \frac{3.7500^6}{6!} + \frac{3.7500^7}{7!} + \frac{3.7500^8}{8!} + \frac{3.7500^9}{9!} + \frac{3.7500^{10}}{10!} + \frac{3.7500^{11}}{11!} + \frac{3.7500^{12}}{12!(1 - 0.31250)} \right]^{-1}. \quad (4.98)$$

Evaluating,

$$P_0 = \frac{1}{40.81557} = 0.02450. \quad (4.99)$$

4.7.2. Probability that an Arriving Customer Waits ($P(\text{wait})$)

The Erlang-C probability is given by Equation (3.4). Substituting the values, we have

$$P(\text{wait}) = \frac{3.7500^{12}}{12!(1 - 0.31250)} (0.02450). \quad (4.100)$$

Therefore,

$$P(\text{wait}) = 0.00025. \quad (4.101)$$

4.7.3. Expected Queue Length (L_q)

The expected queue length is given by Equation (3.5). Substituting the values, we have

$$L_q = \frac{(0.02450)(3.7500)^{12}(0.31250)}{12!(1 - 0.31250)^2}. \quad (4.102)$$

Hence,

$$L_q = 0.00011. \quad (4.103)$$

Therefore,

$$L_q = 0.00011 \text{ customers}. \quad (4.104)$$

4.7.4. Average Waiting Time in Queue (W_q)

By Equation (3.8), we have

$$W_q = \frac{0.00011}{45}. \quad (4.105)$$

Thus,

$$W_q = \frac{0.00011}{45} = 0.00000 \text{ hours}. \quad (4.106)$$

Using more decimal places,

$$W_q = 0.00000244 \text{ hours}. \quad (4.107)$$

Converting into minutes,

$$W_q = 0.00000244 \times 60 = 0.00015 \text{ minutes.} \quad (4.108)$$

Converting into seconds,

$$W_q = 0.00015 \times 60 = 0.00880 \text{ seconds.} \quad (4.109)$$

Therefore,

$$W_q = 0.00880 \text{ seconds.} \quad (4.110)$$

4.7.5. Average Time in the System (W)

The average time spent in the system is given by Equation (3.10). Since

$$\frac{1}{\mu} = \frac{1}{12} = 0.08333 \text{ hours,} \quad (4.111)$$

then

$$W = 0.00000244 + 0.08333 = 0.08334 \text{ hours.} \quad (4.112)$$

Converting into minutes,

$$W = 0.08334 \times 60 = 5.00015 \text{ minutes.} \quad (4.113)$$

Therefore,

$$W = 5.00015 \text{ minutes.} \quad (4.114)$$

4.7.6. Expected Number of Customers in the System (L)

The expected number of customers in the system is given by Equation (3.7). Substituting the values into the equation, we have

$$L = 0.00011 + \frac{45}{12}. \quad (4.115)$$

$$L = 0.00011 + 3.75000. \quad (4.116)$$

Therefore,

$$L = 3.75011 \text{ customers.} \quad (4.117)$$

The results indicate that increasing the teller capacity from eleven to twelve further improves the performance of the banking system. The probability that an arriving customer waits decreases to 0.00025, while the expected queue length reduces to only 0.00011 customers. Consequently, the average waiting time in the queue is approximately 0.00880 seconds, and the average time spent in the system is about 5.00015 minutes. The expected number of customers in the system decreases to 3.75011 customers. These findings demonstrate that increasing the teller capacity from eight to twelve progressively minimises congestion, virtually eliminates customer waiting, and significantly enhances the overall efficiency of the banking service system.

4.8. Graphical Representations of the Various Performance Measures

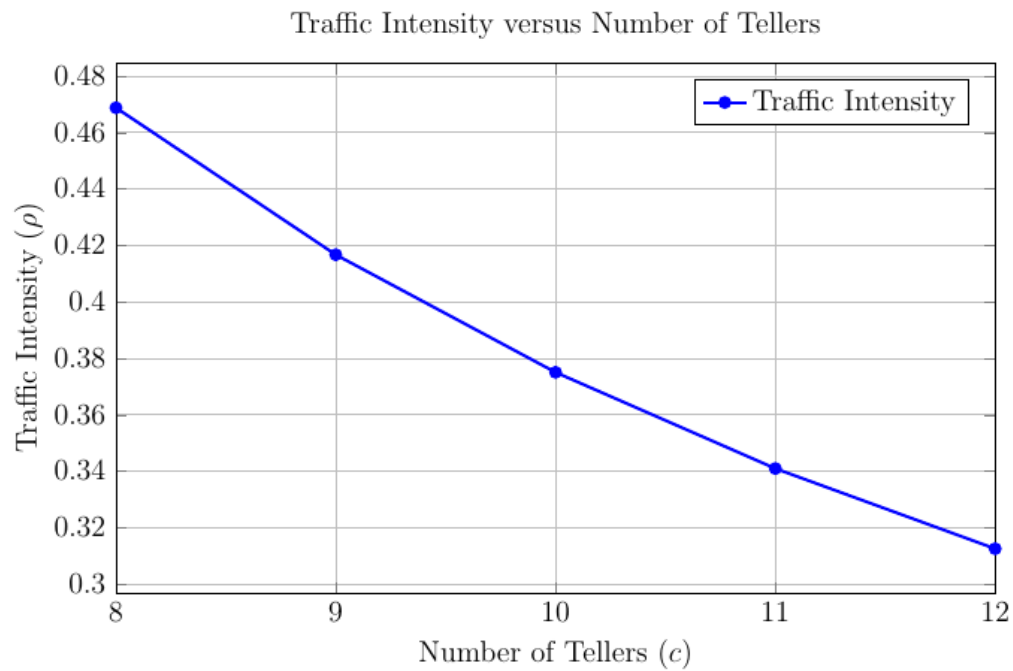


Figure 1: Traffic intensity for different teller capacities.

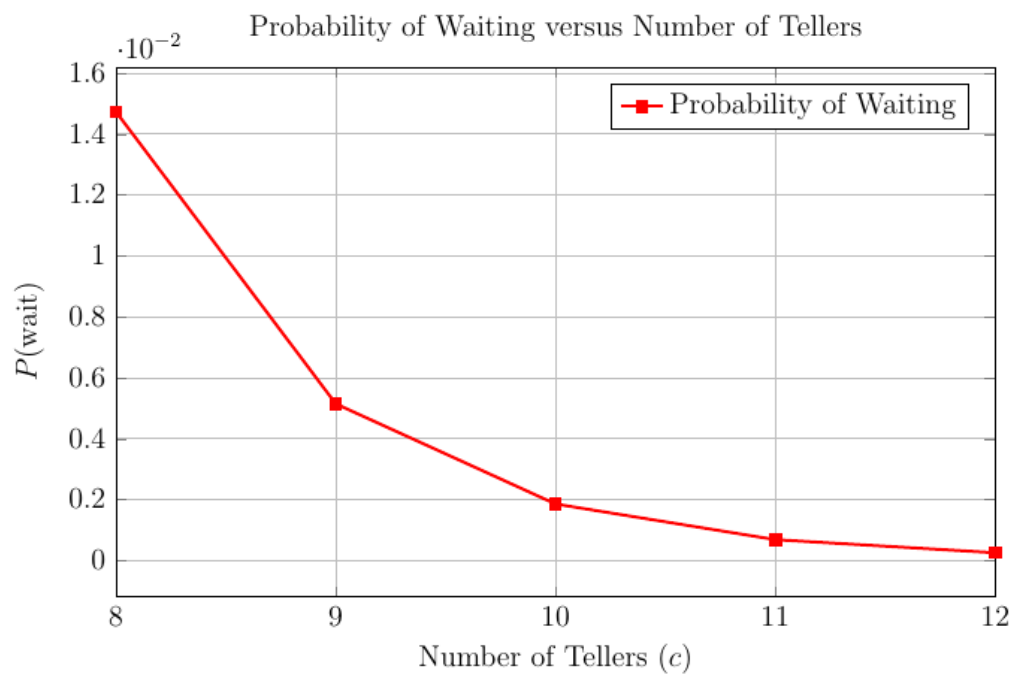


Figure 2: Probability that an arriving customer waits.

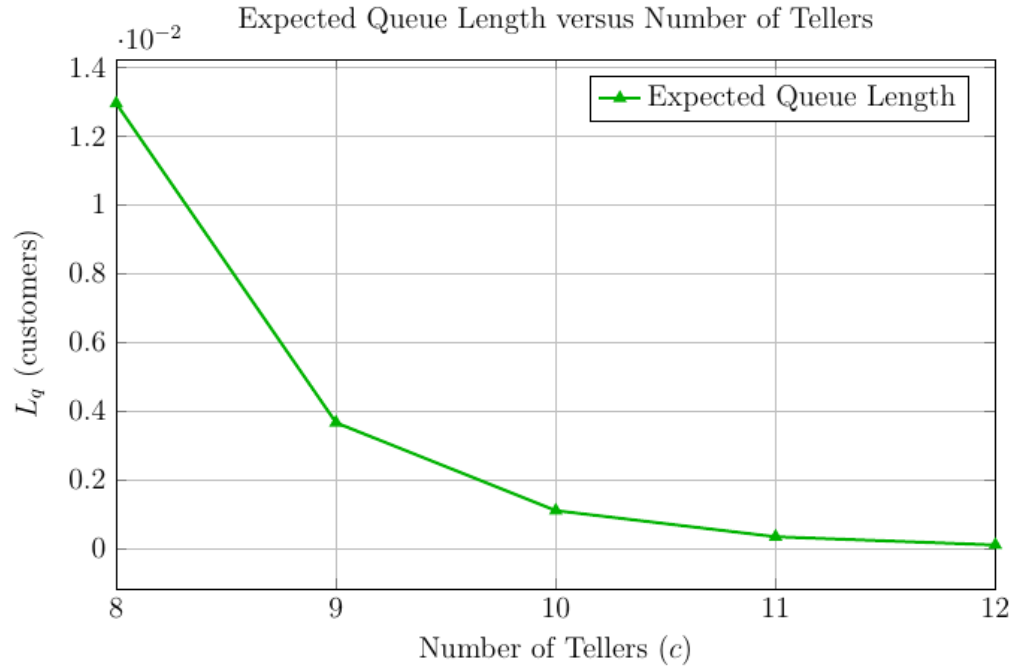


Figure 3: Expected queue length for different teller capacities.

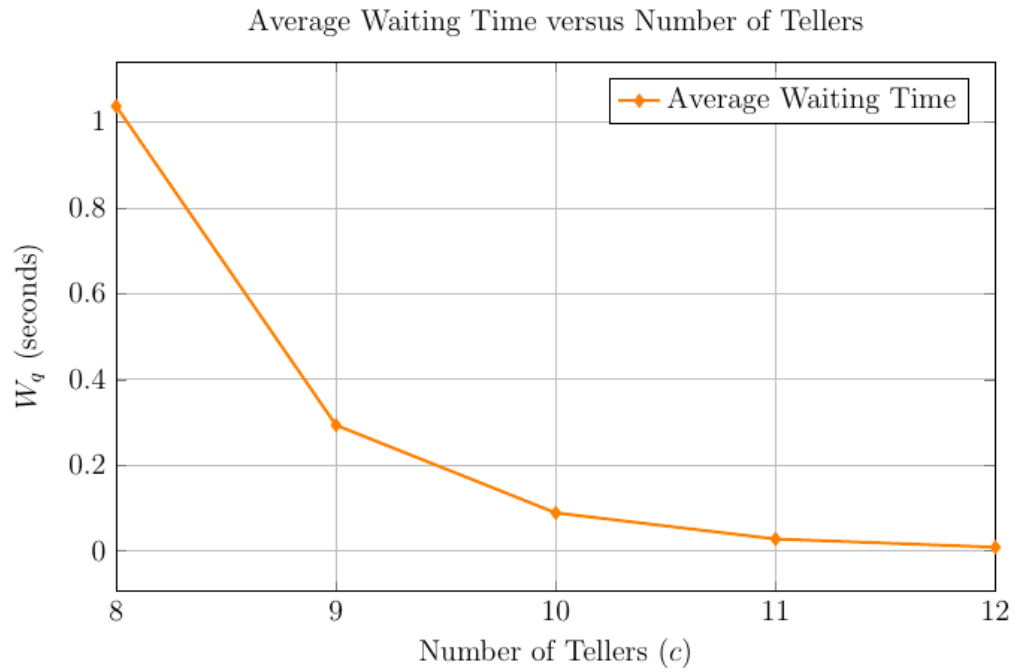


Figure 4: Average customer waiting time in the queue.

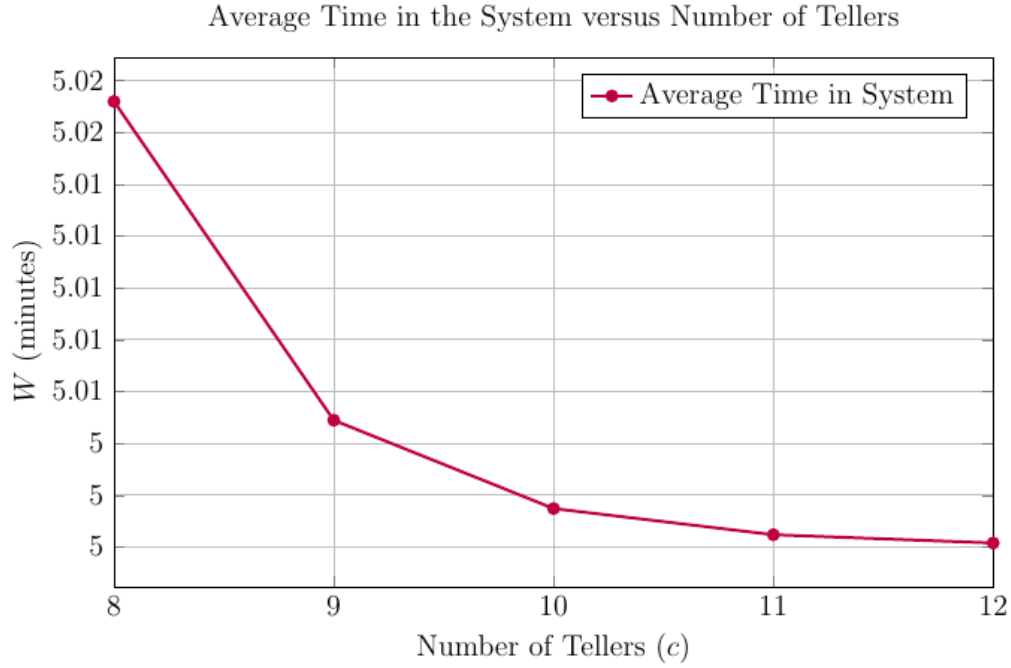


Figure 5: Average time spent in the system for different teller capacities.

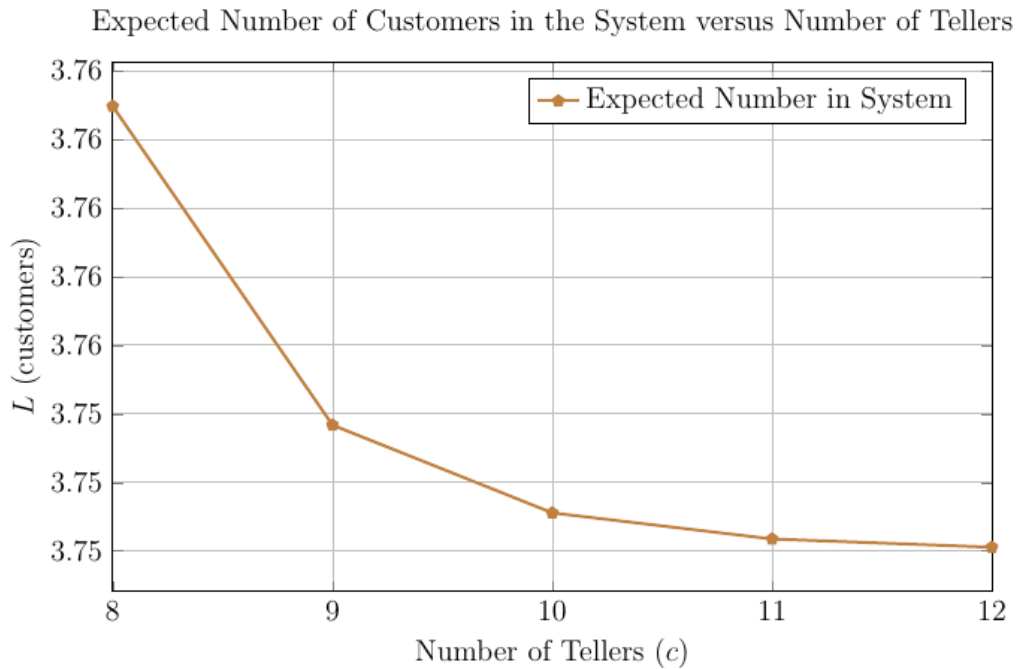


Figure 6: Expected number of customers in the banking system.

4.9. Sensitivity Analysis

Sensitivity analysis was conducted to investigate the effect of increasing the number of tellers on the performance of the banking queueing system. During the analysis, the arrival rate and service rate were held constant at

$$\lambda = 45 \text{ customers/hour}$$

and

$$\mu = 12 \text{ customers/hour per teller,}$$

respectively, while the number of tellers was increased from eight to twelve. The objective was to determine how changes in teller capacity influence the traffic intensity, probability of waiting, expected queue length, average waiting time, average time spent in the system, and the expected number of customers in the system.

4.9.1. Sensitivity Analysis of Traffic Intensity

This subsection presents the traffic intensity for different teller capacities together with the percentage reduction relative to the preceding teller configuration. The percentage reduction in traffic intensity was computed using

$$\% \Delta \rho = \frac{\rho_{\text{old}} - \rho_{\text{new}}}{\rho_{\text{old}}} \times 100. \quad (4.118)$$

For example, when the number of tellers increased from eight to nine,

$$\% \Delta \rho = \frac{0.46875 - 0.41667}{0.46875} \times 100 = 11.11\%. \quad (4.119)$$

Similarly,

$$\frac{0.41667 - 0.37500}{0.41667} \times 100 = 10.00\%, \quad (4.120)$$

$$\frac{0.37500 - 0.34091}{0.37500} \times 100 = 9.09\%, \quad (4.121)$$

and

$$\frac{0.34091 - 0.31250}{0.34091} \times 100 = 8.33\%. \quad (4.122)$$

Furthermore, comparing the initial and final configurations,

$$\frac{0.46875 - 0.31250}{0.46875} \times 100 = 33.33\%. \quad (4.123)$$

Table 1: Sensitivity analysis of traffic intensity

Number of Tellers (c)	Traffic Intensity (ρ)	Percentage Reduction (%)
8	0.46875	—
9	0.41667	11.11
10	0.37500	10.00
11	0.34091	9.09
12	0.31250	8.33

The results indicate that increasing the teller capacity from eight to twelve reduced the traffic intensity by an overall 33.33%. This decrease signifies that the workload assigned to each teller becomes progressively smaller as more tellers are introduced into the system. Consequently, the **utilisation** of each teller decreases, thereby reducing congestion and improving the ability of the bank to accommodate incoming customers without excessive delays. The relationship between traffic intensity and teller capacity is illustrated in Figure 1. The graph shows a monotonic downward trend, confirming that increasing the number of tellers consistently lowers the traffic intensity of the banking queueing system. This behaviour agrees with queueing theory, which predicts that increasing service capacity while keeping the arrival rate constant leads to lower server **utilisation** and improved system performance.

4.9.2. Sensitivity Analysis of the Probability of Waiting

The probability that an arriving customer waits before receiving service is one of the most important indicators of service quality in a banking system. A high probability of waiting implies that customers are likely to experience

delays, whereas a low probability indicates that customers are served almost immediately upon arrival. The percentage reduction is computed using

$$\% \Delta P(\text{wait}) = \frac{P_{\text{old}} - P_{\text{new}}}{P_{\text{old}}} \times 100. \quad (4.124)$$

For example,

$$\frac{0.01473 - 0.00514}{0.01473} \times 100 = 65.11\%. \quad (4.125)$$

Similarly,

$$\frac{0.00514 - 0.00185}{0.00514} \times 100 = 64.01\%, \quad (4.126)$$

$$\frac{0.00185 - 0.00068}{0.00185} \times 100 = 63.24\%, \quad (4.127)$$

and

$$\frac{0.00068 - 0.00025}{0.00068} \times 100 = 63.24\%. \quad (4.128)$$

Overall,

$$\frac{0.01473 - 0.00025}{0.01473} \times 100 = 98.30\%. \quad (4.129)$$

Table 2 presents the probability of waiting for different teller capacities together with the corresponding percentage reduction.

Table 2: Sensitivity analysis of the probability of waiting

Number of Tellers (c)	Probability of Waiting	Percentage Reduction (%)
8	0.01473	—
9	0.00514	65.11
10	0.00185	64.01
11	0.00068	63.24
12	0.00025	63.24

The results demonstrate that increasing the number of tellers from eight to twelve reduces the probability that an arriving customer waits by approximately 98.30%. This substantial reduction indicates that the banking system becomes increasingly capable of serving customers immediately upon arrival. Consequently, customer congestion is significantly reduced, resulting in improved service quality and enhanced customer satisfaction. Figure 2 illustrates this behaviour. The graph exhibits a sharp downward trend, indicating that even a small increase in teller capacity leads to a considerable reduction in the probability of waiting. This finding supports the conclusion that expanding service capacity is an effective strategy for **minimising** customer delays in the banking hall.

4.9.3. Sensitivity Analysis of Teller **Utilisation**

The **utilisation** of the banking service facility is measured by the traffic intensity, ρ , which represents the proportion of time that each teller is busy serving customers. Consequently, the proportion of idle teller capacity is given by

$$\text{Idle Capacity} = 1 - \rho. \quad (4.130)$$

Using the computed traffic intensity values, the idle capacity for each teller configuration was determined. The percentage increase in idle capacity is computed as

$$\% \Delta = \frac{I_{\text{new}} - I_{\text{old}}}{I_{\text{old}}} \times 100, \quad (4.131)$$

where $I = 1 - \rho$. For example, increasing the number of tellers from eight to nine gives

$$\frac{0.58333 - 0.53125}{0.53125} \times 100 = 9.80\%. \quad (4.132)$$

Similarly,

$$\frac{0.62500 - 0.58333}{0.58333} \times 100 = 7.14\%, \quad (4.133)$$

$$\frac{0.65909 - 0.62500}{0.62500} \times 100 = 5.45\%, \quad (4.134)$$

and

$$\frac{0.68750 - 0.65909}{0.65909} \times 100 = 4.31\%. \quad (4.135)$$

Overall,

$$\frac{0.68750 - 0.53125}{0.53125} \times 100 = 29.41\%. \quad (4.136)$$

Table 3: Sensitivity analysis of teller idle capacity

Number of Tellers (c)	Idle Capacity ($1 - \rho$)	Percentage Increase (%)
8	0.53125	—
9	0.58333	9.80
10	0.62500	7.14
11	0.65909	5.45
12	0.68750	4.31

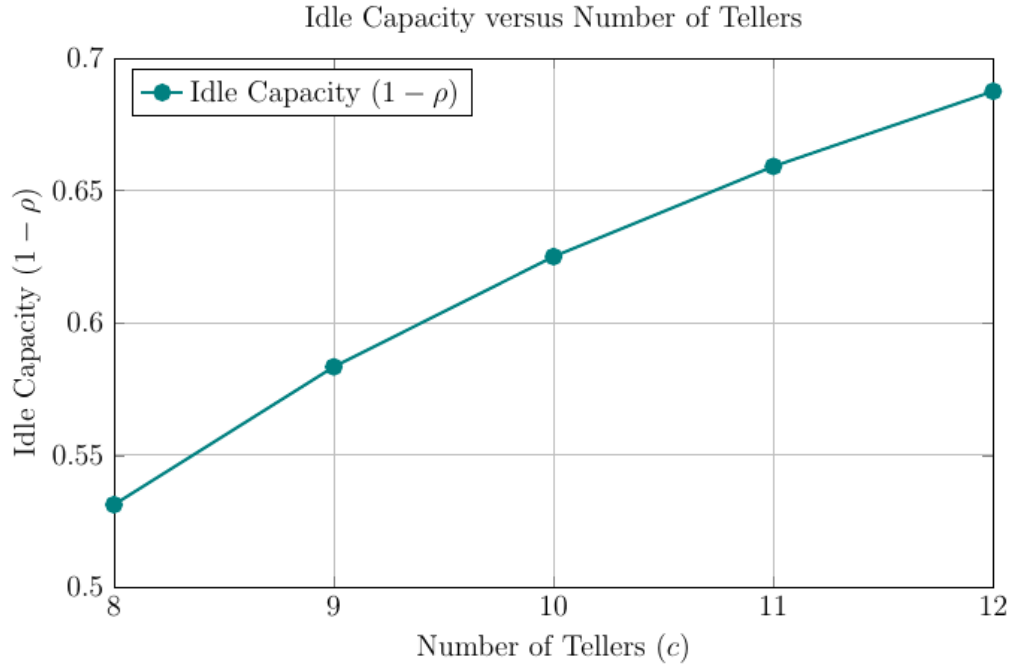


Figure 7: Variation of idle teller capacity with the number of tellers.

The results indicate that increasing the number of tellers from eight to twelve increases the available idle capacity by approximately 29.41%. This implies that the workload assigned to each teller decreases steadily as additional tellers are introduced. Although a higher idle capacity reduces customer waiting time and queue length, excessively high idle capacity may result in **underutilisation** of personnel and increased operating costs. Figure 7 illustrates the relationship between teller capacity and idle capacity. The graph shows a steady upward trend, indicating that additional tellers provide more spare service capacity, thereby reducing the likelihood of congestion within the banking system. It also suggests that a greater proportion of teller time is unused.

Therefore, while increasing the number of tellers improves customer service, excessive idle capacity may lead to **underutilisation** of personnel and higher operating costs. This observation reinforces the need to determine an optimal teller capacity that balances service efficiency with resource **utilisation**.

4.9.4. Comparative Sensitivity Analysis of Queue Performance Measures

The sensitivity analysis presented in the preceding sections demonstrates the effect of increasing the number of tellers on each queue performance measure. For ease of comparison, Table 4 **summarises** all the computed performance measures for teller capacities ranging from eight to twelve.

Table 4: Comparison of queue performance measures under different teller capacities

c	ρ	$P(\text{wait})$	L_q	W_q (sec)	W (min)	L
8	0.46875	0.01473	0.01296	1.03680	5.01720	3.76296
9	0.41667	0.00514	0.00367	0.29333	5.00489	3.75367
10	0.37500	0.00185	0.00111	0.08880	5.00148	3.75111
11	0.34091	0.00068	0.00035	0.02800	5.00047	3.75035
12	0.31250	0.00025	0.00011	0.00880	5.00015	3.75011

Table 4 shows that increasing the number of tellers consistently improves every performance measure. The traffic intensity decreases steadily from 0.46875 to 0.31250, indicating that the workload on each teller reduces as additional tellers are employed. Similarly, the probability that an arriving customer waits decreases by more than 98%, while the expected queue length and the average waiting time in the queue both reduce by approximately 99%. These results demonstrate that increasing service capacity has a substantial impact on eliminating customer congestion. In contrast, the average time spent in the system and the expected number of customers in the system exhibit relatively small reductions. This behaviour is expected because these two measures are dominated by the service component, which remains constant throughout the analysis. As the queue becomes progressively smaller, the total time spent in the system approaches the mean service time of five minutes, while the expected number of customers in the system approaches the theoretical value

$$\frac{\lambda}{\mu} = \frac{45}{12} = 3.7500. \quad (4.137)$$

Furthermore, the improvements obtained from increasing the teller capacity become progressively smaller beyond ten tellers. For example, increasing the number of tellers from ten to eleven reduces the average waiting time by only 0.06080 seconds, while increasing the capacity from eleven to twelve yields an additional reduction of only 0.01920 seconds. Similar diminishing improvements are observed for the probability of waiting and the expected queue length. Therefore, although increasing the number of tellers continuously improves system performance, the marginal benefits decrease after ten tellers. This suggests that operating the banking system with ten or eleven tellers provides a practical balance between customer service efficiency and resource **utilisation**, whereas increasing the teller capacity beyond this level yields only minimal operational improvements.

4.10. Optimal Teller Capacity Analysis

The primary objective of this study was to determine the effect of increasing the number of tellers on the performance of the banking queueing system. Although the preceding sections established that all queue performance measures improve as additional tellers are employed, an important managerial question remains: **“What is the optimal number of tellers required to achieve efficient service delivery without introducing unnecessary excess capacity?”** To answer this question, the improvements in each queue performance measure were examined as the teller capacity increased from eight to twelve. The analysis focused on the marginal improvement obtained by adding one additional teller at a time.

4.11. Marginal Improvement in Queue Performance

Table 5 **summarises** the percentage reduction in the principal queue performance measures resulting from each additional teller.

Table 5: Marginal improvement in queue performance measures

Increase in Tellers	Reduction in $P(\text{wait})$ (%)	Reduction in L_q (%)	Reduction in W_q (%)
8 $\rightarrow$ 9	65.11	71.68	71.71
9 $\rightarrow$ 10	64.01	69.75	69.72
10 $\rightarrow$ 11	63.24	68.47	68.47
11 $\rightarrow$ 12	63.24	68.57	68.57

Table 5 indicates that increasing teller capacity produces substantial reductions in customer waiting and queue congestion. However, the magnitude of these improvements becomes progressively smaller as more tellers are added. The most noticeable improvements occur when the number of tellers is increased from eight to ten. Beyond ten tellers, the reductions in queue length and waiting time become comparatively small because the system is already operating under very light congestion. This behaviour is characteristic of the law of diminishing marginal returns in queueing systems. Initially, adding tellers produces substantial reductions in queue congestion because service capacity increases faster than customer demand. After the queue has been virtually eliminated, additional tellers contribute only marginal improvements since most arriving customers are already receiving immediate service.

4.12. Evaluation of System Performance

The queue performance measures obtained for each teller capacity are summarised in Table 4. Several important observations can be made. First, the traffic intensity decreases steadily from 0.46875 when eight tellers are employed to 0.31250 when twelve tellers are employed, indicating a continuous reduction in teller workload. Secondly, the probability that an arriving customer waits decreases dramatically from 0.01473 to 0.00025, representing an overall reduction exceeding 98%. Similarly, the expected queue length decreases from 0.01296 customers to 0.00011 customers, while the average waiting time decreases from 1.03680 seconds to 0.00880 seconds. In contrast, the average time spent in the system changes only slightly, decreasing from 5.01720 minutes to 5.00015 minutes. Likewise, the expected number of customers in the system approaches

$$\frac{\lambda}{\mu} = \frac{45}{12} = 3.7500, \quad (4.138)$$

which represents customers receiving service rather than waiting in the queue. These observations indicate that once waiting is almost completely eliminated, additional tellers have little influence on the remaining performance measures.

4.13. Determination of the Optimal Number of Tellers

From the foregoing analysis, employing eight tellers results in noticeable customer waiting, although the banking system remains stable. Increasing the teller capacity to nine produces a substantial improvement in all queue performance measures. A further increase to ten tellers reduces the probability of waiting to less than 0.002, while the average waiting time falls below 0.1 seconds. At this stage, customer waiting has become practically negligible. Although increasing the teller capacity to eleven and twelve continues to improve the performance measures, the improvements are relatively small when compared with the additional staffing requirement. Therefore, from an operational perspective, employing more than ten tellers produces only marginal improvements in customer waiting while increasing the number of service facilities. Consequently, ten tellers may be regarded as the optimal operating point for the banking system because it provides an excellent balance between service efficiency and teller utilisation. Where the bank anticipates exceptionally high customer arrivals during peak business periods, increasing the teller capacity to eleven may provide additional operational flexibility with only a modest increase in staffing.

4.14. Summary of Findings

Based on the queueing analysis performed in this chapter, the following conclusions are drawn:

- (1) Increasing teller capacity consistently improves all queue performance measures.
- (2) The probability of customer waiting, expected queue length and average waiting time decrease rapidly as additional tellers are introduced.

- (3) The average time spent in the system and the expected number of customers in the system approach constant values because the service time remains fixed.
- (4) The greatest operational improvements occur when teller capacity is increased from eight to ten.
- (5) Beyond ten tellers, the banking system experiences diminishing marginal improvements in performance.
- (6) Based on the mathematical analysis of all queue performance measures, ten tellers provide the most efficient balance between customer service and resource **utilisation** under the assumed arrival and service rates.

The findings of this chapter therefore provide quantitative evidence that increasing teller capacity substantially improves the efficiency of banking operations while demonstrating that the benefits obtained from additional tellers eventually diminish.

5. Conclusion

The application of the $M/M/c$ queueing model has demonstrated its **importance** in evaluating customer service performance and supporting operational **decision-making** in retail banking. The calculated steady-state performance measures showed that the banking system operates efficiently under the assumed arrival and service conditions. Increasing teller capacity significantly improves customer service by reducing the probability of waiting, queue length and average waiting time. The sensitivity analysis further established that although additional tellers improve queue performance, the corresponding benefits diminish as teller capacity increases. Beyond ten tellers, only negligible improvements in queue performance were observed, whereas teller idle capacity increased steadily. This indicates that simply increasing the number of tellers does not always result in proportional improvements in operational efficiency. The study therefore concludes that an optimal balance between customer service quality and efficient resource **utilisation** is achieved with ten tellers under the estimated arrival and service rates. The $M/M/c$ queueing model therefore serves as an effective **decision-support** tool for determining appropriate teller staffing levels and improving customer service delivery in retail banking.

6. Limitations of the Study

This study was deliberately limited to the application of the classical $M/M/c$ queueing model for evaluating customer service performance in GCB Bank PLC, Tamale branch. The study focused exclusively on a multi-server queueing system in which customers are served by multiple tellers operating under the First-Come, First-Served (FCFS) queue discipline. The analysis was restricted to the determination of the steady-state queue performance measures, namely traffic intensity (ρ), probability of waiting ($P(\text{wait})$), expected queue length (L_q), average waiting time in the queue (W_q), average time spent in the system (W), and the expected number of customers in the system (L).

7. CONFLICT OF INTEREST

The authors declare no competing interests.

Declaration of AI Use

This manuscript was prepared through the intellectual contributions of the author(s) to the study design, data analysis, interpretation of results, and scholarly content. Grammarly and ChatGPT were used solely to assist with grammatical refinement and reference formatting during manuscript revision. The author(s) accept full responsibility for the accuracy, integrity, and final content of the manuscript.

References

- Afolalu, S. A., Babaremu, K. O., Ongbali, S. O., Abioye, A. A., Abdulkareem, A., & Adejuyigbe, S. B. (2019). Overview impact of application of queueing theory model on productivity performance in a banking sector. *Journal of Physics: Conference Series*, 1378, 032033. IOP Publishing. <https://doi.org/10.1088/1742-6596/1378/3/032033>
- Akande, S. A., Mohammed, U. N., Ibrahim, A. M., & Adedayo, O. A. (2022). Application of multi-server queue model ($M/M/c$) for waiting lines management in banking system. *Asian Journal of Mathematical Sciences*, 6(1), 75–83. <https://doi.org/10.22377/ajms.v6i1.418>

- Bolch, G., Greiner, S., de Meer, H., & Trivedi, K. S. (2006). *Queueing networks and Markov chains: Modeling and performance evaluation with computer science applications* (2nd ed.). John Wiley & Sons. <https://doi.org/10.1002/0471791571>
- Bunyaminu, A., Yakubu, I. N., Seidu, A. B., Bashiru, S., & Tuffour, J. K. (2023). Investigating the customer waiting experience in the banking sector in an emerging economy. In *Book of Abstracts: 5th Faculty of Accounting and Finance International Conference (FAFIC)*, Accra, Ghana. University of Professional Studies, Accra (UPSA).
- Chakraborty, P., & Honnappa, H. (2022). Strong embeddings for transitory queueing models. *Mathematics of Operations Research*, 47(2), 1048–1081. <https://doi.org/10.1287/moor.2021.1158>
- Giambene, G. (2014). M/G/1 queueing theory and applications. In *Queueing theory and telecommunications: Networks and applications* (pp. 367–413). Springer. https://doi.org/10.1007/978-1-4614-4084-0_6
- Iwa, L. L., & Audu, M. A. (2023). Analysis of post Covid-19 lockdown offline-customer service delivery in Access Bank Akwanga Branch: A queueing theory approach. *Asian Journal of Probability and Statistics*, 21(3), 23–32. <https://doi.org/10.9734/ajpas/2023/v21i3465>
- Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 24(3), 338–354. <https://doi.org/10.1214/aoms/1177728975>
- Kleinrock, L. (1975). *Queueing systems: Volume I: Theory*. John Wiley & Sons.
- Little, J. D. C., & Graves, S. C. (2008). Little’s law. In D. Chhajed & T. J. Lowe (Eds.), *Building intuition: Insights from basic operations management models and principles* (pp. 81–100). Springer. https://doi.org/10.1007/978-0-387-73699-0_5
- Mamo, T., Tolosa, D., Nagarajan, S., Paramasivam, V., Asfaw, W. H. M., & Umamaheswaran, S. (2026). Optimizing staffing level and waiting time using queueing model in bank service. *The Scientific World Journal*, 2026, 4133358. <https://doi.org/10.1155/tswj/4133358>
- Medhi, J. (2002). *Stochastic models in queueing theory* (2nd ed.). Academic Press. <https://doi.org/10.1016/B978-0-12-487462-6.X5000-0>
- Ngimanang, A. V. (2025). Quantifying the effects of queueing on customer satisfaction in commercial banks. *Asian Journal of Economics, Finance and Management*, 7(1), 1269–1278. <https://doi.org/10.56557/ajefm/2025/v7i1346>
- Odirichukwu, J. C., Lekara, T., & Odii, J. N. (2014). Banking queue system in Nigeria. *Computing, Information Systems, Development Informatics & Allied Research Journal*, 5(1), 95–106. <https://iiste.org/Journals/index.php/CIS/article/view/12008>
- Shortle, J. F., Thompson, J. M., Gross, D., & Harris, C. M. (2018). *Fundamentals of queueing theory* (5th ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119453765>
- Sztrik, J., Tóth, Á., Pintér, Á., & Bács, Z. (2020). The simulation of finite-source retrial queueing systems with two-way communications to the orbit and blocking. In V. M. Vishnevskiy, K. E. Samouylov, & D. V. Kozyrev (Eds.), *Distributed computer and communication networks: Control, computation, communications* (Vol. 1337, pp. 171–182). Springer Nature. https://doi.org/10.1007/978-3-030-66242-4_14
- Yeboah, M. K. (2026). Peak-hour teller staffing optimization in retail banking: An M/M/c queueing analysis. *Asian Journal of Mathematics and Computer Research*, 33(3), 294–317. <https://doi.org/10.56557/ajomcor/2026/v33i310860>
- Yifter, T., Mengstenew, M., Yoseph, S., & Moges, W. (2023). Modeling and simulation of queueing system to improve service quality at commercial bank of Ethiopia. *Cogent Engineering*, 10(2), 2274522. <https://doi.org/10.1080/23311916.2023.2274522>

COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.