International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 10 Issue 7 July - 2026, Pages: 15-28

Analysis and Simulation of an M/M/c Queuing Model with Dynamic Servers in The IM3 Kayoon Retail Outlet Service System

Toha Saifudin[1], Azka Dynina[2], Muhammad Dzaki Aqillah Rudy Katim[3], Ratna Anugrahaningtyas[4], Bashiri Surya Muslimin[5]

[1,2,3,4,5]Statistics Study Program, Department of Mathematics, Faculty of Science and Technology
Airlangga University
Surabaya, Indonesia
Corresponding author: tohasaifudin@fst.unair.ac.id

Abstract: *The IM3 Kayoon outlet is a telecommunications service facility frequently visited by the public to meet various communication needs, such as SIM card activation, purchasing data plans, and resolving customer issues. These service facilities experience high visitor traffic, and customers often have to wait in line before receiving service. Long lines cause customers to wait longer, especially during certain service hours when the number of active servers decreases due to staff break schedules. Therefore, a queueing simulation was conducted to analyze the service system at the IM3 Kayoon Outlet in Surabaya using the M/M/c model with a dynamic servers approach. The data used were collected through direct observation on May 21, 2026, from 11:00 AM to 2:00 PM WIB, with a total of 73 customer records divided into three observation periods. The data collected included customer arrival times, service start times, and service completion times for each customer service representative. The analysis in this study was performed using RStudio software as well as manual calculations. Based on the research results, it was concluded that the queueing system at the IM3 Kayoon outlet in Surabaya follows the (M/M/5):(FCFS/∞/∞) model during periods P1 and P3, and the (M/M/3):(FCFS/∞/∞) model during period P2, with utilization rates of 54.94%, 84.31%, and 42.75%, respectively; average waiting times of 0.4753 minutes, 11.7377 minutes, and 0.1137 minutes; and the probability of no customers in the system of 0.0616, 0.0418, and 0.1167 for each period. The results of this analysis are expected to provide concrete recommendations regarding the configuration of the number of active servers during each operational period to improve efficiency and service quality at the IM3 Kayoon Outlet in Surabaya. This study also contributes to the achievement of the Sustainable Development Goals (SDGs), particularly SDG 9 Industry, Innovation, and Infrastructure which promotes the development of high-quality and sustainable service infrastructure through an innovative approach based on queue simulation.*

Keywords— IM3 Retail Outlet, FCFS, M/M/c Dynamic, Queue Simulation

1. INTRODUCTION

IM3 Kayoon Outlet is one of the service facilities that plays an important role in meeting public communication needs, ranging from service activation, data package purchases, to customer complaint resolution. In this type of service, speed and accuracy are very important because they directly affect customer comfort and the image of the outlet as a public service provider [1]. Efforts to improve the efficiency of telecommunication outlet services are in line with SDG 9, namely Industry, Innovation and Infrastructure, which encourages the development of quality, innovative, and sustainable service infrastructure for society.

In the service process, queues occur because customer arrivals are random while service capacity is limited. Queueing theory is used to understand queue formation, resource utilization, and ways to improve the overall performance of the service system. Through this approach, waiting time, the number of customers in the system, and service efficiency can be analyzed more measurably [2].

IM3 Kayoon Outlet was selected as the research location because it is located in a strategic area and close to two major universities, namely Universitas Airlangga and Institut Teknologi Sepuluh Nopember. A location with high mobility tends to increase the number of customer arrivals at certain hours, which also increases the potential for queues. This condition makes IM3 Kayoon Outlet relevant to be analyzed using a queueing model approach [3].

Several previous studies have shown that queueing models can be used to describe service performance quantitatively. For example, research conducted at SPBU Dharmahusada showed that the M/M/c model was able to measure arrival rate, service rate, probability of an empty system, waiting time, and system utilization effectively [4]. These results show that queueing simulation can be an effective tool for evaluating service efficiency in facilities with high customer traffic.

However, several previous studies still have limitations, mainly because they often use a fixed number of servers throughout the observation period. In reality, service conditions are not always stable, since service needs may change according to peak hours, employee break schedules, and customer arrival patterns. This limitation indicates the need for a more adaptive approach so that the model used can better represent actual field conditions [5].

The novelty of this study lies in the application of the M/M/c queueing model with dynamic servers, where the number of servers changes based on operational hours. In this study, five servers were active from 11:00 to 12:00, then decreased to three servers from 12:00 to 13:00 due to the employee break period, and returned to five servers from 13:00 to 14:00. This approach was analyzed using R-based simulation so that the calculation of queueing parameters could adjust to changing service conditions more realistically [6].

Therefore, this study is expected to provide a more accurate description of the queueing characteristics at IM3 Kayoon Outlet, particularly in terms of waiting time, server utilization, and service efficiency in each observation period. The results are expected to serve as a basis for evaluating the arrangement of active servers at certain hours so that service remains optimal and customer experience can be improved [7].

2. LITERATURE REVIEW

2.1 IM3 Kayoon Outlet

IM3 Outlet is an official face-to-face service channel operated by PT Indosat Tbk (Indosat Ooredoo Hutchison) as one of the largest telecommunications providers in Indonesia. Indosat Ooredoo Hutchison, abbreviated as IOH, is an Indonesian telecommunications company owned by Ooredoo Hutchison Asia, a joint venture between Ooredoo and Hutchison Asia Telecom Group, since 2022, with its main brands including IM3 and 3 (Tri) [8]. IM3 Outlet serves as a tangible manifestation of the company's commitment to providing direct services to customers, complementing the digital channels already available. Services provided at the outlet include SIM card activation, data package purchases, SIM card replacements, complaint resolution, and postpaid services, making the outlet an important point of contact between customers and the company [9].

Pada tahun 2023, Indosat Ooredoo Hutchison meresmikan 32 In 2023, Indosat Ooredoo Hutchison simultaneously inaugurated 32 IM3 Outlets across various cities in Indonesia with a new, more digitalized appearance and concept, adopting the Fast, Simple and Flexible service concept that delivers devices, new systems, and faster solutions through digital services [10]. This concept reflects the direction of physical outlet transformation that functions not only as a transaction point but also as a digital experience center for customers. Service demand in densely populated urban areas continues to rise, particularly for card replacement services, product activation, and fast routine customer services, prompting Indosat to continuously update its outlet concept to be more efficient and accessible [11].

The presence of IM3 Outlet Kayoon in Surabaya is particularly relevant for in-depth study given its strategic position in a high-mobility area. Its proximity to major educational institutions makes customer visit traffic potentially high at certain hours, making efficient queue management a critical aspect of maintaining service quality. This aligns with the finding that service facilities in high-traffic urban areas require careful capacity management so that customer waiting times can be minimized and customer satisfaction maintained.

2.2 Queuing Theory

A queueing system is a service mechanism consisting of customers, servers, and rules that regulate the customer arrival and service process [9]. The purpose of queueing theory is to understand queue formation, resource utilization, and how to improve the performance of a service system [7]. Through queueing theory, customer waiting time, server utilization rate, and the optimal need for service facilities can be analyzed. Previous research states that the M/M/c queueing model can be used to improve system effectiveness by measuring arrival rate, service rate, number of customers in the queue, and average customer waiting time. In the context of IM3 service outlets, the application of queueing theory helps to understand queue density and service efficiency, so that the quality of customer service can be improved [11].

2.3 M/M/c Simulation

The M/M/c queuing model is one of the most widely applied multi-server queuing models in real service system analysis. The first letter M indicates that inter-arrival times follow an exponential distribution (Markovian), the second letter M indicates that service times also follow an exponential distribution, and c denotes the number of servers operating in parallel [12]. This model requires a steady-state condition, in which the system utilization rate $\rho = {}^{\lambda}/_{(c\mu)}$ must be less than one to prevent the system from experiencing unbounded queue accumulation [13].

In the M/M/c model, the customer arrival rate is denoted by λ and the service rate per server is denoted by μ. The probability of no customers in the system (P_0) is calculated using the equation:

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{\left({}^{\lambda}/_{\mu}\right)^n}{n!} + \frac{\left({}^{\lambda}/_{\mu}\right)^c}{c!\,(1-\rho)}\right]^{-1}$$

Based on the value of P_0 several queuing system performance measures can be derived, namely the average number of customers in the queue $\left(L_q\right)$, the average number of customers in the system (L), the average waiting time in the queue $\left(W_q\right)$, and the average time a customer spends in the system (W) [12]:

$$L_q = \frac{(\lambda/\mu)^c \cdot \rho}{c!\,(1-\rho)^2} P_0, \quad L = L_q + \frac{\lambda}{\mu}$$

$$W_q = \frac{L_q}{\lambda}, \quad W = W_q + \frac{1}{\mu}$$

Simulation of the M/M/c queuing system in this study was carried out using the R statistical software. The simulation approach enables researchers to computationally replicate system behavior based on empirical data obtained from field observations. Compared to purely analytical calculations, simulation offers the advantage of ease in

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 10 Issue 7 July - 2026, Pages: 15-28

modifying system parameters, making it possible to evaluate various service scenarios. The data used as simulation inputs include arrival times, service start times, and service completion times, which are then processed to estimate the values of λ and μ as the primary parameters of the M/M/c queuing model.

2.4 Dynamic Servers in Queuing Systems

In the context of queueing systems, a dynamic server refers to an approach in which the number of active servers is not fixed but is adjusted according to operational conditions that change over time. In operations research literature, there are two widely recognized staffing designs: dedicated staffing, in which the number of servers in each service group remains constant over time, and full flexibility, in which server assignments can be adjusted at any time as needed [14]. The dynamic server approach strikes a balance between these two extremes by adjusting the number of active servers based on specific time periods or queue conditions. This adjustment is essentially intended to balance operational efficiency with the quality of service provid [6].

The use of dynamic servers is particularly relevant for service facilities that experience fluctuations in demand throughout the day, such as due to peak hours, staff breaks, or uneven customer arrival patterns. In a congestion-based staffing (CBS) system, the number of active servers is dynamically adjusted based on queue length; when the queue reaches a certain upper threshold, additional servers are activated, while if the queue drops to a lower threshold and some servers become idle, those servers are deactivated. By selecting the appropriate threshold values, the actual queue length can be effectively controlled within the desired range with a high degree of certainty [15]. A similar approach can be applied periodically based on the operational schedule, for example, by setting a different number of servers for each time block according to customer arrival forecasts.

Several studies have demonstrated the effectiveness of the dynamic server approach in improving the performance of queuing systems. A study at Bank XYZ, which used the M/M/s model and WinQSB-based simulation, found fluctuating utilization patterns across periods; therefore, it was recommended that the number of servers be dynamically adjusted each month to avoid both underutilization and overload [16]. These findings indicate that maintaining a fixed number of servers throughout the entire operational period has the potential to create inefficiencies, whether in the form of wasted resources during slow periods or uncontrolled queue buildup during peak periods. Thus, the use of dynamic servers in the M/M/c model represents a more adaptive and realistic approach to describing actual service conditions in the field [17].

2.5 Exponential Distribution

The exponential distribution is a continuous probability distribution that describes the time between events in a Poisson process. It is characterized by a single parameter, λ (the rate of occurrence), such that the probability density function (PDF) is expressed as:

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0$$

while the cumulative distribution function (CDF) is expressed as:

$$F(x;\lambda) = 1 - e^{-\lambda x}, \quad x \geq 0$$

The mean of this distribution is $1/\lambda$, which in the queuing context is interpreted as the average inter-arrival time or the average service time. The simplicity of its mathematical form makes the exponential distribution straightforward to implement and estimate in various stochastic models, including queuing systems [18].

The most fundamental property of the exponential distribution in queuing theory is the memoryless property, formally expressed as:

$$P(\, X > s + t \mid X > s \,) = P(X > t)$$

This property means that the probability of a future event occurring is entirely independent of the time that has already elapsed. In the service context, if the service time of a customer follows an exponential distribution, the remaining service time required is unaffected by how long the customer has already been served. This property underlies the formation of a continuous time Markov chain (CTMC), making the exponential distribution the mathematical foundation that enables queuing systems to be analyzed analytically [19].

The relationship between the exponential distribution and the M/M/c model is particularly close. The letter "M" in that notation stands for Markovian, indicating that both inter-arrival times and service times follow an exponential distribution [20]. In practice, this distribution has been widely used to verify queuing model assumptions across various public service facilities, ranging from railway stations to telecommunications service outlets. Testing the fit of the exponential distribution against empirical data is typically carried out using goodness-of-fit tests such as the Kolmogorov-Smirnov test, so that the validity of the model assumptions can be confirmed before queuing parameters are estimated further [21].

3. RESEARCH METHODOLOGY

3.1 Data and Data Sources

The data used in this study were obtained from direct observation of the service system at IM3 Kayoon Outlet, which was conducted on May 21, 2026. The observation was carried out for three hours, from 11:00 to 14:00 WIB, with a total of 73 observation data. The collected data included customer arrival time, service start time, and service completion time. These data were then used to analyze the characteristics of the queueing system and evaluate service efficiency using the dynamic M/M/c queueing model.

The details of the observations in this study are as follows.

Tabel 1. Research Data Details

Period	Customer	Costumer Service	Arrival	Service Start	Service End
P1	1	CS1	11:00:26	11:00:26	11:01:50
P1	2	CS2	11:01:55	11:01:55	11:03:04
P1	3	CS3	11:02:32	11:02:32	11:08:58
⋮	⋮	⋮	⋮	⋮	⋮
P2	28	CS1	12:01:45	12:01:45	12:05:16
P2	29	CS2	12:02:09	12:02:09	12:09:36
P2	30	CS3	12:06:43	12:06:43	12:53:17
⋮	⋮	⋮	⋮	⋮	⋮
P3	71	CS4	13:59:30	13:59:30	14:00:30
P3	72	CS5	13:59:30	14:00:09	14:01:14
P3	73	CS3	13:59:30	14:00:30	14:05:06

3.2 Research Stages and Algorithms

The analysis stages in this study follow a dynamic M/M/c simulation approach with the following details..

1. Conducting a literature review related to the research topic in terms of both case studies and methods used, particularly regarding M/M/c queuing theory and discrete simulation.
2. Conducting direct observation to collect data in the form of customer arrival times, service start times, and service end times at IM3 Outlet Kayoon Surabaya. The data is divided into three observation periods, namely P1 (11:00–12:00) with 5 active servers, P2 (12:00–13:00) with 3 active servers as 2 servers were on break, and P3 (13:00–14:00) with 5 active servers again.
3. erforming data analysis using the dynamic M/M/c queuing simulation method with the number of servers adjusted to the conditions of each period.

The following is the algorithm used in the program for queuing simulation at IM3 Outlet Kayoon Surabaya.

1. Observational data is read from an Excel file and time columns are converted to seconds, after which service time is calculated as the difference between service end time and service start time in minutes. The data is divided into three periods based on row order..
2. For each period, the interarrival time, arrival rate, $= \frac{Number\ of\ Arrivals}{Observation\ Duration}$, and service rate $\frac{1}{\mu} = \frac{Total\ Service\ Time}{Number\ of\ Customers\ Served}$.
3. A Kolmogorov-Smirnov test is performed on the interarrival time and service time. If the $p - value > 0{,}05$,the exponential distribution is accepted and the analysis proceeds.
4. The queuing model for each period is determined: (M/M/5):(FCFS/∞/∞) for P1 and P3, and (M/M/3):(FCFS/∞/∞) for P2 as two servers were on break.
5. Queuing characteristics of the M/M/c model are calculated, including ρ, P_0, L_q, L_s, W_q, dan W_s for each period.
6. A discrete simulation table is constructed from the original data of 73 customers, producing columns for interarrival time, service time, arrival time, serving server, service start and end times, waiting time, and total time in the system.
7. System performance summaries per period and server occupancy are calculated, after which all results are visualized in an $N(t)$ graph showing the number of customers.

The flow of these research stages is depicted systematically in the form of a flowchart as follows.

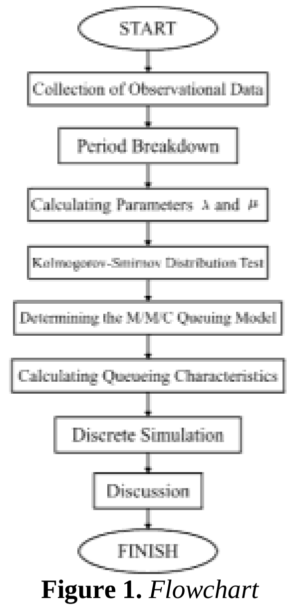


Figure 1. *Flowchart*

4. Results and Discussion

Based on the observational data collected at the IM3 Kayoon Service Center, an analysis was conducted to identify the characteristics of the queueing system and evaluate its service performance. The analysis employed an M/M/c queueing simulation model with a dynamic server approach using customer arrival times, service start times, and service completion times as input data. The analysis included descriptive statistics, exponential distribution testing, steady-state condition analysis, and the evaluation of key queueing

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 10 Issue 7 July - 2026, Pages: 15-28

performance measures, including the arrival rate, service rate, server utilization, average queue length, and average customer waiting time. Furthermore, server occupancy and server activity were analyzed to assess the effectiveness of server utilization during each observation period.

4.1 Descriptive Analysis of Observational Data

A descriptive analysis was conducted to provide an initial overview of the customer distribution within the service system at the IM3 Kayoon Service Center. The analysis was based on the number of customers served by each customer service representative (CS) during the three observation periods, namely P1, P2, and P3.

This analysis illustrates the distribution of customer service workloads among the servers under different staffing conditions. During periods P1 and P3, five servers were active, whereas only three servers were operated during period P2. The variation in the number of active servers affected the distribution of customers served across the observation periods.

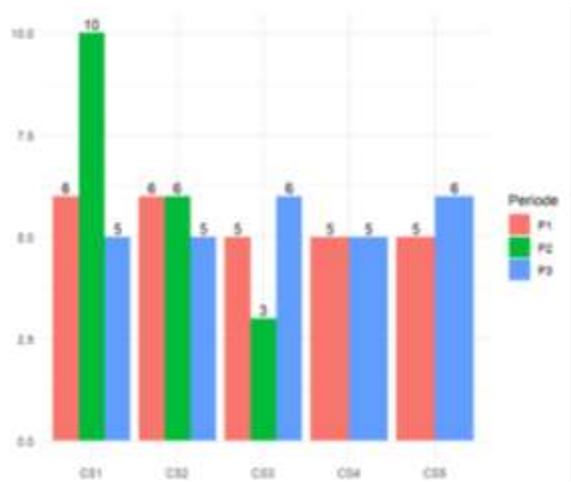


Figure 2. Number of Customers Served by Each Customer Service Representative Across Observation Periods

Figure 2 presents the number of customers served by each customer service representative during the three observation periods. During period P1, the customer distribution was relatively balanced across all servers. CS1 and CS2 each served six customers, while CS3, CS4, and CS5 each served five customers. This indicates that the workload was distributed evenly when all five servers were in operation.

During period P2, the number of active servers decreased to three, resulting in a more concentrated workload. CS1 served the largest number of customers (10), followed by CS2 with six customers and CS3 with three customers. This finding indicates that reducing the number of active servers increased the service workload on the remaining servers, particularly CS1.

In period P3, the customer distribution became more balanced after the number of active servers returned to five. CS1, CS2, and CS4 each served five customers, whereas CS3 and CS5 each served six customers. This suggests that increasing the number of active servers improved the distribution of service workloads among the customer service representatives.

Overall, the bar chart demonstrates that changes in the number of active servers influenced both customer distribution and service workload among the customer service representatives within the queueing system.

4.2 Exponential Distribution Test for Interarrival Time

The goodness-of-fit of the interarrival time data to the exponential distribution was evaluated using the Kolmogorov-Smirnov test. The hypotheses were defined as follows:

H_0 : The interarrival time follows an exponential distribution

H_1 : The interarrival time does not follow an exponential distribution

The Kolmogorov-Smirnov test results for each observation period are presented in Table 2.

Table 2. Kolmogorov-Smirnov Test Results for the Exponential Distribution of Interarrival Time

Period	Test Statistic (D)	*p-value*	Decision
P1	0.15195	0.5857	Fail to reject H_0
P2	0.2708	0.1426	Fail to reject H_0
P3	0.1153	0.8793	Fail to reject H_0

As shown in Table 2, the p-values for all observation periods exceed the significance level of 0.05. Therefore, the null hypothesis cannot be rejected, indicating that the interarrival time data in all observation periods are adequately described by an exponential distribution. These results satisfy one of the fundamental assumptions of the M/M/c queueing model with dynamic servers.

To provide a visual assessment of the distribution, histograms of the interarrival time data for periods P1, P2, and P3 are presented in Figure 3. The histograms exhibit the characteristic right-skewed pattern of an exponential distribution, supporting the statistical results obtained from the Kolmogorov-Smirnov test.

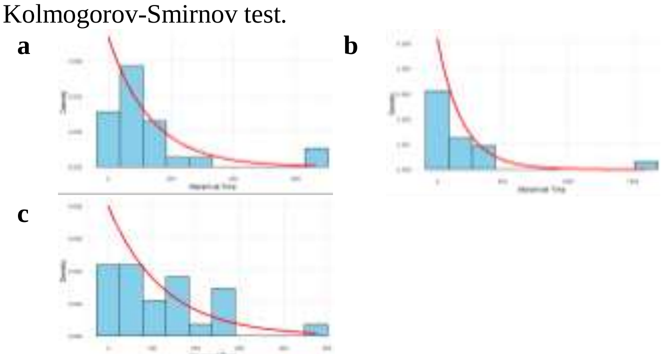


Figure 3. Histogram of Interarrival Time Customer at IM3 Kayoon Outlets for **a.** Period P1 **b.** Period P2 **c.** Period P3

4.3 Exponential Distribution Test for Service Time

The goodness of fit of the service time data to the exponential distribution was also evaluated using the

Kolmogorov-Smirnov test. The hypotheses were defined as follows:

H_0 : The service time follows an exponential distribution

H_1 : The service time does not follow an exponential distribution

The Kolmogorov-Smirnov test results for each observation period are presented in Table 3.

Table 3. Kolmogorov-Smirnov Test Results for the Exponential Distribution of Service Time

Period	Test Statistic (D)	*p-value*	Decision
P1	0.15329	0.5498	Fail to reject H_0
P2	0.13845	0.8121	Fail to reject H_0
P3	0.20773	0.1944	Fail to reject H_0

As presented in Table 3, the *p*-values for all observation periods are greater than the significance level of 0.05. Therefore, the null hypothesis cannot be rejected, indicating that the service time data for periods P1, P2, and P3 are adequately described by an exponential distribution. Consequently, the service process satisfies the exponential service-time assumption required for the M/M/c queueing model with dynamic servers.

Figure 4 presents the histograms of the service time data for the three observation periods. The distributions exhibit the typical right-skewed shape associated with an exponential distribution, providing visual support for the statistical results obtained from the Kolmogorov-Smirnov test.

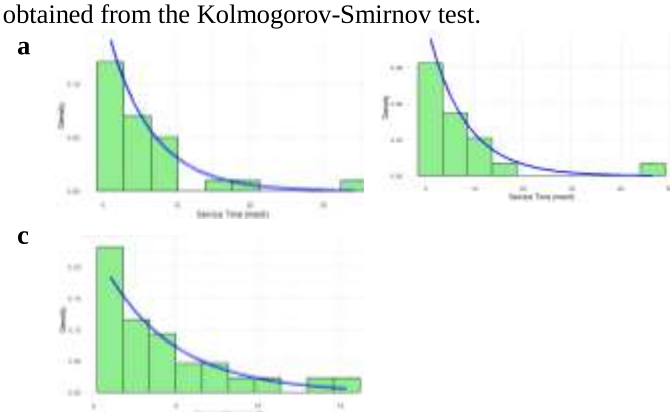


Figure 4. Histogram of Service Time Customer at IM3 Kayoon Outlets for **a.** Period P1 **b.** Period P2 **c.** Period P3

4.4 Steady-State Testing

A steady state is defined as a condition where the system's utilization rate $\rho = \lambda/\mu < 1$. If $\lambda < \mu$ which means $\rho < 1$, this indicates that the arrival rate is lower than the service rate, so the steady-state condition is met. Conversely, if $\rho \geq 1$, it is necessary to increase the number of servers or speed up service. The following is an analysis of the average arrival rate, average service time, and load ratio to determine whether steady-state conditions are met.

4.4.1 Steady-State Testing for Period 1

The average number of customers at the IM3 Kayoon Outlet during Period 1 is as follows.

$$\lambda_1 = \frac{Number\ of\ Arrivals_{P1}}{Observation\ Duration_{P1}}$$

$$\lambda_1 = \frac{27}{59{,}04} = 0.4571 \text{ customers per minute}$$

The average service time at the IM3 Kayoon Outlet during Period 1 is as follows.

$$\frac{1}{\mu_1} = \frac{Service\ Duration_{P1}}{Number\ of\ Customers\ Served_{P1}}$$

$$\frac{1}{\mu_1} = \frac{162{,}2667}{27} = 6{,}0099 \text{ minutes per customer}$$

resulting in the following service rate:

$$\mu_1 = \frac{1}{6{,}0099} = 0{,}1664 \text{ customers per minute}$$

The queueing load ratio at the IM3 Kayoon Outlet during observation period 1 is as follows.

$$\rho_1 = \frac{\lambda_1}{c_1.\mu_1} = \frac{0.4571}{5\times0{,}1664} = 0.5494$$

Based on these calculations, the value of ρ_1 is found to be 0.5494, which means that the ratio between the arrival rate and the system's service capacity or system utilization is 54,94%. Since the value of $\rho_1 < 1$, is less than 1, it can be concluded that the arrival rate is still lower than the system's service capacity, so steady-state conditions have been met and the queueing system can operate stably during the first observation period.

4.4.2 Steady-State Testing for Period 2

The average number of customers at the IM3 Kayoon Outlet during Period 2 is as follows.

$$\lambda_2 = \frac{Number\ of\ Arrivals_{P2}}{Observation\ Duration_{P2}}$$

$$\lambda_2 = \frac{19}{57{,}75} = 0{,}3290 \text{ customers per minute}$$

The average service time at the IM3 Kayoon Outlet during Period 2 is as follows.

$$\frac{1}{\mu_2} = \frac{Service\ Duration_{P2}}{Number\ of\ Customers\ Served_{P2}}$$

$$\frac{1}{\mu_2} = \frac{146{,}0667}{19} = 7{,}6877 \text{ minutes per customer}$$

resulting in the following service rate:

$$\mu_2 = \frac{1}{7{,}6877} = 0{,}1301 \text{ customers per minute}$$

The queueing load ratio at the IM3 Kayoon Outlet during observation period 2 is as follows.

$$\rho_2 = \frac{\lambda_2}{c_2.\mu_2} = \frac{0{,}3290}{3\times0{,}1301} = 0.8431$$

Based on these calculations, the value of ρ_1 was found to be 0.8431 which means that the ratio between the arrival rate and the system's service capacity or system utilization is

84,31%. This value is higher than that of the first period (54,94%), this is because during the second period (12:00–1:00 PM), two staff members were on break, leaving only three servers actively serving customers, while customer arrivals continued. This reduction in the number of servers caused the system load to increase significantly compared to the previous period. Since the value of $\rho_2 < 1$, it can be concluded that the arrival rate is still lower than the system's service capacity, meaning that steady-state conditions were met and the queueing system operated stably during the second observation period.

4.4.3 Steady-State Testing for Period 3

The average number of customers at the IM3 Kayoon Outlet during Period 3 is as follows.

$$\lambda_3 = \frac{Number\ of\ Arrivals_{P3}}{Observation\ Duration_{P3}}$$

$$\lambda_3 = \frac{27}{54{,}25} = 0{,}4997 \text{ customers per minute}$$

The average service time at the IM3 Kayoon Outlet during Period 3 is as follows.

$$\frac{1}{\mu_3} = \frac{Service\ Duration_{P3}}{Number\ of\ Customers\ Served_{P3}}$$

$$\frac{1}{\mu_3} = \frac{115{,}95}{27} = 4{,}2944 \text{ minutes per customer}$$

resulting in the following service rate:

$$\mu_3 = \frac{1}{4{,}2944} = 0{,}2329 \text{ customers per minute}$$

The queueing load ratio at the IM3 Kayoon Outlet during observation period 3 is as follows

$$\rho_3 = \frac{\lambda_3}{c_3 . \mu_3} = \frac{0{,}4997}{5 \times 0{,}2329} = 0.4275$$

Based on these calculations, the value of ρ_1 was found to be 0.4275 which means that the ratio between the arrival rate and the system's service capacity or system utilization is 42,75%. This value shows a significant decrease compared to the second period (84.31%); this is because during the third period (1:00 PM–2:00 PM), all 5 servers became active again after the downtime ended, so the system's service capacity returned to optimal levels, similar to the first period. It was this increase in the number of servers that caused the system load to drop drastically compared to the previous period. Since the value of $\rho_3 < 1$, it can be concluded that the arrival rate is still lower than the system's service capacity, meaning that steady-state conditions have been met and the queuing system was able to operate stably during the third observation period.

4.5 Queuing Model and Simulation for the IM3 Kayoon Outlet

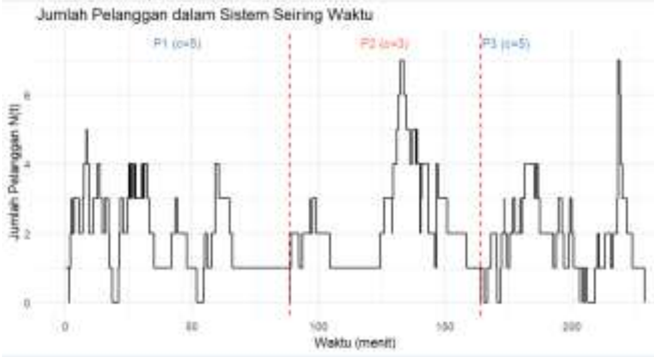
Based on the results of the goodness-of-fit test for the exponential distribution of interarrival times and service times, a queueing model appropriate for the observed system can be determined. The queueing system at the IM3 Kayoon Outlet follows a Multi Channel Single Phase model with a First-Come, First-Served (FCFS) queueing discipline. In this study, the system capacity and the call source are assumed to be unlimited. Since the number of active servers varies between periods due to staff break schedules, the analysis is conducted dynamically on a per-period basis using the M/M/c model. During the first period (11:00 a.m.–12:00 p.m.) and the third period (1:00 p.m.–2:00 p.m.), all 5 servers are active, so the model is (M/M/5): (FCFS/∞/ ∞), whereas in the second period (12:00–13:00), two servers are on break, leaving only three active servers, so the model is (M/M/3): (FCFS/∞/∞). This condition, where servers temporarily stop serving, is known in queuing theory as the dynamic model in a multiserver system. Table 4 below shows the simulation results of the queuing model at the IM3 Kayoon Outlet, based on observations of 73 customers.

Tabel 4. Queue Simulation Results Observation Data

cust	Int_arr	serv_time	arrival	inserver	mulai_serv	end_serv	wait	sys_time
1	0	1,3	0	1	0	1,4	0	1,4
2	1,4833	1,15	1,4833	2	1,4833	2,6333	0	1,15
3	0,6167	6,4333	2,1	3	2,1	8,5333	7	6,4333
4	0,3333	2,85	2,4333	4	2,4333	5,2833	0	2,85
5	1,05	15,1	3,4833	5	3,4833	18,5833	0	15,1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
28	0	3,5167	0	1	0	3,5167	0	3,5167
29	0,4	7,45	0,4	2	0,4	7,85	0	7,45
30	4,5667	46,5667	4,9667	3	4,9667	51,5333	0	46,5667
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
71	0	1	54,25	4	54,25	55,25	0	1
72	0	1,0833	54,25	5	54,9	55,9833	0,65	1,7333
73	0	4,6	54,25	3	55,25	59,85	1	5,6

Based on the results of the queueing simulation using the observed data, the number of customers in the system over time is visually presented in Figure 5 below.

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 10 Issue 7 July - 2026, Pages: 15-28



Gambar 5. Jumlah Pelanggan dalam Sistem

4.6 M/M/c Queueing Model Analysis with Dynamic Servers

Based on the observational data collected at the IM3 Kayoon Service Center, the queueing system was analyzed using the M/M/c model with a dynamic server approach across three observation periods (P1, P2, and P3). The variation in the number of active servers during each period was considered to evaluate its impact on the queueing system performance.

4.6.1 Queueing Model Analysis for Period P1

During Period P1, five servers were actively serving a total of 27 customers. The arrival rate was calculated to be 0.4571 customers per minute, while the service rate for each server was 0.1664 customers per minute. The ratio between the arrival rate and service rate is calculated as follows:

$$\frac{\lambda_1}{\mu_1} = \frac{0.4571}{0.1664} = 2.7470$$

This value indicates that the customer arrival intensity is relatively high compared to the service capacity of a single server. Therefore, multiple servers are required to maintain the stability of the queueing system.

The probability that no customers are present in the system (P_0) is calculated using the following equation:

$$P_0 = \left[\sum\nolimits_{n=0}^{c-1} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} + \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!\,(1-\rho_1)}\right]^{-1}$$

$$P_0 = \left[\frac{(2.7470)^0}{0!} + \frac{(2.7470)^1}{1!} + \frac{(2.7470)^2}{2!} + \frac{(2.7470)^3}{3!} + \frac{(2.7470)^4}{4!} + \frac{(2.7470)^5}{5!\,(1-0.5494)}\right]^{-1}$$

$$P_0 = [1 + 2.7470 + 3.773 + 3.4548 + 2.3726 + 2.8929]^{-1}$$

$$P_0 = \frac{1}{16.2403} = 0.0616$$

The result indicates that the probability of having no customers in the system is 0.0616, or approximately 6.16%, suggesting that the system is occupied most of the time.

The average number of customers waiting in the queue (L_q) is calculated using:

$$L_q = \frac{P_0(\lambda/\mu)^c \rho_1}{c!\,(1-\rho_1)^2}$$

$$L_q = \frac{0.0616(2.7470)^5(0.5494)}{5!\,(1-0.5494)^2}$$

$$L_q = \frac{0.0616(150.4204)(0.5494)}{120(0.4506)^2}$$

$$L_q = \frac{5.0907}{24.3648}$$

$$L_q = 0.2173$$

The value of L_q indicates that, on average, fewer than one customer is waiting in the queue before receiving service. This relatively low queue length suggests that the service capacity is sufficient to accommodate customer arrivals efficiently, resulting in minimal congestion during this observation period.

The average waiting time in the queue (W_q) is calculated as:

$$W_q = \frac{L_q}{\lambda}$$

$$W_q = \frac{0.2173}{0.4571}$$

$W_q = 0.4753$ minutes

This result shows that customers wait for approximately 0.48 minutes before being served, indicating a relatively short waiting time.

The average number of customers in the system (L_s) is obtained by:

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

$$L_s = 0.2173 + 2.7470$$

$$L_s = 2.9643$$

This means that, on average, approximately three customers are present in the system, including both those waiting and those currently being served.

The average time a customer spends in the system (W_s) is calculated as:

$$W_s = \frac{L_s}{\lambda}$$

$$W_s = \frac{2.9643}{0.4571}$$

$W_s = 6.4852$ minutes

Thus, customers spend an average of 6.49 minutes in the system from arrival until the completion of service.

Finally, the server utilization is calculated as:

$$\rho_1 \times 100\% = 0.5494 \times 100\% = 54.94\%$$

This utilization level indicates that the five servers are busy serving customers for approximately 54.94% of the operating time, while they remain idle for the remaining 45.06%. Since the utilization is well below 100% and the average queue length and waiting time are relatively low, the queueing system during Period P1 can be considered stable and capable of providing efficient customer service. Overall, the service capacity in this period is adequate to accommodate customer arrivals without generating significant queues or excessive waiting times.

4.6.2 Queueing Model Analysis for Period P2

During Period P2, three servers were actively serving a total of 19 customers. The customer arrival rate was calculated to be 0.329 customers per minute, while the service rate for each server was 0.1301 customers per minute. The ratio between the arrival rate and service rate is calculated as follows:

$$\frac{\lambda_2}{\mu_2} = \frac{0.329}{0.1301} = 2.5288$$

This value indicates that the customer arrival intensity remains relatively high compared to the service capacity of a single server. Consequently, multiple servers are required to maintain the stability of the queueing system.

The probability that no customers are present in the system (P_0) is calculated using the following equation:

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} + \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!\,(1-\rho_2)}\right]^{-1}$$

$$P_0 = \left[\frac{(2.5288)^0}{0!} + \frac{(2.5288)^1}{1!} + \frac{(2.5288)^2}{2!} + \frac{(2.5288)^3}{3!\,(1-0,8431)}\right]^{-1}$$

$$P_0 = [1 + 2.5288 + 3.1974 + 17.1779]^{-1}$$

$$P_0 = \frac{1}{23.9041} = 0.0418$$

The result indicates that the probability of having no customers in the system is 0.0418, or approximately 4.18%, indicating that the system is occupied almost continuously during this observation period.

The average number of customers waiting in the queue (L_q) is calculated as follows:

$$L_q = \frac{P_0(\lambda/\mu)^c \rho_2}{c!\,(1-\rho_2)^2}$$

$$L_q = \frac{0.0418(2.5288)^3(0.8431)}{3!\,(1-0.8431)^2}$$

$$L_q = \frac{0.0418(16.1712)(0.8431)}{6(0.1569)^2}$$

$$L_q = \frac{0.5699}{0.1477}$$

$$L_q = 3.8617$$

The calculated L_q value indicates that an average of approximately four customers are waiting in the queue before receiving service. Compared with Period P1, this considerably longer queue reflects the higher workload experienced by the service system as a result of the reduced number of active servers. The high queue length suggests that customer arrivals accumulate more rapidly than they can be served, leading to increased congestion.

The average waiting time in the queue (W_q) is calculated as:

$$W_q = \frac{L_q}{\lambda}$$

$$W_q = \frac{3.8617}{0.329}$$

$$W_q = 11.7377 \text{ minutes}$$

This result indicates that customers wait for an average of approximately 11.74 minutes before receiving service. The substantial increase in waiting time compared with Period P1 demonstrates the impact of reduced service capacity on customer experience.

The average number of customers in the system (L_s) is calculated as:

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

$$L_s = 3.8617 + 2.5288$$

$$L_s = 6.3905$$

This result indicates that, on average, approximately six to seven customers are present in the system at any given time, including both those waiting in the queue and those currently being served.

The average time a customer spends in the system (W_s) is calculated using:

$$W_s = \frac{L_s}{\lambda}$$

$$W_s = \frac{6.3905}{0.329}$$

$$W_s = 19.4254 \text{ minutes}$$

Thus, customers spend an average of approximately 19.43 minutes in the system from arrival until the completion of service. This represents a significant increase compared with Period P1 and reflects the effect of higher system congestion.

Finally, the server utilization is calculated as:

$$\rho_2 \times 100\% = 0.8431 \times 100\% = 84.31\%$$

This utilization level indicates that the three servers are busy serving customers for approximately 84.31% of the operating time, leaving only 15.69% of the time idle. Although the utilization is substantially higher than in Period P1 due to the reduced number of active servers during the break schedule, the system remains in a steady-state condition because the utilization factor satisfies the stability requirement ($\rho < 1$). Nevertheless, the high utilization is accompanied by a marked increase in queue length and waiting time, indicating that the system operates much closer to its capacity limit. Consequently, while the available servers are still sufficient to maintain system stability, service performance deteriorates as customers experience longer queues and extended waiting times.

4.6.3 Queueing Model Analysis for Period P3

During Period P3, five servers were actively serving a total of 27 customers. The customer arrival rate was calculated to be 0.4977 customers per minute, while the service rate for each server was 0.2329 customers per minute. The ratio between the arrival rate and service rate is calculated as follows:

$$\frac{\lambda_3}{\mu_3} = \frac{0.4977}{0.2329} = 2.1370$$

This value indicates that the customer arrival intensity is relatively high compared to the service capacity of a single server. Therefore, the simultaneous operation of multiple

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 10 Issue 7 July - 2026, Pages: 15-28

servers is necessary to maintain the stability of the queueing system and minimize customer congestion.

The probability that no customers are present in the system (P_0) is calculated using the following equation:

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} + \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!\,(1-\rho_3)}\right]^{-1}$$

$$P_0 = \left[\frac{(2.1370)^0}{0!} + \frac{(2.1370)^1}{1!} + \frac{(2.1370)^2}{2!} + \frac{(2.1370)^3}{3!} + \frac{(2.1370)^4}{4!} + \frac{(2.1370)^5}{5!\,(1-0.4275)}\right]^{-1}$$

$$P_0 = [1 + 2.1370 + 2.2834 + 1.6265 + 0.8690 + 0.6487]^{-1}$$

$$P_0 = \frac{1}{8.5646} = 0.1167$$

The result indicates that the probability of having no customers in the system is 0.1167, or approximately 11.67%, which is higher than in the previous observation periods. This suggests that the system is idle more frequently due to the increased service capacity provided by the five active servers.

The average number of customers waiting in the queue (L_q) is calculated using:

$$L_q = \frac{P_0(\lambda/\mu)^c \rho_3}{c!\,(1-\rho_3)^2}$$

$$L_q = \frac{0.1167(2.1370)^5(0.4275)}{5!\,(1-0.4275)^2}$$

$$L_q = \frac{0.1167(44.5679)(0.4275)}{120(0.5725)^2}$$

$$L_q = \frac{2.2235}{33.3908}$$

$$L_q = 0.0566$$

The calculated value of $L_q$indicates that, on average, fewer than one customer is waiting in the queue before receiving service. This very small queue length demonstrates that the available service capacity is more than sufficient to accommodate customer arrivals efficiently, resulting in minimal congestion during this period.

The average waiting time in the queue (W_q) is calculated as:

$$W_q = \frac{L_q}{\lambda}$$

$$W_q = \frac{0.0566}{0.4977}$$

$W_q = 0.1137$ minutes

This result indicates that customers wait for an average of only 0.11 minutes before receiving service, reflecting a highly efficient queueing process with almost immediate access to service.

The average number of customers in the system (L_s) is calculated as:

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

$$L_s = 0.0566 + 2.1370$$

$$L_s = 2.1936$$

This result indicates that, on average, approximately two to three customers are present in the system at any given time, including both those waiting and those being served.

The average time a customer spends in the system (W_s) is calculated using:

$$W_s = \frac{L_s}{\lambda}$$

$$W_s = \frac{2.1936}{0.4977}$$

$W_s = 4.4081$ minutes

Thus, customers spend an average of approximately 4.41 minutes in the system from arrival until the completion of service, representing the shortest system time among the three observation periods.

Finally, the server utilization is calculated as:

$$\rho_3 \times 100\% = 0.4275 \times 100\% = 42.75\%$$

This utilization level indicates that the five servers are busy serving customers for approximately 42.75% of the operating time, while they remain idle for the remaining 57.25%. The restoration of all servers after the break period significantly reduces the system workload, as reflected by the substantial decreases in the average queue length ($L_q = 0.0566$) and average waiting time ($W_q = 0.1137$ minutes) compared with Period P2. Although customer arrivals remain relatively high, the increased service capacity enables the system to operate efficiently with minimal waiting time and virtually no queue formation. Overall, Period P3 demonstrates the best queueing performance among the three observation periods, indicating that the allocation of five active servers provides sufficient capacity to maintain stable and efficient service operations.

4.7 Comparison of Queueing System Performance Across Observation Periods

To evaluate the differences in queueing system performance across the three observation periods, a comparison was conducted using several performance measures of the M/M/c queueing model. These measures include the server utilization (ρ), the average number of customers waiting in the queue (L_q), the average number of customers in the system (L_s), the average waiting time in the queue (W_q), and the average time customers spend in the system (W_s). The comparison of queueing system performance is presented in Table 5.

Table 5. Comparison of Queueing System Performance

Period	c	ρ	L_q	L_s	W_q	W_s
P1	5	0,5494	0,2173	2,9643	0,4753	6,4852
P2	3	0,8431	3,8617	6,3905	11,7377	19,4254
P3	5	0,4275	0,0566	2,1936	0,1137	4,4081

Based on Table 5, all observation periods satisfy the stability condition ($\rho < 1$), indicating that the queueing system remained in a steady-state condition throughout the observations. However, the system performance varied according to the number of active servers.

Period P1 exhibited moderate performance with five active servers, resulting in a server utilization of 54.94%, an

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 10 Issue 7 July - 2026, Pages: 15-28

average queue length of 0.2173 customers, an average waiting time of 0.48 minutes, and an average system time of 6.49 minutes. These results indicate that the service system operated efficiently with minimal queue formation.

In contrast, Period P2 showed the highest server utilization at 84.31%, with the average queue length increasing to 3.8617 customers and the average waiting time rising to 11.74 minutes. The average time customers spent in the system also reached 19.43 minutes, indicating that the reduction in the number of active servers significantly increased system congestion.

After five servers resumed operation in Period P3, the server utilization decreased to 42.75%, while the average queue length fell to 0.0566 customers and the average waiting time decreased to only 0.11 minutes. The average system time was also reduced to 4.41 minutes, demonstrating the highest level of service efficiency among all observation periods.

Overall, the results indicate that the number of active servers has a significant impact on queueing system performance. A reduction in the number of servers increases server utilization, queue length, and customer waiting time, whereas increasing the number of active servers improves service efficiency by reducing congestion and minimizing waiting times.

4.8 Server Occupancy Level

In addition to evaluating the queueing system performance during each observation period, this study also analyzes the occupancy level of each server to assess the workload of individual customer service representatives. Server occupancy represents the proportion of operating time during which a server is actively serving customers. A higher occupancy value indicates a greater level of server utilization. The occupancy level of each server is presented in Figure 6.

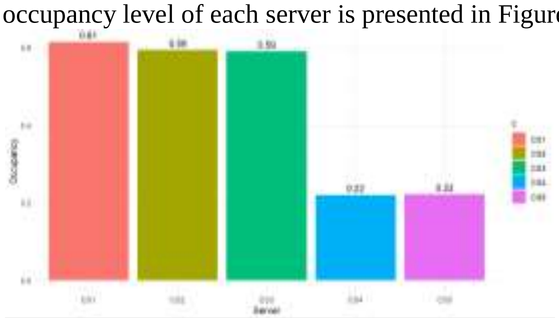


Figure 6. Occupancy Level of Each Server

Based on Figure 6, CS1 recorded the highest occupancy level at 61%, indicating that it was the busiest server during the observation period. Similarly, CS2 and CS3 exhibited relatively high occupancy levels of 59% each, suggesting that these three servers handled a larger share of customer requests than the remaining servers.

In contrast, CS4 and CS5 each recorded an occupancy level of 22%, indicating that they were utilized considerably less frequently. This difference suggests that the service workload was not evenly distributed among all customer service representatives.

The variation in server occupancy is consistent with the queueing performance analysis, particularly during Period P2, when only three servers were active. During this period, the higher workload assigned to CS1, CS2, and CS3 contributed to increased server utilization and longer customer waiting times.

Overall, the server occupancy analysis indicates that changes in the number of active servers under the dynamic server policy directly affect the workload distribution among customer service representatives and influence the overall performance of the queueing system.

4.9 Dynamic Servers Simulation Results

Based on the queueing performance analysis and server occupancy evaluation, the number of active servers was found to have a significant impact on system performance. During Period P2, when only three servers were operating, the system experienced the highest server utilization, accompanied by substantially longer queues and customer waiting times. In contrast, Periods P1 and P3 showed relatively low server occupancy, indicating that some service capacity remained underutilized.

These findings suggest that maintaining a fixed number of active servers throughout each observation period does not effectively accommodate fluctuations in customer demand. Therefore, a dynamic server approach is proposed, allowing the number of active servers to be adjusted according to the current queue conditions. This approach is expected to improve service efficiency while maintaining an acceptable level of customer service.

To evaluate the effectiveness of this approach, a dynamic server simulation was conducted using the actual arrival and service-time data of 73 customers. Unlike the existing system, where the number of active servers remained fixed during each observation period, the simulation dynamically adjusted the number of active servers based on the queue length. The objective was to examine how adaptive server allocation affects customer waiting time, queue length, and server requirements compared with the existing service system.

The server allocation policy used in the simulation is presented in Table 6.

Table 6. Server Allocation Policy in the Dynamic Server Simulation

Number of Waiting Customers	Active Servers
0-1 customer	3 servers
2-3 customer	4 servers
≥4 customer	5 servers

Under this policy, the number of active servers automatically increases or decreases in response to changes in the queue length. Consequently, the service capacity is expected to respond more effectively to variations in customer arrivals while improving resource utilization.

4.9.1 Dynamic Server System Performance

The performance of the dynamic server simulation is presented in Table 7.

Table 7. Dynamic Server System Performance

Indicator	Value
Average customer waiting time (W_q)	0.643 minutes
Maximum waiting time	7.033 minutes
Average queue length	0.192 customer
Maximum queue length	2 customers
Average number of active servers	3.137 server

The simulation results show that the average customer waiting time was 0.643 minutes (approximately 38 seconds), indicating that most customers received service with minimal delay. Furthermore, the average queue length was only 0.192 customers, while the maximum queue length reached 2 customers, suggesting that queue congestion remained low throughout the simulation.

The average number of active servers was 3.137, indicating that the system did not require all five servers to operate continuously. Additional servers were activated only when the queue length exceeded the predefined thresholds, allowing service capacity to adapt to fluctuations in customer demand. As a result, the dynamic server policy maintained a low waiting time and short queues while improving the efficiency of server utilization.

4.9.2 Dynamic Server System Utilization

The utilization of each server in the dynamic server system is presented in Table 8.

Table 8. Dynamic Server System Utilization

Server	Utilization
Server 1	68.6%
Server 2	78.7%
Server 3	63.5%
Server 4	13.3%

Based on Table 8, Servers 1, 2, and 3 served as the primary servers, each achieving utilization levels above 60%. In contrast, Server 4 recorded a utilization of 13.3%, indicating that it was activated only when queue conditions required additional service capacity. Server 5 was not activated during the simulation because the queue length never reached the predefined threshold.

These results indicate that the dynamic server policy utilized service resources efficiently by activating additional servers only when needed. On average, only 3.137 servers were active while maintaining an average customer waiting time of 0.643 minutes, demonstrating that the system was able to balance service quality and resource utilization effectively.

4.9.3 Comparison Between the Actual System and the Dynamic Server System

Since the actual system was observed during three operational periods with different characteristics (P1, P2, and P3), representative performance values were calculated using a weighted average based on the number of customers served in each period. This approach ensures that each period contributes proportionally to the overall performance.

The weighted average customer waiting time was calculated as:

$$W_q = \frac{27(0.4753) + 19(11.7377) + 27(0.1137)}{73}$$

$\approx$ 3.27 minutes

Similarly, the weighted average queue length was calculated as:

$$L_q = \frac{27(0.2173) + 19(3.8617) + 27(0.0566)}{73}$$

$\approx$ 1.11 customers

The weighted average number of active servers was calculated as:

$$c = \frac{27(5) + 19(3) + 27(5)}{73}$$

$\approx$ 4.48 servers

These values were used to represent the performance of the actual system and were compared with the dynamic server simulation, as shown in Table 9.

Table 9. Comparison of the Actual System and the Dynamic Server System

Indicator	Actual System	Dynamic Server
Average number of active servers	4.48	3.14
Average queue length (L_q)	1.11 customers	0.19 customer
Average waiting time (W_q)	3.27 minutes	0.64 minute

The comparison shows that the dynamic server policy improved both service efficiency and system performance. The average number of active servers decreased from 4.48 to 3.14, while the average queue length was reduced from 1.11 to 0.19 customers and the average waiting time from 3.27 to 0.64 minutes. These findings indicate that adjusting the number of active servers according to queue conditions allows the system to use service capacity more efficiently while maintaining a high level of service quality.

5. Conclusion

Based on the queueing system analysis at the IM3 Kayoon Surabaya Service Center, customer arrivals and service times were found to follow exponential distributions, indicating that the system can be modeled as an M/M/c queue with a First Come First Served (FCFS) service discipline. All observation periods satisfied the steady-state condition, indicating that the available service capacity was sufficient to accommodate customer arrivals. However, differences in the number of active servers across operational periods resulted in variations in system performance, while the server occupancy analysis suggested that server utilization was not evenly distributed.

The dynamic server simulation demonstrated that adjusting the number of active servers according to queue conditions improved both operational efficiency and service performance. Compared with the actual system, the proposed approach required fewer active servers while reducing customer waiting time and queue length. These findings indicate that a dynamic server policy can allocate service capacity more effectively according to customer demand, making it a practical strategy for improving resource utilization without compromising service quality.

6. ACKNOWLEDGMENTS

The author would like to thank Airlangga University, particularly the Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, for the academic support and facilities provided during the course of this research. The author would also like to thank the IM3 Kayoon outlet in Surabaya for granting permission to collect observational data.

7. REFERENCES

[1] N. irni N. Akinah, F. S. Titong, and M. A. Alhasni, “Analisis Sistem Antrian pada Stasiun Pengisian Bahan Bakar Umum (Studi Kasus: SPBU Batui II),” *Jurnal Teknik Industri Terintegrasi*, vol. 9, no. 1, pp. 883–891, Jan. 2026, doi: 10.31004/jutin.v9i1.54743.

[2] O. Huda Nur Qodzbari and D. Andesta, “ANALISIS SIMULASI MODEL PADA SISTEM ANTRIAN BENGKEL MOTOR DI CV. XYZ DENGAN SIMULASI ARENA,” *JUSTI (Jurnal Sistem dan Teknik Industri)*, vol. 4, no. 1, p. 91, Oct. 2023, doi: 10.30587/justicb.v4i1.6715.

[3] Y. Setiawannie, “Analisis Sistem Antrean Pengisian Bahan Bakar di SPBU dengan Model Multi Channel Single Phase,” *Jurnal Teknik Industri Terintegrasi*, vol. 8, no. 3, pp. 2493–2503, Jul. 2025, doi: 10.31004/jutin.v8i3.46144.

[4] Fuad Dwi Hanggara and R. D. E. Putra, “Analisis Sistem Antrian Pelanggan SPBU Dengan Pendekatan Simulasi Arena,” *Jurnal INTECH Teknik Industri Universitas Serang Raya*, vol. 6, no. 2, pp. 155–162, Dec. 2020, doi: 10.30656/intech.v6i2.2543.

[5] A. B. Sulistyo, I. Rafauzi, and H. Wijaya, “Analisis Tingkat Kepuasan Konsumen Terhadap Pelayanan Indomaret dengan Metode Servqual dan Importance Performance Analysis (IPA),” *Jurnal Media Teknik dan Sistem Industri*, vol. 6, no. 2, p. 95, Sep. 2022, doi: 10.35194/jmtsi.v6i2.1938.

[6] A. L. F. Afifa, Pradita Eko Prasetyo Utomo, and Hasanatul Iftitah, “ANALISIS EFEKTIVITAS PENAMBAHAN SERVER DI WAKTU TERTENTU PADA SISTEM ANTRIAN TOKO PUPUK SUMBER TANI: PERALIHAN DARI MODEL M/M/1 KE M/M/2 UNTUK MENINGKATKAN KECEPATAN PELAYANAN DAN MENGURANGI ANTRIAN,” *Journal of Management and Innovation Entrepreneurship (JMIE)*, vol. 2, no. 2, pp. 1926–1933, Jan. 2025, doi: 10.70248/jmie.v2i2.1708.

[7] A. S. S. I. Margasiwi and H. Purnomo, “Optimalisasi Antrean Pelayanan Customer Service Nasabah Bank Syariah Indonesia Kantor Cabang Kediri Trade Center,” *Widya Cipta: Jurnal Sekretari dan Manajemen*, vol. 6, no. 1, pp. 41–45, Feb. 2022, doi: 10.31294/widyacipta.v6i1.12170.

[8] Z. Inayah Atifah, Angelina Lahura, Aurelia Viesna Bota Werang, Nur Anita Chandra Putry, and Umi Wahidah, “Perlakuan Akuntansi Merger Padapt Indosat Ooredoo Hutchison,” *Jurnal Mahasiswa Manajemen dan Akuntansi*, vol. 5, no. 1, pp. 461–469, Mar. 2026, doi: 10.30640/jumma45.v5i1.5707.

[9] M. Mursyid, P. Purwanta, and S. Suliantoro, “PENGARUH KUALITAS PELAYANAN CUSTOMER SERVICE TERHADAP KEPUASAN PELANGGAN GRAPARI TELKOMSEL MADIUN,” *ALBAMA: JURNAL BISNIS ADMINISTRASI DAN MANAJEMEN*, vol. 16, no. 1, p. 78, Apr. 2023, doi: 10.56606/albama.v16i1.123.

[10] Indosat Ooredoo Hutchison, “32 Gerai IM3 Siap Layani Pelanggan Seluruh Indonesia,” ANTARA News. Accessed: May 25, 2026. [Online]. Available: https://sumut.antaranews.com/berita/535878/indosat-ooredoo-hutchison-hadirkan-32-gerai-im3-di-berbagai-kota

[11] D. Setiawati, B. P. Hartono, and M. Ahmad, “Analisis Sistem Antrean pada Customer Service Representative (CSR) Plasa Telkom Cilacap menggunakan Model Multi Channel Single Phase,” *Mathematical Sciences and Applications Journal*, vol. 5, no. 1, pp. 7–13, Dec. 2024, doi: 10.22437/msa.v5i1.37490.

[12] L. Kleinrock, *Queueing Systems, Volume 1: Theory*, 1st ed., vol. 1. New York: Wiley-Interscience, 1975.

[13] J. Sztrik, “Basic Queueing Theory,” Debrecen, 2012.

[14] C. W. Chan, M. Huang, and V. Sarhangian, “Dynamic Server Assignment in Multiclass Queues with Shifts, with Applications to Nurse Staffing in Emergency Departments,” *Oper. Res.*, vol. 69, no. 6, pp. 1936–1959, Nov. 2021, doi: 10.1287/opre.2020.2050.

[15] P. Guo and Z. G. Zhang, “Strategic Queueing Behavior and Its Impact on System Performance in Service Systems with the Congestion-Based Staffing Policy,” *Manufacturing & Service Operations Management*, vol. 15, no. 1, pp. 118–131, Feb. 2013, doi: 10.1287/msom.1120.0406.

[16] P. R. Sirait and P. Gultom, “Analisis Sistem Antrian pada Bank Negara Indonesia Kantor Cabang Kawasan Industri Medan,” *EduMatSains : Jurnal Pendidikan, Matematika dan Sains*, vol. 7, no. 2, pp. 292–304, Jan. 2023, doi: 10.33541/edumatsains.v7i2.4283.

[17] O. B. Jennings, A. Mandelbaum, W. A. Massey, and W. Whitt, “Server Staffing to Meet Time-Varying Demand,” *Manage. Sci.*, vol. 42, no. 10, pp. 1383–1394, Oct. 1996, doi: 10.1287/mnsc.42.10.1383.

[18] S. Sugito and M. A. Mukid, “DISTRIBUSI POISSON DAN DISTRIBUSI EKSPONENSIAL DALAM PROSES STOKASTIK,” *MEDIA STATISTIKA*, vol. 4, no. 2, Dec. 2011, doi: 10.14710/medstat.4.2.113-120.

[19] P. O. Nzechukwu, *Internal Audit Practice from A to Z*. Auerbach Publications, 2016. doi: 10.1201/9781315369143.

[20] G. Sapkota and R. P. Ghimire, "Mathematical Analysis of M/M/C Vacation Queueing Model with a Waiting Server and Impatient Customers," *J. Math. Stat.*, vol. 18, no. 1, pp. 36–48, Jan. 2022, doi: 10.3844/jmssp.2022.36.48.

[21] E. P. Setiawan, H. Sukoco, and L. Harini, "SIMULASI PENERAPAN TEORI ANTRIAN DALAM PEMBATASAN PENGUNJUNG OBJEK WISATA," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 4, pp. 719–726, Dec. 2021, doi: 10.30598/barekengvol15iss4pp719-726.