American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-03, Issue-02, pp-63-72
www.ajer.org

Research Paper Open Access

Analysis of a queuing system in an organization (a case study of First Bank PLC, Nigeria)

[1]Dr. Engr. Chuka Emmanuel Chinwuko, [2]Ezeliora Chukwuemeka Daniel , [3]Okoye Patrick Ugochukwu, [4]Obiafudo Obiora J.

[1]Department of Industrial and Production Engineering, Nnamdi Azikiwe University Awka, Anambra State, Nigeria Mobile: 2348037815808,
[2]Department of Industrial and Production Engineering, Nnamdi Azikiwe University Awka, Anambra State, Nigeria Mobile: 2348060480087
[3]Department of ChemicalEngineering, Nnamdi Azikiwe University Awka, Anambra State, Nigeria Mobile: 2348032902484,
[4]Department of Industrial and Production Engineering, Nnamdi Azikiwe University Awka, Anambra State, Nigeria Mobile: 2347030444797,

***Abstract:* -** The analysis of the queuing system shows that the number of their servers was not adequate for the customer's service. It observed that they need 5 servers instead of the 3 at present. It suggests a need to increase the number of servers in order to serve the customer better.

***Key word:* -** *Queuing System, waiting time, Arrival rate, Service rate, Probability, System Utilization, System Capacity, Server*

I. INTRODUCTION

Queuing theory is the mathematical study of waiting lines, or queues [1]. In queuing theory a model is constructed so that queue lengths and waiting times can be predicted [1]. Queuing theory is generally considered a branch of operations research because the results are often used when making business decisions about the resources needed to provide service.

Queuing theory started with research by Agner Krarup Erlang when he created models to describe the Copenhagen telephone exchange [1]. The ideas have since seen applications including telecommunications,[2] traffic engineering, computing[3] and the design of factories, shops, offices and hospitals.[4][5]

Etymology of Queuing System: The word queue comes, via French, from the Latin cauda, meaning tail. The spelling "queuing" over "queuing" is typically encountered in the academic research field. In fact, one of the flagship journals of the profession is named Queuing Systems.

Application of Queuing Theory: The public switched telephone network (PSTN) is designed to accommodate the offered traffic intensity with only a small loss. The performance of loss systems is quantified by their grade of service, driven by the assumption that if sufficient capacity is not available, the call is refused and lost.[13] Alternatively, overflow systems make use of alternative routes to divert calls via different paths — even these systems have a finite traffic carrying capacity.[13]

However, the use of queuing in PSTNs allows the systems to queue their customers' requests until free resources become available. This means that if traffic intensity levels exceed available capacity, customer's calls are not lost; customers instead wait until they can be served.[14] This method is used in queuing customers for the next available operator.

A queuing discipline determines the manner in which the exchange handles calls from customers.[14] It defines the way they will be served, the order in which they are served, and the way in which resources are divided among the customers.[14][15] Here are details of four queuing disciplines:

First in first out: This principle states that customers are served one at a time and that the customer that has been waiting the longest is served first.[15]

Last in first out: This principle also serves customers one at a time; however the customer with the shortest waiting time will be served first.[15] Also known as a **stack**.
Processor sharing: Service capacity is shared equally between customers.[15]
Priority: Customers with high priority are served first.[15]
Queuing is handled by control processes within exchanges, which can be modeled using state equations.[14][15] Queuing systems use a particular form of state equations known as a Markov chain that models the system in each state.[14] Incoming traffic to these systems is modeled via a Poisson distribution and is subject to Erlang's queuing theory assumptions viz.[13]

- Pure-chance traffic – Call arrivals and departures are random and independent events.[13]
- Statistical equilibrium – Probabilities within the system do not change.[13]
- Full availability – All incoming traffic can be routed to any other customer within the network.[13]
- Congestion is cleared as soon as servers are free.[13]

Classic queuing theory involves complex calculations to determine waiting time, service time, server utilization and other metrics that are used to measure queuing performance.[14][15]
Queuing networks: Networks of queues are systems a number of queues are connected by customer routing. When a customer is serviced at one node it can join another node and queue for service, or leave the network. For a network of m the state of the system can be described by an m–dimensional vector $(x_1,x_2,...,x_m)$ where x_i represents the number of customers at each node. The first significant results in this area were Jackson networks, for which an efficient product-form stationary distribution exists and the mean value analysis which allows average metrics such as throughput and sojourn times to be computed.[16]
If the total number of customers in the network remains constant the network is called a closed network and has also been shown to have a product–form stationary distribution in the Gordon–Newell theorem. This result was extended to the BCMP network where a network with very general service time, regimes and customer routing is shown to also exhibit a product-form stationary distribution.
Networks of customers have also been investigated; Kelly networks where customers of different classes experience different priority levels at different service nodes.[17]
Mean field limits: Mean field models consider the limiting behavior of the empirical measure (proportion of queues in different states) as the number of queues (m above) goes to infinity. The impact of other queues on any given queue in the network is approximated by a differential equation. The deterministic model converges to the same stationary distribution as the original model.[18]
Fluid limits: Fluid models are continuous deterministic analogs of queuing networks obtained by taking the limit when the process is scaled in time and space, allowing heterogeneous objects. This scaled trajectory converges to a deterministic equation which allows us stability of the system to be proven. It is known that a queuing network can be stable, but have an unstable fluid limit.[19]
Heavy traffic: In a system with high occupancy rates (utilization) a heavy traffic approximation can be used to approximate the queuing length process by a reflected Brownian motion,[20] Ornstein–Uhlenbeck process or more general diffusion process.[6] The number of dimensions of the RBM is equal to the number of queuing nodes and the diffusion is restricted to the non-negative orthant.
Queuing System Utilization: Utilization is the proportion of the system's resources which is used by the traffic which arrives at it. It should be strictly less than one for the system to function well. It is usually represented by the symbol ρ. If $\rho \geq 1$ then the queue will continue to grow as time goes on. In the simplest case of an M/M/1 queue (Poisson arrivals and a single Poisson server) then it is given by the mean arrival rate over the mean service rate, that is,

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

where λ is the mean arrival rate and μ is the mean service rate. More generally:

$$\rho = \frac{\lambda}{\mu \times c}$$

where λ is the mean arrival rate, μ is the mean service rate, and c is the number of servers, such as in an M/M/c queue.
In general, a lower utilization corresponds to less queuing for customers but means that the system is more idle, which may be considered inefficient.[7]
Role of Poisson process, exponential distributions
A useful queuing model represents a real-life system with sufficient accuracy and is analytically tractable. A queuing model based on the Poisson process and its companion exponential probability distribution often meets these two requirements. A Poisson process models random events (such as a customer arrival, a request for

action from a web server, or the completion of the actions requested of a web server) as emanating from a memoryless process. That is, the length of the time interval from the current time to the occurrence of the next event does not depend upon the time of occurrence of the last event. In the Poisson probability distribution, the observer records the number of events that occur in a time interval of fixed length. In the (negative) exponential probability distribution, the observer records the length of the time interval between consecutive events. In both, the underlying physical process is memoryless.
Models based on the Poisson process often respond to inputs from the environment in a manner that mimics the response of the system being modeled to those same inputs. Even a queuing model based on the Poisson process that does a relatively poor job of mimicking detailed system performance can be useful. The fact that such models often give "worst-case" scenario evaluations appeals to system designers who prefer to include a safety factor in their designs. Also, the form of the solution of models based on the Poisson process often provides insight into the form of the solution to a queuing problem whose detailed behavior is poorly mimicked. As a result, queuing models are frequently modeled as Poisson processes through the use of the exponential distribution. [8]
Limitations of queuing theory: The assumptions of classical queuing theory may be too restrictive to be able to model real-world situations exactly. The complexity of production lines with product-specific characteristics cannot be handled with those models. Therefore specialized tools have been developed to simulate, analyze, visualize and optimize time dynamic queuing line behavior. [9]
For example; the mathematical models often assume infinite numbers of customers, infinite queue capacity, or no bounds on inter-arrival or service times, when it is quite apparent that these bounds must exist in reality. Often, although the bounds do exist, they can be safely ignored because the differences between the real-world and theory is not statistically significant, as the probability that such boundary situations might occur is remote compared to the expected normal situation. Furthermore, several studies show the robustness of queuing models outside their assumptions. In other cases the theoretical solution may either prove intractable or insufficiently informative to be useful. [10]
Alternative means of analysis have thus been devised in order to provide some insight into problems that do not fall under the scope of queuing theory, although they are often scenario-specific because they generally consist of computer simulations or analysis of experimental data. See network traffic simulation. [9]

Research Method Used:
The research method used in this work is a quantitative research approach. The data gathered were the daily record of queuing system over a week. The method used in this research work were the analysis of queuing systems and techniques and also the development of queuing model for the analysis of queuing method and establish a method that will solve the problem of customers arrival rate. The model will establish the actual time it takes to serve the customer as at when due and estimate the actual working serves necessary in the organization. This model developed was used to predict the actual number of servers and time it takes to solve the problem of queuing or waiting before customers are been served as and at when due in the establishment for a week. The model developed was used to test the queuing system against the number of servers and customers arrival rate of the establishment.

Table 1: Day (One) 1 Queuing System Analysis of the Servers

Monday						
	Server 1		Server 2		Server 3	
Time	Arrival Rate	Service Rate	Arrival Rate	Service Rate	Arrival Rate	Service Rate
9:00-10:00am	16	5	17	10	25	13
10:00-11:00am	19	10	24	21	29	31
11:00-12:00 noon	20	19	28	17	37	28
12:00-1:00pm	19	19	24	24	29	29
1:00-2:00pm	14	11	19	17	26	24
2:00-3:00pm	4	4	14	13	20	14

Table 2: Day (Two) 2 Queuing System Analysis of the Servers

Tuesday						
	Server 1		Server 2		Server 3	
Time	Arrival Rate	Service Rate	Arrival Rate	Service Rate	Arrival Rate	Service Rate

9:00-10:00am	17	12	19	21	18	11
10:00-11:00am	25	20	31	23	27	24
11:00-12:00	31	27	31	25	37	29
12:00-1:00pm	30	24	24	24	29	29
1:00-2:00pm	36	35	31	23	26	24
2:00-3:00pm	19	18	19	14	14	8

Table 3: Day (Three) 3 Queuing System Analysis of the Servers

Wednesday						
	Server 1		Server 2		Server 3	
Time	Arrival Rate	Service Rate	Arrival Rate	Service Rate	Arrival Rate	Service Rate
9:00-10:00am	12	11	17	13	24	18
10:00-11:00am	19	15	24	21	29	31
11:00-12:00	29	24	28	24	27	26
12:00-1:00pm	23	23	23	24	27	25
1:00-2:00pm	17	13	21	17	26	23
2:00-3:00pm	10	7	12	13	17	13

Table 4: Day (Four) 4 Queuing System Analysis of the Servers

Thursday						
	Server 1		Server 2		Server 3	
Time	Arrival Rate	Service Rate	Arrival Rate	Service Rate	Arrival Rate	Service Rate
9:00-10:00am	9	6	19	14	27	23
10:00-11:00am	18	15	24	21	29	26
11:00-12:00	22	19	29	23	28	25
12:00-1:00pm	20	21	25	23	31	28
1:00-2:00pm	12	9	21	15	26	23
2:30-3:00pm	8	6	15	11	17	16

Table 4: Day (Five) 5 Queuing System Analysis of the Servers

Friday						
	Server 1		Server 2		Server 3	
Time	Arrival Rate	Service Rate	Arrival Rate	Service Rate	Arrival Rate	Service Rate
9:00-10:00am	13	8	17	11	18	13
10:00-11:00am	20	16	17	16	28	25
11:00-12:00	23	24	29	23	28	24
12:00-1:00pm	20	21	26	22	23	29
1:00-2:00pm	11	9	20	17	22	21
2:00-3:00pm	9	9	11	7	13	11

Table 6: Daily Queuing System Analysis of the Servers

		Server1		Server2		Server3	
		Arrival Rate	Service Rate	Arrival Rate	Service Rate	Arrival Rate	Service Rate
Day 1 (Monday)	Total Arrival or Service Rate	92	68	126	102	166	139
	Average Arrival or Service Rate	15.3333333	11.33333	21	17	27.66667	23.16667
Day 2 (Tuesday)	Total Arrival or Service Rate	158	136	155	130	151	125
	Average Arrival or Service Rate	26.3333333	22.66667	25.83333	21.66667	25.16667	20.83333
Day 3 (Wednesday)	Total Arrival or Service Rate	110	93	125	112	150	136
	Average Arrival or Service Rate	18.3333333	15.5	20.83333	18.66667	25	22.66667
Day 4 (Thursday)	Total Arrival or Service Rate	89	76	133	107	158	139
	Average Arrival or Service Rate	14.8333333	12.66667	22.16667	17.83333	26.33333	23.16667
Day 5 (Friday)	Total Arrival or Service Rate	96	87	120	96	132	123
	Average Arrival or Service Rate	16	14.5	20	16	22	20.5
Total for the Week	Average Total Arrival or Service Rate	545	460	659	547	757	662
	Average System utilization	1.18478261		1.2047532		1.14350453	

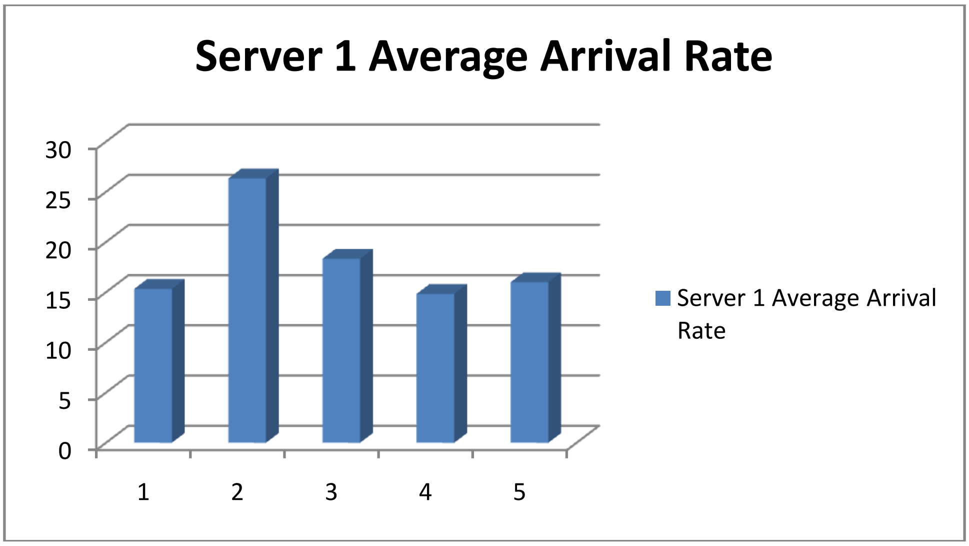

Figure 1: Server 1 Average Arrival Rate

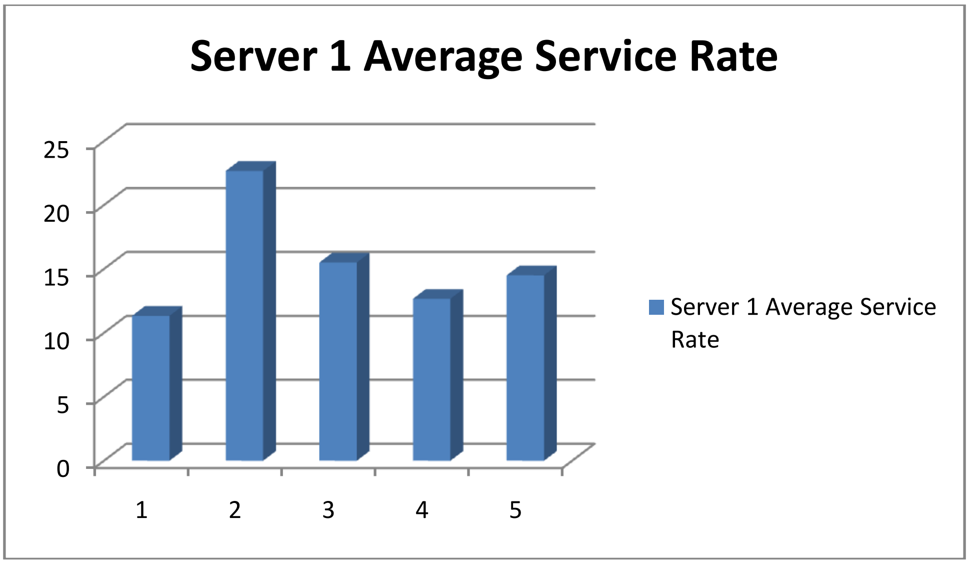

Figure 2: Server 1 Average Service Rate

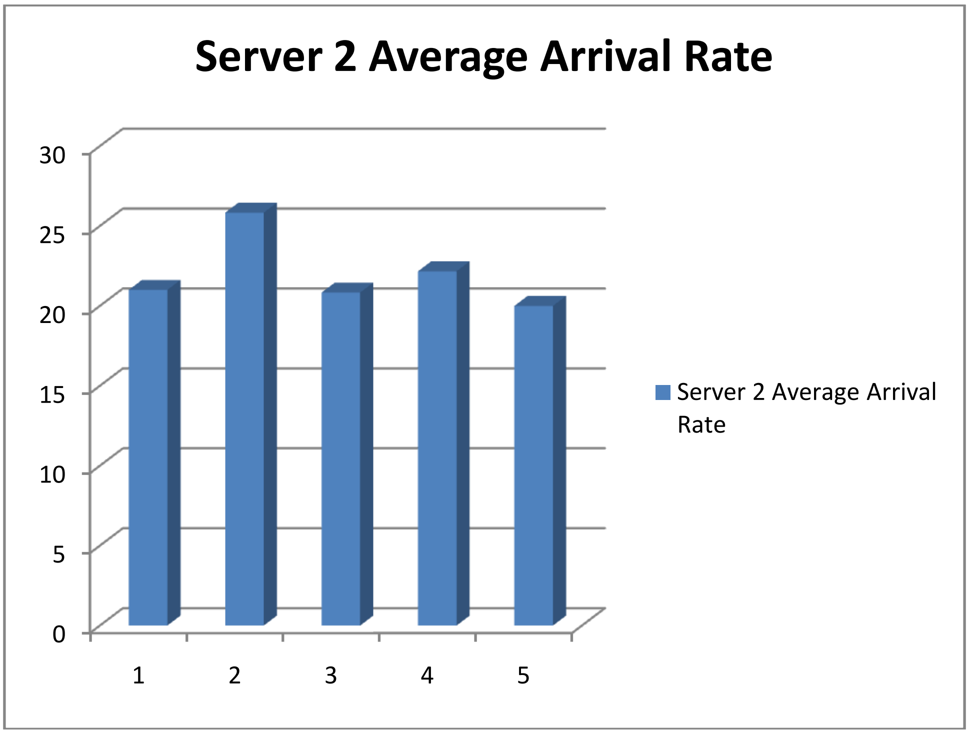

Figure 3: Server 2 Average Arrival Rate

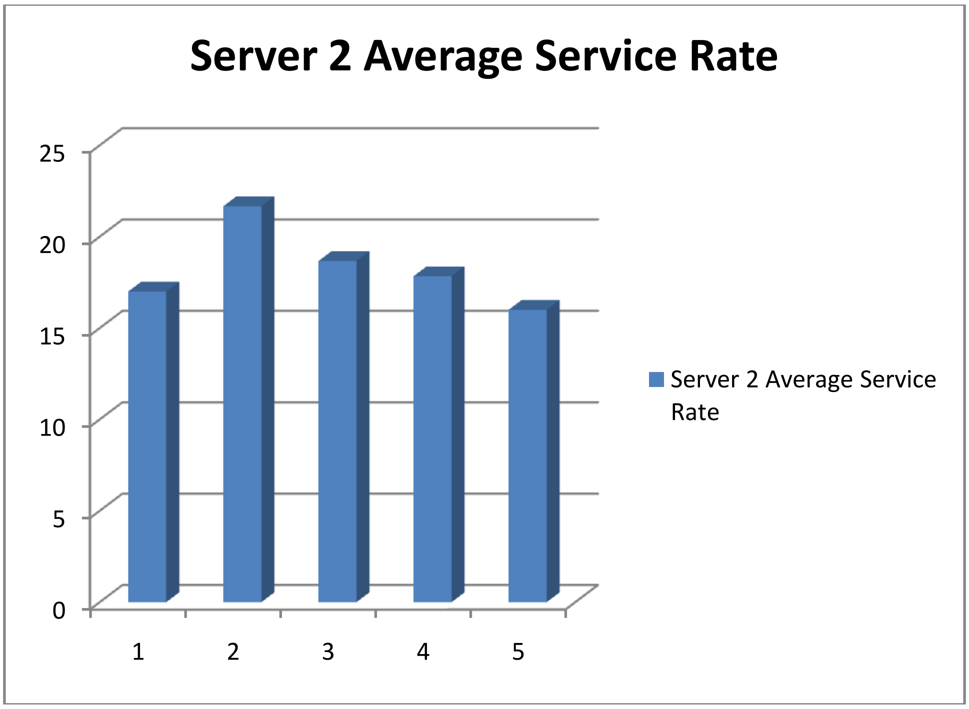

Figure 4: Server 2 Average Service Rate

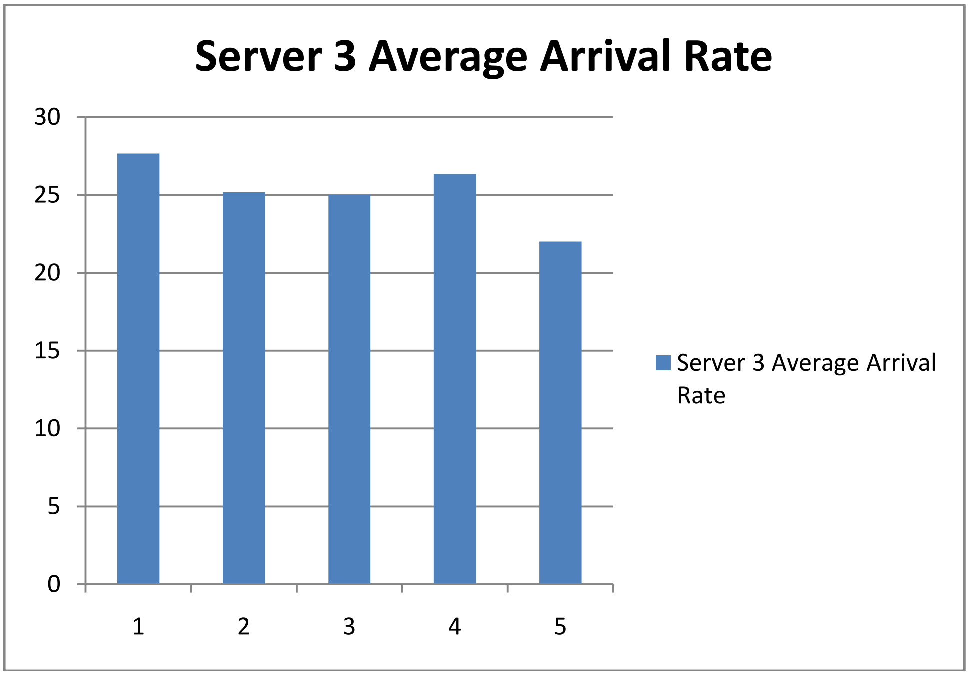

Figure 5: Server 3 Average Arrival Rate

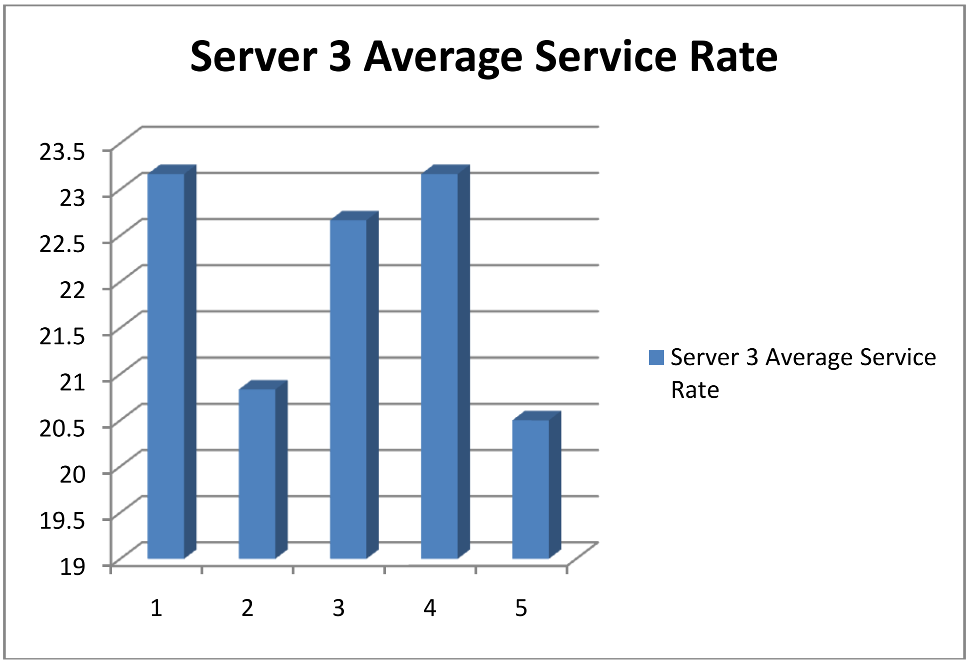

Figure 6: Server 3 Average Service Rate

Table 7: Daily System Utilization for each Server

Daily Record	Server 1	Server 2	Server 3
Day 1	1.352941	1.235294	1.194245
Day 2	1.161765	1.192308	1.208
Day 3	1.182796	1.116071	1.102941
Day 4	1.171053	1.242991	1.136691
Day 5	1.103448	1.25	1.073171

Customers arrival rate for server 1(λ_1) = 18.16667
Customers arrival rate for server 2 (λ_2) = 21.96667
Customers arrival rate for server 3 (λ_3) = 25.23333

Average Customers arrival rate for the servers (λ) = 21.7888889
Service Rate for Server 1(μ_1) = 15.33333
Service Rate for Server 2(μ_2) = 18.23333
Service Rate for Server 3(μ_3) = 22.06667
Average Service Rate for the Servers (μ) = 18.54444
The average number of customers being served(r)

$$R = \frac{\lambda}{\mu} \quad (1)$$

The average number of customers being served in server 1(r_1)

$$R_1 = \frac{\lambda_1}{\mu_1} = \frac{18.16667}{15.33333} = 1.184783 \quad (2)$$

$$R_2 = \frac{\lambda_2}{\mu_2} = \frac{21.96667}{18.23333} = 1.204753 \quad (3)$$

$$R_3 = \frac{\lambda_3}{\mu_3} = \frac{25.23333}{22.06667} = 1.184783 \quad (4)$$

$$R = \frac{\lambda}{\mu} = \frac{21.7888889}{18.54444} = 1.174955 \quad (5)$$

System Utilization for each Channel

$$\rho = \frac{\lambda}{M(\mu)} \quad (6)$$

$$\rho_1 = \frac{\lambda_1}{M_1(\mu_1)} = \frac{18.16667}{1(15.333337)} = 1.184783 \quad (7)$$

$$\rho_2 = \frac{\lambda_2}{M_2(\mu_2)} = \frac{21.96667}{1(18.23333)} = 1.204753 \quad (8)$$

$$\rho_3 = \frac{\lambda_3}{M_3(\mu_3)} = \frac{25.23333}{1(22.06667)} = 1.143505 \quad (9)$$

Average number in line

$$L_q = \frac{\lambda\mu\left(\frac{\lambda}{\mu}\right)^M}{(M-1)!(M\mu-\lambda)^2} P_0 \quad (10)$$

Probability of zero units in the system(P_0)

$$(P_0) = \left[\sum_{n=0}^{M-1} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} + \frac{\left(\frac{\lambda}{\mu}\right)^M}{M!\left(1-\frac{\lambda}{M\mu}\right)}\right]^{-1} \quad (11)$$

Average waiting time for an arrival not immediately served (W_a)

$$(W_a) = \frac{1}{M\mu - \lambda} \quad (12)$$

Probability that an arrival will have to wait for service (P_w)

$$(P_w) = \frac{W_q}{W_a} \quad (13)$$

Using equations (10) and (11) above, we have the results in the table below

Table 7: Results of the three Servers analyses

M	L_Q	P_O
1	0	-0.18048
2	0.836633	0.351096
3	0.13863	0.48456
4	0.050566	0.540665
5	0.025538	0.56806
6	0.015438	0.581313
7	0.015245	0.586285
8	0.007667	0.585774
9	0.00595	0.58128

Table 8: Formulas and Analysis of the Multiple Servers Queuing System

$\frac{\lambda}{\mu}$	M	L_q	P_0	$\frac{\lambda}{\mu}$	M	L_q	P_0
1	2	0.5	0.5	2.1	3	0.968927	0.08071
	3	0.083333	0.666667		4	0.305703	0.081065
	4	0.027778	0.75		5	0.157543	0.061793
	5	0.0125	0.8		6	0.098305	0.041509
	6	0.006667	0.833333		7	0.067568	0.025735
1.1	2	0.672222	0.409091	2.2	3	1.179965	0.064475
	3	0.112625	0.555393		4	0.348496	0.065728
	4	0.039775	0.623107		5	0.175457	0.04853
	5	0.019246	0.66096		6	0.107838	0.031214
	6	0.011106	0.684208		7	0.073286	0.018462
1.2	2	0.9	0.333333	2.3	3	1.45087	0.050809
	3	0.148148	0.462963		4	0.395693	0.053301
	4	0.054518	0.515312		5	0.194224	0.038258
	5	0.025871	0.569136		6	0.117566	0.023635
	6	0.017071	0.548823		7	0.079051	0.013379
	7	0.011719	0.550119	2.4	3	1.811321	0.039308
1.3	2	1.207143	0.269231		4	0.448133	0.043223
	3	0.190613	0.38575		5	0.213999	0.03028
	4	0.072004	0.424122		6	0.127566	0.018023
	5	0.038266	0.434123		7	0.084913	0.009794
	6	0.024414	0.429741	2.5	3	2.314815	0.02963
1.4	2	1.633333	0.214286	ANS	4	0.506894	0.035036
	3	0.240871	0.321027		5	0.234962	0.02406
	4	0.092162	0.347519		6	0.137914	0.01384
	5	0.050189	0.345548		7	0.090917	0.007239
	6	0.032809	0.329288	2.6	3	3.068436	0.021487
1.5	2	2.25	0.166667		4	0.573376	0.028376
	3	0.3	0.266667		5	0.257319	0.019192
	4	0.114894	0.283688		6	1.528019	0.0107
	5	0.063347	0.272506		7	0.075056	0.005401
	6	0.04186	0.248062	2.7	3	4.322134	0.014639
1.6	2	3.2	0.125		4	0.649416	0.022946
	3	0.369408	0.22096		5	0.281304	0.015364
	4	0.14012	0.230911		6	0.159958	0.008326
	5	0.077461	0.213491		7	0.103502	0.004066
	6	0.05123	0.18474	2.8	3	6.825871	0.008884
1.7	2	4.816667	0.088235		4	0.737464	0.018511
	3	0.450982	0.182507		5	0.307187	0.012341
	4	0.167815	0.18757		6	0.171806	0.006519
	5	0.092314	0.166595		7	0.110151	0.003086
	6	0.06069	0.136736	2.9	3	14.32961	0.004052
1.8	2	8.1	0.055556		4	0.840863	0.014881
	3	0.547297	0.15015		5	0.335283	0.009943
	4	0.198026	0.152169		6	0.184311	0.005134
	5	0.107773	0.129788		7	0.117079	0.002361
	6	0.070127	0.101029	3	4	0.964286	0.011905
1.9	2	18.05	0.026316		5	0.365964	0.008032
	3	0.661938	0.122919		6	0.197561	0.004065
	4	0.230895	0.123369		7	0.124318	0.001819
	5	0.123781	0.101139		8	0.051965	0.00076
	6	0.079516	0.074769	3.1	4	0.111448	0.00946
2	3	0.8	0.1		5	0.039967	0.006503
	4	0.266667	0.1		6	0.021165	0.003235
	5	0.140351	0.078947		7	0.01319	0.001411
	6	0.088889	0.055556		8	0.009022	0.000573
	7	0.061836	0.036232		9	0.006563	0.000223

Expected inter arrival time per hour $\frac{1}{\lambda} = \frac{1}{21.7888889} \times 60 = 2.753697$ minutes (14)

Service Time per hour $= \mu = \frac{1}{3.24 \div 60} = 18.54444$ (15)

From the table above: M=5, L_q= 0.025871=0.026 and P_0= 0.569136= 0.569

The average number of customers waiting for service (L_q) = 0.025871=0.026

Average waiting time for an arrival not immediately served (W_a)

$$(W_a) = \frac{1}{M\mu - \lambda} \quad (16)$$

$$(W_a) = \frac{1}{5(18.54) - 21.79} = 0.014 \text{ Hour or } 0.847 \text{ minutes}$$

The average time customers wait in line (W_q)

$$(W_q) = \frac{L_q}{\lambda} = \frac{0.026}{21.79} = 0.00119 hour\ or\ 0.072\ minutes$$

Probability that an arrival will have to wait for service (P_w)

$$(P_w) = \frac{W_q}{W_a} = \frac{0.00119}{0.0.014} = 0.085 \quad (17)$$

The Average Number of Customers in the System (waiting and /or being served)

$$L_S = L_q + R \quad (18)$$

Or $L_S = W_s \times \lambda$ (19)

Using Equation (32) above;

$$L_S = 0.026 + 1.2 = 1.226$$

The average time spend in the system (waiting in line and service time) (W_s)

$$W_s = W_q + \frac{1}{\mu} = \frac{L_s}{\lambda} \quad (20)$$

$$W_s = \frac{L_s}{\lambda} = \frac{1.226}{21.79} = 0.056$$

System Utilization $\rho = \frac{\lambda}{M(\mu)} = \frac{21.79}{5(18.54)} = 0.235$ (21)

The system capacity = Mμ = $5 \times 18.54 = 92.7$ (22)

II. DISCUSSION OF RESULTS

From the analysis, it was observed that number of servers necessary to serve the customers in the case study establishment was five (5) servers (or channels). This was proved in table 7 and 8 above. This is the appropriate number of servers that can serve the customers as and at when due without waiting for long before customers are been served at the actual time necessary for the service. This increase in servers reduces the waiting time, and the probability that an arrival will have to wait for service is 0.056. However, the system utilization was observed to be 0.235 for an hour. Furthermore, the system capacity of the five servers was observed to be 92.7 for an hour.

III. CONCLUSION

The evaluation of queuing system in an establishment is necessary for the betterment of the establishment. As it concerns the case study company, the evaluation or analysis of their queuing system shows that the case study company needs to increase the number of their channels or servers up to five(5) as show in the result analysis. The increase in the number of servers will reduce the time customers have to wait in line before been served. This will also increase the efficiency of the establishment due to the appreciation in their serve to the customers as and at when due.

REFERENCES

[1] Sundarapandian, V. (2009). "7. Queuing Theory". Probability, Statistics and Queuing Theory. PHI Learning. ISBN 8120338448.

[2] Lawrence W. Dowdy, Virgilio A.F. Almeida, Daniel A. Menasce (Thursday Janery 15, 2004). "Performance by Design: Computer Capacity Planning By Example". p. 480

[3] Schlechter, Kira (Monday March 02, 2009). "Hershey Medical Center to open redesigned emergency room". The Patriot-News

[4] Mayhew, Les; Smith, David (December 2006). Using queuing theory to analyse completion times in accident and emergency departments in the light of the Government 4-hour target. Cass Business School. ISBN 978-1-905752-06-5. Retrieved 2008-05-20.

[5] Tijms, H.C, Algorithmic Analysis of Queues", Chapter 9 in A First Course in Stochastic Models, Wiley, Chichester, 2003

[6] 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. doi:10.1214/aoms/1177728975. JSTOR 2236285. edit

[7] http://pass.maths.org.uk/issue2/erlang/index.html
[8] Asmussen, S. R.; Boxma, O. J. (2009). "Editorial introduction". Queuing Systems 63: 1. doi:10.1007/s11134-009-9151-8. edit
[9] "The theory of probabilities and telephone conversations". Nyt Tidsskrift for Matematik B 20: 33–39. 1909.
[10] Kingman, J. F. C. (2009). "The first Erlang century—and the next". Queuing Systems 63: 3–4. doi:10.1007/s11134-009-9147-4. edit
[11] Whittle, P. (2002). "Applied Probability in Great Britain". Operations Research 50: 227–177. doi:10.1287/opre.50.1.227.17792. JSTOR 3088474. edit
[12] Flood, J.E. Telecommunications Switching, Traffic and Networks, Chapter 4: Telecommunications Traffic, New York: Prentice-Hall, 1998.
[13] Bose S.J., Chapter 1 - An Introduction to Queuing Systems, Kluwer/Plenum Publishers, 2002.
[14] Penttinen A., Chapter 8 – Queuing Systems, Lecture Notes: S-38.145 - Introduction to Teletraffic Theory.
[15] Van Dijk, N. M. (1993). "On the arrival theorem for communication networks". Computer Networks and ISDN Systems 25 (10): 1135–2013. doi:10.1016/0169-7552(93)90073-D. edit
[16] Kelly, F. P. (1975). "Networks of Queues with Customers of Different Types". Journal of Applied Probability 12 (3): 542–554. doi:10.2307/3212869. JSTOR 3212869. edit
[17] Bobbio, A.; Gribaudo, M.; Telek, M. S. (2008). "Analysis of Large Scale Interacting Systems by Mean Field Method". 2008 Fifth International Conference on Quantitative Evaluation of Systems. p. 215. doi:10.1109/QEST.2008.47. ISBN 978-0-7695-3360-5. edit
[18] Bramson, M. (1999). "A stable queuing network with unstable fluid model". The Annals of Applied Probability 9 (3): 818. doi:10.1214/aoap/1029962815. edit
[19] Chen, H.; Whitt, W. (1993). "Diffusion approximations for open queuing networks with service interruptions". Queuing Systems 13 (4): 335. doi:10.1007/BF01149260. edit
[20] Yamada, K. (1995). "Diffusion Approximation for Open State-Dependent Queuing Networks in the Heavy Traffic Situation". The Annals of Applied Probability 5 (4): 958. doi:10.1214/aoap/1177004602. JSTOR 2245101