A Simulation of the Optimal Personnel Demand for Banking Hall Services

David Oriedi*, Kennedy Malanga, George Musumba, Gabriel Kamau, and Mosses Ollengo

***Abstract*—The banking industry has seen a revolutionary moment in adopting mobile banking technology in the last few years. The technology has contributed to many customers keeping off the banking hall for basic and routine transactions. This has largely been attributed to the convenience of mobile banking, including time management and increased privacy. However, there are still times when a bank conducts large-scale service provision for many people. For example, a payment agency contracted by the government to disburse funds to a section of the citizens may require that money be channeled only through a particular bank. If the period for such a requirement is limited, there is likely to be pressure on both the bank and the would-be customers because of long queues. For cases like that, there is a need to simulate possible scenarios and plan to avoid delays and frustrations for both parties. This work proposes a queue simulation model that can be used to forecast the number of bank staff that can be deployed for such a scenario vis a vis the expected number of customers seeking service. Results have shown that the simulation model can help the bank management make optimal choices that will avoid waste in the form of human or financial resources.**

***Index Terms*—Bank, model, simulation, model, queue**

I. Introduction

Banking hall services have been around since the introduction of banking services. A typical banking hall service framework constitutes a waiting area, service points, and an array of services, including deposit of funds, local funds transfer, funds withdrawal, account opening, forex services, and funds transfer across borders. The banking sector has recently adopted service mobility through mobile banking technology [1]. Key drivers for faster adoption of mobile banking technology in the banking service sector include security, convenience, and reduction in the time spent on queuing by customers [2]. These factors have resulted in a sudden surge in banking services that have seen customers opting to get services away from the banking hall.

While mobile banking technology has redefined the banking hall experience for customers, it has not eliminated the queue for customers who choose to physically visit the banking hall for banking services [3]. It, therefore, means that as much as the preferred means of bank transactions by most customers is through mobile banking technology, there are still queues at the banking hall. However, there has always been a concern regarding the time customers spend queuing for services at the banking hall [4]. Among others, computer simulation of queuing systems has emerged as one of the ways through which researchers have tried to answer the question of queue management and optimal teller service discharge to help reduce the time customers spend waiting for services at the banking hall.

Simulation of queuing systems in the banking sector is not a recent idea. Among the early works involved in such simulations are [5] and [6], in which a case was made for the need to simulate a banking queuing system to improve the customer experience while queuing for services in the banking hall. Early works like [7] and [8] introduced the idea of simulating banking queuing systems to address the concerns about waiting times at the banking hall. Progressively, multiple works have been dedicated to modeling and simulating queue systems for the banking hall scenario. These works address different aspects of the evolving nature of banking services and the type of queues they attract. For example, while the traditional banking hall queue has been largely based on a first-come-first-served service discipline, requiring the customer to be physically present at a particular queue in the banking hall until served, the current queue system adopted in most banks is automated and does not require the customer to be tied to a specific physical queue. In such automated queue systems, an auto-generated service number determines when the customer will be served, with the advantage of all service disciplines incorporated. Customers are also able to see the service numbers conspicuously displayed on screens. Given that the customer will be seated while waiting to be served in most cases, this system has been more convenient for the customer than the traditional queuing system.

Several research works address the need to make the customer queues in the banking hall easily manageable and probably predictable. For example, the works in [2, 9], and [9] have all been dedicated to figuring out how to improve the experience of the customer while at the banking hall. In most cases, the emphasis has been placed on reducing the time the customer spends waiting for service.

While most simulation works are done to improve the customer's experience in the banking hall, not many research

Manuscript received December 3, 2022; revised January 9 2023; accepted February 25, 2023; published May 10, 2023.

David Oriedi, George Musumba, George Musumba, and Gabriel Kamau are with Department of Computer Science, Dedan Kimathi University of Technology, Nyeri, Kenya, E-mail: kmalanga@kyu.ac.ke (K.M.), george.musumba@dkut.ac.ke (G.M.), gabriel.kamau@dkut.ac.ke (G.K.).

Kennedy Malanga is with Department of Information Technology, Kirinyaga University, Nyeri, Kenya, E-mail: kmalanga@kyu.ac.ke (K.M.)

Mosses Ollengo is with Department of Chemistry, Dedan Kimathi University of Technology, Nyeri-Kenya; E-mail: moses.ollengo@dkut.ac.ke (M.O.).

*Correspondence: david.opondo@dkut.ac.ke (D.O.)

DOI: 10.7763/IJMO.2023.V13.826

works have paid attention to how bank managements deal with fluctuating personnel demands, especially during surges in customer service demand. And as [10] observes, there is a need to model and simulate banking queues that would help bank managers to predict necessary human resources to attend to peak times that come with huge demand surges for banking services. For example, consider a scenario where a government has ordered payments meant for a section of citizens to be channeled through a specific bank. An order like this can result in a sudden unforeseen spike in demand for opening accounts in such a bank. In such a case, it would be necessary for the bank management to prepare upfront for the human and financial resource implications of such a scenario. A simulation of the human and financial resources required to optimally meet the demands of such a surge in customer numbers will go a long way in helping the bank management to plan. This work seeks to simulate a scenario in which the bank management needs to reliably predict the human resource and financial cost implications for demands brought about by a sudden but episodic increase in the number of customers that will require services from the bank. This is an aspect that would contribute positively to the planning needs of the bank.

The rest of this paper is organized as follows: Section I introduces the paper, and Section II summarizes related works. In Section III, we detail the description and the working of our model, while experiments and results are discussed in Section IV. Finally, the paper concludes in Section V with suggestions on open issues for future research.

II. Related Works

Most simulations of queuing systems in banks are intended to find ways of improving the queue experience for the bank customer. This puts more emphasis on the customer's well-being while in the queue. To address the frustrations coming from customers in the Bangkok banking sector, a model was proposed in [11] that substantially reduced the amount of time spent in the queue by customers the whole day in a banking hall. The authors segmented the day into morning and afternoon sessions. They noted that the queues were longer in the morning than in the afternoon and recommended adding service channels during morning hours. In [12], the length of time that a customer spends in the queue was used as a major determinant in designing an automated queuing system in which a machine was used to assign service numbers to customers, and appropriate service discipline was followed. The model recommended changing from multi-queues and multi-desks to single queues and multi-desks.

A model called Quick Pass was proposed by [13], in which an M/M/c queue was analyzed for an optimization strategy. Through the application of a greedy algorithm, this queue system was redesigned with a multiple-optimization model approach. The results saw the average waiting time of customers substantially drop. While proposing no specific model to be used in improving customer queue experience in a banking hall, [7] provided a critical look at the major performance indicators that influence the perception of a customer towards service quality. Through the simulation of the behavioral performance of the bank queuing systems, the authors proposed a criterion for the definition of service quality, in which customer waiting time was a major ingredient. Performance measures defined through such criteria are critical in building queue simulation models for performance improvement. In [2], an intelligent bank branch that uses Virtual Teller Machines and two queuing models, namely the mixed queuing model and the separated queuing model, were proposed for the intelligent bank branches. An intelligent bank is a banking approach that does not involve human tellers. After running simulations on the waiting time and work intensity, the research finding recommended a virtual bank branch over a traditional bank. Additionally, the researchers noted that the customer experience was greatly improved. Finally, the outcome of the simulations brought to the fore that the separated queuing model performs better than the mixed queuing model.

In some instances, the bank management gets concerned about the overall service utilization of the service points. This could happen when there are way too many service personnel than the customers that need service. This scenario may also call for an analysis of the kind of queues being seen in such a bank for possible analysis. A case in point is [14], in which a bank in Malaysia commissioned a simulation study to analyze how effectively the tellers served the customers. Through the analysis of the simulation outputs, the study recommended a reduction in the number of tellers for more optimal service. In [15], the behavior of customers depositing queuing to deposit money in the bank was simulated. The results appreciated how customer behavior while in the queue affects the quality of a queuing system. In other words, the customers' behavior can have either positive or negative consequences on the quality of service. In cases where there is likely to be heavy traffic in the queues, it is recommended in [16] that a predictable way of dealing with such a queue be devised; otherwise, it can easily lead to chaos in the queue, which can frustrate the customers' ability to wait for services.

In conclusion, it appears to be apparent that most of the effort that has been put into modeling and simulating studies of bank queuing systems seeks to improve the experience of the customer. This improvement varies from trying to reduce the time a customer spends in the queue while waiting for service to the time the customer takes to be served at the service point. It also introduces the redefinition of parameters that are used to gauge the quality of service for the customer. However, the simulation of bank queues to help bank management determine personnel demand for a sudden surge in service demand is not as widespread. This work seeks to simulate a bank queuing system to help the bank management plan for instances where demand for services spikes, especially in an unpredictable fashion.

III. Our Model

Our model is a discrete event simulation model. The simulation outputs are determined at separate points in time. We assume a typical banking scenario in which a customer walks in, requests for a service type, and, if there is a queue, proceeds to join the queue as they wait for their turn to be served. However, the assumption also includes a surge in customer numbers that comes suddenly but with human and financial demands in an unforeseen manner. For this purpose, the simulation model is proposed to help the management of a bank quickly forecast what it would mean to undertake such sudden service demand.

The proposed model works by providing different scenarios for personnel deployment based on the expected customer numbers. It is expected that upon simulating several times, the bank's management can settle on a compromise set of variable values to predict the cost of managing such a customer surge. The major inputs to the simulation model will be the number of employees, customer arrival rate, service rate, and the total simulation time. While most of these inputs are common to simulation experiments, considering the number of employees as input is specifically done when estimating how many employees would be needed based on the prevailing circumstances.

The model gives several outputs, including the departure time of a customer from the banking hall, the number of customers served at the end of the simulation, and the time the customer takes in the banking hall. Furthermore, the simulation model traces the time when each customer enters and leaves the banking hall. In this way, we can interrogate the quality of service given to the customer. In real life, it is expected that different customers take a different number of times while being served by the bank tellers. The simulation model is designed to represent this scenario. Accordingly, different customers leave the bank at different times, presumably due to the differences in the kind of services that they are seeking from the bank. Unique ids identify customers to differentiate them from one another. This is another real-life replica of what happens in the simulation model.

The model formulation is based on the works by [17] and [18] in which discrete event model formulations have been extensively described. The discrete event simulation is based on variable state changes that occur only when an event happens in the system. Such events in a banking hall include the arrival of a customer, the beginning and the end of service for a customer, and the departure of a customer from the bank. The formulation of the model is based on the Kendall notation in Eq. (1):

$$X/Y/Z/A/B/C \tag{1}$$

in which X and Y represent customer inter-arrival and service time distributions, respectively. Z represents parallel service stations in the system, A is the service capacity of the system, B is the maximum number of customers allowed in the system. In contrast, C represents the service discipline of the queue system.

$$M/M/c/\infty/\infty/FCFS \tag{2}$$

The banking queue simulation context is that shown in Eq. (2), in which the banking system has an infinite service capacity and population.

In such a system, the utilization of the server - the extent to which the system is productively engaged - is a significant indicator of how good the system is to the population served. The server utilization ρ is calculated by making use of the arrival rate λ and the service rate μ as shown in Equation (3):

$$\rho = \frac{\lambda}{c\mu} \tag{3}$$

The probability of idle service points indicates customers' absence in the queue or the system. This probability is indicated in Eq. (4), with n representing the number of customers.

$$P_0 = \left\{\left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!}\right] + \left[\frac{(c\rho)^c}{c!(1-\rho)}\right]\right\}^{-1} \tag{4}$$

As has already been explained, we simulate a scenario in which there is a sudden surge in the number of customers that the bank is expected to attend to. Under such circumstances, it may be necessary to determine the probability that all service points will be busy and for how long. Such a probability can be used to decide whether to add service points. The determination of this probability is expressed in Eq. (5).

$$P_\infty = \frac{(c\rho)^c P_0}{c!(1-\rho)} \tag{5}$$

Since the model aims to ensure that customers that turn up for service spend as little time as possible in the queue, it is important to know the parameters that act as queue performance indicators. Such parameters include the average length of the queue, the mean number of customers in the system, the average wait time in the system, and the average wait time in the queue. These parameters are summarized in Equations (6) to (9).

$$L_q = \frac{\rho P_\infty}{1-\rho} \tag{6}$$

$$L = c\rho + \frac{\rho P_\infty}{1-\rho} \tag{7}$$

$$W = \frac{L}{\lambda} \tag{8}$$

$$W_q = W - \frac{1}{\mu} \tag{9}$$

While these formulae are not hard to come by in literature on the simulation of queue systems, formulations incorporating the number of staff deployed to attend to the queue and the financial implications of such deployment are not as common. In this model, we introduce the cost factor for the staff serving the customer in the queue as a parameter of interest to the bank's management. Because there are likely to be workers on shift, we introduce the time the workers spend serving during a given shift. So, for a given service point c_i manned by a given staff for a period t_i, paid at a rate k, we compute the total cost T_{ci}, of having one staff attend to a queue during a shift as:

$$T_{ci} = kc_it_i \tag{10}$$

It, therefore, means that the service points c_1 , $c_2, \dots \dots, c_n$, collectively form a set C representing the set of service points. The total cost T_c, can then be expressed as shown in Equation (11).

$$T_c = k\{c_1t_1 + c_2t_2 + c_3t_3 \ldots . c_nt_n\} \quad (11)$$

We argue that Equation (11) is a necessary inclusion in the simulation model to give the management of a bank a prediction of the implications of engaging staff for a high peak service demand that is sudden. More specifically, there is now the option of varying the values of c_i depending on the expected service demand levels. So, for a given total customer population n, a service point can be assigned a threshold number of customers to serve within a unit of time. This will ensure that cases of the service points idling or deliberately performing below expected levels are reduced. The net result is that the service station's satisfactory service will be defined in terms of the total assigned customer numbers within a given unit time.

Therefore, we define a unit of service U_s, as the basic unit used to compute payment for the staff deployed to serve at a particular service point. In this way, the remuneration for the staff is pegged not only on the amount of time the staff has been on shift but, more importantly, on the number of customers served per unit time. A unit of service is expressed as shown in Equation (12).

$$U_s = \frac{n}{t} \quad (12)$$

The service unit then becomes a major determinant in the amount of money paid out to a member of staff that has been sent on such an assignment. For a payment rate p_r, that specifies a certain amount of money for service rendered; we define the payment for a unit of service p_s, as:

$$p_s = U_s \times p_r \quad (13)$$

Using the definition for the payment for a unit of service, Equation (11) can then be rewritten as follows:

$$T_c = p_s\{c_1t_1 + c_2t_2 + c_3t_3 \ldots . c_nt_n\} \quad (14)$$

The norm in the banking sector is that a single bank usually has several branches across regions. In most cases, this points to the bank's good performance in the market [19], while at other times, it shows the desire of the bank to try new and unfamiliar markets [20]. However, the common denominator in both cases is that staff must be deployed to attend to the would-be customers. We assume a similar scenario in this work. The assumption is that several branches of the bank exist. And the anticipated surge is across the bank's various branches. Consequently, there are members of staff deployed to attend to this surge across the branches. Given the number of branches β, we rewrite Equation (14) to reflect the total cost across the bank branches because of this scenario.

$$T_c = \beta . p_s\{c_1t_1 + c_2t_2 + c_3t_3 \ldots . c_nt_n\} \quad (15)$$

This work proposes Equation (15) as representative of the total cost of serving customers during such a demand surge in services rendered by the bank. A variation of the components of this equation can then give the bank a possible picture of the financial implications of attending to the demand surge at such times.

IV. Experiments and Results

The output of the simulation run presents an opportunity to compare different scenarios of staff deployment and the corresponding number of customers that can be served by the such deployment of staff as well as how the service points are utilized. As mentioned, the scenario assumed in this simulation is an anticipated surge in service demand by bank customers for whatever reason. It is also assumed that because of this expected surge, the bank does not depend on its regular service routine, which could easily lead to unexpected delays and frustration for the customers waiting for service. In addition, the demand surge is assumed to affect all the branches of such a bank. The anticipated scenario, therefore, means that several branches are experiencing this sudden service demand. It is, therefore, natural for the bank management to deploy staff across all the branches to handle the high service demand. The core of this article is about helping the bank management, through simulation, to project the financial implications of handling such a surge.

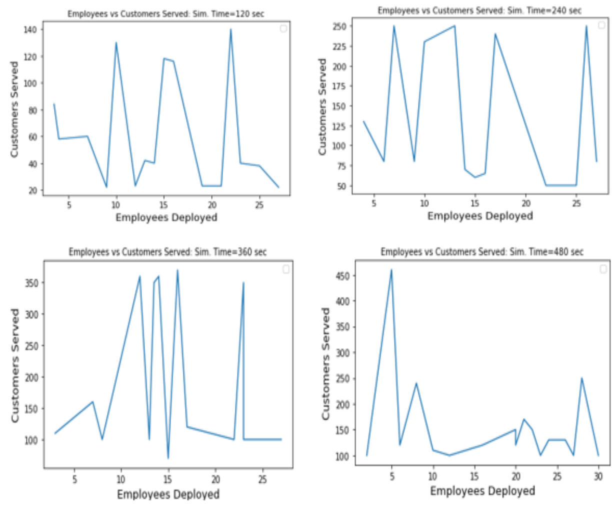

Fig. 1. Employees deployed vs. customers served.

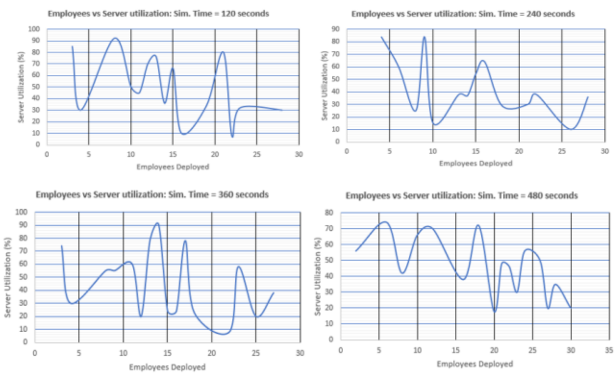

Fig. 2. Employees Deployed vs. Server Utilization.

To avoid a scenario where customers are frustrated through long queues and delayed services, the bank needs to plan for the surge in customer demand. This kind of planning would have financial implications for rewarding the staff on duty for attending to such a surge. This simulation model is intended to provide different scenarios for the financial implication of the deployment of staff on such duty. For example, the number of customers a given number of employees would handle is important information if the number of expected customers can be approximated in advance.

Fig. 1-Fig. 2 show the results of the simulation run. Each figure represents the outcome of the simulation run according to a stated simulation time. The simulation time varies from 120 seconds to 480 seconds in intervals of 120 seconds in each case. As expected, the simulation time represents the total time taken to serve all the customers in the system. As earlier proposed, this is not to be confused with the unit time required to serve a given number of customers. In other words, the simulation time represents the total time during which the simulation experiment was being run. As seen in Fig. 1, an increase in the number of employees deployed initially leads to an increase in the number of customers handled. However, it is notable that an increase in the number of employees does not infinitely lead to an increase in the number of customers served. The number of customers served keeps fluctuating, and even appears to stabilize as the simulation period continues to grow.

Similarly, this fluctuation is seen when the number of employees deployed is plotted against server utilization values in Fig 2. While the fluctuations could be partly attributed to the fact that this work is based on simulated values, bank management can easily use the trends to determine an optimal number of employees that would ensure financial savings. That is, the bank would use the information to avoid either under-staffing or over-staffing for a scenario in which there is a sudden surge in the number of customers needing services.

Moreover, it is notable that fewer employees can serve the same number of customers as more employees. For example, in Fig. 1, 10 employees can serve the same number of customers as 23 employees (t=120), or 8 employees can serve the same number of customers as 23 employees (t=360). With such information, it would be easier for the bank management to deploy reduced staff to serve the same number of customers as more staff. Such a strategy would be beneficial since the bank can save resources and avoid wasting human and financial resources.

V. Conclusion and Further Research

In this work, a simulation model has been used to predict the need for employee deployment for an unpredictable number of customers expected to require services from the bank at a time-of-service demand surge. The work has proposed a model that can be used to predict how the number of employees can be varied alongside service stations among several branches of a bank to determine the optimal number of staff and service stations to deploy.

The variations noted in the results are meant to provide the bank manager with the benefit of foresight while balancing the available resources with the number of staff that need to be deployed. As the results have shown, the simulation runs have predicted the number of employees that should be deployed for the number of customers expected to be served. This way, cases of planning with more staff than necessary or planning with very few staff can be avoided. In addition, a bank's management can avoid the wastage of human and financial resources that would otherwise result in very high expenditure in scenarios like this one.

Further research can be done to determine if this simulation scenario can be replicated for other service provision centers like fuel stations, supermarkets, etc., especially given their slightly different ways of handling customers during surge times. More specifically, it would be interesting to observe how the number of customers served per unit of time affects the whole system's performance.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

David Oriedi carried out the research and wrote the whole paper. Kennedy Malanga contributed technical adjustments to the simulation experiments that were carried out. George Musumba, Gabriel Kamau, and Mosses Ollengo assisted in proofreading the article. All authors had approved the final version.

Funding

This work was supported in part by Dedan Kimathi University of Technology Directorate of Research.

Acknowledgment

We acknowledge the Directorate of Research of the Dedan Kimathi University of Technology for providing the funds needed for presenting this article to a conference and its subsequent publication in the journal.

References

[1] T. N. Thakur and N. Yoshiura, "Modeling and verification of a hybrid model of mobile banking based branchless banking system in e-banking using SPIN," in *Proc. 2022 ACM International Conf.*, 2022, pp. 149–156.

[2] Y. Hu and J. Tian, "Queuing system optimization based on customer experience of Chinese commercial intelligent bank branches," in *Proc. 2019 ACM International Conf.,* 2019, pp. 140–144.

[3] F. Muslimin, A. N. Fajar, and Meyliana, "Business process management and service oriented architecture integration for transactional banking application," presented at the International Conf. on ICT for Smart Society, Bandung - Indonesia, November 19-20, 2020.

[4] R. L. Bennett, D. A. Nuxoll, R. A. Jarrow, M. C. Fu, and H. Zhang, "A loss default simulation model of the federal bank deposit insurance funds," in *Proc. 2005 Winter Simulation Conf.*, 2005, pp. 1835–1843.

[5] T. Taniai and I. Sasase, "Throughput analysis of input queueing packet switch under improved FIFO policy," in *Proc. 1992 Singapore ICCS/ISITA, 1992,* pp. 1282–1285.

[6] N. M. van Dijk, "On hybrid combination of queueing and simulation," in *Proc. 2000 Winter Simulation Conf. 2000*, vol. 1, pp. 147–150.

[7] D. Hammond and S. Mahesh, "A simulation and analysis of bank teller manning," in *Proc. 1995 Winter Simulation Conf. 1995.*, pp. 1077–1080.

[8] P. Y. Liao and L. W. Tyan, "Optimal pricing strategy for queuing systems with balking loss and reneging loss," in *Proc. 2007 IEEE International Conf. on Industrial Engineering and Engineering Management 2007.*, pp. 1738–1742.

[9] S. Connor. (February 2020). Omnithermal Perfect Simulation for Multi-server Queues. *ACM Trans. on Modeling and Computer* Simulation [Online]. 30(1). pp. 1-15. Available: https://dl.acm.org/doi/pdf/10.1145/3361743?download=true

[10] S. A. Bishop, H. I. Okagbue, P. E. Oguntunde, A. A. Opanuga, and O. A. Odetunmibi (August 2018). Survey dataset on analysis of queues in some selected banks in Ogun State, Nigeria. [Online]. Available: https://www.data-in-brief.com/article/S2352-3409(18)30607-3/fulltext

[11] C. Kiataramkul and K. Neamprem, "Simulation of queueing system for commercial bank in university: Case study of Bangkok bank at king mongkut's university of technology north Bangkok," presented at the 2019 Research Invention, and Innovation Congress Conf., Bankok-Thailand, December 11-13, 2019.

[12] H. Song and D. Zhenwei, "Simulation of banks queuing system based on WITNESS," presented at the International Conf. on Computer Application and System Modeling, Taiyuan-China, October 22–24, 2010.

[13] Y. B. Wang, C. Qian, and J. De Cao, "Optimized M/M/c model and simulation for bank queuing system," in *Proc. 2010 IEEE International Conference on Software Engineering and Service Sciences, 2010*, pp. 474–477.

[14] N. Madadi, A. H. Roudsari, K. Y. Wong, and M. R. Galankashi, "Modeling and simulation of a bank queuing system," in *Proc. 2013* International *Conference on Computer Intelligence Modeling and Simulation*, 2013, pp. 209–215.

[15] W. Zandbergen "An empirically-grounded simulation of bank depositors," presented at the Winter Simulations Conference, Washington-USA, December 8-11, 2013.

[16] A. Shah, A. Wikum, and J. Pender, "Using simulation to study the last to enter service delay announcement in multiserver queues with abandonment," in *Proc. 2019 Winter Simulation Conference*, 2019, pp. 2595–2605.

[17] S. Wei, C. Yong-Ke, Y. Ai-Jun, H. Wei, and Z. Chao, "Research on Simulation of Queuing System Based on OOP," in *Proc. 2018 IEEE International Conference on Computer and Communication Engineering Technolology (CCET), 2018*, pp. 34–38.

[18] C. Yin, "Outpatient queue business simulation based on acceptable waiting time," presented at the International Conference on Computer Design and Applications (ICCDA), Qinhuangdao-China, June 25-27, 2010.

[19] E. Berhan. (June 2015). Bank Service Performance Improvements using Multi-Sever Queue System. Business and Management. [Online]. Available: https://www.researchgate.net/profile/Eshetie-Berhan-2/publication/286935591_Bank_Service_Performance_Improvements_using_Multi-Sever_Queue_System/links/5671604108ae5252e6f3edad/Bank-Service-Performance-Improvements-using-Multi-Sever-Queue-System.pdf

[20] D. M. Mathew, É. F. Zulian, S. Kannoth, M. Jung, C. Weis, and N. Wehn, "A bank-wise DRAM power model for system simulations," in *Proc. 2017 ACM International Conference on Rapid Simulation and Performance Evaluation: Methods and Tools,* 2017, pp. 1-7.