Model to Determine Bank Teller Requirements and Predict Transactions

Lawrence Muchemi
Computer Science Lecturer
University of Nairobi
Nairobi, Kenya
Email: lmuchemi@uonbi.ac.ke
Tel: +254 722 234 738

Chege Kennedy R. Muchai
MSc. Computer Science Student
University of Nairobi
Nairobi, Kenya
Email: kcmuchai@yahoo.com
Tel: +254 722 242 282

Abstract: This paper presents an application study to find the optimal Bank teller requirements for providing high quality service levels at reasonable cost and minimizing the waiting times. A model based on the mathematical theory of queues, Little's result, theorem, lemma, law or formula, i.e. $L = \lambda W$, will be developed to calculate the desired teller manning levels. This model will be tested using simulation; data collected from the processing system will also be used to forecast customer transactions for each day of the week and the hourly arrivals. The forecast will be used in the manning model to provide the suggested teller manning levels. The result is a dynamic scheduling option that deploys tellers as per the customer arrivals/demand and a decision tree model for prediction.

Keywords: Bank, Model, Simulation, Teller, Little, Forecast, Transactions, Decision Trees

I. INTRODUCTION

In the highly competitive banking business, Customers demand high levels of service [8]. Now numerous options from the over forty (40) commercial banks available in Kenya, patrons are spoilt of choice.

In spite of the availability of automated teller service, many customers still prefer to use human teller services, but long waits for service are perceived as a major source of customer dissatisfaction [1]

Teller personnel have many tasks to accomplish during the business day besides direct contact with the banking customer. Management's task is to have enough teller stations open to provide quick service while ensuring productive work time is not wasted by having idle tellers.

The banks have conventionally been associated with queues, that anytime one has to visit the Banking Hall, the thought of long wait in the queue deter them away.

The bank's data collection system provided the average number of customer/teller transactions for each time period of the day, grouped into one hour intervals.

With these data, a simulation model with ARENA simulation software would be built and described. The logic model and descriptions of more details about simulation modules would be demonstrated so as to understand the current system and extract relevant performance metrics such as total waiting time. The output of the proposed resourcing model was incorporated in another ARENA model, and the reduced waiting time recorded. Consequently, verification to assure the model behaves as intentions, and validation to certify the model behaves the same as the real system, of the model would be done.

Increasing productivity of Banking Operations has become a major issue in Bank management. Teller line services have been identified as the major area for productivity improvement. This fact has highlighted the need for establishing optimal staffing levels based on standards of customer service. [14] Customers demand higher standards of service and now have numerous choices from where to get served.

A common feature of many service industries ranging from telephone call centers to students in the dining hall and hospital emergency rooms is that, the demand for service often varies greatly by time of day.

The data collected from the bank's system and that gathered via observation will be used to estimate the number of transactions to expect on a particular day of the week and group the same on an hourly basis based on historical arrivals and in so doing advice an optimal number of tellers to deploy so as to serve these customers effectively.

Problem Statement

In businesses where the underlying products have become commodity-like, quality of service depends heavily on the quality of its personnel. This is well documented in a study by [9], who documented that approximately 40 per cent of customers switched banks because of what they considered to be poor service.

Customers expect to be served promptly when they arrive, and therefore there is a need for Optimal Staff Deployment on the customer facing operations. The Teller line especially should take into account varying service demand levels. From the internal survey conducted by this Bank in 2009, long queues were pointed out to be the major source of customer dissatisfaction.

This was also reflected by the customers' feedback and constant complaints that are received via the suggestion boxes, online social media networks such as Tweeter and Facebook.

Optimality is reached when (a mix of) objectives and performance measures (is) are satisfied [11]. Setting staffing requirements is one in a hierarchy of decisions that must be made in the design and management of a service system [3]

Study Objectives

The idea is to have a Teller staffing model, based on work volumes. This study aimed to achieve the following objectives:-

1. Develop a model that could be used to determine teller requirements
2. Predict transaction volumes for each day of a week.
3. Reduce the customer waiting time for service

II. LITERATURE REVIEW

Personnel or staff scheduling problems have been studied for many years due to its importance on the overall performance of a system in terms of quality of service to the customer and cost to the organization [10].

Related Studies

[2] Using simulation and analysis proposed models to find cost effective bank teller management policies for providing high quality service levels at reasonable costs in a modern banking system.

[9] While assessing quality of service to depend heavily on the quality of personnel, documented that approximately 40 percent of customers switched banks because of what they considered to be poor service. Leeds further argued that nearly three-quarters of the banking customers mentioned teller courtesy as a prime consideration in choosing a bank.

[4] In his study of Swedish consumers, noted that although customer satisfaction and quality appear to be important for all firms, satisfaction is more important for loyalty in industries such as banks, insurance, mail order, and automobiles.

[12] In their Call Centers study, exhibited that forecasting is typically driven by a combination of historical data, time series models and expert judgment. They went ahead to demonstrate how to combine queuing theory and simulation for determining optimal staffing levels. The simulation model is used to perform what-if scenarios, because of high flexibility. The queuing model is used to approximate the system performance measures.

Queuing Theory

Queuing Theory is defined as a collection of mathematical models of various queuing systems. Used extensively to analyse production and service processes exhibiting random variability in market demand (arrival times) and service times [5]

The queuing theory with simplest and most-widely used model being the M|M|s queue with parameters λ, μ and s, the primitives:

- The arrival process, assumed Poisson at a constant rate λ
- The service times, assumed exponentially distributed with mean μ
- The number of agents, s

Queues form when the demand for a service facility exceeds the capacity of that facility i.e. the customer do not get served immediately upon arrival but must wait [6]. Waiting lines or queues are a common occurrence both in everyday life and in variety of business and industrial situations. Most waiting line problems are centered about the question of finding the ideal level of services that a firm should provide.

Figure 2.1 shows the configuration of a basic queue system characterized by, the size of the calling population, pattern of arrival at the system and behavior of the arrivals

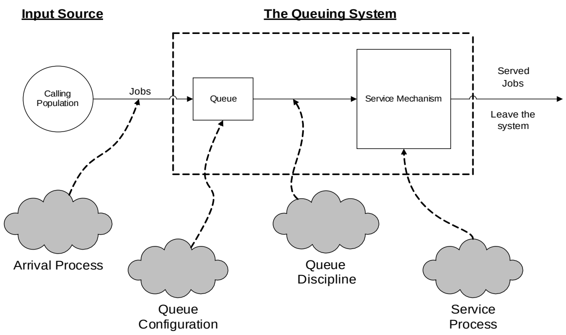

Figure II.1: Components of a Basic Queue System

There are several service types:-

- Single Server – Single Queue
- Single Server – Several Queues
- Several (parallel) Servers – Single Queue
- Several Servers – Several Queues
- Service Facilities in Series

Queue discipline – i.e. the order or manner in which customers from the queue are selected for service.

i. First-Come, First-Served (FCFS) – customers are served in the order of their arrival
ii. Last-Come, First-Served (LCFS) - the customers are served in the reverse order of their entry so that the one who joined last is served first for example the people who enter an elevator last, are the first ones to exit
iii. Dynamic Queue Disciplines are based on the individual customer attributes in the queue. These are: -

a. Service in Random Order (SIRO) - under this rule, customers is selected for service at random, irrespective of their arrival in the service system. Every customer in the queue is equally likely to be selected. The time of arrival of the customer is therefore of no relevance

b. Priority Service – under this rule customers are grouped in priority classes on the basis of some attributes such as service time, urgency or to some identifiable characteristic [15].

Conceptual Model

Consider service system at many commercial Banks, where a customer, upon filling the relevant stationery queues and wait for the next available Teller, they are served and after that, exit the system. When there is no one in the queue, the customer is served immediately and exits the system as depicted in the figure 2.2 below

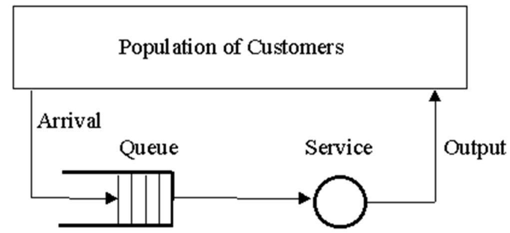

Figure II.2: The Queuing System

- ✓ Population: - Bank's customers
- ✓ Arrivals: - random arrival rates varying with time
- ✓ Queue: - several Parallel servers, single queue
- ✓ Queue discipline: - FCFS
- ✓ Service: - service times (durations for each transaction type)

Simulation

A simulation model is a representation that incorporates time and the changes that occur over time. A discrete model is one whose state changes only at discrete points in time, not continuously. A discrete-event model is one whose state changes only at discrete times called event times. When an event occurs, it may trigger new events, activities and processes [7].

An event is an instantaneous occurrence that changes the model's state. Examples include an arrival event for a customer at a bank, and a service completion event for the same customer. An activity is a duration of time, such as a service time or inter-arrival time that is initiated by an event in conjunction with the model being in a certain state. For example, when arrivals are defined by a probability distribution of inter-arrival times, then when one arrival occurs (an event), the model generates a new inter-arrival time (an activity) which in turn will cause the next arrival event.

A resource is an entity that provides a service to dynamic entities. A resource usually has a finite capacity representing some system constraint. Examples include a worker or a team of workers doing a task, a machine, or a vehicle. In some models, resources may have user-defined resource states and special characteristics such as downtimes and availability schedules.

Almost all discrete-event models are stochastic. That is, they contain some components that are modelled as a statistical distribution. This introduces random variation into a model, making it into a statistical or sampling experiment. More precisely, when one or more components are stochastic (for example, inter-arrival or service times), then model outputs are stochastic, necessitating some kind of statistical analysis to draw valid conclusions [7].

Simulation was based on the interaction perspective, for example, in the bank, the teller serves one customer until completion, and then looks at the queue. If the queue is not empty, the teller “takes” the first entity out of the queue, changes its own state to “busy” and begins a new service activity. If the queue is empty, the teller changes its own state to idle.

The simulation model could not simulate all aspects of the real system, so it would be considered under following assumptions:

1. The working hours (Monday-Friday, 8.30 AM to 4 PM) with a 1 hour lunch break
2. All Tellers are universal i.e. can perform any of the above mentioned function
3. All Tellers take the same time to process a transaction
4. All Tellers have equal capabilities
5. There is no system downtime
6. A customer is served by one Teller at a time
7. A customer cannot renege i.e. once the customer joins the queue they cannot leave without being served
8. Customers wait for service in a single queue

Logic Flow

A general logic flow diagram of the system is depicted as on the Figure 2.3. Random customers enter the system at different times.

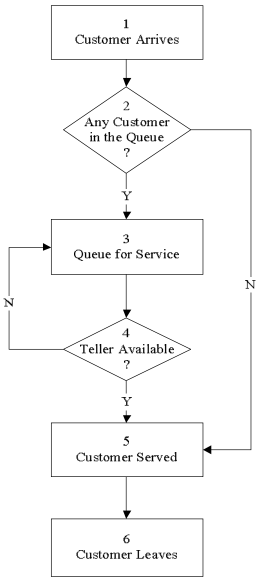

Figure II.3: Logic Flow

Key Performance Measures

We were interested with the following performance measures:

- Average waiting time and
- Quantity in queues of customers

With the problem statement mentioned above, an ARENA simulation model of the studied system is constructed. The ARENA simulation software is used because it has many advantages in simulation and modelling of discrete system.

III. METHODOLOGY

Data Description

The following data formed the input variables to the simulation: -

1. Service times
2. Customers’ inter-arrival times
3. Total transactions
4. Time taken to fill requisite stationery
5. Number of Servers (Tellers)

Service time is defined as the elapsed time from when a teller initializes customer service through the teller transaction platform, to when that customer engagement is ended on the platform. This may exclude some customer engagement time before and after a transaction. Although friendly greetings and small-talk are important for maintaining a positive customer relationship, including them within the service time can be misleading. For instance, tellers may be more likely to have extended conversations with customers during less busy periods in the branch, skewing service time statistics. Our goal in analysing service times is to recommend staffing levels. So it is better to focus on the actual transaction portion of the customer-teller engagement to get an accurate matching of teller performance to staffing needs.

Tools

MS Excel – for data analysis and data clean up so as to have data that can be adapted by ARENA and DTREG and to develop the mathematical/spreadsheet to compute transactions distributions completed on each interval

Modeling and Simulation Software (ARENA form Rockwell) – to provide a less costly approach of experimenting with the model of the real system and also for its great input and output analyzer feature to achieve sufficient level of accuracy.

ARENA is also powerful in simulation and modeling of discrete systems [17].

DTREG - to provide a state of art modeling method and build the decision trees to predict workload/transactions for any given working day of the week (Monday - Saturday) [13].

Visual Basic 6.0 – to develop an application that computes the teller requirements based on hourly arrivals of the customers

Procedure

Our first step is to build a mathematical model based on the mathematical theory of queues. Little's result, theorem, lemma, law or formula says: The long-term average number of customers in a stable system L is equal to the long-term average effective arrival rate, λ, multiplied by the average time a customer spends in the system, W; or expressed algebraically: $L = \lambda W$ [6].

We will use this to calculate the Full Time Equivalents (FTEs) needed in each interval of time of day say one hour (see table 1)

The model is based on the assumption that ours is a stable system, i.e. the rate at which people enter the bank is the rate at which they arrive, and the rate at which they exit as well.

Two tellers should be stationed at the Teller stations on a full time basis so as not to leave the customers unattended to in the event that one takes a break. The model translates the actual number of bank transactions from a past day into required manning.

This branch has deployed four (4) tellers on a full time basis that work from 8.30am until 5.00pm regardless of the customer arrivals (queue length)

The average transactions that were processed by these four tellers for the duration that we sampled is shown in table 3.1

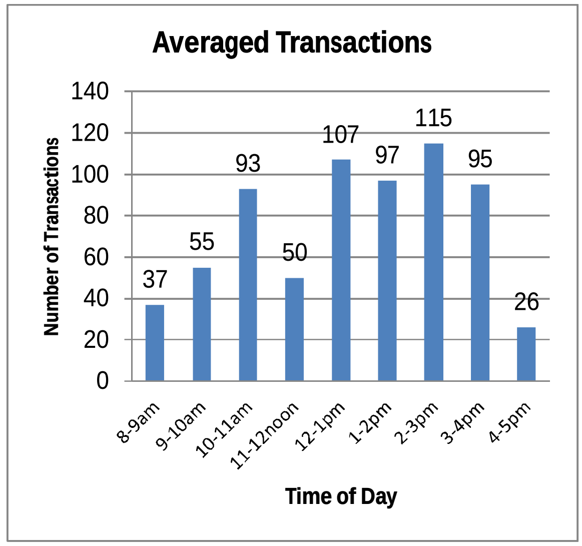

Table III.1: Average hourly entries

The bank's information system is primarily oriented toward data collection, while reports were readily available on the number of transactions; any analysis were accomplished through spread sheets (MS Excel) on a personal computer and ARENA's input analyser that gives the best fitted statistical distribution.

ARENA 12 was the Simulation and modelling software selected for its great input and output analyser feature to achieve sufficient level of accuracy and for its supremacy in simulation and modelling of discrete systems.

Detail about the ARENA Modules

The arrival customer module generates entities entering the system. Entities (customers) are generated and defined by a schedule that translates to hourly arrivals. See figure 3.1.

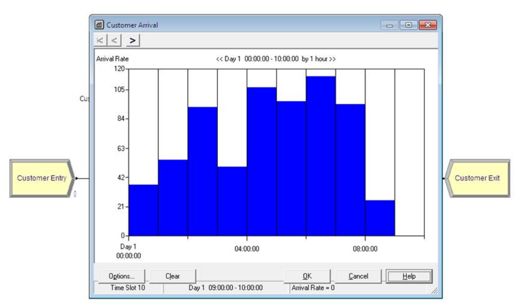

Figure III.1: Arrivals Distribution

The customer is served by one bank teller at a time. The Teller service module presents the transaction times that is a random number with the best fitted statistical distribution.

This model assumed a Poisson distribution with its fixable mean and standard deviation. The logic action of the Process module is set by Seize Delay Release because the entity will temporarily seize the resource, delay it and release after using, See table 4.1 and figure 3.2 [18]

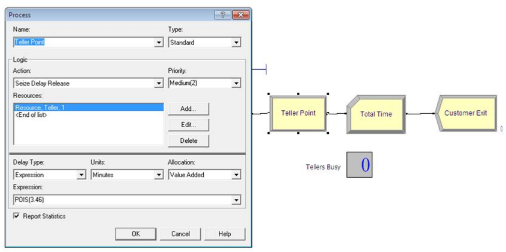

Figure III.2: Teller Service Module

The current system's resource (tellers) was fixed i.e. four (4) whereas the model proposed manning approach was implemented in a schedule (see figure 3.3) in which the required quantity of the resource is one and fluctuates as per the customer arrivals (demand). After being served at the bank teller, the entity will exit the system.

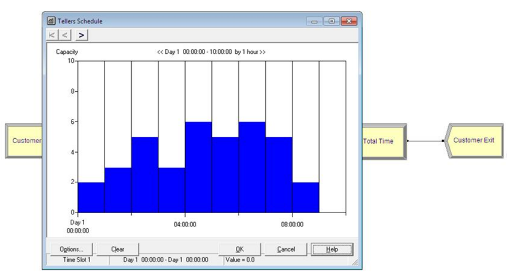

Figure III.3: The teller schedule module

Supervised Learning

Using transactional data for 2011 where a total of 148,429 transactions were completed for the branch for 304 working days. This data was represented monthly (January – December) for each day of the week (Monday – Saturday)

To analyze the modeling data we used supervised learning as:-

- The input data has both predictor (independent) variables and target (dependent) variable whose value is to be estimated and
- The goal is to predict the value of some variable

Since our goal is to learn the data and predict some value, well develop a decision tree using historical transaction data. DTREG will use this data to learn how the value of the target variable is related to the value of predictor variables.

To improve on accuracy level, we modeled the whole set for 2011, and used data for January 2012 to test the accuracy of our estimate.

For example, to predict the value for say, Monday (target variable), all other variables (month, Tuesday – Saturday) will form the predictor variables, see figure 3.4. DTREG performed complex analysis on this data and built decision trees that modeled the data, see figure 4.2.

The analysis also showed the relative importance of predictive variables in relation to the target variable as shown in figure 3.5.

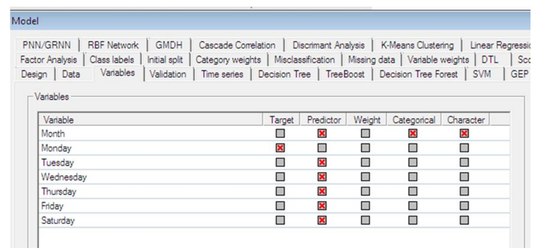

Figure III.4: Predictor & Target variables

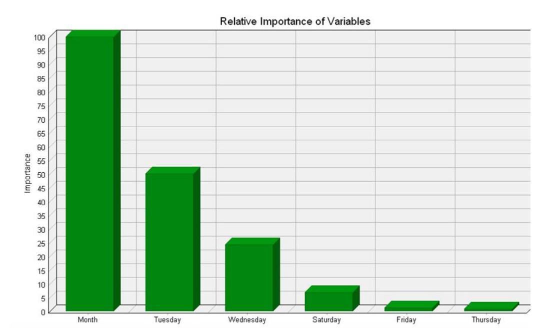

Figure III.5: Predictor variables importance

Decision trees were chosen as:-

1. They are easy to interpret even for non-technical people
2. It's possible to predict target values for specific cases where only the predictor variables are known

Application Development

An application was developed with Microsoft Visual Basic 6.0 which provided an easy to use GUI that the user can click on the predictions of each individual day and the respective decision tree is. The user can then 'walk' the tree to get the prediction based on the input variables.

On getting the estimated value from the decision trees, the same can be entered on the application to get the manning levels of the day in one hour intervals throughout the day.

IV. RESULTS

Simulation

From the data collected of the four (4) full time tellers, using ARENA's input analyser, the best fitted function was a Poisson distribution with a mean of 3.46. Other fitted functions are summarised in table 4.1

Distribution: Poisson
Expression: POIS (3.46)
Square Error: 0.011508

Function	Sq Error
Poisson	0.0115
Weibull	0.0163
Normal	0.0179
Beta	0.0191
Gamma	0.0191
Erlang	0.0201
Triangular	0.0212
Lognormal	0.0328
Uniform	0.0349
Exponential	0.0516

Table IV.1: Service Distribution

In the same breath, the time taken by customers to fill up pre-requisite stationery was fitted to a Gamma distribution with the following parameters

Distribution: Gamma
Expression: 0.5 + GAMM (1.84, 1.63)
Square Error: 0.004378

Simulation results for the current system with four (4) full time tellers, gave the results in table 4.2, while the proposed model results are shown in table 4.3

Current System		
Performance Measure	**Average Time (hrs)**	**Average Time (mins)**
Service & Stationery Time	0.114	7
Queuing Time	0.5933	36
Total Time	**0.7073**	**43**

Table IV.2: Current System Performance

Proposed System		
Performance Measure	**Average Time (hrs)**	**Average Time (mins)**
Service & Stationery Time	0.115	7
Queuing Time	0.3344	20
Total Time	**0.4494**	**27**

Table IV.3: Proposed System Performance

The Proposed Manning Model – Objective 1

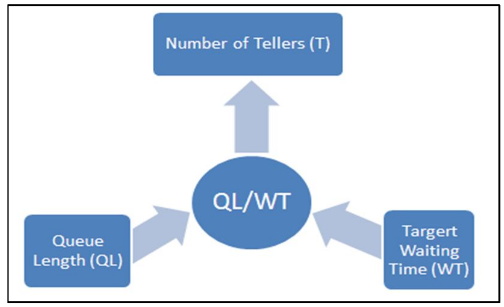

Figure IV.1: The Model

Where:-
Queue Length (QL) – as defined by Little's law
Target waiting time (WT) – the minimum duration set that a customer should wait for service
Number of Tellers (T) – calculated tellers/resources

From the collected data, the average waiting time was 0.7073hrs i.e.42mins, this we got by simulating the as-is (current) system.

This model generated the output shown in table 4.4

Transactional Data for Jan 04 2011 by 4 Full Time Tellers			
Time	**Transactions (λ)**	**Length of Queue (Little's Law) L= λ W**	**Model Resources (Tellers)=Queue Length/W**
8-9am	37	27	2
9-10am	55	39	3
10-11am	93	66	5
11-12noon	50	36	3
12-1pm	107	76	6
1-2pm	97	69	5
2-3pm	115	82	6
3-4pm	95	67	5
4-5pm	26	19	2
Total Txns	**675**		
W =	**0.7073**		

Table IV.4: Model computed results

Transaction Estimation – Objective 2

Using the regression tree models generated by DTREG which sets other input variables for example Month, Monday, Tuesday, Wednesday, Friday & Saturday as Predictor variables while Thursday is set as the Target Variable, Scoring/walking the tree gives the expected values for "Thursday" our target variable as shown in figure 4.2

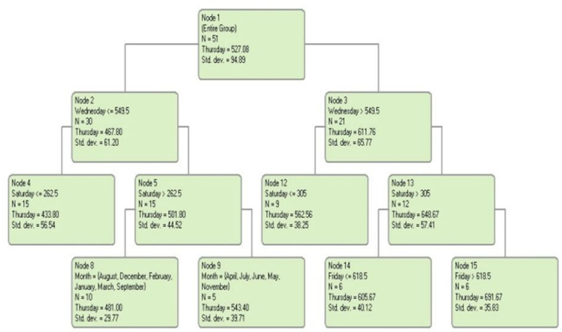

Figure IV.2: Thursday Tree

Accuracy Testing

To test the accuracy of the prediction, we used predicted values of the model (table 4.6) and tested against actual values for January 2012 (table 4.5). The estimation accuracy achieved was 90% as shown in table 4.7

Actual Data for Jan 2012						
Month	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**
Jan	803	757	592	620	714	410
Jan	634	626	434	436	542	333
Jan	525	584	530	595	562	431
Total	**1,962**	**1,967**	**1,556**	**1,651**	**1,818**	**1,174**

Table IV.5: Actual values for Jan 2012

Estimated Values Results						
Month	Mon	Tue	Wed	Thu	Fri	Sat
Jan	746	547	616	691	611	375
Jan	534	547	541	481	474	283
Jan	613	454	576	481	611	315
Total	**1,893**	**1,548**	**1,733**	**1,653**	**1,696**	**973**

Table IV.6: Estimated values for Jan 2012

Accuracy Measure						
Residual value	**69**	**419**	**-177**	**-2**	**122**	**201**
Daily Accuracy	96%	79%	90%	100%	93%	83%
Average Accuracy	**90%**					

Table IV.7: Accuracy testing

Waiting Time reduction – Objective 3

The model reduced the total time spent by customers with an approximate 60% from 36 to 20 min. It scheduled the tellers with intervals of one hour each and is driven by the demand/customer arrivals. The modelled resources are shown in table 4.4

Queue Performance

The performance of the queue was based on: -

- Duration spent on the queue
- Number of people waiting

The duration spent by customers is shown in tables 4.2 and 4.3 for both the current and the proposed systems respectively and the average number of people waiting is shown in table 4.8

System	**Avg. Number Waiting**
Current	36
Proposed	21

Table IV.8: People waiting

V. CONCLUSION

The objectives of this study were:-

1. To develop a model that can be used to determine teller requirements
2. Predict work volumes for each day of a week.
3. Reduce the waiting times

Our efforts produced a model developed from the queuing theory and Little's law principles, illustrating the inadequacy of fixed full time manning system of the teller service stations and highlighted an optimal and effective solution of the scheduling system that would meet the bank's desired service levels.

The literature studied was from scientific literature with key findings such as it is the ideal combination of using simulation models next to queuing theory/models to obtain best results. The simulation model can be used to perform what-if scenarios, because of high flexibility. Apart from the fact that simulation provides the possibility to model more complex processes. The queuing model is used to approximate the system performance measures.

When following an event scheduling perspective, a model developer must define the model logic and system state changes that occur whenever any event occurs. For example, a customer process at a bank consists of an arrival event (to the lobby, perhaps), joining and waiting in a queue (a delay), a service time by a teller, and finally a service completion event (see figure 3.2). In terms of concepts discussed earlier, the service time is an activity and the teller is a resource.

Resources/tellers are computed on the basis of the arrival rate of the customers

The study was based on several assumptions that were made on the onset, a major assumption was that of assuming that all tellers are Universal i.e. are equal in skills and processing speeds.

Banks have come up with generalists and specialists arrangements to hasten the service to customers and reduce on waiting times. Generalists are agents that can handle all types of customers, and specialists are agent that can handle only one type of customer [16]. With this in mind future work could be developed around this area to translate the model to fit the real world and assess the effect on the waiting time for service.

Another interesting factor to look into would be the differences of working experiences of the tellers incorporated with the gender

VI. REFERENCES

[1] David Cohen, Christopher Gan, Hua Hwa Au Yong, and Esther Choong. "Customer Satisfaction: A Study of Bank Customer Retention." New Zealand, 2006.

[2] Donald Hammond, and Sathi Mahesh. "A Simulation and Analysis of Bank Teller Manning." *Proceedings of the 1995 Winter Simulation Conference.* 1995.

[3] Efraim Turban, and Jay E. Aronson. *Decision Support Systems and Intelligent Systems.* 1997.

[4] Fornell. "A National Customer Satisfaction Barometer: The Swedish Exprince." *Journal of Marketing* , 1992: 1-21.

[5] Ivo Adan, and Jacques Resing. *Queuing Theory.* 2002.

[6] John D.C, Little, and Stephen C. Graves. *Little's Law.* 2008.

[7] John S. Carson II. *Introduction to Modelling and Simulation.* USA: Proceedings of 2005 Winter Simulation Conference, 2005.

[8] Kaynak, E., and Kucukemiroglu O. "Bank and Product Selection." *The International Journal of Bank Marketing*, 1999: Vol. 17, No. 11. – pp. 5-19.

[9] Leeds. *'Mystery Shopping' Offers Clues to Quality Service.* Bank Marketing, 1992.

[10] M.A. Centeno, R. Giachetti, R. Linn, and A.M Ismail. "A Simulation-ILP based tool for scheduling ER Staff." *Proceedings for 2003 Winter Simulation Conference* , 2003: 7-10.

[11] Mark .J. Stegeman, and Monique. H. Jansen-Vullers. "Determining Optimal Staffing Levels in Multiple Skill Inbound Call Centers." 2005.

[12] Mehrota V., and J. Fama. "Call Centre Simulation Modeling: Methods Challages and Opportunities." *Proceedings of the 2003 Winter Simulation Conference* , 2003: 135-143.

[13] Philip H. Sherrod. *DTREG Predictive Modeling Software.* 2003 - 2012.

[14] Pihl, W. H., and Wambay M. L. . "The Leading regionals focus on productivity." *Bank Management,* 1990: 66: 52-55.

[15] Robinson Stewart. *Simulation: The Practice of Model Development and Use.* 2004.

[16] Stolletz, Raik. *Performance Analysis and Optimisation of Inbound Call Centers.* New York: Springer-Verlag Berlin Heidelberg, 2003.

[17] Tayfur Altiok, and Benjamin Melamed. *Simulation Modeling and Analysis with Arena.* 2007.

[18] Travis Cogdill, and Michael Monticino. "Analysis of Teller Service Times in Retail Banks." 2007.