

A Comparative Analysis of Waiting Time Routing Rule for Queue Reduction in Call Center

1. Mughele Ese Sophia: Department of Computer Science, Delta State School of Marine Technology Burutu, Delta State Nigeria. Email: prettysophy99@gmail.com
2. Prof. Stella Chiemeké: Professor of Computer Science and Director of Intellectual Property and Technology Transfer Office (IPTTO), University of Benin, Benin City, Edo State Nigeria. Email: schiemeke@yahoo.com
3. Dr. Susan Konyeha: Department of Computer Science, University of Benin, Benin City, Edo State Nigeria. Email: susan.konyeha@gmail.com

Abstract

Satisfying customers need has become an inevitable phenomenon for businesses to survive in this high competitive era. The persistent complaints of waiting queue by customers have continuously posed a threat to call centres globally. Call centres are an important function of most companies' day to day business activities. They are often the link between a company and its customers and hugely impact the customer's perspective or point of view of a company. The queue experienced by customers at call centres is increasingly becoming alarming as many customers are irritated by the long time spent on the queue before their calls are been answered. It is imperative that an investigation be carried out into what criteria are been used to route calls to any particular call centre agent at any given time and how average handling time influences waiting queues at call centres. It is important to conduct a comparative analysis to evaluate existing routing rules for waiting time rules to determine the optimal among the waiting time rule. This study used a collection of java programs to simulate existing rule for waiting time routing rules. The JAVA program evaluator evaluates each of the waiting time routing rules using the call center's raw data, collected from the data logging system of Global Communications, one of the biggest telecommunication company in Nigeria. The study simulated four rules, Java program were developed for each of the four routing rules since their procedure varies from one another. The result from our simulation was able to determine the optimal rule among the existing waiting time routing rules.

Keywords: Call center, Queuing system, Routing rules, Average Speed of Answer

1.0 Introduction

From time immemorial, the most distinctive zeal of professional managers is the ability to achieve high productivity level with emphasis on enhancing customer satisfaction with a view to making profit for the business in the long run. Businesses create value through product offerings or service delivery to their customers. And for customers to have access to these product and services sometimes queue of wait in line to have them. It is therefore the concern of every company management to render prompt delivery of service, eliminate waiting queue and give value for money so as to ensure customer satisfaction and loyalty.

Call centre can be defined as any group whose business is talking to customers or prospects through eh telephone. According to Brizola et al (2001), a call centre is a system that offers complete management of all communication channels between a business and its customers, optimizing polices, eliminating duplicated work and making better use of time. The call centre service has grown a great deal with its application in all sectors of the economy. It serves as a primary contact between businesses and clients. But in recent times, customers waiting for so long in order to lodge a complaint or make an enquiry have become a worrisome phenomenon in the call centres especially in telecommunications.

In contemporary society, satisfying customers need has become a phenomenon seen to be highly inevitable for business that wants to survive in this era of high competition amidst the global financial crisis. A customer's experience during a service encounter consist of two parts namely: the time spent waiting for the service and the service itself. Call centres give priority to the two criteria with emphasis on one more than the

other. Those that place more emphasis on time spent waiting for the service are more concerned with reducing the average time involved in handling a call while those that are concerned with the service itself aims at effective resolution of customer issues. This paper focuses on the spent on queue and techniques to reduce waiting time on queue by customers in call centers. Armony (2005) says for a call centre to reduce waiting lines with emphasis on the reduction of time spent, its best to route calls to agents who can handle customer issues the fastest, sometimes even holding a call in queue to wait for that agent than routing the call to a slower agent. This might lead to further increase in congestion, repeat calls from unreceptive issues and undue burden on some agents.

The Nigerian telecommunications industry is a rapidly growing sector with subscriber base running into millions, and the existence of waiting queue is a common feature of call centres. Consequently customers experience unpleasant waiting in queue develop a negative perception and attitude to a firm's service which may affect the long-term success or prosperity of such firms. In this work, we explore strategies for routing multiple types of calls to a large group of agents, where these assignments are made dynamically based on the specific attributes of the agents and/or the current state of the system. We used data from a telecommunication organisation in Nigeria to conduct simulation on existing waiting time oriented routing rules; the purpose is to determine the routing rule with optimal solution. We believe that this study will make several important contributions to the call centre operations management regarding reduction of queues.

2.0 Related Literature

Enyioko, (2016), defined a queuing system as a birth-death process with a population consisting of customers either waiting for services or currently in service. A birth occurs when a customer arrives at the service facilities. A death occurs when a customer departs from the facility. The state of the system is the number of customers in the facilities. A queue is a situation whereby customers wait in line to be attended to. Sharma (2009) defines it "as any place where a customer (human beings or physical entities) that requires service is made to wait due to the fact that the number of customers exceeds the number of service facilities or when service facilities do not work efficiently and take more time than prescribed to serve a customer.

From the above, it can be deduced that queues arise as a result of either customers exceeding the number of service facilities or service facilities not functioning efficiently where there is a queue, there is a queuing system. Queuing system is a set of customers, set of service and an order which customers arrive and are rendered service to (Enyioko, 2016). Queuing systems are characterized by five components namely: - The arrival pattern of customers, the service pattern, the number of servers, the capacity of the facility to hold customers and the order in which customers are served. Enyioko (2016) explained the five components as follows:-

- a. **Arrival Patterns:-**This is the time between successive customer arrivals to the service facility. It may be deterministic or probabilistic and do they arrive in batches or single.
- b. **Service Patterns:-** This is the time required by one server to serve one customer.
- c. **Number of Servers:-** This is the number of servers available to serve customers. It could be single or multiple.
- d. **Capacity of Facility:-** This is the maximum number of customers both in service and those in the queue permitted in the service facility at the same time. The system capacity can be finite, with a limit or infinite without limit.
- e. **Queue Discipline:-** This is the order in which customers are served. It could be first-in, first out or last-in, first out or on priority basis.

There are two approaches to mathematical analysis of a queuing system:

1. **Analytical model.** The behaviour of all elements in the queue is described by system of (differential) equations.
2. **Simulation model.** The behaviour and relations among elements of the queue is modelled by computer program. In both analytical and simulation model there are four major elements of any queuing situation

which must be considered: arrivals, services, number of servers, capacity of facility and queue discipline (Parson, 2016).

2.1 The concept of Call Centres

Various attempts have been made by several authors and organizations to find a comprehensive and universally accepted definition for the term call centre. Each group defining it as it appears to her. The following are source of such definitions:-

Brizola et al (2001), a call centre is a system that offers complete management of all communication channels between a business and its customers, optimizing process, eliminating duplicated work and making better use of time. A call center can also be perceive as a centralized office used for the purpose of receiving and transmitting a large volume of request by telephone. Avramidis et al (2004) defines it as a set of resources (communication equipment, employees, computers etc.) which enable the delivery of services via the telephone. From the above, it can be seen that call centres are limits that manages an organization communication system. Call centres are known by a variety of names namely: contract centre, customer service centre, customer interaction centre, customer service point etc.

2.2 Customer Satisfaction and Waiting Time

Customer satisfaction has been defined as the difference between the customer's perceptions of the experience and his or her expectations, which is many times based on past experience. Although it is possible to manage and decrease actual waiting time and to some extent to manage customer expectations about customer satisfaction, managing the customer's perception of the queuing experience can be the vital element in satisfaction with the service interaction. The measurement of customer satisfaction as it relates to waiting time is highly qualitative and subjective, and the relationship is generally inverse in nature (i.e., in general, as waiting time decreases, satisfaction increases). This relationship was further expanded by Maister who, in 1985, postulated that satisfaction is dependent on customer perception and customer expectation.

Armory and Maglaras (2004) observed that Customers in a call service center experiences real time delay as a result of queue and call back delay. This metrics affect customer's perception of the product or service and this impact on customer's loyalty. The study deployed Probabilistic choice model and the dynamics of the system are modeled as an $M/M/N$ multiclass system. The result from their study indicated that as the number of agent's increases, the system's load approaches its maximum processing capacity.

2.2.1 Service times

Brown et al. (2005) find the lognormal provides an excellent fit to data, especially after excluding short service times. The excellent fit of the lognormal was also present after conditioning: for all types and priorities of customers, for individual agents, for different days of the week, and for all times of the day. A positive implication is that one can apply standard estimation techniques to relate (regress) log (service time) to various covariates, i.e., observed information, with obvious modelling benefits. Garcia et al (2012) noted that as time spent on queue at the call centers increases, it becomes unacceptable for customers, and this affect their satisfaction level. The researchers conducted a survey using Univariate Analysis of Variance (ANOVA) to determine customer's perception of their wait experience at call centers. From the result obtained the researchers noted that though the time spent on the queue waiting can lead to customer dissatisfaction. Nevertheless it is not as important as the agent's ability.

2.3 Routing Techniques in a Call Centre

Call routing is the sequence of path taken to convey a customer's call to a service agent. Call routing also known as call distribution relates to a set of rules which are applied to isolate the most appropriate resource for a specific call. Call routing is experience by the customer as being guided through a decision tree (Kook 2007). By progressing through that tree the system provides information to and collects user inputs from the caller. The corresponding realization is often referred to as routing path. However having reached the leaf of the decision tree, the collected information is considered as being sufficiently complete and call distribution takes over to determine the most appropriate agent based on agent properties, user input and system load to route the call.

All routing techniques used in call distribution follows a baseline routing rule which serves as a benchmark for routing calls (Methrotra, 2012). The benchmark routing rule usually followed is the first-come, first serve or longest wait rule. Here the rule states that the first customer to arrive on a queue or the customer that has waited the longest on the queue and it follows the sequence until all calls are attended to. The FCFS/LW is the most popularly used routing rule among the waiting time rule yet the queues experienced in our call centers is still enormous based on customers complains. To this end this study conducted a comparative analysis of existing rules for waiting time routing rule using data from the data base system of a telecommunication organisation in Nigeria. In a bit to determine the optimal routing rule among the waiting time rules for the purpose of queue reduction and low waiting time in our call center, thereby increasing customer satisfaction and brand loyalty.

2.4 Related Work

Customer service call centres have obviously become a very integral part of many organisations' business operations today, inbound call centres employ millions of agents across the globe and serve as a primary customer-facing channel for many different industries. There has also been a great deal of research interest in call centre operations management, with the extensive and evolving literature thoroughly analysed (Mehrotra *et al.*, 2009). This study determines whether average handling time and call resolution are true determinants of operational success of a call centre to reduce waiting queue. It also examine whether emphasis should be on reducing handling time or effective call resolution.

Aksin *et al.* (2007) noted that the operational challenges from call centres provide a perspective on both traditional and emerging call centre management challenges and the associated academic research. They deployed literature review method and identified a handful of broad themes for future investigation while also pointing out several very specific research opportunities.

Given the size of the call centre industry and the complexity associated with its operations, call centres have emerged as a fertile ground for academic research. Operations management have paid comparatively little attention to models and methods for managing routing. However, there are many published papers that describe call routing and resource allocation rules for call centres. Armory and Maglaras (2004) observed that customers in a call service centre experiences real time delay as a result of queue and call back delay. This metric affects customer's perception of the product or service and this impact on customer's loyalty. Probabilistic choice model was deployed, and the dynamics of the system are modeled as an $M/M/N$ multiclass system. The study justifies that as the number of agents increases, the system's load approaches its maximum processing capacity but did not consider the Average handling Time in relation to customer decision, routing rules and system design.

Stanley *et al.* (2008) posited that in a service base call centre, the two key challenges are (i. Where should a call be routed to and) (ii. Who should handle the call?) They deployed base case FIFO approach for the simulation to analyse performance-based routing strategies in call centres. Their work shows the potential for significant improvements in call centre performance especially Average Speed to Answer (ASA). This was achieved by using rules based on historic performance data such as Average call Handling Time (AHT) and first call Resolution (FCR) rates.

Garcia. *et al.* (2012), noted that as time spent on queue at the call centres increases, it becomes unacceptable for customers, and this affect their satisfaction level. A study was conducted using Univariate Analysis of Variance (ANOVA) to determine customer's perception of their wait experience at call centres. Their result showed that though the time spent on the queue waiting can lead to customer's dissatisfaction, nevertheless, it is not as important as the agent's ability. More so, the concept of routing rules to be deployed for efficient call resolution rate was not emphasised.

The quality of service accessibility and customer waiting time are dominant performance measures (Vericourt and Zhou 2005b). Hence capacity planning and call routing software system strive to minimize cost while achieving self imposed service level constraints, though considering low average time waiting in queue. Their work was motivated by the fact that a major European telecommunications service provider discovered that customers needed to talk to more than three different agents before their problems are resolved.

Gans *et al.* (2003) and Aksin *et al.* (2007), conducted study on the concept of customer waiting time on the queue, these researchers focused on queues, staffing and performance analysis which are input into personal scheduling and rostering models. Gans *et al.*, (2010), empirically study the agent's heterogeneity in Average Handling Time (AHT). In a related study by Gong *et al.*, (2015), they modeled a call centre as an $M/M/S+M$ queue which is develop to determine the behavioural queue model in which customers arrive in and depart from the system based on their satisfaction with waiting time. Their model has two sectors, representing the feedbacks of repeat behaviour of customer and abandonment rate. The performance metric of abandonment is the loss of customers based on waiting time; they further explained that the metric of satisfaction with waiting experiences is used to build a link between staffing costs and call centre customer revenues. They considered a call centre model with a single class of customers made up of homogeneous and parallel agents, the analysis of Process-

Related Metrics of Call Center. The model of the abandonment behavior was developed by the extension of the Erlang-A formula, which can be viewed as an M/M/s+M queuing system with feedback.

In reducing the challenges of waiting queues experienced by customers at call centres, our model considered wait-time oriented routing rules. The routing rules were simulated using raw data from the call centre of a telecommunication organization in Nigeria. The simulation enabled a comparative analysis of wait-time oriented rule and the optimal routing rule is recommended to reduce wait-time in call centres and improve customer's satisfaction and brand loyalty.

3.0 Research Methodology

The research was conducted using Global communications as a case for the study. A structured Interview was carried out at Global Communications call centre, Lagos in Nigeria. This was necessary for investigating their mode of operation and possible routing rules been adopted by the call centre. Three (3) personnel were interviewed at the call centre i.e. the database administrator, a call centre agent and a call centre Supervisor. The three personnel were interviewed because of the nature of their job description in the organisation and also interviewing and extracting information from them will be relevant to the study. Interview was also conducted at MTN Nigeria, the national radio frequency analyser manager at Lagos was interviewed, this is to determine the routing rule MTN Nigeria is deploying for call center operations. The outcome of the interviews were relevant for the study, having understood the call centre operations, a further request was made for call centre data from its automated data logging system comprising of agent identity, calls attended to, call handling time, call status, etc. These data were used to test each of the four routing rules to determine their performance. A JAVA simulation program was designed for each of the routing rules using the data collected from Global communication. The result from the simulation gave the optimal rules for wait-time oriented routing rules.

The data collected for this study is automated and machine generated from the Global Communications Call Center data logging system. Data drawn from the organization's electronic database was for a period of one (1) month. This data contained information about the volume of calls received, who handled the calls and how they were handled. The data collected from Global communication database was limited to eight categories of call centre agents including:

1. 121 call Agents
2. General call Agents
3. Pidgin call Agents
4. Igbo call Agents
5. Hausa call Agents
6. Premium call Agents
7. Yoruba call Agents
8. Sim registration call Agents

3.1 System Design and Analysis

The study deployed several tools for design, Unified Modeling Language (UML), the use case diagram was used to describe each of the actors and their respective function in the system. Mathematical rules or routing rule was formulated and adapted for each of the four rules been simulated and a collection of Java simulation programs was used to develop four different programs for each of the routing rules. This is because the procedure for each of the routing rule varies from one another.

3.1.1 Use-Case Diagrams

Figures 1 – 3 shows the main actors in our call centre study which comprises of the customer, call centre agents and the system itself. Customers essentially make specific call types to call centres to make complain or enquiries. The call is considered to be a new (fresh call) or a call back. On receiving the call, call centre agent requires the customer location and phone number for record and authentication purposes.



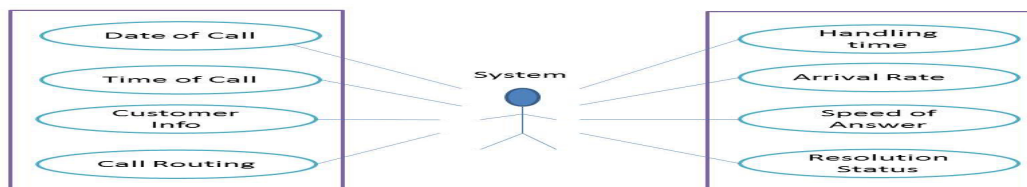
Figures 1: Customer actor and Attributes

Call centre agents are saddled with the responsibility of attending to customer issues. Due to the volume of customer calls, most call centres employ multiple agents to attend to customer inquiries and challenges. Every call centre agent has a unique identification number which helps managers to monitor the progress of each agent and for regular appraisal. Call centres have agent groups who comprises of agent with special trained skill set for handling specific problems ranging from device platform issues to service related issues. The service rates of agents are also recorded. Call centre agents are expected to observe that they are logged into the system and that the system is recording call data such as call date, call time, etc.



Figures 2: Call Centre Agent actor and Attributes

The use case in Figure 3.3 shows the system functions as it is responsible for routing customer calls to available agents using predefined routing rules. These rules will be highlighted later in subsequent section. The system also records the handling time for each call, each call arrival rate, resolution status, speed of answer and other information about the calls. Which are required for computational analysis in order to test the viability of the routing rules.



Figures 3: Call Centre System Actor and Attributes

3.2 Model Approach

In this model, we consider multiple call types (indexed by $i = 1, 2 \dots y$) and multiple agent groups (indexed by $j = 1, 2 \dots z$). Calls of type i arrive at a rate of λ_i . There are n_j agents in group j , with $n_j \in \mathbb{Z}^+$ and each agent in group j serves call type i with rate μ_{ij} . Here we allow agents to handle only a subset of all the call types. If agent group j is not capable of handling call type i then $\mu_{ij} = 0$. When $\mu_{ij} > 0$ we say there is a “match” between call type i and agent group j . In addition, we assume independence of past history each agent of group j has a resolution probability for each call of type i of $p_{ij} \in [0, 1]$.

In the routing rules, $Q_i(t)$ represents the number of type i customers waiting for service at time t and $f_j(t)$ be the number of available agents of type j who are free at time t , where $0 \leq f_j(t) \leq n_j$, for all j, t .

Formally, we use the term “routing rule” to mean both the logic that determines to which agent group an arriving call is assigned if there are no calls in queue and agents from multiple groups are free as well as the logic that

determines which call an agent is assigned to handle when he/she becomes free when calls from more than one type are in queue waiting for service.

3.2.1 Definition and evaluation for waiting time Routing Rule

As adapted from Mehrotra et al. (2012), our benchmark routing rule will be the First-Come-First-Served/Longest-Wait (FCFS/LW) rule, which we specify with the rules as follows.

(1) First Come First Serve/ Longest Waiting (FCFS/LW): When a call arrives and finds no calls of that type in queue and agents of one or more matching group available assigns that call to the agent who has been free the longest, regardless of his/her group.

Let $Q_i(t)$ represents the number of type i customers waiting for service at time t and

Let $f_j(t)$ be the number of available agents of type j who are free at time t ,

Where $0 \leq f_j(t) \leq n_j$, for all j, t .

Let Multiple call types be indexed by $i = 1, 2 \dots I$ and

Let Multiple agent groups be indexed by $j = 1, 2 \dots J$.

Calls of type i arrive at a rate of λ_i .

There are n_j agents in group j , with $n_j \in \mathbb{Z}^+$

Each agent in group j serves call type i with rate μ_{ij}

/Here we allow agents to be trained to handle only a subset of all the call types/

If agent group j is not capable of handling call type I then $\mu_{ij} = 0$

When $\mu_{ij} > 0$ we say there is a “match” between call type i and agent group

In addition, we assume independent of past history each agent of group j has a resolution probability for each call of type i of $p_{ij} \in [0, 1]$.

When an agent of group j becomes free, assign that agent to the call that, among all matching call types, has been waiting the longest regardless of its type. Similarly, if a call arrives and finds no calls of that type in queue and agents of one or more matching group available assigns that call to the agent who has been free the longest, regardless of his/her group.

Below, we introduce several other routing rules whose performance we will compare to that of FCFS/LW.

3.2.2 Waiting-Time Routing Rules

When the system is in a state with multiple routing options – more than one idle server available from the point of view of an arriving customer, or more than one waiting customer available to be served from the point of view of a ready agent the calls are routed such that no call will go unanswered if there are matching agents available (Mehrotra et al 2012).

(2) Fastest Call First Rule (FCF): A call of a particular type that arrives when agents of multiple matching groups are free will be routed to a matching agent group that has the highest service rate for that call type.

Let $Q_i(t)$ represents the number of type i customers waiting for service at time t and

Let $f_j(t)$ be the number of available agents of type j who are free at time t ,

Where $0 \leq f_j(t) \leq n_j$, for all j, t .

Let Multiple call types be indexed by $i = 1, 2 \dots I$ and

Let Multiple agent groups be indexed by $j = 1, 2 \dots J$.

Calls of type i arrive at a rate of λ_i .

There are n_j agents in group j , with $n_j \in \mathbb{Z}^+$

Each agent in group j serves call type i with rate μ_{ij}

/Here we allow agents to be trained to handle only a subset of all the call types/

If agent group j is not capable of handling call type I then $\mu_{ij} = 0$

When $\mu_{ij} > 0$ we say there is a “match” between call type i and agent group

In addition, we assume independent of past history each agent of group j has a resolution probability for each call of type i of $p_{ij} \in [0, 1]$.

When an agent of group j becomes free, select a call of type i ,

Where $i = \operatorname{argmax}_{i: Q_i(t) > 0} \{ \mu_{ij} \mid \mu_{ij} > 0 \}$;

/therefore an agent coming free will choose the matching call type for which he/she has the highest service rate/

If an arriving call of type i find no calls of that type waiting for service and agents of one or more matching group available select an agent of group j

Where $j = \operatorname{argmax}_{j: f_j(t) > 0} \{ \mu_{ij} \mid \mu_{ij} > 0 \}$;

/that is, a call of a particular type that arrives when agents of multiple matching groups are free will be routed to a matching agent group that has the highest service rate for that call type

(3) Shortest Service Time First (SSTF): A call of a particular type that arrives when agents of multiple matching groups are free will be routed to a matching agent group that has the relative Shortest Service Time for that call type.

Let $Q_i(t)$ represents the number of type i customers waiting for service at time t and

Let $f_j(t)$ be the number of available agents of type j who are free at time t ,

Where $0 \leq f_j(t) \leq n_j$, for all j, t .

Let Multiple call types be indexed by $i = 1, 2 \dots I$ and

Let Multiple agent groups be indexed by $j = 1, 2 \dots J$.

Calls of type i arrive at a rate of λ_i .

There are n_j agents in group j , with $n_j \in \mathbb{Z}^+$

Each agent in group j serves call type i with rate μ_{ij}

/Here we allow agents to be trained to handle only a subset of all the call types/

If agent group j is not capable of handling call type I then $\mu_{ij} = 0$

When $\mu_{ij} > 0$ we say there is a “match” between call type i and agent group

In addition, we assume independent of past history each agent of group j has a resolution probability for each call of type i of $p_{ij} \in [0, 1]$.

When an agent of group j becomes free, select $\text{argmax}_{i: Q_i(t) > 0} \{ \mu_{ij} - \max_{k \neq j} \mu_{ik} \mid \mu_{ij} > 0 \}$

/that is, an agent coming free will choose the matching call type for which she has the highest relative service rate/

Similarly, if an arriving call of type i finds no calls of that type waiting for service and agents of one or more matching groups available, select an agent of group j ,

Where $j = \text{argmax}_{j: f_j(t) > 0} \{ \mu_{ij} - \max_{k \neq j} \mu_{ik} \mid \mu_{ij} > 0 \}$

/that is, a call of a particular type that arrives when agents of multiple matching groups are free will be routed to a matching agent group that has the highest relative service rate for that call type/

(4) Highest Service Time First (HSTF): A call of a particular type that arrives when agents of multiple matching groups are free will be routed to a matching agent group that has the highest Service Time for that call type.

Let $Q_i(t)$ represents the number of type i customers waiting for service at time t and

Let $f_j(t)$ be the number of available agents of type j who are free at time t ,

Where $0 \leq f_j(t) \leq n_j$, for all j, t .

Let Multiple call types be indexed by $i = 1, 2 \dots I$ and

Let Multiple agent groups be indexed by $j = 1, 2 \dots J$.

Calls of type i arrive at a rate of λ_i .

There are n_j agents in group j , with $n_j \in \mathbb{Z}^+$

Each agent in group j serves call type i with rate μ_{ij}

/Here we allow agents to be trained to handle only a subset of all the call types/

If agent group j is not capable of handling call type I then $\mu_{ij} = 0$

When $\mu_{ij} > 0$ we say there is a “match” between call type i and agent group

/In addition, we assume independent of past history/

Each agent of group j has a resolution probability for each call of type i of $p_{ij} \in [0, 1]$.

When an agent of group j becomes free, select a call of matching type i ,

Where $i = \text{argmax}_{i: Q_i(t) > 0} \{ p_{ij} \mu_{ij} \mid \mu_{ij} > 0 \}$;

/that is, an agent coming free will choose the matching call type for which she has the highest effective service rate/

Similarly, if an arriving call of type i find no calls of that type waiting for service and agents of one or more matching groups available select a matching agent group j

Where $j = \text{argmax}_{j: f_j(t) > 0} \{ p_{ij} \mu_{ij} \mid \mu_{ij} > 0 \}$

/that is, a call of a particular type that arrives when agents of multiple matching groups are free will be routed to a matching agent group that has the highest effective service rate for that call type/

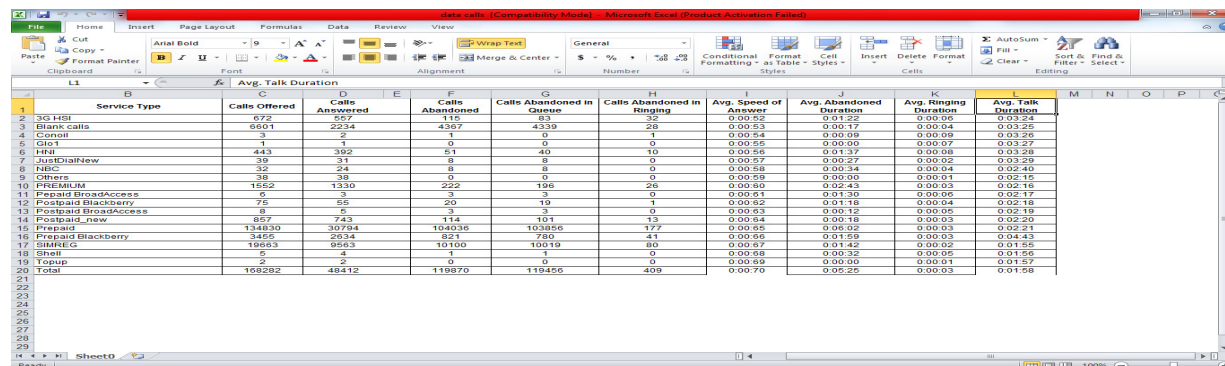
4.0 Simulation Platform

The simulation platform consists of a collection of programs that invoke the simulation library. The library contains all the functionality required to run complex discrete-event simulations of contact centre. Following every service event, the program generates a uniform random variable and compares it to the agent’s Average Speed for Answer to determine if dropped calls dues to undue waiting on the queue. The input data in Table 1,

shows the various service type, number of calls offered, analysis of the number of calls answered, abandoned, average speed of answer, average talk duration and other report from the calls offered.

Simulation run length: For each of the rules described, we simulated for 2000 calls for a period of one (1) hour for some realizations

Table 1: Input call type data for simulation



Service Type	Calls Offered	Calls Answered	Calls Abandoned	Calls Abandoned in Queue	Calls Abandoned in Ringing	Avg. Speed of Answer	Avg. Abandoned Duration	Avg. Ringing Duration	Avg. Talk Duration
3G HSE	672	557	115	83	32	0.0052	0.0122	0.0006	0.0324
Blank calls	6601	2234	4367	4369	28	0.0053	0.0017	0.0004	0.0325
Connet	3	2	1	0	1	0.0054	0.0009	0.0009	0.0326
Glo1	1	1	0	0	0	0.0056	0.0009	0.0007	0.0327
HFI	443	392	51	40	10	0.0056	0.0137	0.0008	0.0328
JuziCallNew	39	31	8	8	0	0.0057	0.0027	0.0002	0.0329
NFC	32	24	8	8	0	0.0058	0.0034	0.0004	0.0240
Others	38	38	0	0	0	0.0059	0.0000	0.0001	0.0216
PREMIUM	1652	1330	222	196	26	0.0040	0.0243	0.0003	0.0216
Prepaid BroadAccess	5	3	2	3	0	0.0051	0.0130	0.0005	0.0217
Postpaid Blackberry	75	55	20	19	1	0.0042	0.0118	0.0004	0.0218
Postpaid BroadAccess	8	5	3	3	0	0.0053	0.0012	0.0005	0.0218
Postpaid_new	857	743	114	101	13	0.0044	0.0018	0.0003	0.0220
Prepaid	134839	30784	104056	103856	177	0.0045	0.0602	0.0003	0.0221
Prepaid Blackberry	3455	2634	821	780	41	0.0046	0.0159	0.0003	0.0443
SIMREG	19863	9583	10100	10019	80	0.0047	0.0142	0.0002	0.0158
Smart	2	2	0	1	0	0.0048	0.0032	0.0005	0.0156
Topup	2	2	0	0	0	0.0049	0.0000	0.0001	0.0157
Total	165282	48412	119870	118450	408	0.0049	0.0522	0.0003	0.0158

4.1 Implementation Interface

A standalone application was developed using Java libraries. The screenshots showing the simulation processes developed in the Java program are shown in Figures 4 – 7 in the Appendix.

5.0 Results and Discussion

The simulation process and a comparative analysis was conducted for the four routing rules to determine the optimal rule that will give the lowest ASA for queue reduction in call center. Table 2 presents the result for various waiting-time oriented routing rules as well as the benchmark FCFS/LW rule.

Table 2: Weighted Average Results for evaluation obtained from simulation Analysis

ROUTING RULE	ASA (seconds)	CALL BACKS	% Call backs
FCFS/LW	47	0.124444444	20.74074074
FCF	36	0.088055556	14.67592593
SSTF	28	0.018055556	3.009259259
HSTF	95	0.203333333	33.88888889

While each of the waiting-time rules results in significantly lower ASA values than the FCFS/LW value except for HSTF (95) seconds. SSTF has the lowest ASA of (28) seconds, this makes it the optimal rule among the waiting time oriented routing rule. The focus of the FCFS/LW rule is clearly on getting calls out of the system as quickly as possible though not as fast as FCF and SSTF, which translates to a significant gap in customer satisfaction and loyalty. We note that the higher the ASA rate the more it is reflected in an increase in system congestion that drives up the mean waiting time under the FCFS/LW rule.

The graph in Figure 8 expresses the result for ASA for each of the waiting time routing rule. From the result of the simulation analysis SSTF routing has ASA of 28 seconds which implies that SSTF performed optimally than other rules, having the lowest waiting time on the queue.

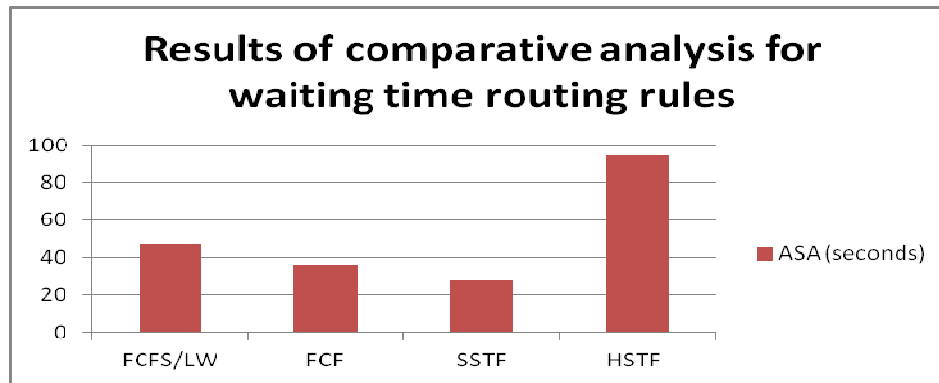


Figure 8: Weighted Average Results for evaluation obtained from simulation Analysis

6.0 Contribution to knowledge

This paper has expanded the frontiers of knowledge because the result from the simulation conducted shows a significant shift in paradigm. Basically queuing system usually follow a FCFS/LW pattern, the FCFS routing rule remove calls from the queue as soon as possible. The Shortest Service Time First (SSTF) does not only handle queue reduction but particularly, considers the highest relative service rate of an agent for a particular call type. When the service rate (μ) is relatively high, the queue in call centre will be reduced faster than any other routing rule. Therefore, we propose a more effective and optimal routing rule for call centre management.

7.0 Conclusion

From literature it has been established that call center generally experience enormous queues this is largely depending on the nature of service deployed by the call center. The type of routing rules used by the call center is also a major factor that impacts on the nature of the services delivered. Hence, this study conducted a comparative analysis of existing routing rule for waiting time routing rule to determine which performs optimally. A collection of JAVA program was deployed to conduct the experiment, java programs was written for each of the four routing rule and data collected from Global communication call center was used to carry out the simulation analysis. The result from the simulation shows that SSTF routing rule is the optimal rule for reduction of queue in call center operations. While there has been a significant amount of research on skill-based routing and agent pooling, to our knowledge this research has not considered the impact of such rules on ASA rates when different agent groups have different Average Handling Time (AHT), this can therefore be a further call for research in this domain.

References

- Aksin .Z, Armony .M. and Mehrotra .V. (2007) The modern call-centre: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, 16(6):665–688, November–December Available at <http://www.stern.nyu.edu/om/faculty/armony/research/CallCentreSurvey.pdf>. (accessed June 2015)
- Armony .M (2005), Dynamic routing in large-scale service systems with heterogeneous servers. *Queueing Systems*, 51(3-4):287–329, December 2005.
- Armony, M. and Maglaras. C. (2004) On Customer Contact Centers with a Call-Back Option: Customer Decisions, Routing Rules, and System Design. *OPERATIONS RESEARCH* Vol. 52, No. 2, March–April 2004, pp. 271–292 ISSN 0030-364X _ EISSN 1526-5463 _ 04 _ 5202 _ 0271
- Avramidis, N, A. Deslauriers, L’Ecuyer, P. (2004). Modeling daily arrivals to a telephone all center. *Management Science* 50(7) 896–908.
- Brizola, N, Costa .S, Pazeto .T, and Freitas P. (2001). Planejamento de Capacidade de Call Center. In : ICIE, Flo-rianoópolis
- Brown, L., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. (2005).

- “Statistical Analysis of a Telephone Call Center: A Queueing-science Perspective.” *Journal of the American Statistical Association* 100(469):36-50.
- Enyioko, N. C. (2016), Relevance of the Queueing Theory to Serviced - Based – Organisations SSRN eLibrary Search Results Service Management eJournal, Medonice Consultiong and Research Institute Date Posted: April 01, 2016 Working Paper Series, [Subshttp://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=992385](http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=992385)
- Gans G. Koole, and Mandelbaum A. (2003), Telephone call centres: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, 5(2):79–141, Spring .
- Gans G, Liu N, Mandelbaum A, Shen H, and Ye H. (2010). Service times in call centers: Agent heterogeneity and learning with Powering Applications—A Festschrift for Lawrence D. Brown, IMS Collections, Institute of Mathematical Statistics, Beachwood, OH, Vol.6(4) 99–123.
- Garcia. D, Archer .T, Moradi .S, and Ghiabi .B (2012), Waiting in Vain: Managing Time and Customer Satisfaction at Call Centers. *Science Research*, <http://dx.doi.org/10.4236/psych.2012.32030>. *Psychology* 2012. Vol.3, No.2, 213-216. Published Online February 2012 in SciRes <http://www.SciRP.org/journal/psych>
- Gong J, Yu M, Tang .J, and Li .M. (2015), Staffing to Maximize Profit for Call Centers with Impatient and Repeat-Calling Customers. *Mathematical Problems in Engineering* Volume 2015, Article ID 926504, 10 pages. Hindawi Publishing Corporation. <http://dx.doi.org/10.1155/2015/926504> (accessed January 2015).
- Maister, D. (1985). *The psychology of waiting lines in the service encounter: managing employee/customer interaction in service businesses*, J. A. Czepiel, M. R. Solomon, C. F. Suprenant. (eds.), D. C. Heath and Company, Lexington Books, Lexington, Massachusetts.
- Mehrotra .V, Ross .K, Ryder .G and Zhou .Y (2009), Routing to Manage Resolution and Waiting Time in Call Centers with Heterogeneous Servers. *Manufacturing & Service Operations Management* 9(4): 167-181, ISSN 1523-4614 j EISSN 1526-5498
- Mehrotra .V, Ross .K, Ryder .G and Zhou .Y (2012), Routing to Manage Resolution and Waiting Time in Call Centers with Heterogeneous Servers. *Manufacturing & Service Operations Management* Vol. 14, No. 1, Winter 2012, pp. 66–81 ISSN 1523-4614 (print) . ISSN 1526-5498 (online) <http://dx.doi.org/10.1287/msom.1110.0349> ©2012 INFORMS
- Parson, L.C. (2016) *Introduction to Queueing Theory*. New York: Mathworks Inc. P.4.
- Sharma (2010) “*Operation Research, theory and application*” 4th edition Macmillian publishers.
- Stanley .J, Saltzman R and Mehrotra V (2008), Improving call center operations using performance-based routing strategies. *CJOM*, 6(1): 24-32.
- V’ericourt P and Zhou Y (2005b). Managing Response Time in a Call-Routing Problem with Service Failure. *OPERATIONS RESEARCH INFORMS* Vol. 53, No. 6, November–December 2005, pp. 268–281 issn 0030-364X _ issn 1526-5463 _ 05 _ 5306 _ 0968 (Accessed June 2015)

Appendix

Screenshots of Simulation Process

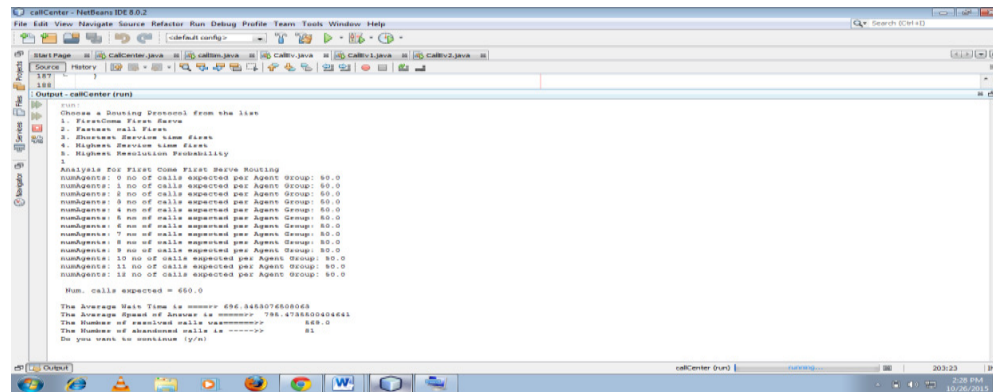


Figure 4: Screen shot of simulation analysis using: First Come First Serve Routing Rule

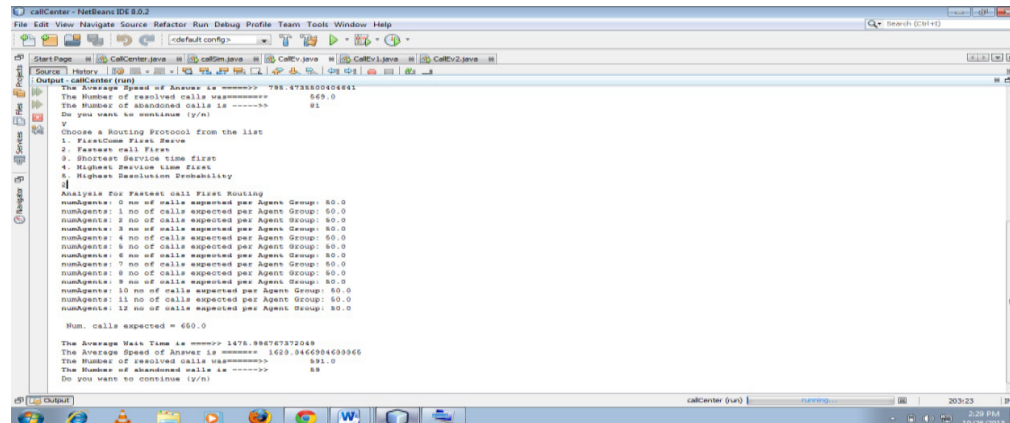


Figure 5: Screen shot of simulation analysis using Fastest Call First Routing

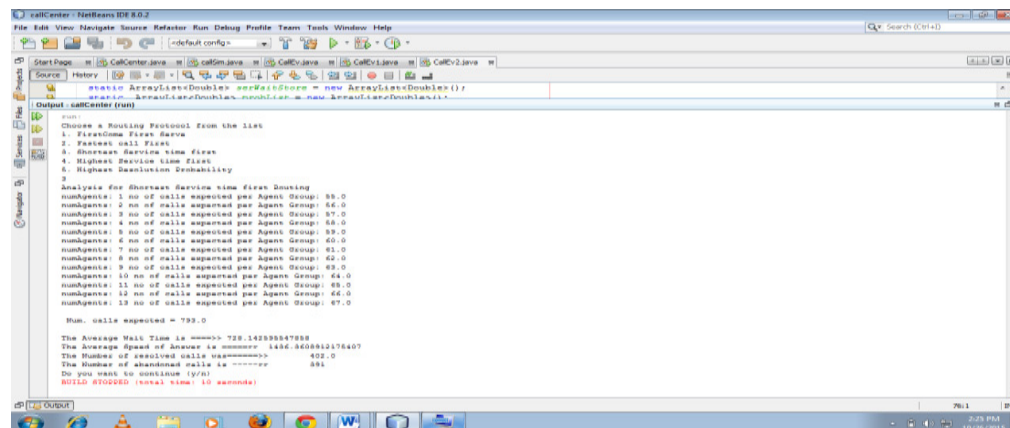


Figure 6: Screen shot of simulation analysis using shortest service time routing

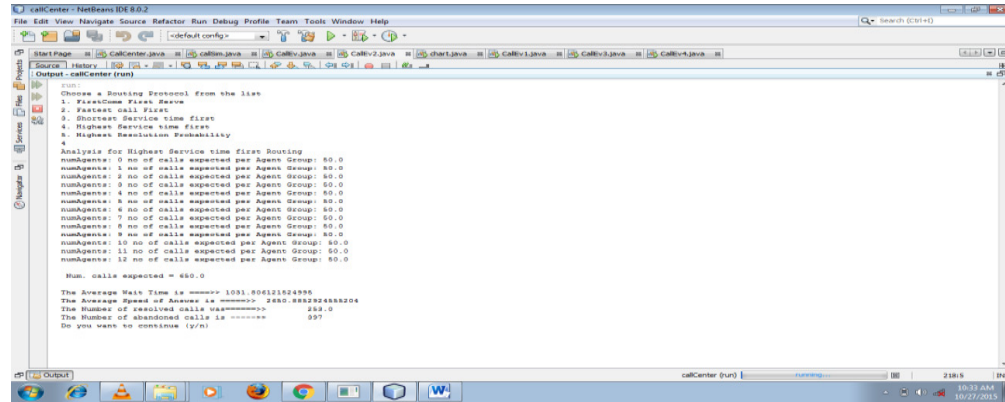


Figure 7: Screen shot of simulation analysis using Highest Service Time First Routing