Transactions of the Nigerian Association of Mathematical Physics
Volume 11, (January - June, 2020 Issue), **pp75 – 80**

OPTIMISATION OF CALL CENTER USING GRAPH THEORY ANALYSIS

Mughele Ese Sophia[1] and Konyeha Susan[2]

[1]Department of Computer Science, School of Science and Technology Delta State School of Marine Technology Burutu, Warri Delta State, Nigeria.
[2] Department of Computer Science, University of Benin, Benin City, Edo State, Nigeria.

Abstract

We studied wait time and call resolution rate oriented routing rules, and established that a hybrid routing rule will enhance performance in call center. There is also the problem associated with effective utilization of the call center system in terms of service rate and waiting time of agent group in a call center. The authors developed a Conceptual Framework for Graph Theory Analysis, in expressing graph in a matrix form; vector y satisfies the equation $Ax = y$ from this equation, mathematical deductions were derived to establish that the features of the optimization rule can enhance the effective utilization of call centers operational system. Maximal service rate (μ) can be obtained and agents waiting time or idle time reduced.

Keywords: Call Center, Optimization, Graph theory, Adjacency matrix, Wait-time,

1. Introduction

In today's fast growing economy, satisfying customers need has become a phenomenon that is highly inevitable for business that wants to survive in this era of high competition amidst the global financial crisis [1].Whenever a customer initiates a call to a call center he encounters the following scenarios namely: the time spent waiting for the service and the service itself [2]. Call centers 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. Call centers have a significant economic importance in today's world, as explained in [3], [4], and [5], for example. Managers face complex optimization problems where the goal is to meet various constraints on quality of service (QoS) at the least possible cost, or optimize a given performance measure (PM) for a given budget. Most of these costs are actually the earnings of the call center agents. In [6], it stated that,for a call center to reduce waiting queue with emphasis on the reduction of time spent, it is 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.

In [7], it was stated that for a call center to reduce waiting queue, emphasis should be on the service itself, that is; call resolution. It is best to route calls to agents who resolve customer issues, sometimes holding a call in queue to wait for such agent. This might also lead to increase in congestion and undue burden on some agents. Once a customer received service from a contact center agent regarding a particular problem, a subsequent call from such a customer regarding that same problem is a clear indication that the problem was not properly resolved in the previous service encounter. The agent's inability to resolve this problem will obviously lead to call-back and customer' dissatisfaction [8].The ideology of this paper is based on the work of [9], the authors developed a hybrid framework which enabled the optimization of wait-time as well as increased resolution rate of the agent in a call center. The hybrid rule was formalized using a set of mathematical equations predicate upon the average handling time (AHT), which is the service time and the resolution ability (the skill of the agent) of the agent for a call type. This paper simply extended the works of [10], [11] and [9]. This paper specifically addresses the problem associated with the utilization of call center system in terms of service rate, waiting time of agent in an agent group. The number of call type i handled by a call center is determined by the maximal service rate of the agents in agent group j.

Correspondence Author: Mughele E.S., Email: prettysophy99@gmail.com, Tel: +2348035859666

2. Call Centers Operations

Call center was defined in [12], as 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. In [13],Call center was defined 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 centers are units that manage an organization communication system. Call centers are known by a variety of names namely: contract center, customer service center, customer interaction center, customer service point. Customer service call centers have obviously become a very integral part of many organizations business operations today, inbound call centers 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 center operations management, with the extensive and evolving literature thoroughly analyzed[14]. This study determines whether average handling time and call resolution are true determinants of operational success of a call center to reduce waiting queue. It also examine whether emphasis should be on reducing handling time or effective call resolution.

The operational challenges from call centers provide a perspective on both traditional and emerging call center management challenges and the associated academic research[4]. 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 center industry and the complexity associated with its operations, call centers have emerged as a fertile ground for academic research. Also [15],provides a complete review of articles on First Call Resolution (FCR), while also pointing out the importance of measuring and using FCR. Resolving customer queries the first time around is a commonly shared goal. A company's business context, human resources strategy, supporting technology and budget constraints influences this KPI in many ways, and makes FCR a difficult measure to benchmark. The study established the differing views on the value and measurement of FCR, identifies the main factors affecting FCR and the relationships among these factors, and relates results in a South African context to academia and practitioner.

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 centers. In [16],it was observed that customers in a call service center 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.

Consider an inbound multi-skill call center in [7], where arriving calls initiated by the customers are categorized by the type of service that they require, the call type. Agents, or customer sale representatives, are partitioned into agent groups, where all agents in each group are trained to answer the same subset of call types, called their skill set. Also [17], worked on determining staffing level on a weekly basis to ensure customer's satisfaction and meeting their needs as well as reducing service cost. They forecasted the system's load which involved predicting both arrival counts and average service time, this enabled them determine staff weekly schedule. The method deployed was the use of arrival count model which is based on a mixed Poisson process approach. Their model also considered billing on the arrival process and they demonstrated this event as an exogenous variables, the model was used to achieve a balance in precision between the forecasting values of the arrival counts and average service time. The authors determined the practical performance of the model using Quality-Efficiency Driven (QED) regime. Their work was able to achieve a desired precision level during most working hour of the day. They tested their model using Israeli Telecommunication Company. Consequently [18], used reinforcement hearing (RH) to select the optimal prompts to maximize the success rate in call routing application. They deployed the use of simulation for dialogue and it was used to experiment with different scenarios, the authors demonstrated that RH can make these systems more robust, the system basically learns from experience through automated voice system. In [7],work was done on routing policies using multiple call types and multiple agent groups. The variable were call type and agent group using the longest waiting time for a call type as against the longest idle time or the number of idle agent in the agent group. The authors optimized the coefficient of the two variable evaluated. They also considered functions responsible for service level, the abandonment ratios and the fairness of occupancy across agent groups. They deployed simulation-based optimization heuristics, establishing that simulation model for the call center can be more detailed and realistic than the model used in literature to study the optimality of certain types of routing rule. The model was based on realistic routing rule which considered the complexities of the real life call center. The developed framework in [7], for the optimization of routing rules used first-come-first serve (FCFS) among wait-time routing rule and Probability Routing (PR), among call resolution routing rule to develop a hybrid rule. In the bid to minimized wait-time and maximize call resolution rate.

In [1], a performance evaluation was further conducted on the three call resolution oriented routing rules discussed by [19]. The authors deployed discrete event driven simulation to evaluate the performance of these rules to determine the optimal routing rules among the resolution rate routing rule. The result from their study showed that Shortest Queue Routing (SQR) was the optimal; they posited that SQR is based on the ability and skill of agent for that particular call type. In a related study by [2], a comparative analysis of waiting time routing rules for queue reduction in call center was conducted, in a bid to determine the optimal routing rule among the available routing rules for queue reduction. The authors also implemented the policy using discrete event driven simulation for the comparison. The result from their study proves that Shortest Service Time First (SSTF) routing rule is the optimal rule. They noted that SSTF is based on the time it takes an agent to accomplish a call time which is the service time. In [8], in their work based on [1] and [2] did a simulation model that can be used for the optimization of a call center. They deployed discrete event driven simulation to optimize the routing rule among call resolution orienting rule which is indicated to be Shortest Queue Routing (SQR)[1]. The optimal for wait-time oriented routing rule as SSTF [2]. The result from the simulation for both SSTF and SQR performed better when compared with other routing rules. Also [9], further developed a framework based on the result from [8]. The framework was implemented using mathematical principles to demonstrate that SSTF and SQR routing rule minimizes wait-time and maximizes call resolution simultaneously. Again [2], in a related study expanded the work of [1], to hybridize the framework as a single routing rule that is able to solve Min/Max problem simultaneously in call center.

This paper proposes a graph theory analysis for the hybrid model formulation of [9]. The graph is expressed using an adjacency matrix form, where vector y satisfies the equation Ax = y. Here A is a block of diagonal with i blocks of A' (Transpose of A). The matrix form establishes that the transpose of A, and element (i, j) of A' is 1 if agent in group j can handle call type i, 0 if agent group are unable to handle call type i.

3. Methodology

The data and design deployed for this study was adopted from [2]. Data used for this study is obtained from Nigerian Telecommunication Company.

The data collected for this study are automated and machine generated from Nigerian's call Center's data logging system. The data obtained from the call center is made up of 60 fields. The data collected from the Telecommunication Company in Nigeria was limited to eight categories of call center 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

Data was also obtained for the entire agent groups for 1month for all 8 agent group comprising of call arrival rates, average handling time for call type in an agent group, service time etc. Depicted in figure 1, is the conceptual framework for the graph theory, which is used to solve the problem of effective utilization of call center systems in terms of service rate and idle time of agents in agent group

3.1 Conceptual Framework for Graph Theory Analysis

Expressing graph in matrix form, vector y satisfies the equation $Ax = y$

where A is block diagonal with i blocks A' (transpose of A), and element (i, j) of A' is 1 if agent in group j can handle call type i, 0 otherwise [10].

Graph theory analysis of model formulation

A graph theory representation of the optimal hybrid model is stated as follows

$(G) = X \cup Y \text{ where}$

$X = \{x_1, x_2, \dots, x_i\} \text{ is the set of Call types}$

$Y = \{y_1, y_2, \dots, y_j\} \text{ is the set of Available agents handling various call types}$

and an edge joins call type x_i to agent y_i if and only if call type x_i can be handled or resolved by Agent y_i. The problem is to determine whether G has a matching which saturate agent Ys. Figure 1 shows the problem graph and it can be represented in an adjacency matrix shown in Table 1

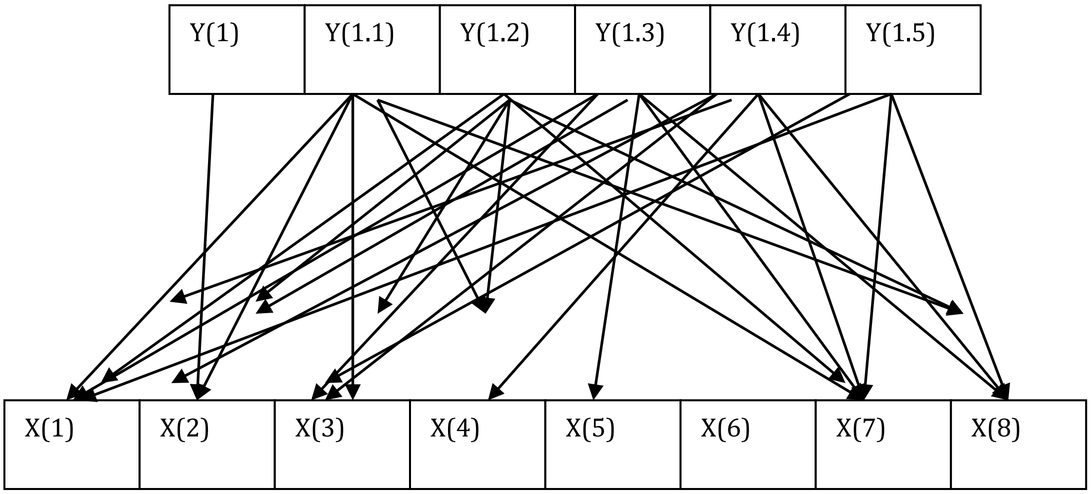

Figure 1: Graph showing interactions between agent group 1s and the various call types Xs

Note: X(1), X(2)..X(8) represent 121 call, General call, Pidgin call, Igbo call, Hausa call, Premium call , Yoruba call and Sim registration call, the eight call types collected from call center (field study 2016).

Table 1: Adjacency matrix Agent 1s and the various call types Xs

	X(1)	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)
Y(1)	0	1	0	0	0	0	0	0
Y(1.1)	1	1	1	1	0	0	1	1
Y(1.2)	1	1	1	1	0	0	1	1
Y(1.3)	1	1	1	0	1	0	1	1
Y(1.4)	1	1	1	1	0	0	1	1
Y(1.5)	1	0	1	0	0	0	1	1

The matrix in Table 1 can be solved as an assignment problem using mathematical programming technique (dynamic programming and Integer Linear programming) or Breath-First Algorithm to determine the optimal solution. Table 1, represent an edge or arc between Xs and the Ys while 0 indicates no link or arc.

4. Results and Deductions from Graph Analysis

As calls arrive into the call center, features or attributes of the calls are extracted based on the SL of agent group designed to resolve the specific call type. The services rate (µ) may be exponential but lognormal which often fits well and are determined via iterative steady state approximations of SL are preferred. The SL is determined by the cost of agent in group j, number of call types routed to an agent in a group and number of agents in that group capable of handling the call type. These attributes are used to compute the service rates of individual agent and agent groups as shown in equation (1)

Service rate of Agent group j for call of type i (u_{ij})

$= \frac{1}{AHT_j} \equiv \frac{1}{\lambda_T}\sum_{i=1}^{j}\frac{\lambda_i}{u_i}$ (Gans*et al.*, 2010) (1)

where $\mu_{ij}$is the service rate of agent group j for call type i, AHT_j is the mean call handling time of all agents in group j and μ_i is the service rate for call type i

The utilization of the system in term of the service rate and waiting time of agent group and number of call type is determined using equation (2)

Maximal service rate of the Agent group j $(\rho_j) = \sum\frac{\beta_{ij}}{\mu_{ij}} = \frac{\left(\sum_i \frac{\beta_i x_{ij}}{\mu_{ij}}\right)}{n_j}$ (2)

The condition for stability is ρ_j<1 that is to say that the mean total arrival rate must be less than the mean.ρ_j

The model also tries to address call types mishandling by constraining the utilisation associated with each call type i. This implies each call type i must be served at total utilisation boundary (). To determine this constrains, the utilization associated with call type I is first determined which require us to calculate the effective service attention given to calls of type i from all agent groups. The total fraction (TF_i) of time spent serving queue i and the total service rate of call type i is given in equations (3) and (4) respectively.

$$TF_{i=}\frac{\frac{\lambda_{ij}}{u_{ij}}}{\sum_{i=1}\frac{\lambda_{i'j}}{u_{i'j}}} \quad (3)$$

Total service rate of call type i $(\mu_z) = \sum_{i=1}^{j} n_j u_{ij}(TF_i)$ (4)

whereμ_z is the total service rate of call type i, $\lambda_{i'j}$ is the arrival rate excluding call type i.

Then, the total effective utilisation associated with call type i is calculated using equation (5).

$$\Gamma_i = \frac{\beta_i}{\mu_z} \quad (5)$$

Every agent in a group has a unique cost expressed as follows:

Costs: c = $(c_{1,1}, \ldots, c_{1,j}, \ldots, c_{i,1}, \ldots, c_{i,j})^t$ where $c_{i,j}$ = cost of an agent of group j handling call type i.

Decision variables: x = $(x_{1,1}, \ldots, x_{1,i}, \ldots, x_{i,1}, \ldots, x_{i,j})$ where $x_{i,j}$ = number of call type i routed to an agent j.

Auxiliary variables: y = $(y_{1,1}, \ldots, y_{1,j}, \ldots, y_{i,1}, \ldots, y_{i,j})$ where $y_{i,j}$ = number of agents of call type i in agent group j.

Expressing in matrix form, vector y satisfies the equation Ax = y

where A is block diagonal with i blocks A' (transpose of A), and element (i, j) of A' is 1 if agent in group j can handle call type i, 0 otherwise [10]

The service level (SL) or fraction of call type i answered by agent in group j is computed using equation (6).

$$g_{i,j}(y) = \frac{E(i,j)}{\sum_1^j \sum_1^i E(i,j)} \quad (6)$$

Where $E(i,j)$ $no\ of\ call\ type\ i\ \ routed\ to\ an\ agent\ in\ group\ j,$

$\sum_1^j \sum_1^i E(i,j)$ is total call type i routed to agent group j.

Similarly, the aggregate service level over call type iis the expected total number of calls of type I answered within by group jover the day, divided by the expected total number of calls of type ireceived over the day.

$g_i(y)$ = aggregated SL for call type i

$g_j(y)$ = aggregated SL for agent group j

$g(y)$ = aggregated SL for both call type i and agent group j

With these features and measures, optimization rule is formulated as.

Max $Z = \sum_{ij} \beta_i \theta_{ij} x_{ij}$ (Maximise total call resolution rate) [19]

Min $C'X = \sum_{j=1}^{J} \sum_{i=1}^{I} c_{i,j} x_{ij}$ (Minimise cost of waiting time) [10]

Ax = y

$g_{i,j}(y) >= Min\ (SL_{i,j}) \forall i,j$ (Service level of call type i in agent group j)

$g_i(y) >= Min\ (SL_i) \forall i$ (Service level of call type i estimated by Erlang C)

$g_{,j}(y) >= Min\ (SL_j) \quad \forall j$ (Service level of agent group j estimated by Erlang C)

$g\ (y) >= Min(SL)$ (Overall Service Level of system estimated by Erlang C)

$0 \leq x_{ij} \leq 1 \quad \forall i,j$ (Fraction of calls bound)

$\sum_j x_{ij} = 1 \quad \forall i$ (Total calls routed to different agent group j)

$\rho_{j-} \leq \rho_{j_i} \leq \rho_{j+} \quad \forall j$ (Utilization of each agent group)

$\Gamma_{i-} \leq \Gamma_i \leq \Gamma_{i_+} \quad \forall i$ (Utilization boundaries of each call type)

$c_{i,j}, x_{ij}, \theta_{ij}, x_{ij}, \Gamma_{i-}, \Gamma_{i_+} \geq 0 \quad and\ Integer \quad non\ negativity\ constraint$

These rules will determine effective system utilization in terms of service rate and agent waiting time, which enforces over all call center optimality in quality service delivery.

5. Conclusion

From the deductions above, it established that the features and measures of the optimization rule formulated above with the challenges associated with effective utilization call center operational system can be achieved. At the time maximal service rate (µ) can be obtained and agents waiting time or idle time will also be reduced. The utilization constraints Γ_{i-}, Γ_{i_+} give the call center freedom to moderate calls during service period, the objective function Max Z help resolve problems maximizing CR. Similarly, Min $C'X$ will help moderate the influence of cost, abandonment and excessive queue length on staff performance and systems overall efficiency. It is on this premise that calls are routed to agents who handles call type i with the highest call resolution rate, lowest call waiting time and optimal performance effectively.

References

[1] Mughele, E.S and Chiemeke C.S. (2016), A Performance Evaluation of Call Resolution Oriented Routing Rules to Enhance Resolution Rates. International Journal of Computer Applications. Foundation of Computer Science, USA. 143(8): 0975 – 8887.Available online http://www.ijcaonline.org/

[2] Mughele, E.S, Chiemeke C.S. and Konyeha .S (2017a), A Comparative Analysis of Waiting Time Routing Rule for Queue Reduction in Call Center. Journal Digital Innovations & Contemporary Research in Science, Engineering & Technology. Available online at http://www.isteams.net/digital-innovations-journal. Volume 5. N0 2. pp. 107-122.

[3] Gans G. Koole, and Mandelbaum A. (2003), Telephone call centers: Tutorial, review, and research prospects. Manufacturing & Service Operations Management, 5(2):79–141.

[4] Aksin .Z, Armony .M. andMehrotra .V. (2007) The modern call-center: A multi-disciplinary perspective on operations management research. Production and Operations Management, 16(6):665–688, Available at http://www.stern.nyu.edu/om/ faculty/armony/research/CallCenterSurvey.pdf. (accessed Jan. 2020)

[5] Chan, Koole, and L'Ecuyer (2000). Dynamic Call Center Routing Policies Using Call Waiting and Agent Idle Times Manufacturing & Service Operations Management.Vol. 00 No. 0, pp. 000–000.

[6] Armony M. (2005), Dynamic routing in large-scale service systems with heterogeneous servers. Queueing Systems, 51(3-4):287–329.

[7] MugheleE.S andChiemeke C.S (2017), A Framework for the Optimisation of Routing Techniques in Call Centers" International Conference on Science, Technology, Education, Arts, Management and Social Science (iSTEAMS). conference proceedings Pp 413-424. Caleb University Imota Lagos State 29th – May to 2nd June 2017.

[8] Mughele, E.S and Chete C. Francis (2017), "Simulation Model for Optimization of Call Center" AIMS Research Journal. 3(2): pp.141-221.Available online at http://www.isteams.net/aims-journal.

[9] Mughele E.S, Chiemeke C.S, Konyeha S., Ukaoha K.C and Akpon-Ebiyomare D. (2017b), Hybridized Optimisation Framework for Routing calls in Call Centers. FTC-2017, Future Technologies Conference, Vancouver, Canada 28-29 November 2017. Peered reviewed and accepted for publication in conference proceeding in IEEE Xplore.

[10] L'Ecuyer P. (2004)``Quasi-Monte Carlo Methods in Finance", Proceedings of the 2004 Winter Simulation Conference, IEEE Press, 2004, 1645-1655.

[11] 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.

[12] Brizola N. Costa S. Pazeto T. and Freitas P. (2001). Planejamento de Capacidade de Call Center. In : ICIE, Florianópolis

[13] Avramidis A. N., Deslauriers A. and L'Ecuyer P. (2004). Modeling daily arrivals to a telephone call center. Management Science 50(7) 896–908.

[14] 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

[15] Hart M, Fichtner B, Fjalestad E, and Langley S. (2006) Contact center performance: In pursuit of first call resolution. Management Dynamics, 15(4):17–28.

[16] 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(2): 271–292.

[17] Aldor-Norman, Feigin D.P and Mandelbaum A. (2009) Workload forecasting for a call center: methodology and a case study. The Annals of Applied Statistics 2009, Vol. 3, No. 4, 1403–1447 DOI: 10.1214/09-AOAS255 c. Institute of Mathematical Statistics.

[18] Lewis C. and Di Fabbrizio, G. (2006) Dialogue Strategy Optimization with Reinforcement Learning in an AT&T Call Routing Application, AAAI Workshop on Statistical and Empirical Approaches for Spoken Dialogue Systems, Boston, Massachusetts, USA.

[19] 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, pp. 66–81 ISSN 1523-4614 (print), ISSN 1526-5498 (online)http://dx.doi.org/10.1287/msom.1110.0349.