Global Journals LaTeX JournalKaleidoscope™

Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Call Center Queue Modeling and Analysis

Nwonye Chukwunoso[1]

[1] University of Port Harcourt, Port Harcourt, Nigeria

Received: 6 December 2012 Accepted: 2 January 2013 Published: 15 January 2013

Abstract

The M/M/c/c model is the most widely applied queueing model in the mathematical analysis of call centers. The M/M/c/c model is also referred to as the Erlang Loss System. The Erlang loss model does not take into consideration system attributes such as blocking and busy signals, balking and reneging, retrials and returns. Although, the Erlang loss model is analytically tractable, it is not easy to obtain insight from its results. The need to develop a more accurate call center model has necessitated the modification of the Erlang loss model. In this research, we model and analyze a call center using M/M/c/N the model. The goal of this paper is to extend existing results and prove new results with regards to the monotonicity and limiting behaviour of the M/M/c/N model with respect to the system capacity N.

Index terms—

1 Introduction

he call center industry has grown explosively in the recent past and that has aroused the interest of researchers from different disciplines. Mandelbaum [11] have provided a comprehensive research bibliography with abstracts in diverse disciplines such as Operations research, Statistics, Engineering, and so on. Call Center research has been reviewed in the tutorial and survey paper by Gans et al. [6]. In this paper, our focus is on the computational rigor of the call center performance metrics using the model.

2 a) Description of a Call Center

A call center is a department of an establishment that attends to customers via telephone conversation often for the purpose of sales and product support, or that makes outgoing telephone calls to customers usually for the purpose of advertisement or telemarketing. Suppose the department also attends to e-mails, faxes, letters, and other similar written correspondence, then, it is called a contact center.

Inbound call center only handle incoming telephone calls initiated by customers while out bound call centers only make outgoing telephone calls to customers. There are call centers that deal with both types of calls. In majority of the call centers, inbound calls form the bulk of contacts with customers. In addition, inbound calls are more time consuming compared to other types of contacting options (e.g. e-mails, faxes, or letters) in terms of waiting times in the telequeue or sojourn times. Hence, we will only focus on inbound call centers. In an inbound call center, there is a group of agents (Customer Sales Representatives, CSRs) who provide the needed service through talking to customers on phones. In this paper, we shall use the terms "agents" and "CSRs" interchangeably. Agents are equipped with equipment, such as a Private Automatic Branch Exchange (PABX or PBX), an Interactive Voice Response Unit (IVRU or VRU), an Automatic Call Distributor (ACD), and computers [16]. See Figure ??.1 for details on the operational process and components of an inbound call center. b) The Operational Process of an Inbound Call Center At some point in our lives, we have all called a call center. We will describe the operational process and components of an inbound call center in line with the description in [6,17]. The process is depicted in Figure ??.1. Customers wanting to receive service from a call center, dial a special number provided by the call center. The Public Service Telephone Network (PSTN) company then uses the Automatic Number Identification (ANI) number (the phone number from which the customer dials) and the customer's Dialed Number Identification Service (DNIS) number (the special number being dialed) to connect the customer to the PABX privately-owned by the call center. The telephone lines (usually called trunk lines) connect the PABX to PSTN. If a trunk line is available, the customer seizes it; else the customer receives a busy signal and will be rejected. Hence, this customer is said to be blocked. Once the call is accepted, the customer will be connected through the PABX to the IVRU. The IVRU provides some automatic service for customers as well as several options for customers to choose from. Upon service completion at the IVRU, some customers leave the system and release the trunk lines. If the customer requires the service of an agent, the call will be passed from the IVRU to the ACD. The ACD is a sophisticated instrument designed to route calls to agents based on the specific needs of calls. If no appropriate CSRs are available, the customer is informed to wait and join a queue at the ACD. The customer is said to be delayed. The ACD decides the next customer to get service according to some preprogrammed queueing discipline (usually First Come First Served, FCFS). Delayed customers may decide to hang up and abandon (or renege) before they are served if they perceive that the service is not worth the wait. Such customers are said to be impatient. Patient customers (who do not abandon service) will eventually be connected to an agent. In serving a customer, the CSR works with a PC furnished with Computer-Telephony Integration (CTI), which is the technology that allows interactions on a telephone and a computer to be integrated. CTI will help ACD to route the call, help the CSR to get the caller's information from the database and hence facilitate the service process. At the completion of service and exit of the customer, the CSR still needs some wrap-up time to finish the whole service process and then may be available for the next customer. The service time is the sum of talk time and wrap-up time. Customers who abandoned and were blocked may try to call again after some random amount of time and these calls are referred to as retrials. Customers who finished talking with anagent may also need further assistance and therefore call back. Hence they become return customers or feedback customers. Notice that these two types of customers are not shown in lines occupied, it gets a busy signal and as such is blocked and cannot access the system. If there is an available truck line, the call is either connected to the system and seizes one of the free trunk lines or it balks. Suppose there is an available trunk line and at least a free agent, then the call is immediately serviced.

Otherwise the call experiences delay and has to wait in a queue at the ACD for a CSR to become available. Calls at the ACD may become impatient and abandon (renege) the system before being served and thus release the trunk line. The ACD usually implements the FCFS queueing discipline. Upon service completion by a CSR, the call leaves the system and then releases both the trunk line and the CSR and these resources become available to other arriving calls. Return (or feedback) calls are calls that return after been served by an agent. Some of those calls who do not get served (blocked, abandon or balk) may call again and they become retrials. The remaining calls become lost calls. Suppose that the call arrivals follow a Poisson process with mean rate and that the service times of the calls are independent and identically distributed ( ) exponential random variables with mean .

Then we can model the system as a queueing system with features such as balking, abandonment, retrial, and feedback.

3 d) Performance Evaluation of the Call Center Queueing Model

In this paper, we will ignore features such as balking, abandonment, retrial, and feedback. Following the above assumptions, we will apply the model in analyzing the call center performance. The queueing system has a closed-form solution for the system state (number of calls in the system), the queue length (number of calls in the queue) distribution and waiting time distribution. Then we can obtain system performance metrics such as average waiting time, average queue length, and probability of blocking. We will apply the performance analysis of the queueing system to call center modeling and in turn show new results. The call center performance measures (metrics or indicators) provide useful information in the design and management of call centers. Performance measures are used in determining the service levels (or quality of service) in call centers. Not all queueing models can be analyzed exactly to obtain performance measures as model. For instance, if we include additional features such as Non-Poisson time varying arrival process, balking, abandonment, retrial, feedback, and nonexponential service times, the model may become insolvable using traditional queueing techniques and other techniques have to be used to analyze the model such as simulation modeling.

4 II.

5 Modeling Call Centers as Single-Node Exponential Queueing Models

In this section, we provide a detailed review of relevant single-node multiserver Markovian queueing models of call centers. Table 2.1 provides a list some main Markovian queueing models and their Let denote the steady-state probability (if it exits) of the system being in state (i.e. having calls in the system).

Applying the modeling techniques of the birth-death processes, we can obtain some interesting system performance measures such as .

Due to the PASTA property, we have for the model, and in the cases of the , we have and respectively.

6 b) Review of the Model and the Erlang B Formula

In this section of the paper we will review the Erlang B model paying attention to the aspects that are relevant to call center modeling. The queue models a single-node system with truck lines and no waiting spaces. Figure 2.1 depicts the queue and figure 3.2, its state transition. An important feature of the Markovian queueing models is that the arrival process follows a Poisson process. Considering the Poisson arrival process, the distribution of customers seen by an arrival to a queueing facility is, stochastically the same as the limiting distribution of customers at that facility. In other words, once the queueing system has reached steady state, each arrival from a Poisson process finds the system at equilibrium. If is the probability that the system contains customers at equilibrium and denotes the probability that an arriving customer finds customers already present, then PASTA states that . This implies that the Poisson process sees the same distribution as a random observer, i.e., at equilibrium, Poisson arrivals take a random look at the system. This result is a direct consequence of the memoryless property of the interarrival time distribution of customers to a queueing system fed by a Poisson process. In particular, it does not depend on the service time distribution. To prove the PASTA property, we proceed as follows. This results from the fact that, since interarrival times possess the memoryless property, is independent of the past history of the arrival process and hence independent of the current state of the queueing system. With the Poisson arrival process having a constant rate , the probability of having an arrival in is equal to that the PASTA property only holds for Poisson arrival processes. The formula for is called "Erlang Loss Formula" and is the fraction of time that all servers are busy. It denotes the probability that an arrival call finds all the truck line busy, (i.e. the blocking probability,).

7 Blocked Calls

It is written as s and is called "Erlang B formula":

Notice that the probability that an arrival is lost is equal to the probability that all channels are busy. Erlang loss formula is also valid for the queue. In other words, the steady-state probabilities are a function only of the mean service time, and not of the complete underlying cumulative distribution function. / ( ) = [1 (,)] = [1 (,)] = [1 (,)] < 1 1 1 < (,)**(0,) = 1, (,) = (1,) + (1,) =**(1,) + (1,) < (1,) [1 (1,)] + (1,) = (1,) [1 (1,)] < 1 < / / / / / / / (2.1)**(,) (,) (,) =**

Owning to the fact that the system is of infinite capacity, the carried load is equal to the offered load, i.e., so that the utilization and as such, we require the stability condition Given that the system is stable, the solution to the balance equations obtained from figure 2 Year013 2 C = , 0 = , 1 1 , ? ? 0 (+ 1) 2 1 1 2 1 + 1 + 1 / / > = = = = < 1 = ! 0 , 0 ! 0 , 0 = ! + ! 1 =0 1 1 1 = 0 ( ) = = = = = (,) = !(1) ! 1 =0 + !(1) = ! (1) 0 0 = (,) ! (1) (,) (,) (,) = (,) (,) + 1 (,) = + (1)/ (1,) = ! 0 = (,)(1) = ! 0 = ! ! , 0 , , = ( ) (

) 0 commonly used in performance modeling and analysis of call centers. In the application of model in call center analysis, it is usually assumed that the arrival and service rate are piece-wise constant and timeindependent. Using the parameters of each interval, the is applied to each time interval. The model is not a realistic tool for modeling call centers due to the following reasons:(2.

It assumes there is no blocking since it has infinite buffer capacity.

It does not consider the impatience (balking and reneging) attributes of customers.

8 d) Review of the Model

When the waiting room in a queueing system has a capacity limit we get a finite queue. In most situations, a finite queue occurs more naturally than a queue with a waiting room of infinite size. However, as the capacity limit gets larger, the behavior of the system approximates that of an infinite-capacity system, and in such cases we are justified in ignoring the size limit. A call center with a finite buffer and several agents is a good example of a finite queueing system. In this section we will review the model and prove new monotonicity properties of performance measures with respect to . Using the concept of total probability, we have that

9 Blocked Calls

> = > | ( ) = = (+ 1 >) = ( ) = = + 1 (+ 1,) ( ) = (>) = (+ 1 >) = ( ) ! =0 > = ( ) ! =0 = !**(1) 0 = ( ) = > 0 = (,) = 0 = 1 (,) = = (,) 1 /**

/ / / The queue is similar to the queue except that the number of buffers is finite. After buffers are full, all arrivals are lost. We assume that is greater than or equal to ;

otherwise, some servers will never be able to operate due to a lack of buffers and the system will effectively operate as a queue.

The state transition diagram for a queue is shown in Figure 2.6. The system can be modeled as a birth-death process using the following respective arrival and service rates:

Solving the balance equations derived from the state diagram, we obtain the following state probabilities. with and i. The Waiting Time Distribution

In this section, we shall provide a mathematical derivation of the waiting time distribution of the model. Due to the finiteness of the capacity of the system, deriving the waiting time distribution of the model is complicated because it results to finite series and also the arrival process is truncated by the system size . The arrival process no longer follows the Poisson process and has necessitated the need to derive the arrival point probabilities, since . In this derivation of , we shall apply the well-known Bayes' theorem.

Taking limits of both sides and using the fact that the probability of an arrival in is we have that C ? ? 0 2 1 1 2 2 1 + 1 / / / / / = / / / / / / = ,**0 1 = , 0 , = ! 0 , 0 ! 0 , 0 = ! 1 =0 + ! 1 +1 1 1 , 1 ! 1 =0**

+ ! (+ For , implies that so that we have= ; ( ) = ; ( ) = 1 ( ) ( ) / / / / / / / / / = (| (, +]) = (( ) = | (, +]) = (( ) = ; (, +]) ((, +]) = ((, +] | ( ) =) (( ) =) ((, +]| ( ) =) (( ) =) =0 = ((, +] | ( ) =) ((, +]| ( ) =) =0 (, +] + ( )**1) 1 , = 1 ( ) = 1**= ! 0 0 = ! = ! ! , 0 , 1 = ! ! =0 + = +1 = ! ! =0 + = +1 1 = 1 (,) +**(1) 1 1**= (1) (,) 1 + (,)**(1) 1, 1, 1, (**)= = = (1) (,) 1 + (,)**(1) ( ) = 1**= = 1 = = 1 1 = 1 1 (1) (,) 1 + (,)**(1)**( ) = (1) (,) 1 + (,)**(1)**( ) = 1 ( ) ( ) = (1 (,))(1) 1 + (,)**(1) ( ), ( ) ( ) = 1.**= 1 = = ! ! =0 + = +1 1 = 1 (,) + 1 = (,) 1 + ( ) (,) ( ) = = = = (,) 1 + ( )**(,) ( ) = 1**= = 1 = = 1 = = ( ) = ( ) (,) 1 + ( ) (,) ( ) = 1 ( ) ( ) = 1 (,) 1 + ( ) (,) = / / / / / / ( ) = (,), ( ) = 0 and ( ) = 1 (,). = ( ) = = = = (,) 1 + ( ) (,) = (,)**(1) 1**+ (0) (,) = (,) ( ) = (1) (,) 1 + (,)(1) = (0) (,) 1 + (,)(0) = 0 ( ) = 1 ( ) ( ) = 1 0 (,) = 1 (,) 1. lim ( ) = 1 , > 1 1, = 1 (,), 0 < < 1 2. lim ( ) = 1 1 , 1 0, 0 < 1 3. lim ( ) = 0, 1**1**

(,), 0 < 1 Year C 0 < < 1 ( ) = (1) (,) 1 + (,)**(1)**(,) 1 + (,) = (,) ( ) = ( ) (,) 1 + ( ) (,) = (,) 1 ( ) + (,) 1 = 1 ( ) = (1) (,) + (,)**(1) 1 =**

(1) (,)1 1 1 + (,)**(1) 1**= 1 lim ( ) = 1 , > 1 0 < < 1 ( ) = = =**(1) (,) 1 + (,)(1) 0 0**, 0 < < 1. = 1 ( ) = = (,) 1 + ( ) (,) 0 > 1, > 1, ( ) = = (1) (,) + (,)**(1) 1 =**(1) (,) 1 + (,)**(1)**

=(,)**(1) (,) 1 +**

(1) (,),) = 1 1 > 1 ( ) = 1 (,) (1) 1 + (,)(1) = 1 (,)(+ (,)**(1)**(1 (,))**(1) 1 =**

Before we proceed to derive the formula for computing an important performance measure , we shall prove some new results that will be useful in the course of our derivations and computations. C 0 < < 1 ( ) = 1 (,) (1) 1 + (,)**(1) 1**(,) (1) 1 + (,)**(1)**= 1 (,) = 1 ( ) = 1 (,) 1 + ( ) (,)**0 1**(|) = > 0 = 1 = = 1 1 = = 1 1 1 = = 1 1 ( ) = ( ) 1 ( ) = 1 + (,)**(1) 1**+ (,)(1 1) (,)**(1) 1**+ (,)**(1)**(|) = > 0 = (,)(1) 1 + (,)(1 1) (|) = = 0 = 1 =0 = 1 (|) = (1)[1 (,)] 1 + (,)(1 1) For , = 1 (|) = ( ) 1 ( ) = ( ) (,)**1+**( ) (,) 1 (,)**1+**( ) (,) = ( ) (,) 1 + (,)[1]

Now, using the principles of conditional probability, we can write For ,

Then for , we have that Where we have used the fact that [15] Global Journal of Computer Science and Technology

Volume XIII Issue X Version I(D D D D D D D D D) Year C > | > 0 = > ; > 0 > 0 = > > 0 > = > | > 0 > 0 > = > | > 0 (|) 1 > | > 0 = > | 1 = > | = + ; 1 (= + | 1) +1 =0 = ( ) ! =0 1 =0 1 + + + 1 = ( ) ! 1 + + + 1 1 = 1 =0 > | > 0 = ( ) ! 1 1 =0 = ( ) ! 1 1 1 =0 1 > = (|) ( ) ! 1 1 1 =0 , 0 (= + | 1) = 1 + + + 1 , 0**1**

For , we also have that= 1 > = (|) ( ) ! 1 1 =0 (2.13) (2.14) (2.15)

In same line of reasoning, we derive the mathematical formula for computing the Average Speed to Answer (ASA) as follows:

By the application of Little's law, we have that III.

10 Limiting Behaviour of the Model Performance Indicators

In this section of the paper we shall prove some limiting properties of the model with respect to . In same way, since IV.1 1 + (1)( ) (1)**(1) = (1 (1 + (1)( )) (,) (1 + (,)(1 1))(1)**

11 Conclusions

In this paper, we have discussed in detail the modeling of a call center as single-node using the Markovian queueing techniques. We considered the Erlang B Loss model and the Erlang C model as well as the more general model. Our emphasis is on the derivation of the exact performance measures of these well-known models. Considering the model, we expressed the system performance measures in terms of Erlang B formula, which facilitates the computation as well as the analysis. Using the results emanating from the analysis, we showed the monotonicity properties for performance measures with respect to and . [1] [2]

[1] On the M/M/c/N Call Center Queue Modeling and Analysis © 2013 Global Journals Inc. (US)

[2] © 2013 Global Journals Inc. (US)

111

Figure 1: Figure 1 Figure 1 . 1 :

12$^{\lambda}$

Figure 2: CFigure 1 . 2 :

21$^{\mu}$

Figure 3: Figure 2 . 1 :

μ

Figure 4: Let = Number of

2$^{\mu}$

Figure 5: Figure 2

2$^{\mu}$

Figure 6: Figure 2

c

Figure 7:

2$^{\lambda}$

Figure 8: Figure 2

2

1 : Some Multiserver Markovian Queueing Models

Figure 9: Table 2 .

[Kleinrock ()] , L Kleinrock . *Queueing Systems* 1975. Wiley. 1.

[Jacques and Queueing Theory ()] , Adan I Jacques , R Queueing Theory . 2002. Eindhoven, Netherlands.

[Mandelbaum (2006)] *Call centers (centres) research bibliography with abstracts*, A Mandelbaum . May 4, 2006. Technion, Haifa, Israel. 7. (Tech. rep.)

[Zhang ()] *Call centres with balking and abandonment: from queueing to queueing network models*, Zhidong Zhang . 2010. Saskatoon, Saskatchewan. University of Saskatchewan (Ph.D. Thesis)

[Borst et al. ()] 'Dimensioning large call centers'. S Borst , A Mandelbaum , M Reiman . *Operations Research* 2004. 52 p. .

[Chen and Yao ()] *Fundamentals of queueing networks: performance, asymptotics and optimization*, H Chen , D Yao . 2001. New York: Springer-Verlag.

[Gross et al. ()] Donald Gross , John F Shortle , James M Thompson , Carl M Harris . *Fundamentals of Queueing Theory*, (New York, NY) 1998. Wiley.

[Ross ()] *Introduction to Probability Models*, S M Ross . 2011. Academic Press. (Tenth Edition)

[Gnedenko and Kovalenko ()] *Introduction to Queueing Theory*, B Gnedenko , I Kovalenko . 1968. Cambridge, MA: Birkhauser Boston Inc.

[Cooper ()] *Introduction to Queueing Theory*, R Cooper . 1981. North-Holland, New York.

[Srinivasan and Talim ()] 'Performance analysis of a call center with interactive voice response units'. R Srinivasan , J Talim , Wang , J . *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research* 2004. 12 p. .

[Boucherie ()] *Product-Form in Queueing Networks*, R Boucherie . 1992. Amsterdam, The Netherlands. Free University (Ph.D. Thesis)

[Koole and Mandelbaum ()] 'Queueing models of call centers: An introduction'. G Koole , A Mandelbaum . *Annals of Operations Research* 2002. 113 p. .

[Bolch et al. ()] *Queueing networks and Markov chains: modeling and performance evaluation with computer science applications*, G Bolch , S Greiner , H De Meer , K S Trivedi . 1998. New York, NY, USA: Wiley Interscience.

[Ross ()] *Stochastic Processes. Second Edition*, S M Ross . 1996. New York: Wiley Inc.

[Gans et al. ()] 'Telephone call centers: Tutorial, review, and research prospects'. N Gans , G Koole , A Mandelbaum . *Manufacturing and Service Operations Management* 2003. 5 p. .

[Mandelbaum and Zeltyn ()] 'The Palm/Erlang-A queue, with applications to call centers'. A Mandelbaum , S Zeltyn . Service Engineering Lecture Notes 2004. (Tech. rep)