Mathematical model of the contact center

ERIK CHROMY, IVAN BARONAK
Slovak University of Technology in Bratislava
Faculty of Electrical Engineering and Information Technology
Ilkovicova 3, 812 19 Bratislava
SLOVAKIA
chromy@ut.fei.stuba.sk, baronak@ut.fei.stuba.sk

Abstract: The paper deals with the contact center modeling with emphasis on the optimal number of agents. The contact center belongs to the queueing systems and its mathematical model can be described by various quality of service parameters. The Erlang C formula is a suitable tool for the modeling of QoS parameters of contact centers. The contact center consist of IVR system and service groups and we propose also two new parameters – downtime and administrative task duration. These parameters are useful for better determination of the optimal number of contact center agents. Based on these parameters we propose a mathematical model for contact centers also with repeated calls.

Key–Words: Contact Center, Erlang C Formula, Interactive Voice Response, Optimization, Quality of Service

1 Introduction

A contact center is in common a typical example of a queueing system [1]–[5]. In a contact center a given number of calls occurs within a defined time interval. These calls originate randomly and are independent from each other. Queueing systems often have to deal with the predetermination of Quality of Service [6]–[12]. For the description and dimensioning of queueing systems various mathematical models are used e.g. Markov models (with various number of servers, with or without a queue), Erlang formulas, Jackson networks and non-Markovian models [13]–[15].

A contact center system consists of various components (Figure 1) where IVR (Interactive Voice Response) system and ACD (Automatic Call Distribution) [16]–[17] can be considered as a basis of contact centers.

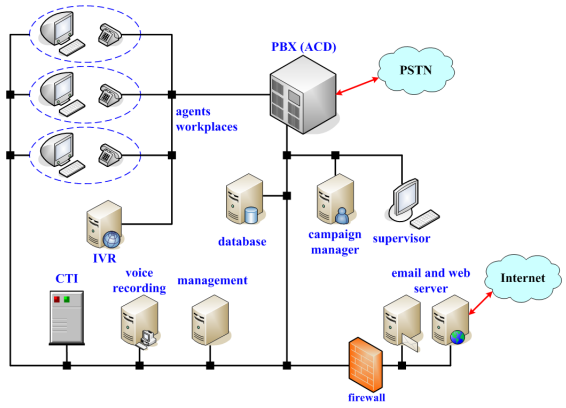

Figure 1: Contact Center technology.

1.1 Interactive Voice Response

The IVR offers exploitation of telephone terminal for realization of communication between the caller and the system that stores records. It is straight contact with calling customer and it serves also as an instrument for callers identification by means of which consecutive spooler system services are possible - administration of information about caller by agent, toward whom is the customer directed, via pop-up screens.

At the same time it allows the caller to serve himself without agent‘s help. Agent is often unnecessary for control of caller‘s requirements. A lots of calls produce the same inquiries (e.g. opening hours of branch offices, range of goods and number of products in stock, solvency, ticket reservations, etc.) and it is possible to automate answer to these questions. As the result it offers customers satisfaction and also it has financial effect because we do not need a lot of agents to handle inquiries from customers.

Success of IVR applications is based on the caller‘s experience how easy is their use. IVR application gives the caller always the possibility of choosing and handling of his course.

Extended package of the modern IVR system functionalities could involve:

- Fax on demand,
- E-mail on demand,

- Text to speech,
- Automatic speech recognition,
- Application mailbox,
- Database Read&Write,
- CTI integration,
- Interactive queuing.

The IVR system automates call processing and many of time-consuming tasks (performed by agents and supervisors on daily basis). Such system allows:

- automatic announcements of a welcome message and estimated waiting time for the caller,
- automatic performs various actions based on the estimated waiting time, e.g. identification of the caller through PIN (Personal Identification Number),
- a caller can leave a message for the agent with a callback inquiry,
- reading of selected text from the database,
- selection of the communication language, etc.

1.2 Automatic Call Distribution

ACD belongs to the basic software needed for contact center realization. It offers various and sophisticated call routing functions, queue creation and productivity control of agents handling these calls. Agents with the same or similar profile form service group. Call routing interconnects a calling customer with a given agent. The following important parameters should be taken into consideration by call routing algorithms:

- caller requirements for service,
- agent profile,
- state of the agent (free or busy),
- state of the queue (number of callers, average waiting time for service),
- system load in real-time,
- alternative resource (in the case of overload).

Although contact centers are effectively utilizing resources we have to assume a situation where no free agents are available to the system. Hence we have to deal with queueing in the case of ACD. Private Branch Exchange (PBX) inserts all incoming calls into a queue. The required service is selected at first (selection is based on the IVR) and based on the selection inside the IVR system the call is routed to the queue for a given service. The call is enqueued according to FIFO algorithm. Special case will occur when the caller is marked as special. In this case the call is put on the beginning of the queue, instead of the end of the queue.

The calls remain in the queue:

- until any agent of the calling group will be free, or
- until expiration of the time interval dedicated for waiting of the caller (parameter is set by the contact center manager), or
- until the call is terminated by the caller.

2 Queueing system and performance measurement

The contact center is typical example of the queueing system with the following characteristics:

- at the contact center input the requests from customers are randomly occurring in random time intervals,
- the handling time of each request is random a variable (hence, it is necessary to determine the average time for handling request by agent),
- contact center can have more service groups, each with custom random parameters (number of requests on input and average handling time),
- the output of the system is the request handled by an agent.

The common model of queueing system consists of three basic parts: process of arrivals, storing and server (handling). The most used model for the arrival process is Poisson arrival process [18]. Handling time is a random variable and it is independent from arrivals. The system inserts customers into the memory until servers become free. There are two extreme cases. The storage can be such large, that it can be regarded as infinite. In the second case the storage is only for customers in the process of handling. Queueing systems are used in many fields (services, industry, storage and maintenance, etc.), but their application is very important specially in the field of telecommunications.

In order to judge individual queueing systems, it is necessary to implement the performance measurement (which will describe such system). Because

the queueing system is a dynamic system, the performance measurement may vary over time. It holds, that the system is in steady state, when all transients in the system are finished, the system is stabilized and the performance measurement values are time independent. The system is in so called statistical equilibrium, i.e. the arrival rate of requests into the system is equal to the rate at which the requests leave the system. Such system is called stable system.

The most important parameters of performance measurement are:

- *probability of number of requests in the system* - the system can be described through the probability vector of number of requests in the system. The average value can be determined from such vector,
- *utilization* ρ - if the queueing system consists of one server, then the utilization is the fraction of time in which the server is active. If there are no limitations on the number of requests in one queue, then the server utilization is as follows:

$$\rho = \frac{arrivalrate}{servicerate} = \frac{\lambda}{\mu} \tag{1}$$

  In the case when the queueing system consists of multiple servers m, then for ρ the following equation is valid:

$$\rho = \frac{\lambda}{m\mu} \tag{2}$$

- *throughput* - throughput of simple queueing system is defined as average number of the handled requests (i.e. leaving rate from the system),
- *response time* - total time the request spent in the system,
- *waiting time in the queue* - represents the time that the request spent in queue before handling (*then response time = waiting time in the queue + service time*),
- *queue length* - number of requests in the queue,
- *number of requests in the system* - total number of requests in the queueing system (i.e. requests waiting and being processed).

3 Contact center model

Our proposed model (Figure 2) of the contact center consist of IVR system and of n service groups. The IVR system is in contact with the caller customer first (calculated QoS parameters have index 1). Service group represents agents that serves a certain group of the customers (calculated QoS parameters have index 2).

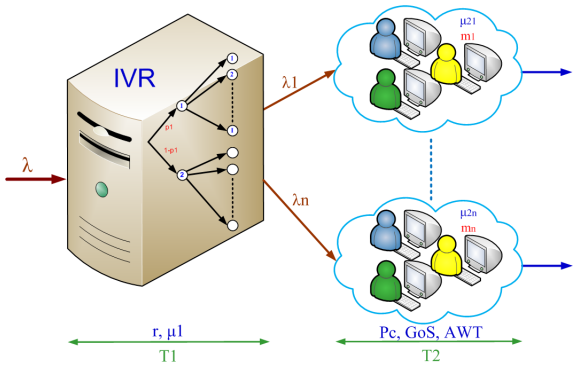

Figure 2: Contact Center model.

The parameters of the IVR system are:

- r - number of calls from range <0,1> handled automatically without need of the contact center agent interaction,
- c - number of calls incoming to the contact center during the busy hour,
- $1/\mu_1$ - average handling time for one request in IVR system.

For the service groups the following parameters are used:

- n - number of service groups,
- $1/\mu_{2,i}$ - average handling time for one calling request in service group i,
- P_C - upper bound from range <0,1> for request enqueue,
- $\boldsymbol{\lambda}$ - number of requests incoming to the given service group during busy hour from IVR system,
- *AWT* - acceptable waiting time (converted to hours),
- *GoS* - request rate from range <0,1> , which the agent must accept within defined AWT period,,
- $T_{p,i}$ - downtime (represents time dedicated for example for the hygienic break) of service group i in minutes,

- $a_{c,i}$ - rate of calls from range <0,1>, which require administrative tasks (activity which the agent has to perform after call) in service group *i*,
- $T_{a_{c,i}}$ - average time of the administrative task duration (in minutes) which the agent has to perform for one call in service group *i*.

In our model, we propose two new parameters: downtime and administrative task duration. After the contact center model definition, two cases of requests distribution can be considered and modeled:

a) Output requests from IVR systems are uniformly distributed among service groups. Then for number of requests incoming to each service group we have following equation:

$$\boldsymbol{\lambda} = \frac{\lambda(1-r)}{n} \tag{3}$$

b) Output requests are non-uniformly distributed among service groups. In this case we have to define probability ($p_{0,1}$ till $p_{0,n}$) by which the service group is selected. These probabilities are set by contact center manager before start of operation based on the detailed analysis of expected interest on services from callers (forecast). Later these probabilities should be set more precisely by measuring of real traffic in contact center.

As we have *n* service groups, then for particular probabilities following condition must be valid:

$$\sum_{i=1}^{n} p_{0,i} = 1 \tag{4}$$

Then for number of the requests incoming into *i-th* service group following equation must be valid:

$$\boldsymbol{\lambda_i} = \lambda.(1-r)p_{0,i} \tag{5}$$

For total average time the caller will spent in contact center system (e.g. IVR + *i-th* service group) we have the following equation:

$$T_i = T_1 + T_{2,i} = \frac{1}{\mu_1} + \frac{1}{\mu_{2,i}} + W_{2,i} = \frac{1}{\mu_1} + \frac{1}{\mu_{(T_p+a_c),i}} + W_{2,i} \tag{6}$$

In our mathematical model, we propose to use the parameter $\mu_{(T_p+a_c),i}$ instead of parameter $\mu_{2,i}$. Because, this parameter takes downtime and administrative task into acount.

The value of $W_{2,i}$ parameter depends on used mathematical model: Erlang C equation or Markov model M/M/m/K.

3.1 Modeling of IVR system

The goal of the input part of the contact center that is represented by IVR system is:

- automatic caller service without agent interaction,
- routing the caller to a suitable service group of agents (based on the choice of caller in the IVR tree).

The input parameters of the IVR system are:

- the number of calls that enter into the system during the busy hour - λ,
- the percentage of calls that are automatically served (i.e. there is no need to communicate with agent) - *r*,
- the probability of transitions between the the states in the IVR tree - $p_{i,j}$.

The output of the IVR system is the number of callers that requires communication with the agent of given service group. The input to the service group (the second part of the SHO) is thus given:

$$\boldsymbol{\lambda} = \lambda.P.(1-r) \tag{7}$$

where the parameter *P* represents the multiplication of probabilities between transitions in the IVR tree (e.g. P=$p_1.p_{1,1}.p_{1,1,1}$).

Thus, the IVR system with so defined parameters can be described by graph theory a markov model M/M/infinity (infinite queue system).

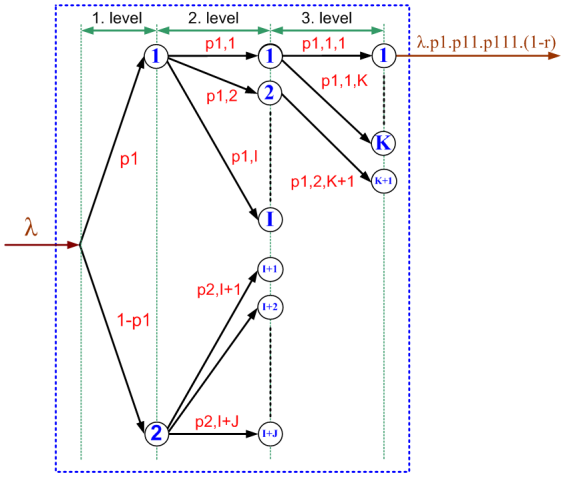

Figure 3: IVR model.

3.2 Important QoS parameters

Based on analyses in papers [19]–[20], the Erlang C formula is a very suitable tool for modeling of important QoS parameters of contact centers. It also allows to calculate the optimal number of agents for given *i-th* service group. Consequently, after defining of all parameters for the calculation of required number of agents we have:

a) equation (8):

$$P_C(m_i, \frac{\boldsymbol{\lambda}}{\mu_{(T_p+a_c),i}}) = \frac{\frac{m_i(\boldsymbol{\lambda}.X)^{m_i}}{m!(m_i-\boldsymbol{\lambda}.X)}}{\sum_{j=0}^{m_j-1}\frac{(\boldsymbol{\lambda}.X)^j}{j!} + \frac{m_i(\boldsymbol{\lambda}.X)^{m_i}}{m!(m_i-\boldsymbol{\lambda}.X)}} \tag{8}$$

where the parameter *X* is represented by equation (9):

$$X = \frac{60 + \mu_{2,i}.a_{c,i}.T_{a_{c,i}}}{\mu_{2,i}(60 - T_{p,i})} \tag{9}$$

b) or equation (9) if we have *GoS* parameter defined:

$$m_i = \frac{AWT.\boldsymbol{\lambda} - ln(\frac{1-GoS}{P_C})}{AWT.\frac{\mu_{2,i}}{X}} \tag{10}$$

where the value of the parameter $\boldsymbol{\lambda}$ in equations (8) and (10) depends on the assumed distribution of output requests from IVR system - uniform distribution (3), or non-uniform distribution (5). In the case of equation (8) the required number of agents can be obtained by iteration.

For calculating of required number of agents according to equation (8) the parameters λ, $1/\mu_1$, n, $1/\mu_{2,i}$, P_C must be known. For equation (8) we also need to know the parameters *GoS* and *AWT*. For other parameters, if they are not known, they get value 0.

On the base of previous considerations, we propose to use the following equation, that is valid for the average number of served callers during the busy hour (taking into acount downtime and administrative task):

$$\mu_{(T_p+a_c),i} = \frac{\mu_{2,i}(60 - T_{p,i})}{60 + \mu_{2,i}.a_{c,i}.T_{a_{c,i}}} \tag{11}$$

After estimation of the number of agents of given service group m_i, another QoS parameters should be verified also:

- load of agent in given *i-th* service group during busy hour is very interesting indicator:

$$\rho = \frac{\boldsymbol{\lambda}_i}{m_i.\mu_{(T_p+a_c),i}} \tag{12}$$

- also the average time the customer will spent in queue of given *i-th* service group can be verified:

$$W_{2,i} = \frac{P_C}{\mu_{(T_p+a_c),i}.(m_i - \frac{\boldsymbol{\lambda}_i}{\mu_{(T_p+a_c),i}})} \tag{13}$$

- further, the average number of callers in queue of *i-th* service group is interesting for contact center manager:

$$Q_{2,i} = \frac{\boldsymbol{\lambda}_i}{\mu_{(T_p+a_c),i}.(m_i - \frac{\boldsymbol{\lambda}_i}{\mu_{(T_p+a_c),i}})}.P_C \tag{14}$$

- for the average time the caller will spent in contact center (IVR + *i-th* service group) following equation is valid:

$$T_{2,i} = \frac{1}{\mu_{(T_p+a_c),i}} + \frac{P_C}{\mu_{(T_p+a_c),i}.(m_i - \frac{\boldsymbol{\lambda}_i}{\mu_{(T_p+a_c),i}})} \tag{15}$$

From the view of contact center manager - if any of the QoS parameters render unacceptable values, the number of agents should be increased.

3.3 Repeated calls in contact center

The situation with repeated calls [14] (in the Fig. 4) can be described with next parameters:

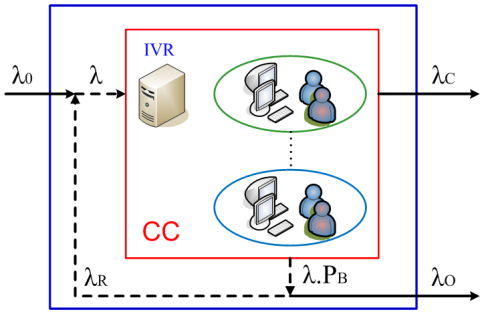

Figure 4: Repeated calls in the contact center.

- λ_0 - the number of calls that enter into the contact center during the busy hour,
- λ_R - the number of repeated calls,

- λ - the whole number of calls that enter into the contact center,
- λ_C - the number of served calls,
- λ_O - the number of non-served calls.

where the parameter λ is represented by equation:

$$\lambda = \lambda_0.\lambda_R \tag{16}$$

and the parameter λ_0 is represented by equation:

$$\lambda_0 = \lambda_C.\lambda_O \tag{17}$$

4 Conclusion

The output of the paper is a mathematical model of contact center. We proposed two parameters - downtime and administrative task duration. These parameters are useful for better determination of the optimal number of contact center agents. Erlang C formula is a suitable tool for modeling of important quality of service parameters of contact centers. When the required number of agents in each service group is determined, it is possible to verify the accuracy of the result through other QoS parameters and then eventually to add another agent into the service group.

Our model consist of IVR system and service groups. Finally, we have shown also the possibility with repeated calls. This represents an interesting situation when a caller tries to call again. Such calls increase the number of incoming calls at the input to the contact center and thus have an impact on the number of required agents.

Acknowledgements: This article was created with the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic within the KEGA agency project - 007STU-4/2016 Progressive educational methods in the field of telecommunications multiservice networks and VEGA agency project - 1/0462/17 Modeling of qualitative parameters in IMS networks.

References:

[1] G. Bolch, "Queueing Networks and Markov Chains - Modeling and Performance Evaluation with Computer Science Applications," 2nd ed., New Jersey: John Wiley & Sons, 2006.

[2] A. Budhiraja, A. Ghosh and X. Liu, "Scheduling control for Markov-modulated single-server multiclass queueing systems in heavy traffic," *Queueing Systems*, vol. 78, no. 1, pp. 57–97, 2014.

[3] B. Buke and H. Chen, "Stabilizing policies for probabilistic matching systems," *Queueing Systems*, vol. 80, no. 1-2, pp. 35–69, 2015.

[4] V. Shah and G. de Veciana, "Asymptotic independence of servers activity in queueing systems with limited resource pooling," *Queueing Systems*, vol. 83, no. 1, pp. 13–28, 2016.

[5] M. Budhiraja, "Invariance of workload in queueing systems," *Queueing Systems*, vol. 83, no. 1, pp. 181–192, 2016.

[6] A. Kovac, M. Halas and M. Orgon, "E-model MOS estimate improvement through jitter buffer packet loss modelling," *Advances in Electrical and Electronic Engineering*, vol. 9, no. 5, pp. 233–242, 2011.

[7] J. Misurec and M. Orgon, "Modeling of power line transfer of data for computer simulation," *International Journal of Communication Networks and Information Security*, vol. 3, no. 2, pp. 104-111, 2011.

[8] H.S. Nguyen, T.-S. Nguyen and M. Voznak, "Successful transmission probability of cognitive device-to-device communications underlaying cellular networks in the presence of hardware impairments," *Eurasip Journal on Wireless Communications and Networking*, vol. 2017, no. 1, 2017, Article number 208.

[9] S. Klucik and M. Lackovic, "Modelling of H.264 MPEG2 TS traffic source, " *Advances in Electrical and Electronic Engineering*, vol. 11, no. 5, pp. 404–409, 2013.

[10] R. Roka, "The environment of fixed transmission media and their negative influences in the simulation," *International Journal of Mathematics and Computers in Simulation*, vol. 9, pp. 190–205, 2015.

[11] F. Certik and R. Roka, "Possibilities for Advanced Encoding Techniques at Signal Transmission in the Optical Transmission Medium," *Journal of Engineering (United States)*, vol. 2016.

[12] J. Frnda, M. Voznak and L. Sevcik, "Impact of packet loss and delay variation on the quality of real-time video streaming," *Telecommunication Systems*, vol. 62, no. 2, pp. 265–275, 2016.

[13] A. Brezavscek and A. Baggia, "Optimization of a Call Centre Performance Using the Stochastic Queueing Models," *Business Systems Research*, vol. 5, no. 3, pp. 6–18, 2014.

[14] S. Ding, G. Koole and R. D. Mei, "On the Estimation of the True Demand in Call Centers with Redials and Reconnects," *European Journal of Operational Research*, vol. 246, no. 1, pp. 250–262, 2015.

[15] J. Zan, J. Hasenbein and D. Morton, "Asymptotically optimal staffing of service systems with joint QoS constraints," *Queueing Systems*, vol. 78, no. 4, pp. 359–386, 2014.

[16] B. Legros, O. Jouini and G. Koole, "Optimal scheduling in call centers with a callback option," *Performance Evaluation*, vol. 95, pp. 1–40, 2016.

[17] G. Koole, B. Nielsen and T. Nielsen, "Optimization of overflow policies in call centers. Probability in the Engineering and Informational Sciences," *Cambridge University Press*, vol. 29, no. 3, pp. 461–471, 2015.

[18] J. F. Hayes, "Modeling and Analysis of Telecommunications Networks," New Jersey: John Wiley & Sons, 2004.

[19] E. Chromy, T. Misuth and M. Kavacky, "Erlang C Formula and Its Use In the Call Centers," *Advances in Electrical and Electronic Engineering*, vol. 9, no. 1, pp. 7–13, 2011.

[20] J. P. C. Blanc, "Queueing Models: Analytical and Numerical Methods. Course 35M2C8," Departement of Econometrics and Operations Research, 226 p., 2011.