SIPARUNTON
International Journal of
Interdisciplinary Research

ISSN 2337-0556 (Print)
ISSN 2337-0572 (Online)
Vol 1, Issue 5, October 2015

==========

Optimizing a call center performance using queueing models – an Albanian Case

Ditila Ekmekçiu
University of Tirana, Faculty of Natural Sciences, Albania,
ditila.ekmekciu@gmail.com

Abstract:
The world of call centers is an important reality nowadays and helping the decision making of the operations management is fundamental for this industry. A call center generally represents the first contact of a customer with a specific company. As a result, the quality of the service offered is of fundamental importance. The objective of this paper is to see how to apply the queueing models in order to optimize the call centers' performance. A crucial factor of the call centers' optimization is determining the proper number of agents, during a working day, considering the chosen performance measure. The experiment is done in one of the Albanian outsourcing call centers. The literature related to the application of such models at call centers is reviewed. A suitable number of agents was determined for different peak periods of the working day, considering the most important performance measures. The obtained results prove how useful and applicable are the stochastic queueing models as a tool for a call centers' performance, in terms of the expected waiting time, number of customers waiting for service and of call centers service levels optimization. Practically, all the data needed for this mathematical/analytical approach is provided. This paper has the purpose to illustrate how such data can be efficiently used to advantage the operations management.
Keywords: Call center, Service Levels, Optimization, Queueing Models

JEL Classification: C02, C44, C61

1. Introduction

==========

A call center is composed by a set of resources (generally staff, computers and telecommunication supplies), which make possible the delivery of services through the telephone. The working environment of a typical large call center (Figure 1) could be imagined as very big room with numerous open-space workstations, in which people with earphones sit in front of computer terminals, providing services/products to "unseen" customers. The majority of call centers also hold up Interactive Voice Response (IVR) units, which are automatic answering machines that include the possibilities of interactions. A current tendency is the extension of the call center into a contact center, which is a call center where the traditional telephone service is reinforced by some additional customer-contact channels, commonly IVR, e-mail, fax, Internet and/or chat. Most important companies have reorganized their communication with customers through one or more call centers, which can be managed internally or outsourced. The tendency towards contact centers has been increased by the societal promotion surrounding the Internet, by customer request for channel variety, and by recognized potential for efficiency profits (requests for e.mail and fax services can be "stored" for later response and, when traffic of the telephone services is lower than forecasted, agents can switch in the other channels). This paper deals exclusively with pure telephone services.

Figure 1: A typical call center

Call centers can be categorized into many different dimensions: functionality (help desk, emergency, telemarketing, teleselling, technical support, market research, information providers, etc.), size (from a few to several thousands of agent workstations), geography (single- or several locations, that can be domestic, nearshore or offshore), agents qualifications (low-skilled or highly-trained, single or multiskilled etc.), industry (telecommunication, finance, TV and internet providers, travel, marketing, sport etc.), line of business (inbound, outbound, backoffice etc.), and so on. The focus here is on inbound call centers (which means that they handle incoming and not outgoing calls). Pure outbound call centers are usually used for advertisement, teleselling or market researches. Modern call/contact centers however are faced with large types of calls, coming in from different communication channels (telephone, internet, fax, e.mail., chat, etc.) and agents have the skill to handle one or more types of calls (for example, they can give technical support for different products in different languages by different channels like phone, e-mail, chat etc.). The organizational architecture of the modern call center varies from the very flat, where all agents are exposed to the incoming calls, to the multi-layered, where a layer represents a level of expertise and customers can potentially be transferred through different layers until being served to satisfaction.

2. Management and quality of service

There exists a large range of literature regarding the management of call centers in the trade literature. As already explained, call centers have evolved into the prime contact bound of businesses with their customers and for this reason, represent highly promising business opportunities. Call center managers are so faced with the conflicting goals and objectives of providing consistent and even increasing high service quality, through an unstable fast expanding service channel, to a frequently huge customer base - a really fertile ground for Operations Research models and analysis, with a significant role reserved for queueing theory and queueing science. Generally, call center goals are formulated as the provision of service at a known given quality, depending on a specific. While Service Quality is a very complexed issue, to which a large number of articles and books have been dedicated [1, 2, 3], a highly simplified approach is sufficient for our

purposes. We measure service quality through two dimensions: qualitative (psychological) and quantitative (operational). The first relates to the way in which service is provided and perceived (customer satisfaction with the agents answer and/or kindness, etc.; like, [4]). The second is related more to service accessibility (how long did the customer have to wait for an answer, was he forced to call back, how much did the call last until the issue was solved etc.). Models supporting the qualitative issues of service quality are generally empirical, originating in the Social Sciences or Marketing, like different surveys made by the end of the call asking some given few questions to measure the customers satisfaction (Sections III, IV and VIII in [5]). Models supporting quantitative management are generally analytical, and here we focus on the subdivision of such models that derives from Operations Research in general and Queueing Theory in particular. The most used practice is that high management decides on the desired service level and then call center managers are called on to defend their specific budget. In a similar way, costs can be associated with service levels (eg. tax-free services pay out-of-pocket for their customers' waiting), and the objective is to minimize total costs. These two approaches are integrated in Borst, Mandelbaum and Reiman [6]. It happens, although, that profit can be associated directly to each single call, for example in sales companies. In that point a direct deal can be made between service level and costs in order to maximize overall profit. Two articles in which we find this approach are Andrews & Parsons [7] and Ak¸sin & Harker [8]. In what follows we concentrate on the service level vs. cost (efficiency) deal. The fact that salaries account for 60–70% of the total direct costs of a call center justifies the focusing mostly at staff costs. This is also the approach embraced by workforce management tools that are used on a large range in call centers. By concentrating on staff, one presumes that other resources are not barriers.

3. Performance measures

Operational service level is generally quantified in terms of some barriers or performance measures, like abandonment, waiting and/or retrials, which underscore the natural fit between queueing models and call centers. Abandonment is measured by the fraction of customers that abandon the

queue before being served. Waiting is measured by the Average Speed of Answering. The typical standard for telephone services usually follows the 80/20 rule, under which at least 80% of the customers must wait no more than 20 seconds. And as last, retrials can be calculated by the fraction of customers whose request is satisfied on first attempt, or by the average number of "visits" needed to resolve a single issue. Performance measures are definitely intercorrelated - see [9] for the remarkably linear relation between the fraction of abandoning customers and average waiting time. They could also transmit more information that actually we can imagine. In contrast to waiting statistics which are something objective, abandonment and retrial measures are in some cases subjective because they incorporate customers' view on whether the offered service is worth its wait (abandonment) and/or returning to (retrials). As another example, it turns out that customers' patience can be quantified in terms of the ratio between the fraction abandoned to the fraction served - indeed, it is explained in [10] that this ratio can be also interpreted as the difference between the average time that customers are willing to wait to the average time that they expect to wait. The practice of service levels must be handled with very much care. Low levels of waiting could correlate highly with short service times which, on the other hand, could result in many retrials. That is to say, service times and delays are definitely long, being accumulated over several visits. It is also not always clear whether service level must hold for each time interval, or it must hold on average over the whole day, or perhaps for an arbitrary customer. The usual situation, and specifically in our case, is that the service level must hold for each time interval. (Koole and van der Sluis [11] emphasize the advantages of looking at an overall service level. Performance measures can be valuable only if archived at a proper resolution and observed at the appropriate frequency. In an ideal situation, one would like to save, for each single transaction at the call center, its operational and business features. This raw data can then be extracted for exploratory purposes, or aggregated into performance measures for management use.

4. Literature Review

The queueing theory is a special field of stochastic process theory. Many books discussing the fundamentals of the queueing theory were published in recent years, e.g. van Dijk and Bouchiere [12]; Gross [13]; Tanner [14];

Tijms [15]. Queues often represent an area where customers wait for service providers to be served, and can be used as inventories in manufacturing or any human related services. Applications of the queueing theory can be found in different areas such as: (1) telecommunication, can be found in Attahirusule [16], Giambene [17]; (2) computer networks traffic studies, can be found in Robertazzi [18]; and (3) road traffic studies (Ismail [19]), and others. Numerous contributions in the literature prove that the queueing theory can be applied successfully also to call centers performance optimization, e.g. Brown [20]; Koole [21]. Koole and Mandelbaum [22] consider that the world of call centers is a challenging fertile area for the applications of queueing models. Call centers (or their modern successors contact centers) are frequently used by organizations for marketing purposes and technical support for end users. They are preferred and prevalent way for many companies to communicate with their customers (Koole and Mandelbaum, [22]). Inbound calls arrive at random according to some complexed stochastic processes, call durations are as well random, waiting calls may abandon after a random patience time, some agents may not show up to work for any reason, and so on (Avramidis [23]). In some cases, emergency calls in call center should be considered, too (especially in police, fireman, hospital call center), where priorities play a determinant role. Additionally, in the research Aksin [24] agent's ability to handle stress is taken into consideration. An essential task of operations managers is to define the appropriate number of agents in the call center to ensure optimal performance. As stated by Aksin [24] an appropriate number of agents for each period of the scheduling perspective in a call center depends on both, how many customers and therefore how much work is arriving into the call center at what times, and how quickly the agents look for to serve these customers. The main performance measures in call centers have to do with the quality of service (service levels) and/or the operating cost. The generally used approach in a call center optimization is direct cost minimization objective with the call center service level constraint (Aksin [24]). Generally, the quality of call center service is related to customer satisfaction with the service, and usually depends on the waiting times before they are answered. For a call center performance optimization many theoretical mathematical models are available in the literature, like Chassioti [25]; Dombacher [26]. Part of them are rather simple and present

closed form expressions for most of its performance measures. Application of this kind of model in practice is therefore quite easy. However, in order to obtain relevant and useful results, it should be assured that fundamental assumptions of the model do not contradict the properties of the call center and its operation.

5. Metodology

The basis for proper selection of a theoretical/mathematical model to describe a specific call center in practice represents the knowledge of the probability density functions of inter-arrival times (for example times between two successive incoming calls) and service times (generally calls length). These functions can be procured if accurate and complete data about the call center operation are available. Since most of modern call centers use contemporary technology, which enables automatic logging of all the events in the call center, the data needed for the mathematical analysis are usually provided. However, the lack of expert knowledge in practice prevents the companies from efficient use of them. Usually the number of operators in a shift is often based on the rule that in wise to follow decision, and is frequently not an optimal solution. This paper represents the practical application of the queueing theory for optimization of a call centers performance. The field data of the call centers operation will be used to analyse the arrival and service arrangements. On the basis of this analysis an appropriate theoretical queueing model will be selected to describe the call center taken into consideration. A suitable number of operators will be determined for different peak periods of the working day regarding different call center performance measures.

A typical queueing system consists of one or more service units (like the servers), arrivals of customers demanding the service, and the service process. When all the customers cannot be served at once, queues are formed. This brings costs (or losses) due to waiting which increase with the number of customers in the queue. To decrease the waiting costs and increase the service level guaranteeing better system performance different improvements can be implemented. However, any kind of improvement often is linked to a certain investment leading to higher costs of the queueing system. Figure 2 shows that it is always possible to establish the

optimal service level which ensures the minimal total costs of the queueing system performance.

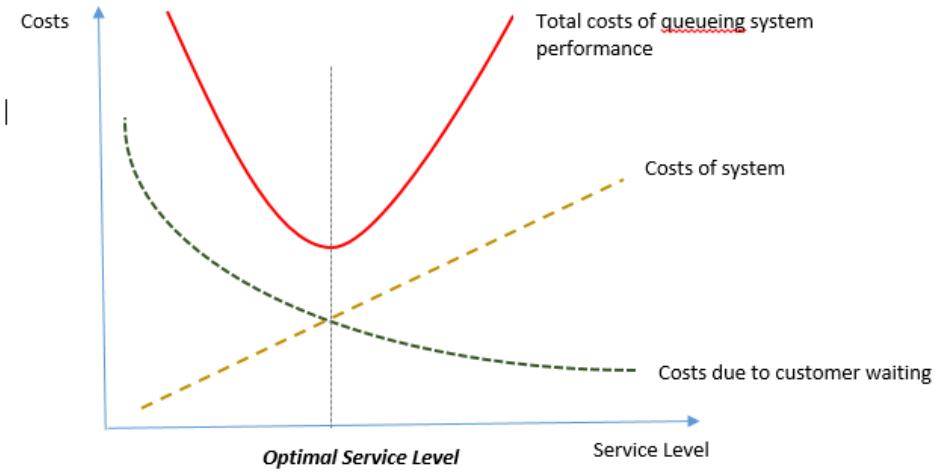

Figure 2: Optimal Service Level of a queueing system

To determine the optimal service level of a queueing system, different quantitative characteristics, like performance measures, can be used. The values of these measures can be calculated using an appropriate mathematical model. A suitable selection of the mathematical model is based on the following elements of the queueing system:

- Arrival process: Population of customers can be considered either limited (closed systems) or unlimited (open systems). Most mathematical models assume individual arrivals of customers and independent identically distributed interarrival times.
- Service mechanism is determined with the system capacity, availability and probability density function of service times. Most mathematical models assume that service times are independent identically distributed random variables.
- Queueing discipline represents the way the queue is organised (First-In-First-Out (FIFO), Last-In-First-Out (LIFO), random selection of customers

or selection based on customer priorities, for examples emergencies first). In the case there is only one server, or there are a number of equivalent and parallel servers, the queueing system is called simple. Simple queueing models use the standard remark for describing the probability density function of inter-arrivals and service times: M – a Poission process of the number of occurrences (i.e. customer arrivals or end of services); and an exponential density function of times between two successive events. G – a general distribution of times between two successive events (with a known mean and variance; for example, a normal density function). D – a deterministic situation; which means times between two successive events are constant. Notation M/M/c {infinity/infinity/FIFO} as a result describes the queueing system with c parallel serving channels, unlimited population, unlimited queue (no restriction for the maximum number of customers allowed to join the queue), First-In-First-Out queueing discipline, meanwhile both of the inter-arrival and the service times are distributed according to the exponential density function, for example Chassioti [25]. For many types of simple queueing models there exists closed formulations for most system performance measures. Assuming that we have the M/M/c {infinity/infinity/FIFO} queueing model the closed form of all four performance measures are accessible, like Tijms [15].

The expected waiting time can be calculated based on the following equation:

$$E(Wq) = \frac{1}{S} \frac{(c\rho)^2}{c!(1-\rho)^2 c\mu}$$

[1]

The expected number of waiting customers can be found according to the expression:

$$E(Nq) = \frac{1}{S} \frac{(c\rho)^2 \rho}{c!(1-\rho)^2}$$

[2]

The probability that one customer is going to wait because all agents are busy can be calculated as follows:

$$P_{wait} = \frac{1}{S} \frac{(c\rho)^2}{c!} \frac{1}{1-\rho}$$

[3]

In the literature, the equation (3) is also known as the Erlang C formula (like Tanner [14]; Garnett [27]) and plays a determinant role in the performance of the telephone systems. The service level is the most frequent measure of quality of the call centers service. It is determined by a given percentile of the waiting time distribution that is given by the following expression:

$$SL(t_0) = P[W_q \leq t_0] = 1 - \frac{1}{S}\frac{(c\rho)^c}{c!(1-\rho)}\exp(-(1-\rho)c\mu t_0)$$

[4]

The equation (4) gives the long-term fraction of customers whose waiting time Wq in the queue is no larger than a given limit t_0 (Avramidis [23]). The symbols used in equations (1), (2), (3) and (4) stand for:

c – number of serving channels

λ – arrival rate; 1/λ is the expected time between two successive arrivals

μ – service rate; 1/μ is the expected service time

ρ – traffic intensity calculated as $\rho = \lambda / c\,\mu$

S – the sum which can be calculated by the following expression:

$$S = 1 + c\rho + \frac{c\rho^2}{2!} + \cdots + \frac{c\rho^{c-1}}{(c-1)!} + \frac{c\rho^c}{c!}\frac{1}{1-\rho}$$

[5]

Equations (1), (2), (3) and (4) make sense when $S < \infty$. This condition stands if $\rho < 1$. The condition $\rho < 1$ ensures that the steady state distribution exists. In this case the infinite queues are not formed and the queueing system still operates after a long run. The minimum number of servers c_{min} needed to satisfy the regular state condition is the lowest integer that fulfil the equation

$$c > \lambda / \mu \quad [6]$$

6. Our case: Call center as a queueing system

The presented research was conducted on the case of an Albanian outsourcing call center, which works with different contractors, providing different kinds of services. The part of the call center that we are going to

analyze is one inbound project, which provides online tickets and customer service to an airline company. The project is opened from 8:00AM until 12:00PM. It employes 10 full time agents. The schedule of agents is defined based on prior experiences and taking into consideration the historical distribution of the calls. No deep analysis of the schedule has been performed until now. Customers call a single phone number. If at least one of the agents is available when a customer calls, he answers the call and serves that customer. If all of the agents are busy the calling customer is not rejected but can wait for a free agent without considering the number of customers in the queue. The principal scheme of the call center is presented in Figure 3.

Figure 3: Call center as queueing system

The call centers project under consideration can be treated as a simple queueing system, where the number of servers is determined with the number of active agents and the queueing discipline is FIFO.

The key element of the call centers optimization is the determination of the adequate number of active agents for different periods of the day. For this purpose the number of calls in different time periods of the day on a typical working week was taken into consideration and analyzed. The working day of the call centers project was divided into the following four slots: from 8:00 AM to 10:00 AM, from 10:00 AM to 1:00 PM, from 1:00 PM to 6:00 PM and from 6:00 PM to 12:00 PM. The number of calls analysis is presented in Figure 4.

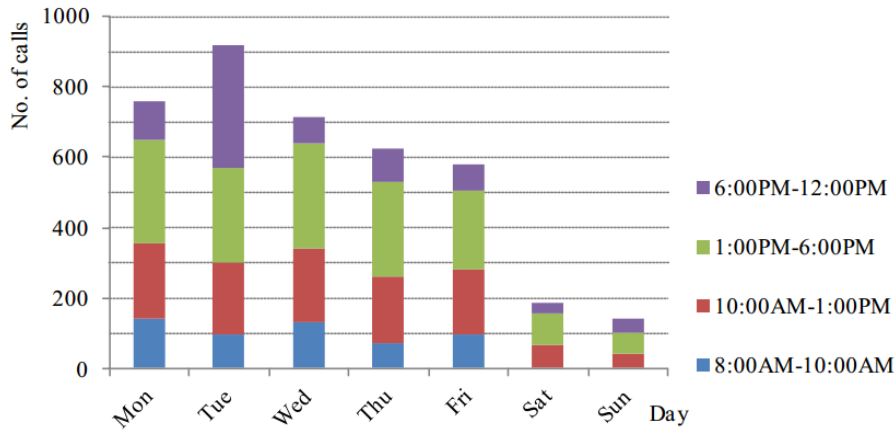

Figure 4: Calls distribution

The analysis showes that the number of incoming calls is significantly lower during weekends than during working days. As a result, weekend days were excluded from the analysis. The number of calls in the specific period is like almost all working days. The exception of the last statement, is the last period during Tuesdays. It was found that this deviation was caused by the unexpected stop-time of one of the principal provider's services. The third period (1:00PM to 6:00PM) is the most common period, whereas the morning and evening periods are less frequent.

7. Queueing model selection

To select a suitable theoretical mathematical model to describe the call center of our case, the distribution of inter-arrival times and the distribution of service times have to be evaluated. Figure 5 demonstrates the frequency distribution of interarrival times while Figure 6 shows the frequency distribution of service times in different periods of a typical working day.

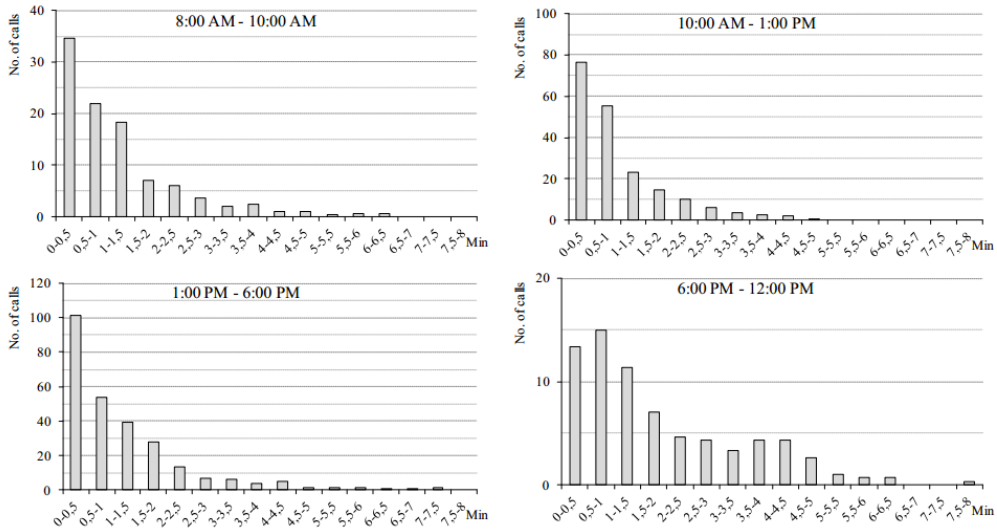

Figure 5: Frequency distribution of interarrival times

We can conclude from Figure 5 that inter-arrivals times in all four slots fit the exponential density function. Additionally, it can be seen from Figure 6 that probability density function of service times follows an asymmetric function. When calls shorter than one minute are excluded, we can presume that also the distribution of service times can be described by the exponential density function in all four periods of a working day. Since these short calls do not cause queues and as a result do not threaten the efficiency of the call center, our assumption is justified. From the description of the call centers organization and from the analysis of arrival and service patterns we can come into the conclusion that the call centers project under consideration can be described by the M/M/c {infinity/infinity/FIFO} queueing model.

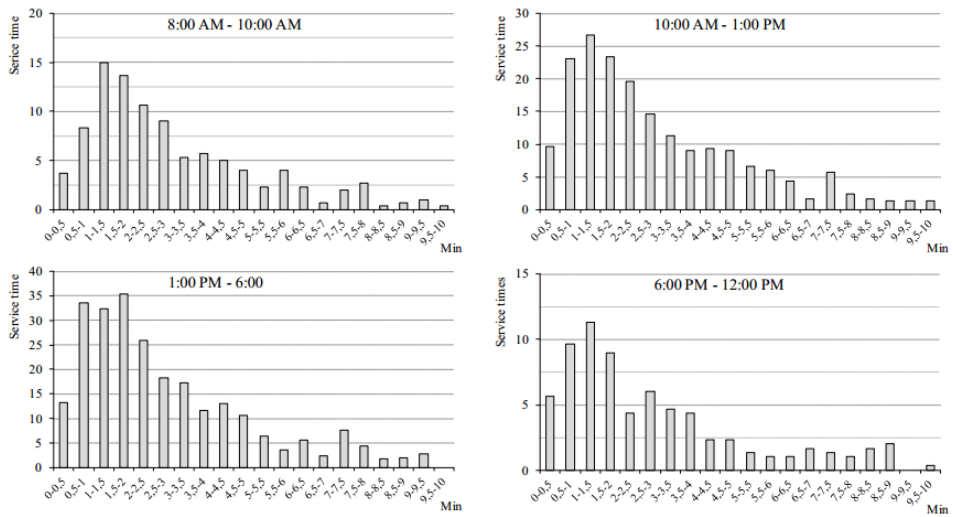

Figure 6: Frequency distribution of service times

8. Definition of performance measures

To analyse the call centers' service quality different performance measures can be used. We limit ourselves to performance measures that depend on customers' waiting time before their calls are handled. The following performance measures will be considered:

- the expected waiting time (1)
- the expected number of waiting customers (2)
- the probability that the customer arriving in the system is going to wait (3)
- the call centers' service level (4)

a. Optimization Results

The first step of the optimization is definition of the regular (steady) state conditions. From the field data the parameters λ and μ were estimated for a specific period of a working day. The value c_{min} needed to satisfy the regular (steady) state condition was determined according to (6). Results are given in Table 1.

Table 1: The steady state conditions for the call center

Period	λ $[min^{-1}]$	μ $[min^{-1}]$	λ/μ	c_{min}
8:00AM - 10:00AM	0.847	0.336	2.25	3
10:00AM - 1:00PM	1.053	0.342	3.08	4
1:00PM - 6:00PM	0.877	0.356	2.46	3
6:00PM - 12:00PM	0.532	0.356	1.49	2

Source: Author's illustration

The following step is setting up the optimization requirement regarding the chosen performance measures. Considering four service measures that we have selected, we define four optimization requirements.
The optimization requirement nr 1: The expected waiting time in a specific time period should not be longer than 20 seconds. The mathematical expression of this requirement is:

$$E(W_q) \leq 20sec = 0.33min$$

Optimization results assuming the requirement 1 are given in Table 2. In the first column of Table 2 the expected waiting time is calculated based on (1) and (5) considering the regular (steady) state condition c_{min} from Table 1. Then the minimal number of servers c needed to accomplish the stated requirement, and corresponding expected waiting time are determined by iteration. Results are shown in the second and in the third column of Table 2.

Table 2: Results of the call center optimization assuming the requirement $E(W_q) \leq 20sec$

Period	$E(W_q)$ **[min]-steady state**	c_{min}	$E(W_q)$ **[min]**
8:00AM - 10:00AM	4.34	5	0.16
10:00AM - 1:00PM	1.72	6	0.11
1:00PM - 6:00PM	3.57	5	0.14
6:00PM - 12:00PM	3.55	4	0.08

Source: Author's illustration

The optimization requirement nr 2: The expected number of waiting customers in a specific time period should not exceed 1,5. The mathematical expression of this requirement is:

$E(N_q) \leq 1.5$

Optimization results assuming the requirement nr 2 are given in Table 3. In the first column of Table 3 the expected number of waiting customers is calculated based on (2) and (5) considering the regular state condition c_{min} from Table 1. Then the minimal number of serving channel c needed to accomplish the optimization requirement stated above, and the corresponding expected number of waiting customers are determined by iteration. Results are shown in the second and in the third column of Table 3.

Table 3: Results of the call center optimization assuming the requirement $E(N_q) \leq 1.5$

Period	$E(N_q)$- steady state	c_{min}	$E(N_q)$
8:00AM - 10:00AM	3.76	4	0.56
10:00AM - 1:00PM	1.82	5	0.41
1:00PM - 6:00PM	3.13	4	0.49
6:00PM - 12:00PM	1.89	3	0.23

Source: Author's illustration

The optimization requirement nr 3: It should be assured that at least 80% of calling customers are served immediately, and will not have to wait for an available agent. This indicates that the probability of waiting should not exceed 20%:

$P_{wait} \leq 0.2$

Optimization results assuming the requirement nr 3 are given in Table 4. In the first column of Table 4 the probability of waiting is calculated based on (3) and (5) considering the regular (steady) state condition c_{min} from Table 1. Then the minimal number of serving channels c needed to accomplish the optimization requirement stated, and corresponding waiting probability are determined by iteration. Results are shown in the second and the third column of Table 4.

Table 4: Results of the call center optimization assuming the requirement $P_{wait} \leq 0.2$

Period	P_{wait} - steady state	c_{min}	P_{wait}
8:00AM - 10:00AM	0.714	5	0.134
10:00AM - 1:00PM	0.543	6	0.11
1:00PM - 6:00PM	0.682	5	0.124
6:00PM - 12:00PM	0.639	4	0.074

Source: Author's illustration

The optimization requirement nr 4: An industry standard for telephone services is the 80/20 rule, under which at least 80% of the customers must wait no more than 20 seconds (Koole and Mandelbaum [22]). The mathematical expression of this requirement is:

$$SL(20sec) = SL(0.33min) \geq 0.8$$

Optimization results assuming the requirement nr 4 are given in Table 5. In the first column of Table 5 the service level SL(20sec) considering the regular (steady) state condition c_{min} from Table 1 is calculated based on (4) and (5). Then the minimal number of serving channels c needed to accomplish the optimization requirement nr 4, and corresponding service level is determined by iteration. Results are shown in the second and in the third column of Table 5.

Table 5: Results of the call center optimization assuming the requirement

$$SL(20sec) = SL(0.33min) \geq 0.8$$

Period	SL(20sec) - steady state	c_{min}	SL(20sec)
8:00AM - 10:00AM	0.323	5	0.898
10:00AM - 1:00PM	0.511	6	0.921
1:00PM - 6:00PM	0.36	5	0.908
6:00PM - 12:00PM	0.398	3	0.803

Source: Author's illustration

9. Discussion of the results

The arrival and the service patterns of the Albanians call centers' project were analysed. It was established that the call centers' project under consideration can be described by the M/M/c {infinity/infinity/FIFO} queueing model. Four performance measures, dependent on customers' waiting time, were used to analyse the call centers' service quality. Taking into consideration the selected performance measures, four optimization requirements were settled as follows:

- the expected waiting time should not surpass 20 seconds,
- the expected number of customers waiting should not surpass 1.5,
- at most 20% calling customers will have to wait to the operator, at least 80% of them should be served immediately,
- at least 80% of the customers should wait no more than 20 seconds.

The goal was to determine the minimal number of serving channels in a particular period of a working day to accomplish the stated requirement. Results obtained prove that all performance measures can be applied to practice quite easily. Particularly useful seem to be the probability that the calling customer will have to wait (differently called Erlang C function) and the service level which is the most frequent measure of the call center service quality. The optimization results requirements showed that the number of serving channels (agents) needed to satisfy the regular (steady) state condition is relatively low (from two to four agents in the particular time period). This assures that the queueing system still operates after a long run without forming infinite queues (so $p<1$). However, if the call centers' project under consideration wants to improve the quality of its service, the number of the agents has to be raised. For each time period, two additional agents have to be added. In the first period, three out of four optimization requirements identified the minimum number of agents to be five. Further, the second period, being the most occupied, six agents are needed to satisfy three out of four requirements. In the third period, again three out of four requirements showed that the minimum number of agents needed is five. The last, night period has the lowest frequency of incoming calls. As a result, in two out of four requirements the minimum number of agents proved to be four. Considering the presented results, the call center

can operate at a satisfying service quality level during the weekdays with 10 agents. A typical working day should employ three variations: a) five agents for the slot from 8:00 AM to 4:00 PM; b) four agents for the slot from 4:00 PM to 24:00 operators; and c) one operator for the shift from 10:00 AM to 6:00 PM

10. Conclusions

The paper represents a case study of a call centers' performance optimization using the queueing theory approach. This kind of research was conducted on the case of an Albanian outsourcing call centers' project. Using the field data on call centers' operation it was found out that the call centers' project under consideration can be described by the M/M/c {infinity/infinity/FIFO} queueing model. Four performance measures, dependent on customers' waiting time, were used to analyse the call centers' service quality. The minimal number of serving channels needed to accomplish the stated requirements in a particular period of a working day was defined.

Results obtained prove that stochastic queueing models represent a useful tool for a call centers' performance optimization. Since all the data necessary for mathematical analysis are usually available, implementation of such models is quite simple while the information they provided is of great importance. Especially, determination of the adequate number of active agents regarding a specific performance measure is a preparatory condition to ensure the optimal service level and as a result the minimal cost of the queueing system performance. Recent studies demonstrate that customers waiting time is not the only measure for the service level quality. As stated by Aksin [24], customers tend to place a high value on other important dimensions of their experience (like first call resolution, perceived agent competency, politeness or kindness). Therefore, if a need to model service quality according to concurrent customer values occurs, the function for optimal service level should be reconsidered. Additional research on scheduling and schedule adjustments, as for example presented by (Mehrotra [28]) has to be conducted to efficiently define the personnel schedules for the entire time of call centers' operation. Discrete event simulation is also an applicable option for accurate performance modelling

and subsequent decision support. Some authors, like Akhtar and Latif [29] argue that the analytical approach is not accurate enough, as it does not resemble randomness. In another research I will simulate the presented case with a discrete event simulation tool, where for explaining the probability density function of service times an asymmetric function will be used. In addition another goal is trying to involve other performance measures considering abandonment and retrials, which are other two important performance measures and that in the analytical model are not taken into consideraton.

11. References

[1] A. Gilmore and L. Moreland. Call centres: How can service quality be managed? Irish Marketing Review, 13:3–11, 2000

[2] L. Bennington, J. Commane, and P. Conn. Customer satisfaction and call centers: an Australian study. International Journal of Service Industry Management, 11:162–173, 2000

[3] R.A. Feinberg, I.-S. Kim, L. Hokama, K. de Ruyter, and C. Keen. Operational determinants of caller satisfaction in the call center. International Journal of Service Industry Management, 11:131–141, 2000

[4] G. Tom, M. Burns, and Y. Zeng. Your life on hold: The effect of telephone waiting time on customer perception. Journal of Direct Marketing, 11:25–31, 1997.

[5] A. Mandelbaum. Call centers (centres): Research bibliography with abstracts. Downloadable from ie.technion.ac.il/serveng/References/ccbib.pdf, 2001.

[6]] S.C. Borst, A. Mandelbaum, and M.I. Reiman. Dimensioning large call centers. Working paper, 2000.

[7] B. Andrews and H. Parsons. Establishing telephone-agent staffing levels through economic optimization. Interfaces, 23(2):14–20, 1993.

[8] O.Z. Ak¸sin and P.T. Harker. Capacity sizing in the presence of a common shared resource: Dimensioning an inbound call center. Working paper, 2001.

[9] E. Zohar, A. Mandelbaum, and N. Shimkin. Adaptive behavior of impatient customers in tele-queues: Theory and emperical support. working paper, 2000.

[10] A. Mandelbaum, A. Sakov, and S. Zeltyn. Empirical analysis of a call center. Downloadable from ie.technion.ac.il/serveng/References, 2001.

[11] G.M. Koole and H.J. van der Sluis. An optimal local search procedure for manpower scheduling in call centers. Technical Report WS-501, Vrije Universiteit Amsterdam, 1998. Electronically available at www.cs.vu.nl/obp/callcenters.

[12] van Dijk, N. M. and Bouchiere, R. J. (2011), Queueing Networks: A

Fundamental Approach, New York, Springer.

[13] Gross, D., Shorttle, J. F., Thompson, J. M. and Harris, C. M. (2008), Fundamentals of Queueing Theory, New Jersey, Wiley Series in Probability and Statistics.

[14] Tanner, M. (1995), Practical Queueing Analysis, London, The IBM McGraw-Hill Series.

[15] Tijms, H. C. (2003), A First Course in Stohastic Models, Chichester, Wiley.

[16] Attahirusule, A. (2010), Queueing Theory for Telecommunications: Discrete Time Modelling of a Single Node System, Springer, New York.

[17] Giambene, G. (2005), Queueing theory and telecommunicaitons: networks and application, New York, Springer.

[18] Robertazzi, T. G. (2000), Computer Networks and Systems: Queueing Theory and Performance Evaluation, 3rd ed, New York, Springer.

[19] Ismail, I. A., Mokaddis, G. S., Metwally, S. A. & Metry, M. K. (2011), „Optimal Treatment of Queueing Model for Highway“, Journal of Computations & Modelling, Vol. 1 No. 1, pp. 61-71.

[20] Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S. and Zhao, L. (2005), „Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective“, Journal of the American Statistical Association, Vol. 100 No. 469, pp. 36-50.

[21] Koole, G. (2007), „Call Center Mathematic: A scientific method for understanding and improving contact centers“, available at http://www.academia.edu/542467/Call_center_mathematics / (12 June 2013).

[22] Koole, G. and Mandelbaum, A. (2001), „Queueing Models of Call Centers: An Introduction“, available at http://www.columbia.edu/~ww2040/cc_review.pdf / (17 June 2013).

[23] Avramidis, A. N., Chan, W., Gendreau, M., L'Ecuyer, P. and Piscane, O. (2010), „Optimizing daily agent scheduling in a multiskill call center. European Journal of Operational Research“, Vol. 200 No. 3, pp. 822-832.

[24] Aksin, Z., Armony, M. and Mehrotra, V. (2007), „The Modern Call Center: A MultiDisciplinary Perspective on Operations Management Research“, Production and Operations anagement, Vol. 16 No. 6, pp. 665-688.

[25] Chassioti, E. (2005), Queueing Models for Call Centres (PhD thesis), Lancaster, Lancaster University Management School.

[26] Dombacher, C. (2010), „Queueing Models for Call Centres“, available at http://www.telecomm.at/documents/Queueing_Models_CC.pdf / (12 June 2013)

[27] Garnett, O., Mandelbaum, A. and Reiman, M. (2001), „Designing a call center with impatient customers“, Manufacturing and Service Operations Management, Vol. 4 No. 3, pp. 208-227.

[28] Mehrotra, V., Ozlük, O. and Saltzman, R. (2010), „Intelligent Procedures for IntraDay Updating of Call Center Agent Schedules“, Production and Operations Management, Vol. 19 No. 3, pp. 353-367.

[29] Akhtar, S. and Latif, M. (2010), "Exploiting Simulation for Call Centre Optimization", in Ao, S.I., Gelman, L., Hukins, D.W.L., Hunter, A., Korsunsky, A.M. (Eds.), Proceedings of the World Congress on Engineering Vol. III, Newswood Limited, London, pp. 2963-2970.