

# Network Stochastic Call Center Model

Tatiana V. Rusilko

*Yanka Kupala State University of Grodno, 22 Ozheshko St, Grodno, 230023, Belarus*

## Abstract

A closed exponential queueing network with impatient customers is considered as a model of a call center. The process of handling calls in the call center is described. A detailed description of the model is given. The stay of the customer at a specific network node and its routing between the nodes correspond to the customer's call status in the call center and the process of its routing between telephone agents of different levels. The process of changing the number of customers at the nodes is studied under the asymptotic assumption of a large number of customers. It is proved that its probability density function satisfies the Fokker–Planck–Kolmogorov equation. The system of differential equations for the expected value of this process is derived. The solution of this system allows predicting the dynamics of the expected number of customers in the network nodes and regulating the parameters of the call center operation. The asymptotic technique used is applicable both in transient and steady-state. The problem of finding the optimal staffing list of agents to maximize call center performance is investigated. The areas of implementation of research results are the design of call centers and improving their performance.

## Keywords 1

Call center, queueing network, impatient customers, mathematical modeling

## 1. Introduction

A call center is a specialized organization or a centralized department located within a company in which telephone calls from customers are handled by a team of advisors, otherwise known as agents. Call centers can handle inbound and outbound calls via voice communication channels, although some may specialize in one or the other. The article [1] concentrates on review the state of research on telephone call centers, this article begins with a tutorial on how call centers function and proceed to survey academic research devoted to the management of their operations.

The quality and operational efficiency of these telephone services must be exceptional to meet the needs of the customers [2]. Agents of a large best-practice call center have to cater to thousands of phone callers per hour. The waiting time of delayed calls has not to exceed a few seconds. To achieve high levels of service quality and efficiency it is a necessary understanding of the scientific principles underlying call-center operations. Undoubtedly sophisticated approaches are needed to accurately describe the reality of contact-center operations, and mathematical modeling of this reality can improve call center performance significantly [1, 3, 4]. Call center capabilities also include the accumulation of a call queue in standby mode, waiting time information, call distribution between agents, call classification, and interaction strategy determination depending on the client's status.

The operational service level of call centers is typically quantified in terms of some efficiency indices or performance measures. Among them, the following can be noted: workload, placing calls by agents, expected call processing time, percentage of successfully processed calls, duration of the service cycle of one call, the fraction of customers that abandon the queue before being served, queue lengths, waiting time.

---

Proceedings of VI International Scientific and Practical Conference Distance Learning Technologies (DLT–2021), September 20–22, 2021, Yalta, Crimea

EMAIL: [tatiana.rusilko@gmail.com](mailto:tatiana.rusilko@gmail.com)

ORCID: 0000-0002-4880-0619



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Having analyzed the above organization of call centers, we conclude that queueing theory can be used to analyze their efficiency, predict the load and optimize the structure. Performance measures listed above also underscore the natural fit between queueing models and call centers [5]. The emergence of the queueing theory is associated with the ordering and optimization of the operation of telephone exchanges. Queueing systems, in particular, Markovian ones, are widely used for call center modeling and performance analysis [6, 7, 8, 9, 10]. Queueing systems with impatient customers [7, 10], retrial queueing systems [6, 9], and many others have originated from the need for mathematical modeling of the behavior of customers of telephone services.

Call centers often have a branched or multilevel structure. The functions of the first-level call center agents include call registration and provision of general information. When the topic of a client request goes beyond the specialization of current level agents, the client is redirected to a higher level, whose specialists are competent in resolving the issue. Thus, calls are routed according to their specificity between different levels of agents. Hence the call center organization structure can be graphically depicted by a network diagram showing the hierarchy of call center agents, routing, and call forwarding according to their specifics between different levels of advisors. Thus, queueing networks are suitable for modeling multilevel or multi-stage call centers. A queueing network is a collection of interconnected queueing systems, between which customers (jobs, requests) circulate. Each queueing system, which is also called a node, has at least one server and a queue. The main classification of queueing networks is based on the location of the customer source, the flow of arriving customers, the service time probability distribution, the scheduling policy at queueing nodes, and the customer routing.

The purpose of this paper is call center mathematical modeling, optimization, and analysis of the effectiveness of the call processing process using closed exponential queueing networks with impatient customers. The network is studied under the asymptotic assumption of a large number of customers [11, 12]. Thus, the greater the number of customers in the network, the greater the precision in the calculation [13].

## 2. Model Description. Problem Formulation

Our focus in this article is on the inbound call center of a large company with a limited number of clients. We denote this number of customers by  $K$ ,  $K \gg 1$ . Inbound call centers handle incoming calls that are initiated by outside callers calling into a center. Typically, these types of centers provide customer support, help-desk services, reservation and sales support for airlines and hotels, and order-taking functions for catalog and Web-based merchants.

A closed queueing network is proposed as a model. The network consists of  $n + 1$  Markov  $m_i$ -linear queueing systems  $S_i$ ,  $i = \overline{1, n + 1}$ , and  $K$ -linear hypothetical system  $S_0$ , which plays the role of an external environment and is considered as a dependable source of arriving customers. Each call to the call center corresponds to the customer's arrival in the queueing network. The stay of the customer at a specific network node and its routing between the nodes correspond to the customer's call status in the call center and the process of its routing between telephone agents of different levels. Let's describe the process of handling calls in the call center.

Consider a company's two-level call center that handles calls on  $n$  topics. Each customer can be in one of the following states:

- $S_0$  – no need to turn to the call center,
- $S_{n+1}$  – call is required service by one of  $m_{n+1}$  first-level agents who register and distribute calls depending on their topic,
- $S_i$  – call is required service by one of the  $m_i$  second-level agents consulting on the  $i$ th topic,  $i = \overline{1, n}$ .

Accordingly, each customer of the queueing network can be located in one of the queueing systems  $S_i$ ,  $i = \overline{0, n + 1}$ .

The customer transition from state  $S_0$  to state  $S_{n+1}$  corresponds to a customer's call to the call center, which occurs at a random time, regardless of the state of other customers, but depends on the time of day. The probability of this transition in the time interval  $[t, t + \Delta t]$  is equal to  $\lambda(t)\Delta t + o(\Delta t)$ , where

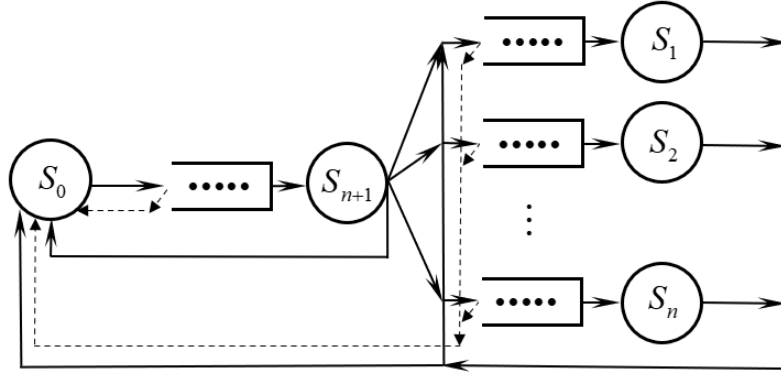
$\lambda(t)$  is the arrival rate (the expected time between each call arriving). The inbound call flow is non-stationary.

From state  $S_{n+1}$  the call is routed to the state  $S_i$  with probability  $p_{(n+1)i}$ ,  $i = \overline{1, n}$ , or with probability  $p_{(n+1)0}$  to stage  $S_0$  in case of an erroneous customer request. After the end of call service in the state  $S_i$ ,  $i = \overline{1, n}$ , the customer goes to  $S_0$  with probability  $p_{i0}$ , if his request is completely fulfilled or redirected to specialists on another topic to stage  $S_j$ ,  $j = \overline{1, n}$ , with probability  $p_{ij}$ ,  $i \neq j$ , if the agents of the state  $S_i$  have a lack of competence to answer the customer's questions. The served customer transition matrix is stochastic  $(p_{ij})_{(n+2) \times (n+2)}$  with the following nonzero entries:

$$p_{0(n+1)} = 1; p_{(n+1)i}, i = \overline{0, n}, \sum_{i=0}^n p_{(n+1)i} = 1;$$

$$p_{ij}, i = \overline{1, n}, j = \overline{0, n}, i \neq j, \sum_{j=0}^n p_{ij} = 1.$$

The customer routing in the network model of the call center is depicted with a solid line in Figure 1.



**Figure 1:** The state-transition diagram of served and impatient customers

If at least one of the agents of the system  $S_i$  is idle, then the incoming call immediately begins to be served at the moment of transition to state  $S_i$ ,  $i = \overline{1, n+1}$ . Calls are served on a first-come, first-served basis. For the queueing network, this means that service requests are selected accordingly to the discipline FIFO. All  $m_i$  agents (servers) at the state (node)  $S_i$  are identical and they have exponentially distributed service time with rate  $\mu_i$ , i.e.,  $\mu_i^{-1}$  is the mean service time,  $i = \overline{1, n+1}$ .

If all  $m_i$  agents in  $S_i$  are busy, then the customer waits in an unlimited queue. However, the client waiting time in the queue of the system  $S_i$  is limited by an exponential random variable  $\tau$  with an expected value of  $\eta_i^{-1}$ ,  $i = \overline{1, n+1}$ . Under this condition, customers in the queueing network are called impatient. If during time  $\tau$  the call is not answered, then it is lost, i.e., with a probability equal to one, the customer after this time passes into the external environment  $S_0$ . The impatient (unfulfilled) customer transition matrix is  $(q_{ij})_{(n+1) \times (n+2)}$  with the following nonzero entries:

$$q_{i0} = 1, i = \overline{1, n+1}.$$

Let  $k_i(t)$  denote the number of customers in the state  $S_i$  at the time  $t$ ,  $i = \overline{1, n+1}$ . It's obvious that  $\sum_{i=0}^{n+1} k_i(t) = K$ . The allocation of customers according to possible states at time  $t$  fully describes the situation in the call center at that time. Accordingly, the allocation of customers by queueing nodes completely determines the state of the queueing network. Thus, the state of the network model under study at time  $t$  is represented by a vector

$$k(t) = (k_1(t), k_2(t), \dots, k_{n+1}(t)).$$

Taking into account the above-described, the process  $k(t)$  of changing the number of customers at queueing nodes is a continuous-time Markov chain with a finite state space. It is clear, the number of customers in the external node  $S_0$  is  $k_0(t) = K - \sum_{i=1}^{n+1} k_i(t)$ .

The following problem is of interest: taking into account the benefits of customer calls and considering agents wage fund, it is required to find such a staffing list (a set of servers in the network nodes) that would ensure the maximum call center performance, provided that  $K, \lambda(t), \mu_i, \eta_i, p_{ij}$  are fixed  $i = \overline{1, n+1}, j = \overline{0, n+1}$ .

Let us introduce the following notation:

- $F_i$  is the monetary effect (or some kind of efficiency) generated by servicing one call in state  $S_i$  per unit of time,  $i = \overline{1, n+1}$ ;
- $L_i$  is the wage of one agent in the system  $S_i$  per unit of time,  $i = \overline{1, n+1}$ ;
- $F_0$  is the monetary effect created by one customer when he does not need to turn to the call center (it can be equal to zero).

Then the total monetary effect of the call center at time  $t$  is given by the following expression

$$F(t) = F_0 \left( K - \sum_{i=1}^{n+1} \min(m_i, k_i(t)) \right) + \sum_{i=1}^{n+1} F_i \min(m_i, k_i(t)) - \sum_{i=1}^{n+1} L_i m_i. \quad (1)$$

Since  $k_i(t)$  are components of a random process, the productivity  $F(t)$  is also a random process. Applying the mathematical expectation operator  $E$  to both sides of the expression (1), we obtain the following expression for the expected monetary effect of a call center:

$$E_F(t) = F_0 K + \sum_{i=1}^{n+1} (F_i - F_0) \min(m_i, E_{k_i}(t)) - \sum_{i=1}^{n+1} L_i m_i. \quad (2)$$

It is assumed that call center operation is worthwhile if  $E_F(t) > 0$ , and its efficiency increases with  $E_F(t)$ . It is interesting to consider the problem of maximizing the functional (2) by the number of servers  $m_i$  in queueing systems  $S_i, i = \overline{1, n+1}$ , provided that the statistical characteristics of customer behavior are fixed. In the simplest case of the assumption about the absence of queues, the mathematical problem of maximizing the average efficiency at the time  $t$  is given by

$$\begin{cases} E_F(t) = F_0 K + \sum_{i=1}^{n+1} (F_i - F_0) E_{k_i}(t) - \sum_{i=1}^{n+1} L_i m_i \rightarrow \max_{m_i, i=\overline{1, n+1}} \\ E_{k_i}(t) \leq m_i, i = \overline{1, n+1}. \end{cases} \quad (3)$$

It is clear that to solve this problem, first of all, it is necessary to determine the probability distribution of the process  $k(t)$ . Only then can the expected values  $E_{k_i}(t)$  be determined,  $i = \overline{1, n+1}$ . The distribution of  $k(t)$  parametrically depends on the number of servers in queueing systems.

### 3. Asymptotic Analysis of the Network Model

Asymptotic analysis implies an approximation method of queueing network study under the assumption of a large number of customers. The mathematical approach used in this paper is based on a discrete model of a continuous Markov process described in many books on the theory of diffusion Markov processes [14, 15].

Let us introduce a  $(n+1)$ -vector of the form  $I_i = (\overbrace{0, 0, \dots, 0}^i, 1, 0, \dots, 0)$  and Heaviside function  $H(x) = \begin{cases} 1, x > 0, \\ 0, x \leq 0. \end{cases}$  As mentioned above, the process  $k(t)$  is a continuous-time Markov chain with a finite state space. The assumptions made in the formulation of the problem determine that in the short time  $\Delta t$  the Markov process  $k(t) = (k, t)$  can make one of the following transitions to the state  $(k, t + \Delta t)$ :

- from the state  $(k - I_{n+1}, t)$  the process transfer to  $(k, t + \Delta t)$  with probability

$$\lambda(t) \left( K - \sum_{i=1}^{n+1} k_i(t) + 1 \right) \Delta t + o(\Delta t),$$

that corresponds to the customer transition from state  $S_0$  to state  $S_{n+1}$ ;

- from the state  $(k + I_i, t)$  the process transfer to  $(k, t + \Delta t)$  with probability  $\eta_i(k_i(t) + 1 - m_i)H(k_i(t) + 1 - m_i)\Delta t +$

$$+\mu_i p_{i0} \min(m_i, k_i(t) + 1)\Delta t + o(\Delta t),$$

that corresponds to the impatient customer transition from state  $S_i$  to state  $S_0$ ,  $i = \overline{1, n+1}$ ;

- from the state  $(k + I_i - I_j, t)$  the process transfer to  $(k, t + \Delta t)$  with probability  $\mu_i p_{ij} \min(m_i, k_i(t) + 1)\Delta t + o(\Delta t)$ ,

that corresponds to the served customer transition from state  $S_i$  to state  $S_j$ ,  $i = \overline{1, n+1}, j = \overline{1, n}$ ;

- from the state  $(k, t)$  to  $(k, t + \Delta t)$  with probability

$$1 - \left( \lambda(t) \left( K - \sum_{i=1}^{n+1} k_i(t) \right) + \sum_{i,j=1}^{n+1} \mu_i p_{ij} \min(m_i, k_i(t)) + \sum_{i=1}^{n+1} (\mu_i p_{i0} \min(m_i, k_i(t)) + \eta_i (k_i(t) - m_i) H(k_i(t) - m_i)) \right) \Delta t + o(\Delta t),$$

that corresponds to no customers transfer;

- from other states the process transfer to  $(k, t + \Delta t)$  with probability  $o(\Delta t)$ .

The probabilities  $p_{ij}$  are entries of the state transition matrix  $(p_{ij})_{(n+2) \times (n+2)}$  that was specified in the model description section. For example,  $p_{i(n+1)} = 0$ ,  $i = \overline{1, n+1}$ .

Having regard to the transitions listed above in the short time  $\Delta t$ , using the law of total probability, the following system of equations is valid for the probability  $P(k, t) = P(k(t) = k)$ :

$$\begin{aligned} P(k, t + \Delta t) &= P(k - I_{n+1}, t) \lambda(t) \left( K - \sum_{i=1}^{n+1} k_i(t) + 1 \right) \Delta t + \\ &+ \sum_{i=1}^{n+1} P(k + I_i, t) (\mu_i p_{i0} \min(m_i, k_i(t) + 1) + \eta_i (k_i(t) + 1 - m_i) H(k_i(t) + 1 - m_i)) \Delta t + \\ &+ \sum_{i,j=1}^{n+1} P(k + I_i - I_j, t) \mu_i p_{ij} \min(m_i, k_i(t) + 1) \Delta t + \\ &+ P(k, t) \left( 1 - \left( \lambda(t) \left( K - \sum_{i=1}^{n+1} k_i(t) \right) + \sum_{i,j=1}^{n+1} \mu_i p_{ij} \min(m_i, k_i(t)) + \sum_{i=1}^{n+1} (\mu_i p_{i0} \min(m_i, k_i(t)) + \eta_i (k_i(t) - m_i) H(k_i(t) - m_i)) \right) \Delta t \right) + o(\Delta t). \end{aligned}$$

We set

$$\begin{aligned} \Delta_{i0} P(k, t) &= P(k + I_i, t) - P(k, t), \\ \Delta_{0(n+1)} P(k, t) &= P(k - I_{n+1}, t) - P(k, t), \\ \Delta_{ij} P(k, t) &= P(k + I_i - I_j, t) - P(k, t), i = \overline{1, n+1}, j = \overline{1, n}. \end{aligned}$$

On the assumption that  $\Delta t \rightarrow 0$ , we pass to the Kolmogorov difference-differential equation for these probabilities:

$$\begin{aligned} \frac{\partial P(k, t)}{\partial t} &= \lambda(t) \left( K - \sum_{i=1}^{n+1} k_i(t) \right) \Delta_{0(n+1)} P(k, t) + \lambda(t) P(k - I_{n+1}, t) + \\ &+ \sum_{i=1}^{n+1} (\mu_i p_{i0} \min(m_i, k_i(t)) + \eta_i (k_i(t) - m_i) H(k_i(t) - m_i)) \Delta_{i0} P(k, t) + \\ &+ \sum_{i=1}^{n+1} \mu_i p_{i0} (\min(m_i, k_i(t) + 1) - \min(m_i, k_i(t))) P(k + I_i, t) + \end{aligned} \quad (4)$$

$$\begin{aligned}
& + \sum_{i=1}^{n+1} \eta_i p_{i0} ((k_i(t) + 1 - m_i)H(k_i(t) + 1 - m_i) - (k_i(t) - m_i)H(k_i(t) - m_i))P(k + I_i, t) + \\
& \quad + \sum_{i,j=1}^{n+1} \mu_i p_{ij} \min(m_i, k_i(t)) \Delta_{ij} P(k, t) + \\
& \quad + \sum_{i,j=1}^{n+1} \mu_i p_{ij} (\min(m_i, k_i(t) + 1) - \min(m_i, k_i(t))) P(k + I_i - I_j, t).
\end{aligned}$$

The last equation defies analytical solution for large  $n$ . Call centers undoubtedly handle a large number of customer calls. In connection with this, consider the important asymptotic case of a large number of customers  $K \gg 1$  and find the probability distribution of the state vector of the call center model with an accuracy of  $O(\varepsilon^2)$ , where  $\varepsilon = 1/K$ . The technique of asymptotic analysis of queueing networks described in the papers [11, 12, 16] is used.

Suppose that we are interested in the properties of the process  $\xi(t) = k(t)/K$  as  $K$  becomes very large. Vector  $\xi(t)$  indicates which part of the company's customers contacted the call center and how the calls were distributed by model states at time  $t$ . During the time  $\Delta t \rightarrow 0$ , the possible changes in the vector  $\xi(t)$  are  $e_i$ , where  $e_i = I_i \varepsilon = I_i/K$ . In the case when  $K \rightarrow \infty$  we have  $e_i \rightarrow 0$  and the vector

$$\xi(t) = \frac{k(t)}{K} = \left( \frac{k_1(t)}{K}, \frac{k_2(t)}{K}, \dots, \frac{k_{n+1}(t)}{K} \right)$$

will be continuous-time continuous-state Markov processes with a probability density function

$$\begin{aligned}
p(x, t) &= \lim_{\varepsilon \rightarrow 0} \frac{P(x_1 \leq \xi_1(t) < x_1 + \varepsilon, x_2 \leq \xi_2(t) < x_2 + \varepsilon, \dots, x_{n+1} \leq \xi_{n+1}(t) < x_{n+1} + \varepsilon)}{\varepsilon^{n+1}} = \\
&= \lim_{\varepsilon \rightarrow 0} \frac{P(Kx_1 \leq k_1(t) < Kx_1 + 1, Kx_2 \leq k_2(t) < Kx_2 + 1, \dots, Kx_{n+1} \leq k_{n+1}(t) < Kx_{n+1} + 1)}{\varepsilon^{n+1}},
\end{aligned}$$

i.e.,

$$K^{n+1} P(k, t) \xrightarrow{K \rightarrow \infty} p(x, t), \quad (5)$$

where  $x \in X$ ,

$$X = \left\{ x = (x_1, x_2, \dots, x_{n+1}): x_i \geq 0, i = \overline{1, n+1}, \sum_{i=1}^{n+1} x_i \leq 1 \right\}.$$

Realizing the asymptotic transition (5) for the equation (4), we obtain the following partial differential equation:

$$\begin{aligned}
\frac{\partial p(x, t)}{\partial t} &= \lambda(t) K \left( 1 - \sum_{i=1}^{n+1} x_i(t) \right) \Delta_{0(n+1)} p(x, t) + \lambda(t) p(x - e_{n+1}, t) + \\
&+ K \sum_{i=1}^{n+1} (\mu_i p_{i0} \min(l_i, x_i(t)) + \eta_i (x_i(t) - l_i) H(x_i(t) - l_i)) \Delta_{i0} p(x, t) + \\
&+ \sum_{i=1}^{n+1} \mu_i p_{i0} \frac{\partial \min(l_i, x_i)}{\partial x_i} p(x + e_i, t) + \sum_{i=1}^{n+1} \eta_i p_{i0} \frac{\partial (x_i(t) - l_i) H(x_i(t) - l_i)}{\partial x_i} p(x + e_i, t) + \\
&+ K \sum_{i,j=1}^{n+1} \mu_i p_{ij} \min(l_i, x_i(t)) \Delta_{ij} p(x, t) + \sum_{i,j=1}^{n+1} \mu_i p_{ij} \frac{\partial \min(l_i, x_i)}{\partial x_i} p(x + e_i - e_j, t).
\end{aligned} \quad (6)$$

If  $p(x, t)$  is twice continuously differentiable function for  $x$ , then

$$\begin{aligned}
p(x \pm e_i, t) &= p(x, t) \pm \varepsilon \frac{\partial p(x, t)}{\partial x_i} + \frac{\varepsilon^2}{2} \frac{\partial^2 p(x, t)}{\partial x_i^2} + o(\varepsilon^2), \\
p(x + e_i - e_j, t) &= p(x, t) + \varepsilon \left( \frac{\partial p(x, t)}{\partial x_i} - \frac{\partial p(x, t)}{\partial x_j} \right) + \\
&+ \frac{\varepsilon^2}{2} \left( \frac{\partial^2 p(x, t)}{\partial x_i^2} - 2 \frac{\partial^2 p(x, t)}{\partial x_i \partial x_j} + \frac{\partial^2 p(x, t)}{\partial x_j^2} \right) + o(\varepsilon^2), \quad i, j = \overline{1, n+1}.
\end{aligned}$$

Substituting this series into equation (6) and grouping terms, we obtain that  $p(x, t)$  satisfies the multidimensional Fokker–Planck–Kolmogorov equation (7):

$$\frac{\partial p(x, t)}{\partial t} = - \sum_{i=1}^{n+1} \frac{\partial}{\partial x_i} (A_i(x, t)p(x, t)) + \frac{\varepsilon}{2} \sum_{i,j=1}^{n+1} \frac{\partial^2}{\partial x_i \partial x_j} (B_{ij}(x, t)p(x, t)), \quad (7)$$

with drifts

$$\begin{aligned}
A_i(x, t) &= \sum_{j=1}^{n+1} \mu_j (p_{ji} - \delta_{ji}) \min(l_j, x_j(t)) - \eta_i (x_i(t) - l_i) H(x_i(t) - l_i) + \\
&+ \lambda(t) p_{0i} \left( 1 - \sum_{j=1}^{n+1} x_j(t) \right), \quad (8)
\end{aligned}$$

and diffusion coefficients

$$\begin{aligned}
B_{ii}(x, t) &= \sum_{j=1}^{n+1} \mu_j (p_{ji} + \delta_{ji}) \min(l_j, x_j(t)) + \eta_i (x_i(t) - l_i) H(x_i(t) - l_i) + \\
&+ \lambda(t) p_{0i} \left( 1 - \sum_{j=1}^{n+1} x_k(t) \right), \\
B_{ij}(x, t) &= -\mu_i p_{ij} \min(l_i, x_i(t)), \quad i \neq j;
\end{aligned}$$

$\delta_{ij}$  is the Kronecker delta,  $i, j = \overline{1, n+1}$ .

The error in this approximation is no more than  $\varepsilon^2 = (1/K)^2$ .

It is known [14, 17] that the expected value  $E_{\xi_i}(t)$  are determined with an accuracy of  $O(1/K^2)$  from a system of ordinary differential equations:

$$\frac{dE_{\xi_i}(t)}{dt} = A_i(E_{\xi_i}(t), t), \quad i = \overline{1, n+1},$$

where  $A_i(x, t)$  is drift given by formula (8). The proof of this statement, based on the application of characteristic functions, is presented in the article [18]. The system of differential equations for the components of vector  $k(t) = K\xi(t)$  is given by

$$\frac{dE_{k_i}(t)}{dt} = KA_i \left( \frac{1}{K} E_{k_i}(t), t \right), \quad i = \overline{1, n+1}. \quad (9)$$

Taking into account (9), (8), and the transition matrix  $(p_{ij})_{(n+2) \times (n+2)}$  of the network model, the expected number of customers  $E_{k_i}(t)$  in state  $S_i$ ,  $i = \overline{1, n+1}$ , can be found by solving the following system:

$$\begin{aligned}
\frac{dE_{k_i}(t)}{dt} &= \sum_{j=1}^{n+1} \mu_j p_{ji} \min(m_j, E_{k_j}(t)) - \mu_i \min(m_i, E_{k_i}(t)) - \\
&- \eta_i (E_{k_i}(t) - m_i) \theta(E_{k_i}(t) - m_i), \quad i = \overline{1, n}, \quad (10)
\end{aligned}$$

$$\begin{aligned}
\frac{dE_{k_{n+1}}(t)}{dt} &= \lambda(t) \left( K - \sum_{i=1}^{n+1} E_{k_i}(t) \right) - \mu_{n+1} \min(m_{n+1}, E_{k_{n+1}}(t)) - \\
&- \eta_{n+1} (E_{k_{n+1}}(t) - m_{n+1}) \theta(E_{k_{n+1}}(t) - m_{n+1}).
\end{aligned}$$

The expected number of customers in the external environment is

$$E_{k_0}(t) = K - \sum_{i=1}^{n+1} E_{k_i}(t).$$

Solution of the system (10) with a certain initial condition, firstly, makes it possible to predict the average number of calls in each model state, and, secondly, it is necessary to solve the problem of maximizing the call center performance (3). To solve (10), it is possible to use numerical methods or to solve the system analytically in the linearity regions of the right-hand side. Having solved (10) for different sets of the number of agents  $m = (m_1, m_2, \dots, m_{n+1})$ , substituting this solution into (2) and comparing the values of the expected productivity  $E_F(t)$  for each  $m = (m_1, m_2, \dots, m_{n+1})$ , we find the optimal set  $m$  for which  $E_F(t)$  takes the largest value at a fixed time  $t$ .

It is well established that under certain conditions, the network reaches "steady-state", the stationary value of the expected productivity is  $E_F = \lim_{t \rightarrow \infty} E_F(t)$ . The steady-state means  $E_{k_i} = \lim_{t \rightarrow \infty} E_{k_i}(t)$ , where  $i = \overline{1, n+1}$ , can be found from a system of algebraic balance equations of the form (10), where the left-hand sides are equal to zero. In this case, the arrival rate should be constant:  $\lambda(t) = \lambda$ . The relative number  $E_{k_i}/K$  is numerically equal to the probability of the customer staying at the state (node)  $S_i, i = \overline{1, n+1}$ , in the steady-state. In the steady-state, problem (3) turns into a linear programming problem.

#### 4. Numerical Example

Consider the company call center that handles calls on three topics,  $n = 3$ . Let the number of customers of the company be equal to  $K = 20\,000$ . Let the work of the call center be specified by the following parameters:

- the transition matrix is  $P = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0.8 & 0 & 0.1 & 0.1 & 0 \\ 0.9 & 0.03 & 0 & 0.07 & 0 \\ 0.75 & 0.1 & 0.15 & 0 & 0 \\ 0.1 & 0.45 & 0.3 & 0.15 & 0 \end{pmatrix}$ ,
- the arrival rate is  $\lambda(t) = 0.002\sin\left(\frac{2\pi t}{24} + 2\right) + 0.0083$  or  $\lambda = 0.01$ ;
- the number of servers in network nodes  $m_1 = 3, m_2 = 4, m_3 = 3, m_4 = 2$ ;
- the service rates  $\mu_1 = 40, \mu_2 = 20, \mu_3 = 15, \mu_4 = 135$ ;
- the mean waiting time  $\eta_1^{-1} = 2, \eta_2^{-1} = 2, \eta_3^{-1} = 2.5, \eta_4^{-1} = 10$ ;
- the initial placement of customers  $E_{k_0}(t) = 20\,000, E_{k_i}(t) = 0, i = \overline{1,4}$ .

Let us solve the system (10) by numerical methods and present its solution graphically in Figure 2.

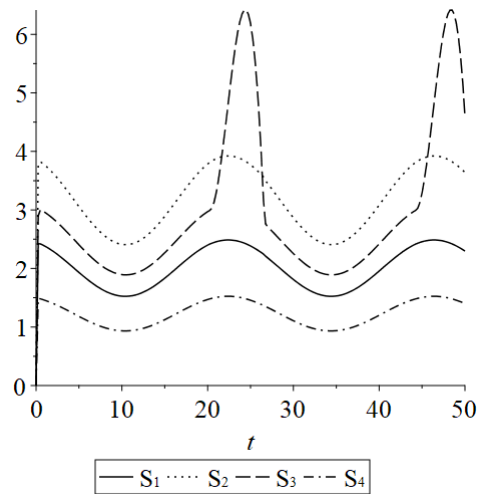
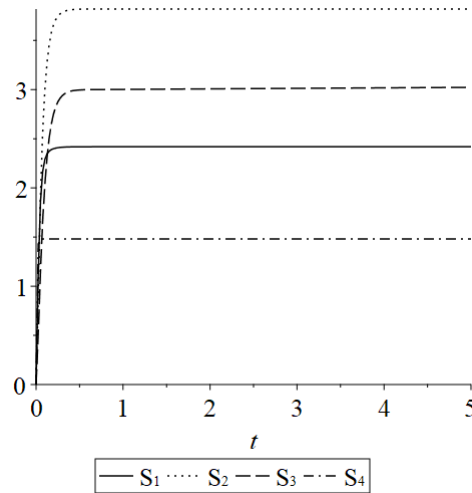


Figure 2: Dynamics of  $E_{k_i}(t), i = \overline{1,4}$ , when the arrival rate depends on the time



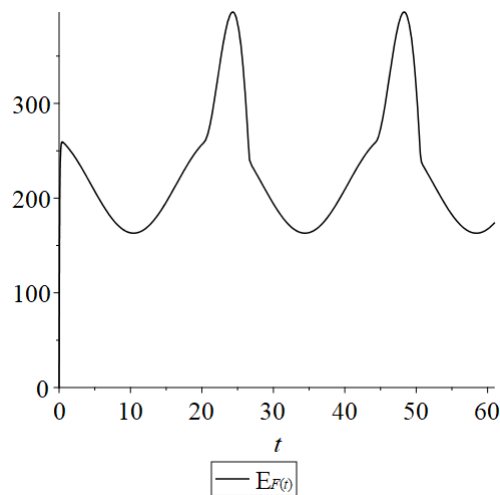
Figure 2 shows the dynamics of the expected number of customers in the network nodes  $S_i$ ,  $i = \overline{1,4}$ , when the arrival rate is the function of time:  $\lambda(t) = 0.002\sin\left(\frac{2\pi t}{24} + 2\right) + 0.0083$ . In this situation, it is advisable to vary the number of agents depending on the arrival rate. In peak situations, node  $S_3$  has a queue.

Figure 3 shows the dynamics of the expected number of customers in the queueing systems  $S_i$ ,  $i = \overline{1,4}$ , when the arrival rate is constant:  $\lambda = 0.01$ . Figure 3 demonstrates that the process quickly reaches a steady state. The largest number of customers accumulates in node  $S_2$ . The servers in the nodes of the network are idle in the mean, and in this situation, the number of agents handling calls is sufficient for the absence of queues.



**Figure 3:** Dynamics of  $E_{k_i}(t)$ ,  $i = \overline{1,4}$ , when the arrival rate is constant

Let us examine the expected monetary effect of a call center  $E_F(t)$ , which is calculated using the formula (2). Let us calculate the value (2) under the assumption that  $F_0 = 0, F_1 = F_2 = 20, F_3 = 40, F_4 = 10, L_1 = L_2 = 0,05, L_3 = 0.07, L_4 = 0.01$  in monetary units. Figure 4 shows the dynamics of  $E_F(t)$  in the cases of the non-stationary arrival flow.



**Figure 4:** Dynamics of  $E_F(t)$  when the arrival rate depends on the time

The results of calculating the expected monetary effect of a call center for different sets of servers  $m = (m_1, m_2, m_3, m_4)$  at the same time  $t = 25$  and the stationary arrival flow with  $\lambda = 0.01$  are presented in Table 1.

As is seen from the calculation results, the most optimal set is  $m = (3, 4, 3, 2)$ , in which there are no queues and the monetary efficiency is maximal.

**Table 1**The average monetary effect of a call center at time  $t=25$ 

No.	$m_1$	$m_2$	$m_3$	$m_4$	$E_F(25)$
1	3	4	3	2	251.59
2	4	4	3	2	251.54
3	3	3	3	2	229.97
4	2	4	3	2	234.67
5	3	4	2	2	207.09

Thus, the calculation results can be useful in analyzing and making decisions regarding the operation of the call center with different parameters.

## 5. Conclusion

In this paper, the network stochastic model of a call center was presented as a queueing network with impatient customers. The model was investigated in the asymptotic case of a large number of customers. This allowed us to calculate the expected number of customers at each queueing node and determine the optimal staffing list of agents to maximize call center performance. The method is applicable both in transient and steady-state. Call center performance indicators can be found using mathematical methods for calculating the nodal characteristics of queueing networks. The areas of implementation of these research results are the design of call centers, the improving their performance, and the using queueing networks as models. In further studies, it is planned to use HM-networks, this type of queueing network first proposed by M. Matalytski [19, 20]. HM-networks assume revenues from changes in network states.

## 6. References

- [1] N. Gans, G. Koole, A. Mandelbaum, Telephone call centers: tutorial, review, and research prospects, *Manufacturing & service operations management* 5(2) (2003) 79-141. DOI: 10.1287/msom.5.2.79.16071.
- [2] L. Bennington, J. Cummane, P. Conn, Customer satisfaction and call centers: an Australian study, *International journal of service industry management* 11(2) (2000) 162-173. DOI: 10.1108/09564230010323723.
- [3] R. Stolletz, *Performance Analysis and Optimization of Inbound Call Centers*, Springer, Berlin, Heidelberg, 2003. DOI:10.1007/978-3-642-55506-0.
- [4] R. Srinivasan, J. Talim, J. Wang, Performance analysis of a call center with interactive voice response units, *Top* 12 (2004) 91–110. DOI:10.1007/BF02578926.
- [5] G. Koole, A. Mandelbaum, Queueing models of call centers: an introduction, *Annals of operations research* 113 (2002) 41–59. DOI:10.1023/A:1020949626017.
- [6] S. Aguir, F. Karaesmen, O. Akşin, F. Chauvet, The impact of retrials on call center performance, *OR Spectrum* 26 (2004) 353–376. DOI:10.1007/s00291-004-0165-7.
- [7] S. Zeltyn, A. Mandelbaum, Call centers with impatient customers: many-server asymptotics of the  $M/M/n + G$  queue, *Queueing Systems* 51 (2005) 361–402. DOI:10.1007/s11134-005-3699-8.
- [8] C. Kim, O. Dudina, A. Dudin, S. Dudin, Queueing system MAP/M/N as a model of the call center with the call-back option, in: K. Al-Begain, D. Fiems, JM. Vincent (Eds.), *Proceedings of 19th international conference on Analytical and Stochastic Modeling Techniques and Applications, ASMTA 2012, Lecture Notes in Computer Science, LNCS*, Springer, Berlin, Heidelberg, 2012, pp. 1–15. DOI:10.1007/978-3-642-30782-9\_1.
- [9] S. V. Pustova, Investigation of call centers as retrial queueing systems, *Cybernetics and System Analysis* 46 (2010) 494–499. DOI:10.1007/s10559-010-9224-z.
- [10] F. Irvani, B. Balcioğlu, On priority queues with impatient customers, *Queueing Systems* 58 (2008) 239. DOI:10.1007/s11134-008-9069-6.

- [11] G. A. Medvedev, Closed queueing systems and their optimization, Proceedings of the USSR Academy of Sciences. Engineering Cybernetics 6 (1978) 199–203.
- [12] M. Matalytski, T. Rusilko, A. Pankov, Asymptotic analysis of the closed queueing structure with time-dependent service parameters and single-type messages, Journal of applied mathematics and computational mechanics 12(2) (2013) 73–80. DOI:10.17512/jamcm.2013.2.09.
- [13] T. V. Rusilko, D. Ya. Kopats, Differential equations for the moments of the state vector of a closed queueing network with impatient customers, Herald of Dagestan State University. Series 1. Natural sciences 36(2) (2021) 20–30. DOI:10.21779/2542-0321-2021-36-2-20-30.
- [14] V. I. Tikhonov, M. A. Mironov, Markov Processes, Soviet Radio, Moscow, 1977.
- [15] K. V. Gardiner, Stochastic Methods in Natural Sciences, Mir, Moscow, 1986.
- [16] M. A. Matalytski, T. V. Rusilko, Approximate Methods for Analysis of Networks with a Central Queueing System and Their Applications, GrSU, Grodno, 2003.
- [17] Y. I. Paraev, Introduction to Statistical Dynamics of Management and Filtering, Soviet Radio, Moscow, 1976.
- [18] T. V. Rusilko, The first two orders moments determination method for the state vector of the queueing network in the asymptotic case, Vesnik of Yanka Kupala State University of Grodno. Series 2 11(2) (2021) 152–161.
- [19] M. A. Matalytski, Forecasting anticipated incomes in the Markov networks with positive and negative customers, Automation and remote control 78(5) (2017) 815–825. DOI:10.1134/S0005117917050046.
- [20] M. A. Matalytski, Analysis and forecasting of expected incomes in Markov networks with unreliable servicing systems, Automation and remote control 76 (2015) 2179–2189. DOI:10.1134/S0005117915120073.