Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2024. Т. 24, вып. 2. С. 287–297
Izvestiya of Saratov University. Mathematics. Mechanics. Informatics, 2024, vol. 24, iss. 2, pp. 287–297
https://mmi.sgu.ru https://doi.org/10.18500/1816-9791-2024-24-2-287-297, EDN: KOUTKP

Article

Queueing network model of a call center with customer retrials and impatient customers

T. V. Rusilko✉, A. V. Pankov

Yanka Kupala State University of Grodno, 22 Ozheshko St., Grodno 230023, Belarus
Tatiana V. Rusilko, tatiana.rusilko@gmail.com, https://orcid.org/0000-0002-4880-0619, AuthorID: 1073719
Andrey V. Pankov, a.pankov@gmail.com, https://orcid.org/0009-0009-5708-4480, AuthorID: 1045760

Abstract. The subject of mathematical study and modelling in this paper is an inbound call center that receives calls initiated by customers. A closed exponential queueing network with customer retrials and impatient customers is used as a stochastic model of call processing. A brief review of published results on the application of queueing models in the mathematical modeling of customer service processes in call centers is discussed. The network model is described. The possible customer states, customer routing, parameters, and customer service features are given. The allocation of customers by network nodes at a fixed time fully describes the situation in the call center at that time. The state of the network model under study is represented by a continuous-time Markov chain on finite state space. The model is studied in the asymptotic case under the critical assumption of a large number of customers in the queueing network. The mathematical approach used makes it possible to use the passage to the limit from a Markov chain to a continuous-state Markov process. It is proved that the probability density function of the model state process satisfies the Fokker – Planck – Kolmogorov equation. Using the drift coefficients of the Fokker – Planck – Kolmogorov equation, a system of ordinary differential equations for calculating the expected number of customers in each network node over time can be written. The solution of this system allows for predicting the dynamics of the expected number of customers at the model nodes and regulating the parameters of the call center operation. The asymptotic technique used is applicable both in transient and steady states. The areas of implementation of research results are the design of call centers and the analysis of their workload.

Keywords: queuing network, call center, mathematical modeling, asymptotic analysis, impatient customer, retrial customer
Acknowledgements: This work was supported by the state program of scientific research of the Republic of Belarus “Convergence-2025”.
For citation: Rusilko T. V., Pankov A. V. Queueing network model of a call center with customer retrials and impatient customers. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics*, 2024, vol. 24, iss. 2, pp. 287–297. https://doi.org/10.18500/1816-9791-2024-24-2-287-297, EDN: KOUTKP

Научная статья
УДК 519.872.5

Сеть массового обслуживания с повторными вызовами и нетерпеливыми клиентами как модель колл-центра

Т. В. Русилко✉, А. В. Паньков

Гродненский государственный университет имени Янки Купалы, Беларусь, 230023, г. Гродно, ул. Ожешко, д. 22

Русилко Татьяна Владимировна, кандидат физико-математических наук, доцент кафедры фундаментальной и прикладной математики, tatiana.rusilko@gmail.com, https://orcid.org/000-0002-4880-0619, AuthorID: 1073719

Паньков Андрей Витальевич, кандидат физико-математических наук, доцент кафедры фундаментальной и прикладной математики, a.pankov@gmail.com, https://orcid.org/0009-0009-5708-4480, AuthorID: 1045760

Аннотация. Предметом математического исследования и моделирования в данной работе является колл-центр, который принимает входящие звонки, инициированные клиентами. В качестве стохастической модели процесса обслуживания звонков предлагается использовать замкнутую экспоненциальную сеть массового обслуживая с повторными вызовами и нетерпеливыми заявками. Приведен краткий обзор опубликованных работ по применению моделей массового обслуживания при математическом моделировании процессов обслуживания клиентов в колл-центрах. Описана сетевая модель, указаны возможные состояния, маршрутизация, параметры и особенности обслуживания заявок. Состояние модели полностью характеризуется распределением заявок по возможным системам массового обслуживания в заданный момент времени. Вектор, определяющий состояние сетевой модели, представляет собой цепь Маркова с непрерывным временем и конечным числом состояний. Модель исследуется в асимптотическом случае — при критическом предположении большого числа заявок в сети массового обслуживания. Используемый математический подход позволяет осуществить предельный переход от цепи Маркова к непрерывному марковскому процессу. Доказано, что плотность распределения вероятностей процесса состояния модели удовлетворяет уравнению Фоккера – Планка – Колмогорова. Используя коэффициенты сноса уравнения Фоккера – Планка – Колмогорова, можно записать систему обыкновенных дифференциальных уравнений для расчета среднего числа заявок в каждом из узлов сетевой модели с течением времени. Решение этой системы позволяет прогнозировать динамику ожидаемого количества клиентов в узлах сети и регулировать параметры работы колл-центра. Преимуществом выбранного метода исследования является возможность расчета средних характеристик модели колл-центра как в переходном, так и в стационарном режиме. Результаты исследования могут быть использованы при проектировании колл-центров и анализе их загруженности.

Ключевые слова: сеть массового обслуживания, колл-центр, математическое моделирование, асимптотический анализ, нетерпеливая заявка, повторный вызов

Благодарности: Работа выполнена в рамках государственной программы научных исследований Республики Беларусь «Конвергенция-2025».

Для цитирования: *Rusilko T. V., Pankov A. V.* Queueing network model of a call center with customer retrials and impatient customers [*Русилко Т. В., Паньков А. В.* Сеть массового обслуживания с повторными вызовами и нетерпеливыми клиентами как модель колл-центра] // Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2024. Т. 24, вып. 2. С. 287–297. https://doi.org/10.18500/1816-9791-2024-24-2-287-297, EDN: KOUTKP

Introduction

A call center is a centralized customer service department for many businesses that deals with incoming and outgoing phone calls via voice communication channels. It is referred to as a “call center” because traditionally, customer service is based on phone support as the main method of contact between customers and companies. The calls are handled by a team of advisors, otherwise known as agents. The article [1] focuses on reviewing the state of call centers research.

Telephone customer service organizations should track key performance indicators to measure the efficiency of call centers and agents. In particular, they strive to improve call handling time, call waiting time, and placing calls by agents. The quality and operational efficiency of these telephone services must be exceptional in order to meet the needs of the customers [2, 3]. Agents of a large best practice call center have to cater to thousands of phone calls per hour. The waiting time for delayed calls must not exceed a few seconds. To achieve high levels of service quality and efficiency, it is necessary to accurately describe the reality of call center operation, and to mathematically model this reality.

Mathematical models of call centers are of great value, but at the same time, each of them is somewhat limited in its ability to characterize system performance. Traditionally, queueing theory can be used to analyze call centers' efficiency and optimize their structure. Indeed, there is a natural fit between standard queueing models and telephone systems [4]. Queueing systems, particularly, Markovian ones, are widely used for call centers' modelling, analyzing their performance, planning, and managing [5–8]. Queueing systems with impatient customers [5, 8], retrial queueing systems [6, 7, 9], and many others have emerged from the need to mathematically model the behavior of telephone service customers, to find the optimal staffing for call centers in order to guarantee maximum profitability and a desired grade-of-service, measured in terms of acceptable waiting and blocking. A tandem queue is suitable for modeling call centers with interactive voice response machines [10]. Queues with general service are widely used to model call centers, e.g., a two-phase hyperexponential approximation can be applied for service time distribution [11]. In the article [12] a multi-server queueing system with a generalized phase-type service time distribution is used as a model of a call center with a call-back option.

The subject of mathematical study and modelling in this paper is an inbound call center that receives calls initiated by customers looking for information support, technical support, billing questions, reservation support, order-taking functions, and other customer service issues. First, an arriving call is routed towards the automatic call distribution switch, whose function is to distribute the inbound calls among the agents according to the customer's need. When the topic of the client's request goes beyond the scope of the current agent's specialization, the client is redirected to specialists who are competent in resolving the issue. Thus, calls are routed according to their specificity between agents of various specializations. Therefore, the organizational structure of a call center can be graphically represented by a network diagram showing the call routing between call center agents on different service issues according to their specifics.

Queueing networks are effective mathematical models for studying discrete probabilistic systems with a network-like structure. A queueing network is a collection of interdependent queuing nodes that provides processing and transfer of jobs or customers. Everyone knows that a telephone customer who receives a busy signal repeats the call until the required connection is made. As a result, the flow of calls circulating in a call center consists of two parts: the flow of primary calls and the flow of repeated calls. It is natural to take into account a flow of impatient customers who find that the residual waiting time is too long and leave the queue forever. These considerations emphasize the need to use the network of retrial queues with impatient customers as a proper modelling of customer behavior in a call center. So, the purpose of this paper is call center mathematical modeling and analysis of call processing efficiency using a closed exponential queueing network with retrial and impatient customers. An asymptotic analysis of the model is performed, which implies an approximation method of queueing network study under the critical assumption of a large but limited number of customers [13–15].

1. Model description

The focus of this paper is the queueing network model of a large call center. We set K as the total number of customers. The closed network consists of $n+1$ exponential m_i-linear queueing nodes (systems) S_i, $i = \overline{1, n+1}$, and K-linear hypothetical node S_0, which plays the role of a dependent source of arriving customers or an external environment. It should be noted that, unlike the others, the node S_{n+1} is the retrial queueing system without a waiting buffer. The node S_{n+1} has the orbit O_{n+2} which plays the role of a virtual waiting room for K customers. Each customer can be in one of the following nodes (states) at any given point in time:

- S_0 — no need to turn to the call center;
- S_{n+1} — a customer call is serviced by one of the m_{n+1} lines of the automatic call distribution switch;
- O_{n+2} — a customer call is waiting in the orbit;
- S_i — a customer call is handled by one of the m_i agents supporting the ith type service issue or is queued in the unlimited waiting buffer, $i = \overline{1, n}$.

The customer transition from the node S_0 to the node S_{n+1} with probability $p_{0(n+1)} = 1$ corresponds to a customer call to the call center. We assume that the arrival of customers from S_0 to S_{n+1} forms the Poisson process of rate $\lambda(t)k_0(t)$, the rate parameter is proportional to the number of customers in the source $k_0(t)$. Inbound call flow is non-stationary.

From the node S_{n+1}, the call is routed to the node S_i with probability $p_{(n+1)i}$, $i = \overline{1,n}$, or with probability $p_{(n+1)0}$ to the external environment S_0 in case of an erroneous customer request. After the end of call service in the node S_i, $i = \overline{1,n}$, the customer is transferred to S_0 with probability p_{i0}, in case his request is completely fulfilled or redirected to specialists on another service issue to the node S_j, $j = \overline{1,n}$, with probability p_{ij}, $i \neq j$, $i,j = \overline{1,n}$. Customers are served according to the FIFO rule. The service facility of the node S_i consists of m_i identical servers and service times are exponentially distributed with rate parameter μ_i, i.e., μ_i^{-1} is the mean service time, $i = \overline{1,n+1}$.

Taking into account the peculiarities of telephone services, it is assumed that the customer waiting time in the queue of the node S_i is limited by an exponential random variable τ_i with an expected value of η_i^{-1}, $i = \overline{1,n}$. At the same time, customers who leave the queue are called impatient. If during time τ_i the call is not answered, then it is lost and transferred to the external environment S_0 with a probability $q_{i0} = 1$, $i = \overline{1,n}$.

It is worth noting some important features of the automatic call distribution switch which is modeled by the retrial queueing node S_{n+1}. If an incoming customer call finds some of m_{n+1} servers free, he instantly occupies one of them and leaves the node after service. On the other hand, any request that finds all servers busy upon arrival is required to leave the service area but it is not lost, it is transferred to the orbit O_{n+2}. The customer in the orbit repeats his call and retries reaching the free server again after an exponentially distributed time τ_{n+2}^r with parameter γ. Thus, we are assuming that the repeated attempts follow the classical retrial policy, where the repetition times of each customer are assumed to be independent and exponentially distributed with the rate parameter γ. In addition, we suppose that the customer's time in the orbit is limited by an exponential random variable τ_{n+2}^o with an expected value of η_{n+2}^{-1}. After this time τ_{n+2}^o, the customer leaves the orbit and is transferred to the external environment S_0 with probability $q_{(n+2)0} = 1$. We also assume that inter-arrival periods, service times, waiting times, and retrial times are mutually independent.

Probabilities p_{ij} are elements of a transition matrix $P = (p_{ij})_{(n+2)\times(n+2)}$ for serviced customers, $i,j = \overline{0,n+1}$. Non-zero elements of the matrix P are $p_{0(n+1)} = 1$ and p_{ij}, $i \neq j$, $i = \overline{1,n+1}$, $j = \overline{0,n}$. The matrix P is a stochastic matrix, so $\sum\limits_{j=0}^{n+1} p_{ij} = 1$. Probabilities q_{ij} are elements of a transition matrix $Q = (q_{ij})_{(n+2)\times(n+2)}$ for impatient customers $i,j = \overline{0,n+1}$. Non-zero elements of the matrix Q are $q_{i0} = 1$, $i = \overline{1,n}$. Retrial customers are circulating between the system S_{n+1} and the orbit O_{n+2}. Impatient retrial customers transfer from O_{n+2} to S_0. We have to keep in mind the non-zero elements of transition matrices. The routing of serviced customers (solid line), impatient customers (dashed line), and retrial customers (long dash) is shown in Fig. 1.

The allocation of customers by 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. Taking into account the above-described, the state of the network model under study at time t is represented by a continuous-time Markov chain on finite state space:

$$k(t) = (k_1(t), k_2(t), ..., k_{n+1}(t), k_{n+2}(t)),$$

where $k_i(t)$ is the number of customers in the node S_i, $i = \overline{1,n+1}$, $k_{n+2}(t)$ is the number of customers in the orbit, at time t, $t \in [0, +\infty)$. Obviously, the number of customers in the external environment S_0 is $k_0(t) = K - \sum\limits_{i=1}^{n+2} k_i(t)$.

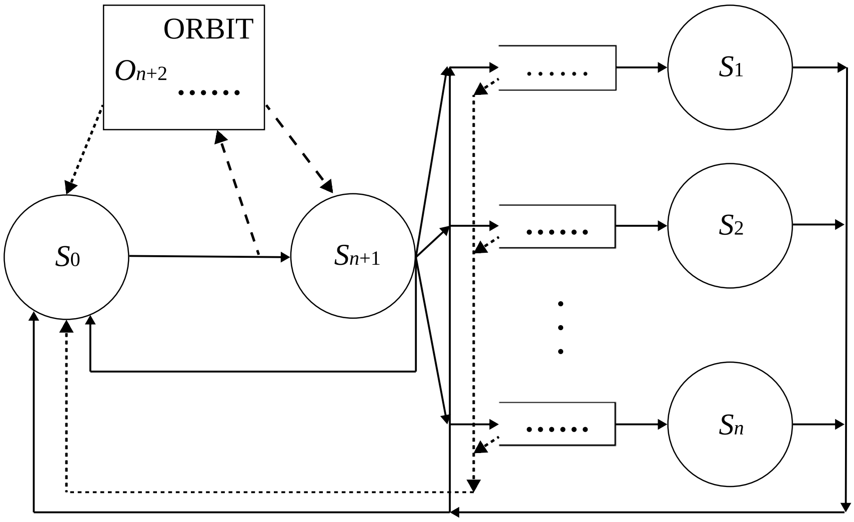

Fig. 1. The state-transition diagram

2. Asymptotic analysis of the network model

Asymptotic analysis implies an approximate method for studying a queueing network under the assumption of a large number of customers K and is based on the theory of diffusion approximation of Markov processes [16]. In this paper, the passage to the limit from a Markov chain $k(t)$ to a continuous-state Markov process $\xi(t)$ is considered. Unlike discontinuous processes, continuous processes in any small-time interval $\Delta t \to 0$ have some small change in the state $\Delta x \to 0$.

Theorem 1. *In the asymptotic case of a large number of customers K the probability density function $p(x,t)$ of the random process*

$$\xi(t) = \frac{k(t)}{K} = \left(\frac{k_1(t)}{K}, \frac{k_2(t)}{K}, \dots, \frac{k_{n+1}(t)}{K}, \frac{k_{n+2}(t)}{K}\right)$$

provided that it is differentiable with respect to t and twice continuously differentiable with respect to x_i, $i = \overline{1, n+2}$, satisfies up to $O(\varepsilon^2)$, where $\varepsilon = \frac{1}{K}$, the multidimensional Fokker – Planck – Kolmogorov equation

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

with drifts

$$A_i(x,t) = \sum_{j=1}^{n+2} \mu_j \left(p_{ji} - \delta_{ji}\right) \min(l_j, x_j) - \eta_i (x_i - l_i)\theta(x_i - l_i), \quad i = \overline{1,n}, \quad (2)$$

$$A_{n+1}(x,t) = -\mu_{n+1}\min(l_{n+1}, x_{n+1}) + \lambda(t)\left(1 - \sum_{i=1}^{n+2} x_i\right)\theta(l_{n+1} - x_{n+1}) + \gamma x_{n+2}\theta(l_{n+1} - x_{n+1}), \quad (3)$$

$$A_{n+2}(x,t) = \lambda(t)\left(1 - \sum_{i=1}^{n+2} x_i\right)\left(1 - \theta(l_{n+1} - x_{n+1})\right) - \gamma x_{n+2}\theta(l_{n+1} - x_{n+1}) - \eta_{n+2}x_{n+2}, \quad (4)$$

δ_{ij} is the Kronecker delta, $l_i = \dfrac{m_i}{K}$, $i, j = \overline{1, n+2}$, $\theta(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leqslant 0 \end{cases}$ is the Heaviside function.

Diffusion coefficients $B_{ij}(x,t)$ are not given in this paper because of their cumbersomeness and also because they are not used for further calculations.

Proof. Let a vector $k = (k_1, k_2, \ldots, k_{n+1}, k_{n+2})$ be formed by components $k_i(t) = k_i$, $i = \overline{1, n+2}$, at some time point t. In this case, we will say that the queueing network under study is in the state (k, t). Let $P(k, t)$ denote the probability that the network is in the state (k, t). Denote I_i as a $(n+2)$-vector with zero components excluding i-th, which equals to 1. Consider all possible state changes of the Markov process $k(t)$ in the short time Δt:

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

$$\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i + 1\right)\theta(m_{n+1} - k_{n+1} + 1)\Delta t + o(\Delta t),$$

that corresponds to the customer call arrival from the nodee S_0 to the nodee S_{n+1} when there are free lines of the automatic call distribution switch;

- from the state $(k - I_{n+2}, t)$ the process transfers to $(k, \Delta t + t)$ with probability

$$\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i + 1\right)(1 - \theta(m_{n+1} - k_{n+1}))\,\Delta t + o(\Delta t),$$

that corresponds to the customer transition to the orbit due to the fact that all lines of the automatic switch are busy;

- from the state $(k + I_{n+2} - I_{n+1}, t)$ the process transfers to $(k, \Delta t + t)$ with probability

$$\gamma(k_{n+2} + 1)\theta(m_{n+1} - k_{n+1} + 1)\Delta t + o(\Delta t),$$

that is, the arrival of a retrial call from the orbit to a free line of the automatic switch;

- from the state $(k + I_{n+2}, t)$ the process transfers to $(k, \Delta t + t)$ with probability

$$\eta_{n+2}(k_{n+2} + 1)\Delta t + o(\Delta t),$$

in case the time spent by an impatient customer in the orbit while trying to reach a free server of the automatic switch has expired;

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

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

that is, the customer departure from the node S_i to the node S_0, when his request is fully satisfied, $i = \overline{1, n+1}$, or the transition of an impatient customer from S_i to S_0, $i = \overline{1, n}$;

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

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

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

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

$$1 - \left(\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i\right)\theta(m_{n+1} - k_{n+1}) + \right.$$
$$+\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i\right)(1 - \theta(m_{n+1} - k_{n+1})) + \gamma k_{n+2}\theta(m_{n+1} - k_{n+1}) + \eta_{n+2}k_{n+2} +$$
$$\left. + \sum_{i=1}^{n+1}(\mu_i p_{i0}\min(m_i, k_i) + \eta_i q_{i0}(k_i - m_i)\theta(k_i - m_i)) + \sum_{i,j=1}^{n+1} \mu_i p_{ij} \min(m_i, k_i)\right)\Delta t + o(\Delta t),$$

if the model state does not change;

- from other model states transfer to $(k, t + \Delta t)$ is possible with probability $o(t)$.

Taking into account state changes listed above and using the law of total probability, the following set of equations is valid for the probability $P(k,t) = P(k(t) = k)$:

$$P(k,t+\Delta t) = \lambda(t)\left(K - \sum_{i=1}^{n+2} k_i + 1\right)\theta(m_{n+1} - k_{n+1} + 1)P(k - I_{n+1}, t)\Delta t +$$
$$+\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i + 1\right)(1 - \theta(m_{n+1} - k_{n+1}))P(k - I_{n+2}, t)\Delta t +$$
$$+\gamma(k_{n+2}+1)\theta(m_{n+1} - k_{n+1} + 1)P(k + I_{n+2} - I_{n+1}, t)\Delta t +$$
$$+\eta_{n+2}(k_{n+2}+1)P(k + I_{n+2}, t)\Delta t + \sum_{i=1}^{n+1} \mu_i p_{i0} \min(m_i, k_i + 1)P(k + I_i, t)\Delta t +$$
$$+\sum_{i=1}^{n+1} \eta_i q_{i0}(k_i + 1 - m_i)\theta(k_i + 1 - m_i)P(k + I_i, t)\Delta t +$$
$$+\sum_{i=1}^{n+1}\sum_{j=1}^{n+1} \mu_i p_{ij} \min(m_i, k_i + 1)P(k + I_i - I_j, t)\Delta t +$$
$$+\Bigg[1 - \Bigg(\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i\right)\theta(m_{n+1} - k_{n+1}) +$$
$$+\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i\right)(1 - \theta(m_{n+1} - k_{n+1})) + \gamma k_{n+2}\theta(m_{n+1} - k_{n+1}) +$$
$$+\eta_{n+2}k_{n+2} + \sum_{i=1}^{n+1} \mu_i p_{i0} \min(m_i, k_i) + \sum_{i=1}^{n+1} \eta_i q_{i0}(k_i - m_i)\theta(k_i - m_i) +$$
$$+\sum_{i,j=1}^{n+1} \mu_i p_{ij} \min(m_i, k_i)\Bigg)\Delta t\Bigg] P(k,t) + o(\Delta t).$$

Let us denote the following differences:

$$\Delta_{0(n+1)}P(k,t) = P(k - I_{n+1}, t) - P(k,t), \quad \Delta_{0(n+2)}P(k,t) = P(k - I_{n+2}, t) - P(k,t),$$
$$\Delta_{i0}P(k,t) = P(k + I_i, t) - P(k,t), \quad \Delta_{ij}P(k,t) = P(k + I_i - I_j, t) - P(k,t), \quad i,j = \overline{1, n+2}.$$

Assuming that $\Delta t \to 0$, we obtain the set of Kolmogorov difference-differential equations for state probabilities:

$$\frac{\partial P(k,t)}{\partial t} = \lambda(t)\left(K - \sum_{i=1}^{n+2} k_i\right)\theta(m_{n+1} - k_{n+1})\Delta_{0(n+1)}P(k,t) +$$
$$+\Bigg(\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i + 1\right)\theta(m_{n+1} - k_{n+1} + 1) -$$
$$-\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i\right)\theta(m_{n+1} - k_{n+1})\Bigg)P(k - I_{n+1}, t) +$$
$$+\lambda(t)\left(K - \sum_{i=1}^{n+2} k_i\right)(1 - \theta(m_{n+1} - k_{n+1}))\Delta_{0(n+2)}P(k,t) +$$
$$+\lambda(t)\Bigg[\left(K - \sum_{i=1}^{n+2} k_i + 1\right)(1 - \theta(m_{n+1} - k_{n+1})) -$$

$$
-\left(K-\sum_{i=1}^{n+2}k_i\right)\left(1-\theta(m_{n+1}-k_{n+1})\right)\Bigg]P(k-I_{n+2},t)+
$$
$$
+\gamma k_{n+2}\theta(m_{n+1}-k_{n+1})\Delta_{(n+2)(n+1)}P(k,t)+
$$
$$
+\left[\gamma(k_{n+2}+1)\theta(m_{n+1}-k_{n+1}+1)-\gamma k_{n+2}\theta(m_{n+1}-k_{n+1})\right]\times
$$
$$
\times P(k+I_{n+2}-I_{n+1},t)+\Delta_{(n+2)0}P(k,t)\eta_{n+2}k_{n+2}+\eta_{n+2}P(k+I_{n+2},t)+
$$
$$
+\sum_{i=1}^{n+1}\mu_i p_{i0}\min(m_i,k_i)\Delta_{i0}P(k,t)+\sum_{i=1}^{n+1}\left(\mu_i p_{i0}\min(m_i,k_i+1)-\mu_i p_{i0}\min(m_i,k_i)\right)P(k+I_i,t)+
$$
$$
+\sum_{i=1}^{n+1}\eta_i q_{i0}(k_i-m_i)\theta(k_i-m_i)\Delta_{i0}P(k,t)+
$$
$$
+\sum_{i=1}^{n+1}\left[\eta_i q_{i0}(k_i+1-m_i)\theta(k_i+1-m_i)-\eta_i q_{i0}(k_i-m_i)\theta(k_i-m_i)\right]\times
$$
$$
\times P(k+I_i,t)+\sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\mu_i p_{ij}\min(m_i,k_i)\Delta_{ij}P(k,t)+
$$
$$
+\sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\left[\mu_i p_{ij}\min(m_i,k_i+1)-\mu_i p_{ij}\min(m_i,k_i)\right]P(k+I_i-I_j,t). \tag{5}
$$

The resulting equation cannot be solved analytically for large n. Call centers undoubtedly handle a large number of customer calls. In this regard, we consider the important asymptotic case of a large number of customers in the model. Let us pose the problem of studying the probability distribution of the state vector $k(t)$ under the critical assumption of a large number of customers K in the queueing network.

Suppose we are interested in the properties of the relative process $\xi(t)=\frac{k(t)}{K}$ when K becomes very large. Vector $\xi(t)$ indicates the relative proportion of the company's customers who contacted the call center and how calls are distributed across model nodes at time t. In time $\Delta t\to 0$, the possible changes in process $\xi(t)$ are e_i, where $e_i=I_i\cdot\varepsilon=\frac{I_i}{K}$. Assuming that $K\to\infty$ we have $e_i\to 0$, and the vector $\xi(t)$ will be a continuous-time continuous-state Markov process. A probability density function of $\xi(t)$ is defined as

$$
p(x,t)=\lim_{\varepsilon\to 0}\frac{P(x_1\leqslant\xi_1(t)<x_1+\varepsilon,\ldots,x_{n+2}\leqslant\xi_r(t)<x_{n+2}+\varepsilon)}{\varepsilon^{n+2}}=
$$
$$
=\lim_{\varepsilon\to 0}\frac{P(Kx_1\leqslant k_1(t)<Kx_1+1,\ldots,Kx_{n+2}\leqslant k_{n+2}(t)<Kx_{n+2}+1)}{\varepsilon^{n+2}},
$$

i.e.

$$
K^{n+2}P(k,t)\to p(x,t),\quad \text{where}\quad x\in X, \tag{6}
$$
$$
X=\left\{x=(x_1,x_2,\ldots,x_{n+2}):x_i\geqslant 0,\quad i=\overline{1,n+2},\quad \sum_{i=1}^{n+2}x_i\leqslant 1\right\}.
$$

Realizing the asymptotic transition (6) for the equation (5), we obtain the following partial differential equation

$$
\frac{\partial p(x,t)}{\partial t}=\lambda(t)K\left(1-\sum_{i=1}^{n+2}x_i\right)\theta(l_{n+1}-x_{n+1})\Delta_{0(n+1)}p(x,t)+
$$
$$
+\lambda(t)\frac{\partial\left(\left(1-\sum_{i=1}^{n+2}x_i\right)\theta(l_{n+1}-x_{n+1})\right)}{\partial x_{n+1}}p(x-e_{n+1},t)+
$$

$$+\lambda(t)K\left(1-\sum_{i=1}^{n+2}x_i\right)(1-\theta(l_{n+1}-x_{n+1}))\Delta_{0(n+2)}p(x,t)+$$
$$+\lambda\frac{\partial\left(1-\sum\limits_{i=1}^{n+2}x_i\right)}{\partial x_{n+2}}(1-\theta(l_{n+1}-x_{n+1}))p(x-e_{n+2},t)+$$
$$+\gamma Kx_{n+2}\theta(l_{n+1}-x_{n+1})\Delta_{(n+2)(n+1)}p(x,t)+\gamma\theta(l_{n+1}-x_{n+1})p(x+e_{n+2}-e_{n+1},t)+$$
$$+\eta_{n+2}Kx_{n+2}\Delta_{(n+2)0}p(x,t)+\eta_{n+2}p(x+e_{n+2},t)+$$
$$+K\sum_{i=1}^{n+1}\mu_ip_{i0}\min(l_i,x_i)\Delta_{i0}p(x,t)+\sum_{i=1}^{n+1}\mu_ip_{i0}\frac{\partial\min(l_i,x_i)}{\partial x_i}p(x+e_i,t)+$$
$$+K\sum_{i=1}^{n+1}\eta_iq_{i0}(x_i-l_i)\theta(x_i-l_i)\Delta_{i0}p(x,t)+\sum_{i=1}^{n+1}\eta_iq_{i0}\frac{\partial((x_i-l_i)\theta(x_i-l_i))}{\partial x_i}p(x+e_i,t)+$$
$$+K\sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\mu_ip_{ij}\min(l_i,x_i)\Delta_{ij}p(x,t)+\sum_{i=1}^{n+1}\sum_{j=1}^{n+1}\mu_ip_{ij}\frac{\partial\min(l_i,x_i)}{\partial x_i}p(x+e_i-e_j,t). \quad (7)$$

If $p(x,t)$ is twice continuously differentiable function with respect to x, then we can use the following Taylor series:

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

Substituting the above-mentioned Taylor series into equation (7), having grouped the terms in the resulting equation, we conclude that the compact mathematical expression (1) is valid. The theorem is proved. □

The probability distribution of the vector $\xi(t)$ given by the probability density $p(x,t)$ is a complete and exhaustive characteristic of the network model state at time t. However, such an exhaustive characteristic cannot be found, since equation (1) is not explicitly solvable. From a practical point of view, it is enough for us to know what the "average value" of $\xi(t)$ is. For this reason, we confine our study to finding expected values $E_{\xi_i}(t)=E(\xi_i(t))$, $i=\overline{1,n+2}$. It is known [15, 17] that the expected value $E_{\xi_i}(t)$ is determined with an accuracy of $O(\varepsilon^2)$ from a set of ordinary differential equations:

$$\frac{dE_{\xi_i}(t)}{dt}=A_i(E_{\xi_i}(t)),\quad i=\overline{1,n+2},$$

here $A_i(\cdot)$ are drifts given by formulas (2)–(4). Then the expected number of customers in each node of the network model $E_{k_i}(t)=E(k_i(t))=E(K\xi_i(t))$ can be found from the set of equations:

$$\frac{dE_{k_i}(t)}{dt}=KA_i\left(\frac{1}{K}E_{k_i}(t)\right),\quad i=\overline{1,n+2}. \quad (8)$$

Taking into account (2)–(4), (8) and the transition matrices of the network model, the expected number of customers $E_{k_i}(t)$ in node S_i, $i=\overline{1,n+2}$, can be found by solving the following set of equations:

$$\frac{dE_{k_i}(t)}{dt}=\sum_{\substack{j=1\\j\neq i}}^{n+1}\mu_jp_{ji}\min\left(m_j,E_{k_j}(t)\right)-\mu_i\min\left(m_i,E_{k_i}(t)\right)-$$

$$-\eta_i(E_{k_i}(t)-m_i)\theta(E_{k_i}(t)-m_i), \quad i=\overline{1,n};$$
$$\frac{dE_{k_{n+1}}(t)}{dt}=-\mu_{n+1}\min(m_{n+1},E_{k_{n+1}}(t))+$$
$$+\lambda(t)\left(K-\sum_{i=1}^{n+2}E_{k_i}(t)\right)\theta(m_{n+1}-E_{k_{n+1}}(t))+\gamma E_{k_{n+2}}(t)\theta(m_{n+1}-E_{k_{n+1}}(t)); \tag{9}$$
$$\frac{dE_{k_{n+2}}(t)}{dt}=\lambda(t)\left(K-\sum_{i=1}^{n+2}E_{k_i}(t)\right)(1-\theta(m_{n+1}-E_{k_{n+1}}(t))-$$
$$-\gamma E_{k_{n+2}}(t)\theta(m_{n+1}-E_{k_{n+1}}(t))-\eta_{n+2}E_{k_{n+2}}(t).$$

3. Numerical example

Consider the queuing network model of a call center with $n=4$. Let the number of customers be equal to $K=50000$.

Let the operation of the call center be specified by the following parameters. The structure of the network is set by the following non-zero elements of the transition matrix: $p_{05}=1$, $p_{10}=0.7$, $p_{12}=p_{13}=p_{14}=0.1$, $p_{20}=0.8$, $p_{21}=p_{23}=0.05$, $p_{24}=0.1$, $p_{30}=0.75$, $p_{31}=0.07$, $p_{32}=0.15$, $p_{34}=0.03$, $p_{40}=0.9$, $p_{41}=0.03$, $p_{42}=0.04$, $p_{43}=0.03$, $p_{50}=0.15$, $p_{51}=0.1$, $p_{52}=0.15$, $p_{53}=0.2$, $p_{54}=0.4$. The number of node servers is $m_1=m_3=6$, $m_2=8$, $m_4=4$, $m_5=20$. The service rates are $\mu_1=70$, $\mu_2=40$, $\mu_3=22$, $\mu_4=60$, $\mu_5=270$. The waiting rates are $\eta_1=\eta_2=5$, $\eta_3=\eta_4=3$, $\eta_6=10$. The retrial rate is $\gamma=50$. The initial placement of customers is $E_{k_0}(0)=50000$.

Firstly, consider the case when the arrival rate is constant $\lambda=0.0117$. The inbound call flow is stationary. Let us solve the set (9) by numerical methods using Maple software. Figure 2 shows a graphical solution of (9). It is possible to predict the mean number of customers at each network node with time. Figure 2 demonstrates that the process quickly reaches a steady state. The largest number of customers accumulates in the node S_4. The servers in most nodes of the network are idle in the mean.

Secondly, consider the case when the arrival rate is not constant and it changes according to a wave-like law (seasonal process) $\lambda(t)=0.002\cos(\pi t/12)+0.0095$. Figure 3 shows the dynamics of the expected number of customers in the network nodes. In this situation, the number of customers varies depending on the arrival rate. Obviously, the number of agents needs to be adjusted.

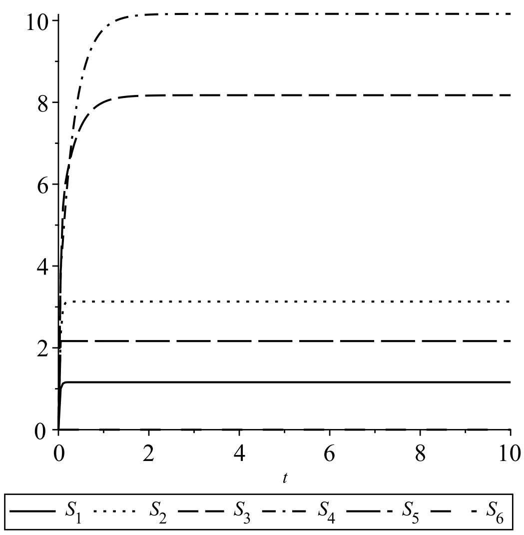

Fig. 2. Dynamics of $E_{k_i}(t)$ when the arrival rate is constant

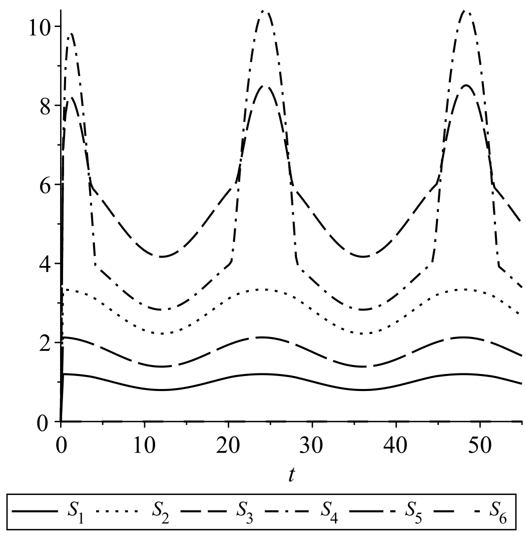

Fig. 3. Dynamics of $E_{k_i}(t)$ when the arrival rate is not constant

Conclusion

In this paper, the network stochastic model of a call center was presented as a queueing network with customer retrials and impatient customers. The model was investigated in the asymptotic case of a large number of customers. The results make it possible to predict the dynamics of the expected number of customers by model states, regulate the parameters of the call center operation, analyze the call center workload, and make decisions. They are applicable with a specified accuracy $O(\varepsilon^2)$ in both the transient and steady state, this is a fundamental advantage of the used asymptotic method.

References

1. Gans N., Koole G., Mandelbaum A. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management*, 2003, vol. 5, iss. 2, pp. 79–141. https://doi.org/10.1287/msom.5.2.79.16071
2. Stolletz R. *Performance Analysis and Optimization of Inbound Call Centers*. Heidelberg, German, Springer-Verlag Berlin, 2003. 219 p. https://doi.org/10.1007/978-3-642-55506-0
3. Srinivasan R., Talim J., Wang J. Performance analysis of a call center with interactive voice response units. *Top*, 2004, vol. 12, pp. 91–110. https://doi.org/10.1007/BF02578926
4. Koole G., Mandelbaum A. Queueing models of call centers: an introduction. *Annals of Operations Research*, 2002, vol. 113, pp. 41–59. https://doi.org/10.1023/A:1020949626017
5. Takagi H., Taguchi Yu. Analysis of a queueing model for a call center with impatient customers and after-call work. *International Journal of Pure and Applied Mathematics*, 2014, vol. 90, iss. 2, pp. 205–237. https://dx.doi.org/10.12732/ijpam.v90i2.10
6. Aguir S., Karaesmen F., Zeynep Aksin O., Chauvet F. The impact of retrials on call center performance. *OR Spectrum*, 2004, vol. 26, pp. 353–376. https://doi.org/10.1007/s00291-004-0165-7
7. Nazarov A. A., Paul S. V., Lizyura O. D. Two-way communication retrial queue with unreliable server and multiple types of outgoing calls. *Discrete and Continuous Models and Applied Computational Science*, 2020, vol. 28, iss. 1, pp. 49–61. https://doi.org/10.22363/2658-4670-2020-28-1-49-61
8. Mandelbaum A., Zeltyn S. Data-stories about (im)patient customers in tele-queues. *Queueing Systems*, 2013, vol. 75, iss. 2, pp. 115–146. https://doi.org/10.1007/s11134-013-9354-x
9. Artalejo J. R., Gomez-Corral A. *Retrial Queueing Systems. A Computational Approach*. Heidelberg, German, Springer-Verlag Berlin, 2008. 318 p. https://doi.org/10.1007/978-3-540-78725-9
10. Kim Ch., Klimenok V., Dudin A. Priority tandem queueing system with retrials and reservation of channels as a model of call center. *Computers and Industrial Engineering*, 2016, vol. 96, pp. 61–71. https://doi.org/10.1016/j.cie.2016.03.012
11. Nazarov A., Moiseev A., Moiseeva S. Mathematical model of call center in the form of multi-server queueing system. *Mathematics*, 2021, vol. 9, iss. 22, art. 2877. https://doi.org/10.3390/math9222877
12. Dudin A., Kim C., Dudina O. Multi-server queueing system with a generalized phase-type service time distribution as a model of call center with a call-back option. *Annals of Operations Research*, 2016, vol. 239, pp. 401–428. https://doi.org/10.1007/s10479-014-1626-2
13. Rusilko T. Asymptotic analysis of a closed G-network of unreliable nodes. *Journal of Applied Mathematics and Computational Mechanics*, 2022, vol. 21, iss. 2, pp. 91–102. https://10.17512/jamcm.2022.2.08
14. Medvedev G. A. Closed queueing systems and their optimization. *Proceedings of the USSR Academy of Sciences. Engineering Cybernetics*, 1975, vol. 6, pp. 65–73 (in Russian).
15. Rusilko T. V. Application of queueing network models in insurance. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics*, 2022, vol. 22, iss. 3, pp. 315–321. https://doi.org/10.18500/1816-9791-2022-22-3-315-321, EDN: ONZHCB
16. Tikhonov V. R., Mironov M. A. *Markovskie protsessy* [Markov Processes]. Moscow, Sovetskoe radio, 1977. 488 p. (in Russian).
17. Paraev Yu. I. *Vvedenie v statisticheskuyu dinamiku protsessov upravleniya i fil'tratsii* [Introduction to Statistical Dynamics of Management and Filtering]. Moscow, Sovetskoe radio, 1976. 185 p. (in Russian).

Поступила в редакцию / Received 18.05.2023
Принята к публикации / Accepted 20.07.2023
Опубликована / Published 31.05.2024