УДК 621.372(075) **DOI:10.30837/rt.2025.2.222.18**

O.J. KADATSKAYA, PhD, C.O. SABUROVA

CONTROL OF CONTACT CENTER MODEL FUNCTIONAL PARAMETERS TO AGENTS LOAD REDUCTION

Introduction

Today, a call center is an integral part of any telecommunications operator. When establishing a contact center, it is essential to choose a platform that supports workforce optimization, centralized call routing, task handling, and interaction history retention. Such a platform should contribute to achieving goals across all service areas, including self-service, agent-assisted service, and proactive customer engagement by the organization [1]. A call center achieves maximum efficiency through the use of digital telecommunications technologies, balanced operator workload distribution, and partial automation of call handling. The level of customer service and call volume handled by the same number of agents varies depending on the applied load distribution management methods. Customer inquiry statistics indicate that most users reach out to centers via alternative multimedia channels. A cloud-based call center structure allows for customer interaction through any channel or device, at any time of day. These processes are integrated into a virtual cloud network and scale rapidly.

In the future, chat will become the primary mode of customer interaction, while social media will serve as a communication channel for the entire organization. A modern contact center must be equipped with multiple customer communication channels, each capable of resolving issues upon first contact. Contact center workflows are integrated with control systems, utilize real-time customer data, interaction history, and resource planning. Operators have access to analytical customer information regardless of the interaction channel. This requires a platform integrated with customer databases, CRM systems, and other components that support the company's business processes.

Evaluation of the contact center model work

The main factors influencing customer satisfaction with a contact center are the speed of request handling and the convenience of accessing necessary information. One of the key objectives of a contact center is to maintain a minimum required number of agents without dropping calls. Each contact center evaluates performance according to its own standards.

To support further analysis and description of a mathematical model for network-based contact center management, we identify key performance indicators (KPIs) for operator efficiency. The primary metrics for assessing contact center performance include:

– Average Speed of Answer (ASA): measures the average time a customer waits before being connected;

– Abandon Rate (AR): the percentage of calls terminated by callers while waiting;

– Service Level (SL): the percentage of calls answered within a predefined time frame (typically not exceeding 3–4 %).

The center is structured so that 80 % of calls are connected within 20 seconds. In cases where an IVR system is deployed, this waiting period may extend to up to 2 minutes.

From a technical perspective, a typical contact center is a software-hardware integrated complex [2, 3]. The equipment handles inbound and outbound calls, records conversations, logs voice traffic, automatically recognizes caller IDs, and populates customer databases. These services are executed on servers running specialized software, often integrated with a CRM system in modern deployments. In a call center, integration with CRM (Customer Relationship Management) software empowers agents to manage customer data efficiently and automate workflows (fig.1). A CRM-enabled call center offers numerous advantages, including real-time access to client profiles, workflow automation, better analytics, omnichannel support, enhanced collaboration [4].

Fig. 1. Call center CRM features

Contact center servers host applications that support extended features. A cloud-based contact center structure consists of the following four components:

– Infrastructure as a Service (IaaS);
– Platform as a Service (PaaS);
– Software as a Service (SaaS);
– Application Integration (enterprise applications).

Traditional contact centers perform three core functions: data exchange, call distribution, and business application management. These are unified within a single operational environment. The CRM system is a pivotal component – it maintains statistical records, manages customer contact logs, and supports customer relationship management.

Statistical parameters affecting contact center performance are aggregated within the Call Management System (CMS), including:

– call type;
– number of calls within a defined time period;
– average queue length;
– average call duration;
– ratio of clients served by IVR versus live agents;
– time during which all lines were busy;
– average operator occupancy time;
– average number of operators active over a time period;
– average time;
– average duration between the end of one call and the start of the next.

Reports based on these statistics are generated at various points across the contact center model. Different vendors provide varying sets of reports, but these levels are common to most systems. Typically, the entry point to a contact center is a virtual extension number of the switchboard, not physically tied to specific hardware. It can be accessed via any method designated for internal extensions. Reporting includes agent performance, agent group efficiency, and queue statistics at the system entry points and trunk lines.

Contact center of operational management mathematical model

One of the core tasks in operating a contact center within a large infocommunications company is establishing a system for real-time resource distribution control, optimal use of network equipment, and efficient coordination between agents and support staff.

CRM platforms are vital. Their wide functionality allows the center to:

– keeping records of customer interactions;
– maintain interaction histories;
– generate reports swiftly.

This feature set gives CRM systems a competitive advantage [5]. While CRM solutions focus on customer relationship processes, ERP systems support the internal organizational structure. Together, they form a synergistic automation system.

A contact center consists of subsystems or functional blocks, which can later be supplemented with new components. The system consolidates all customer interactions, enables selection of optimal request handling algorithms, often relies on complex logic circuits, requiring significant development of programmer effort even to create the simplest scenarios a call center workflow.

To improve data transmission efficiency and terminal management, it is proposed to augment the CRM components with new functionalities aimed at solving tasks such as:

– collecting data on the functioning of operators' terminal, customer profiles, and operator profiles;

– remote resolution of emerging tasks to reduce customer servicing requests times due to the current significant increase in request volume;

– collecting data on the functioning of operators' terminal, customer profiles, and operator profiles;

– remote resolution of tasks, especially under high request volumes;

– centralized updates from the server, considering multiple contextual parameters;

– real-time transmission of terminal parameters for bandwidth optimization and call center channels overload prevention.

Operator terminal management involves their rapid, centralized, remote connection to update real-time agent occupancy data based on call profiles. A vital tool for system monitoring and operational diagnostics is the collection of statistical data and the implementation of user feedback mechanisms.

Different contact centers evaluate their performance based on the specific criteria relevant to their operational context. Nevertheless, during peak hours, performance typically degrades, and maintaining adequate service levels becomes a key performance metric. Incoming call traffic is uneven throughout the day, leading to peak load periods and, consequently, critical agent shortages – an issue common across various centers.

Overloads in a contact center result in queue length increases and call losses during processing. Therefore, it is essential to manage contact center operations proactively to prevent service degradation. Each operator group may face queues of several hundred calls, but such queue lengths should never lead to excessive customer waiting times. Call redistribution among operators is often handled manually, which contributes to queue delays. A more effective approach is a self-adjusting system that dynamically selects the optimal call redistribution algorithm based on total service time estimation.

Let us consider a mathematical model for managing a call center network that includes operator terminals and agent groups. During peak load conditions, the management system receives real-time information and compares the workload of multiple agents or operator groups based on this parameter. It then routes the call to the one with the shortest handling time. It is important to examine the service level, maximum response delays, and the call profile. The key principle of queue organization: maximize the number of processed calls with minimal staff engagement – without compromising service quality or overloading personnel.

In the model, call processing points are grouped into separate states and positioned at designated locations within the center’s hardware-software infrastructure.

The key factors to be considered in the model include:

– the number of operators involved in handling calls during the given time period;

– the volume of statistical data received from the source;

– the intensity of incoming data, which are treated as random variables.

Waiting time can be determined both at the level of each individual call and at the level of a segmented operator group.

To characterize the quality of service, the system's performance indicators include:

– the average number of requests in the system;

– the average number of requests in the queue;

– the average time a request spends on the system;

– the average waiting time before service begins.

To describe the operation of the service system, it is necessary to consider:

– the average number of devices or channels occupied by request processing;

– the utilization factor of servicing devices;

– the idle time factor of servicing devices.

Let there be m operator devices in the network, each associated with a set of parameters B defining call profiles. All devices are grouped into K operator groups.

We define:

– the assignment of parameter i to a block, where i = 1,...,k;

– the assignment of parameter j to a device, where j = 1,...,m.

The resulting call profile state vector and operator profiles are transmitted to a central server. The polling cycle time for each parameter of the j-th reference terminal equipment (TE) depends on the number of parameters involved. When the sever interacts with either a single j-th device or all m devices, and data flows randomly between them.

A mathematical model is used to describe such a hypothetical data stream. To describe a network model with such properties is a stochastic Bernoulli flow [6]. Collection of information about the parameters of control objects carried out for m control objects (CO). Matrix $B = [nxm]$ characterizes the values of all measured parameters of all control objects, i.e. for all agents, including agent profile, call profile. Each parameter $b_{ij}=m_j x n_i$, the number n_i-number of parameters in the i-th CO of the operator's workplace n, the number m_j - number of parameters in the m-th CO operator's workplace information collection systems for the control node of the network monitoring system.

Each i-th control object is characterized by a set of $B_i=\{b_{ij}\}$ parameters, each element of which is a random variable. The distribution function of each parameter $b_{ij,}$ known. The information collection system (ICS) polls each i-th object cyclically. Each request is processed in a random processing time to(i) measurement time of one object parameters, transferred to the MS and evaluated. Each request has a really fixed volume and connects resources for a certain time equal to the average value of the busy request. The data received on demand is transmitted over the network at a certain time, which consists of 2 parts, namely, the time for transmitting a constant value for this process and the time for transmitting changed data. In the general case, the delay of messages in the network [7] is

$$Ti = Twaiting + Tforwarding + Tprocessing \quad (1)$$

Then the characteristic of the delay time of messages in the network is the sum of the time of transmission of the message through the channels of the network Tt, the processing time in the switching nodes Tp and the waiting time in the queue Tq (fig. 2).

The time for which this data is processed consists of the processing time of constant and changed parameter values. Then the total random time for obtaining the value of the b_{ij} parameter from the moment the request received is

$$tpar(i,j) = to(i) + ttr(i,j) + tpr(i,j), \quad (2)$$

where to(i) – request time; ttr(i,j) – request transmission time tpr(i,j) – request processing time.

The average value and variance of the time to obtain the value of the parameter b_{ij} can be determined as follows:

$$M(tr) = M(to) + M(ttr) + M(tpr)\ , \quad (3)$$

$$(tr) = D(to) + D(ttr) + D(tpr)\ . \quad (4)$$

Messages from each i-th device of the operator arrive at random time with the distribution function [8]:

$$A_{ij}(b_{ij}) = F\{b_{ij} \leq t\} \quad (5)$$

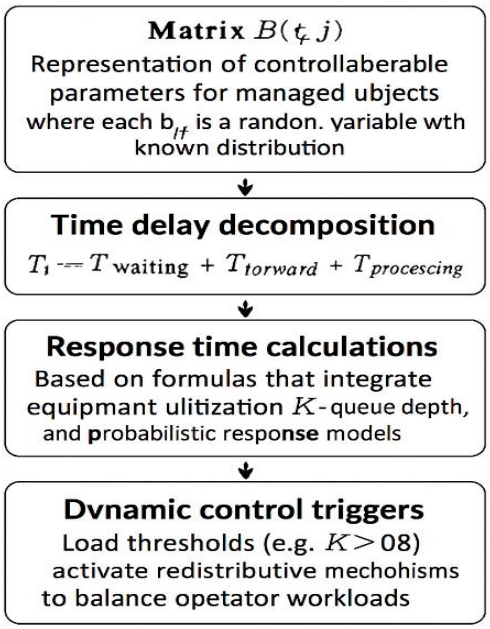

Fig. 2. Flowchart of proposed mathematical model operational control

For the parameters b_{ij} of the control objects, the exponential distribution function is characteristic $A_{ij}(b_{ij}) = 1 - exp(-qt)$. Probability of obtaining a parameter during the poll

$$P(b_{ij}) = 1 - exp[to(i) + t1 + t2] . \quad (6)$$

In the model under consideration, a random data stream comes from the server of the control node to all m devices of the call center. For a mathematical description of the data flow in the call center network configuration using the description of the stochastic Bernoulli flow, the probability of receipt in the time interval Δt of the number of responses y to send requests will be:

$$P_y(td) = C_B^y K^y (1 - K)^{B-y}, \quad (7)$$

where C_B^y – the number of combinations from the number of answers received according to the total number of parameters B: $C_B^y = B!/(y!(B - y))$ [9].

Equipment utilization factor K= tco / td , time tco - time during which the equipment was utilized for processing requests, and time td – the total time during which the equipment was available for operation (the duration of the device's operational cycle within the network). The model is a queuing system and all incoming responses that are not processed by the server are buffered [9]. Using the intensity of responses from devices in the network, we determine the average data processing time in the queuing system for the network, i.e. average processing time for device parameters of m control objects (CO):

$$rpr(i) = n \cdot to + \Sigma_{i=1}^{n}[t1(i) + t2(i)] \quad (8)$$

The total time to receive the value of all parameters b_{ij} for m=1,…,j and n =1,…,i or the time to receive a response

$$tr = (t \cdot rpr/K) \Sigma_{j=1}^{B}[j \cdot C_B^y \cdot K^y (1 - K)^{B-y}] \quad (9)$$

Average load on the host server for one polling cycle

$$Lavr = \Sigma_{j=1}^{B} P_j(td) \cdot j \quad (10)$$

Knowing the network performance parameters, it is possible to determine the time for generating a general server response based on the results of polling all support of devices.

$$K^y(1-K)^{B-y} = T \cdot t_{ob} / \Sigma_{j=1}^{B} j \cdot C_B^y \quad (11)$$

where t_{ob} – the time it takes to process the response; T – total time observation TE.

To verify the performance of the proposed model and visually demonstrate how total service time is formed during each polling cycle, we present a computational example for two matrices B, depending on the equipment utilization coefficient K, for a call center network with ten operators (fig. 3).

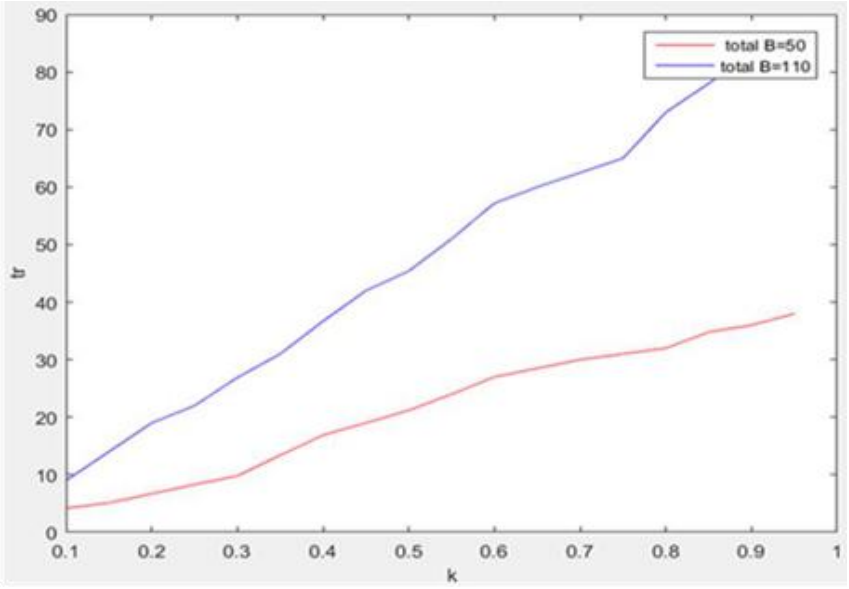

Fig. 3. The total time for obtaining the value of controlled parameters for the control model, depending on the equipment utilization factor

From the figure, it is evident that the total time required to retrieve the values of monitored parameters in the control model increases uniformly as the equipment utilization coefficient K grows. Starting from approximately 80–85 % network load, the growth in response time becomes noticeably more pronounced. This leads to an increase in queue size and necessitates redistribution of operator workloads.

Redistributing agent workload in an overloaded call center isn't merely organizational adjustment – it's a systems-level engineering solution and implementation approaches include:

– redirecting incoming calls to less-loaded agent groups or shifts based on real-time traffic and resource availability.

– reassigning calls on the fly to agents with shorter projected response times.

– shifting agents between tasks (inbound/outbound) or increasing active headcount during peak intervals, using traffic statistics like TD (Traffic Density) and Lavr (Average Waiting Time).

– rerouting calls to backup agents or voice bots when wait time or queue depth exceeds predefined thresholds.

– allocating resources ahead of time using proposed forecast models, based on expected system load K(t).

Conclusions

A contact center is a composite of subsystems that can be configured into a desired operational model. The system is designed to be supplemented with new functions or components based on statistical analysis, the performance of the call redistribution control system, and queue management models. Effective queue length control is achieved through strategic resource planning and the implementation of efficient call handling algorithms, especially under overload conditions.

The call center operator (or group of operators) represents the most resource-intensive element, which makes reducing call handling time via load balancing – through the control system – a critically important task within the model. To obtain the full set of controllable parameters, terminal equipment (TE) at the operator's workstation is used. From the moment a call is received, the management system collects the maximum amount of relevant data and interfaces with all connected information systems. This enables real-time operator workload monitoring and responsive corrective actions. Predictive load balancing use response prediction models and expected load to allocate capacity before overload uses curs.

By incorporating models and tical model for calculating the core probabilistic-temporal performance metrics – supplemented by numerical examples – the proposed framework can extend the functionality of the call processing module (CPM) without complex configurations. Ultimately, this leads to the development of a key performance indicator (KPI) system for the contact center.

References:

1. Call Center Representative Job Description: Top Duties and Qualifications 6, 2022.
2. TechNet Magazine: System Center Operations Manager 2012: Ease of expanding monitoring capabilities. URL: http://technet.microsoft.com (accessed 03 May 2021).
3. How to write a Call Centre Representative Job Description: Top Duties and Qualifications. 2025. 16 p.
4. CRM Architecture https://www.scribd.com/document/225621058/CRM-Architecture
5. CRM Design That Improves Customer Relationships. https://www.eleken.co/blog-posts/how-to-design-a-crm-system-all-you-need-to-know-about-custom-crm
6. H. Altoum, A. Ettaieb, H. Rguigui, Generalized Bernoulli–Wick differential equation, Infin. Dimens. Anal. Quantum Probab. and Relat. Top., 24, Issue 01, 2021. DOI: https://doi.org/10.1142/S0219025721500089
7. Wallace R. et al. Models of Network Delay. In: Einbeck, J., Maeng, H., Ogundimu, E., Perrakis, K. (eds) Developments in Statistical Modelling. IWSM 2024. Contributions to Statistics. Springer, Cham. https://doi.org/10.1007/978-3-031-65723-8_36.
8. Analytic Methods in Applied Probability . EditorsYu.M. Suhov .Providence, RI American Mathematical Society Pages25-36. ISBN 0-8218-3306-5. 2002 .American Mathematical Society Translations, Series 2. Vol. 207
9. S.W. Fuhrmann A. Note on the M/G/1 Queue with Server Vacations12.1984. URL: https://doi.org/10.1287/opre.32.6.1368.

Received 11.06.2025

Відомості про авторів:

Кадацька Ольга Йосипівна – канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри інфокомунікаційної інженерії ім. В.В. Поповського; Україна; e-mail: olga.kadatska@nure.ua; ORCID: https://orcid.org/0000-0002-5331-4324

Сабурова Світлана Олександрівна – Харківський національний університет радіоелектроніки, доцент кафедри інфокомунікаційної інженерії ім. В.В. Поповського; Україна; e-mail: svitlana.saburova@nure.ua; ORCID: https://orcid.org/0000-0003-2214-2440