DOI: http://doi.org/10.25728/cs.2024.4.2

INVESTIGATION OF TANDEM QUEUING SYSTEMS USING MACHINE LEARNING METHODS

V. M. Vishnevsky*, A. A. Larionov**, A. A. Mukhtarov***, and A. M. Sokolov****

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

*✉ vishn@inbox.ru, **✉ larioandr@gmail.com,
✉ mukhtarov.amir.a@gmail.com, *✉ aleksandr.sokolov@phystech.edu

Abstract. This paper considers tandem queuing systems with limited buffer sizes in each phase. The system handles an incoming correlated MAP flow and the service time obeys a PH-distribution. Models of such systems and methods for their investigation are briefly reviewed from the historical perspective. According to the review, the problem statement presented below, the methods proposed for solving this problem, and the corresponding results are novel. An accurate algorithm for calculating the performance characteristics of low-dimensional tandem queuing systems is described, including an estimate of the algorithm's complexity. An approach using both machine learning and simulation modeling is suggested for the investigation of high-dimensional tandem queuing systems. Numerical analysis results are provided to show the effectiveness of machine learning methods for estimating the performance of tandem queuing systems.

Keywords: tandem queuing system, analytical model, simulation modeling, machine learning.

INTRODUCTION

Tandem queuing systems are commonly used to model and optimize the performance of various complex systems, such as technical, economic, industrial, transportation, medical, military, and others [1–4]. Of considerable interest are also the models of tandem queuing systems that adequately describe the operation of modern broadband wireless networks with linear topology [5, 6].

Since the late 1960s [7, 8], tandem queuing systems have been intensively studied up to the present time [9–11]. Initially, analytical results were obtained for two-phase tandem queuing systems. Many authors created methods for studying systems with blocking, incoming Poisson and correlated (Batch Markovian Arrival Process, BMAP) flows, and different distribution functions of the service time; for example, see [12–16]. More detailed descriptions of the works on two-phase systems can be found in the review [17] and the book [18].

However, tandem queuing systems of high dimensionality (with two or more phases) have the greatest practical importance. An analytical solution for these systems has been found only for specific networks that meet the conditions of the BCMP (Baskett–Chandy–Muntz–Palacios) theorem [19, 20] and have a multiplicative probability distribution. When the phases in a tandem network are M/M/1 queuing systems, one can easily calculate the system's steady-state performance characteristics, including the end-to-end delay as an important parameter. The end-to-end delay is the time of transmitting a packet from the first phase to the last.

For other types of high-dimensional tandem networks, finding analytical solutions is practically impossible. Therefore, approximate methods are widely used to study them. A common method for estimating performance characteristics involves dividing a network into separate parts and analyzing their interactions. This analysis can then be used to investigate the performance of the entire system (see [21–24]). In recent years, machine learning methods have been effectively used to study tandem queuing systems [9, 25–27]. In particular, the performance characteristics of tandem queuing systems with an incoming Poisson flow, the exponential distribution of the service time, and multilinear phases $(M/M/S/\infty)$ were analyzed in [9] using artificial neural networks.

This paper presents a new algorithm for accurately calculating the steady-state performance characteristics of tandem networks with an incoming correlated MAP-flow and the phase-type (PH)-distribution of the service time in the system phases. We estimate the complexity of this algorithm and describe its advantages and numerical analysis limitations concerning the number of system phases. A combination of machine learning and simulation modeling methods is used to study large tandem queuing systems. This approach enables rapid calculation of the characteristics of high-dimensional tandems, facilitating the design of complex practical systems [9, 28, 29].

1. PROBLEM STATEMENT

In this paper, we examine a tandem queuing system with $N \geq 2$ phases and a limited buffer size M_i $(i = 1,..., N)$ of each phase. Packets income the system in a MAP-flow governed by a control process $\nu_t (t \geq 0)$. This process is a Markov chain with a state space $\{0,1,..., W\}$ and matrices $\mathbf{D}_0$ and $\mathbf{D}_1$. The matrix $\mathbf{D}_1$ describes the changes in the controlling Markov chain when a packet is generated (observable transitions); the matrix $\mathbf{D}_0$, the changes without packet generation (unobservable transitions). The matrix $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_1$ is an infinitesimal generator of the Markov chain ν_t. The packet arrival rate, denoted by λ, is the product $\lambda = \bar{\boldsymbol{\pi}}\mathbf{D}_1\bar{\mathbf{1}}$, where $\bar{\boldsymbol{\pi}}$ is the steady-state distribution vector of the process ν_t, representing the unique solution of the system of algebraic equations $\bar{\boldsymbol{\pi}}\mathbf{D} = \bar{\mathbf{0}}$, $\bar{\boldsymbol{\pi}}\bar{\mathbf{1}} = 1$. Throughout this paper, $\bar{\mathbf{1}}$ and $\bar{\mathbf{0}}$ stands for a column vector composed of ones and a row vector composed of zeros, respectively (of appropriate dimensions).

The packet service time at the *i*th phase obeys a PH-distribution with an irreducible representation $(\mathbf{S}_i, \bar{\boldsymbol{\tau}}_i)$, $i = 1,..., N$. Here, $\mathbf{S}_i$ is a square matrix of order V_i (the number of process states) and $\bar{\boldsymbol{\tau}}_i$ is a vector defining the probabilities of the initial process state. Thus, the service process at the *i*th phase is controlled by a Markov chain with a state space $\{1,..., V_i, V_i + 1\}$, where $V_i + 1$ is an absorbing state. More detailed information about the PH-distribution and MAP-flow can be found, e.g., in [18]. If a packet arrives when a buffer is full, it will be immediately discarded and considered lost, without being serviced.

In what follows, we analyze two versions of tandem queuing systems: the system with cross-traffic (along with the outgoing flow Z_i, the *i*th phase receives an extra MAP-flow X_i) and the system without cross-traffic (external traffic arrives only at the first phase; see Fig. 1.

The problem is to estimate the steady-state performance characteristics of the described tandem queuing systems, including the end-to-end delay and the probability of packet loss.

2. A PRECISE ALGORITHM FOR CALCULATING THE PERFORMANCE CHARACTERISTICS OF THE TANDEM QUEUING SYSTEM WITH AN INCOMING MAP-FLOW, PH-DISTRIBUTION OF SERVICE TIME, AND LIMITED BUFFER SIZES

Let us examine some properties and characteristics of the MAP/PH/1/*M* system. They can be utilized to develop a precise algorithm for computing the tandem queuing system. The key property of the MAP/PH/1/*M* system is closedness on the set of MAP-flows according to the following theorems [18].

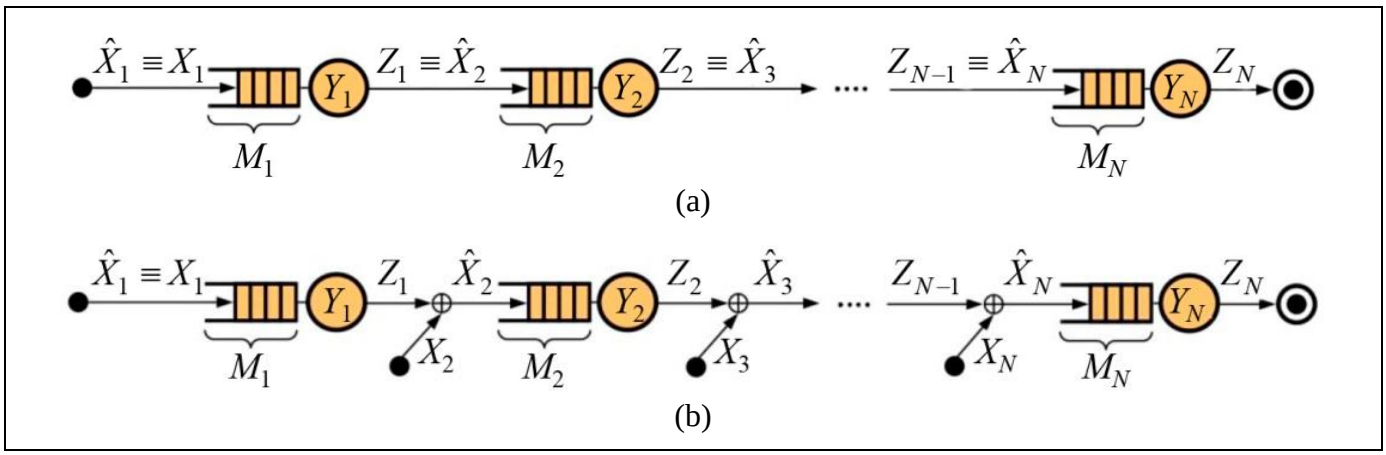

Fig. 1. Packet flows in the tandem queuing system: (a) with cross-traffic and (b) without cross-traffic.

Theorem 1. *The flow of serviced packets in the* MAP/PH/1/*M system, where the MAP-flow is given by X:* $MAP(\mathbf{D}_0, \mathbf{D}_1)$ *and the service time obeys the phase-type distribution Y:* $PH(\mathbf{S}, \bar{\boldsymbol{\tau}})$, *is an MAP-flow* $Z \sim MAP(\mathbf{D}_0', \mathbf{D}_1')$ *characterized by the matrices*

$$\mathbf{D}_0' = \begin{bmatrix} \mathbf{D}_0 \otimes \mathbf{I}_V & \mathbf{D}_1 \otimes (\bar{\boldsymbol{\tau}} \otimes \bar{\mathbf{1}}_V) & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{D}_0 \otimes \mathbf{S} & \mathbf{D}_1 \otimes \mathbf{I}_V & \cdots & 0 & 0 \\ 0 & 0 & \mathbf{D}_0 \otimes \mathbf{S} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{D}_0 \otimes \mathbf{S} & \mathbf{D}_1 \otimes \mathbf{I}_V \\ 0 & 0 & 0 & \cdots & 0 & (\mathbf{D}_0 + \mathbf{D}_1) \otimes \mathbf{S} \end{bmatrix},$$

$$\mathbf{D}_1' = \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 \\ \mathbf{I}_W \otimes \mathbf{C}_t & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & 0 \\ 0 & \cdots & 0 & \mathbf{I}_W \otimes \mathbf{C}_t & 0 \end{bmatrix},$$

where $\mathbf{C}_t = (-\mathbf{S}\bar{\mathbf{1}}_V) \otimes \bar{\boldsymbol{\tau}}$ *and* $\mathbf{I}_V$ *and* $\mathbf{I}_W$ *are identity matrices of order* V *and* W, *respectively.*

Theorem 2. *The superposition of MAP-flows* X_1 *and* X_2, $X_i \sim MAP(\mathbf{D}_0^{(i)}, \mathbf{D}_1^{(i)})$, $i = 1, 2$, *is a MAP-flow*

$$X = X_1 \oplus X_2 \sim MAP(\mathbf{D}_0^{(1)} \oplus \mathbf{D}_0^{(2)}, \mathbf{D}_1^{(1)} \oplus \mathbf{D}_1^{(2)}),$$

where $\oplus$ *denotes the Kronecker sum. If the flows* X_1 *and* X_2 *have orders* W_1 *and* W_2, *respectively, then the order of the total flow* X *is* $W = W_1 W_2$.

Let Z_i be the outgoing flow of the *i*th phase of the tandem queuing system and $\hat{X}_i$ be the total incoming flow of the *i*th phase (Fig. 1). According to Theorems 1 and 2, the flows $\hat{X}_i$ and Z_i are MAP-flows. Therefore, the *i*th phase can be described by a $\mathrm{MAP}_i / \mathrm{PH}_i / 1 / M_i$ queuing system with the packet arrival rate λ_i. For this system, the key performance characteristics are calculated by well-known formulas [18], including the average queue length m_i, the probability of packet loss $P_L^{(i)}$, the average end-to-end delay T_i of a packet at the *i*th phase, and others. By calculating these characteristics, we can determine the desired parameters for the probability of packet loss in the tandem queuing system,

$$P_L = 1 - \prod_{i=1}^{N} (1 - P_L^{(i)}),$$

and the end-to-end delay

$$T = \sum_{i=1}^{N} T_i = \sum_{i=1}^{N} \frac{m_1^{(i)}}{(1 - P_L^{(i)})\lambda_i}.$$

The next subsection presents a formal algorithm for calculating the steady-state performance characteristics of the tandem queuing system.

2.1. An Analytical Algorithm for Calculating the Steady-State Performance Characteristics of the Tandem Queuing System

Step 1. Set $i := 1$.

Step 2. If $i = 1$, set $\hat{X}_i = X_1$. In the case $i > 1$, calculate $\hat{X}_i$: $\hat{X}_i = Z_{i-1}$ if the system has no cross-traffic and $\hat{X}_i = Z_{i-1} \oplus X_i$ otherwise. Denote by $\hat{\mathbf{D}}_{i,0}$ and $\hat{\mathbf{D}}_{i,1}$ the matrices of the flow $\hat{X}_i$, i.e., $\hat{X}_i = \mathrm{MAP}(\hat{\mathbf{D}}_{i,0}, \hat{\mathbf{D}}_{i,1})$.

Step 3. Using Theorem 1, calculate the matrices $\mathbf{D}_{i,0}'$, $\mathbf{D}_{i,1}'$ of the MAP-flow $Z_i = \mathcal{D}(\hat{X}_i, Y_i, M_i)$.

Step 4. For the outgoing MAP-flow Z_i, determine the steady-state distribution $\bar{\boldsymbol{\theta}}^{(i)}$ using the system of linear algebraic equations

$$\begin{cases} \bar{\boldsymbol{\theta}}^{(i)} (\mathbf{D}_{i,0}' + \mathbf{D}_{i,1}') & = 0 \\ \bar{\boldsymbol{\theta}}^{(i)} \bar{\mathbf{1}} & = 1. \end{cases}$$

Step 5. Calculate the average number of packets in the queue at the *i* th phase:

$$m_1^{(i)} = \sum_{k=0}^{M_i + 1} k \sum_{j=1}^{V_i \hat{W}_i} \theta_{kV_i\hat{W}_i + j}^{(i)},$$

where $V_i = |Y_i|$ is the order of the PH-distribution Y_i and $\hat{W}_i = |\hat{X}_i|$ is the order of the MAP-flow $\hat{X}_i$.

Step 6. Determine the steady-state probabilities $\bar{\boldsymbol{\pi}}^{(i)}$ of the incoming flow $\hat{X}_i$. If the system has no cross-traffic and $i > 1$, set $\bar{\boldsymbol{\pi}}^{(i)} \equiv \bar{\boldsymbol{\theta}}^{(i-1)}$. Otherwise,

find $\bar{\boldsymbol{\pi}}^{(i)}$ by solving the system of linear algebraic equations

$$\begin{cases} \bar{\boldsymbol{\pi}}^{(i)}(\hat{\mathbf{D}}_{i,0}+\hat{\mathbf{D}}_{i,1}) &= 0 \\ \bar{\boldsymbol{\pi}}^{(i)}\bar{\mathbf{1}} &= 1. \end{cases}$$

Step 7. Using the steady-state probabilities $\bar{\boldsymbol{\pi}}^{(i)}$ of the incoming MAP-flow $\hat{X}_i$ obtained at the previous step, calculate the arrival rate of packets at the *i*th phase:

$$\lambda_i = \bar{\boldsymbol{\pi}}^{(i)}\hat{\mathbf{D}}_{i,1}\bar{\mathbf{1}}.$$

Step 8. Determine the steady-state distribution of the incoming MAP-flow with M_i+1 packets in the system (i.e., when the system buffer is full):

$$\bar{\boldsymbol{\psi}}^{(i)} = \left(\sum_{j=1}^{V_i} \{\bar{\boldsymbol{\theta}}^{(i)}_{M_i+1}\}_j, \ldots, \sum_{j=1}^{V_i} \{\bar{\boldsymbol{\theta}}^{(i)}_{M_i+1}\}_{(\hat{W}_i-1)V_i+j} \right).$$

Here, the vector $\bar{\boldsymbol{\theta}}^{(i)}_{M_i+1}$ is the part of the vector $\bar{\boldsymbol{\theta}}^{(i)}$ relating to the system states when there are M_i+1 packets in the system.

Step 9. Calculate the probability of packet loss due to the buffer overflow of the *i*th phase:

$$P_L^{(i)} = \bar{\boldsymbol{\psi}}^{(i)}\frac{\hat{\mathbf{D}}_{i,0}}{\lambda_i}\bar{\mathbf{1}}.$$

Step 10. Calculate the average delay at the i th phase:

$$T_i = \frac{m_1^{(i)}}{(1-P_L^{(i)})\lambda_i}.$$

Step 11. If $i < N$, assign $i := i+1$ and return to Step 2. Otherwise, proceed to Step 12.

Step 12. Calculate the probability of packet loss $P_L = 1-\prod_{i=1}^{N}(1-P_L^{(i)})$.

Step 13. Calculate the end-to-end delay (total delay) $T = \sum_{i=1}^{N} T_i$ of the tandem queuing system.

2.2. Estimating the Complexity of the Analytical Algorithm

The proposed scheme is computationally simple. At each step, block matrices are built for the outgoing MAP-flow using Kronecker product operations. In the case of cross-traffic in the system, matrices for the incoming MAP-flow are also constructed using the Kronecker sum. Next, one or two systems of linear algebraic equations are solved to determine the steady-state probabilities of the incoming and outgoing MAP-flow (the cases of cross-traffic and no cross-traffic, respectively). Finally, to calculate the probability of packet loss, the average system size, and the end-to-end delay, these steady-state distributions are multiplied by the flow matrices using several operations. The main disadvantage of this computation scheme is its extremely high computational complexity.

Proposition 1. *Assume that the incoming MAP-flows and the PH-distributions have orders W and V, respectively, the buffer size at each phase is M, and the system contains N phases. Then the complexity of the iterative scheme for calculating the performance characteristics of the tandem queuing system is estimated as follows:*

– $O((MVW)^{3N})$ *if the system has cross-traffic;*

– $O(W^3(MV)^{3N})$ *otherwise.*

Proof. Consider the i th iteration of the algorithm, where $i \leqslant N$, representing the computation of the performance characteristics of the i th phase of the system. Note that for $i>1$, the order of the outgoing MAP-flow from the preceding $(i-1)$th phase is $(M+2)V\hat{W}_i$, where $\hat{W}_i$ denotes the order of the incoming flow at the *i*th phase. If there is cross-traffic in the system, we have $U_i = ((M+2)VW)^i$; otherwise, $U_i = W((M+2)V)^i$.

The complexity of the iteration depends on Steps 4 and 6 (solving systems of linear algebraic equations). The system matrix at Step 4 (the outgoing flow generator) has a higher order than that at Step 6 (the incoming flow generator). Assuming that a Gaussian-like algorithm is used to solve the system, Step 4 requires $O(U_i^3)$ operations. The remaining steps have lower complexity: $O(1)$ for Steps 1, 10, and 11; $O(U_{i-1}^2W^2)$ for Step 2; $O(U_i^2)$ for Step 3; $O(VW+M)$ for Step 5; $O(U_i^2)$, for Steps 7 and 9; $O(VM)$. for Step 8. The complexity of Steps 12 and 13 is $O(N)$. ♦

Hence, if there is cross-traffic in the system, we obtain the algorithmic complexity

$$O((VWM)^3)+O((VWM)^6)+\ldots+O((VWM)^{3N}) +O(N) = O(VWM)^{3N};$$

in the case of no cross-traffic,

$$O(W^3(VM)^3)+O(W^3(VM)^6)+\ldots+O(W^3(VM)^{3N}) +O(N) = O(W^3(VM)^{3N}).$$

Thus, it is hard to find a solution using the algorithm even for relatively small *N*, *V*, and *W*. Table 1 provides the orders of the outgoing MAP-flows calculated under different system parameters. According to this table, a precise solution is feasible only for $N<5$

Table 1

Orders of the outgoing MAP-flows depending on the orders of the PH-distribution (*V*), MAP-flow of arrivals (*W*), and buffer size (*M*)

System parameters			Phase number				
W	*V*	*M*	1	2	3	4	5
Without cross-traffic							
1	1	1	3	9	27	81	243
1	1	3	5	25	125	625	3 125
2	2	2	16	128	1 024	8 192	65 536
3	1	3	15	75	375	1 875	9 375
1	3	3	15	225	3 375	50 625	759 375
3	3	3	45	675	10 125	151 875	2 278 125
With cross-traffic							
1	1	1	3	9	27	81	243
1	1	3	5	25	125	625	3 125
2	2	2	16	256	4 096	65 536	1 048 576
3	1	3	15	225	3 375	50 625	759 375
1	3	3	15	225	3 375	50 625	759 375
3	3	3	45	2 025	91 125	4 100 625	184 528 125

phases. More efficient computational schemes are required to apply high-dimensional tandem queuing systems with MAP/PH/1/*M* phases in practice.

3. ESTIMATING THE PERFORMANCE CHARACTERISTICS OF THE HIGH-DIMENSIONAL TANDEM QUEUING SYSTEM

As has been mentioned above, analytically calculating the performance of a tandem queuing system can be inefficient or impossible due to the high dimensionality of MAP-flows. The high computation time of the steady states and steady-state performance characteristics of the system has a considerable effect in iterative problems. For example, when designing a wireless communication network, it is important to select an optimal topology. This involves evaluating the characteristics and choosing the best option at each iteration step. To do it, we propose a new approach based on a combination of machine learning and simulation methods (Fig. 2). Within this approach, for different sets of input parameters, simulation modeling is used to generate a dataset with the calculated performance characteristics of the tandem queuing system. The generated dataset is then applied in a machine learning algorithm to obtain fast estimates of the performance characteristics. The approach was effectively used to solve problems of queuing theory [28–30]. This section describes the simulation model of the tandem queuing system as well as the procedures for validating this model using an analytical model and estimating the performance characteristics of the tandem queuing system using the combined method.

Calculating the steady-state characteristics of a tandem queuing system using the simulation method involves simulating the process of generating new packets and servicing them until a packet is lost or the packet service ends. Data on the generated and lost packets need to be stored. Average values for performance characteristics (phase delays and the probabilities of packet loss) should be calculated. The model is based on the discrete-event simulation principle: only emerging events (new packets generation and service completion) are processed. The time between successive events changes instantaneously: the model state remains invariable without event processing. During event processing, it is possible to calculate and assign the instants when new events will occur during model execution.

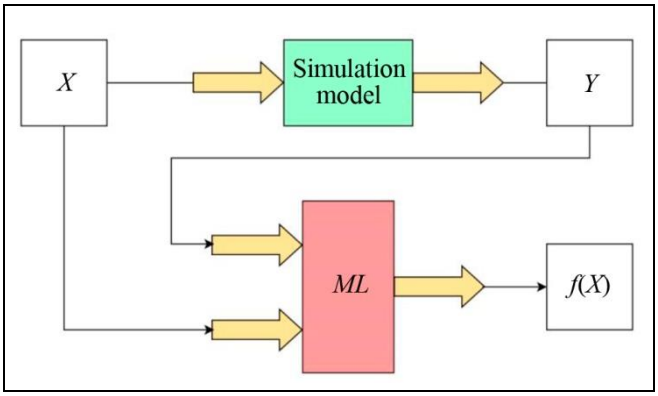

Fig. 2. The fast estimation procedure for the steady-state performance characteristics of the tandem queuing system.

3.1. Estimating the Performance Characteristics of the Tandem Queuing System by Simulation Modeling

Within this study, we developed a simulation model of the tandem queuing system in Python. Some modules of the simulation model were implemented in C++ to improve the performance and efficiency of calculations. Such modules were integrated into the model using Cython. The PyQumo library[1] was applied to implement and study stochastic models, in particular, multiphase queuing systems.

The simulation model has the following input parameters: the dimension N of the tandem queuing system (the number of phases), the buffer sizes $M_i, i = 1,..., N,$ of the phases; the incoming MAP-flow of packets at the first phase of the system, and the PH-

[1] URL: https://github.com/ipu69/pyqumo (Accessed September 16, 2024.)

distributions $PH_i, i = 1,..., N$, of the packet processing time at each phase. If certain phases are available, the incoming MAP cross-traffic flows are specified. If a buffer is empty, the packet goes directly to the server. Each phase receives packets from the previous phase, and the process continues until all packets are serviced at the last phase. If the packet is successfully processed at the last phase, it is considered to be successfully delivered; otherwise, the packet is considered to be lost.

The primary drawback of the simulation method is that the accuracy of the results strongly depends on the number of simulated events. For example, in a tandem queuing system with 10 phases, approximately 100 000 packets must be generated to achieve high accuracy with an error not exceeding 5%. Consequently, this method has limited potential for accelerating computation.

To validate the developed simulation model, we implemented the analytical algorithm for calculating the performance characteristics of the tandem queuing system (see Section 2) in Python. The simulation results were compared with those of the analytical calculations. Validation was performed on a set of 430 random networks, each consisting of 1 to 10 phases. The applicability of the precise algorithm is limited by the dimensionality of the flows of serviced packets (the input flows for the next phases), which grows for the nth phase as $(M+2)^N V^N W$. To ensure the correct operation of the model, the input datasets were generated so that the order of the outgoing flow at the last phase was below 8000.

For different performance characteristics of the tandem network, Fig. 3 shows the dependence of the relative error on the number of packets in the simulation model. The biggest error in the simulation model occurs when calculating the system size at the last phase. In this case, an error within 5% can be achieved by modeling only 25 000 packets; modeling 100 000 packets (see the dataset for training the regression model below) allows reducing the error to 1%.

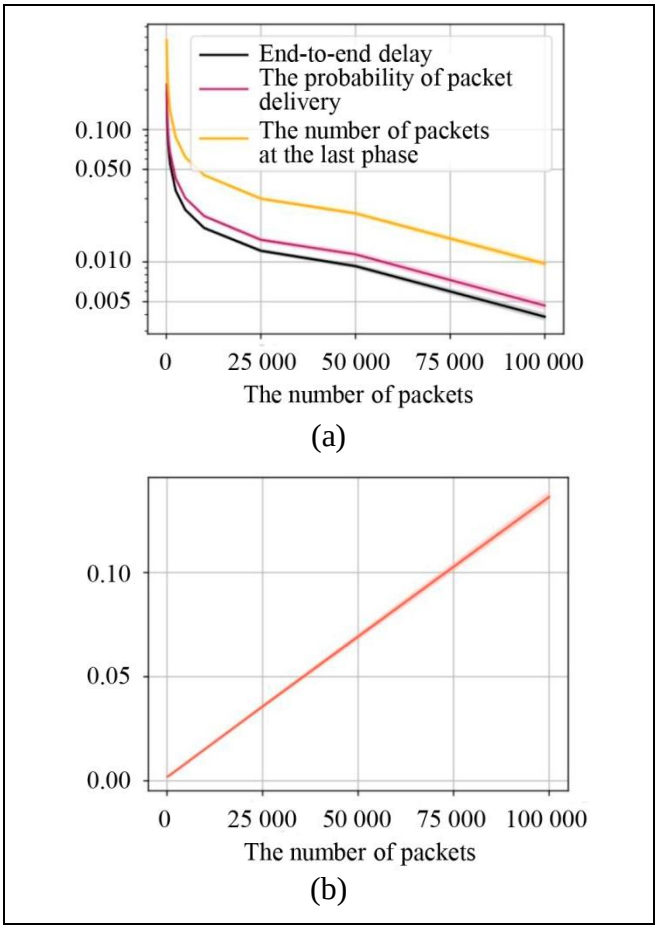

Fig. 3. The convergence of the simulation method and the speed of calculations: (a) relative error and (b) duration.

3.2. Estimating the Steady-State Performance Characteristics of the Tandem Queuing System by Machine Learning Methods

In this paper, we propose to use machine learning methods to accelerate the estimation of the performance characteristics of the tandem queuing system. The proposed methodology is effective in calculating the delay time between two phases and the probability of packet delivery in the tandem queuing system; see numerical examples in Section 4.

The following steady-state performance characteristics were calculated using the simulation model:

- the end-to-end delay in the tandem queuing system, Δt;
- the probability of packet delivery, $1 - P_L$.

Within this study, we built two models for predicting network performance characteristics. The first (regression) model estimates the average end-to-end delay, whereas the second model estimates the probability of successful packet delivery in the system. The models were built under the following constraints:

- the same buffer size for each phase;
- the same PH-distribution of the packet service time for each phase;
- an incoming MAP-flow with the zero autocorrelation coefficient.

Considering the restrictions, the following parameters were taken for the model: the first three moments of the MAP-flow, the first three moments of the PH-distribution, the tandem size (N), and the buffer size (M).

The first three moments were selected as features to characterize the incoming MAP-flow and the PH distributions of the service time. The original distributions can be reconstructed with these features. Let $m_a = \mathbb{E}X$ be the mean time between new packets arrivals and σ_a be its standard deviation. Similarly, let $m_s = \mathbb{E}Y$ be the average service time of a packet, and σ_s be its standard deviation. The service time is described by the mean m_s, the coefficient of variation $c_s = \sigma_s / m_s$, and the coefficient of skewness $\gamma_s = \mathbb{E}[(Y - m_s)^3] / \sigma_s^3$. By analogy, the incoming flow is described by m_a, c_a, and γ_a. As mentioned earlier, the autocorrelation coefficient for the incoming MAP-flow is supposed to be zero. In this case, the first three moments are enough to reconstruct the distributions of the time between packets arrivals and the service time.

In this paper, we employed two methods to reconstruct the distributions. First, we tried to build the second-order acyclic continuous PH-distribution in canonical form (ACPH(2); see Fig. 4a) using the method described in [31]. If the moments fall outside the existence region of ACPH(2), it is necessary to use the more universal method proposed in [32]: the PH-distribution is sought as a mixture of two Erlang distributions $\mathrm{ME}_n(2)$ (see Fig. 4b). In general, the PH-distribution can be found for any values $m > 0$, $c > 0$, and $\gamma > c - 1/c$ of a certain continuous positive distribution [32].

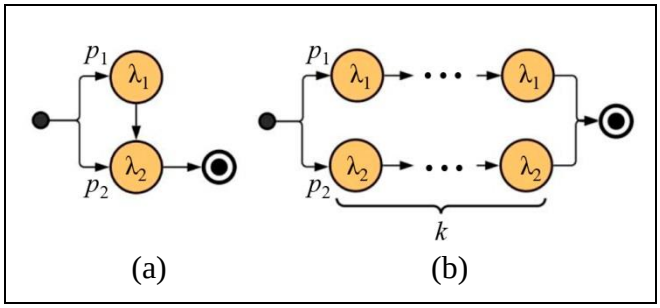

Fig. 4. (a) the acyclic PH-distribution with two ACPH(2) states and (b) the hyper-Erlang distribution with two Erlang distributions of the same order, used to approximate the flows by three moments.

Within this study, the following algorithms and methods were used to forecast the end-to-end delay: least squares method (LSM), tree-based algorithms (decision tree [33–36], gradient boosting [37]), and an artificial neural network with Adam's optimization algorithm [38]. Also, we applied classification models to estimate the probability of packet delivery, $1 - P_L$. Successful delivery was understood in the sense of exceeding a given value B from the interval $(0, 1)$. The algorithm determined the class, $[0, B)$ or $[B, 1]$, that the system belonged to. We used logistic regression, decision tree, gradient boosting, and artificial neural networks with Adam's optimization algorithm to solve the classification problem.

4. NUMERICAL RESULTS

Synthetic data generated by simulation modeling were used to estimate of the performance characteristics of the tandem queuing system. The simulation model received randomly generated input data in the ranges shown in Table 2. The outputs were the average delay time and the average probability of packet delivery, $\{\Delta t, 1 - P_L\}$. Simulation modeling yielded a huge dataset consisting of 101 424 rows of valuable information.

Table 2

Input parameters of the simulation model

Parameter	Range
Packet arrivals	
Mean, m_a	$\sim (0, 10)$
The coefficient of variation, c_a	$\sim (0.5,\ 3)$
The coefficient of skewness, γ_a	$\sim (c_a - \frac{1}{c_a}, 100)$
Service time	
Mean, m_s	$\sim (0{,}10)$
The coefficient of variation, c_s	$\sim (0.5, 3)$
The coefficient of skewness, γ_s	$\sim (c_s - \frac{1}{c_s}, 100)$
Buffer size, M	$\{6, 7, ..., 10\}$
The number of phases, N	$\{1, 2, ..., 20\}$

Since the data were generated randomly, the sample contains values of the loading coefficient $\rho = \frac{m_s}{m_a}$, strongly exceeding the range $(0, 1)$. It is unreasonable to use such data for further model training. Note that the system phases have a limited buffer size, so packets can be lost if a phase is overloaded. Thus, the load on the first phase may be $\rho >> 1$. Therefore, we restrict the analysis to the range $\rho \in (0, 10]$. After eliminating outliers, the sample size became 96 248 (rows).

4.1. Analysis of End-to-End Delays

We used the following metrics to assess the forecast values: the correlation coefficient (R), the

standard deviation (STD), and the coefficient of determination (R^2).

We analyzed the end-to-end delays estimated for different network dimensionalities, $N = 1, 5, 10, 20$. Figure 5 illustrates the architecture of an artificial neural network featuring a single hidden layer with 40 neurons. The sigmoidal activation function was selected, and training was conducted over 1000 epochs.

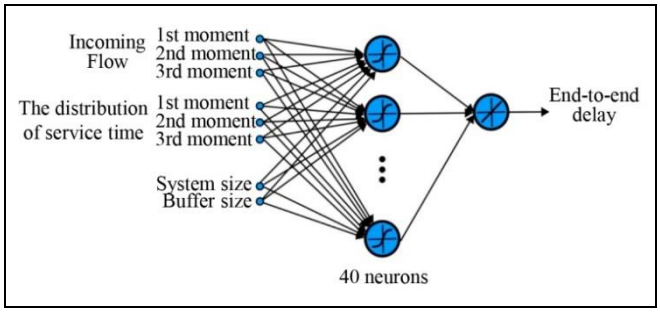

Fig. 5. The neural network architecture for forecasting the end-to-end delay.

The distribution densities of all end-to-end delay estimates in the test sample are presented in Fig. 6. Clearly, the least squares method produced the worst forecasting results; see the green curves in the graph. Due to the linear approximation, most of the estimates turned out to be negative. For the decision tree, the root-mean-square (RMS) error was taken as the partition criterion at each node of the tree. The best estimates were obtained for a tree depth of 36. For the gradient boosting method on the decision tree, a tree depth of 10 was chosen empirically. In the case under consideration, the learning rate was set equal to 0.1 and the number of trees equal to 100. Figure 7 shows the scatter diagrams of all trained models. According to Table 3, the neural network-based model demonstrated the best-quality forecasting.

4.2. Analysis of the Probability of Packet Delivery

Unlike the regression models for estimating the end-to-end delay, the delivery probability model does not need to forecast particular values. It is much more important to predict whether the delivery of packets will be successful or not. Let B be a given threshold for the successful delivery condition. We classified the probabilities of packet delivery into two groups: $1-P_L \in [B, 1]$ (successful delivery) and $P_L \in [0, B)$ (packet loss). In the numerical experiment, we took the threshold $B = 0.9$ for all models. The models were assessed in terms of the following metrics: Precision, Recall, and F_1.

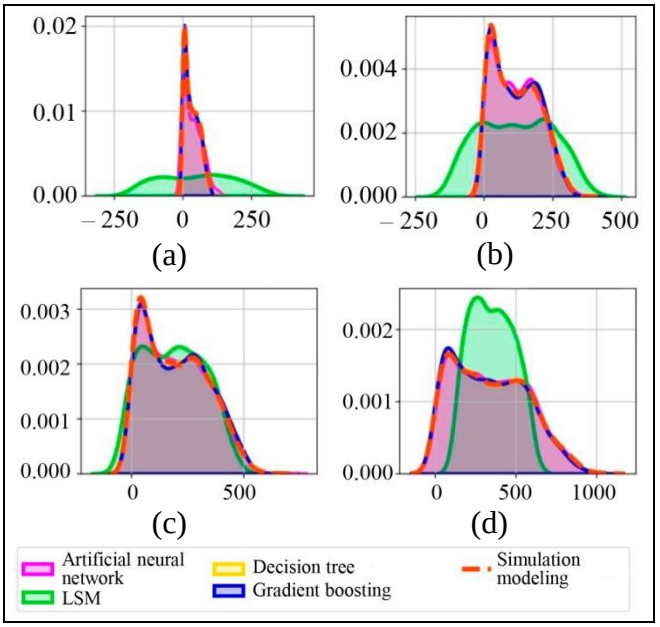

Fig. 6. The distribution density of the end-to-end delay: (a) network size 1, (b) network size 5, (c) network size 10, and (d) network size 20.

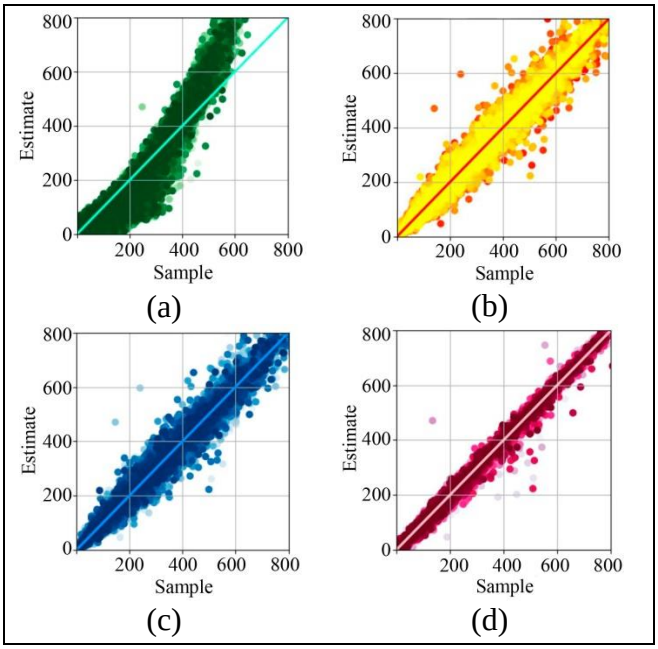

Fig. 7. The scatter diagrams of regression models: (a) LSM, (b) decision tree, (c) gradient boosting, and (d) artificial neural network.

Table 3

Values of the metrics for different forecasting algorithms (the end-to-end delay)

Model	Metrics		
	R	STD	R^2
LSM	0.926	66.31	0.85
Decision tree	0.990	24.48	0.98
Gradient boosting	0.990	24.71	0.98
Artificial neural network	0.998	12.23	0.99

Consider again the tandem queuing system with lengths N = [1, 5, 10, 20] and estimate the probability of packet delivery for the artificial neural network. We selected a multilayer perceptron with 3 hidden layers, each containing 16 neurons (Fig. 8), and the sigmoidal activation function.

Logistic regression forecasted packet delivery with a large share of errors. The decision tree with a depth of 10 demonstrated better results. The best forecasts for the classification problem were obtained using the gradient boosting method and the artificial neural network. The values of the classification metrics are combined in Table 4.

Figure 9 shows the estimated probabilities of packet delivery in all models for the network size of N = 10. The trend indicates the actual values of the probability $1-P_L$ for different loading coefficients ρ. The green color corresponds to the forecasts of packet delivery $(1-P_L \geqslant B)$ and the red to those of packet loss $(1-P_L < B)$. Among all the models, gradient boosting and neural network stand out as the ones providing the most accurate forecasts.

Table 4

Values of the metrics for different classification models (the probability of packet delivery)

Model	Metrics		
	Precision	Recall	F_1
Logistic regression	0.804	0.821	0.813
Decision tree	0.9618	0.905	0.912
Gradient boosting	0.966	0.969	0.968
Artificial neural network	0.977	0.951	0.964

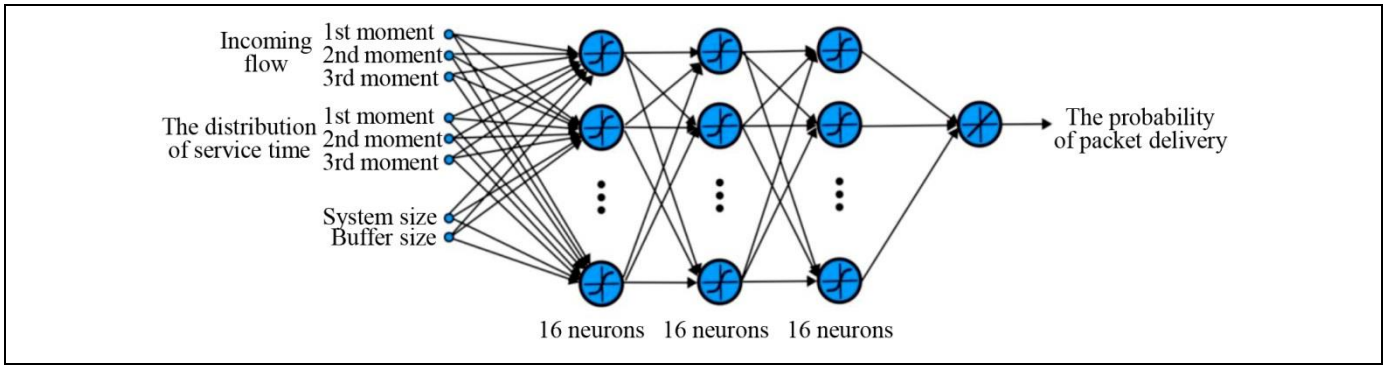

Fig. 8. The neural network architecture for classifying the tandem queuing system by the probability of packet delivery.

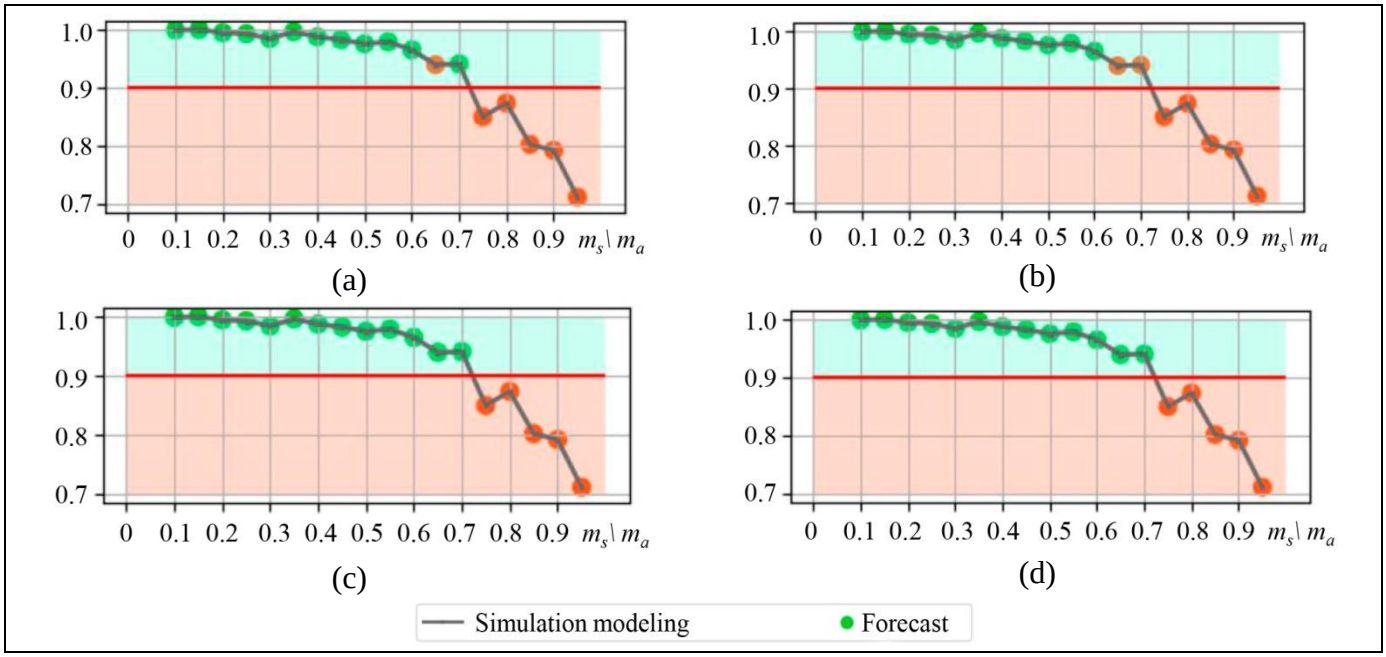

Fig. 9. The estimates of the probability of packet delivery for the tandem queuing system of size 10: (a) logistic regression, (b) decision tree, (c) gradient boosting, and (d) artificial neural network.

4.3. Calculation Time Analysis

For the obtained models, the calculation times of the steady-state performance characteristics by different methods are presented in Table 5. The calculation times of the simulation model depending on the system loading coefficient are also presented for comparison. The sample size is 360 rows. It is reasonable to compare the calculation time for the entire dataset because the time varies in individual cases depending on the tandem length. Note that the calculation time is the time to obtain the final result. Obviously, in each new model run, the time will also vary depending on the CPU occupancy of the PC. Also, the simulation model calculates two network performance characteristics per run. Machine learning models typically take much less time for calculations than simulation models.

Table 5

The time to calculate performance characteristics

Model	Calculation time, s
End-to-end delay	
Simulation model	172.2
LSM	$4.8 \cdot 10^{-6}$
Decision tree	$5.5 \cdot 10^{-6}$
Gradient boosting	$5 \cdot 10^{-6}$
Artificial neural network	$5.7 \cdot 10^{-6}$
The probability of packet delivery	
Simulation model	172.2
Logistic regression	$5.3 \cdot 10^{-6}$
Decision tree	$5 \cdot 10^{-6}$
Gradient boosting	$4.8 \cdot 10^{-6}$
Artificial neural network	$5.3 \cdot 10^{-6}$

CONCLUSIONS

In this paper, we have developed new approaches to studying the performance characteristics of tandem queuing systems. We have described a precise algorithm for calculating a low-dimensional tandem queuing system as well as its complexity estimates and application to the validation of a simulation model. For the investigation of high-dimensional tandem queuing systems, we have proposed an effective approach based on a combination of simulation modeling and machine learning methods. Several machine learning methods, including decision trees, gradient boosting, artificial neural networks, etc., have been comparatively analyzed. According to the numerical examples, machine learning methods demonstrate high effectiveness as well as a sharp reduction in the calculation time.

Acknowledgments. *This work was supported by the Russian Science Foundation, project no. 22-49-02023, https://rscf.ru/project/22-49-02023/.*

REFERENCES

1. Khayyati, S. and Tan, B., Supervised-Learning-Based Approximation Method for Multi-Server Queueing Networks under Different Service Disciplines with Correlated Interarrival and Service Times, *International Journal of Production Research*, 2022, vol. 60, no. 17, pp. 5176–5200. https://doi.org/10.1080/00207543.2021.1951448.
2. Kumar, B.K., Sankar, R., Krishnan, R.N., and Rukmani, R., Performance Analysis of Multi-processor Two-Stage Tandem Call Center Retrial Queues with Non-Reliable Processors, *Methodology and Computing in Applied Probability*, 2022, vol. 24, no. 1, pp. 95–142. https://doi.org/10.1007/s11009-020-09842-6.
3. Oblakova, A., Al Hanbali, A., Boucherie, R.J., van Ommeren, J.C., and Zijm, W.H., An Analytical Model for a Tandem of Two Traffic-Light Intersections under Semi-Actuated and Fixed Control, *Transportation Research Interdisciplinary Perspectives*, 2022, vol. 16, art. no. 100715. https://doi.org/10.1016/j.trip.2022.100715.
4. Rovetto, C., Cruz, E., Nuñez, I., Santana, K., Smolarz, A., Rangel, J., and Cano, E.E., Minimizing Intersection Waiting Time: Proposal of a Queue Network Model Using Kendall's Notation in Panama City, *Applied Sciences*, 2023, vol. 13, no. 18, art. no. 10030. https://doi.org/10.3390/app131810030.
5. Pershin, O.Y., Mukhtarov, A.A., Vishnevsky, V.M., and Larionov, A.A., Optimal Placement of Base Stations in Integrated Design of Wireless Networks, *Programming and Computer Software*, 2023, vol. 49, suppl. 2, pp. S82–S90. https://doi.org/10.1134/S0361768823100055.
6. Vishnevsky, V., Krishnamoorthy, A., Kozyrev, D., and Larionov, A., Review of Methodology and Design of Broadband Wireless Networks with Linear Topology, *Indian Journal of Pure and Applied Mathematics*, 2016, vol. 47, no. 2, pp. 329–342. https://doi.org/10.1007/s13226-016-0190-7.
7. Gnedenko, B.W. and König, D., *Handbuch der Bedienungstheorie II*, Berlin: De Gruyter, 1984. https://doi.org/10.1515/9783112614747.
8. Neuts, M.F., Two Queues in Series with a Finite, Intermediate Waitingroom, *Journal of Applied Probability*, 1968, vol. 5, no. 1, pp. 123–142. https://doi.org/10.2307/3212081.
9. Dieleman, N.A., Berkhout, J., and Heidergott, B., A Neural Network Approach to Performance Analysis of Tandem Lines: the Value of Analytical Knowledge, *Computers and Operations Research*, 2023, vol. 152, no. 3, art. no. 106124. DOI: https://doi.org/10.1016/j.cor.2022.106124.
10. Dudin, S.A., Dudin, A.N., Dudina, O.S., and Chakravarthy, S.R., Analysis of a Tandem Queuing System with Blocking and Group Service in the Second Node, *International Journal of Systems Science: Operations and Logistics*, 2023, vol. 10, no. 1, art. no. 2235270. DOI: https://doi.org/10.1080/23302674.2023.2235270.
11. Dudin, S.A., Dudina, O.S., and Dudin, A.N., Analysis of Tandem Queue with Multi-Server Stages and Group Service at the Second Stage, *Axioms*, 2024, vol. 13, no. 4, art. no. 214. DOI: https://doi.org/10.3390/axioms13040214.

12. Bocharov, P.P., Manzo, R., and Pechinkin, A.V., Analysis of a Two-Phase Queueing System with a Markov Arrival Process and Losses, *Journal of Mathematical Sciences*, 2005, vol. 131, no. 3, pp. 5606–5613. https://doi.org/10.1007/s10958-005-0432-4.
13. Kim, C.S., Klimenok, V., and Taramin, O., A Tandem Retrial Queueing System with Two Markovian Flows and Reservation of Channels, *Computers and Operations Research*, 2010, vol. 37, no. 7, pp. 1238–1246. https://doi.org/10.1016/j.cor.2009.03.030.
14. Kim, C., Klimenok, V.I., and Dudin, A.N., 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.
15. Klimenok, V., Breuer, L., Tsarenkov, G., and Dudin, A., The BMAP/G/1/–>·/PH/1/M Tandem Queue with Losses, *Performance Evaluation*, 2005, vol. 61, no. 1, pp. 17–40. https://doi.org/10.1016/j.peva.2004.09.001.
16. Lian, Z. and Liu, L., A Tandem Network with MAP Inputs, *Operations Research Letters*, 2008, vol. 36, no. 2, pp. 189–195. https://doi.org/10.1016/j.orl.2007.04.004.
17. Vishnevskii, V.M. and Dudin, A.N., Queueing Systems with Correlated Arrival Flows and Their Applications to Modeling Telecommunication Networks, *Automation and Remote Control*, 2017, vol. 78, no. 8, pp. 1361–1403. https://doi.org/10.1134/S000511791708001X.
18. Dudin, A.N., Klimenok, V.I., and Vishnevsky, V.M., *The Theory of Queuing Systems with Correlated Flows*, Cham: Springer, 2019.
19. Bruell, S.C., Balbo, G., and Afshari, P.V., Mean Value Analysis of Mixed, Multiple Class BCMP Networks with Load Dependent Service Stations, *Performance Evaluation*, 1984, vol. 4, no. 4, pp. 241–260. https://doi.org/10.1016/0166-5316(84)90010-5.
20. Vishnevsky, V., Klimenok, V., Sokolov, A., and Larionov, A., Performance Evaluation of the Priority Multi-Server System MMAP/PH/M/N Using Machine Learning Methods, *Mathematics*, 2021, vol. 9, no. 24, art. no. 3236. https://doi.org/10.3390/math9243236.
21. Klimenok, V., Dudin, A., and Vishnevsky, V., On the Stationary Distribution of Tandem Queue Consisting of a Finite Number of Stations, in *Communications in Computer and Information Science*, Springer, 2012, vol. 291, pp. 383–392. https://doi.org/10.1007/978-3-642-31217-5_40.
22. Palomo, S. and Pender, J., Learning the Tandem Network Lindley Recursion, *Proceedings of the 2021 Winter Simulation Conference (WSC)*, Phoenix, USA, December 2021, pp. 1–12. https://doi.org/10.1109/WSC52266.2021.9715530.
23. Rabta, B., A Review of Decomposition Methods for Open Queueing Networks, in *Rapid Modelling for Increasing Competitiveness*, Reiner, G., Ed., London: Springer, 2009, pp. 25–42. https://doi.org/10.1007/978-1-84882-748-6_3.
24. Vishnevsky, V., Larionov, A., Roman, I., and Semenova, O., Estimation of IEEE 802.11 DCF Access Performance in Wireless Networks with Linear Topology Using PH Service Time Approximations and MAP Input, *Proceedings of the 11th IEEE International Conference on Application of Information and Communication Technologies (AICT 2017)*, Moscow, 2017, pp. 1–5. https://doi.org/10.1109/ICAICT.2017.8687247.
25. Gorbunova, A.V., Vishnevsky, V.M., and Larionov, A.A., Evaluation of the End-to-End Delay of a Multiphase Queuing System Using Artificial Neural Networks, in *Lecture Notes in Computer Science*, Cham: Springer, 2021, vol. 12563, pp. 631–642. https://doi.org/10.1007/978-3-030-66471-8_48.
26. Kudou, T., Nii, S., and Okuda, T., A Performance Evaluation of Tandem Queueing Systems by Machine Learning, *Proceedings of the 2022 IEEE International Conference on Consumer Electronics (ICCE-Taiwan 2022)*, Taiwan, 2022, pp. 389–390. https://doi.org/10.1109/ICCE-Taiwan55306.2022.9869030.
27. Kudou, T. And Okuda, T., A Time Series Analysis of Single Server Queueing Systems by Using Machine Learning, *Proceedings of the 2023 IEEE International Conference on Consumer Electronics (ICCE-Taiwan 2023)*, Taiwan, 2023, pp. 327–328. https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226861.
28. Vishnevsky, V.M., *Teoreticheskie osnovy proektirovaniya komp'yuternykh setei* (Theoretical Foundations of Computer Network Design), Moscow: Tekhnosfera, 2003. (In Russian.)
29. Vishnevsky, V.M., Klimenok, V.I., Sokolov, A.M., and Larionov, A.A., Investigation of the Fork–Join System with Markovian Arrival Process Arrivals and Phase-Type Service Time Distribution Using Machine Learning Methods, *Mathematics*, 2024, vol. 12, no. 5, art. no. 659. https://doi.org/10.3390/math12050659.
30. Efrosinin, D., Vishnevsky, V., and Stepanova, N., Optimal Scheduling in General Multi-Queue System by Combining Simulation and Neural Network Techniques, *Sensors*, 2023, vol. 23, no. 12, art. no. 5479. https://doi.org/10.3390/s23125479.
31. Telek, M. and Heindl, A., Matching Moments for Acyclic Discrete and Continuous Phase-Type Distributions of Second Order, *International Journal of Simulation Systems, Science and Technology*, 2002, vol. 3, no. 3, pp. 47–57.
32. Johnson, M.A. and Taaffe, M.R., Matching Moments to Phase Distributions: Mixtures of Erlang Distributions of Common Order, *Communications in Statistics. Stochastic Models*, 1989, vol. 5, no. 4, pp. 711–743. https://doi.org/10.1080/15326348908807131.
33. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, New York: Chapman and Hall/CRC, 1984. https://doi.org/10.1201/9781315139470.
34. Demidova, L.A. and Usachev, P.O., Development and Approbation of the Improved CART Algorithm Version, *Journal of Physics: Conference Series*, 2020, vol. 1479, no. 1, art. no. 012085. https://doi.org/10.1088/1742-6596/1479/1/012085.
35. Gordon, A.D., A Review of *Classification and Regression Trees* by L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Biometrics*, 1984, vol. 40, no. 3, p. 874. https://doi.org/10.2307/2530946.
36. Loh, W.Y., Classification and Regression Trees, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2011, vol. 1, no. 1, pp. 14–23. https://doi.org/10.1002/widm.8.
37. Friedman, J.H., Stochastic Gradient Boosting, *Computational Statistics and Data Analysis*, 2002, vol. 38, no. 4, pp. 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.
38. Kingma, D.P. and Ba, J.L., Adam: A Method for Stochastic Optimization, *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, http://arxiv.org/abs/1412.6980.

This paper was recommended for publication by R.V. Meshcheryakov, a member of the Editorial Board.

Received August 1, 2024, and revised September 6, 2024. Accepted September 13, 2024.

Author information

Vishnevsky, Vladimir Mironovich. Dr. Sci. (Eng.), Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
✉ vishn@inbox.ru
ORCID iD: https://orcid.org/0000-0001-7373-4847

Larionov, Andrei Alekseevich. Cand. Sci. (Eng.), Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
✉ larioandr@gmail.com
ORCID iD: https://orcid.org/0000−0003−0539−0442

Mukhtarov, Amir Amangel'dyevich. Cand. Sci. (Eng.), Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
✉ mukhtarov.amir.a@gmail.com
ORCID iD: https://orcid.org/0000−0002−8191−6381

Sokolov, Aleksandr Mikhailovich. Researcher, Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
✉ aleksandr.sokolov@phystech.edu
ORCID iD: https://orcid.org/0000−0002−3589−5700

Cite this paper

Vishnevsky, V.M., Larionov, A.A., Mukhtarov, A.A., and Sokolov, A.M., Investigation of Tandem Queuing Systems Using Machine Learning Methods. *Control Sciences* **4**, 10–21 (2024). http://doi.org/10.25728/cs.2024.4.2

Original Russian Text © Vishnevsky, V.M., Larionov, A.A., Mukhtarov, A.A., Sokolov, A.M., 2024, published in *Problemy Upravleniya*, 2024, no. 4, pp. 13–25.

Translated into English by the authors;
finally edited by *Alexander Yu. Mazurov*,
Cand. Sci. (Phys.–Math.),
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com