

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ MATHEMATICAL MODELING



УДК 519.872
DOI: 10.37661/1816-0301-2025-22-2-48-62

Оригинальная статья
Original Article

Анализ полуоткрытой сети массового обслуживания с адаптацией скоростей обслуживания к скоростям поступления запросов

А. Н. Дудин, О. С. Дудина, С. А. Дудин[✉]

Белорусский государственный университет,
пр. Независимости, 4, Минск, 220030, Беларусь
✉E-mail: dudins@bsu.by

Аннотация

Цели. Решается задача нахождения основных вероятностных характеристик функционирования полуоткрытой сети массового обслуживания, в которой интенсивность входного потока в узлы имеет несколько возможных уровней. При изменении уровня интенсивности потока возможно изменение скорости обслуживания запросов в узлах сети с целью оптимизации ее функционирования.

Методы. Для решения используется аппарат многомерных цепей Маркова с непрерывным временем и специальной блочной структурой инфинитезимального генератора.

Результаты. Получены алгоритмы и формулы для вычисления блоков генератора этой цепи, что позволяет вычислять ее инвариантное распределение вероятностей. Найдены формулы для вычисления основных характеристик производительности сети по известному инвариантному распределению вероятностей состояний цепи Маркова. Приведен численный пример, иллюстрирующий зависимость основных характеристик производительности сети от пропускной способности ее узлов. При выбранном экономическом критерии качества функционирования сети продемонстрирована возможность оптимизации перераспределения ресурсов сети между ее узлами при изменении уровня входного потока.

Заключение. Полученные результаты могут быть использованы для оптимизации функционирования различных реальных объектов, описываемых полуоткрытыми сетями массового обслуживания, например телекоммуникационных и логистических систем, мобильных роботизированных систем хранения, за счет адаптации распределения ресурсов сети между ее узлами к изменяющейся скорости поступления запросов.

Ключевые слова: полуоткрытая сеть массового обслуживания, изменение скорости поступления запросов, анализ производительности, цепь Маркова, вероятность отказа

Для цитирования. Дудин, А. Н. Анализ полуоткрытой сети массового обслуживания с адаптацией скоростей обслуживания к скоростям поступления запросов / А. Н. Дудин, О. С. Дудина, С. А. Дудин // Информатика. – 2025. – Т. 22, № 2. – С. 48–62. – DOI: 10.37661/1816-0301-2025-22-2-48-62.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 17.04.2025
Подписана в печать | Accepted 02.05.2025
Опубликована | Published 30.06.2025

Analysis of a semi-open queueing network with adaptation of service rates to the rates of arriving requests

Alexander N. Dudin, Olga S. Dudina, Sergei A. Dudin[✉]

Belarusian State University,
av. Nezavisimosti, 4, Minsk, 220030, Belarus
✉E-mail: dudins@bsu.by

Abstract

Objectives. The problem of computation of the main probabilistic characteristics of operation of a semi-open queueing network, in which the intensity of the input flow to the nodes has several possible levels, is considered. When the changing the level of flow intensity occurs, it is possible to change the rate of requests service in the nodes in order to optimize the network functioning.

Methods. The solution is based on the apparatus of multidimensional Markov chains with continuous time and a special block structure of the infinitesimal generator.

Results. The generator blocks of this chain are calculated using algorithms and formulas, which allows the invariant probability distribution to be determined. Formulas for calculating the main characteristics of network performance using the known invariant probability distribution of the states of the Markov chain are derived. A numerical example is provided to illustrate how the dependence of the main characteristics of network performance depend on the throughput of its nodes. Using the selected economic criterion for network performance quality, it is demonstrated that the redistribution of network resources between its nodes can be optimised with the change of the arrival flow level.

Conclusion. The obtained results can be used to optimize the functioning of various real objects described by semi-open queueing networks, for example, telecommunication and logistic systems, mobile robotic fulfillment systems, by adapting the distribution of network resources between its nodes to the changing rate of incoming requests.

Keywords: semi-open queueing network, request arrival rate variation, performance analysis, Markov chain, loss probability

For citation. Dudin A. N., Dudina O. S., Dudin S. A. *Analysis of a semi-open queueing network with adaptation of service rates to the rates of arriving requests*. Informatika [Informatics], 2025, vol. 22, no. 2, pp. 48–62 (In Russ.). DOI: 10.37661/1816-0301-2025-22-2-48-62.

Conflict of interest. The authors declare of no conflict of interest.

Введение. Сеть массового обслуживания представляет собой совокупность конечного числа взаимосвязанных узлов, работа каждого из которых описывается системой массового обслуживания. По сети циркулируют запросы, которые после обслуживания в одном из узлов переходят в другой узел сети в соответствии с маршрутной матрицей или покидают сеть. Такие сети являются адекватными математическими моделями, например, телекоммуникационных, транспортных, производственных, логистических и других реальных систем. Поэтому существует очень обширная литература, посвященная их исследованию. В большинстве работ исследованы так называемые открытые сети, в которых неограниченное число запросов прибывают извне и после получения обслуживания покидают сеть навсегда, и замкнутые сети, в которых непрерывно циркулирует конечное фиксированное число запросов. Относительно недавно (в начале 2000-х гг.) начато исследование так называемых полуоткрытых сетей, в которых запросы так же, как и в открытых сетях, прибывают извне и уходят после получения обслуживания. Однако число запросов, которые могут одновременно присутствовать в системе, ограничено некоторой константой, она иногда называется уровнем мультипрограммирования. Популярность исследования полуоткрытых сетей объясняется их применимостью для анализа и оптимизации многих реальных систем, например для описания работы телекоммуникационных сетей, комплексов для проведения погрузочно-разгрузочных работ, гибких роботизированных производств и т. д. Для подробного ознакомления с тематикой исследования таких сетей можно рекомендовать, например, работы [1–8].

Во всех работах по исследованию полуоткрытых сетей массового обслуживания за исключением работы [8], в которой допускалось изменение параметров сети в зависимости от числа находящихся в ней запросов, предполагается, что параметры сети не изменяются в процессе ее функционирования. Однако хорошо известно, что данные параметры могут изменяться. В частности, может существенно изменяться интенсивность входного потока запросов в различные узлы сети (например, в зависимости от времени суток, погодных условий, возникновения чрезвычайных ситуаций и т. д.). Для обеспечения хорошего качества обслуживания запросов имеет смысл реагировать на изменения интенсивности потока соответствующим изменением интенсивности обслуживания запросов. Например, применительно к телекоммуникационным сетям можно запрашивать большую пропускную способность каналов или оборудования, предоставляемых для обеспечения устойчивой работы узлов. Если это невозможно, необходимо производить перераспределение имеющегося ресурса сети между ее узлами в зависимости от их загруженности.

В настоящей статье проанализирована модель полуоткрытой сети массового обслуживания со случайно меняющейся скоростью поступления запросов в узлы, которая может быть полезна для решения задачи оптимизации стратегии требования необходимого ресурса или перераспределения ресурса сети между ее узлами.

Отметим, что большинство существующих результатов в теории сетей массового обслуживания в силу сложности их анализа относятся к анализу тандемных сетей, имеющих линейную топологию; получению условий мультипликативности стационарного распределения сети, т. е. возможности получения стационарного распределения вероятностей состояний сети как произведения вероятностей состояний узлов сети, либо к получению различного рода приближенных решений.

В данной работе приведены точные алгоритмические результаты для нахождения стационарного распределения вероятностей состояний полуоткрытой сети с произвольной топологией и маршрутизацией запросов.

Математическая модель. Рассмотрим полуоткрытую сеть массового обслуживания, состоящую из K , $K \geq 2$, узлов (рис. 1).

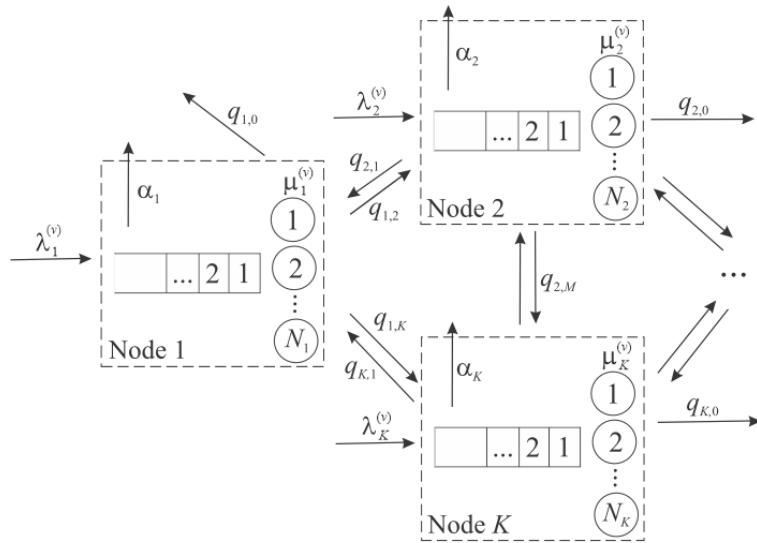


Рис. 1. Структура сети

Fig. 1. The structure of the network

Прибытие запросов в сеть определяется следующим образом. Предполагается, что имеется $V, V \geq 2$, уровней интенсивности входного потока. Переходы между уровнями регулируются цепью Маркова с непрерывным временем $v_t, t \geq 0$, и конечным пространством состояний

$\{1, 2, \dots, V\}$. Интенсивности переходов цепи v_t определяются неприводимым генератором Λ . Инвариантный вероятностный вектор v цепи v_t вычисляется как решение системы уравнений $v\Lambda = \mathbf{0}$, $v\mathbf{e} = 1$. Здесь $\mathbf{0}$ – это вектор-строка, состоящий из нулей, а \mathbf{e} – вектор-столбец, состоящий из единиц.

Предположим, что при уровне v процесса v_t , $t \geq 0$, в узел k , $k = \overline{1, K}$, поступает стационарный пуассоновский поток с интенсивностью $\lambda_k^{(v)}$. Общая интенсивность поступления в сеть при фиксированном уровне v входного процесса определяется как $\lambda^{(v)} = \sum_{k=1}^K \lambda_k^{(v)}$. Общая средняя интенсивность поступления запросов в сеть составляет $\lambda = \sum_{v=1}^V (v)_v \lambda^{(v)}$. Средняя интенсивность поступления в k -й узел $\lambda_k = \sum_{v=1}^V (v)_v \lambda_k^{(v)}$, $k = \overline{1, K}$.

В сети могут одновременно обрабатываться не более N , $N \geq 1$, запросов. Запрос, прибывающий, когда количество запросов в сети равно N , не принимается на обслуживание и считается потерянным. Работа каждого узла сети описывается многолинейной системой массового обслуживания с буфером емкости, гарантирующей невозможность его переполнения. Предполагается, что k -й узел имеет N_k , $k = \overline{1, K}$, идентичных приборов. Время обслуживания прибором в k -м узле при состоянии v процесса поступления запросов распределено экспоненциально с параметром $\mu_k^{(v)}$.

После получения обслуживания в k -м узле, $k = \overline{1, K}$, с вероятностью $q_{k,k'}$ запрос переходит на обслуживание в k' -й узел, $k' = \overline{1, M}$, или с вероятностью $q_{k,0}$ завершает обслуживание в сети, $\sum_{k'=0}^M q_{k,k'} = 1$.

Предполагается, что запросы, ожидающие в буферах, являются нетерпеливыми. Запрос, находящийся в k -м буфере, навсегда покидает сеть после времени, распределенного экспоненциально с параметром α_k , $k = \overline{1, K}$, независимо от других ожидающих запросов.

Целями исследования являются анализ стационарного распределения состояний описанной сети массового обслуживания, вывод выражений для вычисления ее основных характеристик производительности, а также численная иллюстрация поведения этих характеристик.

Случайный процесс, описывающий динамику сети. Для описания поведения исследуемой сети может быть использована следующая цепь Маркова с непрерывным временем:

$$\Psi_t = \{n_t, v_t, n_t^{(1)}, \dots, n_t^{(K)}\}, t \geq 0,$$

где n_t – общее количество запросов в сети, $n_t = \overline{0, N}$;

v_t – уровень входного процесса, $v_t = \overline{1, V}$;

$n_t^{(k)}$ – количество запросов в k -м узле, $n_t^{(k)} = \overline{0, n_k}$, $\sum_{k=1}^K n_t^{(k)} = n_t$, $k = \overline{1, K}$,

в произвольный момент времени t .

Цепь Маркова Ψ_t , $t \geq 0$, является регулярной, неприводимой и имеет конечное пространство состояний. Поэтому стационарные (инвариантные) вероятности состояний цепи

$$\pi(n, v, n^{(1)}, \dots, n^{(K)}) = \lim_{t \rightarrow \infty} P\{n_t = n, v_t = v, n_t^{(1)} = n^{(1)}, \dots, n_t^{(K)} = n^{(K)}\},$$

$$n = \overline{0, N}, v = \overline{1, V}, n^{(k)} = \overline{0, n}, \sum_{k=1}^K n^{(k)} = n, k = \overline{1, K},$$

существуют для любых значений параметров рассматриваемой сети массового обслуживания.

Пусть $\boldsymbol{\pi}_n = (\boldsymbol{\pi}(n, 1), \boldsymbol{\pi}(n, 2), \dots, \boldsymbol{\pi}(n, V))$, $n = \overline{0, N}$, есть векторы-строки, состоящие из вероятностей состояний со значением n первой компоненты, перенумерованных в прямом лексикографическом порядке значений компоненты v_t и в обратном лексикографическом порядке значений компоненты $n_t^{(1)}, \dots, n_t^{(K)}$.

Известно, что векторы $\boldsymbol{\pi}_n$, $n = \overline{0, N}$, удовлетворяют системе линейных алгебраических уравнений

$$(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_N)G = \mathbf{0}, \quad (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_N)\mathbf{e} = 1, \quad (1)$$

где G – инфинитезимальный генератор цепи Маркова ψ_t , $t \geq 0$. Для решения системы (1) нужно определить генератор G .

Теорема. Инфинитезимальный генератор G цепи Маркова ψ_t , $t \geq 0$, имеет блочную трехдиагональную структуру $G = (G_{n,n'})$, $|n - n'| \leq 1$, $n, n' = \overline{0, N}$, с ненулевыми блоками:

$$\begin{aligned} G_{0,0} &= \Lambda - \text{diag}\{\lambda^{(v)}, v = \overline{1, V}\}, \\ G_{n,n} &= \Lambda \otimes I_{J_n} + \text{diag}\{-\lambda^{(v)} I_{J_n} + (\Psi_n^{(v)}(Q^{(v)}) + \Delta_n^{(v)}), v = \overline{1, V}\}, n = \overline{1, N-1}, \\ G_{N,N} &= \Lambda \otimes I_{J_N} + \text{diag}\{(\Psi_N^{(v)}(Q^{(v)}) + \Delta_N^{(v)}), v = \overline{1, V}\}, \\ G_{n,n-1} &= \text{diag}\{S_n^{(v)}(\mathbf{q}^{(v)}) + A_n(\boldsymbol{\alpha}), v = \overline{1, V}\}, n = \overline{1, N}, \\ G_{n,n+1} &= \text{diag}\left\{\sum_{k=1}^K \lambda_k^{(v)} P_n(\boldsymbol{\beta}_k), v = \overline{1, V}\right\}, n = \overline{0, N-1}. \end{aligned}$$

Здесь и далее:

- $\text{diag}\{A_v, v = \overline{1, V}\}$ – диагональная матрица с диагональными элементами, указанными в скобках;
- $\text{diag}^+\{A_v, v = \overline{1, V}\}$ – наддиагональная матрица с наддиагональными элементами, указанными в скобках;
- $\text{diag}^-\{A_v, v = \overline{1, V}\}$ – поддиагональная матрица с поддиагональными элементами, указанными в скобках;
- \otimes – символ Кронекерова произведения матриц [9];
- $J_n = C_{n+K-1}^{K-1} = \frac{(n+K-1)!}{n!(K-1)!}$;
- $\boldsymbol{\alpha}$ – вектор-столбец, определенный как

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_K)^T;$$

- $\boldsymbol{\beta}_k$, $k = \overline{1, K}$, – векторы размера K со всеми нулевыми элементами, кроме k -го элемента, который равен единице;
- $P_n(\boldsymbol{\beta}_k)$, $n = \overline{0, N-1}$, $k = \overline{1, K}$, – матрица, задающая вероятности переходов процесса $\mathbf{n}_t = \{n_t^{(1)}, \dots, n_t^{(K)}\}$, $t \geq 0$, в момент принятия запроса в k -й узел сети, при условии, что в сети

находится n запросов. Данные вероятности не зависят от состояния процесса поступления. Таким образом, для вычисления $P_n(\beta_k)$ можно применить алгоритм из работы [6, лемма 4];

• $A_n(\alpha)$, $n = \overline{1, N}$, – матрицы, содержащие интенсивности переходов процесса \mathbf{n}_t , $t \geq 0$, которые вызваны потерей запроса из-за нетерпеливости в случае нахождения n запросов в сети. Они определяются как $A_n(\alpha) = A_n^{(K-1)}(\alpha)$, где матрицы $A_n^{(K-1)}(\alpha)$ вычисляются из рекурсий:

$$A_n^{(0)}(\alpha) = \max\{0, n - N_K\} \alpha_K, \\ A_n^{(k)}(\alpha) = \begin{pmatrix} \max\{0, n - N_{K-k}\} \alpha_{K-k} I & O & \cdots & O \\ A_1^{(k-1)}(\alpha) & \max\{0, n - 1 - N_{K-k}\} \alpha_{K-k} I & \cdots & O \\ O & A_2^{(k-1)}(\alpha) & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \max\{0, 1 - N_{K-k}\} \alpha_{K-k} I \\ O & O & \cdots & A_n^{(k-1)}(\alpha) \end{pmatrix},$$

$$k = \overline{1, K-1}, n = \overline{2, N};$$

• $\mathbf{q}^{(v)}$, $v = \overline{1, V}$, – вектор-столбец, заданный как

$$\mathbf{q}^{(v)} = (q_1^{(v)}, q_2^{(v)}, \dots, q_K^{(v)})^T = (q_{1,0} \mu_1^{(v)}, q_{2,0} \mu_2^{(v)}, \dots, q_{K,0} \mu_K^{(v)})^T;$$

• $S_n^{(v)}(\mathbf{q}^{(v)})$, $n = \overline{1, N}$, $v = \overline{1, V}$, – матрицы, определяющие интенсивности переходов процесса \mathbf{n}_t , $t \geq 0$, вызванных завершением обслуживания в сети запроса в случае наличия n запросов в сети при уровне v процесса поступления. Они вычисляются следующим образом:

$$S_n^{(v)}(\mathbf{q}^{(v)}) = S_n^{(K-1,v)}(\mathbf{q}^{(v)}),$$

где матрицы $S_n^{(K-1,v)}(\mathbf{q}^{(v)})$ находятся с использованием рекурсий

$$S_n^{(0,v)}(\mathbf{q}^{(v)}) = \min\{n, N_K\} q_K^{(v)}, \\ S_n^{(k,v)}(\mathbf{q}^{(v)}) = \begin{pmatrix} \min\{n, N_{K-k}\} q_{K-k}^{(v)} I & O & \cdots & O \\ S_1^{(k-1,v)}(\mathbf{q}^{(v)}) & \min\{n - 1, N_{K-k}\} q_{K-k}^{(v)} I & \cdots & O \\ O & S_2^{(k-1,v)}(\mathbf{q}^{(v)}) & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \min\{1, N_{K-k}\} q_{K-k}^{(v)} I \\ O & O & \cdots & S_n^{(k-1,v)}(\mathbf{q}^{(v)}) \end{pmatrix},$$

$$k = \overline{1, K-1}, n = \overline{1, N};$$

• $Q^{(v)}$, $v = \overline{1, V}$, – квадратные матрицы размера K со следующими элементами:

$$Q_{k,k}^{(v)} = -\mu_k^{(v)}, k = \overline{1, K}, \quad Q_{k,k'}^{(v)} = q_{k,k'} \mu_k^{(v)}, k, k' = \overline{1, K}, k' \neq k;$$

• $Q_j^{(v)}$, $j = \overline{1, K-2}$, $v = \overline{1, V}$, – матрицы, полученные путем удаления первых $K-2-j$ строк и столбцов из матрицы $Q^{(v)}$;

- $\Psi_n^{(v)}(Q^{(v)})$, $n = \overline{1, N}$, $v = \overline{1, V}$, – матрицы, задающие интенсивности переходов запросов между узлами сети при n запросах в ней и уровне v и определяемые как

$$\Psi_n^{(v)}(Q^{(v)}) = \Psi_n^{(K-2,v)},$$

где матрицы $\Psi_n^{(k,v)} = \Psi_n^{(k,v)}(Q^{(v)})$ вычисляются с использованием рекурсивных формул

$$\Psi_n^{(0,v)} = \text{diag}^+ \{ \min\{i, N_{K-1}\} Q_{K-1,K}^{(v)}, i = \overline{n, 1} \} +$$

$$+ \text{diag}^- \{ \min\{i, N_K\} Q_{K,K-1}^{(v)}, i = \overline{1, n} \}, n = \overline{1, N},$$

$$\Psi_n^{(k,v)} = \text{diag}\{O, \Psi_i^{(k-1,v)}, i = \overline{1, n}\} + \text{diag}^+ \{ \min\{n-i, N_{K-1-k}\} \Gamma_{i,m}^{(v)}, i = \overline{0, n-1} \} +$$

$$+ \text{diag}^- \{ \Omega_{i,k}^{(v)}, i = \overline{1, n} \}, n = \overline{1, N}, k = \overline{1, K-2}.$$

Здесь матрицы $\Omega_{n,j}^{(v)} = \Omega_{n,j}^{(v,j-2,v)}$, $n = \overline{1, N}$, и $\Gamma_{n,j}^{(v)} = \Gamma_{n,j}^{(v,j-2,v)}$, $n = \overline{0, N-1}$, $j = \overline{1, K-2}$, вычисляются следующим образом:

$$\Omega_{n,j}^{(0,v)} = \min\{n, N_K\} q_{v_j,1}^{(j,v)}, n = \overline{1, N}, j = \overline{1, K-2},$$

$$\Omega_{n,j}^{(k,v)} = \begin{pmatrix} \min\{n, N_{K-k}\} q_{v_j,1}^{(j,v)} I & O & \cdots & O \\ \Omega_{1,j}^{(k-1,v)} & \min\{n-1, N_{K-k}\} q_{v_{j-k},1}^{(j,v)} I & \cdots & O \\ O & \Omega_{2,j}^{(k-1,v)} & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \min\{1, N_{K-k}\} q_{v_{j-k},1}^{(j,v)} I \\ O & O & \cdots & \Omega_{n,j}^{(k-1,v)} \end{pmatrix},$$

$$k = \overline{1, v_j - 2}, n = \overline{1, N}, j = \overline{1, K-2},$$

$$\Gamma_{n,j}^{(0,v)} = q_{1,v_j}^{(j,v)}, n = \overline{0, N-1}, j = \overline{1, K-2},$$

$$\Gamma_{n,j}^{(k,v)} = \begin{pmatrix} q_{1,v_j-k}^{(j,v)} I & \Gamma_{0,j}^{(k-1,v)} & O & \cdots & O & O \\ O & q_{1,v_j-k}^{(j,v)} I & \Gamma_{1,j}^{(k-1,v)} & \cdots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \cdots & q_{1,v_j-k}^{(j,v)} I & \Gamma_{n,j}^{(k-1,v)} \end{pmatrix},$$

$$k = \overline{1, v_j - 2}, n = \overline{0, N-1}, j = \overline{1, K-2},$$

где $q_{k,k'}^{(j,v)} = (k, k')$ -й элемент матрицы $Q_j^{(v)}$, а v_j – количество строк матрицы $Q_j^{(v)}$;

- $\Delta_n^{(v)}$, $n = \overline{1, N}$, $v = \overline{1, V}$, – диагональные матрицы, диагональные элементы которых с точностью до знака определяют суммарные интенсивности выхода из соответствующего состояния в случае n запросов в сети при уровне v процесса прибытия. Очевидно, что эти матрицы можно найти с помощью выражения

$$\Delta_n^{(v)} = -\text{diag}\{\Psi_n^{(v)}(Q^{(v)})\mathbf{e} + S_n^{(v)}(\mathbf{q}^{(v)})\mathbf{e} + A_n(\mathbf{a})\mathbf{e}\}, n = \overline{1, N}.$$

Доказательство осуществляется путем анализа всех возможных переходов цепи Маркова $\psi_t, t \geq 0$, с последующей записью соответствующих интенсивностей в виде блоков генератора. Основная трудность здесь – определение матриц, задающих интенсивности переходов процесса $\pi_t, t \geq 0$. Алгоритмы для вычисления данных матриц основаны на результатах работы [6]. В этой работе интенсивности в узлах были фиксированы и не могли изменяться, что существенно облегчало исследование.

Определив вид генератора G , можно подсчитать векторы $\pi_n, n = \overline{0, N}$, стационарных вероятностей состояний цепи Маркова $\psi_t, t \geq 0$, как решение конечной системы линейных алгебраических уравнений (1). Количество уравнений этой системы может быть большим. Для ее решения следует использовать алгоритмы, учитывающие блочную трехдиагональную структуру генератора, например алгоритм, предложенный в работе [10].

Формулы для вычисления характеристик производительности. Вычислив векторы $\pi_n, n = \overline{0, N}$, можно подсчитать значения различных характеристик производительности сети и ее узлов.

Интенсивность выходного потока успешно обслуженных в сети запросов из k -го узла вычисляется как

$$\lambda_k^{succ} = \sum_{n=1}^N \sum_{v=1}^V \pi(n, v) (S_n^{(v)}(\mathbf{q}^{(v, k)})) \mathbf{e}, k = \overline{1, K},$$

где $\mathbf{q}^{(v, k)}$ – вектор-столбец размером K со всеми нулевыми элементами, кроме k -го элемента $(\mathbf{q}^{(v, k)})_k$, который равен $q_k^{(v)}$.

Среднее количество запросов, находящихся в k -м узле, вычисляется как

$$L_k = \sum_{n=1}^N \sum_{v=1}^V \pi(n, v) \Phi_n(\beta_k) \mathbf{e}, k = \overline{1, K},$$

где матрицы $\Phi_n(\beta) = \Phi_n^{(K-1)}(\beta), \beta \in \{\beta_1, \beta_2, \dots, \beta_K\}$, определяются с использованием рекурсии

$$\Phi_n^{(0)}(\beta) = n\beta_K, \\ \Phi_n^{(k)}(\beta) = \begin{pmatrix} n\beta_{K-k}I & O & \cdots & O \\ \Phi_1^{(k-1)}(\beta) & (n-1)\beta_{K-k}I & \cdots & O \\ O & \Phi_2^{(k-1)}(\beta) & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \beta_{K-k}I \\ O & O & \cdots & \Phi_n^{(k-1)}(\beta) \end{pmatrix}, k = \overline{1, K-1}, n = \overline{1, N}.$$

Кроме того, вычисляются:

– среднее количество занятых приборов в k -м узле, $k = \overline{1, K}$,

$$N_k^{\text{serv}} = \sum_{n=1}^N \sum_{v=1}^V \pi(n, v) S_n^{(v)}(\beta_k) \mathbf{e};$$

– среднее количество запросов в буфере k -го узла, $k = \overline{1, K}$,

$$L_k^{\text{buf}} = \sum_{n=1}^N \sum_{v=1}^V \pi(n, v) A_n(\beta_k) \mathbf{e} = L_k - N_k^{\text{serv}};$$

– вероятность потери произвольного запроса при прибытии в k -й узел из-за наличия N запросов в сети

$$P_k^{\text{arr-loss}} = \lambda^{-1} \sum_{v=1}^V \lambda_k^{(v)} \boldsymbol{\pi}(N, v) \mathbf{e}, k = \overline{1, K};$$

– вероятность потери произвольного запроса из-за нетерпеливости в k -м узле

$$P_k^{\text{imp-loss}} = \lambda^{-1} \sum_{n=1}^N \sum_{v=1}^V \boldsymbol{\pi}(n, v) A_n(\boldsymbol{\alpha}^{(k)}) \mathbf{e}, k = \overline{1, K},$$

где $\boldsymbol{\alpha}^{(k)}$ – векторы-столбцы размером K со всеми нулевыми элементами, кроме k -го элемента $(\boldsymbol{\alpha}^{(k)})_k$, который равен α_k ;

– вероятность потери произвольного запроса в k -м узле

$$P_k^{\text{loss}} = P_k^{\text{arr-loss}} + P_k^{\text{imp-loss}}, k = \overline{1, K};$$

– среднее количество запросов в сети

$$L = \sum_{n=1}^N n \boldsymbol{\pi}_n \mathbf{e};$$

– интенсивность выходного потока успешно обслуженных запросов в сети

$$\lambda^{\text{succ}} = \sum_{n=1}^N \sum_{v=1}^V \boldsymbol{\pi}(n, v) S_n^{(v)}(\boldsymbol{q}^{(v)}) \mathbf{e} = \sum_{k=1}^K \lambda_k^{\text{succ}};$$

– среднее количество занятых приборов в сети

$$N^{\text{serv}} = \sum_{k=1}^M N_k^{\text{serv}};$$

– среднее количество запросов во всех буферах сети

$$L^{\text{buf}} = \sum_{k=1}^M L_k^{\text{buf}};$$

– вероятность потери произвольного запроса по прибытии из-за наличия N запросов в сети

$$P^{\text{arr-loss}} = \lambda^{-1} \sum_{v=1}^V \lambda^{(v)} \boldsymbol{\pi}(N, v) \mathbf{e};$$

– вероятность потери произвольного запроса из-за нетерпеливости

$$P^{\text{imp-loss}} = \lambda^{-1} \sum_{n=1}^N \sum_{v=1}^V \boldsymbol{\pi}(n, v) A_n(\boldsymbol{\alpha}) \mathbf{e};$$

– вероятность потери произвольного запроса

$$P^{\text{loss}} = P^{\text{arr-loss}} + P^{\text{imp-loss}} = 1 - \frac{\lambda^{\text{succ}}}{\lambda};$$

– вероятность успешного обслуживания произвольного запроса в сети

$$P^{\text{succ}} = 1 - P^{\text{loss}} = \frac{\lambda^{\text{succ}}}{\lambda}.$$

Численный эксперимент. Приведем численный пример, в котором предполагается, что сеть состоит из $K = 3$ узлов и до $N = 25$ запросов могут находиться в сети одновременно. Количество приборов в узлах составляет $N_1 = 5, N_2 = 4$ и $N_3 = 3$.

Интенсивность нетерпеливости в буферах узлов задана как $\alpha_1 = 0,015, \alpha_2 = 0,008, \alpha_3 = 0,015$.

Процесс прибытия имеет $V = 2$ уровня, переходы между которыми определяются генератором $\Lambda = \begin{pmatrix} -0,002 & 0,002 \\ 0,005 & -0,005 \end{pmatrix}$.

Если процесс прибытия находится на уровне 1, интенсивности прибытия в узлы равны $\lambda_1^{(1)} = 1, \lambda_2^{(1)} = 2, \lambda_3^{(1)} = 0,5$ соответственно. Интенсивности прибытия в узлы сети при уровне 2 процесса прибытия равны $\lambda_1^{(2)} = 1,2, \lambda_2^{(2)} = 1, \lambda_3^{(2)} = 2,5$ соответственно.

Вероятности переходов запросов после завершения обслуживания в соответствующих узлах сети задаются выражениями

$$p_{1,2} = 0,2, \quad p_{1,3} = 0,3, \quad p_{2,1} = 0,4, \quad p_{2,3} = 0,1, \quad p_{3,1} = 0,2, \quad p_{3,2} = 0,1.$$

Предположим, что общая пропускная способность, предоставляемая для передачи информации в сети, составляет $\tilde{B} = 120$ Мбит/с, а запросы идентичны и имеют средний размер $\tilde{v} = 8$ Мбит. При уровне 1 процесса прибытия пропускная способность делится между узлами следующим образом: пропускная способность узла 1 составляет $B_1^{(1)} = 45$ Мбит/с, узла 2 – $B_2^{(1)} = 60$ Мбит/с, а узла 3 – $B_3^{(1)} = 15$ Мбит/с. Таким образом, интенсивности обслуживания приборами в узлах сети при уровне 1 входного процесса определяются как

$$\mu_k^{(1)} = \frac{B_k^{(1)}}{\frac{N_k}{\tilde{v}}}, \quad k = 1, 2, 3, \quad \text{и равны } \mu_1^{(1)} = 9/8, \mu_2^{(1)} = 15/8, \mu_3^{(1)} = 5/8.$$

Целью эксперимента является определение того, как пропускная способность должна оптимально перераспределяться между узлами, когда процесс прибытия переходит на уровень 2. Предполагается, что минимальная пропускная способность, предоставляемая каждому узлу, составляет 5 Мбит/с. Для нахождения оптимального перераспределения пропускной способности при переходе на уровень 2 входного процесса будем изменять пропускную способность узла 1 $B_1^{(2)}$ на интервале $[5, \tilde{B} - 10]$ и пропускную способность узла 2 $B_2^{(2)}$ на интервале $[5, \tilde{B} - 5 - B_1^{(2)}]$ с шагом 5 Мбит/с и исследовать зависимость основных показателей производительности сети от этих параметров. Отметим, что при фиксированной пропускной способности узлов 1 и 2 пропускная способность узла 3 рассчитывается как $B_3^{(2)} = \tilde{B} - B_1^{(2)} - B_2^{(2)}$. Интенсивности обслуживания при уровне 2 входного процесса рассчитываются таким же

образом, как и при уровне 1 входного процесса, и равны $\mu_k^{(2)} = \frac{B_k^{(2)}}{\frac{N_k}{\tilde{v}}}, \quad k = 1, 2, 3$.

На рис. 2 показана зависимость среднего количества L запросов в сети от параметров $B_1^{(2)}$ и $B_2^{(2)}$.

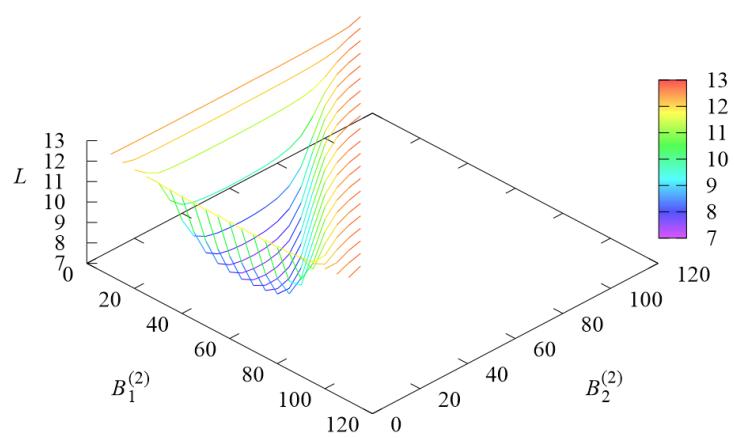


Рис. 2. Зависимость L от $B_1^{(2)}$ и $B_2^{(2)}$
Fig. 2. The dependence of L on $B_1^{(2)}$ and $B_2^{(2)}$

Рис. 3–5 иллюстрируют зависимость вероятностей потерь P_k^{loss} , $k = 1, 2, 3$, произвольного запроса в k -м узле от параметров $B_1^{(2)}$ и $B_2^{(2)}$.

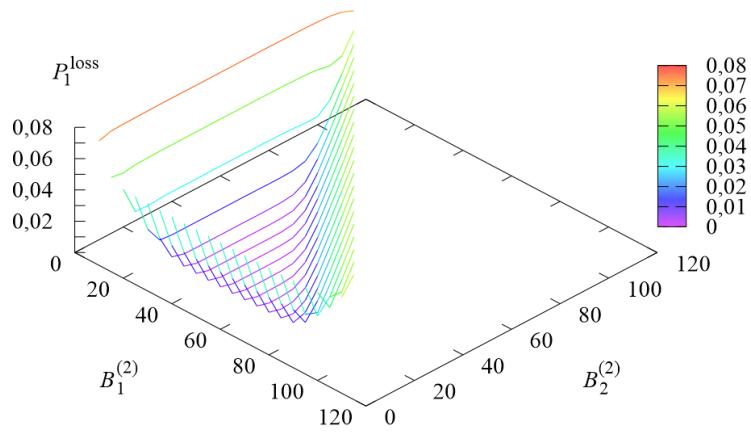


Рис. 3. Зависимость P_1^{loss} от $B_1^{(2)}$ и $B_2^{(2)}$
Fig. 3. The dependence of P_1^{loss} on $B_1^{(2)}$ and $B_2^{(2)}$

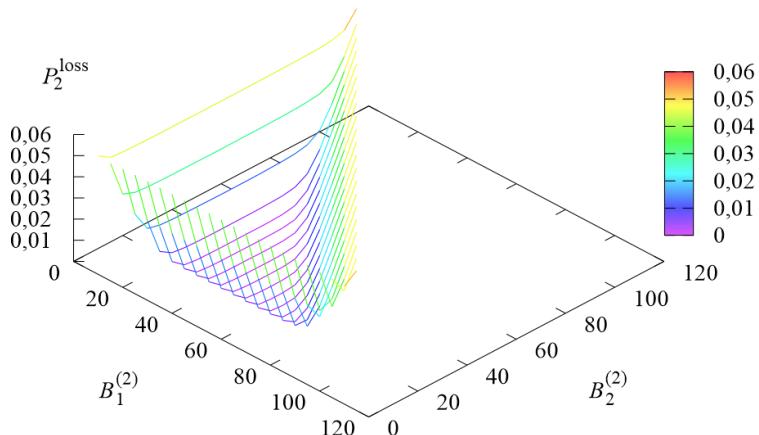


Рис. 4. Зависимость P_2^{loss} от $B_1^{(2)}$ и $B_2^{(2)}$
Fig. 4. The dependence of P_2^{loss} on $B_1^{(2)}$ and $B_2^{(2)}$

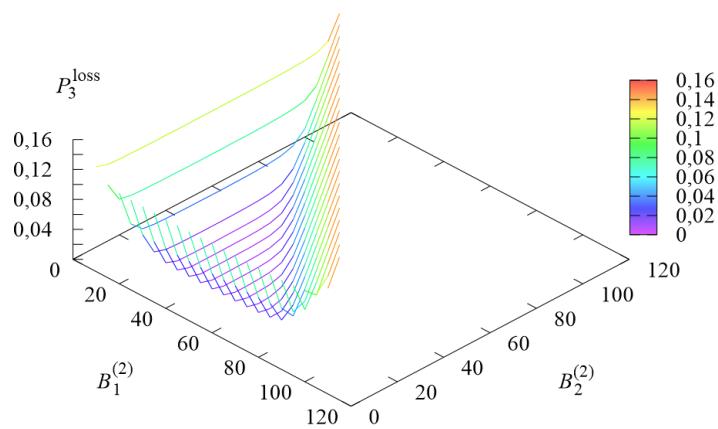


Рис. 5. Зависимость P_3^{loss} от $B_1^{(2)}$ и $B_2^{(2)}$
Fig. 5. The dependence of P_3^{loss} on $B_1^{(2)}$ and $B_2^{(2)}$

На рис. 6–8 изображена зависимость вероятности потери $P^{\text{arr-loss}}$ произвольного запроса по прибытии, вероятности потери $P^{\text{imp-loss}}$ произвольного запроса из-за нетерпеливости и вероятности потери P^{loss} произвольного запроса от параметров $B_1^{(2)}$ и $B_2^{(2)}$.

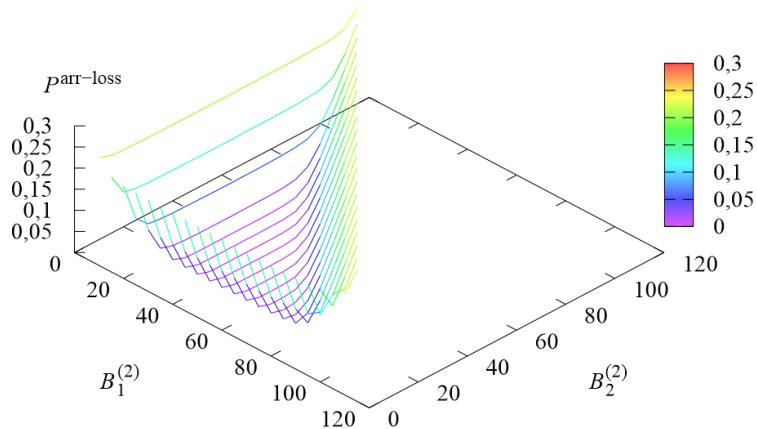


Рис. 6. Зависимость $P^{\text{arr-loss}}$ от $B_1^{(2)}$ и $B_2^{(2)}$
Fig. 6. The dependence of $P^{\text{arr-loss}}$ on $B_1^{(2)}$ and $B_2^{(2)}$

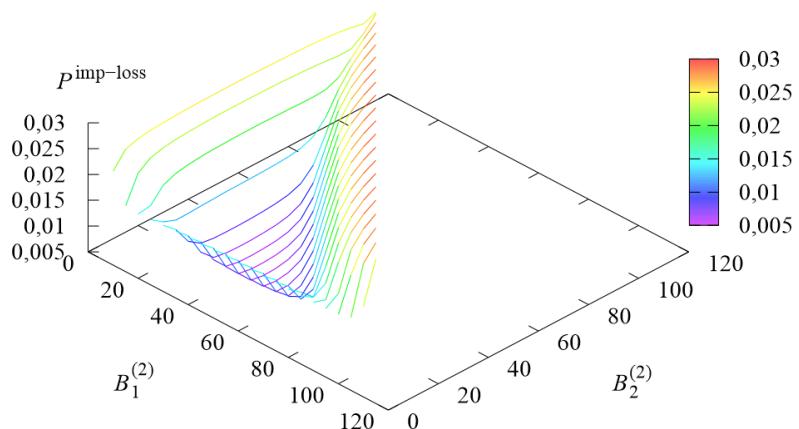


Рис. 7. Зависимость $P^{\text{imp-loss}}$ от $B_1^{(2)}$ и $B_2^{(2)}$
Fig. 7. The dependence of $P^{\text{imp-loss}}$ on $B_1^{(2)}$ and $B_2^{(2)}$

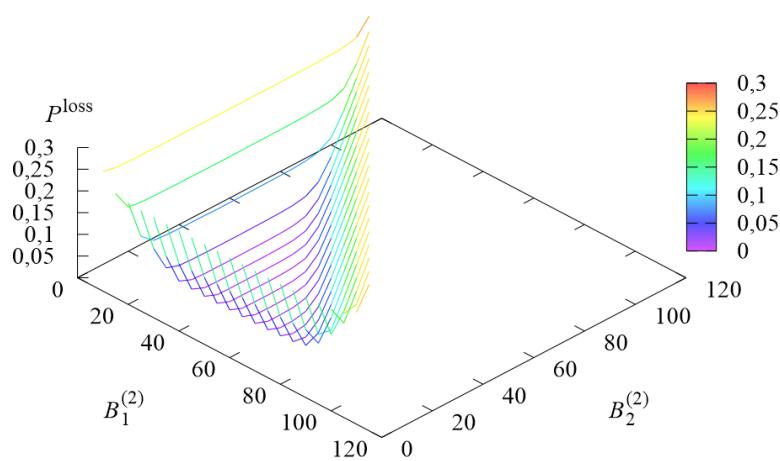


Рис. 8. Зависимость P^{loss} от $B_1^{(2)}$ и $B_2^{(2)}$
Fig. 8. The dependence of P^{loss} on $B_1^{(2)}$ and $B_2^{(2)}$

Как видно из приведенных рисунков, основные характеристики системы существенно зависят от параметров $B_1^{(2)}$ и $B_2^{(2)}$. При слишком больших или слишком маленьких значениях пропускной способности в узлах число запросов в сети, а также вероятности потери запросов принимают большие значения. Поэтому очевидно, что задача перераспределения ресурсов сети при изменении интенсивностей входного потока является важной. Для определения оптимальных значений параметров $B_1^{(2)}$ и $B_2^{(2)}$ необходимо выбрать функционал, описывающий качество функционирования системы.

Предположим, что качество работы сети определяется следующим стоимостным критерием:

$$E = E(B_1^{(2)}, B_2^{(2)}) = a_1 \lambda^{\text{succ}} - b_1 \lambda P^{\text{arr-loss}} - b_2 \lambda P^{\text{imp-loss}},$$

где a_1 – прибыль, получаемая сетью от обслуживания одного запроса; b_1 – штраф за потерю произвольного запроса при поступлении, а b_2 – штраф за потерю произвольного запроса из-за нетерпеливости. Таким образом, стоимостный критерий E определяет прибыль системы в единицу времени. В данном эксперименте зафиксируем следующие значения стоимостных коэффициентов: $a_1 = 2$, $b_1 = 3$, $b_2 = 10$.

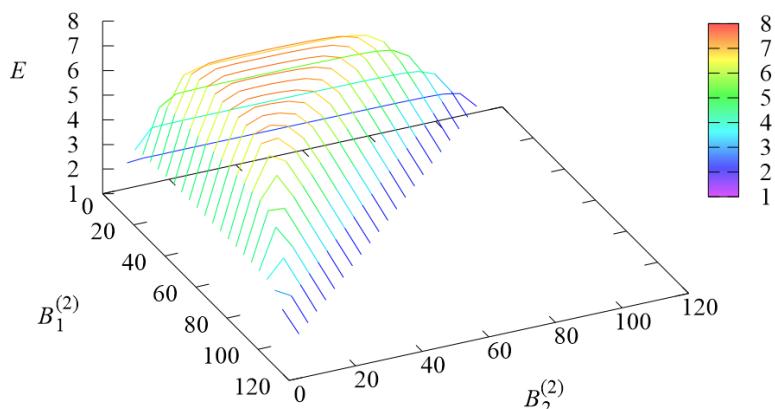


Рис. 9. Зависимость E от $B_1^{(2)}$ и $B_2^{(2)}$
Fig. 9. The dependence of E on $B_1^{(2)}$ and $B_2^{(2)}$

На рис. 9 показана зависимость значений критерия E от параметров $B_1^{(2)}$ и $B_2^{(2)}$. Видно, что при выборе слишком малых и слишком больших значений $B_1^{(2)}$ и $B_2^{(2)}$ прибыль сети за обслуживание запросов существенно уменьшается. Оптимальное значение критерия E равно $E^* = 7,42844$ и достигается при $B_1^{(2)} = 40, B_2^{(2)} = 30$ и, соответственно, $B_3^{(2)} = \tilde{B} - B_1^{(2)} - B_2^{(2)} = 50$. Существенное перераспределение пропускной способности сети в пользу третьего узла легко объясняется пятикратным увеличением интенсивности потока, поступающего в этот узел, при переключении входного потока с уровня 1 на уровень 2. Отметим, что если перераспределение пропускной способности не осуществлять, т. е. оставить значения параметров $B_1^{(2)}$ и $B_2^{(2)}$ равными соответственно $B_1^{(1)} = 45$ и $B_2^{(1)} = 60$, то значение критерия E будет равняться 4,0833, что существенно ниже оптимального значения. Также отметим, что при фиксированных оптимальных значениях $B_1^{(2)} = 40, B_2^{(2)} = 30$ вероятность потери произвольного запроса P^{loss} равна 0,00575, в то время как при отсутствии перераспределения пропускной способности вероятность P^{loss} составляет 0,15118, что более чем в 26 раз больше. Таким образом, перераспределение пропускной способности между узлами имеет смысл.

Вычисления, результаты которых приведены выше, проводились на компьютере с процессором Intel Core i7-8700 3,2 ГГц, 16 ГБ RAM с использованием Wolfram Mathematica 13.2. Время вычисления составило 923 с для расчета 253 точек при максимальном размере блоков генератора, равном 702. В среднем на одну точку тратилось 3,65 с.

Заключение. В статье рассмотрена задача расчета характеристик полуоткрытой сети массового обслуживания, в которой входной поток в узлы имеет несколько возможных уровней, переключение между ними управляет цепью Маркова с конечным пространством состояний. При изменении уровня потока есть возможность изменять и скорости обслуживания запросов в узлах сети. При фиксированных значениях параметров сети ее поведение описано многомерной цепью Маркова. Получены формулы для вычисления генератора этой цепи, имеющего блочную трехдиагональную структуру, и формулы для вычисления основных характеристик производительности сети через векторы стационарных вероятностей состояний цепи Маркова. Приведен численный пример, иллюстрирующий реализуемость предложенных алгоритмов для расчета, а также поведение некоторых характеристик производительности сети. Показана возможность решения задачи оптимизации перераспределения ресурсов сети при изменении уровня входного потока.

В дальнейшем планируется обобщение полученных результатов на случаи более сложного поведения процесса поступления запросов при каждом уровне, рассмотрение зависимости от уровня потока не только скорости обслуживания в узлах сети, но и числа приборов в узлах.

Вклад авторов. *А. Н. Дудин* – построение модели, аналитическое исследование, подготовка текста статьи. *О. С. Дудина и С. А. Дудин* – аналитическое исследование, написание текста, программирование алгоритмов и тестирование программ, проведение и оформление численных экспериментов.

References

1. Roy D. Semi-open queueing networks: a review of stochastic models, solution methods and new research areas. *International Journal of Production Research*, 2016, vol. 54, no. 6, pp. 1735–1752.
2. Otten S., Krenzler R., Xie L., Daduna H., Kruse K. Analysis of semi-open queueing networks using lost customers approximation with an application to robotic mobile fulfilment systems. *OR Spectrum*, 2022, vol. 44, no. 2, pp. 603–648.
3. Jia J., Heragu S. S. Solving semi-open queueing networks. *Operations Research*, 2009, vol. 57, no. 2, pp. 391–401.
4. Kim J., Dudin A., Dudin S., Kim C. Analysis of a semi-open queueing network with Markovian arrival process. *Performance Evaluation*, 2018, vol. 120, pp. 1–19.

5. Kim C., Dudin S., Dudin A., Samouylov K. Analysis of a semi-open queueing network with a state dependent marked Markovian arrival process, customers retrials and impatience. *Mathematics*, 2019, vol. 7, no. 8, p. 715.
6. Dudin S., Dudin A., Manzo R., Rarità L. Analysis of semi-open queueing network with correlated arrival process and multi-server nodes. *Operations Research Forum*. Cham, Springer International Publishing, 2024, vol. 5, no. 4, p. 99.
7. Amjath M., Kerbache L., Elomri A., MacGregor Smith J. Queueing network models for the analysis and optimisation of material handling systems: a systematic literature review. *Flexible Services and Manufacturing Journal*, 2024, vol. 36, no. 2, pp. 668–709.
8. D'Apice C., Dudin A., Dudin S., Manzo R. Study of a semi-open queueing network with hysteresis control of service regimes. *AIMS Mathematics*, 2025, vol. 10, no. 2, pp. 3095–3123.
9. Graham A. *Kronecker Products and Matrix Calculus with Applications*. Courier Dover Publications, 2018, 144 p.
10. Baumann H., Sandmann W. Numerical solution of level dependent quasi-birth-and-death processes. *Procedia Computer Science*, 2010, vol. 1, no. 1, pp. 1561–1569.

Информация об авторах

Дудин Александр Николаевич, доктор физико-математических наук, профессор, заведующий НИЛ прикладного вероятностного анализа, Белорусский государственный университет.

E-mail: dudin@bsu.by

Дудина Ольга Сергеевна, кандидат физико-математических наук, ведущий научный сотрудник НИЛ прикладного вероятностного анализа, Белорусский государственный университет.

E-mail: dudina@bsu.by

Дудин Сергей Александрович, кандидат физико-математических наук, ведущий научный сотрудник НИЛ прикладного вероятностного анализа, Белорусский государственный университет.

E-mail: dudins@bsu.by

Information about the authors

Alexander N. Dudin, D. Sc. (Phys.-Math.), Prof., Head of the Research Laboratory of Applied Probability Analysis, Belarusian State University.
E-mail: dudin@bsu.by

Olga S. Dudina, Ph. D. (Phys.-Math.), Leading Researcher of the Research Laboratory of Applied Probability Analysis, Belarusian State University.
E-mail: dudina@bsu.by

Sergei A. Dudin, Ph. D. (Phys.-Math.), Leading Researcher of the Research Laboratory of Applied Probability Analysis, Belarusian State University.
E-mail: dudins@bsu.by