

АНАЛИЗ СТОХАСТИЧЕСКИХ ТАНДЕМНЫХ СИСТЕМ С ПОВТОРНЫМИ ВЫЗОВАМИ

DOI: 10.36724/2072-8735-2026-20-6-20-29

Manuscript received 21 March 2026;

Accepted 25 May 2026

Вишневский Владимир Миронович,
Институт проблем управления им. В. А. Трапезникова РАН,
Москва, Россия, vishn@inbox.ru

Семёнова Ольга Валерьевна,
Институт проблем управления им. В. А. Трапезникова РАН,
Москва, Россия, olga.semenova@ipu.ru

Ключевые слова: сети массового обслуживания, тандемная система, повторные вызовы (орбита), МАР-поток, распределение фазового типа, имитационное моделирование, машинное обучение

В настоящее время одним из важных направлений проектирования и оценки производительности современных телекоммуникационных систем и сетей является анализ математических моделей тандемных систем. Тандемные системы являются математическими моделями различных технических и социальных систем, в частности, являются адекватными моделями широкополосных беспроводных сетей вдоль протяженных транспортных магистралей (автомобильные и железные дороги, нефтяные и газовые трубопроводы и т.д.). Рассматриваемая в настоящей работе тандемная система включает произвольное, но конечное число обслуживающих устройств (фаз) с ограниченным буфером на каждой фазе. Поток заявок (пакетов) на первую фазу является коррелированным МАР-поток, а случайное время обслуживания описывается фазовым распределением. Если при поступлении заявки на любую фазу буферная память этой фазы заполнена, то заявка отправляется на орбиту (общую память для всех фаз), из которой через случайное время повторно поступает на первую фазу. Оцениваются основные характеристики производительности системы, включая межконцевую задержку (время с момента поступления на первую фазу до момента успешного завершения обслуживания на последней фазе), средние длины очередей на фазах системы и на орбите. Для систем с двумя очередями получены точные формулы для расчета требуемых характеристик. Для систем с произвольным числом фаз использован метод машинного обучения в сочетании с методом имитационного моделирования. Проведены многочисленные численные эксперименты.

Информация об авторах:

Вишневский Владимир Миронович, доктор технических наук, профессор, главный научный сотрудник, лаборатория "Телекоммуникационных систем", Институт проблем управления им. В. А. Трапезникова РАН, Москва, Россия. ORCID 0000-0001-7373-4847

Семёнова Ольга Валерьевна, доктор физико-математических наук, ведущий научный сотрудник, лаборатория "Телекоммуникационных систем", Институт проблем управления им. В. А. Трапезникова РАН, Москва, Россия. ORCID 0000-0001-7846-6212

Для цитирования:

Вишневский В.М., Семёнова О.В. Анализ стохастических тандемных систем с повторными вызовами // Т-Comm: Телекоммуникации и транспорт. 2026. Том 20. №6. С. 20-29.

For citation:

V.M. Vishnevsky, O.V. Semenova, "Analysis of stochastic tandem systems with retrials", T-Comm, 2026, vol. 20, no.6, pp. 20-29. (in Russian)

Введение

Для моделирования современных систем связи транспортных и других систем широко используются стохастические тандемные системы (многофазные системы массового обслуживания) см., например [1-3]. Такие модели представляют особый интерес для оценки характеристик производительности широкополосных беспроводных сетей вдоль протяженных транспортных магистралей (автомобильные и железные дороги, нефтяные и газовые трубопроводы и т.д.) [4-5]. В связи с этим, исследование тандемных систем, начатое в середине прошлого века, интенсивно продолжается до настоящего времени [6-10]. Следует отметить, что большинство работ посвящено исследованию лишь двухфазных тандемных систем, хотя наибольший практический интерес представляют тандемные системы большой размерности (с числом узлов больше двух). Точное аналитическое решение для тандемных систем большой размерности, особенно в случае коррелированного входного потока и произвольной функции распределения времени обслуживания на фазах не было получено. Разработаны лишь многочисленные приближенные методы оценки производительности таких систем [9-10].

Заметим также, что в мировой литературе, посвященной исследованию стохастических тандемных систем, слабо освещены модели тандемных систем с повторными вызовами. В то же время, механизм повторных заявок является неотъемлемой характеристикой большинства современных телекоммуникационных сетей и систем (например, механизм повторной передачи протокола TCP – Transmission Control Protocol). Данная особенность значительно усложняет математический анализ соответствующих моделей, в отличие от моделей без повторных заявок, поскольку марковские цепи, описывающие поведение систем с повторными заявками, являются пространственно неоднородными. В то же время их изучение важно не только для практических приложений, но и представляет значительный теоретический интерес.

Обзоры ранних (до 2016 года) работ по математическому анализу систем массового обслуживания с повторными заявками можно найти в [11]. Тандемные системы с повторными вызовами исследовались в [1, 12-14]. В большинстве данных работ исследовались двухфазные системы, и предполагалась возможность присоединения к орбите только заявок, заблокированных на первой фазе. Хотя в различных реальных протоколах в вычислительных сетях с гарантированной доставкой пакетов разработан механизм повторной передачи пакетов из общей памяти (орбиты). Поэтому рассматриваемая в данной статье задача анализа характеристик тандемной системы большой размерности с общей орбитой является актуальной. Подобные тандемные системы с общей орбитой рассматривались в работах [13, 14].

В работе [13] для тандемной системы с произвольным конечным числом фаз (станций) предложен метод приближенного расчета среднего времени пребывания заявки в системе и среднего числа посещений орбиты. Для рассмотренной модели не накладывается ограничений на распределение между моментами поступления заявок, времени обслуживания и интервалов между повторными попытками заявки обслужиться. Предполагается, что известна лишь вероятность блокировки на каждой фазе. Исходя из результатов, полученных для двухфазной системы, авторы делают вывод о том, что

приближенный метод дает удовлетворительные результаты при относительно низкой интенсивности входного потока.

В статье [14] исследована тандемная система с двумя фазами без буферов и общей орбитой для повторных вызовов с произвольным распределением времени обслуживания на первой фазе и экспоненциальным – на второй. Время между повторными вызовами также распределено по экспоненциальному закону, при этом используется классическая стратегия повторных попыток. Предполагая, что интенсивность повторных вызовов довольно мала, авторы показывают, что масштабированное число заявок на орбите асимптотически близко к диффузионному процессу, что далее используется для приближенной оценки числа заявок на орбите в стационарном режиме. В статье [1] исследуется двухфазная система с орбитой на первой фазе и отдыхами обоих обслуживающих устройств по завершению обслуживания всех заявок в системе.

При проектировании и оценке характеристик современных телекоммуникационных сетей особенно важно учитывать нестационарный и коррелированный характер циркулирующих в них потоков данных. При моделировании учет этих особенностей возможен с использованием математических моделей потоков типа BMAP (Batch Markovian Arrival Process) и MAP (Markovian Arrival Process) [15], которые адекватно описывают потоки в современных вычислительных сетях.

В настоящей статье представлен анализ тандемной системы большой размерности (с числом фаз $N \geq 2$) с общей орбитой, имеющей, в отличие от моделей, исследованных в перечисленных выше работах, более общее предположение о характере входного коррелированного потока и обобщенное фазовое распределение времени обслуживания на фазах системы. Для прогнозирования характеристик производительности тандемной системы большой размерности впервые использована комбинация методов имитационного моделирования и машинного обучения [16].

Статья организована следующим образом. Раздел 1 содержит постановку задачи и описание математической модели. В разделе 2 приведен анализ двухфазной тандемной системы и представлен алгоритм точного расчета характеристик производительности системы. В разделе 3 для системы с произвольным числом фаз впервые применена комбинация методов машинного обучения и имитационного моделирования для оценки межконцевой задержки и других характеристик. Также проведен сравнительный анализ применения различных методов машинного обучения. Раздел 4 содержит заключение к статье.

1 Постановка задачи и описание математической модели

Рассматривается стохастическая тандемная система, состоящая из $N \geq 2$ фаз, представленных системами массового обслуживания типа MAP/PH/1/K, и общей орбитой (рис. 1).

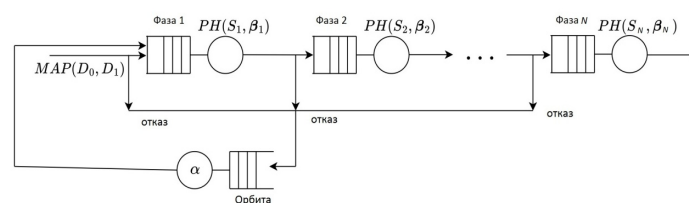


Рис. 1. Стохастическая тандемная система с общей орбитой

Число мест для ожидания в буфере на фазе i ограничено числом K_i , $i = \overline{1, N}$. Заявки поступают извне на первую фазу согласно марковскому входному потоку (МАР-потоку), который описывается матрицами D_0 и D_1 и средней интенсивностью поступления заявок $\lambda = \theta D_1 e$. В последнем равенстве вектор-строка θ находится как единственное решение системы уравнений $\theta(D_0 + D_1) = 0, \theta e = 1$, где $e = (1, 1, \dots, 1)^T$, $0 = (0, \dots, 0)$.

Время обслуживания заявки на k -й фазе имеет распределение фазового типа (РН, Phase-type), которое задается парой (β_k, S_k) , $k = \overline{1, N}$, где $\beta_k = (\beta_k^{(1)}, \dots, \beta_k^{(M_k)})$ – стохастический вектор-строка, а S_k – субгенератор. Интенсивность обслуживания определяется как $\mu_k = -[\beta_k S_k^{-1} e]^{-1}$. Среднее время обслуживания – $b_k = \mu_k^{-1}$, $k = \overline{1, N}$. Более подробное описание МАР-потока как частного случая группового потока ВМАР и фазового распределения и их свойства можно найти в [15].

Если при поступлении заявки на i -ю фазу свободные места для ожидания в буфере отсутствуют, заявка переходит на орбиту для повторного обслуживания на первой фазе. Время пребывания повторной заявки на орбите экспоненциально распределено с параметром α_i , где i – число заявок на орбите. Предполагается, что $\alpha_0 = 0$, $\alpha_i \rightarrow \infty$ при $i \rightarrow \infty$. Такая зависимость включает, в частности, линейную стратегию повторных вызовов $\alpha_i = i\alpha$, $\alpha > 0$.

Задача состоит в отыскании основных характеристик производительности, включая среднее время пребывания заявок в системе (межконцевая задержка) и среднее число заявок на каждой фазе и орбите.

2 Анализ двухфазной тандемной системы с повторными вызовами и общей орбитой

В данном разделе рассмотрим частный случай тандемной системы, описанной в предыдущем разделе, с количеством фаз $N=2$. Ее функционирование может быть описано случайным процессом, принадлежащим классу асимптотически-квазитеплицевых цепей Маркова [15].

Пусть в момент времени t :

- i_t – число заявок на орбите, $i \geq 0$,
- j_t – число заявок на первой фазе, $j_t = \overline{0, K_1}$,
- n_t – число заявок на второй фазе, $n_t = \overline{0, K_2}$,
- v_t – состояние управляющего процесса МАР $v_t = \overline{0, W}$,
- $m_t^{(k)}$ – состояние управляющего процесса РН на k -й фазе, $m_t^{(k)} = \overline{1, M^{(k)}}$, $k = \overline{1, 2}$.

Случайный процесс $\xi_t = \{i_t, j_t, n_t, v_t, m_t^{(1)}, m_t^{(2)}\}$, $t \geq 0$ является регулярной неприводимой цепью Маркова с пространством состояний

$$\Omega = \{(i, j, n, v), i \geq 0, j = 0, n = 0, v = \overline{0, W}\} \cup \{(i, j, n, v, m^{(1)}), i \geq 0, j = \overline{0, K_1}, n = 0, v = \overline{0, W}, m^{(1)} = \overline{1, M^{(1)}}\} \cup \{(i, j, n, v, m^{(2)}), i \geq 0, j = 0, n = \overline{0, K_2}, v = \overline{0, W}, m^{(2)} = \overline{1, M^{(2)}}\} \cup$$

$$\{(i, j, n, v, m^{(1)}, m^{(2)}), i \geq 0, j = \overline{0, K_1}, n = \overline{0, K_2}, v = \overline{0, W}, m^{(1)} = \overline{1, M^{(1)}}, m^{(2)} = \overline{1, M^{(2)}}\}.$$

Упорядочим состояния из множества Ω в лексикографическом порядке согласно возрастанию значений счетной компоненты i_t . Интенсивности переходов между состояниями формируют инфинитезимальный генератор Q , структура которого будет описана далее.

Пусть I – единичная матрица, O – нулевая матрица, $\delta(i, j)$ – символ Кронекера, $\oplus$ и $\otimes$ – символы Кронекеровой суммы и произведения матриц, соответственно. Обозначим через $a = (a_1, \dots, a_n)$ и будем использовать следующие вспомогательные матрицы: $\text{diag}\{a\}$ – диагональная матрица размера $n \times n$ с диагональными элементами, задаваемыми вектором a , $\text{diag}^+(\mathbf{a})$ ($\text{diag}^-(\mathbf{a})$) – матрица размера $n \times n$, имеющая ненулевые элементы с индексами $(i, i+1)$, равные a_i для $i = \overline{1, n-1}$ (соответственно, с индексами $(i-1, i)$, равные a_i для $i = \overline{2, n}$).

Лемма 1. Инфинитезимальный генератор Q цепи Маркова ξ_t , $t \geq 0$ имеет блочную трехдиагональную структуру

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & O & O & \dots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & O & O & \dots \\ O & Q_{2,1} & Q_{2,2} & Q_{2,3} & O & \dots \\ O & O & Q_{3,2} & Q_{3,3} & Q_{3,4} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (1)$$

где матрицы $Q_{i,j}$ также имеют блочную структуру вида $Q_{i,j} = (Q_{i,i'}(j, j'))_{i' \geq 0, j' = \overline{0, K_1}}$ с матрицами $Q_{i,i'}(j, j')$, определяемыми равенствами

$$Q_{i,i-1}(j, j-1) = O, i \geq 1, j = \overline{1, K_1}, \quad Q_{i,i-1}(j, j) = O, i \geq 1, j = \overline{0, K_1}, \\ Q_{i,i-1}(0, 1) = \alpha_i \text{diag}\{I_W \otimes \beta_1, I_W \otimes \beta_1 \otimes I_{M_2}\}_{K_2}, i \geq 1, j = \overline{0, K_1-1},$$

$$Q_{i,i-1}(j, j+1) = \alpha_i I_{WM_1(1+K_2M_2)}, \quad i \geq 1, j = \overline{1, K_1-1},$$

$$Q_{i,i}(1, 0) = \text{diag}^+\{I_W \otimes S_0^{(1)} \otimes \beta_2, I_W \otimes S_0^{(1)} \otimes I_{M_2}\}_{K_2-1}, i \geq 0,$$

$$Q_{i,i}(j, j-1) = \text{diag}^+\{I_W \otimes S_0^{(1)} \beta_1 \otimes \beta_2, I_W \otimes S_0^{(1)} \beta_1 \otimes I_{M_2}\}_{K_2-1},$$

$$i \geq 0, j = \overline{2, K_1},$$

$$Q_{i,i}(0, 0) = \text{diag}\{D_0, D_0 \oplus S_2\}_{K_2} + \text{diag}^-\{I_W \otimes S_0^{(2)}, I_W \otimes S_0^{(2)} \beta_2\}_{K_2-1} - \alpha_i I, i \geq 0,$$

$$Q_{i,i}(j, j) = \text{diag}\{D_0 \oplus S_1, D_0 \oplus S_1 \oplus S_2\}_{K_2} - [1 - \delta(j, K_1)] \alpha_i I + \\ + \text{diag}^-\{I_{WM_1} \otimes S_0^{(2)}, I_{WM_1} \otimes S_0^{(2)} \beta_2\}_{K_2-1}, i \geq 0, j = \overline{1, K_1},$$

$$Q_{i,i}(0, 1) = \text{diag}\{D_1 \otimes \beta_1, D_1 \otimes \beta_1 \otimes I_{M_2}\}_{K_2}, \quad i \geq 0,$$

$$\begin{aligned}
Q_{i,i}(j, j+1) &= \text{diag}\{D_1 \otimes I_{M_1}, \underbrace{D_1 \otimes I_{M_1} \otimes I_{M_2}}_{K_2}\}, i \geq 0, j = \overline{1, K_1 - 1}, \\
Q_{i,i+1}(1, 0) &= \text{diag}\{O_W \otimes S_0^{(1)}, \underbrace{O_W \otimes S_0^{(1)} \otimes I_{M_2}}_{K_2}, I_W \otimes S_0^{(1)} \otimes I_{M_2}\}, i \geq 0, \\
Q_{i,i+1}(j, j-1) &= \text{diag}\{O_{WM_1(1+(K_2-1)M_2)}, I_W \otimes S_0^{(1)} \beta_1 \otimes I_{M_2}\}, \\
&\quad i \geq 0, j = \overline{2, K_1}, \\
Q_{i,i+1}(j, j) &= O, i \geq 0, j = \overline{1, K_1 - 1}, \\
Q_{i,i+1}(K_1, K_1) &= \text{diag}\{D_1 \otimes I_{M_1}, \underbrace{D_1 \otimes I_{M_1 M_2}}_{K_2}\}, i \geq 0, \\
Q_{i,i+1}(j, j+1) &= O, i \geq 0, j = \overline{1, K_1 - 1}.
\end{aligned}$$

Остальные матрицы $Q_{i,i'}(j, j')$, $i' \geq 0, j' = \overline{0, K_1}$, не перечисленные выше, являются нулевыми, $S_0^{(k)} = -S_k e$.

Доказательство. Блочная структура вида (1) следует из анализа возможных переходов цепи Маркова $\xi_t, t \geq 0$ между состояниями. С целью сокращения изложения опишем лишь характеристику этих блоков.

Блоки $Q_{i,i-1}(j, j+1)$ описывают интенсивности переходов цепи Маркова $\xi_t, t \geq 0$, которые сопровождаются успешным переходом заявки с орбиты на первую фазу. Если первая фаза пуста ($j=0$), эти переходы сопровождаются выбором начальной фазы процесса обслуживания на первой фазе в соответствии с вектором вероятностей β_1 .

Блоки $Q_{i,i}(j, j')$ описывают интенсивности переходов цепи $\xi_t, t \geq 0$, которые не приводят к изменению числа заявок на орбите. Пусть $j' = j-1$. Если $j=1$, такие переходы сопровождаются завершением обслуживания заявки на первой фазе (интенсивности переходов его задаются матрицей $S_0^{(1)}$) и переходом этой заявки на вторую фазу. В случае $j > 1$ очередная заявка поступает на обслуживание на первой фазе из буфера. В случае $j' = j$ вид блочных матриц $Q_{i,i}(j, j)$ зависит от значения j и устанавливается путем анализа возможных переходов процесса $\xi_t, t \geq 0$, которые не сопровождаются событиями, изменяющими значение j .

Матрицы $Q_{i,i+1}(j, j')$ описывают интенсивности переходов цепи Маркова $\xi_t, t \geq 0$, которые приводят к увеличению числа заявок на орбите на единицу в случае поступления заявки извне, когда нет мест для ожиданий на первой фазе.

Матрицы $Q_{i,i+1}(j, j-1)$, $j = \overline{1, K_1}$, описывают интенсивности переходов процесса $\xi_t, t \geq 0$ в случае, когда нет мест для ожидания на второй фазе в момент выхода заявки с первой фазы (в момент завершения ее обслуживания).

Следствие 1. Цепь Маркова $\xi_t, t \geq 0$ является асимптотически квазитеплицевой цепью Маркова [15].

Доказательство. Пусть

$$T_i = \text{diag}\left\{\left|(Q_{i,i})_{1,1}\right|, \left|(Q_{i,i})_{2,2}\right|, \dots, \left|(Q_{i,i})_{m,m}\right|\right\},$$

где m – размерность матрицы $Q_{i,i}, i \geq 0$. Покажем, что пределы

$$Y_0 = \lim_{i \rightarrow \infty} T_i^{-1} Q_{i,i-1}, Y_1 = \lim_{i \rightarrow \infty} T_i^{-1} Q_{i,i} + I, Y_2 = \lim_{i \rightarrow \infty} T_i^{-1} Q_{i,i+1}$$

существуют, а матрица $Y_0 + Y_1 + Y_2$ стохастическая. Заметим, что матрица T_i , а также $Q_{i,i}(K_1, K_1 - 1), Q_{i,i}(K_1, K_1), Q_{i,i+1}(K_1, K_1), Q_{i,i+1}(K_1, K_1 - 1)$ не зависят от i и K_1 , поэтому можно ввести обозначения:

$$\begin{aligned}
A &= Q_{i,i}(K_1, K_1 - 1), B = Q_{i,i}(K_1, K_1) + Q_{i,i+1}(K_1, K_1), \\
C &= Q_{i,i+1}(K_1, K_1 - 1).
\end{aligned} \quad (2)$$

С помощью алгебраических преобразований можно показать, что

$$\begin{aligned}
Y_0 &= \text{diag}^+ \{e^T \otimes I_{WM_1(1+K_2M_2)}\}, \\
Y_1 &= \left(\begin{array}{c|c} O & O \\ \hline O & T^{-1}A \end{array} \right), \quad Y_2 = \left(\begin{array}{c|c} O & O \\ \hline O & T^{-1}B \end{array} \right),
\end{aligned}$$

где T – диагональная матрица, сформированная модулями диагональных элементов матрицы C . Нетрудно проверить, что матрица $Y_0 + Y_1 + Y_2$ стохастическая. В силу этого свойства согласно [15] справедливо утверждение следствия.

Далее условие эргодичности цепи Маркова $\xi_t, t \geq 0$ можно сформулировать в терминах матриц Y_0, Y_1, Y_2 . Следуя [15], вначале получим выражение для производящей функции $Y(z)$ этих матриц.

Лемма 2. Матричная производящая функция $Y(z) = Y_0 + Y_1 z + Y_2 z^2$ имеет вид

$$\begin{aligned}
Y(z) &= \text{diag}^+ \{e^T \otimes I_{WM_1(1+K_2M_2)}\} + \\
&\quad + \left(\begin{array}{c|c} O & O \\ \hline O & T^{-1}Az \end{array} \right) + z \left[\begin{array}{c|c} O & O \\ \hline O & T^{-1}(C+Bz) \end{array} \right] + zI,
\end{aligned}$$

где матрицы A, B, C определены в (2).

Теорема 1. 1) Цепь Маркова $\xi_t, t \geq 0$ эргодична, если справедливо следующее неравенство:

$$\lambda < \frac{1}{2} \left\{ \left[y_0 + \sum_{n=1}^{K_2-1} y_n (I_{M_1} \otimes e_{M_2}) \right] S_0^{(1)} + y_{K_2} (e_{M_1} \otimes I_{M_2}) S_0^{(2)} \right\}, \quad (3)$$

где вектор $y = (y_0, y_1, \dots, y_{K_2})$ – единственное решение системы уравнений

$$yV = 0, y e = 1, \quad (4)$$

а матрица V имеет вид

$$V = \text{diag}^+ \{S_0^{(1)} \mathbf{p}_1 \otimes \mathbf{p}_2, \underbrace{S_0^{(1)} \mathbf{p}_1 \otimes I_{M_2}}_{K_2-1}\} + \\ + \text{diag}^- \{I_{M_1} \otimes S_0^{(2)}, \underbrace{I_{M_1} \otimes S_0^{(2)} \mathbf{p}_2}_{K_2-1}\} + \\ + \text{diag} \{S_1, \underbrace{S_1 \otimes S_2}_{K_2-1}, S_1 \otimes S_2 + S_0^{(1)} \mathbf{p}_1 \otimes I_{M_2}\}.$$

2) цепь Маркова $\xi_t, t \geq 0$ не эргодична, если справедливо неравенство (3), взятое с противоположным знаком.

Доказательство. Матрица $Y(1)$ приводима. Обозначим через $Y^{K_2}(1)$ ее нормальную форму (см., например, [15]):

$$Y^{K_2}(z) = \text{diag}^- \{e^T \otimes I_{WM_1(1+K_2M_2)}\} + \\ + \left(\begin{array}{c|c} z[T^{-1}(C+Bz)] + zI & T^{-1}Az \\ \hline O & O \end{array} \right).$$

Заметим, что единственным неприводимым стохастическим блоком на диагонали данной матрицы является блочная структура вида

$$\tilde{Y}(z) = \left(\begin{array}{c|c} z[T^{-1}(C+Bz)] + zI & T^{-1}A \\ \hline I_{WM_1(1+K_2M_2)} & O \end{array} \right).$$

Из [15] (Теорема 3.14) следует, что цепь Маркова $\xi_t, t \geq 0$ эргодична, если

$$\frac{d}{dz} [\det(zI - \tilde{Y}(z))] \Big|_{z=1} > 0. \quad (5)$$

На основании блочной структуры матрицы $\tilde{Y}(z)$, определитель в последнем неравенстве можно свести к следующему виду:

$$\det(zI - \tilde{Y}(z)) = \det(zT^{-1}) \det[-z(C+Bz) - A]. \quad (6)$$

При $z=1$ второй определитель в правой части (6) обращается в ноль в силу свойства инфинитезимального генератора. Заметим, что $\det(zT^{-1}) > 0$ при $z=1$, поэтому выражение (5) принимает вид

$$\frac{d}{dz} [\det(-z(C+Bz) - A)] \Big|_{z=1} > 0,$$

который, в свою очередь, можно свести к неравенству

$$\mathbf{x}(C+2B)\mathbf{e} < 0, \quad (7)$$

где $\mathbf{x}$ - единственное решение системы уравнений

$$\mathbf{x}(C+B+A) = 0, \mathbf{x}\mathbf{e} = 1. \quad (8)$$

Рассмотрим далее вектор $\mathbf{x}$ в виде

$$\mathbf{x} = (\boldsymbol{\theta} \otimes \mathbf{y}_0, \boldsymbol{\theta} \otimes \mathbf{y}_1, \dots, \boldsymbol{\theta} \otimes \mathbf{y}_{K_2}), \quad (9)$$

где вектор $\mathbf{y}_0$ имеет размерность M_1 , а векторы $\mathbf{y}_1, \dots, \mathbf{y}_{K_2}$ - размерность $M_1 M_2$. Далее подставим вектор $\mathbf{x}$ вида (9) и

выражения (2) в (8). С учетом того, что $\boldsymbol{\theta}(D_0 + D_1) = 0, \boldsymbol{\theta}\mathbf{e} = 1$ после преобразований получим (4) для вектора $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{K_2})$. Таким образом, система (8) для вектора $\mathbf{x}$ сводится к (4) для вектора $\mathbf{y}$.

Далее рассмотрим неравенство (7). Подставляя вектор $\mathbf{x}$ в (9), матрицы B, C , определенные соотношением (2), в (7) с учетом $\boldsymbol{\theta} D_1 \mathbf{e} = \lambda$ после алгебраических преобразований получим неравенство (3), эквивалентное (7). Таким образом, утверждение 1 теоремы доказано. Утверждение 2 доказывается аналогично с помощью [15] (Теорема 3.14).

Из утверждения теоремы 1 следует, что в случае двухфазной тандемной системы без буферной памяти (с параметрами $K_1 = K_2 = 1, M_k = 1, S_k = -\mu_k, \mathbf{p}_k = 1, k = 1, 2$) неравенство (3)

$$\text{принимает вид } \lambda < \frac{\mu_1 \mu_2}{\mu_1 + \mu_2}.$$

Получим далее стационарное распределение вероятностей состояний двухфазной тандемной системы. Обозначим через $\mathbf{p}_i, i \geq 0$, векторы стационарных вероятностей цепи Маркова $\xi_t, t \geq 0$. Здесь индекс i обозначает число заявок на орбите. Зная структуру инфинитезимального генератора цепи Маркова $\xi_t, t \geq 0$, можно применить матрично-аналитический подход для расчета векторов $\mathbf{p}_i, i \geq 0$ [15]:

1. Решаем матричное уравнение $G = Y(G)$ с помощью итерационной схемы $G^{(n+1)} = Y(G^{(n)}), n \geq 1$ с начальным условием $G^{(0)} = O$.

2. Вычисляем матрицы $G_{i_0-1}, G_{i_0-2}, \dots, G_0$ с помощью обрточной рекурсии

$$G_i = (-Q_{i+1,i+1} - Q_{i+1,i+2} G_{i+1})^{-1} Q_{i+1,i} \text{ для } i = i_0 - 1, i_0 - 2, \dots, 0$$

с граничным условием $G_i = G, i \geq i_0$, где i_0 - неотрицательное целое число, такое, что для некоторого положительного ε (точности вычисления) выполняется неравенство $\|G_{i_0} - G\| < \varepsilon$.

3. Вычисляем матрицы $\bar{Q}_{i,i} = Q_{i,i} + Q_{i,i+1} G_i, \bar{Q}_{i,i+1} = Q_{i,i+1}$, где $G_i = G, i \geq i_0$.

4. Вычисляем матрицы $F_i, i \geq 1$ с помощью рекурсии

$$F_0 = I, F_i = F_{i-1} \bar{Q}_{i-1,i} (-\bar{Q}_{i,i})^{-1}, i \geq 1.$$

5. Вычисляем вектор $\mathbf{p}_0$ как единственное решение системы

$$\mathbf{p}_0 (-\bar{Q}_{0,0}) = 0, \mathbf{p}_0 \sum_{i=0}^{\infty} F_i \mathbf{e} = 1.$$

6. Вычисляем векторы $\mathbf{p}_i = \mathbf{p}_0 F_i, i \geq 1$.

Вычислив стационарное распределение $\mathbf{p}_i, i \geq 0$, можно получить ряд характеристик производительности рассматриваемой системы и применить закон Литтла для расчета среднего времени пребывания во всей системе:

1. Стационарные вероятности числа заявок на орбите $P_i = \mathbf{p}_i \mathbf{e}, i \geq 0$.

2. Среднее число мест, занятых повторными заявками

$$L = \sum_{i=1}^{\infty} i p_i e.$$

3. Совместное распределение

$$q_i(j, n), i \geq 0, j = \overline{0, K_1}, n = \overline{0, K_2}$$

числа заявок на орбите, на первой и на второй фазах. Здесь $q_i(j, n)$ - вероятность того, что в системе имеется i повторных заявок, на первой фазе ожидают j заявок, а на второй фазе - n заявок. Обозначим через $R = M_1(1 + M_2 K_2)$. Вероятности $q_i(0, 0)$, $q_i(j, 0)$, $q_i(0, n)$, $q_i(j, n)$ вычисляются по формулам:

$$\begin{aligned} q_i(0, 0) &= p_i \begin{pmatrix} e_W \\ 0_{W(M_2 K_2 + R K_1)}^T \end{pmatrix}, i \geq 0, \\ q_i(0, n) &= p_i \begin{pmatrix} 0_{W[1+M_2(n-1)]}^T \\ e_{WM_2} \\ 0_{W[M_2(K_2-n)+R K_1]}^T \end{pmatrix}, i \geq 0, n = \overline{1, K_2}, \\ q_i(j, 0) &= p_i \begin{pmatrix} 0_{W[1+M_2 K_2 + R(j-1)]}^T \\ e_{WM_1} \\ 0_{W[M_1 M_2 K_2 + R(K_1-j)]}^T \end{pmatrix}, \\ q_i(j, n) &= p_i \begin{pmatrix} 0_{W[1+M_2 K_2 + R(j-1)+M_1(1+M_2(n-1))]}^T \\ e_{WM_1 M_2} \\ 0_{W[M_1 M_2(K_2-n)+R(K_1-j)]}^T \end{pmatrix}, \\ i &\geq 0, j = \overline{1, K_1}, n = \overline{1, K_2}. \end{aligned}$$

4. Стационарные вероятности состояний первой фазы:

$$q_j^{(1)} = \sum_{i=0}^{\infty} p_i \sum_{n=0}^{K_2} q_i(j, n), j = \overline{0, K_1}.$$

5. Среднее число заявок, ожидающих на первой фазе:

$$L_1 = \sum_{j=1}^{K_1} j q_j^{(1)}.$$

6. Стационарные вероятности состояний второй фазы:

$$q_n^{(2)} = \sum_{i=0}^{\infty} p_i \sum_{j=0}^{K_1} q_i(j, n), n = \overline{0, K_2}.$$

7. Среднее число заявок, ожидающих на второй фазе:

$$L_2 = \sum_{n=1}^{K_2} n q_n^{(2)}.$$

8. Межконцевая задержка: $W = (L + L_1 + L_2) \lambda^{-1}$.

Для систем, имеющих практическое применение в компьютерных сетях, необходимо, как правило, исследовать модели с более, чем двумя узлами (фазами). Отметим, что результаты, полученные в данном разделе, имеют не только

теоретическое значение, но и будут использованы в следующем разделе для верификации имитационной модели, описывающей функционирование стохастической тандемной системы с произвольным числом фаз ($N > 2$).

3 Анализ тандемных систем с произвольным числом фаз методами машинного обучения

В данном разделе для исследования тандемной системы с повторными вызовами и произвольным числом фаз впервые предложено использование комбинации методов имитационного моделирования и машинного обучения. Имитационная модель позволяет получить набор данных (датасет) о характеристиках тандемной системы, который далее используется для прогнозирования характеристик производительности тандемной системы с использованием методов машинного обучения.

3.1 Генерация синтетического набора данных

Для применения методов машинного обучения [16], на первом этапе необходимо сгенерировать обучающий набор данных, включающий в себя входные параметры тандемной системы с произвольным числом фаз:

- матрицы D_0, D_1 , описывающие MAP-поток;
 - число фаз N ;
 - матрицы (S_i, B_i) , характеризующие фазовое распределение, и размер буфера K_i на i -й фазе, $i = \overline{1, N}$;
 - максимальное количество мест на орбите K_{orbit} и скорость α поступления заявок,
- а также характеристики производительности тандемной системы (межконцевая задержка, среднее количество заявок в системе и т.д.), вычисленных с помощью других методов (аналитический расчет, имитационное моделирование). В данном примере будем рассматривать лишь одну характеристику – межконцевую задержку.

Дискретно-событийная модель реализована на языке программирования C++. Модуль машинного обучения имеет две основные структурные части: предобработка данных и обучение моделей. Каждая из них реализована на языке программирования Python. Задача скрипта обработки данных – считать и преобразовать эти данные в числовой формат, пригодный для обработки библиотеками машинного обучения. Из-за различий в параметрах обучения и используемых библиотеках у каждого метода машинного обучения есть свой собственный скрипт обучения.

Для обучения машинных моделей необходим набор данных, полученных с помощью имитационной модели тандемной системы с большим количеством наборов данных (X, y) , где X соответствует входным параметрам программы имитационного моделирования, а y - прогнозируемый искомый параметр (межконцевая задержка). Набор данных сгенерирован на основе возможных комбинаций входных параметров, включая: $N = 2, 15$; размер матриц, описывающих MAP и PH - от 1 до 16; $K_i = \overline{0, 32}$ для $i = \overline{1, N}$; размер орбиты $K_{orbit} = \overline{0, 512}$. Заметим, что для имитационной модели размер

орбиты должен быть ограничен.

С целью представления более разнообразного набора данных использованы два метода выбора недиагональных значений элементов матриц D_0, D_1, S_i , $i = \overline{1, N}$ и интенсивности α поступления заявок с орбиты на первую фазу:

Способ 1. Недиагональные элементы матрицы D_0 и элементы матрицы D_1 генерируются как непрерывные случайные величины, равномерно распределенные на отрезке $[1; 1024]$. Недиагональные элементы матриц S_i , интенсивности переходов $S_i^{(0)}$ в поглощающее состояние и интенсивность повторных попыток на орбите α генерируются как непрерывные случайные величины, равномерно распределенные на отрезке $[2; 2048]$.

Способ 2. Недиагональные элементы матриц D_0, D_1 генерируются как e^p ; недиагональные элементы матриц S_i , интенсивности переходов $S_i^{(0)}$ в поглощающее состояние и интенсивность повторных заявок α генерируются как $2e^p$, где p – непрерывная случайная величина, равномерно распределенная на отрезке $[-2,5; 2,5]$.

Диагональные же элементы матриц D_0, S_i задаются элементами векторов $-(D_0 + D_1)e$ и $-(S_i e + S_i^{(0)})$, соответственно. Далее входные параметры масштабируются посредством деления элементов матриц D_0, D_1, S_i и интенсивности повторных попыток на орбите α на коэффициент, равный λ (среднее число заявок, поступающих в единицу времени) с целью, чтобы все входные наборы данных имели одинаковую интенсивность входного потока, равную 1.

Данные датасета получены с использованием имитационной модели тандемной системы: *Dataset300k* – набор данных для обучения, содержащих 200 тыс. элементов (X, y) (способ 1), дополненных 100 тыс. элементов, сгенерированных методом 2, и *Test60k* – набор данных для тестирования, содержащий из 60 тыс. элементов (X, y) (способ 2).

С целью сокращения вычислительной сложности на конечном этапе подготовки данных предусмотрена процедура уменьшения размерности МАР-потоков с использованием аппроксимации, описанной в [18].

В следующем разделе проведем сравнительный анализ различных методов машинного обучения: градиентного бустинга, случайного леса и искусственных нейронных сетей [16]. Большая часть обучения методами градиентного бустинга и случайного леса выполнено с помощью библиотеки *scikit-learn* – готовой Python-библиотеки для машинного обучения. Для метода нейронных сетей использована Python-библиотека *PyTorch*.

3.2 Градиентный бустинг

Градиентный бустинг – это ансамблевый метод машинного обучения, который последовательно создаёт набор слабых с точки зрения прогноза требуемых значений моделей, комбинируя их в единую более сильную прогностическую модель.

Ансамбли – это алгоритмы, сочетающие сразу несколько моделей для исправления ошибок, допущенных предыдущими моделями.

Для решения задачи прогнозирования межконцевой задержки стохастической тандемной системы будем обучать модели градиентного бустинга с параметрами, заданными по умолчанию, скоростью обучения 0,01, методом наименьших квадратов для оценки функции потерь и 10000 итераций обучения. Рассмотрим также два варианта: с логарифмическим преобразованием искомого параметра (межконцевой задержки) и без него, что предусматривается возможностями метода и позволяет выбрать способ обучения, дающий более точный результат. Из результатов обучения, представленных в таблице 1, следует, что лучший результат дает опция логарифмического преобразования искомого параметра $\ln(y)$.

Таблица 1

Результаты оценки моделей методом градиентного бустинга

Набор данных	Обучение без логарифмического преобразования		Обучение с логарифмическим преобразованием	
	Score	σ	Score	σ
Dataset300k	0,5988	1,8435	0,0790	0,1351
Test60k	0,5659	1,8155	0,0900	0,2021

3.3 Случайный лес

Случайный лес – это ещё один популярный алгоритм машинного обучения, который так же, как и градиентный бустинг, относится к классу ансамблевых методов. Метод предполагает создание множества отдельных независимых «решающих деревьев», что позволяет получать более точные результаты и, в то же время, снижать тенденцию к переобучению модели.

Результаты прогнозирования межконцевой задержки методом случайного леса представлены в таблице 2. Как и в случае применения градиентного бустинга, результаты получены с использованием логарифмического преобразования искомого параметра (межконцевой задержки), а также без него.

Сравнивая результаты таблиц 1-2, можно отметить более высокую тенденцию к переобучению при использовании метода случайного леса, поскольку на обучающей выборке точность модели случайного леса (4,6%) выше, чем у градиентного бустинга (7,9%), но хуже на тестовой выборке данных *Test60k*.

Существенным недостатком метода случайного леса является значительно больший объем требуемой памяти (таблица 3). Например, даже при количестве деревьев 100, что является значением по умолчанию, модели требуется 2,6 ГБ памяти. Это объясняется тем, что по умолчанию библиотека *Scikit-learn* не предполагает ограничения на глубину деревьев решений (параметр $\maxDepth = 0$). Следовательно, для больших обучающих наборов данных в нашем случае деревья решений будут разрастаться до максимально возможного размера, что приводит к большому размеру результирующей модели. Чтобы уменьшить размер модели, следует обучать ее с параметром $\maxDepth > 0$ за счет снижения точности.

Таблица 2

Результаты оценки моделей методом случайного леса

Набор данных	Обучение без логарифмического преобразования			
	Число деревьев			
	100		300	
	Score	σ	Score	σ
Dataset300k	0,0551	0,1603	0,0551	0,1587
Test60k	0,1176	0,2693	0,1181	0,2703
	500		800	
	Score	σ	Score	σ
	Score	σ	Score	σ
Dataset300k	0,0551	0,1580	0,0551	0,1579
Test60k	0,1180	0,2694	0,1177	0,2686
	Обучение с логарифмическим преобразованием			
	Число деревьев			
	100		300	
	Score	σ	Score	σ
Dataset300k	0,0467	0,0914	0,0458	0,0897
Test60k	0,1097	0,2357	0,1089	0,2344
	500		800	
	Score	σ	Score	σ
	Score	σ	Score	σ
Dataset300k	0,0456	0,0893	0,0455	0,0891
Test60k	0,1085	0,2343	0,1085	0,2343

Таблица 3

Результаты оценки моделей случайного леса, обученных с различными значениями maxDepth

Макс. глубина maxDepth	Набор данных				Объём памяти
	Dataset300k		Test60k		
	Score	σ	Score	σ	
4	0,5205	0,5744	0,4781	0,5450	243 KB
8	0,2601	0,3397	0,2419	0,3417	3,49 MB
16	0,1153	0,1429	0,1399	0,2500	209MB
32	0,0476	0,0905	0,1098	0,2358	2,34 GB
64	0,0467	0,0915	0,1094	0,2350	2,60 GB
0	0,0467	0,0914	0,1097	0,2357	2,6 GB

3.4 Нейронные сети

В современной практике машинного обучения искусственные нейронные сети занимают доминирующее положение, демонстрируя высокую эффективность в широком спектре прикладных задач. Несмотря на то, что данный метод известен довольно давно, его масштабное внедрение стало возможным лишь в последние десятилетия благодаря существенному росту вычислительной мощности оборудования, оптимизированного для выполнения высокопроизводительных операций. Это позволило эффективно обучать глубокие нейронные сети с большим числом параметров на масштабных наборах данных, что, в свою очередь, обеспечило превосходство методов, базирующихся на нейронных сетях, над альтернативными методами в задачах прогнозирования по критерию точности.

Структура нейронной сети, используемой для решения задачи прогнозирования межконцевой задержки в тандемной системе с повторными вызовами, представлена на рис. 2.

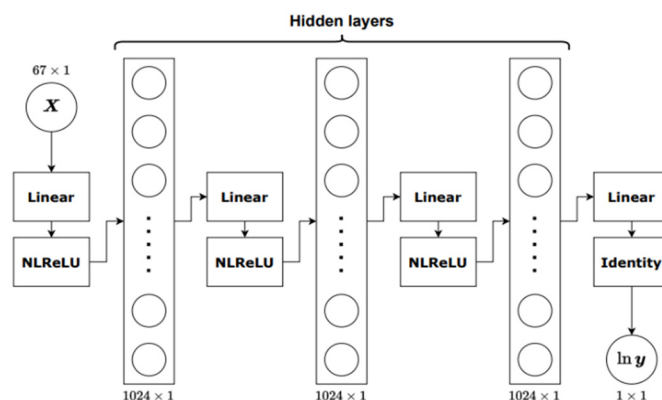


Рис. 2. Структура модели нейронной сети

Сеть содержит три скрытых слоя, каждый из которых состоит из 1024 нейронов. Изначально используется стандартная функция активации ReLU между каждым скрытым слоем, однако во время обучения возникает проблема: значение потерь становится очень высоким (взрывной градиент), что делает модель необучаемой. Простое решение этой проблемы состоит в замене функции активации на ее логарифмический вариант ReLU:

$$NLRReLU(x) = \ln(\text{ReLU}(x) + 1) = \ln(\max(0, x) + 1).$$

В таблице 4 представлены значения межконцевой задержки для некоторых этапов обучения, что демонстрирует более высокую обучающую способность нейронной сети, чем методы градиентного бустинга и случайного леса.

Таблица 4

Результаты оценки модели нейронной сети на некоторых этапах обучения

Набор данных	Момент обучения			
	5670		6912	
	Score	σ	Score	σ
Dataset300k	0,0733	0,2068	0,07070	0,1950
Test60k	0,0646	0,0936	0,0576	0,1005
	8448		9344	
	Score	σ	Score	σ
	Score	σ	Score	σ
Dataset300k	0,0761	0,2048	0,0741	0,1986
Test60k	0,0659	0,1317	0,046	0,0716

На основании приведенных выше результатов использования различных методов машинного обучения можно провести их сравнительный анализ. Модель случайного леса прогнозирует больше значений межконцевой задержки с относительно небольшой ошибкой ($\leq 10\%$), чем градиентный бустинг, но при этом результатов, получаемых с неприемлемой ошибкой ($\geq 90\%$) несоизмеримо больше, чем у других методов. Метод градиентного бустинга для рассматриваемой тандемной системы дает наименьшее число результатов, близких к точному, но при этом число результатов с недопустимой ошибкой также меньше, чем у других рассмотренных методов (табл. 5).

Таблица 5

Число экспериментов при различных значениях ошибки ε

ε	Нейросеть	Градиент. Бустинг	Случайный лес
$\leq 10\%$	532	347	474
$\geq 50\%$	66	90	89
$\geq 90\%$	35	25	43

3.5 Сравнительный анализ времени вычислений

Основная задача применения методов машинного обучения к задачам теории массового обслуживания – сокращение времени вычисления требуемых характеристик модели, возможно, с некоторым снижением точности вычислений. Анализ временных затрат показывает, что генерация одной выборки данных посредством имитационного моделирования занимает в среднем 4 секунды. Для масштабных наборов данных (сотни тысяч или миллионы выборок) это приводит к существенным временным затратам, ограничивающим применимость метода в условиях жёстких временных ограничений. В таблице 6 приведены сравнительные данные о производительности и объеме требуемой памяти для методов машинного обучения, что демонстрирует превосходство данных методов по вычислительной скорости на порядки величин по сравнению с методом имитационного моделирования для оценки производительности модели тандемной системы с повторными вызовами и общей орбитой.

Таблица 6

Сравнение объема памяти и среднего времени обучения

Метод	Объем памяти	Число итераций обучения	Общее время обучения
Градиентный бустинг	34,5 Мб	10000	143 с
Случайный лес	2,6 Гб	100	306 с
Нейронная сеть	24,8 Мб	10000	13 ч

4 Заключение

Для двухфазной тандемной системы разработан алгоритм расчета точных характеристик производительности (межконцевая задержка, средняя число заявок в узлах и на орбите и др.). Такое исследование имеет не только теоретическое значение, но используется для верификации имитационной модели, описывающей функционирование стохастической тандемной системы с произвольным числом фаз. Для исследования системы с произвольным числом фаз впервые предложен подход с использованием методов машинного обучения, обеспечивающий высокую вычислительную эффективность по сравнению с имитационным моделированием.

Проведена оценка достоинств и недостатков различных методов машинного обучения (градиентный бустинг, случайный лес и нейронные сети) с целью их применения для прогнозирования основных характеристик производительности стохастических тандемных систем с повторными вызовами и общей орбитой.

Литература

1. Anitha K., Poongothai V., Godhandaraman P. Performance

analysis of a queueing system with tandem nodes, retrial and server vacations // Results in Control and Optimization. 2025. Vol. 18. Article 100520. DOI: 10.1016/j.rico.2025.100520.

2. Vishnevsky V. M., Larionov A. A., Mukhtarov A. A., Sokolov A. M. Investigation of tandem queueing systems using machine learning methods // Control Sciences. 2024. Vol. 4. P. 10-21. DOI: 10.25728/cs.2024.4.2.

3. Rovetto C., Cruz E., Nuñez I., Santana K., Smolarz A., Rangel J., 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. Article 10030. DOI: 10.3390/app131810030.

4. Вишневецкий В. М., Ларионов А. А., Целикин Ю. В. Анализ и исследование методов проектирования автоматизированных систем безопасности на автодорогах с использованием новых широкополосных беспроводных средств и RFID-технологий // Т-Comm: Телекоммуникации и транспорт. 2012. № 7. С. 48-54.

5. Вишневецкий В. М., Минниханов Р. Н., Барский И. В., Ларионов А. А. Гибридная система идентификации транспортных средств // Вестник НЦБЖД. 2022. № 4 (54). С. 33-42.

6. Dudin S. A., Dudina O. S., Dudin A. N. Analysis of tandem queue with multi-server stages and group service at the second stage // Axioms. 2024. Vol. 13. No. 4. Article 214. DOI: 10.3390/axioms13040214.

7. Dieleman N. A., Berkhout J., 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. Article 106124. DOI: 10.1016/j.cor.2022.106124.

8. Dudin S. A., Dudin A. N., Dudina O. S., Chakravarthy S. R. Analysis of a tandem queueing system with blocking and group service in the second node // International Journal of Systems Science: Operations & Logistics. 2023. Vol. 10. No. 1. Article 2235270. DOI: 10.1080/23302674.2023.2235270.

9. Yeager Y., Boxma O. J., Resing J., Vlasiou M. ASIP tandem queues with consumption // Performance Evaluation Review. 2024., Vol. 51. No. 4, pp. 14-15. DOI: 10.1145/3649477.3649486

10. Вишневецкий В. М., Ларионов А. А., Мухтаров А. А., Соколов А. М. Исследование многофазных систем массового обслуживания с помощью методов машинного обучения // Проблемы управления. 2024. № 4. С. 13-25.

11. Kim J., Kim B. A survey of retrial queueing systems // Annals of Operations Research. 2016. Vol. 247, pp. 3-36. DOI: 10.1007/s10479-015-2038-7.

12. Saravanan V., Poongothai V., Godhandaraman P. Performance analysis of a multi server retrial queueing system with unreliable server, discouragement and vacation model // Mathematics and Computers in Simulation. 2023. Vol. 214. No. C, pp. 204-226. DOI: 10.1016/j.matcom.2023.07.008

13. Avrachenkov K., Yechiali U. On tandem blocking queues with a common retrial queue // Computers & Operations Research. 2010. Vol. 37. P. 1174-1180. DOI: 10.1016/j.cor.2009.10.004.

14. Nazarov A. A., Paul S. V., Phung-Duc T., Morozova M. Analysis of tandem retrial queue with common orbit and MMPP incoming flow // Lecture Notes in Computer Science. 2023. Vol. 13766, pp. 270-283. DOI: 10.1007/978-3-031-23207-7_21

15. Вишневецкий В. М., Дудин А. Н., Клименко В. И. Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях. М.: ТЕХНОСФЕРА, 2018. 564 с.

16. Вишневецкий В. М., Ефросинин Д. В. Теория очередей и машинное обучение. М.: ИНФРА-М, 2024. 370 с.

17. Efrosinin D., Vishnevsky V., Stepanova N. Optimal scheduling in general multi queue system by combining simulation and neural network techniques // Sensors. 2023. Vol. 23. No. 12. Article No. 5479.

18. Vishnevsky V., Larionov A., Ivanov R., 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 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT). 2017, pp. 1-5. DOI: 10.1109/ICAICT.2017.8687247.

ANALYSIS OF STOCHASTIC TANDEM SYSTEMS WITH RETRIALS

Vladimir V. Vishnevsky, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia, vishn@inbox.ru

Olga V. Semenova, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia,

olga.semenova@ipu.ru

Abstract

At present, one of the key areas in the design and performance evaluation of modern telecommunication systems and networks is the analysis of mathematical models of tandem systems. Tandem systems are mathematical models of various technical and social systems, in particular, they are adequate models for broadband wireless networks deployed along extended transportation routes such as highways, railways, oil and gas pipelines, etc. The tandem system considered in this study consists of an arbitrary finite number of service units (stages) with a limited waiting space. The arrival process (packet flow) at the first stage is a correlated Markovian Arrival Process (MAP) and the service times have the phase-type distribution. If the buffer memory of any stage is full upon the customer arrival the customer is placed to an orbit (a shared memory for all stages). From there, after a random period of time, it is tries to get served at the first stage. We evaluate the key performance metrics: end-to-end delay (the time from the moment a customer enters the first stage from outside the system until it is successfully served at the final stage); average queue lengths at the system stages and in the orbit. For systems with two queues, exact formulas are derived to calculate the required performance characteristics. For systems with an arbitrary number of stages, we employ a machine learning approach in combination with a simulation modeling. The study involved conducting numerous numerical experiments across a wide range of parameters.

Keywords: Queueing systems, tandem system, retrials (orbit), MAP-input, phase-type distribution, simulation, machine learning

References

- [1] K. Anitha, V. Poongothai, and P. Godhandaraman, "Performance analysis of a queueing system with tandem nodes, retrial and server vacations," *Results in Control and Optimization*, vol. 18, article 100520, 2025. doi: 10.1016/j.rico.2025.100520.
- [2] V. M. Vishnevsky, A. A. Larionov, A. A. Mukhtarov, and A. M. Sokolov, "Investigation of tandem queueing systems using machine learning methods," *Control Sciences*, vol. 4, pp. 10-21, 2024. doi: 10.25728/cs.2024.4.2.
- [3] C. Rovetto, E. Cruz, I. Nuñez, K. Santana, A. Smolarz, J. Rangel, and E. E. Cano, "Minimizing intersection waiting time: proposal of a queue network model using Kendall's notation in Panama City," *Applied Sciences*, vol. 13, no. 18, article 10030, 2023. doi: 10.3390/app131810030.
- [4] V. M. Vishnevskii, A. A. Larionov, and Yu. V. Tselikin, "Analysis and research of design methods for automated road safety systems using new broadband wireless tools and RFID technologies," *T-Comm*, no. 7, pp. 48-54, 2012.
- [5] V. M. Vishnevskii, R. N. Minnikhanov, I. V. Barskii, and A. A. Larionov, "Hybrid vehicle identification system," *Vestnik NTsBZhD*, no. 4 (54), pp. 33-42, 2022.
- [6] S. A. Dudin, O. S. Dudina, and A. N. Dudin, "Analysis of tandem queue with multi server stages and group service at the second stage," *Axioms*, vol. 13, no. 4, article 214, 2024. doi: 10.3390/axioms13040214.
- [7] N. A. Dieleman, J. Berkhout, and B. Heidergott, "A neural network approach to performance analysis of tandem lines: the value of analytical knowledge," *Computers and Operations Research*, vol. 152, no. 3, article 106124, 2023. doi: 10.1016/j.cor.2022.106124.
- [8] S. A. Dudin, A. N. Dudin, O. S. Dudina, and S. R. Chakravarthy, "Analysis of a tandem queueing system with blocking and group service in the second node," *International Journal of Systems Science: Operations & Logistics*, vol. 10, no. 1, article 2235270, 2023. doi: 10.1080/23302674.2023.2235270
- [9] Y. Yeager, O. J. Boxma, J. Resing, and M. Vlasiov, "ASIP tandem queues with consumption," *ACM SIGMETRICS Performance Evaluation Review*, vol. 51, no. 4, pp. 14-15, Feb. 2024. doi: 10.1145/3649477.3649486.
- [10] V. M. Vishnevskii, A. A. Larionov, A. A. Mukhtarov, and A. M. Sokolov, "Research of multi phase queueing systems using machine learning methods," *Problemy Upravleniya*, no. 4, pp. 13-25, 2024.
- [11] J. Kim and B. Kim, "A survey of retrial queueing systems," *Annals of Operations Research*, vol. 247, pp. 3-36, 2016. doi: 10.1007/s10479-015-2038-7.
- [12] V. Saravanan, V. Poongothai, and P. Godhandaraman, "Performance analysis of a multi server retrial queueing system with unreliable server, discouragement and vacation model," *Mathematics and Computers in Simulation*, vol. 214, pp. 204-226, 2023. doi: 10.1016/j.matcom.2023.07.008.
- [13] K. Avrachenkov and U. Yechiali, "On tandem blocking queues with a common retrial queue," *Computers & Operations Research*, vol. 37, pp. 1174-1180, 2010. doi: 10.1016/j.cor.2009.10.004.
- [14] A. A. Nazarov, S. V. Paul, T. Phung Duc, and M. Morozova, "Analysis of tandem retrial queue with common orbit and MMPP incoming flow," in *Lecture Notes in Computer Science*, vol. 13766, pp. 270-283, 2023. doi: 10.1007/978-3-031-23207-7_21.
- [15] V. M. Vishnevskii, A. N. Dudin, and V. I. Klimenok, *Stochastic Systems with Correlated Flows: Theory and Application in Telecommunication Networks*. Moscow: Tekhnosfera, 2018, 564 pp.
- [16] V. M. Vishnevskii and D. V. Efrosinin, *Queueing Theory and Machine Learning*. Moscow: INFRA-M, 2024, 370 pp.
- [17] D. Efrosinin, V. Vishnevsky, and N. Stepanova, "Optimal scheduling in general multi queue system by combining simulation and neural network techniques," *Sensors*, vol. 23, no. 12, article 5479, 2023.
- [18] V. Vishnevsky, A. Larionov, R. Ivanov, and O. Semenova, "Estimation of IEEE 802.11 DCF access performance in wireless networks with linear topology using PH service time approximations and MAP input," in *Proceedings of the IEEE 11th International Conference on Application of Information and Communication Technologies (AICT)*, 2017, pp. 1-5. doi: 10.1109/ICAICT.2017.8687247.

Information about authors:

Vladimir V. Vishnevsky, Dr. Sci. (Techn.), Professor, Chief Researcher, Laboratory of Telecommunication Systems, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences. ORCID 0000-0001-7373-4847

Olga V. Semenova, Dr. Sci. (Phys.-Math.), Leading Researcher, Laboratory of Telecommunication Systems, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences. ORCID 0000-0001-7846-6212