

ИССЛЕДОВАНИЕ МНОГОФАЗНЫХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ#

В. М. Вишнеvский*, А. А. Ларионов**, А. А. Мухтаров***, А. М. Соколов****

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

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

Аннотация. Рассматривается многофазная система массового обслуживания с входящим коррелированным MAP-потокoм, PH-распределением времени обслуживания и ограниченным размером буфера на фазах системы. Приведен краткий истoрический обзор по анализу моделей таких систем и методов их исследования. На основании проведенного обзора обоснована новизна постановки задачи, рассматриваемой в статье, методов ее решения и результатов. Дано описание алгоритма точного расчета характеристик производительности многофазных систем малой размерности и оценки сложности этого алгоритма. Для исследования многофазных систем большой размерности предложен подход, основанный на комбинации методов имитационного моделирования и машинного обучения. Приведены результаты численного анализа, подтвердившие эффективность применения методов машинного обучения для оценки характеристик производительности тандемных систем, адекватно описывающих функционирование широкополосных беспроводных сетей.

Ключевые слова: многофазные сети, аналитическая модель, имитационное моделирование, машинное обучение.

ВВЕДЕНИЕ

Многофазные системы массового обслуживания (тандемные СМО) традиционно используются для моделирования и оптимизации производительности многих сложных систем, включая технические, экономические, промышленные, транспортные, медицинские, военные и др. [1–4]. Значительный интерес также представляют модели тандемных СМО, адекватно описывающие функционирование современных широкополосных беспроводных сетей с линейной топологией [5, 6].

Исследование многофазных СМО, начатое в конце 1960-х гг. [7, 8], интенсивно продолжается до настоящего времени [9–11]. Первоначально аналитические результаты были получены для оценки характеристик производительности двухфазных СМО. Многочисленные статьи посвящены

разработке методов исследования таких систем с блокировками, пуассоновским и коррелированным (BMAP, *Batch Markovian Arrival Process*) входным потоком и различными функциями распределения времени обслуживания (см., например, работы [12–16]). Более подробно с описанием работ о двухфазных СМО можно ознакомиться в обзоре [17] и книге [18].

Однако наибольшее практическое значение представляют тандемные СМО большой размерности (с числом узлов больше двух). Точное аналитическое решение для таких СМО было получено лишь для небольшого класса сетей, узлы которых удовлетворяют условиям теоремы BSMР (Baskett, Chandy, Muntz, Palacios) [19, 20], а совместное стационарное распределение вероятностей состояний имеет мультипликативную форму. В частном случае, когда узлы (фазы) тандемной сети представляют собой СМО типа $M/M/1$, легко находятся характеристики производительности такой системы, включая важный параметр – межконцевую задержку (время с момента поступ-

Исследование выполнено при поддержке Российского научного фонда, грант № 22-49-02023, <https://rscf.ru/project/22-49-02023/>.

ления пакета на первую фазу до окончания его обслуживания на последней фазе).

Для других типов тандемных сетей большой размерности методы отыскания точных решений практически отсутствуют. Поэтому для их исследования широко применяются приближенные методы. Одним из наиболее распространенных приближенных методов является декомпозиция тандемной сети на отдельные узлы с последующим изучением их взаимодействия друг с другом и обобщением полученных результатов для оценки характеристик производительности сети в целом (см., например, работы [21–24]). В последние годы для исследования тандемных СМО эффективно применяются методы машинного обучения [9, 25–27]. В частности, в статье [9] исследуются характеристики производительности тандемных СМО с пуассоновским входным потоком, экспоненциальным распределением времени обслуживания и многолинейными узлами ($M/M/S/\infty$) с использованием искусственных нейронных сетей.

Новизна настоящей статьи состоит в разработке точного алгоритма расчета характеристик производительности многофазных СМО с коррелированным входным МАР-поток и РН-распределением времени обслуживания на фазах системы. Приводится оценка сложности алгоритма, а также описываются его достоинства и ограничения (на число фаз системы при численном анализе). Для исследования тандемной СМО большой размерности применен эффективный подход, базирующийся на комбинации методов машинного обучения и имитационного моделирования. Указанный подход обеспечивает возможность быстрого расчета характеристик СМО большой размерности при проектировании сложных вычислительных систем и компьютерных сетей [9, 28, 29].

1. ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ

Рассматривается тандемная СМО с числом фаз (узлов) $N \geq 2$ и конечной емкостью буфера на каждой фазе M_i ($i = 1, \dots, N$). Заявки (пакеты) поступают в систему в МАР-поток, который задается управляющим процессом $v_i, (t \geq 0)$, представляющим собой цепь Маркова с пространством состояний $\{0, 1, \dots, W\}$ и матрицами \mathbf{D}_0 и \mathbf{D}_1 . Матрица \mathbf{D}_1 описывает переходы управляющей марковской цепи, сопровождающиеся генерацией заявки (наблюдаемые переходы), а матрица \mathbf{D}_0 – переходы, при которых генерация новой заявки не происходит (невидимые переходы). Матрица $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_1$ является инфинитезимальным генератором мар-

ковской цепи v_i . Интенсивность поступления пакетов λ определяется как $\lambda = \bar{\pi} \mathbf{D}_1 \bar{\mathbf{1}}$, где $\bar{\pi}$ – вектор стационарного распределения процесса v_i , который определяется как единственное решение системы линейных алгебраических уравнений $\bar{\pi} \mathbf{D} = \bar{\mathbf{0}}, \bar{\pi} \bar{\mathbf{1}} = 1$. Здесь и далее $\bar{\mathbf{1}}$ – вектор-столбец, состоящий из единиц, $\bar{\mathbf{0}}$ – вектор-строка, состоящая из нулей.

Время обслуживания пакетов на i -й фазе имеет РН-распределение с неприводимым представлением $(\mathbf{S}_i, \bar{\tau}_i)$, $i = 1, \dots, N$. Здесь \mathbf{S}_i – квадратная матрица порядка V_i (число состояний процесса), а $\bar{\tau}_i$ – вектор, задающий вероятности начального состояния процесса. Таким образом, процесс обслуживания на i -й фазе происходит под управлением цепи Маркова с пространством состояний $\{1, \dots, V_i, V_i + 1\}$, где $V_i + 1$ – поглощающее состояние. Более подробную информацию о РН-распределении и МАР-поток можно найти, например, в монографии [18]. Пакет, поступивший в момент, когда буфер одной из фаз полностью занят, покидает систему без обслуживания (считается потерянным).

В дальнейшем будет проведено исследование двух вариантов тандемной системы: системы с кросс-трафиком, когда в i -ю фазу кроме выходного потока Z_i с $i - 1$ -й фазы поступает и дополнительный МАР-поток X_i , и системы без кросс-трафика, в которой внешний трафик поступает только на первую фазу (рис. 1).

Задача состоит в получении оценок стационарных характеристик производительности описанных тандемных систем, включая время межконцевой задержки (время с момента поступления пакета на первую фазу до завершения его обслуживания на последней фазе), вероятность потери пакета и т. д.

2. АЛГОРИТМ ТОЧНОГО РАСЧЕТА ХАРАКТЕРИСТИК ТАНДЕМНОЙ СЕТИ С ВХОДЯЩИМ МАР-ПОТОКОМ, РН-РАСПРЕДЕЛЕНИЕМ ВРЕМЕНИ ОБСЛУЖИВАНИЯ И ОГРАНИЧЕННЫМИ РАЗМЕРАМИ БУФЕРА НА ФАЗАХ СИСТЕМЫ

Рассмотрим прежде всего свойства и характеристики системы массового обслуживания МАР/РН/1/М, которые могут быть использованы для построения точного алгоритма расчета тандемной системы. Ключевым свойством системы МАР/РН/1/М является замкнутость на множестве МАР-потоков согласно следующим теоремам [18].

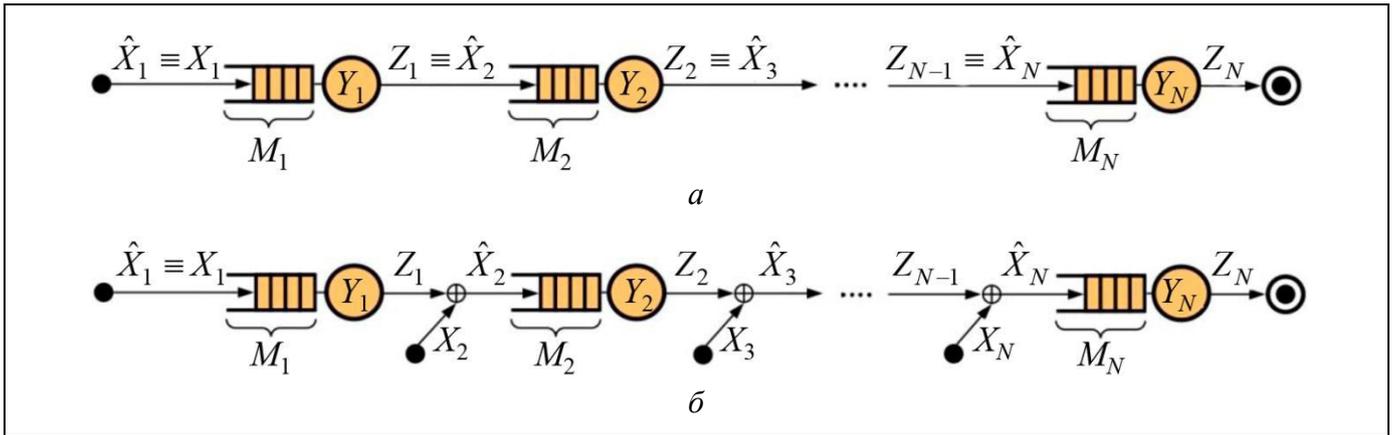


Рис. 1. Потoki заявок в тандемной сети массового обслуживания с кросс-трафиком (б) и без него (а)

Теорема 1. Поток выходных (обслуженных) пакетов в системе $MAP/PH/1/M$, где входной MAP -поток задается $X: MAP(\mathbf{D}_0, \mathbf{D}_1)$, а время обслуживания имеет фазовое распределение $Y: PH(\mathbf{S}, \bar{\tau})$, является MAP -потокom $Z \sim MAP(\mathbf{D}'_0, \mathbf{D}'_1)$, матрицы которого имеют вид

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

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

где $\mathbf{C}_t = (-\mathbf{S}\bar{\mathbf{I}}_V) \otimes \bar{\tau}$, а $\mathbf{I}_V, \mathbf{I}_W$ – единичные матрицы порядков V и W соответственно.

Теорема 2 Суперпозиция MAP -потокoв X_1 и $X_2, X_i \sim MAP(\mathbf{D}_0^{(i)}, \mathbf{D}_1^{(i)})$, $i=1,2$ – MAP -поток

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

где \oplus – сумма Кронекера. Если потоки X_1 и X_2 имеют размерности W_1 и W_2 , то размерность суммарного потока X равна $W = W_1 W_2$.

Обозначим Z_i выходящий поток с i -го узла тандемной системы, а \hat{X}_i – общий входящий поток на i -й узел (см. рис. 1). Тогда согласно теоремам 1 и 2 потоки \hat{X}_i и Z_i являются MAP -потокami. Таким образом, i -й узел представляет собой систему массового обслуживания $MAP_i/PH_i/1/M_i$, интенсивность поступления пакетов в которой λ_i . Для данной системы хорошо известны формулы

расчета основных характеристик производительности [18]: средняя длина очереди m_i ; вероятность потери пакета $P_L^{(i)}$; среднее время пребывания (задержка) пакета на i -м узле T_i и др. Вычисление этих характеристик позволяет определить значения искомым параметров вероятности потери пакетов в тандемной системе

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

и время межконцевой задержки

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

Далее приведен формальный алгоритм нахождения стационарных характеристик производительности тандемной сети.

2.1. Аналитический алгоритм расчета стационарных характеристик производительности тандемной сети

Шаг 1. Положим $i := 1$.

Шаг 2. Если $i = 1$, то положим $\hat{X}_i = X_1$. Если же $i > 1$, то вычисляем \hat{X}_i : $\hat{X}_i = Z_{i-1}$, если в сети нет кросс-трафика, иначе $\hat{X}_i = Z_{i-1} \oplus X_i$. Обозначим матрицы потока \hat{X}_i как $\hat{\mathbf{D}}_{i,0}$ и $\hat{\mathbf{D}}_{i,1}$, т. е. $\hat{X}_i = \text{MAP}(\hat{\mathbf{D}}_{i,0}, \hat{\mathbf{D}}_{i,1})$.

Шаг 3. С помощью теоремы 1 вычисляем матрицы $\mathbf{D}'_{i,0}$, $\mathbf{D}'_{i,1}$ MAP-потока $Z_i = \mathcal{D}(\hat{X}_i, Y_i, M_i)$.

Шаг 4. Для выходящего MAP-потока Z_i вычисляем его стационарное распределение $\bar{\theta}^{(i)}$ с помощью системы линейных алгебраических уравнений

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

Шаг 5. Рассчитываем среднее число заявок в очереди на i -й фазе

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

где $V_i = |Y_i|$ – порядок PH-распределения Y_i , а $\hat{W}_i = |\hat{X}_i|$ – порядок входящего MAP-потока \hat{X}_i .

Шаг 6. Определяем стационарное распределение вероятностей $\bar{\pi}^{(i)}$ входящего потока \hat{X}_i . Если в сети нет кросс-трафика и $i > 1$, то полагаем $\bar{\pi}^{(i)} \equiv \bar{\theta}^{(i-1)}$. В противном случае находим $\bar{\pi}^{(i)}$ как решение системы линейных алгебраических уравнений

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

Шаг 7. С помощью найденного на предыдущем шаге стационарного распределения $\bar{\pi}^{(i)}$ входящего потока \hat{X}_i вычисляем интенсивность поступления заявок на i -й узел

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

Шаг 8. Рассчитываем распределение состояний входящего MAP-потока при наличии в системе $M_i + 1$ заявки (т. е. при заполненной системе)

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

Здесь вектор $\bar{\theta}_{M_i+1}^{(i)}$ – часть вектора $\bar{\theta}^{(i)}$, относящаяся к состояниям системы, когда в ней находится $M_i + 1$ заявка.

Шаг 9. Вычисляем вероятность потери заявки из-за переполнения i -й очереди

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

Шаг 10. Вычисляем среднюю задержку на i -м узле

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

Шаг 11. Если $i < N$, то увеличиваем $i := i + 1$ и переходим на шаг 2. В противном случае переходим далее, на шаг 12.

Шаг 12. Вычисляем вероятность потери заявки

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

Шаг 13. Вычисляем общую задержку $T = \sum_{i=1}^N T_i$.

2.2. Оценка сложности алгоритма нахождения точных характеристик производительности многофазной системы

Предложенная схема проста в вычислении. По сути, на каждом шаге с помощью нескольких операций произведения Кронекера строятся блочные матрицы для выходящего MAP-потока, а также, если в сети есть кросс-трафик, с помощью суммы Кронекера строятся матрицы входящего MAP-потока. Далее решаются две (если в сети есть кросс-трафик) или одна (в противном случае) системы линейных алгебраических уравнений для определения стационарных вероятностей входящего и исходящего потока. Наконец, с помощью нескольких операций умножения найденных распределений на матрицы потоков вычисляются искомые характеристики – вероятность потери пакета, средний размер системы и межконцевая задержка. Главный недостаток этой схемы расчета – чрезвычайно высокая вычислительная сложность.

Утверждение 1. Пусть входящие MAP-потоки имеют порядок W , PH-распределения – порядок V , емкость буфера на каждой фазе равна M и сеть содержит N станций. Тогда итерационная схема расчета характеристик тандемной сети имеет сложность:

– $O((MVW)^{3N})$, если в сети есть кросс-трафик;



Таблица 1

– $O(W^3(MV)^{3N})$, если кросс-трафика в сети нет.

Доказательство. Рассмотрим i -ю итерацию алгоритма, $i \leq N$, т.е. расчет характеристик на i -м узле сети. Отметим сперва, что при $i > 1$ порядок выходящего потока с предыдущего $i-1$ -го узла есть $(M+2)V\hat{W}_i$, где \hat{W}_i – порядок входящего потока на i -й узел. При наличии в сети кросс-трафика $U_i = ((M+2)VW)^i$, а если кросс-трафика нет, то $U_i = W((M+2)V)^i$.

Сложность итерации определяется шагами 4 и 6 алгоритма, на которых необходимо решать системы линейных алгебраических уравнений, причем порядок матрицы системы на шаге 4 (генератор выходящего потока) заведомо выше, чем системы на шаге 6 (генератор входящего потока). Полагая, что для решения системы используется алгоритм наподобие метода Гаусса, на шаге 4 потребуется $O(U_i^3)$ операций. Остальные шаги имеют более низкую сложность: шаги 1, 10 и 11 – $O(1)$, шаг 2 – $O(U_{i-1}^2W^2)$, шаг 3 – $O(U_i^2)$, шаг 5 – $O(VW+M)$, шаги 7 и 9 – $O(U_i^2)$, шаг 8 – $O(VM)$. Сложность шагов 12 и 13 есть $O(N)$. ♦

Таким образом, если в сети есть кросс-трафик, сложность алгоритма составит

$$O((VWM)^3) + O((VWM)^6) + \dots + O((VWM)^{3N}) + O(N) = O(VWM)^{3N},$$

а если кросс-трафика в сети нет, то

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

Таким образом, искать решение с помощью описанного алгоритма становится сложно даже при относительно небольших значениях N , V и W . В табл. 1 приведены примеры значений порядков МАР-потоков в зависимости от значений параметров системы. Из таблицы следует, что получение точного решения возможно лишь при числе узлов $N < 5$. Для практического применения тандемных систем большой размерности с узлами МАР/РН/1/М необходимы более эффективные методы расчета.

3. ПОЛУЧЕНИЕ ОЦЕНОК ХАРАКТЕРИСТИК ПРОИЗВОДИТЕЛЬНОСТИ ТАНДЕМНОЙ СЕТИ БОЛЬШОЙ РАЗМЕРНОСТИ

Как уже упоминалось выше, расчет характеристик производительности тандемной сети с применением аналитического подхода неэффективен, а в

Порядки выходящих МАР-потоков в зависимости от порядков РН-распределения (V), входящих МАР-потоков (W) и емкости буфера (M)

Параметры системы			Номер узла				
W	V	M	1	2	3	4	5
Сети без кросс-трафика							
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
Сети с кросс-трафиком							
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

некоторых случаях и невозможен в силу большой размерности МАР-потоков. Длительное время расчета стационарных состояний системы и стационарных характеристик производительности оказывает существенное влияние в итеративных задачах. Например, при проектировании беспроводной сети связи на стадии выбора топологии, когда на каждом новом шаге итерации необходимо производить оценку характеристик и выбирать оптимальный вариант. Для решения данной задачи предложен новый подход, базирующийся на комбинации методов имитационного моделирования и машинного обучения (рис. 2). В рамках данного метода на разных наборах входных параметрах с помощью имитационного моделирования генерируется набор данных, в котором рассчитаны характеристики производительности тандемной сети. С помощью сгенерированного набора данных реализуется алгоритм машинного обучения для получения быстрых оценок характеристик производительности. Метод эффективно применялся при решении задач теории очередей [28–30]. В настоящем разделе будет описана имитационная модель тандемной сети массового обслуживания, а также методика ее валидации с помощью аналитической модели; представлена методология получения оценок характеристик производительности тандемной сети с помощью комбинированного метода.

Расчет стационарных характеристик тандемной сети массового обслуживания с помощью метода имитационного моделирования состоит в том,

чтобы многократно промоделировать процесс генерации новых заявок и их обслуживания вплоть до потери заявки или окончания ее обслуживания. При этом нужно сохранять данные о числе сгенерированных и потерянных заявок, а также вычислять средние значения различных характеристик производительности, включая значения межконцевых задержек и вероятностей потерь заявок. Модель строится по дискретно-событийному принципу: обрабатываются только возникающие события (появление новых заявок и завершение обслуживания). Время между последовательными событиями меняется моментально, так как вне обработки событий состояние модели не меняется. При обработке событий можно рассчитывать и назначать моменты возникновения новых событий, которые произойдут при дальнейшем выполнении модели.

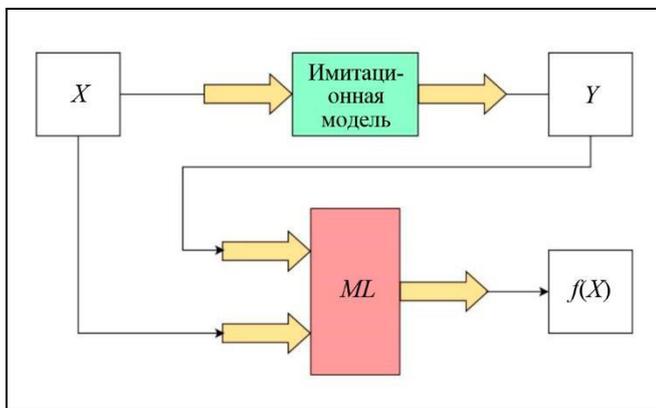


Рис. 2. Методология получения быстрых оценок стационарных характеристик производительности сети

3.1. Оценка характеристик производительности тандемной сети методом имитационного моделирования

В рамках данного исследования авторами разработана имитационная модель тандемной сети с входными МАР-потоками на языке Python. Некоторые модули имитационной модели были реализованы на языке C++ для улучшения производительности и эффективности расчетов. Для интеграции таких модулей в модель использовался язык программирования Cython. В процессе реализации использовалась библиотека PyQumo¹, предназначенная для построения и исследования стохастических моделей, в частности многофазных сетей массового обслуживания.

¹ URL: <https://github.com/ipu69/pyqumo> (дата обращения 16.09.2024).

Входными параметрами имитационной модели являются: размер тандемной сети (число узлов) N , емкость буфера на каждом из узлов $M_i, i = 1, \dots, N$, входной МАР-поток заявок, поступающий на первый узел сети, РН-распределения времени обработки заявок на каждом узле $PH_i, i = 1, \dots, N$. Помимо трафика, поступающего на первый узел сети, модель позволяет учитывать кросс-трафик, поступающий на каждый узел. В имитационной модели учтено, что если очередь узла пустая, то заявки сразу поступают на сервер обслуживания. С выхода каждой фазы заявки поступают на вход каждой последующей фазы, пока заявки не пройдут обслуживание на последней фазе. В случае успешной обработки заявки на последней фазе заявка считается успешно доставленной. В противном случае заявка считается потерянной.

Главный недостаток метода имитационного моделирования – сильная зависимость точности результатов от числа промоделированных событий. Например, при моделировании сети с десятью узлами для получения высокой точности с погрешностью в пределах 5% необходимо сгенерировать порядка 100 000 заявок. Таким образом, у данного метода оказывается весьма ограниченный потенциал к ускорению вычислений.

Для валидации разработанной имитационной модели на языке Python был реализован аналитический алгоритм расчета характеристик тандемной сети, описанный ранее в § 2. После чего было произведено сравнение оценок, полученных с помощью имитационного моделирования, с аналитическими расчетами. Валидация производилась на наборе из 430 различных случайных сетей, содержащих от 1 до 10 фаз. Применимость алгоритма точного расчета ограничена размерностью потоков обслуженных заявок (служащих входящими потоками на следующие фазы), которая растет для n -й фазы как $(M + 2)^N V^N W$. Наборы входных данных для валидации были сгенерированы таким образом, чтобы порядок выходящего потока с последней фазы не превосходил 8 000.

На рис. 3 показана зависимость относительной ошибки для различных характеристик сети массового обслуживания от числа заявок, поступающих в имитационную модель. Можно видеть, что относительная погрешность имитационной модели при расчете среднего числа пакетов на последней фазе больше, чем ошибка при вычислении межконцевой задержки и вероятности доставки для разного числа сгенерированных пакетов в процессе моде-

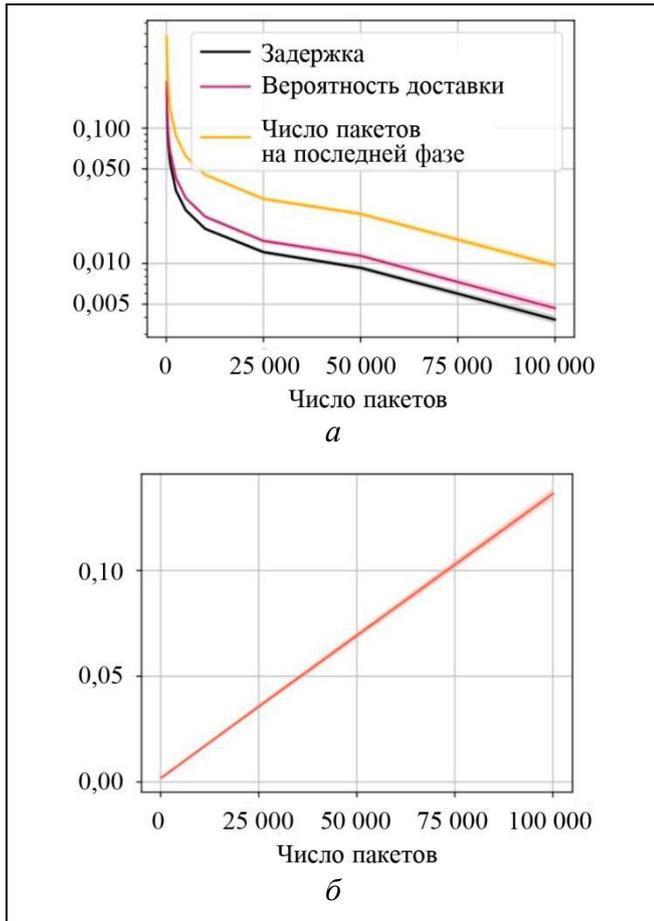


Рис. 3. Сходимость метода имитационного моделирования и скорость выполнения расчетов: a – оценка, b – длительность, c

лирования. При этом погрешности в пределах 5 % можно добиться, моделируя всего 25 000 заявок, а моделирование 100 000 заявок (которое осуществлялось при подготовке набора данных для обучения регрессионной модели) позволяет добиться того, чтобы ошибка не превышала 1 %.

3.2. Применение методов машинного обучения для получения оценок стационарных характеристик производительности тандемной сети

Для ускорения получения оценок характеристик производительности тандемной сети в настоящей работе предложено применять методы машинного обучения. В § 4 статьи на различных численных примерах продемонстрирована эффективность использования предложенной методологии для получения времени межконцевой задержки и вероятности доставки заявки в тандемной сети массового обслуживания.

С помощью имитационной модели рассчитывались следующие характеристики производительности сети:

- время межконцевой задержки в многофазной сети Δt ;
- вероятность доставки заявок $1 - P_L$.

В данной работе были построены две модели прогнозирования характеристик производительности сети. Первая – регрессионная модель для оценки средней межконцевой задержки, вторая – модель классификации для оценки вероятности успешной доставки заявок в сети. В рамках данной работы при построении моделей соблюдались ограничения:

- размер буфера на каждом узле одинаковый;
- распределение времени обслуживания заявки на каждом узле одинаково;
- в качестве входного потока используется частный случай МАР-потока с нулевым коэффициентом автокорреляции.

Учитывая введенные ограничения, для построения моделей использовались следующие параметры: первые три момента МАР-потока, первые три момента РН-распределения, размер сети N , размер буфера M .

В качестве признаков для характеристик входного МАР-потока и РН-распределения времени обслуживания были выбраны первые три момента. С помощью данных характеристик можно восстановить исходные распределения. Пусть $m_a = \mathbb{E}X$ – среднее время между поступлениями новых заявок, σ_a – стандартное отклонение. Аналогично пусть $m_s = \mathbb{E}Y$ – среднее время обслуживания заявки, а σ_s – его стандартное отклонение. Время обслуживания будем задавать средним значением m_s , коэффициентом вариации $c_s = \sigma_s / m_s$ и коэффициентом асимметрии $\gamma_s = \mathbb{E}[(Y - m_s)^3] / \sigma_s^3$. Входящий поток будем задавать аналогично с помощью параметров m_a , c_a и γ_a . Коэффициент автокорреляции для входящего МАР-потока, как было отмечено ранее, равен нулю. Тогда восстановление распределений для времени между поступлениями заявок и времени обслуживания можно произвести по первым трем моментам.

Для восстановления распределений в работе применялись два метода. Сначала была предпринята попытка построения ациклического РН-распределения второго порядка (АСРН(2), *acyclic continuous phase-type*) в каноническом виде (см. рис. 4, а) с помощью метода, описанного в работе [31]. В случае, когда моменты не попадали в область существования АСРН(2), применялся более общий метод, предложенный в работе [32], в котором РН-распределение состоит из двух распределений Эрланга $ME_n(2)$ (см. рис. 4, б). В общем

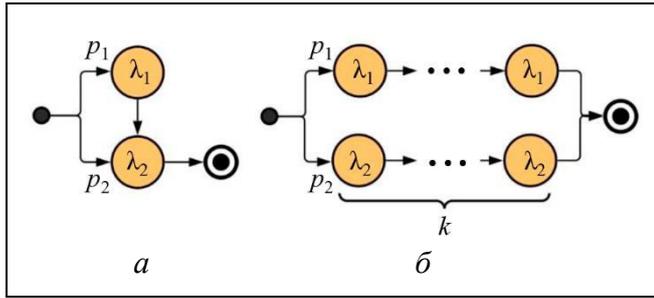


Рис. 4. Ациклическое РН-распределение с двумя состояниями АСРН(2) (а) и гипер-эрланговское распределение с двумя распределениями Эрланга одинакового порядка (б), используемые при аппроксимации потоков по трем моментам

случае РН-распределение можно найти для любого набора значений $m > 0$, $c > 0$ и $\gamma > c - 1/c$, которыми может обладать какое-либо непрерывное положительное распределение [32].

В рамках данной работы для прогнозирования межконцевой задержки применялись следующие алгоритмы и методы: метод наименьших квадратов, алгоритмы на деревьях (решающее дерево [33–36], градиентный бустинг [37]), а также искусственная нейронная сеть с использованием оптимизационного алгоритма Адама [38]. Также в работе проведен эксперимент, в котором тандемная сеть классифицировалась по уровню вероятности успешной доставки пакета $1 - P_L$. Для классификации задавалась некоторая величина B из интервала $(0, 1)$. Алгоритм определял, к какому из классов, $[0, B)$ или $[B, 1]$, принадлежала система. Для решения задачи классификации были применены следующие методы: логистическая регрессия, дерево решений, градиентный бустинг, а также искусственная нейронная сеть с использованием алгоритма оптимизации Адама.

4. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Для получения оценок характеристик производительности тандемной сети использовались синтетические данные, сгенерированные с помощью имитационного моделирования. На вход имитационной модели поступали данные, сгенерированные случайным образом в диапазонах, приведенных в табл. 2. На выходе были получены среднее время межконцевой задержки и средняя вероятность доставки заявки $\{\Delta t, 1 - P_L\}$. С помощью имитационного моделирования были получены данные объемом 101 424 строк.

Так как данные генерировались случайным образом, в выборке присутствуют значения коэффициента загрузки $\rho = m_s / m_a$, существенно выходя-

Таблица 2

Входные параметры имитационной модели

Наименование параметра	Диапазон значений
Время между поступлениями заявок	
Среднее значение m_a	$\sim (0, 10)$
Коэффициент вариации c_a	$\sim (0, 5; 3)$
Коэффициент асимметрии γ_a	$\sim (c_a - \frac{1}{c_a}, 100)$
Время обслуживания заявок на фазах	
Среднее значение m_s	$\sim (0, 10)$
Коэффициент вариации c_s	$\sim (0, 5; 3)$
Коэффициент асимметрии γ_s	$\sim (c_s - \frac{1}{c_s}, 100)$
Размер буфера очередей на фазах M	$\{6, 7; \dots, 10\}$
Число фаз в сети N	$\{1, 2, \dots, 20\}$

щие за пределы диапазона $(0, 1)$. Использовать такие данные для дальнейшего обучения моделей нецелесообразно. Перед тем как начать исключать строки, отметим, что узлы многофазной очереди имеют конечный буфер и заявки при загруженном узле могут теряться. Таким образом, загрузка на первом узле может быть $\rho \gg 1$. Поэтому ограничимся диапазоном $\rho \in (0, 10]$. После исключения выбросов объем выборки составил 96 248 строк.

4.1. Оценка межконцевых задержек

Для оценки прогнозных значений будем использовать следующие метрики: коэффициент корреляции R , стандартное отклонение STD и коэффициент детерминации R^2 .

На примере размерностей сетей $N = 1, 5, 10, 20$ рассмотрим полученные оценки межконцевой задержки. На рис. 5 представлена структура искусственной нейронной сети с одним скрытым слоем, число нейронов в которой равно 40. Для нейронов в скрытом слое использовалась сигмоидальная функция активации. Было задано 1000 эпох для обучения.

Плотности распределения всех оценок межконцевых задержек на тестовой выборке представлены на рис. 6. Худшим прогнозом оказались результаты, полученные с помощью МНК. На графике плотность распределения показана зеленым цветом. В силу линейности аппроксимации большинство оценок получились отрицательными. Для дерева решений в качестве критерия разбиения на каждом узле дерева использовалась среднеквадратичная ошибка. Наилучшие оценки были получе-

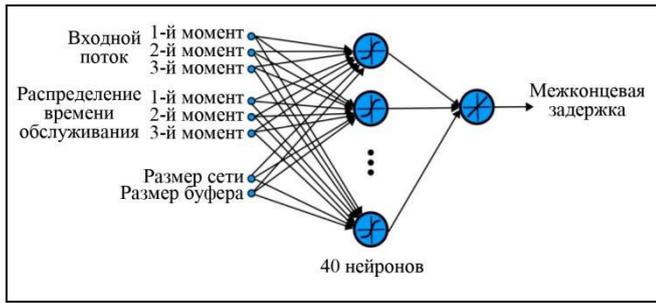


Рис. 5. Структура нейронной сети для предсказания межконцевой задержки

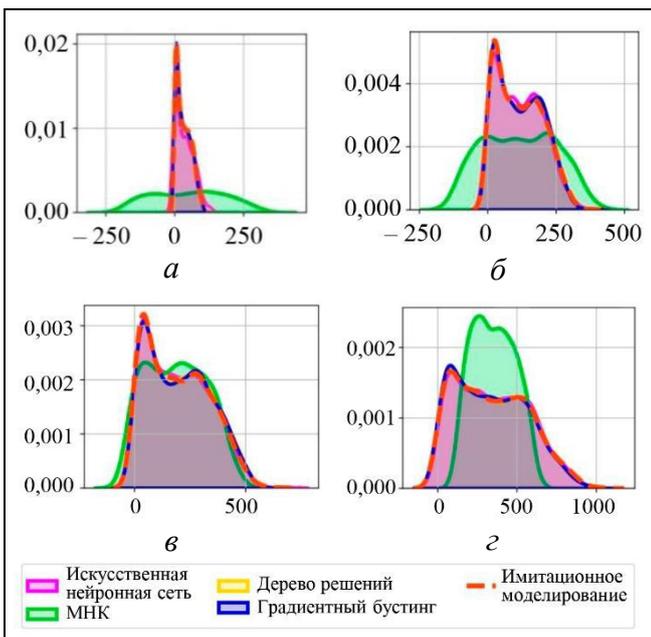


Рис. 6. Плотность распределения вероятности значений межконцевой задержки: а – длина сети 1, б – длина сети 5, в – длина сети 10, з – длина сети 20

ны при глубине дерева, равной 36. Для метода градиентного бустинга на решающем дереве практически путем была выбрана глубина дерева, равная 10. Для рассматриваемого случая скорость обучения задавалась равной 0,1, а число деревьев – равным 100. На рис. 7 представлены диаграммы рассеяния всех обученных моделей. Как видно из табл. 3, наилучшими показателями качества обладает прогнозная модель, построенная на нейронной сети.

4.2. Оценка вероятности успешной доставки

В отличие от оценок времени межконцевой задержки, для которой строились регрессионные модели, модель вероятности доставки не обязательно должна предсказывать конкретные значения. Гораздо важнее дать прогноз, является ли доставка заявки успешной или нет. Зададим граничное ус-

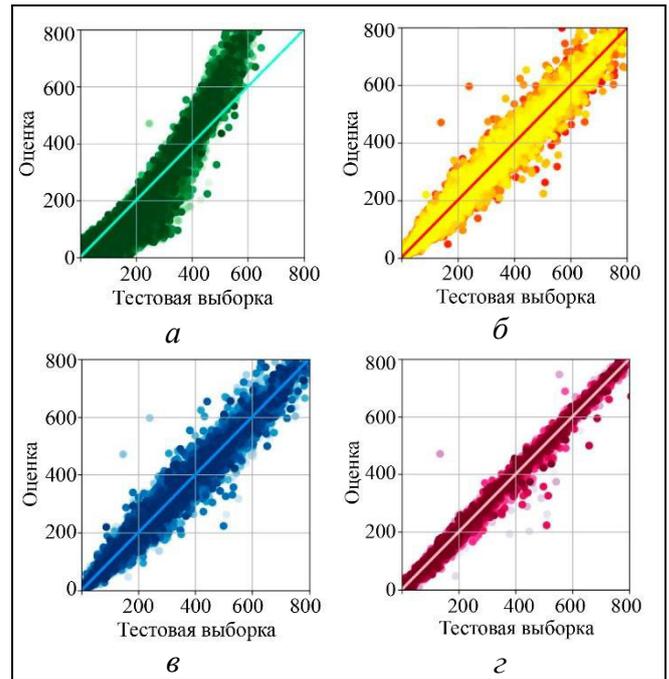


Рис. 7. Диаграммы рассеяния регрессионных моделей: а – МНК, б – дерево решений, в – градиентный бустинг, з – искусственная нейронная сеть

Таблица 3

Значения метрик для различных алгоритмов предсказания межконцевой задержки в сети

Модель	Метрики		
	R	STD	R^2
МНК	0,926	66,31	0,85
Дерево решений	0,990	24,48	0,98
Градиентный бустинг	0,990	24,71	0,98
Нейронная сеть	0,998	12,23	0,99

ловие B для формулировки условия успешной доставки. Будем классифицировать вероятности доставки на две группы: успешная доставка $1 - P_L \in [B, 1]$ и потеря заявок $P_L \in [0, B)$. Для всех моделей в проводимом численном эксперименте зададим границу $B = 0,9$. Для оценки моделей будем использовать следующие метрики: Precision, Recall и F_1 .

Рассмотрим вновь многофазную сеть с длинами $N = 1, 5, 10, 20$ и на примере искусственной нейронной сети оценим вероятность успешной доставки. Структура нейронной сети представляет собой многослойный персептрон с тремя скрытыми слоями по 16 нейронов в каждом (рис. 8). В качестве функции активации выбрана сигмоида.

Логистическая регрессия предсказывает значения успешной доставки с большой долей ошибок.

Дерево решений с глубиной, равной 10, показало несколько лучшие результаты. Наилучшие прогнозы для задачи классификации получены с помощью метода градиентного бустинга и искусственной нейронной сети. Все рассчитанные значения метрик классификации представлены в табл. 4.

На рис. 9 представлены оценки успешной доставки всех моделей для размера сети $N = 10$. Трендом указаны действительные значения вероятности $1 - P_L$ для различных коэффициентов загрузки ρ . Зеленым цветом представлены прогнозы успешной доставки ($1 - P_L \geq B$), красным цветом – вероятности неуспешной доставки ($1 - P_L < B$). Здесь можно выделить градиентный бустинг и

нейронную сеть как модели, предоставляющие наилучшие прогнозы.

Таблица 4

Значени метрик для моделей классификации успешной доставки пакета

Модель	Метрики		
	Precision	Recall	F_1
Логистическая регрессия	0,804	0,821	0,813
Дерево решений	0,9618	0,905	0,912
Градиентный бустинг	0,966	0,969	0,968
Нейронная сеть	0,977	0,951	0,964

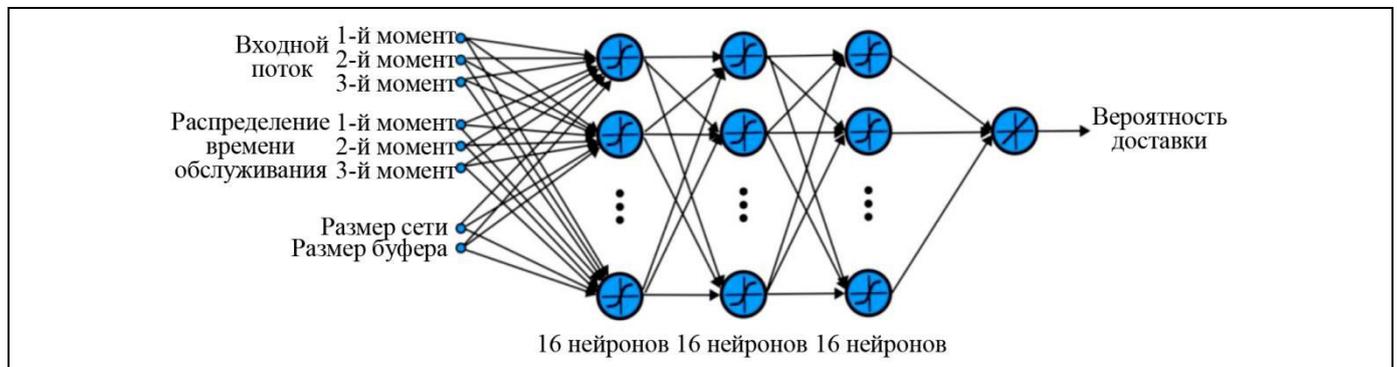
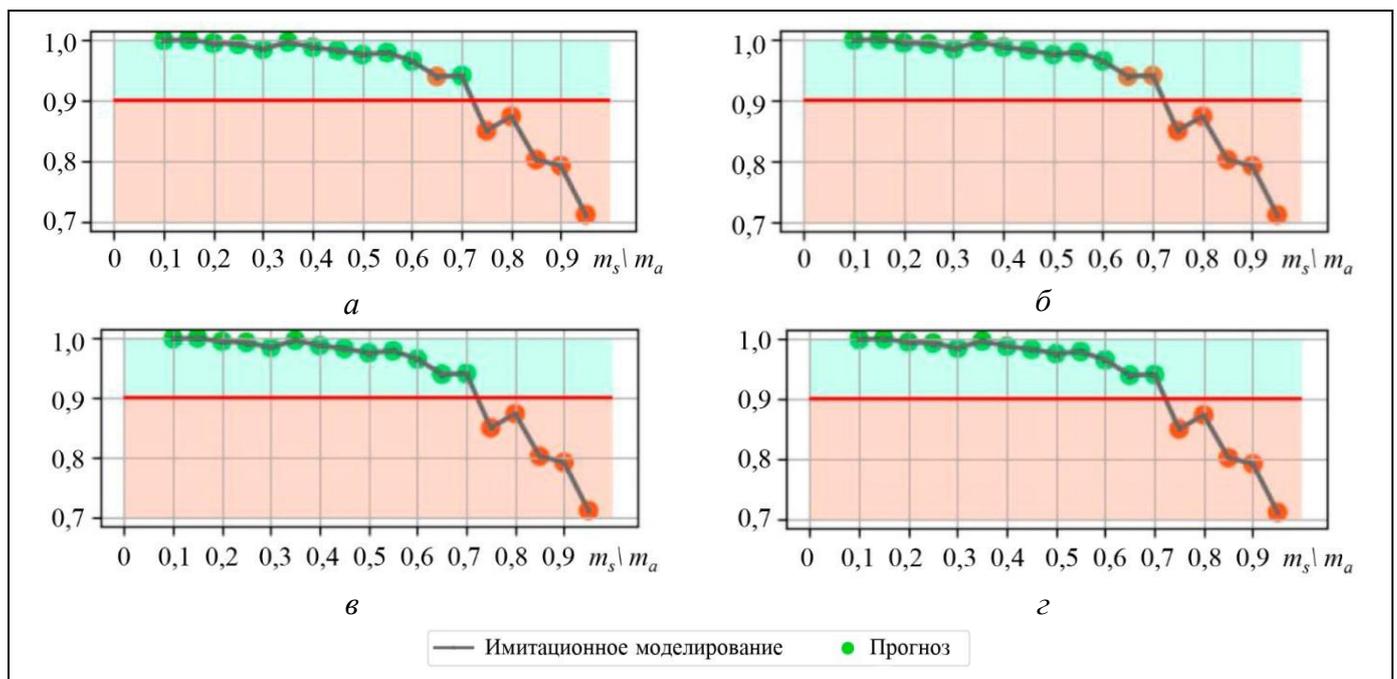


Рис. 8. Структура нейронной сети для классификации системы по вероятности доставки пакета


 Рис. 9. Оценки вероятности доставки, полученные с помощью искусственной нейронной сети: длина тандема 10, a – логистическая регрессия, b – дерево решений, v – градиентный бустинг, z – искусственная нейронная сеть



4.3. Анализ времени выполнения расчетов

Для полученных моделей в табл. 5 представлено время расчета стационарных характеристик производительности различными методами. Для сравнения представлено также время расчета имитационной модели в зависимости от коэффициента загрузки системы. Объем выборки составляет 360 строк. Целесообразно сравнивать среднее время расчета для набора данных, так как для единичных случаев время будет варьироваться в зависимости от длины тандема. Необходимо отметить, что рассчитанное время представлено как время, которое прошло от начало расчета и до получения конечного результата. Очевидно, что при каждой новой прогонке моделей время будет варьироваться также в зависимости от занятости процессора рабочей машины. Кроме того, имитационная модель рассчитывает сразу две характеристики производительности сети. Отметим, что время расчета моделей машинного обучения на несколько порядков ниже, чем время выполнения имитационной модели.

Таблица 5

Время получения характеристик производительности

Модель	Время выполнения расчетов, с
Оценка времени отклика	
Имитационная модель	172,2
МНК	$4,8 \cdot 10^{-6}$
Дерево решений	$5,5 \cdot 10^{-6}$
Градиентный бустинг	$5 \cdot 10^{-6}$
Нейронная сеть	$5,7 \cdot 10^{-6}$
Вероятность успешной доставки	
Имитационная модель	172,2
Логистическая регрессия	$5,3 \cdot 10^{-6}$
Дерево решений	$5 \cdot 10^{-6}$
Градиентный бустинг	$4,8 \cdot 10^{-6}$
Нейронная сеть	$5,3 \cdot 10^{-6}$

ЗАКЛЮЧЕНИЕ

В настоящей статье разработаны новые подходы к исследованию характеристик производительности тандемных сетей. Дано описание алгоритма точного расчета сети малой размерности, оценки сложности этого алгоритма и его применения для валидации имитационной модели. Для исследования сетей большой размерности предложен эффек-

тивный подход, базирующийся на комбинации методов имитационного моделирования и машинного обучения. Проведен сравнительный анализ применения различных методов машинного обучения, включая деревья решений, градиентный бустинг, искусственные нейронные сети и др. На численных примерах показаны высокая эффективность применения методов машинного обучения, а также резкое сокращение времени расчета.

ЛИТЕРАТУРА

1. *Khayyati, S., 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. – P. 5176–5200. – DOI: <https://doi.org/10.1080/00207543.2021.1951448>.
2. *Kumar, B.K., Sankar, R., Krishnan, R.N., 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. – P. 95–142. – DOI: <https://doi.org/10.1007/s11009-020-09842-6>.
3. *Oblakova, A., Al Hanbali, A., Boucherie, R.J., et al.* An Analytical Model for a Tandem of Two Traffic-Light Intersections under Semi-actuated and Fixed Control // *Transportation Research Interdisciplinary Perspectives*. – 2022. – Vol. 16, no. 3. – Art. no. 100715. – DOI: <https://doi.org/10.1016/j.trip.2022.100715>.
4. *Rovetto, C., Cruz, E., Nuñez, I., et al.* 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. – DOI: <https://doi.org/10.3390/app131810030>.
5. *Pershin, O.Y., Mukhtarov, A.A., Vishnevsky, V.M., Larionov, A.A.* Optimal Placement of Base Stations in Integrated Design of Wireless Networks // *Programming and Computer Software*. – 2023. – Vol. 49, suppl. 2. – P. S82–S90. – DOI: <https://doi.org/10.1134/S0361768823100055>.
6. *Vishnevsky, V., Krishnamoorthy, A., Kozyrev, D., 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. – P. 329–342. – DOI: <https://doi.org/10.1007/s13226-016-0190-7>.
7. *Gnedenko, B.W., König, D.* *Handbuch der Bedienungstheorie II*. – Berlin: De Gruyter, 1984. – 608 p. – DOI: <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. – P. 123–142. – DOI: <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., Chakravarthy, S.R.* Analysis of a Tandem Queueing 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., 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., 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. – P. 5606–5613. – DOI: <https://doi.org/10.1007/s10958-005-0432-4>.
 13. *Kim, C.S., Klimenok, V., Taramin, O.* A Tandem Retrial Queueing System with Two Markovian Flows and Reservation of Channels // *Computers and Operations Research*. – 2010. – Vol. 37, no. 7. – P. 1238–1246. – DOI: <https://doi.org/10.1016/j.cor.2009.03.030>.
 14. *Kim, C., Klimenok, V.I., 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. – P. 61–71. – DOI: <https://doi.org/10.1016/j.cie.2016.03.012>.
 15. *Klimenok, V., Breuer, L., Tsarenkov, G., Dudin, A.* The BMAP/G/1/→/PH/1/M Tandem Queue with Losses // *Performance Evaluation*. – 2005. – Vol. 61, no. 1. – P. 17–40. – DOI: <https://doi.org/10.1016/j.peva.2004.09.001>.
 16. *Lian, Z., Liu, L.* A Tandem Network with MAP Inputs // *Operations Research Letters*. – 2008. – Vol. 36, no. 2. – P. 189–195. – DOI: <https://doi.org/10.1016/j.orl.2007.04.004>.
 17. *Vishnevskii, V.M., 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. – P. 1361–1403. – DOI: <https://doi.org/10.1134/S000511791708001X>.
 18. *Dudin, A.N., Klimenok, V.I., Vishnevsky, V.M.* The Theory of Queueing Systems with Correlated Flows // *Cham: Springer*, 2020. – 410 p.
 19. *Bruell, S.C., Balbo, G., Afshari, P.V.* Mean Value Analysis of Mixed, Multiple Class BCMP Networks with Load Dependent Service Stations // *Performance Evaluation*. – 1984. – Vol. 4, iss. 4. – P. 241–260. – DOI: [https://doi.org/10.1016/0166-5316\(84\)90010-5](https://doi.org/10.1016/0166-5316(84)90010-5).
 20. *Vishnevsky, V., Klimenok, V., Sokolov, A., 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. – DOI: <https://doi.org/10.3390/math9243236>.
 21. *Klimenok, V., Dudin, A., Vishnevsky, V.* On the Stationary Distribution of Tandem Queue Consisting of a Finite Number of Stations // *Communications in Computer and Information Science*. – 2012. – Vol. 291. – P. 383–392. – DOI: https://doi.org/10.1007/978-3-642-31217-5_40.
 22. *Palomo, S., Pender, J.* Learning the Tandem Network Lindley Recursion // *Proceedings of 2021 Winter Simulation Conference (WSC)*. – Phoenix, 2021. – P. 1–12. – DOI: <https://doi.org/10.1109/WSC52266.2021.9715530>.
 23. *Rabta, B.* A Review of Decomposition Methods for Open Queueing Networks // *Rapid Modelling for Increasing Competitiveness*. Ed. by G. Reiner – London: Springer, 2009. – P. 25–42. – DOI: https://doi.org/10.1007/978-1-84882-748-6_3.
 24. *Vishnevsky, V., Larionov, A., Roman, I., 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 11th IEEE International Conference on Application of Information and Communication Technologies*. – Moscow, 2017. – P. 1–5. – DOI: <https://doi.org/10.1109/ICAICT.2017.8687247>.
 25. *Gorbulnova, A.V., Vishnevsky, V.M., Larionov, A.A.* Evaluation of the End-to-End Delay of a Multiphase Queueing System Using Artificial Neural Networks // *Lecture Notes in Computer Science*. – Cham: Springer, 2021. – Vol. 12563. – P. 631–642. – DOI: https://doi.org/10.1007/978-3-030-66471-8_48.
 26. *Kudou, T., Nii, S., Okuda, T.* A Performance Evaluation of Tandem Queueing Systems by Machine Learning // *Proceedings of 2022 IEEE International Conference on Consumer Electronics*. – Taiwan, 2022. – P. 389–390. – DOI: <https://doi.org/10.1109/ICCE-Taiwan55306.2022.9869030>.
 27. *Kudou, T., Okuda, T.* A Time Series Analysis of Single Server Queueing Systems by Using Machine Learning // *Proceedings of 2023 International Conference on Consumer Electronics*. – Taiwan, 2023. – P. 327–328. – DOI: <https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226861>.
 28. *Вишневецкий В.М.* Теоретические основы проектирования компьютерных сетей. – М: Техносфера, 2003. – 512 с. [*Vishnevsky, V.M.* Theoretical foundations of computer network design. Technosfera, Moscow. – 2003. – 512 p. (In Russian)]
 29. *Vishnevsky, V.M., Klimenok, V.I., Sokolov, A.M., 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. – DOI: <https://doi.org/10.3390/math12050659>.
 30. *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. – Art. no. 5479. – DOI: <https://doi.org/10.3390/s23125479>.
 31. *Telek, M., 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. – P. 47–57.
 32. *Johnson, M.A., 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. – P. 711–743. – DOI: <https://doi.org/10.1080/15326348908807131>.
 33. *Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.* Classification and Regression Trees. – New York: Chapman and Hall, 1984. – 368 p. – DOI: <https://doi.org/10.1201/9781315139470>.
 34. *Demidova, L.A., Usachev, P.O.* Development and Approbation of the Improved CART Algorithm Version // *Journal of Physics: Conference Series*. – 2020. – Vol. 1479. – Art. no. 012085. – DOI: <https://doi.org/10.1088/1742-6596/1479/1/012085>.
 35. *Gordon, A.D., Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.* Classification and Regression Trees // *Biometrics*. – 1984. – Vol. 40, no. 3. – P. 874. – DOI: <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. – P. 14–23. – DOI: <https://doi.org/10.1002/widm.8>.
 37. *Friedman, J.H.* Stochastic Gradient Boosting // *Computational Statistics and Data Analysis*. – 2002. – Vol. 38, no. 4. – P. 367–378. – DOI: [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).



38. Kingma, D.P., Ba, J.L. Adam: A Method for Stochastic Optimization // arXiv:1412.6980. – 2014. – URL: <http://arxiv.org/abs/1412.6980>.

Статья представлена к публикации членом редколлегии
Р. В. Мецзяковым.

Поступила в редакцию 01.08.2024,
после доработки 06.09.2024.
Принята к публикации 13.09.2024.

Вишневский Владимир Миронович – д-р техн. наук,
✉ vishn@inbox.ru
ORCID iD: <https://orcid.org/0000-0001-7373-4847>

Ларионов Андрей Алексеевич – канд. техн. наук,
✉ larioandr@gmail.com
ORCID iD: <https://orcid.org/0000-0003-0539-0442>

Мухтаров Амир Амангельдыевич – канд. техн. наук,
✉ mukhtarov.amir.a@gmail.com,
ORCID iD: <https://orcid.org/0000-0002-8191-6381>

Соколов Александр Михайлович – науч. сотрудник,
✉ aleksandr.sokolov@phystech.edu,
ORCID iD: <https://orcid.org/0000-0002-3589-5700>

Институт проблем управления им. В. А. Трапезникова РАН,
г. Москва

© 2024 г. Вишневский В. М., Ларионов А. А., Мухтаров А. А.,
Соколов А. М.



Эта статья доступна по [лицензии Creative Commons](https://creativecommons.org/licenses/by/4.0/)
«Attribution» («Атрибуция») 4.0 Всемирная.

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.

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