

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ТЕЗИСЫ
I Международной конференции
ШКОЛА ТЕОРИИ МАССОВОГО
ОБСЛУЖИВАНИЯ (ШТМО-2025)
21–26 апреля 2025 г.

ТОМСК
Издательство Томского
государственного университета
2025

УДК 519.872

ББК 22.18

И74

Школа теории массового обслуживания (ШТМО-2025): тезисы I Международной конференции (21–26 апреля 2025 г.). — Томск: Издательство Томского государственного университета, 2025. — 78 с.

ISBN 978-5-908040-02-0

Сборник содержит тезисы Международной конференции «Школа теории массового обслуживания (ШТМО-2025)» по следующим направлениям: методы стохастического и имитационного моделирования телекоммуникационных, экономических и технических систем.

Для специалистов в области информационных технологий и математического моделирования.

УДК 519.872

ББК 22.18

Р е д к о л л е г и я:

С. П. Моисеева, доктор физико-математических наук, профессор

Е. А. Фёдорова, кандидат физико-математических наук

О. Д. Лизюра

ISBN 978-5-908040-02-0

© Авторы (текст), 2025

© Томский государственный
университет (оформление,
дизайн), 2025

NATIONAL RESEARCH TOMSK STATE UNIVERSITY

**ABSTRACTS
of the 1st International Conference
QUEUEING THEORY SCHOOL (QTS-2025)
2025 April, 21–26**

TOMSK
Tomsk State
University Publishing
2025

UDC 519.872

LBC 22.18

I60

Queueing theory school (QTS-2025): Abstracts of the 1st International Conference (2025 April, 21–26). — Tomsk: Tomsk State University Publishing, 2025. — 78 p.

ISBN 978-5-908040-02-0

This volume presents abstracts from the 1st International Conference "Queueing theory school (QTS-2025)". The papers are devoted to new results in the following areas: stochastic modeling and simulation of telecommunication, economical and technical systems.

UDC 519.872

LBC 22.18

E d i t o r s:

S. P. Moiseeva, Doctor of Physical and Mathematical Sciences,
Professor

E. A. Fedorova, Candidate of Physical and Mathematical Sciences

O. D. Lizyura

ISBN 978-5-908040-02-0

© Authors (Text), 2025

© Tomsk State University

(Publishing, Design), 2025

Пленарные доклады

НАПРАВЛЕНИЯ РАЗВИТИЯ ТЕОРИИ ОЧЕРЕДЕЙ И ЕЁ ПРИМЕНЕНИЕ ПРИ ПРОЕКТИРОВАНИИ ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЕЙ

В. М. Вишневский¹

¹ Институт проблем управления им. В. А. Трапезникова РАН, г. Москва,
Россия

В настоящем докладе дано описание основных направлений и этапов развития теории очередей и их тесной связи с прогрессом в области телекоммуникационных сетей. Следует отметить, что выделение этапов является достаточно условным, но оно отражает основные тенденции развития теории очередей от момента её становления до настоящего времени. На конкретных примерах проектирования телекоммуникационных систем и сетей обоснована необходимость и эффективность применения моделей теории очередей.

Первый этап связан с появлением в конце 19-го в. телефонных сетей и, соответственно, проектированием автоматических телефонных станций (АТС). Впервые возникла проблема доступа к ограниченному коммуникационному ресурсу. Её разрешил русский инженер М.Ф. Фрейденберг, представивший инженерный расчёт по оценке производительности АТС в зависимости от числа абонентов и вероятности отказа в обслуживании абонентов. Строгая математическая модель была разработана основоположником теории очередей датским математиком и инженером А.К. Эрлангом в рамках его работ по исследованию телефонных сетей в период 1909–1922 гг. В дальнейшем развитие теории очередей шло в направлении усложнения моделей СМО. Появились многочисленные статьи и книги по исследованию многолинейных и приоритетных СМО, систем стохастического поллинга и т.д., а также СМО с пуссоновским входным потоком и произвольной функцией распределения времени обслуживания. Были развиты методы анализа стационарных характеристик таких систем, включая метод вложенных цепей Маркова, метод введения дополнительного события и т.д.

Второй этап развития теории очередей связан с появлением и широким внедрением компьютерных сетей, в которых использовался эффективный метод коммутации пакетов в отличие от метода коммутации каналов, применявшегося в телефонных сетях. Возникла проблема разработки нового математического аппарата – исследование моделей

сетей массового обслуживания (СеМО). Решению этой проблемы посвящены работы американского учёного Леонарда Клейнрока [1]. Его исследования по открытым СеМО эффективно использовались для синтеза топологической структуры, управления маршрутизацией, выбора оптимальных параметров сетевых протоколов и т.д. В дальнейшем результаты Д. Клейнрока были расширены и обобщены на класс замкнутых, открытых и смешанных СеМО с мультипликативным представлением вероятности состояний, удовлетворяющих ограничениям теоремы ВСМР.

Новый импульс развития теоретических исследований (3-й этап) связан с появлением цифровых сетей интегрального обслуживания, информационные потоки в которых являются не стационарными и главное – коррелированными. В связи с этим в 80-е гг. возникла проблема разработки математической модели коррелированных потоков, адекватно описывающих информационные потоки в технических и социальных сетях нового поколения. Эта проблема была решена двумя группами научных коллективов – американского под руководством М. Ньютона и российского под руководством Г.П. Башарина, предложивших в качестве такой модели МАР-потоки и их обобщение ВМАР, ММАР и т.д. До настоящего времени ведутся активные исследования СМО с коррелированными входящими потоками. В 2000 г. появилась одна из первых книг этого направления [2], в которой систематизированы методы и подходы анализа СМО с коррелированными входными потоками.

Однако до настоящего времени осталось множество задач по исследованию СМО и СеМО большой размерности с входящим МАР-потоком, для решения которых применение традиционных методов и алгоритмов отыскания численных решений либо затруднено, либо во все невозможно. Поэтому современный этап развития теории очередей характеризуется разработкой новых методов, включая методы машинного обучения. В связи с этим в заключительной части доклада, будет рассмотрен пример проектирования современной широкополосной беспроводной сети с использованием модели многофазной СМО большой размерности, с входящим МАР-потоком и РН-распределением времени обслуживания на фазах, для исследования которой эффективно применяются методы машинного обучения [2, 3].

СПИСОК ЛИТЕРАТУРЫ

1. Клейнрок Л. Вычислительные системы с очередями. М.: Мир, 1979. 600 с.
2. Вишневский В. М., Дудин А. Н., Клименок В. И. Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях. М.: ТЕХНОСФЕРА, 2018. 564 с.

3. *Вишневский В. М., Ефросинин Д. В.* Теория очередей и машинное обучение. М.: ИНФРА-М, 2025. 370 с.

ПАРАДОКС ВОССТАНОВЛЕНИЯ И ПРОЦЕССЫ ОБСЛУЖИВАНИЯ

А. В. Зорин¹

¹ Национальный исследовательский Нижегородский государственный университет им. Н. И. Лобачевского, г. Нижний Новгород, Россия

В теории случайных процессов часто рассматривается следующая конструкция. Пусть X_1, X_2, \dots — независимые, одинаково распределенные случайные величины с функцией распределения $B(x)$ (она предполагается абсолютно непрерывной):

$$S_0 = 0, \quad S_n = X_1 + X_2 + \dots + X_n,$$

$N(t) = \max\{n : S_n \leq t\}$ — число восстановлений. Вводят так называемые функционалы в процессе восстановления (см.: [1]):

$$\gamma_r(t) = S_{N(t)+1} - t, \quad \gamma_e(t) = t - S_{N(t)}$$

есть оставшееся время до следующего после момента t восстановления и прошедшее время с последнего перед моментом t восстановления. В теории восстановления *парадоксом восстановления* называется утверждение о том, что закон распределения величины $\gamma_e(t) + \gamma_r(t)$ отличается от закона распределения, порожденного ф.р. $B(x)$ (см., например, [2]). Совместное распределение $H_t(u, v) = \mathbb{P}(\gamma_e(t) > u, \gamma_r(t) > v)$ изучалось, например, в [1]. Однако следующая теорема в такой формулировке не встречалась автору ни в теоретических монографиях, ни в сборниках задач. Фактически она позволяет строго вычислять вероятность скачка марковского процесса $\{\gamma_e(t); t \geq 0\}$.

Теорема 1. Пусть существует плотность $U'(t)$ для функции восстановления $U(t) = 1 + \sum_{k=1}^{\infty} B^{k*}(t)$, $t \geq 0$. Тогда имеет место равенство $\mathbb{P}(\gamma_r(t) < v | \gamma_e(t) = u) = (B(u+v) - B(u))/(1 - B(u))$.

Необходимость изучения условной вероятности такого вида возникает, когда в системе массового обслуживания с входными потоками рекуррентного вида вводят, следя Коксу [3], дополнительную компоненту вида оставшегося времени до поступления следующего требования. Обычно, напрекор парадоксу восстановления, в ходе рассуждений часто возникает утверждение: «поскольку интервал, накрывающий момент t , имеет функцию распределения $B(x)$, то...». Итоговая формула оказывается верной, но промежуточное утверждение неверно.

Оказывается, строго вычислить вероятность скачка для такой дополнительной компоненты типа прошедшего времени с последнего события, когда известно это прошедшее время в момент t , уже составляет трудоемкую проблему. Мы продемонстрируем трудности на примере системы типа $M/G/1/\infty$.

Пусть в однолинейную систему поступает простейший поток с интенсивностью $\lambda > 0$, очередь без ограничений размера, длительности обслуживания требований н.о.р. с ф.р. $B(x)$. Пусть $L(t)$ — число требований в системе в момент t . В начальный момент очередь пуста, $L(0) = 0$. Требования обслуживаются в порядке поступления согласно дисциплине FIFO.

Следующие случайные величины считаются заданными или определяются явно:

- моменты поступления требований, $\tau_n^{(i)}$, $n = 1, 2, \dots$, таковы, что $(\tau_n^{(i)} - \tau_{n-1}^{(i)})$, $n = 1, 2, \dots$, суть независимые, с показательным распределением с параметром λ и $\tau_0^{(i)} = 0$ (верхний символ i означает input, «входные»);
- длительности обслуживания S_n , $n = 1, 2, \dots$, — н.о.р. с ф.р. $B(x)$;
- длительности ожидания (рекуррентные уравнения Линдли) и моменты выхода

$$W_1 = 0, \quad W_n = \max\{0, W_{n-1} + S_{n-1} - (\tau_n^{(i)} - \tau_{n-1}^{(i)})\}, \\ \tau_n^{(o)} = \tau_n^{(i)} + W_n + S_n,$$

т.е. первое требование не ждет, поступает в пустую систему;

- считающий процесс входного потока

$$\eta(t) = \max\{n : \tau_n^{(i)} \leq t\}, \quad t \geq 0,$$

есть классический процесс Пуассона с непрерывными справа траекториями, $\eta(t)$ считает требования на $[0, t]$;

- считающий процесс выходного потока

$$\xi(t) = \max\{n : \tau_n^{(o)} \leq t\}, \quad t \geq 0,$$

(верхний индекс o означает output, «выходной»), так же непрерывен справа.

Тогда число требований в системе в момент t равно $L(t) = \eta(t) - \xi(t)$ и $\{L(t); t \geq 0\}$ — процесс с непрерывными справа траекториями. Кроме того, введем прошедшее и оставшееся времена для обслуживания требования, находящегося на приборе в момент t :

$$\gamma_e(t) = \max\{0, t - (\tau_{\xi(t)+1}^{(i)} + W_{\xi(t)+1})\}, \quad \gamma_r(t) = \tau_{\xi(t)+1}^{(o)} - t.$$

Обозначим

$$G_n(t, x) = \mathbb{P}(\tau_n^{(o)} < t, L(\tau_n^{(o)}) = x), \quad g_n(t, x) = \frac{d}{dt}g_n(t, x).$$

Теорема 2. Имеют место соотношения ($u > 0, v > 0$):

$$\begin{aligned} & \mathbb{P}(L(t) = x, \gamma_e(t) > v, \gamma_r(t) > u) = \\ & = \lambda e^{-\lambda t} \int_0^{t-v} (1 - B(t+u-y)) \frac{(\lambda(t-y))^{x-1}}{(x-1)!} dy + \\ & + \sum_{n=1}^{\infty} \left(\int_0^{t-v} \left(\lambda e^{-\lambda(t-y)} \int_y^{t-v} (1 - B(u+t-z)) \frac{(\lambda(t-z))^{x-1}}{(x-1)!} dz \right) d_y G_n(y, 0) + \right. \\ & \left. + \sum_{l=1}^x \int_0^{t-v} \left((1 - B(u+t-y)) \frac{(\lambda(t-y))^{x-l}}{(x-l)!} e^{-\lambda(t-y)} \right) d_y G_n(y, l) \right), \\ & \mathbb{P}(\gamma_r(t) > u \mid L(t) = x, \gamma_e(t) = v) = \frac{1 - B(u+v)}{1 - B(u)}, \quad x = 1, 2, \dots; \\ & \mathbb{P}(\gamma_r(t) > u \mid L(t) > 0, \gamma_e(t) = v) = \frac{1 - B(u+v)}{1 - B(u)}. \end{aligned}$$

Как отмечают (без строгого доказательства) немногие авторы (см., например, [4]), распределение величины $\gamma_e(t) + \gamma_r(t)$ при условии $L(t) > 0$ отличается от распределения $B(x)$. Это также легко видеть, используя компьютерную имитационную модель. К сожалению, гораздо чаще при выводе дифференциальных уравнений Колмогорова в модели с дополнительной компонентой явно делается противоположное предположение о совпадении условного распределения с распределением $B(x)$.

СПИСОК ЛИТЕРАТУРЫ

1. Скороход А. В. Элементы теории вероятностей и случайных процессов. Киев: Высшая школа, 1980. 344 с.
2. Феллер В. Введение в теорию вероятностей и ее приложения: в 2 т. Т. 2. М.: Мир, 1984. 738 с.
3. Cox D. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables // Proceedings of Cambridge Philosophical Society. 1955. V. 50. P. 433–441.
4. Kalashnikov V. V. Mathematical methods in queueing theory. Dordrecht: Kluwer academic publishers, 1994. 377 p.

МАРКИРОВАННЫЕ МАРКОВСКИЕ ПРОЦЕССЫ И ИХ ПРИМЕНЕНИЯ

В. В. Рыков^{1,2}, Н. М. Иванова³

¹ Российский государственный университет нефти и газа
НИУ им. И. М. Губкина, г. Москва, Россия

² Российский университет дружбы народов (РУДН), г. Москва, Россия

³ Институт проблем управления им. В. А. Трапезникова РАН, г. Москва,
Россия

Многие случайные явления природы и поведение стохастических систем могут быть описаны с помощью процессов с дискретным вмешательством случая (ПДВС). Марковские, полумарковские процессы с конечным или счётным множеством состояний, регенерирующие, полурегенерирующие, обобщённые полурегенерирующие процессы можно отнести к такому классу процессов.

В работах [1-5] в развитие идеи ПДВС была предложена модель маркированных марковских процессов (ММП), достаточно адекватно описывающих поведение сложных стохастических систем. Под ММП понимается последовательность с дискретным параметром

$$Z(t) = \{(J(t), \mathbf{V}(t)), t = 0.1, \dots\}$$

по моментам $S(t)$ вмешательства случая, где:

- последовательность $J(t)$ описывает состояния системы с множеством $\mathcal{J} \leq \infty$;
- последовательность $\mathbf{V}(t)$ представляет собой набор случайных меток $\mathbf{V}(t) = \{\mathbf{V}_j(t) : j \in \mathcal{J}\}$, дополняющих процесс Z до марковского. Метки принимают значения в измеримых пространствах (E_j, \mathcal{E}_j) .

Такой процесс задаётся:

- переходными вероятностями $p_{ij}(\mathbf{V}_i) = \mathbf{P}\{J(t+1) = j | J(t) = i, \mathbf{V}_i(t)\}$;
- операторами преобразования меток $\Phi_{ij}(\mathbf{V}_i)$;
- набором последовательностей независимых одинаково распределенных случайных величин ξ_t, η_t ($t = 1, 2, \dots$), задающих исходную информацию о модели, и их распределениями.

Данная последовательность определяет вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$, на котором определены все остальные компоненты модели.

Предложенный аппарат показал свою эффективность при исследовании нескольких достаточно сложных стохастических систем, в частности при исследовании системы надёжности $\langle GI_{k \leq n}|GI|l \rangle$ [3, 4] и системы массового обслуживания $(GI|GI|m, n)$ [5]. В работах были доказаны теоремы о преобразовании меток для указанных систем, разработаны процедуры расчета вероятностно-временных характеристик производительности и надежности в терминах меток, предложены алгоритмы численного расчета этих характеристик, основанные на имитационном моделировании преобразования меток. Кроме того, предлагаемый подход позволил исследовать чувствительность характеристик систем к исходным распределениям и их параметрам.

В докладе предлагается более подробный обзор этих работ и полученных результатов, а также задачи для дальнейших исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Рыков В. В., Иванова Н. М. О надёжности восстанавливаемой дублированной системы с произвольными распределениями времени безотказной работы и восстановления её элементов // Материалы XXII-й Международной конференции имени А. Ф. Терпугова, Томск: Том. гос. ун-т, 2023. С. 335 –340.
2. Рыков В. В., Иванова Н. М. О надёжности восстанавливаемой дублированной системы с произвольными распределениями наработка до отказа и времени ремонта ее элементов // Автоматика и телемеханика. 2024. № 9. С. 101 –123. doi: 10.31857/S0005231024090068
3. Рыков В. В., Иванова Н. М. О функции надежности системы $\langle GI_{k \leq n}|GI|l \rangle$. Ч. I // Аналитические результаты. Надежность. 2024. Т. 24, № 3. С. 34 –43. doi: 10.21683/1729-2646-2024-24-3-34-43
4. Рыков В. В., Иванова Н. М. О функции надежности системы $\langle GI_{k \leq n}|GI|l \rangle$. Ч. II // Численное исследование и анализ чувствительности. Надежность. 2024. № 24(4). С. 3 –11. <https://doi.org/10.21683/1729-2646-2024-24-4-3-11>
5. Rykov V., Morozov E., Ivanova N. Performance and Sensitivity Analysis of $(GI|GI|n, m)$ queues using Marked Markov Process // RT&A. 2025 (in print).

ИНВАРИАНТНОСТЬ СТАЦИОНАРНЫХ РАСПРЕДЕЛЕНИЙ ЭКСПОНЕНЦИАЛЬНЫХ СЕТЕЙ С ЗАПРЕТАМИ

Г. Ш. Цициашвили¹, М. А. Осипова¹

¹ Институт прикладной математики ДВО РАН, г. Владивосток, Россия

В работе рассматриваются сети массового обслуживания с запретами на переходы между узлами сети, определяющими протокол их работы [1]. В графе переходных интенсивностей сети выделяется множество базовых вершин (пропорциональное числу ребер) и ставится вопрос об удалении некоторого его подмножества так, чтобы стационарное распределение марковского процесса, описывающего функционирование сети, сохранялось. Чтобы это условие выполнялось, достаточно, чтобы множество вершин графа переходных интенсивностей после удаления подмножества базовых вершин совпадало с множеством состояний марковского процесса и этот граф был связным. Доказано, что отношения числа оставшихся базовых вершин к общему их числу n сходится к $1/2$ при $n \rightarrow \infty$.

СПИСОК ЛИТЕРАТУРЫ

1. Цициашвили Г. Ш. Инвариантные свойства систем массового обслуживания с несколькими потоками // Дальневосточный математический журнал. 2018. Т. 186 № 2. С. 267–270.

Секционные доклады

АНАЛИЗ МОДЕЛИ С МНОГОСЕРВЕРНЫМИ ЗАДАНИЯМИ В ПЕРЕХОДНОМ РЕЖИМЕ

С. Н. Астафьев¹

¹ Институт прикладных математических исследований Карельского научного центра РАН, г. Петрозаводск, Россия

Рассмотрим систему массового обслуживания, состоящую из s одинаковых серверов и общей FCFS очереди [1–5]. Заявки приходят через экспоненциально распределённое время с интенсивностью λ , каждая заявка имеет 2 параметра – случайное число серверов, необходимых для обслуживания, и случайный объём работы, зависящий от необходимого заявке числа серверов. Скорость обслуживания заявок зависит от текущего режима работы системы i , переключение между режимами происходит в моменты прихода и ухода заявок случайным образом согласно марковским цепям с матрицами $P_{n \times n}^{(a)}$ и $P_{n \times n}^{(d)}$ соответственно.

Такая система может быть рассмотрена как обобщённый процесс рождения и гибели, который является Марковской цепью в непрерывном времени с блочной трёхдиагональной матрицей \mathbf{Q} .

Переходные вероятности марковской цепи выражаются как решение системы уравнений Колмогорова–Чепмэна:

$$\begin{aligned}\frac{d\pi(t)}{dt} &= \pi(t)\mathbf{Q}, \\ \pi(0) &= \pi^{(0)},\end{aligned}\tag{1}$$

где $\pi(t)$ – вектор, описывающий вероятностное распределение состояний системы в момент времени $t \geq 0$ при начальном состоянии $\pi^{(0)}$.

Существует несколько методов численного решения уравнения (1). Среди них использование матричной экспоненты, униформизация, метод Дженсена, общие численные методы решения дифференциальных уравнений и их систем.

Более применимым в случае уравнения (1) методом решения является разложение уравнения (1) в ряд Тейлора. Для уравнения (1) метод Тейлора n -го порядка принимает вид

$$\pi^{(i)} = \pi^{(i-1)} \left(I + \frac{h}{1!} \mathbf{Q} + \frac{h^2}{2!} \mathbf{Q}^2 + \cdots + \frac{h^n}{n!} \mathbf{Q}^n \right) = \pi^{(i-1)} \hat{\mathbf{E}}_{(n)},\tag{2}$$

где матрица $\hat{\mathbf{E}}_{(n)}$ – конечная сумма первых n членов ряда Тейлора. Для этого метода можно доказать следующие 3 теоремы:

Теорема 1. Если $h \leq -\alpha^{-1}$, где $\alpha = \min(\mathbf{Q})$, то матрица $\hat{\mathbf{E}}_{(n)}$ является стохастической.

Теорема 2. Пусть $h = \frac{\delta}{-\alpha}$, где $\delta \in (0, 1]$, $\alpha = \min(\mathbf{Q})$. Тогда накопленную к моменту t ошибку аппроксимации для любого i -го компонента вектора $\boldsymbol{\pi}(t)$ можно оценить сверху как

$$|r_{n,i}(t)| \leq \min \left(1, \left\lceil \frac{t}{h} \right\rceil \frac{\delta^{n+1} e_k}{(n+1)!} \right), \quad (3)$$

где $e_k \leq e_{k+1}$ – любое из

$$\begin{aligned} e_1 &= \max \operatorname{abs}(\mathbf{P}^{n+1}), \quad e_2 = |\max(\mathbf{P}^n) - \min(\mathbf{P}^n)|, \\ e_3 &= \max(\mathbf{I} + \mathbf{C})^n, \quad e_4 = (1 + 3 \max \mathbf{C}) 2^{n-2}, \quad e_5 = 2^n, \end{aligned} \quad (4)$$

где

$$\mathbf{P} = -\alpha^{-1} \mathbf{Q}, \quad \mathbf{C} = \mathbf{I} + \mathbf{P}. \quad (5)$$

Теорема 3. Если у цепи Маркова с матрицей переходов \mathbf{Q} существует стационарное распределение $\boldsymbol{\pi}$ и $h \leq -\alpha^{-1}$, где $\alpha = \min(\mathbf{Q})$, то

$$\lim_{t \rightarrow +\infty} \boldsymbol{\pi}(t) = \lim_{i \rightarrow +\infty} \boldsymbol{\pi}^{(i)} = \boldsymbol{\pi}. \quad (6)$$

Объединение всех трёх предыдущих теорем позволяет добиться от метода разложения в ряд Тейлора следующих желаемых, применительно к Марковским цепям, свойств:

- отсутствие отрицательных вероятностей и вероятностей больше 1 в численном решении;
- сумма всех вероятностей, равна 1 на любом шаге численного метода;
- метод сходится к стационарному распределению при $t \rightarrow +\infty$;
- простая оценка для ошибки численного метода.

СПИСОК ЛИТЕРАТУРЫ

1. *Bini D., Latouche G., Meini B.* Numerical methods for structured Markov chains. New York: Oxford University Press, 2005. 327 c.
2. *Latouche G., Ramaswami V.* Introduction to Matrix Analytic Methods in Stochastic Modeling. Philadelphia: Society for Industrial and Applied Mathematics, 1999. 331 c.
3. *Butcher J. C.* Numerical Methods for Ordinary Differential Equations. Chichester: Wiley, 2016. 540 c.
4. *Grassmann W.* Transient solutions in markovian queueing systems // Computers \& Operations Research. 1977. T. 4, № 1. C. 47–53.
5. *Rumyantsev A. S. [et al.]* Three-level modeling of a speed-scaling supercomputer // Annals of Operations Research. 2023. T. 331. C. 649–677.

АСИМПТОТИЧЕСКИЙ АНАЛИЗ ДВУКЛАССОВОЙ RQ-СИСТЕМЫ $M_2|M_2|1$ С ПОЛИТИКОЙ FIFO В УСЛОВИЯХ БОЛЬШОЙ ЗАГРУЗКИ

Е. С. Булгакова¹, Е. А. Фёдорова¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

RQ-системы (Retrial Queue) – системы массового обслуживания с повторными вызовами. Характерной чертой таких систем является наличие повторных обращений заявок к обслуживающему прибору спустя некоторое случайное время после неудачной попытки обслуживания [1].

В работе исследуется RQ-система $M_2|M_2|1$ с двумя источниками повторных вызовов. Это RQ-система с одним обслуживающим прибором, на вход которой поступают простейшие потоки заявок с параметрами λ_1 и λ_2 . Время обслуживания заявок распределено по экспоненциальному закону с параметрами μ_1 и μ_2 соответственно. Если прибор свободен, то поступающая заявка его занимает, если прибор занят, то заявка отправляется на соответствующую орбиту, где они осуществляют случайную задержку, продолжительность которой имеет экспоненциальное распределение с параметрами σ_1 и σ_2 соответственно. Доступ заявок с каждой орбиты описывается политикой FIFO. Ставится задача нахождения стационарного распределения вероятностей числа заявок на каждой орбите.

Для описанной математической модели была составлена система дифференциальных уравнений Колмогорова и были получены уравнения баланса в стационарном режиме. Методом характеристических функций были получены условия стационарного функционирования системы. Выполнен асимптотический анализ системы в условии высокой интенсивности потока одного из классов заявок, т.е. при приближении интенсивности потока первого класса заявок λ_1 к пороговому значению S_1 , соответствующему границе устойчивости. В результате получено выражение для маргинального асимптотического распределения вероятностей числа заявок второго класса на орбите.

СПИСОК ЛИТЕРАТУРЫ

1. Artalejo J. R., Gomez-Corral A. Retrial Queueing Systems. A Computational Approach. Springer-Verlag Berlin Heidelberg, 2008. 267 p.

ОБ АНАЛИЗЕ НЕКОТОРЫХ ХАРАКТЕРИСТИК СИСТЕМЫ С РАЗДЕЛЕНИЕМ И ПАРАЛЛЕЛЬНЫМ ОБСЛУЖИВАНИЕМ

А. В. Горбунова¹

¹ Институт проблем управления им. В. А. Трапезникова Российской академии наук, г. Москва, Россия

Рассматривается система с разделением и параллельным обслуживанием заявок с пуссоновским входящим потоком с параметром $\lambda > 0$ и показательным распределением времени обслуживания на приборах с параметром $\mu > 0$, $\lambda < \mu$. При поступлении в данную систему заявка разделяется на K подзаявок, после чего каждая подзаявка поступает в одну из K очередей на обслуживание (рис. 1). Заявка считается обслуженной только после обслуживания всех изначально составляющих ее подзаявок. С помощью описанной системы массового обслуживания моделируются различные типы физических структур, в которых происходит разделение исходной задачи на K частей для сокращения времени решения. С некоторыми результатами исследования системы можно ознакомиться, например, в [2-4].

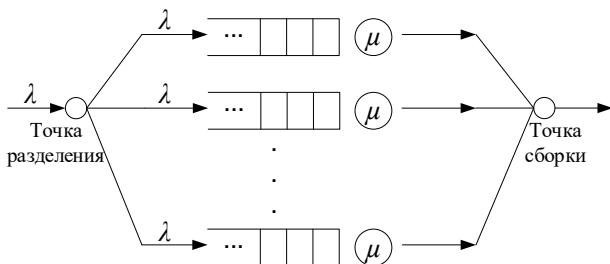


Рис. 1. Схема системы с разделением и параллельным обслуживанием заявок

В работе предлагается аналитическая оценка для квантилей распределения одной из важнейших характеристик производительности подобных систем — времени отклика системы, т.е. времени пребывания заявки в системе. Время отклика системы R_K определяется максимальным временем обслуживания одной из изначально составляющих заявку подзаявок, т.е.

$$R_K = \max\{\xi_1, \dots, \xi_K\}, \quad (1)$$

где ξ_i — время пребывания i -й подзаявки в i -й подсистеме, $i = 1, \dots, K$.

Вывод выражения для оценки квантилей времени отклика основывается на элементах теории копул и их диагональных сечений [5, 6], визуальном анализе данных (как одной из составляющих data mining), применения методов оптимизации, а также на имитационном моделировании системы с разделением и параллельным обслуживанием. Подход был успешно применен для анализа частного случая системы для $K = 2$ [1].

В результате применения подхода получается следующая оценка для квантилей распределения:

$$\hat{R}_{K,p} = -\frac{\ln(1 - p^{\frac{1}{\kappa^{1-(C_1-C_2p^2)\rho}}})}{\mu - \lambda}, \quad (2)$$

где

$$C_1 \approx 0.390797, \quad C_2 \approx 0.221811.$$

Формула (2) демонстрирует хорошее качество приближения на изначально выбранном диапазоне значений параметров, а именно число подсистем (подзаявок) K меняется от 2 до 20, коэффициент загрузки системы $\rho = \lambda/\mu = \{0.10, 0.15, \dots, 0.45\}$, а уровень квантилей, т.е. вероятность $p = \{0.20, 0.25, \dots, 0.90\}$. В частности, модуль максимального значения погрешности приближения не превышает 5.5%, а ее среднее значение составляет примерно 1.1%.

СПИСОК ЛИТЕРАТУРЫ

1. Горбунова А. В., Лебедев А. В. О новом подходе к оценке квантилей времени отклика системы с разделением и параллельным обслуживанием заявок // Управление большими системами: сб. тр.. 2024. № 108. С. 6–21.
2. Nelson R., Tantawi A. N. Approximate analysis of fork/join synchronization in parallel queues // IEEE Transactions on Computers. 1988. Vol. 37, No. 6. P. 739–743.
3. Thomassen A. Analysis of fork/join and related queueing systems // ACM Computing Surveys (CSUR). 2014. Vol. 47, No 2. P. 17:1–17:71.
4. Samyukta Sethuraman Analysis of Fork-Join Systems: Network of Queues with Precedence Constraints. Boca Raton: CRC Press, 2022. 104 p.
5. Nelsen R. An introduction to copulas. Berlin, Germany: Springer, 2006. 269 p.
6. Лебедев А. В. Верхняя граница среднего минимума зависимых случайных величин с известным коэффициентом Кендалла // Теория вероятностей и ее применения. 2019. Т. 64, вып. 3. С. 578–589.

ПОСТРОЕНИЕ И ИССЛЕДОВАНИЕ МОДЕЛИ СИСТЕМЫ УПРАВЛЕНИЯ КОНФЛИКТНЫМИ ПОТОКАМИ С ДВУМЯ ЦИКЛИЧЕСКИМИ РЕЖИМАМИ

В. Е. Гордеев¹, А. В. Зорин¹

¹Национальный исследовательский Нижегородский государственный университет им. Н. И. Лобачевского, г. Нижний Новгород, Россия

В данной работе рассматривается система массового обслуживания с двумя входными потоками Π_1, Π_2 и одним обслуживающим устройством. Поток Π_j – пуассоновский неординарный поток с интенсивностью $\lambda_j > 0$ для поступления групп. Группа из одной машины поступает с вероятностью $1 - \alpha_j$, группа из $n \geq 2$ машин – с вероятностью $\alpha_j(1 - \beta_j)\beta_j^{n-2}$, где $0 < \alpha_j \leq 1$, $0 \leq \beta_j < 1$ [1]. Требования поступают в неограниченные накопители O_1, O_2 . Обслуживающее устройство в каждый момент времени обслуживает не более одного потока. Поток Π_j обслуживается в течение фиксированного времени, после чего включается режим переналадки и управления также в течение фиксированного времени, затем начинается обслуживание другого потока. Обслуживающее устройство имеет 2 режима, режим определяет конкретные длительности обслуживания и переналадок, смена режима происходит только в конце переналадки после обслуживания второго потока. Правило выбора режима определяется некоторой функцией $D(\cdot, \cdot): (\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\}) \rightarrow \{1, 2\}$, где 1 и 2 – номера режимов. Закон обслуживания очередей определяется величинами $\ell_{j,m}$, характеризующими максимально возможное число обслуженных требований из очереди O_j в режиме m . Обслуженные заявки покидают систему.

Для построения математической модели используется представление системы обслуживания в виде кибернетической управляющей системы [2, 3]. Пусть τ_i – момент i -го переключения обслуживающего устройства ($i = 1, 2, \dots$), причем $\tau_0 = 0$. Введем множество состояний обслуживающего устройства $\Gamma = \{\Gamma^{(1,1)}, \Gamma^{(2,1)}, \Gamma^{(3,1)}, \Gamma^{(4,1)}, \Gamma^{(1,2)}, \Gamma^{(2,2)}, \Gamma^{(3,2)}, \Gamma^{(4,2)}\}$. Состояние $\Gamma^{(r,m)}$ означает, что устройство находится в режиме m и обслуживается очередь j , если $r = 2j - 1$; если $r = 2j$, то устройство находится в состоянии переналадки после обслуживания требований из очереди O_j . За $T_{r,m}$ обозначим длительность пребывания обслуживающего устройства в состоянии $\Gamma^{(r,m)}$, причем $T_{r,m}$ – неслучайная положительная величина. Пусть $\eta_{j,i}$ – случайное число требо-

ваний, поступивших по потоку Π_j за временной полуинтервал $(\tau_i, \tau_{i+1}]$, а $\xi_{j,i}$ – число требований потоков насыщения Π_j^* за то же время. Введем вектора $\eta_i = (\eta_{1,i}, \eta_{2,i})$ и $\xi_i = (\xi_{1,i}, \xi_{2,i})$. Входной полюс системы обслуживания будем описывать последовательностью пар $\{(\eta_i, \xi_i); i = 1, 2, \dots\}$ в терминах их условных распределений. Пусть $\varkappa_{j,i}$ – число заявок потока Π_j находящихся в очереди O_j в момент времени τ_i . Введем вектор $\varkappa_i = (\varkappa_{1,i}, \varkappa_{2,i})$. Математическим описанием внешней памяти будем считать последовательность векторов $\{\varkappa_i; i = 0, 1, \dots\}$. Пусть Γ_i есть состояние обслуживающего устройства на промежутке $(\tau_{i-1}, \tau_i]$, Γ_0 – в момент времени τ_0 . Последовательность $\{\Gamma_i; i = 0, 1, \dots\}$ задается рекуррентным образом в зависимости от состояний некоторых других блоков системы.

Результатом работы являются следующие теоремы:

Теорема 1. При заданном начальном распределении вектора $(\Gamma_0, \varkappa_{1,0}, \varkappa_{2,0})$ последовательность $\{(\Gamma_i, \varkappa_{1,i}, \varkappa_{2,i}); i = 0, 1, \dots\}$ является марковской.

Теорема 2. При заданной функции переключения $D(x_1, x_2)$, принимающей значение 1 при $x_2 \leq ax_1 + b$ и значение 2 при $x_2 > ax_1 + b$, где $a, b \in \mathbb{R}$, $a \geq 0$, $b \geq 0$, состояния, принадлежащие множеству $C_u = \{(\Gamma^{(1,2)}, x_1, x_2): x_1 \geq 0, 0 \leq x_2 \leq b\}$, являются несущественными. Состояния, не принадлежащие множеству C_u , образуют один класс существенных с периодом четыре.

Получены рекуррентные соотношения для производящих функций $\psi_i(z_1, z_2, r, m) = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} z_1^{x_{1,i}} z_2^{x_{2,i}} \mathbb{P}(\Gamma_i = \Gamma^{(r,m)}, \varkappa_{1,i} = x_{1,i}, \varkappa_{2,i} = x_{2,i})$, где $z_1, z_2 \in \mathbb{C}$, $|z_1| < 1$, $|z_2| < 1$. Например, $\psi_i(z_1, z_2, 2, m) = q_1(z_1, T_{2,m}) \times q_2(z_2, T_{2,m}) \psi_{i-1}(z_1, z_2, 1, m)$, где $q_j(z, T) = \sum_{x \geq 0} z^x \varphi_j(x, T)$, а $\varphi_j(x, T)$ есть вероятность поступления x требований по потоку Π_j за время T .

СПИСОК ЛИТЕРАТУРЫ

1. Федоткин М. А. Модели в теории вероятностей. М.: ФИЗМАТЛИТ, 2012. 608 с.
2. Зорин А. В. Кибернетический подход к построению и анализу математической модели тандема двух перекрестков // Проблемы теоретической кибернетики: Материалы XVI Междунар. конф. Н. Новгород: Изд-во Нижегород. ун-та, 2011, С. 179–183.
3. Зорин А. В. Кибернетическая модель циклического управления конфликтными потоками с последействием // Ученые записки Казанского университета. 2014. Т. 156, № 3. С. 66–75

МОДЕЛИРОВАНИЕ ЭНЕРГОПОТРЕБЛЕНИЯ В ЗАКРЫТОЙ ДВУХФАЗНОЙ СИСТЕМЕ МАССОВОГО ОБСЛУЖИВАНИЯ С КОНЕЧНЫМ ЧИСЛОМ ПОЛЬЗОВАТЕЛЕЙ

Д. Д. Жамсал¹, И. Л. Лапатин¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

Рассматривается закрытая двухфазная система массового обслуживания с конечным числом пользователей N [1-3]. Каждому пользователю соответствует одна заявка, и общее число заявок в системе постоянно.

Каждая заявка может находиться в одном из двух состояний:

- *активная фаза* — заявка потребляет вычислительные ресурсы и энергию;
- *пассивная фаза* — заявка «отдыхает» и не потребляет ресурсы.

Переходы между фазами происходят независимо для каждой заявки. Время пребывания в активной и пассивной фазах является случайным и подчиняется экспоненциальному закону распределения. Интенсивности переходов из активной фазы в пассивную и обратно обозначаются как $\mu_1(k)$ и μ_2 соответственно, где k — текущее число активных заявок.

В модели предполагается, что только заявки, находящиеся в активной фазе, потребляют энергию. При этом энергопотребление одной активной заявки представляется случайной величиной. Возможны три подхода к заданию энергозатрат:

- *детерминированный случай* — энергия постоянна и известна заранее;
- *дискретный случай* — энергия описывается случайной дискретной величиной с заданными вероятностями;
- *непрерывный случай* — энергия задаётся непрерывной неотрицательной случайной величиной с известной функцией распределения.

Суммарное энергопотребление в системе в каждый момент времени определяется как сумма энергий всех активных заявок. Таким образом, общая нагрузка может существенно варьироваться в зависимости от числа активных пользователей и распределения энергозатрат.

Для более точного моделирования поведения системы при высокой нагрузке вводится функция деградации, которая учитывает влияние количества активных заявок на производительность. Эта функция корректирует интенсивность перехода из активной фазы в пассивную в зависимости от текущей нагрузки.

Интенсивность выхода из активной фазы определяется выражением

$$\mu_1(k) = \mu_1 \cdot f(k),$$

где μ_1 — базовая интенсивность, а $f(k)$ — функция деградации, зависящая от числа активных заявок k .

Функция деградации отражает снижение эффективности системы по мере роста нагрузки. При большом числе одновременно активных заявок производительность падает, что позволяет реалистично учитывать ограниченность ресурсов и ухудшение отклика системы в условиях перегрузки.

На основе построенной математической модели удалось рассчитать стационарные вероятности числа активных заявок в системе. Проведён анализ энергопотребления, учитывающий различные варианты распределения энергозатрат одной заявки — как в дискретной, так и в непрерывной постановке. Получены плотности распределения суммарной энергии в зависимости от числа активных заявок.

Результаты позволяют глубже понять поведение вычислительного узла при различной степени загрузки и дают количественные характеристики энергопотребления. На основе проведённого анализа в дальнейшем возможно построение более точной и адаптивной модели вычислительного узла, учитывающей деградацию производительности и вероятностную природу энергозатрат.

Предложенный подход демонстрирует эффективность применения марковских моделей для описания и оптимизации ресурсов в вычислительных и облачных системах.

СПИСОК ЛИТЕРАТУРЫ

1. Солнышкина И. В. Теория систем массового обслуживания. Комсомольск-на-Амуре, 2015.
2. Теория массового обслуживания. URL: <https://ru.wikipedia.org> (дата обращения: 10.10.2024).
3. Клейнрок Л. Теория массового обслуживания. М.: Машиностроение, 1979.

О ПРИМЕНЕНИИ СЕТЕЙ МАССОВОГО ОБСЛУЖИВАНИЯ ДЛЯ МОДЕЛИРОВАНИЯ УНИВЕРСАЛЬНОГО МОРСКОГО ПОРТА

М. Л. Жарков¹, А. Л. Казаков¹, А. А. Лемперт¹

¹ Институт динамики систем и теории управления имени
В.М. Матросова СО РАН, г. Иркутск, Россия

Теория массового обслуживания еще с 70-х гг. прошлого века считается одним из эффективных математических инструментов для исследования и прогнозирования работы причалов в морских портах [1]. Большая часть современных исследований в данном направлении посвящена контейнерным терминалам [2, 3]. Однако большинство портов относится к классу универсальных, которые обрабатывают разные типы груза и поэтому имеют более сложную технологию работы. Таким образом, изучение этого класса портов является актуальной задачей.

Все порты, как правило, включают четыре подсистемы: якорную стоянку для судов, причальную зону, складские площадки и наземный грузовой фронт. В подсистемах регулярно выполняются однотипные операции с грузами, а на их продолжительность оказывает влияние большое число случайных факторов, в частности погода. Универсальные порты имеют две отличительные особенности. Во-первых, существенно меняется технология разгрузки и погрузки судна в зависимости от типа груза. Во-вторых, размеры партий зависят от типа судна и его вместимости.

Рассмотрим математическую модель функционирования универсального морского порта на основе сетей массового обслуживания (СеМО). Грузы, прибывающие на морском и наземном транспорте, описываются отдельными неординарными потоками заявок. Для моделирования работы различных элементов порта используются узлы СеМО нескольких типов:

- работа складов, бункеров и ворот с автостоянкой $-*/M^X/n/m$;
- работа якорной стоянки, кранов на складах и перевозка грузов между подсистемами на грузовиках $-*/M^X/n/0$;
- разгрузка судов на причале $-*/M/n/m$;
- погрузка судов $-*/M^Y/1/Y$, где Y – случайная величина, которая описывает размер отправляемой партии грузов на судне.

Исследование выполнено за счет гранта Российского научного фонда (проект № 24-21-00264).

Дисциплина обслуживания очереди во всех узлах – FIFO.

Структурная и параметрическая идентификация модели выполнена для универсального порта Куангнинь, который является одним из крупнейших систем данного типа во Вьетнаме. Он расположен в одноименной провинции – ожидаемой конечной точкой планируемого транспортного коридора Россия – Вьетнам.

На рис. 1 представлена схема СeМО, включающая 19 узлов и три входящих неординарных потока заявок, которые описывают поступление грузов на грузовиках, речных баржах и морских судах. Узлы моделируют работу следующих элементов: 1 и 13 – якорных стоянок; 2, 9 – 12, 14 и 15 – причалов, береговых кранов и конвейерной системы на них; 4, 5 и 19 – автомобильной стоянки и пропускных пунктов; 6, 8, 16 и 18 – движение грузовиков; 7 и 17 – открытого и закрытого складов.

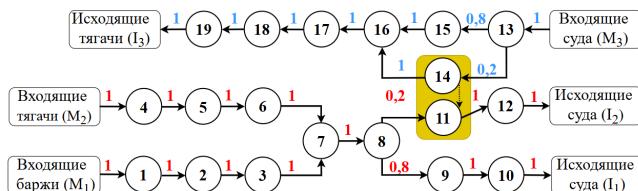


Рис. 1. Схема СeМО

На основе результатов численного исследования СeМО установлено, что порт имеет запас пропускной способности порядка 40% и не нуждается в модернизации. Тем не менее в случае открытия транспортного коридора производительности конвейерной системы осуществляющей загрузку экспортных товаров в суда, будет недостаточно и потребуется увеличение ее скорости работы.

СПИСОК ЛИТЕРАТУРЫ

1. Edmond E. D., Maggs R. P. How useful are queue models in port investment decisions for container berths? // Journal of the Operational Research Society. 1978. Vol. 29. P. 741–750.
2. Legato P., Mazza R. M. Queueing networks for supporting container storage and retrieval // Maritime Business Review. 2023. Vol. 8. P. 301–317.
3. Лемперт А. А., Жарков М. Л., Казаков А. Л., By X. З. Моделирование морского контейнерного терминала с использованием сети массового обслуживания // Управление большими системами. 2024. № 112. С. 741–750.

АСИМПТОТИЧЕСКИЙ АНАЛИЗ МНОГОЛИНЕЙНОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С БЛОКОМ ВОССТАНОВЛЕНИЯ

Я. Е. Измайлова¹, А. Т. Исаев¹

¹ Национальный исследовательский Томский государственный
университет, г. Томск, Россия

Теория массового обслуживания (ТМО) является важным инструментом для анализа систем, в которых происходит обслуживание большого количества заявок. ТМО находит применение в телекоммуникациях, логистике, экономике и других сферах, где необходимо оптимизировать процессы обслуживания. Особый интерес представляют многолинейные системы с двумя потоками разнотипных заявок, поскольку они позволяют учитывать не только стандартные потоки заявок, но и потоки отрицательных заявок, способных уничтожать или изменять поступающие запросы.

Целью исследования является анализ многолинейной системы массового обслуживания, в которую поступают два потока заявок: положительный и отрицательный. Такие системы были впервые предложены Е. Геленбе [1] и первоначально применялись для моделирования биофизических нейронных сетей, но в настоящее время используются в телекоммуникациях и социально-экономических задачах.

Рассматривается многолинейная система массового обслуживания, в которой поступившая положительная заявка переходит в блок обслуживания и встает на прибор. Время обслуживания имеет экспоненциальное распределение. Если в момент обслуживания положительной заявки в систему приходит отрицательная заявка, то она искажает обслуживание положительной, положительная заявка переходит в блок восстановления, а отрицательная заявка уничтожается. Данная модель отличается от классических моделей массового обслуживания где заявка либо успешно завершает обслуживание, либо теряется наличием блока восстановления. Он позволяет учитывать возможность повторной обработки заявок, что приближает модель к реальным системам с механизмом отката или коррекции ошибок. В блоке восстановления положительная заявка находится случайное время, распределенное по экспоненциальному закону, после чего переходит вновь на обслуживание. Если в момент обслуживания положительной заявки в систему не

приходит отрицательная заявка, то после окончания обслуживания положительная заявка покидает систему. Если в блоке обслуживания нет заявок, то поступившая отрицательная заявка, никак не влияя на состояние системы, уничтожается.

Исследование проводилось с помощью метода асимптотического анализа в условиях длительного восстановления заявок, что позволяет оценить поведение системы при различных параметрах. Была получена производящая функция и построены графики распределения вероятностей числа заявок в блоке обслуживания, демонстрирующие динамику системы при различных параметрах входных потоков. Результаты исследования могут быть полезны для проектирования и оптимизации подобных систем в области телекоммуникаций, компьютерных сетей и других технологических процессов.

СПИСОК ЛИТЕРАТУРЫ

1. *Fourneau J. N., Gelenbe E., Suros R.* G-networks with multiple classes of negative and positive customers // Theoret. Comput. Sci. 1996. Vol. 155. P. 141 –156.

ИССЛЕДОВАНИЕ МНОГОПОТОКОВЫХ RQ-СИСТЕМ

Н. С. Кострюков¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

Системы с повторными попытками (Retrial Queue) представляют собой стохастические модели, активно применяемые для анализа и оптимизации процессов в телекоммуникациях, облачных платформах, call-центрах и других распределённых системах [1-4]. Их отличительной чертой является механизм повторного обращения: если заявка не обслуживается из-за занятости ресурса, она не покидает систему, а переходит в орбиту — буфер ожидания, откуда через случайный интервал времени повторно запрашивает обслуживание. Это позволяет более реалистично моделировать работу систем с ограниченными ресурсами и высокой нагрузкой.

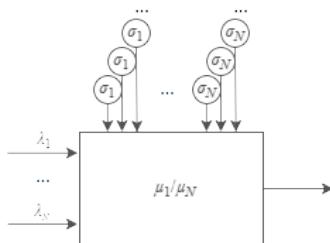
В рамках данной работы рассматривается многоклассовая RQ-система, обрабатывающая заявки различных типов, каждая из которых характеризуется собственными параметрами поступления, обслуживания и задержки в орбите. Такой подход позволяет учитывать гетерогенность трафика — от голосовых вызовов и видеопотоков до IoT-данных и задач облачных вычислений.

Исследуемая модель представляет собой RQ-систему $M_N|M_N|1$ с N простейшими входящими потоками с интенсивностями λ_n , $n = \overline{1, N}$ (рис. 1). Время обслуживания заявок распределено по экспоненциальному закону с параметрами μ_n . Если прибор занят, заявка направляется в орбиту, где ожидает случайное время (экспоненциальное распределение с параметром σ_n), после чего повторно пытается получить обслуживание. Модель реализует множественный случайный доступ: любая заявка с орбиты может обратиться за обслуживанием.

Обозначим через $i_n(t)$ количество заявок класса n в орбите в момент времени t , а $k(t)$ — состояние обслуживающего прибора:

$$k(t) = \begin{cases} 0, & \text{если прибор свободен в момент времени } t, \\ n, & \text{если прибор занят заявкой класса } n. \end{cases}$$

Таким образом, динамика системы описывается $(N+1)$ -мерным марковским процессом $k(t), \mathbf{i}(t)$, где $\mathbf{i}(t) = i_1(t), \dots, i_N(t)$ — вектор числа заявок каждого класса на орбите.

Рис. 1. RQ-система $M_N|M_N|1$

Целью исследования является получение вероятностных характеристик такой многоклассовой RQ-системы. Для этого применяется многомерный гауссовский асимптотический анализ и дискретно-событийный подход имитационного моделирования, что позволяет как теоретически, так и экспериментально оценить поведение системы в условиях высокой нагрузки и конкуренции между потоками.

СПИСОК ЛИТЕРАТУРЫ

1. Artalejo J. R., Gomez-Corral A. Retrial queueing systems // Mathematical and Computer Modelling. 1999. Vol. 30, No. 3–4. P. 13–15.
2. Avrachenkov K., Morozov E., Nekrasova R. Optimal and Equilibrium Retrial Rates in Single-Server Multi-orbit Retrial Systems // Lecture Notes in Computer Science. Springer, Cham, 2015. Vol. 9305. P. 135–146.
3. Krishnamoorthy A., Joshua V. C., Mathew A. P. A Retrial Queueing System with Multiple Hierarchical Orbits and Orbital Search // Communications in Computer and Information Science. Springer, Cham, 2018. Vol. 919. P. 224–233.
4. Рындун А. В. Методика оценки эффективности передачи многомодальной информации // Системный синтез и прикладная синергетика: XI Всерос. науч. конф. 2022. С. 139–144. <https://elibrary.ru/item.asp?id=50348988> – elibrary.ru (дата обращения: 17.11.2024).

МАРКОВСКАЯ СИСТЕМА С ПОВТОРНЫМИ ВЫЗОВАМИ И ПРИОРИТЕТНЫМ ОБСЛУЖИВАНИЕМ

Т. К. Латипов¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

В данной работе рассматривается RQ-система $M_2|M_2|1$ с приоритетными заявками [1]. Система состоит из прибора, к которому на вход поступают 2 простейших потока событий с интенсивностью λ_1 (интенсивность приоритетного потока) и λ_2 . Прибор обслуживает эти заявки с интенсивностью μ_1 и μ_2 соответственно. Если прибор свободен и поступает заявка первого или второго потока, она занимает прибор для обслуживания. Если же прибор не свободен, то возможно одно из следующих событий:

- Обслуживается заявка первого потока, и поступает заявка первого потока. Поступившая заявка теряется [2].
- Обслуживается заявка первого потока, и поступает заявка второго потока. Поступившая заявка уходит на орбиту где она осуществляет случайную задержку в течение экспоненциального времени σ .
- Обслуживается заявка второго потока, и поступает заявка первого потока. Обслуживаемая заявка вытесняется на орбиту.
- Обслуживается заявка второго потока, и поступает заявка второго потока. Поступившая заявка уходит на орбиту.

Пусть $k(t)$ – состояние прибора в момент времени t . Этот процесс может принимать следующие значения: 0 – прибор свободен, 1 – прибор занят обслуживанием заявки первого потока, 2 – прибор занят обслуживанием заявки второго потока. Процесс $i(t)$ – число заявок на орбите в момент времени t . Для распределения вероятностей

$$P_k(i, t) = P\{k(t) = k, i(t) = i\} \quad (1)$$

можно записать систему уравнений Колмогорова в стационарном режиме

$$\begin{aligned}
 & -(\lambda_1 + \lambda_2 + i\sigma)P_0(i) + \mu_1 P_1(i) + \mu_2 P_2(i) = 0, \\
 & -(\lambda_2 + \mu_1)P_1(i) + \lambda_1 P_0(i) + \lambda_2 P_1(i-1) + \lambda_1 P_2(i-1) = 0, \\
 & -(\lambda_1 + \lambda_2 + \mu_2)P_2(i) + \lambda_2 P_0(i) + \sigma(i+1)P_0(i+1) + \\
 & \quad + \lambda_2 P_2(i-1) = 0.
 \end{aligned} \tag{2}$$

Система (2) решена с помощью метода асимптотического анализа для определения распределений вероятностей состояния прибора и числа заявок на орбите.

СПИСОК ЛИТЕРАТУРЫ

1. Wang K. Queueing system with impatient customers: A review // Service Operations and Logistics, and Informatics: Proc. of 2010 IEEE International Conference. 2010. P. 82–87.
2. Шимановская М. В. Модель СМО с неоднородными заявками и абсолютным приоритетом обслуживания // Вестник Пермского университета. Серия: Математика. Механика. Информатика. 2013 № 4. (23). С. 103–107.

АСИМПТОТИЧЕСКИЙ АНАЛИЗ RQ-СИСТЕМЫ С ВЫЗЫВАЕМЫМИ ЗАЯВКАМИ ИЗ ОРБИТЫ

О. Д. Лизюра¹, А. Т. Искакова¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

Рассматривается call-центр, в котором оператор может сам перезванивать клиентам в случаях, когда клиент совершил звонок в момент занятости оператора. Такой подход позволяет сократить время простоя оператора и снизить эксплуатационные расходы. Кроме того, это позволяет уменьшить время ожидания клиента.

В статье [1] рассматривается математическая модель подобного call-центра. Авторы разработали численный метод расчета распределения вероятностей числа заявок на орбите, предполагая, что суммарная интенсивность вызывания заявок из орбиты зависит от текущего числа заявок на орбите. Также Фунг-Дук и его соавторы рассматривали RQ-системы с вызываемыми заявками, где вызывание происходило не из орбиты, а извне [2, 3].

Рассматривается система массового обслуживания с повторными обращениями (RQ-система). На вход поступает простейший поток заявок с интенсивностью λ . Заявка входящего потока, поступившая в систему и обнаружившая прибор свободным, занимает его и обслуживается в течение экспоненциально распределенного времени с параметром μ . Если же в момент поступления заявка застает прибор занятым, она мгновенно уходит на орбиту и повторяет попытку занять прибор по истечении случайного времени, распределенного экспоненциально с параметром σ . Также когда прибор свободен, он вызывает заявки из орбиты с интенсивностью β . Суммарная интенсивность вызывания постоянна и не зависит от числа заявок на орбите.

Обозначим процесс $k(t)$ – состояние прибора в момент времени t . Этот процесс может принимать следующие значения: 0, если прибор свободен; 1, если прибор обслуживает заявку из входящего потока или орбиты. Также введем случайный процесс $i(t)$ – число заявок на орбите в момент времени t . Для нахождения распределения вероятностей случайного процесса $\{k(t), i(t)\}$ применяется метод асимптотического анализа в предельном условии большой задержки заявок на орбите ($\sigma \rightarrow 0$, $\beta \rightarrow 0$). Для согласования порядка малости σ и β представим их в виде $\sigma = \gamma_1 \varepsilon^2$, $\beta = \gamma_2 \varepsilon^2$.

Для такой системы показано, что предельное распределение числа заявок на орбите является нормальным. Получены формулы для параметров распределения вероятностей

$$\varkappa_1 = \frac{\lambda^2}{\sigma(\mu - \lambda)}, \quad \varkappa_2 = \frac{\lambda^2 \mu}{\sigma(\mu - \lambda)^2}.$$

Стационарные вероятности состояний прибора находим по формуле

$$r_0 = 1 - \frac{\lambda}{\mu}, \quad r_1 = \frac{\lambda}{\mu}.$$

Предлагается аппроксимировать распределение вероятностей числа заявок на орбите распределением Гаусса следующего вида:

$$H(u) \approx \exp \left\{ ju\varkappa_1 + \frac{(ju)^2}{2} \varkappa_2 \right\}.$$

СПИСОК ЛИТЕРАТУРЫ

1. Dragieva V., Phung-Duc T. Two-way communication M/M/1 retrial queue with server-orbit interaction // Proceedings of the 11th International Conference on Queueing Theory and Network Applications. Wellington: ACM, 2016. P. 1–7.
2. Sakurai H., Phung-Duc T. Two-way communication retrial queues with multiple types of outgoing calls // Top. 2015. Vol. 23. P. 466–492.
3. Artalejo J. R., Phung-Duc T. Markovian retrial queues with two way communication // Journal of industrial and management optimization. 2012. Vol. 8, № 4. P. 781–806.

НАХОЖДЕНИЕ ПАРАМЕТРОВ МАР-ПОТОКА ПРИ НАЛИЧИИ ИНФОРМАЦИИ О ДЛИНАХ ИНТЕРВАЛОВ

И. Л. Лапатин¹, Н. А. Луценко¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

В работе решается задача нахождения параметров МАР-потока при наличии информации о значениях только некоторых числовых характеристик длин интервалов в потоке. Будем полагать, что известна средняя длина интервалов, коэффициент вариации ($CVar$) длин интервалов и коэффициент корреляции ($CCor$) соседних длин интервалов.

Задача сводится к тому, чтобы найти такие матрицы \mathbf{Q} , $\mathbf{\Lambda}$, \mathbf{D} [1], определяющие МАР-поток, чтобы вычисленные по ним теоретические числовые характеристики ($CVar_T$, $CCov_T$) длин интервалов [2] были близки к заданным.

В качестве меры близости вводится функция ошибки [3]:

$$\frac{1}{N} \sum_{i=0}^N (x_i - y_i)^2.$$

Чтобы данная функция подходила под задачу, её нужно отнормировать, поскольку коэффициент вариации может быть достаточно большим, а коэффициент корреляции изменяется от -1 до 1 . Поэтому финальный вариант функции ошибки

$$fit = \frac{\left(\frac{CVar_T - CVar}{CVar}\right)^2 + \left(\frac{CCov_T - CCov}{1 + |CCov|}\right)^2}{2}.$$

Вычисление теоретических числовых характеристик производится при помощи матриц D_0 , D_1 , предварительно преобразовав матрицы \mathbf{Q} , $\mathbf{\Lambda}$, \mathbf{D} .

Для нахождение параметров МАР-потока предлагается использовать в разных комбинациях три модификации алгоритма перебора параметров:

- 1) базовый генетический алгоритм с одновременным перебором всех параметров;
- 2) последовательный перебор значений элементов матриц \mathbf{Q} , $\mathbf{\Lambda}$, \mathbf{D} ;

- 3) перебор параметров в матрицах \mathbf{Q} , $\mathbf{\Lambda}$, \mathbf{D} в заданной окрестности определенных значений.

В результате получили, что генетический алгоритм работает достаточно долго и точность (значение функции ошибки) найденного решения достигает 5 знаков после запятой.

Алгоритм, который основывается на переборе матриц работает в несколько раз быстрее, и достигается точность в 6 знаков после запятой.

Алгоритм, который перебирает параметры матриц \mathbf{Q} , $\mathbf{\Lambda}$, \mathbf{D} по отдельности по скорости вычисляется примерно так же, как и предыдущий, но точность достигается до 9 знаков после запятой.

СПИСОК ЛИТЕРАТУРЫ

1. Гарайшина И. Р., Моисеева С. П., Назаров А. А. Методы исследования коррелированных потоков и специальных систем массового обслуживания. Томск: Издательство НТЛ, 2010. 204 с.
2. Вишневский В. М., Дудин А. Н., Клименок В. И. Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях. М.: Рекламно-издат. центр «ТЕХНОСФЕРА», 2018. 564 с.
3. URL: <https://www.geeksforgeeks.org/mean-squared-error/> – GeeksForGeeks. 2025

ИМИТАЦИОННАЯ МОДЕЛЬ ИСПОЛЬЗОВАНИЯ ОПЕРАТИВНОЙ ПАМЯТИ ВЫЧИСЛИТЕЛЬНОГО УЗЛА ПРИ ОГРАНИЧЕННОМ ЧИСЛЕ ПОЛЬЗОВАТЕЛЕЙ

К. Г. Малков¹, И. Л. Лапатин¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

Введение. Рассматривается имитационная модель двухфазной закрытой системы массового обслуживания с конечным числом пользователей [1, 2]. Заявки циклически переходят между активным и пассивным состояниями. Учитывается потребление оперативной памяти каждой заявкой, что позволяет исследовать временные характеристики и распределение памяти.

Модель построена на стохастическом подходе с функцией деградации и дискретным или непрерывным распределением потребления памяти. В основе модели лежат принципы анализа марковских процессов.

Заявки циклически переходят из одной фазы в другую, характеризуясь случайными моментами начала и завершения обработки. Каждая заявка потребляет случайное количество памяти, заданное дискретным или непрерывным распределением (например, равномерным или экспоненциальным).

Переходы между фазами:

- λ_0 — интенсивность перехода в пассивное состояние;
- λ_1 — интенсивность перевода в активное состояние.

Состояние системы в момент времени t можно описать парой $(n(t), m(t))$, где

- $n(t) \in [0, N]$ — число активных заявок в системе;
- $m(t) \in \mathbb{R}_+$ — суммарный объём используемой оперативной памяти.

Функция деградации. Для учёта снижения производительности при увеличении нагрузки введена функция деградации $D(n)$, зависящая от числа активных заявок n . Эта функция модулирует интенсивность перехода заявки из активного состояния в пассивное следующим образом:

$$\lambda(n) = \lambda_0 \cdot D(n),$$

где λ_0 — базовая интенсивность обслуживания, $D(n) \leq 1$ — монотонно убывающая функция.

Это позволяет моделировать реалистичное поведение системы, при котором рост числа активных заявок приводит к нелинейному снижению скорости обработки.

Цель работы – построить имитационную модель использования оперативной памяти вычислительного узла в условиях ограниченного числа пользователей с помощью закрытой двухфазной системы массового обслуживания.

Задачи:

- Разработать модель закрытой двухфазной системы массового обслуживания, учитывающую:
 - переход заявок между двумя фазами обслуживания;
 - случайное потребление оперативной памяти заявками.
- Реализовать функцию деградации производительности системы при увеличении числа активных заявок.
- Рассчитать вероятностные характеристики системы:
 - стационарные вероятности числа активных заявок;
 - распределение общего объёма используемой оперативной памяти.

Результаты. На основе разработанной имитационной модели, построенной с использованием стохастического подхода, были получены следующие ключевые результаты:

- Рассчитаны стационарные вероятности числа активных заявок в системе, позволяющие оценить степень загрузки вычислительного узла;
- Построено распределение общего объёма используемой оперативной памяти при различных сценариях потребления ресурсов;
- Реализована функция деградации производительности, отражающая снижение скорости обработки заявок при увеличении числа активных процессов.

СПИСОК ЛИТЕРАТУРЫ

1. Ross S. M. Introduction to Probability Models 10th ed. Burlington, MA: Academic Press, 2010. 371 p.
2. Солнышкина И. В. Теория систем массового обслуживания: учеб. пособие Комсомольск-на-Амуре: ФГБОУ ВПО «КнАГТУ», 2015. 76 с.

АСИМПТОТИЧЕСКИЙ АНАЛИЗ МНОГОПОТОЧНОЙ ГЕТЕРОГЕННОЙ РЕСУРСНОЙ СМО

С. П. Моисеева¹, И. А. Туренова¹

¹Национальный исследовательский Томский государственный
университет, г. Томск, Россия

Многомодальные потоки данных представляют собой интегрированную совокупность информации различных типов, включая голосовые, текстовые и видеоданные, что приводит к необходимости использования непуассоновских моделей для их описания [2, 3]. В качестве математической модели предлагается ресурсная система массового обслуживания, где интенсивность потоков изменяется в зависимости от состояний марковской случайной среды. Требования из различных потоков поступают для обслуживания в один из двух блоков, содержащих неограниченное количество приборов, где обслуживаются за произвольно распределенное случайное время. При этом марковская среда не влияет на процесс обслуживания. Во время обслуживания требование занимает некоторый объем ресурса соответствующего блока. Проводится исследование двумерного случайного процесса суммарно занятых ресурсов в системе в условиях высокой интенсивности входящих требований.

Описание модели. На вход системы поступает три потока, один из которых содержит требования одного типа, поступающие для обслуживания в первый блок, другой – второго типа, поступающие на второй блок обслуживания (рис. 1). Третий поток содержит сдвоенные заявки, поступающие одновременно в оба блока обслуживания. Интенсивности каждого из потоков задаются индивидуально и зависят от состояния некоторой управляющей среды, представленной в модели в виде непрерывной цепи Маркова с конечным числом состояний $k(t) = 1, 2, \dots, K$, заданной матрицей инфинитезимальных характеристик $\mathbf{Q} = \|q_{i,j}\|$.

Таким образом, входящие потоки являются марковски модулированными пуассоновскими потоками с общим управляющим процессом $k(t)$ и определяются диагональными матрицами условных интенсивностей $\Lambda^{(1)}, \Lambda^{(2)}$ и Λ с элементами $\lambda_k^{(1)}, \lambda_k^{(2)}$ и λ_k на главных диагоналях соответственно [1]. Поступающее требование мгновенно занимает свободный прибор в блоке, соответствующий его типу, и обслуживается в течение случайного времени с произвольной функцией распределения

$B_i(t), i = 1, 2$. При этом требование занимает ресурс соответствующего блока случайного объема v_i с функцией распределения $G_i(y), i = 1, 2$.

Обозначим $V_i(t)$ – суммарный объем ресурсов, занимаемый требованиями i -го типа, находящимися в системе в момент времени t .

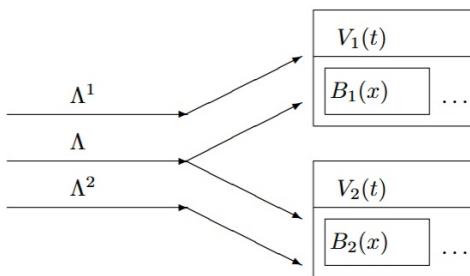


Рис. 1. Ресурсная СМО с входящими ММРР-потоками и двумя блоками обслуживания

Задача исследования двумерного процесса $V(t) = \{V_1(t), V_2(t)\}$ решается с помощью метода динамического просеивания, а также метода асимптотического анализа в условии высокой интенсивности входящих потоков. Построена аппроксимация распределения вероятностей двумерного случайного процесса суммарных занятых объемов ресурсов в системе в виде двумерного нормального распределения с параметрами, определяемыми первыми и вторыми начальными моментами параметров обслуживания и входящих потоков.

СПИСОК ЛИТЕРАТУРЫ

1. Назаров А. А., Моисеева С. П. Методы асимптотического анализа в теории массового обслуживания. Томск: Изд-во НТЛ, 2006. 112 с.
2. Pankratova E. V., Moiseeva S. P., Farhadov M. P., Moiseev A. N. Heterogeneous System MMPP/GI(2)/∞ with Random Customers Capacities // Journal of Siberian Federal University. Mathematics & Physics 2019. Vol. 12. P. 231–239. doi: 10.17516/1997-1397-2019-12-2-231-239
3. Samouylov K., Sopin E., Vikhrova O. Analysis of queueing system with resources and signals. Comm. Com. Inf. Sc. 2017. Vol. 800. P. 358–369. doi: 10.1007/978331968069929

МОДЕЛЬ КОМПОЗИЦИИ ВЕБ-СЕРВИСОВ В ВИДЕ СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ

Р. С. Назаренко¹

¹ Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, г. Саратов, Россия

Рассматривается задача анализа характеристик производительности композиции веб-сервисов, используемой в информационных системах с сервис-ориентированной архитектурой [1]. Описание взаимодействия веб-сервисов и их объединения в единый функциональный процесс выполняется на языке BPEL.

Предлагается математическая модель композиции веб-сервисов в виде открытой экспоненциальной сети массового обслуживания с одним классом требований — сети Джексона [2]. Отдельные веб-сервисы композиции представляются системами массового обслуживания типа $M/M/k$. Процесс обращений к различным веб-сервисам при выполнении запроса пользователя, формализованный в виде программных конструкций языка BPEL, отображается переходами требований между системами массового обслуживания, определяемыми маршрутной матрицей сети. Для того чтобы оценить среднее время отклика композиции веб-сервисов и характеристики отдельных веб-сервисов, вычисляются математическое ожидание длительности пребывания требований в сети массового обслуживания и другие стационарные характеристики сети. Исследуются зависимости характеристик композиции веб-сервисов от ее параметров.

СПИСОК ЛИТЕРАТУРЫ

1. Hoffmann J. Web Service Composition // Encyclopedia of Social Network Analysis and Mining, New York: Springer, 2018. P. 3371–3381.
2. Bolch G., Greiner S., de Meer H., Trivedi K.S. Queueing networks and Markov chains: modeling and performance evaluation with computer science applications. Hoboken: Wiley, 2006. 878 p.

ВЕРОЯТНОСТНЫЕ ХАРАКТЕРИСТИКИ ЗАМКНУТЫХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С ОЖИДАНИЕМ, ОТКАЗАМИ И ОГРАНИЧЕНИЕМ ПО ДЛИНЕ ОЧЕРЕДИ

Т. Э. Петров¹, Н. А. Хасанов¹, А. С. Титовцев¹

¹ Казанский национальный исследовательский технологический университет, г. Казань, Россия

В работе освещаются вероятностные характеристики стационарного режима работы замкнутых СМО с поликомпонентным потоком заявок, в которых имеют место отказы, ожидание и ограничение на длину очереди [1-3]. Рассмотрим замкнутую систему массового обслуживания с входящими ожидающими требованиями, максимальное число которых составляет N . Потоки требований имеют случайную интенсивность λ_k и подчиняются свойствам пуссоновского распределения. Для отображения процессов, происходящих в описанной системе, используется граф состояний СМО (рис 1).

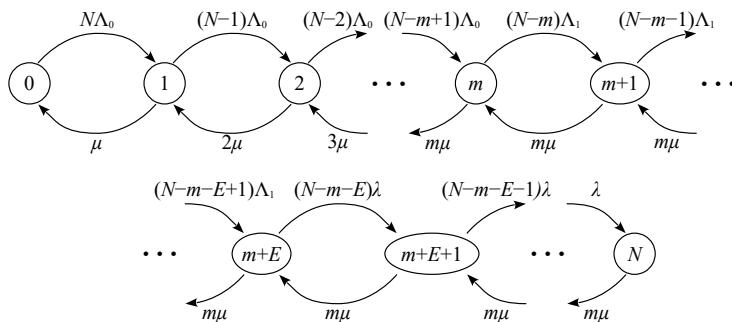


Рис. 1. Граф состояний СМО

Выразив вероятности состояний СМО из системы уравнений Колмогорова через P_0 , получим стационарное распределение вероятностей:

$$P_k = \begin{cases} N^{[k]} \frac{R_0^k}{k!} P_0, & k \leq m; \\ N^{[k]} \left(\frac{R_1}{m} \right)^{k-m} \frac{R_0^m}{m!} P_0, & m < k \leq m + E; \\ N^{[k]} \left(\frac{\rho}{m} \right)^{k-m-E} \left(\frac{R_1}{m} \right)^E \frac{R_0^m}{m!} P_0, & m + E < k \leq N. \end{cases}$$

Вероятность немедленного обслуживания:

$$P_{\text{но}} = \frac{1}{N - \bar{k}} \sum_{k=0}^{m-1} (N - k) P_k = \frac{1}{R_0(N - \bar{k})} \sum_{k=1}^m k P_k.$$

Вероятность ожидания:

$$P_{\text{ож}} = \frac{1}{N - \bar{k}} \left[\frac{\Lambda_1}{\Lambda_0} \sum_{k=m}^{m+E-1} (N - k) P_k + \frac{\lambda}{\Lambda_0} \sum_{k=m+E}^{N-1} (N - k) P_k \right].$$

Вероятность отказа:

$$P_{\text{отк}} = 1 - P_{\text{но}} - P_{\text{ож}}.$$

Результаты данной работы могут быть полезны для моделирования реальных объектов, попадающих под определение замкнутых систем и в которых возникает поток требований неоднородного характера.

СПИСОК ЛИТЕРАТУРЫ

1. Кирпичников А. П., Флакс Д. Б., Галлямова К. Н. Средняя длина очереди в системе массового обслуживания с ограниченным средним временем пребывания заявки в системе // Вестник Технологического университета. 2017. Т. 20, № 2. С. 81–84.
2. Титовцев А. С., Кирпичников А. П. Численно-аналитическое моделирование систем массового обслуживания поликомпонентных потоков // Математические методы в технике и технологиях. 2019. Т. 2. С. 80–84.
3. Гильмутдинов Р. Ф., Кирпичников А. П. Математическая модель замкнутой одноканальной системы массового обслуживания // Вестник Казанского технологического университета. 2012. Т. 15, № 6. С. 189–194.

ИССЛЕДОВАНИЕ ВРЕМЕНИ ОЖИДАНИЯ ЗАЯВКИ В RQ-СИСТЕМЕ С БУНКЕРОМ И ОРБИТОЙ

А. В. Подгайнов¹, А. А. Назаров¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

В современных исследованиях телекоммуникационных и сотовых систем большую роль играют математические модели RQ-систем массового обслуживания [1], обеспечивающие эффективный анализ их функционирования и оптимизацию параметров. В работе исследуется однолинейная RQ-система, которая, помимо орбиты бесконечной длины, имеет бункер длины $N - 1$, используемый для хранения заявок перед обслуживанием на приборе (рис. 1).

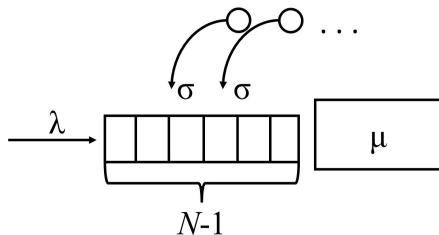


Рис. 1. RQ-система с бункером и орбитой

Входящий поток заявок является простейшим с параметром λ . Время обслуживания заявок на приборе представляет собой экспоненциально распределённую случайную величину с параметром μ , а время ожидания на орбите также распределено экспоненциально с параметром σ . Поступившие в систему заявки обращаются к прибору и начинают обслуживаться, если он свободен. В противном случае они встают в очередь в бункере. Если места в бункере отсутствуют, заявка отправляется на орбиту. После ожидания заявка повторно обращается к прибору и бункеру. Если свободные места снова отсутствуют, заявка вновь отправляется на орбиту, где время её ожидания распределено экспоненциально с тем же параметром σ . Таким образом, в данной системе исключена потеря заявок из-за отсутствия мест в бункере или на при-

боре. Для определения длины бункера в исследуемой системе, помимо распределения вероятностей количества заявок на орбите и среднего числа заявок на орбите, можно использовать такую характеристику, как время ожидания заявки [2] – период с момента её поступления из входящего потока до момента попадания на прибор или в бункер. Таким образом, ставится задача исследования времени ожидания заявки методом асимптотического анализа в условии большой задержки заявок на орбите ($\sigma \rightarrow 0$). По итогу работы найдено распределение вероятностей времени ожидания в виде экспоненциального распределения с параметром $\sigma(1 - \rho)$, где \varkappa_1 – среднее число заявок на орбите и

$$\rho = \frac{1}{1 + \sum_{n=1}^N \left(\frac{\mu}{\lambda + \varkappa_1}\right)^n}.$$

СПИСОК ЛИТЕРАТУРЫ

1. Artalejo J. R., Gomez-Corral A. Retrial Queueing Systems. A Computational Approach. Heidelberg: Springer Berlin, 2008. 267 p.
2. Nazarov A., Samorodova M. Waiting Time Asymptotic Analysis of a M/M/1 Retrial Queueing System Under Two Types of Limiting Condition // Dudin A., Nazarov A., Moiseev A. (eds.) Information Technologies and Mathematical Modelling. Queueing Theory and Applications. ITMM: 2020. Communications in Computer and Information Science. Vol 1391 Cham: Springer, 2021. P. 171–185.

ПОСТРОЕНИЕ И АНАЛИЗ МОДЕЛИ СИСТЕМЫ УПРАВЛЕНИЯ КОНФЛИКТНЫМИ ПОТОКАМИ ПО АЛГОРИТМУ С ОДНОКРАТНЫМ ПРОДЛЕНИЕМ

Т. А. Полетаев¹, А. В. Зорин¹

¹Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, г. Нижний Новгород, Россия

В работе рассматривается СМО с двумя входными потоками и одним обслуживающим устройством (ОУ). Входные потоки $\Pi_j, j = 1, 2$, являются неординарными пуассоновскими потоками и описываются параметрами $\lambda_j \in (0, +\infty)$, $\alpha_j \in (0, 1]$, $\beta_j \in [0, +\infty)$, $\gamma_j \in [0, 1]$. Группа из одного требования прибывает с вероятностью α_j , а из $a = 2, \dots, N$ с вероятностью $\alpha_j \beta_j \gamma_j^{a-1}$ [1]. Требования поступают в накопители неограниченного размера O_1, O_2 . Обслуживание O_j происходит в течение фиксированного времени, затем принимается решение о продлении.

Для построения математической модели используется представление системы обслуживания в виде кибернетической управляющей системы [2, 3]. Зададим дискретную временную шкалу $\{\tau_0, \tau_1, \tau_2, \dots\}$, $\tau_0 = 0$. Моменты времени τ_i являются случайными и соответствуют моменту смены состояния ОУ. Пусть $\Gamma = \{\Gamma^{(1)}, \Gamma^{(2)}, \Gamma^{(3)}, \Gamma^{(4)}, \Gamma^{(5)}, \Gamma^{(6)}\}$ есть множество состояний ОУ. Время нахождения в состоянии $\Gamma^{(s)} \in \Gamma$, $s \in \{1, 2, \dots, 6\}$, есть неслучайная величина $T_s \geq 0$. В состоянии $\Gamma^{(1)}$ обслуживаются требования из первого накопителя. По завершении нахождения в $\Gamma^{(1)}$ возможен переход в состояние продления $\Gamma^{(5)}$ или в состояние переналадки $\Gamma^{(2)}$. Затем в состоянии $\Gamma^{(3)}$ обслуживаются требования из второго накопителя с возможным продлением в $\Gamma^{(6)}$. После переналадки в $\Gamma^{(4)}$ вновь начинается обслуживание первого накопителя. Случайная переменная $\Gamma_i \in \Gamma$ характеризует состояние ОУ на промежутке $(\tau_{i-1}, \tau_i]$, Γ_0 — в момент τ_0 . Наличие продления определяется неслучайными функциями $h_k: X^2 \rightarrow \{0, 1\}$, $h_k(0, 0) = 0$, $X = \mathbf{N} \cup \{0\}$, где 0, 1 соответствуют отсутствию и наличию продления.

Случайный вектор $\varkappa_i = (\varkappa_{1,i}, \varkappa_{2,i}), \varkappa_{j,i} \in X$ характеризует количество требований, накопленных в O_1, O_2 в момент τ_i . Пусть $\eta_{j,i} \in X$ — количество требований поступивших по потоку Π_j за промежуток $[\tau_i, \tau_{i+1})$. Для описания обслуживания требований будем использовать потоки насыщения Π_1^*, Π_2^* . Закон обслуживания определяется величинами ℓ_s , характеризующими максимальное количество требований, об-

служенных за время T_s . Также введём $\xi_{j,i}$ — максимальное количество требований, которые будут обслужены по Π_j^* на $[\tau_i, \tau_{i+1})$.

Результатом работы являются следующие теоремы:

Теорема 1. Последовательность $\{(\Gamma_i, \varkappa_i); i = 0, 1, \dots\}$ является цепью Маркова при заданных распределениях Γ_0, \varkappa_0 .

Теорема 2. Пусть

$$D_1 = \{(\Gamma^{(5)}, x_1, x_2) : x_1 \geq \bar{x}_1 - \ell_5, x_2 \geq \bar{x}_2 : h_1(\bar{x}_1, \bar{x}_2) = 1\},$$

$$D_2 = \{(\Gamma^{(6)}, x_1, x_2) : x_1 \geq \bar{x}_1, x_2 \geq \bar{x}_2 - \ell_6 : h_2(\bar{x}_1, \bar{x}_2) = 1\}.$$

Тогда состояния, принадлежащие

$$M = D_1 \cup D_2 \cup \{(\Gamma^{(r)}, x_1, x_2) : (x_1, x_2) \in X^2, r \in \{1, 2, 3, 4\}\},$$

образуют единственный класс существенных сообщающихся состояний, а состояния из $\Gamma \times X \times X \setminus M$ являются несущественными.

Теорема 3. Существуют $h_1(x_1, x_2), h_2(x_1, x_2)$, такие, что марковская цепь $\{(\Gamma_i, \varkappa_i); i = 0, 1, \dots\}$ имеет различный период.

Также в работе получены рекуррентные соотношения для частичных производящих функций одномерных распределений марковской последовательности $\psi_i(z_1, z_2, s) = \mathbf{M}[z_1^{\varkappa_{1,i}} z_2^{\varkappa_{2,i}} I(\Gamma_i = \Gamma^{(s)})]$. Например, для состояния $\Gamma^{(2)}$ соотношение имеет вид

$$\begin{aligned} \psi_{i+1}(z_1, z_2, 2) &= q_1(z_1, T_2) q_2(z_2, T_2) \times \\ &\times \left(\psi_i(z_1, z_2, 5) + \mathbf{M}[z_1^{\eta_{1,i}} z_2^{\eta_{2,i}} I(\Gamma_i = \Gamma^{(1)}, h_1(\varkappa_{1,i}, \varkappa_{2,i}) = 0)] \right), \end{aligned}$$

где $q_j(z, t) = \mathbf{M}(z^{\eta_{j,i}}) = \sum_{x \in X} z^x \varphi_j(x, t)$, а $\varphi_j(x, t)$ — вероятность поступления x требований по потоку Π_j за время t .

СПИСОК ЛИТЕРАТУРЫ

1. Федоткин М. А., Рачинская М. А. Изучение математической модели трафика автомобилей на основе подхода Ляпунова-Яблонского // Проблемы теоретической кибернетики: материалы XVI Междунар. конф. Н. Новгород: Изд-во Нижегород. ун-та, 2011. С. 508–512.
2. Зорин А. В. Кибернетический подход к построению и анализу математической модели тандема двух перекрестков // Проблемы теоретической кибернетики: Материалы XVI Междунар. конф., Н. Новгород: Изд-во Нижегород. ун-та, 2011, С. 179–183.
3. Зорин А. В. Кибернетическая модель циклического управления конфликтными потоками с последействием // Учёные записки Казанского университета. 2014. Т. 156, № 3. С. 66–75.

ПУАССОНОВСКОЕ СВОЙСТВО МАР-ПОТОКА ДЛЯ ВЫХОДЯЩЕГО ПОТОКА СИСТЕМЫ С ДЕГРАДАЦИЕЙ СКОРОСТИ ОБСЛУЖИВАНИЯ

В. В. Романов¹, И. Л. Лапатин¹

¹Национальный исследовательский Томский государственный университет, г. Томск, Россия

Введение

Для выходящих потоков систем массового обслуживания с неограниченным числом приборов и простейшим входящим потоком есть ряд классических результатов [1, 2]. В них доказано, что если входящий поток является простейшим и времена обслуживаний заявок независимы и одинаково распределены, то выходящий поток будет также простейшим. В данной работе рассматривается выходящий поток системы $M|M_i|\infty$ [3], в котором интенсивность обслуживания зависит от числа заявок в системе $i(t)$ в момент времени t . В этом случае интенсивность обслуживания зависит от состояния системы и меняется для всех заявок одновременно. В этом случае получается, что времена обслуживания не могут быть гарантированно независимыми, соответственно выходящий поток данной системы не удовлетворяет условиям классических результатов середины XX в.

Описание выходящего потока в терминах МАР

Для рассматриваемой системы число заявок в системе $i(t)$ в момент времени t является цепью Маркова. Распределение вероятностей которой \mathbf{r} было получено в работе [3].

Заметим, что события в выходящем потоке наступают только при уменьшении на единицу значения процесса $i(t)$ (один из приборов завершил обслуживание). То есть число заявок в системе является управляющей цепью Маркова для выходящего потока. Соответственно, его можно описать в терминах МАР-потока. Для этого необходимо определить значения элементов матриц, которые определяют МАР-поток. Матрица \mathbf{Q} совпадает с матрицей интенсивностей перехода процесса $i(t)$. Матрица \mathbf{A} является нулевой, поскольку при сохранении значения числа занятых приборов события выходящего потока не наступают. Матрица \mathbf{D}

вероятностей наступления событий при изменении состояния управляющего процесса содержит единицы под главной диагональю, а остальные элементы равны нулю, так как события в выходящем потоке наступают только при уменьшении на единицу значения процесса $i(t)$. Из этих матриц составляется матрица-генератор \mathbf{B} , на диагонали которой стоят элементы матрицы Λ , а внедиагональные элементы равны произведению соответствующих элементов матриц \mathbf{Q} и \mathbf{D} .

Пуассоновское свойство выходящего МАР-потока

В работе [4] сформулированы условия, при которых МАР-поток является пуассоновским. Мы смогли описать наш выходящий поток в виде МАР-потока, поэтому для его параметров можем проверить выполнение этих условий. В результате получили, что для нашего выходящего потока выполняется одно из условий из статьи [4]:

$$\mathbf{r}\mathbf{B} = \lambda\mathbf{B}, \quad (1)$$

где λ – интенсивность входящего (и выходящего) потока системы.

Соответственно, мы показали, что выходящий поток в системе с пуассоновским входящим потоком, неограниченным числом приборов и зависимостью скорости обслуживания от числа заявок в системе, является простейшим.

Данный результат обобщает классические результаты для выходящих потоков систем с неограниченным числом приборов и может применяться при анализе сетевого трафика, телекоммуникационных систем и других стохастических процессов.

СПИСОК ЛИТЕРАТУРЫ

1. D'Auria B. M|M|∞ Queue with ON-OFF Service Speeds // Journal of Mathematical Sciences. 2014. Т. 196, № 5. С. 37–42.
2. Popov S. On transience of M|G|∞ queues // Journal of Applied Probability. 2024.
3. Романов В. В., Лапатин И. Л. Анализ загруженности вычислительного узла при разных распределениях объема поступающих задач // Системы управления, информационные технологии и математическое моделирование. М.: Колос-с, 2024. С. 237–243.
4. Лапатин И. Л. Исследование математических моделей выходящих потоков систем массового обслуживания с неограниченным числом приборов: автореф. дис.... канд. физ.-мат. наук. Томск, 2012. 15 с.

МНОГОЛИНЕЙНАЯ СИСТЕМА С ПОВТОРНЫМИ ВЫЗОВАМИ И ПРИОРИТЕТНЫМИ ЗАЯВКАМИ

А. П. Рычкова¹, С. В. Пауль¹

¹Национальный исследовательский Томский государственный университет, г. Томск, Россия

Рассматривается система с повторными вызовами [2] и отказами, двумя типами входящих потоков и N обслуживающими приборами (рис. 1). На вход системы поступают два простейших потока событий с интенсивностью λ_1 и λ_2 соответственно. Если хотя бы один из N приборов свободен, то поступающие заявки первого потока (приоритетного потока [1]) занимают его и обслуживаются экспоненциальное время с параметром μ_1 . Заявки второго потока могут обслуживаться только на N -ом приборе. Если N -ый прибор свободен, то заявки второго потока, поступающие в систему, занимают его и обслуживаются случайное экспоненциальное время с параметром μ_2 . Если поступившая заявка первого потока застает все приборы занятыми обслуживанием заявки первого потока, она теряется. Если поступившая заявка первого потока застает N -ый прибор занятым обслуживанием заявки второго потока, она вытесняет заявку второго потока на орбиту, где та осуществляет случайную задержку в течение экспоненциального времени с параметром σ , а сама начинает обслуживаться экспоненциальное время с параметром μ_1 . Если поступившая заявка второго потока застает N -ый прибор занятым обслуживанием заявки первого или второго потока, она не теряется, а уходит на орбиту, где осуществляет случайную задержку в течение экспоненциального времени с параметром σ .

Обозначим процесс $k(t)$ как состояние N -го прибора в момент времени t . Этот процесс может принимать следующие значения: 0 – прибор свободен, 1 – прибор занят обслуживанием заявки первого потока, 2 – прибор занят обслуживанием заявки второго потока. Также введем случайный процесс $i(t)$ – число заявок на орбите в момент времени t и процесс $n(t)$ – число занятых приборов заявками первого потока в системе в момент времени t .

Ставится задача нахождения стационарного распределения вероятностей значений процесса $\{k(t), n(t), i(t)\}$ при $k = 0, 1, 2$, при $N = 3$.

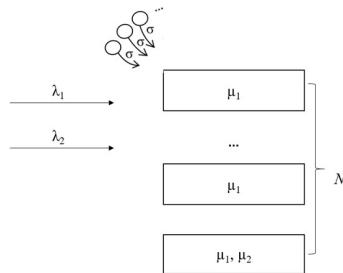


Рис. 1. Система с отказами, двумя типами входящих потоков и N обслуживающими приборами

Для распределения вероятностей $P_k(n, i, t) = P\{k(t) = k, n(t) = n, i(t) = i\}$, $k = 0, 1, 2$, составлена система дифференциальных уравнений Колмогорова. Система была записана в стационарном виде и исследована асимптотическим методом в предельном условии большой задержки заявок на орбите ($\sigma \rightarrow 0$).

СПИСОК ЛИТЕРАТУРЫ

1. Бочаров П. П. Об однолинейной обслуживающей системе с ограниченным числом мест для ожидания и приоритетами // Проблемы передачи информации 1970. С. 70–77.
2. Artalejo J. R. Accessible Bibliography on Retrial Queues // Progress in 2000–2009 Mathematical and Computer Modeling. 2010. Vol. 51. P. 1071–1081.

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ДЛИН ВРЕМЕННЫХ ИНТЕРВАЛОВ В РЕКУРРЕНТНЫХ ПОТОКАХ

Д. Д. Салимзянова¹, С. А. Самойлов¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

Современные телекоммуникационные сети требуют точного анализа потоков событий для оптимизации ресурсов. Оценка параметров распределений интервалов между событиями позволяет прогнозировать нагрузку и снижать задержки. В данной работе рассматривается применение методов машинного обучения для оценки параметров распределений. В работе исследованы длины интервалов между моментами наступления событий в рекуррентном потоке с различными функциями распределения вероятностей (Парето, Фишера, Леви, Фреше, Бура XII, обратное Гамма, Ломакса). Проведено сравнение оценки параметров с использованием двух наборов данных: интервалов между моментами наступления событий и числовых характеристик интервалов. В качестве алгоритмов для оценки параметров был использован градиентный бустинг (CatBoost). Для оценки обобщающей способности моделей использовались следующие метрики: MAPE (средняя абсолютная процентная ошибка) и R^2 (коэффициент детерминации).

Построены модели для каждого распределения на двух наборах данных: интервалах между моментами наступления событий и числовыми характеристиками интервалов (математическое ожидание, среднее квадратичное отклонение, дисперсия, коэффициент вариации, шесть квантилей: 0,1; 0,25; 0,5; 0,75; 0,9; 0,95).

Использование числовых характеристик дало значительное улучшение точности оценок. Например, для параметра μ распределения Леви метрика R^2 увеличилась с 0,032 до 0,851, а метрика MAPE снизилась в 3,6 раза. Для многопараметрических распределений (Фреше и Бура XII) также наблюдается рост точности, например, для параметра t распределения Фреше метрика MAPE снизилась с 235,35 до 52,19.

Использование числовых характеристик длин интервалов распределений в качестве данных для обучения моделей демонстрирует значительное преимущество перед использованием моментов времени наступления событий и интервалов между моментами наступления событий.

ОБ ИССЛЕДОВАНИИ НМ-СЕТИ С ВОЗНАГРАЖДЕНИЯМИ

Д. А. Сальников¹, Т. В. Русилко¹

¹*Гродненский государственный университет имени Янки Купалы,
г. Гродно, Беларусь*

Рассмотрим замкнутую экспоненциальную НМ-сеть массового обслуживания [1], между узлами которой циркулирует K однородных заявок. Перемещение заявок порождает последовательность вознаграждений (доходов), соответствующих переходам сети из одного состояния в другое. Предположим, что сеть включает конечное число узлов с номерами $0, 1, \dots, n$. Узел S_0 является IS-узлом (Infinite Server): число идентичных экспоненциальных линий данного узла равно общему числу заявок в сети – K . IS-узел выступает как конечный источник потока заявок, направляемых на узлы S_1, S_2, \dots, S_n . Интенсивность пуссоновского потока, исходящего из IS-узла S_0 на вход S_i , составляет $\lambda_0 k_0 p_{0i}$, где λ_0 – параметр потока, k_0 – число заявок в S_0 , $i = \overline{1, n}$, $\sum_{i=1}^n p_{0i} = 1$.

Определим характеристики и режим работы узлов S_i , $i = \overline{1, n}$. Каждый из этих узлов представляет собой СМО с t_i идентичными обслуживающими приборами и неограниченным буфером для заявок. Длительности обслуживания заявок каждым прибором СМО S_i имеют экспоненциальное распределение с параметром μ_i , $i = \overline{1, n}$. Очередь заявок в узлах организована по принципу FIFO. Заданы маршрутные вероятности $p_{i0}, p_{i1}, \dots, p_{in}$, которые определяют, с какой вероятностью заявка, обслуженная в узле S_i , $i = \overline{0, n}$, направляется в один из узлов $0, 1, \dots, n$. Матрица $P = (p_{ij})_{i,j=\overline{0,n}}$ является стохастической, причем полагаем, что она неразложима и $p_{ii} = 0$, $i = \overline{0, n}$.

В каждый момент времени t состояние исследуемой сети описывается n -мерным случайным вектором $\mathbf{k}(t) = (k_1(t), k_2(t), \dots, k_n(t))$, где $k_i(t)$ – это число заявок в i -м узле в момент t , $i = \overline{1, n}$. Процесс $\mathbf{k}(t)$ представляет собой цепь Маркова с непрерывным временем и конечным числом состояний.

Предположим, что при переходе заявки с узла S_i на узел S_j НМ-сеть зарабатывает R_{ij} у. е., $i \neq j$, $i, j = \overline{0, n}$. Мы называем R_{ij} вознаграждением.

Работа выполнена при поддержке Государственной программы научных исследований Республики Беларусь «Конвергенция-2025» (подпрограмма «Математические модели и методы»).

граждением, ассоциированным с переходом заявки из S_i в S_j , а $R(\mathbf{k})$ — вознаграждением за единицу времени пребывания НМ-сети в состоянии \mathbf{k} . Обозначим $V(\mathbf{k}, t)$ общее вознаграждение НМ-сети, которое сеть получит, находясь к конечному моменту времени t в состоянии \mathbf{k} .

В отличие от [1] в данной работе получено не обратное, а прямое дифференциально-разностное уравнение для $V(\mathbf{k}, t)$, в котором фигурирует производная по конечному моменту времени t . Однако его решение связано с очевидными техническими затруднениями, что вынуждает прибегнуть к приближенному анализу. В качестве метода исследования выбран асимптотический анализ НМ-сети в случае большого числа заявок K , что позволяет получить приближенные результаты с определенной точностью [1]. В результате осуществлен предельный переход от цепи Маркова $\mathbf{k}(t)$ к непрерывному марковскому процессу, компоненты $k_i(t)/K$ которого отражают число заявок в СМО сети, $i = \overline{1, n}$. Итогом является обобщенное прямое уравнение Фоккера–Планка–Колмогорова для плотности вознаграждения $v(\mathbf{x}, t)$:

$$\begin{aligned} \frac{\partial v(\mathbf{x}, t)}{\partial t} = & - \sum_{i=1}^n \frac{\partial}{\partial x_i} (A_i(\mathbf{x}, t)v(\mathbf{x}, t)) + \\ & + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} (B_{ij}(\mathbf{x}, t)v(\mathbf{x}, t)) + q(\mathbf{x}), \end{aligned} \quad (1)$$

$$A_i(\mathbf{x}, t) = \sum_{j=1}^n \mu_j \min(m_j/K, x_j)(p_{ji} - \delta_{ji}) + \lambda_0 \left(1 - \sum_{i=1}^n x_i \right) p_{0i},$$

$$B_{ii}(\mathbf{x}, t) = \frac{1}{K} \sum_{j=1}^n \mu_j \min(m_j/K, x_j)(\delta_{ji} + p_{ji}) + \frac{1}{K} \lambda_0 \left(1 - \sum_{i=1}^n x_i \right) p_{0i},$$

$$B_{ij}(\mathbf{x}, t) = -\frac{1}{K} \mu_i \min(m_i/K, x_i) p_{ij}, \quad i \neq j,$$

где \mathbf{x} — состояние сети в момент времени t , δ_{ji} — символ Кронекера, $q(\mathbf{x})$ — ставка вознаграждения, установленным образом определяемая вознаграждениями R_{ij} и $R(\mathbf{x})$.

На основе (1) можно определить плотность вознаграждения, интегрируя которую, можно рассчитать вознаграждение сети в момент t .

СПИСОК ЛИТЕРАТУРЫ

1. Rusilko T. V., Salnikov D. A. Asymptotic analysis of a closed G-network with rewards // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2024. № 68. С. 38–47.

ЦИКЛИЧЕСКАЯ СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С Т-СТРАТЕГИЕЙ ПОДКЛЮЧЕНИЯ

В. Д. Слизевич¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

Процесс подачи и обслуживания заявок в технической системе, в которой происходит передача сведений через обслуживающий прибор – шлюз безопасности, занимающийся кодированием данных для взаимодействия с системой межведомственного электронного взаимодействия (СМЭВ), передаваемых ему заявок от разных участников взаимодействия предлагается моделировать в виде циклической модели массового обслуживания (системы поллинга), которая исследуется в работах [1, 2]. На вход системе поступает M простейших потоков с интенсивностями λ_m , $m = 1, \dots, M$ (рис. 1).

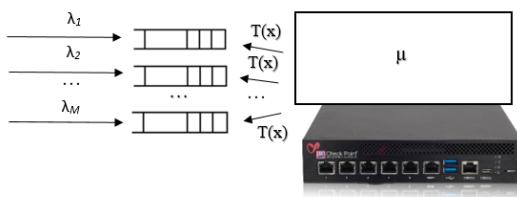


Рис. 1. Модель системы массового обслуживания с одним обслуживающим прибором

Заявки каждого потока образуют свою очередь бесконечной длины. Прибор посещает очереди в циклическом порядке, начиная с первой очереди и заканчивая M -й, потом цикл повторяется. Время пребывания прибора у m -й очереди имеет произвольное распределение с функцией $T(x)$. В течение этого времени прибор обслуживает заявки, которые накопились в очереди с экспоненциальной функцией распределения, заданной параметром μ . Если заявок в очереди нет, то прибор все равно обращен к очереди, пока не истечет время подключения. Будем исследовать исходную систему массового обслуживания методом декомпозиции и рассмотрения системы с прогулками прибора, как представлено в рабо-

те [3], во время прогулки прибор обслуживает другие очереди (рис. 2).

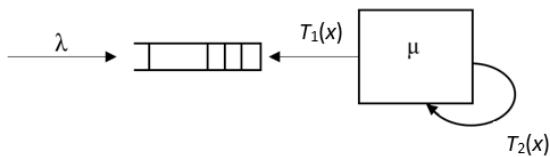


Рис. 2. Модель системы с прогулками прибора

Будем полагать, что продолжительности интервалов обращения к очереди и прогулки случайные и определяются функциями распределения $T_1(x)$ и $T_2(x)$ соответственно. Введем случайный процесс $n(t)$ – число заявок в системе в момент времени t . Для нахождения распределения вероятностей этого процесса и состояний прибора $T_1(x)$ и $T_2(x)$ применяется метод асимптотического анализа в условии большой загрузки системы.

СПИСОК ЛИТЕРАТУРЫ

1. Вишневский В. М., Семенова О. В. Системы поллинга: теория и применение в широкополосных беспроводных сетях. М.: Техносфера, 2007. 312 с.
2. Вишневский В. М., Семенова О. В. Математические методы исследования систем поллинга // Автоматика и телемеханика. 2006. № 2. С. 3–56.
3. Nazarov A. A., Paul S. V. Cyclic Queueing System with Priority Customers and T-Strategy of Service // CCIS. 2016. Vol. 678. P. 182–193.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ГИБРИДНЫХ СИСТЕМ В ВИДЕ СМО С ДВУМЯ БЛОКАМИ ПРИБОРОВ

Я. А. Тюленина¹

¹ Национальный исследовательский Томский государственный университет, г. Томск, Россия

Гибридные системы – это комбинированные решения, которые объединяют элементы различных типов систем для повышения эффективности, гибкости и надежности. Такие системы создают синергетический эффект, позволяя компенсировать недостатки отдельных компонентов за счет их взаимодополняемости. Гибридные системы встречаются в различных сферах, например в сетях связи [1, 2] и энергетических системах [3, 4]. В данном исследовании анализ таких систем проводится с использованием методов теории массового обслуживания (далее СМО).

В данной работе предлагается математическая модель управления гибридными системами в виде СМО с двумя блоками приборов, один из которых ненадежный. В первом блоке (альтернативной электроэнергии) ограниченное число приборов. Во втором блоке (городской электроэнергии) количество приборов не ограничено. Данные блоки могут работать совместно, второй блок работает всегда, а первый включается и выключается с некоторой периодичностью. Время работы первого блока распределено по экспоненциальному закону. Когда накопленная энергия в первом блоке заканчивается, использующий её блок выключается и остается работать только второй блок с городской электрической сетью. В систему поступает простейший поток заявок (запросов на электроэнергию), приоритетным считается обслуживание на первом блоке, если он включен и не переполнен. Если обслуживание на первом блоке невозможно, то заявка начинает обслуживание на втором блоке. Время обслуживания заявки в блоках распределено по экспоненциальному.

Для описанной модели получены формулы моментов первого и второго порядка числа заявок в системе, т.е. объёма потребления. Выполнен сравнительный анализ гибридной системы с системой с одним блоком обслуживания. Сделаны выводы об экономической эффективности гибридной системы в заданных условиях.

Работа выполнена при финансовой поддержке РНФ, грант № 24-71-00022,
<https://rscf.ru/project/24-71-00022/>.

СПИСОК ЛИТЕРАТУРЫ

1. *Вишиневский В. М.* Гибридное оборудование на базе радио- и лазерной технологий // Первая миля. 2007. №1. С. 26–30.
2. *Nadeem F., Gebhart M., Leitgeb E., Kogler W., Awan M. S., Khan M. S., Kandus G.* Simulations and analysis of bandwidth efficient switch-over between FSO and mmW links // IEEE SoftCOM, Split-Dubrovnik, Croatia. 2008. 25–27 September. Р. 356–351.
3. *Новокрещенов О. В., Отмахов Г. С., Хуаде М. Ю.* Комбинированные системы электроснабжения на возобновляемых источниках энергии // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2017. № 132(08). С. 1–12.
4. *Резиньков А. А. , Щекlein С. Е.* Сравнительный анализ эффективности гибридных (солнечно-дизельных) электростанций для регионов России // Энерго- и ресурсосбережение. Энергообеспечение. Нетрадиционные и возобновляемые источники энергии. Атомная энергетика: сб. науч. тр. Екатеринбург: Изд-во Урал. ун-та, 2021. С. 385–388.

ВЕРОЯТНОСТНЫЕ ХАРАКТЕРИСТИКИ ЗАМКНУТЫХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С ТРЕХКОМПОНЕНТНЫМ ПОТОКОМ ЗАЯВОК И ОГРАНИЧЕНИЕМ НА ВРЕМЯ ОЖИДАНИЯ ЗАЯВКИ В ОЧЕРЕДИ

Н. А. Хасанов¹, Т. Э. Петров¹, А. С. Титовцев¹

¹ Казанский национальный исследовательский технологический университет, г. Казань, Россия

В работе приводится математическое описание вероятностных характеристик стационарного режима работы замкнутых СМО с трехкомпонентным потоком заявок и ограничением на время ожидания заявки в очереди. Рассмотрим СМО с количеством обслуживающих устройств m , интенсивностью ухода заявок ν , скоростью обслуживания μ , количеством заявок в источнике N и входящим потоком ожидающих и нетерпеливых требований с интенсивностью Λ_0 и Λ_1 соответственно. При построении математической модели СМО полезным инструментом является граф состояний и переходов системы (рис. 1), учитывающий структуру рассматриваемой СМО.

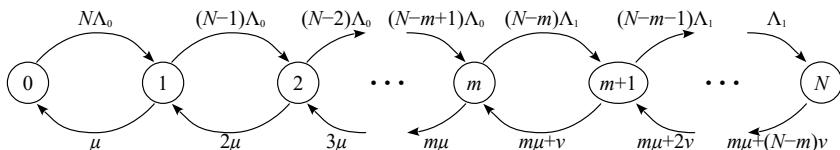


Рис. 1. Граф состояний СМО

Выразив вероятности состояний СМО из системы уравнений Колмогорова через P_0 , получим

$$P_k = \begin{cases} \frac{N!R_0^k}{(N-k)!k!} P_0, & k \leq m; \\ \left(\frac{R_1}{\beta} + 1\right)^{k-m} \left[\left(\frac{m}{\beta} + 1\right)_{k-m} \right]^{-1} \frac{N!R_0^m}{(N-k)!m!} P_0, & m < k \leq N. \end{cases}$$

Вероятность отказа:

$$P_{\text{отк}} = \frac{\lambda_0}{\Lambda_0} \frac{1}{N - \bar{k}} \sum_{k=m}^{N-1} (N - k) P_k = \frac{\lambda_0}{\Lambda_0} \frac{1}{N - \bar{k}} \sum_{k=m}^N (N - k) P_k.$$

Вероятность ожидания:

$$P_{\text{ож}} = \frac{\Lambda_1}{\Lambda_0} \frac{1}{N - \bar{k}} \sum_{k=m}^{N-1} (N - k) P_k = \frac{\Lambda_1}{\Lambda_0} \frac{1}{N - \bar{k}} \sum_{k=m}^N (N - k) P_k.$$

Вероятность немедленного обслуживания:

$$P_{\text{но}} = \frac{1}{N - \bar{k}} \sum_{k=0}^{m-1} (N - k) P_k.$$

Результаты данной работы могут применяться для проектирования систем и объектов, для математического описания которых применяются модели системы массового обслуживания схожего типа.

СПИСОК ЛИТЕРАТУРЫ

1. Кирпичников А. П., Титовцев А. С. Первый и второй моменты общего времени пребывания заявки в системе с произвольным количеством источников и поликомпонентным входным потоком заявок // Вестник Технологического университета. 2016. Т. 19, № 6. С. 120–124.
2. Кирпичников А. П., Титовцев А. С. Характеристики систем дифференцированного обслуживания поликомпонентных потоков // Вестник Казанского технологического университета. 2012. Т. 19, № 8. С. 337–340.
3. Кирпичников А. П., Титовцев А. С. Системы обслуживания с неоднородным входным потоком требований, отказами и очередью // Вестник Казанского технологического университета. 2011. № 5. С. 154–161.

АЛГОРИТМ ПОИСКА ТЕХНИЧЕСКИХ ХАРАКТЕРИСТИК СИСТЕМ С АБСОЛЮТНЫМ И ОТНОСИТЕЛЬНЫМ ПРИОРИТЕТАМИ

Н. Г. Хаустова¹, С. П. Моисеева¹

¹ Национальный исследовательский Томский государственный
университет, г. Томск, Россия

Предлагается модель обработки заявок двух типов — приоритетных и неприоритетных. Время обслуживания заявок случайное с экспоненциальным распределением вероятностей, причем его параметры соответствуют типу заявки. У каждого типа заявок имеется свой бункер ожидания, ограниченный по объему. Информация, находящаяся в очереди, обладает сроком жизни, после истечения которого ее передача может стать неактуальной. Срок жизни — случайная величина, также имеющая экспоненциальное распределение. Для рассматриваемой системы реализован алгоритм поиска технических характеристик путем вычисления стационарного распределения вероятностей, что, в свою очередь, достигается автоматическим построением матриц коэффициентов. Необходимость создания данного алгоритма объясняется нередко возникающими сложностями при моделировании систем с приоритетами [1, 2], что как раз и связано с ограниченной емкостью бункеров ожидания [3] и необходимостью учета нетерпеливости заявок. Представлен обновленный [4] алгоритм для системы с абсолютным приоритетом.

Сформулируем постановку задачи. Пусть $i(t)$ — число приоритетных заявок в бункере в момент времени t , $j(t)$ — число неприоритетных заявок в бункере, $k(t)$ — состояние прибора в момент времени t : $k(t) = 0$ — прибор свободен, $k(t) = 1$ — прибор занят приоритетной заявкой, $k(t) = 2$ — прибор занят неприоритетной заявкой. Ставится задача нахождения стационарного распределения вероятностей трехмерного марковского процесса $\{k(t), i(t), j(t)\}$: $P_k(i, j) = \lim_{t \rightarrow \infty} P\{k(t) = k, i(t) = i, j(t) = j\}$ при следующих параметрах системы: λ_p — интенсивности потоков, μ_p — интенсивности обслуживания, α_p — интенсивности выхода из бункеров, N_p — количество мест для ожидания в бункере, $p = 1, 2$.

Ключевым событием, в котором кроется различие обработки заявок в системе с абсолютным приоритетом сравнительно с относительным, является приход заявки приоритетного типа. Таким образом, если прибор занят неприоритетной заявкой и 1-й бункер свободен, то пришедшая приоритетная заявка вытесняет неприоритетную, и если во 2-м бункере

ре есть место, то число заявок во 2-м бункере увеличивается на 1, состояние прибора равно 1. Способность приоритетной заявки вытеснять неприоритетную при абсолютном приоритете дает хоть и незначительное, но снижение вероятности отказа в обслуживании приоритетных заявок. В сценариях реальной жизни, моделируемых системами с абсолютным приоритетом, счет может идти на секунды — например, оказание медицинской помощи в экстренных случаях.

СПИСОК ЛИТЕРАТУРЫ

1. *Бронштейн О. И., Духовный И. М.* Модели приоритетного обслуживания в информационно-вычислительных системах. М.: Наука, 1976. 220 с.
2. *Степанов С. Н.* Теория телетрафика: концепции, методы, приложения. М.: Горячая линия — Телеком, 2015. 868 с.
3. *Назаров А. А., Змеев О. А.* Ограничение нагрузки в телекоммуникационных сетях // Научное творчество молодежи. Математика. Информатика: материалы XVIII Всерос. науч.-практ. конф. 2014. С. 11–15.
4. *Хаустова Н. Г., Моисеева С. П., Пакулов Е. А.* Алгоритм вычисления стационарного распределения вероятностей системы с приоритетами для произвольного числа мест в очереди // Информационные технологии и математическое моделирование (ИТММ-2024): материалы XXIII Междунар. конф. имени А.Ф. Терпугова, Томск, 20–26 октября г. 2024. 159–165 с.

РАЗВИТИЕ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ С ПОЛИТИКОЙ ПОСТУПЛЕНИЯ ЗАЯВОК «ВЫБОР КРАТЧАЙШЕЙ ОЧЕРЕДИ» («JOIN-THE-SHORTEST-QUEUE»)

Е. А. Хитяев¹, И. С. Зарядов^{1,2}

¹Российский университет дружбы народов, г. Москва, Россия

²Федеральный исследовательский центр «Информатика и управление»
РАН, г. Москва, Россия

Исследование представляет обзор систем массового обслуживания с дисциплиной поступления «выбор кратчайшей очереди» («Join-the-Shortest-Queue» (JSQ)), разработанной для оптимизации времени ожидания начала обслуживания заявки и общей загрузки системы. Кратко сформулирована классическая модель с основными полученными для нее результатами, приведены дальнейшие варианты развития данной дисциплины (функционирование системы в экстремальных условиях; выбор системы для поступления из некоторого подмножества систем). Кроме того, приведены некоторые альтернативы для JSQ-модели.

Классическая JSQ-модель и ее разновидности. Для системы массового обслуживания, состоящей из нескольких параллельных подсистем, каждая из которых включает в себя накопитель и один либо несколько обслуживающих устройств, дисциплина поступления JSQ характеризуется тем, что поступающий поток заявок выбирает подсистему с минимальной длиной очереди [1]. В случае, если при поступлении заявки все подсистемы имеют одинаковые длины очередей, заявка с заданной вероятностью выбирает одну из подсистем [1].

Впервые данная модель, состоящая из двух подсистем с неограниченным числом мест в очереди и отдельным для каждой подсистемы обслуживающим прибором, была рассмотрена в 1958 г. [2].

Анализ классической модели с большим числом накопителей является более трудоемкой задачей в связи с увеличением размерности состояния системы, однако похожие на классическую модели (например, с одним обслуживающим прибором) при трех и более накопителях были исследованы с помощью марковского процесса размножения и гибели [3]. Затем с помощью аналогичных методов было проведено исследование классической модели с большим числом очередей и приборов [4].

На рис. 1 представлены основные виды JSQ-моделей.

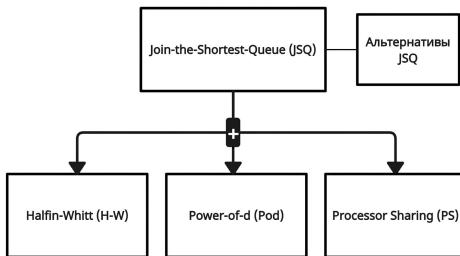


Рис. 1. Развитие JSQ-моделей

В качестве альтернативных моделей JSQ имеются другие подходы для оптимизации обслуживания и увеличения эффективности работы системы: политика Idle-One-First (IIF), требующая меньше информации о состоянии серверов, чем JSQ, но при этом достигающая той же асимптотической производительности за счет отдачи приоритета сначала пустым очередям, а затем очередям с одной заявкой [5]; модель Join-the-Least-Workload (JLW) [6], в которой заявка поступает в очередь с наименьшей текущей загрузкой, и модель RSQ(d) (Replicate to the Shortest Queues), являющаяся обобщением JSQ и JLW моделей [7].

СПИСОК ЛИТЕРАТУРЫ

1. Flatto L., McKean H. P. Two queues in parallel // Communications on Pure and Applied Mathematics. 1977. Vol. 30, № 2. P. 255–263.
2. Haught F. A. Two Queues in Parallel // Biometrika. 1958. Vol. 45, № 3/4. P. 401–410.
3. Lin H.-C., Raghavendra C. S. An analysis of the join the shortest queue (JSQ) policy // Proceedings of the 12th International Conference on Distributed Computing Systems / IEEE. 1992. P. 362–366.
4. Lin H.-C., Raghavendra C. S. An approximate analysis of the join the shortest queue (JSQ) policy // IEEE Transactions on Parallel and Distributed Systems. 1996. Vol. 7, № 3. P. 301–307.
5. Gupta V., Walton N. Load Balancing in the Nondegenerate Slowdown Regime // Operations Research. 2019. Vol. 67, № 1. P. 281–294.
6. Li B., Kong X., Wang L. Optimal Load-Balancing for High-Density Wireless Networks with Flow-Level Dynamics. 2018.
7. Atar R., Keslassy I., Mendelson G. Replicate to the shortest queues // Queueing Syst. 2019. Vol. 92. P. 1–23.

ИССЛЕДОВАНИЕ СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ С ИЗМЕНЯЕМЫМИ ПАРАМЕТРАМИ

Р. Х. Ягудина¹

¹ Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, г. Саратов, Россия

Сети массового обслуживания с изменяемыми параметрами различных классов могут использоваться в качестве математических моделей дискретных стохастических сетевых систем с отказами и восстановлением элементов, а также для анализа разработки методов управления такими системами [1].

Рассматривается экспоненциальная замкнутая сеть систем массового обслуживания типа $M/M/k$ и одним классом требований. Параметры этой сети обслуживания выбраны таким образом, что математическое ожидание числа требований во всех системах обслуживания одинаково. Используя известные методы анализа сетей массового обслуживания [2], проведена серия экспериментов с сетью обслуживания при различных параметрах структуры и интенсивностях обслуживания требований в системах обслуживания.

Проведенное исследование показало, что уменьшение числа приборов или интенсивности обслуживания требований в отдельной системе обслуживания оказывает существенное влияние на стационарные характеристики этой системы. При этом изменение стационарных характеристик в других системах сети незначительно, особенно в сетях большой размерности.

Дальнейшие исследования могут быть направлены на разработку метода адаптивного управления маршрутной матрицей замкнутой сети массового обслуживания.

СПИСОК ЛИТЕРАТУРЫ

1. Шувалов В. П. Обеспечение показателей надежности телекоммуникационных систем и сетей. М.: Горячая линия – Телеком, 2015. 168 с.
2. Митрофанов Ю. И. Анализ сетей массового обслуживания. Саратов.: Научная книга, 2005. 175 с.

A QUEUEING-THEORETIC FRAMEWORK FOR CONTROLLING CUSTOMER FLOW IN LOYALTY PROGRAMS UNDER BUDGET CONSTRAINTS

V. Belolipetskiy¹

¹*Independent Researcher*

In this work, we contribute to queueing theory by applying queueing-theoretic methods as an economic management tool aimed at controlling customer behavior within retail loyalty programs considering system economic state.

Our framework models each loyalty level as a finite-capacity service station within a queueing network. Here, each level represents a section of the customer–product matrix that reflects the ability of customers who have achieved certain shopping milestones to purchase certain products at a discount. Each level has capacity constraints reflecting discount budgets or eligibility limits. Customers, modeled as jobs arriving stochastically (e.g., via a Poisson process with rate λ), are processed sequentially through these levels. The progression from level A_i to level A_{i+1} is governed by the state-update equation:

$$x_{i+1}(t+1) = x_{i+1}(t) + u_i(t) \cdot \mu_i x_i(t) - D_{i+1}(t), \quad (1)$$

where $x_i(t)$ denotes the number of customers at level A_i ; $u_i(t) \in \{0, 1\}$ is a control variable indicating admission to the next level; μ_i is the progression rate (the scaling factor determining the size of the delta from $x_i(t)$); $D_{i+1}(t)$ captures departures from level A_{i+1} . The model enforces capacity constraints:

$$x_i(t) \leq C_i, \quad \forall i, t, \quad (2)$$

which effectively block further transitions when a level is full. This blocking mechanism accounts for additional economic information about the system state and serves as the control lever to regulate customer flow and prevent over-discounting.

By extending the state-dependent F-policy approach, our framework provides a systematic method for dynamically controlling customer progression. Its goal is to optimize key financial metrics by aligning discount allocation with customer behavior, ensuring that discount strategies remain both attractive and financially sustainable [1, 2].

REFERENCES

1. *Jain J., Sanga S. S.* Optimal control F-policy for M/M/R/K queue with an additional server and balking // International Journal of Applied and Computational Mathematics. 2019. Vol. 5. P. 1–16.
2. *Jain J., Sanga S. S.* F-Policy for M/M/1/K Retrial Queueing Model with State-Dependent Rates // Asset Analytics. 2019. P. 127–138.

DISCRETE-TIME MARKOV CHAIN IN CONTAINER VIRTUALISATION

A. A. Chudinova¹

¹ITMO University, Saint-Petersburg, Russian Federation

e-mail: alexandra.a.chudinova@mail.ru

Distributed computer systems will respond to queries reliably and correctly, minimizing the risks of time loss and incorrect behavior of the whole system applying the probabilistic approach to the mathematical modeling. Research is based on the concept of transforming the discrete-time Markov chain formula into the container virtualization distributed computer system under conditions of uncertainty and approval of its applicability. Theoretical model demonstrates the applicability of the discrete-time Markov chain to the container virtualisation.

Introduction. Queuing theory solves the issues with distribution of the requests making it probabilistic and countable. Distributed computing research has led to significant advances in container virtualization, a technology that enables efficient resource management and increased application scalability. However, in contexts characterized by uncertainty, predictability and performance optimization remain open challenges. A promising approach to address these issues is the application of the discrete-time Markov chain, a mathematical technique used to model stochastic systems and predict their future behavior based on previous states. Transforming the Markov chain formula into a container virtualization-based distributed computing system allows for the development of more robust control and optimization algorithms. The main objective of this research is to analyze the feasibility and effectiveness of this approach, assessing its applicability under uncertainty and its impact on dynamic resource management.

Theoretical background. The implementation of the discrete-time Markov chain in container virtualization distributed computing systems is based on a mathematical model that describes the probability of transition between different states of the system over time. We define the system as a series of states $S = \{s_1, s_2, \dots, s_n\}$, where each state represents a specific configuration of the system, such as resource allocation or CPU and memory utilization level. The dynamics of the system is described by a transition

Supported by academic advisor Professor V.V.Bogatyrev, Doctor of technical sciences, ITMO University.

matrix P , where each element p_{ij} represents the probability of transition from state s_i to state s_j in a discrete time interval t [1].

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}, \quad (1)$$

where the sum of the probabilities of each row is equal to 1:

$$\sum_{j=1}^n p_{ij} = 1, \quad \forall i. \quad (2)$$

The state of the system at a given instant t is described by a probability vector $\pi(t)$, where each element $\pi_i(t)$ indicates the probability that the system is in state s_i at time t . The temporal evolution of $\pi(t)$ follows the relation:

$$\pi(t+1) = \pi(t)P. \quad (3)$$

In the long run, the system tends to reach a stationary distribution π^* , which satisfies the following equation:

$$\pi^*P = \pi^*. \quad (4)$$

Under uncertainty, system transitions may be influenced by external factors, such as unexpected workload fluctuations or temporary resource failures. In this case, the model can be extended using a Markov chain with non-deterministic transitions, where the matrix P is replaced by a set of matrices $\{P_1, P_2, \dots, P_m\}$ representing possible scenarios of system condition variations.

An extension of the model includes the use of fuzzy logic to handle uncertainty in state transitions. We define a set of membership functions $\mu(s_i)$ to represent the degree of membership of a state to a given load category (low, medium, high). The update of the probability vector can then be represented as:

$$\pi(t_{11}) = \sum_{k=1}^n \mu_k \pi(t) P_k. \quad (5)$$

Results. The experimental implementation of the proposed approach was conducted in a simulated container virtualization environment using tools such as Kubernetes and Docker. Resource management modeling was

performed by building a Markov chain with states representing different system operating conditions, including variable workload, application response times, and resource failure rate. Results showed that using the Markov chain formula transformation allows for a more accurate prediction of resource allocation needs compared to traditional approaches based on static heuristics. Furthermore, integrating the model with a machine learning-based management system allowed dynamically adapting scheduling decisions to changing system conditions. The effectiveness of the method was evaluated based on metrics such as average response latency, CPU utilization, and load balancing frequency. Simulations highlighted a significant reduction in downtime and increased efficiency in resource use, demonstrating the validity of the proposed approach.

Conclusion. The transformation of the discrete-time Markov chain formula into the container virtualization distributed computing system has proven to be an effective method to address the uncertainty issues in dynamic resource management. Theoretical analysis has confirmed that the model can be adapted to real-world scenarios, while experimental results have demonstrated significant improvements in operational efficiency and system resilience. Integration with machine learning and advanced optimization techniques could further enhance the predictive and adaptive capabilities of the model, opening up new prospects for the use of Markov chain theory in intelligent management of distributed systems. Future research could focus on extending the model to multi-cloud systems and evaluating its performance in large-scale production environments.

REFERENCES

1. Kleinrock L. Queuing theory. Moscow: Mashinostroenie, 1976. 432 p.

VELOCITY-DEPENDENT MOBILE NETWORKS: A CONTINUOUS-STATE MARKOV MODEL WITH PERIODIC BOUNDARY CONDITIONS

A. A. Esin¹

¹IITP RAS, Moscow, Russian Federation

We adopt the classical *wrap-around* model [1, 2] by reducing the infinite 1D lattice of base stations (BS) to a single periodic segment. BS are positioned at $x_j = jD$, $j \in \mathbb{Z}$, with inter-site distance $D > 0$. Owing to spatial homogeneity, we normalise $D \equiv 1$. For any absolute position $x(t) \in \mathbb{R}$ define: $s(t) = \{x(t)\} \in [0, 1] \cong S^1$, $j(t) = \lfloor x(t) \rfloor \in \mathbb{Z}$, so that $x(t) = j(t) + s(t)$. Under uniform motion $x(t) = x(0) + v_0 t$, one obtains: $s(t) = \{x(0) + v_0 t\}$, $j(t) = \lfloor x(0) + v_0 t \rfloor$. Thus, all analysis reduces to motion on the compact manifold S^1 . The mean received power is modelled as a sum of a spatial baseline and a velocity-dependent penalty: $q(s, v_0) = q^*(s) + \Delta_q(v_0)$. To regularise the nondifferentiable minimum function $d(s) = \min\{s, 1 - s\}$, we use a C^∞ -approximation: $d_\varepsilon(s) = \sqrt{s(1 - s)} + \varepsilon^2$, $0 < \varepsilon \ll 1$, and define:

$$q^*(s) = q_0 - a d_\varepsilon(s)^\gamma, \quad \gamma \in [2, 4], \quad a, q_0 > 0. \quad (1)$$

We also take into consideration a *velocity penalty*:

$$\Delta_q(v_0) = -bv_0 - c \ln(1 + v_0), \quad b, c > 0, \quad v_0 \geq 0. \quad (2)$$

Combining (1) and (2) we derived the *total mean signal*:

$$q(s, v_0) = q_0 - a d_\varepsilon(s)^\gamma - bv_0 - c \ln(1 + v_0), \quad (s, v_0) \in S^1 \times \mathbb{R}_+. \quad (3)$$

This function is smooth in s , periodic, and strictly decreasing in v_0 .

Let the state process be

$$X_t := (Q_t, S_t) \in \mathcal{X} := [q_{\min}, q_{\max}] \times S^1, \quad S_t := \{x(0) + v_0 t\},$$

where $q_{\min} < \inf_{s, v_0} q(s, v_0)$ and $q_{\max} > q_0$. We assume reflecting boundaries at q_{\min}, q_{\max} and periodicity in S^1 . The dynamics are:

$$dS_t = v_0 dt, \quad (4)$$

$$dQ_t = -\alpha(Q_t - q(S_t, v_0)) dt + \sigma(Q_t, v_0) dW_t, \quad (5)$$

$$\sigma^2(q, v_0) := \sigma_0^2(1 + \beta v_0^2), \quad \alpha, \sigma_0, \beta > 0.$$

The variable Q_t follows an Ornstein–Uhlenbeck-type diffusion with space–velocity-dependent equilibrium $q(s, v_0)$, see (3). For $f \in C^2(\mathcal{X})$ smooth and periodic in s with Neumann conditions in q , the generator: $\mathcal{L}_{v_0} f(q, s) = v_0 \partial_s f - \alpha(q - q(s, v_0)) \partial_q f + \frac{1}{2} \sigma^2(q, v_0) \partial_{qq} f$. The process X_t is time-homogeneous, as all coefficients are time-independent.

Theorem 1 Stationarity. For any $v_0 > 0$, the process X_t admits a unique stationary density $\pi_{v_0} \in C^\infty(\mathcal{X})$ satisfying:

$$-v_0 \partial_s \pi + \partial_q [\alpha(q - q(s, v_0)) \pi] - \frac{1}{2} \partial_{qq} [\sigma^2(q, v_0) \pi] = 0, \quad (q, s) \in \mathcal{X} \quad (6)$$

with reflecting conditions in q and periodicity in s .

Moreover, π_{v_0} admits the explicit representation:

$$\pi_{v_0}(q, s) = \frac{1}{Z_{v_0}} \int_0^{1/v_0} \frac{1}{\sigma^2(q, v_0)} \exp \left\{ 2 \int_{q_{\min}}^q \frac{-\alpha(x - q(s + v_0 \tau, v_0))}{\sigma^2(x, v_0)} dx \right\} d\tau \quad (7)$$

with normalisation $\int_{\mathcal{X}} \pi_{v_0} = 1$.

Theorem 2 Stability Criterion. Let λ_{in} be the arrival rate. Then the system is stable (positive recurrent) iff

$$\lambda_{\text{in}} < \bar{\mu}(v_0) := \int_{S^1} \int_{q_{\min}}^{q_{\max}} u(q) \pi_{v_0}(q, s) dq ds, \quad (8)$$

i.e. the average service rate exceeds the load, where the throttling policy is given by:

$$u(q) = \begin{cases} \mu_{\text{norm}}, & q \geq q_{\text{good}}, \\ \mu_{\text{slow}} \exp\{-\eta(q_{\text{good}} - q)\}, & q_{\text{poor}} < q < q_{\text{good}}, \\ 0, & q \leq q_{\text{poor}}, \end{cases} \quad q_{\text{poor}} < q_{\text{good}}, \quad \eta > 0, \quad (9)$$

which depends only on Q_t , since $q(S_t, v_0)$ is embedded in the drift.

REFERENCES

1. El-Dolil S. A., Wong W. C., Steele R., Teletraffic performance of highway microcells with overlay macrocell // IEEE J. Sel. Areas Commun. 1989. Vol. 7, № 1. P. 71–78.
2. J. Gilabert et al. Discrete simulation framework for wireless cellular networks // J. Computer Science. 2008. Vol. 4, № 12. P. 982–990.

STOCHASTIC DYNAMICS OF CONNECTIVITY IN QUEUEING NETWORKS WITH MOBILE AGGREGATORS

A. A. Esin¹

¹*IITP RAS, Moscow, Russian Federation*

Introduction

This paper studies a queueing network model with mobile aggregators operating under temporally varying network node availability. The primary focus is the stochastic connectivity dynamics and their impact on buffering parameters for aggregated traffic. Nodes are modeled as queueing systems with Markovian dynamic capacities.

Consider a graph $G = (V, E)$, where V is a set of nodes (base stations), and E are edges indicating mobile aggregator trajectories with speed $p > 0$. Each node $v \in V$ is characterised by:

$$v = \left(q_v, R_v, N(v), S(v), \{\mu_{\max}(v, i)\}_{i=1}^{N(v)} \right),$$

where q_v are node coordinates, R_v is coverage radius, $N(v)$ is the number of channels with random maximum capacities $\mu_{\max}(v, i)$. Channel intensity is given by:

$$\mu(v, i, t) = \mu_{\max}(v, i) \cdot \left[1 - \left(\frac{d(X(t), q_v)}{R_v} \right)^2 \right],$$

where $d(\cdot)$ is Euclidean distance and $X(t)$ the aggregator position.

The aggregator state dynamics are modeled by a finite-state Markov chain with states S_1 (strong signal), S_2 (weak signal), S_3 (no signal), and intensity transition matrix:

$$\mathbf{P} = \begin{bmatrix} -\alpha_1 & \alpha_1 & 0 \\ \beta_1 & -(\beta_1 + \beta_2) & \beta_2 \\ 0 & \gamma_2 & -\gamma_2 \end{bmatrix}, \quad \alpha_1 = \frac{p}{\delta}, \quad \beta_{1,2} = \frac{p}{R - \delta}, \quad \gamma_2 = \frac{p}{\Delta}.$$

Main Results

Theorem 1 (*Stationary Distribution of Queue Length.*) Let the incoming traffic λ_v be distributed among node v channels proportionally to

current intensities $\mu_{v,i}$. Under stability condition:

$$\rho_{v,i} = \frac{\lambda_v \mu_{v,i}}{\sum_j \mu_{v,j} \mu_{v,i}} < 1,$$

the queue length distribution for each channel i at node v is geometric:

$$\pi_{v,i}(n) = (1 - \rho_{v,i}) \rho_{v,i}^n, \quad n \geq 0,$$

with the joint node distribution given by:

$$\mathbb{P}(n_{v,1}, \dots, n_{v,N(v)}) = \prod_{i=1}^{N(v)} (1 - \rho_{v,i}) \rho_{v,i}^{n_{v,i}}.$$

Theorem 2 (*Ergodicity and Exponential Convergence.*) If $\lambda < \bar{\mu}$, the system is ergodic and converges exponentially to stationary distribution π :

$$\|\mathbb{P}(X(t) \in \cdot | X(0)) - \pi(\cdot)\|_{TV} \leq C e^{-(\bar{\mu} - \lambda)t},$$

where $C = \exp(\gamma \|X(0)\|_{\text{init}})$, constants γ , $\|X(0)\|_{\text{init}}$ depend on system parameters.

Theorem 3. (*Buffer Capacity Sufficiency.*) Let the maximum offline duration of the aggregator be:

$$T_{\max} = \frac{\Delta_{\max}}{p}.$$

To prevent data loss, buffer capacity must satisfy:

$$B_{\max} > \lambda T_{\max},$$

ensuring:

$$P\{\sup_{t \geq 0} B(t) \leq B_{\max}\} = 1.$$

Theorem 4. (*Phase Transition with Temporal Connectivity.*) If the coverage probability π exceeds a critical threshold $\pi_c \approx 0.68$, and conditions:

$$B_{\max} > \lambda T_{\max}, \quad \lambda < \mu_{\text{slow}},$$

are satisfied, a temporal connectivity cluster forms, ensuring bounded queue length with probability 1.

СОДЕРЖАНИЕ

Пленарные доклады	5
<i>Вишневский В. М.</i> Направления развития теории очередей и её применение при проектировании телекоммуникационных сетей	6
<i>Зорин А. В.</i> Парадокс восстановления и процессы обслуживания	9
<i>Рыков В. В., Иванова Н. М.</i> Маркованные марковские процессы и их применения	12
<i>Цициашвили Г. Ш., Осипова М. А.</i> Инвариантность стационарных распределений экспоненциальных сетей с запретами	14
Секционные доклады	15
<i>Астафьев С. Н.</i> Анализ модели с многосерверными заданиями в переходном режиме	16
<i>Булгакова Е. С., Фёдорова Е. А.</i> Асимптотический анализ двухклассовой RQ-системы $M_2 M_2 1$ с политикой FIFO в условиях большой загрузки	18
<i>Горбунова А. В.</i> Об анализе некоторых характеристик системы с разделением и параллельным обслуживанием	19
<i>Гордеев В. Е., Зорин А. В.</i> Построение и исследование модели системы управления конфликтными потоками с двумя циклическими режимами	21
<i>Жамсан Д. Д. , Лапатин И. Л.</i> Моделирование энергопотребления в закрытой двухфазной системе массового обслуживания с конечным числом пользователей	23
<i>Жарков М. Л., Казаков А. Л., Лемперт А. А.</i> О применении сетей массового обслуживания для моделирования универсального морского порта	25
<i>Измайлова Я. Е., Исаев А. Т.</i> Асимптотический анализ многолинейной системы массового обслуживания с блоком восстановления	27
<i>Кострюков Н. С.</i> Исследование многопоточных RQ-систем	29
<i>Латипов Т. К.</i> Марковская система с повторными вызовами и приоритетным обслуживанием	31
<i>Лизюра О. Д., Исакова А. Т.</i> Асимптотический анализ RQ-системы с вызываемыми заявками из орбиты	33
<i>Лапатин И. Л., Луценко Н. А.</i> Нахождение параметров МАР-потока при наличии информации о длинах интервалов	35

<i>Малков К. Г., Лапатин И. Л.</i> Имитационная модель использования оперативной памяти вычислительного узла при ограниченном числе пользователей	37
<i>Моисеева С. П., Туренова И. А.</i> Асимптотический анализ многопоточной гетерогенной ресурсной СМО	39
<i>Назаренко Р. С.</i> Модель композиции веб-сервисов в виде сети масштабируемого обслуживания	41
<i>Петров Т. Э., Хасанов Н. А., Титовцев А. С.</i> Вероятностные характеристики замкнутых систем массового обслуживания с ожиданием, отказами и ограничением по длине очереди	42
<i>Подгайнов А. В., Назаров А. А.</i> Исследование времени ожидания заявок в RQ-системе с бункером и орбитой	44
<i>Полетаев Т. А., Зорин А. В.</i> Построение и анализ модели системы управления конфликтными потоками по алгоритму с однократным продлением	46
<i>Романов В. В., Лапатин И. Л.</i> Пуассоновское свойство МАР-потока для выходящего потока системы с деградацией скорости обслуживания	48
<i>Рычкова А. П., Пауль С. В.</i> Многолинейная система с повторными вызовами и приоритетными заявками	50
<i>Салимзянова Д. Д., Самойлов С. А.</i> Применение машинного обучения для анализа длин временных интервалов в рекуррентных потоках	52
<i>Сальников Д. А., Русилко Т. В.</i> Об исследовании НМ-сети с вознаграждениями	53
<i>Слизевич В. Д.</i> Циклическая система массового обслуживания с Т-стратегией подключения	55
<i>Тюленина Я. А.</i> Математическое моделирование гибридных систем в виде СМО с двумя блоками приборов	57
<i>Хасанов Н. А., Петров Т. Э., Титовцев А. С.</i> Вероятностные характеристики замкнутых систем массового обслуживания с трехкомпонентным потоком заявок и ограничением на время ожидания заявки в очереди	59
<i>Хаустова Н. Г., Моисеева С. П.</i> Алгоритм поиска технических характеристик систем с абсолютным и относительным приоритетами	61
<i>Хитяев Е. А., Зарайдов И. С.</i> Развитие систем массового обслуживания с политикой поступления заявок «выбор кратчайшей очереди» («Join-the-Shortest-Queue»)	63

<i>Ягудина Р.Х.</i> Исследование сети массового обслуживания с из- меняемыми параметрами.....	65
<i>Belolipetskiy V.</i> A Queueing-Theoretic Framework for Controlling Customer Flow in Loyalty Programs Under Budget Constraints.....	66
<i>Chudinova A.A.</i> Discrete-time Markov Chain in Container Virtualisation....	68
<i>Esin A. A.</i> Velocity-Dependent Mobile Networks: A Continuous-State Markov Model with Periodic Boundary Conditions	71
<i>Esin A. A.</i> Stochastic Dynamics of Connectivity in Queueing Networks with Mobile Aggregators.....	73

Научное издание

ТЕЗИСЫ
I Международной конференции
ШКОЛА ТЕОРИИ МАССОВОГО
ОБСЛУЖИВАНИЯ (ШТМО-2025)
21–26 апреля 2025 г.

Редактор *Н.А. Афанасьева*
Компьютерная верстка *О.Д. Лизюра*
Дизайн обложки *Л.Д. Кривцовой*

Отпечатано на оборудовании
Издательства Томского государственного университета
634050, г. Томск, пр. Ленина, 36.
Тел. 8+(382-2)-52-98-49
Сайт: <http://publish.tsu.ru>
E-mail: rio.tsu@mail.ru

Подписано к печати 17.05.2025 г.

Формат 60 × 84¹/16. Бумага для офисной техники. Гарнитура «Times».
Печ. л. 4,8. Усл. печ. л. 4,5. Тираж 500 экз. Заказ № 6450.

ISBN 978-5-908040-02-0

