НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ

ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ
им. В.А. ТРАПЕЗНИКОВА РАН

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ (ИТММ-2021)

МАТЕРИАЛЫ
XX Международной конференции
имени А. Ф. Терпугова
1–5 декабря 2021 г.

ТОМСК
Издательство Томского
государственного университета
2022

УДК 519
ББК 22.17
И74
Информационные технологии и математическое моделирование (ИТММ-2021): Материалы XX Международной конференции имени А. Ф. Терпугова (1–5 декабря 2021 г.). — Томск: Издательство Томского государственного университета, 2022. — 391 с.

ISBN 978–5–907572–20–1

Сборник содержит избранные материалы XX Международной конференции имени А.Ф. Терпугова по следующим направлениям: теория массового обслуживания и ее приложения, интеллектуальный анализ данных и визуализация, информационные технологии и программная инженерия, математическое и компьютерное моделирование технологических процессов.

Для специалистов в области информационных технологий и математического моделирования.

УДК 519
ББК 22.17

Конференция проведена при поддержке
международного научно-методического центра
Томского государственного университета по математике,
информатике и цифровым технологиям в рамках
федерального проекта «Кадры для цифровой экономики»
национальной программы
«Цифровая экономика в Российской Федерации»

ISBN 978–5–907572–20–1

NATIONAL RESEARCH TOMSK STATE UNIVERSITY
PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA
V.A. TRAPEZNIKOV INSTITUTE OF CONTROL
SCIENCES OF RUSSIAN ACADEMY OF SCIENCES

INFORMATIONAL TECHNOLOGIES AND MATHEMATICAL MODELLING (ITMM-2021)

PROCEEDINGS
of the 20th International Conference
named after A. F. Terpugov
2021 December, 1–5

TOMSK
Tomsk State
University Publishing
2022

UDC 519
LBC 22.17
I60
Informational technologies and mathematical modelling (ITMM-2021): Proceedings of the 20th International Conference named after A. F. Terpugov (2021 December, 1–5). — Tomsk: Tomsk State University Publishing, 2021. — 391 p.

ISBN 978–5–907572–20–1

This volume presents selected papers from the XIX International Conference named after A.F. Terpugov. The papers are devoted to new results in the following areas: queuing theory and its applications, data mining and visualization, information technology and software engineering, mathematical and computer modeling of technological processes.

UDC 519
LBC 22.17

E d i t o r s:
A.A. Nazarov, Doctor of Technical Sciences, Professor,
S.P. Moiseeva, Doctor of Physical and Mathematical Sciences, Professor,
A.N. Moiseev, Doctor of Physical and Mathematical Sciences, Associate Professor.

The conference was supported by
International Computer Science
Continues Professional Development Center
of the Federal project "Human Resources for the Digital Economy"
of the National program
"Digital Economy of the Russian Federation"

ISBN 978–5–907572–20–1

Информационные технологии и программная инженерия

ЗАДАЧА АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ЗДАНИЙ В ВОДООХРАННЫХ ЗОНАХ НА СПУТНИКОВЫХ СНИМКАХ

С. В. Гилин

Сибирский федеральный университет, г. Красноярск, Россия

Разработка систем распознавания изображений является одним из актуальных и востребованных направлений в области искусственного интеллекта, машинного обучения и анализа данных. В данной работе рассматривается проблема обнаружения на спутниковых снимках зданий и построек, находящихся в водоохранных зонах.

Ключевые слова: *Распознавание образов, распознавание зданий, распознавание водоохранных зон, распознавание на спутниковых снимках, автоматическое распознавание образов.*

Введение

В настоящее время в области развития анализа данных и машинного обучения актуальным является вопрос разработки систем распознавания изображений. Особое внимание уделяется проблеме обнаружения на спутниковых снимках зданий и построек, находящихся в водоохранных зонах. В соответствии с «Водным кодексом Российской Федерации» — «Водоохранными зонами являются территории, которые примыкают к береговой линии морей, рек, ручьев, каналов, озер, водохранилищ и на которых устанавливается специальный режим осуществления хозяйственной и иной деятельности в целях предотвращения загрязнения, засорения, заиления указанных водных объектов и истощения их вод, а также сохранения среды обитания водных биологических ресурсов и других объектов животного и растительного мира», т. е. функционируют ограничения на использование данной территории. Законодательными органами страны принят Водный кодекс, в статьях которого перечислены ограничения на деятельность в водоохранных зонах, а также указаны параметры ширины водоохранной зоны для различных территорий, зависящих от водного объекта и его характеристик. Индивидуальное строительство в водоохранной зоне осуществляется с учетом положений ст. 65 Водного кодекса РФ. Сведения о границах водоохранных зон и границах прибрежных защитных полос водных объектов внесены в Единый государственный реестр недвижимости. Сведения об охранных зонах водных объектов находятся в государственном водном реестре. Строительство на участках, которые полностью или частично входят в водоохранную зону, допускается только при условии, что дом не будет загрязнять водоем и будут соблюдены все санитарные нормы. Строительство любых объектов в водоохраной зоне без согласования влечет за собой ответственность, предусмотренную ст. 8.42 КоАП РФ. К сожалению, как показывает статистика, собранная службами по контролю в сфере природопользования, около 20% застройщиков допускают нарушения при возведении недвижимости в водоохранных зонах. Причём значительная часть нарушителей возводят объекты недвижимости с нарушением границ водоохранных зон, нанося ущерб водным объектам, местной фауне и флоре. Поэтому задача обнаружения зданий и построек, находящихся в водоохранных зонах, является важной и практически значимой. На данный момент самым популярным источником для составления карт местности являются спутниковые снимки, что является следствием повышения качества спутниковых снимков и их доступности. Для удобной работы с картами на спутниковых снимках размечаются различные объекты местности, включая природные и искусственные. В настоящее время самым актуальным способом получения информации является распознавание снимков в автоматическом режиме, что напрямую связано с увеличением объемов участков земли, необходимых для обработки в самые краткие сроки. Машинное обучение при сложившейся специфике обработки информации приводит к необходимости применения машинного метода как базового. При использовании такого метода формируются данные, доступные для аналитики со стороны специалистов, анализирующих результаты обработки снимков.

Целью исследования является разработка нового алгоритма распознавания зданий в водоохранных зонах на спутниковых снимках. Для автоматизации процесса необходимо научить алгоритм добавлять на карты многоугольники домов без участия людей, а также определять — находятся ли постройки на разрешенном расстоянии относительно водоемов.

1. Задача распознавания объектов в водоохранных зонах

Решение поставленной задачи в части снимков, получаемых с космических аппаратов или дронов, которые дают возможность более детального рассмотрения снимков, является частным случаем такой задачи, как распознавание образов. Перечислим основные этапы решения задачи распознавания объектов распознавания зданий в водоохранных зонах на спутниковых снимках — выявление объектов определенных классов на изображениях:

1. Выделение объектов на изображении путем сегментации с дальнейшим преобразованием выделенных сегментов для записи объектов в зависимости от метода дешифрирования.
2. Анализ объектов.
3. Классификация объектов.
4. Нахождение расстояния между распознаваемыми объектами.
5. Вывод результата решения задачи.

При определении распознаваемых объектов, для сокращения выбора построек, предлагается использовать наложение данных Росстата о проверенных зданиях. В данной работе рассматриваются два метода распознавания на спутниковых снимках: автоматическое распознавание построек и распознавание объектов с помощью фильтрации с использованием маски Лапласа.

2. Метод автоматического распознавания контуров зданий

Каждое изображение предварительно обрабатывается, путем извлечения той информации, которая помогает устранить шумы, не подходящие маленькие объекты и разрывы. Здание на изображении — замкнутый контур со сторонами, полученными по значениям следующих составляющих: количество ребер, углов, точек стыковки и отличным от окружения цветом. Алгоритм автоматического распознавания объектов (зданий) выглядит следующим образом:

1. Построение структурной модели изображения.
2. Построение модели в виде графа.
3. Выделение контуров зданий из набора замкнутых контуров, либо их объединений.
4. Постобработка данных для объединения «общей картины» расстановки объектов.

Построение структурной модели изображения

В исследовании в качестве одноименного метода выбран метод попиксельного итеративного утончения, который удовлетворяет критериям выборки. На конечном изображении обозначаются границы, соответствующие принятым критериям:

$$\left(\frac{P_c}{P}\right) \geqslant k_p, \tag{1}$$

$$P_{min} \leqslant P_c \leqslant P_{max}, \tag{2}$$

где P_c — периметр текущего контура; P — периметр минимально описанного прямоугольника вокруг текущего контура P_c, k_p — пороговое значение для определения соответствия текущего контура форме прямоугольника; P_{min}, P_{max} — минимальное и максимальное пороговые значения для текущего периметра. Значение $k_p = \frac{3 \times 3}{10}$, где 3×3 кв. метра размер области на местности, который соответствует одному пикселю изображения.

Построение модели

На этом этапе на обработанном изображении выбираются особые точки (точки, число связности которых не равно двум). Результатом является модель изображения в виде графа $G(P_i, D_{ij})$, где P_i — множество вершин графа $P_i = P_{i_0} \cup P_{i_1}$, P_{i_0} — множество особых точек, каждая из которых соответствует изолированной или точке ветвления на векторной модели изображения (т.е. имеются в виду близкие к водоохранным зонам здания), P_{i_1} — множество промежуточных существенных точек, D_{ij} — множество ребер графа (расстояния между контурами) [1].

Выделение контуров зданий

В данной работе для выделения замкнутых контуров используется пошаговый метод, позволяющий снизить вычислительные затраты. На первом этапе строится базис пространства циклов $\{Z_i\}$ для модели векторного изображения. Определение векторного описания множества базисных циклов $\{Z_i\}$ сводится к поиску ребер и сбору из данных горизонталей целостного объекта [2]. Для определения множества $\{Z_{i_t}\}_{t=0}^{N}$ используется алгоритм эвристического поиска, в котором в качестве функции оценки текущего состояния выступает: $\gamma(C) = \frac{L_0}{L} > \gamma_0$, где C — проверяемый контур, L_0 — суммарная длина ребер, попавших в данные сектора, L — периметр контура, γ_0 — некоторое пороговое значение, γ — правило для принятия решения о том, что контур является зданием [3]. После окончания поиска осуществляется проверка критерия для контура $\acute{C}$ с наилучшей оценкой $\gamma(\acute{C})$ и принятие решения о том, является ли этот контур зданием [4].

Постобработка данных

Задача следующего шага — улучшить полученные данные, заключающиеся в удалении ненужных вершин, выпрямлении углов каждого контура и правильном размещении данных на исходном изображении. Пусть $C = \{C_i\}_{i=1}^{N}$ — множество найденных контуров зданий, $R_i = \{r_j\}_{j=1}^{N_i}$ — множество ребер i-го контура, $A_i =$

$\{(\alpha)_k = \angle(r_k, r_{k+1}) | r_k, r_{k+1} \in R_i\}$ — множество углов i-го контура. Тогда оптимизация векторного представления может быть определена путем минимизации целевой функции следующего вида:

$$R(G) = \sum_{i=0}^{N-1} \sum_{k=0}^{N_i-1} \delta \left(\left| (\alpha)_k - \frac{\pi n}{2} \right| < \varepsilon \right) \cdot \left(\left| (\alpha)_k - \frac{\pi n}{2} \right| \right). \qquad (3)$$

Эта функция, по сути, отвечает за уменьшение отклонения углов контуров от предполагаемых прямых углов. Оптимизация осуществляется путем поиска каждой вершины позиции, в которой обеспечивается минимальное значение вычисляемого $R(G)$ [5]. На последнем шаге выделяем на изображении модели контуров зданий, которые сопоставляются строениям на изображении с помощью полученных данных [6].

3. Метод определения водоохранных зон

Для определения водоохранных зон предлагается при обработке изображения и выделении контуров объекта использовать алгоритм фильтрации с использованием маски оператора Лапласа. При заранее заданных значениях цветов в формате RGB водные ресурсы будут выделены темным цветом, что позволит сократить количество областей распознавания. Алгоритм определения водоохранных зон выглядит следующим образом: Перевод RGB-изображения в цветовую модель YUV (цветовая модель, которая состоит из яркости Y и двух цветоразностных компонент (U и V)), которая задается следующим образом:

$$Y = 0,299R + 0,587G + 0,114B, \qquad (4)$$

$$U = 0,493(B - Y), \qquad (5)$$

$$V = 0,877(R - Y). \qquad (6)$$

Выявление по цвету области водного ресурса с последующим выделением контура. Для данной обработки используется маска оператора Лапласа. К каждой точке изображения с координатами x, y применяется оператор Лапласа, формула которого выглядит следующим образом: $L(f(x,y)) = -f(x-1,y-1) - 2 \cdot f(x,y-1) - f(x+1,y-1) - 2 \cdot f(x-1,y) + 12 \cdot f(x,y) - 2 \cdot f(x+1,y) - f(x-1,y+1) - 2 \cdot f(x,y+1) - f(x+1,y+1)$.

Задание новых значений цвета по адресам фильтра. Данная фильтрация позволяет успешно получать информацию о контурах рельефа местности на спутниковых снимках для последующей векторизации. Распознавание водоохранных зон по заданным характеристикам (контурам и форме, градиенту цвета).

4. Определение расстояния между распознаваемыми объектами

Для решения данной задачи будут использоваться изображения, в которых пикселю изображения соответствует область на местности размером 3×3 кв. метра. Имея данную информацию, вычисляется расстояние от контура объекта проверяемого класса до объекта, который определяется классом водоохранной зоны. В данной работе для определения расстояния используется метрика, введенная Германом Минковским, а именно расстояние городских кварталов:

$$d_j(p,q) = \sum_{j=1}^{m}\sum_{i=1}^{n} |p_{i_j} - q_{i_j}|, \tag{7}$$

где d — расстояние между объектами, p_i и q_i — точки, соединяемые прямой, i — размерность плоскости, j — количество пикселей. С учетом вышесказанного можно построить следующий алгоритм:

1. Выбираются сравниваемые по расстоянию объекты из каждого класса.
2. Определяются всевозможные расстояния между каждой парой точек p и q, которые берутся из каждого класса.
3. С помощью сортировки определяется наименьшее расстояние между объектами.
4. Проверяется соответствие полученного расстояния с контрольными соотношениями ВК РФ. Контрольные соотношения определяются в соответствии с максимальной шириной водоема. Контрольные соотношения задаются формулой:

$$0 \leqslant d_j \leqslant d_w(W), \tag{8}$$

где d_w зависит от ширины водоема W.

5. Если постройка не проходит проверку, то объект можно считать незаконным.

Заключение

В данной статье рассмотрена проблема распознавания зданий в водоохранных зонах на спутниковых снимках. Были описаны метод автоматического распознавания построек и алгоритм определения водоохранных зон с помощью фильтрации на основе маски Лапласа. Метод автоматического обнаружения построек в определенных областях изображений основан на векторной модели и включает в себя поиск замкнутых контуров на изображении с выделением контуров зданий.

СПИСОК ЛИТЕРАТУРЫ

1. *Воронцов К. В.* Лекции по алгоритмам кластеризации и многомерного шкалирования. Москва: МГУ, 2007. 18 с.
2. *Загоруйко Н. Г.* Прикладные методы анализа данных и знаний. Новосибирск: ИМ СО РАН, 1999. 270 с.
3. *Зенин А. В.* Анализ методов распознавания образов // Молодой ученый. 2017. №16. С. 125–130.
4. *Лепский А. Е., Броневич А. Г.* Математические методы распознавания образов. Таганрог: ЮФУ, 2009. 155 с.
5. *Местецкий Л. М.* Математические методы распознавания образов. Москва: МГУ, 2004. 85 с.
6. *Чернов А. В., Чупшев Н. В.* Автоматическое распознавание контуров зданий на картографических изображениях // Компьютерная оптика. 2007. Т. 4. №31. С. 101–103.

Гилин Степан Валентинович — Институт подготовки кадров высшей квалификации. E-mail: *gilin.stepan@mail.ru*

ГИБРИДНЫЙ АЛГОРИТМ ПОИСКА АКАДЕМИЧЕСКОГО ПЛАГИАТА ИСХОДНОГО КОДА С ИСПОЛЬЗОВАНИЕМ ПАРСЕРА ANTLR

А. С. Зоркин, Д. О. Змеев

Национальный исследовательский
Томский государственный университет, г. Томск, Россия

Многие технические высшие учебные учреждения оценивают задачи по программированию студентов на специализированных платформах для проведения контестов, однако при оценке сданных решений преподаватель достаточно много времени тратит на проверку их уникальности. Безусловно, эта рутинная работа может быть автоматизирована путем применения некоторой системы поиска академического плагиата. В данной статье рассматривается гибридный алгоритм сравнения исходных кодов на основе пяти различных методов: по текстовым и бинарным представлениям программы, а также на базе лексем и абстрактных синтаксических деревьев, полученных с использованием парсера ANTLR. Предложенный подход показал высокую эффективность как на детектировании небольших маскирующих изменений, так и на более крупных структурных преобразованиях программного кода.

Ключевые слова: *Плагиат исходного кода, гибридный алгоритм антиплагиата, сравнение схожести программ, способы представления программ, образование в сфере IT, ANTLR.*

Введение

Интерактивная платформа с автоматической проверкой задач «CodeHedgehog» [1] уже несколько лет активно используется преподавателями и студентами Высшей IT-школы ТГУ для изучения основ программирования и алгоритмики. К сожалению, случаи списывания становятся слишком частым явлением, а их обнаружение – довольно трудозатратная операция.

Решением данной проблемы стала разработка модуля антиплагиата для используемой системы, основная цель которого – помощь в вынесении вердикта по оригинальности сданного студентом программного

Авторы выражают благодарность руководству Высшей IT-школы и сотрудникам УЦР ТГУ за оказанную поддержку при создании данной работы

кода для каждой из существующих в рамках контеста задач, например, путем отображения решений, наиболее похожих на целевое. Конечно, существует программное обеспечение, так или иначе решающее данную задачу [2], но оно не удовлетворяет специфике эксплуатируемой платформы и поставленным требованиям на разработку.

Создание описанного модуля усложняется несколькими факторами. Во-первых, многие задачи (например, алгоритмы сортировки) требуют небольшого количества кода и их сложно написать оригинально, поэтому алгоритм должен это учитывать и не давать ложноположительные срабатывания. Во-вторых, студенты часто прибегают к маскирующим преобразованиям с целью скрытия факта списывания, и алгоритм поиска плагиата должен быть предельно устойчив к таким модификациям.

1. Классификация заимствований исходного кода

В исследованиях по данному направлению [3] выделяют около 12 типов заимствования кода: от копирования текста без изменений до структурных модификаций отдельных частей программы. В дальнейшем мы будем придерживаться более общей четырехуровневой классификации [4]:

1) Программный код скопирован без каких-либо изменений (иными словами, идентичен оригиналу с точностью до комментариев);
2) Код скопирован с «косметическими» заменами идентификаторов (имен функций и переменных, типов данных, строковых литералов);
3) Код может включать заимствования второго типа, а также модификации скопированного оригинала путем добавления, редактирования или удаления его фрагментов или изменение порядка их исполнения, не влияющие на логику самой программы;
4) Программа некоторым образом переписана с общим сохранением логики работы и функциональности, однако синтаксически она может абсолютно отличаться от оригинала.

Качественная система антиплагиата должна хорошо определять первые три типа плагиата из классификации выше. Заимствования четвертого типа затруднительны для выявления и часто представляют собой копирование самого алгоритма, а не частей кода оригинальной программы, что не является плагиатом как таковым.

2. Разработка алгоритма антиплагиата

Оценить схожесть двух программ друг с другом можно многими способами. В ходе работы было выявлено и применено пять наиболее универсальных и эффективных подходов:

— Text-based метод заключается в сравнении текстовых представлений программ на основе некоторой метрики, например, расстояния Левенштейна либо Джаро-Виклера. Данный способ отличается своей скоростью и простотой, но результативность быстро снижается даже на «косметических» модификациях.
— Token-based метод основан на преобразовании ключевых слов программы в последовательность лексем языка программирования (далее просто токенов). Полученные токены сравниваются любым доступным способом. Этот алгоритм гораздо точнее предыдущего, так как игнорирует все изменения второго типа, однако он всё ещё не справляется со структурными изменениями.
— Metric-based метод определяет на полученных в предыдущем методе токенах некоторые метрики (например, кол-во используемых циклов и условных конструкций), а далее считает схожесть программ на основе количества совпадающих метрик. В целом, алгоритм может отлично дополняет другие методы антиплагиата, однако он часто даёт ошибочный результат[1].
— AST-based подход требует представления кода в виде абстрактного синтаксического дерева (далее просто AST, Abstract Syntax Tree). В таком формате программы сравниваются любым доступным способом: от «наивного» подсчёта совпадающих узлов до продвинутых методов на основе расстояния Zhang-Shasha [5]. Часто данный способ считается наиболее эффективным [6], но в то же время его сложнее внедрить: помимо построения самих AST, нужна реализация алгоритма их сравнения и могут потребоваться дополнительные преобразования деревьев[2].
— Binary-based метод требует прохождения кодом этапа компиляции (или интерпретации). Полученное бинарное представление (ассемблерный листинг или байткод) служит основой для дальнейшего сравнения программ. Помимо высокой точности на заимствованиях

[1]В частности, метод успешно отрабатывает на программах с переименованием переменных и изменением условий и циклов, но моментально деградирует при добавлении не имеющих смысла операций (по типу объявления пустого цикла или инициализации неиспользуемых переменных)

[2]На практике было выявлено, что парсер может выдать слишком «вытянутые» деревья; такое встречается у языков с большим лексическим словарем (например, C++), а на «сокращенных» деревьях сравнение работает эффективней

1-2 типа[3], алгоритм может дать неплохой результат на модификациях третьего типа[4].

Гибридный алгоритм, заключающийся в использовании комбинации пяти вышеперечисленных подходов, позволяет получить наибольшую эффективность детектирования разного рода заимствований благодаря использованию лучших сторон каждого из данных способов. Метод производит сравнение целевой программы с небольшой выборкой решений других студентов, и его реализация включает в себя несколько простых шагов:

1) Форматирование и нормализация (в т.ч. удаление комментариев) исходного кода для повышения точности текстовых сравнений;
2) Получение промежуточных представлений программы (её токенов, AST и ассемблерного листинг);
3) Непосредственно сравнение полученных форматов вышеперечисленными методами;
4) Вынесение вердикта на основе полученных коэффициентов схожести и вычисление дополнительных полезных статистик, например, дисперсии и квантилей распределения этих коэффициентов для всей выборки решений[5].

В качестве инструмента для получения лексем и AST для каждой программы был использован парсер с открытым исходным кодом ANTLR [7]. Программа широко используется для построения многих языков программирования, утилит и фреймворков, а также может быть настроена для работы с грамматикой любого языка программирования. Сообщество уже добавило поддержку более двух сотен наиболее популярных из них [8]. В модуле антиплагиата данный парсер сыграл решающую роль, позволив реализовать саму логику сравнений, не задумываясь о проблеме получения необходимых для них промежуточных представлений.

[3] Мелкие структурные изменения часто оптимизируются компилятором, а разные реализации одного функционала (к примеру, цикла) дают одинаковый ассемблерный листинг

[4] Однако было выявлено, что некоторые структурные изменения (например, изменение типов данных) могут моментально понизить схожесть программ на несколько десятков процентов

[5] Медианное значение близкое к максимальному и низкая дисперсия говорят о простоте задачи — в этом случае можно уверенно давать отрицательный вердикт на наличие в решении заимствований

3. Полученные результаты

Для тестирования полученного гибридного алгоритма была использована небольшая синтетическая выборка, включающая некоторые модификации программ на языках Java и C++ (750 символов) для всех четырех типов заимствований. В качестве текстовых метрик использовались расстояние Левенштейна и LCS (Longest Common Subsequence), а для сравнения AST использовалась «наивная» стратегия. Усредненные результаты для каждого метода отдельно представлены в табл. 1.

Таблица 1

Средний процент схожести программ с различными модификациями для разных методов антиплагиата

Модификация	Text	Token	Metric	AST	Binary
Изменение комментариев (тип 1)	100%	100%	100%	100%	99%
Реформатирование кода (тип 2)	97%	97%	92%	93%	100%
Добавление неиспользуемых зависимостей (тип 2)	89%	92%	85%	91%	99%
Переименование идентификаторов (тип 2)	85%	100%	100%	96%	98%
Изменение типов данных (тип 2-3)	97%	99%	100%	98%	85%
Изменение порядка исполнения кода (тип 3)	78%	86%	95%	91%	81%
Выделение частей кода в функции (тип 3-4)	65%	74%	72%	51%	67%

Узким местом стало получение промежуточных форматов данных: обращение к ANTLR для каждой программы занимало примерно секунду, в то время как алгоритмическая часть работала почти моментально. Решением проблемы стало использование кэширования представлений, позволившее получать результат антиплагиата примерно за 1-2 секунды для выборки из 50 решений с «прогретым» кэшем.

По результатам итогового тестирования, алгоритм в среднем выдавал следующие проценты схожести программ с заимствованиями различных типов (в зависимости от силы изменений):

1) Заимствования типа 1: **99-100%**
2) Заимствования типа 2: **94-98%**
3) Заимствования типа 3: **87-96%**
4) Заимствования типа 4: **65-82%**

Заключение

Гибридная методика поиска академического плагиата показывает неплохие результаты для детектирования многочисленных кодовых модификаций и довольно устойчива ко многим маскирующим преобразованиям. Стоит отметить и гибкость полученного алгоритма: для сравнения промежуточных представлений, предоставленных парсером ANTLR, можно внедрить любую желаемую стратегию. Значительную часть времени работы может занимать обращение к данному парсеру, однако эта проблема нивелируется использованием кэширования промежуточных форматов программ.

СПИСОК ЛИТЕРАТУРЫ

1. `code.hits.university` — Платформа CodeHedgehog. 2021.
2. *Ahadi A., Mathieson L.* A Comparison of Three Popular Source Code Similarity Tools for Detecting Student Plagiarism // Proceedings of the Twenty-First Australasian Computing Education Conference. 2019. С. 112–117
3. *Roy C., Cordy J.* A Survey on Software Clone Detection Research // Queen's School of Computing TR. 2007. Т. 541, № 115. С. 64–68
4. *Морозов П. А.* Алгоритм выявления плагиата в исходном коде програм // «Информатика: проблемы, методы, технологии» (IPMT). 2021.
5. *Zhang K., Shasha D.* Simple Fast Algorithms For The Editing Distance Between Trees And Related Problems // SIAM Journal on Computing. 1989. Т. 18, № 6. С. 1245–1262
6. *Zheng M., Pan X., Lillis D.* CodEX: Source Code Plagiarism Detection Based on Abstract Syntax Tree // In Proceedings of the 29th Irish Conference on Artificial Intelligence and Cognitive Science (AICS). 2018.
7. `antlr.org` — Официальный сайт парсера ANTLR. 2021.
8. `github.com/antlr/grammars-v4` — Грамматики для ANTLR v4. 2021.

Зоркин Александр Сергеевич — студент НОЦ «Высшая ИТ-школа» НИ ТГУ. E-mail: *somnoynadno@stud.tsu.ru*

Змеев Денис Олегович — ассистент кафедры программной инженерии НИ ТГУ, программист. E-mail: *denis.zmeev@accounts.tsu.ru*

ДИСКРЕТНО-КОСИНУСНОЕ ПРЕОБРАЗОВАНИЕ ДЛЯ СЖАТИЯ ГИПЕРСПЕКТРАЛЬНЫХ ИЗОБРАЖЕНИЙ В ФИТОСАНИТАРНОМ КОНТРОЛЕ ЗЕРНОВЫХ КУЛЬТУР

А. Ж. Саринова, П. А. Дунаев, А. М. Бекбаева

Казахский агротехнический университет им. С. Сейфуллина, г. Нур-Султан, Казахстан

Работа посвящена описанию разработки алгоритмов сжатия гиперспектральных аэрокосмических изображений на основе дискретных ортогональных преобразований с целью последующего сжатия в системах дистанционного зондирования Земли и идентификации заболеваний зерновых культур на основе снимков гиперспектрального изображения. В качестве алгоритмов сжатия, необходимого для сокращения объема передаваемой информации, предлагается использование разработанных методов сжатия на основе преобразований Уолша-Адамара и дискретно-косинусного преобразования.

Ключевые слова: *Гиперспектральные изображения, алгоритмы сжатия, сжатии с потерями, дискретно-косинусное преобразование.*

Введение

Гиперспектральные АИ ДЗЗ необходимы для мониторинга природных ресурсов и последствий чрезвычайных ситуаций и т.д. [1, 2, 3, 4, 5].

Важность проблемы сжатия в том числе связана с проблемой корректного хранения данных дистанционного зондирования, поскольку существующие методы обработки мультиспректральных изображений по фитосанитарному контролю не обеспечивают более точные данные ввиду излишней информации, для этого предлагается использовать гиперспектральные изображения, которые обладают рядом преимуществ для определения заболеваний зерновых культур.

Решение данной проблемы с помощью методов сжатия обеспечивает: повышение эффективности управления сложными техническими систе-

Данное исследование проводилось в рамках грантового финансируемого проекта КОКСОН МОН Республики Казахстана № AP09561922 «Разработка математического аппарата по применению гиперспектральных изображений для фитосанитарного обследования зерновых культур при аэрокосмической съемке».

мами за счет оперативного получения важной информации из сжатых данных; уменьшение больших потоков цифровых данных; повышение памяти выходных устройств записи информации; повышение полосы пропускания канала.

Современные космические системы дистанционного зондирования Земли (ДЗЗ) характеризуются большим разнообразием, в их составе используются космические аппараты (КА), которые ведут съемку в различных диапазонах электромагнитного спектра с разрешением от нескольких километров до нескольких десятков сантиметров [3]. В настоящее время одной из ведущих тенденций развития космических систем ДЗЗ является переход к широкому использованию бортовых многоспектральных и гиперспектральных систем (ГС), работающих в видимом и ИК-диапазонах и осуществляющих съемку в сотнях спектральных каналов [4].

В настоящее время разработка программных систем для сжатия с потерями данных является актуальной задачей. В решении данной задачи существуют различные направления исследования, в которых активно ведутся исследования в области разработки алгоритмов сжатия [6, 7, 8]. Алгоритмы и способы сжатия с потерями данных охватывают широкий спектр сжатия.

1. Этапы алгоритма сжатия

Разработаны алгоритмы обработки гиперспектральных АИ с потерями на основе дискретных преобразований. Последовательность этапов следующая.

1. Преобразование структуры данных на основе исходного гиперспектрального АИ, значения коэффициентов, на основе Уолша-Адамара трех уровней.
2. Оригинальное преобразование структуры данных на основе исходного гиперспектрального АИ, хранящего значения коэффициентов, на основе дискретно-косинусного преобразования с генерированной таблицей квантования.
3. Преобразование полученных структур данных на основе этапов 1-2 путем сгенерированной таблицы квантования коэффициентов.
4. Использование стандартных критериев качества восстановленных изображений.
5. Сжатие полученных структур этапа 4 одним из стандартных энтропийных алгоритмов.
6. Экспериментальное исследование алгоритмов преобразования по степени сжатия и качеству восстановленных данных.

Рассмотрим этапы алгоритмов подробнее.

2. Преобразование Уолша-Адамара

Проведенные эксперименты разработанной программы проиллюстрированы на рисунках 1, 2.

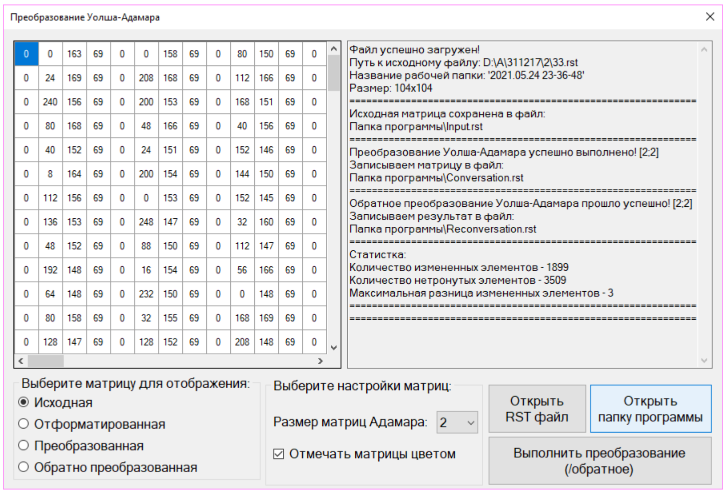

Рис. 1. Преобразование матрицы гиперспектральных АИ

Обратное преобразование

В разработанной программе преобразования Уолша-Адамара для кодирования аэрокосмических изображений включают в себя 2 на 2, 4 на 4, 8 на 8 преобразованных пикселей, в проанализированном обзоре имеющихся научных разработок в таком преобразовании не существует попыток 8 на 8, данный факт является новизной в проведенном исследовании. В дальнейшем планируется исследовать и реализовать вейвлет преобразование Хаара и сравнить экспериментальные исследования с предложенным и разработанным Уолша-Адамара преобразования, а также с существующими научными разработками ученых.

На рисунке 3 представлены исходные и преобразованные изображения после преобразования Уолша-Адамара.

Заключение

На основе проведенных исследований следует сделать следующие выводы:

Рис. 2. Обратное преобразование матрицы гиперспектральных АИ

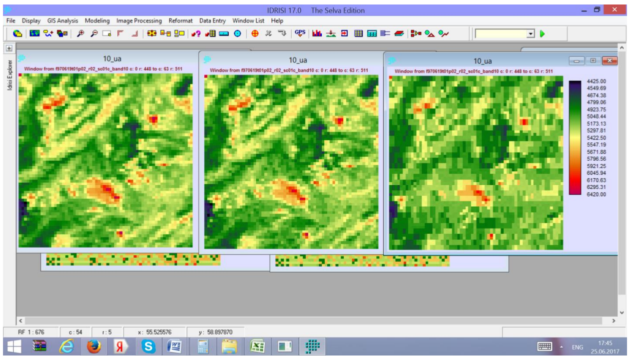

Рис. 3. Исходные и преобразованные изображения

— разработка алгоритмов на основе дискретных преобразований позволяет повысить степень сжатия до $R = 12$;
— предлагаемый подход с потерями определяется в адаптивных преобразованиях, основанных на преобразовании Уолша-Адамара, дискретно-косинусного преобразования и сгенерированной таблицы квантования и последующего адаптивного кодирования;
— полученные результаты сравнения преобразованных гиперспектральных АИ с помощью полученных коэффициентов позволяют

предположить эффективность применения данных исследований с адаптивным кодированием Хаффмана;
- сравнение критериев качества с помощью стандартных метрик качества позволяют отметить, что гипеспектральные АИ восстановлены с высоким качеством и минимальными потерями.

СПИСОК ЛИТЕРАТУРЫ

1. *Гонсалес Р., Вудс Р.* Цифровая обработка изображений. М.: Техносфера, 2012. С. 55–67.
2. *Сэломон Д.* Сжатие данных, изображений и звука. Москва: Техносфера, 2004. 368 с.
3. *Уэлстид С.* Фракталы и вейвлеты для сжатия изображений в действии Название: Фракталы и вейвлеты для сжатия изображений в действии. Издательство: Триумф, 2003.
4. *Штарк Г.-Г.* Применение вейвлетов для ЦОС. М.: Техносфера, 2007. С. 165–177.
5. *Яне Б.* Цифровая обработка изображений. Пер. с англ. М.: Техносфера, 2007. 584 с.
6. *Sujithra D. S., T. Manickam, D. S. Sudheer* Compression of hyperspectral image using discrete wavelet transform and walsh hadamard transform. International Journal of Advanced Research in Electronics and Communication Engineering. 2013. Vol. 2(3). P. 314–319.
7. *JPEG 2000.* [Электронный ресурс] / MainConcept. URL: http://www.mainconcept.com/products/sdks/video/jpeg-2000.html (дата обращения 15.04.2013).
8. *Саринова А.Ж.* Обработка и алгоритмы сжатия гиперспектральных аэрокосмических изображений с учетом междиапазонной корреляции и применения индексирования // Информационные технологии и математическое моделирование (ИТММ-2016) : материалы XVI Междунар. конф. имени А. Ф. Терпугова, 10–15 сент. 2016 г. Ч. 2. Томск: Изд-во НТЛ, 2016. С. 123–127.

Саринова Асия Жумабаевна — к.т.н., PhD, старший преподаватель кафедры эксплуатации электрооборудования. E-mail: *assiya_prog@mail.ru*

Дунаев Павел Александрович — старший преподаватель кафедры «РЭТ». E-mail: *dunayev.kz@mail.ru*

Бекбаева Айгуль Мыктыбаевна — научный сотрудник ГИС центра. E-mail: *bekbaevaaigul@gmail.com*

ОБЗОР СПОСОБОВ РАЗРАБОТКИ КЛИЕНТСКИХ ВЕБ-ПРИЛОЖЕНИЙ И ПРЕИМУЩЕСТВА ИСПОЛЬЗОВАНИЯ ГЕНЕРАТОРА СТАТИЧНЫХ САЙТОВ

С. Ф. Шарапов

Национальный исследовательский
Томский государственный университет, г. Томск, Россия

В настоящее время имеется множество способов разработки клиентской части веб-приложений. Наиболее распространенными среди них являются одностраничные приложения, серверный рендеринг и генератор статичных сайтов. Основными задачами работы являются анализ каждого инструмента с целью выявления их преимуществ и недостатков, а также привести аргументы в пользу использования генератора статичных сайтов.
Ключевые слова: *Веб-приложение, одностраничные приложения, серверный рендеринг, генератор статичных сайтов.*

Для анализа каждого способа разработки в расчет будут приниматься следующие критерии:

— SEO, т.е. насколько хорошо веб-приложение может быть ранжировано в поисковых системах;
— Скорость загрузки и работы веб-приложения;
— Легкость создания динамических пользовательских интерфейсов.

1. Одностраничные веб-приложения

Одностраничные веб-приложения (Single page applications, SPA) имеют только одну HTML-страницу. Во время загрузки веб-приложения, сервер отправляет пустую HTML-страницу, а контент динамически генерируется при помощи специальных JavaScript-файлов в зависимости от действий пользователя [1, 3]. SPA можно создать при помощи фреймворков Vue и Angular или библиотеки React. На рисунке 1 отображена схема работы SPA.

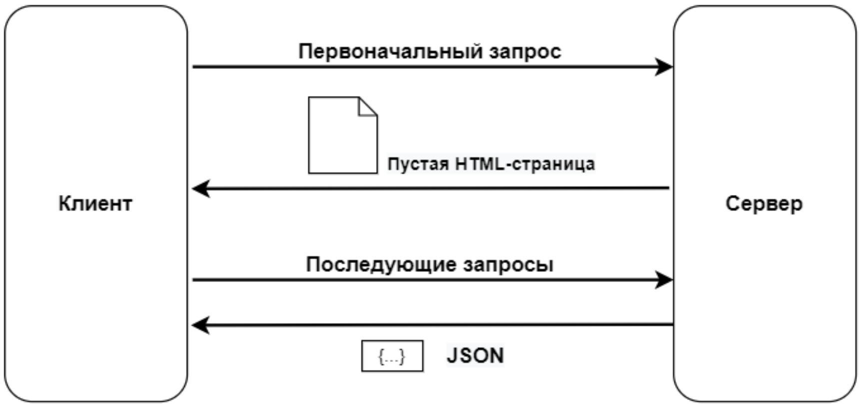

Рис. 1. Схема работы SPA

После первоначального запуска приложения, браузер загружает только контент страницы и динамически отображает их на экране. При этом каждое изменение содержимого страницы не влечет за собой перезагрузку страницы, что существенно повышает скорость приложения и улучшает пользовательский интерфейс [5]. Однако, SPA не являются дружественными к поисковым системам из-за того, что при запуске веб-приложения загружается пустая HTML-страница.

2. Серверный рендеринг

Приложения, разработанные при помощи серверного рендеринга (Server side rendered, SSR-приложения) содержат множество HTML-страниц, которые генерируются на сервере при каждом запросе. Затем, готовые HTML-страницы поступают на браузер и отображаются на экране [3, 4]. При разработке SSR-приложений можно воспользоваться фреймворками Next.js или Nuxt.js. На рисунке 2 представлена схема работы SSR-приложений.

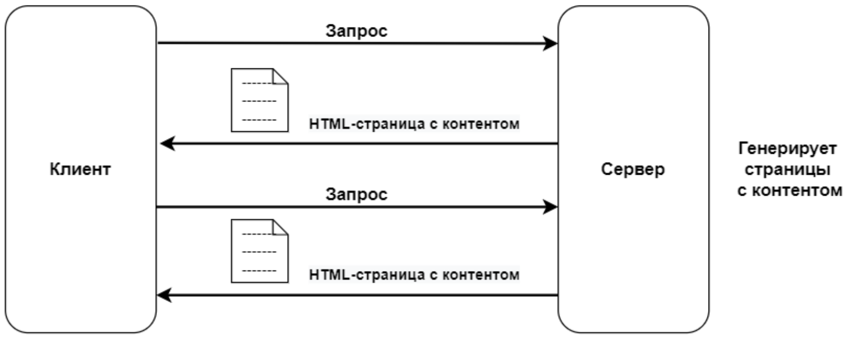

Рис. 2. Схема работы SSR-приложений

При загрузке страницы будут содержать контент, поэтому SSR-приложения являются SEO-дружественными. Сервер может легко встроить динамический контент в эти страницы. Однако необходимость генерации страниц при каждом запросе негативно влияет на скорость загрузки приложения.

3. Генератор статичных сайтов

Генератор статичных сайтов (Static Site Generator, SSG) – инструмент, который генерирует HTML-страницы на основе созданных разработчиком шаблонов. Веб-приложения, созданные посредством SSG, состоят из множества статичных HTML-страниц с нужным контентом, которые компилируются во время сборки приложения до ее развертывания [2, 3]. Затем, HTML-страницы с контентом, а также JavaScript-файлы загружаются на сервер. При первоначальном запросе от клиента, сервер отправляет нужную страницу, а далее приложение будет функционировать как SPA из-за наличия JavaScript-файлов, которые могут вставлять динамический контент в веб-страницу. Это освобождает от необходимости генерировать и загружать HTML-страницы при каждом запросе. На рисунке 3 приведена схема работы SSG-приложений.

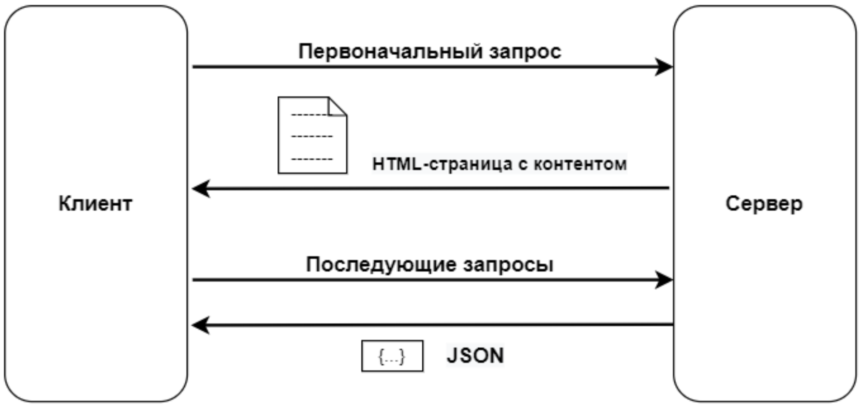

Рис. 3. Схема работы SSG-приложений

В SSG-приложениях поисковые системы могут распознать содержимое страниц из-за наличия контента в них. Кроме того, такие приложения обладают большей производительностью, так как страницы уже сгенерированы в момент сборки. Фреймворки Gatsby и Gridsome являются хорошими инструментами для разработки SSG-приложений.

Заключение

Результаты всех рассмотренных способов разработки веб-приложений можно подытожить в таблице 1.

Таблица 1

Таблица соответствия рассмотренных способов разработки критериям

Способ разработки	SEO	Скорость загрузки и работы	Легкость создания динамических пользовательских интерфейсов
Одностраничные приложения (SPA)	Не соответствует	Соответствует	Соответствует
Серверный рендеринг (SSR)	Соответствует	Не соответствует	Соответствует
Генератор статичных сайтов (SSG)	Соответствует	Соответствует	Соответствует

При необходимости разработки SEO-дружественных веб-приложений с динамическим пользовательским интерфейсом, имеющих с высокую скорость загрузки и работы, целесообразно использовать генератор статичных сайтов.

СПИСОК ЛИТЕРАТУРЫ

1. *Banks A., Porcello E.* Learning React. Sebastopol: O'Reilly Media Inc., 2017. 350 p.
2. `https://gatsbyjs.com/docs/` — Gatsby documentation. 2021.
3. `https://cutt.ly/1YLCTwA` — What is a Single Page Application? Pros and Cons of the Technology. 2019.
4. `https://nextjs.org/docs` — Next.js documentation. 2021.
5. `https://reactjs.org/docs/getting-started.html` — React.js documentation. 2021.

Шарапов Сарвар Фарход угли — магистрант ИПМКН НИ ТГУ. E-mail: *sarvarik883@gmail.com*

Моделирование телекоммуникационных сетей связи

RESPONSE TIME ANALYSIS IN FOG COMPUTING SYSTEM WITH THRESHOLD-BASED OFFLOADING MECHANISM

N. Ashurmetova[1], E. Sopin[1,2]

[1] *Peoples' Friendship University of Russia (RUDN University), Russian Federation*
[2] *Institute of Informatics Problems, FRC CSC RAS, Russian Federation*

Fog computing systems can solve many problems that clouds cannot effectively deal with, such as providing real-time analysis and local management capabilities, connecting and protecting powerful devices with limited network penetration, bandwidth, and availability. In this work, we develop the probabilistic model with three types of nodes that describe the task service process on mobile devices, the fog node and the remote cloud. Taking into account the heterogeneity in the processing volumes of mobile computing tasks, we present the analytical framework to analyze the response time characteristics of the model.

Keywords: *Fog computing, offloading, queuing model, response time.*

Introduction

As fog computing spreads from cloud devices to end users, it has a profound impact on communication with customers [1]. Wireless gateways in homes are becoming hazy nodes that will serve not only as a network gateway, but also as a local computing and storage server [2, 3]. An in-vehicle fog node can make the vehicle an integral part of the end-to-end services provided by fog computing and cloud devices, for example by helping to update software on many microcomputers in the car and allowing the user to reconfigure applications and user interface in the car [4, 5].

The main motivation of the fog computing deployment is strict requirements on the response time of real-time mobile applications. Fog computing brings the computing infrastructure closer to the end user, thus decreasing the response time [6].

The reported study has been funded by RFBR, project 20-07-01052.

In this paper, we assume that the processing volume of mobile tasks is a discrete random variable with a certain probability distribution. A description of the system, a mathematical model is presented, and an analysis of the results is given and conclusions are drawn.

1. System Description

We consider a mobile computing offload scheme with three computing levels: a mobile device, the fog node and the remote cloud computing system. Mobile devices run applications, which consume a large amount of computing resources and power. According to the internal policy of a mobile device, some tasks are offloaded to the fog node. Since the capacity of the fog node is limited, some of the tasks are sent to the remote cloud.

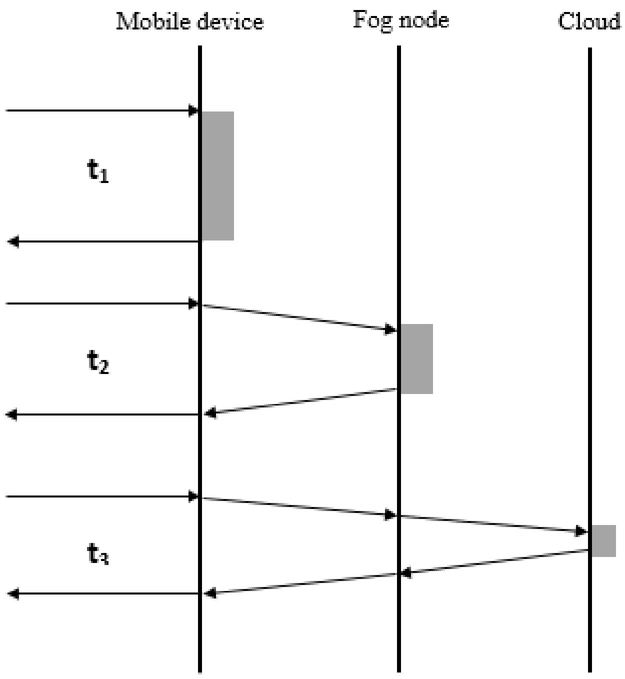

Figure 1. An example for local processing of a task, offloading to the fog node and offloading to the cloud.

Figure 1 shows an example of offloading and sending a request to the fog node. The response time t_1 corresponds to the local processing of a task on mobile device. In this case, the mobile device consumes much power, but has the lowest response time, which consists of only the processing time. The response time t_2 corresponds to the offloading to the fog node. In this case, the mobile device wastes energy on sending a task into the fog over a wireless channel, but saves energy on computations. The resulting response time is the sum of the transmission delay and the processing time at the fog node. The last response time t_3 corresponds to the case when a task is sent to the cloud, but at the same time it takes a lot of time to deliver data.

2. Mathematical model

Figure 2 shows the proposed offloading scheme. As shown in Figure 2, mobile devices are connected to the fog node, and each of mobile devices generates a Poisson flow of tasks with the intensity $\lambda_i, i = 1, 2, ..., N$.

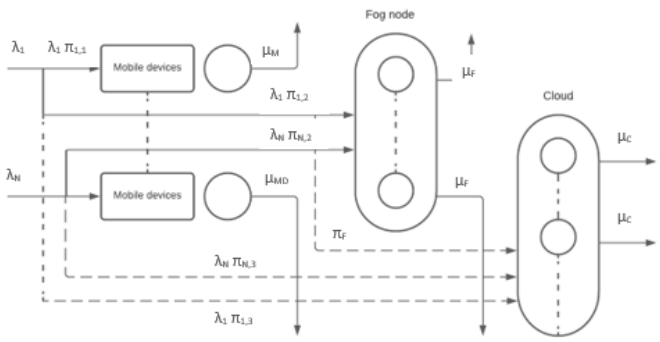

Figure 2. Model in terms of a queuing systems.

Let v_i be a random value of the processing volume of a task from mobile device i with the distribution function $V_i(x)$, and μ_i is the constant task serving rate, measured in MIPS (Million Instructions Per Second). We assume that the "light"tasks (that is, $v_i < W_1$) are processed locally, if the processing volume of a task is between W_1 and W_2 ($W_1 \leqslant v_i < W_2$), then it is sent to the fog node, and "heavy"tasks ($v_i \geqslant W_2$) - to the cloud. Thus, the offloading probability from the i-th mobile device can be obtained using the following formula:

$$\pi_{i,1} = V_i(W_1), \tag{1}$$

$$\pi_{i,2} = V_i(W_2) - V_i(W_1), \tag{2}$$

$$\pi_{i,3} = 1 - V_i(W_2). \tag{3}$$

Let us define the distribution function $V_{MD,i}(x)$ of the processing volume of tasks that are served locally. Using the rules for calculating conditional probabilities, we get

$$V_{MD,i}(x) = P\{v_i < x | v_i < W_1\} = \frac{P\{(v_i < x) \cap (v_i < W_1)\}}{P\{v_i < W_1\}} =$$
$$= \begin{cases} \dfrac{P\{v_i < x\}}{P\{v_i < W_1\}}, x \leqslant W_1; \\ \dfrac{P\{v_i < W_1\}}{P\{v_i < W_1\}}, x > W_1; \end{cases} = \begin{cases} \dfrac{V_i(x)}{\pi_{i,1}}, x \leqslant W_1; \\ 1, x > W_1. \end{cases} . \tag{4}$$

Then, we can evaluate the CDF of the locally processing time:

$$T_{MD,i}(x) = P\left\{\frac{v_i}{\mu_i} < x \mid v_i < W_1\right\} = P\{v_i < \mu_i x | v_i < W_1\} =$$
$$= V_{MD,1}(\mu_i x) = \begin{cases} \frac{V_i(\mu_i x)}{\pi_{i,1}}, x < \frac{W_1}{\mu_i}, \\ 1, x \geqslant \frac{W_1}{\mu_i}. \end{cases} \tag{5}$$

The distribution of the service time makes it easy to obtain the average service time at mobile device i $t_{MD,i}$, as well as the second moment $t^{(2)}_{MD,i}$. These parameters are used to estimate the average sojourn time $\tau_{MD,i}$ for a locally processed task using the well-known Pollachek-Khinchin formula:

$$\tau_{MD,i} = t_{MD,i} + \frac{\lambda_i t^2_{MD,i}}{2(1 - \lambda_i t_{MD,i})}. \tag{6}$$

Consider the CDF of the processing volume of tasks from i-th mobile device on the fog node $V_{F,i}(x)$.

$$V_{F,i}(x) = P\{v_i < x \mid W_1 \leqslant v_i < W_2\} = \frac{P\{v_i < x, W_1 \leqslant v_i < W_2\}}{P\{W_1 \leqslant v_i < W_2\}} =$$
$$= \frac{P\{v_i < x, W_1 \leqslant v_i < W_2\}}{\pi_{i,2}} = \begin{cases} 0, x \leqslant W_1; \\ \frac{V_i(x) - V_i(W_1)}{\pi_{i,2}}, W_1 < x \leqslant W_2, \\ 1, x > W_2. \end{cases} \tag{7}$$

Similarly, the processing time CDF $T_{F,i}$ on a fog node is:

$$T_{F,i}(x) = P\left\{\frac{v_i}{\mu_F} < x | W_1 \leqslant v_i < W_2\right\} = V_{F,i}(\mu_F x). \tag{8}$$

We also calculate the CDF of the processing volume of tasks from the i-th mobile device on the cloud $V_{C,i}(x)$.

$$V_{C,i}(x) = P\{v_i < x \mid v_i \geqslant W_2\} = P\frac{\{v_i < x, v_i \geqslant W_2\}}{P\{v_i \geqslant W_2\}} =$$
$$= \frac{P\{v_i < x,\ v_i \geqslant W_2\}}{\pi_{i,3}} = \begin{cases} 0, x \leqslant W_2, \\ \frac{V_i(x) - V_i(W_2)}{\pi_{i,3}},\ x > W_2. \end{cases} \tag{9}$$

The distribution of the processing time on the cloud for tasks originated from mobile device i has the following form:

$$T_{C,i}(x) = P\left\{\frac{v_i}{\mu_C} < x \mid v_i \geqslant W_2\right\} = V_{C,i}(\mu_C x). \tag{10}$$

The total response time is the conditional sum of processing and transmission delays. The task from the i-th MD is processed locally with the probability $\pi_{i,1}$; on the fog node with probability $\pi_{i,2}$ and on the cloud with probability $\pi_{i,3}$. Thus, the CDF of the response time $T_i(x)$ is

$$T_i(x) = (1 - \pi_{i,1})T_{MD,i}(x) + \pi_{i,2}T_{F,i}(x - 2\Delta_1) +$$
$$+ \pi_3 T_{C,i}(x - 2\Delta_1 - 2\Delta_2), \tag{11}$$

where Δ_1 and Δ_2 are the constant transmission delays from a mobile device to the fog node and from the fog node to the cloud, respectively.

Finally, the average response time τ_i of the i-th mobile device is

$$\tau_1 = (1 - \pi_{i,1})\tau_{MD,i} + \pi_{i,2}(2\Delta_1 + \tau_{F,i}) + \pi_3(2\Delta_1 + 2\Delta_2 + \tau_{C,i}), \tag{12}$$

where $\tau_{F,i}$ and $\tau_{C,i}$ are the average serving times on the fog node and the cloud, respectively.

3. Conclusion

In this article, we develop the analytical framework for the analysis of the fog computing system with the threshold-based offloading mechanism. We take into account that tasks generated by applications on mobile devices can vary significantly in terms of the processing volume. We accurately derived formulas for CDFs of the processing time on each node and obtained the CDF of the response time.

REFERENCES

1. *Kunjan Shah, Sina Khatibi.* An analysis of redundancy schemes on cloud radio access network reliability // Nomor Reseach GmbH Munich, Germany.
2. *Aboobucker Ilmudeen.* Big data-based frameworks for healthcare systems // Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare analytics, 2021. *https://www.sciencedirect.com/topics/engineering/fog-computing.*
3. *Sweta Jain, Ayushi Srivastava.* The fundamentals of Internet of Things: architectures, enabling technologies, and applications // Healthcare Paradigms in the Internet of Things Ecosystem, 2021.
4. *S. Barbarossa, S. Sardellitti, P.D. Lorenzo.* Communicating while computing: distributed mobile cloud computing over 5G heterogeneous networks // IEEE Signal Processing Magazine. 2014. Vol. 31, no. 6. P. 45–55.
5. *Eva Marín Tordera, Xavi Masip-Bruin, Jordi García-Alminana, Admela Jukan, Guang-Jie Ren, Jiafeng Zhu, Josep Farre.* A tutorial on current concepts towards a common Definition // Technical university of Germany.
6. *Sopin E.S., Daraseliya A.V., Correia L.M.* Performance analysis of the offloading scheme in a fog computing system // 10th International congress on ultra modern telecommucations and control system and woprkshops (ICUMT). 2018. P. 1–5.

Ashurmetova Nargis — PhD student at Applied Probability and Informatics Department. E-mail: *mfox262@mail.ru*

Sopin Eduard — candidate of science in mathematics and physics, associate professor at Applied Probability and Informatics Department. E-mail: *sopin-es@rudn.ru*

APPLICATIONS OF k-OUT-OF-n:G SYSTEM AND MACHINE LEARNING METHODS ON RELIABILITY ANALYSIS OF UNMANNED HIGH-ALTITUDE MODULE

N. M. Ivanova[1,2], V. M. Vishnevsky[1]

[1] *Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia*
[2] *Peoples' Friendship University of Russia (RUDN University), Moscow, Russia*

The article is aimed at researching the reliability of an unmanned high-altitude module based on a mathematical model of k-out-of-n:G system. The analytical model of k-out-of-n:G system is considered and the impact of the coefficient of variation of the system elements' lifetime on its operating probability without failure is investigated. Several machine learning methods are used to calculate reliability characteristics for arbitrary input data based on practically significant parameters, the accuracy of which is expressed in terms of estimated mean values.

Keywords: *Tethered unmanned high-altitude platform, k-out-of-n system, system's reliability, coefficient of variation, sensitivity analysis, machine learning, neural network.*

Introduction

Unmanned aerial vehicles (UAVs) are widely used in both civil and military fields. They are used in various situations, for example, for organizing telecommunications infrastructure, monitoring, and video surveillance, creating navigation systems, and others [1]. In this regard, ensuring the reliable and long-term functioning of such systems is an important task. Currently, high-altitude tethered telecommunication platforms based on UAVs have been widely developed [2]. The long-term operation of such platforms is due to the possibility of high-power transmission of energy through a cable-rope that connects the aircraft and ground vehicles.

Due to the multi-rotor architecture of the high-altitude module, k-out-of-n systems, and some others, which are based on Markov process approaches, are often used to describe its behavior with further reliability analysis [3], [4]. These models have been widely studied, various assumptions about the structure of such a model, for example, the dependence and independence

Supported by the Russian Foundation for Basic Research, project no. 19-29-06043 and the RUDN University Strategic Academic Leadership Program.

of the system's elements, the shape of life and repair times distributions, and others were considered [5]. However, the influence of some hidden characteristics such as second moment and coefficient of variation has not been sufficiently studied.

In this paper, the reliability of unmanned high-altitude module of a tethered telecommunication platform is investigated using the k-out-of-n:G system, the influence of the coefficient of variation on the system's characteristics is analyzed, and machine learning, namely neural networks, is applied, which makes it possible to determine with high accuracy a satisfactory level of module reliability at different initial parameters.

1. Problem setting

Due to the multi-rotor architecture of the high-altitude module, which consists of n identical engines, consider homogeneous k-out-of-n:G system. Such a system consists of n elements and remains operational iff at least k out of n elements are operational. Denote by $A_i, \quad i = 1, 2, ...,$ lifetimes of the system elements. Suppose that these random variables are independent and identically distributed, thus the corresponding cumulative density function is defined as $A(t) = \mathbf{P}\{A_i \leqslant t\}$. Suppose also that instantaneous failures are impossible and their mean times are finite:

$$A(0) = 0, \quad a = \int_0^\infty (1 - A(t))dt.$$

For the system investigation introduce process $J = \{J(t),\ t \geq 0\}$ with

$$J(t) = \text{number of working components in time } t$$

with the set of states $E = \{j = \overline{0, k}\}$, where j is number of working units.

Denote also by T time to first system failure $T = \inf\{t : J(t) \in \mathbb{E}_1\}$, where $E_1 = \{j = \overline{0, k-1}\}$ is the set of good (UP) states. Thus, we are interesting in calculation of reliability function

$$R(t) = \mathbf{P}\{T > t\}.$$

2. Analytical Model and Sensitivity Analysis

For homogeneous k-out-of-n:G system we have $A_i(t) = A(t)$ $(i = \overline{1, n})$. It is well known, the probability that at time t exactly i elements of the system from n are in a working state has the form

$$\mathbf{P}(t) = C_n^i (1 - A(t))^i A(t)^{n-i}.$$

Thus, the reliability function of such a system (the probability of the system operating for a certain amount of time t without failure) is

$$R(t) = \mathbf{P}\{T > t\} = \sum_{i \geqslant k}^{n} C_n^i (1 - A(t))^i A(t)^{n-i}. \tag{1}$$

As an example, consider the 4-out-of-6:G system. It is supposed that the lifetime of the system's units have the following distributions:

— Gamma $\left[\Gamma\left(1/v^2, av^2\right)\right]$;

— Gnedenko-Weibull $\left[GW\left(\mu, \dfrac{a}{\Gamma(1+1/\mu)}\right)\right]$;

— Log-normal $\left[LnN\left(\ln\dfrac{a}{\sqrt{1+v^2}}, \sqrt{\ln\left(1+v^2\right)}\right)\right]$,

where a is the expected lifetime of the system's elements and v is its coefficient of variation. μ is the shape parameter of GW distribution, it is selected based on the value of v.

In our experiments we choose $a = 1$ and $v = [0.1, 0.5, 1, 5, 10]$.

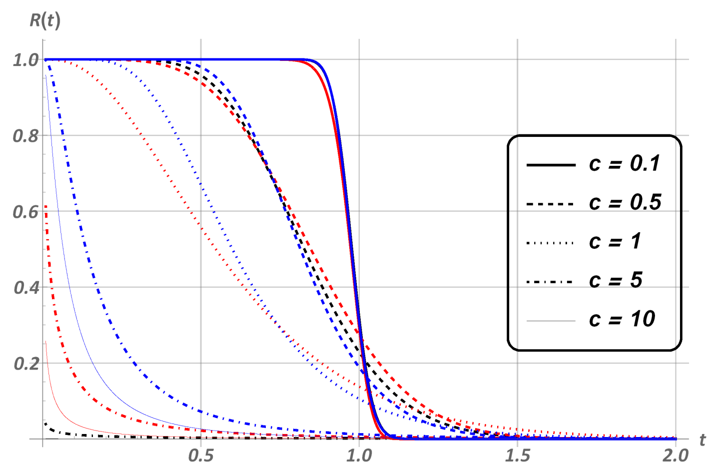

Figure 1. Reliability function $R(t)$ on the time t

Fig. 1 shows the dependence of system reliability function on the time t. Black, red and blue colors denote the Γ, GW, and LnN distributions, respectively. As it can be seen from the curves, the reliability function of the system is asymptotically insensitive to the form of the distribution function of the lifetime at fixed mean and coefficient of variation $v \leq 1$. At the same time for $v > 1$, we can notice that this insensitivity disappears and the system loses its reliability very quickly. We can conclude that the behavior of the system depends on the value of v.

3. Machine learning methods and their application to the task

In queuing and reliability theories, machine learning methods are used to study various probabilistic and time characteristics of complex systems, as well as where it is impossible to obtain results either analytically or using simulation.

The application of machine learning techniques for analyzing the reliability of an unmanned high-altitude module is due to the following factors. Firstly, the conclusion of Section 2 makes it clear that some hidden parameters of the system, namely the coefficient of variation in our case, have a significant impact on its behavior and performance. Secondly, from a practical point of view, the operating time of the system is estimated by its average value, while the shape of the lifetime distribution function is unknown and can only be assumed based on some statistical data. Moreover, the proposed model can be useful and used by engineers at the development stage of such modules to determine a highly reliable system architecture (parameters k, n), select components that satisfy a sufficient value of the parameter v, and also predict how long the module will operate a satisfactory level of system reliability.

As machine learning methods [6], we will consider the followings from scikit-learn (for regressions) and TensorFlow (for neural network) libraries.

— Linear regression (LinReg),
— Polynomial regression (degree = 3) (PolyReg),
— K-nearest neighbors regression (n_neighbors = 5) (KNN),
— Regression chain (Multi-Output Regression) with cross-validation (scoring = MSE) based on Support vector regression (kernel = rbf, order=[1,0]) (MultiReg),
— Artificial neural network with two hidden layers (optimizer = RMSprop(1e-3), loss = MSE, batch size = 96) (ANN).

To train the model using Formula (1) the training dataset was generated, in which $A(t) \sim \Gamma$, and further was divided to train and test sets.

Consider the results of application of machine learning techniques for analyzing the reliability of an unmanned high-altitude module. Tab. 2 shows the mean square error (MSE) for the two predicted values on the training set. Tab. 3 demonstrates MSE, mean absoulute error (MAE) as well as the coefficient variation (R^2).

Table 1

Variables for machine learning models and their ranges

Type	Variables	Symbol	Range
Input	Total number of system's elements	n	$4-10$
	Needed number of elements in operating states	k	$2-(n-1)$
	Mean lifetime	a	$0.1-1$
	Coefficient of variation	v	$0.01-1$
Output	Reliability	R	$0.5-1$
	Time to system's satisfactory level of reliability	t	>0

Table 2

Accuracy of training

	LinReg	PolyReg	KNNg	MultiReg	ANN
MSE of R	0.009382	0.001858	0.001411	0.100988	0.018045
MSE of t	0.032887	0.021047	0.004005	0.022871	0.042639

Table 3

Accuracy of testing

	LinrReg	PolyReg	KNN	MultiReg	ANN
MSE of R	0.00933	0.001875	0.002359	0.015983	0.018042
MSE of t	0.019145	0.022580	0.003869	0.050609	0.042577
MAE of R	0.073982	0.029946	0.020822	0.092072	0.105326
MAE of t	0.109634	0.111218	0.040997	0.177255	0.168281
R^2 of R	0.469286	0.894559	0.897647	0.097334	0.142539
R^2 of t	0.747440	0.706204	0.934491	0.342006	0.239067

Analyzing the results obtained, we can note that MSE estimate for all cases lies in a satisfactory interval. MAE estimate shows the relative value of the prediction error. In our task, MAE $\geqslant 0.05$ is considered unsatisfactory. Therefore, of all cases, only KNN shows the achieved accuracy result. R^2 estimate indicates how well the constructed model adequately describes the initial data. From it, we see that the best result is again shown by K-nearest neighbors regression.

4. Conclusion

The paper investigates the reliability of an unmanned high-altitude module based on a mathematical model of the k-out-of-n system and machine learning methods. Analytical results show the dependence of the system reliability on the coefficient of variation of the lifetime. The application of machine learning methods has shown that K-nearest neighbours regression describes the reliability of the system in the best way. In the future, it is planned to improve the selected model to achieve more accurate predictions.

REFERENCES

1. *Mozaffari M., Saad W., Bennis M., Nam Y.-H., Debbah M.* A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems // IEEE Communications Surveys & Tutorials. 2019. P. 410–438.
2. *Vishnevsky V., Meshcheryakov R.* Experience of Developing a Multifunctional Tethered High-Altitude Unmanned Platform of Long-Term Operation // Lecture Notes in Computer Science, Cham: Springer, 2019. Vol. 11659.
3. *Vishnevsky V.M., Kozyrev D.V., Rykov V.V., Nguyen Z.F.* Reliability modeling of an unmanned high-altitude module of a tethered telecommunication platform // Informatsionnye Tekhnologii i Vychslitel'nye Sistemy. 2020. № 4. P. 26–38 (in Russian).
4. *Aslansefat K., Marques F., Mendonca R., Barata J.* A Markov Process-Based Approach for Reliability Evaluation of the Propulsion System in Multi-rotor Drones // Technological Innovation for Industry and Service Systems. DoCEIS 2019. IFIP Advances in Information and Communication Technology, Cham: Springer, 2019. Vol. 553, P. 5–6.
5. *Ivanova N.M.* Reliability function of a homogeneous k-out-of-n system with dependent failures of its elements // Materialy Vserossiyskoy konferentsii s mezhdunarodnym uchastiyem "Informatsionno-telekommunikatsionnyye tekhnologii i matematicheskoye modelirovaniye vysokotekhnologichnykh sistem", Moscow: RUDN, 2020. P. 34–39 (in Russian).
6. *Bonetto K., Latzko V.* Computing in Communication Networks From Theory to Practice // Chapter 8 - Machine learning, Cambridge: Academic Press, 2020. P. 135–167.

Ivanova Nika M. — Junior Scientific Researcher, ICS RAS; Ph.D. Student, RUDN University. E-mail: *nm_ivanova@bk.ru*

Vishnevsky Vladimir M. — Dr. Sc. (Techn.), Prof., the Head of Network systems control Laboratory, ICS RAS. E-mail: *vishn@inbox.ru*

АСИМПТОТИЧЕСКИЙ АНАЛИЗ G-СЕТИ С НЕНАДЁЖНЫМИ МНОГОЛИНЕЙНЫМИ СИСТЕМАМИ ОБСЛУЖИВАНИЯ

Д. Я. Копать

Гродненский государственный университет имени Янки Купалы, г. Гродно, Республика Беларусь

Объектом исследования данной статьи G-сеть с ненадёжными многолинейными системами массового обслуживания (СеМО), в случае, когда общее число заявок в системах сети, а также число линий обслуживания (ЛО) большое, но ограниченно. Данная сеть является математической моделью биллинговой системы. Для данной СеМО проведен асимптотический анализ. Выведено уравнение Колмогорова–Фоккера–Планка для плотностей распределения вероятностей состояний. С помощью аппарата характеристических функций выведены и решены системы дифференциальных уравнений (ДУ), которым удовлетворяют средние числа заявок и ЛО в системах сети, а также дисперсии данных чисел.

Ключевые слова: *Асимптотческий анализ, среднее число заявок, дисперсии числа заявок, ненадёжное обслуживание, G-сеть.*

Введение

G-сети массового обслуживания (СеМО), впервые введенные в статье [1] в стационарном режиме и в статье [2] переходном имеют широкое применение в качестве математических моделей компьютерных систем и сетей (КС) с учётом действия на них компьютерных вирусов. В статье [3] были найдены доходы данной сети. В статье [4] в стационарном режиме было рассмотрено поведение компьютерных вирусов не только как повреждение пользовательских запросов, но и вывода из троя компьютера пользователя, т.е. так называемых загрузочных вирусов. Это G-сети с ненадёжными системами массового обслуживания (СМО). Однако в работах [5] была рассмотрена данная сеть в переходном режиме методом последовательных приближений. Асимптотический анализ СеМО при большом, но ограниченном числе заявок был впервые введён в работе [6], для G-сетей без особенностей – в работе [7]. Данная статья обобщает результаты работ [5], [7]. Будем использовать методику получения результатов в асимптотическом случае большого числа заявок [8]. Цель работы – рассчитать среднее и дисперсию числа заявок в СМО сети и числа исправных ЛО, исследовать корреляцию между числом заявок в разных системах.

1. Описание сети

Рассмотрим открытую G-сеть [1] , состоящую из n СМО, в каждой из которых функционирует m_i ненадежных линий обслуживания (ЛО) и в которую поступают из внешней среды простейшие потоки положительных и отрицательных заявок с интенсивностями соответственно λ_{0i}^{+} и λ_{0i}^{-} , $i = \overline{1, n}$. Времена исправной работы и восстановлени ЛО имеют показательную функцию распределения (ф.р) с параметрами α_i, γ_i, $i = \overline{1, n}$ соответственно. Длительности восстановления ЛО также считаются независимыми и распределенными по показательной ф.р. Таким образом число исправно работающих ЛО в каждой из СМО может быть описано процессом гибели и размножения. Обозначим через $d_i\,(t)$ количество исправных ЛО в момент времени $t, 0 \leqslant d_i(t) \leqslant m_i$. Процессы, описывающие изменение числа исправных ЛО в разных СМО сети, независимы. Длительности обслуживания положительных заявок одной ЛО в СМО S_i распределены по экспоненциальному закону с параметром μ_i, $i = \overline{1,\ n}$. Отрицательная заявка, поступившая в некоторую непустую СМО сети, уничтожает одну заявку в ней, не требует обслуживания и сразу покидает сеть. Поэтому в каждой СМО сети могут обслуживаться только положительные заявки, и говоря об обслуживании положительных заявок, обычно для краткости их называют просто заявками. Положительная заявка после обслуживания в СМО S_i с вероятностью p_{ij}^{+} переходит в СМО S_j как положительная заявка, с вероятностью p_{ij}^{-} как отрицательная заявка и с вероятностью $p_{i0} = 1 - \sum\limits_{j=1}^{n} \left(p_{ij}^{+} + p_{ij}^{-}\right)$ покидает сеть, $i,\ j = \overline{1, n}$.

Состояние данной сети в момент времени t описывается вектором

$$\mathbf{z}(t) = (\mathbf{d}(t);\ \mathbf{k}(t)) = (d_1(t), ..., d_n(t); k_1(t), ...,\ k_n(t)), \tag{1}$$

где $k_i(t)$ – число заявок в системе S_i в момент времени t. Функционирование сети определяется двумя одновременно протекающими процессами – $d(t)$ и $k(t)$. Процесс $z(t)$ является цепью Маркова (ЦМ) с непрерывным временем. Пусть I_i обозначает n-вектор все компоненты которого нулевые, за исключением i-ой, равной единице, $i = \overline{1, n}$. Можно показать, что вероятности состояний $P(\mathbf{d}, \mathbf{k}, t)$ удовлетворяют следующей системе разностно-дифференциальных уравнений (РДУ):

$$\frac{dP(\mathbf{d},\mathbf{k},t)}{dt} = -\sum_{i=1}^{n} \left[\lambda_{0i}^{+} + \lambda_{0i}^{-} + \mu_i \min(d_i,k_i) + \gamma_i\,(m_i - d_i)\,u\,(d_i)\right) +$$

$$+\alpha_i d_i]\ P\,(\mathbf{d},\mathbf{k},t) + \sum_{i=1}^{n} \lambda_{\ 0\ i}^{\ +} P\,(\mathbf{d},\mathbf{k} - I_i,\ t) + \gamma_i\,(m_i - d_i + 1)\,u\,(d_i) \times$$

$$\times P\,(\mathbf{d} - I_i,\ \ \mathbf{k},\ \ t) + \alpha_i\,(1 + d_i)\,P\,(\mathbf{d} + I_i,\ \ \mathbf{k},\ \ t) + \sum_{i=1}^{n} \left[\lambda_{\ 0i}^{\ -} + \right.$$

$$+\mu_i \min\,(d_i, k_i + 1) \left(p_{i0} + \right.$$

$$\left.\left. + \sum_{j=1}^{n} p_{ij}^{-}\,(\ 1 - u\,(k_j))\right)\right] P\,(\mathbf{d},\mathbf{k} + I_i\ ,\ t) + \sum_{i\ ,\ j\ =\ 1}^{n} \mu_i \min\,(d_i, k_i + 1) \times$$

$$\times \left[\,p_{ij}^{+} u\,(k_j)\,P\,(\mathbf{d},\ \mathbf{k} + I_{\ i} - I_j, t) + p_{ij}^{-} P\,(\mathbf{d},\ \mathbf{k} + I_{\ i} + I_j, t)\right]. \qquad (2)$$

2. Вывод уравнения Колмогорова-Фоккера-Планка

Для решения системы (1) будем рассматривать асимптотический случай большого числа положительных заявок в СеМО и ЛО во внешней среде,которое обозначим через K, т. е. положим, что $K >> 1$. Для того, чтобы найти вид распределения вероятностей состояний случайного вектора $\mathbf{k}(t)$, перейдем к относительным переменным и рассмотрим вектор $\mathbf{x}\,(t) = K^{-1}\mathbf{k}\,(t)\,,\ y_i\,(t) = K^{-1}d_i\,(t)$. Возможные значения этого вектора при фиксированном t принадлежат ограниченному замкнутому множеству

$$G = \Big\{(\mathbf{y},\mathbf{x}) = (y_1,\ ..., y_n, x_1,\ ..., x_n) : x_i \geqslant 0,\, y_i \geqslant 0,\, i = \overline{1,\, n},$$
$$\left.\sum_{i=0}^{n} x_i \leqslant 1,\ \sum_{i=0}^{n} y_i \leqslant m_i K^{-1}\right\}. \qquad (3)$$

Воспользуемся следующей аппроксимацией функции $P\,(\mathbf{d},\mathbf{k},t)$:

$$K^{2n} P\,(\mathbf{d},\ \mathbf{k},\ t) = K^n P\,(\mathbf{y}K,\ \mathbf{x}K,\ t) = p\left(\mathbf{y},\ \vec{x},\ t\right), (y,\ x) \in G,$$

тогда система РДУ (1) перепишется в виде:

$$\frac{\partial p(\mathbf{y},\ \mathbf{x},\ ,t)}{\partial t} = -\ K\sum_{i=1}^{n}\ \left[\varepsilon\left(\lambda_{0i}^{+}+\lambda_{0i}^{-}\right)+\mu_i\min\left(y_i,x_i\right)+\right.$$
$$\left.+\gamma_i\left(l_i-y_i\right)u\left(y_i\right)+\alpha_i y_i\right]\ p\left(\mathbf{y},\ \mathbf{x},\ ,t\right)+$$
$$+K\sum_{i=1}^{n}\left[\lambda_{\ 0\ i}^{\ +}\,\varepsilon p\left(\mathbf{x}-e_i,\ t\right)+\gamma_i\left(l_i-y_i+\varepsilon\right)u\left(y_i\right)p\left(\mathbf{y}-e_i,\ \ \mathbf{x},\ \ t\right)+\right.$$
$$\left.+\alpha_i\left(\varepsilon+y_i\right)p\left(\mathbf{y}+e_i,\ \mathbf{x},\ t\right)\right]+K\sum_{i=1}^{n}\left[\lambda_{\ 0i}^{\ -}\,\varepsilon+\mu_i\min\left(y_i,x_i+\varepsilon\right)\left(p_{i0}+\right.\right.$$
$$\left.\left.+\sum_{j=1}^{n}p_{ij}^{-}\left(\ 1-u\left(x_j\right)\ \right)\right)\right]p\left(\mathbf{y},\ \mathbf{x}\ +e_i,\ t\right)+K\sum_{i\ ,\ j\ =\ 1}^{n}\mu_i\min(y_i,$$
$$x_i+\varepsilon)\left[\ p_{ij}^{+}u\left(x_j\right)p\left(\mathbf{y},\ \mathbf{x}\ +e_{\ i}-e_j,t\right)+p_{ij}^{-}p\left(\mathbf{y},\ \mathbf{x}\ +e_{\ i}+e_j,t\right)\right]\ . \qquad (4)$$

В случае, когда $p\left(\mathbf{x},\ t\right)$ является дважды непрерывно дифференцируемой по $\mathbf{x}$ функцией, то справедливы следующие разложения:

$$p\left(\mathbf{y},\ \mathbf{x}\pm e_i,\ \ t\right)=p\left(\mathbf{y},\ \mathbf{x},\ \ t\right)\pm\varepsilon\tfrac{\partial p(\mathbf{y},\,\mathbf{x},\ t)}{\partial x_i}+\tfrac{\varepsilon^2}{2}\tfrac{\partial^2 p(\mathbf{y},\,\mathbf{x},\,t)}{\partial x_i^2}+o\left(\varepsilon^2\right),$$
$$p\left(\mathbf{y}\pm e_i,\ \mathbf{x},\ \ t\right)=p\left(\mathbf{x},\ \ t\right)\pm\varepsilon\tfrac{\partial p(\mathbf{y},\,\mathbf{x},\ t)}{\partial y_i}+\tfrac{\varepsilon^2}{2}\tfrac{\partial^2 p(\mathbf{y},\,\mathbf{x},\ t)}{\partial y_i^2}+o\left(\varepsilon^2\right),$$

$$p\left(\mathbf{y},\ \mathbf{x}+e_i\pm e_j,\ \ t\right)=p\left(\mathbf{y},\ \mathbf{x},\ \ t\right)+\varepsilon\left(\frac{\partial p\left(\mathbf{y},\ \mathbf{x},\ t\right)}{\partial x_i}\pm\frac{\partial p\left(\mathbf{y},\ \mathbf{x},\ \ t\right)}{\partial x_j}\right)+$$
$$+\frac{\varepsilon^2}{2}\left(\frac{\partial^2 p\left(\mathbf{y},\ \mathbf{x},\ \ t\right)}{\partial x_i^2}\pm 2\frac{\partial^2 p\left(\mathbf{y},\ \mathbf{x},\ \ t\right)}{\partial x_i\partial x_j}+\frac{\partial^2 p\left(\mathbf{y},\ \mathbf{x},\ \ t\right)}{\partial x_j^2}\right)+o\left(\varepsilon^2\right),i,j=\overline{0,\,n},$$
$$\min\left(y_i,x_i+\varepsilon\right)=\min\left(y_i,x_i\right)+\varepsilon u\left(y_i-x_i\right)+o\left(\varepsilon^2\right). \qquad (5)$$

Подставив в (4) данные разложения, получим следующее утверждение:

Теорема 1. Плотность распределения $p\left(\mathbf{x},\ t\right)$ вектора относительных переменных при $K\to\infty$ с точностью до $O\left(\varepsilon^2\right)$, где $\varepsilon=K^{-1}$, удовлетворяет уравнению Колмогорова–Фоккера–Планка

$$\frac{\partial p(\mathbf{y},\mathbf{x},t)}{\partial t}=-\sum_{i=1}^{n}\left(\frac{\partial}{\partial y_i}\left(A_i^{(1)}(\mathbf{y},\mathbf{x},t)p(\mathbf{y},\mathbf{x},t)\right)-\right.$$
$$\left.-\frac{\partial}{\partial x_i}\left(A_i^{(2)}(\mathbf{y},\mathbf{x},t)p(\mathbf{y},\mathbf{x},t)\right)\right)+$$

$$+\frac{\varepsilon}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial x_i \partial x_j}\left(B_{ij}^{(1)}(\mathbf{y},\mathbf{x},t)p(\mathbf{y},\mathbf{x},t)\right)+$$

$$+\frac{\varepsilon}{2}\sum_{i,j=1}^{n}(\frac{\partial^2}{\partial y_i \partial y_j}\left(B_{ij}^{(2)}(\mathbf{y},\mathbf{x},t)p(\mathbf{y},\mathbf{x},t)\right), \qquad (6)$$

где

$$A_i^{(1)}(\mathbf{y},\mathbf{x},t)=\gamma_i\left(l_i-y_i\right)u\left(y_i\right)-\alpha_i y_i, B_{ii}^{(1)}(\mathbf{y},\mathbf{x},t)=\gamma_i\left(l_i-y_i\right)u\left(y_i\right)+\alpha_i y_i,$$

$$A_i^{(2)}(\mathbf{y},\mathbf{x},t)=\left(\lambda_{0i}^{+}-\lambda_{0i}^{-}\right)\varepsilon+\sum_{j=1}^{n}\mu_j\min\left(y_j,x_j\right)p_{ji}^{*}\left(\mathbf{x}\right), \qquad (7)$$

$$B_{ii}^{(2)}(\mathbf{y},\ \mathbf{x},t)=\left(\lambda_{0i}^{+}+\lambda_{0i}^{-}\right)\varepsilon-\mu_j\min\left(y_j,x_j\right)p_{ii}^{**},$$

$$B_{ji}^{(2)}(\mathbf{y},\ \mathbf{x},t)=-\mu_j\min\left(l_j,x_j\right)p_{ji}^{**}\left(\mathbf{x}\right), B_{ij}^{(1)}(\mathbf{y},\mathbf{x},t)=0, j\neq i,$$

где

$$p_{ji}^{*}\left(\mathbf{x}\right)=\begin{cases}\left(p_{ji}^{+}-p_{ji}^{-}\right)u\left(x_i\right)-\left(p_{ij}^{+}+p_{ij}^{-}\right)\left(1-u\left(x_j\right)\right),\ j\neq i,\\ \left(p_{ii}^{+}-p_{ii}^{-}\right)u\left(x_i\right)-1-\left(p_{ij}^{+}+p_{ij}^{-}\right)\left(1-u\left(x_j\right)\right),\ i=j,\end{cases}$$

$$p_{ji}^{**}\left(\mathbf{x}\right)=\begin{cases}\left(p_{ji}^{-}-p_{ji}^{+}\right)u\left(x_i\right)-\left(p_{ij}^{+}+p_{ij}^{-}\right)\left(u\left(x_j\right)-1\right),\ j\neq i,\\ \left(p_{ii}^{-}-p_{ii}^{+}\right)u\left(x_i\right)+1-\left(p_{ij}^{+}+p_{ij}^{-}\right)\left(u\left(x_j\right)-1\right),\ i=j.\end{cases}$$

Известно, что характеристическая функция $\varphi(\lambda,t)=\int_{\mathrm{R}^{2n}}e^{I\lambda\mathbf{z}^{\mathrm{T}}}p(\mathbf{z},t)dz$, $I^2=-1$, также является законом распределения случайного процесса. Помножив обе части (4) на $e^{I\lambda\mathbf{z}^{\mathrm{T}}}$, проинтегрировав по z и учитывая определенные начальные и граничные условия, уравнение (7) сводится к уравнению для характеристической функции:

$$\frac{\partial\varphi(\lambda,t)}{\partial t}=\int_{\mathrm{R}^{2n}}\sum_{i=1}^{n}\Big[I(\lambda_i A_i^{(1)}(\mathbf{y},\mathbf{x},t)+\lambda_{n+i}A_i^{(2)}(\mathbf{y},\mathbf{x},t))-$$

$$-\frac{\varepsilon}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\lambda_i\lambda_j B_{ij}^{(1)}(\mathbf{y},\mathbf{x},t)+\lambda_{n+i}\lambda_{n+j}B_{ij}^{(2)}(\mathbf{y},\mathbf{x},t)\Big]p(\mathbf{z},t)e^{I\lambda\mathbf{z}^{\mathrm{T}}}dx. \qquad (8)$$

Используя известные свойства характеристической функции, в работе [8] были получены системы ДУ для определения начальных моментов первых двух порядков векторов $\mathbf{x}(t),\mathbf{y}(t)$ соответственно:

$$\frac{dn_i(t)}{dt}=MA_i^{(1)},\frac{d\bar{d}_i(t)}{dt}=MA_i^{(1)}(y(t),\ x(t)),i=\overline{1,n}. \qquad (9)$$

$$\frac{dw_{ik}^{(1,1)}(t)}{dt} = A_j^{(1)}\left(\nu_{kj}^{(1,1)}(t)\right) + A_k^{(1)}\left(\nu_{kj}^{(1,1)}(t)\right) + \varepsilon B_{ik}^{(1)}(w(t)),\ i,k=\overline{1,n},$$

$$\frac{d\nu_{ik}^{(1,1)}(t)}{dt} = A_j^{(2)}\left(\nu_{kj}^{(1,1)}(t)\right)+A_k^{(2)}\left(\nu_{kj}^{(1,1)}(t)\right)+\varepsilon B_{ik}^{(2)}(n(t)),\ i,k=\overline{n+1,2n},$$

Рассмотрим первые 2 уравнения системы (7) в явном виде:

$$\frac{dN_i(t)}{dt} = \left(\lambda_{0i}^{+} - \lambda_{0i}^{-}\right) + \sum_{j=1}^{n}\mu_j \min\left(\overline{d}_j(t), N_j(t)\right) p_{ji}^{*}\left(\mathbf{N}\right), \qquad (10)$$

$$\frac{d\overline{d}_i(t)}{dt} = \gamma_i\left(m_i - \overline{d}_i(t)\right)u(\overline{d}_i(t)) - \alpha_i\overline{d}_i(t), i=\overline{1,n},$$

В режиме насыщения, т.е. $\min\left(\overline{d}_j(t), N_j(t)\right) = \overline{d}_j(t), \overline{d}_j(t) > 0, \forall t$ решение последних уравнений имеет вид:

$$\overline{d}_i(t) = d_i(0)e^{-(\alpha_i+\gamma_i)t} + \frac{\gamma_i m_i}{\gamma_i+\alpha_i},$$

$$N_i(t) = \left(\lambda_{0i}^{+} - \lambda_{0i}^{-}\right)t + \sum_{j=1}^{n}\mu_j\int_0^t \overline{d}_j(\tau)d\tau(p_{ji}^{+} - p_{ji}^{-}) =$$

$$= N_i(t) = \left(\lambda_{0i}^{+} - \lambda_{0i}^{-}\right)t + \sum_{j=1}^{n}\mu_j\frac{\overline{d}_j(0)}{-(\alpha_i+\gamma_i)}e^{-(\alpha_i+\gamma_i)t} +$$

$$+\frac{\gamma_i m_i}{\gamma_i+\alpha_i}(p_{ji}^{+} - p_{ji}^{-})t + C.$$

Постоянная C находится из начальных условий. В случае функционирования сети в режиме низкой нагрузки, т.е. когда $\min\left(\overline{d}_j(t), N_j(t)\right) = N_j(t), \forall t$ система уравнений (10) является системой однородных уравнений с постоянными коэффициэнтами, решение которой также не составляет труда.

Решение системы (7) позволяет определить математические ожидания и дисперсии и взаимные корреляционные функции, которые в общем случае – детерминированные функции времени, определяющие соответственно среднюю траекторию процесса $\xi_i(t)$ и рассеяние вокруг нее, $i=\overline{1,n}$. Интервал $(\nu_i^{(1)}(t) \pm \sigma_i(t))$ – интервал, в который с вероятностью около 0,7 попадают реализации $\xi_i(t)$ в случае их нормального распределения, $i=\overline{1,n}$. Смешанные начальные моменты второго порядка $\nu_{ik}^{(1,1)}(t) = M\left(\xi_i(t)\xi_k(t)\right)$ называются взаимными корреляционными функциями или функциями кросс-корреляции $\xi_i(t)$ и $\xi_k(t)$ в момент времени t.

Заключение

В статье проведен асимптотический анализ G-сети с ненадёжными многолинейными СМО и найдены выражения для среднего числа заявок и исправных ЛО и их дисперсий. Так как с большой точностью плотность распределения вероятностей состояний удовлетворяет уравнению Колмогорова-Фоккера-Планка, то вероятности состояний имеют нормальное распределение, параметры которого найдены.

СПИСОК ЛИТЕРАТУРЫ

1. *Gelenbe E.* Product form queueing networks with negative and positive customers // Journal of Applied Probability. 1991. V. 28. № 3. P. 656–663.
2. *Matalytski M., Naumenko V.* Non-stationary analysis of queueing network with positive and negative messages // Journal of Applied Mathematics and Computational Mechanics. 2013. V. 12. № 2. P. 61–71.
3. *Matalytski M., Naumenko V.* Application of a z-transforms method for investigation of Markov G-networks // Journal of Applied Mathematics and Computational Mechanics. 2014. V. 13. № 1. P. 61–73.
4. *Fourneau J. M.* G-networks of unreliable nodes // Probability in the Engineering and Informational Sciences. 2016. V. 30. № 3. P. 361–378.
5. *Копать Д. Я.* Анализ ожидаемых доходов в G-сети с ненадежными многолинейными системами обслуживания // Вестник ГрГУ. Сер. 2. 2018. Т. 8. № 2. Р. 150–159.
6. *Медведев Г. А.* Замкнутые системы массового обслуживания и их оптимизация // Известия АН СССР. Техническая кибернетика. 1978. № 6. Р. 199–203.
7. *Копать Д. Я.* Асимптотический анализ G-сети с многолинейными системами обслуживания // Вестник ГрГУ. Сер. 2. 2021. Т. 11. № 3. Р. 104–113.
8. *Русилко Т. В.* Метод определения моментов первых двух порядков для вектора состояния сети массового обслуживания в асимптотическом случае // Вестник ГрГУ. Сер. 2. 2021. Т. 11. № 2. Р. 152–161.

Копать Дмитрий Ярославович — к.ф.-м.н., старший преподаватель кафедры логистики и методов управления. E-mail: *dk80395@mail.ru*

Методы анализа и визуализации данных

ПРОГНОЗИРОВАНИЕ ПЛАТЕЖЕСПОСОБНОСТИ КЛИЕНТОВ БАНКА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ НА ДАННЫХ, ОТОБРАННЫХ С ПОМОЩЬЮ РАСЧЕТА КОЭФФИЦИЕНТОВ WOE И IV

Д. Д. Бугакова, Е. Ю. Лисовская

Национальный исследовательский Томский государственный университет, г. Томск, Россия

В данной работе рассматриваются основные методы машинного обучения для прогнозирования платежеспособности клиентов банка и методы оценивания качества их работы на предоставленных данных. Для прогнозирования целевого признака в данной работе будут рассмотрены такие методы как: логистическая регрессия, случайный лес, метод ближайших соседей и метод опорных векторов. Для сравнения работы методов будут применены метрики accuracy, recall, precision, F_1, AUC ROC, AUC PR, индекс Джини. На основе значений метрик сделан вывод о том, что рассматриваемые методы примерно одинаково хорошо работают.

Ключевые слова: *Логистическая регрессия, случайный лес, метод опорных векторов, метод ближайших соседей, метрики качества.*

Введение

В наше время область машинного обучения набирает большую популярность в сфере бизнеса, финансов, сфере услуг, промышленности. В частности, в банковском деле для прогнозирования платежеспособности клиента. Своевременный анализ кредитного потенциала заемщика поможет предотвратить невозврат кредитных средств и избежать банкротства банковской организации.

1. Обзор исследуемых данных

Для того, чтобы на практике посмотреть работу алгоритмов воспользуемся набором данных с сайта kaggle.com [1] о клиентах банка «Тинькофф», для которого предлагается по данным из анкеты с использованием алгоритмов машинного обучения предсказать факт наличия дефолта. Набор данных содержит информацию о 205296 клиентах и 17 признаках, 5 количественных (возраст, скоринговый балл, количество обращений в банк, доход, количество отказанных заявок), 7 категориальных (количество связей с другими клиентами, уровень образования, пол, количество лет, которое заявитель является клиентом банка, тип жилплощади, должность, регион), 4 бинарных признаках (наличие автомобиля, иномарки, дохода выше среднего, загранпаспорта) и 1 целевого признака (наличие дефолта у клиента).

2. Отбор признаков

Перед обучением модели на данных предварительно их нужно обработать. Обработка данных состоит из двух этапов: первичная обработка данных и отбор информативных признаков. В данной работе отбор признаков производился с помощью расчета коэффициентов WoE (Weight of Evidence) для признаков с последующей оценкой предсказательной силы отобранных факторов с помощью расчета коэффициента IV (information value) [4]. Подробное описание предварительной обработки данных выходит за рамки данной работы. Результатом применения предложенных методов являются, следующие отобранные признаки: Жилплощадь: студия, Жилплощадь: дом, Должность: начальник, Балл: [-2.387; -2.116], Балл: [-2.116; -1.865], Балл: [-1.865; -1.566], Балл: больше -1.566, Связь с клиентами: 2, Связь с клиентами: 3, Связь с клиентами: более 3.

3. Подготовка данных для обучения моделей

Будем условно называть клиентов «плохими», если значение целевого признака – дефолт, иначе «хорошими». Для построения моделей сначала нужно разделить выборку на тренировочную, на которой модель будет обучаться и тестовую, на которой мы будем проверять качество моделей, в отношении 80/20 (Таблица 1).

Таблица 1

Распределение данных в выборках

Тип выборки	Количество «хороших» клиентов	Количество «плохих» клиентов	Доля «плохих» клиентов
Исходная	144631	18748	11,48 %
Тренировочная	115652	15051	11,52 %
Тестовая	28979	3697	11,31 %

4. Метрики для оценки качества модели

Работа модели может быть охарактеризована с помощью таких критериев качества как: ошибки первого и второго рода, *accuracy*, *recall*, *precision*, F_1, *AUC ROC*, *AUC PR*, индекс Джини. Перед началом рассмотрения метрик введем важное понятие матрицы ошибок (Рис. 1).

Рис. 1. Матрица ошибок

Метрика *accuracy* общая для всех классов и не применима в задачах с несбалансированной выборкой, как и в рассматриваемой задаче [3]

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}.$$

Для правильной оценки качества работы алгоритмов нужно использовать метрики *recall*, *precision* [3]:

$$precision = \frac{TP}{TP + FP}, \qquad recall = \frac{TP}{TP + FN}.$$

Recall, показывает какая доля объектов, положительного класса предсказала модель из всех объектов положительного класса. *Precision* показывает какая доля объектов, которую модель предсказала как положительную действительно является положительной.

Также при обучении модели существуют ошибки I-го и II-го рода *False Positive* и *False Negative*. В рассматриваемой задаче ошибку I-го рода можно интерпретировать как коммерческий риск, связанный с отказом кредитоспособным клиентам. Ошибка II-го рода характеризует кредитный риск, связанный с количеством некредитоспособных клиентов, классифицированных как кредитоспособных. Если *recall* и *precision* являются одинаково значимыми для задачи, используется F_1-мера (среднее гармоническое двух метрик *recall* и *precision*) [3]:

$$F_1 = \frac{2 * precision * recall}{precision + recall}.$$

ROC-кривая – график, показывающий зависимость между верно классифицируемыми объектами положительного класса (TPR) и ложно положительно классифицируемыми объектами негативного класса (FPR) [2]

$$TPR = \frac{TP}{TP + FN}, \qquad FPR = \frac{FP}{FP + TN}.$$

Метрика $ROC\,AUC$ (Area Under Curve) измеряет площадь под кривой ROC (Рис. 2), чем сильнее крутизна ROC-кривой, тем больше площадь под ней и тем лучше работает модель [2].

На основе метрики $ROC\,AUC$ можно вычислить другую метрику – индекс Джини

$$Gini = 2 * (ROC\,AUC - 0.5)\,,$$

чем выше индекс Джини, тем лучше дискриминирующая способность модели.

PR-кривая – график, построенный в координатах $recall$ и $precision$. Площадь под PR-кривой ($AUCPR$) лучше использовать для задач с несбалансированной выборкой (Рис. 3).

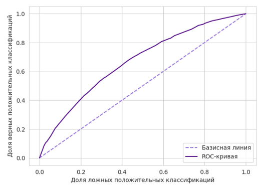

Рис. 2. График ROC-кривой

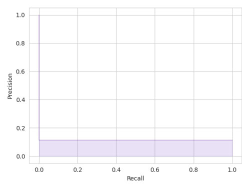

Рис. 3. График PR-кривой

5. Построение моделей

Построим базовые модели для всех алгоритмов. Для удобства обозначений пронумеруем модели. Модель 1 – Логистическая регрессия, Модель 2 – Метод ближайших соседей, Модель 3 – Случайный лес, Модель 4 – Метод опорных векторов. Модели были реализованы с помощью библиотек Python (LogisticRegression, KNeighborsClassifier, RandomForestClassifier, SVC). Результаты работы базовых моделей показывают, что модели не особо сильно отличаются друг от друга своей предсказательной способностью, поэтому для каждой модели нужно подобрать параметры, которые будут улучшать их (Таблица 2). Также для моделей 1, 3 и 4 был применен метод балансировки [3, 2]. Коэффициент регуляризации логистической регрессии получился слишком большой, модель могла переобучиться, поэтому нужно проверить с помощью кросс-валидации на 10 фолдах. После проведения кросс-валидации, выяснилось, что сильный коэффициент регуляризации почти никак не повлиял на предсказательную способность модели, значения метрик изменились совсем немного.

Таблица 2

Подобранные параметры

Модель	Гиперпараметр	Значение
Модель 1	Коэффициент регуляризации	0,0064281
Модель 2	Количество соседей	3
Модель 3	Число деревьев	267
	Число признаков	log2
	Глубина дерева	1200
	Число объектов в листьях	4
Модель 4	Коэффициент регуляризации	1
	Гамма	1
	Вид ядра	гауссово ядро

Подобранные гиперпараметры и применение метода балансировки к некоторым алгоритмам значительно улучшили предсказательную способность моделей (Рис. 4, Таблица 3).

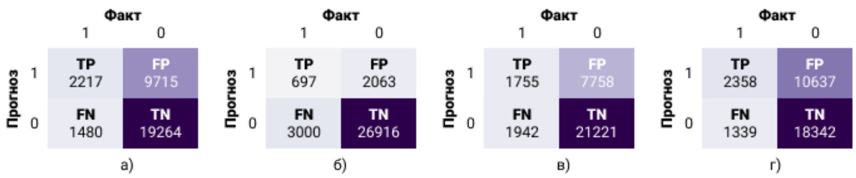

Рис. 4. Матрица ошибок моделей с подобранными гиперпараметрами: а) Модель 1, б) Модель 2, в) Модель 3, г) Модель 4

Таблица 3

Значения метрик базовых моделей после подбора гипперпараметров

Метрика	Модель 1	Модель 2	Модель 3	Модель 4
precision	0,168052	0,125000	0,174900	0,174773
recall	0,667027	0,000541	0,602380	0,602651
F_1	0,268467	0,001077	0,271089	0,270964
AUC PR	0,149768	0,113148	0,1503433	-

Заключение

Применение той или иной модели зависит от конкретной задачи. В задачах прогноза нелинейные модели, такие как случайный лес и метод опорных векторов показывают лучшие результаты.

Рассмотренные в работе алгоритмы показали хорошую предсказательную способность для использования в задачах прогнозирования платежеспособности клиентов банка. В данной работе по метрике F_1 лучше всего себя проявила модель случайного леса. Несмотря на достоинства метода ближайших соседей (простая реализация, хорошая интерпретация, настраивание гиперпараметра), он показал не очень хорошие результаты на данных, по сравнению с другими методами, вероятно из-за проблемы несбалансированности данных.

СПИСОК ЛИТЕРАТУРЫ

1. `https://www.kaggle.com/c/fintech-credit-scoring.` — Kaggle. 2021.
2. *Замятин А. В.* Интеллектуальный анализ данных. Томск: Издательский Дом Томского государственного университета, 2020. 119 с.
3. *Миркин Б. Г.* Введение в анализ данных: учебник и практикум. Москва: Издательство Юрайт университета, 2019. 174 с.
4. *Шунина Ю. С.* Критерии качества работы классификаторов. // Вестник Ульяновского государственного технического университета. 2015. № 2. С. 67–70.

Бугакова Дарья Дмитриевна — студент института прикладной математики и компьютерных наук. E-mail: *bugashka17@inbox.ru*

Лисовская Екатерина Юрьевна — к.ф.-м.н., доцент кафедры теории вероятностей и математической статистики института прикладной математики и компьютерных наук. E-mail: *ekaterina_lisovs@mail.ru*

ОСНОВНЫЕ ЭТАПЫ ОБРАБОТКИ И МЕТОДЫ ОТБОРА ПРИЗНАКОВ ДЛЯ ДАЛЬНЕЙШЕГО ПРОГНОЗИРОВАНИЯ ПЛАТЕЖЕСПОСОБНОСТИ КЛИЕНТОВ БАНКА

Д. Д. Бугакова[1], Е. Ю. Лисовская[1], Г. В. Баймеева[2]

[1]*Национальный исследовательский Томский государственный университет, г. Томск, Россия,*
[2]*X5 Retail Group, г. Москва, Россия*

В данной работе рассматриваются основные этапы обработки и методы отбора признаков для их дальнейшего использования в алгоритмах машинного обучения для построения моделей, которые предназначены для прогнозирования платежеспособности клиентов банка. В работе были рассмотрены такие способы отбора признаков как: расчет коэффициентов WoE (Weight of Evidence) для признаков с последующей оценкой предсказательной силы отобранных факторов с помощью расчета коэффициента IV (information value) и оценка важности признаков с помощью алгоритма случайного леса совместно с методом RFE (recursive feature elimination), основанного на логистической регрессии.
Ключевые слова: *Обработка данных, WoE (Weight of Evidence), IV (information value), квантование (биннинг), логистическая регрессия.*

Введение

Перед обучением модели на данных предварительно их нужно обработать. Обычно, обработка данных состоит из двух этапов: первичная обработка данных (удаление выбросов, удаление дубликатов из выборки, замена или удаление пропущенных значений), отбор информативных признаков (так как для обучения моделей необходимы не все признаки, а только те, которые в большей степени влияют на целевую признак).

1. Обзор исследуемых данных

Для того, чтобы на практике посмотреть работу алгоритмов воспользуемся набором данных с сайта kaggle.com о клиентах банка «Тинькофф» [1], для которого предлагается по данным из анкеты с использованием алгоритмов машинного обучения предсказать факт наличия дефолта.

Набор данных содержит информацию о 205296 клиентах и 17 признаках, 5 количественных (возраст, скоринговый балл, количество обращений в банк, доход, количество отказанных заявок), 7 категориальных (количество связей с другими клиентами, уровень образования, пол, количество лет, которое заявитель является клиентом банка, тип жилплощади, должность, регион), 4 бинарных признаках (наличие автомобиля, иномарки, дохода выше среднего, загранпаспорта) и 1 целевого признака (наличие дефолта у клиента).

2. Первичный анализ данных

После проведения первичного анализа данных был удален признак «даты подачи заявления», т.к. он не несет в себе никакой полезной информации для дальнейшей работы, но возможно этот признак можно было исследовать на сезонность и не удалять этот признак сразу, удалены дубликаты. Были обнаружены и обработаны пропуски во всем наборе данных. Для признака «количество отказанных заявок» была проведена проверка на информативность. Количественный признак будем считать неинформативным, если в нем большинство строк с одинаковыми значениями. Выбранный признак оказался на 82% неинформативным следовательно его можно удалить. В некоторых признаках были обнаружены выбросы на основе межквартильного размаха, значения, которые не попали в этот отрезок были удалены.

3. Отбор признаков

Для лучшего определения того, какие признаки влияют на вероятность дефолта клиента, нужно рассмотреть взаимоотношения между целевым признаком и остальными признаками. На основе графического представления зависимостей были выделены признаки, которые оказывают большее влияние на значение целевого признака (Рис. 1).

После первичного анализа данных, нужно осуществить отбор признаков, потому что для обучения алгоритмов нужны не все признаки, а только те, которые в большей степени влияют на итоговый результат.

Для отбора признаков будут использованы следующие методы:

1) WoE (Weight of Evidence) с последующей оценкой предсказательной силы отобранных факторов с помощью алгоритма IV (information value) [2];
2) Оценка важности признаков с помощью алгоритма случайного леса.

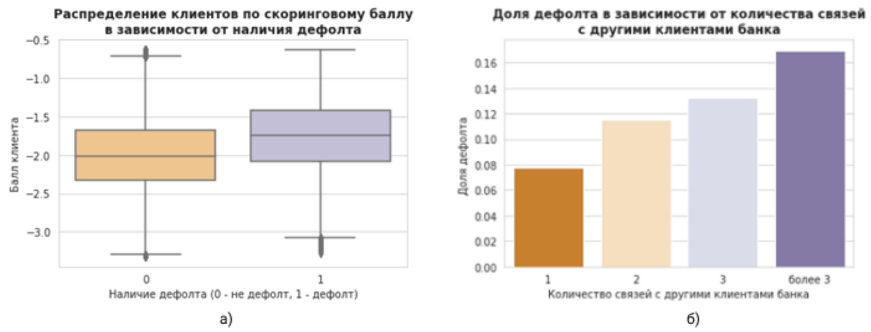

Рис. 1. а) График распределения клиентов по скоринговому баллу в зависимости от наличия дефолта, б) Доля дефолта в зависимости от количества связей с другими клиентами банка

Перед тем как начать работу по отбору признаков первым способом нужно все количественные непрерывные признаки преобразовать в категориальные, используя квантование (биннинг) [4]. Квантование – это процесс обработки данных, который позволяет разбить диапазон количественного признака на заданное количество интервалов (бинов) и присвоить каждому бину название. Квантование бывает двух видов:

1) *Интервальное.* Диапазон значений делится на одинаковые интервалы, в каждом не будет слишком много или мало данных.
2) *Квантильное.* Ширина интервалов будет различна, но в каждый попадет примерно одинаковое количество значений.

В настоящей работе применяется второй вид квантования. С помощью квантования будут преобразованы следующие признаки: возраст клиента, скоринговый балл и доход клиента.

4. Отбор признаков первым способом

Коэффициент WoE относительно данной задачи, характеризует степень отклонения уровня дефолтов в данной группе от среднего значения в выборке.

Для расчета WoE нужно для каждого категориального признака и для каждой группы внутри признака вычислить число клиентов с дефолтом («плохие» клиенты) и без дефолта («хорошие» клиенты) и рассчитать коэффициент по следующей формуле:

$$WoE_i = \ln\left(\frac{p_i}{q_i}\right),$$

где i – номер группы внутри признака, p – доля «хороших» клиентов среди всех «хороших», q – доля «плохих» клиентов среди всех «плохих».

После расчета WoE рассчитывается информационная ценность (коэффициент IV), который характеризует статистическую значимость признака, по следующей формуле:

$$IV = \sum_{i=1}^{n}(p_i - q_i) * WoE_i.$$

Для определения предсказательной силы признака на основе расчета IV воспользуемся следующей классификацией (Значение IV – Предсказательная сила):

— $< 0,02$ – отсутствует,
— $0,02 - 0,1$ – низкая,
— $0,1 - 0,3$ – средняя,
— $> 0,3$ – высокая.

Для признаков, которые были выбраны как наиболее значимые ниже представлена таблица 1 со значениями коэффициента IV. После отбора признаков созданы фиктивные признаки с помощью dummy-кодирования [3].

Пусть один из признаков x_j принимает m значений $\{b_1, \dots b_m\}$, тогда для каждого объекта x^j можно заменить признак x_i^j на $m-1$ признаков со значениями $\{0, 1\}$:

$$Z_i^{b_k} = I\left[x_i^j = b_k\right], \ k \in \{1, \dots, m-1\},$$

где $I\,[A]$ – индикатор события A.

Таблица 1

Значения коэффициента IV для наиболее значимых признаков

Признак	Значение IV
Количество связей с другими клиентами	0,13
Скорринговый балл	0,27
Тип жилплощади	0,11
Тип занимаемой должности	0,10
Число лет, которое заявитель является клиентом	0,10

Последним шагом в отборе признаков первым способом является удаление сильно коррелирующих между собой признаков (из двух признаков нужно оставить те, у которых значение коэффициента WoE больше), для этого были построены матрицы корреляций (Рисунок 2).

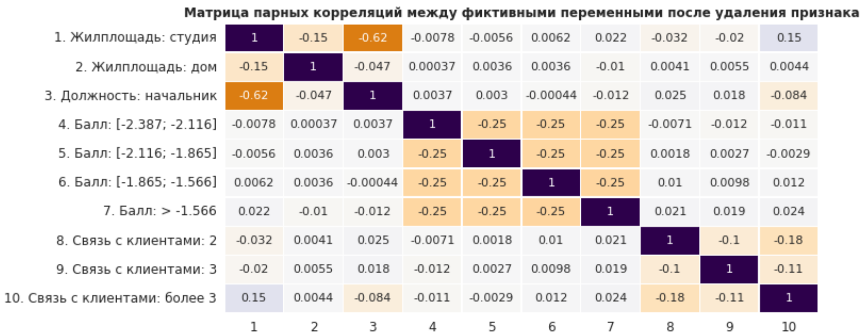

Рис. 2. Матрица корреляций между фиктивными признаками после удаление сильно коррелирующих признаков

Таким образом итоговый набор данных состоит из 10 отобранных первым способом признаков: Жилплощадь: студия; Жилплощадь: дом; Должность: начальник; Балл: [-2.387; -2.116]; Балл: [-2.116; -1.865]; Балл: [-1.865; -1.566]; Балл: больше -1.566; Связь с клиентами: 2; Связь с клиентами: 3; Связь с клиентами: более 3.

5. Отбор признаков вторым способом

Второй способ оценивает важность каждого признака на основе алгоритма случайного леса. Для этого нужно обучить модель на тренировочной выборке и посчитать out-of-bag ошибку для каждого объекта этой выборки. Ошибка усредняется для каждого элемента по всему случайному лесу. Значения каждого признака перемешиваются для всех объектов обучающей выборки и вычисления ошибки производятся заново, чтобы оценить важность признака. Чем больше уменьшается точность предсказаний из-за исключения или перестановки признака, тем важнее этот признак.

$$FI^{(t)}(x_j) = \frac{\sum_{i \in OOB^{(t)}} I\left(y_i = \widehat{y}_i^{(t)}\right)}{|OOB^{(t)}|} - \frac{\sum_{i \in OOB^{(t)}} I\left(y_i = \widehat{y}_{i,\pi_j}^{(t)}\right)}{|OOB^{(t)}|},$$

где $OOB^{(t)}$ – out-of-bag ошибка для дерева $t \in \{1, \ldots, N\}$, N – количество деревьев в случайном лесе, x_j – признак, для которого оценивается важность, $\widehat{y}_i^{(t)}$ – предсказание перед удалением или перестановкой признака, $\widehat{y}_{i,\pi_j}^{(t)}$ – предсказание после удаления или перестановки признака.

Далее производится расчет важности признака по всем деревьям в случайном лесе и может быть представлен в двух формах: ненормализованной и нормализованной

$$FI(x_j) = \frac{1}{N}\sum_{t=1}^{N} FI^{(t)}(x_j),\ z_j = \frac{N \cdot FI(x_j)}{\sigma},$$

где σ – стандартное отклонение разностей.

В таблице 2 приведены расчёты коэффициентов для пяти наиболее важных признаков. Для дальнейшего анализа было выбрано 24 признака для которых нормализованное значение больше 0,02.

На выбранных признаках был использован метод рекурсивного сокращения RFE (recursive feature elimination) в сочетании с логистической регрессией. В данной работе этот метод был использован как дополняющий, но также может быть использован, как и самостоятельный.

Таблица 2

Важность признаков

Признак	Ненормализованное значение	Нормализованное значение
Регион: МСК	227,8	0,043968
Жилплощадь: студия	207,4	0,040031
Пол: мужской	199,0	0,038410
Балл: $>$-1.566	189,4	0,036557
Балл: [-1.865; -1.566]	182,1	0,035148

Суть метода: модель обучается на исходном наборе признаков, оценивает их значимость и исключает наименее важный признак, процесс повторяется до тех пора пока не будет получено оптимальное или заданное количество признаков, каждому признаку присваивается ранг, чем выше ранг, тем важнее признак.

Таким образом итоговый набор данных состоит из 10 отобранных вторым способом признаков: Образование: ВШ; Жилплощадь: студия; Балл: [-2.116; -1.865]; Балл: [-1.865; -1.566]; Балл: $>$ -1.566; Связь с клиентами: более 3; Наличие автомобиля; Клиент банка: более 3; Регион: СПБ; Регион: МСК.

Заключение

После рассмотрения методов по отбору признаков было установлено, что оба метода отбирают примерно одинаковые признаки. Для оценки работы методов в будущем предлагается построить модели на данных отобранных двумя методами, сравнить их, используя метрики качества и сделать вывод о работе моделей на данных, отобранных разными методами. Также предлагается проверить качество моделей на отобранных данных после применения метода undersampling, который предназначен для балансировки данных.

СПИСОК ЛИТЕРАТУРЫ

1. `https://www.kaggle.com/c/fintech-credit-scoring`. — Kaggle. 2021.
2. *Шунина Ю. С.* Критерии качества работы классификаторов. // Вестник Ульяновского государственного технического университета. 2015. № 2. С. 67–70.
3. *Карминский А. М.* Сравнительный анализ методов прогнозирования банкротств российских строительных компаний. // Бизнес-информатика. 2019. № 3. С. 52–66.
4. *Шунина Ю. С.* Прогнозирование кредитоспособности клиентов на основе методов машинного обучения. // Финансы и кредит. 2015. № 27. С. 2–12.

Бугакова Дарья Дмитриевна — студент института прикладной математики и компьютерных наук. E-mail: *bugashka17@inbox.ru*

Лисовская Екатерина Юрьевна — к.ф.-м.н., доцент кафедры теории вероятностей и математической статистики института прикладной математики и компьютерных наук. E-mail: *ekaterina_lisovs@mail.ru*

Баймеева Галина Владимировна — старший менеджер по работе с большими данными. E-mail: *baymeevag@gmail.com*

Математическая теория телетрафика и теория массового обслуживания

DISCRETE TIME QUEUE WITH SELF INTERRUPTION RESULTING REDUCED PRIORITY

M. P. Anilkumar[1], K. P. Jose[2]

[1] *T. M. Govt. College, Tirur - 676502, Kerala, India*
[2] *Dept. of Mathematics, St.Peter's College, Kolenchery - 682 311, Kerala, India*

This paper considers queues with a discrete-time priority that are formed by customers' interruption in service. The interrupted customers are moved to a lower priority queue. Both preemptive and non-preemptive discipline to the service of lower priority customers are taken into account. We use Matrix-Analytic Method to analyze this model. The marginal distributions of the two queue lengths in each discipline are analyzed.
Keywords: *Peemptive, non-preemptive, Markovian arrival process, discrete phase-type distribution, matrix-analytic method.*

Introduction

White and Christie [1] introduced priority in queueing theory in 1958. Priority queues can be classified into two i) preemptive and ii) non-preemptive. In the case of preemptive priority, the service with the lower priority is interrupted when a customer with high priority arrives during the service, while in non-preemptive cases, the arriving customer with high priority is only served after the current service has been completed. An overview of priority queues is done by Brodal [2]. The author attempted to list some of the directions on priority queuing models that have gone over the past 50 years. The matrix geometric method for the priority queue with a discrete-time queue is discussed by Alfa [3], The author has extended the structure of the rate matrix R obtained by Miller [4] to the discrete case.

Customers with high priority or customers with low priority can generate from the queue during a service; these are referred to as self-generated priority or self-generated interruption. The self-generated priority queue with MAP arrivals and service time distribution as phase-type is described by Krishnamoorthy et al. [5], where the capacity of the priority customer intake is one and the remaining priority customers generated are considered lost. Interruptions in queues occur for many reasons, such as server failures, servers taking an emergency brake, and customers who have incomplete information or are distracted. Jacob et al. [6] investigated an infinite capacity queuing system with a single server to which customers arrive according to a Poisson process and the service time follows an exponential distribution. The customer interruption takes place according to a Poisson process and the interruption duration follows an exponential distribution. The self-interrupting customers will enter a finite buffer size and any interrupted customer who finds the buffer full is considered lost. The concept of the self-interrupting infinite buffer for lower priority in continuous time was analyzed by Krishnamoorthy and Manjunath [7]. Anillumar and Jose [8] generalized this model to discrete-time cases with MAP arrivals and phase-type service processes. A discrete-time priority queueing inventory model with customer-induced interruption was also analyzed by Anilkumar and Jose [9]. In this paper, we generalize the service process and interruptions.

1. Modeling

We consider an infinite capacity single server discrete-time queue in which arrival of customers are modeled by Markovian arrival process having n phases with representation (D_0, D_1). Then the arrival rate $\lambda = \psi D_1$, where ψ is the stationary probability vector of $D_0 + D_1$. There are two types of priority queues $\mathcal{P}_1$ and $\mathcal{P}_2$. Customers who arrive in the system first enter the high priority queue. During his service, he may or may not interrupt the service. The primary service is considered to be the time till absorption of a discrete-time Markov chain that has two absorbing states and which is represented by the transition probability matrix

$$T = \begin{bmatrix} S(1) & S^0(1) & S^{02}(1) \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

with initial probability $\beta(1)$ and transient matrix S. If a customer interrupts the service, he is transferred to a low priority queue $\mathcal{P}_2$. The same server serves customers in these two queues one at a time according to their priority. Once an interrupted customer in P_2 receives service, he will have no further interruption during service (except when $\mathcal{P}_1$ customers arrive in preemptive case). After completing the service, customers in both queues leave the system. A customer in $\mathcal{P}_2$ is taken for service only when no customer is present in $\mathcal{P}_1$ and no customer in $\mathcal{P}_2$ ahead of him. The arrival of customers in $\mathcal{P}_1$ during the service of $\mathcal{P}_2$, may or may not affect the service. In other words, service $\mathcal{P}_2$ is either according to preemptive or non-preemptive discipline. The processing times of customers in $\mathcal{P}_2$ are discrete phase-type distributed random variables with parameters $(\beta(2), S(2))$ and dimension m_2 each.

Notations

1) $N_1(n)$: Number of $\mathcal{P}_1$ customers in the sustem at an epoch n.
2) $N_2(n)$: Number of $\mathcal{P}_2$ customers waiting for service an epoch n
3) I(n):The arriving phase of a customer
4) J(n): The service phase
5) $\bar{a} = 1 - a$ where $0 \leqslant a \leqslant 1$
6) $\mathbf{e}$: Column vector of 1's of appropriate order
7) $S^0(2) + S(2)\mathbf{e} = \mathbf{e}$

2. Preemptive Priority

We assume that the priority of service in $\mathcal{P}_2$ is in the preemptive discipline. The arrival of the customer during the service of the non-priority queue affects the arrival of customers in $\mathcal{P}_1$. Now $\{(N_1(n), N_2(n), I(n), J(n)), n = 1, 2, 3, ...\}$ is a Level Independent Quasi-Birth Death process (LIQBD) on the state space $\Delta_1 \cup \Delta_2 \cup \Delta_3$, where

$$\begin{aligned}
&\Delta_1 = \{(0, j); 1 \leqslant j \leqslant n\},\\
&\Delta_2 = \{(0, i_2, j, k_2); i_2 \geqslant 0, 1 \leqslant j \leqslant n, 1 \leqslant k_2 \leqslant m_2\},\\
&\Delta_3 = \{(i_1, i_2, j, k_1); i_1 \geqslant 1, i_2 \geqslant 0, 1 \leqslant j \leqslant n, 1 \leqslant k_1 \leqslant m_1\}.
\end{aligned}$$

The transition probability matrix P of this process is given by

$$P = \begin{bmatrix} B_{00} & B_{01} & 0 & & \\ B_{10} & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{bmatrix},$$

where the elements of P are square matrices of order S and are given by

$$B_{00} = \begin{bmatrix} B_{00}^{00} & & & \\ B_{00}^{10} & B_{00}^{1} & & \\ & B_{00}^{2} & B_{00}^{1} & \\ & & \ddots & \ddots \end{bmatrix}, \; B_{01} = \begin{bmatrix} B_{01}^{00} & & & \\ B_{01}^{2} & B_{01}^{1} & & \\ & B_{01}^{2} & B_{01}^{1} & \\ & & \ddots & \ddots \end{bmatrix},$$

$$B_{10} = \begin{bmatrix} B_{10}^{00} & B_{10}^{0} & & & \\ & B_{10}^{1} & B_{10}^{0} & & \\ & & B_{10}^{1} & B_{10}^{0} & \\ & & & \ddots & \ddots \end{bmatrix}, \; A_2 = \begin{bmatrix} A_2^{1} & A_2^{0} & & & \\ & A_2^{1} & A_2^{0} & & \\ & & A_2^{1} & A_2^{0} & \\ & & & \ddots & \ddots \end{bmatrix},$$

$$A_1 = \begin{bmatrix} A_1^1 & A_1^0 & & & \\ & A_1^1 & A_1^0 & & \\ & & A_1^1 & A_1^0 & \\ & & & \ddots & \ddots \end{bmatrix},\ A_0 = \begin{bmatrix} A_0^1 & & & \\ & A_0^1 & & \\ & & A_0^1 & \\ & & & \ddots \end{bmatrix},$$

$B_{00}^{00} = D_0, B_{00}^{10} = D_0 \otimes S^0(2), B_{00}^1 = D_0 \otimes S(2), B_{00}^2 = D_0 \otimes S^0(2)\beta(2),$

$B_{01}^{00} = D_1 \otimes \beta(1), B_{01}^2 = D_1 \otimes S^0(2)\beta(1), B_{01}^1 = D_1 \otimes S(2)\mathbf{e}\beta(1),$

$B_{10}^{00} = D_0 \otimes S^0(1), B_{10}^0 = D_0 \otimes S^{02}(1)\beta(2), B_{10}^1 = D_0 \otimes S^0(1)\beta(2),$

$A_1^1 = (D_0 \otimes S(1) + D_1 \otimes S^0(1)\beta(1)), A_1^0 = D_1 \otimes S^{02}(1)\beta(1), A_0^1 = D_1 \otimes S(1),$

$A_2^1 = D_0 \otimes S^0(1)\beta(1), A_2^0 = D_0 \otimes S^{02}(1)\beta(1).$

Theorem 1. The system is stable, if and only if,

$$\lambda\beta(1)(I - S(1))^{-1}\boldsymbol{e} + \lambda\beta(1)(I - S(1))^{-1}\boldsymbol{e}\frac{\beta(2)\left(I - S(2)\right)^{-1}\boldsymbol{e}}{\alpha(I - S(1))^{-1}S^{02}(1)} < 1. \quad (1)$$

The proof can follow using intuitive argument.

Steady State Analysis

Since the matrix P has the structure of quasi-birth and death process and the individual phases are infinite, one can use the generalization of matrix geometric method described by Neuts [10] to the case of infinite sub matrix suggested by Miller [4]. For this, first we find the minimal non negative solution R of the matrix quadratic equation,

$$R^2A_2 + RA_1 + A_0 = R,$$

whose spectral radius is less than one. Since, A_0, A_1 and A_2 has upper triangular structure, we can see that the rate matrix R also has the upper triangular structure given by,

$$R = \begin{bmatrix} R_0 & R_1 & R_2 & \dots \\ & R_0 & R_1 & \dots \\ & & R_0 & \dots \\ & & & \ddots \end{bmatrix}.$$

Here, R_0 will be the minimal non negative solution of $R_0 = (A_0^1 + R_0^2)(I - A_1^1)^{-1}$, which can be calculated using iterative method. Then successively find R_1, R_2, $\dots$ recursively calculated.

Steady State Probability Vector

Let $\mathbf{x} = (x_0, x_1, x_2, \dots)$ be the steady state probability vector of P. That is $\mathbf{x}P = \mathbf{x}$ and $\mathbf{xe} = 1$, where $\mathbf{e}$ is the infinite column vector of 1's Then $x_{i+1} = x_i R$ for $i \geqslant 1$.
To find the boundary probability vectors(x_0, x_1), consider the following system of equations

$$x_0 B_{00} + x_1 B_{10} = x_0,$$
$$x_0 B_{01} + x_1(A_1 + RA_2) = x_1.$$

From the second equation, we have

$$x_1 = x_0 B_{01}(I - A1 - RA_2)^{-1}.$$

Substituting x_1 in the normalizing condition, $x_0\mathbf{e} + x_1(I - R)^{-1}\mathbf{e} = 1$, one can solve for x_0 and x_1.

Marginal Probability Distributions

Let $P_i(v)$ be the probability that there are i jobs of in the queue $\mathcal{P}_v$ for $v = 1, 2, i \geqslant 0$

$$\begin{aligned} p_i(1) &= x_i \boldsymbol{e}, \\ p_i(2) &= x_{o(i-1)}\boldsymbol{e}(nm_2) + x_1(I - R)^{-1}\beta_i, \end{aligned} \tag{2}$$

where β_i is an infinite column matrx whose $(inm_1 + 1)^{th}$ to $(i + 1)nm_1^{th}$ entries are one and remainig values are zeros.

3. Non-preemptive Self Generated Interruption

The transition probability matrix has the same structure as discussed earlier but some difference in the finite blocks. There is no change in the stability condition. The rate matrix R possesses upper triangular structure as in the preemptive case and the only difference is that, each R_i is block lower triangular matrix having order $n(m_1 + m_2)$ of the form

$$R_i = \begin{bmatrix} R_{11}^i & \mathbf{0} \\ R_{21}^i & R_{22}^i \end{bmatrix}.$$

4. Conclusion

This paper looked at discrete-time, self-interrupting priority queues of MAP arrivals and service as the time till absorption of an absorbing Markov chain which has two absorbing states through which a customer leaves the system after service or the interrupted customer moved to lower priority queue. Matrix-Analytic Method is used to analyze the model. The marginal probability distribution of queue length is discussed. For future studies, one can consider the chance of abandoning the service if interruption happened. Incorporation of inventory of items is also another interest of study.

REFERENCES

1. *White H., Christie L.* Queuing with preemptive priorities or with breakdown // Operations research. 1958. Vol. 6. № 1. P. 79–95.
2. *Brodal G. S.* A survey on priority queues // Space-Efficient Data Structures, Streams, and Algorithms, Springer, Berlin, Heidelberg, 2013. P. 150–163.
3. *Alfa A. S.* Matrix-geometric solution of discrete time MAP/PH/1 priority queue // Naval Research Logistics (NRL). 1998. Vol. 45. № 1. P. 23–50.
4. *Miller D. R.* Computation of steady-state probabilities for M/M/1 priority queues // Operations Research. 2004. Vol. 29. № 5. P. 945–958.
5. *Krishnamoorthy A., Babu S., Narayanan V. C.* The MAP/(PH/PH)/1 queue with self-generation of priorities and non-preemptive service // European Journal of Operational Research. 2009. Vol. 195. № 1. P. 174–185.
6. *Jacob V., Chakravarthy S. R., Krishnamoorthy A.* On a customer-induced interruption in a service system // Stochastic Analysis and Applications. 2012. Vol. 30. № 6. P. 949–962.
7. *Krishnamoorthy A., Manjunath A. S.* On priority queues generated through customer induced service interruption // Neural, Parallel and Scientific Computations. 2015. Vol. 23. P. 459–486.
8. *Anilkumar M. P., Jose K. P.* Discrete time priority queue with self generated interruption // AIP Conference Proceedings. 2020. Vol. 2263. № 1. P. 030031.
9. *Anilkumar M. P., Jose K. P.* A Geo/Geo/1 Inventory Priority Queue with Self Induced Interruption // International Journal of Applied and Computational Mathematics. 2020. Vol. 6. № 4. P. 1–14.
10. *Neuts M. F.* Matrix-geometric solutions in stochastic models: an algorithmic approach. Courier Corporation, 1994. 340 p.

Anilkumar M. P. — Assistant Professor, Dept. of Mathematics, T. M. Govt. College, Tirur-676502, Kerala, India. E-mail: *anilkumarmp77@gmail.com*

Jose K. P. — Associate Professor, Dept.of Mathematics, St. Peter's College, Kolenchery-682311, Kerala, India. E-mail: *kpjspc@gmail.com*

MODELING TWO-WAY COMMUNICATION SYSTEMS WITH CATASTROPHIC BREAKDOWNS

Attila Kuki, Tamás Bérczes, János Sztrik

Faculty of Informatics University of Debrecen, Hungary

A two-way communication system is modeled in this paper. A retrial queueing system with a finite and an infinite sources is used in the model. Requests from the finite source are the first order or regular customers, while the requests from the infinite source are the second order or the invited customers. In case of an idle server, the second order customers are called for service. The non-reliable server is subject to random breakdowns, and even to catastrophic breakdown, i.e all of the requests at the server, in the buffer, and in the orbit are sent back to the sources. The novelty of this paper is to investigate the effect of catastrophic breakdown in a two-way communication environment. The goal is to determine the steady-state probabilities and the system characteristics with the help of a software package. Figures illustrate the effect of the system parameters on the performance measures.

Keywords: *Two-way communication, catastrophic breakdown.*

Introduction

For modeling various fields of infocommunication and computer sciences, the retrial queueing systems are one of the well-known and widely spread models. Results can be found in works of Falin, Templeton, and more authors [5, 9]. Several models assume, that the customers generate their request from a finite number of population. In some cases, these finite source models are more realistic and give a better description of the considered application [1, 4]. In addition, the considered real-life systems are unfortunately non-reliable, that is the server can lose its efficiency or may break down. These type of a non-reliable systems were investigated e.g. in [10].

Furthermore, a new general model was developed not to lose the requests, who are not able to wait for the server in the queue or in the orbit. They can register for the service and later the idle system can call (outgoing call) for these customers (two-way communication systems). The first and later results can be found e.g. in [2, 3, 7].

The research work was supported by the Austro-Hungarian Cooperation Grant No 106öu4, 2020.

In this paper a special two-way communication system is considered. It is called searching for customers. Two types of customers are considered. The organization has a finite number of regular or goodwill customers. They are the first order customers, making primary calls towards the organization, i.e. the server. These clients are served according to the retrial queueing model. During the busy period of the server, the incoming customers are sent back to the orbit, where they can retry their requests. These customers are represented with a finite number of the source. During the idle periods of the server, outgoing calls are performed towards the customers in the second, infinite source. The clients in this infinite source (second order customers) will contact the organization with some special interest. In case of a busy server (meanwhile another regular customer might be arrived), this special second order customer is treated as a non-preemptive priority client. They are sent into a priority buffer.

The non-reliable server is subject to random breakdowns. The focus of this paper is a special type of breakdowns, the catastrophic breakdown. Retrial queueing models in which customers are removed from the system due to catastrophic or disaster events have been studied extensively in the literature. Modeling special systems, e.g. automatic teller machines needs different types of breakdowns. A catastrophic event can be, for example, mechanical failures or power outages. Disaster events are known also as an arrival of a negative customer. When a negative customer arrives at the system, it immediately removes the positive customer from service. The case, when a negative customer removes all the positive customers from the system at once, is called a disaster. Disaster events not only break the service of the current customer but break down the server. The customers from the server, the orbit, and the priority buffer are sent back to the source. Detailed studies on negative customers can be found in [6, 8], and reference therein.

The remaining parts of this work contain the following. In Section 1 the model definition, the underlying Markovian process with 2 dimensions, and the applied parameters are described. In Section 2 the steady-state probabilities are considered, and some performance measures (utilization, response times, etc.) are provided with the help of MOSEL-2 tool. At the end of the paper, the results are summarized in a Conclusion.

1. Description of the model

The system is modelled by a single-server retrial queueing system with a finite and an infinite sources. The functionality of the model is displayed in Figure 1.

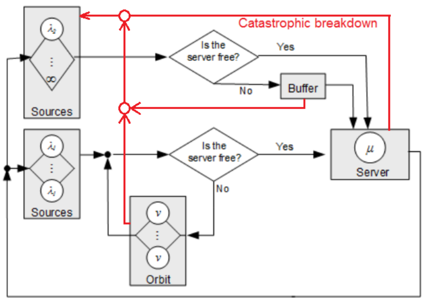

Figure 1. The system model

The system has two sources. The first one is finite with the first order customers, the number of customers is N. These customers generate a job towards the server using the exponential law with parameter λ_1. For the first order customers there is no queue at the server. After the service, the job goes back to the source can generate a new request again. The service time is also exponential distributed with parameter μ_1. When the server is busy, the incoming job is transferred to the orbit. The size of the orbit is N. From the orbit the jobs after an exponential random time interval with parameter ν retry their request to the server until they are served. The model has second order customers in an infinite number of sources. These customers generate triggered requests only. The idle server makes outgoing calls towards this infinite source, and the second order customers generate a request to be served. The generation is also exponential with parameter λ_2. The distribution of service times is also exponential with parameter μ_2. When a second order customer finds the server busy, the job is transferred to a priority buffer. In case of an idle server, a second order customer is called from this buffer. The size of the buffer is one, because in case of an idle server, there is no outgoing call when a customer is in the buffer.

In this model the single server is unreliable, it may subject to breakdown. Here the catastrophic break-down is considered. This is the situation when a disaster event removes all of the customers from the system (from the orbit, from the buffer, and from the server after interrupting the service). The repair of the system starts immediately. The breakdown parameter is γ_0 and γ_1 for idle and busy servers, respectively. γ_2 is the parameter of the repair. The considered times are exponentially distributed. During the breakdown period, the sources are blocked, they are not able to generate requests.

Let us denote $O(t)$ and $S(t)$ the number of requests in the orbit and the state of the server at a given time point of t:

$$S(t) = \begin{cases} 0, & \text{when the server is idle,} \\ 1, & \text{when the server is busy} \\ & \text{with a first order customer,} \\ 2, & \text{when the server is busy} \\ & \text{with a second order customer,} \\ 3, & \text{when the server is down.} \end{cases}$$

The state-space of the underlying Markovian-process $(S(t), O(t))$ can be described as a set of $\{0, 1, 2, 3\} \times \{0, 1, 2, ..., N\}$ elements. Although the system has an infinite source, the maximum number of the customers in the system is $(N + 1)$ (N in the orbit and one second order customer under service), there is no stability problems regarding the system. The state space is finite.

For buffered and non-buffered models the system balance equations can be formulated. The steady-state system probabilities are:

$$p_{i,j} = \lim_{t\to\infty} P(S(t) = i, O(t) = j), i = 0, 1, 2, 3 \text{ and } j = 0, 1, ..N$$

Solving manually these balance equations is rather difficult. There exist several effective tools for performing the background calculations. In this paper, the MOSEL-2 tool was used. When the steady-state probabilities are calculated, this tool provides the well-known performance characteristics. These measures are obtained using the following formulas, e.g. average number of jobs in the orbit

$$\overline{O} = \sum_{s=0}^{3} \sum_{o=0}^{N} oP(s, o).$$

2. Numerical results

The most important goal of these types of stochastic systems is to obtain the performance measures and system characteristics. Usually the throughput, utilization, response times, waiting times, queue length are considered. Here the utilization and response time are focused.

There exist several methods to calculate the system measures. Solving directly the balance equations is rather difficult in most cases. Effective software tools can be used to get the steady-state system probabilities. From these probabilities the performance measures can be computed directly or by the help of the considered tool. Because solving directly the balance equations is rather difficult, the MOSEL-2 tool is used. The system equations are solved by the SPNP (Stochastic Petri Net Program). The following figures illustrate the most interesting numerical results.

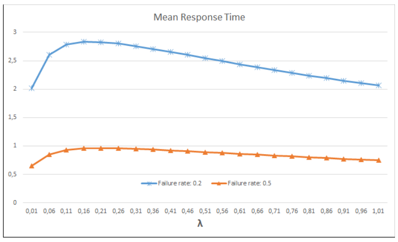

Figure 2. Mean Response Time vs. λ_1

In Figure 2 and 3 the running parameter is $\lambda = \lambda_1$, the first order generation rate. In Figure 2 the mean response time can be seen for two different failure rates ($\gamma_0 = \gamma_1$ in this figure). The catastrophic breakdown is applied. For a higher failure rate lower response time can be observed because the jobs are more often kicked off to the source.

Figure 3 displays the server utilization. $\mu_1 = 4$ and $\mu_2 = 2$. This is the reason, that the utilization is higher in the catastrophic case than in the normal breakdown case.

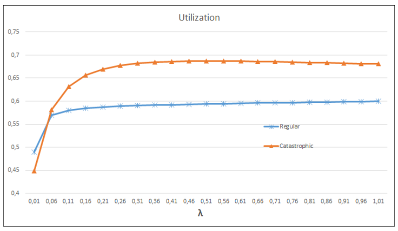

Figure 3. Server utilization vs. λ_1

3. Conclusion

A special two-way communication system was investigated here. First order customers come from a finite source, while in the case of an idle server, second order customers can reach the system via a direct call. Different cases can be considered. Failure rates are set to be equal for idle server, for server with first order customer, and for server with second order customer. Two different cases were considered. The system subject to general breakdown and catastrophic breakdown. In this short abstract only two measures, the mean response time and the server utilization were investigated. Of course, more characteristics with more parameters can also be considered.

REFERENCES

1. *Artalejo J.* Retrial queues with a finite number of sources // Journal of the Korean Mathematical Society. 1998. Vol. 35. P. 503–525.
2. *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. ACM: 2016. Article No: 11, P. 1–7.
3. *Falin G.* Model of coupled switching in presence of recurrent calls // Engineering Cybernetics. 1979. Vol. 17. № 1. P. 53–59.
4. *Falin G., Artalejo J.* A finite source retrial queue // European Journal of Operational Research. 1989. Vol. 108. P. 409–424.
5. *Falin G., Templeton J.* Retrial Queues. London: Chapman and Hall, 1997.
6. *Gupta U., Kumar N., Barbhuiya F.* A queueing system with batch renewal input and negative arrivals // Applied Probability and Stochastic Processes. Springer: 2020. P. 143–157.
7. *Nazarov A. A., Paul S., Gudkova I., et al.* Asymptotic analysis of Markovian retrial queue with two-way communication under low rate of retrials condition // Proceedings 31st European Conference on Modelling and Simulation. 2017.
8. *Subramanian S., et al.* A stochastic model for automated teller machines subject to catastrophic failures and repairs // Queueing Models and Service Management. 2018. Vol. 1. № 1. P. 75–94.
9. *Templeton J.* Retrial queues // Top. 1999. Vol. 7. № 2. P. 351–353.
10. *Wang J., Zhao L., Zhang F.* Performance analysis of the finite source retrial queue with server breakdowns and repairs // Proceedings of the 5th International Conference on Queueing Theory and Network Applications. ACM: 2010. P. 169–176.

Attila Kuki — Assistant Professor, Department of Informatics Systems and Networks. E-mail: *kuki.attila@inf.unideb.hu*

Tamas Berczes — Associate Professor, Department of Informatics Systems and Networks. E-mail: *berczes.tamas@inf.unideb.hu*

Janos Sztrik — Full Professor, Department of Informatics Systems and Networks. E-mail: *sztrik.janos@inf.unideb.hu*

LIMIT BEHAVIOR AND STABILITY OF APPLIED PROBABILITY SYSTEMS

E. Bulinskaya

Lomonosov Moscow State University, Moscow, Russian Federation

The models arising in such applications as insurance, inventory, telecommunications, finance, population dynamics and others are of input-output type. Hence, another interpretation of these elements lets pass from one applied domain to another. Discrete-time models became popular during the last decade because in some situations they give a more appropriate description of system functioning. Therefore we consider two models (discrete- and continuous-time ones) in the framework of reliability and cost approaches, study their limit behavior, stability and reliability.
Keywords: *Limit behavior, stability, optimization, reliability.*

Introduction

It is well known that one has to choose an appropriate mathematical model in order to study a real process or system. The most frequently used models in such applied probability domains as insurance, inventory, telecommunications, finance, population dynamics and many others are described by the six-tuple $(T, Z, Y, U, \Psi, \mathcal{L})$ (see, [3]). Here T is the planning horizon, $Z = (Z(t), t \in [0, T])$ and $Y = (Y(t), t \in [0, T])$ are input and output processes, respectively. A system may be deterministic, stochastic or mixed. That means, either both processes are deterministic (stochastic) or one of the processes is stochastic and the other one is deterministic. Giving different interpretation to input and output we can study the systems from different applied domains using the same model. The system state $X = \Psi(Z, Y, U)$ depends on input, output and control U via functional Ψ describing the system structure and performance mode. The control may be applied to input Z, output Y or system structure Ψ. According to the choice of the objective function $\mathcal{L}$, there exist two approaches (cost and reliability ones), see, [4]. In the first case we calculate the (expected) costs entailed by system functioning and want to minimize them. In the second one the aim is to minimize the probability of system failure (ruin) or maximize the time of uninterrupted system performance.

Supported by the Russian Foundation for Basic Research project № 20-01-00487

Below we consider two models in order to take into account different situations. Since insurance is the oldest applied probability domain we describe the models in terms of insurance company functioning. Thus, Section 1 deals with a mixed-type model, namely, discrete-time model of insurance company using non-proportional reinsurance and bank loans. The premium inflow is deterministic, whereas the claim amounts are stochastic. First of all, we prove that the optimal loan policy is characterized by a sequence of critical levels. After that we establish the system stability using the Kantorovich metric, then prove the strong law of large numbers (SLLN) and central limit theorem (CLT) for the company surplus.

In Section 2 a generalization of the classical Cramér-Lundberg model is considered. The company is supposed to issue several types (say, n) of policies. Premiums, as well as claims, are random. Their flows are described by generalized Poisson processes. The reliability approach is employed. More precisely, the ruin probability is calculated using the martingale technique. Analog of the Lundberg inequality is also obtained. Some special types of claims and premiums distributions are treated.

Due to lack of space the proofs are omitted.

1. Non-proportional reinsurance and bank loans

Suppose that the claims arriving to insurance company are described by a sequence of independent identically distributed (i.i.d.) non-negative random variables (r.v.'s) $\{X_i, i \geqslant 1\}$. Here X_i is the claim amount during the i-th period (year, month or day). Let $F(x)$ be its distribution function (d.f.) having density $\varphi(x)$ and finite expectation. The company uses non-proportional reinsurance with retention a and bank loans. If a loan is taken at the beginning of period (before the claim arrival) the rate is b_1 whereas the emergency loan after the claim arrival is taken at the rate b_2 with $b_2 > b_1$. Our aim is to choose the loans in such a way that the additional payments entailed by loans are minimized. Denote by M the premium acquired by direct insurer (after reinsurance) during each period. Clearly,

$$M = (1+\beta_1)EX - (1+\beta_2)E(X-a)^+.$$

Here β_1 and β_2 are the safety loadings of insurer and reinsurer respectively, usually, $\beta_1 < \beta_2$.

If x is the initial capital and y is the capital after the bank loan, then $f_1(x)$, the minimal expected additional cost during one period, is given by

$$f_1(x) = \min_{y \geqslant x}[b_1(y-x) + b_2 E(\min(X,a) - (y+M))^+]. \tag{1}$$

Clearly, (1) can be rewritten in the form:

$$f_1(x) = -b_1x + \min_{y \geqslant x} G_1(y),\ G_1(y) = b_1y + b_2E[\min(X, a) - (y + M)]^+.$$

Now let $f_n(x)$ be the minimal expected costs during n periods and α the discount factor for future costs. Then, using the dynamic programming, see, e.g., [1], one easily obtains the following relation:

$$f_n(x) = -b_1x + \min_{y \geqslant x} G_n(y),\ G_n(y) = G_1(y) + \alpha E f_{n-1}(y + M - \min(X, a)).$$

It is not difficult to prove by induction the main optimization result.

Theorem 1. Let $F(M) < 1 - \frac{b_1}{b_2} < F(a)$, then there exists an increasing sequence of critical levels $\{y_n\}_{n \geqslant 1}$ such that

$$f_n(x) = -b_1x + \begin{cases} G_n(y_n), & \text{if} \quad x \leqslant y_n, \\ G_n(x), & \text{if} \quad x > y_n. \end{cases} \tag{2}$$

The sequence is bounded by $\bar{y}$ satisfying the equation $H(y) = 0$ where $H(y) = G_1'(y) - b_1\alpha$.

Corollary 1. If $M = 0$ then $\lim_{n\to\infty} y_n = \bar{y}$.

For $M > 0$ such statement is not valid. To verify this fact the numerical analysis was carried out using Python.

Suppose now that we have two different sequences of claims $\{X_n\}$ and $\{Y_n\}$ with distribution functions F_X and F_Y respectively.

The corresponding minimal n-step costs will be denoted by $f_{n,X}$ and $f_{n,Y}$, other functions and constants depending on distribution will be marked by subscripts X and Y, as well.

The distance between distributions will be measured in terms of the Kantorovich metric.

Definition 1. Let random variables X and Y, defined on the same probability space, possess finite expectations. The distance based on the Kantorovich metric is given as follows:

$$\varkappa(X, Y) = \int_{-\infty}^{\infty} |F_X(t) - F_Y(t)|\, dt$$

where F_X and F_Y are the respective distribution functions of X and Y.

The distance between the cost functions is measured in terms of the Kolmogorov uniform metric. Thus, we are going to study

$$\Delta_n = \sup_x |f_{n,X}(x) - f_{n,Y}(x)|.$$

To this end we need the following

Lemma 1. Let functions $g_i(y)$, $i = 1, 2$, be such that $|g_1(y) - g_2(y)| < \delta$ for some $\delta > 0$ and any y, then $\sup\limits_x |\inf\limits_{y \geqslant x} g_1(y) - \inf\limits_{y \geqslant x} g_2(y)| < \delta$.

Now we are able to estimate Δ_1.

Lemma 2. Assume $\varkappa(X, Y) = \rho$, then $\Delta_1 \leqslant b_2 \rho$.

To formulate the main result demonstrating the model stability put $D_n = \frac{b_2(1-\alpha^n)}{1-\alpha} + \frac{b_1(\alpha-\alpha^n)}{1-\alpha}$.

Theorem 2. If $\varkappa(X, Y) = \rho$, then $\Delta_n \leqslant D_n \rho$.

The last problem for this model is the limit behavior of the company surplus as the planning horizon n tends to ∞. Let x be the initial capital. Since we use the reinsurance treaty with retention level a at each step, put $X_k^{(a)} = \min(X_k, a)$ where X_k denotes the claim amount in the k-th period.

According to Theorem 1 the optimal policy of insurer is characterized by the sequence of critical levels y_n as follows. At the first step of n-step process it is necessary to raise the initial capital to level y_n if $x \leqslant y_n$ and take no loan otherwise. Thus, if $Z_k^{(n)}$ is the surplus at the k-th step of the n-step process then $Z_0^{(n)} = x$ and for $k \geqslant 1$

$$Z_k^{(n)} = \begin{cases} y_{n+1-k} + M - X_k^{(a)}, & Z_{k-1}^{(n)} \leqslant y_{n+1-k}, \\ Z_{k-1}^{(n)} + M - X_k^{(a)}, & Z_{k-1}^{(n)} > y_{n+1-k}. \end{cases} \tag{3}$$

Theorem 3 (SLLN for surplus). For $x > a - M$ with probability 1

$$\frac{Z_n^{(n)}}{n} \to \delta(a) = M - EX^{(a)}, \text{ as } n \to \infty.$$

Thus, we can formulate the following

Corollary 2. The ultimate ruin probability of the company is equal to 1 if $\delta(a) \leqslant 0$.

One uses the properties of convergence in distribution, see, e.g., [2], and the properties of Gaussian distributions to derive the asymptotic normality.

Theorem 4 (CLT for surplus). For $x > a - M$

$$\frac{Z_n^{(n)} - EZ_n^{(n)}}{\sqrt{VarZ_n^{(n)}}} \xrightarrow{d} \mathcal{N}, \text{ as } n \to \infty,$$

here $\mathcal{N}$ has Gaussian distribution with parameters $(0, 1)$ and $\xrightarrow{d}$ signifies convergence in distribution.

Hence, it is not difficult to obtain the bounds on the size of surplus with probability $1 - \varepsilon$ for small $\varepsilon > 0$ and choose the appropriate values of retention level a and safety loadings β_i, $i = 1, 2$.

2. Generalized Cramér-Lundberg model

In order to make the model more realistic, we will consider n categories of insurance (not a single one), replace the compound Poisson process by a generalized Poisson process and include the investment. Based on these conditions, a risk model is set up to find (by the martingale methods) the formula of ruin probability and its upper bound. Thus, we put

$$U(t) = u_1 + (u_1 - u_2 - u_3)(b - c + d) + S(t),$$

$$S(t) = u_2 r_1 t + u_3(r_2 t + aB(t)) + \sum_{j=1}^{n} \sum_{i=1}^{N_j(t)} X_i^{(j)} - \sum_{j=1}^{n} \sum_{i=1}^{M_j(t)} Y_i^{(j)}.$$

Here $U(t)$ is the capital of insurance company and $S(t)$ is the gain at time t, u_1 is the initial reserve of insurance company, u_2 is invested in a non-risky asset, u_3 is used in venture capital, r_1 is rate of return of u_2, $u_3(r_2t+aB(t))$ is income from investment of u_3 in a Brownian motion with parameters r_2, a (r_2 is drift parameter, a is volatility, $B(t)$ is a standard Brownian motion), b is interest rate, c is inflation rate, d is exchange rate. Clearly, parameters satisfy the following inequalities $t \geq 0, u_1 > 0, u_2 > 0, u_3 > 0, r_1 > 0, r_2 > 0, a > 0$. Furthermore, $N_j(t)$ and $M_j(t)$ are Generalized Poisson processes providing inhomogeneity of premiums and claims flows, $X_i^{(j)}$ represents the ith premium, while $Y_i^{(j)}$ is the ith claim amount for the jth insurance branch. All r.v.'s possess finite means and variances.

Lemma 3. Surplus process $\{S(t), t \geq 0\}$ has stable independent increments.

Lemma 4. There is a function $h(r)$ satisfying the relation

$$E[exp(-rS(t))] = exp(th(r)).$$

Lemma 5. Equation $h(r) = 0$ has only one positive root R.

Theorem 5. For the Doubly Generalized Poisson multiple risk model, the upper bound of ruin is

$$\varphi(u) \leq \exp[-R(u_1 + (u_1 - u_2 - u_3)(b - c + d))].$$

Theorem 6. For the Doubly Generalized Poisson multiple risk model, the ruin probability is

$$\varphi(u) = \frac{\exp[-R(u_1 + (u_1 - u_2 - u_3)(b - c + d))]}{E[\exp[-RU(T)]|T < \infty]},$$

here T is the ruin time, that is, $T = \inf\{t \geq 0 : U(t) < 0\}$.

Conclusion

Two new insurance models were studied. The first one is discrete-time model of insurance company using non-proportional reinsurance and bank loans. The aim is minimization of expected additional costs associated with loans during n periods. It is proved that the optimal policy is determined by an increasing bounded sequence of critical levels. The model stability is established in terms of Kantorovich metric. For the optimal company surplus SLLN and CLT are proved. Further investigation direction is treatment of incomplete information, see, [5].

The second model describes a company having several business lines. The model is doubly stochastic and uses investment in risky and non-risky assets. Moreover, the input (premiums flow) and output (claims flow) are generalized Poisson processes (not compound ones). The properties of company surplus are studied. This model illustrates the reliability approach, providing the ruin probability. Next step is investigation of system stability and choice of optimal parameters, as in [6].

REFERENCES

1. *Bellman R.* Dynamic programming. Princeton, New Jersey: Princeton University Press, 1957. 340 p.
2. *Billingsley P.* Convergence of Probability Measures. New York: John Wiley and Sons, 1968. 287 p.
3. *Bulinskaya E.* New research directions in modern actuarial sciences // Modern problems of stochastic analysis and statistics – selected contributions in honor of Valentin Konakov (ed. V.Panov). Springer Proceedings in Mathematics & Statistics. 2017. Vol. 208. P. 349–408.
4. *Bulinskaya E.V.* Cost approach versus reliability // Proceedings of International Conference DCCN-2017, Moscow: Technosphera, 2017. P. 382–389.
5. *Bulinskaya E., Gusak J.* Insurance Models under Incomplete Information // Statistics and Simulation (J. Pilz et al. eds.). Springer Proceedings in Mathematics & Statistics. 2018. Vol. 231. P. 171–185.
6. *Bulinskaya E., Gusak J.* Optimal Control and Sensitivity Analysis for Two Risk Models // Communications in Statistics – Simulation and Computation. 2016. Vol. 45. № 5. P. 1451–1466.

Bulinskaya Ekaterina — Dr.Hab.Math.Phys.Sci., Full Professor, Department of Probability Theory. E-mail: *ebulinsk@yandex.ru*

STABILITY ANALYSIS OF CLASSICAL RETRIALS: A REVISED REGENERATIVE PROOF

E. Morozov[1,2,3], S. Rogozin[1,2]

[1] *Institute of Applied Mathematical Research Karelian Research Centre RAS, Petrozavodsk, Russian Federation*
[2] *Petrozavodsk State University, Petrozavodsk, Russian Federation*
[3] *Moscow Center for Fundamental and Applied Mathematics, Moscow State University, Moscow 119991, Russian Federation*

We consider a multiserver multiclass retrial system with classical retrials, and present a new short proof of the sufficient stability (positive recurrence) condition of the system.
Keywords: *Multiserver multiclass system, retrial system, stability condition, regenerative processes, outgoing calls.*

Introduction

The importance of retrial queues in modeling modern wireless telecommunication systems is well-known. To motivate it, we mention the most important sources, such like [1, 2, 3, 4]. However there are still many difficult problems in the performance and stability analysis such systems. In this work we focus on the stability analysis, and using regenerative arguments present a new short proof of the known sufficient stability condition of such retrial system. This proof is not only much shorter that that have been used earlier (see [1, 10]) but also allows easily to cover more general retrial systems with the so-called 'outgoing' calls [8]. Although many papers have investigated steady-state performance of retrial queues, still a little attention is devoted to stability conditions outside the Markovian setting, which is the topic of this research.

In this regard, we first mention a fundamental work [1] in which a detailed stability analysis of a general $G/G/1$-type single-server retrial queue is developed. We note that the authors study the system with a stationary input process and non-exponential retrial times, and also investigate the convergence rate to a stationary regime.

The stability of multiserver retrial systems is studied in a few papers. For instance, the paper [1] studies such a system with a finite buffer, batch Markovian arrival process, phase-type service time distribution and a general retrial rate, and stability analysis is based on the corresponding embedded Markov chains. We note that a comprehensive bibliography on research related to retrial queues can be found in the recognized sources [2, 3, 4].

In the present paper, we consider the stability of a multiserver retrial $M/G/1$ queue with independent Poisson inputs of different classes of primary customers, with generic (exponential) interarrival times τ_i with rates $\lambda_i = 1/\mathsf{E}\tau_i$, with iid service time $\{S_n^{(i)}\}$ of class i-customers (with generic service time $S^{(i)}$ and rate $\mu_i = 1/\mathsf{E}S^{(i)}$) and with class-dependent exponential retrial times, $i = 1, \dots, N$.

If an arriving class-i primary customer finds all servers busy, he joins the corresponding (infinite capacity) orbit i. Then, after exponentially distributed times, the customer attempts to capture server until he finds an idle server. We use the regenerative approach [5, 9] to reprove the stability condition, and this work thus continues our previous works [10] and [8]. In the work [10], the proof has been based on the negative drift of remaining work (in the single) orbit, while in the paper [8], we suggested the approach utilizing the positive drift of the idle time of server, under a predefined condition. In a contrast, in the current analysis we observe the system at the departure instants and evaluate the idle time of servers after each departure. This approach leads to a radical simplification of the stability analysis and, as we show, can be directly applied also to the system with *outgoing calls*. It means that if server is idle, it initiates a private outgoing call [8]. The idea to consider the output process is not new of course. For instance, the analysis of $M/G/1$-type retrial system in [1] is based on the analysis of an embedded Markov chain representing the orbit size at the service completion epochs. It is shown that this chain is ergodic under condition $\lambda/\mu < 1$. It is well-known that the latter condition is stability criterion of a standard buffered $M/G/1$ queueing system with work-conserving service discipline.

The main difference of a retrial system with classical one, from the point-of-view of stability analysis is that, in such a system, after each service completion, the server becomes idle for a random time until the beginning of the next service, implying a loss of server capacity after each departure, and thus the service discipline turns out to be not work-conserving. However, the service discipline in the retrial queueing system with classical retrials approaches the work-conserving discipline in the corresponding buffered system as the orbit size increases, and by this reason, it is called *asymptotically work-conserving* [10].

The main contribution of this paper is to present a new short proof of the sufficient stability condition of this model, which also allows easily to consider the system with the outgoing calls.

The rest of this paper is organized as follows. Section 2 describes the basic model and describes the regenerative structure of the process. In Section 3 we give the basic balance equation and outline the proof of stability condition.

1. Description of the model

In the present paper, we consider the stability of a multiserver retrial $M/G/1$ queue with independent Poisson inputs of different classes of primary customers, with generic (exponential) interarrival times τ_i with rates $\lambda_i = 1/\mathsf{E}\tau_i$, with iid service time $\{S_n^{(i)}\}$ of class i-customers (with generic service time $S^{(i)}$ and rate $\mu_i = 1/\mathsf{E}S^{(i)}$) and with class-dependent exponential retrial times, $i = 1, \dots, N$. (It is straightforward to extend analysis to a renewal input process where each customer belongs to class i with a probability p_i.)

If an arriving class-i primary customer finds all servers busy, he joins the corresponding (infinite capacity) orbit i. Then, after exponentially distributed times, with the rate $\mu_0^{(i)}$, the customer attempts to capture server until he finds a server idle. Denote $\mu_0 = \min \mu_0^{(i)}$. This is the so-called classical retrial policy, and if the orbit size (denoted by $Q(t)$ at instant t) equals N, then the retrial rate at this instant is lower bounded by $\mu_0 N$.

To describe the regenerative structure of the system, we denote by $Q(t)$ the total number of customers in the system at instant t^-, let $\{t_k\}$ be the arrival instants of the superposed input (Poisson) process and denote by $Q(t_k) = \widehat{Q}_k$. Then the regeneration instants $\{T_n\}$ of the process $\{Q(t),\, t \geqslant 0\}$ are recursively defined as

$$T_{n+1} = \inf_k (t_k > T_n : \widehat{Q}_k = 0),\ n \geqslant 0, \tag{1}$$

where we put $T_0 := 0$. The generic regeneration period, the distance between two regeneration points, is denoted by T. We call the regenerative process $\{Q(t)\}$ (and the queueing system) *positive recurrent* (stable) if the mean regeneration period is finite, $\mathsf{E}T < \infty$, and it implies stability of the system, while if $\mathsf{E}T = \infty$ then the system is called *null-recurrent* or unstable. Analogous properties hold for the queue-size process considered at the departure instances $\{d_k\}$, which is analysed below.

2. Stability analysis

Denote by $\rho_i = \lambda/\mu_i$, $\rho = \sum_i \rho_i$ and outline the proof of the following statement.

Theorem 1. If $\rho < m$ then the system under consideration is positive recurrent, $\mathsf{E}T < \infty$.

Denote by $V_i(t)$ the total workload which class-i customers bring in the system in time interval $[0, t]$. Also let $V(t) = \sum_i V_i(t)$. Moreover, denote by $W(t)$ the remaining work in all orbits at instant t, and let $I(t)$ be the total idle time of all servers in interval $[0, t]$. Assume that the system is initially empty. Then the following balance equation holds:

$$V(t) = W(t) + mt - I(t),\ t \geqslant 0. \tag{2}$$

It is easy to find, using the Strong Law of Large Numbers, that with probability 1 (w.p.1)

$$\lim_{t\to\infty} \frac{V(t)}{t} = \sum_i \rho_i = \rho, \tag{3}$$

regardless of whether the system is stable or not. Indeed it follows from theory of cumulative processes [11], that convergence in mean in (3) holds as well. Now we assume that the system is null-recurrent that is $\mathsf{E}T = \infty$, and in this case

$$Q(t) \Rightarrow \infty \text{ in probability as } t \to \infty. \tag{4}$$

This can be proved by contradiction: if we assume that $Q \not\Rightarrow \infty$ then it follows that $\mathsf{E}T < \infty$, see for instance [9]. Then we must show that indeed $\mathsf{E}I(t)/t \to 0$ as $t \to \infty$. However, in this research, we focus on another variant of the proof based on the following form of the balance equation constructed at the departure instances:

$$W(d_n) = V(d_n) - md_n + I(d_n),\ n \geqslant 1.$$

where, recall, $\{d_k\}$ are the departure instances. To simplify notation, denote by

$$W(d_n) = W_n,\ V(d_n) = V_n,\ I(d_n) = I_n,\ Q(d_n) = Q_n,\ n \geqslant 1.$$

Then in particular, in the zero initial state case, when the 1st customer arrives at instant $t_1 = 0$ in an empty system, we obtain

$$I_n = \sum_{k=1}^{n-1} \Delta_k,\ n \geqslant 1,$$

where $\Delta_k = I(d_{k+1}) - I(d_k)$ is the idle time of a server between the kth and $(k+1)$th departures and, by construction, $I_1 = 0$. Note that the following uniform upper bound holds regardless of the states of orbits:

$$\sup_k \mathsf{E}\Delta_k \leqslant 1/\lambda. \tag{5}$$

Then we prove the most challenging step. Namely, we show that, under assumption (3),

$$\lim_{n\to\infty} \mathsf{E}I_n/n = 0. \tag{6}$$

A key observation is that, provided $Q_k \geqslant n$, the mean idle time after the kth departure is upper bounded by the constant

$$C(n) := \frac{1}{\lambda + n\mu_0},$$

and $C(n) \to 0$ as $n \to \infty$. This shows that, if $Q_k \geqslant n$, then $\mathsf{E}\Delta_k \leqslant C(n)$ and can be done arbitrarily small for n large enough. The further analysis is based on the assumption (3). In particular, we show that, for arbitrary $\varepsilon > 0$, the convergence in mean

$$\delta := \limsup_{n\to\infty} \frac{\mathsf{E}I_n}{n} \leqslant \varepsilon \tag{7}$$

holds, implying $\delta = 0$. The final step of analysis is to rewrite the balance equation (2) as

$$\mathsf{E}W_n = \mathsf{E}V_n - d_n + \mathsf{E}I_n,$$

divide both sides of this expression by d_n and letting $n \to \infty$. It eventually gives $\rho \geqslant m$. This contradiction with the main assumption of Theorem shows that assumption (3) is false and $Q_n \not\Rightarrow \infty$. Then we show that the latter results implies that the remaining regeneration time $T(d_n)$ at instant d_n, also satisifies

$$T(d_n) \not\Rightarrow \infty,\ n \to \infty.$$

The rest of the proof follows by a standard argument based on the 'unloading' procedure [9] and 'regeneration' condition

$$\min_{1\leqslant i\leqslant N} \mathsf{P}(\tau > S^{(i)}) > 0,$$

which holds in this system automatically.

We also give a new short regenerative proof of stability condition of the system with the outgoing calls and extend the stability analysis to multi-server system.

3. Conclusion

In this work, we develop a new and short regenerative proof of the stability condition of a multiclass retrial system with classical retrials. Unlike previous proofs, we focus on the departures customers from the system, and provided the number of orbital customers increases infinitely, show a contradiction with a predefined negative drift condition, which thus turns out to be a sufficient stability condition.

REFERENCES

1. *Altman E., Borovkov A.A.* On the stability of retrial queues // Queueing systems. 1997. Vol. 26. P. 343–363.
2. *Artalejo J.R.* Accessible bibliography on retrial queues // Mathematical and Computer Modelling. 1999. Vol. 30. P. 1–6.
3. *Artalejo J.R.* A classified bibliography of research on retrial queues: Progress in 1990-1999 // Top. 1999. Vol. 7. P. 187–211.
4. *Artalejo J.R., Falin G.I.* Standard and retrial queueing systems: A comparative analysis // Revista Matematica Complutense. 2002. Vol. 15. P. 101–129.
5. *Asmussen S.* Applied Probability and Queues. N.Y.: Wiley, 1987. 438 p.
6. *Breuer L., Dudin A., Klimenok V.* A retrial BMAP/PH/N system // Queueing systems. 2002. Vol. 40. P. 433–457.
7. *Falin G.I., Templeton J.G.C.* Retrial Queues. London: Chapman and Hall, 1997. 320 p.
8. *Morozov E., Phung-Duc T.* Stability analysis of a multiclass retrial system with classical retrial policy // Performance Evaluation. 2017. Vol. 112. P. 15–26. (https://doi.org/10.1016/j.peva.2017.03.003)
9. *Morozov E., Steyaert B.* Stability Analysis of Regenerative Queueing Models. Springer, 2021. 185 p. (https://link.springer.com/book/10.1007/978-3-030-82438-9)
10. *Morozov E.* A multiserver retrial queue: Regenerative stability analysis // Queueing Systems. 2007. Vol. 56. P. 157–168.
11. *Smith W.L.* Regenerative stochastic processes // Proceedings of Royal Society, Ser. A. 1955. Vol. 232. P. 6–31.

Evsey Morozov — Doctor (DSc) of Physics and Mathematics, Professor, Chief Researcher in the Laboratory for Mathematical Cybernetics Institute of Applied Mathematical Research KarRC RAS and Professor of Institute of Mathematics and Information Technologies, Petrozavodsk State University. E-mail: *emorozov@karelia.ru*

Stepan Rogozin — PhD student in the Laboratory for Mathematical Cybernetics Institute of Applied Mathematical Research KarRC RAS. E-mail: *ppexa@mail.ru*

A LARGE DEVIATION ANALYSIS OF A RETRIAL QUEUE WITH GENERAL RETRIEVAL TIME

E. Morozov[1,2,3], K. Zhukova[1]

[1] *Karelian Research Centre RAS, Institute of Applied Mathematical Research, Petrozavodsk, Russian Federation*

[2] *Petrozavodsk State University, Petrozavodsk, Russian Federation*

[3] *Moscow Center for Fundamental and Applied Mathematics, Moscow State University, Moscow, Russian Federation*

We consider logarithmic asymptotics of the large deviation probability in a single-server queue with Poisson input, where server, after completion of service, seeks a customer from an orbit (retrial customer) for the next service, unless new arrival captures server. Service times of the input and orbital customers have different distributions. Under stability assumption, the asymptotic of the probability that the number of customers in the system reaches a high threshold within a regeneration cycle is found. The example for exponential retrieval time is given.

Keywords: *Retrial queue, large deviations, overflow probability.*

Introduction

The retrial queues (systems with the orbit) are widely used nowadays for describing modern computer and service systems [1, 2]. In such systems, if an arrival finds server busy, it joins infinite-capacity virtual orbit and repeat attempts to get service or wait until server select him for the next service. In practice it happens for example, when customer try to capture the manager calling several times or when customer leaves contact information and waits for calling back.

In these systems one of the important quality of service parameter is the probability that the number of waiting customers exceeds a given large threshold. The probability that describes this situation is called the overflow probability, and it can be studied by large deviation analysis. Previously this probability has been studied for the classic multi-server buffered system in [3]. Then this result has been adapted for the asymptotic analysis of the overflow probability in the tandem queues in [4], for the multi-server single-class retrial queue in [5] and for the multi-class retrial queue in [6].

In this work we construct the exact logarithmic asymptotics for the overflow probability for the single-server system with Poisson input and one orbit with different service times of the input and orbital customers.

We assume that the sequence of the seeking times is iid with a general distribution. For our analysis we apply the technique developed in [5].

1. Description of the model

We consider a single-server system with Poisson input. There is an infinite-capacity virtual orbit in the system. A new customer joins the orbit if server is busy upon the arrival. When the customer leaves system after being served, server begins to seek a customer from the orbit to be served next according to the First-Come-First-Served discipline: if the retrieval time is completed before the next arrival, the server begins to serve the oldest customer from the orbit. If a new customer arrives during the retrieval time, the server stops seeking process and begins to serve this new customer. The described model is well-motivated by practical applications, see [1, 2, 5].

In the described system we suppose that service time of the customer depends on whether it visits the orbit or not. We consider that the service time of the orbital customer has different distribution than the service time of the input customer that meets empty server upon the first arrival to the system and gets service immediately.

We assume a Poisson input of customers arriving at the instants $\{t_n\}$, with exponential interarrival times $\tau_n = t_{n+1} - t_n$, $t_1 = 0$ with input rate λ. Denote S_n the independent identically distributed (iid) sequence of service times of the customers that meet empty server upon arrival and get service immediately without going to the orbit. If a new arrival meets a busy server, it needs to go to the orbit. This orbital customer will be served later, and we denote its service time S_n^{or}. The sequence of S_n^{or} is also iid. Moreover, retrieval times $\{A_n\}$ are assumed to be iid as well, where A_n is the seeking time after the $(n-1)$-th departure from the system, and we put $A_1 := 0$. Denote input service rate $\mu = 1/\mathsf{E}S$, orbital service rate $\mu^{or} = 1/\mathsf{E}S^{or}$ and retrieval rate $\gamma = 1/\mathsf{E}A$. The difference between distribution of service times S and $S^{(or)}$ is motivated by practical applications. For example, when customers from the orbit require a shorter identification process by the server, we can assume that $S >_{st} S^{(or)}$ [2].

We study the logarithmic asymptotics of the stationary overflow probability P_N that the number of customers in the described system exceeds a fixed level N during regeneration cycle, as $N \to \infty$. The further analysis is based on the approach that has been previously applied in the works [5, 6].

2. Asymptotic of the overflow probability

We remain the number of customers in the system unchanged if assume that each new customer arriving in the system joins the orbit regardless of the state of server. If a new customer arrives during a seeking time, then the server starts to serve the oldest orbital customer immediately instead of the new arrival, which in turn joins the 'end' of the virtual orbit. If the seeking time of the server is not interrupted by the newly arrived customer, then the behaviour of the system remains as in conventional system: the server takes the oldest customer from the orbit. When the served customer leaves the system, server starts to seek the next orbital customer, and the process repeats. We stress that possible change of the order of the customers in this (modified) system does not change distribution of the total number of customers in this system because of the stochastic equivalence of the service times of the customers. On the other hand, this replacement of the customers is convenient for the analysis, because we keep the order of the service the same as the order of arrivals, that is, we preserve FIFO discipline. It allows us to consider the original system with the orbit as a classic buffered system and moreover, applying coupling method, to use the same service times in both system (original and modified) for each nth customer entering server in each system. (For more detail see [5].) Furthermore, the idle time of the server before initiation of the nth service can be interpret as a part of the service time of the next customer, and hence we can define the nth service time in the modified system as

$$\hat{S}_n = S_n \cdot \mathsf{I}_n + S_n^{(or)}(1 - \mathsf{I}_n) + \min(\tau_n, A_n),$$

where

$$\mathsf{I}_n = \begin{cases} 1, \text{ if } \tau_n < A_n, \\ 0, \text{ if } \ \tau_n \geqslant A_n. \end{cases}$$

So, $\hat{S} = S + \tau$ with probability $\mathsf{P}(\tau < A)$, and $\hat{S} = S^{(or)} + \tau$ with probability $\mathsf{P}(\tau \geqslant A)$. Then exponential moment of $\hat{S}$ can be written as

$$\mathsf{E}e^{\theta\hat{S}} = \mathsf{E}[e^{\theta(S+\tau)}|\tau < A]\mathsf{P}(\tau < A) + \mathsf{E}[e^{\theta(S^{(or)}+A)}|\tau \geqslant A]\mathsf{P}(\tau \geqslant A).$$

By independence S and τ,

$$\mathsf{E}[e^{\theta(S+\tau)}|\tau < A]\mathsf{P}(\tau < A) = \mathsf{E}e^{\theta S}\mathsf{E}[e^{\theta\tau}; \tau < A],$$

where, by the independence τ and A,

$$\mathsf{E}[e^{\theta\tau}; \tau < A] = \int_0^\infty e^{\theta x}\lambda e^{-\lambda x}\mathsf{P}(A > x)dx.$$

Assume $\theta < \lambda$. Then, by integration by parts, after some algebra,

$$\int_0^\infty e^{\theta x} \lambda e^{-\lambda x} \mathsf{P}(A > x) dx = \frac{\lambda}{\lambda - \theta}(1 - f_A(\lambda - \theta)),$$

where $f_A(s) := \mathsf{E} e^{-sA}$ is the Laplace transform of the random variable A.

Similarly, by the independence $S^{(or)}$ and A we get

$$\mathsf{E}[e^{\theta(S^{(or)}+A)} | \tau \geqslant A] \mathsf{P}(\tau \geqslant A) = \mathsf{E} e^{\theta S^{(or)}} \mathsf{E}[e^{\theta A}; A \leqslant \tau],$$

where

$$\mathsf{E}[e^{\theta A}; A \leqslant \tau] = \int_0^\infty e^{\theta x} e^{-\lambda x} \mathsf{P}(A \leqslant x) dx = f_A(\lambda - \theta).$$

The log moment generating function for r.v. $\hat{S}$ can be then written as

$$\Lambda_{\hat{S}}(\theta) = \log \mathsf{E} e^{\theta \hat{S}} = \log \left[\frac{\lambda}{\lambda - \theta} \mathsf{E} e^{\theta S}(1 - f_A(\lambda - \theta)) + \mathsf{E} e^{\theta S^{(or)}} f_A(\lambda - \theta) \right].$$

Stability condition in this case becomes

$$\lambda[(1 - f_A(\lambda)) \mathsf{E} S + f_A(\lambda) \mathsf{E} S^{(or)}] + \lambda \mathsf{E} \min(\tau, A) < 1. \tag{1}$$

Now we define the maximal moment indexes

$$\theta_1 = \sup(\theta > 0 : \mathsf{E} e^{\theta S^{or}} < \infty) > 0, \tag{2}$$

$$\theta_2 = \sup(\theta > 0 : \mathsf{E} e^{\theta S} < \infty) > 0, \tag{3}$$

$$\theta_3 = \sup(\theta > 0 : \mathsf{E} e^{\theta A} < \infty) > 0. \tag{4}$$

Then the following result for the logarithmic asymptotic for the P_N are proved as in [5].

Theorem 1. Assume that condition (1) holds true. Then

$$\lim_{N \to \infty} \frac{1}{N} \log \mathsf{P}_N = \Lambda_\tau(-\theta_*), \tag{5}$$

where parameter θ^* is defined as

$$\theta_* = \sup_\theta \Big(0 < \theta < \min(\lambda, \theta_1, \theta_2, \theta_3) : \Lambda_\tau(-\theta) + \Lambda_{\hat{S}}(\theta) \leqslant 0 \Big). \tag{6}$$

3. Example

In this section we calculate function $\Lambda(\theta)$ and parameter θ_* for the case of the exponential distributions of the service and retrieval times.

Since the input is Poisson, let the interval τ be exponential with parameter λ. We consider exponential service time S with parameter μ, service time $S^{(or)}$ of the orbital customers is exponential with parameter $\mu^{(or)}$ and retrieval time A is exponential with parameter γ. In this case, as it is easy to calculate parameters (2)-(4)

$$\theta_1 = \mu^{(or)}, \ \theta_2 = \mu, \ \theta_3 = \gamma.$$

For the Poisson input

$$\Lambda_\tau(-\theta) = \log \frac{\lambda}{\theta + \lambda}.$$

Since retrieval and service times are exponential,

$$f_A(\lambda - \theta) = \frac{\gamma}{\gamma + \lambda - \theta}, \quad \mathsf{E}e^{\theta S} = \frac{\mu}{\mu - \theta}, \quad \mathsf{E}e^{\theta S^{(or)}} = \frac{\mu^{(or)}}{\mu^{(or)} - \theta}.$$

Then

$$\Lambda_{\hat{S}}(\theta) = \log \mathsf{E}e^{\theta \hat{S}} = \log \left[\frac{\lambda}{\lambda - \theta} \mathsf{E}e^{\theta S}(1 - f_A(\lambda - \theta)) + \mathsf{E}e^{\theta S^{(or)}} f_A(\lambda - \theta) \right]$$
$$= \log \left[\frac{\lambda}{\lambda - \theta} \cdot \frac{\mu}{\mu - \theta} \cdot \frac{\lambda - \theta}{\gamma + \lambda - \theta} + \frac{\mu^{(or)}}{\mu^{(or)} - \theta} \cdot \frac{\gamma}{\gamma + \lambda - \theta} \right].$$

Then parameter (6) satisfies

$$\theta_* = \sup \Big(\theta \in (0, \min(\gamma, \mu^{(or)}, \mu)) :$$
$$\frac{\lambda^2 \mu(\lambda - \theta)}{(\lambda^2 - \theta^2)(\gamma + \lambda - \theta)(\mu - \theta)} + \frac{\gamma \lambda \mu^{(or)}}{(\lambda + \theta)(\gamma + \lambda - \theta)(\mu^{(or)} - \theta)} \leqslant 1 \Big).$$

4. Conclusion

In the paper we consider single-server queueing system with an orbit, where the server needs some time to choose an orbital customer. We study the logarithmic asymptotics for overflow probability that the number of customers in the system reaches a high level N within a regeneration cycle. The system is fed by a Poisson input and the server retrieval time is assumed to be generally distributed. Moreover, the different service times for input and orbital customers are considered. The example is included as well.

REFERENCES

1. *Kim J., Kim B.* Tail asymptotics for the queue size distribution in the MAP/G/1 retrial queueTail asymptotics for the queue size distribution in the MAP/G/1 retrial queue // Queueing System. 2010. Vol. 66. P. 79–94.
2. *Baron O., Economou A., Manou A.* The state-dependent M/G/1 queue with orbit // Queueing System. 2018. Vol. 90. P. 89–123.
3. *Sadowsky J.* Large deviations theory and efficient simulation of excessive backlogs in a GI/GI/m queue // IEEE Trans. Autom. Control. 1991. Vol. 36(12). P. 1383–1394.
4. *Buijsrogge A., de Boer P., Rosen K., Scheinhardt W.* Large deviations for the total queue size in non-Markovian tandem queues // Queueing System. 2017. Vol. 85. P. 305–312.
5. *Morozov E., Zhukova K.* A large deviation analysis of retrial models with constant and classic retrial rates // Performance Evaluation. 2019. Vol. 135.
6. *Morozov E., Zhukova K.* The overflow probability asymptotics in amulticlass single-server retrial system // Distributed Computer and Communication Networks. Communications in Computer and Information Science: Springer, 2020. Vol. 1337. P. 394–406.

Morozov Evsey Viktorovich — Doctor of Physics and Mathematics, professor, IPMI KarSC RAS. E-mail: *emorozov@karelia.ru*

Zhukova Ksenia Alekseevna — PhD of Physics and Mathematics, IPMI KarSC RAS. E-mail: *kalininaksenia90@gmail.com*

DIFFUSION APPROXIMATION FOR RETRIAL QUEUE WITH COLLISIONS AND NON-PERSISTENT CUSTOMERS

A. Nazarov, O. Lizyura

Tomsk State University, Tomsk, Russian Federation

This paper is devoted to the analysis of retrial queue with an arbitrary distribution of service times, collisions, and non-persistent customers. Our aim is to investigate the number of customers in the orbit of the system. To this end, we use the asymptotic-diffusion method to build a diffusion approximation for the steady-state distribution of the number of customers in the orbit.
Keywords: *Retrial queue, collision, non-persistent customer, diffusion approximation.*

Introduction

Retrial queues arose as models of communication systems. The basic phenomenon of such systems is the retrial behavior of customers: if the server is busy upon arrival, the customer enters the orbit and repeats the attempt to access the server after a random amount of time.

There are several modifications of retrial queues that reflect the system features such as collisions and non-persistent customers, which appear in various switching communication systems and CSMA-based networks [1]. In recent years, queueing systems with collisions are of interest due to the reborn of IEEE 802.11 wireless LANs. In paper [2], authors describe the markovian retrial queue with collisions and shows applications of persistence to modeling CSMA-CD protocols. In paper [3], the author consider similar markovian model and takes into account the impatience of customers.

Nazarov and Sztrik with their research group have considered several models of finite-source retrial queues with collisions [4, 5, 6]. The phenomena of non-persistent customers in retrial queues was considered in [7]. Lakaour and his colleagues have considered markovian models with collisions, transmission errors and unreliable server [8, 9].

We consider retrial queue with arbitrary distribution of service times, collisions and r-persistent customers. We build diffusion approximation for the number of customers in the orbit and construct the approximation of its probability distribution under the limit condition of growing delay in the orbit.

1. Model Description and Problem Definition

We consider a retrial queue with an arbitrary distribution of service times defined by the distribution function $B(x)$. The input is stationary Poisson process with rate λ. If the server is idle upon arrival, the incoming customer occupies it for service. Otherwise, the collision occurs and one of the customers joins the orbit. The other customer can also join the orbit with probability r or leave the system with probability $(1-r)$.

At the orbit, a customer waits for some random time and tries again to occupy the server. The duration of delay follows an exponential distribution with rate σ.

Let $k(t)$ denote the state of the server at instant t: 0, if the server is idle; 1, if the server is busy. Let $i(t)$ denote the number of customers in the orbit at instant t. We also introduce process $z(t)$, which represents the residual service time. Thus, process $\{k(t),\ i(t),\ z(t)\}$ has variable number of components and exhaustively describes the system state. We denote the probability distribution of process $\{k(t),\ i(t),\ z(t)\}$ as follows:

$$P_0(i,t) = P\{k(t)=0, i(t)=i\},\ P_1(i,z,t) = P\{k(t)=1, i(t)=i, z(t)<z\},$$

and introduce the partial characteristic functions

$$H_0(u,t) = \sum_{i=0}^{\infty} e^{jui}P_0(i,t),\ H_1(u,z,t) = \sum_{i=0}^{\infty} e^{jui}P_1(i,z,t),$$

where j is the imaginary unit. The Kolmogorov system of differential equations for the partial characteristic functions has the following form:

$$\begin{aligned}
&\frac{\partial H_0(u,t)}{\partial t} = -\lambda H_0(u,t) + j\sigma\frac{\partial H_0(u,t)}{\partial u} + \frac{\partial H_1(u,0,t)}{\partial z} + \\
&+\lambda e^{ju}(1+r(e^{ju}-1))H_1(u,t) - j\sigma(1+r(e^{ju}-1))\frac{\partial H_1(u,t)}{\partial u},\\
&\frac{\partial H_1(u,z,t)}{\partial t} = \frac{\partial H_1(u,z,t)}{\partial z} - \frac{\partial H_1(u,0,t)}{\partial z} - \lambda H_1(u,z,t) + \\
&+j\sigma\frac{\partial H_1(u,z,t)}{\partial u} + \lambda H_0(u,t)B(z) - j\sigma e^{-ju}\frac{\partial H_0(u,t)}{\partial u}B(z).
\end{aligned} \qquad (1)$$

After that, we sum up the equations of system (2). Taking the limit by $z \to \infty$, we obtain

$$\frac{\partial H(u,t)}{\partial t} = (e^{ju}-1)\times$$

$$\times \left\{ j\sigma e^{-ju}\frac{\partial H_0(u,t)}{\partial u} + \lambda(1+re^{ju})H_1(u,t) - j\sigma r\frac{\partial H_1(u,t)}{\partial u} \right\}. \qquad (2)$$

Solving system (2) and equation (4) in the limit by $\sigma \to 0$, we derive drift and diffusion coefficients of approximating diffusion process.

2. Asymptotic-Diffusion Analysis

In system (2) and equation (4), we introduce the following notations:

$$\sigma = \varepsilon,\ u = \varepsilon w,\ \tau = \varepsilon t,$$

$$H_0(u,t) = F_0(w,\tau,\varepsilon),\ H_1(u,z,t) = F_1(w,z,\tau,\varepsilon), \qquad (3)$$

and obtain the system of equations

$$\varepsilon\frac{\partial F_0(w,\tau,\varepsilon)}{\partial \tau} = -\lambda F_0(w,\tau,\varepsilon) + j\frac{\partial F_0(w,\tau,\varepsilon)}{\partial w} + \frac{\partial F_1(w,0,\tau,\varepsilon)}{\partial z} +$$

$$+\lambda e^{jw\varepsilon}(1+r(e^{jw\varepsilon}-1))F_1(w,\tau,\varepsilon) - j(1+r(e^{jw\varepsilon}-1))\frac{\partial F_1(w,\tau,\varepsilon)}{\partial w},$$

$$\varepsilon\frac{\partial F_1(w,z,\tau,\varepsilon)}{\partial \tau} = \frac{\partial F_1(w,z,\tau,\varepsilon)}{\partial z} - \frac{\partial F_1(w,0,\tau,\varepsilon)}{\partial z} - \lambda F_1(w,z,\tau,\varepsilon) +$$

$$+j\frac{\partial F_1(w,z,\tau,\varepsilon)}{\partial w} + \lambda F_0(w,\tau,\varepsilon)B(z) - je^{-jw\varepsilon}\frac{\partial F_0(w,\tau,\varepsilon)}{\partial w}B(z),$$

$$\varepsilon\frac{\partial F(w,\tau,\varepsilon)}{\partial \tau} = (e^{jw\varepsilon}-1)\times$$

$$\times \left\{ je^{-jw\varepsilon}\frac{\partial F_0(w,\tau,\varepsilon)}{\partial w} + \lambda(1+re^{jw\varepsilon})F_1(w,\tau,\varepsilon) - jr\frac{\partial F_1(w,\tau,\varepsilon)}{\partial w} \right\}. \qquad (4)$$

We solve system (1) in the limit by $\varepsilon \to 0$ and formulate the following theorem.

Theorem 1. In considered retrial queue, under the limit condition $\sigma \to 0$, the following equality holds:

$$\lim_{\sigma\to 0} \mathbb{E}\left\{ \sigma i\left(\frac{\tau}{\sigma}\right) \right\} = x(\tau),$$

where $x(\tau)$ is a solution of differential equation

$$x'(\tau) = -x(\tau)r_0 + [\lambda + (\lambda + x(\tau))r]r_1, \qquad (5)$$

values r_0, r_1 have the following form:

$$r_0 = \frac{1}{2 - B^*(\lambda + x)},\ r_1 = \frac{1 - B^*(\lambda + x)}{2 - B^*(\lambda + x)}. \qquad (6)$$

Here $B^*(s)$ is the Laplace-Stieltjes transform (LST) of the distribution function of the service times $B(x)$.

From (4), we denote function

$$a(x) = -xr_0 + (\lambda + (\lambda + x)r)r_1. \tag{7}$$

For the second step of analysis, we make the following substitutions in equations (2)-(4):

$$H_0(u,t) = e^{j\frac{u}{\sigma}x(\sigma t)} H_0^{(2)}(u,t),\ H_1(u,z,t) = e^{j\frac{u}{\sigma}x(\sigma t)} H_1^{(2)}(u,z,t).$$

Thus, we obtain the equations for the partial characteristic functions of centered number of customers in the orbit. After that, we introduce the following substitutions:

$$\sigma = \varepsilon^2,\ u = w\varepsilon,\ \tau = t\varepsilon^2,$$
$$H_0^{(2)}(u,t) = F_0^{(2)}(w,\tau,\varepsilon),\ H_1^{(2)}(u,z,t) = F_1^{(2)}(w,z,\tau,\varepsilon), \tag{8}$$

and obtain the system of equations

$$\varepsilon^2 \frac{\partial F_0^{(2)}(w,\tau,\varepsilon)}{\partial \tau} + jw\varepsilon a(x) F_0^{(2)}(w,\tau,\varepsilon) = -(\lambda + x) F_0^{(2)}(w,\tau,\varepsilon) +$$
$$+ j\varepsilon \frac{\partial F_0^{(2)}(w,\tau,\varepsilon)}{\partial w} + \frac{\partial F_1^{(2)}(w,0,\tau,\varepsilon)}{\partial z} +$$
$$+ (\lambda e^{jw\varepsilon} + x)(1 + r(e^{jw\varepsilon} - 1)) F_1^{(2)}(w,\tau,\varepsilon) -$$
$$- j\varepsilon (1 + r(e^{jw\varepsilon} - 1)) \frac{\partial F_1^{(2)}(w,\tau,\varepsilon)}{\partial w},$$

$$\varepsilon^2 \frac{\partial F_1^{(2)}(w,z,\tau,\varepsilon)}{\partial \tau} + jw\varepsilon a(x) F_1^{(2)}(w,z,\tau,\varepsilon) = \frac{\partial F_1^{(2)}(w,z,\tau,\varepsilon)}{\partial z} -$$
$$- \frac{\partial F_1^{(2)}(w,0,\tau,\varepsilon)}{\partial z} - (\lambda + x) F_1^{(2)}(w,z,\tau,\varepsilon) + j\varepsilon \frac{\partial F_1^{(2)}(w,z,\tau,\varepsilon)}{\partial w} +$$
$$+ (\lambda + x e^{-jw\varepsilon}) F_0^{(2)}(w,\tau,\varepsilon) B(z) - j\varepsilon e^{-jw\varepsilon} \frac{\partial F_0^{(2)}(w,\tau,\varepsilon)}{\partial w} B(z),$$

$$\varepsilon^2 \frac{\partial F^{(2)}(w,\tau,\varepsilon)}{\partial \tau} + jw\varepsilon a(x) F^{(2)}(w,\tau,\varepsilon) =$$
$$= (e^{jw\varepsilon} - 1) \left\{ j\varepsilon e^{-jw\varepsilon} \frac{\partial F_0^{(2)}(w,\tau,\varepsilon)}{\partial w} - x e^{-jw\varepsilon} F_0^{(2)}(w,\tau,\varepsilon) + \right.$$

$$\left. +(\lambda + r(\lambda e^{jw\varepsilon} + x))F_1^{(2)}(w,\tau,\varepsilon) - j\varepsilon r \frac{\partial F_1^{(2)}(w,\tau,\varepsilon)}{\partial w} \right\}. \tag{9}$$

Solving system (9) in the limit by $\varepsilon \to 0$, we present Theorem 2.

Theorem 2. Function $\lim\limits_{\varepsilon \to 0} F_k^{(2)}(w,\tau,\varepsilon) = F_k^{(2)}(w,\tau)$ has the following form:

$$F_k^{(2)}(w,\tau) = \Phi(w,\tau) r_k,$$

where r_k is given by (6), function $\Phi(w,\tau)$ is the solution of equation

$$\frac{\partial \Phi(w,\tau)}{\partial \tau} = w \frac{\partial \Phi(w,\tau)}{\partial w} a'(x) + \frac{(jw)^2}{2} \Phi(w,\tau) b(x). \tag{10}$$

Function $a(x)$ is defined by (5), $b(x)$ is determined as follows:

$$b(x) = a(x) + 2[-(\lambda + x)(1 + r)g_0 + x r_0 + r\lambda r_1], \tag{11}$$

where

$$g_0 = \frac{(a(x) + x)(1 - B^*(\lambda + x)) + (\lambda + x)a(x)B^{*\prime}(\lambda + x)}{(\lambda + x)(2 - B^*(\lambda + x))^2}.$$

Here equation (10) is the Fourier transform of the Fokker-Planck equation for the process approximating the number of customers in the orbit of considered retrial queue. If we make the inverse Fourier transform, we can see that the drift coefficient of the obtained diffusion limit is $a(x)$ and diffusion coefficient if $b(x)$.

Theorem 3. Discrete function $PD(i)$ is the approximation of the probability distribution of the number of customers in the orbit and has the following form:

$$PD(i) = \frac{D(i\sigma)}{\sum\limits_{n=0}^{\infty} D(n\sigma)}, \tag{12}$$

where

$$D(z) = \frac{1}{b(z)} \int\limits_0^z \frac{2}{\sigma} \frac{a(x)}{b(x)} dx.$$

3. Conclusion

We have considered the retrial queue with collisions and r-persistent customers. For the number of customers in the orbit, we have derived the approximation of the probability distribution (12). In future, we plan to investigate the stability mode in such system based on the obtained approximation.

REFERENCES

1. *Choi B.D., Shin Y.W., Ahn W.C.* Retrial queues with collision arising from unslotted CSMA/CD protocol // Queueing Systems. 1992. Vol. 11. P. 335–356.
2. *Krishna Kumar B., Vijayalakshmi G., Krishnamoorthy A., Sadiq Basha S.* A single server feedback retrial queue with collisions // Computers & Operations Research. 2010. Vol. 37. P. 1247–1255.
3. *Kim J.* RETRIAL QUEUEING SYSTEM WITH COLLISION AND IMPATIENCE // Commun. Korean Math. Soc. 2010. Vol. 25. № 4. P. 647–653.
4. *Nazarov A., Sztrik J., Kvach A., Toth A.* Some Features of a Finite-Source M/GI/1 Retrial Queuing System with Collisions of Customers // Communications in Computer and Information Science. 2017. Vol. 700. P. 186–200.
5. *Nazarov A., Sztrik J., Kvach A., Toth A.* Asymptotic sojourn time analysis of finite-source M/M/1 retrial queueing system with collisions and server subject to breakdowns and repairs // Annals of Operations Research. 2020. Vol. 288. P. 417–434.
6. *Toth A., Sztrik J.* Simulation of Finite-Source Retrial Queuing Systems With Collisions, Non-Reliable Server and Impatient Customers in the Orbit // Proceedings of the 11 th International Conference on Applied Informatics, Eger: ICAI, 2020. P. 408–419.
7. *Jailaxmi V., Arumuganathan R., Rathinasamy A.* Performance analysis of an Mx/G/1 feedback retrial queue with non-persistent customers and multiple vacations with N-policy // Int. J. Operational Research. 2017. Vol. 29. № 2. P. 149–169.
8. *Lakaour L., Aissani D., Adel-Aissanou K., Barkaoui K.* M/M/1 Retrial Queue with Collisions and Transmission Errors // Methodology and Computing in Applied Probability. 2019. Vol. 21. P. 1395–1406.
9. *Lakaour L., Aissani D., Adel-Aissanou K., Barkaoui K., Ziani S.* An unreliable single server retrial queue with collisions and transmission errors // Communications in Statistics - Theory and Methods. 2020. DOI: 10.1080/03610926.2020.1758943.

Nazarov Anatoly — doctor of technical sciences, full professor, professor of department of probability theory and mathematical statistics. E-mail: *nazarov.tsu@gmail.com*

Lizyura Olga — assistant professor of department of probability theory and mathematical statistics. E-mail: *oliztsu@mail.ru*

RESOURCE QUEUEING SYSTEM WITH FEEDBACK AS A MODEL FOR SERVICING TWO TYPES OF TRAFFIC

A. Nazarov, R. Salimzyanov, E. Lisovskaya

National Research Tomsk State University, Tomsk, Russia

The resource multi-server queue with two arrival processes is considered. The analytical expressions for the means of the number of customers in the orbits are found. The stationary distribution of the two-dimensional stochastic process of numbers of customers in the service block is found, which allows calculating the main probability and numerical characteristics. A numerical example is considered, where the applicability area of approximation was obtained.
Keywords: *Retrial queue, two Poisson arrivals, limited resource, asymptotic analysis method.*

Introduction

With the growing interest in fifth-generation mobile networks, various models are being actively developed to improve the quality of service and various mathematical models are being built to study the impact of various conditions on the quality of service [1] or research additional ways to reduce the load on the network [2]. But there is also the task of optimizing the operation of a specific base station for customers with different communication quality, given that up to 2030 there will be a sharp question of servicing two types of traffic [3]. To solve such an optimization problem, we can use the queueing theory.

In this paper, we consider the resource queue with feetback and two arrivals as a model of a strategy of service two types of traffic on one base station. Such a model allows controlling calls, which are not serviced immediately. We can obtain the main probability and numerical characteristics of such system performance.

1. Problem Statement

Consider the multi-server queue with two Poisson arrivals with intensities λ_1 and λ_2, respectively. Customers' service durations are exponentially distributed random variables with parameters μ_1 and μ_2. Each customer requires x_1 and x_2 resource units, respectively, for the first and second arrivals. The total available resource is limited and equal R units.

Each arrival takes up any free server, if there are a sufficient amount of free resources. In the case, when there is not enough resource, the customer goes to the corresponding orbit, where stay for a random time distributed exponentially with the parameters σ_1 and σ_2, after which it tries to access the system. If there is enough resource, it gets up on service, otherwise, it remains in orbit until the next try to get up on service.

Our goal is to study the four-dimensional stochastic process $\{n_1(t), n_2(t), i_1(t), i_2(t)\}$, where $n_1(t)$, $n_2(t)$ are the number of customers of the first and second arrivals, which be serviced at time t, and $i_1(t)$, $i_2(t)$ are the number of customers in the first and second orbits at time t. The states space of this process is defined as

$$\mathbb{X} = \{(n_1, n_2, i_1, i_2) : x_1 n_1 + x_2 n_2 \leqslant R, i_1, i_2 > 0\}.$$

2. Main System of Equations

We denote the stationary probability distribution of the system states

$$P(n_1, n_2, i_1, i_2) = \lim_{t\to\infty} P(n_1(t) = n_1, n_2(t) = n_2, i_1(t) = i_1, i_2(t) = i_2),$$

for which we write the system of equilibrium equations

$$\begin{gathered}
[\lambda_1 + \lambda_2 + n_1\mu_1 + n_2\mu_2 + i_1\sigma_1 \cdot I(x_1(n_1+1) + x_2 n_2 \leqslant R, i_1 > 0) + \\
+ i_2\sigma_2 \cdot I(x_1 n_1 + x_2(n_2+1) \leqslant R, i_2 > 0)] \cdot P(n_1, n_2, i_1, i_2) = \\
= \lambda_1 \cdot I(n_1 > 0) \cdot P(n_1 - 1, n_2, i_1, i_2) + \lambda_2 \cdot I(n_2 > 0) \cdot P(n_1, n_2 - 1, i_1, i_2) + \\
+ \lambda_1 \cdot I(x_1(n_1+1) + x_2 n_2 > R, i_1 > 0) \cdot P(n_1, n_2, i_1 - 1, i_2) + \\
+ \lambda_2 \cdot I(x_1 n_1 + x_2(n_2+1) > R, i_2 > 0) \cdot P(n_1, n_2, i_1, i_2 - 1) + \\
+ (n_1+1)\mu_1 \cdot P(n_1+1, n_2, i_1, i_2) \cdot I(x_1(n_1+1) + x_2 n_2 \leqslant R) + \\
+ (n_2+1)\mu_2 \cdot P(n_1, n_2+1, i_1, i_2) \cdot I(x_1 n_1 + x_2(n_2+1) \leqslant R) + \\
+ (i_1+1)\sigma_1 \cdot I(n_1 > 0) \cdot P(n_1 - 1, n_2, i_1 + 1, i_2) + \\
+ (i_2+1)\sigma_2 \cdot I(n_2 > 0) \cdot P(n_1, n_2 - 1, i_1, i_2 + 1).
\end{gathered}$$

We introduce the partial characteristic functions

$$H(n_1, n_2, u_1, u_2) = \sum_{i_1=0}^{\infty} e^{ju_1 i_1} \sum_{i_2=0}^{\infty} e^{ju_2 i_2} P(n_1, n_2, i_1, i_2), \quad j = \sqrt{-1},$$

we rewrite the system in the operator form

$$(\mathbf{A}+\lambda_1 e^{ju_1}\mathbf{B_1}+\lambda_2 e^{ju_2}\mathbf{B_2})\mathbf{H}(u_1,u_2)+j\sigma_1(\mathbf{C_1}-e^{-ju_1}\mathbf{D_1})\frac{\partial\mathbf{H}(u_1,u_2)}{\partial u_1}+$$

$$+j\sigma_2(\mathbf{C_2}-e^{-ju_2}\mathbf{D_2})\frac{\partial\mathbf{H}(u_1,u_2)}{\partial u_2}=0, \qquad (1)$$

where $\mathbf{H}(u_1,u_2)$ is a matrix of $H(n_1,n_2,u_1,u_2)$, and $\mathbf{A}$, $\mathbf{B_1}$, $\mathbf{B_2}$, $\mathbf{C_1}$, $\mathbf{C_2}$, $\mathbf{D_1}$, $\mathbf{D_2}$ are operators satisfying the conditions:

$$\mathbf{AH}(u_1,u_2)=\begin{cases}-[\lambda_1+\lambda_2]H(0,0,u_1,u_2)+\mu_1H(1,0,u_1,u_2)+\\ +\mu_2H(0,1,u_1,u_2), & \mathbb{X}_1,\\ -[\lambda_1+\lambda_2+n_1\mu_1+n_2\mu_2]H(n_1,n_2,u_1,u_2)+\\ +\lambda_1H(n_1-1,n_2,u_1,u_2)+\lambda_2H(n_1,n_2-1,u_1,u_2)+\\ +(n_1+1)\mu_1H(n_1+1,n_2,u_1,u_2)+\\ +(n_2+1)\mu_2H(n_1,n_2+1,u_1,u_2), & \mathbb{X}_2,\\ -[\lambda_1+\lambda_2+n_1\mu_1+n_2\mu_2]H(n_1,n_2,u_1,u_2)+\\ +\lambda_1H(n_1-1,n_2,u_1,u_2)+\lambda_2H(n_1,n_2-1,u_1,u_2), & \mathbb{X}_3,\\ -[\lambda_1+\lambda_2+n_1\mu_1+n_2\mu_2]H(n_1,n_2,u_1,u_2)+\\ +\lambda_1H(n_1-1,n_2,u_1,u_2)+\lambda_2H(n_1,n_2-1,u_1,u_2)+\\ +(n_1+1)\mu_1H(n_1+1,n_2,u_1,u_2), & \mathbb{X}_4,\\ -[\lambda_1+\lambda_2+n_1\mu_1+n_2\mu_2]H(n_1,n_2,u_1,u_2)+\\ +\lambda_1H(n_1-1,n_2,u_1,u_2)+\lambda_2H(n_1,n_2-1,u_1,u_2)+\\ +(n_2+1)\mu_1H(n_1,n_2+1,u_1,u_2), & \mathbb{X}_5,\end{cases}$$

$$\mathbf{B_1H}(u_1,u_2)=\begin{cases}0, & \mathbb{X}_1,\mathbb{X}_2,\mathbb{X}_4,\\ H(n_1,n_2,u_1,u_2), & \mathbb{X}_3,\mathbb{X}_5,\end{cases}$$

$$\mathbf{B_2H}(u_1,u_2)=\begin{cases}0, & \mathbb{X}_1,\mathbb{X}_2,\mathbb{X}_5,\\ H(n_1,n_2,u_1,u_2), & \mathbb{X}_3,\mathbb{X}_4,\end{cases}$$

$$\mathbf{C_1H}(u_1,u_2)=\begin{cases}H(n_1,n_2,u_1,u_2), & \mathbb{X}_1,\mathbb{X}_2,\mathbb{X}_4,\\ 0, & \mathbb{X}_3,\mathbb{X}_5,\end{cases}$$

$$\mathbf{C_2H}(u_1,u_2)=\begin{cases}H(n_1,n_2,u_1,u_2), & \mathbb{X}_1,\mathbb{X}_2,\mathbb{X}_5,\\ 0, & \mathbb{X}_3,\mathbb{X}_4,\end{cases}$$

$$\mathbf{D_1H}(u_1,u_2)=\begin{cases}0, & \mathbb{X}_1,\\ H(n_1-1,n_2,u_1,u_2), & \mathbb{X}_2,\mathbb{X}_3,\mathbb{X}_4,\mathbb{X}_5,\end{cases}$$

$$\mathbf{D_2}\mathbf{H}(u_1,u_2)=\begin{cases}0, & \mathbb{X}_1,\\ H(n_1,n_2-1,u_1,u_2), & \mathbb{X}_2,\mathbb{X}_3,\mathbb{X}_4,\mathbb{X}_5,\end{cases}$$

here $\mathbb{X}_1,\ldots,\mathbb{X}_5$ are subspaces of $\mathbb{X}$ defined as follows:

$$\mathbb{X}_1=\{(n_1,n_2,i_1,i_2):x_1n_1+n_2x_2=0\},$$

$$\mathbb{X}_2=\{(n_1,n_2,i_1,i_2):[(n_1+1)x_1+n_2x_2\leqslant R]\cap[n_1x_1+(n_2+1)x_2\leqslant R]\},$$

$$\mathbb{X}_3=\{(n_1,n_2,i_1,i_2):[(n_1+1)x_1+n_2x_2>R]\cap[n_1x_1+(n_2+1)x_2>R]\},$$

$$\mathbb{X}_4=\{(n_1,n_2,i_1,i_2):[(n_1+1)x_1+n_2x_2\leqslant R]\cap[n_1x_1+(n_2+1)x_2>R]\},$$

$$\mathbb{X}_5=\{(n_1,n_2,i_1,i_2):[(n_1+1)x_1+n_2x_2>R]\cap[n_1x_1+(n_2+1)x_2\leqslant R]\}.$$

We introduce the operator $\mathbf{E}$, which sums the equations over all possible values of n_1 and n_2, then given that

$$\mathbf{E}(\mathbf{A}+\lambda_1\mathbf{B_1}+\lambda_2\mathbf{B_2})=0,\quad \mathbf{E}(\mathbf{C_1}-\mathbf{D_1})=0,\quad \mathbf{E}(\mathbf{C_2}-\mathbf{D_2})=0,$$

we obtain an additional scalar equation

$$\mathbf{E}(\lambda_1\mathbf{B_1}(e^{ju_1}-1)+\lambda_2\mathbf{B_2}(e^{ju_2}-1))\mathbf{H}(u_1,u_2)+$$

$$+j\sigma_1\mathbf{E}\mathbf{D_1}(1-e^{-ju_1})\frac{\partial\mathbf{H}(u_1,u_2)}{\partial u_1}+j\sigma_2\mathbf{E}\mathbf{D_2}(1-e^{-ju_2})\frac{\partial\mathbf{H}(u_1,u_2)}{\partial u_2}=0. \quad (2)$$

3. First-Order Asymptotic Analysis

We solve equations (1)–(2) by the asymptotic analysis method under the condition of proportionately increasing delay in orbits [4], denote $\sigma_k=\sigma\cdot\gamma_k$, $\sigma\to 0$. We substitute in the system (1)–(2)

$$\sigma=\varepsilon,\quad u_k=\varepsilon w_k,\quad \mathbf{H}(u_1,u_2)=\mathbf{F}(w_1,w_2,\varepsilon).$$

Then, limiting $\varepsilon\to 0$, using the Taylor series for exponent $e^{j\varepsilon w_k}=1+j\varepsilon w_k+o(\varepsilon)$ and $\lim\limits_{\varepsilon\to 0}\mathbf{F}(w_1,w_2,\varepsilon)=\mathbf{F}(w_1,w_2)$, we can write

$$(\mathbf{A}+\lambda_1\mathbf{B_1}+\lambda_2\mathbf{B_2})\mathbf{F}(w_1,w_2)+j\gamma_1(\mathbf{C_1}-\mathbf{D_1})\frac{\partial\mathbf{F}(w_1,w_2)}{\partial w_1}+$$

$$+j\gamma_2(\mathbf{C_2}-\mathbf{D_2})\frac{\partial\mathbf{F}(w_1,w_2)}{\partial w_2}=0, \quad (3)$$

$$\mathbf{E}\left(w_1\lambda_1\mathbf{B_1}+w_2\lambda_2\mathbf{B_2}\right)\mathbf{F}(w_1,w_2)+j\gamma_1w_1\mathbf{E}\mathbf{D_1}\frac{\partial\mathbf{F}(w_1,w_2)}{\partial w_1}+$$

$$+j\gamma_2 w_2 \mathbf{E}\mathbf{D_2}\frac{\partial \mathbf{F}(w_1,w_2)}{\partial w_2} = 0. \tag{4}$$

We find the solution of (5)–(6) in the form: $\mathbf{F}(w_1,w_2) = \mathbf{R}\cdot\exp\{jw_1\varkappa_1 + jw_2\varkappa_2\}$, where $\mathbf{R}$ is a matrix of $R(n_1,n_2)$ is a stationary joint probability distribution of the two-dimensional process $\{n_1(t), n_2(t)\}$, and $\varkappa_1$, $\varkappa_2$ are normalized means of numbers of customers in orbits. From the operator equation (5), we obtain the operator equation

$$[(\mathbf{A} + \lambda_1\mathbf{B_1} + \lambda_2\mathbf{B_2}) - \varkappa_1\gamma_1(\mathbf{C_1} - \mathbf{D_1}) - \varkappa_2\gamma_2(\mathbf{C_2} - \mathbf{D_2})]\mathbf{R} = 0$$

with normalization condition $\mathbf{E}\mathbf{R} = 1$, and from the scalar equation (6), we obtain

$$\mathbf{E}[\lambda_1\mathbf{B_1} - \varkappa_1\gamma_1\mathbf{D_1}]\mathbf{R} = 0, \quad \mathbf{E}[\lambda_2\mathbf{B_2} - \varkappa_2\gamma_2\mathbf{D_2}]\mathbf{R} = 0.$$

Let us make inverse substitutions, we obtain a first-order approximation of the matrix of characteristic functions of a two-dimensional stochastic process $\{i_1(t), i_2(t)\}$, which coincides with the characteristic function of a deterministic variable and allows us to calculate the means of the number of customers in orbits:

$$\mathbf{H}(u_1,u_2) \approx \mathbf{R}\cdot\exp\left\{ju_1\frac{\varkappa_1}{\sigma} + ju_2\frac{\varkappa_2}{\sigma}\right\}.$$

4. Numerical Example

Let us consider the numerical example to obtain the applicability area of approximation. Let the system parameters be defined as follows

$$\lambda_1 = 3, \lambda_2 = 2, \mu_1 = 2, \mu_2 = 3, \gamma_1 = 3, \gamma_2 = 1, x_1 = 1, x_2 = 2, R = 5.$$

For different values of the approximation parameter σ, we calculate the means of the number of customers in the orbits analytically and using simulation. As a measure of accuracy, we consider the relative error. Table 1 shows the corresponding relative error.

Table 1

Relative error of means of number of customers in the orbits

σ	0.005	0.01	0.05	0.07	0.1	0.3	0.5
Δ_1	**0.001**	**0.007**	**0.029**	**0.045**	**0.061**	0.153	0.231
Δ_2	**0.001**	**0.004**	**0.018**	**0.026**	**0.038**	0.101	0.156

We observe that with a decrease in σ, the accuracy of the approximation increases, and can be applied, for example, with $\sigma < 0.1$ and a threshold relative error of $\Delta = 0.1$.

5. Conclusion

In this paper, the resource multi-server queue with two arrival processes was considered. The first-order asymptotic analysis method was carried out. We found the analytical expressions for the means of the number of customers in the orbits. Also, we found the stationary distribution of the two-dimensional stochastic process of numbers of customers in the service block, which allows calculating the main probability and numerical characteristics, such as means of the number of customers in the service; means total occupied resource amounts; the probability, that customers go to the orbit. We considered the numerical example, which makes it possible to evaluate the system's performance. In the future, it is planned to consider second-order asymptotic analysis for a full analysis of this system and comparison with other strategies for providing access to two types of traffic to one base station.

REFERENCES

1. *Bin Sun, Dudin S., Dudina O., Dudin A.* A Customer Service Model in an Adaptive-Modulation Mobile Communication Cell with Allowance for Random Environment // Automation and Remote Control. 2021. Vol. 82. № 5. P. 812–826. DOI: 10.1134/s0005117921050064.
2. *Tafintsev N. et al* Handling Spontaneous Traffic Variations in 5G+ via Offloading Onto mmWave-Capable UAV "Bridges" // IEEE Transactions on Vehicular Technology. 2021. Vol. 69. № 9. P. 10070-10084. DOI: 10.1109/TVT.2020.3005253.
3. `digital.gov.ru` — Ministry of Digital Development, Communications and Mass Media of the Russian Federation. 2021.
4. *Nazarov A., Phung-Duc T., Izmailova Y.* Multidimensional Central Limit Theorem of the Multiclass M/M/1/1 Retrial Queue // Lecture Notes in Computer Science. 2020. Vol. 12563. P. 298–310. DOI: 10.1007/978-3-030-66471-8_23.

Nazarov Anatoly — Doctor of Technical Science, Professor of Probability Theory and Mathematical Statistics Department. E-mail: *nazarov.tsu@gmail.com*

Salimzyanov Radmir — Student of Institute of Applied Mathematics and Computer Science. E-mail: *rsalimzyanov@yahoo.com*

Lisovskaya Ekaterina — Candidate of Physical and Mathematical Science, Associate Professor of Probability Theory and Mathematical Statistics Department. E-mail: *ekaterina_lisovs@mail.ru*

AN EXPLICIT SOLUTION FOR AN INVENTORY MODEL WITH POSITIVE LEAD TIME AND SERVER INTERRUPTIONS

Sandhya E.[1], C. Sreenivasan[2], Sajeev S. Nair[3], Rajan M.P.[3]

[1] *Government College, Chittur, University of Calicut, India*
[2] *Government Victoria College, Palakkad, India*
[3] *Government Engineering College, Thrissur, India*

A single server queuing system with inventory is considered. Customers arrive according to a Poisson process and service times follow exponential distribution. Inventory is replenished according to (s, S) policy with positive lead time which follows exponential distribution. Interruption to service process and repair of interrupted service are considered, times between two interruptions and repairs both follow exponential distributions.
Keywords: *(s,S) inventory model, server interruptions, positive lead time, explicit solution.*

Introduction

The pioneers in the study of queuing inventory models are Melikov and Molchano [1] and Sigman and Simchi-Levi 1992 [2]. In Sigman and Simchi-Levi customers are allowed to join even when there is no inventory in the system. They also discuss the case of non exponential lead time distribution. Later Berman and et al. [1] considered an inventory system where a processing time is required for serving the inventory. Here they considered deterministic service time and the model was discussed as a dynamic programming model. Berman and Kim [2] and Berman and Sapna [3] later discussed inventory queuing systems with exponential service time distribution and with arbitrary distribution.

There are several papers on inventory queuing models by Krishnamoorthy and his co-authors [4, 5, 6, 7, 8, 9, 10]. They mainly used Matrix Analytic Methods to study these models. In most of the models service time for providing the inventoried item is assumed.

1. Mathematical Model

The system under consideration is described as below. There is a single server counter where inventory is served to which customers arrive for service. The number of arrivals by time t follows a Poisson process with parameter λ. The service times are independently and identically distributed exponential random variables with parameter μ. Inventory is replenished according to (s, S) policy, in the sense that whenever inventory level drops to s an order is placed, order quantity being fixed as $Q = S - s$. The replenishment times follow exponential distribution with parameter η. While a customer is being served by the server, the service may be interrupted, the interruption rate being exponential with rate δ_1. Following a service interruption the service restarts at an exponential rate δ_2.

We make the following assumptions for the model under consideration.

— There is no loss of inventory due to a service interruption.
— The customer being served when interruption occurs waits there until his service is completed.
— No arrival is entertained when the inventory level is zero and while the server is on interruption.
— An order placed if any is cancelled while the server is on interruption.

We denote by $N(t)$ the number of the customers in the system including the one being served (if any), $L(t)$ the inventory level and $S(t)$ the server status at time t. Let $S(t) = 0, 1, 2$; if the server is idle, busy, on interruption.

Then $\Omega = X(t) = ((N(t), S(t), L(t))$ will be a Markov chain. The state space of this Markov chain can be described as $E = \{(0, 0, k) : 0 \leqslant k \leqslant S\} \cup \{(i, 0, 0) : i \geqslant 1\} \cup \{(i, j, k) : i \geqslant 1, j = 1, 2; 1 \leqslant k \leqslant S\}$. The above state space can be partitioned into levels $L(i)$ where $L(0) = ((0, 0, 0), (0, 0, 1), \ldots, (0, 0, S))$ and $L(i) = ((i, 0, 0), (i, 1, 1), (i, 1, 2), \ldots, (i, 1, S), (i, 2, 1), (i, 2, 2), \ldots, (i, 2, S))\,; i \geqslant 1$.

The Markov chain Ω described above is a level independent quasi birth death process whose infinitesimal generator matrix is given by

$$T = \begin{bmatrix} B_0 & B_1 & 0 & 0 & . & . & . \\ B_2 & A_1 & A_0 & 0 & 0 & . & . \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & . \\ & & & . & . & . & \\ & & & & . & . & . \end{bmatrix}$$

Here B_0, B_1, B_2 are matrices of orders $(S+1)\times(S+1), (S+1)\times(2S+1)$ and $(2S+1)\times(S+1)$ respectively. All other matrices are square matrices of order $2S+1$. The different transitions in the Markov chain $\Omega = X(t) = ((N(t), S(t), L(t))$ are given below.

— Due to arrival $(i, j, k) \xrightarrow{\lambda} (i+1, j, k)$; $i \geqslant 0$, $0 < k \leqslant S$, $j = 0, 1$.

— Due to service completions

$$(i, j, k) \xrightarrow{\mu} (i-1, j, k-1);\ i > 0,\ 0 < k \leqslant S,\ j = 1.$$

— Due to replenishments $(i, j, k) \xrightarrow{\eta} (i, j, k+Q)$; $i \geqslant 0$, $0 \leqslant k \leqslant S$, $j = 0, 1$.

— Due to server interruptions $(i, 1, k) \xrightarrow{\delta_1} (i, 2, k)$; $i \geqslant 1$, $0 < k \leqslant S$.

— Due to restart of service $(i, 2, k) \xrightarrow{\delta_2} (i, 1, k)$; $i \geqslant 1$, $0 < k \leqslant S$.

2. Analysis of the Model

Stability condition. Define $A = A_0 + A_1 + A_2$ and

$$\pi = (\pi(0,0), \pi(1,1), \pi(1,2), \ldots, \pi(1,S), \pi(2,1), \pi(2,2), \ldots, \pi(2,S))$$

be the steady state vector of A. We know the QBD process with generator matrix T is stable if and only if $\pi A_0 e < \pi A_2 e$ (see Nuets). That is if and only if $\lambda\,[\pi(1,1) + \pi(1,2) + \ldots + \pi(1,S)] \ < \ \mu\,[\pi(1,1) + \pi(1,2) + \ldots + \pi(1,S)]$, that is if and only if $\lambda < \mu$.

Thus we have the following theorem for the stability of the system under study.

Theorem 1. The Markov chain is stable if and only if $\lambda < \mu$.

3. Computation of steady state vector

We first consider a system identical to the above system except for service time is negligible. For this system $\tilde{\Omega} = \tilde{X}(t) = (S(t), L(t))$ will be a Markov chain where $S(t)$ and $L(t)$ are as defined for the original system. The state space of this Markov chain can be described as $\tilde{E} = \{(0,0), (1,1), (1,2) \ldots, (1,S), (2,1), (2,2), \ldots, (2,S)\}$. The infinitesimal generator matrix of the process is given by $\tilde{T} = \begin{bmatrix} \tilde{B}_0 & \tilde{B}_1 \\ \tilde{B}_2 & \tilde{B}_3 \end{bmatrix}$, where $\tilde{B}_1 = \begin{bmatrix} 0 \\ \delta_1 I_s \end{bmatrix}_{(S+1)\times S}$, $\tilde{B}_2 = \begin{bmatrix} 0 & \delta_2 I_s \end{bmatrix}_{S\times(S+1)}$, $\tilde{B}_3 = -\delta_2 I_s$, $\tilde{B}_0 = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}$.

Here $C_1 = \begin{bmatrix} -\eta & 0 \\ 0 & -(\lambda+\eta+\delta_1) I_{s-1} \end{bmatrix}_{(s+1)\times(s+1)} + \begin{bmatrix} 0 & 0 \\ \lambda I_{s-1} & 0 \end{bmatrix}_{(s+1)\times(s+1)}$,

$C_4 = -(\lambda+\delta_1) I_Q + \begin{bmatrix} 0 & 0 \\ \lambda I_{Q-1} & 0 \end{bmatrix}_{Q\times Q}$, $C_3 = \begin{bmatrix} 0 & \lambda \\ 0 & 0 \end{bmatrix}_{Q\times(s+1)}$,

$C_2 = \begin{bmatrix} 0 & \eta I_{s+1} \end{bmatrix}_{(s+1)\times Q}$.

Let $x = (x(0,0), x(1,1), \ldots, x(1,S), x(2,1), \ldots, x(2,S))$ be the steady state probability vector of the process $\tilde{\Omega}$. Then $x\tilde{T} = 0$ and $xe = 1$ gives

$$x(1,i) = \frac{\eta}{\lambda}\left(\frac{\eta+\lambda}{\lambda}\right)^{i-1} x(0,0)\,;\ 1 \leqslant i \leqslant s+1;$$

$$x(1,s+1)=x(1,s+2)=\ldots=x(1,Q)\,;x(1,Q+i)=x(1,Q)-x(1,i)\,;$$
$$1\leqslant i\leqslant s, x(2,i)=\frac{\delta_1}{\delta_2}x(1,i)\,;\ 1\leqslant i\leqslant S\,;$$
$$\text{where } x(0,0)=\left[1+Q\frac{\eta}{\lambda}\left(\frac{\eta+\lambda}{\lambda}\right)^s\left(\frac{\delta_1+\delta_2}{\delta_2}\right)\right]^{-1}.$$

Let $\pi=(\pi_0,\pi_1,\pi_2,\ldots)$ be the steady state probability vector of the process Ω, where $\pi_0=(\pi(0,0,0),\pi(0,0,1),\ldots,\pi(0,0,S)$ and $\pi_i=(\pi(i,0,0),\pi(i,1,1),\pi(i,1,2),\ldots,\pi(i,1,S),\pi(i,2,1),\pi(i,2,2),\ldots,\pi(i,2,S))$; $i\geqslant 1$. Then π satisfies $\pi T=0$ and $\pi e=1$. We have the equations

$$\pi_0B_0+\pi_1B_2=0\ ;\ \pi_0B_1+\pi_1A_1+\pi_2A_2=0\ ;$$
$$\pi_iA_0+\pi_{i+1}A_1+\pi_{i+2}A_2=0;\ i\geqslant 1.$$

All the above equations are satisfied by taking

$$\pi_0=\zeta(x(0,0),x(1,1),x(1,2),\ldots,x(1,S))$$
$$\pi_i=\zeta\left(\frac{\lambda}{\mu}\right)^i\Big(x(0,0),x(1,1),x(1,2),\ldots,$$
$$x(1,S),\frac{\delta_1}{\delta_2}x(1,1),\frac{\delta_1}{\delta_2}x(1,2),\ldots,\frac{\delta_1}{\delta_2}x(1,S)\Big);\ i\geqslant 1.$$

The value of ζ is obtained from $\pi e=1$ as $\zeta=\dfrac{(\mu-\lambda)\delta_2}{\delta_2\mu+\delta_1\lambda[1-x(0,0)]}$.

4. System Performance Measures

The system performance measures such as expected number of customers in the system (L_S), expected inventory level, expected rate of ordering (E_{OR}), expected replenishment rate (REP_{mean}), expected interruption rate (INT_{mean}), loss rate of customers, probability that server is busy ($P(busy)$) and probability that server is on interruption ($P(int)$) could be calculated explicitly. Mean waiting time of a customer in the system ($WAIT_{mean}$) and variance of waiting time ($WAIT_{var}$) are calculated numerically.

Conclusion

We studied a single server queuing model with positive service time, positive lead time and service interruptions. We could arrive at an explicit expression for the steady state probability vector. We wish to extend this model by considering retrials as well.

Table 1

Effect of service rate on various performance measures
$\lambda = 3 \qquad \eta = 2 \qquad \delta_1 = 6 \qquad \delta_2 = 7$

μ	$P(busy)$	$P(int)$	INT_{mean}	REP_{mean}	L_s	E_{OR}	$WAIT_{mean}$	$WAIT_{var}$
9	0.2999	0.0999	8.5713	0.3856	0.5999	0.3856	0.2191	0.1256
9.2	0.2941	0.098	8.5713	0.3865	0.5819	0.3865	0.2118	0.1178
9.4	0.2884	0.0961	8.5713	0.3873	0.5649	0.3873	0.205	0.1107
9.6	0.2829	0.0943	8.5713	0.388	0.5489	0.3881	0.1986	0.1043
9.8	0.2777	0.0926	8.5713	0.3888	0.5337	0.3888	0.1927	0.0984
10	0.2727	0.0909	8.5713	0.3896	0.5195	0.3895	0.187	0.0931

Table 2

Effect of interruption rate on various performance measures
$\lambda = 3 \qquad \mu = 9 \qquad \eta = 2 \qquad \delta_2 = 7$

δ_1	$P(busy)$	$P(int)$	INT_{mean}	REP_{mean}	L_s	E_{OR}	$WAIT_{mean}$	$WAIT_{var}$
6	0.2726	0.1818	8.5714	0.3506	0.6818	0.3506	0.2918	0.2219
6.2	0.2743	0.1769	8.5714	0.3526	0.6769	0.3526	0.2862	0.2128
6.4	0.2758	0.1723	8.5714	0.3546	0.6723	0.3546	0.281	0.2044
6.6	0.2772	0.168	8.5714	0.3564	0.668	0.3564	0.2761	0.1968
6.8	0.2786	0.1639	8.5714	0.3582	0.6639	0.3582	0.2716	0.1898
7	0.2799	0.1599	8.5714	0.3599	0.6599	0.3599	0.2673	0.1833

Table 3

Effect of re orer level on various performance measures
$\lambda = 3 \qquad \mu = 9 \qquad \eta = 2 \qquad \delta_1 = 7$

s	$P(busy)$	$P(int)$	INT_{mean}	REP_{mean}	L_s	E_{OR}	$WAIT_{mean}$	$WAIT_{var}$
5	0.2727	0.1818	8.5714	0.3506	0.6818	0.3506	0.2918	0.2219
5.2	0.2727	0.1818	8.5714	0.3506	0.6818	0.3506	0.2918	0.2219
5.4	0.2727	0.1818	8.5714	0.3506	0.6818	0.3506	0.2918	0.2219
5.6	0.2727	0.1818	8.5714	0.3506	0.6818	0.3506	0.2918	0.2219
5.8	0.2727	0.1818	8.5714	0.3506	0.6818	0.3506	0.2918	0.2219
6	0.2727	0.1818	9.071	0.409	0.6818	0.4087	0.2916	0.2215

REFERENCES

1. *Melikov A.Z., Molchano A.A.* Stock optimization in transport/storage systems // Cybernetics. 1992. № 27(3). P. 484–487.
2. *Sigman K., Simchi-Levi D.* Light traffic heuristic for a_n M/G/1 queue with limited inventory // Ann Oler. Res. 1992. № 40. P. 371–380.

3. *O. Berman, E. Kim, D.G. Shimshack.* Deterministic approximations for inventory management at service facilities // IEEE Transactions. 1993. № 5. P. 98–104.
4. *O. Berman, E. Kim.* Stochastic models for inventory management at service facilities, Communication in Statistics. // Stochastic Models. 1999. № 4. P. 695–718.
5. *O. Berman, K.P. Sapna.* Inventory management at service facilities for systems with arbitrarily distributed service times // Stochastic Models. 2000. № 3–4. P. 343–360.
6. *T.G. Deepak, A. Krishnamoorthy, C.N. Viswanath, K. Vineetha.* Inventory with service time and transfer of customers and inventory // Annals of Operations Research. 2008. Vol. 160. P. 191– 213.
7. *A. Krishnamoorthy, T.G. Deepak, C.N. Viswanath, K. Vineetha.* Effective utilization of idle time in an (s, S) inventory with positive service time // Journal of Applied Mathematics and Stochastic Analysis. 2006. Vol. 13.
8. *A. Krishnamoorthy, M.E. Islam and C.N. Viswanath.* Retrial inventory with Batch Markovian Arrival and Positive ServiceTime // Stochastic Modelling and Applications. 2006. Vol. 9. № 2. P 38–53.
9. *A. Krishnamoorthy, K.P. Jose, C.N. Viswanath.* Numerical investigation of a $PH/PH/1$ Inventory system with Positive service time and Shortage // Neural, Parallel & Scientific Computations. 2008. Vol. 16. № 4.
10. *A. Krishnamoorthy, S.N. Sajeev, C.N. Viswanath.* Inventory Model with Retrial and Orbital Search // Bulletin of Kerala Mathematical Association. 2009. Vol. 4765.
11. *A. Krishnamoorthy, C.N. Viswanath, T.G. Deepak, K. Vineetha.* Control Policies for Inventory with Service Time // Stochastic analysis and Applications. 2006. Vol. 24. № 4. P. 889–899.
12. *A. Krishnamoorthy, C.N. Viswanath, M.E. Islam.* On a Production Inventory with Service Time and retrial of Customers // Proceedings of the 11th International Conference on Analytical and Stochastic Modelling Techniques and Analysis. 2004. P. 238–247.

Sandhya E. — Research Scholar, Department of Mathematics. E-mail: *esandhya1729@gmail.com*

C. Sreenivasan — Department of Mathematics. E-mail: *sreenikoonathara@gmail.com*

Sajeev S. Nair — Department of Mathematics. E-mail: *sajeev@gectcr.ac.in*

Rajan M.P. — Department of Mathematics.

APPROXIMATE ANALYSIS OF A NETWORK OF RESOURCE LOSS SYSTEMS

E. Sopin[1], V. Begishev[1], D. Moltchanov[2], A. Samuylov[1]

[1] *Peoples' Friendship University of Russia (RUDN University), Russian Federation*
[2] *Tampere University, Finland*

The paper considers a network of resource loss systems (ReLS) with random resource requirements. There are two types of nodes, and customers from the first type nodes are rerouted to the second type node for an exponentially distributed time and then return back to the original node. Customers require random volume of limited resources, so if there are not enough unoccupied resources in the system upon the arrival of a customer, then it is lost. Similarly, if an accepted customer is rerouted to another node and finds that there are not enough resources to meet its requirements, then it is also lost. In this paper, we provide an approach to analyze the stationary behavior of the considered system, as well as formulas for the new customer loss probability and the accepted customer loss probability.

Keywords: *Resource loss system, queuing network, loss probability, wireless network*

Introduction

In addition to much higher propagation losses in millimeter wave (mmWave) band, fifth generation (5G) New Radio (NR) systems will be subject to outages caused by a dynamic blockage by human bodies [1]. This behavior heavily affects quality of service characteristics provided to users and may even cause service interruptions. To alleviate the effect of a dynamic blockage bandwidth reservations technique has been proposed in the past. However, this approach can only be utilized when users do not experience outage conditions in case of blockage. Alternatively, one may use recently standardized 3GPP multiconnectivity functionality [2, 3] to reroute sessions dynamically between locally available 5G NR base stations. Performance characterization of this mechanism naturally calls for queuing network formalism.

The research was funded by the Russian Science Foundation, project no. 21-79-10139, https://rscf.ru/en/project/21-79-10139/

The first work that analyzes a network of resource loss systems (ReLS) is [4]. In this paper, we describe the considered system in more details, define additional performance metrics and provide an iterative approximate algorithm for their evaluation.

1. Model description

We consider a network of resource loss systems (ReLS) with two types of nodes. There are $N-1$ of first type of nodes and one node of the second type. Each node has K_i servers and R_i resources, $i = 1, 2, ..., N$ (see figure 1. Customers arrive according to the Poisson process with intensities λ_1 and λ_2 to the first and second type nodes, respectively. The service times are exponentially distributed with parameters μ_1 and μ_2. Each customer requires not only a free server, but also a random discrete volume of resources, which are determined according to the probability distributions $\{f_{1,j}\}$ and $\{f_{2,j}\}$. Besides, each customer on first type nodes generates a Poisson flow of signals with intensity α that cause rerouting of the customers to the node N. Rerouted customers stay at node N for exponentially distributed time with parameter β and return back to its original node, or leave the system if its service is completed before.

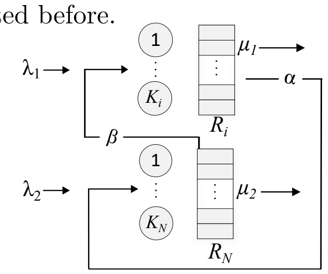

Figure 1. Illustration of the considered queuing system.

We follow the decomposition approach to address this model, which is a powerful methodology for queuing networks [6]. The core assumption here is that the service process at each node in the network is independent of the service processes at other nodes. The relation between the service processes at the nodes of both types is incorporated into the numerical solution algorithm, where the characteristics of the entire network are calculated iteratively until the procedure converges. The stability properties of this class of models were analyzed in [5].

1.1. Service process at the first type nodes

First, consider the first type nodes. Due to the memoryless property of the exponential distribution, the residual service time of returning sessions is also exponential with the same parameter μ_1. Let φ_i be the intensity of the returning session arrivals at node i, $i = 1, 2, ..., N-1$. Then the total session arrival intensity to node i is thus $\lambda_1 + \varphi_i$ and the total intensity of departures is $\mu_1 + \alpha$. The stochastic behavior of the node i can be described by the Markov process $X_i(t) = \{\xi_i(t), \delta_i(t)\}$, where $\xi_i(t)$ is the number of customers at node i at time t and $\delta_i(t)$ is the total volume of occupied resources. Denote the stationary probabilities $q_{i,n}(r)$ as

$$q_{i,n}(r) = \lim_{t\to\infty} P\{\xi_i(t) = n, \delta_i(t) = r\}, \quad n = 0, 1, 2\ldots, K_i, \tag{1}$$
$$r = 0, 1, 2\ldots, R, \quad i = 1, 2, \ldots, N-1.$$

The process $X_i(t)$ describes a ReLS. According to [8, 7], the stationary distribution (1) is given by

$$q_{i,0} = \left(1 + \sum_{n=1}^{K_i} \frac{\rho_i^n}{n!} \sum_{r=0}^{R_i} f_{1,r}^{(n)}\right)^{-1}, \tag{2}$$

$$q_{i,n}(r) = q_{i,0}\frac{\rho_i^n}{n!} f_{1,r}^{(n)}, \quad n = 1, 2, \ldots, K_i, \tag{3}$$

where $\rho_i = (\lambda_1 + \varphi_i)/(\mu_1 + \alpha)$ and $f_{1,j}^{(n)}, j \geqslant 0$ is the n-fold convolution of pmf $\{f_{1,j}\}, j \geqslant 0$. Note that the probability $f_{1,j}^{(n)}$ can be interpreted as the probability that n sessions on a first type node totally occupy j resources.

1.2. Service process at the second type node

The behavior of the node N can also be described in terms of the queuing systems with random resource requirements. As at nodes $1, 2, ..., N-1$, there are also two types of arrivals: customers that arrive initially to the node N with the intensity λ_2 and customers that are rerouted from the first type nodes with intensity φ_N. However, in this case, the service times differ from each other: the service intensity of the initially arriving customers is μ_2, and for the rerouted customers it is $\mu_1 + \beta$. The arrival intensity φ_N for the rerouted customers is obtained by summing up all the rerouting intensities of the first type nodes.

$$\varphi_N = \sum_{i=1}^{N-1} (\lambda_1 + \varphi_i)\frac{\alpha}{\mu_1 + \alpha}. \tag{4}$$

In (4), the term $\alpha/(\mu_1 + \alpha)$ refers to the probability that a customer from a first type node is rerouted to node N before its service completion. The intensities φ_i, $i = 1, 2, ..., N-1$, of customers returning back to their original node have the following form:

$$\varphi_i = (\lambda_1 + \varphi_i)(1 - \pi_{i,1})\alpha/(\alpha + \mu_1)(1 - \pi_{N,1})\beta/(\beta + \mu_1), \tag{5}$$

where $\pi_{i,1}$ is the loss probability of arriving customers at node i, $i = 1, 2, ..., N$ and $\beta/(\beta + \mu_1)$ is the probability that a rerouted customer at node N returns to its original node before service completion.

Observe that (5) implies that the flow of rerouted customers at node i, $i = 1, 2, ..., N-1$ equals the fraction of the accepted flow, which was initially routed to the node N with the probability $\alpha/(\alpha+\mu_1)$, accepted by the node N with the probability $1-\pi_{N,1}$, and finally rerouted back with the probability $\beta/(\beta+\mu_1)$.

According to [7], the stationary distribution of the ReLS with two arrival flows can be represented in the following form:

$$q_{N,n}(r) = q_{N,0}\frac{\rho_N^n}{n!}f_{2,r}^{(n)}, \tag{6}$$

$$q_{N,0} = \left(1+\sum_{n=1}^{K_N}\frac{\rho_N^n}{n!}\sum_{r=0}^{R_N} f_{2,r}^{(n)}\right)^{-1}, \rho_N = \frac{\lambda_2}{\mu_2} + \frac{\varphi_N}{\mu_1+\beta}, \tag{7}$$

which can be evaluated numerically, see, e.g., [9].

1.3. Solution and Performance Metrics

After obtaining the stationary state probabilities for all the nodes, we may proceed with deriving the performance metrics. Recall that our solution is iterative by nature as one needs to add another layer of rerouted customers at each iteration until a parameter converges to its stable value with a given accuracy. The procedure is terminated once the required precision level is achieved. Particularly, at the first iteration, there are no rerouted customers, and thus $\varphi_i = 0$, $i = 1, 2, ..., N$. Then, the algorithm continues as follows:

1) Based on the system parameters, new customer loss probabilities $\pi_{i,1}$ at nodes $i = 1, 2, ..., N$ and arrival intensity of rerouted customers φ_N at node N are evaluated.
2) New values of φ_i, $i = 1, 2, ..., N-1$ are calculated according to (5) by substituting their previous values into the right-hand side; if the difference between the new and the previous value meets the required precision, the algorithm proceeds with 3). Otherwise, it returns to 1).
3) When φ_i, $i = 1, 2, ..., N-1$ converges to a stable value with the desired accuracy, all other performance metrics are evaluated.

The iterative solution outlined above requires new session drop probabilities. These can be evaluated as follows

$$\pi_{i,1} = 1 - q_{i,0}\sum_{n=0}^{K_i-1}\frac{\rho_i^n}{n!}\sum_{r=0}^{R_i} f_{1,r}^{(n+1)}, i = 1, 2, ..., N-1 \tag{8}$$

$$\pi_{N,1} = 1 - q_{N,0}\sum_{n=0}^{K_N-1}\frac{\rho_N^n}{n!}\sum_{r=0}^{R_N} f_{2,r}^{2,(n+1)}. \tag{9}$$

Calculation of the probability that an accepted customer is lost is more complex. Let us introduce the conditional probability Π_i, $i = 1, 2, ..., N-1$ that a customer that originally arrived and accepted at node i and is lost, given that it is rerouted, i.e.,

$$\Pi_i = \pi_{N,1} + (1 - \pi_{N,1})\beta/(\beta + \mu_1)\pi_{i,1}, \tag{10}$$

where the first term corresponds to the case of customer loss at the entering the node N, while the second term is the probability that the rerouted customer is accepted at node N but then lost upon its return to the original node due to insufficient resources.

The average number of accepted customers lost as a result of rerouting during a time interval of length T is $\alpha\tilde{N}_i\Pi_i T$, where $\tilde{N}_i$ is the mean number of customers at node $i = 1, 2, ..., N-1$. The mean number of customers that are accepted during the same time interval is $\lambda_1(1-\pi_{i,1})T$. Hence, the probability that a customer, which was initially accepted at the node i, was dropped is

$$\pi_{i,2} = \lim_{T\to\infty} \frac{\alpha\tilde{N}_i\Pi_i T}{\lambda_1(1-\pi_{i,1})T} = \frac{\alpha\tilde{N}_i\Pi_i}{\lambda_1(1-\pi_{i,1})}. \tag{11}$$

Conclusion

In this paper, we considered a network of resource loss systems. Using the decomposition approach, we derived formulas for single node characteristics, and proposed an iterative algorithm to evaluate the performance measures of the whole system, including the new customer loss probability and the accepted customer loss probability.

REFERENCES

1. *Gapeyenko M., Samuylov A., Gerasimenko M., Moltchanov D., Singh S., Akdeniz M., Aryafa E., Himayat N., Andreev S., and Koucheryavy Y.* On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios // IEEE Transactions on Vehicular Technology. 2017. Vol. 66. No 11. P. 10124–10138.
2. *Moltchanov D., Samuylov A., Lisovskaya E., Kovalchukov R., Begishev V., Sopin E., Gaidamaka Y., and Koucheryavy Y.* Performance Characterization and Traffic Protection in Street Multi-Band Millimeter-Wave and Microwave Deployments // IEEE Transactions on Wireless Communications. 2021. DOI: 10.1109/TWC.2021.3094225.
3. *Begishev V., Sopin E., Moltchanov D., Kovalchukov R., Samuylov A., Andreev S., Koucheryavy Y, and Samouylov K.* Joint Use of Guard Capacity and Multiconnectivity for Improved Session Continuity in Millimeter-Wave 5G NR Systems // IEEE Transactions on Vehicular Technology. 2021. Vol. 70. No 3. P. 2657–2672.

4. *Petrov V., Solomitckii D., Samuylov A., Lema M., Gapeyenko M., Moltchanov D., Andreev S., Naumov V., Samouylov K., Dohler M., and Koucheryavy Y.* Dynamic Multi-Connectivity Performance in Ultra-Dense Urban mmWave Deployments // IEEE Journal on Selected Areas in Communications. 2017. Vol. 35. No 9. P. 2038–2055.
5. *Ageev K., Sopin E., Chursin A., and Shorgin S.* The probabilistic measures approximation of a resource queuing system with signals // Proc. Distributed Computer and Communication Networks (DCCN-2021), Moscow: Lecture Notes in Computer Science, 2021, in print.
6. *Kuehn P.* Approximate analysis of general queuing networks by decomposition // IEEE Transactions on Communications. 1979. Vol. 27(1). P. 113–126.
7. *Vikhrova O.* About Probability Characteristics Evaluation in Queuing System with Limited Resources and Random Requirements // Discrete and Continuous Models and Applied Computational Science. 2017. Vol. 25(3). P. 209–216.
8. *Naumov V., Samuilov K., and Samuilov A.* On the Total Amount of Resources Occupied by Serviced Customers // Automation and Remote Control. 2016. Vol. 77. No 8. P. 1419–1427.
9. *Sopin E., Ageev K., Markova E., Vikhrova O., and Gaidamaka Y.* Performance Analysis of M2M Traffic in LTE Network Using Queuing Systems with Random Resource Requirements // Automatic Control and Computer Sciences. 2018. Vol. 52. No 5. P. 345–353.

Sopin Eduard — Candidate of Science in physics and mathematics, Associate Professor, Department of Applied Probability and Informatics. E-mail: *sopin-es@rudn.ru*

Begishev Vyacheslav — Candidate of Science in physics and mathematics, Associate Professor, Department of Applied Probability and Informatics. E-mail: *begishev-vo@rudn.ru*

Moltchanov Dmitri — Doctor of Science in technology, PhD, University Lecturer, Faculty of Information Technology and Communication Sciences. E-mail: *dmitri.moltchanov@tuni.fi*

Samuylov Andrey — Candidate of Science in physics and mathematics, Associate Professor, Department of Applied Probability and Informatics. E-mail: *samuylov-ak@rudn.ru*

ANALYSIS OF FINITE-SOURCE RETRIAL QUEUEING SYSTEMS WITH IMPATIENT CUSTOMERS AND CATASTROPHIC BREAKDOWNS USING SIMULATION

J. Sztrik, Á. Tóth, Á. Pintér, Z. Bács

University of Debrecen, Debrecen, Hungary

In this paper, a finite-source retrial queueing system is considered with impatient customers and catastrophic breakdowns. The characteristic of the system includes collision which occurs when a new job arrives in the system and the service facility is occupied with a job, they will collide. Both jobs will be forwarded to the virtual waiting room the so-called orbit. Here, the customers initiate other attempts to reach the server after a random time. But they give up retrying after staying in the orbit a while and leave the system which is the impatient attribute of the customers. In case of a negative event, a catastrophic breakdown takes place meaning that all the customers at the server and in the orbit depart from the system. The novelty of this paper is to investigate that feature in a collision environment with impatient customers using different distributions of the service time.

Keywords: *Simulation, catastrophic breakdown, retrial queuing system, collision, impatience, sensitivity analysis.*

Introduction

Designing info-communication systems are essential because of understanding how to optimize a system and also how to handle increasing network traffic. Many tools and mechanisms are available for modeling different systems, and among them, one of the most popular ones is retrial queuing systems. To illustrate real-life problems arising in main telecommunication systems, like telephone switching systems, call centers, computer networks, and computer systems, retrial queues can be effectively applied. In many publications, retrial-queuing systems with repeated calls are utilized to depict their models like in [1]. The specialty of retrial queuing systems relies on the orbit which is assumed to be a virtual waiting room with enough

The research was supported by the Thematic Excellence Programme (TKP2020-IKA-04) of the Ministry for Innovation and Technology in Hungary.

capacity to take in every customer. In this way, a job – whose service can not start – is not lost and may launch numerous attempts to get its service requirement. The source is considered to be finite mainly because in many situations a finite number of entities participate in the operation of the system. Impatient behaviour is a natural occurrence of the customers provoking earlier departure without obtaining its service demand. This phenomenon is experienced in many fields of life just to mention some examples: healthcare applications, call centers, telecommunication networks. Not to mention all the papers where the behaviour of impatience is intensively examined, see for example [2]. Real-life systems tend to be subjected to random breakdowns which can be caused by a power outage, human negligence, or other sudden act. Thus, it is important to examine its effect on the operation of the system and the performance measures because it alters significantly the behaviour of a model. Many papers have studied models having service units assumed to be available all the time which is quite unrealistic. These types of systems have been investigated by many authors for example in [3].

The objective of our investigation is to carry out a sensitivity analysis using different distributions of service times on the main performance measures while catastrophic breakdowns eventuate. In the case of these types of events, customers are forced to leave the system due to sudden acts which can be mechanical failures or power outages. Until repair, it is not allowed for any customer to enter the system and detailed studies on catastrophic breakdowns have been examined by several papers. Because we utilize different distributions for the service time of the customers the results are obtained by our simulation program that is based on Simpack [4]. The basic building blocks of the code are used in which we have the opportunity to calculate any desired measure using numerous values of input parameters. Graphical illustrations are provided depicting the effect of different parameters and distributions on the main performance metrics.

1. System model

A finite-source retrial queueing system of type $M/G/1//N$ is considered with an unreliable service unit, impatient customers, the appearance of collisions, and blocking. This model has one service unit and a finite-source where every individual (altogether N) may generate request towards the system according to exponential law with parameter λ/N meaning that the inter-arrival times are exponentially distributed with mean λ/N. As there are no queues the service of an arriving job starts immediately following gamma, hypo-exponential, hyper-exponential, Pareto, and lognormal distribution with different parameters but with the same mean and variance value. In the case of a busy server, an arriving customer brings about a collision with the customer under service, and both are moved into orbit. Jobs residing in the orbit after an exponentially random time with parameter σ/N initiate other tries to be engaged with the server. Since random breakdowns emerge the failure time is also an exponential random variable with parameter γ_0 when the server is occupied and with γ_1 if idle. Two scenarios are distinguished: general breakdown where the service of a job is interrupted and it is forwarded back to the orbit, other jobs initiated by the individuals of the source can not enter the system until the service unit is functional. The other one is catastrophic breakdown where the service of a job is interrupted but instead of arriving at the orbit it leaves the system as the others from the orbit, no customers are allowed by the system until the server fully recovers.

The repair process starts instantly upon the failure of the service unit which follows an exponential distribution with parameter γ_2. Customers are characterized by impatience implicating that jobs can decide to leave the system after spending an exponentially distributed time with parameter τ in the orbit. These requests return to the source being unserved. In paper of [5] similar models are analyzed by an asymptotic method where N tends to infinity this is why rates λ/N and σ/N are used.

2. Simulation results

To obtain the desired results, our self-developed simulation tool was used in which almost all the performance measures can be estimated. Its statistics package utilizes the batch means method where the useful run is divided into a certain number of batches. Batches are long enough in that way sample averages of the batches are approximately independent thus we have a valid estimation. The following article contains more information about that method [6].The simulations are performed with a confidence level of 99.9%. The relative half-width of the confidence interval required to stop the simulation run is 0.00001. The size of a batch used to detect the initial transient duration is 1000.

Table 1

Numerical values of model parameters

N	γ_0	γ_1	σ/N	τ
100	0.05	1	0.05	0.001

Table 1 consists of every parameter that is applied for all the following figures. The parameters of service time of the customers can be found at

Table 2, every chosen parameter is listed resulting in the same mean and variance in every used distribution. The reason for selecting these values is focusing on the interesting situations it must be noted that this model was tested with other values as well, and in most of the cases, the same phenomenon appeared. It is totally intentional that the squared coefficient of variation is more than one, in another scenario we will run the simulations when it is less than one.

Table 2

Parameters of service time of primary customers

Distribution	**Gamma**	**Hyper-exponential**	**Pareto**	**Lognormal**
Parameters	$\alpha = 0.054$ $\beta = 0.077$	$p = 0.473$ $\lambda_1 = 1.353$ $\lambda_2 = 1.5$	$\alpha = 2.027$ $k = 0.355$	$m = -1.839$ $\sigma = 1.722$
Mean	0.7			
Variance	9			
Squared coefficient of variation	18.367			

On Figure 1 and 2 on the X-axes i represents the number of customers located in the system, and on the Y-axes $P(i)$ denotes the probability that exactly i customer are situated at the server and in the orbit altogether. In both Figure 1 and 2 the distribution of the number of customers in the system is displayed when λ/N is 0.1 using various distributions of service time. Catastrophic breakdown feature is applied and interestingly the mean number of customers in the system differs from each other. In the case of the gamma distribution, customers tend to spend less time in the system compared to Pareto distribution. It is also noticeable that for both types of breakdowns the distribution of the number of customers tends to follow Gaussian distribution.

Figure 2 depicts the comparison of different failure modes besides gamma and hyper-exponential distributions. Naturally more customers are in the system using the general breakdown method but the shape of the curves curiously are slightly disparate. In case of catastrophic breakdown, the peak is not that high and the mean number is fewer but other than that curves follow the same tendency.

3. Conclusion

We simulated a retrial queueing system of type $M/G/1//N$ with impatient customers in the orbit and with an unreliable server using two different failure mechanisms when blocking is applied. Results are obtained by our program to carry out a sensitivity analysis on different performance measures like the distribution of the number of customers in the system. Un-

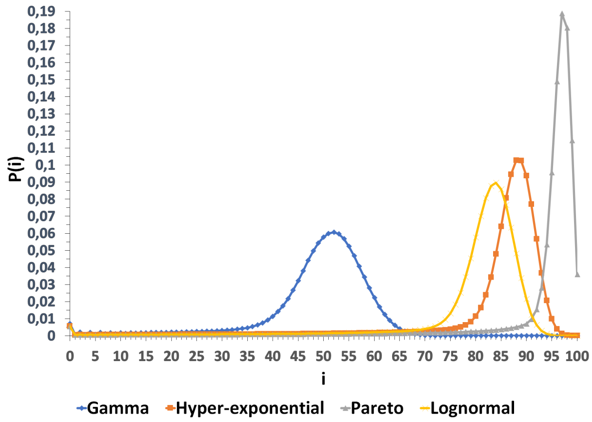

Figure 1. Distribution of the number of customers in the system

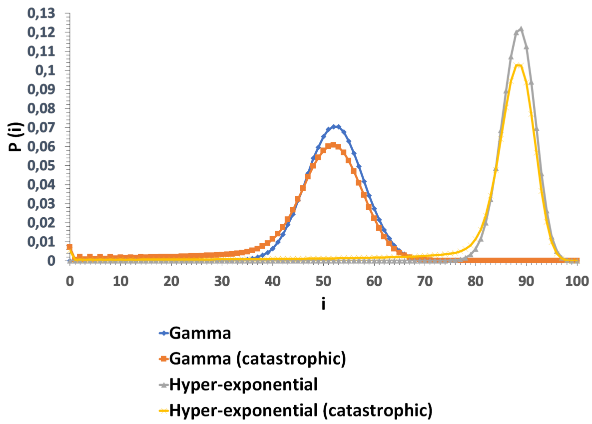

Figure 2. Comparison of distribution of the number of customers in the system using different failure modes

der various parameter settings, the most interesting measures were chosen which were graphically illustrated. When the squared coefficient of variation is more than one significant deviation is experienced between the distributions in almost every aspect of the investigated measures. Consistently, it was also revealed that besides catastrophic breakdown less customer is in the system than in the case of a normal breakdown which is an expected phenomenon but the shape of the curves follows the same tendencies. In future works, the authors aim to carry on investigating the effect of catastrophic breakdown in other models and performing sensitivity analysis for other variables like the failure rate or the impatience of the customers.

REFERENCES

1. *Fiems D., Phung-Duc T.* Light-traffic analysis of random access systems without collisions // Annals of Operations Research. 2019. Vol. 277. P. 311–327.
2. *Kumar R., Jain N., Som B.* Optimization of an $M/M/1/N$ Feedback Queue with Retention of Reneged Customers // Operations Research and Decisions. 2014. Vol. 24. P. 45–58.
3. *Tóth Á., Sztrik J., Pintér Á., Bács Z.* Reliability Analysis of Finite-Source Retrial Queuing System with Collisions and Impatient Customers in the Orbit Using Simulation // 2021 International Conference on Information and Digital Technologies (IDT), 2021, Zilina: IEEE, 2021. P. 230–234.
4. *Fishwick P.A.* SimPack: Getting Started With Simulation Programming In C And C++ // In 1992 Winter Simulation Conference, New York: Association for Computing Machinery, 1992. P. 154–162.
5. *Nazarov A., Sztrik J., Kvach A., Bérczes T.* Asymptotic analysis of finite-source M/M/1 retrial queueing system with collisions and server subject to breakdowns and repairs // Annals of Operations Research. 2019. Vol. 277. № 2. P. 213–229.
6. *Chen J.E., Kelton D.W.* A Procedure for Generating Batch-Means Confidence Intervals for Simulation: Checking Independence and Normality // SIMULATION. 2007. Vol. 83. № 10. P. 683–694.

János Sztrik — Doctor of the Hungarian Academy of Sciences, full professor, Department of Informatics Systems and Networks. E-mail: *sztrik.janos@inf.unideb.hu*

Ádám Tóth — PhD, assistant professor, Department of Informatics Systems and Networks. E-mail: *toth.adam@inf.unideb.hu*

Ákos Pintér — Doctor of the Hungarian Academy of Sciences, full professor, Institute of Mathematics. E-mail: *apinter@science.unideb.hu*

Zoltán Bács — PhD, full professor, Institute of Accounting and Finance. E-mail: *bacs.zoltan@econ.unideb.hu*

A PRODUCTION INVENTORY MODEL WITH N-POLICY

N. J. Thresiamma[1], K. P. Jose[2]

[1] *Govt. Women's Polytechnic College, Thrissur, Kerala, India*
[2] *Dept. of Mathematics, St. Peter's College, Kolenchery, Kerala, India*

This article shows effective inventory management by incorporating N-policy on an (s, S) production inventory system with service time. The customer's arrival pattern follows the Poisson distribution and the service time and production process follow exponential distributions. Matrix Geometric method is used to analyze the system. Optimal values of control variables s, S and N and cost function are obtained.

Keywords: *Production inventory, N-Policy, matrix geometric method, cost function.*

Introduction

In production and manufacturing systems, it is not always advisable to start the service as soon as the first customer arrives at a service station due to large switch-on cost of the server. In such cases, it is imperative that the resulting busy period should be long enough; otherwise the unit operating time cost may be too high compared to customer waiting cost. If we start the service only when a few customers, say N, have arrived then excessive frequent set-ups can be avoided. The so-called N-policy states that as soon as N customers arrive in the system, it starts service, one at a time, and the server goes for a vacation when the system becomes empty. The server will be turned on again only when the queue size reaches N.

Inventory with positive service time is first investigated by Berman et al. [1], where demands and service formed, are two distinct deterministic processes. Krishnamoorthy and Jose [3] analyzed and compared three production inventory systems with positive service time and retrial of customers. Jose and Beena [2] investigated production inventory model with server vacations and positive service time. The concept of N-policy was first introduced in 1963 by Yadin and Naor [8] in queueing literature to minimize the total operational cost in a cycle. Krishnamoorthy and Raju [4] used N-policy in (s, S) inventory system with lead time and negligible service time, involving perishable as well as non-perishable items. Krishnamoorthy et al. [5] were the first to introduce N-policy in (s, S) inventory system with positive service time. In this article, we introduce N-policy in a production inventory system.

1. Mathematical Modelling and analysis of the problem

Consider an (s, S) production inventory system with a single server. The production process is ON when the inventory level reaches s and it is switched OFF when it reaches S. The following assumptions and notations are used in this model.

Assumptions

— The arrival of customers follows Poisson distribution with parameter λ.
— The service pattern and production process follow exponential distributions with parameters μ and β respectively.
— The server is switched OFF when the system is empty and it is turned ON at the instant when there is N customers in the waiting line; and there is a positive on hand inventory.

Notations

$$N(t) : \text{Number of customers in the system at time } t$$
$$I(t) : \text{Inventory level at time } t.$$
$$C(t) : \begin{cases} 0 \text{ if server is idle at time t} \\ 1 \text{ if server is busy at time t} \end{cases}$$
$$J(t) : \begin{cases} 0 \text{ if the production is OFF mode} \\ 1 \text{ if the production is ON mode} \end{cases}$$

The random variable $Z(t) = \{(N(t), C(t), J(t), I(t)), t \geqslant 0\}$ is a QBD Process on the state space $S = \bigcup\limits_{i=0}^{\infty} L(i)$ and is independent for $i \geqslant N+1$, where

$$L(0) = \{(0,0,0,j); s+1 \leqslant j \leqslant S\} \bigcup \{0,0,1,j); 0 \leqslant j \leqslant S-1\},$$

$$L(i) = \{(i,0,0,j); s+1 \leqslant j \leqslant S\} \bigcup \{i,0,1,j); 0 \leqslant j \leqslant S-1\},$$

$$\bigcup \{i,1,0,j); s+1 \leqslant j \leqslant S \bigcup \{i,1,1,j); 1 \leqslant j \leqslant S-1, 1 \leqslant i \leqslant N-1,$$

$$L(i) = (i,0,1,0) \bigcup \{i,1,0,j); s+1 \leqslant j \leqslant S \bigcup \{i,1,1,j); 1 \leqslant j \leqslant S-1, i \geqslant N.$$

Arranging the states in the lexicographic order, infinitesimal generator G of the process $\{Z(t)|t \geqslant 0\}$ can be written as

$$G = \begin{array}{c} \underline{0} \\ \underline{1} \\ \underline{2} \\ \\ \underline{N\_1} \\ \underline{N} \\ \underline{N+1} \\ \\ \end{array} \begin{bmatrix} B_{00} & B_{01} & & & & & & \\ B_{10} & A_1^* & A_0^* & & & & & \\ & A_2^* & A_1^* & A_0^* & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & A_2^* & A_1^* & A_0^{**} & & \\ & & & & A_2^{**} & A_1 & A_0 & \\ & & & & & A_2 & A_1 & A_0 \\ & & & & & & \ddots & \ddots \; \ddots \end{bmatrix},$$

where the block matrices are as follows: $A_0 = \lambda I_{2S-s}$,

$$[A_1](pq) = \begin{cases} -(\lambda+\beta), \text{if } p = q = 1, \\ -(\lambda+\mu), \text{if } p = q,\ p = 2, 3, \dots, S-s+1, \\ -(\beta+\lambda+\mu), \text{if } p = q,\ p = S-s+2, \dots, 2S-s, \\ \beta, \text{if } p = 1,\ q = S-s+2, \\ \qquad q = p+1,\ p = S-s+2, \dots, 2S-s-1, \\ \qquad p = 2S-s,\ q = S-s+1, \\ 0, \text{otherwise}, \end{cases}$$

$$[A_2](pq) = \begin{cases} \mu, \text{if } p = 2,\ q = S+1, \\ \qquad q = p-1,\ p = 3, 4, \dots, S-s+1, \\ \qquad p = S-s+2,\ q = 1, \\ \qquad p = q-1,\ p = S-s+3, \dots, 2S-s, \\ 0, \text{otherwise}, \end{cases}$$

and one can also find other block matrices $B_{00}, B_{01}, B_{10}, A_0^*, A_1^*, A_2^*, A_2^{**}$.

2. Stability

Theorem 1. The process $\{Z(t)|t \geqslant 0\}$ is stable if and only if $\lambda < (1-\pi_1)\mu$, where $\pi_1 = \frac{(\frac{\mu}{\beta})^{s+1}(1-(\frac{\mu}{\beta})^{S-s})(1-\frac{\mu}{\beta})}{(S-s)(1-\frac{\mu}{\beta})+((\frac{\mu}{\beta})^{S+2}-(\frac{\mu}{\beta})^{s+2})}$.

Proof: Since the process $\{Z(t)|t \geqslant 0\}$ is a level independent QBD, it will be stable if and only if $\pi_A A_0 e < \pi_A A_2 e$ (see Neuts [7]).

3. The Steady State Probability vector of G

Let the steady state probability vector $\mathbf{x}$ of G can be partitioned according to the levels as $\mathbf{x} = (x(0), x(1), \dots, x(N-1), x(N), \dots)$, where

$x(i), 1 \leqslant i \leqslant N-1$ contain $4S-2s-1$ elements and all other sub vectors contains $2S-s$ elements. The QBD process $Z(t)$ is state independent for $i \geqslant N+1$. Therefore the steady state solution is of the form (see Latouche and Ramaswami [6]).

$$x_{N+1+j} = x_{N+1}R^j : j \geqslant 1,$$

where R is the minimal non-negative solution of the matrix quadratic equation $R^2A_2 + RA_1 + A_0 = 0$. R can be calculated from the iterative procedure $R_{n+1} = -(R_n^2A_2 + A_0)A_1^{-1}$ (refer Neuts [7]). Also $\mathbf{x}$ satisfies the equations $\mathbf{x}G = 0$ and $\mathbf{x}e = 1$.

The important system performance measures of the system are given by:

— Expected number of customers in the system

$$EC = \sum_{i=1}^{\infty} ix(i)e.$$

— Expected inventory level

$$EI = \sum_{j=s+1}^{S} jx(0,0,0,j) + \sum_{j=0}^{S-1} jx(0,0,1,j) +$$

$$+ \sum_{i=1}^{N-1} \left(\sum_{j=s+1}^{S} jx(i,0,0,j) + \sum_{j=0}^{S-1} jx(i,0,1,j) + \sum_{j=s+1}^{S} jx(i,1,0,j) + \right.$$

$$\left. + \sum_{j=1}^{S-1} jx(i,1,1,j) \right) + \sum_{i=N}^{\infty} \left(\sum_{j=s+1}^{S} jx(i,1,0,j) + \sum_{j=1}^{S-1} jx(i,1,1,j) \right).$$

— Expected number of items produced

$$EP = \beta \left\{ \sum_{j=0}^{S-1} x(0,0,1,j) + \sum_{i=1}^{N-1} \left(\sum_{j=0}^{S-1} x(i,0,1,j) + \sum_{j=1}^{S-1} x(i,1,1,j) \right) + \right.$$

$$\left. + \sum_{i=N}^{\infty} \left(x(i,0,1,0) + \sum_{j=1}^{S-1} x(i,1,1,j) \right) \right\}.$$

— Expected switching rate for production

$$ES_1 = \mu \sum_{i=1}^{\infty} x(i,1,0,s+1).$$

— Expected switching rate for service

$$ES_2 = \lambda \left(\sum_{j=s+1}^{S} x(N-1,0,0,j) + \sum_{j=0}^{S-1} x(N-1,0,1,j) \right).$$

— Expected number of departures after completing the service

$$ED = \mu \left(\sum_{i=1}^{\infty} \left(\sum_{j=1}^{S-1} x(i,1,1,j) + \sum_{j=s+1}^{S} x(i,1,0,j) \right) \right).$$

Cost function: The expected total cost ETC is computed as

$$ETC = c_0 ES_1 + c_1 EP + c_2 ES_2 + c_3 EI + +c_4 EC + c_5 ED,$$

where c_0: fixed cost for production,
c_1: production cost / item / unit time,
c_2: reward cost of customer when the server is idle / customer / unit time,
c_3: holding cost of inventory / unit / unit time,
c_4: holding cost of customer / customer / unit time,
c_5: cost of service / item / unit time.

Numerical Experiments. The optimum value of (s, S) pair and N are obtained by considering suitable parameter values.

Table 1

$N = 4, \lambda = 1.5, \mu = 2, \beta = 2, c_0 = 10, c_1 = 1, c_2 = 1, c_3 = 1, c_4 = 8, c_5 = 1$

S \ s	10	11	12	13	14	15
42	43.0833	43.0822	43.0815	**43.0812**	43.0812	43.0815
43	43.0816	43.0806	43.0800	**43.0797**	43.0798	43.0800
44	43.0804	43.0795	43.0790	**43.0788**	43.0788	43.0791
45	43.0798	43.0790	43.0785	**43.0784**	43.0784	43.0786
46	43.0803	43.0790	43.0787	**43.0785**	43.0786	43.0788
47	43.0816	43.0797	43.0794	**43.0792**	43.0793	43.0795

Concluding Remarks

In this article, an (s, S) production inventory system with positive service time is considered. The optimum values of s and S are obtained. The optimum value of N is also found graphically. This model can further be extended to a production inventory system with multiple servers.

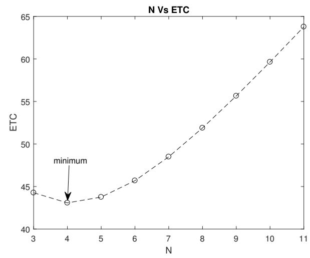

Figure 1. $c_0 = 10, c_1 = c_2 = c_3 = c_5 = 1, c_4 = 8, s = 13, S = 45, \lambda = 1.5, \mu = \beta = 2$

REFERENCES

1. *Berman O., Kaplan E.H., Shimshak D.G.* Deterministic approximations for inventory management at service facilities // IEEE Trans. 1993. Vol. 25. P. 98–104.
2. *Jose K.P., Beena P.* Modelling of an $M/M/2$ production inventory system with multiple vacation // Malaya Journal of Mathematics. 2020. Vol. 8. P. 356–362.
3. *Krishnamoorthy A., Jose K.P.* Three production inventory systems with service, loss and retrial of customers // Internat. J. Inform. Management Sci. 2015. Vol. 19, P. 367–389.
4. *Krishnamoorthy A., Raju N.* (s, S) inventory with lead time-the N Policy // Information and Management Sciences. 1998. Vol. 9. P. 42–52.
5. *Krishnamoorthy A., Viswanath C. Narayanan, Deepak T.G., Vineetha P.* Control Policies for Inventory with Service Time // Stochastic Analysis and Applications. 2006. Vol. 24. P. 889–899.
6. *Latouche G., Ramaswami V.* Introduction to matrix analytic methods in stochastic modelling // Philadelphia: SIAM.: publisher.
7. *Neuts M.F.* Matrix-Geometric Solutions in Stochastic Models – an Algorithmic Approach // John Hopkins University Press: publisher.
8. *Yadin M., Naor. P.* Queueing Systems with a Removable Service Station // Operations Research. 1963. Vol. 14. P. 393–405.

Thresiamma N.J. — Assistant Professor, Dept. of Mathematics.

Jose K.P. — Associate Professor, Dept.of Mathematics. E-mail: *kpjspc@gmail.com*

SYNERGISTIC EFFECTS IN QUEUING SYSTEMS AND RELATED STATISTICAL PROBLEMS

G. Tsitsiashvili

Institute for Applied Mathematics, Far Eastern Branch of Russian Academy Sciences, Vladivostok, Russia

In this paper, a method for detecting synergistic effects of the interaction of elements in multi-server queuing, in construction of queuing systems with large load and small queue. This method is extended to statistical estimates of characteristics of non-uniform Poisson flow, describing distribution of animals in some areas, a resolution of the most powerful decision rule for constructing of technical systems "friend – foe". This allows studying them with mathematical methods of relatively low complexity and to expand possibilities of applications. These methods are based on special techniques of the structural analysis of multi-element stochastic models in combination with upper bounds of their performance indicators. They allow moving to more accurate and rather economical numerical calculations, as they indicate, in which direction it is most convenient to perform these calculations.

Keywords: *Multi-server queuing system, Almost deterministic one-server queuing system, Most powerful decision rule.*

RQ-Queuing Systems with a Large Number of Servers

Consider an RQ-system, i.e., a queuing system, in which, if there is a free server, the customer that has come to the system immediately begins to be served on it. If there are no free servers, then the customer is sent to the orbit, from where it can return to the newly released server in accordance with some protocol [1, 2, 3]. At 12-th International Workshop on Retrial Queues and Related Topics (WRQ 2018) it is marked that in RQ n-channel queuing systems with Poisson input flow and exponentially distributed service times a complexity of its calculations grows significantly with increasing n. In this section we show, how to bypass this problem, using asymptotic analysis of n-channel queuing system with intensity of input flow proportional to n.

Consider the series scheme in which the characteristics of n-server queuing systems are defined by the parameter $n \to \infty$, which characterizes an intensity of input flow tending to infinity. Denote $e_n(t)$ a number of input flow customers arriving before the moment t, $e_n(0) = 0$. Assume that $q_n(t)$

is a number of busy servers in this system at the moment t, $q_n(0) = 0$, τ_j is the service time of j-th incoming customer and τ_j, $j \geqslant 1$, is a sequence of independent and identically distributed random variables (s.i.i.d.r.v.'s) with the distribution function, which has continuous and bounded density. Then in [4, Chapter II, § 1, Theorem 1] there are conditions, in which the relation

$$\Pi_n(T) = P\left(\sup_{0\leqslant t\leqslant T} q_n(t) \geqslant n\right) \to 0,\ n \to \infty,$$

takes place. Denote $P_n(T)$ the probability that there are customers, which will be directed to the orbit at time interval $[0, T]$, then $P_n(T) \leqslant \Pi_n(T)$.

In [5] the conditions of limit relation $\Pi_n(T) \to 0,\ n \to \infty$, were concreted for n-channel queuing systems with: deterministic input flow of customer groups, alternating input flow [6], Erlang input flow [7]. Combining these results with the inequality $P_n(T) \leqslant \Pi_n(T)$ and concentrating our attention to the indicator $P_n(T)$ it is possible to analyse RQ n-channel queuing systems with large n without large calculations.

1. Alternative designs of high load queuing systems with small queue

In this section, two alternative designs are constructed for queuing systems with a large load and a small queue. The first design is an aggregation of a large number of single-channel systems into a multi-channel system. The second design is based on the model of a single-channel system, in which small random fluctuations are defined as the degree of small difference between the unit and the load factor. The exponent of this degree has a critical value, above which the stationary waiting time tends to zero, and below which it tends to infinity. A similar phase transition was found in the multi-channel queuing system.

Multi-channel queuing system $M|M|n|\infty$. Consider n-channel system with a Poisson input flow of intensity $n\lambda$ and exponentially distribution of service times $1 - \exp(-\mu t)$. Such a system can be considered as a combination of n single-channel systems $M|M|1|\infty$ with Poisson input flows of λ intensity and a similar distribution of service times. Denote $\rho = \lambda/\mu$ load factor of the system $M|M|n|\infty$ and put A_n the stationary average waiting time, B_n the stationary average queue length. The following theorem is proved in [8].

Theorem 1. 1) If $\rho < 1$, then for some $c < \infty$, $q < 1$ the relation holds $A_n \leqslant c\, q^n$, $n \geqslant 1$, 2) If $\rho = 1 - n^{-\alpha}$, $0 < \alpha < \infty$, then for $n \to \infty$

$$A_n \to \begin{cases} 0, & \alpha < 1, \\ 1/\mu, & \alpha = 1, \\ \infty, & \alpha > 1. \end{cases} \qquad B_n \to \begin{cases} 0, & \alpha < 1/2, \\ \infty, & \alpha \geqslant 1/2. \end{cases}$$

Thus, the load factor tends to one as $\rho = 1 - n^{-\alpha}$ and depending on the parameter α, the stationary average waiting times and queue lengths tend either to zero or to infinity. Moreover, the critical value of the parameter α varies, depending on the choice of these performance indicators.

Almost deterministic single-channel queuing system. Despite the importance of Theorem 1, such a queuing system design assumes its large size, which is not always convenient from an application point of view. It is clear that an alternative to the described mode of operation of a queuing system with a large load and a small queue can serve as an almost deterministic one channel queuing system (see, for example [9]).

Let's describe the operation of a single-channel queuing system $G|G|1|\infty$ by a sequence of waiting times for the start of service: $w_{i+1} = \max(0, w_i + \eta_i - \tau_i)$. Here τ_i is the interval between the arrival of i-th and $(i+1)$-th customers, $M\tau_i = a$, and η_i – service time of i-th customer, $M\eta_i = b,\ 0 < a - b = \varepsilon$. Assume that random deviations from the distributions means are reduced as follows:

$$\eta_i^\varepsilon = b + \varepsilon^\alpha(\eta_i - b),\ \tau_i^\varepsilon = a + \varepsilon^\alpha(\tau_i - a)$$

and introduce Markov chain $w_i^\varepsilon,\ i \geqslant 0,\ w_0^\varepsilon = 0$, describing almost deterministic single-channel queuing system

$$w_{i+1}^\varepsilon = \max(0, w_i^\varepsilon + \eta_i^\varepsilon - \tau_i^\varepsilon) = \max(0, w_i^\varepsilon + \varepsilon^\alpha \delta_i).$$

Here $\delta_0, \delta_1, \ldots,$ is a sequence of independent and identically distributed random variables, $\delta_i = \eta_i - \tau_i + \varepsilon,\ M\delta_i = 0$. In high load mode, when the load factor $\rho = \dfrac{b}{a}$ is close to one, the positive parameter $\varepsilon = (1-\rho)a$ is small: $\varepsilon \ll 1$. Value $\alpha > 0$ characterizes the rate of decreasing random perturbations with increasing loading.

Due to known results for a single-channel queuing system $G|G|1|\infty$ (see e.g. [10, Chapter 1, \$3]) Markov chain $w_i^\varepsilon,\ i \geqslant 0$ has given for any ε, α : $0 < \varepsilon,\ 0 < \alpha$ the stationary distribution $\lim\limits_{i\to\infty} \mathcal{P}\{w_i^\varepsilon > y\} = \mathcal{P}\{W_\alpha(\varepsilon) > y\},\ y \geqslant 0$. Using results of articles [11, 12, 13, 14, 15, 2, 17, 18, 19, 20, 21] it is possible to prove the following theorems.

Theorem 2. In a single-channel queuing system $G|G|1|\infty$ for some positive constants $\beta,\ c < \infty$ the inequality $M|\delta_1|^{2+\beta} \leqslant c$ is true. Then the following limit relations are valid: $W_\alpha(\varepsilon) \Rightarrow$ (weak convergence) $+\infty,\ 0 \leqslant \alpha < 1/2$;
$W_\alpha(\varepsilon) \Rightarrow 0,\ 1/2 < \alpha;\ W_\alpha(\varepsilon) \Rightarrow \eta,\ \mathcal{P}\{\eta > y\} = e^{-2y/d},\ \alpha = 1/2.$

Theorem 3. Assume that in a single-channel queuing system $G|G|1|\infty$, when for some $\nu,\ 1 < \nu < 2;\ h_\nu > 0$, one of the following relations

with $y \to \infty$ are true for distributions tails: $P(\eta_1 > y) \sim h_\nu y^{-\nu},\ P(\tau_1 > y) = o(P(\eta_1 > y)$,
or $P(\tau_1 > y) \sim h_\nu y^{-\nu},\ P(\eta_1 > y) = o(P(\tau_1 > y)$. Then the following limit relations are valid: $W_\alpha \Rightarrow +\infty,\ 0 \leqslant \alpha < 1/\nu;\ W_\alpha \Rightarrow 0,\ 1/\nu < \alpha;\ \varepsilon \to 0$.

Thus, a parameter α, characterizing either the rate of convergence of the load factor to one in the system $M|M|n|\infty$, or a random fluctuation in the system $G|G|1|\infty$, allows us to detect the convergence of the stationary waiting time to either zero or infinity.

2. Related statistical problems

In this section statistical estimates of characteristics of non-uniform Poisson flow, describing distribution of animals in some areas, and the most powerful decision rule for constructing of technical systems discriminating "friend-foe" are constructed and analysed. Main idea of these considerations is a choice of convenient objective functions for these estimates: relative errors or dependences on combinations of some parameters. These results are based on the classification of statistical problems proposed in the monographs [22, 23] and on the ideas of testing statistical hypotheses in the processing of physical and physics – technical observations [24, 26].

Error in estimating the mean number of Poisson flow points. Specialists in the field of earth sciences have the task of estimating the error of the mean number of Poisson flow points from observations in different cells of a square grid. Let the study area be divided into m cells, and the number of points in them in the area k is $n_k,\ k = 1, \dots, m$. It is natural to assume that the random variables $n_1, \dots, n_m$ are independent and have Poisson distributions with the parameters $\lambda_1, \dots, \lambda_m$. Using the properties of the Poisson distribution, it is easy to establish that the random sum $N = n_1 + \dots + n_m$ has a Poisson distribution with the parameter $\Lambda = \lambda_1 + \dots + \lambda_m$ and consequently $EN = \Lambda,\ VarN = \Lambda$.

Using the known properties of the mathematical expectation and the variance of the Poisson distribution, we proceed to estimate the relative error. To do this, consider the random variable $N/EN = N/\Lambda$. Variance of this random variable $VarN/\Lambda = 1/\Lambda$ and so using Chebyshev inequality it is possible to obtain the following relation $P\left(|N/\Lambda - 1| > \Lambda^{-1/3}\right) \leqslant \Lambda^{-1/3} \to 0,\ \Lambda \to \infty$. Therefore, the relative error decreases with the growth of Λ.

This result does not depend on the non uniformity of the distribution density of the Poisson flow of points, and therefore does not depend on the parameters $\lambda_1, \dots, \lambda_m$. It is based on the well-known models of Poisson point flows in the theory of random sets, which are used in the earth sciences [27].

Resolution of the most powerful decision rule. In the papers [28, 29, 30, 31], a neural network converter "Biometrics access code" is built on the basis of an electroencephalogram. The main indicator of the effectiveness of this converter is the probability of errors of the first kind α_1, when the probability of errors of the second kind α_2 is chosen by experts to distinguish between simple hypotheses "friend-foe". This distinction of hypotheses is made using the most powerful decision rule. A special role here is played by a set of sample characteristics, with the help of which these hypotheses are distinguished.

Consider a sample $x_1, \dots, x_n$, consisting of independent random variables having a normal distribution with an average a and a variance σ^2. From two hypotheses $H_1 = (a = a_1)$, $H_2 = (a = a_2)$, $a_1 > a_2$, the most likely hypothesis is selected. We introduce a characteristic A of the resolution of the most powerful decision rule. The value A is determined by by the following parameters: the probability α_2, the sample size n, the variance σ^2 and the difference of the averages $a_1 - a_2$. Its calculation is based on the Neumann-Pearson lemma and the well-known rule for finding the most powerful solving rule by the Bayesian solving rule.

The most powerful solving rule has the form

$$\sum_{i=1}^{n} x_i/n > C \ \Rightarrow\ H_1,\ \sum_{i=1}^{n} x_i/n \leqslant C \ \Rightarrow\ H_2.$$

Assume that the function $t(\alpha)$, $0 < \alpha < 1$, is defined by the equality $\alpha = \displaystyle\int_{t(\alpha)}^{\infty} \frac{\exp(-u^2/2)}{\sqrt{2\pi}} du$. Put $t(\alpha_1) = A$, then it is not difficult to obtain the relations $C = a_2 + t(\alpha_2)\sigma/\sqrt{n}$, $A = A(a_1, a_2, \alpha_2, n, \sigma) = \sqrt{n}(a_1 - a_2)/\sigma - t(\alpha_2)$. Consequently there is strong dependence of α_1 on $A : \alpha_1 \leqslant \exp(-A^2)/(A\sqrt{2\pi})$. For example, let's take $\alpha_2 = 10^{-9}$, $\alpha_1 = 7 \cdot 10^{-4}$ (these values are taken from [31]), then $A = 9.19246$.

REFERENCES

1. *Dudin A., Nazarov A.* On a tandem queue with retrials and losses and state dependent arrival, service and retrial rates // Int. J. Oper. Res. 2017. Vol. 29. P. 170–182.
2. *Artalejo J., Gomez-Corral A.* Retrial Queueing Systems: A Computational Approach. Springer-Verlag: Berlin/Heidelberg, 2008. 320 p.
3. *Nazarov A., Moiseeva E.* Investigation of the RQ-system $MMPP|M|1$ by the method of asymptotic analysis in the condition of a large load // Izv. Tomsk. Polytech. Univ. 2013. Vol. 322. P. 19–23. (In Russian)

4. *Borovkov A.* Asymptotic Methods in Queueing Theory. J Wiley and Sons: New York, NY, USA, 1984. 292 p.
5. *Tsitsiashvili G., Osipova M.* Synergetic effects for number of busy servers in multiserver queuing systems // Commun. Comput. Inf. Sci. Ser. 2015. Vol. 564. P. 404–414.
6. *Mikosch T., Resnick S., Rootzen H., Stegeman A.* Is network traffic approximated by stable Levy motion or fractional Brownian motion? // Ann. Appl. Probab. 2002. Vol. 12. P. 23–68.
7. *Tsitsiashvili G., Markova N.* Synergistic effects in a multi-channel queuing system with an Erlang input flow // Bull. Pac. State Univ. 2015. Vol. 4. P. 17–22. (In Russian)
8. *Tsitsiashvili G., Osipova M.* Phase Transitions in Multiserver Queuing Systems // Communications in Computer and Information Science 2016. Vol. 638. P. 341–353.
9. *Afanasyeva L.* Queuing Systems with cyclic control processes // Cybernetics and system analysis. 2005. Vol. 41. № 1. P. 54–68. (In Russian)
10. *Borovkov A.* Stochastic processes in queuing theory. Moscow: Science, 1972. 280 p. (In Russian)
11. *Borovkov A.* Some limit theorems in the theory of mass service // Probability Theory & Its Applications. 1964. Vol. 9. № 4. P. 550–565. (In Russian)
12. *Prokhorov Yu.* Transient phenomena in queuing processes // Litov. math. collection. 1963. Vol. 3. № 1. P. 199–205. (In Russian)
13. *Harrison J.* The heavy traffic approximation for single server queues in series // J. Appl. Probab. 1973. Vol. 10. № 3. P. 613–629.
14. *Abate J., Choudhury G., Whitt W.* Calculation of the GI/G/1 waiting-time distribution and its cumulants from Pollaczek's formulas // Archiv für Elektronik und Ubertragungstechnik. 1993. Vol. 47. № 5/6. P. 311–321.
15. *Beran J., Sherman R., Taqqu M., Willinger W.* Long-range dependence in variable-bit-rate video traffic // IEEE Transactions on communications. 1995. Vol. 43. № 2/3/4. P. 1566–1579.
16. *Bingham N., Goldie C., Teugels J., Teugels J.* Regular Variation. Cambridge university, 1989. 479 p.
17. *Boxma O.* Fluid queues and regular variation // Performance Evaluation. 1996. Vol. 27/28. P. 699–712.
18. *Boxma O., Cohen J.* The M/G/l queue with heavy-tailed service time distribution // IEEE, J. Selected Areas Commun. 1998. Vol. 16. P. 749–763.
19. *Cohen J.* Some results on regular variation for distributions in queueing and fluctuation theory // J. Appl. Probab. 1973. Vol. 10. P. 343–353.
20. *Cohen J.* Superimposed renewal processes and storage with gradual input // Stochastic Process. Appl. 1974. Vol. 2. P. 31–58.

21. *Boxma O., Cohen J.* Heavy-traffic analysis for the GI/G/l queue with heavy-tailed distributions // Queueing systems. 1999. Vol. 33. P. 177–204.
22. *Borovkov A.* Mathematical statistics. Evaluation of parameters. Hypothesis testing. Moscow: Nauka, 1984. (In Russian)
23. *Borovkov A.* Mathematical statistics. Additional chapters. Moscow: Nauka, 1984. (In Russian)
24. *Gortsev A., Nezhelskaya L., Shevchenko T.* Estimation of the states of the MC-flow of events in the presence of measurement errors // Izv. Vuzov. Physics. 1993. Vol. 36. № 12. P. 67–85. (In Russian)
25. *Gortsev A., Nezhelskaya L.* Optimization of adapter parameters during observations of the MS flow // Izv. Vuzov. Physics. 1993. Vol. 36. № 12. P. 67–85. (In Russian)
26. *Gortsev A., Nezhelskaya L.* Optimization of adapter parameters during observations of the MS flow // In Sat. Stochastic and deterministic models of complex systems, Novosibirsk: Publishing house of the Computing Center of the Siberian Branch of the USSR Academy of Sciences, 1988. P. 20–32. (In Russian)
27. *Stoyan D., Kendall W., Mecke J.* Stochastic Geometry and Its Applications. New York: J. Wiley and Sons, 1987.
28. *Goncharov S., Borshevnikov A.* Construction of a neural network converter "Biometrics access code" based on the parameters of the visual evoked potential of an electroencephalogram // Reports of the Tomsk State University of Control Systems and Radioelectronics. 2014. Vol. 2. № 32. P. 51–55. (In Russian)
29. *Goncharov S., Borshevnikov A.* Prediction of the output parameters of the neural network converter "Biometrics-access code" based on an electroencephalogram // Information and security. 2018. Vol. 21. № 3. P. 302–307. (In Russian)
30. *Borshevnikov A., Goncharov S., Dobrzhinsky Yu.* On the requirements for the formation of synthetic images of an electroencephalogram for the tasks of highly reliable biometric authentication // Modern science: actual problems of theory and practice. Series: Natural and Technical Science. 2021. Vol. 2. P. 27–29. (In Russian)
31. *Borshevnikov A.* Neural network model of a highly reliable biometric user authentication system based on an encephalogram // Oral report at the general Institute seminar of the IAM FEB RAS. 2021.

Tsitsiashvili Gurami — Doctor of physics and mathematics, Professor, head of laboratory, Institute of Applied Mathematics FEB RAS. E-mail: *guram@iam.dvo.ru*

OPTIMIZATION OF MESSAGE TRANSMISSION WHEN IMPLEMENTING THE SECRET SHARING SCHEME

V. Zadorozhnyi, D. Sagaydak

Omsk State Technical University, Omsk, Russian Federation

We consider the transmission of messages with demultiplexing over two communication channels with different throughputs. Each message is divided into two unequal parts, of which the smaller is transmitted on one channel, the larger – on the other. The problem of calculating the optimal channel throughput is solved, taking into account the characteristics of the transmitted traffic.

Keywords: *Data flow distribution, communication channel throughput, demultiplexing, multiplexing, secret sharing.*

Introduction

With the introduction of the self-isolation regime, there is still an increasing demand for internet connection services, for an increase in data transfer speed, for throughput expansion and for the purchase of additional communication channels. The most popular transmitted content is video data, for example, online broadcasts of cinemas, educational webinars. Broadcasting is carried out using client-server applications, in which the content can first be transformed using any algorithms, and only then transmitted to the user. Pre-transformation of content can be carried out to ensure confidentiality, that is, to perform cryptographic transformations. In such situations, cryptographic algorithms can lead to significant delays in the playback of the video data stream due to the fact that reverse cryptographic conversion must be performed on the client side. The use of an additional communication channel makes it possible to organize distributed data transmission, which makes it possible to solve the problem of ensuring confidentiality, but questions arise related to the efficiency of the use of computing resources, optimization of channel throughput, synchronization of transmitted streams.

1. Formulation of the problem

In such a situation, it is advisable to consider the possibility of using other, not related to classical cryptography, information protection methods – secret sharing schemes (SSS) [1], demultiplexing.

In [2, 3], algorithms for dividing video data into unequal parts are proposed, which allow the sender and the receiver, using two communication channels with different throughputs, to carry out the diversity of the transmitted TCP / IP traffic over these channels, as, for example, it is described in [4, 5]. Further, we assume that SSS can be used to transmit streams of any messages, and all the transformations described in [2, 3] are performed directly on the binary characters of the messages. By dividing messages into unequal shares, a smaller share of each message is transmitted over a lower throughput channel, and a larger share is transmitted over a higher throughput channel.

When using two communication channels at the same time, questions arise related to the efficiency of computing resources, optimization of channels throughput, synchronization of transmitted streams, buffering. These issues can be addressed by implementing appropriate client-server applications and optimizing the throughput of the communication channels. It is advisable to optimize throughputs according to the criteria for minimizing costs, one part of which is associated with message delays in the network, the other part - with the payment of channel throughput, which monotonically grows with the growth of throughput.

As a mathematical model for optimizing a two-channel SSS, a network with splitting requests (S-network) with two single-channel queuing systems (QS) is proposed (see figure 1). In terms of queuing theory (QT), we will call messages and their parts requests, demultiplexing messages – splitting requests, multiplexing messages by assembling requests. Two requests corresponding to two parts of the same divided message will be called conjugate requests. We define the discipline of servicing queues in front of the channels as the FIFO discipline (first in – first out).

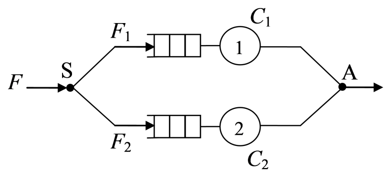

Figure 1. A network with splitting requests. S – split point, A – assemble point

Requests are transmitted through two channels with different throughputs C_1, C_2 (Kbit/s). At point S, the request is split into two requests, one of which arrives at QS1 and the other - in QS2. At point A, two conjugate requests "merge"and turn into one request. The incoming traffic F (Kbit/s) is divided into two parts $F_1 + F_2 = F$. The moment of entry of the request into the network is simultaneously the moment of its splitting and the moment of entry of the resulting conjugate parts into each of the two branches of the network. The intensities λ, λ_1 and λ_2 of the flows of requests entering the network, in the QS1 and in the QS2 are the same, and are described by the same probabilistic law.

Another feature of the S-network, not shown in figure 1, is that two more synchronization queues are formed before point A (one on each branch).

All requests arrive in each synchronization queue in the same sequence in which they enter the network. Therefore, if at least one request is pending in one synchronization queue, then the other synchronization queue is empty. The time u of message transmission (in terms of QT the time spent by the request in the network) is determined by the formula

$$u = \max(u_1, u_2) = \max(w_1 + x_1, w_2 + x_2), \tag{1}$$

where w_1 - waiting time for a request in queue 1; x_1 - service time of the request in channel 1; w_2 - waiting time of the conjugate request in queue 2; x_2 - service time of the adjoint request in channel 2 .
The average time U of staying in the S-network, according to (1):

$$U = M\left[\max(u_1, u_2)\right] = M\left[\max(w_1 + x_1, w_2 + x_2)\right]. \tag{2}$$

The time U depends on the throughputs C_1, C_2: $U = U(C_1, C_2)$.

Let the price of the throughput of any channel is m c.u./(Kbit/s). Then the problem of optimizing the throughputs C_1, C_2 of the S-network channels can be formulated as follows:

$$f = lU\left(C_1, C_2\right) + mC_1 + mC_2 \to \min_{C_1, C_2}, \tag{3}$$

$$\begin{cases} C_1 \geqslant F_1 \\ C_2 \geqslant F_2, \end{cases} \tag{4}$$

where l (c.u./s) is the price of the average delay of an request in the network for one second.

Coefficient l is equal to the losses (occurring due to the increase in the average time U by 1 s), calculated for the period of operation of the network.

Then the problem (3), (4) is posed as the problem of minimizing the average costs over the period of network operation.

2. Exponential network with independent branches

Consider a network with independent branches (figure 2), which can be examined as a first approximation of the S-network shown in figure 1.

In this network with independent branches, QS1 and QS2 operate independently, each serving its own stream of requests. Each of these QSs individually is equivalent to the corresponding QS in the S-network, but a pair of QSs in a network with independent branches (figure 2) is not equivalent to a pair of QSs in an S-network (figure 1).

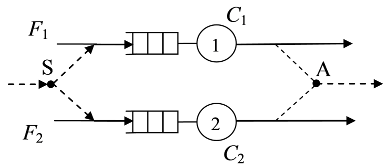

Figure 2. S-network with independent branches

Dashed lines in figure 2 shows a single passage through the network of a single split request arriving in a stationary mode of network operation at a random moment in time. The time it takes for a split request to pass through the network is expressed by formula (1), the network optimization problem has the form (3), (4).

Calculation of the system in figure 2 includes the following steps. First, we find the distribution of the sojourn time of unsplittable request in QS1 and QS2. Further, since the quantities u_1 and u_2 are independent, we can easily find the distribution of the maximum of these quantities, and through it - the required $M\,[max(u_1, u_2)]$.

$$M\,[max(u_1,u_2)] = \frac{1}{\mu_1(1-\rho_1)} + \frac{1}{\mu_2(1-\rho_2)} - \\ - \frac{1}{\mu_1(1-\rho_1)+\mu_2(1-\rho_2)}, \tag{5}$$

where μ_1, μ_2 are the intensity in the first and, respectively, the second channel, ρ_1, ρ_2 are the load factors of the first and second channels.

Passing in expression (5) to the parameters of problem (3), (4) $\rho_1 = F_1/C_1$, $\rho_2 = F_2/C_2$, $\mu_1 = C_1/H_1$, $\mu_2 = C_2/H_2$ and substituting it taking into account (2) in (3) we get the problem:

$$f = \frac{lH_1}{(C_1-F_1)} + \frac{lH_2}{(C_2-F_2)} - \frac{l}{H_1^{-1}(C_1-F_1)+H_2^{-1}(C_2-F_2)} + \\ + mC_1 + mC_2 \to \min_{C_1,C_2}, \tag{6}$$

$$\begin{cases} C_1 \geqslant F_1, \\ C_2 \geqslant F_2. \end{cases} \tag{7}$$

The point (C_1, C_2) of the local minimum of the function f can be found from the system of two equations $\partial f/\partial C_1$, $\partial f/\partial C_2$ i.e.

$$-\frac{lH_1}{x^2}+\frac{lH_1^{-1}}{\left(H_1^{-1}x+H_2^{-1}y\right)^2}+m=0,-\frac{lH_2}{y^2}+\frac{lH_2^{-1}}{\left(H_1^{-1}x+H_2^{-1}y\right)^2}+m=0, \quad (8)$$

where $x = C_1 - F_1 > 0$, $y = C_2 - F_2 > 0$.

Example 1. Let the cost coefficients $l = 1$, $m = 0.001$ and the parameters $H_1 = 100$, $H_2 = 800$, $F_1 = 700$, $F_2 = 5600$. In this case, in the set of solutions of system (8), only one solution contains a pair of positive real numbers: $x \approx 230.461$, $y \approx 848.818$. Accordingly, the minimum of the objective function (6) is attained at the point $C_1 = F_1 + x \approx 930.461$, $C_2 = F_2 + y \approx 6448.818$, and is $f = 8.458558$ c.u. In this case, we obtain $\rho_1 = F_1/C_1 = 0.752$, $\rho_2 = F_2/C_2 = 0.868$.

The results of additional verification of the found optimal values C_1, C_2 are shown in figure 3 (upper curves). On the left, the upper curve is a section of the $f(C_1, C_2)$ surface by the plane $C_2 = 6448.818$, passing through the optimum point, on the right, a section by the plane $C_1 = 930.461$. Markers represent the corresponding simulation results.

Example 2. Given example 1 ($l = 1$, $m = 0.001$, $H_1 = 100$, $H_2 = 800$, $F_1 = 700$, $F_2 = 5600$) we optimize the S-network shown in figure 1, in which channels only transmit split requests. In the simulation model, the share of a request transmitted through the first channel is serviced in H_1/C_1 seconds on average, the servicing time has an exponential distribution. The share of a request transmitted through the second channel has a service time independent of the service time of the conjugate share, and has an exponential distribution with the average H_2/C_2. The intervals for the receipt of requests in the network have an exponential distribution with an average H_1/C_1. It is difficult to find an analytical solution for such an S-network. Optimization of throughputs C_1, C_2 is performed using S-network simulation by coordinate descent method. As an initial approximation, we took a solution that is optimal for a network with independent branches (see example 1). The optimal solution $C_1 \approx 905$, $C_2 \approx 6435$ was found in more than 80 steps, each of which required about 6 minutes of computer work with a 3.1 GHz processor. The results of checking this solution are shown in figure 3 lower curves, which are sections of the $f(C_1, C_2)$ surface by the planes $C_1 = 905$, $C_2 = 6435$, passing through the optimum point.

The time spent on finding the minimum of the function of two variables turned out to be rather large, although, as can be seen from figure 3, the initial approximation is close to the minimum point. The correlation coefficient between the durations u_1 and u_2 of staying in the QS of the conjugate

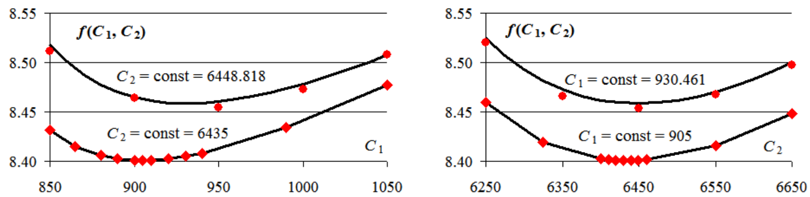

Figure 3. Upper curves - sections of the surface $f(C_1, C_2)$ by planes passing through the opt. point in example 1. Lower curves - sections for example 2

lobes was in the vicinity of the optimum point from 0.51 to 0.53. The load factors were $\rho_1 = 0.773$, $\rho_2 = 0.870$.

3. S-networks with synchronous branches

Consider an S-network in which each incoming request has a random size h (Kbit) and is split into two conjugate requests so that between their sizes h_1 and h_2 (where $h_1 + h_2 = h$) the same proportion always remains: $h_2/h_1 = \gamma = const$. If the throughputs of the channels is connected by the condition $C_2 = \gamma C_1$ then the service time x_1 of the request in channel 1 and the service time x_2 of the conjugate request in channel 2 coincide:

$$x_2 = h_2/C_2 = (\gamma h_1)/(\gamma C_1) = h_1/C_1 = x_1. \tag{9}$$

We call such a network an S-network with synchronous branches, or an Ss-network. The problem (3), (4) optimization of an S-network, whose branches are synchronous, can be solved exactly for any flow of requests for which QS in the branches can be calculated by exact QT methods.

Omitting the proofs, we present the following two theorems.

Theorem 1. Transmission of a regular flow of fixed-size requests through an optimal Ss-network leads to the same costs as transmission through an optimal single-channel system. In this case, the total throughput of the optimal Ss-network is equal to the throughput of the optimal single-channel system.

Theorem 2. The transmission of any flow of requests through the optimal Ss-network leads to the same costs as its transmission through the optimal single-channel system. Then the total throughput of the optimal Ss-network is equal to the throughput of the optimal single-channel system.

4. Networks with splitting requests in constant proportion

S-networks in which the condition $h_2/h_1 = \gamma = const$ is fulfilled during the splitting of requests, but the condition $C_2 = \gamma C_1$ is not imposed, we will call networks with split requests in equal proportions, or Se-networks. The Ss-networks considered above are a subset of Se-networks in which both of these conditions are satisfied. The optimization problem for Se-networks has certain specific features. It is written, like the general problem (3), (4) of optimization of S-networks, in the form

$$f = lU(C_1, C_2) + mC_1 + mC_2 \to \min_{C_1, C_2}, \tag{10}$$

$$\begin{cases} C_1 \geqslant F_1, \\ C_2 \geqslant F_2, \end{cases} \tag{11}$$

where $U(C_1, C_2) = M\,[\max(u_1, u_2)]$, and inherits the property (as opposed to a network with independent branches, see figure 2) that in general case in elementary function the non-expressible dependence of the mathematical expectation $M\,[\max(u_1, u_2)]$ on the varied parameters C_1, C_2 is here determined on stochastically interdependent random variables u_1, u_2 (times of stay in QS1 and QS2). The variable parameters C_1, C_2 in (10), (11) are independent of the optimization problem for Ss-networks, they are not related by the condition $C_2 = \gamma C_1$; therefore, both inequalities are preserved in constraints (11). Therefore, the progress of requests in the branches of the network is generally asynchronous here, which makes it difficult to find an explicit formula that accurately expresses the time $U(C_1, C_2) = M\,[max(u_1, u_2)]$ in terms of the network parameters.

However, in the next theorem it is established that in the class of Se-networks, optimal solutions to problem (10), (11) with independent throughputs always lie on the line $C_2 = \gamma C_1$.

Theorem. The Se-network, which is optimal for the transmission of any flow of requests, is an Ss-network, i.e. the optimal throughputs C_1, C_2 of the Se-network lie on the straight line $C_2 = \gamma C_1$.

Example 3. Using the conditions of example 2, let us establish the constant proportion $h_2 = \gamma H_1 = 8h_1$ for the sizes of the request shares, and describe the size h_1 by the third-order Erlang distribution with mean H_1. Then h_2 and $h = (1 + \gamma)h_1$ are also distributed according to the Erlang law with averages H_2 and H, respectively. Let us describe the intervals of requests arrival by the Erlang distribution with the mean H_1/F_1. The S-network defined in this way is a Se-network, and the optimal C_1 and C_2 for it lie on the line $C_2 = \gamma C_1$.

Its solution $C_{1\,\mathrm{opt}} = 763$ was found in 12 model runs. All three significant digits of this solution are exact, as are the first three digits of the solution $C_{2\,\mathrm{opt}} = \gamma C_{1\,\mathrm{opt}}$. The correlation coefficient between the durations of stay of two conjugated lobes in their QS is 1. Coefficients of channels loading are $\rho_1 = 0.917$, $\rho_2 = 0.917$. The average latency in both sync queues is zero. Then for $C_2 = \gamma C_1$ in Se-networks $u_1 = u_2$ and defined by formula (2) $U = U_1$. The optimal C_1 and the minimum cost (3) can be determined analytically if QS1 is, for example, a system of the class M/M/1, M/G/1.

5. Conclusion

The efficiency of the optimized Se-networks established in the study performed can be achieved by developing algorithms for the implementation of the SSS, which divide messages into shares that are in a constant or approximately constant proportion. Or for shares that have a high positive correlation coefficient. Taking into account the identified possibilities of SSS optimization on two physical channels, it is advisable to study the possibilities of SSS optimization and similar messaging schemes in networks with fractal traffic (using, in particular, the queue reduction methods developed for such networks in [6]), and in 5G networks.

REFERENCES

1. *Shamir A.* How to Share a Secret // Communications of the ACM. 1979. Vol. 22. № 11. P. 612–613.
2. *Sagaydak D.A., Faizullin R.T.* Model secret sharing schemes in systems transmit video // Computer Optics. 2013. Vol. 37. № 1. P. 105–112.
3. *Faizullin R.T, Sagaydak D.A.* Application of the prefix encoding algorithm for the data array in the video data stream secret sharing scheme // Reports of TUSUR. 2013. № 1(25). P. 136–140.
4. *Efimov V.I., Faizullin R.T.* Multiplexing system of distributed TCP/IP traffic // Mathematical structures and modeling. 2002. № 10. P. 170–171.
5. *Guss S.V., Lavrov D.N.* Approaches to the implementation of a network protocol for ensuring guaranteed delivery in multi-route data transmission // Mathematical structures and modeling. 2018. Vol. 2. № 46. P. 95–101.
6. *Zadorozhnyi V.N., Pagano M., Zakharenkova T.R.* Infinite Markings Method in Queueing Systems with the Infinite Variance of Service Time // CCIS. 2021. Vol. 1391. P. 296–311.

Zadorozhnyi Vladimir — doctor of technical sciences, professor. E-mail: *zwn2015@yandex.ru*

Sagaydak Dmitriy — graduate student. E-mail: *sagaydak.dmitriy@gmail.com*

PERFORMANCE INVESTIGATION OF REVERSE BALKING IN COGNITIVE RADIO NETWORKS USING SIMULATION

M. H. Zaghouani, H. Nemouchi, J. Sztrik

The University of Debrecen, Debrecen, Hungary

In this article, the concept of reverse balking is developed and integrated into a Cognitive Radio Network. Reverse balking is a customers pattern in which an arriving user is more likely to join a system if it is more occupied, and vice versa. This type of customer's attitude can be seen in a variety of industries, especially finance. The key performance measures are obtained with the help of simulation.

Keywords: *Finite source queuing systems, Simulation, Cognitive radio networks, Performance measures, Reverse balking.*

Introduction

Our Cognitive Radio Network (CRN) model's principal aim is to enhance the use of the free spaces in the primary frequency band to benefit the secondary. More details can be found in [1, 2, 3, 4, 5, 6]. Our queuing system considers two parts. The first part is developed for Primary Users (PUs) with a finite number of sources who generate primary calls after an exponentially distributed time. All the generated calls are placed in a FIFO queue for service. The second subsystem is dedicated to secondary users (SUs) jobs which are created following exponential distribution and routed to the secondary channel service (SCS) to obtain service. The service time of PUs and SUs is exponentially distributed as well. The generated licensed calls will verify the status of the PCS; if it is available, the service may begin immediately; if it is already in use by a primary call, the latter call will be placed in the FIFO queue. If a secondary customer is occupying the PCS, its service will be interrupted immediately and diverted back to the SCS. The aborted call will be resumed from the beginning of its service or added to the retrial queue (orbit) depending on its present state.

SCS, on the other hand, handles unlicensed queries. If the targeted server is idle, the SU is permitted to start the service; if it is occupied, they

The research work of János Sztrik is supported by the EFOP-3.6.1-16-2016-00022 project. The project is co-financed by the European Union and the European Social Fund. Mohamed Hedi Zaghouani is supported by the Stipendium Hungaricum Scholarship.

may attempt to start opportunistically their service in the PCS. If the last service channel is free, the low priority call may be able to begin the service; otherwise, if it is occupied, the call will be automatically added to the orbit, from where it will retry to be served after an exponentially distributed time.

Several investigations have dealt with CRN in different scenarios. The effect of server unreliability on the CRN, for example, was studied by the authors of [3]. Abandonment was employed in [6], wherein SUs were forced to leave the system if their total waiting time exceeded a random maximum waiting duration. Balking has been investigated in a variety of queuing systems, including [7, 8, 9, 10]. However, after a thorough search of numerous similar topics and reports, we were unable to find any studies that addressed this model in the scenario of revers balking, which is the novelty of our research.

1. System model

The queuing cognitive radio system shown in Fig. 1 is based on the following assumptions. Consider two interconnected subsystems in which primary requests are generated by a finite number of sources N_1 and submitted to the first server using an exponentially dispersed time with an average value of $1/\lambda_1$. If the unit is available, the service may begin; otherwise, the call will be placed in the preemptive priority queue. The principal users' service time is a random variable with an exponential distribution and a parameter μ_1.

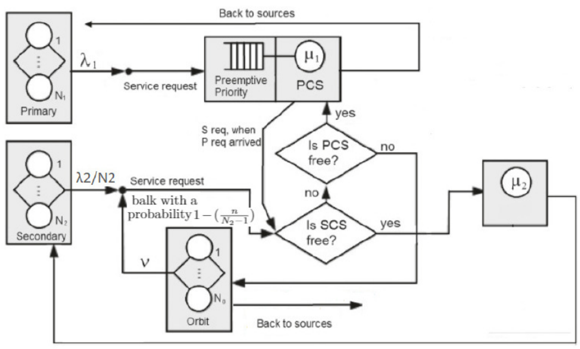

Figure 1. Finite-source retrial queuing system: Modeling the Cognitive Radio Network with reverse balking

Table 1

Simulation input parameters

N_1	N_2	λ_1	λ_2/N_2	μ_1	μ_2	ν	p
20	50	0.03	x-axis	1	1	20	0.5

N_2 denotes the number of sources in the secondary subsystem. Each source produces low priority call according to an exponentially distributed time with parameter λ_2/N_2. SUs service time is generally distributed using hypo-exponential, hyper-exponential, and gamma distributions with the same mean and different variances with a rate μ_2. The retrial time of the secondary customers is supposed to be an exponentially distributed random variable with a parameter ν.

When the system is empty (at the start of the simulation) first customers might balk (do not enter the system) with probability p or join it with $1 - p$. However, when there is at least one customer in the system, new arriving ones balk with a probability 1-q and enter the system with probability q=($\frac{n}{N_2-1}$), while n is the number of SU in the system at a time t. Reverse balking is the term for this type of balking.

2. Simulation results

Assuming that all random variables included in the system are exponentially distributed except the secondary service, we created a stochastic simulation program written in C coding language with SimPack to generate the results of this section. All the numerical results were collected by the validation of the simulation outputs. Table 1 shows the numerical values of the simulation main class input parameters while.

Figure 2 illustrates the influence of secondary service time distribution on the mean residence time of SUs versus secondary request time generation. A high distributions sensitivity can be observed when service times are gamma distributed with a squared coefficient of variation greater than one, especially, in the beginning of the simulation. Furthermore, increasing the arrival intensity of SUs, did not involve a greater mean response time for SUs until value 2.8, where the mean response time was noticeably increased. This was the effect of the reverse balking, as new coming customers are getting more encouraged to enter the system through the time.

The impact of the service time distribution for the secondary subsystem on the mean balking rate versus λ_2 can be observed in Figure 3. Increasing the secondary arrival rate involves a higher discouragement for new arriving secondary customers, this can be seen clearly in the case of Gamma distribution.

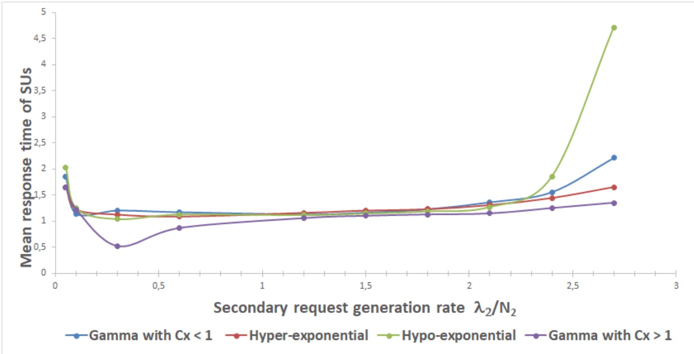

Figure 2. The impact of secondary service time distribution on the mean residence time of SUs vs secondary request time generation

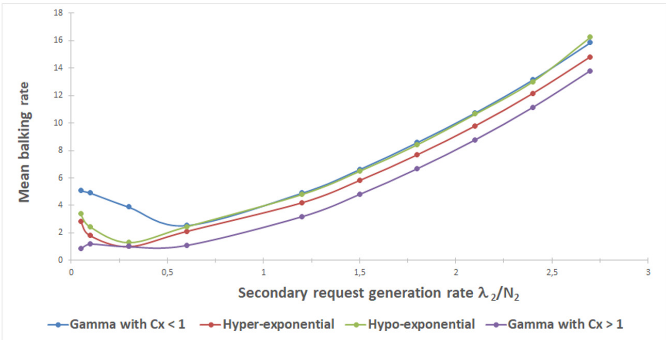

Figure 3. The impact of secondary service time distribution on the mean balking rate of SUs vs secondary request time generation

It is well known according Gamma distribution function that when $c_x^2 > 1$ the generated random service time is great which leads to an overloading of the system.

3. Conclusion

This paper introduces a finite-source retrial queueing system with two non-independent components. Our system was designed to model a cognitive radio network with primary and secondary service units, as well as reverse balking. A thorough review was conducted using simulation to investigate the impact of service time distributions and cognitive technology on the system's key performance measures.

REFERENCES

1. *Devroye N., Vu M. and Tarokh V.* Cognitive radio networks. IEEE Signal Processing Magazine. Vol.25. 2008. 12–23.
2. *Gunawardena S. and Zhuang W.* Modeling and Analysis of Voice and Data in Cognitive Radio Networks // Springer, 2014.
3. *Nemouchi H. and Strik J.* Performance simulation of finite-source cognitive radio networks with servers subjects to breakdowns and repairs // Journal of Mathematical Sciences. Vol.237. 2008. 702–711.
4. *Akyildiz I., Won-Yeol L., Vuran M. and Mohanty S* Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. // Computer networks. Vol.50. 2006. 2127–2159.
5. *Mitola J. and Maguire G* Cognitive radio: making software radios more personal // IEEE personal communications. Vol.6. 1999. 13–18.
6. *Zaghouani M. H. and Sztrik J* Performance evaluation of finite-source Cognitive Radio Networks with impatient customers // Líceum University Press. Vol.51. 2020. 89–99.
7. *Kumar R. and Som B K* An\(M/M/1/N\) queuing system with reverse balking and reverse reneging // Advanced Modeling and Optimization. Vol.6. 2014. 339–353.
8. *Kumar R. and Som B K* Multi-server queue with reverse balking and impatient customers // Pak. J. Statist. Vol.36. 2020. 91–101.
9. *Kumar R. and Som B K* A heterogeneous queuing system with reverse balking and reneging // Journal of Industrial and Production Engineering. Vol.35. 2018. 1–5.
10. *Saikia G. and Choudhury A* A single server Markovian queuing system with limited buffer and reverse balking // Independent Journal of Management & Production. Vol.12. 2021. 1774–1784.

Mohamed Hedi Zaghouani — PhD student, Department of Informatics Systems and Networks, Faculty of Informatics. E-mail: *zaghoueni.hedi@gmail.com*

Dr. Hamza Nemouchi — A PhD degree holder from the Faculty of Informatics. E-mail: *nemouchih@gmail.com*

Dr. János Sztrik — Full Professor at Department of Informatics Systems and Networks, Faculty of Informatics. E-mail: *sztrik.janos@inf.unideb.hu*

ON THE EXISTENCE OF THE STATIONARY DISTRIBUTION IN A POLLING SYSTEM WITH AUTOREGRESSIVE POISSON INPUTS

A. V. Zorine

Nizhni Novgorod State University, Nizhni Novgorod, Russian Federation

In this paper we introduce a new queueing model with a special kind of input processes. It is assumed that the number of arrivals during consecutive time intervals makes an autoregressive sequence with conditional Poisson distributions. A single server serves input flows one by one in cyclic order with instantaneous switching. A d-limited policy is used. The mathematical model of the queueing process takes form of a multidimensional discrete Markov chain. The Markov chain keeps track of the server state, recent arrival numbers and queues' lengths. The necessary and sufficient condition for the existence of the stationary probability distribution is found. A possibility to give an explicit solution for the stationary equations for the probability generating functions is discussed.
Keywords: *Autoregressive Poisson process, polling system, cyclic service, stationarity conditions, probability generating functions.*

Introduction

Popular models for input flows in queueing systems are Poisson, MAP (BMAP), and renewal processes (e.g., [1, 2]). In few cases only, some of the flow constituents are assumed being in an autoregressive relation, e.g. interarrival intervals or batch sizes [3, 4]. So, we will propose and investigate a new queueing model where the numbers of arrivals during certain time intervals are in an autoregressive sort of dependence.

The queueing system belongs to a class of polling systems [5]. Besides the inputs, it differs from classical polling systems by an assumption on the service process. Service time distributions are not known (in real queueing systems service times can be dependent and have different probability distributions), but the server's sojourn time distribution for each node is given together with the upper limit on the number of services customers. It models for example a roads intersection controlled by a fixed-cycle traffic-light, and data transmission nodes governed by a Round Robin algorithm.

We will demonstrate that even under simple assumptions on the queueing system structure the equation for the stationary probability distribution generating function is hard to solve. Still we will obtain conditions for the existence of the stationary probability distribution in the system using the iterative-dominating approach [6, 7].

1. The queueing system

Let us assume that all random variables and random elements in what follows are defined on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Then $\mathbb{E}(\cdot)$ denotes the mathematical expectation with respect to the probability measure $\mathbb{P}$. Set $\varphi(x; a) = a^x e^{-a}/x!$ for $a > 0$ and $x = 0, 1, \ldots$.

Consider a queueing system with $m < \infty$ input flows and a single server. Customers from the j-th flow join an infinite-capacity buffer O_j. Probability properties of the input flows will be defined later. The server spends a constant time $T > 0$ in front of each queue, and then an instant switch-over to the next queue occurs. After the last queue the first queue is visited. The server implements a d-limited policy: during its stay at the j-th queue the server can provide service to $d = \ell_j$ customers at most from that queue, no matter when exactly they arrived if they have arrived before the time T expired.

Let τ_0, $\tau_{i+1} = \tau_i + T = (i+1)T$, $i = 0, 1, \ldots$ be the time instants when the server switches to a next queue. Denote by $\Gamma^{(r)}$ the server state when it is at the r-th queue, $i = 1, 2, \ldots, m$ and let $\Gamma = \{\Gamma^{(1)}, \Gamma^{(2)}, \ldots, \Gamma^{(m)}\}$ be the server state space. Let a random variable $\Gamma_i \in \Gamma$ be the server state during the time interval $(\tau_{i-1}, \tau_i]$ for $i = 1, 2, \ldots$, and $\Gamma_0 \in \Gamma$ be the random server state at time τ_0. Let $r \oplus 1 = r + 1$ for $r < m$ and $m \oplus 1 = 1$. Then $\Gamma_{i+1} = \Gamma_{i+1}(\omega) = \Gamma^{(r \oplus 1)}$ for all $\omega \in \Omega$ such that $\Gamma_i = \Gamma^{(r)}$.

Denote by $\eta_{j,i}$, $i = 1, 2, \ldots$ the random number of new customers arriving from the flow Π_j during the time interval $(\tau_i, \tau_{i+1}]$, $j = 1, 2, \ldots, m$. Let $\eta_{j,-1}$ be a non-negative integer-values random variable, $j = 1, 2, \ldots, m$. Let us assume that the conditional probability distribution of $\eta_{j,i+1}$ for any given $\eta_{j,-1} = x_{-1}$, $\eta_{j,0} = x_0$, $\ldots$, $\eta_{j,i} = x_i$ is the Poisson distribution with parameter $(a_j x_i + b_j)$ for some $a_j > 0$ and $b_j > 0$, so that the regression of $\eta_{j,i+1}$ on past numbers of arrivals equals

$$\mathbb{E}(\eta_{j,i+1} \mid \{\eta_{j,-1} = x_{-1}, \eta_{j,0} = x_0, \ldots, \ldots, \eta_{j,i} = x_i\}) = a_j x_i + b_j.$$

We will call such an input flow an *autoregressive Poisson flow*. In particular, if the flow Π_j is a classical Poisson with intensity λ_j then we will have $a_j = 0$ and $b_j = \lambda_j T$. Further, let us assume that the stochastic sequences

$$\{\eta_{j,i}; i = -1, 0, \ldots\}, \qquad j = 1, 2, \ldots, m$$

are independent.

Denote by $\varkappa_{j,i}$ the random number of customers in the queue O_j at time instant τ_i. Denote by $\xi_{j,i}$ the largest number of customers which can be serviced from O_j during the time interval $(\tau_i, \tau_{i+1}]$. Then the probability

$$\mathbb{P}(\{\xi_{1,i} = y_1, \xi_{2,i} = y_2, \ldots, \xi_{m,i} = y_m\} \mid \{\Gamma_i = \Gamma^{(r)}\})$$

equals 0 for t$y_j > 0$ and $y_k > 0$ for some $k \neq j$; it equals 1 for $y_{r\oplus 1} = \ell_j$. We have

$$\varkappa_{j,i+1} = \max\{0, \varkappa_{j,i} + \eta_{j,i} - \xi_{j,i}\}, \quad i = 0, 1, \ldots; j = 1, 2, \ldots, m.$$

The recurrent equations and probability distributions given above prove the following claims.

Theorem 1. For a given probability distribution of the vertor

$$(\Gamma_0, \varkappa_{1,0}, \varkappa_{2,0}, \ldots, \varkappa_{m,0}, \eta_{1,-1}, \eta_{2,-1}, \ldots, \eta_{m,-1}),$$

random sequences

$$\{(\Gamma_i, \varkappa_{1,i}, \varkappa_{2,i}, \ldots, \varkappa_{m,i}, \eta_{1,i-1}, \eta_{2,i-1}, \ldots, \eta_{m,i-1}); i = 0, 1, \ldots\},$$
$$\{(\Gamma_i, \varkappa_{j,i}, \eta_{j,i-1}); i = 0, 1, \ldots\}, \qquad j = 1, 2, \ldots, m$$

are irreducible periodic Markov chains.

2. Analysis of the model

The main purpose of this section is to establish necessary and sufficient conditions for the existence of the stationary probability distribution of the Markov chain $\{(\Gamma_i, \varkappa_{j,i}, \eta_{j,i-1}); i = 0, 1, \ldots\}$ for $j = 1, 2, \ldots, m$, since it is easy to prove then, that the Markov chain

$$\{(\Gamma_i, \varkappa_{1,i}, \varkappa_{2,i}, \ldots, \varkappa_{m,i}, \eta_{1,i-1}, \eta_{2,i-1}, \ldots, \eta_{m,i-1}); i = 0, 1, \ldots\}$$

has a stationary probability distribution if and only if each single $\{(\Gamma_i, \varkappa_{j,i}, \eta_{j,i-1}); i = 0, 1, \ldots\}$, $j = 1, 2, \ldots, m$ does. In the remainder of this section the value of the index j is fixed.

In the first place, for the existence of the stationary distributions of the Markov chains, the inputs $\{\eta_{j,i}; i = 0, 1, \ldots\}$ need to have statioinary probability distribution. This is possible only if $0 < a_j < 1$ for all $j = 1, 2, \ldots, m$. We assume so in the rest of the section.

Let us define

$$Q_{j,i}(r, x, y) = \mathbb{P}(\{\Gamma_i = \Gamma^{(r)}, \varkappa_{j,i} = x, \eta_{j,i-1} = y\}).$$

Let $I(\cdot)$ denote the indicator random variable for the event given in the parentheses. Let us introduce for $|z| \leqslant 1$, $|w| \leqslant 1$ and $i = 0, 1, \ldots$ the probability generating functions

$$\Psi_{j,i}(z, w; r) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} z^x w^y Q_{j,i}(r, x, y)\mathbb{E}\big(z^{\varkappa_{j,i}} w^{\eta_{j,i-1}} I(\{\Gamma_i = \Gamma^{(r)}\})\big).$$

Theorem 2. The following recurrent equations with respect to $i = 0$, $1, \ldots$ hold:

$$\Psi_{j,i+1}(z, w; r \oplus 1) = e^{b_j(zw-1)}\Psi_{j,i}(z, e^{a_j(zw-1)}; r), \qquad r \oplus 1 \neq j;$$

$$\Psi_{j,i+1}(z, w; r \oplus 1) = z^{-\ell_j} e^{b_j(zw-1)}\Psi_{j,i}(z, e^{a_j(zw-1)}; r) +$$
$$+ \sum_{x=0}^{\ell_j-1}\sum_{n=0}^{\ell_j-x-1}\Big(\sum_{y=0}^{\infty} Q_{j,i}(r, x, y)\varphi(n; a_j y + b_j)\Big)(1 - z^{x+n-\ell_j})w^n$$

for $r \oplus 1 = j$.

Using methods from [6, 7] we get

Theorem 3. For the existence of the stationary probability distribution of the Markov chain $\{(\Gamma_i, \varkappa_{j,i}, \eta_{j,i-1}); i = 0, 1, \ldots\}$ it is necessary and sufficient that

$$\frac{b_j}{1 - a_j}\, m < \ell_j. \tag{1}$$

The condition in the last theorem can be easily interpreted from a physical point of view because the quantity $mb_j(1 - a_j)^{-1}$ is the stationary expected number of arrivals from the flow Π_j during a complete cycle of the server.

In course of the proof of Theorem 3 the following Lemma is essential.

Lemma 1. If $0 < a < 1$ then the equation $w = e^{a(wz-1)}$ has a unique solution

$$w(z) = e^{-a} + \sum_{n=1}^{\infty} z^n \frac{(n + 1)^{n-1} a^n e^{-(n+1)a}}{n!},$$

convergent in the open disk $|z| < a^{-1}e^{a-1}$, such that $w(1) = 1$, $|w(z)| < 1$ for $|z| < 1$.

To obtain equations for the time-stationary probability generating functions we can omit indices i and $i + 1$ in the equations in Theorem 3. Substituting there $w = w(z)$ from Lemma 1 where $a = a_j$ and $b = b_j$, and denoting

$$A_j(x, n) = \sum_{y=0}^{\infty} Q_j(r, x, y)\varphi(n; a_j y + b_j)$$

for r such that $r \oplus 1 = j$ we get

$$\Psi_j(z, w(z); r \oplus 1) = e^{b_j(zw(z)-1)}\Psi_j(z, w(z); r), \qquad r \oplus 1 \neq j; \tag{2}$$

$$\Psi_j(z, w(z); r \oplus 1) = z^{-\ell_j} e^{b_j(zw(z)-1)}\Psi_j(z, w(z); r)+$$
$$+\sum_{x=0}^{\ell_j-1}\sum_{n=0}^{\ell_j-x-1} A_j(x, n)(1 - z^{x+n-\ell_j})(w(z))^n \tag{3}$$

for $r \oplus 1 = j$. To solve these equations for the functions $\Psi_j(z, w(z); r)$, $r = 1, 2, \dots, m$, one needs to identify $\ell_j(\ell_j + 1)/2$ unknown constants $A(x, n)$, $0 \leqslant x + n < \ell_j$, n, x integers. We get

$$(z^{\ell_j} - e^{mb_j(zw(z)-1)})\Psi_j(z, w(z); r) =$$
$$= \sum_{x=0}^{\ell_j-1}\sum_{n=0}^{\ell_j-x-1} A_j(x, n)(z^{\ell_j} - z^{x+n})(w(z))^n, \quad r \oplus 1 = j.$$

Case 1. If $\ell_j = 1$, then the only unknown constant is $A_j(0, 0)$. Recalling that $\Psi_j(z, z; r) = 1/m$ and expanding terms $z - e^{mb_j(zw(z)-1)}$, $(1 - z^{-1})$ in the left neighborhood of $z = 1$, we get

$$\Big(1 - \frac{mb_j}{1 - a_j}\Big)\frac{1}{m} = A_j(0, 0).$$

Case 2. If $\ell_j > 1$, we have $\ell_j(\ell_j + 1)/2 > 1$ unknown constants. Let us study the equation

$$z^{\ell_j} - e^{b_j(zw(z)-1)} = 0.$$

It follows from the modified Rouché theorem [8] and the Lemma below that it has exactly $\ell_j - 1$ zeros inside the unit disk $|z| < 1$ when the stationarity condition (1) is fulfilled.

Lemma 2. If inequality (1) is fulfilled, then $|e^{b_j(zw(z)-1)}| < 1$ for all $|z| = 1$, $z \neq 1$.

Denote these zeros by $\beta_1, \beta_2, \dots, \beta_{\ell_j-1}$.

Theorem 4. If inequality (1) is fulfilled then the following equations take place:

$$\sum_{x=0}^{\ell_j-1}\sum_{n=0}^{\ell_j-x-1} A_j(x, n)(\ell_j - x - n) = \frac{\ell_j}{m} - \frac{b_j}{1 - a_j},$$

$$\sum_{x=0}^{\ell_j-1}\sum_{n=0}^{\ell_j-x-1} A_j(x, n)((\beta_k)^{\ell_j} - (\beta_k)^{x+n})(w(\beta_k))^n = 0, \quad k = 1, 2, \dots \ell_j - 1.$$

The number of linear equations given by Theorem 4 is less than the number of unknown constants. Still, it was to be expected, since equations (2) and (3) are not equivalent to equations of Theorem 2 and by substituting $w = w(z)$ there we lose evidently essential parts of information about the generating functions of interest. Moreover, once we obtain all $A_j(x,n)$, $0 \leqslant x + n < \ell_j$, we still need to solve a functional equation relating $\Psi_j(z,w;r\oplus 1)$ to $\Psi_j(z,e^{a_j(zw-1)};r)$ in the polydisk $\{(z,w)\colon |z|\leqslant 1, |w|\leqslant 1\}\subset\mathbb{C}^2$.

REFERENCES

1. *Lucantoni D. M.* New results on the single server queue with a batch Markovian arrival process // Communications in Statistics: Stochastic Models. 1991. V. 7, N. 1. — P. 1–46.
2. *Klimow G. P.* Bedienungsprozesse. Basel: Springer, 1979. 255 p.
3. *Hwang G. U., Sohraby K.* On the exact analysis of a discrete-time queueing system with autoregressive inputs // Queueing Systems. — 2003. — V. 43, N. 1–2. — P. 29-41.
4. *Leontyev N. D., Ushakov V. G.* Analysis of a queueing system with autoregressive arrivals // Informatics and Applications. — 2014. — V. 8, N.3. — P. 39–44.
5. *Vishnevskii V. M., Semenova O. V.* Mathematical methods to study the polling systems // Automation and Remote Control. — 2006. — V. 67, N. 2. — P. 173–220.
6. *Zorine A.* Stochastic Model for Communicating Retrial Queuing Systems with Cyclic Control in Random Environment // Cybernetics and Systems Analysis. — 2013. — V. 49, N. 6. — P. 890–897.
7. *Zorine A. V.* Study of a Service Process by a Loop Algorithm by Means of a Stopped Random Walk // Communications in Computer and Information Science. — 2019. — V. 1109. — P. 121-–135.
8. *Klimenok V.* On the Modification of Rouche's Theorem for the Queueing Theory Problems // Queueing Systems. 2001. V. 38. P. 431–434.

Zorine Andrei Vladimirovich — Doctor of Physical and Mathematical Sciences, docent, professor in the Department of probability theory and data analysis, Institute of information technologies, mathematics and mechanics. E-mail: *andrei.zorine@itmm.unn.ru*

ИССЛЕДОВАНИЕ ОБЛАСТИ ПРИМЕНИМОСТИ АСИМПТОТИЧЕСКОГО ДВУМЕРНОГО ВЫХОДЯЩЕГО ПОТОКА СИСТЕМЫ С ВЫЗЫВАЕМЫМИ ЗАЯВКАМИ

А. Л. Благинин, И. Л. Лапатин

Национальный исследовательский
Томский государственный университет, г. Томск, Россия

В данной работе мы исследуем область применимости асимптотического приближения характеристической функции числа обслуженных заявок в выходящем потоке RQ–системы с вызываемыми заявками и MMPP. Целью исследования является проверка точности асимптотических результатов при различных параметрах входящего потока, так как вариация длин интервалов поступления заявок может оказывать существенное влияние на выходящий поток.

Ключевые слова: *Системы с повторными обращениями, системы с вызываемыми заявками, выходящий поток, коэффициент вариации, MMPP, имитационное моделирование.*

Введение

В теории массового обслуживания, Марковский модулируемый Пуассоновский поток позволяет точно представлять различные ситуации обработки пакетов в телекоммуникационных сетях. Наличие у входящего потока нескольких состояний, смена которых влечет за собой изменение интенсивности поступления заявок, дает возможность воссоздать взрывной режим прохождения трафика.

Ранее, в [1], было получено асимптотические приближение характеристической функции числа обслуженных заявок в RQ–системе [1, 2] с вызываемыми заявками [2] и MMPP за некоторое время t при большой задержке заявок на орбите. В ввиду специфики модели, а именно наличия двух типов обслуживаемых заявок и переменной интенсивности поступления входящих запросов, требуется провести ряд численных экспериментов для определения пригодности полученного приближения для вычисления характеристик выходящего потока при экстремальных параметрах входящего потока. В данной работе мы акцентируем внимание на параметрах MMPP, которые могут быть заданы таким образом, чтобы имитировать различные паттерны прихода заявок, когда длины интервалов их поступления могут сильно варьироваться.

1. Модель

Рассмотрим RQ–систему с MMPP. Входящая заявка, находя прибор свободным, занимает его. Прибор, в свою очередь, обслуживает ее в течение экспоненциально распределенного времени с параметром μ_1. Если по прибытии заявки прибор занят, она мгновенно перемещается на орбиту, где производит задержку в течении экспоненциально распределенного времени с параметром σ. В свободное от обслуживания входящих заявок время прибор вызывает их самостоятельно с параметром α и обслуживает в течение экспоненциально распределенного времени с параметром μ_2.

Введем обозначения: $i(t)$ — число заявок на орбите в момент времени t, $k(t)$ — состояние прибора: 0 – свободен, 1 – занят входящей заявкой, 2 – занят вызванной заявкой; $m_1(t)$, $m_2(t)$ — количество обслуженных входящих и вызванных заявок к моменту времени t соответственно, $n(t)$ — состояние управляющей Марковской цепи MMPP в момент времени t.

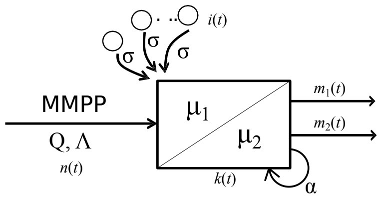

Рис. 1. Модель системы

2. Асимптотическое приближение характеристической функции

Ранее [1] было получено асимптотические приближение характеристической функции числа обслуженных заявок в рассматриваемой системе за время t. Оно имеет вид

$$F(u_1, u_2, t) = \boldsymbol{R} \cdot \exp\{\boldsymbol{G}(u_1, u_2)t\}\boldsymbol{e},$$

где $\boldsymbol{G}(u_1, u_2)$ – матрица коэффициентов уравнений Колмогорова и имеет вид

$$\boldsymbol{G}(u_1, u_2) = \begin{bmatrix} \boldsymbol{Q} - \boldsymbol{\Lambda} - (\alpha + \varkappa)\boldsymbol{I} & \mu_1 e^{ju_1}\boldsymbol{I} & \mu_2 e^{ju_2}\boldsymbol{I} \\ \boldsymbol{\Lambda} + \varkappa\boldsymbol{I} & \boldsymbol{Q} - \mu_1\boldsymbol{I} & 0 \\ \alpha\boldsymbol{I} & 0 & \boldsymbol{Q} - \mu_2\boldsymbol{I} \end{bmatrix}^T,$$

вектор-строка $\boldsymbol{R} = \{\boldsymbol{R}_0, \boldsymbol{R}_1, \boldsymbol{R}_2\}$ — двумерное стационарное распределение вероятностей процесса $\{k(t), n(t)\}$, где $\boldsymbol{R}_k$ имеет размерность N

(числа состояний управляющей цепи MMPP), $\varkappa$ — нормированное среднее число заявок на орбите, а $\boldsymbol{e}$ — единичный вектор-столбец размерности $N \cdot K$, где K — число состояний прибора, и $\boldsymbol{I}$ — единичная матрица размерности N.

3. Метод вычисления вероятностей

Для восстановления распределения вероятностей из характеристической функции используется обратное преобразование Фурье для дискретных случайных величин. Вычисление матричной экспоненты выполняется с помощью преобразования подобия [3]

$$\boldsymbol{G}(u_1, u_2) = \boldsymbol{T}(u_1, u_2)\boldsymbol{GJ}(u_1, u_2)\boldsymbol{T}(u_1, u_2)^{-1},$$

где $\boldsymbol{T}(u_1, u_2)$ — матрица собственных векторов $\boldsymbol{G}(u_1, u_2)$, а $\boldsymbol{GJ}(u_1, u_2)$ — диагональная матрица собственных чисел $\boldsymbol{G}(u_1, u_2)$.

$$e^{\boldsymbol{G}(u_1,u_2)t} = \boldsymbol{T}(u_1, u_2) \cdot \begin{bmatrix} e^{t\Lambda_1(u_1,u_2)} & 0 & 0 \\ 0 & e^{t\Lambda_2(u_1,u_2)} & 0 \\ 0 & 0 & e^{t\Lambda_3(u_1,u_2)} \end{bmatrix} \cdot \boldsymbol{T}(u_1, u_2)^{-1},$$

где Λ_n — собственное число $\boldsymbol{G}(u_1, u_2)$. Тогда получим

$$F(u_1, u_2, t) = \boldsymbol{R} \cdot \boldsymbol{T}(u_1, u_2) \cdot \begin{bmatrix} e^{t\Lambda_1(u_1,u_2)} & 0 & 0 \\ 0 & e^{t\Lambda_2(u_1,u_2)} & 0 \\ 0 & 0 & e^{t\Lambda_3(u_1,u_2)} \end{bmatrix} \cdot$$
$$\cdot \boldsymbol{T}(u_1, u_2)^{-1} \cdot \boldsymbol{e}.$$

Обратное преобразование Фурье

$$P(m_1, m_2, t) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{-i \cdot u_1 \cdot m_1} e^{-i \cdot u_2 \cdot m_2} F(u_1, u_2, t)\, du_1 du_2.$$

4. Численные эксперименты

В ходе экспериментов были зафиксированы следующие параметры системы $\alpha = 0.6, \mu_1 = 5, \mu_2 = 5$. В качестве метрики использовалось расстояние Колмогорова

$$\Delta = \max_{0 \le i < \infty} \left| \sum_{v=0}^{i} (P_0(v) - P_1(v)) \right|,$$

где $P_0(v)$ и $P_1(v)$ — сравниваемые распределения.

Вычисленное распределение сравнивалось с результатами работы имитационной модели при различных значениях σ.

Таблица 1

Расстояние Колмогорова

V/σ	10	1	0,1	0,01
1	0,04513	0,03853	0,01228	0,00231
2	0,14634	0,13380	0,05596	0,00743
4	0,43090	0,39781	0,22850	0,05534

В таблице 1 представлены значения расстояния Колмогорова при различной вариации между моментами наступления событий [4].

В таблице 2 показаны значения расстояния Колмогорова, которые были получены при сравнении распределений вероятностей выходящих процессов системы с такими же параметрами, однако в имитационной модели распределение времени задержки заявок на орбите является равномерным.

В таблицах 3 и 4 представлены результаты аналогичных экспериментов, когда время задержки заявок на орбите имеет гамма-распределение с параметрами формы $k = 1.1$ и $k = 0.5$ соответственно.

Таблица 2

Расстояние Колмогорова

V/σ	10	1	0,1	0,01
1	0,04458	0,03903	0,00995	0,00116
2	0,14649	0,13547	0,04648	0,00599
4	0,43088	0,40184	0,22393	0,05102

Таблица 3

Расстояние Колмогорова

V/σ	10	1	0,1	0,01
1	0,04406	0,03856	0,01162	0,00174
2	0,14650	0,13392	0,05585	0,00721
4	0,43045	0,39841	0,22933	0,05385

Заключение

В данной работе был представлен ряд численных экспериментов, показывающий, что при росте вариации длин интервалов между моментами поступления заявок MMPP скорость сходимости эмпирического распределения вероятностей к асимптотическому распределению замед-

Таблица 4

Расстояние Колмогорова

V/σ	10	1	0,1	0,01
1	0,04455	0,03735	0,01494	0,00426
2	0,14545	0,13059	0,06298	0,01900
4	0,42969	0,38820	0,22864	0,07953

ляется. Также было показано, что скорость сходимости распределения вероятностей к асимптотическому мало зависит от вида распределения времени задержки на орбите.

СПИСОК ЛИТЕРАТУРЫ

1. *Lapatin I. and Nazarov A.* Asymptotic Analysis of the Output Process in Retrial Queue with Markov-Modulated Poisson Input Under Low Rate of Retrials Condition // Communications in Computer and Information Science, 2021. Vol. 1141. P. 315–324.
2. `https://arxiv.org/abs/1906.09560` — Arxiv. 2019.
3. *Bronson R.* Matrix methods: An introduction. London: Gulf Professional Publishing, 1991. 501 p.
4. *Вишневский В., Дудин А., Клименок В.* Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях. М.: Техносфера, 2018. 562 с.

Благинин Алексей Леонидович — лаборант, Научное управление. E-mail: *alex-b.l@yandex.ru*

Лапатин Иван Леонидович — канд. физ.-мат. наук, доцент, Институт прикладной математики и компьютерных наук. E-mail: *ilapatin@mail.ru*

ИССЛЕДОВАНИЕ РЕСУРСНЫХ СИСТЕМ ВИДА $GI^{(2,v)+2M}/GI_2/\infty$ С ПАРАЛЛЕЛЬНЫМ ОБСЛУЖИВАНИЕМ РАЗНОТИПНЫХ ЗАЯВОК

Т. В. Бушкова[1], С. П. Моисеева[1], Е. В. Панкратова[2]

[1] *Национальный исследовательский Томский государственный университет, г. Томск, Россия*
[2] *Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, Россия*

В работе рассматривается математическая модель передачи заявок с реализацией функции разделения ресурсов в виде системы массового обслуживания с входящим рекуррентным потоком и двумя простейшими потоками. Исследование многомерного немарковского случайного процесса суммарных объемов занятых ресурсов в каждом блоке проводится методом асимптотического анализа в условии эквивалентного роста времени обслуживания на приборах. Получена гауссовская аппроксимация стационарного двумерного распределения вероятностей количества занятых ресурсов в параллельных блоках обслуживания.
Ключевые слова: *Система массового обслуживания, метод динамического просеивания, суммарный объем требований.*

Введение

В статье рассматривается математическая модель передачи разнотипных данных по двум каналам (блокам). На вход поступают потоки требований разного типа: два простейших потока с запросами на обслуживание в различных по качеству каналах(например, быстрый и медленный) и рекуррентный поток требований, которые передаются сразу по обоим каналам. Предполагается, что на суммарные объемы находящихся в системе требований нет ограничений. Продолжительности обслуживания являются неотрицательными случайными величинами, стохастически не зависят друг от друга и от количества выделенных ресурсов, и определяются заданными функциями распределения вероятностей с конечными первым и вторым моментами. После завершения обслуживания в одном из блоков ресурс немедленно освобождается в том количестве, что был затребован при поступлении заявки, независимо от того, закончилась ли передача данных в другом блоке или нет.

Ставится задача исследования двумерного случайного процесса суммарных объемов занятых ресурсов в каждом канале. Для исследования используются асимптотические методы [1], которые дают приемлемые для практического использования асимптотические выражения для искомых характеристик системы в случаях, когда точный анализ невозможен. В статье используется асимптотическое условие эквивалентного роста времени обслуживания, практический смысл которого состоит в том, что средняя скорость обслуживания намного меньше интенсивности запросов. Данная работа обобщает исследования [2, 3] на случай с несколькими потоками разнотипных данных.

1. Математическая модель

Рассмотрим ресурсную СМО с двумя блоками обслуживания, в каждом из которых неограниченное число обслуживающих приборов. На вход системы поступает три потока: два пуассоновских с параметрами λ^1 и λ^2 и рекуррентный поток заявок, периоды между наступлениями событий которого имеют функцию распределения $A(z)$. Будем считать, что заявки входящих простейших потоков делятся на два типа (в первом простейшем потоке заявки первого типа, а во втором – второго), а заявка рекуррентного потока расщепляется на две заявки разного типа, одна из которых поступает в первый блок обслуживания, а вторая – во второй. Каждая заявка занимает любой свободный сервер в своем блоке, где обслуживается случайное время с функцией распределения вероятностей $B_k(x)$, $(k = 1, 2)$. Время обслуживания соответствует продолжительности передачи заявки.

При входе в соответствующий блок обслуживания заявка формирует запрос на ресурс случайного объема размера $v_k > 0$, $(k = 1, 2)$ с функцией распределения вероятностей $G_k(y)$, $k = 1, 2$. Занятые ресурсы соответствуют ресурсам канала, захваченным сеансом передачи (например, радиочастоты, полоса пропускания и т.д.). После завершения обслуживания заявка покидает блок системы и освобождает занятый ею ресурс. Обозначим $V_i(t)$ – сумма всех объемов ресурсов, занимаемых заявками i-го типа $(i = 1, 2)$, находящимися на обслуживании в системе в момент времени t. Ставится задача анализа общего количества занятых ресурсов (суммарного объема занятого ресурса) в обоих блоках системы. Следует отметить, что в работе рассматривается модель бесконечного сервера с неограниченными ресурсами, предполагая, что блоки имеют достаточно большую пропускную способность.

Для дальнейших иследований воспользуемся методом динамического просеивания [4], с помощью которого переходим к рассмотрению двумерного процесса $\{W_1(t), W_2(t)\}$ – суммарный объем просеянных заявок за время t. Кроме того, введем случайный процеес $z(t)$ – остаточное время от момента t до момента наступления следующего события в исходном рекуррентном потоке.

2. Интегрально-дифференциальное уравнение Колмогорова

Введем обозначение для распределения вероятностей трёхмерного марковского процесса $\{z(t), W_1(t), W_2(t)\}$:

$$P(z, w_1, w_2, t) = P\{z(t) < z, W_1(t) < w_1, W_2(t) < w_2\}.$$

Для этого распределения составим интегрально-дифференциальное уравнение Колмогорова:

$$\begin{gathered}
\frac{\partial P(z,w_1,w_2,t)}{\partial t} = \frac{\partial P(z,w_1,w_2,t)}{\partial z} + \\
[A(z)(1 - S_1(t) - S_2(t) + S_1(t)S_2(t)) - 1] \frac{\partial P(0,w_1,w_2,t)}{\partial z} + \\
(-\lambda_1 S_1(t) - \lambda_2 S_2(t)) P(z,w_1,w_2,t) + \lambda_1 S_1 \int_0^{w_1} P(z, w_1 - y_1, w_2, t) dG_1(y_1) + \\
\lambda_2 S_2 \int_0^{w_2} P(z, w_1, w_2 - y_2, t) dG_2(y_2) + \\
A(z) \left[S_1(t)(1 - S_2(t)) \int\limits_0^{w_1} \frac{\partial P(0, w_1 - y_1, w_2, t)}{\partial z} dG_1(y_1) + \right. \\
(1 - S_1(t)) S_2(t) \int\limits_0^{w_2} \frac{\partial P(0, w_1, w_2 - y_2, t)}{\partial z} dG_2(y_2) + \\
\left. S_1(t) S_2(t) \int\limits_0^{w_1} \int\limits_0^{w_2} \frac{\partial P(0, w_1 - y_1, w_2 - y_2, t)}{\partial z} dG_1(y_1) dG_2(y_2) \right]
\end{gathered} \tag{1}$$

с начальными условиями:

$$P(z, w_1, w_2, t_0) = \begin{cases} r(z), & w_1 = w_2 = 0, \\ 0, & \text{иначе}. \end{cases} \tag{2}$$

Здесь $r(z)$ – стационарное распределение процесса $z(t)$, удовлетворяющее дифференциальному уравнению

$$r'(z) + r'(0)[A(z) - 1] = 0. \tag{3}$$

Введем частичные характеристические функции

$$h(z,u_1,u_2,t)=\int\limits_0^\infty e^{ju_1w_1}\int\limits_0^\infty e^{ju_2w_2}P(z,dw_1,dw_2,t),$$

где $j=\sqrt{-1}$. Тогда, используя обозначения $G_k^*(u_k)=\int_0^\infty e^{ju_ky_k}dG_k(y_k)$, $k=1,2$, мы можем записать уравнение:

$$\frac{\partial h(z,u_1,u_2,t)}{\partial t}=\frac{\partial h(z,u_1,u_2,t)}{\partial z}+h((z,u_1,u_2,t))\Big[\lambda_1S_1(t)(G_1^*(u_1)-1)+$$

$$\lambda_2S_2(t)(G_2^*(u_2)-1)\Big]+\frac{\partial h(0,u_1,u_2,t)}{\partial z}\Bigg\{A(z)-1+A(z)\Big[S_1(t)(G_1^*(u_1)-1)+$$

$$S_2(t)(G_2^*(u_2)-1)+S_1(t)S_2(t)(G_1^*(u_1)-1)(G_2^*(u_2)-1)\Big]\Bigg\} \tag{4}$$

с начальными условиями

$$h(z,u_1,u_2,t_0)=r(z). \tag{5}$$

Поскольку невозможно найти явный вид решения задачи (5)–(6), будем искать решение при асимптотическом условии эквивалентного роста времени обслуживания в блоках системы. Это асимптотическое условие означает пропорциональный рост среднего времени обслуживания в обоих блоках.

3. Асимптотический анализ

Обозначим $b_k=\int\limits_0^\infty[1-B_k(x)]\,dx$, $k=1,2$ – среднее время обслуживания в k-ом блоке. Тогда мы можем записать асимптотическое условие эквивалентного роста времени обслуживания в виде $b_k\to\infty$, $k=1,2$, а $\lim\limits_{b_1\to\infty}\dfrac{b_1}{b_2}=\mathrm{const}$.

Теорема 1. Асимптотическая характеристическая функция первого порядка $h^{(1)}(z,u_1,u_2,t)\approx h(z,u_1,u_2,t)$ распределения вероятностей процесса $\{z(t),W_1(t),W_2(t)\}$ при условии эквивалентно растущего времени обслуживания имеет вид

$$h^{(1)}(z,u_1,u_2,t)=r(z)\exp\left\{j\sum_{k=1}^{2}(\lambda+\lambda_k)u_ka_k\int\limits_{t_0}^{t}S_k(\tau)d\tau\right\}, \tag{6}$$

где $a_k = \int\limits_0^\infty y_k dG_k(y_k)$ – средний объем ресурса, выделяемых одной заявке в k-м блоке обслуживания, $k = 1, 2$, $\lambda = \left[\int\limits_0^\infty (1 - A(u))du\right]^{-1}$.

Теорема 2. Асимптотическая характеристическая функция второго порядка $h^{(2)}(z, u_1, u_2, t) \approx h(z, u_1, u_2, t)$ распределения вероятностей процесса $\{z(t), W_1(t), W_2(t)\}$ при условии эквивалентно растущего времени обслуживания имеет вид:

$$h^{(2)}(z, u_1, u_2, t) = r(z) \exp\left\{\frac{u_1^2}{2}\big((\lambda + \lambda_1)\alpha_1 S_1(t) + 2\varkappa_1(\lambda + \lambda_1)a_1^2 S_1^2(t)\big) + \frac{u_2^2}{2}\big((\lambda + \lambda_2)\alpha_2 S_2(t) + 2\varkappa_1(\lambda + \lambda_2)a_2^2 S_2^2(t)\big) + \frac{u_1 u_2}{2} S_1(t) S_2(t) a_1 a_2 [\lambda + 2\varkappa_1\{(\lambda + \lambda_1) + (\lambda + \lambda_2)]\right\}, \tag{7}$$

где $\varkappa_1 = \lambda^3(\sigma^2 - a^2)$, $\alpha_k = \int\limits_0^\infty y_k^2 dG_k(y_k)$, $k = 1, 2$, a и σ^2 – среднее и дисперсия случайной величины, заданной функцией распределения $A(x)$.

Полагая $t_0 \to -\infty$, а $t = T$, получаем выражение для асимптотической характеристической функции совместного распределения вероятностей занимаемых ресурсов в обоих блоках серверов в стационарном режиме:

$$h(u_1, u_2) = \exp\{j((\lambda + \lambda_1)u_1 a_1 b_1 + (\lambda + \lambda_2)u_2 a_2 b_2 + \frac{u_1^2}{2}((\lambda + \lambda_1)\alpha_1 b_1 + 2\varkappa_1(\lambda + \lambda_1)a_1^2\beta_1) + \frac{u_2^2}{2}((\lambda + \lambda_2)\alpha_2 b_2 + 2\varkappa_1(\lambda + \lambda_2)a_2^2\beta_2 + \frac{u_1 u_2}{2}\beta_{12} a_1 a_2 [\lambda + 2\varkappa_1\{(\lambda + \lambda_1) + (\lambda + \lambda_2)])\right\}, \tag{8}$$

где $\beta_k = \int\limits_0^\infty [1 - B_k(x)]^2 dx$, $k = 1, 2$, $\beta_{12} = \int\limits_0^\infty [1 - B_1(x)]\,[1 - B_2(x)]\,dx$.

Заключение

В данной статье проведено исследование гетерогенных ресурсных СМО с параллельным обслуживанием потоков разнотипных требований. Решена задача анализа суммарного объема занимаемых ресурсов каждого типа при асимптотическом условии эквивалентного роста времени обслуживания в предположении, что серверы имеют неограниченные ресурсы.Показано, что стационарное распределение вероятностей двумерного процесса суммарных объемов занятых ресурсов является гауссовским.

Очевидно, что на практике обычно нет бесконечного резерва ресурсов для использования, но в случае системы с ограниченными ресурсами можно использовать результаты для решения задачи выбора максимальных значений ресурсов, предоставляемых в каждом блоке и удовлетворяющим определенным условиям, например, заданном уровне потерь запросов из-за отсутствия ресурсов для их обслуживания. Поскольку совместное распределение вероятностей занятых ресурсов является двумерным гауссовским, то оценка оптимальных значений ресурсов в каждом блоке может быть найдена по правилу «трех сигм».

СПИСОК ЛИТЕРАТУРЫ

1. *Назаров А.А., Моисеева С.П.* Метод асимптотического анализа в теории массового обслуживания // Томск, Изд-во НТЛ 2006. С.109.
2. *Moiseev A., Moiseeva S., Lisovskaya E.* Infinite-server queueing tandem with MMPP arrivals and random capacity of customers // European Conference on Modelling and Simulation. Budapest, 2017. P. 673–679.
3. *Galileyskaya A., Lisovskaya E., Pagano M.* On the Total Amount of the Occupied Resources in the Multi-resource QS with Renewal Arrival Process // CCIS. 2019. Vo. 1109. P. 257–269.
4. *Моисеев А.Н., Назаров А.А.* Бесконечнолинейные системы и сети массового обслуживания. Томск: Изд-во НТЛ, 2015.

Бушкова Татьяна Владимировна — аспирант. E-mail: *bushkova70@mail.ru*

Моисеева Светлана Петровна — д.ф.-м.н., профессор, заведующий кафедрой теории вероятностей и математической статистики. E-mail: *smoiseeva @mail.ru*

Панкратова Екатерина Владимировна — к.ф.-м.н., старший научный сотруднк. E-mail: *pankatya86@gmail.com*

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ К РЕШЕНИЮ ЗАДАЧ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

В. М. Вишневский, А. В. Горбунова

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

В настоящем обзоре впервые представлено систематизированное изложение нового метода исследования систем массового обслуживания (СМО) с использованием алгоритмов машинного обучения. Анализ публикаций позволяет сделать вывод о высокой эффективности применения методов машинного обучения, перспективах проведения дальнейших исследований, а также о возможном выделении нового подхода в самостоятельное направление в области решения сложных задач теории очередей.
Ключевые слова: *Теория массового обслуживания, имитационное моделирование, интеллектуальный анализ данных, машинное обучение, искусственные нейронные сети.*

Введение

Математические модели сетей и систем теории очередей широко применяются при проектировании современных телекоммуникационных сетей, включая анализ характеристик существующих и перспективных сетевых протоколов, оптимизацию алгоритмов маршрутизации и топологической структуры сети и т. д. Начиная с пионерских работ А. Эрланга, А. Хиничина, Л. Клейнрока и до настоящего времени, опубликовано огромное количество статей по исследованию различных систем и сетей массового обслуживания и их применению в телекоммуникационных сетях.

В последние годы интерес исследователей в области теории очередей сместился в направлении анализа СМО с коррелированными входными потоками (MAP, BMAP, MMAP), т. к. именно такие потоки характерны для современных компьютерных сетей [1, 2]. Становлению и развитию этого научного направления способствовали работы Н. Ньютса, Г.П. Башарина, Д. Лукантони, А.Н. Дудина, В.И. Клименок и др.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований в рамках научного проекта № 19-29-06043

Традиционно основное внимание при исследовании сложных СМО направлено на анализ стационарного режима функционирования, построение многомерной марковской цепи, вывод системы дифференциальных уравнений Колмогорова и их решение с использованием аппарата преобразования Лапласа-Стилтьеса и/или производящих функций. Однако до настоящего времени остается значительное количество нерешенных задач в теории очередей, аналитическое и численное решение которых либо затруднено, либо вовсе невозможно традиционными методами. Примерами таких задач являются: исследование многолинейных приоритетных СМО с коррелированными входными MMAP-потоками и буфером ограниченного объема; анализ характеристик сетей и многофазных СМО большой размерности с входящим BMAP-потоком и произвольными функциями распределения обслуживания на фазах; исследование систем адаптивного динамического поллинга и сложных СМО типа fork-join, а также ряд задач теории надежности типа k-из-n и т. д. Отсутствие новых методов и подходов к аналитическому и численному решению таких задач сдерживает практическое применение моделей теории очередей при оценке производительности и проектировании телекоммуникационных сетей.

В настоящей работе предлагается один из таких новых подходов, базирующийся на комбинации методов машинного обучения и имитационного моделирования. Приводится краткое описание нового подхода, особенностей его применения на конкретных примерах и, как следствие, возможной перспективы его выделения в самостоятельное направление при проведении дальнейших исследований в области теории очередей. Хотя разработка и реализация нового подхода началась относительно недавно, в настоящем обзоре рассматриваются и ранние публикации, в которых данная тематика была затронута лишь косвенно. Указанные публикации заслуживают внимания, поскольку позволяют проследить развитие нового метода от его исходной точки до современного состояния. При этом строгая формулировка предложенного подхода и демонстрация его эффективности при исследовании различных моделей массового обслуживания отражены преимущественно в недавних работах авторов данного обзора.

1. Обзор публикаций в области применения методов машинного обучения к исследованию характеристик СМО

Хотя методы и алгоритмы машинного обучения нашли широчайшее применение в различных областях науки и техники [3], в том числе при исследовании современных широкополосных беспроводных сетей нового поколения [4, 5, 6], их применение в теории очередей слабо отражено в мировой литературе. В настоящей работе приводится анализ существующих публикаций в этой области, а также описание нового подхода к решению сложных задач теории очередей, базирующегося на комбинации имитационного моделирования и различных методов интеллектуального анализа данных, в частности, искусственных нейронных сетей. Указанный подход, с одной стороны, устраняет один из существенных недостатков имитационного моделирования, который заключается в значительных временных затратах на его проведение при исследовании сложных систем, и, с другой стороны, обладает одним из главных его преимуществ — универсальностью, т. е. может быть использован для исследования практически любых СМО.

Перечислим кратко основные этапы нового подхода с использованием методов машинного обучения:

— получение посредством имитационного моделирования характеристик системы для конечного набора входных параметров;
— обучение интеллектуальной модели на полученных с помощью имитационного моделирования данных одним из методов машинного обучения в рамках решения задачи прогнозирования;
— оценка искомых характеристик производительности для любых других промежуточных значений входных параметров с помощью обученной интеллектуальной модели.

К одному из первых упоминаний применения методов машинного обучения, в частности алгоритма построения дерева решений (ID3), к решению задач теории массового обслуживания можно отнести работу [7]. В более поздней работе [8] алгоритмы нейронных сетей использованы для анализа СМО с «разогревом» вида $H_2|M|M|3$ и $M|H_2|M|3$. Статьи [9, 10] посвящены построению модели классической системы массового обслуживания $M|M|1$ с помощью искусственной нейронной сети и анализу точности этого моделирования. В работах [11, 12, 13, 14] методы машинного обучения применены для оценки характеристик очередей клиентов в медицинских учреждениях и банках.

Новый подход, базирующийся на комбинации методов машинного обучения и имитационного моделирования был апробирован и эффективно использован для исследования сложных СМО в ряде последних работ авторов настоящего обзора [15, 16, 17, 18, 19, 20]. В [15] исследуется межконцевая задержка заявок многофазной системы в которой первый узел представляет собой СМО $G|G|1$, а последующие $(K-1)$ узлов — системы вида $\cdot|G|1$. Новый подход впервые применен в [16] для исследования многолинейной приоритетной СМО с буфером ограниченной емкости и входящим MMAP-потоком. В статьях [17, 18] новым методом анализируется среднее время отклика и его среднеквадратическое отклонение для СМО типа fork-join. В работе [19] исследуется результат применения нейросетей к анализу замкнутой экспоненциальной сети массового обслуживания (СеМО). Новый метод был также применен к широкому классу систем поллинга. В [20] приведено описание результатов машинного обучения для систем поллинга типа $M|M|1$ с циклическим опросом, $MAP|M|1$ с коррелированным входным потоком, а также системы типа $M|M|1$ с адаптивным динамическим опросом. В [21] алгоритмы машинного обучения задействованы при выборе оптимального управления обслуживанием заявок в многолинейной СМО. Отметим также работы [22, 23], в которых методы машинного обучения использованы для исследования тандемной СеМО как с зависимыми, так и с независимыми функциями распределения времени обслуживания пакетов на фазах системы.

Заключение

В статье приводится описание нового, перспективного подхода исследования систем и сетей массового обслуживания, базирующегося на комбинации методов машинного обучения и имитационного моделирования. Приведен обзор опубликованных в мировой литературе работ, где этот метод эффективно используется для отыскания численных характеристик сложных СМО с существенным сокращением трудоемкости и времени вычислений. Показана перспективность применения нового метода для дальнейшего исследования нерешенных задач теории массового обслуживания. Представленный подход в силу своей универсальности может вызвать значительный интерес и спровоцировать проведение множества новых исследований в области теории очередей.

СПИСОК ЛИТЕРАТУРЫ

1. *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.
2. *Lakatos L., Szeidl L., Telek M.* Introduction to Queueing Systems with Telecommunication Applications. Cham: Springer. 2013. 559 p.
3. *Bonetto R., Latzko V.* Computing in Communication Networks. Chapter 8 — Machine learning. Academic Press. 2020. P. 135–167.
4. *Morshedi M., Noll J.* Estimating PQoS of Video Streaming on Wi-Fi Networks Using Machine Learning // Sensors. 2021. Vol. 21. № 2. Art. Num.: 621. P. 1–17.
5. *Laha S., Chowdhury N., Karmakar R.* How Can Machine Learning Impact on Wireless Network and IoT? — A Survey // ICCCNT. 2020. P. 1–7.

6. *Kaur R., Kaur Sandhu J., Sapra L.* Machine Learning Technique for Wireless Sensor Networks // PDGC. 2020. P. 332–335.
7. *Khoshnevis B., Parisay S.* Machine Learning and Simulation: Application in Queuing Systems // Simulation. 1993. Vol. 61. № 5. P. 294–302.
8. *Хомоненко А. Д., Яковлев Е. Л.* Нейросетевая аппроксимация характеристик многоканальных немарковских систем массового обслуживания // Труды СПИИРАН. 2015. Вып. 41. С. 81–93.
9. *Sivakami Sundaria M., Palaniammalb S.* Simulation of M|M|1 Queuing System Using ANN // Malaya Journal of Matematik: Special Issue. 2015. № 1. P. 279–294.
10. *Sivakami Sundaria M., Palaniammalb S.* An ANN Simulation of Single Server with Infinite Capacity Queuing System // International Journal of Innovative Technology and Exploring Engineering. 2019. Vol. 8, № 12. P. 4067–4071.
11. *Mourõo R. N., Carvalho R S., Carvalho R. N. Ramos G. N.* Predicting Waiting Time Overflow on Bank Teller Queues // IEEE ICMLA. 2018. P. 842–847.
12. *Kyritsis A. I., Deriaz M.* Machine Learning Approach to Waiting Time Prediction in Queueing Scenarios // IEEE AI4I. 2019. P. 17–21.
13. *Hermanto R. P. S., Suharjito S., Nugroho A.* Waiting-Time Estimation in Bank Customer Queues using RPROP Neural Networks // Procedia Computer Science. 2018. Vol. 135. P. 35–42.
14. *Curtis C., Liu Ch., Bollerman Th. J., Pianykh O. S.* Machine Learning for Predicting Patient Wait Times and Appointment Delays // Journal of the American College of Radiology. 2018. Vol. 15. № 9. P. 1310–1316.
15. *Gorbunova A. V., Vishnevsky V. M., Larionov A. A.* Evaluation of the End-to-End Delay of a Multiphase Queuing System Using Artificial Neural Networks // Lecture Notes in Computer Science. 2020. Vol. 12563. P. 631–642.
16. *Vishnevsky V., Klimenok V., Sokolov A., Larionov A.* Performance evaluation of the priority multi-server system MMAP/PH/M/N using machine learning methods // Mathematics. 2021. In print.
17. *Gorbunova A. V., Vishnevsky V. M.* Estimating the Response Time of a Cloud Computing System with the Help of Neural Networks // Advances in Systems Science and Applications. 2020. Vol. 20. № 3. P. 105–112.
18. *Gorbunova A. V., Lebedev A. V.* Response Time Estimate for a Fork-join System with Pareto Distributed Service Time as a Model of a Cloud Computing System Using Neural Networks // Communications in Computer and Information Science. 2021. In print.
19. *Gorbunova A. V., Vishnevsky V. M.* Evaluation of the Performance Parameters of a Closed Queuing Network Using Artificial Neural Networks // Lecture Notes in Computer Science. 2021. Vol. 13144. P. 266–280.

20. *Vishnevsky V. M., Semenova O. V., Bui D. T.* Using a machine learning for analysis of polling systems with correlated arrivals // Lecture Notes in Computer Science. 2021. Vol. 13144. P. 338–347.

21. *Efrosinin D., Stepanova N.* Estimation of the Optimal Threshold Policy in a Queue with Heterogeneous Servers Using a Heuristic Solution and Artificial Neural Networks // Mathematics. 2021. Vol. 9. № 11. Art. Num.: 1267.

22. *Лазарева В. Е., Ларионов А. А., Мухтаров А. А.* Расчёт межконцевых задержек и длин очередей в многошаговой тандемной сети с применением методов машинного обучения // ITTMM. М.: РУДН, 2020. С. 43–48.

23. *Вишневский В. М., Ларионов А. А., Мухтаров А. А.* Расчёт характеристик тандемной сети с фиксированными длинами входящих пакетов методом машинного обучения // ICAM. Томск: ТГУ, 2020. С. 82–84.

Вишневский Владимир Миронович — д.т.н., профессор, главный научный сотрудник, заведующий Лабораторией № 69 «Управление сетевыми системами». E-mail: *vishn@inbox.ru*

Горбунова Анастасия Владимировна — к.ф.-м.н., старший научный сотрудник Лаборатории № 69 «Управление сетевыми системами». E-mail: *avgorbunova@list.ru*

ИССЛЕДОВАНИЕ RQ-СИСТЕМЫ MMPP/M/1 С НЕНАДЕЖНЫМ ПРИБОРОМ

Н. М. Воронина[1], С. В. Рожкова[1,2], С. П. Моисеева[2]

[1] *Национальный исследовательский Томский политехнический университет, г. Томск, Россия*
[2] *Национальный исследовательский Томский государственный университет, г. Томск, Россия*

В работе проводится исследование однолинейной RQ-системы с входящим MMPP потоком заявок и экспоненциальным законом их обслуживания с ненадежным прибором. Для исследования этой системы используется метод асимптотического анализа.
Ключевые слова: *RQ-система, MMPP-поток, ненадежный прибор.*

Введение

RQ-системы широко используются в таких системах, как центры обработки вызовов, сотовые сети и протоколы произвольного доступа в локальных сетях. Характерной чертой таких моделей является наличие повторных запросов к прибору после неудачной попытки. Такие ситуации могут быть вызваны не только отсутствием свободных серверов в моменты поступления заявок в систему, но техническими причинами [1, 2]. Если при поступлении заявки, прибор занят обслуживанием, то эта заявка переходит в источник повторных вызовов (на орбиту) и через некоторое случайное время повторяет попытку получить обслуживание. Из-за того, что нагрузка от потока повторных вызовов, как правило, является не учтенной, происходит перегрузка сети, и обслуживающий прибор может выйти из строя, что и говорит о ненадежности сервера. Многочисленные поломки и ограниченные возможности ремонта оказывают значительное влияние на производительность системы. Системы с ненадежными приборами часто являются предметом современных исследований [3, 4].

В данной работе проводится исследование Retrial Queuing System (RQ-системы) с ненадежным прибором и входящим марковски модулированным пуассоновским потоком (MMPP), то есть приборы могут время от времени выходить из строя и требовать восстановления (ремонта), только после которого они могут возобновить обслуживание заявок.

1. Постановка задачи и описание модели

Рассмотрим однолинейную RQ-систему с ненадежным прибором (рис. 1), на вход которой поступает марковский модулированный пуассоновский поток заявок (MMPP). Заявка обслуживается прибором случайное время, распределенное по экспоненциальному закону с параметром μ_1. Ненадежный прибор может быть в одном из состояний: свободен, занят обслуживанием или на ремонте. Если прибор свободен, и на вход поступает заявка, то прибор сразу же начинает обслуживание поступившей заявки. Если заявка поступает в момент, когда прибор занят, то поступившая заявка переходит на орбиту и ожидает возможности занять прибор при следующей попытке. Заявка остается на орбите в течение экспоненциально распределенного времени с параметром σ, затем заявка вновь пробует занять прибор. Время бесперебойной работы сервера распределено по экспоненциальному закону с параметром γ_1, если он простаивает, и с параметром γ_2, если сервер занят обслуживанием. Как только происходит поломка, сервер отправляется на ремонт. Все поступающие заявки переходят на орбиту. Время восстановления после ремонта распределено экспоненциально с параметром μ_2. Требуется найти стационарное распределение вероятностей числа заявок на орбите.

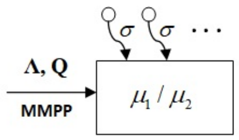

Рис. 1. Модель RQ-системы MMPP/M/1 с ненадежным прибором

Пусть $i(t)$ – число заявок на орбите в момент времени t, $n(t)$ – состояние цепи Маркова, управляющей MMPP потоком, а $k(t)$ определяет состояние прибора: $k(t) = 0$, если прибор свободен, $k(t) = 1$, если прибор занят, $k(t) = 2$, если прибор на ремонте.

Процесс $n(t)$ задается матрицей инфинитезимальных характеристик $\boldsymbol{Q} = [q_{\nu n}]$, $n = \overline{1, N}$ и диагональной матрицей $\boldsymbol{\Lambda} = \mathrm{diag}\,[\lambda_n]$ условных интенсивностей λ_n, $n = \overline{1, N}$. Трехмерный случайный процесс $\{i(t), k(t), n(t)\}$ является цепью Маркова с непрерывным временем.

Обозначим $P_k\,(i, n, t) = P\,\{i(t) = i, k(t) = k, n(t) = n\}$, $k = \{0, 1, 2\}$, $i = \overline{0, \infty}$, $n = \overline{1, N}$.

Составим систему уравнений Колмогорова для стационарных вероятностей $P_k(i, n)$.

$$\begin{cases} -\left(\lambda_n+i\sigma+\gamma_1\right)P_0\left(i,n\right)+\mu_1P_1\left(i,n\right)+\mu_2P_2\left(i,n\right)+\\ \qquad\qquad+\sum\limits_{\nu}P_0(i,\nu)\cdot q_{\nu n}=0,\\ -\left(\lambda_n+\mu_1+\gamma_2\right)P_1\left(i,n\right)+\lambda_nP_0\left(i,n\right)+\left(i+1\right)\sigma P_0\left(i+1,n\right)+\\ \qquad\qquad+\lambda_nP_1\left(i-1,n\right)+\sum\limits_{\nu}P_1(i,\nu)\cdot q_{\nu n}=0,\\ -\left(\lambda_n+\mu_2\right)P_2\left(i\right)+\gamma_1P_0\left(i,n\right)+\gamma_2P_1\left(i-1,n\right)+\lambda_nP_2\left(i-1,n\right)+\\ \qquad\qquad+\sum\limits_{\nu}P_2(i,\nu)\cdot q_{\nu n}=0.\end{cases} \tag{1}$$

2. Асимптотический анализ

Переходя к частичным характеристическим функциям $H_k(u,n)=\sum\limits_{i=0}^{\infty}e^{jui}P_k(i,n)$, $j=\sqrt{-1}$ система (1) принимает вид

$$\begin{cases} -\lambda_nH_0\left(u,n\right)+\sigma j\partial H_0(u,n)/\partial u-\gamma_1H_0\left(u,n\right)+\mu_1H_1\left(u,n\right)+\\ \qquad\qquad+\mu_2H_2\left(u,n\right)+\sum\limits_{\nu=1}^{N}H_0\left(u,\nu\right)\cdot q_{\nu n}=0,\\ -\lambda_nH_1\left(u,n\right)-\mu_1H_1\left(u,n\right)-\gamma_2H_1\left(u,n\right)+\lambda_nH_0\left(u,n\right)-\\ -j\sigma\cdot e^{-ju}\partial H_0(u,n)/\partial u+\lambda_n\cdot e^{ju}H_1\left(u,n\right)+\\ \qquad\qquad+\sum\limits_{\nu=1}^{N}H_1\left(u,\nu\right)\cdot q_{\nu n}=0,\\ -\lambda_nH_2\left(u,n\right)-\mu_2H_2\left(u,n\right)+\gamma_1H_0\left(u,n\right)+\gamma_2\cdot e^{ju}H_1\left(u,n\right)+\\ \qquad\qquad+\lambda_n\cdot e^{ju}H_2\left(u,n\right)+\sum\limits_{\nu=1}^{N}H_2\left(u,\nu\right)\cdot q_{\nu n}=0.\end{cases} \tag{2}$$

Обозначим $\mathbf{H}_n(u)=\{H_n(u,1),\ H_n(u,2),\ \ldots,\ H_n(u,N)\}$ – вектор-строка, $\mathbf{I}$ – единичная матрица, тогда система (2) принимает вид

$$\begin{cases} \mathbf{H}_0\left(u\right)\left(\mathbf{Q}-\mathbf{\Lambda}-\gamma_1\mathbf{I}\right)+j\sigma\partial\mathbf{H}_0(u)/\partial u\\ \qquad\qquad+\mu_1\mathbf{H}_1\left(u\right)+\mu_2\mathbf{H}_2\left(u\right)=0,\\ \mathbf{H}_0\left(u\right)\mathbf{\Lambda}+\mathbf{H}_1\left(u\right)\left[\left(\mathbf{Q}-(1-e^{ju})\mathbf{\Lambda}\right)-(\mu_1+\gamma_2)\mathbf{I}\right]-\\ \qquad\qquad-j\sigma e^{-ju}\partial\mathbf{H}_0(u)/\partial u=0,\\ \gamma_1\mathbf{H}_0\left(u\right)\mathbf{I}+\gamma_2e^{ju}\mathbf{H}_1\left(u\right)+\mathbf{H}_2\left(u\right)\left[\mathbf{Q}-(1-e^{ju})\mathbf{\Lambda}-\mu_2\mathbf{I}\right]=0.\end{cases} \tag{3}$$

$$\mathbf{H}_1\left(u\right)\left[-(\mathbf{\Lambda}+\gamma_2\mathbf{I})\right]\mathbf{e}+\mathbf{H}_2\left(u\right)\left(-\mathbf{\Lambda e}\right)-j\sigma e^{-ju}\mathbf{e}\partial\mathbf{H}_0(u)/\partial u=0 \qquad (4)$$

Уравнение (5) является вспомогательным и получено суммированием всех уравнений системы (4).

Теорема 1. Пусть $i(t)$ – число заявок на орбите в RQ-системе с ненадежным прибором. Тогда для последовательности характеристических функций выполняется предельное равенство

$$\lim_{\sigma\to 0}\mathbf{M}\left\{\exp\left\{jw\sigma i(t)\right\}\right\}=\exp\left\{jwG_1\right\},$$

где $G_1=\lambda(\mu_1\gamma_1+\gamma_1\gamma_2+\gamma_2\mu_2+\lambda\mu_2+\lambda\gamma_2)/(\mu_1\mu_2-\lambda\mu_2-\lambda\gamma_2)$.

Доказательство. В системе уравнений (4) и (5) введем замены $\sigma=\varepsilon,\ u=\varepsilon w,\ \mathbf{H}_k(u)=\mathbf{F}_k(w,\varepsilon)$ и в пределе при $\varepsilon\to 0$, получим

$$\begin{cases}\mathbf{F}_0\left(w,\varepsilon\right)\left(\mathbf{Q}-\mathbf{\Lambda}-\gamma_1\mathbf{I}\right)+j\partial\mathbf{F}_0(w)/\partial w+\mu_1\mathbf{F}_1\left(w,\varepsilon\right)+\\ \qquad\qquad\qquad\qquad+\mu_2\mathbf{F}_2\left(w,\varepsilon\right)=0,\\ \mathbf{F}_0\left(w,\varepsilon\right)\mathbf{\Lambda}+\mathbf{F}_1\left(w,\varepsilon\right)\left[\mathbf{Q}-(\mu_1+\gamma_2)\mathbf{I}\right]-j\partial\mathbf{F}_0(w)/\partial w=0,\\ \gamma_1\mathbf{F}_0\left(w,\varepsilon\right)+\gamma_2\mathbf{F}_1\left(w,\varepsilon\right)+\mathbf{F}_2\left(w,\varepsilon\right)\left[\mathbf{Q}-\mu_2\mathbf{I}\right]=0.\end{cases} \qquad (5)$$

$$\mathbf{F}_1\left(w,\varepsilon\right)\left[-(\mathbf{\Lambda}+\gamma_2\mathbf{I})\right]\mathbf{e}+\mathbf{F}_2\left(w,\varepsilon\right)\left(-\mathbf{\Lambda e}\right)\ -j\mathbf{e}\partial\mathbf{F}_0(w)/\partial w=0 \qquad (6)$$

Решение (6) и (4) будем искать в виде $\mathbf{F}_k(w)=\mathbf{\Phi}(w)\mathbf{R}_k,\quad k=\overline{0,N}$, где $\mathbf{R}_k$ – стационарное распределение вероятностей состояний прибора и состояний цепи Маркова. Тогда

$$\begin{cases}\mathbf{R}_0\mathbf{Qe}-\mathbf{R}_0\mathbf{\Lambda e}-\gamma_1\mathbf{R}_0\mathbf{e}-G_1\mathbf{R}_0\mathbf{e}+\mu_1\mathbf{R}_1\mathbf{e}+\mu_2\mathbf{R}_2\mathbf{e}=0,\\ \mathbf{R}_0\mathbf{\Lambda e}+\mathbf{R}_1\mathbf{Qe}-\mu_1\mathbf{R}_1\mathbf{e}-\gamma_2\mathbf{R}_1\mathbf{e}+G_1\mathbf{R}_0\mathbf{e}=0,\\ \gamma_1\mathbf{R}_0\mathbf{e}+\gamma_2\mathbf{R}_1\mathbf{e}+\mathbf{R}_2\mathbf{Qe}-\mu_2\mathbf{R}_2\mathbf{e}=0.\end{cases} \qquad (7)$$

$$-\mathbf{R}_1\mathbf{\Lambda e}-\gamma_2\mathbf{R}_1\mathbf{e}-\mathbf{R}_2\mathbf{\Lambda e}+G_1\mathbf{R}_0\mathbf{e}=0. \qquad (8)$$

Запишем условие согласованности многомерных распределений для стационарного распределения состояний прибора $\sum_{k=0}^{N}\mathbf{R}_k=\mathbf{R}$, где $\mathbf{R}$ – вектор-строка стационарного распределения цепи Маркова, управляющей входящим потоком, который удовлетворяет системе уравнений $\mathbf{RQ}=0,\quad \mathbf{Re}=1$.

Обозначая $\mathbf{R}_0\mathbf{e}=r_0,\mathbf{R}_1\mathbf{e}=r_1,\mathbf{R}_2\mathbf{e}=r_2$, найдем r_0, r_1, r_2, G_1.

$$r_0=\frac{\mu_1\mu_2-\lambda\mu_2-\lambda\gamma_2}{\mu_1\left(\gamma_1+\mu_2\right)},\quad r_1=\frac{\lambda}{\mu_1},\quad r_2=\frac{\mu_1\gamma_1-\lambda\gamma_1+\lambda\gamma_2}{\mu_1\left(\gamma_1+\mu_2\right)},$$

$G_1=\lambda(\mu_1\gamma_1+\gamma_1\gamma_2+\gamma_2\mu_2+\lambda\mu_2+\lambda\gamma_2)/(\mu_1\mu_2-\lambda\mu_2-\lambda\gamma_2)$. ■

Для того, чтобы вероятности r_0, r_1, r_2 были положительными, добавим ограничения $\mu_1 > \lambda$, $\mu_2 > \frac{\lambda\gamma_2}{\mu_1 - \lambda}$.
Теорема 1 определяет асимптотическое среднее числа заявок на орбите в RQ-системе MMPP/M/1 в предельном условии большой задержки заявок на орбите. Рассмотрим асимптотику второго порядка для более полного исследования процесса.

Теорема 2. Пусть *i(t)* – число заявок на орбите в RQ-системе с ненадежным прибором, тогда для последовательности характеристических функций выполняется равенство

$$\lim_{\sigma\to 0} \mathbf{M}\left\{\exp\left\{jw\sqrt{\sigma}\left(i(t) - G_1/\sigma\ \right)\right\}\right\} = \exp\left\{(jw)^2 G_2/2\right\},$$

где $G_2 = (r_0 G_1\mathbf{I} - G_1\mathbf{u}_0 + \mathbf{\Lambda}\mathbf{u}_1 + \gamma_2\mathbf{u}_1 + \mathbf{\Lambda}\mathbf{u}_2)(G_1\mathbf{z}_0 - \mathbf{\Lambda}\mathbf{z}_2 + r_0\mathbf{I})^{-1}$.

Доказательство. В системе уравнений (7) и уравнении (8) перейдем к характеристической функции $\mathbf{H}_k(u) = e^{ju\frac{G_1}{\sigma}}\mathbf{H}_k^{(2)}(u)$ и сделаем замены $\sigma = \varepsilon^2$, $u = \varepsilon w$, $\mathbf{H}_k^{(2)}(u) = \mathbf{F}_k^{(2)}(w,\varepsilon)$, получим

$$\begin{cases} \mathbf{F}_0^{(2)}(w,\varepsilon)(\mathbf{Q} - \mathbf{\Lambda} - \gamma_1\mathbf{I} - G_1\mathbf{I}) + j\varepsilon\partial\mathbf{F}_0^{(2)}(w,\varepsilon)/\partial w + \\ \qquad + \mu_1\mathbf{F}_1^{(2)}(w,\varepsilon) + \mu_2\mathbf{F}_2^{(2)}(w,\varepsilon) = 0, \\ \mathbf{F}_0^{(2)}(w,\varepsilon)\left(\mathbf{\Lambda} + e^{-jw\varepsilon}G_1\mathbf{I}\right) + \mathbf{F}_1^{(2)}(w,\varepsilon)((\mathbf{Q} - (1 - e^{jw\varepsilon})\mathbf{\Lambda}) - \\ \qquad - (\mu_1 + \gamma_2)\mathbf{I}) - j\varepsilon e^{-jw\varepsilon}\partial\mathbf{F}_0^{(2)}(w,\varepsilon)/\partial w = 0, \\ \gamma_1\mathbf{F}_0^{(2)}(w,\varepsilon) + \gamma_2 e^{jw\varepsilon}\mathbf{F}_1^{(2)}(w,\varepsilon) + \mathbf{F}_2^{(2)}(w,\varepsilon)(\mathbf{Q} - (1 - e^{jw\varepsilon})\mathbf{\Lambda} - \\ \qquad - \mu_2\mathbf{I}) = 0, \\ \mathbf{F}_1^{(2)}(w,\varepsilon)\left[-(\mathbf{\Lambda} + \gamma_2\mathbf{I})\right]\mathbf{e} + \mathbf{F}_2^{(2)}(w,\varepsilon)(-\mathbf{\Lambda}\mathbf{e}) + \\ \qquad + G_1 e^{-jw\varepsilon}\mathbf{F}_0^{(2)}(w,\varepsilon)\mathbf{e} - j\varepsilon e^{-jw\varepsilon}\mathbf{e}\partial\mathbf{F}_0^{(2)}(w,\varepsilon)/\partial w = 0. \end{cases} \quad (9)$$

Решение системы уравнений (9) будем искать в виде
$\mathbf{F}_k(w,\varepsilon) = \mathbf{\Phi}_2(w)\left\{\mathbf{R}_k + jw\varepsilon\mathbf{f}_k\right\} + O(\varepsilon^2)$, тогда

$$\begin{cases} -\mathbf{f}_0\mathbf{\Lambda}\mathbf{e} - \gamma_1\mathbf{f}_0\mathbf{e} - G_1\mathbf{f}_0\mathbf{e} - G_2\mathbf{R}_0\mathbf{e} + \mu_1\mathbf{f}_1\mathbf{e} + \mu_2\mathbf{f}_2\mathbf{e} = 0, \\ -G_1\mathbf{R}_0\mathbf{e} + \mathbf{f}_0\mathbf{\Lambda}\mathbf{e} + G_1\mathbf{f}_0\mathbf{e} + G_2\mathbf{R}_0\mathbf{e} + \mathbf{R}_1\mathbf{\Lambda}\mathbf{e} - \mu_1\mathbf{f}_1\mathbf{e} - \gamma_2\mathbf{f}_1\mathbf{e} = 0, \\ \gamma_1\mathbf{f}_0\mathbf{e} + \gamma_2\mathbf{R}_1\mathbf{e} + \gamma_2\mathbf{f}_1\mathbf{e} + \mathbf{R}_2\mathbf{\Lambda}\mathbf{e} - \mu_2\mathbf{f}_2\mathbf{e} = 0, \\ -\mathbf{f}_1\mathbf{\Lambda}\mathrm{e} - \gamma_2\mathbf{f}_1\mathbf{e} - \mathbf{f}_2\mathbf{\Lambda}\mathbf{e} - G_1\mathbf{R}_0\mathbf{e} + G_1\mathbf{f}_0\mathbf{e} + G_2\mathbf{R}_0\mathbf{e} = 0. \end{cases}$$

Введем обозначения $\mathbf{g}_0\mathbf{e} = \mathbf{z}_0$, $\mathbf{g}_1\mathbf{e} = \mathbf{z}_1$, $\mathbf{g}_2\mathbf{e} = \mathbf{z}_2$, $\mathbf{y}_0\mathbf{e} = \mathbf{u}_0$, $\mathbf{y}_1\mathbf{e} = \mathbf{u}_1$, $\mathbf{y}_2\mathbf{e} = \mathbf{u}_2$. Решение ищем в виде $\mathbf{f}_k = C\mathbf{R}_k + G_2\mathbf{g}_k + \mathbf{y}_k$. Тогда с учетом дополнительных условий $\sum_{k=0}^{N}\mathbf{g}_k\mathbf{e} = 0$, $\sum_{k=0}^{N}\mathbf{y}_k\mathbf{e} = 0$, получим

$$\mathbf{z}_0 = -r_0(\mathbf{\Lambda} + G_1\mathbf{I})^{-1}, \quad \mathbf{z}_1 = 0, \quad \mathbf{z}_2 = -[r_0\gamma_1(\mathbf{\Lambda} + G_1\mathbf{I})^{-1}]/\mu_2,$$
$$\mathbf{u}_0 = [r_0G_1\mathbf{I} - r_1\lambda_1\mathbf{I} - (\mu_1+\gamma_2)(r_1\gamma_2\mathbf{I} + r_2\lambda_2\mathbf{I} - r_0G_1\mathbf{I})/\mu_1] \times$$
$$\times(\mathbf{\Lambda} + G_1\mathbf{I})^{-1},$$
$$\mathbf{u}_1 = (-r_1\gamma_2\mathbf{I} - r_2\lambda_2\mathbf{I} + r_0G_1\mathbf{I})/\mu_1,$$
$$\mathbf{u}_2 = (r_1\gamma_2\mathbf{I} + r_2\lambda_2\mathbf{I} + \gamma_1\mathbf{u}_0 + \gamma_2\mathbf{u}_1)/\mu_2,$$
$$G_2 = (r_0G_1\mathbf{I} - G_1\mathbf{u}_0 + \mathbf{\Lambda}\mathbf{u}_1 + \gamma_2\mathbf{u}_1 + \mathbf{\Lambda}\mathbf{u}_2)(G_1\mathbf{z}_0 - \mathbf{\Lambda}\mathbf{z}_2 + r_0\mathbf{I})^{-1}. \blacksquare$$

Теорема 2 показывает, что распределение числа заявок на орбите является гауссовским с математическим ожиданием G_1/σ, и дисперсией G_2/σ.

Заключение

В работе рассмотрена MMPP/M/1 RQ-система с ненадежным прибором. В ходе исследования была получена асимптотическая характеристическая функция распределения числа заявок на орбите. Доказано, что при условии большой задержки на орбите стационарное распределение вероятностей числа заявок на орбите может быть аппроксимировано гауссовским распределением.

СПИСОК ЛИТЕРАТУРЫ

1. *Artalejo J. R. and Gomez-Corral A.* Retrial queueing systems: a computational approach. Springer, 2008.
2. *Falin G. I. and Templeton J. G. C.* Retrial Queues. London: Chapman and Hall, 1997.
3. *Nazarov A. A., Paul S. V., Lizyura O. D.* Two-way communication retrial queue with unreliable server and multiple types of outgoing calls // Discrete and Continuous Models and Applied Computational Science. 2020. Vol. 28. № 1. P. 49–61.
4. *Lamia Lakaour, Djamil Aissani, Karima Adel-Aissanou, Kamel Barkaoui, Sofiane Ziani* An unreliable single server retrial queue with collisions and transmission errors // Communications in Statistics – Theory and Methods, Taylor & Francis. 2020. P. 1–25.

Воронина Наталия Михайловна — аспирант, старший преподаватель отделения информационных технологий ИШИТР ТПУ. E-mail: *vnm@tpu.ru*

Рожкова Светлана Петровна — д-р физ.-мат. наук, доцент, профессор ОМИ ШБИП ТПУ, кафедры ТВиМС ИПМКН ТГУ. E-mail: *rozhkova@tpu.ru*

Моисеева Светлана Петровна — д-р физ.-мат. наук, профессор, профессор кафедры ТВиМС ИПМКН ТГУ. E-mail: *smoiseeva@mail.ru*

СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ПЕРЕМЕЩАЮЩИМИСЯ НЕТЕРПЕЛИВЫМИ ПОЛЬЗОВАТЕЛЯМИ

А. Н. Дудин[1,2], О. С. Дудина[1], С. А. Дудин[1]

[1] *Белорусский государственный университет, Минск, Беларусь*
[2] *Российский университет дружбы народов (РУДН), Москва, Россия*

Рассматривается модель соты сети мобильной связи. Сота виртуально разделена на зоны в зависимости от силы сигнала от базовой станции и применяемой схемы модуляции. Поступающие пользователи делятся на два типа: новые пользователи и пользователи, перемещающиеся в соту из другой соты в процессе обслуживания (хэндовер пользователи). Для хэндовер пользователей имеется конечный буфер для ожидания. Время обслуживания каждого пользователя имеет экспоненциальное распределение с зависящим от зоны нахождения параметром. Во время обслуживания пользователи могут перемещаться и менять зоны обслуживания. Число пользователей, которые могут одновременно обслуживаться в соте, является конечным. Построена многомерная цепь Маркова, описывающая процесс изменения состояний соты. Выписан инфинитезимальный генератор данной цепи, и найдены основные характеристики производительности соты.

Ключевые слова: *Маркированный марковский входной поток, перемещающиеся пользователи, сота мобильной сети.*

Введение

Системы и сети массового обслуживания хорошо моделируют процессы распределения ресурсов в телекоммуникационных сетях и ввиду этого продолжают быть актуальными объектами для исследования. Относительно недавно начали проводиться работы по аналитическому моделированию сот сетей связи, в которых скорость обслуживания пользователей зависит от силы сигнала, получаемого от базовой станции, за счет использования разных схем модуляции. Обзор литературы по тематике вопроса представлен, например, в работах [1] и [2].

Обычно предполагается, что сота разделена на неперекрывающиеся зоны, где зона определяется как место, в которой используется опреде-

Публикация выполнена при поддержке Программы стратегического академического лидерства РУДН

ленная схема модуляции. Это разделение определяется расстоянием до базовой станции, наличием естественных или искусственных препятствий для распространения сигнала, уровнем существующего шума и т.д. Так как пользователи являются мобильными, во время обслуживания они могут перемещаться и менять неограниченное число раз зоны обслуживания. Когда пользователь переходит из одной зоны в другую во время обслуживания, скорость обслуживания также изменяется. Количество пользователей, которые могут одновременно обслуживаться в соте, предполагается конечным. Если в момент создания соединения от нового пользователя количество пользователей в соте уже достигло предела, то это соединение блокируется, и запрос от пользователя считается потерянным. Необходимо исследовать стационарное поведение системы для выяснения вида зависимостей основных характеристик качества обслуживания пользователей от параметров системы, что позволяет решать определенные оптимизационные задачи.

В данной работе, также как и [2], предполагается, что поступающие пользователи делятся на два вида: новые пользователи, которые совершают попытку начать обслуживание изнутри соты, и хэндовер пользователи (ХП), перемещающиеся в соту из одной из соседних сот в процессе обслуживания. Если количество пользователей достигло предела, то поступающие ХП не получают мгновенный отказ, как новые пользователи. Для кратковременного хранения ХП имеется буфер конечной емкости, поскольку данные пользователи при перемещении в новую соту могут еще какое-то время поддерживать связь с исходной сотой. Время нахождения в данном буфере ограничено в виду того, что в результате перемещения связь с исходной сотой может быть потеряна. Поэтому авторы предполагают, что ХП являются нетерпеливыми.

В отличие от работы [2], где предполагалось, что новые и хэндовер пользователи поступают в разных потоках, в данной работе мы полагаем, что поступления этих видов пользователей являются зависимыми и описываются одним маркированным марковским входным потоком. Данное обобщение является логичным, поскольку времена максимальной активности (дневное время) и пассивности (ночное время) обычных и хэндовер пользователей в реальных системах совпадают. Кроме этого, предложенное обобщение приводит к существенному уменьшению пространства состояний многомерной цепи Маркова, описывающей процесс изменения состояний системы, что в свою очередь уменьшает время, необходимое для проведения расчетов на ПК.

1. Математическая модель и процесс состояний системы

Рассмотрим соту, состоящую из $R (1 < R < \infty)$ непересекающихся зон. Зоны пронумерованы в порядке убывания силы сигнала. В качестве модели соты рассматривается система массового обслуживания, состоящая из N одинаковых приборов и конечного буфера емкости K. Мы делим всех новых пользователей на R типов, в зависимости от зоны поступления, а также считаем, что ХП имеют тип $R+1$. Обслуживаемые пользователи поделены на R типов, в зависимости от зоны нахождения.

Процесс активации новых пользователей в соте описывается маркированным марковским входным потоком ($MMAP$), определяемым следующим образом. Пусть $\nu_t,\ t \geqslant 0$, – неприводимая цепь Маркова с непрерывным временем и пространством состояний $\{1, 2, ..., W\}$. Интенсивности переходов этой цепи Маркова определяются генератором D. Матрица D является суммой квадратных матриц $D_r,\ r = \overline{0, R+1}$, т.е. $D = \sum\limits_{r=0}^{R+1} D_r$. Элементы неотрицательной матрицы D_r, $r = \overline{1, R+1}$, определяют интенсивности переходов цепи ν_t, которые сопровождаются поступлением пользователя типа r, $r = \overline{1, R+1}$. Недиагональные элементы матрицы D_0 определяют интенсивности переходов цепи Маркова ν_t, которые не сопровождаются приходом пользователей. Средняя интенсивность λ_r, $r = \overline{1, R+1}$, пользователей типа r равна $\lambda_r = \boldsymbol{\theta} D_r \mathbf{e}$, где $\boldsymbol{\theta}$ – вектор-строка стационарных вероятностей цепи ν_t, а $\mathbf{e}$ – вектор-столбец, состоящий из единиц. Общая интенсивность λ поступления пользователей определяется как $\lambda = \sum\limits_{r=1}^{R+1} \lambda_r$.

Предлагается, что ХП, который прибывает, когда все приборы заняты, отправляется в буфер. Если буфер заполнен, пользователь теряется. ХП, находящиеся в буфере, нетерпеливы и навсегда покидают систему, если не начали обслуживаться через время, имеющее экспоненциальное распределение с параметром α, $\alpha > 0$. Если ХП принимается на обслуживание, предполагается, что он поступает в самую удаленную зону R.

По аналогии с [2], для определения суммарного времени обслуживания пользователя рассмотрим непрерывную цепь Маркова η_t, $t \geqslant 0$, имеющую R переходных состояний $\{1, 2, \dots, R\}$ и три поглощающих состояния $R+1, R+2$ и $R+3$. Определим полное время обслуживания пользователя в соте как время, за которое цепь Маркова η_t, $t \geqslant 0$, достигает одного из поглощающих состояний. Начальное состояние цепи η_t выбирается в момент начала обслуживания пользователя и определяется зоной поступления пользователя. Т.е., начальное состояние цепи η_t определяется компонентами вектора $\boldsymbol{\beta}^{(r)} = (\underbrace{0,\ldots,0}_{r-1},1,\underbrace{0,\ldots,0}_{R-r})$, $r = \overline{1,R}$, если пользователь начинает обслуживание в r-ой зоне. Моменты смены зоны пользователем соответствуют моментам переходов цепи η_t в другое переходное состояние. Интенсивности переходов процесса η_t между переходными состояниями определяются недиагональными элементами субгенератора S. Переходы цепи η_t в поглощающие состояния $R+1$, $R+2$, и $R+3$ соответствуют успешному завершению обслуживания, уходу пользователя из-за неудовлетворенностью качеством обслуживания и уходу пользователя из соты из-за хэндовера, соответственно. Интенсивности процесса перехода η_t в поглощающее состояние $R+1$ определяются компонентами вектора $\mathbf{S}_{serv}$, в состояние $R+2$ – вектора $\mathbf{S}_{out}$, и в поглощающее состояние $R+3$ – компонентами вектора $\mathbf{S}_{hand}$. Подробное описание этого процесса обслуживания представлено в [2].

Поведение исследуемой системы можно описать цепью Маркова с непрерывным временем $\zeta_t = \{n_t, \nu_t, v_t, \eta_t^{(1)}, \ldots, \eta_t^{(R)}\}$, $t \geqslant 0$, где, в момент t, n_t – число пользователей в системе, $n_t = \overline{0, N+K}$; ν_t – состояние управляющего процесса $MMAP$, определяющего процесс поступления пользователей, $\nu_t = \overline{1,W}$; $\eta_t^{(r)}$ – число пользователей, обслуживаемых в r-ой зоне, $\eta_t^{(r)} = \overline{0, \min\{n_t, N\}}$, $\sum\limits_{r=1}^{R} \eta_t^{(r)} = \min\{n_t, N\}$, $r = \overline{1,R}$.

Цепь Маркова $\zeta_t,\ t \geqslant 0$, регулярная неприводимая и имеет конечное пространство состояний. Следовательно, стационарные вероятности состояний цепи $\pi(n,\nu,\eta^{(1)},\ldots,\eta^{(R)}) = \lim\limits_{t\to\infty} P\{n_t = n, \nu_t = \nu, \eta_t^{(1)} = \eta^{(1)}, \ldots, \eta_t^{(R)} = \eta^{(R)}\}$ существуют для любых значений параметров рассматриваемой системы.

Сформируем векторы-строки $\boldsymbol{\pi}_n$, $n = \overline{0, N+K}$, из этих вероятностей, перенумерованных в обратном лексикографическом порядке компонент $\eta_t^{(1)}, \ldots, \eta_t^{(R)}$ и прямом лексикографическом порядке компоненты ν_t. Хорошо известно, что данные векторы удовлетворяют следующей системе линейных алгебраических уравнений Чепмена-Колмогорова:

$$(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{N+K})Q = \mathbf{0}, \ \ (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{N+K})\mathbf{e} = 1, \tag{1}$$

где Q – инфинитезимальный генератор цепи Маркова $\zeta_t,\ t \geqslant 0$.

Для вычисления векторов стационарных вероятностей состояний системы необходимо получить генератор Q.

Теорема 1. Генератор Q цепи Маркова $\zeta_t,\ t \geqslant 0$, имеет блочно-трехдиагональный вид. Ненулевые блоки определяются как:

$$Q_{0,0} = D_0,\ Q_{n,n} = D_0 \oplus (A_n(S) + \Delta_n),\ n = \overline{1, N-1},$$

$$Q_{n,n} = D_0 \oplus (A_N(S) + \Delta_N) + (D - D_0 - D_{R+1}) \otimes I_{J_N} -$$
$$-(n-N)\alpha I_{WJ_N},\ n = \overline{N, N+K-1},$$
$$Q_{N+K,N+K} = D_0 \oplus (A_N(S) + \Delta_N) + (D - D_0) \otimes I_{J_N} - K\alpha I_{WJ_N},$$
$$Q_{n,n-1} = I_W \otimes (L_n(\mathbf{S}_{serv}) + L_n(\mathbf{S}_{out}) + L_n(\mathbf{S}_{hand})),\ n = \overline{1, N},$$
$$Q_{n,n-1} = I_W \otimes ((L_N(\mathbf{S}_{serv}) + L_N(\mathbf{S}_{out}) + L_N(\mathbf{S}_{hand})) \times P_{N-1}(\boldsymbol{\beta}_R)) +$$
$$+(n-N)\alpha I_{WJ_N},\ n = \overline{N+1, N+K},$$
$$Q_{n,n+1} = \sum_{r=1}^{R} D_r \otimes P_n(\boldsymbol{\beta}_r) + D_{R+1} \otimes P_n(\boldsymbol{\beta}_R),\ n = \overline{0, N-1},$$
$$Q_{n,n+1} = D_{R+1} \otimes I_{J_N},\ n = \overline{N, N+K-1},$$

где матрица $L_n(\tilde{\mathbf{S}})$, $\tilde{\mathbf{S}} \in \{\mathbf{S}_{serv}, \mathbf{S}_{out}, \mathbf{S}_{hand}\}$, определяет интенсивности переходов в момент перехода процесса обслуживания одного из пользователей в соответствующее поглощающее состояние $R + l$, $l = 1, 2, 3$, $n = \overline{1, N}$ (см. [2]); матрица $A_n(S)$, $n = \overline{1, N}$, определяет интенсивности переходов в момент перемещения одного из пользователей в другую зону (см. [2]); матрица $P_n(\boldsymbol{\beta}^{(r)})$ определяет интенсивности переходов в момент начала обслуживания в зоне, r, $r = \overline{1, R}$, в момент, когда n, $n = \overline{0, N-1}$, пользователей находятся на обслуживании (см. [2]); $\Delta^{(n)} = -\mathrm{diag}\{A_n(S)\mathbf{e} + L_n(\mathbf{S}_{serv})\mathbf{e} + L_n(\mathbf{S}_{out})\mathbf{e} + L_n(\mathbf{S}_{hand})\mathbf{e}\}$, $n = \overline{1, N}$, где $\mathrm{diag}\{\dots\}$ обозначает диагональную матрицу с диагональными элементами, заданными вектором, представленным в скобках; $J_n = \frac{(n+R-1)!}{n!(R-1)!}$; $\otimes$ и $\oplus$ – символы Кронекерова произведения и суммы матриц.

2. Характеристики производительности системы

После вычисления векторов $\boldsymbol{\pi}_n$, $n = \overline{0, N+K}$, можно вычислить различные характеристики производительности соты.

Среднее количество пользователей, получающих обслуживание в соте $N_{user} = \sum\limits_{n=1}^{N+K} \min\{n, N\}\boldsymbol{\pi}_n\mathbf{e}$. Среднее количество ХП, находящихся в буфере $N_{hand} = \sum\limits_{n=N+1}^{N+K} (n-N)\boldsymbol{\pi}_n\mathbf{e}$. Среднее количество пользователей, находящихся в системе $L = N_{user} + N_{hand}$. Интенсивность выходного потока успешно обслуженных пользователей $s_+ = \sum\limits_{n=1}^{N+K} \boldsymbol{\pi}_n(I_W \otimes$

$L_{\min\{n,N\}}(\mathbf{S}_{serv}))\mathbf{e}$. Интенсивность выходного потока пользователей, которые прекращают обслуживание из-за неудовлетворительного качества $s_- = \sum\limits_{n=1}^{N+K} \boldsymbol{\pi}_n(I_W \otimes L_{\min\{n,N\}}(\mathbf{S}_{out}))\mathbf{e}$. Выходной поток пользователей, покидающих систему из-за хэндовера $s_h = \sum\limits_{n=1}^{N+K} \boldsymbol{\pi}_n(I_W \otimes L_{\min\{n,N\}}(\mathbf{S}_{hand}))\mathbf{e}$. Вероятность потери нового пользователя по прибытию в систему $P_{new}^{ent-loss} = (\lambda - \lambda_{R+1})^{-1} \sum\limits_{n=N}^{N+K} \boldsymbol{\pi}_n((D - D_0 - D_{R+1}) \otimes I_{J_N})\mathbf{e}$. Вероятность потери нового пользователя типа r по прибытию в систему $P_r^{ent-loss} = \lambda_r^{-1} \sum\limits_{n=N}^{N+K} \boldsymbol{\pi}_n(D_r \otimes I_{J_N})\mathbf{e}$, $r = \overline{1,R}$. Вероятность потери ХП по прибытию в систему $P_{hand}^{ent-loss} = \lambda_{R+1}^{-1} \boldsymbol{\pi}_{N+K}(D_{R+1} \otimes I_{J_N})\mathbf{e}$. Вероятность ухода пользователя из рассматриваемой соты в другую $P^{hand} = \frac{s_h}{\lambda}$. Вероятность вынужденного ухода пользователя из-за неудовлетворенности качеством обслуживания $P^{diss-loss} = \frac{s_-}{\lambda}$. Вероятность потери ХП из буфера из-за нетерпеливости $P_{hand}^{imp-loss} = \frac{\alpha N_{hand}}{\lambda_{R+1}}$. Вероятность потери ХП вычисляется как $P_{hand}^{loss} = P_{hand}^{ent-loss} + P_{hand}^{imp-loss}$. Вероятность потери пользователя $P^{loss} = 1 - \frac{s_+ + s_h}{\lambda}$.

Заключение

В работе исследована система обслуживания с перемещающимися пользователями. Результаты могут быть применены для моделирования и оптимизации соты мобильной сети.

СПИСОК ЛИТЕРАТУРЫ

1. *Sas B., Bernal-Mor E., Spaey K., Pla V., Blondia C., Martinez-Bauset J.* Modelling the time-varying cell capacity in LTE networks // Telecommunication Systems. 2019. T. 55. C. 299-313.
2. *Kim C., Dudin A., Dudin S., Dudina O.* Mathematical Model of Operation of a Cell of a Mobile Communication Network With Adaptive Modulation Schemes and Handover of Mobile Users // IEEE Access. 2021. T. 9. C. 106933-106946.

Дудин Александр Николаевич — д-р физ.-мат. наук, профессор. E-mail: *dudin_alexander@mail.ru*

Дудина Ольга Сергеевна — канд. физ.-мат. наук, ведущий научный сотрудник, НИЛ ПВА БГУ. E-mail: *dudina@bsu.by*

Дудин Сергей Александрович — канд. физ.-мат. наук, ведущий научный сотрудник, НИЛ ПВА БГУ. E-mail: *dudins@bsu.by*

СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ДВУМЯ ТИПАМИ ЗАПРОСОВ И ОГРАНИЧЕННЫМ РАЗДЕЛЕНИЕМ ПРОЦЕССОРА

В. И. Клименок[1], А. Н. Дудин[1,2]

[1] *Белорусский государственный университет, г. Минск, Беларусь*
[2] *Российский университет дружбы народов, г. Москва, Россия*

В данной работе исследуется система массового обслуживания с разделением процессора. Запросы двух типов поступают в систему в маркированном марковском потоке. Предполагается, что число запросов каждого из типов, одновременно находящихся на обслуживании, ограничено. Времена обслуживания запросов распределены по фазовому закону, параметры которого зависят как от типа запроса, так и от числа запросов данного типа, находящихся в системе. Функционирование системы описано в терминах многомерной цепи Маркова. Построен инфинитезимальный генератор цепи, что позволяет вычислить стационарные вероятности и основные характеристики производительности системы.

Ключевые слова: *Система массового обслуживания, маркированный марковский поток, разделение процессора, стационарное распределение.*

1. Введение

Технология разделения процессора очень популярна в компьютерных системах и телекоммуникационных сетях. Чаще всего считается, что процессор может использоваться неограниченным числом пользователей, входной поток является стационарным пуассоновским, а времена обслуживания распределены по экспоненциальному закону. Рассмотренная в данной работе система массового обслуживания существенно расширяет возможности моделирования реальных систем с разделением процессора. Мы считаем, что существуют ограничения на число пользователей разных типов, одновременно находящихся на обслуживании, и не вводим ограничительных предположений, таких как однородность и некоррелированный характер потока задач, а также экспоненциальное распределение времен обслуживания запросов разных типов. Мы предполагаем, что входной поток в систему является неоднородным коррелированным и описывается маркированным марковским потоком (общепринятое название $MMAP$-Marked Markovian Arrival Process), введенным в статье [1]. В целях более адекватного описания процесса обслуживания мы используем распределение фазового типа (PH-Phase type distribution), см., например, [6], успешно используемого для аппроксимации произвольного распределения.

2. Математическая модель

Рассматривается однолинейная система массового обслуживания без буфера. Запросы разных типов поступают в $MMAP$ -потоке под управлением неприводимой цепи Маркова с непрерывным временем ν_t, $t \geq 0$, которая принимает значения в множестве $\{0, 1, 2, \ldots, W\}$ и называется управляющим процессом $MMAP$. Цепь ν_t пребывает в состоянии ν в течение экспоненциально распределенного времени с параметром λ_ν, $\nu = \overline{0, W}$, после чего с вероятностью $p_k(\nu, \nu')$ переходит в состояние ν' и генерируется заявка k-го типа, $k \in \{1, 2, \ldots, K\}$, или, с вероятностью $p_0(\nu, \nu')$, цепь переходит в состояние ν' без генерации заявки, причем $p_0(\nu, \nu) = 0$. Всю информацию и $MMAP$-е удобно хранить в виде набора матриц D_k, $k = \overline{1, K}$, порядка $(W + 1) \times (W + 1)$, элементы которых определяются как

$$(D_k)_{\nu,\nu'} = \lambda_\nu p_k(\nu, \nu'), \quad \nu, \nu' = \overline{0, W}, k = \overline{1, K},$$

$$(D_0)_{\nu,\nu'} = \begin{cases} \lambda_\nu p_0(\nu, \nu'), & \nu \neq \nu', \nu, \nu' = \overline{0, W}, \\ -\lambda_\nu, & \nu = \nu' = \overline{0, W}. \end{cases}$$

Подробное описание $MMAP$ потока можно найти, например, в [1].

В данной работе мы предполагаем, что в систему поступают запросы двух типов, т.е., $K = 2$. На обслуживающем приборе одновременно могут обслуживаться до N запросов 1 типа и до R запросов 2 типа. Если на приборе обслуживается только один запрос l-го типа, $l = 1, 2$, то его время обслуживания имеет PH распределение, заданное неприводимым представлением $(\boldsymbol{\beta}_l, S_l)$ и управляющим процессом $m_t^{(l)}$, $t \geq 0$, с пространством состояний $\{1, \ldots, M_l, M_l + 1\}$, где состояние $M_l + 1$ является поглощающим. Интенсивности переходов в поглощающее состояние определяются вектором-столбцом $\boldsymbol{S}_0^{(l)} = -S_l\mathbf{e}$. Более подробное описание PH распределения можно найти, например, в [6].

Запросы каждого из типов делят отведенную им пропускную способность обслуживающего прибора поровну. Если на приборе одновременно обслуживается n_l запросов l-го типа, то время обслуживания любого из этих запросов имеет PH распределение, заданное неприводимым представлением $(\boldsymbol{\beta}_l, \frac{1}{n_l} S_l)$ и управляющим процессом $m_t^{(l)}$, $t \geq 0$, с пространством состояний $\{1, \ldots, M_l, M_l + 1\}$, где состояние $M_l + 1$ является поглощающим. Интенсивности переходов в поглощающее состояние определяются вектором-столбцом $\frac{1}{n_l}\boldsymbol{S}_0^{(l)}$.

Если поступающий запрос 1 типа застает на приборе $n < N$ запросов, то он поступает на обслуживание. При этом пропускная способность прибора, отведенная запросам 1-го типа, делится поровну между $n+1$ запросами. В противном случае запрос покидает систему необслуженным (теряется). Аналогично, если запрос 2-го типа застает на приборе $r < R$ запросов, то он поступает на обслуживание. При этом пропускная способность прибора, отведенная запросам 2 типа, делится поровну между $r+1$ запросами. В противном случае запрос теряется.

3. Цепь Маркова, описывающая поведение системы

Функционирование системы описывается цепью Маркова

$$\xi_t = \{n_t, r_t, \eta_t^{(1)}, \eta_t^{(2)}, \dots, \eta_t^{(M_1)}, \tau_t^{(1)}, \tau_t^{(2)}, \dots, \tau_t^{(M_2)}, \nu_t\},$$

где

— n_t, r_t – число заявок 1 и 2 типа на приборе, $n_t = \overline{0, N}$, $r_t = \overline{0, R}$;

— $\eta_t^{(m^{(1)})}$ – количество заявок 1 типа, которые обслуживаются на фазе $m^{(1)}$, $\eta_t^{(m^{(1)})} = \overline{0, n_t}$, $m^{(1)} = \overline{1, M_1}$;

— $\tau_t^{(m^{(2)})}$ – количество заявок 2 типа, которые обслуживаются на фазе $m^{(2)}$, $\tau_t^{(m^{(2)})} = \overline{0, r_t}$, $m^{(2)} = \overline{1, M_2}$;

— ν_t – состояние управляющего процесса $MMAP$, $\nu_t = \overline{0, W}$,

в момент времени t.

Введем процессы

$$\mathbf{u}_t^{(1)} = \{\eta_t^{(1)}, \eta_t^{(2)}, \dots, \eta_t^{(M_1)}\}; \ \mathbf{u}_t^{(2)} = \{\tau_t^{(1)}, \tau_t^{(2)}, \dots, \tau_t^{(M_2)}\}.$$

Упорядочим состояния рассматриваемой цепи Маркова $\xi_t, t \geqslant 0$, следующим образом. Перенумеруем компоненты n_t, r_t в прямом лексикографическом порядке и, при фиксированных значениях этих компонент, перенумеруем состояния процессов $\mathbf{u}_t^{(1)}$ и $\mathbf{u}_t^{(2)}$ в обратном лексикографическом порядке. Для дальнейшего описания интенсивностей переходов цепи нам понадобятся матрицы $P_i(\cdot)$, $A_i(\cdot,\cdot)$, и $L_i(\cdot,\cdot)$, введенные в работах [4, 5]. Дадим краткое объяснение вероятностного смысла этих матриц.

Введем в рассмотрение матрицы $\tilde{S}_l = \begin{pmatrix} 0 & O \\ \boldsymbol{S}_0^{(l)} & S_l \end{pmatrix}$, $l = 1, 2$. Тогда: матрица $L_k(n, \tilde{S}_l)$ содержит интенсивности переходов процесса $\mathbf{u}_t^{(l)}$, приводящих к завершению обслуживания одного из $n-k$ запросов l-го типа; матрица $P_n(\boldsymbol{\beta}_l)$ содержит вероятности переходов процесса $\mathbf{u}_t^{(l)}$, приводящих к увеличению на приборе числа запросов l-го типа с n до $n+1$;

матрица $A_n(k,\tilde{S}_l)$ содержит интенсивности переходов процесса $\mathbf{u}_t^{(l)}$ в его пространстве состояний без увеличения или уменьшения числа запросов l-го типа.

Алгоритм вычисления матриц $P_i(\cdot)$, $A_i(\cdot,\cdot)$, и $L_i(\cdot,\cdot)$ следует из результатов В. Рамасвами и Д. Лукантони, опубликованных в статьях [4, 5]. Четко по шагам этот алгоритм описан в [6].

Введем обозначение $Q_{n,n'}$ для интенсивностей переходов цепи из состояний, соответствущих значению n первой компоненты, в состояния, соответствующие значению n' этой компоненты, $n,n'=\overline{0,N}$.

Лемма 1. *Инфинитезимальный генератор цепи Маркова $\xi_t, t\geqslant 0$, имеет блочную трех-диагональную структуру*

$$Q=\begin{pmatrix} Q_{0,0} & Q_{0,1} & O & \dots & O & O\\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \dots & O & O\\ O & Q_{2,1} & Q_{2,2} & \dots & O & O\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ O & O & O & \dots & Q_{N-1,N-1} & Q_{N-1,N}\\ O & O & O & \dots & Q_{N,N-1} & Q_{N,N}\end{pmatrix},$$

где

$$Q_{0,0}=diag^-\{\frac{1}{r}L_{R-r}(R,\tilde{S}_2),r=\overline{1,R}\}\otimes I_{\bar{W}}+$$

$$+diag\{0,\frac{1}{r}A_r(R,S_2),r=\overline{1,R}\}\oplus D_0+diag\{O_{\sum\limits_{r=0}^{R-1}C_{r+M_2-1}^{M_2-1}},I_{C_{R+M_2-1}^{M_2-1}}\}\otimes D_2+$$

$$+diag^+\{P_r(\boldsymbol{\beta}_2),r=\overline{0,R-1}\}\otimes D_2+\Delta_0;$$

$$Q_{n,n+1}=P_n(\boldsymbol{\beta}_1)\otimes I_{\sum\limits_{r=0}^{R}C_{r+M_2-1}^{M_2-1}}\otimes D_1,\ 0\leqslant n\leqslant N-1;$$

$$Q_{n,n-1}=\frac{1}{n}L_{N-n}(N,\tilde{S}_1)\otimes I_{\sum\limits_{r=0}^{R}C_{r+M_2-1}^{M_2-1}}\otimes I_{\bar{W}},\ 1\leqslant n\leqslant N;$$

$$Q_{n,n}=I_{C_{n+M_1-1}^{M_1-1}}\otimes diag^-\{\frac{1}{r}L_{R-r}(R,\tilde{S}_2),r=\overline{1,R}\}\otimes I_{\bar{W}}+$$

$$+\frac{1}{n}A_n(N,\tilde{S}_1)\oplus diag\{0,\frac{1}{r}A_r(R,S_2),r=\overline{1,R}\}\oplus(D_0+\delta_{n,N}D_1)+$$

$$+I_{C_{n+M_1-1}^{M_1-1}}\otimes diag\{O_{\sum\limits_{r=0}^{R-1}C_{r+M_2-1}^{M_2-1}},I_{C_{R+M_2-1}^{M_2-1}}\}\otimes D_2+$$

$$+I_{C_{n+M_1-1}^{M_1-1}}\otimes diag^+\{P_r(\boldsymbol{\beta}_2),r=\overline{0,R-1}\}\otimes D_2+\Delta_n,\ 1\leqslant n\leqslant N,$$

где $\otimes(\oplus)$ – символ кронекорова произведения (суммы) матриц, см. [3], $\delta_{n,N}$ – символ Кронекера, $\Delta_n, n=\overline{0,N}$,–диагональные матрицы, которые строятся так, чтобы выполнялось равенство $Q\mathbf{e}=\mathbf{0}^T$.

4. Стационарное распределение. Характеристики производительности

В соответствии с описанным упорядочением состояний цепи Маркова ξ_t, сформируем векторы-строки $\mathbf{p}_n, n = \overline{0, N}$, стационарных вероятностей состояний цепи. При большом размере системы уравнений для этих вероятностей они могут быть вычислены по алгоритму, разработанному в [7]. На основе стационарного распределения можно получить формулы для расчета стационарных характеристик производительности системы. Ниже приведены некоторые из них.

— Совместное распределение числа запросов 1 типа, находящихся на приборе, числа запросов 1 типа, находящихся на различных фазах обслуживания, и состояний $MMAP$-потока
$$\mathbf{p}_n^* = \mathbf{p}_n(I_{C^{M_1-1}_{n+M_1-1}} \otimes \mathbf{e}_{\sum\limits_{r=0}^{R} C^{M_2-1}_{r+M_2-1}} \otimes I_{\bar{W}}),\ n = \overline{0, N}.$$

— Распределение числа запросов 1 типа в системе $p_n = \mathbf{p}_n^* \mathbf{e}$, $n = \overline{0, N}$.

— Совместное распределение числа запросов 2 типа, находящихся на приборе, числа запросов 2 типа, находящихся на различных фазах обслуживания, и состояний $MMAP$-потока
$$\mathbf{q}_r^* = \sum_{n=0}^{N} \mathbf{p}_n(I^{(n,r)} \otimes I_{\bar{W}}), r = \overline{0, R},$$
где
$$I^{(n,r)} = \begin{pmatrix} O_{C^{M_1-1}_{n+M_1-1} \sum\limits_{m=0}^{r-1} C^{M_2-1}_{m+M_2-1} \times C^{M_2-1}_{r+M_2-1}} \\ \mathbf{e}_{C^{M_1-1}_{n+M_1-1}} \otimes I_{C^{M_2-1}_{r+M_2-1}} \\ O_{C^{M_1-1}_{n+M_1-1} \sum\limits_{m=r+1}^{R} C^{M_2-1}_{m+M_2-1} \times C^{M_2-1}_{r+M_2-1}} \end{pmatrix}.$$

— Распределение числа запросов 2 типа в системе $q_r = \mathbf{q}_r^* \mathbf{e}, r = \overline{0, R}$.

— Вероятность потери запроса k-го типа рассчитывается как
$$P_{loss,k} = \frac{\lambda_k - \varphi_k}{\lambda_k},\ k = 1, 2,$$
где λ_k – интенсивность поступления запросов k-го типа в $MMAP$, φ_k – интенсивность выходящего потока запросов k-го типа из системы. Значение λ_k вычисляется как $\lambda_k = \boldsymbol{\theta} D_k \mathbf{e}$, где вектор $\boldsymbol{\theta}$ единственное решение системы $\boldsymbol{\theta}(D_0 + D_1 + D_2) = \mathbf{0}$, $\boldsymbol{\theta}\mathbf{e} = 1$. Значения φ_k вычисляются как
$$\varphi_1 = \sum_{n=1}^{N} \mathbf{p}_n^*(I_{C^{M_1-1}_{n+M_1-1}} \otimes \mathbf{e}_{\bar{W}})\frac{1}{n} L_{N-n}(N, \tilde{S}_1)\mathbf{e},$$
$$\varphi_2 = \sum_{r=1}^{R} \mathbf{q}_r^*(I_{C^{M_2-1}_{r+M_2-1}} \otimes \mathbf{e}_{\bar{W}})\frac{1}{r} L_{R-r}(R, \tilde{S}_2)\mathbf{e}.$$

5. Заключение

В данной работе исследовано стационарное поведение системы массового обслуживания с $MMAP$ потоком запросов двух типов, разделением процессора и ограниченным числом мест для запросов разных типов. Построена многомерная цепь Маркова, описывающая процесс функционирования системы, получены основные характеристики производительности системы. На основании полученных результатов могут быть найдены преобразования Лапласа-Стилтьеса распределения времен пребывания в системе запросов разных типов.

СПИСОК ЛИТЕРАТУРЫ

1. *He Q.M.* Queues with marked calls // Advances in Applied Probability 1996. V. 28. P. 567–587.
2. *Neuts M. F.* Matrix-Geometric Solutions in Stochastic Models. The Johns Hopkins University Press, Baltimore, 1981.
3. *Graham A.* Kronecker Products and Matrix Calculus with Applications. Ellis Horwood, Cichester, 1981.
4. *Ramaswami V.* Independent Markov processes in parallel // Comm. Statist.-Stochastic Models. 1985. V.1. P. 419–432.
5. *Ramaswami V., Lucantoni D.M.* Algorithms for the multi-server queue with phase-type service // Comm. Statist.-Stoch. Models. 1985. V.1. P. 393–417.
6. *Dudina O., Kim C.S., Dudin S.* Retrial queueing system with Markovian arrival flow and phase type service time distribution // Computers and Industrial Engineering. 2013. V. 66. P. 360–373.
7. *Klimenok V.I., Kim C.S., Orlovsky D.S., Dudin A.N.* Lack of invariant property of Erlang $BMAP/PH/N/0$ model // Queueing Systems. 2005. V.49. P. 187–213.

Клименок Валентина Ивановна — д-р физ.-мат. наук, профессор. E-mail: *vklimenok@yandex.ru*

Дудин Александр Николаевич — д-р физ.-мат. наук, профессор. E-mail: *dudin-alexander@mail.ru*

ТАНДЕМНАЯ СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ $BMAP/G/1 \to \cdot/PH/N/R$

В. И. Клименок, А. Н. Дудин, О. С. Дудина

Белорусский государственный университет, г. Минск, Беларусь

В данной работе исследуется тандемная система массового обслуживания, состоящая из двух последовательных станций. Первая станция представлена однолинейной системой с бесконечным буфером, вторая – многолинейной системой с конечным буфером. Запросы поступают на первую станцию в групповом марковском потоке (общепризнанная аббревиатура $BMAP$ – Batch Markovian Arrival Process). Времена обслуживания на первой станции распределены по произвольному закону, на второй станции – по фазовому закону (общепризнанная аббревиатура PH – Phase Type distribution). В ситуации, когда промежуточный буфер между станциями заполнен в момент окончания обслуживания запроса на первой станции, этот запрос теряется или блокирует прибор первой станции до тех пор, пока не закончится обслуживание на одном из приборов второй станции. Получено условие существования стационарного режима в системе, стационарные распределения состояний во вложенные и произвольные моменты времени, ряд характеристик производительности системы.

Ключевые слова: *Тандемная система массового обслуживания, групповой марковский поток, промежуточный буфер, стационарное распределение.*

1. Введение

Многофазные (тандемные) системы массового обслуживания (СМО) являются простейшими видами сетей массового обслуживания и находят большой интерес среди исследователей в области теории массового обслуживания и телекоммуникаций. Ранние работы в этой области посвящены, в основном, системам со стационарным пуассоновским потоком и экспоненциально распределенным временем обслуживания, см., например, [1]. В последние десятилетия появились работы по тандемным СМО, которые учитывают коррелированный характер входных потоков и неэкспоненциальное распределение времен обслуживания на станциях. В частности, известны работы, в которых входной поток является MAP, допускается произвольное распределение времени обслуживания на второй станции, а буфер первой станции конечен.

Такого рода системы рассмотрены в серии работ А. Гомез-Коррала (Испания), см., например, [2]. Относительная простота исследования таких систем объясняется тем, что первая станция описывается СМО типа $MAP/PH/1/N$. Для такой СМО входящий поток на вторую станцию является поддающимся формальному описанию MAP потоком. Если же произвольное распределение времени обслуживания допускается именно на первой станции, анализ системы усложняется, поэтому значительная часть существующих работ предполагает декомпозицию тандема и приближенный анализ выходящего потока из первой станции. Точный аналитический анализ дуальных тандемов, у которых первая станция представлена системой типа $BMAP/G/1$ или $BMAP/G/1/N$, проводится на основе использования метода вложенных цепей Маркова с последующим применением аппарата процессов марковского восстановления. Такой анализ используется в работах ученых Белорусской школы массового обслуживания, для ссылок см., например, монографию [3].

В данной работе мы обобщаем результаты [4], где исследована тандемная СМО $BMAP/G/1 \to \bullet/M/N/R$ на случай фазового распределени времени обслуживания на второй станции.

2. Математическая модель

Рассматривается тандемная система $BMAP/G/1 \to \bullet/PH/N/R$. Первая станция образована однолинейной СМО с ожиданием. В систему поступает $BMAP$-поток запросов, заданный управляющим процессом ν_t, $t \geqslant 0$, с пространством состояний $\{0, ..., W\}$ и матричной производящей функцией $D(z) = \sum\limits_{k=0}^{\infty} D_k z^k, |z| \leqslant 1$. Средняя интенсивность поступления запросов в $BMAP$ вычисляется по формуле $\lambda = \boldsymbol{\theta} D'(1)\mathbf{e}$, где $\boldsymbol{\theta}$ есть единственное решение системы линейных алгебраических уравнений $\boldsymbol{\theta} D(1) = \mathbf{0}, \boldsymbol{\theta}\mathbf{e} = 1$. Здесь $\mathbf{0}$ есть вектор-строка, составленный из нулей и $\mathbf{e}$ есть вектор-столбец, составленный из единиц. Более подробное описание $BMAP$ можно найти, например, в [3], [5].

Если группа запросов размера $k \geqslant 1$ поступает в систему и застает обслуживающий прибор свободным, то один из запросов начинает обслуживаться, а остальные располагаются в очереди неограниченной длины. Если прибор занят в момент поступления группы, то все запросы этой группы становятся в конец очереди. Времена обслуживания запросов на приборе первой станции являются независимыми случайными величинами с функцией распределения $B(t)$ и конечным первым моментом $b_1 = \int\limits_0^{\infty} t dB(t)$. После обслуживания на первой станции запрос поступает на вторую станцию, образованную N-линейной системой с конечным буфером размера R.

Время обслуживания прибором второй станции имеет PH распределение M-го порядка, заданное неприводимым представлением $(\boldsymbol{\beta}, S)$. Здесь вектор-строка $\boldsymbol{\beta}$ задает первоначальное распределение фаз обслуживания, матрица S- интенсивности переходов управляющего обслуживанием марковского процесса между его состояниями (фазами), вектор $\boldsymbol{S}_0 = -S\mathbf{e}$ – интенсивности переходов, приводящие к окончанию обслуживания. Подробную информацию о PH распределении можно найти, например, в [6]. При отсутствии свободного места в буфере с вероятностью p, $0 \leqslant p \leqslant 1$, запрос уходит из системы недообслуженным (теряется), а с вероятностью $1-p$ прибор первой станции блокируется и не обслуживает следующий запрос до ближайшего окончания обслуживания на второй станции.

3. Стационарное распределение вложенной цепи Маркова

Пусть t_n – n-й момент окончания обслуживания на первой станции, $n \geqslant 1$. Рассмотрим процесс $\xi_n = \{i_n, l_n, m_n^{(1)}, \ldots, m_n^{(\min\{l_n,N\})}, \nu_n\}$, $n \geqslant 1$, где i_n – число запросов на первой станции (не включая запрос, вызвавший блокировку) в момент времени $t_n + 0$; l_n – число запросов на второй станции в момент времени $t_n - 0$; $m_n^{(h)}$ – фаза обслуживания на h-м занятом приборе второй станции в момент времени t_n; ν_n – состояние $BMAP$ в момент времени t_n, $i_n \geqslant 0, l_n = \overline{0, N+R}$, $m_n^{(h)} = \overline{0, M}$, $h = \overline{1, \min\{l_n, N\}}$; $\nu_n = \overline{0, W}$.

Лемма 1. *Матрица вероятностей переходов цепи Маркова* $\xi_n, n \geqslant 1$, *имеет блочную структуру вида*

$$P = \begin{pmatrix} V_0 & V_1 & V_2 & V_3 & \ldots \\ Y_0 & Y_1 & Y_2 & Y_3 & \ldots \\ O & Y_0 & Y_1 & Y_2 & \ldots \\ O & O & Y_0 & Y_1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

где

$$V_i = \sum_{k=1}^{i+1} [-\hat{Q}(\Delta \oplus D_0)^{-1}\tilde{D}_k + (1-p)Q_2 \otimes \int_0^\infty e^{S^{\oplus N} t}\boldsymbol{S}_0^{\oplus N} \otimes P(k,t)dt]\Omega_{i-k+1},$$

$$Y_i = \bar{Q}\Omega_i + (1-p)[Q_2 \otimes \sum_{k=0}^{i} \int_0^\infty e^{S^{\oplus N} t}\boldsymbol{S}_0^{\oplus N} \otimes P(k,t)dt]\Omega_{i-k},\ i \geqslant 0.$$

Здесь $\bar{Q} = \tilde{Q}_1 + p\tilde{Q}_2$; $\hat{Q} = \bar{Q} - (1-p)Q_2 \otimes [(S^{\oplus N} \oplus D_0)]^{-1}[\boldsymbol{S}_0^{\oplus N} \otimes I_{\bar{W}}]$; $\tilde{Q}_1$ *– блочная квадратная матрица порядка* $K(W+1)$, *где* $K =$

$\frac{1-M^{N+1}}{1-M} + RM^N$, *у которой наддиагональные блоки определяются как* $I_{\min\{r,N\}(W+1)}, r = \overline{0, N+R}$, *а остальные элементы равны нулю;* Q_2*– квадратная матрица порядка* $N+R+1$ *вида* $diag\{0,\ldots,0,1\}$ $\tilde{Q}_2$ *– квадратная матрица порядка* $K(W+1)$, *у которой все элементы нулевые, кроме последнего диагонального блока, равного* $I_{M^N(W+1)}$*;* Δ*– квадратная блочно-двухдиагональная матрица порядка* K, *диагональные блоки которой определяются как* $S^{\oplus\min\{r,N\}}$, *а поддиагональные блоки как* $\boldsymbol{S}_0^{\oplus\min\{r,N\}}, r = \overline{0, N+R}$; $\tilde{D}_k = I_K \otimes D_k$, $\Omega_k = \int\limits_0^\infty e^{\Delta t} \otimes P(k,t)dB(t)$, $k \geqslant 0$; *матрицы* $P(k,t)$ *определяются как коэффициенты разложения* $e^{D(z)t} = \sum\limits_{k=0}^{\infty} P(k,t)z^k, |z| \leqslant 1$.

Следствие 1 *Цепь Маркова* ξ_n, $n \geqslant 1$, *принадлежит классу квазитеплицевых цепей Маркова, см. [3, 7].*

Теорема 1. *Необходимым и достаточным условием существования стационарного распределения ЦМ* ξ_n, $n \geqslant 1$, *является выполнение неравенства* $\rho < 1$, *где* $\rho = \lambda[b_1 + (1-p)\hat{\boldsymbol{x}}(-S^{\oplus N}\mathbf{e}]$, $\hat{\boldsymbol{x}}$ *– вектор-строка, компонентами которого являются последние* M^N *компонент вектора* $\boldsymbol{x}$, *который является единственным решением СЛАУ*

$$\boldsymbol{x}(\mathcal{Q}_1 + \mathcal{Q}_2)\int\limits_0^\infty e^{\Delta t}dB(t) = \boldsymbol{x}, \quad \boldsymbol{x}\mathbf{e} = 1,$$

где матрицы $\mathcal{Q}_1, \mathcal{Q}_2$ *имеют тот же вид, что и матрицы* $\tilde{Q}_1, \tilde{Q}_2$, *только порядок каждого из блоков уменьшен в* $W+1$ *раз.*

Обозначим стационарные вероятности состояний цепи Маркова $\xi_t, t \geq 0$, как $\boldsymbol{\pi}(i, l, m^{(1)}, \ldots, m^{(\min\{l,N\})}, \nu)$, перенумеруем эти вероятности в лексикографическом порядке и введем векторы $\boldsymbol{\pi}_i, i \geqslant 0$, этих вероятностей. Эти векторы находятся с использованием алгоритма, описанного в [3, 7].

4. Стационарное распределение в произвольный момент времени

Рассмотрим процесс $\zeta_t = \{i_t, j_t, l_t, m_t^{(1)}, \ldots, m_t^{(\min\{l_t,N\})}, \nu_t\}$, $t \geq 0$, где i_t – число запросов на первой станции (не включая запрос, вызвавший блокировку); j_t– величина, принимающая значения 0, 1, 2 в зависимости от того, свободен, обслуживает запрос или блокирован прибор первой станции; l_t – число запросов на второй станции; $m_t^{(h)}$ – фаза обслуживания на h-м занятом приборе второй станции; ν_t – состояние $BMAP$ в момент времени t.

Обозначим стационарные вероятности состояний процесса $\zeta_t, t \geq 0$, как $p(i, j, l, m^{(1)}, \ldots, m^{(\min\{l,N\}}, \nu)$, перенумеруем эти вероятности в лексикографическом порядке и введем векторы $\boldsymbol{p}_i(j),\ i \geqslant 0,\ j = 0,1,2,$ этих вероятностей.

Теорема 2. *Векторы стационарных вероятностей* $\boldsymbol{p}_i(j),\ i \geqslant 0,\ j = 0,1,2,$ *выражаются через векторы стационарного распределения* $\boldsymbol{\pi}_i,\ i \geqslant 0,$ *вложенной цепи Маркова* $\xi_n,\ n \geqslant 1,$ *следующим образом:*

$$\boldsymbol{p}_0(0) = -\lambda\boldsymbol{\pi}_0\hat{Q}(\Delta \oplus D_0)^{-1},$$

$$\boldsymbol{p}_i(1)=\lambda\{\boldsymbol{\pi}_0\sum_{k=1}^{i}[-\hat{Q}(\Delta\oplus D_0)^{-1}\tilde{D}_k+(1-p)Q_2\otimes\int\limits_0^\infty e^{S^{\oplus N}t}\boldsymbol{S}_0^{\oplus N}\otimes P(k,t)dt]\tilde{\Omega}_{i-k}+$$

$$+\sum_{l=1}^{i}\boldsymbol{\pi}_l[\bar{Q}\tilde{\Omega}_{i-l}+(1-p)(Q_2\otimes\sum_{k=0}^{i-l}\int\limits_0^\infty e^{S^{\oplus N}t}\boldsymbol{S}_0^{\oplus N}\otimes P(k,t)dt)\tilde{\Omega}_{i-k-l}]\},$$

$$\boldsymbol{p}_i(2) = \lambda(1-p)\sum_{l=0}^{i-1}\boldsymbol{\pi}_l[Q_2\otimes\int\limits_0^\infty e^{S^{\oplus N}t}\otimes P(i-l-1,t)dt],$$

где $\tilde{\Omega}_k = \int\limits_0^\infty e^{\Delta t}\otimes P(k,t)(1-B(t))dt,\ k \geqslant 0.$

5. Характеристики производительности системы

Введем обозначения $\boldsymbol{p}_i = \boldsymbol{p}_i(0) + \boldsymbol{p}_i(1) + \boldsymbol{p}_i(2), i \geqslant 0,\ \boldsymbol{P}(z) = \sum\limits_{i=0}^{\infty}\boldsymbol{p}_i z^i, |z| \leqslant 1$. Приведем формулы для некоторых характеристик системы. Среднее число запросов на первой станции $L_1 = \boldsymbol{P}'(1)\mathbf{e}$. Среднее число запросов на второй станции $L_2 = \boldsymbol{P}(1)(I_K\otimes\mathbf{e}_{\bar{W}})R_1\mathbf{e}$, где $R_1 = \mathrm{diag}\,\{rI_{M^{\min\{r,N\}}},\ r = \overline{0,N+R}\}$. Среднее число занятых приборов на второй станции $N_{busy} = \boldsymbol{P}(1)(I_K\otimes\mathbf{e}_{\bar{W}})R_2\mathbf{e}$, где $R_2 = \mathrm{diag}\{rI_{M^{\min\{r,N\}}}, r = 0,1,\dots,N,N,\dots,N\}$. Среднее число занятых мест в конечном буфере $N_{buffer} = \boldsymbol{P}(1)(I_K\otimes\mathbf{e}_{\bar{W}})R_3\mathbf{e}$, где

$$R_3 = \mathrm{diag}\{O_{\frac{1-M^{N+1}}{1-M}}, I_{M^N}, 2I_{M^N},\dots,RI_{M^N}\}.$$

Вероятности того, что произвольный запрос будет потерян или вызовет блокировку прибора после обслуживания на первой станции $P_{loss} = p\boldsymbol{\Pi}(1)\tilde{Q}_2\mathbf{e}$, $P_{block} = (1-p)\boldsymbol{\Pi}(1)\tilde{Q}_2\mathbf{e}$. Вероятности P_{idle}, P_{serve}, $P_{block}^{(server)}$ того, что прибор первой станции свободен, занят обслуживанием или заблокирован

$$P_{idle} = \lambda\boldsymbol{\pi}_0\hat{Q}(-\tilde{D}_0)^{-1}\mathbf{e},\ \ P_{serve} = \lambda b_1,\ \ P_{block}^{(server)} = 1 - P_{idle} - P_{serve}.$$

6. Заключение

В данной работе исследована тандемная система массового обслуживания $BMAP/G/1 \to \bullet/PH/N/R$. В отличие от предыдущих работ по исследованию такого рода систем, мы предполагаем, что распределение времени обслуживания на второй станции не является экспоненциальным. Оно время распределено по довольно общему фазовому закону. Процесс функционирования системы описан в терминах многомерной цепи Маркова. Получены условие ее эргодичности, ствционарное распределение состояний системы в моменты окончания обслуживания на первой станции и в произвольные моменты времени. Полученные распределения использованы при выводе формул для ряда характеристик производительности системы. Результаты исследования могут быть использованы для математического моделирования телекоммуникационных сетей и их фрагментов.

СПИСОК ЛИТЕРАТУРЫ

1. *Gnedenko B. W., Konig D.* Handbuch der Bedienungstheorie. Akademie Verlag, Berlin, 1983.
2. *A. Gomez-Corral A.* A tandem queue with blocking and Markovian arrival process // Queueing Systems. 2002. V. 41. P. 343–370.
3. *Dudin A.N., Klimenok V.I., Vishnevsky V.M.* The theory of queuing systems with correlated flows. Springer Nature, 2019.
4. *Taramin O., Klimenok V., Lee M.H.* Performance analysis of two-node network with multi-server second station and intermediate buffer // Proceedings of the 11th International Workshop on MSPT, Jeonju, Korea, 2008. P. 63–68.
5. *Lucantoni D.* New results on the single server queue with a batch Markovian arrival process // Comm. in Stat.-Stoch. Models. 1991. V. 7. P. 1–46.
6. *Neuts M. F.* Matrix-Geometric Solutions in Stochastic Models. The Johns Hopkins University Press, Baltimore, 1981.
7. *Klimenok V.I., Dudin A.N.* Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory // Queueing Systems. 2006. V. 54. P. 245–259.

Клименок Валентина Ивановна — д-р физ.-мат. наук, профессор. E-mail: *vklimenok@yandex.ru*

Дудин Александр Николаевич — д-р физ.-мат. наук, профессор. E-mail: *dudin-alexander@mail.ru*

Дудина Ольга Сергеевна — канд. физ.-мат. наук, ведущий научный сотрудник. E-mail: *dudina@bsu.by*

ОПТИМИЗАЦИЯ СИСТЕМЫ АДАПТИВНОГО УПРАВЛЕНИЯ КОНФЛИКТНЫМИ ПОТОКАМИ КОКСА–ЛЬЮИСА

Е. В. Кудрявцев, М. А. Федоткин

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

В работе описана система массового обслуживания неординарных пуассоновских потоков с помощью кибернетического подхода Ляпунова–Яблонского. Моделью такой системы является многомерная цепь Маркова. Аналитические свойства данной марковской цепи были изучены в предыдущих работах авторов. В этой работе была построена имитационная модель системы. Исследован переходный процесс и предложен алгоритм поиска момента окончания переходного процесса. С помощью численной оптимизации были найдены квазиоптимальные параметры системы по условию минимума среднего времени ожидания обслуживания.

Ключевые слова: *Конфликтные потоки, кибернетическая система, имитационное моделирование, квазиоптимальные параметры.*

Введение

Данная работа связана с важной проблемой создания алгоритмов в интеллектуальных транспортных системах, которые управляют конфликтными потоками [1] на пересечениях магистралей в крупных городах. Предлагается адаптивный алгоритм управления такого рода потоками. Алгоритм управления учитывает не только длины очередей, но и очередность поступления требований в систему. Построена и изучена математическая модель такой системы управления потоками из неоднородных требований.

В качестве входных потоков рассмотрены два независимых конфликтных неординарных пуассоновских потока Π_1 и Π_2. При $t_0 \geq 0$, $t > 0$ и $j = 1, 2$ получена вероятность $P_j(t,k)$ поступления k требований за промежуток времени $[t_0, t_0 + t)$ по потоку Π_j следующего вида

$$P_j(t,k) = e^{-\lambda_j t} \sum_{n=0}^{[k/2]} \alpha_j^n \frac{(\lambda_j t p_j)^{k-n}}{n!(k-2n)!} + \tag{1}$$

$$+ e^{-\lambda_j t}\sum_{n=0}^{[k/2]}\alpha_j^n\sum_{m=1}^{\min\{k-2n,n\}}\beta_j^m\sum_{l=0}^{k-2n-m}\gamma_j^l\frac{(\lambda_j t p_j)^{k-n-m-l}C_{m+l-1}^l}{(n-m)!m!(k-2n-m-l)!},$$

где α_j, β_j, γ_j, λ_j и $p_j = (1+\alpha_j+\alpha_j\beta_j/(1-\gamma_j))^{-1}$ — параметры распределения. Свойства таких потоков с неоднородными требованиями изучены в [2]. Адекватность представления реальных потоков с помощью распределения вида (1) показана на примерах таблиц из [3].

1. Математическая модель управляющей системы

Для построения и описания математической модели дискретной системы адаптивного управления конфликтными потоками и обслуживания неоднородных требований используется кибернетический подход Ляпунова-Яблонского [4]. Применение данного подхода к изучаемой системе подробно описано в [1].

Согласно кибернетическому подходу будем рассматривать систему в случайные дискретные моменты времени τ_i или на промежутках $[\tau_i,\tau_{i+1})$ при $i=0,1,\ldots$ Здесь величина τ_0 — начальный момент времени, а τ_i, $i>0$ — моменты смены состояний обслуживающего устройства. Пусть $y_0=(0,0)$, $y_1=(1,0)$, $y_2=(0,1)$ и X — целочисленная одномерная неотрицательная решетка. Теперь определим следующие случайные величины и элементы: 1) $\Gamma_i\in\Gamma=\{\Gamma^{(1)},\Gamma^{(2)},\ldots,\Gamma^{(8)}\}$ — состояние обслуживающего устройства на промежутке $[\tau_i,\tau_{i+1})$; 2) $\eta_{j,i}\in X$ — число заявок потока Π_j, поступивших в систему за промежуток $[\tau_i,\tau_{i+1})$, и $\eta_i=(\eta_{1,i},\eta_{2,i})$; 3) η_i' — случайный вектор, принимающий значение y_0, если в систему на i-ом такте $[\tau_i,\tau_{i+1})$ не поступило ни одной заявки, и значение y_j, если на i-ом такте первой пришла заявка (или заявки) потока Π_j; 4) $\varkappa_{j,i}\in X$ — число заявок потока Π_j, которые находятся в системе в момент τ_i, и $\varkappa_i=(\varkappa_{1,i},\varkappa_{2,i})$; 5) $\xi_{j,i}$ — максимально возможное число заявок потока Π_j, которые система может обслужить на интервале $[\tau_i,\tau_{i+1})$, и $\xi_i=(\xi_{1,j},\xi_{2,j})$.

Определим теперь последовательность $\{\tau_i; i>0\}$ моментов смены состояний обслуживающего устройства. С этой целью приведем содержательный смысл каждого состояния из множества Γ. Состояние $\Gamma^{(3j-2)}$ соответствует первому этапу периода обслуживания потока Π_j. Длительность обслуживания одной заявки, поступившей из накопителя, равна постоянной величине $\mu_{j,1}^{-1}$. Длительность пребывания в $\Gamma^{(3j-2)}$ равна T_{3j-2}. Состояние $\Gamma^{(3j-1)}$ соответствует второму этапу периода обслуживания потока Π_j. Длительность обслуживания одной заявки равна величине $\mu_{j,2}^{-1}<\mu_{j,1}^{-1}$. Длительность пребывания в этом состоянии является случайной величиной, принимающей значения из множества

$\{kT_{3j-1}; k = \overline{1, n_j}\}$, где n_j — максимальное число продлений и T_{3j-1} — длительность одного продления. Продление происходит в 2 случаях: 1) длина очереди по потоку Π_j не меньше заданного целочисленного параметра $K_j > 0$, 2) на предыдущем этапе продлений поступили требования, которые необходимо обслужить. Состояние $\Gamma^{(3j)}$ соответствует режиму переналадки для потока Π_j, во время которого возможно только дообслуживание требований потока Π_j. Длительность пребывания в этом состоянии равна T_{3j}. Состояние $\Gamma^{(6+j)}$ соответствует первому этапу периода обслуживания потока Π_j, в случае, когда возможен мгновенный переход в состояние $\Gamma^{(3j)}$. Длительность пребывания в $\Gamma^{(6+j)}$ является случайной величиной. Максимальное время пребывания в этом состоянии равно T_{3j-2}. В данном состоянии очередь по обслуживаемому потоку пуста, и обслуживающее устройство следит за очередностью прихода заявок. Если за время T_{3j-2} первой пришла заявка потока Π_j, то через T_{3j-2} от момента τ_i произойдет переход в состояние $\Gamma^{(3j-1)}$. Если первой пришла заявка другого потока, то произойдет мгновенный переход в состояние $\Gamma^{(3j)}$. И, наконец, если за это время не поступит ни одной заявки по обоим потокам, то обслуживающее устройство также переключится в состояние в $\Gamma^{(3j)}$. Постоянные величины T_k, $k = \overline{1, 6}$, целесообразно выбрать в следующем виде

$$T_{3j-2} = \mu_{j,1}^{-1} + l_{3j-2}\theta_j\mu_{j,1}^{-1}, \quad T_{3j-1} = l_{3j-1}\theta_j\mu_{j,2}^{-1}, \quad T_{3j} = l_{3j}\theta_j\mu_{j,2}^{-1}, \tag{2}$$

где $l_{3j-2} \in X$, l_{3j-1}, $l_{3j} \in \{1, 2, \ldots\}$ и являются параметрами. Величина $0 < \theta_j \leq 1$ обозначает часть обслуживания, которую необходимо пройти требованию, чтобы можно было начать обслуживать следующую заявку. В случае $\theta_j < 1$ одновременно может обслуживаться несколько требований. Соотношение (2) означает, что изменение состояния обслуживающего устройства происходит в момент окончания обслуживания одной из заявок. Получаем, что максимально возможное число обслуженных требований равно $1 + l_{3j-2}$ для состояния $\Gamma^{(3j-2)}$, равно kl_{3j-1} для состояния $\Gamma^{(3j-1)}$ и равно целой части $1/\theta_j$ для состояния $\Gamma^{(3j)}$. Кибернетический подход позволяет получить [1] следующие функциональные рекуррентные по i соотношения

$$\Gamma_{i+1} = u(\Gamma_i, \varkappa_i, \eta_i') =$$
$$= \begin{cases} \Gamma^{(3j-2)}, & \left\{\left[\Gamma_i = \Gamma^{(3s)}\right] \& \left[(\varkappa_{j,i} > 0) \vee (\varkappa_{s,i} \geq K_s) \vee (\eta_i' = y_j)\right]\right\} \vee \\ & \vee \left\{\left[\Gamma_i = \Gamma^{(3j)}\right] \& [\varkappa_{s,i} = 0] \& [\varkappa_{j,i} \leq K_j] \& [\eta_i' = y_j]\right\}, \\ \Gamma^{(3j-1)}, & \left\{\Gamma_i = \Gamma^{(3j-2)}\right\} \vee \left\{\left[\Gamma_i = \Gamma^{(6+j)}\right] \& [\eta_i' = y_j]\right\}, \\ \Gamma^{(3j)}, & \left\{\Gamma_i = \Gamma^{(3j-1)}\right\} \vee \left\{\left[\Gamma_i = \Gamma^{(6+j)}\right] \& [\eta_i' \neq y_j]\right\}, \\ \Gamma^{(6+j)}, & \left[\Gamma_i = \Gamma^{(3s)}\right] \& [\varkappa_{j,i} = 0] \& [\varkappa_{s,i} < K_s] \& [\eta_i' = y_0]; \end{cases} \tag{3}$$

$$\varkappa_{j,i+1} = v_j(\Gamma_i, \varkappa_i, \eta_i, \xi_i) =$$
$$= \begin{cases} \max\{0, \varkappa_{j,i} + \eta_{j,i} - \xi_{j,i}\}, & \text{если } \Gamma_i \in \Gamma \backslash \{\Gamma^{(3)}, \Gamma^{(6)}\}, \\ \eta_{j,i} + \max\{0, \varkappa_{j,i} - \xi_{j,i}\}, & \text{если } \Gamma_i \in \{\Gamma^{(3)}, \Gamma^{(6)}\}; \end{cases} \tag{4}$$

здесь и далее в работе $j, s = 1, 2$, $j \neq s$. Используя соотношение (1), адаптивный алгоритм смены состояний обслуживающего устройства можно отобразить в виде графа, который представлен на рис. 1. Заметим, что

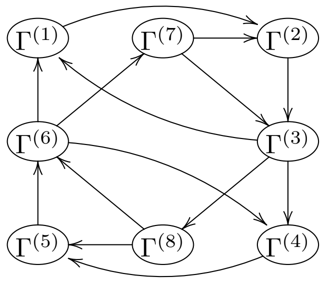

Рис. 1. Граф смены состояний обслуживающего устройства

состояние обслуживающего устройства на следующем шаге зависит от состояния на предыдущем шаге, длин очередей и очередности прихода заявок.

Соотношения (1) и (4) позволяют рассмотреть и изучить предельные свойства векторной последовательности $\{(\Gamma_i, \varkappa_i); i = 0, 1, \ldots\}$, которая является вероятностной моделью рассматриваемой системы адаптивного управления конфликтными потоками и обслуживания неоднородных требований [1].

2. Численное исследование системы

К сожалению, аналитически не удается найти такие важные характеристики, как среднее время ожидания обслуживания произвольного требования и средние длины очередей по потокам.

Для решения поставленных вопросов реализована имитационная модель системы адаптивного управления конфликтными потоками неоднородных требований в виде программы, написанной на языке C++. Каждая реализация функционирования системы задается с помощью следующих входных данных: 1) параметров α_j, β_j, γ_j, λ_j входных потоков; 2) параметров T_1, T_2, $\ldots$, T_6, $\mu_{j,1}$, $\mu_{j,2}$, θ_j, K_j, n_j системы; 3) начальных значений $\Gamma^{(r)}$, x_1, x_2 случайных элементов Γ_0, $\varkappa_{1,0}$, $\varkappa_{2,0}$.

С помощью имитационного моделирования можно построить модель более приближенную к реальной системе, так как можно наблюдать за процессами обслуживания в каждый момент времени. Обозначим через

$\gamma_j^0(l)$ время ожидания обслуживания заявки с номером $l = 1, 2, \ldots$ потока Π_j. Величина $\overline{\gamma}_j^0(n) = \frac{1}{n}\sum_{l=1}^{n}\gamma_j^0(l)$ определяет выборочное среднее время ожидания обслуживания в системе первых n требований потока Π_j. Оценку $\overline{\gamma}^*$ среднего времени ожидания обслуживания произвольного требования будем вычислять по формуле среднего взвешенного

$$\overline{\gamma}^* = \frac{\lambda_1 M_1 \overline{\gamma}_1^0 + \lambda_2 M_2 \overline{\gamma}_2^0}{\lambda_1 M_1 + \lambda_2 M_2},$$

где $M_j = (1 + 2\alpha_j + \alpha_j\beta_j(2/(1-\gamma_j) + 1/(1-\gamma_j)^2))p_j$ — среднее число требований в группе потока Π_j. Выборочные средние времена ожидания обслуживания будем вычислять для системы, в которой достигнут стационарный режим. Момент окончания переходного процесса определяется с помощью алгоритма, описанного в [5].

Приведем пример поиска квазиоптимальных значений параметров при следующем начальном наборе входных параметров: $\alpha_1 = 0.8$, $\beta_1 = 0.7$, $\gamma_1 = 0.5$, $\lambda_1 = 0.6$, $\alpha_2 = 0.6$, $\beta_2 = 0.5$, $\gamma_2 = 0.2$, $\lambda_2 = 0.3$, $T_1 = 1$, $T_2 = 2$, $T_3 = 1$, $T_4 = 1$, $T_5 = 3$, $T_6 = 1$, $\mu_{1,1} = 0.5$, $\mu_{2,1} = 1$, $\mu_{1,2} = 0.3$, $\mu_{2,2} = 0.6$, $\theta_1 = 1$, $\theta_2 = 0.5$, $K_1 = 10$, $K_2 = 10$, $n_1 = 10$, $n_2 = 10$ и $\Gamma^{(r)} = \Gamma^{(1)}$, $x_1 = 0$, $x_2 = 0$ для начального состояния системы. В силу физических ограничений для оптимизации доступны только следующие параметры T_2, T_5, n_1, n_2, K_1, K_2. Оптимизация производится поэтапно по парам параметров (T_2, T_5), (n_1, n_2) и (K_1, K_2) путем сокращенного перебора. Используя алгоритм сокращенного перебора, были получены квазиоптимальные параметры $T_2^* = 5$, $T_5^* = 1$, $n_1^* = 3$, $n_2^* = 10$, $K_1^* = 1$ и $K_2^* = 4$, которым соответствуют оценки $\overline{\gamma}_1^* = 2.47241$, $\overline{\gamma}_2^* = 5.13818$, $\overline{\gamma}^* = 3.17367$ средних времен ожидания обслуживания.

В табл. 1 приведена часть шагов алгоритма оптимизации.

Таблица 1

Поиск квазиоптимальных параметров системы управления

T_2	T_5	n_1	n_2	K_1	K_2	$\overline{\gamma}_1^*$	$\overline{\gamma}_2^*$	$\overline{\gamma}^*$
2	3	10	10	10	10	7.91365	12.9163	9.22965
9	**6**	10	10	10	10	5.10318	11.3382	6.74334
…	…	…	…	…	…	…	…	…
5	1	**3**	**10**	1	4	2.47241	5.13818	**3.17367**
5	1	3	10	**1**	**8**	2.48873	5.21597	3.20615

Заключение

В работе приведены результаты аналитического и численного исследования системы управления потоками Кокса-Льюиса из неоднородных требований. Математическая модель построена с использованием кибернетического подхода и аппарата теории массового обслуживания. Численная оптимизация проводилась на основе программно-реализованной имитационной модели.

СПИСОК ЛИТЕРАТУРЫ

1. *Федоткин М. А., Кудрявцев Е. В.* Анализ дискретной модели системы адаптивного управления конфликтными неоднородными потоками // Вестник московского университета. Серия 15: Вычислительная математика и кибернетика. 2019. №1. С. 19–26.
2. *Fedotkin M. A., Fedotkin A. M., Kudryavtsev E. V.* Construction and Analysis of a Mathematical Model of Spatial and Temporal Characteristics of Traffic Flows // Autom. Control and Comp. Sciences. 2014. **48**. №6. P. 358–367.
3. *Кокс Д., Льюис П.* Статистический анализ последовательностей событий. М.: Мир, 1969, 312 с.
4. *Ляпунов А. А., Яблонский С. В.* Теоретические проблемы кибернетики // Проблемы кибернетики. М.: Физматгиз. 1963. С. 5–22.
5. *Федоткин М. А., Кудрявцев Е. В.* Исследование переходного процесса адаптивного управления потоками неоднородных требований путем имитационного моделирования // Сборник материалов XVIII Международной конференции «Информационные технологии и математическое моделирование (ИТММ-2019)». Томск: НТЛ. 2019. Часть 2. С. 207–212.

Кудрявцев Евгений Владимирович — преподаватель кафедры теории вероятностей и анализа данных, институт ИТММ, ННГУ. E-mail: *evgkudryavcev@gmail.com*

Федоткин Михаил Андреевич — д. ф.-м. н., профессор, профессор кафедры теории вероятностей и анализа данных, институт ИТММ, ННГУ. E-mail: *fma5@rambler.ru*

ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ ХАРАКТЕРИСТИК ВЫХОДЯЩЕГО ПОТОКА ДВУХФАЗНОЙ МОДЕЛИ ОТ ПАРАМЕТРОВ ПЕРВОЙ ФАЗЫ

И. Л. Лапатин, А. А. Воркожоков

Национальный исследовательский
Томский государственный университет, г. Томск, Россия

В работе рассматривается двухфазная система массового обслуживания с буфером конечной ёмкости на первой фазе и дисциплиной с повторными обращениями на второй фазе. Показано, что выходящий поток рассматриваемой двухфазной системы при выполнении асимптотического условия растущей задержки на орбите относится к классу MAP. Проведены численные эксперименты, показывающие зависимость характеристик выходящего потока от параметров первой фазы.
Ключевые слова: *Теория массового обслуживания, выходящий поток, MAP, системы с повторными вызовами, коэффициент вариации.*

Введение

В работе предлагается исследование характеристик двухфазной системы, первая фаза которой представляет собой модель шлюза сбора и обработки трафика пакетов данных, а вторая фаза – модель передачи пакетов данных со шлюза в среде с протоколом случайного множественного доступа. Такая модель рассматривалась, в работе [1], в которой методом асимптотического анализа [2] исследуется распределение вероятностей числа заявок на орбите второй фазы. В работе [3] авторами было получено асимптотическое приближение характеристической функции числа заявок, обслуженных в предложенной двухфазной системе, при асимптотическом условии растущего времени задержки на орбите. Результаты работы [3] используются в текущей работе для определения класса выходящего потока в асимптотическом условии растущего времени задержки на орбите. Определение класса выходящего потока позволяет исследовать его характеристики в зависимости от параметров самой системы. Исследование характеристик выходящего потока позволяет оценить, как меняются характеристики трафика после прохождения узлов связи определённого вида.

1. Описание системы

На вход первой фазы поступает простейший поток заявок с параметром λ. На первой фазе заявки обслуживаются в порядке их поступления (FIFO). Время обслуживания заявок распределено по экспоненциальному закону с параметром μ_1. На первой фазе не может находиться более N заявок, включая заявку на обслуживании. Если на первой фазе уже находиться N заявок, то вновь пришедшая заявка теряется. После окончания обслуживания на первой фазе заявка обращается к прибору на второй фазе. Если прибор на второй фазе простаивает в этот момент, то поступившая заявка занимает его, где обслуживается в течение экспоненциально распределенного времени с параметром μ_2. Если же заявка, обращаюсь к прибору на второй фазе, обнаруживает прибор занятым, она мгновенно уходит на орбиту, где осуществляет случайную задержку в течение экспоненциально распределённого времени с параметром σ, после чего повторяет попытку занять прибор. На второй фазе заявки не теряются и уходят на орбиту после каждой неудачной попытки захватить прибор. После обслуживания на второй фазе заявка попадет в выходящий поток.

Модель данной системы представлена на рисунке 1.

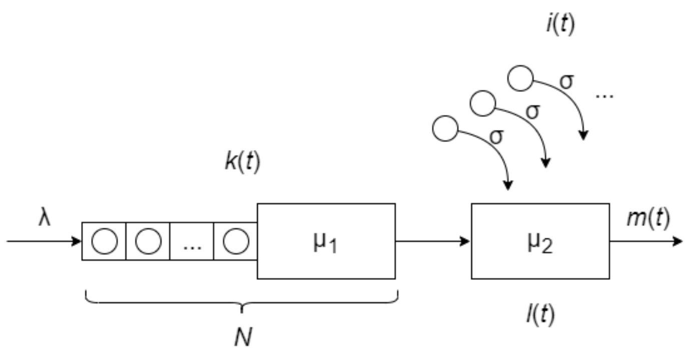

Рис. 1. Модель двухфазной системы

Функционирование системы описывается четырёхмерным Марковским процессом $\{k(t), i(t), l(t), m(t)\}$, где $k(t)$ — число заявок на первой фазе, $i(t)$ — число заявок на орбите в момент времени t, $l(t)$ — состояние прибора на второй фазе в момент времени t, $m(t)$ — число заявок в выходящем потоке за время t.

2. Выходящий поток

В работе [3] было получено асимптотическое приближение характеристической функции числа событий, наступивших в выходящем потоке за некоторое время t, при условии растущей задержки заявок на орбите:

$$\lim_{\sigma\to 0} M\{e^{jum(t)}\} = \mathbf{H}(u,t)\mathbf{e}, \tag{1}$$

где $\mathbf{e}$ — единичный вектор размерности $2N+2$, а вектор-строка $\mathbf{H}(u,t)$ той же размерности удовлетворяет матричному дифференциальному уравнению:

$$\frac{\partial \mathbf{H}(u,t)}{\partial t} = \mathbf{H}(u,t)\mathbf{J}(u), \tag{2}$$

где

$$\mathbf{J}(u) = \begin{bmatrix} \lambda\mathbf{C}^T - \mathbf{A} - \varkappa\mathbf{I} & \mu_1\mathbf{C} + \varkappa\mathbf{I} \\ e^{ju}\mu_2\mathbf{I} & \lambda\mathbf{C}^T - \mathbf{A} - \mu_2\mathbf{I} + \mu_1\mathbf{C} \end{bmatrix}_{[2N+2,2N+2]} \tag{3}$$

$$\mathbf{A} = \begin{bmatrix} \lambda & 0 & \dots & 0 & 0 \\ 0 & \lambda+\mu_1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda+\mu_1 & 0 \\ 0 & 0 & \dots & 0 & \mu_1 \end{bmatrix}_{[N+1,N+1]}, \tag{4}$$

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}_{[N+1,N+1]}.$$

$\varkappa$ – среднее число заявок на орбите [1], $\mathbf{I}$ – единичная матрица.

Известно [4], что уравнение для вектор-строки, определяющей характеристическую функцию числа заявок в MAP-потоке имеет вид:

$$\frac{\partial \mathbf{H}(u,t)}{\partial t} = \mathbf{H}(u,t)[\mathbf{Q} - (e^{ju} - 1)\mathbf{B}] \tag{5}$$

где матрицы $\mathbf{Q}$ и $\mathbf{B}$ определяют MAP-поток [2, 4].

Таким образом выходящий поток двухфазной системы в асимптотическом условии большой задержки на орбите можно считать MAP-потоком, определяемым матрицами (6) и (4).

$$\mathbf{Q} = \begin{bmatrix} \lambda\mathbf{C}^T - \mathbf{A} - \varkappa\mathbf{I} & \mu_1\mathbf{C} + \varkappa\mathbf{I} \\ \mu_2\mathbf{I} & \lambda\mathbf{C}^T - \mathbf{A} - \mu_2\mathbf{I} + \mu_1\mathbf{C} \end{bmatrix}, \tag{6}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ \mu_2\mathbf{I} & 0 \end{bmatrix}. \tag{7}$$

3. Численные эксперименты

Рассмотрим характеристики выходящего потока: C_{cor} – коэффициент корреляции длин интервалов между событиями, C_{var} – коэффициент вариации длин интервалов между событиями и λ_g – среднее число заявок в выходящем потоке за единицу времени. Поскольку мы показали, что наш выходящий поток в асимптотике относится к классу MAP, то воспользуемся формулами для вычисления C_{cor}, C_{var} и λ_g для MAP-потока [5]. Так как мы будем менять параметры первой фазы, вместе с выходящим потоком рассмотрим изменение характеристик первой фазы: $Ploss$ – вероятность потери заявки на первой фазе и $T0$ – среднее время ожидания в очереди. Зафиксируем значения $\lambda = 1$, $\mu_2 = 3$. Результаты представлены в табл. 1, 2, 3, 4, 5.

Таблица 1

C_{cor} – коэффициент корреляции длин интервалов в выходящем потоке

μ_1 \ N	3	5	7	9
2.5	$-7 \cdot 10^{-3}$	$-1 \cdot 10^{-3}$	$-2 \cdot 10^{-4}$	$-3 \cdot 10^{-5}$
2	$-12 \cdot 10^{-3}$	$-29 \cdot 10^{-4}$	$-7 \cdot 10^{-4}$	$-17 \cdot 10^{-5}$
0.5	$-23 \cdot 10^{-3}$	$-52 \cdot 10^{-4}$	$-13 \cdot 10^{-4}$	$-3 \cdot 10^{-4}$
0.1	$-56 \cdot 10^{-5}$	$-6 \cdot 10^{-6}$	$-6 \cdot 10^{-8}$	$-6 \cdot 10^{-10}$

Из табл. 1 видно, что значения корреляции длин интервалов в выходящем потоке для рассматриваемой модели не принимают абсолютных значений больше 10^{-3} для различных значений параметров системы. В общем случае корреляция незначительна.

Таблица 2

C_{var} – коэффициент вариации длин интервалов в выходящем потоке

μ_1 \ N	3	5	7	9
2.5	0.745	0.746	0.745	0.745
2	0.745	0.745	0.745	0.745
0.5	0.838	0.847	0.849	0.850
0.1	0.967	0.967	0.967	0.967

В табл. 2 видно, что значения вариации практически не зависят от числа мест в очереди. Коэффициент вариации во всех случаях принял значение меньше 1. Стоит отметить зависимость коэффициента вариации от коэффициента загрузки первой фазы: после того как интенсивность обслуживания становится больше интенсивности входящего потока, коэффициент вариации не изменяется.

Таблица 3

$Ploss$ – вероятность потери заявки на первой фазе

μ_1 \ N	3	5	7	9
2.5	0.039	0.006	0.001	0.001
2	0.067	0.016	0.004	0.001
0.5	0.533	0.508	0.502	0.500
0.1	0.900	0.900	0.900	0.900

Из табл. 3 можно сделать вывод о том, что при коэффициенте загрузки первой фазы меньше 0.5 и минимальной размерности буфера вероятность потери заявок становится приемлемой.

Таблица 4

$T0$ – среднее время ожидания на первой фазе

μ_1 \ N	3	5	7	9
2.5	0.462	0.615	0.655	0.664
2	0.571	0.839	0.945	0.982
0.5	1.429	3.161	5.055	7.018
0.1	1.892	3.889	5.889	7.889

Из табл. 4 можно сделать вывод о том, что время ожидания на первой фазе зависит от обоих параметров первой фазы. При увеличении числа мест в очереди, вновь пришедшая заявка будет вынуждена ждать, пока большее число заявок обслужится перед ней.

Таблица 5

λ_g – среднее число заявок в выходящем потоке

μ_1 \ N	3	5	7	9
2.5	0.961	0.994	0.999	1
2	0.933	0.984	0.996	0.999
0.5	0.467	0.492	0.498	0.5
0.1	0.1	0.1	0.1	0.1

В табл. 5 видно, что среднее число заявок в выходящем потоке практически не зависит от числа мест в буфере на первой фазе. Очевидно, что при увеличении интенсивности обслуживания на первой фазе, большее число заявок в среднем будет попадать в выходящий поток системы.

Заключение

В данной работе были представлены результаты исследования выходящего потока двухфазной системы, показано, что выходящий поток при условии большой задержки на орбите относится к классу MAP. Проведены численные эксперименты, иллюстрирующие зависимость характеристик системы и выходящего потока от параметров первой фазы.

СПИСОК ЛИТЕРАТУРЫ

1. *Назаров А. А., Анисимова А. А.* Асимтпотический анализ первого порядка двухфазной СМО с конечной очередью и орбитой в условии большой задержки заявок на орбите // Молодежная научная школа по прикладной теории вероятностей и телекоммуникационным технологиям (АРТСТ–2017) Москва: РУДН 2017. С. 175–177.
2. *Назаров А. А., Моисеева С. П.* Метод асимптотического анализа в теории массового обслуживания. Томск: Изд-во НТЛ, 2006. 112 с.
3. *Воркожоков А. А., Лапатин И. Л.* Исследование характеристик двухуровневой модели обработки запросов // Материалы VIII Международной молодежной научной конференции «Математическое и программное обеспечение информационных, технических и экономических систем». Томск: Издательство Томского государственного университета 2021. С. 118–121.
4. *Лопухова С. В.* Асимптотические и численные методы исследования специальных потоков однородных событий. Автореферат диссертации на соискание ученой степени кандидата физико-математических наук. Томск: Издательство Томского государственного университета 2008. 15 с.
5. *Вишневский В. М., Дудин А. Н.* Системы массового обслуживания с коррелированными входными потоками и их применение для моделирования телекоммуникационных сетей. // Автомат. и телемех., 2017, №8, 3–59.

Лапатин Иван Леонидович — канд. физ.-мат. наук, доцент кафедры прикладной информатики Томского государственного университета. E-mail: *ilapatin@mail.ru*

Воркожоков Андрей Алексеевич — студент кафедры прикладной информатики Томского государственного университета. E-mail: *vorkozhokov andrey@gmail.com*

ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ ХАРАКТЕРИСТИК МАРКОВСКОЙ МОДЕЛИ УЗЛА ПЕРЕДАЧИ ДАННЫХ ОТ ПАРАМЕТРОВ ПЕРЕКРЫТИЯ КАНАЛА

И. Л. Лапатин, О. А. Решетников

Национальный исследовательский
Томский государственный университет, г. Томск, Россия

В работе рассматривается модель узла связи в виде системы массового обслуживания с повторными обращениями и с «прогулками» прибора. «Прогулки» прибора моделируют ситуацию, когда узел связи периодически прекращает свою работу на случайное время. В ходе работы было показано, что выходящий поток рассматриваемой системы в асимптотическом условии большой задержки на орбите относится к классу MAP. Используя результаты численных экспериментов, показано влияние значений параметров системы на характеристики выходящего потока.
Ключевые слова: *Системы с повторными вызовами, выходящий поток, MAP, коэффициент вариации.*

Введение

Системы массового обслуживания с повторными вызовами моделируют работу узла передачи данных со случайным множественным доступом к каналу, где входящий поток является моделью агрегированного трафика от всех пользователей канала. Модели с повторными вызовами в различных модификациях достаточно много исследуются в последние десятилетия, причем основным объектом исследования является количество заявок на орбите. Так, например, в работе [1] была рассмотрена Марковская система массового обслуживания с «прогулками» прибора, то есть с периодическим выходом из строя обслуживающего прибора на случайное время. В работе [1] используется метод асимптотического анализа [2] для исследования числа заявок, ожидающих повторного обращения к прибору. В работе [3] для модели с «прогулками» прибора авторами получена формула вычисления асимптотической характеристической функции числа заявок в выходящем потоке при условии большой задержки заявок на орбите. Исследование характеристик выходящего потока всегда является актуальной темой, так как после прохождения одного узла связи обслуженные заявки поступают на другой, то есть выходящий поток одной системы является входящим для другой. И в общем случае характеристики потока могут меняться после прохождения системы. В представленной работе выходящий поток системы с повторными обращениями и «прогулками» прибора (которые можно так же назвать перекрытием канала связи) описывается в терминах MAP-потока при выполнении асимптотического условия растущего времени задержки заявок на орбите. В ходе численного эксперимента было показано влияние параметров системы на характеристики выходящего потока.

1. Математическая модель

На вход системы поступает простейший поток заявок с интенсивностью λ. Если прибор свободен, то поступающая заявка занимает его, где обслуживается в течение экспоненциально распределенного времени с параметром μ. Если же заявка, обращаясь к прибору, обнаруживает прибор занятым, она мгновенно уходит на орбиту, где осуществляет случайную задержку в течение экспоненциально распределённого времени с параметром σ, после чего повторяет попытку занять прибор. В системе заявки не теряются и уходят на орбиту после каждой неудачной попытки захватить прибор. Так же прибор в случайные моменты времени может уходить на «прогулку», то есть выходить из строя, с интенсивностью α_1.

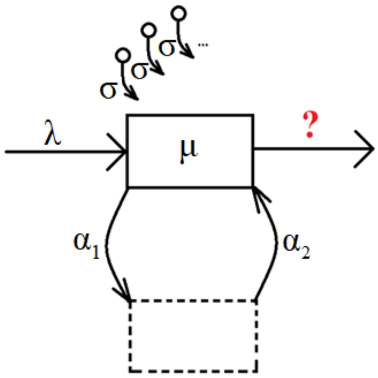

Рис. 1. Система с повторными обращениями и «прогулками» прибора

Если в момент выхода прибора из строя заявка была на обслуживании, то она отправляется на орбиту. Длительность «прогулки» имеет экспоненциальное распределение с параметром α_2. Во время «прогулки» прибора все поступающие заявки отправляются на орбиту. Обслуженные до конца заявки формируют события выходящего потока. Он характеризуется числом заявок, обслуженных к моменту времени t, т.е. за промежуток времени $[0,\ t]$.

Работа системы описывается рядом процессов:

1. $k(t)$ – состояние прибора в момент времени t.
2. $i(t)$ – количество заявок на орбите в момент времени t.
3. $m(t)$ – число обслуженных заявок за время t.

Для описания приведенной модели можно использовать теорию марковских процессов. Рассматриваемый процесс $m(t)$ сам по себе не является марковским, т.к. зависит еще и от состояния прибора в момент времени t, а следовательно и процесса $k(t)$, который в свою очередь зависит от числа заявок, находящихся на орбите – процесс $i(t)$. В таком случае, зная текущее состояние компонент трехмерного процесса $\{k(t),\ m(t),\ i(t)\}$, можно с вероятностной точки зрения определить будущее состояние, что согласуется с определением марковского процесса.

2. Выходящий поток

В работе [3] было получено асимптотическое приближение характеристической функции числа событий, наступивших в выходящем потоке за некоторое время t, при условии растущей задержки заявок на орбите:

$$\lim_{\sigma\to 0} M\{e^{ju_2 m(t)}\} = \mathbf{F}(u_2,t)\mathbf{e}, \tag{1}$$

где $\mathbf{e}$ – единичный вектор-столбец размерности 3, $\mathbf{F}(u_2,t)$ – вектор-строка размерности 3, удовлетворяющая матричному дифференциальному уравнению:

$$\frac{\partial \mathbf{F}(u_2,t)}{\partial t} = \mathbf{F}(u_2,t)\mathbf{M}(u_2), \tag{2}$$

где

$$\mathbf{M}(u_2) = \begin{bmatrix} -(\lambda+\alpha_1+\varkappa) & (\varkappa+\lambda) & \alpha_1 \\ \mu e^{ju_2} & -(\mu+\alpha_1) & \alpha_1 \\ \alpha_2 & 0 & -\alpha_2 \end{bmatrix},$$

$\varkappa$ – нормированное среднее число заявок на орбите.

Известно [4], что уравнение для вектор-строки, определяющей характеристическую функцию числа заявок в MAP-потоке, имеет вид схожий с (2):

$$\frac{\partial \mathbf{F}(u_2,t)}{\partial t} = \mathbf{F}(u_2,t)[\mathbf{Q}+(e^{ju_2}-1)\mathbf{B}], \tag{3}$$

где матрицы $\mathbf{Q}$ и $\mathbf{B}$ определяют MAP-поток[2,4].

Представим матрицу $\mathbf{M}(u_2)$ в виде $[\mathbf{Q}+(e^{ju_2}-1)\mathbf{B}]$. Тогда матрицы $\mathbf{Q}$ и $\mathbf{B}$ равны

$$\mathbf{Q}=\begin{bmatrix}-(\lambda+\alpha_1+\varkappa) & (\varkappa+\lambda) & \alpha_1\\ \mu & -(\mu+\alpha_1) & \alpha_1\\ \alpha_2 & 0 & -\alpha_2\end{bmatrix} \tag{4}$$

$$\mathbf{B}=\begin{bmatrix}0 & 0 & 0\\ \mu & 0 & 0\\ 0 & 0 & 0\end{bmatrix}. \tag{5}$$

Таким образом, выходящий поток системы с повторными обращениями и «прогулками» прибора в асимптотическом условии большой задержки на орбите можно считать MAP-потоком, определяемым матрицами (4) и (5).

3. Численные эксперименты

В ходе численных экспериментов путем изменения параметров частоты (α_1) и длительности (α_2) прогулки прибора проводилось наблюдение изменения нормированного среднего числа заявок на орбите – $\varkappa$, вероятности простоя прибора R_0, при котором прибор свободен и готов обслуживать заявки, коэффициента вариации длин интервалов между наступлениями событий в выходящем потоке – C_{var} и коэффициента корреляции длин интервалов между наступлениями событий в выходящем потоке – C_{cor}.

Для расчетов C_{var} и C_{cor} использовались формулы для MAP-потока, приведенные в [5]. При параметрах $\lambda = 1$, $\mu = 2$:

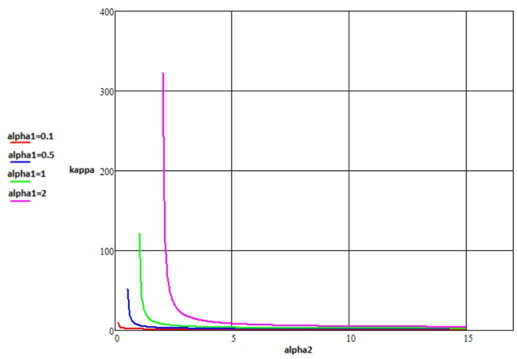

Рис. 2. Изменение нормированного среднего числа заявок на орбите $\varkappa$

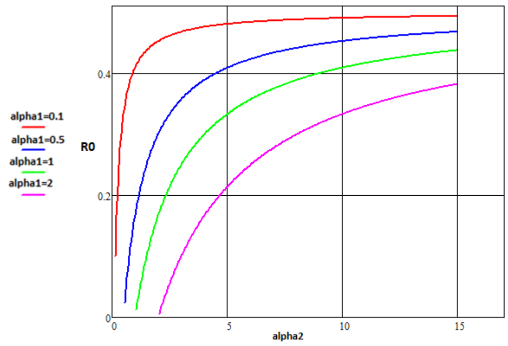

Рис. 3. Изменение вероятности простоя прибора R_0

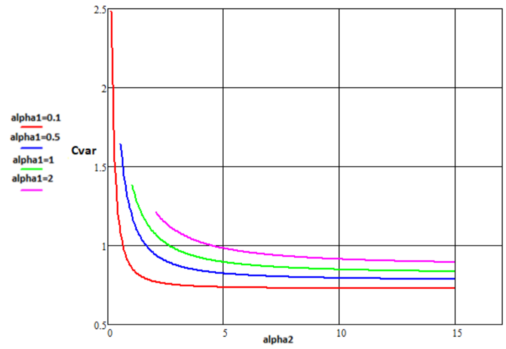

Рис. 4. Изменение коэффициента вариации длин интервалов C_{var}

Изменение коэффициента корреляции C_{cor} приняло нулевые значения во всех рассмотренных случаях.

Заключение

В ходе проделанной работы была рассмотрена модель узла связи в виде системы массового обслуживания с повторными обращениями и с «прогулками» прибора. Было показано, что асимптотический выходящий поток, при условии большой задержки на орбите, относится к классу MAP. Используя результаты проделанных численных экспериментов, можно проследить влияние изменений параметров перекрытия канала на характеристики выходящего потока.

СПИСОК ЛИТЕРАТУРЫ

1. *Рожкова С., Воронина Н., Семашко А.* Исследование RQ-системы M/M/1 с ненадежным прибором асимптотическим и матричным методами // ИТММ-2020 Материалы XIX Международной конференции имени А.Ф. Терпугова. Город: Томск Изд-во НТЛ, 2021. С. 244–251.
2. *Назаров А. А., Моисеева С. П.* Метод асимптотического анализа в теории массового обслуживания // Город: Томск Изд-во НТЛ, 2006. С. 112.
3. *Решетников О. А., Лапатин И. Л.* Исследование характеристик работы ненадежного узла обработки запросов // Материалы VIII Международной молодежной научной конференции «Математическое и программное обеспечение информационных, технических и экономических систем» Томск: Издательство Томского государственного университета, 2021. 140–144 с.
4. *Лопухова С. В.* Асимптотические и численные методы исследования специальных потоков однородных событий // Автореферат диссертации на соискание ученой степени кандидата физико-математических наук. Томский государственный университет. Город: Томск, 2008.
5. *Вишневский В. М., Дудин А. Н., Клименок В. И.* Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях // Москва. 2018. 115 с.

Лапатин Иван Леонидович — канд. физ.-мат. наук, доцент кафедры прикладной информатики. E-mail: *ilapatin@mail.ru*

Решетников Олег Алексеевич — студент института прикладной математики и компьютерных наук. E-mail: *oleg.reshetnikov.0411@gmail.com*

ВЛИЯНИЕ ПАРАМЕТРОВ ПОТОКА БАРТЛЕТТА НА ПРОЦЕСС УПРАВЛЕНИЯ КОНФЛИКТНЫМИ ПОТОКАМИ НА ПЕРЕКРЁСТКЕ

С. А. Лембриков, Е. В. Кувыкина

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

В работе рассмотрен специальный класс алгоритмов с обратной связью. Такие алгоритмы предназначены для управления потоками транспорта и пешеходов на перекрёстке. Специфической особенностью системы управления дорожным движением является изменение вероятностной структуры входных потоков под влиянием случайных факторов. Целью данного исследования является оптимизация работы системы для случая входных потоков Пуассона и Бартлетта, а также изучение влияния параметров потока Бартлетта на оптимальные параметры управления и минимальные средние задержки произвольного требования на перекрёстке.

Ключевые слова: *Управление движением, конфликтные потоки, поток Бартлетта, поток Пуассона, транспортные потоки, пешеходные потоки, интенсивность входных потоков, средняя длина пачки, оптимальные параметры управления, транспортная пачка.*

Введение

Задачи регулирования транспортных перекрёстков являются актуальными как с точки зрения безопасности, так и с точки зрения экологической ситуации [1–3]. При этом, в качестве участников дорожного движения могут выступать как транспортные единицы, так и пешеходы. Особенностью транспортных систем является изменение вероятностной структуры входных потоков под влиянием внешних факторов, которые зачастую носят случайный характер. К таким факторам можно отнести погоду. Так, при хороших погодных условиях, движение машин в потоке можно считать независимым. При плохих — возможности маневрирования ограничены, что приводит к образованию транспортных пачек. Потоки первого типа являются пуассоновскими, а потоки пачек описываются распределением Бартлетта [4]. Поэтому, при изучении транспортных систем важно установить, с одной стороны, влияние вероятностной структуры входных потоков, а с другой — влияние параметров потока Бартлетта на работу системы.

1. Постановка задачи

Рассматривается система массового обслуживания (СМО) с ожиданием. Она представляет собой «Т—образный» перекрёсток. В систему поступает три независимых конфликтных потока Π_1, Π_2 и Π_3, где Π_1 и Π_2 являются потоками машин, а Π_3 — потоком пешеходов. Движение происходит в пересекающихся направлениях, поэтому одновременное обслуживание потоков невозможно. Таким образом, потоки являются попарно конфликтными [5].

Обслуживающее устройство (ОУ) имеет восемь режимов работы $\Gamma^{(1)}, \Gamma^{(2)}, \ldots, \Gamma^{(8)}$. В режимах $\Gamma^{(1)}$ и $\Gamma^{(3)}$ разрешён переезд машинам потоков Π_1 и Π_2 соответственно. В $\Gamma^{(5)}$ и $\Gamma^{(7)}$ обслуживается поток пешеходов Π_3. Режимы $\Gamma^{(2)}, \Gamma^{(4)}, \Gamma^{(6)}$ и $\Gamma^{(8)}$ соответствуют жёлтому свету. Граф смены режимов работы ОУ представлен на рис. 1. Переходы $\Gamma^{(1)} \to \Gamma^{(2)}, \Gamma^{(3)} \to \Gamma^{(4)}, \Gamma^{(5)} \to \Gamma^{(6)}, \Gamma^{(6)} \to \Gamma^{(3)}, \Gamma^{(7)} \to \Gamma^{(8)}, \Gamma^{(8)} \to \Gamma^{(1)}$ определяются однозначно. Из режима $\Gamma^{(2)}$ ОУ переключается в режим $\Gamma^{(3)}$, если в момент окончания $\Gamma^{(2)}$ на перекрёстке скопилось не более, чем N_1 пешеходов. В противном случае, включается режим $\Gamma^{(5)}$. Аналогично, после режима $\Gamma^{(4)}$, включается режим $\Gamma^{(1)}$, если в момент окончания $\Gamma^{(4)}$ на перекрёстке скопилось не более, чем N_2 пешеходов. В противном случае, ОУ переходит в режим $\Gamma^{(7)}$. Таким образом, пешеходный поток Π_3 является информативным, а управляющий алгоритм — алгоритмом с обратной связью.

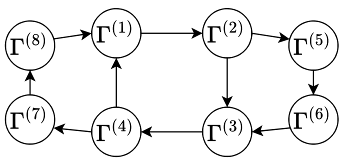

Рис. 1. Граф смены режимов работы ОУ

В каждом из режимов $\Gamma^{(r)}$ ОУ находится некоторое фиксированное время $T^{(r)}, r = \overline{1,8}$. Длительности режимов жёлтого света T_2, T_4, T_6 и T_8 выбираются из условий безопасности движения на перекрёстке. Параметры T_5 и T_7 также полагаются заданными. Длительности T_1 и T_3 являются управляющими параметрами алгоритма. В качестве оптимальных параметров рассматриваются те, при которых достигается минимальное значение средних задержек γ произвольного требования на перекрёстке. Данная характеристика является основным критерием качества работы системы с ожиданием и вычисляется по следующей формуле:

$$\gamma = (\lambda_1 \cdot \gamma_1 + \lambda_2 \cdot \gamma_2 \cdot \lambda_3 \cdot \gamma_3)/(\lambda_1 + \lambda_2 + \lambda_3), \tag{1}$$

где λ_j — интенсивность Π_j, а γ_j — средние задержки произвольной заявки в системе по потоку $\Pi_j, j = \overline{1,3}$. Задержки произвольного требования складываются из времени ожидания начала обслуживания и времени пересечения перекрёстка и представляют собой время пребывания требования в системе. Средние задержки γ_j по потоку $\Pi_j, j = \overline{1,3}$ рассчитываются как среднее арифметическое задержек заявок данного потока.

В качестве потоков Π_1 и Π_2 рассматриваются потоки Пуассона и Бартлетта. Поскольку возможности маневрирования и обгона у пешеходов значительно выше, чем у транспорта, их движение гораздо чаще является независимым. В следствие этого поток Π_3 в работе предполагается пуассоновским. Потоки Бартлетта описываются нелокально с помощью векторной случайной последовательности $\{(\tau_i, \eta_{1,i}, \eta_{2,i}), i \geqslant 0\}$ [4, 6]. В качестве моментов наблюдения $\tau_i, i \geqslant 0$ выбираются моменты поступления первых требований в пачках. Случайные величины $\eta_{j,i}$ описывают длину i-ой пачки потока Π_j и имеют распределение Бартлетта:

$$\begin{cases} P(\eta_{j,i} = 1) = 1 - r_j, \\ P(\eta_{j,i} = k) = r_j \cdot (1 - g_j) \cdot {g_j}^{k-2}, k \geqslant 2, \end{cases} \tag{2}$$

где r_j,g_j — параметры распределения, $0 < r_j, g_j < 1, j = 1,2$.

Под длиной пачки понимается количество заявок, поступивших в систему за i-ый такт наблюдения $[\tau_i, \tau_{i+1}), i \geqslant 0$. Параметр r_j — это вероятность того, что в систему поступит пачка из двух и более заявок по потоку $\Pi_j, j = 1,2$. Параметр g_j влияет на среднюю длину пачки:

$$M_{\eta_{j,i}} = 1 + r_j/(1 - g_j), j = 1,2. \tag{3}$$

Таким образом, параметры r_j и g_j связаны следующим соотношением:

$$r_j = (M_{\eta_{j,i}} - 1) \cdot (1 - g_j), j = 1,2. \tag{4}$$

Нетрудно видеть, что при фиксированной средней длине пачки, с ростом параметра g_j, параметр r_j убывает, $j = 1,2$.

2. Численные исследования

Численные исследования данной системы проведены с применением метода имитационного моделирования. Расчёты показали, что вероятностная структура входных потоков существенно влияет на поведение системы. При управлении потоками Пуассона и Бартлетта одинаковой интенсивности, во втором случае наблюдается рост минимальных средних задержек произвольной заявки на перекрёстке. Установлено, что с ростом параметра g_j происходит резкое увеличение средних задержек. Для иллюстрации данного положения рассмотрим пример 1.

Пример 1. Пусть длительности режимов жёлтого света $T_2 = T_4 = T_6 = T_8 = 3$ (с), а режимов обслуживания пешеходов — $T_5 = T_7 = 30$ (с). Критические значения для очередей по потоку пешеходов Π_3 примем равными нулю: $N_1 = N_2 = 0$ (заявок). Пусть интенсивности потоков Π_1, Π_2 и Π_3 совпадают: $\lambda_1 = \lambda_2 = \lambda_3 = 0,1$ (заявок в секунду). Рассмотрим два случая входных потоков. Если Π_1, Π_2 и Π_3 — пуассоновские потоки, то минимальное значение средних задержек произвольной заявки в системе составляет $33,72$ (с). Данные результаты представлены на рис. 2 в виде прямой I, помеченной оранжевым цветом. Во втором случае Π_1 и Π_2 — потоки Бартлетта со средней длиной пачки $M_{\eta_{j,i}} = 2$ (заявок), $j = 1, 2$. При этом, значение параметра $g_j, j = 1, 2$ изменяется в пределах от $0,1$ до $0,99$: на интервале $[0,1; 0,9)$ — с шагом $0,1$, а на интервале $[0,9; 0,99)$ — с шагом $0,01$. Параметр $r_j, j = 1, 2$ определяется из формулы (4). Увеличение параметра g_j приводит к росту минимальных средних задержек с $34,72$ (с) при $g_j = 0,1$ до $161,41$ (с) при $g_j = 0,99, j = 1, 2$. Причём, начиная с уровня $g_j = 0,8, j = 1, 2$, наблюдаются существенные отличия от случая пуассоновских входных потоков. Данные результаты также представлены на рис. 2 (кривая II, помеченная синим цветом).

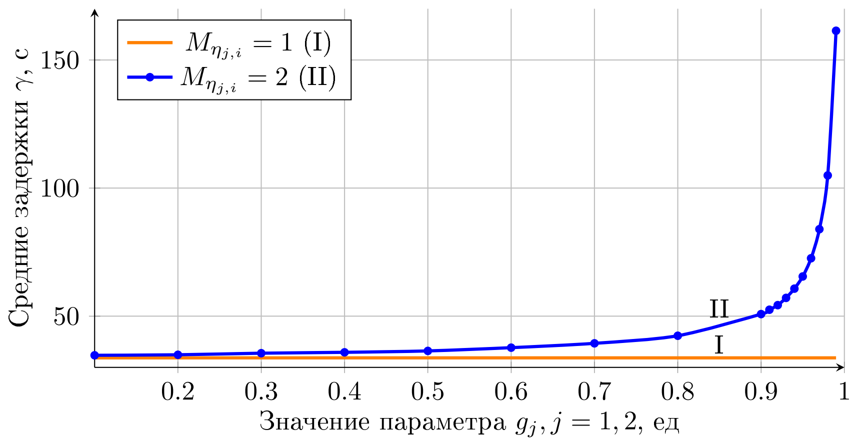

Рис. 2. Зависимость минимальных средних задержек от параметра $g_j, j = 1, 2$

Причину такого влияния параметра $g_j, j = 1, 2$ на работу системы продемонстрируем на примере 2.

Пример 2. Зафиксируем среднюю длину пачки: $M_{\eta_{j,i}} = 11$ (заявок) и рассмотрим два набора параметров. Пусть $g_j = 0,91$. Тогда, согласно формуле (4), параметр $r_j = 0,9, j = 1, 2$. Этому случаю соответствует кривая I на рис. 3, помеченная синим цветом. Тогда вероятность того, что в систему поступит пачка из 2 заявок, будет равна $P(\eta_{j,i} = 2) = r_j \cdot (1 - g_j) \cdot (g_j)^0 = 0,081$ согласно формуле (2). Увеличим значение параметра g_j до $0,99$, тогда $r_j = 0,1$ согласно (4), $j = 1, 2$. Данный набор параметров соответствует кривой II, помеченной оранжевым цветом. В этом случае, вероятность того, что в систему поступит пачка из 2 заявок, будет гораздо меньше: $P(\eta_{j,i} = 2) = 0,001$. Действительно, кривая II изначально лежит ниже кривой I. В первую очередь, это связано со значением параметра $r_j, j = 1, 2$, так как именно он определяет вероятность того, что в систему поступит пачка из двух и более заявок. Согласно (2), вероятность $P(\eta_{j,i} = k)$ представляет собой степенную функцию параметра g_j, и параметр g_j принимает значения строго меньше единицы. В связи с этим, с увеличением k, вероятность $P(\eta_{j,i} = k)$ убывает гораздо быстрее для первого набора параметров. Именно из-за большей вероятности поступления пачек большой длины минимальное значение средних задержек произвольной заявки в системе увеличивается.

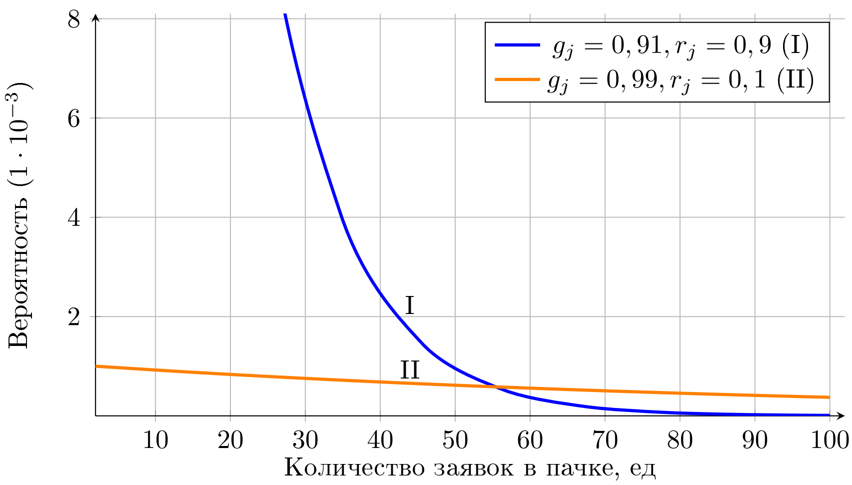

Рис. 3. Зависимость вероятности поступления пачки от длины k

Заключение

Проведённые исследования показали, что изменение вероятностной структуры входных потоков влияет на работу системы — наименьшее значение минимальных средних задержек произвольного требования на перекрёстке достигается при управлении пуассоновскими входными потоками. Также удалось выяснить, что соотношение параметров потока Бартлетта оказывает значительное воздействие на работу перекрёстка. Установлено, что преимущественно параметр $g_j, j = 1, 2$ влияет на вероятность поступления в систему пачек большой длины, а потому с его ростом происходит резкое увеличение средних задержек γ.

СПИСОК ЛИТЕРАТУРЫ

1. *Кудрявцев Е. В.* Анализ дискретной модели системы адаптивного управления конфликтными неоднородными потоками // Вестник Московоского университета. 2019. № 1. С. 19–26.
2. *Загутин Д. С., Скудина А. А., Бахтеев О. А., Миронов С. А.* Исследование параметров установки транспортных и пешеходных светофоров // Инженерный вестник Дона. 2019. № 1.
3. *Бабичева Т. С.* Методы теории массового обслуживания при исследовании и оптимизации движения на управляемых перекрёстках // Труды МФТИ. 2015. № 2. С. 119–130.
4. *Федоткин М. А.* Неполное описание потоков неоднородных требований. Теория массового обслуживания. Москва: МГУ, ВНИИСИ, 1981. 118 с.
5. *Зорин А. В.* Оптимизация параметров управления конфликтными потоками в классе циклических алгоритмов // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. № 3. С. 70–77.
6. *Кувыкина Е. В.* Общая постановка задачи об алгоритмическом управлении конфликтными потоками заявок при их нелокальном описании // Моделирование динамических систем. Сборник научных трудов Нижегородского филиала института машиноведения РАН. 2007. С. 65–71.

Лембриков Степан Андреевич — магистрант, кафедра теории вероятностей и анализа данных института информационных технологий математики и механики ННГУ. E-mail: *stepan.lembrikov.ru@mail.ru*

Кувыкина Елена Вадимовна — кандидат физико—математических наук, доцент, кафедра теории вероятностей и анализа данных института информационных технологий математики и механики ННГУ. E-mail: *e.kuvykina@yandex.ru*

ИССЛЕДОВАНИЕ СМО ВИДА $M|M|1|N-1$ С ТРЕМЯ ТИПАМИ ОБРАТНОЙ СВЯЗИ МЕТОДОМ АСИМПТОТИЧЕСКИ ДИФФУЗИОННОГО АНАЛИЗА

А. А. Назаров, Е. А. Павлова, А. А. Павлова

Национальный исследовательский
Томский государственный университет, г. Томск, Россия

В данной работе исследована система массового обслуживания с единственным обслуживающим прибором, а также источником повторных вызовов - орбитой, имеющей неограниченное количество мест. На вход системы поступает простейший поток заявок, обслуживание которых осуществляется в соответствии с экспоненциальным законом распределения. Для данной системы характерно наличие трех типов обратной связи для обрабатываемой заявки. Для нахождения распределения вероятностей числа заявок на орбите был применен метод асимптотически диффузионного анализа.

Ключевые слова: *Система массового обслуживания с обратной связью, асимптотически диффузионный анализ.*

Введение

При исследовании классических СМО не учитывается возможность получения обратной связи заявки, что в значительной степени отдаляет систему от реальных процессов в области коммуникации. Системы массового обслуживания с обратной связью характеризуются тем, что при получении первичного обслуживания заявка может, не покидая системы, либо мгновенно получить повторное обслуживание, либо осуществить задержку в течение случайного времени, а затем вновь совершить попытку получить обслуживание (отложенная обратная связь) [1].

Большое количество работ посвящено системам массового обслуживания такого типа [2-5], что обусловлено оптимизацией всевозможных граней описываемой области. В данной работе предлагается рассмотреть модель с тремя типами обратной связи: мгновенной, отложенной через бункер и отложенной через орбиту.

1. Описание модели и постановка задачи

Будем исследовать СМО вида $M|M|1|N-1$ (рис. 1) с тремя типами обратной связи, на вход которой поступает простейший поток заявок с параметром λ. При условии наличия свободного места в бункере, имеющем $N-1$ место, поступившая заявка встает в очередь. Если свободен прибор, заявка отправляется на него и обслуживается там в течение случайного времени, распределенного экспоненциально с параметром μ. Если же все места в бункере заняты, заявка уходит на орбиту с неограниченным количеством мест и задерживается там в течение случайного времени, распределенного экспоненциально с параметром σ. По завершении обслуживания заявки на приборе она покидает систему с вероятностью r_0; с вероятностью r_1 заявка мгновенно получает повторное обслуживание; с вероятностью r_2 – уходит в бункер, а с вероятностью r_3 – на орбиту.

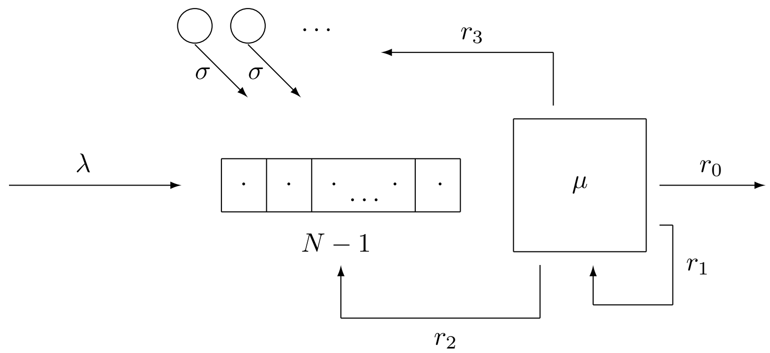

Рисунок 1. Система вида $M|M|1|N-1$ с тремя типами обратной связи

Обозначим $n(t)$ – число заявок на приборе и в очереди в момент времени t, $n(t)=0,1,...,N$,$i(t)$ – число заявок на орбите. Случайный двумерный марковский процесс, который будем исследовать – $\{n(t),i(t)\}$.

Ставится задача нахождения двумерного распределения вероятностей $P(n,i,t)=P\{n(t)=n,i(t)=i\}$ в стационарном режиме.

2. Система уравнений Колмогорова

Для распределения вероятностей $P(n,i,t)$ запишем систему дифференциальных уравнений Колмогорова

$$\frac{\partial P(0,i,t)}{\partial t}=-P(0,i,t)(i\sigma+\lambda)+P(1,i,t)\mu r_0+P(1,i-1,t)\mu r_3,\ n=0;$$

$$\frac{\partial P(n,i,t)}{\partial t}=-P(n,i,t)(i\sigma+\lambda+\mu)+P(n-1,i,t)\lambda+$$

$$+P(n-1,i+1,t)\sigma(i+1)+P(n+1,i,t)\mu r_0+P(n,i,t)\mu(r_1+r_2)+$$

$$+P(n+1,i-1,t)\mu r_3,\ 0<n<N;$$

$$\frac{\partial P(N,i,t)}{\partial t} = -P(N,i,t)(\lambda+\mu) + P(N-1,i,t)\lambda + P(N,i-1,t)\lambda +$$
$$+P(N-1,i+1,t)\sigma(i+1) + P(N,i,t)\mu(r_1+r_2),\ n=N.$$

Введем частичные характеристические функции и перейдем к ним

$$H(n,u,t) = \sum_{i=0}^{\infty} e^{jui} P(n,i,t),$$

здесь j – мнимая единица ($j=\sqrt{-1}$).

$$\frac{\partial H(0,u,t)}{\partial t} = -\lambda H(0,u,t) + j\sigma \frac{\partial H(0,u,t)}{\partial u} + r_0 \mu H(1,u,t) +$$
$$+r_3 \mu e^{ju} H(1,u,t),\ n=0;$$
$$\frac{\partial H(n,u,t)}{\partial t} = -(\lambda+\mu)H(n,u,t) + j\sigma \frac{\partial H(n,u,t)}{\partial u} + \lambda H(n-1,u,t) -$$
$$-j\sigma e^{-ju} \frac{\partial H(n-1,u,t)}{\partial u} + \mu r_0 H(n+1,u,t) + (r_1+r_2)\mu H(n,u,t) +$$
$$+r_3 \mu e^{ju} H(n+1,u,t),\ 0<n<N;$$
$$\frac{\partial H(N,u,t)}{\partial t} = -(\lambda+\mu)H(N,u,t) +$$
$$+\lambda H(N-1,u,t) + \lambda e^{ju} H(n,u,t) - j\sigma e^{-ju} \frac{\partial H(N-1,u,t)}{\partial u} +$$
$$+(r_1+r_2)\mu H(N,u,t),\ n=N. \tag{1}$$

Для удобства записи перейдем в (1) к матричной форме, обозначив

$$\mathbf{H}(u,t) = \{H(0,u,t)), H(1,u,t), \ldots, H(N,u,t)\},$$
$$\frac{\partial \mathbf{H}(u,t)}{\partial t} = \mathbf{H}(u,t)(\mathbf{A} + e^{ju}\mathbf{B}) + j\sigma \frac{\partial \mathbf{H}(u,t)}{\partial u}(\mathbf{I_0} - e^{-ju}\mathbf{I_1}), \tag{2}$$

где $\mathbf{A}, \mathbf{B}, \mathbf{I_0}, \mathbf{I_1}$ – матрицы размерности $(N+1)\times(N+1)$ вида

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & \ldots & 0 \\ \mu r_0 & \mu((r_1+r_2)-1)-\lambda & \ldots & 0 \\ 0 & \mu r_0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda \\ 0 & 0 & \ldots & \mu((r_1+r_2)-1)-\lambda \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \mu r_3 & 0 & \dots & 0 & 0 \\ 0 & \mu r_3 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mu r_3 & \lambda \end{pmatrix}, \mathbf{I_0} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix},$$

$$\mathbf{I_1} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Домножим уравнение в матричном виде (1) на единичный вектор-столбец $\mathbf{e}$ и получим скалярное уравнение, принимая во внимание равенства $(\mathbf{A}+\mathbf{B})\mathbf{e} = 0$ и $(\mathbf{I_0} - \mathbf{I_1})\mathbf{e} = 0$. Выпишем полученную систему

$$\frac{\partial \mathbf{H}(u,t)}{\partial t} = \mathbf{H}(u,t)(\mathbf{A} + e^{ju}\mathbf{B}) + j\sigma \frac{\partial \mathbf{H}(u,t)}{\partial u}(\mathbf{I_0} - e^{-ju}\mathbf{I_1}),$$

$$\frac{\partial \mathbf{H}(u,t)}{\partial t}\mathbf{e} = \mathbf{H}(u,t)\mathbf{B}\mathbf{e} + j\sigma \frac{\partial \mathbf{H}(u,t)}{\partial u} e^{-ju}\mathbf{I_1}\mathbf{e}. \tag{3}$$

Для решения задачи (2) методом асимптотически диффузионного анализа введем асимптотическое условие $\sigma \to 0$, описывающее большую задержку заявки на орбите.

3. Асимптотически диффузионный анализ

Обозначим $\sigma = \varepsilon$ и введем следующие замены

$$u = \varepsilon w, \tau = t\varepsilon, \mathbf{H}(u,t) = \mathbf{F}(w,\tau,\varepsilon).$$

Перепишем (2) согласно введенным заменам

$$\varepsilon \frac{\partial \mathbf{F}(w,\tau,\varepsilon)}{\partial \tau} = \mathbf{F}(w,\tau,\varepsilon)(\mathbf{A} + e^{jew}\mathbf{B}) + j\frac{\partial \mathbf{F}(w,\tau,\varepsilon)}{\partial w}(\mathbf{I_0} - e^{-jew}\mathbf{I_1}),$$

$$\varepsilon \frac{\partial \mathbf{F}(w,\tau,\varepsilon)}{\partial \tau}\mathbf{e} = \mathbf{F}(w,\tau,\varepsilon)\mathbf{B}\mathbf{e} + j\frac{\partial \mathbf{F}(w,\tau,\varepsilon)}{\partial w} e^{-jew}\mathbf{I_1}\mathbf{e}.$$

Теорема 1. Пусть имеется система массового обслуживания с тремя типами обратной связи, тогда в асимптотическом условии $\sigma \to 0$ предельную характеристическую функцию при условии нормировки числа заявок на орбите можно представить в следующем виде

$$\lim_{\sigma \to 0} M\left\{e^{jw\sigma i(t)}\right\} = e^{jwx(\tau)}. \tag{4}$$

Функцию $x(\tau)$ в (4) можно найти из решения следующего дифференциального уравнения $x'(\tau) = a(x)$, здесь

$$a(x) = \mathbf{R}(\mathbf{B} - x(\tau)\mathbf{I_1})\mathbf{e}, \tag{5}$$

а вектор стационарного распределения числа заявок в исследуемой системе $\mathbf{R} = \{R(0), R(1), \ldots, R(N)\}$, компоненты которого это вероятности того, что в системе находится k заявок ($k = 0, 1, \ldots, N$), причем вектор $\mathbf{R}$ есть решение системы

$$\mathbf{R}\left\{(\mathbf{A} + \mathbf{B}) - x(\tau)(\mathbf{I_0} - \mathbf{I_1})\right\} = 0,$$

$$\mathbf{Re} = 1.$$

Выполняя центрирование случайного процесса $i(t)$, можно сформулировать следующее утверждение.

Теорема 2. В рассматриваемой СМО в предельном условии $\sigma \to 0$ асимптотическая характеристическая функция представима в виде

$$\lim_{\sigma \to 0} Mexp\left\{jw\sqrt{\sigma}(i(\frac{\tau}{\sigma}) - \frac{x(\tau)}{\sigma})\right\} = \Phi(w, \tau),$$

здесь функция $\Phi(w, \tau)$ является решением следующего уравнения

$$\frac{\partial \Phi(w, \tau)}{\partial \tau} = \frac{(jw)^2}{2} b(x)\Phi(w, \tau) + w\frac{\partial \Phi(w, \tau)}{\partial w} a'(x),$$

где функция $b(x)$ определяется как

$$b(x) = a(x) + 2\mathbf{g}\left\{\mathbf{B} - x(\tau)\mathbf{I_1}\right\}\mathbf{e} + \mathbf{R}x(\tau)\mathbf{I_1}\mathbf{e}, \tag{6}$$

здесь вектор $\mathbf{g}$ определяется единственным образом из решения системы

$$\mathbf{g}[\mathbf{A} + \mathbf{B} - x(\tau)(\mathbf{I_0} - \mathbf{I_1})] = \mathbf{R}(x(\tau)\mathbf{I_1} - \mathbf{B}) + a(x)\mathbf{R},$$

$$\mathbf{ge} = 0.$$

Можно показать, что имеет место следующее утверждение.

Теорема 3. В рассматриваемой СМО стационарная предельная плотность распределения вероятностей числа заявок на орбите представима в виде

$$\Pi(z) = \frac{C}{b(z)} \cdot exp\left\{\frac{2}{\sigma}\int_0^z \frac{a(x)}{b(x)}dx\right\},$$

где функция $a(x)$ имеет вид (5), а функция $b(x)$ – (4), C – это нормирующая константа.

Заключение

В данной работе исследована система массового обслуживания вида $M|M|1|N-1$ с тремя типами обратной связи методом асимптотически диффузионного анализа в предельном условии большой задержки заявки на орбите ($\sigma \to 0$). В результате получено асимптотическое распределение вероятностей нормированного числа заявок на орбите.

СПИСОК ЛИТЕРАТУРЫ

1. *Назаров А. А., Федорова Е. А.* Исследование RQ-системы $MMPP|GI|1$ методом асимптотического анализа второго порядка в условии большой загрузки // Изв. ТПУ. Информационные технологии. 2014. Т. 325. № 5. С. 6–15.
2. *Шахбазов А. А.* Обслуживание с разогревом и зависимостью времени обслуживания от длины очереди. // Изв. АН АзССРР, сер. физ.-техн. и матем. наук. 1974. № 1. С. 32–35.
3. *Назаров А. А, Рожкова С. В., Титаренко Е. Ю.* Исследование $M|M|1$ системы с повторными вызовами, обратой связью и неординарным пуассоновским входящим потоком. // Математика, ее приложения и математическое образование, Улан-Удэ: ВСГУТУ, 2020. С. 160–162.
4. *Громов А. И.* О двухфазной системе массового обслуживания с обратной связью // Автоматика и механика. 1988. № 12. С. 93–97.
5. *Королюк В. С., Меликов А. З., Пономаренко Л. А., Рустамов А. М.* Методы анализа многоканальной системы обслуживания с мгновенной и отсроченной обратными связями. // Кибернетика и системный анализ. 2016. Т. 52. № 1. С. 64–77.

Назаров Анатолий Андреевич — д.т.н., профессор, заведующий кафедрой теории вероятностей и математической статистики. E-mail: *nazarov.tsu @gmail.com*

Павлова Екатерина Алексеевна — аспирант института прикладной математики и компьютерных наук. E-mail: *pavlovakatya_ 2010@mail.ru*

Павлова Анна Алексеевна — студент института прикладной математики и компьютерных наук. E-mail: *Annapavlova161199@mail.ru*

АСИМПТОТИЧЕСКИЙ АНАЛИЗ RQ-СИСТЕМЫ M/M/1/N–1 С ОБРАТНЫМИ СВЯЗЯМИ

А. А. Назаров[1], С. В. Рожкова[1,2], Е. Ю. Титаренко[1,2]

[1] *Национальный исследовательский Томский государственный университет, г. Томск, Россия*
[2] *Национальный исследовательский Томский политехнический университет, г. Томск, Россия*

В работе асимптотическим методом исследована RQ-система M/M/1/N–1 с мгновенной и отложенной обратными связями. В системе имеется один обслуживающий прибор и буфер, содержащий N–1 мест для ожидания. Входящая заявка, заставшая прибор занятым, а буфер заполненным, попадает на орбиту, где ожидает случайное время и повторяет попытку попасть на обслуживание или в очередь. После обслуживания заявка либо покидает систему, либо повторно поступает на обслуживание, либо переходит на орбиту. В работе показано, что асимптотическое распределение вероятностей числа заявок на орбите при условии растущего среднего времени ожидания на орбите является гауссовским, получены уравнения для нахождения параметров распределения.
Ключевые слова: *Асимптотический анализ, RQ-система, очередь, обратная связь.*

Введение

В теории массового обслуживания рассматриваются системы с потерей заявок (с отказами) и системы с ожиданием, имеющие конечный или бесконечный буфер для ожидания. Во втором случае заявка, поступившая в систему и заставшая прибор занятым, становится в очередь. Особое значение имеют системы с повторными вызовами (RQ-системы), в которых заявка, заставшая прибор занятым, ожидает на орбите случайное время и повторяет попытку попасть на обслуживание. Поскольку эффект повторных попыток типичен для многих телекоммуникационных сетей, исследованию RQ-систем посвящено большое количество работ [1], [2]. Кроме того, в реальных сетях связи происходит повторная передача ошибочно переданных данных. В системах массового обслуживания такая возможность обеспечивается обратной связью. Модели с обратными связями мало изучены. В [3] рассматривается модель системы с бесконечным буфером, мгновенной обратной связью матрично-геометрическим методом.

В данной работе рассматривается система с повторными вызовами, имеющая один обслуживающий прибор и буфер, содержащий $N-1$ мест для ожидания. Входящая заявка, заставшая прибор занятым, а буфер заполненным, попадает на орбиту, где ожидает случайное время и повторяет попытку попасть на обслуживание или в очередь. В системе учитывается возможность повторного обслуживания в виде мгновенной и отсроченной обратной связи. Система исследуется методом асимптотического анализа при условии растущего среднего времени ожидания заявок на орбите.

1. Описание модели и постановка задачи

Рассмотрим систему массового обслуживания M/M/1/N–1 с повторными вызовами (рис. 1). На вход поступает простейший поток заявок с интенсивностью λ. Если обслуживающий прибор свободен, то заявка поступает на обслуживание. Если прибор занят, то при наличии мест в буфере для ожидания заявка становится в очередь, а при отсутствии мест попадает на орбиту. Время обслуживания заявки распределено по экспоненциальному закону с параметром μ. Заявка, обслуживание которой завершено, покидает систему с вероятностью r_0, мгновенно поступает на повторное обслуживание с вероятностью r_1 или переходит на орбиту с вероятностью r_2, таким образом $r_0 + r_1 + r_2 = 1$. На орбите заявки ожидают повторного обслуживания в течение времени, распределенного по экспоненциальному закону с параметром σ, после чего делают попытку встать в очередь. В случае отсутствия мест в буфере заявки остаются на орбите.

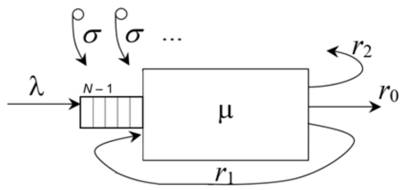

Рис. 1. Схема системы

Обозначим $i(t)$ – число заявок на орбите в момент времени t, процесс $n(t)$ определяет состояние прибора и буфера для ожидания следующим образом: $n(t) = 0$, если прибор свободен; $n(t) = n$, если прибор занят и в очереди $n-1$ заявка, $n = \overline{1, N}$. Двумерный процесс

$\{i(t), n(t)\}$ является цепью Маркова с непрерывным временем. Обозначим вероятности числа заявок на орбите с учетом состояния прибора $P_n(i,t) = P\{i(t) = i,\ n(t) = n\}, n = \overline{0,N};\ i = \overline{0,\infty}$. Требуется найти стационарное распределение вероятностей числа заявок на орбите.

2. Уравнения Колмогорова

Для стационарного распределения вероятностей $P_n(i) \equiv P_n(i,t)$, запишем систему уравнений

$$
\begin{aligned}
&-(\lambda + i\sigma)\,P_0(i) + \mu r_0 P_1(i) + \mu r_2 P_1(i-1) = 0;\\
&\lambda P_{n-1}(i) + (i+1)\sigma P_{n-1}(i+1) - (\lambda + \mu r_0 + \mu r_2 + i\sigma)\,P_n(i) + \mu r_0 P_{n+1}(i) +\\
&+ \mu r_2 P_{n+1}(i-1) = 0, \quad n = \overline{1, N-1};\\
&(i+1)\sigma P_{N-1}(i+1) - (\lambda + \mu r_0 + \mu r_2)\,P_N(i) + \lambda P_{N-1}(i) + \lambda P_N(i-1) = 0.
\end{aligned}
$$

Введем частичные характеристические функции числа заявок на орбите $H_n(u) = \sum_{i=0}^{\infty} e^{jui} P_n(i)$ и преобразуем систему к виду

$$
-\lambda H_0(u) + j\sigma \frac{\partial H_0(u)}{\partial u} + \left(\mu r_0 + \mu r_2 e^{ju}\right) H_1(u) = 0; \tag{1}
$$

$$
\begin{aligned}
&\lambda H_{n-1}(u) - j\sigma e^{-ju} \frac{\partial H_{n-1}(u)}{\partial u} - (\lambda + \mu r_0 + \mu r_2)\,H_n(u) + j\sigma \frac{\partial H_n(u)}{\partial u} +\\
&+ \left(\mu r_0 + \mu r_2 e^{ju}\right) H_{n+1}(u) = 0, \quad n = \overline{1, N-1};\\
&\lambda H_{N-1}(u) - j\sigma e^{-ju} \frac{\partial H_{N-1}(u)}{\partial u} - \left(\lambda - \lambda e^{ju} + \mu r_0 + \mu r_2\right) H_N(u) = 0.
\end{aligned}
$$

Полная характеристическая функция числа заявок на орбите $H(u) = \sum_{n=0}^{N} H_n(u)$. Сложим уравнения системы (1), получим

$$
\mu r_2 \sum_{n=1}^{N-1} H_n(u) + j\sigma e^{-ju} \sum_{n=0}^{N-1} \frac{\partial H_n(u)}{\partial u} + (\lambda + \mu r_2)\,H_N(u) = 0. \tag{2}
$$

3. Асимптотика первого порядка

Решим уравнения для характеристической функции (1), (2) при асимптотическом условии растущего среднего времени ожидания на орбите, то есть будем полагать, что $\sigma \to 0$. Результат сформулируем в виде следующей теоремы.

Теорема 1. Пусть $i(t)$ – число заявок на орбите в RQ-системы $M/M/1/N-1$ с обратными связями, тогда для последовательности характеристических функций выполняется равенство $\lim\limits_{\sigma\to 0} M\left\{e^{jwi(t)\sigma}\right\} = e^{jw\varkappa_1}$, где $\varkappa_1$ является решением уравнения

$$\frac{d^N-1}{d-1}\varkappa_1 = \mu r_2 \frac{d(d^N-1)}{d-1} + \lambda d^N, d = \frac{\lambda+\varkappa_1}{\mu r_0 + \mu r_2}. \tag{3}$$

Доказательство. Обозначим $\sigma = \varepsilon$ и сделаем в системе (1), (2) замены: $u = \varepsilon w$, $H_n(u) = F_n(w,\varepsilon)$, $n = \overline{0,N}$. В полученной системе уравнений выполним предельный переход при $\varepsilon \to 0$ и обозначим $F_n(w) = \lim\limits_{\varepsilon\to 0} F_n(w,\varepsilon)$. Будем искать решение системы в виде $F_n(w) = R_n\Phi(w) + O(\varepsilon)$. После преобразований получим систему

$$-\lambda R_0 + jR_0\frac{\Phi'(w)}{\Phi(w)} + (\mu r_0 + \mu r_2)R_1 = 0; \tag{4}$$

$$\lambda R_{n-1} - jR_{n-1}\frac{\Phi'(w)}{\Phi(w)} - (\lambda + \mu r_0 + \mu r_2)R_n + jR_n\frac{\Phi'(w)}{\Phi(w)} +$$
$$+ (\mu r_0 + \mu r_2)R_{n+1} = 0, \quad n = \overline{1, N-1};$$
$$\lambda R_{N-1} - jR_{N-1}\frac{\Phi'(w)}{\Phi(w)} - (\mu r_0 + \mu r_2)R_n = 0;$$
$$\mu r_2\sum_{n=1}^{N} R_n + j\frac{\Phi'(w)}{\Phi(w)}\sum_{n=0}^{N-1} R_n + \lambda R_N = 0.$$

Из системы (4) видно, что $\Phi'(w)/\Phi(w)$ не зависит от w, тогда можно обозначить $\Phi'(w)/\Phi(w) = j\varkappa_1$ и записать $\Phi(w)$ в виде $\Phi(w) = \exp\{jw\varkappa_1\}$, что и утверждается в формулировке теоремы. Решая первые $N+1$ уравнений системы (4) с учетом условия $\sum_{n=0}^{N} R_n = 1$, получим $R_n = \frac{d^n(d-1)}{d^{N+1}-1}$, а из последнего уравнения системы получим алгебраическое уравнение (3) для $\varkappa_1$. В общем случае алгебраическое уравнение N степени имеет N корней. Многочисленные численные эксперименты показали, что среди корней только один действительный положительный корень, который и будет значением $\varkappa_1$. ■

Из теоремы 1 следует, что $F_n(w) = R_n e^{j\varkappa_1 w}, n = \overline{0,N}$, тогда асимптотическая характеристическая функция числа заявок на орбите имеет вид

$$H(u) = \sum_{n=0}^{N} F_n(w,\varepsilon) \approx \sum_{n=0}^{N} F_n(w) = \exp\{j\varkappa_1 w\} = \exp\left\{j\frac{\varkappa_1}{\sigma}u\right\}.$$

4. Асимптотика второго порядка

Основной результат анализа асимптотики второго порядка представим в виде теоремы.

Теорема 2. Пусть $i(t)$ – число заявок на орбите в RQ-системы $M/M/1/N-1$ с обратными связями, тогда для последовательности характеристических функций выполняется равенство

$$\lim_{\sigma\to 0} M\left\{e^{jw\sqrt{\sigma}(i(t)-\varkappa_1/\sigma)}\right\} = \exp\left\{(jw)^2\varkappa_2/2\right\},$$

где

$$\varkappa_2 = \frac{\varkappa_1\left(R_N-1\right)-\left(\varkappa_1+\lambda\right)g_N+\mu r_2 g_0}{R_N-1+\left(\varkappa_1+\lambda\right)\varphi_N-\mu r_2\varphi_0}, \tag{5}$$

$\varphi_n = \frac{(n-N)d^{n+N+1}+(N-n+1)d^{n+N}-(n+1)d^n+nd^{n-1}}{(d^{N+1}-1)^2(\mu r_0+\mu r_2)}, n=\overline{0,N}$, а функции g_0 и g_N являются решением системы уравнений

$$\begin{aligned}
&-(\lambda+\varkappa_1)g_0+(\mu r_0+\mu r_2)g_1=-\mu r_2 R_1;\\
&(\lambda+\varkappa_1)\,g_{n-1}-(\lambda+\mu r_0+\mu r_2+\varkappa_1)\,g_n+(\mu r_0+\mu r_2)\,g_{n+1}=\\
&=\varkappa_1 R_{n-1}-\mu r_2 R_{n+1},\quad n=\overline{1,N-1};\\
&(\lambda+\varkappa_1)\,g_{N-1}-(\mu r_0+\mu r_2)\,g_N=\varkappa_1 R_{N-1}-\lambda R_N.
\end{aligned}$$

Доказательство. В системе уравнений (1), (2) выполним замену $H_n(u) = H_n^{(2)}(u)e^{ju\varkappa_1/\sigma}$, $n=\overline{0,N}$. Здесь $H_n^{(2)}(u)$ – характеристическая функция центрированной случайной величины $i(t)-\varkappa_1/\sigma$. Затем обозначим $\sigma=\varepsilon^2$ и введем замену $u=\varepsilon w$, $H_n^{(2)}(u)=F_n^{(2)}(w,\varepsilon)$, $n=\overline{0,N}$. Решение полученной системы запишем в виде разложения $F_n^{(2)}(w,\varepsilon)=\Phi_2(w)\left(R_n+j\varepsilon w f_n\right)+O(\varepsilon^2)$. После преобразований, аналогичных проведенным при доказательстве теоремы 1, можно обозначить $\frac{\Phi'_2(w)}{w\Phi_2(w)}=-\varkappa_2$ и записать $\Phi_2(w)=exp\left\{\frac{(jw)^2}{2}\varkappa_2\right\}$. Тогда после замены $f_n=C\cdot R_n+g_n+\varkappa_2\varphi_n$, $n=\overline{0,N}$ получим (5). ■

Теорема 2 показывает, что асимптотическая функция числа заявок на орбите в RQ-системы $M/M/1/N-1$ с обратными связями является характеристической функцией гауссовской случайной величины с математическим ожиданием $\varkappa_1/\sigma$ и дисперсией $\varkappa_2/\sigma$.

Отметим, что результат теоремы 1 и 2 при $N=1$ и отсутствии буфера для ожидания согласуется с результатом, полученным в [4].

Заключение

В работе проведено исследование системы массового обслуживания $M/M/1/N-1$ с обратными связями и повторными вызовами. Показано, что асимптотическое распределение вероятностей числа заявок на орбите при условии растущего среднего времени ожидания на орбите является гауссовским с параметрами $\varkappa_1/\sigma$ и $\sqrt{\varkappa_2/\sigma}$. Получены уравнения для нахождения параметров распределения. Многочисленные численные эксперименты показали, что уравнения позволяют найти единственное положительное значение параметров распределения.

СПИСОК ЛИТЕРАТУРЫ

1. *Вишневский В. М., Дудин А. Н.* Системы массового обслуживания с коррелированными входными потоками и их применение для моделирования телекоммуникационных сетей //Автоматика и телемеханика. 2017. №8. С. 3–59.
2. *Степанов С. Н.* Теория телетрафика. Концепции, модели, приложения. М.: НТИ «Горячая линия–Телеком», 2015. С. 860.
3. *Melikov A., Divya V., Aliyeva S.* Analyses of feedback queue with positive server setup time and impatient calls //Информационные технологии и математическое моделирование: материалы XIX Международной конференции имени А. Ф. Терпугова, г. Томск: Издательство НТЛ, 2021. С. 77–81.
4. *Nazarov A. A., Rozhkova S. V., Titarenko E. Y.* Asymptotic Analysis of RQ-System with Feedback and Batch Poisson Arrival Under the Condition of Increasing Average Waiting Time in Orbit //Distributed Computer and Communication Networks: Control, Computation, Communications. DCCN 2020. Communications in Computer and Information Science, vol 1337. Springer, Cham., 2020. https://doi.org/10.1007/978-3-030-66242-4_26.

Назаров Анатолий Андреевич — д-р техн. наук, профессор, профессор, кафедра теории вероятностей и математической статистики, Институт прикладной математики и компьютерных наук ТГУ. E-mail: *nazarov.tsu@gmail.com*

Рожкова Светлана Владимировна — д-р физ.-мат. наук, доцент, профессор, отделение математики и информатики ШБИП ТПУ, кафедра теории вероятностей и математической статистики, Институт прикладной математики и компьютерных наук ТГУ. E-mail: *rozhkova@tpu.ru*

Титаренко Екатерина Юрьевна — аспирант, старший преподаватель, отделение информационных технологий ИШИТР ТПУ, физический факультет ТГУ. E-mail: *teu@tpu.ru*

АСИМПТОТИЧЕСКИЙ АНАЛИЗ НЕОДНОРОДНОЙ СМО $MR|M|\infty$, ФУНКЦИОНИРУЮЩЕЙ В ПОЛУМАРКОВСКОЙ СЛУЧАЙНОЙ СРЕДЕ, В УСЛОВИИ ЭКВИВАЛЕНТНОГО РОСТА ВРЕМЕНИ ОБСЛУЖИВАНИЯ НА ПРИБОРАХ

Е. П. Полин[1,2], С. П. Моисеева[1]

[1] *Национальный исследовательский Томский государственный университет, г. Томск, Россия*
[2] *Национальный исследовательский Томский политехнический университет, г. Томск, Россия*

Рассматривается неоднородная система массового обслуживания (СМО) с неограниченным числом обслуживающих приборов, функционирующая в условиях изменяющейся внешней среды. На вход СМО поступает поток марковского восстановления, моменты наступления событий в котором определяются полумарковской матрицей. Каждая заявка обслуживается случайное время, распределенное по экспоненциальному закону с параметром, зависящим от состояния вложенной по моментам наступления событий цепи Маркова. Решается задача исследования двумерного случайного процесса – числа заявок, обслуживаемых с разной интенсивностью в системе методом асимптотического анализа.

Ключевые слова: *Поток марковского восстановления, система массового обслуживания, случайная среда, метод асимптотического анализа.*

Введение

На текущий момент значительная часть информационных, телекоммуникационных и других систем функционирует в условиях изменяющейся внешней среды. Воздействие случайной среды может выражаться, например, в изменении параметров функционирования системы. В связи с этим возникают вопросы устойчивости таких систем к внешним воздействиям. Поэтому исследование систем, функционирующих в случайной среде [1], является актуальной задачей.

В различных работах, посвященных исследованию систем в марковских [2] и полумарковских [3] случайных средах, рассматривались разные варианты реакции системы на смену состояния внешней среды. В данной работе рассматривается случай, предполагающий, что режим обслуживания заявок не меняется до тех пор, пока они не покинут систему.

Одним из способов исследования систем массового обслуживания в случайной среде является применение асимптотических методов [4] при различных условиях. Метод асимптотического анализа реализуется в построении последовательности асимптотик возрастающего порядка, в котором асимптотика первого порядка определяет асимптотическое среднее значение числа занятых приборов. Асимптотика второго порядка позволяет построить гауссовскую аппроксимацию распределения вероятностей числа занятых приборов в системе.

1. Постановка задачи

Рассматривается система массового обслуживания $MR|M|\infty$ с неограниченным числом обслуживающих приборов двух типов, функционирующая в полумарковской случайной среде. На вход поступает поток марковского восстановления, заданный матрицей функций распределения длин интервалов $\mathbf{A}(z) = diag\,[A_i(z)]$, $i = 1,2$ и матрицей вероятностей переходов $\mathbf{P} = [p_{ij}]$, $i,j = 1,2$ — вложенной по моментам наступления событий цепи Маркова с конечным числом состояний $k(t) = 1,2$. Дисциплина обслуживания определяется следующим образом: если вложенная цепь Маркова находится в состоянии $k(t) = i$, то поступающая заявка будет обслуживаться на приборе i-го типа в течение случайного времени, распределенного по экспоненциальному закону $F_i(x) = 1 - e^{-\mu_i x}$.

Ставится задача исследования двумерного случайного процесса $\mathbf{i}(t) = [i_1(t), i_2(t)]$ — числа занятых приборов разного типа в системе в момент времени t. Процесс $\mathbf{i}(t)$ не является марковским. Определим четырехмерный марковский случайный процесс $\{k(t), z(t), i_1(t), i_2(t)\}$, где $z(t)$ — длина интервала от момента времени t до момента наступления очередного события в потоке марковского восстановления, $k(t)$ — вложенная по моментам восстановления цепь Маркова.

2. Метод характеристических функций

Для распределения вероятностей $P(k, z, i_1, i_2, t) = P\{k(t) = k, z(t) < z, i_1(t) = i_1, i_2(t) = i_2\}$ запишем систему дифференциальных уравнений Колмогорова в стационарном режиме:

$$\frac{\partial P(1,z,i_1,i_2)}{\partial z} - \frac{\partial P(1,0,i_1,i_2)}{\partial z} -$$

$$-(i_1\mu_1+i_2\mu_2)P(1,z,i_1,i_2)+\frac{\partial P(1,0,i_1-1,i_2)}{\partial z}p_{11}A_1(z)+$$
$$+\frac{\partial P(2,0,i_1-1,i_2)}{\partial z}p_{21}A_1(z)+P(1,z,i_1+1,i_2)(i_1+1)\mu_1+$$
$$+P(1,z,i_1,i_2+1)(i_2+1)\mu_2=0,$$

$$\frac{\partial P(2,z,i_1,i_2)}{\partial z}-\frac{\partial P(2,0,i_1,i_2)}{\partial z}-$$
$$-(i_1\mu_1+i_2\mu_2)P(2,z,i_1,i_2)+\frac{\partial P(2,0,i_1,i_2-1)}{\partial z}p_{22}A_2(z)+$$
$$+\frac{\partial P(1,0,i_1,i_2-1)}{\partial z}p_{12}A_2(z)+P(2,z,i_1+1,i_2)(i_1+1)\mu_1+$$
$$+P(2,z,i_1,i_2+1)(i_2+1)\mu_2=0.$$

Введем частичные характеристические функции вида

$$H(k,z,u_1,u_2)=\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}e^{ju_1i_1}e^{ju_2i_2}P(k,z,i_1,i_2),\ \text{где } j=\sqrt{-1}.$$

Запишем систему дифференциальных уравнений для частичных характеристических функций в векторно-матричной форме

$$\frac{\partial \mathbf{h}(z,u_1,u_2)}{\partial z}+\frac{\partial \mathbf{h}(0,u_1,u_2)}{\partial z}\left(\mathbf{PA}(z)\mathbf{B}(\mathbf{u})-\mathbf{I}\right)+$$
$$+j\mu_1\left(1-e^{-ju_1}\right)\frac{\partial \mathbf{h}(z,u_1,u_2)}{\partial u_1}+j\mu_2\left(1-e^{-ju_2}\right)\frac{\partial \mathbf{h}(z,u_1,u_2)}{\partial u_2}=0, \tag{1}$$

где $\mathbf{h}(z,u_1,u_2)=[H(1,z,u_1,u_2),H(2,z,u_1,u_2)]$, $\mathbf{B}(\mathbf{u})=\begin{bmatrix}e^{ju_1} & 0\\ 0 & e^{ju_2}\end{bmatrix}$, $\mathbf{I}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$.

Полученная система уравнений (1) является основной для дальнейших исследований. Предлагается провести анализ характеристик рассматриваемой СМО с помощью метода асимптотического анализа.

3. Асимптотика первого порядка

Обозначим $\mu_1 = \varepsilon$, $\mu_2 = q\varepsilon$, $q = const$ (ε – бесконечно малая величина). Решение системы (1) будем находить в асимптотическом условии эквивалентно растущего времени обслуживания, то есть при $\mu_1, \mu_2 \to 0$. В (1) выполним замены

$$\mathbf{h}(z, u_1, u_2) = \mathbf{f}(z, w_1, w_2, \varepsilon), \ u_1 = \varepsilon w_1, \ u_2 = \varepsilon w_2,$$

получим матричное уравнение для $\mathbf{f}(z, w_1, w_2, \varepsilon)$

$$\begin{gathered}\frac{\partial \mathbf{f}(z, w_1, w_2, \varepsilon)}{\partial z} + \frac{\partial \mathbf{f}(0, w_1, w_2, \varepsilon)}{\partial z} \left(\mathbf{PA}(z)\mathbf{B}(\mathbf{w}, \varepsilon) - \mathbf{I}\right) + \\ + j\left(1 - e^{-j\varepsilon w_1}\right) \frac{\partial \mathbf{f}(z, w_1, w_2, \varepsilon)}{\partial w_1} + jq\left(1 - e^{-j\varepsilon w_2}\right) \frac{\partial \mathbf{f}(z, w_1, w_2, \varepsilon)}{\partial w_2} = 0,\end{gathered} \tag{2}$$

Теорема 1. Предельное при $\varepsilon \to 0$ решение уравнения (2) $\mathbf{f}(z, w_1, w_2)$ имеет вид

$$\mathbf{f}(z, w_1, w_2) = \mathbf{r}(z) \exp\left\{ j\lambda \left(r_1 w_1 + \frac{r_2 w_2}{q} \right) \right\}, \tag{3}$$

где $\lambda = \dfrac{1}{\int_0^\infty (1 - \mathbf{rA}(x)\mathbf{e})\, dx}$, $\mathbf{e}$ — единичный вектор-столбец, $\mathbf{r}(z) = [r_1(z), r_2(z)]$ — вектор-строка распределения вероятностей значений вложенной цепи Маркова, $\mathbf{r} = [r_1, r_2]$ — вектор-строка стационарного распределения вероятностей значений вложенной цепи Маркова.

В силу замены и равенства (3) запишем приближенное (асимптотическое) равенство

$$\mathbf{h}(z, u_1, u_2) \approx \mathbf{f}(z, w_1, w_2) = \mathbf{r}(z) \exp\left\{ j\lambda \left(r_1 w_1 + \frac{r_2 w_2}{q} \right) \right\} =$$

$$= \mathbf{r}(z) \exp\left\{ j\lambda \left(\frac{r_1 u_1}{\mu_1} + \frac{r_2 u_2}{\mu_2} \right) \right\}.$$

Определим характеристическую процесса $\{i_1(t), i_2(t)\}$ в стационарном режиме

$$h(u_1, u_2) = \exp\left\{ j\lambda \left(\frac{r_1 u_1}{\mu_1} + \frac{r_2 u_2}{\mu_2} \right) \right\},$$

которую будем называть асимптотикой первого порядка характеристических функций числа занятых приборов в системе.

4. Асимптотика второго порядка

В уравнении (1) выполним замену

$$\mathbf{h}(z, u_1, u_2) = \mathbf{h}_2(z, u_1, u_2) \exp\left\{ j\lambda \left(\frac{r_1 u_1}{\mu_1} + \frac{r_2 u_2}{\mu_2} \right) \right\},$$

получим уравнение для $\mathbf{h}_2(z,u_1,u_2)$

$$\begin{gathered}
\frac{\partial\mathbf{h}_2(z,u_1,u_2)}{\partial z}+\frac{\partial\mathbf{h}_2(0,u_1,u_2)}{\partial z}\left(\mathbf{PA}(z)\mathbf{B}(\mathbf{u})-\mathbf{I}\right)+\\
+j\mu_1\left(1-e^{-ju_1}\right)\frac{\partial\mathbf{h}_2(z,u_1,u_2)}{\partial u_1}+j\mu_2\left(1-e^{-ju_2}\right)\frac{\partial\mathbf{h}_2(z,u_1,u_2)}{\partial u_2}-\\
-\lambda r_1\left(1-e^{-ju_1}\right)\mathbf{h}_2(z,u_1,u_2)-\lambda r_2\left(1-e^{-ju_2}\right)\mathbf{h}_2(z,u_1,u_2)=0.
\end{gathered}\tag{4}$$

Обозначив $\mu_1=\varepsilon^2$, $\mu_2=q\varepsilon^2$, в (1) выполним замены

$$\mathbf{h}_2(z,u_1,u_2)=\mathbf{f}_2(z,w_1,w_2,\varepsilon),\ u_1=\varepsilon w_1,\ u_2=\varepsilon w_2,$$

получим матричное уравнение для $\mathbf{f}_2(z,w_1,w_2,\varepsilon)$

$$\begin{gathered}
\frac{\partial\mathbf{f}_2(z,w_1,w_2,\varepsilon)}{\partial z}+\frac{\partial\mathbf{f}_2(0,w_1,w_2,\varepsilon)}{\partial z}\left(\mathbf{PA}(z)\mathbf{B}(\mathbf{w},\varepsilon)-\mathbf{I}\right)+\\
+j\varepsilon\left(1-e^{-j\varepsilon w_1}\right)\frac{\partial\mathbf{f}_2(z,w_1,w_2,\varepsilon)}{\partial w_1}+j\varepsilon q\left(1-e^{-j\varepsilon w_2}\right)\frac{\partial\mathbf{f}_2(z,w_1,w_2,\varepsilon)}{\partial w_2}-\\
-\lambda r_1\left(1-e^{-j\varepsilon w_1}\right)\mathbf{f}_2(z,w_1,w_2,\varepsilon)-\lambda r_2\left(1-e^{-j\varepsilon w_2}\right)\mathbf{f}_2(z,w_1,w_2,\varepsilon)=0.
\end{gathered}\tag{5}$$

Теорема 2. Предельное при $\varepsilon\to0$ решение уравнения (4) $\mathbf{F}_2(z,w_1,w_2)$ имеет вид

$$\begin{gathered}
\mathbf{f}_2(z,w_1,w_2)=\mathbf{r}(z)\exp\left\{\frac{j^2}{2}\left(\lambda\left(r_1w_1^2+r_2\frac{w_2^2}{q}\right)+\right.\right.\\
\left.\left.+\varkappa\left(r_1^2w_1^2+4r_1r_2\frac{w_1w_2}{q+1}+r_2^2\frac{w_2^2}{q}\right)\right)\right\},
\end{gathered}\tag{6}$$

где $\varkappa=\lambda^2\displaystyle\int_0^\infty\left(\mathbf{rA}(x)-\mathbf{r}(x)\right)\mathbf{e}dx$.

В силу замены и равенства (6) для функции $\mathbf{h}_2(z,u_1,u_2)$ запишем приближённое (асимптотическое) равенство

$$\begin{gathered}
\mathbf{h}_2(z,u_1,u_2)\approx\mathbf{f}_2(z,w_1,w_2)=\\
=\mathbf{r}(z)\exp\left\{\frac{j^2}{2}\left(\lambda\left(r_1\frac{u_1^2}{\mu_1}+r_2\frac{u_2^2}{\mu_2}\right)+\right.\right.\\
\left.\left.+\varkappa\left(r_1^2\frac{u_1^2}{\mu_1}+4r_1r_2\frac{u_1u_2}{\mu_1+\mu_2}+r_2^2\frac{u_2^2}{\mu_2}\right)\right)\right\}.
\end{gathered}$$

Таким образом, характеристическая функция числа занятых приборов в рассматриваемой системе имеет вид

$$h_2(u_1,u_2) =$$
$$= \exp\left\{j\lambda\left(\frac{r_1u_1}{\mu_1}+\frac{r_2u_2}{\mu_2}\right)+\frac{j^2}{2}\left(\lambda\left(r_1\frac{u_1^2}{\mu_1}+r_2\frac{u_2^2}{\mu_2}\right)+\right.\right.$$
$$\left.\left.+\varkappa\left(r_1^2\frac{u_1^2}{\mu_1}+4r_1r_2\frac{u_1u_2}{\mu_1+\mu_2}+r_2^2\frac{u_2^2}{\mu_2}\right)\right)\right\}.$$

Заключение

В данной работе методом асимптотического анализа исследована математическая модель системы $MR|M|\infty$, функционирующей в условии изменяющейся внешней среды. Рассмотрен случай, когда полумарковская случайная среда имеет 2 различных состояния. Доказано, что асимптотическая характеристическая функция числа занятых приборов каждого типа в рассматриваемой системе является гауссовской с вектором математических ожиданий $\mathbf{a} = \left[\lambda\frac{r_1}{\mu_1}, \lambda\frac{r_2}{\mu_2}\right]$ и матрицей ковариации $\mathbf{K} = \begin{bmatrix}\lambda\frac{r_1}{\mu_1}+\varkappa\frac{r_1^2}{\mu_1} & 4\frac{r_1r_2}{\mu_1+\mu_2}\\ 4\frac{r_1r_2}{\mu_1+\mu_2} & \lambda\frac{r_2}{\mu_2}+\varkappa\frac{r_2^2}{\mu_2}\end{bmatrix}$.

СПИСОК ЛИТЕРАТУРЫ

1. *Таташев А. Г.* Система массового обслуживания с переменной интенсивностью входного потока // Автоматика и телемеханика. 1995. № 12. С. 78–84.
2. *Королюк Д. В.* Статистические эксперименты в сбалансированной марковской случайной среде // Кибернетика и системный анализ. 2015. Т. 51. № 5. С. 111–116.
3. *Коротаев И. А., Спивак Л. Р.* Системы массового обслуживания в полумарковской случайной среде // Автомат. и телемех. 1992. № 7. С. 86–92.
4. *Назаров А. А., Моисеева С. П.* Метод асимптотического анализа в теории массового обслуживания. Томск: Издательство НТЛ, 2006. 112 с.

Полин Евгений Павлович — ассистент кафедры ТВиМС. E-mail: *polin_evgeny@mail.ru*

Моисеева Светлана Петровна — профессор кафедры ТВиМС. E-mail: *smoiseeva@mail.ru*

МОДЕЛИ ОБСЛУЖИВАНИЯ ТРАФИКА НА БАЗОВОЙ СТАНЦИИ В ВИДЕ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

Р. Р. Салимзянов, Е. Ю. Лисовская

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

В работе рассматриваются математические модели функционирования одной базовой станции, обслуживающей два типа сетевого трафика. В качестве классической модели предлагается система массового обслуживания с ограниченным числом обслуживающих приборов и двумя входящими потоками заявок, каждая из которых занимает один прибор на время обслуживания. Кроме того, предлагается модификация базовой модели в виде ресурсной системы массового обслуживания, где каждой пользовательской сессии для обслуживания потребуется некоторое (не единичное) количество ресурсов. В работе находится стационарное совместное распределение вероятностей числа сессий обоих типов трафика на обслуживании, с помощью которого проводится численный анализ основных метрик производительности сети. В заключении делается вывод о необходимости дальнейшей модификации модели и предложения новой стратегии доступа.

Ключевые слова: *Базовая станция, имитационная модель, сетевой трафик, теория массового обслуживания.*

Введение

Проблема телефонных сетей впервые была сформулирована и решена А.К. Эрлангом в начале XX века [1]. Он рассчитал количество клиентов, ожидающих соединения с кем-то за пределами населенного пункта из-за занятости линии другими абонентами. Модель Эрланга является простой, но ее используют для анализа и улучшения показателей систем в различных областях, а также на основе этой модели функционируют сложные телефонные сети.

С ростом числа пользователей мобильных устройств [2] все серьезней стоит вопрос об оптимизации и улучшении стратегий доступа трафика к базовой станции. К сожалению, очень часто изменения конфигураций базовых станций для улучшения качества обслуживания не включает в себя теоретического или численного решения для оптимизации выбора новой конфигурации. В связи с этим растет актуальность исследования уже используемых стратегий доступа и разработка новых стратегий на основе систем массового обслуживания для оценки нагрузки системы и распределения ресурсов занятых в ней.

1. Классическая СМО

В качестве первой модели стратегии доступа рассмотрим классическую СМО с N приборами, на вход в систему поступает два простейших потока заявок с параметрами λ_1 и λ_2. Длительности обслуживания заявок являются экспоненциально распределенными случайными величинами с параметрами μ_1 и μ_2 соответственно, для каждого из потоков.

Входящая заявка занимает любой свободный прибор, где обслуживается в течение случайного времени, в случае, когда для вновь прибывшей заявки нет свободного прибора в системе, то заявка теряется, не начав обслуживание.

Пусть $n_1(t)$ – число приборов, занятых заявками первого потока в момент времени t, а $n_2(t)$ – число приборов, занятых заявками второго потока в момент времени t. Случайный процесс $\{n_1(t), n_2(t)\}$ является марковским и определен на множестве состояний:

$$\mathbb{N} = \{(n_1, n_2) : 0 \leqslant n_1 + n_2 \leqslant N\}$$

при этом $n_1 \in \{0, \cdots, N\}$ и $n_2 \in \{0, \cdots, N\}$. Рассматриваемая система относится к классу марковских моделей массового обслуживания и поэтому для ее исследования можно применить теорию цепей Маркова с непрерывным временем и составить систему уравнений равновесия для стационарного распределения вероятностей вида:

$$p(n_1, n_2) = \lim_{t\to\infty} P\{n_1(t) = n_1, n_2(t) = n_2\}.$$

Запишем инфинитезимальные характеристики:

$$q(n_1, n_2)(n_1^{'}, n_2^{'}) = \begin{cases} \lambda_1, & n_1^{'} = n_1 + 1, n_2^{'} = n_2, n_1^{'} + n_2^{'} \leqslant N, \\ \lambda_2, & n_1^{'} = n_1, n_2^{'} = n_2 + 1, n_1^{'} + n_2^{'} \leqslant N, \\ n_1\mu_1, & n_1^{'} = n_1 - 1, n_2^{'} = n_2, n_2 > 0, \\ n_2\mu_2, & n_1^{'} = n_1, n_2^{'} = n_2 - 1, n_1 > 0, \\ **, & n_1^{'} = n_1, n_2^{'} = n_2, \\ 0, & \text{другое}, \end{cases}$$

$$** = -\lambda_1 \cdot I(n_1 + n_2 < N) - \lambda_2 \cdot I(n_1 + n_2 < N) - n_1\mu_1 - n_2\mu_2.$$

Запишем СЛАУ:

$$p(n_1,n_2)[\lambda_1 \cdot I(n_1+n_2<N)+\lambda_2 \cdot I(n_1+n_2<N)+n_1\mu_1+n_2\mu_2]=$$
$$=p(n_1+1,n_2)(n_1+1)\mu_1 \cdot I(n_1+n_2<N)+$$
$$+p(n_1,n_2+1)(n_2+1)\mu_2 \cdot I(n_1+n_2<N)+$$
$$+p(n_1-1,n_2)\lambda_1 I(n_1>0)+p(n_1,n_2-1)\lambda_2 I(n_2>0),$$

решение которой имеет вид [3]:

$$p(n_1,n_2)=\left(\sum_{i=0}^{N}\sum_{j=0}^{N-i}\frac{\rho_1^{i_1}}{i!}\frac{\rho_2^{i_2}}{j!}\right)^{-1}\cdot\frac{\rho_1^{n_1}}{n_1!}\frac{\rho_2^{n_2}}{n_2!},$$

где $\rho_1=\dfrac{\lambda_1}{\mu_1}$ и $\rho_2=\dfrac{\lambda_2}{\mu_2}$.

2. Ресурсная СМО

В качестве второй модели стратегии доступа рассмотрим СМО, на вход которой аналогично поступает два простейших потока заявок с параметрами λ_1 и λ_2, а длительности обслуживания заявок являются экспоненциально распределенными случайными величинами с параметрами μ_1 и μ_2, соответственно. Кроме этого, каждая входящая заявка требует на время обслуживания x_1 и x_2 единиц ресурса соответственно. Входящая заявка занимает свободный прибор при достаточном количестве свободного ресурса из общего ресурса R, в случае, когда для вновь прибывшей заявки нет достаточного количества ресурса – заявка покидает систему, не начав обслуживание.

Рассмотрим двумерный марковский процесс с непрерывным временем $\{n_1(t),n_2(t)\}$. Пусть $n_1(t)$ – количество заявок первого потока на обслуживании в момент времени t, $n_2(t)$ – количество заявок второго потока на обслуживании в момент времени t.

Пространство состояний определяется следующим образом:

$$\mathbb{N}=\{(n_1,n_2): n_1x_1+n_2x_2\leqslant R\}.$$

Инфинитезимальные характеристики:

$$q(n_1,n_2)(n_1^{'},n_2^{'})=\begin{cases}\lambda_1, & n_1^{'}=n_1+1, n_2^{'}=n_2, n_1^{'}x_1+n_2x_2\leqslant R,\\ \lambda_2, & n_1^{'}=n_1, n_2^{'}=n_2+1, n_1x_1+n_2^{'}x_2\leqslant R,\\ n_1\mu_1, & n_1^{'}=n_1-1, n_2^{'}=n_2, n_1>0,\\ n_2\mu_2, & n_1^{'}=n_1, n_2^{'}=n_2-1, n_2>0,\\ **, & n_1^{'}=n_1, n_2^{'}=n_2,\\ 0, & \text{другое,}\end{cases}$$

$$** = -\lambda_1 \cdot I((n_1+1)x_1 + n_2x_2 \leqslant R) -$$
$$-\lambda_2 \cdot I(n_1x_1 + (n_2+1)x_2 \leqslant R) - n_1\mu_1 - n_2\mu_2.$$

Так как процесс марковский, апериодический и неразложимый, подразумевая, что существуют финальные вероятности состояний процесса, совпадающие с вероятностями стационарного распределения [4], можем использовать принцип локального равновесия для записи СЛАУ стационарного распределения вероятностей:

$$p(n_1, n_2) = \lim_{t\to\infty} P\{n_1(t) = n_1, n_2(t) = n_2\},$$

получим

$$p(n_1, n_2)[\lambda_1 I((n_1+1)x_1 + n_2x_2 \leqslant R) +$$
$$+\lambda_2 I(n_1x_1 + (n_2+1)x_2 \leqslant R) + n_1\mu_1 + n_2\mu_2] =$$
$$= p(n_1+1, n_2)(n_1+1)\mu_1 I((n_1+1)x_1 + n_2x_2 \leqslant R) +$$
$$+p(n_1, n_2+1)(n_2+1)\mu_2 I(n_1x_1 + (n_2+1)x_2 \leqslant R +$$
$$+p(n_1-1, n_2)\lambda_1 I(n_1 > 0) + p(n_1, n_2-1)\lambda_2 I(n_2 > 0).$$

Решение системы имеет вид:

$$p(n_1, n_2) = \left(\sum_{i=0}^{N_1} \sum_{j=0}^{N_2} \frac{\rho_1^i}{i!} \frac{\rho_2^j}{j!} \cdot I(ix_1 + jx_2 \leqslant R) \right)^{-1} \times$$
$$\times \frac{\rho_1^{n_1}}{n_1!} \frac{\rho_1^{n_2}}{n_2!} \cdot I(n_1x_1 + n_2x_2 \leqslant R),$$

где $N_1 = \left\lfloor \dfrac{R}{x_1} \right\rfloor$ и $N_2 = \left\lfloor \dfrac{R}{x_2} \right\rfloor$, $\rho_1 = \dfrac{\lambda_1}{\mu_1}$ и $\rho_2 = \dfrac{\lambda_2}{\mu_2}$.

3. Численный эксперимент

В ходе работы были проведены численные эксперименты для сравнения работы двух систем в одинаковых условиях. Для удобства проведения эксперимента, можно представить классическую СМО в виде ресурсной, в которой каждая заявка требует по 1 единице ресурса ($x_1 = x_2 = 1$). Рассмотрим один из экспериментов. В таблице 1 приведены значения параметров системы. Параметры для проведения эксперимента были выбраны исходя из технических характеристик сетей пятого и четвертого поколения.

На рисунках 1–2 представлены графики зависимости вероятности потери заявок от интенсивности поступления заявок 4G трафика (λ_1).

Таблица 1

Параметры СМО

Параметр	Значение
Емкость базовой станции (R)	250
Количество требуемого ресурса для трафика 4G (x_1)	1
Количество требуемого ресурса для трафика 5G (x_2)	2
Интенсивность входящего потока для трафика 5G (λ_2)	198,315 1/сек
Интенсивность обслуживания сессии 4G (μ_1)	0,033 1/сек
Интенсивность обслуживания сессии 5G (μ_2)	5,261 1/сек

С увеличением интенсивности поступления заявок 4G трафика (λ_1), среднее количество заявок на обслуживании 4G трафика растет, а среднее количество заявок на обслуживании 5G трафика падает (Рисунки 3–4), аналогично изменяется среднее количество занятого ресурса.

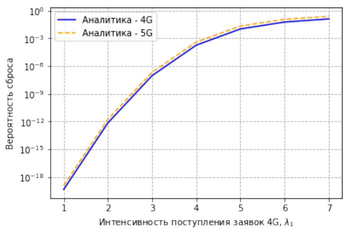

Рис. 1. Вероятность потери – ресурсная СМО

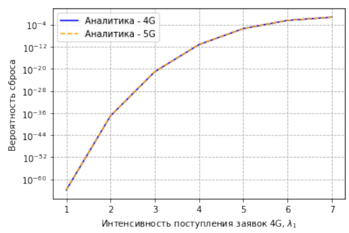

Рис. 2. Вероятность потери – классическая СМО

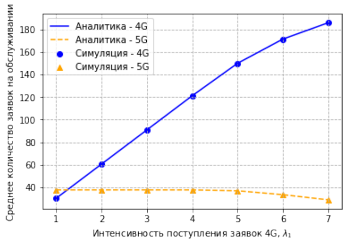

Рис. 3. Среднее количество заявок на обслуживании – ресурсная СМО

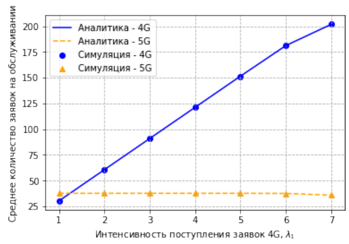

Рис. 4. Среднее количество заявок на обслуживании – классическая СМО

Заключение

В результате проделанной работы были исследованы две СМО, была построена имитационная модель, которая обладает достаточной точностью для дальнейшего исследования, проведены численные эксперименты по разным возможным сценариям работы базовых станций.

После проведения численных экспериментов можно сделать вывод о том, что для полноценной разработки и сравнения стратегий доступа двух типов трафика к базовой станции необходимо рассмотреть модификацию системы массового обслуживания в виде ресурсной RQ-системы, где при нехватке свободного ресурса для заявки, она не покидает систему, а уходит на орбиту на случайное время, после которого совершает повторную попытку встать на обслуживание. После нахождения вероятностных характеристик для такой системы и проведения экспериментов с помощью имитационной модели можно будет сделать вывод о наиболее эффективной стратегии предоставления доступа.

СПИСОК ЛИТЕРАТУРЫ

1. *Erlang A. K.* Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges // American Journal of Operations Research. 1917.
2. `https://rosstat.gov.ru/folder/70843/document/100659` — Федеральная служба государственной статистики. 2020.
3. *Назаров А. А.,Терпугов А. Ф.* Теория массового обслуживания. Томск, 2010.
4. *Моисеева С. П.* Теория случайных процессов. Часть 2. Марковские процессы. Томск, 2014.

Салимзянов Радмир Ренатович — студент института прикладной математики и компьютерных наук НИ ТГУ. E-mail: *rsalimzyanov@yahoo.com*

Лисовская Екатерина Юрьевна — канд. физ.-мат. наук, доцент кафедры теории вероятности и математической статистики НИ ТГУ. E-mail: *ekaterina_lisovs@mail.ru*

МОДЕЛИРОВАНИЕ И АНАЛИЗ СЕРВЕРНЫХ ФЕРМ С РАЗОГРЕВОМ

К. Д. Сизов, А. В. Зорин

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

Развитие сети Интернет привело к появлению большого числа крупных серверных систем, которые потребляют много электроэнергии. При определенных предположениях используя классические методы теории массового обслуживания можно построить и проанализировать математическую модель серверной фермы, чтобы понять удастся ли улучшить энергоэффективность данной системы, если отключать обслуживающие устройства, не занятые обработкой требований. В данной работе в рамках теории массового обслуживания удалось получить численный метод нахождения стационарных характеристик системы, в которой время до поступления требования и время его обслуживания распределены по экспоненциальному закону, а время включения сервера по закону Эрланга. Для наглядности было проведено сравнение поведения системы с «отключениями» и без них.

Ключевые слова: *Серверные фермы, теория массового обслуживания, марковский процесс, обобщенные процессы рождения и гибели с зависимыми уровнями, M/M/c/ системы, стационарное распределение.*

Введение

В рамках классической теории массового обслуживания уже хорошо изучены системы класса $M|M|c|\infty$. Такие системы обслуживания имеют пуассоновский входной поток, экспоненциальное время обслуживания требования и очередь с бесконечной емкостью и терпеливыми заявками. Стоит отметить, что в рамках этой модели предполагается, что все обслуживающие устройства готовы моментально принять на обработку пришедшие требование. Однако на практике встречаются системы, в которых, незанятые обслуживанием, устройства отключаются, и для обратного приведения этих устройств во включенный режим требуется некоторое время.

Одним из примеров таких систем являются серверные фермы с разогревом. Поскольку Интернет-сервера потребляют большое количество электроэнергии. Отключение устройств, которые в данный момент не заняты обработкой запроса, является одним из очевидных способов снизить энергопотребление таких систем.

В статье [4] приводится построение математической модели серверной фермы с разогревом, в которой случайные величины: время до прибытия следующего запроса, время его обслуживания и время включения сервера распределены по экспоненциальному закону. Автору статьи удалось получить стационарное распределение такой системы. При этом в решении матрично-геометрическим методом в силу структуры, рассматриваемой системы, удалось получить явные выражения для элементов вспомогательных матриц метода.

Целью данной работы является анализ серверной фермы с разогревом, в которой время включения (разогрева) сервера есть случайная величина распределенная по закону Эрланга с известными параметрами. Под анализом подразумевается построение математической модели, вычисление основных характеристик качества и сравнение полученных характеристик с характеристиками для системы без выключений.

1. Математическая модель

Ниже будет проводится построение и анализ математической модели системы массового обслуживания со следующими характеристиками:

— *Поток входящих требований* — пуассоновский с параметром λ.
— *Время обслуживания требования* — случайная величина, имеющая экспоненциальное распределение с параметром μ.
— *Дисциплина формирования очереди* — в порядке поступления требований.
— *Структура системы* — многолинейная система с ожиданием.
— *Число обслуживающих устройств (серверов)* — c.
— Сервер после обслуживания требования, в случае отсутствия других требований в очереди, ожидающих обслуживания, незамедлительно отключается.
— При поступлении нового требования в систему, включается одно из выключенных устройств, при этом время включения устройства — случайная величина имеющая распределение Эрланга с параметром масштаба α и параметром формы r.
— Если устройство завершило обслуживание, и в очереди имеется не более c требований, то только что завершившее обслуживание, устройство берёт на обслуживание требование из очереди, при этом если имеются сервера находящиеся на стадии включения, то сервер, который к данному моменту потратил на включение меньше всего

времени, отключается. В случае если требований в очереди больше c, отключение включающегося устройства не происходит.

Время включения сервера распределено по закону Эрланга $E(r,\alpha)$, поэтому будем рассматривать последовательно проходимые фазы включения, в соответствии с «методом Эрланга». Состояние системы опишем в виде $\boldsymbol{\nu}(t) = (i, j, y_1, \ldots, y_r)$, где

$$i \in \{0, 1, \ldots, c\};\ j \in \{i+d, i+d+1, \ldots\};\ \sum_{k=1}^{r} y_k \in \{0, 1, \ldots, c-i\}, \quad (1)$$

Компоненты вектора состояния имеют следующий физический смысл: i — число занятых устройств, j — число требований в системе, y_k — число устройств на k-ой фазе включения $k = 1,\ 2,\ \ldots,\ r$. Множество всевозможных состояний удовлетворяющих условиям (1) обозначим через S.

Будем рассматривать случайный процесс с непрерывным временем:

$$\{\boldsymbol{\nu}(t) \in S;\ t \geqslant 0\}.$$

Интенсивности переходов для данного процесса удобно записать в виде матрицы Q интенсивностей переходов, которая имеет блочный вид:

$$\begin{pmatrix}
N_0 & \Lambda_0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\
M_1 & N_1 & \Lambda_1 & 0 & \ldots & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\
0 & M_2 & N_2 & \Lambda_2 & \ldots & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ldots \\
0 & 0 & 0 & 0 & \ldots & M_{c-1} & N_{c-1} & \Lambda_{c-1} & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & \ldots & 0 & M_c & N & \Lambda & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & \ldots & 0 & 0 & M & N & \Lambda & 0 & \ldots \\
0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & M & N & \Lambda & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}. \quad (2)$$

Блокам N_k, Λ_k, M_k соответствуют подмножества состояний:

$$S_k = \{(i, j, y_1, \ldots, y_r) \in S : j = k\} \subset S.$$

Далее это подмножество будем назвать k-ым слоем системы. Внутри каждого слоя строки и столбцы матрицы Q упорядочиваются в лексикографическом порядке со словом вида $(y_1, y_2, \ldots, y_r, i)$. Именно благодаря такому упорядочиванию состояний системы, матрица Q имеет блочно-трехдиагональный вид. Для блоков матрицы Q справедливы следующие утверждения:

1) $N_k \in \mathbb{R}^{n_k \times n_k}$, $\Lambda_k \in \mathbb{R}^{n_k \times n_{k+1}}$, $M_k \in \mathbb{R}^{n_k \times n_{k-1}}$, где $n_k = |S_k|$ и выполняются неравенства: $n_0 < n_1 < \dots < n_c = n_{c+1} = n_{c+2} = \dots$;
2) $\Lambda_k = \begin{pmatrix} \lambda E & 0 \end{pmatrix}$, где E — единичная матрица размерности n_k;
3) матрицы N_k — квадратные верхнетреугольные со строго отрицательными элементами на главной диагонали.

Запись всех блоков матрицы Q в общем виде имеет громоздкий вид и была опущена для краткости.

2. Стационарное распределение

Обозначим через $\boldsymbol{\pi}_k$ вектор стационарных вероятностей для состояний k-ого слоя. Искомый же вектор стационарных вероятностей всех состояний системы будет иметь вид:

$$\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \dots \boldsymbol{\pi}_c, \dots)$$

Используя тот факт, что блоки матрицы Q перестают изменяться от слоя к слою и опираясь на результаты изложенные в [1], [3], можно сформулировать теорему:

Теорема 1. Для процесса $\{\boldsymbol{\nu}(t) \in S;\ t \geqslant 0\}$, имеющего матрицу интенсивностей (2), стационарные вероятности для состояний k-ого слоя должны удовлетворять следующим условиям:

1) выполняется рекуррентное соотношение:

$$\boldsymbol{\pi}_{c+i} = \boldsymbol{\pi}_{c+i-1} R, (i \in \mathbb{N}), \tag{3}$$

где матрица R — минимальный неотрицательный корень квадратного матричного уравнения:

$$\Lambda + RN + R^2 M = 0; \tag{4}$$

2) справедлива система уравнений:

$$\begin{cases} \boldsymbol{\pi}_0 \cdot \mathbf{1}^T + \boldsymbol{\pi}_1 \cdot \mathbf{1}^T + \boldsymbol{\pi}_c \cdot (E + R + R^2 + R^3 + \dots) & = 1 \\ \boldsymbol{\pi}_0 N_0 + \boldsymbol{\pi}_1 M_1 & = \mathbf{0} \\ \boldsymbol{\pi}_0 \Lambda_0 + \boldsymbol{\pi}_1 N_1 + \boldsymbol{\pi}_2 M_2 & = \mathbf{0} \\ \boldsymbol{\pi}_1 \Lambda_1 + \boldsymbol{\pi}_2 N_2 + \boldsymbol{\pi}_3 M_3 & = \mathbf{0} \\ \dots \\ \boldsymbol{\pi}_{c-2} \Lambda_{c-2} + \boldsymbol{\pi}_{c-1} N_{c-1} + \boldsymbol{\pi}_c M_c & = \mathbf{0} \\ \boldsymbol{\pi}_{c-1} \Lambda_{c-1} + \boldsymbol{\pi}_c N + \boldsymbol{\pi}_c RM & = \mathbf{0} \end{cases}, \tag{5}$$

где E — единичная матрица размерности равной размерности R.

В случае, когда ряд $\sum\limits_{i=0}^{\infty} R^i$ расходится стационарного распределения не существует.

Решение данного квадратного уравнения (4) можно найти [2] как предел последовательности, заданной рекуррентным соотношением:

$$R_0 = 0;\ R^{(k)} = \frac{1}{\omega}[\Lambda + R^{(k-1)}(N + \omega E) + R^{(k-1)}R^{(k-1)}M]\ k \geqslant 1,$$

где $\omega = |\min\{\mathrm{diag}\,(N)\}|$.

Систему (5), в случае сходимости матричного ряда в первом уравнении, можно решить известными численными методами для конечных систем линейных алгебраических уравнений. В частности, используя сингулярное разложение, можно найти решение однородной подсистемы и потом учесть условие нормировки. Те компоненты ветора $\boldsymbol{\pi}$, которые не входят в систему (5), могут быть получены при помощи соотношения (3).

Для вычисление среднего энергопотребления системы в стационарном режиме будем применять формулу:

$$C = C_\alpha \mathbb{E}Y + C_\mu \mathbb{E}X,$$

здесь $\mathbb{E}X, \mathbb{E}Y$ — среднее число активных и разогревающихся серверов в стационарном режиме соответственно, C_μ, C_α — энергия потребляемая в единицу времени одним сервером, находящимся в активном состоянии и в состоянии разогрева соответственно. Величины $\mathbb{E}X, \mathbb{E}Y$ вычисляются на основе получаемого стационарного распределения системы.

3. Результаты численных экспериментов

На основе изложенного метода получения стационарного распределения была написана его численная реализация, которая позволяет вычислять стационарные характеристики системы при любых допустимых входных параметрах. Входными параметрами в данном случае являются параметры системы $c, \lambda, \mu, \alpha, r, C_\alpha, C_\mu$. Получение блоков матрицы Q также проводилось численно: поскольку каждому элементу матрицы соответствует пара состояний системы, то вычисляя для каждой такой пары интенсивность перехода, мы получаем значения всех элементов каждого блока матрицы Q. Примеры результатов численных экспериментов при фиксированных параметрах: $r = 2, \mu = 1.0, C_\alpha = 1.0, C_\mu = 1.0$, приведены ниже.

Как видно из графиков, не при всех параметрах системы удаётся добиться экономии электроэнергии при отключении не занятых серверов.

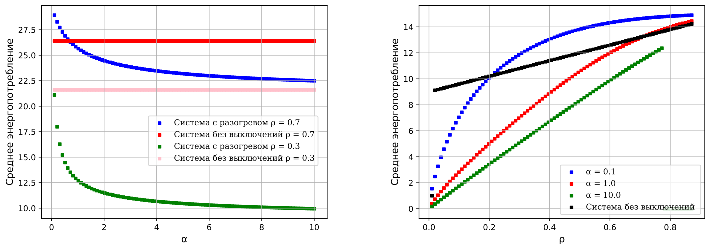

Рис. 1. Зависимость потребляемой энергии от параметров системы

Заключение

В данной работе впервые была проанализирована система массового обслуживания, в которой обслуживающие устройства отключаются, а время их включения распределено по закону Эрланга. Для сконструированной математической модели был адаптирован матрично-геометрический метод. С помощью численной реализации данного метода удалось найти стационарные характеристики системы, а также сравнить их с характеристиками классической модели.

СПИСОК ЛИТЕРАТУРЫ

1. *Neuts M. F.* Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach, Baltimore: Johns Hopkins University Press, 1981. 529 с.
2. *Бочаров П. П.,Печинкин А. В.* Теория массового обслуживания. Москва: РУДН, 1995. 332 с.
3. *Rupert G. Miller* Stationarity Equations in Continuous Time Markov Chains // Transactions of the American Mathematical Society. 1963. Т. 109. № 1. С. 35–44.
4. *Tuan Phung-Duc* Exact solutions for M/M/c/Setup queues // Telecommun Syst. 2017. Т. 64. № 1. С. 309–324.

Сизов Кирилл Дмитриевич — студент магистратуры, ИИТММ, ННГУ им. Н.И. Лобачевского. E-mail: *sizow.k.d@gmail.com*

Зорин Андрей Владимирович — д.ф.-м.н., доцент, профессор кафедры Теории вероятности и анализа данных, ИИТММ, ННГУ им. Н.И. Лобачевского. E-mail: *andrei.zorine@itmm.unn.ru*

МЕТОД АНАЛИЗА ОТКРЫТОЙ СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ С ГРУППОВЫМ ОБСЛУЖИВАНИЕМ

Е. П. Станкевич[1], И. Е. Тананко[1], М. Пагано[2]

[1] *Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, г. Саратов, Россия*
[2] *Пизанский университет, г. Пиза, Италия*

В данной работе рассматривается открытая сеть массового обслуживания с групповым обслуживанием требований. В сеть из источника поступает пуассоновский поток требований. Каждая система сети содержит один обслуживающий прибор и очередь бесконечной длины. Требования обслуживаются группой заданного размера. Если в очереди количество требований меньше заданного числа, то обслуживающий прибор простаивает, иначе выбирает на обслуживание заданное число требований, которые в дальнейшем обслуживаются группой с экспоненциально распределенной длительностью обслуживания. Для получения стационарного распределения открытой сети обслуживания с групповым обслуживанием используется эквивалентная сеть массового обслуживания с обслуживанием требований по одному.

Ключевые слова: *Сети массового обслуживания, групповое обслуживание, стационарное распределение.*

Введение

Простота и удобство использования сетей массового обслуживания при решении задач, возникающих в сетях связи, компьютерных и телекоммуникационных системах, способствовали интенсивному развитию методов анализа сетей массового обслуживания. В последнее время одним из перспективных направлений в теории сетей массового обслуживания является развитие методов анализа сетей массового обслуживания с групповым обслуживанием и групповыми переходами требований [1, 2, 3, 4, 5, 6]. Сети данного класса являются удобным инструментом для решения задач проектирования и анализа многопользовательских систем [4, 5], производственных [2] и транспортных систем [6], а также для исследования эффективности алгоритмов многопутевой

Работа выполнена при поддержке Минобрнауки России в рамках выполнения государственного задания (проект № FSRR-2020-0006)

маршрутизации. В данной работе показано, что стационарное распределение открытых сетей массового обслуживания с групповым обслуживанием требований можно получить, используя известные результаты для сетей Джексона.

1. Описание модели

Рассматривается открытая сеть массового обслуживания N с непрерывным временем, состоящая из L систем массового обслуживания S_i, $i \in I$, $I = \{1, \ldots, L\}$. В сеть из источника S_0 поступает пуассоновский поток требований одного класса с интенсивностью λ_0. Требования между системами сети и источником переходят в соответствии с маршрутной матрицей $\Theta = (\theta_{ij})$, $i, j = 0, \ldots, L$. Состояние сети определяется вектором $s = (s_i)$, $i = 1, \ldots, L$, где s_i– число требований, находящихся в системе S_i, $i = 1, \ldots, L$. Множество всех состояний сети N обозначим через X.

Система S_i, $i = 1, \ldots, L$, состоит из очереди бесконечной длины и одного обслуживающего прибора, одновременно обслуживающего группу, состоящую из b_i, $b_i \geqslant 1$, требований. Длительность обслуживания группы требований в системе S_i является экспоненциально распределенной случайной величиной с параметром μ_i. Если в очереди находится меньше, чем b_i требований, то обслуживающий прибор будет простаивать до тех пор, пока в очередь не поступит, по крайней мере, b_i требований. После завершения обслуживания в системе S_i, каждое требование независимо от остальных требований группы переходит в систему S_j, $j \in I$, с вероятностью θ_{ij}. Полагаем, что в каждой системе S_i сети N размер обслуживаемой группы требований b_i значительно меньше числа систем S_j, в которые могут перейти требования, после завершения обслуживания в системе S_i, а также, что вероятности переходов требований в смежные системы сравнимы.

2. Метод анализа

Из предположения большой размерности сети N и индивидуальной маршрутизации требований после обслуживания группы в системе S_i, $i = 1, \ldots, L$, следует, что вероятность одновременного поступления двух и более требований в любую из систем обслуживания настолько мала, что этой вероятностью можно пренебречь. Поэтому будем считать, что в каждую систему обслуживания сети N поступает пуассоновский поток требований с некоторой интенсивностью. Исследование сети N начнем с изучения изолированной от остальных систем обслуживания системы S_i, $i = 1, \ldots, L$. Известно, что система уравнений равновесия для этой системы обслуживания имеет вид

$$\begin{cases} \lambda_i \pi_i(0) = \mu_i \pi_i(b_i), \\ \lambda_i \pi_i(n) = \lambda_i \pi_i(n-1) + \mu_i \pi_i(b_i+n), \ 1 \leqslant n \leqslant b_i - 1, \\ (\lambda_i + \mu_i)\pi_i(n) = \lambda_i \pi_i(n-1) + \mu_i \pi_i(b_i+n), \ n \geqslant b_i. \end{cases} \tag{1}$$

где λ_i – интенсивность потока требований в систему S_i, $i = 1, \dots, L$.

Построим процесс ξ_i размножения и гибели, который является эквивалентным по стационарному распределению вероятностей состояний системы S_i. Пусть процесс ξ_i определен на множестве состояний $\{0, 1, \dots\}$, $\lambda_i = \lambda_i(n)$ – интенсивность перехода процесса ξ_i из состояния n в состояние $n+1$ не зависит от состояния n, $n \in \{0, 1, \dots\}$, $\widetilde{\mu}_i(n)$ – интенсивность перехода процесса ξ_i из состояния n в состояние $n-1$, где $n \in \{1, 2, \dots\}$. Состояния $\{0, 1, \dots\}$ и параметр λ_i процесса ξ_i соответствуют состояниям $\{0, 1, \dots\}$ и параметру λ_i системы S_i. Определим интенсивности $\widetilde{\mu}_i(n)$, $n = 1, 2, \dots$. Стационарное распределение вероятностей состояний процесса ξ_i размножения и гибели

$$\pi_i(k) = \pi_i(0) \prod_{n=1}^{k} \frac{\lambda_i}{\widetilde{\mu}_i(n)}, \ k = 1, 2, \dots, \tag{2}$$

где

$$\pi_i(0) = \left(1 + \sum_{k=1}^{\infty} \prod_{n=1}^{k} \frac{\lambda_i}{\widetilde{\mu}_i(n)}\right)^{-1}, \ i = 1, \dots, L.$$

Подставим (2) в (1) и получим выражения для вычисления $\widetilde{\mu}_i(n)$, $n = 1, 2, \dots$,

$$\begin{cases} \widetilde{\mu}_i(n) = \lambda_i - \mu_i \dfrac{\lambda_i^{b_i}}{\widetilde{\mu}_i(n+1) \cdot \ldots \cdot \widetilde{\mu}_i(b_i+n)}, \ 1 \leqslant n \leqslant b_i - 1, \\ \widetilde{\mu}_i(n) = \lambda_i + \mu_i - \mu_i \dfrac{\lambda_i^{b_i}}{\widetilde{\mu}_i(n+1) \cdot \ldots \cdot \widetilde{\mu}_i(b_i+n)}, \ n \geqslant b_i. \end{cases} \tag{3}$$

Обозначим $M_i = \lim\limits_{n \to \infty} \widetilde{\mu}_i(n)$. Тогда уравнения (4) при $n \to \infty$ примут вид

$$M_i^{b_i+1} - (\lambda_i + \mu_i) M_i^{b_i} + \lambda_i^{b_i} \mu_i = 0. \tag{4}$$

В уравнении (2) интересующий нас единственный вещественный корень расположен в интервале $\left(\frac{b_i(\lambda_i+\mu_i)}{b_i+1}, \frac{(\lambda_i+\mu_i)^{b_i+1} - \lambda_i^{b_i}\mu_i}{(\lambda_i+\mu_i)^{b_i}}\right)$ и может быть определен численно [7]. Из системы уравнений (4) следует, что

$$\widetilde{\mu}_i(b_i) = \widetilde{\mu}_i(b_i+1) = \widetilde{\mu}_i(b_i+2) = \dots = M_i.$$

Используя уравнения (4), определяются последовательно интенсивности обслуживания $\widetilde{\mu}_i(b_i-1)$, $\widetilde{\mu}_i(b_i-2)$,..., $\widetilde{\mu}_i(1)$. Таким образом, для каждого состояния n процесса ξ_i определены интенсивности $\widetilde{\mu}_i(n)$.

Воспользуемся результатами, полученными для процесса ξ_i, и построим открытую экспоненциальную сеть обслуживания $\widetilde{N}$, состоящую из систем обслуживания $\widetilde{S}_i$ с интенсивностями обслуживания $\widetilde{\mu}_i(n)$, где n – число требований в системе $\widetilde{S}_i$, $n=1,2,\ldots$, $i=1,\ldots,L$. Остальные параметры сети $\widetilde{N}$ совпадают с параметрами сети N.

Построенная сеть обслуживания $\widetilde{N}$ эквивалентна по стационарному распределению сети массового обслуживания N с групповым обслуживанием требований и представляет собой сеть Джексона.

Интенсивности потоков требований в системы S_i определяются из соотношений
$$\lambda_i=\frac{\omega_i}{\omega_0}\lambda_0,\ i=1,\ldots,L,$$
где вектор относительных интенсивностей потоков $\omega=(\omega_i)$, $i=0,\ldots,L$, является решением уравнения $\omega\Theta=\omega$ с условием нормировки $\sum_{i=0}^{L}\omega_i=1$.

Стационарный режим сети N и эквивалентной ей сети $\widetilde{N}$ существует, если для каждой системы S_i выполнятся условие $\lambda_i<b_i\mu_i$, $i=1,\ldots,L$. Стационарные вероятности $\pi(s)$ состояний сети $\widetilde{N}$, а значит и сети N, определяются из выражения
$$\pi(s)=\prod_{i=1}^{L}\pi_i(s_i),\ s\in X,$$
где
$$\pi_i(s_i)=\pi_i(0)\prod_{n=1}^{s_i}\frac{\lambda_i}{\widetilde{\mu}_i(n)}.$$

Математическое ожидание (м. о.) числа требований в системе S_i, $i=1,\ldots,L$,
$$\bar{s}_i=\sum_{n=1}^{\infty}n\pi_i(n),$$
м. о. длительности пребывания требований в сети
$$\bar{\tau}=\frac{1}{\lambda_0}\sum_{i=1}^{L}\bar{s}_i.$$

3. Пример

Рассмотрим сеть обслуживания N с параметрами: $L = 14$, $\lambda_0 = 1$, $b = (3,\ 2,\ 2,\ 3,\ 2,\ 2,\ 3,\ 3,\ 2,\ 2,\ 3,\ 2,\ 2,\ 3)$, $\mu = (0.8,\ 0.6,\ 0.9,\ 0.6,\ 0.8,\ 0.8,\ 0.9,\ 0.6,\ 0.7,\ 0.8,\ 0.9,\ 1.0,\ 0.7,\ 0.7)$,

$$\Theta = \begin{pmatrix}
0.0 & 0.3 & 0.4 & 0.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.2 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.2 & 0.1 & 0.2 & 0.1 & 0.1 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.2 & 0.2 & 0.1 \\
0.2 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.2 \\
0.2 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.2 & 0.1 & 0.1 & 0.1 & 0.1 \\
0.2 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 \\
0.2 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.1 \\
0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.2 & 0.1 & 0.1 \\
0.3 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.0 \\
0.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 \\
0.3 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 & 0.0 \\
0.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.0 & 0.1 & 0.1 & 0.0 & 0.1 \\
0.4 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0
\end{pmatrix}.$$

Данная сеть удовлетворяет сделанным выше предположениям, а именно, сеть состоит из большого количества систем, размер группы требований в каждой системе значительно меньше числа смежных выходных систем. С использованием предложенного метода анализа были вычислены вектор м. о. числа требований в системах сети $\bar{s} = (1.67, 1.98, 1.11, 1.37, 0.88, 0.96, 1.37, 1.69, 1.32, 1.15, 1.62, 1.07, 1.20, 1.48)$, вектор интенсивностей потоков в системы $\lambda = (0.50, 0.57, 0.47, 0.22, 0.28, 0.33, 0.33, 0.39, 0.46, 0.44, 0.52, 0.50, 0.41, 0.32)$, м.о. длительности пребывания требований в сети $\bar{\tau} = 18.87$. С помощью имитационного моделирования систем сети обслуживания получены соответствующие стационарные характеристики с доверительным интервалом 0.001 и доверительной вероятностью не ниже 0.95 : $\bar{s}^I = (1.69, 1.97, 1.11, 1.41, 0.90, 0.98, 1.40, 1.72, 1.35, 1.17, 1.65, 1.10, 1.23, 1.50)$, $\lambda^I = (0.50, 0.57, 0.47, 0.22, 0.28, 0.33, 0.33, 0.39, 0.46, 0.44, 0.52, 0.50, 0.41, 0.32)$ и $\bar{\tau}^I = 19.19$. Сравнивая результаты аналитического и имитационного моделирования видно, что интенсивности входящих потоков в системы совпадают, значения м. о. числа требований с системах сети различается максимум на 2,9% в системе S_4, а значения м.о. длительности пребывания требований в сети различаются на 1,7%.

Заключение

В работе показано, что для открытой сети массового обслуживания с групповым обслуживанием требований большой размерности, в системах которой размеры обслуживаемых групп требований значительно меньше числа смежных выходных систем, стационарное распределение может быть вычислено в мультипликативной форме.

СПИСОК ЛИТЕРАТУРЫ

1. *Klünder W.* Decomposition of Open Queueing Networks with Batch Service // Operations Research Proceedings. 2016. P. 575–581.
2. *Rabta B., Reiner G.* Batch Sizes Optimisation by Means of Queueing Network Decomposition and Genetic Algorithm // International Journal of Production Research. 2012. V. 50. P. 2720–2731.
3. *Stankevich E. P., Tananko I. E., Dolgov V. I.* Analysis of Closed Queueing Networks with Batch Service // Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform. 2020. V. 20. P. 527–533.
4. *Baetens J., Steyaert B., Claeys D., Bruneel H.* System Occupancy in a Multiclass Batch-Service Queueing System with Limited Variable Service Capacity // Annals of Operations Research. 2020. V. 293. P. 3–26.
5. *Bellalta B., Oliver M.* A Space-Time Batch-Service Queueing Model for Multi-User MIMO Communication Systems // Proceedings of the 12th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. 2009. P. 357—364.
6. *Grippa P., Schilcher U., Bettstetter Ch.* On Access Control in Cabin-Based Transport Systems // IEEE Transactions on Intelligent Transportation Systems. 2019. V. 20. P. 2149–2156.
7. *Станкевич Е. П., Тананко И. Е, Пагано М.* Анализ системы массового обслуживания с групповым обслуживанием требований // Компьютерные науки и информационные технологии: Материалы Междунар. науч. конф. Саратов, СГУ, 18 ноября – 20 ноября 2021 г. Саратов (в печати).

Станкевич Елена Петровна — к. ф.-м. н., доцент, кафедра системного анализа и автоматического управления. E-mail: *StankevichElena@mail.ru*

Тананко Игорь Евстафьевич — к. ф.-м. н., доцент, заведующий кафедрой, кафедра системного анализа и автоматического управления. E-mail: *TanankoIE.sgu@gmail.com*

Пагано Микеле — PhD, профессор, департамент информационной инженерии. E-mail: *Michele.Pagano@iet.unipi.it*

МОМЕНТЫ ПРИРАЩЕНИЙ И ПЕРИОДА РЕГЕНЕРАЦИИ ОБОБЩЕННЫХ РЕГЕНЕРИРУЮЩИХ СЛУЧАЙНЫХ ПРОЦЕССОВ В СИСТЕМЕ МАССОВОГО ОБСЛУЖИВАНИЯ M/G/1 С ПОЛУМАРКОВСКИМ ПОСТУПЛЕНИЕМ ТРЕБОВАНИЙ

Г. Т. Турсунов

Ташкентский университет информационных технологий, г. Ташкент, Узбекистан

В данной работе рассмотрена система массового обслуживания с пуассоновским входящим потоком, с одним обслуживающим прибором. Требование, заставшее обслуживающий прибор занятым обслуживанием становится в очередь, места для ожидания обслуживания не ограничены. Время обслуживания имеет произвольный закон распределения. В систему поступают требования m различных типов, которые образуют однородный цепь Маркова. Составлены уравнения для моментов приращений и периода регенерации обобщенных регенерирующих случайных процессов (интеграл от величины очереди в системе, включая обслуживаемое требование, число обслуженных требований за время t и число закончившихся периодов занятостей к моменту t), получены условия существования их решений.

Ключевые слова: *Система массового обслуживания, пуассоновский поток, цепь Маркова, полумарковский процесс, матричные уравнения, величина очереди, число обслуженных требований.*

Рассмотрим систему массового обслуживания с пуассоновским входящим потоком с параметром λ и с одним обслуживающим прибором. Требование, заставшее обслуживающий прибор занятым обслуживанием становится в очередь, места для ожидания обслуживания не ограничены.

В систему поступают требования m различных типов, причем первое требование типа J_1 поступает в момент $t = 0$ и обслуживается в течении времени τ_1. Обозначим через τ_k время обслуживания и J_k тип k-го поступившего в систему требования, $k \geqslant 1$. Предположим, что $\{\tau_k, k \geqslant 1\}$ является последовательностью независимых случайных величин, $\{J_k, k \geqslant 1\}$ является однородной, неприводимой и эргодической цепью Маркова с конечным множеством состояний $H = \{1, 2, \ldots, m\}$, матрицей переходных вероятностей $P = \|p_{i,j}\|_{i,j \in H}$, начальным распределением $\{p_i, i \in H\}, p_i = P\{J_1 = i\}$ и со стационарным распределением $\{\pi_k, k \in H\}$, где $p_{i,j} = P\{J_{k+1} = j/J_k = i\}$.

Пусть $\{(\tau_k, J_k), k \geqslant 1\}$ — полумарковский процесс с пространством состояний $Z = [0, \infty) \times \{1, 2, \ldots, m\}$ и полумарковским стохастическим ядром $P\{\tau_{k+1} \leqslant x, J_{k+1} = j/\tau_k = t, J_k = i\} = P\{\tau_{k+1} \leqslant x, J_{k+1} = j/J_k = i\} = F_{i,j}(x)$ для $k \geqslant 1, 0 \leqslant x \leqslant \infty$, $i \in H, j \in H$ [1].

Обозначим матрицы $F(x) = \|F_{i,j}(x)\|_{i,j \in H}$ и $M(l) = \|\alpha_{i,j}(l)\|_{i,j \in H}$, где $\alpha_{i,j}(l) = \int_0^\infty x^l dF_{i,j}(x)$ моменты порядка $l \geqslant 1$ времени обслуживания, загрузку системы $\rho = \lambda \sum_{i,j=1}^m \alpha_{i,j}(1)\pi_j$ и $\alpha = \max\limits_{1 \leqslant i \leqslant m} \sum_{j=1}^m \alpha_{i,j}(1)$.

Полумарковский процесс $\{(\tau_r, J_k), k \geqslant 1\}$ является правильным $\left(\sum_{j=1}^m p_{i,j} = 1\right)$ и непрерывным, если $\rho \leqslant 1$ и положительно возвратным, если $\rho < 1$ [1].

Обозначим через $\eta(t)$, $\nu(t)$ и $N(t) = \max(n \geqslant 1 : \sum_{i=1}^n \omega_i \leqslant t)$, соответственно, величину очереди в системе, включая обслуживаемое требование, число обслуженных требований за время t и число закончившихся периодов занятостей к моменту t.

Для изучения асимптотического поведения обобщенных регенерирующих случайных процессов $\xi(t) = \int_0^t \eta(t)dt$, $\nu(t)$ и $N(t)$, $0 \leqslant t \leqslant \infty$ при возрастании времени функционирования системы или при $\rho \to \rho_0 \leqslant 1$ возникает необходимость по мере возможности явно указать центрирующие, нормирующие и предельные константы, а эти постоянные выражаются через моменты приращений вышеуказанных процессов за периоды занятости системы, которые являются периодами регенерации, поскольку в периодах простоя системы, следующих за периодами занятости системы, значения случайных процессов $\xi(t)$ и $\nu(t)$ равны нулю.

Составлены уравнения для моментов приращений и периода регенерации вышеуказанных обобщенных регенерирующих случайных процессов и получены условия существования их решений.

Если $\{\omega_k, k \geqslant 1\}$ последовательность периодов занятости системы, $I_1 = J_1$ и I_k тип первого обслуженного требования в k-том периоде занятости ω_k, то в наших предположениях $\{I_k, k \geqslant 1\}$ является однородной, неприводимой и эргодической цепью Маркова с конечным множеством состояний H, матрицей переходных вероятностей $Q = \|q_{i,j}\|_{i,j=1}^m$, начальным распределением $\{q_i, i \in H\}, q_i = P\{I_1 = i\}$ и со стационарным распределением $\{\upsilon_k, k \in H\}$, где $q_{i,j} = P\{I_{k+1} = j/I_k = i\}$, а последовательность $\{(\omega_k, I_k), k \geqslant 1\}$ является полумарковским процессом с пространством состояний $Z = [0, \infty) \times \{1, 2, \ldots, m\}$

и полумарковским стохастическим ядром $Q(x) = \|Q_{i,j}(x)\|_{i,j\in H}$, где $P\{\omega_{k+1} \leqslant x,\ I_{k+1} = j/I_k = i\} = Q_{i,j}(x)$, причем если $\rho < 1$, то он положительно возвратный.

Обозначим $P_k = \|p_{i,j}(k)\|_{i,j\in H}, p_{i,j}(k) = \int_0^\infty e^{-\lambda x}\frac{(\lambda x)^k}{k!}dF_{i,j}(x), k \geqslant 0$.

Теорема 1. 1) Матрица переходных вероятностей Q цепи Марков $\{I_k, k \geqslant 1\}$ удовлетворяет уравнению $Q = \sum_{k=1}^\infty P_k Q^k$.

2) Если $\lambda\alpha < 1$, то это матричное уравнение имеет одно и только одно решение в виде стохастической матрицы и это решение может быть получено как предел последовательности матриц $Q(n) = \sum_{k=1}^\infty P_k Q(n-1)^k$, $n \geqslant 1$, где $Q(0)$произвольная стохастическая матрица, причем оценка ошибки n-го приближения

$$\Delta_n \equiv \rho(Q(n), Q) \leqslant \frac{(\lambda\alpha)^n}{1-\lambda\alpha}\rho\left(Q(0),\ \sum_{k=1}^\infty P_k Q(0)^k\right).$$

Введем приращения случайных процессов $\xi(t)$ и $\nu(t)$: $\xi_k = \xi(t_k) - \xi(t_{k-1})$ и $\nu_k = \nu(t_k) - \nu(t_{k-1})$, где $t_k = \sum_{i=1}^k \omega_i, k \geqslant 1, t_0 = 0$.

В наших предположениях последовательности $\{(\xi_k, I_k), k \geqslant 1\}$ и $\{(\nu_k, I_k), k \geqslant 1\}$ являются полумарковскими процессами, причем если $\rho < 1$, то они положительно возвратные.

Введем матрицы моментов приращений и периода регенерации обобщенных регенерирующих случайных процессов $\xi(t)$ и $\nu(t)$ для $l \geqslant 1$:

$$X(l) = \|x_{i,j}(l)\|_{i,j\in H}, \quad x_{i,j}(l) = \int_0^\infty u^l dQ_{i,j}(u),$$

$$Y(l) = \|y_{i,j}(l)\|_{i,j\in H}, \quad y_{i,j}(l) = \int_0^\infty u^l dP\{\xi_k \in du, I_{k+1} = j/I_k = i\},$$

$$Z(l) = \|z_{i,j}(l)\|_{i,j\in H}, \quad z_{i,j}(l) = \int_0^\infty u^l dP\{\nu_k \leqslant du, I_{k+1} = j/I_k = i\},$$

$$T = \|t_{i,j}\|_{i,j\in H}, \quad t_{i,j} = \int_0^\infty\int_0^\infty usdP\{\xi_k \in du, \omega_k \in ds, I_{k+1} = j/I_k = i\}.$$

Обозначим через $\tilde{\Pi}_m$ пространство всех m-мерных квадратных матриц с неотрицательными элементами и расстоянием $\rho(A,B) = \max\limits_{0\leqslant i,j\leqslant m}|a_{i,j} - b_{i,j}|$, где $A = \|a_{i,j}\|, B = \|b_{i,j}\| \in \tilde{\Pi}_m$. Тогда ρ-сходимость эквивалентна по элементной сходимости матриц с неотрицательными элементами и предел каждой сходящейся в себе последовательности матриц с неотрицательными элементами является также матрицей с неотрицательными элементами и следовательно $\left(\tilde{\Pi}_m, \rho\right)$ полное метрическое пространство. На этом пространстве для $l \geqslant 1$ и

$A_k, B, C \in \tilde{\Pi}_m, k \geqslant 1$ рассмотрим операторы $f_l : \tilde{\Pi}_m \to \tilde{\Pi}_m$

$$f_1(A_k, B) = \sum_{k=1}^{\infty} A_k \sum_{r=0}^{k-1} Q^r B Q^{k-r-1},$$

$$f_2(A_k, B) = \sum_{k=1}^{\infty} A_k \sum_{r=0}^{k-1} (r-1) Q^r B Q^{k-r-1},$$

$$f_3(A_k, B) = \sum_{k=1}^{\infty} A_k \sum_{r=0}^{k-1} (r-1)^2 Q^r B Q^{k-r-1},$$

$$f_4(A_k, B) = \sum_{k=2}^{\infty} A_k \sum_{\substack{r_1 + r_2 + r_3 = k-2 \\ 0 \leqslant r_1, r_2, r_3 \leqslant \infty}} Q^{r_1} B Q^{r_2} B Q^{r_3},$$

$$f_5(B, C) = \sum_{k=2}^{\infty} P_k \sum_{\substack{r_1 + r_2 + r_3 = k-2 \\ 0 \leqslant r_1, r_2, r_3 \leqslant \infty}} Q^{r_1} B Q^{r_2} C Q^{r_3},$$

$$f_6(B) = \sum_{k=2}^{\infty} P_k \sum_{\substack{r_1 + r_2 + r_3 + r_4 = k-3 \\ 0 \leqslant r_1, r_2, r_3, r_4 \leqslant \infty}} Q^{r_1} B Q^{r_2} B Q^{r_3} B Q^{r_4}.$$

С помощью этих операторов и матриц $F(x) = \|F_{i,j}(x)\|_{i,j \in H}$, $Q = \|q_{i,j}\|_{i,j=1}^m$ введем следующие матрицы

$$L_1 = \int_0^{\infty} u \sum_{k=0}^{\infty} e^{-\lambda u} \frac{(\lambda u)^k}{k!} dF(u) Q^k,$$

$$L_2 = \int_0^{\infty} u^2 \sum_{k=0}^{\infty} e^{-\lambda u} \frac{(\lambda u)^k}{k!} dF(u) Q^k +$$
$$+2 \int_0^{\infty} u \sum_{k=0}^{\infty} e^{-\lambda u} \frac{(\lambda u)^k}{k!} dF(u) f_1(P_k, X(1)) +$$
$$+ \sum_{k=1}^{\infty} P_k \sum_{r_1=0}^{k-1} \left\{ \sum_{r_2=0}^{r_1-1} Q^k X(1) Q^{r_1-r_2-1} X(1) Q^{k-r_1-1} + \right.$$
$$\left. + \sum_{r_3=0}^{k-r_1-2} Q^{r_1} X(1) Q^{r_3} X(1) Q^{k-r_3-r_1-2} \right\},$$

$K_1 = ||1||$ m-мерная матрица состоящая из 1,

$$K_2 = K_1 + 2f_1(P_k, Z(1)) + f_4(P_k, Z(1))$$
$$K_3 = K_1 + 3f_1(P_k, Z(1)) + 3f_1(P_k, Z(2)) + 3f_4(P_k, Z(1)) +$$
$$+\frac{1}{2}\left(f_5(Z(1), Z(2) + f_5(Z(2), Z(1)) + f_6(Z(1)),\right.$$
$$N_1 = (M(1) + \frac{\lambda}{2}M(2))K_1 + f_2(P_k, X(1)),$$
$$N_2 = (M(2) + \frac{4\lambda}{3}M(3) + \frac{\lambda}{4}M(4))K_1 + 2f_1(B_k, Y(1)) +$$
$$+f_4(P_k, Y(1)) + f_4(P_k, X(1)) + \frac{1}{2}\left(f_5(X(1), Y(1) + f_5(Y(1), X(1)) +\right.$$
$$+2f_2(P_k, T) + f_3(P_k, X(2)),$$
$$N_3 = (M(2) + \frac{\lambda}{2}M(3))K_1 + f_1(B_k, X(1)) + f_2(P_k, X(1)) + f_1(C_k, Y(1)) +$$
$$+f_2(P_k, X(2)) + f_4(P_k, X(1)) + \frac{1}{2}\left(f_5(X(1), Y(1) + f_5(Y(1), X(1))\right),$$

здесь $B_k = C_k + G_k$, $C_k = \|\alpha_{i,j}(1)p_{i,j}(k)\|_{i,j\in H}$, а матрица $G_k = \|\gamma_{i,j}p_{i,j}(k)\|_{i,j\in H}$ имеет производящую функцию

$$\sum_{k=0}^{\infty} z^k G_k = \frac{1}{\lambda(1-z)^3}[\lambda(1-z)M(1) + \Phi(\lambda - \lambda z) - K_1], \Phi(s) =$$
$$= \|f_{i,j}(s)\|_{i,j\in H}, f_{i,j}(s) = \int_0^\infty e^{-sx} dF_{i,j}(x).$$

Теорема 2. 1) Матрицы моментов приращений обобщенных регенерирующих случайных процессов $\xi(t)$ и $\nu(t)$ и периода их регенерации удовлетворяют следующим уравнениям:

$$f(A) = f_1(P_k, A),$$
$$X(l) = f_1(P_k, X(l)) + L_l,$$
$$Y(l) = f_1(P_k, Y(l)) + N_l,\ l = 1, 2;$$
$$Z(l) = f_1(P_k, Z(l)) + K_l,\ l = 1, 2, 3.$$

2) Если $\lambda\alpha < 1$, то эти матричные уравнения имеют одно и только одно решение и эти решения могут быть получены как пределы после-

довательностей матриц

$$T^{(n)} = f_1(P_k, T^{(n-1)}) + N_3,$$
$$X^{(n)}(l) = f_1(P_k, X^{(n-1)}(l)) + L_l, Y^{(n)}(l) = f_1(P_k,$$
$$Y^{(n-1)}(l)) + N_l,\ l = 1, 2;$$
$$Z^{(n)}(l) = f_1(P_k, Z^{(n-1)}(l)) + K_l,\ l = 1, 2, 3, n \geqslant 1,$$

соответственно, где $T^{(0)}, X^{(0)}, (l), Y^{(0)}(l), Z^{(0)}(l)$ произвольные матрицы с положительными элементами, причем оценки ошибок n-го приближения равны, соответственно

$$\Delta_{n,1} \equiv \rho(T^{(n)}, T) \leqslant \frac{(\lambda\alpha)^n}{1-\lambda\alpha}\rho\left(T^{(0)}, f_1\left(P_k, T^{(0)}\right) + N_3\right),$$
$$\Delta_{n,2} \equiv \rho\left(X^{(n)}(l), X(l)\right) \leqslant \frac{(\lambda\alpha)^n}{1-\lambda\alpha}\rho\left(X^{(0)}(l), f_1(P_k, X^{(0)}(l)) + L_l\right),$$
$$\Delta_{n,4} \equiv \rho(Z^{(n)}(l), Z(l)) \leqslant \frac{(\lambda\alpha)^n}{1-\lambda\alpha}\rho(Z^{(0)}(l), f_1(P_k, Z^{(0)}(l)) + K_l)).$$

СПИСОК ЛИТЕРАТУРЫ

1. *Сильвестров Д. С.* Полумарковские процессы с дискретным множеством состояний, Москва, «Советское радио», 1980.

Турсунов Гафур — . E-mail: *tursunovgafur52@gmail.com*

АСИМПТОТИЧЕСКИЙ АНАЛИЗ RQ-СИСТЕМЫ MMPP/M/N В УСЛОВИИ БОЛЬШОЙ ЗАГРУЗКИ

Е. А. Фёдорова

Национальный исследовательский
Томский государственный университет, г. Томск, Россия

В работе исследуется многолинейная RQ-система с входящим MMPP-потоком. Для нахождения стационарного распределения числа заявок на орбите предлагается метод асимптотического анализа в условии большой загрузки. Доказано, что асимптотическая характеристическая функция имеет вид гамма-распределения с найденными параметрами.

Ключевые слова: *RQ-система, асимптотический анализ, большая загрузка.*

RQ-системы являются важными математическими моделями, широко используемыми для анализа и оптимизации телекоммуникационных систем, управляемыми протоколами множественного доступа, когнитивных сетей, сетей мобильной связи, call-центров и др. Описание RQ-систем и основные результаты исследований с простейшим входящим потоком приведены в монографиях Г. И. Фалина [1] и J. Artalejo [2]. Однако реальные информационные потоки имеют более сложную структуру, в связи с этим, в работе исследована RQ-система с входящим MMPP-потоком.

1. Математическая модель

Рассмотрим многолинейную RQ-систему (рисунок 1), на вход которой поступает MMPP-поток заявок (Markov Modulated Poisson Process), который описывается матрицами $\mathbf{D_0}$ и $\mathbf{D_1}$.

Обозначим: $n(t)$ – цепь Маркова, управляющая MMPP-потоком ($n = 1, 2, \ldots, W$). Матрица инфинитезимальных характеристик управляющего процесса $n(t)$ равна $\mathbf{Q} = \mathbf{D_0} + \mathbf{D_1}$. Матрица $\mathbf{D_1} = \mathbf{\Lambda}$ – диагональная матрица с элементами λ_n (условными интенсивностями). Очевидно, что интенсивность MMPP-потоком равна $\lambda = \mathbf{r\Lambda e}$, где $\mathbf{e}$ – единичный вектор-столбец, $\mathbf{r}$ – стационарное распределение вероятностей состояний $n(t)$, определяемое из системы уравнений:

Исследование выполнено при финансовой поддержке РФФИ и Томской области в рамках научного проекта № 19-41-703002.

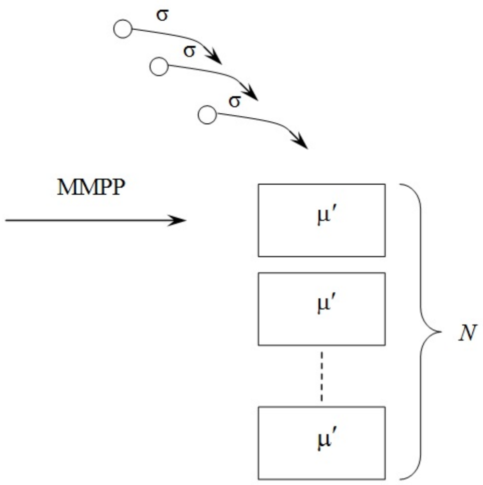

Рис. 1. RQ-система MMPP|M|N

$$\begin{cases} \mathbf{rQ} = \mathbf{0}, \\ \mathbf{re} = 1. \end{cases} \tag{1}$$

Если поступившая заявка застает любой прибор свободным, то она занимает его для обслуживания. Время обслуживания каждой заявки распределено по экспоненциальному закону с параметром μ'. Если все приборы заняты, то заявка переходит на орбиту, где осуществляет случайную задержку, распределенную экспоненциально с параметром σ. После случайной задержки заявка с орбиты вновь обращается блоку обслуживающих приборов: если хотя бы один прибор свободен, то заявка занимает его для обслуживания; если все приборы заняты, заявка мгновенно возвращается на орбиту для реализации следующей задержки. Введем параметр ρ, характеризующий загрузку системы. Очевидно, что $\rho = \lambda/(N\mu')$. Тогда введем обозначение $\mu = \rho N \mu'$, такое что $\lambda/\mu \equiv 1$ или

$$\mathbf{r\Lambda e} = \mu. \tag{2}$$

Тогда в последующих уравнениях будем использовать замену:

$$\mu' = \mu/(\rho N).$$

Пусть (t) – число заявок на орбите, а $k(t)$ – состояние блока обслуживания, определяемое следующим образом:

$$k(t) = \begin{cases} 0, \text{если все приборы свободны в момент времени } t, \\ 1, \text{если 1 прибор занят в момент времени } t, \\ \ldots \\ N, \text{если все приборы заняты в момент времени } t. \end{cases}$$

Обозначим $P(k,n,i,t) = P\{k(t) = k, n(t) = n, i(t) = i\}$ – вероятность того, что прибор в момент времени t находится в состоянии k, цепь Маркова, управляющая MMPP-потоком, находится в состоянии n и на орбите i заявок. Очевидно, что процесс $\{k(t), n(t), i(t)\}$ изменения состояний данной системы является Марковским. Для распределения вероятностей $P(k,i,n,t)$ составим систему уравнений Колмогорова (для $i \geqslant 0$, $n = \overline{1,W}$):

$$\begin{cases} \dfrac{\partial P(0,n,i,t)}{\partial t} = -(\lambda_n + i\sigma - q_{nn})P(0,n,i,t) + \dfrac{\mu}{\rho N}P(1,n,i,t) + \\ \quad + \sum\limits_{v \neq n} P(0,v,i,t)q_{vn}, \\ \dfrac{\partial P(k,n,i,t)}{\partial t} = -(\lambda_n + i\sigma + \dfrac{k\mu}{\rho N} - q_{nn})P(k,n,i,t) + \\ \quad + \lambda_n P(k-1,n,i,t) + \dfrac{(k+1)\mu}{\rho N}P(k+1,n,i,t) + \\ \quad + (i+1)\sigma P(k-1,n,i+1,t) + \sum\limits_{v \neq n} P(k,v,i,t)q_{vn}, \\ \quad k = \overline{1,N-1}, \\ \dfrac{\partial P(N,n,i,t)}{\partial t} = -(\lambda_n + \dfrac{\mu}{\rho} - q_{nn})P(N,n,i,t) + \\ \quad + \lambda_n P(N-1,n,i,t) + (i+1)\sigma P(N-1,n,i+1,t) + \\ \quad + \lambda_n P(N,n,i-1,t) + \sum\limits_{v \neq n} P(N,v,i,t)q_{vn}. \end{cases} \tag{3}$$

2. Метод асимптотического анализа

Обозначим вектор-строки $\mathbf{P}(k,i) = \{P(k,1,i), P(k,2,i), \ldots, P(k,N,i)\}$, где $P(k,n,i) = \lim\limits_{t\to\infty} P(k,n,i,t)$. Тогда в стационарном режиме в матричной форме система (1) примет вид:

$$\begin{cases} \mathbf{P}(0,i)(\mathbf{Q}-\mathbf{\Lambda}-i\sigma\mathbf{I})+\dfrac{\mu}{\rho N}\mathbf{P}(1,i)=\mathbf{0}, \\ \mathbf{P}(k,i)(\mathbf{Q}-\mathbf{\Lambda}-\dfrac{k\mu}{\rho N}\mathbf{I}-i\sigma\mathbf{I})+\mathbf{P}(k-1,i)\mathbf{\Lambda}+ \\ \quad +\sigma(i+1)\mathbf{P}(k-1,i+1)+\dfrac{(k+1)\mu}{\rho N}\mathbf{P}(k+1,i)=\mathbf{0}, \\ \quad k=\overline{1,N-1}, \\ \mathbf{P}(N,i)(\mathbf{Q}-\mathbf{\Lambda}-\dfrac{\mu}{\rho}\mathbf{I})+\mathbf{P}(N-1,i)\mathbf{\Lambda}+ \\ \quad +\sigma(i+1)\mathbf{P}(N-1,i+1)+\mathbf{P}(N,i-1)\mathbf{\Lambda}=\mathbf{0}, \end{cases} \tag{4}$$

где $\mathbf{I}$ – единичная матрица.

Перейдем в системе (4) к частичным характеристическим функциям:

$$\mathbf{H}(k,u)=\sum_i e^{jui}\mathbf{P}(k,i).$$

Получим следующую систему матричных уравнений:

$$\begin{cases} \mathbf{H}(0,u)(\mathbf{Q}-\mathbf{\Lambda})+j\sigma\dfrac{\partial\mathbf{H}(0,u)}{\partial u}+\dfrac{\mu}{\rho N}\mathbf{H}(1,u)=\mathbf{0}, \\ \mathbf{H}(k,u)(\mathbf{Q}-\mathbf{\Lambda}-\dfrac{k\mu}{\rho N}\mathbf{I})+j\sigma\dfrac{\partial\mathbf{H}(k,u)}{\partial u}+\mathbf{H}(k-1,u)\mathbf{\Lambda}- \\ \quad -j\sigma e^{-ju}\dfrac{\partial\mathbf{H}(k-1,u)}{\partial u}+\dfrac{(k+1)\mu}{\rho N}\mathbf{H}(k+1,u)=\mathbf{0}, \\ \quad k=\overline{1,N-1}, \\ \mathbf{H}(N,u)(\mathbf{Q}-\mathbf{\Lambda}-\dfrac{\mu}{\rho}\mathbf{I})+\mathbf{H}(N-1,u)\mathbf{\Lambda}+ \\ \quad -j\sigma e^{ju}\dfrac{\partial\mathbf{H}(N-1,u)}{\partial u}+e^{ju}\mathbf{H}(N,u)\mathbf{\Lambda}=\mathbf{0}. \end{cases} \tag{5}$$

Для решения системы (4) предложен метод асимптотического анализа [3, 4] в условии большой загрузки $\rho\to 1$. В результате была сформулирована и доказана следующая теорема.

Теорема 1. Асимптотическая характеристическая функция числа заявок на орбите в условии большой загрузки RQ-системе MMPP|M|N имеет вид характеристической функции гамма-распределения

$$h(u)=\left(1-\frac{ju}{(1-\rho)\beta}\right)^{-\alpha} \tag{6}$$

с параметрами

$$\alpha=1+\frac{\mu}{\sigma}\beta,\quad \beta=\frac{\mu}{\mu+\mathbf{v\Lambda e}-\mu\mathbf{ve}},$$

где вектор $\mathbf{v}$ является решением неоднородной системы $\mathbf{vQ} = \mathbf{r}(\mu\mathbf{I} - \mathbf{\Lambda})$.

Заключение

Таким образом, в работе рассмотрена многолинейная RQ-система с входящим MMPP-потоком. Доказано, что асимптотическая характеристическая функция числа заявок на орбите в условии большой загрузки имеет вид гамма-распределения, как и для однолинейных RQ-систем.

СПИСОК ЛИТЕРАТУРЫ

1. *Falin G. I., Templeton J. G. C.* Retrial queues. London: Chapman & Hall, 1997. 328 p.
2. *Artalejo J. R., Gomez-Corral A.* Retrial Queueing Systems. A Computational Approach. Springer-Verlag Berlin Heidelberg, 2008. 318 p.
3. *Гарайшина И. Р., Моисеева С. П., Назаров А. А.* Методы исследования коррелированных потоков и специальных систем массового обслуживания. Томск : Изд-во НТЛ, 2010. 202 с.
4. *Моисеев А.Н.* Асимптотический анализ системы массового обслуживания MAP/GI/∞ с высокоинтенсивным входящим потоком Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 3 (32). С. 56–65.

Фёдорова Екатерина Александровна — кандидат физико-математических наук, доцент кафедры Теории вероятностей и математической статистики ИПМКН ТГУ. E-mail: *moiskate@mail.ru*

Математическое и компьютерное моделирование в технике, естествознании, экономике, образовании, гуманитарных науках и других областях

ON ASYMPTOTICS OF A PROBABILITY OF THE EVENT: EACH CELL CONTAINS EVEN NUMBER OF PARTICLES

F.A. Abdushukurov[1,2]

[1] *Yeoju Technical Institute in Tashkent, Tashkent, Uzbekistan*
[2] *Institute of Mathematics named after V.I.Romanovsky of the Academy of Sciences of the Republic of Uzbekistan, Tashkent, Uzbekistan*

We consider the probability $P(n,N)$ of the event: under allocation of $2n$ distinguishable particles by N different cells each cell contains even number of particles and the probability $P'(n,N)$ of the event: under allocation of $2n$ non-distinguishable particles by N different cells each cell contains even number of particles. For various type of the convergence $N,n\to\infty$ we study assimptotical bihavior of $P(n,N)$ and $P'(n,N)$.

Keywords: *Allocation scheme, Gaussian random variable, Poisson limit theorem, Berry-Essen inequelity, limit theorem, local limit theorem.*

Introduction

Let n,N be integer numbers. A homogeneous allocation scheme of n distinguishable particles by N different cells is named the random variables $\eta_1,\dots,\eta_N$, with the joint distribution defined by formula

$$P\{\eta_1=k_1,\dots\eta_N=k_N\}=\frac{n!}{k_1!k_2!\cdots k_N!}\left(\frac{1}{N}\right)^n,$$

where $k_1,k_2,\dots k_N$ are nonnegative integer number such that $k_1+k_2+\cdots+k_N=n$. Many papers deal with limit theorems for allocation scheme of distinguishable paticles by different cells (see [1] and its references).

Denote by $P(n,N)$ the probability of the event: in allocation scheme of $2n$ distinguishable particles by N different cells each cell contains even number of particles.

Observe that

$$P(n,N)=\sum_{k_i\in\mathbf{N},\ 1\leqslant i\leqslant N,\ k_1+k_2+\cdots+k_N=n}\frac{(2n)!}{(2k_1)!(2k_2)!\cdots(2k_N)!}\frac{1}{N^{2n}}=$$

$$=\sum_{k_i\in\mathbf{N},\ 1\leqslant i\leqslant N,\ k_1+k_2+\cdots+k_N=n}\frac{\frac{\alpha^{2k_1}}{(2k_1)!}\frac{\alpha^{2k_2}}{(2k_2)!}\cdots\frac{\alpha^{2k_N}}{(2k_N)!}}{\frac{(\alpha N)^{2n}}{(2n)!}}=$$

$$= \left(\frac{ch(\alpha)}{e^{\alpha}}\right)^N \sum_{k_i\in\mathbf{N},\ 1\leqslant i\leqslant N,\ k_1+k_2+\dots+k_N=2n} \frac{\frac{\alpha^{2k_1}}{ch(\alpha)(2k_1)!}\frac{\alpha^{2k_2}}{ch(\alpha)(2k_2)!}\cdots\frac{\alpha^{2k_N}}{ch(\alpha)(2k_N)!}}{e^{-\alpha N}\frac{(\alpha N)^{2n}}{(2n)!}}.$$

Therefore, using Stirling formula for estimation of $(2n)!$, for $\alpha = \frac{2n}{N}$ we obtain

$$P(n,N) = (1+o(1))\sqrt{4\pi n}\left(\frac{1+e^{-2\alpha}}{2}\right)^N \mathbf{P}(\xi_1+\dots+\xi_N = 2n),$$

where $\xi_1, \xi_2, \dots \xi_N$ are independent random variables with the distribution

$$\mathbf{P}(\xi_i = 2k) = \frac{\alpha^{2k}}{(2k)!ch(\alpha)},\ \ k = 0,1,2\dots,$$

or

$$P(n,N) = (1+o(1))\sqrt{4\pi n}\left(\frac{1+e^{-2\sqrt{\alpha_1}}}{2}\right)^N \mathbf{P}(\xi'_1+\dots+\xi'_N = n),$$

where $\xi'_1, \xi'_2, \dots \xi'_N$ are independent random variables with the distribution

$$\mathbf{P}(\xi'_i = k) = \frac{\alpha_1^k}{(2k)!ch(\sqrt{\alpha_1})},\ \ k = 0,1,2\dots,\ \ \alpha_1 = \alpha^2.$$

The expectation and the variance of ξ_i are

$$e(\alpha) = \alpha th(\alpha),\quad \sigma^2(\alpha) = \frac{\alpha^2}{ch^2(\alpha)} + \alpha th(\alpha).$$

Using Poisson limit theorem for an estimation of the probability

$$\mathbf{P}(\xi_1+\dots+\xi_N = 2n),$$

we obtain the following theorem.

Theorem 1. *Let $n, N \to \infty$ such that the family of numbers*

$$N\frac{\alpha^2}{2} = \lambda$$

is bounded. Then we have

$$P(n,N) = (1+o(1))\sqrt{4\pi n}e^{-4n}\left(e^{-\lambda}\frac{\lambda^n}{n!} + O\left(\frac{1}{N}\right)\right).$$

In the next theorem for an estimation of the probability $\mathbf{P}(\xi_1+\dots+\xi_N = 2n)$ we use the local limit theorem from [2].

Theorem 2. *Let $0 < \alpha' < \alpha'' < \infty$. Let $n, N \to \infty$ such that $\alpha' < \alpha < \alpha''$. Then we have*

$$P(n,N) = (1+o(1))\sqrt{\frac{1}{\frac{\alpha}{ch^2(\alpha)} + th(\alpha)}}\left(\frac{1+e^{-2\alpha}}{2}\right)^N e^{-NC(\alpha)},$$

where

$$C(\alpha) = \frac{N\alpha e^{-2\alpha}}{2(\alpha + sh(\alpha)ch(\alpha))}.$$

In the next theorem for an estimation of the probability $\mathbf{P}(\xi_1+\cdots+\xi_N = 2n)$ we use the local limit theorems from [3].

Theorem 3. *Let $n, N \to \infty$ such that $\alpha \to 0$ and $N\alpha^2 \to \infty$. Then we have*

$$P(n,N) = (1+o(1))\sqrt{\frac{1}{2\alpha}}\left(\frac{1+e^{-2\alpha}}{2}\right)^N \left(e^{-NC(\alpha)} + o(1)\right).$$

In the next theorem we use a local limit theorem tor the sum $\xi_1 + \cdots + \xi_N = 2n$ as $\alpha \to \infty$. Since ξ_i are latticed random variables with the step 2, the probability $\mathbf{P}(\xi_1+\cdots+\xi_N = 2n)$ has the same estimation as in classical local limit theorem but with the coefficient 2.

Theorem 4. *Let $n, N \to \infty$ such that $\alpha \to \infty$. Then we have*

$$P(n,N) = (1+o(1))\left(\frac{1+e^{-2\alpha}}{2}\right)^N \left(2e^{-NC(\alpha)} + o(1)\right).$$

Theorem 5. *Let $n, N \to \infty$ such that $Ne^{-2\alpha} \to \beta$, $0 \leqslant \beta < \infty$. Then we have*

$$P(n,N) = (1+o(1))\frac{e^{\beta}}{2^{N-1}}.$$

Theorem 6. *Let $n, N \to \infty$ such that $ne^{-2\alpha} < \beta$, for some $\beta < \infty$. Then we have*

$$P(n,N) = (1+o(1))\frac{1}{2^{N-1}}.$$

Allocation scneme of n non-distingushing paticles by N different cells are the random variables $\eta_1, \ldots, \eta_N$ with the joint distribution

$$P\{\eta_1 = k_1, \ldots \eta_N = k_N\} = \frac{1}{C_{n+N-1}^{N-1}},$$

where $k_1, k_2, \ldots k_N$ are nonnegative integer numbers such that $k_1 + k_2 + \cdots + k_N = n$.

Denote by $P'(n,N)$ the probability of the event: under allocation of $2n$ non-distingusning particles by N different cells each cell contains even number of particles. We have

$$P'(n,N) = \sum_{k_i \in \mathbf{N},\ 1 \leqslant i \leqslant N,\ k_1+k_2+\cdots+k_N = n} \frac{1}{C_{2n+N-1}^{N-1}} = \frac{C_{n+N-1}^{N-1}}{C_{2n+N-1}^{N-1}}.$$

Theorem 7. *Let $n, N \to \infty$ such that $\frac{N^2}{4n} \to \beta$, where $0 \leqslant \beta < \infty$. Then we have*

$$P'(n,N) = (1+o(1))\frac{e^{\beta}}{2^{N-1}}.$$

Consider N independent allocation schemes of $2n$ non-distinguishing particles by 2 different cells. Let A_i, $1 \leqslant i \leqslant N$, be the event: 1-th cell contains even number of particles from i-th scheme. Then $\cap_{i=1}^{N} A_i$, $1 \leqslant i \leqslant N$, is the event: 1-th cell contains even number of particles from i-th scheme for all $1 \leqslant i \leqslant N$. Denote $P^*(n, N) = \mathbf{P}\left(\cap_{i=1}^{N} A_i\right)$. We have

$$P^*(n,N) = \left(\frac{n+1}{2n+1}\right)^N = \frac{1}{2^N}\left(1+\frac{1}{2n+1}\right)^N = \frac{1}{2^N}e^{\frac{N}{2n}(1+o(1))},$$
$$\text{as} \quad n \to \infty.$$

REFERENCES

1. *Kolchin V. F., Sevast'ynov B. P., Chistiakov V. P.* Random allocation // Scripta Series in Mathematics. V. H. Winston & Sons, 1978. City: Washington, DC, 1978.
2. *Timashev A. N.* Asimptotic expansions in probabilistic combinatorics // M.: TVP. 2011.
3. *Kolchin A. V. and Kolchin V. F.* On transition of distributions of sums of independent identically distributed random variables from one lattice to another in the generalised allocation scheme // Discrete Math. Appl. 2006. Vol. 16, No. 6. P. 527–540.

F.A. Abdushukurov — Senior Lecturer of the Department of Information Technologies. E-mail: *f.abdushukurov@ytit.uz*

ON ESTIMATES OF THE DISTRIBUTION FUNCTION IN INFORMATIVE MODEL OF RANDOM CENSORSHIP FROM BOTH SIDES

A. A. Abdushukurov[1], D. R. Mansurov[2]

[1]*Moscow State University Tashkent branch, Tashkent, Uzbekistan*
[2]*Navoi State Pedagogical Institute, Navoi, Uzbekistan*

In the paper, we investigate two semiparametrical estimators of distribution function in informative model of random censorship from both sides. Under investigating of estimators, the characterization properties of the considered informative model is used. The properties of the semiparametric estimators by using methods of numerical modeling are discussed.
Keywords: *Random censorship, informative model, proportional hazards, exponential distribution.*

Introduction

In biomedical studies of individuals for survival, in engineering tests of technical devices for reliability, there may be cases when the tested objects coming under observation after a certain random time after the start of testing. This phenomenon is called delayed entry or left random censoring. The random variable (r.v.) X that characterizes the lifetime of the tested object becomes to observation under the condition $X \geqslant L$. Here L is the moment when the object was placed under observation. In addition to this, r.v. X may also be censored from the right by some other r.v. Y. We are interested in r.v. X which will be subjected to random censorship from both sides by random vector (L, Y). Let r.v.-s L, X and Y are mutually independent with continuous distribution functions (d.f.) K, F and G respectively.

1. Informative model of random censorship from both sides

In this paper we present some asymptotic properties of power-type semiparametric estimator of survival function of the lifetime in informative model of random censorship from both sides.

Let $\{(X_k, L_k, Y_k), \ k \geqslant 1\}$ be a sequence of independent realizations of the triple (X, L, Y) and

$$S^{(n)} = \{(\mathrm{Z}_i, \Delta_i), \, i = 1, \dots, n\}$$

is the observed sample, where $\mathrm{Z}_i = \max\{L_i, \min\{X_i, Y_i\}\}$, $\Delta_i = \left(\delta_i^{(0)}, \delta_i^{(1)}, \delta_i^{(2)}\right)$, with $\delta_i^{(0)} = I(\min(X_i, Y_i) < L_i)$, $\delta_i^{(1)} = I(L_i \leqslant X_i < Y_i)$, $\delta_i^{(2)} = I(L_i \leqslant Y_i < X_i)$ and $I(A)$ standing for indicator of the event A. Note that in sample $S^{(n)}$ the number of observed r.v.-s X_i is equal to $\delta_1^{(1)} + ... + \delta_n^{(1)}$. The statistical task is consist in estimating of d.f. F from a sample $S^{(n)}$. However, such a general statement of the estimating problem, d.f.-s K and G are considered as a nuisance. In this paper, we will investigate the evaluation of d.f. F in the case of informative censoring from two sources, when d.f.-s K and G functionally depend on F. To describe such a model by H and N we denote d.f.-s of r.v.-s Z_i and $V_i = \min(X_i, Y_i)$. Then it is easy to see, that

$$H(x) = K(x)N(x), \quad N(x) = 1 - (1 - F(x))(1 - G(x)), \quad x \in \mathrm{R}^1. \quad (1)$$

Assume that there are positive unknown parameters θ, β such that following representations are valid for all $x \in \mathrm{R}^1$:

$$\begin{cases} 1 - G(x) = (1 - F(x))^{\theta}, \\ K(x) = (N(x))^{\beta}, \end{cases} \quad (2)$$

where the parameter β is responsible for the power of censoring from the left, and $\theta-$from the right. The proximity of the values of these parameters to zero determines the weakness of the corresponding censorship. Such a special model of random censorship on both sides was introduced in [1] and generalized for the case of competing risks in [2]. In this model, the censoring parameters θ and β, which determine the deepness of censorship from both sides. From formulas (1) and (2), it is not difficult to derive the following representation for d.f. F :

$$1 - F(x) = \left[1 - (H(x))^{\lambda}\right]^{\gamma}, \quad x \in \mathrm{R}^1, \quad (3)$$

where $\lambda = \frac{1}{1+\beta}$, $\gamma = \frac{1}{1+\theta}$ and therefore, the closeness of the parameters λ and γ to 1, denotes the weakness of the censoring. Using representation (3), we can construct a semiparametric estimate for F over a sample $S^{(n)}$ by estimating a triple $(H(x), \lambda, \gamma)$. For this purpose, we also use the characterization properties of the model under consideration. In this regard, we define the sub-distributions $\left\{T^{(m)}(x) = P\left(\mathrm{Z}_i \leqslant x, \delta_i^{(m)} = 1\right), m = 0,1,2\right\}$ by the following formulas:

$$T^{(0)}(x) = \int_{-\infty}^{x} N(s)\,dK(s), \quad T^{(1)}(x) = \int_{-\infty}^{x} K(s)(1 - G(s))\,dF(s),$$

$$T^{(2)}(x) = \int_{-\infty}^{x} K(s)(1 - F(s))\, dG(s).$$

It's easy to see that $T^{(0)}(x) + T^{(1)}(x) + T^{(2)}(x) = H(x)$, $x \in \mathrm{R}^1$ and because of representation (3) in the model (2):

$$T_F = T_G = T_N = T_K = T_H = \inf\left\{x \in \mathrm{R}^1 : H(x) = 1\right\},$$
$$\tau_F = \tau_G = \tau_N = \tau_K = \tau_H = \sup\left\{x \in \mathrm{R}^1 : H(x) = 0\right\}.$$

The following theorem characterizes the model (2). It firstly announced without the proof in paper [1].

Theorem 1. Equalities (2) hold if and only if the r.v.-s Z_i and Δ_i are independent.

Suppose that conditions (2) hold. Then the calculation of subdistribution functions easy gives

$$T^{(0)}(x) = (1-\lambda) H(x), \ T^{(1)}(x) = \gamma\lambda H(x), \ T^{(2)}(x) = (1-\gamma)\lambda H(x) \quad (4)$$

and under $x \to +\infty$ from (4) we have

$$P\left(\delta_i^{(0)} = 1\right) = 1 - \lambda, \ P\left(\delta_i^{(1)} = 1\right) = \gamma\lambda, \ P\left(\delta_i^{(2)} = 1\right) = (1-\gamma)\lambda. \quad (5)$$

Estimating the probabilities $p^{(m)} = P\left(\delta_i^{(m)} = 1\right)$, $m = 0, 1, 2$ by the corresponding frequencies $p_n^{(m)} = \frac{1}{n}\sum_{i=1}^{n} \delta_i^{(m)}$, $m = 0, 1, 2$ from formulas (5), we find the following estimates of the parameters λ and γ : $\lambda_n = 1 - p_n^{(0)}$, $\gamma_n = p_n^{(1)}\left(1 - p_n^{(0)}\right)^{-1}$. For d.f. $H(x)$ we use an empirical estimator

$$H_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\mathrm{Z}_i \leqslant x), \quad x \in \mathrm{R}^1.$$

Now, because of (3), we construct the corresponding estimate for d.f. $F(x)$ by the plug in method:

$$F_n(x) = 1 - \left[1 - (H_n(x))^{\lambda_n}\right]^{\gamma_n}, \quad x \in \mathrm{R}^1. \quad (6)$$

2. Studies of the dependence of the estimator on unknown parameters of censorship

Numerical studies were conducted using the python programming language to clarify how much the estimate $F_n(x)$ depends on the parameters, demonstrating its proximity to the estimated function $F(x)$ (see, figures 1-4). As $F(x)$ we use the standard normal and exponential d.f.-s. Simulated volume data was used for this purpose is $n = 500$. From figures 1-4, we can see that when $0 < \beta, \theta \leqslant 1$ the estimator $F_n(x)$ is well approximated to $F(x)$ (figures 1-2). In other cases, with an increasing of censoring from the left and right the discrepancy between the estimate and d.f. is noticeable.

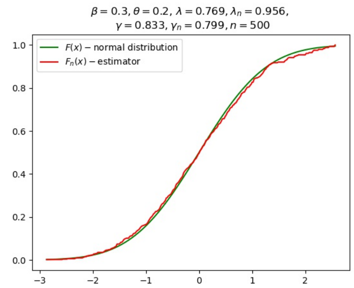

Figure 1. $\beta = 0.3, \quad \theta = 0.2, \quad n = 500$

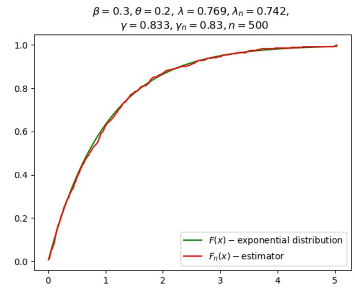

Figure 2. $\beta = 0.3, \quad \theta = 0.2, \quad n = 500$

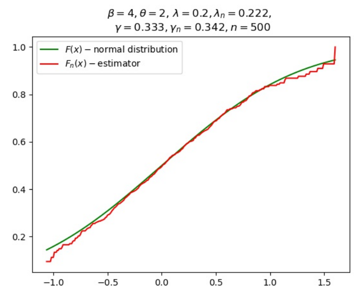

Figure 3. $\beta = 4, \quad \theta = 2, \quad n = 500$

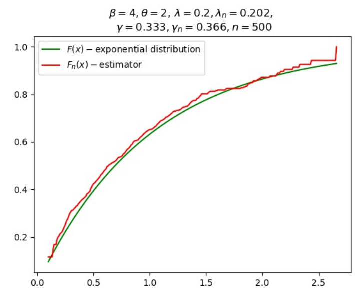

Figure 4. $\beta = 4, \quad \theta = 2, \quad n = 500$

3. Parametric estimation of exponential distribution

Now let's look at the case in model (2) $X_i \sim F(x, \alpha) = 1 - e^{-\frac{x}{\alpha}}$, $x \geqslant 0$, $\alpha > 0$, $Y_i \sim G(x, \alpha) = 1 - (1 - F(x, \alpha))^{\theta}$, $\theta > 0$ and $L_i \sim K(x, \alpha) = (N(x, \alpha))^{\beta}$, $\beta > 0$. Then in accordance with (3) we have $1 - e^{-\frac{x}{\alpha}} = 1 - \left[1 - (H(x))^{\lambda}\right]^{\gamma}$ and we get for $H(x)$ the expression $H(x) = \left(1 - e^{-\frac{x}{\alpha \cdot \gamma}}\right)^{\frac{1}{\lambda}}$. Differentiate this distribution over by x we have density of $H(x)$ as $h(x) =$

$\frac{1}{\alpha\lambda\gamma}e^{-\frac{x}{\alpha\gamma}}\left(1-e^{-\frac{x}{\alpha\gamma}}\right)^{\frac{1-\lambda}{\lambda}}$. Now we construct a maximum likelihood function for $h(x)$ using the sample from Z_i :

$$L(Z,\,\alpha)=\ln\Bigg((\alpha\lambda\gamma)^{-n}\cdot e^{-\frac{\sum_{i=1}^{n}Z_i}{\alpha\gamma}}\left(\left(1-e^{-\frac{Z_1}{\alpha\gamma}}\right)\left(1-e^{-\frac{Z_1}{\alpha\gamma}}\right)\times\;\ldots\times\right.$$

$$\left.\times\left(1-e^{-\frac{Z_n}{\alpha\gamma}}\right)\right)^{\frac{1-\lambda}{\lambda}}\Bigg)=-n\ln(\alpha\lambda\gamma)-\frac{\sum_{i=1}^{n}Z_i}{\alpha\gamma}+\frac{1-\lambda}{\lambda}\sum_{i=1}^{n}\ln\left(1-e^{-\frac{Z_i}{\alpha\gamma}}\right).$$

Now calculate $\dfrac{\partial L(Z,\,\alpha)}{\partial\alpha}$:

$$\frac{\partial L(Z,\,\alpha)}{\partial\alpha}=-\frac{n}{\alpha}+\frac{\sum_{i=1}^{n}Z_i}{\alpha^2\gamma}+\frac{1-\lambda}{\lambda}\sum_{i=1}^{n}\frac{-\frac{Z_i}{\alpha^2\gamma}e^{-\frac{Z_i}{\alpha\gamma}}}{1-e^{-\frac{Z_i}{\alpha\gamma}}}.$$

We solve maximum likelihood equation $\dfrac{\partial L(Z,\,\alpha)}{\partial\alpha}=0$ by α and get

$$\alpha=\frac{1}{\lambda\gamma n}\cdot\sum_{i=1}^{n}Z_i\cdot\left(1-\frac{1-\lambda}{(H(Z_i))^{\lambda}}\right). \tag{7}$$

From (7), instead of (H,λ,γ) of supplying appropriate estimates (H_n,λ_n,γ_n), we obtain an semiparametric estimate for the parameter α :

$$\alpha_n=\frac{1}{\lambda_n\gamma_n n}\cdot\sum_{i=1}^{n}Z_i\cdot\left(1-\frac{1-\lambda_n}{(H_n(Z_i))^{\lambda_n}}\right). \tag{8}$$

From (8), we obtain an estimate $F(x,\,\alpha_n)$ for $F(x,\,\alpha)$. It is not difficult to see that under absence of censoring from the left (i.e. $\lambda_n=1$), estimator (8) is coincides with well-known maximum likelihood estimator of parameter α of exponential distribution. Now we draw the following graphs using numerical methods, in order to compare the above estimate (6) and estimate constructed using (8). In conclusion, we can say that the estimate $F(x,\alpha_n)$ would be very good (Figures 5-8).

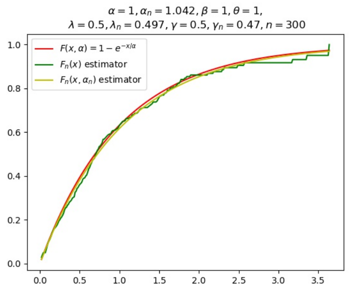

Figure 5. $\beta = 1, \quad \theta = 1, \quad n = 300$

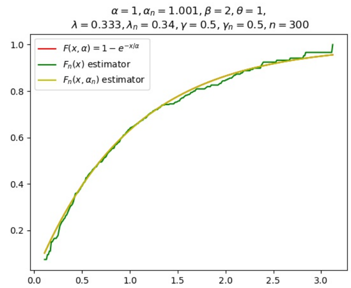

Figure 6. $\beta = 2, \quad \theta = 1, \quad n = 300$

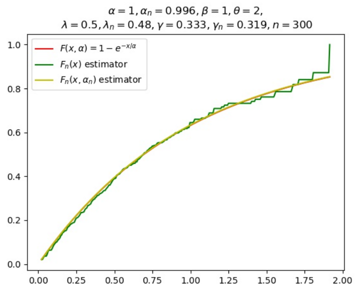

Figure 7. $\beta = 1, \quad \theta = 2, \quad n = 300$

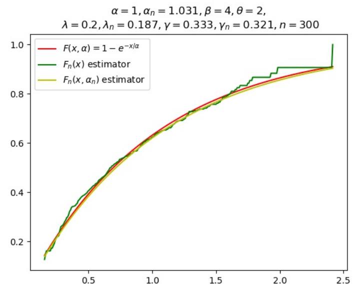

Figure 8. $\beta = 4, \quad \theta = 2, \quad n = 300$

REFERENCES

1. *Abdushukurov A. A.* Random cencorship model from both sides and independence test for it // Report of Acad. Sci. Rep. Uz. 1994. Vol. 11. P. 8-9. (In Russian).
2. *Abdushukurov A. A.* Nonparametric estimation of the distribution function based on relative risk function // Commun. Statist: Th. & Meth. 1998. Vol. 27. No. 8. P. 1991-2012.

Abdushukurov Abdurakhim Akhmedovich — Prof. (DSc) of Dpt. Applied Mathematics and Informatics of Tashkent Branch of Moscow State University, Tashkent, Uzbekistan. E-mail: *a_abdushukurov@rambler.ru*

Mansurov Dilshod Ravilovich — Doctoral Student of Navoi State Pedagogical Institute, Navoi, Uzbekistan. E-mail: *mathematicianmd@gmail.com*

TRANSITION PHENOMENON IN BRANCHING RANDOM PROCESS WITH STATE-DEPENDENT IMMIGRATIONS

J.B. Azimov

Tashkent State Transport University, Tashkent, Uzbekistan

Branching processes models are important field of stochastic processes and extensively used in various parts of natural sciences, among others in biology, epidemiology, physics, computer sciences and so forth. In this paper we consider the continuous time branching process with state-dependent immigration. Transition phenomenon in branching process with state-dependent immigration are investigated.
Keywords: *branching random process, Galton-Watson process, state-dependent immigration, transition phenomenon, generating function, uniform distribution.*

We consider the branching process with state-dependent immigration. Foster J.H. [1] consider Galton-Watson processes modified to allow immigration of particles whenever the number of particles is zero. Let $\{Z_n;\, n = 0, 1, ...\}$ be such a process. The asymptotic properties of such processes were also studied by Pakes [2] and many other authors. Asymptotic properties of the branching processes with decreasing immigration are investigated in [3, 4]. Under some additional assumptions, J.H. Foster showed that the limit law of this processes with offspring law having mean 1 is quite different from that of the original critical Galton-Watson processes $\{Z_n^*;\, n = 0, 1, ...\}$. He proved that for any $k > 0$

$$\lim_{n\to\infty} P\left(\frac{\log Z_n}{\log n} \leqslant \beta | Z_0 = k\right) = \beta$$

for $0 < \beta < 1$, i.e., the limit distribution of $\frac{\log Z_n}{\log n}$ is the uniform distribution on [0,1]. On the other hand, it is known for original Galton-Watson process that

$$\lim_{n\to\infty} P\left(\frac{Z_n^*}{n} \leqslant \frac{Bx}{2}\, | Z_n^* > 0, Z_0^* = 1\right) = 1 - e^{-x}$$

for $x \geqslant 0$, where $B = \operatorname{var} Z_1^*$.

We consider continuous time branching processes. Earlier such processes were studied in [5, 6, 7, 8].

Let $\{Z^*(t); t \geqslant 0\}$ be a time continuous Markov chain on the non-negative integers. The chain is called continuous time Markov branching process if its transition probabilities

$$P_{ij}^*(t) = P\left\{Z^*(t+s) = j | Z^*(s) = i\right\}$$

satisfy a) $P_{ij}^*(t) = \sum\limits_{j_1+\ldots+j_i=j} P_{ij_1}^*(t)\ldots P_{1j_i}^*(t)$ for all $i > 0$, $j \geqslant 0$ and $P_{00}^*(t) = 0$.

b) $P_{1i}^*(s) = \delta_{1i} + p_i s + o(s)$ as $s \downarrow 0$ for $i \geqslant 0$, p_i finite, $\{\delta_{1i}\}$ – Kronecker delta.

We use an abbreviation CTBP for such a chain. Note that $p_1 \leqslant 0$ and $p_i \geqslant 0$ for $i \neq 1$. From now on we assume $p_1 < 0$.

For CTBP we define a generating function $F(t,s)$ and an infinitesimal generating function $f(s)$ by $F(t,s) = \sum\limits_{i=0}^{\infty} P_{1i}^*(t) s^i$ and $f(s) = \sum\limits_{i=0}^{\infty} p_i s^i$.

Now, we recall some of their properties

1. $f(1) = 0$.
2. $\begin{cases} \frac{\partial}{\partial t} F(t,s) = f\left(F(t,s)\right), \\ F(0,s) = s. \end{cases}$
3. $\begin{cases} \frac{\partial}{\partial t} F(t,s) = f(s) \frac{\partial}{\partial s} F(t,s), \\ F(0,s) = s. \end{cases}$
4. $\sum\limits_{j=0}^{\infty} P_{ij}^*(t) s^j = F(t,s)^i$

The infinitesimal generator A^* of CTBP is easily obtained from the definition of CTBP and is given by $A^* = \left(a_{ij}^*\right)$, where

$$a_{ij}^* = \begin{cases} i p_{j-i+1} & if\ j \geqslant i-1, \\ 0 & \text{if } \ j < i-1. \end{cases}$$

It is known that if $a_1 = f'(1)$ is finite, then solution of (2) satisfies $F(t,1) = 1$ for all $t \geqslant 0$.

That is, $P_{ij}^*(t)$ is uniquely determined by $\{p_i\}$. Usually CTBP is called supercritical if $0 < f'(1) < \infty$, critical if $f'(1) = 0$, subcritical if $f'(1) < 0$. $P_{10}^*(t)$ satisfies

$$\lim_{t \to \infty} P_{10}^*(t) = \begin{cases} 1 & \text{in critical or subcritical case}, \\ q < 1 & \text{in supercritical case.} \end{cases}$$

For the generator A^* of CTBP described above, the state 0 is an absorbing barrier. We modify this state so that the resulting infinitesimal generator is $A = (a_{ij})$, where

$$a_{ij} = \begin{cases} ip_{j-i+1} & \text{if } j \geqslant i-1 \text{ and } i \geqslant 1, \\ q_j & \text{if } i = 0, \\ 0 & \text{otherwise.} \end{cases}$$

From now on, we assume that $q_0 < 0, \quad q_i \geqslant 0, \quad i > 0$ and $\sum\limits_{i=0}^{\infty} q_i = 0.$

The process generated by A is a continuous time analogue of "branching process with state-dependent immigration" investigated by J.H.Foster, A.G.Pakes.

Let $\{Z(t), \ 0 \leqslant t < \infty\}$ be a continuous time branching process with state-dependent immigrations. We can easily determine by $a_1 = f'(1)$ whether $\{Z(t); t < \infty\}$ is transient or recurrent.

Theorem 1. (Yamazato). $\{Z(t); \ 0 \leqslant t < \infty\}$ is recurrent if and only if $a_1 \leqslant 0$.

Now we define one more generating function $g(s) = \sum\limits_{i=0}^{\infty} q_i s^i$. As is seen, $a_1 = f'(1)$ determines whether the process is recurrent or transient. Further, does a_1 determine null-recurrence and positive-recurrence of the process? The answer is "no". We must into consideration the function $g(s)$.

Theorem 2. (Yamazato). Let $a_1 \leqslant 0$. If $-\int\limits_0^1 \frac{g(s)}{f(s)} ds < \infty$, then $\{Z(t); \ 0 \leqslant t < \infty\}$ is positive recurrent. If this integral diverges, then the process is null-recurrent.

Let $a_1 = 0$. We assume that $Z(0) = 0$.

Theorem 3. (Yamazato). If $a_1 = 0, f''(1) = b_1 < \infty$ and $g'(1) = a_2 < \infty$, then, for $0 < \beta < 1$

$$\lim_{n \to \infty} P\left(\frac{\log Z(t)}{\log t} \leqslant \beta\right) = \beta.$$

Now we consider transition phenomenon in branching process $Z(t)$.

Analogously to usual branching processes, phenomenon arising from $a \to 0, \ \ t \to \infty$, are called transition. Transition phenomenon for usual branching processes are investigated by B.A.Sevastyanov [9], C.V.Nagaev and R.I.Mukhamedhanova [10] and others. Following to B.A. Sevastyanov, generating function $f(t)$ attribute to the class $K(b_0, c_0), if b_1 > b_0, f'''(1) \leqslant c_0, \ \ 0 < b_0 < \infty, \ \ 0 \leqslant c_0 < \infty.$

Theorem 4. Let $a \to 0,\ \ t \to \infty,\ \ at \to const,\ \ a_2 < \infty$. Then

$$\lim_{t\to\infty} P\left\{\frac{\log Z(t)}{\log t} \leqslant \beta\right\} = \beta,\ \ 0 < \beta < 1$$

uniformly by all $f(t) \in K(b_0, c_0)$.

Theorem 5. Let $a \to 0+,\ \ t \to \infty$. Then

$$EZ(t) \sim \frac{b_1}{2}\,\frac{t}{\log\frac{b_1}{2a}};\quad DZ(t) \sim \frac{b_1^2}{4}\cdot\frac{t^2}{\log\frac{b_1}{2a}},\quad \left(g^{''}(1) < \infty\right)$$

uniformly by all $f(t) \in K(b_0, c_0)$.

REFERENCES

1. *Foster J.* A limit theorem for a branching process with state-dependent immigration // Ann. of Math. Statist., 1971. Vol. 42. P. 1773–1776.
2. *Pakes A.* A branching process with state-dependent immigration component // Adv. Appl. Prob., 1971. Vol. 3. № 2. P. 301–314.
3. *Mitov K, Yanev N.* Critical Galton-Watson processes with decreasing state-dependent immigrations. // J.Appl.Prob., 1984. Vol. 21. P. 22–39.
4. *Azimov J.* Asymptotic properties of the branching processes with state-Dependent immigrations // 2012 IV International Conference"Problems of Cybernetics and Informatics"(PCI). - IEEE, 2012. - P. 1–2.
5. *Yamazato M.* A some results continuous time branching processes with state-dependent immigration // Journ.of Math. Soc. of Japan., 1975. Vol. 27. № 3. P. 479–496.
6. *Formanov Sh.K., Azimov J.B.* Markov branching processes with regular varying generating function and immigration of special form // Theory Probability and Math. Statist., 2002. Vol. 65. № 3. P. 181–188.
7. *Imomov A.A.* Limit properties of transition functions of continuous-time markov branching processes // Inter.Journal of Stoch.Anal., 2014. P. 1–10.
8. *Imomov A.A.* On the limit structure of continuous-time Markov branching process // Journal of SFU, Ser. Math. and fiz., Vol. 10. № 1. P. 117–127.
9. *Sevastyanov B.* Brancching processes. Moscow: Nauka, 1971. 436 p.
10. *Nagaev S., Mukhamedkhanova R.* Transition phenomena in branching random processes with discrete time. // In "Limit theorems and stat. conclusions". Tashkent: "Fan 1966. P. 83–89.

Azimov Jakhongir — candidate of science, docent. E-mail: *azimovjb@gmail.com*

RESOURCE REQUIREMENT DISTRIBUTION EVALUATION FOR TRAFFIC OFFLOADING IN WIRELESS NETWORKS

A. Daraseliya[1], E. Sopin[1,2]

[1] *Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation*
[2] *Institute of Informatics Problems, Federal Research Center Computer Science and Control of Russian Academy of Sciences, Moscow, Russian Federation*

Today, the research community and standardization bodies seek for a systematic answer to address the effects of temporal variability in mobile traffic. One of a viable option for mitigating the impact of traffic fluctuations is offloading in unlicensed bands. In this paper, we have described model of offloading customers on unlicensed frequency range of wireless network. We obtained the resource requirement distribution of offloading customers onto unlicensed band.
Keywords: *Wireless network, queuing theory.*

Introduction

5G New Radio (NR) technology, standardized as a part of 3GPP efforts, promises drastic boost in the access rate at the last mile [1]. This is specifically the case for NR operating in millimeter wave frequency band, where a large set of resources has been made available worldwide [2]. Similarly to the respective LTE specifications, NR-U documentation has been extended to include the possibility of operation over the unlicensed bands.

In our previous study [3], we developed the model for collocated NR-U design explicitly capturing the random access behavior in unlicensed band and characterizing the NR-U customer loss probability. However, we utilized very simple M/M/K/0 queuing model to capture the specifics of resource allocation in the licensed band. In [4], we suggested a more accurate model of the service process in the licensed band. However, we described only a simple offloading strategy, where where customer offloading onto unlicensed spectrum was determined by an only insufficient amount of resources in the licensed band. In this study, we propose a strategy in which, in addition, resource-intensive customers can be initially redirected to the unlicensed range based on their "weight".

The reported study was funded by RSF, project numbers 20-71-00124.

1. Offloading Scheme

Since the coverage radii of the licensed and unlicensed bands are different we consider two types of customers. The first type of customer can be served in the licensed band only. Therefore, if there is not enough resources to service this customer, the customer is lost. The second type of customers can be potentially offloaded to the unlicensed band.

The system receives customers arrival flow with rate λ, which can be represented as $\lambda = \lambda_1 + \lambda_2$, where λ_1 and λ_2 are the arrival rates of first and second types of customers, respectively.

For "weight" based offloading strategy (Fig. 1) the pmfs of resource requirements at NR-U BS depend on the threshold R_1. The arrival flow of the second type customers is divided according to the "weight" of the customer. "Heavier" customers are initially directed to the unlicensed band with a probability $\pi_{2,1}$, and with probability $(1 - \pi_{2,1})$ "lighter" customers are directed to licensed band. Thus, the overall rate of both types of customers to the licensed band is $\lambda_1 + \lambda_2 (1 - \pi_{2,1})$.

Observe, that the second type customers arrive to the unlicensed band in two cases: (i) when the "weight" of the customer is more than a certain threshold R_1, (ii) when there are no sufficient amount of resources or servers available for a customer that has been initially routed to licensed band. The probability $\pi_{1,2}$ that the second type customer will be directed to the unlicensed spectrum is the sum of the probability $\pi_{2,1}$ that the customer was "heavy" and the probability $\pi_{2,2}$ that a "light" customer cannot be handled at the licensed band and thus offloaded to unlicensed one, i.e.,

$$\pi_2 = \pi_{2,1} + (1 - \pi_{2,1})\pi_{2,2}. \tag{1}$$

In this way, the arrival rate to the unlicensed part of NR-U BS can be calculated as $\lambda_2\pi_2^*$.

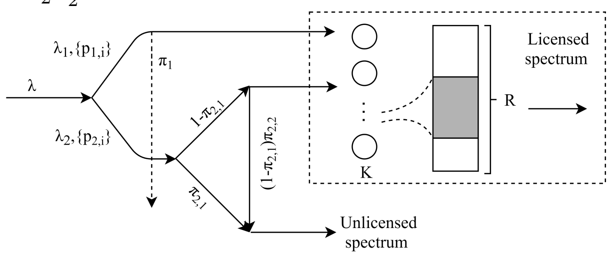

Figure 1. Illustration of the queuing model.

2. Model description

To model the customer service process in the licensed band we apply the framework of resources queuing systems [5, 6, 7]. For this purpose, consider a multiserver queuing system with $K < \infty$ servers and $R < \infty$ resource units, where K denotes the maximum number of UEs in the system, i.e., the maximum number of customers that can be simultaneously served in the licensed band. Customers of two types arrive to the system, both according to the Poisson processes with arrival rates λ_1 for the first type and λ_2 for the second one. Thus, the total arriving flow is Poisson with parameter $\lambda = \lambda_1 + \lambda_2$. The service time distribution is exponential with the rate μ.

Service process of each customer requires a server and a random amount of resources, $0 \leqslant r \leqslant R$. The distributions of resource requirements for considered customer types are given by $\{p_{l,j}\}_{j \geqslant 0}$, $l = 1, 2$, where $p_{l,j}$ is the probability that a customer of type l requires j resources. According to [5] resource-based queuing system with two flows can be analyzed as a system with one aggregated flow assuming the following

$$\tilde{p}_{j,1} = \frac{\rho_1^*}{\rho^*} p_{1,j} + \frac{\rho_2^*}{\rho^*} \tilde{p}_{j,1,2}, \tag{2}$$

where $\rho^* = \rho_1^* + \rho_2^*$, and $\rho_1^* = \lambda_1/\mu$, $\rho_2^* = \lambda_2(1 - \pi_{2,1})/\mu$.

The system operates as follows. An arriving customer is accepted to the system if at the moment of arrival there are sufficient amount of resources available. Alternatively, an arriving customer is dropped. In this case, a first type of customer is lost while the customer of the second type is being redirected to the unlicensed band. When the service time of a customer is over, it leaves the system releasing all the occupied resources.

Denote by $P_k(r)$ the stationary probability that there are k customers in the system that totally occupy r resources. According to [8], the stationary distribution is given by

$$\begin{aligned} P_k(r) &= P_0 \frac{\rho^i}{k!} \tilde{p}_{r,1}^{(k)}, \; k = 1, 2, \ldots, K, \\ P_0 &= \left(1 + \sum_{k=1}^{K} \frac{\rho_s^k}{k!} \sum_{r=0}^{R} \tilde{p}_{r,1}^{(k)} \right)^{-1}, \end{aligned} \tag{3}$$

where $\{\tilde{p}_{r,1}^{(k)}\}_{r \geqslant 0}$ is k-fold convolution of pmf $\{\tilde{p}_{r,1}\}_{r \geqslant 0}$.

The probability that the second type customer requires j resources in the licensed band is given by

$$\tilde{p}_{j,1,2} = \left(\sum_{i=0}^{R_1} p_{2,i}\right)^{-1} p_{2,j}, \ 0 \leqslant j \leqslant R_1. \tag{4}$$

The probability that there is no sufficient amount of resources in the licensed band to serve a session of the first type is

$$\pi_1 = 1 - P_0 \sum_{k=0}^{K-1} \frac{(\rho^*)^k}{k!} \sum_{r=0}^{R} \tilde{p}_{r,1}^{(k+1)}. \tag{5}$$

In the case of large values of K and R according to (5) calculations are computationally demanding. For this reason, we can adopted a recurrent computational algorithm proposed in [6]. Let us introduce an auxiliary function $G(K,R)$ as

$$G(n,r) = \sum_{i=0}^{n} \frac{(\rho^*)^i}{i!} \sum_{j=0}^{r} \tilde{p}_{j,1}^{(i)}, \ P_0 = G^{-1}(K,R). \tag{6}$$

According to it, the the probability π_1 from (5) can be rewritten as

$$\pi_1 = 1 - G^{-1}(K,R) \sum_{i=0}^{R} p_{1,i} G(K-1, R-i). \tag{7}$$

3. Resource requirement distribution of offloading customers

In this section, we will specify the probability distribution of resource requirements and the intensities of second type customer offloads to the unlicensed band.

The customer is considered “heavy” if it requires more than R_1 resources, and is thus originally routed to the unlicensed spectrum. Then the probability $\pi_{2,1}$ that the customer is “heavy”, can be calculated as follows

$$\pi_{2,1} = 1 - \sum_{i=0}^{R_1} p_{2,i}, \tag{8}$$

The probability $\pi_{2,2}$ that a “light” customer cannot be served in the licensed band and thus offloaded to the unlicensed band is calculated similarly to (5) as

$$\pi_{2,2} = 1 - P_0 \sum_{k=0}^{K-1} \frac{(\rho^*)^k}{k!} \sum_{r=0}^{R} \tilde{p}_{r,1}^{(k+1)}. \tag{9}$$

By analogy to (7), the probability $\pi_{2,2}$ (9) is given by

$$\pi_{2,2} = 1 - G^{-1}(K,R)\sum_{i=0}^{R}\tilde{p}_{i,2}G(K-1,R-i). \tag{10}$$

The probability that the customer requires j resources in the unlicensed band needs to be calculated separately for two cases: (i) when a customer is "heavy" and thus initially routed to the unlicensed band, and (ii) when a customer is first routed to the licensed band but there are not enough of resources available for its service. Reflecting on these cases we arrive at

$$\begin{aligned}&\frac{1-\pi_{2,1}}{\pi_2}p_{2,j}\left(\sum_{r=0}^{R}P_K(r)+\sum_{k=0}^{K-1}\sum_{r=R-j+1}^{R}P_k(r)\right), j \leqslant R_1,\\ &\frac{1}{\pi_2}p_{2,j}, j > R_1.\end{aligned} \tag{11}$$

After substituting the function $G(n,r)$ into (11), the probability that the customer requires j resources in the unlicensed band is

$$p_{j,2} = \begin{cases}\frac{1-\pi_{2,1}}{\pi_{1,2}}p_{2,j}\frac{G(K,R)-G(K-1,R-j)}{G(K,R)}, j \leqslant R_1,\\ \frac{1}{\pi_2}p_{2,j}, j > R_1.\end{cases} \tag{12}$$

4. Conclusion

In this paper, by utilizing the tools of queuing theory we have described resource model with the "weight" based strategy of offloading customers onto unlicensed band and obtained a resource requirement distribution of offloaded customers. In further studies, we will consider a full-fledged two-component model including service process in the unlicensed band in terms of Markov chains and an additional offloading strategy with the priority of "light" customers.

REFERENCES

1. *Parkvall S. and all.* NR: The new 5G radio access technology // IEEE Communications Standards Magazine. 2017. Vol. 1, No. 4. P. 24–30.
2. *Wang T. and all.* Spectrum analysis and regulations for 5G // 5G Mobile Communications, Springer. 2017. P. 27–50.
3. *Daraseliya A., Korshykov M., Sopin E., Moltchanov D., Koucheryavy Y., Samouylov K.* Handling Overflow Traffic in Millimeter Wave 5G NR Deployments using NR-U Technology // 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, United Kingdom. 2020. P. 1–7.

4. *Daraseliya A., Sopin E.S., Moltchanov D.A., Samouylov K.E.* Analysis of 5G NR Base stations offloading by means of NR-U technology // Informatics and Applications. 2021. Vol. 15, № 3. P. 98–111.
5. *Samouylov K., Sopin E., Vikhrova O.* Analyzing blocking probability in LTE wireless network via queuing system with finite amount of resources // International Conference on Information Technologies and Mathematical Modelling, Springer. 2015. P. 393–403.
6. *Sopin E.S., Ageev K.A., Markova E.V. and all.* Performance Analysis of M2M Traffic in LTE Network Using Queuing Systems with Random Resource Requirements // Automatic Control and Computer Sciences, Springer. 2018. Vol. 52, P. 345–353.
7. *Begishev V., Moltchanov D., Sopin E., Samuylov A., Andreev S., Koucheryavy Ye., Samouylov K.* Quantifying the impact of guard capacity on session continuity in 3GPP new radio systems // IEEE Transactions on Vehicular Technology. 2019. Vol. 68, № 12. P. 12345–12359.
8. *Naumov V.A., Samuilov K.E., Samuilov A.K.* On the total amount of resources occupied by serviced customers // Automation and Remote Control, Springer. 2016. Vol. 77, P. 1419–1427.

Daraseliya Anstasia — M.S. degree, Peoples' Friendship University of Russia (RUDN University), assistant, Department of Applied Probability and Informatics. E-mail: *daraselia-av@rudn.ru*

Eduard Sopin — Ph.D. degree, Peoples' Friendship University of Russia (RUDN University), Associate Professor, Department of Applied Probability and Informatics. E-mail: *sopin-es@rudn.ru*

LOCAL LIMIT THEOREMS FOR SOME CLASS OF MARKOV BRANCHING PROCESSES ALLOWING IMMIGRATION WITHOUT HIGHER-ORDER MOMENTS

Azam A. Imomov, Abror Kh. Meyliev

Karshi State University, Karshi, Uzbekistan

We consider the Markov branching process allowing immigration. We are dealing with the critical case, in which the immigration law has infinite mean and the offspring law have an infinite variance. We study the asymptotic properties of transition functions and their asymptotic expansion, in the case when the process is transient. The main mathematical tool of our research is the generating functions with regularly varying tails in the sense of Karamata.
Keywords: *Slow variation, Markov branching process, immigration, transition functions, state space classification, generating functions, local limit theorem.*

1. Background, aim and basic assumptions

We consider a continuous-time Markov branching process allowing immigration. Our main analytical tool is the slow variation (or more general, a regular variation) conception in the sense of Karamata. The slow variation property arises in many issues, but it usually remains rather hidden. For example, denoting by $p(n)$ the perimeter of an equilateral polygon with n sides inscribed in a circle with a diameter of length d, one can check that the function $\boldsymbol{\pi}(n) := p(n)/d$ converges to π in the sense of Archimedes, but it slowly varies at infinity in the sense of Karamata. In fact, it is known that $p(n) = dn \sin(\pi/n)$ and then it follows $\boldsymbol{\pi}(\lambda x)/\boldsymbol{\pi}(x) \to 1$ as $x \to \infty$ for each $\lambda > 0$. Thus, $\boldsymbol{\pi}(x)$ is so slowly approaching π that it can be suspected that "π is not quite constant".

Application of Karamata functions in the branching processes theory allows one to bypass severe constraints concerning existence of the higher-order moments of the infinitesimal characteristics of the process under study. Zolotarev was one of the first who demonstrated an encouraging prospect of application of the slow variation conception in the theory of Markov branching processes and has obtained principally new results on asymptote of the survival probability of the process without immigration.

We deal with the model of population growth called continuous-time Markov branching process allowing immigration (MBPI). This process can have a simple physical interpretation: a population size changes not only as a result of reproduction and disappearance of existing individuals, but also at the random stream of inbound "extraneous" individuals from outside. The population of individuals evolves as follows. Each individual existing at time $t \in \mathcal{T} := [0, +\infty)$ independently of his history and of each other for a small time interval $(t, t+\varepsilon)$ transforms into $j \in \mathbb{N}_0 \backslash \{1\}$ individuals with probability $a_j\varepsilon + o(\varepsilon)$ and, with probability $1 + a_1\varepsilon + o(\varepsilon)$, stays to live or makes evenly one descendant as $\varepsilon \downarrow 0$, where $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$ and $\mathbb{N}$ is the set of natural numbers. Here $\{a_j\}$ are the intensities of individuals transformation such that $a_j \geqslant 0$ for $j \in \mathbb{N}_0 \backslash \{1\}$ and

$$0 < a_0 < -a_1 = \sum_{j \in \mathbb{N}_0 \backslash \{1\}} a_j < \infty.$$

Independently of these processes, for each time interval $j \in \mathbb{N}$, new individuals leave the population with a probability $b_j\varepsilon + o(\varepsilon)$ and the immigration does not occur with a probability $1 + b_0\varepsilon + o(\varepsilon)$. The immigration intensities $b_j \geqslant 0$ for $j \in \mathbb{N}$ and

$$0 < -b_0 = \sum_{j \in \mathbb{N}} b_j < \infty.$$

Newly arrived individuals undergo transformation in accordance with the reproduction law generated by the intensities $\{a_j\}$, see [9, Ch. VII, Sect. 1]. Thus, the process under consideration is completely determined by infinitesimal generating functions

$$f(s) = \sum_{j \in \mathbb{N}_0} a_j s^j \quad \text{and} \quad g(s) = \sum_{j \in \mathbb{N}_0} b_j s^j \quad \text{for} \quad s \in [0,1).$$

Let $X(t)$ be the population size at a time $t \in \mathcal{T}$ in MBPI. This is a homogenous continuous-time Markov chain with the state space $\mathcal{S} \subset \mathbb{N}_0$ and transition functions

$$p_{ij}(t) := \mathbb{P}_i \{X(t) = j\} = \mathbb{P}\left\{X(t+\tau) = j \,\middle|\, X(\tau) = i\right\} \quad \tau, t \in \mathcal{T}$$

for all $i, j \in \mathcal{S}$. An appropriate probability generating function, see [4],

$$\sum_{j \in \mathcal{S}} p_{ij}(t) s^j = \Big(F(t;s)\Big)^i \exp\left\{ \int_0^t g\big(F(u;s)\big) du \right\},$$

where $F(t;s)$ is generating function of $Z(t)$ is the Markov branching process initiated by a single particle without immigration components.

It is known that the classification of the state space $\mathcal{S}$ depends on a value of the parameter $m := f'(1-)$. According to the general classification of continuous-time Markov chains, the process $X(t)$ is called *subcritical*, *critical*, and *supercritical* if respectively $m < 0$, $m = 0$ and $m > 0$, see [9, Ch VII, Sect. 2]. We consider the critical case only.

We will substantially use the approaches of the theory of regularly varying functions in the sense of Karamata, see, for instance, [2] and [8]. We recall that a function $L(x)$ is called *slowly varying at infinity* ($\mathbf{SV}_\infty$) if it is defined on $(0,\infty)$, positive and

$$\lim_{x\to\infty}\frac{L(\lambda x)}{L(x)} = 1$$

for each $\lambda > 0$. The representation theorem states that each $\mathbf{SV}_\infty$-function may be written in the form

$$L(x) = c(x)\exp\left(\int_a^x\left(\frac{\varepsilon(u)}{u}\right)du\right)$$

for some $a > 0$, where $c(x)$ is a bounded function such that $c(x) \to c > 0$ and $\varepsilon(x)$ – continuous function and $\varepsilon(x) \to 0$ as $x \to \infty$. If $c(x) \equiv c$ then $L(x)$ is said to be the *normalised* $\mathbf{SV}_\infty$. A function $V(x)$ is called *regularly varying at infinity* ($\mathbf{RV}_\infty$) with index ρ if it is defined on $(0,\infty)$ and

$$\lim_{x\to\infty}\frac{V(\lambda x)}{V(x)} = \lambda^\rho \quad \text{for all} \quad \lambda > 0.$$

This is why each $\mathbf{RV}_\infty$-function can be expressed as $V(x) = x^\rho L(x)$ for some $L(x)$ being an $\mathbf{SV}_\infty$-function.

We make the following assumptions on $f(s)$ and $g(s)$:

$$[f_\nu] \qquad f(s) = (1-s)^{1+\nu}\mathcal{L}\left(\frac{1}{1-s}\right)$$

and

$$[g_\delta] \qquad g(s) = -(1-s)^\delta\ell\left(\frac{1}{1-s}\right)$$

for all $s \in [0,1)$, where $0 < \nu, \delta < 1$ and functions $\mathcal{L}(\cdot)$, $\ell(\cdot)$ are $\mathbf{SV}_\infty$. By the criticality of our process, the assumption $[f_\nu]$ implies $2b := f''(1-) = \infty$.

If $b < \infty$ then $[f_\nu]$ holds with $\nu = 1$ and $\mathcal{L}(t) \to b$ as $t \to \infty$. Similarly, a generating function $g(s)$ of the form $[g_\delta]$ generates the law of immigrants arrival having the moment of δ-order. If $g'(1-) < \infty$ then $[g_\delta]$ holds with $\delta = 1$ and $\ell(t) \to g'(1-)$ as $t \to \infty$.

Throughout the paper $[f_\nu]$ and $[g_\delta]$ will be our Basic assumptions.

2. Main Results

Let

$$\mathsf{L}(x) := \frac{\ell(x)}{\mathcal{L}(x)},$$

and $\gamma = \delta - \nu$.

All appearances, the three cases can be divided concerning the classification of the state space $\mathcal{S}$, depending on a sign of γ. It was shown in [6], that $\mathcal{S}$ is positive-recurrent if $\gamma > 0$, and it is transient if $\gamma < 0$. The special case $\gamma = 0$ implies that $g(s) = f'(s)$ and that $\mathsf{L}(t) \to 1 + \nu$ as $t \to \infty$. And we get another population process called *Markov Q-process* instead of MBPI. We refer the reader to [3] and [5] for the details on the Markov Q-process; see also [1, Ch I, Part D, §14] and [10] for the discrete-time case.

Our theorems observe asymptotic expansion of transition function $p_{00}(t)$. Further we use a designation

$$\tau(t) := \frac{(\nu t)^{1/\nu}}{\mathcal{N}(t)},$$

where the function $\mathcal{N}(x)$ is $\mathbf{SV}_\infty$ such that

$$\mathcal{N}^{\nu}(t)\mathcal{L}\big(\tau(t)\big) \to 1 \qquad \text{as} \qquad t \to \infty.$$

Theorem 1. *Let Basic assumptions hold. If $\gamma > 0$, then*

$$\lim_{t\to\infty} p_{00}(t) = constant \neq 0.$$

More precisely there exists a $\mathbf{SV}_\infty$-function $\mathsf{L}_0(t)$, such that $\mathsf{L}_0(t)L^{-1}(t) \to 1$ and

$$-\ln p_{00}(t) = \frac{1}{\gamma}\mathsf{L}_0(1)\big(1 + o(1)\big) \qquad \text{as} \qquad t \to \infty.$$

Theorem 2. *Let Basic assumptions hold. If $\gamma < 0$, then*

$$-\big(\tau(t)\big)^{-|\gamma|} \ln p_{00}(t) = \frac{1}{|\gamma|}\mathsf{L}\big(\tau(t)\big)\big(1 + o(1)\big) \qquad \text{as} \qquad t \to \infty.$$

The above limit theorems allow us to obtain complete information about all possible states of the process under consideration. After all, it is known that the ratio limit

$$\left\{\mu_j := \lim_{t\to\infty}\frac{p_{ij}(t)}{p_{00}(t)}\right\} \qquad \text{for} \qquad i, j \in \mathcal{S}$$

is an invariant measure, which is unique up to multiplicative constants. Besides, it is known that the GF $\mathcal{M}(s) = \sum_{j\in\mathbb{N}} \mu_j s^j$ satisfies the Schroeder type functional equation; see [4]. And therefore, the asymptotic form of transition functions $p_{ij}(t)$ follows from that of $p_{00}(t)$.

REFERENCES

1. *K. B. Athreya, P. E. Ney.* Branching processes. New York: Springer, 1972.
2. *N. H. Bingham, C. M. Goldie, J. L. Teugels.* Regular Variation. Cambridge, 1987.
3. *A. A. Imomov.* On conditioned limit structure of the Markov branching process without finite second moment // Malaysian Jour. Math. Sci. 2017. Vol. 11, №3. P. 393–422.
4. *A. A. Imomov.* On long-term behavior of continuous-time Markov branching processes allowing immigration // J Siberian Fed. Univ.: Math. and Phys. 2014. Vol. 7, №4. P. 443–454.
5. *A. A. Imomov.* On Markov analogue of Q-processes with continuous time // Theory of Prob. and Math. Stat. 2012. Vol. 84. P. 57–64.
6. *J. Li, A. Chen, A. G. Pakes.* Asymptotic properties of the Markov branching process with immigration // J of Theor. Prob. 2012. Vol. 25, №1. P. 122–143.
7. *A. G. Pakes.* Revisiting conditional limit theorems for the mortal simple branching process // Bernoulli. 1999. Vol. 5, №6. P. 969–998.
8. *E. Seneta.* Regularly varying functions. Berlin: Springer, 1976.
9. *B. A. Sevastyanov.* Branching processes. Moscow: Nauka, 1971. (in Russian).

Azam A.Imomov — Doctor of Sciences, Head of Department, Faculty of Mathematics and Computer Sciences. E-mail: *imomov_azam@mail.ru*

Abror Kh.Meyliev — Researcher. E-mail: *abror_meyliyev@mail.ru*

LOCAL LIMIT THEOREM FOR THE GALTON-WATSON BRANCHING-IMMIGRATION SYSTEM WITHOUT FINITE VARIANCES

Azam A. Imomov, Erkin E. Tukhtaev

Karshi State University, Karshi, Uzbekistan

We consider the Galton-Watson Branching-Immigration System. We are dealing with the critical case, in which the immigration law has infinite mean and the offspring law have an infinite variance. We study the asymptotic properties of transition probabilities and their asymptotic expansion, in the case when a state space of considering system is transient.

Keywords: *Branching-immigration system, transition probabilities, generating functions, slow variation, local limit theorem.*

We are dealing with well-known Galton-Watson Branching-Immigration System (GWBI) describing the evolution of the number of particles of one type. An evolution of the system will occur by following scheme. An initial state is empty and the the evolutionary process begins due to the possible flow of immigrants. Let $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$ and $\mathbb{N}$ is the set of natural numbers. Each particle at any time $n \in \mathbb{N}$ produces j progeny with probability p_j, $j \in \mathbb{N}_0$ independently of each other so that $p_0 > 0$. Simultaneously into the population i immigrants possibly arrive with probability h_i, $i \in \mathbb{N}_0$ in each moment $n \in \mathbb{N}$. Particles entering the system undergo further transformation in accordance with the offspring probability $\{p_j\}$. We denote X_n the population size of GWBI in time n. Then we can write the following recursive equation:

$$X_0 = 0, \quad X_{n+1} = \xi_1 + \xi_2 + \cdots + \xi_{X_n} + \eta_{n+1} \quad \text{for any} \quad n \in \mathbb{N}, \tag{1}$$

where ξ_k and η_k are independent random variables with $\mathbb{P}\{\xi_k = j\} = p_j$ and $\mathbb{P}\{\eta_k = j\} = h_j$ for all $k \in \mathbb{N}$. Throughout the paper be assumed that $\sum_{j\in\mathbb{N}_0} h_j = 1$. Letting

$$h(s) := \sum_{j\in\mathbb{N}_0} h_j s^j \qquad \text{and} \qquad f(s) := \sum_{j\in\mathbb{N}_0} p_j s^j$$

are the probability generating functions (GFs), it follows from (1) that transition probabilities of the GWBI are

$$p_{ij} = \text{coefficient of } \; s^j \; \text{ in } \; h(s)\big(f(s)\big)^i \quad \text{for} \quad s \in [0,1).$$

It is known that the sequence $\{X_n\}$ is an homogeneous Markov chain with state space $\mathcal{S} \subset \mathbb{N}_0$ and its n-step transition probabilities

$$p_{ij}(n) := \mathbb{P}\left\{X_{n+k} = j \middle| X_k = i\right\} \quad \text{for any} \quad k \in \mathbb{N}$$

are given by GF

$$\sum_{j\in\mathcal{S}} p_{ij}(n)s^j = \bigl(f_n(s)\bigr)^i \prod_{k=0}^{n-1} h\bigl(f_k(s)\bigr) \quad \text{for any} \quad i \in \mathcal{S}$$

where $f_n(s)$ is n-fold iteration of GF $f(s)$ for $s \in [0,1)$; see [1]. Thus the transition probabilities $\{p_{ij}(n)\}$ are completely defined by the probabilities $\{p_j\}$ and $\{h_j\}$.

The above defined population process $\{X_n\}$ was first considered by Heathcote [2] in 1965. Long-time properties of the system state space $\mathcal{S}$ and existence problem and uniqueness properties of invariant measures for GWBI were investigated by Seneta [3, 4], Pakes [5, 6] and by many famous authors. In the above-mentioned articles, high-order moment conditions for offspring law $\{p_j\}$ and immigration law $\{h_j\}$ was mainly required.

In 1957 Zolotarev [7] demonstrated the encouraging prospect of application the Karamata's slow variation conception in the theory of continuous-time Markov branching processes. He obtained a fundamentally new result on the probability of population survival. Afterwards due to the SV theory principally new results were proved, for the simple Galton-Watson Process without immigration by Slack [8] and Seneta [9], for GWBI by Pakes [10, 11]; see, also [12, 13, 14, 15].

Further we consider the critical case, i.e. the mean per-capita offspring number $\sum_{j\in\mathbb{N}} jp_j = f'(1-) = 1$. Discussing this case, we proceed from the assumption that the offspring law GF $f(s)$ admits for $s \in [0,1)$ the following representation:

$$f(s) - s = (1-s)^{1+\nu}\mathcal{L}\left(\frac{1}{1-s}\right), \tag{f_ν}$$

where $0 < \nu < 1$ and $\mathcal{L}(x)$ is slowly varying at infinity (SV_∞).

And also we everywhere will consider the case that immigration GF $h(s)$ for $s \in [0,1)$ has the following form:

$$1 - h(s) = (1-s)^{\delta}\ell\left(\frac{1}{1-s}\right), \tag{h_δ}$$

where $0 < \delta < 1$ and $\ell(x)$ is SV_∞.

In our previous paper [15] we considered the case $\delta > \nu$, in which $\mathcal{S}$ is ergodic. Now we consider the case and $\delta < \nu$ which implies that $\mathcal{S}$ is transient; see [11] .

Further we use a designation

$$\tau(n) := \frac{(\nu n)^{1/\nu}}{\mathcal{N}(n)},$$

where the function $\mathcal{N}(n)$ is SV_∞ such that

$$\mathcal{N}^{\,\nu}(n)\mathcal{L}\big(\tau(n)\big) \to 1 \quad \text{as} \quad n \to \infty.$$

The following theorem is a main result of the paper. Put

$$\mathsf{L}(x) := \frac{\ell(x)}{\mathcal{L}(x)}.$$

Theorem 1. *Let conditions $[f_\nu]$, $[h_\delta]$ hold and $\gamma := \delta - \nu < 0$. Then*

$$-\big(\tau(n)\big)^{-|\gamma|} \cdot \ln p_{00}(n) = \frac{1}{|\gamma|}\mathsf{L}\big(\tau(n)\big) + \mathcal{O}\left(\frac{1}{n^{1-\delta/\nu}}\right) \quad \textit{as} \quad n \to \infty.$$

Theorem 1 allows us to obtain complete information on system state space $\mathcal{S}$. Actually, it is known that the ratio limit

$$\left\{\pi_j := \lim_{n\to\infty} \frac{p_{ij}(n)}{p_{00}(n)}\right\} \qquad \text{for} \qquad i, j \in \mathcal{S}$$

is an invariant measure, which is unique up to multiplicative constants. Besides, it is known that the GF $\mathcal{P}(s) = \sum_{j\in\mathbb{N}} \pi_j s^j$ satisfies the Schroeder type functional equation; see [14]. And therefore, the asymptotic form of transition functions $p_{ij}(n)$ follows from that of $p_{00}(n)$.

In the case of $\delta = \nu$ it follows that $\ell(u) \equiv (1+\nu)\mathcal{L}(u)\,(1+o(1))$. Therefore $\mathsf{L}(u) \to 1+\nu$ as $u \to \infty$. This case lends us another population growth system called *Q-process* instead of GWBI; see [16].

REFERENCES

1. *Pakes A. G.* Limit theorems for the simple branching process allowing immigration, I. The case of finite offspring mean // Adv. Appl. Prob. 1979. Vol. 11. P. 31–62.
2. *Heatcote C. R.* A branching process allowing immigration // J Royal Stat. Soc. 1965. Vol. 27. P. 138–143.

3. *Seneta E.* Functional equations and the Galton-Watson process // Adv. Appl. Prob. 1969. Vol. 1. P. 1–42.
4. *Seneta E.* The stationary distribution of a branching process allowing immigration: A remark on the critical case // J Royal Stat. Soc. 1968. Vol. 30, №1. P. 176–179.
5. *Pakes A. G.* On the critical Galton-Watson process with immigration // Jour. Austral. Math. Soc. 1971. Vol. 12. P. 476–482.
6. *Pakes A. G.* Branching processes with immigration // Jour. Appl. Prob. 1971. Vol. 8, №1. P. 32–42.
7. *Zolotarev V. M.* More exact statements of several theorems in the theory of branching processes // Theory Prob. and Appl. 1957. Vol. 2. P. 256–266.
8. *Slack R. S.* A branching process with mean one and possible infinite variance // Wahrscheinlichkeitstheor. und Verv. Geb. 1968. Vol. 9. P. 139–145.
9. *Seneta E.* Regularly varying functions in the theory of simple branching process // Adv. Appl. Prob. 1974. Vol. 6. P. 408–420.
10. *Pakes A. G.* Revisiting conditional limit theorems for the mortal simple branching process // Bernoulli. 1999. Vol. 5, №6. P. 969–998.
11. *Pakes A. G.* Some results for non-supercritical Galton-Watson process with immigration // Math. Biosci. 1975. Vol. 24. P. 71–92.
12. *Imomov A. A.* On a limit structure of the Galton-Watson branching processes with regularly varying generating functions // Prob. and Math. Stat. 2019. Vol. 39, №1. P. 61–73.
13. *Imomov A. A., Tukhtaev E. E.* On application of Slowly Varying Functions with Remainder in the theory of Galton-Watson Branching Process // Jour. Siber. Fed. Univ.: Math. Phys. 2019. Vol. 12, №1. P. 51–57.
14. *Imomov A. A.* On long-time behaviors of states of Galton-Watson Branching Processes allowing Immigration // J Siber. Fed. Univ.: Math. Phys. 2015. Vol. 8, №4. P. 394–405.
15. *Imomov A. A., Tukhtaev E. E.* On Asymptotic Structure of the Critical Galton-Watson Branching Processes with Infinite Variance and Allowing Immigration // In book: Applied Modeling Techniques and Data Analysis 2; Volume 8 – Big Data, Artificial Intelligence and Data Analysis, Chapter 13. 2021. P. 185–194.
16. *Imomov A. A.* Limit Theorem for the Joint Distribution in the Q-processes // Jour. Siber. Fed. Univ.: Math. Phys. 2014. Vol. 7, №3. P. 289–296.

Azam A.Imomov — Doctor of Sciences, Head of Department, Faculty of Mathematics and Computer Sciences. E-mail: *imomov_azam@mail.ru*

Erkin E.Tukhtaev — Graduate student. E-mail: *erkin.tuxtayev@mail.ru*

ON PROPERTIES OF A STRUCTURAL PARAMETER OF THE MARKOV Q-PROCESS

Azam A. Imomov[1,3], Zukhriddin A. Nazarov[2], Misliddin S. Murtazaev[3]

[1] *Karshi State University, Karshi, Uzbekistan*
[2] *National University of Uzbekistan, Tashkent, Uzbekistan*
[3] *V.I.Romanovskiy Institute of Mathematics, Uzbekistan*

The paper considers the Markov Q-processes. An unbiased estimator function for the structural parameter of this process is proposed. The asymptotic properties of variance of estimator function are investigated.

Keywords: *Markov branching processes, Markov Q-processes, structural parameter, transition probabilities, unbiased estimator.*

Estimation theory of the parameters is an interesting and important part of the progress of theory of stochastic branching processes from the point of view of statistics. Several estimators have been developed to estimate the per-capita offspring mean based on the observation of population size function. One of the most popular and reasonable functions for estimating the offspring mean is the classical Lotka-Nagaev estimator and its possible extensions; see [1, 2, 3].

Our interest in this paper is focused on the estimate of the structural parameter of the so-called Q-process.

Let $Z(t)$ be the population size at time $t \geqslant 0$ in a homogeneous continuous time Markov branching process (MBP) with transition probabilities

$$P_{ij}(t) := \mathsf{P}\left\{Z(t+\tau) = j \,|Z(\tau) = i\right\}, \quad \tau \geqslant 0,$$

for all $i, j \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers. These probabilities are equal to the i-fold convolution of $P_{1j}(t)$, i.e.

$$P_{ij}(t) = \sum_{j_1+j_2+\cdots+j_i=j} P_{1j_1}(t) \cdot P_{1j_2}(t) \cdot \ \ldots \ \cdot P_{1j_i}(t).$$

Herein transition probabilities $P_{1j}(t)$ are expressed using the local densities $\{a_j, j \in \mathbb{N}_0\}$ by relation

$$P_{1j}(\varepsilon) = \delta_{1j} + a_j\varepsilon + o(\varepsilon) \quad \text{as} \quad \varepsilon \downarrow 0, \tag{1}$$

where δ_{ij} is Kronecker's delta function and $a_j \geqslant 0$ for $j \in \mathbb{N}_0 \backslash \{1\}$ such that

$$0 < a_0 < -a_1 = \sum_{j \in \mathbb{N}_0 \backslash \{1\}} a_j < \infty;$$

see [4, pp. 11–12, 26].

Define the stochastic matrix $\{Q_{ij}(t), j \in \mathbb{N}_0\}$ by passing to the limit as follows:

$$Q_{ij}(t) := \lim_{k\to\infty} \mathsf{P}\left\{Z(t+\tau) = j \mid Z(\tau) = i,\, Z(t+\tau+k) > 0\right\}$$

for $t, \tau \geqslant 0$. Then

$$Q_{ij}(t) = \frac{jq^{j-i}}{i\beta^t} P_{ij}(t), \tag{2}$$

where $\beta = \exp\left\{\sum_{j\in\mathbb{N}} j a_j q^{j-1}\right\}$, and q is an extinction probability of the MBP $Z(t)$. Now define stochastic process $\{W(t), t \geqslant 0\}$ with transition functions

$$\mathsf{P}\left\{W(t+\tau) = j \mid W(\tau) = i\right\} = Q_{ij}(t).$$

This is so-called Markov Q-process (MQP) $\{W(t), t \geqslant 0\}$, defined as a "long-living" Markov population growth system with transition functions $\{Q_{ij}(t)\}$; see [6]. The asymptotic and structural properties of MQP were studied also in [7, 8, 9].

Using (1) and (2) the probabilities $Q_{1j}(\varepsilon)$, for $\varepsilon \downarrow 0$, can be presented as

$$Q_{1j}(\varepsilon) = \delta_{1j} + p_j\varepsilon + o(\varepsilon) \quad \text{as} \quad \varepsilon \downarrow 0$$

with transition probability densities

$$p_0 = 0, \quad p_1 = a_1 - \ln\beta < 0, \quad p_j = jq^{j-1}a_j \geqslant 0, \quad j \in \mathbb{N}\backslash\{1\}.$$

Therefore, the generating function (GF)

$$g(x) := \sum_{j\in\mathbb{N}} p_j x^j = x\left[f'(qx) - f'(q)\right]$$

completely determines the MQP, where $f(x)$ is the infinitesimal GF generating the MQP $Z(t)$, that is

$$f(x) = \sum_{j\in\mathbb{N}_0} a_j x^j.$$

In this notation $\beta := \exp\{f'(q)\}$ and $f(q) = 0$; see [6].

It is known that the regulating parameter for the MBP is the value $f'(1)$ and, according to the trajectory property, three types of processes are distinguished, characterized by its value: $f'(1) < 0$, $f'(1) = 0$ and $f'(1) > 0$. Note that the evolution of the MQP is controlled (regulated) by an essentially positive parameter β and it is known that $\beta = 1$ for $f'(1) = 0$ and $\beta < 1$ if $f'(1) \neq 0$; see [6] and the literature therein. Thus, there are two types of MQP, depending on the values of the parameter β. The above is clearly described in the following local limit theorem.

Theorem A [5]. *Let* $b := g'(1) < \infty$.
1) *If* $\beta = 1$, *then*

$$t^2 Q_{11}(t) \to \frac{2}{bf(0)} \quad \text{as} \quad t \to \infty.$$

2) *If* $\beta < 1$, *then*

$$Q_{11}(t) \to \frac{|\ln \beta|\, A}{f(0)} \quad \text{as} \quad t \to \infty,$$

where the constant is

$$A = q \exp\left\{ \int_0^q \left[\frac{1}{s-q} - \frac{f'(q)}{f(s)} \right] ds \right\}.$$

It is easy to verify that for $q = 1$ the positive constant A is the Sevastyanov constant from theory of subcritical MBPs ($f'(1) < 0$); see [5].

Let us now consider the following GF:

$$G_i(t;x) := \mathsf{E}_i x^{W(t)} = \mathsf{E}\left[x^{W(t)} \middle| W(0) = i \right] = \sum_{j \in \mathbb{N}} Q_{ij}(t) x^j.$$

It was proved in [6] that

$$G_i(t;x) = x \left[\frac{\Phi(t;qx)}{q} \right]^{i-1} \cdot \exp\left\{ \int\limits_0^t b\left(\frac{\Phi(\tau;qx)}{q} \right) d\tau \right\}, \tag{3}$$

where

$$\Phi(t;x) = \sum\nolimits_{j \in \mathbb{N}_0} P_{1j}(t) x^j$$

and

$$b(x) = \frac{g(x)}{x} = f'(qx) - f'(q).$$

Differentiating at the point $x = 1$, it follows from (3) that

$$\begin{aligned} \mathsf{E}_i W(t) &= (i-1)\,\beta^t + \mathsf{E}_1 W(t) \\ &= \begin{cases} (i-1)\,\beta^t + bt + 1, & \text{if} \quad \beta = 1, \\ (i-1)\,\beta^t + 1 + \gamma(1-\beta^t), & \text{if} \quad \beta < 1, \end{cases} \end{aligned}$$

and

$$\mathsf{Var}_i W(t) = \begin{cases} bti, & \text{if} \quad \beta = 1, \\ \left[\gamma + (i-1)\,(1+\gamma)\,\beta^t\right](1-\beta^t), & \text{if} \quad \beta < 1, \end{cases}$$

where $\mathsf{Var}_i W(t) = \mathsf{Var}\left[W(t) \,|\, W(0) = i\right]$ and $\gamma = b/|\ln \beta|$.

In view of all of the above, let us pose the problem of estimating the parameter β from the observed values of $W(t)$. It is obvious from (3) that

$$\mathsf{E}\left[x^{W(t+1)} \middle| W(t) \right] = \left[\frac{\Phi(1;qx)}{q} \right]^{W(t)-1} G(1;x),$$

where $G(t;x) := G_1(t;x)$. Hence we have

$$\mathsf{E}\left[W(t+1)\,|W(t)\right] = [W(t)-1]\cdot\beta + \mathsf{E}_1 W(1).$$

The last relation allows us to write down the equation

$$W(t+1) = [W(t)-1]\cdot\beta + W(1) + \varepsilon(t)$$

with an error $\varepsilon(t)$, having a zero mean $\mathsf{E}\varepsilon(t) = 0$. Based on this equation, we propose the following estimator for β with a known $\mathsf{E}_1 W(1)$:

$$\widehat{\beta}(t) = \frac{W(t+1) - \mathsf{E}_1 W(1)}{W(t)-1}, \qquad t > 1.$$

The estimator $\widehat{\beta}(t)$ is unbiased for the parameter β. Indeed, according to the Law of total expectation, taking into account the homogeneity of MQP and (4), we have

$$\begin{aligned}
\mathsf{E}\widehat{\beta}(t) &= \sum_{j\in\mathbb{N}} \mathsf{P}\left\{W(t)=j\right\} \mathsf{E}\left[\frac{W(t+1)-\mathsf{E}_1 W(1)}{W(t)-1}\,\middle|\, W(t)=j\right] \\
&= \sum_{j\in\mathbb{N}} \frac{\mathsf{E}_j W(1) - \mathsf{E}_1 W(1)}{j-1} Q_{1j}(t) = \beta \sum_{j\in\mathbb{N}} Q_{1j}(t) = \beta.
\end{aligned}$$

The following theorems characterize the properties of the variance of the estimator $\widehat{\beta}(t)$.

Theorem 1. *Let $b < \infty$. If $\beta = 1$, then*

$$\frac{t}{2}\cdot \mathsf{Var}\widehat{\beta}(t) = 1 + \mathcal{O}\left(\frac{\ln^2 t}{t}\right) \quad \text{as} \quad t \to \infty.$$

Theorem 2. *Let $b < \infty$. If $\beta < 1$, then*

$$\mathsf{Var}\widehat{\beta}(t) = \mathcal{O}\left(1\right) \quad \text{as} \quad t \to \infty.$$

Theorem 3. *Let $b < \infty$ and $\beta = 1$. Then*

$$\mathsf{P}\left\{\left(\widehat{\beta}(t) - 1\right) t < x\right\} \to D(x) \quad \text{as} \quad t \to \infty,$$

where

$$\int_{-\infty}^{+\infty} e^{i\theta x} dD(x) = \int_{-\infty}^{+\infty} x e^{-x + 2i\theta/x} dx.$$

The method used in the proof of the above theorems is based on arguments from the paper [2].

REFERENCES

1. *Badalbaev I.S., Muxitdinov A.A.* Statisticheskie zadachi mnogotipnix vetvyashixsya protsessov // Tashkent: Fan, 1990. (in Russian)
2. *Nagaev A.B.* Ob otsenke srednego chisla neposredstvennix potomkov chastitsi v vetvyashemsya sluchaynom protsesse // Teoriya veroyatnostey i primen. 1967. Vol. 67, №. 2. P. 363–369. (in Russian)
3. *Gaoa Z., Qiu L.* Large Deviations for Lotka-Nagaev Estimatorof a Randomly Indexed Branching Process // Filomat. 2018. Vol. 32, № 17. P. 5803–5808.
4. *Sevastyanov B.A.* Vetvyashiesya protsessi // –M.: Nauka, 1971. (in Russian)
5. *Imomov A.A.* A differential analogue of the main lemma of the theory of Markov branching processes and its applications // Ukrainian Mathematical Journal. 2005. Vol. 57, № 2. P. 307–315.
6. *Imomov A.A.* On Markov continuous time analogue of Q-processes // Journal Theory of Probability and Mathematical Statistics. 2012. Vol. 84. P. 57–64.
7. *Imomov A.A., Meyliev A.Kh.* On asymptotic structure of continuous-time Markov branching processes allowing immigration without higher-order moments // Ufa Math. Jour. 2021. Vol. 13, № 1. P. 137–147.
8. *Imomov A.A., Meyliev A.Kh.* Ob asimptoticheskoy strukture nekritechiskix markovskix vetvyashixsya sluchaynix protsessov s neprerivnim vremenem // Vestnik Tomskogo gosudarstvennogo universiteta. Matematika i mexanika. 2021. Vol. 69. P. 22–36. (in Russian)
9. *Imomov A.A.* On estimation of the convergence rate to invariant measures in Markov branching processes with possibly infinite variance and immigration // Journ. Siberian Federal University: Math. and Fiz. 2021. Vol. 14, № 5. P. 573–583.

Azam Imomov — Doctor of Sciences, Head of Department, Faculty of Mathematics and Computer Sciences. E-mail: *imomov_azam@mail.ru*

Zukhriddin Nazarov — Graduate student, National University of Uzbekistan, Faculty of Mathematics. E-mail: *zuhrov13@gmail.com*

Misliddin Murtazaev — Graduate student, V.I. Romanovskiy Institute of Mathematics, Stochastic analysis laboratory. E-mail: *misliddin1991@mail.ru*

EMPIRICAL KAC PROCESSES UNDER GENERAL RANDOM CENSORSHIP MODEL

G. S. Sayfulloyeva

Navoi State Pedagogical Institute, Navoi, Uzbekistan

In this paper we consider a general random censorship model and prove approximation results for empirical Kac processes. This model includes such important special case as random censorship on the right. These results are used also for estimating of characteristic functions in random censorship model on the right.
Keywords: *censored data, empirical estimates, Kac estimate, strong approximation, Gaussian process, characteristic function.*

Introduction

Following of [1, 2, 3] we define a general random censorship model in the following way. Let Z be a real random variable (r.v.) with distribution function (d.f.) $H(x) = P(Z \leqslant x)$, $x \in \mathbb{R}$. For a fixed integer $k \geqslant 1$ let $A^{(1)}, ..., A^{(k)}$ be pairwise disjoint random events, and define the subdistribution functions $H(x;i) = P(Z \leqslant x, A^{(i)}), i \in \Im = \{1, ..., k\}$. Suppose that when observing Z we are interested in the joint behaviour of the pairs $(Z, A^{(i)}), i \in \Im$. Let $\{(Z_j, A_j^{(1)}, ..., A_j^{(k)}), j \geqslant 1\}$ be a sequence of independent replicas of the $(Z, A^{(1)}, ..., A^{(k)})$defined on some probability space $\{\Omega, A, P\}$. We assume throughout that the functions $H(x), H(x;1), ..., H(x;k)$ are continuous. Let $H_n(x)$ denote the ordinary empirical d.f. of $Z_1, ..., Z_n$ and introduce the empirical sub d.f. $H_n(x;i)$, $i \in \Im$

$$H_n(x;i) = \frac{1}{n}\sum_{j=1}^{n} \delta_j^{(i)} I(Z_j \leqslant x), (x;i) \in \overline{\mathbb{R}} \times \Im,$$

where $\overline{\mathbb{R}} = [-\infty; \infty]$, $\delta_j^{(i)} = I(A_j^{(i)})$ is an indicator of event $A_j^{(i)}$ and

$$H_n(x;1) + ... + H_n(x;k) = \frac{1}{n}\sum_{j=1}^{n} I(Z_j \leqslant x) = H_n(x), x \in \overline{\mathbb{R}},$$

is the ordinary empirical d.f.. Properties of many biometrical estimates depends on limit behaviours of proposed empirical statistics. The following results are a straightforward consequences of exponential inequality of Dvoretzky-Kiefer-Wolfowitz with exactly constant $D = 2$ from [5, 6] :

For all $n = 1, 2, ...$ and $\varepsilon > 0$:

$$P\left(\sup_{|x|<\infty}\left|H_n(x) - H(x)\right| > \left(\frac{(1+\varepsilon)}{2}\cdot\frac{\log n}{n}\right)^{1/2}\right) \leqslant 2n^{-(1+\varepsilon)}, \quad (1)$$

and

$$P\Big(\sup_{|x|<\infty}\left|H_n(x;i) - H(x;i)\right| > 2\big(\frac{(1+\varepsilon)}{2}\frac{\log n}{n}\big)^{1/2}\Big) \leqslant 4n^{-(1+\varepsilon)}. \quad (2)$$

A crucial role is played the vector-valued empirical process $\Big\{a_n(t) = \Big(a_n^{(0)}(t_0), a_n^{(1)}(t_1), ..., a_n^{(k)}(t_k)\Big), t = (t_0, ..., t_k) \in \overline{\mathbb{R}}^{k+1}\Big\}$; where $a_n^{(0)}(x) = \sqrt{n}(H_n(x) - H(x))$, $a_n^{(i)}(x) = \sqrt{n}(H_n(x;i) - H(x;i))$, $i \in \Im$.

1. Kac processes under general censoring

Following of [5] we introduce the modified empirical d.f. of Kac by the following way. Along with sequence $\{Z_j, j \geqslant 1\}$ on a probability space $\{\Omega, \mathcal{A}, P\}$ consider also a sequence $\{\nu_n, \quad n \geqslant 1\}$ of r.v.-s having Poisson distribution with parameter $E\nu_n = n, \quad n = 1, 2, ...$ Assume throughout that the two sequences $\{Z_j, j \geqslant 1\}$ and $\{\nu_n, n \geqslant 1\}$ are independent. Kac's empirical d.f. is

$$H_n^*(x) = \begin{cases} \frac{1}{n}\sum_{j=1}^{\nu_n} I(Z_j \leqslant x), & if \quad \nu_n \geqslant 1 \quad a.s., \\ 0, \qquad \nu_n = 0 \quad a.s., \end{cases}$$

while the empirical sub-d.f. one is

$$H_n^*(x;i) = \begin{cases} c\frac{1}{n}\sum_{j=1}^{\nu_n} I\big(Z_j \leqslant x, A_j^{(i)}\big), \quad i \in \Im, \quad if \quad \nu_n \geqslant 1 \quad a.s., \\ 0, \quad i \in \Im \quad if \quad \nu_n = 0 \quad a.s., \end{cases}$$

with $H_n^*(x;1) + ... + H_n^*(x;k) = H_n^*(x)$ for all $x \in \overline{\mathbb{R}}$. Here we suppose that sequence $\{\nu_n, \quad n \geqslant 1\}$ is independent of random vectors $\big\{\big(Z_j, \delta_j^{(1)}, ..., \delta_j^{(k)}\big), j \geqslant 1\big\}$, where $\delta_j^{(i)} = I(A_j^{(i)})$. Note that statistics $H_n^*(x;i)$ (consequently also $H_n^*(x)$) are unbiased estimators of $H(x;i), \quad i \in \Im$ (consequently also of $H(x)$):

$$E(H_n^*(x;i)) = \frac{1}{n}E\Bigg\{\sum_{m=1}^{\infty} E\Bigg[\sum_{k=1}^{n}\delta_k^{(i)}\cdot I(Z_k \leqslant x)\Bigg], \nu_n = m\Bigg\} =$$
$$= \frac{1}{n}E\Bigg\{\sum_{m=1}^{\infty} E\Bigg[\sum_{k=1}^{n}\delta_k^{(i)}\cdot I(Z_k \leqslant x)/\nu_n = m\Bigg]\cdot P(\nu_n = m)\Bigg\} =$$
$$= \frac{1}{n}\sum_{m=1}^{\infty} H(x;i)mP(\nu_n = m) = \frac{1}{n}H(x;i)\sum_{m=1}^{\infty} m\cdot\frac{n^m e^{-n}}{m!} =$$
$$= H(x;i)e^{-n}\sum_{m=0}^{\infty}\frac{n^m}{m!} = H(x;i), \quad (x;i)\in\overline{\mathbb{R}}\times\Im.$$

Consequently,

$$E\big[H_n^*(x)\big] = \sum_{i=1}^{k} E\big[H_n^*(x;i)\big] = \sum_{i=1}^{k} H(x;i) = H(x), \quad x\in\overline{\mathbb{R}}.$$

Let's define $a_n^{(i)*}(x) = \sqrt{n}\big(H_n^*(x;i) - H(x;i)\big), \quad i\in\Im$ and $a_n^{(0)*}(x) = \sqrt{n}\big(H_n^*(x) - H(x)\big)$ the empirical Kac processes.

Theorem 1. *If the underlying probability space $\{\Omega,\mathcal{A},P\}$ is rich enough, then one can define $k+1$ sequences of Gaussian processes $W_n^{(0)}(x), W_n^{(1)}(x), ..., W_n^{(k)}(x)$ such that for $a_n^*(t) = \big(a_n^{(0)*}(t_0), a_n^{(1)*}(t_1), \ldots, a_n^{(k)*}(t_k)\big)$ and $W_n^*(t) = \big(W_n^{(0)}(t_0), W_n^{(1)}(t_1), ..., W_n^{(k)}(t_k)\big)$, $t = (t_0, t_1, ..., t_k)$, we have*

$$P\Bigg\{\sup_{t\in\overline{\mathbb{R}}^{k+1}}\Bigg\|a_n^*(t) - W_n^*(t)\Bigg\|^{(k+1)} > C^* n^{-\frac{1}{2}}\log n\Bigg\} \leqslant K^* n^{-r}, \tag{3}$$

where $r\geqslant 2$ is an arbitrary integer, $C^* = C^*(r)$- depends only on r and K^* is an absolute constant. Moreover, $W_n^*(t)$ itself is a $(k+1)$- dimensional vector-valued Gaussian process with expectation $EW^{(i)}(x) = 0$, $(x,i)\in\overline{\mathbb{R}}\times\overline{\Im}$ and for any $i,j\in\Im, \quad i\neq j, \quad x,y\in\overline{\mathbb{R}}$:

$$\begin{aligned} &EW_n^{(0)}(x)W_n^{(0)}(y) = \min\big\{H(x), H(y)\big\},\\ &EW_n^{(i)}(x)W_n^{(j)}(y) = \min\big\{H(x;i), H(y;j)\big\},\\ &EW_n^{(i)}(x)W_n^{(0)}(y) = \min\big\{H(x;i), H(y)\big\}. \end{aligned} \tag{4}$$

In so far as $\lim\limits_{x\uparrow+\infty} H_n^*(x) = H_n^*(+\infty) = \frac{\nu_n}{n}$, then by Stirlings formula

$$P(\nu_n = n) = P\big(H_n^*(+\infty) = 1\big) = \frac{n^n e^{-n}}{n!} = \frac{1}{\sqrt{2\pi n}}(1 + o(1)), \quad n\to\infty,$$

and

$$P\big(H_n^*(+\infty) > 1\big) = P(\nu_n > n) = \sum_{k=n+1}^{\infty}\frac{n^k e^{-n}}{k!} = o(1), \quad n\to\infty.$$

Thus $H_n^*(x)$ with positive probability to be greater than 1. In order to avoid these undesirable properties, we propose following modifications of Kac statistics:

$$\begin{aligned} \widetilde{H}_n(x) &= 1-\big(1-H_n^*(x)\big)I\big(H_n^*(x)<1\big), \quad x\in\overline{\mathbb{R}}, \\ \widetilde{H}_n(x;i) &= 1-\big(1-H_n^*(x;i)\big)I\big(H_n^*(x;i)<1\big), \quad (x;i)\in\overline{\mathbb{R}}\times\Im. \end{aligned} \tag{5}$$

The following inequalities are useful in investigating of Kac processes.

Theorem 2. *Let $\{\nu_n, n\geqslant 1\}$ be a sequence of Poisson r.v.-s with $E\nu_n = n$. Then for any $\varepsilon>0$ such that*

$$\frac{n}{\log n}\geqslant\frac{\varepsilon}{8(1+\frac{e}{3})^2}, \quad e=\exp(1), \tag{6}$$

we have

$$P\bigg(\big|\nu_n-n\big|>\frac{1}{2}\Big(\frac{\varepsilon}{2}n\log n\Big)^{\frac{1}{2}}\bigg)\leqslant 2n^{-\varepsilon w}, \tag{7}$$

$$P\bigg(\sup_{|x|<\infty}\big|H_n^*(x;i)-H(x;i)\big|>2\Big(\frac{\varepsilon\log n}{2n}\Big)^{\frac{1}{2}}\bigg)\leqslant 4n^{-4\varepsilon w}, \quad i\in\Im, \tag{8}$$

$$P\bigg(\sup_{|x|<\infty}\big|\widetilde{H}_n(x;i)-H(x;i)\big|>2\Big(\frac{\varepsilon\log n}{2n}\Big)^{\frac{1}{2}}\bigg)\leqslant 4n^{-4\varepsilon w}, \quad i\in\Im, \tag{9}$$

where $w=\big[16(1+\frac{e}{3})\big]^{-1}$. Let $\widetilde{a}_n(t)=\Big(\widetilde{a}_n^{(0)}(t_0),\widetilde{a}_n^{(1)}(t_1),\ldots,\widetilde{a}_n^{(k)}(t_k)\Big)$, where $\widetilde{a}_n^{(0)}(x) = \sqrt{n}\left(\widetilde{H}_n(x)-H(x)\right)$, $\widetilde{a}_n^{(i)}(x) = \sqrt{n}\left(\widetilde{H}_n(x;i)-H(x;i)\right)$, $(x;i)\in\overline{\mathbb{R}}\times\Im$.

We shall prove an approximation theorem of the vector-valued modified empirical Kac process $\widetilde{a}_n(t)$ by the appropriate Gaussian vector-valued process $W_n^*(t), t\in\overline{\mathbb{R}}^{k+1}$ from theorem 2.

Theorem 3. *Let $\{T_n, n\geqslant 1\}$ be a numerical sequence satisfying, for each n, the condition $T_n<T_H=\inf\{x: H(x)=1\}\leqslant\infty$ such that*

$$\min_{i\in\Im}\left\{P(A^{(i)})-H(T_n,i)\right\}\geqslant 1-H(T_n)\geqslant 2\left(\frac{r\log n}{2wn}\right)^{1/2}. \tag{10}$$

If for any $\varepsilon>0$ condition (6) *hold, then on a probability space of theorem 2 one can define $k+1$ sequences of mean zero Gaussian processes*

$W_n^{(0)}(x), W_n^{(1)}(x), ..., W_n^{(k)}(x)$ *with the covariance structure* (8) *such that for* $\widetilde{a}_n(t)$ *and* $W_n^*(t) = \left(W_n^{(0)}(t_0), W_n^{(1)}(t_1), ..., W_n^{(k)}(t_k)\right)$ *we have*

$$P\left\{\sup_{t\in(-\infty;T_n]^{(k+1)}} \|\widetilde{a}_n(t) - W_n^*(t)\|^{(k+1)} > \widetilde{C}n^{\frac{1}{2}}\log n\right\} \leqslant \widetilde{K}n^{-\beta}, \tag{11}$$

where $\widetilde{K}$ *is an absolute constant,* $\widetilde{C} = \widetilde{C}(\varepsilon)$ *and* $\beta = \min(r, \varepsilon w)$ *for any* $\varepsilon > 0$.

2. Estimation of characteristic function under random right censoring

Let $X_1, X_2, ...$ be independent identically distributed r.v.-s with common continuous d.f. F. We interpret them as an infinite sample of the random lifetime X. Another sequence independent of $\{X_j\}$, of independent and identically distributed r.v.-s $Y_1, Y_2, ...$ with common continuous d.f. G censors on the right the preceding one, so that the observations available at the n-th stage consist of the pairs $\{(Z_j, \delta_j), 1 \leqslant j \leqslant n\} = \mathbb{C}^{(n)}$, where $Z_j = \min(X_j, Y_j)$ and δ_j is the indicator of the event $A_j = \{Z_j = X_j\} = \{X_j \leqslant Y_j\}$. Let

$$C(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

is the characteristic function of d.f. F. The problem consist in estimating of d.f. F from censored sample $\mathbb{C}^{(n)}$. In a number of situations it is more desirable to aim at estimating $C(t)$ rather that F. We consider estimator for $C(t)$ in this model as Fourier-Stieltjes transform of estimator $F_n(x) = 1 - S_{1n}(x) = 1 - \exp\left(-\Lambda_n^{(1)}(x)\right)$:

$$C_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x), \quad t \in \mathbb{R}.$$

Using the estimation above it follows that as $n \to \infty$

$$\sup_{x\leqslant T_n} |F_n(x) - F(x)| \stackrel{a.s.}{=} O\left(b_n^2\left(\frac{\log n}{n}\right)^{\frac{1}{2}}\right), \tag{12}$$

where $b_n^{-1} = 1 - H(T_n)$. Now from (12) also follows that as $n \to \infty$

$$1 - F_n(T_n) \stackrel{a.s.}{=} O(1 - F(T_n)), \quad F_n(-T_n) \stackrel{a.s.}{=} O(F(-T_n)). \tag{13}$$

It is trivial that $\Delta_n(\tau) \xrightarrow{a.s.} 0$ as $n \to \infty$ for any $\tau < \infty$, where $\Delta_n(\tau) = \sup_{|t| \leqslant \tau} |C_n(t) - C(t)|$. Consider quantity $\Delta_n(\tau_n)$ for some special choosed numerical sequence τ_n tending to $+\infty$ as $n \to \infty$. In the following theorem we prove result of uniform convergence for the empirical characteristic function.

Theorem 4. *Let $\{\tau_n, n \geqslant 1\}$ be a numerical sequence tending to $+\infty$ slowly as $n \to \infty$. Then, $\Delta_n(\tau_n) \xrightarrow{a.s.} 0$ as $n \to \infty$.*

3. Conclusion

In conclusion, this theme is such important in stoxastic analysis. Our results uses strong approximation theory and build optimal approximation rates. Cumulative hazard processes also investigated in a similar manner in the general setting. These results can also be used also for estimating of characteristic functions in random censorship model on the right.

REFERENCES

1. *Abdushukurov A.A* Statistics of incomplete observations. University Press. Tashkent, 2009. 269 p. (in Russian).
2. *Burke M.D., Csörgő S., Horváth L.* Strong approximations of some biometric estimates under random censorship // Z. Wahrschein. Verw. Gebiete. 1981. Vol. 56. P. 87–112.
3. *Burke M.D., Csörgő S., Horváth L.* A correction to and improvement of "Strong approximations of some biometric estimates under random censorship" // Probab.Th.Ret.Fields. 1988. Vol. 79. P. 51–57.
4. *Csörgő S.* Strong approximation of empirical Kac processes // Carleton Math. Lect. Note. 1980. Vol. 26. P. 71–86.
5. *Dvoretzky A., Kiefer J., Wolfowitz J.* Asymptotic minimax character of the sample distribution function and of the multinomial estimator // Ann. Math. Statist. 1956. Vol. 27. P. 642–669.
6. *Massart P.* The tight constant in the Dvoretzky-Kiefer-Wolfowits inequality // Ann.Probab. 1990. Vol. 18, № 3. P. 1269–1283.

Sayfulloyeva Gulnoz — Navoi State Pedagogical Institute, Navoi, Uzbekistan.
E-mail: *sayfullayevagulnoz@gmail.com*

NEURAL NETWORKS ARCHITECTURE APPLIED TO FPGA

E. Solis Romeu, D. Shashev

Tomsk State University, Tomsk, Russian Federation

Field Programmable Field Arrays distinguish themselves from other integrated circuits, by their reconfigurability. This, along with their parallelism opens possibilities for implementing artificial neural networks with a high efficiency regarding cost and energy use. Research conducted on the implementation of Neural Network Architectures like VGG16 and AlexaNet in FPGAs can result in systems that are faster and more power efficient than GPU implementations.
Keywords: *FPGA, ASIC, Convolutional Neural Networks, AlexaNet, VGG16, Winograd, Matrix multiplication*

Introduction

FPGA stands for Field Programmable Field Arrays. It is a device made out of semiconductors that can be configured after being manufactured. This characteristic gives it the advantage of allowing flexibility in its design, which saves time and costs at the moment of introducing changes to it. FPGAs consist of programmable logic elements that are connected one to another, as well as memory blocks. They stand in contrast to ASICs (Application Specific Integrated Circuits) and ASSPs (Application Specific Standard Products). ASICs as their name implies, have a specific customized particular use. This means that they are designed with a particular functionality in mind. ASICs usually are sold to a specific user and ASSPs, while also having a specific use, are sold to many users. The three are integrated circuits, with the main difference being that FPGAs are reconfigurable while ASICs and ASSPs are not [1].

1. Neural Networks

Neural Networks are algorithms that mimic the functionality of human neural networks to create systems capable of analysing data. They can be used for Artificial Intelligence tasks like image recognition and video analysis. These networks are composed of at least three node layers: the input layer, the hidden layer and the output layer. Each node is a neuron that uses an activation function. An activation function is used to determine

Supported by FRBR, project number 19-29-06078.

the output of the node, given the inputs coming from previous layers and the weights assigned to them. The design of neural networks requires forward and backward processes. The forward process focuses on the amount of layers and the forward flow of the input and output information, while the backward process focuses on defining the loss function to calculate the gap between the prediction value and the actual values of the data samples during the training phase. The result of the backward process is used to alter the weights given to each node to make the final output more accurate. When the system runs a forward and backward process it is known as an epoch. Neural Networks require hundreds of epochs of training to reach high levels of accuracy [2].

2. Why Use a FPGAs for Designing Neural Networks?

The idea of creating a semi custom high performance system for neural networks that is able to outperform conventional processors is the main reason behind the efforts to develop these systems. The results so far have been that FPGAs are outperformed by ASICs. Nevertheless the structure of a FPGA presents an alternative to the software inside a general purpose processor, thus providing the possibility for delivering a superior performance in relation to cost and energy consumption on specific applications. Another apparent alternative would be the use of ASIC systems, nevertheless they have disadvantages over FGPAs.

Two important reasons exist as to why FPGAs are a better alternative than ASICs for developing systems capable of beating general purpose processors in performance. The first one is that when developing ASIC neurocomputers with a significant amount of flexibility, the final result is a system with a structure that closely resembles a FPGA, while never reaching the flexibility that a FPGA can attain. This is important because neural networks require to be easily reconfigured since they are meant to be used for different applications. The second fact is that the design and production of custom neurocomputers has a limited user base, thus the development of software for ASIC neurocomputers will lag behind FPGAs that will be more widely used [3].

3. Parallelism

Other reason for the drive towards the use of FPGA for creating neural networks systems are the perceived abilities that this technology has in the realm of parallelism. Parallelism is the feature that allows one to run two or more computational processes at the same time. FPGAs are more suitable for parallelism than ASICs systems. These capabilities are suitable for working with neural networks. To better understand this, we can take a look at different types of parallelism applicable to this topic. The first one is Training parallelism, this refers to the ability of running different training sessions for the neural network at the same time. Node parallelism refers to the idea of going through the nodes of each layer of the neural network in parallel, and this could be a great advantage for a FPGA system. Nevertheless it is necessary to take into account that neural networks can have up to a million nodes, and a system with that high amount will face limitations that will not allow it to reach those high levels of parallelism. Another promise of FPGA systems is the idea of computing the layers of the neural network in a parallel way [3].

4. Implementation with Winograd Algorithm

In the article Evaluating Fast Algorithms for Convolutional Neural Network, The authors evaluated the performance of two FPGA devices (Xilinx ZC760 and ZCU102) on which they used the Winograd algorithm for matrix multiplication as a way of making more efficient the implementation of two CNN architectures (AlexaNet and VGG16). The Winograd algorithm speeds the convolution process by simplifying the matrix multiplication. This plays an important role in convolutional algorithms since they require the multiplication of matrix filters in each layer of the neural network. During the experiment, other two FPGAs (VX485T and GSD8) were used as a control group without the Winograd algorithm being implemented in them. All models used the AlexaNet Architecture for CNN. [4] The results show that the Winograd implementation improved the average convolution performance from 61.6 Gop/s to 1006.4 Gop/s, while the CNN average performance rose from 72.4 Gop/s to 854.6 Gop/s. The implementation of the modified algorithms resulted in the use of less resources; this can be seen in the energy efficiency rising from 3.79 Gop/s/W to 36.2 Gop/s/w.

5. Conclusion

The advantages of FPGAs over ASICs and ASSPs is their reconfigurability, which opens the possibility for their flexible application by Artificial Intelligence specialists. Processors have the advantage of using their power to deliver systems with a higher analytical capacity as they are more accurate, but the advantage of the FPGA comes from delivering a more efficient use of resources (both computing power and energy). A recurring theme on the implementation and modification of neural network algorithms for FPGA, is that advantages in resource and power efficiency come at the price of less accuracy. How much one is willing to sacrifice for the other will affect the choice of using a FPGA or a GPU for the implementation of a CNN. In the end, the existence of tasks that demand fast systems with low power consumption creates incentives for the use of FPGA. Another of the promises of FPGA systems is parallelism, which would enable them to fulfil several tasks of an algorithm at the same time. These reasons justify the experimentation with CNN architectures and algorithms with the aim of creating competitive configurable systems based on FPGAs.

REFERENCES

1. *More A., Wilson R.* FPGAs for Dummies. City: John Wiley & Sons Inc, 2017. P. 3-6.
2. *Yufeng H.* A General Neural Network Hardware Architecture on FPGA // Dept. of Electronic, Electrical and Systems Engineering University of Birmingham, City: Birminham, 2017.
3. *Omondi A., Rajapakse J., Bajger M.* FPGA Neurocomputers // FPGA Implementations of Neural Networks, Publisher: Springer, 2006.
4. *Lu L., Liang Y., Xiao Q., Yan S.* Evaluating Fast Algorithms for Convolutional Neural Networks on FPGAs // IEEE 25th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), 2017.

Solis Romeu Edgar — PhD Student, Master, Faculty of Innovation Technology. E-mail: *solisromeu@mail.ru*

Shashev Dmitry — Ph.D in Engineering, Deputy Dean for R&D at the Faculty of Innovative Technologies , Faculty of Innovation Technology. E-mail: *dshashev@mail.tsu.ru*

КРАЕВАЯ ЗАДАЧА ДЛЯ ОДНОГО КЛАССА УРАВНЕНИЙ ТРЕТЬЕГО ПОРЯДКА

М. О. Абулов

Каршинский государственный университет, г. Карши, Узбекистан

Рассматривается краевая задача для уравнения третьего порядка смешанного типа в четырехугольной области. С помощью метода Галеркина при определенных условиях на коэффициенты и правую часть уравнения доказано существование слабо обобщенного решения в пространстве Соболева. При этих же условиях доказана единственность обобщенного решения.

Ключевые слова: *Краевая задача, уравнения третьего порядка, уравнения смешанного типа.*

Введение

Теория краевых задач для уравнений смешанного типа в силу ее прикладной и теоретической значимости стала одним из важнейших разделов теории дифференциальных уравнений с частными производными. Начало исследований краевых задач для уравнений смешанного типа было положено в работах Ф. Трикоми и С. Геллерстедта, где впервые ставились и изучались краевые задачи для модельных уравнений смешанного типа. Библиографию вопроса можно найти в монографиях [1, 2]. В дальнейшем, в работах [3, 4] предложенные подстановки обобщались для уравнений высокого порядка, для различных краевых условий для операторных уравнений и для уравнений неклассического типа. Среди последних работ можно отметить [5, 6].

В данной работе будут использоваться обычные пространства L_p, Соболева W_p^l. Определения и свойства этих пространств можно найти в монография [7].

В области $Q = \{(x,t) : \; -1 \leqslant x \leqslant 1, \; 0 \leqslant t \leqslant T\}$ рассмотрим уравнение

$$\begin{aligned} Lu \equiv u_{ttt} + \mu(x)u_{xxx} + a(x,t)u_{xx} + b(x,t)u_{tt} + \\ + c(x,t)u_x + d(x,t)u_t + e(x,t)u = f(x,t) \quad (1) \end{aligned}$$

где $x\mu(x) > 0$ при $x \neq 0$, $\mu(0) = 0$.

Краевая задача

Найти в области Q решение уравнения (1), удовлетворяющее краевым условиям

$$u\,|_{\partial Q} = 0, \qquad u_t\,|_{t=0} = 0. \tag{2}$$

Для простоты изложения будем предполагать, что коэффициенты уравнения (1) являются бесконечно дифференцируемыми функциями.

Определение 1. Обозначим через $H(Q)$ пространство функций, полученное замыканием функций из $C^\infty(Q)$, удовлетворяющих условиям (2) по норме

$$\|u\|^2_{H(Q)} = \int\limits_Q \left(u_{xx}^2 + u_{xt}^2 + u_x^2 + u_t^2 + u^2\right) dQ.$$

Определение 2. Функцию $u \in H(Q)$ будем называть слабым обобщенным решением задачи (1)–(2), если для всех $v \in C_0^\infty(Q)$ выполняется тождество

$$\int\limits_Q (u_t v_{tt} - \mu u_{xx} v_x - \mu_x u_{xx} v + a u_{xx} v - b u_t v_t -$$

$$- b_t u_t v + c u_x v + d u_t v + e u v) dQ = \int\limits_Q f v dQ. \tag{3}$$

Теорема 1. Пусть выполнены условия

$$a(x,t) - \frac{3}{2}\,|\mu_x| \geqslant \delta > 0, \quad b(x,t) \geqslant 0. \tag{4}$$

Тогда для любой функции $f(x,t)$, такой, что $f \in L_2(Q)$, существует единственное решение задачи (1)–(2) из $H(Q)$.

Доказательство. Решение задачи (1)–(2) будем искать методом Галеркина

$$u_m(x,t) = \sum_{i=1}^m j_i(t) \varphi_i(x),$$

где функции $\varphi_i(x)$ являются решениями задачи

$$\varphi_i{}'' = -\lambda_i \varphi_i, \quad \varphi_i(-1) = \varphi_i(1) = 0,$$

а коэффициенты $j_i(t)$ находятся из решения системы обыкновенных дифференциальных уравнений

$$(u_{mttt},\varphi_i)_0+(\mu u_{mxxx},\varphi_i)_0+(au_{mxx},\varphi_i)_0+(bu_{mtt},\varphi_i)_0+ \\ +(cu_{mx},\varphi_i)_0+(du_{mt},\varphi_i)_0+(eu_m,\varphi_i)_0=(f,\varphi_i)_0, \quad (5)$$

$$j_i(0)=j_i(T)=j_{it}(0)=0, \quad i=1,2,\ldots,m. \quad (6)$$

Разрешимость задачи (5)–(6) при фиксированном m вытекает из общей теории обыкновенных дифференциальных уравнений.

Получим равномерные по m оценки для Галеркинских приближений. Для этого умножим (5) на $(-j_i(t))$ и, суммируя по i, получим

$$(u_{mttt},-u_m)_0+(\mu u_{mxxx},-u_m)_0+(au_{mxx},-u_m)_0+(bu_{mtt},-u_m)_0+ \\ +(cu_{mx},-u_m)_0+(du_{mt},-u_m)_0+(eu_m,-u_m)_0=(f,-u_m). \quad (7)$$

Отсюда, интегрируя по tи интегрируя по частям, в силу (2), (4), после некоторых преобразований приходим к неравенству

$$\int\limits_Q (u_{mt}^2+u_{mx}^2+u_m^2)dQ\leqslant C. \quad (8)$$

Далее, рассмотрим следующее равенство

$$(u_{mttt},u_{mxx})_0+(\mu u_{mxxx},u_{mxx})_0+(au_{mxx},u_{mxx})_0+ \\ +(bu_{mtt},u_{mxx})_0=(f-cu_{mx}-du_{mt}-eu_m,u_{mxx})_0. \quad (9)$$

Из тождества (9), в силу (2), (3) и оценки (8), интегрируя по t и интегрируя по частям, после несложных преобразований вытекает следующая оценка

$$\int\limits_Q (u_{mxx}^2+u_{mxt}^2)dQ\leqslant C. \quad (10)$$

Из оценок (8), (10) следует ограниченность последовательности приближенных решений $\{u_m(x,t)\}$ в пространстве $H(Q)$, можно выбрать подпоследовательность $\{u_{m_k}(x,t)\}$ и перейти к пределу при $m_k\to\infty$ в системе (5). Нетрудно проверить, что предельная функция принадлежит пространству $H(Q)$ и удовлетворяет тождеству (3). Поскольку система $\{\varphi_i(x)\}$ плотна в $L_2(-1,\ 1)$.

Докажем, что решение задачи (1)–(2) единственно.

Если u, v – два решения задачи (1)–(2), то $w=u-v$ удовлетворяет уравнению

$$w_{ttt}+\mu(x)w_{xxx}+a(x,t)w_{xx}+b(x,t)w_{tt}+c(x,t)w_x+d(x,t)w_t+e(x,t)w=0.$$

Рассмотрим интеграл

$$\int\limits_Q (w_{ttt} + \mu(x)w_{xxx} + a(x,t)w_{xx} + b(x,t)w_{tt} +$$
$$+ c(x,t)w_x + d(x,t)w_t + e(x,t)w)wdQ = 0$$

и интегрируя по частям, в силу (2), получим

$$\int\limits_Q (w_t^2 + w_x^2 + w^2)dQ \leqslant 0,$$

откуда следует, что $w = 0$ в Q. ■

В данной работе доказаны новые теоремы существования и единственности решения краевой задачи (1)–(2), которые позволяет расширить круг решаемых проблем в теории краевых задач для неклассических уравнений математической физики.

СПИСОК ЛИТЕРАТУРЫ

1. *Сабитов К. Б.* К теории уравнений смешанного типа. М., Физматлит, 2014. 290 с.
2. *Врагов В. Н.* Краевые задачи для неклассических уравнений математической физики. Новосибирск: НГУ, 1983. 86 с.
3. *Егоров И. Е., Федоров В. Е.* Неклассические уравнения математической физики высокого порядка. Новосибирск, 1995. 134 с.
4. *Абулов М. А.* Краевая задача для смешанно-составного типа третьего порядка. // Новосибирск, НГУ, 2021. – 5с. Деп. в ВИНИТИ, №1436 – В89.
5. *Артюшин А. Н.* О регулярной разрешимости краевой задачи для уравнения третьего порядка с меняющимся направлением эволюции в весовых пространствах Соболева // Сибирские электронные математические известия. 2019. Т. 16. С. 2003–2012.
6. *Кожанов А. И.* Начально-граничные задачи для вырождающихся гиперболических уравнений // Сибирские электронные математические известия. 2021. Т. 18. С. 43–53.
7. *Соболев С. Л.* Некоторые применения функционального анализа в математической физике. М. Наука, 1988. 336 с.

Абулов Мумин Орзикулович — к.ф.-м.н., доцент кафедры Алгебры и геометрии. E-mail: *abulov1959@mail.ru*

НЕЛОКАЛЬНАЯ ЗАДАЧА ДЛЯ ОДНОГО УРАВНЕНИЯ ЧЕТВЕРТОГО ПОРЯДКА

М. О. Абулов, Б. А. Турсунов

Каршинский госудорственный университет, г. Карши, Узбекистан

В данной работе изучена нелокальная задача для одного уравнения четвертого порядка. При выполнении некоторых условий на правую часть уравнения доказаны существование и единственность решения поставленной задачи

Ключевые слова: *нелокальная задача, сопряженная задача, уравнение четвертого порядка.*

Введение

Нелокальные задачи в настоящее время являются интенсивно развивающимся разделом теории дифференциальных уравнений. Интерес к нелокальной задаче (кроме теоретического значения) вызван, очевидно, возможностью ее физической интерпретации: если дифференциальные уравнения описывают некоторый физической процесс, то нелокальные краевые условия являются некоторыми алгебраическими выражениями, связывающими искомое решение и его производные в двух и более точках наблюдения физического процесса. Проблемы современной науки и техники выдвинули решение более реальных практических задач, связанных с исследованием разнообразных классов математических моделей. Известно, что математическое моделирование многих биологических и технологических процессов приводит к изучению нелокальных краевых задач для различных классов дифференциальных уравнений. Поэтому, изучение нелокальных задач для различных классов дифференциальных уравнений привлекало внимание многих математиков. Библиографию вопроса можно найти в монографиях [1, 2]. Среды последних работ можно отметить [3, 4, 5, 6, 7].

Нелокальная задача

В области

$$Q = \{(x, y, t) : 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ 0 \leqslant t \leqslant T\}$$

рассмотрим уравнение четвертого порядка

$$Lu \equiv u_{ttxx} - u_{xxxx} - u_{yyxx} + u_{xx} = f(x, y, t) \tag{1}$$

Поставим задачу нахождения в области Q решения уравнения (1), удовлетворяющее следующим условиям

$$u\,|_{t=0} = e^{\frac{\lambda T}{2}} u\,|_{t=T}\ , \quad u_t\,|_{t=0} = e^{\frac{\lambda T}{2}} u_t\,|_{t=T}\ ,$$
$$u\,|_{x=0} = u\,|_{x=1} = u_x\,|_{x=0} = u_x\,|_{x=1} = 0,\ u\,|_{y=0} = u\,|_{y=1} = 0 \quad (2)$$

где $\lambda > 0-$некоторые числа.

Нетрудно видеть, что для (1)–(2) сопряженная задача имеет вид:

$$L^* v \equiv v_{ttxx} - v_{xxxx} - v_{yyxx} + v_{xx} = g(x, y, t) \quad (3)$$

$$v\,|_{t=0} = e^{-\frac{\lambda T}{2}} v\,|_{t=T}\ , \quad v_t\,|_{t=0} = e^{-\frac{\lambda T}{2}} v_t\,|_{t=T}\ ,$$
$$v\,|_{x=0} = v\,|_{x=1} = v_x\,|_{x=0} = v_x\,|_{x=1} = 0,\ v\,|_{y=0} = v\,|_{y=1} = 0 \quad (4)$$

Обозначим через C_L (соответственно C_{L^*}) класс функций, четырежды непрерывно дифференцируемых в области Q и удовлетворяющих условию (2) (соответственно (4)).

Определение 1. Обозначим через $H(Q)$ (соответственно $H(Q)^*$) пространство функций, полученное замыканием функций из C_L (соответственно из C_{L^*}) по норме

$$\|u\|^2 = \int\limits_Q (u_{ttx}^2 + u_{txx}^2 + u_{txy}^2 + u_{tt}^2 + u_{tx}^2 +$$
$$+ u_{xx}^2 + u_{xy}^2 + u_t^2 + u_x^2 + u_y^2 + u^2) dQ.$$

Определение 2. Функцию $u \in H(Q)$ будем называть слабым обобщенным решением задачи (1)–(2), если для всех $v \in C_{L^*}$ выполняется тождество

$$B(u, v) = \int\limits_Q (u_{xtt} v_x + u_{xx} v_{xx} + u_{xy} v_{xy} + u_{xx} v) dQ = \int\limits_Q f v dQ, \quad \forall v \in C_{L^*}.$$

Теорема 1. Пусть для функции $f(x, y, t) \in W_2^1(Q)$ выполнены условии

$$f\,|_{t=0} = e^{\frac{\lambda T}{2}} f\,|_{t=T}\ , \quad \int\limits_0^x f(w, y, 0) dw = e^{\frac{\lambda T}{2}} \int\limits_0^x f(w, y, T) dw. \quad (5)$$

Тогда существует единственное решение задачи (1)–(2) из $H(Q)$.

Доказательство. Умножая уравнение (1) на $e^{\lambda t}u_t$, интегрируя по области Q, и в силу условий (2), (3) получим

$$\int\limits_Q (u_{tx}^2 + u_{xx}^2 + u_{xy}^2 + u_x^2)dQ \leqslant C\int\limits_Q f^2 dQ. \tag{6}$$

Уравнение (1) можем переписать в виде

$$u_{ttx} = u_{xxx} + u_{xyy} + u_x + \int\limits_0^x f(w,y,t)dw + \varphi(y,t), \tag{7}$$

где $\varphi(x,y)-$неизвестная функция. Из (7), в силу (3), получим

$$u_{xtt}\,|_{x=0} = u_{xxx}\,|_{x=0} + \varphi(y,t) = 0 \quad \text{т.е} \quad \varphi(y,t) = -u_{xxx}\,|_{x=0}$$

так как

$$\varphi(y,0) = -u_{xxx}(0,y,0) \quad \text{и} \quad \varphi(y,T) = -u_{xxx}(0,y,T).$$

Отсюда, в силу (2), получим

$$\varphi(y,0) = e^{\frac{\lambda T}{2}}\varphi(y,T).$$

Тогда из уравнения (7) выводим, что

$$u_{xtt}\,|_{t=0} = e^{\frac{\lambda T}{2}}u_{xtt}\,|_{t=T}. \tag{8}$$

Далее, уравнение (1) дифференцируя по t, после этого умножая на $e^{\lambda t}u_{tt}$ и интегрируя по области Q, в силу условий (2), (6), (8), получим

$$\int\limits_Q (u_{ttx}^2 + u_{txx}^2 + u_{txy}^2 + u_{tx}^2)dQ \leqslant C\int\limits_Q (f^2 + f_t^2)dQ. \tag{9}$$

Так как, в силу (2), (3), для решения задачи (1)–(2) верна оценка

$$\int\limits_Q (u_{ttx}^2 + u_{txx}^2 + u_{txy}^2 + u_{tt}^2 + u_{tx}^2 + u_{xx}^2 + u_{xy}^2 +$$

$$+ u_t^2 + u_x^2 + u_y^2 + u^2)dQ \leqslant C\int\limits_Q (f^2 + f_t^2)dQ. \tag{10}$$

Теперь рассмотрим сопряженную задачу (3)–(4). Аналогично, как в (5), умножая уравнения (3) на $(-e^{-\lambda t}v_t)$ и интегрируя по области Q и в силу (4), (8) получим

$$\int\limits_Q (v_{tx}^2 + v_{xx}^2 + v_{xy}^2 + v_t^2)dQ \leqslant C\int\limits_Q g^2 dQ. \tag{11}$$

Заметим, что как в (8), верно равенство

$$v_{xtt}\,|_{t=0} = e^{-\frac{\lambda T}{2}} v_{xtt}\,|_{t=T}. \tag{12}$$

Далее, уравнение (3) дифференцируя по t, после этого умножая на $(-e^{-\lambda t}v_{tt})$ и интегрируя по области Q, интегрируя по частям в силу (4), (11), (12) получим

$$\int\limits_Q (v_{ttx}^2 + v_{txx}^2 + v_{txy}^2 + v_{tt}^2)dQ \leqslant C\int\limits_Q g_t^2 dQ. \tag{13}$$

Из оценок (10), (11), (11) стандартным образом вытекает существования и единственность решения задачи (1), (2) (см. например, [9, с. 19, с. 35]) в пространстве $H(Q)$. ■

Нетрудно видеть, что задачу (1)–(2) можно трактовать как обратную задачу для определения правой части волнового уравнения.

Заключение

В данной статье доказаны новые теоремы существования и единственности решения нелокальной задачи (1)–(2), которые позволяют расширить круг решаемых проблем в теории нелокальных задач для неклассических уравнений математической физики.

СПИСОК ЛИТЕРАТУРЫ

1. *Тахиров Ж. О.* Неклассические нелинейные задачи и задачи со свободной границей. Ташкент, 2014, 240 с.
2. *Мамедов И. Г.* Решение многомерных локальных и нелокальных краевых задач для гиперболических уравнений высокого порядка с негладкими коэффициентами и их применение к задачам оптимального управления. Баку, 2015. 49 с.
3. *Балкизов Ж. А.* Нелокальная краевая задача для уравнения параболо-гиперболического типа третьего порядка с вырождением типа и порядка в области его гиперболичности // Итоги науки и техники. Современная математика и ее приложения. 2018. Т. 149. С. 14–24.

4. *Моисеев Е. И., Лихоманенко Т. Н.* Об одной нелокальной краевой задаче для уравнения Лаврентьева–Бицадзе // Доклады Академии наук. 2012. Т. 446. № 3. С. 256–258.
5. *Сабитов К. Б.* Критерий единственности решения нелокальной задачи для вырождающегося уравнения смешанного типа в прямоугольной области // Дифференциальные уравнения. 2010. Т. 46. № 8. С. 1205–1208.
6. *Вадакова В. А., Гучаева З. Х.* Нелокальная задача для нагруженного уравнения третьего порядка с кратными характеристиками // Успехи современного естествознания. 2014. № 7. С. 90–92.
7. *Репин О. А., Тарасенко А. В.* Нелокальная задача для уравнения с частной производной дробного порядка // Вестник Самарского государственного технического университета. Физико-математические науки. 2015. Т. 19. № 1. С. 78–86.
8. *Беразанский В. М.* Разложение по собственным функциям самосопряженных операторов. Киев, 1965. 800 с.
9. *Врагов В. Н.* Краевые задачи для неклассических уравнений математической физики. Новосибирск: НГУ, 1983. 84 с.
10. *Абулов М. О.* Разрешимость смешанных, нелокальных и краевых задач для некоторых классов третьего порядка. Ташкент, 2020. 82 с.

Абулов Мумин Орзикулович — к.ф.-м.н., доцент кафедры Алгебры и геометрии. E-mail: *abulov1959@mail.ru*

Турсунов Байрамали Акбарович — к.ф.-м.н. (Ph.D.), ст. преподаватель кафедры Алгебры и геометрии. E-mail: *bakbarovich@mail.ru*

ЗАДАЧА RCSP ДЛЯ ИНВЕСТИЦОННЫХ ПРОЕКТОВ В УСЛОВИЯХ ОГРАНИЧЕННОСТИ РЕСУРСОВ И НЕОПРЕДЕЛЁННОСТИ ЗАТРАТ И ДОХОДОВ

А. М. Булавчук[1], Д. В. Семенова[1,2]

[1] *Сибирский федеральный университет, г. Красноярск, Россия*
[2] *Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого, г. Красноярск, Россия*

Исследуется задача календарного планирования (RCPSP) для инвестиционных проектов. Рассматриваемая постановка предполагает нахождение максимума чистой приведенной прибыли в условиях ограниченности ресурсов и неопределённости затрат и доходов по проекту. Запасы и затраты ресурсов в модели оцениваются в денежном выражении.

Ключевые слова: *Календарное планирование, инвестиционный проект, NPV.*

Введение

Календарное планирование инвестиционных проектов — важная практическая задача, возникающая в сфере управления производством. Рассмотрим инвестиционный проект, предполагающий выполнение заданного количества работ за определенный промежуток времени. Каждая из работ характеризуется длительностью её выполнения, затратами ресурсов и доходами, которые она генерирует. Кроме того, известны технологические и организационные взаимосвязи между работами. Будем считать, что все ресурсы проекта могут быть сведены к денежным. Для оценки результатов реализации проекта воспользуемся показателем чистой приведенной стоимости (NPV). Основными проявлениями факторов неопределенности, действующих в процессе реализации проекта, будут недополучение запланированных доходов и рост издержек реализации проекта.

1. Постановка задачи

Будем опираться на модель RCSP (Resource-constrained project scheduling problem), предложенную В. В. Сервахом и Е. А. Казаковцевой [2]. Пусть $V = \{1, 2, \ldots, N\}$ — множество запланированных работ, связанных друг с другом отношением частичного порядка E. Каждая из работ характеризуется сроками $Q = (q_1, q_2, \dots, q_N)$, то есть целым числом периодов, необходимых для ее выполнения. Обозначим через $c_j(\tau) \in \mathcal{F}$ случайные величины поступлений от j-й работы в момент времени $\tau \in \{0, 1, \dots, q_j\}$. Для отрицательных значений $c_j(\tau)$ речь будет идти о расходах на выполнение работы. Пусть T — максимальная длительность проекта в целых периодах. Бюджет проекта в момент $t \in \{0, 1, \dots, T-1\}$ обозначим через $K(t)$, а ставку дисконтирования для соответствующего периода — через r_0. Кроме того, обозначим через N_t множество работ, выполняемых в интервале $[t; t+1)$. Будем искать оптимальные сроки начала каждой из работ, которые образуют расписание выполнения проекта $S = (s_1, s_2, \dots, s_N)$, где s_j — момент начала j-й работы. Очевидно, что $s_j \in \{0, 1, \dots T - q_j\}$.

При применении формулы сложных процентов чистый приведенный доход от j-й работы рассчитывается по формуле: $NPV_j = \sum\limits_{\tau=0}^{q_j} \dfrac{c_j(\tau)}{(1+r_0)^\tau}$.

В каждый момент времени должно выполнятся условие достаточности средств. Источником средств выступают бюджет проекта, прибыль, полученная в предыдущие периоды, а также доходы от реинвестирования неизрасходованной прибыли. Поскольку часть указанных величин носит случайный характер, то допустимое расписание, соответствующее какой-либо одной реализации случайной величины, для другой реализации окажется недопустимым. На практике проблема недоспустимости выбранного на старте проекта расписания решается путем привлечения заёмных средств. При этом сумма дополнительных расходов складывается не только из размера дефицита во все моменты времени, но и из платежей за пользование деньгами. Пусть r^* – размер ставки за пользование деньгами в течение одного периода. Тогда дополнительные издержки на покрытие затрат по проекту могут быть расчитаны по формулам:

$$D(S) = \sum_{t=0}^{T-1} \delta_t D_t, \tag{1}$$

$$\text{где } D_t = \sum_{\tau=0}^{t} \frac{K(\tau)}{(1+r_0)^\tau} + \sum_{\tau=0}^{t} \sum_{j \in N_\tau} \frac{c_j(\tau - s_j)}{(1+r_0)^\tau}, \tag{2}$$

$$\delta_t = \begin{cases} 1 + r^*, & \text{если } D_t < 0, \\ 0, & \text{если } D_t \geq 0. \end{cases} \tag{3}$$

Найденная для данного расписания величина дефицита средств $D(S)$ должна учитываться при расчете целевой функции оптимизационной задачи. Обозначим через $NPV_D(S)$ текущую стоимость прибыли с учётом затрат на привлечение кредитов. Тогда оптимизационная модель примет вид:

$$NPV_D^*(S) = \max_S \left[\sum_{j \in V} \frac{NPV_j}{(1+r_0)^{s_j}} + D(S) \right] = \max_S \left[NPV(S) + D(S) \right] \quad (4)$$

$$s_i + q_i \le s_j, \quad (i,j) \in E. \quad (5)$$

Ограничениями в модели (4) выступают условия частичного порядка (5).

Величина $NPV(S)$ характеризует приведенную стоимость прибыли за весь срок реализации проекта при выбранном расписании S без учёта затрат на кредитование. Очевидно, что $NPV(S)$ и $D(S)$ представляют собой случайные величины, распределение которых зависит от распределений случайных характеристик работ. Отметим также, что при заданном расписании S функция $NPV(S)$ есть линейная комбинация величин поступлений от каждой из работ:

$$NPV(S) = \sum_{j=1}^{N} \sum_{\tau=0}^{q_j} \alpha_j(\tau) c_j(\tau), \quad (6)$$

где $\alpha_j(\tau) = (1+r_0)^{-(\tau+s_j)}$.

Задача календарного планирования с критерием максимизации приведённой прибыли является NP-трудной в сильном смысле [2].

2. Случайные параметры проекта

Пусть компоненты потока платежей для каждой из работ в каждый момент времени представляют собой независимые случайные величины. Поведение целевой функции будет зависеть от того, какое именно распредление имеют указанные величины. Представляется целесообразным выбирать одно распределение для всех случайных характеристик проекта, поскольку все они имееют одну и ту же экономическую природу. Нами были рассмотрены несколько вариантов непрерывных распределений. Приведем пример анализа для нормального распределения.

Если все компоненты потока платежей подчиняются нормальному закону распределения $c_j(\tau) \in \mathcal{N}\left(\mu_j(\tau), \sigma_j(\tau)\right)$, то для каждого расписания S верно, что $NPV(S) \in \mathcal{N}(\mu(S), \sigma(S))$, где

$$\mu(S) = \sum_{j=1}^{N}\sum_{\tau=0}^{q_j} \alpha_j(\tau)\mu_j(\tau), \quad \sigma(S) = \sqrt{\sum_{j=1}^{N}\sum_{\tau=0}^{q_j} \alpha_j^2(\tau)\sigma_j^2(\tau)}. \qquad (7)$$

3. Вычислительные эксперименты

Рассмотрим пример проекта, состоящего из 6 работ. Порядок следования работ, а также число периодов выполнения каждой работы представлены на рисунке 1.

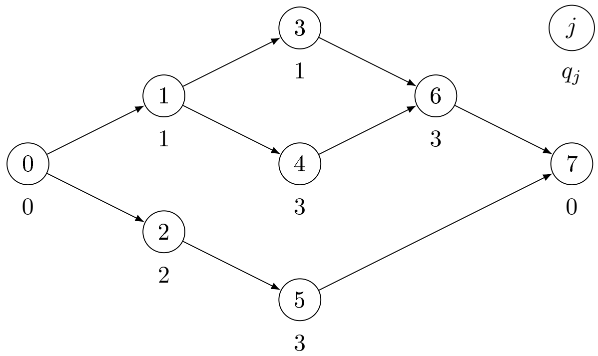

Рис. 1. Сетевой график проекта

Максимальный срок реализации проекта – 8 лет. Примем запас капитала в каждом году равным 5 единицам, ставку дисконтирования равной 10 % годовых, а ставку по кредитам – 20 % годовых. В качестве характеристик денежного потока для каждой из работ возьмём верхнюю и нижнюю границы величины поступлений $\left(c_j^l(\tau), c_j^h(\tau)\right)$. Указанные характеристики приведены в таблице 1.

Рассмотрим случай нормального распределения характеристик денежного потока. Параметры распределений $\mu_j(\tau)$ и $\sigma_j(\tau)$ расчитаем по формулам:

$$\mu_j(\tau) = \frac{c_j^l(\tau) + c_j^h(\tau)}{2}, \sigma_j(\tau) = \frac{\mu_j(\tau) - c_j^l(\tau)}{3}. \qquad (8)$$

Проведем серию из 1000 экспериментов по нахождению оптимального расписания для данного проекта с использованием ранее разработанного авторами генетического алгоритма [3]. Результаты экспериментов приведены в таблице 2.

Таблица 1

Границы изменений величины поступлений $(c_j^l(\tau), c_j^h(\tau))$

j	τ			
	0	1	2	3
1	$(-7,-3)$	$(-3,-1)$		
2	$(-7,-3)$	$(-2,-1)$	$(7,3)$	
3	$(-2,-1)$	$(8,12)$		
4	$(-6,-4)$	$(-6,-4)$	$(1,3)$	$(6,10)$
5	$(4,7)$	$(5,7)$	$(3,6)$	$(1,3)$
6	$(-7,-3)$	$(-3,-1)$	$(-4,-2)$	$(5,11)$

Таблица 2

Результаты вычислительных экспериментов

Расписание	Число исходов	$NPV(S)$	$D(S)$	$NPV^*(S)$
$(1,0,2,2,2,5)$	685	$8,035$	$-1,221$	$6,814$
$(0,2,1,2,4,5)$	226	$5,091$	$-0,177$	$4,914$
$(1,3,2,2,5,5)$	89	$3,383$	0	$3,383$

В зависимости от тех значений, которые принимают параметры проекта, оптимальным оказывается одно из трёх расписаний. Более высокому среднему значению прибыли соответствует и более высокая средняя потребность в кредитных ресурсах. На рисунке 2 приведены гистограммы плотности распределения $NPV_D^*(S)$ каждого из расписаний.

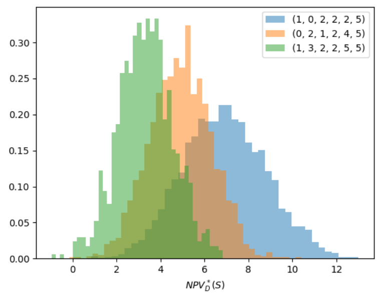

Рис. 2. Распределение прибыли для различных расписаний

Только одно расписание $(1,3,2,2,5,5)$ оказалось доступно без привлечения заёмных средств, однако оно же демонстрирует худшие показатели среднего размера прибыли. По данным о распределении прибыли от проекта была произведена оценка риска отклонения $NPV_D(S)$ от ожидаемого значения.

Заключение

Предложенная модель задачи календарного планирования позволяет учитывать неопределенность потока платежей проекта. При оптимизации принимаются во внимание также затраты на привлечение заёмных средств. На основе анализа модели сделаны выводы о распределении целевых показателей. Проведенные вычислительные эксперименты подтверждают сделанные выводы и демонстрируют подход к выбору расписания в условиях неопределённости.

СПИСОК ЛИТЕРАТУРЫ

1. *Джонсон Н. Л., Коц С., Балакришнан Н.* Одномерные непрерывные распределения: справочник: в 2 частях, Ч. 1. Москва: Лаборатория знаний, 2017. 706 с.
2. *Казаковцева Е. А., Servах В. В.* Сложность задачи календарного планирования с кредитованием // Дискретный анализ и исследование операций. 2015. Т. 22. №4. С. 35–49.
3. *Bulavchuk A. M.,Semenova D. V.* Genetic Algorithm Based on Idempotent Algebra Methods for RCPSP // 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT), 2021, pp. 1-4.

Булавчук Александр Михайлович — старший преподаватель, Институт экономики, государственного управления и финансов СФУ. E-mail: *abulavchuk@sfu-kras.ru*

Семенова Дарья Владиславовна — кандидат физико-математических наук, доцент, доцент, Институт математики и фундаментальной информатики СФУ. E-mail: *DVSemenova@sfu-kras.ru*

СРАВНЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ ЗАТРАТ НА ОСНОВНЫЕ МАТЕМАТИЧЕСКИЕ ОПЕРАЦИИ С ПОМОЩЬЮ СИСТЕМЫ NMPUD

А. В. Войтишек[1,2,3], М. Г. Гаджиахмедов[2], И. А. Рыжов[3], И. А. Трофимов[3]

[1]*Институт вычислительной математики и математической геофизики СО РАН, г. Новосибирск, Россия*
[2]*Новосибирский государственный университет, г. Новосибирск, Россия*
[3]*Лицей № 130 города Новосибирска, г. Новосибирск, Россия*

В работе описано новое полезное применение специальной компьютерной системы моделирования вероятностных одномерных распределений NMPUD: это возможность объективного сравнения трудоемкостей основных арифметических операций и вычислений значений функций.

Ключевые слова: *Трудоемкость метода Монте-Карло, компьютерная система NMPUD, генератор псевдослучайных чисел, компьютерная трудоемкость математических операций.*

Введение

При проведении научных исследований многих реальных процессов и явлений важным является построение соответствующих математических моделей. Эти модели позволяют как уточнять параметры натурных экспериментов (а иногда и замещать эти эксперименты), так и предсказывать новые эффекты. В свою очередь, методы вычислительной математики позволяют проводить приближенные расчеты, связанные с той или иной моделью, на современных компьютерах.

Важным классом современных математических моделей являются стохастические (вероятностные) модели. Им соответствуют вычислительные (компьютерные) алгоритмы статистического моделирования (или методы Монте-Карло; см., например, [1, 2]). В этих алгоритмах так или иначе присутствуют алгоритмы (формулы) вида

$$\xi_0 = \Psi_\xi\left(\alpha_0; \boldsymbol{\lambda}\right) \tag{1}$$

для компьютерного моделирования выборочных значений ξ_i одномерных случайных величин $\xi \in (a, b); -\infty \leqslant a < b \leqslant +\infty$, имеющих заданные (выбранные) плотности распределения

Работа выполнена в рамках государственного задания ИВМиМГ СО РАН (проект 0251–2021–0002).

$$f_\xi\,(u;\boldsymbol{\lambda})\,, \quad u \in (a,b); \tag{2}$$

здесь $\alpha_0 \in U(0,1)$ – стандартное случайное число (см. раздел 1.1 книги [1] и раздел 2.4 книги [2]), т. е. выборочное значение случайной величины $\alpha \in U(0,1)$, равномерно распределенной в интервале $(0,1)$ (это выборочное значение реализуется на компьютере с помощью соответствующего генератора – специальной подпрограммы, именуемой в языках программирования *RAND* или *RANDOM*), а $\boldsymbol{\lambda}$ – параметр (набор параметров) распределения.

В теории и приложениях статистического моделирования алгоритмы (формулы) (1) чаще всего используются в последовательных процедурах моделирования выборочных значений $\boldsymbol{\xi}_i = \left(\xi_i^{(1)}, ..., \ldots, \xi_i^{(d)}\right); i = 1, ..., n$ компонент случайного вектора $\boldsymbol{\xi} = \left(\xi^{(1)}, ..., \ldots, \xi^{(d)}\right)$ для приближенного вычисления средних вида $I_q^{(\boldsymbol{\xi})} = \mathbf{E}q(\boldsymbol{\xi})$ по формуле

$$I_q^{(\boldsymbol{\xi})} = \mathbf{E}q(\boldsymbol{\xi}) \approx Z_n = \frac{q(\boldsymbol{\xi}_1) + ... + q(\boldsymbol{\xi}_n)}{n} \tag{3}$$

(см. введение книги [1] и раздел 1.4 книги [2]).

Известно, что погрешность $\delta_n = \left| I_q^{(\boldsymbol{\xi})} - Z_n \right|$ приближения (3) убывает при росте числа n выборочных значений $\left\{q(\boldsymbol{\xi}_j); j = 1, ..., n\right\}$ достаточно медленно, а точнее, выполнено соотношение

$$\delta_n \approx H \frac{\sqrt{\mathbf{D}q(\boldsymbol{\xi})}}{\sqrt{n}}; \;\; \mathbf{D}q(\boldsymbol{\xi}) = \mathbf{E}q^2(\boldsymbol{\xi}) - [\mathbf{E}q(\boldsymbol{\xi})]^2; \;\; H \approx 0.80...$$

(см. подраздел 1.9 книги [2]). Поэтому в практических расчетах берется большое количество $n \gg 1$ выборочных значений $\left\{q(\boldsymbol{\xi}_j); j = 1, ..., n\right\}$, а значит и число $d \times n$ обращений к формулам (алгоритмам) (1) является весьма большим. Отсюда следует, что формулы (1) следует выбирать (делать) максимально экономичными.

1. Компьютерная система NMPUD

Конструирование, изучение и накопление экономичных моделирующих формул (алгоритмов) (1) происходит в компьютерной системе NMPUD (Numerical Modelling of Probabilistic Univariate Distributions), разрабатываемой в лаборатории математического моделирования лицея № 130 города Новосибирска (подробное описание системы см. в работах [3, 4]). Эта система достаточно активно используется исследователями и студентами, занимающимися изучением, разработкой и (или) использованием компьютерных стохастических моделей для решения важных прикладных проблем.

В настоящее время в банк системы NMPUD входят формулы *элементарных вероятностных плотностей* (2) (см. раздел 1.4.1 книги [1] и разделы 2.5, 2.6 книги [2]), т. е. таких, для которых имеются формулы т. н. *метода обратной функции распределения*

$$\xi_0 = F_\xi^{-1}(\alpha_0; \boldsymbol{\lambda}) \tag{4}$$

(здесь $F_\xi(x; \boldsymbol{\lambda}) = \mathbf{P}\{\xi < x\} = \int_{-\infty}^{x} f_\xi(u; \boldsymbol{\lambda})\, du$ – функция распределения случайной величины ξ), получаемые с помощью решения уравнения $\int_a^{\xi_0} f_\xi(u; \boldsymbol{\lambda})\, du = \alpha_0$ относительно верхнего предела ξ_0 интеграла из левой части этого соотношения. В банке фиксируется также такая важная характеристика формулы (алгоритма) (1) или (4) как затраты на моделирование (среднее время одного обращения к формуле (1) или (4) – в наносекундах).

Пополнение банка системы NMPUD производится с помощью *технологии последовательных (вложенных) замен*, подробно описанной в работе [4] (см. также подраздел 14.2 книги [2]). Активно разрабатываются отдельные блоки по автоматизации технологии последовательных замен, а также по использованию различных вариаций модифицированного метода дискретной суперпозиции (см., в частности, статью [5] из настоящего сборника).

2. О генераторе стандартных случайных чисел, используемом в системе NMPUD

При разработке компьютерной системы NMPUD крайне важными оказались следующие исследования, проведенные нами совместно с Т. Е. Булгаковой и С. В. Булгаковым и касающиеся принципиального вопроса о том, какой генератор стандартных случайных чисел $\alpha_1 \in U(0,1)$ следует использовать в системе. Здесь весьма важным был учет следующего соображения.

Замечание 1. (см. раздел 1.1.4 книги [1] и раздел 9.5 книги [2]). Компьютерные генераторы стандартных случайных чисел α_i, даже реализованные оптимально для используемых языков программирования, являются относительно трудоемкими. Поэтому при оптимизации алгоритмов метода Монте-Карло целесообразно по возможности уменьшать число обращений к подпрограмме типа RAND.

Мы следовали весьма разумным рекомендациям из раздела 1.1 книги [1] и раздела 9 книги [2] о целесообразности использования одного из вариантов мультипликативного метода вычетов

$$\alpha_{j+1} = \{Q\alpha_j\}; \quad j = 0, 1, 2, \ldots \tag{5}$$

для заданного $\alpha_0 \in (0,1)$; здесь множитель $Q \in \mathbb{N}$ является положительным целым (натуральным; как правило, $Q \gg 1$), а $\{A\}$ обозначает дробную часть числа A. Мы также учитывали, что в новосибирской школе методов Монте-Карло для вычислительных процедур с числом испытаний порядка 10^9 и меньше (а именно такие используются в системе NMPUD) долгие годы вполне удовлетворительным считался и считается рекуррентный генератор метода вычетов вида (5) с контролируемой мантиссой длины $m = 40$, множителем $Q = 5^{17}$ и длиной периода 2^{38} (см. разделы 1.1.4 и 1.1.5 книги [1] и разделы 9.5, 9.6 книги [2]), прошедший всестороннее многолетнее тестирование. Подробное сравнение время работы этого генератора с затратами других датчиков подтвердило целесообразность использования этого датчика в системе NMPUD.

3. Технология сравнения затрат на математические операции и ее применение

С учетом замечания 1 и анализа использования генераторов в формулах и алгоритмах моделирования случайных величин, нами были проведены замеры затрат на выполнение арифметических операций и получения значений основных элементарных функций. В качестве итога проведенных исследований можно сформулировать следующее уточнение замечания 1.

Замечание 2. Использование генераторов стандартных случайных чисел является достаточно трудоемким, однако соответствующие затраты не превышают затрат на вычисление значений основных элементарных функций.

В ходе исследований выяснилось, что наиболее объективные результаты замеров затрат получаются при комбинировании операций и функций со стандартными случайными числами. Поэтому возникает дополнительная полезная функция компьютерной системы NMPUD: система позволяет получать объективные соотношения между трудоемкостями упомянутых основных компьютерных процедур.

Мы выбрали вероятностные распределения (2), для которых формулы метода обратной функции распределения (4) имеют вид комбинации стандартного случайного числа α_0 и одной из важнейших компьютерных процедур (арифметическая операция, вычисление элементарной функции) и провели соответствующие расчеты на многих компьютерах. В таблице показаны результаты характерных замеров затрат Δ соответствующих моделирующих формул в системе NMPUD и соотношение затрат Δ с компьютерным временем A обращения к генератору псевдослучайных чисел (эти результаты получены на персональном компьютере *Acer Aspire 3* с процессором *AMD Ryzen 3 2200U with Radeon Vega Mobile Gfx (4 CPUs) 2.5GHz, 6 GB RAM).*

В ходе экспериментов выяснилось, что компьютерные затраты на сложение, умножение, деление и взятие квадратного корня невелики и примерно одинаковы; «умеренно большими» можно назвать затраты на вычисление тригонометрических функций (и обратных к ним) и логарифмов; «экстремально большими» являются затраты на вычисление степенной и показательной функций.

Таблица 1

Сравнения трудоемкостей основных арифметических операций и вычисления значений элементарных функций средствами системы NMPUD.

Плотность и интервал распределения	**Моделирующая формула**	**Затраты Δ в системе NMPUD**	**Соотношение $\frac{\Delta}{A}$**
$f_\xi(u) \equiv 1;\ 0 < u < 1$	$\xi_0 = \alpha_0$	$A = 170.00$	1.00
$f_\xi(u) \equiv 1;\ 1 < u < 2$	$\xi_0 = 1 + \alpha_0$	211.43	1.24
$f_\xi(u) \equiv \frac{1}{2};\ 0 < u < 2$	$\xi_0 = 2\alpha_0$	187.88	1.11
$f_\xi(u) \equiv 2;\ 0 < u < \frac{1}{2}$	$\xi_0 = \frac{\alpha_0}{2}$	205.41	1.21
$f_\xi(u) = 2u;\ 0 < u < 1$	$\xi_0 = \sqrt{\alpha_0}$	211.11	1.24
$f_\xi(u) = \frac{1}{\sqrt{1-u^2}};$ $0 < u < \sin 1$	$\xi_0 = \sin\alpha_0$	337.50	1.99
$f_\xi(u) = \cos u;\ 0 < u < \frac{\pi}{2}$	$\xi_0 = \arcsin\alpha_0$	442.31	2.60
$f_\xi(u) = 3u^2;\ 0 < u < 1$	$\xi_0 = \sqrt[3]{\alpha_0}$	1262.50	7.42
$f_\xi(u) = (\pi+1)u^2;\ 0 < u < 1$	$\xi_0 = \sqrt[(\pi+1)]{\alpha_0}$	1261.25	7.42

Заключение

В работе описано новое полезное применение специальной компьютерной системы моделирования вероятностных одномерных распределений NMPUD: это возможность объективного сравнения трудоемкостей основных арифметических операций и вычислений значений функций.

СПИСОК ЛИТЕРАТУРЫ

1. *Михайлов Г. А., Войтишек А. В.* Численное статистическое моделирование. Методы Монте-Карло. М.: Изд. центр «Академия», 2006.

2. *Войтишек А. В.* Лекции по численным методам Монте-Карло. Новосибирск: ИПЦ НГУ, 2018.
3. *Васильев Т. В., Постовалов Я. С., Черкашин Д. А.* Проект компьютерной системы для выбора и исследования моделируемых вероятностных распределений // Материалы 58-й Международной студенческой конференции «Студент и научно-технический прогресс». Математика, Новосибирск: ИПЦ НГУ, 2020. С. 112.
4. *Войтишек А. В., Постовалов Я. С., Черкашин Д. А.* Система численного моделирования одномерных случайных величин NMPUD: формирование банка плотностей, автоматизация математических выкладок и приложения // Материалы XIX Международной конференции имени А. Ф. Терпугова «Информационные технологии и математическое моделирование» (Томск, 2–5 декабря 2020 года), Томск: Изд-во НТЛ, 2021. С. 36–368.
5. *Войтишек А. В., Гаджиахмедов М. Г., Черкашин Д. А.* Компьютерное моделирование случайных величин по приближениям их плотностей распределения // Настоящий сборник.

Войтишек Антон Вацлавович — доктор физико-математических наук, профессор; ведущий научный сотрудник Института вычислительной математики и математической геофизики Сибирского отделения РАН; профессор кафедры вычислительной математики Новосибирского государственного университета; заведующий лабораторией математического моделирования лицея № 130 города Новосибирска. E-mail: *vav@osmf.sscc.ru*

Гаджиахмедов Мурад Гюлахмедович — студент механико-математического факультета Новосибирского государственного университета. E-mail: *m.gadzhiakhmedov@g.nsu.ru*

Рыжов Иван Антонович — ученик 10-М класса лицея № 130 города Новосибирска. E-mail: *ruzhovi2005@gmail.com*

Трофимов Иван Александрович — ученик 10-М класса лицея № 130 города Новосибирска. E-mail: *ivantrofimov052@gmail.com*

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН ПО ПРИБЛИЖЕНИЯМ ИХ ПЛОТНОСТЕЙ РАСПРЕДЕЛЕНИЯ

А. В. Войтишек[1,3], М. Г. Гаджиахмедов[2], Д. А. Черкашин[3]

[1]*Институт вычислительной математики и математической геофизики СО РАН, г. Новосибирск, Россия*
[2]*Новосибирский государственный университет, г. Новосибирск, Россия*
[3]*Лицей № 130 города Новосибирска, г. Новосибирск, Россия*

В данной работе проведен сравнительный анализ возможностей использования вместо вероятностной плотности распределения с трудоемкой моделирующей формулой кусочно-постоянной или кусочно-постоянной аппроксимаций этой плотности. Несмотря на экономичность моделирующего алгоритма для кусочно-постоянной плотности, следует рекомендовать использование более точной кусочно-линейной аппроксимации и соответствующего ей моделирующего алгоритма метода дискретной суперпозиции.

Ключевые слова: *трудоемкие формулы и алгоритмы моделирования одномерных случайных величин, кусочно-линейные и кусочно-постоянные приближения вероятностных плотностей.*

Введение

Представляемый здесь материал определенным образом связан со статьей [1] из данного сборника, в которой обосновывается необходимость построения и использования экономичных алгоритмов компьютерного моделирования выборочных значений ξ_i одномерных случайных величин $\xi \in (a, b)$ с заданными плотностями распределения

$$f_\xi(u), \quad u \in (a, b); \tag{1}$$

В свою очередь, эти алгоритмы применяются для численного решения обширного класса задач стохастического моделирования (см., например, [2, 3]).

Работа выполнена в рамках государственного задания ИВМиМГ СО РАН (проект 0251–2021–0002).

В работе [1], в том числе, показано, что даже для элементарной плотности (1), т. е. такой, что имеется возможность вывести аналитическую формулу метода обратной функции распределения

$$\xi_0 = F_\xi^{-1}(\alpha_0); \;\; F_\xi(x) = \int_{-\infty}^{x} f_\xi(u)\,du \tag{2}$$

(см. раздел 1.4.1 книги [2] и разделы 2.5, 2.6 книги [3]), моделирующая формула (2) может оказаться весьма трудоемкой (ярким примером здесь является степенное распределение). В формуле (2) $\alpha_0 \in U(0,1)$ – это стандартное случайное число (см. раздел 1.1 книги [2] и раздел 2.4 книги [3]).

В описанной ситуации может оказаться целесообразным использование вместо плотности (1) ее численных приближений. В работе [4] показано, что из всего спектра вычислительных конструкций приближения функций (см., например, [5]) по критерию, подразумевающему сочетание малой погрешности используемой аппроксимации с экономичностью моделирования случайной величины по приближению исходной плотности, наилучшей видится использование кусочно-линейной аппроксимации исходной плотности (1). Как показано в [1, 4], использование более гладких приближений ведет к увеличению трудоемкости соответствующего алгоритма метода суперпозиции, а также к сложностям в определении аппроксимационных коэффициентов с необходимыми свойствами (в частности, они должны быть положительными).

Можно, наоборот, использовать более экономичное при моделировании случайной величины кусочно-постоянное приближение исходной плотности, однако здесь ухудшаются аппроксимационные свойства приближения. Сравнительному анализу использования кусочно-линейного и кусочно-постоянного приближения плотности (1) при моделировании выборочных значений нужной случайной величины посвящена данная работа. В разделе 2 представлены алгоритмы модифицированного метода суперпозиции для кусочно-постоянной и кусочно-линейной плотностей. Особо выделен случай равных вероятностей в используемых методах суперпозиции. В разделе 3 приведен анализ погрешностей изучаемых приближений плотности (1). В разделе 4 (в заключении) приведены основные выводы работы.

1. Моделирование случайных величин с кусочно-постоянной и кусочно-линейной плотностями

Пусть требуется построить экономичные алгоритмы компьютерного моделирования выборочного значения $\bar{\xi}_0$ случайной величины $\bar{\xi} \in [a,b)$, имеющей кусочно-постоянную плотность распределения

$$f_{\bar{\xi}}(u) \equiv \bar{v}_i, \quad a_i \leqslant u < b_i; \quad i = 1, ..., M; \quad \bar{v}_i \geqslant 0 \tag{3}$$

и выборочного значения $\hat{\xi}_0$ случайной величины $\hat{\xi} \in [a, b)$, имеющей кусочно-линейную плотность распределения

$$f_{\hat{\xi}}(u) = \hat{v}_{i-1} + (u - a_i)\frac{\hat{v}_i - \hat{v}_{i-1}}{b_i - a_i}, \quad a_i \leqslant u < b_i; \quad i = 1, ..., M, \tag{4}$$

где $\hat{v}_i \geqslant 0;\ i = 1, ..., M$ и

$$-\infty < a = a_1 < b_1 = a_2 < b_2 < ... < b_{M-1} = a_M < b_M < +\infty. \tag{5}$$

Плотности (3), (4) являются частными случаями составной плотности вида

$$f_\xi(u) = \sum_{i=1}^{M} g_i(u)\chi^{[a_i, b_i)}(u); \quad u \in (a, b) \tag{6}$$

для разбиения (5); здесь $\chi^A(u)$ – индикатор множества A.

По аналогии с рассуждениями раздела 11.3 из книги [3], представим плотность (6) в виде

$$f_\xi(u) = \sum_{i=1}^{M} p_i f_i(u)\chi^{[a_i, b_i)}(u); \quad u \in (a, b) \tag{7}$$

где $f_i(u) = \frac{g_i(u)}{\int_{a_i}^{b_i} g_i(v)\,dv}$, $p_i = \int_{a_i}^{b_i} g_i(v)\,dv$; $u \in [a, b)$.

Для плотности (3) имеем $p_i = \bar{v}_i(b_i - a_i)$, $f_i(u) \equiv \frac{1}{b_i - a_i}$, а для плотности (4) $p_i = \frac{(\hat{v}_{i-1} + \hat{v}_i)(b_i - a_i)}{2}$, $f_i(u) = A_i u + B_i$ для $a_i \leqslant u < b_i$; $i = 1, ..., M$;

$$A_i = \frac{2(\hat{v}_i - \hat{v}_{i-1})}{(\hat{v}_{i-1} + \hat{v}_i)(b_i - a_i)^2}, \quad B_i = \frac{2(\hat{v}_{i-1} b_i - \hat{v}_i a_i)}{(\hat{v}_{i-1} + \hat{v}_i)(b_i - a_i)^2}.$$

Учитывая, что для составных плотностей (7) метод обратной функции распределения совпадает с модифицированным методом суперпозиции, в разделе 1.8.1 книги [2] и разделе 11.3 книги [3] для моделирования случайных величин согласно плотностям (3) и (4) предложено использовать следующие алгоритмы.

АЛГОРИТМ 1.

1. Моделируя стандартное случайное число $\alpha_0 \in U(0, 1)$ и используя наиболее эффективный (экономичный) из алгоритмов моделирования *целочисленной случайной величины* $\bar{\mu}$ *с распределением* $\mathbf{P}\{\bar{\mu} = i\} = p_i = \bar{v}_i(b_i - a_i);\ i = 1, ..., M$ *(см. раздел* 1.2 *книги* [2] *и раздел* 10 *книги* [3]*), моделируем значение* $\bar{\mu}_0(\alpha_0) = \bar{m}$.

2. Моделируем выборочное значение $\bar{\xi}_0$ *случайной величины* $\bar{\xi}$ *по формуле* $\bar{\xi}_0 = a_{\bar{m}} + (b_{\bar{m}} - a_{\bar{m}})\bar{\beta}_0(\alpha_0; \bar{m});\ \bar{\beta}_0(\alpha_0; \bar{m}) = \frac{\alpha_0 - \sum_{i=1}^{\bar{m}-1} p_i}{p_{\bar{m}}}$.

Для важного частного случая, когда значения $\bar{v}_i, a_i, b_i$ подобраны таким образом, что $p_i = \bar{v}_i(b_i - a_i) \equiv \frac{1}{M};\ i = 1, ..., M$, имеем следующие экономичные формулы для алгоритма 1: $\bar{m} = [M\alpha_0] + 1,\ \bar{\xi}_0 = a_{\bar{m}} + (b_{\bar{m}} - a_{\bar{m}})(M\alpha_0 - \bar{m} + 1)$; здесь $[A]$ – целая часть числа A.

АЛГОРИТМ 2.

1. Моделируя стандартное случайное число $\alpha_0 \in U(0, 1)$ *и используя наиболее эффективный (экономичный) из алгоритмов моделирования целочисленной случайной величины* $\hat{\mu}$ *с распределением* $\mathbf{P}\{\hat{\mu} = i\} = p_i = \frac{(\hat{v}_{i-1} + \hat{v}_i)(b_i - a_i)}{2};\ i = 1, ..., M$, *моделируем значение* $\hat{\mu}_0(\alpha_0) = \hat{m}$.

2. Моделируем выборочное значение $\hat{\xi}_0$ *случайной величины* $\hat{\xi}$ *по формуле* $\hat{\xi}_0 = \frac{-B_{\hat{m}} + \sqrt{(B_{\hat{m}} + A_{\hat{m}} a_{\hat{m}})^2 + 2A_{\hat{m}} \hat{\beta}_0(\alpha_0; \hat{m})}}{A_{\hat{m}}};\ \hat{\beta}_0(\alpha_0; \hat{m}) = \frac{\alpha_0 - \sum_{i=1}^{\hat{m}-1} p_i}{p_{\hat{m}}}$.

Для важного частного случая, когда значения $\hat{v}_i, a_i, b_i$ подобраны таким образом, что $p_i = \frac{(\hat{v}_{i-1} + \hat{v}_i)(b_i - a_i)}{2} \equiv \frac{1}{M};\ i = 1, ..., M$, имеем следующие экономичные формулы для алгоритма 2: $\hat{m} = [M\alpha_0] + 1$;

$$\hat{\xi}_0 = \frac{-B_{\hat{m}} + \sqrt{(B_{\hat{m}} + A_{\hat{m}} a_{\hat{m}})^2 + 2A_{\hat{m}}(M\alpha_0 - \hat{m} + 1)}}{A_{\hat{m}}}.$$

Достаточно очевидно, что для одного и того же M алгоритм 1 является более экономичным по сравнению с алгоритмом 2.

2. Сравнение погрешностей кусочно-постоянного и кусочно-линейного приближений плотности

Для простоты рассмотрим конечный интервал (т. е. $-\infty < a < b < +\infty$) и его разбиение (5) такое, что $b_i - a_i \equiv h = \frac{b-a}{M};\ i = 1, ..., M$ Полагаем также, что для плотности (1) формула (2) является трудоемкой, а также $f_\xi \in \mathbb{C}^2[a, b]$.

Рассмотрим для функции $f_\xi(u)$ кусочно-постоянное приближение

$$f_\xi(u) \approx \bar{g}(u) \equiv f_\xi\left(\frac{a_i + b_i}{2}\right) \tag{8}$$

и кусочно-линейное приближение

$$f_\xi(u) \approx \hat{g}(u) = f_\xi(a_i) + (u - a_i)\frac{f_\xi(b_i) - f_\xi(a_i)}{b_i - a_i} \tag{9}$$

для $a_i \leqslant u < b_i$; $i = 1, ..., M$. Заметим, что функции $\bar{g}(u)$ и $\hat{g}(u)$ из соотношений (8) и (9) не являются, вообще говоря, плотностями – они требуют нормировки. Реализуя соответствующую нормировку, получаем плотности вида (3) и (4):

$$f_{\bar{\xi}}(u) = \frac{f_\xi\left(\frac{a_i+b_i}{2}\right)}{\bar{G}}; \;\; \bar{G} = h\sum_{k=1}^{M} f_\xi\left(\frac{a_k+b_k}{2}\right) \;\; \text{и} \tag{10}$$

$$f_{\hat{\xi}}(u) = \frac{f_\xi(a_i)}{\hat{G}} + (u-a_i)\frac{f_\xi(b_i) - f_\xi(a_i)}{\hat{G}(b_i - a_i)}; \; \hat{G} = \frac{h}{2}\sum_{k=1}^{M}\left[f_\xi(a_k) + f_\xi(b_k)\right], \tag{11}$$

для $a_i \leqslant u < b_i$; $i = 1, ..., M$.

Моделирование выборочного значения $\bar{\xi}_0$ случайной величины $\bar{\xi}$, распределенной согласно плотности (10) следует производить согласно алгоритму 1 для $\bar{v}_i = \frac{f_\xi\left(\frac{a_i+b_i}{2}\right)}{\bar{G}}$. В свою очередь, моделирование выборочного значения $\hat{\xi}_0$ случайной величины $\hat{\xi}$, распределенной согласно плотности (11) следует производить согласно алгоритму 2 для $\hat{v}_{i-1} = \frac{f_\xi(a_i)}{\hat{G}}$, $\hat{v}_i = \frac{f_\xi(b_i)}{\hat{G}}$.

Очевидно, что замена плотности $f_\xi(u)$ на плотность $f_{\bar{\xi}}(u)$ из соотношения (10) или на плотность $f_{\hat{\xi}}(u)$ из соотношения (11) дает некоторое искажение исходного распределения.

Оценим соответствующие погрешности $\delta^{(\mathbb{C})}_{appr} = \|f_\xi - f_\eta\|_{\mathbb{C}[a,b]} = \sup\limits_{u\in[a,b]} |f_\xi(u) - f_\eta(u)|$; $\eta = \bar{\xi} \vee \hat{\xi}$. Для $\eta = \bar{\xi}$ получаем неравенство

$$\bar{\delta}^{(\mathbb{C})}_{appr} = \|f_\xi - f_{\bar{\xi}}\|_{\mathbb{C}[a,b]} \leqslant \frac{2\max\limits_{u\in[a,b]}|f_\xi(u)| \times Hh^2 + \max\limits_{u\in[a,b]}|f'_\xi(u)|h}{2(1-Hh^2)}, \tag{12}$$

где $H = \frac{(b-a)\max\limits_{u\in[a,b]}|f''_\xi(u)|}{24}$. Верхняя граница погрешности из правой части соотношения (12) имеет первый порядок по h при $h \downarrow 0$.

В свою очередь, для $\eta = \hat{\xi}$ имеем

$$\hat{\delta}^{(\mathbb{C})}_{appr} \leqslant \frac{\max\limits_{u\in[a,b]}|f''_\xi(u)|\left[2(b-a)\max\limits_{u\in[a,b]}|f_\xi(u)| + 3\right]h^2}{24(1-2Hh^2)}, \tag{13}$$

Верхняя граница погрешности из правой части соотношения (13) имеет второй порядок по h при $h \downarrow 0$.

Заключение

В данной работе проведен сравнительный анализ возможностей использования вместо вероятностной плотности распределения (1) с трудоемкой моделирующей формулой (2) кусочно-постоянной или кусочно-постоянной аппроксимаций этой плотности. Приведенные здесь соображения и проведенные нами тестовые расчеты показывают, что, несмотря на экономичность алгоритма 1 для кусочно-постоянной плотности (10), следует рекомендовать использование более точной кусочно-линейной аппроксимации (11) и соответствующего ей алгоритма 2.

СПИСОК ЛИТЕРАТУРЫ

1. *Войтишек А. В., Гаджиахмедов М. Г., Рыжов И. А., Трофимов И. А.* Сравнение вычислительных затрат на основные математические операции с помощью системы NMPUD // Настоящий сборник.
2. *Михайлов Г. А., Войтишек А. В.* Численное статистическое моделирование. Методы Монте-Карло. М.: Изд. центр «Академия», 2006.
3. *Войтишек А. В.* Лекции по численным методам Монте-Карло. Новосибирск: ИПЦ НГУ, 2018.
4. *Булгакова Т. Е.* Оптимизация функциональных вычислительных статистических оценок и алгоритмов. Новосибирск, 2020 (Диссертация на соискание ученой степени кандидата физико-математических наук / ИВМиМГ СО РАН).
5. *Бахвалов Н. С.* Численные методы. М.: Наука, 1975.

Войтишек Антон Вацлавович — доктор физико-математических наук, профессор; ведущий научный сотрудник Института вычислительной математики и математической геофизики Сибирского отделения РАН; профессор кафедры вычислительной математики Новосибирского государственного университета; заведующий лабораторией математического моделирования лицея № 130 города Новосибирска. E-mail: *vav@osmf.sscc.ru*

Гаджиахмедов Мурад Гюлахмедович — студент механико-математического факультета Новосибирского государственного университета. E-mail: *m.gadzhiakhmedov@g.nsu.ru*

Черкашин Данил Андреевич — ученик 11-ЕН класса лицея № 130 города Новосибирска. E-mail: *cherkashin.daniel@gmail.com*

МОДЕЛИРОВАНИЕ ПЕРЕХОДНЫХ ПРОЦЕССОВ ГРУЗОВЫХ ЖЕЛЕЗНОДОРОЖНЫХ СТАНЦИЙ НА ОСНОВЕ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

М. Л. Жарков, А. Л. Казаков, А. А. Лемперт

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

Работа продолжает проведенные ранее авторами исследования, в ходе которых была предложена методика моделирования работы грузовых железнодорожных станций, основанная на теории массового обслуживания. Для выбранного объекта строится математическое описание, имеющее вид многофазной системы массового обслуживания с BMAP-потоком, для которой численно определяются основные стационарные характеристики. Однако известно, что при исследовании работы сложных технических систем нередки случаи, когда стационарных характеристик недостаточно, и требуется также рассмотреть свойства протекающих в них переходных процессов. В связи с этим было решено выполнить соответствующее исследование для разработанной модели. С этой целью проведено построение и исследование систем дифференциальных уравнений Колмогорова, описывающих зависимость вероятностей состояний от времени. Предложенный подход применен при анализе работы типовой грузовой железнодорожной станции, для которой определяются характер переходных процессов и скорость их выхода на стационарный режим.

Ключевые слова: *Система массового обслуживания, BMAP-поток, система дифференциальных уравнений Колмогорова, грузовая железнодорожная станция, переходный процесс.*

Введение

Теория массового обслуживания (ТМО) является эффективным инструментом при исследовании работы технических систем, подверженных воздействию случайных факторов, возникающих в различных сфе-

Исследование выполнено в рамках госзадания Минобрнауки России по проекту «Теоретические основы, методы и высокопроизводительные алгоритмы непрерывной и дискретной оптимизации для поддержки междисциплинарных научных исследований» (№гос регистрации: 121041300065-9), при поддержке РФФИ и Правительства Иркутской области в рамках проекта № 20-47-383002.

рах человеческой деятельности: информационно-телекоммуникационные системы, производство, транспорт и т.д. [1, 2, 3]. Ранее авторами с использованием данного математического аппарата была разработана методика моделирования работы грузовых и сортировочных железнодорожных станций (ЖДС) [4]. Применяющиеся в ней многофазные систем массового обслуживания (СМО) [1] с $BMAP$-потоком [5] позволяют детально описать несколько входящих транспортных потоков и маршрут движения транспорта внутри объекта с иерархической структурой, благодаря чему, во-первых, достигается приемлемая точность моделей; во-вторых, полученные модели обладают универсальностью и могут быть адаптированы под разные типы ЖДС за минимальное время.

В ходе предыдущих исследований моделировалась работа ЖДС в стационарном режиме. Однако данные объекты представляют собой динамические системы, параметры которых меняются со временем. Прежде всего это касается интенсивности входящих транспортных потоков. Как результат, в работе станций возникают переходные процессы. Соответственно, исследователю необходимо определить: а) как долго эти процессы будут длиться до того, как система перейдет в установившийся (стационарный) режим работы; б) насколько сильно характеристики работы в переходном режиме отличаются от стационарных. В данной статье строится модель работы грузовой ЖДС в виде трехфазной СМО, находятся вероятности ее состояний как функции времени, и на основе их анализа делаются выводы о свойствах переходных процессов в рассматриваемой технической системе.

1. Объект исследования

Грузовые ЖДС выполняют погрузочно-разгрузочные и маневровые работы по подаче вагонов на грузовые фронты и их уборке. Типовая схема грузовой ЖДС включает следующие основные подсистемы: парк прибытия и отправления (ППО), сортировочный парк (СП) и грузовой двор (ГД). На крупных станциях парк прибытия и отправления может быть разделен на две отдельные подсистемы. При этом парка отправления может и не быть, а его функции выполнять соседняя ЖДС.

Поезда поступают на ЖДС с двух и более направлений и принимаются только в ППО. Далее они обрабатываются в СП, переставляются в ГД для погрузки/разгрузки, затем обратно в ППО для отправления со станции. Подсистемы имеют различные параметры: число обслуживаемых устройств, тип выполняемых в них операций, вместимость вагонов. В ППО основным устройством является сортировочная горка, в СП – маневровые локомотивы, в ГД – устройства погрузки/разгрузки. Входящий поездопоток и работа подсистем подвержены влиянию множества различных негативных случайных факторов: поломки техники, ошибки персонала и т.д., что приводит к нарушению ритмичности работы и вызывает необходимость рассмотрения вероятностных моделей.

2. Обобщенная математическая модель и система дифференциальных уравнений Колмогорова

Согласно предложенной ранее методике [4] каждый вагон считается заявкой на обслуживание, а поезд — группой заявок. Они моделируются с помощью $BMAP$-потока, что позволяет описать единой моделью вагонопотоки, поступающие с нескольких направлений и имеющие различные параметры. Моделями подсистем являются СМО, в которых количество каналов определяется числом обслуживающих устройств, а длина очереди — суммарной вместимостью вагонов в подсистеме. Рассмотрим только одно направление движения вагонов: ППО–СП–ГД. Здесь учитываются основные операции с поездами: принятие, обработка в СП и погрузка/разгрузка. Операции по отправлению поезда со станции выполняются значительно быстрее и не влияют на работу других подсистем. Тогда в терминах ТМО модель грузовой ЖДС будет иметь вид трехфазной СМО $BMAP/M/n_1/m_1 \ \rightarrow \ */M/n_2/m_2 \ \rightarrow \ */M/n_3/m_3$.

Для исследования переходных процессов в СМО рассмотрим систему дифференциальных уравнений уравнений Колмогорова (СДУК) [6], в которой неизвестными функциями являются вероятности состояний системы $p_i(t), i = 0, 1, \ldots, K$ (далее — переходные вероятности состояний), где K — максимальное число заявок в выбранной СМО, а независимой переменной — время. Начальными данными является состояние системы в нулевой момент времени. Ранее авторами были получены уравнения для составления СДУК для трехфазных СМО с $BMAP$-потоком [7]. Отметим, что они могут быть применены для СМО, в которых потоки между фазами являются пуассоновскими. Далее рассматривается СМО, в которой а) интенсивность входящего потока ниже, чем интенсивность обслуживания на всех фазах, т.е. практически все заявки будут обслужены; б) времена между поступлением заявок, обслуживания в фазах и блокировки каналов подчиняется экспоненциальному распределению. В этом случае сделаем упрощающее модельное предположение, что потоки между фазами обладают требуемыми свойствами. Поэтому эти формулы используются при составлении СДУК. Поскольку число каналов измеряется десятками, а длина очереди — сотнями единиц, необходимо понизить детализацию модели, чтобы уменьшить размерность СДУК. Далее при построении модели ЖДС примем, что весь поезд является одной заявкой на обслуживание. Даже в этом случае СДУК будет включать более сотни уравнений, поэтому для ее анализа применяются численные методы. Наиболее популярный из них — методы Рунге-Кутты [8].

3. Вычислительный эксперимент

Численное исследование подразумевает анализ модели конкретного объекта. Поэтому рассмотрим модельную грузовую ЖДС, параметры которой близки к техническим характеристикам действующей в Иркутской области типовой грузовой ЖДС Суховская (ВСЖД). Она является вполне типичной для Российских железных дорог, кроме того авторам удалось получить данные о ее структуре и входящих поездо- и вагонопотоках. В парке прибытия имеется одна сортировочная горка, его вместимость — 5 поездов. В СП действуют два маневровых локомотива, вместимость — 8 поездов. В ГД имеется 3 канала, а очередь отсутствует. В терминах ТМО получим трехфазную СМО $BMAP/M/1/5 \to */M/2/8 \to */M/3/0$. По натурному обследованию установлено, что на станцию в среднем в сутки поступает 3 поезда с четного и 2 поезда с нечетного направлений. Тогда матрицы $BMAP$ будут иметь вид $D_0 = \begin{pmatrix} -0,208 & 0 \\ 0 & -0,208 \end{pmatrix}$, $D_1 = \begin{pmatrix} 0,1248 & 0,0832 \\ 0,1248 & 0,0832 \end{pmatrix}$.

СДУК для данной трехфазной СМО будет включать $(K_1+1)(K_2+2)(K_3+3)-K_2 = 408$ уравнений. Конкретный вид полученной системы приводить не будем из-за ее большого размера.

Для построения решения использовался метод Рунге-Кутты четвертого порядка при параметрах $\mu_1 = 0{,}74, \mu_2 = 0{,}222, \mu_3 = 0{,}083$ (поездов в час), шаге 0,05 и начальном распределении вероятностей, которое определяет условия Коши, $p_{000}(0) = 1, p_{i,j,z}(0) = 0, i = \overline{0,6}, j = \overline{0,11}, z = \overline{0,4}$, где i, j и z — число заявок на фазе 1, 2 и 3 соответственно. Состояния $p_{i,11,z}(t)$ и $p_{i,j,4}(t)$ обозначают блокировку каналов фаз 1 и 2 соответственно. В результате были найдены переходные вероятности состояний рассматриваемой СМО. Далее на рисунке представлены графики изменения вероятности отказа за $t = [0,200]$ при различных начальных состояниях системы: 1) $p_{1,9,3}(0) = 1$, 2) $p_{000}(0) = 1$, 3) $p_{2,5,0}(0) = 1$, 4) $p_{3,5,0}(0) = 1$, 5) $p_{5,11,3}(0) = 1$ — блокировка каналов фазы 1, 6) $p_{6,10,1}(0) = 1$. В правом верхнем углу приведены линии 1)–4) в увеличенном по оси ординат масштабе.

Анализ переходных вероятностей состояний показывает, что их поведение и скорость сходимости к стационарным параметрам существенно зависит от начального состояния системы. Во-первых, минимальное время, за которое переходные вероятности приблизятся к стационарным с отклонением 0,04%, достигается при средней загрузке системы, а максимальное — при наименьшей/наибольшей. Так, при начальном

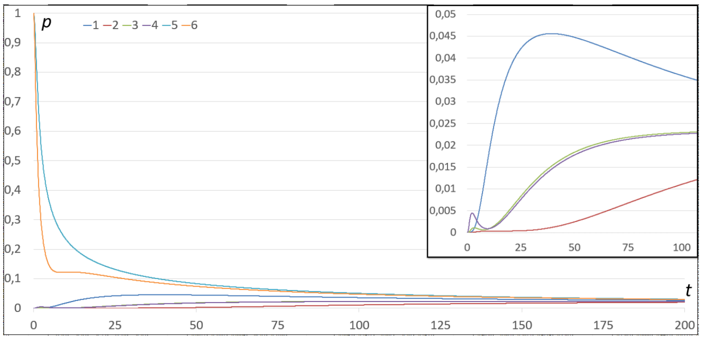

Рис. 1. Вероятность отказа при различных начальных состояниях СМО

состоянии $p_{3,5,0}(0) = 1$ время сходимости составит $t = 92,5$ (часов), при $p_{000}(0) = 1$ или $p_{6,10,1}(0) = 1$ оно будет равно $t = 200$. Кроме того при средней загрузке системы график вероятности отказа имеет экстремумы (см. линии 3 и 4), как правило, локальные. Однако имеется и особый случай: если при $t = 0$ первая фаза свободна, а остальные загружены, в частности, при $p_{1,9,3}(0) = 1$, то появляется глобальный максимум, значение которого может в два раза превышать стационарное (линия 1).

Перенося результаты исследования на модельную станцию, получаем следующее. Во-первых, продолжительность переходных процессов является значимым фактором для такого типа систем. Так, минимальное время входа в стационарный режим составляет более четырех суток. Во-вторых, необходимо учитывать загрузку системы в начальный момент времени для оценки ее пропускной способности. В частности, после устранения аварии на грузовом дворе, что соответствует случаю полной загрузки фаз 2 и 3, период восстановления работы станции составляет более пяти суток, а вероятность отказа увеличивается в два раза по сравнению с предельными значениями. Таким образом, при изучении работы ЖДС можно не учитывать переходные процессы а) при введении станции в эксплуатацию, т.е. система практически пуста, б) в периоды бесперебойной ее работы. В остальных случаях для полноты исследования необходим анализ переходных процессов.

Заключение

В работе был проведен анализ переходных процессов для трехфазной СМО, моделирующей работу грузовой железнодорожной станции, в результате которого установлено, что поведение и скорость сходимости к стационарным параметрам существенно зависят от начальных состояний систем. Так, при средней или максимальной загрузке станции в начальный момент времени необходимо учитывать влияние переходных процессов. В случае же минимальной загрузки ими можно пренебречь при прогнозировании работы системы.

Представленные результаты составляют только часть проведенного исследования. В докладе помимо более детального рассмотрения модели грузовой станции, будут изучено влияние временного увеличения интенсивности входящего поездопотока на поведение переходных процессов в модели данного объекта. Дальнейшая работа предполагает сопоставление полученных результатов с реальными данными и применение моделей для решения практических задач.

СПИСОК ЛИТЕРАТУРЫ

1. *Вишневский В. М.* Теоретические основы проектирования компьютерных сетей. М.: Техносфера, 2003. 512 с.
2. *Поттгофф Г.* Учение о транспортных потоках. М.: Транспорт, 1979. 324 с.
3. *Wilson N, Fourie C. J., Delmistro R.* Mathematical and simulation techniques for modelling urban train networks // South Afr. J. Ind. Eng. 2016. № 27. pp. 109–119.
4. *Лемперт А. А., Павидис М. М., Жарков М. Л.* Моделирование зарубежных сортировочных станций на основе теории массового обслуживания // Вестник УрГУПС. 2019. № 3. С. 24–32.
5. *Дудин А. Н., Клименок В. И.* Системы массового обслуживания с коррелированными потоками. Мн.: БГУ, 2000. 176 с.
6. *Bolch G, (eds.)* Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications. New York: John Wiley and sons, 2006. 726 p.
7. *Zharkov M. L., Kazakov A. L., Lempert A. A.* Transient process modeling in micrologistic transport systems // IOP Conf. Series: Earth and Environmental Science. 2021. № 629. pp. 012023.
8. *Butcher J, C.* Numerical Methods for Ordinary Differential Equations. New York: John Wiley and Sons, 2006. 482 p.

Жарков Максим Леонидович — к.т.н., н.с., Лаб. 6.2. Логических и оптимизационных методов анализа сложных систем. E-mail: *zharkm@mail.ru*

Казаков Александр Леонидович — д.ф.-м.н., г.н.с., Отд. 2. Прикладных проблем математической физики и теории поля. E-mail: *kazakov@icc.ru*

Лемперт Анна Ананьевна — к.ф.-м.н., зав.лаб., Лаб. 6.2. Логических и оптимизационных методов анализа сложных систем. E-mail: *lempert@icc.ru*

ОБ АЛГОРИТМЕ КЛАСТЕРИЗАЦИИ ЗНАКОВОГО ГРАФА

Э. И. Ибрагимова

Сибирский федеральный университет, г. Красноярск, Россия

В работе рассматривается задача кластеризации знакового графа. Задача кластеризации предполагает нахождение разбиения знакового графа с минимальной суммарной ошибкой. Ошибка складывается из внутрикластерной и межкластерной ошибок в зависимости от параметра. Предложен алгоритм решения данной задачи для произвольного значения параметра. Алгоритм является развитием двух предложенных ранее алгоритмов решения двух частных случаев решения рассматриваемой задачи кластеризации. Приведены вычислительные эксперименты, демонстрирующие эффективность предложенного алгоритма.

Ключевые слова: *Кластеризация, знаковый граф, знаковые сети.*

Введение

Знаковые графы были впервые введены F. Harary в 50-х годах XX-го века при изучении проблем социальной психологии. Важнейшие результаты в исследовании знаковых графов были получены такими учеными как F. Harary [1], T. Zaslavsky [2], P. Doreian [3, 4] и многими другими. В работах этих авторов исследовались следующие задачи знакового баланса: распознавание сбалансированности, поиск максимального сбалансированного подграфа, определение индекса и меры сбалансированности, задача кластеризации знакового графа. На практике задачи знакового баланса находят свое применение при анализе мультиагентных систем [5], социальных сетей [6] и тематическом моделировании текстов [7].

В работе рассматривается одно из востребованных приложений знакового графа задача разбиения (кластеризации) знакового графа [3, 4]. Вид решения задачи кластеризации существенно зависит от выбора функционала качества кластеризации или функционала ошибки. Выделяют следующие вариации задачи. Когда количество кластеров k задано, то k может выступать параметром кластеризации, тогда решается задача о разбиении ровно на k кластеров или не более (не менее) чем на k кластеров. Когда количество кластеров k не задано, то оно является решением задачи [8]. В данной работе рассматривается случай, когда количество кластеров является решением задачи. В общем виде задача кластеризации знакового графа является NP-полной [7]. В работе [4] предложены эвристические методы для ее решения.

В разделе 1 приведены основные определения и постановки задач распознавания k-сбалансированности и кластеризации знакового графа. Введена функция ошибки, зависящая от параметра α. Параметр α регулирует важность внутрикластерной и межкластерной ошибок. В разделе 2 приведён алгоритм для решения задачи кластеризации знакового графа, основанный на разработанных ранее алгоритмах для решения задачи при $\alpha = 0$ и $\alpha = 1$. В разделе 3 представлены результаты вычислительных экспериментов.

1. Общие положения

Знаковым графом называется пара $\Sigma = (G, \sigma)$, где $G = (V, E)$ — неориентированный граф ($n = |V| \geq 2, m = |E| \geq 1$), на ребрах которого задана функция знака $\sigma : \ \rightarrow \{+, -\}$. Знаковый граф называется k-сбалансированным, если множество его вершин можно разбить на k попарно непересекающихся непустых кластеров так, что все положительные рёбра находятся внутри кластеров, а отрицательные между кластерами [2, 3]. Задача распознавания k-сбалансированности знакового графа формулируется следующим образом [3].

Условие: задан знаковый граф $\Sigma = (G, \sigma)$, где $G = (V, E)$ — неориентированный граф, $n = |V| \geq 2, m = |E| \geq 1$.
Вопрос: при каких k граф $\Sigma = (G, \sigma)$ является k-сбалансированным?

Данную задачу можно решить за полиномиальное время поиском в ширину [9]. Так как знаковые графы не всегда являются k-сбалансированными, возникает задача поиска такого разбиения, для которого некоторый функционал ошибки был бы минимальным.

Введем необходимые обозначения. Будем говорить, что система множеств $C = \{V_1, V_2, \ldots, V_k\}$ образует разбиение множества вершин V на k подмножеств, где k — параметр разбиения, если подмножества непустыеи: $V_i = \varnothing$, $i = 1, \ldots, k$, попарно не пересекаются: $V_i \cap V_j = \varnothing$, $i \neq j$, и их объединение равно V: $V = V_1 \cup V_2 \cup \cdots \cup V_k$.

Обозначим $\Phi = \cup_{k=1}^{n} \Phi_k$ множество всех возможных разбиений V, где Φ_k — множество возможных разбиений V ровно на k подмножеств. Пусть C — какое-либо разбиение графа из Φ. Под отрицательной ошибкой $N(C)$ будем понимать число отрицательных ребер внутри подмножеств, а под положительной $P(C)$ — число положительных ребер между подмножествами. Критерием качества разбиения будет являться функция суммарной ошибки — выпуклая комбинация положительной и отрицательной ошибок:

$$E_\alpha(C) = \alpha N(C) + (1-\alpha)P(c), \alpha \in [0,1] \tag{1}$$

Тогда задача кластеризации знакового графа может быть сформулирована следующим образом [3].

Условие: задан знаковый граф $\Sigma = (G, \sigma)$, где $G = (V, E)$ — неориентированный граф, $n = |V| \geq 2, m = |E| \geq 1$.

Вопрос: для заданного $\alpha \in [0,1]$ требуется найти разбиение C множества вершин V знакового графа Σ с минимальной суммарной ошибкой $E_\alpha(C)$

Данная задача может иметь множественное решение.

В работе [7] показано, что задача минимизации суммарной ошибки является NP-полной.

В [4] предложены алгоритмы решения данной задачи. Основная идея предложенных методов состоит в том, что берется некоторое начальное разбиение множества вершин графа, а затем вершины перемещаются из одного кластера в другой, либо две вершины из разных кластеров меняются местами пока не будет минимизирована функция суммарной ошибки для некоторого α. В зависимости от метода меняется правило, по которому выбираются перемещаемые вершины и кластеры.

Решением задачи кластеризации знакового графа будет множество кластеров C^*, доставляющих минимум функции ошибки:

$$C^* = \arg\min_{C \in \Phi} \big[\alpha N(C) + (1-\alpha)P(c)\big] \tag{2}$$

и $k = |C^*|$ — количество кластеров. При $\alpha = 0$ и $\alpha = 1$ данная задача вырождается в полиномиально разрешимые случаи минимизации положительной и отрицательной ошибок, соответственно. Более подробно данные случаи рассмотрены в [9].

2. Алгоритм SGClustα для решения задачи кластеризации

На рисунке 1 приведена схема алгоритма SGClustα для решения задачи кластеризации знакового графа.

Алгоритм является двухфазным. На первой фазе *Initialization* для заданного графа Σ с помощью удаления отрицательных ребер строится граф Σ^+ и выделяются компоненты связности $\mathbf{C}$. Для этого необходимо выполнить не более чем $O(n+m)$ операций.

На второй фазе *Minimalization* функция *Separate*($\mathbf{C}$) просматривается каждый кластер и выделяет в отдельные кластеры вершины с наибольшим количеством инцидентных отрицательных ребер внутри кластера, если эти вершины еще не были посещены этой функцией,

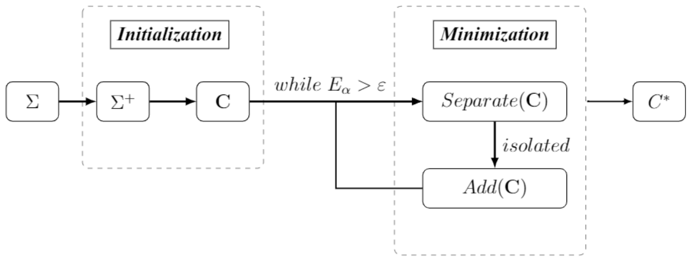

Рис. 1. Схема работы алгоритма SGClustα

а также составляет список изолированных вершин $isolated$. Для этого необходимо не более чем $O(nm)$ операций.

После этого, если список изолированных вершин непустой, то для каждой вершины из этого списка функция $Add(\mathbf{C})$ проверяет, уменьшится ли ошибка при добавлении этой вершины к какому-либо кластеру. Если ошибка уменьшится, то вершина присоединяется к этому кластеру. Для этого необходимо просмотреть все инцидентные вершине ребра, что требует для всего графа не более чем $O(m)$ операций.

Функции $Separate(\mathbf{C})$ и $Add(\mathbf{C})$ чередуются до тех пор, пока не будет достигнута заданная точность ε, либо не будут посещены все вершины, либо не будет выполнено заданное количество итераций. В худшем случае необходмо будет повторить n раз.

Теорема 1. Вычислительная сложность алгоритма $SGClust\alpha$ $O()$ равна $O(n + m + n(nm + n)) = O(n^4)$.

3. Вычислительные эксперименты

Для оценки эффективности алгоритмов были проведены вычислительные эксперименты, результаты которых приведены в таблицах 1, 2, 3. Для проверки использовались графы, сгенерированные псевдослучайным образом. Эксперименты для алгоритма $SGClust\alpha$ проводились при $\alpha = 0,5$. Эксперименты проводились на ПК с процессором Intel®Core i7-10510U CPU @ 1.80ГГц и оперативной памятью объемом 8ГБ.

В таблице 1 представлены результаты работы алгоритма $SGClust\alpha$ при различных значениях параметра α. В таблице 2 сравнивается алгоритм $SGClust0$ и общий алгоритма $SGClust\alpha$ при $\alpha = 0$. В таблице 3 сравнивается алгоритм $SGClust1$ и общий алгоритма $SGClust\alpha$ при $\alpha = 1$.

Результаты экспериментов показывают, что для частных случаев алгоритмы $SGClust0$ и $SGClust1$ работают быстрее. В некоторых случаях алгоритм $SGClust\alpha$ не успевает найти оптимальное решение за приемлемое время.

Таблица 1

Результаты работы алгоритма $SGlust\alpha$ для разных значений α

n	m	$SGClust\alpha$, $\alpha = 0.25$			$SGClust\alpha$, $\alpha = 0.5$			$SGClust\alpha$, $\alpha = 0.75$		
		k	$E_{0.25}$	t	k	$E_{0.5}$	t	k	$E_{0.75}$	t
5	6	4	0.75	938	4	0.5	943	4	0.25	943
7	9	3	0.75	1	3	0.5	2	3	0.25	2
10	19	6	0	2	6	0	3	6	0	5
14	10	10	0	4	10	0	4	10	0	9
15	27	12	39.75	3815	12	26.5	3841	7	8.5	3
20	42	9	7.5	2719	9	5	2746	9	2.5	2755

Таблица 2

Результаты сравнения алгоритмов $SGClust0$ и $SGClust\alpha$ при $\alpha = 0$

n	m	$SGClust0$			$SGClust\alpha$, $\alpha = 0$		
		k	E_0	t	k	E_0	t
7	21	2	0	1	2	0	3
8	28	1	0	1	1	0	1
8	28	1	0	1	2	5	14552
15	12	11	0	1	11	0	5
15	105	1	0	1	1	0	4
30	50	5	0	1	5	0	5

Таблица 3

Результаты сравнения алгоритмов $SGClust0$ и $SGClust\alpha$ при $\alpha = 1$

n	m	$SGClust1$			$SGClust\alpha$, $\alpha = 1$		
		k	E_1	t	k	E_1	t
7	21	7	0	3	4	0	4
15	12	11	0	2	11	0	8
15	105	15	0	1	7	0	10
30	50	11	0	1	8	0	8
100	9750	93	0	2	99	0	165
500	24925	488	0	36	494	0	26649

Заключение

Для рассматриваемой задачи кластеризации знакового графа при произвольном параметре $0 < \alpha < 1$ разработан алгоритм $SGClust\alpha$. Приведенные вычислительные эксперименты, демонстрируют сложность данного алгоритма. Одним из перспективных направлений дальнейших исследований является оптимизация разработанного алгоритма с целью повышения его производительности, а также исследование задачи кластеризации при заданном количестве кластеров.

СПИСОК ЛИТЕРАТУРЫ

1. *Kartwright D., Harary F.* Structural balance: a generalization of Heider's theory // The Psychological Review. 1956. Vol. 63. № 5. P. 277–293.
2. *Zaslavsky T.* Balance and clustering in signed graphs // Unpublished manuscript, New York. 2010.
3. *Doreian P., Mrvar A.* Structural balance and partitioning signed graphs // In A. Ferligoj & A. Kramberger (Eds.), Developments in data analysis (Metodoloskizvezki 12, P. 195–208). Ljubljana, Slovenia: FDV.
4. *Brusco M. J., Doreian P.* Partitioning signed networks using relocation heuristics, tabu search, and variable neighbor-hood search // Social Networks. 2019. № 56. P. 70–80.
5. *Bo Liu, Housheng Su, Licheng Wu, Xixi Shen* Observability of Leader-Based Discrete-Time Multi-Agent Systems Over Signed Networks // IEEE Transactions On Network Science And Engineering. January-March 2021. Vol. 8. № 1. P. 25–39.
6. *Aparna Sankaran Srinath, Hannah Johnson, Gaby G. Dagher, Min Long* BullyNet: Unmasking Cyberbullies on Social Networks // IEEE Transactions On Computational Social Systems. April 2021. Vol. 8. № 2. P. 332–344.
7. *Nikhil Bansal, Avrim Blum, Shuchi Chawla* Correlation clustering // Machine Learning. 2004. Vol. 56. P. 89–113.
8. *Ильев В. П., Ильева С. Д., Моршиин А. В.* Алгоритм приближенного решения одной задачи кластеризации графа // Прикладная дискретная математика. 2019. Vol. 45. P. 64–77.
9. *Ибрагимова Э. И., Семенова Д. В.* Распознавание k-кластеризуемости знаковых графов // Информационные технологии и математическое моделирование (ИТММ-2020). Материалы XIX Международной конференции имени А. Ф. Терпугова. 2020. P. 468–472.

Ибрагимова Эллада Ибрагимовна — магистрант 2-го года обучения института математики и фундаментальной информатики Сибирского федерального университета. E-mail: *ibragimovaei@mail.ru*

ОБ ОДНОМ КРИТЕРИИ СХОДИМОСТИ К ЭКСПОНЕНЦИАЛЬНОМУ ЗАКОНУ

Азам А. Имомов, Эркин Э. Тухтаев, Абдимумин Абдирахманов

Каршинский государственный университет,
г. Карши, Узбекистан

В настоящей заметке исследуются один критерий справедливости теоремы о сходимости к экспоненциальному закону и локально-дифференциальный аналог основной леммы теории критических процессов Гальтона-Ватсона. Обсуждается некоторые их применения.
Ключевые слова: *Экспоненциальный закон, локально- дифференциальный аналог основной леммы теории критических процессов Гальтона-Ватсона, Q-процессы.*

В современных исследованиях, при доказательстве предельных теорем, наряду с теоремой непрерывности часто используется характеристические свойства закона распределения. Когда эти свойства представляют собой дифференциальное уравнение, которому удовлетворяет характеристическая функция или преобразование Лапласа (ПЛ) предельного закона, утверждение теоремы заключается в устойчивости решения этого уравнения. В случае нормальной аппроксимации такой подход к задаче составляет метод Стейна-Тихомирова (МСТ); см. [1]. В работе [2] Ш. К. Формановым был предложен упрощенный вариант МСТ, реализованный путем определения одного дифференциального оператора для характеристической функции. С помощью этого метода им был установлен критерий справедливости неклассической центральной предельной теоремы.

В настоящей заметке будет установлен критерий сходимости к экспоненциальному закону, основываясь на идеи работы [2]. Будет доказан также так называемый локально-дифференциальный аналог основной леммы теории критических ветвящихся процессов Гальтона-Ватсона (ГВП). Мы заметим полезность применения полученных результатов в исследованиях теории ГВП и $Q-$процессов.

Рассмотрим функцию распределения экспоненциального закона

$$\Gamma_\alpha(x) = \begin{cases} 1 - e^{-\alpha x}, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

где $\alpha > 0$. Следуя Форманову [2], введем следующий класс ПЛ:

$$\Phi = \left\{ \psi(\theta) : \ |\psi'(0)| = \frac{1}{\alpha}, \ \theta > 0 \right\}.$$

В классе Φ рассмотрим оператор

$$\Delta\left(\psi(\theta)\right) := \psi'(\theta) \ + \ \frac{1}{\alpha}\, \psi^2(\theta). \tag{1}$$

Поскольку ПЛ экспоненциального закона

$$\psi_\alpha(\theta) := \int_0^\infty e^{-\theta x} d\Gamma_\alpha(x) = \frac{1}{[1+\theta/\alpha]},$$

то мы легко заметим, что

$$\Delta\left(\psi_\alpha(\theta)\right) = 0, \tag{2}$$

то есть оператор $\Delta\,(*)$ «аннулирует» ПЛ распределения $\Gamma_\alpha(x)$.

Теорема 1. Пусть $\{G_n(x), \ n \in \mathbf{N}\}$ $(\mathbf{N} = 1, 2, \ldots)$, $x \geqslant 0$, есть последовательность функций распределений и соответствующие им ПЛ $\psi_n(\theta)$ принадлежат классу Φ. Для того, чтобы при $n \to \infty$

$$\sup_x |G_n(x) - \Gamma_\alpha(x)| \to 0, \tag{3}$$

необходимо и достаточно, чтобы

$$\sup_{\theta \leqslant \Theta} |\Delta\left(\psi_n(\theta)\right)| \to 0, \quad n \to \infty, \tag{4}$$

при любом $\Theta > 0$.

Доказательство. Рассуждение по части необходимости условия (4) основывается на свойствах ПЛ. Из (1) и (2) следует, что

$$\begin{aligned} |\Delta(\psi_n(\theta))| &= |\Delta(\psi_n(\theta)) - \Delta(\psi_\alpha(\theta))| \\ &\leqslant |\psi_n'(\theta) - \psi_\alpha'(\theta)| + \frac{1}{\alpha}\left|\psi_n^2(\theta) - \psi_\alpha^2(\theta)\right|. \end{aligned} \tag{5}$$

Поскольку функция распределения и ПЛ являются ограниченными функциями, то дифференцирование и интегрирование по частям нам дает

$$\begin{aligned} |\psi_n'(\theta) - \psi_\alpha'(\theta)| &= \left| \int_0^\infty x e^{-\theta x} d\left(G_n(x) - \Gamma_\alpha(x)\right) \right| \\ &= \left| \int_0^\infty (1-\theta x) e^{-\theta x} \left[G_n(x) - \Gamma_\alpha(x)\right] dx \right| \\ &\leqslant L_1 \cdot \sup_x |\, G_n(x) - \Gamma_\alpha(x)|\,, \end{aligned} \tag{6}$$

где $L_1 = L_1(\theta) \equiv const > 0$, при любом $\theta \leqslant \Theta$. В свою очередь, очевидно, что

$$\left|\psi_n^2(\theta) - \psi_\alpha^2(\theta)\right| \leqslant 2\left|\psi_n(\theta) - \psi_\alpha(\theta)\right|. \tag{7}$$

Из соотношений (5)–(7), при наличии условия (3), мы приходим к (4).

Чтобы доказать достаточность условия (4), рассмотрим (1) как дифференциальное уравнение с начальным условием $\psi(0) = 1$. Тогда легко убедимся в том, что

$$\psi_n(\theta) - \psi_\alpha(\theta) = \psi_n(\theta)\psi_\alpha(\theta)\int_0^\theta \frac{\Delta\left(\psi_n(\tau)\right)}{\psi_n^2(\tau)}d\tau.$$

Откуда мы убедимся в справедливости неравенства

$$\sup_{\theta\leqslant\Theta}\left|\psi_n(\theta) - \psi_\alpha(\theta)\right| \leqslant L_2\cdot\Theta\cdot\sup_{\theta\leqslant\Theta}\left|\Delta\left(\psi_n(\theta)\right)\right|, \tag{8}$$

для любого $\Theta > 0$, здесь $L_2 = L_2(\theta) \equiv const > 0$, при любом $\theta \leqslant \Theta$.

Неравенство (8) доказывает достаточность условия (4) для сходимости (3). ■

Далее мы обсуждаем применение доказанной Теоремы 1.

Обозначим Z_n ($Z_0 = 1$) число частиц к моменту времени $n \in \mathbf{N}_0$ ($\mathbf{N}_0 = \{0\} \cup \{\mathbf{N}\}$) в ГВП, заданный производящей функцией (п.ф.)

$$F(x) = \sum_{k\in\mathbf{N}_0} p_k x^k, \quad p_k = \{Z_1 = k\}, \quad |x| \leqslant 1.$$

Будем рассматривать случай, когда среднее число потомков одной частицы за одно поколение равно 1, при котором ГВП называется критическим; см. [3]. Последнее равносильно тому, что $F'(1) = 1$; предположим, что $B := F''(1)$ конечен.

Пусть $F_n(x) = \mathbf{M}x^{Z_n}$. Введем функцию $R_n(x) := 1 - F_n(x)$, и заметим, что

$$g_n(x) := \sum_{j\in\mathbf{N}} \mathbf{P}\left\{Z_n = j\,|Z_n > 0\right\}x^j = 1 - \frac{R_n(x)}{R_n(0)}, \tag{9}$$

где величина $R_n(0) = \{Z_n > 0\}$ представляет собой вероятность продолжения ГВП.

Утверждение об асимптотическом представлении $R_n(x)$, по своей значимости, называется основной леммой теории критических ГВП; см. [3, стр. 19].

В нашем случае немаловажное значение представляет и асимптотическое разложение функции $R'_n(x)$. Мы обнаружим это разложение в окрестности точки $x = 1$. Имеет место следующий локально-дифференциальный аналог основной леммы (см. [4] в непрерывном случае).

Лемма 1. Если $B < \infty$, то при $x \to 1$ справедливо соотношение

$$R'_n(x) \sim -g_n^2(x), \quad n \to \infty, \tag{10}$$

где п.ф. $g_n(x)$ задана равенством (9).

Доказательство. В условиях теоремы имеем разложение

$$F(t) = t + \frac{B}{2}(t-1)^2\left(1 + o(t-1)^2\right), \quad t \to 1. \tag{11}$$

С помощью итерации для $F(x)$, с учетом того, что $|R_n(x)| \to 0$, $n \to \infty$ (см. [3, стр. 8]), из (11) получим

$$F_n(F(x)) - F_n(x) = \frac{B}{2}R_n^2(x)(1 + o(1)), \quad n \to \infty. \tag{12}$$

Используя теорему Лагранжа, из (12) находим

$$F'_n(c(x)) = \frac{B}{2(F(x) - x)}R_n^2(x)(1 + o(1)), \; n \to \infty, \tag{13}$$

где $c(x) = x + (F(x) - x)\theta$, $0 < \theta < 1$. Учитывая асимптотическое представлении 11 имеем $c(x) \sim x$, $x \to 1$. Опять-таки, равенство (11) утверждает $F(x) - x \sim \frac{B}{2}(1-x)^2$, $x \to 1$. Сказанные вместе с (13) позволяет написать, что при $x \to 1$

$$F'_n(x) \sim \left[\frac{1}{1-x}R_n(x)\right]^2, \quad n \to \infty.$$

С помощью основной леммы (см. [3, стр. 19]), последнее представление мы легко преобразуем к виду (10). ∎

Теперь используя Теоремы 1 в сочетании с Леммой, приведем новое доказательство следующей известной предельной теоремы Яглома.

Теорема 2. [3, стр. 20]. Пусть $B < \infty$. Тогда для любого $x > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{Z_n}{\mathbf{M}[Z_n | Z_n > 0]} \leqslant x | Z_n > 0\right\} = 1 - e^{-x}. \tag{14}$$

Доказательство. . Нам известно, что $[Z_n|Z_n > 0] = 1/R_n(0)$. (см. [3, стр. 19]). Тогда ПЛ рассматриваемого условного распределения

$$\varphi_n(\theta) := \left[e^{-\theta Z_n/[Z_n|Z_n>0]}|Z_n > 0\right]$$

мы можем записать в виде $\varphi_n(\theta) = g_n(\theta_n)$, где $\theta_n = \exp\{-\theta R_n(0)\}$.

Поскольку $R_n(0) \sim O(1/n)$, $n \to \infty$, то дифференцируя его, с учетом 10, получим

$$\varphi_n'(\theta) \sim -\varphi_n^2(\theta), \quad n \to \infty.$$

Последнее равносильно тому, что

$$\sup_{\theta \leqslant \Theta} |\Delta(\varphi_n(\theta))| \to 0, \quad n \to \infty,$$

при любом $\Theta > 0$. Согласно Теореме 1, этого достаточно для сходимости (14). ■

В конце мы продемонстрируем одно применение Леммы в теории $Q-$процессов. Пусть величина W_n обозначает состояние $Q-$процесса к моменту времени n, описываемый последовательностью п.ф. (см. напр., [3, стр.58])

$$\left\{W_n(x) := \mathbf{M}x^{W_n} = -x\frac{R_n'(x)}{\beta^n},\ n \in \mathbf{N}_0\right\},$$

где $\beta = F'(q)$ и число $q \in (0;1]$ есть вероятность вырождения ГВП.

Теорема 3. Пусть $F'(1) = 1$ и $B < \infty$. Тогда

$$n^2 W_n(x) = \mu(x)(1 + o(1)),\ \ n \to \infty, \tag{15}$$

где предельная п.ф. $\mu(x) = \sum_{k\in\mathbf{N}} \mu_k x^k$ и последовательность неотрицательных чисел $\{\mu_n,\ n \in \mathbf{N}\}$ имеет свойство

$$\lim_{n\to\infty} \frac{1}{n^2}[\mu_1 + \mu_2 + \cdots + \mu_n] = \frac{2}{B^2} \tag{16}$$

Доказательство. В рассматриваемом случае $q = 1$ и $\beta = 1$; см. [3, стр. 7]. Тогда

$$W_n(x) = -xR_n'(x). \tag{17}$$

Поскольку $R_n(0) \sim 2/Bn$ при $n \to \infty$, то с помощью 9 и основной леммы (см. [3, стр. 19]), соотношение 10 преобразуем к виду

$$R_n'(x) \sim -\frac{4}{B^2n^2}\frac{1}{(1-x)^2}, \quad n \to \infty, \tag{18}$$

при $x \to 1$. Рассмотрев вместе (17) и (18), мы получим (15), где

$$\mu(x) \sim -\frac{4}{B^2}\frac{1}{(1-x)^2}, \quad x \to 1. \tag{19}$$

Теперь нам остается вспомнить Тауберову теорему Харди-Литтлвуда, согласно которой каждое из соотношений (16) и (19) влечет другое. ■

СПИСОК ЛИТЕРАТУРЫ

1. *Тихомиров А. Н.* О скорости сходимости в центральной предельной теореме для слабо зависимых величин // Теория вероятностей и ее применения. 1980. Т. 25. №4. С. 800–818.
2. *Форманов Ш. К.* Метод Стейна-Тихомирова и неклассическая центральная предельная теорема // Математические заметки. 2002. Т. 71. №4. С. 604–610.
3. *Athreya K. B., Ney P. E.* Branching processes. New York: Springer, 1972.
4. *Imomov A. A.* A differential analog of the main lemma of the theory of Markov branching processes and its applications // Ukrainian Mathematical Journal. 2005. Т. 57. №2. С. 307–315.

Имомов Аъзам Абдурахимович — д.ф.-м.н., зав.кафедрой Алгебры и геометрии. E-mail: *imomov_azam@mail.ru*

Тухтаев Эркин Эгамбердиевич — преподаватель кафедры Алгебры и геометрии. E-mail: *tukhtaev_erkin@mail.ru*

Абдирахмонов Абдимумин — к.п.н., ст. преподаватель кафедры Алгебры и геометрии.

О ЧИСЛЕ СОБСТВЕННЫХ ЗНАЧЕНИЙ ДВУХЧАСТИЧНОГО ГАМИЛЬТОНИАНА НА РЕШЕТКЕ

М. Э. Муминов, А. М. Хуррамов

Самаркандский государственный университет, г. Самарканд, Узбекистан

Рассматривается система двух произвольных квантовых частиц на трехмерной решетке с некоторыми дисперсионными функциями (описывающими перенос частицы с узла на соседний узел), частицы взаимодействуют с помощью потенциала притяжения только на ближайших соседних узлах. Изучена зависимость числа собственных значений семейства операторов $h(k)$ от энергии взаимодействия частиц и полного квазиимпульса $k \in T^3$ (T^3- 3-мерный тор). Введен вспомогательной оператор и полностью опесан его спектр. С помощью спектральных свойств этого оператора найдены условия сущесвования собственных значений рассматриваемого оператора.

Ключевые слова: *Двухчастичный гамильтониан на решетке, собственное значение, положительный оператор.*

Введение

В непрерывном случае изучение спектральных свойств полного гамильтониана системы двух частиц сводится к изучению двухчастичного оператора Шрёдингера с помощью выделения энергии движения центра масс, при этом одночастичные связанные состояния суть собственные векторы оператора энергии с отделенным полным импульсом [1] (такой оператор фактически не зависит от значений полного импульса). На решетке выделению центра масс системы отвечает реализация гамильтониана как «расслоенного» оператора, т. е. прямого интеграла семейства операторов $h(k)$ энергии двух частиц, зависящих от значений полного квазиимпульса $k \in T^d$ (T^d – d-мерный тор) [2, 3]. Является ли число дискретных собственных значений многочастичной системы конечным или бесконечным, зависит от числа виртуальных уровней подсистем.

1. Постановка задачи и формулировка основных результатов

Пусть $T = (-\pi, \pi]$, $L_2(T^3)$ – гильбертово пространство всех квадратично-интегрируемых функций, определенных на T^3.

Рассмотрим $h(k)$ – самосопряженный оператор, действующий в $L_2(T^3)$ по формуле

$$h(k) = h_0(k) - \mathbf{v},\ k = (k_1, k_2, k_3) \in T^3,$$

$h_0(k)$ – оператор умножения на функцию

$$\mathcal{E}_k(p) = \frac{1}{m_1}\varepsilon(p) + \frac{1}{m_2}\varepsilon(p-k), \quad \varepsilon(p) = \sum_{i=1}^{3}(1 - \cos p_i),$$

$\mathbf{v}$ – интегральный оператор с ядром $v(p-s) = \mu_0 + \sum\limits_{\alpha=1}^{3} \mu_\alpha \cos(p_\alpha - s_\alpha)$.

Отметим, что из теоремы Вейля о существенном спектре [4] следует, что существенный спектр $\sigma_{ess}(h(k))$ оператора $h(k)$ не меняется при компактном возмущении $\mathbf{v}$ и совпадает со спектром невозмущенного оператора $h_0(k)$. При этом $\sigma_{ess}(h(k))$ сопадает с областью значений функции $\mathcal{E}_k(\cdot)$:

$$\sigma_{ess}(h(k)) = \sigma(h_0(k)) = [m(k), M(k)],$$

$$m(k) = \min_p \mathcal{E}_k(p), \quad M(k) = \max_p \mathcal{E}_k(p).$$

Поскольку $\mathbf{v} \geq 0$, мы имеем

$$\sup(h(k)f, f) \leqslant \sup(h_0(k)f, f) = M(k)(f, f), \qquad f \in L_2(T^3).$$

Поэтому оператор $h(k)$ не имеет собственного значения, лежащего правее существенного спектра, т.е.

$$\sigma(h(k)) \cap (M(k), \infty) = \varnothing.$$

Положим

$$\mu_i^{\pm}(k; z) = \frac{c_i(k;z) + s_i(k;z) \pm \sqrt{(c_i(k;z) - s_i(k;z))^2 + 4\xi_i^2(k;z)}}{2[c_i(k;z)s_i(k;z) - \xi_i^2(k;z)]},$$

где

$$c_i(k;z) = \int\limits_{T^3} \frac{\cos^2 s_i ds}{\tilde{\mathcal{E}}_k(s) - z},$$

$$s_i(k;z) = \int\limits_{T^3} \frac{\sin^2 s_i ds}{\tilde{\mathcal{E}}_k(s) - z},$$

$$\xi_i(k;z) = \int\limits_{T^3} \frac{\sin s_i \cos s_i ds}{\tilde{\mathcal{E}}_k(s) - z}, \; z \leqslant \tilde{m}(k),$$

здесь

$$\tilde{\mathcal{E}}_k(p) = \frac{1}{m_1}\tilde{\varepsilon}(p) + \frac{1}{m_2}\tilde{\varepsilon}(p-k), \quad \tilde{\varepsilon}(p) = \sum_{i=1}^{3}(1 + |\cos p_i|),$$

$$\tilde{m}(k) = \min_p \tilde{\mathcal{E}}_k(p), \quad \tilde{M}(k) = \max_p \tilde{\mathcal{E}}_k(p).$$

Отметим, что $c_i(k;z)s_i(k;z) - \xi_i^2(k;z) > 0$ (см. [5]). Функции $b(k;z), c_i(k;z), s_i(k;z)$ и $\xi_i^2(k;z)$ непрерывный в $(-\infty), \tilde{m}(k)]$. Заметим, что $m(k) < \tilde{m}(k)$. Поэтому

$$b(k;m(k)), \quad c_i(k;m(k)), \quad s_i(k;m(k)), \quad \xi_i^2(k;m(k))$$

существуют, где $b(k;z) = \int\limits_{T^3} \frac{ds}{\tilde{\mathcal{E}}_k(s)-z}$.

Положим

$$\mu^0(k) = \frac{1}{b(k;m(k))} \quad , \mu_i^c(k) = \frac{1}{c_i(k;m(k))}, \quad \mu_i^s(k) = \frac{1}{s_i(k;m(k))}.$$

$$\mu_i^{\pm}(k) = \mu_i^{\pm}(k;m(k)), \quad i = 1,2,3.$$

При этом

$$\mu_i^-(k) \leqslant \mu_i^+(k) \quad \text{при} \quad k \in T^3, \quad i = 1,2,3.$$

Определим функции

$$\alpha(\mu;k) = \begin{cases} 0 & \text{при} \quad \mu_0 \in (0;\mu^0(k)], \\ 1 & \text{при} \quad \mu_0 \in (\mu^0(k);\infty), \end{cases}$$

$$\beta_i(\mu;k) = \begin{cases} 0 & \text{при} \quad \mu_i \in (0, \mu_i^-(k)], \\ 1 & \text{при} \quad \mu_i \in (\mu_i^-(k), \mu_i^+(k)], \\ 2 & \text{при} \quad \mu_i \in (\mu_i^+(k), \infty) \end{cases}$$

для всех $i = 1,2,3$.

Теорема 1. Пусть $\mu = (\mu_0, \mu_1, \mu_2, \mu_3) \in R_+^4$. Тогда оператор $h(k)$ имеет не менее чем с учетом кратности ровно

$$\alpha(\mu;k) + \sum_{i=1}^{3} \beta_i(\mu;k)$$

собственных значений, лежащих левее существенного спектра.

В частности $\mu_0 > \mu^0(k)$ и $\mu_i > \mu_i^+(k)$, $i = 1, 2, 3$. Тогда оператор $h(k)$ имеет с учетом кратности ровно 7 собственных значений, лежащих левее существенного спектра.

Заметим, что в случае $k = 0$ теорема 1 формулеруется более компактно.

Теорема 2. Для любого $\mu \in R_+^4$, $\mu = (\mu_0, \mu_1, \mu_2, \mu_3)$ оператор $h(\mathbf{0})$ имеет не менее чем с учетом кратности ровно

$$\alpha(\mu, \mathbf{0}) + \sum_{i=1}^{3} \theta(i)$$

собственных значений, лежащих левее существенного спектра, где

$$\alpha(\mu, \mathbf{0}) = \begin{cases} 0 & \text{при} \quad \mu_0 \in (0, \mu^0(\mathbf{0})], \\ 1 & \text{при} \quad \mu_0 \in (\mu^0(\mathbf{0}), \infty), \end{cases}$$

$$\theta(i) = \begin{cases} 0 & \text{при} \quad \mu_i \in (0, \mu_i^c(\mathbf{0})], \\ 1 & \text{при} \quad \mu_i \in (\mu_i^c(\mathbf{0}), \mu_i^s(\mathbf{0})], \\ 2 & \text{при} \quad \mu_i \in (\mu_i^s(\mathbf{0}), \infty). \end{cases}$$

СПИСОК ЛИТЕРАТУРЫ

1. *Фаддеев Л. Д.* Математические вопросы квантовой теории рассеяния для системы трех частиц // Труды Математического института имени В.А. Стеклова. 1963. Т. 69. № 0. С. 3–122.
2. *Mattis D. C.* The few-body problem on lattice // Rew. mod. Phys. 1986. T. 58. C. 361–379.
3. *Albeverio S., Lakaev S. N., Makarov K. A., Muminov Z. I.* The threshold effects for the two-particle Hamiltonians // Commun. Math. Phys. 2006. T. 262. C. 91–115.
4. *Рид М., Саймон Б.* Методы современной математической физики. М.: Мир, 1982.
5. *Муминов М. Э., Хуррамов А. М.* Спектральные свойства двухчастичного гамильтониана на решетке // Теор. Мат. Физика. 2013. Т. 177. № 3. С. 480–493.

Муминов М. Э. — доцент Самаркандского государственного университета Республики Узбекистан. E-mail: *mmuminov@mail.ru*

Хуррамов А. М. — докторант Самаркандского государственного университета Республики Узбекистан. E-mail: *xurramov@mail.ru*

О ЗАДАЧАХ В АНАЛИЗЕ ДОРОЖНЫХ СЕТЕЙ С ТОЧКАМИ ИНТЕРЕСА

А. А. Солдатенко[1], Д. В. Семенова[1,2]

[1] *Сибирский федеральный университет, г. Красноярск, Россия*
[2] *Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого, г. Красноярск, Россия*

В работе представлены новые результаты анализа дорожных сетей. Исследована загруженность дорожных участков сети при проезде в точки интереса — корпуса высших учебных заведений. Анализ выполнен на основе многократных решений задачи о кратчайшем пути в нестационарной метрической сети с условием FIFO и в ресурсоограниченной сети с одним ресурсом. Для получения связных областей дорожной сети предложен новый метод применения максимальных индуцированных биклик. Анализ дорожных сетей производился с применением алгоритмов, ранее разработанных авторами.

Ключевые слова: *Анализ сетей, дорожные сети, маршрутизация, максимальные индуцированные биклики.*

Введение

В настоящее время, развитие городской инфраструктуры и расширение сети Интернет неизбежно влечет за собой реструктуризацию существующих моделей, описывающих дорожные и телекоммуникационные сети. Это затрагивает методы анализа и проектирования сетей, в которые входят решение таких задач как поиск кратчайших путей в сетях с ограничениями, задача о покрытии, задачи перечисления специфичных конфигураций [1, 2, 3].

В данной работе рассматриваются задачи о кратчайшем пути в нестационарной метрической сети и в ресурсоограниченной сети с одним ресурсом, а также поиска всех максимальных индуцированных биклик. Приводятся результаты вычислительных экспериментов для дорожных сетей реальных городов, с использованием ранее разработанных авторами алгоритмов [4, 5, 6, 7].

Работа поддержана Красноярским математическим центром, финансируемым Минобрнауки РФ в рамках мероприятий по созданию и развитию региональных НОМЦ (Соглашение 075-02-2020-1534/1).

1. Теоретико-графовые задачи для анализа сетей

В работе рассматриваются две задачи о кратчайшем пути и задача перечислительного типа о поиске всех максимальных индуцированных биклик. Для формулировок задач будем использовать понятия и определения из работ [4, 8].

Задача о кратчайшем пути в нестационарной сети (TDSP). Заданы нестационарная сеть $G = (V, E)$ и (s, d, t_s)-запрос. Требуется найти значение $dist(s, d, t_s)$ для кратчайшего (s, d, t_s)-пути и последовательность вершин, образующих его. Здесь величина $dist(s, d, t_s)$ определяет наименьшее время движения из вершины s в вершину d, если выход из вершины s совершен в момент времени t_s.

Задача о кратчайшем пути в ресурсоограниченной сети (RCSP). Задан ориентированный граф $G = (V, E)$, на дугах которого определены положительные вещественнозначные функции w_e, $r(e)$, величина R. Требуется найти оптимальный (s, d)-маршрут.

Задача поиска всех максимальных индуцированных биклик (MIBGP). Задан граф $G = (V, E)$ без кратных ребер. Требуется найти множество всех максимальных индуцированных биклик.

Для решения задач использовались разработанные авторами алгоритмы RevTree, HFindMIB и модифицированный алгоритм ALT [5, 6, 7]. Класс сложности задач и вычислительная сложность алгоритмов представлена в табл. 1.

Таблица 1

Рассматриваемые задачи и алгоритмы их решения

Задача		Алгоритм	
Название	Класс сложности	Название	Сложность
TDSP	P	ALT	$\mathcal{O}(n^2)$
RCSP	NP-трудная	RevTree	$\mathcal{O}(n^2)$
MIBGP	Не легче NP-трудной	HFindMIB	$\mathcal{O}\big(2^{2\Delta} \cdot \Delta \cdot (\lvert\mathcal{MBC}\rvert + \Delta^3 \cdot \log(2^{2\Delta})) + n^2\big)$

2. Вычислительные эксперименты

В работе представлены новые вычислительные эксперименты для анализа дорожных сетей городов: Томска (TOF), Красноярска (KJA), Новосибирска (NSK), Баку (GYD). Модифицированный алгоритм ALT и алгоритм RevTree применялся к сгенерированным запросам на поиск кратчайших путей до точек интереса. Алгоритм HFindMIB применялся для неориентированных версий соответствующих дорожных сетей.

Вычислительные эксперименты были проведены на ЭВМ с процессором AMD Ryzen 5 3600 6-Core Processor 3,60 ГГц и ОЗУ объемом 16 Гбайт. Результаты экспериментов представлены в табл. 2.

Таблица 2

Результаты работы алгоритмов ALT, RevTree, HFindMIB

Дорожная сеть	n	m	ALT	RevTree	HFindMIB		
			Время, с.	Время, с.	Δ	$\mathcal{MBC}$	Время, с.
TOF	3840	9639	4,566	1,890	5	4084	26,527
KJA	4362	6005	5,975	6,038	5	4823	36,898
NSK	7357	19106	14,141	7,957	6	7927	103,115
GYD	12458	28936	50,722	12,575	6	11844	252,244

В качестве точек интереса в сети KJA были выбраны корпуса Сибирского федерального университета. В остальных сетях в качестве точек интереса рассматривались высшие учебные заведения. На основе кратчайших путей, полученных алгоритмами ALT и RevTree, построены тепловые карты дорог, при движении к точками интереса. Дорога отмечена красным цветом, если она часто является частью кратчайшего пути, синим цветом в противном случае. Точки интереса отмечены фиолетовым цветом. Карты для сети KJA представлены на рис. 1 и 2, а для сети TOF на рис. 3 и 4.

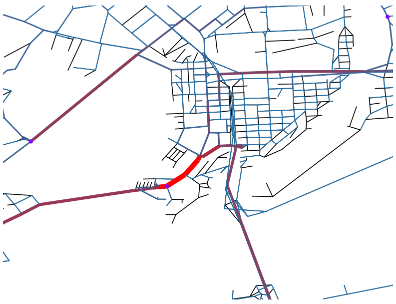

Рис. 1. Фрагмент тепловой карты для сети KJA по результатам решения задачи TDSP с точками интереса

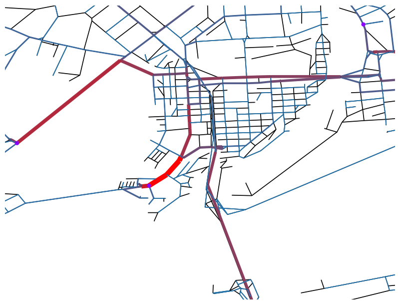

Рис. 2. Фрагмент тепловой карты для сети KJA по результатам решения задачи RCSP с точками интереса

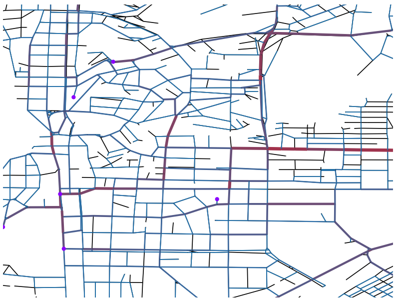

Рис. 3. Фрагмент тепловой карты для сети TOF по результатам решения задачи TDSP с точками интереса

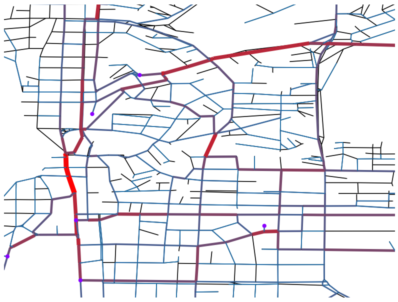

Рис. 4. Фрагмент тепловой карты для сети TOF по результатам решения задачи RCSP с точками интереса

Согласно представленным результатам, красные участки дороги можно интерпретировать как наиболее загруженные.

Результатом работы алгоритма HFindMIB является множество всех максимальных индуцированных биклик дорожной сети. Ранее было отмечено, что максимальные индуцированные биклики различного размера могут определять разные реальные структуры [4].

Вершины, которые имели маршрут через максимальные индуцированные биклики одного размера, имеющие хотя бы одну общую вершину, окрашивались в один цвет. Результаты для биклик размера $(2, 2)$ и $(4, 1)$ в сетях KJA и TOF представлены на рис. 5–8.

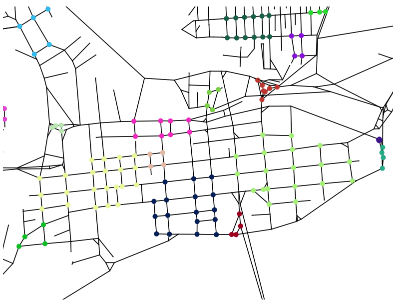

Рис. 5. Области сети KJA, полученные из смежных максимальных индуцированных биклик размера $(2, 2)$

Рис. 6. Области сети KJA, полученные из смежных максимальных индуцированных биклик размера $(4, 1)$

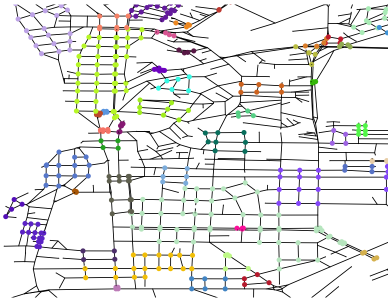

Рис. 7. Области сети TOF, полученные из смежных максимальных индуцированных биклик размера $(2, 2)$

Рис. 8. Области сети TOF, полученные из смежных максимальных индуцированных биклик размера $(4, 1)$

Данный результат может иметь существенное приложение при кластеризации и атомизации различных сетей, однако требует дополнительных исследований.

Заключение

В работе представлены новые результаты для анализа дорожных сетей с использованием решений задач о кратчайшем пути в нестационарной сети с условием FIFO и в ресурсоограниченной сети с одним ресурсом, а также задачей поиска всех максимальных индуцированных биклик в сети.

Алгоритм RevTree и модифицированный алгоритм ALT используют накопленные данные о загруженности дорожной сети в разное время суток, чтобы строить оптимальный маршрут с учетом этих данных. Получение кратчайших путей для серии запросов позволяет находить наиболее востребованные дороги в дорожной сети. Использование модельной дорожной сети с добавлением новых дорог или перекрытием существующих может позволить анализировать последствия таких решений. Основываясь на информации о наиболее востребованных ребрах в сети можно выполнять реорганизацию дорожной сети и моделировать новые объездные дороги.

Алгоритм HFindMIB для нахождения всех максимальных индуцированных биклик позволяет выделять в дорожной сети биклики, которые можно интерпретировать как отдельные дорожные структуры (перекрестки, протяженные дороги, городские кварталы и пр.), тем самым это может позволить оценивать скопление тех или иных структур в областях сети. Метод объединения смежных биклик в единые множества является интересным для дальнейшего исследования с точки зрения методов кластеризации и декомпозиции сетей различных предметных областей.

СПИСОК ЛИТЕРАТУРЫ

1. *Faure N., Chretienne P., Gourdin E., Sourd F.* Biclique completion problems for multicast network design // Discrete Optimization. 2007. V. 4, Issues 3–4. P. 360–377.
2. *Goldberg A., Kaplan H., Werneck R.* Better landmarks within reach // International Workshop on Experimental Algorithms (WEA). Springer, 2007. 4525 of LNCS. P. 38–51.
3. *Nejad M. M., Mashayekhy L., Chinnam R. B., Phillips A.* Hierarchical time-dependent shortest path algorithms for vehicle routing under ITS // J. IIE Transactions. 2016. V. 48. №. 2. P. 158–169.
4. *Soldatenko A. A., Semenova D. V.* On Two Problems of Combinatorial Optimization in the Analysis of Road Networks // International Conference on Problems of Logistics, Management and Operation in the East-West Transport Corridor (PLMO), October 27-29, 2021. P. 101–107. (in press)
5. *Быкова В. В., Солдатенко А. А.* Оптимальная маршрутизация по ориентирам в нестационарных сетях // Прикладная дискретная математика. 2017. № 37. С. 114–123.
6. *Soldatenko A. A.* On accuracy of approximation for the resource constrained shortest path problem // Journal of Siberian Federal University. Mathematics & Physics. 2019. V. 12. № 5. P. 621–627.
7. *Soldatenko A. A., Semenova D. V.* On Problem of Finding all Maximal Induced Bicliques of Hypergraph // Journal of Siberian Federal University. Mathematics & Physics. 2021. V. 14. № 5. P. 638–648.
8. *Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И.* Лекции по теории графов. М.: Книжный дом «Либроком», 2012. 390 с.

Солдатенко Александр Александрович — ассистент, Институт математики и фундаментальной информатики, Сибирский федеральный университет. E-mail: *ASoldatenko@sfu-kras.ru*

Семенова Дарья Владиславовна — кандидат физико-математических наук, доцент, доцент, Институт математики и фундаментальной информатики, Сибирский федеральный университет. E-mail: *DVSemenova@sfu-kras.ru*

АНАЛИЗ РАСПРЕДЕЛЕНИЯ ВРЕМЕНИ РЕАЛИЗАЦИИ РЕСУРСА С НУЛЕВЫМ КОНЕЧНЫМ ЗАПАСОМ И ЭКСПОНЕНЦИАЛЬНО РАСПРЕДЕЛЁННЫМИ ЗАПРОСАМИ

Н. В. Степанова[1], А. В. Китаева[2], М. П. Фархадов[1]

[1] *ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, Россия*
[2] *Томский государственный университет, г. Томск, Россия*

В работе проводится анализ времени реализации скоропортящегося продукта, имеющего заданный срок хранения, при условии поступления заказов фиксированного объема. Поступление заявок на ресурс происходит в соответствии с неоднородным пуассоновским потоком, где интенсивность зависит от текущей цены ресурса. Размер партии в поступившей заявке предполагается экспоненциально распределённым. Исследуется модель контроля розничных цен, позволяющая нам почти наверняка продать партию без остатков в течение срока годности продукта. Получены преобразования Лапласа плотности распределения условного времени до окончания сеанса и условного среднего значения этого времени. В работе представлены также результаты имитационного моделирования продолжительности времени реализации ресурса. Эти результаты сравниваются затем со значениями, полученными ранее с помощью аппроксимационной модели.

Ключевые слова: *Динамическое ценообразование, нулевой конечный запас, продолжительность продажи, продукт с фиксированным сроком службы, неоднородный пуассоновский спрос, преобразование Лапласа.*

Введение

Сокращение образования отходов – это один из основных способов уменьшения нашего воздействие на окружающую среду. Варианты сокращения отходов имеют высокий приоритет, поскольку утилизация отходов может быть дорогостоящей и может вызвать экологические проблемы; см. [2], [5] и [9]. Экологические и социальные последствия управления скоропортящимися продуктами недавно обсуждались в [1] и [11], где также подчеркивалось влияние динамического ценообразования на объем выручки и испорченной продукции. Модели динамического ценообразования особенно подходят для поставщиков свежих товаров в электронной коммерции, где издержки связанные с изменением цен можно практически не учитывать. Такие модели могут обеспечить значительное увеличение розничной прибыли; см., например, работу [3]. Недавний краткий обзор литературы по моделям динамического ценообразования можно найти в [10]. В работе [6] мы рассмотрели динамическую модель управления розничными ценами, которая вытекает из требования равенства средней скорости, с которой количество товара $Q(t)$ должно быть продано в течение временного интервала $[t, T]$, чтобы быть полностью проданным к концу периода жизни T, и мгновенной скорости, с которой он продается в момент времени t

$$a_1\lambda(c(t)) = \frac{Q(t)}{T-t}. \tag{1}$$

Здесь $\lambda(c)$ - зависящая от цены интенсивность пуассоновского потока покупателей, - розничная цена за единицу товара; значения объема заказов - независимые одинаково распределенные (н.о.р.) непрерывные случайные величины с конечными первым и вторым моментами, равными соответственно a_1 и a_2. Мы предполагаем, что спрос очень чувствителен к цене, и продукт обязательно будет продан, если цена будет достаточно низкой. Это может быть характерно, например, для товаров первой необходимости. Из-за сложности такого рода задач широко используются численные или приближенные аналитические методы. В работе [6] мы проанализировали диффузионную аппроксимацию этой модели, предполагая, что процесс уровня запасов $Q(\cdot)$ может быть описан следующим стохастическим дифференциальным уравнением:

$$dQ(t) = -a_1\lambda(c(t))dt - \sqrt{a_2\lambda(c(t))}dw(t).$$

Здесь $w(\cdot)$ - процесс Винера и $Q(0) = Q_0$. Эта аппроксимация хорошо работает при большом количестве закупок. Предложенная модель управления ценами (1) решает проблему остатков, то есть конечные запасы почти наверняка равны нулю. В то же время вероятность отсутствия запасов в течение срока годности продукта достаточно мала для вышеприведенной диффузионной аппроксимации.

В данной работе мы рассматриваем точное распределение продолжительности времени, необходимого для продажи продукта, и его среднего значения в случае экспоненциального распределения размера партии, то есть предполагается, что значения покупок являются н.о.р. случайными величинами, имеющими экспоненциальное распределение со средним a. В этом случае $a_1 = a$ и $a_2 = 2a^2$. Мы выводим соответствующие дифференциальные уравнения методом и используем преобразования Лапласа для решения этих уравнений. Нахождение решения дифференциальных уравнений является одним из традиционных применений преобразования Лапласа. В работе [4], где метод преобразования используется в задачах предпочтения риска, упоминается множество областей его применения. В [7] получены точные распределения спроса и времени сеанса продажи для неоднородного пуассоновского процесса спроса с экспоненциальным распределением размера партии в рамках задачи продавца газет для постоянной интенсивности потока покупателей. Распределения удалось найти в явной форме благодаря свойству экспоненциального распределения отсутствия последействия.

1. Среднее время реализации ресурса

Обозначим через $m(Q,t)$ ожидаемое время до окончания процесса продажи, если в момент t количество товара составляет $Q(t) = Q$. Для вывода уравнения воспользуемся Δt методом, при котором рассматривается приращение интересующей нас переменной за интервал времени $\Delta t << 1$, а затем $\Delta t \to 0$. В интервале времени $[t, t + \Delta t]$ возможны следующие варианты:

а) В течение времени Δt не было совершено ни одной покупки. Вероятность этого события равна $1 - \lambda(Q,t)\Delta t + o(\Delta t)$ при $\Delta t \to 0$, а оставшееся среднее время до конца сессии равно $m(Q, t + \Delta t)$.

б) Есть покупка за время Δt. Вероятность этого события равна $\lambda(Q,t)\Delta t + o(\Delta t)$ при $\Delta t \to 0$. Если стоимость покупки $\xi < Q$, то оставшееся время равно $m(Q - \xi, t + \Delta t)$; в противном случае сессия продаж завершается.

в) Вероятность всех остальных сценариев равна $o(\Delta t)$ при $\Delta t \to 0$.

Таким образом, можно получить следующее уравнение:

$$\frac{\partial m(Q,t)}{\partial t} - \frac{Q}{a(T-t)} m(Q,t) + \frac{Q}{a(T-t)} \int_0^Q m(Q-x, t+\Delta t) \frac{1}{a} e^{-\frac{x}{a}} dx = -1, \tag{2}$$

при условии $m(0,t) = 0$ и $m(Q,T) = 0$. Обозначим теперь через $\tilde{m}(q,t) = \int_0^\infty m(Q,t) e^{-qQ} dQ$ преобразование Лапласа (ПЛ) среднего времени.

Утверждение 1. ПЛ $\tilde{m}(q,t)$ представляется в виде

$$\tilde{m}(q,t) = (T-t)(aq+1)e^{aq} \int_q^\infty \frac{e^{-ax}}{x^2} dx. \tag{3}$$

Доказательство. Утверждение следует путём умножения (2) на e^{-qQ} и последующего вычисления интеграла по Q на интервале от 0 до ∞ с учётом имеющихся граничных условий. ■

Следствие 1. ПЛ средней длительности сессии имеет вид

$$\tilde{m}(q,0) = T(aq+1)e^{aq}\int_q^\infty \frac{e^{-ax}}{x^2}dx = aT(aq+1)e^{aq}\Gamma(-1,aq),$$

где $\Gamma(\cdot,\cdot)$ верхняя неполная гамма-функция.

2. Распределение времени реализации ресурса

Рассмотрим распределение продолжительности времени, необходимого для реализации заказа $Q(t)$, для спроса, моделируемого сложным пуассоновским процессом с экспоненциальным распределением размера партии со средним a и управляемой ценовой интенсивностью (1). Обозначим через $G(s,Q,t) = \mathbb{E}[e^{-s\tau}|Q,t]$ преобразование Лапласа условной плотности $p_\tau(\cdot|Q(t) = Q)$. Теперь получим уравнение для $G(s,Q,t)$. В интервале $[t, t+\Delta t]$ возможны следующие случаи:

a) Нет покупок на интервале $[t, t+\Delta t]$. В этом случае оставшееся время равно $\tau(Q, t+\Delta t)$.
b) Есть покупка $\xi < Q$ на интервале $[t, t+\Delta t]$, тогда оставшееся время равно $\tau(Q-\xi, t+\Delta t)$, где ξ - экспоненциально распределенная случайная величина со средним a.
c) Есть покупка $\xi \geqslant Q$ на интервале $[t, t+\Delta t]$, тогда оставшееся время равно 0.

Вероятности всех остальных вариантов равны $o(\Delta t)$ при $\Delta t \to 0$. Таким образом, с учетом (1), справедливо следующее соотношение:

$$\begin{aligned} &aT(1-\frac{t}{T})\frac{\partial G(s,Q,t)}{\partial t} - (saT(1-\frac{t}{T}) + Q)G(s,Q,t) \\ &+ Qe^{-\frac{Q}{a}} + Q\int_0^Q G(s,Q-x,t)\frac{1}{a}e^{-\frac{x}{a}}dx = 0 \end{aligned} \tag{4}$$

Рассмотрим теперь преобразование Лапласа по аргументу Q, т.е. $\tilde{G}(s,q,t) = \int_0^\infty e^{-qQ}dQ$.

Утверждение 2. Функция $\bar{G}(s,q,t) = e^{s(T-t)}\tilde{G}(s,q,t)$ удовлетворяет неоднородному дифференциальному уравнению частной производной первого порядка,

$$aT(1-\frac{t}{T})\frac{\partial\bar{G}(s,q,t)}{\partial t}+\frac{aq}{aq+1}\frac{\partial\bar{G}(s,q,t)}{\partial q} \tag{5}$$
$$+\frac{a}{(aq+1)^2}\bar{G}(s,q,t)=-\frac{a^2}{(aq+1)^2}e^{s(T-t)}$$

Следствие 2. Частное решение (5) имеет вид

$$\bar{G}(s,q,t)=\sum_{r=0}^{\infty}\varphi_r(s,q)(T-t)^r, \quad \text{где}$$
$$\varphi_r(s,q)=\frac{a(aq+1)}{r!}e^{raq}s^r q^{r-1}\int_q^{\infty}\frac{e^{-rax}x^{-r}}{(ax+1)^2}dx.$$

3. Имитационное моделирование

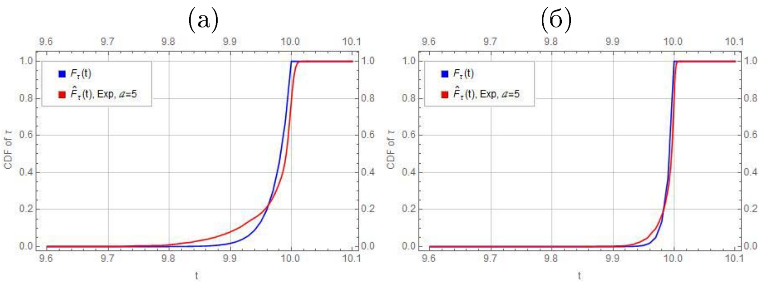

Рис. 1. Функция $F_\tau(t)$ для экспоненциально распределённого размера заявки, $a=5$, $Q_0=2000$, $\hat{\tau}=9.973$ (а) и $Q_0=5000$, $\hat{\tau}=9.999$

Здесь рассматривается сравнительный анализ диффузионной аппроксимации и имитационного моделирования. Для кумулятивной функции распределения (CDF) $F_\tau(\cdot)$ длительности времени, необходимого для продажи фиксированной партии товара Q_0, в случае большого числа заявок имеем

$$F_\tau(t)=\mathbb{P}[\tau\leqslant t]=e^{-\beta\frac{T-t}{t}Q_0}, \quad \text{где } \beta=\frac{2a_1}{a_2}=\frac{1}{a}. \tag{6}$$

На рисунках 1 и 2 сравниваются теоретические (приближенные) CDF и сглаженные эмпирические CDF для экспоненциального и равномерного распределений закупок и двух различных значений коэффициента βQ_0. Также вычислены выборочные средние $\hat{\tau}$ при всех наборах параметров.

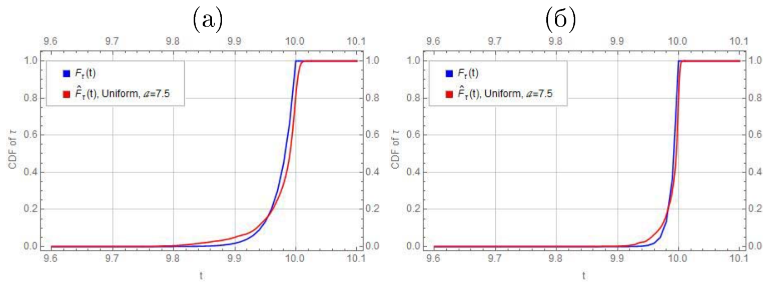

Рис. 2. Функция $F_\tau(t)$ для равномерно распределённого размера заявки на интервале $[0, 15]$, $a_1 = a = 7.5$, $Q_0 = 2000, \hat{\tau} = 9.977$ (а) и $Q_0 = 5000, \hat{\tau} = 9.989$ (б)

На рисунках $T = 10$. В этом случае большинство реализаций процесса будут заканчиваться опустошением склада до момента порчи товара. Количество итераций для моделирования τ равно 1000. На Рис. 1(б) и Рис. 2(б) значение βQ_0 в 2,5 раза больше, чем на Рис. 1(а) и Рис. 2(а), поэтому кривые на рисунках, обозначенных буквой (б), расположены ближе друг к другу.

Для генерации неоднородного пуассоновского процесса применяется инверсионный подход, см., например, [8].

Заключение

Получено преобразование Лапласа условного среднего времени пересечения нулевого уровня процессом уровня запасов для предложенной модели управления через зависящую от цены интенсивность пуассоновского потока покупателей (1) с экспоненциально распределенными закупками, а также преобразование Лапласа условной плотности вероятности этого времени. Моделирование подтверждает аналитические результаты для диффузионной аппроксимации процесса уровня запасов при большом количестве закупок. Модель управления ценами (1) обеспечивает высокий уровень обслуживания в течение срока годности T.

СПИСОК ЛИТЕРАТУРЫ

1. *Andenso-Diaz B., Lozano S., Palacio A.* Effects of dynamic pricing of perishable products on revenue and waste // Applied Mathematical Modelling. 2017. № 45. С. 148–164.
2. *Curran T., Williams I. D.* A zero waste vision for industrial networks in Europe // Journal of Hazardous Materials. 2012. № 207-208. С. 3–7.

3. *Gallego G., van Ryzin G.* Optimal dynamic pricing of inventories with stochastic demand over finite horizons // Management science. 1994. T. 40. № 8. C. 999–1020.
4. *Grubbström R. W.* Risk preference evaluation – a fourth dimension of the application of the Laplace transform // International Journal of Production Research. 2018. T. 56. № 1-2. C. 344–373.
5. *Hannon J., Zaman S. U.* Exploring the Phenimenon of Zero Waste and Future Cities // Urban Science. 2018. T. 2. № 3. C. 90.
6. *Kitaeva A. V., Stepanova N. V., Zhukovskaya A. O.* Zero ending Inventory Dynamic Pricing Model under Sto-chastic Demand, Fixed Lifetime Product, and Fixed Order Quantity // 9th IFAC Conference on Manufacturing Modelling, Management and Control (MIM 2019), Germany, Berlin, 2019. C. 2482–2487.
7. *Kitaeva A. V., Subbotina V. I., Zhukovskaya A. O.* Compound Poisson Demand Inventory Models with Exponential Batch Size's Distribution // Communication in Computer and Information Science. Springer, 2015. T. 564. C. 240–249.
8. *Pasupathy R.* Generating Nonhomogeneous Poisson Process. Wiley Encyclopedia of Operations Research and Management Science, John Wiley & Sons, 2011.
9. *Singh S., Ramakrishna S., Gupta A. K.* Towards zero waste manufacturing: A Multidisciplinary Review // Journal of Cleaner Production. 2017. T. 168. № 1. C. 1230–1243.
10. *Stasinski K.* A Literature Review on Dynamic Pricing-State of Current research and New Directions // Communication in computer and Information Science, Springer. 2020. T. 1287. C. 465–477.
11. *Tekin P., Erol R.* A New Dynamic Pricing Model for the Effective Sustainability of Perishable Product Life Cycle // Sustainability. 2017. T. 9. C. 1330.

Степанова Наталья Викторовна — к.т.н., с.н.с. E-mail: *natalia0410@rambler.ru*

Китаева Анна Владимировна — д.ф.-м.н. E-mail: *kit1157@yandex.ru*

Фархадов Маис Пашаевич — д.т.н., зав.лаб. E-mail: *mais.farkhadov@gmail.com*

О ДИНАМИКЕ ПОПУЛЯЦИЙ В РЕЧНОЙ ЭКОЛОГИИ

Ж. О. Тахиров, М. И. Боборахимова

Институт математики АН РУз, г. Ташкент, Узбекистан

Пространственно-временная динамика популяции представляет собой один из самых интересных аспектов и проблем для экологического моделирования. В работе исследуется математическая модель, основанный на одномерных уравнениях реакции-диффузии-адвекции, для роста популяции в гетерогенной среде обитания. Мы изучаем часто противоположные роли адвекции и диффузии в устойчивости популяции. Когда это возможно, мы демонстрируем основные математические методы и даем критические условия, обеспечивающие выживание популяции в простых системах и в более сложных ресурсах.

Ключевые слова: *динамика популяции, адвекция и диффузия, устойчивое состояние.*

Введение

В последнее время многие исследователи сосредоточились на моделировании пространственной динамики популяций в адвективной среде. Очевидно, это происходит в реках, где особи уносятся вниз по течению водным потоком. В [1] авторы предписали разные граничные условия на концах вверх и вниз по течению, мотивированные разными экологическими сценариями. Когда условия адвекции включены в классические конкурентно-диффузионные системы Лотки-Вольтерры, изучение глобальной динамики образующихся систем становится очень сложной задачей. В настоящее время уравнения реакции-диффузии часто используются в качестве стандартных моделей для решения задач, связанных с пространственной экологией и эволюцией (см.напр. [2]). В частности, А. Гастингс [3] рассматривал два конкурирующих вида с единственной разницей, заключающейся в их случайных скоростях диффузии. Авторы работы [4] исследовали более общую модель в том смысле, что двум популяциям разрешено вести себя по-разному с точки зрения скорости расселения, темпов роста и конкурентоспособности, и они успешно установили полную классификацию всех возможных долгосрочных динамических характеристик поведение. Этот успех во многом обусловлен ключевой априорной оценкой устойчивого состояния сосуществования, согласно которой каждое устойчивое состояние сосуществования, если оно существует, является линейно устойчивым. Однако, когда условия адвекции включаются в эти классические конкурентно-диффузионные системы Лотки-Вольтерры, глобальная динамика возникающих в результате систем реакции-диффузии остается далекой от полного понимания. Основная причина, с точки зрения теории операторов, заключается в том, что линеаризованный оператор, такой как диффузионная адвекция, больше не является самосопряженным, и, следовательно, различные методы, разработанные в области диффузионных моделей Лотки-Вольтерра, по-прежнему неэффективны. В [5] исследуется динамика популяций двухвидовой конкурентной системы Лотка-Вольтерра, возникающей в речной экологии. Интересная особенность этой системы моделирования заключается в граничных условиях в нижнем течении реки, где популяции могут быть подвержены гибели особей разного размера. Аналогичная задача исследована и в работе [6].

1. Постановка проблемы и о результатах

В этой работе мы попытаемся исследовать динамическое поведение следующих двух типов системы реакция-диффузия-адвекция, которая является развитием вышеупомянутых результатов. Рассматривается задача

$$\begin{cases} u_t = d_1 u_{xx} - \alpha_1 u_x + u(r-u-v), \quad 0 < x < L, t > 0 \\ v_t = d_2 v_{xx} - \alpha_2 v_x + v(r-u-v), 0 < x < L, t > 0 \\ d_1 u_x(0,t) - \alpha_1 u(0,t) = 0, d_2 v_x(0,t) - \alpha_2(0,t) = 0, t > 0, \\ d_1 u_x(L,t) - \alpha_1 u(L,t) = -b_1 \alpha_1 u(L,t), t > 0, \\ d_2 v_x(L,t) - \alpha_2(L,t) = -b_2 \alpha_2 v(L,t), t > 0, \\ \not\equiv 0, v(x,0) = v_0(x) \geqslant, \not\equiv 0, 0 < x < L, \\ u(x,0) = u_0(x) \geqslant, \not\equiv 0, v(x,0) = v_0(x) \geqslant, \not\equiv 0, 0 < x < L, \end{cases} \tag{1}$$

где $u(x,t)$ и $v(x,t)$ – плотности популяции двух конкурирующих водных видов в точке x и времени $t > 0$ соответственно. Предполагается, что два вида живут в реке с однонаправленным потоком воды, который здесь обозначен открытым интервалом $(0, L)$, и что обе популяции совершают некоторые диффузные движения за счет самодвижения со скоростью $d_1 > 0, d_2 > 0$, а также некоторые перемещения из – для потока воды вниз по потоку со скоростью адвекции $\alpha_1 > 0, \alpha_2 > 0$, а константа $r > 0$ означает внутреннюю скорость роста.Предполагается, что

(H1)

$$0 < d_1 < d_2;$$

(H2)

$$1 \leqslant b_1 \leqslant b_2 \leqslant \infty.$$

Поскольку система (1) порождает монотонную динамическую систему, глобальная динамика таких систем во многом определяется установившимися состояниями и их качественными свойствами. Основной проблемой при применении монотонных динамических систем является наличие или отсутствие сосуществующих устойчивых состояний. Для этого необходимо выяснить структуру решения следующей эллиптической задачи, соответствующей системе (1)

$$\begin{cases} d_1 u_{xx} - \alpha_1 u_x + u(r - u - v) = 0, 0 < x < L, \\ d_2 v_{xx} - \alpha_2 v_x + v(r - u - v) = 0, 0 < x < L, \\ d_1 u_x(0) - \alpha_1 u(0) = 0, d_2 v_x(0) - \alpha_2 v(0) = 0, \\ d_1 u_x(L) - \alpha_1 u(L) = -b_1 \alpha_1 u(L), \\ d_2 v_x(L) - \alpha_2 v(L) = -b_2 \alpha_2 v(L), \end{cases} \tag{2}$$

Наша основная цель – изучить, как различные стратегии распространения влияют на результат конкуренции в однородной среде. Применяя теорию главных собственных значений и монотонных динамических систем, устанавливается полное представление о глобальной динамике. Мы исследовали систему реакция-диффузия-адвекция, которую можно использовать для описания конкуренции между двумя водными видами, живущими в однонаправленной реке. Предполагается, что эти два вида конкурируют за одни и те же ресурсы, которые все реже и реже распределяются в пространстве и претерпевают разные стратегии распределения, что отражается в их скорости распространения и адвекции. Основываясь на этом предположении, мы исследуем комбинированное влияние стратегии путешествий и неоднородности окружающей среды на результат конкуренции. Оказывается, принцип конкурентного исключения выполняется в большинстве ситуаций, и какая форма имеет больше конкурентных преимуществ, зависит от соотношения значений их скоростей адвекции α_1 и α_2 при фиксированном $0 < d_1 < d_2$.

Заключение

В работе нами изучены система реакции-диффузии-адвекции, которую можно применить для описания конкуренции между двумя водными видами, живущими в реке с однонаправленным потоком воды. Предполагается, что два вида конкурируют за одни и те же ресурсы, которые постепенно распределяются в пространстве и претерпевают разные стратегии распространения, что отражено в их скорости распространения и/или адвекции. Исходя из этого предположения, мы исследуем совместное влияние стратегии передвижения и неоднородности окружающей среды на результат конкуренции. Мы разрабатываем новые идеи и аргументы для анализа качественных свойств как полутривиальных, так и сосуществующих устойчивых состояний.

СПИСОК ЛИТЕРАТУРЫ

1. *Lou Y., Lutscher F.* Evolution of dispersal in open advective environments // Math. Biol. 2014. Vol. 39. P. 1319–1342.
2. *He X., Ni W. M.* Global dynamics of the Lotka-Volterra competition-diffusion system with equal amount of total resources // Calc. Var. Partial Differential Equations. 2016. Vol. 55, № 2. P. 20–25.
3. *Hastings A.* Can spatial variation alone lead to selection for dispersal? // Theor. Popul. Biol. 1983. Vol. 24. P. 244–251.
4. *He X., Ni W. M.* Global dynamics of the Lotka-Volterra competition-diffusion system:Diffusion and spatial heterogeneity // Comm. Pure. Appl. Math. 2016. Vol. 69. № 1. P. 981–1014.
5. *Li K., Xu F.* Global dynamics of a population model from river ecology // Journal of Applied Analysis and Computation. 2020. Vol. 10. № 4. P. 1698–1707.
6. *Tang D., Chen Y. M.* Global dynamics of a Lotka-Volterra competitiondiffusion system in advective homogeneous environments // Differential Equations. 2020. T. 269.C. 1465–1483.

Тахиров Жозил Останович — доктор физико-математических наук, профессор, заведующий лабораторией, Математическое моделирование нелинейных систем. E-mail: *prof.takhirov@yahoo.com*

Боборахимова Махбуба Ихтиёровна — базовый докторант, Математическое моделирование нелинейных систем. E-mail: *kamina9314@mail.ru*

ЗАДАЧА MARL ДЛЯ СВЕТОФОРНОЙ СЕТИ

Т. И. Тисленко[1], Д. В. Семенова[1,2]

[1] *Сибирский федеральный университет, г. Красноярск, Россия*
[2] *Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого, г. Красноярск, Россия*

Рассматривается проблема оптимизации планов координаций на основе управления выбором задержки сигнала для светофорных объектов участка дорожной сети. В качестве математической модели светофорной сети используется управляемый марковский процесс с конечным числом действий и состояний, а задача минимизации задержек движения транспортных средств сводится к задаче мультиагентного обучения с подкреплением (MARL). Для исследования представленной модели была разработана программа имитационного моделирования в системе AnyLogic и проведены серии вычислительных экспериментов.
Ключевые слова: *MARL, Q-learning, MDP, система управления светофорами.*

Введение

Проблема уменьшения задержек в автомобильном движении актуальна и востребована в современных мегаполисах. Одним из подходов к решению этой проблемы является оптимизации планов координаций на основе управления выбором задержки сигнала для светофорных объектов участка дорожной сети. На данный момент существует множество адаптивных систем управления трафиком, основанных на интенсивности транспортного потока — числе транспортных средств, проезжающих через сечение дороги за единицу времени [1, 2]. В качестве альтернативного способа координации светофоров в последнее время исследуются модели, основанные на марковских процессах принятия решений [3]. Данные модели могут не оперировать интенсивностью потока. В работе рассматривается подход, основанный на мультиагентном программировании. В качестве агентов рассматриваются светофоры, взаимодействующие в среде, представленной сетью перекрёстков с машинами. Предполагается, что состояние среды отражает активность фаз светофоров. В обучении с подкреплением используются вознаграждения агента, определенные на каждом шаге и зависящие от текущего состояния среды и совершенного действия. В качестве вознаграждения в модели используется суммарное время, затраченное машинами на преодоление перекрестка. Отсчёт времени проезда начинается за 100 м до стоп-полосы и заканчивается при преодолении перекрестка. Задача каждого агента — выбрать допустимый сигнал, который обеспечит максимальное суммарное вознаграждение. В качестве математической модели светофорной сети используется управляемая однородная марковская цепь с конечным числом действий и состояний, а задача минимизации задержек движения транспортных средств сводится к задаче мультиагентного обучения с подкреплением (MARL — Multiagent Reinforcement Learning) [3, 4]. Для нахождения решения используется алгоритм Q-обучения [5], определяющий оптимальный выбор действий каждого из агентов в текущем состоянии среды с учетом совокупных вознаграждений. Для исследования представленной модели была разработана программа имитационного моделирования в системе AnyLogic и проведены серии вычислительных экспериментов.

1. Задача MARL для светофорной сети

Пусть светофор — это агент, не располагающий ресурсами. В качестве среды будем рассматривать перекресток с машинами, где на отрезках дорог за 100 м до стоп-линий засекается время проезда через него. Состояние среды отражает активность фазы светофора. Обозначим пространство состояний агента как S, а множество решений агента как A. В работе исследуются двухфазные светофоры как агенты с пространством состяний $S = \left\{s^{(0)} = 0, s^{(1)} = 1\right\}$, где $s^{(i)}$ интерпретируется как «активна фаза i», и с множеством решений $A = \left\{a^{(0)} = 0, a^{(1)} = 1\right\}$, здесь $a^{(0)}$ интерпретируется как «оставить фазу», а $a^{(1)}$ — «сменить фазу». Фазы светофора меняются последовательно. Состояние s', в которое перейдёт система из состояния s при решении a, однозначно определяется по формуле $s' = (a + s) \bmod 2$. Поведение агента в момент времени t будем представлять парой объектов (s_t, a_t), где $s_t \in S$ — активная фаза светофора, а $a_t \in A$ — соответствующее t решение. Будем считать, что смена фаз светофора описывается марковской цепью с дискретным временем $\xi(t)$. А семейство решений a_t образует стратегию $\delta = \{a_t, 0 \le t < \infty\}$. Тогда поведение агента будем описывать управлемым однородным марковским процессом (ξ, δ) с конечным числом действий A и состояний S и дискретным временем [6]. На рис. 1 представлен стохастический граф управляемого процесса смены фазы для двухфазного светофора.

Под задержкой на фазе будем понимать суммарное засечённое время всех машин, проходящих через отрезок дороги. Задержка на фазе s_t после действия a_t есть функция $r(s_t, a_t; s_{t+1})$, которая всецело опреде-

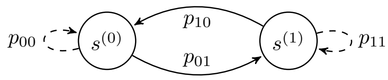

Рис. 1. Стохастический граф управляемого процесса смены фаз для двухфазного светофора. Здесь пунктиром обозначены переходы при действии $a^{(0)}$, а сплошной линией переходы при решении $a^{(1)}$. Вероятность p_{ij} есть вероятность перехода из состояния $s^{(i)}$ в $s^{(j)}$, $i, j = 0, 1$.

ляется текущим состоянием s_t, выбранным решением a_t и состоянием s_{t+1}, в которое перейдет процесс на следующем шаге. Тогда функция суммарных доходов/вознаграждений агента при выбранной стратегии примет вид $V\Big(\{S_t, \delta\}\Big) = \sum\limits_{t=0}^{\infty} \gamma^t r(s_t, a_t)$, где $0 \leq \gamma \leq 1$ — коэффициент переоценки, $\{S_t\}$ — реализация случайного процесса $\xi(t)$ при выбранной стратегии $\delta = \{a_t, 0 \leq t < \infty\}$.

Задача обучения с подкреплением (RL — Reinforcement Learning) для одного агента в общем виде формулируется следующим образом: найти такую оптимальную стратегию $\delta^* = \{a_t^*, 0 \leq t < \infty\}$ поведения агента, при которой функция суммарных доходов $V\Big(\{S_t, \delta\}\Big)$ достигает своего масимума.

Решение задачи RL для одного светофора ищется методом динамического программирования на основе принципа оптимальности Беллмана и базируется на идее Q-обучения, которая заключается в оценке невычислимой правой части [6, 5, 4]:

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha_t \left(r(s, a) + \gamma \max_{a' \in A} [Q_t(s', a')] - Q_t(s, a) \right). \quad (1)$$

В [4] установлен критерий сходимости итерационного процесса (1). Таким образом, оптимальное управление для одного светофора в момент времени t ищется в виде $a_t = \arg\max\limits_{a' \in A} Q_t(s, a')$.

Расширим задачу обучения с подкреплением на сеть светофоров. Рассмотрим множество из K агентов, для каждого из которых описаны множество состояний S^k и множество решений A^k, $k = 1, \ldots, K$. Обозначим $\mathbf{s}_t = \{s_t^1, s_t^2, \ldots, s_t^K\} \in S^1 \times S^2 \times \ldots \times S^K$ — совокупное состояние среды в момент времени t, а $\mathbf{a}_t = \{a_t^1, a_t^2, \ldots, a_t^K\} \in A^1 \times A^2 \times \ldots \times A^K$ — совокупное управление в момент времени t. Отметим, что смена фазы любым агентом приводит к изменению общего состояния среды $\mathbf{s}$. Для упрощения реализации принято рассматривать взаимодействие агентов попарно. Для каждого светофора k определим множество соседних агентов $N(k)$, с которыми он взаимодействует. Для фиксированной пары агентов k и $j \in N(k)$ множество их совместных действий $\mathbf{a}^{kj} \in A^k \times A^j$ и совместных состояний $\mathbf{s}^{kj} \in S^k \times S^j$ представимы в виде матриц размерности $|A^k| \times |A^j|$ и $|S^k| \times |S^j|$ соответственно. Для случая, когда рассматриваются сеть светофоров с двумя состояниями, следующую фазу светофора можно найти с помощью операции «исключающего или»: $\mathbf{s}' = \mathbf{a} \oplus \mathbf{s}$. Оптимальное управление для фиксированного агента k будем искать как решение задачи MARL в виде [3]: $a_k = \arg\max\limits_{a_k \in A^k} \sum\limits_{j \in N(k)} \sum\limits_{a_j \in A^j} Q_t^{kj}(\mathbf{s}, \mathbf{a}^{kj}) p(\mathbf{s}'|(\mathbf{s}, \mathbf{a}^{kj}))$. Здесь функция $Q_t^{kj}(s, a)$ определяется аналогично (1) с учётом множеств совместных состояний и действий агента k с множеством его соседей $N(k)$. Вероятность $p\left(\mathbf{s}'|(\mathbf{s}, \mathbf{a}^{kj})\right)$ — это вероятность того, что агент j выберет действие a_j с учётом текущего совместного состояния $\mathbf{s}$ и выбранного агентом k действия a_k.

2. Вычислительные эксперименты

Для исследования представленной модели была разработана программа имитационного моделирования в системе AnyLogic и проведены серии вычислительных экспериментов. Эксперименты проводились на ПК с процессором Intel Core i7-10510U CPU @ 1.80ГГц и оперативной памятью объемом 8ГБ. Целью экспериментов было сравнение времени задержки машин для двух моделей. Для первой модели длительность фаз (TrafficLight state) для первого (TLs0, TLs1) и второго светофора (TL1s0, TL1s1) получена перебором. Для второй модели длительность фаз получена как решение задачи MARL на основе Q-обучения. На рис. 2 представлены вторая модель светофроной сети и соответствующий ей стохастический граф управляемого однородного марковского процесса смены фазы для сети из двух двухфазных светофоров. Имитационное моделирование процесса управления светофором проводилось с учётом следующих условий: машины прибывают на каждый из перекрестоков с трёх направлений, отмеченных зеленым цветом, с интенсивностью 1000 в час и уничтожаются в областях, отмеченных синим; направления движений машин указаны белыми стрелками на перекрестках; в модели управляемой однородной марковской цепи с конечным числом действий и состояний коэффициенты скидки α и переоценки γ подобраны эмпирическим путём; дискретизация по времени равна 5 секундам. При пересечении полосы, на расстоянии 100 метров до стоп-линии, пары, состоящие из указателей на объект машины и текущего времени модели, добавляются в одну из коллекций $tcf, \ldots, tcf7$ (time collection forward). Далее, машины удаляются из коллекции, при проезде через перекресток и разница времени системы и хранящегося в

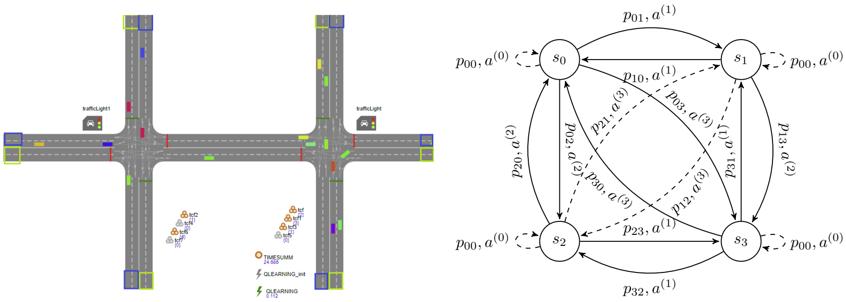

Рис. 2. Модель сети из двух двухфазных светофоров и соответствующий ей стохастический граф управляемого процесса смены фаз. Здесь s_l отражает совокупное состояние светофоров l (l =0b00, 0b01, 0b10, 0b11). При действии $a^{(m)}$ (m =0b00, 0b01, 0b10, 0b11) процесс перейдет в состояние $n = l \oplus m$, с вероятностью p_{ln}.

коллекции прибавляется к величине TIMESUMM. Для второй модели в течении 5 секунд вызывается событие QLEARNING, реализующее решение задачи MARL. На его основе принимается решение об остановке или продлении фаз светофоров. Модель, управляемая марковским процессом, показала уменьшение суммарной задержки в среднем в 1.2 раза по сравнению с системой управления светофором, длительность фаз которой подобрана перебором (рис. 3). Синим цветом указано лучшее допустимое значение $TIMESUMM$. Серые точки показывают некоторые наиболее частые отклонения от лучшего значения.

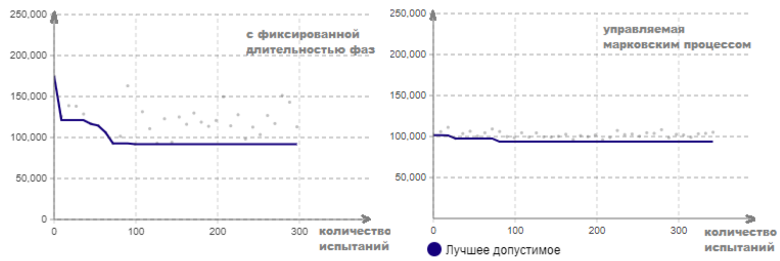

Рис. 3. Значение $TIMESUMM$ для моделей

Рис. 4. Справа на оси ординат показано значение $TIMESUMM$ для модели, управляемой марковским процессом, слева напоказано значение $TIMESUMM$ для модели перебора. На оси абсцисс указано количество экспериментов.

Заключение

В статье исследуется математическая модель процесса выбора фаз сети светофоров, которая учитывает состояния светофоров и их загруженность. Задача MARL формулируется как для двух, так и для любого числа светофоров на перекрестке. Решение задачи минимизации задержки трафика рассматривается как итерационный процесс, для которого получен критерий сходимости и представлен алгоритм поиска решения.

СПИСОК ЛИТЕРАТУРЫ

1. *Goel S., Bush S., Gershenson C.* Self-Organization in Traffic Lights: Evolution of Signal Control with Advances in Sensors and Communications // University at Albany. ArXiv, 2017. 1708 c.
2. *Manolis D., Diakaki C., Papamichail I.* Simulation investigations of the coordinated traffic-responsive signal control strategy TUC with actuation at the local junction level // European Transport Research Review. 2018. C.25.
3. *El-Tantawy S., Abdulhai B. and Abdelgawad H.* Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto // IEEE Transactions on Intelligent Transportation Systems. 2013. T. 14, №. 3, C. 1140-1150.
4. *Тисленко Т. И.* Задача MARL для светофора на перекрёстке // Материалы VIII Международной молодежной научной конференции «Математическое и программное обеспечение информационных, технических и экономических систем»,Томск: Издательство Томского государственного университета, 2021. С. 144–234
5. *Гасников А. В., Горбунов Э. А.,Гуз С. А.* Лекции по случайным процессам: учебное пособие // Москва: «МОСКВА»: МФТИ, 2019. 285 с.
6. *Рыков В. В.* Управляемые марковские процессы с конечными пространствами состояний и управлений // Теория вероятн. и ее примен. 1966. Т. 11, выпуск 2. С. 343–351.

Семенова Дарья Владиславовна — кандидат физико-математических наук, доцент, доцент кафедры высшей и прикладной математики, Институт математики и фундаментальной информатики СФУ. E-mail: *DVSemenova@sfu-kras.ru*

Тисленко Тимофей Иванович — студент 4-го курса бакалавриата, Институт математики и фундаментальной информатики СФУ. E-mail: *timtisko@mail.ru*

СОДЕРЖАНИЕ

Научное издание

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ (ИТММ-2021)

МАТЕРИАЛЫ
XX Международной конференции
имени А. Ф. Терпугова
1–5 декабря 2021 г.

Редактор *В.Г. Лихачева*
Компьютерная верстка *Д.В. Семенова, Е.Ю. Лисовская, 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 12.09.2022 г.
Формат $60 \times 84^{1}/_{16}$. Бумага для офисной техники. Гарнитура «Times».
Печ. л. 24.5. Усл. печ. л. 22.7. Тираж 500 экз. Заказ № 5145.

ISBN 978-5-907572-20-1
9 785907 572201 >