

Информатика и её применения

Том 12 Выпуск 4 Год 2018

СОДЕРЖАНИЕ

Применение моделей случайного блуждания при моделировании перемещения устройств в беспроводной сети	
К. Е. Самуйлов, Ю. В. Гайдамака, С. Я. Шоргин	2
Comparison of two active queue management schemes through the $M/D/1/N$ queue	
M. G. Konovalov and R. V. Razumchik	9
Определение экстремальности объемов осадков на основе модифицированного метода превышения порогового значения	
А. К. Горшенин, В. Ю. Королев	16
Оптимизация объема буферной памяти узла коммутации при схеме полного разделения памяти	
Я. М. Агаларов	25
Система обслуживания с относительным приоритетом и профилактиками прибора	
Е. С. Кондранин, В. Г. Ушаков	33
Методы идентификации захвата хоста в распределенной информационно-вычислительной системе, защищенной с помощью метаданных	
А. А. Грушо, Н. А. Грушо, М. В. Левыкин, Е. Е. Тимонина	39
Использование вероятностной модели вычислений для тестирования одного класса готовых к использованию программных компонентов локальных и сетевых систем	
Б. М. Басок, В. Н. Захаров, С. Л. Френкель	44
Seamless route updates in software-defined networking via quality of service compliance verification	
S. L. Frenkel and D. Khankin	52
Выбор оптимальной модели рекуррентной сети в задачах поиска парафразы	
А. Н. Смердов, О. Ю. Бахтеев, В. В. Стрижов	63
Теоретико-информационная характеристика совершенно уравновешенных функций	
О. А. Логачев	70
Новые представления обобщенного распределения Миттаг-Леффлера в виде смесей и их приложения	
В. Ю. Королев, А. К. Горшенин, А. И. Зейфман	75
О неравномерных оценках точности нормальной аппроксимации для распределений некоторых случайных сумм при ослабленных моментных условиях	
В. Ю. Королев, А. В. Дорофеева	86
Гамма-вейбулловский случай априорных распределений в байесовских моделях массового обслуживания	
Е. Н. Арутюнов, А. А. Кудрявцев, А. И. Титова	92
Количественный анализ результатов машинного перевода с использованием надкорпусных баз данных	
Н. В. Бунтман, А. А. Гончаров, И. М. Зацман, В. А. Нуриев	96
Развитие сервисов цифровых платформ для преодоления нефинансовых барьеров	
А. К. Горшенин	106
Об авторах	113
Авторский указатель за 2018 г.	115
2018 Author Index	119
Правила подготовки рукописей	124
Requirements for manuscripts	127

Технический редактор *Л. Кокушкина* Художественный редактор *М. Седакова*

Сдано в набор 06.12.18. Подписано в печать 24.12.18. Формат 60 x 84 / 8

Бумага офсетная. Печать цифровая. Усл.-печ. л. 14,88. Уч.-изд. л. 14,5. Тираж 100 экз.

Заказ № 1055

Издательство «ТОРУС ПРЕСС», Москва 121614, ул. Крылатская, 29-1-43

Отпечатано в НИПКЦ «Восход-А» с готовых файлов

Москва 109052, ул. Смирновская, д. 25, стр. 3

ПРИМЕНЕНИЕ МОДЕЛЕЙ СЛУЧАЙНОГО БЛУЖДЕНИЯ ПРИ МОДЕЛИРОВАНИИ ПЕРЕМЕЩЕНИЯ УСТРОЙСТВ В БЕСПРОВОДНОЙ СЕТИ*

К. Е. Самуйлов¹, Ю. В. Гайдамака², С. Я. Шоргин³

Аннотация: Выполнен обзор моделей случайного блуждания объектов, применяемых при моделировании передвижения приемопередающих устройств пользователей беспроводной сети пятого поколения (5G). Рассмотрены модели мобильности, характерные для имитационного моделирования поведения пользователей беспроводной самоорганизующейся сети. Обсуждаются особенности различных моделей индивидуального движения объектов, а также моделей движения групп объектов с точки зрения применения к анализу интерференции в беспроводных сетях. Цель статьи — предложить ряд моделей мобильности для принятия обоснованного решения при выборе модели случайного блуждания для анализа качества предоставления услуг в беспроводных сетях. В качестве иллюстрации применения разработанного авторами комплекса аналитических и имитационных моделей проведен анализ отношения сигнал/интерференция, определяющего качество предоставления услуг в сетях пятого поколения, для сценария случайного блуждания мобильных абонентов в торгово-развлекательном центре при использовании модели случайного блуждания с остановками Random Waypoint.

Ключевые слова: модель случайного блуждания; модель мобильности; отношение сигнал/интерференция; отношение сигнал/шум

DOI: 10.14357/19922264180401

1 Введение

В беспроводных сетях 5G интерференция служит одним из существенных источников помех, влияющим на показатели качества функционирования сети, к которым относятся пиковые скорости передачи данных между устройствами, задержка начала передачи данных, отношение сигнал/интерференция (signal to interference ratio, SIR), энергопотребление, эффективность использования частотного спектра и др. [1]. При анализе интерференции следует учитывать особенности современных беспроводных сетей, которые при использовании в них технологии прямого взаимодействия оконечных устройств (device-to-device, D2D) образуют самоорганизующиеся сети (mobile ad hoc network, MANET) с перемещающимися в зоне покрытия узлами. Относительно небольшие расстояния между подвижными узлами, соответствующими приемопередающим устройствам, делают необходимым при анализе интерференции между соседними источниками сигнала в таких сетях учитывать траектории

перемещения узлов, которые фактически определяют динамику показателя SIR в радиоканале между приемником и передатчиком.

В работах [2, 3] перемещение беспроводных устройств моделировалось с помощью кинетического уравнения Фоккера–Планка [4], регулирование параметров (снос, диффузия) которого позволяет исследовать различные типы движения большого числа объектов, не строя индивидуальную траекторию перемещения каждого объекта.

Однако такой подход не применим при анализе интерференции в задачах, где необходимо принимать во внимание особенности предоставления услуг передачи данных в сети, например учитывать препятствия в зоне перемещения устройств, наличие нескольких сред распространения сигнала и другие ограничения. Для решения таких задач необходимо детальное моделирование траектории движения каждого беспроводного устройства с помощью аналитических [5–7] и имитационных [8, 9] моделей.

* Исследование выполнено при финансовой поддержке Российского научного фонда (проект 16-11-10227).

¹ Российский университет дружбы народов; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, samouylov_ke@rudn.university

² Российский университет дружбы народов; Федеральный исследовательский центр «Информатика и управление» Российской академии наук, gaydamaka_yuv@rudn.university

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, ssorgin@ipiran.ru

В статье проведен обзор моделей мобильности объектов, применяемых при моделировании перемещения устройств в беспроводных самоорганизующихся сетях. В разд. 2 обсуждаются особенности различных моделей индивидуального движения объектов, а также моделей движения групп объектов с точки зрения применения к анализу интерференции в беспроводных сетях. В качестве иллюстрации в разд. 3 на примере прикладной задачи анализа движения мобильных абонентов в торгово-развлекательном центре с помощью разработанного авторами комплекса аналитических и имитационных моделей на наборе исходных данных, близких к реальным, проведен расчет показателя SIR, определяющего качество предоставления услуг в сетях пятого поколения.

2 Модели случайного блуждания

В настоящее время для имитационного моделирования передвижения беспроводных устройств в сетях MANET традиционно используются как модели мобильности объектов, основанные на сборе и анализе статистики движения абонентов в реальных беспроводных сетях (traces), так и синтетические (synthetic) модели [8–10]. Первые построены на основе обработки данных от большого числа узлов сети, собранных в течение длительного периода наблюдения, поэтому обеспечивают достоверное моделирование, однако их применение возможно лишь для анализа уже существующих сетей. Поскольку сети 5G в полной мере еще не реализованы, востребованными оказались синтетические модели, с помощью которых можно реалистично воспроизводить поведение абонентов беспроводной сети, регулируя правила изменения скорости и направления движения мобильных узлов. Например, мобильные узлы не должны иметь прямую траекторию движения и постоянную скорость в течение всего времени моделирования, потому что в реальных сетях движение абонентов имеет более сложный характер. Как правило, для моделирования движения абонентов беспроводной сети используются как модели мобильности, описывающие независимое друг от друга движение абонентов беспроводной сети, модели так называемой «индивидуальной» мобильности (entity mobility models), так и модели «групповой» мобильности (group mobility models), в которых, например, движение группы основано на траектории логического центра [9–11].

Примерами моделей индивидуальной мобильности, используемых для описания перемещения независимых объектов, являются модель RW (Random Walk), для которой в каждой точке направ-

ление и скорость движения разыгрываются случайным образом (рис. 1, а), и ее расширение — модель RWP (Random WayPoint), в которой предусмотрено время остановки в каждой точке перед продолжением движения (рис. 1, б).

Модели RW и RWP чаще других используются для моделирования движения мобильных устройств в самоорганизующейся беспроводной сети. Особенностью обеих моделей является отсутствие «памяти» — текущие значения параметров модели (направление, скорость, длительность остановки) не зависят от значений этих параметров на прошлом шаге, что приводит к генерации траекторий с внезапными остановками и резкими поворотами. При небольших значениях скорости в модели RW движение объектов становится броуновским; следовательно, эту модель можно рекомендовать для анализа интерференции в статической сети.

Интересным развитием модели RWP является ее вероятностная версия, в которой следующая позиция мобильного узла определяется в соответствии с заданными вероятностями. Недостатком модели RWP является замеченная особенность — при достаточно длительном периоде работы имитационной модели плотность объектов, в начале моделирования распределенных равномерно, по краям области моделирования становится заметно ниже, чем в центре. Поскольку расстояние до передатчика является ключевым фактором, оказывающим влияние на мощность фиксируемого на приемнике сигнала, безосновательное увеличение числа ближайших интерферирующих передатчиков при оценке показателя SIR исказит вывод о качестве соединения в моделируемой сети. Однако при моделировании некоторых сценариев поведения пользователей, например осмотра музея в соответствии с предлагаемой схемой знакомства с экспозицией, модель случайного блуждания RWP за счет своей гибкости создает реалистичные траектории движения объектов. Кроме того, описанный эффект скопления объектов в центре области моделирования практически исчезает для случая долгих остановок даже при высоких значениях параметра скорости [10].

Указанного для модели RWP недостатка лишены модель RD (Random Direction), для которой направление и скорость движения меняются при достижении объектом границы области моделирования (рис. 1, в), а также модель движения Гаусса–Маркова (Gauss–Markov, GM), которая позволяет получить плавную траекторию движения объекта (рис. 1, г). В [10] описан метод, который для модели GM принудительно меняет направление движения объекта при приближении к границе области моделирования, что позволяет избежать нежелатель-

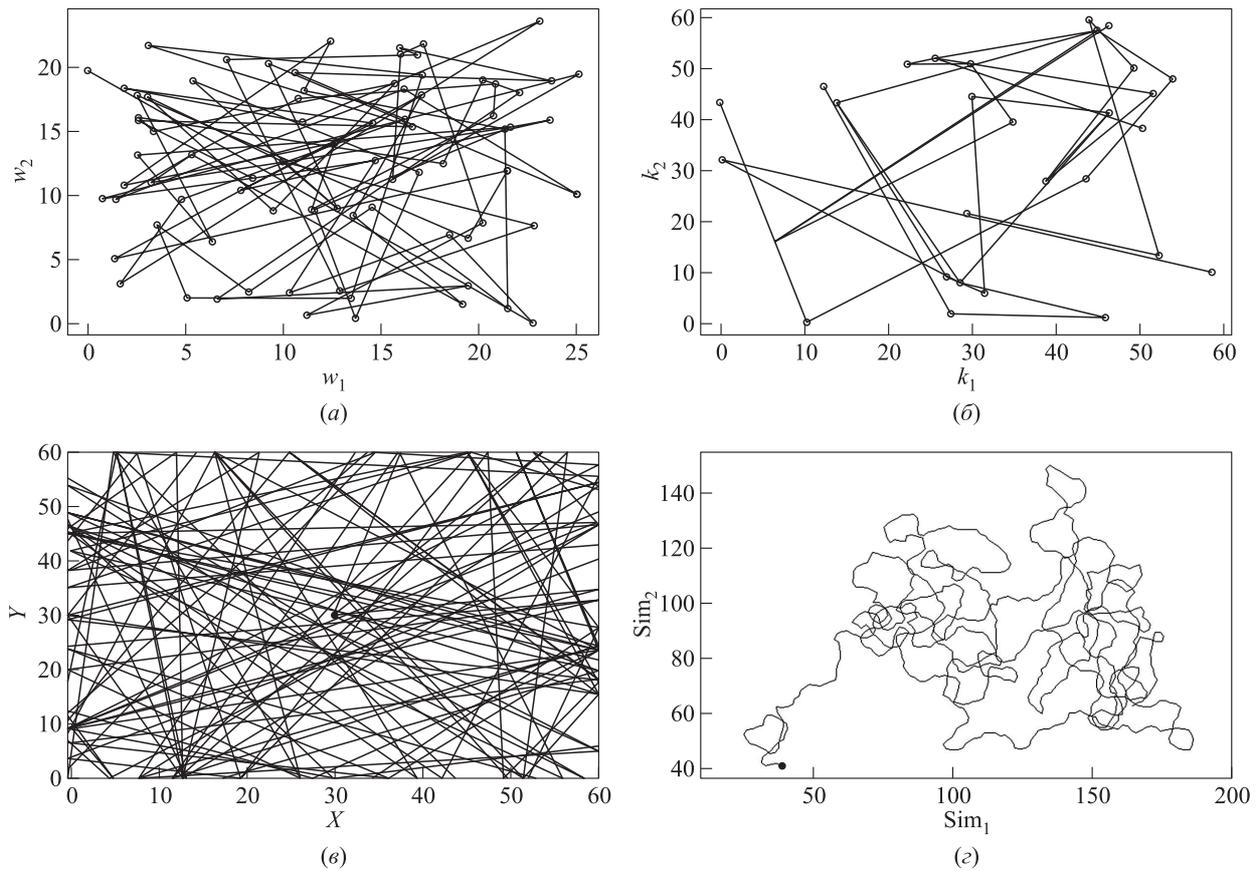


Рис. 1 Примеры траекторий перемещения объекта при случайном блуждании: (а) модель RW; (б) модель RWP; (в) модель RD; (г) модель GM

ных эффектов «прилипания» объекта к краю области.

Характерная для всех упомянутых выше моделей проблема «краевого эффекта» при приближении объекта к границе области отсутствует в модели Boundless Simulation Area (BSA), область моделирования которой представляет собою тор. В этой модели текущие значения направления и скорости движения зависят от значений этих параметров на прошлом шаге, что создает реалистичную траекторию движения объекта. Однако при моделировании беспроводной сети с помощью модели BSA не избежать искажения динамики показателя SIR, поскольку движущееся по поверхности тора беспроводное устройство регулярно становится соседом какого-либо неподвижного беспроводного устройства.

Еще одной моделью индивидуальной мобильности является так называемое «вероятностное блуждание по сетке» (Probabilistic Grid, PG) — модель в дискретном времени, согласно которой на каждом временном такте объект делает шаг единичной длины, а выбор одного из четырех направлений задается вероятностной матрицей [10]. Благодаря

простоте реализации эта модель также широко применяется при моделировании движения, однако задание вероятностной матрицы для конкретного сценария поведения пользователей представляет определенную трудность.

К моделям групповой мобильности [10, 11] относятся модель ECRM (Exponential Correlated Random Mobility), основанная на экспоненциальной зависимости скорости движения объектов; модель CM (Column Mobility), в которой моделируется движение объектов, выстроенных в линию; модель перемещения кочевников NCM (Nomadic Community Mobility); модель преследования (Pursue Mobility), в которой группа следует за лидером, передвигающимся по заданной траектории, а также наиболее общая модель групповой мобильности с опорной точкой RPGM (Reference Point Group Mobility), в которой предусмотрено случайное движение группы с одновременным случайным перемещением каждого отдельного объекта внутри группы (рис. 2). Недостатком модели ECRM, которая позволяет описать практически все виды групповой мобильности, является существенная сложность подбора параметров модели.

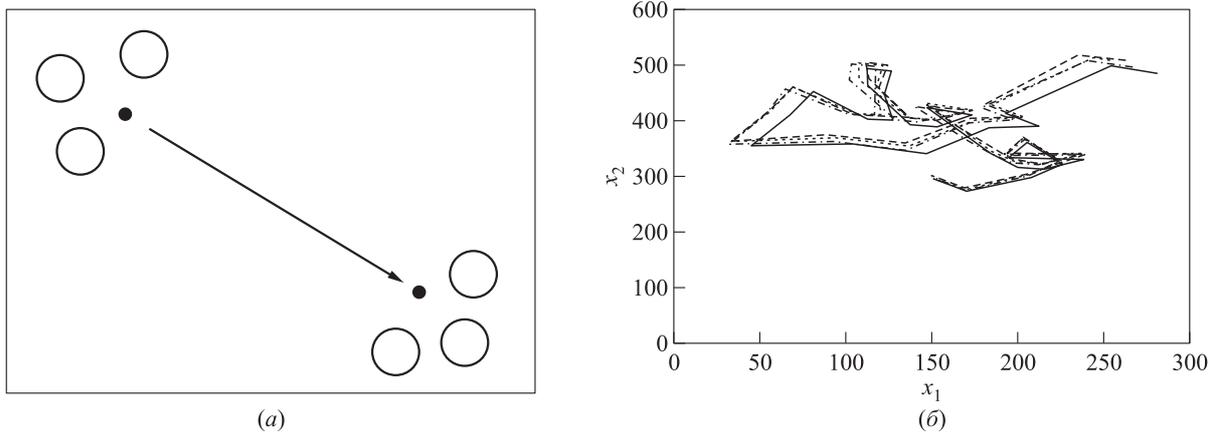


Рис. 2 Пример траекторий для модели групповой мобильности с опорной точкой RPGM для трех объектов: (а) перемещение опорной точки; (б) траектории перемещения объектов

Перечисленные модели движения традиционно используются при исследовании производительности различных сетевых протоколов, применяемых в самоорганизующихся беспроводных сетях, при этом сравнение проводится по таким показателям, как переданные полезная и служебная нагрузка, джиттер, межконцевая задержка, затраты на маршрутизацию [11]. При выборе модели мобильности с целью исследования интерференции важно учитывать сценарий поведения пользователей. Наиболее универсальными моделями индивидуальной мобильности являются модель Random Waypoint и модель Гаусса–Маркова, настройка параметров которых позволяет гибко подстроиться под большинство сценариев. Для воспроизведения перемещения группы пользователей беспроводной сети рекомендуется использовать модель групповой мобильности с опорной точкой Reference Point Group Mobility, которая при соответствующих значениях параметров позволяет реализовать модели Column, Nomadic Community и Pursue.

3 Пример анализа интерференции при случайном блуждании

Одной из основных характеристик качества канала в беспроводных сетях связи служит отношение уровня сигнала к уровню интерференции и шума (ОСШ, *англ.* Signal to Interference and Noise Ratio, SINR), которое определяется отношением мощности принимаемого сигнала от соответствующего передатчика к суммарной мощности шума и принимаемого сигнала от интерферирующих передатчиков [12]. При этом мощности фиксиру-

емого на приемнике сигнала как от целевого, так и от каждого из интерферирующих передатчиков определяются с соответствием с классической моделью распространения сигнала, а именно: прямо пропорционально базовой мощности сигнала передатчика и обратно пропорционально расстоянию между передатчиком и приемником в некоторой постоянной степени, зависящей от среды распространения сигнала. Как и в [12, 13], для оценки отношения SIR далее используется формула

$$SIR = \frac{r_0^{-\gamma_0}}{\sum_{n=1}^N d_n^{-\gamma_n}},$$

где r_0 — расстояние между приемником и передатчиком в исследуемой целевой паре; d_n — расстояние между приемником целевой пары и передатчиком n -й интерферирующей пары; γ — коэффициент распространения сигнала, характеризующий среду передачи (от 2 в условиях прямой видимости до 6 в худшем случае, при котором возможна связь). Расчет проведен в предположении о равных излучаемых мощностях и коэффициентах усиления приемной и передающей антенн для всех устройств. Для моделирования препятствий для прохождения сигнала использовались различные значения коэффициентов распространения сигнала γ_n , $n = 0, 1, \dots, N$ [13].

На рис. 3 для одной из моделей индивидуальной мобильности, модели случайного блуждания RWP, приведена кривая, показывающая изменение показателя SIR в течение 500 с. Для наглядности выбран случай блуждания по сетке четырех мобильных устройств — целевой пары приемник-передатчик и двух интерферирующих передатчиков, работающих на близких частотах.

Моделировалось целенаправленное движение, когда пользователи, носители мобильных

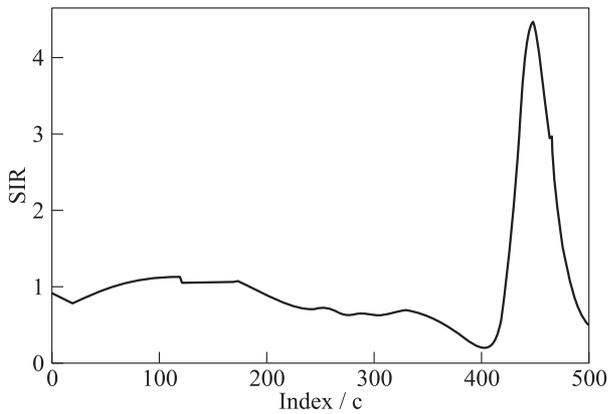


Рис. 3 Динамика SIR

устройств, перемещались по кратчайшему пути между заранее выбранными точками своих маршрутов с постоянной скоростью 1 м/с независимо друг от друга в квадрате 500×500 м. Такой сценарий характерен, например, для последовательного посещения магазинов торгово-развлекательного центра по заранее намеченному маршруту. Траектории устройств показаны на рис. 4, где перемещение целевого приемника, на котором оценивалось отношение сигнал/интерференция, показано сплошной линией, перемещение передатчиков — пунктирными линиями, при этом целевому передатчику соответствует траектория с самым длинным размером штриха.

На рис. 4 точки, отмеченные крестиками на соответствующих траекториях, позволяют судить о взаимном расположении устройств. Для наглядности целевые передатчик и приемник соединены показывающим направление передачи в радиоканале вектором, модуль которого равен расстоянию

в целевой паре. Так, на рис. 4, а, который соответствует 402-й секунде моделирования, один из интерферирующих передатчиков расположен значительно ближе к приемнику, чем целевой передатчик, а на рис. 4, б отражена обратная ситуация, когда на 449-й секунде расстояние в целевой паре становится минимальным. Соответствующие локальные экстремумы показателя SIR в указанные моменты отражены на рис. 3.

Предложенный метод расчета SIR, основанный на моделировании траекторий движения устройств, позволяет оценивать качество предоставления услуг в сети при заданных для каждой услуги требованиях к минимальному допустимому значению этого показателя.

4 Заключение

Проведенный в работе обзор моделей случайного блуждания, традиционно применяемых для моделирования перемещения мобильных узлов в беспроводных самоорганизующихся сетях, позволяет при выборе модели для проведения эксперимента учесть особенности каждой модели, существенные с точки зрения сценария поведения пользователей. Универсальной модели мобильности, способной воспроизвести поведение пользователя при любом сценарии, не существует, поэтому анализ интерференции рекомендуется проводить, применяя несколько моделей движения объектов. Также задачей дальнейших исследований может стать разработка новой комбинированной модели мобильности для воспроизведения перемещения пользователей беспроводной самоорганизующейся сети, сочетающей подход модели Gauss–Markov на границе области моделирования и принцип перемещения объектов

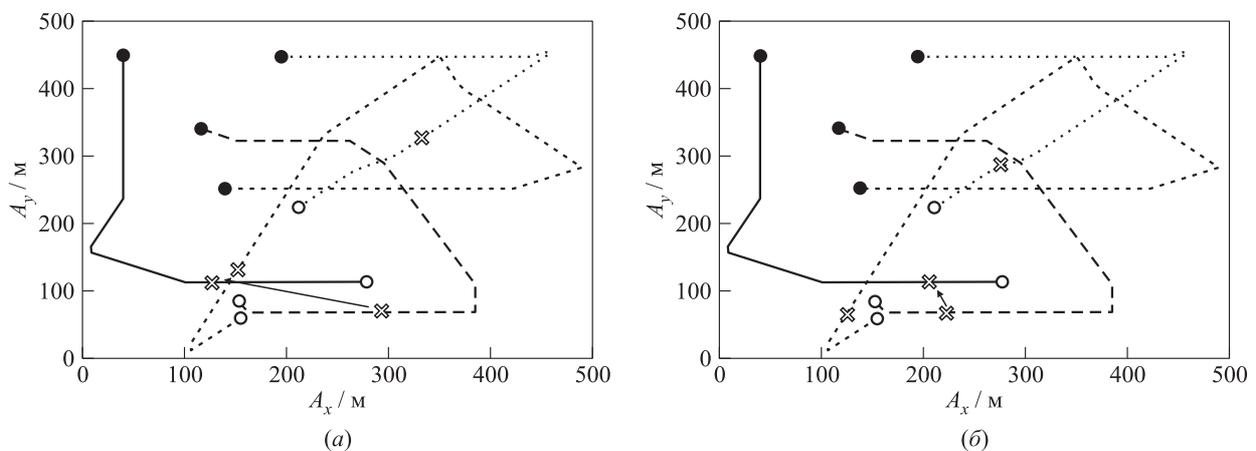


Рис. 4 Траектории и взаимное расположение устройств: (а) минимальное значение SIR (402-я секунда); (б) максимальное значение SIR (449-я секунда)

модели Random Waypoint внутри области, таким образом сохраняя преимущества и компенсируя недостатки каждой из этих моделей.

Авторы выражают благодарность магистрам кафедры прикладной информатики и теории вероятностей РУДН А. Жданкову и О. Крупко за подготовку иллюстраций к статье по разработанному комплексу аналитических и имитационных моделей.

Литература

1. Andrews J. G., Buzzi S., Choi W., Hanly S. V., Lozano A., Soong A. C., Zhang J. C. What will 5G be? // IEEE J. Sel. Area. Comm., 2014. Vol. 32. No. 6. P. 1065–1082. doi: 10.1109/JSAC.2014.2328098.
2. Orlov Yu. N., Fedorov S. L., Samuylov A. K., Gaidamaka Yu. V., Molchanov D. A. Simulation of devices mobility to estimate wireless channel quality metrics in 5G networks // AIP Conf. Proc., 2017. Vol. 1863. P. 090005-1–090005-3. doi: 10.1063/1.4992270.
3. Гайдамака Ю. В., Орлов Ю. Н., Молчанов Д. А., Самуйлов А. К. Моделирование отношения сигнал/интерференция в мобильной сети со случайным блужданием взаимодействующих устройств // Информатика и её применения, 2017. Т. 11. Вып. 2. С. 50–58. doi: 10.14357/19922264170206.
4. Risken H., Frank T. The Fokker–Planck equation: Methods of solution and applications. — Springer ser. in synergetics. — Berlin–Heidelberg: Springer-Verlag, 1996. Vol. 18. 486 p. doi: 10.1007/978-3-642-61544-3.
5. Тоффоли Т., Марголюс Н. Машины клеточных автоматов / Пер. с англ. — М.: Мир, 1991. 283 с. (Toffoli T., Margolus N. Cellular automata machines. — The MIT Press, 1987. 276 p.)
6. Grewal M. S., Andrews A. P. Kalman filtering: Theory and practice using MATLAB. — 2nd ed. — John Wiley & Sons, Inc., 2001. 410 p.
7. Семушин И. В., Цыганов А. В., Цыганова Ю. В., Голубков А. В., Винокуров С. Д. Моделирование и оценивание траектории движущегося объекта // Вестник ЮУрГУ. Сер. Матем. моделирование и программирование, 2017. Т. 10. № 3. С. 108–119. doi: 10.14529/mmp170309.
8. The VINT Project (Virtual InterNetwork Testbed). The Network Simulator — ns-2. <http://www.isi.edu/nsnam/ns>.
9. Wolfram Demonstrations Project. Constrained Random Walk. <http://demonstrations.wolfram.com>.
10. Camp T., Boleng J., Davies V. A survey of mobility models for ad hoc network research // Wirel. Commun. Mob. Com., 2002. No. 2. P. 483–502. doi: 10.1002/wcm.72.
11. Talwar G., Narang H., Pandey K., Singhal P. Analysis of different mobility models for ad hoc on-demand distance vector routing protocol and dynamic source routing protocol. — Lecture notes in electrical engineering ser. — New York, NY, USA: Springer, 2013. Vol. 131. P. 579–588. doi: 10.1007/978-1-4614-6154-8_57.
12. Отт Г. Методы подавления шумов и помех в электронных системах / Пер. с англ. — М.: Мир, 1979. 318 с. (Ott G. Noise reduction techniques in electronic systems. — New York, NY, USA: Wiley, 1976. 312 p.)
13. Гайдамака Ю. В., Андреев С. Д., Сопин Э. С., Самуйлов К. Е., Шоргин С. Я. Анализ характеристик интерференции в модели взаимодействия устройств с учетом среды распространения сигнала // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 2–10. doi: 10.14357/19922264160401.

Поступила в редакцию 11.09.18

MODELING MOVEMENT OF DEVICES IN A WIRELESS NETWORK BY RANDOM WALK MODELS

K. E. Samouylov^{1,2}, Yu. V. Gaidamaka^{1,2}, and S. Ya. Shorgin³

¹Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

²Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The authors overview mobility models which are applicable for simulation of movement of users' devices in a fifth generation (5G) wireless network. Mobility patterns that are typical for simulating the behavior of users of a wireless ad hoc network are considered. The features of the models are discussed, both for individual motion of objects and for motion of groups of objects, from the point of view of appliance to the analysis of interference in wireless networks. The purpose of the paper is to propose a number of mobility models for making an informed decision when choosing a model for evaluating the quality of service in 5G wireless networks. The authors present

simulation results that illustrate the method of estimation of the key performance quality parameter, i. e., signal to interference ratio. For illustration, the developed complex of analytical and simulation models is used for simulation of movement of shopping mall customers with the help of the grid random walk mobility model.

Keywords: entity mobility model; group mobility model; ad hoc network simulation; signal to interference and noise ratio; SINR

DOI: 10.14357/19922264180401

Acknowledgments

The reported study was supported by the Russian Science Foundation, research project No. 16-11-10227.

References

1. Andrews, J. G., S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang. 2014. What will 5G be? *IEEE J. Sel. Area. Comm.* 32(6):1065–1082. doi: 10.1109/JSAC.2014.2328098.
2. Orlov, Yu. N., S. L. Fedorov, A. K. Samuylov, Yu. V. Gaidamaka, and D. A. Molchanov. 2016. Simulation of devices mobility to estimate wireless channel quality metrics in 5G networks. *AIP Conf. Proc.* 1863:090005-1–090005-3. 2017. doi: 10.1063/1.4992270.
3. Gaidamaka, Yu. V., Yu. N. Orlov, D. A. Molchanov, and A. K. Samuylov. 2017. Modelirovanie otnosheniya signal/interferentsiya v mobil'noy seti so sluchaynym bluzhdaniem vzaimodeystvuyushchikh ustroystv [Modeling the signal–interference ratio in a mobile network with moving devices]. *Informatika i ee Primeneniya — Inform. Appl.* 11(2):50–58. doi: 10.14357/19922264170206.
4. Risken, H., and T. Frank. 1996. *The Fokker–Planck equation: Methods of solution and applications*. Springer ser. in synergetics. Berlin–Heidelberg: Springer-Verlag. Vol. 8. 486 p. doi: 10.1007/978-3-642-61544-3.
5. Toffoli, T., and N. Margolus. 1987. *Cellular automata machines*. The MIT Press. 276 p.
6. Grewal, M. S., and A. P. Andrews. 2001. *Kalman filtering: Theory and practice using MATLAB*. 2nd ed. John Wiley & Sons, Inc. 410 p.
7. Semushin, I. V., A. V. Tsyganov, Yu. V. Tsyganova, A. V. Golubkov, and S. D. Vinokurov. 2017. Modelirovanie i otsenivanie traektorii dvizhushchegosya ob"ekta [Modelling and estimation of a moving object trajectory]. *South Ural State University Bulletin. Ser. Mathematical Modelling, Programming & Computer Software* 10(3):108–119. doi: 10.14529/mmp170309.
8. The VINT Project (Virtual InterNetwork Testbed). The Network Simulator ns-2. Available at: <http://www.isi.edu/nsnam/ns/> (accessed September 10, 2018).
9. The Wolfram Demonstrations Project. Constrained Random Walk. Available at: <http://demonstrations.wolfram.com> (accessed September 10, 2018).
10. Camp, T., J. Boleng, and V. Davies. 2002. A survey of mobility models for ad hoc network research. *Wirel. Commun. Com.* (2):483–502.
11. Talwar, G., H. Narang, K. Pandey, and P. Singhal. 2013. Analysis of different mobility models for ad hoc on-demand distance vector routing protocol and dynamic source routing protocol. Lecture notes in electrical engineering ser. New York, NY: Springer. 131:579–588. doi: 10.1007/978-1-4614-6154-8_57.
12. Ott, G. 1976. *Noise reduction techniques in electronic systems*. New York, NY: Wiley. 312 p.
13. Gaidamaka, Yu. V., S. D. Andreev, E. S. Sopin, K. E. Samouylov, and S. Ya. Shorgin. 2016. Analiz kharakteristik interferentsii v modeli vzaimodeystviya ustroystv s uchetom sredy rasprostraneniya signala [Analysis of the characteristics of the interference in the model of interaction between devices taking into account the signal propagation environment]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):2–10. doi:10.14357/19922264160401.

Received September 11, 2018

Contributors

Samouylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Director of Institute of Applied Mathematics and Telecommunications, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation, samuylov_ke@rudn.university

Gaidamaka Yuliya V. (b. 1971) — Doctor of Science in physics and mathematics, professor, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation, gaydamaka_yuv@rudn.university

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sshorgin@ipiran.ru

COMPARISON OF TWO ACTIVE QUEUE MANAGEMENT SCHEMES THROUGH THE $M/D/1/N$ QUEUE

M. G. Konovalov¹ and R. V. Razumchik²

Abstract: The paper focuses on giving the first in the literature numerical evidence that the stationary performance characteristics of single-server queues with the general renovation mechanism may be as good as of single-server queues with the RED-type active queue management mechanisms (AQM). Comparison is made in the queueing theory context: the basic model is the $M/D/1/N$ queue. The characteristics reported are: the loss ratio, average system size, and average number of consecutive losses along with the standard deviations. Numerical results are based on the well-known facts and some new analytic results, presented in the paper.

Keywords: queueing system; active queue management; RED; renovation

DOI: 10.14357/19922264180402

1 Introduction

A large number of AQM mechanisms have been developed up to nowadays and quite a lot of efforts have been devoted to the studies of their efficiency. These mechanisms may be applicable in different contexts but historically, they are more often related to communication networks in the context of mitigation of congestion and congestion avoidance. This problem, as highlighted in the latest RFC 7567 [1], still does not have a satisfying solution. An AQM mechanism is an advanced rejection discipline, when an arriving customer (packet, job, etc.) is lost randomly with a probability that may depend on the (current, past, average, etc.) system state or performance. The most popular class of AQM mechanisms seems to be the Random Early Detection (RED) and its ramifications like GRED (Gentle RED), REM (Random Exponential Marking), etc. (a recent survey on the AQM can be found in [2]). The goals of AQM are usually diverse and conflicting: prevent queues from growing too long, maintain high server (processor) utilization and low variance of the queue size, ensure fairness among competing flows, and others. These are discussed in detail in [1] in the context of communications network but most of the goals are applicable in other contexts as well (buffer-bloat problems in data-center, etc.).

Besides simulation, analytic performance evaluation of systems with AQM is quite often carried out in the queueing theory context (see, for example, [3–8] and

references therein). Usually, the system with an AQM mechanism is modeled as a queueing system or network and then its performance characteristics are studied using known analytic techniques. Throughout the paper, we stay within the queueing theory context.

In the series of recent papers [9–13], the authors have proposed the new type of AQM mechanism which they call *renovation*. Roughly speaking, renovation implies that each customer, having received service, may remove some additional work from the system (i. e., may renovate it). We will make this definition more precise in the next section but for now, note that queue management *after service completions* is what makes the renovation different from the most known AQM schemes³, in which the decisions are made *upon arrivals*. To our best knowledge, there are no studies, which tell whether the performance of the systems with renovation is better/same/worse than that of the same systems but with the implemented AQM mechanisms. Thus, there is a lack of bridge between available theoretical results for renovation and its practical perspective.

The scope of this paper is to give the first in the literature numerical evidence that the stationary performance of single-server queueing systems with the implemented renovation mechanism can be as good as of the same single-server queues but the well-known packed dropping procedures like RED. The emphasis is primarily on the reporting of this finding, complemented with some

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44/2 Vavilov Str., Moscow 119333, Russian Federation, mkonovalov@ipiran.ru

²Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44/2 Vavilov Str., Moscow 119333, Russian Federation; Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; rrazumchik@ipiran.ru

³Indeed, renovation and most of the known AQM mechanisms are conceptually different. One of the main goals of AQM mechanisms is to prevent queue from growing too large leaving space for potential new arrivals. In systems with renovation, the queue can become full (meaning that fewer customers are lost) but after a service completion, several customers may be removed from it. In this way, the content of the queue can be preserved at a certain average level but the loss pattern becomes intricate.

new insights into queueing systems with renovation. The relation to other AQM mechanisms like CoDel [14] is not discussed here. Moreover, in the numerical experiments presented here, we did not use any benchmarks to generate the traffic profiles but used the theoretical distributions instead.

The main stationary performance characteristics reported are: the loss ratio, the average number in the system (average system size), and the average number of consecutive losses along with their standard deviations. After introducing the renovation mechanism and the analytic setting, in which renovation mechanism is compared with RED, we give the new analytic results for computing system size moments and the loss ratio under the renovation mechanism. The results presented in the numerical section are based on the analytic results. Monte-Carlo simulation is used only for the average (and standard deviation) number of consecutive losses in the system with renovation.

2 Settings and the Model

We follow the queueing theoretic approach and as the basic model, we use $M/D/1/N$ queue, i. e., queue of finite capacity N fed at rate λ by a Poisson flow of customers, which are served on a first-come-first-served basis by a single server with constant service time $d > 0$. We assume that the system is in the steady state. When an arriving customer sees that the queue is full, it is lost. If no other type of losses occur in the system, we say that the Tail Drop mechanism is implemented in it.

If an arriving customer is lost with probability d_n where n is the total number of customers it sees in the system on arrival, then we say that an AQM mechanism is implemented in the system. Various dropping functions can be obtained by specifying the values of d_n (see, for example, RED dropping function in [5, Example 1]). Important notice should be made here. In practice, RED-type mechanisms may use moving averages of the queue size instead of its instantaneous value. Thus, the way d_n introduced above is a simplified way of thinking. Yet, this trade-off is important because it allows to keep the mathematical models of RED-type AQM tractable. Luckily, as noticed in [3, Section II.C], such approximation may not lead to significant bias, when the weight of the moving average scheme is small (which is claimed to be the case sometimes in practice).

The renovation mechanism, which is implemented in a system with Tail Drop, works as follows. Define $N + 1$ numbers, say, $q_i \geq 0$, $0 \leq i \leq N$, satisfying $\sum_{i=0}^N q_i = 1$. If upon service completion there are i , $1 \leq i \leq N$, customers waiting in the queue, then the served customer leaves the system and

- with probability $q_0 + Q_i$ nothing else happens, where $Q_i = q_i + q_{i+1} + \dots + q_N$; and
- with probability q_j , $0 < j < i$, exactly j customers from the queue leave the system and those customers are chosen successively *starting from the head of the queue*.

The served customer, which sees the empty queue, leaves the system. Thus, after the renovation (if it happened), the system never becomes empty.

In the numerical section, we rank the systems with RED and renovation according to the stationary loss ratio, average system size, and average number of consecutive losses along with their standard deviations. The system with the Tail Drop is the standard $M/D/1/N$ queue, for which all these performance characteristics follow from the classical results in queueing theory (see, for example, [15]). Analytic results for the systems of $M/G/1/N$ type with relatively arbitrary dropping functions are given in [5]. Yet, for the system with renovation, we need to derive these performance characteristics anew, since the available results in [11, 12] are not valid for the renovation mechanism introduced above. We briefly sketch the derivations in the next section and omit most of the details since they are based on the methodology, developed in [11, 12], and reviewed in [16].

3 Performance Characteristics

Consider the $M/D/1/N$ queue with the renovation mechanism introduced above. Since a customer is served for constant service time d , then for the cumulative distribution function $B(x)$ of its service time, one has:

$$B(x) = \begin{cases} 0 & \text{if } x \leq d; \\ 1 & \text{if } x > d. \end{cases}$$

Let $N(t)$ be the total number of customers in the system at instant t and $E(t)$ be the elapsed service time of the customer in server (if there is one). In order to compute the stationary system size moments, one needs to know the stationary distribution:

$$P_n = \lim_{t \rightarrow \infty} \mathbf{P}\{N(t) = n\}, \quad 0 \leq n \leq N + 1.$$

For the computation of the loss ratio, due to the PASTA (Poisson Arrivals See Time Averages) property, it is sufficient to know the stationary probability densities $p_n(x) = P'_n(x)$ where

$$P_n(x) = \lim_{t \rightarrow \infty} \mathbf{P}\{N(t) = n, E(t) < x\}, \quad 1 \leq n \leq N, x \in [0, d].$$

Since we are dealing with the finite-capacity queue and work conserving service discipline, the introduced

stationary distributions exist. The probabilities P_n and the densities $p_n(x)$ can be computed as follows. Let t_n denotes the n th service completion epoch and $N_n = N(t_n + 0)$ denotes the total number of customers in the system. Clearly, $\{N_n, n \geq 1\}$ is the finite-state Markov chain. The entries of the transition probability matrix $\mathbb{P} = (p_{ij})$ of this chain have the form:

$$p_{0j} = p_{1j} = \begin{cases} \beta_0, & j = 0; \\ \beta_j Q_j + \sum_{k=j}^N \beta_k q_{k-j} + B_N q_{N-j}, & 1 \leq j \leq N-1; \\ (q_0 + q_N) B_{N-1}, & j = N; \end{cases}$$

$$p_{ij} = \begin{cases} 0, & j = 0; \\ \sum_{k=j-1}^{N-1} \beta_k q_{k-j+1} + B_{N-1} q_{N-j}, & 1 \leq j \leq i-2; \\ \beta_{j-i+1} Q_j + \sum_{k=j-1}^{N-1} \beta_k q_{k-j+1} + B_{N-1} q_{N-j}, & i-1 \leq j \leq N-1; \\ (q_0 + q_N) B_{N-i}, & j = N, \end{cases} \quad 2 \leq i \leq N.$$

Here, $B_0 = 1 - \beta_0$; $B_k = B_{k-1} - \beta_k$; and $\beta_k = [(\lambda d)^k / k!] e^{-\lambda d}$. The matrix \mathbb{P} does not have any special structure. So, the stationary distribution $\{P_n^+, 0 \leq n \leq N\}$ may be found in a straightforward manner by solving the system of linear algebraic equations

$$\vec{P}^+ = \vec{P}^+ \mathbb{P}; \quad \vec{P}^+ \vec{1} = 1$$

where $\vec{P}^+ = (P_0^+, \dots, P_N^+)$ and $\vec{1}$ is the vector of ones.

Once the probabilities P_n^+ are found, the stationary distribution $\{P_n, 0 \leq n \leq N + 1\}$ is computed from the relation¹

$$P_n = \sum_{i=0}^N P_i^+ \frac{f_{in}}{f^*}$$

where f_{in} is the average time during which there were n customers in the system provided that the system started with i customers in it; and f^* is the mean time between transitions of the embedded Markov chain $\{N_n, n \geq 1\}$.

Finally, the stationary probability densities $p_n(x) = P_n'(x)$ can be computed using the fact that the relation for $p_n(x)$ coincides with the relation for $p_n(x)$ in the standard $M/D/1/N$ queue. Thus, $p_n(x)$ are given by (see, for example, [15, p. 72])

$$p_n(x) = e^{-\lambda x} (1 - B(x)) \sum_{k=0}^{n-1} p_{n-k}(0) \frac{(\lambda x)^k}{k!}, \quad 1 \leq n \leq N, x \in [0, d]. \quad (1)$$

Even though (1) holds, due to the presence of renovation, the boundary conditions $p_n(0)$ for the considered queue do not coincide with boundary conditions $p_n(0)$ for the standard $M/D/1/N$ queue. By integration (1) from 0 to d , one gets the following relation between $p_n(0)$ and $P_n = \int_0^d p_n(x) dx$:

$$p_n(0) = \frac{1}{B_0} \left(\lambda P_n - \sum_{k=1}^{n-1} B_k p_{n-k}(0) \right), \quad 1 \leq n \leq N. \quad (2)$$

Since the stationary distribution $\{P_n, 0 \leq n \leq N + 1\}$ is already known, the values of $p_n(0)$ are computed recursively from (2). The closed-form expressions for the average and the standard deviation of the system size are, in the most cases, not available and thus, they can be computed, respectively, by $\sum_{n=0}^{N+1} n P_n$ and $\sqrt{\sum_{n=0}^{N+1} n^2 P_n - (\sum_{n=0}^{N+1} n P_n)^2}$.

The computation of the loss ratio π , i. e., the probability that the arriving customer is lost, is more involved. This is due to the fact that the accepted customer may be lost either after the first service completion or the second, etc. and the chance to be lost varies, depending on the number of new customers that have arrived between successive service completions.

Let us introduce two quantities:

- (1) γ_{ij} , $1 \leq i \leq N, j \geq 0$, — probability that the arriving customer finds i customers in the system and until the next service completion, exactly j new customers arrive at the system; and
- (2) r_{ij} , $0 \leq j \leq N - 1, 0 \leq i \leq N - j - 1$, — probability that the tagged customer waiting in the queue *will not* be served (i. e., will be lost), if there are j customers in front of it in the queue (excluding the one in server) and i behind.

Given that γ_{ij} and r_{ij} are known, the loss ratio π can be computed as

$$\pi = P_{N+1} + \sum_{i=1}^N \left[\sum_{j=0}^{N-i} \gamma_{ij} \left(\sum_{k=0}^{i-2} q_k r_{j,i-2-k} + \sum_{k=i}^{i+j-1} q_k + Q_{j+i} r_{j,i-2} \right) \right] +$$

¹This follows from the well-known results for the Markov regenerative processes (see, for example, [17, Theorem 9.19]).

$$+ \sum_{j=N-i+1}^{\infty} \gamma_{ij} \left(\sum_{k=0}^{i-2} q_k r_{N-i, i-2-k} + \sum_{k=i}^{N-1} q_k + Q_N r_{N-i, i-2} \right).$$

Due to the PASTA property of Poisson arrivals, the expression for γ_{ij} follows from the law of total probability:

$$\gamma_{ij} = \int_0^d p_i(x) \frac{(\lambda(d-x))^j}{j!} e^{-\lambda(d-x)} dx, \quad 1 \leq i \leq N, j \geq 0.$$

Again, by applying the law of total probability, one gets the relations for the recursive computation of r_{ij} , $0 \leq j \leq N-1, 0 \leq i \leq N-j-1$:

$$\begin{aligned} r_{i0} &= \sum_{m=0}^{N-i-1} \beta_m \sum_{k=1}^{m+i} q_k + B_{N-i-1} \sum_{k=1}^{N-1} q_k; \\ r_{ij} &= \sum_{m=i}^{N-1-j} \beta_{m-i} \left(\sum_{k=0}^{j-1} q_k r_{m, j-1-k} + \sum_{k=j+1}^{m+j} q_k + Q_{j+m+1} r_{m, j-1} \right) + \\ &+ B_{N-j-i-1} \left(\sum_{k=0}^{j-1} q_k r_{N-j-1, j-1-k} + \sum_{k=j+1}^{N-1} q_k + Q_N r_{N-j-1, j-1} \right). \end{aligned}$$

The expressions above can be further simplified¹ by computing the integrals explicitly, but we do not dwell on it here. For small and moderate values of d, N , and λ , they can be directly used for numerical implementation. In the numerical section, precisely these expressions are used to calculate the loss ratio. The expressions for the average and the standard deviation of consecutive losses are much harder to derive and we leave it for a separate study. The values of these two parameters were taken from the Monte-Carlo simulation.

¹There are no principal difficulties in generalizing the results for the BMAP/G/1/N queue. Yet, this would obscure the goal of the paper and thus, we remain with the simple model.

²For the sake of reproducibility of the results presented in the paper, we also report the obtained values of the renovation probabilities: for $\rho = 0.5$, $\vec{q} = (0.2544, 0.0037, 0.0053, 0.0065, 0.0122, 0.0352, 0.1108, 0.1898, 0.2129, 0.1691)$; for $\rho = 1$, $q_0 = 0.0551, q_6 = 0.051, q_7 = 0.7166, q_8 = 0.0917$, and $q_9 = 0.0856$; for $\rho = 2$, $q_0 = 0.1078, q_1 = 0.6374, q_4 = 0.0042, q_6 = 0.0084$, and $q_9 = 0.2422$; and for $\rho = 3$, $q_1 = 0.4608$, and $q_2 = 0.5392$.

4 Numerical Example

As the reference point, we have chosen the numerical results in [5] which are based on the analytic expressions and which show the performance characteristics of the $M/D/1/N$ queue with four different AQM mechanisms. Since RED scheme is one of the best among the four, our goal here is to rank the $M/D/1/N$ queue with RED from [5] and the $M/D/1/N$ queue with renovation. Comparison is made according to the stationary loss ratio, average system size, and average number of consecutive losses along with their standard deviations.

The initial conditions are: the maximum queue size is $N = 9$ and the service time is $d = 1$. Thus, the offered load is $\rho = \lambda d$. The RED dropping function is given by (see [5, Eq. (59)]):

$$d_n = \begin{cases} 0, & n \leq 3; \\ 0.11917n - 0.35752, & 4 \leq n \leq 9; \\ 1, & n = 10. \end{cases} \quad (3)$$

The performance of the $M/D/1/N$ queue with this RED dropping function is given in [5, Tables 1, 3, and 4]. In order to find out whether there exists a renovation mechanism under which the $M/D/1/N$ queue can perform at least as good as under RED, one needs to perform exhaustive search over the possible values of the renovation parameters $\{q_i, 0 \leq i \leq N\}$. Since we are unaware of any analytic way of choosing these values, adaptive search algorithms for partially observable Markov decision processes from [18] were used instead. Metaheuristics (like particle swarm optimization), which are also applicable here, were not used.

In Tables 1 and 2, one can see the numerical results for the four different cases of the offered load² ρ : $\rho = 0.5$ — underloaded system; $\rho = 1$ — critically loaded system; and $\rho = 2$ and 3 — overloaded system. The values displayed are the values from the numerical experiments rounded to three decimal digits.

Data in the tables show that with respect to the loss ratio, renovation can perform as good as RED in the wide range of the offered load ρ . The only exception is the case $\rho = 1$: here, renovation can keep only the average system size at the same level as RED; other four performance characteristics are worse.

If we fix the loss ratio, then the renovation mechanism can guarantee at least the same value of the average system size as guaranteed by RED. It is worth noticing that as the offered load increases, the average system size under the renovation mechanism becomes smaller

Table 1 Performance characteristics of the $M/D/1/9$ system with the RED mechanism (3) and the $M/D/1/9$ the renovation mechanism (ren.) under the offered load $\rho = 0.5$ and $\rho = 1$

		$\rho = 0.5$			$\rho = 1$		
		Tail Drop	RED	ren.	Tail Drop	RED	ren.
loss ratio		0	0.002	0.002	0.051	0.091	0.104
system size	average	0.750	0.741	0.744	5.064	3.000	2.999
	standard deviation	0.946	0.920	0.935	2.897	1.887	2.091
consecutive losses	average	1.152	1.053	1.800	1.359	1.239	6.876
	standard deviation	0.403	0.240	1.260	0.647	0.561	0.852

Table 2 Performance characteristics of the $M/D/1/9$ system with the RED mechanism (3) and the $M/D/1/9$ the renovation mechanism (ren.) under the offered load $\rho = 2$ and $\rho = 3$

		$\rho = 2$			$\rho = 3$		
		Tail Drop	RED	ren.	Tail Drop	RED	ren.
loss ratio		0.500	0.500	0.502	0.667	0.667	0.667
system size	average	9.372	7.146	7.142	9.646	8.390	7.114
	standard deviation	0.744	1.436	2.387	0.523	1.090	2.246
consecutive losses	average	1.884	1.996	1.592	2.542	2.876	2.141
	standard deviation	1.069	1.366	1.100	1.454	2.064	1.236

than the average system size under RED. Yet, renovation keeps the queue less stable than RED: the standard deviation of system size is always smaller for RED.

If we fix the loss ratio and the average system size, then the renovation mechanism spreads out the losses worse than RED when the system is underloaded and better than RED when it is overloaded. This can be seen from the values of the averages and standard deviations in the last two rows of Tables 1 and 2.

5 Concluding Remarks

Even though the idea behind the renovation-type AQMs is completely different from the idea behind RED-type AQMs, renovation-type AQMs may allow one to achieve in some cases at least the same system performance level as guaranteed by RED-type AQMs. Although the comparison presented here is based only on a single RED dropping function (3), our numerical experiments show that the results remain qualitatively the same for RED-type AQMs with other dropping functions. Being defined by N parameters, the renovation mechanism is very flexible and this constitutes its strength and weakness. By varying the values of the renovation probabilities q_i , it is possible to carry out conditional optimization, but good searching procedures are required here.

Implementation of the renovation as a packet dropping mechanism requires *a priori* tuning and/or operational configuration of its parameters. Thus, whether it is appropriate to use renovation as a packet dropping mechanism or not in practice heavily depends on the

use case. Although the tuning of the renovation parameters q_i can be made on the fly during operation, with respect to the recommendations of the RFC 7567 [1], renovation mechanism is not the proper choice for the network congestion control unless simple recommendations on how to set up the renovation parameters are given. We believe this can be done based on more deep and insightful numerical experiments.

There remain a large number of unresolved issues related to the renovation mechanism (e.g., can renovation ensure fairness among competing flows? may the average queue size instead of its instantaneous value increase the efficiency of the renovation mechanism?) and this motivates its further analysis. Furthermore, evaluation of the renovation mechanism with parameters adapted to a realistic router/switch use case and/or evaluation which includes TCP feedback loops of several flows remains an open issue as well.

Acknowledgments

The reported study was partially funded by the Russian Foundation for Basic Research according to the research project No. 18-07-00692.

The authors would like to thank the anonymous referees for their valuable comments which helped to improve the paper.

References

1. Baker, F., and G. Fairhurst. 2015. IETF recommendations regarding active queue management. Available at:

- <https://tools.ietf.org/html/7567> (accessed October 4, 2018).
- Adams, R. 2013. Active queue management: A survey. *IEEE Commun. Surv. Tut.* 15(3):1425–1476.
 - Bonald, T., M. May, and J. C. Bolot. 2000. Analytic evaluation of RED performance. *IEEE Conference on Computer Communications Proceedings* 3:1415–1424.
 - Hao, W., and Y. Wei. 2005. An extended $GI^X/M/1/N$ queueing model for evaluating the performance of AQM algorithms with aggregate traffic. *Networking and mobile computing*. Eds. Xicheng Lu and Wei Zhao. Lecture notes in computer science ser. Springer. 3619:395–404.
 - Chydziański, A., and L. Chróst. 2011. Analysis of AQM queues with queue size based packet dropping. *Int. J. Appl. Math. Comp.* 21(3):567–577.
 - Chydziański, A., and P. Mrozowski. 2016. Queues with dropping functions and general arrival processes. *PLoS ONE* 11(3):e0150702. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150702> (accessed October 4, 2018).
 - Tikhonenko, O., and W. Kempa. 2016. Performance evaluation of an $M/G/n$ -type queue with bounded capacity and packet dropping. *Int. J. Appl. Math. Comp.* 26(4):841–854.
 - Konovarov, M. G., and R. V. Razumchik. 2018. Numerical analysis of improved access restriction algorithms in a $GI/G/1/N$ system. *J. Commun. Technol. El.* 63(6):616–625.
 - Kreinin, A. Y. 1997. Queueing systems with renovation. *J. Appl. Math. Stochastic Analysis* 10(4):431–441.
 - Zaryadov, I. S. 2009. Queueing systems with general renovation. *Conference (International) on Ultra Modern Telecommunications Proceedings*. 1–4.
 - Zaryadov, I. S., and A. V. Pechinkin. 2009. Stationary time characteristics of the $GI/M/n/\infty$ system with some variants of the generalized renovation discipline. *Automat. Rem. Contr.* 70(12):2085–2097.
 - Zaryadov, I. S. 2010. The $GI/M/n/\infty$ queueing system with generalized renovation. *Automat. Rem. Contr.* 71(4):663–671.
 - Korolkova, A., and I. Zaryadov. 2010. The mathematical model of the traffic transfer process with a rate adjustable by RED. *Conference (International) on Ultra Modern Telecommunications Proceedings*. 1046–1050.
 - Nichols, K., V. Jacobson, A. McGregor, and J. Iyengar. 2018. Controlled delay active queue management. Available at: <https://datatracker.ietf.org/doc/rfc8289> (accessed October 4, 2018).
 - Riordan, J. 1962. *Stochastic service systems*. SIAM ser. in applied mathematics. New York, NY: Wiley. 139 p.
 - Konovarov, M., and R. Razumchik. 2017. Queueing systems with renovation vs. queues with RED. Supplementary material. *ArXiv e-prints*. Available at: <https://arxiv.org/abs/1709.01477/> (accessed October 4, 2018).
 - Kulkarni, V. G. 2016. *Modeling and analysis of stochastic systems*. 3rd ed. Chapman & Hall/CRC texts in statistical science ser. Chapman & Hall/CRC. 606 p.
 - Konovarov, M. G. 2007. *Metody adaptivnoy obrabotki informatsii i ikh prilozheniya* [Methods of adaptive information processing and their applications]. Moscow: Institute of Informatics Problems of RAS. 212 p.

Received October 9, 2018

Contributors

Konovarov Mikhail G. (b. 1950) — Doctor of Science in technology, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mkonovarov@ipiran.ru

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; rrazumchik@ipiran.ru

СРАВНЕНИЕ ДВУХ МЕХАНИЗМОВ АКТИВНОГО УПРАВЛЕНИЯ ОЧЕРЕДЬЮ В СИСТЕМЕ $M/D/1/N^*$

М. Г. Коновалов¹, Р. В. Разумчик^{1,2}

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

²Российский университет дружбы народов

Аннотация: Представлены некоторые результаты численных экспериментов, подтверждающие следующее обстоятельство: параметры механизма обобщенного обновления могут быть подобраны таким образом,

*Исследование выполнено при частичной финансовой поддержке РФФИ (проект 18-07-00692).

что уровень производительности однолинейных систем массового обслуживания с обобщенным обновлением может быть не ниже уровня производительности систем с RED-подобными механизмами активного управления очередями. Механизмы сравниваются на примере системы $M/D/1/N$ по стационарным значениям следующих характеристик: вероятность потери заявки, среднее число заявок в системе, среднее число последовательных потерь заявок и их средние квадратические отклонения. Расчеты основаны на известных фактах, а также на ряде новых аналитических результатов для систем с обобщенным обновлением, полученных в данной работе.

Ключевые слова: система массового обслуживания; алгоритмы активного управления очередями; обобщенное обновление

DOI: 10.14357/19922264180402

Литература

1. Baker F., Fairhurst G. IETF recommendations regarding active queue management, 2015. <https://tools.ietf.org/html/7567>.
2. Adams R. Active queue management: A survey // IEEE Commun. Surv. Tut., 2013. Vol. 15. No. 3. P. 1425–1476.
3. Bonald T., May M., Bolot J. C. Analytic evaluation of RED performance // IEEE Conference on Computer Communications Proceedings, 2000. Vol. 3. P. 1415–1424.
4. Hao W., Wei Y. An extended $GI^X/M/1/N$ queueing model for evaluating the performance of AQM algorithms with aggregate traffic // Networking and mobile computing / Eds. Xicheng Lu and Wei Zhao. — Lecture notes in computer science ser. — Springer, 2005. Vol. 3619. P. 395–404.
5. Chydzński A., Chróst L. Analysis of AQM queues with queue size based packet dropping // Int. J. Appl. Math. Comp., 2011. Vol. 21. No. 3. P. 567–577.
6. Chydzński A., Mrozowski P. Queues with dropping functions and general arrival processes // PLoS ONE, 2016. Vol. 11. No. 3. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150702>.
7. Tikhonenko O., Kempa W. Performance evaluation of an $M/G/n$ -type queue with bounded capacity and packet dropping // Int. J. Appl. Math. Comp., 2016. Vol. 26. No. 4. P. 841–854.
8. Konovalov M. G., Razumchik R. V. Numerical analysis of improved access restriction algorithms in a $GI/G/1/N$ system // J. Commun. Technol. El., 2018. Vol. 63. No. 6. P. 616–625.
9. Kreinin A. Y. Queueing systems with renovation // J. Appl. Math. Stochastic Analysis, 1997. Vol. 10. No. 4. P. 431–441.
10. Zaryadov I. S. Queueing systems with general renovation // Conference (International) on Ultra Modern Telecommunications Proceedings, 2009. P. 1–4.
11. Зарядов И. С., Печинкин А. В. Стационарные временные характеристики системы $GI/M/n/\infty$ с некоторыми вариантами дисциплины обобщенного обновления // Автоматика и телемеханика, 2009. Вып. 12. С. 161–174.
12. Зарядов И. С. Система массового обслуживания $GI/M/n/\infty$ с обобщенным обновлением // Автоматика и телемеханика, 2010. Вып. 4. С. 130–139.
13. Korolkova A., Zaryadov I. The mathematical model of the traffic transfer process with a rate adjustable by RED // Conference (International) on Ultra Modern Telecommunications Proceedings, 2010. P. 1046–1050.
14. Nichols K., Jacobson V., McGregor A., Iyengar J. Controlled delay active queue management, 2018. <https://datatracker.ietf.org/doc/rfc8289>.
15. Riordan J. Stochastic service systems. — SIAM ser. in applied mathematics. — New York, NY, USA: Wiley, 1962. 139 p.
16. Konovalov M., Razumchik R. Queueing systems with renovation vs. queues with RED. Supplementary material // ArXiv e-prints, 2017. <https://arxiv.org/abs/1709.01477/>.
17. Kulkarni V. G. Modeling and analysis of stochastic systems. 3rd ed. — Chapman & Hall/CRC texts in statistical science ser. — Chapman & Hall/CRC, 2016. 606 p.
18. Коновалов М. Г. Методы адаптивной обработки информации и их приложения. — М.: ИПИ РАН, 2007. 212 с.

Поступила в редакцию 09.10.2018

ОПРЕДЕЛЕНИЕ ЭКСТРЕМАЛЬНОСТИ ОБЪЕМОВ ОСАДКОВ НА ОСНОВЕ МОДИФИЦИРОВАННОГО МЕТОДА ПРЕВЫШЕНИЯ ПОРОГОВОГО ЗНАЧЕНИЯ*

А. К. Горшенин¹, В. Ю. Королев²

Аннотация: Задача корректного определения того, какие наблюдения следует признавать экстремальными, чрезвычайно важна при изучении метеорологических явлений. Предложены восходящий и нисходящий методы определения порогового (экстремального) уровня на основе использования теорем Реньи для редящихся потоков и результатов Пикандса – Балкемы – Де Хаана. На примере данных за 60 лет наблюдений в Потсдаме и Элисте продемонстрировано, что восходящий метод показывает отличные результаты для суточных объемов осадков, однако для дождливых периодов необходимо использовать нисходящий метод. Приведено сравнение результатов подобного непараметрического подхода с параметрическим критерием, который был предложен авторами в предшествующих работах.

Ключевые слова: осадки; дождливые периоды; экстремальные наблюдения; пороговые значения; теорема Реньи; теорема Пикандса – Балкемы – Де Хаана; проверка статистических гипотез; анализ данных

DOI: 10.14357/19922264180403

1 Введение

Оценки закономерностей и тенденций в наблюдениях аномально экстремальных метеорологических явлений важны для понимания процесса изменения климата. Однако известно, что различные методы оценки экстремальных осадков в применении к разным моделям дают существенно отличающиеся результаты. Чаще всего решение основано на выборе порогового значения, определяемого как квантиль некоторого распределения [1]. Однако изменения в максимальных значениях объемов осадков (и, соответственно, большее число превышений порога) не всегда ведут к качественному изменению того, какие именно события действительно должны быть признаны экстремальными. В частности, увеличение доли осадков подобного рода может быть связано с изменениями суммарного объема или с увеличением интенсивности осадков в сочетании с уменьшением числа дождливых дней [2].

Для решения задачи определения аномальных наблюдений в климатологических задачах часто используется подход теории экстремальных значений, называемый *Peaks over Threshold* (PoT; пики, превышающие порог) [3]. В частности, известно использование подобного подхода в задачах,

связанных со штормовыми волнами [4], дневными температурами [5], осадками [6–8]. Для выявления критических/экстремальных наблюдений в информационных потоках различной природы авторами данной статьи ранее была предложена методология поиска порогового значения на основе модификации PoT-метода с использованием результатов теоремы Реньи для редящихся потоков [9]. В настоящей работе подобный подход будет развит для определения аномально экстремальных значений объемов осадков, в том числе выпавших в течение так называемого дождливого периода (подряд идущих дней, в которые наблюдались осадки). Кроме того, приведено сравнение результатов указанного непараметрического подхода с параметрическим критерием, который был ранее предложен авторами в статье [10] для решения подобного класса задач.

2 Использование PoT-методологии для суточных объемов осадков

Воспользуемся подходом на основе двух фундаментальных результатов — теоремы Реньи для

*Работа выполнена при поддержке РФФИ (проект 17-07-00851) и Стипендии Президента Российской Федерации молодым ученым и аспирантам (СП-538.2018.5).

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, agorshenin@frcsc.ru

²Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vkorolev@cs.msu.ru

редующих потоков [11] и классического результата теории экстремальных значений, связанного с именами Пикандса, Балкемы и Де Хаана [12, 13], которые позволяют избегать априорных предположений о данных [14]. Из указанных выше теорем следует, что распределение разностей моментов превышения порогового значения должно соответствовать экспоненциальному закону, а величины превышений данного порога — обобщенному распределению Парето, которое имеет следующий вид ($\xi \in \mathbb{R}$ — параметр формы, $\mu \in \mathbb{R}$ — сдвига, $\sigma > 0$ — масштаб):

$$F_{\xi,\sigma,\mu}(y) = \begin{cases} 1 - \left(1 + \frac{\xi(y - \mu)}{\sigma}\right)^{-1/\xi}, & \text{если } \xi \neq 0; \\ 1 - e^{-(y-\mu)/\sigma} & \text{иначе.} \end{cases}$$

Таким образом, пороговое значение может быть определено в рамках статистической процедуры, в которой для каждого уровня, начиная с некоторого значения, например минимума в данных, с заранее заданным шагом должны проверяться последовательно две гипотезы об экспоненциальности и паретовости описанных выше объектов. В случае

принятия обеих текущий уровень может считаться экстремальным пороговым значением. Назовем данный метод, следуя статьям [9, 14], *восходящим* (в соответствии с направлением сдвига порогового значения в процессе анализа данных).

Продемонстрируем описанный подход на примере наблюдений за объемом дневных осадков за почти 60-летний период (1950–2008 гг.) в городах Потсдаме и Элисте. Процедура поиска автоматизирована с помощью программного решения, созданного на встроенном языке программирования пакета MATLAB. Получено, что для Потсдама критический уровень составляет 30,2 мм (шаг изменения уровня — 0,01 мм, уровень значимости статистического критерия при проверке по χ^2 -тесту выбран равным 0,01). Для экспоненциального распределения $p_{\text{знач}} = 0,07$ (параметр оценивается значением 0,002), для обобщенного Парето $p_{\text{знач}} = 0,29$ (параметры: 0,226, 9,196 и 30,2).

На рис. 1 продемонстрировано визуальное соответствие между экспериментальными данными и подогаанными распределениями (см. гистограммы и аппроксимирующие кривые на рис. 1, а и 1, б).

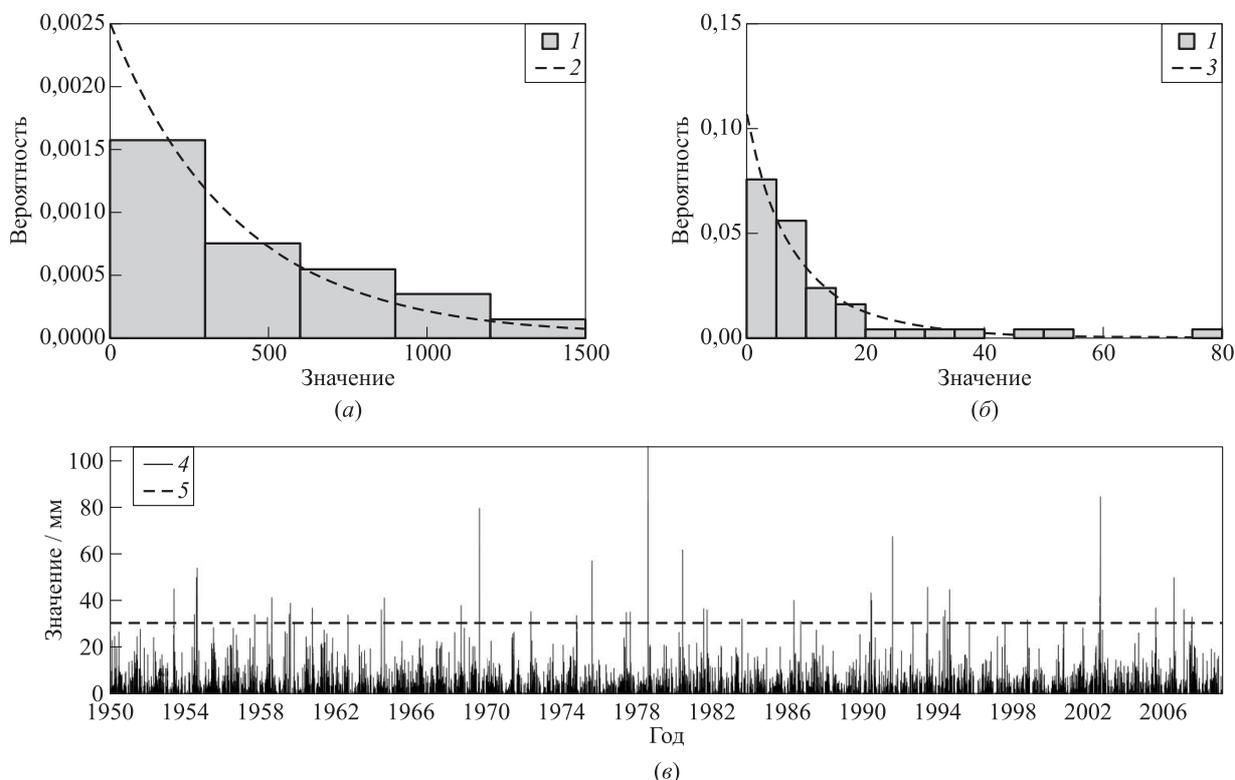


Рис. 1 Пороговый уровень, полученный для дневных объемов восходящим методом, Потсдам: (а) распределение интервалов между превышениями порога; (б) распределение превышений порога; (в) суточные объемы; 1 — гистограмма; 2 — экспоненциальное распределение с параметром 0,002; 3 — обобщенное распределение Парето с параметрами 0,226, 9,196 и 30,2; 4 — данные; 5 — порог 30,2 мм

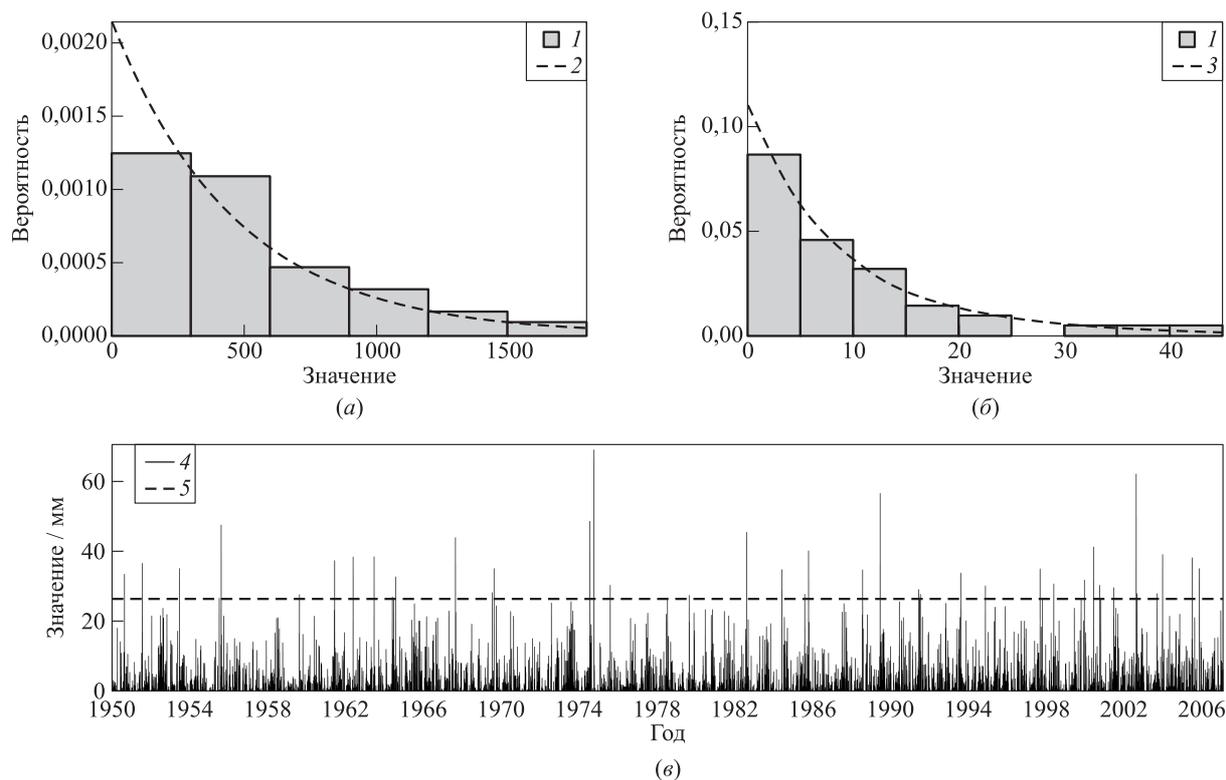


Рис. 2 Пороговый уровень, полученный для дневных объемов восходящим методом, Элиста: (а) распределение интервалов между превышениями порога; (б) распределение превышений порога; (в) суточные объемы; 1 — гистограмма; 2 — экспоненциальное распределение с параметром 0,002; 3 — обобщенное распределение Парето с параметрами 0,038, 9,009 и 26,5; 4 — данные; 5 — порог 26,5 мм

На рис. 1, в продемонстрировано, что полученное пороговое значение превышает относительно небольшое число раз за весь период наблюдений; таким образом, данные пики могут рассматриваться как «подозрительные на экстремальность».

Для Элисты критический уровень составляет 26,5 мм. Для экспоненциального распределения $p_{\text{знач}} = 0,84$ (параметр оценивается значением 0,002), для обобщенного Парето $p_{\text{знач}} = 0,6$ (параметры: 0,038, 9,009 и 36,5). На рис. 2 отображено соответствие между экспериментальными данными и подогнанными распределениями (рис. 2, а и 2, б). На рис. 2, в продемонстрировано, что полученное пороговое значение снова превышает относительно небольшое число раз за весь период наблюдений.

3 Использование RoT-методологии для объемов осадков за дождливые периоды

Отметим, что описанный в разд. 2 восходящий метод может быть реализован и в обратном направ-

лении (с точки зрения смещения уровня в процессе анализа данных). Для этого необходимо задать верхнюю границу (например, совпадающую с максимумом наблюдений) и последовательно проверять гипотезы об экспоненциальности и паретовости.

В начале работы метода число превышений порога будет недостаточным для построения гистограмм и корректного оценивания параметров, поэтому необходимо обеспечить минимально приемлемый объем соответствующей выборки.

В работе [9] описана схожая процедура, однако в качестве примера рассмотрены кумулятивные данные. В данном разделе откажемся от ограничения на тип рассматриваемых данных и будем считать *нисходящим* методом описанный выше алгоритм.

В качестве примера рассмотрим суммарные объемы осадков, зарегистрированные в течение дождливого периода, однако данный метод может быть использован и для исходных суточных наблюдений.

Для объемов осадков за дождливые периоды в Потсдаме верхний критический уровень (полу-

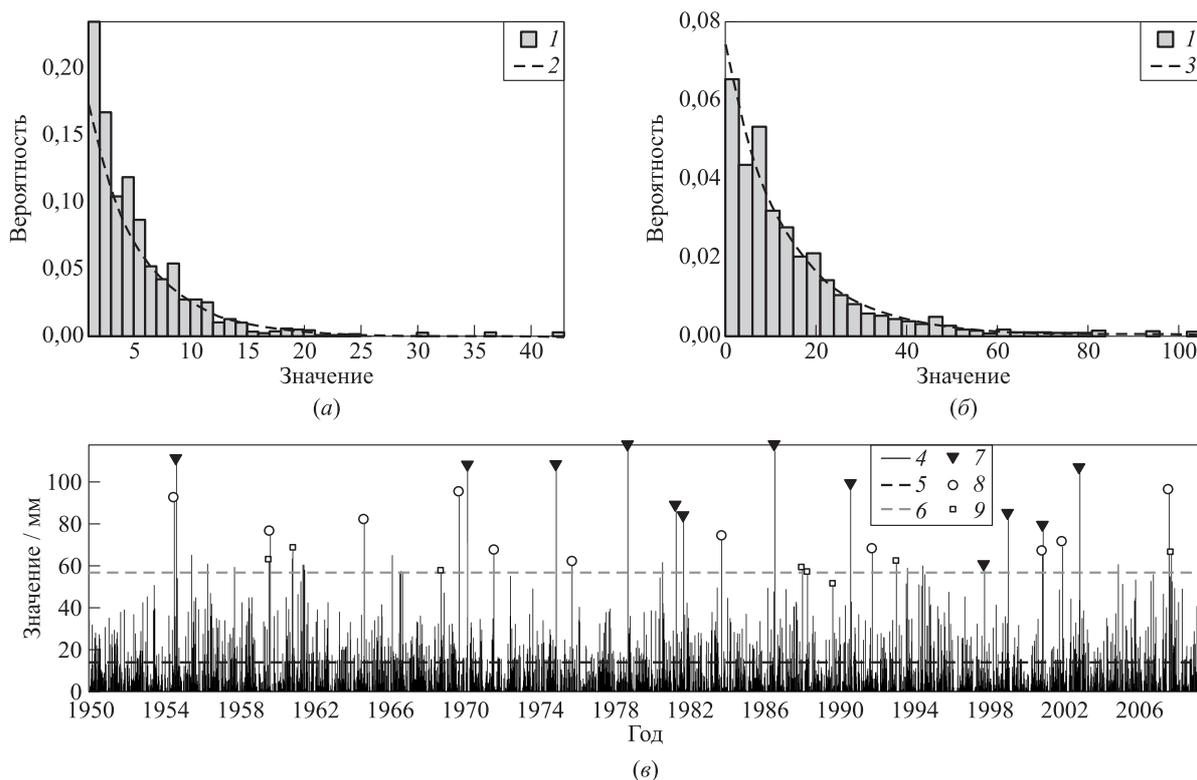


Рис. 3 Нисходящий, восходящий и параметрический методы определения экстремальности: (а) распределение интервалов между превышениями порога; (б) распределение превышений порога; (в) объемы осадков за дождливые периоды, Потсдам; 1 — гистограмма; 2 — экспоненциальное распределение с параметром 0,216; 3 — обобщенное распределение Парето с параметрами 0,076, 13,406 и 14,41; 4 — данные; 5 — порог 14,41 мм; 6 — порог 57,2 мм; 7 — абсолютно экстремальные значения; 8 — промежуточные; 9 — относительно экстремальные значения

чен нисходящим методом) составляет 57,2 мм (шаг изменения уровня — 0,01 мм, уровень значимости статистического критерия при проверке по χ^2 -тесту выбран равным 0,01). Для экспоненциального распределения $p_{\text{знач}} = 0,1$ (параметр оценивается значением 0,014), для обобщенного Парето $p_{\text{знач}} = 0,03$ (параметры: 0,097, 16,95 и 57,2). Нижний критический уровень (получен восходящим методом) составляет 14,41 мм. Для экспоненциального распределения $p_{\text{знач}} = 0,058$ (параметр оценивается значением 0,216), для обобщенного Парето $p_{\text{знач}} = 0,29$ (параметры: 0,076, 13,406 и 14,41).

На рис. 3 продемонстрировано визуальное соответствие между экспериментальными данными и подогаанными распределениями (см. гистограммы и аппроксимирующие кривые на рис. 3, а и 3, б) для восходящего метода. Рисунок 3, в демонстрирует разницу пороговых значений, полученных двумя способами. Очевидно, что в отличие от случая разд. 2 (исходные данные) восходящий метод устанавливает критическую планку слишком низко и в рассмотрении окажется избыточное количество пиков.

Для объемов осадков за дождливые периоды в Элисте верхний критический уровень (получен нисходящим методом) составляет 28 мм (шаг изменения уровня — 0,01 мм, уровень значимости статистического критерия при проверке по χ^2 -тесту выбран равным 0,01). Для экспоненциального распределения $p_{\text{знач}} = 0,082$ (параметр оценивается значением 0,029), для обобщенного Парето $p_{\text{знач}} = 0,44$ (параметры: -0,095, 13,66 и 28). Нижний критический уровень (получен восходящим методом) составляет 10,71 мм. Для экспоненциального распределения $p_{\text{знач}} = 0,062$ (параметр оценивается значением 0,183), для обобщенного Парето $p_{\text{знач}} = 0,21$ (параметры: 0,066, 9,586 и 10,71).

На рис. 4 продемонстрировано визуальное соответствие между экспериментальными данными и подогаанными распределениями (см. гистограммы и аппроксимирующие кривые на рис. 4, а и 4, б) для восходящего метода. Рисунок 4, в демонстрирует разницу пороговых значений, полученных двумя способами. Снова восходящий метод устанавливает критическую планку слишком низко, и более интересным представляется рассмотрение уровня для нисходящего метода.

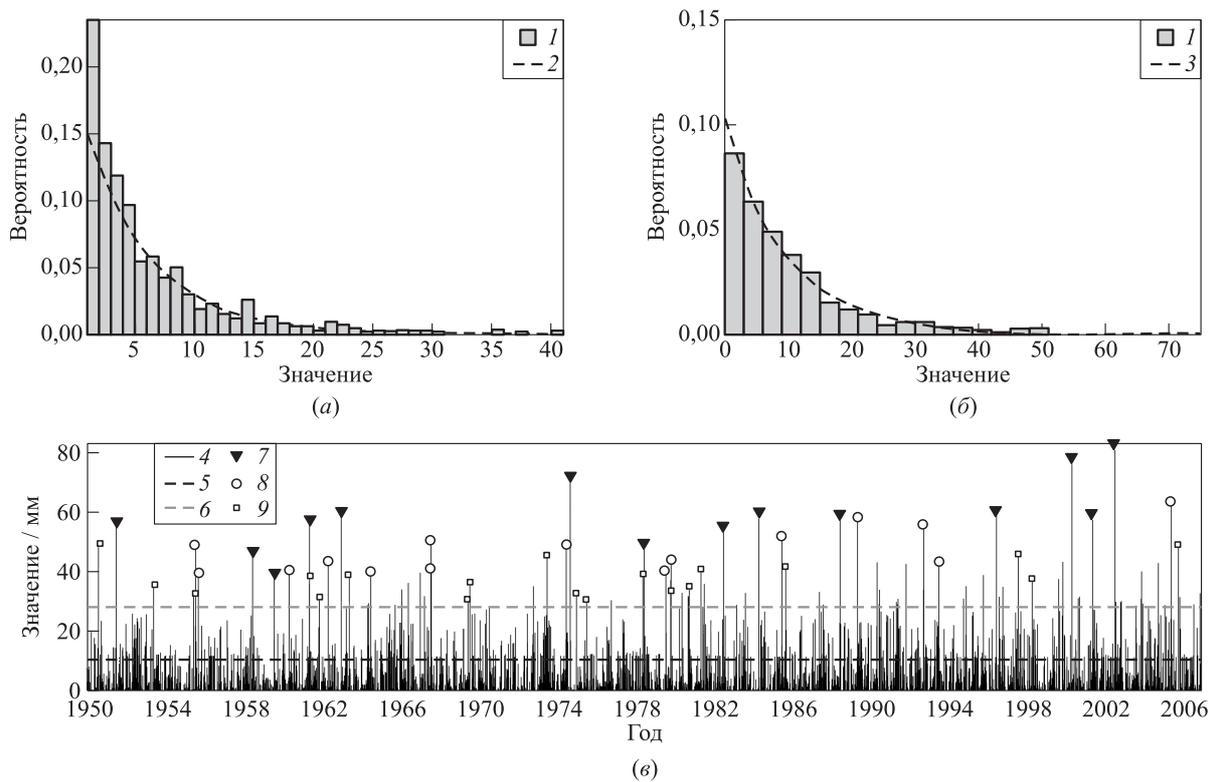


Рис. 4 Нисходящий, восходящий и параметрический методы определения экстремальности: (а) распределение интервалов между превышениями порога; (б) распределение превышений порога; (в) объемы осадков за дождливые периоды, Элиста; 1 — гистограмма; 2 — экспоненциальное распределение с параметром 0,183; 3 — обобщенное распределение Парето с параметрами 0,066, 9,586 и 10,71; 4 — данные; 5 — порог 10,71 мм; 6 — порог 28 мм; 7 — абсолютно экстремальные значения; 8 — промежуточные; 9 — относительно экстремальные значения

4 Статистический тест экстремальности объемов, основанный на распределении Снедекора–Фишера

На рис. 3, в и 4, в нанесены дополнительные маркеры (треугольники, круги и квадраты), которыми отмечены экстремальные наблюдения, полученные в соответствии со статистическим критерием, описанным в работе [10].

Для перехода к естественной временной шкале на соответствующие графики наносятся пики, по величине совпадающие с объемами осадков, выпавших за дождливые периоды, а их расположение на временной шкале выбрано совпадающим с днем начала выпадения осадков. Используемый для разметки параметрический критерий существенным образом использует установленные в статьях [14–19] факты о соответствии распределений длительностей дождливых периодов отрицательному биномиальному распределению, а их объемов — гамма-распределению. Ниже опишем

соответствующую процедуру, предложенную в работе [10].

Пусть $m \in \mathbb{N}$ и $G_{r,\mu}^{(1)}, G_{r,\mu}^{(2)}, \dots, G_{r,\mu}^{(m)}$ — независимые случайные величины с общим гамма-распределением с параметрами $r > 0$ и $\mu > 0$. Рассмотрим статистику

$$R_0 = \frac{(m-1)G_{r,\mu}^{(1)}}{G_{r,\mu}^{(2)} + \dots + G_{r,\mu}^{(m)}} \stackrel{d}{=} \frac{k}{r} \frac{G_{r,\mu}}{G_{k,\mu}} \stackrel{d}{=} \frac{k}{r} \frac{G_{r,1}}{G_{k,1}} \stackrel{d}{=} Q_{r,k},$$

где $k = (m-1)r$; $Q_{k,r}$ — случайная величина с распределении Снедекора–Фишера с параметрами $k > 0$ и $r > 0$. Предположим, что V_1, \dots, V_m — суммарные объемы осадков, выпавших за m дождливых периодов. Рассмотрим следующий выборочный аналог величины R_0 :

$$SR_{0,i} = \frac{(m-1)V_i}{\sum_{j \neq i} V_j}.$$

Сформулируем гипотезу H_0 в следующем виде: «объем осадков V_i не является аномально большим относительно остальных $m-1$ величин». То-

гда статистика $SR_{0,i}$ имеет распределение Снедекора–Фишера с параметрами $r > 0$ и $k = (m - 1)r$. Обозначим через $q_{r,k}(1 - \alpha)$ квантиль распределения Снедекора–Фишера уровня $(1 - \alpha)$, $\alpha \in (0, 1)$. В случае, если $SR_{0,i} > q_{r,k}(1 - \alpha)$, гипотеза H_0 отвергается и объем V_i должен быть признан экстремально большим. Уровень значимости критерия установлен на уровне α .

Описанная процедура может быть расширена с помощью метода скользящего окна. Задавая его ширину (т.е. число элементов в выборке объемов дождливых периодов) равной m и сдвигая каждый раз на один элемент вправо (направление астрономического времени на графике с эксперимен-

тальными данными), можно последовательно проверить, является ли каждый конкретный объем экстремальным относительно других в смысле описанного выше критерия. Данная процедура может быть полезна в ситуации, когда в одно окно попадают достаточно близкие по абсолютной величине объемы, мало отличающиеся от максимального.

Таким образом, в рамках данного подхода каждый элемент (начиная с m) проверяется в точности m раз. Тогда возможны следующие варианты:

- (1) элемент признается аномальным во всех m случаях и соответствующее значение может считаться *абсолютно* экстремальным;

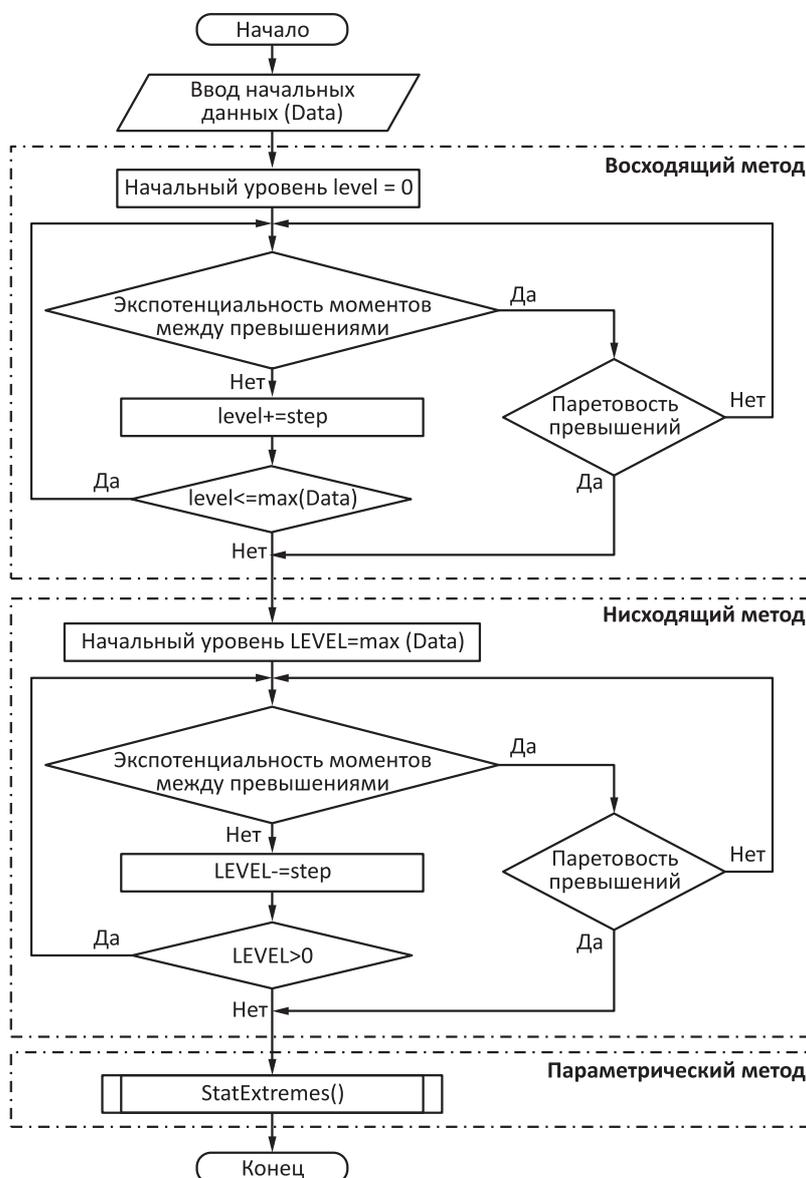


Рис. 5 Алгоритм определения экстремальных значений

- (2) элемент признается аномальным более чем в половине случаев (т.е. не меньше чем на $\lceil m/2 \rceil$ положениях окна) и соответствующее значение может считаться *промежуточным* экстремумом;
- (3) элемент признается аномальным менее чем в половине случаев и соответствующее значение может считаться *относительно* экстремальным.

Для приведения к астрономическому времени необходимо найти среднюю продолжительность дождливых и сухих периодов (используется модель на основе отрицательного биномиального распределения) и умножить на размер окна. На рис. 3 и 4 расчеты проводились для периода в 360 дней (при этом $m = 62$).

5 Алгоритм определения экстремальности значений объемов

Таким образом, алгоритм определения экстремальных значений (и сравнения результатов, полученных разными способами) может быть представлен в виде блок-схемы (рис. 5).

К данным Data последовательно применяются восходящий, нисходящий и параметрический методы поиска экстремальных значений. Для первых двух из них задается начальное значение ($level = 0$ для нисходящего и $LEVEL = \max(Data)$ для восходящего). Далее последовательно проверяются условия для теорем Реньи и Пикандса – Балкемы – Де Хаана при некотором заданном уровне значимости α . Процедура StatExtremes() реализует метод, описанный в разд. 4. В результате можно выявить наблюдения, которые признаются экстремальными с помощью каждого из этих методов, как продемонстрировано на рис. 3 и 4. Стоит отметить, что восходящий и нисходящий методы могут быть использованы для любых данных, а параметрический подход ориентирован на ряд дополнительных предположений о распределениях характеристик наблюдений (а также неотрицательность исходных значений).

6 Заключение

Итак, в работе рассмотрены восходящий и нисходящий методы определения порогового (экстремального) уровня в рамках идеологии PoT. Восходящий метод показывает отличные результаты

для суточных объемов осадков, однако для дождливых периодов необходимо использовать нисходящий метод. Кроме того, предложено сравнение результатов со статистическим критерием, который построен в рамках определенной (при этом весьма хорошо согласующейся с реальными данными) вероятностной модели. Указанный метод позволяет улучшить полученные результаты, однако требует ряда априорных предположений о распределениях характеристик исследуемых данных.

Литература

1. Groisman P. Y., Karl T. R., Easterling D. R., et al. Changes in the probability of heavy precipitation: Important indicators of climatic change // J. Climate, 1999. Vol. 42. P. 243–285.
2. Zolina O., Simmer C., Kapala A., Bachner S., Gulev S., Maechel H. Seasonally dependent changes of precipitation extremes over Germany since 1950 from a very dense observational network // J. Geophys. Res., 2008. Vol. 113. Art. No. D06110.
3. Leadbetter M. R. On a basis for “Peaks over Threshold” modeling // Stat. Probabil. Lett., 1991. Vol. 12. Iss. 4. P. 357–362.
4. Mendez F. J., Menendez M., Luceno A., Losada I. J. Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak over Threshold (PoT) model // J. Geophys. Res. Oceans, 2006. Vol. 111. Iss. C7. Art. No. C07024.
5. Kysely J., Picek J., Beranova R. Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold // Global Planet. Change, 2010. Vol. 72. Iss. 1–2. P. 55–68.
6. Begueria S., Vicente-Serrano S. M. Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques // J. Appl. Meteorol. Clim., 2006. Vol. 45. Iss. 1. P. 108–124.
7. Begueria S., Angulo-Martinez M., Vicente-Serrano S. M., Lopez-Moreno I. J., El-Kenawy A. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: A case study in northeast Spain from 1930 to 2006 // Int. J. Climatol., 2011. Vol. 31. Iss. 142. P. 2102–2114.
8. Roth M., Buishand T. A., Jongbloed G., Tank A. M. G., van Zanten J. H. A regional peaks-over-threshold model in a nonstationary climate // Water Resour. Res., 2012. Vol. 48. Art. No. W11533.
9. Gorshenin A. K., Korolev V. Yu. A methodology for the identification of extremal loading in data flows in information systems // Comm. Com. Inf. Sc., 2016. Vol. 638. P. 94–103.
10. Korolev V. Yu., Gorshenin A. K., Belyaev K. P. Statistical tests for extreme precipitation volumes // ArXiv: 1802.02928v3 [stat.ME], 2018.

11. Gnedenko B. V., Korolev V. Yu. Random summation: Limit theorems and applications. — Boca Raton, FL, USA: CRC Press, 1996. 288 p.
12. Balkema A., de Haan L. Residual life time at great age // *Annals Probability*, 1974. Vol. 2. No. 5. P. 792–804.
13. Pickands J. Statistical inference using extreme order statistics // *Ann. Stat.*, 1975. Vol. 3. No. 1. P. 119–131.
14. Горшенин А. К. О некоторых математических и программных методах построения структурных моделей информационных потоков // *Информатика и её применения*, 2017. Т. 11. Вып. 1. С. 58–68.
15. Горшенин А. К. Анализ вероятностно-статистических характеристик осадков на основе паттернов // *Информатика и её применения*, 2017. Т. 11. Вып. 4. С. 38–46.
16. Korolev V. Yu., Gorshenin A. K., Gulev S. K., Belyaev K. P., Grusho A. A. Statistical analysis of precipitation events // *AIP Conf. Proc.*, 2017. Vol. 1863. P. 090011-1–090011-4.
17. Королев В. Ю., Горшенин А. К. О распределении вероятностей экстремальных осадков // *Докл. РАН*, 2017. Т. 477. Вып. 5. С. 604–609.
18. Gorshenin A. K., Kuzmin V. Yu. Neural network forecasting of precipitation volumes using patterns // *Pattern Recogn. Image Anal.*, 2018. Vol. 28. No. 3. P. 450–461.
19. Gorshenin A. K., Korolev V. Yu. Scale mixtures of Frechet distributions as asymptotic approximations of extreme precipitation // *J. Math. Sci.*, 2018. Vol. 234. No. 6. P. 886–903.

Поступила в редакцию 15.10.18

DETERMINING THE EXTREMES OF PRECIPITATION VOLUMES BASED ON THE MODIFIED “PEAKS OVER THRESHOLD” METHOD

A. K. Gorshenin^{1,2} and V. Yu. Korolev^{1,2}

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Leninskie Gory, Moscow 119991, GSP-1, Russian Federation

Abstract: The problem of correct determination of extreme observations is very important when studying meteorological phenomena. The paper proposes ascending and descending methods for finding the threshold for extremes based on the Rényi theorem for thinning flows and the Pickands – Balkema – De Haan results. Using the observation data for 60 years for Potsdam and Elista, it is demonstrated that the ascending method can present excellent results for daily precipitation but for volumes of wet periods, the descending method should be used. The results of such nonparametric approaches are compared with the parametric criterion proposed in the previous papers by the authors.

Keywords: precipitation; wet periods; extreme values; thresholds; Rényi theorem; Pickands – Balkema – de Haan theorem; testing statistical hypotheses; data analysis

DOI: 10.14357/19922264180403

Acknowledgments

The research is supported by the Russian Foundation for Basic Research (project 17-07-00851) and the RF Presidential scholarship program (No. 538.2018.5).

References

1. Groisman, P. Y., T. R. Karl, D. R. Easterling, *et al.* 1999. Changes in the probability of heavy precipitation: Important indicators of climatic change. *J. Climate* 42:243–285.
2. Zolina, O., C. Simmer, A. Kapala, S. Bachner, S. Gulev, and H. Maechel. 2008. Seasonally dependent changes of precipitation extremes over Germany since 1950 from a very dense observational network. *J. Geophys. Res.* 113: D06110.
3. Leadbetter, M. R. 1991. On a basis for “Peaks over Threshold” modeling. *Stat. Probabil. Lett.* 12(4):357–362.
4. Mendez, F. J., M. Menendez, A. Luceno, and I. J. Losada. 2006. Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak over Threshold (PoT) model. *J. Geophys. Res. Oceans* 111(C7):C07024.
5. Kysely, J., J. Picek, and R. Beranova. 2010. Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global Planet. Change* 72(1–2):55–68.
6. Begueria, S., and S. M. Vicente-Serrano. 2006. Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *J. Appl. Meteorol. Clim.* 45(1):108–124.

7. Begueria, S., M. Angulo-Martinez, S. M. Vicente-Serrano, I. J. Lopez-Moreno, and A. El-Kenawy. 2011. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: A case study in northeast Spain from 1930 to 2006. *Int. J. Climatol.* 31(142):2102–2114.
8. Roth, M., T. A. Buishand, G. Jongbloed, A. M. G. Tank, and J. H. van Zanten. 2012. A regional peaks-over-threshold model in a nonstationary climate. *Water Resour. Res.* 48:W11533.
9. Gorshenin, A. K., and V. Yu. Korolev. 2016. A methodology for the identification of extremal loading in data flows in information systems. *Comm. Com. Inf. Sc.* 638:94–103.
10. Korolev, V. Yu., A. K. Gorshenin, and K. P. Belyaev. 2018. Statistical tests for extreme precipitation volumes. 1802.02928 [stat.ME]. Available at: <https://arxiv.org/abs/1802.02928v3> (accessed December 3, 2018).
11. Gnedenko, B. V., and V. Yu. Korolev. 1996. *Random summation: Limit theorems and applications*. Boca Raton, FL: CRC Press. 288 p.
12. Balkema, A., and L. de Haan. 1974. Residual life time at great age. *Ann. Probab.* 2(5):792–804.
13. Pickands, J. 1975. Statistical inference using extreme order statistics. *Ann. Stat.* 3(1):119–131.
14. Gorshenin, A. K. 2017. O nekotorykh matematicheskikh i programmnykh metodakh postroeniya strukturnykh modeley informatsionnykh potokov [On some mathematical and programming methods for construction of structural models of information flows]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):58–68.
15. Gorshenin, A. K. 2017. Analiz veroyatnostno-statisticheskikh kharakteristik osadkov na osnove pater-nov [Pattern-based analysis of probabilistic and statistical characteristics of precipitations]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):38–46.
16. Korolev, V. Yu., A. K. Gorshenin, S. K. Gulev, K. P. Belyaev, and A. A. Grusho. 2017. Statistical analysis of precipitation events. *AIP Conf. Proc.* 1863:090011-1–090011-4.
17. Korolev, V. Yu., and A. K. Gorshenin. 2017. The probability distribution of extreme precipitation. *Dokl. Earth Sci.* 477(2):1461–1466.
18. Gorshenin, A. K., and V. Yu. Kuzmin. 2018. Neural network forecasting of precipitation volumes using patterns. *Pattern Recogn. Image Anal.* 28(3):450–461.
19. Gorshenin, A. K., and V. Yu. Korolev. 2018. Scale mixtures of Frechet distributions as asymptotic approximations of extreme precipitation. *J. Math. Sci.* 234(6):886–903.

Received October 15, 2018

Contributors

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; leading scientist, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Leninskie Gory, Moscow 119991, GSP-1, Russian Federation; agorshenin@frcsc.ru

Korolev Victor Yu. (b. 1954) — Doctor of Science (PhD) in physics and mathematics, professor, Head of Department, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Leninskie Gory, Moscow 119991, GSP-1, Russian Federation; leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China; vkorolev@cs.msu.ru

ОПТИМИЗАЦИЯ ОБЪЕМА БУФЕРНОЙ ПАМЯТИ УЗЛА КОММУТАЦИИ ПРИ СХЕМЕ ПОЛНОГО РАЗДЕЛЕНИЯ ПАМЯТИ*

Я. М. Агаларов¹

Аннотация: Рассматривается задача оптимизации объема буферной памяти (БП) узла коммутации (УК) сети с коммутацией пакетов (КП), в котором используется схема полного разделения памяти между выходными линиями передачи, с учетом потерь пакетов из-за переполнения памяти и средних задержек пакетов. Для модели УК с пуассоновскими входящими потоками, общими распределениями времени обслуживания и одноканальными выходными линиями передачи получены свойства решения задачи и разработан алгоритм поиска оптимального объема. В рамках указанной модели разработан алгоритм поиска оптимального плана распределения БП при заданном ее объеме. Приведены результаты вычислительных экспериментов с использованием разработанных алгоритмов. Показана применимость полученных в работе подходов и алгоритмов к решению рассматриваемой задачи в рамках модели УК с рекуррентными входящими потоками, экспоненциальными временами обслуживания и одноканальными линиями передачи.

Ключевые слова: узел коммутации; управление потоками; распределение буферной памяти; оптимальный объем буферной памяти

DOI: 10.14357/19922264180404

1 Введение

Технико-экономические характеристики сети КП в значительной степени зависят от управления ресурсами и потоками требований в узлах-источниках и УК сети. Управление потоками предназначено для эффективной загрузки основных ресурсов (БП и каналов связи) в условиях их ограниченного объема с целью обеспечения требований к качеству обслуживания (Quality of Service — QoS) в сети. Зависимость качества обслуживания от объема памяти УК возникает из того, что при ограниченной пропускной способности каналов связи и небольшом объеме памяти могут быть недопустимо большие потери пакетов, что вызывает рост числа повторных передач пакетов и, как следствие, резкий рост нагрузки на отдельных участках сети (возможно, и в сети в целом), а при большом объеме памяти возможны недопустимо большие задержки пакетов в УК. Отсюда и возникает задача выбора оптимального объема БП УК с точки зрения выполнения определенных требований к качеству обслуживания пакетов и затратам, связанным со стоимостью оборудования и технической эксплуатации УК [1, 2]. При этом следует учитывать зависимость качества обслуживания в узле коммутации (и во всей сети в целом) и от схем (планов) распределения БП, из которых наиболее часто применяемыми на прак-

тике являются статические схемы управления [3, 4]: CS (Complete Sharing); CP (Complete Partitioning); SMQ (Sharing with Maximum Queue Length); SMA (Sharing with Minimum Allocation); SMQMA (Sharing with Maximum Queue Length and Minimum Allocation). Зависимость качества обслуживания в УК от схем распределения БП вызвана тем, что при бесконтрольном допуске пакетов в УК происходит захват буферов пакетами более интенсивных потоков, что приводит к неэффективному использованию выходных линий.

С момента зарождения сетей разработано великое множество алгоритмов локального управления потоками, многие из которых успешно применяются и в современных сетях. Основные результаты по управлению потоками в УК получены методом математического моделирования с использованием различных типов систем массового обслуживания (СМО). Некоторые из этих результатов описаны в работах [4–6], которые посвящены сравнительному анализу различных схем распределения БП, и в работах [7–13], в которых получены явные или алгоритмические решения задач по оптимизации параметров управления потоками в УК (объема БП, схем распределения БП УК).

В работе [5] приводится подробное описание методов анализа и расчета некоторых наиболее известных схем распределения БП и анализ известных

*Работа выполнена при частичной финансовой поддержке РФФИ (проекты 18-07-00692, 17-07-000717 и 17-07-00851).

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, agglar@yandex.ru

результатов оптимизации схем СР и SMQ в предположении пуассоновских входящих потоков, экспоненциального времени обслуживания и одноканальных линий передачи.

В работе [7] при этом же предположении рассмотрено динамическое управление распределением БП и получен для случая трех каналов вид допустимого пространства состояний, соответствующий оптимальному решению. В работе [8] рассмотрена двухступенчатая схема управления потоками в УК и приведен алгоритм оптимизации интенсивностей потоков пакетов, допускаемых на первой ступени в УК, при схеме SMQ на второй ступени и в предположении пуассоновских входных потоков, экспоненциальных времен обслуживания и многоканальных выходных линий передачи.

В работе [9] предложен эвристический алгоритм решения задачи совместного управления объемом БП и пропускной способностью линий передачи, который рассматривается в следующих аспектах:

- (1) при заданной степени загрузки каналов и заданной пропускной способности найти необходимый объем БП УК, обеспечивающий заданную среднюю задержку в сети и требуемую вероятность потери ячеек вследствие переполнения буферов;
- (2) при заданной степени загрузки каналов и известных объемах БП найти необходимую пропускную способность, обеспечивающую заданную среднюю задержку в сети и требуемую вероятность потери ячеек вследствие переполнения буферов.

В работе [10] авторы решили задачу выбора объема БП для СМО типа $M/M/1/r$, сформулировав ее как задачу нелинейного программирования:

- (1) для фиксированной входной нагрузки найти объем накопителя (БП), при котором средняя задержка достигает минимума, а интенсивность потерь не превышает заданной величины;
- (2) для фиксированного объема накопителя (БП) найти значение входной нагрузки, при которой средняя задержка достигает минимума, а интенсивность потерь не превышает заданной величины.

Для СМО типа $G/M/1$ и $M/G/1$ с интегральной целевой функцией, учитывающей среднюю задержку в БП и вероятности потерь пакетов из-за переполнения БП, оптимизационная задача ограничения длины очереди решена в работах [11–13].

Ниже рассматривается экономическая модель УК с L выходными линиями передачи, представленная в виде L СМО типа $M/G/1/N$, в кото-

рой качество обслуживания регулируется методом штрафных функций. В рамках этой модели решается задача выбора оптимального объема БП УК и оптимальной схемы распределения типа СР.

2 Модель узла коммутации и постановка задачи

В качестве модели УК рассматривается многопотоковая СМО типа $M_L/G/L$ с накопителем емкости K и L приборами, на которую поступает $L \leq K$ пуассоновских потоков заявок $\lambda_j > 0$, $j = 1, \dots, L$, к каждому из которых для обслуживания заявок прикреплен отдельный прибор, время обслуживания на котором является независимой случайной величиной с заданной функцией распределения $G_j(t)$, $j = 1, \dots, L$. Накопитель СМО моделирует БП, приборы — линии передачи. Заявки каждого потока обслуживаются в порядке поступления, места в накопителе распределяются по схеме СР, т. е. за каждой j -й линией закрепляется $k_j > 0$ мест накопителя, где $\sum_{i=1}^L k_i = K$. В дальнейшем вектор $\bar{k} = (k_1, \dots, k_L)$, удовлетворяющий указанному условию, будем называть планом распределения БП. Поступившая заявка j -го потока занимает любое свободное место в накопителе, если в момент ее поступления число занятых мест в накопителе меньше k_j . Если заявка поступила в накопитель, она занимает любое свободное место в накопителе и обслуживается на приборе в порядке поступления. Заявка покидает систему только при завершении обслуживания, освободив одновременно прибор и накопитель, а на освобожденный прибор поступает очередная заявка соответствующего типа (если таковая есть в накопителе). Качество работы системы оценивается с помощью интегральной целевой функции (функцией дохода системы) со стоимостными параметрами (весовыми коэффициентами), зависящей от среднего времени ожидания и вероятности потерь заявок. Функция дохода имеет следующие стоимостные параметры:

$C_{0j} > 0$ — плата, получаемая системой за обслуживание единицы длины заявки j -го потока;

$C_{1j} > 0$ — величина штрафа, который платит система, если отклонена (потеряна) заявка j -го потока;

$C_{2j} > 0$ — величина штрафа системы за единицу времени задержки заявки j -го потока;

$C_{3j} > 0$ — величина штрафа системы за единицу времени простоя j -й линии;

$C_{4j} > 0$ — затраты системы в единицу времени на техническое обслуживание j -й линии.

Считается, что плату за обслуживание система получает в момент завершения обслуживания каждой заявки в зависимости от длины заявки (длительности занятия прибора).

Как следует из определения схемы СР, для каждого фиксированного значения K и фиксированного плана распределения БП рассматриваемая СМО представляет собой совокупность L независимых СМО типа $M/G_j/1/k_j$ (в работе используются обозначения СМО по классификации Кендалла). Это делает возможным использование в дальнейшем результатов работ [11–13], полученных для систем $M/G/1$ и $G/M/1$ с ограничением на длину очереди.

Введем обозначения:

$\pi_{ji}^{k_j}$, $0 \leq i \leq k_j = 1$, — стационарная вероятность того, что в системе находится i заявок потока j в момент завершения обслуживания заявки j -м прибором;

$g_j^{k_j}$ — значение предельного дохода подсистемы $M/G_j/1/k_j$, усредненного по числу обслуженных j -й линией заявок, $j = 1, \dots, L$;

$g^{\bar{k}}$ — значение суммарного предельного дохода всей системы, усредненного по числу обслуженных заявок при плане $\bar{k} = (k_1, \dots, k_L)$;

$Q_j^{k_j}$ — предельный средний доход, получаемый системой $M/G_j/1/k_j$ в единицу времени;

$Q^{\bar{k}}$ — предельный суммарный средний доход, получаемый всей системой в единицу времени при заданном плане \bar{k} ;

$\bar{v}_j = \int_0^\infty t dG_j(t)$ — среднее время обслуживания заявки потока j , $0 < \bar{v}_j < \infty$, $j = 1, \dots, L$;

$\rho_j = \lambda_j \bar{v}_j$ — нагрузка на линию j , $j = 1, \dots, L$.

Пусть $A_K = \{\bar{k} > 0 : \sum_{i=1}^L k_i = K\}$, $U_{CP} = \{A_K, K \geq L\}$, $U_{CP,N} = \{\bar{k} : L \leq \sum_{j=1}^L k_j \leq N\}$,

$Q^k = \max_{\bar{k}: \bar{k} \in A_K} Q^{\bar{k}}$.

Ставятся следующие две задачи:

1. Для множества планов U_{CP} найти значение $K^* \geq L$ такое, что

$$\max_{K \geq L} Q^K = Q^{K^*}. \quad (1)$$

2. Найти $\bar{k}^0 \in U_{CP,N}$ такой, что

$$\max_{\bar{k} \in U_{CP,N}} k^{\bar{k}} = g^{\bar{k}^0}. \quad (2)$$

Ниже всюду через K^* будем обозначать решение задачи (1), через $\bar{k}^0 \in U_{CP,N}$ — решение задачи (2).

3 Решение задачи (1)

Так как при схеме СР рассматриваемая модель УК есть совокупность независимых СМО типа $M/G_j/1/k_j$, $j = 1, \dots, L$, то ее доход при заданном $\bar{k} \in U_{CP,K}$ представляется в виде суммы:

$$Q^{\bar{k}} = \sum_{j=1}^L Q_j^{k_j}.$$

Тогда, очевидно, верны равенства:

$$\begin{aligned} \max_{K \geq L} Q^K &= \max_{K \geq L} \max_{\bar{k}: \bar{k} \in A_K} Q^{\bar{k}} = \\ &= \max_{K \geq L} \max_{\bar{k}: \bar{k} \in A_K} \sum_{j=1}^L Q_j^{k_j} = \sum_{j=1}^L \max_{k_j > 0} Q_j^{k_j}, \quad (3) \end{aligned}$$

т. е. решением задачи (1) является $K^* = \sum_{j=1}^L k_j^*$, где k_j^* — решение задачи

$$\max_{k_j > 0} Q_j^{k_j}, \quad j = 1, \dots, L. \quad (4)$$

Таким образом, задача (1) сводится к задаче, аналогичной (4), рассмотренной в работе [12] для СМО типа $M/G/1$ с ограничением на длину очереди.

В работе [12] для дохода СМО типа $M/G/1$ при заданном пороговом значении длины очереди k получено выражение:

$$Q^k = \lambda (1 - \theta^k) g^k,$$

где g^k — значение предельного дохода системы $M/G/1/k$, усредненного по числу обслуженных линией заявок; θ^k — вероятность того, что поступившая заявка будет потеряна (для величин g^k и θ^k в [12] приведены явные аналитические выражения). Эта формула для системы $M/G_j/1/k_j$ ($j = 1, \dots, L$) в новых обозначениях записывается в виде:

$$Q_j^{k_j} = \lambda_j (1 - \theta_j^{k_j}) g_j^{k_j}.$$

Здесь (см. [12])

$$\theta_{k_j}^{k_j} = 1 - \frac{1}{\frac{k_j}{\pi_{j0}^{k_j}} + \rho}, \quad g^{k_j} = \sum_{i=0}^{k_j-1} q_{ji}^{k_j} \pi_{ji}^{k_j},$$

где

$$\begin{aligned} q_{ji}^{k_j} &= C_{0j} \bar{v} - C_{1j} \sum_{m=k_j-i+1}^{\infty} (m - k_j + i) r_{jm} - \\ &- \frac{C_{2j}}{\lambda_j} \left[\frac{1}{2} \sum_{m=2}^{k_j-i+1} (m-1) m r_{jm} + (k_j - i) \sum_{m=k_j-i+2}^{\infty} m r_{jm} - \right. \\ &- \left. \frac{1}{2} (k_j - i) (k_j - i + 1) \sum_{m=k_j-i+2}^{\infty} r_{jm} \right] - \\ &- C_{2j} (i-1) \bar{v}_j - C_4 \bar{v}_j, \quad 1 \leq i \leq k_j - 1; \end{aligned}$$

$$q_{j0}^{k_j} = q_{j1}^{k_j} - \frac{C_{3j} + C_{4j}}{\lambda_j};$$

$$r_{jm} = \int_0^\infty \frac{(\lambda_j v)^m}{m!} e^{-\lambda_j v} dG_j(v) \text{ при } m \geq 0.$$

Для системы $M/G_j/1/k_j$ функции $\tilde{G}(k)$ и $\tilde{F}(k)$, введенные в [12], в новых обозначениях примут вид:

$$H_j(k_j) = C_{0j} - \frac{C_{1j}}{\bar{v}_j} (\rho_j - 1) - C_{4j} - \frac{C_{2j}}{\rho_j} [\tilde{F}_j(k_j) + (k_j - 1)\rho_j]; \quad (5)$$

$$\tilde{F}(k_j) = \left(\sum_{i=1}^{k_j} \pi_{ji}^{k_j+1} \sum_{m=k_j-i+1}^\infty [m - (k_j - i + 1)] k_m + \pi_{j0}^{k_j+1} \sum_{m=k_j}^\infty (m - k_j) k_m \right) / \pi_{jk_j}^{k_j+1}.$$

Из (3), (5) и утверждений 1 и 2 работы [12] вытекают следующие два свойства решения задачи (1).

Свойство 1. При любых значениях параметров $0 \leq C_{ij} < \infty$, $i = 0, 1, 3, 4$, и $0 < C_{2j} < \infty$, $j = 1, \dots, L$, существует оптимальное значение $0 < K^* < \infty$, причем

$$K^* = \sum_{j=1}^L k_j^*,$$

где $k_j^* = \min \{k_j > 0 : H_j(k_j) \leq Q_j^{k_j}\}$. Если $C_{2j} = 0$ и $Q_j^1 < H_j(1)$ хотя бы для одного $j \in \overline{1, L}$, то $k_j^* = \infty$, $K^* = \infty$.

Свойство 2. Значение $0 < K^* < \infty$ удовлетворяет условию $\max_{K>0} Q^K = Q^{K^*}$ тогда и только тогда, когда

$$K^* = \sum_{j=1}^L k_j^*$$

и $\bar{k}^* = (k_1^*, \dots, k_L^*)$ удовлетворяет одному из двух условий:

(а) для каждого $j \in \overline{1, L}$

$$k_j^* = \begin{cases} \min \{r_j > 0 : H_j(k_j) \leq Q_j^{k_j}\} \\ \text{при } 0 < C_{2j} < \infty \text{ и } H_j(1) > Q_j^1; \\ 1 & \text{при } H_j(1) \leq Q_j^1; \end{cases}$$

(б) для каждого $j \in \overline{1, L}$

$$Q_j^{k_j^*-1} < Q_j^{k_j^*}; \quad Q_j^{k_j^*+1} \leq Q_j^{k_j^*}.$$

Из условия (а) свойства 2 вытекает следующий алгоритм нахождения оптимальных значений \bar{k}^* и K^* [12].

1. Положить $j = 1$.
2. Положить $k_j = 1$.
3. Вычислить $H_j(k_j)$ и $Q_j^{k_j}$.
4. Если $C_{2j} = 0$ и $Q_j^{k_j} < H_j(k_j)$, то положить $k_j^* = \infty$ и перейти к п. 9.
5. Если выполняется неравенство $Q_j^{k_j} \geq H_j(k_j)$, то перейти к п. 8.
6. Увеличить k_j на единицу.
7. Вычислить $H_j(k_j)$ и $Q_j^{k_j}$ и перейти к п. 4.
8. Положить $k_j^* = k_j$.
9. Если $j < L$, то увеличить j на единицу и перейти к п. 2.
10. Положить $K^* = \sum_{j=1}^L k_j^*$.
11. Конец алгоритма.

Приведем некоторые выражения, упрощающие расчет величин $\tilde{F}_j(k_j)$ и $H_j(k_j)$, $j = 1, \dots, L$, и показывающие некоторые их свойства. Используя соотношения для $\pi_{ji}^{k_j}$ и $\bar{F}_j(k_j)$, полученные в [12] (см. (4) и (12)), найдем:

$$\tilde{F}_j(k_j) = \bar{F}_j(k_j) + \rho_j - 1 = \frac{\pi_{j0}^{k_j+1} + \rho_j - 1}{\pi_{jk_j}^{k_j+1}} - r_{j0}. \quad (6)$$

Отсюда получаем:

$$\begin{aligned} \tilde{F}_j(k_j) - \tilde{F}_j(k_j + 1) &= \\ &= \frac{\pi_{j0}^{k_j+1}}{\pi_{jk_j}^{k_j+1}} + \frac{\rho_j - 1}{\pi_{jk_j}^{k_j+1}} - \frac{\pi_{j0}^{k_j+2}}{\pi_{jk_j}^{k_j+2}} - \frac{\rho_j - 1}{\pi_{jk_j}^{k_j+2}} = \frac{\pi_{j0}^{k_j+1}}{\pi_{jk_j}^{k_j+1}} - \\ &- \frac{\pi_{j0}^{k_j+1} (1 - \pi_{jk_j}^{k_j+2})}{\pi_{jk_j}^{k_j+2}} + (\rho_j - 1) \left(\frac{1}{\pi_{jk_j}^{k_j+1}} - \frac{1}{\pi_{jk_j}^{k_j+2}} \right) = \\ &= \left(\pi_{j0}^{k_j+1} + \rho_j - 1 \right) \left(\frac{1}{\pi_{jk_j}^{k_j+1}} - \frac{1}{\pi_{jk_j}^{k_j+2}} \right) + \pi_{j0}^{k_j+1}. \end{aligned}$$

Из правой части последнего равенства следует, например, что величины $\tilde{F}_j(k_j) - \tilde{F}_j(k_j + 1) < \rho_j$ (т. е. $H_j(k_j)$ — убывающая функция по k_j) при условии $\pi_{j0}^{k_j+1} \leq \rho_j$ (заметим, что выполняются неравенства $\pi_{j0}^{k_j+1} + \rho_j - 1 > 0$, $1/\pi_{jk_j}^{k_j+1} - 1/\pi_{jk_j}^{k_j+2} < 0$, так как $\pi_{j0}^{k_j+1}$ и $\pi_{jk_j}^{k_j+1}$ убывают по $k_j \geq 0$ и $\lim_{k_j \rightarrow \infty} \pi_{j0}^{k_j+1} = 1 - \rho_j$ [14]).

Из формул (5), (6) данной статьи и (4) из [12] следует, что при известных значениях величин r_{ji} , $i = 1, \dots, k_j$, $j = 1, \dots, L$, для вычисления значений \bar{k}^* и K^* предложенный алгоритм требует порядка $\sum_{j=1}^L k_j^{*2}$ арифметических операций.

4 Решение задачи (2)

Для решения задачи (2) предлагается алгоритм, состоящий из следующих пунктов.

1. Если $N = L$, то положить $k_j^0 = 1, j = 1, \dots, L, K^0 = L$ и перейти к п. 12.
2. Положить $K = L, m = 0, k_j = 1, \varphi_j = 0, j = 1, \dots, L$.
3. Вычислить $g_j^{k_j}, j = 1, \dots, L$.
4. Для $j = 1, \dots, L$ выполнить:
 - 1.1 Вычислить $g_j^{k_j+1}, \Delta_j(k_j) = g_j^{k_j+1} - g_j^{k_j}$;
 - 2.1 Если $\Delta_j(k_j) \leq 0$, то положить $\varphi_j = 1, m = m + 1$.
5. Если $m = L$, то перейти к п. 11.
6. Найти $j_0 : \Delta_{j_0}(k_{j_0}) = \max\{\Delta_j(k_j), \varphi_j \neq 1, j = 1, \dots, L\}$.
7. Положить $K = K + 1, k_j = k_j + 1$ для $j = j_0$.
8. Если $K = N$, то перейти к п. 11.
9. Вычислить $g_j^{k_j+1}, \Delta_j(k_j) = g_j^{k_j+1} - g_j^{k_j}$ для $j = j_0$.
10. Если $\Delta_{j_0}(k_{j_0}) \leq 0$ то положить $\varphi_{j_0} = 1, m = m + 1$ и перейти к п. 5.
11. Положить $k_j^0 = k_j, j = 1, \dots, L, K^0 = \sum_{j=1}^L k_j^0$.
12. Конец алгоритма.

Справедливо следующее

Утверждение 1. *Результат работы алгоритма является решением задачи (2).*

Доказательство. Справедливость утверждения при $N = \infty$ следует непосредственно из условия (б) свойства 2 решения задачи (1). Действительно, согласно алгоритму в этом случае для каждого $j = 1, \dots, L$ величина k_j увеличивается до тех пор, пока не выполняются неравенства упомянутого выше условия свойства 2. В этом случае $K^0 = K^*$.

Докажем для случая $N < \infty$. Если $K^* = \sum_{j=1}^L k_j^* < N$, то, как легко видеть, алгоритм работает так же, как и при $N = \infty$, и, следовательно, утверждение справедливо и $K^0 = K^*$.

Пусть $K^* \geq N$. В работах [11, 12] для $g_j^{k_j}, j = 1, \dots, L$, приводится равенство:

$$g_j^{k_j} - g_j^{k_j+1} = \pi_{j k_j}^{k_j+1} \left[g_j^{k_j} - G_j(k_j) \right], \quad (7)$$

где $G_j(k_j) = H_j(k_j) \bar{v}_j$.

Согласно следствию теоремы 1 из [11] (см. пп. 1 и 3) верны неравенства $g_j^{k_j} < g_j^{k_j+1}$ при $0 < k_j < k_j^*, j = 1, \dots, L$. Тогда, так как $\pi_{j k_j}^{k_j+1} < 1$ и $H_j(k_j) \bar{v}_j -$

убывающие по k_j функции, из (6) получаем неравенства:

$$\Delta_j(k_j) < \Delta_j(k_j + 1) \quad \text{для } 0 < k_j \leq k_j^* - 2 \text{ при } k_j^* > 2. \quad (8)$$

Рассмотрим последовательности $\{\Delta_j(k_j), k_j > 0\}, j = 1, \dots, L$, и объединим их в одну последовательность $\{\Delta(k), k > 0\}$, упорядоченную по убыванию элементов. Обратим внимание, что алгоритм каждый раз, перед тем как увеличить K на единицу, в п. 6 находит элемент $\Delta(K + 1) = \Delta_{j_0}(k_{j_0})$ объединенной последовательности и выполняет п. 7. Заметим также, что

$$\begin{aligned} g^{\bar{k}} &= \sum_{j=1}^L g_j^{k_j} = \sum_{j=1}^L \left[\sum_{m=1}^{k_j-1} \Delta_j(m) + g_j^1 \right] = \\ &= \sum_{j=1}^L \sum_{m=1}^{k_j-1} \Delta_j(m) + \sum_{j=1}^L g_j^1, \end{aligned}$$

и, если $\bar{k}' = (k'_1, \dots, k'_L) \in U_{\text{СР}, K}$ удовлетворяет условию $g^{\bar{k}'} = \max_{\bar{k} \in U_{\text{СР}, K}} g^{\bar{k}}$, то $\max_{\bar{k} \in U_{\text{СР}, K+1}} g^{\bar{k}'} = g^{\bar{k}'} + \Delta_{j_0}(k'_{j_0})$, где j_0 вычисляется согласно п. 6 алгоритма. Тогда, так как $K^* \geq N$ и имеет место (8), согласно алгоритму по завершении п. 11 имеют место равенства:

$$K^0 = \sum_{j=1}^L k_j^0 = N;$$

$$\begin{aligned} g^{\bar{k}^0} &= \sum_{j=1}^L g_j^{k_j^0} = \sum_{j=1}^L \sum_{m=1}^{k_j^0-1} \Delta_j(m) + \sum_{j=1}^L g_j^1 = \\ &= \sum_{k=L+1}^N \Delta(k) + \sum_{j=1}^L g_j^1 = \\ &= \max_{\bar{k} \in U_{\text{СР}, N}} \left[\sum_{j=1}^L \sum_{m=1}^{k_j-1} \Delta_j(m) + \sum_{j=1}^L g_j^1 \right] = \max_{\bar{k} \in U_{\text{СР}, N}} g^{\bar{k}}. \end{aligned}$$

Следовательно, утверждение 1 справедливо.

5 Примеры

Рассмотрим СМО $M_2/G/2$ с функциями распределения времени обслуживания

$$\begin{aligned} G_1(t) &= f_{11} (1 - e^{-\mu_{11}t}) + f_{12} (1 - e^{-\mu_{12}t}); \\ G_2(t) &= f_{21} (1 - e^{-\mu_{21}t}) + f_{22} (1 - e^{-\mu_{22}t}), \end{aligned}$$

стоимостными параметрами $C_{10} = C_{20} = 30, C_{11} = C_{21} = 15, C_{12} = C_{22} = 4, C_{13} = C_{23} = 0, C_{14} = C_{24} = 0,4$ и одинаковыми входными потоками

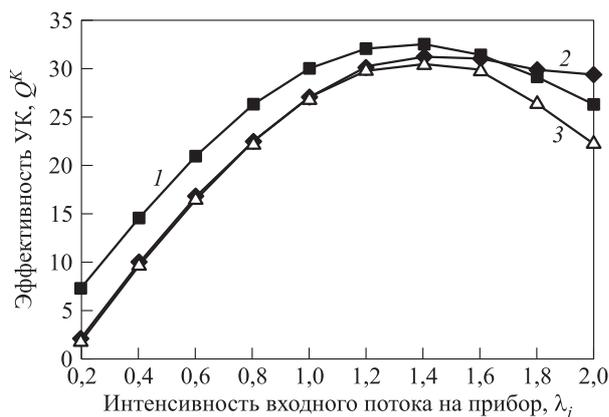


Рис. 1 Зависимость величины дохода системы от интенсивности входного потока: 1 — оптимальная схема СР; 2 — схема квадратного корня; 3 — схема СС

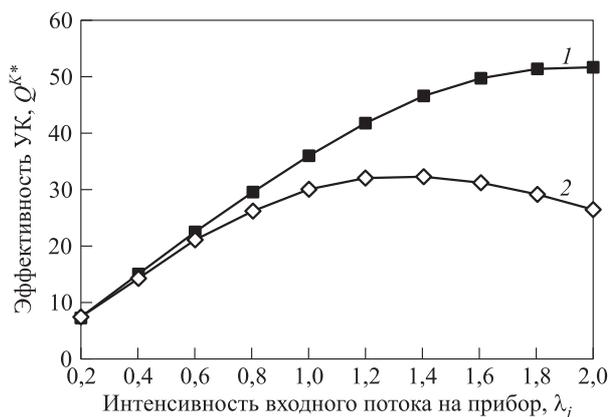


Рис. 2 Зависимость дохода системы от интенсивности входного потока и времени обслуживания приборами: 1 — $\mu_{11} = 1,5, \mu_{12} = 3, \mu_{21} = 2$ и $\mu_{22} = 4$; 2 — $\mu_{11} = 1, \mu_{12} = 2, \mu_{21} = 1,5$ и $\mu_{22} = 3$

на приборы (выходные линии передачи) $\lambda_1 = \lambda_2$. На графиках рис. 1 проиллюстрированы зависимости величины дохода Q^k рассматриваемого СМО от интенсивности входного потока для трех схем распределения БП: оптимальной схемы СР, схемы квадратного корня при $K = K^*$, схемы СС при $K = K^*$. На всех трех графиках доход обозначен через Q^k . Значения параметров функций $G_1(t)$ и $G_2(t)$ соответственно равны: $f_{11} = f_{12} = f_{21} = f_{22} = 0,5; \mu_{11} = 1; \mu_{12} = 2; \mu_{21} = 1,5; \mu_{22} = 3$.

На рис. 2 изображена зависимость значения дохода Q^{k*} системы от интенсивности входного потока и времени обслуживания приборами.

На рис. 3 приведена диаграмма зависимости оптимального значения K^* от интенсивности входного потока.

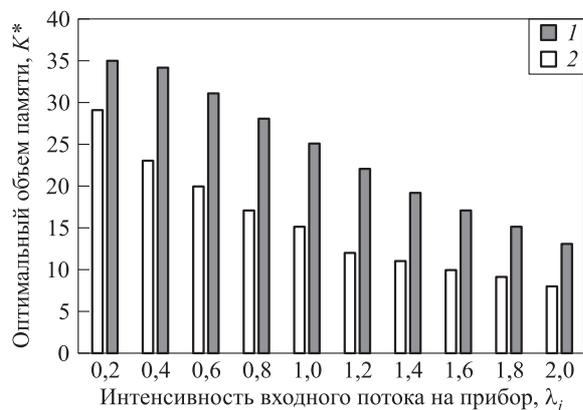


Рис. 3 Зависимость оптимального объема БП от интенсивности входного потока и времени обслуживания приборами: 1 — $\mu_{11} = 1,5, \mu_{12} = 3, \mu_{21} = 2$ и $\mu_{22} = 4$; 2 — $\mu_{11} = 1, \mu_{12} = 2, \mu_{21} = 1,5$ и $\mu_{22} = 3$

6 Заключение

На основании проведенных выше исследований можно сделать следующие выводы:

- при схеме СР, если параметр стоимости для времени задержки — положительная величина, то существует конечный объем БП, при котором функция дохода Q^k достигнет максимального значения на множестве планов $U_{СР}$;
- функция Q^k является унимодальной функцией от переменной $K > 0$;
- функция дохода УК зависит от объема БП, интенсивностей входных потоков и распределения БП между выходными линиями (см. рис. 1 и 2), т.е. оптимальная процедура локального управления потоками в УК должна управлять всеми тремя указанными параметрами;
- оптимальный объем БП существенно зависит от интенсивности входной нагрузки, механизма распределения БП и пропускной способности каналов связи (см. рис. 1 и 3).

Отметим, что соотношение вида (7), использованное при доказательстве приведенного выше утверждения 1, получено также и в работе [13] для СМО типа $G/M/1$ с пороговым значением длины очереди. Указанное соотношение (см. (12) в [13]) имеет вид:

$$g^k - g^{k+1} = \pi_0^{k+1} [g^k - G(k)], \quad k > 0,$$

где k соответствует объему БП; π_0^{k+1} и $G(k)$ обладают аналогичными свойствами величин $\pi_{jk}^{k_j+1}$ и $G_j(k_j)$, использованным выше в доказательстве

утверждения 1. Следовательно, алгоритм и утверждение 1, приведенные в разд. 3, применимы для рассмотренной выше модели УК также при предположениях: входящие потоки — рекуррентные; времена обслуживания — экспоненциальные. Однако заметим, что в этом случае, в отличие от рассмотренного выше, согласно работе [13] g^k — предельный доход, усредненный по числу поступивших в систему заявок.

Отметим также, что каждая из схем CP, CS, SMQ, SMA и SMQMA предпочтительней в смысле эффективности, чем остальные, только в соответствующей области значений параметров системы и выбранного критерия эффективности [5]. Полученные с помощью предложенных выше алгоритмов оптимальные при схеме CP значения объема БП и плана распределения БП могут рассматриваться для других схем как оценочные.

Литература

1. Клейнрок Л. Вычислительные системы с очередями / Пер. с англ. — М.: Мир, 1979. 600 с. (*Kleinrock L. Queueing systems. Vol. II: Computer applications.* — New York, NY, USA: Wiley, 1976. 549 p.)
2. Дэвис Д., Барбер Д., Прайс У., Соломонидес С. Сети связи для вычислительных систем / Пер. с англ. — М.: Мир, 1982. 563 с. (*Davies D. W., Barber D. L. A., Price W. L., Solomonides C. M.* 1979. *Computer networks and their protocols.* — New York, NY, USA: Wiley. 487 p.)
3. Irland M. Buffer management in a packet switch // IEEE T. Commun., 1978. Vol. 26. No. 3. P. 328–337.
4. Kamoun F., Kleinrock L. Analysis of shared finite storage in a computer networks node environment under general traffic conditions // IEEE T. Commun., 1980. Vol. 28. No. 7. P. 992–1003.
5. Башарин Г. П., Самуйлов К. Е. Об оптимальной структуре буферной памяти в сетях передачи данных с коммутацией пакетов. — М., 1982. 70 с.
6. Михеев П. А. Анализ стратегий разделения конечной буферной памяти маршрутизатора между выходными каналами // Автоматика и телемеханика, 2014. № 10. С. 125–128.
7. Forschini G., Gopinath B. Sharing memory optimally // IEEE T. Commun., 1983. Vol. 31. No. 3. P. 352–359.
8. Агаларов Я. М. Двухступенчатое адаптивное управление потоками в узле коммутации телекоммуникационной сети // Вестник РУДН. Сер. Прикладная математика и информатика, 2003. № 1. С. 134–141.
9. Линец Г. И. Управление объемом буферной памяти и пропускной способностью каналов в мультисервисных сетях // Инфокоммуникационные технологии, 2008. Т. 6. № 2. С. 62–64.
10. Жерновский Ю. В. Решение задач оптимального синтеза для некоторых марковских моделей обслуживания // Информационные процессы, 2010. Т. 10. № 3. С. 257–274.
11. Агаларов Я. М., Агаларов М. Я., Шоргин В. С. Об оптимальном пороговом значении длины очереди в одной задаче максимизации дохода СМО типа $M/G/1$ // Информатика и её применения, 2016. Т. 10. Вып. 2. С. 70–79.
12. Агаларов Я. М. Максимизация среднего стационарного дохода СМО типа $M/G/1$ // Информатика и её применения, 2017. Т. 11. Вып. 2. С. 25–32.
13. Агаларов Я. М., Шоргин В. С. Об одной задаче максимизации дохода системы массового обслуживания типа $G/M/1$ с пороговым управлением очередью // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 55–64.
14. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 529 с.

Поступила в редакцию 10.06.18

OPTIMIZATION OF BUFFER MEMORY SIZE OF SWITCHING NODE IN MODE OF FULL MEMORY SHARING

Ya. M. Agalarov

Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The problem of optimizing the volume of the buffer memory of the network switching node with the CP (Complete Partitioning) is considered, which uses a scheme for the complete memory sharing between the output transmission lines, taking into account packet loss due to memory overflow and average packet delays. The properties of the solution of the problem are obtained, and an algorithm for searching for the optimal volume is developed for the switching node model with Poisson incoming flows, common service time distributions, and single-channel output transmission lines. Within the given model, an algorithm for finding the optimal buffer

memory allocation plan for a given volume is developed. The results of computational experiments using the developed algorithms are presented. The applicability of the proposed algorithms and approaches is obtained in the framework of the switching node model with recurrent incoming flows, exponential service times, and single-channel transmission lines.

Keywords: switching node; flow management; buffer memory allocation; optimal buffer memory capacity

DOI: 10.14357/19922264180404

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (projects 18-07-00692, 17-07-000717, and 17-07-00851).

References

1. Kleinrock, L. 1976. *Queueing systems. Vol. II: Computer applications*. New York, NY: Wiley. 549 p.
2. Davies, D.W., D.L.A. Barber, W.L. Price, and C.M. Solomonides. 1979. *Computer networks and their protocols*. New York, NY: Wiley. 487 p.
3. Irland, M. 1978. Buffer management in a packet switch. *IEEE T. Commun.* 26(3):328–337.
4. Kamoun, F., and L. Kleinrock. 1980. Analysis of shared finite storage in a computer networks node environment under general traffic conditions. *IEEE Trans. Commun.* 28(7):992–1003.
5. Basharin, G. P., and K. E. Samuilov. 1982. *Ob optimal'noy strukture bufernoy pamyati v setyakh peredachi dannykh s kommutatsiy paketov* [On the optimal structure of BP in data transmission networks with packet commutation]. — Moscow. 70 p.
6. Mikheev, P.A. 2014. Analyzing sharing strategies for finite buffer memory in a router among outgoing channels. *Automat. Rem. Contr.* 75(10):1814–1825.
7. Forschini, G., and B. Gopinath. 1983. Sharing memory optimally. *IEEE T. Commun.* 31(3): 352–359.
8. Agalarov, Ya. M. 2003. Dvukhstupenchatoe adaptivnoe upravlenie potokami v uzle kommutatsii telekommunikatsionnoy seti [Two-level flow control for a switching node of communication network]. *Vestnik RUDN. Ser. Prikladnaya matematika i informatika* [PFUR Bull., Applied Mathematics and Computer Science] 1:134–141.
9. Linets, G. I. 2010. Upravlenie ob'emom bufernoy pamyati i propusknoy sposobnost'yu kanalov v mul'tiservisnykh setyakh [Volume management of buffer memory and throughput of channels in multiservice networks]. *Infokommunikatsionnye tekhnologii* [Information Communication Technologies] 6(2):62–64.
10. Zhernovyy, Ju. V. 2010. Reshenie zadach optimal'nogo sinteza dlya nekotorykh markovskikh modeley obsluzhivaniya [Solution of optimum synthesis problem for some Markov models of service]. *Informatsionnye protsessy* [Information Processes] 10(3):257–274.
11. Agalarov, Ya.M., M. Ya. Agalarov, and V. S. Shorgin. 2016. Ob optimal'nom porogovom znachenii dliny ocheredi v odnoy zadache maksimizatsii dokhoda SMO tipa $M/G/1$ [About the optimal threshold of queue length in particular problem of profit maximization in $M/G/1$ queueing system]. *Informatika i ee Primeneniya — Inform. Appl.* 10(2):70–79.
12. Agalarov, Ya. M. 2017. Maksimizatsiya srednego statsionarnogo dokhoda SMO tipa $M/G/1$ [Maximization of average stationary profit in $M/G/1$ queueing system]. *Informatika i ee Primeneniya — Inform. Appl.* 11(2):25–32.
13. Agalarov, Ya.M., and V.S. Shorgin. 2016. Ob odnoy zadache maksimizatsii dokhoda sistemy massovogo obsluzhivaniya tipa $G/M/1$ s porogovym upravleniem ochered'yu [Profit maximization in $G/M/1$ queueing system on a set of threshold strategies with two switch points]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4): 55–64.
14. Bocharov, P.P., and A.V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 529 p.

Received June 10, 2018

Contributor

Agalarov Yaver M. (b. 1952) — Candidate of Science (PhD) in technology, associate professor; leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; agglar@yandex.ru

СИСТЕМА ОБСЛУЖИВАНИЯ С ОТНОСИТЕЛЬНЫМ ПРИОРИТЕТОМ И ПРОФИЛАКТИКАМИ ПРИБОРА*

Е. С. Кондранин¹, В. Г. Ушаков²

Аннотация: Изучена одноканальная система массового обслуживания с двумя типами требований, бесконечным числом мест для ожидания, гиперэкспоненциальным входящим потоком и профилактиками обслуживающего прибора при освобождении системы. Тип требования определяется случайно с заданными вероятностями в момент его поступления в систему обслуживания. Требования первого типа имеют относительный приоритет перед требованиями второго типа. Найдено нестационарное совместное распределение числа требований каждого типа в системе. Профилактики прибора заключаются в том, что в момент освобождения системы от требований прибор на случайное время с заданным распределением становится недоступным для обслуживания. Если за время профилактики поступает хотя бы одно требование, то начинается нормальное функционирование системы. Если требования не поступают, то прибор отправляется на новую профилактику. Такие системы хорошо описывают функционирование большого числа реальных вычислительных и информационных систем.

Ключевые слова: гиперэкспоненциальный поток; профилактики обслуживающего прибора; одноканальная система; относительный приоритет; длина очереди

DOI: 10.14357/19922264180405

1 Введение

В классической системе массового обслуживания ожидание требований в очереди связано только с занятостью обслуживающего прибора. В то же время в реальных системах сам прибор может пребывать как в активном, так и в неактивном состоянии. Такое неактивное состояние прибора (в литературе на английском языке используется термин *vacation*, а на русском — профилактика или прогулка) может быть связано со многими причинами. В частности, системы обслуживания с профилактиками прибора хорошо описывают функционирование реальных вычислительных и информационных систем, в которых наряду с основными требованиями имеются второстепенные. Второстепенные требования всегда присутствуют в системе, а их обслуживание может проводиться только тогда, когда нет основных, т. е. в фоновом режиме.

С точки зрения самого процесса профилактики прибора существует несколько ее разновидностей. Во-первых, могут быть разными правила, задающие условия начала профилактики: прибор может брать перерыв только при полном исчерпании требований в очереди (*exhaustive service*) либо при наличии определенного их числа (*nonexhaustive service*). Во-

вторых, могут быть разными правила возвращения прибора в работу. С этой точки зрения различают случаи однократного (*single vacation*) и многократного (*multiple vacation*) перерыва в работе. В первом случае ушедший на профилактику прибор после ее окончания находится в рабочем состоянии независимо от наличия требований в системе. Во втором случае прибор, не обнаружив новых требований в очереди, уходит на новую профилактику.

В работах [1–4] можно найти обзор известных результатов, большое число постановок задач, описание различных приложений и обширную библиографию по анализу систем с профилактиками обслуживающего прибора.

В настоящей работе исследуется совместное распределение длин очередей в нестационарном режиме в однолинейной системе с ожиданием, гиперэкспоненциальным входящим потоком, двумя типами требований и относительным приоритетом. Аналогичная неприоритетная система обслуживания исследована в [5].

2 Описание модели

Рассматривается однолинейная система массового обслуживания с двумя приоритетными класса-

*Работа выполнена при финансовой поддержке РФФИ (проект 18-07-00678).

¹Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, ekondranin@yandex.ru

²Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vgushakov@mail.ru

ми требований. Входящий поток — гиперэкспоненциальный с функцией распределения интервалов между поступлениями требований вида:

$$A(t) = \sum_{i=1}^k c_i (1 - e^{-a_i t}), \quad t > 0, \quad a_i > 0, \quad c_i > 0,$$

$$a_i \neq a_j, \quad i \neq j, \quad \sum_{i=1}^k c_i = 1.$$

Каждое поступившее требование направляется в первый класс с вероятностью p и во второй класс с вероятностью $1 - p$ независимо от остальных требований. Требования первого класса обладают относительным приоритетом перед требованиями второго класса. Длительности обслуживания требований i -го приоритетного класса — независимые в совокупности и не зависящие от входящего потока случайные величины с функцией распределения $B_i(x)$, $i = 1, 2$. Если в некоторый момент времени система освободилась от требований, то обслуживающий прибор отправляется на профилактику, которая длится случайное время с функцией распределения $C(x)$. Не ограничивая общности, будем считать, что $B_i(x) < 1$ и $C(x) < 1$ для любого x и существуют плотности распределения $b_i(x)$ и $c(x)$. Обозначим:

$$\beta_i(s) = \int_0^{\infty} e^{-sx} b_i(x) dx; \quad \gamma(s) = \int_0^{\infty} e^{-sx} c(x) dx.$$

Пока прибор находится на профилактике, он не доступен для обслуживания. Если за время профилактики поступают требования, после ее завершения начинается их обслуживание. Если ни одно требование не поступает, то прибор отправляется на новую профилактику. Длительности различных профилактик являются независимыми случайными величинами и не зависят от входящего потока и времен обслуживания.

3 Вспомогательные результаты

Рассмотрим многочлен по μ степени k вида:

$$\prod_{i=1}^k (\mu + a_i) - (pz_1 + (1 - p)z_2) \sum_{j=1}^k c_j a_j \prod_{i \neq j} (\mu + a_i). \quad (1)$$

Занумеруем его корни $\mu_1(z_1, z_2), \dots, \mu_k(z_1, z_2)$ таким образом, чтобы они были непрерывными

функциями и $\mu_1(1, 1) = 0$. Тогда $\operatorname{Re} \mu_j(z_1, z_2) < 0$, $|z_1| < 1$, $|z_2| < 1$, $\mu_i(z_1, z_2) \neq \mu_j(z_1, z_2)$, $i \neq j$, $j = 1, \dots, k$. Обозначим:

$$\alpha_m(z_1, z_2) = \prod_{j \neq m} (\mu_m(z_1, z_2) - \mu_j(z_1, z_2)).$$

Справедливы следующие леммы.

Лемма 1. Для любого $l = 1, \dots, k$ система уравнений

$$z_j = \beta_j(s - \mu_l(z_1, z_2)), \quad j = 1, 2,$$

имеет единственное решение $z_i = z_{il}(s)$ такое, что $|z_{il}(s)| < 1$ при $l = 2, \dots, k$, $\operatorname{Re} s \geq 0$, $a z_{i1}(0) = 1$, $|z_{i1}(s)| < 1$ при $\operatorname{Re} s > 0$, $i = 1, 2$.

Лемма 2. При каждом $l = 1, \dots, k$ уравнение

$$z_1 = \beta_1(s - \mu_l(z_1, z_2))$$

имеет единственное решение $z_1 = z_{1l}(z_2, s)$, аналитическое в области $\operatorname{Re} s > 0$, $|z_2| < 1$.

Положим

$$\lambda_l(s) = \mu_l(z_{1l}(s), z_{2l}(s)).$$

4 Распределение длины очереди

Гиперэкспоненциальный поток можно рассматривать как пуассоновский поток со случайной интенсивностью a , которая принимает k различных значений a_1, \dots, a_k с вероятностями c_1, \dots, c_k . Текущее значение a разыгрывается в момент поступления требования и не меняется между двумя соседними поступлениями. Введем случайный процесс $j(t)$ такой, что если $a = a_j$ в момент времени t , то $j(t) = j$.

Целью работы является нахождение распределения случайного процесса $(L_1(t), L_2(t))$, где $L_i(t)$ — число требований из i -го приоритетного класса, находящихся в системе в момент времени t .

При сделанных предположениях относительно параметров изучаемой системы обслуживания процесс $(L_1(t), L_2(t))$ не является, вообще говоря, марковским. Пусть $i(t) = i$, $i = 1, 2$, если в момент времени t обслуживается требование из i -го приоритетного класса, и $i(t) = 0$, если в момент времени t прибор находится на профилактике. Случайный процесс $x(t)$ определим следующим образом. Если $i(t) \neq 0$, то $x(t)$ есть время, прошедшее с начала обслуживания требования, находящегося на приборе, до момента t . Если $i(t) = 0$,

то $x(t)$ есть время, прошедшее с начала профилактики прибора до момента t . Случайный процесс $(L_1(t), L_2(t), i(t), j(t), x(t))$ является однородным марковским процессом. Положим

$$P_{ij}(n_1, n_2, x, t) = \frac{\partial}{\partial x} \mathbf{P}(L_1(t) = n_1, L_2(t) = n_2, \\ i(t) = i, j(t) = j, x(t) < x), \quad x \geq 0, \\ j = 1, \dots, k, \quad i = 0, 1, 2;$$

$$\eta_i(x) = \frac{b_i(x)}{1 - B_i(x)}, \quad i = 1, 2; \quad \eta_0(x) = \frac{c(x)}{1 - C(x)};$$

$$\delta_{i,j} = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$$

Функции $P_{ij}(n_1, n_2, x, t)$ удовлетворяют при $x > 0$ системам дифференциальных уравнений:

$$\frac{\partial P_{ij}(n_1, n_2, x, t)}{\partial t} + \frac{\partial P_{ij}(n_1, n_2, x, t)}{\partial x} = \\ = -(a_j + \eta_i(x))P_{ij}(n_1, n_2, x, t) + \\ + c_j \sum_{l=1}^k a_l (p P_{il}(n_1 - 1, n_2, x, t) + \\ + (1 - p)P_{il}(n_1, n_2 - 1, x, t)) \quad (2)$$

и краевым условиям при $x = 0$:

$$P_{0j}(n_1, n_2, 0, t) = 0, \quad n_1 + n_2 > 0; \\ P_{0j}(0, 0, 0, t) = \int_0^\infty P_{0j}(0, 0, x, t)\eta_0(x) dx + \\ + \int_0^\infty P_{1j}(1, 0, x, t)\eta_1(x) dx + \\ + \int_0^\infty P_{2j}(0, 1, x, t)\eta_2(x) dx; \quad (3)$$

$$P_{1j}(n_1, n_2, 0, t) + P_{2j}(n_1, n_2, 0, t) = \\ = \int_0^\infty P_{1j}(n_1 + 1, n_2, x, t)\eta_1(x) dx + \\ + \int_0^\infty P_{2j}(n_1, n_2 + 1, x, t)\eta_2(x) dx + \\ + \int_0^\infty P_{0j}(n_1, n_2, 0, t)\eta_0(x) dx. \quad (4)$$

Будем предполагать, что в начальный момент времени $t = 0$ система свободна от требований,

а с начала профилактики прибора прошло случайное время с заданным распределением с плотностью $d(x)$. Таким образом,

$$P_{ij}(n_1, n_2, x, 0) = 0, \quad i = 1, 2; \\ P_{0j}(n_1, n_2, x, 0) = c_j d(x) \delta_{n_1+n_2, 0}, \quad j = 1, \dots, k.$$

Положим

$$p_{ij}(z_1, z_2, x, s) = \\ = \sum_{n_1=0}^\infty \sum_{n_2=0}^\infty z_1^{n_1} z_2^{n_2} \int_0^\infty e^{-st} P_{ij}(n_1, n_2, x, t) dt; \\ \psi(s) = \int_0^\infty e^{-sx} dx \int_0^\infty \frac{c(u+x)d(u)}{1-C(u)} du.$$

Тогда, учитывая начальные условия, из (2) получаем:

$$\frac{\partial p_{ij}(z_1, z_2, x, s)}{\partial x} = \\ = -(s + a_j + \eta_i(x)) p_{ij}(z_1, z_2, x, s) + \\ + c_j (pz_1 + (1-p)z_2) \sum_{l=1}^k a_l p_{il}(z_1, z_2, x, s), \\ i = 1, 2; \quad (5)$$

$$\frac{\partial p_{0j}(z_1, z_2, x, s)}{\partial x} = \\ = -(s + a_j + \eta_0(x)) p_{0j}(z_1, z_2, x, s) + \\ + c_j (pz_1 + (1-p)z_2) \sum_{l=1}^k a_l p_{0l}(z_1, z_2, x, s) + \\ + c_j d(x). \quad (6)$$

Решения (5) и (6) имеют вид:

$$p_{ij}(z_1, z_2, x, s) = (1 - B_i(x)) c_j \times \\ \times \sum_{m=1}^k \frac{\gamma_i^{(m)}(z_1, z_2, s)}{\mu_m(z_1, z_2) + a_j} e^{-(s - \mu_m(z_1, z_2))x}, \\ i = 1, 2, \quad (7)$$

$$p_{0j}(z_1, z_2, x, s) = \\ = (1 - C(x)) c_j \sum_{m=1}^k e^{-(s - \mu_m(z_1, z_2))x} \left(\delta^{(m)}(z_1, z_2, s) + \right. \\ \left. + \alpha_m^{-1}(z_1, z_2) \prod_{l=1}^k (\mu_m(z_1, z_2) + a_l) \times \right. \\ \left. \times \int_0^x e^{(s - \mu_m(z_1, z_2))u} \frac{d(u)}{1 - C(u)} du \right) / (\mu_m(z_1, z_2) + \\ + a_j), \quad (8)$$

где функции $\gamma_i^{(m)}(z_1, z_2, s)$ и $\delta^{(m)}(z_1, z_2, s)$ являются произвольными функциями указанных переменных и определяются из краевых условий. Из (3) и (4) получаем:

$$\begin{aligned}
 & p_{1j}(z_1, z_2, 0, s) + p_{2j}(z_1, z_2, 0, s) = \\
 & = z_1^{-1} \int_0^\infty p_{1j}(z_1, z_2, x, s) \eta_1(x) dx + \\
 & + z_2^{-1} \int_0^\infty p_{2j}(z_1, z_2, x, s) \eta_2(x) dx + \\
 & + \int_0^\infty p_{0j}(z_1, z_2, x, s) \eta_0(x) dx - p_{0j}(z_1, z_2, 0, s). \quad (9)
 \end{aligned}$$

Заметим, что $p_{0j}(z_1, z_2, 0, s)$ не зависит от z_1 и z_2 , т. е. $p_{0j}(z_1, z_2, 0, s) = q_j(s)$. Подставляя (7) и (8) в (9), получаем:

$$\begin{aligned}
 & \gamma_1^{(m)}(z_1, z_2, s) (1 - z_1^{-1} \beta_1(s - \mu_m(z_1, z_2))) + \\
 & + \gamma_2^{(m)}(z_1, z_2, s) (1 - z_2^{-1} \beta_2(s - \mu_m(z_1, z_2))) = \\
 & = \delta^{(m)}(z_1, z_2, s) (\gamma(s - \mu_m(z_1, z_2)) - 1) + \\
 & + \alpha_m^{-1}(z_1, z_2) \prod_{l=1}^k (\mu_m(z_1, z_2) + a_l) \psi(s - \mu_m(z_1, z_2)), \\
 & j = 1, \dots, k. \quad (10)
 \end{aligned}$$

В силу леммы 1 левая часть (10) обращается в 0 при $z_1 = z_{1m}(s)$ и $z_2 = z_{2m}(s)$, $m = 1, \dots, k$. Следовательно,

$$\begin{aligned}
 & \delta^{(m)}(z_{1m}(s), z_{2m}(s), s) = \\
 & = \frac{\psi(s - \lambda_m(s))}{\alpha_m(z_{1m}(s), z_{2m}(s))(1 - \gamma(s - \lambda_m(s)))} \times \\
 & \times \prod_{l=1}^k (\lambda_m(s) + a_l). \quad (11)
 \end{aligned}$$

Из (8) следует, что

$$q_j(s) = c_j \sum_{m=1}^k \frac{\delta^{(m)}(z_1, z_2, s)}{\mu_m(z_1, z_2) + a_j}, \quad j = 1, \dots, k.$$

Решая эту систему уравнений относительно $\delta^{(m)}(z_1, z_2, s)$, получаем:

$$\begin{aligned}
 & \delta^{(m)}(z_1, z_2, s) = (pz_1 + (1-p)z_2) \times \\
 & \times \frac{\prod_{j=1}^k (\mu_m(z_1, z_2) + a_j)}{\alpha_m(z_1, z_2)} \sum_{l=1}^k \frac{a_l q_l(s)}{\mu_m(z_1, z_2) + a_l}. \quad (12)
 \end{aligned}$$

Подставляя в (12) $z_1 = z_{1m}(s)$ и $z_2 = z_{2m}(s)$, имеем:

$$\begin{aligned}
 & \delta^{(m)}(z_{1m}(s), z_{2m}(s), s) = \\
 & = (pz_{1m}(s) + (1-p)z_{2m}(s)) \times \\
 & \times \frac{\prod_{j=1}^k (\lambda_m(s) + a_j)}{\alpha_m(z_{1m}(s), z_{2m}(s))} \sum_{l=1}^k \frac{a_l q_l(s)}{\lambda_m(s) + a_l}. \quad (13)
 \end{aligned}$$

Сравнивая два представления (11) в (13) для $\delta^{(m)}(z_m(s), s)$, получаем систему уравнений для $q_l(s)$:

$$\begin{aligned}
 & \sum_{l=1}^k \frac{a_l q_l(s)}{\lambda_m(s) + a_l} = \\
 & = \frac{\psi(s - \lambda_m(s))}{(pz_{1m}(s) + (1-p)z_{2m}(s))(1 - \gamma(s - \lambda_m(s)))}, \\
 & m = 1, \dots, k,
 \end{aligned}$$

из которой находим

$$\begin{aligned}
 & q_l(s) = c_l \prod_{j=1}^k (\lambda_l(s) + a_j) \sum_{m=1}^k \frac{\psi(s - \lambda_m(s))}{(1 - \gamma(s - \lambda_m(s)))(\lambda_m(s) + a_l)} \times \\
 & \times \prod_{n \neq m} (\lambda_m(s) - \lambda_n(s)). \quad (14)
 \end{aligned}$$

Подставляя (14) в (12) и учитывая (1), получаем:

$$\begin{aligned}
 & \delta^{(m)}(z_1, z_2, s) = \frac{(pz_1 + (1-p)z_2)}{\alpha_m(z_1, z_2)} \times \\
 & \times \sum_{j=1}^k \frac{\psi(s - \lambda_j(s)) \prod_{l=1}^k (\lambda_j(s) + a_l)}{(pz_{1j}(s) + (1-p)z_{2j}(s))(1 - \gamma(s - \lambda_j(s)))} \times \\
 & \times \prod_{\nu \neq j} \frac{\mu_m(z_1, z_2) - \lambda_\nu(s)}{\lambda_j(s) - \lambda_\nu(s)}.
 \end{aligned}$$

Положим

$$\lambda_m(z_2, s) = \mu_m(z_{1m}(z_2, s), z_2), \quad m = 1, \dots, k.$$

Подставляя в (10) $z_1 = z_{1m}(z_2, s)$, имеем:

$$\begin{aligned}
 & \gamma_2^{(m)}(z_{1m}(z_2, s), z_2, s) = \\
 & = \frac{\delta^{(m)}(z_{1m}(z_2, s), z_2, s) (\gamma_m(s - \lambda_m(z_2, s)) - 1)}{1 - z_2^{-1} \beta_2(s - \lambda_m(z_2, s))} + \\
 & + \alpha_m^{-1}(z_{1m}(z_2, s), z_2) \psi(s - \lambda_m(z_2, s)) \prod_{l=1}^k (\lambda_m(z_2, s) + a_l) / (1 - z_2^{-1} \beta_2(s - \lambda_m(z_2, s))). \quad (15)
 \end{aligned}$$

Далее, из (7) следует:

$$p_{2j}(z_1, z_2, 0, s) = c_j \sum_{m=1}^k \frac{\gamma_2^{(m)}(z_1, z_2, s)}{\mu_m(z_1, z_2) + a_j}.$$

Отсюда

$$\begin{aligned} \gamma_2^{(m)}(z_1, z_2, s) &= \frac{pz_1 + (1-p)z_2}{\alpha_m(z_1, z_2)} \times \\ &\times \prod_{j=1}^k (\mu_m(z_1, z_2) + a_j) \sum_{l=1}^k \frac{a_l p_{2l}(z_1, z_2, 0, s)}{\mu_m(z_1, z_2) + a_l}. \end{aligned} \quad (16)$$

Так как $p_{2j}(z_1, z_2, 0, s)$ не зависит от z_1 , то

$$\begin{aligned} p_{2j}(z_1, z_2, 0, s) &= \\ &= c_j \sum_{m=1}^k \frac{\gamma_2^{(m)}(z_{1m}(z_2, s), z_2, s)}{\lambda_m(z_2, s) + a_j}. \end{aligned} \quad (17)$$

Таким образом, соотношения (15)–(17) полностью определяют $\gamma_2^{(m)}(z_1, z_2, s)$ при любых z_1 и z_2 . Теперь из (10) можно найти $\gamma_2^{(m)}(z_1, z_2, s)$.

Все функции, необходимые для вычисления $p_{ij}(z_1, z_2, x, s)$, $i = 0, 1, 2$, $j = 1, \dots, k$, найде-

ны. Искомая производящая функция процесса $(L_1(t), L_2(t))$ равна:

$$\begin{aligned} \int_0^\infty e^{-st} \mathbf{E} z_1^{L_1(t)} z_2^{L_2(t)} dt &= \\ &= \sum_{i=0}^2 \sum_{j=1}^k \int_0^\infty p_{ij}(z_1, z_2, x, s) dx. \end{aligned}$$

Литература

1. Doshi B. T. Queueing systems with vacations — a survey // Queueing Syst., 1986. Vol. 1. P. 29–66.
2. Takagi H. Time-dependent analysis of $M|G|1$ vacation models with exhaustive service // Queueing Syst., 1990. Vol. 6. P. 369–390.
3. Li J., Tian N., Zhang Z. G., Luh H. P. Analysis of the $M|G|1$ queue with exponentially working vacations — a matrix analytic approach // Queueing Syst., 2009. Vol. 61. P. 139–166.
4. Bouman N., Borst S. C., Boxma O. J., Leeuwaarden J. S. H. Queues with random back-offs // Queueing Syst., 2014. Vol. 77. P. 33–74.
5. Ушаков В. Г. Система обслуживания с гиперэкспоненциальным входящим потоком и профилактиками прибора // Информатика и её применения, 2016. Т. 10. Вып. 2. С. 93–98.

Поступила в редакцию 11.05.18

A HEAD OF THE LINE PRIORITY QUEUE WITH WORKING VACATIONS

E. S. Kondranin¹ and V. G. Ushakov^{1,2}

¹Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, Moscow 119991, GSP-1, Russian Federation

²Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The authors analyze the single-server queueing system with two types of customers, head of the line priority, hyperexponential input stream, and working vacations. The authors obtain the Laplace transform (with respect to an arbitrary point in time) of the joint distribution of server state, queue size, and elapsed time in that state. The authors restrict themselves to a system with exhaustive service (the queue must be empty when the server starts a vacation) and multiple vacations. The queueing systems with vacations have been well studied because of their applications in modeling computer networks, communication, and manufacturing systems. For example, in many digital systems, the processor is multiplexed among a number of jobs and, hence, is not available all the time to handle one job type. Besides such an application, theoretical interest in vacation models has been aroused with respect to their relationship with polling models.

Keywords: hyperexponential input stream; working vacations; single server; head of the line priority; queue length

DOI: 10.14357/19922264180405

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (project 18-07-00678).

References

1. Doshi, B. T. 1986. Queueing systems with vacations — a survey. *Queueing Syst.* 1:29–66.
2. Takagi, H. 1990. Time-dependent analysis of $M|G|M|1$ vacation models with exhaustive service. *Queueing Syst.* 6:369–390.
3. Li, J., N. Tian, Z. G. Zhang, and H. P. Luh. 2009. Analysis of the $M|G|1$ queue with exponentially working vacations — a matrix analytic approach. *Queueing Syst.* 61:139–166.
4. Bouman, N., S. C. Borst, O. J. Boxma, and J. S. H. Leeuwaarden. 2014. Queues with random back-offs. *Queueing Syst.* 77:33–74.
5. Ushakov, V. G. 2016. Sistema obsluzhivaniya s gipereksponentsialnym vkhodyashchim potokom i profilaktikami pribora [Queueing system with working vacations and hyperexponential input stream]. *Informatika i ee Primeneniya — Inform. Appl.* 10(2):93–98.

Received May 11, 2018

Contributors

Kondranin Egor S. (b. 1995) — MSc student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, Moscow 119991, GSP-1, Russian Federation; ekondranin@yandex.ru

Ushakov Vladimir G. (b. 1952) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, Moscow 119991, GSP-1, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vgushakov@mail.ru

МЕТОДЫ ИДЕНТИФИКАЦИИ ЗАХВАТА ХОСТА В РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЕ, ЗАЩИЩЕННОЙ С ПОМОЩЬЮ МЕТАДАННЫХ*

А. А. Грушо¹, Н. А. Грушо², М. В. Левыкин³, Е. Е. Тимонина⁴

Аннотация: Рассматривается модель распределенной информационно-вычислительной системы (РИВС), в которой разрешения на сетевые соединения определяются с помощью метаданных (МД). Метаданные служат упрощением моделей бизнес-процессов (БП). Доказано, что нарушитель информационной безопасности (ИБ), захвативший хост и аккуратно атакующий систему с помощью изменений выходных данных решаемых на этом хосте задач, не выявляется на уровне МД. Проблема связана с тем, что модели БП и, следовательно, МД оперируют переменными, для которых изменения конкретных значений не отражаются в их описании. Исключения составляют случаи выхода на запрещенные значения, например за пределы областей определения и множества значений функций, из которых построены информационные технологии (ИТ). Предложены дополнительные варианты защиты информации, учитывающие «невидимость» подобных нарушений ИБ на уровне МД.

Ключевые слова: информационная безопасность; информационные технологии; распределенные информационно-вычислительные системы; метаданные; запреты; угроза захвата хоста

DOI: 10.14357/19922264180406

1 Введение

Концепция защиты сетевых соединений в РИВС с помощью МД рассматривалась в работах [1–3]. Основная идея концепции состоит в том, чтобы в сети допускать только разрешенные соединения хостов. С этой целью реализуемые в РИВС информационные технологии должны быть описаны математическими моделями (Unified Modeling Language (UML), графы зависимостей и др.) [4], которые чаще всего сложны для управления соединениями. Упрощая эти модели с помощью набора функций, можно получить быстроисчисляемые ограничения на запрашиваемые соединения, которые называем МД. Такой подход к организации взаимодействий в сети решает целый ряд проблем ИБ, в том числе переносит управление сетью на уровень выше обмена данными.

Вместе с тем угрозы захвата хоста в сети не полностью исследованы [5]. В этой работе будет рассмотрено несколько не рассмотренных раньше вариантов действий нарушителя ИБ в случае захвата

им хоста. Оказалось, что некоторые виды нарушений ИБ при захвате хоста не могут быть выявлены на уровне МД по любым данным мониторинга. Необходимо получать дополнительные данные и создавать дополнительные технологии для решения рассматриваемой задачи.

2 Возможные стратегии нарушителя информационной безопасности

Некоторые стратегии нарушителя ИБ легко обнаружить на уровне МД. Получив информацию о перечне задач, с которыми нарушитель может установить соединение, но не зная скрытого экземпляра идентификатора текущей ИТ, которая должна исполняться на захваченном хосте, нарушитель может неудачно запросить соединение. В этом случае механизмы управления соединениями на уровне МД (задача 9, см. [2]) сразу идентифицируют захваченный хост.

* Работа поддержана РФФИ (проект 18-07-00274).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, grusho@yandex.ru

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, info@itake.ru

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, de_shiko@yahoo.com

⁴ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

В работе [5] рассматривается более изощренная атака, создающая задержки в выполнении всех действующих ИТ. Однако при правильном распределении задач по хостам (задача \mathcal{M} , см. [2]) эта атака компенсируется.

Наиболее сложные проблемы возникают, если нарушитель начинает аккуратно менять выходные данные задач, которые легально решаются на данном хосте. Простейший вариант этой задачи рассмотрен в [5]. Если нарушитель меняет выходные данные решаемой задачи на уже использованные старые данные этой задачи, то возможно создание баз данных (БД) значений хеш-функций выходов всех задач. Это могут быть централизованные или локальные БД [5]. Сравнение очередных значений хеш-функций выходных данных с сохраняющимися в БД позволяет выявить использование старых данных и захваченный хост. Это возможно, так как повторение значений является маловероятным событием и связано с повторами значений других переменных. Появление одиночного повторения значения переменной и отсутствие повторений в других переменных указывает на атаку.

3 Условия невидимости изменения данных на уровне метаданных

Рассмотрим проблему замены данных нарушителем в более общей постановке и определим условия, когда выявление подмены невозможно.

Существуют несколько хороших средств построения математических моделей БП, из упрощения которых получают МД. Назовем некоторые из них [4]:

- UML;
- BPEL (Business Processes Execution Language);
- ARIS (Architecture of Integration Information Systems).

Детальный анализ существующих средств описания моделей БП показал, что они не предполагают использования конкретных значений данных в этих моделях и связанных с ними ИТ. В этих моделях всюду используются переменные величины и описания доменов этих переменных, но нигде в моделях не используются значения данных. Это легко понять, так как все БП и ИТ предполагают многократное использование. Поэтому множественный характер значений данных не может входить в описание преобразований, реализуемых БП и ИТ, кроме как в виде переменных, областей определения и областей значений.

Отсюда можно сделать следующий вывод. Любые отображения преобразований, входящих в модели БП и ИТ, не зависят от значений данных. Результаты преобразований БП и ИТ также описываются переменными, поэтому в описании МД не входят конкретные значения переменных.

Отсюда вытекает следующее

Утверждение. *Метаданные не могут отслеживать намеренные изменения данных в задачах, кроме как с помощью выхода значений переменных в области запрещенных значений (запреты [6]).*

Доказательство. Изменение значений переменных злоумышленником соответствует изменению значений параметров, определенных с помощью некоторых переменных в описаниях БП и ИТ. Метаданные являются функциями от преобразований, участвующих в описаниях БП и ИТ. Поэтому все переменные, встречающиеся в МД, не содержат конкретных значений, кроме как в описаниях доменов или запретов.

Следствие. Нарушитель ИБ, захвативший хост и выбравший стратегию намеренного изменения данных, не может быть выявлен без дополнительной информации о доменах значений переменных или запретах, получаемых с помощью наблюдений данных мониторинга.

Это означает, что для выявления нарушителя, выбравшего стратегию изменения данных (далее просто нарушителя), необходимо знать запреты на значения переменных и надеяться, что изменения данных приведут к появлению запретов.

4 Стратегия выявления нарушителя с помощью запретов

Можно предложить несколько способов описания доменов и запретов при различных стратегиях изменения данных нарушителем.

Наиболее общий случай — это установление запретов на недопустимые значения выполняемых функций, когда появление запрета означает некорректное появление данных в каких-то задачах.

Поясним этот метод на следующем примере. Пусть A — задача, получающая исходные данные от задач B и C . Пусть хост $H(B)$ захвачен и изменяет данные, а хост $H(C)$ не захвачен. Тогда задача A получает от задачи B искаженные исходные данные x , а от задачи C — правильные исходные данные y . Выполнение задачи A означает вычисление некоторой функции f_A . Положим $f_A(x, y) = z$.

Значения x и y часто связаны между собой исходными условиями. Поэтому измененные данные x , возможно, нарушают эти связи, что отражается на значениях z . Если $z \notin D_A$, где D_A — множество допустимых значений функции f_A , то этот факт можно интерпретировать как сбой исходных данных. Отсюда следует, что необходимо исследовать хосты $H(B)$ и $H(C)$ на захват.

Исследование множеств значений всех функций, используемых в ИТ, — это сложная задача. Обычно происходит предварительное тестирование наиболее часто встречающихся исходных данных и для ряда задач A строятся $D'_A \subseteq D_A$, где D'_A — эмпирическая оценка множества D_A [7]. Использование множеств D'_A вместо множеств D_A порождает ложные тревоги, когда исходные данные легально отличаются от часто встречающихся данных. Поэтому эта методика обычно применяется для агрегированных задач, а их крайние значения получаются при использовании нагрузочных тестов.

5 Технология «ловушек»

«Ловушки» (Honey Pot) в ИТ отличаются от «ловушек», используемых для борьбы с вредоносным кодом. Предлагается следующая идея выявления нарушителя, изменяющего данные.

На этапе опытной эксплуатации запоминаются все траектории процессов исполнения ИТ, т. е. входные и выходные данные всех задач (множество U). Элементы множества U считаются «правильными». В случае изменения данных хотя бы для одной из задач появятся аномалии [8]. В соответствии с некоторым (случайным) расписанием вместо очередного экземпляра ИТ запускается экземпляр ИТ $u \in U$. Если нарушитель постоянно заменяет данные на захваченном хосте, то этот хост выявляется на запуске тестовой технологии u , так как на траектории u полностью определены выходные данные всех задач. Если нарушитель применяет «византийскую» стратегию, т. е. изменяет правильные данные не на каждом экземпляре ИТ, а на случайно выбранных, то возможность выявления захваченного хоста можно определить на основе следующей вероятностной модели.

Пусть p — вероятность запуска контрольного экземпляра ИТ в последовательности запускаемых ИТ ($n = 1, 2, \dots$). Пусть q — вероятность изменения данных в рассматриваемом экземпляре ИТ. Тогда вероятность выявления изменения данных в очередном экземпляре ИТ равна pq . Согласно модели геометрического распределения вероятность выявления изменения данных за n шагов равна $1 - (1 - pq)^n$. Математическое ожидание

числа шагов до выявления изменения данных равно $1/(pq)$.

В рассматриваемом методе предполагается, что повторений контрольных траекторий нет, иначе возможно выявление контроля нарушителем, так как нарушитель запоминает проходящие через захваченный им хост значения параметров.

Реализация метода «ловушек» требует специальной архитектуры РИВС. Это может быть клиент-серверная архитектура или варианты, близкие к ней [3, 9]. Однако в клиент-серверной архитектуре исчезает необходимость непосредственного взаимодействия хостов между собой.

6 Заключение

Работа посвящена проблеме выявления нарушителя ИБ в условиях, когда на захваченном хосте им выбирается стратегия изменения данных в задачах ИТ, решаемых на этом хосте. Показано, что традиционные средства защиты информации не могут выявить нарушителя, кроме как путем выявления запретов значений вычисляемых функций.

Наиболее экономный и эффективный способ выявления захваченного хоста представляет собой «ловушку», которая основана на экземплярах правильных траекторий процессов исполнения ИТ. Однако реализация этого подхода требует ограничений на архитектуру РИВС.

Литература

1. Grusho A., Grusho N., Zabezhalo M., Zatsarinny A., Timonina E. Information security of SDN on the basis of meta data // Computer network security / Eds. J. Rak, J. Bay, I. V. Kotenko, et al. — Lecture notes in computer science ser. — Springer, 2017. Vol. 10446. P. 339–347. https://link.springer.com/chapter/10.1007/978-3-319-65127-9_27.
2. Grusho A. A., Timonina E. E., Shorgin S. Ya. Modelling for ensuring information security of the distributed information systems // 31th European Conference on Modelling and Simulation Proceedings. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2017. P. 656–660. http://www.scs-europe.net/dlib/2017/ecms2017acceptedpapers/0656-probstat_ECMS2017_0026.pdf.
3. Грушо А. А., Тимонина Е. Е., Шоргин С. Я. Иерархический метод порождения метаданных для управления сетевыми соединениями // Информатика и её применения, 2018. Т. 12. Вып. 2. С. 44–49.
4. Самуйлов К. Е., Чукарин А. В., Яркина Н. В. Бизнес-процессы и информационные технологии в управлении телекоммуникационными компаниями. — М.: Альпина Паблишерс, 2009. 442 с.

5. Grusho A., Timonina E., Shorgin S. Security models based on stochastic meta data // Analytical and computational methods in theory probability / Eds. V. Rykov, N. Singpurwalla, A. Zubkov. — Lecture notes in computer science ser. — Springer, 2017. Vol. 10684: P. 388–400. https://link.springer.com/chapter/10.1007/978-3-319-71504-9_32.
6. Грушо А. А., Тимонина Е. Е. Запреты в дискретных вероятностно-статистических задачах // Дискретная математика, 2011. Т. 23. № 2. С. 53–58.
7. Грушо А. А., Грушо Н. А., Тимонина Е. Е. Статистические методы определения запретов вероятностных мер на дискретных пространствах // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 54–57.
8. Grusho A., Grusho N., Timonina E. Detection of anomalies in non-numerical data // 8th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings. — Piscataway, NJ, USA: IEEE, 2016. P. 273–276. <https://ieeexplore.ieee.org/document/7765370/>.
9. Грушо А. А., Грушо Н. А., Забежайло М. И., Тимонина Е. Е. Защита ценной информации в информационных технологиях // Проблемы информационной безопасности. Компьютерные системы, 2018. № 2. С. 22–26.

Поступила в редакцию 24.09.18

METHODS OF IDENTIFICATION OF HOST CAPTURE IN A DISTRIBUTED INFORMATION SYSTEM WHICH IS PROTECTED ON THE BASIS OF META DATA

A. A. Grusho, N. A. Grusho, M. V. Levykin, and E. E. Timonina

Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The model of a distributed information system in which permissions on network connections are based on meta data is considered. Meta data are simplification of business process models. It is proved that the adversary of information security who captured a host and accurately attacked a system by means of changes of output data of tasks solved on this host cannot be detected at the level of meta data. The problem is connected with the fact that a business process model and, therefore, meta data operate with variables for which changes of specific values are not reflected in their description. Exceptions are output cases on forbidden values, for example, out of limits of definition ranges and a set of values of functions of which information technologies are constructed. Additional variants of information security measures which consider “invisibility” of similar violations of information security at the level of meta data are suggested.

Keywords: information security; information technologies; distributed information system; meta data; ban; threat of host capture

DOI: 10.14357/19922264180406

Acknowledgments

The paper was supported by the Russian Foundation for Basic Research (project 18-07-00274).

References

1. Grusho, A., N. Grusho, M. Zabezhailo, A. Zatsarinny, and E. Timonina. 2017. Information security of SDN on the basis of meta data. *Computer network security*. Eds. J. Rak, J. Bay, I. V. Kotenko, et al. Lecture notes in computer science ser. Springer. 10446:339–347. Available at: https://link.springer.com/chapter/10.1007/978-3-319-65127-9_27 (September 23, 2018).
2. Grusho, A. A., E. E. Timonina, and S. Ya. Shorgin. 2017. Modelling for ensuring information security of the distributed information systems. *31th European Conference on Modelling and Simulation Proceedings*. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 656–660. Available at: http://www.scs-europe.net/dlib/2017/ecms2017acceptedpapers/0656-probstat_ECMS2017_0026.pdf (accessed September 23, 2018).
3. Grusho, A. A., E. E. Timonina, and S. Ya. Shorgin. 2018. Ierarkhicheskiy metod porozhdeniya metadannykh dlya upravleniya setevymi soedineniyami [Hierarchical method of meta data generation for control of network con-

- nections]. *Informatika i ee Primeneniya — Inform. Appl.* 12(2):44–49.
4. Samuylov, K. E., A. V. Chukarin, and N. V. Yarkina. 2009. *Biznes-protsessy i informatsionnye tekhnologii v upravlenii telekommunikatsionnymi kompaniyami* [Business processes and information technologies in management of the telecommunication companies]. Moscow: Alpina Pabls. 442 p.
 5. Grusho, A., E. Timonina, and S. Shorgin. 2017. Security models based on stochastic meta data. *Analytical and computational methods in theory probability*. Eds. V. Rykov, N. Singpurwalla, and A. Zubkov. Lecture notes in computer science ser. Springer. 10684:388–400. Available at: https://link.springer.com/chapter/10.1007/978-3-319-71504-9_32 (accessed September 23, 2018).
 6. Grusho, A., and E. Timonina. 2011. Prohibitions in discrete probabilistic statistical problems. *Discrete Mathematics Applications* 21(3):275–281.
 7. Grusho, A. A., N. A. Grusho, and E. E. Timonina. 2013. Statisticheskie metody opredeleniya zapretov veroyatnostnykh mer na diskretnykh prostranstvakh [Statistical methods of definition of the bans of probability measures on discrete spaces]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):54–57.
 8. Grusho, A., N. Grusho, and E. Timonina. 2016. Detection of anomalies in non-numerical data. *8th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings*. Piscataway, NJ: IEEE. 273–276. Available at: <https://ieeexplore.ieee.org/document/7765370/> (accessed September 23, 2018).
 9. Grusho, A., N. Grusho, M. Zabezhaiko, and E. Timonina. 2018. Zashchita tsennoy informatsii v informatsionnykh tekhnologiyakh [Protection of valuable information in information technologies]. *Information Security Problems. Computer Systems* 2:22–26.

Received September 24, 2018

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; grusho@yandex.ru

Grusho Nikolai A. (b. 1982) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; info@itake.ru

Levykin Mikhail V. (b. 1985) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; de_shiko@yahoo.com

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; eltimon@yandex.ru

ИСПОЛЬЗОВАНИЕ ВЕРОЯТНОСТНОЙ МОДЕЛИ ВЫЧИСЛЕНИЙ ДЛЯ ТЕСТИРОВАНИЯ ОДНОГО КЛАССА ГОТОВЫХ К ИСПОЛЬЗОВАНИЮ ПРОГРАММНЫХ КОМПОНЕНТОВ ЛОКАЛЬНЫХ И СЕТЕВЫХ СИСТЕМ*

Б. М. Басок¹, В. Н. Захаров², С. Л. Френкель³

Аннотация: Обсуждается и анализируется возможность обеспечения эффективного тестирования готовых к использованию программных продуктов (ГИПП), решающих задачи вычисления функций, в условиях отсутствия полной информации, необходимой для традиционного тестирования. Под эффективностью понимается возможность обеспечения сколь угодно высокой вероятности обнаружения возможных ошибок вычислений, не выявленных при выходном контроле, по мере роста числа проверок. В качестве концептуальной модели предлагается использовать свойства функций со случайной самоприводимостью (random self-reducible function — RSR), т. е. функций, вычисление которых на конкретном входном наборе можно свести к вычислению на нескольких случайно выбранных входных наборах. Обосновывается рациональность обеспечения свойств самотестируемости в коммерческих вычислительных ГИПП.

Ключевые слова: тестирование программ; самотестирование

DOI: 10.14357/19922264180407

1 Введение

При приобретении ГИПП перед пользователем стоит задача убедиться в его полной исправности, в его способности без ошибок выполнять те функции, которые указаны в прилагаемой к нему документации. Для этого непосредственно перед эксплуатацией программный продукт (ПП) должен быть тщательно протестирован пользователями или передан для этого фирме, специализирующейся на тестировании ПП.

Данные действия должны осуществляться независимо от того, прилагаются к ГИПП тесты разработчика или они отсутствуют. В первом случае следует убедиться, что ПП работает и с другими данными, выполнить проверку наиболее сложных и труднореализуемых функций, особенно тех, ошибки в которых могут привести к непредсказуемым последствиям. Во втором случае тестирование, безусловно, необходимо, поскольку только в этом случае можно убедиться в работоспособности ГИПП и его возможностях.

Технология тестирования ГИПП отличается от технологии тестирования ПП на этапе его разработки и имеет свои особенности. Эти особенности подробно были рассмотрены в работе [1]. Среди данных

особенностей можно в первую очередь выделить следующие:

- доступ к коду ГИПП для пользователей и тестеров в подавляющем большинстве случаев исключен. Также отсутствует возможность непосредственного контакта с разработчиками ПП. Поэтому при тестировании используется метод «черного ящика» (BB — Black Box), основанный на системном функциональном или поведенческом тестировании. Кроме того, отсутствуют возможности создания и эксплуатации модульных тестов и тестов проверки межмодульных связей (интеграционных тестов);
- ошибок в программе существенно меньше, чем было на этапе ее разработки. По данным, приводимым в [2], это число составляет примерно 0,1–0,3 ошибки на 1000 строк кода;
- оставшиеся в ГИПП ошибки труднее выявляются, поскольку большинство ошибок было уже найдено при тестировании программы на этапе ее разработки и естественно предположение о том, что оставшиеся ошибки «труднотестируемые» (hard to test [3]). Цена указанных ошибок достаточно велика, существенно выше цены найденных ранее ошибок примерно

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 18-07-00669, 18-29-03100).

¹ МИРЭА — Российский технологический университет, VM.E@mail.ru

² Федеральный исследовательский центр «Информатика и управление» Российской академии наук, VZakharov@ipiran.ru

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, fsergei51@gmail.com

на 1–2 порядка. Поэтому от тестеров требуется повышенное внимание к организации тестирования ГИПП перед началом его эксплуатации.

Обычно тесты — это последовательность шагов, каждый из которых содержит описание действий, предпринимаемых тестером, и описание ожидаемых результатов, используемых при тестировании в качестве эталонов. Разработка тестов ведется тестером на основе изучения работы ГИПП и эксплуатационной документации.

В связи с тем что тестер может смоделировать лишь небольшую часть функций программы (так называемое остаточное моделирование) [4], возможны ошибки при задании как входных данных, так и ожидаемых результатов (эталонных значений), получаемых вручную. Эти ошибки проявляются при реализации процедур сравнения полученных результатов тестирования и эталонных данных и принятия на их основе решений о наличии ошибок в ПП, называемых в специальной литературе оракулами [5]. При этом основную часть этих ошибок составляют ошибки, связанные с заданием эталонов, поскольку их получение осуществляется тестером, как правило, вручную [2].

Кроме того, в связи с тем что входные данные тестов ПП, как правило, представляют собой искусственно подобранные наборы данных, нет гарантии, что они достаточно полно проверяют все особенности ПП. В то же время при выполнении реальных задач с произвольными данными можно обнаружить некоторые отказы в программе, поскольку при их выполнении осуществляется максимальный охват программного кода. Например, при тестировании текстового редактора реальные тексты могут быть не менее полезными, чем специально подобранные буквосочетания. Однако определение ожидаемых результатов при обработке программой реальных данных, особенно при реализации сложных вычислительных процессов, оказывается достаточно трудоемкой процедурой и, как следствие этого, с возможными ошибками.

Таким образом, задача получения правильных эталонных данных при тестировании ГИПП является достаточно сложной и актуальной.

В предлагаемой статье рассматривается ряд принципиальных трудностей практики тестирования, связанных с отсутствием полной спецификации программы и информации о тестах разработчика, об эталонных значениях в частности.

Будет показано, что для программ, реализующих функции, обладающие свойствами случайной самоприводимости (RSR [6]), т. е. функций, вычисление которых на конкретном входном наборе можно свести к вычислению на случайно выбранных вход-

ных наборах, часть трудностей, связанных с отсутствием эталонов, можно преодолеть.

Заметим, что вопрос использования свойств RSR далеко не новый, он широко обсуждался в начале 1990-х гг. в рамках задач самотестирования/самопроверяемости (self-testing/self-checking) программ [7]. Принимая во внимание опыт по практическому тестированию программ [1, 2], авторы считают полезным рассмотреть применение данной модели к проблеме тестирования готовых к использованию программ, особенно в связи с отмеченной проблемой обеспечения тестера информацией об эталонных значениях вычисляемых функций.

2 Основные подходы к тестированию готового к использованию программного продукта

Подходы к тестированию ГИПП зависят в основном от того, для решения каких задач они используются. Возможны следующие ситуации:

- (1) ГИПП используется как продукт, предназначенный для самостоятельного решения конкретной прикладной задачи. В этом случае тестирование состоит в проверке правильности (согласно некоторому протоколу) его функционирования в конкретной операционной среде и в проверке правильности решения задачи на рабочих данных;
- (2) ГИПП представляет собой некоторый пакет программ, используемый для решения различных задач в связанных прикладных областях (например, САПР (система автоматизированного проектирования) для строительно-монтажных работ);
- (3) ГИПП должен использоваться как компонент некоторой сложной разрабатываемой программной или программно-аппаратной системы, в частности в виде сервисов в сервис-ориентированных архитектурах.

Основные современные подходы к разработке тестов для различных вариантов данного списка задач можно разделить на методы «черного ящика» и инъекции ошибок (который может включать в себя и метод «мутаций» кодов и входных данных [8]).

2.1 Метод «черного ящика»

Для тестирования методом ВВ [9] требуется исполняемый компонент, входные данные и «оракул», представляющий собой спецификацию поведения

компонента при разных входных данных, в частности эталонные значения выходов (результатов работы) программы.

Поскольку исходный код ГИПП, как правило, у пользователя отсутствует, единственный источник информации для тестирования — это записи («журналы», «лог-файлы») различных активностей при выполнении программы на целевой платформе на некоторых тестовых входах, на которых значения вычисляемых функций должны играть роль эталонных значений при обнаружении ошибок рассматриваемым тестом. Однако при этом часто возникают указанные выше трудности, связанные с отсутствием информации об истинных эталонных выходных значениях и их связи с конкретным операционным профилем, который может меняться в зависимости от модификации операционной среды или от специфики работы рассматриваемой программы в составе конкретной сервис-ориентированной архитектуры.

Возможным выходом из этой ситуации считают использование исполняемой версии приложения с аналогичной функциональностью, часть результатов которой можно принять за эталонные для новой ГИПП [2, 10]. Это соответствует ситуации, когда существуют прежние версии новой ГИПП или приложения, которые можно рассматривать как подобные рассматриваемой новой ГИПП (с точки зрения результатов вычислений). Часто, однако, такие версии не доступны приобретателю ГИПП.

2.2 Инъекция возможных программных ошибок и системных сбоев

Под инъекцией ошибок понимают введение в код программы на том или ином уровне представления (от исходного до исполнимого кода) искажений, которые могут приводить к тем или иным видам некорректного поведения [8].

Существует множество типов инъекций, которые могут отражать либо некоторые гипотетические сбои в операционной системе (включая вызванные сбоями в аппаратной части системы), либо списки возможных ошибок в исходном коде программы.

Очевидно, что ввиду отсутствия у пользователя ГИПП исходного кода второй способ инъекций интереса не представляет. Реальным приложением инъекций может быть имитация режимов отказа системы, таких как искажение данных операционной среды, повреждение данных, проходящих между вызовами компонент.

В [11] описана система инъекции ошибок при выполнении программы (runtime binary injection tool). Этот инструмент имитирует ошибки в запросах в вызываемых функциях и возвращает значения

выполняемых при этом функций. Общий принцип использования таких инструментов основан на предположении, что тесты, обнаруживающие инжектированные ошибки во время выполнения программы, с высокой вероятностью обнаружат также и ошибки в операторах исходного кода [10], результатом которых может быть, например, неправильная обработка того или иного системного сбоя. Следовательно, входные данные, определенные с использованием инструмента инъекции как тестовые (т. е. обнаруживающие по результатам выполнения программы возможные ошибки), могут использоваться для тестирования ГИПП в эксплуатационных условиях (например, при проведении регламентных работ).

Трудности использования данного подхода связаны прежде всего с высокой стоимостью систем инъекции, а также высокими требованиями к квалификации тестера. При этом, разумеется, остается проблема определения правильных выходов выполняемых функций и программы в целом, используемых в качестве эталонов при оценке результатов тестирования на специфических тестах.

3 О возможности тестирования программ с неизвестной структурой и не полностью специфицированными функциями

Существенной трудностью тестирования ГИПП является отсутствие точных значений выходов программы при конкретных условиях ее выполнения («оракулов», в терминах современной практики тестирования программ [5]). Интересно было бы рассмотреть вопрос о принципиальной возможности решения этой проблемы.

Представляется, что в качестве теоретической и концептуальной основы тестирования при неполной спецификации программы можно было бы взять концепцию тестирования случайных самоприводимых функций, применяемую в теории самотестируемых программ. Поскольку самотестируемость означает, что программа сама может определить отличие правильного результата от неправильного, естественно использовать правильное значение как эталон в задаче внешнего (т. е. с выбором исходных данных самим пользователем программы) тестирования.

Функция является случайной самоприводимой на некотором множестве, если ее значение в данной точке может быть эффективно реконструировано из ее оценки в случайных точках [6, 12].

Определение. Функция $f(x)$, определенная над множеством D , называется случайной самоприводимой функцией, если существует функция φ и множество функций $\sigma_1, \dots, \sigma_n$ таких, что

$$f(x) = \varphi(x, r, f(\sigma_1(x, r)), \dots, f(\sigma_k(x, r))), \quad (1)$$

где случайные переменные $r = \{r_1, \dots, r_m\}$ имеют известные распределения, а функция φ и множество функций $\sigma_1, \dots, \sigma_k$ могут быть вычислены за полиномиальное время.

Иными словами, если для функции $f(x)$ оценить ее значения на любом входе x можно свести за полиномиальное время к оценке ее значения на одном или более случайных экземплярах входных переменных, то такая функция является самоприводимой.

По сути, можно говорить о *восстановлении* значения функции $f(x)$ по случайно выбранным входным переменным $\{r\}$, так как это свойство позволяет восстановить значение функции, используя конечное число элементов, взятых из ее области определения, без каких-либо знаний о реализации программы, которая выполняет вычисление.

Например, пусть программа A должна вычислять линейную функцию $f(x) = wx$, где w — любое действительное число (кроме 0), и вычислениям функций $\sigma_1, \dots, \sigma_4$ соответствуют вызовы программы: $A(w - r_1, x - r_2)$, $A(w - r_1, r_2)$, $A(r_1, x - r_2)$ и $A(r_1, r_2)$, где r_1 и r_2 — равномерно распределенные в области задания переменных w и x случайные величины.

Если программа правильно вычисляет функцию с указанными аргументами, то из приведенного ниже очевидного тождества следует:

$$y_c = A(w - r_1, x - r_2) + A(w - r_1, r_2) + A(r_1, x - r_2) + A(r_1, r_2) = (w - r_1)(x - r_2) + (w - r_1)r_2 + r_1(x - r_2) + r_1r_2 = wx, \quad (2)$$

что демонстрирует выполнение условий определения RSR при любых случайных r_1 и r_2 .

В данном примере функции φ в (1) соответствуют сумма функций $\sigma_1, \dots, \sigma_4$ с соответствующими аргументами.

Если каждый вызов программы выполняется с небольшой вероятностью ошибки $\leq \alpha$ (скажем, 0,01), где α — это доля входных наборов, на которых возможны ошибки вычисления (гарантированные, например, разработчиком программы), то вероятность ошибки вычисления суммы будет $\leq 4\alpha$ и, соответственно, большинство значений y_c при достаточном большом числе случайно выбранных r_1 и r_2 будут равны истинному значению wx , что означает возможность определения правильного значения выхода программы (эталона) как значения, которое получается на большинстве входных

наборов программы, вычисляющей данную функцию.

Таким образом, для реализации данного подхода необходимо выбрать статистически обоснованное число пар r_1, r_2 , чтобы обеспечить требуемую вероятность правильного вычисления с достаточным уровнем доверия (величиной доверительного интервала [12]). Для этого, предполагая, что последовательность правильных результатов образует последовательность Бернулли (поскольку случайные переменные r_1 и r_2 — независимые, равномерно распределенные числа), получим вероятность того, что значение функции, получаемое на большинстве из n входных наборов переменных (т. е. $\geq [n/2] + 1$), есть точное значение функции:

$$\Pr(k \geq L + 1) = 1 - \sum_{k=0}^L C_n^k p^k q^{n-k}, \quad (3)$$

где $p > 1/2$ — вероятность успеха (правильного вычисления функции, $1 - 4\alpha$ в рассматриваемом примере); $q = 1 - p$; k — число правильных ответов в данной последовательности n испытаний, $L = [n/2] + 1$.

Границы этой вероятности определяются различными формами неравенства Чернова (Chernoff inequality), например:

$$\Pr\left(k > \frac{n}{2}\right) \geq 1 - \exp\left(-2n\left(p - \frac{1}{2}\right)^2\right), \quad (4)$$

где k , как и выше, — число успешных исходов в n испытаниях Бернулли.

Очевидно, что с ростом числа испытаний n растет вероятность того, что большинство (больше половины всех исходов) будут успешными (т. е. функция будет вычислена правильно).

Задавая требуемый статистический уровень доверия $1 - \beta$ [12] (например, 0,99, что означает, что вероятность $\Pr(k > n/2)$ на конечной выборке испытаний длиной n в рассматриваемом примере будет оценена с вероятностью ошибки не более 0,01), получают число испытаний Бернулли (т. е. число тестовых прогонов для вычисления функции), необходимых для определения вероятности большинства правильных ответов при данном уровне β [13, 14]:

$$n \geq \frac{\ln(1/\beta)}{(p - 1/2)^2}. \quad (5)$$

Иными словами, n — это число случайных наборов, которые дают достаточную вероятность того, что большая часть результатов вычислений тестируемой функции $f(x)$ будет правильной и можно определить этот результат просто по большинству

ответов при данной вероятности α ошибки выполнения программы A .

Формула (5) применима для оценки необходимого числа запусков программы, вычисляющей любые функции, представимые как (1).

В настоящее время доказано [15, 16], что значительный класс вычислительных задач может быть представлен как вычисление функций со свойствами RSR. Сюда, например, относятся:

- преобразование Фурье;
- матричные вычисления (например, перемножение матриц, построение обратной матрицы, вычисление определителей, вычисление перманента матрицы);
- вычисление тригонометрических, степенных, логарифмических функций;
- операции (умножение-деление) с полиномами нескольких переменных;
- значительное число теоретико-числовых функций, используемых в поисковых алгоритмах и криптографии [15].

Оставляя вопрос об автоматизированном построении функций φ и σ_i ($i = 1, \dots, k$) в формуле (1), отметим, что в настоящее время в литературе приведены различные примеры φ и $\{\sigma_1, \dots, \sigma_k\}$ для всех перечисленных выше RSR-функций, т. е. с практической точки зрения можно говорить о возможности использования «библиотеки» таких функций [7].

Например, в [16] доказана следующая

Теорема. Для каждого набора $\{a_1, \dots, a_{d+1}\}$ попарно различных элементов конечного поля F с числом элементов $|F| > d + 1$ существует конечное число чисел $\{c_1, \dots, c_{d+1}\}$, таких что для каждого полинома $P(x)$ степени d , определенного на F , справедливо

$$\forall x, r \in F, P(x) = \sum_{i=1}^{d+1} c_i P(x + a_i r).$$

В этом случае функции $\sigma_i = x + a_i r$, где r , как и выше, — случайное число. Если $(a_1, \dots, a_{d+1}) = (1, 2, \dots, d + 1)$, то $c_i = (-1)^{i+1} C_{d+1}^i$ — биномиальные коэффициенты.

Существенно, что единственное, что надо знать в данном случае о вычисляемой функции, — это то, что она полином степени d .

Учитывая, что большинство непрерывных функций вычисляется через их полиномиальное представление, можно говорить о широких возможностях использования RSR в задачах тестирования.

Заметим, что, поскольку формально вычисления в любых аппаратных средах являются приближенными (конечность разрядной сетки процессоров), в теории RSR-функций введено также

понятие «приближенной случайной самоприводимости» (approximate-random self-reducible) [13] для функции, значение в некоторой точке x которой может быть представлено по формуле (1) лишь с некоторой точностью ε . С формальной точки зрения это вполне естественно для вычислений с фиксированной и плавающей точкой.

В [13] приведены примеры выполнения (1) при вычислении показательной и логарифмической функций для k -битных чисел с l_m -битной мантиссой и l_{exp} -битной экспонентой.

Соответственно, все сказанное выше о вычислении значений функций тестируемыми программами может быть переформулировано в терминах вычисления приближенных значений.

Итак, если поставщик ГИПП (функциональное назначение которой состоит в вычислении тех или иных функций) гарантирует достаточно высокую вероятность правильной работы программы, вычисляющей требуемую функцию, то, организовав вычисление функции на n случайных наборах из области определения тестируемой функции, где n определяется согласно (5), с высокой вероятностью можно определить неизвестный из спецификации правильный результат вычисления функции.

Очевидно, что при необходимости многократного тестирования ГИПП, например для оценки необходимости дальнейшей модификации программы, или при тестировании в процессе регламентной проверки оборудования нет необходимости прогона на указанных случайных наборах, поскольку достаточно использовать полученные ранее значения выхода программ как эталонные.

4 О самотестировании готового к использованию программного продукта

Как отмечалось во введении, подход, основанный на теории RSR-функций, был первоначально предложен для задач самотестирования и самокоррекции программ. Соответственно, он может быть использован при решении задач самотестирования и ГИПП [13], поскольку для самопроверки нужен лишь перебор достаточного числа входных значений без использования эталонных значений работы программы. Иными словами, можно использовать известные в теории и практике самокоррекции программ методы выбора правильного решения по большинству результатов (например, N-version programming). Более того, ввиду роста вероятности правильного вычисления по мере роста числа n случайных входных наборов $\{r_1, \dots, r_k\}$

(формулы (2)–(5)), можно говорить о повышении надежности программы (этот эффект называют «усилением» (amplification) [17]).

Обычно используют два вида процедур самотестирования программ: следящее и активное [5, 18]. При реализации первого вида самотестирования осуществляется проверка результатов обработки данных в процессе эксплуатации. Примерами следящего самотестирования могут служить подстановка найденных корней в уравнение и оценка разности между левой и правой частью, проверка правильности сортировки путем анализа выходного массива, вычисление и сравнение контрольных сумм и т. д. Главным недостатком данного вида самотестирования ПП является невозможность хранения эталонных данных работы программы, необходимых для анализа результатов ее работы и диагностирования отказов программы.

При активном самотестировании осуществляется проверка результатов обработки специальных тестовых наборов входных данных путем сравнения полученных результатов с хранимыми в программе эталонными данными.

Активное самотестирование может быть полезно при оценке работы ПП с различной аппаратурой или в различных операционных средах, при использовании различных браузеров, при проверке правильности работы программы после обновления некоторых ее функций. Недостатком данного вида самотестирования является то, что работа с ограниченным объемом фиксированных входных данных тестирования не позволяет оценить полностью тестирования.

Самотестирование на основе предлагаемого в статье подхода включает в себя ряд достоинств как первого, так и второго вида самотестирования и в то же время свободно от присущих им недостатков. Прежде всего это обусловлено отсутствием необходимости заранее вычислять ожидаемые выходы, и тем самым исключаются ошибки разработчиков тестов при подготовке эталонных данных для самотестирования.

5 Заключение

Специфика тестирования приобретенных у производителя ПП (ГИПП) широко обсуждается в современных публикациях по программному обеспечению (именуемая в англоязычной литературе как проблема тестирования ПП Ready to Use Software Product (RUSP) [4] или как Commercial off-the-shelf (COTS) [19]).

Одна из основных проблем при этом состоит в отсутствии достаточно подробных спецификаций продаваемых программ, что создает трудности по-

купателю в тестировании этих программ, которое он, например, обязан выполнить перед включением их как компонентов в более сложные системы с повышенными требованиями к надежности (авиационные, медицинские и т. п.).

В данной статье рассмотрена возможность преодоления проблемы отсутствия надежных данных о вычисляемых приобретаемыми программами функциях, используя при этом свойства случайной самоприводимости значительной части функций и дополняя тем самым функциональное тестирование техникой случайного перебора входных наборов, что позволяет вычислять эталонные значения программ с высокой достоверностью.

Данный подход может быть использован для значительного числа вычислительных задач, таких как преобразование Фурье, матричные вычисления, операции с полиномами нескольких переменных, вычисление различных полиномов над конечными полями, широко используемых при вычислении хеш-функций в задачах поиска и хранения информации [15] и функций шифрования, используемых в компьютерных сетях для обеспечения кибербезопасности. Существенно, что свойства случайной самоприводимости обеспечивают также самотестирование и самокоррекцию соответствующих программ, поскольку вычисление правильных значений функций (определяемых по большинству используемых наборов) можно интерпретировать и как подтверждение появления правильных результатов вычисления (тестирование), и как исправление (за счет использования «правила большинства»). А если в составе предлагаемой на рынок программы будет предусмотрен генератор случайных чисел и указанный механизм принятия решения, это может существенно снизить затраты на тестирование после приобретения (например, перед интегрированием в более сложные системы) и, соответственно, повысить их привлекательность для потенциального покупателя.

Литература

1. Басок Б. М., Головин С. А., Захаров В. Н., Френкель С. Л. Тестирование готового к использованию программного продукта // ИТ-Стандарт: Электронный научный журнал, 2018. № 1. 7 с. http://journal.tc22.ru/wp-content/uploads/2018/05/testirovanie_gotovogo_k_ispolzovaniyu_programmnogo_produkta.pdf
2. Лунаев В. В. Тестирование компонентов и комплексов программ. — Москва—Берлин: Директ-Медиа, 2015. 528 с.
3. 7 Types of software errors, that every tester should know // Software Testing Help, 2018. www.softwaretestinghelp.com/types-of-software-errors.

4. ГОСТ Р ИСО/МЭК 25051-2017. Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuARE). Требования к качеству готового к использованию программного продукта (RUSP) и инструкции по тестированию. — М.: Стандартинформ, 2017. 32 с.
5. Barr E. T., Harman M., McMinn P., Shahbaz M., Yoo S. The oracle problem in software testing: A survey // IEEE T. Software Eng., 2015. Vol. 41. No. 5. P. 507–525.
6. Lipton R. New directions in testing // Distributed computing and cryptography / Eds. J. Feigenbaum, M. J. Merritt. — DIMACS ser. in discrete mathematics and theoretical computer science. — AMS, 1991. Vol. 2. P. 191–202.
7. Blum M., Luby M., Rubinfeld R. Self-testing/correcting with applications to numerical problems // 22nd ACM Symposium on Theory of Computing Proceedings. — New York, NY, USA: ACM Press, 1990. P. 73–83.
8. Natella R., Cotroneo D., Duraes J. A., Madeira H. On fault representativeness of software fault injection // IEEE T. Software Eng., 2013. Vol. 39. No. 1. P. 80–96.
9. Canfora G., Di Penta M. Testing services and service-centric systems: Challenges and opportunities // IT Prof., 2006. Vol. 8. No. 2. P. 10–17.
10. Buck D., Hollingsworth J. An API for runtime code patching // Int. J. High Perform. C., 2000. Vol. 14. No. 4. P. 317–329.
11. Barrantes E. G., Ackley D. H., Forrest S., Palmer T. S., Stefanovic D., Zovi D. D. Randomized instruction set emulation to disrupt binary code injection attacks // 10th ACM Conference on Computer and Communications Security Proceedings. — New York, NY, USA: ACM Press, 2003. P. 281–289.
12. Смирнов Н. В., Душин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. — Л.: Наука, 1969. 512 с.
13. Gemmell P., Lipton R., Rubinfeld R., Sudan M., Wigderson A. Self-testing/correcting for polynomials and for approximate functions // 23rd ACM Symposium on the Theory of Computing Proceedings. — New York, NY, USA: ACM Press, 1991. P. 32–43.
14. Bhattacharyya A., Dey P. Sample complexity for winner prediction in elections // arXiv.org, 2016. arXiv: 1502.04354 [cs.DS].
15. Carter L., Wegman M. Universal hash functions // J. Comput. Syst. Sci., 1979. Vol. 18. P. 143–154.
16. Nouber G., Nussbauer H. Self-correcting polynomial programs // Reliab. Comput., 1996. Vol. 2. No. 2. P. 139–145.
17. Dolev Sh., Frenkel S. Extending the scope of self-correcting // 13th Conference (International) on Applied Stochastic Models and Data Analysis Proceedings. P. 458–462.
18. Басок Б. М., Красовский В. Е. Тестирование программного обеспечения. — М.: МИРЭА, 2010. 120 с.
19. Voas J., Charron F., Miller K. Robust software interfaces: Can COTS-based systems be trusted without them? // 15th Conference (International) on Computer Safety, Reliability and Security Proceedings. — Vienna: Springer Verlag, 1996. P. 126–135.

Поступила в редакцию 10.06.18

USING A PROBABILISTIC CALCULATION MODEL TO TEST ONE CLASS OF READY-TO-USE SOFTWARE COMPONENTS OF LOCAL AND NETWORK SYSTEMS

B. M. Basok¹, V. N. Zakharov², and S. L. Frenkel³

¹MIREA — Russian Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation

²Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper discusses and analyzes the possibility of providing effective testing of ready-to-use software products that solve the task of calculating functions, in the absence of complete information necessary for traditional testing. Efficiency means the possibility of providing an arbitrarily high probability of detecting possible computational errors that were not detected by the output control as the number of inspections increases. As a conceptual model of the proposed approach, the properties of functions with the random self-reducible function are used, that is, functions whose calculation on a particular input set can be reduced to calculation on several randomly selected input sets. The rationality of providing self-testability properties in ready-to-use software is substantiated.

Keywords: software testing; self-testing

DOI: 10.14357/19922264180407

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (projects 18-07-00669 and 18-29-03100).

References

1. Basok, B. M., S. A. Golovin, V. N. Zakharov, and S. L. Frenkel. 2018. Testirovanie gotovogo k ispol'zovaniyu programmogo produkta [Testing of ready-to-use software product]. *It-standart: Elektronnyy nauchnyy zh.* [IT-Standart: Electronic Scientific J.] 1. 7 p. Available at: http://journal.tc22.ru/wp-content/uploads/2018/05/testirovanie_gotovogo_k_ispolzovaniyu_programmnogo_produkta.pdf (accessed October 30, 2018).
2. Lipaev, V. V. 2015. *Testirovanie komponentov i kompleksov program* [Testing of components and software packages]. Moscow—Berlin: Direkt-Media. 528 p.
3. 7 Types of software errors, that every tester should know. *Software Testing Help*. Available at: www.softwetestinghelp.com/types-of-software-errors (accessed October 30, 2018).
4. GOST R ISO/MEK 25051-2017. Informatsionnyye tekhnologii. Sistemnaya i programmaya inzheneriya. Trebovaniya i otsenka kachestva sistem i programmogo obespecheniya (SQuaRE). Trebovaniya k kachestvu gotovogo k ispol'zovaniyu programmogo produkta (RUSP) i instruktsii po testirovaniyu [Information technology. System and software engineering. Requirements and quality assessment of systems and software (SQuaRE). Requirements for the quality of ready-to-use software product (RUSP) and instructions for testing]. Moscow: Standardinform Pubs. 32 p.
5. Barr, E. T., M. Harman, P. McMinn, M. Shahbaz, and S. Yoo. 2015. The oracle problem in software testing: A survey. *IEEE T. Software Eng.* 41(5):507–525.
6. Lipton, R. 1991. New directions in testing. *Distributed computing and cryptography*. Eds. J. Feigenbaum and M. J. Merritt. DIMACS ser. in discrete mathematics and theoretical computer science. AMS. 2:191–202.
7. Blum, M., M. Luby, and R. Rubinfeld. 1990. Self-testing/correcting with applications to numerical problems. *22nd ACM Symposium on Theory of Computing Proceedings*. New York, NY: ACM Press. 73–83.
8. Natella, R., D. Cotroneo, J. A. Duraes, and H. Madeira. 2013. On fault representativeness of software fault injection. *IEEE T. Software Eng.* 39(1):80–96.
9. Canfora, G. and V. Di Penta. 2006. Testing services and service-centric systems: Challenges and opportunities. *IT Prof.* 8(2):10–17.
10. Buck, D., and J. Hollingsworth. 2000. An API for runtime code patching. *Int. J. High Perform. C.* 14(4):317–329.
11. Barrantes, E. G., D. Y. Ackley, T. S. Palmer, D. Stefanovic, and D. Zovi. 2003. Randomized instruction set emulation to disrupt binary code injection attacks. *10th ACM Conference on Computer and Communications Security Proceedings*. New York, NY: ACM Press. 281–289.
12. Smirnov, N. V., and I. V. Dunin-Barkovskiy. 1969. *Kurs teorii veroyatnostey i matematicheskoy statistiki dlya tekhnicheskikh prilozheniy* [Course in the probabilities theory and mathematical statistics for technical applications]. Leningrad: Nauka. 512 p.
13. Gemmell, P., R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson. 1991. Self-testing/correcting for polynomials and for approximate functions. *23rd ACM Symposium on the Theory of Computing Proceedings*. New York, NY: ACM Press. 32–43.
14. Bhattacharyya, A., and P. Dey. 2016. Sample complexity for winner prediction in elections. *arXiv.org*. arXiv: 1502.04354 [cs.DS].
15. Carter, L., and M. Wegman. 1979. Universal hash functions. *J. Comput. Syst. Sci.* 18:143–154.
16. Noubert, G., and H. Nussbauer. 1996. Self-correcting polynomial programs. *Reliab. Comput.* 2(2):139–145.
17. Dolev, Sh., and S. Frenkel. 2009. Extending the scope of self-correcting. *13th Conference (International) on Applied Stochastic Models and Data Analysis Proceedings*. 458–462.
18. Basok, B. M., and V. E. Krasovskiy. 2010. Testirovanie programmogo obespecheniya [Software testing]. Moscow: MIREA. 120 p.
19. Voas J., F. Charron, and K. Miller. 1996. Robust software interfaces: Can COTS-based systems be trusted without them? *15th Conference (International) on Computer Safety, Reliability and Security Proceedings*. Vienna: Springer Verlag. 126–135.

Received September 20, 2018

Contributors

Basok Boris M. (b. 1948) — Candidate of Science (PhD) in technology, associate professor, MIREA — Russian Technological University, 78 Vernadskogo Ave., Moscow 119454, Russian Federation; VM_E@mail.ru

Zakharov Victor N. (b. 1948) — Doctor of Science in technology, associate professor; Scientific Secretary, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vzakharov@ipiran.ru

Frenkel Sergey L. (b. 1951) — Candidate of Science (PhD) in technology, associate professor, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; fsergei51@gmail.com

SEAMLESS ROUTE UPDATES IN SOFTWARE-DEFINED NETWORKING VIA QUALITY OF SERVICE COMPLIANCE VERIFICATION

S. L. Frenkel¹ and D. Khankin²

Abstract: In software-defined networking (SDN), the control plane and the data plane are decoupled. This allows high flexibility by providing abstractions for network management applications and being directly programmable. However, reconfiguration and updates of a network are sometimes inevitable due to topology changes, maintenance, or failures. In the scenario, a current route C and a set of possible new routes $\{N_i\}$, where one of the new routes is required to replace the current route, are given. There is a chance that a new route N_i is longer than a different new route N_j , but N_i is a more reliable one and it will update faster or perform better after the update in terms of quality of service (QoS) demands. Taking into account the random nature of the network functioning, the present authors supplement the recently proposed algorithm by Delaet *et al.* for route updates with a technique based on Markov chains (MCs). As such, an enhanced algorithm for complying QoS demands during route updates is proposed in a seamless fashion. First, an extension to the update algorithm of Delaet *et al.* that describes the transmission of packets through a chosen route and compares the update process for all possible alternative routes is suggested. Second, several methods for choosing a combination of preferred subparts of new routes, resulting in an optimal, in the sense of QoS compliance, new route is provided.

Keywords: software-defined networking; Markov chains; quality of service

DOI: 10.14357/19922264180408

1 Introduction

Software-defined networking is an emerging network paradigm, in which the control plane is decoupled from the data plane enabling centralized control logic. Such a dynamic network may require frequent modifications and updates to the network topology and configuration. Also, the network topology is available to the centralized control entity, there, due to this flexibility, it is possible to perform offline optimized calculations.

Network functions virtualization (NFV) allows replacing traditional network devices with software that is running on commodity servers. This software implements the functionality that was previously provided by dedicated hardware. Network functions virtualization allows services to be composed of virtual network functions (VNF) hosted on different data centers. Software-defined networking, when applied to NFV, helps in addressing challenges of dynamic resource management and intelligent service orchestration [1]. Sometimes, traffic is often required to pass through and be processed by an ordered sequence of possibly remote VNFs [2]. For example, traffic may be required to pass through intrusion detection system, proxy, load balancer, or a firewall.

Such concatenation of services is called *service function chaining* (SFC).

Consider, for example, two communicating parties in a network featuring complex network topology (e. g., Small-world network), and the communication flow is passed over a series of VNFs. It may be the case that the network operator is required to move the communicating flow to a different path due to QoS requirements or other possible cost considerations. We are interested to model the anticipated expected number of steps until the update is complete given a possible new route following the required QoS demands, e. g., delay, communication rounds, cost, etc.

Let us consider a pair $(C, \{N_i\})$ where a current route C from s to d is scheduled to be replaced by a new route from the set $\{N_i\}$, each from s to d either. Let us model each route as an ordered list of network elements, such as VNFs (SFCs) or generally saying routers. Each new route N_i is constructed during the update process, and thus, certain delays may be introduced due to initial packet processing or due to possible losses.

The design goals must be achieved by constructing effective algorithms for efficient packet QoS routing in NFV/SDN computer network. Depending on the QoS

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation, fsergei51@gmail.com

²Computer Science Department, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel, danielkh@post.bgu.ac.il

metric, the lower (e. g., for reliability) or upper (e. g., for a delay) constraints represent the desired bounds that the orchestration must meet. Since different configurations could meet these bounds, the designer should also optimize against a specific metric by using these both ends of the extreme.

Methods based on integer linear programming (ILP) were proposed in several works (see section 2). The difficulty of using tools based on ILP in the operational work of an administrator is that in view of the possible infeasibility of the resulting solution, it may take not a few resources (time, efforts) until acceptable QoS values can be ensured.

We consider the use of “design via verification” approach, suggesting a method for complying QoS demands. The method is based on augmenting the update algorithm with a verification logic. Namely, we suggest the use of *Probabilistic real-time Computation Tree Logic* (PCTL) [3] for expressing real-time and probability in systems. Using PCTL, we can express the probability for a process to complete after a certain number of steps along an execution path and verify the selected route for the update.

Delaet *et al.* proposed a multicast-based scheme for seamlessly updating a current route to a new one [4]. According to the multicast scheme, the controller instructs a router to temporarily have two (s, d) entries in the routing table. When a router $r \neq d$ receives a packet from s to d , it sends the packet according to the forwarding instructions of all of its (s, d) routing table entries. When two identical copies of a packet that was multicast over the current and new portion of a route arrive, the controller can dismantle the current route, as the new route is ready. During the update process, packets should not be lost, no cycles should be formed, and communication should not be disrupted.

Our contribution is a model for a successful route update, including its intermediate steps, as MC states, each with a given probability. With our model, we are able to characterize the quality of an update by expected number of steps in the MC.

We suggest an enhanced update method for the network administrator to augment his decision regarding QoS demands in terms of various network parameters and possible failure of the update process. Moreover, in contrast to other works, we are able to provide a version of an algorithm that can perform real-time QoS assessment during a route update, for each subpart of a route. At last, using our method, it is possible that the active new route will be comprised of subparts of different new routes, providing optimal route update service in regard of required network QoS.

Extended abstract of this work appeared as a conference paper in [5] which presented preliminary results. In this work, we describe in detail the system settings

and bring new results by providing two additional algorithms.

In the following section, we overview the related work. Next, we provide the required definitions and the system settings and describe the MC characterization of the network. Further, we describe different update setting, accordingly accompanying algorithms and data structures, used for QoS assessment during route updates.

2 Related Work

Quality of service routing using multipath was proposed in [6]. The routing algorithm, initially, eliminates all links that do not meet the bandwidth requirements. Then, it finds disjoint shortest paths based on the residual network graph in each iteration.

The work [7] proposed a QoS optimized routing over multidomain OpenFlow networks managed by a distributed control plane, where each controller performs optimal routing within its domain. The QoS routing problem was posed as a constrained shortest path (CSP) problem, and the proposed solution computes a near-optimal route, based on LARAC (Lagrange relaxation based aggregated cost) algorithm [8]. The proposed algorithm is an approximation algorithm; it always gives a suboptimal solution.

For traditional network architecture, a routing strategy approach based on ILP was introduced in [9]. The main disadvantage of using ILP is that the problem is NP-hard. Additionally, ILP cannot be applied to probabilistic values. Using linear programming (not limited to integers) rounded to integer solutions will not yield an optimal solution.

Route updates are extensively researched in SDN [10], standing on the work by Reitblatt *et al.* where requirements for SDN updates were examined. This work focused on per-packet consistency property, stating that packets have to be forwarded either using the initial configuration or the final configuration but never a mixture of them, throughout the update process [11]. The authors proposed a 2-phase commit technique which relies on packets tagging so that either of the rules is applied. However, such technique wastes critical network resources and complications are formed due to packet tagging [10]. Further, Delaet *et al.* showed in [4] that using a careful multicast during route updates provides a better working solution.

Hogan and Esposito propose in [12] the use of Bayesian networks for delay estimation as a traffic engineering tool and model the path selection problem using a risk minimization technique. However, the authors state that the accuracy of their model is limited by its ability to correctly identify dependencies in the data. In our work, we suggest a general tool for probabilistic

verification of any network parameter, which does not depend on variance within the dataset.

In [13], an update protocol proposed where packets are sent to the controller during updates; such approach adds a significant cost to the control plane bandwidth [4]. In [14], an algorithm to find a safe update sequence expressed as a logic circuit has been proposed. However, the algorithm requires a dedicated protocol which is not currently supported [10]. The authors of [15] propose to perform the 2-phase update scheme from [11] incrementally, making the update longer.

Software-defined networking allows the involvement of the network administrator into the network management during route updates and, in particular, during packet transmission. Thus, it would be highly desirable to support the decision making process with the right tools. Our novelty is exactly such tool, for augmenting online decision making of the network administrator during network management in a stochastic environment.

The work by Delaet *et al.* [4] introduced the Make&Activate-Before-Break approach for seamless route update in SDN. The authors described in a high-level the multicasting-based update, which we employ in this work. Also, they introduced a controller-based method for verifying the correctness of a new route before the traffic redirection. Dinitz *et al.* [16] extended the work [4] to the general case of several dependent (via shared links) routes pairs. The routes update problem was proved to be NP-hard [17]. The authors of [16] suggested the use of artificial intelligence (AI) methods for solving the problem. As a basis for AI-based solutions, Dinitz *et al.* proposed a dependence graph model describing the current state of the problem instance at any replacement stage. In addition, route readiness verification similar to that in [4] was implemented in [16] as a high-level network protocol.

In this work, we investigate a different problem; we consider the route updates problem from a QoS perspective and describe in high-level both the prediction and the update processes.

3 Preliminaries and Definitions

The basic system settings are as follows. For a (route) sequence X , we denote by x_i the i th element in it. In a (directed) communication network, we are given a route C from source s to destination d . Additionally, we are given a set of different new routes N_i , each going from s to d . We model each route as an ordered set of network nodes connected by network links. We assume that neither of the routes contains cycles. Each router in a route matches a packet from s to d and forwards the packet to the next router in order. After the update is complete, each router in the new route should forward

the packets from s to d to the next router in order along the new route.

In our work, we consider the route replacement problem as a sequence of subroutes replacements. The routes replacement subsystem was in great detail described by Dinitz *et al.* in [16]. We borrow from [16] the relevant parts which we briefly describe here.

Definition 1. We define a subset from $a \in X$ to $b \in X$ of an ordered set X , when a precedes b , as a subroute from a to b , and denote such subroute by $[a, b]$.

Subroutes. The current route C subdivides each new route to k common subroutes (a subroute may consist of one router in the simplest case) and $k - 1$ noncommon subroutes. For illustration, see Fig. 1. In Fig. 1 and figures below, the current route is depicted in a light grey color full nodes, connected with solid edges. The new route is depicted in white colored nodes, connected with dashed edges. The common nodes are depicted as shaded. If there are several new routes, the nodes of each route are filled with a designating pattern. Additionally, for easier reading, when it is possible, we denote subroutes of some route X as X' , X'' , etc. In other cases, a subroute j of a new (current) route i is denoted as $N_j^i(C_i^j)$. Similarly, routers of some route X are denoted by r' , r'' , etc.

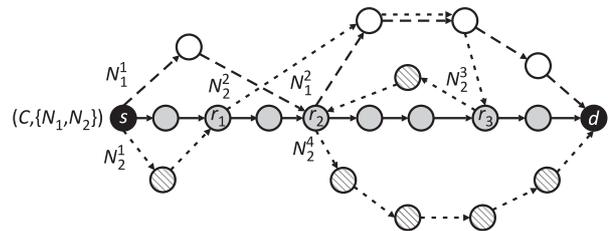


Figure 1 Route C with two possible new routes sharing a link

In the example in Fig. 1, noncommon new subroutes of route N_1 are denoted by $N_1^1 = [s, r_2]$ and $N_1^2 = [r_2, d]$, while the noncommon new subroutes of N_2 are denoted by $N_2^1 = [s, r_1]$, $N_2^2 = [r_1, r_3]$, $N_2^3 = [r_3, r_2]$, and $N_2^4 = [r_2, d]$.

Note that in general, the order of common subroutes along C and along N can be different. See, for example, the common subroutes of C and N_2 in Fig. 1.

Definition 2. A new noncommon subroute of N from router a to router b is legitimate for update only if a precedes b on the route C .

Definition 2 guides us on which subroutes can be launched without creating routing cycles in the network system. (See [4] for details.)

When an update of a subroute N' from router r to r' is finished, the update flow goes along C from s to r , continues along N' up to r' , and finishes along C from r' to d . For illustration, see the result of launching N_2^4 in Fig. 2.

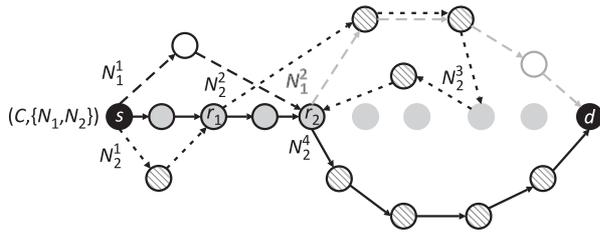


Figure 2 N_2^4 was launched

Note that launching a currently nonlegitimate new subroute, for example, N_2^3 in Fig. 1, is forbidden since it will form a cycle resulting in packets circulating and overwhelming the network.

Dynamics of the system. Dinitz *et al.* performed a detailed analysis on the dynamics of a subroutes system. After an update of a subroute is complete, the set of current subroutes C and the set of new subroutes N are recalculated. This may result in different system of subroutes. For example, see Fig. 2 where after the launch of N_2^4 from the example in Fig. 1, the sets of subroutes are recalculated. As a result, we obtain different subroutes (for clarity, the previous labels are kept). See also [16] for details and extensive analysis.

3.1 Markov chain characterization of the network states

We characterize execution of some (sub)route in the network by a packet delay time between the (sub)route's common sender and common destination routers as well the probability of a packet drop. Let us for now define our network routing model (conceptual model) informally in the following terms. Delay of a packet is obtained using a physical delay and the total processing time in the router. We consider that transmission of packets in a network can have a random behavior, caused by the random character of both, the input, and possible loss of packets. There we are interested in a probabilistic model, namely, a Markov model. In order to fully characterize the network as an MC, the internal state of each router (and, in particular, the buffer occupancies), as well as the characteristics of all flows, need to be expressed as states in the chain.

However, such approach would result in an enormous and intractable number of states. Therefore, to simplify these computations, let us characterize the delay time as an abstract variable t . This abstract variable can be interpreted in different ways, e.g., the current processing queue length and a packet transmission rate of the link, or possibly a fixed value, such as an interval between the beginning of a packet transmission after being processed in some node and the end of processing at the next node.

We describe the functioning of the network in the transmission of packets as transitions of a discrete-time MC (DTMC). The state space corresponds to the set of nodes such that the transmission of a packet from a node that has finished processing the packet to the next node corresponds to the transition of the chain to the next state.

Discrete-time MC is defined as a tuple $D = (S, s_0, P)$. In the tuple, S is the finite set of states, $s_0 \in S$ is the initial state, $P : S \times S \rightarrow [0, 1]$ is the transition probability matrix in which $\forall s \in S, \sum_{s' \in S} P(s, s') = 1$. For any two states $s, s' \in S$, if $P(s, s') > 0$, then s' is the successor of s . For a subset of states $T \subseteq S$, the probability of moving from a state s to any state $t \in T$ in a single step is denoted by $P(s, T)$ and is given by $P(s, T) = \sum_{t \in T} P(s, t)$.

3.2 Verification syntax

For implementation of our PCTL-based model, we use PRISM — probabilistic model checker [18]. There, we follow PRISM property specification language. Here, we briefly describe the essential syntax while more details can be found in [19].

Given a property Ψ , we say that Ψ is true with probability p and write that as $P_p[\Psi]$. If the probability p is unknown, PRISM allows, for DTMC, writing properties queries of the form $P_{=?}[\Psi]$, meaning “what is the probability that Ψ is true?”. Additionally, it is possible to use a time bound and write properties queries such as $P_{=?}[F^{\leq T} \Psi]$, meaning “what is the probability that Ψ is true after less than T steps?”. At last, it is possible to compute properties such as expected time or expected number of steps. For example, $R_{=?}[F \Psi]$, meaning “what is the expected number of steps until Ψ is true?”.

4 Prediction of Preferred Update

The states of a DTMC describe the nodes in the new route and the transition probabilities in the chain represent the possible delay or a packet loss in the routers along the new route. The states are defined as $\{s_1, \dots, s_n\}$ where n is the number of nodes in the new route. The network achieves the state s_i if a packet has reached the i th node. For example, in Fig. 3, the self-transition edge represents the probability for a delay due to packet loss, rules installation at the router, or congestion on the router-controller link, while the forward transition edge represents the probability for a successful transition to the next state. These probabilities can be estimated from network statistics (see, for example, [12]). The labels on edges are the probability values, when edge has no label means probability 1.

The initial probability distribution of states is given by the vector P_0 of size n . We can determine the prob-

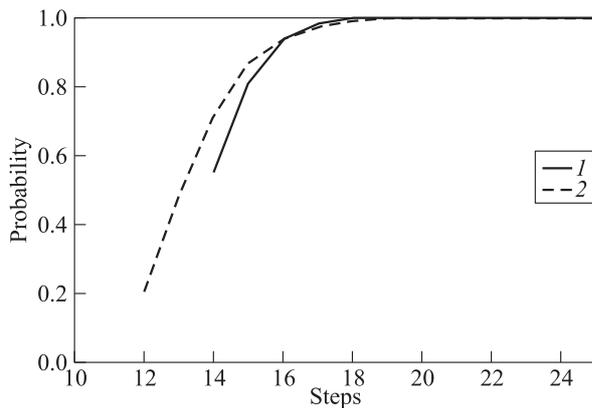


Figure 3 Probability as a function of number of steps to update routes N_1 (1) and N_2 (2)

ability that a particular route delays the update process by k , that is, the number of steps required for a successful update is given by $p(k) = P_0 P^k$. Using this characteristic, which is, in fact, the probability distribution of the number of steps $P(k < x)$, one can calculate various properties like average delay time for the new route, maximum or minimum number of steps to update, etc.

Consider the example illustrated in Fig. 4. Figure 4a illustrates the current route C and a candidate new route N_1 . Figure 4b shows the same current route C with another candidate new route N_2 . Figures 4c and 4d show the MCs for new routes N_1 and N_2 , accordingly, with given transition probabilities.

During the update process, packets are sent along the current and the new routes. Since the new route is

not operational yet, packets can be delayed due to congestion on certain nodes or due to switch configurations. For example, if routing rules have not yet been installed in some switch, then an arriving packet is sent to the controller [20]. The controller then decides reactively on further actions whether to install an appropriate rule for the packet. Also, the controller may be busy with other work and not respond immediately. Those packet processing actions may delay the update process. In the case buffer becomes full, for example, if the network is being congested, packets may be dropped. There, the transition to the next state during the update process depends on the likelihood of a delay or a loss of a packet in the current state.

In the example, the number of steps required for launching N_2 is smaller than the number of steps required for launching N_1 . However, due to a higher likelihood of delays along the route N_2 , it is possible that N_1 is preferred having a higher probability for a successful update. The network administrator may ask which new route is recommended for the update process, considering the expected number of steps required for the update. That is, updating paths requires the operator to decide on the possible choice of a subroute for the next step. One should consider the possibility of including a decision tool augmenting the controller during route updates.

There were many attempts to use the LP/ILP approach, as it was already mentioned above (see, e. g., [8]), but they have encountered the same difficulties, especially when taking into account online implementation. We show that it is possible to describe the routing process

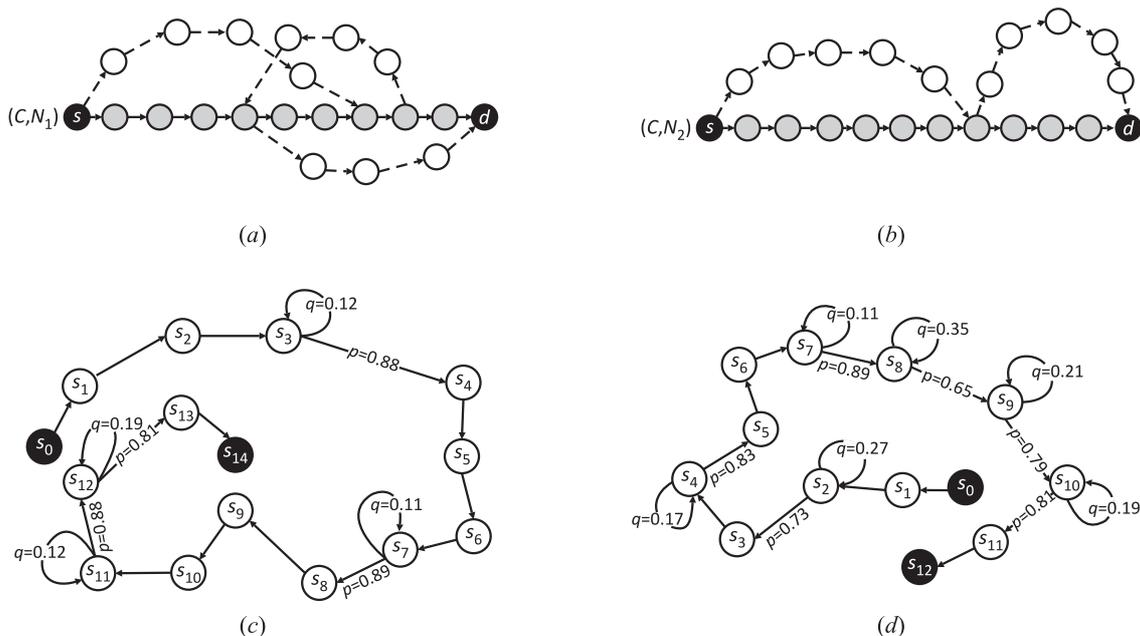


Figure 4 New routes N_1 (a) and N_2 (b) and MC states for N_1 (c) and N_2 (d)

as DTMC. Thus, taking into consideration $O(n^3)$ worst case computation complexity, we consider using the “design via verification” mentioned above based on PCTL verification, similar to the one used in PRISM [18].

We have calculated the probability for a successful update as a function of number of steps for routes N_1 and N_2 from the example in Fig. 4. See Fig. 3 where this function is shown. Curve 1 represents the plot for N_1 and curve 2 represents the plot for N_2 .

Observe that after 20 steps, both new routes will be launched with probability 1 which can be written as

$$P_1 [F^{>20} N_1] = P_1 [F^{>20} N_2] = 1 .$$

The expected number of steps required for N_1 is smaller than the required for N_2 :

$$R [F N_1] < R [F N_2] .$$

However, the probability for successfully updating in less than 15 steps is higher for route N_2 (0.55 ± 0.040 for N_1 and 0.717 ± 0.036 for N_2 , based on 99% confidence level): $P_{0.717 \pm 0.036} [F^{\leq 15} N_2]$.

5 Route Updates per Quality of Service

In this section, we show algorithm that we propose for various settings. First, we show an enhancement for the sequential update algorithm from [4], which during the update process decides on preferred subroute from the set of possible subroutes as part of QoS requirements. In the multicast-based update, several methods were proposed in [4] for eliminating duplicated packets. In the case the common destination router is not able to immediately eliminate duplicated packets, the algorithm begins the update from the end, ensuring a correct update process [4].

After that, we show an algorithm that chooses the subroutes for update arbitrary, assuming that the common destination node will not leak duplicated packets. However, the packets sending rate along the new subroute need to be temporarily limited [4].

At last, we present a supplementing algorithm that suggests which subroutes can be updated in parallel.

5.1 Sequential update

Let us begin the update from the end, namely, from the last alternative subroute of any new route. Provably, this prevents the formation of cycles [4]. In order to represent all possible choices of a path from a current state of the update process to the end of the update process, we propose to use a directed graph which nodes are the new, legitimate for launching, subroutes of the

network. The edges of the graph represent a legal order of launching new subroutes. Each path in this graph from a current node to the last node in the path represents a legal combination of chosen subroutes. The update process is continued as long as there is a possible node to transition to.

Let us examine the two possible new routes N_1 and N_2 that can replace the current route C from the example depicted in Fig. 1. The new route N_1 is composed of N_1^1 and N_1^2 , while the new route N_2 composed of N_2^1 , N_2^2 , N_2^3 , and N_2^4 . Starting from the end, the only new subroutes that are allowable to launch are N_1^2 and N_2^4 . Assume that based on the DTMC calculations performed as described in section 4, the subroute N_2^4 is chosen for update. After the update of the subroute is complete, the current route C is composed of not updated yet part of the old route and N_2^4 . See Fig. 2 where the change in C is depicted.

After the subroute N_2^4 is launched, we arrive at a smaller problem in which less subroutes are left to update. Due to dynamics of the system (see section 3), some new subroutes can merge into a single new subroute. See Fig. 2 where after N_2^4 was launched, the new subroutes N_2^3 and N_2^2 are merged into a single subroute. Now, one can launch either N_1^1 or N_2^2 merged with N_2^3 . Assume that we choose to launch N_1^1 , which launch finishes the update. The route C updated to N_1^1 and N_2^4 . See Fig. 5 illustrating that.

Figure 6 shows the directed graph that represents the possible update sequences. Initially, the subroutes that are legal for launch are N_1^2 and N_2^4 . As such, these are the only subroutes that have in-degree 0. Launching N_2^3 is forbidden; hence, there is no node in the graph G that represents this subroute. After launching N_2^4 , we

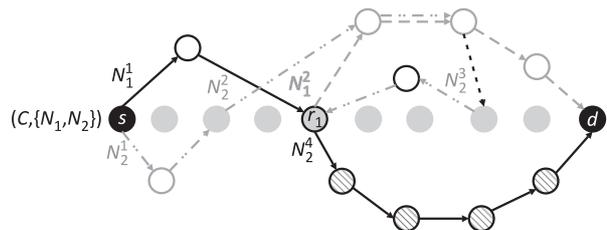


Figure 5 N_1^1 was launched

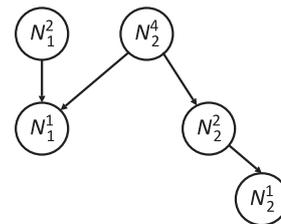


Figure 6 Graph representation for possible update paths for routes update example from Fig. 1

Algorithm 1: Update per QoS Algorithm

```

1 directed graph  $G$ 
  /*  $A$  is a collection of nodes */
2  $A \leftarrow$  choose nodes from  $G$  with in-degree 0
3 repeat
4   foreach  $v \in A$  do
5     calculate  $R[F v]$ 
6   end
7    $N_i^t \leftarrow \arg \max_v (R[F v])$ 
8   launch  $N_i^t$ 
9   update  $C$  accordingly
10  merge any new and common subroutes as described in section 3
11   $A \leftarrow$  choose nodes neighboring to  $N_i^t$ 
12 until out-degree of node  $N_i^t > 0$ ;

```

can proceed by launching N_1^1 or N_2^2 . However, if N_1^2 was launched first, it would be forbidden to launch N_2^2 since it shares a common edge with N_1^2 . This is reflected in the graph G by not having a directed edge from the node N_1^2 to the node N_2^2 . We finish the update process by arriving either to N_1^1 or to N_2^1 . Notably, these nodes have out-degree 0.

Algorithm 1 updates subroutes according to calculated QoS for each new subroute, by choosing at each step the new subroute that maximizes QoS.

The algorithm starts by selecting the initial set of subroute nodes. These are nodes with in-degree 0. The algorithm continues traversing the graph up to arrival at a node with out-degree 0 which would be the last subroute to launch. The inner loop at lines 4–6 calculates the QoS for each neighboring node. Afterward, at line 7, the algorithm chooses the node that maximizes QoS. Then launches this node and updates the route C , accordingly (see Figs. 1–5 for illustration). Afterward, the algorithm selects the next neighboring nodes.

After execution of Algorithm 1, the resulting new route maximally complies QoS requirements.

5.2 Arbitrary subroutes selection

In this subsection, we assume that immediate duplicate packets elimination is possible. It may be that some of the subroutes are not ready for an update yet. Thus, meanwhile, the administrator may want to proceed with the update process to other subroutes or see possible variations of the update. For such scenario, we provide an algorithm which can select a subroute for update arbitrary and continue the update process from there. We create a forest graph of all possible update combinations from which the desired update sequence can be chosen.

Figure 7 shows all possible combinations from example in Fig. 1. Noticeable, as mentioned earlier, some

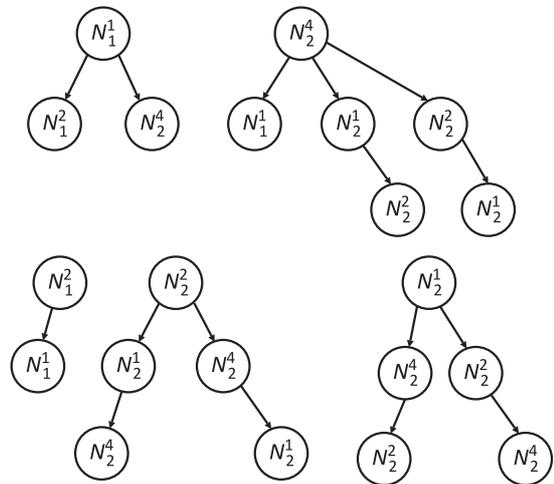


Figure 7 Forest graph representing execution combinations for example from Fig. 1

combinations exhibit fewer steps, though possible that its QoS compliance is worse than others.

Algorithm 2 starts by iterating over all roots of the forest graph and calculating QoS using Algorithm 1 each tree. Afterward, launch the update of the tree that maximizes QoS.

5.3 Parallel update

In certain cases, it is possible to update in parallel several subroutes and, as such, decrease update time. However, launching subroutes in parallel is not always possible since subroute may share a link and, thus, leads to congestion during the update process, close a cycle, or lead to an inconsistent state of the system. In [4], it was shown that two new subroutes N' from a to b and N'' from c to d can be launched in parallel only if c succeeds b or a succeeds d .

Algorithm 2: Arbitrary Selection Update

```

1 directed graph  $G$ 
2  $A_0 \leftarrow$  choose nodes from  $G$  with in-degree 0
3  $Q \leftarrow \{\}$ 
  /* iterate over all roots of trees in the forest  $G$  */
4 foreach  $v_r \in A_0$  do
5    $q \leftarrow$  get the expected QoS using Algorithm 1 for  $v_r$ 
6    $Q \leftarrow Q \cup \{q \rightarrow \text{root}\}$ 
7 end
8  $q_{\max} \leftarrow \max_{\text{QoS}}(Q)$ 
9 launch maximizing QoS update order in root =  $Q[q_{\max}]$ 

```

We create a supplementary graph G_S , in which nodes are the new legitimate for launching subroutes, and edges represent restrictions on parallel launching of subroutes. See Fig. 8 for illustration, depicting subroutes from example in Fig. 1 and their parallel restrictions. For example, N_2^4 and N_2^1 can be launched in parallel since there is no edge connecting them.

Clearly, any independent set of subroutes from the supplementary graph contains subroutes that can be launched in parallel. This can be further enhanced by setting QoS calculated values as weights on nodes of the graph and finding the subroutes that can be launched in parallel by finding a maximum-weight independent set of the graph G_S . Since G_S has few number of nodes (several tens), it is possible to find the maximum-weight independent set even by enumerating all possible independent sets [21] and comparing their total weights.

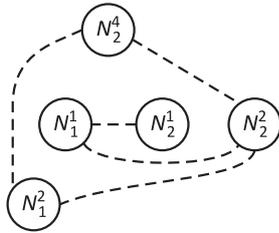


Figure 8 Supplementary graph of the example in Fig. 1, showing which subroutes cannot be run in parallel

Important, the parallel method should not be launched on its own. For example, assume that at the first iteration of Algorithm 3, the independent sets of nodes are A_1 and A_2 . Let us assume that A_1 complies better to QoS demands than A_2 and, thus, A_1 will be selected. Also, let us assume that B_1 is the next independent set in the graph if A_1 was selected and B_2 if A_2 was selected. Also, let us assume that B_1 is the next independent set in the graph if A_1 was selected and B_2 if A_2 was selected. It is possible that due to the dynamics of the system (see section 3), we could obtain overall higher QoS results if we initially launched the subroutes from the sets A_2 and B_2 afterwards than from the sets A_1 and B_1 .

Therefore, the graph that we create in this section for parallelization constraints is a supplementary graph which must be used in conjunction with the graphs from previous sections. Optimal results will be obtained when used in conjunction with the forest graph from subsection 5.2.

It is also important to note that, in the worst case, when there are no disjoint subroutes, the parallel method is reduced to the sequential method thought with a higher running time.

6 Implementation

We implemented the update algorithms from [4] as services for our QoS verification module. The update

Algorithm 3: Parallel Update

```

1 weighted graph  $G_S$ 
2 while there are still current subroutes to update do
3    $A \leftarrow$  find maximum-weight independent set in  $G_S$ 
  /* do in parallel */
4   foreach  $N_i^t \in A$  do
5     launch  $N_i^t$ 
6   end
7 end

```

algorithm itself was not modified. In other words, we treated the update itself as an atomic action. The route updates algorithms are implemented as applications interacting with the northbound interface of an SDN controller. We used POX [22] as a platform for controller development and Mininet [23] for network topology emulation. Figure 9 depicts the schematic arrangement of the functional elements.

We created networks with topology of random graph and small-world features. During each simulation trial, a pair of common source and destination nodes (s, d) were selected. A path connecting s and d was selected as a current route and a set of 4 new routes connecting (s, d), to replace the current route, were selected, possibly with shared links among themselves and the current route.

We considered latency due to the formed congestion as QoS demands for the update, implemented by forming congestion on randomly selected subroutes. Route update was executed by the update algorithm from [4] for each pair of current and new routes. Further, one of the enhanced versions was executed, updating to the preferred combination of subroutes, by identifying the congested subroutes (e. g., by estimating latency).

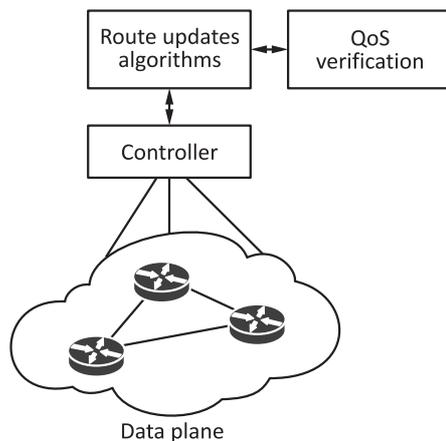


Figure 9 Description of the system

7 Concluding Remarks

The study in this paper illustrates a feasibility of modeling and designing the route update process via verification using DTMC. The goal was to strengthen the network administrator involvement in management and decision making during route update. In the present model, the network administrator is able to consider network parameters such as packet losses, delay, communication rounds, flow table updates, congestion, and other inherent unreliabilities of the network.

We extended the updating algorithm with the ability to compute QoS as the MC characteristics, where the

MC corresponds to the states of the update process. Using this MC computation ability, it is possible to predict the expected number of steps (delay time) required to complete the update process. These prediction results allow the administrator to make a decision whether a new route can satisfy the user requirements per QoS or a more reliable route will be selected.

We provided sequential update algorithm and an arbitrary order algorithm when for the later, it is assumed that immediate duplicate packets elimination is possible. Further, we suggest a supplementary graph and algorithm for launching updates in parallel when it is possible.

This paper proposes a conceptual approach. In future research, we will focus on optimization of predictions supplementing the network administrator with a powerful tool which will be able to enhance the update process with fine grained analysis of the network.

Acknowledgments

The first author has partially been supported by the Russian Foundation for Basic Research under grants RFBR 18-07-00669 and 18-29-03100. The second author has partially been supported by the Rita Altura Trust Chair in Computer Sciences; The Lynne and William Frankel Center for Computer Science.

The authors thank Prof. Shlomi Dolev for his valuable input and Prof. Yefim Dinitz for his comments.

References

1. Rao, S. K. 2014. SDN and its use-cases — NV and NFV: A state-of-the-art survey. NEC Technologies India Ltd. 25 p.
2. Ghaznavi, M., N. Shahriar, R. Ahmed, and R. Boutaba. 2016. Service function chaining simplified. arXiv.org. arXiv:1601.00751.
3. Hansson, H., and B. Jonsson. 1994. A logic for reasoning about time and reliability. *Form. Asp. Comput.* 6(5):512–535.
4. Delaet, S., S. Dolev, D. Khankin, S. Tzur-David, and T. Godinger. 2015. Seamless SDN route updates. *IEEE 14th Symposium (International) on Network Computing and Applications*. IEEE. 120–125.
5. Frenkel, S., D. Khankin, and A. Kutsyy. 2017. Predicting and choosing alternatives of route updates per QoS VNF in SDN. *IEEE 16th Symposium (International) on Network Computing and Applications*. IEEE. 1–6.
6. Devi, G., and S. Upadhyaya. 2015. An approach to distributed multi-path QoS routing. *Indian J. Sci. Technol.* 8(20):1–14. doi: 10.17485/ijst/2015/v8i20/49253.
7. Egilmez, H. E., S. Civanlar, and A. M. Tekalp. 2012. A distributed QoS routing architecture for scalable video streaming over multi-domain OpenFlow networks. *19th IEEE Conference (International) on Image Processing*. IEEE. 2237–2240.

8. Juttner, A., B. Szviovski, I. Mecs, and Z. Rajko. 2001. Lagrange relaxation based method for the QoS routing problem. *IEEE Conference on Computer Communications. 20th Annual Joint Conference of the IEEE Computer and Communications Society Proceedings*. IEEE. 2:859–868.
9. Yu, Z., F. Ma, J. Liu, B. Hu, and Z. Zhang. 2013. An efficient approximate algorithm for disjoint QoS routing. *Math. Probl. Eng.* 2013:489149. 9 p. doi: 10.1155/2013/489149.
10. Foerster, K.-T., S. Schmid, and S. Vissicchio. 2016. A survey of consistent network updates. Arxiv.org. arXiv: 1609.02305.
11. Reitblatt, M., N. Foster, J. Rexford, and D. Walker. 2011. Consistent updates for software-defined networks: Change you can believe in! *10th ACM Workshop on Hot Topics in Networks Proceedings*. New York, NY: ACM. Art. No. 7. doi: 10.1145/2070562.2070569.
12. Hogan, M., and F. Esposito. 2017. Stochastic delay forecasts for edge traffic engineering via Bayesian networks. *IEEE 16th Symposium (International) on Network Computing and Applications*. IEEE. 1–4.
13. McGeer, R. 2012. A safe, efficient Update Protocol for Openflow Networks. *1st Workshop on Hot Topics in Software Defined Networks Proceedings*. New York, NY: ACM. 12:61–66.
14. McGeer, R. 2013. A correct, zero-overhead protocol for network updates. *2nd ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking Proceedings*. New York, NY: ACM. 13:161–162.
15. Katta, N. P., J. Rexford, and D. Walker. 2013. Incremental consistent updates. *2nd ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking Proceedings*. New York, NY: ACM. 13:49–54.
16. Dinitz, Y., S. Dolev, and D. Khankin. 2017. Dependence graph and master switch for seamless dependent routes replacement in SDN. *IEEE 16th Symposium (International) on Network Computing and Applications*. IEEE. 1–7.
17. Amiri, S. A., S. Dudycz, S. Schmid, and S. Wiederrecht. 2016. Congestion-free rerouting of flows on DAGs. ArXiv.org. arXiv:1611.09296.
18. Kwiatkowska, M., G. Norman, and D. Parker. 2011. PRISM 4.0: Verification of probabilistic real-time systems. *Computer aided verification*. Eds. G. Gopalakrishnan and S. Qadeer. Lecture notes in computer science ser. Springer. 6806:585–591.
19. Kwiatkowska, M., G. Norman, and D. Parker. 2018. PRISM manual. Available at: <http://www.prismmodelchecker.org/manual/> (accessed December 10, 2018).
20. Open Networking Foundation. 2015. OpenFlow Switch Specification Ver 1.5.1.
21. Wu, Q., and J.-K. Hao. 2015. A review on algorithms for maximum clique problems. *Eur. J. Oper. Res.* 242(3):693–709.
22. Kaur, S., J. Singh, and N. S. Ghumman. 2014. Network programmability using POX controller. *Conference (International) on Communication, Computing and Systems*. 138.
23. Lantz, B., B. Heller, and N. McKeown. 2010. A network in a laptop: Rapid prototyping for software-defined networks. *9th ACM SIGCOMM Workshop on Hot Topics in Networks Proceedings*. New York, NY: ACM. Art. No. 19. doi: 10.1145/1868447.1868466.

Received October 9, 2018

Contributors

Frenkel Sergey L. (b. 1951) — Candidate of Science (PhD) in technology, associate professor, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; fsergei51@gmail.com

Khankin D. (b. 1983) — MSc, doctorate student, Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel; danielkh@post.bgu.ac.il

НЕПРЕРЫВНЫЕ ОБНОВЛЕНИЯ МАРШРУТА В SDN С ИСПОЛЬЗОВАНИЕМ ПРОВЕРКИ СООТВЕТСТВИЯ КАЧЕСТВУ ОБСЛУЖИВАНИЯ*

С. Л. Френкель¹, Д. Ханкин²

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

²Университет им. Бен-Гуриона в Негеве, Беэр-Шева, Израиль

Аннотация: В программно-определяемой сети (SDN — software-defined networking) уровень управления и уровень данных разделены. Это обеспечивает высокую гибкость эксплуатации, предоставляя абстракции для управления сетью приложений и возможность непосредственного программирования маршрутов. Однако из-за изменений топологии, процедуры обслуживания или происходящих сбоев иногда необходима реконфигурация и обновление сети. В предлагаемом сценарии рассматривается текущий маршрут C и набор возможных новых маршрутов $\{N_i\}$, где для замены текущего маршрута требуется один из

* Работа была частично поддержана РФФИ (гранты 18-07 00669 и 18-29-03100), а также Rita Altura Trust Chair in Computer Sciences; The Lynne and William Frankel Center for Computer Science.

новых маршрутов. Существует вероятность того, что новый маршрут N_i окажется длиннее некоторого другого нового маршрута N_j , но при этом N_i будет более надежным и он будет обновляться быстрее или работать лучше после обновления с точки зрения требований качества обслуживания (QoS — quality of service). Принимая во внимание случайный характер функционирования сети, авторы дополнили недавно предложенный алгоритм обновления маршрута Delaet с соавт. методом оценки соблюдения требований QoS во время непрерывного обновления маршрута, основанным на использовании цепей Маркова. При этом, во-первых, предлагается расширить алгоритм передачи пакетов по выбранному маршруту, сравнивая процесс обновления для возможных альтернатив маршрута. Во-вторых, предлагается несколько способов выбора комбинаций предпочтительных отрезков путей новых маршрутов, что приводит к оптимальному в смысле соответствия QoS маршруту.

Ключевые слова: программно-определяемые сети; цепи Маркова; качество обслуживания

DOI: 10.14357/19922264180408

Литература

1. Rao S. K. SDN and its use-cases — NV and NFV: A state-of-the-art survey. — NEC Technologies India Ltd., 2014. 25 p.
2. Ghaznavi M., Shahriar N., Ahmed R., Boutaba R. Service function chaining simplified // Arxiv.org, 2016. arXiv:1601.00751cs.
3. Hansson H., Jonsson B. A logic for reasoning about time and reliability // Form. Asp. Comput., 1994. Vol. 6. No. 5. P. 512–535.
4. Delaet S., Dolev S., Khankin D., Tzur-David S., Godinger T. Seamless SDN route updates // IEEE 14th Symposium (International) on Network Computing and Applications. — IEEE, 2015. P. 120–125.
5. Frenkel S., Khankin D., Kutsyy A. Predicting and choosing alternatives of route updates per QoS VNF in SDN // IEEE 16th Symposium (International) on Network Computing and Applications. — IEEE, 2017. P. 1–6.
6. Devi G., Upadhyaya S. An approach to distributed multi-path QoS routing // Indian J. Sci. Technol., 2015. Vol. 8. Iss. 20. P. 1–14. doi: 10.17485/ijst/2015/v8i20/49253.
7. Egilmez H. E., Civanlar S., Tekalp A. M. A distributed QoS routing architecture for scalable video streaming over multi-domain OpenFlow networks // 19th IEEE Conference (International) on Image Processing. — IEEE, 2012. P. 2237–2240.
8. Juttner A., Szviatovski B., Mecs I., Rajko Z. Lagrange relaxation based method for the QoS routing problem // IEEE INFOCOM 2001 Conference on Computer Communications. 20th Annual Joint Conference of the IEEE Computer and Communications Society Proceedings. — IEEE, 2001. Vol. 2. P. 859–868.
9. Yu Z., Ma F., Liu J., Hu B., Zhang Z. An efficient approximate algorithm for disjoint QoS routing // Math. Probl. Eng., 2013. Vol. 2013. Art. No. 489149. 9 p. doi: 10.1155/2013/489149.
10. Foerster K.-T., Schmid S., Vissicchio S. A survey of consistent network updates // Arxiv.org, 2016. arXiv:1609.02305.
11. Reitblatt M., Foster N., Rexford J., Walker D. Consistent updates for software-defined networks: Change you can believe in! // 10th ACM Workshop on Hot Topics in Networks Proceedings. — New York, NY, USA: ACM, 2011. Art. No. 7. doi: 10.1145/2070562.2070569.
12. Hogan M., Esposito F. Stochastic delay forecasts for edge traffic engineering via Bayesian Networks // IEEE 16th Symposium (International) on Network Computing and Applications. — IEEE, 2017. P. 1–4.
13. McGeer R. A safe, efficient Update Protocol for Openflow Networks // 1st Workshop on Hot Topics in Software Defined Networks Proceedings. — New York, NY, USA: ACM, 2012. Vol. 12. P. 61–66.
14. McGeer R. 2013. A correct, zero-overhead protocol for network updates // 2nd Workshop on Hot Topics in Software Defined Networking Proceedings. — New York, NY, USA: ACM, 2013. Vol. 13. P. 161–162.
15. Katta N. P., Rexford J., Walker D. Incremental consistent updates // 2nd Workshop on Hot Topics in Software Defined Networking Proceedings. — New York, NY, USA: ACM, 2013. Vol. 13. P. 49–54.
16. Dinitz Y., Dolev S., Khankin D. Dependence graph and master switch for seamless dependent routes replacement in SDN // IEEE 16th Symposium (International) on Network Computing and Applications. — IEEE, 2017. P. 1–7.
17. Amiri S. A., Dudycz S., Schmid S., Wiederrecht S. Congestion-free rerouting of flows on DAGs // Arxiv.org, 2016. arXiv:1611.09296.
18. Kwiatkowska M., Norman G., Parker D. PRISM 4.0: Verification of probabilistic real-time systems // Computer aided verification / Eds. G. Gopalakrishnan, S. Qadeer. — Lecture notes in computer science ser. — Springer, 2011. Vol. 6806. P. 585–591.
19. Kwiatkowska M., Norman G., Parker D. PRISM manual, 2018. <http://www.prismmodelchecker.org/manual>.
20. Open Networking Foundation. OpenFlow Switch Specification Ver 1.5.1, 2015.
21. Wu Q., Hao J.-K. A review on algorithms for maximum clique problems // Eur. J. Oper. Res., 2015. Vol. 242. No. 3. P. 693–709.
22. Kaur S., Singh J., Ghumman N. S. Network programmability using POX controller // Conference (International) on Communication, Computing and Systems, 2014. P. 138.
23. Lantz B., Heller B., McKeown N. A network in a laptop: Rapid prototyping for software-defined networks // 9th ACM SIGCOMM Workshop on Hot Topics in Networks Proceedings. — New York, NY, USA: ACM, 2010. Art. No. 19. doi: 10.1145/1868447.1868466.

Поступила в редакцию 09.10.2018

ВЫБОР ОПТИМАЛЬНОЙ МОДЕЛИ РЕКУРРЕНТНОЙ СЕТИ В ЗАДАЧАХ ПОИСКА ПАРАФРАЗА*

А. Н. Смердов¹, О. Ю. Бахтеев², В. В. Стрижов³

Аннотация: Рассматривается задача выбора оптимальной рекуррентной нейронной сети. В качестве критерия оптимальности используется нижняя оценка правдоподобия модели. Исследование сконцентрировано на применении вариационного подхода к аппроксимации апостериорного распределения параметров модели. Частным случаем аппроксимации выступает нормальное распределение параметров с различными видами матрицы ковариаций. Для увеличения правдоподобия модели предлагается метод удаления параметров с наибольшей плотностью вероятности в нуле. В качестве иллюстративного примера рассматривается задача многоклассовой классификации на выборке пар схожих и несхожих предложений SemEval 2015.

Ключевые слова: глубокое обучение; выбор оптимальной модели; рекуррентная нейросеть; разреживание нейросети; вариационный вывод

DOI: 10.14357/19922264180409

1 Введение

Целью работы является выбор оптимальной нейросетевой модели из класса рекуррентных нейронных сетей. Рекуррентной нейросетью называется нейросеть со связью между нейронами одного слоя. В качестве критерия оптимальности используется нижняя оценка правдоподобия модели.

Число параметров в моделях глубокого обучения может достигать миллионов [1]. Большое число параметров влечет сложность оптимизации параметров и переобучение моделей [2]. Предлагается уменьшить число параметров рекуррентной сети. Это обеспечит большую устойчивость модели и снизит время оптимизации ее параметров. Для решения поставленной задачи используются как байесовские методы [3], так и методы прореживания переусложненной нейросети, наращивания простой нейросети и их комбинации [4].

Для построения модели рекуррентной сети рассматривается модель из [5], решающая задачу определения сходства предложений. Модель принимает на вход векторизованные представления слов. Векторизация выполняется с помощью алгоритма GloVe (Global Vectors for word representation), основанного на факторизации матрицы слов-контекстов и использовании весовой функции для уменьшения значимости редких слов [6]. Альтерна-

тивной этому алгоритму выступает линейная модель Word2vec, комбинирующая в себе Continuous Bag-of-Words, skip-gram и negative sampling [7]. Несмотря на разные подходы к проблеме, GloVe и Word2vec оптимизируют схожие функционалы [8]. Упрощенной линейной моделью Word2vec, предназначенной для классификации документов, является fastText — метод, работающий на символьных n -граммах [9].

В работе предлагается подход, основанный на получении вариационной нижней оценки правдоподобия модели. Подобная задача решалась в [10] аппроксимацией апостериорного распределения нормальным, получением аналитических формул для нижней границы правдоподобия модели и удалением параметров с наибольшей плотностью вероятности в нуле. Описанный ниже подход продолжает это исследование. Априорное и апостериорное распределения параметров аппроксимируются нормальным со скалярным, диагональным и блочным видами матрицы ковариаций. После оптимизации гиперпараметров выполняется прореживание сети.

Предлагаемый подход сравнивается с методом удаления параметров Optimal Brain Damage, базирующимся на анализе функции ошибки [11]. Его обобщенной версией выступает алгоритм Optimal Brain Surgeon [12], не предполагающий диагонального вида гессиана функции ошибки.

* Работа выполнена при финансовой поддержке РФФИ (проект 16-07-01160) и Правительства Российской Федерации (соглашение № 05.Y09.21.0018).

¹Московский физико-технический институт, anton.smerdov1@gmail.com

²Московский физико-технический институт, bakhteev@phystech.edu

³Московский физико-технический институт; Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, strizov@ccas.ru

Вычислительный эксперимент проводится на выборке размеченных пар предложений SemEval 2015. Для каждой пары предложений из корпуса дана экспертная оценка их семантической близости. Требуется построить модель, оценивающую семантическую близость двух предложений. Проблема рассматривается как задача многоклассовой классификации аналогично [5]. Критерием качества служит F1-мера, учитывающая как полноту, так и точность предсказаний. В качестве базовой модели рассматривается пара соединенных рекуррентных сетей с общим вектором параметров и softmax-классификатором на выходе.

2 Задача выбора оптимальной нейросетевой модели

Для построения выборки используем набор пар предложений SemEval 2015 [13]. Каждому слову сопоставим вектор размерности n . Обозначим через l число слов в самом длинном предложении. Предложения длины, меньшей l , дополним нулевыми векторами. Построим выборку

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}, \quad i = 1, \dots, N,$$

где $\mathbf{x}_i = [\mathbf{x}_i^1, \mathbf{x}_i^2]$ — пары последовательностей векторов слов, соответствующих i -й паре предложений, $\mathbf{x}_i^1, \mathbf{x}_i^2 \in \mathbb{R}^{n \times l}$; $y_i \in Y = \{0, \dots, 5\}$ — экспертная оценка семантической близости.

Требуется построить модель $f(\mathbf{w}) : \mathbb{R}^{n \times l} \times \mathbb{R}^{n \times l} \rightarrow Y$, сопоставляющую паре предложений \mathbf{x}_i^1 и \mathbf{x}_i^2 класс семантической близости, где $\mathbf{w} \in \mathbb{W} \subseteq \mathbb{R}^s$ — пространство параметров модели. Искомая модель выбирается из множества \mathcal{F} рекуррентных нейронных сетей с функцией активации tanh. Модель

$$f(\mathbf{w}) : \mathbb{R}^{n \times l} \times \mathbb{R}^{n \times l} \rightarrow Y$$

принадлежит искомому классу моделей \mathcal{F} , если существуют такие матрицы перехода $\mathbf{W} \in \mathbb{R}^{n \times m}$, $\mathbf{U} \in \mathbb{R}^{n \times n}$, $\mathbf{V} \in \mathbb{R}^{(|Y| \times 2n)}$ и вектор смещения $\mathbf{b} \in \mathbb{R}^n$, что для j -х элементов $\mathbf{x}_{ij}^1, \mathbf{x}_{ij}^2 \in \mathbb{R}^m$ последовательностей \mathbf{x}_i^1 и \mathbf{x}_i^2 определены векторы скрытого слоя $\mathbf{h}_{ij}^1, \mathbf{h}_{ij}^2 \in \mathbb{R}^n$:

$$\begin{aligned} \mathbf{h}_{ij}^1 &= \tanh(\mathbf{W} \cdot \mathbf{x}_{ij}^1 + \mathbf{U} \cdot \mathbf{h}_{i,j-1}^1 + \mathbf{b}); \\ \mathbf{h}_{ij}^2 &= \tanh(\mathbf{W} \cdot \mathbf{x}_{ij}^2 + \mathbf{U} \cdot \mathbf{h}_{i,j-1}^2 + \mathbf{b}). \end{aligned}$$

Для определения класса семантической близости используются последние значения скрытого слоя \mathbf{h}_{il}^1 и \mathbf{h}_{il}^2 , сконкатенированные в один вектор. После l -й итерации пару предложений будем относить

к классу с наибольшим значением, полученным после l -й итерации, $j = 1, \dots, l$:

$$y = \arg \max_{k \in Y} \left(\mathbf{V} \begin{bmatrix} \mathbf{h}_{il}^1 \\ \mathbf{h}_{il}^2 \end{bmatrix} \right)_k,$$

где $(\cdot)_k$ — k -я компонента вектора. Для каждой модели и соответствующего ей вектора параметров $\mathbf{w} \in \mathbb{W}$ определим логарифмическую функцию правдоподобия выборки $L_{\mathcal{D}}(\mathcal{D}, \mathbf{f}, \mathbf{w})$:

$$\begin{aligned} L_{\mathcal{D}}(\mathcal{D}, \mathbf{f}, \mathbf{w}) &= \log p(\mathbf{y}|\mathbf{x}, \mathbf{f}, \mathbf{w}) = \log p(\mathcal{D}|\mathbf{f}, \mathbf{w}) = \\ &= \sum_{(\mathbf{x}_i, y_i) \in \mathcal{D}} \log p(y_i|\mathbf{x}_i, \mathbf{f}, \mathbf{w}), \quad (1) \end{aligned}$$

где $p(\mathbf{y}|\mathbf{x}, \mathbf{f}, \mathbf{w})$ — апостериорная вероятность вектора \mathbf{y} при заданных \mathbf{x} , \mathbf{f} и \mathbf{w} . Здесь и далее используется обозначение $p(\mathbf{x}|\mathbf{y}) = p(\mathcal{D})$.

Оптимальная модель \mathbf{f} находится максимизацией логарифма ее правдоподобия:

$$\begin{aligned} L_{\mathbf{f}}(\mathcal{D}, \mathbf{f}) &= L_{\mathbf{f}}(\mathcal{D}|\mathbf{f}) = \log p(\mathbf{y}|\mathbf{x}, \mathbf{f}) = \\ &= \log p(\mathcal{D}|\mathbf{f}) = \log \int_{\mathbf{w} \in \mathbb{W}} p(\mathcal{D}|\mathbf{w}, \mathbf{f}) p(\mathbf{w}|\mathbf{f}) d\mathbf{w}. \quad (2) \end{aligned}$$

Апостериорное распределение параметров модели находится из уравнения:

$$p(\mathbf{w}|\mathcal{D}, \mathbf{f}) = \frac{p(\mathcal{D}|\mathbf{w}, \mathbf{f}) p(\mathbf{w}|\mathbf{f})}{p(\mathcal{D}|\mathbf{f})}. \quad (3)$$

Приближим интеграл (2) вариационной нижней оценкой. Воспользуемся оценкой [2, разд. 10.2–10.4], полученной из неравенства Йенсена:

$$\begin{aligned} L_{\mathbf{f}}(\mathcal{D}, \mathbf{f}) &= \log \int_{\mathbf{w} \in \mathbb{W}} p(\mathcal{D}|\mathbf{w}) p(\mathbf{w}|\mathbf{f}) d\mathbf{w} = \\ &= \int_{\mathbf{w} \in \mathbb{W}} p(\mathbf{w}|\mathcal{D}, \mathbf{f}) \log \frac{p(\mathbf{w}|\mathcal{D}, \mathbf{f})}{p(\mathcal{D}|\mathbf{f}, \mathbf{w})} d\mathbf{w} + \\ &\quad + D_{\text{KL}}(p(\mathcal{D}|\mathbf{w}) || p(\mathcal{D}|\mathbf{f})), \end{aligned}$$

где $D_{\text{KL}}(q(\mathbf{w}) || p(\mathbf{w}))$ — расстояние Кульбака–Лейблера между $q(\mathbf{w})$ и $p(\mathbf{w})$,

$$D_{\text{KL}}(q(\mathbf{w}) || p(\mathbf{w})) = - \int_{\mathbf{w}} q(\mathbf{w}) \log \frac{q(\mathbf{w})}{p(\mathbf{w})} d\mathbf{w}.$$

Учитывая неотрицательность расстояния Кульбака–Лейблера, получаем:

$$L_{\mathbf{f}}(\mathcal{D}, \mathbf{f}) \geq \int_{\mathbf{w} \in \mathbb{W}} p(\mathbf{w}|\mathcal{D}, \mathbf{f}) \log \frac{p(\mathbf{w}|\mathcal{D}, \mathbf{f})}{p(\mathcal{D}|\mathbf{f}, \mathbf{w})} d\mathbf{w}. \quad (4)$$

Упростим интеграл в левой части (4):

$$\begin{aligned} \int_{\mathbf{w} \in \mathbb{W}} p(\mathbf{w}|\mathcal{D}, \mathbf{f}) \log \frac{p(\mathbf{w}|\mathcal{D}, \mathbf{f})}{p(\mathcal{D}|\mathbf{f}, \mathbf{w})} d\mathbf{w} = \\ = -D_{\text{KL}}(p(\mathbf{w}|\mathcal{D}, \mathbf{f})||p(\mathbf{w}|\mathbf{f})) + \\ + \int_{\mathbf{w} \in \mathbb{W}} p(\mathbf{w}|\mathcal{D}, \mathbf{f}) \log p(\mathcal{D}|\mathbf{f}, \mathbf{w}) d\mathbf{w}. \quad (5) \end{aligned}$$

Обозначим сумму в левой части (5) через $-L(\mathcal{D}, \mathbf{f}, \mathbf{w})$:

$$\begin{aligned} L(\mathcal{D}, \mathbf{f}, \mathbf{w}) = \underbrace{D_{\text{KL}}(p(\mathbf{w}|\mathcal{D}, \mathbf{f})||p(\mathbf{w}|\mathbf{f}))}_{L_{\mathbf{w}}(\mathcal{D}, \mathbf{f}, \mathbf{w})} - \\ - \underbrace{\int_{\mathbf{w} \in \mathbb{W}} p(\mathbf{w}|\mathcal{D}, \mathbf{f}) \log p(\mathcal{D}|\mathbf{f}, \mathbf{w}) d\mathbf{w}}_{L_E(\mathcal{D}, \mathbf{f})}. \quad (6) \end{aligned}$$

Первое слагаемое формулы (6) интерпретируется как минимальная длина описания распределения $p(\mathbf{w}|\mathcal{D}, \mathbf{f})$ с помощью $p(\mathbf{w}|\mathbf{f})$. Эту величину назовем сложностью модели $L_{\mathbf{w}}(\mathcal{D}, \mathbf{f}, \mathbf{w})$:

$$L_{\mathbf{w}}(\mathcal{D}, \mathbf{f}, \mathbf{w}) = D_{\text{KL}}(p(\mathbf{w}|\mathcal{D}, \mathbf{f})||p(\mathbf{w}|\mathbf{f})).$$

Второе слагаемое формулы (6) есть минус матожидание правдоподобия выборки $L_{\mathcal{D}}(1)$, и оно тем меньше, чем выше правдоподобие выборки, поэтому интерпретируется как функционал ошибки $L_E(\mathcal{D}, \mathbf{f})$ в ходе вычислительного эксперимента:

$$L_E(\mathcal{D}, \mathbf{f}) = \mathbb{E}_{\mathbf{w} \sim p(\mathbf{w}|\mathcal{D}, \mathbf{f})} L_{\mathcal{D}}(\mathbf{y}, \mathcal{D}, \mathbf{f}, \mathbf{w}).$$

Запишем суммарную функцию потерь $L(\mathcal{D}, \mathbf{f}, \mathbf{w})$ как сумму функционала сложности модели $L_{\mathbf{w}}(\mathcal{D}, \mathbf{f}, \mathbf{w})$ и функционала ошибки $L_E(\mathcal{D}, \mathbf{f})$:

$$L(\mathcal{D}, \mathbf{f}, \mathbf{w}) = L_{\mathbf{w}}(\mathcal{D}, \mathbf{f}, \mathbf{w}) + L_E(\mathcal{D}, \mathbf{f}). \quad (7)$$

Искомая модель минимизирует суммарный функционал потерь

$$\mathbf{f} = \arg \min_{\mathbf{f} \in \mathcal{F}} L(\mathcal{D}, \mathbf{f}, \mathbf{w}).$$

3 Предлагаемое решение оптимизационной задачи

Так как апостериорное распределение $p(\mathbf{w}|\mathcal{D}, \mathbf{f})$ (3) невозможно получить аналитически, минимизация функционала потерь $L(\mathcal{D}, \mathbf{f}, \mathbf{w})$ (7) затруднена. Для решения этой проблемы применим вариационный подход. Он заключается

в аппроксимации неизвестного распределения распределением из известного класса. В качестве приближения $p(\mathbf{w}|\mathcal{D}, \mathbf{f})$ выберем нормальное распределение:

$$p(\mathbf{w}|\mathcal{D}, \mathbf{f}) \sim \mathcal{N}(\mathbf{m}, \mathbf{A}_{\text{ps}}^{-1}),$$

где \mathbf{m} и $\mathbf{A}_{\text{ps}}^{-1}$ — вектор средних и матрица ковариации этого распределения. Априорное распределение $p(\mathbf{w}|\mathbf{f})$ вектора параметров \mathbf{w} будем считать нормальным с параметрами $\boldsymbol{\mu}$ и $\mathbf{A}_{\text{pr}}^{-1}$:

$$p(\mathbf{w}|\mathbf{f}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{A}_{\text{pr}}^{-1}),$$

где $\boldsymbol{\mu}$ — вектор средних; $\mathbf{A}_{\text{pr}}^{-1}$ — матрица ковариаций. Расстояние Кульбака–Лейблера между нормальными распределениями $\mathcal{N}(\boldsymbol{\mu}, \mathbf{A}_{\text{pr}}^{-1})$ и $\mathcal{N}(\mathbf{m}, \mathbf{A}_{\text{ps}}^{-1})$ вычисляется по формуле:

$$\begin{aligned} D_{\text{KL}}(\mathcal{N}(\boldsymbol{\mu}, \mathbf{A}_{\text{pr}}^{-1})||\mathcal{N}(\mathbf{m}, \mathbf{A}_{\text{ps}}^{-1})) = \\ = \frac{1}{2} \left(\log \frac{|\mathbf{A}_{\text{ps}}^{-1}|}{|\mathbf{A}_{\text{pr}}^{-1}|} - d + \text{tr}(\mathbf{A}_2 \mathbf{A}_{\text{pr}}^{-1}) + \right. \\ \left. + (\boldsymbol{\mu} - \mathbf{m})^T \mathbf{A}_2 (\boldsymbol{\mu} - \mathbf{m}) \right). \end{aligned}$$

Рассмотрим частные случаи вида матриц ковариаций $\mathbf{A}_{\text{pr}}^{-1}$ и $\mathbf{A}_{\text{ps}}^{-1}$. Так как априори нет предпочтений при выборе параметров, то априорное распределение для всех параметров считаем одинаковым, т. е. вектор средних $\boldsymbol{\mu} = \mu \mathbf{1}$, матрица ковариаций скалярна: $\mathbf{A}_{\text{pr}}^{-1} = \sigma \mathbf{I}$. После получения информации о выборке получаем апостериорный вектор средних \mathbf{m} .

Алгоритм решения оптимизационной задачи заключается в выполнении градиентного шага при заданном априорном распределении, вычислении апостериорного распределения и аппроксимации нового априорного распределения полученным апостериорным. Рассмотрим различные виды апостериорной матрицы ковариаций $\mathbf{A}_{\text{ps}}^{-1}$.

1. Матрица ковариаций скалярна: $\mathbf{A}_{\text{ps}}^{-1} = \alpha \mathbf{I}$. В этом случае

$$\begin{aligned} D_{\text{KL}}(\mathcal{N}(\boldsymbol{\mu}, \mathbf{A}_{\text{pr}}^{-1})||\mathcal{N}(\mathbf{m}, \mathbf{A}_{\text{ps}}^{-1})) = \\ = \sum_{i=1}^W \left(\log \frac{\sigma}{\alpha} + \frac{(\mu - m_i)^2 + \alpha^2 + \sigma^2}{2\sigma^2} \right). \end{aligned}$$

По значениям параметров α и \mathbf{m} апостериорного распределения вычислим параметры априорного. Число элементов вектора \mathbf{m} обозначим W . Из условия

$$\frac{\partial}{\partial \mu} D_{\text{KL}} = \sum_{i=1}^W \frac{\mu - m_i}{\sigma^2} = 0$$

получаем выражения для μ на следующей итерации $\hat{\mu} = (1/W) \sum_{i=1}^W m_i$. Аналогично

$$\frac{\partial}{\partial \sigma^2} D_{\text{KL}} = \sum_{i=1}^W \frac{1}{2\sigma^2} - \frac{(\mu - m_i)^2 + \alpha^2}{2\sigma^4} = 0 \Rightarrow$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{W} \sum_{i=1}^W (\mu - m_i)^2 + \alpha^2.$$

2. Матрица ковариаций диагональна: $\mathbf{A}_{\text{ps}}^{-1} = \text{diag}(\sigma^2)$. В этом случае

$$D_{\text{KL}}(\mathcal{N}(\boldsymbol{\mu}, \mathbf{A}_{\text{pr}}^{-1}) || \mathcal{N}(\mathbf{m}, \mathbf{A}_{\text{ps}}^{-1})) =$$

$$= \sum_{i=1}^W \left(\log \frac{\sigma}{\sigma_i} + \frac{(\mu - m_i)^2 + \sigma_i^2 + \sigma^2}{2\sigma^2} \right).$$

Значения параметров априорного распределения для следующей итерации вычисляются следующим образом:

$$\text{из } \frac{\partial}{\partial \mu} D_{\text{KL}} = \sum_{i=1}^W \frac{\mu - m_i}{\sigma^2} = 0$$

получаем $\hat{\mu} = \frac{1}{W} \sum_{i=1}^W m_i$;

$$\text{из } \frac{\partial}{\partial \sigma^2} D_{\text{KL}} = \sum_{i=1}^W \frac{1}{2\sigma^2} - \frac{(\mu - m_i)^2 + \sigma_i^2}{2\sigma^4} = 0$$

получаем $\hat{\sigma}^2 = \frac{1}{W} \sum_{i=1}^W (\mu - m_i)^2 + \sigma_i^2$.

Оптимизация параметров сводится к следующему алгоритму.

Инициализировать $\boldsymbol{\sigma} = \mathbf{1}$, $\mathbf{m} = \mathbf{0}$, $\mu = 0$, $\sigma^2 = 1$.

Повторять:

Сделать градиентный шаг $\boldsymbol{\sigma} := \boldsymbol{\sigma} - \eta \nabla \boldsymbol{\sigma}$,
 $\mathbf{m} := \mathbf{m} - \eta \nabla \mathbf{m}$, $\mathbf{w} := \mathbf{w} - \eta \nabla \mathbf{w}$.

Обновить параметры априорного распределения $\mu := \hat{\mu}$, $\sigma^2 := \hat{\sigma}^2$.

Пока значение L не стабилизируется.

Результаты вычислительного эксперимента представлены в таблице.

Результаты вычислительного эксперимента

Метод	F1-мера	
	Валидация	Эксперимент
Logistic Regression	0,286	0,286
SVC	0,290	0,290
DecisionTreeClassifier	0,316	0,316
KNeighborsClassifier	0,322	0,322
RNN	0,393	0,362
RNN+variational, I, I	—	0,311
RNN+variational, D, I	—	0,330

4 Удаление параметров из сети

Введем множество индексов активных параметров модели $\mathcal{A} = \{i | w_i \neq 0\}$. Для увеличения правдоподобия модели предлагается уменьшить ее сложность, т.е. уменьшить число параметров $|\mathcal{A}|$. Для удаления выберем параметры, имеющие наибольшую плотность апостериорной вероятности ρ в нуле. Если апостериорная матрица ковариаций скалярна, то

$$\rho_i = \exp\left(-\frac{\mu_i^2}{2\sigma^2}\right).$$

Чем больше ρ , тем меньше $|\mu_i/\sigma|$, поэтому удаляются параметры со значением $|\mu_i/\sigma| < \lambda$, где λ — пороговое значение. Варьируя пороговое значение λ , выбираем оптимальное число неудаленных параметров. Для диагонального вида матрицы ковариаций критерий удаления параметров записывается как $|\mu_i/\sigma_i| < \lambda$.

5 Вычислительный эксперимент

Цель эксперимента — проверка работоспособности предложенного алгоритма и сравнение результатов с ранее полученными. В качестве данных использовалась выборка SemEval 2015, состоящая из 8331 пары схожих и несхожих предложений. Слова преобразовывались в векторы размерности 50 при помощи алгоритма GloVe [14]. Для базовых алгоритмов тренировочная, валидационная и тестовая выборки составили 70%, 15% и 15% соответственно. Для рекуррентной нейронной сети, полученной вариационным методом, валидационная выборка отсутствовала, а тренировочная и тестовая выборки составили 85% и 15% соответственно. Критерием качества была выбрана F1-мера. В качестве базовых алгоритмов использовались линейная регрессия, метод ближайших соседей, решающее дерево и модификация метода опорных векторов SVC (support vector clustering). Базовые алгоритмы взяты из библиотеки sklearn. Дополнительно были построены рекуррентная нейросеть с одним скрытым слоем [5] и нейросеть с одним скрытым слоем и вариационной оптимизацией параметров [10, 15].

На рис. 1, а представлены зависимости оценки правдоподобия L (7) от параметра λ . Для обоих случаев существует оптимальное значение λ , минимизирующее L ; модели с таким параметром будут оптимальными. На рис. 1, б и 1, в отображены зависимости качества модели от λ и доли выброшенных параметров. Видно, что даже при удалении большинства параметров из сети качество предсказаний

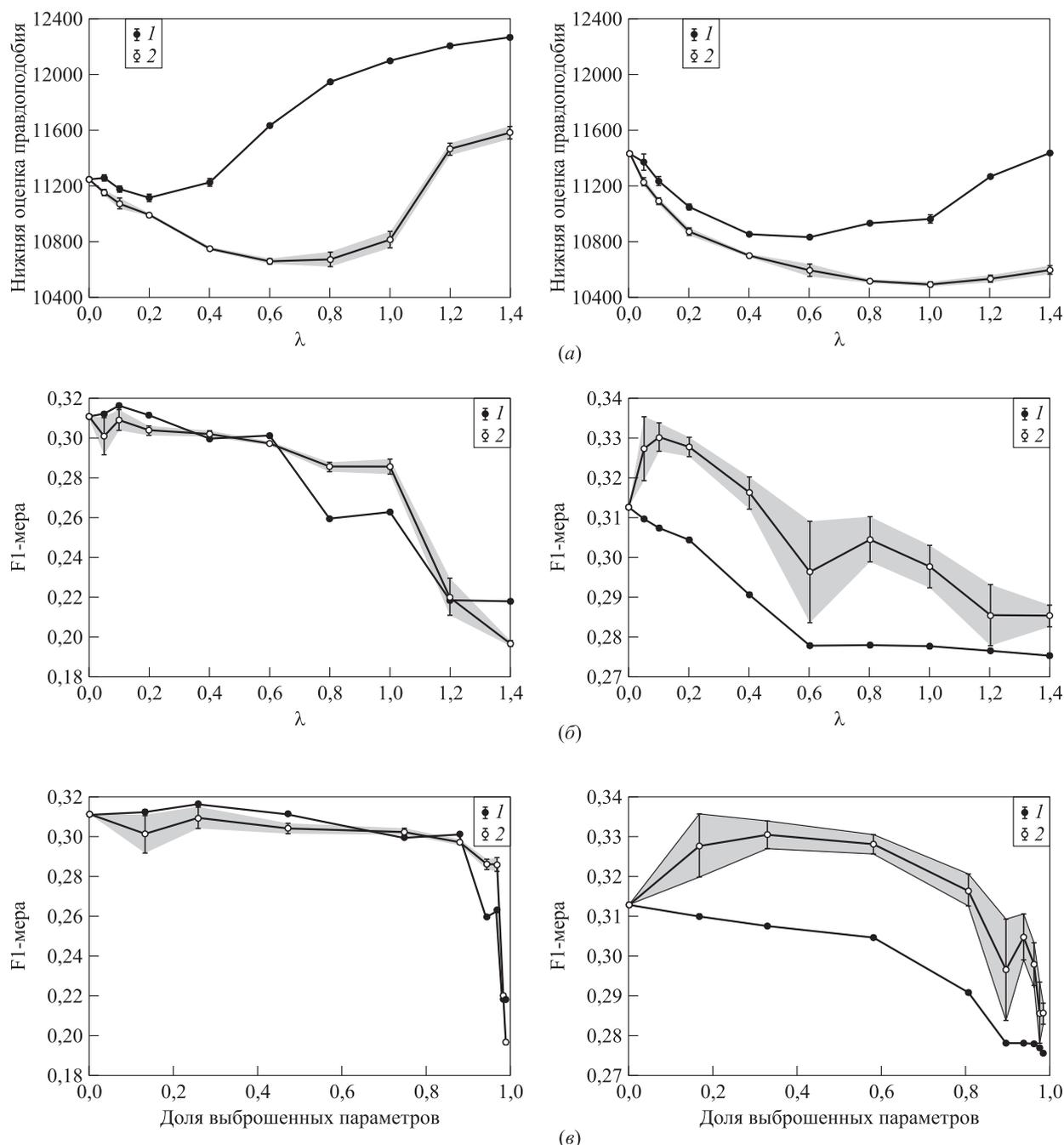


Рис. 1 Зависимость нижней оценки правдоподобия модели от λ (а) и F1-меры от λ (б) и доли выброшенных параметров (в) для скалярной (левый столбец) и диагональной (правый столбец) матриц: *I* — первоначальная оценка; *2* — полученная оценка

меняется несущественно, что говорит о слишком большом числе параметров исходной модели.

Из рис. 2 видно, что при малых λ из сети с диагональной апостериорной матрицей ковариаций удаляется больше весов, а при больших λ — меньше, что говорит о лучшем отборе параметров такой моделью.

6 Заключение

С помощью вариационного байесовского подхода был построен набор моделей глубинного обучения с оптимальной нижней оценкой правдоподобия, отличающихся различными предположениями о виде априорного и апостериорного

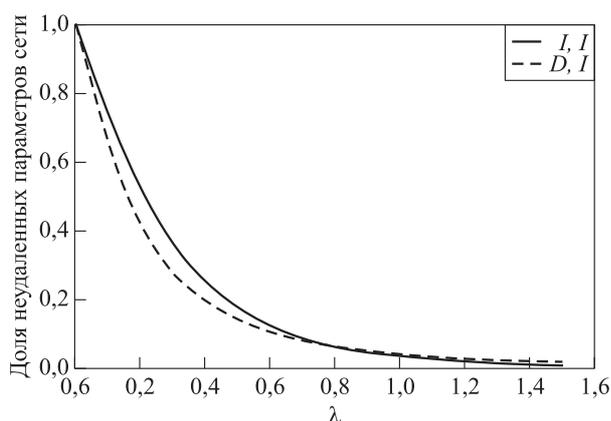


Рис. 2 Доля неудаленных параметров сети в зависимости от порогового значения λ для скалярного (I) и диагонального (D) вида апостериорной матрицы ковариаций

распределения параметров. Из случайности распределения параметров был получен критерий их удаления, что позволило увеличить нижнюю оценку правдоподобия моделей. Результаты полученных нейросетей в вычислительном эксперименте оказались близки к результатам других алгоритмов.

Литература

1. Sutskever I., Vinyals O., Le Q. V. Sequence to sequence learning with neural networks // Adv. Neur. In., 2014. Vol. 27. P. 3104–3112. <https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf>.
2. Bishop C. M. Pattern recognition and machine learning. — Springer, 2006. 758 p.
3. Kuznetsov M. P., Tokmakova A. A., Strijov V. V. Analytic and stochastic methods of structure parameter estimation // Informatica, 2016. Vol. 27. No. 3. P. 607–624.

4. Попова М. С., Стрижов В. В. Выбор оптимальной модели классификации физической активности по измерениям акселерометра // Информатика и её применения, 2015. Т. 9. Вып. 1. С. 76–86.
5. Sanborn A., Skryzalin J. Deep learning for semantic similarity // Deep learning for natural language processing. — Stanford, CA, USA: Stanford University, 2015. Vol. CS224d. P. 1–7. <https://cs224d.stanford.edu/reports/SanbornAdrian.pdf>.
6. Pennington J., Socher R., Manning C. D. GloVe: Global vectors for word representation // Conference on Empirical Methods in Natural Language Processing Proceedings, 2014. Vol. 12. P. 1532–1543. <https://nlp.stanford.edu/pubs/glove.pdf>.
7. Rong X. Word2vec parameter learning explained // arXiv:1411.2738, 2014. 21 p.
8. Shi T., Liu Z. Linking GloVe with word2vec // arXiv: 1411.5595, 2014. 5 p.
9. Zolotov V., Kung D. Analysis and optimization of fastText linear text classifier // arXiv:1702.05531, 2017. 9 p.
10. Graves A. Practical variational inference for neural networks // Adv. Neur. In., 2011. Vol. 24. P. 2348–2356. <http://papers.nips.cc/paper/4329-practical-variational-inference-for-neural-networks.pdf>.
11. Le Cun Y., Denker J. S., Solla S. A. Optimal brain damage // Adv. Neur. In., 1989. Vol. 2. P. 598–605. <https://papers.nips.cc/paper/250-optimal-brain-damage.pdf>.
12. Hassibi B., Stork D. G., Wolff G. J. Optimal brain surgeon and general network pruning // Neural Comput., 1992. Vol. 4. P. 1–8.
13. Выборка пар предложений различной степени схожести. <http://alt.qcri.org/semEval2015/task2/index.php?id=data-and-tools>.
14. Библиотека GloVe. Python. <https://github.com/stanfordnlp/GloVe>.
15. Смердов А. Н. Код вычислительного эксперимента. <https://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group474/Smerdov2017Paraphrase/code>.

Поступила в редакцию 08.05.18

OPTIMAL RECURRENT NEURAL NETWORK MODEL IN PARAPHRASE DETECTION

A. N. Smerdov¹, O. Y. Bakhteev¹, and V. V. Strijov^{1,2}

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation

²A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: This paper addresses the problem of optimal recurrent neural network selection. It asserts the neural network evidence lower bound as the optimal criterion for selection. It investigates variational inference methods to approximate the posterior distribution of the network parameters. As a particular case, the normal distribution

of the parameters with different types of the covariance matrix is investigated. The authors propose a method of pruning parameters with the highest probability density in zero to increase the model marginal likelihood. As an illustrative example, a computational experiment of multiclass classification on the SemEval 2015 dataset was carried out.

Keywords: deep learning; recurrent neural network; neural network pruning; variational approach

DOI: 10.14357/19922264180409

Acknowledgments

This research was supported by the Russian Foundation for Basic Research, project 16-07-01160, and by Government of the Russian Federation, agreement 05.Y09.21.0018.

References

1. Sutskever, I., O. Vinyals, and Q. V. Le. 2014. Sequence to sequence learning with neural networks. *Adv. Neur. In.* 27:3104–3112. Available at: <https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf> (accessed December 29, 2017).
2. Bishop, C. M. 2006. *Pattern recognition and machine learning*. Springer. 758 p.
3. Kuznetsov, M. P., A. A. Tokmakova, and V. V. Strijov. 2016. Analytic and stochastic methods of structure parameter estimation. *Informatika* 27(3):607–624.
4. Popova, M. S., and V. V. Strijov. 2015. Vybor optimal'noy modeli klassifikatsii fizicheskoy aktivnosti po izmereniyam akselerometra [Selection of optimal physical activity classification model using measurements of accelerometer]. *Informatika i ee Primeneniya — Inform. Appl.* 9(1):76–86.
5. Sanborn, A., and J. Skryzalin. 2015. Deep learning for semantic similarity. *Deep learning for natural language processing*. Stanford, CA: Stanford University. CS224d:1–7. Available at: <https://cs224d.stanford.edu/reports/SanbornAdrian.pdf> (accessed December 29, 2017).
6. Pennington, J., R. Socher, and C. D. Manning. 2014. GloVe: Global vectors for word representation. *Conference on Empirical Methods in Natural Language Processing Proceedings*. 12:1532–1543. <https://nlp.stanford.edu/pubs/glove.pdf> (accessed December 29, 2017).
7. Rong, X. 2014. Word2vec parameter learning explained. Arxiv. Available at: <https://arxiv.org/abs/1411.2738> (accessed December 29, 2017).
8. Shi, T., and Z. Liu. 2014. Linking GloVe with word2vec. Arxiv. Available at: <http://arxiv.org/abs/1411.5595> (accessed December 29, 2017).
9. Zolotov, V., and D. Kung. 2017. Analysis and optimization of fastText linear text classifier. Arxiv. Available at: <https://arxiv.org/ftp/arxiv/papers/1702/1702.05531.pdf> (accessed December 29, 2017).
10. Graves, A. 2011. Practical variational inference for neural networks. *Adv. Neur. In.* 24:2348–2356. Available at: <http://papers.nips.cc/paper/4329-practical-variational-inference-for-neural-networks.pdf> (accessed December 29, 2017).
11. Le Cun, Y., J. S. Denker, and S. A. Solla. 1989. Optimal brain damage. *Adv. Neur. In.* 2:598–605. Available at: <https://papers.nips.cc/paper/250-optimal-brain-damage.pdf> (accessed December 29, 2017).
12. Hassibi, B., D. G. Stork, and G. J. Wolff. 1992. Optimal brain surgeon and general network pruning. *Neural Comput.* 4:1–8.
13. Dataset of sentences with different types of similarity. Available at: <http://alt.qcri.org/semEval2015/task2/index.php?id=data-and-tools> (accessed December 29, 2017).
14. GloVe Python library. Available at: <https://github.com/stanfordnlp/GloVe> (accessed December 29, 2017).
15. Smerdov, A. N. Computational experiment code. Available at: <https://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group474/Smerdov2017Paraphrase/code/> (accessed December 29, 2017).

Received May 5, 2018

Contributors

Smerdov Anton N. (b. 1995) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; anton.smerdov1@gmail.com

Bakhteev Oleg Y. (b. 1991) — graduate student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; bakhteev@phystech.edu

Strijov Vadim V. (b. 1967) — Doctor of Science in physics and mathematics, professor, Department of Intelligent Systems, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; strijov@ccas.ru

ТЕОРЕТИКО-ИНФОРМАЦИОННАЯ ХАРАКТЕРИЗАЦИЯ СОВЕРШЕННО УРАВНОВЕШЕННЫХ ФУНКЦИЙ*

О. А. Логачев¹

Аннотация: Совершенно уравновешенные дискретные функции являются объектом исследований для ряда математических дисциплин, близких к информатике, таких как комбинаторика, теория кодирования, криптография, символическая динамика, теория автоматов и др. Данный класс дискретных функций оказался удобным математическим инструментом для синтеза и описания сверточных кодов, некоторых криптографических примитивов, сюръективных эндоморфизмов дискретных динамических систем, а также для конечных автоматов без потери информации. Ранее Хеллундом и Сумароковым были доказаны критерии, связывающие свойство совершенной уравновешенности со свойствами функции быть дефекта нуль и без потери информации. В данной статье доказывается новый критерий совершенной уравновешенности функции, носящий теоретико-информационный характер, а также рассмотрены некоторые алгоритмические свойства совершенно уравновешенных функций как преобразователей информации.

Ключевые слова: конечный алфавит; дискретная функция; случайная величина; закон распределения; взаимная энтропия; совершенная уравновешенность

DOI: 10.14357/19922264180410

1 Введение

В работах [1, 2] показана эквивалентность следующих свойств дискретных функций над конечным алфавитом:

- совершенная уравновешенность функции (сильная равновероятность);
- свойство функции быть дефекта нуль (без запрета);
- свойство функции быть без потери информации.

В данной работе рассмотрен и доказан критерий совершенной уравновешенности дискретных функций над конечным алфавитом, использующий понятия и терминологию теории информации. Полученный результат характеризует процессы обработки информации, осуществляемые неавтономными регистрами сдвига с фильтрующими функциями.

2 Дискретные функции

Пусть $A = \{a_1, \dots, a_p\}$ — конечное множество символов (алфавит) мощности p , $p \geq 2$, и $\mathbb{N} = \{1, 2, \dots\}$ — множество натуральных чисел. Для произвольного натурального числа n через $A^n = \underbrace{A \times \dots \times A}_n$ будем обозначать множество слов

длины n (n -наборов) в алфавите A . Для произвольных натуральных чисел n и m через $\mathcal{F}_{n,m,p}$ будем обозначать множество всех функций вида $f: A^n \rightarrow A^m$. Переменные x_1 и x_n для функции $f(x_1, \dots, x_n)$ будем называть крайними. Функция $f(x_1, \dots, x_n)$ из $\mathcal{F}_{n,1,p}$ называется перестановочной по переменной x_i , $1 \leq i \leq n$, если для любого набора $c = (c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_n) \in A^{n-1}$ функция $\varphi_c(x_i) = f(c_1, \dots, c_{i-1}, x_i, c_{i+1}, \dots, c_n)$

является перестановкой элементов алфавита A .

Для произвольной функции $f \in \mathcal{F}_{n,1,p}$ и произвольного натурального числа m через f_m обозначим функцию из $\mathcal{F}_{n,m,p}$ вида:

$$\begin{aligned} f_m(x_1, \dots, x_{m+n-1}) &= \\ &= (f(x_1, \dots, x_n), f(x_2, \dots, x_{n+1}), \dots \\ &\quad \dots, f(x_m, \dots, x_{m+n-1})). \end{aligned}$$

Для конечного множества S через $\#S$ будем обозначать его мощность.

3 Взаимная энтропия

Для случайной величины ξ , принимающей значения из A в соответствии с законом распределения \mathcal{P} , будем пользоваться обозначением $\xi \in \mathcal{P}A$. В частности, для равномерного распределения далее используется символ \mathcal{U} .

* Работа выполнена при поддержке РФФИ (проект 16-01-00470-А).

¹ Московский государственный университет им. М. В. Ломоносова, logol@iisi.msu.ru

Энтропия Шеннона (далее — энтропия, см. [3]) случайной величины ξ (или распределения $\mathcal{P} = \{\Pr[\xi = x], x \in S\}$) определяется соотношением:

$$H(\xi) = - \sum_{x \in S} \Pr[\xi = x] \log_2 \Pr[\xi = x],$$

где всегда полагают $0 \cdot \log_2 0 = 0$. Для пары случайных величин ξ и η , принимающих значения в множествах S и T соответственно, известны [3] следующие виды энтропии:

– совместная энтропия:

$$\begin{aligned} H(\xi, \eta) &= \\ &= - \sum_{x \in S} \sum_{y \in T} \Pr[\xi = x, \eta = y] \log_2 \Pr[\xi = x, \eta = y]; \end{aligned}$$

– условная энтропия случайной величины η относительно случайной величины ξ :

$$\begin{aligned} H(\eta|\xi) &= \\ &= - \sum_{x \in S} \sum_{y \in T} \Pr[\xi = x, \eta = y] \log_2 \Pr[\eta = y|\xi = x]. \end{aligned}$$

Далее понадобится величина, называемая средней взаимной информацией случайных величин ξ и η :

$$\begin{aligned} I(\eta|\xi) &= I(\xi|\eta) = H(\xi) + H(\eta) - H(\xi, \eta) = \\ &= H(\xi) - H(\xi|\eta) = H(\eta) - H(\eta|\xi). \end{aligned}$$

Средняя взаимная информация характеризует среднее значение информации о случайной величине η , содержащейся в случайной величине ξ (и наоборот).

4 Совершенно уравновешенные функции

Свойство совершенной уравновешенности первоначально исследовалось для булева случая ($p = 2$). Был установлен ряд свойств совершенно уравновешенных функций. Подробнее с результатами и библиографией в данной области можно ознакомиться в [4].

Определение 1. Функция $f \in \mathcal{F}_{n,1,p}$ называется совершенно уравновешенной, если для любого натурального m и для любого набора $y \in A^m$ выполнено равенство:

$$\#f_m^{-1}(y) = p^{n-1}.$$

Пример 1. Функция $f \in \mathcal{F}_{n,1,p}$, перестановочная по переменной x_1 (либо перестановочная по переменной x_n), является совершенно уравновешенной функцией.

Необходимо отметить, что свойство перестановочности функции по некоторой переменной

в булевом случае соответствует линейности булевой функции по данной переменной.

5 Средняя на символ взаимная информация

Пусть $f \in \mathcal{F}_{n,1,p}$ и последовательность случайных величин $\{\xi_m\}_{m=1}^{\infty}$, $\xi_m = (\xi_{m,1}, \dots, \xi_{m,m+n-1})$, удовлетворяет условию:

$$\xi_m \in_{\mathcal{U}} A^{m+n-1}, \quad m = 1, 2, \dots$$

Для каждой ξ_m определим случайную величину:

$$\eta_m^f = f_m(\xi_m). \quad (1)$$

Справедливо следующее утверждение.

Предложение 1. Для любого натурального числа m случайная величина η_m^f удовлетворяет условию:

$$\eta_m^f \in_{\mathcal{U}} A^m$$

тогда и только тогда, когда $f \in \mathcal{F}_{n,1,p}$ — совершенно уравновешенная функция.

Доказательство. Непосредственно вытекает из определения совершенно уравновешенной функции и соотношения (1). \square

Определим среднюю на символ взаимную информацию для случайных величин ξ_m и η_m^f как

$$i_m(f) = m^{-1} I(\eta_m^f|\xi_m).$$

Теорема 1. Пусть $f \in \mathcal{F}_{n,1,p}$, $\xi_m \in_{\mathcal{U}} A^{m+n-1}$ и $\eta_m^f = f_m(\xi_m)$, $m = 1, 2, \dots$. Функция f — совершенно уравновешенная тогда и только тогда, когда для любого натурального m выполнено $i_m(f) = \log_2 p$.

Доказательство. Заметим, что для любых $u \in A^{m+n-1}$ и $v \in A^m$ справедливы следующие соотношения:

$$\begin{aligned} \pi_u &= \Pr[\xi_m = u] = p^{-(m+n-1)}; \\ \pi_{v/u} &= \Pr[\eta_m^f = v|\xi_m = u] = \\ &= \begin{cases} 1, & \text{если } u \in f_m^{-1}(v); \\ 0, & \text{если } u \notin f_m^{-1}(v); \end{cases} \\ \pi_{v,u} &= \Pr[\eta_m^f = v, \xi_m = u] = \pi_u \pi_{v/u} = \\ &= \begin{cases} p^{-(m+n-1)}, & \text{если } u \in f_m^{-1}(v); \\ 0, & \text{если } u \notin f_m^{-1}(v); \end{cases} \\ \tilde{\pi}_v &= \sum_{u \in A^{m+n-1}} \pi_{v,u} = p^{-(m+n-1)} \#f_m^{-1}(v). \end{aligned}$$

Воспользовавшись этими соотношениями, легко показать справедливость следующей цепочки равенств:

$$\begin{aligned}
 i_m(f) &= m^{-1} I(\eta_m^f | \xi_m) = \\
 &= m^{-1} [H(\eta_m^f) - H(\eta_m^f | \xi_m)] = \\
 &= m^{-1} \sum_{v \in A^m} \sum_{u \in f_m^{-1}(v)} \pi_{v,u} [\log_2 \pi_{v/u} - \log_2 \tilde{\pi}_v] = \\
 &= m^{-1} \sum_{v \in A^m} \sum_{u \in f_m^{-1}(v)} p^{-(m+n-1)} \times \\
 &\times [(m+n-1) \log_2 p - \log_2 (\#f_m^{-1}(v))] = \\
 &= m^{-1} \sum_{v \in A^m} (\#f_m^{-1}(v)) p^{-(m+n-1)} \times \\
 &\times [(m+n-1) \log_2 p - \log_2 (\#f_m^{-1}(v))] = \\
 &= m^{-1} \left[(m+n-1) \log_2 p - \right. \\
 &\left. - p^{-(m+n-1)} \sum_{v \in A^m} (\#f_m^{-1}(v)) \log_2 (\#f_m^{-1}(v)) \right]. \quad (2)
 \end{aligned}$$

Пусть f — совершенно уравновешенная функция из $\mathcal{F}_{n,1,p}$. Тогда для любого натурального числа m и любого набора $v \in A^m$ справедливо равенство

$$\#f_m^{-1}(v) = p^{n-1}. \quad (3)$$

Воспользовавшись соотношениями (2) и (3), для произвольного m получаем $i_m(f) = \log_2 p$.

Докажем теперь обратное утверждение. Пусть $f \in \mathcal{F}_{n,1,p}$ и для любого натурального числа m выполняется равенство $i_m(f) = \log_2 p$. Тогда из соотношения (2) следует, что

$$-\sum_{v \in A^m} \frac{\#f_m^{-1}(v)}{p^{m+n-1}} \log_2 \left(\frac{\#f_m^{-1}(v)}{p^{m+n-1}} \right) = m \log_2 p.$$

Кроме того,

$$\sum_{v \in A^m} \frac{\#f_m^{-1}(v)}{p^{m+n-1}} = 1.$$

Следовательно, энтропия распределения $\{\#f_m^{-1}(v)/p^{m+n-1}\}_{v \in A^m}$ имеет максимальное значение. Это возможно (см. [3]) лишь в случае, когда

$$\frac{\#f_m^{-1}(v)}{p^{m+n-1}} = p^{-m}$$

для любого $v \in A^m$, т.е. $\#f_m^{-1}(v) = p^{n-1}$. Поскольку в приведенных рассуждениях m — произвольное, то функция f — совершенно уравновешенная. \square

Следствие 1. Если функция $f \in \mathcal{F}_{n,1,p}$ не является совершенно уравновешенной, то существует такое натуральное число m_0 , что $i_{m_0}(f) < \log_2 p$.

Доказательство. Непосредственно вытекает из определения совершенно уравновешенной функции и утверждения теоремы 1. \square

Характеризации классов дискретных функций (т.е. необходимые и достаточные условия принадлежности функции данному классу) имеют разнообразную математическую природу. В соответствующих формулировках используются алгебраические, комбинаторные, метрические, спектральные и другие параметры этих функций. Наличие нескольких характеристик для конкретного класса дискретных функций позволяет рассматривать их с различных позиций, а также оценивать их особенности с точки зрения возможных приложений.

6 Полиномиальный алгоритм обращения совершенно уравновешенных функций

Пусть f — произвольная функция из $\mathcal{F}_{n,1,p}$ и $F(f) = \{f_m \in \mathcal{F}_{n,m,p}, m \in \mathbb{N}\}$ — счетное семейство функций, порожденное f . Задачу обращения (см. [5]) функций из семейства $F(f)$ можно сформулировать следующим образом: для произвольного натурального m по известному набору $y = (y_1, y_2, \dots, y_m)$ из $f_m(A^{m+n-1})$ найти по крайней мере один набор $x = (x_1, x_2, \dots, x_{m+n-1})$ из $f_m^{-1}(y)$. Рассмотрим простейший алгоритм решения этой задачи.

Алгоритм \mathfrak{A}_f обращения функций семейства $F(f)$. Пусть $m \in \mathbb{N}$, $y = (y_1, y_2, \dots, y_m) \in f_m(A^{m+n-1})$. Строим последовательно m наборов из p непересекающихся подмножеств A^n :

$$\begin{aligned}
 &(S_1(1), S_1(2), \dots, S_1(p)), \\
 &(S_2(1), S_2(2), \dots, S_2(p)), \\
 &\quad \vdots \quad \quad \quad \ddots \quad \quad \quad \vdots \\
 &(S_m(1), S_m(2), \dots, S_m(p));
 \end{aligned}$$

если $i = 1$, то

$$\begin{aligned}
 S_1(1) &= \\
 &= \{(u_1, \dots, u_{n-1}, a_1) \in A^n : f(u_1, \dots, u_{n-1}, a_1) = y_1\};
 \end{aligned}$$

\vdots

$$\begin{aligned}
 S_1(p) &= \\
 &= \{(u_1, \dots, u_{n-1}, a_p) \in A^n : f(u_1, \dots, u_{n-1}, a_p) = y_1\};
 \end{aligned}$$

$$S_1 = \bigcup_{j=1}^p S_1(j) \neq \emptyset;$$

если набор подмножеств $(S_{i-1}(1), \dots, S_{i-1}(p))$, $1 < i \leq m$, построен, то

$$\begin{aligned}
 S_i(1) &= \{(v_2, \dots, v_n, a_1) : \exists v = \\
 &= (v_1, v_2, \dots, v_n) \in S_{i-1} (f(v_2, \dots, v_n, a_1) = y_i)\}; \\
 &\quad \vdots \\
 S_i(p) &= \{(v_2, \dots, v_n, a_p) : \exists v = \\
 &= (v_1, v_2, \dots, v_n) \in S_{i-1} (f(v_2, \dots, v_n, a_p) = y_i)\}; \\
 S_i &= \bigcup_{j=1}^p S_i(j) \neq \emptyset, \quad i = 2, 3, \dots, m;
 \end{aligned}$$

вычисление набора $x \in f_m^{-1}(y)$: для y_m , поскольку $S_m \neq \emptyset$, существует

$$(x_m, x_{m+1}, \dots, x_{m+n-1}) \in S_m,$$

такой что $f(x_m, x_{m+1}, \dots, x_{m+n-1}) = y_m$; для y_{m-1} множество $S_{m-1} \neq \emptyset$ и существует

$$(x_{m-1}, x_m, \dots, x_{m+n-2}) \in S_{m-1},$$

такой что $f(x_{m-1}, x_m, \dots, x_{m+n-2}) = y_{m-1}$ и т.д. для $i = m-2, m-3, \dots, 1$; полученные в результате m наборов

$$\begin{aligned}
 &(x_m, x_{m+1}, \dots, x_{m+n-1}), \\
 &(x_{m-1}, x_m, \dots, x_{m+n-2}), \\
 &\quad \vdots \\
 &(x_1, x_2, \dots, x_n)
 \end{aligned}$$

определяют набор $x = (x_1, x_2, \dots, x_{m+n-1}) \in A^{m+n-1}$, такой что $f_m(x) = y$.

Остановимся подробнее на некоторых параметрах описанного выше алгоритма. Основными операциями, используемыми в ходе реализации алгоритма \mathcal{A}_f , являются следующие:

- (1) запись в память (считывание из памяти) n -наборов в алфавите A ;
- (2) переход от набора (v_1, v_2, \dots, v_n) к набору $(v_2, \dots, v_n, a), v_1, v_2, \dots, v_n, a \in A$;
- (3) проверка выполнения соотношения $f(v_2, \dots, v_n, a) = b, v_2, \dots, v_n, a, b \in A$.

Будем называть их единичными операциями (ед. оп.). Используемую память будем представлять в виде массивов ячеек памяти, необходимых для хранения n -наборов в алфавите A , т.е. способных хранить $n \log_2 p$ бит информации.

Обозначим через $T_f: \mathbb{N} \rightarrow \mathbb{N}$ и $M_f: \mathbb{N} \rightarrow \mathbb{N}$ трудоемкость алгоритма \mathcal{A}_f и память, необходимую для его реализации, соответственно.

Из описания алгоритма \mathcal{A}_f легко видеть, что

$$\begin{aligned}
 T_f &\leq C_1(n, p)m \text{ [ед. оп.]}; \\
 M_f &\leq C_2(n, p)m \text{ [бит]}.
 \end{aligned} \tag{4}$$

Очевидно, что алгоритм \mathcal{A}_f является детерминированным полиномиальным (по m) и улучшить асимптотическую оценку (4) не представляется возможным. Действительно, в ходе реализации любого детерминированного алгоритма обращения функций из $F(f)$ необходимо будет по крайней мере просмотреть набор $y = (y_1, y_2, \dots, y_m)$. Нетрудно заметить, что $C_1(n, p) \leq 5p^n + 1$ и $C_2(n, p) \leq p^n n \log_2 p$.

В случае если функция f является совершенно уравновешенной, из утверждения теоремы 1 следует, что набор $y = f_m(x)$ содержит в среднем максимально возможное количество информации о наборе x . Естественно ожидать, что для совершенно уравновешенных функций величины $C_1(n, p)$ и $C_2(n, p)$ могут быть значительно меньшими. Например, воспользовавшись эквивалентностью свойства функций быть без потери информации (см. [4]) свойству совершенной уравновешенности, несложно получить оценку $C_1(n, p) \leq 5p^{n-1} + 1$, $C_2(n, p) \leq p^{n-1} n \log_2 p$.

Замечание 1. Описание алгоритма \mathcal{A}_f легко может быть перенесено на теоретико-автоматную модель с соответствующей заменой единичных операций на операции, соответствующие функциям переходов и выходов конечного автомата. Следовательно, для любой ограниченно-детерминированной функции над конечным алфавитом (см. [6]) существует полиномиальный алгоритм обращения.

Литература

1. Hedlund G. A. Endomorphisms and automorphisms of the shift dynamical system // Math. Syst. Theory, 1969. No. 3. P. 320–375.
2. Сумароков С. Н. Запреты двоичных функций и обратимость для одного класса кодирующих устройств // Обзорные прикладной и промышленной математики, 1994. № 1. С. 33–55.
3. Файнштейн А. Основы теории информации / Пер. с англ. — М.: ИЛ, 1960. 140 с. (Feinstein A. Foundations of information theory. — New York, NY, USA: McGraw-Hill, 1958. 137 p.)
4. Логачев О. А., Сальников А. А., Смышляев С. В., Яценко В. В. Булевы функции в теории кодирования и криптологии. — М.: Ленанд, 2015. 576 с.
5. Goldreich O. Foundations of cryptography. Vol. I: Basic tools. — Cambridge: Cambridge University Press, 2003. 372 p.
6. Кудрявцев В. Б., Алешин С. В., Подколзин А. С. Введение в теорию автоматов. — М.: Наука, 1985. 320 с.

Поступила в редакцию 03.09.18

AN INFORMATION BASED CRITERION FOR PERFECTLY BALANCED FUNCTIONS

O. A. Logachev

Institute of Information Security Issues, M. V. Lomonosov Moscow State University, 1 Michurinsky Pr., Moscow 119192, Russian Federation

Abstract: The class of perfectly balanced functions is important for some areas of mathematics, e. g., combinatorics, coding theory, cryptography, symbolic dynamics, and automata theory. It turns out that perfectly balanced functions provide a suitable mathematical tool for description and studying of convolutional codes, cryptographic primitives, surjective endomorphisms of discrete dynamical systems, and information-lossless finite-state automata. Previously, Hedlund and Sumarokov proved criteria of perfect balancedness of functions, which are related to the property of being defect zero and information-lossless. The present author proves a new criterion of the perfect balancedness property in terms of average mutual information. The author also describes a polynomial-time inverting algorithm for perfectly balanced functions.

Keywords: finite alphabet; discrete function; average mutual information; perfect balancedness; perfectly balanced function; function of defect zero

DOI: 10.14357/19922264180410

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 16-01-00470-A).

References

1. Hedlund, G. A. 1969. Endomorphisms and automorphisms of the shift dynamical system. *Math. Syst. Theory* 3:320–375.
2. Sumarokov, S. N. 1994. Zaprety dvoichnykh funktsiy i obratimost' dlya odnogo klassa kodiruyushchikh ustroystv [Prohibitions of Boolean functions and invertibility for a coding devices class]. *Obozrenie prikladnoy i promyshlennoy matematiki* [Surveys in Applied and Industrial Mathematics] 1:33–55.
3. Feinstein, A. 1958. *Foundations of information theory*. New York, NY: Mcgraw-Hill. 137 p.
4. Logachev, O. A., A. A. Sal'nikov, S. V. Smyshlyaev, and V. V. Yashchenko. 2015. *Bulevy funktsii v teorii kodirovaniya i kriptologii* [Boolean functions in coding theory and cryptography]. Moscow: Lenand. 576 p.
5. Goldreich, O. 2003. *Foundations of cryptography. Vol. I: Basic tools*. Cambridge, U.K.: Cambridge University Press. 372 p.
6. Kudryavcev, V. B., S. V. Aleshin, and A. S. Podkolzin. 1985. *Vvedenie v teoriyu avtomatov* [Introduction to automata theory]. Moscow: Nauka. 320 p.

Received September 3, 2018

Contributor

Logachev Oleg A. (b. 1950) — Candidate of Science(PhD) in physics and mathematics, Head of Department, Institute of Information Security Issues, M. V. Lomonosov Moscow State University, 1 Michurinsky Pr., Moscow 119192, Russian Federation; logol@iisi.msu.ru

НОВЫЕ ПРЕДСТАВЛЕНИЯ ОБОБЩЕННОГО РАСПРЕДЕЛЕНИЯ МИТТАГ-ЛЕФФЛЕРА В ВИДЕ СМЕСЕЙ И ИХ ПРИЛОЖЕНИЯ*

В. Ю. Королев¹, А. К. Горшенин², А. И. Зейфман³

Аннотация: Приведены новые представления обобщенного распределения Миттаг-Леффлера в виде смесей. В частности, показано, что при значениях «обобщающего» параметра, не превосходящих единицы, обобщенное распределение Миттаг-Леффлера является масштабной смесью полунормальных законов, масштабной смесью «обычных» распределений Миттаг-Леффлера или масштабной смесью обобщенных распределений Миттаг-Леффлера с большими значениями характеристического показателя. Во всех случаях приведены явные выражения для смешивающих величин. Полученные представления позволяют предложить новые алгоритмы моделирования случайных величин (с.в.) с обобщенным распределением Миттаг-Леффлера и сформулировать новые предельные теоремы, в которых указанное распределение выступает в качестве предельного.

Ключевые слова: обобщенное распределение Миттаг-Леффлера; масштабная смесь; обобщенное гамма-распределение; полунормальное распределение; устойчивое распределение

DOI: 10.14357/19922264180411

1 Введение

Данная статья продолжает исследования, начатые в работах [1–4]. В статье приведены новые представления обобщенного распределения Миттаг-Леффлера в виде смесей. Это распределение представляет особый интерес как «тяжелохвостая» модель статистических закономерностей, при которых большие значения наблюдаемых характеристик встречаются намного чаще, чем предписывает классическая экспоненциальная модель. Оно возникает в некоторых задачах, связанных с дифференциальными уравнениями дробного порядка в физике, астрономии, финансовой математике и других областях [5–8]. «Обычное» распределение Миттаг-Леффлера традиционно рассматривается вместе с распределением Линника, поскольку характеристическая функция (х.ф.) распределения Линника имеет такой же аналитический вид, как преобразование Лапласа–Стилтьеса (п. Л.–С.) распределения Миттаг-Леффлера. Поэтому эти распределения обладают многими сходными свойствами. В частности, они геометрически устойчивы, так как являются предельными для геометрических случайных сумм независимых одинаково распреде-

ленных с.в. с бесконечными дисперсиями и потому представимы в виде масштабных смесей устойчивых законов, в которых смешивающим выступает распределение Вейбулла. Соответственно, обобщенные распределения Линника и Миттаг-Леффлера представляют собой масштабные смеси устойчивых законов, в которых смешивающим служит обобщенное гамма-распределение.

В данной работе приведены альтернативные представления обобщенного распределения Миттаг-Леффлера в виде смесей. В частности, показано, что при значениях «обобщающего» параметра, не превосходящих единицы, обобщенное распределение Миттаг-Леффлера — это масштабная смесь полунормальных законов, «обычных» распределений Миттаг-Леффлера или обобщенных распределений Миттаг-Леффлера с большими значениями характеристического показателя. Во всех случаях приведены явные выражения для смешивающих величин. Полученные представления позволяют предложить новые алгоритмы моделирования с.в. с обобщенным распределением Миттаг-Леффлера и сформулировать новые предельные теоремы, в которых указанное распределение выступает в качестве предельного.

*Работа выполнена при поддержке РФФИ (проект 17-07-00717).

¹Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Hangzhou Dianzi University, Китай, vkorolev@cs.msu.ru

²Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, agorshenin@frccsc.ru

³Вологодский государственный университет; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Вологодский научный центр Российской академии наук, a_zeifman@mail.ru

Аналогичные результаты относительно обобщенного распределения Линника приведены в [4], где, в частности, показано, что обобщенное распределение Линника — масштабная смесь нормальных законов со смешивающим распределением типа обобщенного распределения Миттаг-Леффлера. Здесь этот результат будет использован для вывода некоторых свойств обобщенного распределения Миттаг-Леффлера.

2 Распределения Миттаг-Леффлера и Линника

Пусть $\alpha \in (0, 1]$ и M_α — неотрицательная с.в. с п. Л.—С.:

$$\psi_\alpha(s) \equiv \mathbb{E} \exp\{-sM_\alpha\} = (1 + s^\alpha)^{-1}, \quad s \geq 0. \quad (1)$$

Распределения с п. Л.—С. (1) принято называть *распределениями Миттаг-Леффлера*. Происхождение этого названия связано с тем, что плотность, соответствующая п. Л.—С. (1), имеет вид:

$$f_\alpha^M(x) = \frac{1}{x^{1-\alpha}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{\alpha n}}{\Gamma(\alpha n + 1)} = -\frac{d}{dx} E_\alpha(-x^\alpha), \quad x \geq 0,$$

где $E_\alpha(z)$ — функция Миттаг-Леффлера индекса α , определяемая как степенной ряд

$$E_\alpha(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + 1)}, \quad \alpha > 0, \quad z \in \mathbb{Z}.$$

Функция распределения (ф.р.), соответствующая плотности $f_\alpha^M(x)$, будет обозначаться $F_\alpha^M(x)$.

При $\alpha = 1$ распределение Миттаг-Леффлера превращается в стандартное показательное распределение: $M_1 \stackrel{d}{=} W_1$. Но при $\alpha < 1$ плотность (1) имеет хвост, убывающий степенным образом: если $0 < \alpha < 1$, то $\lim_{x \rightarrow \infty} x^{\alpha+1} f_\alpha^M(x) = \pi^{-1} \Gamma(\alpha + 1) \sin \pi \alpha$ (см., например, [8]).

Моменты с.в. с распределением Миттаг-Леффлера порядков $\beta \geq \alpha$ бесконечны, но если $0 < \beta < \alpha < 1$, то $\mathbb{E} M_\alpha^\beta = \Gamma(1 + \beta/\alpha) \Gamma(1 - \beta/\alpha)$.

Пусть $\nu > 0$, $\alpha \in (0, 1]$. Распределение неотрицательной с.в. $M_{\alpha, \nu}$, соответствующее п. Л.—С.

$$\psi_{\alpha, \nu}(s) \equiv \mathbb{E} e^{-sM_{\alpha, \nu}} = (1 + s^\alpha)^{-\nu}, \quad s \geq 0,$$

называется *обобщенным распределением Миттаг-Леффлера* (см. [9, 10] и ссылки в этих работах).

Распределения с х.ф. $f_\alpha^L(t) = (1 + |t|^\alpha)^{-1}$, $t \in \mathbb{R}$, где $0 < \alpha \leq 2$, принято называть *распределениями Линника* (в работе [11] предложено альтернативное менее употребительное название α -Laplace

distribution). Они были введены Ю. В. Линником в 1953 г. [12]. При $\alpha = 2$ распределение Линника превращается в распределение Лапласа, соответствующее плотности

$$f^\Lambda(x) = \frac{1}{2} e^{-|x|}, \quad x \in \mathbb{R}. \quad (2)$$

Лапласовская с.в. с плотностью (2) и ее ф.р. будут соответственно обозначаться Λ и $F^\Lambda(x)$.

Случайная величина, имеющая распределение Линника с параметром α , ее ф.р. и плотность будут соответственно обозначаться L_α , F_α^L и f_α^L . При этом $F_\alpha^L(x) \equiv F^\Lambda(x)$, $x \in \mathbb{R}$.

Распределения Линника обладают многими интересными свойствами, которые описаны, например, в работах [13–17]. Абсолютные моменты порядков $\beta < \alpha$ с.в. L_α имеют вид:

$$\mathbb{E} |L_\alpha|^\beta = \frac{2^\beta}{\sqrt{\pi}} \frac{\Gamma(1 + \beta/\alpha) \Gamma((1 + \beta)/2) \Gamma(1 - \beta/\alpha)}{\Gamma(1 - \beta/2)}.$$

В работе [18] показано, что при $0 < \alpha < 2$ хвосты распределения Линника убывают степенным образом:

$$\lim_{x \rightarrow \infty} x^\alpha [1 - F_\alpha^L(x)] = \pi^{-1} \Gamma(\alpha) \sin\left(\frac{\pi \alpha}{2}\right).$$

В работах [2, 3, 19, 20] получены разнообразные представления распределений Линника в виде смесей. Некоторые из этих представлений будут приведены и использованы ниже.

В работе [20] замечено, что *обобщенные распределения Линника*, задаваемые х.ф.

$$\phi_{\alpha, \nu, \theta}(t) = (1 + e^{-i\theta \operatorname{sgn} t} |t|^\alpha)^{-\nu}, \quad t \in \mathbb{R}, \quad |\theta| \leq \min\left\{\frac{1}{2} \pi \alpha, \pi - \frac{1}{2} \pi \alpha\right\}, \quad \nu > 0, \quad (3)$$

играют видную роль в некоторых характеристизационных задачах математической статистики. Среди работ, посвященных свойствам этих распределений и их применениям, следует упомянуть [5, 6, 13, 19, 21–25].

В данной работе будут рассматриваться симметричные распределения, для которых в соотношении (3) $\theta = 0$.

3 Вспомогательные сведения

В дальнейшем удобнее вести изложение не в терминах распределений, а в терминах с.в., предполагая, что все они заданы на одном вероятностном пространстве $(\Omega, \mathfrak{A}, \mathbb{P})$.

Случайная величина со стандартной показательной ф.р. будет обозначаться W_1 :

$$P(W_1 < x) = [1 - e^{-x}] \mathbf{1}(x \geq 0)$$

(здесь и далее символ $\mathbf{1}(C)$ обозначает индикатор множества C). Случайная величина со стандартной нормальной ф.р. $\Phi(x)$ будет обозначаться X :

$$P(X < x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz, \quad x \in \mathbb{R}.$$

Функция распределения и плотность строго устойчивого распределения с характеристическим показателем α и параметром формы θ , определяемого характеристической функцией

$$f_{\alpha,\theta}(t) = \exp\{-|t|^\alpha \exp\{-(1/2)i\pi\theta \operatorname{sign} t\}\}, \quad t \in \mathbb{R},$$

где $0 < \alpha \leq 2$, $|\theta| \leq \min\{1, 2/\alpha - 1\}$, будут соответственно обозначаться $G_{\alpha,\theta}(x)$ и $g_{\alpha,\theta}(x)$ (см., например, [26]). Любую с.в. с ф.р. $G_{\alpha,\theta}(x)$ будем обозначать $S_{\alpha,\theta}$. Симметричным строго устойчивым распределениям соответствует значение $\theta = 0$ и х.ф. $f_{\alpha,0}(t) = e^{-|t|^\alpha}$, $t \in \mathbb{R}$. Отсюда несложно видеть, что $S_{2,0} \stackrel{d}{=} \sqrt{2}X$.

Односторонним строго устойчивым законам, сосредоточенным на неотрицательной полуоси, соответствуют значения $\theta = 1$ и $0 < \alpha \leq 1$. Пары $\alpha = 1$, $\theta = \pm 1$ отвечают распределениям, вырожденным в ± 1 соответственно. Остальные устойчивые распределения абсолютно непрерывны. Явные выражения устойчивых плотностей в терминах элементарных функций отсутствуют за четырьмя исключениями (нормальный закон ($\alpha = 2$, $\theta = 0$), распределение Коши ($\alpha = 1$, $\theta = 0$), распределение Леви ($\alpha = 1/2$, $\theta = 1$) и распределение, симметричное к распределению Леви ($\alpha = 1/2$, $\theta = -1$)). Выражения устойчивых плотностей в терминах функций Фокса (обобщенных G -функций Мейера) можно найти в [27, 28].

Хорошо известно, что если $0 < \alpha < 2$, то $E|S_{\alpha,\theta}|^\beta < \infty$ для любого $\beta \in (0, \alpha)$, но моментов с.в. $S_{\alpha,\theta}$ порядков $\beta > \alpha$ не существует (см., например, [26]). Несмотря на отсутствие явных выражений плотностей устойчивых распределений в терминах элементарных функций, можно показать [29], что для $0 < \beta < \alpha < 2$

$$E|S_{\alpha,0}|^\beta = \frac{2^\beta}{\sqrt{\pi}} \frac{\Gamma((\beta+1)/2) \Gamma(1-\beta/\alpha)}{\Gamma(2/\beta-1)} \quad (4)$$

и для $0 < \beta < \alpha < 1$

$$E S_{\alpha,1}^\beta = \frac{\Gamma(1-\beta/\alpha)}{\Gamma(1-\beta)}. \quad (5)$$

Символы $\stackrel{d}{=}$ и \implies будут соответственно обозначать совпадение распределений и сходимость по распределению.

Лемма 1 [26, теорема 3.3.1]. Пусть $\alpha \in (0, 2]$, $\alpha' \in (0, 1]$. Тогда $S_{\alpha\alpha',0} \stackrel{d}{=} S_{\alpha,0} S_{\alpha',1}^{1/\alpha}$, где с.в. в правой части независимы.

Следствие 1. Симметричное строго устойчивое распределение с характеристическим показателем α является масштабной смесью нормальных законов, в которой смешивающим служит одностороннее строго устойчивое распределение с характеристическим показателем $\alpha/2$:

$$S_{\alpha,0} \stackrel{d}{=} X \sqrt{2S_{\alpha/2,1}}, \quad (6)$$

где с.в. в правой части независимы.

Случайная величина, имеющая гамма-распределение с параметром формы $\nu > 0$ и параметром масштаба $\lambda > 0$, будет обозначаться $G_{\nu,\lambda}$,

$$P(G_{\nu,\lambda} < x) = \int_0^x g(z; \nu, \lambda) dz,$$

где

$$g(x; \nu, \lambda) = \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}, \quad x \geq 0,$$

В этих обозначениях, очевидно, $G_{1,1} \stackrel{d}{=} W_1$.

Гамма-распределение — это частный случай обобщенных гамма-распределений, введенных в работе [30] как единый класс, одновременно содержащий гамма-распределение и распределение Вейбулла. *Обобщенное гамма-распределение* — это абсолютно непрерывное распределение, плотность которого имеет вид:

$$\bar{g}(x; \nu, \alpha, \lambda) = \frac{|\alpha| \lambda^\nu}{\Gamma(\nu)} x^{\alpha\nu-1} e^{-\lambda x^\alpha}, \quad x \geq 0,$$

где $\alpha \in \mathbb{R}$, $\lambda > 0$, $\nu > 0$.

Случайная величина с плотностью $\bar{g}(x; \nu, \alpha, \lambda)$ будет обозначаться $\bar{G}_{\nu,\alpha,\lambda}$. Легко видеть, что

$$\bar{G}_{\nu,\alpha,\mu} \stackrel{d}{=} G_{\nu,\mu}^{1/\alpha} \stackrel{d}{=} \mu^{-1/\alpha} G_{\nu,1}^{1/\alpha} \stackrel{d}{=} \mu^{-1/\alpha} \bar{G}_{\nu,\alpha,1}.$$

Для с.в. с *распределением Вейбулла*, частным случаем обобщенных гамма-распределений, соответствующим плотности $\bar{g}(x; 1, \alpha, 1)$ и ф.р. $[1 - e^{-x^\alpha}] \mathbf{1}(x \geq 0)$ с $\alpha > 0$, будет использовано особое обозначение W_α . Таким образом, $G_{1,1} \stackrel{d}{=} \bar{G}_{1,1} \stackrel{d}{=} W_1$. Очевидно, $W_1^{1/\alpha} \stackrel{d}{=} W_\alpha$.

Несложно убедиться, что если $\gamma > 0$ и $\gamma' > 0$, то

$$P(W_{\gamma'}^{1/\gamma} \geq x) = P(W_{\gamma'} \geq x^\gamma) = e^{-x^{\gamma\gamma'}} = P(W_{\gamma\gamma'} \geq x), \quad x \geq 0,$$

т. е. при любых $\gamma > 0$ и $\gamma' > 0$

$$W_{\gamma\gamma'} \stackrel{d}{=} W_{\gamma'}^{1/\gamma}. \quad (7)$$

В статье [31] было показано, что каждое гамма-распределение с параметром формы, не превосходящим единицы, является смешанным показательным. Сформулируем этот результат в виде следующей леммы.

Лемма 2 [31]. *Плотность гамма-распределения $g(x; \nu, \mu)$ с $0 < \nu < 1$ может быть представлена в виде*

$$g(x; \nu, \mu) = \int_0^\infty ze^{-zx} p(z; \nu, \mu) dz,$$

где

$$p(z; \nu, \mu) = \frac{\mu^\nu}{\Gamma(1-\nu)\Gamma(\nu)} \frac{\mathbf{1}(z \geq \mu)}{(z-\mu)^\nu z}. \quad (8)$$

Более того, гамма-распределение с параметром формы $\nu > 1$ не может быть представлено в виде смешанного показательного закона.

Лемма 3 [32]. *Для $\nu \in (0, 1)$ пусть $G_{\nu,1}$ и $G_{1-\nu,1}$ — независимые гамма-распределенные с.в. Пусть $\mu > 0$. Тогда плотность $p(z; \nu, \mu)$, определенная выражением (8), соответствует с.в.*

$$Z_{\nu,\mu} = \frac{\mu(G_{\nu,1} + G_{1-\nu,1})}{G_{\nu,1}} \stackrel{d}{=} \mu Z_{\nu,1} \stackrel{d}{=} \mu \left(1 + \frac{1-\nu}{\nu} Q_{1-\nu,\nu} \right),$$

где $Q_{1-\nu,\nu}$ — с.в. с распределением Снедекора–Фишера, соответствующим плотности

$$q(x; 1-\nu, \nu) = \frac{(1-\nu)^{1-\nu} \nu^\nu}{\Gamma(1-\nu)\Gamma(\nu)} \frac{1}{x^\nu [\nu + (1-\nu)x]}, \quad x \geq 0.$$

Несложно видеть, что $G_{\nu,1} + G_{1-\nu,1} \stackrel{d}{=} W_1$. Однако числитель и знаменатель в определении с.в. $Z_{\nu,\mu}$ не являются независимыми с.в.

Фактически леммы 2 и 3 означают, что если $\nu \in (0, 1)$, то

$$G_{\nu,\mu} \stackrel{d}{=} W_1 Z_{\nu,\mu}^{-1}, \quad (9)$$

где с.в. в правой части независимы.

Следующее утверждение уже стало фольклором. Без претензий на первенство его доказательство приведено в [3] как упражнение.

Лемма 4. *При каждом $\delta \in (0, 1]$ распределение Миттаг-Леффлера с параметром δ является масштабной смесью одностороннего устойчивого закона, в которой смешивающее распределение — распределение Вейбулла с параметром $\delta/2$:*

$$M_\delta \stackrel{d}{=} S_{\delta,1} W_\delta \stackrel{d}{=} S_{\delta,1} \sqrt{W_{\delta/2}},$$

где с.в. в правой части независимы.

Пусть $\rho \in (0, 1)$. В статье [24] было показано, что функция

$$f_\rho^K(x) = \frac{\sin(\pi\rho)}{\pi\rho[x^2 + 2x \cos(\pi\rho) + 1]}, \quad x \in (0, \infty), \quad (10)$$

является плотностью вероятностей на $(0, \infty)$. Случайную величину с плотностью (10) обозначим K_ρ .

Лемма 5 [24]. *Пусть $0 < \delta < \delta' \leq 1$ и $\rho = \delta/\delta' < 1$. Тогда $M_\delta \stackrel{d}{=} M_{\delta'} K_\rho^{1/\delta}$, где с.в. в правой части независимы.*

В статье [3] было показано, что при любом $\delta \in (0, 1)$

$$K_\delta^{1/\delta} \stackrel{d}{=} \frac{S_{\delta,1}}{S'_{\delta,1}}, \quad (11)$$

где $S'_{\delta,1} \stackrel{d}{=} S_{\delta,1}$ и где с.в. в правой части независимы. Таким образом, при $\delta' = 1$ из леммы 5 вытекает

Следствие 2 [3, 24]. Пусть $0 < \delta < 1$. Тогда распределение Миттаг-Леффлера с параметром δ является смешанным показательным, т. е. справедливо представление:

$$M_\delta \stackrel{d}{=} W_1 K_\delta^{1/\delta} \stackrel{d}{=} W_1 \frac{S_{\delta,1}}{S'_{\delta,1}},$$

где с.в. в правой части независимы.

Пусть $0 < \alpha < \alpha' \leq 2$. В статье [19] было показано, что функция

$$f_{\alpha,\alpha'}^Q(x) = \frac{\alpha' \sin(\pi\alpha/\alpha') x^{\alpha-1}}{\pi[1 + x^{2\alpha} + 2x^\alpha \cos(\pi\alpha/\alpha')]}, \quad x > 0,$$

является плотностью вероятностей на $(0, \infty)$. Пусть $Q_{\alpha,\alpha'}$ — с.в. с плотностью $f_{\alpha,\alpha'}^Q(x)$.

Лемма 6 [19]. *Пусть $0 < \alpha < \alpha' \leq 2$. Тогда $L_\alpha \stackrel{d}{=} L_{\alpha'} Q_{\alpha,\alpha'}$, где с.в. в правой части независимы.*

При $\alpha' = 2$ получаем

Следствие 3 [19]. Пусть $0 < \alpha < 2$. Тогда распределение Линника с параметром α является масштабной смесью распределений Лапласа, соответствующих плотности (2): $L_\alpha \stackrel{d}{=} \Lambda Q_{\alpha,2}$, где с.в. в правой части независимы.

В работе [13] было доказано следующее утверждение. Здесь оно уточнено с учетом (7).

Лемма 7 [13]. *При любом $\alpha \in (0, 2]$*

$$L_\alpha \stackrel{d}{=} S_{\alpha,0} W_1^{1/\alpha},$$

где с.в. в правой части независимы.

Лемма 8 [3]. Пусть $\alpha \in (0, 2]$, $\alpha' \in (0, 1]$. Тогда

$$L_{\alpha\alpha'} \stackrel{d}{=} S_{\alpha,0} M_{\alpha'}^{1/\alpha}.$$

Следствие 4 [3]. При каждом $\alpha \in (0, 2]$ распределение Линника с параметром α является масштабной смесью нормальных законов, в которой смешивающее распределение — распределение Миттаг-Леффлера с параметром $\alpha/2$:

$$L_{\alpha} \stackrel{d}{=} X \sqrt{2M_{\alpha/2}}, \quad (12)$$

где с.в. в правой части независимы.

Лемма 9 [3]. При каждом $\alpha \in (0, 2]$ справедливо представление:

$$L_{\alpha} \stackrel{d}{=} \Lambda \sqrt{\frac{S_{\alpha/2,1}}{S'_{\alpha/2,1}}},$$

где с.в. в правой части независимы.

В статье [3] показано, что если $S_{\alpha,1}$ и $S'_{\alpha,1}$ — независимые с.в. с одним и тем же односторонним строго устойчивым распределением с характеристическим показателем $\alpha \in (0, 1)$, то $S_{\alpha,1}/S'_{\alpha,1} \stackrel{d}{=} K_{\alpha}^{1/\alpha} \stackrel{d}{=} Q_{2\alpha,2}^2$, т.е. плотность $p_{\alpha}(x)$ отношения $S_{\alpha,1}/S'_{\alpha,1}$ имеет вид:

$$p_{\alpha}(x) = f_{\alpha,1}^Q(x) = \frac{\sin(\pi\alpha)x^{\alpha-1}}{\pi[1+x^{2\alpha}+2x^{\alpha}\cos(\pi\alpha)]}, \quad x > 0.$$

Лемма 10 [3]. При любом $\delta \in (0, 1]$ распределение Миттаг-Леффлера с параметром δ является масштабной смесью полунормальных законов:

$$M_{\delta} \stackrel{d}{=} |X| \sqrt{2W_1} \frac{S_{\delta,1}}{S'_{\delta,1}},$$

где с.в. в правой части независимы.

4 Представления обобщенных распределений Миттаг-Леффлера и Линника в виде смесей

Представленные здесь теоремы обобщают и уточняют некоторые результаты работ [9, 10, 20, 33, 34]. Некоторые известные результаты сформулируем в виде лемм.

Лемма 11 [13, 20]. Пусть $\alpha, \nu \in (0, 2]$, $\nu > 0$. Тогда

$$L_{\alpha,\nu} \stackrel{d}{=} S_{\alpha,0} G_{\nu,1}^{1/\alpha} \stackrel{d}{=} S_{\alpha,0} \overline{G}_{\nu,\alpha,1}.$$

Из леммы 11 и соотношения (4) получаем

Следствие 5. Для $0 < \beta < \alpha < 2$

$$E|L_{\alpha,\nu}|^{\beta} = \frac{2^{\beta}\Gamma((\beta+1)/2)\Gamma(1-\beta/\alpha)\Gamma(\nu+\beta/\alpha)}{\sqrt{\pi}\Gamma(2/\beta-1)\Gamma(\nu)}.$$

Лемма 12 [9, 10, 34]. Пусть $\delta \in (0, 1]$ и $\nu > 0$. Тогда

$$M_{\delta,\nu} \stackrel{d}{=} S_{\delta,1} \overline{G}_{\nu,\delta,1}.$$

Из леммы 12 и соотношения (5) получаем

Следствие 6. Для $0 < \beta < \delta < 1$

$$E M_{\delta,\nu}^{\beta} = \frac{\Gamma(1-\beta/\delta)\Gamma(\nu+\beta/\delta)}{\Gamma(1-\beta)\Gamma(\nu)}.$$

Из следствия 1 (см. соотношение (6)) вытекает, что для $\nu > 0$ и $\alpha \in (0, 2]$

$$L_{\alpha,\nu} \stackrel{d}{=} X \sqrt{2S_{\alpha/2,1} G_{\nu,1}^{1/\alpha}} \stackrel{d}{=} X \sqrt{2S_{\alpha/2,1} \overline{G}_{\nu,\alpha/2,1}},$$

т.е. обобщенное распределение Линника является масштабной смесью нормальных законов. При этом согласно лемме 12 смешивающим распределением служит обобщенное распределение Миттаг-Леффлера. Таким образом, по аналогии со следствием 4 получаем следующее утверждение.

Теорема 1. Если $\alpha \in (0, 2]$ и $\nu > 0$, то $L_{\alpha,\nu} \stackrel{d}{=} X \sqrt{2M_{\alpha/2,\nu}}$, где с.в. в правой части независимы.

Пусть $\alpha \in (0, 2]$, $\alpha' \in (0, 1]$ и $\nu > 0$. Используя леммы 1 и 11, получаем следующую цепочку соотношений:

$$L_{\alpha\alpha',\nu} \stackrel{d}{=} S_{\alpha\alpha',0} G_{\nu,1}^{1/\alpha} \stackrel{d}{=} S_{\alpha,0} S_{\alpha',1}^{1/\alpha} G_{\nu,1}^{1/\alpha} \stackrel{d}{=} L_{\alpha,\nu} S_{\alpha',1}^{1/\alpha}.$$

Следовательно, справедливо следующее утверждение.

Теорема 2. Пусть $\alpha \in (0, 2]$, $\alpha' \in (0, 1]$ и $\nu > 0$. Тогда обобщенное распределение Линника является масштабной смесью обобщенных распределений Линника с большим характеристическим параметром:

$$L_{\alpha\alpha',\nu} \stackrel{d}{=} L_{\alpha,\nu} S_{\alpha',1}^{1/\alpha},$$

где с.в. в правой части независимы.

Пусть теперь $\nu \in (0, 1]$. Из представления (9) и леммы 7 получаем цепочку соотношений:

$$L_{\alpha,\nu} \stackrel{d}{=} S_{\alpha,0} G_{\nu,1}^{1/\alpha} \stackrel{d}{=} S_{\alpha,0} W_1^{1/\alpha} Z_{\nu,1}^{-1/\alpha} \stackrel{d}{=} S_{\alpha,0} W_{\alpha} Z_{\nu,1}^{-1/\alpha} \stackrel{d}{=} L_{\alpha} Z_{\nu,1}^{-1/\alpha},$$

из которой вытекает следующее утверждение, связывающее обобщенное и «обычное» распределения Линника.

Теорема 3. Если $\nu \in (0, 1)$ и $\alpha \in (0, 2]$, то

$$L_{\alpha,\nu} \stackrel{d}{=} L_{\alpha} Z_{\nu,1}^{-1/\alpha}, \quad (13)$$

где с.в. в правой части независимы. Другими словами, при $\nu \in (0, 1]$ и $\alpha \in (0, 2]$ обобщенное распределение Линника является масштабной смесью обычных распределений Линника.

Из (13) и леммы 9 получаем следующее представление обобщенного распределения Линника в виде масштабной смеси распределений Лапласа:

$$L_{\alpha,\nu} \stackrel{d}{=} \Lambda Z_{\nu,1}^{-1/\alpha} \sqrt{\frac{S_{\alpha/2,1}}{S'_{\alpha/2,1}}}.$$

Более того, из следствия 4 (см. соотношение (12)) вытекает, что при $\nu \in (0, 1)$ и $\alpha \in (0, 2]$

$$L_{\alpha,\nu} \stackrel{d}{=} X Z_{\nu,1}^{-1/\alpha} \sqrt{2M_{\alpha/2}}. \quad (14)$$

Поскольку масштабные смеси нормальных законов идентифицируемы [35], из (14) и теоремы 1 получаем следующее представление обобщенного распределения Миттаг-Леффлера в виде масштабной смеси «обычных» распределений Миттаг-Леффлера.

Теорема 4. Пусть $\nu \in (0, 1)$ и $\delta \in (0, 1]$. Тогда

$$M_{\delta,\nu} \stackrel{d}{=} Z_{\nu,1}^{-1/\delta} M_{\delta},$$

где с.в. в правой части независимы.

Пусть $\delta \in (0, 1]$. Из леммы 12 вытекает, что

$$M_{\delta,\nu} \stackrel{d}{=} S_{\delta,1} \overline{G}_{\nu,\delta,1} \stackrel{d}{=} S_{\delta,1} G_{\nu,1}^{1/\delta}.$$

Теперь есть все инструменты, позволяющие получить аналог теоремы 2 для распределений Миттаг-Леффлера.

Пусть $\alpha \in (0, 2]$, $\alpha' \in (0, 1)$ и $\nu > 0$. Из теоремы 1 вытекает, что $L_{\alpha\alpha',\nu} \stackrel{d}{=} X \sqrt{2M_{\alpha\alpha'/2,\nu}}$ и $L_{\alpha,\nu} \stackrel{d}{=} X \sqrt{2M_{\alpha/2,\nu}}$. Из теоремы 2 вытекает, что

$$\begin{aligned} X \sqrt{2M_{\alpha\alpha'/2,\nu}} &\stackrel{d}{=} L_{\alpha\alpha',\nu} \stackrel{d}{=} \\ &\stackrel{d}{=} L_{\alpha,\nu} S_{\alpha',1}^{1/\alpha} \stackrel{d}{=} X \sqrt{2M_{\alpha/2,\nu}} S_{\alpha',1}^{1/\alpha}. \end{aligned}$$

Следовательно, в силу идентифицируемости масштабных смесей нормальных законов $M_{\alpha\alpha'/2,\nu} \stackrel{d}{=} \stackrel{d}{=} M_{\alpha/2,\nu} S_{\alpha',1}^{2/\alpha}$. Поэтому, переобозначив $\alpha/2 = \delta$, $\alpha' = \delta'$, получаем следующий результат.

Теорема 5. Пусть $\delta \in (0, 1]$, $\delta' \in (0, 1)$ и $\nu > 0$. Тогда

$$M_{\delta\delta',\nu} \stackrel{d}{=} M_{\delta,\nu} S_{\delta',1}^{1/\delta},$$

где с.в. в правой части независимы.

Другими словами, любое обобщенное распределение Миттаг-Леффлера является масштабной смесью обобщенных распределений Миттаг-Леффлера с большим характеристическим параметром.

Комбинируя утверждения теорем 4 и 5, получаем

Следствие 7. Пусть $\delta \in (0, 1]$, $\delta' \in (0, 1)$ и $\nu \in (0, 1)$. Тогда

$$M_{\delta\delta',\nu} \stackrel{d}{=} M_{\delta} \left(\frac{S_{\delta',1}}{Z_{\nu,1}} \right)^{1/\delta},$$

где с.в. в правой части независимы.

Из теоремы 4, лемм 4 и 5 вытекает, что обобщенное распределение Миттаг-Леффлера допускает представление в виде смешанного распределения Вейбулла.

Теорема 6. Если $\nu \in (0, 1)$ и $0 < \delta < \delta' \leq 1$, то

$$M_{\delta,\nu} \stackrel{d}{=} W_{\delta'} S_{\delta',1} \left(\frac{K_{\delta/\delta'}}{Z_{\nu,1}} \right)^{1/\delta},$$

где все с. в. в правой части независимы.

Из теоремы 4 и леммы 10 вытекает, что обобщенное распределение Миттаг-Леффлера допускает представление в виде масштабной смеси полунормальных законов.

Теорема 7. Пусть $\nu \in (0, 1)$ и $\delta \in (0, 1]$. Тогда

$$M_{\delta,\nu} \stackrel{d}{=} |X| \frac{\sqrt{2W_1}}{Z_{\nu,1}^{1/\delta}} \frac{S_{\delta,1}}{S'_{\delta,1}},$$

где с.в. в правой части независимы.

5 Сходимость распределений экстремальных порядковых статистик в выборках случайного объема к обобщенному распределению Миттаг-Леффлера

Хорошо известно, что при достаточно общих условиях распределение Вейбулла может быть предельным для линейно преобразованных экстремальных порядковых статистик. Этот факт вкпе

с теоремой 6 позволяет убедиться, что обобщенное распределение Миттаг-Леффлера может служить предельным для экстремальных порядковых статистик в выборках случайного объема.

В книге [36] предложено описывать эволюцию неоднородных хаотических стохастических процессов при помощи моделей вида обобщенных дважды стохастических пуассоновских процессов (обобщенных процессов Кокса). В соответствии с таким подходом поток информативных событий, каждое из которых генерирует очередное наблюдение, описывается стохастическим точечным процессом $P(U(t))$, где $P(t)$, $t \geq 0$, — однородный пуассоновский процесс с единичной интенсивностью, а $U(t)$, $t \geq 0$, — независимый от $P(t)$ случайный процесс, такой что $U(0) = 0$, $P(U(t) < \infty) = 1$ для любого $t > 0$, траектории $U(t)$ не убывают и непрерывны справа. Процесс $P(U(t))$, $t \geq 0$, называется дважды стохастическим пуассоновским процессом (процессом Кокса) [37].

В рамках такой модели при каждом t с.в. $P(U(t))$ имеет смешанное пуассоновское распределение. Для наглядности рассмотрим ситуацию с дискретным временем t : $U(t) = U(n) = U_n$, $n \in \mathbb{N}$, где $\{U_n\}_{n \geq 1}$ — неограниченно возрастающая последовательность неотрицательных с.в. такая, что $U_{n+1}(\omega) \geq U_n(\omega)$ для каждого $\omega \in \Omega$, $n \geq 1$. При этом асимптотика $n \rightarrow \infty$ может быть интерпретирована как то, что интенсивность потока информативных событий неограниченно возрастает.

Из сделанных выше предположений вытекает, что с.в. U_n независима от стандартного пуассоновского процесса $P(t)$, $t \geq 0$. Для каждого $n \in \mathbb{N}$ положим $N_n = P(U_n)$, $n \geq 1$. Очевидно, что так определенная с.в. N_n имеет смешанное пуассоновское распределение

$$P(N_n = k) = P(P(U_n) = k) = \int_0^\infty e^{-nz} \frac{(nz)^k}{k!} dP(U_n < z), \quad k = 0, 1, \dots$$

Пусть X_1, X_2, \dots — независимые одинаково распределенные с.в. с общей ф.р. $F(x) = P(X_i < x)$, $x \in \mathbb{R}$, $i \geq 1$. Обозначим $\text{lext}(F) = \inf\{x : F(x) > 0\}$. Предположим, что при каждом $k \in \mathbb{N}$ с.в. N_k независима от последовательности X_1, X_2, \dots . В книге [38] доказано следующее утверждение.

Лемма 13. *Предположим, что существуют неограниченно возрастающая последовательность положительных чисел $\{d_n\}_{n \geq 1}$ и неотрицательная с.в. U такие, что $U_n/d_n \Rightarrow U$. Также пусть $\text{lext}(F) > -\infty$ и ф.р. $A_F(x) = F(\text{lext}(F) - x^{-1})$ удовлетворяет условию: при каждом $x > 0$*

$$\lim_{y \rightarrow \infty} \frac{A_F(yx)}{A_F(y)} = x^{-\delta'} \tag{15}$$

для некоторого положительного числа δ' . Тогда существуют числа a_n и b_n такие, что

$$P\left(\min_{1 \leq j \leq N_n} X_j - a_n < b_n x\right) \Rightarrow \left[1 - \int_0^\infty e^{-ux^{\delta'}} dP(U < u)\right] \mathbf{1}(x \geq 0).$$

При этом числа a_n и b_n можно определить как

$$\left. \begin{aligned} a_n &= \text{lext}(F); \\ b_n &= \sup\{x : F(x) \leq d_n^{-1}\} - \text{lext}(F), \quad n \geq 1. \end{aligned} \right\} \tag{16}$$

Теорема 8. *Пусть $\nu \in (0, 1)$, $\delta \in (0, 1)$. Для того чтобы существовали числа $a_n \in \mathbb{R}$ и $b_n > 0$ такие, что*

$$\frac{1}{b_n} \left(\min_{1 \leq j \leq N_n} X_j - a_n\right) \Rightarrow M_{\delta, \nu},$$

достаточно, чтобы были выполнены следующие условия:

- (1) *существует число $\delta' \in (\delta, 1]$ такое, что ф.р. F принадлежит области \min -притяжения распределения Вейбулла с параметром формы $\delta' \in (0, 1]$, т. е. $\text{lext}(F) > -\infty$ и выполнено условие (15);*
- (2) *существует неограниченно возрастающая последовательность $\{d_n\}_{n \geq 1}$ такая, что $U_n/d_n \Rightarrow S_{\delta', 1}^{-\delta'} (K_{\delta/\delta'} Z_{\nu, 1})^{\delta'/\delta}$.*

При этом числа a_n и b_n могут быть определены в соответствии с (16).

Доказательство. Требуемое утверждение является непосредственным следствием леммы 13 и теоремы 6 с учетом соотношения $K_{\delta/\delta'}^{-1} \stackrel{d}{=} K_{\delta/\delta'}$, вытекающего из (11).

6 Сходимость распределений максимальных случайных сумм к обобщенному распределению Миттаг-Леффлера

В этом разделе будет показано, что обобщенное распределение Миттаг-Леффлера может быть предельным для максимальных или минимальных случайных сумм, а также абсолютных величин случайных сумм независимых с.в. с конечными дисперсиями. Основную роль будет играть теоре-

ма 7, устанавливающая возможность представления обобщенного распределения Миттаг-Леффлера в виде масштабной смеси полунормальных распределений.

Рассмотрим независимые, не обязательно одинаково распределенные с.в. X_1, X_2, \dots с $EX_i = 0$ и $0 < \sigma_i^2 = DX_i < \infty, i \in \mathbb{N}$. Для $n \geq 1$ обозначим

$$\overline{S}_n^* = \max_{1 \leq i \leq n} S_i^*; \underline{S}_n^* = \min_{1 \leq i \leq n} S_i^*; \\ B_n^2 = \sigma_1^2 + \dots + \sigma_n^2.$$

Предположим, что с.в. X_1, X_2, \dots удовлетворяют условию Линдберга: для любого $\tau > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{i=1}^n \int_{|x| \geq \tau B_n} x^2 dP(X_i < x) = 0. \quad (17)$$

Как известно, при таких условиях

$$P(\overline{S}_n^* < B_n x) \implies \Psi(x) \equiv P(|X| < x) = 2\Phi(x) - 1; \\ P(\underline{S}_n^* < B_n x) \implies 1 - \Psi(-x).$$

Пусть N_1, N_2, \dots — последовательность неотрицательных целочисленных с.в. таких, что при каждом $n \in \mathbb{N}$ с.в. N_n, X_1, X_2, \dots независимы. Для $n \in \mathbb{N}$ положим

$$S_{N_n}^* = X_1 + \dots + X_{N_n}; \\ \overline{S}_{N_n}^* = \max_{1 \leq i \leq N_n} S_i^*; \underline{S}_{N_n}^* = \min_{1 \leq i \leq N_n} S_i^*$$

(для определенности считаем, что $S_0^* = \overline{S}_0^* = \underline{S}_0^* = 0$).

Пусть $\{d_n\}_{n \geq 1}$ — неограниченно возрастающая последовательность положительных чисел.

Лемма 14 [39]. *Предположим, что с.в. X_1, X_2, \dots и N_1, N_2, \dots удовлетворяют приведенным выше условиям. В частности, пусть выполнено условие Линдберга (17). Более того, пусть $N_n \xrightarrow{P} \infty$. Тогда распределение нормированных экстремальных случайных сумм и абсолютных величин случайных сумм сходятся к некоторым распределениям, т.е. существуют с.в. Y, \overline{Y} и \underline{Y} такие, что*

$$\frac{\overline{S}_{N_n}^*}{d_n} \implies \overline{Y}; \frac{\underline{S}_{N_n}^*}{d_n} \implies \underline{Y}; \frac{|S_{N_n}^*|}{d_n} \implies |Y|$$

тогда и только тогда, когда существует неотрицательная с.в. U такая, что $d_n^{-2} B_{N_n}^2 \implies U$. При этом

$$P(\overline{Y} < x) = P(|Y| < x) = E\Psi\left(\frac{x}{\sqrt{U}}\right); \\ P(\underline{Y} < x) = 1 - E\Psi\left(-\frac{x}{\sqrt{U}}\right), \quad x \in \mathbb{R}.$$

Из леммы 14 и теоремы 7 вытекает следующее утверждение.

Теорема 9. *Пусть $\delta \in (0, 1], \nu \in (0, 1)$. Предположим, что с.в. X_1, X_2, \dots и N_1, N_2, \dots удовлетворяют приведенным выше условиям. В частности, пусть выполнено условие Линдберга (17). Более того, пусть $N_n \xrightarrow{P} \infty$. Тогда следующие утверждения эквивалентны:*

$$\frac{\overline{S}_{N_n}^*}{d_n} \implies M_{\delta, \nu}; \frac{\underline{S}_{N_n}^*}{d_n} \implies -M_{\delta, \nu}; \frac{|S_{N_n}^*|}{d_n} \implies M_{\delta, \nu}; \\ \frac{B_{N_n}^2}{d_n^2} \implies \frac{2W_1}{Z_{\nu, 1}^{2/\delta}} \left(\frac{S_{\delta, 1}^*}{S'_{\delta, 1}}\right)^2.$$

Литература

1. Королев В. Ю., Зейфман А. И., Корчагин А. Ю. Несимметричные двусторонние распределения Миттаг-Леффлера как предельные законы для случайных сумм независимых случайных величин с конечными дисперсиями // Статистические методы оценивания и проверки гипотез. — Пермь: Пермский гос. ун-т, 2016. Т. 27. С. 69–89.
2. Korolev V. Yu., Zeifman A. I. A note on mixture representations for the Linnik and Mittag-Leffler distributions and their applications // J. Math. Sci., 2017. Vol. 218. No. 3. P. 314–327.
3. Korolev V. Yu., Zeifman A. I. Convergence of statistics constructed from samples with random sizes to the Linnik and Mittag-Leffler distributions and their generalizations // J. Korean Stat. Soc., 2017. Vol. 46. No. 2. P. 161–181.
4. Korolev V. Yu., Gorshenin A. K., Zeifman A. I. On mixture representations for the generalized Linnik distribution and their applications in limit theorems // arXiv, 2018.
5. Mittnik S., Rachev S. T. Modeling asset returns with alternative stable distributions // Economet. Rev., 1993. Vol. 12. P. 261–330.
6. Kotz S., Kozubowski T. J., Podgorski K. The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance. — Boston, MA, USA: Birkhauser, 2001. 349 p.
7. Gorenflo R., Mainardi F. Continuous time random walk, Mittag-Leffler waiting time and fractional diffusion: Mathematical aspects // Anomalous transport: Foundations and applications / Eds. R. Klages, G. Radons, I. M. Sokolov. — Weinheim, Germany: Wiley-VCH, 2008. P. 93–127.
8. Gorenflo R., Kilbas A. A., Mainardi F., Rogosin S. V. Mittag-Leffler functions, related topics and applications. — Berlin/New York: Springer, 2014. 443 p.
9. Jose K. K., Uma P., Lekshmi V. S., Haubold H. J. Generalized Mittag-Leffler distributions and processes for applications in astrophysics and time series modeling // Astrophysics Space, 2010. Iss. 202559. P. 79–92.

10. *Mathai A. M., Haubold H. J.* Matrix-variate statistical distributions and fractional calculus // *Fract. Calc. Appl. Anal.*, 2011. Vol. 14. No. 1. P. 138–155.
11. *Pillai R. N.* Semi- α -Laplace distributions // *Commun. Stat. Theory*, 1985. Vol. 14. P. 991–1000.
12. *Линник Ю. В.* Линейные формы и статистические критерии. I, II // *Украинский математический ж.*, 1953. Т. 5. Вып. 2. С. 207–243; Вып. 3. С. 247–290.
13. *Devroye L.* A note on Linnik's distribution // *Stat. Probabil. Lett.*, 1990. Vol. 9. P. 305–306.
14. *Anderson D. N.* A multivariate Linnik distribution // *Stat. Probabil. Lett.*, 1992. Vol. 14. P. 333–336.
15. *Lin G. D.* Characterizations of the Laplace and related distributions via geometric compound // *Sankhya Ser. A*, 1994. Vol. 56. P. 1–9.
16. *Kotz S., Ostrovskii I. V., Hayfavi A.* Analytic and asymptotic properties of Linnik's probability densities, I // *J. Math. Anal. Appl.*, 1995. Vol. 193. P. 353–371.
17. *Kotz S., Ostrovskii I. V., Hayfavi A.* Analytic and asymptotic properties of Linnik's probability densities, II // *J. Math. Anal. Appl.*, 1995. Vol. 193. P. 497–521.
18. *Jacques C., Rémillard B., Theodorescu R.* Estimation of Linnik law parameters // *Statistics Risk Modeling*, 1999. Vol. 17. No. 3. P. 213–236.
19. *Kotz S., Ostrovskii I. V.* A mixture representation of the Linnik distribution // *Stat. Probabil. Lett.*, 1996. Vol. 26. P. 61–64.
20. *Pakes A. G.* Mixture representations for symmetric generalized Linnik laws // *Stat. Probabil. Lett.*, 1998. Vol. 37. P. 213–221.
21. *Anderson D. N., Arnold B. C.* Linnik distributions and processes // *J. Appl. Probab.*, 1993. Vol. 30. P. 330–340.
22. *Jayakumar K., Kalyanaraman K., Pillai R. N.* α -Laplace processes // *Math. Comput. Model.*, 1995. Vol. 22. P. 109–116.
23. *Baringhaus L., Grubel R.* On a class of characterization problems for random convex combinations // *Ann. I. Stat. Math.*, 1997. Vol. 49. P. 555–567.
24. *Kozubowski T. J.* Mixture representation of Linnik distribution revisited // *Stat. Probabil. Lett.*, 1998. Vol. 38. P. 157–160.
25. *Lin G. D.* A note on the Linnik distributions // *J. Math. Anal. Appl.*, 1998. Vol. 217. P. 701–706.
26. *Золотарев В. М.* Одномерные устойчивые распределения. — Теория вероятностей и математическая статистика сер. — М.: Наука, 1983. 304 с.
27. *Schneider W. R.* Stable distributions: Fox function representation and generalization // *Stochastic processes in classical and quantum systems* / Eds. S. Albeverio, G. Casati, D. Merlini. — Berlin: Springer, 1986. P. 497–511.
28. *Uchaikin V. V., Zolotarev V. M.* Chance and stability. — Utrecht: VSP, 1999. 596 p.
29. *Korolev V. Yu.* Product representations for random variables with the Weibull distributions and their applications // *J. Math. Sci.*, 2016. Vol. 218. No. 3. P. 298–313.
30. *Stacy E. W.* A generalization of the gamma distribution // *Ann. Math. Stat.*, 1962. Vol. 33. P. 1187–1192.
31. *Gleser L. J.* The gamma distribution as a mixture of exponential distributions // *Am. Stat.*, 1989. Vol. 43. P. 115–117.
32. *Королев В. Ю.* Аналогии теоремы Глезера для отрицательных биномиальных и обобщенных гамма-распределений и некоторые их приложения // *Информатика и её применения*, 2017. Т. 11. Вып. 3. С. 2–17.
33. *Lim S. C., Teo L. P.* Analytic and asymptotic properties of multivariate generalized Linnik's probability densities // *J. Fourier Anal. Appl.*, 2010. Vol. 16. Iss. 5. P. 715–747.
34. *Mathai A. M.* Some properties of Mittag-Leffler functions and matrix-variate analogues: A statistical perspective // *Fract. Calc. Appl. Anal.*, 2010. Vol. 13. No. 2. P. 113–132.
35. *Teicher H.* Identifiability of mixtures // *Ann. Math. Stat.*, 1961. Vol. 32. P. 244–248.
36. *Gnedenko B. V., Korolev V. Yu.* Random summation: Limit theorems and applications. — Boca Raton, FL, USA: CRC Press, 1996. 288 p.
37. *Grandell J.* Doubly stochastic Poisson processes. — Lecture notes in mathematics book ser. — Berlin–Heidelberg–New York: Springer, 1976. Vol. 529. 244 p.
38. *Королев В. Ю., Соколов И. А.* Математические модели неоднородных потоков экстремальных событий. — М.: ТОРУС ПРЕСС, 2008. 192 с.
39. *Королев В. Ю.* Сходимость случайных последовательностей с независимыми случайными индексами. I // *Теория вероятностей и ее применения*, 1994. Т. 39. № 2. С. 313–333.

Поступила в редакцию 15.10.18

NEW MIXTURE REPRESENTATIONS OF THE GENERALIZED MITTAG-LEFFLER DISTRIBUTION AND THEIR APPLICATIONS

V. Yu. Korolev^{1,2,3}, A. K. Gorshenin^{1,2}, and A. I. Zeifman^{2,4,5}

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

²Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China

⁴Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation

⁵Vologda Research Center of the Russian Academy of Sciences, 56-A Gorky Str., Vologda 160001, Russian Federation

Abstract: The article provides new mixture representations for the generalized Mittag-Leffler distribution. In particular, it is shown that for values of the “generalizing” parameter not exceeding one, the generalized Mittag-Leffler distribution is a scale mixture of the half-normal distribution laws, classic Mittag-Leffler distributions, or generalized Mittag-Leffler distributions with the larger values of the characteristic index. The explicit expressions for mixing quantities are given for all cases. The obtained representations allow proposing new algorithms for modeling random variables with the generalized Mittag-Leffler distribution and formulating new limit theorems in which such distributions appear as the limit ones.

Keywords: generalized Mittag-Leffler distribution; scale mixture; generalized gamma distribution; half-normal distribution; stable distribution

DOI: 10.14357/19922264180411

Acknowledgments

The research is supported by the Russian Foundation for Basic Research (project 17-07-00717).

References

1. Korolev, V. Yu., A. I. Zeifman, and A. Yu. Korcha-gin. 2016. Nesimmetrichnye dvustoronnie raspredeleniya Mittag-Lefflera kak predel'nye zakony dlya sluchaynykh summ nezavisimyykh sluchaynykh velichin s konechnymi dispersiyami [Nonsymmetric two-sided Mittag-Leffler distributions as limit laws for random sums of independent random variables with finite variances]. *Statisticheskie metody otsenivaniya i proevpki gipotez* [Statistical methods for evaluating and testing hypothesis]. Perm: Perm State University. 27:69–89.
2. Korolev, V. Yu., and A. I. Zeifman. 2017. A note on mixture representations for the Linnik and Mittag-Leffler distributions and their applications. *J. Math. Sci.* 218(3):314–327.
3. Korolev, V. Yu., and A. I. Zeifman. 2017. Convergence of statistics constructed from samples with random sizes to the Linnik and Mittag-Leffler distributions and their generalizations. *J. Korean Stat. Soc.* 46(2):161–181.
4. Korolev, V. Yu., A. K. Gorshenin, and A. I. Zeifman. 2018. On mixture representations for the generalized Linnik distribution and their applications in limit theorems. *arXiv*.
5. Mittnik, S., and S. T. Rachev. 1993. Modeling asset returns with alternative stable distributions. *Economet. Rev.* 12:261–330.
6. Kotz, S., T. J. Kozubowski, and K. Podgorski. 2001. *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*. Boston, MA: Birkhauser. 349 p.
7. Gorenflo, R., and F. Mainardi. 2008. Continuous time random walk, Mittag-Leffler waiting time and fractional diffusion: Mathematical aspects. *Anomalous transport: Foundations and applications*. Eds. R. Klages, G. Radons, and I. M. Sokolov. Weinheim, Germany: Wiley-VCH. 93–127.
8. Gorenflo, R., A. A. Kilbas, F. Mainardi, and S. V. Rogosin. 2014. *Mittag-Leffler functions, related topics and applications*. Berlin–New York: Springer. 443 p.
9. Jose, K. K., P. Uma, V. S. Lekshmi, and H. J. Haubold. 2010. Generalized Mittag-Leffler distributions and processes for applications in astrophysics and time series modeling. *Astrophysics Space* 202559:79–92.
10. Mathai, A. M., and H. J. Haubold. 2011. Matrix-variate statistical distributions and fractional calculus. *Fract. Calc. Appl. Anal.* 14(1):138–155.
11. Pillai, R. N. 1985. Semi- α -Laplace distributions. *Commun. Stat. Theory* 14:991–1000.
12. Linnik, Yu. V. 1953. Lineynye formy i statisticheskie kriterii. I, II [Linear forms and statistical criteria. I. II]. *Ukr. Math. J.* 5(2):207–243; 5(3):247–290.
13. Devroye, L. 1990. A note on Linnik’s distribution. *Stat. Probabil. Lett.* 9:305–306.

14. Anderson, D. N. 1992. A multivariate Linnik distribution. *Stat. Probabil. Lett.* 14:333–336.
15. Lin, G. D. 1994. Characterizations of the Laplace and related distributions via geometric compound. *Sankhya Ser. A* 56:1–9.
16. Kotz, S., I. V. Ostrovskii, and A. Hayfavi. 1995. Analytic and asymptotic properties of Linnik’s probability densities, I. *J. Math. Anal. Appl.* 193:353–371.
17. Kotz, S., I. V. Ostrovskii, and A. Hayfavi. 1995. Analytic and asymptotic properties of Linnik’s probability densities, II. *J. Math. Anal. Appl.* 193:497–521.
18. Jacques, C., B. Rémillard, and R. Theodorescu. 1999. Estimation of Linnik law parameters. *Statistics Risk Modeling* 17(3):213–236.
19. Kotz, S., and I. V. Ostrovskii. 1996. A mixture representation of the Linnik distribution. *Stat. Probabil. Lett.* 26:61–64.
20. Pakes, A. G. 1998. Mixture representations for symmetric generalized Linnik laws. *Stat. Probabil. Lett.* 37:213–221.
21. Anderson, D. N., and B. C. Arnold. 1993. Linnik distributions and processes. *J. Appl. Probab.* 30:330–340.
22. Jayakumar, K., K. Kalyanaraman, and R. N. Pillai. 1995. α -Laplace processes. *Math. Comput. Model.* 22:109–116.
23. Baringhaus, L., and R. Grubel. 1997. On a class of characterization problems for random convex combinations. *Ann. I. Stat. Math.* 49:555–567.
24. Kozubowski, T. J. 1998. Mixture representation of Linnik distribution. *Stat. Probabil. Lett.* 38:157–160.
25. Lin, G. D. 1998. A note on the Linnik distributions. *J. Math. Anal. Appl.* 217:701–706.
26. Zolotarev, V. M. 1986. *One-dimensional stable distributions*. Translation of mathematical monographs ser. Providence, RI: American Mathematical Society. Vol. 65. 284 p.
27. Schneider, W. R. 1986. Stable distributions: Fox function representation and generalization. *Stochastic processes in classical and quantum systems*. Eds. S. Albeverio, G. Casati, and D. Merlini. Berlin: Springer. 497–511.
28. Uchaikin, V. V., and V. M. Zolotarev. 1999. *Chance and stability*. Utrecht: VSP. 596 p.
29. Korolev, V. Yu. 2016. Product representations for random variables with the Weibull distributions and their applications. *J. Math. Sci.* 218(3):298–313.
30. Stacy, E. W. 1962. A generalization of the gamma distribution. *Ann. Math. Stat.* 33:1187–1192.
31. Gleser, L. J. 1989. The gamma distribution as a mixture of exponential distributions. *Am. Stat.* 43:115–117.
32. Korolev, V. Yu. 2017. Analogi teoremy Glezera dlya otritsatel’nykh binomial’nykh i obobshchennykh gamma-raspredeleniy i nekotorye ikh prilozheniya [Analogues of Gleser’s theorem for negative binomial and generalized gamma distributions and some their applications]. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):2–17.
33. Lim, S. C., and L. P. Teo. 2010. Analytic and asymptotic properties of multivariate generalized Linnik’s probability densities. *J. Fourier Anal. Appl.* 16(5):715–747.
34. Mathai, A. M. 2010. Some properties of Mittag-Leffler functions and matrix-variate analogues: A statistical perspective. *Fract. Calc. Appl. Anal.* 13(2):113–132.
35. Teicher, H. 1961. Identifiability of mixtures. *Ann. Math. Stat.* 32:244–248.
36. Gnedenko, B. V., and V. Yu. Korolev. 1996. *Random summation: Limit theorems and applications*. Boca Raton, FL: CRC Press. 288 p.
37. Grandell, J. 1976. *Doubly stochastic poisson processes*. Lecture notes in mathematics book ser. Berlin – Heidelberg – New York: Springer. Vol. 529. 244 p.
38. Korolev, V. Yu., and I. A. Sokolov. 2008. *Matematicheskie modeli neodnorodnykh potokov ekstremal’nykh sobytii* [Mathematical models of nonhomogeneous flows of extremal events]. Moscow: TORUS PRESS. 192 p.
39. Korolev, V. Yu. 1994. Convergence of random sequences with the independent random indexes. I. *Theor. Probab. Appl.* 39(2):282–297.

Received October 15, 2018

Contributors

Korolev Victor Yu. (b. 1954) — Doctor of Science (PhD) in physics and mathematics, professor, Head of Department, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China; vkorolev@cs.msu.ru

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; leading scientist, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; agorshenin@frcsc.ru

Zeifman Alexander I. (b. 1954) — Doctor of Science in physics and mathematics, professor, Head of Department, Vologda State University, 15 Lenin Str., Vologda 160000, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; principal scientist, Vologda Research Center of the Russian Academy of Sciences, 56-A Gorky Str., Vologda 160001, Russian Federation; a_zeifman@mail.ru

О НЕРАВНОМЕРНЫХ ОЦЕНКАХ ТОЧНОСТИ НОРМАЛЬНОЙ АППРОКСИМАЦИИ ДЛЯ РАСПРЕДЕЛЕНИЙ НЕКОТОРЫХ СЛУЧАЙНЫХ СУММ ПРИ ОСЛАБЛЕННЫХ МОМЕНТНЫХ УСЛОВИЯХ*

В. Ю. Королев¹, А. В. Дорофеева²

Аннотация: Представлены неравномерные оценки скорости сходимости в центральной предельной теореме для сумм случайного числа независимых одинаково распределенных случайных величин для случаев, когда индекс суммирования (число слагаемых в сумме) имеет биномиальное или пуассоновское распределение и стохастически независим от слагаемых. Рассматривается ситуация, в которой доступна информация лишь о существовании моментов второго порядка у слагаемых. Указаны конкретные числовые значения абсолютных констант, входящих в оценки. Попутно анонсируется уточнение абсолютной константы в неравномерной оценке скорости сходимости в центральной предельной теореме для сумм неслучайного числа независимых одинаково распределенных случайных величин с моментами порядков не выше второго.

Ключевые слова: центральная предельная теорема; нормальная аппроксимация; случайная сумма; биномиальное распределение; распределение Пуассона; теорема Пуассона

DOI: 10.14357/19922264180412

1 Введение

Оценки точности нормальной аппроксимации для распределений сумм случайных величин традиционно являются объектом пристального внимания среди специалистов в области теории вероятностей, поскольку они играют важную роль во многих прикладных задачах. Такие оценки помогают осознанно принимать решения об адекватности или неадекватности нормальной модели для наблюдаемых статистических закономерностей. При этом особый интерес представляет ситуация, в которой доступна лишь минимальная информация о существовании моментов второго порядка у слагаемых. Именно такой случай и рассматривается в данной заметке.

Пусть X_1, X_2, \dots — независимые случайные величины с $EX_i = 0$ и $0 < EX_i^2 \equiv \sigma_i^2 < \infty$, $i = 1, 2, \dots$. Существование моментов случайных величин X_1, X_2, \dots порядков выше второго не предполагается. Для $n \in \mathbb{N}$ обозначим $S_n = X_1 + \dots + X_n$ и $B_n^2 = \sigma_1^2 + \dots + \sigma_n^2$. Стандартную нормальную функцию распределения обозначим $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz, \quad x \in \mathbb{R}.$$

Обозначим

$$\begin{aligned} \Delta_n(x) &= |\mathbb{P}(S_n < xB_n) - \Phi(x)|; \\ \Delta_n &= \sup_x |\mathbb{P}(S_n < xB_n) - \Phi(x)|. \end{aligned}$$

Всюду далее символ $\mathbb{I}(A)$ будет обозначать индикаторную функцию события A . Для $\varepsilon \in (0, \infty)$ обозначим:

$$L_n(\varepsilon) = \frac{1}{B_n^2} \sum_{i=1}^n EX_i^2 \mathbb{I}(|X_i| \geq \varepsilon B_n); \quad L_n = L_n(1);$$

$$M_n(\varepsilon) = \frac{1}{B_n^3} \sum_{i=1}^n E|X_i|^3 \mathbb{I}(|X_i| < \varepsilon B_n);$$

$$M_n = M_n(1).$$

Оценкам величины Δ_n при указанных выше минимально возможных моментных условиях посвящены работы [1–13] (см. также книги [14, 15]). В частности, для любого $\varepsilon \in (0, \infty)$ справедлива оценка:

$$\Delta_n \leq 1,86 (L_n(\varepsilon) + M_n(\varepsilon)) \leq 1,86 (L_n(\varepsilon) + \varepsilon). \quad (1)$$

*Работа выполнена при поддержке РФФИ (проект 18-07-01405).

¹Факультет вычислительной математики и кибернетики, Московский государственный университет им. М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Hangzhou Dianzi University, China, vkorolev@cs.msu.ru

²Факультет вычислительной математики и кибернетики, Московский государственный университет им. М. В. Ломоносова, alex.dorofeyeva@gmail.com

Детальная история уточнения верхних оценок величины Δ_n , изобилующая интересными результатами и курьезами, описана в работах [13, 16], в которых подчеркнута, что оценки типа (1) разумно считать *естественными*, поскольку они связывают скорость сходимости в центральной предельной теореме с критерием сходимости.

В данной работе сосредоточимся на верхних оценках величины $\Delta_n(x)$. В 1979 г. В. В. Петров [17] показал, что существует конечная положительная постоянная C , гарантирующая выполнение неравенства:

$$\Delta_n(x) \leq C \sum_{k=1}^n \left[\frac{EX_k^2 \mathbf{1}(|X_k| \geq (1+|x|)B_n)}{B_n^2(1+|x|)^2} + \frac{E|X_k|^3 \mathbf{1}(|X_k| < (1+|x|)B_n)}{B_n^3(1+|x|)^3} \right]. \quad (2)$$

В 2001 г. неравенство (2) было передоказано другим методом в статье [9]. В нескольких работах предпринимались попытки оценить значение константы C в неравенстве (2). В частности, в работах [18, 19] была получена оценка $C \leq 76,17$. Эта оценка была существенно уточнена в работе [12], где было показано, что в случае одинаково распределенных слагаемых, рассматриваемом в настоящей статье, константа не превосходит 39,25. Следующее утверждение содержит уточненную оценку абсолютной константы.

Теорема 1. Пусть X_1, X_2, \dots — независимые одинаково распределенные случайные величины с $EX_1 = 0$ и $0 < EX_1^2 \equiv \sigma^2 < \infty$. Тогда для любого $x \in \mathbb{R}$ справедливо неравенство:

$$\Delta_n(x) \leq 36,62 \left[\frac{EX_1^2 \mathbf{1}(|X_1| \geq (1+|x|)\sigma\sqrt{n})}{\sigma^2(1+|x|)^2} + \frac{E|X_1|^3 \mathbf{1}(|X_1| < (1+|x|)\sigma\sqrt{n})}{\sigma^3\sqrt{n}(1+|x|)^3} \right].$$

Доказательство. Для уточнения абсолютной константы здесь частично использовались методы, описанные в [12], с учетом текущих наилучших оценок констант в неравенстве Берри–Эссена [20], его неравномерном аналоге [21] и неравенстве (1) [13]. Подробное описание алгоритма будет представлено в одной из следующих статей.

Цель настоящей работы — распространить утверждение теоремы 1 на случайные суммы, в которых число слагаемых имеет биномиальное или пуассоновское распределение. При этом будет существенно использоваться подход, развитый в работе [13].

2 Неравномерные оценки для биномиальных случайных сумм

Всюду далее рассматриваются независимые одинаково распределенные случайные величины X_1, X_2, \dots с $EX_i = 0$ и $0 < EX_i^2 \equiv \sigma^2 < \infty$. Пусть $p \in (0, 1]$ — произвольно. Пусть ξ_1, \dots, ξ_n — независимые случайные величины, такие что

$$\xi_j = \begin{cases} 1 & \text{с вероятностью } p, \\ 0 & \text{с вероятностью } 1-p, \end{cases} \quad j = 1, \dots, n.$$

Случайная величина $N_{n,p} = \xi_1 + \dots + \xi_n$ может интерпретироваться как число успехов в схеме испытаний Бернулли с вероятностью успеха p . Случайная величина $N_{n,p}$ имеет биномиальное распределение с параметрами n и p :

$$P(N_{n,p} = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

Предположим, что при каждом $n \in \mathbb{N}$ случайные величины $N_{n,p}, X_1, X_2, \dots$ взаимно независимы. В данном разделе основным объектом изучения будут *биномиальные случайные суммы* вида

$$S_{N_{n,p}} = X_1 + \dots + X_{N_{n,p}}.$$

При этом если $N_{n,p} = 0$, то $S_{N_{n,p}} = 0$.

Для $j \in \mathbb{N}$ введем случайные величины \tilde{X}_j , полагая

$$\tilde{X}_j = \begin{cases} X_j & \text{с вероятностью } p, \\ 0 & \text{с вероятностью } 1-p. \end{cases}$$

Несложно видеть, что $\tilde{X}_j \stackrel{d}{=} \xi_j X_j$, где сомножители в правой части независимы (здесь и далее символ $\stackrel{d}{=}$ обозначает совпадение распределений).

Пусть $F(x)$ — общая функция распределения случайных величин X_j , $E_0(x)$ — функция распределения с единственным единичным скачком в нуле. Тогда, очевидно,

$$P(\tilde{X}_j < x) = pF(x) + (1-p)E_0(x), \quad x \in \mathbb{R}, \quad j \in \mathbb{N}.$$

При этом $E\tilde{X}_j = 0$,

$$D\tilde{X}_j = E\tilde{X}_j^2 = p\sigma^2. \quad (3)$$

Лемма 1. Для любых $n \in \mathbb{N}$ и $p_j \in (0, 1]$

$$S_{N_{n,p}} \stackrel{d}{=} \tilde{X}_1 + \dots + \tilde{X}_n, \quad (4)$$

где случайные величины в правой части (4) независимы.

Доказательство. Доказательство представляет собой простое упражнение на свойства характеристических функций.

С учетом (3) и (4) легко заметить, что

$$DS_{N_{n,p}} = n\sigma^2.$$

Обозначим

$$\Delta_{n,p}(x) = \left| P(S_{N_{n,p}} < x\sigma\sqrt{np}) - \Phi(x) \right|.$$

Теорема 2. Для любых $n \in \mathbb{N}$ и $p \in (0, 1]$ справедливо неравенство:

$$\begin{aligned} \Delta_{n,p}(x) &\leq 36,62 \left[\frac{EX_1^2 \mathbf{1}(|X_1| \geq (1+|x|)\sigma\sqrt{np})}{\sigma^2(1+|x|)^2} + \right. \\ &\quad \left. + \frac{E|X_1|^3 \mathbf{1}(|X_1| < (1+|x|)\sigma\sqrt{np})}{\sigma^3\sqrt{np}(1+|x|)^3} \right] \leq \\ &\leq \frac{36,62}{\sigma^2(1+|x|)^2} \cdot EX_1^2 \min \left\{ 1, \frac{|X_1|}{\sigma\sqrt{np}(1+|x|)} \right\}. \end{aligned}$$

Доказательство. Из леммы 1 и соотношения (3) вытекает, что

$$\Delta_{n,p} = \left| P(\tilde{X}_1 + \dots + \tilde{X}_n < x\sigma\sqrt{np}) - \Phi(x) \right|.$$

Правую часть этого соотношения оценим с помощью теоремы 1 и получим:

$$\begin{aligned} &\left| P(\tilde{X}_1 + \dots + \tilde{X}_n < x\sigma\sqrt{\theta_n}) - \Phi(x) \right| \leq \\ &\leq 36,62 \left[\frac{E\tilde{X}_1^2 \mathbf{1}(|\tilde{X}_1| \geq (1+|x|)\sigma\sqrt{np})}{p\sigma^2(1+|x|)^2} + \right. \\ &\quad \left. + \frac{E|\tilde{X}_1|^3 \mathbf{1}(|\tilde{X}_1| < (1+|x|)\sigma\sqrt{np})}{\sigma^3 p^{3/2} \sqrt{n}(1+|x|)^3} \right] = \\ &\leq 36,62 \left[\frac{EX_1^2 \mathbf{1}(|X_1| \geq (1+|x|)\sigma\sqrt{np})}{\sigma^2(1+|x|)^2} + \right. \\ &\quad \left. + \frac{E|X_1|^3 \mathbf{1}(|X_1| < (1+|x|)\sigma\sqrt{np})}{\sigma^3\sqrt{np}(1+|x|)^3} \right] = \\ &= \frac{36,62}{\sigma^2(1+|x|)^2} EX_1^2 \min \left\{ 1, \frac{|X_1|}{\sigma\sqrt{np}(1+|x|)} \right\}, \end{aligned}$$

что и требовалось доказать.

3 Неравномерные оценки для пуассоновских случайных сумм

Для $\lambda > 0$ пусть N_λ — случайная величина, имеющая распределение Пуассона с параметром λ :

$$P(N_\lambda = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N} \cup \{0\}.$$

Предположим, что при каждом $\lambda > 0$ случайные величины $N_\lambda, X_1, X_2, \dots$ независимы. Рассмотрим пуассоновскую случайную сумму

$$S_{N_\lambda} = X_1 + \dots + X_{N_\lambda}.$$

Если $N_\lambda = 0$, то полагаем $S_{N_\lambda} = 0$. Несложно убедиться, что $ES_{N_\lambda} = 0$ и $DS_{N_\lambda} = \lambda\sigma^2$. Точность нормальной аппроксимации для распределений пуассоновских случайных сумм изучалась многими авторами (см. исторические обзоры в работах [22, 23]). Равномерные оценки точности нормальной аппроксимации для распределений пуассоновских случайных сумм при ослабленных моментных условиях получены в статье [13]. Насколько известно авторам, неравномерные оценки для такой ситуации еще не проводились.

В данном разделе будут построены верхние оценки величины

$$\Delta_\lambda(x) = \left| P(S_{N_\lambda} < x\sigma\sqrt{\lambda}) - \Phi(x) \right|.$$

Зафиксируем λ и наряду с N_λ рассмотрим случайную величину $N_{n,p}$, имеющую биномиальное распределение с произвольной парой параметров n и $p \in (0, 1]$, удовлетворяющей условию $np = \lambda$. При этом

$$DS_{N_\lambda} = DS_{N_{n,p}} = \sigma^2\lambda = \sigma^2np.$$

Поэтому для любого $x \in \mathbb{R}$ по неравенству треугольника

$$\begin{aligned} \Delta_\lambda(x) &\leq \\ &\leq \Delta_{n,p}(x) + \left| P(S_{N_\lambda} < x) - P(S_{N_{n,p}} < x) \right|. \end{aligned} \quad (5)$$

Первое слагаемое в правой части (5) оценим с помощью теоремы 2 и получим:

$$\begin{aligned} \Delta_{n,p}(x) &\leq \\ &\leq \frac{36,62}{\sigma^2(1+|x|)^2} EX_1^2 \min \left\{ 1, \frac{|X_1|}{\sigma\sqrt{\lambda}(1+|x|)} \right\}. \end{aligned} \quad (6)$$

Рассмотрим второе слагаемое в правой части (5).
Имеем:

$$\begin{aligned} & |P(S_{N_\lambda} < x) - P(S_{N_{n,p}} < x)| \leq \\ & \leq \sup_x \sum_{k=0}^{\infty} P\left(\sum_{j=1}^k X_j < x\right) |P(N_{n,p} = k) - \\ & - P(N_\lambda = k)| \leq \sum_{k=0}^{\infty} |P(N_{n,p} = k) - P(N_\lambda = k)|. \end{aligned} \quad (7)$$

Правую часть неравенства (7) оценим с помощью неравенства Барбура–Холла [24], в соответствии с которым

$$\sum_{k=0}^{\infty} |P(N_{n,p} = k) - P(N_\lambda = k)| \leq 2p \min\{1, \lambda\}. \quad (8)$$

Таким образом, из (5), (6) и (8) вытекает, что для любых n и p , удовлетворяющих условию $np = \lambda$, и для любого $x \in \mathbb{R}$ справедливо неравенство:

$$\begin{aligned} \Delta_\lambda(x) \leq \frac{36,62}{\sigma^2(1+|x|)^2} EX_1^2 \min\left\{1, \frac{|X_1|}{\sigma\sqrt{\lambda}(1+|x|)}\right\} + \\ + \frac{2}{n} \lambda \min\{1, \lambda\}. \end{aligned} \quad (9)$$

Теперь, устремляя в (9) $n \rightarrow \infty$, получаем окончательный результат.

Теорема 3. Для любых $\lambda > 0$ и $x \in \mathbb{R}$ справедлива оценка:

$$\Delta_\lambda(x) \leq \frac{36,62}{\sigma^2(1+|x|)^2} EX_1^2 \min\left\{1, \frac{|X_1|}{\sigma\sqrt{\lambda}(1+|x|)}\right\}.$$

Литература

1. Katz M. Note on the Berry–Esseen theorem // Ann. Math. Stat., 1963. Vol. 39. No. 4. P. 1348–1349.
2. Петров В. В. Одна оценка отклонения распределения суммы независимых случайных величин от нормального закона // Докл. АН СССР, 1965. Т. 160. Вып. 5. С. 1013–1015.
3. Осипов Л. В. Уточнение теоремы Линдберга // Теория вероятностей и ее применения, 1966. Т. 11. Вып. 2. С. 339–342.
4. Feller W. On the Berry–Esseen theorem // Z. Wahrscheinlichkeit., 1968. Bd. 10. S. 261–268.
5. Paditz L. Bemerkungen zu einer Fehlerabschätzung im zentralen Grenzwertsatz // Wiss. Z. Hochsch. Verkehrswesen Friedrich List Dres., 1980. Vol. 27. No. 4. P. 829–837.
6. Paditz L. On error-estimates in the central limit theorem for generalized linear discounting // Math. Operationsforsch. Stat. Ser. Stat., 1984. Vol. 15. No. 4. P. 601–610.
7. Barbour A. D., Hall P. Stein’s method and the Berry–Esseen theorem // Aust. J. Stat., 1984. Vol. 26. P. 8–15.
8. Paditz L. Über eine Fehlerabschätzung im zentralen Grenzwertsatz // Wiss. Z. Hochsch. Verkehrswesen Friedrich List Dres., 1986. Vol. 33. No. 2. P. 399–404.
9. Chen L. H. Y., Shao Q. M. A non-uniform Berry–Esseen bound via Stein’s method // Probab. Theory Rel., 2001. Vol. 120. P. 236–254.
10. Королев В. Ю., Понов С. В. Уточнение оценок скорости сходимости в центральной предельной теореме при отсутствии моментов порядков, больших второго // Теория вероятностей и ее применения, 2011. Т. 56. Вып. 4. С. 797–805.
11. Королев В. Ю., Понов С. В. Уточнение оценок скорости сходимости в центральной предельной теореме при ослабленных моментных условиях // Докл. РАН, 2012. Т. 445. Вып. 3. С. 265–270.
12. Понов С. В. Оценки скорости сходимости в центральной предельной теореме при ослабленных моментных условиях: Дис. . . . канд. физ.-мат. наук. — М.: МГУ, 2012.
13. Korolev V., Dorofeeva A. Bounds of the accuracy of the normal approximation to the distributions of random sums under relaxed moment conditions // Lith. Math. J., 2017. Vol. 57. No. 1. P. 38–58.
14. Петров В. В. Суммы независимых случайных величин. — М.: Наука, 1972. 416 с.
15. Петров В. В. Предельные теоремы для сумм независимых случайных величин. — М.: Наука, 1987. 320 с.
16. Шевцова И. Г. Моментное неравенство с применением к оценкам скорости сходимости в глобальной ЦПТ для пуассон-биномиальных случайных сумм // Теория вероятностей и ее применения, 2017. Т. 62. Вып. 2. С. 345–364.
17. Петров В. В. Одна предельная теорема для сумм независимых неодинаково распределенных случайных величин // Записки научных семинаров ЛОМИ, 1979. Т. 85. С. 188–192.
18. Thongtha P., Neammanee K. Refinement of the constants in the non-uniform version of the Berry–Esseen theorem // Thai J. Math., 2007. Vol. 5. P. 1–13.
19. Neammanee K., Thongtha P. Improvement of the non-uniform version of the Berry–Esseen inequality via Paditz–Shiganov theorems // J. Inequalities Pure Appl. Math., 2007. Vol. 8. No. 4. Art. 92.
20. Шевцова И. Г. Об абсолютных константах в неравенствах Берри–Эссеена // Докл. РАН, 2014. Т. 456. № 6. С. 650–654.
21. Нефедова Ю. С., Шевцова И. Г. О неравномерных оценках скорости сходимости в центральной предельной теореме // Теория вероятностей и ее применения, 2012. Т. 57. № 1. С. 62–97.

22. Korolev V. Yu., Shevtsova I. G. An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums // *Scand. Actuar. J.*, 2012. No. 2. P. 81–105.
23. Шевцова И. Г. О точности нормальной аппроксимации для обобщенных пуассоновских распределений // *Теория вероятностей и ее применения*, 2013. Т. 58. № 1. С. 152–176.
24. Barbour A. D., Hall P. On the rate of Poisson convergence // *Math. Proc. Cambridge*, 1984. Vol. 95. P. 473–480.

Поступила в редакцию 15.10.18

ON NONUNIFORM ESTIMATES OF ACCURACY OF NORMAL APPROXIMATION FOR DISTRIBUTIONS OF SOME RANDOM SUMS UNDER RELAXED MOMENT CONDITIONS

V. Yu. Korolev^{1,2,3} and A. V. Dorofeeva¹

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

²Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China

Abstract: Nonuniform estimates are presented for the rate of convergence in the central limit theorem for sums of a random number of independent identically distributed random variables. Two cases are studied in which the summation index (the number of summands in the sum) has the binomial or Poisson distribution. The index is assumed to be independent of the summands. The situation is considered where the information that only the second moments of the summands exist is available. Particular numerical values of the absolute constants are presented explicitly. Also, the sharpening of the absolute constant in the nonuniform estimate of the rate of convergence in the central limit theorem for sums of a nonrandom number of independent identically distributed random variables is announced for the case where the summands possess only second moments.

Keywords: central limit theorem; normal approximation; random sum; binomial distribution; Poisson distribution; Poisson theorem

DOI: 10.14357/19922264180412

Acknowledgments

This work was financially supported by the Russian Foundation for Basic Research (grant No. 118-07-01405).

References

1. Katz, M. 1963. Note on the Berry–Esseen theorem. *Ann. Math. Stat.* 39(4):1348–1349.
2. Petrov, V. V. 1965. Odnа otsenka otkloneniya raspredeleniya summy nezavisimyykh sluchaynykh velichin ot normal'nogo zakona [One estimate of the deviation of distribution of the sum of independent random variables from the normal law]. *Sov. Math.* 160(5):1013–1015.
3. Osipov, L. V. 1966. Refinement of Lindeberg's theorem. *Theor. Probab. Appl.* 11(2):299–302.
4. Feller, W. 1968. On the Berry–Esseen theorem. *Z. Wahrscheinlichkeit.* 10:261–268.
5. Paditz, L. 1980. Bemerkungen zu einer Fehlerabschätzung im zentralen Grenzwertsatz. *Wiss. Z. Hochsch. Verkehrswesen Friedrich List Dres.* 27(4):829–837.
6. Paditz, L. 1984. On error-estimates in the central limit theorem for generalized linear discounting. *Math. Operationsforsch. Stat. Ser. Stat.* 15(4):601–610.
7. Barbour, A. D., and P. Hall. 1984. Stein's method and the Berry–Esseen theorem. *Aust. J. Stat.* 26:8–15.
8. Paditz, L. 1986. Über eine Fehlerabschätzung im zentralen Grenzwertsatz. *Wiss. Z. Hochsch. Verkehrswesen Friedrich List Dres.* 33(2):399–404.
9. Chen, L. H. Y., and Q. M. Shao. 2001. A non-uniform Berry–Esseen bound via Stein's method. *Probab. Theory Rel.* 120:236–254.
10. Korolev, V. Yu., and S. V. Popov. 2012. An improvement of convergence rate estimates in the central limit theorem under absence of moments higher than the second. *Theor. Probab. Appl.* 56(4):682–691.

11. Korolev, V. Yu., and S. V. Popov. 2012. Improvement of convergence rate estimates in the central limit theorem under weakened moment conditions. *Dokl. Math.* 86(1):506–511.
12. Popov, S. V. 2012. Utochneniye ocnok skorosti skhodomosti v tsentral'noy predel'noy teoreme pri oslablennykh momentnykh usloviyakh [Improvement of convergence rate estimates in the central limit theorem under weakened moment conditions]. Moscow: MSU. PhD Diss.
13. Korolev, V., and A. Dorofeeva. 2017. Bounds of the accuracy of the normal approximation to the distributions of random sums under relaxed moment conditions. *Lith. Math. J.* 57(1):38–58.
14. Petrov, V. V. 1972. *Summy nezavisimykh sluchaynykh velichin* [Sums of independent random variables]. Moscow: Nauka. 416 p.
15. Petrov, V. V. 1987. *Predel'nye teoremy dlya summ nezavisimykh sluchaynykh velichin* [Limit theorems for sums of independent random variables]. Moscow: Nauka. 320 p.
16. Shevtsova, I. G. 2018. A moment inequality with application to convergence rate estimates in the global CLT for Poisson-binomial random sums. *Theor. Probab. Appl.* 62(2):278–294.
17. Petrov, V. V. 1979. Odnа predel'naya teorema dlya summ nezavisimykh neodnakovo raspredelennykh sluchaynykh velichin [One limit theorem for sums of independent unequally distributed random variables]. *J. Mathematical Sciences* 85:188–192.
18. Thongtha, P., and K. Neammanee. 2007. Refinement of the constants in the non-uniform version of the Berry–Esseen theorem. *Thai J. Math.* 5:1–13.
19. Neammanee, K., and P. Thongtha. 2007. Improvement of the non-uniform version of the Berry–Esseen inequality via Paditz–Shiganov theorems. *J. Inequalities Pure Appl. Math.* 8(4): 92.
20. Shevtsova, I. G. 2014. On the absolute constants in the Berry–Esseen-type inequalities. *Dokl. Math.* 89(3):378–381.
21. Nefedova, Yu. S., and I. G. Shevtsova. 2013. On nonuniform convergence rate estimates in the central limit theorem. *Theor. Probab. Appl.* 57(1):28–59.
22. Korolev, V. Yu., and I. G. Shevtsova. 2012. An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scand. Actuar. J.* 2:81–105.
23. Shevtsova, I. G. 2014. On the accuracy of the normal approximation to compound Poisson distributions. *Theor. Probab. Appl.* 58(1):138–158.
24. Barbour, A. D., and P. Hall. 1984. On the rate of Poisson convergence. *Math. Proc. Cambridge* 95:473–480.

Received October 15, 2018

Contributors

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, professor, Head of the Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China; vkorolev@cs.msu.su

Dorofeeva Alexandra V. (b. 1991) — PhD student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; alex.dorofeyeva@gmail.com

ГАММА-ВЕЙБУЛЛОВСКИЙ СЛУЧАЙ АПРИОРНЫХ РАСПРЕДЕЛЕНИЙ В БАЙЕСОВСКИХ МОДЕЛЯХ МАССОВОГО ОБСЛУЖИВАНИЯ*

Е. Н. Арутюнов¹, А. А. Кудрявцев², А. И. Титова³

Аннотация: Статья посвящена байесовскому подходу к задачам теории массового обслуживания и надежности. В байесовских моделях для классических постановок задач предполагается, что основные параметры системы, например интенсивности входящего потока и обслуживания, являются случайными величинами с известными априорными распределениями. Байесовский подход эффективен при изучении больших совокупностей однотипных систем или одной системы, характеристики которой меняются в моменты времени, неизвестные исследователю. Рандомизация основных параметров системы приводит к тому, что различные ее характеристики, такие как коэффициент загрузки, также становятся случайными. Приводятся аналитические результаты для вероятностных характеристик коэффициента загрузки, выражаемые в терминах гамма-экспоненциальной функции, в случае, когда в качестве пары априорных распределений параметров системы $M|M|1|0$ рассматриваются гамма-распределение и распределение Вейбулла.

Ключевые слова: байесовский подход; системы массового обслуживания; смешанные распределения; распределение Вейбулла; гамма-распределение; гамма-экспоненциальная функция

DOI: 10.14357/19922264180413

1 Введение

При изучении и математическом моделировании функционирования различных объектов среди параметров моделируемого объекта можно выделить «способствующие» и «препятствующие» функционированию факторы. При применении такого разделения в теории массового обслуживания к группе параметров, «способствующих» функционированию, можно отнести интенсивность обслуживания запросов, а к параметрам, «препятствующим» функционированию, — интенсивность входящего потока требований.

Зачастую в силу структурной сложности моделируемых систем и стохастичности среды функционирования определение точного значения характеристик каждого компонента системы или совокупности систем представляется ресурсоемкой задачей. В таких ситуациях целесообразно воспользоваться байесовским подходом. При байесовском подходе исходные характеристики системы считаются заданными в определенном смысле неточно,

однако предполагается, что имеется информация об их априорном распределении. За счет рандомизации исходных параметров системы происходит и рандомизация зависящих от них показателей, которые, как правило, представляют больший интерес для исследования функционирования системы.

В данной работе рассматривается система массового обслуживания $M|M|1|0$, одним из основных показателей которой является коэффициент ее загрузки ρ . Значение коэффициента загрузки определяется как отношение параметра входящего потока λ к параметру обслуживания μ . Подробное описание предпосылок для исследования, особенностей и библиографии байесовских моделей в теории массового обслуживания и надежности можно найти в книге [1].

Ниже для модели $M|M|1|0$ приводятся вероятностные характеристики коэффициента загрузки ρ в случае, когда в качестве пары априорных распределений параметров системы λ и μ рассматриваются гамма-распределение и распределение Вейбулла.

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 17-07-00577).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, enaroleon@mail.ru

² Факультет вычислительной математики и кибернетики, Московский государственный университет им. М. В. Ломоносова, nubigena@mail.ru

³ Факультет вычислительной математики и кибернетики, Московский государственный университет им. М. В. Ломоносова, onkelskroot@gmail.com

2 Основные результаты

Обозначим через $G(q, \theta)$ гамма-распределение с плотностью

$$g_{q,\theta}(x) = \frac{\theta^q x^{q-1} e^{-\theta x}}{\Gamma(q)}, \quad x > 0, \quad q > 0, \quad \theta > 0,$$

а через $W(p, \alpha)$ — распределение Вейбулла с плотностью

$$w_{p,\alpha}(x) = \frac{p x^{p-1} e^{-(x/\alpha)^p}}{\alpha^p}, \quad x > 0, \quad p > 0, \quad \alpha > 0.$$

Рассмотрим гамма-экспоненциальную функцию [2]:

$$\text{Ge}_{\alpha,\beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \Gamma(\alpha k + \beta), \quad x \in \mathbb{R}, \quad \alpha \geq 0, \quad \beta > 0.$$

Хорошо известен следующий результат.

Лемма 1. Для случайных величин ξ и η , имеющих гамма-распределение $G(q, \theta)$ и распределение Вейбулла $W(p, \alpha)$ соответственно, для $z \in \mathbb{R}$ выполняются следующие соотношения:

$$\mathbb{E} \xi^z = \frac{\Gamma(q+z)}{\theta^z \Gamma(q)}, \quad z > -q;$$

$$\mathbb{E} \eta^z = \alpha^z \Gamma\left(1 + \frac{z}{p}\right), \quad z > -p.$$

Для дальнейших вычислений потребуется следующее утверждение.

Лемма 2. Для некоторых $q > 0, p > 0, a > 0$ и $b > 0$ справедливо:

$$\int_0^{\infty} y^{q+p-1} e^{-ay} e^{-(y/b)^p} dy = \begin{cases} b^{q+p} p^{-1} \text{Ge}_{1/p, q/p+1}(-ab), & p > 1; \\ a^{-q-p} \text{Ge}_{p, q+p}(-(ab)^{-p}), & p < 1; \\ \left(a + \frac{1}{b}\right)^{-q-1} \Gamma(q+1), & p = 1. \end{cases}$$

Доказательство. Рассмотрим случай $p > 1$. Используя теорему Лебега о предельном переходе, получаем:

$$\begin{aligned} \int_0^{\infty} y^{q+p-1} e^{-ay} e^{-(y/b)^p} dy &= \\ &= \int_0^{\infty} y^{q+p-1} \sum_{k=0}^{\infty} \frac{(-a)^k y^k}{k!} e^{-(y/b)^p} dy = \\ &= \sum_{k=0}^{\infty} \frac{(-ab)^k}{k!} \int_0^{\infty} t^{k/p+q/p} e^{-t} dt = \\ &= \frac{b^{q+p}}{p} \sum_{k=0}^{\infty} \frac{(-ab)^k}{k!} \Gamma\left(\frac{k}{p} + \frac{q}{p} + 1\right). \end{aligned}$$

Случай $p < 1$ рассматривается аналогично. Случай $p = 1$ напрямую следует из определения гамма-функции. Лемма доказана.

Леммы 1 и 2 дают возможность находить вероятностные характеристики коэффициента загрузки системы.

Теорема 1. Пусть параметр входящего потока λ имеет гамма-распределение $G(q, \theta)$, а параметр обслуживания μ имеет распределение Вейбулла $W(p, \alpha)$, причем λ и μ независимы. Тогда при $x > 0$ плотность, функция распределения и моменты коэффициента загрузки $\rho = \lambda/\mu$ имеют вид:

$$\begin{aligned} f_{\rho}(x) &= \begin{cases} \frac{(\alpha\theta)^q x^{q-1}}{\Gamma(q)} \text{Ge}_{1/p, q/p+1}(-\alpha\theta x), & p > 1; \\ \frac{p}{x(\alpha\theta x)^p \Gamma(q)} \text{Ge}_{p, q+p}(-(\alpha\theta x)^{-p}), & p < 1; \end{cases} \\ F_{\rho}(x) &= \begin{cases} \frac{(\alpha\theta x)^q}{p\Gamma(q)} \text{Ge}_{1/p, q/p}(-\theta\alpha x), & p > 1; \\ \frac{1}{\Gamma(q)} \int_{(\alpha\theta x)^{-p}}^{\infty} \text{Ge}_{p, p+q}(-z) dz, & p < 1; \end{cases} \\ \mathbb{E} \rho^z &= \frac{\Gamma(q+z)}{(\alpha\theta)^z \Gamma(q)} \Gamma\left(1 - \frac{z}{p}\right), \quad z < p. \end{aligned}$$

При $p = 1$ распределение коэффициента загрузки ρ совпадает с распределением Дагума [3] с параметрами $(1, (\alpha\theta)^{-1}, q)$.

Доказательство. Поскольку

$$f_{\rho}(x) = \int_0^{\infty} \frac{p\theta^q x^{q-1} y^{p+q-1}}{\alpha^p \Gamma(q)} e^{-\theta xy} e^{-(y/\alpha)^p} dy,$$

вид плотности $f_{\rho}(x)$ вытекает из леммы 2 для всех $p > 0$.

Для функции распределения ρ при $p > 1$ справедливо для $x > 0$:

$$\begin{aligned} F_{\rho}(x) &= \int_0^x \frac{(\alpha\theta)^q u^{q-1}}{\Gamma(q)} \text{Ge}_{1/p, q/p+1}(-\theta\alpha u) du = \\ &= \frac{(\alpha\theta)^q}{\Gamma(q)} \sum_{k=0}^{\infty} \frac{(-\alpha\theta)^k}{k!} \Gamma\left(\frac{k}{p} + \frac{q}{p} + 1\right) \int_0^x u^{q+k-1} du = \\ &= \frac{(\alpha\theta x)^q}{p\Gamma(q)} \sum_{k=0}^{\infty} \frac{(-\alpha\theta x)^k}{k!} \Gamma\left(\frac{k}{p} + \frac{q}{p}\right). \end{aligned}$$

В случае $p < 1$ имеем:

$$\begin{aligned} F_p(x) &= \int_0^x \frac{p \text{Ge}_{p, q+p}(-(\alpha\theta u)^{-p})}{u(\alpha\theta u)^p \Gamma(p)} du = \\ &= \int_0^x \sum_{k=0}^{\infty} \frac{p(-(\alpha\theta u)^{-p})^k \Gamma(kp + q + p)}{(\alpha\theta)^p \Gamma(p) k! u^{p+1}} du = \\ &= \frac{1}{\Gamma(p)} \int_{(x/\alpha\theta)^{-p}}^{\infty} \sum_{k=0}^{\infty} \frac{(-z)^k}{k!} \Gamma(kp + q + p) dz. \end{aligned}$$

Для нахождения моментов ρ достаточно воспользоваться независимостью случайных величин λ и μ и леммой 1. Теорема доказана.

Рассмотрим симметричный случай априорных распределений параметров входящего потока и обслуживания.

Теорема 2. Пусть параметр входящего потока λ имеет распределение Вейбулла $W(q, \theta)$, а параметр обслуживания μ имеет гамма-распределение $G(p, \alpha)$, причем λ и μ независимы. Тогда при $x > 0$ плотность, функция распределения и моменты коэффициента загрузки $\rho = \lambda/\mu$ имеют вид:

$$\begin{aligned} f_\rho(x) &= \begin{cases} \frac{(\alpha\theta)^p}{x^{p+1} \Gamma(p)} \text{Ge}_{1/q, p/q+1} \left(-\frac{\alpha\theta}{x} \right), & q > 1; \\ \frac{qx^{q-1}}{(\alpha\theta)^q \Gamma(p)} \text{Ge}_{q, p+q} \left(-\left(\frac{x}{\alpha\theta}\right)^q \right), & q < 1; \end{cases} \\ F_\rho(x) &= \begin{cases} \frac{1}{p\Gamma(p)} \int_{(x/\alpha\theta)^{-p}}^{\infty} \text{Ge}_{1/q, p/q+1}(-z^{1/p}) dz, & q > 1; \\ 1 - \frac{1}{\Gamma(p)} \text{Ge}_{q, p} \left(-\left(\frac{x}{\theta\alpha}\right)^q \right), & q < 1; \end{cases} \\ E \rho^z &= (\alpha\theta)^z \frac{\Gamma(1 + z/q) \Gamma(p - z)}{\Gamma(p)}, \quad z < p. \end{aligned}$$

При $q = 1$ распределение коэффициента загрузки ρ совпадает с распределением Ломакса [4] с параметрами $\alpha\theta$ и p .

Доказательство. Аналогично предыдущей теореме для получения выражения для плотности ρ при всех $q > 0$ достаточно воспользоваться леммой 2.

Найдем функцию распределения в случае $q > 1$. Имеем:

$$F_\rho(x) = \frac{(\alpha\theta)^p}{\Gamma(p)} \int_0^x u^{-p-1} \text{Ge}_{1/q, p/q+1} \left(-\frac{\alpha\theta}{u} \right) du =$$

$$\begin{aligned} &= \frac{(\alpha\theta)^p}{\Gamma(p)} \int_0^x u^{-p-1} \sum_{k=0}^{\infty} \frac{(-\alpha\theta/u)^k}{k!} \Gamma\left(\frac{k}{q} + \frac{p}{q} + 1\right) du = \\ &= \frac{1}{p\Gamma(p)} \int_{(x/\alpha\theta)^{-p}}^{\infty} \sum_{k=0}^{\infty} \frac{(-z^{1/p})^k}{k!} \Gamma\left(\frac{k}{q} + \frac{p}{q} + 1\right) dz. \end{aligned}$$

Для $q < 1$ функция распределения ρ имеет вид:

$$\begin{aligned} F_\rho(x) &= \int_0^x \frac{qu^{q-1}}{(\alpha\theta)^q \Gamma(p)} \text{Ge}_{q, p+q} \left(-\left(\frac{u}{\alpha\theta}\right)^q \right) du = \\ &= \int_0^x \frac{qu^{q-1}}{(\alpha\theta)^q \Gamma(p)} \sum_{k=0}^{\infty} \frac{(-u/(\theta\alpha))^k}{k!} \Gamma(qk + q + p) du = \\ &= -\frac{1}{\Gamma(p)} \sum_{k=0}^{\infty} \frac{(-x/(\theta\alpha))^k}{(k+1)!} \Gamma(q(k+1) + p) = \\ &= 1 - \frac{1}{\Gamma(p)} \sum_{m=0}^{\infty} \frac{(-x/(\theta\alpha))^m}{m!} \Gamma(qm + p). \end{aligned}$$

Соотношение для моментов следует из леммы 1 и независимости случайных величин λ и μ . Теорема доказана.

3 Заключение

Результаты статьи в очередной раз свидетельствуют о том, что гамма-экспоненциальная функция $\text{Ge}_{\alpha, \beta}(x)$ является удобным инструментом исследования различных характеристик вероятностных смесей законов гамма-типа.

Литература

1. Кудрявцев А. А., Шоргин С. Я. Байесовские модели в теории массового обслуживания и надежности. — М.: ФИЦ ИУ РАН, 2015. 76 с.
2. Кудрявцев А. А., Титова А. И. Гамма-экспоненциальная функция в байесовских моделях массового обслуживания // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 104–108.
3. Dagum C. A new model of personal income-distribution-specification and estimation // Econ. Appl., 1977. Vol. 30. No. 3. P. 413–437.
4. Lomax K. S. Business failures: Another example of the analysis of failure data // J. Am. Stat. Assoc., 1954. Vol. 49. No. 268. P. 847–852.

Поступила в редакцию 19.08.18

GAMMA-WEIBULL *A PRIORI* DISTRIBUTIONS IN BAYESIAN QUEUING MODELS

E. N. Arutyunov¹, A. A. Kudryavtsev², and A. I. Titova²

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

Abstract: This paper considers the Bayesian approach to the queuing theory and reliability theory. Within the framework of the Bayesian approach, it is assumed that the key parameters of classical systems, such as the input flow intensity and the service intensity, are random variables with known *a priori* distributions. It is reasonable to use the Bayesian approach in the studying of the systems which are of the same type as well as in the studying of one system where changes of its characteristics happen at unpredictable moments of time. The randomization of the system’s key parameters leads to the randomization of the system’s characteristics, for instance, the traffic intensity. In the paper, the analytical results for the traffic intensity probability characteristics in the case of the gamma and Weibull *a priori* distributions of $M/M/1/0$ system’s parameters are presented. The obtained results are formulated using the gamma-exponential function.

Keywords: Bayesian approach; queuing system; mixed distribution; Weibull distribution; gamma distribution; gamma-exponential function

DOI: 10.14357/19922264180413

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 17-07-00577).

References

1. Kudryavtsev, A. A., and S. Ya. Shorgin. 2015. *Bayesovskie modeli v teorii massovogo obsluzhivaniya i nadezhnosti* [Bayesian models in mass service and reliability theories]. Moscow: FRC CSC RAS. 76 p.
2. Kudryavtsev, A. A., and A. I. Titova. 2017. Gamma-eksponentsial’naya funktsiya v bayesovskikh modelyakh massovogo obsluzhivaniya [Gamma-exponential function in Bayesian queuing models]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):104–108.
3. Dagum, C. 1977. A new model of personal income-distribution-specification and estimation. *Econ. Appl.* 30(3):413–437.
4. Lomax, K. S. 1954. Business failures: Another example of the analysis of failure data. *J. Am. Stat. Assoc.* 49(268):847–852.

Received August 19, 2018

Contributors

Arutyunov Evgeny N. (b. 1952) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; enapoleon@mail.ru

Kudryavtsev Alexey A. (b. 1978) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; nubigena@mail.ru

Titova Anastasiia I. (b. 1995) — student, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; onkelskroot@gmail.com

КОЛИЧЕСТВЕННЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ МАШИННОГО ПЕРЕВОДА С ИСПОЛЬЗОВАНИЕМ НАДКОРПУСНЫХ БАЗ ДАННЫХ*

Н. В. Бунтман¹, А. А. Гончаров², И. М. Зацман³, В. А. Нуриев⁴

Аннотация: Рассматривается информационная технология, которая поддерживает экспертизу результатов машинного перевода литературных текстов. Технология разрабатывалась и апробировалась на примерах переводов коннекторов при следующих условиях. Во-первых, объектом исследования выступают переводы предложений как с однословными (например, *хотя, а, кстати* и т.д.), так и с неоднословными коннекторами (например, *да еще, но зато, и вообще, и притом, хотя и* и т.д.). Во-вторых, между словами, входящими в состав коннекторов, может быть фрагмент текста, например: *если (расстояние) так, когда (расстояние) то, не только (расстояние) но и, так как (расстояние) то* и т.д. Технология поддержки экспертизы результатов машинного перевода коннекторов и их контекстов с использованием надкорпусных баз данных (НБД); (2) количественная обработка результатов аннотирования; (3) лингвистический анализ сформированных аннотаций и полученных числовых данных. Статья посвящена описанию технологических аспектов поддержки экспертизы, относящихся к ее первым двум стадиям. Экспериментальный материал включает примеры только с неоднословными коннекторами, части которых могут располагаться как дистантно, так и контактно.

Ключевые слова: надкорпусная база данных; машинный перевод; классификация ошибок; технология поддержки экспертизы; лингвистическое аннотирование; корпусная лингвистика; коннекторы

DOI: 10.14357/19922264180414

1 Введение

В 2018 г. в издательстве «Шпрингер» началась публикация новой серии книг, посвященных технологиям и применению систем машинного перевода [1, 2]. Первый том этой серии полностью посвящен анализу результатов машинного перевода и включает обзор подходов к классификации и аннотированию его ошибок [3], что говорит об актуальности этой проблематики.

Одна из задач российско-швейцарского проекта «Контрастивное корпусное исследование коннекторов русского языка»⁵ заключается в том, чтобы создать классификацию и систему аннотирования ошибок машинного перевода коннекторов русского языка, функция которых состоит в обеспечении связности текста [4].

В качестве основного инструмента для решения этой задачи использовалась НБД, предназначен-

ная для лингвистического аннотирования языковых единиц разных категорий, включая коннекторы [5–8]. В качестве исходного информационного ресурса для классификации и аннотирования ошибок использовались параллельные выровненные тексты [9].

Для поддержки процесса создания классификации и аннотирования ошибок в рамках проекта была разработана информационная технология, которая охватывает три основные стадии этого процесса:

- (1) лингвистическое аннотирование результатов машинного перевода коннекторов и их контекстов с использованием НБД;
- (2) количественная обработка результатов аннотирования;
- (3) лингвистический анализ сформированных аннотаций и полученных числовых данных.

* Работа выполнена в Институте проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук при поддержке РФФИ (проект № 16-24-41002).

¹ Факультет иностранных языков и регионоведения, Московский государственный университет им. М. В. Ломоносова, nabunt@hotmail.com

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, a.gonch48@gmail.com

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, izatsman@yandex.ru

⁴ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, nurieff.v@gmail.com

⁵ Проект поддержан РФФИ (грант № 16-24-41002) и Швейцарским национальным научным фондом (грант № IZLRZ1.164059).

Таблица 1 Пример результата машинного перевода

Когда бываю в таком состоянии, то становлюсь нахальным и наглым до крайности.	Quand je suis dans cet état, je deviens impudent et insolent à l'extrême. (translate.google.com, дата обращения 05.02.2018 19:48)
---	---

Примечание. С момента записи аннотации изменился вариант перевода, предлагаемый системой GNMT: «Quand je suis dans un tel état, je deviens impudent et impudent à l'extrême» (11.10.2018 19:35).

Цель статьи — описать технологические аспекты поддержки экспертизы результатов машинного перевода с использованием системы Google (translate.google.com), которые относятся к первым двум стадиям.

Система машинного перевода Google применялась для перевода с русского языка на французский фрагментов художественного текста, состоящих из одного или более предложений. На момент начала эксперимента в 2018 г. переводчик Google уже работал по принципу нейронной сети, т. е. представлял собой систему нейронного машинного перевода (Google's Neural Machine Translation system — GNMT). Его архитектура отличается от устройства автоматических переводчиков предыдущего поколения — статистических машинных переводчиков. Система состоит из рекуррентной нейросети с долгой краткосрочной памятью (long short-term memory). В этой сети задействованы два восьмислойных сегмента — кодирующий (encoder) и декодирующий (decoder) в сочетании с механизмом внимания (attention mechanism). Свойство рекуррентности позволяет вычислять значение слова или словосочетания с учетом предыдущих значений в последовательности, с которыми сеть уже имела дело. Единицей перевода в такой системе является не словосочетание, как в статистических машинных переводчиках, а предложение. Система GNMT умеет работать с фрагментами слов: слова в исходном тексте и переводе представляются в форме набора составных элементов (wordpieces), что направлено на улучшение точности при обработке редких слов (подробнее о системе GNMT см. [10, 11]).

Источником используемых в эксперименте текстов послужил русско-французский параллельный подкорпус Национального корпуса русского языка (НКРЯ) [12], где переводы на французский язык выполнены профессиональными переводчиками. Поэтому кроме аннотирования и экспертизы результатов машинного перевода была возможность провести сопоставление «человеческих» (далее — референтных) переводов с машинным, что заслуживает освещения в отдельной публикации.

Приведем краткое описание используемых информационных ресурсов. Из НКРЯ были получены тексты нескольких десятков книг на русском языке объемом около 1,7 млн словоупотреблений и их переводы на французский язык объемом около 2,3 млн словоупотреблений (существенное превышение объема перевода связано с тем, что одно произведение может иметь несколько переводов). Параллельные тексты были выровнены, т. е. одному или нескольким предложениям на русском языке (объект перевода) соответствует одно или несколько предложений на французском (результат перевода).

Использованные технологические решения были во многом обусловлены следующими факторами. Во-первых, предметом исследования выступают переводы предложений с неоднословными коннекторами (например, *да еще, но зато, и вообще, и притом, хотя и* и т. д.). Во-вторых, между словами, входящими в состав коннекторов, может быть фрагмент текста, например: *если (расстояние) так, когда (расстояние) то, не только (расстояние) но и, так как (расстояние) то* и т. д.

Из текстов НКРЯ был отобран экспериментальный массив объемом 1500 объектов перевода с коннекторами русского языка, которые и были переведены с помощью системы GNMT (см. табл. 1 с примером машинного перевода предложения из пьесы А. П. Чехова «Дядя Ваня»). Критерии отбора объектов перевода будут описаны в разд. 2.

На первой стадии экспертизы было выполнено лингвистическое аннотирование результатов машинного перевода коннекторов и их контекстов, включая оценку их качества. Для этого использовался подход, основанный на анализе ошибок перевода, что подразумевает их локализацию и классификацию [13]. Ориентация проекта на исследование коннекторов не позволяла использовать уже существующие системы классификации ошибок перевода [3, 14], поскольку в них не учитывается специфика этих языковых единиц, в частности структура коннектора, его позиция (конечная, начальная, неначальная) и расположение частей коннектора (дистантное, контактное).

2 Динамические фасетные классификации

Система классификации ошибок машинного перевода коннекторов, учитывающая их специфику, формировалась в процессе лингвистического аннотирования на основе следующей категоризации коннекторов (в конце каждого элемента списка указана доля коннекторов каждой категории, аннотированных в НБД, в процентах от общего их числа 1276 по состоянию на 20.08.2018) [15]:

- одноэлементные (однословные) — состоят из одного элемента (например, *и, или, но, а, хотя, иначе* и т. д.) — 5%;
- многоэлементные — состоят из нескольких элементов (например, *и|вообще, к тому же, а особенно, но при всем при том* и т. д.) — 55%;
- двухкомпонентные — коннекторы, части которых вводят два разных текстовых фрагмента (например, *даже если бы||то, как только||тут же, не только||а прежде всего, раз уж||так уж* и т. д.)¹ — 35%;
- многокомпонентные — коннекторы, части которых вводят более двух разных текстовых фрагментов (например, *не только||но даже и||и даже, хотя||хотя||однако все же* и т. д.) — 5%.

Из определений этих четырех категорий следует, что однословные коннекторы относятся к последним трем категориям и только в этих категориях фрагмент текста может находиться между частями коннекторов. В качестве первоначальной версии классификации ошибок результатов машинного перевода использовался список из 8 рубрик [16], приведенный в табл. 2.

В процессе аннотирования результатов машинного перевода первоначальная классификация ошибок итерационно уточнялась. Технологи-

чески это обеспечивалось НБД коннекторов за счет использования в ней динамической фасетной классификации признаков (= рубрик) аннотирования [17–19]. До начала аннотирования в нее был добавлен новый фасет для описания ошибок машинного перевода коннекторов, включавший 8 рубрик (см. табл. 2). В процессе аннотирования НБД обеспечивала внесение необходимых лингвистам изменений в классификацию непосредственно в ходе этого процесса (включая переименование, удаление и добавление, деление и объединение рубрик), что иногда приводило к переклассификации уже сформированных аннотаций (более подробно о переклассификации см. [20]).

После завершения аннотирования обновленная система классификации ошибок увеличилась до 15 рубрик (табл. 3). Метод аннотирования с использованием динамических фасетных классификаций описан в работах [17, 19, 21].

Отметим, что использование в НБД метода аннотирования с применением динамических фасетных классификаций позволяет начинать этот процесс, не имея в фасете ошибок ни одной рубрики, т. е. можно формировать систему классификации с нуля в процессе аннотирования [22]. В любом варианте формирования (с нуля или с использованием первоначальной версии) на вариант классификации, получаемый после завершения аннотирования, влияет степень представительности отобранного массива объектов перевода.

Отбор объектов перевода осуществлялся с учетом следующего условия, которое диктуется самой языковой реальностью: частотность употребления разных коннекторов непосредственно коррелирует с их структурой. Согласно собранному экспериментальному материалу, на однословные коннекторы приходится 47,4% всех зарегистрированных случаев употребления, доля многоэлементных составляет 40,1%, двухкомпонентных — 12%, а мно-

Таблица 2 Версия классификации ошибок до начала эксперимента

Название рубрики	Код рубрики
1. Все предложение аграмматично	AgramTotal
2. Коннектор переведен несуществующей языковой единицей	ErrorTotal
3. Локальная аграмматичность во фрагменте текста, вводимом коннектором	AgramPostCNT
4. Локальная аграмматичность во фрагменте текста, не вводимом коннектором	AgramLocal
5. Орфографическая ошибка в форме коннектора	ErrorOrth
6. Семантическая ошибка в выборе коннектора	ErrorCNT
7. Слова кириллицей	Cyrillic
8. Часть однословного коннектора переведена ошибочно	ErrorPart

Примечание. В работе [16] рубрика «Орфографическая ошибка в форме коннектора» имела код AgramOrth.

¹Двойная вертикальная черта «||» в списке указывает на то, что коннектор состоит из двух или более компонентов, а одиночная вертикальная черта «|» — что языковые единицы, составляющие коннектор или его компонент, разделены текстом.

Таблица 3 Обновленная классификация ошибок

Название рубрики	Код рубрики
1. Все предложение аграмматично	AgramTotal
2. Локальная аграмматичность во фрагменте текста, не вводимом коннектором	AgramLocal
3. Локальная аграмматичность во фрагменте текста, вводимом коннектором	AgramPostCNT
4. Лексическая ошибка во фрагменте текста	ErrorLex
5. Слова кириллицей	Cyrillic
6. Русское слово латинским шрифтом	Latin
7. Пропуск фрагмента текста	Lacuna
8. Переведена первая часть неоднословного коннектора	TradPartI
9. Переведена вторая часть неоднословного коннектора	TradPartII
10. Первая часть неоднословного коннектора переведена ошибочно	ErrorPartI
11. Вторая часть неоднословного коннектора переведена ошибочно	ErrorPartII
12. Коннектор переведен несуществующей единицей	ErrorTotal
13. Коннектор ошибочно заменен языковой единицей, не являющейся коннектором	ErrorMorph
14. Орфографическая ошибка в форме коннектора	ErrorOrth
15. Семантическая ошибка в выборе коннектора	ErrorCNT

гокомпонентных — всего 0,5%. Однако при формировании массива объектов перевода было решено только частично руководствоваться ограничениями, которые накладывает естественный язык. Исходя из условий, обозначенных в исследовательском проекте, объем массива должен был составлять 1500 случаев употребления, т. е. 1500 русскоязычных контекстов¹, где зафиксированы коннекторы разной структуры.

Чтобы обеспечить репрезентативность экспериментального материала, с одной стороны, было решено не рассматривать случаи употребления многокомпонентных коннекторов (в силу недостаточной репрезентативности имеющейся выборки) и отобрать по 500 контекстов для остальных трех категорий (более частотных по употреблению коннекторов), с другой стороны, стояла задача обеспечить многообразие коннекторов каждой из этих трех категорий. При этом отбирались те коннекторы, которые характеризуются высокой долей употребления, что позволяет наблюдать, как система машинного перевода работает с ними в разных контекстах их употребления. Третья категория (двухкомпонентные коннекторы), в отличие от первых двух, включает в себя большее число коннекторов (табл. 4)²: поскольку эти языковые единицы гораздо менее частотны, приходится увеличивать количество коннекторов, чтобы обеспечить требуемое число контекстов — объектов перевода (500).

Следует отметить главный недостаток такого подхода к отбору объектов перевода — в массив

не попадают низкочастотные коннекторы, однако особенности их машинного перевода могут стать предметом отдельного исследования.

3 Количественный анализ с помощью надкорпусной базы данных

По рубрикам, перечисленным в табл. 3, НБД позволяет получать широкий спектр количественных характеристик, необходимых для проведения лингвистического анализа ошибок перевода. Отметим, что именно рубрики фасетной классификации, используемые в процессе аннотирования, позволяют получать количественные характеристики ошибок машинного перевода коннекторов, что является отличительной чертой функциональных возможностей НБД. Проиллюстрируем это на примере рубрики TradPartI, наличие которой в аннотации говорит о том, что в результате машинного перевода была переведена только первая часть неоднословного коннектора.

Как отмечалось выше, в процессе аннотирования использовались результаты машинного перевода сформированного экспериментального массива, который состоял из 1500 объектов перевода (см. табл. 4). В итоге полученные 1500 аннотаций включали 54 коннектора. Из них 15 относились к категории одноэлементных, а это значит, что аннотации с ними не могли иметь рубрики TradPartI. Таким

¹ По факту объем массива объектов перевода составляет 1530 контекстов, т. е. на 30 контекстов больше запланированного по условиям проекта. Это увеличение объема обеспечивает свободу последующей коррекции и отбора необходимых 1500 контекстов.

² В табл. 4 представлен массив объектов перевода в его первоначальной — плановой форме (1530 объектов). После завершения аннотирования и проверки аннотаций число отобранных объектов перевода по отдельным категориям незначительно изменилось, также менялось число отобранных коннекторов.

Таблица 4 Коннекторы в массиве объектов машинного перевода

Одноэлементные	Число объектов перевода	Многоэлементные	Число объектов перевода	Двухкомпонентные	Число объектов перевода
только	35	то есть	41	не только но и	63
впрочем	40	и вообще	41	если то	63
вообще	40	одним словом	41	не а	56
а	40	при этом	41	когда то	47
или	40	да еще	41	как так и	42
также	25	а впрочем	41	не только но даже	30
тут	29	лишь только	41	хотя но	27
например	42	как только	41	если так	27
хотя	42	а также	37	так как то	24
кстати	40	в общем	32	не только но	20
зато	42	между прочим	26	как так	15
причем	42	да и вообще	25	не успеть + Inf как	13
притом	31	и притом	23	не только и	10
если	22	но зато	22	не только ∅	16
		если только	17	не только даже	8
				если б то бы	9
				не то что а	9
				когда так	8
				или или	8
				не только	8
				раз то	7

образом, оставалось 39 коннекторов, аннотации с которыми могли иметь эту рубрику. Однако после завершения аннотирования 1500 объектов перевода эту рубрику имели только 22 коннектора (см. первый столбец табл. 5), а именно: 8 многоэлементных и 14 двухкомпонентных.

В табл. 5 эти коннекторы упорядочены в соответствии с числом (по мере убывания) сформированных для них аннотаций с рубрикой TradPartI. Во втором столбце указано общее число аннотаций, сформированных для соответствующего коннектора из первого столбца, в третьем — число тех из них, где была проставлена эта рубрика, а в четвертом — процентное отношение числа аннотаций с рубрикой TradPartI к их общему числу. Всего для 22 коннекторов было сформировано 574 аннотации (т. е. 926 аннотаций были сформированы для других коннекторов, не имеющих этой рубрики), из которых 177 имеют эту рубрику.

Отметим, что рубрика TradPartI не всегда говорит об ошибке машинного перевода. Строго говоря, она применяется в тех аннотациях, где при переводе был опущен второй элемент многоэлементного коннектора или второй компонент двухкомпонентного коннектора. Поэтому в интересах лингвистического анализа предлагается разделить случаи проставления этой рубрики на три группы:

- (1) потеря второго элемента (или компонента) коннектора при машинном переводе не влечет

за собой изменения его семантики. В массиве из 1500 аннотаций это наблюдается для следующих коннекторов: *когда||то, если||то, так как||то, если||так, когда||так, если б||то|бы, раз||то*. Аннотация в табл. 6 иллюстрирует случай, когда при машинном переводе не меняются ни семантика коннектора, ни смысл всего предложения;

- (2) потеря второго элемента или компонента при переводе влечет незначительные изменения в передаче семантики коннектора (табл. 7), когда в оригинале второй элемент коннектора конкретизирует семантику первого (*да еще*) или добавляет к выражаемому им логико-семантическому отношению близкое по семантике (*но зато*);
- (3) потеря второго элемента или компонента коннектора при переводе ведет к изменению выражаемого им логико-семантического отношения (см. аннотацию в табл. 8), а также может сопровождаться сбоями на уровне грамматики или смысла переводимого фрагмента (см. аннотацию в табл. 9).

В заключение этого раздела отметим, что в процессе лингвистического анализа используются таблицы с более широким спектром количественных характеристик, чем табл. 5, так как одновременно рассматривается, как правило, сочетание несколь-

Таблица 5 Количественные характеристики, полученные с помощью НБД*

Коннектор русского языка	Всего аннотаций	Из них число аннотаций с TradPartI	Доля аннотаций с TradPartI, %
1) когда то	47	39	82,98
2) если то	59	20	33,90
3) но зато	22	20	90,91
4) так как то	24	15	62,50
5) а впрочем	41	11	26,83
6) да еще	41	11	26,83
7) если так	25	9	36,00
8) а также	37	8	21,62
9) не успеть + Inf как	13	6	46,15
10) если б то бы	9	5	55,56
11) как так и	37	5	13,51
12) когда так	8	5	62,50
13) и вообще	44	4	9,09
14) и притом	23	4	17,39
15) хотя и	3	3	100,00
16) хотя но	27	3	11,11
17) как так	13	2	15,38
18) не только и	10	2	20,00
19) не только но и	61	2	3,28
20) если только	17	1	5,88
21) не только даже	8	1	12,50
22) раз то	5	1	20,00
ИТОГО	574	177	—

* Авторы благодарят М. Г. Кружкова за подготовку данной таблицы.

Таблица 6 Логико-семантическое отношение не изменяется при потере второй части коннектора в процессе машинного перевода

Когда же взгляды встретились, то дверь вдруг захлопнулась. . . (Ф. М. Достоевский «Преступление и наказание»)	когда то (временные) (CNT p CNT q) (CNT) (Дистант)	Quand les regards se rencontrèrent, la porte se referma brusquement. . . (translate.google.com, дата обращения 05.02.2018 18:37)	quand (временные) (с предикацией) (начальная) (CNT q p) (CNT) (TradPartI) (NoError)
--	---	---	--

Примечания: (1) с момента записи аннотации изменился вариант перевода, предлагаемый системой GNMT: «Quand les yeux se croisèrent, la porte claqua soudainement. . .» (12.10.2018 19:22); (2) описание фасетов и их рубрик во втором и четвертом столбцах приведено в работах [8, 17].

Таблица 7 Логико-семантическое отношение незначительно изменяется при потере второй части коннектора в процессе машинного перевода

Он дал ей понять, что догадался о ее любви к нему, да еще, может быть, догадался невпопад. (И. А. Гончаров. «Обломов»)	да еще (аддитивные пропозициональные) (с предикацией) (начальная) (p CNT q) (CNT) (Контакт)	Il lui fit savoir qu'il avait deviné son amour pour lui et, peut-être, l'avait deviné au mauvais moment. (translate.google.com, дата обращения 31.01.2018 01:27)	et (соединительные) (с предикацией) (начальная) (p CNT q) (CNT) (TradPartI) (NoError)
---	---	---	--

Примечание. С момента записи аннотации изменился вариант перевода, предлагаемый системой GNMT: «Il lui fit savoir qu'il avait deviné son amour pour lui et peut-être aussi l'avait-il deviné au hasard» (12.10.2018 19:38).

Таблица 8 Логико-семантическое отношение значительно изменяется при потере второй части коннектора в процессе машинного перевода

<p><i>Вот, как</i> придет человек, <i>так</i> отдам.</p> <p>(И. А. Гончаров. «Обломов»)</p>	<p>как так ⟨временное/условное⟩ ⟨CNT p CNT q⟩ ⟨CNT⟩ ⟨Дистант⟩</p>	<p><i>Alors, comme</i> un homme vient, je vais le donner.</p> <p>(translate.google.com, дата обращения 27.02.2018 19:31)</p>	<p>comme ⟨причина⟩ ⟨с предикацией⟩ ⟨начальная⟩ ⟨CNT q p⟩ ⟨CNT⟩ ⟨SubCNT⟩ ⟨TradPartI⟩ ⟨Error CNT⟩</p>
---	---	--	--

Примечание. С момента записи аннотации изменился вариант перевода, предлагаемый системой GNMT: «Voilà comment un homme vient, alors je vais donner» (12.10.2018 19:58).

Таблица 9 Логико-семантическое отношение значительно изменяется и сопровождается аграмматичностью при потере второй части коннектора в процессе машинного перевода

<p>Целое отделение таких девочек... как ехали в вагоне, так и лежали...</p> <p>(С. А. Алексиевич. «Время секунд хэнд»)</p>	<p>как так и ⟨аналогия⟩ ⟨CNT p CNT q⟩ ⟨CNT⟩ ⟨Дистант⟩</p>	<p>L'ensemble du département de ces filles... comme ils sont montés dans la voiture, ils étaient...</p> <p>(translate.google.com, дата обращения 06.02.2018 12:59)</p>	<p>comme ⟨сравнительные⟩ ⟨с предикацией⟩ ⟨начальная⟩ ⟨CNT q p⟩ ⟨CNT⟩ ⟨AgramPostCNT⟩ ⟨ErrorLex⟩ ⟨TradPartI⟩ ⟨Error CNT⟩</p>
--	---	---	--

Примечание. С момента записи аннотации изменился вариант перевода, предлагаемый системой GNMT: «Toute une branche de ces filles... comme elles étaient dans la voiture, elles gisaient comme ça...» (13.10.2018 21:55).

ких рубрик. Здесь была рассмотрена только одна рубрика TradPartI, чтобы продемонстрировать функциональные возможности НБД по аннотированию и количественной обработке результатов машинного перевода.

4 Заключение

Разработанная информационная технология, которая поддерживает экспертизу результатов машинного перевода, существенно расширяет спектр задач компьютерной лингвистики, которые можно решать с использованием НБД. Сначала НБД использовались в основном для лингвистического аннотирования языковых единиц в параллельных текстах с референтными переводами. Затем НБД была адаптирована для поддержки экспертизы результатов машинного перевода.

В процессе применения разработанной технологии с использованием адаптированной НБД были получены следующие новые результаты в компьютерной лингвистике. Во-первых, появилась возможность интегрировать в рамках единой фасетной классификации общезыковые рубрики оши-

бок (например, пропуск слов, нарушение их порядка, ошибки в пунктуации и др. [3]) и рубрики, характеризующие ошибки перевода только одного вида языковых единиц. Такой подход позволяет выделить отдельный фасет классификации ошибок для каждого вида языковых единиц. Во-вторых, появилась возможность непосредственно в процессе аннотирования формировать единую систему классификации и общезыковых, и видовых ошибок машинного перевода с нуля или начиная с некоторой ее первоначальной (базовой) версии.

Литература

1. Translation quality assessment: From principles to practice / Eds. J. Moorkens, S. Castilho, F. Gaspari, S. Doherty. — Machine translation: Technologies and applications ser. — Cham: Springer International Publishing, 2018. Vol. 1. 299 p.
2. Scott B. Translation, brains and the computer: A neurolinguistic solution to ambiguity and complexity in machine translation. — Machine translation: Technologies and applications ser. — Cham: Springer International Publishing, 2018. Vol. 2. 241 p.

3. *Popović M.* Error classification and analysis for machine translation quality assessment // Translation quality assessment: From principles to practice / Eds. J. Moorkens, S. Castilho, F. Gaspari, S. Doherty. — Machine translation: Technologies and applications ser. — Cham: Springer International Publishing, 2018. Vol. 1. P. 129–158.
4. Семантика коннекторов: контрастивное исследование / Под ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2018. 368 с.
5. *Кружков М. Г.* Информационные ресурсы контрастивных лингвистических исследований: электронные корпуса текстов // Системы и средства информатики, 2015. Т. 25. № 2. С. 140–159.
6. *Зализняк Анна А., Зацман И. М., Инькова О. Ю., Кружков М. Г.* Надкорпусные базы данных как лингвистический ресурс // Корпусная лингвистика-2015: Труды 7-й Междунар. конф. — СПб.: СПбГУ, 2015. С. 211–218.
7. *Попкова Н. А., Инькова О. Ю., Зацман И. М., Кружков М. Г.* Методика построения моноэквиваленций в надкорпусной базе данных коннекторов // Задачи современной информатики: Труды 2-й конф. — М.: ФИЦ ИУ РАН, 2015. С. 143–153.
8. *Зацман И. М., Инькова О. Ю., Кружков М. Г., Попкова Н. А.* Представление кросс-языковых знаний о коннекторах в надкорпусных базах данных // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 106–118.
9. *Добровольский Д. О., Кретов А. А., Шаров С. А.* Корпус параллельных текстов: архитектура и возможности использования // Национальный корпус русского языка: 2003–2005. — М.: Индрик, 2005. С. 263–296.
10. *Wu Y., Schuster M., Chen Z., et al.* Google's neural machine translation system: Bridging the gap between human and machine translation // arXiv.org, 2016. <https://arxiv.org/pdf/1609.08144.pdf>.
11. *Johnson M., Schuster M., Le Q. V., Krikun M., Wu Y., Chen Zh., Thorat N., Viégas F., Wattenberg M., Corrado G., Hughes M., Dean J.* Google's multilingual neural machine translation system: Enabling zero-shot translation // T. Assoc. Computational Linguistics, 2017. Vol. 5. P. 339–351.
12. Национальный корпус русского языка. <http://www.ruscorpora.ru>.
13. *Улиткин И. А.* Автоматическая оценка качества машинного перевода научно-технического текста // Вестник МГОУ. Сер. Лингвистика, 2016. № 4. С. 174–182.
14. *Vilar D., Xu J., D'haro L., Ney H.* Error analysis of statistical machine translation output // 5th Conference (International) on Language Resources and Evaluation Proceedings. — Genoa, Italy: European Language Resources Association, 2006. P. 697–702. http://www.lrec-conf.org/proceedings/lrec2006/pdf/413_pdf.pdf.
15. *Inkova O. Yu., Krzhkov M. G.* Statistical analysis of language specificity of connectives based on parallel texts // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 83–90.
16. *Nuriev V., Buntman N., Inkova O.* Machine translation of Russian connectives into French: Errors and quality failures // Информатика и её применения, 2018. Т. 12. Вып. 2. С. 105–113.
17. *Зализняк Анна А., Зацман И. М., Инькова О. Ю.* Надкорпусная база данных коннекторов: построение системы терминов // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 100–106.
18. *Inkova O. Yu., Popkova N. A.* Statistical data as information source for linguistic analysis of Russian connectors // Информатика и её применения, 2017. Т. 11. Вып. 3. С. 123–131.
19. *Зацман И. М., Кружков М. Г., Лоцилова Е. Ю.* Методы анализа частотности моделей перевода коннекторов и обратимость генерализации статистических данных // Системы и средства информатики, 2017. Т. 27. № 4. С. 164–176.
20. *Зацман И. М.* Стадии целенаправленного извлечения знаний, имплицитированных в параллельных текстах // Системы и средства информатики, 2018. Т. 28. № 3. С. 169–182.
21. *Дурново А. А., Зацман И. М., Лоцилова Е. Ю.* Кросс-лингвистическая база данных для аннотирования логико-семантических отношений в тексте // Системы и средства информатики, 2016. Т. 26. № 4. С. 124–137.
22. *Zatsman I.* Goal-oriented creation of individual knowledge: Model and information technology // 19th European Conference on Knowledge Management Proceedings. — Reading: Academic Publishing International Ltd., 2018. Vol. 2. P. 947–956.

Поступила в редакцию 15.10.18

USING SUPRACORPORA DATABASES FOR QUANTITATIVE ANALYSIS OF MACHINE TRANSLATIONS

N. V. Buntman¹, A. A. Goncharov², I. M. Zatsman², and V. A. Nuriev²

¹M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

²Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper discusses an information technology that supports expertise of machine translations. The technology has been developed to meet the following conditions: (i) there are connectives in all translated contexts; (ii) the connectives may be both one-word (*khotya* ‘although,’ *a* ‘and’) and multiword (*da esche* ‘and beside this,’ *no zato* ‘but instead’); and (iii) between words making up a given connective, there may be a space (*esli* (space) *tak* ‘if (space) then’). With this technology, expertise of machine translations develops through three main stages: (i) linguistic annotation of machine translations in a supracorpora database; (ii) quantitative processing of annotations; and (iii) linguistic analysis of annotations and quantitative data. The paper describes technological aspects of the first two stages. The examples given are only those with multiword connectives. Source sentences chosen for machine translation have been collected from literary texts.

Keywords: supracorpora database; machine translation; classification of errors; technology supporting expertise; linguistic annotation; corpus linguistics; connectives

DOI: 10.14357/19922264180414

Acknowledgments

The work was fulfilled at the Institute of Informatics Problems of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences and supported by the Russian Foundation for Basic Research (project No. 16-24-41002).

References

1. Moorkens, J., S. Castilho, F. Gaspari, and S. Doherty, eds. 2018. *Translation quality assessment: From principles to practice*. Machine translation: Technologies and applications ser. Cham: Springer International Publishing. Vol. 1. 299 p.
2. Scott, B. 2018. *Translation, brains and the computer: A neurolinguistic solution to ambiguity and complexity in machine translation*. Machine translation: Technologies and applications ser. Cham: Springer International Publishing. Vol. 2. 241 p.
3. Popović, M. 2018. Error classification and analysis for machine translation quality assessment. *Translation quality assessment: From principles to practice*. Eds. J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty. Machine translation: Technologies and applications ser. Cham: Springer International Publishing. 1:129–158.
4. Inkova, O. Yu., ed. 2018. *Semantika konnektorov: kontrastivnoe issledovanie* [Semantics of connectives: A contrastive study]. Moscow: TORUS PRESS. 368 p.
5. Kruzhkov, M. G. 2015. Informatsionnye resursy kontrastivnykh lingvisticheskikh issledovaniy: elektronnye korpusa tekstov [Information resources for contrastive studies: Electronic text corpora]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(2):140–159.
6. Zaliznyak, Anna A., I. M. Zatsman, O. Yu. Inkova, and M. G. Kruzhkov. 2015. Nadkorpurnye bazy dannykh kak lingvisticheskiy resurs [Supracorpora databases as linguistic resource]. *Conference (International) “Corpus linguistics-2015” Proceedings*. St. Petersburg: St. Petersburg State University. 211–218.
7. Popkova, N. A., O. Yu. Inkova, I. M. Zatsman, and M. G. Kruzhkov. 2015. Metodika postroeniya monoekvivalentnykh v nadkorpurnoy baze dannykh konnektorov [Methodology of constructing monoequivalences in the Supracorpora database of connectors]. *Tr. 2-y nauchn. konf. “Zadachi sovremennoy informatiki”* [2nd Scientific Conference “Problems of Modern Informatics” Proceedings]. Moscow: FRC CSC RAS. 143–153.
8. Zatsman, I. M., O. Yu. Inkova, M. G. Kruzhkov, and N. A. Popkova. 2016. Predstavlenie kross-yazykovykh znaniy o konnektorakh v nadkorpurnykh bazakh dannykh [Representation of cross-lingual knowledge about connectors in Supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):106–118.
9. Dobvol’skiy, D. O., A. A. Kretov, and S. A. Sharov. 2005. Korpus paralel’nykh tekstov: arkhitektura i vozmozhnosti ispol’zovaniya [Corpus of parallel texts: Architecture and applications]. *Natsional’nyy korpus russkogo yazyka: 2003–2005* [Russian National Corpus: 2003–2005]. Moscow: Indrik. 263–296.
10. Wu, Y., M. Schuster, Z. Chen, et al. 2016. Google’s neural machine translation system: Bridging the gap

- between human and machine translation. Available at: <https://arxiv.org/pdf/1609.08144.pdf> (accessed September 3, 2018).
11. Johnson, M., M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Zh. Chen, N. Thorat, F. Viégas, M. Wattenberg, G. Corrado, M. Hughes, and J. Dean. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *T. Assoc. Computational Linguistics* 5:339–351.
 12. Natsional'nyy korpus russkogo yazyka [Russian National Corpus]. Available at: <http://www.ruscorpora.ru> (accessed November 30, 2018).
 13. Ulitkin, I. A. 2016. Avtomaticheskaya otsenka kachestva mashinnogo perevoda nauchno-tehnicheskogo teksta [Automatic evaluation of machine translation quality of a scientific text]. *B. MRSU. Ser. Linguistics* 4:174–182.
 14. Vilar, D., J. Xu, L. D'haro, and H. Ney. 2006. Error analysis of statistical machine translation output. *5th Conference (International) on Language Resources and Evaluation Proceedings*. Genoa, Italy: European Language Resources Association. Available at: <http://www.lrec-conf.org/proceedings/lrec2006/pdf/413.pdf.pdf> (accessed September 3, 2018).
 15. Inkova, O. Yu., and M. G. Kruzhev. 2018. Statistical analysis of language specificity of connectives based on parallel texts. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):83–90.
 16. Nuriev, V., N. Buntman, and O. Inkova. 2018. Machine translation of Russian connectives into French: Errors and quality failures. *Informatika i ee Primeneniya — Inform. Appl.* 12(2):105–113.
 17. Zaliznyak, Anna A., I. M. Zatsman, and O. Yu. Inkova. 2017. Nadkorpurnaya baza dannykh konnektorov: postroenie sistemy terminov [Supracorpora database on connectives: Term system development]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):100–106.
 18. Inkova, O. Yu., and N. A. Popkova. 2017. Statistical data as information source for linguistic analysis of Russian connectors. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):123–131.
 19. Zatsman, I. M., M. G. Kruzhev, and E. Yu. Loshchilova. 2017. Metody analiza chastotnosti modeley perevoda konnektorov i obratimost' generalizatsii statisticheskikh dannykh [Methods of frequency analysis of connectives translations and reversibility of statistical data generalization]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(4):164–176.
 20. Zatsman, I. M. 2018. Stadii tselenapravlennoy izvlecheniya znaniy, implitsirovannykh v parallel'nykh tekstakh [Stages of goal-oriented discovery of knowledge implied in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):169–182.
 21. Durnovo, A. A., I. M. Zatsman, and E. Yu. Loshchilova. 2016. Kross-lingvisticheskaya baza dannykh dlya annotirovaniya logiko-semanticheskikh otnosheniy v tekste [Cross-lingual database for annotating logical-semantic relations in the text]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(4):124–137.
 22. Zatsman, I. 2018. Goal-oriented creation of individual knowledge: Model and information technology. *19th European Conference on Knowledge Management Proceedings*. Reading: Academic Publishing International Ltd. 2:947–956.

Received October 15, 2018

Contributors

Buntman Nadezhda V. (b. 1957) — Candidate of Science (PhD) in philology, associate professor, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; nabunt@hotmail.com

Goncharov Alexander A. (b. 1994) — engineer, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; a.gonch48@gmail.com

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; izatsman@yandex.ru

Nuriev Vitaly A. (b. 1980) — Candidate of Science (PhD) in philology, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; nurieff.v@gmail.com

РАЗВИТИЕ СЕРВИСОВ ЦИФРОВЫХ ПЛАТФОРМ ДЛЯ ПРЕОДОЛЕНИЯ НЕФИНАНСОВЫХ БАРЬЕРОВ*

А. К. Горшенин¹

Аннотация: Рассматриваются примеры различных нефинансовых барьеров, препятствующих эффективному развитию молодежи в научной и образовательной сферах. Их преодоление не всегда возможно исключительно за счет привлечения дополнительных ресурсов, так как зачастую требуется изменение среды. Предложены концептуальные способы преодоления подобных барьеров за счет создания и развития сервисов платформ цифровой экономики — одной из основных современных парадигм в информационных технологиях (ИТ). На примере создаваемой в Федеральном исследовательском центре «Информатика и управление» Российской академии наук цифровой платформы «Наука и образование» продемонстрировано соответствие сервисов основным направлениям реализации Стратегии научно-технологического развития Российской Федерации (СНТР). Предложенная концепция является эффективной и для решения более широкого класса задач.

Ключевые слова: цифровые платформы; нефинансовые барьеры; стратегия научно-технологического развития; цифровая экономика; молодежная политика

DOI: 10.14357/19922264180415

1 Введение

Российская Федерация по относительному показателю — доле затрат на проведение научных исследований и разработок от величины внутреннего валового продукта (ВВП) — уступает лидерам (США, Китай, Япония, Южная Корея), однако находится на приемлемом мировом уровне [1], опережая по ряду показателей страны БРИКС (включая Индию и Бразилию), а также европейские государства (Италия, Испания). В частности, с точки зрения такого важного наукометрического показателя, как число опубликованных в международных базах работ, согласно данным The SCImago Journal & Country Rank (<https://www.scimagojr.com/countryrank.php>), Россия входит в топ-15 стран мира (рис. 1).

В то же время влияние отечественных исследований на общемировые тенденции не столь велико. На рис. 2 продемонстрирована зависимость «странового» индекса Хирша от среднего числа цитирований документа (<https://www.scimagojr.com/worldreport.php>) для различных стран. Диаметр кругов соответствует числу работ, Россия выделена более крупным полужирным шрифтом.

Указанное обстоятельство связано как с дефицитом финансирования отрасли (только 1,52% ВВП

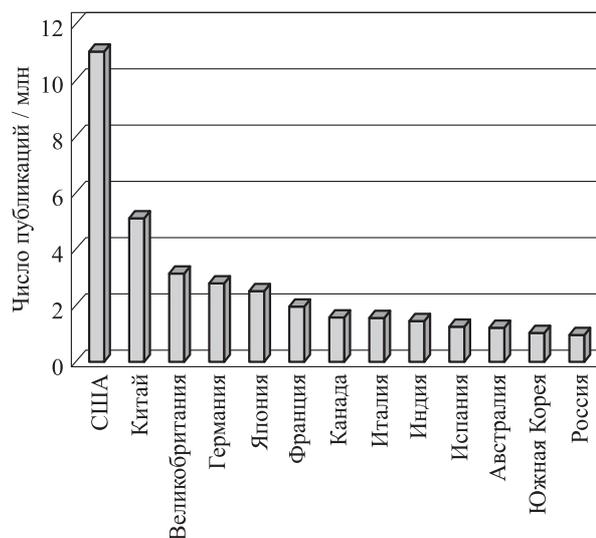


Рис. 1 Общее число опубликованных статей за 1996–2017 гг. в странах-лидерах

инвестируется в научные исследования [1], в то время как в США — 2,83%, при общем отличии отечественного ВВП в 6–8 раз по сравнению с США и Китаем), так и с достаточно небольшим числом ученых и инженеров в пересчете на миллион населения (около 3300 человек) — по этому показателю

* Работа выполнена при поддержке РНФ (проект 18-71-00156).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, agorshenin@frccsc.ru

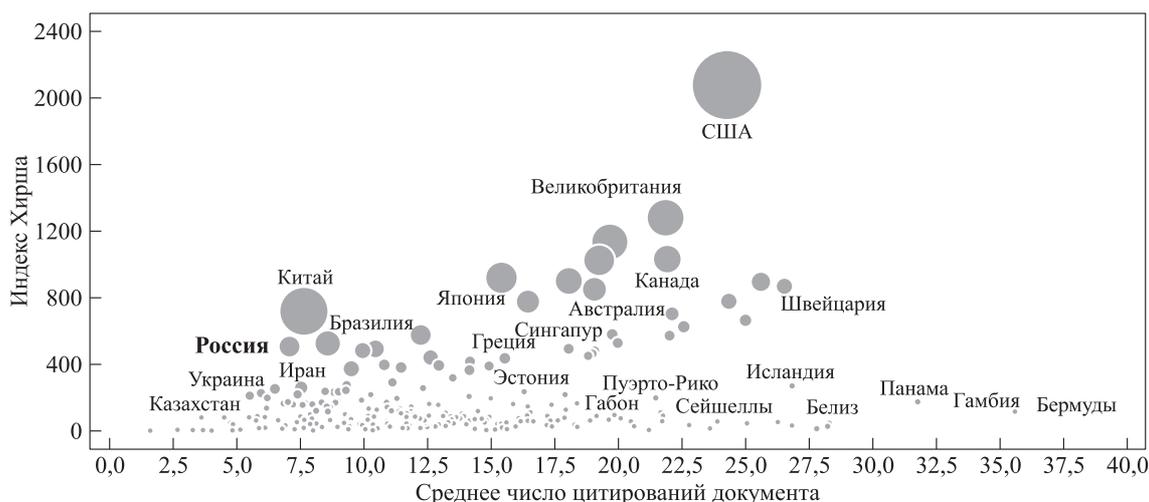


Рис. 2 Зависимость «странового» индекса Хирша от среднего числа цитирований

Россия уступает Тайваню, Сингапuru, Дании, Норвегии, Финляндии, Швеции. Очевидно, что в данной области необходимо достижение позитивной динамики, которая невозможна без привлечения новых молодых исследователей и улучшения условий для уже задействованных в научно-образовательной сфере.

Развитие кадрового и образовательного потенциала — один из ключевых драйверов роста согласно программе «Цифровая экономика Российской Федерации» (утверждена Распоряжением Правительства РФ от 28.07.2017 № 1632-р).

В качестве барьеров для развития молодых исследователей могут выступать и нормативные документы, и сложившиеся практики, и отсутствие должного уровня автоматизации процессов. Их преодоление не всегда возможно исключительно за счет привлечения дополнительных ресурсов, поэтому в данном контексте они будут называться нефинансовыми барьерами.

В настоящей статье рассматриваются соответствующие примеры по основным направлениям реализации СНТР (утверждена Указом Президента РФ № 642 от 01.12.2016) и предлагаются возможные способы их преодоления путем создания и использования специальных сервисов цифровых платформ.

В качестве конкретного примера рассмотрено соответствие между сервисами создаваемой в настоящий момент Федеральным исследовательским центром «Информатика и управление» Российской академии наук цифровой платформы «Наука и образование» [2] и соответствующими направлениями СНТР.

2 Концепция использования сервисов цифровых платформ

Переход к цифровой экономике и поэтапное развитие подходов Индустрии 4.0 [3] приводят к необходимости подготовки кадров, обладающих новым набором компетенций. Значительное влияние на данные процессы оказывает и стремительное развитие ИТ, предоставляющих инструменты формирования единой цифровой институциональной среды для различных заинтересованных сторон. В частности, одним из наиболее востребованных на сегодняшний день решений являются цифровые платформы, составляющие ИТ-базис для задач современной цифровой экономики.

Преодоление нефинансовых барьеров в определенной степени может осуществляться за счет нормативно-правовых (внесение изменений в действующие акты, положения) или финансовых (расширение инвестиций в научно-образовательную сферу целевым образом на поддержку и развитие молодежных кадров) мер. Однако с учетом экономической ситуации и потенциальных сложностей при изменении законодательства на первый план должны выходить более гибкие инструменты решения задач. Поэтому в данной работе предлагается подход на основе использования цифровых платформ, предполагающий создание или настройку сервисов, ориентированных на решение соответствующих задач (при этом, возможно, в рамках расширения базового функционала).

Безусловно, для создания таких решений требуется вовлечение широкого спектра материальных и организационных ресурсов, однако развитие так

называемой цифровой науки — одна из наиболее современных парадигм [4]. Это означает, что подобные решения будут так или иначе создаваться, а значит, в их рамках можно выделить необходимые инструменты, актуальные для развития молодежи.

Создание научно-образовательных цифровых платформ для решения задач государственного уровня необходимо осуществлять на основе так называемых центров превосходства [5], осуществляющих прорывные фундаментальные и прикладные исследования в наиболее важных и инновационных областях знания, обладающих уникальными интеллектуальными и материально-техническими ресурсами. В следующих разделах будут рассмотрены существующие барьеры в соответствии с направлениями СНТР, а также предложено использование или развитие соответствующих сервисов цифровой платформы [2] для преодоления существующих сложностей по каждому из них на примере ряда кейсов для научной молодежи.

3 Кадры и человеческий капитал

На сегодняшний день в научных и образовательных организациях редко внедряются наиболее современные технологии развития молодежного персонала, включая инструментарий формирования широкого спектра «непрофильных» компетенций (например, управленческих, юридических, финансовых) и иных инструментов создания полноценного кадрового резерва. Для молодых сотрудников необходимо проводить различные мероприятия (от интенсивов до полноценных образовательных курсов) с целью развития их компетенций в административном, инновационном и бизнес-направлениях. Подобное дополнительное образование (в том числе за счет организации) распространено в коммерческом секторе, однако в условиях формирования цифровой экономики необходима реализация данного подхода в более широком спектре направлений для формирования специалистов на стыке фундаментальной науки, инженерии и технологического предпринимательства.

Решение данной задачи возможно в рамках все более распространенного в мировой практике подхода цифровизации образования [6] и развертывания на базе цифровой платформы [2] образовательного сервиса как площадки для электронного и дистанционного обучения. Это позволит сформировать *цифровые* кадры во взаимодействии с университетами, федеральными органами исполнительной власти, институтами развития, государственными корпорациями. В рамках такого сервиса возможно внедрение в образовательный процесс

технологий искусственного интеллекта. Уже сейчас существуют решения для определения индивидуальных методов эффективного электронного обучения [7], планирования востребованности образовательных курсов [8] и самообучения [9]. Реализация и развитие таких подходов направлены на создание современной образовательной ИТ-эко-системы [10].

4 Инфраструктура и среда

Крайне важно, чтобы исследования молодых ученых находились в русле мировых тенденций и были ориентированы на достижение прорывных результатов. Для этого необходимо иметь возможность познакомиться с накопленными знаниями в профильных областях, а также смежных или даже просто потенциально интересных. В настоящий момент многие научно-образовательные организации получили доступ к базам Web of Science (преимущественно к Core Collection) и Scopus, однако в них представлены тексты статей только для журналов с открытым доступом. Таким образом, очевидна необходимость внедрения инструментов доступа к источникам научной информации, например с помощью подписки на ведущие мировые журналы. Сейчас такие механизмы реализуются, в частности, через конкурсы Российского фонда фундаментальных исследований (РФФИ), однако максимальное удобство и наиболее широкий инструментарий для дальнейшего использования возможно обеспечить только в рамках *информационного* сервиса цифровой платформы.

Еще один важный мировой тренд, ориентированный в том числе и на развитие национальных инновационных систем [11], — использование в научной деятельности наиболее современной исследовательской инфраструктуры, в том числе и установок класса мегасайенс. Столь сложное оборудование должно быть задействовано и в целях методологического и технологического обеспечения научно-образовательного процесса. В России в настоящий момент подобные объекты представлены центрами коллективного пользования (ЦКП) и уникальными научными установками (УНУ). Они достаточно активно используются в том числе и коммерческими заказчиками. При этом для молодых исследователей существуют определенные сложности с доступом к соответствующим ресурсам, так как они не могут гарантировать оплату или получение прорывных результатов, публикации в престижных международных журналах (один из отчетных показателей деятельности ЦКП и УНУ). Более того, здесь важен и образователь-

ный эффект — молодым коллективам нужно научиться грамотно использовать подобные инфраструктурные решения, прежде чем будут получены результаты (мировой опыт по «удаленным» лабораторным работам описан, например, в статье [12]). Таким образом, необходимо предусмотреть специальные механизмы поддержки исследований молодых ученых на ЦКП и УНУ, например с помощью введения специальных квот. Их выделение, использование, а также наукометрический эффект могут быть учтены с помощью *системы управления научными сервисами (СУС)* — специализированного сервиса платформы [2], который ориентирован на решение подобного класса задач для интегрированного оборудования ЦКП и УНУ. Финансовые процедуры (например, выделение субсидий на оплату использования) реализуются в *обеспечивающих* сервисах, а современная интеллектуальная обработка научных данных исследовательскими коллективами проводится с помощью специализированных *научных сервисов*.

5 Взаимодействие и кооперация

В настоящее время для молодых исследователей доступны несколько потенциальных возможностей коммерциализации их научных исследований: создание малых инновационных предприятий (МИП), спин-офф компаний, стартапов в рамках проектов институтов развития, включая технопарки и бизнес-инкубаторы. Однако вопросы привлечения дополнительного финансирования могут серьезно отвлекать молодежные коллективы именно от содержательной составляющей научных и инновационных исследований. В подобных условиях достаточно трудно получить столь необходимый опыт в рамках схемы «идея — научные исследования — разработка прототипа — запуск в производство», так как поддержку получают только отдельные отобранные коллективы. Поэтому оправдан проект по созданию массовых технопарков и бизнес-инкубаторов, в которых одной из основных целей стало бы обучение механизму трансфера технологий [13], правильному распределению ролей в команде, поддержке института техноброкерства. Такую практику можно было бы назвать *научными стартапами*. Сами научные организации в силу разных причин зачастую не заинтересованы в создании МИП и соответствующей поддержке. В таких условиях консультационно-организационная помощь от представителей бизнеса (менторов), а также более опытных ученых чрезвычайно важна — и может быть развернута в рамках сервиса *коммуникаций*. Полезным может оказаться и предоставление базовой организацией льготного доступа

к уникальному дорогостоящему оборудованию, которое обсуждалось в разд. 4.

6 Управление и инвестиции

Даже в рамках предложенного в разд. 5 формата научных стартапов и инкубаторов существенное финансирование будет выделяться по-настоящему интересным и действительно прорывным проектам. Однако при этом появится пространство для экспериментов, а по итогам функционирования таких площадок вполне естественно ожидать и увеличения общего числа качественных разработок, и понимания молодежью соответствующих инструментов. Вопрос финансирования здесь может взять на себя как бизнес, так и научные фонды, при этом цифровая платформа [2] предоставит инструменты эффективного взаимодействия различных субъектов цифровой экономики, а также систему учета и коммуникации лидеров и энтузиастов направлений.

Поддержка молодых ученых должна осуществляться в том числе и в рамках целевого субсидирования их научных исследований в рамках грантов и программ. Известно, что практика сочетания индивидуальных исследовательских проектов с решением прорывных задач на базе центров превосходства показала себя весьма эффективной с различных точек зрения [14]. Целевые молодежные проекты предлагаются в настоящий момент как Российским научным фондом в рамках Президентской программы, так и РФФИ, включая и инициативы сотрудничества с региональными властями. Существенная сложность заключается в том, что предпочтение зачастую отдается заявкам с хорошим научным (в том числе и публикационным) заделом — и начинающим исследователям достаточно сложно приобрести нужный опыт. Для этого и предложен инструмент научных стартапов, в рамках которых вопрос поддержки проектов может решаться с учетом иных подходов. Вторая задача — поиск молодых соисполнителей (или руководителя проекта). Эта проблема также может быть решена с помощью сервиса коммуникаций, а *аналитические* сервисы обеспечат сбор, систематизацию и эффективную обработку наукометрической информации, а также предоставят рекомендации по дальнейшим проектам активным представителям научной молодежи.

7 Сотрудничество и интеграция

Необходимо отметить значимый экономический эффект, возникающий на стыке научно-про-

мышленного сотрудничества [15]. Прорывные достижения в рамках научных исследований стимулируют развитие технологических инновационных процессов, формирование и задействование новых компетенций, проявляется эффект от трансфера технологий.

Это позволяет сочетать вклад в фундаментальную науку с формированием новых индустриально-инновационных отраслей бизнеса. Поэтому необходима поддержка образовательной и инновационной мобильности молодых ученых в ведущих отечественных и зарубежных организациях, выполнение совместных научных исследований (например, делегирование молодых исследователей для участия в мегасайенс-проектах, таких как Compact Linear Collider, European XFEL (X-ray Free Electron Laser), Facility for Antiproton and Ion Research, Human Proteom Project, ITER (International Thermonuclear Experimental Reactor) и др.), их участие в международных конференциях. Очевидно, что данный пункт существенным образом связан с дополнительным финансированием и заключением партнерских соглашений, однако необходимо выделить и сугубо организационные моменты, связанные с деталями оформления командировок.

Данные процессы могут быть полностью автоматизированы в рамках обеспечивающих сервисов, что упростит саму процедуру как для молодых ученых, так и для участвующих организаций, а также позволит избежать возможных ошибок при оформлении документов, финансовых и временных затрат на их исправление и пересогласование. Кроме того,

цифровые платформы предоставляют эффективные инструменты для организации и проведения совместных исследований распределенными научными коллективами, включая создание виртуальных лабораторий.

8 Заключение

В статье выделены некоторые нефинансовые барьеры для развития молодежи в сфере науки и образования в соответствии с направлениями СНТР. Концептуальная схема соответствия представлена на рис. 3.

Данная схема отражает тесную взаимосвязь между направлениями СНТР и означает, что предложенное разделение сервисов цифровой платформы «Наука и образование» по ним носит несколько условный характер. Некоторые сервисы могут быть использованы для решения задач сразу из нескольких отраслей, при этом значительный эффект достигается именно при условии их комбинации. Часть рассмотренных барьеров может быть устранена и с помощью дополнительных инвестиций, однако в рамках идеологии «больших вызовов» СНТР такие меры не всегда оказываются однозначно эффективными, так как зачастую требуется изменение среды. Подход на основе сервисов цифровой платформы представляется наиболее универсальным и современным. Кроме того, он является актуальным и для решения задач, необязательно связанных непосредственно с молодежной политикой.

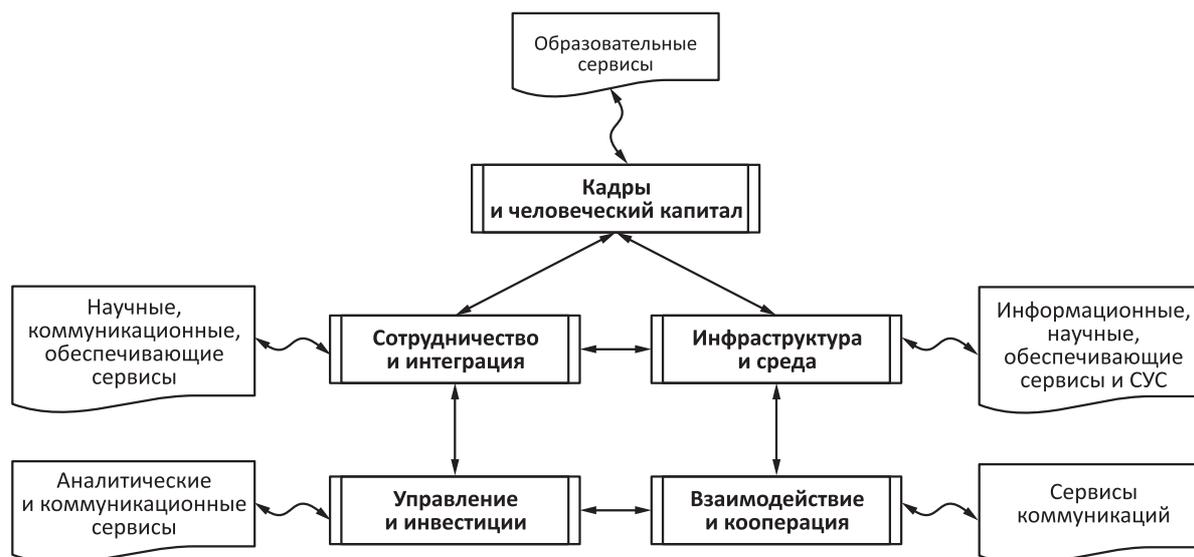


Рис. 3 Концептуальное соответствие между направлениями СНТР и сервисами цифровой платформы «Наука и образование»

Литература

1. 2018 Global R&D Funding Forecast // R&D Mag., Winter 2018. 36 p.
2. Зацаринный А. А., Горшенин А. К., Волович К. И., Коллин К. К., Кондрашев В. А., Степанов П. В. Управление научными сервисами как основа национальной цифровой платформы «Наука и образование» // Стратегические приоритеты, 2017. № 2(14). С. 103–113.
3. Шваб К. Четвертая промышленная революция. — М.: Эксмо, 2016. 208 с.
4. Зацаринный А. А., Горшенин А. К., Волович К. И., Кондрашев В. А. Основные направления развития информационных технологий в условиях вызовов цифровой экономики // Цифровая обработка сигналов, 2018. № 1. С. 3–7.
5. Заиченко С. А. Центры превосходства в системе современной научной политики // Форсайт, 2008. Т. 2. № 1. С. 42–50.
6. Paulsen M. F. Experiences with Learning Management Systems in 113 European institutions // Educ. Technol. Soc., 2003. Vol. 6. Iss. 4. P. 134–148.
7. Villaverde J. E., Godoy D., Amandi A. Learning styles' recognition in e-learning environments with feed-forward neural networks // J. Comput. Assist. Lear., 2006. Vol. 22. Iss. 3. P. 197–206.
8. Kardan A. A., Sadeghi H., Ghidary S. S., Sani M. R. F. Prediction of student course selection in online higher education institutes using neural network // Comput. Educ., 2013. Vol. 65. P. 1–11.
9. Kose U., Arslan A. Optimization of self-learning in computer engineering courses: An intelligent software system supported by artificial neural network and vortex optimization algorithm // Comput. Appl. Eng. Educ., 2017. Vol. 25. Iss. 1. P. 142–156.
10. Gorshenin A. Toward modern educational IT-ecosystems: From learning management systems to digital platforms // 10th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings. — Piscataway, NJ, USA: IEEE, 2018. P. 329–333.
11. Autio E., Hameri A.-P., Vuola O. A framework of industrial knowledge spillovers in big-science centers // Res. Policy, 2004. Vol. 33. Iss. 1. P. 107–126.
12. Rapuano S., Zoino F. A learning management system including laboratory experiments on measurement instrumentation // IEEE T. Instrum. Meas., 2006. Vol. 55. Iss. 5. P. 1757–1766.
13. Grosse R. International technology transfer in services // J. Int. Bus. Stud., 1996. Vol. 27. Iss. 4. P. 781–800.
14. Fortin J.-M., Currie D. J. Big science vs. little science: How scientific impact scales with funding // PLoS One, 2013. Vol. 8. Iss. 6. Art. No. e65263.
15. Vuola O., Hameri A.-P. Mutually benefiting joint innovation process between industry and big-science // Technovation, 2006. Vol. 26. Iss. 1. P. 3–12.

Поступила в редакцию 20.09.18

DEVELOPMENT OF SERVICES OF DIGITAL PLATFORMS TO OVERCOME NONFINANCIAL BARRIERS

A. K. Gorshenin^{1,2}

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

Abstract: The article discusses examples of various nonfinancial barriers that impede the effective development of young people in science and education. Overcoming them is not always possible solely by attracting additional resources, since changes in the environment are often required. Conceptual ways to overcome such barriers through creation and development of digital platform services, one of key paradigms in modern information technologies, are proposed. Using the example of the digital platform “Science and Education” created in the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, the compliance of services with the main directions of the Scientific and Technological Development Strategy of the Russian Federation is demonstrated. The proposed concept is effective for solving a wider class of problems.

Keywords: digital platforms; nonfinancial barriers; strategy of scientific and technological development; digital economy; youth policy

DOI: 10.14357/19922264180415

Acknowledgments

The research was supported by the Russian Science Foundation (project 18-71-00156).

References

1. 2018 Global R&D Funding Forecast. Winter 2018. *R&D Mag.* 36 p.
2. Zatsarinnyy, A. A., A. K. Gorshenin, K. I. Volovich, K. K. Kolin, V. A. Kondrashev, and P. V. Stepanov. 2017. Upravlenie nauchnymi servisami kak osnova natsional'noy tsifrovoy platformy "Nauka i obrazovanie" [Management of scientific services as the basis of the national digital platform "Science and Education"]. *Strategic Priorities* 2(14):103–113.
3. Schwab, K. M. 2017. *The fourth industrial revolution*. London: Penguin U.K. 192 p.
4. Zatsarinnyy, A. A., A. K. Gorshenin, K. I. Volovich, and V. A. Kondrashev. 2018. Osnovnye napravleniya razvitiya informatsionnykh tekhnologiy v usloviyakh vyzovov tsifrovoy ekonomiki [The main trends of the development of information technologies within the challenges of the digital economy]. *Digital Signal Processing* 1:3–7.
5. Zaichenko, S. 2008. Tsentry prevoskhodstva v sisteme sovremennoy nauchnoy politiki [Centres of excellence in the system of contemporary science policy]. *Foresight and STI Governance* 2(1):42–50.
6. Paulsen, M. F. 2003. Experiences with Learning Management Systems in 113 European institutions. *Educ. Technol. Soc.* 6(4):134–148.
7. Villaverde, J. E., D. Godoy, and A. Amandi. 2006. Learning styles' recognition in e-learning environments with feed-forward neural networks. *J. Comput. Assist. Lear.* 22(3):197–206.
8. Kardan, A. A., H. Sadeghi, S. S. Ghidary, and M. R. F. Sani. 2013. Prediction of student course selection in online higher education institutes using neural network. *Comput. Educ.* 65:1–11.
9. Kose, U., and A. Arslan. 2017. Optimization of self-learning in computer engineering courses: An intelligent software system supported by artificial neural network and vortex optimization algorithm. *Comput. Appl. Eng. Educ.* 25(1):142–156.
10. Gorshenin, A. 2018. Toward modern educational IT-ecosystems: From learning management systems to digital platforms. *10th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops Proceedings*. Piscataway, NJ: IEEE. 329–333.
11. Autio, E., A.-P. Hameri, and O. Vuola. 2004. A framework of industrial knowledge spillovers in big-science centers. *Res. Policy* 33(1):107–126.
12. Rapuano, S., and F. Zoino. 2006. A learning management system including laboratory experiments on measurement instrumentation. *IEEE T. Instrum. Meas.* 55(5):1757–1766.
13. Grosse, R. 1996. International technology transfer in services. *J. Int. Bus. Stud.* 27(4):781–800.
14. Fortin, J.-M., and D. J. Currie. 2013. Big science vs. little science: How scientific impact scales with funding. *PLoS One* 8(6):e65263.
15. Vuola, O., and A.-P. Hameri. 2006. Mutually benefiting joint innovation process between industry and big-science. *Technovation* 26(1):3–12.

Received September 20, 2018

Contributor

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, leading scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; leading scientist, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; agorshenin@frccsc.ru

Агаларов Явер Мирзабекович (р. 1952) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Арутюнов Евгений Наполеонович (р. 1952) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Басок Борис Моисеевич (р. 1948) — кандидат технических наук, доцент МИРЭА — Российского технологического университета

Бахтеев Олег Юрьевич (р. 1991) — аспирант Московского физико-технического института

Бунтман Надежда Валентиновна (р. 1957) — кандидат филологических наук, доцент Московского государственного университета им. М. В. Ломоносова

Гайдамака Юлия Васильевна (р. 1971) — доктор физико-математических наук, профессор кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гончаров Александр Анатольевич (р. 1994) — инженер Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, доцент, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; ведущий научный сотрудник факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Николай Александрович (р. 1982) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Дорофеева Александра Владимировна (р. 1991) — аспирант кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Захаров Виктор Николаевич (р. 1948) — доктор технических наук, доцент, ученый секретарь Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зейфман Александр Израилевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой прикладной математики Вологодского государственного университета; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; главный научный сотрудник Вологодского научного центра Российской академии наук

Кондранин Егор Сергеевич (р. 1995) — студент магистратуры кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Коновалов Михаил Григорьевич (р. 1950) — доктор технических наук, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; ведущий научный сотрудник

Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор Университета Дианьзи города Ханчжоу (Китай)

Кудрявцев Алексей Андреевич (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Левыкин Михаил Владимирович (р. 1985) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Логачев Олег Алексеевич (р. 1950) — кандидат физико-математических наук, заведующий отделом Института проблем информационной безопасности Московского государственного университета им. М. В. Ломоносова

Нуриев Виталий Александрович (р. 1980) — кандидат филологических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Самуйлов Константин Евгеньевич (р. 1955) — доктор технических наук, профессор, заведующий кафедрой прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Смердов Антон Николаевич (р. 1995) — студент Московского физико-технического института

Стрижов Вадим Викторович (р. 1967) — доктор физико-математических наук, профессор кафедры интеллектуальных систем Московского физико-технического института, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Титова Анастасия Игоревна (р. 1995) — студентка кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Ушаков Владимир Георгиевич (р. 1952) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Френкель Сергей Лазаревич (р. 1951) — кандидат технических наук, доцент, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ханкин Даниэль (р. 1983) — магистр, докторант отделения информатики университета им. Бен-Гуриона в Негеве, Беер-Шева, Израиль

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2018 г.

	Вып.	Стр.
Агаларов Я. М. Оптимизация объема буферной памяти узла коммутации при схеме полного разделения памяти	4	25–32
Агасандян Г. А. Континуальный критерий VaR на сценарных рынках	1	31–39
Алешин И. С. О формальной постановке задач поиска сгущений в разреженных булевых матрицах	1	40–48
Арутюнов Е. Н., Кудрявцев А. А., Титова А. И. Гамма-вейбулловский случай априорных распределений в байесовских моделях массового обслуживания	4	92–95
Атаева О. М., Серебряков В. А. Онтология цифровой семантической библиотеки LibMeta	1	2–10
Басок Б. М., Захаров В. Н., Френкель С. Л. Использование вероятностной модели вычислений для тестирования одного класса готовых к использованию программных компонентов локальных и сетевых систем	4	44–51
Батенков А. А., Маньяков Ю. А., Гасилов А. В., Яковлев О. А. Математическая модель оптимальной триангуляции	2	69–74
Бахтеев О. Ю. см. Огальцов А. В.		
Бахтеев О. Ю. см. Смердов А. Н.		
Борисов А. В. Фильтрация состояний марковских скачкообразных процессов по дискретизованным наблюдениям	3	115–121
Босов А. В., Игнатов А. Н., Наумов А. В. Модель передвижения поездов и маневровых локомотивов на железнодорожной станции в приложении к оценке и анализу вероятности бокового столкновения	3	107–114
Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. I. Оптимальное решение методом динамического программирования	3	99–106
Бунтман Н. В., Гончаров А. А., Зацман И. М., Нуриев В. А. Количественный анализ результатов машинного перевода с использованием надкорпусных баз данных	4	96–105
Бунтман Н. В. см. Нуриев В. А.		
Быковец Е. В., Лаврентьев В. В., Назаров Л. В. Вероятностная модель влияния книги заказов на процесс цены	2	29–34
Васильева С. Н., Кан Ю. С. Алгоритм визуализации плоского ядра вероятностной меры	2	60–68
Виноградов Д. В. Учет предварительных оценок скорости порождения сходств спаривающей цепью Маркова	1	49–54
Вышинский Л. Л., Флеров Ю. А., Широков Н. И. Автоматизированная система весового проектирования самолетов	1	18–30
Гайдамака Ю. В. см. Горбунова А. В.		
Гайдамака Ю. В. см. Самуйлов К. Е.		
Гасилов А. В. см. Батенков А. А.,		
Гончаров А. А. см. Бунтман Н. В.		
Горбунова А. В., Наумов В. А., Гайдамака Ю. В., Самуйлов К. Е. Ресурсные системы массового обслуживания как модели беспроводных систем связи	3	48–55
Горшенин А. К. Зашумление данных конечными смесями нормальных и гамма-распределений с применением к задаче округления наблюдений	3	28–34
Горшенин А. К. Развитие сервисов цифровых платформ для преодоления нефинансовых барьеров	4	106–112
Горшенин А. К., Королев В. Ю. Определение экстремальности объемов осадков на основе модифицированного метода превышения порогового значения	4	16–24
Горшенин А. К. см. Королев В. Ю.		

	Вып.	Стр.
Грушо А. А., Грушо Н. А., Забежайло М. И., Смирнов Д. В., Тимонина Е. Е. Параметризация в прикладных задачах поиска эмпирических причин	3	62–66
Грушо А. А., Грушо Н. А., Левыкин М. В., Тимонина Е. Е. Методы идентификации захвата хоста в распределенной информационно-вычислительной системе, защищенной с помощью метаданных	4	39–43
Грушо А. А., Забежайло М. И., Зацаринный А. А., Тимонина Е. Е. О некоторых возможностях управления ресурсами при организации проактивного противодействия компьютерным атакам	1	62–70
Грушо А. А., Тимонина Е. Е., Шоргин С. Я. Иерархический метод порождения метаданных для управления сетевыми соединениями	2	44–49
Грушо Н. А. см. Грушо А. А.		
Грушо Н. А. см. Грушо А. А.		
Дорофеева А. В. см. Королев В. Ю.		
Егоров А. Ю. см. Шнурков П. В.		
Егоров А. Ю. см. Шнурков П. В.		
Жуков Д. О., Хватова Т. Ю., Лесько С. А., Зальцман А. Д. Влияние плотности связей на кластеризацию и порог перколяции при распространении информации в социальных сетях	2	90–97
Забежайло М. И. см. Грушо А. А.		
Забежайло М. И. см. Грушо А. А.		
Зальцман А. Д. см. Жуков Д. О.		
Захаров В. Н. см. Басок Б. М.		
Захаров В. Н. см. Шанин И. А.		
Зацаринный А. А., Сучков А. П. Система ситуационного управления как мультисервисная технология в облачной среде	1	78–88
Зацаринный А. А. см. Грушо А. А.		
Зацман И. М. Имплицированные знания: основания и технологии извлечения	3	74–82
Зацман И. М. см. Бунтман Н. В.		
Зейфман А. И. см. Королев В. Ю.		
Зубарев Д. В. см. Соченков И. В.		
Игнатов А. Н. см. Босов А. В.		
Инькова О. Ю., Кружков М. Г. Статистический анализ лингвоспецифичности коннекторов (на материале параллельных корпусов)	3	83–90
Инькова О. Ю. см. Нуриев В. А.		
Кан Ю. С. см. Васильева С. Н.		
Ковалёв С. П. Теория категорий как математическая прагматика модельно-ориентированной системной инженерии	1	95–104
Козеренко Е. Б., Кузнецов К. И., Романов Д. А. Семантическая обработка неструктурированных текстовых данных на основе лингвистического процессора PullEnti	3	91–98
Кондранин Е. С., Ушаков В. Г. Система обслуживания с относительным приоритетом и профилактиками прибора	4	33–38
Коновалов М. Г., Разумчик Р. В. Сравнение двух механизмов активного управления очередью в системе $M/D/1/N$	4	9–15
Коновалов М. Г., Разумчик Р. В. Управление случайным блужданием с эталонным стационарным распределением	3	2–13
Королев В. Ю., Горшенин А. К., Зейфман А. И. Новые представления обобщенного распределения Миттаг-Леффлера в виде смесей и их приложения	4	75–85
Королев В. Ю., Дорофеева А. В. О неравномерных оценках точности нормальной аппроксимации для распределений некоторых случайных сумм при ослабленных моментных условиях	4	86–91
Королев В. Ю. см. Горшенин А. К.		
Кривенко М. П. Обучаемая классификация данных с учетом анализа главных компонент	3	56–61
Кривенко М. П. Реконструкция осей главных компонент	1	71–77
Кружков М. Г. см. Инькова О. Ю.		

	Вып.	Стр.
Кудрявцев А. А. Байесовские модели баланса	3	18–27
Кудрявцев А. А., Шестаков О. В. Байесовские модели тестирования больших групп обслуживающих приборов	1	105–108
Кудрявцев А. А., Шестаков О. В. Минимизация ошибок вычисления вейвлет-коэффициентов при решении обратных задач	2	17–23
Кудрявцев А. А. см. Арутюнов Е. Н.		
Кузнецов К. И. см. Козеренко Е. Б.		
Лаврентьев В. В. см. Быковец Е. В.		
Лебедев А. В. Максимальные ветвящиеся процессы в случайной среде	2	35–43
Левыкин М. В. см. Грушо А. А.		
Лери М. М., Павлов Ю. Л. Об устойчивости конфигурационных графов в случайной среде	2	2–10
Лесько С. А. см. Жуков Д. О.		
Логачев О. А. Теоретико-информационная характеристика совершенно уравновешенных функций	4	70–74
Малашенко Ю. Е., Назарова И. А., Новикова Н. М. Анализ разрезных повреждений в многополюсных сетях	3	35–41
Малашенко Ю. Е., Назарова И. А., Новикова Н. М. Диаграммы уязвимости потоковых сетевых систем	1	11–17
Маньяков Ю. А. см. Батенков А. А.		
Мирзабеков Я. М., Шихиев Ш. Б. Дискретный анализ в синтаксическом анализе	2	98–104
Мистрюков А. В., Ушаков В. Г. Достаточные условия эргодичности приоритетных систем массового обслуживания	2	24–28
Назаров Л. В. см. Быковец Е. В.		
Назарова И. А. см. Малашенко Ю. Е.		
Назарова И. А. см. Малашенко Ю. Е.		
Наумов А. В. см. Босов А. В.		
Наумов В. А. см. Горбунова А. В.		
Наумов В. А. см. Сопин Э. С.		
Новикова Н. М. см. Малашенко Ю. Е.		
Новикова Н. М. см. Малашенко Ю. Е.		
Нуриев В. А., Бунтман Н. В., Инькова О. Ю. Ошибки и неточности машинного перевода русских коннекторов на французский язык	2	105–113
Нуриев В. А. см. Бунтман Н. В.		
Огальцов А. В., Бахтеев О. Ю. Автоматическое извлечение метаданных из научных PDF-документов	2	75–82
Павлов Ю. Л. см. Лери М. М.		
Разумчик Р. В. см. Коновалов М. Г.		
Разумчик Р. В. см. Коновалов М. Г.		
Романов Д. А. см. Козеренко Е. Б.		
Самуйлов К. Е., Гайдамака Ю. В., Шоргин С. Я. Применение моделей случайного блуждания при моделировании перемещения устройств в беспроводной сети	4	2–8
Самуйлов К. Е. см. Горбунова А. В.		
Самуйлов К. Е. см. Сопин Э. С.		
Серебряков В. А. см. Атаева О. М.		
Синицын И. Н. Метод интерполяционного аналитического моделирования одномерных распределений в стохастических системах	1	55–61
Смердов А. Н., Бахтеев О. Ю., Стрижов В. В. Выбор оптимальной модели рекуррентной сети в задачах поиска парафразы	4	63–69
Смирнов Д. В. см. Грушо А. А.		
Сопин Э. С., Наумов В. А., Самуйлов К. Е. Об инвариантности стационарного распределения системы массового обслуживания с ограниченными ресурсами и с интенсивностями поступления и обслуживания, зависящими от состояния системы	3	42–47
Соченков И. В., Зубарев Д. В., Тихомиров И. А. Эксплоративный патентный поиск	1	89–94

	Вып.	Стр.
Стефанович А. И. см. Босов А. В.		
Стрижов В. В. см. Смердов А. Н.		
Ступников С. А. см. Шанин И. А.		
Сурина А. А. см. Тырсин А. Н.		
Сучков А. П. см. Зацаринный А. А.		
Сюнтюренко О. В. Финансирование фундаментальных исследований: концептуальный облик системы поддержки принятия решений с использованием методов наукометрии и анализа данных	1	118–127
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Титова А. И. см. Арутюнов Е. Н.		
Тихомиров И. А. см. Соченков И. В.		
Тырсин А. Н., Сурина А. А. Модели управления риском в гауссовских стохастических системах	2	50–59
Ушаков В. Г. см. Кондранин Е. С.		
Ушаков В. Г. см. Мистрюков А. В.		
Флеров Ю. А. см. Вышинский Л. Л.		
Френкель С. Л., Ханкин Д. Непрерывные обновления маршрута в SDN с использованием проверки соответствия качеству обслуживания	4	52–62
Френкель С. Л. см. Басок Б. М.		
Ханкин Д. см. Френкель С. Л.		
Хватова Т. Ю. см. Жуков Д. О.		
Шанин И. А., Ступников С. А., Захаров В. Н. Методы и средства обнаружения нештатных ситуаций, возникающих на элементах жилищно-коммунальной инфраструктуры	3	67–73
Шестаков О. В. Несмещенная оценка риска стабилизированной жесткой пороговой обработки в модели с долгосрочной зависимостью	2	11–16
Шестаков О. В. Среднеквадратичный риск пороговой обработки при случайном объеме выборки	3	14–17
Шестаков О. В. см. Кудрявцев А. А.		
Шестаков О. В. см. Кудрявцев А. А.		
Широков Н. И. см. Вышинский Л. Л.		
Шихиев Ш. Б. см. Мирзабеков Я. М.		
Шнурков П. В., Егоров А. Ю. Разработка и предварительное исследование стохастической полумарковской модели управления запасом непрерывного продукта при постоянно происходящем потреблении	1	109–117
Шнурков П. В., Егоров А. Ю. Решение проблемы оптимального управления запасом непрерывного продукта при постоянно происходящем потреблении в стохастической полумарковской модели	2	83–89
Шоргин С. Я. см. Грушо А. А.		
Шоргин С. Я. см. Самуйлов К. Е.		
Яковлев О. А. см. Батенков А. А.		

2018 AUTHOR INDEX

	Issue	Page
Agalarov Ya. M. Optimization of buffer memory size of switching node in mode of full memory sharing	4	25–32
Agasandyan G. A. Continuous VaR-criterion in scenario markets	1	31–39
Aleshin I. S. On the formalization of tasks searching dense submatrices in boolean sparse matrices	1	40–48
Arutyunov E. N., Kudryavtsev A. A., and Titova A. I. Gamma-Weibull <i>a priori</i> distributions in Bayesian queuing models	4	92–95
Ataeva O. M. see Serebryakov V. A.		
Bakhteev O. Y. see Ogaltsov A. V.		
Bakhteev O. Y. see Smerdov A. N.		
Basok B. M., Zakharov V. N., and Frenkel S. L. Using a probabilistic calculation model to test one class of ready-to-use software components of local and network systems	4	44–51
Batenkov A. A., Maniakov Yu. A., Gasilov A. V., and Yakovlev O. A. Mathematical model of optimal triangulation	2	69–74
Borisov A. V. Filtering of Markov jump processes by discretized observations	3	115–121
Bosov A. V., Ignatov A. N., and Naumov A. V. Model of transportation of trains and shunting locomotives at a railway station for evaluation and analysis of side-collision probability	3	107–114
Bosov A. V. and Stefanovich A. I. Stochastic differential system output control by the quadratic criterion. I. Dynamic programming optimal solution	3	99–106
Buntman N. V., Goncharov A. A., Zatsman I. M., and Nuriev V. A. Using supracorpora databases for quantitative analysis of machine translations	4	96–105
Buntman N. V. see Nuriev V. A.,		
Bykovets E. V. see Nazarov L. V.		
Dorofeeva A. V. see Korolev V. Yu.		
Egorov A. Y. see Shnurkov P. V.		
Egorov A. Y. see Shnurkov P. V.		
Flerov Yu. A. see Vyshinsky L. L.		
Frenkel S. L. and Khankin D. Seamless route updates in software-defined networking via quality of service compliance verification	4	52–62
Frenkel S. L. see Basok B. M.		
Gaidamaka Yu. V. see Gorbunova A. V.		
Gaidamaka Yu. V. see Samouylov K. E.		
Gasilov A. V. see Batenkov A. A.		
Goncharov A. A. see Buntman N. V.		
Gorbunova A. V., Naumov V. A., Gaidamaka Yu. V., and Samouylov K. E. Resource queuing systems as models of wireless communication systems	3	48–55
Gorshenin A. K. Data noising by finite normal and gamma mixtures with application to the problem of rounded observations	3	28–34
Gorshenin A. K. Development of services of digital platforms to overcome nonfinancial barriers	4	106–112
Gorshenin A. K. and Korolev V. Yu. Determining the extremes of precipitation volumes based on the modified “Peaks over Threshold” method	4	16–24
Gorshenin A. K. see Korolev V. Yu.		
Grusho A. A., Grusho N. A., Levykin M. V., and Timonina E. E. Methods of identification of host capture in a distributed information system which is protected on the basis of meta data	4	39–43
Grusho A. A., Grusho N. A., Zabezhailo M. I., Smirnov D. V., and Timonina E. E. Parametrization in applied problems of search of empirical reasons	3	62–66

	Issue	Page
Grusho A. A., Timonina E. E., and Shorgin S. Ya. Hierarchical method of meta data generation for control of network connections	2	44–49
Grusho A. A., Zabezhailo M. I., Zatsarinny A. A., and Timonina E. E. On some possibilities of resource management for organizing active counteraction to computer attacks	1	62–70
Grusho N. A. see Grusho A. A.		
Grusho N. A. see Grusho A. A.		
Ignatov A. N. see Bosov A. V.		
Inkova O. Yu. and Kruzhkov M. G. Statistical analysis of language specificity of connectives based on parallel texts	3	83–90
Inkova O. Yu. see Nuriev V. A.,		
Kan Yu. S. see Vasil'eva S. N.		
Khankin D. see Frenkel S. L.		
Khvatova T. Yu. see Zhukov D. O.		
Kondranin E. S. and Ushakov V. G. A head of the line priority queue with working vacations	4	33–38
Konovalov M. G. and Razumchik R. V. Comparison of two active queue management schemes through the $M/D/1/N$ queue	4	9–15
Konovalov M. G. and Razumchik R. V. Finding control policy for one discrete-time Markov chain on $[0, 1]$ with a given invariant measure	3	2–13
Korolev V. Yu. and Dorofeeva A. V. On nonuniform estimates of accuracy of normal approximation for distributions of some random sums under relaxed moment conditions	4	86–91
Korolev V. Yu., Gorshenin A. K., and Zeifman A. I. New mixture representations of the generalized Mittag-Leffler distribution and their applications	4	75–85
Korolev V. Yu. see Gorshenin A. K.		
Kovalyov S. P. Category theory as a mathematical pragmatics of model-based systems engineering	1	95–104
Kozerenko E. B., Kuznetsov K. I., and Romanov D. A. Semantic processing of unstructured textual data based on the linguistic processor PullEnti	3	91–98
Krivenko M. P. Principal axes reconstruction	1	71–77
Krivenko M. P. Supervised learning classification of data taking into account principal component analysis	3	56–61
Kruzhkov M. G. see Inkova O. Yu.		
Kudryavtsev A. A. Bayesian balance models	3	18–27
Kudryavtsev A. A. and Shestakov O. V. Bayesian models for testing large groups of service devices	1	105–108
Kudryavtsev A. A. and Shestakov O. V. Minimization of errors of calculating wavelet coefficients while solving inverse problems	2	17–23
Kudryavtsev A. A. see Arutyunov E. N.		
Kuznetsov K. I. see Kozerenko E. B.		
Lavrentyev V. V. see Nazarov L. V.		
Lebedev A. V. Maximal branching processes in random environment	2	35–43
Leri M. M. and Pavlov Yu. L. On the robustness of configuration graphs in a random environment	2	2–10
Lesko S. A. see Zhukov D. O.		
Levykin M. V. see Grusho A. A.		
Logachev O. A. An information based criterion for perfectly balanced functions	4	70–74
Malashenko Yu. E., Nazarova I. A., and Novikova N. M. Analysis of cutting damages to multipolar networks	3	35–41
Malashenko Yu. E., Nazarova I. A., and Novikova N. M. Diagrams of the functional vulnerability of flow network systems	1	11–17
Maniakov Yu. A. see Batenkov A. A.		
Mirzabekov Ya. M. and Shihiev Sh. B. Discrete analysis in parsing	2	98–104
Mistryukov A. V. and Ushakov V. G. Sufficient ergodicity conditions for priority queues	2	24–28
Naumov A. V. see Bosov A. V.		
Naumov V. A. see Gorbunova A. V.		
Naumov V. A. see Sopin E. S.		

	Issue	Page
Nazarov L. V., Lavrentyev V. V., and Bykovets E. V. A probability model of the influence of the order book on the price process	2	29–34
Nazarova I. A. see Malashenko Yu. E.		
Nazarova I. A. see Malashenko Yu. E.		
Novikova N. M. see Malashenko Yu. E.		
Novikova N. M. see Malashenko Yu. E.		
Nuriev V. A., Buntman N. V., and Inkova O. Yu. Machine translation of russian connectives into french: Errors and quality failures	2	105–113
Nuriev V. A. see Buntman N. V.		
Ogaltsov A. V. and Bakhteev O. Y. Automatic metadata extraction from scientific PDF documents	2	75–82
Pavlov Yu. L. see Leri M. M.		
Razumchik R. V. see Konovalov M. G.		
Razumchik R. V. see Konovalov M. G.		
Romanov D. A. see Kozerenko E. B.		
Samouylov K. E., Gaidamaka Yu. V., and Shorgin S. Ya. Modeling movement of devices in a wireless network by random walk models	4	2–8
Samouylov K. E. see Gorbunova A. V.		
Samouylov K. E. see Sopin E. S.		
Serebryakov V. A. and Ataeva O. M. Ontology of the digital semantic library LibMeta	1	2–10
Shanin I. A., Stupnikov S. A., and Zakharov V. N. Methods and tools for fault detection on elements of housing and utility infrastructure	3	67–73
Shestakov O. V. Mean-square thresholding risk with a random sample size	3	14–17
Shestakov O. V. Unbiased risk estimate of stabilized hard thresholding in the model with a long-range dependence	2	11–16
Shestakov O. V. see Kudryavtsev A. A.		
Shestakov O. V. see Kudryavtsev A. A.		
Shihiev Sh. B. see Mirzabekov Ya. M.		
Shirokov N. I. see Vyshinsky L. L.		
Shnurkov P. V. and Egorov A. Y. Development and preliminary study of a stochastic semi-Markov model of continuous supply of product management under the condition of constant consumption	1	109–117
Shnurkov P. V. and Egorov A. Y. Solution to the problem of optimal control of a stochastic semi-Markov model of continuous supply of product management under the condition of constantly happening consumption	2	83–89
Shorgin S. Ya. see Grusho A. A.		
Shorgin S. Ya. see Samouylov K. E.		
Sinitsyn I. N. Method of interpolational analytical modeling of processes in stochastic systems	1	55–61
Smerdov A. N., Bakhteev O. Y., and Strijov V. V. Optimal recurrent neural network model in paraphrase detection	4	63–69
Smirnov D. V. see Grusho A. A.		
Sochenkov I. V., Zubarev D. V., and Tikhomirov I. A. Exploratory patent search	1	89–94
Sopin E. S., Naumov V. A., and Samouylov K. E. On the insensitivity of the stationary distribution of the limited resources queuing system with state-dependent arrival and service rates	3	42–47
Stefanovich A. I. see Bosov A. V.		
Strijov V. V. see Smerdov A. N.		
Stupnikov S. A. see Shanin I. A.		
Suchkov A. P. see Zatsarinny A. A.		
Surina A. A. see Tyrsin A. N.		
Syntyurenko O. V. Financing of basic research: Conceptual shape of a system of support of decision-making with use of methods of scientometrics and analysis of data	1	118–127
Tikhomirov I. A. see Sochenkov I. V.		
Timonina E. E. see Grusho A. A.		
Timonina E. E. see Grusho A. A.		

	Issue	Page
Timonina E. E. see Grusho A. A.		
Timonina E. E. see Grusho A. A.		
Titova A. I. see Arutyunov E. N.		
Tyrsin A. N. and Surina A. A. A model of risk management in Gaussian stochastic systems	2	50–59
Ushakov V. G. see Kondranin E. S.		
Ushakov V. G. see Mistryukov A. V.		
Vasil'eva S. N. and Kan Yu. S. A visualization algorithm for the plane probability measure kernel	2	60–68
Vinogradov D. V. Influence of preliminary estimates on the speed of search of similarities by the coupling Markov chain	1	49–54
Vyshinsky L. L., Flerov Yu. A., and Shirokov N. I. Computer-aided system of aircraft weight design	1	18–30
Yakovlev O. A. see Batenkov A. A.		
Zabekhailo M. I. see Grusho A. A.		
Zabekhailo M. I. see Grusho A. A.		
Zakharov V. N. see Basok B. M.		
Zakharov V. N. see Shanin I. A.		
Zaltsman A. D. see Zhukov D. O.		
Zatsarinny A. A. and Suchkov A. P. The situational management system as a multiservice technology in the cloud	1	78–88
Zatsarinny A. A. see Grusho A. A.,		
Zatsman I. M. Implied knowledge: Foundations and technologies of explication	3	74–82
Zatsman I. M. see Buntman N. V.		
Zeifman A. I. see Korolev V. Yu.		
Zhukov D. O., Khatova T. Yu., Lesko S. A., and Zaltsman A. D. The influence of the connections' density on clusterization and percolation threshold during information distribution in social networks	2	90–97
Zubarev D. V. see Sochenkov I. V.		



Станислав Васильевич Емельянов

18.05.1929–15.11.2018

Редакционная коллегия журнала «Информатика и её применения» с глубоким прискорбием извещает, что 15 ноября 2018 года на 90-м году жизни скончался первый Главный редактор журнала «Информатика и её применения», сопредседатель Редакционного совета и член Редколлегии журнала Станислав Васильевич Емельянов — академик РАН, профессор, выдающийся ученый в области автоматики, системного анализа и информатики, научный руководитель ФИЦ ИУ РАН.

Станислав Васильевич Емельянов — основоположник многих новых научных направлений, в том числе теории систем с переменной структурой, теории бинарного управления и теории новых типов обратной связи, ориентированных на решение задач управления в условиях неопределенности и сильной изменчивости характеристик объекта. Эти теории получили широкое мировое признание, активно развиваются и используются при решении актуальных практических задач в важнейших отраслях техники и промышленного производства. Он — основатель крупной научной школы. Им подготовлена большая группа докторов и кандидатов наук. Среди его учеников — академики и члены-корреспонденты РАН. С. В. Емельянов — автор 25 книг, 278 статей в ведущих научных журналах, 72 авторских свидетельств на изобретения.

Станислав Васильевич Емельянов — лауреат: Ленинской премии (1972), Государственной премии СССР (1980), Премии Совета министров СССР (1981), Государственной премии РФ (1994), Премии Правительства РФ в области науки и технологий (2009), Премии Правительства РФ в области образования (2012). Он награжден орденами: Октябрьской революции (1974), Дружбы народов (1979), За заслуги перед Отечеством IV степени (1999) и III степени (2004), Почета (2010).

Станислав Васильевич вложил много сил в создание, становление и развитие нашего журнала, активно работал в Редакционном совете и Редколлегии журнала. Глубокая научная эрудиция, разносторонние знания, высокий профессионализм снискали любовь и уважение к нему коллег и друзей.

Память о Станиславе Васильевиче Емельянове навсегда сохранится в наших сердцах.

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или \LaTeX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 15 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на **русском и английском языках**:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01_rus/authors.asp и
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’—kompozit” [Experimental study of the strength of joints “steel—composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursosberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (Timoshenko S. P., Young D. H., Weaver W. *Vibration problems in engineering*. — 4th ed. — N.Y.: Wiley, 1974. 521 p.)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

Описание интернет-ресурса:

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

Описание диссертации или автореферата диссертации:

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: rust@ipiran.ru (Сейфуль-Мулюков Рустем Бадриевич)

<http://www.ipiran.ru/journal/issues/>

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or \LaTeX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 15 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English**:

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format: (see site)

http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp and

http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
 - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences;
 - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/bgn/>).
- Please take into account the following examples of Russian references appearance:
- Article in journal:**
Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.
- Journal article in electronic format:**
Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).
- Article from the continuing publication (collection of works, proceedings):**
Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental’noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.
- Conference proceedings:**
Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursosberegayushchie tekhnologii nedropol’zovaniya i povysheniya neftegazoidachi” [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]*. Moscow. 267–272.
- Books and other monographs:**
Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem [Operation of turbine generators with direct cooling]*. Moscow: Energy Publs. 352 p.
- Dissertation and Thesis:**
Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovarya informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.
- State standards and patents:**
GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch’yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.
Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel’nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.
- References in Latin transcription are presented in the original language.
References in the text are numbered according to the order of their first appearance; the number is placed in square brackets.
All items from the reference list should be cited.
10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.
11. Submissions of files by e-mail must include:
- the journal title and author’s name in the “Subject” field;
 - an article and additional materials have to be attached using the “attach” function;
 - an electronic version of the article should contain the file with the text and a separate file with figures.
12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia
Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05
e-mail: rust@ipiran.ru (to Prof. Rustem Seyful-Mulyukov)
<http://www.ipiran.ru/english/journal.asp>