

Информатика и её применения

Том 9 Выпуск 4 Год 2015

СОДЕРЖАНИЕ

Статистическое моделирование турбулентных потоков тепла между океаном и атмосферой с помощью метода скользящего разделения конечных нормальных смесей <i>В. Ю. Королев, А. К. Горшенин, С. К. Гулев, К. П. Беляев</i>	3
Обобщенные дисперсионные гамма-распределения как модели статистических закономерностей на финансовых рынках <i>В. Ю. Королев, А. Ю. Корчагин, И. А. Соколов</i>	14
Потоковая модель вычислений как парадигма программирования будущего <i>Д. Н. Змеев, А. В. Климов, Н. Н. Левченко, А. С. Окунев, А. Л. Стемповский</i>	29
Аналитическое моделирование процессов в динамических системах с цилиндрическими бесселевыми нелинейностями <i>И. Н. Синицын</i>	37
К анализу показателей качества обслуживания в современных беспроводных сетях <i>О. Г. Вихрова, К. Е. Самуйлов, Э. С. Сопин, С. Я. Шоргин</i>	48
Обзор моделей и алгоритмов размещения заданий в системах с параллельным обслуживанием <i>М. Г. Коновалов, Р. В. Разумчик</i>	56
Algebraic method for approximating joint stationary distribution in finite capacity queue with negative customers and two queues <i>R. Razumchik</i>	68
Performance improvement of Lempel–Ziv–Welch compression algorithm <i>S. Frenkel, M. Kopeetsky, R. Molotkovski, and P. Borovsky</i>	78
Оценки скорости передачи информации и пропускной способности в скрытых каналах с метками <i>А. А. Грушо, Н. А. Грушо, Е. Е. Тимонина</i>	85
Контроль и управление информационными потоками в облачной среде <i>А. А. Грушо, М. И. Забежайло, А. А. Зацаринный</i>	91

Информатика и её применения

Том 9 Выпуск 4 Год 2015

СОДЕРЖАНИЕ

Мелкозернистые гибридные интеллектуальные системы. Часть 1. Лингвистический подход И. А. Кириков, А. В. Колесников, С. В. Листопад, С. Б. Румовская	98
Об усреднении округленных данных В. Г. Ушаков, Н. Г. Ушаков	106
Об авторах	110
Авторский указатель за 2015 г.	112
2015 Author Index	115
Правила подготовки рукописей	118
Requirements for manuscripts	121

Технический редактор *Л. Кокушкина* Художественный редактор *М. Седакова*
Сдано в набор 01.10.15. Подписано в печать 18.12.15. Формат 60 x 84 / 8
Бумага офсетная. Печать цифровая. Усл.-печ. л. 14,2. Уч.-изд. л. 13,5. Тираж 100 экз.
Заказ № 2028к
Издательство «ТОРУС ПРЕСС», Москва 121614, ул. Крылатская, 29-1-43
Отпечатано в Академиздатцентре «Наука» РАН с готовых файлов
Москва 121099, Шубинский пер., д. 6

СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТУРБУЛЕНТНЫХ ПОТОКОВ ТЕПЛА МЕЖДУ ОКЕАНОМ И АТМОСФЕРОЙ С ПОМОЩЬЮ МЕТОДА СКОЛЬЗЯЩЕГО РАЗДЕЛЕНИЯ КОНЕЧНЫХ НОРМАЛЬНЫХ СМЕСЕЙ*

В. Ю. Королев¹, А. К. Горшенин², С. К. Гулев³, К. П. Беляев⁴

Аннотация: Метод скользящего разделения смесей применяется к задаче статистического моделирования закономерностей в явных и скрытых турбулентных тепловых потоках. В качестве исходных данных используются результаты шестичасовых наблюдений в Атлантике (NCEP-NCAR, 1948–2008 гг.). В качестве математической модели используется аппроксимация конечными смесями нормальных распределений с параметрами, зависящими от времени. Применение методологии скользящего разделения смесей направлено на выявление закономерности в изменении этих параметров и определение изменчивости, которая может быть ассоциирована как с трендом, так и с нерегулярной изменчивостью. Предложен соответствующий подход к определению доли экстремальных наблюдений в исходной выборке.

Ключевые слова: конечные смеси нормальных распределений; метод скользящего разделения смесей; вероятностные модели; интеллектуальный анализ данных

DOI: 10.14357/19922264150401

1 Введение

Изучение явных и скрытых поверхностных турбулентных тепловых потоков между атмосферой и океаном, определяющих их взаимодействие, является чрезвычайно важным во многих областях наук о Земле. Данные о потоках могут быть получены из нескольких источников, каждый из которых обладает собственными достоинствами и недостатками. Наиболее подробные временные ряды (с периодами наблюдений в 100 и более лет) доступны с помощью программы наблюдений Voluntary Observing Ship (VOS) [1], в то время как данные за последние несколько десятилетий собираются с помощью спутников, повторного анализа и комбинированных продуктов (например, OA-Flux [2, 3]) с высоким пространственным и временным разрешением.

Информация об изменчивости поверхностных турбулентных тепловых потоков в большинстве случаев ограничивается первым (в некоторых

случаях еще и вторым) моментом вероятностного распределения потоков. Они традиционно вычисляются по временному ряду, соответствующему потоку, и составляют основу для климатологических исследований и проверок [4, 5]. Тем не менее детальная оценка характеристик теплового потока, в том числе определение экстремальных значений, требует точного знания вероятностного распределения, а также анализа изменений его параметров во времени и пространстве. Отсутствие подобных знаний при построении океанологических и климатических моделей серьезно снижает качество основного на них прогнозирования.

Еще одна важная причина для изучения вероятностных распределений турбулентных потоков тепла — необходимость количественного оценивания и минимизации выборочных ошибок в продуктах, построенных по базам VOS [6, 7]. Большие по величине ошибки в выборках влияют как на оценки средних величин потоков, так и на характеристики экстремальных потоков.

* Работа выполнена при поддержке гранта Президента РФ МК-4103.2014.9 и РФФИ (проекты 15-07-02652 и 15-37-20851).

¹ Факультет 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.su

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский государственный университет информационных технологий, радиотехники и электроники, a.k.gorshenin@gmail.com

³ Институт океанологии им. П. П. Ширшова РАН; географический факультет Московского государственного университета им. М. В. Ломоносова; Кильский университет, Германия, gul@sail.msk.ru

⁴ Институт океанологии им. П. П. Ширшова РАН; Федеральный университет штата Баия, Сальвадор, Бразилия, kosbel55@gmail.com

Попытка выбрать подходящий тип вероятностного распределения для турбулентных потоков была осуществлена в работе [8], в которой показано, что в качестве достаточно хорошей аппроксимации для данных может быть использовано так называемое двухпараметрическое распределение Фишера–Типпета (FT-распределение). В указанной работе были оценены параметры сдвига и масштаба данного распределения, осуществлена проверка гипотез о качестве подбора модели, а также была предложена глобальная климатологическая интерпретация параметров FT-распределения. Кроме того, FT-распределение было использовано для анализа временных рядов значительного объема для поверхностных турбулентных потоков, реконструированных по наблюдениям из базы VOS с 1880 г. [9]. Тем не менее многие вопросы, связанные с вероятностным распределением поверхностных турбулентных потоков, все еще остаются открытыми. В частности, с помощью FT-распределения часто не удается корректно аппроксимировать экстремальные турбулентные потоки тепла, а также в полной мере учесть случай так называемых «тяжелых хвостов» в распределениях потоков.

В настоящей работе используется идея представления вероятностного распределения приращений теплового потока (разности первого порядка) в виде конечной смеси нескольких нормальных (гауссовских) распределений с параметрами, зависящими от времени. В частности, вес каждого из распределений в смеси также может изменяться во времени. Для поиска оценок неизвестных параметров модели используется метод скользящего разделения смесей. Данный подход позволяет выявить закономерности в изменении параметров и определить изменчивость, которая может быть ассоциирована с трендом, а также динамику, связанную с диффузией или нерегулярной изменчивостью.

2 Однородность данных

Статистический анализ стохастических закономерностей в наблюдаемых временных рядах традиционно подразумевает работу со всеми имеющимися данными без какой-либо предварительной обработки с целью получения однородных данных. Например, в работе [8] FT-распределение применялось для аппроксимации исходного временного ряда. Однако такой подход, вероятно, не может быть использован для анализа очень длинных временных рядов и эволюции параметров распределения во времени. При этом выборка, используемая для статистического анализа, не является однородной, так как отдельные ее элементы не являются независимыми. Чтобы пояснить это обстоятельство, рассмотрим следующий модельный пример.

Всюду в дальнейшем обозначения $\varphi(x)$ и $\Phi(x)$ будут использованы для плотности и функции распределения стандартного нормального закона, а именно:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}; \quad \Phi(x) = \int_{-\infty}^x \varphi(z) dz, \quad x \in \mathbb{R}.$$

Предположим, что n — натуральное число, $\xi_1, \xi_2, \dots, \xi_n$ — независимые одинаково распределенные случайные величины с общей функцией распределения $F(x) = \Phi(x - a)$ (т. е. каждая случайная величина ξ_j имеет нормальное распределение со средним a и единичной дисперсией). Определим новый набор случайных величин $\zeta_1, \zeta_2, \dots, \zeta_n$ следующим образом:

$$\zeta_k = \xi_1 + \dots + \xi_k, \quad k = 1, \dots, n.$$

Очевидно, что для каждого $k \in \{1, \dots, n\}$ элемент ζ_k имеет нормальное распределение со средним ka и дисперсией k . Таким образом, выборка

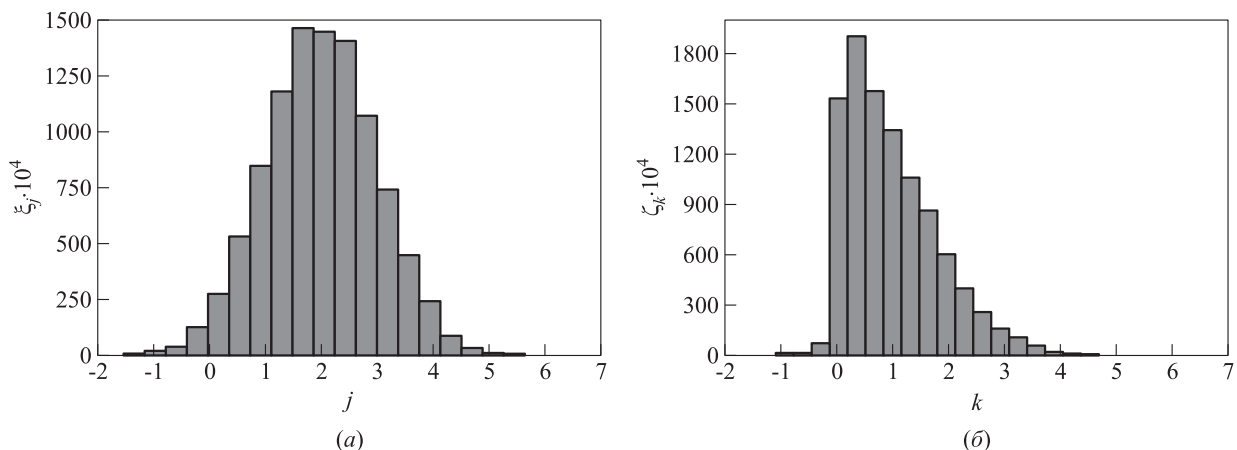


Рис. 1 Гистограммы, построенные по выборкам ξ_1, \dots, ξ_n (а) и ζ_1, \dots, ζ_n (б) с параметрами $n = 1000$ и $a = 2$

$\zeta_1, \zeta_2, \dots, \zeta_n$ не является однородной и независимой.

На рис. 1 данный эффект проиллюстрирован с помощью гистограмм, построенных по смоделированной выборке ξ_1, \dots, ξ_n с $n = 1000$ и $a = 2$ (см. рис. 1, а) и соответствующей выборке ζ_1, \dots, ζ_n (см. рис. 1, б). Гистограмма на рис. 1, б существенно скошена вправо с малым числом отрицательных значений. Данная картина в точности соответствует форме распределения, предложенного в [8].

Стохастический характер ζ_k в значительной степени определяется суммами $\xi_1 + \dots + \xi_{k-1}$ и слабо зависит от ξ_k . Чем больше величина k , тем меньший вклад случайной величины ξ_k в ζ_k . Таким образом, любой анализ статистических закономерностей $\xi_i, i = 1, \dots, n$, непосредственно по выборке $\zeta_1, \zeta_2, \dots, \zeta_n$ может выполняться только в рамках серьезных дополнительных допущений. Кроме того, с математической точки зрения стандартные статистические процедуры для выборки $\zeta_1, \zeta_2, \dots, \zeta_n$ не применимы. Практическая ценность результатов, основанных на таком анализе, вызывает существенные вопросы.

Чтобы избежать влияния упомянутых проблем, возникающих при использовании традиционных способов, следует проанализировать преобразованный временной ряд, рассмотрев ряд *приращений* турбулентных тепловых потоков.

3 Основы применения метода скользящего разделения смесей на основе конечных нормальных смесей для анализа временных рядов

Для выявления структурных изменений в наблюдаемых стохастических процессах с течением времени может быть успешно использован метод скользящего разделения смесей (СРС-метод), предложенный в книге [10]. В статьях [11–13] приводятся примеры эффективной работы этого метода при анализе финансовых рынков, трафика в информационных системах и процессов в турбулентной плазме. Ключевая особенность СРС-метода состоит в представлении изменчивости временных рядов динамической и диффузионной компонентами.

В рамках этого метода одномерные распределения приращений основного процесса аппроксимируются конечными смесями нормальных распределений. Теоретические основы для этих моделей основаны на утверждении, что конечные

нормальные смеси могут эффективно приблизить сдвиг-масштабные нормальные смеси общего вида или нормальные дисперсионно-сдвиговые смеси. Они появляются в качестве предельных законов для распределений сумм случайного числа независимых случайных величин или неоднородных и нестационарных случайных блужданий (подробнее см. [10, 14, 15]).

Для проведения анализа изменений в стохастическом процессе необходимо последовательно решать задачу статистического оценивания неизвестных параметров распределений для движущегося отрезка выборки (некоторого заранее заданного размера), представляющего собой подвыборку для дальнейшего анализа. Оценивая параметры для данного отрезка (также называемого окном), можно получить временной ряд этих параметров, который может быть использован для анализа эволюции диффузионной и динамической компонент процесса во времени.

Предположим, что функция распределения для части данных (окна) в момент времени t может быть представлена в виде

$$F_t(x) = \sum_{i=1}^k \frac{p_i(t)}{\sigma_i(t)\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{(t - a_i(t))^2}{2\sigma_i^2(t)}\right\} dt, \quad (1)$$

где

$$\sum_{i=1}^k p_i(t) = 1, \quad p_i(t) \geq 0 \quad (2)$$

(для любых $x \in \mathbb{R}, a_i(t) \in \mathbb{R}, \sigma_i(t) > 0, i = 1, \dots, k$). Модель (1) называется конечной смесью нормальных законов. Параметры $p_1(t), \dots, p_k(t)$, удовлетворяющие соотношению (2), называются весами, параметр k определяет число компонент в смеси.

Параметры модели (1) существенным образом зависят от времени. Данный факт проиллюстрирован на рис. 2, изображающем гистограммы, построенные на различных окнах (размер каждого окна составляет 200 элементов), и приближающие плотности вида (1), параметры которых оценены на соответствующих окнах.

Параметр k также может зависеть от времени. Тем не менее с точки зрения выбора эффективных настроек метода, и с точки зрения интерпретации результатов желательно задать предельно допустимое значение k заранее. Как правило, типичное число компонент не превышает 6 или 7 — именно такое количество обеспечивает качественное приближение любой модели. При оценивании параметров модели (1) в режиме скользящего окна не-

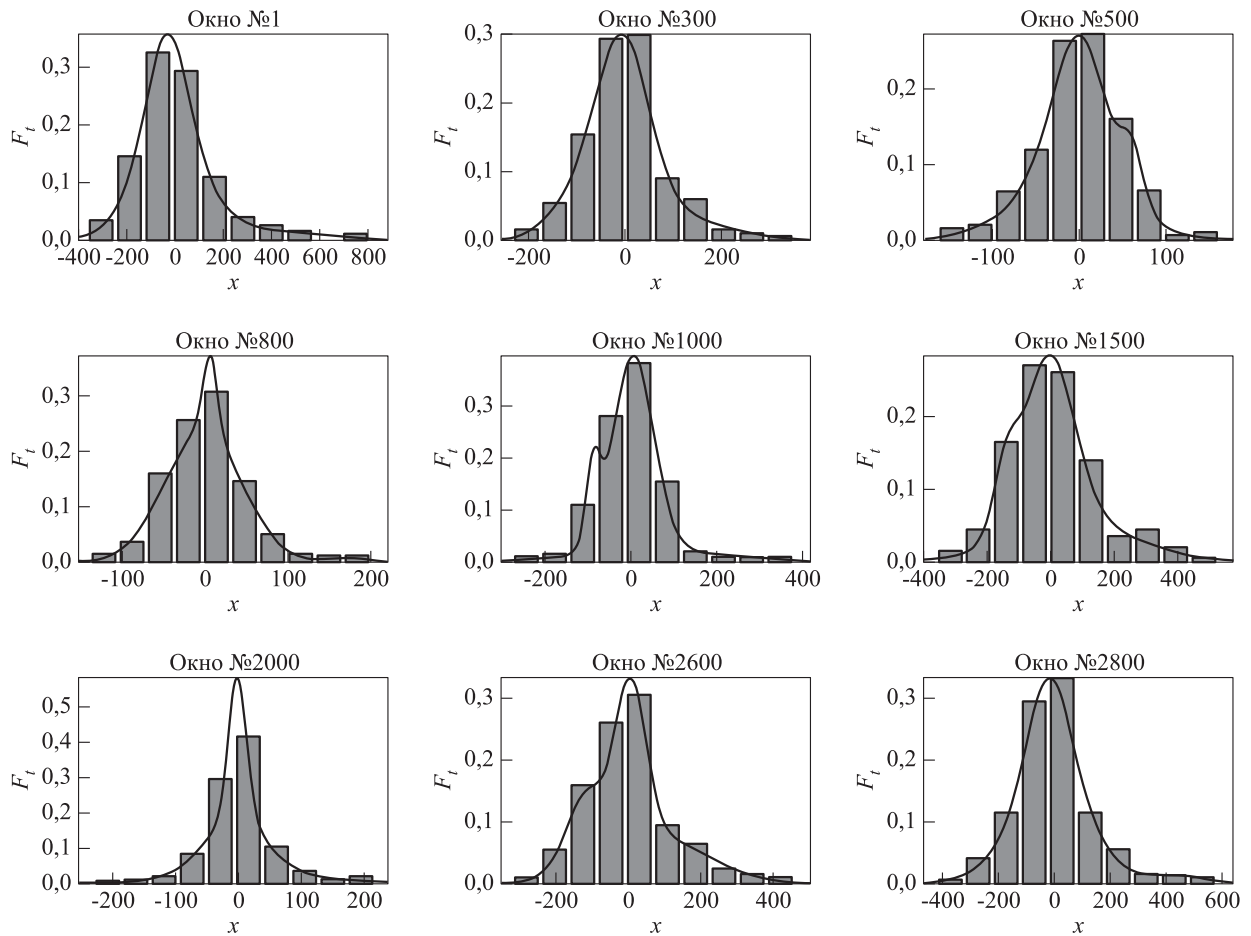


Рис. 2 Гистограммы, построенные для различных положений окна, с подгонкой плотностями типа конечных смесей нормальных законов

которые веса могут быть достаточно близки к нулевому значению. Это влечет необходимость отбрасывания из рассмотрения соответствующей компоненты и уменьшение общего числа компонент в смеси.

Параметры $a_1(t), \dots, a_k(t)$ ассоциированы с динамической компонентой внутренней изменчивости процесса, параметры $\sigma_1(t), \dots, \sigma_k(t)$ связаны с диффузионной компонентой [10]. Если переменная Z_t представляет собой случайную величину с функцией распределения, определяемой выражением (1), то ее дисперсия может быть представлена в виде суммы двух компонент:

$$DZ_t = \sum_{i=1}^k p_i(t) [a_i(t) - \bar{a}(t)]^2 + \sum_{i=1}^k p_i(t) \sigma_i^2(t), \quad (3)$$

где

$$\bar{a}(t) = \sum_{i=1}^k p_i(t) a_i(t).$$

Первый член в правой части выражения (3) зависит только от весов $p_i(t)$ и математических ожи-

даний $a_i(t)$ компонент смеси (1). Так как Z_t представляет собой приращения исходного процесса, то $a_i(t)$ является математическим ожиданием значения приращения, т.е. «трендовой» компонентой.

Таким образом, первое слагаемое определяет часть общей дисперсии (изменчивости), обусловленной локальными элементарными трендами. Его обычно называют *динамической* компонентой дисперсии. Кроме того, $a_i(t)$ является математическим ожиданием случайной величины, распределение которой соответствует i -й компоненте смеси (1). По построению эта случайная величина является приращением исходного процесса на *единичном* временном интервале, т.е. $a_i(t)$ представляет собой среднюю скорость изменения i -й компоненты. Таким образом, множество пар $(a_1(t), p_1(t)), \dots, (a_k(t), p_k(t))$ определяет распределение скоростей изменения локальных трендов в момент времени t .

Второе слагаемое в правой части выражения (3) зависит только от весов $p_i(t)$ и дисперсий $\sigma_i^2(t)$

компонентов и представляет собой чисто стохастическую *диффузионную* составляющую общей дисперсии.

4 Оценивание неизвестных параметров распределений

Для оценивания параметров модели (1) на каждом окне можно воспользоваться классическим EM (expectation-maximization) алгоритмом — итерационным численным методом для максимизации функции правдоподобия нескольких переменных. Данный метод был предложен в [16] и подробно описан в [10]. Хотя существует значительное число всевозможных модификаций этого алгоритма, классический EM-алгоритм остается весьма надежным инструментом для поиска оценок параметров смеси конечных нормальных законов (1).

При этом данный алгоритм не лишен и ряда существенных недостатков. Вероятно, основной из них — его неустойчивость по отношению к выбору начального приближения. Функция правдоподобия для модели конечной нормальной смеси имеет существенно негладкую поверхность с пиками, впадинами или разрывами. Поэтому, будучи «жадным» алгоритмом, EM-алгоритм сходится к тому локальному максимуму, который является ближайшим к начальному приближению. Стандартные способы преодоления данной особенности представлены ниже:

- начальное приближение следует выбирать случайным образом;
- нужно выбрать несколько начальных приближений, а оценку на окне определить как усреднение результатов нескольких независимых запусков EM-алгоритма;
- следует выбрать несколько начальных приближений, а затем в качестве оценки на окне выбрать набор параметров, максимизирующих значение функции правдоподобия.

При использовании EM-алгоритма в скользящем режиме (как это сделано в данной работе) значительное внимание должно быть уделено визуализации результатов. Для удобной интерпретации результатов полученные кривые, изображающие эволюцию во времени параметров $p_i(t)$, $a_i(t)$, $\sigma_i(t)$, $i = 1, \dots, k$, должны быть гладкими. В частности, это может быть достигнуто в случае использования оценок EM-алгоритма на текущем окне в качестве значений начальных приближений для следующего окна. Однако такой способ увеличивает риск выбора локального экстремума вместо глобального.

Результаты, полученные для трех вышеуказанных способов выбора начального приближения, представлены на рис. 3.

На рис. 3, *a* (верхний график) представлена эволюция во времени параметров $a_i(t)$ для локальных трендов, оцененных с помощью EM-алгоритма со случайным выбором начальных приближений. На каждом окне расчеты с помощью EM-алгоритма проводятся пять раз, при этом начальные значения выбираются случайным образом для каждого запуска. Результаты усредняются по всему набору значений, полученному при работе алгоритмов на данном окне. Веса компонентов изображены в соответствии с цветовой шкалой, изображенной справа вне графиков. Рисунок 3, *a* (нижний график) демонстрирует изменение во времени локальных параметров диффузии $\sigma_i^2(t)$, оцениваемых с помощью описанной версии EM-алгоритма.

На рис. 3, *b* (верхний график) продемонстрирована эволюция во времени параметров $a_i(t)$ для локальных трендов, оцениваемых с помощью «классического» EM-алгоритма со случайным выбором начальных значений для весов. На каждом окне начальные приближения для параметров сдвига и масштаба выбираются единым образом (как среднее и выборочная дисперсия для текущего окна соответственно). На рис. 3, *b* (нижний график) приведены изменения во времени параметров локальных диффузий $\sigma_i^2(t)$, оцениваемых с помощью «классического» EM-алгоритма.

На рис. 3, *в* (верхний график) приведены изменения во времени параметров $a_i(t)$ для локальных трендов, оцененных с помощью EM-алгоритма с выбором случайного начального приближения на каждом окне. Для каждого положения окна EM-алгоритм запускается пять раз, начальные значения выбираются каждый раз случайным образом. При этом в качестве оценки выбирается набор параметров, соответствующий функции правдоподобия с наибольшим значением среди всех запусков. Рисунок 3, *в* (нижний график) демонстрирует изменение во времени локальных параметров диффузии $\sigma_i^2(t)$, оцениваемых с помощью описанной версии EM-алгоритма.

Можно сделать вывод, что третья версия EM-алгоритма со случайным выбором начальных приближений и поиском набора, соответствующего максимальному значению функции правдоподобия на каждом окне, дает наиболее четкие результаты.

Данный алгоритм был использован для анализа временной изменчивости параметров распределения приращений тепловых потоков. Полученные квантили распределения представлены на рис. 4 изотопными линиями (горизонтальная ось времени соответствует периоду длительностью око-

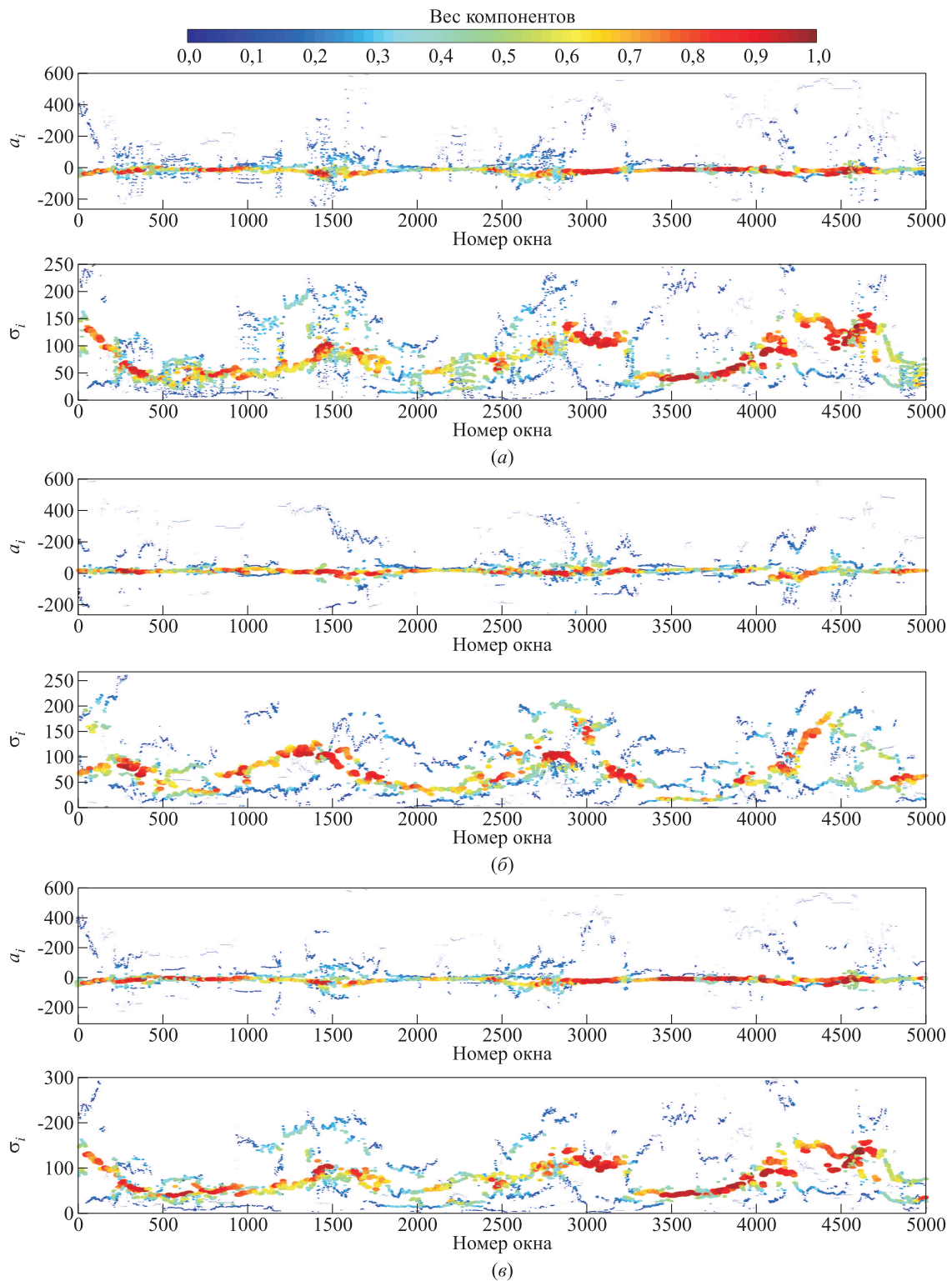


Рис. 3 Графики, изображающие оценки параметров, полученные в трех режимах: (а) EM-алгоритм со случайным выбором начальных приближений; (б) «классический» EM-алгоритм со случайным выбором начальных значений для весов; (в) EM-алгоритм с выбором случайного начального приближения на каждом окне

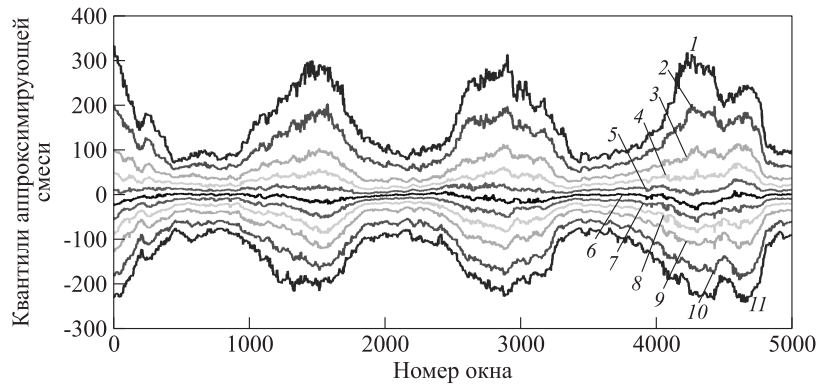


Рис. 4 Квантили вероятностного распределения приращений процесса потоков тепла: 1 — 0,05; 2 — 0,1; 3 — 0,2; 4 — 0,3; 5 — 0,4; 6 — 0,5; 7 — 0,6; 8 — 0,7; 9 — 0,8; 10 — 0,9; 11 — 0,95

ло 3,5 лет). На графиках явно выделяется сезонная периодичность.

На рис. 5 представлена эволюция моментных характеристик распределения вероятностей прира-

щений процесса тепловых потоков. Видно, что математическое ожидание заметно колеблется во времени с изменяющейся амплитудой. Кроме того, для каждого периода амплитуда меньше для перио-

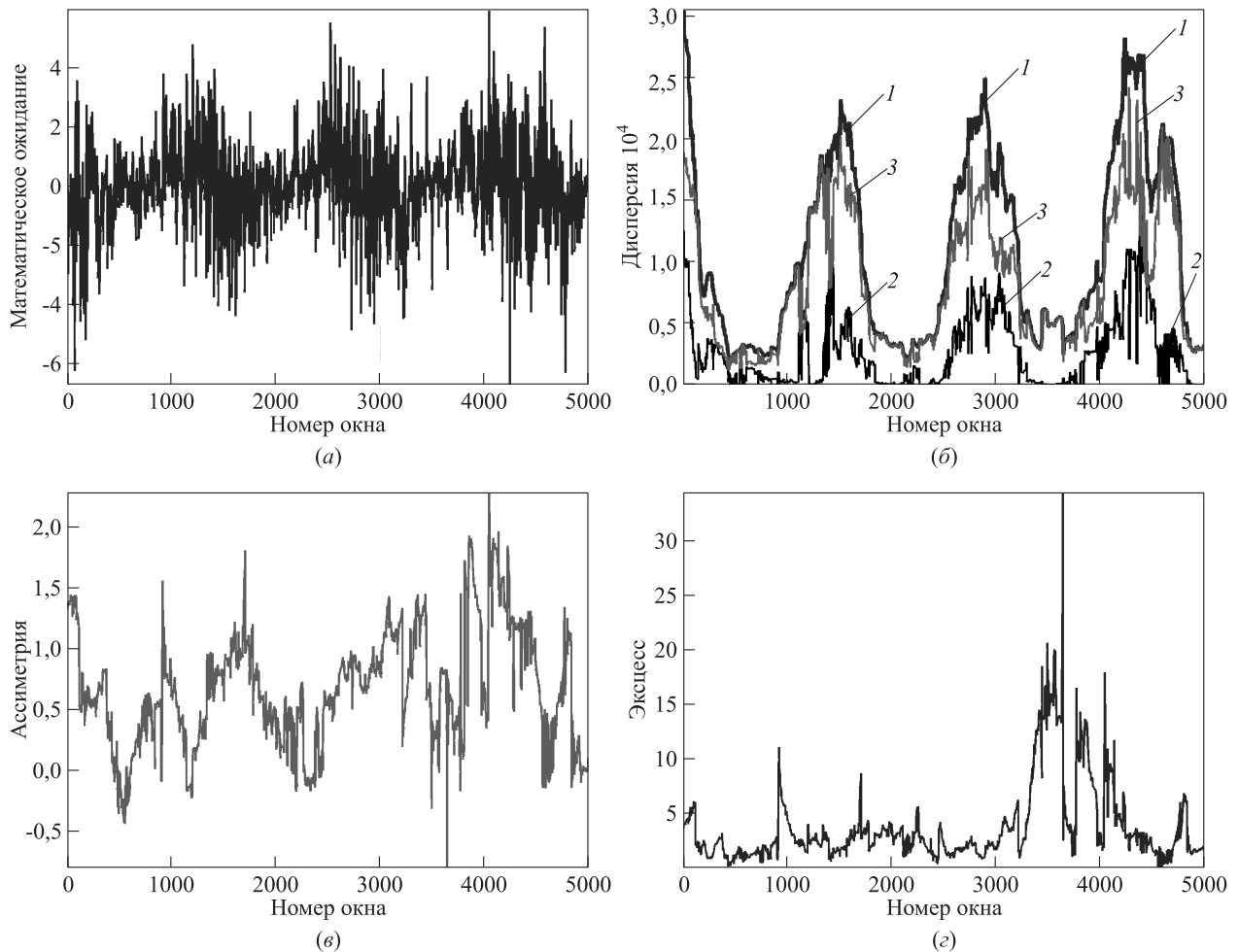


Рис. 5 Моментные характеристики вероятностного распределения приращений тепловых потоков: (а) математическое ожидание; (б) дисперсия (1 — дисперсия аппроксимирующей смеси нормальных распределений; 2 — динамическая компонента дисперсии; 3 — диффузионная компонента дисперсии); (в) асимметрия; (г) эксцесс

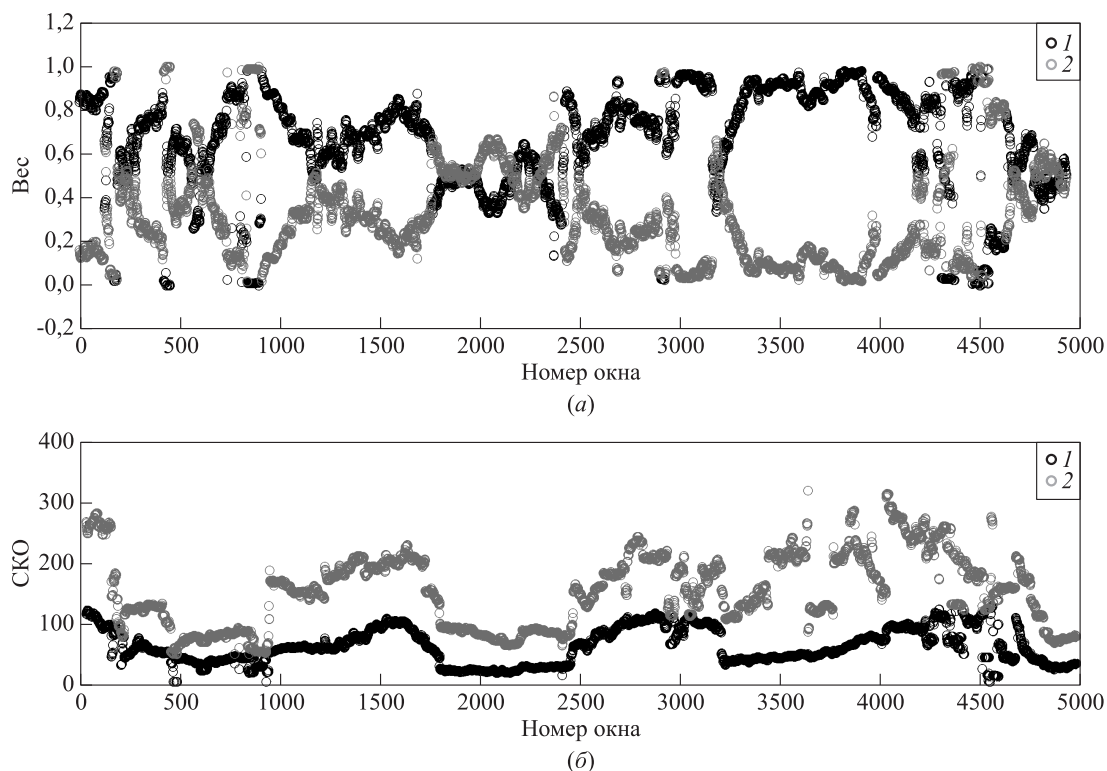


Рис. 6 Определение доли экстремальных наблюдений: (а) вес; (б) СКО ($\sigma_1 \leq \sigma_2$)

да сезонного увеличения общего среднего по сравнению с амплитудой колебаний для периода сезонного снижения общего среднего значения. Хорошо прослеживается сезонный характер изменения дисперсии. Достаточно интересным представляется и тот факт, что вклад в общую дисперсию чисто стохастической диффузионной компоненты дисперсии (см. второе слагаемое в правой части соотношения (3)) — зеленая кривая на правом верхнем графике — больше, чем динамической составляющей (см. первый член в правой части соотношения (3)), изображенной синей кривой. Можно отметить, что правый хвост распределения приращений тяжелее левого. Еще одно интересное наблюдение заключается в том, что эксцесс этого распределения максимален во время периода «спокойствия».

5 Определение доли экстремальных наблюдений в данных

На основании рассмотренной выше методологии в настоящем разделе обсудим возможный способ решения задачи определения доли экстремальных наблюдений в исходной выборке, не

требующий привлечения теории экстремальных значений.

Предположим, что в модели (1) параметр k выбирается равным двум (с одной стороны, для ускорения вычислений, с другой — для получения более контрастной общей картины). На рис. 6 представлены два графика для рассмотренного ранее ряда. На верхнем изображена эволюция весов каждой из компонент, а на нижнем — соответствующие им среднеквадратические отклонения (СКО). При этом справедливо соотношение $\sigma_1 \leq \sigma_2$ (отметим, что один параметр превосходит другой минимум в 2 раза для каждого положения скользящего окна), а цвета значков на обоих графиках соответствуют друг другу: серым обозначены графики для компоненты с наибольшей дисперсией на каждом окне, черным — с наименьшей.

Наблюдения, соответствующие компоненте с наибольшей дисперсией, могут быть проинтерпретированы как экстремальные (относительно наблюдений другой компоненты). Стоит отметить, что веса этой компоненты лежат в диапазоне 0,2–0,4, т. е. можно сказать, что указанные наблюдения составляют примерно треть от общего числа. Отметим, что указанная картина сохраняется для всех сезонов, при этом характер экстремальных наблюдений меняется.

6 Выводы

В данной работе метод скользящего разделения смесей был использован для анализа статистических закономерностей во временной эволюции тепловых потоков. Вычисления проводились с помощью нескольких модификаций EM-алгоритма, из которых в качестве наиболее подходящей была выбрана специальная версия, основанная на максимизации функции правдоподобия в классе конечных смесей нормальных законов. Было показано, что в стохастическом поведении эволюции тепловых потоков основная компонента с небольшой дисперсией может сопровождаться стохастически развивающимися и исчезающими компонентами с большой дисперсией. Также отмечен ряд закономерностей во временной изменчивости моментных характеристик приращений значений процесса тепловых потоков. На основании упорядочивания весов и дисперсий предложен метод определения доли экстремальных наблюдений в рассматриваемых временных рядах.

Литература

1. *Berry D. I., Kent E. C.* A new air–sea interaction gridded dataset from ICOADS with uncertainty estimates // *Bull. Am. Meteorol. Soc.*, 2009. Vol. 90. No. 5. P. 645–656.
2. *Yu L.* Global variations in oceanic evaporation (1958–2005): The role of the changing wind speed // *J. Climate*, 2007. Vol. 20. P. 5376–5390.
3. *Yu L., Weller R. A.* Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005) // *Bull. Am. Meteorol. Soc.*, 2007. Vol. 88. P. 527–539.
4. *Josey S. A.* A comparison of ECMWF, NCEP-NCAR and SOC surface heat fluxes with moored buoy measurements in the subduction region of the Northeast Atlantic // *J. Climate*, 2001. Vol. 14. P. 1780–1789.
5. *Bouras D.* Comparison of five satellite-derived latent heat flux products to moored buoy data // *J. Climate*, 2006. Vol. 19. P. 6291–6313.
6. *Gulev S. K., Jung T., Ruprecht E.* Estimation of the impact of sampling errors in the VOS observations on air–sea fluxes. Part I. Uncertainties in climate means // *J. Climate*, 2007. Vol. 20. P. 279–301.
7. *Gulev S. K., Jung T., Ruprecht E.* Estimation of the impact of sampling errors in the VOS observations on air–sea fluxes. Part II. Impact on trends and interannual variability // *J. Climate*, 2007. Vol. 20. P. 302–315.
8. *Gulev S. K., Belyaev K. P.* Probability distribution characteristics for surface air–sea turbulent heat fluxes over the global ocean // *J. Climate*, 2012. Vol. 25. No. 1. P. 184–206.
9. *Gulev S. K., Latif M., Keenlyside N., Park W., Koltermann K. P.* North Atlantic Ocean control on surface heat flux on multidecadal timescales // *Nature*, 2013. Vol. 499. P. 464–467.
10. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Моск. ун-та, 2011. 512 с.
11. *Горшенин А. К., Королев В. Ю., Турсунбаев А. М.* Медицинские модификации EM- и SEM-алгоритмов для разделения смесей вероятностных распределений и их применение к декомпозиции волатильности финансовых временных рядов // *Информатика и её применения*, 2008. Т. 2. Вып. 4. С. 12–47.
12. *Gorshenin A., Korolev V.* Modeling of statistical fluctuations of information flows by mixtures of gamma distributions // *27th European Conference on Modelling and Simulation Proceedings*. — Dudweiler, Germany: Digitaldruck Pirrot GmbH, 2013. P. 569–572.
13. *Горшенин А. К.* Информационная технология исследования тонкой структуры хаотических процессов в плазме с помощью анализа спектров // *Системы и средства информатики*, 2014. Т. 24. Вып. 1. С. 116–127.
14. *Королев В. Ю.* Обобщенные гиперболические распределения как предельные для случайных сумм // *Теория вероятностей и её применения*, 2013. Т. 58. Вып. 1. С. 117–132.
15. *Королев В. Ю., Закс Л. М.* Обобщенные дисперсионные гамма-распределения как предельные для случайных сумм // *Информатика и её применения*, 2013. Т. 7. Вып. 1. С. 105–115.
16. *Dempster A., Laird N., Rubin D.* Maximum likelihood estimation from incompleting data // *J. Roy. Stat. Soc. B*, 1977. Vol. 39. No. 1. P. 1–38.

Поступила в редакцию 04.09.15

STATISTICAL MODELING OF AIR–SEA TURBULENT HEAT FLUXES BY THE METHOD OF MOVING SEPARATION OF FINITE NORMAL MIXTURES

V. Yu. Korolev^{1,2}, A. K. Gorshenin^{2,3}, S. K. Gulev^{4,5,6}, and K. P. Belyaev^{5,7}

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 Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³Moscow State University of Information Technologies, Radioengineering, and Electronics, 78 Vernadskogo Ave., Moscow 119454, Russian Federation

⁴Faculty of Geography, M. V. Lomonosov Moscow State University, 1 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

⁵P. P. Shirshov Institute of Oceanology, 36 Nakhimovski Prosp., Moscow 117997, Russian Federation

⁶University of Kiel, Christian-Albrechts-Universität zu Kiel, 4 Christian-Albrechts-Platz, Kiel 24098, Germany

⁷Federal University of Bahia, Rua Adhemar de Barros, no 500, Ondina, 40.710-110, Salvador, Bahia, Brazil

Abstract: The method of moving separation of mixtures is applied to the problem of statistical modeling of regularities in explicit and latent turbulent heat fluxes. The six-hour observations in the Atlantic region (NCEP-NCAR, 1948–2008) are used as initial data. The basic approximate mathematical model is a finite normal mixture with parameters depending on time. The methodology of moving separation of mixtures allows one to analyze the regularities in the variation of parameters and to capture the variability which can be associated with the trend as well as the irregular variation. An approach is proposed to the determination of the proportion of extreme observations in the original sample.

Keywords: finite normal mixtures; moving separation of mixtures; probabilistic models; data mining

DOI: 10.14357/19922264150401

Acknowledgments

The work was supported by the President of the Russian Federation (grant MK-4103.2014.9) and the Russian Foundation for Basic Research (projects 15-07-02652 and 15-37-20851).

References

1. Berry, D. I., and E. C. Kent. 2009. A new air–sea interaction gridded dataset from ICOADS with uncertainty estimates. *Bull. Am. Meteorol. Soc.* 90(5):645–656.
2. Yu, L. 2007. Global variations in oceanic evaporation (1958–2005): The role of the changing wind speed. *J. Climate* 20:5376–5390.
3. Yu, L., and R. A. Weller. 2007. Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). *Bull. Am. Meteorol. Soc.* 88:527–539.
4. Josey, S. A. 2001. A comparison of ECMWF, NCEP-NCAR and SOC surface heat fluxes with moored buoy measurements in the subduction region of the Northeast Atlantic. *J. Climate* 14:1780–1789.
5. Bouras, D. 2006. Comparison of five satellite-derived latent heat flux products to moored buoy data. *J. Climate* 19:6291–6313.
6. Gulev, S. K., T. Jung, and E. Ruprecht. 2007. Estimation of the impact of sampling errors in the VOS observations on air–sea fluxes. Part I. Uncertainties in climate means. *J. Climate* 20:279–301.
7. Gulev, S. K., T. Jung, and E. Ruprecht. 2007. Estimation of the impact of sampling errors in the VOS observations on air–sea fluxes. Part II. Impact on trends and interannual variability. *J. Climate* 20:302–315.
8. Gulev, S. K., and K. P. Belyaev. 2012. Probability distribution characteristics for surface air–sea turbulent heat fluxes over the global ocean. *J. Climate* 25(1):184–206.
9. Gulev, S. K., M. Latif, N. Keenlyside, W. Park, and K. P. Koltermann. 2013. North Atlantic Ocean control on surface heat flux on multidecadal timescales. *Nature* 499:464–467.
10. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh protsessov* [Probabilistic and statistical methods of decomposition of volatility of chaotic processes]. Moscow: Moscow University Publ. 512 p.
11. Gorshenin, A. K., V. Yu. Korolev, D. V. Malakhov and A. M. Tursunbaev. 2008. Mediannye modifikatsii EM-i SEM-algoritmov dlya razdeleniya smesey veroyatno-

- snykh raspredeleniy i ikh primenenie k dekompozitsii volatil'nosti finansovykh vremennykh ryadov [Median modification of EM- and SEM-algorithms for separation of mixtures of probability distributions and their application to the decomposition of volatility of financial time series]. *Informatika i ee Primeneniya — Inform. Appl.* 2(4):12–47.
12. Gorshenin, A. K., and V. Yu. Korolev. 2013. Modeling of statistical fluctuations of information flows by mixtures of gamma distributions. *27th European Conference on Modelling and Simulation Proceedings*. Dudweiler, Germany: Digitaldruck Pirrot GmbH. 569–572.
 13. Gorshenin, A. K. 2014. Informatsionnaya tekhnologiya issledovaniya tonkoy struktury khaoticheskikh protsessov v plazme s pomoshch'yu analiza spektrov [The information technology to research the fine structure of chaotic processes in plasma by the analysis of the spectra]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(1):116–127.
 14. Korolev, V. Yu. 2013. Obobshchennye giperbolicheskie raspredeleniya kak predel'nye dlya sluchaynykh summ [Generalized hyperbolic laws as limit distributions for random sums]. *Teoriya Veroyatnostey i ee Primeneniya* [Theory of Probability and Its Applications] 58(1):117–132.
 15. Korolev, V. Yu., and L. M. Zaks. 2013. Obobshchennye dispersionnyye gamma-raspredeleniya kak predel'nye dlya sluchaynykh summ [Variance-generalized-gamma-distributions as limit laws for random sums]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):105–115.
 16. Dempster, A., N. Laird, and D. Rubin. 1977. Maximum likelihood estimation from incompleated data. *J. Roy. Stat. Soc. B* 39(1):1–38.

Received September 04, 2015

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; vkorolev@cs.msu.ru

Gorshenin Andrey K. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; associate professor, Moscow State University of Information Technologies, Radioengineering, and Electronics, 78 Vernadskogo Ave., Moscow 119454, Russian Federation; a.k.gorshenin@gmail.com

Gulev Sergey K. (b. 1958) — Corresponding Member of the Russian Academy of Sciences, Doctor of Science in physics and mathematics, Head of Laboratory, P.P. Shirshov Institute of Oceanology, 36 Nakhimovski Prosp., Moscow 117997, Russian Federation; professor, Department of Oceanology, Faculty of Geography, M. V. Lomonosov Moscow State University, 1 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; professor, University of Kiel, Christian-Albrechts-Universität zu Kiel, 4 Christian-Albrechts-Platz, Kiel 24098, Germany; gul@sail.msk.ru

Belyaev Konstantin P. (b. 1955) — Doctor of Science in physics and mathematics, leading scientist, P.P. Shirshov Institute of Oceanology, 36 Nakhimovski Prosp., Moscow 117997, Russian Federation; visiting professor, Federal University of Bahia, Rua Adhemar de Barros, no 500, Ondina, 40.710-110, Salvador, Bahia, Brazil; kosbel55@gmail.com

ОБОБЩЕННЫЕ ДИСПЕРСИОННЫЕ ГАММА-РАСПРЕДЕЛЕНИЯ КАК МОДЕЛИ СТАТИСТИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ НА ФИНАНСОВЫХ РЫНКАХ*

В. Ю. Королев¹, А. Ю. Корчагин², И. А. Соколов³

Аннотация: Обсуждаются различные вопросы, связанные с применением обобщенных дисперсионных гамма-распределений для моделирования статистических закономерностей на финансовых рынках. Описаны простейшие свойства обобщенных дисперсионных гамма-распределений как специальных дисперсионно-сдвиговых смесей нормальных законов, в которых смешивающими являются обобщенные гамма-распределения. Приведены предельные теоремы для сумм случайного числа независимых случайных величин — аналоги закона больших чисел и центральной предельной теоремы, — обосновывающие возможность использования обобщенных дисперсионных гамма-распределений в качестве асимптотических аппроксимаций. Приведены результаты практической подгонки обобщенных дисперсионных гамма-распределений к реальным данным о поведении финансовых индексов и обобщенных гамма-распределений к наблюдаемым интенсивностям информационных потоков в современных финансовых информационных системах. Результаты сравнения обобщенных дисперсионных гамма-моделей с обобщенными гиперболическими моделями свидетельствуют о преимуществе первых над вторыми. Также обсуждаются методы оценивания параметров обобщенных дисперсионных гамма-моделей и их применение при прогнозировании процессов, протекающих на финансовых рынках.

Ключевые слова: обобщенные дисперсионные гамма-распределения; дисперсионно-сдвиговые смеси нормальных законов; обобщенные гамма-распределения; суммы случайного числа случайных величин; закон больших чисел; центральная предельная теорема

DOI: 10.14357/19922264150402

1 Введение

Обобщенные дисперсионные гамма-распределения (GVG-distributions) являются специальными дисперсионно-сдвиговыми смесями нормальных законов. Понятие дисперсионно-сдвиговой смеси нормальных законов (normal variance-mean mixture) введено в 1970–1980 гг. в работах О.-Е. Барндорфф-Нильсена и его коллег [1–3] как довольно гибкое обобщение нормального распределения.

Пусть $\beta \in \mathbb{R}$, $\alpha \in \mathbb{R}$, $0 < \sigma < \infty$, $\Phi(x)$ — стандартная нормальная функция распределения, $G(x)$ — функция распределения, все точки роста которой сосредоточены на \mathbb{R}_+ . Дисперсионно-сдвиговыми смесями нормальных законов называется функция распределения

$$F(x) = \int_0^{\infty} \Phi\left(\frac{x - \beta - \alpha z}{\sigma\sqrt{z}}\right) dG(z), \quad x \in \mathbb{R}. \quad (1)$$

Обратим внимание, что в соотношении (1) смешивание происходит одновременно и по параметру сдвига, и по параметру масштаба, но так как эти параметры в (1) связаны жесткой зависимостью, при которой параметры положения (*сдвига*) смешиваемых нормальных законов пропорциональны их *дисперсиям*, то фактически смесь (1) является однопараметрической. Именно поэтому смеси вида (1) называются *дисперсионно-сдвиговыми*.

Вероятностные модели типа дисперсионно-сдвиговых смесей нормальных законов рассматриваются в качестве базовых во многих практических задачах, связанных с изучением статистических закономерностей поведения информационных потоков. Наиболее известными представителями таких моделей являются *обобщенные гиперболические распределения*, введенные в упомянутых работах О.-Е. Барндорфф-Нильсена и его коллег. Обобщенные гиперболические модели имеют вид (1), в кото-

* Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 14-07-00041а.)

¹ Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vkorolev@cs.msu.su

² Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова, sasha.korchagin@gmail.com

³ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ISokolov@ipran.ru

ром $G(z) = P_{\text{GIG}}(z; \nu, \mu, \lambda)$ — функция обобщенного обратного гауссовского распределения, соответствующая плотности

$$p_{\text{GIG}}(x; \nu, \mu, \lambda) = \frac{\lambda^{\nu/2}}{2\mu^{\nu/2} K_{\nu}(\sqrt{\mu\lambda})} \cdot x^{\nu-1} \times \exp\left\{-\frac{1}{2}\left(\frac{\mu}{x} + \lambda x\right)\right\}, \quad x > 0.$$

Здесь $K_{\nu}(z)$ — модифицированная бесселева функция третьего рода порядка ν ,

$$K_{\nu}(z) = \frac{1}{2} \int_0^{\infty} y^{\nu-1} \exp\left\{-\frac{z}{2}\left(y + \frac{1}{y}\right)\right\} dy, \quad z \in \mathbb{C}, \operatorname{Re} z > 0.$$

При этом $\nu \in \mathbb{R}$, и если $\nu < 0$, то $\mu > 0, \lambda \geq 0$; если $\nu = 0$, то $\mu > 0, \lambda > 0$; если $\nu > 0$, то $\mu \geq 0, \lambda > 0$. Обобщенные гиперболические модели демонстрируют очень высокую адекватность при анализе финансовых данных.

В работе [4] было введено еще одно семейство дисперсионно-сдвиговых смесей нормальных законов вида (1) — семейство *обобщенных дисперсионных гамма-распределений*, в котором $G(z) = P_{\text{GG}}(z; \nu, \kappa, \delta)$ — функция обобщенного гамма-распределения, соответствующая плотности

$$p_{\text{GG}}(x; \nu, \kappa, \delta) = \frac{|\nu|}{\delta^{|\nu|} \Gamma(\kappa)} x^{\kappa\nu-1} \exp\left\{-\frac{x^{\nu}}{\delta^{|\nu|}}\right\}, \quad x \geq 0, \quad (2)$$

с параметрами $\nu \in \mathbb{R}, \kappa, \delta \in \mathbb{R}^+$, отвечающими соответственно за *степень, форму и масштаб*. Функцию обобщенного гамма-дисперсионного распределения будем обозначать $P_{\text{GVG}}(x; \alpha, \sigma, \nu, \kappa, \delta)$,

$$P_{\text{GVG}}(x; \alpha, \sigma, \nu, \kappa, \delta) = \int_0^{\infty} \Phi\left(\frac{x - \beta - \alpha z}{\sigma\sqrt{z}}\right) p_{\text{GG}}(x; \nu, \kappa, \delta) dz, \quad x \in \mathbb{R}. \quad (3)$$

Класс обобщенных гамма-распределений (GG-распределений) был впервые описан как единое семейство в 1962 г. в работе [5] в качестве семейства вероятностных моделей, включающего в себя одновременно гамма-распределения и распределения Вейбулла. Семейство GG-распределений содержит практически все наиболее популярные абсолютно непрерывные распределения. В частности, семейство GG-распределений содержит гамма-распределение ($\nu = 1$) и его частные случаи: показательное (экспоненциальное) распределение ($\nu = 1, \kappa = 1$),

распределение Эрланга ($\nu = 1, \kappa \in \mathbb{N}$), распределение хи-квадрат ($\nu = 1, \delta = 2$); распределение Накагами ($\nu = 2$); полунормальное распределение, иначе называемое сложенным (folded) нормальным (распределение максимума винеровского процесса на отрезке $[0, 1]$) ($\nu = 2, \kappa = 1/2$); распределение Рэлея ($\nu = 2, \kappa = 1$); хи-распределение ($\nu = 2, \delta = \sqrt{2}$); распределение Максвелла (распределение модулей скоростей движения молекул в разреженной газе) ($\nu = 2, \kappa = 3/2$); распределение Вейбулла–Гнеденко ($\kappa = 1, \nu > 0$); обратное гамма-распределение ($\nu = -1$); распределение Леви (одностороннее устойчивое распределение с характеристическим показателем $\alpha = 1/2$ — распределение времени достижения процессом броуновского движения единичного уровня) ($\nu = -1, \kappa = 1/2$); логнормальное распределение ($\kappa \rightarrow \infty$) и другие законы.

Обоженные гамма-распределения широко применяются во многих прикладных задачах, прежде всего связанных с обработкой сигналов и изображений. Работы, посвященные применению GG-распределений в качестве моделей реально наблюдаемых закономерностей, исчисляются сотнями. Большая популярность GG-распределений, по-видимому, обусловлена возможностью их использования в качестве адекватных асимптотических аппроксимаций, поскольку все перечисленные выше отдельные представители GG-распределений выступают в качестве предельных в различных предельных теоремах теории вероятностей. Эти свойства GG-распределений обосновывают, в частности, целесообразность моделирования с их помощью распределения случайной интенсивности потока запросов в информационных системах. Ниже будет сформулирована предельная теорема для случайных сумм независимых случайных величин, в которой предельными законами оказываются GG-распределения в общем их виде.

Необходимость рассмотрения GG-распределений в качестве смешивающих в дисперсионно-сдвиговых смесях нормальных законов обусловлена тем, что, в отличие от GIG-распределений, класс GG-распределений содержит законы, хвосты которых убывают вейбулловским образом. В частности, при $0 < \nu < 1$ GG-распределения занимают промежуточное место между распределениями с экспоненциальным убыванием хвостов (показательное распределение, гамма-распределение) и «тяжелохвостыми» распределениями со степенным убыванием хвостов типа Ципфа–Парето. В частности, возможно, именно поэтому распределения Вейбулла–Гнеденко с $\nu < 1$ в некоторых работах (см., например, [6–8]) называются *растянутыми (stretched) показательными распределениями*. Вместе с тем,

как показано в работе [9], асимптотическое поведение хвостов смесей нормальных законов в определенном смысле совпадает с аналогичным поведением хвостов смешивающих законов. Поэтому, в отличие от обобщенных гиперболических законов, класс обобщенных дисперсионных гамма-распределений содержит распределения с хвостами, убывающими экспоненциально-степенным образом. Такие особенности эмпирических распределений характерны для финансовых индексов с нерегулярным (перемежающимся) поведением, когда на фоне «обычных» стохастических флуктуаций возможны (кратковременные) периоды повышенной активности («волатильности») или отдельные выбросы.

Статья построена следующим образом. В разд. 2 приведены элементарные свойства дисперсионно-сдвиговых смесей нормальных законов. Основная часть этого раздела посвящена обсуждению возможности использования GVG-распределений в качестве асимптотических аппроксимаций. Высокую адекватность моделей типа дисперсионно-сдвиговых смесей нормальных законов можно легко объяснить довольно большим числом настраиваемых параметров в указанных моделях. Однако на самом деле их адекватность имеет гораздо более глубокие теоретические обоснования. В прикладной теории вероятностей принято считать, что та или иная модель может быть в достаточной мере обоснованной (адекватной) только тогда, когда она является *асимптотической аппроксимацией*, т. е. когда существует довольно простая предельная схема (например, схема максимума или схема суммирования) и соответствующая предельная теорема, в которой рассматриваемая модель выступает в качестве предельного распределения. Наличие такой формальной асимптотической схемы может дать дополнительную информацию о реальных механизмах, формирующих те или иные наблюдаемые статистические закономерности. Оказывается, что дисперсионно-сдвиговые смеси нормальных законов являются предельными законами в довольно простых предельных теоремах для случайно остановленных случайных блужданий. Такие теоремы позволяют однозначно связать конкретный смешивающий закон в дисперсионно-сдвиговых смесях с поведением интенсивности потока информативных событий, в результате которых накапливаются данные, характеризующие анализируемый случайный процесс. Тем самым эти теоремы как бы позволяют разделить вклады внешних и внутренних факторов в случайность поведения анализируемого процесса. В разд. 2 приведены и прокомментированы довольно простые предельные теоремы для случайных сумм типа центральной предельной те-

оремы, в которых предельными законами являются обобщенные дисперсионные гамма-распределения. Особое внимание в этом разделе уделено механизму формирования GG-распределений общего вида как асимптотических аппроксимаций. В качестве такого механизма предложен аналог закона больших чисел для отрицательных биномиальных случайных сумм независимых неодинаково распределенных случайных величин.

В разд. 3 приведены примеры подгонки GG-распределений к реальным данным об интенсивности потоков событий на финансовых рынках, а также результаты сравнения таких моделей с GIG-распределениями.

Для оценивания параметров упомянутых выше моделей недавно разработан, реализован и теоретически и экспериментально исследован новый комбинированный сеточный метод разделения дисперсионно-сдвиговых смесей нормальных законов, описываемый в разд. 4. Этот метод превосходит известные методы по точности и эффективности. Последнее обстоятельство особенно важно при применении этого метода для анализа реальных процессов с целью прогнозирования сопутствующих рисков, возможности и результаты которого на примере индекса Shanghai Composite также обсуждаются в разд. 4.

2 Обобщенные дисперсионные гамма-распределения как асимптотические аппроксимации

2.1 Элементарные свойства дисперсионно-сдвиговых смесей нормальных законов

Будем считать, что все случайные величины, о которых пойдет речь ниже, заданы на одном вероятностном пространстве $(\Omega, \mathfrak{A}, P)$. Если Y и U — независимые случайные величины с функциями распределения $\Phi(x)$ и $G(x)$ (при этом, очевидно, $P(U \geq 0) = 1$), а Z — случайная величина, удовлетворяющая соотношению $Z \stackrel{d}{=} \sigma Y \sqrt{U} + \alpha U + \beta$, то $P(Z < x) = F(x)$, где $F(x)$ — функция распределения дисперсионно-сдвиговой смеси нормальных законов, определяемая соотношением (1), $x \in \mathbb{R}$.

Легко убедиться, что если $EU < \infty$, то

$$EZ = \beta + \alpha EU, \quad (4)$$

а если при этом и $EU^2 < \infty$, то $EZ^2 = \beta^2 + (\sigma^2 + 2\alpha\beta)EU + \alpha^2 EU^2$, $DZ = \alpha^2 DU + \sigma^2 EU$.

При вычислении моментов более высоких порядков можно использовать следующее соотношение между характеристическими функциями f_Z и f_U случайных величин Z и U :

$$f_Z(t) = e^{it\beta} f_U\left(\alpha t + \frac{1}{2} i\sigma^2 t^2\right), \quad t \in \mathbb{R}.$$

Чтобы убедиться в справедливости последнего соотношения, достаточно заметить, что по теореме Фубини

$$\begin{aligned} f_Z(t) &= \mathbb{E} e^{itZ} = \mathbb{E} \exp\{it(\beta + \sigma\sqrt{U} \cdot Y + \alpha U)\} = \\ &= \int_0^\infty \mathbb{E} \exp\{it\beta + \sigma\sqrt{u} \cdot Y + \alpha u\} dG(u) = \\ &= e^{it\beta} \int_0^\infty \exp\left\{it\alpha u - \frac{1}{2} \sigma^2 t^2 u\right\} dG(u) = \\ &= e^{it\beta} \int_0^\infty \exp\left\{iu\left(\alpha + \frac{1}{2} i\sigma^2 t^2\right)\right\} dG(u) = \\ &= e^{it\beta} f_U\left(\alpha + \frac{1}{2} i\sigma^2 t^2\right), \quad t \in \mathbb{R}. \end{aligned}$$

2.2 Критерий сходимости распределений случайных сумм к обобщенным дисперсионным гамма-распределениям

Пусть $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$, — семейство последовательностей одинаково распределенных в каждой последовательности (при каждом фиксированном n) случайных величин. Пусть $\{N_n\}_{n \geq 1}$ — последовательность целочисленных неотрицательных случайных величин, таких что при каждом $n \geq 1$ случайные величины $N_n, X_{n,1}, X_{n,2}, \dots$ независимы. Положим $S_{n,k} = X_{n,1} + \dots + X_{n,k}$. Во избежание недоразумений полагаем $\sum_{j=1}^0 = 0$. Символ \implies будет обозначать слабую сходимость (сходимость по распределению). Всюду далее сходимость подразумевается при $n \rightarrow \infty$. Будем говорить, что $N_n \rightarrow \infty$ по вероятности, если $P(N_n \leq m) \rightarrow 0$ для любого $m > 0$.

В работе [10] доказана следующая базовая теорема.

Теорема 1. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\alpha \in \mathbb{R}$ такие, что*

$$P(S_{n,k_n} < x) \implies \Phi(x - \alpha). \quad (5)$$

Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость распределений

случайных сумм к некоторой функции распределения $F(x)$:

$$P(S_{n,N_n} < x) \implies F(x),$$

необходимо и достаточно, чтобы существовала функция распределения $G(x)$ такая, что

$$G(0) = 0;$$

$$F(x) = \int_0^\infty \Phi\left(\frac{x - \alpha z}{\sqrt{z}}\right) dG(z);$$

$$P(N_n < x k_n) \implies G(x).$$

Условие (5) выполняется в следующей довольно общей ситуации. Предположим, что случайные величины $X_{n,j}$ имеют конечные дисперсии. Также предположим, что величины $X_{n,j}$ могут быть представлены в виде $X_{n,j} = X_{n,j}^* + \alpha_n$, где $\alpha_n \in \mathbb{R}$, а $X_{n,j}^*$ — случайная величина с $\mathbb{E} X_{n,j}^* = 0$, $\mathbb{D} X_{n,j}^* = \sigma_n^2 < \infty$, так что $\mathbb{E} X_{n,1} = \alpha_n$ и $\mathbb{D} X_{n,1} = \sigma_n^2$. Предположим, что $\alpha_n k_n \rightarrow a$ и $k_n \sigma_n^2 \rightarrow 1$ при $n \rightarrow \infty$. Тогда вследствие хорошо известного результата о необходимых и достаточных условиях сходимости к нормальному закону распределений сумм независимых случайных величин с конечными дисперсиями в схеме серий (см., например, [11]) можно заметить, что соотношение (5) имеет место тогда и только тогда, когда выполнено условие Линдберга:

$$\lim_{n \rightarrow \infty} k_n \mathbb{E}(X_{n,1}^*)^2 \mathbb{I}(|X_{n,1}^*| \geq \varepsilon) = 0$$

для любого $\varepsilon > 0$ (здесь $\mathbb{I}(A)$ — индикаторная функция множества (события) A), т.е. квадратичные хвосты распределений слагаемых должны убывать достаточно быстро.

Следующее следствие теоремы 1, приведенное в [12], представляет собой критерий сходимости распределений случайных сумм к обобщенным дисперсионно-сдвиговым распределениям.

Следствие 1. *Предположим, что существуют последовательность натуральных чисел $\{k_n\}_{n \geq 1}$ и число $\alpha \in \mathbb{R}$ такие, что имеет место сходимость (5). Предположим, что $N_n \rightarrow \infty$ по вероятности. Для того чтобы имела место сходимость распределений случайных сумм к обобщенным дисперсионным гамма-распределениям:*

$$P(S_{n,N_n} < x) \implies P_{\text{GVG}}(x; \alpha, \sigma, \nu, \kappa, \delta), \quad (6)$$

необходимо и достаточно, чтобы

$$P(N_n < x k_n) \implies P_{\text{GG}}(x; \nu, \kappa, \delta), \quad (7)$$

где $P_{\text{GG}}(x; \nu, \kappa, \delta)$ — функция распределения обобщенного гамма-распределения, соответствующая плотности $p_{\text{GG}}(x; \nu, \kappa, \delta)$ (см. (2)).

В соотношениях (5)–(7) предельные функции распределения непрерывны, поэтому в них сходимость по распределению эквивалентна равномерной сходимости функций распределения.

2.3 Сходимость распределений отрицательных биномиальных случайных сумм к обобщенным гамма-распределениям

В следствии 1 ключевым условием является соотношение (7). В данном параграфе будет приведен пример асимптотической конструкции, в рамках которой справедлива предельная теорема, влекущая сходимость (7).

Рассмотрим случайную величину $\overline{N}_{r,p}$, имеющую отрицательное биномиальное распределение с параметрами $r > 0$ и $p \in (0, 1)$:

$$P(\overline{N}_{r,p} = k) = C_{r+k-2}^{k-1} p^r (1-p)^{k-1}, \quad k = 1, 2, \dots, \quad (8)$$

где для нецелых r

$$C_{r+k-2}^{k-1} = \frac{\Gamma(r+k-1)}{(k-1)! \Gamma(r)},$$

а $\Gamma(x)$ — эйлерова гамма-функция. В таком случае $E\overline{N}_{r,p} = r/p$.

Лемма 1. Если $p \rightarrow 0$, то

$$\lim_{p \rightarrow 0} \sup_x \left| P(p\overline{N}_{r,p} < x) - \frac{1}{\Gamma(r)} \int_0^x z^{r-1} e^{-z} dz \right| = 0. \quad (9)$$

Доказательство. Характеристическая функция $f_{r,p}(t)$ случайной величины $\overline{N}_{r,p}$ имеет вид:

$$f_{r,p}(t) = \left(\frac{pe^{it}}{1 - (1-p)e^{it}} \right)^r, \quad t \in \mathbb{R},$$

поэтому для каждого $t \in \mathbb{R}$

$$\begin{aligned} Ee^{itpN_{r,p}} &= f_{r,p}(pt) = \left(\frac{pe^{ipt}}{1 - (1-p)e^{ipt}} \right)^r = \\ &= \left(\frac{p}{e^{-ipt} - 1 + p} \right)^r = \left(\frac{p}{p - ipt + o(p)} \right)^r = \\ &= [1 - it + o(1)]^{-r} \rightarrow (1 - it)^{-r} \quad (10) \end{aligned}$$

при $p \rightarrow 0$. Но правая часть (10) — это характеристическая функция гамма-распределения с параметром формы r и параметром масштаба $\lambda = 1$. Таким образом, по теореме непрерывности имеет

место поточечная сходимость функций распределения, участвующих в (9), которая является равномерной в силу непрерывности предельной функции распределения. Лемма доказана.

Заметим, что случайная величина $\overline{N}_{1,p}$ имеет геометрическое распределение, а предельным в (9) является стандартное показательное распределение.

Рассмотрим последовательность случайных величин ζ_1, ζ_2, \dots . Пусть τ_1, τ_2, \dots — натуральнозначные случайные величины такие, что при каждом n случайная величина τ_n независима от последовательности ζ_1, ζ_2, \dots . В следующей лемме сходимость подразумевается при $n \rightarrow \infty$.

Лемма 2. Предположим, что существуют неограниченно возрастающая (убывающая к нулю) последовательность положительных чисел $\{b_n\}_{n \geq 1}$ и случайная величина ζ такие, что

$$b_n^{-1} \zeta_n \Rightarrow \zeta.$$

Если существуют неограниченно возрастающая (убывающая к нулю) последовательность положительных чисел $\{d_n\}_{n \geq 1}$ и случайная величина N такие, что

$$d_n^{-1} b_{\tau_n} \Rightarrow N, \quad (11)$$

то

$$d_n^{-1} \zeta_{\tau_n} \Rightarrow \zeta N, \quad (12)$$

причем случайные множители в правой части (12) независимы. Если дополнительно $\tau_n \rightarrow \infty$ по вероятности и семейство масштабных смесей функции распределения случайной величины ζ идентифицируемо, то условие (11) не только достаточно для (12), но и необходимо.

Доказательство см. в [13] (случай $b_n, d_n \rightarrow \infty$), [14] (случай $b_n, d_n \rightarrow 0$) или [15] (теорема 3.5.5).

Перейдем к постановке задачи, в рамках которой GG-распределения в общем их виде (2) оказываются предельными для случайных сумм независимых случайных величин. Пусть M_1, M_2, \dots — независимые случайные величины. Не ограничивая общности, можно считать, что эти величины принимают натуральные значения. Предположим, что случайная величина \overline{N}_{r,p_n} имеет отрицательное биномиальное распределение (3) с параметрами $r > 0$ и $p_n \in (0, 1)$, причем при каждом $n \in \mathbb{N}$ случайная величина \overline{N}_{r,p_n} независима от последовательности независимых случайных величин $\{M_j\}_{j \geq 1}$ и $p_n \rightarrow 0$ при $n \rightarrow \infty$. В качестве случайной величины N_n рассмотрим отрицательную биномиальную случайную сумму

$$N_n = \sum_{j=1}^{\overline{N}_{r,p_n}} M_j. \quad (13)$$

В частном случае, когда $r = 1$, N_n является геометрической суммой, свойства которой хорошо изучены для ситуации, в которой все величины M_1, M_2, \dots имеют одинаковое распределение [16]. Особо отметим, что здесь не предполагается, что распределения случайных величин M_1, M_2, \dots совпадают. Используя лемму 2, получаем следующее утверждение.

Теорема 3. *Предположим, что случайные величины M_j , $j \geq 1$, удовлетворяют условию статистической устойчивости:*

$$\frac{1}{n^{1/\nu}} \sum_{j=1}^n M_j \longrightarrow 1 \quad (14)$$

по вероятности при $n \rightarrow \infty$ с некоторым $\nu > 0$. Предположим, что случайная величина \bar{N}_{r,p_n} имеет отрицательное биномиальное распределение (3) с параметрами $r > 0$ и $p_n \rightarrow 0$ и при каждом $n \in \mathbb{N}$ независима от последовательности $\{M_j\}_{j \geq 1}$. Пусть случайные величины N_n имеют вид (16). Тогда

$$\lim_{n \rightarrow \infty} \sup_{x > 0} \left| \mathbb{P} \left(p_n^{1/\nu} N_n < x \right) - \frac{\nu}{\Gamma(r)} \int_0^x z^{\nu r - 1} e^{-z^\nu} dz \right| = 0.$$

Доказательство. Воспользуемся леммой 2, в которой для каждого $n \in \mathbb{N}$ положим $b_n = n^{1/\nu}$, $d_n = p_n^{-1/\nu}$, $\zeta_n = M_1 + \dots + M_n$, $\tau_n = \bar{N}_{r,p_n}$. Тогда $\zeta_{\tau_n} = N_n$ и в силу (14) $\zeta = 1$. Очевидно, что

$$\begin{aligned} \mathbb{P}(d_n^{-1} b_{\tau_n} < x) &= \mathbb{P} \left((p_n \tau_n)^{1/\nu} < x \right) = \\ &= \mathbb{P} \left(p_n \bar{N}_{r,p_n} < x^\nu \right), \end{aligned}$$

поэтому в соответствии с леммами 2 и 1

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left(p_n^{1/\nu} N_n < x \right) &= \lim_{n \rightarrow \infty} \mathbb{P} \left(p_n \bar{N}_{r,p_n} < x^\nu \right) = \\ &= \frac{1}{\Gamma(r)} \int_0^{x^\nu} z^{r-1} e^{-z} dz = \frac{\nu}{\Gamma(r)} \int_0^x z^{\nu r - 1} e^{-z^\nu} dz \end{aligned}$$

равномерно по $x > 0$. Лемма доказана.

В частном случае $r = 1$ распределение случайной величины \bar{N}_{r,p_n} является геометрическим, а предельное обобщенное гамма-распределение в теореме 3 является распределением Вейбулла с параметром ν . Таким образом, теорема 3, по сути, представляет собой закон больших чисел для отрицательных биномиальных случайных сумм необязательно одинаково распределенных независимых случайных величин и устанавливает условия сходимости распределений таких сумм к обобщенным

гамма-распределениям. Тем самым оно обобщает известные утверждения о сходимости геометрических сумм (см., например, [16]). В частности, при $r = 1$ теорема 3 описывает сходимость геометрических случайных сумм неодинаково распределенных независимых слагаемых к распределению Вейбулла.

Пусть k_n — натуральные числа, фигурирующие в условии (8). Поскольку параметр p_n в теореме 3 может быть произвольным, полагая $p_n = k_n^{-\nu}$, можно получить конкретный пример ситуации, в которой асимптотическим распределением числа слагаемых в сумме, фигурирующей в следствии 1, является GG-распределение. В этом примере одним из важнейших является условие статистической устойчивости (14), смысл которого в том, что интенсивность событий имеет нетривиальный тренд (затухание или, наоборот, возрастание).

3 Подгонка моделей обобщенных гамма-распределений к реальным данным об интенсивности потоков событий на финансовых рынках

В данном разделе будут приведены результаты статистического анализа конкретных финансовых данных, подтверждающие перспективность применения GVG-распределений в качестве моделей наблюдаемых статистических закономерностей. Из общей теории формирования финансовых индексов, основанной на предельных теоремах для случайных блужданий со случайными интенсивностями скачков [17], вытекает, что распределения (логарифмических) приращений хорошо аппроксимируются дисперсионно-сдвиговыми смесями нормальных законов. Этот вывод подтвержден многочисленными работами, в которых указанные закономерности хорошо описываются с помощью конкретных смесей, например обобщенных гиперболических распределений или их частных случаев [3, 18–28]. Как показано в теореме 1 и следствии 1, основным условием (критерием) адекватности той или иной дисперсионно-сдвиговой смеси нормальных законов является адекватность конкретного смешивающего распределения как модели распределения случайной интенсивности торгов. Здесь на нескольких примерах будет продемонстрировано, что GG-распределения лучше описывают статистические закономерности поведения случайной интенсивности торгов, нежели GIG-модели, и, следовательно, обобщенные

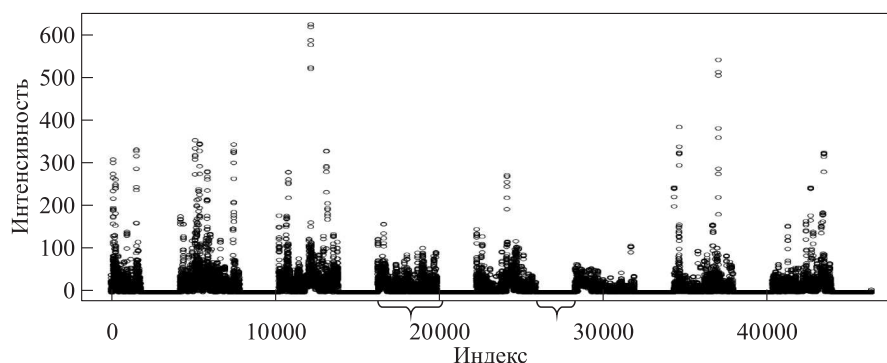


Рис. 1 Интенсивность торгов по акциям Сбербанка в течение дня

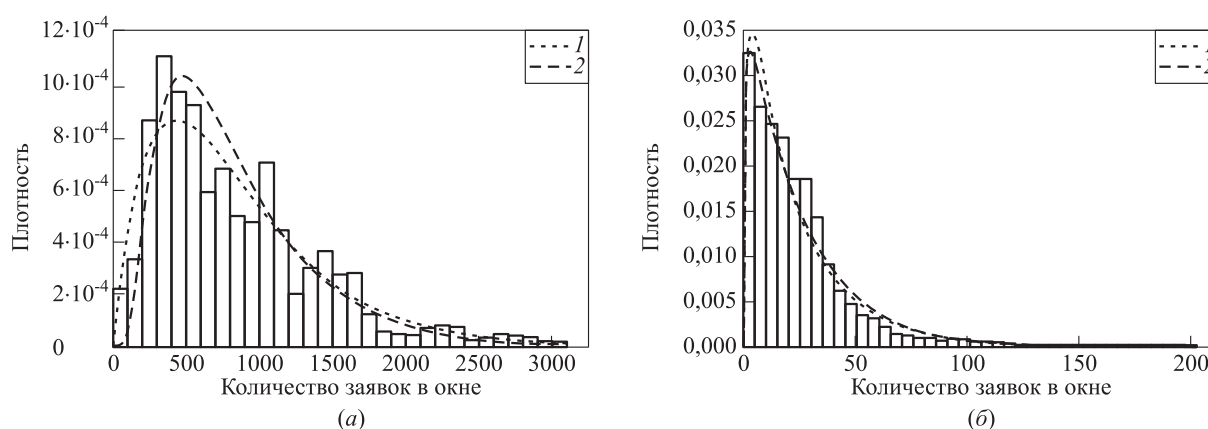


Рис. 2 Сравнение GG (1) и GIG (2) моделей по «склеенным» данным (а) и на одном «активном» периоде (б)

дисперсионные гамма-распределения лучше описывают статистические закономерности поведения (логарифмических) приращений соответствующих финансовых индексов.

Рассмотрим интенсивности потоков заявок (ордеров) на торгах акциями Сбербанка. Эти акции торгуются не очень активно, но торгуются интересно — в течение дня заметны «пробелы» с нулевой интенсивностью (рис. 1).

Если формально убрать периоды «отдыха» торгующихся, «склеить» оставшиеся интервалы и вычислить интенсивность с окном в 60 с, то получается картина, представленная на рис. 2, а, из которой видно, что GG-распределение (кривая 1) дает слегка лучшее приближение, чем GIG (кривая 2) (P -значения критерия хи-квадрат составляют соответственно 0,283 и 0,278).

Если же взять отдельно период активности (выделенный на рис. 1 фигурной скобкой снизу) и вычислить для него интенсивность торгов с окном в 5 с, то преимущество GG-модели станет неоспоримым (см. рис. 2, б). Видно, что, хотя визуально GG- и GIG-модели близки, P -значение для GIG-распределения практически равно нулю, поскольку хвост подогнанного GIG-распределения убывает

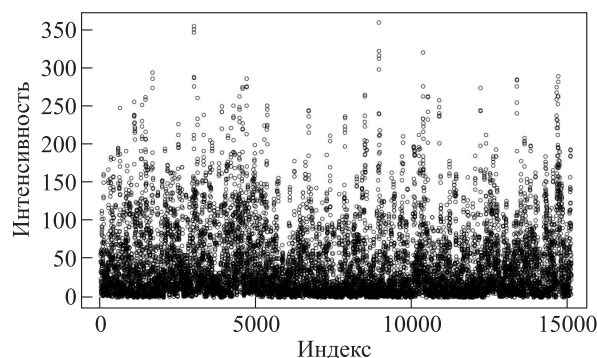


Рис. 3 Текущая интенсивность потока заявок на LSE

слишком быстро по сравнению с GG-моделью, для которой P -значение равно 0,393.

Аналогичным образом были также проанализированы очень высокочастотные данные о торгах среднеактивным инструментом на лондонской бирже LSE (London Stock Exchange). Здесь поток заявок гораздо плотнее, данные о текущей интенсивности представлены на рис. 3. Из-за высокой плотности заявок за все время торгов интенсивность, даже вычисленная для 5-секундных окон, не опускается до нуля. На рис. 4, а представлены

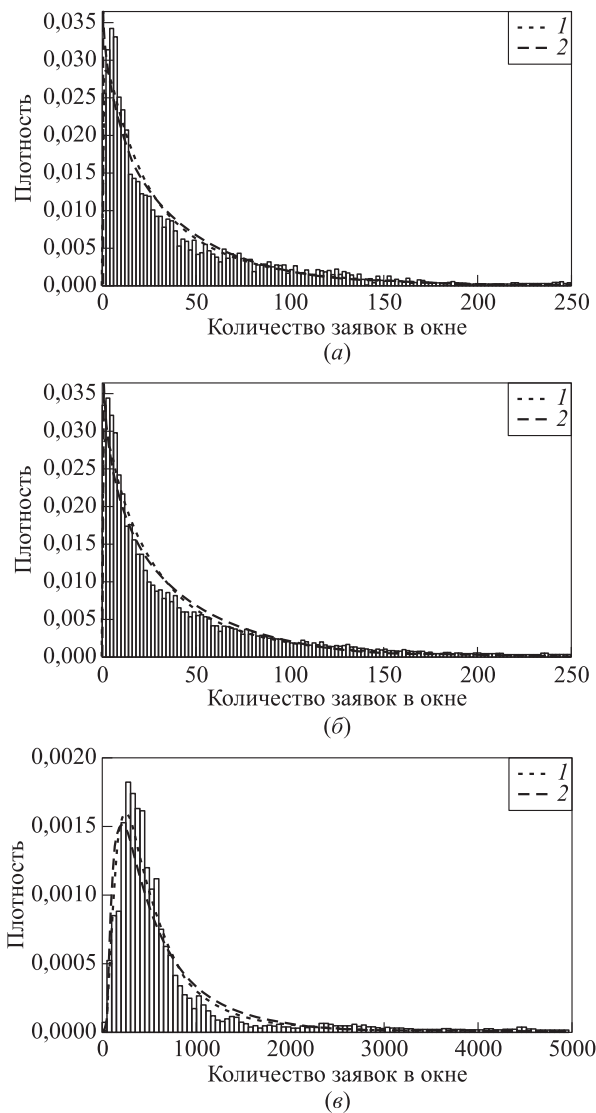


Рис. 4 Результаты подгонки GG (1) и GIG (2) моделей к данным об интенсивности потока заявок, вычисленных по 5-секундным окнам для интервала длиной 1 ч с 10:00 до 11:00 (а) и к данным об интенсивности потока заявок для всего торгового дня, вычисленных по 5-секундным (б) и 60-секундным (в) окнам

результаты подгонки GG- и GIG-моделей к данным об интенсивности потока заявок, вычисленных по 5-секундным окнам для интервала длиной 1 ч с 10:00 до 11:00.

На рис. 4, б и 4, в приведены результаты подгонки GG- и GIG-моделей к данным об интенсивности потока заявок для всего торгового дня, вычисленных по 5-секундным (б) и 60-секундным (в) окнам. Везде заметно преимущество GG-моделей, подкрепленное большими P -значениями.

4 Прогнозирование финансовых рисков с помощью модифицированного сеточного метода скользящего разделения обобщенных дисперсионных гамма-распределений

Одной из важнейших практических задач на финансовых рынках является задача прогнозирования рисков, связанных с превышением теми или иными показателями критических порогов. Помимо непосредственного исследования распределений любая финансовая организация заинтересована в получении более-менее достоверных прогнозов на основе наблюдаемых данных. Прогнозирование несет в себе большой спекулятивный фактор, но некоторые жесткие требования к любому осмысленному методу прогнозирования известны заранее: метод должен работать достаточно быстро, чтобы прогноз оставлял время для принятия решения, и он должен показывать хорошие результаты на случайно выбранных исторических данных.

Задача прогнозирования рисков сводится к прогнозированию распределений. Для упрощения задачи оценки и прогнозирования распределений часто используется подход снижения размерности путем априорного сужения классов допустимых смесей. В данном разделе описывается алгоритм прогнозирования параметров дисперсионно-сдвиговых смесей (в частности, оценки рисков) на примере обобщенных дисперсионных гамма-распределений. Этот алгоритм был ранее отработан на обобщенных гиперболических распределениях [29].

Возьмем интересующий нас временной ряд и применим к нему стандартный подход разделения смесей нормальных законов с использованием скользящего окна. Для этого зафиксируем размер окна w и сдвиг окна $s \leq w$. Далее на каждом окне применим модифицированный двухэтапный сеточный метод разделения дисперсионно-сдвиговых смесей нормальных законов, предложенный в работе [30].

Без ограничения общности будем считать, что параметр сдвига β (см. (1)) равен нулю.

На первом этапе на положительной полупрямой выделим основную часть носителя смешивающего распределения, т. е. ограниченный интервал, вероятность которого, вычисленная в соответствии со смешивающим распределением, практически равна единице. На этот интервал накинём конечную сетку, содержащую, возможно, очень много

известных узлов u_1, \dots, u_K . Приближим искомое обобщенное дисперсионное гамма-распределение конечной смесью нормальных законов:

$$P_{\text{GGV}}(x; \alpha, \sigma, \nu, \kappa, \delta) \approx \sum_{i=1}^K p_i \Phi\left(\frac{x - \alpha u_i}{\sqrt{u_i}}\right), \quad x \in \mathbb{R}. \quad (15)$$

В смеси, стоящей в правой части соотношения (15), неизвестными являются только параметры $p_1, \dots, p_{K-1}, \alpha$.

Пусть x_1, \dots, x_n — анализируемая выборка значений случайной величины с оцениваемым обобщенным гиперболическим распределением. Итерационный процесс, определяющий сеточный ЕМ (expectation-maximization) алгоритм для данной задачи, задается следующим образом. Пусть $p_1^{(m)}, \dots, p_{K-1}^{(m)}, \alpha^{(m)}$ — оценки параметров p_1, \dots, p_{K-1} и α на m -й итерации, $p_K^{(m)} = 1 - p_1^{(m)} - \dots - p_{K-1}^{(m)}$. Обозначим

$$\phi_{ij}^{(m)} = \frac{1}{\sqrt{u_i}} \phi\left(\frac{x_j - \alpha^{(m)} u_i}{\sqrt{u_i}}\right);$$

$$g_{ij}^{(m)} = \frac{p_i^{(m)} \phi_{ij}^{(m)}}{\sum_{r=1}^K p_r^{(m)} \phi_{rj}^{(m)}}, \quad i = 1, \dots, K; \quad j = 1, \dots, n.$$

Тогда, используя стандартные рассуждения, определяющие вычислительные формулы ЕМ-алгоритма для параметров конечной смеси нормальных законов (см., например, [17, разд. 5.3.7–5.3.8]), следует положить

$$p_i^{(m+1)} = \frac{1}{n} \sum_{j=1}^n g_{ij}^{(m)}, \quad i = 1, \dots, K. \quad (16)$$

Обозначим $\bar{x} = (1/n) \sum_{j=1}^n x_j$. Используя соотношение (5.3.24) в [17], с учетом очевидного равенства $\sum_{i=1}^K g_{ij}^{(m)} = 1$ можно заметить, что уточненная оценка параметра α имеет вид:

$$\alpha^{(m+1)} = \frac{\bar{x}}{\sum_{i=1}^K u_i p_i^{(m+1)}}, \quad (17)$$

т. е. равна отношению генерального выборочного среднего и текущего эмпирического среднего смешивающего распределения, что вполне согласуется с соотношением (4).

Как известно, классический ЕМ-алгоритм обладает свойством монотонности. Поэтому если узлы u_1, \dots, u_K сетки различны, неотрицательны и известны, то итерационный процесс (16), (17) является монотонным, т. е. каждая его итерация не уменьшает целевую сеточную функцию правдоподобия

$$L(p_1, \dots, p_K, \alpha; x_1, \dots, x_n) = \prod_{j=1}^n \left[\sum_{i=1}^K \frac{p_i}{\sqrt{u_i}} \phi\left(\frac{x_j - \alpha^{(m)} u_i}{\sqrt{u_i}}\right) \right].$$

В [17, разд. 5.7.4] показано, что при каждом фиксированном значении параметра α сеточная функция правдоподобия $L(p_1, \dots, p_{K-1}, \alpha; x_1, \dots, x_n)$ вогнута по аргументам p_1, \dots, p_{K-1} . Поэтому на каждом шаге итерационного процесса вместо соотношения (3) можно использовать любую более быстрый алгоритм максимизации функции $L(p_1, \dots, p_{K-1}, \alpha^{(m)}; x_1, \dots, x_n)$ по переменным p_1, \dots, p_{K-1} . Например, оценки весов p_1, \dots, p_K можно искать методом условного градиента [17, 31].

Таким образом, на первом этапе получают оценки параметра α и весов всех узлов u_i конечной сетки, накинутой на носитель смешивающего обобщенного гамма-распределения $P_{\text{GG}}(x; \nu, \kappa, \delta)$.

На втором этапе остается применить какой-либо стандартный метод подгонки GG-распределения $P_{\text{GG}}(x; \nu, \kappa, \delta)$ к эмпирическим данным типа гистограммы $(u_1, p_1), \dots, (u_K, p_K)$. Например, параметры ν, κ и δ можно оценить, решая задачу наименьших квадратов

$$(\nu^*, \kappa^*, \delta^*) = \arg \min_{\nu, \kappa, \delta} \sum_{i=1}^K \left[p_i - \int_{(u_{i-1} + u_i)/2}^{(u_i + u_{i+1})/2} p_{\text{GG}}(u; \nu, \kappa, \delta) du \right]^2,$$

где $u_0 = 0, u_{K+1} = \infty$. Или же эти параметры можно найти из условия

$$(\nu^*, \kappa^*, \delta^*) = \arg \min_{\nu, \kappa, \delta} D_{\text{KL}} [p_{\text{GG}}(u; \nu, \kappa, \delta), h(u)],$$

где $h(u)$ — гистограмма, построенная по значениям $(u_1, p_1), \dots, (u_K, p_K)$:

$$h(u) = \begin{cases} 0, & u \leq \frac{1}{2} u_1; \\ \frac{u_{i+1} - u_{i-1}}{2p_i}, & \frac{1}{2} (u_{i-1} + u_i) < u \leq \frac{1}{2} (u_i + u_{i+1}); \\ 0, & u > \frac{1}{2} (3u_K - u_{K-1}), \end{cases}$$

а $D_{\text{KL}} [p_{\text{GG}}(u; \nu, \kappa, \delta), h(u)]$ — расстояние (дивергенция) Кульбака–Лейблера.

Конкретный алгоритм выбора сетки на первом этапе описан в работе [30].

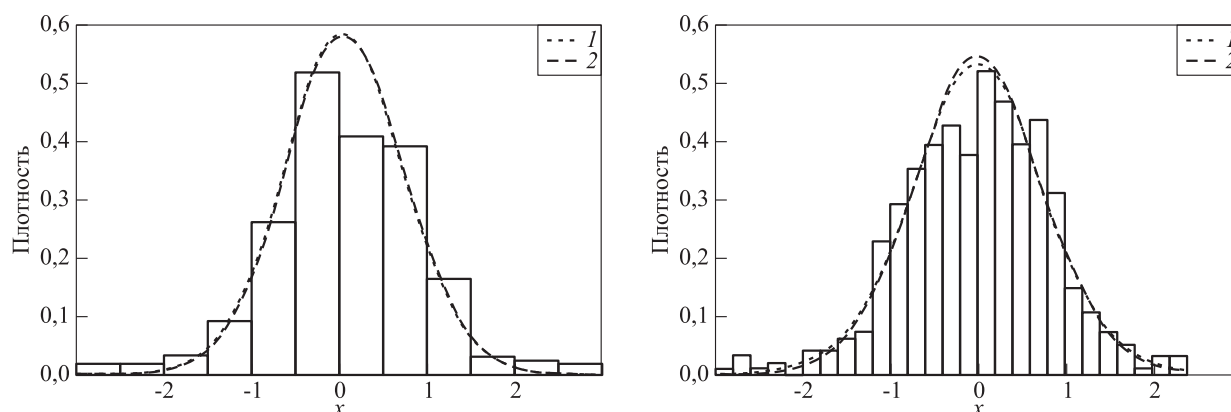


Рис. 5 Результаты подгонки GVG (1) и GH (2) распределений к приращениям логарифмов индекса Shanghai Composite при двух разных положениях скользящего окна

В качестве входных данных для метода прогнозирования используется результат работы описанного выше двухэтапного сеточного метода разделения дисперсионно-сдвиговых смесей нормальных законов, т. е. ряд оценок параметров распределений, полученных на \hat{N} известных окнах с историческими наблюдениями, $\theta_1, \theta_2, \dots, \theta_N$, где каждое $\theta_i = (\alpha_i, \nu_i, \kappa_i, \delta_i)^T$.

Конечная задача прогнозирования — получить оценки $\theta_{n+1}, \theta_{n+2}, \dots$ для окон, которые будут частично или полностью состоять из будущих наблюдений. Для этого рассмотрим соотношение

$$\tilde{\theta}_{i+1} = F_1 \theta_i + F_2 \theta_{i-1} + \dots + F_r \theta_{i-r+1},$$

где $r \in \mathbb{N}$ — заранее фиксированный параметр, имеющий смысл порядка прогноза; $F_j \in \mathbb{R}^{5 \times 5}$ — матрицы-регрессоры.

Процедура поиска параметров модели имеет вид

$$(F_1, \dots, F_r) = \arg \min \sum_{i=r+1}^{N-1} (\theta_{i+1} - \tilde{\theta}_{i+1})^2. \quad (18)$$

По сути, имеем авторегрессионную модель порядка r , где поиск матриц F_j (обучение модели) производится с использованием минимизации суммарного RSS (residual sum of squares) по $\hat{N} - r$ предсказаниям модели на известных данных. Для простоты обозначений примем $N = \hat{N} - r$.

Соотношение (18) представляет собой разновидность линейной регрессии, и ее программная реализация не представляется сложной. Более того, во многих статистических пакетах есть встроенные функции расчета матриц F_j для случаев $r = 1, 2$.

В работе [32] детально описаны особенности использования модифицированного сеточного алгоритма и приведены практические советы для получения входного ряда $\theta_1, \theta_2, \dots, \theta_n$.

В качестве исходных данных возьмем индекс Shanghai Composite — основной индикатор китайской биржи, включающий в себя все торгующиеся на этой бирже компании. Индекс является ключевым для китайского рынка, поэтому он достаточно часто подвергается анализу и попыткам предсказания своей динамики.

Будет рассматриваться изменение логарифмов значений индекса (цены) с частотой (тиком), равной 1 мин, на протяжении трех рабочих дней, начиная с открытия биржи 5 января 2015 г. Размер окна установим равным 3 ч: $w = 180$, сдвиг окна минимальный — одно наблюдение: $s = 1$ тик.

Для начала проверим целесообразность использования именно GVG-моделей для описания статистических закономерностей поведения приращений логарифмов индекса Shanghai Composite. Результаты этой проверки иллюстрируются на рис. 5, где приведены результаты подгонки GVG и GH (generalized hyperbolic) распределений к указанным приращениям при двух разных положениях скользящего окна. Видно, что оба распределения удовлетворительно описывают данные, но P -значение для GVG-распределения в обоих случаях выше. Аналогичная проверка была произведена для 410 положений окна. В 403 случаях P -значение для GVG-распределения оказалось выше, чем для GH-распределения.

Применяя модифицированный сеточный алгоритм к исходным данным, получаем ряд оценок параметров $\theta_i = (\alpha_i, \nu_i, \kappa_i, \delta_i)^T$. Так как алгоритм устойчив к входным данным, а сдвиг окна минимальный, ряд θ_i дает сглаженную картину. Далее применяется алгоритм прогнозирования, описанный в [29]. Результаты прогнозирования для разных временных горизонтов представлены на рис. 6, где показано изменение истинного и прогнозируемого распределения с расширением горизонта прогно-

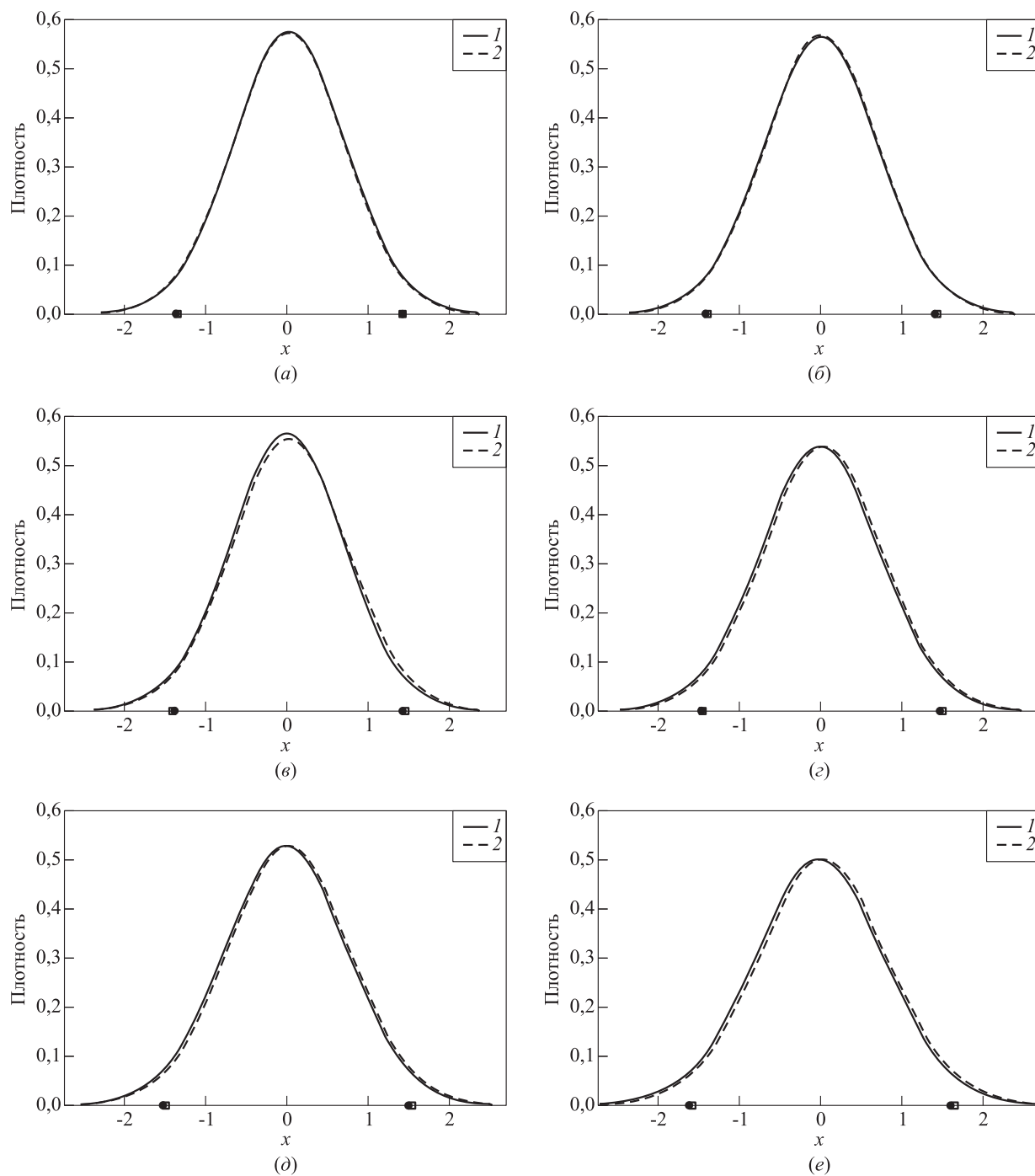


Рис. 6 Прогнозируемое (1) и истинное (2) распределения для следующих горизонтов: (а) 10 мин, $T + 10$; (б) 30 мин, $T + 30$; (в) 1 ч, $T + 60$; (г) 1,5 ч, $T + 90$; (д) 2 ч, $T + 120$; (е) 3 ч, $T + 180$

зирования, а также приведены значения прогнозированных и «истинных» параметров указанных распределений.

Особый интерес представляет сравнение динамики отдельных параметров распределения прогноза и истинных исторических значений без использования самого вида распределения. На рис. 7

отдельно показаны прогнозируемые значения параметра ν (после момента времени $T = 100$) и его истинные значения.

Обратим внимание, что и истинные, и прогнозируемые значения параметра ν лежат между нулем и единицей, т. е. подогнанные распределения занимают промежуточное место между законами с экс-

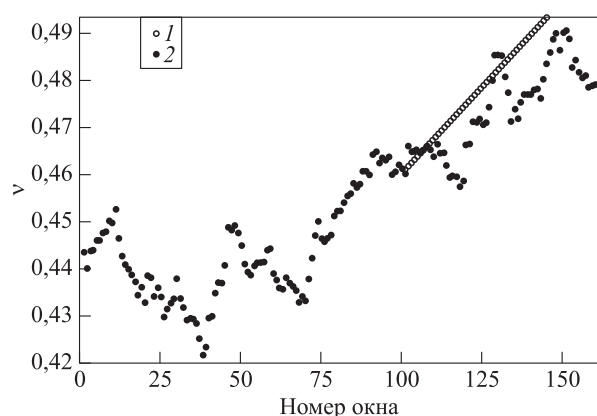


Рис. 7 Прогнозируемые (1) и истинные (2) значения параметра ν

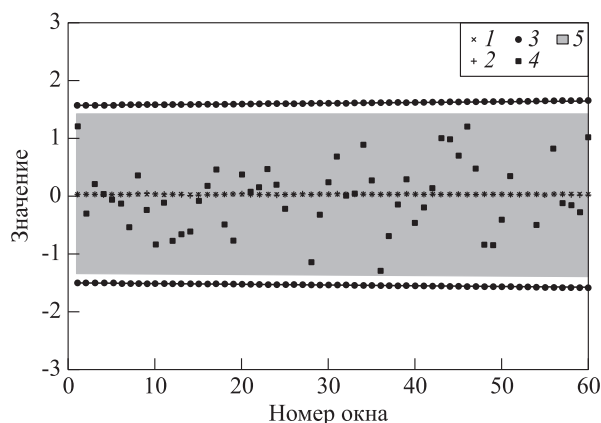


Рис. 8 Прогнозируемые значения медианы (1), математического значения (2) и квантилей логарифмических приращений (3) индекса Shanghai Composite, а также истинные значения индекса (4) и квантилей (5)

понижающимися хвостами и распределениями с хвостами, убывающими степенным образом.

Наконец, на рис. 8 приведены прогнозируемые значения базовых числовых характеристик процесса (среднее, медиана, квантили порядков 0,025 и 0,975), а также истинные значения процесса.

На основании полученных результатов можно сделать вывод, что модель типа GVG-распределений хорошо подходит для прогнозирования как минимум на 2 ч вперед. Важно отметить, что найденная модель достаточно точно оценивает хвосты распределений, позволяя полагаться на полученные прогнозы при оценке рисков.

При практическом использовании модели на поступающих в реальном времени данных постепенно становятся доступными новые исходные параметры, что позволяет проводить повторное обучение модели (пересчитывая регрессионные

матрицы). Это особенно удобно, поскольку поиск необходимых матриц происходит очень быстро и не представляет большой вычислительной сложности.

Литература

1. Barndorff-Nielsen O. E. Exponentially decreasing distributions for the logarithm of particle size // Proc. Roy. Soc. Lond. A, 1977. Vol. 353. P. 401–419.
2. Barndorff-Nielsen O. E., Kent J., Sørensen M. Normal variance-mean mixtures and z -distributions // Int. Statist. Rev., 1982. Vol. 50. No. 2. P. 145–159.
3. Barndorff-Nielsen O. E. Hyperbolic distributions and distributions of hyperbolae // Scand. J. Statist., 1978. Vol. 5. P. 151–157.
4. Королев В. Ю., Соколов И. А. Скошенные распределения Стьюдента, дисперсионные гамма-распределения и их обобщения как асимптотические аппроксимации // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 2–10.
5. Stacy E. W. A generalization of the gamma distribution // Ann. Math. Stat., 1962. Vol. 33. P. 1187–1192.
6. Laherrère J., Sornette D. Stretched exponential distributions in nature and economy: “Fat tails” with characteristic scales // Eur. Phys. J. B, 1998. Vol. 2. P. 525–539.
7. Malevergne Y., Pisarenko V., Sornette D. Empirical distributions of stock returns: Between the stretched exponential and the power law? // Quantitative Finance, 2005. Vol. 5. P. 379–401.
8. Malevergne Y., Pisarenko V., Sornette D. On the power of generalized extreme value (GEV) and generalized Pareto distribution (GDP) estimators for empirical distributions of stock returns // Appl. Financial Economics, 2006. Vol. 16. P. 271–289.
9. Антонов С. Н., Кокиаров С. Н. Об асимптотическом поведении хвостов масштабных смесей нормальных распределений // Статистические методы оценивания и проверки гипотез. — Пермь: Изд-во Пермского ун-та, 2006. С. 90–105.
10. Королев В. Ю. Обобщенные гиперболические законы как предельные для распределений случайных сумм // Теория вероятностей и ее применения, 2013. Т. 58. Вып. 1. С. 117–132.
11. Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. — М.-Л.: ГИТТЛ, 1949. 264 с.
12. Закс Л. М., Королев В. Ю. Обобщенные дисперсионные гамма-распределения как предельные для случайных сумм // Информатика и её применения, 2013. Т. 7. Вып. 1. С. 105–115.
13. Королев В. Ю. Сходимость случайных последовательностей с независимыми случайными индексами. I // Теория вероятностей и ее применения, 1994. Т. 39. Вып. 2. С. 313–333.

14. *Korolev B. Yu.* Сходимость случайных последовательностей с независимыми случайными индексами. II // Теория вероятностей и ее применения, 1995. Т. 40. Вып. 4. С. 907–910.
15. *Bening V. E., Korolev V. Yu.* Generalized Poisson models and their applications in insurance and finance. — Utrecht: VSP, 2002. 456 p.
16. *Kalashnikov V. V.* Geometric sums: Bounds for rare events with applications. — Dordrecht: Kluwer Academic Publs., 1997. 270 p.
17. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского ун-та, 2011. 510 с.
18. *Barndorff-Nielsen O. E.* Models for non-Gaussian variation, with applications to turbulence // Proc. Roy. Soc. Lond. A, 1979. Vol. 368. P. 501–520.
19. *Madan D. B., Seneta E.* The variance gamma (V.G.) model for share market return // J. Business, 1990. Vol. 63. P. 511–524.
20. *Eberlein E., Keller U.* Hyperbolic distributions in finance // Bernoulli, 1995. Vol. 1. No. 3. P. 281–299.
21. *Prause K.* Modeling financial data using generalized hyperbolic distributions. — Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1997. Preprint No. 48. 14 p.
22. *Carr P. P., Madan D. B., Chang E. C.* The Variance Gamma process and option pricing // Eur. Finance Rev., 1998. Vol. 2. P. 79–105.
23. *Eberlein E., Keller U., Prause K.* New insights into smile, mispricing and value at risk: The hyperbolic model // J. Business, 1998. Vol. 71. P. 371–405.
24. *Barndorff-Nielsen O. E.* Processes of normal inverse Gaussian type // Finance Stochastics, 1998. Vol. 2. P. 41–68.
25. *Eberlein E., Prause K.* The generalized hyperbolic model: Financial derivatives and risk measures. — Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1998. Preprint No. 56. 168 p.
26. *Ширяев А. Н.* Основы стохастической финансовой математики. Т. 1. Факты. Модели. — М.: Фазис, 1998. 512 с.
27. *Eberlein E.* Application of generalized hyperbolic Levy motions to finance. — Freiburg: Universität Freiburg, Institut für Mathematische Stochastic, 1999. Preprint No. 64. P. 319–336.
28. *Barndorff-Nielsen O. E., Blæsild P., Schmiegel J.* A parsimonious and universal description of turbulent velocity increments // Eur. Phys. J., 2004. Vol. B41. P. 345–363.
29. *Корчагин А. Ю., Королев В. Ю.* Прогнозирование финансовых рисков с помощью модифицированного сеточного метода скользящего разделения дисперсионно-сдвиговых смесей нормальных законов // Вестник Карагандинского ун-та, 2015. Т. 78. Вып. 2. С. 65–74.
30. *Королев В. Ю., Корчагин А. Ю.* Модифицированный сеточный метод разделения дисперсионно-сдвиговых смесей нормальных законов // Информатика и её применения, 2014. Т. 8. Вып. 4. С. 11–19.
31. *Королев В. Ю., Назаров А. Л.* Разделение смесей вероятностных распределений при помощи сеточных методов моментов и максимального правдоподобия // Автоматика и телемеханика, 2010. Вып. 3. С. 98–116.
32. *Корчагин А. Ю., Ярошенко И. И.* О практическом использовании модифицированного сеточного метода разделения дисперсионно-сдвиговых смесей нормальных законов // Информатика и её применения, 2015. Т. 9. Вып. 1. С. 2–10.

Поступила в редакцию 10.11.15

MODELING OF STATISTICAL REGULARITIES IN FINANCIAL MARKETS BY GENERALIZED VARIANCE GAMMA DISTRIBUTIONS

V. Yu. Korolev^{1,2}, A. Yu. Korchagin², and I. A. Sokolov³

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

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

Abstract: Some aspects of the application of generalized variance gamma distributions for modeling statistical regularities in financial markets are discussed. The paper describes elementary properties of generalized variance gamma distributions as special normal variance-mean mixtures in which mixing distributions are the generalized gamma laws. Limit theorems for sums of a random number of independent random variables are presented that are analogs of the law of large numbers and the central limit theorem. These theorems give grounds for the possibility of using generalized variance gamma distributions as asymptotic approximations. The paper presents

the results of practical fitting of generalized variance gamma distributions to real data concerning the behavior of financial indexes as well as of fitting generalized gamma distributions to the observed intensities of information flows in contemporary financial information systems. The results of comparison of generalized gamma models with generalized hyperbolic models demonstrate the superiority of the former over the latter. The methods for parameter estimation of generalized gamma models are also discussed as well as their application for predicting processes in financial markets.

Keywords: random sum; normal mixture; normal variance-mean mixture; generalized hyperbolic distribution; generalized variance-gamma distribution; generalized gamma distribution; law of large numbers; central limit theorem

DOI: 10.14357/19922264150402

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project No. 14-07-00041a).

References

1. Barndorff-Nielsen, O. E. 1977. Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. Lond. A* 353:401–419.
2. Barndorff-Nielsen, O. E., J. Kent, and M. M. Sørensen. 1982. Normal variance-mean mixtures and z -distributions. *Int. Statist. Rev.* 50(2):145–159.
3. Barndorff-Nielsen, O. E. 1978. Hyperbolic distributions and distributions of hyperbolae. *Scand. J. Statist.* 5:151–157.
4. Korolev, V. Yu., and I. A. Sokolov. 2012. Skoshennye raspredeleniya St'yudenta, dispersionnyye gamma-raspredeleniya i ikh obobshcheniya kak asimptoticheskie approksimatsii [Skewed Student's distributions, variance gamma distributions and their generalizations as asymptotic approximations]. *Informatika i ee Primeneniya — Inform. Appl.* 6(1):2–10.
5. Stacy, E. W. 1962. A generalization of the gamma distribution. *Ann. Math. Stat.* 33:1187–1192.
6. Laherrère, J., and D. Sornette. 1998. Stretched exponential distributions in nature and economy: “Fat tails” with characteristic scales. *Eur. Phys. J. B* 2:525–539.
7. Malevergne, Y., V. Pisarenko, and D. Sornette. 2005. Empirical distributions of stock returns: Between the stretched exponential and the power law? *Quantitative Finance* 5:379–401.
8. Malevergne, Y., V. Pisarenko, and D. Sornette. 2006. On the power of generalized extreme value (GEV) and generalized Pareto distribution (GDP) estimators for empirical distributions of stock returns. *Appl. Financial Economics* 16:271–289.
9. Antonov, S. N., and S. N. Koksharov. 2006. Ob asimptoticheskom povedenii khvostov masshtabnykh smesey normalnykh raspredeleniy [On asymptotic behavior of tails of normal variance mixtures]. *Statisticheskie metody otsenivaniya i proverki gipotez* [Statistical methods for estimation and hypothesis validation]. Perm': Perm' University Press. 2–10.
10. Korolev, V. Yu. 2013. Obobshchennyye giperbolicheskie raspredeleniya kak predel'nye dlya sluchaynykh summ [Generalized hyperbolic distributions as limiting for random sums] *Theory Probab. Appl.* 58(1):117–132.
11. Gnedenko, B. V., and A. N. Kolmogorov. 1949. *Predel'nye raspredeleniya dlya summ nezavisimyykh sluchaynykh velichin* [Limit distributions for sums of independent random variables]. Moscow–Leningrad: GITTL. 264 p.
12. Zaks, L. M., and V. Yu. Korolev. 2013. Obobshchennyye dispersionnyye gamma-raspredeleniya kak predel'nye dlya sluchaynykh summ [Generalized variance gamma distributions as limiting for random sums]. *Informatika i ee Primeneniya — Inform. Appl.* 7(1):105–115.
13. Korolev, V. Yu. 1994. Skhodimost' sluchaynykh posledovatel'nostey s nezavisimymi sluchaynymi indeksami. I [Convergence of random sequences with independent random indices. I]. *Theory Probab. Appl.* 39(2):313–333.
14. Korolev, V. Yu. 1995. Skhodimost' sluchaynykh posledovatel'nostey s nezavisimymi sluchaynymi indeksami. II [Convergence of random sequences with independent random indices. II]. *Theory Probab. Appl.* 40(4):907–910.
15. Bening, V. E., and V. Yu. Korolev. 2002. *Generalized Poisson models and their applications in insurance and finance*. Utrecht: VSP. 456 p.
16. Kalashnikov, V. V. 1997. *Geometric sums: Bounds for rare events with applications*. Dordrecht: Kluwer Academic Publs. 270 p.
17. Korolev, V. Yu. 2011. *Veroyatnostno-statisticheskie metody dekompozitsii volatil'nosti khaoticheskikh protsessov* [Probabilistic and statistical methods for the decomposition of volatility of chaotic processes]. Moscow: Moscow University Press. 510 p.
18. Barndorff-Nielsen, O. E. 1979. Models for non-Gaussian variation, with applications to turbulence. *Proc. Roy. Soc. Lond. A* 368:501–520.
19. Madan, D. B., and E. Seneta. 1990. The variance gamma (V.G.) model for share market return. *J. Business* 63:511–524.
20. Eberlein, E., and U. Keller. 1995. Hyperbolic distributions in finance. *Bernoulli* 1(3):281–299.
21. Prause, K. 1997. Modeling financial data using generalized hyperbolic distributions. Freiburg: Universität Freiburg, Institut für Mathematische Stochastic. Preprint No. 48. 14 p.

22. Carr, P. P., D. B. Madan, and E. C. Chang. 1998. The Variance Gamma process and option pricing. *Eur. Finance Rev.* 2:79–105.
23. Eberlein, E., U. Keller, and K. Prause. 1998. New insights into smile, mispricing and value at risk: The hyperbolic model. *J. Business* 71:371–405.
24. Barndorff-Nielsen, O. E. 1998. Processes of normal inverse Gaussian type. *Finance Stochastics* 2:41–68.
25. Eberlein, E., and K. Prause. 1998. The generalized hyperbolic model: Financial derivatives and risk measures. Freiburg: Universität Freiburg, Institut für Mathematische Stochastic. Preprint No. 56. 168 p.
26. Shiryaev, A. N. 1998. *Osnovy stokhasticheskoy finansovoy matematiki. T. 1. Fakty. Modeli* [Basics of stochastic financial mathematics. Vol. 1. Facts. Models]. Moscow: Fazis. 512 p.
27. Eberlein, E. 1999. Application of generalized hyperbolic Levy motions to finance. Freiburg: Universität Freiburg, Institut für Mathematische Stochastic. Preprint No. 64. 319–336.
28. Barndorff-Nielsen, O. E., P. Blæsild, and J. Schmiegel. 2004. A parsimonious and universal description of turbulent velocity increments. *Eur. Phys. J. B* 41:345–363.
29. Korchagin, A. Yu., and V. Yu. Korolev. 2015. Prognozirovanie finansovykh riskov s pomoshch'yu modifitsirovannogo setochnogo metoda skol'zyashchego razdeleniya dispersionno-sdvigovykh smesey normal'nykh zakonov [Forecasting financial risks using modified grid method of moving window decomposition of variance-mean normal mixtures]. *Vestnik Karagandinskogo Un-ta* 78(2):65–74.
30. Korolev, V. Yu., and A. Yu. Korchagin. 2014. Modifitsirovanny setochnyy metod razdeleniya dispersionno-sdvigovykh smesey normal'nykh zakonov [Modified grid-based method for decomposition of variance-mean normal mixtures]. *Informatika i ee Primeneniya — Inform. Appl.* 8(4):11–19.
31. Korolev, V. Yu., and A. L. Nazarov. 2010. Razdelenie smesey veroyatnostnykh raspredeleniy pri pomoshchi setochnykh metodov momentov i maksimal'nogo pravdopodobiya [Separation of mixtures using grid moment-based methods and maximum likelihood]. *Avtomatika i Telemekhanika* [Automatics Telemechanics] 3:98–116.
32. Korchagin, A. Yu., and I. I. Yaroshenko. 2015. O prakticheskom ispol'zovanii modifitsirovannogo setochnogo metoda razdeleniya dispersionno-sdvigovykh smesey normal'nykh zakonov [On practical usage of modified grid method of decomposition of normal variance-mean mixtures.] *Informatika i ee Primeneniya — Informatics Appl.* 9(1):2–10.

Received November 10, 2015

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; vkorolev@cs.msu.su

Korchagin Alexander Yu. (b. 1989) — junior scientist, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; sasha.korchagin@gmail.com

Sokolov Igor A. (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, Director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; isokolov@ipiran.ru

ПОТОКОВАЯ МОДЕЛЬ ВЫЧИСЛЕНИЙ КАК ПАРАДИГМА ПРОГРАММИРОВАНИЯ БУДУЩЕГО

Д. Н. Змеев¹, А. В. Климов², Н. Н. Левченко³, А. С. Окунев⁴, А. Л. Стемповский⁵

Аннотация: Описана потоковая модель вычислений с динамически формируемым контекстом, а также особенности ее архитектурной реализации. Потоковая модель вычислений позволяет решать проблемы, возникающие при создании и применении суперкомпьютеров. Одной из таких проблем является то, что с ростом возможности максимального распараллеливания, предоставляемой пользователям создателями современной аппаратуры, все сложнее становится загрузить постоянно увеличивающееся число функциональных устройств вычислительных ядер, оставаясь в рамках традиционного программирования. Описаны преимущества предлагаемой модели вычислений. Сравняются парадигма «раздачи», в которой работает потоковая модель вычислений, и традиционная парадигма «сбора». Приведены основные особенности архитектуры параллельной потоковой вычислительной системы (ППВС) «Буря» и ее отличия от классических потоковых вычислительных систем. Проведенные исследования позволяют надеяться, что предложенная модель вычислений станет в будущем основной парадигмой программирования для масштабных параллельных вычислений.

Ключевые слова: потоковая модель вычислений; парадигма раздачи; ассоциативная память; локализация вычислений

DOI: 10.14357/19922264150403

1 Введение

В области высокопроизводительных вычислений существует проблема, которая заключается в том, что с ростом возможности максимального распараллеливания, предоставляемой пользователям создателями современной аппаратуры, все сложнее становится загрузить постоянно увеличивающееся число функциональных устройств вычислительных ядер, оставаясь в рамках традиционного программирования по фон Нейману. Можно выделить две основные причины этой проблемы:

- (1) это представление об алгоритме как о линейной последовательности действий — нити управления. Нить при работе имеет всегда ровно одну точку управления, поэтому будем называть ее *одноточечной*.
- (2) представление о взаимодействии по линии «процессор—память» как о состоящем, по сути, только из двух операций: чтения и записи по адресу. При этом есть только одна воз-

можность гарантированно прочитать то, что записано, — выполнить чтение позже записи.

Когда нить одна, проблем нет, не считая необходимости в поддержке механизма кэширования. Ситуация кардинально меняется, когда таких нитей становится много.

Данные представления восходят еще к Тьюрингу. Только там вместо программной нити — конечный автомат, а вместо памяти — бесконечная в две стороны лента. Работа выполняется пошагово, т. е. последовательно.

Посмотрим, к чему это ведет в связи с задачей загрузить работой систему из большого числа элементов, способных работать параллельно.

По первому пункту можно представить параллельный алгоритм в виде семейства одноточечных нитей, которые «взаимодействуют» друг с другом — одни передают сообщения, другие ждут и принимают. Таким образом возникает модель SPMD с MPI⁶. В модели GPGPU⁷ — та же нить, только выполняемая над длинным вектором, или иногда много

¹Институт проблем проектирования в микроэлектронике Российской академии наук, zmejevndn@ippm.ru

²Институт проблем проектирования в микроэлектронике Российской академии наук, klimov@ippm.ru

³Институт проблем проектирования в микроэлектронике Российской академии наук, nick@ippm.ru

⁴Институт проблем проектирования в микроэлектронике Российской академии наук, oku@ippm.ru

⁵Институт проблем проектирования в микроэлектронике Российской академии наук, ippm@ippm.ru

⁶SPMD — модель вычислений (Single-Program, Multiple-Data — одна программа, множество потоков данных) с протоколом передачи MPI (Message Passing Interface — интерфейс передачи сообщений).

⁷GPGPU — модель вычислений (general-purpose computation on graphics processing units — универсальные вычисления на графических процессорах).

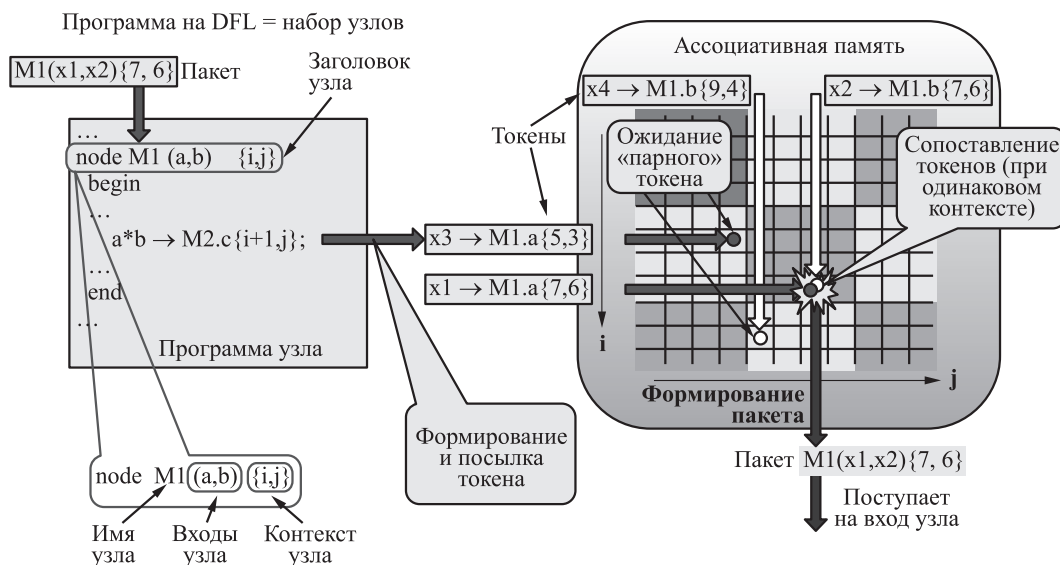


Рис. 1 Поточковая модель вычислений с динамически формируемым контекстом

нитей, по одной для каждого элемента, но уже не взаимодействующих. И как развитие возникает представление о вложенном параллелизме: одна нить распадается на многие, причем все они должны быть завершены, прежде чем родительская нить продолжит работу. Последнее означает наличие барьера, который может приводить к простоям.

Второй пункт в этом случае ставит следующие вопросы: как обеспечить прочтение другой нитью записанные данные и как узнать в другой нити, что данные записаны и их можно читать?

Авторы видят альтернативу в том, чтобы алгоритм сразу представлять как многоточечный, содержащий много точек управления. Одна активная операция (команда) может «передать управление» нескольким операциям, тогда как новая операция для начала своей работы может требовать нескольких передач управления ей. Вместе с «передачей управления» логично всегда передавать некоторый объем требуемых данных. Так возникает модель программирования dataflow со специфическим механизмом управления — по готовности входных данных. Для обеспечения многократности повторения однотипных действий, на чем стоит всякое программирование, вводим для всякой команды (или программы узла) *контекст*, он же адрес, или набор индексов. Передача управления на начало узла должна содержать кроме имени (номера) узла еще и необходимый набор индексов, определяющих конкретный экземпляр узла-получателя.

2 Поточковая модель вычислений

Такая модель существует, это предлагаемая авторами поточковая модель вычислений с динамически формируемым контекстом (рис. 1), воплощенная в языке программирования DFL (dynamic fuzzy logic) [1, 2]. Программа на этом языке представляет собой набор описаний узлов, состоящих из заголовка узла и программного кода. Заголовок содержит имя программного узла, список входов и список атрибутов программы узла — *контекст*.

Активация узла происходит тогда и только тогда, когда на все входы одного узла с определенным именем и контекстом придут все требуемые элементы данных — токены¹. Это является главным принципом управления потоком данных. При активации выполняется код программы узла, в котором вычисляются новые данные (исключительно на основе значений входов и атрибутов ключа) и посылаются специальными операторами на другие узлы, причем атрибуты ключа адресата вычисляются непосредственно в этом же коде. Это означает, что работа производится в парадигме «раздачи» [3].

Как видится авторам, корень проблемы фоннеймановского программирования заключается в том, что в ней реализуется *парадигма «сбора»*: для нее характерно, что только потребитель данных знает, какие данные ему нужны и где их взять, и сам их запрашивает, указывая имя переменной (рис. 2, а). В этих условиях аппаратуре трудно предвидеть, какие данные будут нужны. Для более

¹Токен — это структура данных, в состав которой входит передаваемый операнд (данное), ключ с контекстом и адресом программного узла, маска и ряд других специальных признаков.

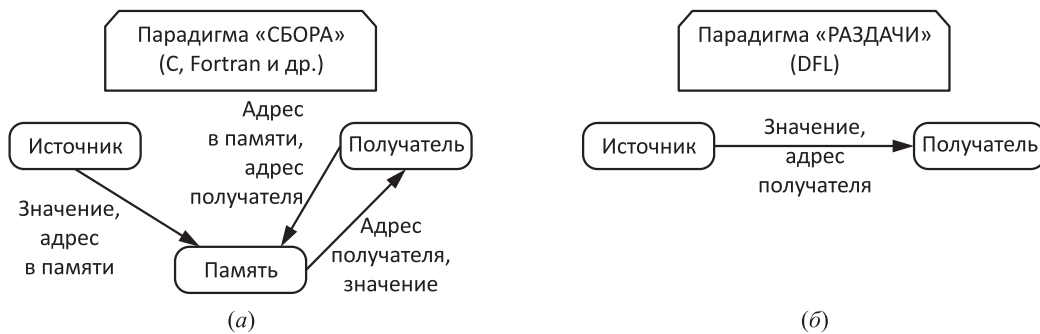


Рис. 2 Парадигмы «сбора» (а) и «раздачи» (б)

качественной стратегии перемещения данных была бы более продуктивной противоположная традиционная парадигма «раздачи», когда производитель каждого нового значения знает, кому оно требуется, и самостоятельно обеспечивает рассылку по нужным адресам. А получателю тогда остается просто «пассивно дожидаться» прихода данных, ничего не зная об их источнике (рис. 2, б).

Парадигма «раздачи» является более экономной в плане числа обменов сообщениями: одно сообщение на каждое использование, — тогда как в парадигме «сбора» одно сообщение на запись и еще по два на каждое использование (запрос и ответ).

Наиболее эффективным образом это реализуется через ассоциативную память¹. Заметим, что в этом случае обычная память не нужна. Всякое значение, необходимое для совершения операции узла, просто должно быть отправлено ему на один из входов другим каким-то узлом. С другой стороны, когда на одни входы данные уже пришли, то до прихода остальных они должны где-то находиться в ожидании. Это и есть память, только теперь — ассоциативная, в ней «адресом» является имя/номер узла с набором его индексов. Этот «адрес» называется *ключом*.

Кроме хранения, в памяти также происходит сопоставление (сравнение) ключей токенов, необходимое для формирования групп токенов, направленных на один и тот же узел (с одинаковыми именем и атрибутами ключа) [4].

Теперь интерфейс между процессором — исполнителем (многих) готовых активностей — и этой «памятью» выглядит сложнее: процессор посылает в память данные с указанием ключа и номера-имени входа, а «память» добавляет их к накапливаемым на входах узлов и, если возникает полный комплект, отправляет в исполнительное устройство задание на выполнение операции узла. Эта операция может представлять собой небольшую программу в обычной парадигме, которая с переданными ей

данными отработает от начала до конца, без каких-либо ожиданий, и, возможно, породит посылки новых данных в другие узлы, одновременно стирая (по умолчанию) использованные данные со своих входов.

Таким образом решается много проблем: программа узла всегда имеет нужные ей данные локально, при этом программа не прерывается на подкачку дополнительных данных: активность возникает только тогда, когда все нужные для выполнения программы узла данные имеются в наличии. Затем активность исчезает, передав потенциалы активностей в виде результирующих данных другим узлам, причем ответа ждать не надо: все такие сообщения (передачи токенов) односторонние.

3 Сравнение параллельной потоковой вычислительной системы «Буран» с классическими dataflow-системами

Классические потоковые системы (dataflow) [5], которые разрабатывались в 1980-х гг. с целью преодоления проблем фон-неймановской модели вычислений, представляли собой по большей части вычислительные системы для выполнения традиционных (обычных) программ, но в новой модели вычислений. С этой же позиции создавались и новые высокоуровневые языки программирования с однократным присваиванием. Работы в этой области были свернуты потому, что в те времена еще не стояли так остро проблемы распараллеливания и был недостаточно высок технологический уровень развития вычислительной техники. В настоящее время проблема создания параллельных программ с высокой степенью масштабируемости

¹Ассоциативная память — память с выборкой по содержанию.

Отличия ППВС «Буран» от классических dataflow-систем

Классические dataflow-системы	Модель вычислений ППВС «Буран»
Языки высокого уровня Sisal, Id, VAL и др. «работают» в парадигме «сбора»	Новая парадигма программирования — парадигма «раздачи», выраженная в языке DFL
Заранее (статически) создается граф выполнения программы	Граф выполнения программы создается (разворачивается) в процессе вычислений
Процессора сопоставления не было (примитивные команды)	Имеется процессор сопоставления с развитой системой команд
Не были решены проблемы управления локализацией и планирования вычислений	Развитые (удобные) средства управления локализацией и планированием вычислений
Были трудности с реализацией ассоциативной памяти ключей и токенов большого объема	Благодаря реализации функций управления локализацией и планированием вычислений удалось избежать построения физической памяти большого объема (иерархия памяти)
Не была реализована поддержка динамического формирования контекста	Реализована аппаратная поддержка динамического формирования контекста и функций взаимодействия (формирование токенов и пакетов)
Целевые узлы определялись в основном статически (были только элементы динамики)	Динамический выбор целевых узлов в узлах-источниках

выходит на первый план при наличии многоядерных (многопроцессорных) систем огромной мощности и интеграции вычислительной аппаратуры.

В основе системы, которая разрабатывается в ИППМ РАН (ППВС «Буран») лежит, прежде всего, совершенно новая модель вычислений, для которой создаются новые алгоритмы в рамках нового языка DFL и новой парадигмы программирования.

Еще одним ключевым отличием от классических dataflow-систем (см. таблицу) является аппаратно поддерживаемая работа с динамически формируемым контекстом, что дает возможность программисту реализовывать новые принципы организации параллельных вычислительных процессов, повысить эффективность программирования, использовать контекст для распределения вычислений, а также использовать все поля контекста для программирования. В классических потоковых системах работа с контекстом была ограничена и не давала нужной степени свободы. В ППВС [6–8] контекст полностью доступен программисту.

Существенной проблемой классических dataflow-систем была локализация вычислений. В ППВС она решается при помощи созданных функций распределения вычислений, которые в динамике обеспечивают равномерность распределения вычислений по физическим вычислительным ядрам системы, а также минимизацию обменов между вычислительными ядрами и тем самым повышают масштабируемость вычислительной системы и рост ее производительности [9].

Планирование вычислений, которое тесно связано с построением иерархии памяти сопоставления, являлось еще одной проблемой классических потоковых систем. В ППВС она решается благодаря применению функций распределения для этапов¹ (локализация во времени), которые в динамике разбивают вычислительный процесс на части с помощью небольшой настройки программистом [10].

Кроме этого удалось избежать построения сопоставляющей памяти большого объема для хранения ключей и выстроена глобальная иерархия памяти.

В отличие от предыдущих классических потоковых систем в ППВС реализуется развитая система команд процессора сопоставления², который аппаратно поддерживает функции языка высокого уровня.

4 Особенности архитектуры параллельной потоковой вычислительной системы «Буран»

Можно отметить, что в архитектуре исполняющей системы ППВС «Буран» отчетливо проявляются следующие современные тенденции развития в области высокопроизводительных вычислений:

Многопоточность. Каждый узел — это отдельный поток, и связь между узлами осуществляется

¹Этапом называется группа токенов, которая необходима для выполнения определенного интервала вычислений. Обычно этап связан с более или менее независимым действием, например итерация в программе.

²Процессор сопоставления — основной узел вычислительного ядра ППВС «Буран», обеспечивающий эффективную работу ассоциативной памяти ключей, синхронизацию вычислительных процессов и ряд других функций.

только через передачу входных данных, причем потоки не приостанавливаются, а только завершаются. На одном многопоточном исполнительном устройстве могут одновременно выполняться несколько активаций узлов, которые никак не взаимодействуют между собой.

Мелкозернистость. Обычно в современных вычислительных системах стремятся увеличить размер гранулы, поскольку накладные расходы на запуск треда (потока), прием сообщения, приостановку треда и т. п. высоки. В предлагаемой системе эти накладные расходы минимизируются (перекладываются на аппаратуру).

Активные сообщения. Когда значения на других входах узла уже есть, токен, несущий входное значение, активирует вычисление. Поскольку любой токен может оказаться последним, каждый токен в предлагаемой системе может считаться активным сообщением.

Процессор в памяти. Токен несет адрес и данные, и он может рассматриваться как обращение к памяти, сопровождаемое запросом на обработку:

- при поступлении последнего значения на входы узла с одним и тем же контекстом (адресом) происходит активация нужной программы узла;
- отсутствует проблема ожидания завершения записи данных в память;
- один из входов может принять адрес получателя результата;
- возможна непосредственная обработка на входе типа редукции.

Память сопоставления (ассоциативная). Обеспечивается функциональность, аналогичная механизму работы памяти с использованием full-empty bits.

Разделение доступа к данным и обработки данных (decoupled access-execute). В предлагаемой системе такая работа заложена непосредственно в самой модели вычислений.

5 Заключение

Решение проблем, которые возникают при создании суперкомпьютеров с огромными возможностями по распараллеливанию вычислений, но работающих в рамках традиционного фон-неймановского программирования, безусловно, необходимо. В статье в общих чертах показана альтернатива современным тенденциям развития вычислительной

техники, а именно: реализация высокопроизводительных систем на базе нетрадиционной потоковой модели вычислений с динамически формируемым контекстом. Описаны ключевые отличия разрабатываемой ППВС «Буран» от традиционных вычислительных систем и классических потоковых (dataflow) систем.

Показано, как проявляются в предложенной модели вычислений современные тенденции развития в области высокопроизводительных вычислений.

Проведенные эксперименты [11–15] на программных моделях ППВС «Буран» (разной степени абстракции) [16–18] позволяют надеяться на то, что предлагаемая модель вычислений с ее уникальными особенностями и преимуществами действительно станет в будущем основной реализуемой парадигмой программирования в области масштабных параллельных вычислений, а разрабатываемая аппаратура ППВС «Буран» для этой модели вычислений позволит в дальнейшем создавать отечественные вычислительные системы общего и специального назначения на базе перспективных архитектур.

Литература

1. Климов А. В., Левченко Н. Н., Окунев А. С. Преимущества потоковой модели вычислений в условиях неоднородных сетей // Информационные технологии и вычислительные системы, 2012. № 2. С. 36–45.
2. Левченко Н. Н., Окунев А. С., Стемпковский А. Л. Использование модели вычислений с управлением потоком данных и реализующей ее архитектуры для систем экзафлопсного уровня производительности // Проблемы разработки перспективных микро- и наноэлектронных систем (МЭС-2012): Сб. тр. V Всероссий. научн.-технич. конф. — М.: ИППМ РАН, 2012. С. 459–462.
3. Климов А. В., Левченко Н. Н., Окунев А. С., Стемпковский А. Л. Суперкомпьютеры, иерархия памяти и потоковая модель вычислений // Программные системы: теория и приложения. Электронный научный журнал, 2014. Т. 5. № 1(19). С. 15–36. http://psta.psiras.ru/read/psta2014_1_15-36.pdf.
4. Стемпковский А. Л., Левченко Н. Н., Окунев А. С., Цветков В. В. Параллельная потоковая вычислительная система — дальнейшее развитие архитектуры и структурной организации вычислительной системы с автоматическим распределением ресурсов // Информационные технологии, 2008. № 10. С. 2–7.
5. Silc J., Robic B., Ungerer T. Asynchrony in parallel computing: From dataflow to multithreading // Parallel Distributed Computing Practices, 1998. Vol. 1. No. 1. P. 3–30.
6. Стемпковский А. Л., Климов А. В., Левченко Н. Н., Окунев А. С. Методы адаптации параллельной потоковой вычислительной системы под задачи отдельных

- классов // Информационные технологии и вычислительные системы, 2009. № 3. С. 12–21.
7. *Стемпковский А. Л., Левченко Н. Н., Окунев А. С.* Архитектура сверхпетафлопной вычислительной системы с высокой реальной производительностью, базирующейся на нетрадиционной модели вычислений // Результаты целевых ориентированных фундаментальных исследований и их использование в российской промышленности: Мат-лы научн. конф. — Таганрог: ТТИ ЮФУ, 2010. С. 68–72.
 8. *Климов А. В., Левченко Н. Н., Окунев А. С., Стемпковский А. Л.* Автоматическое распараллеливание для гибридной системы с потоковым ускорителем // Информационные технологии и вычислительные системы, 2011. № 2. С. 3–11.
 9. *Змеев Д. Н., Левченко Н. Н., Окунев А. С., Климов А. В.* Способы регулирования вычислений в параллельной потоковой вычислительной системе // Проблемы разработки перспективных микро- и наноэлектронных систем-2014: Сб. тр. / Под общ. ред. академика РАН А. Л. Стемпковского. — М.: ИППМ РАН, 2014. Ч. IV. С. 79–82.
 10. *Змеев Д. Н., Левченко Н. Н., Окунев А. С., Климов А. В.* Архитектура планировщика процессора сопоставления ППВС «Буран» // Проблемы разработки перспективных микро- и наноэлектронных систем-2014: Сб. тр. / Под общ. ред. академика РАН А. Л. Стемпковского. — М.: ИППМ РАН, 2014. Ч. IV. С. 75–78.
 11. *Левченко Н. Н., Окунев А. С., Климов А. В.* Реализация задачи «Повышение контрастности изображения» на параллельной потоковой вычислительной системе // Многопроцессорные вычислительные и управляющие системы (МВУС-2009): Мат-лы Междунар. научн.-технич. конф. — Дивноморское: ТТИ ЮФУ, 2009. Т. 1. С. 197–200.
 12. *Левченко Н. Н., Окунев А. С.* Анализ прохождения задачи «обнаружение дефектов» на параллельной потоковой вычислительной системе // Методы и средства обработки информации: Тр. III Всеросс. научн. конф. — М.: МАКС Пресс, 2009. С. 346–350.
 13. *Левченко Н. Н., Окунев А. С.* Специализация архитектуры многоядерной параллельной потоковой вычислительной системы для решения задачи быстрого преобразования Фурье // Проблемы разработки перспективных микро- и наноэлектронных систем (МЭС-2010): Сб. научн. тр. IV Всеросс. научн.-технич. конф. — М.: ИППМ РАН, 2010. С. 458–461.
 14. *Климов А. В., Левченко Н. Н., Окунев А. С., Змеев Д. Н., Стемпковский А. Л.* Исследование возможности асинхронной реализации задачи молекулярной динамики на ППВС «Буран» // Качество. Инновации. Образование, 2014. № 10. С. 46–51.
 15. *Змеев Д. Н., Климов А. В., Левченко Н. Н., Окунев А. С.* Применение параллельной потоковой вычислительной системы «Буран» для задач молекулярной динамики // Суперкомпьютерные технологии (СКТ-2014): Мат-лы III Всеросс. научн.-технич. конф. — Ростов-на-Дону: ЮФУ, 2014. Т. 2. С. 222–224.
 16. *Левченко Н. Н., Окунев А. С., Змеев Д. Н.* Расширение возможностей поведенческой блочно-регистровой модели параллельной потоковой вычислительной системы // Высокопроизводительные вычислительные системы (ВПВС-2008): Мат-лы V Междунар. молодежной научн.-технич. конф. — Таганрог: ТТИ ЮФУ, 2008. С. 371–374.
 17. *Змеев Д. Н., Окунев А. С., Левченко Н. Н., Климов А. В.* Реализация параллельной модели вычислений с управлением потоком данных на кластерных суперкомпьютерах // Научный сервис в сети Интернет: все грани параллелизма: Тр. Междунар. суперкомпьютерной конф. — М.: МГУ, 2013. С. 375–377.
 18. *Змеев Д. Н., Левченко Н. Н., Окунев А. С.* Исследование принципов работы параллельной потоковой вычислительной системы на кластерных вычислительных системах // Суперкомпьютерные технологии (СКТ-2014): Мат-лы III Всеросс. научн.-технич. конф. — Ростов-на-Дону: ЮФУ, 2014. Т. 1. С. 45–48.

Поступила в редакцию 21.08.15

DATAFLOW COMPUTING MODEL AS A PARADIGM OF FUTURE MAINSTREAM OF SOFTWARE DEVELOPMENT

D. N. Zmejey, A. V. Klimov, N. N. Levchenko, A. S. Okunev, and A. L. Stempkovsky

Institute for Design Problems in Microelectronics, Russian Academy of Sciences, 3 Sovetskaya Str., Moscow 124365, Russian Federation

Abstract: The article describes the parallel dataflow computing system with dynamic forming of context and the architectural features of its implementation. The dataflow computing model allows to solve the problems that arise when creating and using supercomputers. One of these problems is the difficulty of loading the rising number of functional units of cores, while remaining in the borders of traditional programming. The article describes

the advantages of the proposed computing model. The article compares the paradigm of “distribution” with the traditional paradigm of “collection.” The dataflow computing model works in the paradigm of “distribution.” The description of basic features of the architecture of the “Buran” parallel dataflow computing system and its differences from the classical flow computing systems are given. The studies give hope that the proposed computing model will become in the future the main programming paradigm for large-scale parallel computing.

Keywords: dataflow computing system; paradigm of distribution; content-addressable memory; localization of computation

DOI: 10.14357/19922264150403

References

1. Klimov, A. V., N. N. Levchenko, and A. S. Okunev. 2012. Preimushchestva potokovoy modeli vychisleniy v usloviyakh neodnorodnykh setey [The advantages of dataflow calculation model in heterogeneous networks conditions]. *Zh. Informatsionnye Tekhnologii i Vychislitel'nye Sistemy* [J. Information Technologies Computing Systems] 2:36–45.
2. Levchenko, N. N., A. S. Okunev, and A. L. Stempkovskiy. 2012. Ispol'zovanie modeli vychisleniy s upravleniem potokom dannykh i realizuyushchey ee arkhitektury dlya sistem eksaflopsnogo urovnya proizvoditel'nosti [Dataflow calculation model and its architecture usage for exaflop's computing systems]. *Vseross. nauch.-tekhnich. konf. "Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem" (MES-2012): Sb. tr.* [5th All-Russia Scientific and Technical Conference “Problems of Advanced Micro- and Nanoelectronic Systems Development” (MES-2012) Proceedings]. Moscow. 459–462.
3. Klimov, A. V., N. N. Levchenko, A. S. Okunev, and A. L. Stempkovskiy. 2014. Superkomp'yutery, ierarkhiya pamyati i potokovaya model' vychisleniy [Supercomputers, memory hierarchy, and the dataflow computing system]. *Programmnye Sistemy: Teoriya i Prilozheniya. Elektron. Nauchn. Zh.* [Program Systems: Theory and Applications. Electron. Sci. J.] 5:1(19):15–36.
4. Stempkovskiy, A. L., N. N. Levchenko, A. S. Okunev, and V. V. Tsvetkov. 2008. Parallel'naya potokovaya vychislitel'naya sistema — dal'neyshee razvitie arkhitektury i strukturnoy organizatsii vychislitel'noy sistemy s avtomaticheskim raspredeleniem resursov [Parallel dataflow computing system: Further development of architecture and structural organization of the computing system with automatic distribution of resources]. *Zh. Informatsionnye Tekhnologii* [J. Inform. Technol.] 10:2–7.
5. Silc, J., B. Robic, and T. Ungerer. 1998. Asynchrony in parallel computing: From dataflow to multithreading. *Parallel Distributed Computing Practices* 1(1):3–30.
6. Stempkovskiy, A. L., A. V. Klimov, N. N. Levchenko, and A. S. Okunev. 2009. Metody adaptatsii parallel'noy potokovoy vychislitel'noy sistemy pod zadachi otdel'nykh klassov [Methods of parallel dataflow computing system adaptation for problems of individual classes]. *Zh. Informatsionnye Tekhnologii i Vychislitel'nye Sistemy* [J. Inform. Technol. Comput. Syst.] 3:12–21.
7. Stempkovskiy, A. L., N. N. Levchenko, and A. S. Okunev. 2010. Arkhitektura sverkhpetaflopnoy vychislitel'noy sistemy s vysokoy real'noy proizvoditel'nost'yu, bazi-ruyushchey na netraditsionnoy modeli vychisleniy [Abovepetaflop's computer systems architecture with high real performance, based on nontraditional computing model]. *Mat-ly nauchn. konf. "Rezultaty Tselevykh Orientirovannykh Fundamental'nykh Issledovaniy i Ikh Ispol'zovanie v Rossiyskoy Promyshlennosti"* [Annual Scientific Conference “Results of the Target-Oriented Fundamental Researches and Their Usage in Russian Industry” Proceedings]. Taganrog, Russia. 68–72.
8. Klimov, A. V., N. N. Levchenko, A. S. Okunev, and A. L. Stempkovskiy. 2011. Avtomaticheskoe rasparallelivanie dlya gibridnoy sistemy s potokovym uskoritelem [Automatic parallelization of sequential programs for hybrid systems with accelerator-based dataflow]. *Zh. Informatsionnye Tekhnologii i Vychislitel'nye Sistemy* [J. Inform. Technol. Comput. Syst.] 2:3–11.
9. Zmeev, D. N., N. N. Levchenko, A. S. Okunev, and A. V. Klimov. 2014. Sposoby regulirovaniya vychisleniy v parallel'noy potokovoy vychislitel'noy sisteme [Methods of control of computations in the parallel dataflow computing system]. *Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem-2014: Sb. tr.* [All-Russia Scientific and Technical Conference “Problems of Advanced Micro- and Nanoelectronic Systems Development-2014” Proceedings]. Ed. A. L. Stempkovskiy. Moscow. IV:79–82.
10. Zmeev, D. N., N. N. Levchenko, A. S. Okunev, and A. V. Klimov. 2014. Arkhitektura planirovshchika protsessora sopostavleniya PPVS “Buran” [Architecture of the scheduler of mapping processor of PDCS “Buran”]. *Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem-2014. Sb. tr.* [All-Russia Scientific and Technical Conference “Problems of Advanced Micro- and Nanoelectronic Systems Development-2014” Proceedings]. Ed. A. L. Stempkovskiy. Moscow. IV:75–78.
11. Levchenko, N. N., A. S. Okunev, and A. V. Klimov. 2009. Realizatsiya zadachi “Povyshenie kontrastnosti izobrazheniya” na parallel'noy potokovoy vychislitel'noy sisteme [Realization of the task of “Improving the contrast of the image” on the parallel dataflow computing system]. *Scientific and Technical Conference (International) “Multiprocessor Computing and Control Systems (MCCS-2009)” Proceedings*. Divnomorskoe, Russia. 1:197–200.
12. Levchenko, N. N., and A. S. Okunev. 2009. Analiz prokhozheniya zadachi “obnaruzhenie defektov” na parallel'noy potokovoy vychislitel'noy sisteme [Analysis of the running of defects detection tasks on the parallel dataflow computing system]. *Tr. III Vseross. nauchn. konf. Metody i Sredstva Obrabotki Informatsii* [3rd All-Russia

- Scientific Conference “Methods and Tools of Information Processing” Proceedings]. Moscow. 346–350.
13. Levchenko, N. N., and A. S. Okunev. 2010. Spetsializatsiya arkhitektury mnogoyadernoy parallel'noy potokovoy vychislitel'noy sistemy dlya resheniya zadachi bystrogo preobrazovaniya Fur'e [Specialization of the architecture of multicore parallel dataflow computing system to compute the task of Fast Fourier Transform]. *Sb. nauch. tr. IV Vseross. nauchn.-tekhnich. konf. “Problemy Razrabotki Perspektivnykh Mikro- i Nanoelektronnykh Sistem-2010* [4th All-Russia Scientific and Technical Conference “Problems of Advanced Micro- and Nanoelectronic Systems Development-2010” Proceedings]. Moscow. 458–461.
 14. Klimov, A. V., N. N. Levchenko, A. S. Okunev, D. N. Zmeev, and A. L. Stempkovskiy. 2014. Issledovanie vozmozhnosti asinkhronnoy realizatsii zadachi molekulyarnoy dinamiki na PPVS “Buran” [The research of the asynchronous implementation of molecular dynamics on PDCS “Buran”]. *Zh. Kachestvo. Innovatsii. Obrazovanie* [J. Quality. Innovation. Education] 10:46–51.
 15. Zmeev, D. N., A. V. Klimov, N. N. Levchenko, and A. S. Okunev. 2014. Primenenie parallel'noy potokovoy vychislitel'noy sistemy “Buran” dlya zadach molekulyarnoy dinamiki [Application of the parallel dataflow computing system “Buran” for molecular dynamics]. *Mat-ly 3-y Vseross. nauchn.-tekhnich. konf. “Superkomp'yuternye Tekhnologii” (SKT-2014)* [3rd All-Russia Scientific and Technical Conference “Supercomputer Technologies” (SCT-2014) Proceedings]. Moscow. 371–374.
 16. Levchenko, N. N., A. S. Okunev, and D. N. Zmeev. 2008. Rasshirenie vozmozhnostey povedencheskoy blochno-registrovoy modeli parallel'noy potokovoy vychislitel'noy sistemy [Enhancement of capability of the behavioral block-rated model of the parallel dataflow computing system]. *5th Youth Research and Technical Conference (International) “High-Performance Computing Systems (HPCS 2008)” Proceedings*. Taganrog, Russia. 371–374.
 17. Zmeev, D. N., A. S. Okunev, N. N. Levchenko, and A. V. Klimov. 2013. Realizatsiya parallel'noy modeli vychisleniy s upravleniem potokom dannykh na klasternykh superkomp'yuterakh [The implementation of parallel computing model with dataflow control management on cluster supercomputers]. *Supercomuter Conference (International) “Scientific Service in the Internet: All Sides of Parallelism” Proceedings*. Moscow: MSU. 375–377.
 18. Zmeev, D. N., N. N. Levchenko, and A. S. Okunev. 2014. Issledovanie printsipov raboty parallel'noy potokovoy vychislitel'noy sistemy na klasternykh vychislitel'nykh sistemakh [Study of the principles of parallel dataflow computing system on the cluster computing systems]. *Mat-ly 3-y Vseross. nauchn.-tekhnich. konf. “Superkomp'yuternye Tekhnologii” (SKT-2014)* [3rd All-Russia Scientific and Technical Conference “Supercomputer Technologies” (SCT-2014) Proceedings]. 1:45–48.

Received August 21, 2015

Contributors

Zmejev Dmitry N. (b. 1978) — scientist, Institute for Design Problems in Microelectronics, Russian Academy of Sciences, 3 Sovetskaya Str., Moscow 124365, Russian Federation; zmejevdn@ippm.ru

Klimov Arkady V. (b. 1953) — senior scientist, Institute for Design Problems in Microelectronics, Russian Academy of Sciences, 3 Sovetskaya Str., Moscow 124365, Russian Federation; klimov@ippm.ru

Levchenko Nikolay N. (b. 1978) — Candidate of Science (PhD) in technology, Head of Department, Institute for Design Problems in Microelectronics, Russian Academy of Sciences, 3 Sovetskaya Str., Moscow 124365, Russian Federation; nick@ippm.ru

Okunev Anatoly S. (b. 1951) — Candidate of Science (PhD) in technology, leading scientist, Institute for Design Problems in Microelectronics, Russian Academy of Sciences, 3 Sovetskaya Str., Moscow 124365, Russian Federation; oku@ippm.ru

Stempkovsky Alexander L. (b. 1950) — Doctor of Science in technology, professor, Academician of the Russian Academy of Sciences, Director, Institute for Design Problems in Microelectronics, Russian Academy of Sciences, 3 Sovetskaya Str., Moscow 124365, Russian Federation; ippm@ippm.ru

АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ В ДИНАМИЧЕСКИХ СИСТЕМАХ С ЦИЛИНДРИЧЕСКИМИ БЕССЕЛЕВЫМИ НЕЛИНЕЙНОСТЯМИ*

И. Н. Сеницын¹

Аннотация: Рассмотрены методы аналитического моделирования (МAM) процессов в динамических системах со сложными бесселевыми нелинейностями (БН) при гармонических и стохастических, узкополосных и широкополосных возмущениях. Даны необходимые сведения из теории бесселевых функций и сложных БН. Приведено методическое и алгоритмическое обеспечение МAM на основе методов статистической линеаризации (МСЛ) и нормальной аппроксимации (МНА) для стохастических широкополосных процессов типа белого шума. Рассмотрены особенности МAM для гармонических и узкополосных стохастических процессов (СтП). В приложении приведены формулы для коэффициентов МСЛ для типовых БН. В качестве тестовых примеров рассмотрены задачи МAM процессов в одномерных системах с аддитивными и мультипликативными белыми шумами. Особое внимание уделено процессам в бесселевом осцилляторе в условиях различных возмущений. Заключение содержит основные выводы и некоторые обобщения.

Ключевые слова: бесселева нелинейность (БН); гармонический процесс; метод аналитического моделирования (МAM); метод нормальной аппроксимации (МНА); метод статистической линеаризации (МСЛ); осциллятор Бесселя; сложные бесселевы нелинейности (СБН); стохастический нормальный гауссовский процесс; стохастические системы на многообразиях (МСтС); узкополосный стохастический процесс; фазовый портрет; формула Гиббса; функции Куммера

DOI: 10.14357/19922264150404

1 Введение

В [1–4] рассмотрены вопросы аналитического моделирования процессов в дифференциальных стохастических системах (СтС) на многообразиях (МСтС) с винеровскими и пуассоновскими шумами, а также со сложными трансцендентными нелинейностями (СТН). Такие модели описывают поведение многих современных нано- и квантовооптических технических средств информатики и управления. Приведены уравнения МНА и МСЛ для моделирования нестационарных и стационарных нормальных процессов в МСтС. Рассмотрены методы вычисления типовых интегралов для одно- и многомерных ТН скалярного и векторного аргумента, получающихся из суперпозиции типовых ТН. Обсуждается алгоритмическое обеспечение МAM.

Продолжим исследования [1–4] применительно к МСтС с СТН в виде (цилиндрических) БН. В разд. 2 приводятся необходимые сведения из теории бесселевых функций первого рода и целого порядка. В разд. 3 излагается МСЛ для МСтС с БН,

а соответствующий МAM на основе МНА и МСЛ дается в разд. 4. В разд. 5 рассматриваются особенности МAM для гармонических и узкополосных стохастических процессов. Приложение 1 содержит формулы для коэффициентов МСЛ для типовых БН. Приложение 2 содержит тестовые примеры. Особое внимание уделено моделированию процессов в осцилляторе Бесселя в условиях различных возмущений.

2 Цилиндрические функции Бесселя и бесселевы нелинейности

Многие сложные трансцендентные нелинейные преобразования вида

$$z = \varphi^{\text{СТН}}(y)$$

выражаются через функции Бесселя $J_n(y)$ первого рода порядка n ($n = 0, 1, 2, \dots$). Функции Бесселя,

* Работа выполнена при поддержке РФФИ (проект 15-07-02244).

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

порядок которых равен половине нечетного целого числа, выражаются через элементарные функции [5–7].

Как известно [5–7], функции $J_n(y)$ целого порядка n удобно определить следующим бесконечным рядом:

$$J_n(y) = \left(\frac{y}{2}\right)^n \sum_{k=0}^{\infty} \frac{1}{k!(n+k+1)!} \left(-\frac{y^2}{4}\right)^k,$$

причем $J_{-n} = (-1)^n J_n(y)$.

Функции $J_n = J_n(y)$ являются решением следующего линейного дифференциального уравнения с переменными коэффициентами:

$$y^2 \frac{d^2 J_n}{dy^2} + y \frac{dJ_n}{dy} + (y^2 - n^2) J_n = 0,$$

допускают интегральные представления вида

$$\begin{aligned} J_n(y) &= \frac{(y/2)^n}{\sqrt{\pi}\Gamma(n+1/2)} \int_0^\pi \cos(y \cos \theta) \sin^{2n} \theta d\theta = \\ &= \frac{2(y/2)^n}{\sqrt{\pi}\Gamma(n+1/2)} \int_0^1 (1-t^2)^{n-1/2} \cos yt dt \\ &\quad \left(\operatorname{Re} n > -\frac{1}{2}\right), \end{aligned}$$

а также удовлетворяют рекуррентным соотношениям

$$\left. \begin{aligned} J_{n-1}(y) + J_{n+1}(y) &= \frac{2n}{y} J_n(y); \\ J_{n-1}(y) - J_{n+1}(y) &= 2J'_n(y); \\ J'_n(y) &= J_{n-1}(y) - \frac{n}{y} J_n(y); \\ J'_n(y) &= -J_{n+1}(y) + \frac{n}{y} J_n(y), \end{aligned} \right\} (1)$$

где $J'_n(y) = dJ_n(y)/dy$; $\Gamma(n)$ — гамма-функция. При этом верхние границы для функции $J_n(y)$ определяются следующими неравенствами:

$$\left. \begin{aligned} |J_n(y)| &< 1 \quad (\forall n \geq 0); \\ |J_n(y)| &\leq \frac{1}{\sqrt{2}} \quad (n \geq 1); \\ 0 < J_n(y) &< \frac{2^{1/3}}{3^{2/3}\Gamma(2/3)n^{1/3}} \quad (n > 0); \\ |J_n(ny)| &\leq \left| y^n \exp \frac{n\sqrt{1-y^2}}{[1+\sqrt{1-y^2}]^n} \right|. \end{aligned} \right\} (2)$$

Нули $j_{n,s}$ и $j'_{n,s}$ для $J_n(y)$ и $J'_n(y) = dJ_n(y)/dy$ действительны, причем они разделяются для $y \neq 0$:

$$\left. \begin{aligned} 0 < j_{n,1} < j_{n+1,1} < j_{n,2} < j_{n+1,2} < \dots \quad (n > 0); \\ n \leq j'_{n,1} < j_{n,1} < j'_{n,2} < j_{n,2} < \dots \quad (n > 0); \\ j'_{0,1} &= 0; \quad j'_{0,s} = j_{1,s-1} \quad (s = 2, 3, \dots); \\ j_{n,s} &= \rho_n(s); \quad j'_{n,s} = \sigma_n(s-1), \end{aligned} \right\} (3)$$

где ρ_n и σ_n — нули $J_n(y)$ и $J'_n(y)$, $\rho_n(0) = 0$, $\sigma_n(0) = j'_{n,1}$.

Теоремы умножения и сложения для функции $J_n(y)$ записываются в виде:

$$J_n(\lambda y) = \lambda^n \sum_{k=0}^{\infty} \frac{(-1)^k (\lambda^2 - 1)^k (y/2)^k}{k!} J_{n+k}(y); \quad (4)$$

$$J_n(y_1 \pm y_2) = \sum_{k=-\infty}^{\infty} J_{n \mp k}(y_1) J_k(y_2).$$

Дифференцирование $J_n(y)$ проводится согласно следующим формулам:

$$J'_0(y) = -J_1(y); \quad yJ'_n(y) = nJ_n(y) - yJ_{n+1}(y);$$

$$\frac{2n}{y} J_n(y) = J_{n-1}(y) + J_{n+1}(y);$$

$$\frac{d}{dy} [y^n J_n(y)] = y^n J_{n-1}(y);$$

$$4J''_n(y) = 2J'_{n-1}(y) - 2J'_{n+1}(y) = J_{n-2}(y) - 2J_n(y) + J_{n+2}(y);$$

$$2^s J_n^{(s)}(y) = [J_{n-s}(y) - C_s^1 J_{n-s+2}(y) + C_s^2 J_{n-s+4}(y) - \dots - (-1)^s J_{n+s}(y)] \quad (s = 0, 1, 2, \dots).$$

В основе интегрирования функции $J_n(y)$ лежат следующие формулы:

$$\left. \begin{aligned} \int_0^y J_{2r}(\eta) d\eta &= \int_0^y J_0(\eta) d\eta - 2 \sum_{k=0}^{r-1} J_{2k+1}(y); \\ \int_0^y J_{2r+1}(\eta) d\eta &= 1 - J_0(y) - 2 \sum_{k=0}^r J_{2k}(y); \\ \int_0^y J_{n+1}(\eta) d\eta &= \int_0^y J_{n-1}(\eta) d\eta - 2J_n(y). \end{aligned} \right\} (5)$$

Замечание 1. Цилиндрической функцией Бесселя мнимого аргумента порядка n называется функция вида $I_n(y) = (i)^{-n} J_n(iy)$.

Замечание 2. Примерами СБН, получаемых посредством сумм типовых БН, могут служить следующие:

$$\varphi^{\text{СБН}}(Y, t) = \sum_{r=1}^n l_{rt} \varphi_r^{\text{БН}}(Y);$$

$$\varphi^{\text{СБН}}(Y, t) = \sum_{r=1}^n l_{rt} \varphi_r^{\text{БН}}(Y) \varphi_r^{\text{АН}}(Y),$$

а также дробно-рациональные представления:

$$\varphi^{\text{СБН}}(Y, t) = \frac{\sum_{r=1}^{n'} l'_{rt} \varphi_r^{\text{БН}}(Y)}{\sum_{r=1}^{n''} l''_{rt} \varphi_r^{\text{БН}}(Y)};$$

$$\varphi^{\text{СБН}}(Y, t) = \frac{\sum_{r=1}^{n'} l'_{rt} \varphi_r^{\text{БН}}(Y) \varphi_r^{\text{АН}}(Y)}{\sum_{r=1}^{n''} l''_{rt} \varphi_r^{\text{БН}}(Y) \varphi_r^{\text{АН}}(Y)},$$

где $\varphi_r^{\text{БН}}(Y)$, $\varphi_r^{\text{АН}}(Y)$ и $\varphi_r^{\text{АН}}(Y)$ — типовые бesselевы функции; l_{rt} , l'_{rt} и l''_{rt} — коэффициенты, зависящие от времени t ; $\varphi_r^{\text{АН}}(Y)$, $\varphi_r^{\text{АН}}(Y)$ и $\varphi_r^{\text{АН}}(Y)$ — алгебраические нелинейности (многочлены, степенные, иррациональные, дробно-рациональные и другие функции).

Другими примерами СБН являются нелинейности, получаемые путем соответствующего преобразования аргумента:

$$\varphi^{\text{СБН}}(Y, t) = \varphi^{\text{АН}}(\psi^{\text{БН}}(Y, t), t);$$

$$\varphi^{\text{СБН}}(Y, t) = \varphi^{\text{БН}}(\psi^{\text{АН}}(Y, t), t),$$

где $\varphi^{\text{АН}}$, $\psi^{\text{АН}}$, $\varphi^{\text{БН}}(Y, t)$ и $\psi^{\text{БН}}(Y, t)$ — типовые алгебраические и бesselевы трансцендентные нелинейности.

Замечание 3. В качестве примеров скалярных СБН векторного аргумента $Y = [Y_1 \dots Y_p]^T$ рассмотрим следующие:

$$\varphi^{\text{СБН}}(Y, t) = \sum_{r=1}^n \prod_{h=1}^H l_{rh,t} \varphi_{rh}^{\text{БН}}(Y_h);$$

$$\varphi^{\text{СБН}}(Y, t) = \frac{\sum_{r=1}^{n'} \prod_{h=1}^{H'} l'_{rh,t} \varphi_{rh}^{\text{БН}}(Y_h)}{\sum_{r=1}^{n''} \prod_{h=1}^{H''} l''_{rh,t} \varphi_{rh}^{\text{БН}}(Y_h)}.$$

В случае векторных и матричных СБН приведенные формулы имеют место для соответствующих компонент.

Функции $J_n(y)$ для неотрицательных $n = 0, 1, 2, \dots$ представляют собой однозначные целые функции y и определяются как коэффициенты следующих рядов Фурье:

$$\cos(y \sin \xi) = J_0(y) + 2 \sum_{k=1}^{\infty} J_{2k}(y) \cos 2k\xi;$$

$$\sin(y \sin \xi) = 2 \sum_{k=1}^{\infty} J_{2k-1}(y) \sin(2k-1)\xi.$$

Пусть ρ_i и ρ_k — два нуля функции Бесселя $J_n(y)$ целого порядка, тогда имеют место следующие условия ортогональности:

$$\int_0^1 J_n(\rho_i y) J_n(\rho_k y) dy = \begin{cases} 0 & \text{при } i \neq k; \\ \frac{1}{2} [J'_n(\rho_i)]^2 = \frac{1}{2} [J_{n+1}(\rho_i)]^2 & \text{при } i = k. \end{cases}$$

При достаточно общих условиях нелинейную функцию $\varphi^{\text{СБН}}(y)$ можно разложить в следующий ряд Фурье–Бесселя [5–7]:

$$\left. \begin{aligned} \varphi(y) &= \sum_{k=1}^{\infty} \alpha_k J_n(\rho_k y); \\ \alpha_k &= \frac{2}{[J_{n+1}(\rho_k)]^2} \int_0^1 \xi \varphi^{\text{СБН}}(\xi) J_n(\rho_k \xi) d\xi. \end{aligned} \right\} (6)$$

Здесь ρ_1, ρ_2, \dots — положительные нули функции $J_n(y)$, расположенные в порядке их возрастания, а α_k — коэффициенты разложения.

3 Статистическая линейзация бesselевых нелинейностей

Рассмотрим статистическую линейризацию БН по Казакову [8]:

$$Z = J_n(Y) \quad (7)$$

при несимметричном гауссовском стохастическом входном сигнале Y :

$$Y_t = Y(t) = m_y + Y_t^0, \quad (8)$$

где m_y — математическое ожидание, $Y_t^0 = Y^0(t) = Y(t) - m_y$. В соответствии с МСЛ зависимость аппроксимируется следующим выражением [8–10]:

$$Z_t = \varphi_0(m_y, D_y) + k_1(m_y, D_y) Y_t^0. \quad (9)$$

Здесь φ_0 и k_1 — коэффициенты статистической линейризации, зависящие от m_y , дисперсии D_y и определяемые по формулам:

$$\varphi_0 = \varphi_0(m_y, D_y) = \frac{1}{\sqrt{2\pi D_y}} \int_{-\infty}^{\infty} J_n(\eta) e^{-(\eta - m_y)^2 / 2D_y} d\eta; \quad (10)$$

$$k_1 = k_1(m_y, D_y) = \frac{1}{\sqrt{2\pi D_y}} \int_{-\infty}^{\infty} (\eta - m_y) J_n(\eta) e^{-(\eta - m_y)^2 / 2D_y} d\eta = \frac{\partial \varphi_0(m_y, D_y)}{\partial m_y}. \quad (11)$$

При $n = 2r + 1$ для функции $J_{2r+1}(Y)$ в силу ее нечетности имеем:

$$\varphi_0^{J_{2r+1}}(m_y, D_y) = k_0^{J_{2r+1}}(m_y, D_y)m_y; \quad (12)$$

$$k_0^{J_{2r+1}}(m_y, D_y) = \frac{1}{m_y \sqrt{2\pi D_y}} \int_{-\infty}^{\infty} J_{2r+1}(\eta) e^{-(\eta - m_y)^2 / 2D_y} d\eta. \quad (13)$$

Теорема 1. Если существуют вероятностные моменты второго порядка для уравнений (8) и (9), то в основе МСЛ для БН четного порядка ($n = 2r$) лежат уравнения (8)–(11), а для БН нечетного порядка ($n = 2r + 1$) – (9), (12), (13).

Коэффициенты статистической линейризации типовых БН приведены в приложении П1.

4 Аналитическое моделирование гауссовских процессов в стохастических системах с бесселевыми нелинейностями

Как известно [8–10], уравнения конечномерных непрерывных нелинейных систем со стохастическими возмущениями путем расширения вектора состояния СтС могут быть записаны в виде следующего векторного стохастического дифференциального уравнения Ито:

$$dY_t = a(Y_t, t) dt + b(Y_t, t) dW_0 + \int_{R_0} c(Y_t, t, v) P^0(dt, dv), \quad Y(t_0) = Y_0. \quad (14)$$

Здесь Y_t – $(p \times 1)$ -мерный вектор состояния, $Y_t \in \Delta_y$ (Δ_y – многообразие состояний); $a = a(Y_t, t)$ и $b = b(Y_t, t)$ – известные $(p \times 1)$ -мерная и $(p \times m)$ -мерная функции Y_t и t ; $W_0 = W_0(t)$ – $(r \times 1)$ -мерный винеровский СтП интенсивности $\nu_0 = \nu_0(t)$; $c(Y_t, t, v)$ – $(p \times 1)$ -мерная функция Y_t, t и вспомогательного $(q \times 1)$ -мерного параметра v ; $\int_{\Delta} dP^0(t, A)$ – центрированная пуассоновская мера, определяемая

$$\int_{\Delta} dP^0(t, A) = \int_{\Delta} dP(t, A) = \int_{\Delta} \nu_P(t, A) dt.$$

При этом принято: \int_{Δ} – число скачков пуассоновского СтП в интервале времени $\Delta = (t_1, t_2]$; $\nu_P(t, A)$ – интенсивность пуассоновского СтП $P(t, A)$; A – некоторое борелевское множество

пространства R_0^q с выколотым началом. Начальное значение Y_0 представляет собой случайную величину, не зависящую от приращений $W_0(t)$ и $P(t, A)$ на интервалах времени, следующих за $t_0, t_0 \leq t_1 \leq t_2$, для любого множества A . Предполагается, что элементы функций $a(y, t)$, $b(y, t)$ и $c(y, t, v)$: $a_i(y_t, t)$, $b_{ij}(y_t, t)$ и $c_i(y_t, t, v)$ – допускают разложения Фурье–Бесселя вида (6):

$$\begin{aligned} a_i(y_t, t) &= \sum_k \prod_l \alpha_{lk}^{a_i} J_n(\rho_k y_{kt}); \\ b_{ij}(y_t, t) &= \sum_k \prod_l \beta_{lk}^{b_{ij}} J_n(\rho_k y_{kt}); \\ c_i(y_t, t, v) &= \sum_k \prod_l \gamma_{lk}^{c_i} J_n(\rho_k y_{kt}). \end{aligned}$$

Замечание 4. Для аддитивных гауссовских (нормальных) и обобщенных пуассоновских возмущений уравнение (14) принимает вид [8–10]:

$$\dot{Y} = a(Y_t, t) + b_0(t)V, \quad V = \dot{W}, \quad Y(t_0) = Y_0. \quad (15)$$

Здесь W – СтП с независимыми приращениями, представляющий собой смесь нормального и обобщенного пуассоновского СтП.

Если предположить существование конечных вероятностных моментов второго порядка для моментов времени t_1 и t_2 , то уравнения МНА примут следующий вид [8–10]:

– для характеристических функций:

$$\left. \begin{aligned} g_1^N(\lambda; t) &= \exp \left[i\lambda^T m_t - \frac{1}{2} \lambda^T K_t \lambda \right]; \\ g_{t_1, t_2}^N(\lambda_1, \lambda_2; t_1, t_2) &= \\ &= \exp \left[i\bar{\lambda}^T \bar{m}_2 - \frac{1}{2} \bar{\lambda}^T \bar{K}_2 \lambda \right], \end{aligned} \right\} \quad (16)$$

где

$$\begin{aligned} \bar{\lambda} &= [\lambda_1^T \lambda_2^T]^T; \quad \bar{m}_2 = [m_{t_1}^T m_{t_2}^T]^T; \\ \bar{K}_2 &= \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) \\ K(t_2, t_1) & K(t_2, t_2) \end{bmatrix}; \end{aligned}$$

– для математических ожиданий m_t , ковариационной матрицы K_t и матрицы ковариационных функций $K(t_1, t_2)$:

$$\left. \begin{aligned} \dot{m}_t &= a_1(m_t, K_t, t), \quad m_0 = m(t_0); \\ \dot{K}_t &= a_2(m_t, K_t, t), \quad K_0 = K(t_0); \\ \frac{\partial K(t_1, t_2)}{\partial t_2} &= K(t_1, t_2) a_{21}(m_{t_2}, K_{t_2}, t_2)^T, \\ &K(t_1, t_1) = K_{t_1}. \end{aligned} \right\} \quad (17)$$

Здесь приняты следующие обозначения:

$$\begin{aligned}
 m_t &= M_{\Delta_y}^N Y_t; \quad Y_t^0 = Y_t - m_t; \\
 K_t &= M_{\Delta_y}^N Y_t^0 Y_t^{0T}; \quad K(t_1, t_2) = M_{\Delta_y}^N Y_{t_1}^0 Y_{t_2}^{0T}; \\
 a_1 &= a_1(m_t, K_t, t) = M_{\Delta_y}^N a(Y_t, t); \\
 a_2(m_t, K_t, t) &= a_{21}(m_t, K_t, t) + \\
 &\quad + a_{21}(m_t, K_t, t)^T + a_{22}(m_t, K_t, t); \\
 a_{21} &= a_{21}(m_t, K_t, t) = M_{\Delta_y}^N a(Y_t, t) Y_t^{0T}; \\
 a_{22} &= a_{22}(m_t, K_t, t) = M_{\Delta_y}^N \bar{\sigma}(Y_t, t); \\
 \sigma(Y_t, t) &= b(Y_t, t) \nu_0(t) b(Y_t, t)^T; \\
 \bar{\sigma}(Y_t, t) &= \\
 &= \sigma(Y_t, t) + \int_{R^q} c(Y_t, t, v) c(Y_t, t, v)^T \nu_P(t, dv), \quad (18)
 \end{aligned}$$

где $M_{\Delta_y}^N$ — символ вычисления математического ожидания для нормальных распределений (16).

Для стационарных СтС нормальные стационарные СтП — если они существуют, то $m_t = m^*$, $K_t = K^*$, $K(t_1, t_2) = k(\tau)$ ($\tau = t_1 - t_2$) — определяются уравнениями [8–10]:

$$\left. \begin{aligned}
 a_1(m^*, K^*) &= 0; \quad a_2(m^*, K^*) = 0; \\
 \dot{k}_\tau(\tau) &= a_{21}(m^*, K^*) K^{*-1} k(\tau); \\
 k(0) &= K^* \quad (\forall \tau > 0); \\
 k(\tau) &= k(-\tau)^T \quad (\forall \tau < 0).
 \end{aligned} \right\} \quad (19)$$

При этом необходимо, чтобы матрица $a_{21}(m^*, K^*) = a_{21}^*$ была асимптотически устойчивой.

Уравнения МНА в случае СтС (15) переходят в уравнения МСЛ [8–10], если принять

$$\left. \begin{aligned}
 a(Y_t, t) &= a_0(m_t, K_t) + k_1^a(m_t, K_t) Y_t^0; \\
 k_1^a(m_t, K_t, t) &= \left[\left(\frac{\partial}{\partial m_t} \right) a_0(m_t, K_t, t)^T \right]^T; \\
 b(Y_t, t) &= b_0(t); \\
 \sigma(Y_t, t) &= b_0(t) \nu(t) b_0(t)^T = \sigma_0(t); \\
 \dot{m}_t &= a_1(m_t, K_t, t), \quad m_0 = m(t_0); \\
 \dot{K}_t &= k_1^a(m_t, K_t, t) K_t + K_t k_1^a(m_t, K_t, t)^T + \\
 &\quad + \sigma_0(t), \quad K_0 = K(t_0); \\
 \frac{\partial K(t_1, t_2)}{\partial t_2} &= K(t_1, t_2) K_{t_2} k_1^a(m_{t_2}, K_{t_2}, t_2)^T, \\
 K(t_1, t_2) &= K_{t_1}.
 \end{aligned} \right\} \quad (20)$$

Для стационарных СтС (15) при условии асимптотической устойчивости матрицы $k_1^a(m^*, K^*)$

в основе МСЛ лежат уравнения (19), записанные в виде:

$$\left. \begin{aligned}
 a_0(m^*, K^*) &= 0; \\
 k_1^a(m^*, K^*) K^* + K^* k_1^a(m^*, K^*)^T + \bar{\sigma}_0 &= 0; \\
 \dot{k}_\tau(\tau) &= k_1^a(m^*, K^*) k(\tau); \\
 k(0) &= K^* \quad (\forall \tau > 0); \\
 k(\tau) &= k(-\tau)^T \quad (\forall \tau < 0).
 \end{aligned} \right\} \quad (21)$$

Теорема 2. Если существуют интегралы (18), то уравнения (17) лежат в основе МАМ на базе МСЛ для СтС (14), а уравнения (20) — на базе МСЛ для СтС (15). Для моделирования стационарных СтП согласно МНА служат соотношения (19), а МСЛ — (21) при условии асимптотической устойчивости матрицы a_{21}^* .

Для алгоритмизации МНА необходимо уметь вычислять следующие интегралы:

$$\begin{aligned}
 I_0^a &= I_0^a(m_t, K_t, t) = \\
 &= a_1(m_t, K_t, t) = M_{\Delta_y}^N a(Y_t, t); \quad (22)
 \end{aligned}$$

$$\begin{aligned}
 I_1^a &= I_1^a(m_t, K_t, t) = \\
 &= a_{21}(m_t, K_t, t) = M_{\Delta_y}^N a(Y_t, t) Y_t^{0T}; \quad (23)
 \end{aligned}$$

$$\begin{aligned}
 I_0^{\bar{\sigma}} &= I_0^{\bar{\sigma}}(m_t, K_t, t) = \\
 &= a_{22}(m_t, K_t, t) = M_N \bar{\sigma}(Y_t, t), \quad (24)
 \end{aligned}$$

а для МСЛ достаточно вычислить интеграл в (22), причем интеграл I_1^a вычисляется по формуле [8–10]:

$$k_1^a = k_1^a(m_t, K_t, t) = \left[\left(\frac{\partial}{\partial m_t} \right) I_0^a(m_t, K_t, t)^T \right]^T.$$

Численные методы определения интегралов типа (22)–(24) описаны в [7, 11].

Замечание 5. Важно иметь в виду, что уравнения МНА (МСЛ) содержат интегралы I_0^a , I_1^a и $I_0^{\bar{\sigma}}$ в виде соответствующих коэффициентов, поэтому процедура вычисления интегралов должна быть согласована с методом численного решения обыкновенных дифференциальных уравнений для m_t , K_t и $K(t_1, t_2)$. Эти коэффициенты допускают дифференцирование по m_t и K_t , так как под интегралом стоит сглаживающая нормальная плотность.

В [1] изложены алгоритмы дискретного аналитического и статистического моделирования типовых распределений (в том числе нормальных) в нелинейных МССтС. Алгоритмы дискретного аналитического и статистического моделирования для СтС с СБН, а также смешанные алгоритмы различной степени точности относительно шага интегрирования также представлены в [1].

5 Особенности аналитического моделирования гармонических и узкополосных стохастических процессов

Рассмотрим сначала гармоническую линеаризацию бесселевой функции $J_n(y)$ целого порядка (7) при несимметричном гармоническом входном процессе Y вида

$$Y = A_0 + A \sin \psi, \quad \psi = \omega t. \quad (25)$$

В соответствии с методом гармонической линеаризации (МГЛ) [8] зависимость (7) при условии (25) аппроксимируется функцией

$$Z \approx A_{z0} + q_z A \sin \psi + q'_z A \cos \psi. \quad (26)$$

Входящие в (26) коэффициенты МГЛ определяются формулами:

$$A_{z0} = A_{z0}(A_0, A) = \frac{1}{2\pi} \int_0^{2\pi} J_n(A_0 + A \sin \psi) d\psi, \quad (27)$$

$$\left. \begin{aligned} q_z &= q_z(A_0, A) = \\ &= \frac{1}{\pi A} \int_0^{2\pi} J_n(A_0 + A \sin \psi) \sin \psi d\psi; \\ q'_z &= q'_z(A_0, A) = \\ &= \frac{1}{\pi A} \int_0^{2\pi} J_n(A_0 + A \sin \psi) \cos \psi d\psi. \end{aligned} \right\} \quad (28)$$

Для $n = 2r + 1$, учитывая нечетность функции $J_{2r+1}(y)$, вместо A_{z0} вводят следующий коэффициент:

$$\begin{aligned} q_{z0} &= q_{z0}(A_0, A) = \\ &= \frac{1}{2\pi A_0} \int_0^{2\pi} J_{2r+1}(A_0 + A \sin \psi) d\psi. \end{aligned} \quad (29)$$

Теорема 3. Если существуют интегралы (27) и (28), то уравнения (26) при условиях (27) и (28) лежат в основе МГЛ для БН (7) четного порядка $J_{2r}(y)$, а (28) и (29) — для БН нечетного порядка $J_{2r+1}(y)$.

Рассмотрим конечномерную динамическую систему вида

$$\dot{Y}_t = F(Y_t, X_t, t), \quad Y(t_0) = Y_0. \quad (30)$$

Здесь Y_t — вектор состояния размерности p^y ; $X_t = A_0 + A_1 \sin \psi$ ($\psi = \omega t$) — несимметричный гармонический процесс размерности p^x (A_0 и A_1 —

постоянные или медленно меняющиеся векторные функции на интервале $2\pi/\omega$); $F(Y_t, X_t, t)$ — векторная нелинейная функция, допускающая разложение Фурье–Бесселя вида (6):

$$F(Y_t, X_t, t) = \sum_l \prod_k \alpha_{lk}(X_{l,k}, t) J_n(\rho_l Y_{lk}). \quad (31)$$

Здесь α_{lk} — известные коэффициенты; ρ_l — положительные корни $J_n(y)$, расположенные в порядке возрастания.

Используя формулы (1)–(4) и применяя метод медленно меняющихся амплитуд, положив

$$Y_t = B_0 + B_1 \sin \psi + B_2 \cos \psi \quad (\psi = \omega t), \quad (32)$$

придем к следующим дифференциальным уравнениям для векторов B_i ($i = 0, 1, 2$):

$$\left. \begin{aligned} \dot{B}_0 &= F_0(A_0, A_1, t); \\ \dot{B}_1 &= B_2 \omega + F_1(A_0, A_1, t); \\ \dot{B}_2 &= -B_1 \omega + F_2(A_0, A_1, t). \end{aligned} \right\} \quad (33)$$

Здесь $F_i = F_i(A_0, A_1, t)$ ($i = 0, 1, 2$) — векторные коэффициенты МГЛ, вычисляемые согласно разд. 3 по формулам:

$$\left. \begin{aligned} F_0(A_0, A_1, t) &= \frac{1}{2\pi} \int_0^{2\pi} F(B_0 + B_1 \sin \psi + \\ &+ B_2 \cos \psi, A_0 + A_1 \sin \psi, t) d\psi; \\ F_1(A_0, A_1, t) &= \frac{1}{\pi} \int_0^{2\pi} F(B_0 + B_1 \sin \psi + \\ &+ B_2 \cos \psi, A_0 + A_1 \sin \psi, t) \sin \psi d\psi; \\ F_2(A_0, A_1, t) &= \frac{1}{\pi} \int_0^{2\pi} F(B_0 + B_1 \sin \psi + \\ &+ B_2 \cos \psi, A_0 + A_1 \sin \psi, t) \cos \psi d\psi. \end{aligned} \right\} \quad (34)$$

Для определения стационарных B_0^* , B_1^* и B_2^* следует приравнять правые части (34) нулю:

$$F_i(B_0^*, B_1^*, B_2^*) = 0 \quad (i = 0, 1, 2). \quad (35)$$

Теорема 4. Если существуют интегралы (34), то в основе МАМ для медленноменяющихся гармонических процессов в динамической системе (30) при условии (31) лежат уравнения (32)–(34), а для стационарных гармонических процессов — уравнения (35).

Теперь рассмотрим узкополосный гауссовский СтП, допускающий представление вида [12]:

$$X_t = A_t \sin[\omega t + \Theta(t)].$$

Здесь A_t и Θ_t — медленноменяющиеся процессы на интервале времени $2\pi/\omega$ и в диапазоне частот $\Delta\omega \ll \omega$. Одномерные плотности $a = A_t$ и $\theta = \Theta_t$ определяются следующими формулами:

$$f(\dot{a}) = \frac{1}{\Delta\omega\sqrt{2\pi D}} \exp\left(-\frac{\dot{a}}{2\Delta^2 D}\right);$$

$$f(\dot{\theta}) = \frac{\Delta^2}{2(\dot{\theta}^2 + \Delta^2)^{3/2}}.$$

В этом случае непосредственно применяется метод медленноменяющихся амплитуд (25)–(35) совместно с формулами вычисления функций случайного аргумента [8–10].

В приложении П2 приведены тестовые примеры к разд. 4 и 5.

6 Заключение

На основе приближенных методов МНА и МГЛ развито методическое и алгоритмическое обеспечение аналитического моделирования процессов в динамических системах со сложными (цилиндрическими первого рода и целого порядка) БН в условиях гармонических, стохастических узкополосных и широкополосных возмущений.

В приложениях П1 и П2 приведены выражения для коэффициентов МНА (МСЛ) для типовых БН, а также тестовые примеры. Особое внимание уделено аналитическому моделированию процессов в осцилляторе Бесселя в условиях различных возмущений.

Для учета негауссовости стохастической динамики могут использоваться методы моментов [9, 10, 13] и ортогональных разложений [9, 10, 14, 15].

Алгоритмы положены в основу разрабатываемого инструментального программного обеспечения для решения задач надежности и безопасности систем и средств информатики и управления.

Результаты допускают обобщение на случай других бesselевских нелинейностей, а также непрерывно-дискретных, дискретных бesselевых динамических систем, в том числе с автокоррелированными возмущениями.

Приложения

П1. Коэффициенты статистической и гармонической линеаризации типовых бesselевых нелинейностей

Формулы (10)–(13) для коэффициентов статистической линеаризации $\varphi_0 = \varphi_0(m_y, D_y)$, $k_0 = k_0(m_y, D_y)$ и $k_1 = k_1(m_y, D_y)$ с учетом замечаний 2 и 3 при $m_y \neq 0$,

$D_y \neq 0$ позволяют проводить расчеты для СБН численными методами [11], используя в качестве параметров D_y и отношение сигнал–шум $\zeta = m_y D_y^{-1/2}$.

Для случая $m_y = 0$, используя известные формулы для определенных интегралов от типовых БН [6, 7]

$$\int_0^\infty e^{-a^2 \eta^2} \eta^{\mu-1} J_n(b\eta) d\eta =$$

$$= \frac{\Gamma((n+\mu)/2) (b/(2a))^n}{2a^\mu \Gamma(n+1)} \mathcal{K}\left(\frac{n+\mu}{2}, n+1, -\frac{b^2}{4a^2}\right)$$

(Re $(\mu+n) > 0$, Re $a^2 > 0$);

$$\int_0^\infty e^{-a^2 \eta^2} J_n^2(b\eta) d\eta = \frac{1}{2a^2} \exp\left(-\frac{b^2}{2a^2}\right) I_n\left(\frac{b^2}{2a^2}\right)$$

(Re $n > 1$, $b > 0$)

($\mathcal{K}(x, y, z)$ — вырожденная функция Куммера [7], $I_n(x)$ — функция Бесселя мнимого аргумента (замечание 1)), приходим к следующим результатам:

$$Z = J_n(Y); \varphi_0(0, D) =$$

$$= \frac{\Gamma((n+1)/2)}{\sqrt{\pi}\Gamma(n+1)} \left(\frac{D}{2}\right)^{n/2} \mathcal{K}\left(\frac{n+1}{2}, n+1, -\frac{D}{2}\right); \quad (36)$$

$$Z = J_n(bY); \varphi_0(0, D) =$$

$$= \frac{\Gamma((n+1)/2)}{\sqrt{\pi}\Gamma(n+1)} \left(b\sqrt{\frac{D}{2}}\right)^n \mathcal{K}\left(\frac{n+1}{2}, n+1, -\frac{b^2 D}{2}\right);$$

$$Z = Y^{\mu-1} J_n(bY); \varphi_0(0, D) =$$

$$= \frac{2^{(\mu-1)/2}}{\sqrt{\pi}\Gamma(n+1)} D^{(\mu-1)/2} \mathcal{K}\left(\frac{n+\mu}{2}, n+1, -\frac{b^2 D}{2}\right);$$

$$k_1(0, D) =$$

$$= \frac{2^{\mu/2} D^\mu}{\sqrt{\pi}\Gamma(n+1)} \mathcal{K}\left(\frac{n+\mu+1}{2}, n+1, -\frac{b^2 D}{2}\right); \quad (37)$$

$$Z = J_n^2(Y); \varphi_0(0, D) = \sqrt{\frac{2D}{\pi}} e^{-D} I_n(D);$$

$$Z = J_n^2(bY); \varphi_0(0, D) = \sqrt{\frac{2D}{\pi}} e^{-b^2 D} I_n(b^2 D). \quad (38)$$

Для расчета коэффициентов гармонической линеаризации БН и СБН используются формулы (27)–(29) и (34). При этом обычно используются численные методы [11].

П2. Тестовые примеры

Пример 1. Для одномерной бesselевой системы с аддитивным гауссовским шумом

$$\dot{Y}_t = \alpha + \beta J_{2r+1}(bY_t) + \gamma V$$

(α, β, b и γ — постоянные; V — белый шум интенсивности ν) МАМ на основе МСЛ (теорема 2) приводит к следующим результатам:

$$\left. \begin{aligned} \dot{m}_y &= \alpha + \beta \varphi_0^{2r+1}(m_y, D_y); \\ \dot{D}_y &= 2\beta k_1^{2r+1}(m_y, D_y) D_y + \sigma. \end{aligned} \right\} \quad (39)$$

Здесь введены следующие обозначения:

$$\begin{aligned} \sigma &= \gamma^2 \nu; \\ \varphi_0^{2r+1}(m_y, D_y) &= \\ &= \frac{1}{\sqrt{2\pi D_y}} \int_{-\infty}^{\infty} J_{2r+1}(b\eta) e^{-(\eta - m_y)^2 / 2D_y} d\eta; \\ k_1^{2r+1}(m_y, D_y) &= \frac{\partial \varphi_0^{2r+1}(m_y, D_y)}{\partial m_y}. \end{aligned}$$

При $\beta < 0$, приравняв нулю правые части (39), получим уравнения для стационарных m^* и D^* :

$$\alpha + \beta \varphi_0^{2r+1}(m^*, D^*) = 0; \quad 2\beta k_1^{2r+1}(m^*, D^*) D^* + \gamma^2 \nu = 0.$$

Отсюда при $\alpha = 0$ находим, что $m^* = 0$, а D^* с учетом (37) при $\mu = 1$ определяется из уравнения

$$k^{2r+1}(0, D^*) D^* + \delta^* = 0 \quad \left(\delta^* = \frac{\gamma^2 \nu^*}{2\beta} \right).$$

Пример 2. Для одномерной бesselевой системы с мультипликативным шумом

$$\dot{Y}_t = \alpha + \beta Y_t + \gamma J_{2r+1}(bY_t) V$$

(α, β, γ и b — постоянные; V — гауссовский белый шум интенсивности ν) МАМ на основе МНА (теорема 2) приводит к следующим результатам:

$$\left. \begin{aligned} \dot{m}_y &= \alpha + \beta m_y; \\ \dot{D}_y &= 2\beta D_y + \sigma(m_y, D_y); \\ \sigma(m_y, D_y) &= \\ &= \frac{\gamma^2 \nu}{\sqrt{2\pi D_y}} \int_{-\infty}^{\infty} J_{2r+1}^2(b\eta) e^{-(\eta - m_y)^2 / (2D_y)} d\eta. \end{aligned} \right\} \quad (40)$$

При $\beta < 0$, приравняв нулю правые части (40), придем к следующим уравнениям для стационарных m^* и D^* :

$$\alpha + \beta m^* = 0; \quad 2\beta D^* + \sigma(m^*, D^*) = 0.$$

Отсюда при $\alpha = 0$ следует, что $m^* = 0$, а D^* с учетом (38) определяется из уравнения

$$\sqrt{D^*} I_{2r+1}(b^2 D^*) e^{-b^2 D^*} = \sqrt{\frac{\pi}{2}}.$$

Пример 3. Рассмотрим сначала двумерную СтС, описывающую стохастическую динамику осциллятора Бесселя:

$$\left. \begin{aligned} \dot{Y} + \omega_0^2 J_{2r+1}(Y) &= 0 \text{ при } Y_1 = Y, \dot{Y}_1 = Y_2; \\ \dot{Y}_2 &= \Pi'(Y_1); Y_1(t_0) = Y_{10}; Y_2(t_0) = Y_{20}. \end{aligned} \right\}$$

Здесь Y_1 и Y_2 — координата и скорость; $\Pi'_{Y_1}(Y_1) = \partial \Pi(Y_1) / \partial Y_1 = -\omega_0^2 J_{2r+1}(Y_1)$; $\Pi''(Y_1) = -\omega_0^2 J'_{2r+1}(Y_1) = -(\omega_0^2 / 2) [J_{2r}(Y_1) - J_{2r+2}(Y_1)]$.

Пользуясь формулами (5) для интеграла от бesselевой функции, находим точное выражение для потенциальной энергии $\Pi(Y_1)$:

$$\begin{aligned} \Pi_{2r+1}(Y_1) &= -\omega_0^2 \int_0^{Y_1} J_{2r+1}(\eta) d\eta = \\ &= -\omega_0^2 \left[1 - J_0(Y_1) - 2 \sum_{k=0}^r J_{2k}(Y_1) \right]. \end{aligned}$$

При этом функция Гамильтона будет иметь вид:

$$\mathcal{H}(Y_1, Y_2) = \frac{Y_2^2}{2} - \omega_0^2 \left[1 - J_0(Y_1) - 2 \sum_{k=0}^r J_{2k}(Y_1) \right]. \quad (41)$$

Нули $J_n(Y_1)$ действительны, разделяются нулями $J_{n+1}(Y_1)$ и $J'_n(Y_1)$. При этом выполняются условия (3).

Фазовый портрет осциллятора Бесселя при произвольных начальных условиях определяется следующими дифференциальными или интегральными соотношениями:

$$\frac{dY_2}{dY_1} = \frac{-\omega_0^2 J_{2r+1}(Y_1)}{Y_2};$$

$$Y_2^2 = h - 2\Pi_{2r+1}(Y_1), \quad h = Y_{20}^2 + 2\omega_0^2 \int_0^{Y_{10}} J_{2r+1}(\eta) d\eta.$$

Особыми точками являются множества $\{Y_2 = 0, J_{2r+1}(Y_1) = 0\}$. Для их устойчивости достаточно, чтобы $\Pi''(Y_1) > 0$.

Полученные соотношения лежат в основе МАМ при детерминированных и случайных начальных условиях.

Наличие сил сопротивления (возбуждения), зависящих от параметра U , в виде постоянной составляющей $L_0 = L_0(U)$ и переменной составляющей $L_1 = L_1(Y_1, Y_2, U)$ в уравнениях осциллятора, когда

$$\dot{Y}_1 = Y_2; \quad \dot{Y}_2 = \Pi'(Y_1, U) + L_0(U) - L_1(Y_1, Y_2, U),$$

приводит к следующему дифференциальному уравнению фазового портрета:

$$\frac{dY_2}{dY_1} = \frac{-\omega_0^2 J_{2r+1}(Y_1) + L_0(U) - L_1(Y_1, Y_2, U)}{Y_2}.$$

Особыми точками фазового портрета являются множества $\{Y_2 = 0, -\omega_0^2 J_{2r+1}(Y_1) + L_0(U) - L_1(Y_1, Y_2, U) = 0\}$. Полученные соотношения лежат в основе МАМ при детерминированных и случайных значениях начальных условий Y_{10}, Y_{20} и параметра U .

Теперь рассмотрим аналитическое моделирование процессов в осцилляторе Бесселя при наличии сил вязкого трения и гармонического возмущения, основываясь на уравнениях

$$\dot{Y}_1 = Y_2, \quad \dot{Y}_2 = -\omega_0^2 J_{2r+1}(Y_1) - 2\varepsilon\omega_0 Y_2 + A \sin \omega t. \quad (42)$$

Выполним в (42) замену переменных $Y_1 = B \sin \psi$ и $Y_2 = B\omega \cos \psi$ ($\psi = \omega t + \theta$) и проведем гармоническую линеаризацию $J_{2r+1}(Y_1)$, положив

$$\begin{aligned} J_{2r+1}(Y_1) &\approx q_{2r+1}(B) B \sin \psi; \\ q_{2r+1}(B) &= \frac{1}{\pi B} \int_0^{2\pi} J_{2r+1}(B \sin \psi) \sin \psi \, d\psi = \\ &= \frac{1}{\pi B} \int_0^{2\pi} \left[\frac{(A \sin \psi)^{2r+1}}{2^{2r+1}} \sum_{k=0}^{\infty} \left(-\frac{A^2 \sin^2 \psi}{4} \right)^k \times \right. \\ &\quad \left. \times \frac{1}{k!(2r+1+k)!} \right] d\psi. \end{aligned}$$

В результате уравнения (42) примут следующий вид:

$$\left. \begin{aligned} \dot{B} &= L_c \cos^2 \psi + L_s \sin \psi \cos \psi; \\ B\dot{\theta} &= -L_c \sin \psi \cos \psi - L_s \sin^2 \psi, \end{aligned} \right\} \quad (43)$$

где

$$\begin{aligned} L_c &= -\frac{2\varepsilon\omega_0}{\omega} B + \frac{A}{\omega} \sin \theta; \\ L_s &= \frac{\omega^2 - \omega_0^2}{\omega} q_{2r+1}(B) B + \frac{A}{\omega} \cos \theta. \end{aligned}$$

Далее после осреднения по ψ правых частей (43) приходим к искомым уравнениям для осредненных B и θ :

$$\dot{B} = \frac{1}{2} \langle L_c(B, \theta, A, \omega) \rangle; \quad \dot{\theta} = -\frac{1}{2} \frac{\langle L_s(B, \theta, A, \omega) \rangle}{B}, \quad (44)$$

где $\langle \dots \rangle$ — символ осреднения по ψ . Для определения стационарных B^* и θ^* имеем конечные уравнения:

$$\langle L_c(B^*, \theta^*, A, \omega) \rangle = 0; \quad \langle L_s(B^*, \theta^*, A, \omega) \rangle = 0.$$

При этом, как следует из (44), резонансное соотношение имеет следующий вид:

$$\left(\frac{\omega_0}{\omega} \right)^2 = q_{2r+1}(B). \quad (45)$$

Полученные соотношения лежат в основе МАМ вынужденных гармонических колебаний.

Рассмотрим уравнения

$$\dot{Y}_1 = Y_2; \quad \dot{Y}_2 = -\omega_0^2 J_{2r+1}(Y_1) + L_0 - 2\varepsilon\omega_0 Y_2 + V, \quad (46)$$

описывающие процессы в осцилляторе Бесселя при наличии постоянного возмущения L_0 , сил вязкого трения и широкополосного гауссовского стохастического процесса вида белого шума интенсивности $\nu^V = \nu$.

При $L_0 = 0$ и $\varepsilon\omega_0 > 0$ точное стационарное распределение Y_1 и Y_2 определяется формулой Гиббса [9, 10] для функции Гамильтона (41):

$$f^*(y_1, y_2) = c \exp \left[-\frac{4\varepsilon\omega_0}{\nu} \mathcal{H}(y_1, y_2) \right],$$

где c — постоянная нормировки плотности. Отсюда следует, что распределение Y_2 — гауссовское с параметрами $m_{y_2}^* = 0$, $D_2^* = \nu/(2\varepsilon\omega_0)$, а распределение Y_1 — негауссовское:

$$f_{y_1}^*(y_1) = c \exp \left[-\frac{4\varepsilon\omega_0}{\nu} \Pi_{2r+1}(y_1) \right].$$

Точные значения математического ожидания и дисперсии для Y_1 равны:

$$m_{y_1}^* = 0; \quad D_1^* = c \int_{-\infty}^{\infty} \eta \exp \left[-\frac{4\varepsilon\omega_0}{\nu} \Pi_{2r+1}(\eta) \right] d\eta.$$

Для уравнений (45) при $L_0 \neq 0$, согласно теореме 2, находим:

$$\left. \begin{aligned} \dot{m}_1 &= m_2; \\ \dot{m}_2 &= -\omega_0^2 \varphi_0^{2r+1}(m_1, D_1) + L_0 - 2\varepsilon\omega_0 m_2; \\ \dot{D}_1 &= 2K_{12}; \\ \dot{D}_2 &= -2 [\omega_0^2 k_1^{2r+1}(m_1, D_1) K_{12} + 2\varepsilon\omega_0 D_2] + \nu; \\ \dot{K}_{12} &= D_2 - \omega_0^2 k_1^{2r+1}(m_1, D_1) D_1 - 2\varepsilon\omega_0 K_{12}. \end{aligned} \right\} \quad (47)$$

Здесь

$$\begin{aligned} \varphi_0^{2r+1}(m_1, D_1) &= \\ &= \frac{1}{\sqrt{2\pi D_1}} \int_{-\infty}^{\infty} J_{2r+1}(\eta) e^{-(\eta - m_y)^2 / (2D_y)} d\eta; \\ k_1^{2r+1}(m_1, D_1) &= \frac{\partial \varphi_0^{2r+1}(m_1, D_1)}{\partial m_1}. \end{aligned}$$

Отсюда при $\varepsilon > 0$ и $L_0 = 0$ в стационарном случае имеем:

$$\left. \begin{aligned} m_1^* &= 0; \quad m_2^* = 0; \quad K_{12}^* = 0; \\ D_2^* &= \frac{\nu^*}{4\varepsilon\omega_0}; \quad k_1^{2r+1}(0, D_1^*) D_1^* = \frac{\nu^*}{4\varepsilon\omega_0^2}. \end{aligned} \right\} \quad (48)$$

Для входящего в (48) коэффициента статистической линеаризации в силу (36) находим следующее выражение:

$$\begin{aligned} k_1^{2r+1}(h) &= \frac{4\Gamma((2r+3)/2)}{\Gamma(2r+2)(2h)^{2r+3}} \mathcal{K} \left(\frac{2r+3}{2}, 2r+2, \xi \right) \\ &\quad \left(h^2 = \frac{1}{2D_1^*} \right). \end{aligned} \quad (49)$$

Пользуясь асимптотическими разложениями функции Куммера [7], из (49) получаем следующие формулы:

$$k_1^{2r+1}(h) \sim \begin{cases} \frac{4\Gamma((2r+3)/2)}{\Gamma(2r+2)(2h)^{2r+3}} & \text{при } D_1 \rightarrow 0; \\ \frac{\Gamma(2r+2)}{\Gamma((2r+1)/2)} \cdot (2h)^{2r+3} & \text{при } D_1 \rightarrow \infty, \end{cases} \quad (50)$$

где $\Gamma(n+1) = n!$; $h^2 = 1/(2D_1)$.

Формулы (50) для $n = 2r + 1$ при $r = 0$ и 1 имеют соответственно вид:

$$k_1^{(1)} \sim \frac{2\sqrt{\pi}}{(2h)^3}; \quad k_1^{(1)} \sim \frac{2}{\sqrt{\pi}}(2h)^3,$$

$$k_1^{(3)} \sim \frac{\sqrt{\pi}}{12(2h)^5}; \quad k_1^{(3)} \sim \frac{12}{\sqrt{\pi}}(2h)^5.$$

На основе уравнений (47) строятся эллипсы рассеивания для зависимостей $D_1 = D_1(t)$, $D_2 = D_2(t)$ и $K_{12} = K_{12}(t)$ в переходном и стационарном режимах.

Замечание 6. Аналогично МНА (МСЛ) используется для моделирования динамики осциллятора Бесселя в случае аддитивных и мультипликативных возмущений вида $[1 + \gamma(Y_1, Y_2, U)]V$ в (46).

Литература

1. Сеницын И. Н. Параметрическое статистическое и аналитическое моделирование распределений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2013. Т. 7. Вып. 2. С. 4–16.
2. Сеницын И. Н., Сеницын В. И. Аналитическое моделирование нормальных процессов в стохастических системах со сложными нелинейностями // Информатика и её применения, 2014. Т. 8. Вып. 3. С. 12–18.
3. Сеницын И. Н., Сеницын В. И., Сергеев И. В., Белосусов В. В., Шоргин В. С. Математическое обеспечение аналитического моделирования стохастических систем со сложными нелинейностями // Системы и средства информатики, 2014. Т. 24. № 3. С. 4–17.
4. Сеницын И. Н., Сеницын В. И., Корепанов Э. Р. Моделирование процессов в стохастических системах со сложными трансцендентными нелинейностями // Информатика и её применения, 2015. Т. 9. Вып. 2. С. 23–29.
5. Грей Э., Мэтьюз Г. Функции Бесселя и их приложения к физике и механике. — М.: ИЛ, 1953. 371 с.
6. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. — М.: ГИФМЛ, 1963. 1100 с.
7. Справочник по специальным функциям / Под ред. М. Абрамовича, И. Стигана. — М.: Наука, 1979. 832 с.
8. Сеницын И. Н., Сеницын В. И. Лекции по нормальной и эллипсоидальной аппроксимации распределений в стохастических системах. — М.: ТОРУС ПРЕСС, 2013. 488 с.
9. Пугачёв В. С., Сеницын И. Н. Стохастические дифференциальные системы. Анализ и фильтрация. — М.: Наука, 1990. 632 с. (Pugachev V. S., Sinitsyn I. N. Stochastic differential systems. Analysis and filtering. — Chichester — New York, NY, USA: John Wiley, 1987. 549 p.)
10. Пугачёв В. С., Сеницын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с. (Pugachev V. S., Sinitsyn I. N. Stochastic systems. Theory and applications. — Singapore: World Scientific, 2001. 908 p.)
11. Попов Б. А., Теслер Г. С. Вычисление функций на ЭВМ: Справочник. — Киев: Наукова Думка, 1984. 599 с.
12. Свешников А. А. Прикладные методы теории случайных функций. — М.: Наука, 1968. 462 с.
13. Сеницын И. Н. Методы моментов в задачах аналитического моделирования распределений в нелинейных стохастических системах на многообразиях // Системы и средства информатики, 2015. Т. 25. № 3. С. 24–43.
14. Сеницын И. Н. Аналитическое моделирование распределений методом ортогональных разложений в нелинейных стохастических системах на многообразиях // Информатика и её применения, 2015. Т. 9. Вып. 3. С. 17–24.
15. Сеницын И. Н. Применение ортогональных разложений для аналитического моделирования многомерных распределений в нелинейных стохастических системах на многообразиях // Системы и средства информатики, 2015. Т. 25. № 3. С. 4–23.

Поступила в редакцию 13.08.15

ANALYTICAL MODELING OF PROCESSES IN DYNAMICAL SYSTEMS WITH CYLINDRIC BESSEL NONLINEARITIES

I. N. Sinitsyn

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

Abstract: Methods of analytical modeling (MAM) for processes in dynamical systems with complex Bessel nonlinearities with harmonically and stochastically wide and narrow band disturbances are given. Necessary elements of the cylindrical Bessel functions theory and complex Bessel nonlinearities are presented. Methodological and algorithmical support for MAM based on the statistical linearization method (SLM) and the normal approximation method for wide-band stochastic processes (white noise) is developed. Peculiarities of MAM for

harmonical and narrow band stochastic processes are discussed. Test examples of one-dimensional systems with additive and multiplicative noises and Bessel nonlinearities and for Bessel oscillator with various disturbances are given. Conclusions and some generalizations are mentioned.

Keywords: Bessel nonlinearity; Bessel oscillator; complex Bessel nonlinearity; Gibbs formula; harmonical process; Kummer function; method of analytical modeling; narrow band stochastic processes; normal approximation method (NAM); statistical linearization method (SLM); stochastic system on manifold (MStS); white noise; wide band stochastic process

DOI: 10.14357/19922264150404

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 15-07-02244).

References

1. Sinitsyn, I. N. 2013. Parametricheskoe statisticheskoe i analiticheskoe modelirovanie raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Parametrical statistical and analytical modeling of distributions in nonlinear stochastic systems on manifolds]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):4–16.
2. Sinitsyn, I. N., and V. I. Sinitsyn. 2014. Analiticheskoe modelirovanie normal'nykh protsessov v stokhasticheskikh sistemakh so slozhnymi nelineynostyami [Analytical modeling of normal processes in stochastic systems with complex nonlinearities]. *Informatika i ee Primeneniya — Inform. Appl.* 8(3):12–18.
3. Sinitsyn, I. N., V. I. Sinitsyn, I. V. Sergeev, V. V. Belousov, and V. S. Shorgin. 2014. Matematicheskoe obespechenie analiticheskogo modelirovaniya stokhasticheskikh sistem so slozhnymi nelineynostyami [Mathematical software tools for analytical modeling in stochastic systems with complex nonlinearities]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(3):4–17.
4. Sinitsyn, I. N., V. I. Sinitsyn, and E. R. Korepanov. 2015. Modelirovanie protsessov v stokhasticheskikh sistemakh so slozhnymi transtsendentnymi nelineynostyami [Modeling of normal processes in stochastic systems with complex transcendental nonlinearities]. *Informatika i ee Primeneniya — Inform. Appl.* 9(2):23–29.
5. Grey, E., and G. Met'yuz. 1953. *Funktsii Besselya i ikh prilozheniya k fizike i mekhanike*. — Moscow: IL. 371 p.
6. Gradshteyn, I. S., and I. M. Ryzhik. 1963. *Tablitsy integralov, summ, ryadov i proizvedeniy* [Tables of integrals, sums, series and products]. Moscow: GIFML. 1100 p.
7. Abramovich, M., and I. Stigan, eds. 1979. *Spravochnik po spetsial'nym funktsiyam* [Handbook of mathematical functions]. Moscow: Nauka. 832 p.
8. Sinitsyn, I. N., and V. I. Sinitsyn. 2013. *Leksii po normal'noy i ellipsoidal'noy aproksimatsii raspredeleniy v stokhasticheskikh sistemakh* [Lectures on normal and ellipsoidal approximation of distributions in stochastic systems]. Moscow: TORUS PRESS. 488 p.
9. Pugachev, V. S., and I. N. Sinitsyn. 1987. *Stochastic differential systems. Analysis and filtering*. Chichester — New York, NY: John Wiley. 549 p.
10. Pugachev, V. S., and I. N. Sinitsyn. 2001. *Stochastic systems. Theory and applications*. Singapore: World Scientific. 908 p.
11. Popov, B. A., and G. S. Tesler. 1984. *Vychislenie funktsiy na EVM: Spravochnik* [Calculation of functions on the computer: Handbook]. Kiev: Naukova Dumka. 599 p.
12. Sveshnikov, A. A. 1968. *Prikladnye metody teorii sluchaynykh funktsiy* [Applied methods in theory of random functions]. Moscow: Nauka. 462 p.
13. Sinitsyn, I. N. 2015. Metody momentov v zadachakh analiticheskogo modelirovaniya raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Moments methods for analytical modeling of stochastic systems on manifolds]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):24–43.
14. Sinitsyn, I. N. 2015. Analiticheskoe modelirovanie raspredeleniy metodom ortogonal'nykh razlozheniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Analytical modeling in stochastic systems on manifolds based on orthogonal expansions]. *Informatika i ee Primeneniya — Inform. Appl.* 9(3):17–24.
15. Sinitsyn, I. N. 2015. Primenenie ortogonal'nykh razlozheniy dlya analiticheskogo modelirovaniya mnogomernykh raspredeleniy v nelineynykh stokhasticheskikh sistemakh na mnogoobraznykh [Applications of orthogonal expansions for analytical modeling of multidimensional distributions in stochastic systems on manifold]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):4–23.

Received August 13, 2015

Contributor

Sinitsyn Igor N. (b. 1940) — Doctor of Science in technology, professor, Honored scientist of RF, 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; sinitsin@dol.ru

К АНАЛИЗУ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ОБСЛУЖИВАНИЯ В СОВРЕМЕННЫХ БЕСПРОВОДНЫХ СЕТЯХ

О. Г. Вихрова¹, К. Е. Самуйлов², Э. С. Сопин³, С. Я. Шоргин⁴

Аннотация: Аналитики предсказывают, что в период с 2014 по 2019 г. темпы роста мирового мобильного трафика втрое превысят темпы роста мирового фиксированного трафика. Число мобильных пользователей увеличится с 4,1 млрд до 4,9 млрд, число мобильных устройств и подключений может достигнуть 10 млрд. К 2019 г. скорость передачи в мировых мобильных сетях может увеличиться с 1,7 до 4,0 Мбит/с. Особое внимание стоит обратить на долю мобильного видео, которая составит 72% мирового мобильного трафика. Тенденции роста числа мобильных подключений и объема мобильного трафика на сегодняшний день ставят перед операторами мобильных услуг задачу эффективного и адаптивного использования доступных радиоресурсов. В связи с этим в работе исследуется упрощенная математическая модель, которая позволяет аналитически оценить вероятность блокировки системы и среднее значение занятого ресурса в соответствии с политикой распределения ресурсов технологии LTE-Advanced.

Keywords: LTE-Advanced; политика распределения ресурсов; СМО с ограниченными ресурсами

DOI: 10.14357/19922264150405

1 Введение

Популярность мультимедийных услуг обусловлена стремительным ростом числа разнообразных мобильных устройств, подключенных к сети. Пользователи все чаще предпочитают просматривать медиаконтент на смартфонах, планшетах и отдают предпочтение беспроводному типу связи. Провайдеры телекоммуникационных услуг вынуждены искать способы эффективного и адаптивного распределения и использования ограниченных радиоресурсов. Технология высокоскоростной беспроводной передачи данных LTE-Advanced позволяет выделять каждой новой сессии диапазоны частот, основываясь на принципах мультидиапазонности, принимая во внимание тип запрашиваемой пользователем услуги и учитывая географическую зону обслуживания.

Каждое мобильное устройство сети LTE может передавать данные базовой станции с определенной скоростью. Мгновенная скорость в этом случае определяется политикой распределения ресурсов и ограничена максимальным значением, которое определяется по формуле $c_i^{\max} = w \log_2(1 + \gamma_i p_{\max})$ для некоторой i -й сессии, где w — доступная ширина полосы частот; p_{\max} — мощность передачи сигнала усилителем; γ_i — отношение сигнала к шуму для данной сессии, которое зависит от расстояния между устройством и станцией. Сессия может

быть принята только в том случае, если ресурсов системы достаточно, т. е. должно выполняться неравенство $\sum_i c_i / c_i^{\max} \leq 1$, где c_i — запрашиваемая i -й сессией скорость передачи данных. Для анализа вероятности блокировки вызова и доли используемых радиоресурсов системы в данной работе построена математическая модель в терминах систем массового обслуживания (СМО) с ограниченными ресурсами [1].

2 Система массового обслуживания с дискретным ресурсом

В работе [2] рассматривается многолинейная система без очереди, в которой поступившая заявка занимает некоторые ресурсы на все время ее обслуживания. По завершении обслуживания весь занятый данной заявкой объем ресурсов освобождается (рис. 1). Если в системе недостаточно свободных ресурсов, необходимых для обслуживания поступившей заявки, то такая заявка теряется. Предполагается, что случайные векторы, описывающие требования заявок к ресурсам, не зависят от процессов поступления и обслуживания заявок, независимы в совокупности и одинаково распределены. При анализе данной системы необходимо

¹Российский университет дружбы народов, o.vikhrova@gmail.com

²Российский университет дружбы народов, ksam@sci.pfu.edu.ru

³Российский университет дружбы народов, esopin@sci.pfu.edu.ru

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

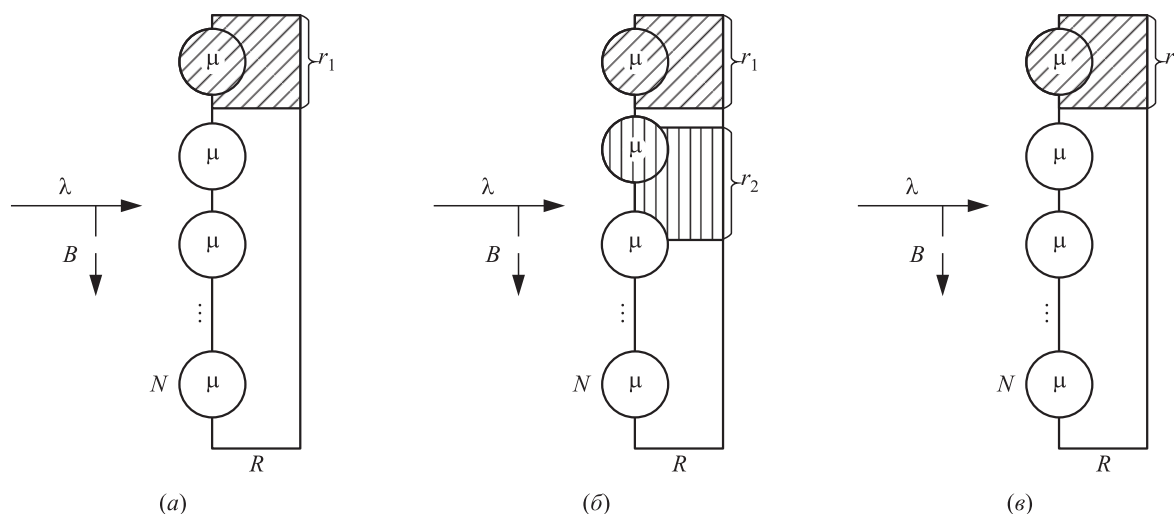


Рис. 1 Функционирование исходной СМО: (а) в системе одна заявка: занят один прибор и занято r_1 единиц ресурса; (б) в системе две заявки: заняты два прибора; 1-я заявка занимает r_1 единиц ресурса и 2-я заявка занимает r_2 единиц ресурса; (в) обслужилась вторая заявка: занят один прибор и 1-я заявка занимает r_1 единиц ресурса

для каждой заявки запоминать вектор занятых ею ресурсов.

Для упрощения анализа модели предлагается отслеживать только суммарный объем ресурсов, занятый всеми заявками. В связи с тем, что в упрощенной системе не отслеживаются объемы ресурсов, занятых каждой заявкой, объемы ресурсов, освобождаемых по завершении обслуживания, могут отличаться от тех, что были выделены заявкам при поступлении. При заданных суммарных объемах занятых ресурсов и числе заявок в системе в момент окончания обслуживания объемы освобождаемых ресурсов не зависят от поведения системы до этого момента и имеют функцию распределения, вычисляемую по формуле Байеса.

Средние значения объема занятого ресурса в исходной и упрощенной системе оказались очень близки при пуассоновском входящем потоке и экспоненциальном обслуживании [2]. Позднее в [3] было показано, что близкие значения имеют не только средние, но и стационарные распределения объема занимаемых ресурсов. И наконец, в статье [4] было строго доказано, что распределение стационарных вероятностей исходной и упрощенной модели совпадают.

Рассматривается упрощенная СМО заявок с N приборами и ограниченным дискретным ресурсом объемом R . В систему поступает пуассоновский поток заявок с параметром λ . Длительности обслуживания заявок независимы между собой, также не зависят от поступающего потока, экспоненциально распределены с параметром μ . В системе в некоторый момент времени t находится $\xi(t)$ заявок, при этом занято $\delta(t)$ ресурса, $\delta(t) < R$. Поступающей

в систему для обслуживания i -й заявке необходимо выделить целое число единиц ресурса $r_i \geq 0$ (рис. 2). Случайные величины r_i независимы в совокупности и одинаково распределены с функцией распределения $F(x)$, средним значением m и дисперсией σ^2 . Значение выделяемого объема ресурса не зависит от процессов поступления и обслуживания заявок.

Поступившая i -я заявка теряется, если на момент ее поступления объем доступного ресурса меньше объема ресурса, необходимого для ее обслуживания: $R - \delta(t) < r_i$, а также в случае, если все приборы заняты: $\xi(t) = N$. Объем занятого ресурса системы $\delta(t)$ увеличивается на величину ресурса $r_i \geq 0$, выделенного i -й заявке в момент начала ее обслуживания.

Объем занятых ресурсов $\delta(\tau_i)$ уменьшается на случайную величину v_i в момент τ_i завершения обслуживания i -й заявки. Случайная величина v_i при некотором заданном числе заявок в системе $\xi(\tau_i) = k$ и объеме занятого ресурса $\delta(\tau_i) = y$ не зависит от поведения системы в прошлом и согласно формуле Байеса имеет функцию распределения $f_k(x|y) = P(v_i \leq x | \xi(\tau_i) = k; \delta(\tau_i) = y)$, $0 \leq x \leq y$.

Случайные величины r_i требуемого заявке ресурса принимают значения $j = \overline{0, R}$. Пусть $p_j = P(r_i = j)$, $j = \overline{0, R}$, — распределение случайных величин r_i .

Введем обозначение $p_j^{(k)}$, $j = \overline{0, R}$, для k -кратной свертки этого распределения, которое позволяет определить, какие значения принимает случайная величина $y = \sum_{i=0}^k r_i$.

Тогда функция распределения объема всего занятого ресурса системы k заявками $F_k(x|j)$ является

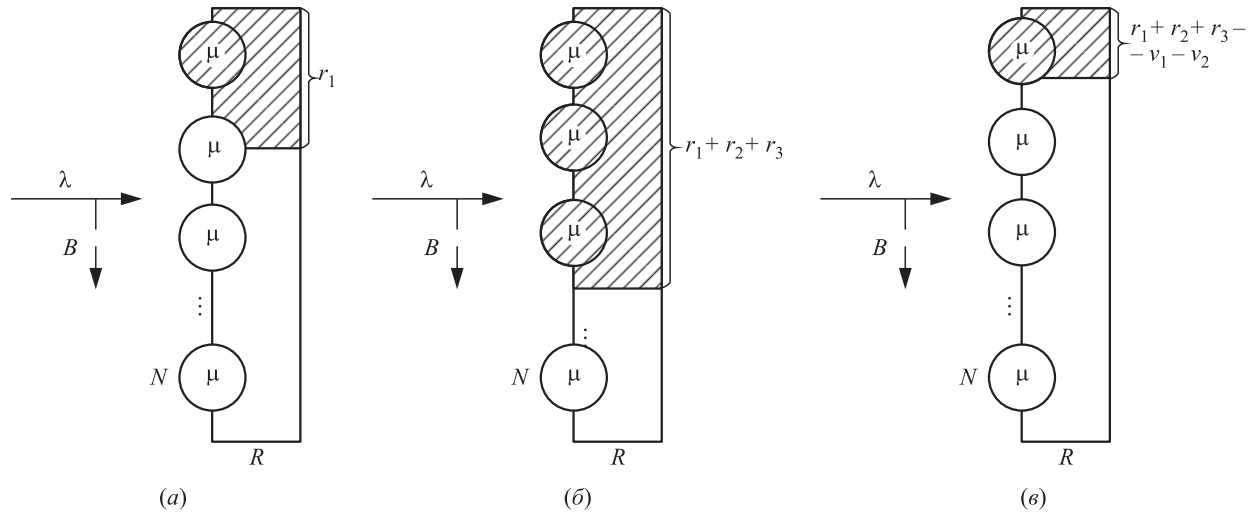


Рис. 2 Функционирование упрощенной СМО: (а) в системе одна заявка: занят один прибор и занято r_1 единиц ресурса; (б) в системе три заявки: заняты три прибора и занято $r_1 + r_2 + r_3$ единиц ресурса; (в) две заявки обслужились: занят один прибор и занято $r_1 + r_2 + r_3 - v_1 - v_2$ единиц ресурса

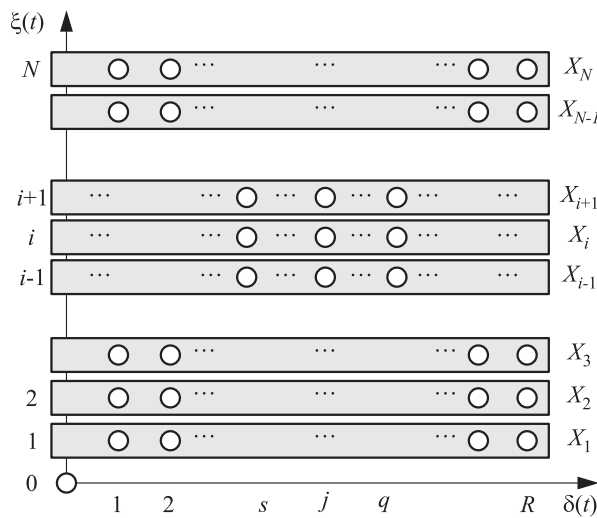


Рис. 3 Пространство состояний СП $X(t)$

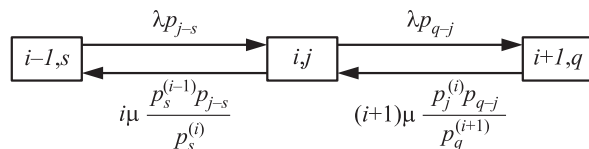


Рис. 4 Диаграмма переходов между состояниями системы

кусочно-постоянной на всей области ее определения, в точке $x = i$, $i = \overline{1, R}$, имеет скачок на величину

$$p_{ij}^k = \frac{p_i p_{j-i}^{(k-1)}}{p_j^{(k)}}, \quad 0 < i \leq j, \quad i = \overline{0, R}.$$

Свертки вероятностей $p_j^{(k)}$, $k \geq 2$, в общем случае могут быть вычислены на основе известного распределения p_i , $0 \leq i \leq R$, по рекуррентной формуле

$$p_j^{(k)} = \sum_{i=0}^j p_i p_{j-i}^{(k-1)}. \quad (1)$$

Случайный процесс (СП) $X(t) = (\xi(t), \delta(t))$ на множестве состояний $X_k = \{(k, i) | 0 \leq i \leq R, p_i^{(k)} > 0\}$ является цепью Маркова, заданной на пространстве состояний $X = \bigcup_{k=0}^N X_k$, где $X_k = \{(k, i) | 0 \leq i \leq R, p_i^{(k)} > 0\}$, и изображенной на рис. 3.

На рис. 4 представлена диаграмма переходов между состояниями системы при условии, что $1 \leq i < N$, $0 \leq s \leq j \leq q \leq R$.

Матрица интенсивностей переходов

$$A = [a((i, j), (k, r))]$$

является блочной трехдиагональной матрицей с диагональными блоками $\Psi_0, \Psi_1, \dots, \Psi_N$, наддиагональными блоками $\Lambda_1, \dots, \Lambda_N$ и поддиагональными блоками M_0, \dots, M_{N-1} , где

$$\Psi_0 = -\lambda P_R; \quad (2)$$

$$\Lambda_1 = (\lambda_{0,j})_{(j|(1,j) \in X_1)} = \lambda p_j; \quad (3)$$

$$M_0 = (\mu_{i,0})_{(i|(1,i) \in X_1)} = \mu; \quad (4)$$

$$\Psi_n = (\psi_{i,j})_{(i,j|(n,i),(n,j) \in X_k)} \begin{cases} -(\lambda P_{R-i} + n\mu), & i = j; \\ 0, & i \neq j \end{cases} \quad (5)$$

при $n = \overline{1, N-1}$;

$$\Psi_N = (\psi_{i,j})_{(i,j)|(N,i),(N,j) \in X_N} \begin{cases} -N\mu, & i = j; \\ 0, & i \neq j; \end{cases} \quad (6)$$

$$\Lambda_n = (\lambda_{i,j})_{(i,j)|(n,i) \in X_{k-1}, (n,j) \in X_k} \begin{cases} \lambda p_{j-i}, & i \leq j \leq R; \\ 0 & \end{cases} \quad (7)$$

при $n = \overline{2, N}$;

$$M_n = (\mu_{i,j})_{(i,j)|(n,i) \in X_{k+1}, (n,j) \in X_k} \begin{cases} (n+1)\mu \frac{p_j^{(n)} p_{i-j}}{p_i^{(n+1)}}, & j \leq i \leq R; \\ 0 & \end{cases} \quad (8)$$

при $n = \overline{1, N-1}$.

Решив систему уравнений равновесия (СУР) с матрицей A , можно найти стационарное распределение процесса $X(t) = (\xi(t), \delta(t))$, где

$$q_{0,0} = \lim_{t \rightarrow \infty} P\{\xi(t) = 0, \delta(t) = 0\};$$

$$q_{k,i} = \lim_{t \rightarrow \infty} P\{\xi(t) = k, \delta(t) = i\},$$

$$1 \leq k \leq N, 0 \leq i \leq R.$$

На основании формул (2)–(8) СУР можно записать в следующем виде:

$$\lambda P_R q_{0,0} - \mu \sum_{j=0}^R q_{1,j} = 0;$$

$$(\lambda P_{R-j} + i\mu) q_{i,j} - \lambda \sum_{s=0}^j p_{j-s} q_{i-1,s} -$$

$$- (i+1)\mu \sum_{s=j}^R \frac{p_j^{(i)} p_{s-j}}{p_s^{(i+1)}} q_{i+1,s} = 0,$$

$$0 \leq j \leq R, 1 \leq i \leq N-1;$$

$$N\mu q_{N,j} - \lambda \sum_{s=0}^j p_{j-s} q_{N-1,s} = 0, 0 \leq j \leq R.$$

Введем подвекторы стационарных вероятностей $q_0 = q_{0,0}$, $q_i = (q_{i,j})_{j=\overline{0,R}}$, $i = 1, \dots, N$. Система уравнений равновесия в матричном виде с учетом блочно-тредиагонального вида матрицы интенсивностей переходов A выглядит следующим образом:

$$q_0 \Psi_0 - q_1^T M_0 = 0;$$

$$q_i^T \Psi_i - q_{i+1}^T M_i - q_{i-1}^T \Lambda_i = 0, i = 1, \dots, N-1;$$

$$q_N^T \Psi_N - q_{N-1}^T \Lambda_N = 0.$$

Матрица A является неразложимой консервативной трехдиагональной матрицей по построению, поэтому для решения СУР применим метод UL-разложения [5]

$$b = \sum_{k=0}^N b_k \sum_{i=k}^N q_{k,i},$$

где b_k — условное среднее число занятого ресурса при k заявках в системе:

$$b_k = \frac{\sum_{i=0}^N i p_i^{(k)}}{\sum_{i=0}^N p_i^{(k)}}.$$

Вероятность потери заявки из-за нехватки ресурса или свободных приборов для обслуживания определяется согласно формуле:

$$B = 1 - \sum_{k=0}^N \sum_{i=k}^N q_{k,i} \sum_{j=1}^{N-i} p_j. \quad (9)$$

Среднее число заявок в системе \overline{N} можно определить по формуле:

$$\overline{N} = \sum_{i=0}^N i q_i, \quad (10)$$

где $q_i = \sum_{j=0}^R q_{i,j}$.

В работе [4] было найдено аналитическое решение СУР (2)–(8) в общем виде. Для частного случая, когда объем выделяемых заявке ресурсов имеет дискретное распределение, авторами были получены более простые формулы для определения стационарных вероятностей:

$$q_{k,\bullet} = \lim_{t \rightarrow \infty} P\{\xi(t) = k\} = p_0 \frac{\rho^k}{k!} \sum_{i=0}^R p_i^{(k)}, 0 < k < N;$$

$$q_{k,j} = \lim_{t \rightarrow \infty} P\{\xi(t) = k; \delta(t) = j\} = p_0 \frac{\rho^k}{k!} p_j^{(k)},$$

$$0 \leq j \leq R, 0 < k \leq N,$$

где

$$p_0 = \left(1 + \sum_{k=1}^N \frac{\rho^k}{k!} \sum_{i=0}^R p_i^{(k)} \right)^{-1}.$$

Тогда средний объем занятых ресурсов b можно найти по формуле:

$$b = \sum_{k=0}^N b_k q_{k,\bullet} = p_0 \sum_{k=0}^N \frac{\rho^k}{k!} \sum_{i=0}^R i p_i^{(k)}. \quad (11)$$

Вероятность потери заявки определяется согласно формуле:

$$B = 1 - \sum_{k=0}^{N-1} \sum_{i=0}^R q_{k,i} \sum_{j=0}^{R-i} p_j = 1 - p_0 \sum_{k=0}^{N-1} \frac{\rho^k}{k!} \sum_{i=0}^R p_i^{(k+1)}. \quad (12)$$

3 Численный анализ

Рассмотрим различные частные случаи, когда объем ресурса, выделяемого заявке, имеет биномиальное, смещенное биномиальное и геометрическое распределение, а общий объем ресурса системы R совпадает с количеством приборов N .

Для вычисления вероятностно-временных характеристик модели (9)–(12) необходимо вычислить свертку распределения вероятностей $p_j^{(k)}$ по формуле (1) для выбранного дискретного распределения объема занимаемых заявкой ресурсов.

Рассмотрим частный случай, когда объем ресурса, выделяемого одной заявке $r_j \geq 0$, имеет биномиальное распределение

$$p_i = \binom{r}{i} p^i (1-p)^{r-i}, \quad 0 \leq i \leq r,$$

где $p = m/r$;

$$p_j^{(k)} = \binom{kr}{j} p^j (1-p)^{kr-j}.$$

Тогда распределение объема ресурса, который освободится после завершения обслуживания i -й заявки

$$p_{ij}^k = \frac{p_i p_{j-i}^{(k-1)}}{p_j^{(k)}} = \binom{r}{i} \binom{kr-r}{j-i} / \binom{kr}{j}, \quad 0 \leq i \leq r, \quad i \leq j, \quad 0 \leq j \leq kr,$$

зависит только от максимального числа r ресурса, доступного одной заявке.

Важно для приложений частном случае, когда объем выделяемого заявке ресурса больше нуля, $r_j > 0$ имеет смещенное биномиальное распределение

$$p_i = \binom{r-1}{i-1} p^{i-1} (1-p)^{r-i}, \quad 1 \leq i \leq r,$$

где $p = (m-1)/(r-1)$, $m = np + (1-p)$ и $\sigma^2 = rp(1-p)$;

$$p_j^{(k)} = \binom{k(r-1)}{j-k} p^{j-k} (1-p)^{kr-j}.$$

Соответственно, распределение

$$p_{ij}^k = \binom{r-1}{i-1} \binom{k(r-1)-(r-1)}{(j-k)-(i-1)} / \binom{k(r-1)}{j-k}, \quad 1 < i \leq r, \quad i \leq j, \quad k \leq j \leq k(r-1), \quad 1 < k \leq j.$$

Для сравнения полученных для первых двух частных случаев результатов рассмотрим пример, когда число единиц ресурса, выделяемого поступившей заявке, распределено согласно геометрическому закону:

$$p_i = p^i (1-p), \quad 1 \leq i \leq r,$$

где

$$m = \frac{1-p}{p}; \quad \sigma^2 = \frac{1-p}{p^2}; \quad p_j^{(k)} = \binom{k+j-1}{k} p^j (1-p)^k,$$

распределение

$$p_{ij}^k = \binom{k+j-i-2}{k-1} / \binom{k+j-1}{k}, \quad i \leq j, \quad k \leq j, \quad 1 < k \leq j.$$

В качестве исходных данных для численного анализа были выбраны значения $N = R = 100$, $\mu = 1$, $\lambda = \{12, 13, \dots, 20\}$. На рис. 5 представлены графики зависимости вероятностных характеристик от нагрузки ρ для выбранных дискретных распределений объема ресурса. Параметры распределений были подобраны таким образом, что математическое ожидание $m = 5,4$ и объем доступного одной заявке ресурса $r = 18$ для всех трех распределений, а дисперсии были различными.

На рис. 5, а заметим, что для случаев, когда выделяемые заявкам ресурсы имеют геометрическое распределение, среднее число заявок в системе при $\rho > 1$ растет быстрее, чем в случаях с биномиальными распределениями. График зависимости среднего числа занятого ресурса для геометрического распределения, наоборот, приближается к кривой для биномиального распределения. Так как геометрическое распределение имеет большую дисперсию, чем биномиальное и смещенное биномиальное распределения, то была исследована зависимость тех же вероятностных характеристик при различной нагрузке от дисперсии.

На рис. 6 представлены графики зависимости среднего числа заявок, среднего числа занятых ресурсов системы и вероятности блокировки СМО от дисперсии σ^2 . С ростом дисперсии геометрическо-

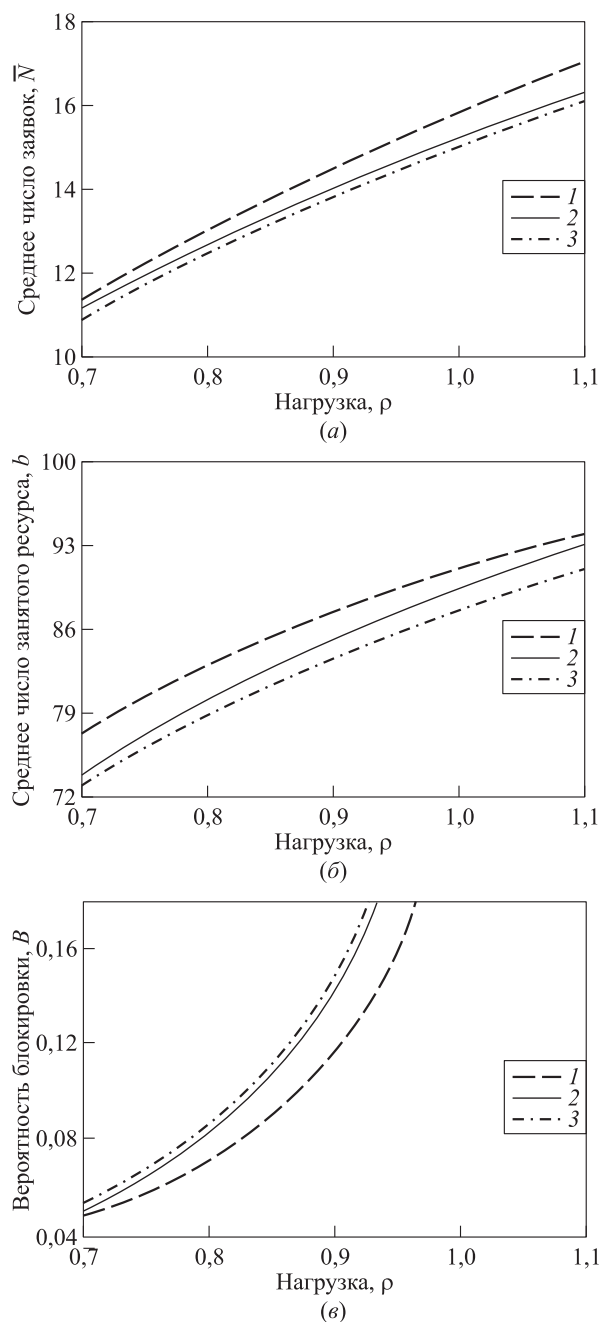


Рис. 5 Зависимость среднего числа заявок в системе (а), среднего числа занятого ресурса (б) и вероятности блокировки СМО (в) от нагрузки для геометрического (1); биномиального (2) и смещенного биномиального (3) распределений

го распределения среднее число занятых приборов увеличивается, а среднее число занятого ресурса уменьшается.

Заявки, которые требуют больше ресурсов для обслуживания, будут блокироваться, и в систему будет поступать больше заявок, требующих меньшего объема ресурса.

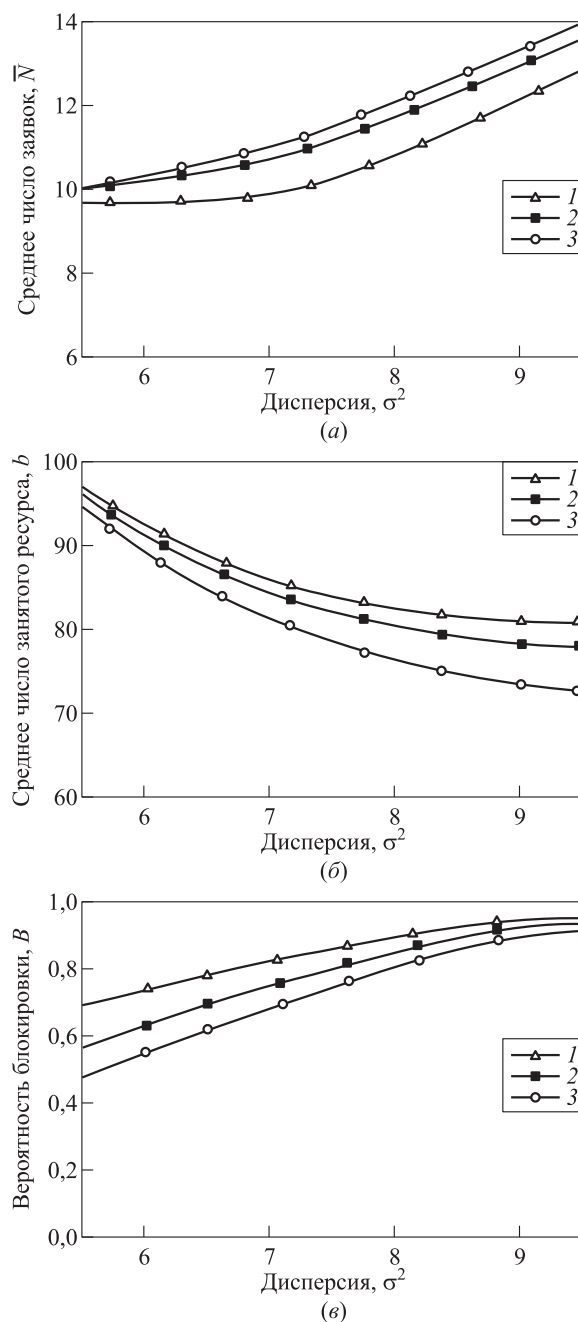


Рис. 6 Зависимость среднего числа заявок в системе (а), среднего числа занятого ресурса (б) и вероятности блокировки СМО (в) от дисперсии: 1 — $\rho = 0,9$; 2 — $0,8$; 3 — $\rho = 0,7$

4 Заключение

В работе исследуется упрощенная математическая модель установления сессий между мобильными устройствами конечных пользователей мультимедийных услуг и базовой станцией с распределением ограниченных ресурсов беспроводной

высокоскоростной сети LTE. Для расчетов предложены численный и аналитический способы расчета вероятностных характеристик системы.

Также был проведен сравнительный анализ зависимости вероятностных характеристик системы от параметров и типов распределений.

Было замечено, что при увеличении дисперсии геометрического распределения объема выделяемого поступившей заявке ресурса среднее число заявок в системе и среднее число занятого ресурса уменьшаются. В условиях нагрузки, близкой к единице, $\rho = \{0,7; 0,8; 0,9\}$, на обслуживание будут приниматься заявки, которые для обслуживания требуют наименьшего объема ресурсов. Более требовательные заявки будут блокироваться.

Дальнейшие исследования будут включать разработку эффективных вычислительных алгоритмов для расчета вероятностно-временных характеристик модели.

Литература

1. Pyattaev A., Johnsson K., Surak A., Florea R., Andreev S., Koucheryavy Y. Network-assisted D2D communications: Implementing a technology prototype for cellular traffic offloading // 2014 IEEE Wireless Communications and Networking Conference (WCNC). — Istanbul: IEEE, 2014. P. 3266–3271.
2. Наумов В. А., Самуйлов К. Е. О моделировании систем массового обслуживания с множественными ресурсами // Вестник РУДН. Сер. Математика. Информатика. Физика, 2014. № 3. С. 58–62.
3. Naumov V., Samouylov K., Sopin E., Andreev S. Two approaches to analysis of queuing systems with limited resources // 6th Congress (International) on Ultra Modern Telecommunications and Control Systems (ICUMT-2014) Proceedings. — СПб: IEEE, 2014. P. 485–488.
4. Наумов В. А., Самуйлов К. Е., Самуйлов А. К. О суммарном объеме ресурсов, занимаемых заявками // Автоматика и телемеханика, 2015 (в печати).
5. Наумов В. А. Численные методы анализа марковских систем. — М.: РУДН, 1985. 36 с.

Поступила в редакцию 5.10.15

ON PERFORMANCE ANALYSIS OF MODERN WIRELESS NETWORKS

O. G. Vikhrova¹, K. E. Samouylov¹, E. S. Sopin¹, and S. Ya. Shorgin²

¹Peoples' Friendship University of Russia, 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

Abstract: Analytics predict that worldwide mobile traffic growth rate will exceed fixed traffic approximately three times from 2014 to 2019. Number of mobile users will increase up to 4.9 billions and mobile devices number will exceed 10 billions. The average mobile network connection speed (1.7 Mbps in 2014) will reach nearly 4.0 Mbps by 2019. Special attention should be paid to mobile video traffic that will reach three-fourths of the whole mobile traffic by 2019. These tendencies bring new challenges for mobile communication providers to increase efficiency and additivity of radio resource allocation. In this connection, the paper analyzes a simplified model that allows one to obtain analytical estimates of the blocking probability and the average value of occupied resources according to the resource allocation policy of the LTE-Advanced technology.

Keywords: LTE-Advance; resource allocation policy; limited resource queue

DOI: 10.14357/19922264150405

References

1. Pyattaev, A., K. Johnsson, A. Surak, R. Florea, S. Andreev, and Y. Koucheryavy. 2014. Network-assisted D2D communications: Implementing a technology prototype for cellular traffic offloading. *Wireless Communications and Networking Conference (WCNC)*. Istanbul: IEEE. 3266–3271.
2. Naumov, V. A., and K. E. Samouylov. 2014. O modelirovani sistem massovogo obsluzhivaniya s mnozhestvennymi resursami [Modelling queuing systems with different types of resources]. *Vestnik RUDN. Ser. Matematika. Informatika. Fizika* [PFUR Bulletin, Mathematics, Informatics, Physics]. Moscow. 3:58–62.
3. Naumov, V., K. Samouylov, E. Sopin, and S. Andreev. 2014. Two approaches to analysis of queuing systems with limited resources. *6th Congress (International) on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) Proceedings*. St. Petersburg: IEEE. 485–488.
4. Naumov, V. A., K. E. Samouylov, and A. K. Samuylov. 2015 (in press). O summarnom ob"eme resursov, zanimaemykh zayavkami [Resources aggregation in queuing system]. *Automation and Remote Control*.
5. Naumov, V. A. 1985. *Chislennyye metody analiza markovskikh sistem* [Numerical analysis of Markov systems]. Moscow: PFUR Publ. 36 p.

Received October 5, 2015

Contributors

Vikhrova Olga G. (b. 1990) — PhD student, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; o.vikhrova@gmail.ru

Samouylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; ksam@sci.pfu.edu.ru

Sopin Eduard S. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, associate professor, Peoples' Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; esopin@sci.pfu.edu.ru

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

ОБЗОР МОДЕЛЕЙ И АЛГОРИТМОВ РАЗМЕЩЕНИЯ ЗАДАНИЙ В СИСТЕМАХ С ПАРАЛЛЕЛЬНЫМ ОБСЛУЖИВАНИЕМ*

М. Г. Коновалов¹, Р. В. Разумчик²

Аннотация: Представлен обзор зарубежных публикаций по проблеме эффективного распределения заданий в системах обслуживания (проблема диспетчеризации). Отобранные для обзора модели касаются систем с независимыми параллельными безотказными серверами и случайными потоками заданий, не имеющих внутренней структуры. Общая трактовка проблемы сформулирована как задача оптимизации, которая имеет многочисленные варианты в зависимости от дополнительных предположений. Изложение моделей следует классификации этих предположений по объему и характеру априорной информации, возможностям наблюдения в процессе функционирования системы и критериям эффективности. Приведено описание основных отмеченных в литературе и используемых на практике алгоритмов диспетчеризации и их сравнительных свойств. Отмечены основные методы, используемые при анализе моделей и построении стратегий размещения заданий. Обзор призван акцентировать внимание на одной из важных и далеких от окончательного разрешения проблем обработки больших объемов информации.

Ключевые слова: системы с параллельным обслуживанием; стратегии размещения заданий; диспетчеризация

DOI: 10.14357/19922264150406

1 Введение

Многие современные технические системы представляют собой параллельно работающие обслуживающие ресурсы — серверы, на которые поступают заявки на выполнение заданий. Каждое задание может быть выполнено на каком-либо сервере, однако выбор заранее однозначно не определен. Требуется построить процедуру выбора сервера для выполнения задания. Необходимость в таких процедурах возникает в разнообразных компьютерных, сетевых, производственных и иных системах, т. е. практически везде, где есть необходимость или возможность параллельного обслуживания. Появление новых возможностей (услуг, сервисов и др.) влечет усложнение техники, оставляющей эти возможности, и одновременно рост числа пользовательских запросов. Качественная обработка запросов невозможна без управляющих механизмов, которые определяют, какой запрос, где и когда должен быть выполнен. Приведем несколько примеров реальных систем, в которых возникает потребность в эффективных алгоритмах размещения заданий:

- распределенные компьютерные системы передачи, обработки и хранения данных: телеком-

муникационные сети, грид- и клауд-системы, базы данных;

- узлы, представляющие собой совокупности серверов, соединенных сетью передачи данных и работающих как единое целое (веб-фер, центры обработки данных, компьютерные кластеры);
- разнородные объекты: суперкомпьютеры, веб-провайдеры, машины с распределенной памятью, хеш-таблицы, когнитивные радиосистемы;
- человеко-машинные системы: транспортные, ремонтные и иные обслуживающие системы.

В таких системах механизмы размещения заданий являются неотъемлемой составной частью, причем для соблюдения требований к качеству обслуживания распределение заданий должно происходить как можно быстрее. Например, компании Google и Amazon сообщили, что увеличение среднего времени отклика для веб-поиска на 500 мс приводит к потере 1,2% пользователей [1].

Важность рассматриваемой проблемы подтверждается большим числом публикаций, в настоящее время почти исключительно зарубежных. Сама же проблема является сложной, и относящиеся к ней

* Работа выполнена при частичной поддержке РФФИ (проекты 15-07-03406 и 13-07-00223).

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

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Российский университет дружбы народов, grazumchik@ipiran.ru

задачи чаще всего не удается решить исключительно математическими методами. Тем не менее, признавая практическую ценность и плодотворность эвристических идей и инженерного подхода, для настоящего обзора отбирались работы, в которых математическая модель все же присутствует. Во всех отобранных моделях рассматриваемые системы состоят из нескольких параллельных и независимых серверов; их функционирование рассматривается без учета возможных отказов; отсутствует внутренняя структура у поступающих заданий. Общая постановка задачи сформулирована в разд. 2. В разд. 3 содержится описание наиболее известных и употребительных на практике алгоритмов. Изучаемые системы разделяются на два класса в зависимости от того, составлены ли они из одинаковых или неодинаковых серверов. Обзор моделей, относящихся к этим классам, помещен соответственно в разд. 4 и 5.

2 Общая задача о размещении заданий

Рассматривается модель системы, состоящей из N параллельных серверов $\mathbb{R}_1, \dots, \mathbb{R}_N$, где \mathbb{R} — набор параметров n -го сервера. В систему поступают задания в моменты, порожденные входными потоками $\mathbb{F}_1, \dots, \mathbb{F}_M$, где \mathbb{F}_m — набор параметров, определяющий m -й поток. В момент поступления задание направляется в очередь одного из серверов, и в дальнейшем, если не оговорено противное, будем считать, что переход заявок из одной очереди в другую невозможен. Каждый сервер работает с поступившими заданиями автономно от остальной части системы.

В число параметров сервера входят дисциплина обслуживания заданий в очереди (английские термины: *scheduling, service discipline*), а также характеристики его производительности. В число параметров потока входят характеристики случайного процесса, задающего моменты поступления заданий, а также характеристики заданий, поступающих из этого потока.

Время пребывания задания в системе совпадает со временем его пребывания на выбранном сервере и складывается из времени ожидания в очереди и времени (непосредственного) выполнения. Время ожидания зависит от дисциплины обслуживания. Время выполнения задания зависит от характеристик задания и сервера. Таким образом, рассматриваемая модель задается набором параметров $\mathbb{S} = (\mathbb{F}, \dots, \mathbb{F}_M; \mathbb{R}, \dots, \mathbb{R}_N)$.

Распределение заданий по серверам в системе \mathbb{S} осуществляется согласно стратегии размещения за-

даний, которую будем называть также диспетчеризацией (*dispatching*) и которая является основным предметом обсуждения в этой статье.

Система эволюционирует в непрерывном времени. Ее полная предыстория складывается из моментов поступления заданий, из принятых решений о выборе сервера и из моментов начала и окончания выполнения заданий. Пусть h_k — наблюдаемая часть полной предыстории системы к моменту поступления k -го по счету задания. Диспетчеризация в самом общем виде представляется набором правил назначения сервера для поступающего задания в зависимости от наблюдаемой предыстории:

$$\sigma = \{\sigma_k(h_k), k = 1, 2, \dots\},$$

где правило $\sigma_k(h_k)$ — это вероятностное распределение на множестве $\{1, \dots, N\}$. Пусть Ξ — заданное подмножество систем \mathbb{S} ; Σ — заданное подмножество диспетчеризаций σ , которое назовем множеством допустимых стратегий размещения; $W = W(\mathbb{S}, \sigma)$ — заданный числовой показатель, который характеризует затраты, связанные с применением диспетчеризации σ к системе \mathbb{S} .

Проблему диспетчеризации можно сформулировать как задачу минимизации, т. е. нахождения такой стратегии размещения σ^* , которая обеспечит бы выполнение неравенства

$$W(\sigma^*) \leq \inf_{\Sigma} W(\sigma) + \varepsilon$$

для любой системы из множества Ξ и для заданного значения $\varepsilon \geq 0$.

Задание множества Ξ равносильно указанию априорной информации о системе, которая имеется при конструировании стратегии размещения заданий. Например, если множество Ξ состоит из одного элемента, то это означает, что при построении диспетчеризации можно считать известными все параметры системы \mathbb{S} . Напротив, неполнота априорной информации о параметрах из набора \mathbb{S} означает, что диспетчер, размещающий задания, может иметь дело с любой из систем в диапазоне неопределенности, заданном множеством Ξ , и при этом не знать в точности, каким именно объектом он управляет. Согласно определению, оптимальная диспетчеризация σ^* должна в этом случае обладать определенной «инвариантностью» по отношению к множеству систем Ξ .

Задание множества Σ тесно связано с описанием возможностей наблюдения в процессе функционирования системы. Например, нельзя считать допустимой стратегию, опирающуюся на длины очередей, если эти очереди невозможно отслеживать в реальном времени. Однако происхождение множества Σ вызвано не только фактором наблюдаемости. Ограничение множества допустимых стра-

тегий может быть вызвано техническими или иными дополнительными соображениями, которые могут потребовать, скажем, использовать только стратегии с минимальным объемом сопутствующих вычислений и ограниченными требованиями к памяти.

Конкретные задачи диспетчеризации содержат дополнительные предположения, которые фактически выделяют множества Ξ и Σ , обычно не указывая эти множества явно. Перечислим основные условия, которые обычно накладываются на общую модель, с тем чтобы получить более доступную для решения задачу.

Последовательность моментов поступления заданий предполагается чаще всего порожденной единственным пуассоновским процессом. В более общем варианте рассматривается рекуррентный поток.

Задания, как правило, характеризуются функцией распределения F , определяющей размер задания или время его выполнения. Относительно функции F делаются дополнительные предположения.

Серверы часто представляют собой готовые технические изделия со встроенной схемой управления очередью, которую нельзя легко изменить. Наиболее употребительными дисциплинами обслуживания очереди являются¹:

- FIFO (First In First Out);
- LIFO (Last In First Out) или LIFO-PR (Last In First Out Preemptive Resume);
- PS (Processor Sharing).

Производительность сервера характеризуется иногда распределением времени выполнения задания, а иногда скоростью $u > 0$. В последнем случае время выполнения задания размером s равно s/u .

Большая вариативность имеет место в части предположений о возможностях наблюдения. Диапазон этих предположений простирается от полной наблюдаемости до фактически полной невозможности что-либо наблюдать, кроме моментов выбора серверов (или, что то же самое, кроме моментов поступления заданий) и номеров выбора. Впрочем, условия на наблюдаемость часто вообще явно не формулируются, а только отражаются в том объеме информации, которую используют рассматриваемые стратегии.

Хотя сформулированную оптимизационную задачу можно назвать основной в проблематике диспетчеризации, ее постановка, разумеется, не исчерпывает всех вопросов, которые возникают в рассматриваемой области. Существенное вни-

мание уделяется также, например, анализу стационарных характеристик системы, которые имеют место при той или иной стратегии размещения заданий.

Надо также отметить, что изложенный взгляд на проблему диспетчеризации с чисто математической точки зрения является до некоторой степени идеализированным, поскольку формальная задача, как правило, весьма сложна как для аналитического, так и для численного решения. Этим объясняется малое число работ, в которых оптимальная диспетчеризация была бы найдена как решение задачи оптимизации (по этому поводу см., например, краткий обзор в [2]). Гораздо чаще используется «инженерный» подход, когда вначале на основании, как правило, интуитивно ясных эвристических соображений создаются алгоритмы размещения заданий, а затем исследуются и сравниваются их свойства, в том числе оптимизационные. По этой причине обзор методов диспетчеризации начинается в следующем разделе с описания наиболее употребительных алгоритмов.

3 Основные стратегии размещения заданий

Простейшая диспетчеризация Random состоит в том, что каждый раз сервер выбирается равновероятно. Она не использует никакой информации о системе — ни априорной, ни текущей (кроме значения N). Диспетчеризация Random является частным случаем более общей диспетчеризации PAP (Probabilistic Allocation Policy), которая параметризована дискретным распределением $p = (p_1, \dots, p_N)$. Задание направляется на сервер n с вероятностью p_n . Реализация стратегии не требует наблюдений. Однако для того, чтобы найти, хотя бы приближенно, наилучшее распределение p , надо знать априорную информацию о параметрах заданий и серверов. Известен также подход, в котором PAP оптимизируется в процессе работы системы, но в этом случае необходимы дополнительные наблюдения за траекторией системы. Обе стратегии Random и PAP относятся к классу, который в литературе часто называется Bernoulli splitting и характеризуется тем, что для принятия решения не нужно знания текущего состояния системы. Естественным обобщением стратегии Bernoulli splitting для систем с несколькими входящими потоками является стратегия MCPAP (MultiClass Probabilistic Allocation Policy), когда задание класса i направляется на сервер j с вероятностью p_{ij} .

¹ Более широко класс наиболее часто рассматриваемых дисциплин можно определить так: рассматриваются консервативные (сохраняющие работу) дисциплины, которые не допускают прерывания уже выполняемого задания.

Детерминированным аналогом Random является диспетчеризация Round-Robin, которая действует циклически, направляя k -е задание на сервер $k \bmod N$. Для ее реализации необходимо помнить предыдущий выбор.

Диспетчеризация JSQ (Join the Shortest Queue) направляет задание на сервер с минимальной очередью. Диспетчеризация JSQ(d) действует аналогичным образом, но не на всем множестве серверов, а в группе из случайно выбранных d серверов. Если же группа формируется так: сначала случайным образом формируется группа серверов размера $M \leq N$, а затем из этой группы наугад выбираются d серверов, то диспетчеризация называется HJSQ(d) (Hybrid JSQ(d)).

Диспетчеризация MEST (Minimum Expected Service Time) направляет задание в конец очереди с наименьшим ожидаемым временем начала обслуживания. Эта стратегия неявно предполагает, что размер задания становится известным только при его поступлении на обработку на сервер и требует, чтобы была доступна следующая информация:

- размер очереди, т. е. число заданий, ожидающих выполнения;
- производительность серверов;
- распределение времени обслуживания задания;
- среднее остаточное время выполнения задания при условии, что оно уже выполняется в течение некоторого времени.

Динамическая диспетчеризация MERL (Minimum Expected Residual Load), также известная, как LWL (Least Work Left), направляет задание на сервер с наименьшим объемом невыполненной нагрузки, т. е. с минимальной остаточной работой, включая заявки в очереди. Стратегия неявно предполагает, что размер задания становится известным в момент его поступления в систему, и в остальном требует той же информации и тех же наблюдений, что и стратегия MEST.

Диспетчеризация Муорис предполагает, что вновь поступающее задание поступает в ту очередь, которая минимизирует время, необходимое для освобождения от работы всей системы целиком, в предположении, что в дальнейшем задания в систему не поступают.

Три предыдущие стратегии, по сути, используют одинаковую идею: посылать задание туда, где «лучше» с сиюминутной точки зрения. Они представляют собой разновидности из известного семейства «жадных» алгоритмов.

Диспетчеризация SITA-E (Size Interval Task Assignment with Equal Load) устроена так, что каждый

сервер обслуживает задания, размер которых попадает в назначенный для данного сервера диапазон. Выбор диапазонов для серверов осуществляется так, чтобы средняя нагрузка была одинакова. Более точно, пусть область определения функции F сосредоточена в диапазоне от x_0 до x_h . Тогда разбиение $0 \leq x_0 < x_1 < \dots < x_h \leq \infty$ выбирается так, чтобы выполнялись равенства:

$$\int_{x_{i-1}}^{x_i} dF(x) = \frac{1}{h} \int_{x_0}^{x_h} x dF(x).$$

Диспетчеризация SITA-V (Size Interval Task Assignment with Variable Load) является модификацией стратегии SITA-E. Ее реализация на примере системы из двух серверов выглядит следующим образом. Пусть размер заданий принимает значения из отрезка $[k, p]$ и пусть x — некоторая точка из этого отрезка. Маршрутизация происходит таким образом, что задания, размер которых меньше x , направляются на первый сервер, а остальные задания — на второй сервер. Таким образом, идея заключается в том, чтобы рассредоточивать многочисленные короткие задания и редкие длинные задания. Отметим, что при этом нарушается так называемый принцип балансировки нагрузки, который соблюдается в той или иной степени в стратегиях Random, Round-Robin, MERL и SITA-E. Рандомизированная балансировка является очень простой и одновременно эффективной с точки зрения уменьшения коллизий, времени ожидания, очередей и т. д.

Если вместо размеров заданий рассматривается их «значимость» (т. е. некоторая числовая характеристика), то диспетчеризация, устроенная по принципу SITA, носит название VITA (Value Interval Task Assignment). Другими диспетчеризациями, использующими информацию о значимости заданий (value of job), являются C-MU и LAVA (Length and Value Aware). Обе диспетчеризации для принятия решения используют информацию только о находящихся в момент принятия решения заданиях в системе. Согласно диспетчеризации C-MU задание направляется на сервер с максимальным значением произведения $c_i \mu_i$, где c_i — «стоимость хранения» (holding cost), а μ_i — интенсивность обслуживания на сервере i . Точнее, задание с характеристикой значимости ν направляется на сервер $\operatorname{argmin}_i (\bar{\nu}_i - \nu) \mu_i / (n_i + 1)$, где $\bar{\nu}_i = \sum_{j=1}^{n_i} \nu_{i,j} / n_i$, $\nu_{i,j}$ — значимость выполнения задачи j на сервере i , n_i — число заявок на сервере i в момент принятия решения. Если же задание с характеристикой ν направляется на сервер $\operatorname{argmin}_i \left((n_i + 2) \nu + \sum_{j=1}^{n_i} \nu_{i,j} \right) / (2 \mu_i)$, то это диспетчеризация LAVA.

Диспетчеризация TDP (Threshold Dispatching Policy) предполагает использование «порогового» правила для принятия решения. Согласно TDP в момент поступления задания оно направляется на сервер с номером $\underset{i}{\operatorname{argmin}}(u_i + \xi_i)$, где u_i — время (начиная с момента поступления), необходимое для завершения всей имеющейся работы на сервере i , а ξ_i — некоторые константы (пороги). От значения порогов ξ_i и зависит значение целевой функции. Отметим, что применение данного класса пороговых диспетчеризаций можно часто наблюдать на практике (например, в оборудовании сетей связи). Однако доказать его оптимальность удастся лишь в некоторых весьма частных случаях.

Диспетчеризация FPI (First Policy Iteration), в отличие от всех изложенных выше, основана не на интуитивных соображениях, а на методах марковского процесса принятия решений. Для ее построения используется первый шаг итеративного алгоритма улучшения марковской стратегии. За начальное приближение берется одна из диспетчеризаций, например Random, или VITA, или любая другая, при которой входные потоки на серверы остаются пуассоновскими. Последнее обстоятельство позволяет выписать явные формулы, необходимые для одного итеративного шага (последующие шаги технически невыполнимы). Согласно FPI поступившая заявка направляется в узел $\underset{i}{\operatorname{argmin}}(\gamma \lambda_i M \nu_i (n_i + 1) / (\mu_i - \lambda_i) + \sum_{j=1}^{n_i} \nu_{i,j} + (n_i + 2) \nu) / (2\mu_i - \lambda_i)$, где γ — коэффициент дисконтирования.

Следующие четыре стратегии допускают переход между очередями для заданий и прерывание обслуживания.

Диспетчеризация TAGS (Task Assignment by Guessing Size) каждому серверу назначает допустимый временной интервал ожидания обслуживания, т. е. интервал, в течение которого задание может ожидать обработки. Длины интервалов различны для разных серверов. Если задание не выполнено в течение отведенного времени, то его выполнение прерывается и задание отправляется в очередь другого сервера, на котором будет выполняться заново. Подробнее, пусть Δ_n — время, которое задание может провести на сервере n , включая ожидание в очереди и непосредственное выполнение, и пусть $\Delta_1 < \Delta_2 < \dots < \Delta_N$. Все поступающие в систему задания сначала отправляются на сервер 1. Если задание не выполнено за время Δ_1 , то оно направляется на сервер 2. Там оно может пробыть время, не превышающее Δ_2 , и в случае невыполнения отправляется на сервер 3 и т. д.

Диспетчеризация TAPTF (Task Assignment based on Prioritizing Traffic Flows) является модификацией

стратегии TAGS. Для этой стратегии аналогичным образом определяются числа Δ_n . Но у каждого сервера имеется не одна, а две очереди: основная и вспомогательная (так называемая очередь повторных заданий). Когда новое задание поступает в систему, оно немедленно направляется в основную очередь одного из серверов в соответствии с некоторым статическим правилом. Если сервер не успел выполнить задание за отведенный интервал времени, задание перемещается во вспомогательную очередь того сервера, у которого допустимый интервал ожидания выше. При этом задания в основной очереди имеют более высокий приоритет обслуживания, чем во вспомогательной.

Диспетчеризация TAPTF-WC (Task Assignment based on Prioritizing Traffic Flows with Work-Conserving migration) отличается от предыдущей тем, что задания из очереди повторных заданий выполняются с того места, на котором было прервано их обслуживание на предыдущем сервере.

Диспетчеризация MTTMEL (MultiTier Task assignment policy with Minimum Excess Load) является обобщением предыдущих трех стратегий. В этом алгоритме все серверы разбиваются на группы различного размера. Допустимый временной интервал ожидания обслуживания теперь назначается не каждому серверу в отдельности, а сразу на всю группу. Новое задание направляется на один из серверов группы наименьшего размера. Если там оно не выполнено за отведенное время, то оно переходит на один из серверов следующей по размеру группы и т. д.

4 Модели с одинаковыми серверами

Наиболее изучены в теоретическом отношении задачи, в которых все серверы идентичны по своим параметрам. В этом разделе представлены несколько моделей с одинаковыми серверами, которые отличаются дисциплинами обслуживания очередей, функцией распределения времени выполнения заданий F и — главное отличие — возможностями наблюдения. Каноническим методом для эффективного распределения заданий между ресурсами в случае одинаковых серверов является балансировка нагрузки. Согласно этому принципу действуют стратегии Random, Round-Robin, JSQ, JSQ(d), MERL и SITA-E.

4.1 Наблюдаемые очереди

4.1.1 Случай, когда неизвестен размер заданий, приводит к системам, носящим иногда название

«модели супермаркета». Эти модели изучались в [3–10] при различных дополнительных условиях. Существующая методология исследования заключается в следующем:

- система описывается с помощью марковского процесса, для которого устанавливается наличие предельного распределения;
- система описывается с помощью дифференциальных уравнений, для которых исследуют вопросы существования и единственности решения, дающего стационарное распределение очередей.

Такой подход становится затруднителен, когда распределения времен обслуживания не экспоненциальные, поскольку приходится усложнять пространство состояний марковской цепи. В [10] предлагается способ выхода из затруднения. В этой работе дополнительно предполагается, что

- интенсивность пуассоновского входящего потока прямо пропорциональна числу серверов;
- дисциплина обслуживания очереди одинаковая для всех серверов;
- функция распределения времени выполнения задания произвольная;
- наблюдаемы размеры очередей.

Основой подхода является диспетчеризация JSQ(d) и гипотеза об асимптотической независимости размеров очередей при больших значениях N . Для системы с дисциплиной обслуживания очереди LIFO-PR и функцией распределения времени обслуживания общего вида с помощью предложенного подхода доказано, что при $n \rightarrow \infty$ вероятность P_k того, что в очереди к любому серверу будет не менее k заданий, удовлетворяет асимптотическому соотношению $P_k \sim \lambda(d^k - 1)/(d - 1)$. Для системы с дисциплиной обслуживания очереди FIFO зависимость P_k от d имеет пороговый характер. При различных значениях d характер зависимости может быть полиномиальным, экспоненциальным и дважды экспоненциальным. При неограниченном росте числа серверов совместное распределение числа заявок в очередях (если используются дисциплины обслуживания PS или LIFO-PR) являются инвариантными относительно распределения времени обслуживания. В вычислительном отношении использованный в [10] подход позволил уменьшить на два порядка время для имитационного моделирования.

4.1.2 Известны характеристики времени выполнения заданий, в частности доступны оценки оста-

точных времен выполнения заданий [11, 12]. Дополнительные предположения:

- входной поток — пуассоновский;
- дисциплина обслуживания очереди — FIFO;
- интенсивность обслуживания $h(t) = f(t)/(1 - F(t))$ является неубывающей функцией.

Анализ диспетчеризации JSQ для системы из параллельных серверов проведен в [14, 13] в предположении пуассоновского входного потока и экспоненциального распределения времени обслуживания. В этих предположениях ее оптимальность доказана в [11] для серверов с дисциплиной FIFO. В [15] JSQ анализируется для обслуживания с дисциплиной PS, где показано (с помощью различного рода приближений и численных экспериментов), что в системах серверов с произвольным распределением времени обслуживания схема JSQ хотя и не является оптимальной в смысле минимума задержки, но имеет хорошую производительность в сравнении с другими, более сложными стратегиями. Оптимальность JSQ для некоторых более сложных систем доказана в [16].

В [12] изучается диспетчеризация MEST, которая оказывается оптимальной в смысле стохастического упорядочивания. Это означает, что она максимизирует вероятность того, что для произвольных k и t не менее чем k заявок будут обслужены за время t .

4.1.3 Модели с известным размером заданий (или с известным временем выполнения заданий) ценны тем, что в них делаются реалистичные предположения о функции распределения размера задания. Данное предположение имеет практическую предпосылку. В некоторых случаях (например, в сетевых запросах на скачивание файлов) удается оценить распределение размеров (времен выполнения) поступающих заданий [17]. Анализ статистики в реальных системах показал, что распределение длины задач в современных информационно-телекоммуникационных системах хорошо описывается степенной функцией $x^{-\alpha}$, где $0,9 \leq \alpha \leq 2$ [17–21].

Стратегия «наименьшей очереди» (JSQ), применяемая при неизвестных размерах задач, не всегда является оптимальной [22], как, например, в случае распределения времени обслуживания с тяжелым хвостом и с большой дисперсией. В [17, 23] проведен подробный сравнительный анализ нескольких диспетчеризаций для таких функций F при следующих дополнительных предположениях:

- входной поток — пуассоновский;
- дисциплина обслуживания очереди — FIFO;

- распределение времени обслуживания F имеет плотность

$$f(x) = \frac{\alpha k^\alpha}{1 - (k/p)^\alpha} x^{-\alpha-1}, k \leq x \leq p;$$

- размер задания известен в момент поступления.

Рассмотрены следующие диспетчеризации: Random, Round-Robin, Dynamic, SITA-E и SITA-V. Выводы относительно производительности стратегий сделаны на основании методов теории массового обслуживания и результатов численных экспериментов по критериям среднего времени выполнения задач и среднего значения для отношения времени выполнения заданий к их размеру. Последний критерий считается важным в рассматриваемых моделях и носит специальное название: замедление (slowdown). Лейтмотивом всего исследования служит то очевидное соображение, что разные критерии (разные целевые функции) могут требовать разных стратегий. Установлены, в частности, следующие факты.

Первые две стратегии обладают свойством выравнивать нагрузку. Однако если тестировать стратегии для различных времен между поступлениями, то результаты для циклической стратегии обладают меньшим разбросом, чем результаты для стратегии Random.

Привязка сервера к определенному диапазону размера задач позволяет избежать дискриминации коротких задач. Динамическая стратегия «оптимальна» с точки зрения отдельной задачи, а также с точки зрения мгновенного выравнивания нагрузки во всей системе. Если дисперсия размера задач не очень высока, то предпочтительнее динамическая стратегия, в противном случае лучшей является стратегия SITA-E.

Различие между четырьмя стратегиями может быть велико. Стратегии Random и Round-Robin хуже, чем SITA-E и Dynamic на несколько порядков. SITA-E превосходит Dynamic почти в два раза.

Преимущество SITA-E объясняют следующим образом:

- уменьшается дисперсия поступающих на отдельный сервер задач;
- если нагрузка сбалансирована, то большинство задач направляется на сервер с «лучшими условиями»;
- улучшается среднее значение замедления.

Стратегия SITA-E улучшает замедление, однако при этом увеличивает среднее время обслуживания. Вычислительные эксперименты показали целесообразность применения алгоритма SITA-V на веб-фермах.

Диспетчеризации SITA-E и SITA-V являются представителями целого класса стратегий SITA, которые регулируют размер заданий, поступающих на отдельные серверы (более подробно см. [24, 25]).

4.2 Ненаблюдаемые очереди

При неизвестном заранее времени обслуживания диспетчеризация может быть либо самой простой (вроде Random), либо может использоваться каким-то достаточно сложным образом наблюдения за реальным временем выполнения заданий. В [26–29] следуют вторым путем. В этих работах в предположении, что все очереди управляются согласно дисциплине FIFO, последовательно предложены четыре нетривиальных диспетчеризации (все на основе эвристических соображений): TAGS, TAPTF, TAPTF-WC и MTTMEL. Во всех четырех случаях необходимо задать значения Δ_n времени, которое может провести задание на сервере n , включая ожидание в очереди и непосредственное выполнение (описание диспетчеризаций — см. разд. 3). Эти значения, выбранные изначально наугад или по экспертной оценке, актуализируются путем наблюдения как за системой, так и за поведением выбранной целевой функции. Оптимизация заключается в выборе значений Δ_n , при которых целевая функция достигает значения, близкого к оптимальному.

Упомянутые диспетчеризации снижают дисперсию времени выполнения заданий. На практике это означает, что на каждом сервере скапливаются примерно одинаковые по времени выполнения задания. Хорошие результаты обнаруживаются для функций F , имеющих тяжелые хвосты (при этом существенно, что времена обслуживания неизвестны). Однако имеются и серьезные с практической точки зрения недостатки. Реализация таких стратегий связана с «накладными расходами»: необходимы механизмы, обеспечивающие аккуратное прерывание заданий, перенаправление на другие серверы и возобновление обслуживания. Кроме того, возможны плохие результаты при высокой загрузке системы. Например, TAGS вносит несправедливость в распределение заданий, так как длинные задания чаще прерываются.

5 Модели с неодинаковыми серверами

5.1 Наблюдаемые очереди

В [30] исследуется система, в которой каждое задание характеризуется парой независимых в со-

вокупности случайных величин (X, V) , где X — размер (время обслуживания), имеющий экспоненциальное распределение, V — значимость задачи (некоторая числовая характеристика), распределение которой задано заранее. Значение V наблюдается в момент поступления задания в систему. Напротив, значения X не наблюдаются¹. В работе с помощью аналитических методов теории массового обслуживания, численных методов и имитационного моделирования проведен сравнительный анализ диспетчеризаций, оптимизирующих среднее значение произведения VT , где T — время пребывания задания в системе. В экспериментах диспетчеризации Random и VITA уступали JSQ, в то время как C-MU и LAVA превосходили ее. Чаше всего наилучшие результаты показывала LAVA. Однако применение более сложной диспетчеризации FPI позволило улучшить LAVA, особенно заметно в случае большой загрузки. Аналитический раздел [30] содержит ряд утверждений, касающихся режимов устойчивости диспетчеризаций, а также сравнения LAVA с JSQ в частном случае двух одинаковых серверов.

В [31] рассматривается система, в которую могут поступать несколько пуассоновских потоков задач интенсивности λ_i , $i = 1, \dots, M$. Поступив на сервер, они обслуживаются одинаково, вне зависимости от типа потока. Каждый сервер j описывается функцией $D_j = D_j(\gamma_j)$, которая характеризует «затраты» на обслуживание на нем задания, включая ожидание в очереди, в зависимости от нагрузки γ_j на этот сервер. С каждым классом заданий i связано число β_i , которое является количественной (экспертной) оценкой того, насколько критично качество обслуживания для заданий класса i . В работе показано с помощью методов математического программирования, что диспетчеризация MCFAP доставляет оптимальное значение целевой функции:

$$W(P) = \sum_{i=1}^M \sum_{j=1}^N \beta_i \lambda_i p_{ij} D_j(\gamma_j).$$

В связи с интересом к децентрализованному управлению наряду с минимизацией целевой функции рассматривается и игровая постановка. Целью обслуживания заданий из потока i объявляется выбор такого сервера j , который минимизировал бы собственные затраты $V_{ij} = c_j + \beta_i D_j$ ($c_j > 0$ — заданные коэффициенты). Вводится понятие равновесной стратегии, обладающей свойством $\forall i, j, k : p_{ij} > 0 \Rightarrow V_{ij} \leq V_{ik}$. Если задания размещаются в соответствии с равновесной стратегией, то сервер, на который направляются задания с положительной

вероятностью, должен обходиться не дороже, чем любой другой сервер. Равновесная стратегия всегда существует как частный случай равновесия по Нэшу.

Похожая постановка задачи рассмотрена в [32]. Основной результат работы тот же: оптимальной с точки зрения минимизации функции типа $W(P)$ является диспетчеризация MCFAP. Кроме того, эксперименты показали, что при росте нагрузки имеет значение дисциплина обслуживания на серверах. Так, дисциплина, в соответствии с которой заявки на сервере переупорядочиваются некоторым определенным образом, оказалась лучше, чем FIFO.

Исследование систем, в которых имеется возможность заранее задать производительность c_i каждого сервера (обслуживающего задания по дисциплине PS), проведено в [1]. Задание, направленное на сервер i , обслуживается на нем со скоростью $c_i/x_i(t)$, где $x_i(t)$ — общее число задач на этом сервере в момент t . Используя в качестве целевой функции среднее время пребывания задания в системе, авторы показали, что нахождение оптимальной диспетчеризации сводится к задаче математического программирования. В качестве наилучших вариантов предлагается использовать MCFAP, JSQ и HJSQ.

В [33] рассмотрена модель, в которой точные размеры (будущие времена обслуживания) заданий предполагаются известными в моменты их поступления в систему. Целевая функция — среднее время пребывания в системе. Для случая двух серверов доказано, что оптимальная диспетчеризация найдется среди стратегий TDP. Заметим, что в случае одинаковых серверов оптимальной является стратегия LWL вне зависимости от числа серверов [34, 35]. Для нахождения порога предложен алгоритм типа FPI. Для большего числа серверов вопрос оптимальности TDP не исследован, но эксперименты показывают, что TDP может давать результаты, близкие к оптимальным. Однако вопрос нахождения точных пороговых значений остается открытым [36].

5.2 Ненаблюдаемые очереди

В [37] рассмотрена модель с известными скоростями серверов и размерами поступающих заданий. При этом нет возможности наблюдать помимо очередей еще и остаточные (прошедшие) времена обслуживания заданий на серверах. Дисциплина обслуживания на каждом сервере — FIFO. Опи-

¹На самом деле, даже если они наблюдаются, то в силу предположения об экспоненциальности размеров заданий не несут никакой информации из-за свойства отсутствия последствия.

сан класс легко реализуемых статических диспетчеризаций, позволяющих достигать близких к оптимальным значений времени ожидания начала обслуживания (и времени пребывания в системе). В работе сравниваются диспетчеризации Random, SITA и NSI (Nested Size-Interval). Последний вариант диспетчеризации, NSI, отличается от SITA лишь тем, что интервалы, которые определяют, задания какого размера (времени выполнения) поступают на каждый сервер, могут как перекрываться, так и быть вложенными друг в друга. Диспетчеризация SITA, являясь оптимальной среди всех статистических стратегий в случае одинаковых серверов, уже не является таковой в случае различных серверов. В [37] доказано, что в данном случае оптимальная статистическая стратегия находится в классе диспетчеризаций NSI, и предложен метод ее нахождения, использующий аппарат теории массового обслуживания (результаты для системы массового обслуживания типа $M/G/1$).

6 Заключение

Проведенный обзор свидетельствует о наличии заметного раздела современной науки, который сформировался в связи с проблемами развития информационных технологий. Многообразие реальных приложений и ситуаций порождает большое число моделей в области оптимального размещения заданий в системах с параллельным обслуживанием. Основные успехи в создании алгоритмов диспетчеризации на данном этапе связаны с реализацией эвристических идей. Эти достижения вполне ощутимы, и в качестве примера действующих реализаций механизмов для размещения заданий можно привести такие системы, как Cisco's Local Director, IBM's Network Dispatcher, F5's Big IP [38–42]. Однако общий подход, который связал бы отдельные многочисленные постановки задач в единую теорию, пока отсутствует. Среди аналитических методов, использованных в рассмотренных моделях, выделяются методы теории массового обслуживания и теории марковского процесса принятия решений. Применяются также методы оптимизации (линейное, нелинейное и целочисленное программирование), униформизация (переход от дискретного времени к непрерывному). В некоторых постановках, не отмеченных явным образом в статье, оказываются полезными методы теории чисел (цепные дроби, последовательности Фарея и др.). Можно, однако, наблюдать, как перечень применимых аналитических методов анализа стремительно сокращается по мере усложнения рассматриваемой системы и остается единственный

подход — метод статистических испытаний. Со статистическим моделированием связаны исследования по созданию самонастраивающихся алгоритмов, корректирующих диспетчеризацию по наблюдениям за траекторией поведения системы. Это направление представляется перспективным и требует отдельного дополнительного освещения.

Литература

1. Mukhopadhyay A., Mazumdar R. R. Analysis of load balancing in large heterogeneous processor sharing systems. <http://arxiv.org/abs/1311.5806>.
2. Hyttiä E. Optimal routing of fixed size jobs to two parallel servers // *INFOR: Inform. Syst. Oper. Res.*, 2013. Vol. 51. No. 4. P. 215–224.
3. Mitzenmacher M. The power of two choices in randomized load balancing. — Berkeley, 1996. Ph.D. Thesis.
4. Vvedenskaya N. D., Dobrushin R. L., Karpelevich F. I. Queueing system with selection of the shortest of two queues: An asymptotic approach // *Probl. Inf. Transm.*, 1996. Vol. 32. No. 1. P. 20–34.
5. Martin J. B., Suhov Y. M. Fast jackson networks // *Ann. Appl. Probab.*, 1999. Vol. 9. No. 4. P. 840–854.
6. Graham C. Chaoticity on path space for a queueing network with selection of the shortest queue among several // *J. Appl. Probab.*, 2000. Vol. 37. P. 198–211.
7. Mitzenmacher M., Richa A., Sitaraman R. K. The power of two random choices: A survey of techniques and results // *Handbook of randomized computing* / Eds. S. Rajasekaran, P. M. Pardalos, J. H. Reif, J. Rolim. — Norwell, MA, USA: Kluwer Academic Publs., 2001. Vol. 1. P. 255–312.
8. Luczak M., McDiarmid C. On the power of two choices: Balls and bins in continuous time // *Ann. Appl. Probab.*, 2005. Vol. 15. No. 3. P. 1733–1764.
9. Luczak M., McDiarmid C. On the maximum queue length in the supermarket model // *Ann. Probab.*, 2006. Vol. 34. No. 2. P. 493–527.
10. Bramson M., Lu Y., Prabhakar B. Randomized load balancing with general service time distributions // *ACM Special Interest Group on Computer Systems Performance, SIGMETRICS Proceedings*, 2010. Vol. 38. Iss. 1. P. 275–286.
11. Winston W. Optimality of the shortest line discipline // *J. Appl. Probab.*, 1977. Vol. 14. P. 181–189.
12. Weber R. On the optimal assignment of customers to parallel servers // *J. Appl. Probab.*, 1978. Vol. 15. P. 406–413.
13. Haight A. Two queues in parallel // *Biometrika*, 1958. Vol. 45. P. 401–410.
14. Kingman J. F. C. Two similar queues in parallel // *Biometrika*, 1961. Vol. 48. P. 1316–1323.
15. Gupta V., Harchol-Balter M., Sigman K., Whitt W. Analysis of join-the-shortest-queue routing for Web Server Farms // *Perform. Evaluation*, 2007. Vol. 64. P. 1062–1081.

16. *Akgun O., Richter R., Wolff R.* Multiple server system with flexible arrivals // *Adv. Appl. Probab.*, 2011. Vol. 43. P. 985–1004.
17. *Harchol-Balter M., Crovella M., Murta C.* On choosing a task assignment policy for a distributed server system // *J. Parallel Distr. Comp.*, 1999. Vol. 59. Iss. 2. P. 204–228.
18. *Paxson V., Floyd S.* Wide area traffic: The failure of Poisson modeling // *IEEE/ACM Trans. Netw.*, 1995. Vol. 3. No. 3. P. 226–244.
19. *Peterson D., Adams D.* Fractal patterns in DASD I/O traffic // *22nd Computer Measurement Group Conference (International) Proceedings.* — San Diego, CA, USA, 1996. P. 560–571.
20. *Crovella M., Taqqu M., Bestavros A.* Heavy-tailed probability distributions in the World Wide Web // *A practical guide to heavy tails* / Eds. R.J. Adler, R.E. Feldman, and M.S. Taqqu. — Cambridge, MA, USA: Birkhauser Boston Inc., 1998. P. 3–25.
21. *Feitelson D.* Workload modeling for computer systems performance evaluation. — Cambridge, MA, USA: Cambridge University Press, 2015. 597 p.
22. *Whitt W.* Deciding which queue to join: Some counterexamples // *Oper. Res.*, 1986. Vol. 34. No. 1. P. 55–62.
23. *Crovella M., Harchol-Balter M., Murta C.* Task assignment in a distributed system: Improving performance by unbalancing load. — Boston: Boston University, 1997. Boston University Computer Science Department Technical Reports. BUCS-TR-1997-018.
24. *Crovella M., Harchol-Balter M., Murta C.* Task assignment in a distributed system: Improving performance by unbalancing load // *ACM Sigmetrics '98 Conference on Measurement and Modeling of Computer Systems Poster Session Proceedings.* — Madison, WI, USA, 1998. P. 268–269.
25. *Harchol-Balter M., Crovella M., Murta C.* On choosing a task assignment policy for a distributed server system // *J. Parallel Distr. Comp.*, 1999. Vol. 59. P. 204–228.
26. *Harchol-Balter M.* Task assignment with unknown duration // *J. ACM*, 2002. Vol. 49. P. 260–288.
27. *Broberg J., Tari Z., Zeephongsekul P.* Task assignment based on prioritising traffic flows // *Principles of distributed systems* / Ed. T. Higashino. — Lecture notes in computer science ser. — Grenoble, France: Springer, 2004. Vol. 3544. P. 415–430.
28. *Broberg J., Tari Z., Zeephongsekul P.* Task assignment with work-conserving migration // *J. Parallel Computing*, 2006. Vol. 32. P. 808–830.
29. *Jayasinghe M., Tari Z., Zeephongsekul P.* A scalable multi-tier task assignment policy with minimum excess load // *IEEE Symposium on Computers and Communications Proceedings.* — Riccione, Italy: IEEE, 2010. P. 913–918.
30. *Doroudi S., Hyytiä E., Harchol-Balter M.* Value driven load balancing // *Perform. Evaluation*, 2014. Vol. 79. P. 306–327.
31. *Bodas T., Ganesh A., Manjunath D.* Tolls and welfare optimization for multiclass traffic in multiqueue systems. <http://arxiv.org/abs/1409.7195>.
32. *Becker K., Gaver D., Glazebrook K., Jacobs P., Lawphongpanich S.* Allocation of tasks to specialized processors: A planning approach // *Eur. J. Oper. Res.*, 2000. Vol. 126. P. 80–88.
33. *Hyytiä E.* Optimal routing of fixed size jobs to two parallel servers // *INFOR: Inform. Syst. Oper. Res.*, 2013. Vol. 51. No. 4. P. 215–224.
34. *Harchol-Balter M., Crovella M., Murta C.* On choosing a task assignment policy for a distributed server system // *J. Parallel Distr. Comp.*, 1999. Vol. 59. P. 204–228.
35. *Hyytiä E., Penttinen A., Aalto S., Virtamo J.* Dispatching problem with fixed size jobs and processor sharing discipline // *23rd Teletraffic Congress (International) (ITC'23).* — San Francisco, CA, USA, 2011. P. 190–197.
36. *Konovalov M., Razumchik R.* Iterative algorithm for threshold calculation in the problem of routing fixed size jobs to two parallel servers // *J. Telecommunications Inform. Technol.*, 2015. Vol. 3. P. 32–38.
37. *Feng H., Misra V., Rubenstein D.* Optimal state-free, size-aware dispatching for heterogeneous $M/G/1$ -type systems // *Performance Evaluation*, 2005. Vol. 62. P. 475–492.
38. Cisco LocalDirector 400 series. <http://www.cisco.com/c/en/us/products/routers/localdirector-400-series>.
39. *Pistoia M., Letilley C.* IBM web sphere performance pack: Load balancing with IBM secure way network dispatcher // *IBM Redbooks*, 1999.
40. F5 Products. Big-IP. <http://www.f5.com/products/big-ip>.
41. *Schurman E., Brutlag J.* The user and business impact on server delays, additional bytes and http chunking in web search // *O'Reilly Velocity Web Performance and Operations Conference*, 2009. <http://velocityconf.com/velocity2009/public/schedule/detail/8523>.
42. Microsoft sharepoint 2010 load balancer. <http://loadbalancer.org/applications/microsoft-apps/microsoft-sharepoint>.

Поступила в редакцию 19.10.15

METHODS AND ALGORITHMS FOR JOB SCHEDULING IN SYSTEMS WITH PARALLEL SERVICE: A SURVEY

M. G. Konovalov¹ and R. V. Razumchik^{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

²Peoples’ Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

Abstract: The review of research papers devoted to the analysis of the dispatching problem in queueing systems is presented. The analysis is restricted to the class of systems with independent, operating in parallel, fully reliable servers, stochastic incoming flows of customers without any preceding constraints. The general goal of the analysis carried out in most of the papers is the solution of an optimization problem, which specification heavily depends on additional assumptions made. The models considered in the review are classified into several classes depending on the amount of *a priori* information and observability at decision times and performance criteria. The description of the dispatching algorithms most commonly found in literature and their properties is given. The main methods used for the analysis of the systems under these dispatching algorithms are reviewed. This review is intended to draw attention of the research community to one of the important problems in the field of information processing.

Keywords: dispatching; scheduling; parallel service; queueing system; optimization

DOI: 10.14357/19922264150406

Acknowledgments

The research was partly supported by the Russian Foundation for Basic Research (projects 15-07-03406 and 13-07-00223).

References

1. Mukhopadhyay, A., and R. R. Mazumdar. Analysis of load balancing in large heterogeneous processor sharing systems. Available at: <http://arxiv.org/abs/1311.5806> (accessed October 17, 2015).
2. Hyytiä, E. 2013. Optimal routing of fixed size jobs to two parallel servers. *INFOR: Inform. Syst. Oper. Res.* 51(4):215–224.
3. Mitzenmacher, M. 1996. The power of two choices in randomized load balancing. Berkeley. Ph.D. Thesis.
4. Vvedenskaya, N. D., R. L. Dobrushin, and F. I. Karpelevich. 1996. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Probl. Inf. Transm.* 32(1):20–34.
5. Martin, J. B., and Y. M. Suhov. 1999. Fast jackson networks. *Ann. Appl. Probab.* 9(4):840–854.
6. Graham, C. 2000. Chaoticity on path space for a queueing network with selection of the shortest queue among several. *J. Appl. Probab.* 37:198–211.
7. Mitzenmacher, M., A. Richa, and R. Sitaraman. 2001. The power of two random choices: A survey of techniques and results. *Handbook of randomized computing*. Eds. S. Rajasekaran, P. M. Pardalos, J. H. Reif, and J. Rolim. Norwell, MA: Kluwer Academic Publs. 1:255–312.
8. Luczak, M., and C. McDiarmid. 2005. On the power of two choices: Balls and bins in continuous time. *Ann. Appl. Probab.* 15(3):1733–1764.
9. Luczak, M., and C. McDiarmid. 2006. On the maximum queue length in the supermarket model. *Ann. Probab.* 34(2):493–527.
10. Bramson, M., Y. Lu, and B. Prabhakar. 2010. Randomized load balancing with general service time distributions. *ACM Special Interest Group on Computer Systems Performance, SIGMETRICS Proceedings*. 38(1):275–286.
11. Winston, W. 1977. Optimality of the shortest line discipline. *J. Appl. Probab.* 14:181–189.
12. Weber, R. 1978. On the optimal assignment of customers to parallel servers. *J. Appl. Probab.* 15:406–413.
13. Haight, A. 1958. Two queues in parallel. *Biometrika* 45:401–410.
14. Kingman, J. F. C. 1961. Two similar queues in parallel. *Biometrika* 48:1316–1323.
15. Gupta, V., M. Harchol-Balter, K. Sigman, and W. Whitt. 2007. Analysis of join-the-shortest-queue routing for Web Server Farms. *Performance Evaluation* 64:1062–1081.
16. Akgun, O., R. Righter, and R. Wolff. 2011. Multiple server system with flexible arrivals. *Adv. Appl. Probab.* 43:985–1004.
17. Harchol-Balter, M., M. Crovella, and C. Murta. 1999. On choosing a task assignment policy for a distributed server system. *J. Parallel Distr. Comp.* 59(2):204–228.
18. Paxson, V., and S. Floyd. 1995. Wide area traffic: The failure of Poisson modeling. *IEEE/ACM Trans. Netw.* 3(3):226–244.

19. Peterson, D., and D. Adams. 1996. Fractal patterns in DASD I/O traffic. *22nd Computer Measurement Group Conference (International) Proceedings*. San Diego, CA. 560–571.
20. Crovella, M., M. Taqqu, and A. Bestavros. 1998. Heavy-tailed probability distributions in the World Wide Web. *A practical guide to heavy tails*. Eds. R. J. Adler, R. E. Feldman, and M. S. Taqqu. Cambridge, MA: Birkhauser Boston Inc. 3–25.
21. Feitelson, D. 2015. *Workload modeling for computer systems performance evaluation*. Cambridge, MA: Cambridge University Press. 597 p.
22. Whitt, W. 1986. Deciding which queue to join: Some counterexamples. *Oper. Res.* 34(1):55–62.
23. Crovella, M., M. Harchol-Balter, and C. Murta. 1997. Task assignment in a distributed system: Improving performance by unbalancing load. Boston: Boston University. Boston University Computer Science Department Technical Reports. BUCS-TR-1997-018.
24. Crovella, M., M. Harchol-Balter, and C. Murta. 1998. Task assignment in a distributed system: Improving performance by unbalancing load. *ACM Sigmetrics'98 Conference on Measurement and Modeling of Computer Systems Poster Session Proceedings*. Madison, WI. 268–269.
25. Harchol-Balter, M., M. Crovella, and C. Murta. 1999. On choosing a task assignment policy for a distributed server system. *J. Parallel Distr. Comp.* 59:204–228.
26. Harchol-Balter, M. 2002. Task assignment with unknown duration. *J. ACM* 49:260–288.
27. Broberg, J., Z. Tari, and P. Zeephongsekul. 2004. Task assignment based on prioritising traffic flows. *Principles of distributed systems*. Ed. T. Higashino. Lecture notes in computer science ser. Grenoble, France: Springer. 3544:415–430.
28. Broberg, J., Z. Tari, and P. Zeephongsekul. 2006. Task assignment with work- conserving migration. *J. Parallel Computing* 32:808–830.
29. Jayasinghe, M., Z. Tari, and P. Zeephongsekul. 2010. A scalable multi-tier task assignment policy with minimum excess load. *IEEE Symposium on Computers and Communications Proceedings*. Riccione, Italy: IEEE. 913–918.
30. Doroudi, S., E. Hyttiä, and M. Harchol-Balter. 2014. Value driven load balancing. *Perform. Evaluation* 79:306–327.
31. Bodas, T., A. Ganesh, and D. Manjunath. 2014. Tolls and welfare optimization for multiclass traffic in multiqueue systems. Available at: <http://arxiv.org/abs/1409.7195> (accessed October 17, 2015).
32. Becker, K., D. Gaver, K. Glazebrook, P. Jacobs, and S. Lawphongpanich. 2000. Allocation of tasks to specialized processors: A planning approach. *Eur. J. Oper. Res.* 126:80–88.
33. Hyttiä, E. 2013. Optimal routing of fixed size jobs to two parallel servers. *INFOR: Inform. Syst. Oper. Res.* 51(4):215–224.
34. Harchol-Balter, M., M. Crovella, and C. Murta. 1999. On choosing a task assignment policy for a distributed server system. *J. Parallel Distr. Comp.* 59:204–228.
35. Hyttiä, E., A. Penttinen, S. Aalto., and J. Virtamo. 2011. Dispatching problem with fixed size jobs and processor sharing discipline. *23rd Teletraffic Congress (International) (ITC'23)*. San Fransisco, CA. 190–197.
36. Kononov, M., and R. Razumchik. 2015. Iterative algorithm for threshold calculation in the problem of routing fixed size jobs to two parallel servers. *J. Telecommunications Inform. Technol.* 3:32–38.
37. Feng, H., V. Misra, and D. Rubenstein. 2005. Optimal state-free, size-aware dispatching for heterogeneous M/G/-type systems. *Perform. Evaluation* 62:475–492.
38. Cisco LocalDirector 400 Series. Available at: <http://www.cisco.com/c/en/us/products/routers/localdirector-400-series> (accessed December 2, 2015).
39. Pistoia, M., and C. Letilley. 1999. IBM websphere performance pack: Load balancing with IBM secure way network dispatcher. *IBM Redbooks*.
40. F5 Products, Big-IP. Available at: <http://www.f5.com/products/big-ip> (accessed December 2, 2015).
41. Schurman, E., and J. Brutlag. 2009. The user and business impact on server delays, additional bytes and http chunking in web search. *O'Reilly Velocity Web Performance and Operations Conference*. Available at: <http://velocityconf.com/velocity2009/public/schedule/detail/8523> (accessed December 2, 2015).
42. Microsoft sharepoint 2010 load balancer. Available at: <http://loadbalancer.org/applications/microsoft-apps/microsoft-sharepoint> (accessed December 2, 2015).

Received October 19, 2015

Contributors

Kononov Mikhail G. (b. 1950) — Doctor of Science in technology, Head of Laboratory, 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; mkononov@ipiran.ru

Razumchik Rostislav V. (b. 1984) — 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; associate professor, Peoples’ Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; rrazumchik@ipiran.ru

ALGEBRAIC METHOD FOR APPROXIMATING JOINT STATIONARY DISTRIBUTION IN FINITE CAPACITY QUEUE WITH NEGATIVE CUSTOMERS AND TWO QUEUES*

R. Razumchik¹

Abstract: Consideration is given to the single-server queueing system (QS) with a Poisson flow of (ordinary) customers and Poisson flow of negative customers. There is a queue of capacity k ($0 < k < \infty$), where ordinary customers wait for service. If an ordinary customer finds the queue full upon an arrival, it is considered to be lost. Each negative customer upon arrival moves one ordinary customer from the queue, if it not empty, to another queue (bunker) of capacity r ($0 < r < \infty$) and after that it leaves the system. If upon arrival of a negative customer the queue is not empty and the bunker is full, the negative customer and one ordinary customer from the queue leave the system. In all other cases, an arrival of a negative customer has no effect on the system. Customers from bunker are served with relative priority (i. e., a customer from bunker enters server if only there are no customers in the queue to be served). Service times of customers from both the queue and the bunker are exponentially distributed with the same parameter. Purely algebraic method based on generating functions, Chebyshev and Gegenbauer polynomials for approximate calculation of joint stationary probability distribution is presented for the case $k = r$. Numerical examples, showing both pros and cons of the method are provided.

Keywords: queueing system; negative customers; Gegenbauer polynomials; stationary distribution; approximation

DOI: 10.14357/19922264150407

1 Introduction

Queueing systems and networks with negative customers have received significant attention from researches since they appeared in early 1990s [1–4]. This topic has already collected a big body of knowledge and it is neither possible nor intended to state here the comprehensive list of related papers. For review of the publications up to 2011, one can be referred to [5–7]. Detailed bibliography up to 2011 can be found in [8]. Among the latest research results, one can mention, for example, [7–19]. The novel approach to the analysis of networks with negative customers one can find in [20, 21].

The concept of negative customer turned out to be versatile. Quite often, it is assumed that the arrival of a negative customer leads to permanent removal of other (ordinary) customers residing in the system. It can also be the case that negative arrivals move ordinary customers between queues or nodes in the network. One of the problems' statements introduced in [10] implies that negative customers do not remove ordinary customers waiting for service in the buffer from the system, but delay their service by displacing them into another queue (bunker) wherefrom they are served according to a certain discipline (say, for example, relative priority). Latest results of the analysis of QS with such type of negative

customers are presented in [12, 13, 16, 19]. Particular interest in the analysis of such QS presents the case when queues' capacities (of buffer and bunker) are finite.

It is shown in [17] that apart from matrix analytic approach, Gaussian elimination method, and known numerical methods, it is possible to obtain joint stationary distribution utilizing properties of generating functions and special functions (Chebyshev and Gegenbauer polynomials). Note that such approach is not new as it was already used in [22] (and later in some other papers, for example, in [23]) but for the analysis of another types of QS. The method presented in [17] which allows one to do it is direct and suitable for exact arithmetics implementation though still suffers from computational problem. It is expressed in the need to efficiently perform basic operations with arbitrary big numbers.

In this paper, the modification of method from [17] is presented, which makes it, from the one hand, approximate and, from the other hand, allows one to carry out the computations of stationary performance characteristics for higher values of queues' capacities almost without losing accuracy. One interesting feature of the proposed method revealed during numerical experiments (see section 4) indicates that depending on initial

*This work was supported in part by the Russian Foundation for Basic Research (grants 15-07-03007 and 13-07-00223).

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; Peoples' Friendship University, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation, rrazumchik@ipiran.ru

values, one can obtain exact values for several performance characteristics using worst-case approximation (which substantially cuts computation time).

The rest of the paper is organized as follows. Section 2 is devoted to the detailed description of the system, equilibrium equations for joint stationary probabilities, and several useful relations that follow from rate-in-rate-out principle. In section 3, it is shown how to use generating functions and Chebyshev and Gegenbauer polynomials in order to obtain expressions for joint stationary probability distribution and to calculate stationary performance characteristics. Computational results are provided in section 4. In conclusion, the summary of the paper and future research plans are given.

2 Description of the System

Consideration is given to the QS with incoming Poisson flow of ordinary customers of intensity λ and Poisson flow of negative customers of intensity λ^- . For ordinary customers, there is a finite buffer of size k . If upon arrival of ordinary customer buffer is full, it is lost. A negative customer upon arrival moves one ordinary customer from buffer (if it is not empty; negative customers cannot remove customer which is in service) to another queue of finite capacity r (bunker). If upon arrival of a negative customer the queue in the buffer is not empty but bunker is full, the customer that is displaced from buffer is lost. Customers from both queues are served according to exponential distribution with the same parameter μ , FIFO (first in, first out) discipline, but the customers from the bunker are served with relative priority. This means that the customer from the bunker goes to the server only when server finished service of all customers from the buffer.

Denote by p_{ij} stationary probability of the fact that there are i customers in buffer, j customers in bunker, and server is busy. By p_0 , denote stationary probability of empty system. Stationary distribution satisfies the following system of equilibrium equations:

$$\lambda p_0 = \mu p_{00}; \quad (1)$$

$$(\lambda + \mu)p_{00} = \lambda p_0 + \mu p_{10} + \mu p_{01}, \quad i = 0, j = 0; \quad (2)$$

$$(\lambda + \mu + \lambda^-) p_{i0} = \lambda p_{i-1,0} + \mu p_{i+1,0}, \quad i = 1, \dots, k-1, j = 0; \quad (3)$$

$$(\mu + \lambda^-) p_{kj} = \lambda p_{k-1,j}, \quad i = k, j = 0, \dots, r; \quad (4)$$

$$(\lambda + \mu) p_{0j} = \mu p_{1j} + \lambda^- p_{1,j-1} + \mu p_{0,j+1}, \quad i = 0, j = 1, \dots, r-1; \quad (5)$$

$$(\lambda + \mu) p_{0r} = (\mu + \lambda^-) p_{1r} + \lambda^- p_{1,r-1}, \quad i = 0, j = r; \quad (6)$$

$$(\lambda + \mu + \lambda^-) p_{ij} = \lambda p_{i-1,j} + \mu p_{i+1,j} + \lambda^- p_{i+1,j-1}, \quad i = 1, \dots, k-1, j = 1, \dots, r-1; \quad (7)$$

$$(\lambda + \mu + \lambda^-) p_{ir} = \lambda p_{i-1,r} + \lambda^- p_{i+1,r-1} + (\mu + \lambda^-) p_{i+1,r}, \quad i = 1, \dots, k-1, j = r, \quad (8)$$

with the normalization condition

$$p_0 + \sum_{i=0}^k \sum_{j=0}^r p_{ij} = 1.$$

Although for the sake of clarity further the case $k = r$ and $r \geq 6$ will be considered, there appears nothing new (but only technical details) in the general case $0 < k < \infty, 0 < r < \infty$.

One of the key relations that allows the computation of stationary joint probability distribution is the one which shows how probability of empty system p_0 depends on boundary probabilities $\{p_{0j}, j = 1, \dots, r\}$, $\{p_{i0}, i = 1, \dots, r-1\}$, $\{p_{ir}, i = 1, \dots, r-1\}$, and $\{p_{rj}, j = 0, \dots, r\}$. It is shown in [17] using rate-in-rate-out principle that for probability p_0 , it holds

$$p_0 = 1 - \frac{(1 - q^{r+1})\rho}{(1 - \rho + \lambda^-/\mu)q^{r+1}} \sum_{j=0}^r p_{rj} \quad (9)$$

where here and henceforth $\rho = \lambda/\mu, q = \lambda/(\mu + \lambda^-)$. Thus, for the considered QS, the probability p_0 , besides initial parameters r, λ, λ^- , and μ , depends only on the boundary probabilities $\{p_{rj}, j = 0, \dots, r\}$. In [17], it is shown how it is possible to find such values X_{rj} that $p_{rj} = X_{rj}p_0$ and how then, using (9), one can compute the probability p_0 and, consequently, the whole joint probability distribution. In the next section, the author follows the argumentation from [17] and suggests modifications for several steps that lead to approximate solution.

3 Approximation of Joint Stationary Probability Distribution

Introduce double probability generating function (PGF)

$$P(u, v) = \sum_{i=0}^k \sum_{j=0}^r p_{ij} u^i v^j, \quad 0 \leq u \leq 1, \quad 0 \leq v \leq 1.$$

Multiplying (1)–(8) by $u^i v^j$ and summing over all values of i and j , having collected common terms, one obtains

$$\begin{aligned} & - [\lambda u^2 - (\lambda + \mu + \lambda^-) u + \mu + \lambda^-] P(u, v) \\ & = \frac{\mu u(v-1)}{v} p_{00} + \lambda u^{r+1} (1-u) \sum_{j=0}^r p_{rj} v^j \end{aligned}$$

$$\begin{aligned}
 & + \frac{(\mu + \lambda^- v)(u - v)}{v} \sum_{j=0}^r p_{0j} v^j \\
 & + \lambda^- v^r (1 - v) \sum_{i=1}^r p_{ir} u^i. \quad (10)
 \end{aligned}$$

The expression in the square brackets in the left part of the Eq. (10) is a polynomial of a second degree in u . Its roots have the form:

$$\begin{aligned}
 u_{1,2} & = u_{1,2}(v) \\
 & = \frac{\lambda + \mu + \lambda^- \mp \sqrt{(\lambda + \mu + \lambda^-)^2 - 4\lambda(\mu + \lambda^- v)}}{2\lambda}.
 \end{aligned}$$

It can be seen that $u_2(v) > 1$ and $0 < u_1(v) \leq 1$ for $0 \leq v \leq 1$. Generating function $P(u, v)$ is the ratio of two polynomial functions. For each value of v , PGF $P(u, v)$ is the continuous function of u on the whole set \mathbf{R} of real numbers. Then, as left part in (10) vanishes at points $(u_1(v), v)$ and $(u_2(v), v)$, the right part must vanish at these points too. Therefore, one obtains two equations:

$$\begin{aligned}
 & \frac{\mu u_1(v - 1)}{v} p_{00} + \lambda u_1^{r+1} (1 - u_1) \sum_{j=0}^r p_{rj} v^j \\
 & + \frac{(\mu + \lambda^- v)(u_1 - v)}{v} \sum_{j=0}^r p_{0j} v^j \\
 & + \lambda^- v^r (1 - v) \sum_{i=1}^r p_{ir} u_1^i = 0; \quad (11)
 \end{aligned}$$

$$\begin{aligned}
 & \frac{\mu u_2(v - 1)}{v} p_{00} + \lambda u_2^{r+1} (1 - u_2) \sum_{j=0}^r p_{rj} v^j \\
 & + \frac{(\mu + \lambda^- v)(u_2 - v)}{v} \sum_{j=0}^r p_{0j} v^j \\
 & + \lambda^- v^r (1 - v) \sum_{i=1}^r p_{ir} u_2^i = 0. \quad (12)
 \end{aligned}$$

If one now expresses the term with p_{00} from (11) and put it into (12), after collecting common terms, one gets the following equation:

$$\begin{aligned}
 & \left(\frac{u_2^r - u_1^r}{u_2 - u_1} - \frac{u_2^{r+1} - u_1^{r+1}}{u_2 - u_1} \right) \sum_{j=0}^r p_{rj} v^j + \\
 & + \frac{\lambda^- v^r (1 - v)}{\lambda} \sum_{i=1}^r p_{ir} \frac{u_2^{i-1} - u_1^{i-1}}{u_2 - u_1} \\
 & + \sum_{j=0}^r p_{0j} v^j = 0. \quad (13)
 \end{aligned}$$

But if one expresses term with $\sum_{j=0}^r p_{0j} v^j$ from (11) and puts it into (12), one obtains:

$$\begin{aligned}
 & \mu(1 - v)p_{00} + \mu \left(\frac{u_2^r - u_1^r}{u_2 - u_1} - \frac{u_2^{r+1} - u_1^{r+1}}{u_2 - u_1} \right) \sum_{j=0}^r p_{rj} v^j \\
 & + \left(\lambda^- \frac{u_2^r - u_1^r}{u_2 - u_1} - (\lambda + \lambda^-) \frac{u_2^{r+1} - u_1^{r+1}}{u_2 - u_1} \right. \\
 & \quad \left. + \lambda \frac{u_2^{r+2} - u_1^{r+2}}{u_2 - u_1} \right) \sum_{j=0}^r p_{rj} v^{j+1} \\
 & + \frac{\mu \lambda^- v^r (1 - v)}{\lambda} \sum_{i=1}^r p_{ir} \frac{u_2^{i-1} - u_1^{i-1}}{u_2 - u_1} \\
 & - \lambda^- v^{r+1} (1 - v) \sum_{i=1}^r p_{ir} \frac{u_2^i - u_1^i}{u_2 - u_1} + \\
 & + \frac{(\lambda^-)^2 v^{r+1} (1 - v)}{\lambda} \sum_{i=1}^r p_{ir} \frac{u_2^{i-1} - u_1^{i-1}}{u_2 - u_1} = 0. \quad (14)
 \end{aligned}$$

Though the expressions for the roots $u_1(v)$ and $u_2(v)$ are known, use the Lagrange interpolating polynomials $(L_n^1(v)$ and $L_n^2(v)$, respectively), which coincide with $u_1(v)$ and $u_2(v)$ at $n + 1$ different points. It is known (see, e. g., [9]) that the approximation error $u_1(v) \approx L_n^1(v)$ (and $u_2(v) \approx L_n^2(v)$) is minimum if the interpolation nodes \bar{v}_k are the zeros of the Chebyshev polynomials $T_{n+1}(v)$ of the first kind of degree $n + 1$, i. e.,

$$\bar{v}_k = \cos \left(\frac{(2k - 1)\pi}{2n + 2} \right), \quad k = 1, 2, \dots, n + 1.$$

Given that Chebyshev polynomials of the first kind are defined on $[-1, 1]$ and functions $u_{1,2}(v)$ are defined for $v \in [0, 1]$, then the values of nodes must be recalculated using linear transformation, i. e., the new nodes have the values

$$v_k = \frac{\bar{v}_k + 1}{2}, \quad k = 1, 2, \dots, n + 1.$$

Now, the expressions $(u_2^i - u_1^i)/(u_2 - u_1)$, $i \geq 1$, that enter Eqs. (13) and (14), can be rewritten in the form:

$$\begin{aligned}
 \frac{u_2^i - u_1^i}{u_2 - u_1} & = \frac{u_2(v)^i - u_1(v)^i}{u_2(v) - u_1(v)} = \frac{L_n^2(v)^i - L_n^1(v)^i}{L_n^2(v) - L_n^1(v)} \\
 & = f_{in}(v), \quad i \geq 1, v \in [0, 1].
 \end{aligned}$$

It is known (see, e. g., [24]) that the function $f_{in}(v)$ that interpolates the fraction $(u_2^i - u_1^i)/(u_2 - u_1)$ at $n + 1$ zeros of Chebyshev polynomials of the first kind can be written as their combination, that is,

$$f_{in}(v) \approx \sum_{k=0}^n w_{ik} T_k(2v - 1) \quad (15)$$

where the coefficients w_k

$$w_{i1} = \frac{\sum_{j=1}^{n+1} f_{in}(v_j)T_1(\bar{v}_j)}{n+1};$$

$$w_{ik} = \frac{2 \sum_{j=1}^{n+1} f_{in}(v_j)T_k(\bar{v}_j)}{n+1}, \quad k = 1, \dots, n.$$

In [17], there was found the closed-form expression for $(u_2^i - u_1^i)/(u_2 - u_1)$ which is a polynomial of integer degree $\lfloor i/2 \rfloor$ in v with real coefficients. Here is the main point of modification. In (15), the **same** number of interpolation nodes n for every $i > 1$ in such way that $n < \lfloor r/2 \rfloor$. This reduces the degrees of the polynomials that need to be computed and, consequently, leads to the simplification of calculations (of course, at the expense of accuracy loss).

Using the exact relations between the Chebyshev polynomials $T_k(v)$ and $U_k(v)$ of the first and the second kind, respectively, $T_k(v) = U_k(v) - vU_{k-1}(v)$ and noticing that $U_{-1}(v) = 0$, one can rewrite (15) in the form:

$$f_{in}(v) \approx \sum_{k=0}^n w_{ik}U_k(2v-1) - (2v-1) \sum_{k=0}^{n-1} w_{i,k+1}U_k(2v-1). \quad (16)$$

Denote by $U_n^{(m)}(v)$ the m th derivative of $U_n(v)$ at point v . Derivatives $U_n^{(m)}(v)$ are related with Gegenbauer polynomials $C_n^m(v)$ by equality (see, e.g., [25, p. 186])

$$U_n^{(m)}(v) = 2^m m! C_{n-m}^{m+1}(v). \quad (17)$$

The values of $C_n^m(v)$ at point $v = 0$ are given by (see, e.g., [25, p. 175]):

$$C_n^m(0) = \begin{cases} 0, & n \neq 2l; \\ \frac{(-1)^{n/2} \Gamma(m+n/2)}{(n/2)! \Gamma(m)}, & n = 2l. \end{cases}$$

Here, $\Gamma(n)$ denotes gamma function. Expanding in (16) functions $U_k(2v-1)$ in Maclaurin series and using (17), one obtains

$$f_{in}(v) \approx \sum_{k=0}^n w_{ik} \sum_{m=0}^k 2^m (2v-1)^m C_{k-m}^{m+1}(0) - \sum_{k=0}^{n-1} w_{i,k+1} \sum_{m=0}^k 2^m (2v-1)^{m+1} C_{k-m}^{m+1}(0).$$

The latter expression is the polynomial of integer degree n in v , that is,

$$f_{in}(v) \approx \sum_{j=0}^n h_{ij} v^j \quad (18)$$

where the coefficients h_j are computed using the following relations:

$$h_{i0} = \sum_{m=0}^{n-1} (-2)^m \sum_{k=m}^{n-1} [w_{ik} C_{k-m}^{m+1}(0) + w_{i,k+1} C_{k-m}^{m+1}(0)] + \sum_{m=0}^n (-2)^m w_{i,n} C_{n-m}^{m+1}(0);$$

$$h_{ij} = 2^j \left(\sum_{m=j}^n 2^m \binom{m}{j} (-1)^{m-j} \sum_{k=m}^n w_{ik} C_{k-m}^{m+1}(0) - 2^{j-1} \sum_{k=j-1}^{n-1} w_{i,k+1} C_{k-j+1}^j(0) - \sum_{m=j}^{n-1} 2^m \binom{m+1}{j} (-1)^{m+1-j} \sum_{k=m}^{n-1} w_{i,k+1} C_{k-m}^{m+1}(0) \right);$$

$$h_{in} = w_{in} 2^{2n-1} (2C_0^{m+1}(0) - C_0^m(0)).$$

Here, $\binom{n}{k}$ is the number of the k th combination from n .

Now, let go back to the simplification of equalities (13) and (14). The coefficient in front of the sum in the first term of (13) can be represented, with respect to (18), as

$$\frac{u_2^r - u_1^r}{u_2 - u_1} - \frac{u_2^{r+1} - u_1^{r+1}}{u_2 - u_1} = \sum_{j=0}^n h_{rj} v^j - \sum_{j=0}^n h_{r+1,j} v^j = \sum_{j=0}^n A_j v^j \quad (19)$$

where $A_j = h_{rj} - h_{r+1,j}$. The second term in (13) can be rewritten as

$$\sum_{i=1}^r p_{ir} \frac{u_2^{i-1} - u_1^{i-1}}{u_2 - u_1} = p_{2r} + \sum_{j=0}^n B_j v^j \quad (20)$$

where $B_j = \sum_{i=3}^r p_{ir} h_{i-1,j}$. Now, if one puts (19) and (20) into (13) and collects the common terms, one obtains the following equation:

$$\sum_{m=0}^n v^m \left[\sum_{i=0}^m p_{r,m-i} A_i + p_{0m} \right] + \sum_{m=n+1}^r v^m \left[\sum_{i=0}^n p_{r,m-i} A_i + p_{0m} \right] + \sum_{m=r+1}^{r+n} v^m \left[\sum_{i=m-r}^n p_{r,m-i} A_i + \frac{\lambda^-}{\lambda} B_{m-r} \right] + \frac{\lambda^-}{\lambda} v^r [p_{2r} + B_0] - \frac{\lambda^-}{\lambda} v^{r+1} [p_{2r} + B_0] - \frac{\lambda^-}{\lambda} \sum_{m=r+2}^{r+n+1} B_{m-r-1} v^m = 0.$$

This is polynomial in a single variable v of integer degree $r + n + 1$. From the fact that it equals zero $\forall v \in [0, 1]$ and, therefore, its coefficients are all equal to zero, one obtains the following system of algebraic equations with constant coefficients:

$$\sum_{i=0}^m p_{r,m-i} A_i + p_{0m} = 0, \quad 0 \leq m \leq n; \quad (21)$$

$$\sum_{i=0}^n p_{r,m-i} A_i + p_{0m} = 0, \quad n + 1 \leq m \leq r - 1; \quad (22)$$

$$\sum_{i=0}^n p_{r,r-i} A_i + p_{0r} + \frac{\lambda^-}{\lambda} [p_{2r} + B_0] = 0, \quad m = r; \quad (23)$$

$$\sum_{i=1}^n p_{r,r+1-i} A_i + \frac{\lambda^-}{\lambda} B_1 - \frac{\lambda^-}{\lambda} [p_{2r} + B_0] = 0, \quad m = r + 1; \quad (24)$$

$$\sum_{i=m-r}^n p_{r,m-i} A_i + \frac{\lambda^-}{\lambda} B_{m-r} - \frac{\lambda^-}{\lambda} B_{m-r-1} = 0, \quad r + 2 \leq m \leq r + n; \quad (25)$$

$$-\frac{\lambda^-}{\lambda} B_n = 0, \quad m = r + n + 1. \quad (26)$$

If one sums up all equations starting from $m = r + 1$, one obtains

$$\frac{\lambda^-}{\lambda} [p_{2r} + B_0] = \sum_{m=r+1}^{r+n} \sum_{i=m-r}^n p_{r,m-i} A_i. \quad (27)$$

Now, simplify relation (14) by analogy with (13). It is straightforward to show that the coefficient in front of the sum in the third term, with respect to (18), equals

$$\lambda^- \frac{u_2^r - u_1^r}{u_2 - u_1} - (\lambda + \lambda^-) \frac{u_2^{r+1} - u_1^{r+1}}{u_2 - u_1} + \lambda \frac{u_2^{r+2} - u_1^{r+2}}{u_2 - u_1} = \sum_{j=0}^n D_j v^j \quad (28)$$

where $D_j = \lambda^- h_{rj} - (\lambda + \lambda^-) h_{r+1,j} + \lambda h_{r+2,j}$. The last term in (14) simply equals

$$\sum_{i=1}^r p_{ir} \frac{u_2^i - u_1^i}{u_2 - u_1} = p_{1r} + \sum_{j=0}^n E_j v^j \quad (29)$$

where $E_j = \sum_{i=2}^r p_{ir} h_{ij}$. By putting the representations (19), (20), (28) and (29) in (14), having collected the common terms, one arrives at the following equation:

$$\mu(1 - v)p_{00} + \mu \left(\sum_{m=0}^n v^m \sum_{i=0}^m p_{r,m-i} A_i + \sum_{m=n+1}^r v^m \sum_{i=0}^n p_{r,m-i} A_i + \sum_{m=r+1}^{n+r} v^m \sum_{i=m-r}^n p_{r,m-i} A_i \right)$$

$$+ \left(\sum_{m=1}^{n+1} v^m \sum_{i=0}^{m-1} p_{r,m-i-1} D_i + \sum_{m=n+2}^{r+1} v^m \sum_{i=0}^n p_{r,m-i-1} D_i + \sum_{m=r+2}^{n+r+1} v^m \sum_{i=m-r-1}^n p_{r,m-i-1} D_i \right) + \frac{\mu \lambda^-}{\lambda} v^r [p_{2r} + B_0] + v^{r+1} \lambda^- \left([p_{2r} + B_0] \frac{(\lambda^- - \mu)}{\lambda} - [p_{1r} + E_0] \right) + v^{r+2} \lambda^- \left(p_{1r} + E_0 - \frac{\lambda^-}{\lambda} [p_{2r} + B_0] \right) + \frac{\mu \lambda^-}{\lambda} \sum_{j=1}^n B_j v^{r+j} - \frac{\mu \lambda^-}{\lambda} \sum_{j=1}^n B_j v^{r+j+1} + \frac{(\lambda^-)^2}{\lambda} \sum_{j=1}^n B_j v^{r+j+1} - \frac{(\lambda^-)^2}{\lambda} \sum_{j=1}^n B_j v^{r+j+2} - \lambda^- \sum_{j=1}^n E_j v^{r+j+1} + \lambda^- \sum_{j=1}^n E_j v^{r+j+2} = 0.$$

This is a polynomial in a single-variable v of integer degree $r + n + 2$. From the fact that it equals zero $\forall v \in [0, 1]$ and, therefore, its coefficients are all equal to zero, one can obtain the system of algebraic equations with constant coefficients. Only first (starting from the lowest degree of v) $r + 1$ equations which are

$$\mu p_{00} + \mu p_{r0} A_0 = 0, \quad m = 0; \quad (30)$$

$$-\mu p_{00} + \mu \sum_{i=0}^1 p_{r,1-i} A_i + p_{r0} D_0 = 0, \quad m = 1; \quad (31)$$

$$\mu \sum_{i=0}^m p_{r,m-i} A_i + \sum_{i=0}^{m-1} p_{r,m-i-1} D_i = 0, \quad 2 \leq m \leq n; \quad (32)$$

$$\mu \sum_{i=0}^n p_{r,n+1-i} A_i + \sum_{i=0}^n p_{r,n-i} D_i = 0, \quad m = n + 1; \quad (33)$$

$$\mu \sum_{i=0}^n p_{r,m-i} A_i + \sum_{i=0}^n p_{r,m-i-1} D_i = 0, \quad n + 2 \leq m \leq r - 1; \quad (34)$$

$$\mu \sum_{i=0}^n p_{r,r-i} A_i + \sum_{i=0}^n p_{r,r-i-1} D_i + \frac{\mu \lambda^-}{\lambda} [p_{2r} + B_0] = 0, \quad m = r, \quad (35)$$

are needed. The last term in Eqs. (35) depends on p_{ir} , $i = 2, \dots, r$. But it can be replaced according to (27) by the expression that contains only probabilities p_{rj} , $j = 0, \dots, r$, thus making the whole system (30)–(35) solvable (up to a constant). Let one now express from Eq. (30) probability p_{r0} , from the Eq. (31) – probability p_{r1} , from the Eq. (32) – probability p_{r2} , etc. Consequently, one obtains the following recursive formulas for probabilities p_{rj} , $j = 0, \dots, r$:

$$\left. \begin{aligned}
 p_{r0} &= -\frac{\lambda}{\mu A_0} p_0, \quad m = 0; \\
 p_{r1} &= -\frac{1}{\mu A_0} (\mu A_1 + D_0 + \mu A_0) p_{r0}, \\
 &\quad m = 1; \\
 p_{rm} &= -\frac{1}{\mu A_0} \sum_{i=0}^{m-1} p_{r,m-i-1} (\mu A_{i+1} + D_i), \\
 &\quad 2 \leq m \leq n; \\
 p_{r,n+1} &= -\frac{1}{\mu A_0} \left(\mu \sum_{i=0}^{n-1} p_{r,n-i} A_{i+1} \right. \\
 &\quad \left. + \sum_{i=0}^n p_{r,n-i} D_i \right), \quad m = n + 1; \\
 p_{rm} &= -\frac{1}{\mu A_0} \left(\mu \sum_{i=1}^n p_{r,m-i} A_i \right. \\
 &\quad \left. + \sum_{i=0}^n p_{r,m-i-1} D_i \right), \quad n + 2 \leq m \leq r - 1; \\
 p_{rr} &= -\frac{1}{\mu \sum_{m=0}^n A_m} \left(\sum_{i=0}^n p_{r,r-i-1} D_i \right. \\
 &\quad \left. + \mu \sum_{m=0}^{n-1} \sum_{i=1}^{n-m} p_{r,r-i} A_{m+i} \right), \quad m = r.
 \end{aligned} \right\} (36)$$

Notice that here, A_i and D_i are the constants defined in (19) and (28), and their values depend on the quality of approximation (18).

From Eqs. (21)–(23), taking into account (27), one immediately obtains recurrence relations for probabilities $\{p_{0j}, j = 0, \dots, r\}$:

$$\left. \begin{aligned}
 p_{0,m} &= -\sum_{i=0}^m p_{r,m-i} A_i, \quad 0 \leq m \leq n; \\
 p_{0,m} &= -\sum_{i=0}^n p_{r,m-i} A_i, \quad n + 1 \leq m \leq r - 1; \\
 p_{0,r} &= -\sum_{m=r}^{i=0} \sum_{i=m-r}^n p_{r,m-i} A_i.
 \end{aligned} \right\} (37)$$

Now, the algorithm for the computation of the stationary joint probability distribution p_{ij} is straightforward¹:

- compute values $X_{rj} = p_{rj}/p_0, j = 0, \dots, r$, using Eqs. (36);
- compute probability p_0 from (9) using representation $p_{rj} = p_0 X_{rj}$;
- compute probabilities $\{p_{rj}, j = 0, \dots, r\}$, using equation $p_{rj} = X_{rj} p_0$;
- compute probabilities $\{p_{0j}, j = 0, \dots, r\}$, using Eqs. (37);

- compute probabilities $\{p_{r-1,j}, j = 0, \dots, r\}$ using Eq. (4);
- compute probabilities $\{p_{i0}, j = r - 2, \dots, 1\}$ using Eq. (3);
- compute probabilities $\{p_{r-2,j}, j = r - 1, \dots, 1\}$ using Eq. (7); and
- for each $i = r - 2, \dots, 1$, compute probabilities $\{p_{i-1,j}, j = r - 1, \dots, 1\}$ using Eq. (7) and probabilities $p_{i,r}$ using Eq. (8).

For some important performance characteristics of the system, one needs to perform only the first three steps of the algorithm, For example, the probability that the arriving customer is lost due to the full buffer equals $\pi_1 = \sum_{j=0}^r p_{rj}$. The probability π_2 that the ordinary customer will be lost due to the full bunker equals $\pi_2 = \lambda^- \sum_{i=1}^r p_{ir} / (\lambda(1 - \pi_1))$. Other quantities such as the expected number and the variance of number of customers in the buffer and in the bunker can be computed using the corresponding formulas given in [17].

In the next section, the author presents some numerical examples which show how performance characteristics behave when one varies the quality of approximation (18).

4 Numerical Examples

Here, the results of numerical computations of the following stationary performance characteristics are presented: the probability that the arriving customer is lost due to the full buffer (π_1), probability customer displaced from buffer is lost due to the full bunker (π_2), mean waiting time in buffer (ω_{buff}) and bunker (ω_{bunk}).

Four different combinations of initial parameters are considered:

- (1) $r = 26, \lambda = 7, \lambda^- = 5, \mu = 10$ (Table 1);
- (2) $r = 26, \lambda = 12, \lambda^- = 5, \mu = 10$ (Table 2);
- (3) $r = 26, \lambda = 20, \lambda^- = 5, \mu = 10$ (Table 3); and
- (4) $r = 26, \lambda = 20, \lambda^- = 12, \mu = 10$ (Table 4).

For each of these cases, numerical computations were carried out using different number of interpolation nodes n , specifically, $3 \leq n < \lfloor r/2 \rfloor$. The results are presented in Tables 1–4. Last line in each table states exact values of performance characteristics, obtained using algorithm from [17].

From the results in the tables, one can see that it is possible to obtain (though not always) the accurate values for the considered performance characteristics if the number of interpolation points is less than the degree of polynomials used in direct method (i. e., less than $\lfloor r/2 \rfloor$). Numerical experiments show that compu-

¹In fact, it is the same as in [17]; so, here, it is just repeated for the sake of completeness.

Table 1 Performance characteristics for $r = 26$, $\lambda = 7$, $\lambda^- = 5$, and $\mu = 10$

n	π_1	π_2	ω_{buff}	ω_{bunk}
3	0.000000	-0.426874	0.124851	-40.620061
4	0.000000	-0.427355	0.124894	-0.053369
5	0.000000	-0.743357	0.152544	3.467104
6	0.000000	-0.084505	0.094894	1.839610
7	0.000000	0.007251	0.086866	0.364169
8	0.000000	-0.000211	0.087518	0.637415
9	0.000000	0.000004	0.087500	0.624550
	0.000000	0.000007	0.087499	0.624747

Table 2 Performance characteristics for $r = 26$, $\lambda = 12$, $\lambda^- = 5$, and $\mu = 10$

n	π_1	π_2	ω_{buff}	ω_{bunk}
3	0.000606	0.166082	0.328110	-14.288452
4	0.000606	0.166201	0.328063	-14.598858
5	0.000605	0.166759	0.327844	-15.622581
6	0.000606	0.166190	0.328068	3.012986
7	0.000606	0.166627	0.327896	12.247301
8	0.000606	0.166654	0.327885	11.335891
9	0.000606	0.166656	0.327884	11.366810
	0.000606	0.166656	0.327884	11.366303

Table 3 Performance characteristics for $r = 26$, $\lambda = 20$, $\lambda^- = 5$, and $\mu = 10$

n	π_1	π_2	ω_{buff}	ω_{bunk}
3	0.250106	0.333239	1.534312	-14.854582
4	0.250106	0.333240	1.534311	-12.811167
5	0.250106	0.333239	1.534312	-19.689504
6	0.250106	0.333239	1.534312	36.261106
7	0.250106	0.333239	1.534312	-892.561062
8	0.250106	0.333239	1.534312	10237.695931
9	0.250106	0.333239	1.534312	18727.217279
	0.250106	0.333239	1.534312	18414.922135

Table 4 Performance characteristics for $r = 26$, $\lambda = 20$, $\lambda^- = 12$, and $\mu = 10$

n	π_1	π_2	ω_{buff}	ω_{bunk}
3	0.008258	0.495837	0.391757	-3.781687
4	0.008258	0.495837	0.391757	-3.828509
5	0.008258	0.495837	0.391756	-3.784435
6	0.008258	0.495837	0.391757	-3.584416
7	0.008258	0.495837	0.391757	-5.855355
8	0.008258	0.495837	0.391757	12.740076
9	0.008258	0.495837	0.391757	26.932748
	0.008258	0.495826	0.391757	23.646577

tational time decreases substantially with slow decrease of n . Nevertheless, it remains an open questing how to choose appropriate number of nodes n so as not to lose much in accuracy but noticeably gain in computation time and whether the value of n is the same for all performance characteristics independently of combination of initial parameters.

5 Concluding Remarks

In this paper, the algebraic method for approximate calculation of joint stationary probability distribution in single server QS with Poisson flows of ordinary and negative customers, finite buffer, and finite bunker for customers that were displaced from the buffer is presented. Obtained relations were verified by comparing with the results provided by direct methods. The method heavily relies on the opportunity to solve the equilibrium equations recursively. This opportunity is established by considering additional equations obtained from the properties of the generation function of the joint distribution. Though the results are presented for the special case when the queues' capacities are equal, it is straightforward to generalize the method for different (finite) capacities and heterogeneous servers. Future research efforts are worth concentrating in two directions. The first one is the in-depth analysis of the approximation and the study of cases in which it works either well or bad. The second one is the application of the proposed method for the analysis of more complex QS which behavior can be represented as multidimensional Markov processes with finite-state space.

References

- Gelenbe, E., P. Glynn, and K. Sigman. 1991. Queues with negative arrivals. *J. Appl. Probab.* 28:245–250.
- Gelenbe, E. 1993. G-networks with instantaneous customer movement. *J. Appl. Probab.* 30(3):742–748.
- Harrison, P. G., and E. Pitel. 1993. Sojourn times in single-server queues with negative customers. *J. Appl. Probab.* 30:943–963.
- Fourneau, J., E. Gelenbe, and R. Suros. 1996. G-networks with multiple classes of negative and positive customers. *Theor. Comput. Sci.* 155(1):141–156.
- Artalejo, J. R. 2000. G-networks: A versatile approach for work removal in queueing systems. *Eur. J. Oper. Res.* 126:233–249.
- Bocharov, P. P., and V. M. Vishnevskii. 2003. G-networks: Development of the theory of multiplicative networks. *Automat. Rem. Contr.* 64(5):714–739.
- Boucherie, R. J., and N. van Dijk, eds. 2011. *Queueing networks: A fundamental approach*. International ser. in operations research and management science. New York, NY: Springer. 154. 800 p.
- Van Do, T. 2011. An initiative for a classified bibliography on G-networks. *Perform. Evaluation* 68(4):385–394.
- Suetin, P. K. 1979. *Klassicheskie ortogonal'nye mnogochleny* [Classical orthogonal polynomials]. Moscow: Nauka. 416 p.
- Manzo, R., N. Cascone, and R. V. Razumchik. 2008. Exponential queueing system with negative customers and bunker for ousted customers. *Automat. Rem. Contr.* 69:1542–1551.

11. Klimenok, V., and A. Dudin. 2012. A BMAP/PH/ N queue with negative customers and partial protection of service. *Commun. Stat. Simulat.* 41(7):1062–1082.
12. Pechinkin, A. V., and R. V. Razumchik. 2012. Stationary waiting time distribution in queueing system with negative customers and bunker for ousted customers under Last-LIFO-LIFO service discipline. *J. Commun. Technol. El.* 57(12):1331–1339.
13. Pechinkin, A. V., and R. V. Razumchik. 2012. A method for calculating stationary queue distribution in a queueing system with flows of ordinary and negative claims and a bunker for superseded claims. *J. Commun. Technol. El.* 57(8):882–891.
14. Dao-Thi, T., J. Fourneau, and M. Tran. 2013. Networks of order independent queues with signals. *21st Symposium (International) on Modeling, Analysis and Simulation of Computer and Telecommunication Systems Proceedings.* 131–140.
15. Krishna Kumar, B., S. Pavai Madheswari, and S. R. Anantha Lakshmi. 2013. An $M/G/1$ Bernoulli feedback retrial queueing system with negative customers. *Oper. Res.* 13(2):187–210.
16. Razumchik, R. V. 2013. Stacionarnoe raspredelenie vremeni ozhidaniya v sisteme obsluzhivaniya s otritsatel'nymi zayavkami, bunkerom dlya vytesnennykh zayavok, razlichnymi intensivnostyami obsluzhivaniya pri discipline First-FIFO-FIFO [Stationary waiting time distribution in queueing system with negative customers and bunker for ousted customers under First-FIFO-FIFO service discipline]. *Informatika i ee Primeneniya — Inform. Appl.* 7(2):34–39.
17. Razumchik, R. V. 2014. Analysis of finite capacity queue with negative customers and bunker for ousted customers using chebyshev and gegenbauer polynomials. *Asia Pac. J. Oper. Res.* 31(4):1450029. 21 p.
18. Van Do, T., D. Papp, R. Chakka, J. Sztrik, and J. Wang. 2014. $M/M/1$ retrial queue with working vacations and negative customer arrivals. *Int. J. Adv. Intelligence Paradigms* 6(1):52–65.
19. Razumchik, R., and M. Telek. 2015. Delay analysis of a queue with re-sequencing buffer and Markov environment. *Queueing Syst.* 22 p. doi: 10.1007/S11134-015-9444-Z.
20. Balsamo, S., P. G. Harrison, and A. Marin. 2010. A unifying approach to product-forms in networks with finite capacity constraints. *SIGMETRICS* 38(1):25–36.
21. Harrison, P. G., and A. Marin. 2014. Product-forms in multi-way synchronisations. *Comput. J.* 57(11):1693–1710.
22. Avrachenkov, K. E., N. O. Vilchevsky, and G. L. Shevljakov. 2003. Priority queueing with finite buffer size and randomized push-out mechanism. *ACM Conference (International) on Measurement and Modelling of Computer Proceedings.* San Diego. 324–335.
23. Ilyashenko, A., O. Zayats, V. Muliukha, and L. Laboshin. 2014. Further investigations of the priority queueing system with preemptive priority and randomized push-out mechanism. *Internet of things, smart spaces, and next generation networks and systems.* Eds. S. Balandin, S. D. Andreev, and Y. Koucheryavy. Lecture notes in computer science ser. Springer. 8638:433–443.
24. Gil, A., J. Seguram, and N. M. Temme. 2007. *Numerical methods for special functions.* Philadelphia, PA: Society for Industrial and Applied Mathematics. 431 p.
25. Erdelyi, A., and H. Bateman. 1985. *Higher transcendental functions.* II. Malabar: Robert E. Krieger Publishing Co. 396 p.

Received October 19, 2015

Contributor

Razumchik Rostislav V. (b. 1984) — 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; associate professor, Peoples’ Friendship University of Russia, 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; rrazumchik@ipiran.ru

АЛГЕБРАИЧЕСКИЙ МЕТОД ПРИБЛИЖЕННОГО РАСЧЕТА СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ В СИСТЕМЕ ОБСЛУЖИВАНИЯ КОНЕЧНОЙ ЕМКОСТИ С ОТРИЦАТЕЛЬНЫМИ ЗАЯВКАМИ И ДВУМЯ ОЧЕРЕДЯМИ

Р. В. Разумчик

Институт проблем информатики Федерального исследовательского центра «Информатика и управление»
Российской академии наук, Российский университет дружбы народов, grazumchik@ipiran.ru

Аннотация: Рассматривается система массового обслуживания с пуассоновским потоком обычных и пуассоновским потоком отрицательных заявок. Для обычных заявок имеется накопитель конечной емкости k . Если обычная заявка при поступлении застаёт накопитель полностью заполненным, она теряется. Отрицательная заявка при поступлении вытесняет одну обычную заявку из очереди в накопителе (если он не пуст) в другую очередь (бункер) конечной емкости r , после чего покидает систему, не оказывая на нее никакого воздействия. Если в момент вытеснения обычной заявки из накопителя бункер полностью заполнен, обе заявки (обычная и отрицательная) покидают систему. В других случаях поступления отрицательной заявки не оказывают влияния на функционирование системы. Заявки из бункера обслуживаются с относительным приоритетом. Времена обслуживания заявок как из накопителя, так и из бункера имеют экспоненциальное распределение с одинаковым параметром. Предложен алгебраический метод приближенного расчета совместного стационарного распределения очередей для случая $k = r$. Представлены некоторые результаты численных экспериментов, показывающие достоинства и недостатки метода.

Ключевые слова: система обслуживания; отрицательные заявки; многочлены Гегенбауэра; стационарное распределение

DOI: 10.14357/19922264150407

Литература

1. Gelenbe E., Glynn P., Sigman K. Queues with negative arrivals // J. Appl. Probab., 1991. Vol. 28. P. 245–250.
2. Gelenbe E. G-networks with instantaneous customer movement // J. Appl. Probab., 1993. Vol. 30. No. 3. P. 742–748.
3. Harrison P. G., Pitel E. Sojourn times in single-server queues with negative customers // J. Appl. Probab., 1993. Vol. 30. P. 943–963.
4. Fourneau J., Gelenbe E., Suros R. G-networks with multiple classes of negative and positive customers // Theor. Comput. Sci., 1996. Vol. 155. No. 1. P. 141–156.
5. Artalejo J. R. G-networks: A versatile approach for work removal in queueing systems // Eur. J. Oper. Res., 2000. Vol. 126. P. 233–249.
6. Bocharov P. P., Vishnevskii V. M. G-networks: Development of the theory of multiplicative networks // Automat. Rem. Contr., 2003. Vol. 64. No. 5. P. 714–739.
7. Queueing networks: A fundamental approach / Eds. Buncherie R. J., N. Van Dijk. — International ser. in operations research and management science. — New York, NY, USA: Springer, 2011. Vol. 154. 800 p.
8. Van Do T. An initiative for a classified bibliography on G-networks // Perform. Evaluation, 2011. Vol. 68. No. 4. P. 385–394.
9. Суетин П. К. Классические ортогональные многочлены. — М.: Наука, 1979. 416 с.
10. Manzo R., Cascone N., Razumchik R. V. Exponential queueing system with negative customers and bunker for ousted customers // Automat. Rem. Contr., 2008. Vol. 69. P. 1542–1551.
11. Klimenok V., Dudin A. A BMAP/PH/N queue with negative customers and partial protection of service // Commun. Stat. Simul., 2012. Vol. 41. No. 7. P. 1062–1082.
12. Pechinkin A. V., Razumchik R. V. Stationary waiting time distribution in queueing system with negative customers and bunker for ousted customers under Last-LIFO-LIFO service discipline // J. Commun. Technol. El., 2012. Vol. 57. No. 12. P. 1331–1339.
13. Pechinkin A. V., Razumchik R. V. A method for calculating stationary queue distribution in a queueing system with flows of ordinary and negative claims and a bunker for superseded claims // J. Commun. Technol. El., 2012. Vol. 57. No. 8. P. 882–891.
14. Dao-Thi T., Fourneau J., Tran M. Networks of order independent queues with signals // 21st Symposium (International) on Modeling, Analysis and Simulation of Computer and Telecommunication Systems Proceedings, 2013. P. 131–140.
15. Krishna Kumar B., Pavai Madheswari S., Anantha Lakshmi S. R. An M/G/1 Bernoulli feedback retrial queue-

- ing system with negative customers // *Oper. Res.*, 2013. Vol. 13. No. 2. P. 187–210.
16. *Разумчик Р. В.* Стационарное распределение времени ожидания в системе обслуживания с отрицательными заявками, бункером для вытесненных заявок, различными интенсивностями обслуживания при дисциплине First-FIFO-FIFO // *Информатика и её применения*, 2013. Т. 7. Вып. 2. С. 34–39.
 17. *Razumchik R. V.* Analysis of finite capacity queue with negative customers and bunker for ousted customers using chebyshev and gegenbauer polynomials // *Asia Pac. J. Oper. Res.*, 2014. Vol. 31. No. 1450029. 21 p.
 18. *Van Do T., Papp D., Chakka R., Sztrik J., Wang J. M.* M/M/1 retrial queue with working vacations and negative customer arrivals // *Int. J. Adv. Intelligence Paradigms*, 2014. Vol. 6. No. 1. P. 52–65.
 19. *Razumchik R., Telek M.* Delay analysis of a queue with resequencing buffer and Markov environment // *Queueing Syst.*, 2015. 22 p. doi: 10.1007/S11134-015-9444-Z.
 20. *Balsamo S., Harrison P. G., Marin A.* A unifying approach to product-forms in networks with finite capacity constraints // *SIGMETRICS*, 2010. Vol. 38. No. 1. P. 25–36.
 21. *Harrison P. G., Marin A.* Product-forms in multi-way synchronisations // *Comput. J.*, 2014. Vol. 57. No. 11. P. 1693–1710.
 22. *Avrachenkov K. E., Vilchevsky N. O., Shevljakov G. L.* Priority queueing with finite buffer size and randomized push-out mechanism // *ACM Conference (International) on Measurement and Modelling of Computer Proceedings*. San Diego, 2003. P. 324–335.
 23. *Ilyashenko A., Zayats O., Muliukha V., Laboshin L.* Further investigations of the priority queueing system with preemptive priority and randomized push-out mechanism // *Internet of things, smart spaces, and next generation networks and systems* / Eds. S. Balandin, S. D. Andreev, and Y. Koucheryavy. — *Lecture notes in computer science ser.* — Springer, 2014. Vol. 8638. P. 433–443.
 24. *Gil A., Seguram J., Temme N. M.* Numerical methods for special functions. — Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007. 431 p.
 25. *Erdelyi A., Bateman H.* Higher transcendental functions. II. — Malabar: Robert E. Krieger Publishing Co., 1985. 396 p.

Поступила в редакцию 19.10.2015

PERFORMANCE IMPROVEMENT OF LEMPEL–ZIV–WELCH COMPRESSION ALGORITHM

S. Frenkel^{1,2}, M. Kopeetsky³, R. Molotkovski³, and P. Borovsky³

Abstract: The paper proposes two novel schemes which improve the dictionary-based Lempel–Ziv–Welch (LZW) compression algorithm. The first scheme proposes an improvement over the LZW algorithm by applying an exponential decay (ED) technique as a tool to manage and remove infrequently used entries in the LZW dictionary. The presented results demonstrate that ED may be an efficient tool to manage and refresh the LZW dictionary. The achieved compression ratio (CR) is higher than in the traditional methods like Dictionary Reset (DR) and Least Recently Used (LRU). Another approach uses the Distance from Last Use (DLU) method. The DLU can be compressed by Huffman coding based on the frequencies of the phrases. The compression scheme, called HCD (Huffman Coding of Distance), was tested on different real-life data types such as text, programming code, audio, video and image files, characterized by different Shannon entropy. The experimental results demonstrate that the ED and HCD scheme may provide higher CR, compared with the LZW algorithm.

Keywords: (LZW) Dictionary Compression; dynamic dictionary; dictionary reset DR; least recently used LRU; exponential decay ED

DOI: 10.14357/19922264150408

1 Introduction

Data compression based on dynamic dictionary is a commonly used form of lossless data compression algorithms. A dictionary that starts from a predetermined state, nevertheless its values can be changed during the encoding or decoding process, is used. One of the most well known dynamic dictionary based compression algorithms is the LZW algorithm which was proposed by Abraham Lempel, Jacob Ziv, and Terry Welch as a modification of the ZL77 algorithm [1–3]. The LZW algorithm is a general purpose algorithm in the sense that it can work with almost any type of data. The dictionary that is used in LZW stores sets of codes and strings of characters. At the initial state, only one length strings are stored (usually, ASCII characters) in the dictionary; however, during the encoding process, strings that have not been encountered previously are added to the dictionary. Compression is achieved by replacing strings with codewords (indexes) that are associated with them. Similarly, decompression is achieved by replacing sets of codewords (indexes) with the associated strings.

One of the main factors that should be taken into account in the LZW compression is the amount of bits that should be allocated to each codeword, namely, codeword length, which determines the amount of unique codewords that can be stored in the dictionary at any given time. The drawback of large dictionaries is that the larger the dictionary is, the more space each codeword

will require that can effect negatively the achievable CR. To avoid this problem, there are several solutions that are designed to manage the dictionary after it becomes full. This paper proposes some improvements over LZW algorithm by employing new methods of LZW dictionary management, namely, ED as a tool to manage and remove infrequently used entries in the LZW dictionary, and by using of knowledge about the distance between two sequential phrases, DLU. Distance from last use is compressed by Huffman coding based on the frequencies of the codewords.

Paper organization. The paper is organized as follows: Section 2 introduces several implementations of LZW, which are designed to manage dictionary size. The ED scheme, its time performance, and optimization are discussed in section 2. Improvement of LZW Performance by using of the DLU is presented in section 3. Conclusions of section 4 complete the paper.

2 Lempel–Ziv–Welch Algorithm and Dynamic Dictionaries

The LZW algorithm is a lossless data compression algorithm that is designed to compress any kind of data. Some of its main uses are in image compression like GIF (graphics interchange format) and TIFF (tagged image file format). The LZW algorithm is one of the compression algorithms that take advantage of a dynamic

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

²Moscow State University of Information Technologies, Radioengineering, and Electronics, 78 Vernadskogo Ave., Moscow 119454, Russian Federation

³Department of Software Engineering, Shamoon College of Engineering, Basel/Bialik Sts, Beer-Sheva, Israel

dictionary during the encoding and decoding processes. A dictionary is a table where each its entry holds pairs of indexes and strings. At the initial state, a dictionary will contain only one length strings; however, during encoding and decoding processes, new sets of strings and indexes are added dynamically. Compression is achieved by replacing sets of strings with their corresponding indexes. Similarly, during decompression, in order to recover the original file, the indexes are replaced with their corresponding strings. A resulting CR is a main factor that determines the compression quality and it is defined as $CR = L_0/L_{LZW}$ where L_0 and L_{LZW} are the lengths of the original and compressed file, respectively.

Limiting the size of the dictionary. The key factors that affect the achievable CR are the dictionary's size and managing the dictionary after it becomes full. For a dictionary that can store n unique indexes, each index is of length $\log_2 n$ bits. Therefore, the larger dictionary is, the more space each index will require, which, in turn, can result in a low CR. This means that for achieving a more efficient compression, the dictionary's size should be limited. One of the most used methods in handling a full dictionary is known as DR [2]. Every time the dictionary becomes full, it is restored to its initial state in which it stores only one length strings. The main disadvantage of the DR method is that all of the dictionary's entries are treated equally, regardless of whether they are repeated frequently or not. Generally, the methods that can remove infrequently used entries while keeping the frequent ones, usually achieve higher CR. Some of the existing frequency based update methods are LRU and Least Frequently Used (LFU) [4].

This paper introduces and analyzes a new ED method for removing infrequently used entries which makes use of exponential decay as a tool to sort the entries dynamically. Note that the exponentially decaying probabilities for Huffman coding were used in [5] for input rather than for a dictionary content.

2.1 Exponential decay and other methods for managing dictionary size

Dictionary reset. This method employs the removal of all data from dictionary except one length strings which together determine the initial state when the dictionary becomes full.

Least recently used. In the LRU method, each dictionary entry holds a value that indicates the last time it was used (LTU). In order to free space in a full dictionary, the entry which was LRU is removed (it is an entry with the lowest LTU value).

Exponential decay. This method works similarly to the LRU method in the sense that when the dictionary becomes full, a single entry is removed in order

to free space. However, unlike the LRU implementation, each dictionary entry holds two values in addition to the entry's code and string: entry creation time (ECT) and number of times used (NTU). When choosing which entry should be removed from the dictionary, a decay value $N(t)$ is calculated for each entry as $N(t) = N(0)e^{-\lambda t} + NTU$ where $N(0)$ is the initial value; λ is the decay rate; t is the current time minus ECT; and NTU is equal to the number of times a string was used. The entry with the lowest $N(t)$ value is removed from the dictionary.

Encoding and decoding runtime. In the complexity analysis of all of the mentioned LZW-based algorithms, string lookup implementation was used by means of a binary search tree. Therefore, lookup, insertion, and deletion operations are implemented in the cost of $O(\log_2 n)$. Dictionary reset is the fastest method since entry removal is done only every n steps where n is the maximum number of entries in the dictionary. In the present LRU implementation, minimal NTU value lookup was done using binary search tree. Although being slightly more costly than the basic default LZW implementation, overall time complexity for each operation remained the same $O(\log_2 n)$. Unlike DR and LRU, ED method is more costly since minimal $N(t)$ value lookup requires iteration over every entry in the dictionary with iteration cost of $O(n)$.

Testing and simulation. In order to investigate performance of LZW based on DR, LRU, and ED methods, the files of different formats have been tested. The results are presented in Fig. 1.

Evaluation of CR for DR, LRU and ED is performed for $N(0) = 1$ and $\lambda = 0.001$. Since dictionary size significantly influences the CR, the following results demonstrate CR as a function of the dictionary size for the range of codewords lengths from 9 to 19 bit (for each method).

Figure 1 demonstrates that ED provides higher CR compared with DR and LRU methods. Note that LZW with ED can provide higher CR compared with basic LZW algorithm for rather short files (see Fig. 1c) since the asymptotic properties of Lempel–Ziv algorithms [6, 7] are still not satisfied.

Optimization of the ED method. Unlike DR and LRU, ED method is characterized by two additional parameters $N(0)$ and λ that may affect the achieved CR. It should be reminded that $N(0)$ determines the initial value when the entry is created and λ determines how fast an entry's value will decay as it ages.

Optimization of $N(0)$. The provided tests demonstrate that $N(0)$ has an impact on CR only in the small dictionaries with codewords at most 13 bit length (Table 1). The shorter codeword is, the more significant impact of $N(0)$ on the CR is. For the codewords longer than 13 bit,

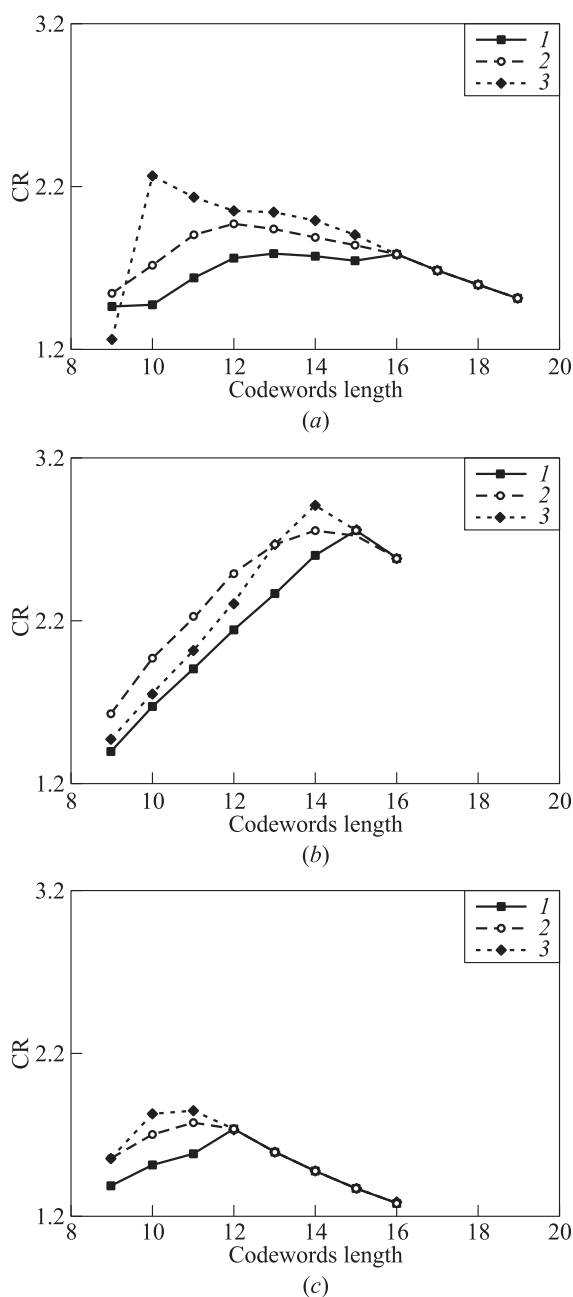


Figure 1 Compression ratios: (a) BW_Photo.bmp (233 KB); (b) combo002.css (136 KB); and (c) aooles.txt (8.75 KB)

the impact of $N(0)$ on CR is negligible. In fact, the value of $N(0)$ is not relevant for the files with the large dictionary size (while codeword length is larger than 13 bit in the present tests).

Optimization of λ . Parameter λ can affect the CR significantly since a different decay ratio can cause a different order of entry removal during encoding and decoding process. In the tests for the optimal value of λ , the optimal dictionary size (codeword length) was used for each file (Table 2).

Table 1 Compression ratio for files BW_Photo.bmp, apples.txt, and earth.avi with $N(0)$ values for codeword length ranging between 1 and 10 bit

$N(0)$	BW_Photo.bmp at 10 bit	apples.txt at 11 bit	earth.avi at 13 bit
1	2.2746	1.9227	2.1574
2	2.2998	1.9227	2.1572
3	2.2070	1.9227	2.1554
4	2.1249	1.9227	2.1528
5	2.0943	1.9215	2.1492
6	2.0301	1.9162	2.1458
7	1.9798	1.9080	2.1410
8	1.9510	1.8975	2.1373
9	1.9156	1.8939	2.1337
10	1.8941	1.8859	2.1327

Table 2 Compression ratios for Yahoo UK.htm with dictionary size at 15 bits, combo002.css with dictionary size at 14 bits, and earth.avi file with codeword length 13 bits

λ	Yahoo UK.htm	combo002.css	earth.avi
10^{-5}	2.0022	2.9414	2.1574
10^{-3}	2.0022	2.9414	2.1574
10^{-1}	2.0010	2.9401	2.1602
10^1	1.9681	2.9048	2.1534

The presented results demonstrate that λ has small or negligible influence on the CR. Generally, smaller λ values tend to produce slightly better compression. Additionally, during testing, it was noticed that using large λ values can be problematic since they can cause too small $N(t)$ values for any floating point variables to handle. Hence, unpredictable behavior and even failure to reconstruct the original file correctly can occur.

3 Improvement of Lempel–Ziv–Welch Performance by Using of the Distance from Last Use

A Huffmann Coding of Distance scheme that combines a dictionary-based LZW algorithm [2] and well-known sliding window LZ77 algorithm [6] has been proposed. For each phrase to be encoded, the distance between the current position and the previous one (namely, the DLU) is calculated and encoded by the Huffman algorithm. As a result, the generated codewords are shorter and the CR is higher. The experimental results presented in Fig. 2 demonstrate that as short the mean distance between phrase occurrences is, the improvement of compression ratio is more significant.

A new phrase is inserted each time to the dictionary in position “0” and all the other phrases are moved to

FileName	File Size	Classic LZW	LZW with DCH	Avarage Distance	Better Algorithm	Difference
Thumbs.db	37.0KB	0.948851	1.122085	146.127749	LZW with DCH	15%
ColorfullPicture4.bmp	269.0KB	0.954555	1.108645	191.393848	LZW with DCH	14%
combo.css	49.0KB	1.338224	1.54723	283.49459	LZW with DCH	14%
speech_32kbps_fb.wav	1013.0KB	1.1686	1.317057	263.724276	LZW with DCH	11%
ColorfullPicture1.bmp	326.0KB	1.177771	1.320835	236.798932	LZW with DCH	11%
Yahoo UK.htm	288.0KB	1.540207	1.703037	328.864892	LZW with DCH	10%
6_Channel_ID.wav	2.0MB	5.398812	5.938102	319.662332	LZW with DCH	9%
r-csc.htm	6.0KB	1.487954	1.613798	298.465927	LZW with DCH	8%
ColorfullPicture3.bmp	250.0KB	1.373437	1.486304	253.739556	LZW with DCH	8%
M1F1-int24-AFsp.wav	137.0KB	1.578284	1.700655	299.00218	LZW with DCH	7%
g-r-min.js	125.0KB	1.587094	1.705418	359.698917	LZW with DCH	7%
sample.wav	68.0KB	1.643014	1.764959	393.150951	LZW with DCH	7%
Microsoft – Official Home Page.htm	469.0KB	1.986475	2.120882	400.274717	LZW with DCH	6%
ColorfullPicture2.bmp	605.0KB	2.293332	2.43224	431.536804	LZW with DCH	6%
combo_003.js	35.0KB	1.755063	1.858265	388.352804	LZW with DCH	6%
combo.js	238.0KB	1.813116	1.917195	397.508152	LZW with DCH	5%
r-sf.htm	4.0KB	1.377003	1.453412	315.215466	LZW with DCH	5%
combo_002.js	292.0KB	1.83563	1.936179	408.413156	LZW with DCH	5%
earth.avi	162.0KB	1.968642	2.075158	296.739646	LZW with DCH	5%
combo_002.css	136.0KB	2.239666	2.351608	443.292855	LZW with DCH	5%
BW_Photo3.bmp	241.0KB	2.151351	2.220986	403.563923	LZW with DCH	3%
stereofl.wav	228.0KB	2.25647	2.318941	366.399494	LZW with DCH	3%
adventur.txt	7.0KB	2.073223	2.111595	447.488644	LZW with DCH	2%
914bbs.txt	3.0KB	1.746479	1.778575	343.817932	LZW with DCH	2%
apples.txt	8.0KB	1.835313	1.865112	375.986172	LZW with DCH	2%
BW_Photo.bmp	233.0KB	1.840946	1.870299	403.954537	LZW with DCH	2%
M1F1-float64-AFsp.wav	367.0KB	3.379871	3.433245	400.50319	LZW with DCH	2%
BW_Photo2.bmp	265.0KB	2.881363	2.916726	442.58681	LZW with DCH	1%
balls.txt	2.0KB	1.745399	1.764888	332.101227	LZW with DCH	1%
cs_0.js	1.0KB	1.240175	1.229437	273.25655	Classic LZW	1%
BW_PictureOfText1.bmp	330.0KB	7.338325	7.315622	479.859728	Classic LZW	0%
BW_PictureOfText2.bmp	541.0KB	12.28593	12.279669	466.579322	Classic LZW	0%

Figure 2 Comparison of CR of original implementation of LZW vs. LZW with HDC

the right, and the last phrase is removed in the HCD scheme. In order to encode a phrase that is already in the dictionary, its position was simply encoded. So, if the dictionary is maintained in this manner, the position of the phrase is the distance from the last occurrence of this phrase.

The HCD scheme is based on the Kac's lemma [8] that states that the average distance between occurrences of a pattern is equal to the inverse of the probability of the pattern. Hence, it is reasonable to assume that if a phrase has been recently used, there is a higher probability that it will appear again. Based on the Kac's lemma, the HCD scheme permanently computes the frequencies of different phrases and generates the Huffman codes by representing the more frequent phrases by a smaller codeword. As a result, if most of the phrases appear close to each other in a particular file, then this file will be efficiently compressed, providing high CR. As a matter of fact, the similar approach has achieved great success in the window-based LZ77 compression algorithm [6]. The experimental results presented in section 4 demonstrate that this approach is effective in the dictionary-based LZW [2] compression algorithm.

Related work in combination of LZ, and Huffmann coding.

There are several papers that use the similar approach to improve LZ77 compression algorithm. The Huffman encoding is used for a range that will group several DLUs together. Only the range is compressed by the Huffman code and the index is encoded as is. The idea of the HCD scheme is basically the same; nevertheless, HCD

is designed for the sliding window LZ77-like scheme. A similar approach was used in Deflate [9], a data compression algorithm that uses a combination of the LZ77 algorithm and the Huffman coding. It was originally defined by Phil Katz for version 2 of his PKZIP archiving tool and was later specified in RFC 1951. This scheme is also based on the sliding window LZ77 algorithm [6], while the HCD scheme is focused on the modification of the dictionary-based LZW algorithm [6].

There are several research papers that propose a combination of the LZW algorithm and Huffman coding [10, 11]. Nevertheless, unlike HCD scheme, none of them mention the method for dictionary maintenance (such as LRU/First In First Out (FIFO) approach) that is the basic idea in HCD.

Paper [12] proposes the similar solution with one difference that the generation of the Huffman codes is based on the frequency of specific entries in the dictionary and not on the distance from last use divided into ranges.

In order to maintain the dictionary generated by the LZW algorithm, the FIFO method was used, which is quite similar approach to LRU. The reason is that FIFO allows one to calculate the distance between elements immediately instead of looping on the linked LRU list each time (in order to calculate a distance). The same issue was described in the USA patent [13]. The proposed solution is applying a Move Ahead Technique: every time if a phrase occurs in the dictionary, it is moved ahead. On the one hand, the proposed method could possibly

decrease the performance in terms of the compression speed which is very important for the compression algorithm. On the other hand, FIFO approach used in the described scheme, does not cause decrease of the compression speed.

Architecture of the Huffman coding of distance scheme.

The presented scheme is designed from the layer added on top of the LZW scheme. This approach allows one to modify and improve each layer independently on other layers:

- (1) input–output stream is responsible to communicate with the file system, read the input to be processed, and write the generated output;
- (2) LZW layer is the original implementation of the LZW algorithm and it does not differ from the original LZW code;
- (3) the distance of the phrase generated by the LZW from its last use is computed in the HCD component and the records of phrases use are maintained; and
- (4) the Huffman coding layer is the part where the HCD improvement is introduced and the code generated by the LZW scheme is compressed by utilizing the frequency of the different distance ranges. Figure 3 describes the pseudocode of the HCD scheme for the case when the codeword length is 10 bit.

```

1: HCD scheme: Encoder
2: Calculate the distance of the code from the FIFO head
3: Push the code into a FIFO
4: Divide the distance into two parts:
5: length 5 is a half):
6: 5 least significant bits
7: Encode the most significant bits with Huffman code
   and write it to the output
8: Write the least significant bits to the output
9: HCD scheme: Decoder
10: Decode 5 most significant bits by Huffman code
11: Read 5 least significant bits from input
12: Return the resulted code to the LZW part
    
```

Figure 3 The HCD scheme

The HCD scheme provides higher CR for the majority of the tested files. Specifically, the experimental results of Fig. 2 demonstrate that the HCD scheme can significantly improve the compression quality for most file types. According to present experiments, three factors were identified that are crucial to the CR of the LZW algorithm with the HCD scheme. These factors are the size of the LZW dictionary, the division of the distance to ranges, and the period of recalculating the Huffman code.

4 Discussion of Experimental Results and Concluding Remarks

The presented results demonstrate that there are various ways for improvement of LZW algorithm by the additional techniques for maintenance of a dictionary. In particular, ED may be an efficient tool to manage and refresh the LZW dictionary. The achieved CR is higher than in the traditional methods like DR and LRU. However, the encoding and decoding time is larger. It was demonstrated that the codeword length (dictionary size) has a significant impact on the achieved CR. Moreover, different files may require different codeword lengths in order to achieve an optimal CR. As for addition of the LZW with HCD, there is usually increase of the CR. Nevertheless, usually, ED technique is slightly better. Hence, there is a need for analytical and experimental research of the ED method in order to determine better relationship between file structure and values of parameters $N(0)$ and λ . Taking into account that both considered techniques differently affect the LZW mechanisms, it would be interesting to consider their joint implementation.

Acknowledgments

The first author has partially been supported by the Russian Foundation for Basic Research under grant 15-07-05316. The second author has partially been supported by the internal research program of the Shamoon College of Engineering. The authors thank Shlomi Dolev for the idea to apply the exponential decay method in the LZW algorithm.

References

1. Ziv, J., and A. Lempel. 1978. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inform. Theory* 5(24):530–536.
2. Welch, T. 1984. A technique for high-performance data compression. *Computer J.* 17(6):8–19.
3. Solomon, D. 2004. *Data compression: The complete reference*. 3rd ed. Springer. P. 206–208.
4. Asit, D., and T. Don. 1990. An approximate analysis of the LRU and FIFO buffer replacement schemes. *ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems SIGMETRICS Proceedings*. 143–152.
5. *Tcheslavski, G. V.* 2008. Basic Image compression methods. Available at <http://ee.lamar.edu/gleb/dip/index.htm> (accessed December 7, 2015).
6. Ziv, J., and A. Lempel. 1977. A universal algorithm for sequential data compression. *IEEE Trans. Inform. Theory* 3(23):337–343.

7. Ziv, J. 2009. The universal LZ77 compression algorithm is essentially optimal for individual finite-length-blocks. *IEEE Trans. Inform. Theory* 5(55):1941–1944.
8. Kac, M. 1947. On the notion of recurrence in discrete stochastic processes. *Bull. Am. Math. Soc.* 53.10. 1002–1010.
9. Katz, P. W. 1991. String searcher, and compressor using same. U.S. Patent No. 551745.
10. Perl, Y., and A. Mehta. 1991. Cascading LZW algorithm with Huffman coding: A variable to variable length compression algorithm. *Computing in the 90's*. Eds. N. A. Sherwani, E. de Doncker, and J. A. Kapenga. Lecture notes in computer science ser. Springer. 507:170–178.
11. Yikun, Z. 2010. An lidar data compression method based on improved LZW and Huffman algorithm. *IEEE Conference (International) on Electronics and Information Engineering (ICEIE)*. 2:250–254.
12. Raja, P., and D. Saraswathi. 2011. An effective two stage text compression. *Int. J. Electron. Commun. Eng.* 4(2):233–241.
13. Jaquette, G. A. 1995. Literal handling in LZ compression employing MRU/LRU encoding. U.S. Patent No. 6218970.

Received October 5, 2015

Contributors

Frenkel Sergey L. (b. 1951) — 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 119333, Russian Federation; associate professor, Moscow State University of Information Technologies, Radioengineering, and Electronics, 78 Vernadskogo Ave., Moscow 119454, Russian Federation; fsergei51@gmail.com

Kopeetsky Marina (b. 1966) — PhD in Computer Sciences, senior lectures, Department of Software Engineering, Shamoon College of Engineering, Basel/Bialik Sts, Beer-Sheva 84100, Israel; marinako@sce.ac.il

Molotkovsky Roman (b. 1989) — Bachelor of Sciences in Computer Sciences, Department of Software Engineering, Shamoon College of Engineering, Basel/Bialik Sts, Beer-Sheva 84100, Israel; megaclaff@gmail.com

Borovsky Pavel (b. 1986) — Master of Science in Computer Science, Department of Software Engineering, Shamoon College of Engineering, Basel/Bialik Sts, Beer-Sheva 84100, Israel; pavelbsky@gmail.com

УЛУЧШЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ АЛГОРИТМА СЖАТИЯ ДАННЫХ ЛЕМПЕЛЯ–ЗИВА–УЭЛЧА

С. Френкель¹, М. Копицкая², Р. Молотковский³, П. Боровский⁴

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский государственный университет информационных технологий, радиотехники и электроники; fsergei51@gmail.com

²Академический инженерный колледж Шамуна, Беер-Шева, Израиль; marinako@sce.ac.il

³Академический инженерный колледж Шамуна, Беер-Шева, Израиль; megaclaff@gmail.com

⁴Академический инженерный колледж Шамуна, Беер-Шева, Израиль; pavelbsky@gmail.com

Аннотация: Предлагаются две новые схемы, улучшающие сжатие данных при использовании алгоритма Лемпеля–Зива–Уэлча (LZW). Первая схема основана на управлении размером словаря методом “exponential decay” (ED) путем удаления редко встречающихся записей словарной таблицы. Представленные результаты показывают, что ED является эффективным инструментом управления и обновления словаря. Достигнутый коэффициент сжатия выше, чем при использовании традиционных методов повышения эффективности LZW, например, таких как Dictionary Reset (DR) и Least Recently Used (LRU). Вторая схема, названная Huffman Coding of Distance (HCD), основана на учете расстояния в словаре до слова, использованного при прошлом обращении, называемого Distance from Last Use (DLU). Величина DLU может быть сжата кодом Хаффмана. Эта схема тестировалась на различных типах данных, таких как текстовые, коды программ, графические, аудио- и видеоформаты. Экспериментальные результаты показывают, что как ED, так и HCD обеспечивают более существенное сжатие, чем обычный LZW.

Ключевые слова: сжатие данных; алгоритмы Лемпеля–Зива; динамический словарь; LRU; экспоненциальное затухание

DOI: 10.14357/19922264150408

Литература

1. Ziv J., Lempel A. Compression of individual sequences via variable-rate coding // IEEE Trans. Inform. Theory, 1978. Vol. 5. No. 24. P. 530–536.
2. Welch T. A technique for high-performance data compression // Computer J., 1984. Vol. 17. No. 6. P. 8–19.
3. Solomon D. Data compression: The complete reference. — 3rd ed. — Springer, 2004. P. 206–208.
4. Asit D., Don T. An approximate analysis of the LRU and FIFO buffer replacement schemes // ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems SIGMETRICS Proceedings, 1990. P. 143–152.
5. Tcheslavski G. V. Basic Image compression methods. 2008. <http://ee.lamar.edu/gleb/dip/index.htm>.
6. Ziv J., Lempel A. A universal algorithm for sequential data compression // IEEE Trans. Inform. Theory, 1977. Vol. 3. No. 23. P. 337–343.
7. Ziv J. The universal LZ77 compression algorithm is essentially optimal for individual finite-length-blocks // IEEE Trans. Inform. Theory, 1977. Vol. 5. No. 55. P. 1941–1944.
8. Kac M. On the notion of recurrence in discrete stochastic processes // Bull. Amer. Math. Soc. 53.10, 1947. P. 1002–1010.
9. Katz P. W. String searcher, and compressor using same. U.S. Patent No. 551745, 1991.
10. Perl Y., Mehta A. Cascading LZW algorithm with human coding: A variable to variable length compression algorithm // Computing in the 90's / Eds. N. A. Sherwani, E. de Doncker, J. A. Kapenga. — Lecture notes in computer science ser. — Springer, 1991. Vol. 507. P. 170–178.
11. Yikun Z. An lidar data compression method based on improved LZW and Huffman algorithm // IEEE Conference (International) on Electronics and Information Engineering (ICEIE), 2010. Vol. 2. P. 250–254.
12. Raja P., Saraswathi D. An effective two stage text compression // Int. J. Electron. Commun. Eng., 2011. Vol. 4. No. 2. P. 233–241.
13. Jaquette G. A. Literal handling in LZ compression employing MRU/LRU encoding. U.S. Patent No. 6218970, 1995.

Поступила в редакцию 05.10.2015

ОЦЕНКИ СКОРОСТИ ПЕРЕДАЧИ ИНФОРМАЦИИ И ПРОПУСКНОЙ СПОСОБНОСТИ В СКРЫТЫХ КАНАЛАХ С МЕТКАМИ*

А. А. Грушо¹, Н. А. Грушо², Е. Е. Тимонина³

Аннотация: Работа посвящена оценкам пропускной способности и скорости передачи информации в скрытых каналах специального вида. Эти скрытые каналы порождаются с помощью сигналов (меток), не несущих самостоятельной информации, но легко выделяемых на приемном конце. Такие метки служат своего рода запретами с точки зрения допустимых значений всех параметров, связанных с передачей легальной информации. Скрываемая информация кодируется длинами выделяемых метками фрагментов данных в легальной передаче.

Ключевые слова: скрытый канал; теоретико-вероятностные модели скрытых каналов; скрытые каналы, порожденные метками; пропускная способность скрытого канала; скорость передачи

DOI: 10.14357/19922264150409

1 Введение

Впервые понятие скрытого канала было введено в работе Лэмпсона [1] в 1973 г. Канал называется скрытым, если он не проектировался, не предполагался для передачи информации в электронной системе обработки данных.

Свойствам скрытых каналов посвящено большое количество научных работ. История исследований скрытых каналов отражена, например, в обзоре [2]. Основные проблемы, связанные со скрытыми каналами, следующие:

- разработка способа передачи и получения скрываемой информации;
- обеспечение невыявляемости («невидимости») скрытого канала;
- оценка пропускной способности и скорости передачи информации по скрытому каналу;
- обеспечение устойчивости скрытой передачи информации к методам защиты от скрытых каналов.

Рассматриваемый в статье класс скрытых каналов относится к каналам по времени [2]. В каналах этого типа используются метки, т. е. не несущие самостоятельной информации сигналы, позволяющие выделять (помечать) символы в легальной

передаче данных. Метки могут быть отделены от символов легальной передачи данных, но появление меток должно синхронизироваться с появлением символов легальной передачи [3, 4].

Для организации скрытых каналов в компьютерных системах и сетях необходимо, как правило, наличие программно-аппаратных агентов (сущностей) отправителя и получателя скрываемой информации. В рамках данной статьи предполагается наличие таких агентов. Кроме задач по организации взаимодействия с помощью скрытого канала агенты должны обладать единым способом кодирования информации на передающем конце и декодирования на приемном конце.

Передача информации по скрытому каналу, использующему метки, основана на том, что агент-отправитель может устанавливать метку синхронно с появлением выбранного символа легальной передачи данных, а агент-получатель должен выявлять присланную метку и на основании этого выделять символ, к которому она относится. Реализация передачи основана на системе счетчиков, которые вычисляют длины фрагментов легальной передачи данных, определяемых метками. Эти длины могут записываться в память без декодирования и передаваться по другим скрытым каналам с использованием, может быть, других меток. Таким образом

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 13-01-00215).

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

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

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

может быть организован транзит скрытой передачи в различных средах.

Пример скрытого канала данного типа внутри компьютерной системы приведен в работе [3]. В этом примере в качестве меток использовались изменения формы электрических сигналов, с помощью которых передавались данные в компьютере. Передающие и принимающие микропроцессоры могут формировать и идентифицировать изменения формы электрических сигналов. Однако в цифровых данных, используемых компьютером, такие изменения не идентифицируются.

Исследованиям пропускной способности скрытых каналов также посвящено значительное число работ. Наиболее близкой к проблематике данной статьи является работа [5]. В ней также рассматривается организация скрытого канала с помощью генерации объектов синхронизации. Однако сам канал строится на других принципах, а именно: на принципах удаления и вставки бит в битовую легальную передачу данных.

2 Скрытые каналы с метками

Предположим, что время дискретно и легальная информация представляет собой последовательность данных, каждый элемент которых является буквой некоторого конечного алфавита. Буквы на разных местах передаваемой последовательности могут принадлежать различным алфавитам. Метки определяются некоторым способом выделения символов в передаваемой последовательности легальных данных. Метки разбивают последовательность данных на фрагменты. В скрытых каналах с метками информация передается такими фрагментами [3, 4]. Длина фрагмента — число знаков в нем.

Определим код в рассматриваемом скрытом канале. Кодироваться слова и символы некоторого языка, описывающего скрываемые данные. Эти элементы будем называть кодовыми величинами (КВ). Будем считать, что код состоит из N КВ. Каждая КВ кодируется наборами длин фрагментов $L = (l_1, \dots, l_k)$, которые будем называть кодовыми значениями (КЗ). Длины отсчитываются от одного начала фрагмента. Для каждой КВ выбирается свой набор длин, и число их, вообще говоря, может быть различным.

Для простоты в рамках данной статьи будем считать, что число k для всех КЗ одинаково и все $l_i \leq n, i = 1, \dots, k$. Тогда передача КЗ с помощью скрытого канала описывается блочным кодом [6] с длиной блока n . Таким образом, начало всех КЗ в последовательности точно определено.

В скрытом канале с метками предполагается наличие шума, который выражается в случайном появлении меток в случайных местах. Будем считать, что метки появляются случайно и независимо друг от друга с одинаковой вероятностью p . Появление на одном месте двух меток: случайной и неслучайной — равносильно появлению одной метки.

Согласно [6], скорость передачи одной КВ в данной модели равна n . В работе строится оценка сверху для минимальной скорости передачи при заданном ограничении ε на вероятность ошибочного декодирования.

Определим подробнее процедуру кодирования и декодирования в скрытом канале с метками. Каждая из N возможных КВ источника скрытого сообщения по заданной таблице кодируется вектором $L^{(i)} = (l_1^{(i)}, \dots, l_k^{(i)})$, $i = 1, \dots, N$. При этом выполняется следующая цепочка неравенств:

$$l_1^{(i)} < l_2^{(i)} < \dots < l_k^{(i)} \leq n, \quad i = 1, \dots, N.$$

В передаваемой легальной последовательности данных, начиная от точки отсчета, кратной n (начало блока), ставятся неслучайные метки на символах, расположенных от начала блока на расстояниях $l_1^{(i)}, \dots, l_k^{(i)}$, если передается КЗ $L^{(i)}$. На приемном конце известны точки начала каждого КЗ. Тогда приемник скрытого сообщения определяет наличие меток на местах $l_1^{(i)}, \dots, l_k^{(i)}$ и идентифицирует полученное КЗ $L^{(i)}$.

Возможный шум может породить ложное принятое КЗ. Возникшая неопределенность декодирования в рассматриваемом случае считается ошибкой при передаче по скрытому каналу. Именно вероятность появления ложных КЗ за счет шума ограничивается заданным параметром $\varepsilon, 0 < \varepsilon < 1$. Следовательно, истинное КЗ обязательно выявляется в рассматриваемой модели, но вероятность возникновения неопределенности при передаче каждого КЗ в скрытом канале ограничена ε .

3 Кодирование с помощью непересекающихся длин в кодовых значениях

Рассмотрим простейшую модель кода в скрытом канале с метками и построим оценку скорости передачи в нем.

Код записывается в виде матрицы $\mathcal{L} = \|a_{ij}\|$ размера $N \times n$, где $a_{ij} = 1$ тогда и только тогда, когда $L^{(i)}$ имеет фрагмент длины j . В противном случае $a_{ij} = 0$. В каждой строке матрицы \mathcal{L} находится ровно k единиц.

Простейший код характеризуется тем, что в каждом столбце матрицы \mathcal{L} не больше одной единицы. Это условие означает, что длины фрагментов в кодовых словах для различных КЗ не пересекаются. Легко подсчитать число матриц \mathcal{L} , удовлетворяющих таким условиям: $n!/((k!)^N(n - kN)!)$.

Таким образом, множество возможных кодов не пусто, когда выполнено следующее неравенство:

$$n \geq Nk. \quad (1)$$

Вероятность случайного порождения данного КЗ равна p^k .

Из непересекаемости длин фрагментов в кодовых словах следует независимость случайного возникновения различных ложных КЗ, поэтому вероятность появления хотя бы одного ложного КЗ равна $1 - (1 - p^k)^{N-1}$.

По условию эта вероятность ограничена параметром ε ; следовательно, получаем уравнение для k , обеспечивающего заданную малость вероятности ошибки:

$$\varepsilon = 1 - (1 - p^k)^{N-1}.$$

Отсюда k имеет следующий вид:

$$k = \left\lceil \frac{\ln \left(1 - (1 - \varepsilon)^{1/(N-1)} \right)}{\ln p} \right\rceil, \quad (2)$$

где $[x]$ означает наименьшее целое, большее или равное x .

Из формул (1) и (2) следует оценка скорости передачи по скрытому каналу:

$$n = N \left\lceil \frac{\ln \left(1 - (1 - \varepsilon)^{1/(N-1)} \right)}{\ln p} \right\rceil.$$

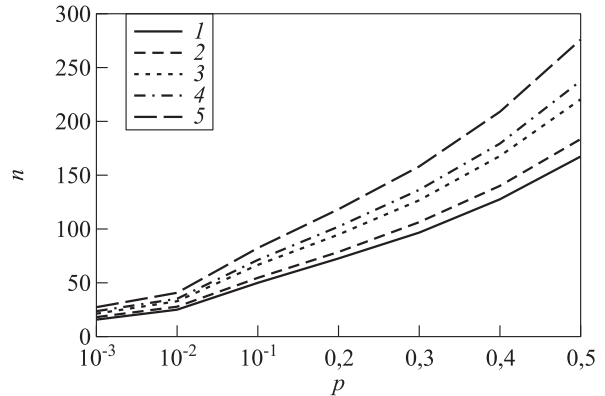
Приведем примеры численных расчетов характеристик скрытого канала в зависимости от мощности алфавита скрытого канала N , допустимой вероятности неоднозначного декодирования ε и вероятности случайного появления меток p .

Из рисунка следует, что построение скрытого канала указанного типа возможно в достаточно широком спектре значений определяющих параметров.

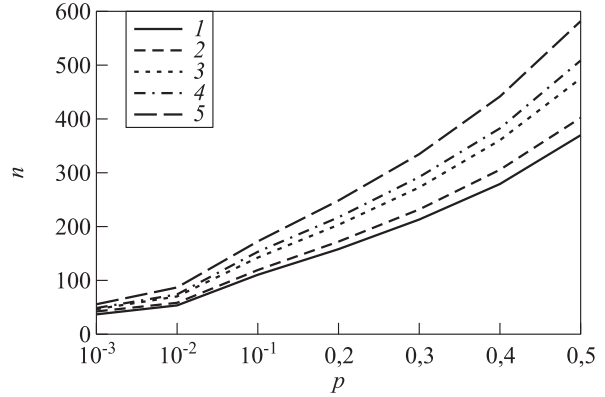
Алгоритм декодирования в рассматриваемом случае можно уточнить следующим образом.

1. От начала блока отсчитываются расстояния $l_1^{(1)}, \dots, l_1^{(N)}$.
2. Выделяются фрагменты, в которых на этих расстояниях есть метки. Пусть эти фрагменты определяются длинами $l_1^{(i_1)}, \dots, l_1^{(i_j)}$. Верхние индексы определяют потенциальные КЗ:

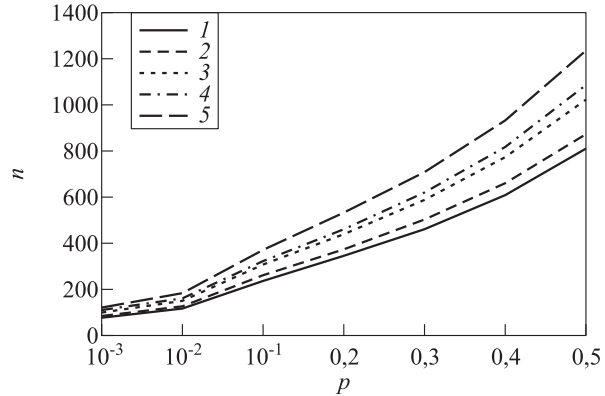
$$\mathcal{L}_1 = \{L^{(i_1)}, \dots, L^{(i_j)}\}.$$



(а)



(б)



(в)

Число знаков легальной передачи на один знак скрытого канала при размере кода скрытой передачи $N = 16$ (а), 32 (б) и 64 (в): 1 — $\varepsilon = 0,01$; 2 — 0,005; 3 — 0,001; 4 — 0,0005; 5 — 0,0001

3. Для выделенных КЗ находятся метки на местах $l_2^{(i_1)}, \dots, l_2^{(i_j)}$. Эта процедура выделяет новый набор КЗ:

$$\mathcal{L}_2 = \{L^{(t_1)}, \dots, L^{(t_s)}\} \subseteq \mathcal{L}_1.$$

4. Продолжая такой отсев, можно получить единственное кодовое слово, которое было передано, или несколько КЗ, что означает ошибку.

В зависимости от свойств исходных сообщений (например, перечисление некоторых свойств) в одном блоке можно передавать несколько КЗ. В этом случае следует считать, что алгоритм декодирования определяет несколько КЗ, не считающихся ошибкой. Для этого случая вероятность ошибки оценивается несколько иначе.

Пусть допускается передача в одном блоке ровно r КЗ. Тогда среднее появлений хотя бы одного ложного КЗ равна $r(1 - (1 - p^k)^{N-r})$. Приравняв эту величину к уровню допустимой ошибки ε , получим:

$$k = \left\lceil \frac{\ln \left(1 - (1 - \varepsilon/r)^{1/(N-r)} \right)}{\ln p} \right\rceil. \quad (3)$$

Из (1) и (3) следует оценка скорости передачи в скрытом канале:

$$n = kN = N \left\lceil \frac{\ln \left(1 - (1 - \varepsilon/r)^{1/(N-r)} \right)}{\ln p} \right\rceil.$$

Из возможности передачи (используя особенности семантики) одновременно нескольких кодовых слов в одном блоке следует, что полученная оценка скорости передачи может оказаться строгой оценкой сверху. Полученные оценки являются оценками пропускной способности [6].

4 Оценки скорости передачи скрытого канала при использовании специальных кодов

Пропускную способность скрытого канала с метками можно увеличить, если построить код в соответствии с некоторыми дополнительными ограничениями.

Определение 1 [7]. *Уравновешенной неполной блок-схемой (далее — просто блок-схемой) называется такое размещение n различных элементов по N блокам, что каждый блок содержит точно k различных элементов, каждый элемент появляется точно в r различных блоках и каждая пара различных элементов появляется точно в λ блоках.*

Определение 2 [7]. *Блок-схема называется симметричной, если $n = N$.*

Для симметричных блок-схем [7] $k = r$ и любые два различных блока имеют точно λ общих элементов.

Пусть кодовые слова \mathcal{L} образуют симметричную блок-схему. Тогда вероятность того, что кроме истинного КЗ появится случайное ложное кодовое слово, равна $p^{k-\lambda}$. Используя неравенство Маркова, получим оценку вероятности P появления хотя бы одного ложного КЗ:

$$P \leq (N - 1)p^{k-\lambda}.$$

Приравнявая P к ε , получим следующую оценку для k :

$$k \geq \frac{\ln(\varepsilon/(N - 1))}{\ln p} + \lambda. \quad (4)$$

Эта оценка справедлива для тех p , для которых $(N - 1)p^{k-\lambda} < 1$.

Дополнительным условием неравенства (4) служит ограничение

$$k \leq n = N. \quad (5)$$

Из вида матрицы \mathcal{L} следует [8], что она представима в виде суммы подстановочных матриц. Каждая такая подстановочная матрица однозначно определяет каждое кодовое слово, а k обеспечивает необходимый уровень ошибки.

Условие (5) позволяет получить дополнительное ограничение на вероятность p , при котором задача передачи с заданной ошибкой разрешима. Из (4) и (5) получаем

$$N \geq \frac{\ln(\varepsilon/(N - 1))}{\ln p} + \lambda.$$

Отсюда получаем еще одну оценку для p :

$$p \leq \left(\frac{\varepsilon}{N - 1} \right)^{1/N} e^{\lambda/N}.$$

В проведенных расчетах не учитывались дополнительные соотношения между параметрами симметричных блок-схем, при которых эти блок-схемы существуют. Например, если в симметричной блок-схеме n четно, то $k - \lambda$ есть точный квадрат [7]. В противном случае искомой блок-схемы не существует.

Это и другие ограничения делают задачу построения кода со скоростью передачи N достаточно сложной. К примеру, существует симметричная блок-схема с параметрами $N = n = 16$, $k = 6$, $\lambda = 2$. Для этого кода уровень неоднозначного декодирования $\varepsilon = 0,0001$ при $p = 0,1$ и $\varepsilon = 0,001$ при $p = 0,2$.

Подробнее с этими вопросами можно ознакомиться в книге [7].

Еще один класс ограничений накладывает требование «невидимости» скрытого канала. Число искусственных меток должно быть на уровне стандартного отклонения от числа случайных меток, т. е. k имеет порядок $\sqrt{np(1-p)}$.

Единственным способом защиты от рассмотренного класса скрытых каналов является нарушение согласования счетчиков на передающем и приемном концах скрытого канала. Такое рассогласование можно создать с помощью вставок неинформативных символов в легальное сообщение. Если изъятие этих вставок происходит перед декодированием скрытой передачи, то рассогласование длин может помешать правильному декодированию информации в скрытом канале.

Однако в указанном методе борьбы с рассмотренным скрытым каналом необходимо учитывать следующее. Если размеры вставок малы, то за счет соотношения между длинами $l_1^{(i)}, \dots, l_k^{(i)}$ и увеличения k можно нейтрализовать влияние защитных вставок. Если вставки делать достаточно длинными, то резко снижается пропускная способность легальной передачи, что противоречит функциональности легального канала.

5 Заключение

В работе получены оценки пропускной способности скрытых каналов, построенных с помощью меток. Полученные оценки показывают соизмеримость передаваемой по скрытому каналу информации с объемом информации, передаваемой легаль-

но. Полученные оценки являются оценками сверху, так как скрытые каналы допускают одновременную передачу нескольких кодовых слов в одном блоке.

Обсуждаются вопросы защиты от скрытых каналов рассматриваемого класса.

Литература

1. *Lampson B. W.* A note of the confinement problem // Commun. ACM, 1973. Vol. 16. No. 10. P. 613–615.
2. *Тимонина Е.* Скрытые каналы (обзор) // Jet Info, 2002. Т. 14. Вып. 114. С. 3–11.
3. *Грушо Н.* Скрытые каналы, основанные на метках // Системы и средства информатики, 2013. Т. 23. № 1. С. 7–13.
4. *Грушо А., Грушо Н., Тимонина Е.* Анализ меток в скрытых каналах // Информатика и её применения, 2014. Т. 8. Вып. 4. С. 12–16.
5. *Wang Z., Lee R. B.* Capacity estimation of non-synchronous covert channels // 2nd Workshop (International) on Security in Distributed Computing Systems (SDCS'05) Proceedings. — Columbus, OH, USA, 2005. P. 170–176.
6. *Чисар И., Кёрнер Я.* Теория информации: теоремы кодирования для дискретных схем без памяти / Пер. с англ. — М.: Мир, 1985. 400 с. (*Csiszar I., Korner J.* Information theory: Coding theorems for discrete memoryless systems. — Budapest: Akademiai Kiado, 1981. 472 p.).
7. *Холл М.* Комбинаторика / Пер. с англ. — М.: Мир, 1970. 424 с. (*Hall M., Jr.* Combinatorial theory. — Waltham, MA – Toronto – London: Blaisdell Publishing Co., 1967. 310 p.)
8. *Сачков В. Н., Тараканов В. Е.* Комбинаторика неотрицательных матриц. — М.: ТВП, 2000. 448 с.

Поступила в редакцию 25.07.15

RATE OF INFORMATION TRANSFER AND CAPACITY IN COVERT CHANNELS DEFINED BY TAGS

A. A. Grusho, N. A. Grusho, and E. E. Timonina

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

Abstract: The paper is devoted to estimations of capacity and rate of information transfer in covert channels of a special type. These covert channels are generated with the help of signals (tags) which do not bear semantic information, but are easily allocated on the reception end. Such tags are a kind of bans from the point of view of admissible values of all parameters connected with transfer of legal information. The covert information is coded by lengths of fragments of the data in legal transfer allocated by tags.

Keywords: covert channels; probability-theoretic models of covert channels; covert channels generated by tags; capacity of a covert channel; rate of information transfer

DOI: 10.14357/19922264150409

Acknowledgments

The paper was partly supported by the Russian Foundation for Basic Research (project 13-01-00215).

References

1. Lampon, B. W. 1973. A note of the confinement problem. *Commun. ACM* 16(10):613–615.
2. Timonina, E. 2002. Skrytye kanaly (obzor) [Covert channels (review)]. *Jet Info* 14(114):3–11.
3. Grusho, N. 2013. Skrytye kanaly, osnovannye na metkakh [The covert channels based on tags]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 23(1):7–13.
4. Grusho, A., N. Grusho, and E. Timonina. 2014. Analiz metok v skrytykh kanalakh [The analysis of tags in the covert channels]. *Informatika i ee Primeneniya — Inform. Appl.* 8(4):12–16.
5. Wang, Z., and R. B. Lee. 2005. Capacity estimation of non-synchronous covert channels. *2nd Workshop (International) on Security in Distributed Computing Systems (SDCS) Proceedings*. Columbus, OH. 170–176.
6. Csiszar, I., and J. Korner. 1981. *Information theory: Coding theorems for discrete memoryless systems*. Budapest: Akademiai Kiado. 472 p.
7. Hall, M., Jr. 1967. *Combinatorial theory*. Waltham, MA – Toronto – London: Blaisdell Publishing Co. 310 p.
8. Sachkov, V. N., and V. E. Tarakanov. 2000. *Kombinatorika neotritsatel'nykh matrits* [Combination theory of nonnegative matrixes]. Moscow: TVP. 448 p.

Received July 25, 2015

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, 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 119333, Russian Federation; grusho@yandex.ru

Grusho Nick 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 119333, Russian Federation; info@itake.ru

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 119333, Russian Federation; eltimon@yandex.ru

КОНТРОЛЬ И УПРАВЛЕНИЕ ИНФОРМАЦИОННЫМИ ПОТОКАМИ В ОБЛАЧНОЙ СРЕДЕ*

А. А. Грушо¹, М. И. Забейло², А. А. Зацаринный³

Аннотация: Контроль и управление информационными потоками в облачной среде служат важным механизмом информационной безопасности. Основная проблема в реализации такого механизма — скорость коммутации пакетов в соответствии с заданными правилами. В статье рассмотрены методы контроля и управления информационными потоками в облачной среде, базирующиеся на программно-конфигурируемых сетях. Построены оценки скорости информационного обмена в рамках облачных вычислений. Обсуждается система процедур, позволяющая реконфигурировать коммутационные таблицы (КТ) в соответствии с текущими изменениями топологии сетевых взаимодействий в облачной среде.

Ключевые слова: облачные вычисления; программно-конфигурируемые сети; скорость управления и безопасность информационных потоков в облаке; управление масштабируемостью и динамическими изменениями в топологии сети

DOI: 10.14357/19922264150410

1 Введение

Сегодня cloud computing — один из наиболее востребованных и интенсивно развиваемых комплексов информационных технологий. В русскоязычной специальной литературе наиболее распространенным термином для понятия *cloud computing* стали *облачные вычисления*, хотя в этой формулировке в явном виде, к сожалению, не прослеживаются некоторые другие важные смысловые оттенки исходного понятия. Например, восприятие существительного *cloud* как *покров* (*cover, covering*) и как *множество* (*multiplicity, plurality, variety*) позволяет указать на одну из принципиально важных функциональных особенностей *облачных вычислений* — *масштабируемость* (так называемую *вычислительную эластичность*) таких решений.

Ответом на потребность в виртуализации трех базовых ресурсов — вычислений, хранения данных и сетевых взаимодействий — стал целый спектр новых подходов, например Software Defined Networks (SDN — программно-конфигурируемые сети) [1] и Network Function Virtualization (NFV — виртуализация сетевых функций) [2], воплотившиеся в эффективные прикладные информационно-технологические решения нового поколения. Они вошли в проекты Internet2 [3] и развивающуюся в настоящий момент уже как международная коопера-

ция американскую национальную программу GENI (Global Environment for Network Innovations) [4].

Критически важными характеристиками облачных решений оказались *скорость* и *защищенность* информационного обмена. При этом названные два показателя оказываются «противоположно направленными». Высокий уровень защищенности, как правило, существенным образом снижает скорость облачных взаимодействий, так как понижает скорость обмена данными. Наоборот, высокие скорости информационного обмена предъявляют весьма жесткие требования к архитектуре и функциональности соответствующих систем информационной безопасности.

Традиционно решения по безопасности в больших информационных системах основаны на разграничении доступа. Например, политика безопасности RBAC (Role Based Access Control — контроль доступа, основанный на ролях) [5] может быть построена с помощью изолированных доменов, отвечающих дереву привилегий в организации.

Вторым важнейшим механизмом обеспечения информационной безопасности может стать контроль и управление информационными потоками (см., например, [6]).

Отметим, что с помощью SDN можно контролировать информационные потоки и управлять ими. С помощью контроля информационных потоков

* Работа поддержана РФФИ (проект 15-29-07981).

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

² ВИНТИ РАН, m.zabzhailo@yandex.ru

³ Федеральный исследовательский центр «Информатика и управление» Российской академии наук, azatsarinny@ipiran.ru

можно создавать изолированные домены. Таким образом, SDN может обеспечить важнейшие функции защиты от несанкционированного доступа.

Еще одной важной функцией в системе безопасности информационной системы является мониторинг событий безопасности [6]. Одним из самых эффективных механизмов решения этой проблемы служит DPI (Deep Packet Inspection — глубокий анализ содержимого пакетов). Отметим, что DPI на уровне контроллера SDN решает также задачу защиты системы мониторинга.

Приведем некоторые технологические решения, позволяющие SDN решать перечисленные выше задачи обеспечения информационной безопасности облачных вычислений. За счет специального (проблемно-ориентированного) разделения уровня исполнения трафика данных и уровня управления этим трафиком на две области можно сформировать дополнительные по сравнению с использованием классических сетевых технологий возможности управления сетевыми взаимодействиями. Так, в частности, удается получить существенные преимущества за счет переноса ряда важных функций на SDN-контроллер, который (как программная система, обеспечивающая виртуализацию необходимых функций по управлению компьютерной сетью) может располагаться на отдельной (в том числе распределенной) вычислительной установке требуемой производительности. Как следствие, открываются возможности гибкого управления выполняемыми на контроллере вычислениями и безопасностью.

Интерес, в первую очередь, представляют возможности вынести ряд важных, однако второстепенных по отношению к собственно передаче пакетов и потоков данных вычислительных процедур на внешний по отношению к механизмам организации трафика контроллер. Это позволяет реализовать ряд ресурсоемких процедур управления трафиком вне механизма организации трафика (вне SDN-коммутатора), используя внешний по отношению к нему вычислитель необходимой производительности. Назовем его SDN-контроллером. Такой прием позволяет разделить решение ряда задач организации и управления облачным трафиком, «чувствительных» к объемам требуемых вычислений, на две части:

- (1) однократное исполнение требующей существенных объемов вычислений специальной части целостного комплекса процедур анализа и управления трафиком;
- (2) многократно выполняемый набор действий, которые достаточно малозатратны в части объемов реализующих их вычислений.

Первая из них может выполняться на внешнем SDN-контроллере, а вторая — в том числе и на SDN-коммутаторе.

В качестве иллюстративного примера такого разделения можно обратиться к задаче поиска заголовка входного пакета в ТК (подробно эта задача будет рассмотрена в следующих разделах). Исполнению здесь подлежит достаточно простая по постановке формальная процедура — проверить, содержится ли входной булев вектор фиксированной длины k в содержащей n строк заданной КТ булевых векторов той же длины k . Однако при проверке таких входящих приходится учитывать следующие существенные соображения. В реальных приложениях КТ оказываются очень большими. На текущий момент известны предлагаемые рынку промышленные коммутаторы, оперирующие с КТ суммарным размером в 150–200 тыс. строк. При этом необходима высокая скорость решения задачи.

При решении этой формальной задачи можно использовать ряд достаточно хорошо изученных вариантов ее разделения на две подзадачи, одна из которых будет решаться однократно, порождая приведение КТ к некоторому специальному виду и требуя при этом существенно больших (по сравнению с решением второй подзадачи) объемов вычислений. Вторая подзадача обеспечит многократную быструю проверку, использующую этот специальный вид КТ, собственно вложимости входного заголовка h в текущее состояние КТ.

Хорошо известными (см., в частности, классическую работу [7]) примерами подобных разделений являются:

- лексикографическое упорядочение исходной КТ, когда проверка вложимости h в КТ (как и внесение динамических изменений в текущее состояние КТ — о чем речь пойдет чуть позже) оказывается достаточно простой вычислительной процедурой;
- приведение КТ к виду бинарного дерева (диаграммы), когда порождение такого дерева оказывается в общем случае экспоненциально сложной вычислительной процедурой, а собственно проверка вложимости входного заголовка h в уже построенное дерево представляет собою процедуру, которая требует всего лишь линейно быстро растущего с ростом длины вектора h объема вычислений.

Однако при переходе от анализа заголовков пересылаемых пакетов данных также и к анализу их содержания (DPI, IDS/IPS (Intrusion Detection/Protection Systems — системы обнаружения/предотвращения вторжений)) рассматриваемая ситуация

существенно усложняется. Множество признаков, проверяемых на встречаемость в «теле» передаваемых пакетов, многократно возрастает. Например, актуальные базы сигнатур вирусов сегодня оперируют уже сотнями тысяч объектов. На «техническом» уровне эта проблема «размерности» объекта анализа отражается в том, что

- формализация ее в уже представленной выше «технике» операций с булевыми векторами приводит к необходимости работать с объектами очень большой длины, содержащими при этом существенно больше нулей, чем единиц;
- в подобных булевых векторах вся полезная информация уже представлена элементами неупорядоченных множеств.

Ранее [8] уже рассматривались некоторые варианты статического механизма контроля сетевого трафика с использованием возможностей SDN-подхода. В данной работе предлагается целенаправленное расширение представленного подхода на случай управления динамическими изменениями актуальной конфигурации SDN-сети, учитывающий помимо обработки заголовков пакетов данных также и потребности в организации быстрого анализа их содержания. При этом, как и прежде, основная роль в предлагаемой процедурной конструкции отводится эффективной организации (на функционирующем по предлагаемым специальным правилам SDN-контроллере) вспомогательных вычислений, которые и обеспечивают поддержку принятия решений по реконфигурированию в соответствующей сети обмена информацией.

2 Методы проверки вложимости векторов в коммутационной таблице

В промышленных приложениях «ширина» КТ (число ее столбцов) k , как правило, существенно меньше числа ее строк n . Однако следует иметь в виду, что $k \leq \log_2 n$. Заметим, что прямой поиск — перебор строк коммутационной таблицы — также может быть оптимизирован за счет использования лексикографических упорядочений строк исходной КТ или же, например, приведения исходной КТ к виду бинарного дерева. Рассмотрим эти соображения более подробно.

Пусть $L(\text{КТ})$ есть лексикографически упорядоченная по строкам версия анализируемой КТ. Обозначим число операций для лексикографического упорядочения КТ через S_L^1 . Несложно убедиться, что процедура приведения КТ к целевому виду

требует при порождении $L(\text{КТ})$ $S_L^1 \leq C_L^1 k n \log_2 n$ операций, где $C_L^1 = \text{const}$. Таким образом, имеет место следующая теорема.

Теорема 1. *Сложность приведения к виду $L(\text{КТ})$ таблицы коммутаций, имеющей n строк и k столбцов, ограничена сверху величиной $C_L^1 k n \log_2 n$.*

При проверке вложимости входного заголовка $h_i = \{a_{i_1}^1, a_{i_2}^2, \dots, a_{i_k}^k\}$ в $L(\text{КТ})$ в каждом из ее столбцов с номером $j + 1$, $j \in \{1, 2, \dots, k - 1\}$, придется искать соответствующую пару содержащих только 1 и только 0 «блоков» (множеств строк КТ), на которую «распадается» множество значений $a_{i_j}^j$ в предыдущем столбце номер j . Чтобы сократить объем перебора значений в каждом из столбцов K (по сравнению с прямым просмотром каждого столбца целиком), можно воспользоваться алгоритмом последовательного разбиения множеств элементов столбца, например пополам или же по правилу золотого сечения. Однако число таких блоков с ростом номера j будет расти как $\log_2 j$, т. е. не быстрее, чем $C_L^2 \log_2 n$ для некоторой константы C_L^2 .

Пусть S_L^2 — объем перебора при проверке вложимости заголовка h в $L(\text{КТ})$. Таким образом, имеет место следующая теорема.

Теорема 2. *Объем перебора при проверке вложимости заголовка h в $L(\text{КТ})$ таблицы коммутаций можно оценить как $S_L^2 \leq C_L^2 \log_2 n$.*

Определим бинарное дерево, соответствующее КТ. Корень дерева находится на уровне (слое) 0, и из каждой вершины в следующий слой могут идти не более двух ребер, помеченных 0 и 1. Каждая вершина, кроме корневой, имеет метку 0 или 1, соответствующую метке входящего ребра. Таким образом, каждый путь из корня в вершину слоя i взаимно однозначно соответствует двоичному вектору длины i (как по меткам ребер, так и по меткам вершин). Число слоев в бинарном дереве, представляющем КТ, равно k . Максимальное число вершин в слое k равно 2^k , а число всех вершин в полном бинарном дереве равно $2^{k+1} - 1$.

Выделим пути в полном бинарном дереве с k слоями, соответствующие векторам из КТ. Полученное дерево обозначим $D(\text{КТ})$. Проверка наличия вектора в КТ с использованием построенного дерева $D(\text{КТ})$ состоит в прохождении пути длины k , причем из каждой вершины переход осуществляется по ребру с меткой, соответствующей очередному знаку в проверяемом векторе. Если на очередном шаге необходимое ребро отсутствует, то это означает, что рассматриваемого вектора в КТ нет. Отметим, что по построению дерева $D(\text{КТ})$, если КТ не пуста, из любой достигнутой вершины

есть хотя бы одно выходящее ребро. Проверка с помощью дерева $D(KT)$ присутствия векторов в KT не превосходит k шагов.

Для построения $D(KT)$ на некотором макете полного двоичного дерева необходимо $2^{k+1} - 1$ элементов памяти и nk операций выделения фрагментов $D(KT)$ в этом макете.

Теорема 3. *Определение принадлежности вектора таблице коммутаций с помощью $D(KT)$ требует не более k операций прохождения по ребрам $D(KT)$, но построение $D(KT)$ требует nk операций выделения фрагментов $D(KT)$ в макете полного двоичного дерева с k слоями.*

Положим $k = \alpha \log_2 n$, где $\alpha \geq 1$. Тогда объем памяти для макета полного двоичного дерева равен $2n^\alpha - 1$. Сложность построения $D(KT)$ равна $\alpha n \log_2 n$, а проверка принадлежности к KT требует не больше чем $\alpha \log_2 n$ операций.

Метод бинарного дерева можно усовершенствовать так, что скорость проверки вложимости существенно сократится при небольших ухудшениях остальных параметров.

Наименьшим запретом в двоичном корневом дереве $D(KT)$ назовем путь (соответственно, вектор) из корня, заканчивающийся отсутствующим ребром. Запрет — это путь (вектор), соответствующий в $D(KT)$ отсутствующему пути в концевую вершину на уровне k . Для каждого запрета в полном двоичном дереве существует наименьший запрет. Вложимость вектора в KT можно проверять по дереву наименьших запретов $M(KT)$. Если начало проверяемого вектора совпадает с соответствующим наименьшим запретом в дереве $M(KT)$, то проверяемый вектор не лежит в KT .

На простейшей модели оценим влияние коротких и длинных наименьших запретов. Предположим, что \bar{x} — проверяемый вектор, а KT формируется случайно с помощью независимого равновероятного выбора векторов длины k . Обозначим через P_i вероятность того, что случайный вектор войдет в противоречие с \bar{x} до координаты с индексом i . Тогда $1 - P_i$ означает вероятность совпадения случайного вектора с вектором \bar{x} по всем i первым координатам, при этом $1 - P_i = 1/2^i$. Вероятность того, что различие векторов произойдет на координате i , равна $2^{-(i+1)}$. Число векторов, для которых различие происходит в координате с индексом i , распределено по биномиальному закону. Отсюда следует, что среднее число векторов из n независимых случайных векторов, в которых различие происходит на шаге $i + 1$, равно $n/2^{i+1}$. Это число убывает с ростом i , что служит доводом в пользу того, что основное количество наименьших запретов имеет маленькую

длину. Эти рассуждения показывают целесообразность использования дерева наименьших запретов для определения принадлежности вектора KT .

Скорость определения принадлежности вектора KT с помощью бинарных деревьев можно повысить с помощью распараллеливания поиска наименьших запретов. С этой целью рассмотрим пару полных корневых бинарных деревьев с k слоями, которые будем называть «левое» и «правое». Коммутационная таблица порождает в левом полном дереве поддерево $D_L(KT)$, как это было описано выше при построении $D(KT)$. Аналогично по KT на основе правого полного дерева строится дерево $D_R(KT)$. Разница между деревьями состоит в том, что пути в $D_R(KT)$ соответствуют векторам из KT , читаемым справа налево. Соответственно строятся два дерева наименьших запретов $M_L(KT)$ и $M_R(KT)$. Если вектор \bar{x} не входит в KT , то в $M_L(KT)$ и $M_R(KT)$ есть два пути, соответствующие вектору \bar{x} при его чтении слева направо в $M_L(KT)$ и справа налево в $M_R(KT)$. Вектору \bar{x} соответствуют наименьшие запреты \bar{x}_L и \bar{x}_R . Поиск наименьшего запрета для проверяемого вектора \bar{x} параллельно по $M_L(KT)$ и $M_R(KT)$ может ускорить процесс отбраковки.

При наличии других признаков отбраковки (требования политики безопасности) можно также рассматривать параллельно несколько деревьев наименьших запретов, позволяющих ускорить процесс разрешения информационного потока.

3 Процедура реконfigurирования коммутационной таблицы

Пусть на момент времени T KT имела конфигурацию $H(T)$. Будем считать, что через промежуток времени ΔT_1 в таблице произошли ее первые структурные изменения:

- удалены строки (заголовки), образующие множество $H^-(T + \Delta T_1)$;
- добавлены строки, образующие множество $H^+(T + \Delta T_1)$.

При этом с формальной точки зрения представляется естественным считать, что хотя бы одно из множеств $H^+(T + \Delta T_1)$ и $H^-(T + \Delta T_1)$ не является пустым. Таким образом,

$$H(T + \Delta T_1) = (H(T) \setminus H^-(T + \Delta T_1)) \cup H^+(T + \Delta T_1).$$

Исключение векторов или включение новых в лексикографически упорядоченную KT по сложности равносильно поиску векторов в такой таблице. Это замечание определяет сложность вставки

или удаления новых векторов из лексикографически упорядоченной КТ.

Удаление пути из дерева равносильно удалению ближайшего к корню ребра и всех его продолжений. Добавление пути в дерево равносильно добавлению новых ребер. Отсюда сложность одного изменения не превосходит k .

4 Контроль и управление информационными потоками рассмотренными методами

Несложно убедиться, что мониторинг контента и управление потоками данных в облачной среде могут быть основаны не только на анализе структуры заголовков (булевых векторов фиксированной длины). В частности, можно учитывать также те или иные характеристики содержания передаваемых пакетов. В КТ обсуждалась обработка заголовков, где необходимо было оперировать векторами фиксированной и при этом достаточно малой по сравнению с размерами пакета данных длины. В случае учета контентных характеристик пакетов возникает необходимость обрабатывать подмножества, вообще говоря, достаточно большого множества признаков, например множества контентных характеристик пакетов, выделяемых DPI- или же IDS/IPS-средствами. Таким образом, появляется возможность обрабатывать в едином процессе множество параметров информационной безопасности.

Все эти признаки можно собрать в некоторый шаблон описания, где наличие конкретного признака у конкретного пакета данных будет кодироваться единицей, а его отсутствие — нулем. Однако размер такого шаблона (число позиций в соответствующем этому шаблону булевом векторе) может оказаться большим.

Представленные возможности могут оказаться весьма полезными при оптимизации функционирования подсистемы обеспечения информационной безопасности в облачном ландшафте, отвечающей, например, за мониторинг исполнения соответствующей политики безопасности компьютерной сети, в том числе за контроль:

- (а) «непересекаемости» заданных подсетей облака (т.е. физического отделения используемой каждой из двух соответствующих подсетей системно-технических ресурсов, если это дано политикой безопасности, при имеющих место динамических изменениях топологии виртуальной инфраструктуры облака);
- (б) доступности (в том числе соответствия заданной политике безопасности) конкретных сете-

вых взаимодействий типа (узел X — узел Y) или (множество узлов Q — множество узлов R);

- (в) доступности пакетов заданного типа лишь в наперед заданном перечне узлов.

При динамических изменениях текущей топологии виртуальной архитектуры облачной сети недостаточно только отслеживать (и пресекать) возможные нарушения заранее заданных требований политики безопасности. К примеру, функция (в) может поддерживаться средствами DPI- или же IDS/IPS-решений, которые удобно запускать именно на SDN-контроллере. Не менее важно выполнять быстрее, т.е. не порождающее критичных задержек трафика в сети, реконfigurирование КТ к виду, который обеспечивает решение текущих задач в данном облачном ландшафте и при этом удовлетворяет требованиям заранее сформулированной политики безопасности.

5 Заключение

Разумеется, для эффективного использования в тех или иных реальных приложениях предложенная процедурная схема управления реконfigurированием КТ в каждом конкретном случае требует аккуратной настройки, которая учитывала бы доступные в имеющемся системно-техническом ландшафте соответствующей облачной среды функциональные возможности.

Для определения значений таких «калибровочных» параметров требуется проведение специальных тестовых экспериментов на соответствующем компьютерном оборудовании. Развитие представленных здесь работ именно в этом направлении экспериментальных исследований будет следующей целью в ближайшем будущем.

Еще одним не менее интересным направлением перспективного развития обсуждаемого здесь подхода является изучение возможностей распараллеливания вычислений при реализации предложенных процедур реконfigurирования таблиц коммутации.

Наконец, определенные перспективы имеет и углубленное изучение тех дополнительных возможностей, которые предлагаемый подход к управлению вычислительными ресурсами облачной среды предоставляет для аутентификации «статуса» и «полномочий» передаваемых пакетов, а также контроля корректности их адресации как важных функций поддержания устойчивости и безопасности информационного обмена в облачной среде вычислений при динамическом реконfigurировании ее ресурсов.

Литература

1. Сетевые технологии SDN — Software Defined Networking. <http://habrahabr.ru/company/muk/blog/251959>.
2. ETSI. Network functions virtualization. <http://www.etsi.org/technologies-clusters/technologies/nfv>.
3. The Internet2 community: Enabling the future. <http://www.internet2.edu>.
4. GENI: Exploring networks of the future. <https://www.geni.net>.
5. Ferraiolo D. F., Kuhn D. R. Role-based access controls // 15th National Computer Security Conference. Baltimore, 1992. P. 554–563. <http://csrc.nist.gov/rbac/rbacSTD-ACM.pdf>.
6. ГОСТ Р ИСО/МЭК 15408-1-2008. Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Часть 1. Введение и общая модель. <http://docs.cntd.ru/document/gost-r-iso-mek-15408-1-2008>.
7. Кнут Д. Искусство программирования. Т. 3. Сортировка и поиск / Пер. с англ. — М: Вильямс, 2007. 832 с. (Knuth D. E. 1998. The art of computer programming. Vol. 3: Sorting and searching. — 2nd ed. — Reading, MA, USA: Addison-Wesley, 1998. 794 p.)
8. Зabeжaйлo M. И. К задаче анализа вложимости подслов в заголовки пакетов данных // Системы и средства информатики, 2013. Т. 23. № 1. С. 58–68.

Поступила в редакцию 23.10.15

INFORMATION FLOW MONITORING AND CONTROL IN THE CLOUD COMPUTING ENVIRONMENT

A. A. Grusho¹, M. I. Zabezhailo², and A. A. Zatsarinny³

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

²All-Russian Institute for Scientific and Technical Information of Russian Academy of Sciences, 20 Usievicha Str., Moscow 125190, Russian Federation

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

Abstract: Control of information flows is one of the most important mechanisms of security implementation in cloud computing. Some SDN (Software Defined Networks) based technologies for packet forwarding in the cloud computing environment are presented and analyzed. The special problem-oriented procedure for monitoring and quick control of forwarding table reconfiguration is discussed. This procedure is responsible for control of requirements of network security policy in the dynamically changing cloud computing environment. The presented control procedures are based on Lexical Ordering (LO) of forwarding table (FT) and FT transformation in Binary Tree (BT). Some estimations of computational complexity of the presented header analysis by BT and LO are discussed. Possibilities to extend the presented algorithmic approach from header analysis to Deep Packet Inspection (DPI) and some related problems are announced and evaluated.

Keywords: cloud computing; software defined networks; scalability of network resources; dynamic changes in network topology; information security in cloud computing environment; header analysis

DOI: 10.14357/19922264150410

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 15-29-07981).

References

1. Software Defined Networking. Available at: <http://habrahabr.ru/company/muk/blog/251959/> (accessed December 11, 2015).
2. ETSI. Network functions virtualization. Available at: <http://www.etsi.org/technologies-clusters/technologies/nfv> (accessed December 11, 2015).
3. The Internet2 community: Enabling the future. Available at: <http://www.internet2.edu/> (accessed December 11, 2015).
4. GENI: Exploring networks of the future. Available at: <https://www.geni.net> (accessed December 11, 2015).
5. Ferraiolo, D.F., and D.R. Kuhn. 1992. Role-based access controls. *15th National Computer Security Conference*. Baltimore. 554–563. Available at: <http://csrc.nist.gov/rbac/rbacSTD-ACM.pdf>.

- nist.gov/rbac/rbacSTD-ACM.pdf (accessed December 11, 2015).
6. GOST R ISO/IEC 15408-1-2008. Informatsionnaya tekhnologiya. Metody i sredstva obespecheniya informatsionnoy bezopasnosti. Kriterii otsenki bezopasnosti informatsionnykh tekhnologiy. Chast' 1. Vvedenie i obshchaya model' [Information technology. Security techniques. Evaluation criteria for IT security. Part 1. Introduction and general model]. Available at: <http://docs.cntd.ru/document/gost-r-iso-mek-15408-1-2008> (accessed December 11, 2015).
 7. Knuth, D. E. 1998. *The art of computer programming. Vol. 3: Sorting and searching*. 2nd ed. Reading, MA: Addison-Wesley. 794 p.
 8. Zabezhailo, M. I. 2013. K zadache analiza vlozhimosti podslov v zagolovki paketov dannykh [On the problem of subsequences inclusion into the data packages headers]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 23(1):58–68.

Received October 23, 2015

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, 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 119333, Russian Federation; grusho@yandex.ru

Zabezhailo Michael I. (b. 1956) — Candidate of Science (PhD) in physics and mathematics, senior scientist, All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences, 20 Usievicha Str., Moscow 125190, Russian Federation; m.zabezhailo@yandex.ru

Zatsarinny Alexander A. (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; azatsarinny@ipiran.ru

МЕЛКОЗЕРНИСТЫЕ ГИБРИДНЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ. ЧАСТЬ 1: ЛИНГВИСТИЧЕСКИЙ ПОДХОД

И. А. Кириков¹, А. В. Колесников², С. В. Листопад³, С. Б. Румовская⁴

Аннотация: Рассматривается проблематика междисциплинарных инструментариев и свойства «зернистости» гибридов в информатике. Результаты исследований показаны в рамках лингвистического подхода, суть которого состоит в трансформации вербализованной информации об объектах-оригиналах (сложных задачах) и объектах-прототипах (методах моделирования), имеющейся в полях профессиональной деятельности, в объекты-результаты (функциональные гибридные интеллектуальные системы (ФГИС)). Трансформация направляется эвристиками — схемами ролевых концептуальных моделей (КМ) в неформальной аксиоматической теории. Категориальное ядро теории — «ресурс—свойство—действие—отношение». Над его расширением специфицированы одно-, двух- и трехролевые конструкторы — основные элементы, из которых по правилам склеивания построены схемы отображения информации о ресурсах, действиях, ситуациях, состояниях объекта управления, сложных задачах субъекта управления и мелкозернистых гибридах субъекта моделирования.

Ключевые слова: логико-математический интеллект; гибридные интеллектуальные системы; лингвистический подход; теория ролевых концептуальных моделей

DOI: 10.14357/19922264150411

1 Введение

Идея построения многомодельной семиотической системы с логико-лингвистической, имитационной и аналитической компонентой впервые высказана Я. А. Гельфандбейном, А. В. Колесниковым и И. Д. Рудинским (Рига, Калининград, СССР) в 1981 г. в проекте «Ситуация» Научного совета по комплексной проблеме «Кибернетика» при Президиуме АН СССР. Четыре года понадобилось на реализацию и апробацию сменного-суточного планирования. Спустя более чем десять лет, в 1994–1995 гг. в работах L. Medsker (Вашингтон, США) были анонсированы гибридные интеллектуальные системы (ГИС), по существу совпадающие с интеллектуальными гибридными системами [1], гибридными интегрированными системами [2, 3], гибридными информационными системами [4] и гибридными интеллектуальными адаптивными системами [5].

Параллельно развивалась и научно-практическая «платформа» методов ГИС — компьютерные системы поддержки принятия решений (СППР) [2]. На рубеже XX и XXI вв. знания

в этой междисциплинарной области изучены, обобщены и опубликованы в работах А. Н. Борисова, А. В. Гаврилова [6], А. В. Колесникова [7], Д. А. Поспелова, Г. В. Рыбиной [3], В. Б. Тарасова, Н. Г. Ярушкиной [8], S. Goonatilake и S. Khebbal [1], L. Medsker [4]. В России сложилось несколько школ в этой области междисциплинарных знаний: Д. А. Поспелова — В. Б. Тарасова; В. Н. Вагина — А. П. Еремеева; Г. В. Рыбиной в Москве; Н. Г. Ярушкиной в Ульяновске; В. Ф. Пономарева — А. В. Колесникова в Калининграде.

Обобщение многолетнего опыта автоматизированного решения сложных практических задач методами междисциплинарных инструментариев показывает их сегодняшнюю проблематику:

- (1) наличие противоречия между свойствами объективной реальности, процессами, явлениями и событиями окружающего мира и научной картиной мира как части мировоззрения ученых и специалистов, работающих в информатике, управлении и проектировании;
- (2) отображение только логико-математического интеллекта, языковой коммуникации, ле-

¹ Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, baltbipiran@mail.ru

² Балтийский федеральный университет им. И. Канта; Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, avkolesnikov@yandex.ru

³ Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, ser-list-post@yandex.ru

⁴ Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, sophiyabr@gmail.com

восторонней составляющей рассуждений экспертов и лица, принимающего коллективные решения в условиях неоднородности и неопределенности информации, в то время как специалисты из разных предметных областей убеждены, что именно сочетание естественно-го и визуального языков в логико-математических и визуально-пространственных, правосторонних рассуждениях, соответственно, сочетание языковой и визуальной коммуникации над проблемами релевантно феномену человеческого мышления;

- (3) несмотря на более чем 15-летние исследования, мало изучены свойства «зернистости» гибридов и, как следствие, перспективные цели гибридизации, в частности мелкозернистые ФГиИС (МФГиИС);
- (4) не исследовано свойство устойчивости ГиИС управления, в частности применением методов гомеостатики к взаимодействующим, балансирующим и компенсирующим недостатки друг друга элементам-моделям экспертов и лиц, принимающих решения (ЛПР), управлению их внутренним противоречием, что приведет к созданию более совершенных с точки зрения живучести, помехоустойчивости структур систем управления;
- (5) информатика не завершила выработку принципов, переводящих создание гибридов из уникальной ремесленной мастерской в проектно-конструкторскую деятельность;
- (6) малочисленность конкурирующих методологий гибридизации в информатике;
- (7) неразвитость, а для МФГиИС и отсутствие технологий гибридизации и инструментальных средств поддержки;
- (8) чрезвычайно узкий охват автоматизацией только задач пассивных фаз управления [7]: учета, контроля, режиссуры, анализа, нормирования, прогнозирования, — и непонимание актуальности автоматизированного решения сложных задач организации, планирования и регулирования, а отсюда — слабая востребованность гибридизации практикой компьютерных СППР; положение усугубляется терминологической путаницей, компьютерным жаргоном, несоблюдением нормативов стандартизации.

Рассмотренная проблематика частично затрагивалась в работах [1, 4, 6–9], четвертая проблема — известная, ее решение открывает путь к новому классу информационных систем — гомеостатическим ГиИС.

Настоящая работа состоит из двух частей. В первой части, представленной в данной статье, в рамках лингвистического подхода для молодых ученых, специалистов по информатике и научно-практического сообщества, а также практиков СППР исследуется «зернистость» гибридов. Во второй части [10] формулируется в теории схем ролевых КМ класс мелкозернистых ГиИС, представлены результаты исследования свойств функциональной, инструментальной неоднородности сложных задач и введено понятие двунаправленной гибридизации.

2 Лингвистический подход к гибридизации

Разработку ГиИС называют гибридизацией [4, 7]. Это трудоемкий процесс, требующий широкого спектра знаний о предметной области, задачах, методах их решения, длительной по времени, сложной обработки информации и экспериментов. Двадцатилетний опыт авторов статьи показывает, что на разработку «вручную» ФГиИС, решающей сложную задачу, затрачивается 3–5 чел.-лет.

Выполняя гибридизацию, разработчик имеет дело с тремя сущностями:

- (1) сложной задачей (объектом-оригиналом);
- (2) несколькими методами решения задач (объектами-прототипами);
- (3) моделью ГиИС (объектом-результатом).

Первые два объекта — источники информации для гибридизации, а последний — основа направленного формообразования, изначально программирующая то, что должно появиться в конце, после трансформации информации.

Взаимодействие субъектов в процессе решения сложных задач может осуществляться только с использованием механизма обеспечения такого взаимодействия — языка. Вместе с системами формирования и манипулирования образами, запахами и т.п. язык формирует изученный к настоящему времени в науке инструментарий повседневной и профессиональной деятельности человека. Будучи одновременно средством мысленного эксперимента и моделью внешнего мира, язык профессиональной деятельности (ЯПД) — объект анализа и получения информации для гибридизации при разработке ФГиИС. В этой связи излагаемый ниже подход к разработке ГиИС будем называть *лингвистическим*.

Профессиональные языки управления и моделирования — это полиязыки, сформировавшиеся в результате совместной деятельности различных профессиональных групп. Трансформация

информации об объектах-оригиналах с полиязыка профессиональной коллективной деятельности по принятию решений и профессионального языка разработчика об объектах-прототипах в объекты-результаты на сегодняшний день не изучается математикой, и ожидать в обозримом будущем гибридизации по принципу трансляции текстов оригиналов и прототипов во внутримашинное представление с получением исполняемой программы-приложения не приходится. Поэтому современная методология предусматривает последовательное, а зачастую итерационное, моделирование на все более и более искусственных языках, свободных от недостатков ЯПД.

Один из таких искусственных языков построения моделей — язык КМ. Концептуальная модель (англ. *concept* — понятие, идея; общее представление; концепция) — модель предметной области, состоящая из перечня всех понятий, используемых для описания этой области, вместе со свойствами и характеристиками, классификацией этих понятий по типам, ситуациям, признакам в данной области и законов протекания процессов в ней [7]. Такой смысл близок к термину «концепция» (от лат. *conceptio* — понимание, система) как определенный способ трактовки какого-либо предмета, процесса, явления.

Таким образом, КМ имеет двоякую сущность. С одной стороны, она ограничивает предметную область как совокупность реальных или абстрактных объектов (сущностей), связей и отношений между объектами, а также процедур преобразования этих объектов при решении задач. С другой стороны, она предполагает внесение в гибридизацию субъективных представлений разработчика в виде его знаний и опыта — концепций. Это эвристики, специфицирующие понятия ресурса, свойства, действия, структуры, состояния, поведения объекта управления, цели, оценки, критерия, плана, задачи субъекта управления, а также метода, модели, алгоритма, программы субъекта моделирования и раскрывающие содержание, логическую связанность отдельных видов деятельности при гибридизации. Отсутствие таких эвристик у разработчика делает гибридизацию нецеленаправленной, вносит в эту сложную деятельность стохастичность. Извлечение, представление, хранение, тиражирование и применение таких эвристик чрезвычайно важны при гибридизации. Естественно-языковые концепции возбуждают и правополушарное мышление, хранящее концепции-образы. Следуя Д. А. Поспелову [11], предложен и развит класс ролевых КМ [7], а следуя А. И. Умову [12], триада «вещь—свойство—отношение» положена в основу структурирования ЯПД экспертов и ЛПР, ЯПД разработчи-

ка, а также неформальной аксиоматической теории концептуального моделирования функциональных ГиИС [13].

Концептуализация основывается на описании предметной области в виде неформальной (содержательной) аксиоматической теории. Большинство аксиоматических теорий — неформальные в том смысле, что они предполагают использование результатов общей теории множеств и следуют правилам вывода в логике.

Неформальная аксиоматическая теория Te схем ролевых КМ включает в себя четыре списка [14]:

$$Te = \langle Te^{TH}, Te^{TO}, Te^A, Te^{BT} \rangle, \quad (1)$$

где Te^{TH} , Te^{TO} — неопределяемые и определяемые термины соответственно; Te^A — аксиомы; Te^{BT} — высказывания-теоремы, выводимые из Te^A по некоторым фиксированным логическим правилам.

Упорядоченная пара $\langle Te^{TH}, Te^A \rangle$ — формулировка теории, которая совместно с Te^{BT} определяет ее предметную область. В концептуальном моделировании понятийная структура предметной области описывается неформальными аксиоматическими теориями, формулировки которых называются концептуальными моделями.

Категориальное ядро теории (Te^{TH} , Te^{TO}). Один из результатов в теории систем и системном анализе получен А. И. Умовым [12], предожившим в качестве базиса системного подхода к анализу и синтезу сложных объектов триаду категорий «вещь—свойство—отношение» (рис. 1, а).

Главная особенность триады «вещь—свойство—отношение» — соотносительность, все эти категории специфицируются друг через друга, причем центральная, основная категория — категория вещи [12], и в категориальном ядре не действует правило о запрещении круга в определениях. Непосредственное применение ядра «вещь—свойство—отношение» к системному анализу предметной области затруднительно в силу нескольких причин:

- (1) категория вещи — верхний предел абстракции, что вызывает трудности с ее интерпретацией участниками коллективного принятия решений и разработчиком;
- (2) трудности с интерпретацией понятия «зависимость» между категориями ядра; поскольку зависимость между ними — это отношение, то возникает неоднозначность в толковании отношений;
- (3) категория «отношение» носит общий характер, и целесообразно выбрать категорию, которая, являясь отношением, была бы в то же время предметно-ориентированной.

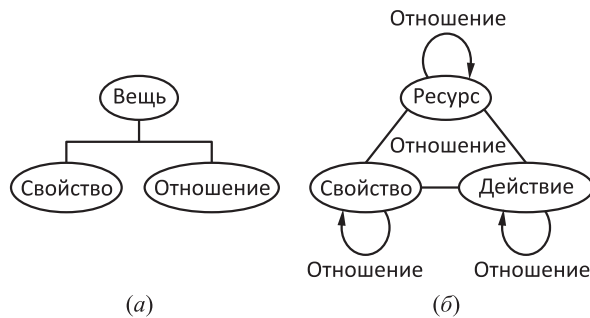


Рис. 1 Категориальный базис аксиоматической теории ролевых КМ: (а) триада «вещь–свойство–отношение»; (б) тетрада «ресурс–свойство–действие–отношение»

Трудности преодолены следующим образом. Множество вещей ограничено вещами-ресурсами (ресурсами), поскольку прагматичны вещи из арсенала ЛПР для решения задач (природные явления — также ресурс). Множество отношений ограничим отношениями-действиями (действиями), подчеркивая интерес к отношениям на ресурсах, установление или разрыв которых может изменить состояние объекта управления, т. е. операциям и процессам.

Концептуальная модель на рис. 1, б дает простое и прозрачное структурирование внешнего мира, утверждая, что это есть мир ресурсов, действий их свойств и девяти классов отношений «ресурс–ресурс», «действие–действие», «свойство–свойство», «ресурс–свойство» и «свойство–ресурс», «ресурс–действие» и «действие–ресурс», а также «действие–свойство» и «свойство–действие», которые ниже обозначены как r , а их множества — R . Особенности этой модели — ролевой характер и инвариантность к специфике предметной области.

Введем множество $X = \{RES, PR, ACT\}$, где $RES = X^{res}$, $PR = X^{pr}$, $ACT = X^{act}$ — множества базисных понятий, обозначающих ресурсы, свойства и действия соответственно, и построим на X полный граф $G = \langle X, R \rangle$. Тогда имеем классификацию базовых отношений $R = \{R^{res\ res}, R^{pr\ pr}, R^{act\ act}, R^{res\ pr}, R^{pr\ res}, R^{res\ act}, R^{act\ res}, R^{act\ pr}, R^{pr\ act}\}$, где $R^{res\ res}, R^{pr\ pr}, R^{act\ act}, R^{res\ pr}, R^{pr\ res}, R^{res\ act}, R^{act\ res}, R^{act\ pr}, R^{pr\ act}$ — множества отношений «ресурс–ресурс», «свойство–свойство», «действие–действие», «ресурс–свойство», «свойство–ресурс», «ресурс–действие», «действие–ресурс», «действие–свойство», «свойство–действие» соответственно.

Определив граф G , имеем $Te^{TH} = \{\text{«вещь»}, \text{«свойство»}, \text{«отношение»}\}$, $Te^{TO} = \{\text{«вещь»}, \text{«свойство»}, \text{«отношение»}, \text{«ресурс»}, \text{«действие»}\}$ и можно ввести аксиоматику теории (1).

Аксиоматика ролевых конструктов (Te^A)

Аксиома 1. Если $X^\alpha \subseteq X$, а $R^{\alpha\alpha} \subseteq R$, то существует одноролевой конструкт $co^1 = R^{\alpha\alpha}(X^\alpha, X^\alpha)$, где $\alpha \in \{res, pr, act\}$, имеющий в качестве ролей одну и ту же категорию ядра. Одноролевой конструкт — это точечная диаграмма. Например, в ЯПД диспетчера морского порта есть одноролевые конструкты: **склад рядом причал, камера часть холодильника**. Семантическая сила одноролевых конструктов невелика. Они выражают смысл на абстрактном уровне.

Аксиома 2. Если $X^\alpha, X^\beta \subseteq X$, а $R^{\alpha\alpha}, R^{\beta\beta}, R^{\alpha\beta} \subseteq R$, то существует двухролевой конструкт $co^2 = R^{\alpha\alpha}(X^\alpha, X^\alpha) \circ R^{\beta\beta}(X^\beta, X^\beta) \circ R^{\alpha\beta}(X^\alpha, X^\beta)$, где $\alpha, \beta \in \{res, pr, act\}$; $\alpha \neq \beta$; \circ — конкатенация. Двухролевой конструкт включает два одноролевых конструкта, связанных отношениями $R^{\alpha\beta}$. Двухролевых конструктов шесть, по числу ребер на графе G . Переход к двухролевой модели — это переход от точечной к линейной диаграмме. В ЯПД диспетчера морского порта есть двухролевые конструкты: **судно Агат стоит у причала № 16, ускоренная выгрузка используется на судне Топаз**. Линейная диаграмма, какой бы длинной она ни была, не обеспечивает целостную картину внешнего мира.

Аксиома 3. Если $X^\alpha, X^\beta, X^\gamma \subseteq X$, а $R^{\alpha\alpha}, R^{\beta\beta}, R^{\gamma\gamma}, R^{\alpha\gamma}, R^{\beta\alpha}, R^{\gamma\beta} \subseteq R$, то существует трехролевой конструкт co^3 , которым назовем выражение $co^3 = R^{\alpha\alpha}(X^\alpha, X^\alpha) \circ R^{\beta\beta}(X^\beta, X^\beta) \circ R^{\gamma\gamma}(X^\gamma, X^\gamma) \circ R^{\alpha\gamma}(X^\alpha, X^\gamma) \circ R^{\beta\alpha}(X^\beta, X^\alpha) \circ R^{\gamma\beta}(X^\gamma, X^\beta)$, где $\alpha, \beta, \gamma \in \{res, pr, act\}$; $\alpha \neq \beta \neq \gamma$. Трехролевых конструктов девять. Переход от двухролевой модели к трехролевой — это переход от линейной разомкнутой диаграммы к фигурной (треугольной) замкнутой диаграмме, концептуально полной для триады «ресурс–свойство–действие» картине мира. Например, в ЯПД диспетчера есть трехролевой конструкт: **судно стоит у причала № 15, и на нем выполняется ускоренная выгрузка мороженой рыбы**.

В итоге имеем $Te^A = \{A_1, A_2, A_3\}$ и три множества конструктов $CO^1 = \{co_1^1, \dots, co_3^1\}$, $CO^2 = \{co_1^2, \dots, co_6^2\}$, $CO^3 = \{co_1^3, \dots, co_9^3\}$. На категориальном ядре, т. е. множестве X , невозможно задать никаких других конструктов.

Расширение категориального ядра (Te^{TH}, Te^{TO}).

Для моделирования предметных знаний об объекте управления категориальное ядро расширим $PAR = X^{par} \subseteq X^{pr}$ — физическими свойствами (параметрами), $CH = X^{ch} \subseteq X^{pr}$ — характеристическими свойствами (характеристиками), $NAM = X^{nam} \subseteq X^{pr}$ — именными свойствами (именами), $MES = X^{mes}$ — единицами измерения (мерами), $VAL = X^{val}$ — значениями и $ST = X^{st}$ —

состояниями. Для моделирования предметных знаний о субъекте управления в X включены $EST = X^{est}$ — оценки и $PRB = X^{prb} = \{prb - \text{«задача»}\}$, а о субъекте моделирования — $MET = X^{met}$ — «метод», mod — «модель», prg — «программа». Чтобы иметь возможность расширять категориальное ядро, введем и $EX = X^{ex}$ — экзотические понятия.

Правило склеивания ролевых конструктов. Для построения множества высказываний-теорем Te^{BT} теории (1) сформулируем на X правило склеивания (во всех правилах $\alpha, \beta \in \{res, pr, act, mes, val, st, est, prb, met, mod, prg, par, ch, nam\}$) следующим образом:

- (1) трехролевой конструкт $co^3 \in CO^3$ есть схема $sch \in SCH$ ролевых КМ;
- (2) если $sch = A_1 \circ R^{\alpha\alpha}(X^\alpha, X^\alpha) \circ B_1$ — схема ролевых КМ, где A_1 и B_1 — любые фрагменты (в том числе и пустые) модели, не содержащие $R^{\alpha\alpha}(X^\alpha, X^\alpha)$, исключая случай $A_1 = B_1 = \emptyset$, и $co = A_2 \circ R^{\alpha\alpha}(X^\alpha, X^\alpha) \circ B_2$ — двух- или трехролевой конструкт, где A_2 и B_2 — любые фрагменты (в том числе и пустые), не содержащие $R^{\alpha\alpha}(X^\alpha, X^\alpha)$, исключая случай $A_2 = B_2 = \emptyset$, то $sch = A_1 \circ R^{\alpha\alpha}(X^\alpha, X^\alpha) \circ B_1 \circ A_2 \circ B_2$ — также схема ролевых КМ (склеивание треугольников вершиной);
- (3) если $sch_1 = A_1 \circ R^{\alpha\beta}(X^\alpha, X^\beta) \circ B_1$ и $sch_2 = A_2 \circ R^{\beta\alpha}(X^\beta, X^\alpha) \circ B_2$ — схемы ролевых КМ, причем $A_1 = A_2$ и $B_1 = B_2$ — непустые фрагменты, то $sch_3 = A_1 \circ R^{\alpha,\beta}(X^\alpha, X^\beta) \circ R^{\beta,\alpha}(X^\beta, X^\alpha) \circ B_1$ — также схема ролевых КМ (добавление обратного отношения);
- (4) если $sch_1 = A_1 \circ R^{\alpha\beta}(X^\alpha, X^\beta) \circ B_1$ и $sch_2 = A_2 \circ R^{\alpha\beta}(X^\alpha, X^\beta) \circ B_2$ — схемы ролевых КМ, причем A_1, B_1, A_2 и B_2 — непустые фрагменты, то $sch_3 = A_1 \circ A_2 \circ R^{\alpha\beta}(X^\alpha, X^\beta) \circ B_1 \circ B_2$ — также схема ролевых КМ (склеивание треугольников стороной);
- (5) в схеме ролевых КМ не может быть более одного несвязанного одноролевого конструкта;
- (6) никаких других схем ролевых КМ нет.

Таким образом, схемы ролевых КМ $sch \in SCH$ строятся из конструктов-аксиом, и есть высказывания-теоремы теории (1), полученные по правилу склеивания. Схемы — это эвристики, знания и опыт разработчика в виде концепций. Они, с одной стороны, повторяют структурирование ЯПД коллективного интеллекта, вырабатывающего и принимающего решения, специфицируют понятия ресурса, свойства, действия, структуры, состояния, поведения объекта управления, цели, оценки, критерия,

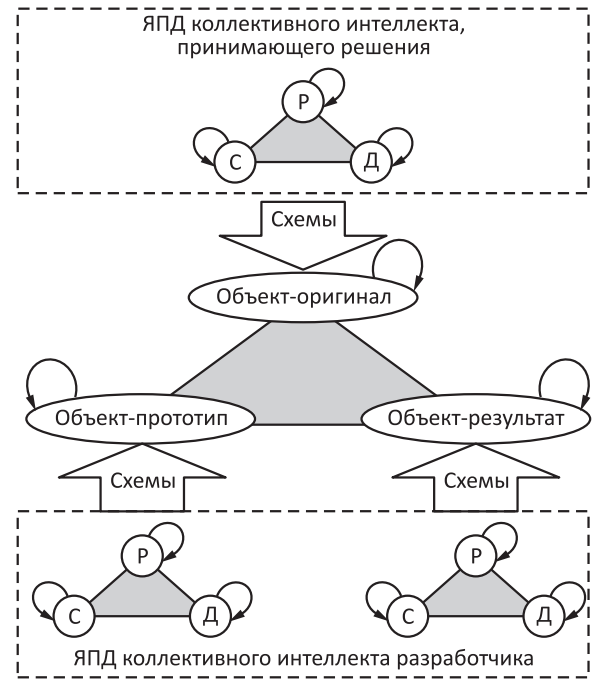


Рис. 2 Лингвистический подход к гибридизации

плана, задачи субъекта управления. С другой стороны, повторяют структурирование ЯПД коллективного интеллекта разработчика и специфицируют понятия метода, модели, алгоритма, программы и, с третьей стороны, повторяют объекты конструирования и целеполагания — модель-элемента, гетерогенное модельное поле, функциональную ГиИС, мелкозернистую ГиИС, поведение ГиИС, раскрывают содержание, логическую связанность деятельности при гибридизации.

Суть лингвистического подхода к гибридизации показана на рис. 2, а основные положения изложены ниже.

1. Базируется на компьютерном моделировании логико-математического интеллекта — одного из множества интеллектов человека по Гарднеру.
2. Данные, знания, опыт, связанные с решением сложных задач (объектами-оригиналами) в системе управления, должны быть вербализованы, извлекаемы из экспертов, ЛПР и других источников — текстов на ЯПД, баз данных. Поиск, извлечение знаний об объектах-оригиналах и конструирование знаков информационного языка выполняются по соответствующим схемам-эвристикам теории (1).
3. Данные, знания, опыт, связанные с методами моделирования (объектами-прототипами) и их

применением в научно-исследовательских работах, проектно-конструкторской и эксплуатационной деятельности, должны быть вербализованы, извлекаемы из экспертов и других источников — научной, учебно-методической литературы, проектно-конструкторской документации.

4. Данные, знания, опыт, связанные с ГиИС (объектами-результатами) и их применением для автоматизированного решения сложных задач, должны быть вербализованы, извлекаемы из экспертов и представлены методологическим учением (методологией). Объекты-результаты конструируются, а их поведение специфицируется как знаки информационного языка по соответствующим схемам-эвристикам теории (1).
5. Данные, знания, опыт, связанные с действиями по преобразованию вербальной информации об объектах-оригиналах (функциональных родительских признаках) и объектах-прототипах (инструментальных родительских признаках) в объекты-результаты (целевые комбинации родительских функциональных и инструментальных признаков), должны быть вербализованы, извлекаемы из экспертов и представлены методологическим учением (методологией) — логически увязанной и целенаправленной последовательностью действий.

3 Заключение

В рамках лингвистического подхода к гибридизации рассмотрено моделирование МФГиИС в неформальной аксиоматической теории схем ролевых КМ. Такие системы обладают значительным потенциалом имитации рассуждений специалистов, решающих сложные задачи. С одной стороны, модельная архитектура МФГиИС отображает на компьютере функциональную структуру сложной задачи, области ее однородных параметров, представленных множествами подзадач, что позволяет имитировать несколько линий рассуждений. С другой стороны, в таких системах для решения подзадач используются не только известные в теории принятия решений, исследовании операций, искусственном интеллекте, математической статистике методы моделирования (методы-родители),

но и конструируются из набора инструментальных средств (зерен) методы-потомки, лишенные родительских недостатков, что позволяет вести рассуждения эффективнее.

Литература

1. *Goonatilake S., Khebbal S.* Intelligent hybrid systems // 1st Singapore Conference (International) on Intelligent Systems Proceedings, 1992. P. 356–364.
2. *Поспелов Г. С.* Искусственный интеллект — основа новой информационной технологии. — М.: Наука, 1988. 280 с.
3. *Рыбина Г. В.* Интегрированные экспертные системы: современное состояние, проблемы и тенденции // Известия РАН. Теория и системы управления, 2002. № 5. С. 111–126.
4. *Medsker L. R.* Hybrid intelligent systems. — Kluwer Academic Publ., 1995. 295 p.
5. *Kasabov N., Kozma R.* Hybrid intelligent adaptive systems: A framework and a case study on speech recognition // Int. J. Intell. Syst., 1998. Vol. 13. Iss. 6. P. 455–466.
6. *Гаврилов А. В.* Гибридные интеллектуальные системы. — Новосибирск: НГТУ, 2003. 164 с.
7. *Колесников А. В.* Гибридные интеллектуальные системы. Теория и технология разработки / Под ред. А. М. Яшина. — СПб: СПбГТУ, 2001. 711 с.
8. *Ярушкина Н. Г.* Основы теории нечетких и гибридных систем. — М.: Финансы и статистика, 2004. 320 с.
9. *Колесников А. В., Кириков И. А.* Методология и технология решения сложных задач методами функциональных гибридных интеллектуальных систем. — М.: ИПИ РАН, 2007. 387 с.
10. *Кириков И. А., Колесников А. В., Листопад С. В., Румовская С. Б.* Мелкозернистые гибридные интеллектуальные системы. Часть 2. Двухнаправленная гибридизация // «Информатика и её применения», 2016. Вып. 1 (в печати).
11. *Поспелов Д. А.* Ситуационное управление: теория и практика. — М.: Наука, 1986. 288 с.
12. *Уемов А. И.* Вещи, свойства, отношения. — М.: Институт философии АН СССР, 1963. 184 с.
13. *Колесников А. В.* Моделирование естественных гетерогенных систем коллективного принятия решений // Системный анализ и информационные технологии (САИТ-2015): Тр. 6-й Междунар. конф. — М.: ИСА РАН, 2015. Т. 1. С. 7–16.
14. *Лелюк В. А.* Концептуальное проектирование систем с базами знаний. — Харьков: Основа, 1990. 144 с.

Поступила в редакцию 20.09.15

FINE-GRAINED HYBRID INTELLIGENT SYSTEMS. PART 1: LINGUISTIC APPROACH

I. A. Kirikov¹, A. V. Kolesnikov^{1,2}, S. V. Listopad¹, and S. B. Rumovskaya¹

¹Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation

²Immanuel Kant Baltic Federal University, 14 Nevskogo Str., Kaliningrad 236041, Russian Federation

Abstract: The paper considers the problematic of interdisciplinary tools and the property of “grain” for hybrids in informatics. The results are presented within the linguistic approach, the core of which is transformation of the verbalized information about objects—originals (complex subjects) and objects—prototypes (modeling approaches) to objects—results (functional hybrid intelligent system). It exists in polylanguages of professional activity. The transformation is directed by heuristics, which are the schemes of the conceptual role models in the informal axiomatic theory. The category core of the theory is “resource—property—operation—relation.” Mono-, bi-, and tri-role constructs are specified over its extension, which are the basic elements. On the basis of these elements the schemes of reflection of information about resources, operations, situations, state of object of management, complex tasks of management entity, and fine-grained hybrids of modeling entity are built.

Keywords: logical-mathematical intelligence; hybrid intelligent systems; linguistic approach; theory of role conceptual models

DOI: 10.14357/19922264150411

References

1. Goonatilake, S., and S. Khebbal. 1992. Intelligent hybrid systems. *1st Singapore Conference (International) on Intelligent Systems Proceedings*. 356–364.
2. Pospelov, G. S. 1988. *Iskusstvennyy intellekt — osnova novoy informatsionnoy tekhnologii* [Artificial intelligence — the base of the new information processing technology]. Moscow: Nauka. 280 p.
3. Rybina, G. V. 2002. Integrirovannyye ekspertnyye sistemy: Sovremennoe sostoyanie, problemy i tendentsii [Integrated expert systems: Actual status, problems and scientific trends]. *Izvestiya RAN. Teoriya i Sistemy Upravleniya* [J. Computer Systems Sciences International] 5:111–126.
4. Medsker, L. R. 1995. *Hybrid intelligent systems*. Kluwer Academic Publ. 295 p.
5. Kasabov, N., and R. Kozma. 1998. Hybrid intelligent adaptive systems: A framework and a case study on speech recognition. *Int. J. Intell. Syst.* 13(6):455–466.
6. Gavrilov, A. V. 2003. *Gibridnye intellektual'nye sistemy* [Hybrid intelligent systems]. Novosibirsk: NNSTU Pubs. 164 p.
7. Kolesnikov, A. V. 2001. *Gibridnye intellektual'nye sistemy. Teoriya i tekhnologiya razrabotki* [Hybrid artificial systems. Theory and development technology]. St. Petersburg: SPbGTU. 711 p.
8. Jarushkina, N. G. 2004. *Osnovy teorii nechetkikh i gibridnykh sistem* [Foundations of theory of fuzzy and hybrid systems]. Moscow: Finance and Statistics. 320 p.
9. Kolesnikov, A. V., and I. A. Kirikov. 2007. *Metodologiya i tekhnologiya resheniya slozhnykh zadach metodami funktsional'nykh gibridnykh intellektual'nykh sistem* [Methodology and technology for solving of complex problems using the methodology of functional hybrid artificial systems]. Moscow: IPI RAN. 387 p.
10. Kirikov, I. A., A. V. Kolesnikov, S. V. Listopad, and S. B. Rumovskaya. 2016 (in press). *Melkozernistyie gibridnye intellektual'nye sistemy. Chast' 2: Dvunapravlen'naya gibridizatsiya* [Fine-grained hybrid intelligent systems. Part 2: Bidirectional hybridization]. *Informatika i ee Primeneniya — Inform. Appl.* 1.
11. Pospelov, D. A. 1986. *Situatsionnoe upravlenie: Teoriya i praktika* [Situation control: Theory and practice]. Moscow: Nauka. 288 p.
12. Uemov, A. I. 1963. *Veshchi, svoystva, otnosheniya* [Items, properties, relations]. Moscow: Institute of Philosophy of the Academy of Sciences of the USSR. 184 p.
13. Kolesnikov, A. V. 2015. Modelirovanie estestvennykh geterogennykh sistem kollektivnogo prinyatiya resheniy [Modeling of the natural heterogeneous systems of collective decision-making]. *6th Conference (International) “Systems Analysis and Information Technology” Proceedings*. Moscow: ISA RAS. 1:7–16.
14. Lejduk, V. A. 1990. *Kontseptual'noe proektirovanie sistem s bazami znaniy* [Substructured developmet of the systems with knowledge base]. Kharkov: Osнова. 144 p.

Received September 20, 2015

Contributors

Kirikov Igor A. (b. 1955) — Candidate of Sciences (PhD) in technology; director, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; baltbipiran@mail.ru

Kolesnikov Alexander V. (b. 1948) — Doctor of Sciences in technology; professor, Department of telecommunications, Immanuel Kant Baltic Federal University, 14 Nevskogo Str., Kaliningrad 236041, Russian Federation; senior scientist, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; avkolesnikov@yandex.ru

Listopad Sergey V. (b. 1984) — Candidate of Sciences (PhD) in technology; scientist, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; ser-list-post@yandex.ru

Rumovskaya Sophiya B. (b. 1985) — programmer I, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; sophiyabr@gmail.com

ОБ УСРЕДНЕНИИ ОКРУГЛЕННЫХ ДАННЫХ*

В. Г. Ушаков¹, Н. Г. Ушаков²

Аннотация: Значения каждой наблюдаемой величины регистрируются с конечной точностью, определяемой разрешающей способностью измерительного инструмента. Естественно ожидать, что ошибки округления могут существенно повлиять на точность восстановления математического ожидания наблюдаемой величины. С другой стороны, часто исследователь имеет возможность воздействовать на наблюдения перед их регистрацией, например добавлять аддитивную или мультипликативную шумовую составляющую. В работе изучается связь между ошибкой измерения, погрешностью округления и точностью восстановления измеряемой величины при усреднении многократно проводимых измерений. Показано, что при одном и том же уровне округления точность восстановления тем выше, чем больше, в определенном смысле, ошибка измерения.

Ключевые слова: округленные данные; закон больших чисел; полная вариация; разложение вероятностных распределений

DOI: 10.14357/19922264150412

1 Введение

Округление экспериментальных данных является неизбежным этапом их статистической обработки. Так как аналоговые сигналы хранятся и обрабатываются в цифровом виде, они подвергаются аналого-цифровому преобразованию. Разрядность современных аналого-цифровых преобразователей большая, но конечная.

Исследования показали, что игнорирование этого факта может привести к значительным ошибкам при решении задач статистической обработки данных. В последнее время появилось много работ, где исследуется влияние погрешностей округления на решение различных статистических задач и предлагаются методы, которые можно применять в этой ситуации. В работах [1–5] можно найти как решения задач статистического анализа округленных данных, так и обширную библиографию.

В настоящей работе рассматривается следующая задача. Предположим, что помимо случайных ошибок, сопутствующих каждому реальному измерению, наблюдаемые значения содержат еще одну погрешность, связанную с ограниченной точностью измеряющего прибора (т. е. подвергаются округлению). При отсутствии этой погрешности и при наличии достаточно большого числа независимых измерений можно сколь угодно точно определить

измеряемую величину, если распределение случайной ошибки таково, что выполняется закон больших чисел. Можно ли надеяться на это, если есть еще и погрешность округления? В работе дается положительный ответ на этот вопрос. Средство достижения неограниченной точности кажется парадоксальным: наблюдаемая величина должна быть дополнительно зашумлена до округления, причем дополнительный шум не должен зависеть от наблюдаемой величины и чем больше его дисперсия, тем большей точности можно достигнуть.

Существуют различные типы округления: к меньшему, к большему, к ближайшему, случайное округление, чередующееся округление и др. Будем рассматривать округление к ближайшему. Кроме того, не ограничивая общности, будем считать, что шаг округления равен единице. Таким образом, для действительного числа x с целой частью $[x]$ и дробной частью $\{x\}$ округленное значение (обозначим его x^*) равно $[x]$, если $\{x\} < 0,5$, и $[x] + 1$, если $\{x\} \geq 0,5$. Заметим, что $x^* = [x + 1/2]$.

Итак, пусть случайные величины X_1, \dots, X_n являются результатами измерений величины μ . Предположим теперь, что сами случайные величины X_1, \dots, X_n наблюдаются с некоторыми дополнительными ошибками $\varepsilon_1, \dots, \varepsilon_n$, а затем еще регистрируются в округленном виде. Таким образом, наблюдения имеют вид $(X_1 + \varepsilon_1)^*, \dots, (X_n + \varepsilon_n)^*$.

* Работа выполнена при финансовой поддержке РФФИ (проект 14-11-00364).

¹ Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vgushakov@mail.ru

² Институт проблем технологии микроэлектроники и особо чистых материалов Российской академии наук, Черноголовка; Норвежский научно-технологический университет, Тронхейм, ushakov@math.ntnu.no

2 Вспомогательные результаты

В данном разделе докажем несколько вспомогательных результатов. Будем говорить, что функция распределения $G(x)$ является компонентой функции распределения $F(x)$, если существует функция распределения $H(x)$ такая, что

$$F(x) = \int_{-\infty}^{\infty} G(x-t) dH(t).$$

Полной вариацией действительной функции $f(x)$, определенной на конечном или бесконечном отрезке, называется величина

$$V(f) = \sup \sum_{i=0}^n |f(x_{i+1}) - f(x_i)|,$$

где верхняя грань берется по всем значениям n и всевозможным наборам $x_0 < x_1 < \dots < x_n$.

Лемма 1. Если равномерное на отрезке $[0, 1]$ распределение является компонентой распределения случайной величины X , то дробная часть $\{X\}$ имеет равномерное на $[0, 1]$ распределение.

Доказательство. Обозначим функции распределения случайных величин X и $\{X\}$ соответственно $F(x)$ и $F_1(x)$, а равномерную на $[0, 1]$ функцию распределения — $U(x)$. По условию леммы существует функция распределения $G(x)$ такая, что

$$F(x) = \int_{-\infty}^{\infty} G(x-t) dU(t) = \int_0^1 G(x-t) dt.$$

Следовательно, для любых $a < b$

$$\begin{aligned} P(a \leq X \leq b) &= F(b) - F(a) = \\ &= \int_0^1 G(b-t) dt - \int_0^1 G(a-t) dt \end{aligned}$$

(функция распределения $F(x)$ непрерывна, поскольку содержит в качестве компоненты равномерное распределение). Пусть $0 < x < 1$. Легко видеть, что

$$F_1(x) = \sum_{k=-\infty}^{\infty} P(k \leq X \leq k+x),$$

поэтому

$$\begin{aligned} F_1(x) &= \\ &= \sum_{k=-\infty}^{\infty} \left[\int_0^1 G(k+x-t) dt - \int_0^1 G(k-t) dt \right] = \end{aligned}$$

$$\begin{aligned} &= \sum_{k=-\infty}^{\infty} \left[\int_{x+k-1}^{x+k} G(y) dy - \int_{k-1}^k G(y) dy \right] = \\ &= \sum_{k=-\infty}^{\infty} \left[\int_k^{k+x} G(y) dy - \int_{k-1}^{k-1+x} G(y) dy \right] = \\ &= \lim_{m \rightarrow \infty} \sum_{k=-\infty}^m \left[\int_k^{k+x} G(y) dy - \int_{k-1}^{k-1+x} G(y) dy \right] = \\ &= \lim_{m \rightarrow \infty} \int_m^{m+x} G(y) dy = x, \end{aligned}$$

т. е. $F_1(x)$ равномерна на $[0, 1]$.

Лемма 2. Пусть X — абсолютно непрерывная случайная величина с непрерывной плотностью распределения $f(x)$. Тогда

$$\left| E\{X\} - \frac{1}{2} \right| \leq \frac{V(f)}{2}.$$

Доказательство. Обозначим

$$p_k = \int_k^{k+1} f(x) dx = \int_0^1 f(x+k) dx, \quad k = 0, \pm 1, \pm 2, \dots$$

Для каждого k пусть x'_k, x''_k и $x_k^{(0)}$ — точки интервала $[0, 1]$ такие, что

$$\begin{aligned} f(x'_k + k) &= \min_{0 \leq x \leq 1} f(x+k); \\ f(x''_k + k) &= \max_{0 \leq x \leq 1} f(x+k); \end{aligned}$$

$$\int_0^1 x f(x+k) dx = f(x_k^* + k) \int_0^1 x dx = \frac{f(x_k^* + k)}{2}.$$

В силу равенства

$$P(\{X\} \leq x) = \sum_{k=-\infty}^{\infty} P(k \leq X \leq k+x)$$

при $0 < x < 1$, получаем

$$E\{X\} = \sum_{k=-\infty}^{\infty} \int_0^1 x f(x+k) dx$$

и, следовательно,

$$\begin{aligned} \left| E\{X\} - \frac{1}{2} \right| &= \\ &= \left| \sum_{k=-\infty}^{\infty} \left(\int_0^1 x f(x+k) dx - p_k \int_0^1 x dx \right) \right| \leq \\ &\leq \sum_{k=-\infty}^{\infty} \left| \int_0^1 x f(x+k) dx - p_k \int_0^1 x dx \right| = \end{aligned}$$

$$= \frac{1}{2} \sum_{k=-\infty}^{\infty} |f(x_k^* + k) - p_k| \leq \leq \frac{1}{2} \sum_{k=-\infty}^{\infty} |f(x_k' + k) - f(x_k'' + k)| \leq \frac{V(f)}{2}.$$

3 Закон больших чисел

Итак, имеется последовательность независимых и одинаково распределенных случайных величин X_1, X_2, \dots с математическим ожиданием μ и последовательность независимых (и независимых от X_1, X_2, \dots) и одинаково распределенных абсолютно непрерывных случайных величин $\varepsilon_1, \varepsilon_2, \dots$ с нулевым математическим ожиданием.

Следующая теорема дает простое достаточное условие, когда μ может быть восстановлено с неограниченной точностью (разумеется, при неограниченном числе измерений).

Теорема 1. Если равномерное на отрезке единичной длины распределение является компонентой распределения случайных величин ε_i , то

$$\frac{1}{n} \sum_{i=1}^n (X_i + \varepsilon_i)^* \xrightarrow{\text{a.s.}} \mu$$

при $n \rightarrow \infty$.

Доказательство. Из условия теоремы следует, что равномерное на отрезке $[0, 1]$ распределение является компонентой распределения случайных величин $X_i + \varepsilon_i + 1/2$; следовательно, в силу леммы 1 распределение дробной части $\{X_i + \varepsilon_i + 1/2\}$ равномерно на $[0, 1)$ и

$$E \left\{ X_i + \varepsilon_i + \frac{1}{2} \right\} = \frac{1}{2}.$$

С учетом этого получаем

$$E (X_i + \varepsilon_i)^* = E \left[X_i + \varepsilon_i + \frac{1}{2} \right] = = E \left(X_i + \varepsilon_i + \frac{1}{2} \right) - E \left\{ X_i + \varepsilon_i + \frac{1}{2} \right\} = \mu.$$

Результат теперь следует из классического усиленного закона больших чисел.

Теорема 2. Если плотность распределения случайных величин ε_i имеет ограниченную вариацию и $\lambda \rightarrow \infty$, то

$$\frac{1}{n} \sum_{i=1}^n (X_i + \lambda \varepsilon_i)^* \xrightarrow{\text{a.s.}} \mu$$

при $n \rightarrow \infty$, т. е. предел (почти наверное) левой части при $n \rightarrow \infty$ стремится к μ при стремлении λ к бесконечности.

Доказательство. В силу усиленного закона больших чисел теорема будет доказана, если показать, что

$$\lim_{\lambda \rightarrow \infty} E (X_i + \lambda \varepsilon_i)^* = \mu. \quad (1)$$

В свою очередь, в силу равенства

$$E (X_i + \lambda \varepsilon_i)^* = E \left(X_i + \lambda \varepsilon_i + \frac{1}{2} \right) - - E \left\{ X_i + \lambda \varepsilon_i + \frac{1}{2} \right\}$$

уравнение (1) будет доказано, если показать, что

$$\lim_{\lambda \rightarrow \infty} E \left\{ X_i + \lambda \varepsilon_i + \frac{1}{2} \right\} = \frac{1}{2}. \quad (2)$$

Пусть $f(x)$ и $f_\lambda(x)$ — плотности распределения случайных величин $X_i + \lambda \varepsilon_i + 1/2$ и $\lambda \varepsilon_i$ соответственно. Поскольку

$$f_\lambda(x) = \frac{1}{\lambda} f_1 \left(\frac{x}{\lambda} \right),$$

имеем

$$V(f_\lambda) = \frac{1}{\lambda} V(f_1).$$

С другой стороны, $V(f) \leq V(f_\lambda)$. Таким образом,

$$\lim_{\lambda \rightarrow \infty} V(f) = 0.$$

Отсюда в силу леммы 2 следует (2) и, следовательно, (1).

Отметим, что дисперсия дополнительного зашумления пропорциональна λ^2 . Таким образом, верхняя граница точности восстановления оказывается тем меньше, чем больше дисперсия ошибки. Цена, которую приходится за это платить, — необходимость увеличения числа наблюдений.

Литература

1. Hall P. The influence of rounding errors on some nonparametric estimators of a density and its derivatives // SIAM J. Appl. Math., 1982. Vol. 42. P. 390–399.
2. Zhidong B., Shurong Z., Baoxue Z., Guorong H. Statistical analysis for rounded data // J. Stat. Plan. Infer., 2009. Vol. 139. P. 2526–2542.
3. Schneeweiss H., Komlos J., Ahmad A.S. Symmetric and asymmetric rounding: A review and some new results // AStA Adv. Stat. Anal., 2010. Vol. 94. P. 247–271.
4. Weiming L., Tianqing L., Zhidong B. Rounded data analysis based on ranked set sample // Stat. Papers, 2012. Vol. 53. P. 439–455.
5. Ningning Z., Zhidong B. Analysis of rounded data in mixture normal model // Stat. Papers, 2012. Vol. 53. P. 895–914.

Поступила в редакцию 10.11.15

ON AVERAGING OF ROUNDED DATA

V. G. Ushakov^{1,2} and N. G. Ushakov^{3,4}

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 Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

³Institute of Microelectronics Technology and High-Purity Materials of the Russian Academy of Sciences, 6 Academician Osipyan Str., Chernogolovka, Moscow Region 142432, Russian Federation

⁴Norwegian University of Science and Technology, 15A S.P. Andersensvei, Trondheim 7491, Norway

Abstract: Each observed value is registered with finite accuracy which is determined by the sensitivity of the equipment. It is expected that rounding errors could play an important role in the estimation of the mean of the observed value. On the other hand, the researcher usually has a possibility to affect the observation before its registration, for example, to intensify it or to add some additional component. This paper studies the relationship between the measurement error, rounding error, and the accuracy of the reconstruction of the observed value for the case of averaging of repeated measurements. It is demonstrated that under a fixed rounding level, in some sense, the greater the measurement error, the higher the reconstruction accuracy.

Keywords: rounded data; law of large numbers; total variation; decomposition of probability distributions

DOI: 10.14357/19922264150412

Acknowledgments

The work was supported by the Russian Science Foundation (project 14-11-00364).

References

- Hall, P. 1982. The influence of rounding errors on some nonparametric estimators of a density and its derivatives. *SIAM J. Appl. Math.* 42:390–399.
- Zhidong, B., Z. Shurong, Z. Baoxue, and H. Guorong. 2009. Statistical analysis for rounded data. *J. Stat. Plan. Infer.* 139:2526–2542.
- Schneeweiss, H., J. Komlos, and A. S. Ahmad. 2010. Symmetric and asymmetric rounding: A review and some new results. *ASTA Adv. Stat. Anal.* 94:247–271.
- Weiming, L., L. Tianqing, and B. Zhidong. 2012. Rounded data analysis based on ranked set sample. *Stat. Papers* 53:439–455.
- Ningning, Z., and B. Zhidong. 2012. Analysis of rounded data in mixture normal model. *Stat. Papers* 53:895–914.

Received November 10, 2015

Contributors

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 Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vgushakov@mail.ru

Ushakov Nikolai G. (b. 1952) — Doctor of Science in physics and mathematics; leading scientist, Institute of Microelectronics Technology and High-Purity Materials of the Russian Academy of Sciences, 6 Academician Osipyan Str., Chernogolovka, Moscow Region 142432, Russian Federation; professor, Norwegian University of Science and Technology, 15A S.P. Andersensvei, Trondheim 7491, Norway; ushakov@math.ntnu.no

Беляев Константин Павлович (р. 1955) — доктор физико-математических наук, ведущий научный сотрудник Института океанологии им. П. П. Ширшова; приглашенный профессор, Федеральный университет штата Баия, Сальвадор, Бразилия

Боровский Павел (р. 1986) — магистрант кафедры программной инженерии Академического инженерного колледжа Шамуна, Беер-Шева, Израиль

Вихрова Ольга Геннадиевна (р. 1990) — аспирантка Российского университета дружбы народов

Горшенин Андрей Константинович (р. 1986) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент Московского государственного университета информационных технологий, радиотехники и электроники

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, ведущий 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

Гулев Сергей Константинович (р. 1958) — член-корреспондент Российской академии наук, доктор физико-математических наук, заведующий лабораторией Института океанологии им. П. П. Ширшова; профессор кафедры океанологии географического факультета Московского государственного университета имени М. В. Ломоносова; профессор Кильского университета, Германия

Забейайло Михаил Иванович (р. 1956) — кандидат физико-математических наук, доцент, старший научный сотрудник Всероссийского института научной и технической информации Российской академии наук (ВИНИТИ РАН)

Запаринный Александр Алексеевич (р. 1951) — доктор технических наук, профессор, заместитель директора Федерального исследовательского цент-

ра «Информатика и управление» Российской академии наук

Змеев Дмитрий Николаевич (р. 1978) — научный сотрудник Института проблем проектирования в микроэлектронике Российской академии наук

Кириков Игорь Александрович (р. 1955) — кандидат технических наук, директор Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Климов Аркадий Валентинович (р. 1953) — старший научный сотрудник Института проблем проектирования в микроэлектронике Российской академии наук

Колесников Александр Васильевич (р. 1948) — доктор технических наук, профессор кафедры телекоммуникаций Российского государственного университета имени Иммануила Канта, старший научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Коновалов Михаил Григорьевич (р. 1950) — доктор технических наук, заведующий сектором Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Копицкая Марина (р. 1966) — PhD по информатике, старший лектор кафедры программной инженерии Академического инженерного колледжа Шамуна, Беер-Шева, Израиль

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Корчагин Александр Юрьевич (р. 1989) — младший научный сотрудник факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Левченко Николай Николаевич (р. 1978) — кандидат технических наук, заведующий отделом Институт

проблем проектирования в микроэлектронике Российской академии наук

Листопад Сергей Викторович (р. 1984) — кандидат технических наук, научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Молотковский Роман (р. 1989) — студент кафедры программной инженерии Академического инженерного колледжа Шамуна, Беер-Шева, Израиль

Окунев Анатолий Семенович (р. 1951) — кандидат технических наук, ведущий научный сотрудник Института проблем проектирования в микроэлектронике Российской академии наук

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, доцент Российского университета дружбы народов

Румовская София Борисовна (р. 1985) — программист 1-й категории Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Самуйлов Константин Евгеньевич (р. 1955) — доктор технических наук, профессор, заведующий кафедрой Российского университета дружбы народов

Синицын Игорь Николаевич (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Соколов Игорь Анатольевич (р. 1954) — академик Российской академии наук, доктор технических наук, директор Федерального исследовательского центра «Информатика и управление» Российской академии наук

Сопин Эдуард Сергеевич (р. 1987) — кандидат физико-математических наук, доцент Российского университета дружбы народов

Стемпковский Александр Леонидович (р. 1950) — доктор технических наук, профессор, академик Российской академии наук, директор Института проблем проектирования в микроэлектронике 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

Ушаков Владимир Георгиевич (р. 1952) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ушаков Николай Георгиевич (р. 1954) — доктор физико-математических наук, ведущий научный сотрудник Института проблем технологии микроэлектроники и особо чистых материалов Российской академии наук (Черноголовка); профессор Норвежского научно-технологического университета (г. Тронхейм)

Френкель Сергей Лазаревич (р. 1951) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук; доцент Московского государственного университета информационных технологий, радиотехники и электроники

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, заместитель директора Федерального исследовательского центра «Информатика и управление» Российской академии наук

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2015 г.

	Выпуск	Стр.
Абрамов М. О., Катаев М. Ю. Влияние морфологических операций на распознавание фигуры движущегося человека по потоку изображений	3	124
Агаларов Я. М. Пороговая стратегия ограничения доступа к ресурсам в системе массового обслуживания $M/D/1$ с функцией штрафов за несвоевременное обслуживание заявок	3	55
Беляев К. П. см. Королев В. Ю.		
Беляев К. П., Тучкова Н. П. Предельные распределения для характеристик при усвоении данных наблюдений в стационарном режиме	2	50
Беляев Ю. К., Кристрём Б. Анализ обзорных обследований, содержащих цензурированные данные в округленных интервалах	3	2
Березин С. В., Заяц О. И. Применение уравнения Пугачёва–Свешникова к решению задачи Бакстера о длительности выбросов	2	39
Боровский П. см. Френкель С.		
Бронштейн Е. М., Вагапова Д. М. Сравнительный анализ применения эвристического и метаэвристического алгоритмов к задаче о школьном автобусе	2	56
Вагапова Д. М. см. Бронштейн Е. М.		
Васильев Н. С. Коалиционно устойчивые эффективные равновесия в моделях коллективного поведения с обменом информацией	2	2
Вихрова О. Г., Самуйлов К. Е., Сопин Э. С., Шоргин С. Я. К анализу показателей качества обслуживания в современных беспроводных сетях	4	48
Гайдамака Ю. В., Самуйлов А. К. Метод расчета характеристик интерференции двух взаимодействующих устройств в беспроводной гетерогенной сети	1	9
Галина И. В., Козеренко Е. Б., Морозова Ю. И., Сомин Н. В., Шарнин М. М. Ассоциативные портреты предметной области — инструмент автоматизированного построения систем big data для извлечения знаний: теория, методика, визуализация, возможное применение	2	93
Горбунова А. В., Зарядов И. С., Матюшенко С. И., Самуйлов К. Е., Шоргин С. Я. Аппроксимация времени отклика системы облачных вычислений	3	32
Горшенин А. К. см. Королев В. Ю.		
Грушо А. А., Грушо Н. А., Тимонина Е. Е. Оценки скорости передачи информации и пропускной способности в скрытых каналах с метками	4	85
Грушо А. А., Забейайло М. И., Зацаринный А. А. Контроль и управление информационными потоками в облачной среде	4	91
Грушо Н. А. см. Грушо А. А.		
Гулев С. К. см. Королев В. Ю.		
Долев Ш., Коган-Садецкая М. Эвристические сертификаты посредством приближений	1	15
Жворонкова Ю. В., Кудрявцев А. А., Шоргин С. Я. Байесовская рекуррентная модель роста надежности: бета-равномерное распределение параметров	1	98
Забейайло М. И. см. Грушо А. А.		
Зарядов И. С. см. Горбунова А. В.		
Зацаринный А. А. см. Грушо А. А.		
Зацман И. М. Процессы целенаправленной генерации и развития кросс-языковых экспертных знаний: семиотические основания моделирования	3	106
Зацман И. М. см. Минин В. А.		
Заяц О. И. см. Березин С. В.		
Змеев Д. Н., Климов А. В., Левченко Н. Н., Окунев А. С., Стемповский А. Л. Поточковая модель вычислений как парадигма программирования будущего	4	29

	Выпуск	Стр.
Зыкин В. С. Ссылочная целостность данных в корпоративных информационных системах	3	97
Калиниченко Л. А., Ковалев Д. Ю., Ковалева Д. А., Малков О. Ю. Методы и средства поддержки исследований, движимых гипотезами: обзор	1	28
Кантор О. Г. см. Спивак С. И.		
Катаев М. Ю. см. Абрамов М. О.		
Кириков И. А., Колесников А. В., Листопад С. В., Румовская С. Б. Мелкозернистые гибридные интеллектуальные системы. Часть 1: Лингвистический подход	4	98
Климов А. В. см. Змеев Д. Н.		
Ковалев Д. Ю. см. Калиниченко Л. А.		
Ковалева Д. А. см. Калиниченко Л. А.		
Ковалёв С. П. Формальный аксиоматический подход к аспектно-ориентированному расширению технологий программирования	1	55
Коган-Садецкая М. см. Долев Ш.		
Козеренко Е. Б. см. Галина И. В.		
Колесников А. В. см. Кириков И. А.		
Колесов С. В. см. Спивак С. И.		
Коновалов М. Г., Разумчик Р. В. Обзор моделей и алгоритмов размещения заданий в системах с параллельным обслуживанием	4	56
Копицкая М. см. Френкель С.		
Корепанов Э. Р. см. Сеницын И. Н.		
Корепанов Э. Р. см. Сеницын И. Н.		
Корепанов Э. Р. см. Сеницын И. Н.		
Корепанов Э. Р. см. Сеницын И. Н.		
Королев В. Ю., Горшенин А. К., Гулев С. К., Беляев К. П. Статистическое моделирование турбулентных потоков тепла между океаном и атмосферой с помощью метода скользящего разделения конечных нормальных смесей	4	3
Королев В. Ю., Корчагин А. Ю., Соколов И. А. Обобщенные дисперсионные гамма-распределения как модели статистических закономерностей на финансовых рынках	4	14
Корчагин А. Ю. см. Королев В. Ю.		
Корчажкина О. М. К оценке эффективности учебно-познавательной деятельности учащихся с использованием информационных технологий	1	106
Кривенко М. П. Модели для представления и обработки референсных значений	2	63
Кристрём Б. см. Беляев Ю. К.		
Кудрявцев А. А. см. Жаворонкова Ю. В.		
Кузнецов С. И. см. Спивак С. И.		
Лебедев А. В. Экстремальные индексы в схеме серий и их приложения	3	39
Левченко Н. Н. см. Змеев Д. Н.		
Лери М. М. Пожар на конфигурационном графе со случайными переходами огня по ребрам	3	65
Листопад С. В. см. Кириков И. А.		
Малков О. Ю. см. Калиниченко Л. А.		
Матюшенко С. И. см. Горбунова А. В.		
Мейханаджян Л. А., Милованова Т. А., Разумчик Р. В. Время ожидания в системе обслуживания с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом	2	14
Милованова Т. А. см. Мейханаджян Л. А.		
Минин В. А., Зацман И. М., Хавансков В. А., Шубников С. К. Индикаторы тематических взаимосвязей отраслей науки и информационно-компьютерных технологий в начале XXI века	2	111
Миронов А. М. Реализуемость вероятностных реакций конечными вероятностными автоматами	3	85
Молотковский Р. см. Френкель С.		
Морозова Ю. И. см. Галина И. В.		
Окунев А. С. см. Змеев Д. Н.		
Печинкин А. В., Разумчик Р. В. Совместное стационарное распределение числа заявок в m очередях в N -канальной системе обслуживания с переупорядочением заявок	3	25

	Выпуск	Стр.
Попова М. С., Стрижов В. В. Выбор оптимальной модели классификации физической активности по измерениям акселерометра	1	76
Разумчик Р. В. Алгебраический метод приближенного расчета стационарного распределения в системе обслуживания конечной емкости с отрицательными заявками и двумя очередями	4	68
Разумчик Р. В. см. Коновалов М. Г.		
Разумчик Р. В. см. Мейханаджян Л. А.		
Разумчик Р. В. см. Печинкин А. В.		
Румовская С. Б. см. Кириков И. А.		
Самуйлов А. К. см. Гайдамака Ю. В.		
Самуйлов К. Е. см. Вихрова О. Г.		
Самуйлов К. Е. см. Горбунова А. В.		
Синицын В. И. см. Синицын И. Н.		
Синицын В. И. см. Синицын И. Н.		
Синицын И. Н. Аналитическое моделирование процессов в динамических системах с цилиндрическими бесселевыми нелинейностями	4	37
Синицын И. Н. Аналитическое моделирование распределений методом ортогональных разложений в нелинейных стохастических системах на многообразиях	3	17
Синицын И. Н., Корепанов Э. Р. Нормальные условно-оптимальные фильтры Пугачёва для дифференциальных стохастических систем, линейных относительно состояния	2	30
Синицын И. Н., Корепанов Э. Р. Устойчивые линейные условно оптимальные фильтры и экстраполяторы для стохастических систем с мультипликативными шумами	1	70
Синицын И. Н., Синицын В. И., Корепанов Э. Р. Моделирование нормальных процессов в стохастических системах со сложными иррациональными нелинейностями	1	2
Синицын И. Н., Синицын В. И., Корепанов Э. Р. Моделирование нормальных процессов в стохастических системах со сложными трансцендентными нелинейностями	2	23
Соколов И. А. см. Королев В. Ю.		
Сомин Н. В. см. Галина И. В.		
Сопин Э. С. см. Вихрова О. Г.		
Спивак С. И., Кантор О. Г., Юнусова Д. С., Кузнецов С. И., Колесов С. В. Оценка погрешности и значимости измерений для линейных моделей	1	87
Стемпковский А. Л. см. Змеев Д. Н.		
Стенина М. М., Стрижов В. В. Согласование прогнозов при решении задач прогнозирования иерархических временных рядов	2	75
Стрижов В. В. см. Попова М. С.		
Стрижов В. В. см. Стенина М. М.		
Тимонина Е. Е. см. Грушо А. А.		
Тучкова Н. П. см. Беляев К. П.		
Ушаков В. Г., Ушаков Н. Г. Об усреднении округленных данных	4	106
Ушаков Н. Г. см. Ушаков В. Г.		
Френкель С., Копицкая М., Молотковский Р., Боровский П. Улучшение производительности алгоритма сжатия данных Лемпеля–Зива–Уэлча	4	78
Хавансков В. А. см. Минин В. А.		
Чичагов В. В. Асимптотические разложения высокого порядка для несмещенных оценок и их дисперсий в модели однопараметрического экспоненциального семейства	3	72
Шарнин М. М. см. Галина И. В.		
Шестаков О. В. Непараметрическое оценивание многомерной плотности с помощью вейвлет-оценок одномерных проекций	2	88
Шоргин С. Я. см. Вихрова О. Г.		
Шоргин С. Я. см. Горбунова А. В.		
Шоргин С. Я. см. Жаворонкова Ю. В.		
Шубников С. К. см. Минин В. А.		
Юнусова Д. С. см. Спивак С. И.		

2015 AUTHOR INDEX

	Issue	Page
Abramov M. O. and Kataev M. Yu. Influence of morphological operations on recognition of a moving human figure from a set of images	3	124
Agalarov Ya. M. The threshold strategy for restricting access in the $M/D/1$ queueing system with penalty function for late service	3	55
Belyaev K. P. see Korolev V. Yu.		
Belyaev K. P. and Tuchkova N. P. On a limit distribution of characteristics in stationary regime for the linear assimilation problem	2	50
Belyaev Yu. K. and Kristrom B. Analysis of survey data containing rounded censoring intervals	3	2
Berezin S. V. and Zayats O. I. Application of the Pugachev–Sveshnikov equation to the Baxter occupation time problem	2	39
Borovsky P. see Frenkel S.		
Bronshtein E. M. and Vagapova D. M. Comparative analysis of application of heuristic and metaheuristic algorithms to the school bus routing problem	2	56
Charnine M. M. see Galina I. V.		
Chichagov V. V. Higher-order asymptotic expansions of unbiased estimators and their variances on the one-parameter exponential family model	3	72
Dolev Sh. and Kogan-Sadetsky M. Heuristic certificates via approximations	1	15
Frenkel S., Kopeetsky M., Molotkovski R., and Borovsky P. Performance improvement of Lempel–Ziv–Welch compression algorithm	4	78
Gaidamaka Yu. V. and Samuylov A. K. Method for calculating numerical characteristics of two devices interference for device-to-device communications in a wireless heterogeneous network	1	9
Galina I. V., Kozerenko E. B., Morozova Yu. I., Somin N. V., and Charnine M. M. Associative portraits of subject areas as a tool for automated construction of big data systems for knowledge extraction: Theory, methods, visualization and application	2	93
Gorbunova A. V., Zaryadov I. S., Matyushenko S. I., Samouylov K. E., and Shorgin S. Ya. The approximation of response time of a cloud computing system	3	32
Gorshenin A. K. see Korolev V. Yu.		
Grusho A. A., Grusho N. A., and Timonina E. E. Rate of information transfer and capacity in covert channels defined by tags	4	85
Grusho A. A., Zabezhailo M. I., and Zatsarinny A. A. Information flow monitoring and control in the cloud computing environment	4	91
Grusho N. A. see Grusho A. A.		
Gulev S. K. see Korolev V. Yu.		
Havanskov V. A. see Minin V. A.		
Kalinichenko L. A., Kovalev D. Yu., Kovaleva D. A., and Malkov O. Yu. Methods and tools for hypothesis-driven research support: A survey	1	28
Kantor O. G. see Spivak S. I.		
Kataev M. Yu. see Abramov M. O.		
Kirikov I. A., Kolesnikov A. V., Listopad S. V., and Rumovskaya S. B. Fine-grained hybrid intelligent systems. Part 1: Linguistic approach	4	98
Klimov A. V. see Zmejev D. N.		
Kogan-Sadetsky M. see Dolev Sh.		
Kolesnikov A. V. see Kirikov I. A.		
Kolesov S. V. see Spivak S. I.		
Konovalov M. G. and Razumchik R. V. Methods and algorithms for job scheduling in systems with parallel service: A survey	4	56

	Issue	Page
Kopeetsky M. see Frenkel S.		
Korchagin A. Yu. see Korolev V. Yu.		
Korchazhkina O. M. On access to the efficiency of students' cognitive activities while using the new information technologies	1	106
Korepanov E. R. see Sinitsyn I. N.		
Korepanov E. R. see Sinitsyn I. N.		
Korepanov E. R. see Sinitsyn I. N.		
Korepanov E. R. see Sinitsyn I. N.		
Korolev V. Yu., Gorshenin A. K., Gulev S. K., and Belyaev K. P. Statistical modeling of air–sea turbulent heat fluxes by the method of moving separation of finite normal mixtures	4	3
Korolev V. Yu., Korchagin A. Yu., and Sokolov I. A. Modeling of statistical regularities in financial markets by generalized variance gamma distributions	4	14
Kovalev D. Yu. see Kalinichenko L. A.		
Kovaleva D. A. see Kalinichenko L. A.		
Kovalyov S. P. Formal axiomatic approach to aspect-oriented extension of programming technologies	1	55
Kozerenko E. B. see Galina I. V.		
Kristrom B. see Belyaev Yu. K.		
Krivenko M. P. Models for representation and treatment of reference values	2	63
Kudryavtsev A. A. see Zhavoronkova Iu. V.		
Kuznetsov S. I. see Spivak S. I.		
Lebedev A. V. Extremal indices in a series scheme and their applications	3	39
Leri M. M. Forest fire on a configuration graph with random fire propagation	3	65
Levchenko N. N. see Zmejev D. N.		
Listopad S. V. see Kirikov I. A.		
Malkov O. Yu. see Kalinichenko L. A.		
Matyushenko S. I. see Gorbunova A. V.		
Meykhanadzhyan L. A., Milovanova T. A., and Razumchik R. V. Stationary waiting time in a queueing system with inverse service order and generalized probabilistic priority	2	14
Milovanova T. A. see Meykhanadzhyan L. A.		
Minin V. A., Zatsman I. M., Havanskov V. A., and Shubnikov S. K. Indicators for thematic linkages between science and information and computer technologies at the beginning of the XXI century	2	111
Mironov A. M. Realizability of probabilistic reactions by finite probabilistic automata	3	85
Molotkovski R. see Frenkel S.		
Morozova Yu. I. see Galina I. V.		
Okunev A. S. see Zmejev D. N.		
Pechinkin A. V. and Razumchik R. V. Joint stationary distribution of m queues in the N -server queueing system with reordering	3	25
Popova M. S. and Strijov V. V. Selection of optimal physical activity classification model using measurements of accelerometer	1	76
Razumchik R. V. Algebraic method for approximating joint stationary distribution in finite capacity queue with negative customers and two queues	4	68
Razumchik R. V. see Konovalov M. G.		
Razumchik R. V. see Meykhanadzhyan L. A.		
Razumchik R. V. see Pechinkin A. V.		
Rumovskaya S. B. see Kirikov I. A.		
Samouylov K. E. see Gorbunova A. V.		
Samouylov K. E. see Vikhrova O. G.		
Samuylov A. K. see Gaidamaka Yu. V.		
Shestakov O. V. Nonparametric estimation of multidimensional density with the use of wavelet estimates of univariate projections	2	88

	Issue	Page
Shorgin S. Ya. see Gorbunova A. V.		
Shorgin S. Ya. see Vikhrova O. G.		
Shorgin S. Ya. see Zhavoronkova Iu. V.		
Shubnikov S. K. see Minin V. A.		
Sinitsyn I. N. Analytical modeling in stochastic systems on manifolds based on orthogonal expansions	3	17
Sinitsyn I. N. Analytical modeling of processes in dynamical systems with cylindric Bessel nonlinearities	4	37
Sinitsyn I. N. and Korepanov E. R. Normal Pugachev filters for state linear stochastic systems	2	30
Sinitsyn I. N. and Korepanov E. R. Stable linear conditionally optimal filters and extrapolators for stochastic systems with multiplicative noises	1	70
Sinitsyn I. N., Sinitsyn V. I., and Korepanov E. R. Analytical modeling of normal processes in stochastic systems with complex irrational nonlinearities	1	2
Sinitsyn I. N., Sinitsyn V. I., and Korepanov E. R. Modeling of normal processes in stochastic systems with complex transcendental nonlinearities	2	23
Sinitsyn V. I. see Sinitsyn I. N.		
Sinitsyn V. I. see Sinitsyn I. N.		
Sokolov I. A. see Korolev V. Yu.		
Somin N. V. see Galina I. V.		
Sopin E. S. see Vikhrova O. G.		
Spivak S. I., Kantor O. G., Yunusova D. S., Kuznetsov S. I., and Kolesov S. V. Evaluation of measurement accuracy and significance for linear models	1	87
Stempkovsky A. L. see Zmejev D. N.		
Stenina M. M. and Strijov V. V. Forecasts reconciliation for hierarchical time series forecasting problem	2	75
Strijov V. V. see Popova M. S.		
Strijov V. V. see Stenina M. M.		
Timonina E. E. see Grusho A. A.		
Tuchkova N. P. see Belyaev K. P.		
Ushakov N. G. see Ushakov V. G.		
Ushakov V. G. and Ushakov N. G. On averaging of rounded data	4	106
Vagapova D. M. see Bronshtein E. M.		
Vasilyev N. S. On availability of Pareto effective equilibrium situations in collective behavior models with data exchange	2	2
Vikhrova O. G., Samouylov K. E., Sopin E. S., and Shorgin S. Ya. On performance analysis of modern wireless networks	4	48
Yunusova D. S. see Spivak S. I.		
Zabekhailo M. I. see Grusho A. A.		
Zaryadov I. S. see Gorbunova A. V.		
Zatsarinny A. A. see Grusho A. A.		
Zatsman I. M. Goal-oriented processes of cross-lingual expert knowledge creation: Semiotic foundations for modeling	3	106
Zatsman I. M. see Minin V. A.		
Zayats O. I. see Berezin S. V.		
Zhavoronkova Iu. V., Kudryavtsev A. A., and Shorgin S. Ya. Bayesian recurrent model of reliability growth: Beta-uniform distribution of parameters	1	98
Zmejev D. N., Klimov A. V., Levchenko N. N., Okunev A. S., and Stempkovsky A. L. Dataflow computing model as a paradigm of future mainstream of software development	4	29
Zykin V. S. Referential integrity of data in corporate information systems	3	97

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 не должна нарушать закон об авторских правах.

Направляя рукопись в редакцию, авторы сохраняют все права собственников данной рукописи и при этом передают учредителям и редколлегии неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на ее распространение в России и за рубежом. Авторы должны представить в редакцию письмо в следующей форме:

Соглашение о передаче права на публикацию:

«Мы, нижеподписавшиеся, авторы рукописи «. . .», передаем учредителям и редколлегии журнала «Информатика и её применения» неисключительное право опубликовать данную рукопись статьи на русском языке как в печатной, так и в электронной версиях журнала. Мы подтверждаем, что данная публикация не нарушает авторского права других лиц или организаций, а также не содержит сведений, запрещенных к опубликованию в открытой печати.

Подписи авторов: (ф. и. о., дата, адрес)».

Это соглашение может быть представлено в бумажном виде или в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации представленной 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 L^AT_EX (.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 20 страниц указанного формата.

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://translit.ru/> бесплатной программой транслитерации русского текста в латиницу, при этом в закладке «варианты. . .» следует выбрать опцию BGN.

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 resursoberegayushchie 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 Publ. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. I: 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 slovarya 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 Publ. 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 L^AT_EX (.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 20 pages of the specified format.

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.ru>, option 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 resursoberegayushchie 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.
- 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 slovary 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>