

Информатика и её применения

Том 6 Выпуск 3 Год 2012

СОДЕРЖАНИЕ

Предисловие 3

**Статьи, являющиеся развитием докладов, представленных
на конференции «Электронные библиотеки:
перспективные методы и технологии, электронные коллекции»
(RCDL'2011, г. Воронеж, 19–22 октября 2011 г.)**

Исследование радиоисточников средствами виртуальной обсерватории
О. П. Желенкова 5

Методологические аспекты разработки электронного русско-английского
тезауруса по компьютерной лингвистике
**Ю. А. Загорулько, О. И. Боровикова, И. С. Кононенко,
Е. Г. Соколова** 22

Классификация и использование семантических связей между
информационными объектами в научных электронных библиотеках
М. Р. Когаловский, С. И. Паринов 32

Об интеграции географических метаданных посредством
ретроспективного тезауруса
Д. М. Скачков, О. Л. Жижимов 43

Универсальная система проверки текстов на плагиат «Автор.NET»
Е. В. Шарапова, Р. В. Шарапов 52

**Статьи, являющиеся развитием докладов, представленных
на XXIX Международном семинаре
по проблемам устойчивости стохастических моделей
(г. Светлогорск Калининградской области России,
10–16 октября 2011 г.)**

Fractional Levy motion with dependent increments and its application
to network traffic modeling
C. De Nikola, Y. S. Khokhlov, M. Pagano, and O. I. Sidorova 59

About the rate of convergence of one U-statistic
O. Yanushkevichiene and R. Yanushkevichius 64

Технический редактор Л. Кокушкина

Художественный редактор М. Седакова

Сдано в набор 01.07.12. Подписано в печать 20.09.12. Формат 60 x 84 / 8

Бумага офсетная. Печать цифровая. Усл.-печ. л. 16. Уч.-изд. л. 12. Тираж 100 экз. Заказ №

Издательство «ТОРУС ПРЕСС», Москва 119991, ул. Косыгина, д. 4

torus@torus-press.ru; <http://www.torus-press.ru>

Отпечатано в Академиздатцентре «Наука» РАН с готовых файлов, Москва 121099, Шубинский пер., д. 6.

Информатика и её применения

Том 6 Выпуск 3 Год 2012

СОДЕРЖАНИЕ

Оценки скорости сходимости распределений случайных сумм к дисперсионным гамма-распределениям В. Е. Бенинг, Л. М. Закс, В. Ю. Королев	69
Задачи оптимального планирования межуровневого интерфейса в беспроводных сетях Ю. В. Гайдамака, Т. В. Ефимушкина, А. К. Самуйлов, К. Е. Самуйлов	74
Асимптотика максимума процесса нагрузки для некоторого класса гауссовских очередей О. В. Лукашенко, Е. В. Морозов	81
Об оценивании вероятности переполнения конечного буфера в регенеративных системах обслуживания Е. В. Морозов, Р. С. Некрасова	90
Вероятностные модели многопроцессорных систем: стационарность и моментные свойства Е. В. Морозов, А. С. Румянцев	99
Ограничение на суммарный объем заявок в дискретной системе $Geo/G/1/\infty$ А. В. Печинкин, И. А. Соколов, С. Я. Шоргин	107
Анализ системы обслуживания с гиперэкспоненциальным входящим потокм в условиях критической загрузки А. В. Ушаков	114
Abstracts	119
Об авторах	124
About Authors	126

Предисловие

Вниманию читателей предлагается очередной выпуск журнала «Информатика и её применения», полностью посвященный публикации расширенных версий докладов, представленных на ряде научных мероприятий, которые были проведены в 2011 г. с активным организационным и научным участием Российской академии наук (и прежде всего Института проблем информатики Российской академии наук).

Серия Всероссийских конференций с международным участием «*Электронные библиотеки: перспективные методы и технологии, электронные коллекции*» (Russian Conference on Digital Libraries — RCDL) направлена на формирование сообщества специалистов России, ведущих исследования и разработки в области электронных библиотек и близких областях. Для такого сообщества конференции RCDL предоставляют возможность обмена накопленным опытом, идеями и полученными результатами, а также установления контактов для более тесного сотрудничества. Конференции RCDL проводятся начиная с 1999 г. В конференциях RCDL приняли участие многочисленные российские и зарубежные специалисты из различных научных областей — не только представители университетов и исследовательских центров, но также разработчики и пользователи электронных библиотек, студенты и аспиранты. Конференция RCDL активно сотрудничает с близкими по тематике зарубежными конференциями ECDL (European Conference on Digital Libraries) и JCDL (Joint Conference on Digital Libraries), что способствует развитию международного сотрудничества в области электронных библиотек. Конференция активно поддерживается Российским фондом фундаментальных исследований (РФФИ) и Московской секцией ACM SIGMOD (Special Interest Group on Management Of Data).

В 2011 г. конференция RCDL проводилась в г. Воронеже с 19 по 22 октября. Среди организаторов конференции — Российская академия наук, РФФИ, Воронежский государственный университет, Институт проблем информатики РАН, Московская секция ACM SIGMOD.

Программный комитет конференции RCDL-2011 рекомендовал к публикации в журнале «Информатика и её применения» пять статей, являющихся развернутыми вариантами докладов, представленных на конференции:

- О. П. Желенкова «Исследование радиоисточников средствами виртуальной обсерватории»;
- Ю. А. Загорулько, О. И. Боровикова, И. С. Кононенко, Е. Г. Соколова «Методологические аспекты разработки электронного русско-английского тезауруса по компьютерной лингвистике»;
- М. Р. Когаловский, С. И. Паринов «Классификация и использование семантических связей между информационными объектами в научных электронных библиотеках»;
- Е. В. Шарапова, Р. В. Шарапов «Универсальная система проверки текстов на плагиат «АВТОР.NET»;
- Д. М. Скачков, О. Л. Жижимов «Об интеграции географических метаданных посредством ретроспективного тезауруса».

В число авторов этих статей входят ученые из организаций Российской академии наук (Специальная астрофизическая обсерватория РАН, Институт систем информатики им. А. П. Ершова СО РАН, Институт проблем рынка РАН, Центральный экономико-математический институт РАН, Институт вычислительных технологий СО РАН), а также из Российского государственного гуманитарного университета и Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых.

Международный семинар по проблемам устойчивости стохастических моделей имеет давние традиции. Он был основан профессором В. М. Золотарёвым в 1970 г.

В семинарах принимали и принимают участие ведущие специалисты в области теории вероятностей и ее приложений, представляющие разные страны мира. В XXI в. семинары проходили в Венгрии, Болгарии, Испании, Латвии, Италии, Израиле, Румынии и Польше.

В 2011 г. XXIX Международный семинар по проблемам устойчивости стохастических моделей проводился в г. Светлогорске Калининградской области России с 10 по 16 октября. Организаторы семинара: Московский государственный университет им. М. В. Ломоносова и Институт проблем информатики РАН.

Программный комитет XXIX Международного семинара по проблемам устойчивости стохастических моделей рекомендовал к публикации в журнале «Информатика и её применения» девять статей, являющихся развернутыми вариантами или дальнейшим развитием докладов, представленных на конференции:

- С. De Nikola, Y. S. Khokhlov, M. Pagano, and O. I. Sidorova “Fractional Levy motion with dependent increments and its application to network traffic modeling”;
- O. Yanushkevichiene and R. Yanushkevichius “About the rate of convergence of one U-statistic”;
- В. Е. Бенинг, Л. М. Закс, В. Ю. Королев «Оценки скорости сходимости распределений случайных сумм к дисперсионным гамма-распределениям»;
- Ю. В. Гайдамака, Т. В. Ефимушкина, А. К. Самуйлов, К. Е. Самуйлов «Задачи оптимального планирования межуровневого интерфейса в беспроводных сетях»*;
- О. В. Лукашенко, Е. В. Морозов «Асимптотика максимума процесса нагрузки для некоторого класса»*;
- Е. В. Морозов, Р. С. Некрасова «Об оценивании вероятности переполнения конечного буфера в регенеративных системах обслуживания»*;
- Е. В. Морозов, А. С. Румянцев «Вероятностные модели многопроцессорных систем: стационарность и моментные свойства»*;
- А. В. Печинкин, И. А. Соколов «Ограничение на суммарный объем заявок в дискретной системе Geo/G/1/∞»*;
- А. В. Ушаков «Анализ системы обслуживания с гиперэкспоненциальным входящим потоком в условиях критической загрузки»*.

В число авторов этих статей входят исследователи из организаций Российской академии наук (Институт проблем информатики РАН, Институт проблем рынка РАН, Центральный экономико-математический институт РАН, Институт вычислительных технологий СО РАН), представители факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, Российского университета дружбы народов, Тверского государственного университета, отдела моделирования и математической статистики Альфа-банка (все — Россия), университетов гг. Салерно и Пиза (оба — Италия), Института математики и информатики Вильнюсского университета и Литовского университета эдукологии (оба — Литва).

Статьи, отмеченные звездочкой (*), являются расширенными вариантами или научным развитием докладов, представленных на осенней сессии V Международного семинара «Прикладные задачи теории вероятностей и математической статистики, связанные с моделированием информационных систем», проведенной в качестве сопутствующего мероприятия в рамках XXIX Международного семинара по проблемам устойчивости стохастических моделей.

Организатором Международного семинара «Прикладные задачи теории вероятностей и математической статистики, связанные с моделированием информационных систем», проводящегося с 2006 г., является Институт проблем информатики РАН.

Статьи, являющиеся развитием докладов, представленных
на конференции «Электронные библиотеки:
перспективные методы и технологии, электронные коллекции»
(RCDL'2011, г. Воронеж, 19–22 октября 2011 г.)

ИССЛЕДОВАНИЕ РАДИОИСТОЧНИКОВ СРЕДСТВАМИ ВИРТУАЛЬНОЙ ОБСЕРВАТОРИИ*

О. П. Желенкова¹

Аннотация: В течение ряда лет с использованием разных подходов на базе средств виртуальной обсерватории в САО РАН проводились исследования радиоисточников обзоров, выполненных на крупнейшем российском радиотелескопе РАТАН-600 в 1980–1999 гг. Проведено их массовое отождествление с максимальным использованием имеющихся в открытом доступе данных разных диапазонов электромагнитного спектра. С применением программного инструментария виртуальной обсерватории реализован подход по автоматической подготовке и предварительной обработке данных. Для полученного компилятивного каталога разработана информационно-поисковая система, которая применялась при анализе информации о каждом источнике и принятии решения об отождествлении. Исходя из полученного опыта при работе с многочисленными разнородными ресурсами, можно подытожить, что программные средства виртуальной обсерватории обеспечивают удобный доступ к астрономическим данным и существенно повышают эффективность научных исследований. Однако все еще нет развитого инструментария для дальнейшего анализа, актуализации и публикации собранных исследователем данных. Рядом проектов ведутся разработки по реализации большей связности данных на базе уже существующих веб-технологий, что переведет сервисы виртуальной обсерватории на новый уровень, обеспечивающий обмен знаниями посредством аннотирования записей каталогов и реализацией связей между ними.

Ключевые слова: виртуальная обсерватория; распределенные информационные системы; информационные технологии в научных исследованиях; интеграция неоднородных информационных ресурсов; базы данных

1 Введение

Сегодня виртуальная обсерватория — прежде всего средство удобного и эффективного доступа к разнообразным астрономическим данным. Первые шаги в сторону организации этой распределенной инфраструктуры были сделаны в 1990-х гг., когда в США была создана сеть центров данных для поддержки информации, полученной космическими миссиями НАСА. Достижения в области информационных технологий обеспечили основу, на которой распределенные коллекции данных стали рассматриваться как интегрированная информационная система. Виртуальная обсерватория открыла новые направления научных исследований, опирающиеся на статистический анализ, поиск новых закономерностей и объединение данных разных диапазонов.

Обычной практикой в астрономии были отдельные и/или повторные наблюдения индивидуальных объектов, что хорошо работало при открытии фундаментальных законов. Но по мере того как понимание разных астрономических феноменов и закономерностей становилось более точным, число вопросов, на которые можно ответить с помощью одного наблюдения, становилось все меньшим. Методы наблюдений в настоящее время все больше смещаются в сторону накопления больших объемов данных, а подход к решению астрофизических задач — к статистическим методам. Большой объем данных может привести к обнаружению процессов, чьи наблюдательные проявления маскируются разными эффектами, которые трудно бывает учесть из-за недостаточного количества данных. Наблюдательное время самых мощных

* Работа поддержана грантом РФФИ № 10-07-00412.

¹ Специальная астрофизическая обсерватория РАН, zhe@sao.ru

инструментов было, остается и будет оставаться весьма ограниченным, поэтому многие астрофизические вопросы, для решения которых требуется большое количество наблюдательных данных, часто просто не могут рассматриваться.

Взрывообразный рост объема и сложности данных вызван прогрессом в получении цифровых изображений (основной источник данных в астрономии), способов обработки, хранения и доступа к информации. В астрономии происходит сдвиг в сторону науки, базирующейся на обзорах, которые становятся все более важным методом в исследовании Вселенной. Сейчас имеются средства для проведения обзоров практически во всем диапазоне электромагнитного спектра, пространственных масштабов и временных эпох. Каталоги, получаемые из панхроматических обзоров, дают возможности для обнаружения новых явлений, которые могут фундаментально изменить наши представления о физике звезд и их эволюции, ближнем космосе и планетных системах, формировании галактик и природе активных ядер галактик. Например, изображение одного и того же участка неба в оптическом и радиодиапазоне привело к открытию квазаров в 1960-е гг., а данные в инфракрасном диапазоне позволили исследовать скрытые от наблюдателя пылью области звездообразования и активные ядра галактик, чего невозможно сделать по оптическим изображениям.

Повторные наблюдения областей неба привели к открытию транзиентных событий — сверхновых и более редких явлений — микролинзирования. Панхроматические наборы данных позволяют сравнивать теоретические модели и реальные данные. Такие исследования предъявляют определенные требования как к постановке задачи, так и к методам их решения, которые все больше опираются на информационные технологии и в первую очередь на веб-сервисы, системы управления базами данных, грид и облачные вычисления.

Исследования радиогалактик важны для понимания механизмов излучения внегалактических объектов в радиодиапазоне и относятся к широкому кругу задач, связанному с изучением проявлений активности галактических ядер. Феномен мощного радиоизлучения ядра галактики является кратковременной по космологическим масштабам (до $\sim 10^8$ лет) эволюционной фазой самых массивных звездных систем. Хотя мощные радиогалактики — редкие объекты (пространственная плотность оценивается в $\sim 10^{-6}$ Мпс $^{-3}$), их наблюдение в радиодиапазоне возможно практически на любых космологических расстояниях, что используется при изучении крупномасштабной структуры Вселенной, проверке гипотез формирования самых

первых звездных систем и решении других космологических задач.

Отождествление радиоисточников с объектами в других диапазонах — обязательная процедура при многочастотных исследованиях, и не такая простая, как это кажется на первый взгляд. Кросс-идентификация по координатам (около объекта в области с заданным радиусом ищется объект другого каталога) оптических и радиокаталогов при разном угловом разрешении, предельной чувствительности и координатной точности последних, а также морфологической структуре самих источников дает от 5% до 30% совпадений. Отметим, что только по оптическим данным можно оценить расстояние до родительских галактик радиоисточников, которые по большей части — весьма слабые оптические объекты, особенно на больших красных смещениях. Их оптические наблюдения требуют больших затрат наблюдательного времени на крупнейших оптических телескопах при наилучших погодных условиях.

В качестве примера приведем результаты исследований выборки источников с крутыми спектрами (SS, Steep Spectra) из каталога RC (RATAN Cold), полученного по материалам глубокого обзора полосы неба на радиотелескопе РАТАН-600 [1–3]. Для кандидатов в выборку учитывались угловые размеры, морфологическая структура, а также яркость объекта в радиодиапазоне. Так из ~ 1000 источников каталога RC в выборку с крутыми спектрами вошли ~ 100 объектов. Потребовалось 15 лет фотометрических и спектральных наблюдений на 6-м оптическом телескопе БТА (Большой телескоп азимутальный), чтобы отождествить и получить спектры для объектов выборки [4]. Из этих объектов у четырех источников оказалось $1 \leq Z < 2$, у трех $2 \leq Z < 3$, у одного радиоисточника $3 \leq Z < 4$ и самый далекий объект выборки с $Z = 4.51$.

Не только поиск далеких радиогалактик, но и статистические свойства радиоисточников в разных диапазонах электромагнитного спектра важны для понимания природы активных галактических ядер. Массовое исследование радиоисточников позволяет уточнять существующие и открывать новые селекционные критерии, которые можно использовать при классификации этих объектов. И такие исследования проводятся с привлечением современных цифровых обзоров в разных диапазонах.

На крупнейшем российском радиотелескопе РАТАН-600 в 1980–1999 гг. была проведена серия глубоких обзоров полосы неба шириной около 40 угловых минут. По данным этих обзоров получен каталог RC, а затем RCR (RATAN Cold Revised) [5]. С появлением глубоких цифровых обзоров в оптическом и инфракрасном диапазо-

не, таких как SDSS (Sloan Digital Sky Survey) [6] и UKIDSS (United Kingdom Infra-red Deep Sky Survey) [7], появилась возможность провести отождествление этих каталогов. Для выполнения этой задачи были максимально использованы все имеющиеся в открытом доступе данные — два оптических обзора: DSS2 (Digital Sky Survey) и SDSS (полосы u, g, r, i, z), включая каталоги GSC (Guide Star Catalog) [8] и USNO-B1 [9], обзоры ближнего инфракрасного диапазона 2MASS (Two Micron All Sky Survey) [10] и UKIDSS (полосы J, H, K), а также проведены многочастотные исследования радиоисточников по радиообзорам VLSS (VLA Low-frequency Sky Survey, 74 МГц) [11], TXS (Texas Survey of radio sources, 365 МГц) [12], NVSS (NRAO-VLA Sky Survey, 1,4 ГГц) [13], FIRST (Faint Images of the Radio Sky at Twenty centimeters, 1,4 ГГц) [14], GB6 (Green Bank survey, 4,85 ГГц) [15]. Радиоисточники каталога RC [16–18], а затем RCR [19, 20] были отождествлены с данными этих обзоров. Радиоисточники, у которых не обнаруживались кандидаты в оптических/инфракрасных каталогах, дополнительно отождествлялись с суммарными изображениями обзора SDSS в трех фильтрах (g, r, i) и/или инфракрасного обзора UKIDSS в фильтрах (J, H, K) для достижения более глубокого предела кадров.

Для многочастотного исследования выборки источников каталога RC, а затем каталога RCR разработана методика детального отождествления радиоисточников, включающая подбор информационных ресурсов, автоматическую подготовку данных из выбранных ресурсов для каждого источника, морфологическую классификацию, визуальную инспекцию подготовленных данных для принятия решения об отождествлении [21, 22]. Средствами интерактивного атласа неба Aladin [23] (программного интерфейса Perl для командного интерфейса и макроконтроллера), а также с помощью программного интерфейса Python к SAOImage DS9 [24] реализованы потоки работ по списку радиоисточников для подготовки данных и визуализации результатов. Для полученного компилятивного каталога разработана информационно-поисковая система, которая использовалась при отождествлении радиоисточников.

2 Программные средства виртуальной обсерватории

Сейчас активно развиваются программные средства (протоколы, метаданные и программы, функционирующие на их основе), учитывающие особенности информации, относящейся к определенной сфере человеческой деятельности, в част-

ности к научным исследованиям. Такие разработки в астрономии объединены в виртуальную организацию, которая носит название виртуальной обсерватории и координируется международным альянсом IVOA (International Virtual Observatory Alliance) [25–27]. В IVOA объединены виртуальные обсерватории разных стран, включая и Российскую виртуальную обсерваторию [28]. Рабочими группами IVOA ведется разработка стандартов более чем по десятку направлений, среди которых: представление и формализация данных и знаний предметной области, разработка протоколов доступа к данным, стандарты программных сервисов для распределенных вычислений, протокол обмена сообщениями для программных клиентских приложений виртуальной обсерватории, описание и публикация ресурсов, формат обмена данными, язык запросов, поддержка сохранности данных и пр. С момента появления в 2003 г. альянса IVOA разработано около полусотни спецификаций протоколов, форматов и соглашений, используемых при создании программных продуктов виртуальной обсерватории. Подробный обзор стандартов виртуальной обсерватории и применяемых технологий можно найти в обзоре Брюхова и др. [28].

2.1 Текущий статус

Инфраструктура виртуальной обсерватории является сервисно-ориентированной. Веб-сервисы IVOA разделены на три класса: обнаружение и публикация ресурсов, передача данных и организация запросов, а также сервисы для распределенных вычислений. Обнаружение данных выполняется через регистры виртуальной обсерватории. Для сервисов, предоставляемых регистрами виртуальной обсерватории, и спецификаций описания ресурсов были рассмотрены несколько промышленных стандартов, обеспечивающих механизмы обмена метаданными в Интернете, и был выбран протокол OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) [29]. Для описания астрономических ресурсов (каталогов, цифровых обзоров, баз данных, архивов наблюдений, программных средств, функционирующих как веб-сервисы) в регистрах используется стандарт описания сетевых ресурсов Dublin Core [30]. Веб-сервисы ориентированы на то, чтобы операции над данными в сети выполнялись без участия человека. Повторное использование простых сервисов и комбинирование их для выполнения более сложных действий реализует поток работ. Поток выполнения использует принципы интероперабельности, когда компоненты потока работ взаимодействуют друг с другом посредством протоколов, определяющих правила

запуска сервиса и структуру входных и выходных данных. Реализация таких протоколов опирается на модели данных.

Сервисы данных, кроме стандартных графических форматов (gif, jpeg), работают с двумя астрономическими форматами — FITS (Flexible Image Transport System) [31], который является с 1982 г. астрономическим стандартом для хранения и обмена данными, и VOTable [32]. VOTable-формат используется в сервисах ВО для представления результатов запросов. Основой VOTable является индустриальный стандарт XML и опыт разработок астрономических форматов FITS и CDS Astrores.

Астрономы для обозначения одних и тех же физических величин и параметров используют разные названия. Чтобы избежать неоднозначности при интерпретации величины, необходимо определить, что именно обозначают разные идентификаторы. В VOTable-формате используется семантический описатель UCD (Uniform Content Descriptor) [33], который устанавливает смысловую связь между обозначениями величин и астрономическими понятиями и/или физическими величинами. IVOA поддерживает и контролирует словарь дескрипторов.

Доступ к данным DAL (Data Access Layer) [34] включает стандарты, описывающие механизм доступа к распределенным астрономическим данным, и программные средства, обеспечивающие такой доступ. Для реализации запросов используется расширенное подмножество SQL — ADQL (Astronomical Data Query Language) [35]. Хотя SQL можно использовать для запросов к большинству современных астрономических баз данных, астрономическая специфика требует расширения возможностей языка. ADQL, кроме координатных запросов, поддерживает доступ по протоколам ВО к таблицам, изображениям и спектрам.

Так в общих чертах можно описать основы виртуальной обсерватории, которая за десять с небольшим лет превратилась в действующую инфраструктуру, и для астрономов сейчас нет проблем доступа к данным цифровых обзоров неба, архивам наблюдательных данных и каталогам. Есть удобные клиентские приложения и программные интерфейсы к ним, веб-интерфейсы к основным базам астрономических данных для поиска информации по одиночному объекту или списку объектов, запросов по списку объектов и визуализации полученной информации.

2.2 Новые задачи

На передний план выходит следующая цель виртуальной обсерватории — обеспечить профессио-

нальных астрономов возможностью получать информацию о небесных объектах. Сейчас это еще в слабой степени решается средствами виртуальной обсерватории, так как данные в основном не структурированы и не являются связанной по смысловому содержанию информацией. По этой причине невозможно выполнить запрос типа «найти все источники в каталогах, которые являются квазарами» и т. п. Это ограничивает пользователя эффективную работу с информацией. Нет прямого способа воспользоваться знаниями об объекте, полученными другими исследователями.

Вся имеющаяся у астрономического сообщества информация о небесных объектах собрана в каталогах и используется для статистических исследований объектов, подбора интересующих объектов, для поиска аномальных объектов или объектов с особой комбинацией свойств. Каталоги публикуются различными способами: от таблиц в журналах до публикации сервисами Vizier [36]. Самые крупные каталоги доступны через специализированные веб-интерфейсы для архивов и центров данных, таких как WFCAM [37], SDSS, IRSA [38], SkyView, MAST [39] и др. Есть следующие проблемы при работе с каталогами:

- в каталогах содержатся измеренные величины, а результаты их анализа и интерпретации публикуются обычно в журнальных статьях. Через гиперссылки, предоставляемые информационной системой ADS (Astrophysics Data System), в которой находится большая часть астрономических полнотекстовых статей, журнальная статья может указывать на каталог, используемый в ней, но в архивах данных не всегда реализованы аналогичные указатели на литературу;
- небесные объекты не имеют уникальных идентификаторов. Объединение информации в различных диапазонах электромагнитного спектра требует операций кросс-идентификации, при которой строки из двух каталогов, содержащих различную, но при связывании вдвойне полезную информацию, объединяются на основе перекрытия координатных положений с допусками, учитывающими ошибки сравниваемых каталогов. Кросс-идентификация проводится пользователями многократно, поскольку эта важная связь между объектами каталогов не сохраняется;
- каталоги являются статическими объектами. Если создается новый каталог, полученный из одного или нескольких существующих каталогов, но с добавлением какой-то новой информации, то для него, как правило, не отслеживаются родительские каталоги;

- поиск данных в каталогах может быть трудоемким, а объединение данных из двух каталогов трудоемко и неудобно. Сложно работать с компилятивными каталогами, полученными на базе объединения нескольких каталогов разных диапазонов, даже если эта информация собирается по небольшому списку объектов;
- если появляются новые релизы каталогов или новые каталоги, то исследователю надо заново выполнять одни и те же запросы для интересующих его объектов. Нет средств для оповещения о появлении новой информации и обновления данных пользователя.

2.3 Новый качественный уровень виртуальной обсерватории

В последнее время появилось несколько проектов, направленных на дальнейшее развитие инфраструктуры виртуальной обсерватории и, в частности, на решение проблем, которые возникают при работе с каталогами. Далее приведем сведения о тех из них, которые, вероятнее всего, могут помочь в работе с разнородными данными, полученными при массовом отождествлении радиоисточников. В этих проектах ведутся разработки, связанные с внедрением инновационных информационных технологий (грид, облачные вычисления, Семантический Веб) в астрофизические исследования.

Цель проекта AstroDAbis [40] — создание независимого механизма публикации пояснений (комментариев, аннотаций). Пояснения могут создаваться пользователем для одиночного объекта («объект X есть квазар») или для нескольких объектов («объект с номером 123 в каталоге A есть то же самое, что объект с номером 456 в каталоге B»). Как полагают авторы AstroDAbis, этим решаются проблемы передачи знаний, создания компилятивных каталогов и реализации их связи с родительскими каталогами. Авторы статей, где представлена информация, полученная на основе анализа каталогов, с помощью аннотаций могут передать знания о небесном объекте в форме, которая может быть использована в последующих запросах к каталогу. Когда возникнет потребность объединить два каталога и создать компилятивный каталог (например, слияние оптических данных SDSS и инфракрасных данных тех же источников из UKIDSS), такая связь позволит обойтись без повторной кросс-идентификации ресурсов. С помощью аннотации такие каталоги сохраняют связи с исходными каталогами, и связи будут однозначно зафиксированы.

Результатом проекта AstroDAbis является прототип сервиса, который, в общем-то, является кросс-

идентификацией нескольких существующих каталогов. Однако он вносит новое качество в работу с имеющимися ресурсами. Целевая аудитория этого проекта — в первую очередь астрономы, которые являются довольно-таки небольшим по количеству сообществом, но взаимосвязанную информацию в базах данных можно предоставить и более широкой публике посредством API-сервисов, что облегчит будущим разработчикам создание удобной системы поиска информации об астрономических объектах для любой группы пользователей.

Аналогичные разработки не являются новыми в науке (аннотирование данных в генетике — Distributed Annotation System, <http://www.biodas.org>) или в Интернете — RDF (Resource Description Framework) [41] и LOD (Linking Open Data) [42]. Система AstroDAbis разработана так, чтобы естественным образом использовать TAP-factory [43] на базе OGSA-DAI (Open Grid Services Architecture Data Access Interface) [44], где TAP (Table Access Protocol) [45] — протокол IVOA для работы с таблицами. Используя TAP-factory, можно создать сервис, который позволит выполнять запросы, обращающиеся к другим сервисам, совместимым по протоколу TAP. AstroDAbis также имеет LOD-интерфейс, который обеспечивает создание URI для аннотируемых объектов, что подготавливает платформу для будущих экспериментов с Семантическим Вебом в астрономии.

Чтобы найти и получить данные пользователь сам инициирует взаимодействие с инфраструктурой виртуальной обсерватории посредством клиентских приложений (TOPCAT [46], ALADIN, DS9 SAOImage и др.) или веб-интерфейсов к базам данных. Всякий раз, когда пользователь хочет узнать о возможно уже появившихся обновлениях, ему надо повторить первоначальный запрос, сравнить полученный результат с существующим и скопировать, если это требуется, данные. Постоянно растущие объемы данных, включающие новые релизы существующих обзоров, и публикации новых каталогов требуют другого подхода при отслеживании новой информации о небесных объектах, интересной пользователю. Особенно это полезно при обновлении и актуализации компилятивных каталогов и баз данных. Решение этой задачи предлагается с помощью веб-приложения для поддержки данных пользователя Vodka (VO Data Keeping-up Agent) [47], который ретранслирует запросы пользователей в инфраструктуру виртуальной обсерватории и рассылает уведомления об обновлениях. При выбранном пользователем темпе опроса агент асинхронно посылает один и тот же запрос, сформулированный пользователем, и фиксирует результа-

ты, отражающие временной срез информации, выполняет сравнение этих срезов и оповещает пользователя по электронной почте. У пользователя есть возможность просматривать результаты запросов, сохраненные в *snapshot*-файлах, журналы сравнения этих файлов, копировать снимки и новые появившиеся данные, а также инкрементальные файлы, включающие старые, новые и пропущенные данные.

Во многих областях научных исследований имеется насущная потребность работы с большими по объему распределенными массивами данных и выполнения над ними разнообразных задач по извлечению знаний. Итало-американский проект DAME (DAta Mining & Exploration) [48] направлен на создание междисциплинарной распределенной среды, специализированной под исследования больших массивов данных (MDS, Massive Data Set) с помощью машинно-обучаемых алгоритмов и методов добычи данных, которая реализована на унифицированной технологической платформе. DAME включает несколько проектов по решению разных астрофизических задач и может предложить для разных e-science сообществ широкий спектр вычислительных мощностей для применения машинно-обучаемых и статистических алгоритмов к астрономическим данным. Эти проекты используют единую технологическую платформу, базирующуюся на архитектуре сервисно-ориентированных приложений и совместимую со стандартами виртуальной обсерватории.

Обнаружение знаний в базах данных KDD (Knowledge Discovery in Data Bases) сейчас связывают с новым семейством научных дисциплин, называемым X-Informatics. Оно считается четвертой парадигмой в науке после теории, эксперимента и моделирования. В таком контексте проект DAME призван:

- обеспечить сообщество расширяемой интегрированной средой для добычи данных и исследований на базе технологий Web 2.0;
- поддерживать стандарты и форматы виртуальной обсерватории для интероперабельности приложений;
- обеспечить виртуальную обсерваторию общей вычислительной платформой, использующей современные технологии (грид, облачные вычисления и т. п.).

При происходящем в настоящее время росте сложности данных и необходимости проведения исследований с большими массивами данных альянсом IVOA было принято решение о создании группы по интересам, связанной с обнаружением знаний в базах данных (KDD-IG), которая должна

согласовывать стандарты IVOA и потребности научных исследований с использованием баз данных.

В этом проекте находятся в стадии разработки несколько научных сценариев, которые оформляются в виде веб-приложений, базирующихся на архитектуре системы DAME. Из них для исследования интересующей автора выборки радиоисточников наиболее привлекательны разработки по оценке фотометрического красного смещения галактик и селекции квазаров на основе фотометрических данных, поскольку определение спектральных красных смещений требуют больших затрат наблюдательного времени в отличие от фотометрических оценок.

Разработки проекта ADSASS (The ADS All-Sky Survey) [49] направлены на превращение системы NASA ADS (Astrophysics Data System), широко используемой среди астрономов в качестве полнотекстового библиографического ресурса, в карту неба. Система ADS не является источником наблюдательных данных, но является неявным хранилищем ценной астрономической информации в форме изображений, таблиц и ссылок на небесные объекты, которые являются частью публикации. Необходимо сделать эту информацию доступной для запросов и просмотра. Рассматриваются три категории данных:

- (1) ссылки на небесные объекты, которые предполагается собрать из внешних баз данных и добавить в виде аннотации (*astrotag*) связь со статьями в ADS. Так же, как это сделано в *geotags* для объектов на земной поверхности, *astrotags* являются пространственными и временными аннотациями для небесных объектов;
- (2) оптические и изображения в других диапазонах, имеющиеся в статьях, также получают связывающие ссылки (*astroreference*). Так это сделано для геоданных (*georeferencing*), которые ссылаются на карты, имеющие привязку к системе земных координат, ссылки (*astroreferencing*) свяжут изображения, которые будут приведены к одной небесной координатной системе с учетом ориентации, координатной привязки и масштаба пикселей каждого кадра;
- (3) другого сорта данные, такие как текст или подписи под рисунками, будут привязаны к координатам или имени источника.

В результате выполнения проекта будет получена карта всего неба, которая будет активировать ссылки на статьи, показывая, какая часть неба в них описывается, а также слой исторических данных на базе хранилища *astroreference*-ссылок и изображений, извлеченных из статей, которые можно

использовать для анализа. Для визуализации этой информации будут использоваться приложения, в которых можно отображать полностью все небо, а именно: WorldWide Telescope (Microsoft), ALADIN (CDS), Google Sky (Google) и др. Система AD-SASS будет опираться на постоянно обновляемую базу данных тегов, которая предназначена как для обнаружения новой информации о небесных объектах по любой тематике, так и для поиска событий переменного характера по данным исторического слоя.

3 Массовое отождествление радиоисточников

3.1 Научная мотивация

В противоположность начальной стадии своего возникновения Вселенная сегодня богата структурами — галактиками, скоплениями галактик, сверхскоплениями и пустотами — войдами. Все эти структуры эволюционируют с гравитационным расширением из небольших первоначальных неоднородностей плотности. В иерархических космогониях первые гравитационно-связанные системы могли быть звездами и/или небольшими звездоформирующими системами, при слиянии которых формируются галактики. Возникающие из конечных продуктов звездной эволюции и мерджинга (слияния) центральные черные дыры продолжают расти. В любом случае аккреция, питающая массивные черные дыры, проявляет себя как феномен активного галактического ядра (Active Galaxy Nuclear, AGN). Из-за своей экстремальной светимости AGN являются подходящими реперами для исследования Вселенной. Хотя почти все AGN имеют схожие источники энергии, их наблюдаемые свойства сильно различаются. К примеру, одни AGN имеют мощное радиоизлучение, а другие нет. Могут наблюдаться еще разные проявления активности ядра — широкие эмиссионные линии в оптике, высокая степень поляризации оптического излучения, переменность, рентгеновское или гамма-излучение. В моделях, объясняющих наблюдаемые свойства активных галактик, предполагается, что только несколько физических процессов обеспечивают наблюдаемый диапазон AGN-характеристик. Полагают, что разнообразие типов AGN возникает из: (1) отсутствия или наличия пыли вокруг ядра; (2) направления релятивистского джета относительно наблюдателя; (3) цикла активности; (4) полной светимости галактики. Еще разделение по мощности радиоизлучения зависит от углового момента и массы центральной черной дыры.

Популяцию мощных радиогалактик с помощью имеющихся радиотелескопов можно наблюдать практически на любом расстоянии. Это позволяет изучать их эволюцию в радиодиапазоне от момента образования до наших дней. От низких до умеренных красных смещений ($Z \sim 1$) мощные радиоисточники связывают с гигантскими эллиптическими галактиками, поэтому радиогалактики можно использовать для изучения формирования и эволюции самых массивных звездных систем, из истории звездообразования которых можно получить важные ограничения на модели формирования галактик и космологические параметры. Радиоисточники часто ассоциируются с центральными галактиками скоплений, поэтому далекие радиогалактики могут быть индикаторами первых протоскоплений. Так выглядят в общих чертах те задачи, которые решаются при исследовании радиогалактик.

3.2 Предметный посредник для поиска далеких радиогалактик

Известны разные техники селекции объектов для поиска далеких галактик. К ним относятся: глубокая спектроскопия пустых полей, узкополосные снимки, спектроскопия объектов вокруг радиогалактик с большим Z , использование показателей цвета (разность звездных величин объекта в различных фотометрических фильтрах/полосах) для выбора кандидатов. Известно, что спектральное распределение энергии небесных объектов в зависимости от красного смещения сдвигается в красную область спектра. Из-за этого галактика может быть ярче или существенно слабее в каком-либо фильтре, быть видимой только в одном фильтре из-за того, что в эту полосу попадают особенности спектрального распределения объекта — излучение в водородной линии Лайман-альфа $\lambda = 1216 \text{ \AA}$ или завал спектра на $\lambda = 912 \text{ \AA}$. Чтобы поймать это усиление или, наоборот, ослабление яркости, используют ограничения на цветовые индексы, например, следующие: $(u+r)/2 - g > 1$; $(g+i)/2 - r > 1$ и т. п., где u, g, r, i, z — звездные величины оптического объекта в полосах обзора SDSS. Это и есть так называемые dropout-техники, используемые при отборе кандидатов в далекие галактики. Затем для кандидатов проводят спектральные исследования, чтобы определить по смещению спектральных линий, действительно ли объект является далеким. Применение этих техник привело к обнаружению далеких объектов с $Z \sim 6-7$. Однако радиогалактики продолжают оставаться интересными для изучения объектами, так как они являются самыми массивными звездными системами, во многих

случаях указывают на скопления галактик, а на космологических расстояниях — на протоскопления.

С появлением оптического обзора неба SDSS и радиообзора FIRST, которые обладают надежной координатной привязкой, достаточной глубиной и угловым разрешением, а также программных средств виртуальной обсерватории исследования природы радиоисточников стало возможным проводить не только по небольшим по числу объектов выборкам, но и по любым каталогам/спискам.

Был предложен научный сценарий поиска далеких галактик по радиоисточникам каталога RC, который использует обзоры FIRST, NVSS и SDSS [50]. Для каждого RC-источника из обзора SDSS выбираются объекты, которые попадают в область, размерами равную боксу ошибок определения координат ($\pm 3\sigma$). При средней плотности объектов обзора SDSS (~ 7 – 8 объектов на кв. угл. минуту) в область поиска, размеры которой варьируются от $45''$ до $2'$ в зависимости от положения RC-источника относительно центральной части диаграммы направленности телескопа, попадают сотни объектов. Поскольку низкая точность координат каталога RC не позволяет выполнить отождествление по позиционному совпадению, можно провести дополнительную селекцию в оптике, а именно использовать ограничения для разности показателя цвета. Если оптический объект, попадающий по координатам в область поиска, еще и удовлетворяет ограничениям по показателям цвета, то он является наиболее вероятным кандидатом для отождествления.

Для сценария отождествления списка радиоисточников был создан предметный посредник [51], архитектура которого была реализована как объединение системы AstroGrid [52], которая разработана в Великобритании и совместима с протоколами IVOA, и средств поддержки предметных посредников, созданных в ИПИ РАН. Для прототипа гибридной архитектуры было выполнено сопряжение исполнительных механизмов двух инфраструктур (AstroGrid и предметных посредников). Эта разработка выполнялась для решения задач, связанных с разработкой прототипа Российской виртуальной обсерватории (РВО) [28].

Сценарий автоматического отождествления выборки каталога RC в области, пересекающейся с обзорами SDSS и FIRST, разделен на два этапа: подготовка данных и визуальная инспекция результатов для принятия решения об отождествлении радиоисточника с оптическим кандидатом.

Поток задач для подготовки данных состоит из следующей последовательности:

- (1) выбор координат радиоисточника из каталога RC;

- (2) извлечение списка объектов из области заданного размера из базы данных обзора SDSS;
- (3) кросс-идентификация результатов запросов с использованием селекционного ограничения по цветовым индексам;
- (4) извлечение радиоизображений из обзора FIRST;
- (5) извлечение оптических изображений SDSS;
- (6) суперпозиция изображений;
- (7) сохранение результатов запросов.

Чтобы можно было работать с каталогом RC из потока задач, он оформлен как компонент AstroGrid для доступа к данным — DSA (Data Set Access). Запрос к базе данных, содержащей каталог RC (шаг 1), выполняется CEA (Common Execution Architecture) приложением на узле AstroGrid, развернутом в ИПИ РАН. Результат запроса в формате VOTable автоматически запоминается в MySpace (виртуальная область памяти системы AstroGrid). Запрос на языке ADQL выглядит следующим образом:

```
SELECT crd.ra, crd.de, cat.name
FROM RCCatalog as cat, CoordEQJ as crd
WHERE cat.coord_id = crd.coord_id
```

Запрос к каталогу SDSS (шаг 2) выполняется CEA-приложением с помощью веб-сервиса, работающего на сервере SDSS (http://voservices.net/CasService/ws_v1_0/CasService.asmx). Оно запрашивает данные и возвращает результат в файл в формате VOTable, сохраняемый в MySpace. Запрос к SDSS записывается следующим образом:

```
SELECT ra=cast(ra as real),
dec=cast(dec as real),
objid, u, g, i, r, z,
colorIndexURG = (u+r)/2.0-g .GT. 1.
FROM PhotoPrimary
WHERE ra BETWEEN 225.0 AND 225.5 AND dec
BETWEEN 4.0 AND 5.61 AND r BETWEEN 15.0 AND 23.0
```

Кросс-идентификация (шаг 3) результатов двух предыдущих шагов выполняется веб-сервисом AstroGrid CrossMatchFull (ivo://org.astrogrid/CrossMatcher) в UK. Результатом служит таблица в VOTable формате, которая также помещается в MySpace.

Извлечение и суперпозиция изображений (шаги 4–6) производится в ИПИ РАН. Приложение CEA вызывает ALADIN для каждого объекта из каталога RC, используя его координаты в качестве центра области. ALADIN извлекает изображения из DSS и FIRST, после чего контуры радиоизображения совмещаются с оптическим. Дополнительно извлекаются объекты из каталогов SDSS, 2MASS,

FIRST, NVSS, попавшие в область поиска. Выполняющая эти запросы программа, написанная на языке команд ALADIN, показана ниже:

```
get DSS.ES0(DSS1,14.1,14.1), address, 5'; sync;
/* извлечь изображение из обзора DSS.
/* Переменная address содержит координаты
/* в формате 'hh:mm:ss sdd:mm:ss'
get NVSS(0.2,15.0,Stokes I, Sine), address, 5';
/* извлечь изображение из радиообзора NVSS
sync; contour 4;
/* синхронизировать и построить контуры
get FIRST(10), address, 5';
/* извлечь изображение из радиообзора FIRST
sync; contour 4;
get SDSSDR3cat, address, 1';
/* извлечь данные из каталога SDSS
sync;
get Vizier(2mass), address, 1';
/* извлечь данные из каталога 2MASS
sync;
backup st.aj
/* сохранить данные
```

После выполнения программы данные запоминаются (шаг 7) в стеке ALADIN и сохраняются в MySpace. Подготовка данных на этом закончена.

На втором этапе для просмотра подготовленных данных запускается ALADIN и Workbench (клиентское приложение для работы с AstroGrid). Данные, сохраненные в MySpace AstroGrid, открываются в ALADIN, проводится визуальная инспекция подготовленных данных и принимается решение об отождествлении объекта.

Подробное описание реализации этого сценария на основе AstroGrid и средств поддержки предметных посредников приведено в [51].

3.3 Оптическое отождествление каталога RC

Подходящих кандидатов в далекие галактики среди источников каталога RC не было обнаружено, но была разработана и опробована методика массового отождествления радиоисточников. Исследования были продолжены дальше, но уже с другой целью — отождествление всех источников каталога RC, попадающих в область обзоров SDSS и FIRST, и определение типов родительских галактик.

Сценарий массового отождествления радиоисточников включает следующие этапы: (1) подготовка данных; (2) предварительная обработка и визуальная инспекция; (3) анализ — уточнение координат RC-радиоисточников; (4) анализ — определение морфологических типов радиоисточников; (5) анализ — оптическое отождествление.

1. Подготовка данных. Этот этап выполняется так же, как в сценарии поиска далеких радиоисточников, но список используемых каталогов существенно расширен. Подготовка данных производится автоматически с помощью perl-программы, которая использует средства программного интерфейса ALADIN. Для каждого источника каталога RC по координатам извлекаются изображения из оптических [DSS-II (сервер Space Telescope Science Institute) и SDSS (сервер SkyView — the Internet's Virtual Telescope)] и радиообзоров неба [NVSS, FIRST (сервер National Radio Astronomy Observatory)], данные из оптических (USNO-B1, SDSS) и инфракрасных (2MASS) каталогов, а также из радиокаталогов [VLSS, TXS, NVSS, FIRST, GB6 и RC (сервер Vizier)], выполняется суперпозиция изображений посредством наложения контуров радиоизображения на оптическое, извлеченные данные сохраняются для последующих фаз сценария. Пример программы приведен ниже:

```
#!/usr/bin/perl
die "Usage: $myname Catalog\n" unless (@ARGV);
$CATALOG = $ARGV[0];
open (READ_CAT, "<$CATALOG")
or die ("Cannot open file");
/* открытие файла с координатами
open(ALADIN,"| java
-Dhttp.proxyHost=192.168.2.33 -
/* запуск ALADIN
Dhttp.proxyPort=8080
-jar /Data/users/zhe/Aladin/Aladin.jar");
$rcN = 0;
while ($str = readline (*READ_CAT))
{
($pref,$name,$last) = split (/ /, $str, 3);
/* преобразование координат
$rcname = join('',$pref,$name);
$raJ = substr($last,0,11);
$deJ = substr($last,12,11);
$obj = join('',$raJ,$deJ);
$stack = join('','s',$rcname,'.aj');
print ALADIN "reset; \n";
/* передача команд ALADIN
print ALADIN "get Vizier(VIII/42/txs)
$obj 5\';\n";
/* извлечь данные из каталога TXS
print ALADIN "get aladin(DSS2,F) $obj 5\';\n";
/* -- изображение из обзора DSS
print ALADIN "get NVSS(0.2,15.0,Stokes I,Sine)
$obj 5\';\n";
/* -- -- из радиообзора NVSS
print ALADIN "sync; contour 4; \n";
/* построить контуры радиоизображения
print ALADIN "get Vizier(VIII/65/nvss) $obj 5\';
sync;\n";
/* извлечь данные из каталога NVSS
print ALADIN "get Vizier(sdss) $obj 1\';\n";
```



```

\* -"- -"- из каталога SDSS
print ALADIN "sync \n";
print ALADIN "get Vizier(2mass) $obj 1\';\n";
\* -"- -"- из каталога 2MASS
print ALADIN "sync \n";
print ALADIN "get Vizier(USNOB) $obj 1\';\n";
\* -"- -"- из каталога USNO-B1
print ALADIN "sync \n";
print ALADIN "get FIRST(100) $obj 5\';\n";
\* извлечь изображение из обзора FIRST
print ALADIN "sync; contour 4; \n";
\* построить контуры радиоизображения
print ALADIN "get Vizier(VIII/71/first)
$obj 5\'; sync;\n";
\* извлечь данные из каталога FIRST
print ALADIN "get Vizier(J/A+AS/87/1/table1)
$obj 5\'; sync;\n";
\* -"- -"- из каталога RC
print ALADIN "backup $stack\n";
\* сохранение подготовленных данных
};

```

2. Предварительная обработка и визуальная инспекция. На этом этапе используется макроконтроллер ALADIN. Этот сервис производит интерпретацию двух файлов. Один файл содержит скрипт с командами ALADIN, второй — данные, которые являются параметрами команд. Предварительная обработка производится средствами графического интерфейса ALADIN (рис. 1, а). Она состоит в под-

боре уровней контуров и диапазона отображаемых данных для изображений.

3. Уточнение координат. Поскольку координаты каталога RC грубы для оптического отождествления, то требуется уточнение координат по радиообзорам с более высокой координатной точностью. Отметим, эта задача легче решается, если у сравниваемых каталогов близкое угловое разрешение, при этом необходимо учитывать изменение плотности потока источника на разных частотах, а также предельную чувствительность каждого каталога.

Идентификация RC-источников проводилась сначала с источниками обзора NVSS, поскольку каталог RC имеет близкое к NVSS угловое разрешение по прямому восхождению ($45''$), а координатная точность NVSS ($1''$) существенно лучше, чем у RC ($15'' \times 40''$).

Перечислим условия в порядке убывания значимости, выполнение которых принималось во внимание при отождествлении источника каталога RC с источником обзора NVSS:

- координатное совпадение по прямому восхождению ($r < 3\sigma$, где σ — приведенная в каталоге RC ошибка координат по прямому восхождению);
- координатное совпадение по склонению;
- совпадение плотностей потока для RC-источника и NVSS-источника. Вызывают сомнения

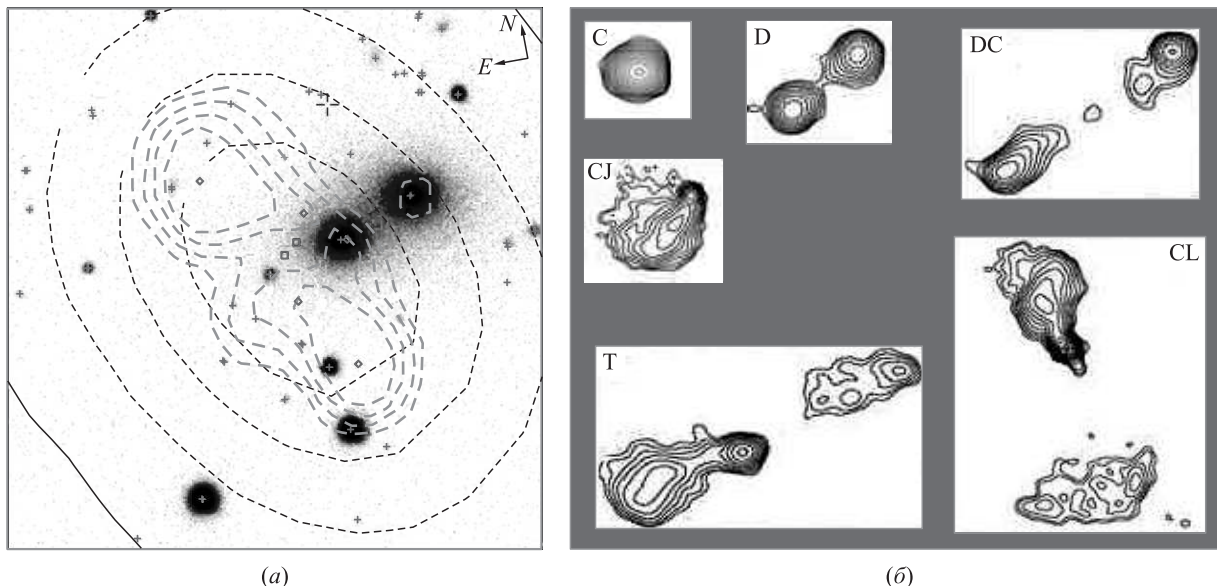


Рис. 1 (а) Рисунок, полученный средствами ALADIN. Полутоновое изображение — данные в полосе r из оптического цифрового обзора неба SDSS; штриховые контуры — контурная радиокарта из обзора NVSS (угловое разрешение — $45''$); сплошные контуры — радиокарта обзора FIRST с более высоким угловым разрешением ($5''$), которая позволяет определить детальную структуру радиоисточника; крестиками отмечены данные из каталога SDSS, ромбами — радиокаталоги. (б) Морфологические типы радиоисточников: C (core) — точечный; D и DC (double) — двойной; CJ (core-jet) — ядро с выбросом; CL (core-lobe) — ядро с компонентами; T (triple) — тройной

случаи, когда при координатном совпадении RC-источник не согласуется по плотности потока с NVSS (при пересчете плотностей потоков полагаем, что спектральный индекс источника $\alpha \sim 0.7$, $S(\nu) \sim \nu^{-\alpha}$);

- присутствие соседних источников. Если рядом с RC-источником есть не один, а два или несколько источников NVSS, которые попадают в диаграмму направленности РАТАН-600, то возникает неоднозначная ситуация при идентификации. В этом случае принималось, что наибольший вклад дает самый яркий NVSS-источник, с которым и отождествлялся RC-источник.
- когда плотность потока источника на 3,9 ГГц оказывается больше, чем плотность потока на 1,4 ГГц, требуется дополнительная информация, подтверждающая рост плотности потока к более высоким частотам. В этих случаях использовался как каталог, так и радиообзор GB6 на 4,85 ГГц. В каталог обычно включаются объекты с плотностью потока выше 5σ уровня отношения сигнал/шум. Источники с плотностью потока на уровне 3σ – 4σ , отсутствующие в каталоге GB6, обнаруживаются при визуальной инспекции изображений обзора GB6.

Эта дополнительная информация помогала при неоднозначных случаях отождествления.

По такому алгоритму [17] были отождествлены и уточнены координаты у 75% источников каталога RC, для которых на следующем шаге проводилось оптическое отождествление.

4. Определение морфологического типа радиоисточника. Корректность оптического отождествления радиоисточника зависит от правильного определения его морфологического типа, поскольку есть связь между типом и предполагаемым положением родительской галактики. Для этого использовался обзор FIRST, где из-за более высокого углового разрешения имеется более подробная информация о структуре источника.

Было использовано 5 морфологических типов радиоисточников (рис. 1, б): точечные (core), двойные (double, double-core, double-double), тройные (triple), ядро с джетом (core-jet), ядро с компонентами (core-lobe). Тип радиоисточника определялся по радиоизображениям (рис. 2, а) и дополнительно по картам с контурами интенсивности (рис. 2, б), если структуру сложно классифицировать.

5. Определение оптического кандидата. После определения морфологического типа радиоисточника определялось предполагаемое положение

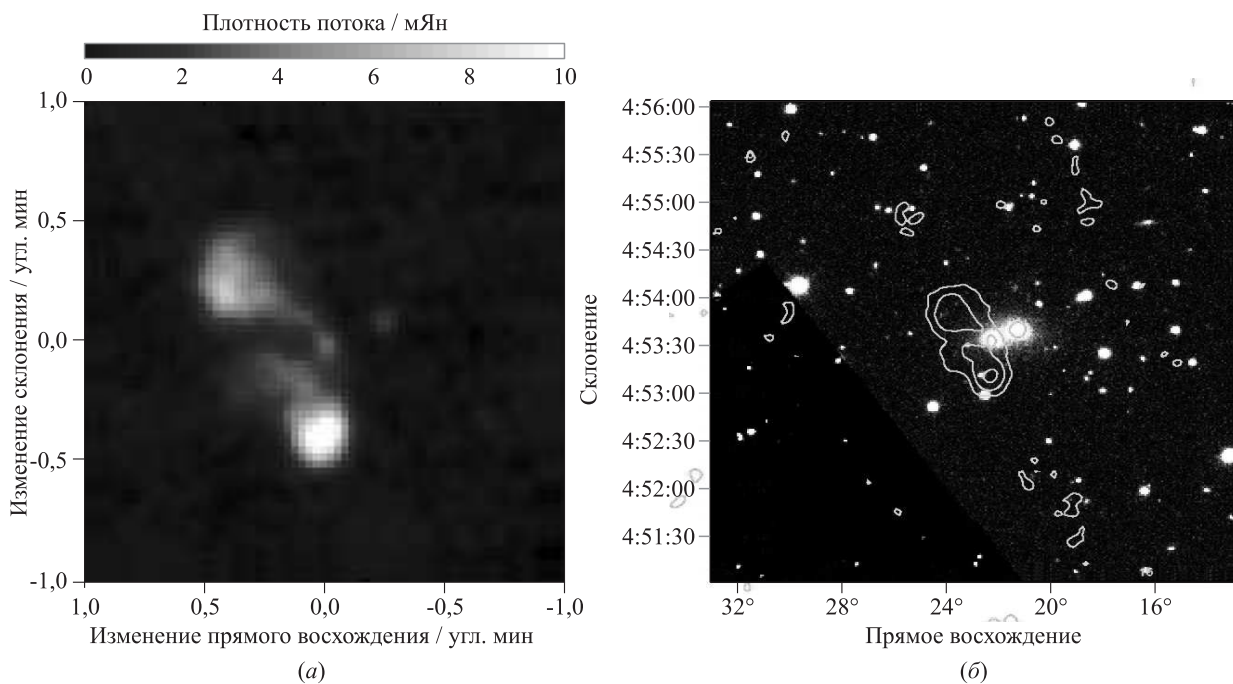


Рис. 2 (а) Изображение радиоисточника каталога RCR из обзора FIRST. (б) Составное изображение оптика—радио. Контур, построенные по изображению радиообзора FIRST, наложены на цветное RGB-изображение, полученное из трех кадров обзора SDSS в фотометрических полосах g , r , i , где фильтр g соответствует B , r — G и i — R . Рисунок получен с помощью python-скрипта, использующего программный интерфейс к приложению для визуализации и доступа к данным виртуальной обсерватории DS9 SAOImage

родительской галактики. Именно эти координаты затем использовались при выборе оптического кандидата. Оптический объект считался надежным кандидатом на отождествление, если его положение (по каталогу SDSS) было не дальше 3σ от предполагаемого положения родительской галактики, где σ — ошибка координат. К возможным отождествлениям отнесены следующие случаи:

- источник точечный или двойной с ядром, а оптический объект расположен дальше чем 3σ от центра радиоисточника;
- два оптических объекта рядом с предполагаемым положением оптического кандидата, и по имеющейся фотометрической и спектральной информации нельзя сделать уверенного выбора между объектами;
- источник двойной, положение ядра определяется неуверенно, оптический объект сдвинут в сторону от линии, соединяющей максимумы плотностей потока компонентов;
- сложно сделать выводы о структуре радиоисточника по радиокарте FIRST.

Надежные оптические кандидаты были найдены для 70% радиоисточников, для 10% радиоисточников имеются возможные кандидаты, а для 20% не обнаружены оптические объекты, так как родительские галактики радиоисточников слабее предельной глубины обзора SDSS. Для 75% оптических кандидатов было проведено разделение на галактики и звездные объекты [18]. Результаты отождествления представлены на <http://www.sao.ru/fetch/cgi-bin/SkyObj/rc.cgi>.

3.4 Оптическое отождествление каталога RCR

Каталог RCR (RATAN Cold Refined) [5] получен в результате обработки 7 циклов наблюдений, проведенных на радиотелескопе РАТАН-600 с 1987 по 1999 гг., и повторной обработки данных 1980–1981 гг. для интервала прямых восхождений $7^h \leq \alpha_{2000} \leq 17^h$ с целью улучшения координат и уточнения плотностей потоков источников каталога RC. Отметим, что в результате отождествления каталога RC $\sim 25\%$ из 432 объектов, попавших в область обзоров FIRST и SDSS, не отождествились с источниками NVSS. Собственно говоря, это и послужило толчком для подготовки следующего релиза каталога RC — каталога RCR, по которому работа по отождествлению радиоисточников обзоров «Холод» 1980–1999 гг. была продолжена. В каталоге RCR 550 источников, что в 1,7 раза больше,

чем количество RC-источников, для которых уже проводилось оптическое отождествление.

Обычно радиоисточники отождествляются с достаточно слабыми объектами в оптике. Чем слабее по потоку радиоисточники, тем более глубокие снимки в оптике необходимы для обнаружения объекта, ответственного за радиоизлучение. При предельной глубине обзора SDSS 22.6 звездной величины по оценкам можно отождествить $\sim 30\%$ – 50% радиоисточников с плотностью потока ярче 1 мЯн на 1,4 ГГц. В каталоге RCR источники ярче — слабые объекты имеют плотность потока на частоте 3,9 ГГц 10–15 мЯн, что при пересчете на 1,4 ГГц будет соответствовать 17–25 мЯн. Так по каталогу SDSS удалось отождествить 70% RC-источников. Чтобы отождествить оставшиеся объекты, нужны были более глубокие снимки, которые можно получить, сложив изображения в трех фильтрах обзора SDSS — g , r и i , а также используя обзоры инфракрасного диапазона. Чем больше диапазонов электромагнитного спектра привлекается для исследования радиоисточников, тем больше информации для определения типа родительской галактики, а также ее физических характеристик. Поэтому для исследования радиоисточников каталога RCR был добавлен глубокий обзор неба в ближнем инфракрасном диапазоне UKIDSS и в среднем инфракрасном диапазоне — WISE.

Сценарий для отождествления источников каталога RCR включает: (1) подготовку данных; (2) визуализацию и предварительную обработку; (3) определение морфологического типа радиоисточника; (4) оптическое отождествление. Эти этапы мало отличаются по методике от описанных в предыдущем подразделе шагов.

Был расширен список каталогов и обзоров, а также использованы появившиеся в 6-й версии ALADIN возможности макроконтроллера по выполнению арифметических операций с изображениями. Ниже приведен пример скрипта для макроконтроллера с суммированием изображений:

```
G = get Skyview(300,Default,"SDSS G",Tan,J2000)
$1 $2
R = get Skyview(300,Default,"SDSS R",Tan,J2000)
$1 $2
\* извлечение изображений SDSS в фильтрах g, r, i
I = get Skyview(300,Default,"SDSS I",Tan,J2000)
$1 $2
#
R_n = norm -cut R
I_n = norm -cut I
\* нормализация изображений
G_n = norm -cut G
sync
RI = R_n + I_n
GRI = RI + G_n
```

* суммирование изображений
supc

С помощью программного интерфейса к приложению DS9 SAOImage была реализована программа на языке Python для построения рисунков в формате postscript по списку радиоисточников (рис. 2, б). Рисунки использовались для морфологической классификации радиоисточников.

В результате были получены надежные отождествления для 82% радиоисточников, для 10% обнаружены возможные кандидаты, а для 8% так и не были найдены оптические и/или инфракрасные объекты.

4 Информационно-поисковая система по результатам отождествления радиоисточников каталога RCR

По результатам отождествления радиоисточников каталога RCR автором накоплен материал, который предполагается использовать для дальнейших исследований, а именно: определения физических характеристик родительских галактик радиоисточников, их классификации, подготовки

выборки источников со схожими свойствами, поиска далеких объектов, поиска переменности в оптическом и радиодиапазоне. В табл. 1 приведены используемые при отождествлении информационные ресурсы, указаны их основные характеристики и спектральные полосы.

Для работы с компилятивным каталогом разработана информационная система, включающая базу данных по радиоисточникам и их родительским галактикам и веб-интерфейс для отображения разнообразной информации об объектах. Схема таблиц информационно-поисковой системы включает таблицы оригинальных каталогов (в описываемом случае это 12 каталогов), компилятивные таблицы, включающие материал по результатам отождествления радиоисточников. В схему включены представления:

- `v_rflux` — блеск объекта в радио-, инфракрасном и оптическом диапазоне;
- `v_rcparamr` — параметры радиоисточника;
- `v_rcparamo` — оптические параметры родительской галактики.

Веб-интерфейс (рис. 3) позволяет отображать по имени радиоисточника весь материал, относящийся к объекту — изображения, данные каталогов и вычисленные параметры. Адрес ре-

Таблица 1 Каталоги и обзоры, использовавшиеся для отождествления радиоисточников каталога RCR

Диапазон	Каталоги, обзоры	Спектральный диапазон	Разрешение/ошибки	Предел чувствительности
Радио	VLSS	74 МГц	80''	500 мДжанки
	TXS	365 МГц	$\sim 10''$	150 мДжанки
	NVSS	1400 МГц	45''	2.5 мДжанки
	FIRST	1400 МГц	5.4''	1 мДжанки
	GB6	4850 МГц	3.5''	28–37 мДжанки
Оптика	DSS-II	blue, red, IR (J, F, N)		$\sim 21^m$ ($\sum \text{BRI} \sim 21.2^m R$)
	SDSS	u, g, r, i, z ($g + r + i$)	$\pm 0.1''$	$22.0^m, 22.2^m, 22.2^m, 21.3^m,$ 20.5^m ($\sum \text{gri} \sim 22.6^m$)
	USNO-B1	B_1, R_1, B_2, R_2, I (O, E, J, F, N)	0.2'' 0.3 ^m	$V = 21^m$
	GSC 2.3.2	J, F, N (B_J, R_F, I_N)	0.2''–0.28'' 0.13 ^m –0.22 ^m	$R_F = 20^m$
ИК	2MASS	J, H, K	0.2'' 10%	$15.8^m, 15.1^m, 14.3^m$
	UKIDSS	Y (1.02 мкм), J (1.25 мкм), H (1.63 мкм), K (2.2 мкм) ($H + K$)	$< 0.1''$	$20.5^m, 20.0^m,$ $18.8^m, 18.4^m$ ($\sum HK \sim 22.8^m$)
	WISE	3.4 мкм, 4.6 мкм, 12 мкм, 22 мкм	6.1'', 6.4'', 6.5'', 12''	$16.5^m, 15.5^m,$ $11.2^m, 7.9^m$

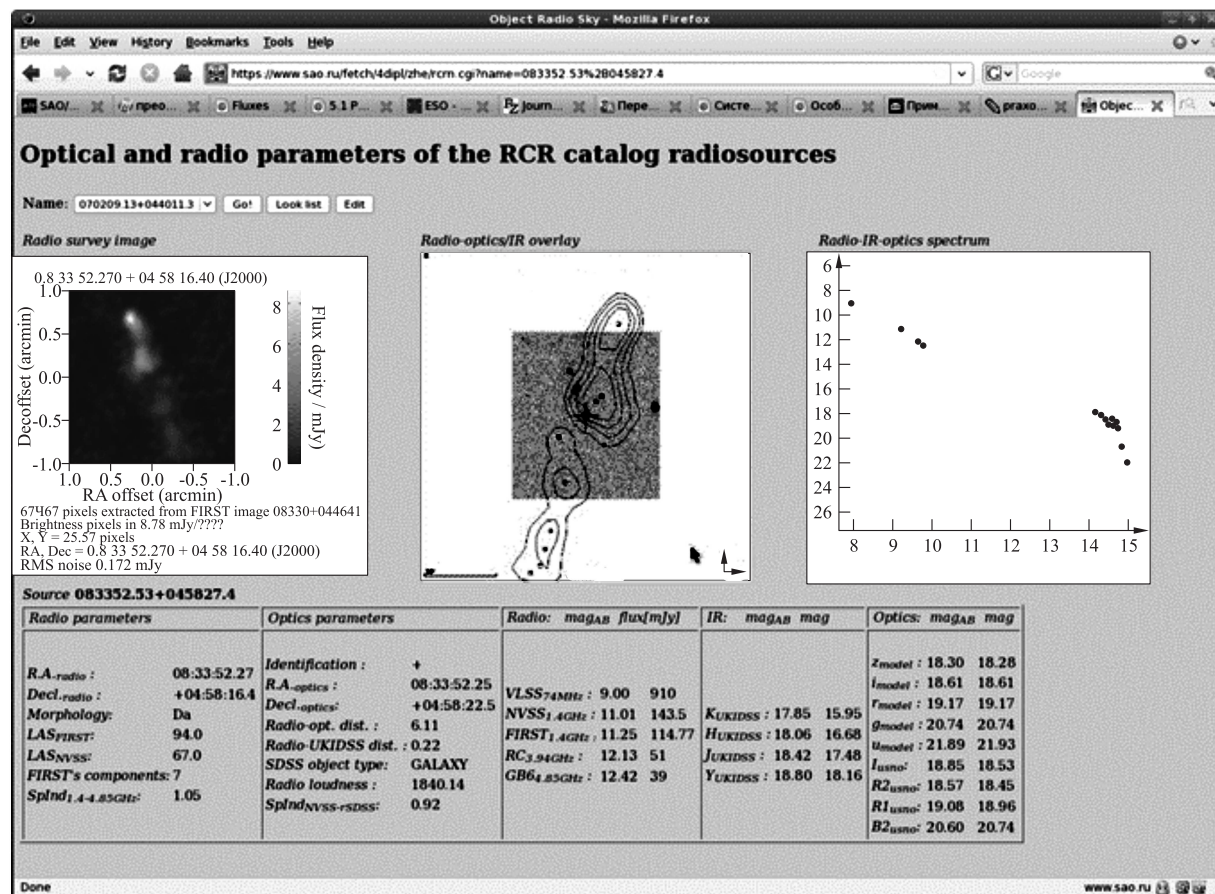


Рис. 3 Веб-интерфейс информационно-поисковой системы с результатами отождествления радиоисточников каталога RCR

сурса — <http://www.sao.ru/fetch/cgi-bin/SkyObj/rcrn.cgi>. Интерфейс реализован в архитектуре «клиент — сервер приложений — сервер СУБД». В качестве клиента используется браузер, сервер приложений — Apache, сервер базы данных — PostgreSQL. При написании скриптов использовался Python со стандартными пакетами для поддержки CGI-интерфейса, графическая библиотека PIL и модуль PyGreSQL для связи с СУБД по DBD/DBI-интерфейсам.

Информация, которая представлена в веб-интерфейсе, разделяется на следующие части: графическая статическая (подготовленные предварительно рисунки), графическая динамическая (веб-сервисы извлекают на лету изображение из обзора и помещают рисунок на страничку) и параметры радиоисточника, хранящиеся в информационно-поисковой системе.

Динамически выполняется построение спектрального распределения энергии радиоисточника (колонка «Radio-IR-optics spectrum»). Если источник отождествлен или есть возможный кандидат,

то спектр строится по данным радио-, оптического и инфракрасного (ИК) диапазонов. Поскольку в радиодиапазоне плотность потока от объекта измеряется в янских на заданной частоте, а блеск в оптическом и ИК диапазонах измеряется в звездных величинах в полосе длин волн (ангстремы или нанометры), то все величины пересчитываются в звездные величины в фотометрической системе AB [53]. На графике со спектральным распределением энергии объекта по оси абсцисс отложена величина десятичного логарифма частоты в герцах, а по оси ординат — звездная величина. Точки на графике обозначены разными цветами. Каждому цвету соответствуют данные определенного каталога. Поскольку ширина полос в оптическом и инфракрасном диапазоне обычно несколько сотен ангстрем, то звездная величина приписывается эффективной длине волны фильтра.

В радиопараметрах приводятся координаты центра источника, морфологический тип, угловые размеры источника в угловых секундах, число компонент в обзоре FIRST и спектральные индексы для

радиодиапазона. В оптических параметрах приведены координаты оптического кандидата, разница оптических и радиokoординат, тип оптического объекта и др. В третьей колонке приведены звездные величины в системе АВ и величины из каталогов.

Разработанная информационно-поисковая система упростила просмотр разнородных данных по радиоисточникам и использовалась при принятии решения об отождествлении.

5 Заключение

Виртуальная обсерватория, объединяющая астрономические данные в распределенную инфраструктуру, за несколько лет своего существования обеспечила новый качественный уровень работы с цифровыми коллекциями. В большинстве публикуемых статей по разным областям исследований в астрофизике присутствуют данные, которые получены с применением средств и ресурсов виртуальной обсерватории.

В течение ряда лет автором проводилось массовое изучение радиоисточников с применением разных приложений виртуальной обсерватории и накоплен как положительный, так и отрицательный опыт работы с многочастотными данными. Эффективность оптического отождествления радиоисточников несравненно выросла, как только появились глубокие цифровые обзоры неба и средства для работы с ними. То, на что раньше требовались годы, сейчас выполняется за обозримое время, однако идентификация списка объектов с каталогами остается трудоемкой. Когда появляются новые обзоры и релизы существующих обзоров, выборку информации из каталогов и обзоров приходится повторять, накапливая тем самым варианты компилятивных таблиц. Эффективность автоматической кросс-идентификации по координатам радиокаталогов с оптическими в худших случаях составляет всего несколько процентов, в лучших — около 30%. Причем в одних случаях для установления связи между объектами каталогов и/или обзоров кросс-идентификации достаточно, а в других случаях нужен более глубокий анализ с привлечением любых доступных данных, включая и информацию из опубликованных статей.

Установление связи между записями разных каталогов по смысловому содержанию является знанием, которое нужно сохранять и поддерживать. Проекты AstroDAbis и ADSASS нацелены на решение этой задачи, что послужит дальнейшему развитию виртуальной обсерватории как универсального инструмента исследований.

Литература

1. *Parijskij Yu. N., Bursov N. N., Lipovka N. M., Soboleva N. S., Temirova A. V.* The RATAN-600 7.6-cm catalog of radio sources from 'Experiment Cold-80' // *Astronomy Astrophys. Supplement Ser.*, 1991. Vol. 87. P. 1–32.
2. *Parijskij Yu. N., Soboleva N. S., Goss W. M., Kopylov A. I., Verkhodanov O. V., Temirova A. V., Zhelenkova O. P.* The RATAN-600 – VLA – 6 m Russian telescope: Early Universe Project // *175th Symposium of the International Astronomical Union.* — Bologna: Kluwer Acad. Publ., 1996. P. 591–602.
3. *Verkhodanov O. V., Parijskij Yu. N., Soboleva N. S., Temirova A. V., Zhelenkova O. P.* Color redshifts and the age of the stellar population of distant RC radio galaxies // *Astronomy Rep.*, 2002. Vol. 46. P. 531–542.
4. *Parijskij Yu. N., Kopylov A. I., Temirova A. V., Soboleva N. S., Zhelenkova O. P., Verkhodanov O. V., Goss W. M., Fatkhullin T. A.* Spectroscopy of "Big Trio" objects using the "Scorpio" spectrograph of the 6-m telescope of the Special Astrophysical Observatory // *Astronomy Rep.*, 2010. Vol. 54. P. 675–695.
5. *Soboleva N. S., Majorova E. K., Zhelenkova O. P., Temirova A. V., Bursov N. N.* RATAN-600 7.6-cm deep sky strip surveys at the declination of the SS433 source during the 1980–1999 period. Data reduction and the catalog of radio sources in the right-ascension interval $7^h < R.A. < 17^h$ // *Astrophys. Bull.*, 2010. Vol. 65. P. 42–59.
6. *Abazajian K. N., Adelman-McCarthy J. K., Agüeros M. A., et al.* The seventh data release of the Sloan Digital Sky Survey // *Astrophys. J. Suppl.*, 2009. Vol. 182. P. 543–558.
7. *Dye S., Warren S. J., Hambly N. C., et al.* The UKIRT infrared deep sky survey early data release // *Monthly Notices of the Royal Astronomical Society*, 2006. Vol. 372. P. 1227–1252.
8. *Lasker B. M., Lattanzi M. G., McLean B. J., et al.* The Second-Generation Guide Star Catalog: Description and properties // *Astronomical J.*, 2008. Vol. 136. P. 735–766.
9. *Monet D. G., Levine S. E., Canzian B., et al.* The USNO-B Catalog // *Astronomical J.*, 2003. Vol. 125. P. 984–993.
10. *Skrutskie M. F., Cutri R. M., Stiening R., et al.* The two micron all sky survey (2MASS) // *Astronomical J.*, 2006. Vol. 131. P. 1163–1183.
11. *Cohen A. S., Lane W. M., Cotton W. D., et al.* The VLA low-frequency sky survey // *Astronomical J.*, 2007. Vol. 134. P. 1245–1262.
12. *Douglas J. N., Bash F. N., Bozyan F. A., Torrence G. W., Wolfe C.* The Texas survey of radio sources covering $-35.5 < \delta < 71.5$ at 365 MHz // *Astronomical J.*, 1996. Vol. 111. P. 1945–1963.
13. *Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J.* The NRAO VLA sky survey // *Astronomical J.*, 1998. Vol. 115. P. 1693–1716.

14. *Becker R. H., Helfand D. J., White R. L., Gregg M. D., Laurent-Muehleisen S. A.* The FIRST Survey Catalog, Version 2003Apr11 // *Astrophysical J.*, 1997. Vol. 475. P. 479–493.
15. *Gregory P. C., Scott W. K., Douglas K., Condon J. J.* The GB6 catalog of radio sources // *Astrophys. J. Suppl.*, 1996. Vol. 103. P. 427–432.
16. *Zhelenkova O. P., Kopylov A., Chernenkov V.* The investigation of the RC catalog radiosources in the SDSS and FIRST crossing area with IVOA program tools // *JENAM. Our non-stable universe.* — Yerevan: BAO, 2007. P. 84–85.
17. *Zhelenkova O. P., Kopylov A. I.* Analysis of the RC catalog sample in the region overlapping with the regions of the FIRST and SDSS surveys: I. Identification of sources with the VLSS, TXS, NVSS, FIRST, and GB6 catalogs // *Astrophys. Bull.*, 2008. Vol. 63. P. 346–356.
18. *Zhelenkova O. P., Kopylov A. I.* Analysis of a sample of RC catalog objects in the region overlapping with the areas covered by FIRST and SDSS surveys. II: Optical identification with the SDSS survey and USNO-B1 and 2MASS catalogs // *Astrophys. Bull.*, 2009. Vol. 64. P. 111–122.
19. *Желенкова О. П., Майорова Е. К., Соболева Н. С., Темирова А. В.* Многочастотное исследование радиоисточников средствами виртуальной обсерватории // *Радиотелескопы, аппаратура и методы радиоастрономии: Мат-лы Всеросс. радиоастрономической конф. (БПК-2011).* — СПб.: ИПА РАН, 2011. С. 179–184.
20. *Zhelenkova O. P., Soboleva N. S., Majorova E. K., Temirova A. V.* Multiband study of radiosources of the RCR catalogue with the virtual observatory tools // *Baltic Astronomy*, 2012. Vol. 21. P. 5–13.
21. *Желенкова О. П., Майорова Е. К., Соболева Н. С., Темирова А. В.* Методы виртуальной обсерватории в задаче оптического отождествления радиоисточников // *Электронные библиотеки*, 2010. Т. 13. Вып. 4.
22. *Желенкова О. П.* Исследование радиоисточников средствами виртуальной обсерватории // *Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL'2011): Труды XIII Всеросс. науч. конф.* — Воронеж: ВГУ, 2011. С. 326–333.
23. *Boch T., Oberto A., Fernique P., Bonnarel F.* Aladin: An open source all-sky browser // *Astronomical Data Analysis Software and Systems XX.* — Boston: ASP, 2011. Vol. 442. P. 683–691.
24. *Joye W. A.* New features of SAOImage DS9 // *Astronomical Data Analysis Software and Systems XV.* — San Lorenzo de El Escorial: ASP, 2006. Vol. 351. P. 574–579.
25. *Brunner R. J., Djorgovski S. G., Lonsdale C., Madore B., Prince T., Szalay A. S.* Multi-wavelength cross-identification of the extragalactic sky: An NVO cornerstone // *Bull. Amer. Astronomical Soc.*, 2000. Vol. 32. P. 1601–1605.
26. *Quinn P. J., Benvenuti P., Diamond P. J., Genova F., Lawrence A., Mellier Y.* Astrophysical virtual observatory (AVO): A progress report // *SPIE Proceedings*, 2002. Vol. 4846. P. 1–5.
27. *Quinn P. J., Barnes D. G., Csabai I., et al.* The International Virtual Observatory Alliance: Recent technical developments and the road ahead // *SPIE Proceedings*, 2004. Vol. 5493. P. 137–145.
28. *Briukhov D. O., Kalinichenko L. A., Zakharov V. N., Panchuk V. E., Vitkovsky V. V., Zhelenkova O. P., Dluzhnevskaya O. B., Malkov O. Yu., Kovaleva D. A.* Information infrastructure of the Russian Virtual Observatory (RVO). — 2nd ed. — M.: IPI RAS, 2005.
29. The open archives initiative protocol for metadata harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.htm>.
30. Dublin core metadata initiative. <http://dublincore.org/documents>.
31. *Pence W. D., Chiappetti L., Page C. G., Shaw R. A., Stobie E.* Definition of the flexible image transport system (FITS), version 3.0 // *Astronomy Astrophys.*, 2010. Vol. 524. P. 42–82.
32. *Ochsenbein F., Williams R., Davenhall C., et al.* IVOA recommendation: VOTable format definition Version 1.2 // E-print, 2011. Arxiv:1110.0524. P. 1–35.
33. *Derriere S., Gray N., Mann R., Martinez A. P., McDowell J., McGlynn T., Ochsenbein F., Osuna P., Rixon G., Williams R.* An IVOA standard for unified content descriptors. Version 1.1. <http://www.ivoa.net/Documents/REC/UCD/UCD-20050812.pdf>.
34. IVOA data access layer. <http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaDAL>.
35. *Ortiz I., Lusted J., Dowler P., et al.* IVOA recommendation: IVOA astronomical data query language Version 2.00 // E-print, 2011. Arxiv:1110.0503. P. 1–36.
36. *Ochsenbein F., Bauer P., Marcout J.* The Vizier database of astronomical catalogues // *Astronomy Astrophys. Suppl.*, 2000. Vol. 143. P. 23–32.
37. *Hambly N. C., Collins R. S., Cross N. J. G., et al.* The WFCAM science archive // *Monthly Notices Roy. Astronomical Soc.*, 2008. Vol. 384. P. 637–662.
38. *Berriman G. B.* The NASA/IPAC Infrared Science Archive (IRSA) as a resource in supporting observatory operations // *SPIE Proceedings*, 2008. Vol. 7016. P. 701618–701618-9.
39. *Kamp I., Thompson R., Conti A., et al.* MAST in the context of VO activities // *Astronomical Data Analysis Software and Systems XIV.* — Pasadena: ASP, 2005. Vol. 347. P. 208–212.
40. *Gray N., Mann R. G., Morris D., Holliman M., Noddle K.* AstroDabIs: Annotations and cross-matches for remote catalogues // E-print, 2011. ArXiv:1111.6116. P. 1–4.
41. Resource description framework. <http://www.w3.org/standards/techs/rdf>.
42. Linked data. <http://www.w3.org/DesignIssues/LinkedData.html>.
43. *Holliman M., Alemu T., Hume A., van Hemert J., Mann R. G., Noddle K., Valkonen L.* Service infrastructure for cross-matching distributed datasets using OGSA-DAI and TAP // *Astronomical Data Analysis Software and Systems XX.* — Boston: ASP, 2011. Vol. 442. P. 579–583.

44. OGSA-DAI. <http://www.ogsadai.org.uk/about/index.php>.
45. *Dowler P., Rixon G., Tody D.* Table access protocol (TAP, v1.0), IVOA recommendation // ArXiv:1110.0497. <http://www.ivoa.net/Documents/TAP>.
46. *Taylor M. B.* TOPCAT & STIL: Starlink Table/VOTable Processing Software // Astronomical Data Analysis Software and Systems XIV. — Pasadena: ASP, 2005. Vol. 347. P. 29–32.
47. *Laurino O., Smareglia R.* Vodka: A data keeping-up agent for the virtual observatory // Astronomical Data Analysis Software and Systems XX. — Boston: ASP, 2011. Vol. 442. P. 571–574.
48. *Brescia M., Longo G., Djorgovski G. S., et al.* DAME: A web oriented infrastructure for scientific data mining & exploration // E-print, 2011. Arxiv:1111.3983. P. 1–16.
49. *Pepe A., Goodman A., Muench A.* The ADS all-sky survey // E-print, 2011. Arxiv:1111.6116. P. 1–4.
50. *Zhelenkova O., Vitkovskij V. V., Briukhov D., Kalinichenko L. A.* Search of distant radio galaxies as a subject mediator example // Astronomical Data Analysis Software and Systems XV. — San Francisco: ASP, 2006. Vol. 351. P. 244–249.
51. *Брюхов Д. О., Вовченко А. Е., Захаров В. Н., Желенкова О. П., Калиниченко Л. А., Мартынов Д. О., Скворцов Н. А., Ступников С. А.* Архитектура промежуточного слоя предметных посредников для решения задач над множеством неоднородных распределенных информационных ресурсов в гибридной грид-инфраструктуре виртуальных обсерваторий // Информатика и её применения, 2008. Т. 2. Вып. 1. С. 2–34.
52. *Walton N. A.* The AstroGrid Consortium. The AstroGrid Virtual Observatory Service // Astronomical Data Analysis Software and Systems XVII. — London: ASP, 2008. Vol. 394. P. 251–255.
53. *Oke J. B., Gunn J. E.* Secondary standard stars for absolute spectrophotometry // Astrophys. J., 1983. Vol. 266. P. 713–717.

МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ РАЗРАБОТКИ ЭЛЕКТРОННОГО РУССКО-АНГЛИЙСКОГО ТЕЗАУРУСА ПО КОМПЬЮТЕРНОЙ ЛИНГВИСТИКЕ*

Ю. А. Загорулько¹, О. И. Боровикова², И. С. Кононенко³, Е. Г. Соколова⁴

Аннотация: Обсуждаются методологические аспекты разработки русско-английского электронного тезауруса по компьютерной лингвистике (КЛ). Обосновывается необходимость разработки такого тезауруса и принципы его построения. Описываются состав тезауруса, структура тезаурусной статьи и набор связей между терминами. Обсуждается методика выбора терминов для включения в тезаурус, а также проблемы выбора основного термина-дескриптора из множества синонимичных терминов и подбора парных терминов-эквивалентов. Рассматриваются особенности реализации электронной версии тезауруса, при этом особое внимание уделяется проблеме поддержания логической целостности терминологической системы тезауруса и обеспечению удобного доступа к его содержанию.

Ключевые слова: многоязычный тезаурус; компьютерная лингвистика; методология разработки тезаурусов; онтология; концептуальная схема тезауруса; технология построения порталов научных знаний

1 Введение

Обеспечить обработку и эффективное использование постоянно растущих объемов неструктурированной информации становится уже невозможно без привлечения методов КЛ. Чтобы успешно применять эти методы для решения задач индексирования и содержательного поиска документов, извлечения информации из текстов, машинного перевода и построения естественно-языковых интерфейсов, требуется знание терминологии КЛ, причем не только русскоязычной, но и англоязычной, так как большинство учебников и научных публикаций по тематике КЛ представлено на английском языке. Однако на данный момент в КЛ не существует четкой и общепринятой системы научной терминологии, причем многие термины современной КЛ не представлены на русском языке ни в одном из лингвистических источников.

Так, тезаурус по теоретической и прикладной лингвистике, созданный в 1978 г. С. Е. Никитиной [1], уже устарел. К тому же он одноязычный и не содержит определений понятий. Англо-русский терминологический словарь В. З. Демьянкова [2] содержит толкования, но не отражает современную картину этого научного направления.

Собственно лингвистика представлена в нескольких фундаментальных источниках, в част-

ности в Лингвистическом энциклопедическом словаре (ЛЭС) [3], словаре О. С. Ахмановой [4], а также интернет-энциклопедии «Кругосвет» [5], содержащей статьи по новым для традиционной лингвистики понятиям. Разработанный в 2007 г. в ИНИОН РАН тезаурус по языкознанию [6] содержит около 3000 терминов, однако только около 4% из них относятся к области КЛ.

Определения терминов КЛ можно найти в толковом словаре по искусственному интеллекту [7]. Однако он отражает терминологию на конец 1980-х гг. и содержит довольно мало терминов КЛ.

Так как КЛ имеет междисциплинарный характер, то некоторые ее термины можно найти в общих энциклопедиях, например в Большом энциклопедическом словаре [8]. Популярным источником знаний по КЛ сейчас является Википедия [9], в которой можно найти объяснения, классификации и ссылки на источники по многим понятиям КЛ, однако эти сведения часто страдают односторонностью, неполнотой и эскизностью.

Таким образом, на данный момент не существует источника, в котором вся терминология КЛ была бы приведена в единую систему. Это вызывает необходимость разработки двуязычного тезауруса, содержащего английские и русские термины КЛ и их толкования. Двуязычность тезауруса даст возможность отечественным ученым и специалистам

* Работа выполнена при финансовой поддержке РГНФ (проект № 10-04-12108в).

¹ Институт систем информатики имени А. П. Ершова СО РАН, zagor@iis.nsk.su

² Институт систем информатики имени А. П. Ершова СО РАН, olesya@iis.nsk.su

³ Институт систем информатики имени А. П. Ершова СО РАН, irina_k@cn.ru

⁴ Российский государственный гуманитарный университет, minegot@rambler.ru

быстрее и эффективнее ориентироваться в мировой ситуации в данной области. Составление такого тезауруса позволит выявлять различия и сходства между понятиями, используемыми в отечественной и зарубежной науке, а также вводить новые понятия и лингвистические термины, отсутствующие в русском языке [10].

В данной работе обсуждаются методологические аспекты разработки русско-английского тезауруса по компьютерной лингвистике. В разд. 2 описываются принципы разработки и состав тезауруса, структура тезаурусной статьи и набор связей между терминами. В разд. 3 обсуждается методика выбора терминов для включения в тезаурус, а также проблемы выбора основного термина-дескриптора из множества синонимичных терминов и подбора парных терминов-эквивалентов. В разд. 4 рассматриваются особенности реализации электронной версии тезауруса.

2 Проектирование структуры тезауруса

Проектирование структуры русско-английского тезауруса по компьютерной лингвистике выполнялось в соответствии с существующими отечественными и международными стандартами [11–15], регламентирующими построение информационно-поисковых тезаурусов (ИПТ), а также на основе анализа и обобщения накопленного к этому времени опыта разработки ряда отечественных тезаурусов ИНИОН [16], РуТез [17] и др.

Упомянутые выше стандарты определяют основные единицы тезауруса и возможный набор отношений между ними, устанавливают общие правила сбора массива лексических единиц, формирования словника, построения словарных статей и оформления ИПТ.

В зависимости от назначения ИПТ могут включать в свой состав либо только дескрипторы (предпочтительные термины), либо дескрипторы и аскрипторы (обычные термины). Во втором случае дескрипторы могут использоваться при индексировании документов и в поисковых запросах, а аскрипторы (как текстовые входы) подлежат замене одним или несколькими дескрипторами [17].

Тезаурусы делятся на одноязычные и многоязычные. Многоязычный информационно-поисковый тезаурус (МИПТ) содержит термины из нескольких естественных языков и представляет эквивалентные по смыслу понятия на каждом из них.

Построение русско-английского тезауруса по КЛ выполнялось в соответствии с требованиями межгосударственного стандарта ГОСТ 7.24-2007 [15], который разработан с учетом основных нормативных положений международного стандарта ISO 5964-1985 [11] и устанавливает состав, структуру и основные требования к построению МИПТ. Тезаурус разрабатывался как набор одноязычных версий МИПТ, при этом выполнялось согласованное построение одновременно двух версий тезауруса — русскоязычной и англоязычной. Разработка одноязычных версий тезауруса выполнялась на основе международного стандарта ISO 2788-1986 [12], межгосударственного стандарта ГОСТ 7.25-2001 [13] и американского стандарта Z39.19-2005 [14].

2.1 Выбор структуры словарной статьи

Основными единицами разработанного тезауруса являются термины предметной области (ПрО), подразделяемые на дескрипторы и аскрипторы. В тезаурус включаются следующие типы лексических единиц: одиночные слова (преимущественно существительные), именные словосочетания, лексически значимые компоненты сложных слов, сокращения слов и словосочетаний. Близкие по смыслу лексические единицы образуют класс эквивалентности, при этом одна из них выбирается в качестве представителя этого класса и получает статус дескриптора, остальные лексические единицы получают статус аскриптора. Статус аскриптора получают также и термины, представляемые аббревиатурами или иными вариантами написания (через дефис, с пробелом и т. п.).

В состав словарной статьи термина, вне зависимости от его статуса, входят следующие элементы:

- *название термина*, т. е. лексическая единица, представленная в нормальной форме (для одиночного существительного или опорного слова словосочетания это форма именительного падежа единственного числа);
- *язык*, на котором дано название термина;
- *комментарий*, включающий правила и рекомендации использования термина, а также замечания и пояснения автора словарной статьи;
- *автор словарной статьи*, т. е. фамилия и имя разработчика словарной статьи (задается для контроля процесса коллективной разработки тезауруса).

Для описания терминов-дескрипторов, кроме перечисленных выше атрибутов, вводятся следующие дополнительные атрибуты:

- *определение термина*, поясняющее на языке термина его смысл или значение. Наличие в тезаурусе определений терминов делает возможным его использование не только в качестве инструмента для ручного или автоматизированного индексирования, но и в качестве источника систематизированных знаний о данной ПрО;
- *релятор*, представляющий собой помету, введенную для различения омонимичных терминов (омографов) в рамках описываемой ПрО. Он является частью термина и поясняет его значение, относя его к определенной понятийной категории или предметно-тематической области (в контексте данной статьи — подобласти КЛ или смежной с ней области/подобласти знаний). Например, для различения двух понятий, образованных на основе словосочетания РАЗМЕТКА ТЕКСТА, могут быть использованы реляторы ПРОЦЕСС и ОБЪЕКТ. В результате получается два разных термина-дескриптора РАЗМЕТКА ТЕКСТА (ПРОЦЕСС) и РАЗМЕТКА ТЕКСТА (ОБЪЕКТ);
- *область/подобласть знаний*, к которой относится данный термин-дескриптор;
- *признак корневого термина*, указывающий на то, что дескриптор находится на самом верхнем уровне одной из представленных в тезаурусе иерархий понятий.

Термины тезауруса связываются различными семантическими отношениями, отражающими место каждого термина в системе понятий выбранной ПрО.

Для связи дескрипторов с аскрипторами используются отношения синонимии нескольких типов. Так, если дескриптор может однозначно во всех контекстах заменить какой-то аскриптор, то он связывается с ним отношением «Синоним»; при этом также устанавливается обратное отношение от аскриптора к дескриптору — «Смотри». Для моделирования других соотношений между аскрипторами и дескрипторами в соответствии с ГОСТ 7.25-2001 в тезаурус вводятся отношения, позволяющие задавать связи между аскрипторами и альтернативными дескрипторами или представлять аскриптор комбинацией дескрипторов.

В тех случаях, когда нет однозначного соответствия между дескрипторами и аскрипторами, используются отношения «Используй альтернативно» или «Используй комбинацию», задающие соответствие между аскриптором и заменяющими его дескрипторами; при этом вводятся обратные им отношения «Сравни альтернативный выбор» и «Сравни комбинацию».

Например, аскриптор ПАРТИЦИПАНТ может быть связан отношением «Используй альтернативно» с дескрипторами СЕМАНТИЧЕСКАЯ ВАЛЕНТНОСТЬ и УЧАСТНИК СИТУАЦИИ. В то же время аскриптор СИСТЕМА СТАТИСТИЧЕСКОГО МАШИННОГО ПЕРЕВОДА может быть представлен с помощью связи «Используй комбинацию» как комбинация (сочетание) двух дескрипторов — СИСТЕМА МАШИННОГО ПЕРЕВОДА и СТАТИСТИЧЕСКИЙ МАШИННЫЙ ПЕРЕВОД.

Для отражения семантических связей между понятиями, выражаемыми дескрипторами, устанавливаются иерархические и ассоциативные отношения. (Следует заметить, что такого типа отношения связываются только дескрипторы, входящие в одну и ту же одноязычную версию тезауруса.)

В тезаурусе допускается использование таких иерархических отношений, как недифференцированная иерархическая связь «Выше», направленная от нижестоящего дескриптора к вышестоящему; родовидовая связь «Выше род», устанавливаемая между двумя дескрипторами, когда объем понятия нижестоящего дескриптора входит в объем понятия вышестоящего дескриптора; партонимическая связь «Выше целое», задаваемая между двумя дескрипторами в том случае, когда нижестоящий дескриптор представляет компонент объекта, обозначаемого вышестоящим дескриптором. Вводятся также обратные им отношения: «Ниже», «Ниже вид», «Ниже часть».

Для задания отношений между дескрипторами, представляющими класс понятий и экземпляр этого класса, были выбраны связи «Выше класс» и «Экземпляр класса».

При установлении иерархических отношений для некоторых дескрипторов можно указать признак «Аспект деления иерархии». Так, например, в иерархии, построенной по отношению «Ниже вид», МАШИННЫЙ ПЕРЕВОД по признаку «подход» разделяется на СТАТИСТИЧЕСКИЙ МАШИННЫЙ ПЕРЕВОД, МАШИННЫЙ ПЕРЕВОД НА ОСНОВЕ ПРАВИЛ и МАШИННЫЙ ПЕРЕВОД, ОСНОВАННЫЙ НА ПРЕЦЕДЕНТАХ, а по признаку «степень участия человека» — на ПОЛНОСТЬЮ АВТОМАТИЧЕСКИЙ ПЕРЕВОД и ЧЕЛОВЕКО-МАШИННЫЙ ПЕРЕВОД.

Таким образом, один и тот же дескриптор одновременно может входить в несколько иерархий понятий, построенных по различным отношениям («Выше», «Выше род», «Выше целое») и по различным аспектам деления иерархии.

Для задания произвольных ассоциативных связей между дескрипторами, например отношений, выражающих зависимости вида «процесс—объект»,

«причина—следствие» и др., вводится отношение «Ассоциируется с».

Для связывания эквивалентных по смыслу дескрипторов, входящих в разные одноязычные версии, служит отношение «Эквивалент на другом языке».

2.2 Представление источников терминов

Для подтверждения актуальности введенных в тезаурус терминов и ознакомления пользователей тезауруса с практикой их употребления для каждого термина задаются его связи с источниками, т. е. текстовыми документами или коллекциями текстовых документов, в которых данный термин встречается или определяется.

Этим целям служат отношения «Встречается в», «Встречается в части документа» и «Дается определение в».

Отношение «Встречается в» служит для связывания любого термина с источником; при этом, если источник — коллекция текстов, то в качестве значения специального атрибута этого отношения можно указать частоту встречаемости термина в источнике.

С помощью отношения «Встречается в части документа» можно сообщить, что данный термин встречается в предметном указателе или глоссарии источника, что указывает на важность термина и повышает степень доверия к нему.

С помощью отношения «Дается определение в» термины-дескрипторы, снабженные толкованиями-определениями, связываются с источниками определений.

В тезаурусе источники описываются следующими параметрами: название, библиографическая ссылка, язык, тип (книга, монография, научная статья, документация, учебник, словарь, тезаурус, интернет-ресурс, коллекция текстов и др.), краткое описание и адрес в сети Интернет. Для коллекции текстов дополнительно задается число текстов и словоупотреблений.

3 Методика выбора терминов для включения в тезаурус

Важным моментом при построении тезауруса является методика подбора терминов — кандидатов на включение в тезаурус, — выбор терминов-дескрипторов из множеств синонимичных терминов, а также подбор иноязычных эквивалентов.

Выбор терминов для включения в русско-английский тезаурус по КЛ сопряжен с трудностями,

которые обусловлены особенностями самой КЛ как новейшей науки и состоянием ее развития в России. Здесь важно отметить следующие факторы, характеризующие КЛ в целом и русскоязычную КЛ (РКЛ) в частности:

- междисциплинарный характер КЛ;
- неоднородность Про «Компьютерная лингвистика»;
- неравномерность развития отдельных направлений КЛ;
- отличие русскоязычной КЛ от англоязычной (в частности, отставание отдельных направлений РКЛ).

Ранее КЛ рассматривалась как часть исследовательского направления «искусственный интеллект» (ИИ), терминология которого считается зрелой: «Специальная терминология по искусственному интеллекту и интеллектуальным системам начала формироваться в 60-е годы XX в. Первый этап формирования терминологии всегда отличается наличием многих синонимических терминов, которые используют различные школы и группы специалистов. На этом этапе термины быстро возникают и часть из них так же быстро исчезает. К середине 1970-х гг. терминология в области искусственного интеллекта стала устанавливаться. Появились термины, которые признало подавляющее большинство специалистов. Все эти термины (за редким исключением) по происхождению англоязычные, так как именно в США проводились интенсивные исследования в этой области. Окончательно основная терминология закрепились в первой половине 1980-х гг.» [7].

Искусственный интеллект — это методологическая область, методы которой применимы к разным Про, в частности активно применяются в КЛ в последнее десятилетие. Терминология КЛ в отдельных разделах продолжает сохранять черты первого этапа (наличие большого числа синонимов, например в разделе семантических отношений). Искусственный интеллект тоже считается междисциплинарной областью, однако по этому параметру ИИ и КЛ противоположны: ИИ междисциплинарна, потому что ее методы применяются в разных дисциплинах, КЛ — потому что она вбирает в себя разные дисциплины, такие как лингвистика (разделы, связанные с обработкой текстов и речи), психология, некоторые разделы ИИ.

Следствием указанных выше факторов является отсутствие русскоязычных учебных и лексикографических источников, достаточно полно отражающих структуру современной КЛ, в отличие от англоязычных источников, где она представлена детально и отчетливо. До сих пор термины РКЛ

входили лишь в состав словарей и глоссариев по лингвистике и смежным ей областям знаний. Так, имеются источники только по отдельным разделам смежных областей и КЛ, например по искусственному интеллекту, информационному поиску, и почти полностью отсутствуют русскоязычные термины по другим разделам КЛ, например по «Оценке эффективности систем и методов» (*Evaluation*). Кроме того, один и тот же термин, например *синтаксический анализ*, в таких смежных науках, как ИИ и КЛ, имеет разное толкование.

Учитывая вышеперечисленные особенности КЛ и связанный с ними недостаток современной справочной русскоязычной литературы по КЛ, при разработке тезауруса использовались источники «живых» терминов РКЛ и их толкований, и именно они фиксировались в словарных статьях тезауруса.

В качестве основного источника русскоязычных терминов была выбрана коллекция текстов докладов, представленных на международной конференции «Диалог» [18] в 2000–2010 гг., как «зеркала», отражающего термины РКЛ в их реальном употреблении. Собранный коллекция имеет следующие характеристики: число документов — 1193, объем — 4 610 694 словоупотреблений, суммарный размер — 27,5 МБ.

К данной коллекции была применена словарная технология [19], с помощью которой на базе лингвистических моделей (морфологического и локального синтаксического анализа) и статистических показателей был создан список статистически значимых в данной ПрО слов и словосочетаний — кандидатов в термины ПрО. Затем этот список был обработан (отфильтрован) экспертами в области КЛ, которые существенно опирались не только на знания о предмете и направлениях КЛ, но и на общелингвистические представления о терминологичности и путях формирования терминологических словников. Таким образом, избранный авторами подход, учитывающий предварительное структурирование ПрО, согласуется с общей методикой формирования словников на базе классификационных схем предметных областей (см., например, [20]).

Для английской части словника, с учетом русско-английской направленности создаваемого тезауруса, выбирались переводные эквиваленты из доступных англоязычных источников по КЛ.

С другой стороны, чтобы дополнить картину РКЛ в тех ее разделах, где имеются пробелы, при сборе терминов по таким разделам пришлось опираться преимущественно на англоязычные источники. Так, учитывая скачок, совершенный в течение последних нескольких лет в такой высокотехнологичной подобласти КЛ, как «Речевые тех-

нологии», а также тот факт, что это направление слабо представлено в коллекции «Диалог», при сборе терминов для этой подобласти была применена обратная методика, т. е. в качестве основных использовались англоязычные источники: предметные указатели нескольких современных и наиболее авторитетных англоязычных книжных источников обзорно-учебного профиля и глоссарии, входящие в документацию известных звуковых анализаторов. На данной терминологической базе был составлен англо-русский словник параллельных терминов.

Достаточно сложной оказалась и проблема выбора основного термина-дескриптора из множества синонимичных терминов. Прежде всего, эта проблема связана с появлением новых понятий и соответствующих им терминов. Так, появление систем *translation memory* в сфере автоматизированного перевода привело к широкому использованию практиками-переводчиками термина *память переводов*, который не был принят научным сообществом, противопоставившим ему термин *переводческая память* (синонимический ряд терминов с частотными характеристиками из коллекции «Диалог»: *переводческая память* — 8, *память переводов* — 0, *архив переводов* — 1, *накопитель переводов* — 0, *копилка переводов* — 0).

Развитие некоторых направлений КЛ (например, таких как *автоматический перевод в режиме онлайн*) приводит к столкновению вариантов старых терминов. Так, тезаурус ИНИОН [6] и ЛЭС [3] основным термином в паре *автоматический перевод* и *машинный перевод* считают *автоматический перевод*, присвоив ему статус дескриптора. Однако показатели встречаемости в коллекции «Диалог» говорят в пользу термина *машинный перевод*: 318 против 58. Интернет-энциклопедии «Википедия» и «Кругосвет», а также учебники придерживаются этой же традиции. На сайте Европейской ассоциации машинного перевода [21] также отмечается, что термин *machine translation*, хоть и звучит архаично, но, тем не менее, сохраняется как основной общий термин для всей области. В данном случае эксперты согласились с этой, соответствующей традиции, точкой зрения.

Проблема выбора дескриптора возникает на фоне незрелости системы понятий КЛ, приводящей в некоторых случаях к очень широкой вариативности терминов, с одной стороны, и к их многозначности, с другой. Так, термин *валентная структура* с частотностью 20 имеет целый ряд вариантов: *валентная рамка* — 14, *рамка валентностей* — 64, *валентностная структура* — 3, *схема валентностей* — 3. В то же время исследование реального употребления термина *валентная структура* показало, что он, как и термин *модель управления*, имеет

как узкое толкование (множество синтаксических валентностей предикатного слова), так и более широкое толкование (описание соответствия семантических валентностей слова их грамматическому оформлению, т. е. синтаксическим валентностям). В этой ситуации в тезаурус вводится два одинаковых дескриптора, один из которых снабжается релятором.

Серьезные трудности возникают при подборе парных терминов (англо-русских эквивалентов). В качестве примера можно привести термин *spoken language machine translation*. Задача автоматического перевода устной речи возникла на стыке «Машинного перевода» и «Речевых технологий». *Spoken language processing* обычно переводится как *автоматическая обработка устного языка*, одной из задач которой является автоматический устный перевод (АУП) с его разновидностями, соответствующими АУП типа «Речь(L_1) → Текст(L_2)» и АУП типа «Речь(L_1) → Речь(L_2)». Вторая разновидность представлена английским термином *speech-to-speech translation*. В русскоязычной литературе такой традиции нет, как нет (или практически нет) и такого типа приложений. Поиск в Интернете дал в качестве эквивалента для *spoken language machine translation* единично встретившийся вариант *автоматический перевод устной речи*. Этот русский переводной эквивалент и был выбран в качестве парного русскоязычного термина- дескриптора.

Таким образом, при выборе терминов-дескрипторов авторы опирались не только на статистику, но и на традиции словоупотребления, сложившиеся к настоящему времени в лингвистическом научном сообществе. Что же касается выбора парных терминов для новейших подобластей КЛ, не представленных в русскоязычной литературе, соответствующие дескрипторы предлагались как переводные эквиваленты, а в качестве основных критериев выбора перевода выступили знания и интуиция эксперта.

4 Подход к реализации электронной версии тезауруса

Для реализации электронной версии тезауруса было решено использовать методологию и программные компоненты технологии построения порталов научных знаний [22, 23], уже ранее применявшиеся при создании порталов знаний по археологии [24] и компьютерной лингвистике [25]. Данная технология базируется на онтологии и предоставляет средства настройки на предметную область и управления контентом информационной системы, а также средства навигации и поиска.

Средства настройки на предметную область и поддерживаемая ими методология достаточно хорошо подходят для разработки концептуальной схемы тезауруса, а остальные из перечисленных средств могут выполнять роль его основных программных компонентов, обеспечивающих создание, сопровождение и использование тезауруса.

4.1 Разработка онтологии представления тезауруса

В используемой технологии в качестве информационной модели портала знаний (информационной системы) используется онтология, которая, обеспечивая формальное описание предметной области системы, не только определяет структуры для его информационного наполнения (контента), но и задает базис для организации содержательного доступа к знаниям и данным, содержащимся в нем.

Для описания онтологии данная технология предоставляет формализм, который назовем онтологией представления знаний, и поддерживающий его редактор онтологии. С помощью этих средств была построена онтология представления тезауруса O_{Th} , задающая его концептуальную схему:

$$O_{Th} = \langle C, R, T, D, At, P, Axt \rangle$$

где $C = \{Tr, S_T, S_K\}$ — конечное непустое множество классов, представляющих основные сущности тезауруса; здесь $Tr = Asc\ Tr = Asc \cup Des$ — класс терминов, представляющих понятия ПрО «Компьютерная лингвистика», включающий два подкласса — Asc (термины-аскрипторы) и Des (термины-дескрипторы); S_T — класс источников терминов; S_K — класс областей/подобластей знаний; $R = R^{TT} \cup R^{TST} \cup R^{TSK}$ — конечное множество отношений, где $R^{TT} = \{R_a^{TT}, \dots, R_m^{TT}\}$, $R_i^{TT} \subseteq Tr \times Tr$ — конечное множество бинарных отношений, заданных на терминах, $R^{TST} = \{R^{TSF}, R^{TSP}, R^{TSD}\}$, $R_i^{TST} \subseteq Tr \times S_T$ — бинарные отношения, связывающие термины тезауруса с источниками, причем R^{TSF} связывает термин с источником, где он встречается, R^{TSP} связывает термин с источником, где он встречается в предметном указателе или глоссарии, а R^{TSD} — связывает термин с источником, где дается его определение; $R^{TSK} = \{R^{SKT}, R^{SKS}\}$ — бинарные отношения, служащие для встраивания областей знаний в тезаурус, где $R^{SKT} \subseteq Tr \times S_K$ связывает термины тезауруса с областями знаний, а $R^{SKS} \subseteq S_K \times S_K$ задает иерархию на подобластях знаний; T — множество стандартных типов; $D = \{d_1, \dots, d_n\}$ — множество доменов $d_i = \{s_1, \dots, s_k\}$, где s_i — значение стандартного типа *string*; $at = \{at_1, \dots, at_w\}$ — конечное множество атрибутов, описывающих свойства

основных сущностей тезауруса и отношений между ними; значения этих свойств определены на множестве $T \cup D$; $P = \{P_1, \dots, P_n\}$ — множество формальных свойств отношений R^{TT} ; Axt — множество аксиом, задающих дополнительные ограничения на связи между терминами.

Таким образом, онтология представления тезауруса описывает классы, представляющие основные сущности тезауруса (термины тезауруса, их источники, области/подобласти знаний), отношения, связывающие объекты этих классов между собой, свойства понятий и отношений, а также аксиомы, определяющие их дополнительную семантику. Кроме того, в онтологии задается множество доменов, т. е. возможных значений атрибутов классов и отношений, что позволяет уменьшить число ошибок при создании/редактировании конкретного тезауруса.

Для отношений в онтологии задаются математические свойства (симметричность, рефлексивность, транзитивность, асимметричность, антирефлексивность) и обратные отношения.

Так, для введенных в подразд. 2.1 иерархических отношений («Выше», «Выше род», «Выше класс», «Выше целое») задаются математические свойства «транзитивность» и «асимметричность» и соответствующие обратные отношения («Ниже», «Ниже вид», «Экземпляр класса», «Ниже часть»). Отношения «Эквивалент на другом языке» и «Ассоциируется с» объявляются симметричными и антирефлексивными. Для отношений, выражающих синонимию терминов («Синоним», «Используй альтернативно», «Используй комбинацию»), задаются обратные отношения (соответственно «Смотри», «Сравни альтернативный выбор», «Сравни комбинацию»).

4.2 Организация управления контентом тезауруса

Для описания конкретных терминов, их источников, областей знаний, а также для установления связей между ними используется редактор данных, предоставляемый технологией построения порталов знаний и управляемый онтологией представления тезауруса. Этот редактор реализован как веб-приложение и доступен зарегистрированным пользователям через Интернет. (Заметим, что сразу после завершения ввода и/или редактирования описаний терминов, источников и связей между ними, новая информация становится доступной через пользовательский веб-интерфейс тезауруса.)

С целью обеспечения распределенной коллективной разработки используемая технология поддерживает механизм делегирования прав экспертам

разных уровней. В соответствии с этим механизмом только эксперты самого высокого уровня могут редактировать структуры тезауруса (с помощью редактора онтологий), а эксперты более низких уровней — только его содержание (с помощью редактора данных).

Кроме того, действует правило, по которому редактировать словарную статью может только ее автор. Если кто-то из экспертов захочет внести изменения в «чужую» статью, он должен согласовать такую возможность с ее автором, в частности, через специальный форум, на который имеется ссылка в электронном тезаурусе.

Для того чтобы тезаурус мог использоваться при индексировании и поиске текстовых документов, он должен представлять целостную и непротиворечивую систему понятий ПрО. Это обеспечивается встроенными в редактор механизмами вывода и поддержки логической целостности системы понятий тезауруса, работа которых базируется на описаниях свойств классов и отношений, заданных в онтологии представления тезауруса.

В частности, на основе этих свойств происходит корректное установление связей между терминами тезауруса, при необходимости осуществляется их автоматическое добавление и/или удаление. Кроме того, контролируются ограничения на существование и число связей между терминами тезауруса в зависимости от принадлежности терминов к тем или иным классам.

Например, если для рассмотренного выше отношения «Смотри» задано обратное отношение («Синоним») и ограничение на существование связей («только одна связь данного типа для каждого термина-аскриптора»), то при связывании аскриптора ПАМЯТЬ ПЕРЕВОДОВ и дескриптора ПЕРЕВОДЧЕСКАЯ ПАМЯТЬ отношением *Смотри*(ПАМЯТЬ ПЕРЕВОДОВ, ПЕРЕВОДЧЕСКАЯ ПАМЯТЬ) произойдет создание обратной связи *Синоним*(ПЕРЕВОДЧЕСКАЯ ПАМЯТЬ, ПАМЯТЬ ПЕРЕВОДОВ) (если таковой еще не существует), а также для аскриптора ПАМЯТЬ ПЕРЕВОДОВ будет обеспечиваться запрет на создание связей «Смотри» и «Синоним» с другими дескрипторами.

4.3 Обеспечение доступа к контенту тезауруса

Удобный доступ к терминам тезауруса обеспечивается пользовательским веб-интерфейсом, также предоставляемым технологией построения порталов научных знаний. В этом интерфейсе содержимое тезауруса представляется пользователю в виде сети взаимосвязанных информационных

Дескриптор	
название	переводческая память
язык	русский
релятор	
определение 1	Инструмент, который предлагает варианты перевода предложений или более крупных сегментов текста. Программа представляет собой базу данных сегментов текста на языке оригинала и их эквивалентов на языке перевода. Фактически, это компьютеризированные параллельные корпуса и инструменты их использования.
автор словарной статьи	Кононенко И.С.
комментарий	Среди переводчиков наиболее распространен лобовой перевод с англ. «память переводов», на Диалоге он вообще не представлен, этот более удачный.
Связи объекта	
Встречается дескриптор в (SourceDescriptor)	
Источник	частота
Коллекция текстов Диалог 2000-2010	8
Сайт Центра переводов СПбГУ им. А.В. Федорова	
Выше род (BTG)	
Дескриптор	Аспект деления иерархии
средство автоматизации перевода	
Выше целое (BTP)	
Дескриптор	
автоматизированное рабочее место переводчика	
Дается определение в (SourceDef)	
Источник	определение
Сайт кафедры перевода и переводоведения ТюмГУ	1
Подобласть знаний (Sub Area)	
Подобласть знаний	
2.3.1. Машинный перевод – Machine translation	
2.4.1. Корпуса текстов – Text corpora	
Синоним (Syn)	
Аскриптор	
архив переводов	
база прецедентных переводов	
банк переводных соответствий	
накопитель переводов	
память переводов	
Эквивалент на другом языке (Trans)	
Дескриптор	
translation memory	

Представление термина «Переводческая память»

объектов — элементов тезауруса: терминов (дескрипторов и аскрипторов) и описаний источников терминов и их определений. Набор атрибутов терминов и связей, установленных между ними, соответствует структуре тезауруса, описанной в подразд. 2.1.

При навигации по тезаурусу обеспечивается возможность выбора необходимых пользователю терминов, детального просмотра их описаний (тезаурусных статей), а также описаний источников (публикаций или коллекций текстов), в которых встречается термин и/или его определение.

Пользователь может указать, какой тип информации его интересует — все термины, дескрипторы, аскрипторы, подобласти знаний или источники терминов. При этом ему выдается упорядоченный по алфавиту полный список имеющихся в тезаурусе объектов выбранного класса, который отображается в виде html-страницы, содержащей набор ссылок на эти объекты.

Информация о конкретном объекте и его связях также отображается в виде html-страницы (см. рисунок). При этом объекты, связанные с данным объектом, представляются на его странице в ви-

де гиперссылок, по которым можно перейти к их детальному описанию.

Дальнейшая навигация по тезаурусу представляет собой процесс перехода от одних объектов тезауруса к другим по заданным между ними связям, отражающим существующие между ними — тезаурусные (между терминами) или библиографические (между терминами и источниками) — отношения.

Для обеспечения доступа к содержимому тезауруса из внешних систем разработан программный интерфейс, благодаря которому тезаурус может использоваться при решении задач индексирования и поиска текстовых документов по КЛ.

Интерфейс поддерживает поиск терминов в тезаурусе по типу, наименованию или части наименования. Для каждого термина-дескриптора или аскриптора можно получить список связанных терминов по выбранному отношению (синонимия, эквивалентность, ассоциация и т. п.). Дополнительно для каждого термина-дескриптора можно получить его перевод, список терминов-дескрипторов, связанных родовидовыми отношениями в соответствии с аспектом организации иерархии, список подобластей знаний, к которым относится данный термин, а также список источников (текстовых документов или коллекций текстов), в которых описан данный термин.

5 Заключение

В статье рассмотрены методологические аспекты построения русско-английского электронного тезауруса по компьютерной лингвистике, разработанного в соответствии с международными и отечественными стандартами. Описана методика выбора терминов для включения в тезаурус, а также предложены подходы к выбору основного термина-дескриптора из множества синонимичных терминов и подбору парных терминов-эквивалентов. Рассмотрены особенности реализации электронной версии тезауруса, обусловленные использованием в качестве инструмента разработки методологии и программных компонентов технологии построения порталов научных знаний [22, 23].

В настоящее время ведется активная разработка тезаурусных статей и заполнение ими контента электронного тезауруса, который на данный момент включает более 1600 терминов, связанных примерно 8000 семантических отношений, а также описания более 180 источников терминов из 50 подобластей знаний.

Тезаурус ориентирован как на непосредственное использование людьми, желающими обратиться к системе понятий из области КЛ, так и для решения

задач индексирования, тематического рубрицирования и информационного поиска (для этого он снабжен программным интерфейсом).

Тезаурус может использоваться в учебном процессе — в тех вузах страны, где изучается компьютерная лингвистика и/или используются ее результаты. Использование тезауруса в вузах повысит уровень профессиональной подготовки будущих специалистов в сфере КЛ и информационных технологий. По существу, это необходимый и профессионально выполненный инструмент и ресурс обучения, особенно ценный в виду междисциплинарной природы КЛ и полного отсутствия каких-либо русскоязычных учебников и даже методических пособий в этой новой и быстро развивающейся области знаний.

Литература

1. Никитина С. Е. Тезаурус по теоретической и прикладной лингвистике. — М.: Наука, 1978.
2. Демьянков В. З. Англо-русские термины по прикладной лингвистике и автоматической переработке текста. Вып. 2. Методы анализа текста // Тетради новых терминов. № 39. — М.: ВЦП, 1982.
3. Лингвистический энциклопедический словарь / Под ред. В. Н. Ярцевой. — М.: Советская энциклопедия, 1990.
4. Ахманова О. С. Словарь лингвистических терминов. — 3-е изд., стер. — М.: КомКнига, 2005.
5. Кругосвет: Онлайн-энциклопедия, 2001–2009. <http://www.krugosvet.ru>.
6. Языкознание: Информационно-поисковый тезаурус ИНИОН. — М.: ИНИОН РАН, 2007.
7. Аверкин А. Н., Гаазе-Рапопорт М. Г., Поспелов Д. А. Толковый словарь по искусственному интеллекту. — М.: Радио и связь, 1992.
8. Большой энциклопедический словарь (БЭС) / Гл. ред. А. М. Прохоров. — 2-е изд., перераб. и доп. — СПб.: Норинт, 2004.
9. Википедия: Свободная энциклопедия. <http://ru.wikipedia.org>.
10. Соколова Е. Г., Семенова С. Ю., Кононенко И. С., Загоруйко Ю. А., Кривнова О. Ф., Захаров В. П. Особенности подготовки терминов для русско-английского тезауруса по компьютерной лингвистике // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам ежегодной междунар. конф. «Диалог» (Бекасово, 25–29 мая 2011). — М.: РГГУ, 2011. Вып. 10(17). С. 644–655.
11. ISO 5964-1985. Documentation — Guidelines for the establishment and development of multilingual thesauri, IDT. (Revised by: ISO/DIS 25964-1. Under development.)

12. ISO 2788-1986. Documentation — Guidelines for the establishment and development of monolingual thesauri. Ed. 2.
13. ГОСТ 7.25-2001. Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления. Введен в действие с 1 июля 2002 г.
14. ANSI/NISO Z39.19-2005 Guidelines for the construction, format, and management of monolingual controlled vocabularies: Periodic review.
15. ГОСТ 7.24-2007. Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению. Введен в действие с 1 июля 2008 г.
16. Мдивани Р. Р. О разработке серии тезаурусов по социальным и гуманитарным наукам // НТИ, 2004. Сер. 2. № 7. С. 1–9.
17. Лукашевич Н. В. Тезаурусы в задачах информационного поиска. — М.: Изд-во Московского ун-та, 2011.
18. Диалог: Сайт международной конференции. <http://www.dialog-21.ru>.
19. Сидорова Е. А. Многоцелевая словарная подсистема извлечения предметной лексики // Компьютерная лингвистика и интеллектуальные технологии: Труды междунар. конф. Диалог'2008. — М.: РГГУ, 2008. Вып. 7(14). С. 475–481.
20. Перерва В. М. О принципах и проблемах отбора терминов и составления словника терминологических словарей // Проблематика определений терминов в словарях разных типов. — Л., 1976. С. 190–204.
21. EAMT (The European Association for Machine Translation). <http://www.eamt.org>.
22. Загорулько Ю. А., Боровикова О. И. Подход к построению порталов научных знаний // Автометрия, 2008. Т. 44. № 1. С. 100–110.
23. Загорулько Ю. А. Технология разработки порталов научных знаний // Программные продукты и системы, 2009. № 4. С. 25–29.
24. Андреева О. А., Боровикова О. И., Булгаков С. В., Загорулько Ю. А., Сидорова Е. А., Циркин Б. Г., Холушкин Ю. П. Археологический портал знаний: содержательный доступ к знаниям и информационным ресурсам по археологии // КИИ-2006: Труды 10-й Национальной конф. по искусственному интеллекту с международным участием. — М.: Физматлит, 2006. Т. 3. С. 832–840.
25. Боровикова О. И., Загорулько Ю. А., Загорулько Г. Б., Кононенко И. С., Соколова Е. Г. Разработка портала знаний по компьютерной лингвистике // КИИ-2008: Труды 11-й Национальной конф. по искусственному интеллекту с международным участием. — М.: ЛЕНАНД, 2008. Т. 3. С. 380–388.

КЛАССИФИКАЦИЯ И ИСПОЛЬЗОВАНИЕ СЕМАНТИЧЕСКИХ СВЯЗЕЙ МЕЖДУ ИНФОРМАЦИОННЫМИ ОБЪЕКТАМИ В НАУЧНЫХ ЭЛЕКТРОННЫХ БИБЛИОТЕКАХ*

М. Р. Когаловский¹, С. И. Паринов²

Аннотация: Обсуждается подход, обеспечивающий повышение информационной ценности контента научной электронной библиотеки благодаря поддержке классифицированных семантических связей между содержащимися в ней информационными объектами. Рассматривается реализация предлагаемого подхода на основе отечественной системы Соционет, объединяющей большое число научных электронных библиотек и являющейся де-факто институциональным исследовательским информационным пространством Отделения общественных наук Российской академии наук.

Ключевые слова: электронная библиотека; информационный объект; коллекция информационных ресурсов; семантическая связь; классификатор связей; онтология; наукометрия

1 Введение

Информационные объекты, содержащиеся в научных электронных библиотеках (статьи, книги, персональные профили авторов, профили организаций и др.), имеют многие другие связи (отношения) друг с другом, помимо обычно поддерживаемых средствами управления информационными ресурсами и отображаемых программными интерфейсами электронных библиотек. Большое число видов связей из-за неразвитости технологий пока остается за рамками электронных библиотек, часто только в сознании исследователей. Такие связи являются ненаблюдаемыми и не фиксируются в цифровой форме.

К обычно отображаемым связям между информационными объектами электронных библиотек относятся, например, связи между научными публикациями (статья, книга и т. п.) и персональными профилями их авторов, а также с профилями организаций, в которых были получены соответствующие результаты исследований. Кроме этого, статьи, снабженные кодами тематических классификаторов, имеют связи с тематическими рубриками классификаторов научных дисциплин. Все чаще в электронных библиотеках встречаются связи между статьями и комментариями их читателей. В некоторых крупных электронных библиотеках поддерживаются связи цитирования, профили авторов и связи публикаций с ними. Механизмы для этой цели имеются, в частности, в электронной библиотеке ACM (Association for Computing Machin-

ery) [1], научной библиотеке eLibrary РФФИ [2], в Академии Google (Google Scholar) [3]. Однако в этих и в других аналогичных случаях связи цитирования не несут никакой информации, кроме самого факта цитирования, не характеризуют семантики отношения между цитирующим и цитируемым текстовым документом. Будем называть такие связи «*немymi*».

Связи между информационными объектами в электронных библиотеках, в том числе и связи цитирования, обладают семантикой, и она может быть явно описана способом, доступным для пользователей и системных механизмов, и продуктивно использована. На ее основе может, в частности, формироваться более дифференцированная наукометрическая статистика, учитывающая позитивное, негативное или иное отношение автора цитирующего документа к цитируемому. Связи с явно описанной семантикой будем называть далее *семантическими связями*.

Как уже отмечалось, между содержащимися в научной электронной библиотеке информационными объектами могут поддерживаться не только семантические связи, такие как «*немые*» связи цитирования и другие, упоминаемые ранее. Например, семантическая связь может создаваться для выражения мнения автора одного из информационных объектов или экспертов о существовании некоторого отношения между контентом двух информационных объектов в ситуации, даже когда этот факт не отмечен явным образом в контенте рассматриваемых информационных объектов.

* Работа поддерживается РГНФ, проект 11-02-12026-в.

¹ Институт проблем рынка Российской академии наук, kogalov@cemi.rssi.ru

² Центральный экономико-математический институт Российской академии наук, sparinov@gmail.com

Явным образом описанные и поддерживаемые в библиотеке семантические связи могут быть *классифицированы* на основе характера отношений между информационными объектами — участниками связей. Введение классификации приводит к образованию многослойной семантической структуры контента электронной библиотеки, каждый слой которой соответствует некоторому классу семантических связей. Такая структура может служить источником информации для проведения качественно новых наукометрических измерений, для исследования структурных свойств корпуса знаний в различных областях науки, представительным образом отраженного в контенте электронной библиотеки.

Онлайновый режим функционирования научной электронной библиотеки позволяет реализовать ее систему управления таким образом, чтобы не только поддерживалась семантическая структура контента и обрабатывались пользовательские запросы, касающиеся ее характеристик, но и предоставлялась пользователям возможность самостоятельно в децентрализованном режиме описывать и создавать семантические связи. Могут быть также предусмотрены мониторинг состояния структуры связей и автоматическое оповещение авторов информационных объектов о том, что некоторый их объект стал участником вновь учрежденной связи или что ликвидирована существующая связь, в которой этот объект являлся участником. Благодаря этому автор информационного объекта, получивший указанное оповещение, стимулируется тем самым реагировать на эту ситуацию, если событие, о котором он информируется, противоречит его представлениям.

Таким образом, в онлайн-электронной библиотеке, в которой поддерживаются классифицированные семантические связи, может быть обеспечен комплекс новых возможностей:

- поддержка многослойной структуры семантических связей;
- создание новых связей и аннулирование существующих связей не только администраторами информационных ресурсов, но и пользователями системы в децентрализованном режиме;
- формирование дифференцированной по классам семантических связей статистики связей, в частности, касающейся связей цитирования;
- оповещение авторов представленных в электронной библиотеке информационных объектов об их включении в новые связи или об аннулировании некоторых связей, в которых они были участниками.

Обладающая такими возможностями информационная среда обеспечивает качественно новые технологии для научной и научно-организационной деятельности, открывает новые возможности для коммуникаций в научном сообществе. В предлагаемой статье обсуждается подход авторов к созданию такой среды, реализуемый в научном информационном пространстве Соционет [4].

Остальная часть статьи организована следующим образом. В разд. 2 уточняется постановка рассматриваемой в статье проблемы и предлагаются пути ее решения. В разд. 3 обсуждается вопрос о классификации связей и дается краткий обзор известных исследований в рассматриваемой области. В разд. 4 обсуждаются принципы представления семантических связей в электронной библиотеке как самостоятельных информационных объектов. Свойства семантических связей рассматриваются в разд. 5. В разд. 6 обсуждается реализация предлагаемого в статье подхода в среде системы Соционет. В заключении подводятся итоги обсуждения.

2 Уточнение постановки проблемы

Коллекции информационных ресурсов традиционных научных электронных библиотек состоят из множества объектов определенных типов: электронных версий публикаций, изданных типографским способом, научных отчетов, рабочих записок, рецензий, авторефератов диссертаций, полных текстов диссертационных работ, таблиц научных данных, карт звездного неба и др. Коллекции могут содержать также сведения об авторах представленных в них публикаций, об организациях, в которых они работают, и информационные объекты других типов.

В последние годы на основе библиографических ссылок, содержащихся в публикациях, которые выпускаются в авторитетных периодических изданиях, начали создаваться индексы цитирования, обеспечивающие формирование библиометрической статистики. Связи цитирования в текстовых публикациях обычно представляются неструктурированным образом в виде списка используемой литературы. Они не являются при этом носителями какой-либо информации, кроме указания целевой публикации ссылки и существования самого факта ссылки. Однако с фактом цитирования связана еще и некоторая не отображаемая при этом семантика, выражающая отношение автора цитирующего документа к цитируемому источнику или какое-либо иное семантическое отношение между цити-

рующей и цитируемой публикацией. Как правило, связи цитирования аннотируются в тексте публикации, и в таких случаях семантика связи все-таки описана, но в неструктурированном виде. Это создает значительные сложности для ее анализа, и в существующих системах такой анализ обычно не производится.

Наряду со связями цитирования между информационными объектами научных электронных библиотек существуют разнообразные другие семантические связи. Например, связь может указывать, что ее целевой информационный объект содержит научные результаты, базирующиеся на результатах, описанных в исходном объекте связи, или что в исходном объекте связи опровергается результат, изложенный в ее целевом объекте. Связь может также указывать, что ее исходный информационный объект является новой редакцией целевого объекта или представляет собой его составную часть, например аннотацию.

Существует большое разнообразие семантических связей, которые можно при необходимости поддерживать между информационными объектами в библиотеке. Эти связи выявляются в результате участия представителей научного сообщества в процессах, реализующих их научную и научно-организационную деятельность. К числу основных видов таких процессов можно отнести процессы систематизации, классификации и упорядочения корпуса научных знаний (например, при подготовке аналитических обзоров), процессы научной оценки опубликованных результатов (рецензирование работ), процессы продуцирования нового научного знания, процессы создания научных произведений, научно-организационные процессы. Именно на основе информации, рождающейся в результате участия в процессах перечисленных видов, пользователь электронной библиотеки может прийти к выводу о целесообразности создания тех или иных семантических связей между представленными в ней некоторыми информационными объектами.

Определяемые явным и структурированным образом семантические связи могут быть представлены и могут динамически поддерживаться как самостоятельные информационные объекты электронной библиотеки. Такие объекты содержат идентификаторы участвующих в них информационных объектов и значения других атрибутов. Объекты-связи могут быть классифицированы, и их свойства определяются значениями атрибутов, специфических для каждого класса.

В результате определения явным образом описанных классифицированных семантических связей, как уже отмечалось, порождается многослойная семантическая структура контента библиотеки.

При этом каждому классу связей соответствует некоторый слой этой структуры, который наряду с полной структурой связей может служить для наукометрических измерений и анализа. В частности, могут поддерживаться слои, отображающие структуру продуцирования научных результатов и другие содержательные отношения между научными публикациями, например связи оценки публикаций научными сотрудниками, связи между частями научных публикаций, связи научно-организационного характера (научное учреждение – сотрудники – авторы публикаций, авторы – публикации) и др.

Анализ структуры таких связей в научной электронной библиотеке позволяет решать также ряд задач, связанных с поддержкой научно-организационной деятельности, позволяет авторам публикаций более продуктивно использовать имеющиеся в электронной библиотеке научные информационные ресурсы, дает возможность извлекать из контента библиотеки ценную информацию, не содержащуюся в отдельных информационных объектах. Например, можно получать полезные наукометрические сведения, а также сведения, основанные на анализе топологии структуры связей, которые достаточно сложно получить иным путем. Исследование топологии связей научных публикаций позволяет, в частности, анализировать процесс формирования научных направлений и школ, влияние публикаций тех или иных исследователей на формирование научных направлений или теорий. Поддержка структуры семантических связей обеспечивает также дополнительные (навигационные) пути доступа пользователей к информационным объектам библиотеки. Другое направление, где необходима поддержка семантических связей между информационными объектами электронной библиотеки, — это технология «живых» публикаций, подробно рассмотренная в [5, 6].

Для эффективного использования новых возможностей, которые обеспечиваются благодаря поддержке в онлайн-электронной библиотеке многослойной структуры семантических связей представленных в ней информационных объектов, необходимо, чтобы система управления электронной библиотекой удовлетворяла определенным требованиям. В частности, она должна быть способна не только обрабатывать запросы относительно семантической структуры контента, но и располагать механизмами, позволяющими пользователям самостоятельно устанавливать, модифицировать или удалять семантические связи в рамках их полномочий, а также обеспечивать мониторинг изменений состояния структуры семантических связей. Механизмы мониторинга позволяют автоматически оповещать авторов информацион-

ных объектов о том, что некоторый их информационный объект стал участником вновь учрежденной связи или что ликвидирована существующая связь, в которой он являлся участником, либо об изменениях значений ее атрибутов.

Семантическое структурирование контента научных электронных библиотек представляет значительно больший интерес, если оно поддерживается на представительном репозитории научных информационных объектов. Одним из популярных подходов к созданию крупных репозитория научных публикаций, позволяющих интегрировать коллекции ряда научных и образовательных учреждений, является подход, основанный на технологии открытых архивов [7]. Поддержка и исследование семантической структуры в создаваемом на ее основе крупном интегрированном контенте, формируемом на федеративных принципах рядом исследовательских организаций, дает возможность изучать структуру результатов научных исследований не только отдельных научных коллективов или школ, но и целых направлений науки и областей знаний.

Обеспечение возможностей поддержки в научных электронных библиотеках явно представленных классифицированных семантических связей между содержащимися в них информационными объектами в сочетании с методами мониторинга изменений структуры этих связей и основанными на такой структуре новыми функциональными возможностями является, по мнению авторов, весьма перспективным новым направлением развития научных электронных библиотек. Для эффективного практического использования обсуждаемых возможностей необходимо решить следующие задачи:

- разработать способы и конкретные форматы представления семантических связей между информационными объектами электронной библиотеки в виде самостоятельных информационных объектов специального типа;
- создать классификатор семантических связей, которые целесообразно поддерживать в научных электронных библиотеках;
- определить операционные возможности, которые должна обеспечивать система управления научной электронной библиотекой для того, чтобы извлекать в достаточно полной мере информацию, содержащуюся в структуре семантических связей представленных в ней информационных объектов.

В данной работе обсуждается предлагаемый авторами подход к решению этих задач и его реализация в среде крупного отечественного онлайн-ового

научно-образовательного пространства, поддерживаемого системой Соционет [4], основанного на технологии открытых архивов и содержащего большой объем информационных ресурсов по социально-экономической тематике. Соционет функционирует уже более десяти лет и приобрела в последние годы де-факто институциональный статус в Отделении общественных наук. Информационное пространство Соционет содержит также публикации ряда образовательных учреждений и других организаций. Соционет стала полигоном для проведения исследований в области перспективных технологий электронных библиотек. Постоянно проводятся работы по расширению разнообразия типов представляемых в этой системе информационных ресурсов и развитию функциональности механизмов управления библиотекой. Основные идеи данной работы сформировались на основе более ранних публикаций [8–11] и были детально представлены на конференции RCDL-2011 [12].

3 Классификация связей и известные работы в данной области

Проблемы структуризации крупных коллекций информационных ресурсов электронных библиотек и классификации семантических связей в последние годы привлекают большое внимание исследователей. Известны попытки систематической классификации семантических связей между единицами информационных ресурсов и/или их компонентами, предпринятые для использования в электронных библиотеках, издательских системах, для представления знаний в среде Семантического Веба. Рассмотрим наиболее известные разработки в этой области.

Прежде всего следует упомянуть работы по распознаванию и классификации используемых в научных статьях языковых конструкций (для английского языка и отдельных научных дисциплин), проводимые средствами программного обеспечения компании Xerox. Они позволили эмпирическим путем выявить некоторые устойчивые виды семантических отношений, существующих как между разделами внутри научной статьи, так и между статьями и цитируемыми в ней материалами [13, 14]. Эмпирическая классификация поводов цитирования в научных статьях предлагается также в [15]. В этой работе выделен ряд их типичных вариантов: «слабость цитируемого подхода», «автор использует цитируемую работу как основу или исходную точку» и др. Другой подход к развитию

классификации семантических связей реализуется в исследованиях модульности научных документов [16].

К рассматриваемому направлению примыкает также рекомендация SKOS (Simple Knowledge Organization System) [17] консорциума W3C. Эта спецификация предназначена для поддержки использования систем организации знаний, таких как тезаурусы, схемы классификации, таксономии и рубрикаторы (Subject Heading Systems) в среде Семантического Веба. Для этой цели определяется концептуальная схема (в спецификации она называется *общей моделью данных*) для совместного использования и связывания систем организации знаний средствами Веба. Унификация концептуальной схемы, определяемой спецификацией SKOS, создает возможности для относительно нетрудоемкой интеграции существующих систем организации знаний в Семантический Веб.

Специалистами в области биомедицины из Оксфордского и Болонского университетов разработан модульный онтологический комплекс SPAR (the Semantic Publishing and Referencing Ontologies) [18, 19]. Он состоит из восьми независимых повторно используемых детализированных онтологий. Фактически каждая из них представляет собой таксономию, описанную на языках OWL2 DL и RDF консорциума W3C. Первые четыре из них (FaBiO — FRBR-aligned Bibliographic Ontology, где FRBR — Functional Requirements for Bibliographic Records); CiTO — Citation Typing Ontology; BiRO — Bibliographic Reference Ontology; C4O — Citation Counting and Context Characterization Ontology) полезны для описания библиографических объектов, библиографических записей и источников в списках литературы в публикациях, связей цитирования, контекстов цитирования и их связей с релевантными разделами цитируемых публикаций, а также для организации ссылок в библиографиях, в списках источников и в библиотечных каталогах. Остальные онтологии (DoCO — Document Components Ontology; PRO — Publishing Roles Ontology; PSO — Publishing Status Ontology; PWO — Publishing Workflow Ontology) служат для создания структурированных управляемых словарей классов компонентов документов, ролей публикаций, состояний публикаций и потоков работ в издательских процессах.

В Главном госпитале Массачусетса и в Медицинской школе в Гарварде разработана онтология SWAN (Semantic Web Applications in Neuromedicine) [20]. Как и SPAR, эта онтология состоит из набора онтологий-модулей. Онтологии, входящие в состав SWAN, также описаны на языке описания онтологий OWL DL. Как указывается в спецификации SWAN, цель этой онто-

логии — обеспечение в рамках Семантического Веба комфортной среды, называемой авторами *социально-технической экосистемой*, которая позволяет создавать и сохранять семантический контекст научных коммуникаций, обеспечивает доступ к нему, его интеграцию, а также обмен неструктурированной или слабоструктурированной цифровой научной информацией.

Нужно отметить здесь важную тенденцию конструирования сложных онтологий, предназначенных для достаточно широкой сферы применения: они строятся по модульному принципу. Такой подход облегчает их повторное использование. Обычно не требуется использовать полную онтологию и берется только нужный ее модуль. При этом модульность облегчает также интеграцию с другими онтологиями. Примером такой интеграции может служить комплекс SPAR, в котором использованы элементы SWAN. В свою очередь, в SWAN используется SKOS.

Следует, наконец, упомянуть также имеющий отношение к обсуждаемому в этом разделе вопросу проект CERIF (Common European Research Information Format) [21], который в 1980–1990-е годы реализовывался при поддержке Европейской комиссии, а в 2000 г. был передан ею под опеку международной научной организации euroCRIS. Главная цель этого проекта фактически заключается в создании стандарта так называемой *полной модели данных* (Full Data Model), которая рассматривается как единая основа создания информационных систем (Current Research Information Systems, CRIS) для поддержки научно-организационной деятельности в разных странах и научных организациях. Благодаря стандартизации модели данных обеспечивается интероперабельность таких систем. В последнее время в проекте CERIF уделяется большое внимание семантическим аспектам созданной модели. Для этой цели разработаны онтологии CERIF [22, 23].

Рассмотренные результаты в области классификации возможных семантических связей между научными публикациями и/или другими продуктами научной деятельности могут использоваться в качестве основы для семантического структурирования контента научных электронных библиотек. В разработке классификатора семантических связей в обсуждаемом в этой работе проекте авторы использовали фрагменты рассмотренных онтологий — CiTO, DoCo, SWAN, SKOS и CERIF. Наиболее существенную часть предлагаемого классификатора определяют фрагменты онтологий CiTO и DoCo.

Онтология CiTO [24, 25] обеспечивает возможности для характеристики природы связей цитиро-

вания, как фактологических (например, «цитирует как источник данных» или «цитирует как основополагающую»), так и риторических (например, «уточняет» или «опровергает»). При этом учитываются как непосредственные и явные связи цитирования, так и косвенные и неявные. Онтология DoCO [26] классифицирует составные части документов. Она предоставляет структурированный управляемый словарь классов их компонентов, например «Введение», «Обсуждение», «Благодарности», «Список использованных источников», «Приложение» и т. д.

Результаты рассмотренных исследований могут быть использованы для классификации некоторых видов связей на множестве не только текстовых научных информационных объектов. Это обстоятельство имеет в рассматриваемом случае существенное значение, поскольку, как отмечалось ранее, интерес представляют также связи, участниками которых являются профили организаций и их сотрудников — авторов и пользователей библиотеки, а также информационные объекты других типов, не являющиеся текстовыми документами.

Классификатор связей в системе Соционет предусматривает разбиение множества семантических связей информационных объектов (текстовых информационных объектов, профилей пользователей и организаций и информационных объектов других типов) на категории (оценочные связи, научно-организационные связи и др.). Каждой категории соответствует некоторый набор классов связей. Эти наборы представляются в виде словарей классов связей. При необходимости в процессе функционирования системы может пополняться состав категорий и словари могут дополняться новыми классами связей. Более подробно принципы организации и содержание классификатора семантических связей, используемого в системе Соционет, обсуждается в [12].

Помимо рассмотренных выше работ, появляются также новые публикации, посвященные затронутой проблеме. Однако авторам не известны проекты, в которых реализован описанный выше комплекс возможностей использования классифицированных семантических связей между информационными объектами научных электронных библиотек.

4 Семантические связи как информационные объекты библиотеки

В электронных библиотеках традиционно с помощью гиперссылок поддерживаются связи между

каталогами и описываемыми в них информационными объектами. В системе Соционет таким же образом поддерживаются связи цитирования, связи с профилями авторов и организаций и некоторые другие. Для этого в Соционет имеются метаданные, описывающие информационные объекты, их авторов (профили авторов), коллекции информационных ресурсов, организации — места работы авторов (профили организаций) и др. В таком случае связи между информационными объектами представляются как атрибуты метаданных, описывающих эти информационные объекты. С использованием связей такого вида можно анализировать структуру связей, осуществлять наукометрические измерения, визуализировать структуру связей.

Однако при таком традиционном способе представления связей явным образом не отображается семантика связей. Например, для связи цитирования одного информационного объекта с другим отсутствует информация, характеризующая цель цитирования, оценку цитируемой работы и другие характеристики. Предлагаемая далее модель связей между информационными объектами научной электронной библиотеки устраняет это ограничение.

В общем случае связи могут представляться двумя способами. При использовании первого способа, представленного выше, данные, описывающие связи, содержатся в метаданных одного из связываемых объектов, например в метаданных исходного объекта связи. Однако поскольку в электронной библиотеке, построенной на федеративных принципах, изменять метаданные может только их автор или уполномоченный автором администратор информационных ресурсов, то при этом способе только они и могут создавать связи этого объекта с другими информационными объектами.

При втором способе создаваемые связи представляются как самостоятельные информационные объекты. Такой способ является более универсальным и предпочтительным, так как он охватывает все многообразие возможных ситуаций, обеспечивает более богатые возможности анализа структуры связей, которые значительно проще реализуются, и он позволяет создавать связи любому пользователю, поскольку при этом не затрагиваются недоступные ему метаданные связываемых объектов.

Описание связи в обоих представлениях должно включать уникальный идентификатор целевого объекта связи, а также может включать атрибуты, характеризующие семантику связи, различного рода комментарии и пр. Если связь создается как самостоятельный информационный объект,

то ее описание в дополнение к уже перечисленному должно включать: уникальный идентификатор данного объекта-связи в библиотеке; уникальный идентификатор пользователя, создающего данную связь; уникальный идентификатор исходного объекта связи (рассматриваются ориентированные бинарные связи), а также дата создания или изменения связи. Для описания семантики связи используется имя класса связи, выбираемое из поддерживаемых контролируемых словарей, а также значения свойств конкретного экземпляра связи, определяемые пользователем. Полномочия на создание связей между информационными объектами предоставляются только зарегистрированным в библиотеке пользователям, что обеспечивает автоматическую фиксацию идентификатора пользователя, создающего связь, при его входе в систему.

Рассмотрим процедуру создания связи между двумя информационными объектами в системе Соционет, в которой реализованы оба способа представления связей. При первом способе создание связи осуществляется автором исходного информационного объекта связи или его представителем. Рассмотрим процедуру второго способа.

Множество параметров, влияющих на создание связи, включает: тип исходного объекта связи; тип целевого объекта связи; множество категорий связей, учрежденных в системе для заданной пары типов связываемых объектов; множество словарей классов связей, предусмотренных в системе для связей заданной категории; множество классов связей в словаре, выбранном для создания связи между объектами заданных типов.

Рассматриваемая процедура состоит из следующих шагов:

1. Пользователь выбирает пару связываемых информационных объектов.
2. Из множества категорий связей, предусмотренных в системе для выбранной пары типов объектов, выбирается конкретная категория. Если подходящей категории не существует, пользователь имеет возможность предложить новую категорию и предоставить соответствующий ей словарь классов связей для включения в систему. Это предложение вступит в силу только после одобрения администратором системы.
3. Если подходящая категория связей выбрана, то открывается соответствующий словарь классов связей.
4. Если в словаре имеется подходящий класс связей, характеризующий требуемое семантическое отношение между заданной парой объектов, то пользователь его выбирает. Если же такой класс отсутствует в данном словаре, пользователь может предложить подходящий класс связей для пополнения данного словаря. Предложение вступит в силу только после одобрения его администратором системы или соответствующего словаря.
5. По желанию пользователь может привести в описании связи комментарий, объясняющий мотивы ее создания.
6. Сформированный информационный объект-связь сохраняется. При этом система запрашивает у пользователя, в какую его коллекцию следует поместить созданный объект, а также уникальный идентификатор этого объекта в соответствующей коллекции.

Рассмотренная процедура обеспечивает создание информационного объекта, представляющего требуемую связь среди других объектов библиотеки. При этом также осуществляется проверка непротиворечивости семантики новой связи с уже существующими связями между данными объектами, созданными тем же пользователем.

Хотя формирование семантических связей между информационными объектами требует определенных трудозатрат, в результате информативность контента научной электронной библиотеки существенно повышается. Создаются также дополнительные возможности для анализа семантической структуры контента.

Поддержка развитой структуры семантических связей в научной электронной библиотеке с достаточно представительным контентом позволяет в результате их анализа осуществлять наукометрические измерения, использовать технологии «живых» публикаций [5, 6], а также получать качественно новую информацию о развитии научных знаний в конкретных областях исследований и о вкладе отдельных ученых.

В описанной выше процедуре предполагается, что любой зарегистрированный пользователь научной электронной библиотеки может создавать связи между любыми ее информационными объектами. При определении семантики связей их создатель выражает свое субъективное мнение, которое в некоторых случаях может вызывать несогласие или протест как авторов объектов, участвующих в данных связях, так и других членов научного сообщества. Например, могут вызывать протесты случаи, когда устанавливаются семантические связи, несущие негативную оценку некоторого научного произведения (опровержение, высмеивание, обвинение в плагиате и т. п.).

Как известно, научная истина устанавливается в борьбе мнений. Поэтому если научное сообщество начинает использовать подобные технические средства, то с учетом потенциального конфликта интересов научная среда должна предоставлять ученым равные права и одинаковый доступ к использованию этих средств, а также надежную фиксацию профессиональной и социально-этической ответственности ученого за характер использования им данных средств.

Для выполнения данных принципов, по мнению авторов, крайне важно обеспечить модерирование всех создаваемых связей с точки зрения соблюдения авторами научной этики, а также наличия в создаваемых связях признаков добавленной научной «стоимости» или научного вклада (исключение связей с чисто эмоциональным или ненаучным содержанием).

В системе Соционет пользователи создают связи в своем личном (закрытом от свободного доступа) пространстве. Такое пространство с сервисами для его использования предусматривается для авторов или администраторов информационных ресурсов в системе и называется их Личной зоной. Создаваемые в Личной зоне объекты-связи предлагаются далее для включения в общедоступные информационные ресурсы. Они становятся общедоступными только после одобрения модератором.

5 Свойства семантических связей

Обсуждаемая в данной работе структура семантических связей, формируемая и поддерживаемая над контентом электронной библиотеки, порождается бинарными ориентированными семантическими связями между информационными объектами библиотеки, составляющими ее коллекции информационных ресурсов.

Как уже отмечалось, семантические связи, определяемые в библиотеке явным образом в виде структурированных данных, представляются и могут динамически поддерживаться как самостоятельные информационные объекты.

Информационные объекты-связи категоризируются, как было описано выше, и в рамках каждой категории классифицируются в соответствии с их семантикой. Таким образом, каждый экземпляр создаваемых в библиотеке связей относится к какой-либо категории, а в рамках категории — к какому-либо классу связей этой категории. Свойства экземпляров объектов-связей задаются значениями атрибутов, определенных для соответствующих классов связей. Между двумя информационными

объектами библиотеки может быть определено несколько связей одной или нескольких категорий.

Каждому экземпляру объекта-связи при его создании присваивается некоторое значение уникального идентификатора, а значения его атрибутов, наряду с другими возможными свойствами, указывают категорию и класс представляемой им связи, идентификатор пользователя, который создает этот объект-связь, идентификаторы исходного и целевого информационных объектов библиотеки, участвующих в данной связи, дату ее создания.

Структура семантических связей, поддерживаемых в электронной библиотеке, динамична. Могут создаваться новые, а также обновляться или ликвидироваться существующие связи — мнения авторов связей могут изменяться с течением времени. Динамичность структуры связей обусловлена и пополнением контента библиотеки новыми информационными объектами — потенциальными участниками связей.

В некоторых категориях связей существуют классы связей с противоречивой семантикой. Например, к категории оценочных связей относятся связи между информационными объектами, которые выражают одобрение или согласие исходного объекта с целевым, а также связи, выражающие опровержение результатов, представленных в целевом информационном объекте. Естественно, что между двумя информационными объектами не могут одновременно существовать связи этих двух классов, созданные одним и тем же пользователем. Возникновение таких ситуаций должны предотвращать системные механизмы библиотеки. В то же время вполне возможны семантически противоречивые связи между двумя информационными объектами, созданные разными пользователями. Системные механизмы должны контролировать выполнение и некоторых других ограничений на создание, обновление и ликвидацию экземпляров связей. К ним относятся, в частности, ограничения доступа — для выполнения таких операций пользователь должен обладать необходимыми полномочиями.

Каждая семантическая категория связей между информационными объектами библиотеки и каждый относящийся к ней класс связей, как уже отмечалось, образуют некоторые слои в структуре связей. Таким образом, в электронной библиотеке, механизмы которой обладают рассматриваемой функциональностью, поддерживается многослойная структура семантических связей принадлежащих ей информационных объектов, которая при достижении достаточной ее представительности становится весьма значимым полигоном для ана-

лиза и поддержки научной и научно-организационной деятельности.

6 Реализация предлагаемого подхода в системе Соционет

Для формирования в электронной библиотеке и продуктивного использования многослойной структуры семантических связей информационных объектов ее контента необходимо, чтобы система управления библиотекой включала механизмы, предоставляющие необходимые операционные возможности. Кратко рассмотрим состав и функции таких механизмов, которые предусмотрены в системе Соционет.

Механизмы формирования и поддержки словарей связей. Классификатор семантических связей в системе Соционет, как уже отмечалось, имеет модульную структуру и представлен в виде совокупности управляемых словарей.

Основу разработанных словарей составляют упоминавшийся выше комплекс онтологий SPAR (в частности, онтологии CiTO и DoCo), спецификация SKOS консорциума W3C, онтология проекта SWAN, а также один из разделов CERIF, определяющий семантику связей.

Каждый из словарей соответствует некоторой предусмотренной в классификаторе категории связей и содержит имена относящихся к ней классов связей. Рассматриваемые механизмы позволяют системному администратору формировать и модифицировать эти словари. Пользовательский интерфейс механизмов создания связей предоставляет доступ к словарям и справочной информации, необходимой для их корректного использования.

Механизмы управления связями. Эти механизмы позволяют авторизованному пользователю создавать в модерируемом режиме связи между информационными объектами библиотеки. Как указывалось выше, связи создаются как информационные объекты специального типа. При создании новой связи используются управляемые словари классов связей. Новая связь создается только при условии, если ее создание не нарушает заданных ограничений (см. разд. 5).

В системе Соционет поддерживается множество типов информационных объектов — статьи, монографии, диссертации или авторефераты диссертаций, профили пользователей, профили организаций, рубрикаторы, научные артефакты, цитаты, информационные объекты-связи и т. п. Для каждой пары типов информационных объектов допустимы только определенные классы связей. При попытке

создания конкретного экземпляра связи проверяется его допустимость.

Механизмы управления связями позволяют также ликвидировать существующие связи и обновлять значения их атрибутов. Можно, например, изменить текст комментария. В рассматриваемой группе механизмов важное место занимает механизм мониторинга состояния структуры связей. При появлении новой связи, удалении связи или некоторых изменениях атрибутов связей этот механизм генерирует сообщения авторам информационных объектов — участников таких связей, стимулируя тем самым их реакцию на эти события.

Фактически предлагаемый подход предусматривает создание в системе Соционет наряду с уже много лет функционирующим открытым крупным репозиторием метаданных научных статей, монографий, персональных профилей, профилей организаций и других информационных объектов также и открытого репозитория семантических связей, который является ценным информационным источником структурного анализа представленного в системе корпуса научных знаний.

Механизмы обработки запросов. Эти механизмы выполняют довольно большой набор функций, позволяющих получать разнообразную информацию о структуре связей в библиотеке. Прежде всего, это статистическая информация. В системе Соционет имеются развитые сервисы для наукометрических измерений. Они обсуждаются подробно в [8, 27]. Измерения на основе структуры семантических связей существенно обогащают аналитические возможности системы. Можно, например, запросить количество связей заданного класса или некоторой категории, исходящих из данного информационного объекта библиотеки либо входящих в него. Например, можно узнать, сколько имеется положительных или негативных оценок данной работы.

Другая группа запросов позволяет получить перечень информационных объектов, связанных с заданным объектом как исходным или целевым в связях заданных классов или категорий. Запросы этого вида позволяют, например, выяснить, на результаты каких публикаций опирается некоторая конкретная работа или, наоборот, в каких публикациях получены результаты, основанные на данной работе. При этом можно учитывать как непосредственные, так и транзитивные связи. В качестве критерия отбора связей или одного из термов критерия может также использоваться идентификатор автора связей. Таким образом может быть получена разнообразная аналитическая информация о структуре различных областей исследований, вкладе в их

развитие конкретных ученых, о процессе эволюции интересующих областей знаний и т.д. Исследования в этой области планируется развивать на основе системы Соционет.

Следует здесь упомянуть проект SciVal компании Elsevier [28]. Функциональный модуль SciVal Spotlight созданного компанией программного продукта позволяет осуществлять анализ научной деятельности исследовательского учреждения или страны в целом, на основе которого может оцениваться эффективность исследований и могут приниматься стратегические решения. Принятый в этом интересном проекте подход основан на анализе структуры связей цитирования публикаций субъектов научной деятельности, поддерживаемых в индексе цитирования Scopus. Однако при этом используются традиционные «немые» связи — связи, не несущие семантики. В этом смысле предлагаемый авторами подход выгодно отличается, так как обеспечивает более дифференцированный анализ, результаты которого учитывают семантику связей.

Механизмы визуализации и анализа графа связей.

Важную группу запросов составляют запросы операций над полным графом связей. Здесь можно решать различные задачи, связанные как с анализом топологии графа и вычленением подграфов с заданными свойствами, так и с визуализацией подграфов полного графа. Например, можно вычлениить и визуализировать из многослойной структуры связей слой, соответствующий связи некоторого класса, такой как связь, указывающая на использование одного информационного объекта как основополагающего для других объектов. Можно также запросить подграф, образованный связями, относящимися к категории развития научных результатов, и указать, что ему должна принадлежать некоторая имеющаяся в библиотеке общепризнанная основополагающая публикация в некоторой области исследований. Полученный подграф будет характеризовать логику развития данной области науки, если, конечно, в библиотеке будут достаточно основательно представлены публикации, относящиеся к этой области. Еще одним примером операций над полным графом связей библиотеки является операция вычленения из него подграфа связей, созданных данным пользователем, возможно, с указанием в запросе также категории или конкретного класса связей.

Отметим, наконец, что визуализация графа связей или некоторого его подграфа может быть использована для навигации в структуре связей, а также просмотра свойств отдельных экземпляров связей и участвующих в них информационных объектов.

7 Заключение

В работе предложен подход, обеспечивающий повышение информационной ценности контента научной электронной библиотеки путем поддержки в ней классифицированных семантических связей между содержащимися в ее коллекциях информационными объектами, а также создания механизмов управления связями и обработки информации, носителями которой они являются.

Реализующая этот подход технология позволяет более эффективно использовать существующий корпус электронных знаний благодаря визуализации семантических связей между научными произведениями, навигации в такой многослойной семантической структуре, созданию основы для получения качественно новых наукометрических измерений, а также для структурного исследования электронного корпуса научных знаний.

Предлагаемая технология обеспечивает также естественный механизм мотивации научных коммуникаций в исследовательском сообществе в процессе создания и обсуждения новых научных результатов. Она хорошо согласуется также с технологией «живых» публикаций, для поддержки которой применимы реализующие ее механизмы.

Литература

1. ACM Digital Library. <http://dl.acm.org>.
2. Научная электронная библиотека eLibrary.ru. <http://elibrary.ru>.
3. Академия Google. <http://scholar.google.com>.
4. *Паринов С. И.* СОЦИОНЕТ.РУ как модель информационного пространства 2-го поколения // Информационное общество, 2001. Вып. 1. С. 43–45.
5. *Паринов С. И., Коголовский М. Р.* Технология поддержки электронных научных публикаций как «живых» документов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL-2009): Труды XI Всеросс. науч. конф. (Петрозаводск, 17–21 сентября 2009). — Петрозаводск: КарНЦ РАН, 2009. С. 53–58.
6. *Паринов С. И., Коголовский М. Р.* «Живые» документы в электронных библиотеках // Прикладная информатика, 2009. № 6. С. 123–131.
7. Open Archives Initiative. <http://www.openarchives.org>.
8. *Коголовский М. Р., Паринов С. И.* Метрики онлайн-овых информационных пространств // Экономика и математические методы, 2008. Вып. 2. С. 108–120.
9. *Коголовский М. Р., Паринов С. И.* Использование связей цитирования для наукометрических измерений в системе Соционет // Депонировано в Соционет, 2009. <http://socionet.ru/publication.xml?h=repec:rus:rssalc:web-32>.

10. *Parinov S.* The electronic library: using technology to measure and support Open Science // World Library and Information Congress: 76th IFLA General Conference and Assembly Proceedings. (10–15 August 2010, Gothenburg, Sweden). P. 1–13.
11. *Паринов С. И.* Концепция виртуальной научной среды «Открытая Наука» // Научный сервис в сети Интернет: суперкомпьютерные центры и задачи: Труды междунар. суперкомпьютерной конф. (Новосибирск, 20–25 сентября 2010). — М.: МГУ, 2010. С. 473–481.
12. *Паринов С. И., Коголовский М. Р.* Технология семантического структурирования контента научных электронных библиотек // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL'2011): Труды XIII Всеросс. науч. конф. (Воронеж, 19–22 октября 2011). — Воронеж: ВГУ, 2011. С. 197–206.
13. *Åström F., Sándor Á.* Models of scholarly communication and citation analysis // ISSI 2009: 12th Conference (International) of the International Society for Scientometrics and Informetrics Proceedings. Vol. 1. <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=1459018&fileId=1883080>.
14. *Sándor Á., Kaplan A., Rondeau G.* Discourse and citation analysis with concept-matching, CiteSeer. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.67.7518&rep=rep1&type=pdf>.
15. *Teufel S., Siddharthan A., Tidhar D.* Automatic classification of citation function // 2006 Conference on Empirical Methods in Natural Language Processing Proceedings. <http://portal.acm.org/citation.cfm?id=1610091>.
16. *De Waard A., Kircz J.* Modeling scientific research articles — shifting perspectives and persistent issues // ELPUB 2008 Conference on Electronic Publishing Proceedings. — Toronto, Canada, 2008. http://elpub.scix.net/data/works/att/234_elpub2008.content.pdf.
17. SKOS — Simple Knowledge Organization System. <http://www.w3.org/TR/skos-reference>.
18. *Shotton D.* Introducing the semantic publishing and referencing (SPAR) ontologies. October 14, 2010. <http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies>.
19. *Shotton D., Peroni S.* Semantic annotation of publication entities using the SPAR (Semantic Publishing and Referencing) ontologies // Beyond the PDF Workshop. La Jolla. January 19, 2011. http://imageweb.zoo.ox.ac.uk/pub/2010/Publications/Shotton&Peroni_semantic_annotation_of_publication_entities.pdf.
20. Semantic Web Applications in Neuromedicine (SWAN) ontology. W3C Interest Group Note. October 20, 2009. <http://www.w3.org/TR/2009/NOTE-hcls-swan-20091020>.
21. CERIF 2008 — Final Release (1.2). <http://www.eurocris.org/Index.php?page=CERIF2008&t=19>.
22. CERIF-2008-1.3 Ontology. <http://spi-fm.uca.es/neologism/cerif#>.
23. CERIF-2008-1.3 Semantic Vocabulary. <http://spi-fm.uca.es/neologism/semcerif#>.
24. *Shotton D., Peroni S.* CiTO, the Citation Typing Ontology, v. 2.0. <http://purl.org/spar/cito>.
25. *Shotton D.* CiTO, the Citation Typing Ontology // J. Biomedical Semantics, 2010. Vol. 1. Suppl. 1. P. 6. <http://www.jbiomedsem.com/content/1/S1/S6>.
26. *Shotton D., Peroni S.* DoCO, the Document Components Ontology. <http://speroni.web.cs.unibo.it/cgi-bin/lode/req.py?req=http://purl.org/spar/doco>.
27. *Коголовский М. Р., Паринов С. И.* Использование связей цитирования для наукометрических измерений в системе Соционет. Депонировано в Соционет, 2009. <http://socionet.ru/publication.xml?h=repec:rus:rssalc:web-32>.
28. SciVal. http://www.elsevier.com/wps/find/electronicproductdescription.cws_home/720941/description#description.

ОБ ИНТЕГРАЦИИ ГЕОГРАФИЧЕСКИХ МЕТАДААННЫХ ПОСРЕДСТВОМ РЕТРОСПЕКТИВНОГО ТЕЗАУРУСА*

Д. М. Скачков¹, О. Л. Жижимов²

Аннотация: Обсуждаются вопросы, связанные с построением интероперабельного тезауруса географических наименований, включающего геометрические данные географических объектов, в том числе и ретроспективные. Определяются основные требования к подобному тезаурусу, производится обзор существующих решений исходя из описанных требований, формулируются основные позиции соответствующего профиля для организации доступа к тезаурусу, приводится реляционная схема, предназначенная для хранения данных тезауруса.

Ключевые слова: географические метаданные; интеграция; ретроспективное геокодирование; тезаурус

1 Введение

В настоящее время в связи с возрастающей потребностью общества в информационном обеспечении, в том числе и связанном с географическим аспектом информации, все большую актуальность приобретают разработки, направленные на интеграцию «негеографических» информационных систем с информационными системами, изначально ориентированными на обработку географической информации. Под «негеографическими» информационными системами здесь и в дальнейшем будем понимать информационные системы, для которых изначально не предполагалось использование пространственных данных. К таким системам относятся, например, электронные библиотеки. Добавление географического аспекта к информации, хранящейся в таких системах, позволило бы существенно повысить функциональность их навигационных, поисковых и визуализационных сервисов. Подобная интеграция даст возможность, к примеру, производить поиск по заданному географическому региону [1], отображать на карте материалы, относящиеся к соответствующим точкам на поверхности Земли (как это делается на Google Maps), повысить релевантность результатов поиска.

Следует заметить, что существующие в настоящее время программные комплексы, явным образом не связанные с географическими информационными системами (ГИС), не содержат необходимой функциональности по хранению и обработке географических данных. Наделение же их требуемой функциональностью осложняется отсутствием единых стандартов на поиск и представление данных, связанных с географическим аспектом,

а также отсутствием четкого описания технологии интеграции как таковой [2].

Таким образом, разработка технологии, обеспечивающей обработку географического аспекта информации в «негеографических» информационных системах общего назначения, является актуальной и перспективной.

2 Пути интеграции географических данных

Прежде чем описывать варианты внедрения географической информации в объекты информационной системы, разграничим две важные составляющие любого объекта, характерного для рассматриваемых систем.

Вся информация об объекте может быть разделена на две составляющие:

- (1) контент — информационное наполнение объекта;
- (2) контекст — среда, в которой существует объект.

В дальнейшем будем считать, что географический аспект информации может быть зафиксирован на уровне метаданных, описывающих контент и контекст. При этом «географические» метаданные объекта могут быть заданы двумя способами:

- (1) с помощью геометрического описания географического объекта на основе координат;
- (2) с помощью ссылки на элемент некоторого тезауруса, включающего географические названия.

* Работа выполнена при поддержке РФФИ, грант № 10-07-00302-а.

¹ Институт вычислительных технологий СО РАН, danil.skachkov@gmail.com

² Институт вычислительных технологий СО РАН, zhizhim@sbras.ru

ния соответствующих объектов. Так как термин «тезаурус» может употребляться в различных значениях, в данной работе под тезаурусом будем понимать информационно-поисковый тезаурус. Информационно-поисковый тезаурус — это нормативный словарь, явно указывающий отношения между терминами и предназначенный для описания содержания документов и поисковых запросов [3].

Оба варианта в применении к задаче интеграции имеют как положительные, так и отрицательные стороны.

Первый вариант исключает неоднозначные толкования, но в то же время он не очень удобен по причине необходимости внесения существенных изменений в уже существующие информационные системы. *Второй* вариант не является однозначным, но может быть реализован на базе существующих парадигм информационных систем при условии их небольшой модернизации, а также обладает большей гибкостью. Хотя реализация тезауруса географических названий сопряжена с большим объемом работ, но возможность его повторного использования оправдывает все затраты. Более того, при реализации первого варианта интеграции тезаурус географических наименований также необходим для определения координат географических объектов, имеющих отношение к записям информационной системы. Поэтому в данной работе речь пойдет именно о втором варианте.

3 Препятствия при интеграции посредством тезауруса

Существует множество тезаурусов географических наименований, но сложность их использования применительно к данной задаче заключается в том, что географический аспект объектов, хранящихся в негеографических информационных системах, зачастую относится не к текущему моменту, а к моментам времени прошедшим. Однако с течением времени могут изменяться как географические названия, так и границы географических объектов. Будем называть это изменение свойств с течением времени *ретроспективным аспектом информации*. В то же время большинство тезаурусов содержит сведения, относящиеся только к текущему моменту времени, т. е. не учитывает ретроспективный аспект информации. Данная особенность препятствует использованию существующих тезаурусов географических наименований в подобных системах.

Следует заметить, что любые изменения географических названий и геометрических объектов, ассоциированных с ними, как правило, связаны с каким-либо нормативным документом.

Более того, в существующих тезаурусах координаты географического объекта чаще всего задаются в виде точки, в то время как реальные координаты объекта представляют собой далеко не точку, а, в общем случае, некоторую область. Что, конечно же, также уменьшает полезность таких тезаурусов при проведении поиска. Поэтому более предпочтительным будет тезаурус, где положение объектов задано с помощью координат границ области, занимаемой объектом.

Для задач поиска полезными будут также данные о том, как географические объекты расположены относительно друг друга. Например, если производится поиск по некоему региону, целесообразно считать релевантными также и элементы, относящиеся к географическим объектам, лежащим в целевом регионе.

Таким образом, для использования в информационных системах общего назначения географического аспекта в его любом виде необходим справочный аппарат (тезаурус), который включает в себя не только географический аспект информации, но и ее временной (ретроспективный) аспект.

В данной работе сделана попытка сформулировать основные требования к подобному тезаурусу географических названий, который мог бы удовлетворить потребности существующих информационных систем по обработке географического и исторического аспекта информации. В работе приводится обзор некоторых схем данных, а также существующих тезаурусов, анализ их сильных и слабых сторон (в контексте применения к задаче привязки географических метаданных к объектам информационных систем). Формулируются требования к тезаурусу географических наименований, подходящему для использования в рамках задачи интеграции географических метаданных. Также приводится вариант реляционной схемы данных тезауруса.

4 Основные требования к тезаурусу

На основе всего вышеизложенного сформулируем список требований к тезаурусу, подходящему для использования в рассматриваемой задаче.

Тезаурус должен:

- (1) обеспечивать прямое и обратное геокодирование;

- (2) обеспечивать ретроспективное прямое и обратное геокодирование;
 - (3) позволять включать информацию в технологию поиска в существующих информационных массивах;
 - (4) содержать внутренние связи:
 - (а) по географическим объектам,
 - (б) по временным характеристикам,
 - (в) по документам;
 - (5) быть представлен в схеме, максимально приближенной к какой-либо стандартной;
 - (6) однозначно отображаться на другие схемы тезаурусов, в частности необходимо однозначное соответствие профилю Zthes [4], быть может расширенному, для интеграции с существующими информационными системами.
- (1) наличие ретроспективных данных. Возможность извлечь данные, относящиеся к прошлому;
 - (2) наличие связей с нормативными документами. Возможность определить, согласно какому документу было изменено название или координаты объекта;
 - (3) описание координат географического объекта сообразно его форме. Представление географического объекта не только в виде точки, а также в виде замкнутого контура, линии, композиции примитивов;
 - (4) наличие связей, отражающих относительное расположение географических объектов.

В первую очередь рассмотрим существующие схемы данных.

5.1 ГОСТ Р 52573-2006

Для пояснения сформулированных требований рассмотрим основные сценарии использования тезауруса географических наименований. Работа с тезаурусом включает два основных сценария: запрос координат геометрического примитива объекта по имени этого объекта и запрос всех имен объектов по заданному координатно-привязанному геометрическому примитиву. Обычно это называется *прямым* и *обратным геокодированием*. Для целевых систем, которые потенциально могут содержать ретроспективную информацию, отличительной особенностью становится необходимость указания времени, для которого соответствующее геокодирование будет актуальным. При этом отсутствие задания момента времени может служить указанием на использование текущего момента времени в качестве параметра запроса.

Существенным моментом прямого геокодирования является тот факт, что заданные в запросе имя и время могут быть взаимно противоречивы. Например, запрос на координаты объекта (Новосибирск, 1920-05-20) должен возвращать ответ (Новониколаевск, координаты геометрического примитива, 1920-05-20). Запросы обратного геокодирования в этом смысле более просты, так как задаваемые в запросе координаты не связаны с действующей топонимикой.

5 Обзор существующих решений

Рассмотрим имеющиеся на данный момент схемы представления данных и существующие тезаурусы, которые могут представлять интерес в рамках этой задачи.

При рассмотрении будем обращать внимание на следующие свойства:

Национальный стандарт Российской Федерации «Географическая информация. Метаданные». Стандарт предназначен для специалистов в области информационных технологий, разработчиков геоинформационных систем, баз и банков пространственных данных, а также прикладных информационных систем различного назначения. Стандарт разработан в соответствии с правилами создания профилей, указанными в стандарте ISO 19115 [5].

Данный стандарт содержит рекомендацию к использованию ретроспективных данных (Сущность EX_Extent). Координаты объекта задаются с помощью одной из сущностей:

- (1) EX_BoundingPolygon — многоугольник (задается множеством точек);
- (2) EX_GeographicBoundingBox — прямоугольная область (задается координатами углов);
- (3) EX_GeographicDescription — описание объекта с использованием географического идентификатора.

Есть сведения о документе-источнике, но они привязаны к объекту в целом, а не к данным о его координатах и наименовании.

Сведения о связях между объектами отсутствуют в данной схеме.

5.2 CIDOC Conceptual Reference Model

Committee on Documentation Conceptual Reference Model (CRM) не является схемой тезауруса, но

рассматривается в данной работе, так как представляет собой формальную онтологию, предназначенную для улучшения интеграции и обмена гетерогенной информацией по культурному наследию. Более конкретно, CIDOC CRM определяет семантику схем баз данных и структур документов, используемых в культурном наследии и музейной документации, в терминах формальной онтологии. Модель не определяет терминологию, появляющуюся в конкретных структурах данных, но имеет характерные отношения для ее использования.

Модель может служить как руководством для разработчиков информационных систем, так и общим языком для экспертов предметной области и специалистов по информационным технологиям. Она предназначена для покрытия контекстной информации исторического, географического и теоретического характера об отдельных экспонатах и музейных коллекциях в целом [6].

В CIDOC CRM представляет интерес сущность E53_Place (место), которая как раз описывает географические метаданные объекта. Данная сущность является экземпляром E44_Place_Appellation. E44_Place_Appellation содержит данные о координатах, адресе, географическом наименовании. Координаты могут задаваться в любом виде (не только географические). Имеется возможность задать ссылки на родительский элемент (иерархические связи).

Но в онтологии CIDOC CRM не учтено изменение свойств географических объектов с течением времени, отсутствует связь географических метаданных с нормативными документами.

5.3 Getty Thesaurus of Geographic Names

Тезаурус географических имен института Getty — англоязычный тезаурус, содержащий более чем миллион географических имен, информацию о континентах, физических объектах, административных сущностях и нациях современного политического мира, а также сведения об исторически значимых областях [7].

Схеме тезауруса Getty, естественно, присущи как положительные, так и отрицательные черты.

Из отрицательных моментов можно отметить отсутствие информации об изменении координат географических объектов с течением времени. Координаты объекта могут быть представлены либо точкой, либо прямоугольником, что недостаточно для полного описания области на земной поверхности.

В то же время в схеме данного тезауруса учтено изменение названия объекта с течением времени (Term_Date). Также учтены нормативные докумен-

ты (Subject_Sources) для данного объекта и для его наименований (Term_Source) [8]. Записи содержат данные об иерархии.

5.4 Тезаурус Российской государственной библиотеки

Справочник географических названий Российской государственной библиотеки содержит наименования географических объектов (городов, рек, и т. д.) на территории Российской Федерации [9, 10].

Тезаурус не содержит ретроспективных данных в записях. Невозможно получить ни данных о предыдущих названиях, ни данных о предыдущих координатах объектов.

В записях присутствуют ссылки на нормативные документы, определяющие наименование объекта.

Координаты географических объектов заданы в виде координат точек, что не совсем соответствует действительности.

В записях тезауруса также есть данные о взаиморасположении объектов.

5.5 Служба геокодирования API сервиса Google Maps

Позволяет определить координаты объекта, а также найти адрес, наиболее близкий к указанным координатам [11].

В записях, предоставляемых данной службой, отсутствуют ретроспективные данные. Отсутствуют связи с нормативными документами.

Координаты объектов указаны в виде точек.

В записях содержатся иерархические связи.

В то же время стоит отметить, что тезаурус содержит данные не только о крупных географических объектах, но также и об адресах. Есть возможность произвести обратное геокодирование. Но использовать службу геокодирования можно только вместе с картами Google, что делает невозможным использование сервиса в данной задаче.

5.6 Служба геокодирования API сервиса Яндекс.Карты

Имеет функциональность, аналогичную геокодеру Google [12]. Обладает практически теми же достоинствами и недостатками, но из дополнительных достоинств можно выделить более обширную базу российских наименований географических объектов. Сервис обладает аналогичными ограничениями по использованию, что также делает невозможным использование сервиса в данной задаче.

Таблица 1 Сравнительная таблица схем тезаурусов

Схема/тезаурус	Содержит ретроспективные сведения	Содержит ссылки на документы-источники	Координаты географических объектов заданы соответственно их размерам и форме	Наличие иерархических связей
ГОСТ Р 52573-2006	+	+	±	–
CIDOC CRM	–	–	±	+
Getty	+	+	±	+
Российская государственная библиотека	–	+	–	+
Геокодер сервиса Google Maps	–	–	–	+
Геокодер сервиса Яндекс.Карты	–	–	–	+

5.7 Сравнительная таблица

По итогам сравнения составим сравнительную таблицу рассмотренных схем (табл. 1).

Таким образом, проанализировав существующие решения, приходим к выводу, что схемы тезауруса с необходимой функциональностью нет. Но есть достаточно близкие схемы, которыми можно руководствоваться при разработке собственного тезауруса.

6 Разработка схемы тезауруса

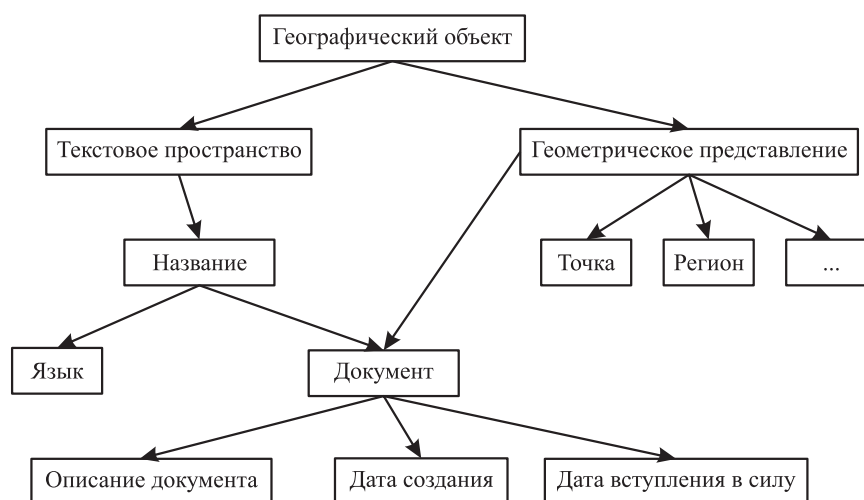
На основании приведенных выше данных была построена онтология тезауруса, отвечающего сформулированным ранее требованиям (рис. 1) [13].

Заметим, что правильно организованный тезаурус географических названий может служить осно-

вой и для получения информации, отличной от результатов прямого и обратного геокодирования, в частности:

- информации о документах, связанных с конкретным географическим объектом;
- информации о времени актуальности названий объектов;
- информации о времени актуальности координат объектов;
- информации о временных характеристиках производных параметров.

Следует также заметить, что любой тезаурус является лишь дополнительной базой данных, которая может быть задействована при обработке запросов к различным информационным массивам.

**Рис. 1** Онтология тезауруса

Ретроспективный тезаурус географических названий может быть задействован при обработке запросов, включающих ретроспективные географические названия.

Можно выделить три вида условий в запросе к тезаурусу:

- (1) по имени;
- (2) по координатам;
- (3) по времени.

Перечисленные условия могут комбинироваться друг с другом.

Для интеграции с существующими информационными системами и обеспечения интероперабельности необходимо зафиксировать профиль доступа к обсуждаемому тезаурусу (RGeoThes). Этот профиль, несомненно, должен являться расширением профиля ZThes [4] для доступа к тезаурусам по протоколам Z39.50 и SRW/SRU и включать необходимые компоненты для временного и географического поиска. При этом профиль должен определять:

- схему данных;
- структуру записи и наборы элементов;
- обязательные и дополнительные индексы (точки доступа);
- синтаксис поисковых запросов и поисковые атрибуты;
- форматы представления данных;
- протоколы доступа к ресурсу.

6.1 Протоколы доступа

Для обеспечения интероперабельности доступ к RGeoThes должен обеспечиваться по протоколам:

- Z39.50;
- HTTP/XML/SOAP/SRW;
- HTTP/SRU.

Каждый из указанных способов доступа имеет свои специфические особенности, которые должны быть определены общим профилем.

6.2 Форматы представления данных

В качестве основного обязательного формата представления записи RGeoThes для всех способов доступа является формат XML. Дополнительным необязательным форматом является HTML. Для доступа по Z39.50 обязательным форматом также является GRS-1. В качестве дополнительных (необязательных) форматов могут использоваться RUSMARC, MARC21 и др.

6.3 Схема данных

Схема данных определяется в терминах XML (XSD) и должна соответствовать онтологии, схематично представленной на рис. 1.

6.4 Индексы и точки доступа

Точками доступа записи RGeoThes должны быть элементы, представленные в табл. 2.

6.5 Синтаксис поисковых запросов и поисковые атрибуты

Для доступа по Z39.59 обязательным синтаксисом запросов должен являться RPN-1, необязательным — CQL. Для доступа по SRW/SRU обязательным синтаксисом запросов должен являться CQL, необязательным — RPN-1 (x-pquery).

Поисковые атрибуты RPN для доступа по Z39.50 для обеспечения интероперабельности должны соответствовать поисковым атрибутам профиля Z-Thes из наборов zthes-1, utility, cross-domain (xd-1). Для поиска по времени и координатам должны использоваться атрибуты из набора cip-1. Аналогичное требование справедливо и для запросов CQL.

Соответствие поисковых атрибутов точкам доступа приведено в табл. 2 [14].

7 Реляционная схема данных тезауруса

Рассмотрим вариант схемы для хранения записей тезауруса в случае использования реляционной СУБД PostgreSQL в качестве хранилища данных. Учитывая все вышеперечисленное, построим реляционную схему данных. Схема представлена на рис. 2.

Основной в данной схеме является таблица «**Запись тезауруса**» (далее — **главная таблица**), в которой находится список квалификаторов записей тезауруса. Строка из данной таблицы может содержать ссылку на предыдущий вариант записи и на родительскую запись. Связи между записями тезауруса содержатся в таблице «**Связь между записями**». Каждая связь содержит квалификаторы двух записей тезауруса, которые она связывает. Также связь характеризуется двумя документами (что представлено в виде внешних ключей). «Начальный документ» определяет документ, в котором

Таблица 2 Точки доступа записи RGeoThes

Точка доступа	Набор	Тип	Значение
Локальный номер	utility	1	4
Название термина	cross-domain	1	1
Квалификатор термина	zthes-1	1	1
Тип термина	zthes-1	1	2
Статус термина	zthes-1	1	7
Категория термина	zthes-1	1	6
Язык названия	utility	1	3
Дата начала действия названия	cip-1		
Дата окончания действия названия	cip-1	1	2073
		2	14, 15, 16, 17, 18
Документ, фиксирующий название	cross-domain	1	6
Тип геометрического объекта	cip-1	4	201, 202
Координаты геометрического объекта	cip-1	1	2059, 2060
		2	7, 8, 9, 10
Дата начала действия определения геометрии	cip-1	1	2072
		2	14, 15, 16, 17, 18
Дата окончания действия определения геометрии	cip-1	1	2073
		2	14, 15, 16, 17, 18
Документ, фиксирующий определения геометрии	cross-domain	1	6
Комментарий	cross-domain	1	4
Идентификатор связанного термина	zthes-1	1	4

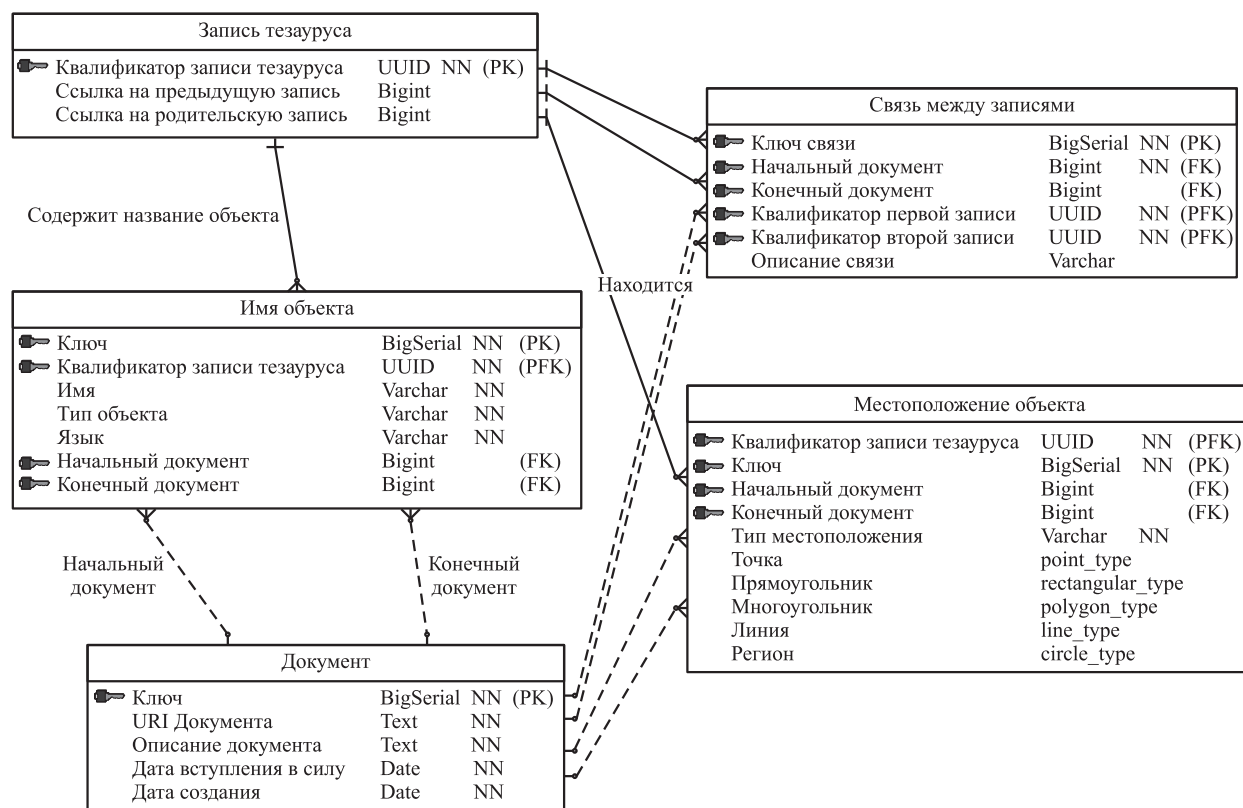


Рис. 2 Реляционная схема данных тезауруса

зафиксировано появление связи. «Конечный документ», который может быть не указан, определяет документ, в котором зафиксировано исчезновение связи.

В таблице «Имя объекта» задаются наименования географических объектов, содержащихся в тезаурусе. Каждая из записей главной таблицы может быть связана с несколькими строками имен. В свою очередь, имя может быть связано только с одной записью из главной таблицы. Каждое из имен характеризуется собственно именем, а также типом объекта и языком. Под типом объекта понимается, например, тип населенного пункта. Каждая запись таблицы «Имя объекта» содержит идентификаторы двух документов — начального и конечного. «Начальный документ» определяет документ, в котором зафиксировано присвоение данного имени географическому объекту. «Конечный документ» определяет документ, в котором зафиксировано окончание срока действия данного имени.

Таблица «Местоположение объекта» содержит данные о координатах географических объектов тезауруса. Каждая из записей главной таблицы может быть связана с несколькими строками данной таблицы. В то же время запись таблицы «Местоположение объекта» может быть связана только с одной строкой из главной таблицы. Каждая запись таблицы «Местоположение объекта» содержит идентификаторы двух документов — начального и конечного. «Начальный документ» определяет документ, в котором зафиксировано присвоение данного местоположения географическому объекту. «Конечный документ» определяет документ, в котором зафиксировано окончание срока действия местоположения применительно к данному объекту. Также каждая запись содержит поле «тип местоположения», содержащее идентификатор типа местоположения объекта. Таким типом может быть точка, прямоугольник, многоугольник, линия, регион и прочие. Благодаря использованию отображения «Одна иерархия — одна таблица» для набора типов местоположений объекта становится возможным легко добавлять новые типы местоположения в уже работающую схему. Для хранения координатных данных используются поля «Точка», «Прямоугольник», «Многоугольник», «Линия», «Регион» с типами данных `point_type`, `rectangle_type`, `polygon_type`, `line_type`, `circle_type` соответственно. Типы данных для этих полей являются композитными типами, содержащими всю координатную информацию, характерную для представления. Например, тип данных `rectangle_type`, соответствующий прямоугольной области на поверхности Земли, содержит поле `rect` встроенного типа `box`.

Данное разделение на типы для различных видов географических объектов сделано в целях повышения гибкости схемы.

Таблица «Документ» содержит данные о документах, регистрирующих изменение характеристик объектов с течением времени. Каждый документ содержит описание, уникальный идентификатор ресурса (URI), дату создания и дату вступления в силу. Именно датой вступления в силу документов определяются временные рамки существования той или иной характеристики географического объекта.

8 Заключение

В данной работе была показана необходимость реализации информационно-поискового тезауруса географических наименований в рамках задачи интеграции географических метаданных в информационные системы общего назначения.

Изложенные выше основные положения процесса интеграции, а также организации ретроспективного геокодирования и соответствующего тезауруса географических названий будут в дальнейшем использованы для построения модели информационной системы с возможностями геометрического и ретроспективного поиска информации. Поиск предполагается организовать на основе картографических интерфейсов в соответствии с описанным выше профилем.

Приведен вариант реляционной схемы для хранения данных тезауруса. Схема в настоящее время используется для хранения данных, собираемых в тезаурус в рамках данной работы. В дальнейшем планируется реализация доступа к тезаурусу на основе протоколов SRW/SRU, Z39.50, HTTP, а также экспериментальная интеграция метаданных в работающую систему.

Литература

1. Жижимов О. Л., Мазов Н. А. География и стандарты метаданных для электронных библиотек: содержание, применение, проблемы // Электронные библиотеки, 2009. Т. 12. № 1. <http://www.elbib.ru/index.php?page=elbib/rus/journal/2009/part1/ZM>.
2. Жижимов О., Мазов Л. А. Проблемы географической привязки цифровых объектов в электронных библиотеках // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL'2010): Труды XII Всеросс. научн. конф. — Казань: КГУ, 2010. С. 207–214.
3. Лукашевич Н. В. Тезаурусы в задачах информационного поиска. — М.: Изд-во Московского ун-та, 2011.

4. The Zthes specifications for thesaurus representation, access and navigation. <http://zthes.z3950.org>.
5. ГОСТ Р 7.24-2007. Тезаурус Информационно-поисковый многоязычный. Состав, структура и основные требования к построению. — М.: Стандартинформ, 2006.
6. Онтология в области документации в сфере культурного наследия: CIDOC CRM. <http://www.intuit.ru/department/expert/ontoth/5>.
7. Getty Thesaurus of Geographic Names® Online. <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>.
8. Contribute to the Getty Vocabularies. <http://www.getty.edu/research/tools/vocabularies/contribute.html>.
9. Тезаурус РГБ. http://aleph.rsl.ru/F/?func=file&file_name=find-b&local_base=tst11.
10. Лаврёнова О. А. Многоязычный доступ к данным на основе тезауруса географических названий // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL'2007): Труды IX Всеросс. научн. конф. — Переславль-Залесский: Ун-тет г. Переславля, 2007. С. 57–62.
11. Геокодирование — Службы API Карт Google. <http://code.google.com/intl/ru/apis/maps/documentation/geocoding>.
12. Поиск по карте — Яндекс.Карты. <http://api.yandex.ru/maps/geocoder>.
13. Соловьев В. Д., Добров Б. В., Иванов В. В., Лукашевич Н. В. Онтологии и тезаурусы: Учебное пособие. — Казань, М., 2006.
14. Catalogue Interoperability Protocol (CIP) Specification — Release B // CEOS/WGISS/ICS/CIP-B. April 2005. Issue 2.4.75.

УНИВЕРСАЛЬНАЯ СИСТЕМА ПРОВЕРКИ ТЕКСТОВ НА ПЛАГИАТ «АВТОР.NET»

Е. В. Шарапова¹, Р. В. Шарапов²

Аннотация: Обсуждается проблема обнаружения в текстах заимствований из других источников. Рассматриваются основные подходы к обнаружению заимствований, проводится обзор существующих на сегодняшний день программ. Дается обзор методов к сокрытию фактов заимствований. Дается описание разработанной системы «Автор.NET», способной проводить проверку заимствований по внутренним источникам и сети Интернет.

Ключевые слова: плагиат; обнаружение плагиата; заимствование

1 Введение

Бурное развитие вычислительной техники привело к глубокому проникновению компьютеров в нашу жизнь. Компьютеры окружают нас везде — на работе, дома, в магазинах и общественных местах. Современное развитие информационных технологий и глобальной сети Интернет предоставило широким кругам пользователей доступ к огромным массивам информации. Появилось большое число онлайн-библиотек, содержащих в электронном виде художественную и научно-техническую литературу. Стало возможным читать книги, новости и газеты непосредственно с экрана компьютера.

В сети Интернет стало доступно множество методических указаний, курсов лекций, учебников и т. д. Кроме того, появились огромные коллекции рефератов, готовых лабораторных работ, курсовых и дипломных проектов и даже диссертаций. Использование компьютерной техники сильно облегчило задачу поиска и копирования подобной информации. Если раньше для написания реферата или контрольной работы информацию было нужно, по крайней мере, найти в книгах и переписать (вручную, перепечатать или ввести в компьютер с помощью сканера и программ распознавания текстов), то теперь достаточно ввести название темы в поисковую систему и скопировать найденные материалы. Стал распространяться метод написания работ, получивший название «Copy & Paste». Метод заключается в простом копировании информации из одного или нескольких источников с минимальным редактированием получающегося таким образом текста.

Аналогичная ситуация наблюдается с отчетными материалами внутри учебных заведений. В связи с тем, что большое число пояснительных записок по курсовым и дипломным проектам выполняется с использованием компьютеров, происходит их распространение и повторное использование среди учащихся.

В последнее время наблюдается бурный рост использования в учебном процессе подобной заимствованной информации. Ситуация усугубляется тем, что учащиеся иногда не знают (не читают) то, что написано в «их» работах.

Плагиат — умышленное присвоение авторства на чужое произведение литературы, науки, искусства, изобретение или рационализаторское предложение (полностью или частично) [1].

Как можно убедиться из определения, подобные заимствованные работы можно отнести к разряду плагиата. Задача обнаружения недобросовестного использования заимствованных текстов в учебных и ученых кругах (фактов плагиата) приобретает высокую актуальность.

2 Формы заимствований текстов

Рассмотрим формы заимствований, встречающиеся в практике учебных заведений и подлежащие выявлению.

1. Полное или частичное копирование текста из одного источника.
2. Копирование и компоновка текста из нескольких источников.

¹Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых, mivlgu@mail.ru

²Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых, info@vanta.ru

Таблица 1 Формы плагиата

Форма плагиата	Доля
Полное или частичное копирование текста из одного источника	36%
Копирование и компоновка текста из нескольких источников	62%
Копирование текста из другого источника и изменение порядка следования частей текста	2%

Таблица 2 Частота использования подходов к сокрытию фактов плагиата

Подходы к сокрытию плагиата	Доля
Корректировка родов, чисел и времен, входящих в текст слов	32%
Незначительное изменение текста	38%
Сокращение заимствованного текста	44%
Замена букв	4%
Синонимизация текста	2%

3. Копирование текста из другого источника и его частичное редактирование.

Для того чтобы скрыть факт заимствований, могут применяться следующие подходы:

1. Корректировка родов, чисел и времен входящих в текст слов. Например, замена слова «выполнил» на «выполнила» или «выполнили», использование местоимения «я» вместо «мы» в оригинальном тексте и т. д.
2. Незначительное изменение заимствованного текста.
3. Сокращение заимствованного текста путем удаления слов, предложений, абзацев, рисунков, формул и т. д.
4. Обход систем проверки на плагиат путем замены русских букв на аналогичные по написанию английские и т. д.
5. Осуществление ручной или автоматической синонимизации текста.

Все вышеописанное должно учитываться при создании и использовании систем проверки на заимствования. О правомочности того или иного заимствования решение выносит сам проверяющий.

Для оценки частоты использования тех или иных форм плагиата мы провели следующий эксперимент. Студентам двух групп гуманитарных специальностей было предложено написать статьи на тему экологической ситуации в регионе (Владимирская область). Студенты были предупреждены о том, что статьи будут проверяться на наличие плагиата. Из полученного набора были исключены оригинальные статьи. Анализ статей, содержащих заимствованный контент, показал, что большинство из них скомпонованы из нескольких (реже одного) источников, чаще всего из учебников, статей

из сети Интернет и публикаций региональной прессы (табл. 1). Тот факт, что доля статей, полностью или частично скопированных только из одного источника, составила всего 36% (в реальных условиях она часто бывает больше), вероятно связан со знанием авторов о том, что работы будут проверяться. Доля работ, составленных путем копирования текста из другого источника и изменения порядка следования частей текста, оказалась незначительной (2%).

Анализ подходов, используемых студентами для сокрытия факта плагиата, показал, что в 32% работ осуществлялась корректировка родов, чисел и времен слов (табл. 2). В 38% работ (составленных как из одного, так и из нескольких источников) осуществлялось незначительное изменение заимствованного текста. Так, например, делались вставки слов и предложений в заимствованный текст, подвергались изменению названия населенных пунктов и рек (р. Волга в оригинале заменялась на р. Ока в статье). Надо заметить, что часть работ кроме заимствованных текстов содержала оригинальные блоки, чаще всего введение и заключение. Приведенная выше доля статей, подвергавшихся изменению, учитывает только заимствованные части таких текстов. Из работ, скопированных из одного источника, 44% подвергались сокращению. В данном случае под сокращением подразумевалось исключение части предложений, графиков, рисунков из заимствованных текстов, а также исключение начальных или конечных блоков текста, по смыслу составляющих единое целое с заимствованным фрагментом. Копирование законченного фрагмента из текста (например, раздела или главы) сокращением не считалось. Замена букв осуществлялась в 4% работ. В одной из работ замене подверглись практически все русские буквы, сходные по написанию с английскими буквами. В осталь-

ных работах заменялись одна—две гласные буквы. Ручная синонимизация проводилась только в 2% работ. Применения автоматической синонимизации в статьях замечено не было. Надо заметить, что около 40% рассматриваемых работ вообще не подвергались каким-либо изменениям, призванным скрыть факты плагиата.

3 Подходы к обнаружению заимствований

Существует несколько подходов к обнаружению заимствований (или, как их еще называют, нечетких дублей текстов). Достаточно подробный обзор приведен в [2].

Наибольшую известность получил метод «шинглов» [3]. Метод основан на представлении текстов в виде множества последовательностей фиксированной длины, состоящих из соседних слов. При значительном пересечении таких множеств документы будут похожи друг на друга. Одна из модификаций метода, получившая название «супершинглов», используется для быстрого обнаружения подобных документов [2].

Существует ряд методов, использующих сигнатурную лексическую информацию документов. В [4] для этих целей используется I-Match сигнатура, вычисляемая для слов со средним значением IDF (инверсной частоты слов в документах). Другим сигнатурным подходом, основанным на лексических принципах, является метод «опорных» слов [5]. В данном случае для документов составляются по определенным правилам наборы опорных слов, для которых строятся сигнатуры документов. Совпадение сигнатур говорит о подобии самих документов. Эта группа методов, несмотря на большую сложность реализации, показывает более хорошие результаты в обнаружении похожих документов [2].

Для обнаружения заимствований иногда используются алгоритмы, построенные на классических принципах информационного поиска, таких как TF, TF*IDF и т. д. [6]. В [7] предлагается использовать функцию схожести Джаккарда, применение которой позволяет добиться неплохих результатов даже в текстах с использованием синонимов и наличием орфографических ошибок.

4 Обзор существующих систем

Рассмотрим практическое использование описанных подходов в задачах обнаружения плагиата.

В настоящее время существует достаточно большое количество сервисов и программ, позволяющих так или иначе выявить заимствованный контент. Большую известность получила система «Антиплагиат», разработанная компанией «Форексис» [8]. Система осуществляет поиск по большому количеству коллекций рефератов, контрольных работ и учебников, хранящихся в собственной базе системы. Тем не менее система имеет ряд недостатков. Во-первых, она не осуществляет поиск по всем документам, доступным в сети Интернет. Особенно это касается тематических сайтов и новостных порталов: большое число заимствований осуществляется именно из таких источников. Соответственно, даже при полном дублировании подобной информации, система «Антиплагиат» соответствий не обнаружит. Во-вторых, присутствует ограничение размера проверяемого текста 3000 или 5000 символами (доступно после регистрации). В-третьих, ограничен просмотр документов, частично соответствующих проверяемому тексту. Кроме того, система ограничивает возможность проверки по базе имеющихся работ.

Программа Advengo Plagiatus осуществляет проверку с применением поисковых систем [9]. Использует разные поисковые системы и проверяет их доступность. В отличие от аналогичных систем, Advengo Plagiatus не использует Яндекс.XML, а обращается напрямую к таким поисковым системам, как Яндекс, Google, Bing, Рамблер, Yahoo, Поиск@Mail.ru, Nigma, QIP. Качество обнаружения плагиата достаточно высокое. Программа выдает процент совпадения текста и выводит найденные источники. Недостатком является отсутствие преобразования букв, отсутствие поддержки поиска по собственной базе. Из-за особенностей работы программы возникают ситуации, когда результаты проверки отличаются от раза к разу.

Сервис www.miratools.ru позволяет осуществлять онлайн-проверку текста на плагиат [10]. Система использует результаты выдачи поисковых систем. К достоинствам можно отнести возможность замены английских букв на русские. Имеются возможности изменять длину и шаг шинглов (используемых для проверки). По результатам проверки выдается процент совпадений и найденные источники. Система не работает с собственной базой. Присутствует ограничение на длину текста в 3000 символов и на число проверок в течение суток.

Сервис www.istio.com осуществляет проверку текста на наличие заимствованного контента с использованием поисковых систем [11]. Для этих целей используют Яндекс.XML и Yahoo.com. Возможности сервиса несколько слабее по сравнению с Miratools. По результатам проверки выдается со-

Таблица 3 Сравнение функциональности сервисов проверки текстов на плагиат

Система	Поиск в Интернете	Поиск в локальной базе	Обработка замены букв	Подробный отчет
Advego Plagiatus	+	—	—	+
«Антиплагиат»	—	+	—	—
Istio	+	—	—	—
Miratools	+	—	+	+
Plagiat-inform	+	+	—	+
Praide Unique Content Analyser II	+	—	—	+

общение о том, является ли текст уникальным или нет, и выдается список подобных сайтов. Преобразование букв и поддержка поиска по собственной базе отсутствует. Сервис предоставляет дополнительные средства для анализа текстов, например проверку орфографии, анализ наиболее частотных слов и т. д.

Программа Praide Unique Content Analyser II [12] имеет широкие возможности по проверке текстов с использованием поисковых систем. Имеется возможность выбора используемых поисковых систем, содержит средства добавления новых поисковых систем. Проверка осуществляется пассажами и шинглами, длину которых можно изменять. Можно задавать количества слов перекрытия шинглов. Выводится подробный отчет по проверке в каждой поисковой системе. К недостаткам можно отнести отсутствие замены букв и обработки стоп-слов. Нет поддержки работы с собственной базой.

Система Plagiatinform, по заверениям авторов, имеет наиболее широкий функционал [13, 14]. Она умеет проверять документы на наличие заимствований как в локальной базе, так и в сети Интернет. Система умеет обрабатывать документы, скомпонованные из перемешанных кусков текста нескольких источников. Проверка может осуществляться с использованием быстрого или углубленного поиска. Результаты проверки выдаются в виде наглядного отчета. Авторы не предоставляют возможности свободного использования или тестирования системы, и оценить качество ее работы невозможно.

Результаты сравнения функциональности рассмотренных сервисов проверки на плагиат приведены в табл. 3. Несмотря на большое число существующих решений, ни одно из них не может служить универсальным средством проверки на плагиат. Основной недостаток большинства существующих систем — это направленность поиска либо на сеть Интернет, либо на собственную базу. Очевидно, что более точная и универсальная проверка будет обеспечена при использовании обоих видов источников. Кроме того, большинство систем не способны обрабатывать замену букв, чем

часто пользуются недобросовестные авторы (чаще всего студенты).

В большинстве рассмотренных систем используется метод «шинглов». По исследованиям [2] этот метод демонстрирует высокую точность обнаружения дублированных текстов. Тем не менее из-за особенностей реализации результаты проверки в каждой системе сильно отличаются от других. Минусом метода является отсутствие возможности обработки синонимов [7]. Это является значительным недостатком существующих систем.

5 Практическая реализация

На базе Владимирского государственного университета авторами была разработана система проверки текстов на наличие заимствований из других источников (проверки на плагиат) «Автор.NET». Система осуществляет проверку как по источникам, доступным в сети Интернет, так и по собственным источникам (базам статей, курсовых и контрольных работ, дипломных проектов и т. д.). По результатам проверки формируется отчет с подсветкой найденных заимствований и возможностью просмотра найденных источников.

Рассмотрим структуру системы (рис. 1).

Проверяемый исходный текст подвергается предварительной обработке, в которую входят:

- (1) исключение из текста знаков препинания и спецсимволов;
- (2) преобразование регистра;
- (3) обработка замены символов (преобразование латинских букв в русских словах на аналогичные буквы русского алфавита для текстов на русском языке);
- (4) удаление стоп-слов и знаков препинания (предлоги, наречия и т. д.);
- (5) фильтрация текста (удаление неинформативных слов);
- (6) стемминг (обработка окончаний слов).

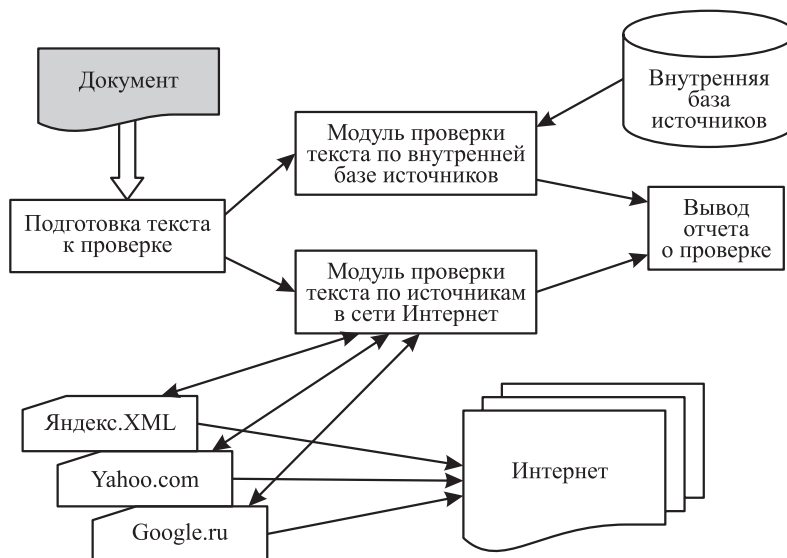


Рис. 1 Структура системы проверки текстов на заимствования

Фильтрация текста заключается в удалении наиболее частотных слов, неинформативных слов и т. д. Кроме того, фильтрации подвергаются цифры, спецсимволы и т. д. Эта процедура позволяет существенно сократить объемы вычислений (длину проверяемого текста).

Стемминг заключается в обработке окончаний слов. В описываемой системе они просто отбрасываются. Это позволяет исключить влияние таких модификаций текста, как изменение единственного и множественного числа, мужского и женского рода, настоящего и прошедшего времени и т. д.

Система проверки на плагиат «Автор.NET» состоит из двух модулей, которые функционируют независимо друг от друга.

Первый модуль осуществляет проверку по внутренней базе источников. База источников включает в себя статьи, курсовые и контрольные работы, дипломные проекты, а также учебники и курсы лекций. Источники хранятся как в виде полных текстов, необходимых для оценки значимости заимствований (по результатам проверки), так и в виде специально организованного поискового индекса. Последний необходим для быстрой проверки на совпадения текста и базы источников. Нет необходимости при каждой проверке просматривать все имеющиеся тексты и производить их достаточно трудоемкую обработку. Вся необходимая для поиска информация уже включена в структурированный поисковый индекс, с которым и работает модуль. Поисковый индекс формируется из текстов, прошедших описанную выше предварительную обработку.

Второй модуль осуществляет проверку по источникам сети Интернет. Для этих целей текст проверяемого документа разбивается на информативные куски (разбиение проводится по полному тексту документа без проведения фильтрации и стемминга). Число таких кусков зависит от размера документа. Далее с использованием поисковых систем проводится поиск источников, содержащих указанные информативные куски. Для осуществления поиска модуль использует Яндекс.XML, а также доступ к онлайн-поиску систем Google.ru, Rambler.ru, Aport.ru, Поиск.Mail.ru, Nigma.ru и т. д. Полученные таким образом источники проверяются затем на соответствие исходному документу. Для этого определяется формат источника (html-документ, txt-файл, doc- или rtf-документ, pdf-файл). В случае html-документа из источника удаляются теги разметки. Файлы *.doc, *.rtf и *.pdf преобразуются, если это возможно, в обычный текстовый формат без разметки. Далее источники проходят предварительную обработку, и затем проводится оценка их сходства с исходным документом (рис. 2).

Для оценки сходства исходного документа и источников используется некая модификация алгоритма «шинглов». Модификация алгоритма заключается в том, что рассматривается не оригинальный документ, а его обработанная и отфильтрованная копия с исключением неинформативных объектов. Основное требование к системе — полнота и точность оценки совпадений. Авторы не ставили задачей сокращение времени проверки, проведение экспресс-оценки на полные дубли и т. д.

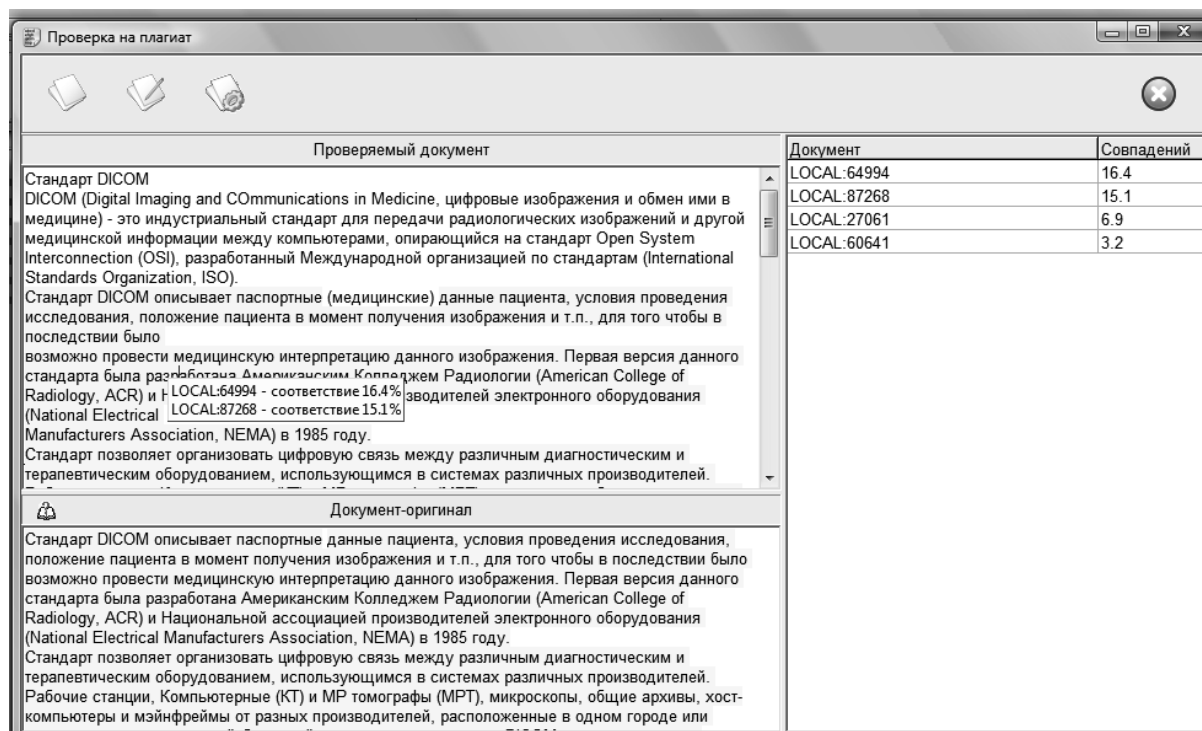


Рис. 2 Интерфейс программы «Автор.NET»

В настоящее время локальная база системы содержит дипломные проекты, выполненные за последние 6 лет, и курсовые проекты, выполненные за последние 3 года студентами одной из специальностей. Также в базе содержится ряд контрольных работ, выполненных студентами заочной формы обучения.

6 Результаты исследования

Для проверки работоспособности системы «Автор.NET» были составлены тесты трех видов:

1. Заимствования с изменением в тексте времен и родов слов (T_1).
2. Заимствования из одного источника с измененным порядком следования предложений и добавлением оригинального текста между предложениями (T_2).
3. Заимствования, взятые из нескольких источников, с измененным порядком следования предложений (T_3).

Все тесты имели приблизительно одинаковый размер в 2000 символов и содержали в среднем по 400 слов. В качестве источника текстов для составления тестов использовалась коллекция рефератов, широко доступная в сети Интернет. Было составлено по 10 тестов каждого вида.

Для оценки качества обнаружения заимствований сравнивались результаты работы системы с результатами систем «Антиплагиат», Advengo Plagiatius и Miratools. В связи с тем, что каждая система имеет свои принципы подсчета оригинальности документа, в качестве метрики оригинальности использовалось процентное отношение оригинальных слов в документе к общему количеству слов.

Для оценки качества обнаружения заимствований использовался показатель полноты (Recall), показывающий, какой процент заимствований был обнаружен (табл. 4). Точность обнаружения (Precision) во всех системах была на высоком уровне и стремилась к 1 [15].

Как можно заметить, ни одна из трех рассматриваемых систем не справилась с тестом на замену окончаний (T_1). Показатель Advengo Plagiatius объясняется наличием в измененном тексте цепочек из 5 слов, для которых окончания не менялись. Применение стемминга в системе «Автор.NET» поз-

Таблица 4 Результаты тестирования

Система	T_1	T_2	T_3
«Антиплагиат»	0	1	0,97
Miratools	0	0.9	0,83
Advengo Plagiatius	0,14	1	0,62
«Автор.NET»	0,99	1	0,98

волило ей справиться с указанной задачей и обнаружить заимствования.

С задачей обнаружения изменения порядка следования предложений, взятых из одного источника (T_2), справились все системы. Чуть худший результат Miratools (полнота 0,9) объясняется, видимо, особенностями реализации алгоритма сравнения в этой системе.

С задачей обнаружения предложений, взятых из разных источников с изменением порядка их следования (T_3), рассматриваемые системы справились немного хуже. Система «Антиплагиат» показала хорошее значение полноты (0,97). Результаты системы Miratools оказались более скромными (полнота 0,83). В системе Advego Plagiatius полнота иногда опускалась до 0,45 при среднем значении в 0,62. Система «Автор.NET» хорошо справилась с указанной задачей, продемонстрировав полноту в 0,98.

Как можно заметить, система «Автор.NET» успешно справилась со всеми видами тестов и показала результаты, не уступающие, а иногда и превосходящие результаты работы существующих систем.

7 Выводы

Таким образом, разработанная система «Автор.NET» проверки текстов на плагиат показала достаточно хорошие результаты. Использование фильтрации текста, стемминга и преобразования символов позволило системе находить заимствованные тексты даже при их незначительной модификации.

Система позволяет работать не только с русскоязычными текстами, но с текстами на иных языках.

Особенностью системы является возможность проведения проверки как по внутренней базе источников, так и по источникам сети Интернет. Это делает ее достаточно универсальным средством проверки текстов и выгодно отличает от существующих систем. Выдаваемые системой отчеты позволяют оценивать правомерность найденных заимствований текстов.

Система «Автор.NET» может использоваться для проверки уникальности студенческих работ (курсовых и дипломных проектов, рефератов и контрольных работ). Еще одной областью применения может служить использование системы для проверки докладов, представляемых на студенческие и молодежные научные конференции.

Литература

1. Большой энциклопедический словарь. — М.: АСТ, Астрель, 2008. 1248 с.
2. Зеленков Ю. Г., Сегалович И. В. Сравнительный анализ методов определения нечетких дубликатов для WEB-документов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL'2007): Труды IX Всеросс. научн. конф. — Переславль-Залесский: Ун-т г. Переславля, 2007. Т. 1. С. 166–174.
3. Broder A. On the resemblance and containment of documents // Compression and Complexity of Sequences (SEQUENCES'97). — IEEE Computer Society, 1998. P. 21–29.
4. Kolcz A., Chowdhury A., Alspector J. Improved robustness of signature-based near-replica detection via lexicon randomization // KDD 2004 Proceedings. — Seattle, 2004.
5. Ilyinsky S., Kuzmin M., Melkov A., Segalovich I. An efficient method to detect duplicates of Web documents with the use of inverted index // WWW'2002: 11th World Wide Web Conference (International) Proceedings. — New York: ACM Press, 2002.
6. Шаранов Р. В., Шарапова Е. В. Пути расширения булевой модели поиска // Информационные системы и технологии. Известия ОрелГТУ. — Орел: ОрелГТУ, 2009. № 6(56). С. 74–78.
7. Неелова Н. В., Сычугов А. А. Сравнение результатов детектирования дублей методом шинглов и методом Джаккарда // Вестник РГРТУ, 2010. № 4(34). С. 72–78.
8. Антиплагиат. <http://www.antiplagiat.ru>.
9. Advego Plagiatius — проверка уникальности текста. <http://advego.ru/plagiatius>.
10. Сервис проверки уникальности контента. <http://www.miratools.ru>.
11. Анализировать текст, поиск плагиата. <http://istio.com/rus/text/analyz>.
12. Проверка уникальности текста в Интернете — очень полезная программа для качественной раскрутки сайтов. <http://www.nado.su/downloads.html>.
13. SearchInform Плагиат-Информ — система для определения плагиата в документах. <http://www.searchinform.ru/main/full-text-search-plagiarism-search-plagiatinform.html>.
14. Ширяев М. А., Мустакимов В. Plagiatinform избавит от плагиата в научных работах // Educational Technol. Soc., 2009. № 11(1). С. 367–374.
15. Шаранов Р. В., Шарапова Е. В. Система проверки текстов на заимствования из других источников // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL'2011): Труды XIII Всеросс. научн. конф. — Воронеж: ВГУ, 2011. С. 233–238.

Статьи, являющиеся развитием докладов, представленных
на XXIX Международном семинаре
по проблемам устойчивости стохастических моделей
(г. Светлогорск Калининградской области России, 10–16 октября 2011 г.)

FRACTIONAL LEVY MOTION WITH DEPENDENT INCREMENTS AND ITS APPLICATION TO NETWORK TRAFFIC MODELING

C. De Nikola¹, Y. S. Khokhlov², M. Pagano³, and O. I. Sidorova⁴

Abstract: Since the beginning of the 1990s, accurate traffic measurements carried out in different network scenarios highlighted that Internet traffic exhibits strong irregularities (*burstiness*) both in terms of extreme variability and long-term correlations. These features, which cannot be captured in a parsimonious way by traditional Markovian models, have a deep impact on the network performance and lead to the introduction of α -stable distribution and self-similar processes into the network traffic modeling. In this paper, a generalization of fractional Brownian motion (fBm), which is able to capture both above-mentioned features of the real traffic, is considered.

Keywords: fractional Brownian motion; α -stable subordinator; self-similar processes; buffer overflow probability

1 Introduction

The application of probabilistic methods in the modeling and the analysis of telecommunication systems has a long history. Namely, the first researches in this framework date back to the beginning of the last century when A. K. Erlang (1878–1929), as a scientific collaborator and the head of the newly-established physico-technical laboratory of the Copenhagen Telephone Company, studied the issues related to loss and waiting time in automatic telephone exchanges. In the 1930s, the interests for these topics grew from a practical as well as theoretical point of view. Indeed, Erlang's results were soon used by telephone companies in several countries and gave birth to a new branch in the framework of probability theory, known as queueing theory, which attracted the interests of well-known probabilists such as Palm, Pollachek, Lindly, Khincine, Gnedenko, to name just a few.

In the 1920–1930s, many empirical works showed that, in case of telephone traffic, a suitable model is represented by the Poisson process. At the same time, Poisson flows have many “useful” mathematical properties:

- the superposition of independent Poisson processes is still a Poisson process;
- it has independent and stationary increments; and

- under some mild regularity conditions, the superposition of independent flows converges to a Poisson flow, if the number of flows grows, but the individual rates become infinitesimal so that the overall rate stays constant.

Because of the last property, in many works it has been proposed that the amount of traffic in global telecommunication backbones can be modelled as a Poisson process. For several decades, such model has been used without any further experimental validation and applied to new network scenarios, such as packet-switching networks.

At the beginning of the 1990s, a lot of empirical studies have been conducted in order to better understand the statistical features of packet traffic in global networks, such as Internet, as well as in local area networks inside research institutes, university campuses, and corporates [1–3]. Statistical studies of the collected data highlighted their radical differences with respect to the ubiquitous Poisson process and other traditional (typically Markovian, for the sake of analytical tractability) models. For instance, it is enough to visually check the behavior of real traffic data under different level of aggregations [4]. It is easy to see that at all the aggregation levels (in the range from milliseconds to hours) the data keep a random behavior, which appears to be almost the

¹University of Salerno, denicola@diima.unisa.it

²People's Friendship University of Russia, yskhokhlov@yandex.ru

³University of Pisa, m.pagano@iet.unipi.it

⁴Tver State University, Oksana.I.Sidorova@yandex.ru

same at all the different scales (apart from a normalization factor, related to the length of the observation window).

More accurate mathematical analyses [1] pointed out that real data presents **fractal** properties, i. e., they can be interpreted as trajectories of so-called **automodel** or **self-similar processes**. Moreover, it was showed that traffic flows, unlike the Poisson model, presents **long range dependence**, which has a huge impact on queuing performance. The third important characteristic of traffic data is that the distribution of many different traffic features (such as file length, duration of on and off periods of single sources) presents **heavy tails**.

These properties of actual traffic flows pointed out the necessity of new traffic models, able to captures them in a parsimonious way. It is worth mentioning that similar models were already known in the field of probability theory since they have been successfully applied in different frameworks, such as turbulence modeling and statistical physics.

The rationale behind the fractal nature of traffics and the links among the above-mentioned characteristics of measured traces have been widely investigated [5]. In particular, it has been shown that if locally the traffic load presents heavy tails, then under a sufficiently high level of aggregation it converges to a self similar process (for a precise formulation of the problem and the related scaling conditions (see [6–9])). According to the considered aggregation regime, two different models might arise: fBm and α -stable Levy motion, which, as will be clarified in the following, present “opposite” features. In more detail, fBm presents long range dependence, but the tails of its marginal distribution decay fast (by definition, according to Gaussian law!). On the contrary, α -stable Levy motion is characterised by independent increments (i. e., no long memory at all!), but has heavy tailed distribution (i. e., its tails decay as a power law).

The goal is to build a model, able to take into account both these features of real traffic. Moreover, using such model as input to a queuing system, it would be also interesting to determine relevant queueing parameters, such as the probability of buffer overflow, which gives an upper bound for the loss probability in finite buffer queues.

From the historical point of view, the first attempt to apply the fractional concept to traffic modelling was to use fBm $B_H(t)$ instead of traditional Poisson-based models. Compared to standard Brownian motion (BM), fBm has one extra parameter, the Hurst parameter H , which quantifies the strength of the fractional scaling. It is said usually, that fBm is self-similar, or fractional, with Hurst parameter H . In [10], Norros has proposed the following model for cumulative traffic

$$A(t) = mt + (\sigma m)^{1/2} B_H(t)$$

where $m > 0$ is the mean input rate, σ is the scale factor. This model has been widely studied and have been proposed asymptotic lower bounds [10] as well as exact asymptotics in the case of large buffers [11, 12].

It is important to point out that in this case, one has a long-range correlation, but not heavy tails of marginal distributions.

To deal with this issue, several papers extended Norros model by modelling the input traffic as α -stable Levy motion [13, 14] or, to take into account also the long range correlations, fractional α -stable Levy motion (see [15, 16]).

In the paper, a new variant of fractional Levy motion is suggested and, following the approach proposed in [10], an asymptotic lower bound for the overflow probability is determined.

2 Stable Distributions and Processes

Levy processes have been popular in modeling the teletraffic. Below, some definitions are given and some properties of such processes are considered.

Definition 1. A stochastic process $Y = (Y(t), t \geq 0)$ is a Levy process if

- (1) $Y(0) = 0$ almost surely;
- (2) Y has independent increments; and
- (3) Y has stationary increments.

Usually, for the sake of regularity, the following property is required: with probability one all trajectories of Y are right-continuous and have finite limits from the left.

The distributions of the process Y is defined uniquely by the distribution of random variable $Y(1)$, which is infinitely divisible.

The most familiar example of Levy process is the BM (Weiner process).

Definition 2. A Levy process $B = (B(t), t \geq 0)$ is called *Brownian Motion* if for any $t \geq 0, h > 0$ the increment $B(t+h) - B(t)$ has Gaussian distribution with zero mean and variance $\sigma^2 h$.

If $\sigma^2 = 1$, one has a standard BM. It is easily seen that

$$K(t, s) = \text{Cov}(Y(t), Y(s)) = \sigma^2 \min(t, s).$$

By definition, BM has Gaussian distributions. Such distributions have been got for normalized sums of independent identically distributed random variables with finite variance. In the case of infinite variance, the so-called stable distributions are considered.

Definition 3. A random variable Y is said to have an α -stable distribution if its characteristic function has the following form:

$$\begin{aligned}\varphi(\omega) &:= E[e^{j\omega X}] \\ &= \exp\{j\mu\omega - \sigma|\omega|^\alpha[1 - j\beta \operatorname{sgn}(\omega)\theta(\omega, \alpha)]\}\end{aligned}$$

where $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$, $\mu \in \mathbb{R}^1$, and

$$\theta(\omega, \alpha) = \begin{cases} \tan\left(\frac{\alpha\pi}{2}\right), & \alpha \neq 1; \\ -\frac{2}{\pi} \ln|\omega|, & \alpha = 1. \end{cases}$$

Parameter α is called *characteristic exponent* and specifies the level of burstiness in distribution, i.e., it specifies the weight of the tails of the distribution. σ and μ are called *scale* and *location parameters*. β is called *skewness parameter*. If $\beta = 0$ then X is symmetrically distributed around μ . If $0 < \alpha < 1$, $\mu = 0$ and $\beta = 1$ then X has positive values with probability 1. In what follows, a random variable Y is said to have standard α -stable distribution if $\mu = 0$ and $\sigma = 1$.

The α -stable distribution is infinitely divisible. So, it generates some Levy process.

Definition 4. A stochastic process $L_\alpha = (L_\alpha(t), t \geq 0)$ is said to be an α -stable Levy motion if it is a Levy process such that $L_\alpha(1)$ has a given α -stable distribution.

If the distribution of $L_\alpha(1)$ is totally positive skewed ($0 < \alpha < 1$, $\beta = 1$), then all trajectories of the process L_α are nondecreasing and nonnegative. Such process is called α -stable subordinator.

If $\alpha = 2$, $\mu = 0$, one has again BM B .

There exists very interesting relation between α -stable Levy motions with different α .

Theorem 1. If $(L_{\alpha_1}(t), t \geq 0)$, $0 < \alpha_1 \leq 2$, is a α_1 -stable Levy motion with symmetric distributions, $(L_{\alpha_2}(t), t \geq 0)$, $0 < \alpha_2 < 1$, is a α_2 -stable subordinator, then stochastic process $Y = (Y(t) := L_{\alpha_1}(L_{\alpha_2}(t)), t \geq 0)$ is $\alpha_1\alpha_2$ -stable Levy motion with symmetric distributions.

This theorem is a corollary of the following result by Zolotarev [17, theorem 3.3.1].

Theorem 2. If Y_1 has symmetric α_1 -stable distribution, $0 < \alpha_1 \leq 2$, Y_2 has one-sided α_2 -stable distribution, $0 < \alpha_2 < 1$, then random variable $Y = Y_1 Y_2^{1/\alpha_1}$ has symmetric $\alpha_1\alpha_2$ -stable distribution.

In particular, for $\alpha_1 = 2$ and $0 < \alpha_2 = \alpha/2 < 1$, one gets the following

Theorem 3. If $B = (B(t), t \geq 0)$ is the Brownian motion, $L_{\alpha/2} = (L_{\alpha/2}(t), t \geq 0)$ is a $\alpha/2$ -stable subordinator, then $L_\alpha = (L_\alpha(t) := B(L_{\alpha/2}(t)), t \geq 0)$, $0 < \alpha < 2$, is an α -stable Levy motion with symmetric distributions.

3 Self-Similar Processes

Definition 5. A process $Y = (Y(t), t \geq 0)$ is self-similar, with Hurst parameter $H \geq 0$, if it satisfies the condition

$$Y(t) \stackrel{d}{=} c^{-H} Y(ct), \quad \forall t \geq 0, \quad \forall c > 0,$$

where the equality is the sense of finite-dimensional distributions.

Two of the most popular examples of self-similar processes are fBm and α -stable Levy motion.

Definition 6. The fractional Brownian motion with Hurst parameter H is a Gaussian process $(B_H(t), t \geq 0)$ with zero mean and correlation function

$$K_H(t, s) = \frac{1}{2} [|t|^{2H} + |s|^{2H} - |t - s|^{2H}].$$

The definition of α -stable Levy motion see above.

More information about stable and self-similar processes can be found in [18, 19].

4 New Variant of Fractional Levy Motion

Above, it was shown how to get symmetric α -stable Levy motion using BM and $\alpha/2$ -stable subordinator. Below, it is proposed to use the same construction to get fractional Levy motion from fBm B_H and $\alpha/2$ -stable subordinator $L_{\alpha/2}$.

Let $(B_H(t), t \geq 0)$ be the fBm with Hurst parameter H , $(L_\alpha^1(t), t \geq 0)$, $(L_\alpha^2(t), t \geq 0)$ be standard α -stable subordinators, $0 < \alpha < 1$, and B_H , L_α^1 and L_α^2 are independent. Consider the new process

$$X(t) := \begin{cases} B_H(L_\alpha^1(t)), & t \geq 0; \\ -B_H(L_\alpha^2(t)), & t < 0. \end{cases}$$

Theorem 4. The above process X is self-similar with Hurst parameter $H_1 = H/\alpha$.

Proof. The processes $(L_\alpha^k(t), t \geq 0)$, $k = 1, 2$, are α -stable and self-similar with Hurst parameter $1/\alpha$. So, one has

$$(L_\alpha^k(ct), t \geq 0) \stackrel{d}{=} (c^{1/\alpha} L_\alpha^k(t), t \geq 0).$$

Then,

$$\begin{aligned}(X(ct), t \in \mathbb{R}^1) &= \pm B_H(L_\alpha^k(c|t|), t \in \mathbb{R}^1) \\ &\stackrel{d}{=} (\pm B_H(c^{1/\alpha} L_\alpha^k(|t|), t \in \mathbb{R}^1)).\end{aligned}$$

Using self-similarity of B_H for fixed $\tau = L_\alpha^k(|t|)$, for any $a > 0$, one has

$$(B_H(a\tau), \tau \geq 0) \stackrel{d}{=} (a^H B_H(\tau), \tau \geq 0)$$

or

$$(\pm B_H(c^{1/\alpha}\tau), \tau \geq 0) \stackrel{d}{=} (\pm c^{H/\alpha} B_H(\tau), \tau \geq 0).$$

Due to the complete probability formula, the result is obtained.

Corollary 1. For any $t > 0$,

$$X(t) \stackrel{d}{=} (L_\alpha^1(t))^H Y$$

where Y has standard normal distribution and $L_\alpha^1(t)$ and Y are independent.

Remark. Hurst parameter H_1 for above process X can be any positive number. But for traffic applications, it is more interesting the case where $1/2 < H_1 < 1$. So, it is assumed in what follows.

Theorem 5. The above process X has stationary increments.

Proof. Fractional Brownian motion B_H has stationary increments. So for any $t_1 < t_2$

$$B_H(t_2) - B_H(t_1) \stackrel{d}{=} B_H(t_2 - t_1).$$

Then, for any $t \geq 0$, $h > 0$ and fixed $L_\alpha^k(t+h) = t_2$, $L_\alpha^k(t) = t_1$, one has

$$B_H(L_\alpha^k(t+h)) - B_H(L_\alpha^k(t)) \stackrel{d}{=} B_H(L_\alpha^k(t+h) - L_\alpha^k(t)).$$

Due to the complete probability formula, one has the same for random moments of time. The process $L_\alpha^k(t)$ has stationary increments too. So, one gets

$$B_H(L_\alpha^k(t+h) - L_\alpha^k(t)) \stackrel{d}{=} B_H(L_\alpha^k(h)).$$

5 Application to Traffic Modeling

Define the cumulative traffic (or arrival) process $A(t)$, i. e., the total amount of load produced by a source in the time interval $[0, t]$, $t > 0$, by

$$A(t) := mt + (\sigma m)^{1/\beta} X(t),$$

where $m > 0$ is the mean input rate, σ is the scale factor, $\beta = \alpha/H = 1/H_1$, X is the process defined above.

Consider a single server queue with constant service rate $r > 0$ and infinite buffer space, where input is the stable self-similar process defined above ($r > m$ for stability). The buffer occupancy $Q(t, r)$ at time $t \in \mathbb{R}^1$ (queue size or queue length) can be written as

$$Q(t, r) = \sup_{s \leq t} (A(t) - A(s) - r(t - s)).$$

Due to theorem 2, the process $Q = (Q(t, r), t \in \mathbb{R}^1)$, is stationary. So, the most interesting is the following probability of overflow:

$$\varepsilon(b) = P(Q(0, r) > b) = P\left(\sup_{\tau \geq 0} (A(\tau) - r\tau) > b\right).$$

Using the technique elaborated in papers [1, 13], one can get the lower bound for the probability of buffer overflow for large b .

It is easily seen that

$$\begin{aligned} \varepsilon(b) &\geq \sup_{\tau \geq 0} P((A(\tau) - r\tau) > b) \\ &= \sup_{\tau \geq 0} P(m\tau + (\sigma m)^{1/\beta} X(\tau) - r\tau > b) \\ &= \sup_{\tau \geq 0} P\left(X(\tau) > \frac{b + (r - m)\tau}{(\sigma m)^{1/\beta}}\right) \\ &= \sup_{\tau \geq 0} P\left(\tau^{1/\beta} X(1) > \frac{b + (r - m)\tau}{(\sigma m)^{1/\beta}}\right) \\ &= \sup_{\tau \geq 0} P\left(X(1) > \frac{b + (r - m)\tau}{(\sigma m\tau)^{1/\beta}}\right). \end{aligned}$$

Last probability under supremum is a decreasing function of the value

$$f(\tau) = \frac{b + (r - m)\tau}{(\sigma m\tau)^{1/\beta}}.$$

Elementary calculations give us that the minimal value of this function is achieved at the point

$$\tau_0 = \frac{b}{\beta(1 - 1/\beta)(r - m)} = \frac{bH_1}{(1 - H_1)(r - m)}.$$

It follows

$$\varepsilon(b) \geq P(X(1) > f(\tau_0) = b_1)$$

where

$$b_1 = \frac{(r - m)^{H_1}(1 - H_1)^{-(1-H_1)}}{(\sigma m H_1)^{H_1}} b^{1-H_1}.$$

Using corollary 1, one gets

$$\begin{aligned} P(X(1) > b_1) &= P((L_\alpha^1(1))^H Y > b_1) \\ &\geq P((L_\alpha^1(1))^H Y > b_1, Y > 1) \\ &\geq P((L_\alpha^1(1))^H > b_1, Y > 1) \\ &= P(L_\alpha^1(1) > (b_1)^{1/H}) P(Y > 1). \end{aligned}$$

For large $x > 0$ (see [20, theorem 2.4.1]), one has

$$P(L_\alpha^1(1) > x) \sim C(\alpha) x^{-\alpha}$$

where

$$C(\alpha) = \frac{\sin(\pi\alpha)}{\pi} \Gamma(\alpha).$$

It follows for large b

$$\begin{aligned} \varepsilon(b) &\geq C(\alpha) (b_1)^{-1/H_1} P(Y > 1) \\ &= C_1(\alpha, H_1) \sigma \frac{m}{r - m} b^{-(1-H_1)/H_1}. \end{aligned}$$

Finally, one has the following

Theorem 6. *An asymptotic lower bound for the overflow probability is given by*

$$\varepsilon(b) \geq C_1(\alpha, H_1) \sigma \frac{m}{r-m} b^{-(1-H_1)/H_1}, \quad b \rightarrow \infty.$$

References

1. Leland W., Taqqu M., Willinger W., Wilson D. On the selfsimilar nature of Ethernet traffic (extended version) // IEEE/ACM Trans. Networking, 1994. P. 1–15.
2. Park K., Kim G., Crovella M. On the relationship between file sizes, transport protocols, and self-similar network traffic // Conference (International) on Network Protocols Proceedings, October 1996. P. 171–180.
3. Crovella M. E., Bestavros A. Self-similarity in world wide web traffic: Evidence and possible causes // IEEE/ACM Transactions on Networking, December 1997. Vol. 5. No. 6. P. 835–846.
4. Willinger W., Paxson V. Where mathematics meets the Internet // Notices of the AMS, 1998. Vol. 45. No. 8. P. 961–970.
5. Park K., Willinger W. Self-similar network traffic and performance evaluation. — Wiley, 2000.
6. Taqqu M. S., Levy J. B. Using renewal processes to generate long-range dependence and high variability // Dependence in probability and statistics / Eds. E. Eberlein and M. S. Taqqu. — Boston: Birkhauser, 1986. P. 73–89.
7. Taqqu M. S., Willinger W., Serman R. Proof of a fundamental result in self-similar traffic modeling // Computer Communications Rev., 1997. Vol. 27. No. 2. P. 5–23.
8. Levy J. B., Taqqu M. S. Renewal reward processes with heavy-tailed inter-arrival times and heavy tailed rewards // Bernoulli, 2000. Vol. 6. No. 1. P. 23–44.
9. Mikosch Th., Resnick S., Rootzen H., Stegeman A. Is network traffic approximated by stable Levy motion or fractional Brownian motion? // Ann. Appl. Probab., 2002. Vol. 12. No. 1. P. 23–68.
10. Norros I. A storage model with self-similar input // Queuing Syst., 1994. Vol. 16. P. 387–396.
11. Narayan O. Exact asymptotic queue length distribution for fractional Brownian traffic // Adv. Perf. Anal., 1998. Vol. 1. P. 39–63.
12. Hüsler J., Piterbarg V. Extremes of a certain class of Gaussian processes // Stoch. Proc. Appl., 1999. Vol. 83. P. 257–271.
13. Laskin N., Lambadaris I., Harmantzis F. C., Devetsikiotis M. Fractional Levy motion and its application to network traffic modeling // Computer Networks, 2002. Vol. 40. P. 363–375.
14. Garroppo R. G., Giordano S., Pagano M., Procissi G. Testing α -stable processes in capturing the queuing behavior of broadband teletraffic // Signal Proc., 2002. Vol. 82. P. 1861–1872.
15. Gallardo J. R., Makrakis D., Orozco-Barbosa L. Use of α -stable self-similar stochastic processes for modeling traffic // Performance Eval., 2000. Vol. 40. P. 71–98.
16. Karasaridis A. Network heavy traffic modeling using α -stable self-similar processes // IEEE Transactions on Communications, July 2001. Vol. 49. No. 7. P. 1203–1214.
17. Zolotarev V. M. One-dimensional stable distributions. — Translations of mathematical monographs. AMS, 1986. Vol. 65.
18. Samorodnitsky G., Taqqu M. S. Stable non-Gaussian random processes. — Chapman & Hall, 1994.
19. Embrechts P., Maejima M. Self-similar process. — Princeton University Press, 2002.
20. Ibragimov I. A., Linnik Yu. V. Independent and stationary sequences of random variables. — Groningen: Wolters-Noordhoff, 1971.

ДРОБНОЕ ДВИЖЕНИЕ ЛЕВИ С ЗАВИСИМЫМИ ПРИРАЩЕНИЯМИ И ЕГО ПРИЛОЖЕНИЕ К МОДЕЛИРОВАНИЮ СЕТЕВОГО ТРАФИКА

К. Де Никола¹, Ю. С. Хохлов², М. Пагано³, О. И. Сидорова⁴

¹Университет г. Салерно, denicola@diima.unisa.it

²Российский университет дружбы народов, yskhokhlov@yandex.ru

³Университет г. Пиза, m.pagano@iet.unipi.it

⁴Тверской государственный университет, Oksana.I.Sidorova@yandex.ru

Аннотация: С начала 1990-х гг. были проведены многочисленные высокоточные измерения для различных сетевых сценариев, которые показали, что трафик в Интернете проявляет сильную иррегулярность, выраженную в чрезвычайной вариабельности, а также в наличии долговременной зависимости. Эти новые особенности, которые не удастся описать экономным образом с помощью традиционных марковских моделей, имеют сильное влияние на поведение сети, и это привело к необходимости введения в моделирование сетевого трафика α -устойчивых распределений и самоподобных процессов. В настоящей работе рассматривается некоторое обобщение дробного броуновского движения, которое позволяет охватить одновременно обе отмеченные выше особенности реального трафика.

Ключевые слова: дробное броуновское движение; α -устойчивый субординатор; самоподобные процессы; вероятность переполнения буфера

ABOUT THE RATE OF CONVERGENCE OF ONE U-STATISTIC

O. Yanushkevichienė¹ and R. Yanushkevichius²

Abstract: This paper pertains to the analysis of the asymptotic behavior of U-statistics, which are important in the construction and application of modern statistical methods for studying information systems.

Keywords: rates of convergence; Berry–Esseen bound; quadratic form; second-degree U-statistics; Kolmogorov distance

1 Introduction and Formulation of the Result

Analysis of modern information systems is impossible without the use of statistical methods. Function exploring the characteristics of the traffic is usually characterized by a symmetry property with respect to the arguments. This makes it especially important to study the behavior of U-statistics. This paper pertains to the investigation of the limit distribution of the second-order U-statistic.

Let X, X_1, \dots, X_n be independent identically distributed (i.i.d.) random variables taking values in a measurable space (Θ, \mathcal{R}) . Let $h : \Theta^2 \rightarrow \mathbf{R}$ be real-valued measurable functions. Let h be symmetric, that is, $h(x, y) = h(y, x)$ for all $x, y \in \Theta$. Consider the U-statistic

$$Z = Z(X_1, \dots, X_n) = n^{-1} \sum_{1 \leq i < k \leq n} h(X_i, X_k) \quad (1)$$

assuming that $\beta_2 = \mathbf{E}|h(X, X_1)|^2 > 0$, $\beta_3 = \mathbf{E}|h(X, X_1)|^3 < \infty$, and $\mathbf{E}h(x, X) = 0$ for all $x \in \Theta$.

The condition $\beta_2 > 0$ ensures that the quadratic part of the statistic is not asymptotically negligible and, therefore, Z is not asymptotically normal. To be more precise, the asymptotic distribution of Z is non-Gaussian and is given by the distribution of the random variable

$$Z_0 = \frac{1}{2} \sum_{i \geq 1} q_i (\eta_i^2 - 1).$$

Here, η_i is the sequence of i.i.d. standard normal random variables and q_1, q_2, \dots denote the eigenvalues of the Hilbert–Schmidt operator, say Q , associated with the kernel h (see [1] for detailed definitions). Without loss of generality, let assume throughout that $|q_1| \geq |q_2| \geq \dots$

Let

$$F(x) = \mathbf{P}\{Z \leq x\}; \quad F_0(x) = \mathbf{P}\{Z_0 \leq x\}.$$

Let write

$$\Delta_n = \rho(Z, Z_0) = \sup_x |F(x) - F_0(x)|.$$

Korolyuk and Borovskikh proved [2] that $\Delta_n = o(n^{-1/2})$ if $|q_i| > 0$ for all i . Under the condition $|q_9| > 0$, an optimal bound $\Delta_n = O(n^{-1})$ was established by Bentkus and Götze [1]. From a construction in [1], it is clear that the bounds for Δ_n are related to the estimates of the convergence rate in Hilbert spaces. Therefore, if $|q_k| > 0$, a result of Senatov [3] implies a lower bound for Δ_n , namely, $\liminf_{n \rightarrow \infty} \Delta_n n^{k/12} (q_1 \cdots q_k)^{1/2} > 0$, then $k \leq 6$. In the considered case, $k = 1$ and the lower bound is $n^{k/12}$. The rate of convergence of such order was got in [4] for the case, when the expression (1) has not only the quadratic part, but also the linear part. Now, the question is: what order of the rate of convergence will be, if the linear part does not exist? The theorem below answers this question.

By c , positive absolute constants which may differ from line to line or from formula to formula are denoted. The following theorem is the authors' result.

Theorem 1. *If $q_1 > 0$, one gets*

$$\Delta_n \leq \frac{c\beta_3^{1/6}}{\sqrt{q_1}n^{1/14}} + \frac{c}{n^{1/4}} \left(\sum_{i \geq 1} \left| \frac{q_i}{q_1} \right| \right)^{1/2} + \frac{c}{\sqrt{q_1}n}. \quad (2)$$

Remark 1. If on the right-hand side of (2) the sum $\sum_{i \geq 1} |q_i|$ is divergent, i. e.,

$$\sum_{i \geq 1} |q_i| = \infty, \quad (3)$$

then $\Delta_n = O(n^{-1})$.

Indeed, in principle, the series on the right-hand side of (2) can diverge. Then, the estimate (2) is naturally true, but useless. However, if (3) holds, then $|q_9| > 0$, since $|q_1| \geq |q_2| \geq \dots \geq |q_9| \geq \dots \geq |q_n| \geq 0$. Then, let use Theorem 1.1 from [1]. The theorem states that if $|q_9| > 0$, then $\Delta_n = O(n^{-1})$. That is, if the series mentioned is divergent, then estimate (2) can be replaced by the improved estimate from [1].

In the one-dimensional case, that is, in the case $q_2 = q_3 = \dots = 0$, the bound (2) $\Delta_n = O(n^{-1/14})$

¹Vilnius University, Institute of Mathematics and Informatics, Lithuania, olgjan@zebra.lt

²Lithuanian University of Educational Sciences, Vilnius, Lithuania

improves to $\Delta_n = O(n^{-1/4})$, and the rate $O(n^{-1/4})$ is best possible [5]. The result of Senatov [3] shows that the optimal order of bound (2) is $O(n^{-1/12})$.

The authors' results answer the question of V. Benktus, who also suggested an idea of the proof.

2 Proof

2.1 Lemma

To prove the theorem, one needs the following lemma.

Lemma 1. *Let η be a standard Gaussian random variable and $q_1 > 0$. Then the distribution function $H(x) = P\{q_1\eta^2 < x\}$ satisfies the Lipschitz condition*

$$\sup_x |H(x + \varepsilon) - H(x)| \leq c \sqrt{\frac{\varepsilon}{q_1}} \quad (4)$$

where $\varepsilon > 0$.

Proof of Lemma 1. The distribution of η^2 has the density

$$f(x) = \begin{cases} \frac{1}{\sqrt{2}\Gamma(1/2)\sqrt{x}} e^{-x/2} & \text{for } x > 0; \\ 0 & \text{for } x \leq 0. \end{cases}$$

The function $f(u)$ is decreasing for $u > 0$. One can write:

$$\sup_x |H(x + \varepsilon) - H(x)| \leq c \int_0^{\varepsilon/q_1} \frac{e^{-u}}{\sqrt{u}} du \leq c \sqrt{\frac{\varepsilon}{q_1}}.$$

The lemma is proved.

Lemma 2. *Let η be a standard Gaussian random variable and $q_1 > 0$. Then the the following inequality is true:*

$$\sup_x \left| P(q_1\eta_1^2 - \frac{1}{n}q_1\eta_1^2 \leq x) - P(q_1\eta_1^2 \leq x) \right| \leq \frac{c}{\sqrt{q_1 n}}.$$

Proof of Lemma 2. It is easy to see that

$$\begin{aligned} & \left| P\left(q_1\left(1 - \frac{1}{n}\right)\eta_1^2 \leq x\right) - P(q_1\eta_1^2 \leq x) \right| \\ & \leq c \int_0^{2/\sqrt{q_1 n}} e^{-t^2/2} dt \leq \frac{c}{\sqrt{q_1 n}}. \end{aligned}$$

The lemma is proved.

It is necessary to prove the authors' theorem, i.e., the bound (2). Let $G, G_i, 1 \leq i \leq n$, be i.i.d. Gaussian random vectors $G_i = (G_{1,i}, G_{2,i}, \dots)$ with values in \mathbb{R}^∞ where $G_{1,i}, G_{2,i}, \dots$ denote i.i.d. standard normal random variables. Let assume throughout that

$$\left. \begin{aligned} \mathbf{E}h(x, G) &= 0; \\ \mathbf{E}h(x, G)h(y, G) &= \mathbf{E}h(x, X)h(y, X) \end{aligned} \right\} \quad (5)$$

for all $x \in \Theta$. Note that the possibility of selecting Gaussian random variables so that Eqs. (5) are valid is substantiated in [1, p. 461].

According to the triangle inequality, one has

$$\rho(Z, Z_0) \leq \rho(Z(X_1, \dots, X_n), Z(G_1, \dots, G_n)) + \rho(Z(G_1, \dots, G_n), Z_0). \quad (6)$$

In view of (6), to prove (2), it is sufficient to establish that

$$\begin{aligned} \rho(Z(X_1, \dots, X_n), Z(G_1, \dots, G_n)) \\ \leq cq_1^{-1/2} \beta_3^{1/6} n^{-1/14} \end{aligned} \quad (7)$$

and

$$\begin{aligned} \rho(Z(G_1, \dots, G_n), Z_0) \\ \leq \frac{c}{n^{1/4}} \left(\sum_{i \geq 1} \left| \frac{q_i}{q_1} \right| \right)^{1/2} + \frac{c}{\sqrt{q_1 n}}, \end{aligned} \quad (8)$$

2.2 Proof of inequality (7)

Let estimate $\rho(Z(X_1, \dots, X_n), Z(G_1, \dots, G_n))$. It is to verify using lemma 1 that the distribution function of $Z(G_1, \dots, G_n)$ satisfies the Lipschitz condition with the exponent 1/2.

Let prove that, for any $\varepsilon > 0$, one has

$$\rho(Z(X_1, \dots, X_n), Z(G_1, \dots, G_n)) \leq \frac{c}{\sqrt{q_1}} \sqrt{\varepsilon} + \Delta \quad (9)$$

where

$$\Delta = \max_{\varphi} |\mathbf{E}\varphi(Z(X_1, \dots, X_n)) - \mathbf{E}\varphi(Z(G_1, \dots, G_n))|,$$

and maximum is taken over all infinite differentiable φ such that $0 \leq \varphi(u) \leq 1$ and $|\varphi^{(k)}(u)| \leq c_1/\varepsilon^k$, $k = 1, 2, 3$:

$$\varphi(u) = \begin{cases} 1, & \text{if } u \leq x - \varepsilon; \\ 0, & \text{if } u \geq x, \end{cases} \quad (10)$$

or

$$\varphi(u) = \begin{cases} 1, & \text{if } u \leq x; \\ 0, & \text{if } u \geq x + \varepsilon. \end{cases} \quad (11)$$

Let write $\delta^* = P\{Z(X_1, \dots, X_n) \leq x\} - P\{Z(G_1, \dots, G_n) \leq x\}$. Let prove (9) in the case $\delta^* \geq 0$. Let take a function φ such that (10) holds. Then

$$\begin{aligned}\delta^* &= EI\{Z(X_1, \dots, X_n) \leq x\} \\ &\quad - P\{Z(G_1, \dots, G_n) \leq x\} \\ &\leq |E\varphi(Z(X_1, \dots, X_n)) - E\varphi(Z(G_1, \dots, G_n))| \\ &\quad + |E\varphi(Z(G_1, \dots, G_n)) - P\{Z(G_1, \dots, G_n) \leq x\}| \\ &\leq \Delta + P\{x \leq Z(G_1, \dots, G_n) \leq x + \varepsilon\}.\end{aligned}$$

Using the Lipschitz condition, one gets (9).

The proof is similar if $\delta^* < 0$, taking a function φ satisfying (11) instead of (10).

Let estimate

$$\Delta^*(\varphi) = |E\varphi(Z(X_1, \dots, X_n)) - E\varphi(Z(G_1, \dots, G_n))|.$$

It is easy to see that

$$\begin{aligned}\Delta^*(\varphi) &\leq |E\varphi(Z(X_1, \dots, X_n)) \\ &\quad - E\varphi(Z(X_1, \dots, X_{n-1}, G_n))| \\ &\quad + |E\varphi(Z(X_1, \dots, X_{n-1}, G_n)) \\ &\quad - E\varphi(Z(X_1, \dots, X_{n-2}, G_{n-1}, G_n))| + \dots \\ &\quad \dots + |E\varphi(Z(X_1, G_2, \dots, G_n)) \\ &\quad - E\varphi(Z(G_1, \dots, G_n))| = \Delta_{1,n}^* + \dots + \Delta_{n,n}^*.\end{aligned}\quad (12)$$

One has

$$\begin{aligned}Z(X_1, \dots, X_n) &= \frac{1}{n}\{h(X_1, X_2) + \dots \\ &\quad \dots + h(X_1, X_{n-1}) + h(X_1, X_n) + \dots \\ &\quad \dots + h(X_{n-2}, X_{n-1}) + h(X_{n-2}, X_n) + \\ &\quad \quad + h(X_{n-1}, X_n)\}.\end{aligned}$$

one designates:

$$\begin{aligned}w &= \frac{1}{n}\{h(X_1, X_2) + \dots + h(X_1, X_{n-1}) + \dots \\ &\quad \dots + h(X_{n-2}, X_{n-1})\}\end{aligned}$$

and

$$\begin{aligned}l_n &= \frac{1}{n}\{h(X_1, X_n) + \dots + h(X_{n-2}, X_n) \\ &\quad \quad + h(X_{n-1}, X_n)\}.\end{aligned}$$

Replacing X_n by G_n , one gets:

$$\begin{aligned}Z(X_1, \dots, X_{n-1}, G_n) &= w + \frac{1}{n}\{h(X_1, G_n) + \dots \\ &\quad \dots + h(X_{n-2}, G_n) + h(X_{n-1}, G_n)\} = w + l_n^*.\end{aligned}$$

Let expand into the Taylor series:

$$\begin{aligned}\varphi(x+y) &= \varphi(x) + \varphi'(x)y + \frac{1}{2}\varphi''(x)y^2 \\ &\quad + \frac{1}{2}E\varphi'''(x+\tau y)(1-\tau)^2y^3.\end{aligned}$$

Here, τ is the random variable uniformly distributed in $[0, 1]$ and independent of all the other random variables. Let replace x by w and y by l_n , respectively. Let write

$$\begin{aligned}\Delta_{1,n}^* &= |E\varphi(Z(X_1, \dots, X_n)) - E\varphi(Z(X_1, \dots, X_{n-1}, G_n))| \\ &= |E\varphi(w) - E\varphi(w) + E\varphi'(w)l_n - E\varphi'(w)l_n^* \\ &\quad + \frac{1}{2}E\varphi''(w)l_n^2 - \frac{1}{2}E\varphi''(w)(l_n^*)^2 \\ &\quad + \frac{1}{2}E\varphi'''(w+l_n\tau)l_n^3(1-\tau)^2 \\ &\quad \quad - \frac{1}{2}E\varphi'''(w+l_n^*\tau)(l_n^*)^3(1-\tau)^2|.\end{aligned}$$

Now, let condition $X_1, \dots, X_{n-1}, \tau$:

$$\begin{aligned}\Delta_{1,n}^* &= |E_{X_1, \dots, X_{n-1}, \tau}E_{X_n}\varphi'(w)l_n \\ &\quad - E_{X_1, \dots, X_{n-1}, \tau}E_{X_n}\varphi'(w)l_n^* \\ &\quad + \frac{1}{2}E_{X_1, \dots, X_{n-1}, \tau}E_{X_n}\varphi''(w)l_n^2 \\ &\quad - \frac{1}{2}E_{X_1, \dots, X_{n-1}, \tau}E_{X_n}\varphi''(w)(l_n^*)^2 \\ &\quad + \frac{1}{2}E\varphi'''(w+l_n\tau)l_n^3(1-\tau)^2 \\ &\quad \quad - \frac{1}{2}E\varphi'''(w+l_n^*\tau)(l_n^*)^3(1-\tau)^2|.\end{aligned}$$

Using (5), one gets:

$$\begin{aligned}\Delta_{1,n}^* &= \left| \frac{1}{2}E\varphi'''(w+l_n\tau)l_n^3(1-\tau)^2 \right. \\ &\quad \left. - \frac{1}{2}E\varphi'''(w+l_n^*\tau)(l_n^*)^3(1-\tau)^2 \right|.\end{aligned}\quad (13)$$

It is easy to see that

$$\begin{aligned}|E\varphi'''(w+l_n\tau)l_n^3(1-\tau)^2| &\leq \frac{c}{\varepsilon^3}E\mathbf{I}(x \leq w + \tau l_n \\ &\leq x + \varepsilon)|l_n|^3 \leq \frac{c}{\varepsilon^3}E|l_n|^3.\end{aligned}$$

From theorem 20 in [5, p. 89], one gets:

$$\begin{aligned}E|l_n|^3 &= |E(h(X_1, X_n) + \dots + h(X_{n-2}, X_n) \\ &\quad + h(X_{n-1}, X_n))|^3 \leq c(n-1)^{3/2}\beta_3.\end{aligned}\quad (14)$$

Combining (9), (12)–(14), one has

$$\begin{aligned}\rho(Z(X_1, \dots, X_n), Z(G_1, \dots, G_n)) &\leq \frac{c}{\sqrt{q_1}}\sqrt{\varepsilon} + \frac{c}{\varepsilon^3}(n-1)^{3/2}\beta_3.\end{aligned}$$

Let $\varepsilon = n^{-1/7} \beta_3^{1/3}$, then one gets:

$$\rho(Z(X_1, \dots, X_n), Z(G_1, \dots, G_n)) \leq c_2 q_1^{-1/2} \beta_3^{1/6} n^{-1/14}.$$

2.3 Proof of inequality (8)

Now, let estimate $\rho(Z(G_1, \dots, G_n), Z_0)$.

It has been shown in [1] that it is possible to represent the statistic $Z(G_1, \dots, G_n)$ in the form:

$$Z(G_1, \dots, G_n) = n^{-1} \sum_{1 \leq i < k \leq n} \langle QG_i, G_k \rangle$$

where $a = (a_i)_{i \geq 1}$ is some constant. Let $G_{i,j}$, $i \geq 1$ be the components of vector G_j . Then, one can rewrite this expression in the form:

$$Z(G_1, \dots, G_n) = n^{-1} \sum_{i \geq 1} q_i \sum_{1 \leq j < k \leq n} G_{i,j} G_{i,k}$$

where $G_{i,j}$ are the normally distributed random variables with the moments $(0, 1)$. It is easy to see that

$$n^{-1} q_i \sum_{1 \leq j < k \leq n} G_{i,j} G_{i,k} = (n-1) q_i \bar{G}_i^2 - q_i S_i^2$$

where

$$\bar{G}_i = n^{-1} \sum_{1 \leq j \leq n} G_{i,j} \text{ and } S_i^2 = n^{-1} \sum_{1 \leq j \leq n} G_{i,j}^2 - \bar{G}_i^2.$$

It is well known [6] that random variables \bar{G}_i and S_i are independent. The random variables \bar{G}_i can be written in the form $\bar{G}_i = \eta_i / \sqrt{n}$; so, one can write:

$$n^{-1} q_i \sum_{1 \leq j < k \leq n} G_{i,j} G_{i,k} = \frac{n-1}{n} q_i \eta_i^2 - q_i S_i^2.$$

Then

$$\begin{aligned} Z(G_1, \dots, G_n) &= \sum_{i \geq 1} \left(\frac{n-1}{2n} q_i \eta_i^2 - \frac{q_i}{2} S_i^2 \right) \\ &= \sum_{i \geq 1} \left(\frac{1}{2} q_i (\eta_i^2 - 1) - \frac{1}{2n} q_i \eta_i^2 - \frac{q_i}{2} (S_i^2 - 1) \right). \end{aligned}$$

Assume

$$\begin{aligned} \rho(Z(G_1, \dots, G_n), Z_0) &= \sup_x \left| \mathbf{P} \left(\sum_{i \geq 1} \left(\frac{1}{2} q_i (\eta_i^2 - 1) - \frac{1}{2n} q_i \eta_i^2 - \frac{q_i}{2} (S_i^2 - 1) \right) \leq x \right) \right. \\ &\quad \left. - \mathbf{P} \left(\frac{1}{2} \sum_{i \geq 1} q_i (\eta_i^2 - 1) \leq x \right) \right| \end{aligned}$$

$$\begin{aligned} &= \sup_x \left| \mathbf{P} \left(\sum_{i \geq 1} \left(\frac{1}{2} q_i (\eta_i^2 - 1) - \frac{1}{2n} q_i \eta_i^2 - \frac{q_i}{2} (S_i^2 - 1) \right) \leq x \right) \right. \\ &\quad \left. - \mathbf{P} \left(\sum_{i \geq 1} \left(\frac{1}{2} q_i (\eta_i^2 - 1) - \frac{1}{2n} q_i \eta_i^2 \right) \leq x \right) \right| \\ &+ \sup_x \left| \mathbf{P} \left(\sum_{i \geq 1} \left(\frac{1}{2} q_i (\eta_i^2 - 1) - \frac{1}{2n} q_i \eta_i^2 \right) \leq x \right) \right. \\ &\quad \left. - \mathbf{P} \left(\frac{1}{2} \sum_{i \geq 1} q_i (\eta_i^2 - 1) \leq x \right) \right| = v_1 + v_2. \quad (15) \end{aligned}$$

Let estimate v_1 . Using the independence of S_i , η_i , and η_j , $i \neq j$ and inequality (4), one obtains:

$$v_1 \leq c \left(|q_1|^{-1} \sum_{i \geq 1} |q_i| \mathbf{E} |S_1^2 - 1| \right)^{1/2}.$$

It follows that

$$\begin{aligned} (\mathbf{E} |S_1^2 - 1|)^{1/2} &= \left(\mathbf{E} |n^{-1} \sum_{1 \leq j \leq n} (G_{1,j}^2 - 1) - \bar{G}_1^2| \right)^{1/2} \\ &\leq c \left(\mathbf{E} |n^{-1} \sum_{1 \leq j \leq n} (G_{1,j}^2 - 1)| \right)^{1/2} + (\mathbf{E} |\bar{G}_1^2|)^{1/2}. \end{aligned}$$

Let estimate the summands. One has:

$$(\mathbf{E} |\bar{G}_1^2|)^{1/2} = n^{-1/4} (\mathbf{E} |\eta_1^2|)^{1/2} = n^{-1/4}.$$

Also, it is easy to see that

$$\begin{aligned} &\left(\mathbf{E} |n^{-1} \sum_{1 \leq j \leq n} (G_{1,j}^2 - 1)| \right)^{1/2} \\ &= n^{-1/4} \left(\mathbf{E} \left| \sum_{1 \leq j \leq n} \frac{(G_{1,j}^2 - 1)}{\sqrt{n}} \right| \right)^{1/2} \\ &\leq n^{-1/4} \left(\mathbf{E} \left(\sum_{1 \leq j \leq n} \frac{(G_{1,j}^2 - 1)}{\sqrt{n}} \right)^2 \right)^{1/4} \leq c n^{-1/4}. \end{aligned}$$

Consequently,

$$v_1 \leq c \left(\sum_{i \geq 1} \left| \frac{q_i}{q_1} \right| \right)^{1/2} n^{-1/4}.$$

Let estimate the second summand in (15):

$$v_2 = \sup_x \left| \mathbf{P} \left(\sum_{i \geq 1} \left(q_i (\eta_i^2 - 1) - \frac{1}{n} q_i \eta_i^2 \right) \leq x \right) - \mathbf{P} (q_i (\eta_i^2 - 1) \leq x) \right|.$$

Denote

$$\theta = -\frac{1}{n} q_1 \eta_1^2; \quad T = \sum_{i \geq 2} q_i (\eta_i^2 - 1); \quad Y = -\frac{1}{n} \sum_{i \geq 2} q_i \eta_i^2.$$

Let $F_1(u) = \mathbf{P}(q_1(\eta_1^2 - 1) + \theta \leq x)$, $F(u) = \mathbf{P}(q_1(\eta_1^2 - 1) \leq x)$, then

$$v_2 = \sup_x |\mathbf{E} F_1(x - T - Y) - \mathbf{E} F(x - T)|.$$

Using triangle inequality, one finds:

$$v_2 \leq \sup_x |\mathbf{E} F_1(x - T - Y) - \mathbf{E} F_1(x - T)| + \sup_x |\mathbf{E} F_1(x - T) - \mathbf{E} F(x - T)|. \quad (16)$$

One has:

$$\begin{aligned} \sup_x |\mathbf{E} F_1(x - T - Y) - \mathbf{E} F_1(x - T)| &\leq c |q_1|^{-1/2} \sqrt{\mathbf{E} Y} \\ &\leq c |q_1|^{-1/2} (\mathbf{E} Y^2)^{1/4}. \end{aligned}$$

It is easy to see that

$$\mathbf{E} (Y^2)^{1/4} \leq c n^{-1/2} \left(\sum_{i \geq 2} q_i^2 \right)^{1/4}.$$

For the second summand in (16), let apply Lemma 2:

$$\sup_x |\mathbf{E} F_1(x - T) - \mathbf{E} F(x - T)| \leq \frac{c}{\sqrt{q_1 n}}.$$

Finally, one has:

$$v_2 \leq \frac{c}{\sqrt{q_1 n}} + c n^{-1/2} \left(\sum_{i \geq 2} \left(\frac{q_i}{q_1} \right)^2 \right)^{1/4}.$$

Consequently,

$$\rho(Z(G_1, \dots, G_n), Z_0) \leq \frac{c}{n^{1/4}} \left(\sum_{i \geq 1} \left| \frac{q_i}{q_1} \right| \right)^{1/2} + \frac{c}{\sqrt{q_1 n}}.$$

The theorem is proved.

References

1. Bentkus V., Götze F. Optimal bounds in non-Gaussian limit theorems for U-statistics // Ann. Prob., 1999. Vol. 27. No. 1. P. 454–521.
2. Korolyuk V. S., Borovskikh Yu. V. Rate of convergence of degenerate von Mises functionals // Theory Prob. Appl., 1988. Vol. 33. No. 1. P. 125–135.
3. Senatov V. V. Qualitative effects in the estimates of the convergence rate in the central limit theorem in multidimensional spaces // Proceedings of the Steklov Inst. of Math., 1996. Vol. 215. No. 4.
4. Yanushkevichiene O. Asymptotic rate of convergence in the degenerate U-statistics of second order // Banach Center Pubs., 2010. Vol. 90. P. 275–284.
5. Yanushkevichiene O. Optimal rates of convergence of second degree polynomials in several metrics // J. Math. Sci., 2006. Vol. 138, No. 1. P. 5472–5479.
6. Cramer H. Mathematical methods of statistics. — Stockholm, 1946.

О СКОРОСТИ СХОДИМОСТИ НЕКОТОРОЙ U-СТАТИСТИКИ

О. Л. Янушкявичене¹, Р. Янушкявичюс²

¹Институт математики и информатики Вильнюсского университета, Литва, olgjan@zebra.lt

²Литовский университет эдукологии, Вильнюс, Литва

Аннотация: Анализируется асимптотическое поведение U-статистик, которые важны для конструирования статистических методов, применяемых для изучения информационных систем.

Ключевые слова: скорость сходимости; неравенство Берри–Эссеена; квадратичная форма, U-статистика второго порядка; метрика Колмогорова

ОЦЕНКИ СКОРОСТИ СХОДИМОСТИ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ СУММ К ДИСПЕРСИОННЫМ ГАММА-РАСПРЕДЕЛЕНИЯМ*

В. Е. Бенинг¹, Л. М. Закс², В. Ю. Королев³

Аннотация: Строятся оценки точности приближения распределений отрицательных биномиальных случайных сумм независимых случайных величин с ненулевыми средними несимметричными дисперсионными гамма-распределениями.

Ключевые слова: случайная сумма; отрицательное биномиальное распределение; смешанное пуассоновское распределение; гамма-распределение; оценка скорости сходимости; дисперсионное гамма-распределение

1 Введение

Дисперсионные гамма-распределения являются традиционными математическими моделями статистических закономерностей, используемыми для описания эффекта наличия тяжелых или полутяжелых хвостов. Такие модели очень важны для адекватного описания статистических закономерностей поведения различных характеристик открытых сложных систем, эволюция которых в значительной мере зависит от информационных потоков, к примеру телекоммуникационных сетей или финансовых рынков. В частности, в финансовой математике хорошо известны так называемые *дисперсионные гамма-процессы* (variance gamma processes или VG-processes). В развитие идей теоретического моделирования таких эффектов с помощью предельных теорем для обобщенных дважды стохастических пуассоновских процессов — в определенном смысле наилучших моделей хаотических процессов со случайной интенсивностью, — изложенных в работах [1–3], в данной статье приводятся оценки скорости сходимости распределений специальных смешанных пуассоновских случайных сумм независимых случайных величин с ненулевыми средними к несимметричным дисперсионным гамма-распределениям (variance gamma-distributions). Как показано в работе [3], такие распределения возникают как асимптотические аппроксимации для аддитивных характеристик информационных потоков в том случае, когда случайная интенсивность

соответствующего потока информативных событий имеет гамма-распределение. Таким образом, приводимые в данной статье оценки могут быть полезны при определении адекватности смешанных вероятностных моделей статистических закономерностей, в которых смешивающим является гамма-распределение.

Дисперсионные гамма-процессы, предложенные в работах [4, 5], — это процессы Леви (случайные процессы с независимыми стационарными приращениями), одномерные распределения которых являются дисперсионными гамма-распределениями. Плотность дисперсионного гамма-распределения имеет вид:

$$f_{VG}(x; \alpha, \sigma, \mu, \lambda) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\infty} \exp \left\{ -\frac{1}{2} \left(\frac{x - \alpha u}{\sigma \sqrt{u}} \right)^2 \right\} \frac{g(u; \mu, \lambda)}{\sqrt{u}} du, \quad (1)$$

где $\alpha \in \mathbb{R}$, $\sigma > 0$, $\mu > 0$, $\lambda > 0$, а $g(x; \mu, \lambda)$ — плотность гамма-распределения с параметрами μ и λ ,

$$g(x; \mu, \lambda) = \frac{\lambda^\mu}{\Gamma(\mu)} x^{\mu-1} e^{-\lambda x}, \quad x \geq 0. \quad (2)$$

Здесь и далее $\Gamma(\cdot)$ — эйлерова гамма-функция,

$$\Gamma(z) = \int_0^{\infty} e^{-y} y^{z-1} dy, \quad z > 0.$$

* Работа поддержана РФФИ, гранты 12-07-00109а, 12-07-00115а.

¹ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики РАН, bening@cs.msu.ru

² Альфа-банк, отдел моделирования и математической статистики, lily.zaks@gmail.com

³ Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; Институт проблем информатики РАН, vkorolev@cs.msu.ru

В дальнейшем функцию гамма-распределения, соответствующую плотности $g(x; \mu, \lambda)$ (см. (2)), будем обозначать $G(x; \mu, \lambda)$, а функцию дисперсионного гамма-распределения, соответствующую плотности $f_{VG}(x; \alpha, \sigma, \mu, \lambda)$ (см. (1)), будем обозначать $F_{VG}(x; \alpha, \sigma, \mu, \lambda)$.

Как отмечено в упомянутых работах, подобные модели также демонстрируют высокую адекватность при описании динамики цен финансовых активов.

Вместе с тем в прикладной теории вероятностей хорошо известен принцип, согласно которому та или иная модель может считаться в достаточной мере обоснованной только тогда, когда она является *асимптотической аппроксимацией*, т.е. когда существует довольно простая предельная теорема, в которой рассматриваемая модель выступает в качестве предельного распределения [6]. В книге [7] прослежена глубокая связь этого принципа с универсальным принципом неубывания энтропии в замкнутых системах. Обе рассматриваемые в данной статье модели имеют вид сдвиг-масштабных смесей нормальных законов. Как известно, нормальное распределение обладает максимальной энтропией среди всех распределений, носителем которых является вся числовая прямая, и имеющих конечный второй момент. Если бы моделируемая сложная система была информационно изолирована от окружающей среды, то в соответствии с принципом неубывания энтропии, который в теории вероятностей проявляется в виде предельных теорем [7], наблюдаемые статистические распределения ее характеристик были бы неотличимы от нормального. Но поскольку любая математическая модель по своему определению не может учесть все факторы, влияющие на состояние или эволюцию моделируемой системы, то параметры этого нормального закона изменяются в зависимости от состояния среды, внешней по отношению к моделируемой системе. Другими словами, эти параметры являются случайными и изменяются под влиянием информационных потоков между системой и внешней средой. Таким образом, во многих ситуациях разумные модели статистических закономерностей изменения параметров сложных систем должны иметь вид сдвиг-масштабных смесей нормальных законов, частным случаем которых является (1).

В первоисточниках упомянутые выше модели вводились чисто умозрительно как распределения процесса броуновского движения со случайным временем, в каждый момент имеющим гамма-распределение. «Асимптотическое» обоснование моделей типа дисперсионных гамма-распределений было дано лишь недавно в работе [3], где показано, что такие распределения могут выступать в качестве

предельных в довольно простых предельных теоремах для регулярных статистик, построенных по выборкам случайного объема, в частности, в схеме случайного суммирования случайных величин и, следовательно, могут считаться *естественными* асимптотическими аппроксимациями для распределений многих процессов, например, сходных с неоднородными случайными блужданиями.

Изучению условий и скорости сходимости распределений специальных случайных сумм к частному случаю дисперсионных гамма-распределений — несимметричному распределению Лапласа посвящен раздел 12.7.7 в книге [8]. Свойства самого несимметричного распределения Лапласа и его практическое применение описаны в работе [9].

Пусть $\{X_{n,j}\}_{j \geq 1}$, $n = 1, 2, \dots$ — последовательность серий независимых и одинаково в каждой серии распределенных случайных величин, а N_n , $n = 1, 2, \dots$ — положительные целочисленные случайные величины такие, что при каждом n случайная величина N_n независима от последовательности $\{X_{n,j}\}_{j \geq 1}$. Для натуральных k обозначим

$$S_{n,k} = X_{n,1} + \dots + X_{n,k}.$$

Для определенности будем считать, что $\sum_{j=1}^0 = 0$ и все функции распределения, о которых пойдет речь ниже, непрерывны справа. Символ \implies будет обозначать сходимость по распределению. Пусть $\Phi(x)$ — стандартная нормальная функция распределения:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz.$$

В работе [3] доказана следующая теорема.

Теорема А. *Предположим, что существуют числа $\alpha \in \mathbb{R}$, $\sigma^2 \in (0, \infty)$, $\mu \in (0, \infty)$, и последовательность натуральных чисел $\{m_n\}_{n \geq 1}$ такие, что при $n \rightarrow \infty$*

$$P(S_{n,m_n} < x) \longrightarrow \Phi\left(\frac{x - \alpha}{\sigma}\right); \quad (3)$$

$$P(N_n < m_n x) \longrightarrow G(x; \mu, \lambda). \quad (4)$$

Тогда

$$P(S_{n,N_n} < x) \longrightarrow F_{VG}(x; \alpha, \sigma, \mu, \lambda),$$

причем предельная функция дисперсионного гамма-распределения может быть представлена как свертка функций распределения $G(x; \mu, \lambda_1)$ и $1 - G(-x; \mu, \lambda_2)$, где

$$\lambda_1 = \left(\frac{1}{2} \sqrt{\frac{\alpha^2}{\lambda^2} + \frac{2\sigma^2}{\lambda}} - \frac{\alpha}{2\lambda} \right)^{-1};$$

$$\lambda_2 = \left(\frac{1}{2} \sqrt{\frac{\alpha^2}{\lambda^2} + \frac{2\sigma^2}{\lambda}} + \frac{\alpha}{2\lambda} \right)^{-1}.$$

В данной статье будет рассмотрена скорость сходимости в теореме А для одного довольно иллюстративного частного случая, демонстрирующего один из возможных механизмов формирования несимметричных предельных законов для случайных блужданий.

2 Основной результат

Пусть ξ_1, ξ_2, \dots — независимые одинаково распределенные случайные величины с $E\xi_1 = 0$, $0 < D\xi_1 = \sigma^2 < \infty$, $\beta^3 = E|\xi_1|^3 < \infty$, $a \in \mathbb{R}$, n — натуральное число. Положим

$$X_{n,j} = \frac{\xi_j}{\sqrt{n}} + \frac{a}{n}.$$

В терминах случайных блужданий рассматриваемая конструкция слагаемых предполагает *одинаковый порядок малости* элементарных трендов и *дисперсий*, что характерно, например, для приращений винеровского процесса со сносом. Обозначим

$$S_n = \sum_{j=1}^n X_{n,j} \left(= \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_j + a \right).$$

В силу классической центральной предельной теоремы имеем

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} \left| P(S_n < x) - \Phi \left(\frac{x - a}{\sigma} \right) \right| = 0,$$

т.е. так определенные случайные величины $X_{n,j}$ удовлетворяют условию (3) с $m_n = n$.

Пусть $V_{\mu,\lambda}$ — случайная величина, имеющая гамма-распределение $G(x; \mu, \lambda)$, независимая от стандартного пуассоновского процесса $M(t)$, $t \geq 0$. Для каждого натурального n положим

$$N_n = M(nV_{\mu,\lambda}).$$

Очевидно, что так определенная случайная величина N_n имеет смешанное пуассоновское распределение:

$$\begin{aligned} P(N_n = k) &= P(M(nV_{\mu,\lambda}) = k) = \\ &= \int_0^\infty e^{-nz} \frac{(nz)^k}{k!} g(z; \mu, \lambda) dz, \quad k = 0, 1, \dots \end{aligned}$$

Более того, несложно проверить, что это распределение является отрицательным биномиальным (распределением Паскаля) с параметрами μ и $p_n = \lambda/(n + \lambda)$:

$$\begin{aligned} P(M(nV_{\mu,\lambda}) = k) &= \\ &= \int_0^\infty e^{-nz} \frac{(nz)^k}{k!} \frac{\lambda^\mu}{\Gamma(\mu)} e^{-\lambda z} z^{\mu-1} dz = \\ &= \frac{n^k \lambda^\mu \Gamma(k + \mu)}{k! \Gamma(\mu) (\lambda + n)^{k+\mu}} = \\ &= \frac{\Gamma(k + \mu)}{k! \Gamma(\mu)} \left(\frac{\lambda}{\lambda + n} \right)^\mu \left(1 - \frac{\lambda}{\lambda + n} \right)^k, \\ & \quad k = 0, 1, \dots \quad (5) \end{aligned}$$

Подобные случайные величины хорошо описывают число несчастных случаев в некоторой большой группе, отдельные индивидуумы которой в разной степени подвержены риску [10], число страховых выплат по неоднородным страховым портфелям [8], число сделок на больших биржах, на которых интенсивность торгов неоднородна в силу разных интересов разных групп игроков в зависимости от времени [11].

Обозначим $A_n(z) = A_n(z; \mu, \lambda) = P(N_n < nz)$, $z \geq 0$ ($A_n(z) = 0$ при $z < 0$). Несложно видеть, что

$$A_n(x) \Rightarrow G(x; \mu, \lambda) \quad (n \rightarrow \infty).$$

Действительно, как известно, если $\Pi(x; \ell)$ — функция распределения Пуассона с параметром $\ell > 0$ и $E(x; c)$ — функция распределения с единственным единичным скачком в точке $c \in \mathbb{R}$, то

$$\Pi(\ell x; \ell) \Rightarrow E(x; 1) \quad (\ell \rightarrow \infty).$$

Так как для $x \in \mathbb{R}$

$$A_n(x) = \int_0^\infty \Pi(nx; nz) dG(z; \mu, \lambda),$$

то по теореме Лебега о мажорируемой сходимости при $n \rightarrow \infty$

$$\begin{aligned} A_n(x) &\Rightarrow \int_0^\infty E(x/z; 1) dG(z; \mu, \lambda) = \\ &= \int_0^x dG(z; \mu, \lambda) = G(x; \mu, \lambda), \end{aligned}$$

т.е. так определенные случайные величины N_n удовлетворяют условию (4) с $m_n = n$. Впредь будем считать, что при каждом $n \geq 1$ случайная величина N_n независима от последовательности $\{\xi_j\}_{j \geq 1}$,

что гарантирует независимость случайных величин $N_n, X_{n,1}, X_{n,2}, \dots$.

Таким образом, в силу непрерывности функции распределения $F_{VG}(x; \alpha, \sigma, \mu, \lambda)$ из теоремы А следует, что

$$D_n \equiv \sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\sum_{j=1}^{N_n} X_{n,j} < x \right) - F_{VG}(x; \alpha, \sigma, \mu, \lambda) \right| \longrightarrow 0 \quad (n \rightarrow \infty).$$

Скорость стремления D_n к нулю описывается следующим утверждением.

Теорема 1. Пусть $\mu > 1/2$. Для любого $n \geq 1$ справедлива оценка

$$D_n \leq \frac{0,4532}{\sqrt{n}} \frac{\beta^3}{\sigma^3} \frac{\sqrt{\lambda} \Gamma(\mu - 1/2)}{\Gamma(\mu)} + \frac{0,1210}{n} \frac{\alpha^2}{\sigma^2}.$$

Доказательство. Как уже было показано, распределение случайной величины N_n является смешанным пуассоновским (см. (5)). Следовательно, по теореме Фубини

$$\begin{aligned} \mathbb{P} \left(\sum_{j=1}^{N_n} X_{n,j} < x \right) &= \mathbb{P} \left(\sum_{j=1}^{M(nU_{\mu,\lambda})} X_{n,j} < x \right) = \\ &= \int_0^\infty \mathbb{P} \left(\sum_{j=1}^{M(nz)} X_{n,j} < x \right) g(z; \mu, \lambda) dz. \end{aligned}$$

При этом

$$\begin{aligned} \mathbb{E} X_{n,j} &= \frac{\alpha}{n}; \quad \mathbb{D} X_{n,j} = \frac{\sigma^2}{n}; \\ \mathbb{E} |X_{n,j} - \mathbb{E} X_{n,j}|^3 &= \frac{\beta^3}{n^{3/2}}. \end{aligned}$$

Таким образом, при каждом $z \in (0, \infty)$

$$\begin{aligned} \mathbb{E} \sum_{j=1}^{M(nz)} X_{n,j} &= \alpha z; \\ \mathbb{D} \sum_{j=1}^{M(nz)} X_{n,j} &= nz \left(\frac{\alpha^2}{n^2} + \frac{\sigma^2}{n} \right) = z\sigma^2 \left(1 + \frac{\alpha^2}{n\sigma^2} \right). \end{aligned}$$

Согласно определению дисперсионного гамма-распределения,

$$F_{VG}(x; \alpha, \sigma, \mu, \lambda) = \int_0^\infty \Phi \left(\frac{x - \alpha z}{\sigma \sqrt{z}} \right) g(z; \mu, \lambda) dz.$$

Поэтому

$$\begin{aligned} D_n &= \sup_x \left| \int_0^\infty g(z; \mu, \lambda) \left[\mathbb{P} \left(\sum_{j=1}^{M(nz)} X_{n,j} < x \right) - \right. \right. \\ &\quad \left. \left. - \Phi \left(\frac{x - \alpha z}{\sigma \sqrt{z(1 + \alpha^2/(n\sigma^2))}} \right) + \right. \right. \\ &\quad \left. \left. + \Phi \left(\frac{x - \alpha z}{\sigma \sqrt{z(1 + \alpha^2/(n\sigma^2))}} \right) - \Phi \left(\frac{x - \alpha z}{\sigma \sqrt{z}} \right) \right] dz \right| \leq \\ &\leq I_1 + I_2, \quad (6) \end{aligned}$$

где

$$\begin{aligned} I_1 &= \int_0^\infty g(z; \mu, \lambda) \sup_x \left| \mathbb{P} \left(\sum_{j=1}^{M(nz)} X_{n,j} < x \right) - \right. \\ &\quad \left. - \Phi \left(\frac{x - \alpha z}{\sigma \sqrt{z(1 + \alpha^2/(n\sigma^2))}} \right) \right| dz; \\ I_2 &= \int_0^\infty g(z; \mu, \lambda) \sup_x \left| \Phi \left(\frac{x - \alpha z}{\sigma \sqrt{z(1 + \alpha^2/(n\sigma^2))}} \right) - \right. \\ &\quad \left. - \Phi \left(\frac{x - \alpha z}{\sigma \sqrt{z}} \right) \right| dz. \end{aligned}$$

В дальнейшем понадобится следующее утверждение.

Лемма 1. Пусть случайные величины X_1, X_2, \dots одинаково распределены. Пусть N_λ — пуассоновская случайная величина с параметром $\lambda > 0$. Предположим, что случайные величины $N_\lambda, X_1, X_2, \dots$ независимы в совокупности. Обозначим

$$S_\lambda = X_1 + \dots + X_{N_\lambda}.$$

Тогда

$$\begin{aligned} \sup_x \left| \mathbb{P}(S_\lambda < x) - \Phi \left(\frac{x - \mathbb{E} S_\lambda}{\sqrt{\mathbb{D} S_\lambda}} \right) \right| &\leq \\ &\leq \frac{0,4532}{\sqrt{\lambda}} \frac{\mathbb{E} |X_1 - \mathbb{E} X_1|^3}{(\mathbb{D} X_1)^{3/2}}. \end{aligned}$$

Доказательство леммы 1 приведено в работе [12] (см. также [8, с. 144]).

Продолжим доказательство теоремы 1. Рассмотрим I_1 . Применяя лемму 1 с учетом (5), получаем:

$$\begin{aligned} I_1 &\leq \frac{0,4532}{\sqrt{n}} \frac{\beta^3}{\sigma^3} \frac{\lambda^\mu}{\Gamma(\mu)} \int_0^\infty e^{-\lambda z} z^{\mu-3/2} dz = \\ &= \frac{0,4532}{\sqrt{n}} \frac{\beta^3}{\sigma^3} \frac{\lambda^\mu \Gamma(\mu - 1/2)}{\lambda^{\mu-1/2} \Gamma(\mu)} = \\ &= \frac{0,4532}{\sqrt{n}} \frac{\beta^3}{\sigma^3} \frac{\sqrt{\lambda} \Gamma(\mu - 1/2)}{\Gamma(\mu)}. \quad (7) \end{aligned}$$

Рассмотрим I_2 . В дальнейшем понадобится еще одно вспомогательное утверждение.

Лемма 2. Пусть $b \in \mathbb{R}$, $0 < c < \infty$, $0 < d < \infty$. Тогда

$$\sup_y |\Phi(y) - \Phi(cy)| \leq \frac{1}{\sqrt{2\pi e}} \left| \max \left\{ c, \frac{1}{c} \right\} - 1 \right|; \quad (8)$$

$$\sqrt{1+d} - 1 \leq \frac{d}{2}. \quad (9)$$

Элементарное доказательство этого утверждения основано на применении формулы Лагранжа.

Продолжим доказательство теоремы 1. В лемме 2 положим

$$y = \frac{x - \alpha z}{\sigma \sqrt{z} \sqrt{1 + \alpha^2/(n\sigma^2)}}; \quad c = \sqrt{1 + \frac{\alpha^2}{n\sigma^2}}.$$

Тогда $c \geq 1$ и в силу утверждения (8) леммы 2 имеем:

$$I_2 \leq \frac{1}{\sqrt{2\pi e}} \left(\sqrt{1 + \frac{\alpha^2}{n\sigma^2}} - 1 \right).$$

При этом в силу утверждения (9) леммы 2

$$\sqrt{1 + \frac{\alpha^2}{n\sigma^2}} - 1 \leq \frac{\alpha^2}{2n\sigma^2}.$$

Окончательно получаем:

$$I_2 \leq \frac{\alpha^2}{2\sqrt{2\pi e} n \sigma^2}. \quad (10)$$

Подставляя (7) и (10) в (6), получаем утверждение теоремы. Теорема доказана.

Если $\alpha = 0$, то оценку, представленную в теореме 1, можно уточнить не только за счет того, что в таком случае второе слагаемое в правой части обнуляется, но и за счет уменьшения коэффициента при первом слагаемом. Аналогом леммы 1 в таком случае может служить следующее утверждение.

Лемма 3. Пусть случайные величины X_1, X_2, \dots одинаково распределены. Пусть N_ℓ — пуассоновская случайная величина с параметром $\ell > 0$. Предположим, что случайные величины N_ℓ, X_1, X_2, \dots независимы в совокупности. Обозначим

$$S_\ell = X_1 + \dots + X_{N_\ell}.$$

Тогда

$$\sup_x \left| P(S_\ell < x) - \Phi \left(\frac{x - ES_\ell}{\sqrt{DS_\ell}} \right) \right| \leq \frac{0,3041}{\sqrt{\ell}} \frac{E|X_1|^3}{(EX_1^2)^{3/2}}.$$

Доказательство этого утверждения приведено в [13], также см. [8], теорема 2.4.3.

Если в случае $\alpha = 0$ в доказательстве теоремы 1 вместо леммы 1 воспользоваться леммой 3, то в результате получится следующее утверждение.

Следствие 1. Пусть в дополнение к условиям теоремы 1 $\alpha = 0$. Тогда для любого $n \geq 1$ справедлива оценка

$$D_n \leq \frac{0,3041}{\sqrt{n}} \frac{\beta^3}{\sigma^3} \frac{\sqrt{\lambda} \Gamma(\mu - 1/2)}{\Gamma(\mu)}.$$

Это утверждение обобщает следствие 12.2.2 из [8] на случай произвольных параметров μ и λ .

Литература

1. Королев В. Ю., Соколов И. А. Математические модели неоднородных потоков экстремальных событий. — М.: Торус Пресс, 2008.
2. Королев В. Ю., Шоргин С. Я. Математические методы анализа стохастической структуры информационных потоков. — М.: ИПИ РАН, 2011.
3. Королев В. Ю., Соколов И. А. Скошенные распределения Стюдента, дисперсионные гамма-распределения и их обобщения как асимптотические аппроксимации // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 2–10.
4. Madan D. B., Seneta E. The variance gamma (V.G.) model for share market return // J. Business, 1990. Vol. 63. P. 511–524.
5. 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.
6. Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. — М.—Л.: ГИТТЛ, 1949.
7. Gnedenko B. V., Korolev V. Yu. Random summation: Limit theorems and applications. — Boca Raton: CRC Press, 1996.
8. Королев В. Ю., Бенинг В. Е., Шоргин С. Я. Математические основы теории риска. — 2-е изд., перераб. и доп. — М.: Физматлит, 2011. 620 с.
9. Kotz S., Kozubowski T. J., Podgorski K. The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering and finance. — Boston: Birkhauser, 2001.
10. Greenwood M., Yule G. U. An inquiry into the nature of frequency-distributions of multiple happenings, etc. // J. Royal Statist. Soc., 1920. Vol. 83. P. 255–279.
11. Королев В. Ю. Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — М.: Изд-во Московского ун-та, 2011. 510 с.
12. Королев В. Ю., Шевцова И. Г., Шоргин С. Я. О неравенствах типа Берри—Эсееена для пуассоновских случайных сумм // Информатика и её применения, 2011. Т. 5. Вып. 3. С. 64–66.
13. Korolev V., Shevtsova I. An improvement of the Berry—Esseen inequality with applications to Poisson and mixed Poisson random sums // Scandinavian Actuarial J., June 04, 2010. DOI:10.1080/03461238.2010.485370.

ЗАДАЧИ ОПТИМАЛЬНОГО ПЛАНИРОВАНИЯ МЕЖУРОВНЕВОГО ИНТЕРФЕЙСА В БЕСПРОВОДНЫХ СЕТЯХ*

Ю. В. Гайдамака¹, Т. В. Ефимушкина², А. К. Самуйлов³, К. Е. Самуйлов⁴

Аннотация: В данном обзоре проведено исследование современного состояния задач оптимального планирования межуровневого интерфейса на базе механизма мультиплексирования с ортогональным частотным разделением (OFDM, Orthogonal Frequency Division Multiplexing) для нисходящего канала в сетевой технологии LTE (Long-Term Evolution). При этом рассматривается понятие межуровневой оптимизации, подробно описаны оптимизационные задачи и ограничения, возникающие при разделении радиоресурсов в нисходящем канале, дан краткий обзор планировщиков и соответствующих им функций полезности, определяющих уровень удовлетворенности пользователей схемой распределения радиоресурсов при заданных ограничениях.

Ключевые слова: технология OFDM; межуровневая оптимизация; функция полезности; планировщик; эффективное распределение частот

1 Введение

К основным задачам в беспроводных сетях относится оптимизация распределения ограниченного числа радиоресурсов между пользователями. Различные типы пакетного трафика, передаваемого по сети, предполагают динамическое выделение ресурсов пользователям. Решением задач планирования ресурсов, назначения приоритетов доступа в зависимости от типа трафика с заданными требованиями к качеству обслуживания (QoS, Quality of Service) занимаются модули управления радиоресурсами, называемые планировщиками (англ. *schedulers*).

Динамичное изменение загруженности канала в беспроводной сети определяет требования к планировщику, одним из которых является межуровневый (англ. *crosslayer*) подход к решению задачи оптимального распределения ресурсов. Основным принципом межуровневой оптимизации является комплексное решение задачи эффективного использования ограниченного числа радиоресурсов, учитывающее ряд первостепенных факторов: повышение пропускной способности; обеспечение равнодоступности — справедливого (англ. *fair*) разделения ресурсов между пользователями; достижение требуемого или, по крайней мере, наилучшего возможного качества обслуживания [1].

В обзоре в общем виде сформулированы основные задачи оптимизации, возникающие при планировании ресурсов в беспроводных сетях в целях повышения эффективности работы сети с большим числом несущих, сети, построенной на базе механизма OFDM, характерного для нисходящего канала в технологии LTE [2, 3]. Вводится понятие межуровневой оптимизации, подробно описаны оптимизационные задачи и ограничения, возникающие при разделении ресурсов в нисходящем канале. Исследованы два алгоритма межуровневой оптимизации, предназначенных для максимизации функции полезности в различных условиях — алгоритм динамического назначения поднесущих DSA (Dynamic Subcarrier Assignment) и алгоритм адаптивного распределения мощности APA (Adaptive Power Allocation). Для рассматриваемых алгоритмов сформулированы задачи максимизации функции полезности и получены их решения.

2 Виды межуровневой оптимизации

С точки зрения терминологии межуровневая оптимизация заключается в объединении нескольких уровней модели взаимодействия открытых систем (OSI, Open Systems Interconnection) для получения более качественных решений и эффективных

* Работа выполнена при частичной поддержке РФФИ (гранты 10-07-00487-а и 12-07-00108) и Рособразования (проект 020619-1-174).

¹Российский университет дружбы народов, кафедра систем телекоммуникаций, ygaidamaka@sci.pfu.edu.ru

²Российский университет дружбы народов, кафедра систем телекоммуникаций, tefimushkina@gmail.com

³Российский университет дружбы народов, кафедра систем телекоммуникаций, asam1988@gmail.com

⁴Российский университет дружбы народов, кафедра систем телекоммуникаций, ksam@sci.pfu.edu.ru

алгоритмов без лишних межуровневых обменов информации. Межуровневый подход к решению задачи оптимального распределения ресурсов позволяет в динамическом режиме учесть изменения типов трафика в беспроводной сети, потребности в услугах, значений параметров в канале связи и мобильность абонентов.

Выделим три основных вида межуровневой оптимизации. Главной задачей канально-ориентированного вида межуровневой оптимизации является эффективное использование ограниченного числа изменяющихся во времени радиоресурсов с целью обеспечения высокой пропускной способности, заданных требований к качеству и равнодоступности. С точки зрения модели OSI данная задача рассматривается между физическим и канальными уровнями.

Второй вид, ориентированный на качество предоставления услуг абоненту «из конца в конец», решает задачи адаптации протоколов верхних уровней к нестабильным, изменяющимся во времени канальным ресурсам для достижения заданных требований к QoS параметрам, например к производительности и задержкам. Заметим, что эффективность функционирования протокола TCP в беспроводных сетях связи — одна из типичных задач данного вида оптимизации [4–6].

Выбор наилучшего маршрута определяет третий вид межуровневой оптимизации, рассматриваемый в [7–9]. При этом поиск наиболее эффективного маршрута происходит с учетом сетевого и физического или канального уровней. Далее в статье рассматриваются задачи оптимизации только с точки зрения канально-ориентированного межуровневого подхода.

3 Задачи оптимизации

3.1 Задача минимизации мощности

В [10] сформулирована задача минимизации общей выделяемой пользователям сети мощности, учитывающая распределение поднесущих с определением числа бит и уровня выделяемой мощности для каждой из поднесущих на основе мгновенных (*instantaneous*) характеристик состояния канала, измеренных для каждого из пользователей сети. В рамках данной задачи предложен и реализован итерационный алгоритм распределения поднесущих между пользователями, а также обобщенный алгоритм задания числа бит и уровня мощности для поднесущих, назначенных пользователям. В [10] рассматривается сеть связи с K пользователями, в которой k -пользователь имеет

скорость передачи, равную R_k бит на OFDM-символ (далее бит/символ), $k = \overline{1, K}$. На передатчике реализован алгоритм назначения n -поднесущей k -пользователю, после применения которого на основе характеристик состояния канала для k -пользователя применяется обобщенный алгоритм задания $c_{k,n}$ числа бит/символ для n -поднесущей (здесь и далее $n = \overline{1, N}$).

В зависимости от числа назначенных $c_{k,n}$ бит из множества $\mathcal{D} = \{0, 1, 2, \dots, M\}$, где M — максимально возможное для передачи число бит/символ, адаптивный модулятор выбирает соответствующую схему модуляции, при этом уровень выделяемой мощности адаптируется согласно обобщенному алгоритму. Заметим, что n -поднесущая предоставляется только одному пользователю, т. е. при $c_{k',n} \neq 0$, $c_{k,n} = 0$ для всех $k \neq k'$.

В частотно-селективном канале с замираниями n -поднесущая характеризуется уровнем мощности сигнала $a_{k,n}$ по отношению к k -пользователю. При этом дисперсия уровня спектральной плотности шума $\sigma_{k,n}^2$ принята равной единице для всех поднесущих. Для поддержания требуемого качества услуги на приемнике выделяемая мощность для передачи k -пользователю на n -поднесущей рассчитывается по формуле

$$P_{k,n} = \frac{f_k(c_{k,n})}{a_{k,n}^2},$$

где $f_k(c_{k,n})$ — требуемая мощность для приема данных и их последующей демодуляции. Таким образом, задача минимизации мощности представляется в виде

$$P^* = \min_{c_{k,n} \in \mathcal{D}} \sum_{n=1}^N \sum_{k=1}^K \frac{f_k(c_{k,n})}{a_{k,n}^2}$$

с ограничением для k -пользователя по числу бит для передачи $R_k = \sum_{n=1}^N c_{k,n}$.

Исследованный в [10–12] алгоритм задания числа бит и уровня мощности в сети с одним пользователем служит основой для решения задачи минимизации мощности для случая многопользовательской сети. Этот алгоритм относится к так называемым «жадным» алгоритмам и назначает бит поднесущей, требующей выделения наименьшей дополнительной мощности. Процесс назначения происходит по одному биту за один раз до тех пор, пока R бит не будут распределены между N поднесущими.

Решение оптимизационной задачи для случая многопользовательской сети предусматривает использование действительных значений для числа

бит/символ $c_{k,n} \in \mathbb{R}[0, M]$, а также введение функции назначения k -пользователю n -поднесущей, $\rho_{k,n} \in \mathbb{R}[0, 1]$. Тогда задача минимизации мощности принимает вид:

$$p^* = \min_{c_{k,n}, \rho_{k,n}} \sum_{n=1}^N \sum_{k=1}^K \frac{\rho_{k,n} f_k(c_{k,n})}{a_{k,n}^2}$$

с ограничениями $\sum_{n=1}^N \rho_{k,n} c_{k,n} = R_k$ и $\sum_{k=1}^K \rho_{k,n} = 1$.

Данное предположение позволяет решить задачу назначения поднесущей k -пользователю методом множителей Лагранжа по алгоритму, приведенному в [10]. В рамках решения задачи оптимизации полученные величины определяют нижнюю границу искомого минимального значения выделяемой мощности. Однако из-за принятого ранее предположения о величинах $c_{k,n}^* \notin \mathbb{Z}$ и $\rho_{k,n}^* \in \mathbb{R}[0, 1]$ предполагается разделение поднесущей между несколькими пользователями. Решение данной проблемы методом квантования полученных величин может не удовлетворять требованию k -пользователя к скорости передачи R_k . В предположении, что $\rho_{k',n}^* = 1$, $k' = \arg \max_k \rho_{k,n}^*$ и $\rho_{k,n}^* = 0, k \neq k'$, алгоритм назначения n -поднесущей в [10] дополняется алгоритмом задания числа бит и уровня мощности в сети с одним пользователем. В результате в [10] предложен многопользовательский адаптивный алгоритм (MAO, Multiuser Adaptive OFDM).

Данный подход к решению задачи оптимизации, согласно [13], относят к методу релаксаций. Использование нецелого числа бит и разделения поднесущей между пользователями позволяет эффективно решать задачу оптимизации, однако требует применения дополнительных процедур для получения целых величин, являющихся целесообразными с точки зрения функционирования сети. Двумя другими методами, предложенными в [13], являются разбиение задачи на несколько более простых и эвристический алгоритм. Первый метод предполагает определение числа поднесущих для k -пользователя с учетом требований к скорости передачи R_k и далее назначение конкретных, выбранных по некоторому алгоритму, поднесущих. Эвристический подход основывается на методе сортировки и представляет собой реализацию двухэтапного аналитического метода, описанного выше. Решение задачи оптимизации с помощью эвристического метода представлено также в [14].

3.2 Задача максимизации скорости передачи

В [15] нелинейная оптимизационная задача преобразована в линейную задачу максимизации ско-

рости передачи путем равномерного разделения общей мощности p_{tot} между пользователями в сети для каждой из поднесущих:

$$p_{k,n} = \frac{p_{tot}}{N}.$$

Задача максимизации общей пропускной способности сети представлена в виде:

$$R^* = \max_{c_{k,n} \in \mathcal{D}} \sum_{n=1}^N \sum_{k=1}^K c_{k,n} \rho_{k,n},$$

принимая во внимание требование r_k , предъявляемое k -пользователем к минимальной скорости передачи числа бит на один OFDM-символ,

$$\sum_{n=1}^N c_{k,n} \rho_{k,n} \geq r_k.$$

Фиксируя уровень выделяемой мощности $p_{k,n}$ и допуская, что значения заданного для k -пользователя коэффициента ошибок BER (Bit Error Rate) и состояния канала $a_{k,n}$ известны на базовой станции для всех пользователей, находятся значения числа бит/символ

$$c_{k,n} = f(\text{BER}, p_{k,n}, a_{k,n}).$$

Данный подход позволяет эффективно решить линейную задачу оптимизации методом целочисленного линейного программирования, однако предусматривает экспоненциальный рост уровня сложности с увеличением числа поднесущих и пользователей в сети.

В [15] предложен алгоритм понижения сложности, состоящий из двух этапов: назначение поднесущих пользователям с наибольшим возможным числом бит для передачи и перераспределение поднесущих для соблюдения требований r_k . Поднесущие на первом этапе распределяются между пользователями с целью достижения максимальной пропускной способности без учета требований r_k к минимальной скорости передачи. Для перераспределения пользователей на втором этапе требуется соблюдение следующих условий:

- (1) выделенная на первом этапе k_n^* -пользователю n -поднесущая не может быть переназначена другим пользователям, если переназначение повлечет возможное нарушение требования r_k к минимальной скорости передачи k_n^* -пользователя, $R_{k_n^*,n} - c_{k_n^*,n} < r_{k_n^*}$;
- (2) каждое переназначение поднесущих должно минимально сокращать общую пропускную способность системы;
- (3) число переназначений должно быть наименьшим.

Для выполнения последних двух условий вводится функция $e_{k,n} = (c_{k,n}^* - c_{k,n})/c_{k,n}$ оценки переназначения n -поднесущей k -пользователю. Согласно алгоритму [15] перераспределение происходит поочередно для всех пользователей в сети, которым назначены поднесущие после первого этапа, не удовлетворяющие требованиям по скорости передачи. При этом для k -пользователя выбирается поднесущая с наименьшей функцией оценки переназначения. Перераспределение k -пользователя на n' -поднесущую происходит, если $R_{k,n',n'} - c_{k,n'}^* \geq r_{k,n'}$. В противном случае выбирается другая поднесущая с минимальной функцией перераспределения.

3.3 Задача обеспечения равнодоступности

В [16] приведены три из наиболее известных схем распределения поднесущих между пользователями.

Согласно первой из них обеспечение максимальной пропускной способности (maxBR, maximum bit-rate) достигается за счет предоставления n -поднесущей k -пользователю, находящемуся в лучших канальных условиях, т.е. обладающему наибольшим частотным откликом (англ. *frequency response*) $H_{k,n}$. Следует отметить, что данный метод не решает задачу обеспечения равнодоступности. Однако, рассматривая величину частотного отклика в качестве единственного параметра при распределении поднесущих, метод maxBR определяет верхнюю границу возможной скорости передачи данных.

Вторая схема распределения канальных ресурсов, изложенная в [17], предполагает решение задачи обеспечения равнодоступности путем предоставления одинаковой скорости передачи всем пользователям. Помимо ограничения по мощности передачи $\sum_{k=1}^K \sum_{n=1}^N p_{k,n} \leq p_{\text{tot}}$ в [17] также вводится пропорциональное ограничение:

$$N_1 : N_2 : \dots : N_K = \phi_1 : \phi_2 : \dots : \phi_K,$$

где $N_k = \phi_k N$ — число поднесущих, назначенных k -пользователю, и ϕ_k — нормированная пропорциональная постоянная скорости передачи данных для k -пользователя. При этом $\tilde{n} = N/K$ определяет максимально возможное число выделяемых пользователю поднесущих.

Третья схема назначения поднесущих, предложенная в [18], заключается в выборе \tilde{n} поднесущих с наибольшими значениями частотных откликов для пользователя. Данная процедура повторяется для всех пользователей в сети.

Введем величину отношения сигнал-шум (SNR, Signal-to-Noise Ratio), используемую для постановки оптимизационной задачи:

$$\text{SNR}_{k,n} = \frac{|H_{k,n}|^2 p_{k,n}}{\sigma_{k,n}^2}.$$

Задача максимизации общей пропускной способности всех пользователей в сети формулируется в виде:

$$F^* = \max_{\text{SNR}_{k,n}} \sum_{k=1}^K \sum_{n=1}^N f(\text{SNR}_{k,n}).$$

Следует отметить, что подобные задачи не учитывают улучшения пропускной способности отдельных пользователей. В [16] приводится решение данной оптимизационной задачи с учетом обеспечения равнодоступности путем назначения равного числа поднесущих по одному из двух алгоритмов, кратко охарактеризованных ниже, а далее путем использования обобщенного алгоритма задания числа бит/символ для каждой поднесущей.

Первый алгоритм заключается в сравнении и выборе поднесущей с наибольшим частотным откликом, а также ее назначении пользователю. В ходе данного назначения пользователь, получивший оптимальное число поднесущих, удаляется из рассматриваемого множества. Данная процедура продолжается для всех оставшихся пользователей и поднесущих.

Второй алгоритм заключается в поиске для первого пользователя максимального частотного отклика и назначении ему соответствующей поднесущей. После выделения по одной поднесущей каждому из K пользователей данный алгоритм повторяется в противоположном порядке: от K -го до 1-го пользователя.

3.4 Задача максимизации полезности

Следует отметить, что все рассмотренные выше задачи подразумевают оптимизацию некоторой функции полезности, описывающей тот или иной уровень удовлетворенности пользователей для определенной схемы распределения радиоресурсов при некоторых ограничениях. Тем не менее, как будет показано ниже, оптимизация полезности может оказаться задачей, представляющей самостоятельный интерес.

В [19, 20] предложены два алгоритма межуровневой оптимизации, предназначенные для максимизации функции полезности (Utility Function) в различных условиях — алгоритм динамического назначения поднесущих DSA (Dynamic Subcarrier Assignment) и алгоритм адаптивного распределения мощности APA, а также комбинация этих

алгоритмов. Эффективность алгоритмов оценивалась с помощью имитационного моделирования, при котором они сравнивались с алгоритмом фиксированного назначения поднесущих FSA (Fixed Subcarrier Assignment). Ниже для алгоритмов DSA и APA сформулированы две задачи нелинейного целочисленного программирования, для которых получены условия оптимальности. Вводятся следующие обозначения: \mathcal{N} — множество поднесущих $\{1, \dots, N\}$; \mathcal{K} — множество пользователей $\{1, \dots, K\}$; β — коэффициент побитовой ошибки (BER); $\mathbf{p} = (p[1], \dots, p[N])$ — вектор мощностей поднесущих; ρ — состояние поднесущей (отношение сигнал—шум); Δf — ширина полосы пропускания поднесущей; $c_k^p[n]$ — достижимая эффективность передачи (бит/символ); r_k — скорость передачи k -пользователя, $k \in \mathcal{K}$.

Предполагается, что нисходящий канал базовой станции ячейки сети OFDM используется всеми пользователями, причем базовой станции известно состояние назначенной пользователю поднесущей. Достижимая скорость передачи данных зависит от отношения сигнал—шум и мощности передачи:

$$c_k^p[n] = f(\log_2(1 + \beta p[n] \rho_k[n])),$$

где функция $f(\cdot)$ зависит от выбранной схемы адаптации скорости. Например, если применять непрерывную адаптацию скорости, то $f(x) = x$ и $c_k^p[n] = \log_2(1 + \beta p[n] \rho_k[n])$.

Введем $x_{kn} \in \{0, 1\}$ состояние n -поднесущей так, что $x_{kn} = 1$, если n -поднесущая назначена k -пользователю, и $x_{kn} = 0$ в противном случае. Тогда $x_k = (x_{kn})_{n \in \mathcal{N}}$ — вектор состояния поднесущих для k -пользователя, причем условие $\sum_{k \in \mathcal{K}} x_{kn} = 1$ означает, что n -поднесущая может быть назначена только одному пользователю. Множество $D_k(\mathbf{x}_k) = \{n : x_{kn} = 1\}$ включает все поднесущие, назначенные k -пользователю в состоянии \mathbf{x}_k , а набор множеств $D(\mathbf{x}) = (D_k(\mathbf{x}_k))_{k \in \mathcal{K}}$ определяет распределение поднесущих по пользователям, когда система находится в состоянии $\mathbf{x} = (\mathbf{x}_k)_{k \in \mathcal{K}}$. Тогда множество состояний системы можно определить в виде:

$$\mathcal{X} = \left\{ \mathbf{x} = (\mathbf{x}_k)_{k \in \mathcal{K}} : D_i(\mathbf{x}_i) \cap D_j(\mathbf{x}_j) = \emptyset, \right. \\ \left. i \neq j \in \mathcal{K}, \bigcup_{k \in \mathcal{K}} D_k(\mathbf{x}_k) \subseteq \mathcal{N} \right\}.$$

Введем множество всех возможных наборов поднесущих $\mathcal{D} = \{D(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ и множество возможных вариантов распределения мощностей

$$\mathcal{P} = \{\mathbf{p} : 0 \leq p(n) \leq P, \sum_{n \in \mathcal{N}} p(n) = P, n \in \mathcal{N}\}.$$

Скорость передачи данных r_k [бит/с] для k -пользователя в состоянии \mathbf{x}_k представима в виде:

$$r_k := r_k(\mathbf{x}_k, \mathbf{p}) = \sum_{n \in \mathcal{N}} c_k^p(n) \Delta f x_{kn} = \\ = \sum_{n \in D_k(\mathbf{x}_k)} c_k^p(n) \Delta f = r_k(D_k(\mathbf{x}_k), \mathbf{p}) =: r_k(D_k, \mathbf{p}), \\ k \in \mathcal{K}.$$

Пусть $U_k(\cdot)$ — функция полезности для k -пользователя, $k \in \mathcal{K}$. Будем рассматривать в качестве основного блага для пользователя величину скорости передачи данных r_k и определим функцию полезностей в следующем виде:

$$U(r(d, \mathbf{p})) := \sum_{k \in \mathcal{K}} U_k(r_k(D_k, \mathbf{p})).$$

Таким образом, задача межуровневой оптимизации в общем случае может быть сформулирована как максимизация функции полезности для ячейки сети OFDM в виде:

$$\max_{d, \mathbf{p}} \sum_{k \in \mathcal{K}} U_k(r_k(d, \mathbf{p}))$$

с ограничениями $d \in \mathcal{D}$ и $\mathbf{p} \in \mathcal{P}$.

Пусть вектор \mathbf{p} распределения мощностей фиксирован, т. е. $\mathbf{p} = \tilde{\mathbf{p}}$. Тогда функцию полезности для алгоритма динамического назначения поднесущих DSA можно определить в виде:

$$U(r(d)) := U(r(d, \tilde{\mathbf{p}})).$$

В случае алгоритма адаптивного распределения мощностей APA фиксированным является набор d множеств поднесущих $d = \tilde{d}$, и, следовательно, функция полезности имеет вид:

$$U(r(\mathbf{p})) := U(r(\tilde{d}, \mathbf{p})).$$

Отметим, что данная задача относится к классу задач целочисленного нелинейного программирования. В [19] для решения этой задачи используется метод релаксаций, идея которого заключается в том, что при разработке метода решения задач отбрасывается требование к целочисленности переменных. Предполагается, что функция полезности $U_k(r_k)$ для k -пользователя является неубывающей выпуклой и существует ее производная $U'_k(r_k)$.

Алгоритм динамического назначения поднесущих DSA

С учетом введенных обозначений задача максимизации полезности для алгоритма DSA записывается в виде $\max_{d \in \mathcal{D}} U(r(d))$.

Используя метод математической индукции можно доказать, что максимум функции полезности $U(r(d))$ достигается в сформулированных ниже достаточных условиях оптимальности.

Утверждение 1. Если для набора $d^* = (D_k^*)_{k \in \mathcal{K}}$ выполняется условие:

$$U'_k(r_k^*)c_k^{\tilde{p}}(n) \geq U'_j(r_j^*)c_j^{\tilde{p}}(n), \quad k \neq j \in \mathcal{K}, \quad n \in D_k^*,$$

где $r_k^* = \sum_{n \in D_k^*} c_n^{\tilde{p}}(n) \Delta f$, тогда функция полезности $U(r(d))$ достигает глобального максимума на наборе $d = d^* \in \mathcal{D}$.

Из полученных условий оптимальности получаем правило назначения поднесущей пользователю. Для заданного вектора $\mathbf{p} = \tilde{\mathbf{p}}$ распределения мощности номер пользователя, которому назначается n -поднесущая, определяется формулой:

$$k(n) = \arg \max_{k \in \mathcal{K}} \left\{ U'_k(r_k^*)c_k^{\tilde{p}}(n) \right\}.$$

Алгоритм адаптивного распределения мощности АРА

Алгоритм адаптивного распределения мощности АРА состоит в назначении каждой поднесущей $n \in \mathcal{N}$ определенной мощности передачи при условии фиксированного набора $d = \tilde{d}$ распределения поднесущих между пользователями.

Задача межуровневой оптимизации для алгоритма АРА может быть сформулирована в виде $\max_{\mathbf{p} \in \mathcal{P}} U(r(\mathbf{p}))$. Ниже сформулированы необходимые условия достижения максимума функцией полезности $U(r(\mathbf{p}))$.

Утверждение 2. Если $p^*(n)$ является решением задачи $\max_{\mathbf{p} \in \mathcal{P}} U(r(\mathbf{p}))$, тогда

$$p^*(n) = \left[\frac{U'_k(r_k^*)}{\lambda} - \frac{1}{\beta \rho_k(n)} \right]^*, \quad k \in \mathcal{K}, \quad n \in \tilde{D}_k,$$

где $\lambda > 0$ — нормирующая константа оптимального распределения мощностей.

Из полученных условий оптимальности функции $U(r(\mathbf{p}))$ очевидным образом следует правило назначения мощностей поднесущих для алгоритма АРА.

4 Заключение

Планирование межуровневого интерфейса является наиболее эффективным подходом к согласованию возможностей современных беспроводных технологий и возрастающих требований по

обслуживанию больших объемов трафика пользователей с заданным качеством.

В обзоре технические и алгоритмические проблемы создания планировщиков межуровневого интерфейса иллюстрированы постановками оптимизационных задач, возникающих при распределении ресурсов в сетях с большим числом несущих радиочастот. Рассмотрены наиболее известные задачи и ограничения, характерные для технологии LTE, даны краткие комментарии по их решению, алгоритмам поиска оптимального решения и условиям оптимальности. Приведен типичный пример задачи оптимизации функции полезности как наиболее общей задачи оптимального планирования межуровневого интерфейса. Исследовано и сформулировано достаточное условие нахождения глобального максимума функции полезности для задачи DSA, а также необходимое условие для задачи АРА.

Литература

1. Shariat M., Qudus A. U., Ghorashi S. A., Tafazolli R. Scheduling as an important cross-layer operation for emerging broadband wireless systems // IEEE Commun. Surveys Tuts., 2009. Vol. 11. No. 2. P. 74–86.
2. Вишневецкий В. М., Ляхов А. И., Портной С. Л., Шахнович И. В. Широкополосные беспроводные сети передачи информации. — М.: Техносфера, 2005. 597 с.
3. Тихвинский В. О., Терентьев С. В., Юрчук А. Б. Сети мобильной связи LTE: технология и архитектура. — М.: Эко-Трендз, 2010. 284 с.
4. Wu G., Bai Y., Lai J., Ogielski A. Interaction between TCP and RLP in wireless Internet // IEEE Global Communication Conference Proceedings, 1999. Vol. 1b. P. 661–666.
5. Kim B. J. A network service providing wireless channel information for adaptive mobile applications: Part I: Proposals // IEEE Conference (International) on Communications Proceedings, 2001. Vol. 5. P. 1345–1351.
6. Sudame P., Badrinath B. R. On providing support for protocol adaptation in mobile networks // Mobile Networks Applications, 2001. Vol. 6. No. 1. P. 43–55.
7. Chiang M. To layer or not to layer: Balancing transport and physical layers in wireless multihop networks // IEEE J. Selected Areas Commun., 2005. Vol. 23. No. 1. P. 104–116.
8. Kawadia V., Kumar P. R. A cautionary perspective on cross-layer design // IEEE Wireless Commun., 2005. Vol. 12. No. 1. P. 3–11.
9. Iannone L., Fdida S. Evaluating a cross-layer approach for routing in wireless mesh networks // Telecommunication Systems J. (Springer) Special issue: Next Generation Networks — Architectures, Protocols, Performance, 2006. Vol. 31. No. 2–3. P. 173–193.
10. Wong C. Y., Cheng R. S., Letaief K. B. Multiuser OFDM with adaptive subcarrier, bit, and power allocation //

- IEEE J. Selected Areas Commun., 1999. Vol. 17. No. 10. P. 1747–1757.
11. *Hughes-Hartogs D.* Ensemble modem structure for imperfect transmission media. U.S. Patents No. 4679227, July 1987; No. 4731816, March 1988; No. 4833796, May 1989.
 12. *Lai S. K., Cheng R. S., Letaief K. B., Murch R. D.* Adaptive trellis coded MQAM and power optimization for OFDM transmission // IEEE Trans. Commun., 1999. Vol. 47. P. 538–545.
 13. *Bohge M., Gross J., Wolisz A., Meyer M.* Dynamic resource allocation in OFDM Systems: an overview of cross-layer optimization principles and techniques // IEEE Networks, 2007. Vol. 21. No. 1. P. 53–59.
 14. *Kivanc D., Li G., Liu H.* Computationally efficient bandwidth allocation and power control for OFDMA // IEEE Trans. Wireless Commun., 2003. Vol. 2. No. 6. P. 1150–1158.
 15. *Zhang Y. J., Letaief K. B.* Multiuser adaptive subcarrier and bit allocation with adaptive cell selection for OFDM systems // IEEE Trans. Wireless Commun., 2004. Vol. 3. No. 5. P. 1566–1575.
 16. *Otani Y., Ohno S., Teo K., Teo D., Hinamoto T.* Subcarrier allocation for multi-user OFDM system // Asia-Pacific Communication Conference Proceedings, 2005. P. 1073–1077.
 17. *Wong C., Shen Z., Evans L., Andrews J. G.* A low complexity algorithm for proportional resource allocation in OFDMA systems // IEEE Workshop on Signal Processing Systems Proceedings. — Texas, USA, 2004. P. 1–6.
 18. *Fu J., Karasawa Y.* Fundamental analysis on throughput characteristics of orthogonal frequency division multiple access OFDMA in multipath propagation environments // IEICE Trans., 2002. Vol. J85-B. No. 11. P. 1884–1894.
 19. *Song G., Li Ye.* Cross-layer optimization for OFDM wireless networks — Part I: Theoretical framework // IEEE Trans. Wireless Commun., 2005. Vol. 4. No. 2. P. 614–624.
 20. *Song G., Li Y.* Cross-layer optimization for OFDM wireless networks — Part II: Algorithm development // IEEE Trans. Wireless Commun., 2005. Vol. 4. No. 2. P. 625–634.

АСИМПТОТИКА МАКСИМУМА ПРОЦЕССА НАГРУЗКИ ДЛЯ НЕКОТОРОГО КЛАССА ГАУССОВСКИХ ОЧЕРЕДЕЙ*

О. В. Лукашенко¹, Е. В. Морозов²

Аннотация: Изучается асимптотическое поведение максимума процесса нагрузки в жидкостной системе обслуживания, на вход которой поступает процесс, содержащий случайную компоненту, описываемую центрированным гауссовским процессом. Предполагается, что дисперсия этого процесса является регулярно меняющейся на бесконечности функцией с показателем $V \in (0, 2)$. К такому классу процессов, в частности, относится сумма независимых дробных броуновских движений (ДБД). Показано, что при соответствующей нормировке максимум процесса нагрузки на интервале $[0, t]$ сходится по вероятности при $t \rightarrow \infty$ к некоторой явно выписанной константе.

Ключевые слова: гауссовская система обслуживания; максимум процесса нагрузки; дробное броуновское движение; асимптотический анализ; правильное изменение

1 Введение

В последнее время значительно возрос интерес исследователей к анализу гауссовских жидкостных моделей телекоммуникационных систем. В таких моделях входной поток, задающий величину поступившей в систему работы, является гауссовским процессом (далее — гауссовский входной процесс). Основная причина этого интереса состоит в том, что, как было выяснено рядом исследователей, гауссовские процессы позволяют учесть при моделировании современных телекоммуникационных систем такие важные характеристики сетевого трафика, как самоподобие (инвариантность по времени) и долговременную зависимость (долгую память) [1, 2]. Наличие таких свойств существенно затрудняет вероятностный анализ и, как правило, не позволяет получить в явном виде ключевые характеристики системы, такие как вероятность переполнения буфера, вероятность потери сообщения и т. д. Эти характеристики критически важны для определения качества обслуживания (QoS), обеспечиваемого данной системой. С другой стороны, гауссовские процессы достаточно хорошо изучены, и это обстоятельство позволяет в ряде случаев осуществлять, по крайней мере, асимптотический анализ систем с гауссовским входным процессом.

Наиболее важным (входным) гауссовским процессом в телекоммуникационных системах является ДБД, обладающее самоподобием и долгой па-

мятью. Важность этого процесса обусловлена, в частности, тем, что ДБД возникает при суперпозиции большого числа независимых так называемых on/off-источников с тяжелыми хвостами на больших масштабах времени.

Известно, что в системе с бесконечным буфером и входным процессом ДБД стационарный процесс загрузки Q^* (текущая незавершенная работа в системе) распределен как максимум гауссовского процесса с отрицательным линейным сносом на положительной полуоси [3].

Отсутствие точных аналитических результатов для Q^* вызывает необходимость исследования асимптотик соответствующих характеристик системы. Такие результаты получены, например, для вероятности $\mathbb{P}(Q^* > b)$ переполнения буфера размера b [4, 5] либо для логарифма этой вероятности (логарифмические асимптотики) при $b \rightarrow \infty$ [6–8]. В большинстве упомянутых работ в качестве входного процесса рассматривается ДБД.

Наряду с вероятностью переполнения другой важной характеристикой систем обслуживания является максимум процесса загрузки на конечном интервале $[0, t]$. Для этой характеристики в работах [9, 10] найдены асимптотики (при $t \rightarrow \infty$) в случае единственного входного процесса ДБД. Эти результаты легли в основу представленного в данной статье асимптотического анализа максимума процесса загрузки в более общей модели. Рассматри-

*Работа поддержана РФФИ (проект 10-07-00017). Работа выполнена при поддержке Программы стратегического развития на 2012–2016 гг. «Университетский комплекс ПетрГУ в научно-образовательном пространстве Европейского Севера: стратегия инновационного развития».

¹Институт прикладных математических исследований КарНЦ РАН, Петрозаводский государственный университет, lukashenko-oleg@mail.ru

²Институт прикладных математических исследований КарНЦ РАН, Петрозаводский государственный университет, emorozov@karelia.ru

вается система с входным процессом, содержащим гауссовский процесс со стационарными приращениями, дисперсия которого принадлежит к классу *правильно меняющихся на бесконечности функций*. Частным случаем такого процесса является суперпозиция независимых ДБД. Прежде всего дадим мотивировку такой постановки задачи. Рассмотрим N независимых on/off-источников, причем k -й источник описывается процессом $\{I_k(t), t \geq 0\}$, $k = 1, \dots, N$, где

$$I_k(t) = \begin{cases} 1, & t \in \text{on-период}; \\ 0, & t \in \text{off-период}. \end{cases}$$

Поясним, что on-период означает период непрерывной работы источника, а следующий за ним (независимый) off-период есть время простоя. Таким образом, on/off-периоды каждого источника образуют альтернирующий процесс восстановления. По условию, процессы для разных источников независимы. Суммарная нагрузка (совокупный агрегированный трафик), поступившая в систему от всех источников на интервале $[0, t]$, равна

$$A_N(t) := \int_0^t \left(\sum_{k=1}^N I_k(u) \right) du,$$

т. е. это суммарное время работы (активности) всех N источников на интервале $[0, t]$. Предположим, что имеется n типов источников, среди которых N_i источников типа $i = 1, \dots, n$. Предположим также, что хвосты функций распределения on/off-периодов источника i -го типа имеют следующую асимптотику при $x \rightarrow \infty$:

$$\left. \begin{aligned} 1 - F_{\text{on}}^i(x) &\sim \ell_{\text{on}}^i x^{-\alpha_{\text{on}}^i} L_{\text{on}}^i(x), \\ 1 - F_{\text{off}}^i(x) &\sim \ell_{\text{off}}^i x^{-\alpha_{\text{off}}^i} L_{\text{off}}^i(x), \end{aligned} \right\} \quad (1)$$

где $\ell_{\text{on}}^i, \ell_{\text{off}}^i$ — положительные константы, показатели $\alpha_{\text{on}}^i, \alpha_{\text{off}}^i \in (1, 2)$, а функции $L_{\text{on}}^i, L_{\text{off}}^i$ медленно меняются на бесконечности, т. е.

$$\lim_{x \rightarrow \infty} \frac{L^i(tx)}{L^i(x)} = 1, \quad i = 1, \dots, n,$$

для любого фиксированного $t > 0$. (Условия (1) означают, что функции распределения F_{on}^i и F_{off}^i имеют *тяжелые хвосты*.) Обозначим через μ_{on}^i и μ_{off}^i математическое ожидание on- и off-периода соответственно для источника i . (Заметим, что эти величины конечны, поскольку $\alpha_{\text{on}}^i, \alpha_{\text{off}}^i > 1$.) В работе [11] доказана функциональная предельная теорема, согласно которой распределение агрегированного трафика $\{A_N(tT), t \geq 0\}$

с ростом сначала величин N_i , а затем параметра T (этот порядок важен) сближается с распределением процесса $T \left(\sum_{i=1}^n N_i \mu_{\text{on}}^i / (\mu_{\text{on}}^i + \mu_{\text{off}}^i) \right) t + \sum_{i=1}^n T^{H_i} \sqrt{L_i(T) N_i c_i} B_{H_i}(t)$, где c_i — положительные константы, L_i — медленно меняющиеся на бесконечности функции, выраженные через исходные параметры, а B_{H_i} — независимые ДБД с параметрами Херста H_i , определяемыми как

$$H_i = \frac{3 - \min(\alpha_{\text{on}}^i, \alpha_{\text{off}}^i)}{2} \in \left(\frac{1}{2}, 1 \right), \quad i = 1, \dots, n.$$

Таким образом, суммарный трафик, порожденный большим числом (независимых) источников, у которых распределения on/off-периодов имеют тяжелые хвосты, приближенно описывается процессом, включающим сумму независимых ДБД. Данный результат обосновывает большой интерес к моделям телекоммуникационных систем, на вход которых поступает один или несколько независимых ДБД.

2 Гауссовские очереди

Вначале дадим описание гауссовской системы обслуживания в целом, а затем описание новой модели, изучаемой в данной статье. Пусть $A(t)$ — суммарная работа, поступившая в систему за время $[0, t]$. При анализе гауссовских очередей входной процесс обычно задается в следующем виде:

$$A(t) = mt + \sigma X(t), \quad (2)$$

где m, σ — положительные константы, а $X := \{X(t), t \geq 0\}$ — центрированный гауссовский процесс со стационарными приращениями, $X(0) = 0$ [12]. Подчеркнем, что такая форма входного процесса является общепринятой при описании гауссовской системы обслуживания, однако более полное описание случайной компоненты X зависит от специфики модели. Будем считать, что система имеет одно обслуживающее устройство с постоянной скоростью обслуживания C , причем $r := C - m > 0$. Обозначим также $W(t) = \sigma X(t) - rt$ и пусть $Q(t)$ — величина нагрузки (незавершенная работа в системе) в момент времени t . Если $Q(0) = 0$, то для $Q(t)$ справедливо выражение [3]:

$$\begin{aligned} Q(t) &= \sup_{0 \leq s \leq t} (A(t) - A(s) - C(t - s)) = \\ &= \sup_{0 \leq s \leq t} (\sigma(X(t) - X(s)) - r(t - s)) = \\ &= \sup_{0 \leq s \leq t} (W(t) - W(s)). \end{aligned} \quad (3)$$

Заметим, что $EA(1) = m$, поэтому условие $r > 0$ обеспечивает существование стационарного про-

цесса нагрузки, который определяется следующим образом [12]:

$$Q = \sup_{t \geq 0} (A(t) - Ct) = \sup_{t \geq 0} (\sigma X(t) - rt) = \sup_{t \geq 0} W(t). \quad (4)$$

Хорошо известно [12] (и легко установить), что ковариационная функция $\Gamma(s, t)$ процесса X имеет вид

$$\Gamma(s, t) = E[X(s)X(t)] = \frac{1}{2}[v(t) + v(s) - v(|t - s|)],$$

где $v(t)$ есть дисперсия $X(t)$.

Основное предположение, принятое в данной статье, состоит в том, что функция $v(t)$ *правильно меняется на бесконечности с индексом* $0 < V < 2$, т. е. для любого $y > 0$

$$\lim_{t \rightarrow \infty} \frac{v(yt)}{v(t)} = y^V.$$

Известно, что любая правильно меняющаяся на бесконечности функция с индексом V может быть представлена в виде

$$v(t) = t^V L(t),$$

где $L(t)$ — медленно меняющаяся на бесконечности функция [13]. Обозначим $\beta = 1/(2 - V)$, а также выберем и зафиксируем любое $\varepsilon \in (0, 2 - V)$. Будем далее считать, что функция $L(t)$ является *дважды дифференцируемой* на \mathbb{R}_+ . (Вообще говоря, достаточно, чтобы это условие было выполнено на сколь угодно удаленном от начала координат луче $[a, \infty)$, $a > 0$.) Кроме того, предположим, что также выполнены следующие условия (при $t \rightarrow \infty$):

$$L(tL^\beta(t)) \sim L(t), \quad (5)$$

$$L''(t) = o\left(\frac{1}{t^{V+\varepsilon}}\right). \quad (6)$$

Из условия (6) следует (с использованием правила Лопиталя), что $L'(t) = o(t^{-V-\varepsilon+1})$. Отметим, что класс медленно меняющихся функций, удовлетворяющих условиям (5) и (6), достаточно обширен и, в частности, включает функции, имеющие (на бесконечности) конечный ненулевой предел, а также функции вида $(\ln t)^a$, $(\ln \ln t)^a$, $a \geq 0$ и т. д.

Из (4) следует, что вероятность превышения процессом стационарной нагрузки некоторого уровня $b > 0$ определяется следующим образом:

$$\mathbb{P}(Q > b) = \mathbb{P}\left(\sup_{t \geq 0} W(t) > b\right).$$

В работе [8] показано, что для центрированного гауссовского процесса со стационарными приращениями и с дисперсией $v(t)$, правильно меняющейся на бесконечности с индексом $0 < V < 2$, справедлива такая (логарифмическая) асимптотика:

$$\lim_{b \rightarrow \infty} \frac{v(b)}{b^2} \ln \mathbb{P}(Q > b) = -\theta, \quad (7)$$

где параметр $\theta > 0$ имеет вид

$$\theta = \frac{2}{\sigma^2(2 - V)^{2-V}} \left(\frac{r}{V}\right)^V. \quad (8)$$

Кроме того, в [14] показано, что на одном вероятностном пространстве можно задать процесс $W(t) = \sigma X(t) - rt$ и стационарный процесс $Q^* := \{Q^*(t), t \geq 0\}$ таким образом, что одновременно выполнены условия

$$Q^*(t) =_d Q \text{ для всех } t \geq 0, \quad (9)$$

$$Q^*(t) = W(t) + \max\{Q^*(0), L^*(t)\}, \quad t \geq 0, \quad (10)$$

где $=_d$ означает равенство по распределению, а $L^*(t) = -\min_{0 \leq s \leq t} \{W(s)\}$. Обозначим

$$M(t) = \max_{0 \leq s \leq t} Q(s), \quad M^*(t) = \max_{0 \leq s \leq t} Q^*(s).$$

Таким образом, $M^*(t)$ есть максимум стационарного процесса нагрузки Q^* (удовлетворяющего условиям (9) и (10)), а $M(t)$ — максимум исходного (нестационарного) процесса нагрузки (3) на интервале $[0, t]$. Далее будем изучать асимптотическое (при $t \rightarrow \infty$) поведение этих максимумов.

3 Асимптотика максимума процесса нагрузки

В данном разделе сформулирован и доказан основной результат об асимптотическом поведении максимумов $M^*(t)$, $M(t)$ при $t \rightarrow \infty$. Этот результат обобщает работу [9], где процесс $X = B_H$, т. е. является ДБД с параметром Херста $H \in (1/2, 1)$. В частности, для модели из [9] соотношения (11) и (12) ниже выполняются при $V = 2H$ (см. (8)) и $\gamma(t) = \ln t$. Подчеркнем, что доказательство в данной статье в целом следует подходу, использованному в [9].

Для удобства обозначим далее

$$\gamma(t) = L[(\ln t)^\beta] \ln t.$$

Справедлива следующая теорема.

Теорема 3.1. Пусть дисперсия гауссовской компоненты X входного процесса (2) удовлетворяет условиям (5) и (6), а также $r > 0$. Тогда

$$\frac{M^*(t)}{\gamma^\beta(t)} \Rightarrow \left(\frac{1}{\theta}\right)^\beta, \quad t \rightarrow \infty; \quad (11)$$

$$\frac{M(t)}{\gamma^\beta(t)} \Rightarrow \left(\frac{1}{\theta}\right)^\beta, \quad t \rightarrow \infty, \quad (12)$$

где параметр θ удовлетворяет соотношению (8), а знак \Rightarrow означает сходимость по вероятности.

Доказательство. Для доказательства (11) достаточно показать, что для любого $\delta > 0$ выполняются следующие два соотношения:

$$\mathbb{P}\left(\frac{M^*(t)}{\gamma^\beta(t)} \geq \left(\frac{1-\delta}{\theta}\right)^\beta\right) \rightarrow 1, \quad t \rightarrow \infty; \quad (13)$$

$$\mathbb{P}\left(\frac{M^*(t)}{\gamma^\beta(t)} \geq \left(\frac{1+\delta}{\theta}\right)^\beta\right) \rightarrow 0, \quad t \rightarrow \infty. \quad (14)$$

Действительно,

$$\frac{(1 \pm \delta)^\beta}{\theta^\beta} = \frac{1}{\theta^\beta} \pm \frac{\beta\delta}{\theta^\beta} + o(\delta), \quad \delta \rightarrow 0.$$

Поскольку $\epsilon := \beta\delta\theta^{-\beta} + o(\delta) \rightarrow 0$ при $\delta \rightarrow 0$, то (13) и (14) означают, что для любого $\epsilon > 0$ выполнены соответственно условия

$$\begin{aligned} \mathbb{P}\left(\frac{M^*(t)}{\gamma^\beta(t)} - \left(\frac{1}{\theta}\right)^\beta > -\epsilon\right) &\rightarrow 1, \quad t \rightarrow \infty; \\ \mathbb{P}\left(\frac{M^*(t)}{\gamma^\beta(t)} - \left(\frac{1}{\theta}\right)^\beta > \epsilon\right) &\rightarrow 0, \quad t \rightarrow \infty, \end{aligned}$$

которые вместе эквивалентны (11). Докажем вначале соотношение (13). Для этого возьмем некоторое (пока произвольное) $\Delta \in (0, t)$. (Ниже величина Δ будет выбрана специальным образом, зависящим от t .) В силу свойства (10) имеем

$$\begin{aligned} Q^*(t) &= W(t) + \max\{Q^*(0), L^*(t)\} \geq \\ &\geq W(t) - \inf_{0 \leq s \leq t} W(s) \geq W(t) - W(t - \Delta). \end{aligned} \quad (15)$$

Это дает следующую цепочку неравенств:

$$\begin{aligned} M^*(t) &= \max_{0 \leq s \leq t} Q^*(s) \geq \max_{k=1, \dots, \lfloor t/\Delta \rfloor} Q^*(k\Delta) \geq \\ &\geq \max_{k=1, \dots, \lfloor t/\Delta \rfloor} [W(k\Delta) - W((k-1)\Delta)] := \\ &:= \max_{1 \leq k \leq \lfloor t/\Delta \rfloor} Y_k^{(\Delta)}, \end{aligned}$$

где $Y_k^{(\Delta)} := W(k\Delta) - W((k-1)\Delta)$, а $\lfloor x \rfloor$ есть наибольшее целое, не превосходящее x . Таким образом, справедливо неравенство

$$\begin{aligned} \mathbb{P}\left(\frac{M^*(t)}{\gamma^\beta(t)} \geq \left(\frac{1-\delta}{\theta}\right)^\beta\right) &\geq \\ &\geq \mathbb{P}\left(\max_{i=1, \dots, \lfloor t/\Delta \rfloor} Y_i^{(\Delta)} \geq \left(\frac{1-\delta}{\theta}\right)^\beta \gamma(t)\right). \end{aligned} \quad (16)$$

Напомним, что $W(t) = \sigma X(t) - rt$. Поэтому

$$\begin{aligned} Y_k^{(\Delta)} &= W(k\Delta) - W((k-1)\Delta) =_d \\ &= \sigma X(\Delta) - r\Delta =_d \sigma \sqrt{v(\Delta)} \mathcal{N}(0, 1) - r\Delta, \end{aligned}$$

где $\mathcal{N}(0, 1)$ — нормальная случайная величина (с. в.). Поскольку

$$Z_i := \frac{Y_i^{(\Delta)} + r\Delta}{\sigma \sqrt{v(\Delta)}} =_d \mathcal{N}(0, 1), \quad (17)$$

то $\{Z_i, i = 1, \dots, \lfloor t/\Delta \rfloor\}$ является стационарной последовательностью нормальных с. в. Кроме того, для автоковариационной функции этой последовательности получаем

$$\begin{aligned} \rho(k) &:= \mathbb{Cov}(Z_1, Z_{1+k}) = \\ &= \frac{1}{\sigma^2 v(\Delta)} \mathbb{Cov}(Y_1^{(\Delta)}, Y_{1+k}^{(\Delta)}) = \\ &= \frac{1}{v(\Delta)} \mathbb{Cov}(X(\Delta), X((k+1)\Delta) - X(k\Delta)) = \\ &= \frac{1}{v(\Delta)} [\Gamma(\Delta, (k+1)\Delta) - \Gamma(\Delta, k\Delta)] = \frac{1}{2v(\Delta)} \times \\ &\times [v((k+1)\Delta) - 2v(k\Delta) + v((k-1)\Delta)]. \end{aligned} \quad (18)$$

Из формулы конечных приращений и (18) следует, что

$$\rho(k) = \frac{\Delta}{2v(\Delta)} v''(u_3)(u_1 - u_2)$$

для некоторых $u_1 \in (k\Delta, (k+1)\Delta)$, $u_2 \in ((k-1)\Delta, k\Delta)$, $u_3 \in (u_2, u_1)$. (Вообще говоря, величины u_i зависят от k .) Легко проверить, что

$$\begin{aligned} v''(k) &= k^V L''(k) + 2Vk^{V-1} L'(k) + \\ &+ V(V-1)k^{V-2} L(k). \end{aligned} \quad (19)$$

Далее, с учетом свойства (6) легко показать, что при $k \rightarrow \infty$

$$k^V \ln k L''(k) \rightarrow 0; \quad k^{V-1} \ln k L'(k) \rightarrow 0. \quad (20)$$

Поскольку $V < 2$, то также

$$k^{V-2} \ln k L(k) \rightarrow 0. \quad (21)$$

Таким образом, из (18)–(21) следует, что

$$\rho(k) \ln k \rightarrow 0, \quad k \rightarrow \infty. \quad (22)$$

Теперь воспользуемся леммой из [15], которая в адаптации к рассматриваемой ситуации примет следующий вид:

Лемма 1. Пусть имеется стационарная последовательность $\{Z_i\}_{i=1}^m$ стандартных нормальных с. в. с ковариационной функцией, удовлетворяющей соотношению (22). Тогда для любой последовательности действительных чисел u_m условие

$$\mathbb{P}\left(\max_{i=1, \dots, m} Z_i \geq u_m\right) \rightarrow 1, \quad m \rightarrow \infty,$$

выполнено тогда и только тогда, когда

$$\lim_{m \rightarrow \infty} m \mathbb{P}(Z > u_m) = \infty. \quad (23)$$

Далее покажем, что можно выбрать $m := m(t)$ и $u_{m(t)} := u(t)$ таким образом, что $m(t) \rightarrow \infty$ при $t \rightarrow \infty$ и условие (23) оказывается выполненным. Прежде всего положим $\Delta := \Delta(t) = A\gamma^\beta(t)$, где $A > 0$ — некоторая постоянная. (Таким образом, величина $\Delta(t)$ растет вместе с t .) Далее будет доказано, что постоянную A можно выбрать так, чтобы обеспечить указанную выше сходимость.

Поскольку любая медленно меняющаяся функция растет медленнее степенной, то существует такое t_0 , что $\Delta(t) < t$ при всех $t \geq t_0$. (Разумеется, t_0 зависит от параметров, определяющих $\Delta(t)$.) Поэтому далее (где это требуется) предполагается, что $t \geq t_0$. Фиксируем произвольное $\delta \in (0, 1)$ и введем обозначения

$$\alpha(t) = \left(\frac{1-\delta}{\theta} \gamma(t)\right)^\beta;$$

$$u(t) = \frac{\alpha(t) + r\Delta(t)}{\sigma\sqrt{v(\Delta(t))}};$$

$$\tau(t) = m(t)\mathbb{P}(Z_1 > u(t)),$$

где $m(t) := \lfloor t/\Delta(t) \rfloor$. Заметим, что ввиду (16) и (17)

$$\begin{aligned} \mathbb{P}\left(\max_{i=1, \dots, m(t)} Y_i^{(\Delta)} \geq \alpha(t)\right) &= \\ &= \mathbb{P}\left(\max_{i=1, \dots, m(t)} Z_i \geq u(t)\right). \end{aligned} \quad (24)$$

Для удобства дальнейшего анализа выпишем связь параметров

$$\beta = \frac{1}{2-V}; \quad \beta - \frac{\beta V}{2} = \frac{1}{2}; \quad \frac{\beta V + 1}{2} = \beta,$$

а также введем обозначение $(\ln t)^\beta = \phi(t)$. Используя свойство (5), можно получить следующие соотношения:

$$\begin{aligned} u(t) &= \frac{\alpha(t) + r\Delta(t)}{\sigma\sqrt{v(\Delta(t))}} = \frac{1}{\sigma} \times \\ &\times \frac{[(1-\delta)/\theta]^\beta + rA \phi(t) [L(\phi(t))]^\beta}{\sqrt{A^V [\phi(t)]^V [L(\phi(t))]^{\beta V} L(A\phi(t) [L(\phi(t))]^\beta)}} = \\ &= \frac{[(1-\delta)/\theta]^\beta + rA \phi(t) [L(\phi(t))]^\beta}{\sigma A^{V/2} [\phi(t)]^{V/2} [L(\phi(t))]^{(\beta V + 1)/2}} (1 + o(1)) = \\ &= C(A) \sqrt{\ln t} (1 + o(1)), \quad t \rightarrow \infty, \end{aligned}$$

где использовано обозначение

$$\begin{aligned} C(A) &= \frac{(1-\delta)^\beta}{\sigma\theta^\beta} A^{-V/2} + \frac{r}{\sigma} A^{1-V/2} := \\ &:= C_1 A^{-V/2} + C_2 A^{1-V/2}, \end{aligned}$$

а также

$$C_1 = \frac{(1-\delta)^\beta}{\sigma\theta^\beta}; \quad C_2 = \frac{r}{\sigma}. \quad (25)$$

Несложно проверить, что справедлива следующая оценка для вероятности большого отклонения стандартной нормальной с. в. (см., например, [16]):

$$\begin{aligned} \mathbb{P}(Z_1 > u(t)) &= \frac{1}{\sqrt{2\pi}} \int_{u(t)}^{\infty} e^{-x^2/2} dx \geq \\ &\geq \frac{1 - u^{-2}(t)}{u(t)\sqrt{2\pi}} e^{-u^2(t)/2} \sim \\ &\sim \frac{1}{u(t)\sqrt{2\pi}} e^{-u^2(t)/2}, \quad t \rightarrow \infty. \end{aligned}$$

Далее с учетом того, что $u(t) = C(A)\sqrt{\ln t} (1 + o(1))$, при достаточно больших t получим:

$$\begin{aligned} \tau(t) &= m(t)\mathbb{P}(Z_1 > u(t)) \geq \\ &\geq m(t) \frac{1 - u^{-2}(t)}{u(t)\sqrt{2\pi}} e^{-u^2(t)/2} \sim \\ &\sim t \left[\sqrt{2\pi} \Delta(t) u(t) e^{u^2(t)/2} \right]^{-1} \sim \\ &\sim t \left[\sqrt{2\pi} A \gamma^\beta(t) C(A) (\ln t)^{1/2} t^{C^2(A)/2 + o(1)} \right]^{-1} \sim \\ &\sim t^{1 - (C^2(A))/2 + o(1)} [L_1(t)]^{-1} := g(t), \end{aligned} \quad (26)$$

где функция

$$L_1(t) := \sqrt{2\pi} A C(A) (\ln t)^{\beta + 1/2} [L(\phi(t))]^\beta$$

медленно меняется на бесконечности. Из (26) следует, что если можно выбрать $C(A) < \sqrt{2}$, то будет иметь место $g(t) \rightarrow \infty, t \rightarrow \infty$. Покажем, что та-

кой выбор действительно возможен. Обозначим $u = V/2$, рассмотрим функцию

$$f(x) = C_1 x^{-u} + C_2 x^{1-u}$$

и заметим, что $f(A) = C(A)$. Легко убедиться, что функция $f(x)$ достигает минимума в точке

$$x_* = \frac{C_1 u}{C_2(1-u)},$$

причем с учетом (25)

$$f(x_*) = \frac{C_1^{1-u} C_2^u}{u^u (1-u)^{1-u}} = \sqrt{(1-\delta)2} < \sqrt{2}.$$

Таким образом, выбор $A = x_*$ обеспечивает требуемую сходимость и ввиду (26)

$$\tau(t) \geq g(t) \rightarrow \infty, \quad t \rightarrow \infty.$$

Поэтому с учетом леммы 1 и соотношений (16), (24) следует утверждение (13):

$$\begin{aligned} \mathbb{P}\left(\frac{M^*(t)}{\gamma^\beta(t)} \geq \left(\frac{1-\delta}{\theta}\right)^\beta\right) &\geq \\ &\geq \mathbb{P}\left(\max_{i=1, \dots, m(t)} Z_i \geq u(t)\right) \rightarrow 1, \quad t \rightarrow \infty. \end{aligned}$$

Теперь обратимся к доказательству соотношения (14). Положим

$$Y_i = \sup_{s \in [i-1, i)} Q^*(s), \quad i = 1, \dots, [t],$$

где $[x]$ — наименьшее целое число, превышающее x . Очевидно, что

$$M^*(t) \leq \max\{Y_i : i = 1, \dots, [t]\}.$$

Ввиду стационарности процесса Q^* с.в. $\{Y_i\}$ одинаково распределены (как некоторая с.в. Y). Далее зафиксируем произвольное $\delta > 0$ и заметим, что

$$\begin{aligned} \mathbb{P}\left(M^*(t) \geq \left(\frac{1+\delta}{\theta}\right)^\beta \gamma^\beta(t)\right) &\leq \\ &\leq \mathbb{P}\left(\max_{i=1, \dots, [t]} Y_i \geq \left(\frac{1+\delta}{\theta}\right)^\beta \gamma^\beta(t)\right) \leq \\ &\leq [t] \mathbb{P}\left(Y \geq \left(\frac{1+\delta}{\theta}\right)^\beta \gamma^\beta(t)\right). \quad (27) \end{aligned}$$

Для последующего анализа необходимо оценить хвост распределения с.в. Y . Для этого заметим, что верны следующие соотношения:

$$\begin{aligned} Y = \max_{s \in [0,1)} \left[W(s) + \right. \\ \left. + \max \left\{ Q^*(0), - \min_{0 \leq \tau \leq s} W(\tau) \right\} \right] = \end{aligned}$$

$$\begin{aligned} &= \max_{s \in [0,1)} \left[\max \left\{ W(s) + Q^*(0), W(s) - \right. \right. \\ &\quad \left. \left. - \min_{0 \leq \tau \leq s} W(\tau) \right\} \right] \leq \\ &\leq \max_{s \in [0,1)} \left[Q^*(0) + W(s) - \min_{0 \leq \tau \leq s} W(\tau) \right] \leq \\ &\leq \max_{s \in [0,1)} \left[Q^*(0) + W(s) - \min_{0 \leq \tau \leq 1} W(\tau) \right] = \\ &= Q^*(0) + \max_{0 \leq s \leq 1} W(s) - \min_{0 \leq s \leq 1} W(s). \quad (28) \end{aligned}$$

Для получения первого неравенства выше был использован тот факт, что для любого a и любых $b, c \geq 0$ имеет место неравенство $\max\{a+b, a+c\} \leq a+b+c$.

Приведем вспомогательную лемму, которая понадобится при уточнении оценки хвоста распределения с.в. Y . (Простое доказательство этой леммы опущено.)

Лемма 2. Для двух неотрицательных с.в. ξ, η и любых чисел $0 < a < b$ справедливо неравенство

$$\mathbb{P}(\xi + \eta \geq b) \leq \mathbb{P}(\xi \geq b) + \mathbb{P}(\eta \geq b - a).$$

Вернемся к доказательству теоремы. Запишем $(1+\delta)^\beta = (1+\delta/2)^\beta + R(\delta)$ и обозначим $R_1 := R(\delta)/(2\theta^\beta) > 0$. Введем также обозначения

$$\begin{aligned} Q_1 &:= \mathbb{P}\left(Q^*(0) \geq \left(\frac{1+\delta/2}{\theta}\right)^\beta \gamma^\beta(t)\right); \\ Q_2 &:= \mathbb{P}\left(\max_{0 \leq s \leq 1} W(s) \geq R_1 \gamma^\beta(t)\right); \\ Q_3 &:= \mathbb{P}\left(-\min_{0 \leq s \leq 1} W(s) \geq R_1 \gamma^\beta(t)\right) \end{aligned}$$

для хвостов функций распределения (случайных) слагаемых в формуле (28). Из неравенства (28) и леммы 2 следует цепочка неравенств

$$\begin{aligned} \mathbb{P}\left(Y \geq \left(\frac{1+\delta}{\theta}\right)^\beta \gamma^\beta(t)\right) &\leq \\ &\leq \mathbb{P}\left(Q^*(0) + \max_{0 \leq s \leq 1} W(s) - \min_{0 \leq s \leq 1} W(s) \geq \right. \\ &\quad \left. \geq \left(\frac{1+\delta}{\theta}\right)^\beta \gamma^\beta(t)\right) \leq Q_1 + Q_2 + Q_3. \quad (29) \end{aligned}$$

В свою очередь, из соотношений (27) и (29) следует неравенство

$$\begin{aligned} 0 \leq \mathbb{P}\left(M^*(t) \geq \left(\frac{1+\delta}{\theta}\right)^\beta \gamma^\beta(t)\right) &\leq \\ &\leq [t](Q_1 + Q_2 + Q_3). \end{aligned}$$

Чтобы получить неравенство (14), докажем, что $[t]Q_i \rightarrow 0, t \rightarrow \infty, i = 1, 2, 3$.

Начнем с анализа Q_2 , для чего заметим, что

$$\max_{0 \leq s \leq 1} W(s) = \max_{0 \leq s \leq 1} (\sigma X(s) - rs) \leq \max_{0 \leq s \leq 1} \sigma X(s). \quad (30)$$

Пусть $a = \mathbb{E} \max_{0 \leq s \leq 1} X(s) < \infty$. Теперь воспользуемся следствием из неравенства Бореля–Судакова–Цирельсона для максимума на конечном интервале центрированного гауссовского процесса со стационарными приращениями [17]:

$$\mathbb{P} \left(\max_{0 \leq s \leq 1} X(s) > x \right) \leq e^{-(1/(2v))(x-a)^2}, \quad x > a, \quad (31)$$

где дисперсия $v := DX(1)$. Введем обозначение $\lambda(t) = R_1 \gamma^\beta(t) \sigma^{-1}$. Из соотношений (30) и (31) следует, что при всех достаточно больших t

$$\begin{aligned} [t]Q_2 &\leq (t+1)\mathbb{P} \left(\max_{0 \leq s \leq 1} X(s) \geq \lambda(t) \right) \leq \\ &\leq (t+1)t^{-(\lambda^2(t)/\ln t) \cdot (1+o(1))/(2v)} := g_1(t). \end{aligned}$$

Известно [13], что $t^\varepsilon L(t) \rightarrow \infty$ при $t \rightarrow \infty$ для любого $\varepsilon > 0$, и поэтому $L(t) > t^{-\varepsilon}$ для всех достаточно больших t . Также заметим, что $2\beta - 1 = V/(2 - V) > 0$ и выберем произвольное $\varepsilon \in (0, (2\beta - 1)/(2\beta^2))$. С учетом этого получаем, что при $t \rightarrow \infty$

$$\begin{aligned} \frac{\lambda^2(t)}{\ln t} &= \left(\frac{R_1}{\sigma} \right)^2 (\ln t)^{2\beta-1} [L((\ln t)^\beta)]^{2\beta} > \\ &> \left(\frac{R_1}{\sigma} \right)^2 [\ln t]^{2\beta-1-2\beta^2\varepsilon} \rightarrow \infty. \end{aligned}$$

Тогда легко увидеть, что $g_1(t) \rightarrow 0$ и, значит, $[t]Q_2 \rightarrow 0, t \rightarrow \infty$.

Теперь рассмотрим слагаемое Q_3 и заметим, что

$$\begin{aligned} -\min_{0 \leq s \leq 1} W(s) &= -\min_{0 \leq s \leq 1} (\sigma X(s) - rs) =_d \\ &=_d \max_{0 \leq s \leq 1} (rs + \sigma X(s)) \leq r + \sigma \max_{0 \leq s \leq 1} X(s), \end{aligned}$$

поскольку для центрированного гауссовского процесса $X(s) =_d -X(s)$. Поэтому справедливо неравенство

$$Q_3 \leq \mathbb{P} \left(\max_{0 \leq s \leq 1} X(s) \geq \frac{R_1}{\sigma} \gamma^\beta(t) - \frac{r}{\sigma} \right).$$

Рассуждая далее, как и при анализе Q_2 , получим, что

$$[t]Q_3 \rightarrow 0, \quad t \rightarrow \infty.$$

Осталось показать, что $[t]Q_1 \rightarrow 0, t \rightarrow \infty$. Это эквивалентно тому, что

$$a(t) := \ln[t] + \ln Q_1 \rightarrow -\infty.$$

Из асимптотики для вероятности переполнения (7) следует, что

$$b^{V-2} L(b) \ln \mathbb{P}(Q^*(0) > b) \rightarrow -\theta, \quad b \rightarrow \infty. \quad (32)$$

В рассматриваемом случае возьмем

$$b = b(t) = \left(\frac{1 + \delta/2}{\theta} \right)^\beta [\gamma(t)]^\beta.$$

Заметим, что тогда

$$b^{V-2} = \frac{\theta}{1 + \delta/2} \frac{1}{\gamma(t)}. \quad (33)$$

Используя условие (5), из (32) и (33) несложно получить, что

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\ln \mathbb{P}(Q^*(0) \geq ((1 + \delta/2)/\theta)^\beta \gamma^\beta(t))}{\ln t} &= \\ &= -\left(1 + \frac{\delta}{2}\right). \end{aligned}$$

Следовательно, при $t \rightarrow \infty$

$$\begin{aligned} a(t) &\sim \ln t + \ln Q_1 = \\ &= \ln t \left(1 + \frac{\ln \mathbb{P}(Q^*(0) \geq ((1 + \delta/2)/\theta)^\beta \gamma^\beta(t))}{\ln t} \right) \rightarrow \\ &\rightarrow -\infty. \end{aligned}$$

Итак, доказательство соотношения (11) для $M^*(t)$ завершено.

Обратимся к доказательству соотношения (12). С учетом предыдущего анализа для этого достаточно сделать лишь несколько пояснений. Во-первых, как и выше достаточно показать, что

$$\mathbb{P} \left(\frac{M(t)}{\gamma^\beta(t)} \geq \left(\frac{1 - \delta}{\theta} \right)^\beta \right) \rightarrow 1, \quad t \rightarrow \infty; \quad (34)$$

$$\mathbb{P} \left(\frac{M(t)}{\gamma^\beta(t)} \geq \left(\frac{1 + \delta}{\theta} \right)^\beta \right) \rightarrow 0, \quad t \rightarrow \infty. \quad (35)$$

Для доказательства нижней границы (34) заметим, что

$$\begin{aligned} Q(t) &= \sup_{0 \leq s \leq t} [W(t) - W(s)] = \\ &= W(t) - \inf_{0 \leq s \leq t} W(s) \geq W(t) - W(t - \Delta). \end{aligned}$$

Далее все выкладки остаются без изменения с заменой лишь $M^*(t)$ на $M(t)$.

Для доказательства верхней границы (35) заметим, что в силу начального условия $Q(0) = 0$ имеем $Q(t) \leq Q^*(t)$ и, следовательно, $M(t) \leq M^*(t)$. В этом случае справедливо неравенство

$$\mathbb{P} \left(\frac{M(t)}{\gamma^\beta(t)} \geq \left(\frac{1+\delta}{\theta} \right)^\beta \right) \leq \mathbb{P} \left(\frac{M^*(t)}{\gamma^\beta(t)} \geq \left(\frac{1+\delta}{\theta} \right)^\beta \right) \rightarrow 0, \quad t \rightarrow \infty.$$

Таким образом, соотношения (34) и (35), а значит, и сходимости (12) доказаны. \square

Замечание. Если $L(t) \rightarrow c \in (0, \infty)$, $t \rightarrow \infty$, то $\gamma(t) \sim c \ln t$ и нормировка в утверждении теоремы 3.1 становится более простой, а именно: формулы (11) и (12) принимают соответственно вид

$$\frac{M^*(t)}{(c \ln t)^\beta} \Rightarrow \left(\frac{1}{\theta} \right)^\beta, \quad \frac{M(t)}{(c \ln t)^\beta} \Rightarrow \left(\frac{1}{\theta} \right)^\beta, \quad t \rightarrow \infty.$$

Доказанная теорема позволяет непосредственно получить асимптотику максимума процесса нагрузки для специального важного случая, когда стохастическая компонента входного процесса является суммой независимых ДБД, т. е.

$$X(t) = \sum_{i=1}^n B_{H_i}(t), \quad t \geq 0, \quad (36)$$

где параметры Херста $H_i \in (0, 1)$. Без ограничения общности будем считать, что $H_1 > \max_{i>1} H_i$. Тогда дисперсия $v(t)$ процесса $\{X(t)\}$ имеет вид

$$v(t) = \sum_{i=1}^n t^{2H_i} = t^{2H_1} L(t),$$

где медленно меняющаяся функция $L(t) = 1 + \sum_{i>1} t^{2(H_i-H_1)} \rightarrow 1$, $t \rightarrow \infty$. Таким образом, дисперсия $v(t)$ является правильно меняющейся на бесконечности функцией с показателем $V = 2H_1 \in (0, 2)$. Обозначим $(2 - 2H_1)^{-1} = \delta$. Тогда в силу замечания выше имеет место такое

Следствие. Если компонента X входного процесса имеет вид (36), то

$$\frac{M^*(t)}{(\ln t)^\delta} \Rightarrow \left(\frac{1}{\theta} \right)^\delta, \quad \frac{M(t)}{(\ln t)^\delta} \Rightarrow \left(\frac{1}{\theta} \right)^\delta, \quad t \rightarrow \infty.$$

Данный результат говорит о том, что в асимптотическом анализе максимума процесса нагрузки доминирующую роль играет ДБД с наибольшим значением параметра Херста.

4 Заключение

В данной статье исследовано асимптотическое поведение максимума процесса нагрузки в жидкостной системе обслуживания с одним сервером. На вход системы поступает процесс, содержащий линейную (детерминированную) компоненту и случайную компоненту, описываемую центрированным гауссовским процессом, у которого дисперсия является регулярно меняющейся функцией с показателем $V \in (0, 2)$. К такому классу процессов, в частности, относится сумма независимых ДБД. Показано, что при соответствующей нормировке максимум процесса нагрузки на интервале $[0, t]$ сходится по вероятности (при $t \rightarrow \infty$) к некоторой явно выписанной константе. Этот результат обобщает соответствующий результат, полученный ранее в работе [9] для случая единственного входного процесса ДБД.

Литература

1. Leland W. E., Taqqu M. S., Willinger W., Wilson D. V. On the self-similar nature of ethernet traffic (extended version) // IEEE/ACM Transactions of Networking, 1994. Vol. 2. No. 1. P. 1–15.
2. Willinger W., Taqqu M. S., Leland W. E., Wilson D. Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements // Statistical Sci., 1995. Vol. 10. No. 1. P. 67–85.
3. Reich E. On the integrodifferential equation of Takacs I // Ann. Math. Stat., 1958. Vol. 29. P. 563–570.
4. Narayan O. Exact asymptotic queue length distribution for fractional Brownian traffic // Advances in Performance Analysis, 1998. Vol. 1. P. 39–63.
5. Hüsler J., Piterbarg V. I. Extremes of a certain class of Gaussian processes // Stochastic Processes and Their Applications, 1999. Vol. 83. P. 257–271.
6. Duffield N., O'Connell N. Large deviations and overflow probabilities for the general single server queue, with applications // Proc. Cambridge Philosophical Society, 1995. Vol. 118. P. 363–374.
7. Debicki K. A note on LDP for supremum of Gaussian processes over infinite horizon // Stat. Probab. Lett., 1999. Vol. 44. P. 211–220.
8. Duffy K., Lewis J. T., Sullivan W. G. Logarithmic asymptotics for the supremum of a stochastic process // Ann. Appl. Probab., 2003. Vol. 13. No. 2. P. 430–445.
9. Zeevi A., Glynn P. On the maximum workload in a queue fed by fractional Brownian motion // Ann. Appl. Probab., 2000. Vol. 10. P. 1084–1099.
10. Hüsler J., Piterbarg V. I. Limit theorem for maximum of the storage process with fractional Brownian as input // Stochastic Processes and their Applications, 2004. Vol. 114. P. 231–250.

11. *Taqqu M. S., Willinger W., Sherman R.* Proof of a fundamental result in self-similar traffic modeling // *Computer Communication Rev.*, 1997. Vol. 27. P. 5–23.
12. *Mandjes M.* Large Deviations of Gaussian Queues. — Chichester: Wiley, 2007. 339 p.
13. *Сенета Е.* Правильно меняющиеся функции. — М.: Наука, 1985. 143 с.
14. *Konstantopoulos T., Zazanis M., De Veciana G.* Conservation laws and reflection mappings with application to multiclass mean value analysis for stochastic fluid queues // *Stochastic Processes and their Applications*, 1996. Vol. 65. P. 139–146.
15. *Лидбеттер М., Линдгрен Г., Ротсен Х.* Экстремумы случайных последовательностей и процессов. — М.: Мир, 1989. 392 с.
16. *Лифшиц М. А.* Гауссовские случайные функции. — Киев: ТвиМС, 1995. 248 с.
17. *Adler R. J.* An introduction to continuity, extrema, and related topics for general Gaussian processes. — Hayward, CA: Institute of Mathematical Statistics, 1990. 170 p.

*Посвящается 70-летию со дня рождения В. В. Калашикова (1942–2001),
внесшего большой вклад в развитие регенеративного метода*

ОБ ОЦЕНИВАНИИ ВЕРОЯТНОСТИ ПЕРЕПОЛНЕНИЯ КОНЕЧНОГО БУФЕРА В РЕГЕНЕРАТИВНЫХ СИСТЕМАХ ОБСЛУЖИВАНИЯ*

Е. В. Морозов¹, Р. С. Некрасова²

Аннотация: Рассмотрены вопросы оценивания стационарной вероятности переполнения конечного буфера на основе регенеративного моделирования. Приведен вывод общего соотношения, связывающего в стационарном режиме вероятность потери с вероятностью простоя обслуживающего канала. Показано его применение для широкого класса систем с потерями, а также для системы с повторными вызовами и постоянной скоростью возвращения заявок с орбиты на обслуживание. Исследована эффективность этого соотношения при регенеративном оценивании вероятности потери при различных режимах загрузки системы, а также при использовании k -регенераций, возникающих при анализе немарковских систем с потерями. Приведены результаты численного моделирования.

Ключевые слова: системы с конечным буфером; вероятность потери; вероятность простоя; регенеративный метод оценивания; k -регенерации; система с повторными вызовами

1 Введение

Модели с ограничениями, в частности с конечным буфером, играют важную роль в анализе современных телекоммуникационных систем. В таких системах поток, образованный потерянными заявками, часто является входным потоком для другого узла коммуникационной системы, а вероятность потери является ключевым показателем качества обслуживания.

В данной работе представлено доказательство весьма общей формулы, связывающей стационарную вероятность потери P_{loss} со стационарной вероятностью простоя канала P_0 . Этот результат позволяет сравнить эффективность оценки P_{loss} , полученной с помощью оценки P_0 , и стандартной оценки, равной доле потерянных заявок. Формула через P_0 применяется также к системе обслуживания с повторными вызовами и с постоянной скоростью возвращения заявок с орбиты [1]. В статье также приведено условие стационарности такой модели в случае двух серверов. Исследования по регенеративному оцениванию вероятности блокировки в такой системе с повторными вызовами (как модели телефонных протоколов множественного доступа) были начаты в работе [2]. В данной

работе оценивание также опирается на регенеративный метод с использованием так называемых k -регенераций (возникающих в системах с входным пуассоновским потоком и/или экспоненциальным временем обслуживания), когда в моменты прихода (или ухода) заявок в системе находится k других заявок. В статье исследуется эффективность оценивания при использовании разных типов k -регенераций (в том числе классической 0-регенерации). Как известно, доверительные интервалы, использующие разные последовательности точек регенерации, асимптотически (с ростом числа наблюдений) эквивалентны [3]. Однако выбор последовательности может существенно повлиять на *скорость получения оценки*, и этот вопрос также рассматривается в статье.

Статья организована следующим образом. В разд. 2 описана регенеративная структура процессов в системах с потерями и входным процессом восстановления. Рассмотрены как классические, так и k -регенерации, возникающие в немарковских системах $M/G/1/n$ и $GI/M/m/n$. Приведена процедура построения доверительного интервала при регенеративном оценивании. Также в разд. 2 затрагивается вопрос асимптотической эквивалентности длин доверительных интервалов, использу-

*Работа поддержана РФФИ (проект 10-07-00017). Работа выполнена при поддержке Программы стратегического развития на 2012–2016 гг. «Университетский комплекс ПетрГУ в научно-образовательном пространстве Европейского Севера: стратегия инновационного развития».

¹Институт прикладных математических исследований КарНЦ РАН, Петрозаводский государственный университет, emorozov@karelia.ru

²Институт прикладных математических исследований КарНЦ РАН, Петрозаводский государственный университет, ruslana.nekrasova@mail.ru

ющих различные последовательности регенераций. В разд. 3 выведено основное соотношение между вероятностями P_{loss} и P_0 . В разд. 4 рассматривается двухсерверная система с конечным буфером, повторными вызовами и постоянной скоростью возвращения заявок с орбиты на обслуживание. Для такой системы в явном виде получено условие стационарности, включающее вероятность P_{loss} в некоторой тесно связанной с ней системе с потерями. В заключительном разделе 5 приведены некоторые результаты численного оценивания вероятности потери P_{loss} с использованием k -регенераций, а также формулы, связывающей P_{loss} и P_0 .

2 Регенеративная структура систем с потерями

При отсутствии точной формулы для вычисления вероятности P_{loss} возникает необходимость в ее надежном оценивании. В данной работе с этой целью в системе вида $GI/G/m/n$ применяется регенеративный метод. Обозначим через $\{t_n\}$ моменты прихода заявок в систему, и пусть $\{\tau_i := t_{i+1} - t_i\}$ — независимые одинаково распределенные (н.о.р.) интервалы входного потока, а $\{S_i\}$ — н.о.р. интервалы обслуживания (с типичными элементами τ, S соответственно). Пусть ν_n — число заявок в системе в момент прихода заявки $n \geq 1$. Пусть $\{\nu(t)t \geq 0\}$ есть (непрерывный справа) процесс числа заявок в системе, т.е. $\nu_n = \nu(t_n^-)$. Тогда классические регенерации процесса $\{\nu(t)\}$ (и других процессов в непрерывном времени в данной системе), порождаемые приходом заявок в пустую систему, рекурсивно определяются хорошо известным образом:

$$T_{n+1} = \min\{t_k > T_n : \nu_k = 0\}, \quad n \geq 0, \quad T_0 := 0. \quad (1)$$

Заметим, что регенерации (1) не являются событиями в потоке отказов и поэтому являются *скрытыми* по отношению к этому потоку. Регенерирующий процесс $\{\nu(t)\}$ называется *положительно возвратным*, если $ET < \infty$, где через T обозначена типичная длина цикла регенерации [4, 5]. Положительная возвратность является необходимым условием применимости регенеративного метода оценивания [6]. Отметим, что более общая конструкция так называемой однозависимой регенерации возникает в цепях Маркова, возвратных по Харрису. Анализ такой регенерации, в том числе в системах с потерями и в сетях обслуживания, содержится, например, в [7, 8].

Для более специальных систем существуют альтернативные (классические) регенерации. Сохраняя введенные обозначения, в системе $GI/M/m/n$

зафиксируем любое целое $k \in [0, m+n]$ и определим рекурсивно моменты k -регенерации в *дискретном времени* как номера тех заявок, которые находят в системе k других заявок, т.е.

$$\beta_{n+1}^{(k)} = \inf_l \{l > \beta_n^{(k)} : \nu_l = k\}, \quad \beta_0^{(k)} := 0. \quad (2)$$

Моменты (2) связаны с (1), как $T_n = t_{\beta_n^{(0)}}$, $n \geq 0$. Процесс $\{\nu_n\}$ образует апериодическую, неприводимую цепь Маркова с состояниями $\{0, \dots, n+m\}$, когда на события $\{\nu_l = k\}$ в момент прихода заявки времени обслуживания в каналах разыгрываются заново.

Заметим, что *мера регенерации* — распределение процесса в момент регенерации — является невырожденной при $k > 0$. Для пояснения рассмотрим событие $\{\beta_n^{(k)} = r\}$, на котором заявка r встречает k других заявок, т.е. $\nu_r = k$ (и это n -я подобная заявка среди первых r заявок). Рассмотрим $(m+1)$ -мерный процесс $W(l) := (Q_l, W_l^1, \dots, W_l^m)$, $l \geq 1$, где Q_l — число заявок в буфере, а W_l^i — i -е в порядке возрастания остаточное время обслуживания заявки в момент t_l прихода заявки l ($i = 1, \dots, m$). Отметим очевидное соотношение $\nu_l = Q_l + \sum_{i=1}^m I(W_l^i > 0)$,

где I — индикатор. На событии $\{\beta_n^{(k)} = r\}$, если $k < m$, то $W(r) =_{\text{st}} (0, 0, \dots, \phi_{k+1}, \dots, \phi_m)$, а если $k \geq m$, то $W(r) =_{\text{st}} (k-m, \phi_1, \dots, \phi_m)$, где $\{\phi_i\}$ есть н.о.р. случайные величины (с.в.), а знак $=_{\text{st}}$ означает стохастическое равенство. (С.в. $\{\phi_i\}$ имеют показательное распределение времени обслуживания.) Поэтому распределение $W(r)$ зависит лишь от типа регенерации k , но не от момента r .

Аналогично, если ν_n^* есть число заявок, которое оставляет n -я (уходящая) заявка в системе $M/G/1/n$, то для фиксированного $k \in [0, n]$

$$\alpha_{n+1}^{(k)} = \inf_l \{l > \alpha_n^{(k)} : \nu_l^* = k\}, \quad \alpha_0^{(k)} := 0$$

есть моменты k -регенерации для вложенной цепи Маркова $\{\nu_n^*\}$, мера регенерации которой является невырожденной при $k > 0$. В системах с конечным буфером незавершенная работа (стохастически) ограничена. Кроме того, в системе $GI/M/m/n$ длина k -цикла непериодическая, поскольку

$$\begin{aligned} P(A_k = 1) &= P(\nu_{l+1}^* = k | \nu_l^* = k) = \\ &= P(\tau_1 + \tau_2 > S > \tau_1) > 0. \end{aligned}$$

Аналогично непериодичность длины k -цикла в системе $M/G/1/n$ следует из условия $P(S_1 + S_2 > \tau > S_1) > 0$. Это обеспечивает положительную возвратность k -регенераций [9].

Дадим теперь строгое определение вероятности P_{loss} . Пусть $R(t)$ — число потерянных заявок, а $A(t)$ — число приходов в систему в интервале $[0, t]$. Поскольку поведение системы на цикле k -регенерации (далее — k -цикл), зависит от k , то будем обозначать через A_k, R_k, T_k число приходов, число потерь на k -цикле и длину k -цикла в непрерывном времени соответственно. Из теории регенерации следует, что в системе $GI/G/m/n$ с условием $P(\tau > S) > 0$ существуют и равны следующие пределы:

$$\lim_{t \rightarrow \infty} \frac{R(t)}{A(t)} = \frac{ER_k}{EA_k} := P_{\text{loss}} = \lim_{n \rightarrow \infty} P(I_n = 1), \quad (3)$$

где $I_n = 1$, если n -я заявка получает отказ ($I_n = 0$, иначе). Заметим, что условие $P(\tau > S) > 0$ влечет неперIODичность дискретной длины цикла занятости A_0 и существование слабого предела $\lim P(I_n = 1)$.

Обозначим $Z_k := R_k - P_{\text{loss}}A_k$. Регенеративный метод позволяет построить доверительный интервал для стационарной вероятности потери P_{loss} , если дисперсия $DZ_k := D(R_k - P_{\text{loss}}A_k) < \infty$.

Рассмотрим систему $M/G/1/n$ в предположении $ES^2 < \infty$, что влечет $DT_0 < \infty$ [4]. Пусть I_k есть индикатор события $\{\tau_1 + \dots + \tau_{k+1} > S_1 \geq \tau_1 + \dots + \tau_k\}$, вероятность которого равна

$$P_k = \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^k}{k!} dB(x) > 0,$$

где B — распределение времени обслуживания. Заметим, что при $I_k = 1$ за время обслуживания одной заявки, начинающей 0-цикл, в системе появится ровно k других заявок, т.е. начнется k -цикл. Следовательно, $T_0 \geq I_k T_k$ и неравенство $ET_k^2 \leq ET_0^2/P_k < \infty$ влечет $DT_k < \infty$. Поскольку $T_k \stackrel{\text{st}}{=} \tau_1 + \dots + \tau_{A_k}$, то

$$DT_k = EA_k D\tau + E\tau^2 DA_k \geq E\tau^2 DA_k, \quad k = 1, \dots, n.$$

Это дает $DA_k < \infty$, а также (поскольку $R_k \leq_{\text{st}} A_k$) и $\text{cov}(R_k, A_k) \leq EA_k^2$. Поэтому условие $ES^2 < \infty$ влечет $DZ_k < \infty$ и позволяет применить в системе $M/G/1/n$ оценивание вероятности P_{loss} на основе k -регенераций.

Предположим, что в рассматриваемой системе $DZ_k \in (0, \infty)$, и обозначим через p_k число k -циклов, полученных в процессе моделирования системы. Тогда стандартным образом можно получить $(1 - \gamma)\%$ доверительный интервал для P_{loss} в следующей форме:

$$\left[P_{\text{loss}}(p_k) - \frac{z_\gamma \sigma(p_k)}{\hat{A}(p_k) \sqrt{p_k}}, P_{\text{loss}}(p_k) + \frac{z_\gamma \sigma(p_k)}{\hat{A}(p_k) \sqrt{p_k}} \right], \quad (4)$$

где $P_{\text{loss}}(p_k)$ — оценка P_{loss} , $\hat{A}(p_k)$ — выборочная длина k -цикла, $\sigma^2(p_k)$ — выборочная оценка дисперсии DZ_k , а квантиль z_γ находится из условия $z_\gamma = \phi^{-1}((1 - \gamma)/2)$ ($\phi(x)$ — функция Лапласа). При этом $\hat{A}(p_k) \rightarrow EA_k$, $\sigma^2(p_k) \rightarrow DZ_k$ при $p_k \rightarrow \infty$ с в. 1.

Рассмотрим подробнее построение доверительного интервала по случайному числу $p_k(t)$ k -циклов, завершённых на периоде моделирования $[0, t]$. Введем последовательность н.о.р. с.в. $\{Z_k^{(i)} = R_k^{(i)} - P_{\text{loss}}A_k^{(i)}, i \geq 1\}$ с типичным элементом $Z_k = R_k - P_{\text{loss}}A_k$ и заметим, что

$$R(t) - P_{\text{loss}}A(t) = \sum_{i=1}^{p_k(t)} Z_k^{(i)} + o(t), \quad t \rightarrow \infty, \quad (5)$$

где величина $o(t)$ описывает число потерь на оставшейся части цикла, *накрывающего* момент t [6, 10]. С учетом $EZ_k = 0$ получаем такую центральную предельную теорему [6]:

$$\frac{R(t) - P_{\text{loss}}A(t)}{\sqrt{A(t)}} \Rightarrow N\left(0, \frac{DZ_k}{EA_k}\right), \quad t \rightarrow \infty, \quad (6)$$

где N обозначает нормальную с.в. Таким образом, отношение DZ_k/EA_k является константой, не зависящей от k (см. также [3]). Следующий результат также известен [3], однако представляется полезным дать краткое пояснение к его выводу. Пусть $p_i(t), p_j(t)$ — число i -циклов и j -циклов соответственно, полученных в интервале $[0, t]$. Отметим, что в рассматриваемом случае доверительный интервал строится как и (4), но с использованием случайного (а не детерминированного) числа циклов. Обозначим через $|I_i(t)|$ длину доверительного интервала, построенного по i -циклам и пусть $\hat{A}_i(t)$ — выборочная длина i -цикла. Обозначая $DZ_i = \sigma_i^2$, получаем

$$\begin{aligned} \frac{|I_i(t)|}{|I_j(t)|} &= \frac{\sigma_i \hat{A}_j(t) \sqrt{p_j(t)}}{\sigma_j \hat{A}_i(t) \sqrt{p_i(t)}} = \\ &= \sqrt{\frac{\sigma_i^2 \hat{A}_j(t)}{\hat{A}_i(t) \sigma_j^2}} \sqrt{\frac{\hat{A}_j(t) p_j(t)}{\hat{A}_i(t) p_i(t)}} \rightarrow 1 \text{ при } t \rightarrow \infty, \quad (7) \end{aligned}$$

где использовано постоянство отношения DZ_k/EA_k и усиленный закон больших чисел

$$\left. \begin{aligned} \hat{A}_i(t) &\rightarrow EA_i, \quad \hat{A}_j(t) \rightarrow EA_j; \\ \hat{A}_j(t) p_j(t) &\sim \hat{A}_i(t) p_i(t) \sim t, \quad t \rightarrow \infty, \end{aligned} \right\} \quad (8)$$

($a \sim b$ означает $a/b \rightarrow 1$). Таким образом, интервалы, построенные по различным последовательностям k -регенераций, асимптотически эквивалентны. Однако существенное различие в числе циклов, полученных при ограниченном времени моделирования, может дать преимущество одной последовательности перед другими. Этот вопрос рассматривается в секции 5. Подробное описание построения доверительных интервалов при регенеративном оценивании можно найти в [3, 11–14]. (Доказательство сходимости (6) для случайных сумм вида (5) содержится также в [6, 15].)

3 Соотношение между вероятностью потери и вероятностью простоя

В данном разделе для широкого класса регенеративных систем с потерями доказано соотношение, связывающее вероятность P_{loss} со стационарной вероятностью простоя канала P_0 . При этом используются лишь 0-регенерации. Сохраняя прежние обозначения, рассмотрим систему $GI/G/m/n$ с коэффициентом загрузки $\rho = ES/E\tau$. Заметим, что в формулировке теоремы 1 предполагается, что выбор свободного канала для новой заявки происходит равновероятно, если таких каналов несколько.

Теорема 1. В системе $GI/G/m/n$ при условии $P(\tau > S) > 0$ стационарная вероятность потери P_{loss} определяется формулой (3) и связана со стационарной вероятностью простоя (любого) канала P_0 следующим образом:

$$P_{\text{loss}} = 1 - \frac{m}{\rho}(1 - P_0). \quad (9)$$

Доказательство. Определим процесс накопленной работы в системе $W(t) = \sum_{i=1}^m W_i(t)$, где $W_i(t)$ есть не завершённая в момент t работа, предназначенная для канала i . Поскольку буфер конечен, то процесс $\{W(t)\}$ является стохастически ограниченным, а из условия $P(\tau > S) > 0$ следует положительная возвратность $ET_0 < \infty$. (Подробный анализ стационарности регенеративных систем, в том числе с конечным буфером, содержится в [5, 9].) Пусть $V(t)$ — суммарная нагрузка, поступившая в систему, $B(t)$ — обслуженная нагрузка, а $L(t)$ — потерянная нагрузка (время, которое требовалось для обслуживания потерянных заявок), все в интервале $[0, t]$. Получаем следующее уравнение баланса:

$$V(t) = W(t) + B(t) + L(t), \quad (10)$$

где, возможно, $V(0) = W(0) \neq 0$. Заметим, что $B(t) = \sum_{i=1}^m B_i(t)$, а

$$B_i(t) = \int_0^t I(W_i(u) > 0) du$$

есть время занятости канала i в интервале $[0, t]$, $i = 1, \dots, m$. Очевидно, что

$$V(t) = \sum_{k=1}^{A(t)} S_k, \quad L(t) = \sum_{k=1}^{R(t)} S_k. \quad (11)$$

Из усиленного закона больших чисел при $t \rightarrow \infty$ легко следует

$$\frac{V(t)}{t} \rightarrow \rho, \quad \frac{L(t)}{V(t)} \rightarrow \frac{ER}{EA} = P_{\text{loss}}. \quad (12)$$

Поскольку $W(t) \leq_{\text{st}} \sum_{i=1}^n S_i + \sum_{i=1}^m S_i(t)$, где не завершённое в канале i в момент t время обслуживания $S_i(t) = o(t)$ [10], то $W(t) = o(t)$, $t \rightarrow \infty$. (Последний результат можно также вывести непосредственно из положительной возвратности 0-регенераций.) Так как каналы идентичны, то получаем также

$$\frac{B(t)}{t} = \frac{\sum_{i=1}^m B_i(t)}{t} \rightarrow m(1 - P_0), \quad t \rightarrow \infty, \quad (13)$$

что вместе с (12) даёт

$$\frac{B(t)}{V(t)} \rightarrow \frac{m(1 - P_0)}{\rho}.$$

Таким образом, предел $(t - B_i(t))/t \rightarrow P_0$ является стационарной вероятностью простоя любого канала. Поделив обе части (10) на $V(t)$ и перейдя к пределу при $t \rightarrow \infty$, получаем (9). Наконец, как было отмечено (ввиду условия $P(\tau > S) > 0$), P_{loss} также является стационарной вероятностью потери приходящей заявки. \square

В случае одного сервера статистический аналог формулы (9) используется в работах [16, 17] при построении так называемой не прямой оценки вероятности P_{loss} (см. также [18]). Однако в случае нескольких серверов представление суммарного времени занятости в виде $B(t) = \sum_{i=1}^m B_i(t)$ там не используется. Формула (9) для системы с потерями вида $M/G/1/n$ также получена другим методом в [19, с. 333].

Как показано в разд. 5, формула (9) может быть полезна для оценки вероятности P_{loss} через оценку

вероятности простоя (или занятости) канала в случае большой нагрузки, когда прямая оценка P_{loss} с помощью метода Монте-Карло оказывается неэффективной.

В заключение этого раздела отметим, что соотношение (9) можно непосредственно применить к системам с буфером *случайного размера*, который регенерирует на периодах занятости системы. Такие системы могут представлять интерес для моделирования узлов связи в некоторых современных коммуникационных сетях. Кроме того, соотношение (9) верно для предложенной в [20] системы со случайным объемом поступающих заявок и ограничением на суммарный объем. В такой системе P_{loss} равна предельной доле потерянного объема. Также соотношение (9) верно для систем с ограниченным ожиданием/пребыванием заявки.

Нетрудно видеть, что для *жидкостной системы* со скоростью обслуживания C соотношение (9) принимает вид

$$P_{\text{loss}} = 1 - \frac{Cm(1 - P_0)}{\rho}. \quad (14)$$

Заметим, что в классических моделях $C = 1$, а в жидкостных моделях поступающая нагрузка $V(t)$ не разделяется на отдельные заявки, а следует, например, процессу Леви (с независимыми стационарными приращениями) с заданной интенсивностью ρ . В этом случае величина ρ/C является коэффициентом загрузки и, таким образом, формулы (14) и (9) эквивалентны.

Отметим, что если рассматривать P_{loss} , P_0 лишь как пределы в среднем, то соотношение (9) имеет место для гораздо более широкого класса систем, чем регенеративные.

4 Система с повторными вызовами и постоянной скоростью возвращения заявок с орбиты

В данном разделе рассматривается система с повторными вызовами типа $M/G/2/0$ без буфера с входным (пуассоновским) потоком первичных заявок с интенсивностью λ и произвольным временем обслуживания с интенсивностью μ . Эта система, далее обозначенная через Σ_O , успешно применена для моделирования занятости телефонных линий в мобильных системах [21], протоколов множественного доступа ALOHA [22], а также протокола ТСП с короткими сообщениями [23]. Когда серверы заняты, первичные заявки уходят на орбиту бесконечного объема, а затем вновь пытаются

попасть на серверы. Поток повторных попыток (при непустой орбите) является пуассоновским с параметром μ_0 и не зависит от величины орбиты, в отличие от классических систем с повторными вызовами. Как показано в работах [1, 18, 24], для анализа стационарности системы Σ_O можно использовать следующую систему с потерями Σ_L (без орбиты). Система Σ_L имеет тот же входной поток первичных заявок, то же распределение времени обслуживания, что и система Σ_O , но имеет еще один (независимый) пуассоновский входной поток заявок с параметром μ_0 . Если в момент прихода система Σ_L занята, то заявка теряется. Обозначим через P_{loss} стационарную вероятность потери в системе Σ_L . В работе [1] доказано, что условие

$$(\lambda + \mu_0)P_{\text{loss}} < \mu_0 \quad (15)$$

является достаточным (а в марковском случае также и необходимым) условием стационарности системы Σ_O . Очевидно, что вероятность P_{loss} удовлетворяет основному соотношению (9). Система с потерями Σ_L , однако, важна не только для определения зоны стационарности Σ_O . Как показано в [2, 18], в зоне нестационарности системы Σ_O оценка вероятности P_{loss} сходится с в. 1 к вероятности блокировки вызова P_{orb} в системе Σ_O . Более того, в работе [24] доказано, что в односерверной системе Σ_O в нестационарном режиме

$$P_{\text{orb}} = 1 - \frac{P_b}{\tilde{\rho}}, \quad (16)$$

где $\tilde{\rho} := (\lambda + \mu_0)/\mu$, а P_b есть стационарная вероятность занятости системы Σ_L . Очевидно, (16) является аналогом (9), поскольку $P_{\text{orb}} = P_{\text{loss}}$. Утверждение (16) позволяет провести анализ стационарности в *частично нестационарной* системе и достоверно оценить вероятность блокировки при неограниченно растущей орбите. Такая тесная связь двух систем мотивирует анализ условий стационарности системы Σ_O , представленный ниже.

Напомним формулу Эрланга для вероятности потери в системе $M/G/m/0$ [25]

$$P_{\text{loss}} = \frac{\rho^m/m!}{\sum_{k=0}^m \rho^k/k!}, \quad \rho := \lambda/\mu. \quad (17)$$

В работе [24] для системы Σ_O типа $M/G/1/0$ использование (17) позволило получить условие (15) при $\lambda = 1$ в форме $1/\mu_0 + 1 < \mu$. Ниже приведено в явном виде условие стационарности *двухсерверной* системы Σ_O .

Теорема 2. Система с повторными вызовами типа $M/G/2/0$ (с интенсивностью первичных заявок $\lambda = 1$) стационарна, если $\mu > 1/2$ и если

$$\mu_0 > \frac{\mu^2 + \mu - 1 - \mu\sqrt{\mu^2 + 2\mu - 1}}{1 - 2\mu} := \mu_{01}.$$

Доказательство. Используя формулу (17) для системы Эрланга с коэффициентом загрузки $\rho = (1 + \mu_0)/\mu$ и $m = 2$, получим условие стационарности (15) в форме неравенства

$$-\mu^2\mu_0 - \mu(\mu_0 + \mu_0^2) + \frac{1}{2} + \mu_0 + \frac{\mu_0^2}{2} < 0. \quad (18)$$

Разложение на множители при условии $\mu \geq \sqrt{2} - 1$ имеет вид

$$\left(\frac{1}{2} - \mu\right)(\mu_0 - \mu_{01})(\mu_0 - \mu_{02}) < 0. \quad (19)$$

При $\mu < \sqrt{2} - 1$ неравенство (18) не выполняется и система нестационарна. Таким образом, стационарность может иметь место только при $\mu \geq \sqrt{2} - 1$, когда справедливо разложение в левой части (19). Анализ значения параметров (μ, μ_0) в области $(\sqrt{2} - 1, \infty) \times (0, \infty)$, при которых выполняется неравенство (18), легко приводит к утверждению теоремы. \square

Отметим, что условие (15) совпадает с критерием стационарности системы с повторными вызовами вида $M/M/2/0$, полученным в [26] на основе традиционной для марковских моделей техники (см. также [27]). В принятых обозначениях (и при $\lambda = 1$) оба эти условия могут быть записаны, например, в форме

$$(1 + \mu_0)^2 < 2\mu\mu_0(1 + \mu + \mu_0).$$

Область значений параметров μ и μ_0 , в которой система с повторными вызовами типа $M/M/2/0$ обладает стационарностью, представлена на рис. 1.

Рассмотрим более общую m -серверную систему с повторными вызовами (с постоянной скоростью возвращения с орбиты) с конечным буфером и с входным потоком восстановления с интенсивностью λ , изученную в [1]. Предполагается, что серверы стохастически эквивалентны (и S — типичное время обслуживания), так что, в частности, вновь поступающие заявки распределяются равномерно по свободным серверам (если их несколько). Пусть $V(t)$ — поступившая, а $B(t)$ — обслуженная нагрузка в интервале $[0, t]$, причем $B(t) = \sum_{i=1}^m B_i(t)$, где $B_i(t)$ есть время занятости сервера i в интервале $[0, t]$, $i = 1, \dots, m$.

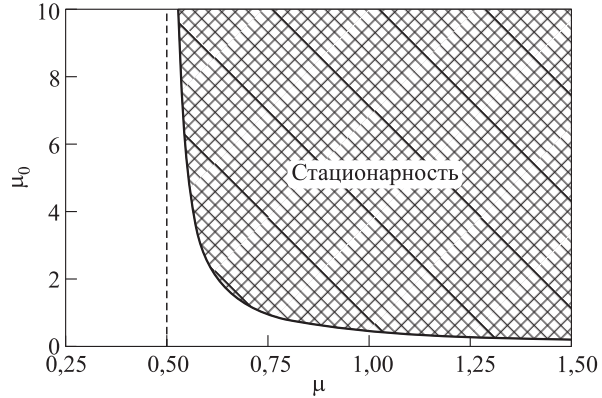


Рис. 1 Область стационарности системы с повторными вызовами вида $M/M/2/0$

Пусть $W(t)$ есть не завершенная в момент t работа в системе (в буфере, в серверах и на орбите). Для данной системы имеет место такой аналог уравнения баланса (10), в котором отсутствует слагаемое, связанное с потерями:

$$V(t) = W(t) + B(t) = W(t) + \sum_{i=1}^m B_i(t), \quad t \geq 0. \quad (20)$$

Предположим теперь, что система стационарна, т.е. вложенный процесс регенераций положительно возвращен. Обозначим $\rho = \lambda ES$. Тогда, в частности, $W(t) = o(t)$, $t \rightarrow \infty$, и из (20) следует предельное соотношение (см. (11)–(13))

$$\rho = mP_b, \quad (21)$$

где P_b есть стационарная вероятность занятости любого сервера. В частности, $P_b = \rho$ при $m = 1$. Более того, если входной процесс пуассоновский, то (по свойству PASTA) ρ есть также стационарная вероятность блокировки (ухода на орбиту) вновь поступающей первичной заявки (при $m = 1$). Интересно отметить, что соотношение (21) верно также и для общей системы $GI/G/m$ (с бесконечным буфером и без орбиты). (Здесь важно лишь отсутствие потерь.) Однако критерий стационарности $\rho/m < 1$ этой системы, разумеется, отличается от (достаточного) условия стационарности (15) системы с повторными вызовами. Например, для системы с повторными вызовами вида $M/G/1/K$ условие стационарности имеет вид

$$\lambda\rho < \mu_0(1 - \rho). \quad (22)$$

С учетом свойства PASTA [6] условие (22) имеет ясную физическую интерпретацию: левая часть неравенства есть интенсивность ухода заявок на орбиту, а правая часть есть максимальная интенсивность (успешного) ухода заявок с орбиты на сервер,

поскольку множитель $1 - \rho$ есть стационарная вероятность незанятости сервера (когда успешный уход с орбиты возможен).

5 Результаты численного моделирования

В данном разделе представлены некоторые численные результаты оценивания вероятности P_{loss} по методу Монте-Карло с использованием (в очевидных обозначениях) традиционной оценки $\hat{P}_{\text{loss}}(i_0) = \hat{R}(i_0)/\hat{A}(i_0)$ и оценки на основе формулы (9)

$$\hat{P}_l := \hat{P}_l(i_0) = 1 - \frac{m}{\rho} \hat{P}_b(i_0)$$

соответственно в зависимости от числа 0-регенераций i_0 , где $\hat{P}_b(i_0)$ — оценка стационарной вероятности занятости. (Обозначение $\hat{P}_l(i_0)$ и $\hat{P}_b(i_0)$ подчеркивает, что используется *шкала циклов*, а не *отдельных наблюдений*.) В табл. 1 приведены результаты оценивания для системы $M/M/m/n$ в зависимости от коэффициента загрузки ρ , а также числа серверов m и величины буфера n . Величина $VR = D[\hat{P}_l(i_0)]/D[\hat{P}_{\text{loss}}(i_0)]$ равна отношению выборочных дисперсий оценок. Наблюдения проводились до получения заданной точности $\varepsilon > 0$. Значения оценок также сравниваются с точным значением, вычисленным по формуле [25]:

$$P_{\text{loss}} = \frac{\rho^m}{m!} \theta^n P_0,$$

где $\theta = \rho/m$, а

$$P_0 = \left(\sum_{k=0}^m \frac{\rho^k}{k!} + \frac{\rho^m}{m!} \sum_{k=1}^n \theta^k \right)^{-1}.$$

Критерием эффективности является необходимое число 0-циклов для получения заданной точности, а также дисперсия оценки. Как видно из табл. 1, в случае *малой нагрузки* ($\rho/m < 1$) классическая оценка \hat{P}_{loss} более эффективна, чем \hat{P}_l , как по скорости построения, так и по величине дисперсии. При

средней нагрузке ($\rho/m = 1$) и буфере $n = 0$ эффективность оценок близка, однако с ростом величины буфера оценка \hat{P}_{loss} становится предпочтительнее. При *большой нагрузке* ($\rho/m = 4$) и $n = 0$ оценка \hat{P}_l оказывается эффективней как по времени оценивания, так и по величине дисперсии. Это согласуется с результатами работ [17, 24], где показана эффективность *непрямой* оценки \hat{P}_l при большой нагрузке (а также и ее отрицательная корреляция с оценкой \hat{P}_{loss}). При увеличении величины буфера оценка \hat{P}_{loss} строится быстрее, однако имеет большую дисперсию, чем \hat{P}_l и поэтому нельзя сделать однозначный вывод об эффективности оценок.

Аналогичные численные результаты получены для системы $\text{Pareto}/M/m/n$ (с распределением Парето входного потока).

Исследовалась также эффективность применения различных k -регенераций. Результаты приведены на рис. 2 и в табл. 2, где оценка $\hat{P}_{\text{loss}}(i) = R(i)/A(i)$ строилась по *числу наблюдений* i . Число наблюдений, как правило, пропорционально времени моделирования и позволяет оценить реальную скорость получения оценки, в отличие от числа циклов, различие в длинах которых может быть очень значительным. На рис. 2 представлена зависимость от *числа наблюдений* границ доверительного интервала для вероятности потери в системе $M/M/2/4$, где рассматривались k -регенерации при $k = 0, 3, 6$. С ростом числа наблюдений доверительные интервалы для всех типов регенераций сближаются между собой, что соответствует (8). Однако при малых k оценивание происходит быстрее, поэтому (при малой нагрузке) такие регенерации (в том числе 0-регенерации) более эффективны. Отметим, что скорость оценивания возрастает с ростом k , однако при $k \geq 7$ она растет незначительно. При максимальном $k = m + n = 10$ оценка имеет наименьшую дисперсию, и поэтому (при большой нагрузке) такие регенерации в моменты потерь наиболее эффективны. Причина состоит в том, что классические регенерации в данном случае редки и их использование для оценки вероятности P_{loss} оказывается неэффективным. В табл. 3. представлены результаты анализа системы $\text{Pareto}/M/4/6$ при $\rho/m = 4$ и нескольких типах k -регенераций.

Таблица 1 Системы $M/M/m/n$, $\lambda = 4$

ρ	m	n	P_{loss}	\hat{P}_{loss}	i_0	\hat{P}_l	i_0	VR	ε
1	2	0	0.200	0.203	1242	0.228	4439	6.269	0.05
2	2	0	0.400	0.419	1790	0.391	1791	1.007	0.05
4	4	4	0.139	0.144	3810	0.041	5914	11.318	0.05
8	2	0	0.781	0.783	2193	0.779	1332	0.034	0.05
4	1	3	0.751	0.734	1840	0.692	8531	0.059	0.05

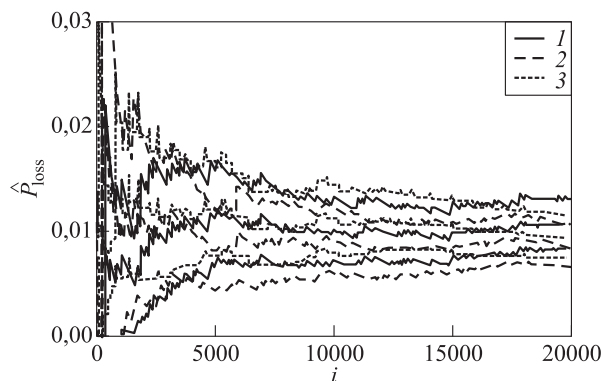


Рис. 2 Доверительное оценивание P_{loss} на основе k -регенераций, $k = 0$ (1), 3 (2), и 6 (3), в системе $M/M/2/4$ при $\rho/m = 0.5$

Таблица 2 Система $M/M/2/4$ при $\rho/m = 0.5$

k	i	$\hat{P}_{\text{loss}}(i)$	$\text{Var}[\hat{P}_{\text{loss}}(i)]$	ε
0	78	0,013	0,001	0,05
3	291	0,028	0,005	0,05
6	1979	0,005	0,228	0,05

Таблица 3 Система $\text{Pareto}/M/4/6$ при $\rho/m = 4$

k	i	$\hat{P}_{\text{loss}}(i)$	$\text{Var}[\hat{P}_{\text{loss}}(i)]$	ε
4	14489	0,756	0,131	0,05
6	2933	0,738	0,120	0,05
7	1927	0,753	0,135	0,05

6 Заключение

В статье рассмотрен ряд вопросов регенеративного оценивания стационарной вероятности потери в системах обслуживания с конечным буфером. Доказано общее соотношение, связывающее в стационарном режиме вероятность потери с вероятностью простоя обслуживающего канала для широкого класса систем, где потери могут быть вызваны различными причинами. Исследована эффективность этого соотношения при регенеративном оценивании вероятности потери при использовании k -регенераций, возникающих при анализе немарковских систем на основе вложенных цепей Маркова. Получены в явном виде условия стационарности для двухсерверной системы с повторными вызовами и постоянной скоростью возвращения заявок с орбиты на обслуживание. Приведены некоторые результаты численного моделирования.

Литература

1. Avrachenkov K., Morozov E. V. Stability analysis of $GI/G/c/K$ retrial queue with constant retrial rate. INRIA (Sophia Antipolis): Research Report, 2010. No. 7335.

2. Avrachenkov K., Goricheva R. S., Morozov E. V. Verification of stability region of a retrial queueing system by regenerative method // Modern Probabilistic Methods for Analysis and Optimization of Information and Telecommunication Networks: International Conference Proceedings. — Minsk, 2011. P. 22–28.
3. Shedler G. S. Regeneration and networks of queues. — New-York: Springer-Verlag, 1987.
4. Wolff R. W. Stochastic modeling and the theory of queues. — Englewood Cliffs, NJ: Prentice Hall, 1989.
5. Morozov E. V. Weak regeneration in modeling of queueing processes // Queueing Syst., 2004. Vol. 46. P. 295–315.
6. Asmussen S. Applied probability and queues. — 2nd ed. — New York: Springer-Verlag, 2003.
7. Sigman K. Queues as Harris recurrent Markov chains // Queueing Syst., 1988. No. 3. P. 179–198.
8. Sigman K. One-dependent regenerative processes and queues in continuous time // Math. Oper. Res., 1990. No. 15. P. 175–189.
9. Морозов Е., Делгадо Р. Анализ стационарности регенеративных систем обслуживания // Автоматика и телемеханика, 2009. No. 12. С. 42–58.
10. Smith W. L. Regenerative stochastic processes // Proc. Roy. Soc. Ser. A, 1955. Vol. 232. P. 6–31.
11. Крэйн М., Лемуан О. Введение в регенеративный метод анализа моделей. — М.: Наука, 1982. 104 с.
12. Иглехарт Д. Л., Шедлер Д. С. Регенеративное моделирование сетей массового обслуживания. — М.: Радио и связь, 1984. 136 с.
13. Glynn P. W., Iglehart D. L. Conditions for the applicability of the regenerative method // Management Sci., 1993. Vol. 39. P. 1108–1111.
14. Glynn P. W. Some topics in regenerative steady-state simulation // Acta Appl. Math., 1994. No. 34. P. 225–236.
15. Биллингсли П. Сходимость вероятностных мер. — М.: Наука, 1977. 352 с.
16. Whitt W. A review of $L = \lambda W$ and extensions // Queueing Syst., 1991. Vol. 9. P. 235–268.
17. Srikant R., Whitt W. Variance reduction in simulations of loss models // Oper. Res., 1999. Vol. 47. No. 4. P. 509–523.
18. Горичева Р. С., Морозов Е. В. Регенеративное моделирование вероятности потери в системах обслуживания с конечным буфером // Труды Карельского научного центра РАН, 2010. № 3. С. 20–29.
19. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1955. 529 с.
20. Тихоненко О. М. Модели массового обслуживания в системах обработки информации. — Минск: Университетское, 1990. 191 с.
21. Fayolle G. A simple telephone exchange with delayed feedback // Teletraffic Anal. Comp. Performance Evaluation, 1986. Vol. 7. P. 245–253.
22. Choi B. D., Rhee K. H., Park K. K. The $M/G/1$ retrial queue with retrial rate control policy // Prob. Engng. Informational Sci., 1993. Vol. 7. P. 29–46.

23. *Avrachenkov K., Yechiali U.* Retrial networks with finite buffers and their application to Internet data traffic // *Prob. Engng. Informational Sci.*, 2008. Vol. 22. P. 519–536.
24. *Морозов Е. В., Некрасова Р. С.* Оценивание вероятности блокировки в системе с повторными вызовами и постоянной скоростью возвращения заявок с орбиты // *Труды Карельского научного центра РАН*, 2011. № 5. С. 63–74.
25. *Гнеденко Б. В., Коваленко И. Н.* Введение в теорию массового обслуживания. — М.: Наука, 1987. 336 с.
26. *Artalejo J. R.* Stationary analysis of the characteristics of the $M/M/2$ queue with constant repeated attempts // *Opsearch*, 1996. Vol. 33. P. 83–95.
27. *Artalejo J. R., Gómez-Corral A., Neuts M. F.* Analysis of multiserver queues with constant retrial rate // *Eur. J. Oper. Res.*, 2001. Vol. 135. P. 569–581.

ВЕРОЯТНОСТНЫЕ МОДЕЛИ МНОГОПРОЦЕССОРНЫХ СИСТЕМ: СТАЦИОНАРНОСТЬ И МОМЕНТНЫЕ СВОЙСТВА*

Е. В. Морозов¹, А. С. Румянцев²

Аннотация: Дан анализ основных моделей многопроцессорных систем (МС), где для обработки заявки требуется случайное число процессоров. Предложена и исследована новая модель таких систем, в которой времена обработки заданий данной заявки на всех требуемых процессорах являются идентичными. Это предположение, отражающее реальный процесс обработки, существенно усложняет анализ. Для исследования данной модели построены минорантная и мажорантная (классические) модели, с помощью которых для ряда важных частных случаев удалось получить как условия стационарности исходной модели, так и моментные свойства стационарного процесса нагрузки.

Ключевые слова: многопроцессорные системы; групповое занятие процессоров; идентичные времена обработки; условия стационарности; моментные свойства; стационарный процесс нагрузки; вычислительный кластер

1 Введение

В настоящее время растет интерес к моделированию МС, в первую очередь систем с массово-параллельной архитектурой, таких как, например, вычислительные кластеры (ВК). Для разработки эффективных алгоритмов управления очередями доступа к МС нужна соответствующая (вероятностная) модель, которая позволяет экспериментально определять качество обслуживания, обеспечиваемое данной МС. В этой связи отметим работу [1], в которой описаны особенности работы с МС и представлен хороший обзор современных способов моделирования загрузки в таких системах.

Как известно, даже для классической системы $GI/G/m$ отсутствуют явные формулы для основных стационарных характеристик, а асимптотические оценки, как правило, неточны [2]. Еще большие трудности возникают при моделировании современных МС. Отметим ряд таких проблем. Достаточно часто времена вычисления задач в МС адекватно описываются распределениями с *тяжелыми хвостами* (например, распределением Парето) [1, 2]. Наличие тяжелых хвостов требует иного подхода, чем традиционно используемый в теории очередей, и этот подход активно развивается. Кроме того, как правило, некоторые характеристики являются зависимыми (например, время вычисления задачи и ее размер [1, 3]). Также во многих современных МС условия конечности моментов

незавершенной работы [4], распределение хвоста времени ожидания [5], точность асимптотических оценок [2], алгоритмы «справедливого» распределения задач по процессорам [6] существенно зависят от коэффициента загрузки системы, в частности от наличия так называемых «резервных процессоров». Кроме того, важной особенностью многих современных МС является свойство *долгой памяти* входного процесса, т.е. расхожимость ряда автокорреляций, что существенно усложняет процедуру оценивания [1, 3]. Поэтому анализ моделей, отражающих эти новые аспекты современных МС является весьма важным. В данной статье проанализированы известные ранее модели МС с групповым занятием процессоров приходящей заявкой. Кроме того, исследуется новый класс таких моделей, где учитываются некоторые важные особенности современных МС, а также изучены моментные свойства процесса загрузки в таких системах.

Статья организована следующим образом. В разд. 2 дан обзор основных моделей МС, в которых каждой заявке требуется несколько процессоров для обслуживания. Приведены условия стационарности таких систем. Раздел 3 посвящен развитию модели ВК на основе модифицированной рекурсии Кифера–Вольфовица, предложенной в работах [7–9]. В частности, на основе этой рекурсии строятся классические системы, которые в определенном смысле являются минорирующей и мажорирующей для исходной модели. Затем с по-

*Работа поддержана РФФИ (проект 10-07-00017). Работа выполнена при поддержке Программы стратегического развития на 2012–2016 гг. «Университетский комплекс ПетрГУ в научно-образовательном пространстве Европейского Севера: стратегия инновационного развития».

¹Институт прикладных математических исследований КарНЦ РАН, Петрозаводский государственный университет, emorozov@karelia.ru

²Институт прикладных математических исследований КарНЦ РАН, ar0@krc.karelia.ru

мощью этих систем получены условия стационарности основной модели, а также найдены моментные свойства ее стационарного процесса нагрузки. В разд. 4 рассматриваются вопросы численного анализа данной модели на основе реальных данных лог-файла кластера ЦКП КарНЦ РАН «Центр высокопроизводительной обработки данных».

2 Многопроцессорные системы, в которых для обработки заявки требуется случайное число процессоров

Для дальнейшего анализа важно разделять системы, в которых заявке с номером i одновременно требуется случайное число $N_i \geq 1$ процессоров, на системы с *независимым освобождением процессоров* и системы с *одновременным освобождением процессоров*. В первом случае времена обслуживания на всех N_i процессорах являются независимыми одинаково распределенными (н.о.р.), а во втором случае на всех N_i процессорах используется одна и та же реализация времени обслуживания (т.е. времена обслуживания *идентичны*). Системы второго типа существенно более сложны для анализа [10]. Для таких систем известны лишь численные результаты [11], а при отсутствии буфера — также некоторые аналитические результаты [12–14].

Следуя работам [10, 15, 16], рассмотрим важные для дальнейшего анализа системы с независимым освобождением процессоров. Рассмотрим систему типа $M/M/m$ с интенсивностью входного (пуассоновского) потока λ , дисциплиной FIFO, где заявке i требуется N_i процессоров, на каждом из которых независимо реализуется экспоненциальное время обслуживания с параметром μ . Предполагается, что $\{N_i\}$ являются н.о.р. случайными величинами (с.в.) с заданным распределением $p_k = P(N = k)$, где N — типичный элемент последовательности. Для такой (изначально свободной) системы в работе [15] получено следующее *достаточное условие* существования стационарного режима:

$$\lambda \sum_{k=1}^m \sum_{j=0}^{k-1} \frac{p_k}{\mu(m-j)} < 1. \quad (1)$$

В [16] исследована двухпроцессорная система с двумя независимыми пуассоновскими потоками с интенсивностями λ_i , $i = 1, 2$. Заявки первого класса обслуживаются на одном процессоре, а заявки второго — на двух (с одинаковой интенсивностью μ). Очевидно, эта модель является частным случаем

предыдущей. С помощью минорантной системы $M/M/2$ (с заявками только 1-го типа) и мажорантной системы типа $M/M/1$ (с заявками только 2-го типа) в работе [16] получено стационарное распределение исходной системы, а также *критерий ее стационарности* в виде

$$\frac{\lambda}{2\mu}(2 - p_1^2) < 1. \quad (2)$$

(Заметим, что (1) влечет (2).) В работе [10] обсуждается неоптимальность дисциплины FIFO в рассматриваемых моделях, поскольку некоторые процессоры могут простаивать при наличии очереди. (Иными словами, используемая дисциплина не является сохраняющей работу.) В этом состоит основная трудность аналитического исследования таких моделей МС.

Модель с независимым освобождением процессоров может описывать поведение узлов вычислительной грид-системы (объединения слабосвязанных вычислителей), так как в этом случае времена вычисления заданий, запускаемые на отдельных узлах и связанные с данной заявкой, являются независимыми.

В работе [12] описана система хранения с потерями, в которой на вход в блочное устройство хранения данных поступают пуассоновские потоки заявок k классов. Заявке i класса k требуется случайное число b_i единиц хранения на случайное время t_i . (Эти величины имеют заданные распределения, зависящие от k .) Отказ в обслуживании (уход без возвращения) определяется при помощи множества допустимых состояний $\Omega = \{(n_1, \dots, n_k)\}$, описывающих распределение числа заявок по классам. В [12] получено стационарное распределение состояний системы для произвольного множества Ω и произвольных распределений времени хранения. Если Ω допускает полное разделение ресурсов (все ресурсы могут быть заняты любым классом заявок), то в [12] получена эффективная с вычислительной точки зрения рекурсия для приближенного расчета вероятности отказа. В работе [13] рассматриваются вопросы аппроксимации стационарного распределения для данной модели. Описанная модель может быть полезна для моделирования ВК, в котором пользователи часто разделяются на группы с разными правами и ограничениями на доступ к ресурсам. Вместо блочного устройства можно рассматривать МС, в которой заявки занимают одновременно несколько процессоров.

Современные ВК имеют сотни тысяч процессоров, но, как правило, решаемые на них задачи не имеют столь же высокой степени распараллеливания. Поэтому в качестве подходящей модели таких

ВК обосновано рассматривать системы с бесконечным числом процессоров и групповым поступлением заявок. (Такие модели адекватно описывают системы с большим числом процессоров [17, 18].)

В работе [19] рассмотрена система с бесконечным числом процессоров, в которую поступает поток групп заявок, представляющий собой процесс восстановления. Группа n характеризуется маркой Y_n , причем последовательность $\{Y_n\}$ образует цепь Маркова. Группа n требует выполнения N_n задач, причем распределение N_n зависит от пары (Y_n, Y_{n-1}) , а времена вычисления (в каждой группе) являются н. о. р. с. в. В этой модели процессоры освобождаются независимо, и поэтому она также может быть использована для описания вычислительной грид-системы. Важнейшей характеристикой является также число процессоров $\nu_i(n)$, занятых в момент прихода n -й заявки класса $i \in [1, M]$, $n \geq 1$. В работе [19] найдены автокорреляционные функции числа занятых процессоров для непрерывного времени и в моменты прихода заявок каждого класса.

В работе [20] получено преобразование Лапласа–Стилтьеса функции распределения стационарного числа ν занятых процессоров в системе типа $M/G/\infty$ (с интенсивностью входного потока λ и временем обслуживания S), где каждой заявке (единственного) класса требуется случайное число N процессоров (с заданным распределением). Показано, что условие стационарности имеет вид

$$\rho := \lambda E N E S < \infty \quad (3)$$

и что $E\nu = \rho$. (Последний результат хорошо известен в случае ординарного потока, т.е. при $P(N=1)=1$.) В работе [21] изучаются моментные свойства, свойство долгой памяти процесса величины очереди в системе $M/G/\infty$, а в работе [22] исследуется свойство долгой памяти периодов занятости такой системы. В [23] получено стационарное распределение числа занятых процессоров в моменты прихода в системе с групповым поступлением (размер группы — постоянная величина) и независимым экспоненциальным временем обслуживания каждой заявки внутри группы. Наконец, отметим работу [3], которая содержит обзор имеющихся в литературе моделей массово-параллельных МС, а также анализ некоторой новой модели на основе марковских цепей.

3 Модель вычислительного кластера

В данном разделе изучается модель МС, в которой новая заявка занимает несколько процессоров

на *идентичное* время, что существенно усложняет анализ по сравнению со случаем независимых времен обработки. (Анализ этой модели был начат в работах [7, 8].) В данной модели в момент освобождения нескольких процессоров на обслуживание может поступить одновременно несколько заявок и это не позволяет применить анализ из работы [15].

Рассмотрим следующее обобщение классической m -процессорной системы $GI/G/m$ с н.о.р. интервалами между заявками $\{T_n\}$ и н.о.р. временами обслуживания $\{S_n\}$, в которой i -й приходящей заявке требуется одновременно случайное число процессоров $N_i \in [1, m]$. Если число свободных процессоров меньше N_i , то она ожидает в буфере освобождения недостающего числа процессоров. Соответствующая модификация рекурсии Кифера–Вольфовица для вектора процесса нагрузки $W_i := (W_i(1), \dots, W_i(m))$ [24] принимает вид:

$$W_{i+1} = R(W_i(N_i) + S_i - T_i, \dots, W_i(N_i) + S_i - T_i, \\ W_i(N_i + 1) - T_i, \dots, W_i(m) - T_i)^+, \quad (4)$$

где оператор R упорядочивает компоненты в порядке возрастания, $(\cdot)^+ = \max(0, \cdot)$ (для вектора операция применяется покомпонентно). Заметим, что первые N_i компонент вектора одинаковы, так как заявка i занимает сразу N_i процессоров. (Если в момент прихода $k > N_i$ процессоров свободны, то заявка не ждет в очереди.) По условию, каждая заявка освобождает все занимаемые ею процессоры одновременно. Обозначим через $D_i := W_i(N_i)$ время ожидания заявки i в очереди.

Мы будем использовать так называемый каплинг-метод, позволяющий сравнивать траектории случайных процессов [25]. Нетрудно доказать следующее утверждение об операторе R из (4). (Доказательства можно также найти в [26–28].)

Лемма 3.1. Пусть векторы $X \leq Y$ (покомпонентно). Тогда

$$R(X) \leq R(Y).$$

Ниже доказано, что для предложенной модели ВК минорантной будет система, в которой i -я заявка заменяется на группу из N_i заявок, причем каждая заявка из данной группы имеет одно и то же время обслуживания S_i . (Таким образом, очередная заявка из группы немедленно занимает освободившийся процессор.) Обозначим эту систему через $\Sigma^{(\text{low})}$ и снабдим ее характеристики верхним индексом low. Отметим, что в системе $\Sigma^{(\text{low})}$ на один процессор может распределяться несколько заявок из группы.

Теорема 3.2. Пусть $W_0^{(\text{low})} = W_0 = 0$. Тогда

$$W_{i+1}^{(\text{low})} \leq W_{i+1}, \quad i \geq 0. \quad (5)$$

Доказательство. Будем считать, что в момент прихода заявки i процессоры нумеруются в порядке возрастания нагрузки, как и компоненты вектора W_i в результате применения оператора R . Ввиду эквивалентности процессоров такая процедура не меняет вероятностных свойств процесса нагрузки. В системе $\Sigma^{(\text{low})}$ на процессор с номером N_i из i -й группы может быть распределено не более одной заявки. Действительно, предположим, что на процессор N_i распределено не менее двух заявок (со временем обслуживания S_i). Тогда на некоторый процессор с номером $k < N_i$ не поступает заявка. Это означает, что после размещения заявки на процессоре N_i выполнено неравенство $W_i^{(\text{low})}(k) > W_i^{(\text{low})}(N_i) + S_i$. Но это противоречит условию $W_i^{(\text{low})}(k) \leq W_i^{(\text{low})}(N_i)$. Далее, если на некоторый процессор с номером $k < N_i$ распределено $n \in (1, N_i]$ заявок, то

$$W_i^{(\text{low})}(k) + nS_i \leq W_i^{(\text{low})}(N_i) + S_i. \quad (6)$$

Действительно, в момент распределения заявки с номером $n \leq N_i$ из i -й группы на процессор k должно быть выполнено неравенство $W_i^{(\text{low})}(k) + (n-1)S_i \leq W_i^{(\text{low})}(N_i)$.

Так как обе системы изначально свободны, то до применения оператора R первые N_i компонент векторов $W_0^{(\text{low})}$ и W_0 совпадают. Предположим (по индукции), что $W_i^{(\text{low})} \leq W_i$ для некоторого i . Поскольку процессор с номером N_i в системе $\Sigma^{(\text{low})}$ получит не более одной заявки, то в силу предположения индукции

$$(W_i^{(\text{low})}(N_i) - T_i)^+ \leq (W_i^{(\text{low})}(N_i) + S_i - T_i)^+ \leq (W_i(N_i) + S_i - T_i)^+,$$

а для компонент с номерами $k < N_i$ по свойству (6) и предположению индукции следует

$$(W_i^{(\text{low})}(k) + nS_i - T_i)^+ \leq (W_i^{(\text{low})}(N_i) + S_i - T_i)^+ \leq (W_i(N_i) + S_i - T_i)^+.$$

Теперь, применив оператор R , в силу леммы 3.1 получаем (5).

Замечание. Система $\Sigma^{(\text{low})}$ в действительности может быть использована как модель вычислительной грид-системы. На практике поступающая на грид заявка часто является группой заданий, требующих перебора в пространстве параметров модели.

Задания в группе, как правило, независимы и поступают на свободные вычислители, не дожидаясь одновременного освобождения всех требуемых данной группе вычислителей. Полученный выше результат показывает, что для подобных задач переборного типа целесообразнее использовать архитектуру грид-системы, чем ВК, так как время ожидания заявки в такой системе в среднем оказывается меньше.

Поскольку в настоящее время мощности ВК исчисляются сотнями тысяч процессоров, а задач, масштабируемых на такое количество процессоров, немного, то ограничение $P(N \leq N_{\max}) = 1$ для некоторого $N_{\max} \ll m$ представляется вполне мотивированным. Обозначим

$$j = \min \left\{ k \geq 1 : P \left(N \leq \left\lfloor \frac{m}{k} \right\rfloor \right) = 1 \right\}, \quad N_{\max} = \left\lfloor \frac{m}{j} \right\rfloor. \quad (7)$$

(Здесь $\lfloor x \rfloor$ означает наибольшее целое число, не превосходящее x .) Случай $j = 1$ (когда каждой заявке разрешено занимать все процессоры) типичен для небольших ВК.

Как показано ниже, для исходной системы мажорантной будет система (в которой соответствующие величины снабжены верхним индексом up), где каждая заявка занимает ровно $N_{\max} = \lfloor m/j \rfloor$ процессоров. Для такой системы рекурсия (4) примет вид

$$\begin{aligned} W_{i+1}^{(\text{up})} &= R \left(W_i^{(\text{up})}(N_{\max}) + S_i - T_i, \dots \right. \\ &\dots, W_i^{(\text{up})}(N_{\max}) + S_i - T_i, W_i^{(\text{up})}(N_{\max} + 1) - T_i, \dots \\ &\dots, W_i^{(\text{up})}(m) - T_i \left. \right)^+. \quad (8) \end{aligned}$$

Теорема 3.3. Пусть $W_0 = W_0^{(\text{up})} = 0$. Тогда

$$W_{i+1} \leq W_{i+1}^{(\text{up})}, \quad i \geq 0. \quad (9)$$

Доказательство. Очевидно, при нулевых начальных условиях неравенство $W_1 \leq W_1^{(\text{up})}$ выполнено. Предположим, что $W_i \leq W_i^{(\text{up})}$ для некоторого $i > 1$. Рассмотрим рекурсии (4), (8) до применения оператора R и докажем индуктивный переход. Для компонент с номерами $1 \leq k \leq N_i$ имеем

$$(W_i(N_i) + S_i - T_i)^+ \leq (W_i^{(\text{up})}(N_{\max}) + S_i - T_i)^+.$$

Для компонент с номерами $N_i + 1 \leq k \leq N_{\max}$ получаем неравенства

$$(W_i(k) - T_i)^+ \leq (W_i^{(\text{up})}(N_{\max}) + S_i - T_i)^+,$$

а для компонент с номерами $k > N_{\max}$ выполнены неравенства

$$(W_i(k) - T_i)^+ \leq (W_i^{(\text{up})}(k) - T_i)^+.$$

В силу леммы 3.1 это влечет (9).

На самом деле каждую группу из N_{\max} процессоров можно считать одним процессором, так как все они занимаются и освобождаются заявкой одновременно. Таким образом, мажорирующая система эквивалентна стандартной системе обслуживания $GI/G/j$ (с теми же интервалами между приходами и временами обслуживания). Хорошо известное (достаточное) условие стационарности такой системы

$$\lambda ES < j \quad (10)$$

является, таким образом, также условием стационарности модели ВК. (При $m \rightarrow \infty$ условие (10) переходит в условие стационарности $\lambda ES < \infty$ модели с бесконечным числом процессоров для ординарного входного потока (см. (3)).)

Получим теперь достаточное условие неустойчивости (нестационарности) системы $\Sigma^{(\text{low})}$ (опустив для простоты в обозначениях индекс low). Обозначим через $A(t)$ число приходов в систему $\Sigma^{(\text{low})}$ в интервале $[0, t]$. (Это процесс восстановления с интенсивностью $\lambda = 1/ET$.) Очевидно, что $M(t) = \sum_{i=1}^{A(t)} N_i$ есть суммарное число процессоров, требуемых заявкам, поступившим в интервале $[0, t]$. Пусть $D(t)$ обозначает число уходов (освобожденных процессоров) в системе $\Sigma^{(\text{low})}$ в интервале $[0, t]$. Тогда $\nu(t) = M(t) - D(t)$ есть число процессоров, требуемых заявкам, находящимся в системе $\Sigma^{(\text{low})}$ в момент t . Обозначим $\rho = \lambda ENES$.

Лемма 3.4. Если $\rho > m$, то $\nu(t) \rightarrow \infty$ с вероятностью 1.

Доказательство. Обозначим через $\hat{D}(t)$ число уходов из системы $\Sigma^{(\text{low})}$ в интервале $[0, t]$ в предположении, что каждый процессор работает без простоев. Таким образом, $\{\hat{D}(t), t \geq 0\}$ есть суперпозиция m независимых процессов восстановления, каждый с интенсивностью $\mu = 1/ES$. Очевидно, $\hat{D}(t) \geq D(t)$, $t \geq 0$, и поэтому $\nu(t) \geq M(t) - \hat{D}(t)$. Кроме того,

$$\frac{M(t)}{t} = \frac{A(t)}{t} \frac{\sum_{i=1}^{A(t)} N_i}{A(t)} \rightarrow \lambda EN, \quad t \rightarrow \infty.$$

Из теории восстановления следует, что с вероятностью 1

$$\frac{\hat{D}(t)}{t} \rightarrow \mu m, \quad t \rightarrow \infty.$$

Поэтому

$$\liminf \frac{\nu(t)}{t} \geq \mu(\rho - m) > 0.$$

Таким образом, необходимым условием устойчивости минорантной системы $\Sigma^{(\text{low})}$, а следовательно, и исходной модели ВК является условие $\rho < m$.

Замечание. Полученный результат означает *сильную неустойчивость*, в отличие от *слабой неустойчивости*, когда неограниченный рост очереди происходит по вероятности. Различные виды неустойчивости процессов обслуживания рассматриваются, например, в [29–31]. В частности, в [30] с использованием регенеративного анализа показано, что минорирующая система слабо неустойчива при $\rho = m$.

По аналогии с работой [32] построим рекурсию для (скалярной) компоненты $D_i = W_i(N_i)$, являющейся задержкой (временем ожидания в очереди) заявки i . Пусть для системы ВК (4)

$$\left. \begin{aligned} P_i &:= W_i(N_i + N_{i+1}) - W_i(N_i), \\ Q_i &:= W_i(N_{i+1}) - W_i(N_i), \\ (P_i &:= \infty \text{ при } N_i + N_{i+1} > m), \\ U_i &:= \max(Q_i, \min(P_i, S_i)) = \\ &= \min(P_i, \max(Q_i, S_i)). \end{aligned} \right\} \quad (11)$$

Теорема 3.5. В модели ВК величина задержки удовлетворяет рекурсии

$$D_{i+1} = (D_i + U_i - T_i)^+, \quad i \geq 0. \quad (12)$$

Доказательство. В нижеследующем анализе нумерация процессоров соответствует их состоянию *перед приходом* заявки. Нетрудно увидеть, что для любого i

$$P_i \geq Q_i. \quad (13)$$

Рассмотрим возможные случаи.

1. Пусть $N_i + N_{i+1} \leq m$ и выполнено неравенство $(W_i(N_i + N_{i+1}) - T_i)^+ \leq (W_i(N_i) + S_i - T_i)^+$. Это означает, что в момент прихода заявки $i+1$ процессоры с номерами $1, \dots, N_i$ заняты заявкой i и для обслуживания заявки $i+1$ будут использованы процессоры с номерами $N_i+1, \dots, N_i+N_{i+1}$. Поэтому

$$D_{i+1} = (W_i(N_i + N_{i+1}) - T_i)^+, \quad i \geq 0. \quad (14)$$

Поскольку в данном случае $W_i(N_i + N_{i+1}) \leq W_i(N_i) + S_i$, то с учетом (13) выполнено неравенство $Q_i \leq P_i \leq S_i$. Поэтому с учетом (11) $U_i = P_i$ и, следовательно, (14) влечет (12).

2. Пусть теперь $N_{i+1} > N_i$ и $(W_i(N_{i+1}) - T_i)^+ > (W_i(N_i) + S_i - T_i)^+$. Это означает, что в момент прихода заявки $i + 1$ процессоры с номерами $1, \dots, N_i$ заняты заявкой i . Однако для заявки $i + 1$ этих процессоров недостаточно, и она ожидает освобождения самого загруженного (из требуемых ей) процессора с номером N_{i+1} . В этом случае

$$D_{i+1} = (W_i(N_{i+1}) - T_i)^+, \quad i \geq 0. \quad (15)$$

Тогда с учетом (13) выполнено $P_i \geq Q_i \geq S_i$. Следовательно, $U_i = Q_i$ и (15) снова влечет (12). Полезно отметить, что эти два рассмотренных случая несовместны, т. е.

$$\{W_i(N_{i+1}) > W_i(N_i) + S_i\} \cap \{W_i(N_i + N_{i+1}) < W_i(N_i) + S_i\} = \emptyset.$$

3. Наконец, рассмотрим ситуацию, когда ни один из рассмотренных выше случаев не имеет места. Тогда, как нетрудно понять, величина задержки определяется нагрузкой на любом из процессоров с номерами $1, \dots, N_i$, т. е.

$$D_{i+1} = (W_i(N_i) + S_i - T_i)^+ = (D_i + S_i - T_i)^+, \quad i \geq 0. \quad (16)$$

Легко проверить, что в данном случае $U_i = S_i$ и поэтому (16) влечет (12).

Заметим, что в классической системе GI/G/m $N_i \equiv 1$ и поэтому $Q_i = 0$, а условие $N_i + N_{i+1} \leq m$ всегда верно для $m \geq 2$. Поэтому $P_i = W_i(2) - W_i(1)$, $U_i = \min(P_i, S_i)$ и, как легко проверить, рекурсия из [32] оказывается частным случаем рекурсии (12). А поскольку в классической системе $D_i = W_i(1)$, то

$$D_{i+1} = \min((W_i(1) + S_i - T_i)^+, (W_i(2) - T_i)^+) = (D_i + U_i - T_i)^+.$$

Заметим, что доказанная в теореме 3.3 монотонность верна и для соответствующих моментов компонент вектора нагрузки в модели ВК. Обозначим через $W = (W_1, \dots, W_m)$ стационарный вектор нагрузки в модели ВК, т. е. предполагаем, что слабый предел $W_n \Rightarrow W$ существует.

Напомним обозначения (7) и обозначим также $k(i) = \lfloor i/N_{\max} \rfloor$, $1 \leq i \leq m$. Прямым следствием теоремы 3.3 являются следующие моментные свойства стационарного вектора нагрузки для модели ВК, полученные в работе [4] для системы GI/G/j.

Теорема 3.6. Пусть $\rho := ES/ET < j$ и $\alpha \geq 1$. Тогда имеют место следующие импликации:

1. Для компонент вектора W с индексами $1 \leq i \leq \lceil \rho \rceil N_{\max}$

$$ES^{1+\alpha/(j-\lfloor \rho \rfloor)} < \infty \Rightarrow EW_i^\alpha < \infty.$$

2. Для компонент вектора W с индексами $\lceil \rho \rceil N_{\max} < i \leq m$

$$ES^{1+\alpha/(j-k(i))} < \infty \Rightarrow EW_i^\alpha < \infty.$$

В случае $j = 1$, $N_{\max} = m$ все компоненты вектора загрузки имеют одинаковые моментные свойства, а именно: условие $ES^{\alpha+1} < \infty$ влечет $EW_i^\alpha < \infty$ для $1 \leq i \leq m$. (Это классический результат из работы [33].) Следствием теоремы 3.6 являются моментные свойства стационарной задержки заявки в модели ВК.

Теорема 3.7. Пусть выполнены условия теоремы 3.6. Тогда

$$ES^{1+\alpha/(j-\lfloor \rho \rfloor)} < \infty \Rightarrow ED^\alpha < \infty.$$

Доказательство. Заметим, что первые $\lceil \rho \rceil N_{\max}$ компонент вектора нагрузки имеют одинаковые моментные свойства. Далее, поскольку $\lceil \rho \rceil N_{\max} \geq N_{\max}$ при любом $\rho > 0$, то условие $P(N \leq N_{\max}) = 1$, следующее из (7), гарантирует, что задержка $D_i := W_i(N_i)$ окажется среди первых $\lceil \rho \rceil N_{\max}$ координат вектора нагрузки. В то же время условие $ED^\alpha < \infty$ для этих компонент следует из теоремы 3.6.

4 Заключение

В статье дан краткий обзор основных моделей МС, в которых заявке требуется для обслуживания случайное число процессоров. Кроме того, приведена новая модель ВК на основе модифицированной рекурсии Кифера–Вольфовица. Эта модель исследована с помощью классических многоканальных систем, одна из которых является минорантной, а другая мажорантной для процесса нагрузки в модели ВК. На основе этого подхода, в частности, получены достаточные условия стационарности, условие (сильной) неустойчивости, а также моментные свойства компонент стационарного вектора нагрузки в модели ВК.

На основе данных лог-файла запусков задач кластера ЦКП КарНЦ РАН [34] была проведена апробация модели ВК. (Анализ некоторых численных результатов, относящихся к данной модели, представлен также в работах [9, 35, 36].) Эксперименты проводились с использованием разработанного

авторами пакета расширения для вычислительной среды R. Сформулируем краткие выводы из проведенных исследований, которые согласуются с результатами авторитетных исследований в данной области [1]. В частности, обнаружено, что интервалы между приходами заявок хорошо описываются с помощью лог-нормального распределения (*с тяжелым хвостом*). Для моделирования времени обслуживания заявок использовалось усеченное распределение Парето, что соответствует принятой практике [2]. Обнаружена зависимость времени обработки S_i от числа требуемых процессоров N_i (что, однако, не отражено в рассматриваемой модели ВК). Кроме того, (визуально) обнаружено медленное убывание автокорреляционной функции последовательности N_i , что может отражать наличие тяжелого хвоста у распределения с. в. N . В ходе исследования также обнаружено доминирование значений $N_i = 2^k$ для $k = 0, \dots, 8$ (около 87%). (В этой связи укажем работы [3, 37].) Наконец отметим, что почти 50% всех задач были однопроцессорными. В целом модель показала хорошее согласие с экспериментальными данными лог-файла, что говорит об определенном потенциале ее практического применения для анализа существующих и проектирования новых МС.

Литература

1. Talby D., Feitelson D., Raveh A. A co-plot analysis of logs and models of parallel workloads // ACM Transactions on Modeling and Computer Simulation (TOMACS), 2007. Vol. 17. No. 3. Article 12.
2. Gupta V., Harchol-Balter M., Dai J.G., Zwart B. On the inapproximability of M/G/K: Why two moments of job size distribution are not enough // Queueing Syst., 2010. Vol. 64. P. 5–48.
3. Krampe A., Lepping J., Sieben W. A hybrid Markov chain modeling architecture for workload on parallel computers // HPDC'10: 19th ACM Symposium (International) on High Performance Distributed Computing Proceedings. — New York: ACM, 2010. P. 589–596.
4. Scheller-Wolf A., Vesilo R. Sink or swim together: Necessary and sufficient conditions for finite moments of workload components in FIFO multiserver queues // Queueing Syst., 2011. Vol. 67. No. 1. P. 47–61.
5. Foss S., Korshunov D. Heavy tails in multi server queue // Queueing Syst., 2006. Vol. 52. No. 1. P. 31–48.
6. Harchol-Balter M., Schroeder B. Evaluation of task assignment policies for supercomputing servers: The case for load unbalancing and fairness // HPDC'00: 9th IEEE Symposium on High Performance Distributed Computing Proceedings. — New York: ACM, 2001. P. 211–219.
7. Морозов Е. В., Румянцев А. С. Некоторые модели многопроцессорных систем обслуживания с тяжелыми хвостами // Параллельные вычислительные технологии 2011: Сборник трудов междунар. научн. конф. — Челябинск: ЮУрГУ, 2011. С. 555–566.
8. Румянцев А. С. О стохастическом моделировании вычислительного кластера // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем: Тезисы докладов Всеросс. конф. с международным участием (18–22 апреля 2011). — М.: РУДН, 2011. С. 46–47.
9. Морозов Е. В., Румянцев А. С. Модели многосерверных систем для анализа вычислительного кластера // Труды Карельского научного центра Российской академии наук, 2011. № 5. С. 75–86.
10. Green L. Comparing operating characteristics of queues in which customers require a random number of servers // Management Sci., 1980. Vol. 27. No. 1. P. 65–74.
11. Kim S. M/M/s queueing system where customers demand multiple server use. Ph.D. Dissertation. — Southern Methodist University, 1979.
12. Kaufman J. Blocking in a shared resource environment // IEEE Transactions on Communications, 1981. Vol. 29. No. 10. P. 1474–1481.
13. Whitt W. Blocking when service is required from several facilities simultaneously // AT&T Techn. J., 1985. Vol. 64. No. 8. P. 1807–1856.
14. Van Dijk N., Smeitink E. A non-exponential queueing system with batch servicing // Research memorandum. — Amsterdam: Free University, 1988. No. 13.
15. Green L. A queueing system in which customers require a random number of servers // Operations Res., 1980. Vol. 28. No. 6. P. 1335–1346.
16. Brill P., Green L. Queues in which customers receive simultaneous service from a random number of servers: A system point approach // Management Sci., 1984. Vol. 30. No. 1. P. 51–68.
17. Боровков А. А. Вероятностные процессы в теории массового обслуживания. — М.: Наука, 1972.
18. Гнеденко Б. В., Коваленко И. Н. Введение в теорию массового обслуживания. — М.: Наука, 1987.
19. Liu L., Templeton J. Autocorrelations in infinite server batch arrival queues // Queueing Syst., 1993. Vol. 14. P. 313–337.
20. Тихоненко О. М. Модели массового обслуживания в системах обработки информации. — Минск: Университетское, 1990.
21. Eliazar I. The M/G/∞ system revisited: Finiteness, summability, long range dependence, and reverse engineering // Queueing Syst., 2007. Vol. 55. P. 71–82.
22. Daley D. The busy period of the M/GI/∞ queue // Queueing Syst., 2001. Vol. 38. P. 195–204.
23. Brandt A., Sulanke H. On the GI/M/∞ queue with batch arrivals of constant size // Queueing Syst., 1987. Vol. 2. P. 187–200.
24. Kiefer J., Wolfowitz J. On the theory of queues with many servers // Trans. Amer. Math. Soc., 1955. Vol. 78. P. 1–18.
25. Ширяев А. Н. Вероятность. — М.: Наука, 1989. 640 с.

26. *Jacobs D. R., Schach S.* Stochastic order relationships between $GI/G/k$ queues // *Ann. Math. Stat.*, 1972. Vol. 43. P. 1623–1633.
27. *Scheller-Wolf A.* Further delay moment results in FI-FO multiserver queues // *Queueing Syst.*, 2000. Vol. 34. P. 387–400.
28. *Morozov E. V.* Coupling and monotonicity of queues. Sci. Report. No. 779. — Barcelona: CRM, 2008. P. 1–29.
29. *El-Taha M.* Pathwise rate-stability for input-output processes // *Queueing Syst.*, 1996. Vol. 22. P. 47–63.
30. *Morozov E.* Instability of nonhomogeneous queueing networks // *J. Math. Sci.*, 2002. Vol. 112. No. 2. P. 4155–4167.
31. *Morozov E.* Stability of Jackson-type network output // *Queueing Syst.*, 2002. Vol. 40. P. 383–406.
32. *Scheller-Wolf A., Sigman K.* Delay moments for FIFO $GI/GI/s$ queues // *Queueing Syst.*, 1997. Vol. 25. P. 77–95.
33. *Kiefer J., Wolfowitz J.* On the characteristics of the general queueing process applications to random walks // *Ann. Math. Statist.*, 1956. Vol. 27. P. 147–161.
34. Центр высокопроизводительной обработки данных. — ЦКП КарНЦ РАН. <http://cluster.krc.karelia.ru>.
35. *Morozov E. V., Rumyantsev A. S.* Stability analysis of a multiprocessor model describing a high performance cluster // *Applied problems in theory of probabilities and mathematical statistics related to modeling of information systems: Book of Abstracts of the 29th Seminar (International) on Stability Problems for Stochastic Models and 5th Workshop (International)*. — Moscow: Institute of Informatics Problems, RAS, 2011. P. 82–83.
36. *Румянцев А. С.* Моделирование процесса нагрузки вычислительного кластера на примере кластера ЦКП КарНЦ РАН «Центр высокопроизводительной обработки данных» // *Высокопроизводительные параллельные вычисления на кластерных системах: Мат-лы XI Всеросс. конф.* / Под ред. В.П. Герге-ля. — Нижний Новгород: Изд-во Нижегородского госуниверситета, 2011. С. 272–275.
37. *Downey A., Feitelson D.* The elusive goal of workload characterization // *Performance Evaluation Rev.*, 1999. Vol. 26. No. 4. P. 14–29.

ОГРАНИЧЕНИЕ НА СУММАРНЫЙ ОБЪЕМ ЗАЯВОК В ДИСКРЕТНОЙ СИСТЕМЕ $\text{Geo}/G/1/\infty^*$

А. В. Печинкин¹, И. А. Соколов², С. Я. Шоргин³

Аннотация: Рассматривается функционирующая в дискретном времени однолинейная система массового обслуживания $\text{Geo}/G/1$ с инверсионным порядком обслуживания без прерывания обслуживания, в которой каждая заявка наряду с (дискретной) случайной длиной имеет также (дискретный) случайный объем. Суммарный объем находящихся в системе заявок ограничен некоторым (неслучайным) числом. Получены алгоритмы, позволяющие вычислять основные стационарные показатели функционирования этой системы.

Ключевые слова: система массового обслуживания; дискретное время; длина и объем заявки

1 Введение. Описание системы

Задача исследования систем массового обслуживания (СМО), в которых каждая поступающая в систему заявка наряду со случайной длиной имеет случайный объем, причем суммарный объем всех находящихся в системе заявок ограничен, как было замечено еще в работах [1–3], играет важную роль при моделировании работы самых разнообразных технических устройств, в частности современных информационно-вычислительных систем. Однако аналитических решений этой задачи при дисциплине выбора заявок из очереди на обслуживание в порядке поступления (FIFO) до сих пор не найдено, поскольку для корректного построения соответствующего марковского процесса, описывающего функционирование СМО, необходимо учитывать объемы всех заявок в системе. Фактически приходится сталкиваться с теми же самыми трудностями, что и при исследовании многолинейных СМО, для которых также не найдено удовлетворительных аналитических решений.

В работах [4–11] были исследованы СМО с ограничением на суммарный объем заявок, но при инверсионном порядке обслуживания (дисциплина LIFO). Оказалось, что в этом случае можно получить алгоритмы, пригодные для численных расчетов стационарных характеристик.

Тем не менее во всех этих работах, в том числе и в [9–11], где рассматривались СМО в дискретном времени, распределение объема заявки предполагалось непрерывным, а тогда алгоритмы расчета опирались на решения довольно сложных интегральных уравнений, что снижало практическую

значимость полученных результатов. В рассматриваемой в настоящей статье СМО $\text{Geo}_m/G/1$ с ограничением на суммарный объем находящихся в ней заявок, в отличие от цитированных выше работ, объем каждой заявки является дискретной случайной величиной. Это позволяет получить более простые и эффективные алгоритмы расчета основных стационарных показателей функционирования.

Рассмотрим функционирующую в дискретном времени однолинейную СМО $\text{Geo}_m/G/1$, в которую поступает геометрический поток заявок с вероятностью a поступления заявки на такте.

Каждая поступающая в систему заявка наряду с длиной имеет случайный целочисленный (неотрицательный) объем. Совместное распределение длины и объема заявки задается вероятностью $b_{k,l}$, $k, l \geq 0$, того, что ее длина (число тактов обслуживания) равна k , а объем равен l . Будем предполагать выполненным естественное условие, что длина заявки и ее объем не могут равняться нулю, т. е. $b_{k,0} = b_{0,l} = 0$ для любых k и l .

Общий объем находящихся в системе заявок ограничен (неслучайным) числом L , $0 < L < \infty$. Если объем поступающей в систему заявки в сумме с объемами остальных находящихся в системе заявок больше L , то она теряется. Будем предполагать, что если в момент поступления новой заявки систему покидает обслуженная заявка, то ее объем при определении суммарного объема не учитывается.

В системе реализован инверсионный порядок обслуживания без прерывания обслуживания, при котором принятая в систему заявка становится на

*Работа выполнена при поддержке РФФИ (гранты 11-07-00112, 12-07-00108 и 11-01-12026-офи-м).

¹Институт проблем информатики Российской академии наук, apchinkin@ipiran.ru

²Институт проблем информатики Российской академии наук, isokolov@ipiran.ru

³Институт проблем информатики Российской академии наук, sshorgin@ipiran.ru

первое место в очереди. Будем считать для определенности, что если в момент поступления новой заявки систему покидает обслуженная заявка, то на прибор становится новая заявка.

Введем также обозначения:

$b(l) = \sum_{k=1}^{\infty} b_{k,l}$, $l \geq 1$ — вероятность того, что объем заявки равен l ;

$B(l) = \sum_{j=1}^l b(j) = \sum_{j=1}^l \sum_{k=1}^{\infty} b_{k,j}$, $l \geq 1$ — вероятность того, что объем заявки не более l ;

$b(k|l) = b_{k,l}/b(l)$, $k, l \geq 1$ — условная вероятность того, что длина заявки равна k , при условии, что ее объем равен l ;

$B(k|l) = \sum_{i=k}^{\infty} b(i|l)$, $k, l \geq 1$ — условная вероятность того, что длина заявки не менее k , при условии, что ее объем равен l ;

$\beta(z|l) = \sum_{k=1}^{\infty} z^k b(k|l)$, $l \geq 1$ — производящая функция (ПФ) длины заявки при условии, что ее объем равен l ;

$\beta^*(z|l) = \sum_{k=1}^{\infty} z^{k-1} B(k+1|l) = [z - \beta(z|l)]/[z(1 - z)]$, $l \geq 1$;

$\bar{m} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} k b_{k,l}$ — математическое ожидание длины заявки.

Далее будем предполагать, что выполнено условие $\bar{m} < \infty$. Это условие является необходимым и достаточным для существования стационарного режима функционирования системы. Кроме того, чтобы избежать непринципиальных трудностей в изложении, будем считать, что объем заявки, во-первых, не превосходит L , а во-вторых, с ненулевой вероятностью может принимать значение единица.

2 Стационарные вероятности состояний

Обозначим через p_0 стационарную вероятность отсутствия заявок в системе, а через $p_{k,i}(l_1, \dots, l_i)$, $i \geq 1$, $k \geq 1$, $l_1, \dots, l_i \geq 1$, — стационарную вероятность того, что в системе находится i заявок, причем (обслуженная) длина и объем заявки на приборе равны k и l_1 , а объемы остальных находящихся в системе заявок равны (в порядке очереди) l_2, \dots, l_i .

Заметим, что поскольку объем каждой заявки — целое положительное число, то суммарный объем $l_1 + \dots + l_i$ заявок в системе не может быть меньше числа i заявок, т. е. неравенство $p_{k,i}(l_1, \dots, l_i) > 0$ может выполняться только при

$$i \leq l_1 + \dots + l_i \leq L, \quad i \geq 1. \quad (1)$$

Поэтому всюду далее, не оговаривая этого особо, будем предполагать, что условие (1) выполнено.

Используя метод исключения состояний [12, с. 22], получаем систему уравнений:

$$p_{k,1}(l) = [1 - aB(L-l)] \frac{B(k+1|l)}{B(k|l)} p_{k-1,1}(l), \quad k \geq 1, \quad (2)$$

$$\begin{aligned} p_{k,i}(l_1, \dots, l_i) &= [1 - aB(L-l_1 - \dots \\ &\dots - l_i)] \frac{B(k+1|l_1)}{B(k|l_1)} p_{k-1,i}(l_1, \dots, l_i) + \\ &+ ab(l_2) \frac{B(k+1|l_1)}{B(k|l_1)} p_{k-1,i-1}(l_1, l_3, \dots, l_i), \end{aligned} \quad i \geq 2, \quad k \geq 1, \quad (3)$$

с начальными условиями

$$\begin{aligned} p_{0,1}(l) &= \\ &= p_0 ab(l) + ab(l) \sum_{s=1}^{L-l} \sum_{k=1}^{\infty} \frac{B(k+1|s)}{B(k|s)} p_{k-1,1}(s) + \\ &+ ab(l) \sum_{s=1}^L \sum_{k=1}^{\infty} \frac{b(k|s)}{B(k|s)} p_{k-1,1}(s); \end{aligned} \quad (4)$$

$$\begin{aligned} p_{0,i}(l_1, \dots, l_i) &= ab(l_1) \times \\ &\times \sum_{l=1}^{L-l_1-\dots-l_i} \sum_{k=1}^{\infty} \frac{B(k+1|l)}{B(k|l)} p_{k-1,i}(l, l_2, \dots, l_i) + \\ &+ ab(l_1) \sum_{l=1}^{L-l_2-\dots-l_i} \sum_{k=1}^{\infty} \frac{b(k|l)}{B(k|l)} p_{k-1,i}(l, l_2, \dots, l_i), \end{aligned} \quad i \geq 2. \quad (5)$$

Из уравнений (2) и (4) получаем:

$$p_{k,1}(l) = [1 - aB(L-l)]^k B(k+1|l) p_{0,1}(l); \quad k \geq 1; \quad (6)$$

$$\begin{aligned} p_{0,1}(l) &= p_0 ab(l) + \\ &+ ab(l) \sum_{s=1}^{L-l} \beta^*([1 - aB(L-s)] | s) p_{0,1}(s) + \\ &+ ab(l) \sum_{s=1}^L \beta([1 - aB(L-s)] | s) p_{0,1}(s). \end{aligned} \quad (7)$$

Из уравнения (3) имеем:

$$\begin{aligned} p_{k,i}(l_1, \dots, l_i) &= \\ &= [1 - aB(L-l_1 - \dots - l_i)]^k B(k+1|l_1) p_{0,i}(l_1, \dots, l_i) + \\ &+ r_{k,i}(l_1, \dots, l_i), \quad i \geq 2, \quad k \geq 1, \end{aligned} \quad (8)$$

где

$$r_{0,i}(l_1, \dots, l_i) = 0, \quad i \geq 2, \quad (9)$$

$$\begin{aligned} r_{k,i}(l_1, \dots, l_i) = & [1 - aB(L - l_1 - \dots \\ & \dots - l_i)] \frac{B(k+1|l_1)}{B(k|l_1)} r_{k-1,i}(l_1, \dots, l_i) + \\ & + ab(l_2) \frac{B(k+1|l_1)}{B(k|l_1)} p_{k-1,i-1}(l_1, l_3, \dots, l_i), \\ & i \geq 2, \quad k \geq 1, \quad (10) \end{aligned}$$

а из уравнения (5) находим:

$$\begin{aligned} p_{0,i}(l_1, \dots, l_i) = & ab(l_1) \sum_{l=1}^{L-l_1-\dots-l_i} \beta^* \times \\ & \times ([1 - aB(L - l - l_2 - \dots - l_i)] | l) p_{0,i}(l, l_2, \dots, l_i) + \\ & + ab(l_1) \sum_{l=1}^{L-l_2-\dots-l_i} \beta([1 - aB(L - l - l_2 - \dots - l_i)] | l) \times \\ & \times p_{0,i}(l, l_2, \dots, l_i) + ab(l_1) \times \\ & \times \sum_{l=1}^{L-l_1-\dots-l_i} \sum_{k=1}^{\infty} \frac{B(k+1|l)}{B(k|l)} r_{k-1,i}(l, l_2, \dots, l_i) + \\ & + ab(l_1) \sum_{l=1}^{L-l_2-\dots-l_i} \sum_{k=1}^{\infty} \frac{b(k|l)}{B(k|l)} r_{k-1,i}(l, l_2, \dots, l_i), \\ & i \geq 2. \quad (11) \end{aligned}$$

Соотношения (6)–(11) позволяют последовательно, начиная с $i = 1$ и кончая $i = L$, с точностью до p_0 определять вероятности $p_{k,i}(l, l_1, \dots, l_{i-1})$.

Вероятность p_0 , как обычно, определяется из условия нормировки.

Однако нахождение $p_{k,i}(l, l_1, \dots, l_{i-1})$ из-за большой размерности вычисляемых вероятностей невозможно уже при совсем небольших значениях i даже для современной вычислительной техники. Поэтому в следующем разделе будет определено более простое маргинальное распределение и приведены формулы для его расчета, а в разд. 4 описан удобный алгоритм вычислений.

3 Маргинальное распределение стационарных вероятностей состояний

С точки зрения практики вполне достаточно знать не совместное распределение объемов всех находящихся в системе заявок, а только лишь совместное распределение объема обслуживаемой на приборе заявки и суммарного объема остальных находящихся в системе заявок. Более того, в системе без ограничения на число находящихся в ней

заявок при инверсионном порядке обслуживания число заявок в системе также не представляет особого интереса. Поэтому обозначим через

$$\begin{aligned} p_k(l, m) = & \sum_{i=1}^{\infty} \sum_{l_1+\dots+l_{i-1}=m} p_{k,i}(l, l_1, \dots, l_{i-1}), \\ & k \geq 0, \quad m \geq 0, \quad l = \overline{L-m}, \end{aligned}$$

стационарную вероятность того, что (обслуженная) длина и объем заявки на приборе равны k и l , а суммарный объем остальных находящихся в системе заявок равен m (напомним, что в силу условия (1) обязательно должно выполняться двойное неравенство $i-1 \leq m \leq L-l$). Примем соглашение, что значение $m = 0$ соответствует отсутствию заявок в накопителе ($i = 1$).

Из соотношений (6)–(10) имеем:

$$\begin{aligned} p_k(l, m) = & [1 - aB(L - l - m)]^k B(k+1|l) p_0(l, m) + \\ & + r_k(l, m), \quad k \geq 1, \quad m \geq 0, \quad (12) \end{aligned}$$

где

$$r_k(l, 0) = 0, \quad k \geq 1; \quad (13)$$

$$r_0(l, m) = 0, \quad m \geq 1; \quad (14)$$

$$\begin{aligned} r_k(l, m) = & [1 - aB(L - l - m)] \frac{B(k+1|l)}{B(k|l)} r_{k-1}(l, m) + \\ & + a \frac{B(k+1|l)}{B(k|l)} \sum_{s=1}^{m-1} b(s) p_{k-1}(l, m-s), \\ & k \geq 1, \quad m \geq 1, \quad (15) \end{aligned}$$

а из соотношений (7) и (11) находим:

$$\begin{aligned} p_0(l, m) = & q(l, m) + \\ & + ab(l) \sum_{s=1}^{L-l-m} \beta^*([1 - aB(L - s - m)] | s) p_0(s, m) + \\ & + ab(l) \sum_{s=1}^{L-m} \beta([1 - aB(L - s - m)] | s) p_0(s, m), \\ & m \geq 0, \quad (16) \end{aligned}$$

где

$$q(l, 0) = p_0 ab(l); \quad (17)$$

$$\begin{aligned} q(l, m) = & ab(l) \sum_{s=1}^{L-l-m} \sum_{k=1}^{\infty} \frac{B(k+1|s)}{B(k|s)} r_{k-1}(s, m) + \\ & + ab(l) \sum_{s=1}^{L-m} \sum_{k=1}^{\infty} \frac{b(k|s)}{B(k|s)} r_{k-1}(s, m), \quad m \geq 1. \quad (18) \end{aligned}$$

4 Алгоритм решения системы уравнений

В этом разделе приведем простой алгоритм численного решения систем линейных алгебраических уравнений (12)–(18), который состоит в последовательном по m от $m = 0$ до $m = L - 1$ вычислении стационарных вероятностей $p_k(l, m)$.

Начнем с определения $p_k(l, 0) = p_{k,1}(l)$.

Для сокращения записи введем обозначения:

$$x_l = p_0(l, 0), \quad l = \overline{1, L};$$

$$b_l = q(l, 0) = ab(l), \quad l = \overline{1, L};$$

$$\beta_l^* = \beta^*([1 - aB(L - l)] | l), \quad l = \overline{1, L};$$

$$\beta_l = \beta([1 - aB(L - l)] | l), \quad l = \overline{1, L};$$

$$y_l = \sum_{s=1}^{L-l} \beta^*([1 - aB(L - s)] | s) p_0(s, 0), \quad l = \overline{1, L};$$

$$y = \sum_{s=1}^L \beta([1 - aB(L - s)] | s) p_0(s, 0).$$

Тогда систему (16) можно записать следующим образом:

$$x_l = b_l p_0 + b_l y_l + b_l y, \quad l = \overline{1, L}; \quad (19)$$

$$y_l = \sum_{s=1}^{L-l} \beta_s^* x_s, \quad l = \overline{1, L-1}; \quad y_L = 0; \quad (20)$$

$$y = \sum_{s=1}^L \beta_s x_s. \quad (21)$$

Заметим, что при $l = \overline{2, L-1}$ каждое y_l представимо в виде:

$$y_l = y_{l-1} - \beta_{L-l+1}^* x_{L-l+1}, \quad (22)$$

а при $l = \overline{1, L-2}$ — в виде:

$$y_l = y_{l+1} + \beta_{L-l}^* x_{L-l}. \quad (23)$$

Сначала выразим x_l , $l = \overline{1, L}$, и y_l , $l = \overline{2, L-1}$, через y_1 и y по формулам:

$$x_l = c_l + d_l y_1 + e_l y, \quad l = \overline{1, L};$$

$$y_l = f_l + g_l y_1 + h_l y, \quad l = \overline{2, L-1}.$$

Подставляя в (19) $l = L$, имеем:

$$x_L = b_L p_0 + b_L y,$$

т. е.

$$c_L = b_L p_0; \quad d_L = 0; \quad e_L = b_L.$$

При $l = 1$ из (19) и (20) находим:

$$x_1 = b_1 p_0 + b_1 y_1 + b_1 y; \quad y_{L-1} = b_1^* x_1;$$

$$c_1 = b_1 p_0; \quad d_1 = b_1; \quad e_1 = b_1; \quad f_{L-1} = b_1^* c_1;$$

$$g_{L-1} = b_1^* d_1; \quad h_{L-1} = b_1^* e_1.$$

При $l = L - 1$ из (19) и (22) получаем:

$$x_{L-1} = b_{L-1} p_0 + b_{L-1} y_{L-1} + b_{L-1} y;$$

$$y_2 = y_1 - \beta_{L-1}^* x_{L-1};$$

$$c_{L-1} = b_{L-1} p_0 + b_{L-1} f_{L-1}; \quad d_{L-1} = b_{L-1} g_{L-1};$$

$$e_{L-1} = b_{L-1} h_{L-1} + b_{L-1};$$

$$f_2 = -b_{L-1}^* c_{L-1}; \quad g_2 = 1 - b_{L-1}^* d_{L-1};$$

$$h_2 = -b_{L-1}^* e_{L-1}.$$

При $l = 2$ из (19) и (23) имеем:

$$x_2 = b_2 p_0 + b_2 y_2 + b_2 y; \quad y_{L-2} = y_{L-1} + \beta_2^* x_2;$$

$$c_2 = b_2 p_0 + b_2 f_2; \quad d_2 = b_2 g_2; \quad e_2 = b_2 h_2 + b_2;$$

$$f_{L-2} = f_{L-1} - b_2^* c_2; \quad g_{L-2} = g_{L-1} - b_2^* d_2;$$

$$h_{L-2} = h_{L-1} - b_2^* e_2.$$

Продолжая эту процедуру, получаем, что x_l при четном l в пределах от $l = L/2 + 1$ до $L - 2$ и при нечетном l в пределах от $l = (L + 1)/2$ до $L - 2$ вычисляются по формулам:

$$x_{L-s} = b_{L-s} p_0 + b_{L-s} y_{L-s} + b_{L-s} y;$$

$$y_{s+1} = y_s - \beta_{L-s}^* x_{L-s};$$

$$c_{L-s} = b_{L-s} p_0 + b_{L-s} f_{L-s}; \quad d_{L-s} = b_{L-s} g_{L-s};$$

$$e_{L-s} = b_{L-s} h_{L-s} + b_{L-s};$$

$$f_{s+1} = f_s - b_{L-s}^* c_{L-s}; \quad g_{s+1} = g_s - b_{L-s}^* d_{L-s};$$

$$h_{s+1} = h_s - b_{L-s}^* e_{L-s},$$

а при четном l в пределах от $l = 3$ до $L/2$ и при нечетном l в пределах от $l = 1$ до $(L - 1)/2$ — по формулам:

$$x_s = b_s p_0 + b_s y_s + b_s y; \quad y_{L-s} = y_{L-s+1} + \beta_s^* x_s;$$

$$c_s = b_s p_0 + b_s f_s; \quad d_s = b_s g_s; \quad e_s = b_s h_s + b_s;$$

$$f_{L-s} = f_{L-s+1} - b_s^* c_s; \quad g_{L-s} = g_{L-s+1} - b_s^* d_s;$$

$$h_{L-s} = h_{L-s+1} - b_s^* e_s.$$

Подставляя найденные значения x_l в равенство (20) при $l = 1$ и в равенство (21), приходим к системе из двух линейных алгебраических уравнений относительно y_1 и y , решая которую, находим эти величины, затем стационарные вероятности $x_l = p_0(l, 0)$, $l = \overline{1, L}$, и далее с помощью формул (12) и (13) стационарные вероятности $p_k(l, 0)$, $l = \overline{1, L}$, $k \geq 1$.

Предположим теперь, что вероятности $p_k(l, j)$, $k \geq 0$, $l = \overline{1, L-j}$, $j = \overline{j-1, L}$, уже найдены для всех $j = \overline{0, m-1}$, $m = \overline{0, L-1}$. Найдем эти вероятности для m .

Как и прежде, для сокращения записи введем обозначения:

$$x_l = p_0(l, m), \quad l = \overline{1, L-m};$$

$$b_l = ab(l), \quad l = \overline{1, L-m};$$

$$\beta_l^* = \beta^*([1 - aB(L - l - m)] | l), \quad l = \overline{1, L - m};$$

$$\beta_l = \beta([1 - aB(L - l - m)] | l), \quad l = \overline{1, L - m};$$

$$y_l = \sum_{s=1}^{L-l-m} \beta^*([1 - aB(L - s - m)] | s) p_0(s, m),$$

$$l = \overline{1, L - m};$$

$$y = \sum_{s=1}^{L-m} \beta([1 - aB(L - s - m)] | s) p_0(s, m);$$

$$\tilde{b}_l = q(l, m) =$$

$$= ab(l) \left[\sum_{s=1}^{L-l-m} \sum_{k=1}^{\infty} \frac{B(k+1|s)}{B(k|s)} r_{k-1}(s, m) + \right.$$

$$\left. + \sum_{s=1}^{L-m} \sum_{k=1}^{\infty} \frac{b(k|s)}{B(k|s)} r_{k-1}(s, m) \right], \quad l = \overline{1, L - m}.$$

При этом, как видно из (14), (15), (17) и (18), \tilde{b}_l выражается через уже известные величины.

С учетом введенных обозначений система (16) записывается следующим образом:

$$x_l = \tilde{b}_l + b_l y_l + b_l y, \quad l = \overline{1, L - m}; \quad (24)$$

$$y_l = \sum_{s=1}^{L-l-m} \beta_s^* x_s, \quad l = \overline{1, L - 1 - m}; \quad y_{L-m} = 0; \quad (25)$$

$$y = \sum_{s=1}^{L-m} \beta_s x_s. \quad (26)$$

Нетрудно видеть, что при каждом фиксированном m алгоритм решения системы (24)–(26) полностью совпадает с алгоритмом решения системы (19)–(21).

Оставшиеся неизвестными вероятности $p_k(l, m)$, $k \geq 1$, вычисляются по формуле (12).

Приведенный здесь алгоритм позволяет вычислить стационарные вероятности $p_{k,i}(l, m)$ с точностью до вероятности p_0 , которая, как уже говорилось, определяется из условия нормировки

$$p_0 + \sum_{k=0}^{\infty} \sum_{m=0}^{L-1} \sum_{l=1}^{L-m} p_k(l, m) = 1.$$

5 Некоторые стационарные показатели, связанные с числом заявок в системе

Выпишем выражения для некоторых стационарных характеристик, связанных со стационарными вероятностями состояний.

Стационарная вероятность $p(m)$, $m = \overline{1, L}$, того, что суммарный объем находящихся в системе заявок равен m , задается формулой

$$p(m) = \sum_{k=0}^{\infty} \sum_{l=1}^{m-1} p_k(l, m-l), \quad m = \overline{1, L}.$$

Стационарная вероятность p_0^* того, что в момент поступления новой заявки система будет свободна (в том числе на приборе закончится обслуживание единственной находящейся в системе заявки), задается формулой

$$p_0^* = p_0 + \sum_{k=1}^{\infty} \sum_{l=1}^L \frac{b(k|l)}{B(k|l)} p_{k-1}(l, 0).$$

Стационарная вероятность $p_k^*(l, m)$, $m = \overline{0, L-1}$, $l = \overline{1, L-m}$, $k \geq 1$, того, что поступающая (не обязательно принятая в систему) заявка застанет в системе, по крайней мере, одну заявку, причем длина и объем заявки на приборе равны k и l и она продолжит обслуживаться, а суммарный объем остальных находящихся в системе заявок равен m , определяется выражением:

$$p_k^*(l, m) = \frac{B(k+1|l)}{B(k|l)} p_{k-1}(l, m),$$

$$m = \overline{0, L-l}, \quad l = \overline{1, L-m}, \quad k \geq 1.$$

Стационарная вероятность $p^*(m)$, $m = \overline{1, L-1}$, того, что в момент поступления новой заявки на приборе закончится обслуживание заявки, а суммарный объем оставшихся заявок будет равен m , задается формулой

$$p^*(m) = \sum_{k=1}^{\infty} \sum_{l=1}^{L-m} \frac{b(k|l)}{B(k|l)} p_{k-1}(l, m), \quad m = \overline{1, L-1}.$$

Наконец, стационарная вероятность $\pi(l)$ того, что поступающая заявка объема l будет принята в систему, и стационарная вероятность π того, что поступающая заявка произвольной длины будет принята в систему, имеют вид:

$$\pi(l) = p_0^* + \sum_{k=1}^{\infty} \sum_{m=1}^{L-l} \sum_{j=1}^m p_k^*(j, m-j) +$$

$$+ \sum_{m=1}^{L-l} p^*(m), \quad l = \overline{1, L};$$

$$\pi = \sum_{l=1}^L b(l) p_0^*(l).$$

6 Стационарное распределение времени пребывания заявки в системе

Будем называть M -системой систему, аналогичную исходной, но с ограничением M , $1 \leq M \leq L$, на суммарный объем заявок и вероятностью a поступления заявки на такте. Нетрудно видеть, что M -система представляет собой исходную СМО, но при условии, что в ней постоянно находятся заявки суммарного объема $L - M$.

Обозначим через $g_i(k, l; M)$, $i \geq 1$, $k \geq 0$, $l = \overline{1, M}$, вероятность того, что период занятости (ПЗ) M -системы, открываемый заявкой (обслуженной) длины k и объема l , продлится i тактов. Для $g_i(k, l; M)$ справедливы следующие соотношения:

$$g_1(k, l; M) = [1 - aB(M)] \frac{b(k+1|l)}{B(k+1|l)}, \quad l = \overline{1, M}, \quad k \geq 0; \quad (27)$$

$$\begin{aligned} g_2(k, l; M) = & [1 - aB(M-l)] \frac{B(k+2|l)}{B(k+1|l)} g_1(k+1, l; M) + \\ & + a \frac{b(k+1|l)}{B(k+1|l)} \sum_{m=1}^M b(m) g_1(0, m; M), \\ & l = \overline{1, M}, \quad k \geq 0; \quad (28) \end{aligned}$$

$$\begin{aligned} g_i(k, l; M) = & [1 - aB(M-l)] \frac{B(k+2|l)}{B(k+1|l)} g_{i-1}(k+1, l; M) + \\ & + a \frac{B(k+2|l)}{B(k+1|l)} \sum_{m=1}^{M-l} b(m) \times \\ & \times \sum_{j=1}^{i-2} g_j(k+1, l; M-m) g_{i-1-j}(0, m; M) + \\ & + a \frac{b(k+1|l)}{B(k+1|l)} \sum_{m=1}^M b(m) g_{i-1}(0, m; M), \\ & l = \overline{1, M}, \quad i \geq 3, \quad k \geq 0. \quad (29) \end{aligned}$$

В последнем соотношении принято соглашение, что $\sum_{j=1}^0 (\cdot) = 0$.

Система уравнений (27)–(29) задает рекуррентную по M от $M = 1$ до $M = L$ процедуру определения вероятностей $g_i(k, l; M)$, которые из этих уравнений вычисляются при каждом M последовательно по i от $i = 1$ для всех возможных значений k и l .

Обозначим через $w_k(l)$, $l = \overline{1, L}$, $k \geq 0$, стационарную вероятность того, что заявка объема l будет принята в систему и будет ожидать начала обслуживания k тактов, а через $v_k(l)$, $l = \overline{1, L}$, $k \geq 1$, — стационарную вероятность того, что заявка объема l будет принята в систему и будет находиться в системе k тактов. Тогда

$$\begin{aligned} w_0(l) &= p_0^* + \sum_{m=1}^{L-l} p^*(m), \quad l = \overline{1, L}, \\ w_k(l) &= \sum_{i=1}^{\infty} \sum_{j=1}^{L-l} \sum_{m=0}^{L-l-j} p_i^*(j, m) g_k(i, j; L-m-l), \\ & \quad l = \overline{1, L}, \quad k \geq 1, \\ v_k(l) &= \sum_{i=0}^{k-l} w_i(l) b(k-i|l), \quad l = \overline{1, L}, \quad k \geq 1. \end{aligned}$$

Наконец, обозначая через w_k , $k \geq 0$ условную стационарную вероятность того, что заявка объема l , принятая в систему, будет ожидать начала обслуживания k тактов и через v_k , $k \geq 1$, — стационарную вероятность того, что эта заявка будет находиться в системе k тактов, имеем:

$$\begin{aligned} w_k &= \frac{1}{\pi} \sum_{l=1}^L b(l) w_k(l), \quad k \geq 0, \\ v_k &= \frac{1}{\pi} \sum_{l=1}^L b(l) v_k(l), \quad k \geq 1. \end{aligned}$$

7 Заключение

В настоящей статье получены математические соотношения, позволяющие вычислять основные стационарные характеристики функционирующей в дискретном времени СМО, в которой каждая заявка наряду с длиной (временем обслуживания) имеет (дискретный) случайный объем и суммарный объем находящихся в системе заявок ограничен. Приведены просто реализуемые алгоритмы для численных расчетов по этим соотношениям.

Литература

1. Ромм Э. Л., Скитович В. В. Об одном обобщении задачи Эрланга // Автоматика и телемеханика, 1971. № 6. С. 164–167.
2. Александров А. М., Кац Б. А. Обслуживание потоков неоднородных требований // Изв. АН СССР. Технич. кибернетика, 1973. № 2. С. 47–53.
3. Тихоненко О. М. Модели массового обслуживания в системах обработки информации. — Минск: Университетское, 1990.

4. Печинкин А. В., Печинкина О. А. Система $M_k/G/1/n$ с дисциплиной LIFO с прерыванием и ограничением на суммарный объем требований // Вестник Российского ун-та дружбы народов. Сер. Прикладная математика и информатика, 1996. № 1. С. 86–93.
5. Печинкин А. В. Система обслуживания с дисциплиной LIFO и ограничением на суммарный объем требований // Вестник Российского ун-та дружбы народов. Сер. Прикладная математика и информатика, 1996. № 2. С. 85–99.
6. Печинкин А. В. Система $M_1/G/1/n$ с дисциплиной LIFO и ограничением на суммарный объем требований // Автоматика и телемеханика, 1998. № 4. С. 106–116.
7. Абрамушкина Т. В., Апарина С. В., Кузнецова Е. Н., Печинкин А. В. Численные методы расчета стационарных вероятностей состояний системы $M/G/1/n$ с дисциплиной LIFO PR и ограничением на суммарный объем требований // Вестник Российского ун-та дружбы народов. Сер. Прикладная математика и информатика, 1998. № 1. С. 40–47.
8. Manzo R., Cascone A., Razumchik R. V. Exponential queueing system with negative customers and bunker for ousted customers // Automation Remote Control, 2008. Vol. 69. No. 9. P. 1542–1551.
9. Cascone A., Manzo R., Pechinkin A. V., Shorgin S. Ya. A $Geo_m/G/1/n$ queueing system with LIFO discipline, service interruptions and resumption, and restrictions on the total volume of demands // World Congress on Engineering 2010 Proceedings. Vol. III. WCE 2010. — London, U.K., 2010. P. 1765–1769. ISBN (Vol. III): 978-988-18210-8-9 ISSN: 2078-0958 (Print) ISSN: 2078-0966 (Online).
10. Pechinkin A., Shorgin S. A $Geo_m/G/1/n$ queueing system with LIFO discipline, service interruptions and repeat again service, and restrictions on the total volume of demands // Multiple Access Communication (MACOM 2010): Proceedings of the 3rd Workshop (International). — Barcelona, Spain, 2010. P. 98–106.
11. Cascone A., Manzo R., Pechinkin A. V., Shorgin S. Ya. $Geo_m/G/1/n$ system with LIFO discipline without interrupts and constrained total amount of customers // Automation Remote Control, 2011. Vol. 72, No. 1. P. 99–110.
12. Bocharov P. P., D'Apice C., Pechinkin A. V., Salerno S. Queueing theory. Modern probability and statistics ser. — Utrecht, Boston: VSP Publ., 2004.

АНАЛИЗ СИСТЕМЫ ОБСЛУЖИВАНИЯ С ГИПЕРЭКСПОНЕНЦИАЛЬНЫМ ВХОДЯЩИМ ПОТОКОМ В УСЛОВИЯХ КРИТИЧЕСКОЙ ЗАГРУЗКИ*

А. В. Ушаков¹

Аннотация: Найдены предельные распределения виртуального времени ожидания при критической загрузке в одноканальной системе обслуживания с относительным приоритетом и рекуррентным входящим потоком с гиперэкспоненциальным распределением интервалов между поступлениями требований.

Ключевые слова: виртуальное время ожидания; относительный приоритет; гиперэкспоненциальный поток; критическая загрузка

1 Введение

В данной статье изучено предельное поведение виртуального времени ожидания в одноканальной системе с гиперэкспоненциальным входящим потоком и относительным приоритетом в условиях критической загрузки. Рассмотрен случай, когда не существует момент второго порядка длительностей обслуживания требований, но существует момент порядка $1 < \gamma < 2$. Показано, что в зависимости от соотношения скоростей стремления времени к бесконечности и загрузки к единице возможны три различных предельных распределения. Доказательство предельной теоремы основано на результатах работы [1], в которой найдено нестационарное распределение времени ожидания, работ [2–5], в которых подробно изучены свойства решений ряда функциональных уравнений, а также асимптотических разложения этих решений, полученных в данной работе.

2 Описание системы. Основные обозначения и предположения

Рассматривается последовательность одноканальных систем обслуживания. Каждое поступившее в систему требование направляется в один из r , $r \geq 1$, приоритетных классов. В n -й системе (а) длительности обслуживания — независимые в совокупности и не зависящие от входящего потока случайные величины с функцией распределения $B_i^{(n)}(x)$ для требований i -го класса; (б) входящий поток требований — рекуррентный, определяемый

плотностью распределения интервалов между поступлениями требований вида

$$a^{(n)}(x) = \begin{cases} \sum_{j=1}^N c_j^{(n)} a_j^{(n)} \exp(-a_j^{(n)} x), & x \geq 0; \\ 0, & x < 0, \end{cases}$$

где $a_i^{(n)} \neq a_j^{(n)}$ при $i \neq j$, $c_j^{(n)} > 0$, $\sum_{i=1}^N c_i^{(n)} = 1$; (в) поступившее требование направляется в i -й приоритетный класс с вероятностью $p_i^{(n)}$, $i = 1, \dots, r$, независимо от остальных требований.

Будем предполагать, что системы обслуживания в рассматриваемой последовательности функционируют независимо друг от друга, требования из класса с меньшим номером имеют относительный приоритет перед требованиями из класса с большим номером. Требования из одного приоритетного класса обслуживаются в порядке их поступления в систему (дисциплина FIFO — first in, first out). Пусть, кроме того, в начальный момент $t = 0$ системы свободны от требований.

В дальнейшем изложении для сокращения записи будем опускать индекс n (номер в серии). При этом \lim будет означать $\lim_{n \rightarrow \infty}$. Предельные значения параметров входящего потока и длительностей обслуживания (функции распределения, моменты и т. п.) будем обозначать теми же символами, что и допредельные, но с дополнительным индексом «*». Пусть $b_i(x)$, $\beta_i(s)$, β_{ij} — соответственно плотность распределения, преобразование Лапласа–Стилтьеса и j -й момент случайной величины с функцией распределения $B_i(x)$,

* Работа выполнена при финансовой поддержке РФФИ, гранты 11-07-00112а, 12-07-00109а.

¹ Институт проблем информатики Российской академии наук; ushakov@akado.ru

$$a = \left(\sum_{j=1}^N c_j a_j^{-1} \right)^{-1}; \quad \rho_{k1} = a \cdot \sum_{i=1}^k p_i \beta_{i1};$$

$$\rho_k = 1 - \rho_{k1}; \quad \rho = \rho_r.$$

Введем случайный процесс $w(t)$ — виртуальное время ожидания для требований r -го приоритетного класса в момент времени t .

Положим

$$W_j(s, t) = \int_0^\infty e^{-sy} d_y \mathbf{P}(w(t) < y, j(t) = j),$$

$$\omega_j(s, v) = \int_0^\infty e^{-vt} W_j(s, t) dt.$$

Сделаем следующие предположения:

- (1) существует момент порядка $1 < \gamma < 2$ длительностей обслуживания требований всех приоритетов, причем справедливы разложения

$$\beta_i(s) = 1 - \beta_{i1}s + \hat{\beta}_{i\gamma}s^\gamma + o_n(s^\gamma), \quad (1)$$

где $(o_n(s^\gamma)/s^\gamma) \rightarrow 0$ при $s \rightarrow 0$ равномерно по n ;

- (2) для любого $n \geq 1$ $\rho_{r1} < 1$;
- (3) существуют пределы $\lim c_j = c_j^*$, $\lim a_j = a_j^*$, $j = 1, \dots, N$, $\lim \beta_{i1} = \beta_{i1}^*$, $\lim \hat{\beta}_{i\gamma} = \hat{\beta}_{i\gamma}^*$, $i = 1, \dots, r$, $\lim p_i = p_i^*$, $i = 1, \dots, r$, $\lim \rho_{r-11} < 1$, $\lim \rho_{r1} = 1$.

3 Предварительные результаты

Из результатов работы [1] вытекают следующие соотношения для нахождения распределения виртуального времени ожидания:

$$\omega_j(s, v) = \sum_{\nu=1}^N \sum_{k=1}^N \prod_{l \neq \nu} \frac{\mu_{k,r-1}^{(k)}(s) + a_l}{a_l} \times$$

$$\times \prod_{p \neq k} \frac{\mu_{p,r-1}^{(p)}(s)}{\mu_{p,r-1}^{(p)}(s) - \mu_{k,r-1}^{(k)}(s)} c_\nu a_j \left(\sum_{m=1}^{r-1} p_m z_{r-1,m}^{(k)}(s) + \right.$$

$$\left. + p_r \right) \sum_{c=1}^N \left(\prod_{q \neq j} \left(\mu_{c,r-1}^{(k)}(s) + a_q \right) / \left(\left(\mu_{c,r-1}^{(k)}(s) + \right. \right. \right.$$

$$\left. \left. + a_\nu \right) \alpha_c \left(z_{r-1,1}^{(k)}(s), \dots, z_{r-1,r-1}^{(k)}(s), 1 \right) \right) \times$$

$$\times \omega_{rj} \left(s - \mu_{c,r-1}^{(k)}(s), v \right), \quad (2)$$

где

$$\omega_{rj}(s, v) = \frac{c_j}{v - s + a_j} - \frac{s}{v - s + a_j} p_{0j}(v) +$$

$$+ \frac{c_j}{v - s + a_j} \sum_{k=1}^N a_k \omega_{rk}(s, v) \sum_{m=1}^r p_m \beta_m(s); \quad (3)$$

$$\left(1 - \sum_{j=1}^N \frac{c_j a_j}{v - s + a_j} \sum_{m=1}^r p_m \beta_m(s) \right) \sum_{k=1}^N a_k \omega_{rk}(s, v) =$$

$$= \sum_{j=1}^N \frac{c_j a_j}{v - s + a_j} - s \sum_{j=1}^N \frac{a_j p_{0j}(v)}{v - s + a_j}; \quad (4)$$

функции $p_{0k}(v)$, $k = 1, \dots, N$, определяются по формулам

$$a_k p_{0k}(v) = \sum_{l=1}^N \sum_{\nu=1}^N \frac{c_\nu a_\nu}{(v + \psi_{lr}(v))(a_\nu - \psi_{lr}(v))} \times$$

$$\times \frac{\prod_{j=1}^N ((a_k - \psi_{jr}(v))(a_j - \psi_{lr}(v)))}{(a_k - \psi_{lr}(v)) \prod_{n \neq l} (\psi_{nr}(v) - \psi_{lr}(v)) \prod_{i \neq k} (a_k - a_i)},$$

а функции $\psi_{li}(v)$, $l = 1, \dots, N$, $i = 0, \dots, r$, являются решениями (относительно s) уравнения

$$\sum_{j=1}^N \frac{c_j a_j}{a_j - s} \left(\sum_{m=1}^i p_m \beta_m(s + v) + \sum_{m=i+1}^r p_m \right) = 1.$$

Пусть далее $\alpha_k(z) = \prod_{i \neq k} (\mu_k(z) - \mu_i(z))$, а

$\mu_1(z), \dots, \mu_N(z)$ — корни многочлена $\prod_{i=1}^N (\mu + a_i) -$

$$- (p, z) \sum_{j=1}^N c_j a_j \prod_{i \neq j} (\mu + a_i),$$

$$z_{ij}^{(k)}(s) = \beta_j(s + \psi_{ki}(s));$$

$$\mu_{ci}^{(k)}(s) = \mu_c(z_{i1}^{(k)}(s), \dots, z_{ii}^{(k)}(s), 1, \dots, 1);$$

$$(p, z) = \sum_{m=1}^r p_m z_m.$$

Замечание 1. Легко показать, что только одна из функций $\psi_{li}(v)$, $l = 1, \dots, N$, обращается в нуль при $v = 0$ и только одна из функций $\mu_k(z)$, $k = 1, \dots, N$, обращается в нуль при $z = 1$. Не ограничивая общности, будем считать, что $\psi_{1i}(0) = 0$ и $\mu_1(1) = 0$.

4 Основные результаты

Для доказательства основной предельной теоремы понадобится асимптотика при $n \rightarrow \infty$ функций, определенных в предыдущем параграфе.

Лемма 1. *Справедливы следующие асимптотические разложения функции $\psi_{1r}(v)$:*

$$\begin{aligned} \psi_{1r}(v\rho^\alpha) = \\ = \begin{cases} \left(\frac{v}{u}\right)^{1/\gamma} \rho^{\alpha/\gamma} + o(\rho^{\alpha/\gamma}), & \alpha < \frac{\gamma}{\gamma-1}; \\ \psi(v)\rho^{1/(\gamma-1)} + o(\rho^{1/(\gamma-1)}), & \alpha = \frac{\gamma}{\gamma-1}; \\ v\rho^{\alpha-1} + o(\rho^{\alpha-1}), & \alpha > \frac{\gamma}{\gamma-1}, \end{cases} \quad (5) \end{aligned}$$

где $u = a \sum_{m=1}^r p_m \hat{\beta}_{m\gamma}$, а $\psi(v)$ — решение уравнения $u\psi^\gamma(v) + \psi(v) - v = 0$, удовлетворяющее условию $\psi(0) = 0$.

Доказательство. Так как

$$\sum_{j=1}^N \frac{c_j a_j}{a_j - \psi_{1r}(v\rho^\alpha)} \left(\sum_{m=1}^r p_m \beta_m (v\rho^\alpha + \psi_{1r}(v\rho^\alpha)) \right) = 1,$$

то, используя (1), имеем:

$$\begin{aligned} & \left(1 + a^{-1} \psi_{1r}(v\rho^\alpha) + \sum_{j=1}^N c_j a_j^{-2} \psi_{1r}(v\rho^\alpha)^2 + \right. \\ & \left. + o(\psi_{1r}(v\rho^\alpha)^2) \right) (1 - a^{-1} \rho_{r1}(v\rho^\alpha + \psi_{1r}(v\rho^\alpha)) + \\ & + a^{-1} u (v\rho^\alpha + \psi_{1r}(v\rho^\alpha))^\gamma + \\ & + o((v\rho^\alpha + \psi_{1r}(v\rho^\alpha))^\gamma)) = 1. \end{aligned}$$

Отсюда

$$\begin{aligned} -v\rho^\alpha + \rho\psi_{1r}(v\rho^\alpha) + u(v\rho^\alpha + \psi_{1r}(v\rho^\alpha))^\gamma = \\ = o((v\rho^\alpha + \psi_{1r}(v\rho^\alpha))^\gamma). \quad (6) \end{aligned}$$

Положим $\psi_{1r}(v\rho^\alpha) = \psi(v)\rho^\varphi$. Рассмотрим три случая.

1. Пусть $\alpha < \gamma/(\gamma-1)$. Тогда из (6) имеем $\varphi = \alpha/\gamma$ и $u\psi^\gamma(v) = v$.
2. Пусть $\alpha = \gamma/(\gamma-1)$. Тогда $\varphi = 1/(\gamma-1)$, а $\psi(v)$ является решением уравнения $u\psi^\gamma(v) + \psi(v) - v = 0$.
3. Пусть $\alpha > \gamma/(\gamma-1)$. Тогда $\varphi = \alpha-1$ и $\psi(v) = v$.

Отсюда вытекают разложения (5).

Лемма 2. *Справедливо следующее асимптотическое разложение функции $\psi_{1,r-1}(s)$:*

$$\psi_{1,r-1}(s\rho^\delta) = \frac{\rho_{r-1,1}}{\rho_{r-1}} s\rho^\delta + o(\rho^\delta).$$

Доказательство. Из определения функции $\psi_{1,r-1}(s)$ имеем:

$$\begin{aligned} & \sum_{j=1}^N \frac{c_j a_j}{a_j - \psi_{1,r-1}(s\rho^\delta)} \times \\ & \times \left(\sum_{m=1}^{r-1} p_m \beta_m (s\rho^\delta + \psi_{1,r-1}(s\rho^\delta)) + p_r \right) = 1. \end{aligned}$$

Отсюда и из (1) получаем:

$$\begin{aligned} & (1 + a^{-1} \psi_{1,r-1}(s\rho^\delta) + o(\psi_{1,r-1}(s\rho^\delta))) \times \\ & \times (1 - a^{-1} \rho_{r-1,1}(s\rho^\delta + \psi_{1,r-1}(s\rho^\delta)) + \\ & + o(s\rho^\delta + \psi_{1,r-1}(s\rho^\delta))) = 1. \end{aligned}$$

Отсюда следует утверждение леммы.

В дальнейшем

$$\delta = \begin{cases} \frac{\alpha}{\gamma}, & \alpha \leq \frac{\gamma}{\gamma-1}; \\ \frac{1}{\gamma-1}, & \alpha > \frac{\gamma}{\gamma-1}. \end{cases}$$

Лемма 3. *Справедливы следующие асимптотические разложения:*

$$\begin{aligned} & 1 - \sum_{j=1}^N \frac{c_j a_j}{v\rho^\alpha + a_j - (s\rho^\delta + \psi_{1,r-1}(s\rho^\delta))} \times \\ & \times \sum_{m=1}^r p_m \beta_m (s\rho^\delta + \psi_{1,r-1}(s\rho^\delta)) = \\ & = \begin{cases} a^{-1} \rho^\alpha \left(v - \frac{us^\gamma}{\rho_{r-1}^\gamma} \right) + o(\rho^\alpha), & \alpha < \frac{\gamma}{\gamma-1}; \\ a^{-1} \rho^{\gamma/(\gamma-1)} \left(v - \frac{s}{\rho_{r-1}} - \frac{us^\gamma}{\rho_{r-1}^\gamma} \right) + o(\rho^{\gamma/(\gamma-1)}), & \alpha = \frac{\gamma}{\gamma-1}; \\ -a^{-1} \frac{s}{\rho_{r-1}} \rho^{\gamma/(\gamma-1)} \left(1 + \frac{s^{\gamma-1}u}{\rho_{r-1}^{\gamma-1}} \right) + \\ + o(\rho^{\gamma/(\gamma-1)}), & \alpha > \frac{\gamma}{\gamma-1}. \end{cases} \end{aligned}$$

Доказательство. Используя разложения (1) и результаты леммы 2, имеем:

$$\begin{aligned}
 & 1 - \sum_{j=1}^N \frac{c_j a_j}{v \rho^\alpha + a_j - (s \rho^\delta + \psi_{1,r-1}(s \rho^\delta))} \times \\
 & \quad \times \sum_{m=1}^r p_m \beta_m (s \rho^\delta + \psi_{1,r-1}(s \rho^\delta)) = \\
 & = 1 - \left(1 + a^{-1} (s \rho^\delta + \psi_{1,r-1}(s \rho^\delta) - v \rho^\alpha) + \right. \\
 & \quad \left. + \sum_{j=1}^N c_j a_j^{-2} (s \rho^\delta + \psi_{1,r-1}(s \rho^\delta) - v \rho^\alpha)^2 + o(\rho^{2\delta}) \right) \times \\
 & \quad \times \left(1 - \sum_{m=1}^r p_m \beta_{m1} (s \rho^\delta + \psi_{1,r-1}(s \rho^\delta)) + \right. \\
 & \quad \left. + \sum_{m=1}^r p_m \hat{\beta}_{m\gamma} (s \rho^\delta + \psi_{1,r-1}(s \rho^\delta))^\gamma + o(\rho^{\delta\gamma}) \right) = \\
 & = a^{-1} v \rho^\alpha - a^{-1} \rho (s \rho^\delta + \psi_{1,r-1}(s \rho^\delta)) - \\
 & \quad - a^{-1} u (s \rho^\delta + \psi_{1,r-1}(s \rho^\delta))^\gamma + o(\rho^{\delta\gamma}) = \\
 & = a^{-1} \left(v \rho^\alpha - \frac{s \rho^{\delta+1}}{\rho_{r-1}} - \frac{s^\gamma u}{\rho_{r-1}^\gamma} \rho^{\delta\gamma} \right) + o(\rho^{\delta\gamma}).
 \end{aligned}$$

Отсюда вытекает утверждение леммы.

Лемма 4. Для функций $p_{0k}(v)$ справедливы следующие асимптотические разложения:

$$\begin{aligned}
 & p_{0k}(v \rho^\alpha) = \\
 & = \begin{cases} \left(\frac{u}{v} \right)^{1/\gamma} f_k \rho^{-\alpha/\gamma} + o(\rho^{-\alpha/\gamma}), & \alpha < \frac{\gamma}{\gamma-1}; \\ \psi^{-1}(v) f_k \rho^{-1/(\gamma-1)} + o(\rho^{-1/(\gamma-1)}), & \alpha = \frac{\gamma}{\gamma-1}; \\ \frac{f_k}{v} \rho^{1-\alpha} + o(\rho^{1-\alpha}), & \alpha > \frac{\gamma}{\gamma-1}, \end{cases}
 \end{aligned}$$

где

$$f_k = \prod_{i \neq k} \frac{a_i}{a_k - a_i} \prod_{j=2}^N \frac{a_k - \psi_{jr}(0)}{\psi_{jr}(0)}.$$

Доказательство непосредственно вытекает из определения функций $p_{0k}(v)$ и результатов леммы 1.

Лемма 5.

$$\begin{aligned}
 & \lim \rho^\alpha \sum_{k=1}^N a_k \omega_{rk} (s \rho^\delta - \mu_{1,r-1}^{(1)}(s \rho^\delta), v \rho^\alpha) = \\
 & = \begin{cases} \frac{a^* (1 - s/(\rho_{r-1}^* (u^*/v)^{1/\gamma}))}{v (1 - u^*/v (s/(\rho_{r-1}^*))^\gamma)}, & \alpha < \frac{\gamma}{\gamma-1}; \\ \frac{a^* (1 - s/(\rho_{r-1}^* \psi(v)))}{v - s/(\rho_{r-1}^*) - u^* s^\gamma / (\rho_{r-1}^*)^\gamma}, & \alpha = \frac{\gamma}{\gamma-1}; \\ \frac{a^*}{v (1 + s^{\gamma-1} u^* / (\rho_{r-1}^*)^{\gamma-1})}, & \alpha > \frac{\gamma}{\gamma-1}. \end{cases}
 \end{aligned}$$

Доказательство. Прежде всего заметим, что $\sum_{k=1}^N f_k = 1$. Далее из (4) имеем:

$$\begin{aligned}
 & \rho^\alpha \sum_{k=1}^N a_k \omega_{rk} (s \rho^\delta - \mu_{1,r-1}^{(1)}(s \rho^\delta), v \rho^\alpha) = \\
 & = \left(\sum_{j=1}^N \frac{c_j a_j}{v \rho^\alpha - s \rho^\delta + \mu_{1,r-1}^{(1)}(s \rho^\delta) + a_j} - \right. \\
 & \quad \left. - (s \rho^\delta - \mu_{1,r-1}^{(1)}(s \rho^\delta)) \times \right. \\
 & \quad \left. \times \sum_{j=1}^N \frac{a_j p_{0j}(v \rho^\alpha)}{v \rho^\alpha - s \rho^\delta + \mu_{1,r-1}^{(1)}(s \rho^\delta) + a_j} \right) \times \\
 & \quad \times \rho^\alpha / \left(1 - \sum_{j=1}^N \frac{c_j a_j}{v \rho^\alpha - s \rho^\delta + \mu_{1,r-1}^{(1)}(s \rho^\delta) + a_j} \times \right. \\
 & \quad \left. \times \sum_{m=1}^r p_m \beta_m (s \rho^\delta - \mu_{1,r-1}^{(1)}(s \rho^\delta)) \right).
 \end{aligned}$$

Рассмотрим отдельно три случая.

1. $\alpha < \gamma/(\gamma-1)$. В силу лемм 2–4

$$\begin{aligned}
 & \rho^\alpha \sum_{k=1}^N a_k \omega_{rk} (s \rho^\delta - \mu_{1,r-1}^{(1)}(s \rho^\delta), v \rho^\alpha) = \\
 & = a \left(v - \frac{u s^\gamma}{\rho_{r-1}^\gamma} + o(1) \right)^{-1} \times \\
 & \quad \times \left(1 - \frac{s \rho^{\alpha/\gamma}}{\rho_{r-1}} \left(\frac{u}{v} \right)^{1/\gamma} \rho^{-\alpha/\gamma} + o(1) \right) = \\
 & = \frac{a (1 - (s/(\rho_{r-1})) (u/v)^{1/\gamma})}{v (1 - (u/v) (s/(\rho_{r-1}))^\gamma)} + o(1).
 \end{aligned}$$

2. $\alpha = \gamma/(\gamma-1)$. В этом случае

$$\begin{aligned}
 & \rho^\alpha \sum_{k=1}^N a_k \omega_{rk} (s \rho^\delta - \mu_{1,r-1}^{(1)}(s \rho^\delta), v \rho^\alpha) = \\
 & = \left(1 - \frac{s}{\rho_{r-1}} \psi^{-1}(v) + o(1) \right) \times \\
 & \quad \times a \left(v - \frac{s}{\rho_{r-1}} - \frac{u s^\gamma}{\rho_{r-1}^\gamma} + o(1) \right)^{-1} = \\
 & = \frac{a (1 - s/(\rho_{r-1} \psi(v)))}{v - s/(\rho_{r-1}) - u s^\gamma / (\rho_{r-1}^\gamma)} + o(1).
 \end{aligned}$$

3. $\alpha > \gamma/(\gamma-1)$. В этом случае

$$\rho^\alpha \sum_{k=1}^N a_k \omega_{rk} (s \rho^\delta - \mu_{1,r-1}^{(1)}(s \rho^\delta), v \rho^\alpha) =$$

$$= \frac{-a\rho^{\alpha-\gamma/(\gamma-1)} + o\left(\rho^{\alpha-\gamma/(\gamma-1)}\right)}{(s/\rho_{r-1})\left(1 + s^{\gamma-1}u/\rho_{r-1}^{\gamma-1}\right)} \times \\ \times \left(1 - \frac{s\rho^{1/(\gamma-1)}}{\rho_{r-1}} \frac{\rho^{1-\alpha}}{v} + o\left(\rho^{1/(\gamma-1)+1-\alpha}\right)\right) = \\ = \frac{a}{v\left(1 + s^{\gamma-1}u/\rho_{r-1}^{\gamma-1}\right)} + o(1).$$

Переходя в полученных соотношениях к пределу при $n \rightarrow +\infty$, получаем утверждение леммы.

Теорема. При $n \rightarrow +\infty$ существует предел

$$\lim P\left(\rho^\delta w(t\rho^{-\alpha}) < x\right) = G(x, t), \quad x > 0, \quad t > 0,$$

где функция $G(x, t)$ определяется своими преобразованиями Лапласа и Лапласа–Стилтьеса

$$\int_0^\infty \int_0^\infty e^{-vt-sx} d_x G(x, t) = \\ = \begin{cases} \frac{\left(1 - (s/\rho_{r-1}^*) (u^*/v)^{1/\gamma}\right)}{v\left(1 - (u^*/v) (s/\rho_{r-1}^*)^\gamma\right)}, & \alpha < \frac{\gamma}{\gamma-1}; \\ \frac{\left(1 - s/(\rho_{r-1}^* \psi(v))\right)}{v - s/\rho_{r-1}^* - u^* s^\gamma/(\rho_{r-1}^*)^\gamma}, & \alpha = \frac{\gamma}{\gamma-1}; \\ \frac{1}{v\left(1 + s^{\gamma-1}u^*/(\rho_{r-1}^*)^{\gamma-1}\right)}, & \alpha > \frac{\gamma}{\gamma-1}. \end{cases}$$

Доказательство. Заметим, что

$$\int_0^{+\infty} e^{-vt} \mathbf{E} \left(\exp \left(-s\rho^\delta w(t\rho^{-\alpha}) \right) \right) dt = \\ = \sum_{j=1}^N \rho^\alpha \omega_j(s\rho^\delta, v\rho^\alpha).$$

Поэтому достаточно найти $\lim \sum_{j=1}^N \rho^\alpha \omega_j(s\rho^\delta, v\rho^\alpha)$.

Из (2) и свойств функций $\mu_{c,r-1}^{(p)}(s)$ и $\alpha_c \left(z_{r-1,1}^{(k)}(s), \dots, z_{r-1,1}^{(k)}(s), 1 \right)$ следует, что

$$\lim \sum_{j=1}^N \rho^\alpha \omega_j(s\rho^\delta, v\rho^\alpha) = \lim \sum_{\nu=1}^N c_\nu a_\nu^{-1} \times \\ \times \frac{\prod_{q=1}^N a_q}{\alpha_1 \left(z_{r-1,1}^{(k)}(0), \dots, z_{r-1,1}^{(k)}(0), 1 \right)} \rho^\alpha \times \\ \times \sum_{j=1}^N \omega_{rj} \left(s\rho^\delta - \mu_{1,r-1}^{(1)}(s\rho^\delta), v\rho^\alpha \right).$$

Но $\alpha_1 \left(z_{r-1,1}^{(k)}(0), \dots, z_{r-1,1}^{(k)}(0), 1 \right) = a^{-1} \prod_{q=1}^N a_q$. Следовательно,

$$\lim \sum_{j=1}^N \rho^\alpha \omega_j(s\rho^\delta, v\rho^\alpha) = \\ = \lim \rho^\alpha \sum_{j=1}^N \omega_{rj} \left(s\rho^\delta - \mu_{1,r-1}^{(1)}(s\rho^\delta), v\rho^\alpha \right).$$

Теперь из (3) и из лемм 4 и 5 получаем утверждение теоремы.

Литература

1. Ушаков А. В. О виртуальном времени ожидания в системе с относительным приоритетом и гиперэкспоненциальным входящим потоком // Информатика и её применения, 2012. Т. 6. Вып. 1. С. 2–6.
2. Ушаков В. Г. Система обслуживания с эрланговским входящим потоком и относительным приоритетом // Теория вероятн. и её примен., 1977. Т. 22. С. 860–866.
3. Матвеев В. Ф., Ушаков В. Г. Системы массового обслуживания. — М.: МГУ, 1984.
4. Ушаков В. Г. Аналитические методы анализа системы массового обслуживания $GI|G_r|1$ с относительным приоритетом // Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн., 1993. № 4. С. 57–69.
5. Ушаков А. В., Ушаков В. Г. О длине очереди в системе с абсолютным приоритетом и гиперэкспоненциальным входящим потоком // Вестн. Моск. ун-та. Сер. 15. Вычисл. матем. и киберн., 2012. № 1. С. 27–34.

RADIO SOURCES STUDY WITH THE VIRTUAL OBSERVATORY TOOLS

O. P. Zhelenkova

Special Astrophysical Observatory, RAS, zhe@sao.ru

During a number of years (1980–1999), using different approaches on the basis of the Virtual Observatory means, the study of sources of the surveys was conducted on the largest Russian radio telescope RATAN-600. Their bulk identification is executed with the maximum use of data for different ranges of an electromagnetic spectrum available in open access. With the use of the Virtual Observatory program tools, the approach on automatic preparation and preliminary data processing is realized. The information system is developed for the resulting compilative catalogue preparation that was applied in the analysis of information on each source and making decision on its identification. Based on the experience gained while working with numerous heterogeneous resources, the author summarizes that the Virtual Observatory infrastructure provides convenient access to astronomical resources and significantly increases the efficiency of scientific research. However, there are still no advanced tools for further analysis, updating and publication of data collected by the researcher. A number of projects is being developed to implement the most linked data based on already existing Web technologies and that will transfer the Virtual Observatory services to a new level of knowledge exchange through annotation of catalogue entries and the implementation of links between them.

Keywords: virtual observatory; distributed information system; computer science technology in scientific research; integration of diverse information resources; databases

METHODOLOGICAL ASPECTS OF DEVELOPMENT OF THE ELECTRONIC RUSSIAN–ENGLISH THESAURUS ON COMPUTATIONAL LINGUISTICS

Yu. A. Zagorulko¹, O. I. Borovikova², I. S. Kononenko³, and E. G. Sokolova⁴

¹A. P. Ershov Institute of Informatics Systems, SB RAS, zagor@iis.nsk.su

²A. P. Ershov Institute of Informatics Systems, SB RAS, olesya@iis.nsk.su

³A. P. Ershov Institute of Informatics Systems, SB RAS, irina_k@cn.ru

⁴Russian State University for Humanities, minegot@rambler.ru

The paper discusses methodological aspects of development of the electronic Russian–English thesaurus on computational linguistics. Rationale for the creation of the thesaurus, and the design principles are substantiated. The general structure of the thesaurus, composition of the thesaurus entries, and set of relations between the thesaurus terms are described. The procedure of choice of terms for inclusion in the thesaurus is depicted and the problems with selection of the preferred term (descriptor) from a synonyms list as well as translation equivalents are discussed. Features of implementation of online digital version of the thesaurus are considered. The paper gives a particular attention to maintenance of a logical consistency of the thesaurus terminology system and to providing a convenient access to the thesaurus content.

Keywords: multilingual thesaurus; computational linguistics; methodology for thesaurus development; ontology; conceptual model of the thesaurus; scientific knowledge portal technology

CLASSIFICATION AND UTILIZATION OF SEMANTIC LINKAGES BETWEEN INFORMATION OBJECTS IN SCIENTIFIC DIGITAL LIBRARIES

M. R. Kogalovsky¹ and S. I. Parinov²

¹Market Economy Institute, RAS, kogalov@gmail.com

²Central Economics and Mathematics Institute, RAS, sparinov@gmail.com

An approach to increase the information value of scientific digital library by supporting the classified semantic linkages between its information objects is discussed. An implementation of the proposed approach is based on the domestic Socionet system that integrates a large number of scientific digital libraries and which is used *de facto* as a research information space of the Social Science Division of the Russian Academy of Sciences.

Keywords: digital library; information object; collection of information resources; semantic linkage; linkage classifier; ontology; scientometrics

INTEGRATION OF GEOGRAPHIC METADATA USING RETROSPECTIVE THESAURUS

D. M. Skachkov¹ and O. L. Zhizhimov²

¹Institute of Computational Technologies, SB RAS, danil.skachkov@gmail.com

²Institute of Computational Technologies, SB RAS, zhizhim@sbras.ru

The questions, related with the development of interoperable thesaurus of geographic names, with geometric and retrospective data included in, are discussed. The basic requirements to the thesaurus are defined. An overview of existing solutions is provided and the main positions for thesaurus access profile have been formulated. The relational database schema for thesaurus data is described.

Keywords: geographical metadata; integration; retrospective geocoding; thesaurus

UNIVERSAL SYSTEM OF PLAGIARISM DETECTION "AUTOR.NET"

E. V. Sharapova¹ and R. V. Sharapov²

¹Vladimir State University, mivlgu@mail.ru

²Vladimir State University, info@vanta.ru

The problem of duplicate texts detecting is discussed. The basic approaches to detection of text duplicates are given. The existing programs of duplicate texts detecting have been reviewed. An overview of methods to conceal the evidence of plagiarism is presented. A system "Autor.NET" which checks the text duplications founded on the internal base of the sources and on Internet has been created.

Keywords: plagiarism; detection of plagiarism; copying

FRACTIONAL LEVY MOTION WITH DEPENDENT INCREMENTS AND ITS APPLICATION TO NETWORK TRAFFIC MODELING

C. De Nikola¹, Y. S. Khokhlov², M. Pagano³, and O. I. Sidorova⁴

¹University of Salerno, denicola@diima.unisa.it

²People's Friendship University of Russia, yskhokhlov@yandex.ru

³University of Pisa, m.pagano@iet.unipi.it

⁴Tver State University, Oksana.I.Sidorova@yandex.ru

Since the beginning of the 1990s, accurate traffic measurements carried out in different network scenarios highlighted that Internet traffic exhibits strong irregularities (*burstiness*) both in terms of extreme variability and long-term correlations. These features, which cannot be captured in a parsimonious way by traditional Markovian

models, have a deep impact on the network performance and lead to the introduction of α -stable distribution and self-similar processes into the network traffic modeling. In this paper, a generalization of fractional Brownian motion, which is able to capture both above-mentioned features of the real traffic, is considered.

Keywords: fractional Brownian motion; α -stable subordinator; self-similar processes; buffer overflow probability

ABOUT THE RATE OF CONVERGENCE OF ONE U-STATISTIC

O. Yanushkevichienė¹ and R. Yanushkevichius²

¹Vilnius University, Institute of Mathematics and Informatics, Lithuania, olgjan@zebra.lt

²Lithuanian University of Educational Sciences, Vilnius, Lithuania

This paper pertains to the analysis of the asymptotic behavior of U-statistics, which are important in the construction and application of modern statistical methods for studying information systems.

Keywords: rates of convergence; Berry–Esseen bound; quadratic form; second-degree U-statistics; Kolmogorov distance

ESTIMATES OF THE RATE OF CONVERGENCE OF THE DISTRIBUTIONS OF RANDOM SUMS TO VARIANCE-GAMMA DISTRIBUTIONS

V. E. Bening¹, L. M. Zaks², and V. Yu. Korolev³

¹Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, bening@cs.msu.su

²Department of Modeling and Mathematical Statistics, Alpha-Bank, lily.zaks@gmail.com

³Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University; IPI RAN, vkorolev@cs.msu.su

The estimates are constructed for the accuracy of the approximation of the distributions of special random sums of independent random variables with nonzero means by variance-gamma distributions.

Keywords: random sum; negative binomial distribution; gamma-distribution; convergence rate estimate; mixed Poisson distribution; variance-gamma distribution

CROSS-LAYER OPTIMIZATION PLANNING PROBLEMS IN WIRELESS NETWORKS

Y. V. Gaidamaka¹, T. V. Efimushkina², A. K. Samuylov³, and K. E. Samouylov⁴

¹Peoples' Friendship University of Russia, Telecommunication Systems Department, ygaidamaka@sci.pfu.edu.ru

²Peoples' Friendship University of Russia, Telecommunication Systems Department, tefimushkina@gmail.com

³Peoples' Friendship University of Russia, Telecommunication Systems Department, asam1988@gmail.com

⁴Peoples' Friendship University of Russia, Telecommunication Systems Department, ksam@sci.pfu.edu.ru

The review provides the state-of-the-art of the cross-layer optimization problems for the Orthogonal Frequency Division Multiplex technology of Long-Term Evolution wireless networks. The notion of cross-layer optimization is considered, the detailed optimization problems and its constraints arising from the radio resources allocation in the downlink are investigated, and the brief discussion of schedulers and its utilization functions are presented.

Keywords: OFDM; cross-layer optimization; utility function; efficient frequency allocation

ASYMPTOTICS OF THE MAXIMUM WORKLOAD FOR A CLASS OF GAUSSIAN QUEUEING SYSTEMS

O. V. Lukashenko¹ and E. V. Morozov²¹Institute of Applied Mathematical Research of Karelian Research Center, RAS; Petrozavodsk State University, lukashenko-oleg@mail.ru²Institute of Applied Mathematical Research of Karelian Research Center, RAS; Petrozavodsk State University, emorozov@karelia.ru

The asymptotics of the maximum workload in a fluid queueing system fed by a process containing a random component are described by a centered Gaussian process. It is assumed that the variance of the process is a regularly varying at infinity function with index belonging to interval $(0, 2)$. Such class of processes includes, in particular, a sum of independent fractional Brownian motions. It is shown that, under an appropriate scaling, the maximum workload over interval $[0, t]$ converges in probability to an explicitly given constant as t increases.

Keywords: Gaussian queueing system; maximum workload; fractional Brownian motion; asymptotical analysis; regular variation

ON THE ESTIMATION OF THE OVERFLOW PROBABILITY IN REGENERATIVE FINITE BUFFER QUEUEING SYSTEMS

E. V. Morozov¹ and R. S. Nekrasova²¹Institute of Applied Mathematical Research of Karelian Research Center, RAS; Petrozavodsk State University, emorozov@karelia.ru²Institute of Applied Mathematical Research of Karelian Research Center, RAS; Petrozavodsk State University, ruslana.nekrasova@mail.ru

A few aspects of the regenerative estimation of the stationary overflow probability in finite buffer system are considered. A proof of a general relation connecting the loss probability and idle probability in stationary regime is given. An application of this result to a broad class of loss systems, in particular, to a retrial system with constant retrial rate, is shown. An efficiency of this relation for regenerative estimation of the loss probability under different traffic regimes and also with use of k -regenerations, appearing in the analysis of non-Markovian systems, is studied. The numerical results of estimation are given.

Keywords: finite buffer systems; loss probability; idle probability; regenerative estimation; k -regenerations; retrial queueing system

STOCHASTIC MODELS OF MULTIPROCESSOR SYSTEMS: STABILITY AND MOMENT PROPERTIES

E. V. Morozov and A. S. Rumyantsev²¹Institute of Applied Mathematical Research of Karelian Research Center, RAS; Petrozavodsk State University, emorozov@karelia.ru²Institute of Applied Mathematical Research of Karelian Research Center, RAS, ar0@krc.karelia.ru

The basic models of multiprocessor systems in which each job requires a random number of processors have been studied. A new model of such system is proposed and studied, where, for a given job, all processing times are identical. This assumption reflecting real service process makes the analysis of the system more difficult. To investigate the model, some minorant and majorant (classical) systems are constructed, which allow, for a few important particular cases, to obtain both stability conditions and the moment properties of the original stationary workload process.

Keywords: multiprocessor systems; random number of processors; identical processing times; stability conditions; moment properties; stationary workload; high performance cluster

A RESTRICTION ON THE TOTAL VOLUME OF DEMANDS IN THE DISCRETE-TIME SYSTEM
 $\text{Geo}/G/1/\infty$

A. V. Pechinkin¹, I. A. Sokolov², and S. Ya. Shorgin³

¹IPI RAN, apechinkin@ipiran.ru

²IPI RAN, isokolov@ipiran.ru

³IPI RAN, sshorgin@ipiran.ru

Consideration is given to a discrete-time queueing system $\text{Geo}/G/1/\infty$ with inverse service discipline without service interruption, where each demand has random volume besides its length. The total volume of the demands in the queue is limited by a certain nonrandom value. The algorithms for the system main stationary characteristics evaluation are developed.

Keywords: queueing system; discrete time; length and volume of a demand

HEAVY-TRAFFIC ANALYSIS FOR THE QUEUEING SYSTEM WITH HYPEREXPONENTIAL STREAM

A. V. Ushakov

IPI RAN, ushakov@akado.ru

The heavy traffic limiting distributions of the virtual waiting time in a single server queue with hyperexponential input stream and head-of-the-line priority discipline are obtained.

Keywords: virtual waiting time; head-of-the-line priority; hyperexponential input stream; heavy traffic

Об авторах

Бенинг Владимир Евгеньевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник ИПИ РАН

Боровикова Олеся Игнатьевна (р. 1977) — младший научный сотрудник Института систем информатики им. А. П. Ершова Сибирского отделения Российской академии наук

Гайдамака Юлия Васильевна (р. 1971) — кандидат физико-математических наук, доцент Российского университета дружбы народов

Де Никола Кармине (р. 1972) — доктор математических наук (PhD), инженер-электронщик, научный сотрудник Университета г. Салерно, Италия

Ефимушкина Татьяна Владимировна (р. 1988) — аспирант Российского университета дружбы народов

Желенкова Ольга Петровна (р. 1955) — кандидат физико-математических наук, старший научный сотрудник Специальной астрофизической обсерватории РАН

Жижимов Олег Львович (р. 1956) — доктор технических наук, заведующий лабораторией Института вычислительных технологий СО РАН

Загорulyко Юрий Алексеевич (р. 1957) — кандидат технических наук, заведующий лабораторией Института систем информатики им. А. П. Ершова Сибирского отделения Российской академии наук

Закс Лилия Михайловна (р. 1989) — главный специалист отдела моделирования и математической статистики Альфа-банка

Когаловский Михаил Рувимович (р. 1939) — кандидат технических наук, доцент, ведущий научный сотрудник Института проблем рынка РАН; доцент Московского государственного университета им. М. В. Ломоносова

Кононенко Ирина Семеновна (р. 1953) — научный сотрудник Института систем информатики им. А. П. Ершова Сибирского отделения Российской академии наук

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; ведущий научный сотрудник ИПИ РАН

Лукашенко Олег Викторович (р. 1986) — аспирант Института прикладных математических исследований КарНЦ РАН; преподаватель Петрозаводского государственного университета

Морозов Евсей Викторович (р. 1947) — доктор физико-математических наук, профессор, ведущий научный сотрудник Института прикладных математических исследований КарНЦ РАН; профессор Петрозаводского государственного университета

Некрасова Руслана Сергеевна (р. 1986) — аспирант Института прикладных математических исследований КарНЦ РАН; преподаватель Петрозаводского государственного университета

Пагано Микеле (р. 1968) — доктор математических наук (PhD), профессор Университета г. Пиза, Италия

Паринов Сергей Иванович (р. 1956) — доктор технических наук, главный научный сотрудник Центрального экономико-математического института РАН

Печинкин Александр Владимирович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник ИПИ РАН; профессор Российского университета дружбы народов

Румянцев Александр Сергеевич (р. 1986) — младший научный сотрудник Института прикладных математических исследований КарНЦ РАН

Самуйлов Андрей Константинович (р. 1988) — аспирант Российского университета дружбы народов

Самуйлов Константин Евгеньевич (р. 1955) — доктор физико-математических наук, профессор, заведующий кафедрой Российского университета дружбы народов

Сидорова Оксана Игоревна (р. 1977) — кандидат физико-математических наук, доцент Тверского государственного университета

Скачков Данил Михайлович (р. 1988) — аспирант Института вычислительных технологий СО РАН

Соколов Игорь Анатольевич (р. 1954) — академик (действительный член) Российской академии наук, доктор технических наук, директор ИПИ РАН

Соколова Елена Григорьевна (р. 1952) — кандидат филологических наук, доцент Российского государственного гуманитарного университета, г. Москва

Ушаков Андрей Владимирович (р. 1987) — аспирант ИПИ РАН

Хохлов Юрий Степанович (р. 1952) — доктор физико-математических наук, профессор, заведующий кафедрой Российского университета дружбы народов

Шарапов Руслан Владимирович (р. 1976) — кандидат технических наук, доцент Муромского института (филиала) Владимирского государственного уни-

верситета им. Александра Григорьевича и Николая Григорьевича Столетовых

Шарапова Екатерина Викторовна (р. 1981) — инженер Муромского института (филиала) Владимирского государственного университета им. Александра Григорьевича и Николая Григорьевича Столетовых

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, заместитель директора ИПИ РАН

Янушкявичене Ольга Леонидовна (р. 1955) — доктор физико-математических наук, профессор, старший научный сотрудник Института математики и информатики Вильнюсского университета, Литва

Янушкявичюс Романас (р. 1953) — доктор физико-математических наук, профессор, заведующий кафедрой математики Литовского университета эдукологии, г. Вильнюс, Литва

About Authors

Bening Vladimir E. (b. 1954) — Doctor of Science in physics and mathematics; professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; senior scientist, Institute of Informatics Problems, Russian Academy of Sciences

Borovikova Olesya I. (b. 1977) — junior scientist, A. P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences

De Nicola Carmine (b. 1972) — PhD in mathematics, electronic engineer, scientist, University of Salerno, Italy

Efimushkina Tatiana V. (b. 1988) — PhD student, Peoples' Friendship University of Russia

Gaidamaka Yuliya V. (b. 1971) — Candidate of Science (PhD) in physics and mathematics, associate professor, Peoples' Friendship University of Russia

Khokhlov Yuri S. (b. 1952) — Doctor of Science (PhD) in physics and mathematics, professor, Head of Department, People's Friendship University of Russia

Kogalovsky Mikhail R. (b. 1939) — Candidate of Science (PhD) in technology, leading scientist of Market Economy Institute of the Russian Academy of Sciences; associate professor, M.V. Lomonosov Moscow State University

Kononenko Irina S. (b. 1953) — scientist, A. P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences

Lukashenko Oleg V. (b. 1986) — PhD student, Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences; lecturer, Petrozavodsk State University

Morozov Evsei V. (b. 1947) — Doctor of Science in physics and mathematics, professor, leading scientist, Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences; professor, Petrozavodsk State University

Nekrasova Ruslana S. (b. 1986) — PhD student, Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences; lecturer, Petrozavodsk State University

Pagano Michele (b. 1968) — PhD in mathematics, professor, University of Pisa, Italy

Parinov Sergey I. (b. 1956) — Doctor of Science in technology; principal scientist, Central Economics and Mathematics Institute, Russian Academy of Sciences

Pechinkin Alexander V. (b. 1946) — Doctor of Science in physics and mathematics; principal scientist, Institute of Informatics Problems, Russian Academy of Sciences; professor, Peoples' Friendship University of Russia

Rumyantsev Alexander S. (b. 1986) — junior scientist, Institute of Applied Mathematical Research of Karelian Research Center, Russian Academy of Sciences

Samouylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples' Friendship University of Russia

Samuylov Andrey K. (b. 1988) — PhD student, Peoples' Friendship University of Russia

Sharapov Ruslan V. (b. 1976) — Candidate of Science (PhD) in technology, associate professor, Murom Institute of Vladimir State University

Sharapova Ekaterina V. (b. 1981) — engineer, Murom Institute of Vladimir State University

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, Deputy Director, Institute of Informatics Problems, Russian Academy of Sciences

Sidorova Oksana I. (b. 1977) — Candidate of Science (PhD) in physics and mathematics, associate professor, Tver State University

Skachkov Danil M. (b. 1988) — PhD student, Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences

Sokolov Igor A. (b. 1954) — Academician of the Russian Academy of Sciences, Doctor of Science in technology, Director, Institute of Informatics Problems, Russian Academy of Sciences

Sokolova Elena G. (b. 1952) — Candidate of Science (PhD) in philology, associate professor, Russian State University for Humanities

Ushakov Andrey V. (b. 1987) — PhD student, Institute of Informatics Problems, Russian Academy of Sciences

Yanushkeviciene Olga L. (b. 1955) — Doctor of Science in physics and mathematics, senior scientist, professor, Vilnius University, Institute of Mathematics and Informatics, Lithuania

Yanushkevichius Romanas (b. 1953) — Doctor of Science in physics and mathematics, professor, Head of the Department of Mathematics, Lithuanian University of Educational Sciences, Vilnius, Lithuania

Zagorulko Yury A. (b. 1957) — Candidate of Science (PhD) in technology, Head of Laboratory, A. P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences

Zaks Lily M. (b. 1989) — principal officer, Department of Modeling and Mathematical Statistics, Alpha-Bank

Zhelenkova Olga P. (b. 1955) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Special Astrophysical Observatory, Russian Academy of Sciences

Zhizhimov Oleg L. (b. 1956) — Doctor of Science in technology, Head of Laboratory, Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences

Правила подготовки рукописей статей 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, в виде исключения, 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 результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях. Публикация не должна нарушать закон об авторских правах. Направляя свою рукопись в редакцию, авторы автоматически передают учредителям и редколлегии неисключительные права на издание данной статьи на русском языке и на ее распространение в России и за рубежом. При этом за авторами сохраняются все права как собственников данной рукописи. В связи с этим авторами должно быть представлено 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. Подписи авторов: (ф. и. о., дата, адрес)».

Указанное соглашение может быть представлено как в бумажном виде, так и в виде отсканированной копии (с подписями авторов).

Редколлегия вправе запросить у авторов экспертное заключение о возможности опубликования представленной статьи в открытой печати.

2. Статья подписывается всеми авторами. На отдельном листе представляются данные автора (или всех авторов): фамилия, полные имя и отчество, телефон, факс, e-mail, почтовый адрес. Если работа выполнена несколькими авторами, указывается фамилия одного из них, ответственного за переписку с редакцией.
3. Редакция журнала осуществляет самостоятельную экспертизу присланных статей. Возвращение рукописи на доработку не означает, что статья уже принята к печати. Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.
4. Решение редакционной коллегии о принятии статьи к печати или ее отклонении сообщается авторам. Редколлегия не обязуется направлять рецензию авторам отклоненной статьи; дискуссия с авторами по поводу отклоненных статей не ведется.
5. Корректурa статей высылается авторам для просмотра. Редакция просит авторов присылать свои замечания в кратчайшие сроки.
6. 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 25 страниц указанного формата. Ознакомиться с шаблонами, содержащими примеры оформления, можно по адресу в Интернете: <http://www.ipiran.ru/journal/template.doc>.
7. К рукописи, предоставляемой в 2-х экземплярах, обязательно прилагается электронная версия статьи (как правило, в форматах MS WORD (.doc) или L^AT_EX (.tex), а также — дополнительно — в формате .pdf) на дискете, лазерном диске или по электронной почте. Сокращения слов, кроме стандартных, не применяются. Все страницы рукописи должны быть пронумерованы.
8. Статья должна содержать следующую информацию на русском и английском языках: название, Ф.И.О. авторов, места работы авторов и их электронные адреса, подробные сведения об авторах, оформленные в соответствии с форматом, определяемым файлами http://www.ipiran.ru/journal/issues/2011_05_01/authors.asp и http://www.ipiran.ru/journal/issues/2011_01_eng/authors.asp, аннотация (не более 100 слов), ключевые слова. Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в порядке их первого упоминания. В списке литературы не должно быть позиций, на которые нет ссылки в тексте статьи. Все фамилии авторов, заглавия статей, названия книг, конференций и т. п. даются на языке оригинала, если этот язык использует кириллический или латинский алфавит.
9. Присланные в редакцию материалы авторам не возвращаются.
10. При отправке файлов по электронной почте просим придерживаться следующих правил:
 - указывать в поле subject (тема) название журнала и фамилию автора;
 - использовать attach (присоединение);
 - в случае больших объемов информации возможно использование общеизвестных архиваторов (ZIP, RAR);
 - в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.
11. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию с авторов не взимается, гонорар авторам не выплачивается.

Адрес редакции: Москва 119333, ул. Вавилова, д. 44, корп. 2, ИПИ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05 E-mail: rust@ipiran.ru