

Федеральное государственное автономное образовательное учреждение  
высшего образования «Уральский федеральный университет имени первого  
Президента России Б.Н. Ельцина»

На правах рукописи



БОЖАЛКИН Даниил Александрович

МАТЕМАТИЧЕСКОЕ И АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ  
ДЛЯ АНАЛИЗА ХАРАКТЕРИСТИК ИНФОРМАЦИОННЫХ ПОТОКОВ  
В МАГИСТРАЛЬНЫХ ИНТЕРНЕТ-КАНАЛАХ

05.13.01 – Системный анализ, управление и обработка информации  
(информатизация и связь)

АВТОРЕФЕРАТ  
диссертации на соискание ученой степени кандидата  
технических наук

Екатеринбург – 2020

Работа выполнена в учебно-научном центре «Информационная безопасность» Института радиоэлектроники и информационных технологий – РтФ ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина».

Научный руководитель – доктор технических наук, профессор,  
ПОРШНЕВ Сергей Владимирович

Официальные оппоненты: НИКУЛЬЧЕВ Евгений Витальевич, доктор технических наук, профессор, ФГБОУ ВО «МИРЭА – Российский технологический университет», г. Москва, профессор кафедры «Управление и моделирование систем»;

БЕРГ Дмитрий Борисович, доктор физико-математических наук, профессор, ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», профессор кафедры анализа систем и принятия решений Института экономики и управления Школы экономики и менеджмента;

БУДЫЛДИНА Надежда Вениаминовна, кандидат технических наук, доцент, Уральский технический институт связи и информатики (филиал) ФГБОУ ВО «Сибирский государственный университет телекоммуникаций и информатики» в г. Екатеринбурге, доцент кафедры общепрофессиональных дисциплин технических специальностей

Защита диссертации состоится «15» сентября 2020 г. в 11:00 часов на заседании диссертационного совета УрФУ 05.07.18 по адресу: 620002, г. Екатеринбург, ул. Мира, 19, ауд. И-420 (зал Ученого совета).

С диссертацией можно ознакомиться в библиотеке и на сайте ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», <https://dissovet2.urfu.ru/mod/data/view.php?d=12&rid=1264>

Автореферат разослан «\_\_\_» \_\_\_\_\_ 2020 г.

Ученый секретарь  
диссертационного совета



Сафиуллин Николай Тахирович

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования и степень ее разработанности.** Сегодня компьютерные сети (КС) широко используются для передачи различной информации, контроля и управления различными сервисами в реальном времени, просмотра телепередач, онлайн покупок и т.д. В связи с увеличением новых классов телекоммуникационных устройств и соответствующих сервисов быстрыми темпами увеличиваются объемы информации, передаваемой через сеть Интернет (Интернет-трафик). Например, по данным Cisco Visual Networking Index (наглядные показатели Сети) объемы передаваемого Интернет-трафика увеличились со 100 Г/сутки в 1992 г. до 16 000 Гб/с в 2014 г. При этом существенно усложнилась собственно структура передаваемой информации, которая создается и используется многочисленными пользователями персональных компьютеров, смартфонов, планшетов, телевизоров, бытовой техникой (Интернет вещей) и др.

В этой ситуации закономерно возрастают требования к гибкости и масштабируемости современных КС свойства, которых оказываются существенно отличными от свойств КС с классической архитектурой (по сути, статических). Например, традиционные архитектуры/дизайны КС оказываются неэффективными в динамических средах. При этом классические подходы, ориентированные на распределенное управление устройствами традиционных КС (например, виртуальные сети (VLAN)), не соответствуют современному уровню развития виртуализации серверов и систем хранения данных, а также требованиям крупного бизнеса и сервис-провайдеров (например, AT&T, Verizon, Google, Facebook, Microsoft и др.). Сложившаяся ситуация в телекоммуникационной отрасли подтверждается, том числе, данными, содержащимися в аналитическом отчете за 2016 г. компании KPMG, где введены понятия «разрушающий трафик» (т.е. трафик таких объемов, с которыми не справляется используемое сетевое оборудование) и «разрушительные технологии», которые создают разрушающий трафик (виртуальная реальность, облачные сервисы, искусственный интеллект, анализ данных в реальном времени и др.).

Для эффективного решения проблем «разрушающего трафика», вопросов проектирования оборудования отвечающего потребностям современных КС, а так же проектирования КС нового поколения, в том числе и виртуальных программно-конфигурируемых сетей – SDN (Software-defined Networking), необходимо понимать особенности информационных потоков, передаваемых в современных КС, механизмы их взаимодействия друг с другом и влияния на загрузку канала.

Анализ состояния современной теории телетрафика показывает, что имеет место определенный разрыв между современными уровнями развития телекоммуникационных технологий и математических моделей информационных процессов в КС, который, в известной мере, пытаются восполнить большим количеством экспериментальных результатов, проведенных исследований особенностей информационных потоков в КС, в особенности, высокоскоростных магистральных Интернет-каналах (см. работы О.И. Шелухина, В.В. Петрова, Н.Г. Треногина, Е.В. Никульчева, М. Soysal, K. Fukuda, W. Leland, W. Willinger, D. Wilson и др.).

Однако объективный анализ этих работ показал, что проводимые экспериментальные исследования, зачастую, имеют бессистемный характер. Это проявляется: в отсутствии общепринятой методики исследований Интернет-трафика в магистральном высокоскоростном интернет-канале; их направленности не на проверку, но на подтверждение тех или иных изначально выбранных математических моделей Интернет-трафика (в первую очередь, самоподобных моделей).

**Цель работы** разработка и применение математического и алгоритмического обеспечения для анализа характеристик информационных потоков в высокоскоростных магистральных Интернет-каналах.

Для достижения поставленной цели были поставлены и решены следующие основные задачи исследования:

1. Анализ методов исследования информационных потоков в КС сетей с точки зрения их применимости для исследования трафика в высокоскоростных магистральных Интернет-каналах.

2. Разработка программного инструмента, обеспечивающего автоматическое извлечение информации из rсар-файлов в выбранном измерении.
3. Разработка методики анализа первичной информации, извлеченной в соответствующем измерении из rсар-файлов, обеспечивающей получение количественных характеристики информационных потоков, переданных в магистральном высокоскоростном Интернет-канале.
4. Изучение особенностей информационных потоков в магистральном Интернет-канале, создаваемых выбранными классами пользователей («Слоны», «Мулы», «Мыши»), и их взаимного влияния друг на друга.

**Научная новизна.** В диссертации получены следующие новые научные результаты:

1. Разработано математическое и алгоритмическое обеспечение, а также соответствующее программное обеспечение (ПО) для анализа характеристик информационных потоков в высокоскоростных магистральных Интернет-каналах.
2. Предложена методика анализа Интернет-трафика и доказана ее работоспособность при исследовании информационных потоков в высокоскоростном магистральном Интернет-канале.
3. Проведено исследование взаимного влияния информационных потоков, создаваемых в магистральном Интернет-канале выбранными классами пользователей («Слоны», «Мулы», «Мыши»), и доказано, что связи между объемами информации, переданной в магистральном Интернет-канале каждым из выбранных классов пользователей, описываются детерминированными линейными моделями.
4. Предложен алгоритм управления загрузкой канала передачи информационных потоков позволяющая, за счет отслеживания глобального показателя Херста накопленных сумм случайных последовательностей объема переданной информации класса «Мыши» и соответствующего проактивного ограничения скоростей потоков класса «Слоны» и «Мыши», минимизировать кол-во сбросов скользящего окна для каждого потока, обеспечивая тем самым использование пропускной способности канала близкой к максимальной.

**Методология и методы исследования.** В работе проведено экспериментальное исследование свойств информационных потоков в высокоскоростных магистральных Интернет-каналов на основе системного подхода с использованием методов математической статистики, генетических алгоритмов и методов анализа временных рядов.

**Теоретическая и практическая значимость диссертационного исследования:**

1. Создан программно-аппаратный комплекс, обеспечивающий анализ дампов Интернет-трафика, размещенных в rсар-файлов, адаптированный для использования на суперкомпьютере «Уран» Института математики и механики им. академика Н.Н. Красовского УрО РАН.
2. Проведен анализ дампов Интернет-трафика, зарегистрированных в магистральном Интернет-канале, проложенном между США и Японией, результаты которого подтвердили работоспособность разработанной методики анализа Интернет-трафика и созданного программно-аппаратного комплекса.
3. Предложен механизм балансировки объемов передаваемой информации, каждым из выделенных классов пользователей («Слоны», «Мулы», «Мыши»), устанавливающий скорость передачи информации для каждого класса пользователей, исходя из значений показателей Херста накопленных сумм зависимостей «мгновенного» числа переданных пакетов, «мгновенного» объема переданной информации, «мгновенного» объема информации, переданной одним пакетом, от времени. (Здесь и далее под «мгновенными» значения понимаются значения соответствующих параметров, подсчитанные в течение заданного временного интервала.)

### **Основные научные положения, выносимые на защиту.**

1. Для количественного описания информационных потоков в высокоскоростных магистральных Интернет-каналах целесообразно использовать зависимости «мгновенного» (вычисленного в течение выбранного временного интервала) числа переданных пакетов ( $N_i$ ), «мгновенного» объема переданной информации ( $V_i$ ), «мгновенного» объема информации, переданной одним пакетом ( $\tilde{V}_i$ ), от времени.

2. Зависимости  $N_i$ ,  $V_i$ ,  $\tilde{V}_i$ , представляющие собой случайные последовательности (СП), члены которых есть случайные величины с ограниченной областью рассеяния, вопреки устоявшимся представлениям, не являются самоподобными СП.

3. СП, вычисленные в соответствии с известными алгоритмами генерации ФБД, в которых вместо нормально распределенных случайных величин используются случайные величины ограниченной областью рассеяния, являются самоподобными.

4. Накопленные суммы ВР  $N_i$ ,  $V_i$ ,  $\tilde{V}_i$  являются самоподобными СП.

**Достоверность полученных результатов**, научных положений и выводов, изложенных в диссертации, подтверждается использованием адекватных методов анализа первичной информации и выбранных количественных показателей процесса передачи данных в высокоскоростных магистральных интернет каналах и согласованностью полученных результатов с моделью OSI (Open Systems Interconnection – Модель взаимодействия открытых систем), технологией Ethernet (технология организации сетей) на базе которых построено большинство современных сетей, а также с результатами математического моделирования фрактального броуновского движения.

**Апробация работы.** Материалы работы докладывались на следующих научных конференциях: Международной IEEE Сибирской конференции по управлению и связи SIBCON-2015, Омск, 21-23 мая 2015 г.; Международной Крымской конференции «СВЧ-техника и телекоммуникационные технологии» (КрыМи-Ко'2015), Севастополь, 6-12 сентября 2015 г.; Международной IEEE конференции АICT, Ростов-на-Дону, 14-16 октября 2015 г.; Международная научная конференция «Современные методы прикладной математики, теории управления и компьютерных технологий», Воронеж, 20-26 сентября 2016 г.; Международной научной конференции «Динамика систем, механизмов и машин» Dynamics 2016, Омск 15-17 Ноября 2016 г.; Международной IEEE конференции АICT, Москва, 20-22 сентября 2017 г.

**Публикации.** По теме диссертации опубликовано 14 работ, отражающих основные положения исследования, среди которых 9 статей в журналах, определенных ВАК и Аттестационным советом УрФУ, включая 4 статьи в изданиях, индексируемых в международных цитатно-аналитических базах Web of Science и Scopus, 3 текстов докладов в материалах международных научно-практических конференций. Получено 2 свидетельства о государственной регистрации программ для ЭВМ.

**Структура и объем диссертации.** Диссертация состоит из введения, четырех глав, заключения, списка литературы из 105 наименований, 7 приложений, содержит 42 рисунков и 55 таблиц. Основной текст работы составляет 148 страницы, общий объем – 196 страниц.

**Личный вклад автора.** Основные результаты теоретических и прикладных исследований получены автором самостоятельно. В работах, опубликованных в соавторстве, лично соискателю принадлежат: разработка технологии семантического анализа дампа трафика информационных потоков в компьютерных сетях, исследование самоподобия трафика в высокоскоростной КС.

### **СОДЕРЖАНИЕ РАБОТЫ**

**Во введении** обосновывается актуальность проблемы, рассматривается состояние исследуемого вопроса, формулируется цель работы, перечисляются основные научные результаты диссертации и её краткое содержание. Приводятся основные положения, выносимые на защиту,

сформулирована научная новизна, теоретическая и практическая значимость проведенного исследования.

**Первая глава** работы посвящена анализу состояния предметной области, в том числе: моделей и протоколов, регламентирующих передачу данных в компьютерных сетях; структуры файлов, содержащих Dump(ы) Интернет-трафика – основного источника информации о параметрах информационных потоков вычислительных сетей; современных инструментов для работы с Dump(ами) трафика и методов анализа трафика, результаты которого свидетельствуют о необходимости разработки собственного программного инструмента, обеспечивающего не только извлечение в выбранном разрезе информации из pcap-файлов, но и ее количественный анализ.

На основании результатов анализа состояния предметной области сформулированы цель и задачи исследования.

**Вторая глава** посвящена разработке математического и программного обеспечения для анализа трафика, передаваемого в высокоскоростных магистральных Интернет-каналах. В качестве объекта исследований выбраны ежедневные пятнадцатиминутные дампы трафика магистрального канала между США и Японией, находящиеся в общедоступном хранилище – MAWI (Measurement and Analysis on the WIDE Internet), в которых находится как актуальная информация о текущем состоянии Интернет-канала, а также архивные данные, начиная с 1999 г. Для сохранения конфиденциальности информации в дампе обезличены сведения о конкретных пользователях (номерах портов отправителей и получателей пакетов), а также содержание пакетов заменено на сгенерированные случайным образом данные, при этом сохранена вся необходимая для исследований техническая информация (сохранено соответствие между виртуальными номерами портов отправителя и получателя, протоколы, размер пакета и т.д.).

Проведен сравнительный анализ известных программных продуктов для обработки дампов(ов) Интернет-трафика, результаты которого представлены в таблице 1.

Таблица 1

Сравнительная таблица средств математической обработки данных

<i>Критерии сравнения</i>	<i>Наименования программных пакетов</i>			
	<i>Scilab</i>	<i>Octave</i>	<i>Sage</i>	<i>MATLAB</i>
Возможность построения графиков	+	Использует дополнительные программы: Gnuplot и Grace	+	+
Возможность имитационного моделирования	+	–	–	+
Функционал	Большое количество встроенных математических функций	Большое количество встроенных математических функций	Ограниченное количество встроенных математических готовых	Большое количество встроенных функций математических и пакетов расширений
Максимально возможный размер системы линейных алгебраических уравнений	256×256	2048×2048	1024×1024	2048×2048
Возможность компилирования других языков	C, C++, Fortran	C, C++, Fortran	Phyton	C, C++, Visual C, Fortran

Критерии сравнения	Наименования программных пакетов			
	Scilab	Octave	Sage	MATLAB
Удобство установки и настройки	Не требует дополнительных настроек после установки	Необходима дополнительная настройка для запуска графического интерфейса	Необходима дополнительная настройка шрифтов для корректного отображения в графическом окне	Не требует дополнительных настроек после установки
Совместимость с операционными системами	Linux, MacOS X, Windows	Linux, UNIX, Cygwin, Windows, Android	Linux, Unix, Windows	Microsoft Windows, Mac OS, Linux, Unix
Разработчик	Scilab Enterprises	John W. Eaton	Уильям Стей	The MathWorks
Техническая поддержка	Нерегулярные обновления. Получение информации на форумах и в сообществах.	Нерегулярные обновления. Получение информации на форумах и в сообществах.	Нерегулярные обновления. Получение информации на форумах и в сообществах.	Регулярные обновления. Получение информации онлайн по телефону, а также на форумах и в сообществах.
Стоимость	Бесплатно. Лицензия GNU GPL	Бесплатно. Лицензия CeCILL (совместима с GNU GPL v.2)	Бесплатно. Лицензия GNU GPL	1500 рублей для студентов. Лицензия проприетарная

Из табл. 1 видно, что наиболее перспективными с точки зрения минимизации финансовых затрат на их приобретение являются свободно распространяемые пакеты Scilab, Octave, Sage. Однако у данных пакетов отсутствует техническая поддержка, кроме того, существуют проблемы совместимости с некоторыми языковыми таблицами (кодировками), в них, например, в сравнении с пакетом MATLAB, который сегодня активно используется как членами научного сообщества, так и профессионалами различных сферах человеческой деятельности, в них реализовано значительно меньше математических функций, пакетов-расширений, дополнений. Основным недостатком пакета MATLAB состоит в том, что он является платным коммерческим продуктом. Однако для студентов и аспирантов плата за использование пакета MATLAB, работающего под управлением операционных систем (ОС) семейства Windows является вполне приемлемой (45\$), а соответствующая версия пакета, работающая под управлением ОС семейства UNIX, является бесплатной. В этой связи в качестве базового программного средства, используемого для анализа дампа Интернет-трафика в высокоскоростном магистральном канале, был выбран пакет MATLAB.

Также на основе анализа программных инструментов для математической обработки данных, лингвистического анализа, сетевых анализаторов и операционных систем обоснован подход к разработке лингвистического анализатора дампа трафика, который предложено реализовать в виде самостоятельной мех-функции MATLAB (мех-функции, используемые в MATLAB для вызова подпрограмм, написанных на языках С или Фортран, аналогично функциям, включенным в ядро пакета MATLAB) matshark; выбор оптимальной операционной системы и технических характеристик оборудования, необходимых для анализа damp(ов) Интернет-трафика, и программного обеспечения:

1. операционная система – Ubuntu Server 12.04.4 LTS;
2. инструмент математического анализа – MATLAB 2013b;
3. Сетевой анализатор – Wireshark 1.10.05;
4. ЭВМ следующей конфигурации:
  - 4.1. процессор Intel Xeon X5670 2.93GHz (2ядра);

- 4.2. 16 Гб оперативной памяти;
- 4.3. жесткий диск объёмом 500 Гб.

Выбранный подход обеспечил использование в качестве готовых списков параметров запросов фильтров, реализованных в программных инструментах tcpdump или Wireshark, поскольку списки параметров фильтров данных программных продуктов покрывают большинство возможных запросов к файлам дампов Интернет-трафика. Структурная схема работы реализованной мех-функции, представлена на рис. 1.

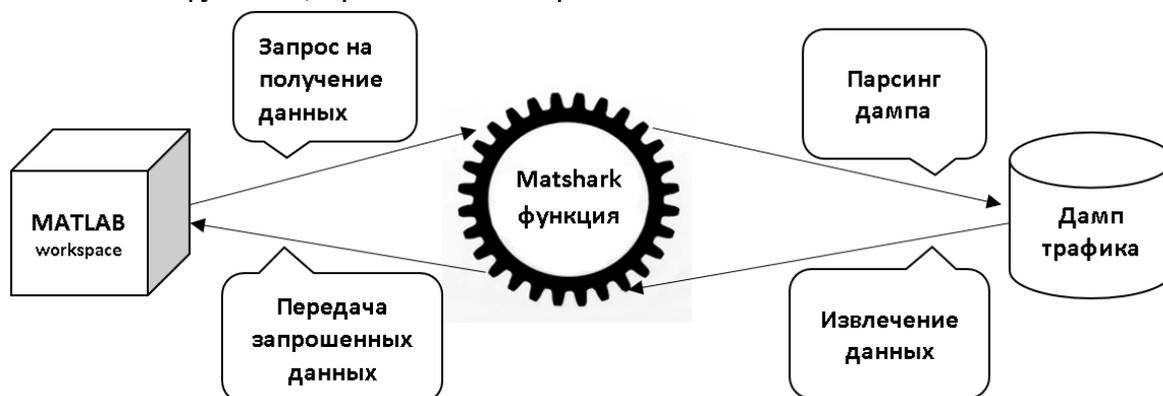


Рис. 1. Организация импорта данных из дампа в среду обработки MATLAB

Блок-схема алгоритма, реализованного в данной мех-функции *matshark*, представлена на рис. 2. Авторские права на данный программный инструмент защищены государственным свидетельством о регистрации программы для ЭВМ № 2015611426.

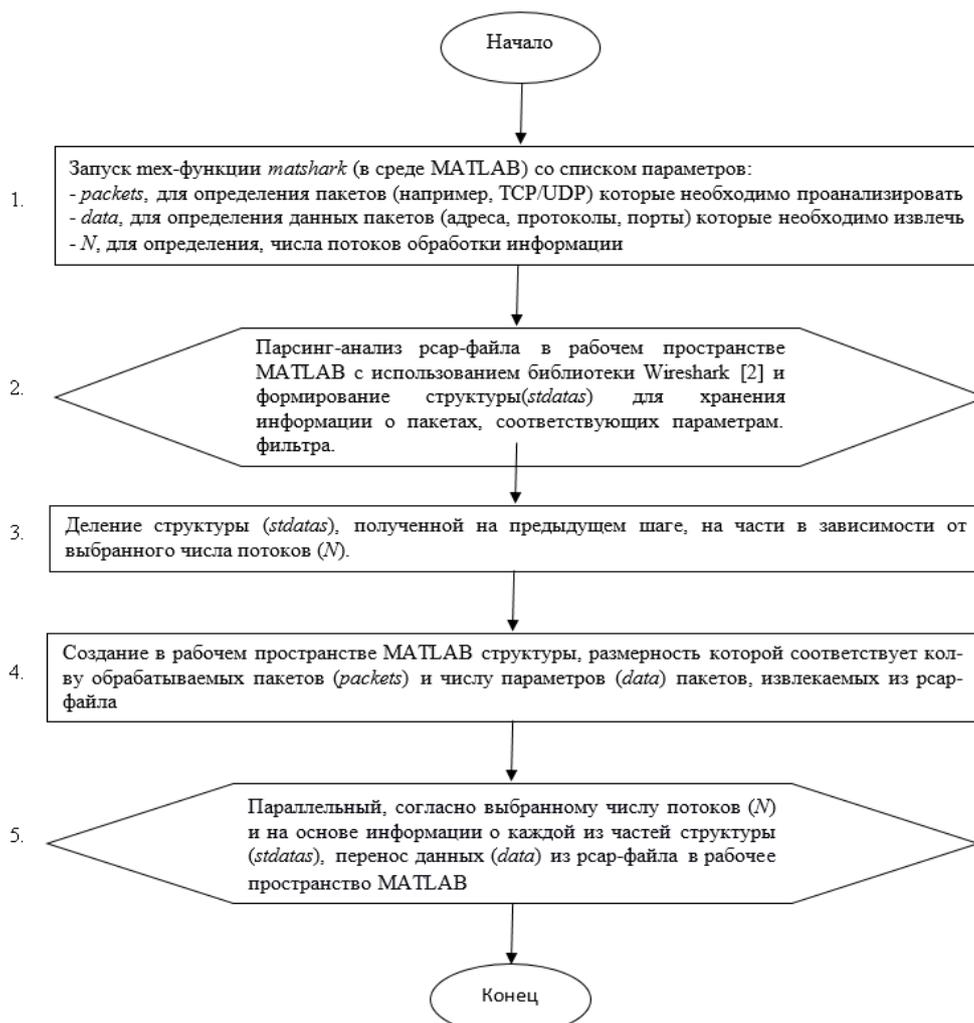


Рис. 2. Блок-схема алгоритма, реализованного в мех-функции *matshark*

Работоспособность разработанной тех-функции matshark подтверждена результатами анализа Интернет-трафика из MAWI-архива, зарегистрированного 27.10.14 (файл 201410271400.dump), представленного на рис. 3.

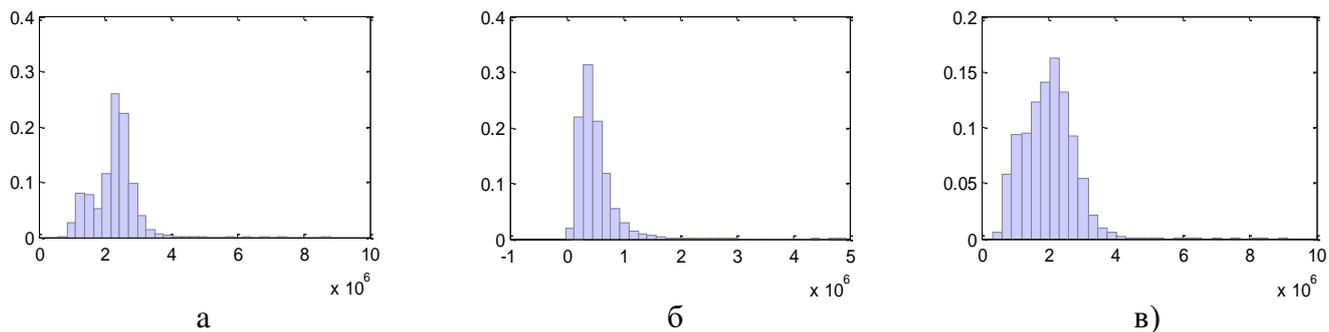


Рис. 3. Гистограммы распределений «мгновенных» объёмов данных, переданных 27.10.14 потоками различных размеров за 0,1 с:

а) – потоками объемом до 0,3 Мбайт («Мыши»), б) – потоками объемом от 0,3 до 10 Мбайт («Мулы»), в) – потоками объемом более 10 Мбайт («Слоны»)

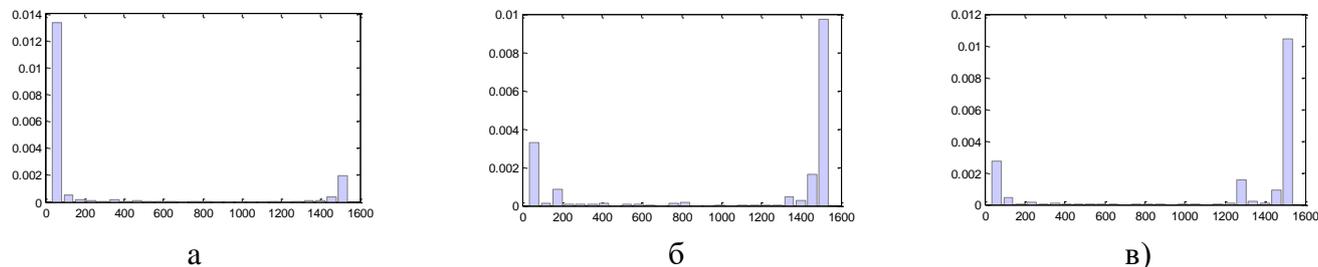


Рис. 4. Гистограммы распределений размеров пакетов, переданных потоками различных размеров в течении 15 минут:

а) – «Мышами», б) – «Мулами», в) – «Слонами»

Далее в главе 2 предложена научно обоснованная методика выделения из дампов информации, относящейся к информационным потокам (поток трафика – совокупность данных, переданная приложением, инициировавшим данный поток, например, запрос клиента к серверу, как в прямом, так и в обратном направлениях), создаваемых выбранными классами пользователей: «Слоны», «Мулы», «Мыши», предусматривающая на первом этапе семантический анализ Интернет-трафика, на втором этапе – идентификацию потоков. Блок-схема данной методики представлена на рис. 5, блок-схема процесса переноса данных из pcap-файла в рабочее пространство пакета MATLAB – на рис. 6.

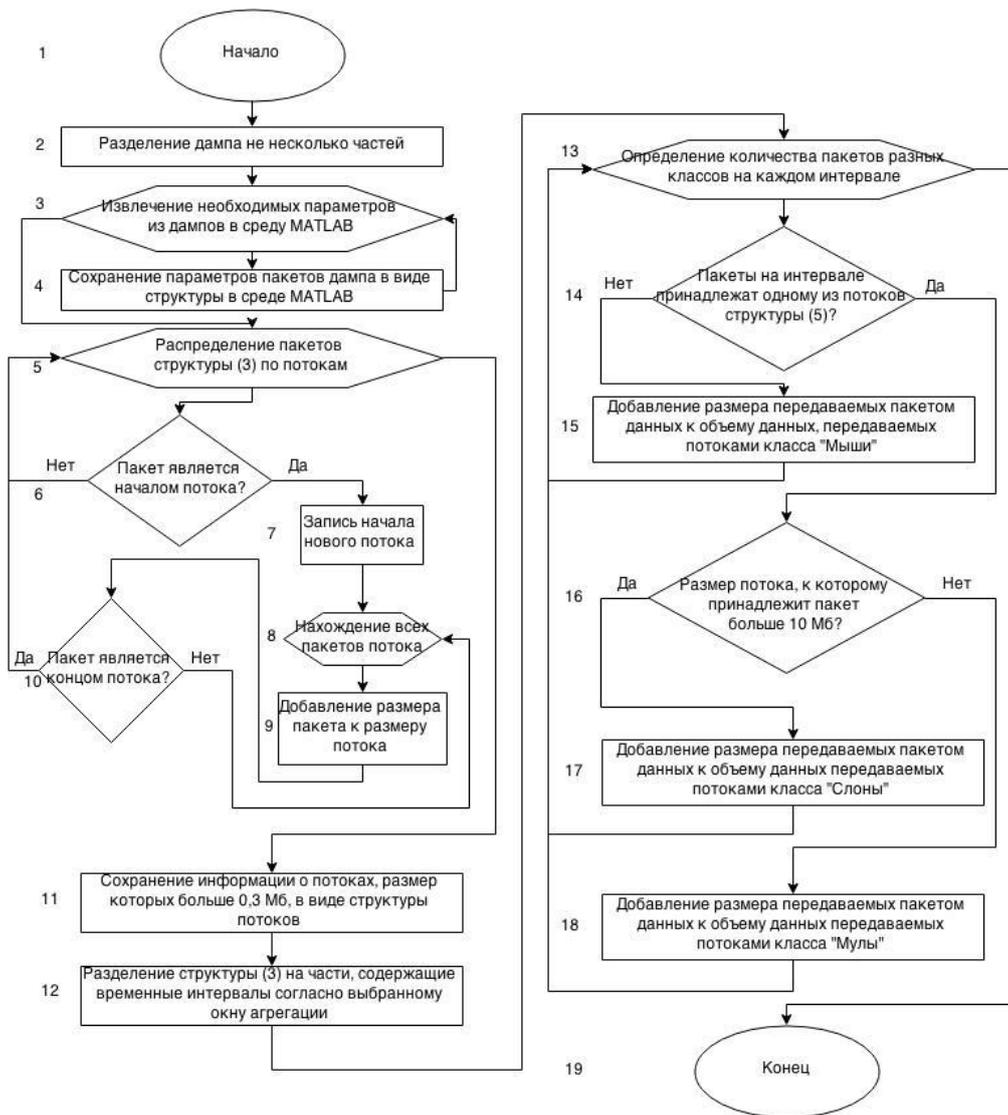


Рис. 5. Блок-схема методики выделения из дампа Интернет-трафика информации потоков, относящейся к информационным потокам, создаваемых «Мышами», «Мулами» и «Слонами»



Рис. 6. Схема процесс переноса данных из pcap-файла в рабочее пространство пакета MATLAB

В третьей главе изложены результаты исследования особенностей информационных потоков в магистральном Интернет-канале, в котором изучались СП, содержащие «мгновенные» объемы информации, переданной данным классом пользователей,  $V_i^{(\tau)}$ , на последовательных во временных интервалах длительностью  $\tau = 0,1; 1,0$  с, а также аналогичные СП  $N_i^{(\tau)}$ , содержащие «мгновенное» число пакетов, переданных данным классом пользователей. Анализ СП  $V_i^{(0,1)}$ ,  $N_i^{(0,1)}$  показал, что изученные СП относятся к классу СП с ограниченной областью рассеяния, например, область рассеяния СП  $V_i^{(0,1)}$ : у «Мышей» –  $[0,8; 8,0] \cdot 10^6$  байт, у «Мулов» –  $[0,3,0] \cdot 10^6$  байт, у «Слонов» –  $[0; 8,0] \cdot 10^6$  байт; область рассеяния СП  $N_i^{(0,1)}$ : у «Мышей» –  $[4800; 14900]$ ; у «Мулов» –  $[0; 4000]$ ; у «Слонов» –  $[800; 10000]$ .

В этой связи для аппроксимации плотностей распределений изучаемых случайных последовательностей был использован метод мнимых источников, в соответствие с которым плотность распределения вычисляется по формуле:

$$f_{LAD}(x; \mu, \sigma, l) = A \left[ \varphi(x; \mu, \sigma, l) + \sum_{n=0}^{\infty} \varphi_{2n+1}^{\pm}(x; \mu, \sigma, l) + \sum_{n=1}^{\infty} \varphi_{2n}^{\pm}(x; \mu, \sigma, l) \right], \quad 1)$$

где  $A$  – нормировочный коэффициент, определяемый из условия

$$\int_a^b f_{LAD}(\xi; \mu, \sigma, l) d\xi = 1, \quad 2)$$

$$\varphi(x; \mu, \sigma, l) = \exp\left[-(x - \mu)^2 / 2\sigma^2\right],$$

$$\varphi_{2n+1}^{\pm}(x; \mu, \sigma, l) = \exp\left[-(x - x_{2n+1}^{\pm})^2 / 2\sigma^2\right],$$

$$\varphi_{2n}^{\pm}(x; \mu, \sigma, l) = \exp\left[-(x - x_{2n}^{\pm})^2 / 2\sigma^2\right],$$

$$x_{2n}^{\pm} = \pm 4nl + \mu, \quad x_{2n+1}^{\pm} = \pm (4n + 2)l - \mu - \text{координаты центров фиктивных источников;}$$

$l$  – размер области рассеяния,

$\mu$  – математическое ожидание случайной величины при отсутствии области ограничения,

$\sigma$  – СКО случайной величины при отсутствии области ограничения.

Здесь значения параметров  $\mu$  и  $\sigma$  являются в зависимости от выбранной меры близости эмпирической и расчетной ПР решением той или иной системы нелинейных уравнений, которое может быть найдено с помощью известных численных методов решения систем нелинейных уравнений и генетических алгоритмов и метода аппроксимации Розенблатта-Парзена.

Анализ дампов Интернет-трафика, зарегистрированных в период с 27.10.14 по 02.11.14 показал, что СП  $V_i^{(\tau)}$ ,  $N_i^{(\tau)}$  в разные дни вне зависимости от размера окна агрегации  $\tau$  принадлежат различным генеральным совокупностям, отличающимся друг от друга значениями параметров  $\mu, \sigma$ . При этом максимальные вариации параметров значений  $\mu_i, \sigma_i, i = \overline{1, 7}$  относительно значений  $\bar{\mu}, \bar{\sigma}$  составили:

1. у «Мышей»:  $\Delta\mu_{0,1} = 17,25\%$ ,  $\Delta\sigma_{0,1} = 32,70\%$ ;  $\Delta\mu_{1,0} = 17,23\%$ ,  $\Delta\sigma_{1,0} = 59,77\%$ ;

2. у «Мулов»:  $\Delta\mu_{0,1} = 77,74\%$ ,  $\Delta\sigma_{0,1} = 28,93\%$ ;  $\Delta\mu_{1,0} = 68,95\%$ ,  $\Delta\sigma_{1,0} = 16,44\%$ ;

3. у «Слонов»:  $\Delta\mu_{0,1} = 43,09\%$ ,  $\Delta\sigma_{0,1} = 58,04\%$ ;  $\Delta\mu_{1,0} = 43,41\%$ ,  $\Delta\sigma_{1,0} = 68,46\%$ ,

для  $\tau = 0.1$  и  $1.0$  с, соответственно (здесь значение нижнего индекса соответствует размеру окна агрегации  $\tau$ ). Таким образом, вне зависимости от размера окна агрегации  $\tau$  наиболее стабильными значения параметров распределения  $\mu, \sigma$  оказались у «Мышей», наименее стабильными – у «Мулов».

Значения  $\bar{\mu}_{1,0}, \bar{\sigma}_{1,0}$  и  $\bar{\mu}_{0,1}, \bar{\sigma}_{0,1}$  оказались связанными между собой следующими соотношениями:

1. «Мыши»:  $\bar{\mu}_{1,0} \approx 9,99\bar{\mu}_{0,1}$ ,  $\bar{\sigma}_{1,0} \approx 7.35\bar{\sigma}_{0,1}$ ;

2. «Мулы»:  $\bar{\mu}_{1,0} \approx 10,52\bar{\mu}_{0,1}$ ,  $\bar{\sigma}_{1,0} \approx 7.25\bar{\sigma}_{0,1}$ ;

3. «Слоны»:  $\bar{\mu}_{1,0} \approx 9,99\bar{\mu}_{0,1}$ ,  $\bar{\sigma}_{1,0} \approx 7.94\bar{\sigma}_{0,1}$ .

Из приведенных соотношений видно, что использованная процедура агрегации размера потоков не является масштабно-инвариантным преобразованием по параметру  $\sigma$ , а потому статистические свойства СП, содержащих значения объемов переданной информации каждым из выбранных классов пользователей в течении данного временного интервала оказываются зависящими от размера окна агрегации  $\tau$ .

Максимальные отклонения значений параметров  $\Delta\mu_{\tau,k} = \max_i \left( \left| \mu_{\tau,i,k} - \bar{\mu}_{\tau,i} \right| / \bar{\mu}_{\tau,i} \right)$ ,

$\Delta\sigma_{\tau,k} = \max_i \left( \left| \sigma_{\tau,i,k} - \bar{\sigma}_{\tau,i} \right| / \bar{\sigma}_{\tau,i} \right)$  для выбранных классов потоков составили:

1. «Мыши»:

$$\Delta\mu_{0,1k=1,2} = 9,73\%, \quad \Delta\sigma_{0,1k=1,2} = 19,52\%; \quad \Delta\mu_{0,1k=\overline{1,4}} = 55,32\%, \quad \Delta\sigma_{0,1k=\overline{1,4}} = 30,69\%;$$

$$\Delta\mu_{1,0k=1,2} = 9,72\%, \quad \Delta\sigma_{1,0k=1,2} = 25,79\%; \quad \Delta\mu_{1,0k=\overline{1,4}} = 41,78\%, \quad \Delta\sigma_{1,0k=\overline{1,4}} = 83,48\%;$$

2. «Мулы»:

$$\Delta\mu_{0,1k=1,2} = 49,12\%, \quad \Delta\sigma_{0,1k=1,2} = 23,47\%; \quad \Delta\mu_{0,1k=\overline{1,4}} = 69,81\%, \quad \Delta\sigma_{0,1k=\overline{1,4}} = 58,15\%;$$

$$\Delta\mu_{1,0k=1,2} = 46,50\%, \quad \Delta\sigma_{1,0k=1,2} = 40,38\%; \quad \Delta\mu_{1,0k=\overline{1,4}} = 45,22\%, \quad \Delta\sigma_{1,0k=\overline{1,4}} = 61,28\%;$$

3. «Слоны»:

$$\Delta\mu_{0,1k=1,2} = 19,16\%, \quad \Delta\sigma_{0,1k=1,2} = 19,87\%; \quad \Delta\mu_{0,1k=\overline{1,4}} = 28,34\%, \quad \Delta\sigma_{0,1k=\overline{1,4}} = 26,70\%;$$

$$\Delta\mu_{1,0k=1,2} = 20,39\%, \quad \Delta\sigma_{1,0k=1,2} = 31,04\%; \quad \Delta\mu_{1,0k=\overline{1,4}} = 28,14\%, \quad \Delta\sigma_{1,0k=\overline{1,4}} = 42,77\%.$$

Таким образом, наиболее вариативным, оказывается параметр  $\mu$  у «Мулов» даже при его оценке по СП  $V_i^{(0.1)}$ ,  $V_i^{(1.0)}$ , зарегистрированным в один и тот же день.

Анализ результатов вычисления параметров распределения СП  $V_i^{(0.1)}$ ,  $V_i^{(1.0)}$ , позволил сделать следующие выводы.

1. СП  $V_i^{(0.1)}$ ,  $V_i^{(1.0)}$  вне зависимости от размера окна агрегации являются СП с ограниченными областями рассеяния.

2. Для аппроксимации ПР изученных СП  $V_i^{(0.1)}$ ,  $V_i^{(1.0)}$  целесообразно использовать обобщенный закон распределения нормальной с ограниченной областью рассеяния.

3. Оценки параметров ПР СП  $V_i^{(0.1)}$ ,  $V_i^{(1.0)}$  не являются стационарными величинами, но могут меняться как день ото дня, так и в течение одного дня.

4. Наиболее стабильными параметры ПР изученных СП  $V_i^{(0.1)}$ ,  $V_i^{(1.0)}$ , как в течение одного дня, так и недели, оказываются у потоков, создаваемых классом пользователей «Мыши».

5. Наименее стабильные параметры ПР изученных СП  $V_i^{(0.1)}$ ,  $V_i^{(1.0)}$ , как в течение одного дня, так и в течение недели, имеют потоки классов «Мулы» и «Слоны», у которых отклонения параметра  $\mu$  от среднего значения в течение недели могут достигать 60% и 40%, соответственно.

Анализ результатов вычисления параметров распределения СП  $N_i^{(0.1)}$ ,  $N_i^{(1.0)}$ , позволил сделать следующие выводы:

1. Число СП  $N_i^{(0.1)}$ ,  $N_i^{(1.0)}$ , созданным каждым из классов пользователей вне зависимости от размера окна агрегации являются СП с ограниченными областями рассеяния.

2. Оценки параметров ПР СП  $N_i^{(0.1)}$ ,  $N_i^{(1.0)}$  не являются стационарными величинами, но могут меняться по дням, так и в течение одного дня.

3. Наиболее стабильные параметры ПР СП  $N_i^{(0.1)}$ ,  $N_i^{(1.0)}$ , как в течение одного дня, так и недели, оказываются у класса потоков «Мыши».

4. Наименее стабильные параметры ПР СП  $N_i^{(0.1)}$ ,  $N_i^{(1.0)}$ , как в течение одного дня, так и в течение недели, имеют потоки классов «Мулы» и «Слоны», у которых отклонения параметра  $\mu$  от среднего значения в течение недели могут достигать 60% и 40%, соответственно.

Также была исследована стационарность статистических характеристик СП  $V_i^{(\tau)}$ ,  $N_i^{(\tau)}$  для каждого из выбранных классов пользователей. Для этого была использована методика, реализующаяся следующей последовательностью действий:

1. Для каждой СП  $N_i^{(0.1)}$ ,  $N_i^{(1.0)}$  вычислялись ПР в соответствие с методом мнимых источников и аппроксимации Розенблатта-Парзена.

2. Каждая СП  $N_i^{(0.1)}$ ,  $N_i^{(1.0)}$  делилась на 2 части и для каждой из полученных СП2,1, СП2,2 выполнялся п.1.

3. Каждая СП  $N_i^{(0,1)}, N_i^{(1,0)}$  делилась на 4 части и для каждого из полученных СП4,1–СП4,4 выполнялся п.1.

Для каждого класса потоков в каждый из дней были получены по 7 ПР СП  $N_i^{(\tau)}$  для окон агрегации размером 0,1 и 1,0 с.

Результаты изучения стационарности статистических характеристик СП  $V_i^{(\tau)}, N_i^{(\tau)}$  позволяют сделать следующие выводы.

1. При  $\tau=0,1$  с. гипотеза о возможности аппроксимации ПР изученных СП  $N_i^{(0,1)}, V_i^{(0,1)}$  (объемы данных и кол-во пакетов) с помощью метода мнимых источников в большинстве случаев отвергается критерием типа Колмогорова-Смирнова.

2. При  $\tau=0,1$  с. аппроксимации ПР СП  $N_i^{(0,1)}, V_i^{(0,1)}$ , построенные для частей ВР (объемы данных и кол-во пакетов) как с помощью метода мнимых источников, так и с помощью аппроксимации Розенблатта-Парзена, изученных ВР оказываются отличными друг от друга, что свидетельствует о нестационарности СП  $N_i^{(0,1)}, V_i^{(0,1)}$ .

3. При использовании окна агрегации 1,0 с. для аппроксимации ПР методом мнимых источников изученных СП  $N_i^{(0,1)}, N_i^{(1,0)}$  обеспечивается удовлетворительное качество аппроксимации.

Проведенный анализ Интернет-трафика в разрезе размера пакетов, переданных каждым из классов пользователей, показал:

- в информационных потоках, создаваемых, классом «Мыши» преобладающими являются пакеты размером не более 200 байт;
- в информационных потоках, создаваемых, классами «Мулы» и «Слоны» преобладающими являются пакеты размером  $\sim 1500$  байт;
- информационные потоки, создаваемые каждым из классов пользователей, на пятнадцатиминутных интервалах имеют стационарные во времени распределения по размеру передаваемых пакетов.

Наиболее стабильные во времени параметры распределения информационных потоков по размеру пакетов в течение недели оказались у класса пользователей «Слоны» и «Мулы», наименее стабильные – у пользователей класса «Мыши». Данный результат объясняется тем, что «Слоны» и «Мулы» передают пакеты отличные от максимального размера только во время инициации и завершения соединения, далее для эффективного использования полосы пропускания передаются пакеты максимального размера, «Мыши» же в свою очередь при передаче информации в большинстве случаев используют пакеты меньших отличных друг от друга размеров.

**Четвертая глава** содержит результаты проверки по экспериментальным результатам гипотезы о самоподобии Интернет-трафика. Для этого получены оценки показателя Херста СП  $N_i^{(\tau)}, V_i^{(\tau)}$  по зависимости накопленной дисперсии приращений обобщенного броуновского движения от длины приращения (алгоритм HСALC) для каждого СП  $N_i^{(\tau)}, V_i^{(\tau)}$  каждого из выбранных классов пользователей по полным последовательностям («глобальное» значение показателя Херста),  $\tau=0,1$  и 1,0 с, а также усредненные по ансамблю, содержащему оценки показателя Херста, вычисленные на 9-ти непересекающихся частях СП.

При этом использовались следующие параметры алгоритма HСALC:

- длина приращения  $P_{\max} = 20$ ;
- число отчетов, по которым осуществляется вычисление накопленной дисперсии,  $L=9000$  для полного ВР и  $L=1000$  для частей ВР.
- координата вектора, начиная с которой вычисляется накопленная дисперсия,  $N_{start} = 1$

для полного ВР и  $N_{start} = 1, 8001$  с шагом в тысячу для частей ВР.

Также были вычислены средние за весь период наблюдений значения глобальных и усредненных значений показателей Херста СП  $N_i^{(\tau)}, V_i^{(\tau)}$ . Оказалось, что показатель Херста СП  $N_i^{(\tau)}, V_i^{(\tau)}$  является некоторой случайной величиной, изменяющейся как в течение недели (27.10.14–02.11.14), так и в течение одного дня. При этом его значения оказались существенно зависящими от размера окна агрегации, что также является свидетельством, подтверждающим обоснованность отклонения гипотезы о самоподобии СП  $N_i^{(\tau)}, V_i^{(\tau)}$ .

Для объяснения причин обнаруженных случайных изменений значений показателя Херста были изучены статистические свойства первых разностей 4200 частичных последовательностей, выбираемых из СП  $N_i^{(\tau)}, V_i^{(\tau)}$  (3 класса потоков  $\times$  2 случайные последовательности (объем информации и кол-во пакетов)  $\times$  7 дней (27.10.14–02.11.14)  $\times$  100 окон агрегации (от 0,1 до 10 с) = 4200). Для каждой частичной СП были вычислены первые разности и для них проверялась статистическая гипотеза о нормальности их распределений с помощью критерия типа Колмогорова-Смирнова и критерия  $\chi^2$ . Далее была составлена бинарная таблица, в которой при принятии статистической гипотезы о нормальности закона распределения частичной СП устанавливалось значение равное 1 и 0 в противоположном случае. Затем по ней были вычислены относительные доли данного типа частичных последовательностей, для которых гипотеза о нормальности распределения первых разностей была принята. Зависимости относительной доли СП, у которых первые разности имеют нормальный закон распределения  $D_{\tau}^{[V]}$ , от размера окна агрегации представлены на рис. 7.

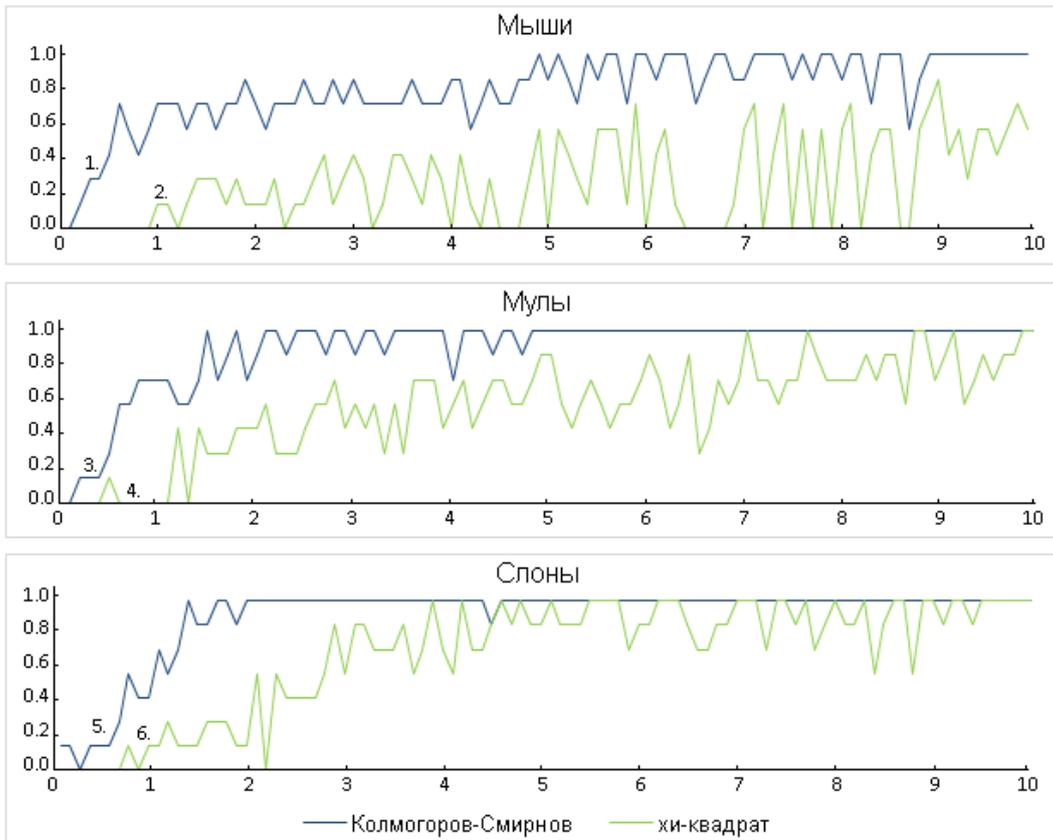


Рис. 7. Зависимости относительной доли  $D_{\tau}^{[V]}$  СП, у которых первые разности имеют нормальный закон распределения (1,3,5 – критерий Колмогорова-Смирнова, критерий 2,4,6 – Хи-квадрат), от размера окна агрегации (0,1–10 с)

Анализ плотностей распределений и выборочных плотностей СП  $N_i^{(t)}, V_i^{(t)}$  показал, что для их описания можно использовать модель случайных броуновских блужданий в ограниченной области рассеяния, ПР и ФР, которого вычисляются в соответствии с (1) и (2). Данный результат позволяет выдвинуть гипотезу о том, что свойствами ФБД могут обладать накопленные суммы СП  $N_i^{(t)}, V_i^{(t)}$ . Для подтверждения справедливости данной гипотезы или ее отклонения необходимо была проверена гипотеза о возможности генерации фрактального броуновского движения в случае использования случайных величин с ограниченной областью рассеяния. Здесь для генерации случайных последовательностей был использован известный алгоритм генерации ФБД на основе Фурье-фильтрации, в котором вместо случайных чисел с нормальным законом распределения были использованы случайные числа, генерируемые в соответствии с распределением (1). Анализ показателей Херста синтезированных СП подтвердил гипотезу о том, что, накопленные суммы СП, сгенерированные с помощью алгоритма Фурье-фильтрации и использования случайных чисел с ограниченной областью рассеяния, действительно, являются самоподобными случайными процессами. Таким образом, вопреки общепринятой точки зрения, для анализа самоподобных свойств Интернет-трафика следует использовать накопленные суммы СП  $N_i^{(t)}, V_i^{(t)}$ .

Наличие разработанной методика анализа информационных потоков дампов Интернет-трафика в магистральном Интернет-канале позволила комплексное исследование особенностей Интернет-трафика, передаваемого в магистральном Интернет-канале, и выявить следующие закономерности:

1. Вне зависимости от размера окна агрегации размеры потоков, создаваемые каждым из рассмотренных классов вне зависимости от размера окна агрегации, а также число пакетов, переданных каждым из классов пользователей, являются СП с ограниченными областями рассеяния.

2. Оценки параметров распределения не являются стационарными величинами. Наиболее стабильные параметры ПР, как в течение одного дня, так и недели, оказываются у класса потоков «Мыши». Наименее стабильные параметры распределений, как в течение одного дня, так и в течение недели, имеют потоки классов «Мулы» и «Слоны», у которых отклонения параметра  $\mu$  от среднего значения в течение недели могут достигать 60% и 40%, соответственно.

3. При использовании окна агрегации 0,1 с. аппроксимации ПР, построенные для частей ВР  $(N_i, V_i)$ , как с помощью метода мнимых источников, так и с помощью аппроксимации Розенблатта-Парзена, изученных СП оказываются отличными друг от друга, что свидетельствует о нестационарности данных СП. При использовании окна агрегации 1,0 с. для аппроксимации ПР изученных ВР методом мнимых источников на различных временных интервалах параметры аппроксимирующих функций оказываются отличными друг от друга, что свидетельствует о том, что даже для больших значений окон агрегации Интернет-трафик является нестационарным процессом.

4. Информационные потоки, создаваемые каждым из классов пользователей, на пятнадцатиминутных интервалах имеют стационарные во времени распределения по размеру передаваемых пакетов. Наиболее стабильные во времени параметры распределения информационных потоков по размеру пакетов в течение недели оказываются у класса пользователей «Слоны» и «Мулы», наименее стабильные – у пользователей класса «Мыши».

5. Полученные оценки значений показателя Херста случайных последовательностей  $N_i^{(t)}, V_i^{(t)}$  и их первых разностей, позволяют сделать обоснованный вывод о несоответствии изученных процессов модели фрактального броуновского движения. Анализ плотностей распределений и выборочных плотностей случайных последовательностей  $N_i^{(t)}, V_i^{(t)}$  показал, что для их описания следует использовать модель случайных броуновских блужданий с ограниченной областью рассеяния.

Для выявления возможных взаимосвязей информационных потоков, передаваемых в магистральном Интернет-канале, были использованы показатели Херста накопленных сумм СП

$N_i, V_i$  и  $\tilde{V}_i$  – СП, состоящей из отношений соответствующих членов СП  $N_i, V_i$ . Физический смысл СП  $\tilde{V}_i$  – средний объем переданной информации, переданной одним пакетом.

В ходе проведенных исследований проверялась гипотеза о том, что взаимосвязь информационных потоков, передаваемых в магистральном Интернет-канале, проявляется в наличии зависимостей показателей Херста каждого из классов пользователей как от интегральных характеристик (вычисляемых по всему пятнадцатиминутному дампу Интернет-трафика информационного потока, создаваемого данным классом пользователей: объема переданной информации  $V$ , числа переданных пакетов  $N$ , объема переданной информации одним пакетом  $\tilde{V}$ , так и информационных характеристик потоков, создаваемых другими классами пользователей, то есть:

$$\begin{aligned}
 H_V^{\{Mice\}} &= F_{V1}(V^{\{Mice\}}, V^{\{Mules\}}, V^{\{Elephants\}}), & H_V^{\{Mules\}} &= F_{V2}(V^{\{Mice\}}, V^{\{Mules\}}, V^{\{Elephants\}}), \\
 H_V^{\{Elephants\}} &= F_{V3}(V^{\{Mice\}}, V^{\{Mules\}}, V^{\{Elephants\}}), & H_N^{\{Mice\}} &= F_{N1}(N^{\{Mice\}}, N^{\{Mules\}}, N^{\{Elephants\}}), \\
 H_N^{\{Mules\}} &= F_{N2}(N^{\{Mice\}}, N^{\{Mules\}}, N^{\{Elephants\}}), & H_N^{\{Elephants\}} &= F_{N3}(N^{\{Mice\}}, N^{\{Mules\}}, N^{\{Elephants\}}), \\
 H_V^{\{Mice\}} &= F_{V4}(\tilde{V}^{\{Mice\}}, \tilde{V}^{\{Mules\}}, \tilde{V}^{\{Elephants\}}), & H_V^{\{Mules\}} &= F_{V5}(\tilde{V}^{\{Mice\}}, \tilde{V}^{\{Mules\}}, \tilde{V}^{\{Elephants\}}), \\
 H_V^{\{Elephants\}} &= F_{V6}(\tilde{V}^{\{Mice\}}, \tilde{V}^{\{Mules\}}, \tilde{V}^{\{Elephants\}}), & H_N^{\{Mice\}} &= F_{N4}(\tilde{V}^{\{Mice\}}, \tilde{V}^{\{Mules\}}, \tilde{V}^{\{Elephants\}}), \\
 H_N^{\{Mules\}} &= F_{N5}(\tilde{V}^{\{Mice\}}, \tilde{V}^{\{Mules\}}, \tilde{V}^{\{Elephants\}}), & H_N^{\{Elephants\}} &= F_{N6}(\tilde{V}^{\{Mice\}}, \tilde{V}^{\{Mules\}}, \tilde{V}^{\{Elephants\}}).
 \end{aligned}
 \tag{3}$$

Пример исходных данных, использовавшиеся для идентификации функций  $F_{V1}, F_{V2}, F_{V3}$  в проведенном исследовании, представлены в таблице 2.

Таблица 2

Исходные данные для идентификации функций  $F_{V1}, F_{V2}, F_{V3}$

День	Класс пользователей	Объем переданной информации	$H_V^{\{0.1\}}$	$H_V^{\{1.0\}}$	$\bar{H}_V^{\{0.1\}}$	$\bar{H}_V^{\{1.0\}}$
1	Мыши	30996017241	0,926	0,98	0,911	0,973
	Мулы	8118004515	0,852	0,948	0,840	0,944
	Слоны	37638020936	0,936	0,981	0,915	0,974
2	Мыши	28494531094	0,957	0,986	0,944	0,978
	Мулы	7462853381	0,8500	0,948	0,834	0,942
	Слоны	34600502042	0,907	0,974	0,893	0,967
3	Мыши	22566213503	0,864	0,962	0,835	0,950
	Мулы	6769864050	0,810	0,945	0,814	0,943
	Слоны	29336077554	0,825	0,946	0,818	0,939
4	Мыши	26623867248	0,828	0,949	0,801	0,942
	Мулы	7987160174	0,856	0,960	0,843	0,953
	Слоны	34611027422	0,879	0,967	0,864	0,961
5	Мыши	24346132026	0,824	0,954	0,793	0,941
	Мулы	6376367911	0,860	0,952	0,839	0,944
	Слоны	29563160317	0,893	0,968	0,868	0,956
6	Мыши	18022343530	0,811	0,944	0,775	0,928
	Мулы	4095987165	0,852	0,956	0,835	0,944
	Слоны	20479935829	0,917	0,979	0,865	0,958
7	Мыши	18424280953	0,784	0,938	0,758	0,923
	Мулы	4187336580	0,823	0,948	0,812	0,940
	Слоны	20936682901	0,873	0,964	0,862	0,956

На первом этапе были построены многомерные линейные модели зависимостей вида

$$z = a_1x + a_2y + a_3z,$$

коэффициенты которых при соответствующих переменных есть коэффициенты корреляции между показателям Херста накопленных сумм СП  $N_i, V_i$ , представленных в таблице 5, и количеством переданных пакетов (СП  $N_i$ ), объемами переданных данных (СП  $V_i$ ), средними объемами переданных данных на пакет (СП  $\tilde{V}_i$ ). Выбор данного вида модели обусловлен очевидными физическими соображения, состоящими в том, что значение показателя Херста при  $x = y = z \equiv 0$ , то есть отсутствии информационных потоков, должно равняться 0. Полученные результаты позволили сделать следующие выводы:

1. Переменная, соответствующая объему информации, переданной классом «Слоны» отсутствует в регрессионных моделях, описывающих зависимости между значениями показателей Херста накопленных сумм СП  $V_i$  и объемами переданной информации каждым из рассматриваемых классов пользователей, а также в регрессионных моделях, описывающую зависимости показателя Херста накопленных сумм СП  $N_i$  от объемов переданной информации каждым из классов пользователей, что свидетельствует о наличии близкой к функциональной зависимости между объемами переданной информации классами «Мыши», «Мулы» и «Слоны».

2. Среднее значение скорректированного коэффициента детерминации  $R_{adj}^2 = 0,964$ , что свидетельствует о статистической значимости построенных регрессионных моделей, а также о наличии сильной зависимости между показателями Херста накопленных сумм изученных СП и объемами переданной пользователями информации.

Для объяснения причины, обусловившей отсутствие в регрессионных моделях, описывающих зависимости между значениями показателей Херста накопленных сумм СП  $V_i, N_i$  и объемами переданной информации каждым из рассматриваемых классов пользователей, переменных, соответствующих «Слонам», были изучены корреляционные связи между объемами информации, переданной «Слонами», и объемами информации, переданной «Мышами» и «Мулами».

В результате было установлено, что между значениями объемов переданной информации каждым из классов пользователей имеет место линейная зависимость вида:

$$V^{\{Elephants\}} = 0,625 \cdot V^{\{Mice\}} + 2,25 \cdot V^{\{Mules\}} \quad (4)$$

с коэффициентом детерминации  $R^2$  равным 1, что объясняет отсутствие одной из переменных в построенных регрессионных моделях. С физической точки зрения (4) является следствием «закона сохранения общего объема переданной информации» в исследованном высокоскоростном магистральном канале, являющегося суммой объемов переданной информации каждым из выбранных классов пользователей, который, как очевидно, не может превосходить некоторого максимального объема, определяющегося техническими характеристиками канала передачи. Полученный результат свидетельствует о возможности разработки механизмов балансировки объемов передаваемой информации для повышения качества обслуживания пользователей каналом (проактивного контроля загрузки канала), например, за счет управления скоростью передачи информации каждого из выделенных классов пользователей в зависимости от текущих значений характеристик информационных потоков, вычисляемых в реальном режиме времени. Для этого, принимая во внимание полученные подтверждения существования зависимостей показателей Херста накопленных сумм временных рядов  $N_i, V_i, \tilde{V}_i$ , от объемов переданной информации каждым из классов пользователей, а также соответствующих количеств переданных пакетов, представляется перспективным использование этих показателей при решении рассматриваемой задачи.

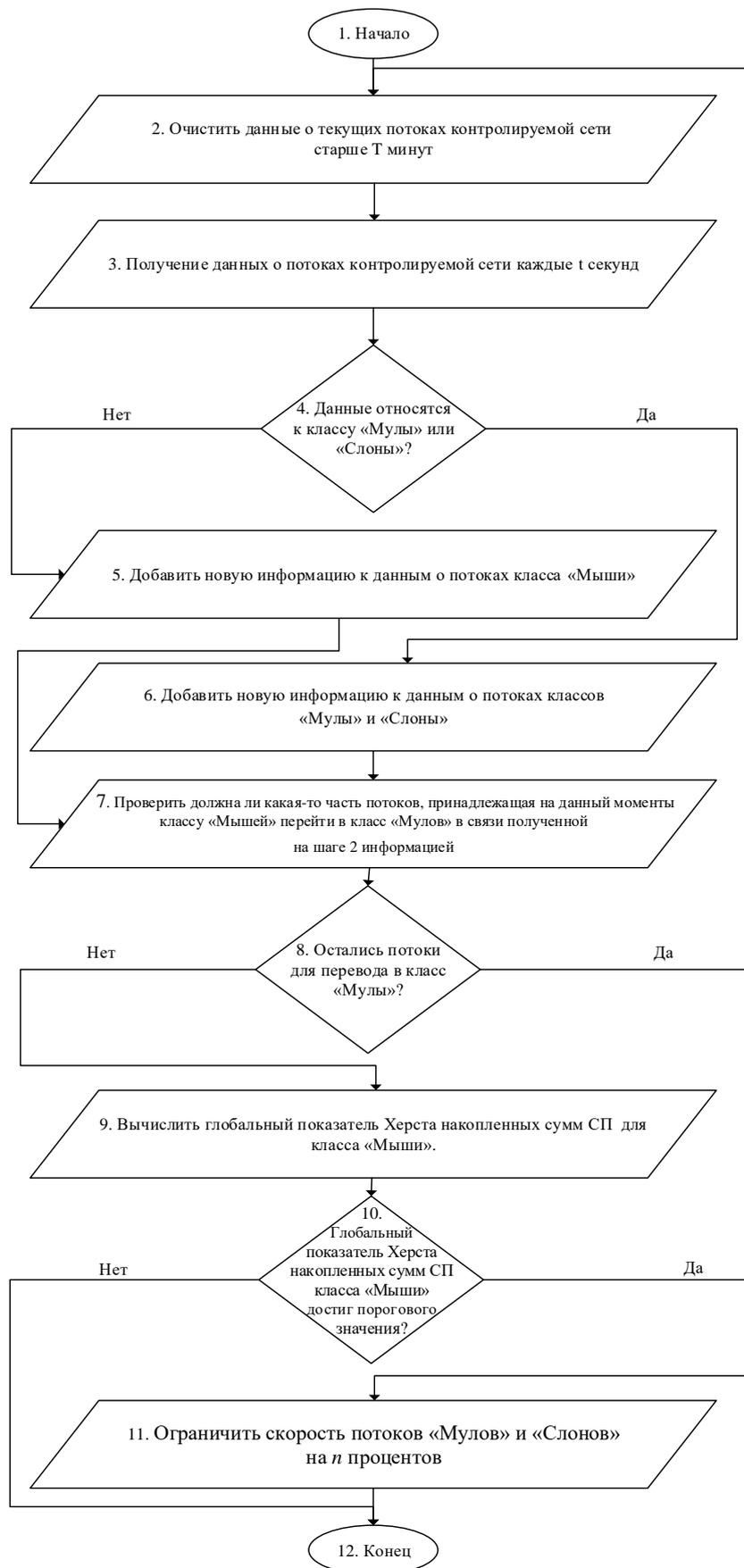


Рис. 8. Блок-схема алгоритма проактивного контроля нагрузки канала передачи информационных потоков

Блок-схема одного из возможных алгоритмов проактивного управления загрузкой канала передачи информационных потоков с целью обеспечения его максимально полной утилизации,

основанного на использования глобального показателя Херста накопленных сумм СП  $V_i$ , порожденных активностью пользователей класса «Мыши», который оказывается более чувствительным к локальным изменениям показателей балансируемых информационных потоков, представлена на рис. 4.7. Данный алгоритм обеспечивает минимизацию кол-ва сбросов скользящего окна за счет сравнения глобального показателя Херста накопленных сумм СП  $V_i$ , порожденных активностью пользователей класса «Мыши», с известным пороговым значением выбранного показателя, при превышении которого соответствующим образом регулируется скорость потоков класса «Слоны» и «Мулы». Пороговое значение глобального показателя Херста накопленных сумм СП  $V_i$ , значение скорости передачи данных для каждого из выбранных классов пользователей для каждой конкретной сети, а так же длительность временных интервалов должны определяться экспериментально. Эффективность в данном случае будет равна проценту недоутилизации сети ввиду сброса скользящего окна и может достигать до 25%. Так как логика работы скользящего окна является стандартом для сетей, использующих TCP/IP модель передачи данных (наиболее распространенный на данный момент стандарт) она может быть применена на любых сетях использующих данную модель передачи данных.

### ЗАКЛЮЧЕНИЕ

1. Проведен анализ методов исследования информационных потоков в КС сетей с точки зрения их применимости для исследования трафика в высокоскоростных магистральных Интернет-каналах, и соответствующих программных инструментов, результаты, которого подтвердили необходимость разработки специализированного математического и алгоритмического программного обеспечения, обеспечивающего автоматическое извлечение информации из рсар-файлов в выбранном измерении.

2. Создан программно-аппаратный комплекс, обеспечивающий извлечение в автоматизированном режиме из дампов трафика, передаваемого в высокоскоростном магистральном Интернет-канале, количественные его количественные характеристики, адаптированный для суперкомпьютера «Уран» ИММ УрО РАН.

3. Разработана методика анализа первичной информации, извлекаемой из рсар-файлов, обеспечивающая получение количественных характеристики информационных потоков, переданных в магистральном высокоскоростном Интернет-канале.

4. Изучены в соответствие с авторской методикой особенности информационных потоков в магистральном Интернет-канале, создаваемых выбранными классами пользователей («Слоны», «Мулы», «Мыши»), и их взаимное влияния друг на друга, в том числе, установлено, что:

4.1. в информационных потоках, создаваемых, классом «Мыши» преобладающими являются пакеты размером не более 200 байт, в информационных потоках, создаваемых, классами «Мулы» и «Слоны» преобладающими являются пакеты размером ~1500 байт;

4.2. наиболее стабильными во времени параметры распределения информационных потоков по размеру пакетов в течение недели оказываются у класса пользователей «Слоны» и «Мулы», наименее стабильные – у пользователей класса «Мыши»;

4.3. временные ряды  $N_i$ ,  $V_i$ ,  $\tilde{V}_i$ , содержащие упорядоченные во времени значения числа пакетов и объемов информации, переданных в течение заданного временного интервала, а также среднего объема переданной информации, приходящейся на один пакет. представляют собой случайные последовательности с ограниченной областью рассеяния, накопленные суммы которых обладают свойством самоподобия;

4.4. существует детерминированная линейная связь между объемами информации, переданной в течении 15 минут каждым из классов пользователей, что свидетельствует о возможности создания на основе управления скоростью передачи информации каждого из выделенных классов пользователей механизмов балансировки объемов передаваемой информации, призванных повысить качество обслуживания пользователей данным каналом, и предложена структурная схема подобного алгоритма.

## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи, опубликованные в рецензируемых научных журналах и изданиях, определённых ВАК и Аттестационным советом УрФУ:

1) Поршнеv С.В., **Божалкин Д.А.** Технология семантического анализа дампа трафика информационных потоков в компьютерных сетях // Информационные технологии. 2014. – №11. – С. 12-19. 0,5 п.л. / 0,25 п.л.

2) Porshnev S.V., Kopusov A. S., **Bozhalkin D.A.** Features of information flows in the backbone Internet-channel: the analysis of the statistical characteristics of the relationship between the number of packets and the time //2015 9th International Conference on Application of Information and Communication Technologies (AICT). – IEEE, 2015. – С. 437-440. 0,25 п.л. / 0,08 п.л. (Scopus, WoS).

3) Porshnev S.V., Kopusov A.S., Bozhalkin D.A. The research of a network traffic in a backbone internet channel //2015 International Siberian Conference on Control and Communications (SIBCON). – IEEE, 2015. – С. 1-5. 0,31 п.л. / 0,10 п.л. (Scopus, WoS).

4) Поршнеv С.В., **Божалкин Д.А.** К вопросу о самоподобии трафика, передаваемого в магистральном Интернет-канале // Фундаментальные исследования. – 2016. – № 2(32). – С. 301-310. 0,62 п.л. / 0,31 п.л.

5) Поршнеv С.В., **Божалкин Д.А.**, Копосов А.С. Исследование особенностей потоков сетевого трафика в магистральном интернет-канале // Электросвязь. – 2016. – №2. – С. 16-23. 0,5 п.л. / 0,25 п.л.

6) Поршнеv С.В., **Божалкин Д.А.**, Копосов А.С. Опыт использования суперкомпьютера для обработки дампов сетевого трафика магистрального интернет-канала // Информационные технологии. – 2016. – №1. – С. 42-47. 0,37 п.л. / 0,18 п.л.

7) Porshnev S.V., **Bozhalkin D.A.** The study of self-similarity of the traffic transmitted in the backbone Internet channel // Dynamics of Systems, Mechanisms and Machines (Dynamics), 2016. – IEEE, 2016. – С. 1-7. 0,43 п.л. / 0,21 п.л. (Scopus, WoS).

8) Поршнеv С.В., **Божалкин Д.А.** О возможности использования случайных величин с ограниченной областью рассеяния для генерации фрактального броуновского движения //Информатика и системы управления. – 2017. – №. 1. – С. 23-32. 0,62 п.л. / 0,31 п.л.

9) Porshnev S.V., Kopusov A.S., **Bozhalkin D.A.** Research of Information Flows Features in the Backbone Internet-Channel: Distribution of Packet Size //2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT). – IEEE, 2017. – С. 1-5. 0,31 п.л. / 0,10 п.л. (Scopus, WoS).

### Патенты и программы:

10) Поршнеv С.В., **Божалкин Д.А.** Семантический анализатор дампов трафика информационных потоков в компьютерных сетях // Свидетельство о государственной регистрации программы для ЭВМ № 2015611426 (Заявка № 2014662922. Дата поступления 12 декабря 2014 г. Дата государственной регистрации в Реестре программ для ЭВМ 29 января 2015 г.). –1 с.

11) Поршнеv С.В., **Божалкин Д.А.** Анализатор-классификатор информационных потоков дампов трафика компьютерных сетей // Свидетельство о государственной регистрации программы для ЭВМ № 2015661799 (Заявка № 2015618601. Дата поступления 17 сентября 2015 г. Дата государственной регистрации в Реестре программ для ЭВМ 09 ноября 2015 г.). –1 с.

### Другие публикации:

12) Поршнеv С.В., **Божалкин Д.А.**, Копосов А.С. Особенности информационных потоков в магистральном интернет-канале: анализ статистических характеристик зависимостей объема пакетов от времени //Материалы 25-й Международной Крымской конференция СВЧ-техника и телекоммуникационные технологии. – 2015. – С. 333-336. 0,25 п.л./ 0,08 п.л.

13) Поршнеv С.В., **Божалкин Д.А.**, Копосов А.С. Исследование особенностей потоков сетевого трафика в магистральном интернет-канале: распределения размеров пакетов // Современные методы прикладной математики, теории управления и компьютерных технологий (ПМТУКТ-2016) Сборник трудов IX международной конференции. — 2016. — №1. — С. 283-286. 7819062. 0,25 п.л. / 0,08 п.л.

14) Поршнеv С.В., Божалкин Д.А., Овечкина Е.В. Анализ взаимного влияния информационных потоков в магистральном интернет-канале // Cloud of science. – 2019. – Т. 6. – №. 1. – С. 83-108. 1,62 п.л. / 0,54 п.л.

---

**Подписано к печати \_\_.\_\_.2020.**  
**Объем 1 печ. л. Тираж 80 экз. Зак. №01**

---