

На правах рукописи



Тамазян Араик Симакович

**СТАТИСТИЧЕСКИЕ МОДЕЛИ ТРАФИКА
КОМПЬЮТЕРНЫХ СЕТЕЙ НА ОСНОВЕ
ДОЛГОВРЕМЕННО-ЗАВИСИМЫХ
ДИНАМИЧЕСКИХ ПОТОКОВ**

Специальность 05.13.18 —
«Математическое моделирование, численные методы и
комплексы программ»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Санкт-Петербург — 2016

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» имени В.И. Ульянова (Ленина)» на кафедре радиотехнических систем

Научный руководитель: кандидат технических наук, ведущий научный сотрудник

Богачев Михаил Игоревич

Официальные оппоненты: **Васильев Константин Константинович**,
доктор технических наук,
ФГБОУ ВПО «Ульяновский государственный технический университет»,
заведующий кафедрой телекоммуникаций

Симонина Ольга Александровна,

кандидат технических наук,
ФГБОУ ВО «Санкт-Петербургский государственный университет телекоммуникаций имени проф. М.А. Бонч-Бруевича»,
доцент кафедры радиосвязи и вещания

Ведущая организация: ФГБОУ ВПО «Новгородский государственный университет имени Ярослава Мудрого», г. Великий Новгород

Защита состоится 21 декабря 2016 г. в 14:00 часов на заседании диссертационного совета Д 212.238.01 на базе Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» имени В.И. Ульянова (Ленина) по адресу: 197376, Санкт-Петербург, ул. Проф. Попова, д. 5.

С диссертацией можно ознакомиться в библиотеке Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» имени В.И. Ульянова (Ленина) и на сайте www.eltech.ru.

Автореферат разослан 20 октября 2016 года.

Ученый секретарь

диссертационного совета

Д 212.238.01, кандидат технических наук



Щеголева Н.Л.

Общая характеристика работы

Актуальность темы. Интенсификация обмена информацией в сетях общего пользования обуславливает актуальность оптимизации распределения сетевых ресурсов и динамического управления ими в интересах минимизации вероятности возникновения перегрузок. С начала 90-х годов по настоящее время результаты ряда исследований указывают на несоответствие описания динамики сетевого трафика традиционными моделями, предполагающими пуассоновский характер потоков пользовательских запросов, результатам эмпирических наблюдений. Ускорение обмена информацией в многопользовательских сетях приводит к росту относительной ошибки описания и предсказания динамики трафика такими моделями, что обуславливает актуальность поиска альтернативных моделей динамики трафика в широком диапазоне временных масштабов и размеров сетей. Существующие модели и алгоритмы имитации сетевого трафика в основном отражают технические аспекты передачи информации в сетях на различных уровнях, при этом за скобками часто остается существенное влияние на эмпирическую динамику трафика совместной активности пользователей, причем последний фактор оказывает возрастающее влияние в связи с интенсификацией обмена информацией в сетевых сообществах. Технические решения для сетей различного масштаба и протоколов различного уровня существенно различаются, в то время как динамика пользовательской активности зависит от размеров и связности сообщества пользователей. Представляется целесообразным исследование возможности описания динамики пользовательских обращений к сетевым ресурсам в рамках универсальных масштабируемых моделей. Существующие модели динамики сетевого трафика хорошо работают на коротких временных интервалах, но не учитывают факторы интенсификации обмена информацией в сетях на значительных временных интервалах. Существует потребность в оценке данных параметров и потребности в масштабировании сетевых ресурсов на длительных временных интервалах.

Целью данной работы является исследование статистических свойств трафика компьютерных сетей и построение его масштабируемой математической модели, учитывающей динамику совместной активности пользователей и отражающей текущие потребности в сетевых ресурсах.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Осуществить сбор и статистический анализ эмпирических данных трафика в сетях различного масштаба (отдельные узлы, локальные и опорные сети) и организации на различных уровнях съема данных (пакетном, файловом, сессионном).

2. Разработать феноменологическую и имитационную модели сетевого трафика с учетом эффектов долговременной зависимости и временной кластеризации выбросов трафика.

3. Выполнить статистический анализ и моделирование долговременных изменений динамики сетевого трафика и валидация предложенной модели с помощью методов теории массового обслуживания.

Основные положения, выносимые на защиту:

1. Потоки пользовательских запросов и агрегированного трафика в узлах и каналах многопользовательских компьютерных сетей различного масштаба на различных уровнях съема данных могут быть представлены в рамках единой *феноменологической модели* на основе суперпозиции стационарных пуассоновских потоков с долговременно-зависимым потоком локальных интенсивностей.

2. Оценка вероятности присутствия и характеристик линейного тренда при его наблюдении на фоне случайного процесса с долговременной зависимостью в окне ограниченной длины может быть выполнена с использованием предложенного *численного метода* с произвольной заранее заданной доверительной вероятностью при известной длине окна наблюдения и известном показателе Херста.

3. Оценка статистических характеристик и моделирование потоков данных в многопользовательских компьютерных сетях различного масштаба на различных уровнях съема данных, а также анализ характеристик эмпирического и модельного трафика с позиций теории массового обслуживания может быть выполнена с использованием разработанного *комплекса программ*.

Научная новизна:

1. Предложена оригинальная феноменологическая модель сетевого трафика, позволяющая описать потоки пользовательских запросов и агрегированный трафик в узлах и каналах компьютерных сетей различного масштаба на различных уровнях съема на основе суперпозиции стационарных пуассоновских потоков с долговременно-зависимым потоком локальных интенсивностей.

2. Получена оценка вероятности наличия и параметров линейного тренда в долговременно-зависимых рядах данных на примере синтезированных выборок с различными законами распределениями и эмпирических потоков пользовательских запросов и динамики агрегированного трафика высоконагруженного узла компьютерной сети.

3. Разработано оригинальное программное обеспечение для статистического и имитационного моделирования потоков пользовательских запросов и агрегированного трафика в узлах и каналах компьютерных сетей с учетом свойств долговременной зависимости и непуассоновского характера потоков пользовательских запросов.

Практическая значимость заключается в том, что предложенные модели трафика ИКС, численные методы оценки его характеристик и разработанный на их основе комплекс программ могут быть использованы для оценки потребности в ресурсах и оптимизации управления ресурсами ИКС с учетом непуассоновского характера пользовательских запросов и эффектов долговременной зависимости.

Достоверность полученных результатов обеспечивается результатами анализа большого объема эмпирических потоков данных в ИКС различного масштаба и организации на различных уровнях съема, результатами статистического моделирования, результатами аналитических расчетов, и не противоречат литературным данным.

Апробация работы. Основные результаты работы докладывались на научно-технической школе-семинаре “Инфокоммуникационные технологии в цифровом мире” (СПб, 2012 г.); 68-й и 69-й научно-технической конференциях СПбНТОРЭС им. А.С. Попова (СПб, 2013, 2014 гг.); международных конференциях IEEE Young Researchers in Electrical and Electronic Engineering Conference (СПб, 2015, 2016 гг.).

Диссертационные исследования автора были поддержаны стипендией Президента РФ для обучения студентов и аспирантов за рубежом 2013-2014 гг., специальными стипендиями Ученого Совета СПбГЭТУ “ЛЭТИ”, присужденными по результатам конкурсов научных достижений аспирантов 2015-2016 гг., поддержкой персонального научно-исследовательского проекта по результатам конкурса научно-исследовательских и инновационных проектов для аспирантов и молодых научно-педагогических работников СПбГЭТУ “ЛЭТИ”

2016 г., а также грантом РФФИ (шифр проекта 16-37-00374), где диссертант выступает руководителем исследований. Результаты исследований были использованы при выполнении НИР в рамках Госзадания Минобрнауки РФ на кафедре РС СПбГЭТУ “ЛЭТИ” (шифр проекта 8.324.2014/К). Результаты диссертационной работы, включая разработанное автором программное обеспечение, апробированы и используются в учебном процессе СПбГЭТУ “ЛЭТИ” при проведении лабораторных и практических занятий по дисциплине “Математическое моделирование радиотехнических устройств и систем”.

Личный вклад. Основные исследования выполнены автором лично. Соавторы публикаций принимали участие в выполнении исследований на этапах постановки задач и обсуждения результатов исследования.

Публикации. Основные результаты по теме диссертации изложены в 12 работах, включая 3 статьи в журналах, включенных в список ВАК [1–3], 2 статьи в международных рецензируемых журналах, включенных в базу цитирования Web of Science [4;5] 3 статьи в материалах международных конференций IEEE, индексируемых Scopus [6–8], 2 публикации в тезисах докладов на российских конференциях [9;10], 2 заявки о регистрации программ для ЭВМ [11;12].

Объем и структура работы. Диссертация состоит из введения, четырёх глав, заключения и двух приложений. Полный объём диссертации составляет 118 страниц с 55 рисунками и 2 таблицами. Список литературы содержит 72 наименования.

Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, отмечаются научная новизна и практическая значимость представляемой работы.

Первая глава посвящена обзору существующих подходов к моделированию трафика компьютерных сетей [Таненбаум, 2012; Соколов, 2003; Наумов, 2007]. Рассматриваются феноменологические модели с учетом эффектов долговременной зависимости [Шелухин, 2003, 2005, 2011], используемые для моделирования потоков данных в каналах и узлах инфокоммуникационных сетей (ИКС). Приводятся имитационные модели, основанные на аппарате теории

массового обслуживания, и используемые для оценки показателей эффективности обслуживания пользователей с учетом пропускной способности в узлах и каналах инфокоммуникационных сетей [Саати, 1961; Лифшиц, 1978; Степанов, 2010; Вишневецкий, 2003].

Характерной особенностью трафика инфокоммуникационных сетей является его самоподобие [Leland, 1994; Paxson, 1995; Feldman, 1998; Riedi, 1999; Park & Willinger, 2000; Шелухин, 2003, 2005, 2011; Иванов, Колесников, 2015 и др.]. Понятие самоподобия было введено в работах Б. Мандельброта, и означает, что некоторое свойство временного ряда $x(i)$ сохраняется при масштабировании аргумента $i \rightarrow bi$, причем уравнение самоподобия для x будет иметь вид $x(bi) = ax(i) = b^H x(i)$. Показатель $H = \ln a / \ln b$ называется показателем Херста и является основной статистической характеристикой самоподобия. Если показатель Херста процесса $0.5 < H < 1$, то он является положительно коррелированным стационарным процессом с АКФ вида $C(s) \sim s^{-\gamma}$, где γ — корреляционная экспонента, взаимно однозначно связанная с показателем Херста H выражением $\gamma = 2 - 2H$. Для процессов с такой автокорреляционной функцией (АКФ) характерен неограниченный рост оценки интервала корреляции $C_\times = \int_0^N C(s) ds$ с ростом объема выборки N , поэтому их часто называют процессами с долговременной зависимостью (ДВЗ). Спектральная плотность такого процесса имеет вид $S(f) \sim f^{-\beta}$, где $\beta = 1 - \gamma$.

В условиях затрудненности прямого оценивания АКФ, в частности, при наблюдении флуктуационного поведения исследуемого процесса на фоне регулярных циклов, трендов и шумов, показатель Херста обычно оценивается с помощью методов флуктуационного анализа, среди которых наибольшее распространение в последние годы получил метод флуктуационного анализа с исключением тренда (DFA). Этот метод основан на зависимости оценки $F(s)$ среднего квадрата отклонения суммы с накоплением процесса от ее полиномиальной аппроксимации во временном окне от его длины s , называемой флуктуационной функцией (ФФ). Для рассмотренного класса процессов $F(s) \sim s^H$. При моделировании случайных процессов в ИКС в последние годы широкое распространение получили математические модели рядов с ДВЗ, основной характеристикой которых является H , которые могут быть сформированы методом линейной фильтрации. В случаях, когда распределение отсчетов формируемого ряда статистически значительно отличается от нормального, для синтеза рядов с

ДВЗ используется итерационная процедура, известная как алгоритм Шрайбера-Шмитца [Schreiber & Schmitz, 1996].

Вторая глава посвящена статистическому анализу эмпирических данных и построению феноменологической модели пользовательских запросов и агрегированного трафика ИКС различного масштаба. В многопользовательских сетях в силу особенностей обмена информацией между отдельными пользователями характерны нерегулярные колебания пользовательской активности. Предлагается модель, учитывающая влияние динамики совместного поведения пользователей на общий поток пользовательских запросов и агрегированный трафик ИКС.

В качестве эмпирических данных используются данные трафика отдельных узлов ИКС со стационарной и нестационарной динамикой трафика, включая высоконагруженный узел WorldCup 98; данные трафика магистрального канала японской опорной академической сети WIDE и данные трафика канала, соединяющего локальную сеть комплекса общежитий СПбГЭТУ «ЛЭТИ» с интернет-провайдером, приведенные к уровню пользовательских сеансов. Данная процедура весьма распространена при анализе трафика [Nuzman, 2002], поскольку она позволяет во многом абстрагироваться от конкретных технических решений и сосредоточиться на характеристиках активности пользователей в сети. Соответственно, на сеансовом уровне становится возможным непосредственное сравнение статистических свойств трафика, который исходно был собран в каналах и узлах сетей различного масштаба и организации.

Далее был выполнен анализ статистических свойств трафика для кратковременных фрагментов, выбранных таким образом, чтобы в большинстве случаев обеспечивалась стационарность трафика в рамках одного фрагмента. Исходя из этих соображений, по результатам анализа эмпирических данных была выбрана длительность фрагмента трафика.

Результаты анализа эмпирических данных указывают на то, что плотность вероятности (ПВ) межсеансового времени в отдельных фрагментах может быть приближенно описана убывающей экспоненциальной функцией $p(\tau/\bar{\tau})$, что соответствует пуассоновскому потоку со случайным поступлением запросов. Таким образом, динамика агрегированного трафика в каждом фрагменте может быть охарактеризована единственным параметром — интенсивностью поступления запросов $\beta \cong 1/\bar{\tau}$. Также было установлено, что $p(\beta)$,

будучи перемасштабированной к единицам локального β в течение данного дня, показывает близкую к универсальной форму распределения для каждой из рассматриваемых сетей. Установлено, что $p(\beta)$ асимптотически убывает по экспоненте квадрата аргумента, при этом ни одно из широко используемых распределений не обеспечивает достаточно точное описание распределения β с учетом его центральной части. Однако подобное асимптотическое поведение характерно для нормального распределения. Поскольку временные интервалы по определению являются неотрицательными, на практике может быть использовано усеченное нормальное распределение, либо сходящиеся к нему Γ - и χ^2 -распределения (последнее — при большом количестве степеней свободы k).

Соответственно, динамика агрегированного трафика может быть описана в виде суперпозиции статистик для множества отдельных коротких фрагментов, каждый из которых описывается экспоненциально распределенным межсеансовым временем с локальной интенсивностью β . Так как β — случайная величина, описываемая своим собственным распределением $p(\beta)$, общее распределение межсеансового времени для агрегированного трафика может быть выражено как

$$P(\tau) = \int_0^{\infty} P(\beta)P(\tau|\beta)d\beta = \int_0^{\infty} P(\beta)\beta e^{-\beta\tau} d\beta. \quad (1)$$

Подобная суперпозиция случайных величин с экспоненциальным распределением приводит к распределению, которое носит название q -экспоненциального распределения [Tsallis, 1988; Beck, 2003], ПВ которого может быть выражена как

$$P(\tau) \propto [1 + \lambda(q - 1)\tau]^{1/(1-q)}, \quad (2)$$

где $q = 1 + 2/(k + 2)$ и β распределена по закону χ^2 с числом степеней свободы k [Beck, 2005; Briggs, 2006]. Для больших k ПВ распределения χ^2 убывает по квадрату экспоненты. Для рассмотренных эмпирических данных коэффициент вариации $\rho = \sqrt{2k}/k$ ведет к $k \cong 16$ и $k \cong 200$, из чего предполагается $q \cong 1.11$ и $q \cong 1.01$ для каналов локальной и опорной сетей, соответственно. Кроме того, q -экспоненциальное распределение наблюдается также в эмпирических данных межпакетных интервалов в трафике опорной инфокоммуникационной сети [8]. Для высоконагруженного кластера с нестационарной динамикой получены

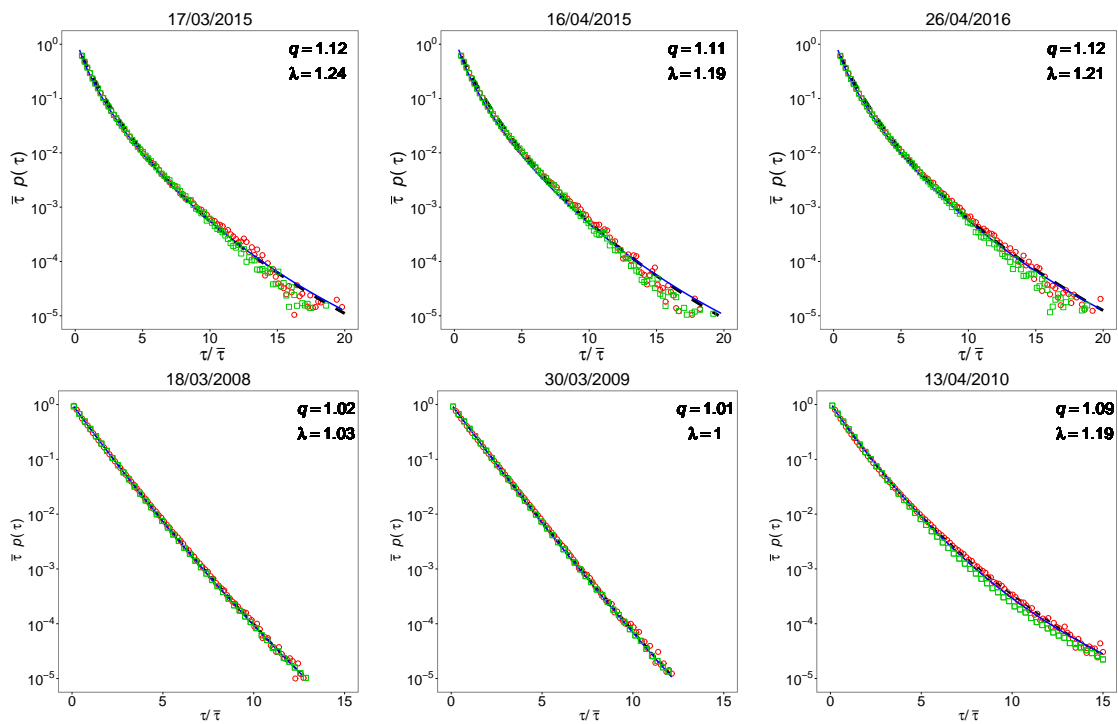


Рис. 1 — Оценки ПВ нормированных межсеансового интервалов для эмпирических данных трафика локальной (верхний ряд) и опорной (нижний ряд) сетей, их аппроксимации q -экспоненциальным распределением, результаты численного интегрирования согласно (1).

аппроксимации интервалов между запросами q -экспоненциальным распределением с эффективным показателем q .

Для проверки сформулированных выше гипотез получены оценки ПВ для эмпирического (нормированного к среднему) межсеансового времени $p(\tau)$ для агрегированного трафика за суточные интервалы наблюдений часа. На рис. 1 показано, что q -экспоненциальное распределение с высокой степенью точности описывает оценки ПВ со значениями q , что согласуется с прогнозом на основе предложенной модели по критерию согласия Колмогорова-Смирнова.

Как и для межсеансового времени, q -экспоненциальное распределение характерно для распределения объема информации, переданной за сеанс, с $q \cong 1.5$ для локальной сети общежитий и $q \cong 1.65$ для магистральной сети WIDE (см. рис. 2). Стоит отметить, что данное наблюдение близко соотносится с предыдущими наблюдениями закона Зипфа в сетевом трафике [Breslau, 1999]. Аналогичные закономерности были выявлены для данных трафика односторонних WWW-серверов, что указывает на широкую применимость предложенного подхода для описания потоков запросов и агрегированного трафика узлов и каналов инфокоммуникационных сетей различного масштаба [4].

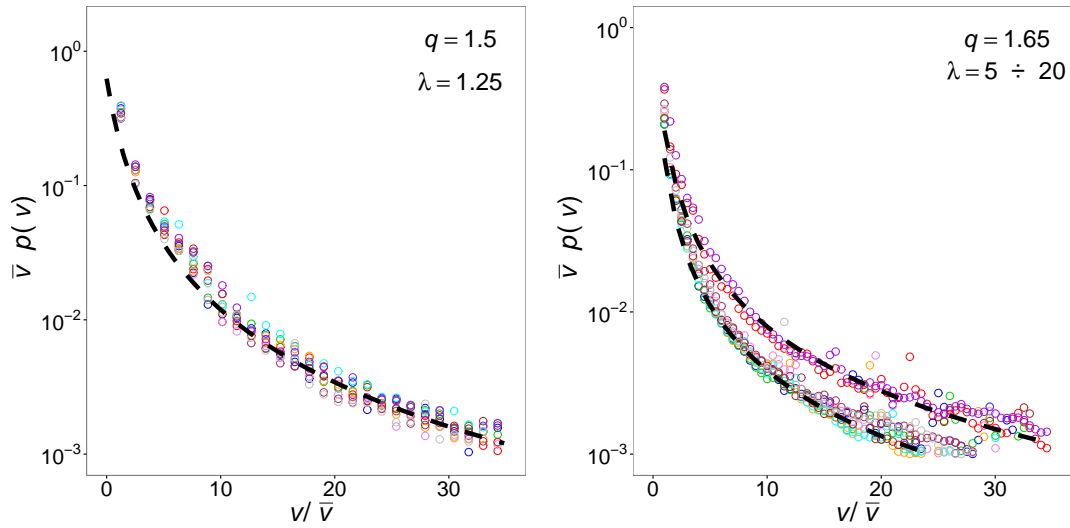


Рис. 2 — Оценки ПВ нормированных объемов информации, переданной за сеанс, для локальной и опорной сетей и их аппроксимации q -экспоненциальными распределениями.

Флуктуации локальной интенсивности потока пользовательских запросов β соответствуют интенсивности обмена информацией между пользователями и могут быть воспроизведены с использованием диффузионной модели распространения информации в сообществе. Полученные результаты отражают влияние флуктуаций характеристик модели распространения информации, включая динамический размер сообщества — число активно обменивающихся информацией пользователей, на характеристики потока пользовательских запросов и агрегированного трафика в заданной ИКС.

Третья глава посвящена решению задачи обнаружения и оценки параметров тренда в трафике ИКС. Ввиду того, что долговременные тренды и ДВЗ могут как имитировать, так и маскировать друг друга, актуальной является задача оценки вероятности присутствия относительного тренда в данных длительного наблюдения трафика ИКС, наблюдаемого на фоне наличия ДВЗ.

Отметим, что проблема обнаружения внешнего тренда не связана с его происхождением. Рассмотрим значения временного ряда $y_i, i = 1, \dots, L$. Для аппроксимации тренда в рассматриваемом временном окне длины L обычно выполняется регрессионный анализ, в простейшем случае — линейный. Из уравнения регрессии $r_i = bi + d$ можно получить относительное смещение регрессионной линии на интервале L $\delta = b(L - 1)$ и показатель флуктуаций в районе тренда, описываемый стандартным отклонением $\sigma = [(1/L) \sum_{i=1}^L (y_i - r_i)^2]^{1/2}$.

Оценкой относительного тренда будет величина

$$x = \delta/\sigma \quad (3)$$

При оценке тренда в конечной выборке данных возникает задача различения истинного тренда и тренда, имитированного ДВЗ. Для решения этой задачи необходимо получить оценку вероятности $\hat{P}(x; L)$. В дальнейшем без потери общности рассматривается случай $x > 0$. Из \hat{P} мы получаем оценку дополнительной функции распределения $\hat{W}(x; L) = \int_{x'}^{\infty} P(x; L)dx'$ и статистическую значимость тренда

$$S(x; L) = \int_{-x}^x P(x'; L)dx' = 1 - 2W(x; L). \quad (4)$$

Соотношение $S(x_{95}; L) = 0.95$ задает верхние и нижние пределы $\pm x_{95}$ 95% доверительного интервала. Если значимость относительного тренда x выше 0.95, считается, что рассматриваемый тренд с вероятностью 0.95 не может быть объяснен эффектами, связанными с наблюдением ДВЗ в рассматриваемом временном ряде. В противном случае, относительные тренды x между $-x_{95}(L)$ и $x_{95}(L)$ считаются с вероятностью 0.95 имитированными ДВЗ, присутствующей во временном ряде.

Если x выше x_{95} , то часть $x - x_{95}$ не может быть объяснена эффектами ДВЗ во временном ряде и, таким образом, может быть расценена как присутствие внешнего тренда, минимальное относительное смещение которого определяется как

$$x_{\text{ext}}^{\min} = x - x_{95}. \quad (5)$$

С другой стороны, внешний тренд с доверительной вероятностью 95% не превышает значение

$$x_{\text{ext}}^{\max} = x + x_{95}, \quad (6)$$

которое представляет максимальный внешний относительный тренд.

По определению, x_{ext}^{\min} представляет собой нижнюю границу наблюдаемого внешнего тренда, который не может быть объяснен только эффектами ДВЗ, в то время как x_{ext}^{\max} — наибольший возможный внешний относительный тренд, соответствующий эффектам ДВЗ во временном ряде. Согласно (5) и (6),

$\pm x_{95}(L)$ могут рассматриваться как границы 95% доверительного интервала относительного тренда временном ряде длины L .

Для белого гауссовского шума предполагается [Santer, 2000], что отношение t_b между оценкой коэффициента наклона b и его стандартной ошибкой s_b [Bronstein, 2004] может быть описано распределением Стьюдента. Так как $b = \Delta/(L - 1)$ и $s_b = [\sum_{l=1}^L (y_l - r_l)^2 / ((L - 2) \sum_{l=1}^L (l - \bar{l})^2)]^{1/2}$ с $\bar{l} = 1/l \sum_{l=1}^L l$, ПВ $P(x; L)$ имеет вид

$$P(x; L) = \frac{\Gamma([l(L) + 1]/2)}{\Gamma(l(L)/2) \sqrt{\pi l(L)} a} (1 + (x/a)^2/l(L))^{-[l(L)+1]/2} \quad (7)$$

с числом степеней свободы

$$l(L) = L - 2 \quad (8)$$

и коэффициентом масштаба

$$a = \frac{\sqrt{12}(L - 1)}{\sqrt{L^2 + 2}} \frac{1}{\sqrt{L - 2}}, \quad (9)$$

где Γ — гамма-функция. При больших значениях L a стремится к $a \cong \sqrt{12}/\sqrt{l(L)}$.

Из (7) и (4) можно выразить значимость S как функцию от x/a и $l(L)$

$$S(x; L) = 2 \frac{x}{a} \frac{\Gamma(\frac{1}{2}(l(L) + 1))}{\sqrt{\pi l} \Gamma(\frac{l(L)}{2})} {}_2F_1 \left(\frac{1}{2}, \frac{1}{2}(l(L) + 1); \frac{3}{2}; -\frac{(x/a)^2}{l(L)} \right) \quad (10)$$

где ${}_2F_1$ — гипергеометрическая функция.

Для численной оценки $P(x; \alpha; L)$ мы используем аппарат математического моделирования, а именно - метод Монте-Карло, для генерации 800 реализаций длины $L_{tot} = 2^{21}$, для 241 значения глобального показателя Херста α^* , изменяющегося от $\alpha^* = 0.1$ до $\alpha^* = 2.5$. Однако наибольший интерес представляют значения α от 0.5 до 1.5, которые наиболее часто встречаются в выборках эмпирических данных. В многих работах [Rybski, 2008; Lennartz, 2009, 2011] было показано, что оценки локального показателя Херста α в каждом фрагменте выборки не равны глобальному показателю Херста α^* всей выборки, а лежат в некоторой окрестности α^* . На рис. 3 показаны функции $1 - S(x; \alpha, L) = 2W(x; \alpha, L)$ для трех репрезентативных выборок длины

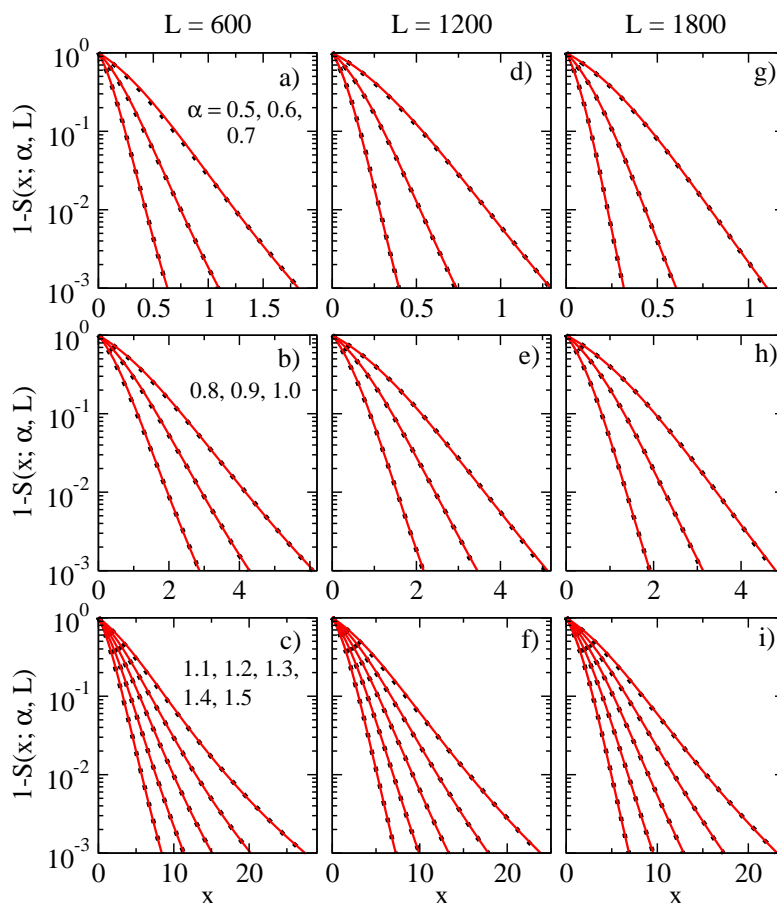


Рис. 3 — Значимость $S(x; \alpha, L)$ для относительных трендов x , возникающих в выборках длины L с ДВЗ с показателем Херста α . Выборки обладают нормальным распределением. (a) $L = 600$ и $\alpha = 0.5, 0.6, 0.7$ (сверху вниз). Сплошной линией показаны соответствующие аппроксимации t -распределением Стьюдента. (b,c) Аналогично (a) но для значений показателя Херста $\alpha = 0.8, 0.9, 1.0$ и $\alpha = 1.1, 1.2, 1.3, 1.4, 1.5$, соответственно. (d-f) и (g-i): То же, что и (a-c), но для выборок длины $L = 1200$ и $L = 1800$, соответственно.

$L = 600, 1200$, и 1800 . ДВЗ в данных характеризуется значениями локального показателя Херста $\alpha = 0.5, 0.6, \dots, 1.5$. Точками показаны результаты математического моделирования по методу Монте-Карло. Сплошными линиями показаны аппроксимации W t -распределением Стьюдента согласно (10) с соответствующими значениями коэффициента масштаба a и эффективной длины l . Видно, что аппроксимация $1 - S$ статистически неразличима относительно результатов математического моделирования, что подтверждается оценками по критерию согласия Колмогорова-Смирнова.

Четвертая глава посвящена имитационному моделированию канала ИКС как системы массового обслуживания (СМО) с использованием как эмпирических данных, так и синтезированных выборок с q -экспоненциальным распределением. Схема работы СМО показана на рис. 4.

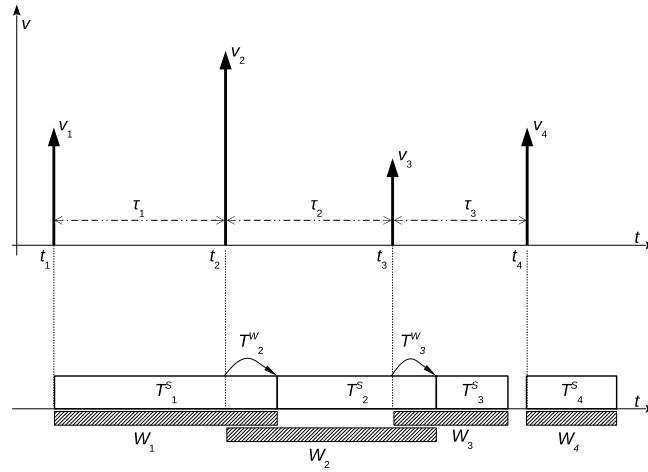


Рис. 4 — Диаграмма модели системы массового обслуживания. Пользовательские сеансы характеризуются временем начала t_i и объемом информации, переданной за сеанс, v_i . T_i^W , T_i^S и W — времена ожидания, обслуживания и пребывания в системе соответственно.

При моделировании предполагается, что пользовательские сеансы обслуживаются непрерывно и последовательно. Соответственно, если канал свободен в момент прибытия пользовательского запроса, начинается сеанс, продолжающийся до тех пор, пока не будет передана вся запрошенная пользователем информация. Таким образом среднее время пребывания сеанса равно времени обслуживания $W_i = T_i^S = v_i/c$, где c — пропускная способность канала.

В противном случае среднее время пребывания сеанса будет равно $W_i = T_i^W + T_i^S$, где T_i^W — время ожидания. Основным параметром, характеризующим производительность СМО, будет являться средняя время пребывания пользовательского сеанса.

На рис. 5 показаны оценки среднего времени пребывания в зависимости от коэффициента использования U . Сплошной линией показан результат моделирования для эмпирических данных, незакрашенными точками — для модели с агрегированием пуассоновских потоков с эмпирическими интенсивностями β , ромбами — для модели с агрегированием пуассоновских потоков с перемешанными эмпирическими интенсивностями с введением корреляции, квадратами — для модели с агрегированием пуассоновских потоков с перемешанными эмпирическими интенсивностями, пунктирной линией — оценка, рассчитанная по формуле Кингмана. Видно, что оценка среднего времени пребывания, полученная с помощью формулы Кингмана, недооценивает \bar{W} на один–два порядка, что абсолютно неприемлемо.

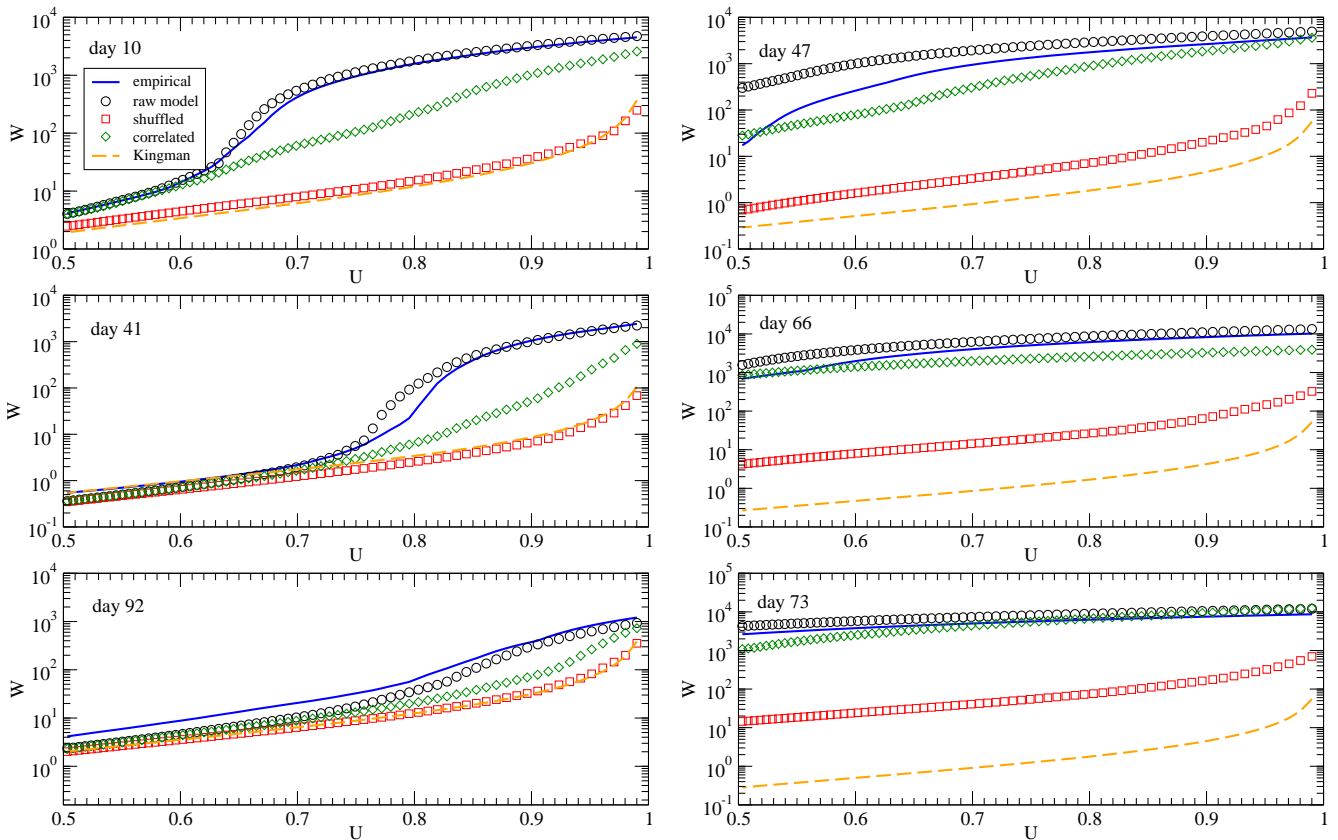


Рис. 5 — Среднее время пребывания запроса в высоконагруженном узле от его коэффициента использования для первой группы дней, где наблюдается стационарная динамика трафика (слева), и для второй группы дней, где наблюдается нестационарная динамика трафика (справа).

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

1. На основе представления агрегированных потоков данных в многопользовательских сетях в виде суперпозиции локальных пуассоновских потоков с различной интенсивностью предложена универсальная модель интервалов между требованиями, описываемая q -экспоненциальным распределением, являющимся частным случаем обобщенного распределения Парето. При этом высокая корреляция параметров распределения q и λ позволяет уменьшить параметризацию модели.

2. Предложен оригинальный численный метод оценки вероятности присутствия и параметров тренда при его наблюдении на фоне процесса с долговременной зависимостью в короткой выборке данных. Предложена модель агрегированных потока требований и исходящего трафика высоконагруженного узла на основе суперпозиции пуассоновских потоков с долговременной

зависимыми значениями локальных интенсивностей, получены эффективные аппроксимации распределений интервалов между событиями в таких потоках.

3. Методом статистического моделирования с использованием аппарата теории массового обслуживания показана эффективность предложенных моделей для описания потоков данных в узлах и каналах ИКС различного масштаба и организации. По эмпирическим данным проанализированы статистические характеристики имитационной и феноменологической моделей потоков данных в высоконагруженном узле ИКС, сформулированы рекомендации по их имитационному моделированию в стационарном и нестационарном режимах.

Публикации в изданиях, входящих в перечень ВАК

1. Тамазян А.С., Богачев М.И. Анализ интервальных статистик между абонентскими запросами к узлу глобальной информационной сети // *Известия высших учебных заведений России. Радиоэлектроника* — 2013. — Т. 2. — С. 35–38.

2. Тамазян А.С., Богачев М.И. Оценка эффективного времени обслуживания пользователей в узле глобальной информационной сети // *Известия высших учебных заведений России. Радиоэлектроника* — 2013. — Т. 3. — С. 51–54.

3. Тамазян А.С., Богачев М.И. Исследования влияния корреляционных свойств входного потока требований на показатели эффективности системы массового обслуживания // *Известия высших учебных заведений России. Радиоэлектроника* — 2014. — Т. 4. — С. 44–47.

Публикации в изданиях, входящих в Web of Science и Scopus

4. Tamazian A., Ludescher J., Bunde A. Significance of trends in long-term correlated records // *Phys. Rev. E*. — 2015. — Vol. 91. — Pp. 032806(1–10).

5. Tamazian A., Nguyen V.D., Bogachev M., Markelov O. Universal model for collective access patterns in the internet traffic dynamics // *Europhys. Lett.*. — 2016. — Vol. 115. — Pp. 10008(1–7).

6. Tamazian A., Bogachev M. Analytical and numerical estimates of the Weibull/M/1 and Weibull/Weibull/1 queues efficiency // *Proceedings of the Young Researchers in Electrical and Electronic Engineering Conference (EIconRusNW)*. — 2015. — Pp. 97–100.

7. Nguyen V.D., Tamazian A., Markelov O., Bogachev M. Temporal Clustering Effects in the Network Traffic Evaluated by Queueing System Performance //

Proceedings of the Young Researchers in Electrical and Electronic Engineering Conference (EIconRusNW). — 2016. — Pp. 197–199.

8. *Tamazian A., Markelov O., Bogachev M.* Traffic Analysis on the WIDE Backbone Link: From Transport Level to End User Activity //Proceedings of the Young Researchers in Electrical and Electronic Engineering Conference (EIconRusNW). — 2016. — Pp. 197–199.

Другие публикации

9. *Тамазян А.С., Богачев М.И.* Моделирование трафика многопользовательских инфокоммуникационных систем //Сборник докладов научно-технической школы–семинара «Инфокоммуникационные технологии в цифровом мире» — 2012. — С. 14–15.

10. *Тамазян А.С., Богачев М.И.* Оценка распределения времени между запросами и времени обслуживания в системе массового обслуживания с непугасоновским входным потоком //Сборник трудов 68-й научно-технической конференции СПбНТОРЭС им. А.С. Попова. — 2013. — С. 35-36.

Патенты и свидетельства о регистрации программ для ЭВМ

11. *Тамазян А.С.* Заявка на регистрацию программы для ЭВМ №2016619231 от 30 августа 2016 г. //Программа статистического анализа потоков пользовательских запросов по данным трафика инфокоммуникационных сетей. — 2016.

12. *Тамазян А.С.* Заявка на регистрацию программы для ЭВМ №2016619233 от 30 августа 2016 г. //Программа оценки требуемой пропускной способности каналов инфокоммуникационных сетей с непугасоновскими потоками запросов с позиций теории массового обслуживания. — 2016.