

УДК 004.727.4

МЕТОД СБОРА ДАННЫХ О ТЕКУЩИХ ХАРАКТЕРИСТИКАХ В ВЫСОКОСКОРОСТНЫХ КАНАЛАХ ПАКЕТНОЙ ПЕРЕДАЧИ ДАННЫХ

**Крюков Юрий Алексеевич¹, Кубарский Максим Александрович²
Чернягин Денис Викторович³**

¹ *Кандидат технических наук, доцент Института системного анализа и управления;*

*ГОУ ВПО «Международный Университет природы, общества и человека «Дубна»,
Институт системного анализа и управления;*

141980, Московская обл., г. Дубна, ул. Университетская, 19;

e-mail: kua@uni-dubna.ru.

² *Ведущий разработчик, ООО «СТИКЛЕР»;*

ООО «СТИКЛЕР»;

141986, Московская область, г. Дубна, ул. Приборостроителей, д.5, корп.2;

e-mail: 3dmax_k@list.ru.

³ *Старший преподаватель;*

*ГОУ ВПО «Международный Университет природы, общества и человека «Дубна»,
Институт системного анализа и управления;*

141980, Московская обл., г. Дубна, ул. Университетская, 19;

e-mail: dancher2000@mail.ru.

В век бурного развития Интернет-технологий и технологий пакетной передачи данных, задачи мониторинга каналов связи и управления информационным потоком становятся наиболее актуальными. Чтобы понять и оценить процессы, происходящие в высокоскоростном магистральном канале, был разработан сетевой анализатор на основе библиотеки winpcap с использованием таймера высокого разрешения. Для оценки характеристик трафика с учетом самоподобных процессов использовался коэффициент Херста. Исследования показали, что трафик в магистральном канале обладает свойством самоподобия с достаточно большим значением коэффициента Херста.

Ключевые слова: сетевой трафик, самоподобие, параметр Херста.

МЕТОД СБОРА ДАННЫХ О ТЕКУЩИХ ХАРАКТЕРИСТИКАХ В ВЫСОКОСКОРОСТНЫХ КАНАЛАХ ПАКЕТНОЙ ПЕРЕДАЧИ ДАННЫХ

**Krukov Yury Alekseevich¹, Kubarsky Maxim Aleksandrovich²,
Chernyagin Denis Victorovich³**

¹ *PhD, docent of Institute of system analysis and management;*

International university of the nature, society and man «Dubna», Institute of system analysis and management;

141980, Dubna, Moscow reg., Universitetskaya str., 19;

e-mail: kua@uni-dubna.ru.

² *Senior developer, STIKLER company;*

«STIKLER»

141986 Moscow Region, Dubna, Priborostroiteley str. 5, hall 2;

e-mail: 3dmax_k@list.ru.

³ Senior teacher;

International university of the nature, society and man «Dubna», Institute of system analysis and management;

141980, Dubna, Moscow reg., Universitetskaya str., 19.

e-mail: dancher2000@mail.ru.

In the century of rapid development of the Internet technology and technology of packet transmission data, network monitoring and control data flow became the goal task. To understand the process that occurs in a high-speed backbone channel, network analyzer with high-resolution timer was created. For estimating traffic characteristics with a glance of self-similar processes Hurst's parameter was applied. According to the investigation the backbone traffic shows significant levels of self-similarity.

Keywords: Network traffic, self-similarity, Hurst's parameter.

Введение

Конец XX — начало XXI века ознаменовались революционными изменениями в области электросвязи. Теория конвергенции, сформулированная к концу 90-х годов XX столетия, практически сразу нашла свою реализацию и как результат этой реализации сегодня происходят радикальные изменения в различных типах сервисов и телекоммуникационных технологиях. Интегрирование гетерогенных сетевых сервисов и услуг, основанных на протоколах Интернета, приводит к стремительным изменениям в телекоммуникационных технологиях, а также к увеличению пропускной способности каналов связи. Благодаря этому сближение голосовых сервисов, видео сервисов и данных стало действительно возможным и многие телекоммуникационные компании сегодня производят замену оборудования традиционной телефонии на IP, также появилась возможность внедрения потокового видео, peer-to-peer видео, видео по требованию, интернет радио, услуги хранения и предоставления данных, web-сервисы и т. д.

Существование огромного количества разнородных сервисов в одном физическом канале в часы наивысшей нагрузки могут приводить к перегрузке коммутирующих и маршрутизирующих устройств на магистральных линиях связи и как следствие к частичной или полной деградации сетевой инфраструктуры и отказу широкого спектра предоставляемых услуг. Для предотвращения ситуаций ведущих к отказу магистрального сетевого оборудования наиболее значимой становятся задача инженерии трафика. Очевидно, что задача управления трафиком необходима не только для предотвращения возможных перегрузок в сети, но и для оптимизации использования сетевых ресурсов для извлечения максимальной прибыли при минимальной утилизации канала связи. Таким образом, емкости магистральных и субмагистральных каналов должно быть достаточно не только для существующих сетевых сервисов, но и для развития и внедрения новых сетевых услуг, обеспечивая при этом необходимое качество доставки.

В целях реализации поставленных задач, на основе магистрального канала передачи данных Единой информационно-образовательной сети г. Дубны в университете «Дубна» разработана система захвата сетевого трафика с последующим хранением больших объемов данных. Созданная система сетевого мониторинга обеспечивает захват и проведение статистического анализа трафика магистрального сегмента сети, выявление статистических характеристик состояний трафика построенного по технологии Ethernet, со скоростью 1 Гбит/сек и работающего с загрузкой 80-95% от максимальной пропускной способности.

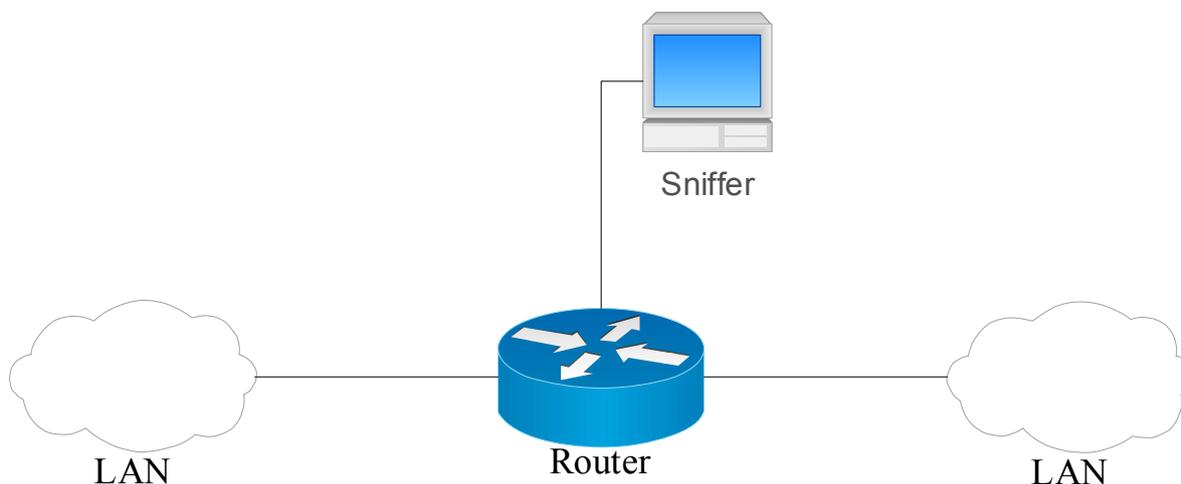


Рис. 1. Логическая схема организации мониторинга сетевого трафика

1. Постановка эксперимента

Эксперимент по исследованию статистических характеристик трафика сети был поставлен следующим образом. В сети города был выбран коммутатор, обеспечивающий трансляцию максимального объема трафика. К свободному порту коммутатора по интерфейсу RJ-45 был подключен персональный компьютер (ПК) с измерительным ПО. На канальном уровне связь осуществлялась по протоколу Gigabit Ethernet. Выбранный порт устройства был сконфигурирован в виде SPAN-порта, зеркально отображающий данные, входящие в коммутатор, т.е. симплексный Rx-канал. В качестве измерительного ПО использовалась разработанная автором программа-снифер на основе открытой библиотеки WinPcap с использованием таймера высокого разрешения. Программное обеспечение позволяет фиксировать событие прихода пакета с точностью до 10 нс. После захвата четырех миллиардов пакетов снифер останавливался и производил обработку данных, приводя полученные реализации в эквидистантную форму с постоянным шагом m при помощи процедуры агрегирования. В ремя начала захвата трафика 22:07:52, а общее время измерения составило 22:17:27 Логическая схема организации сети отображена на Рис. .

Расчет минимального межкадрового интервала в стандарте Gigabit Ethernet составляет 608 нс [1]. Минимальный кадр Gigabit Ethernet состоит из преамбулы (служит для синхронизации передатчика и приемника на физическом уровне), служебной части, полезной нагрузки, поля контрольной последовательности кадра и поля расширения несущей. Размер преамбулы зафиксирован в 64 бит, служебной части в 144 бит, полезной нагрузки в 368 бит, контрольной последовательности в 32 бит, а поле расширения несущей дополняет размер кадра до 512 бит. Минимальное значение «чистого» межкадрового интервала в точности равно времени передачи 96 бит и составляет для Gigabit Ethernet 96 нс, а максимальное значение не ограничено. Учитывая, что преамбула на канальном уровне не обрабатывается, получаем значение «реального» минимального межкадрового интервала в 608 нс. Это свидетельствует о том, что выбранная интервальная шкала в 10 нс, разработанной программы-снифера, способна различить два последовательно приходящих кадра.

Для исследования структуры трафика и иллюстрации его фрактального (самоподобного) характера использовался математический пакет Matlab. В результате процесса агрегирования профилей сетевого трафика были получены временные ряды с временем агрегации: 0,1с; 1с; 10с (Рис. 2). Агрегирование профиля сетевого трафика с временем агрегации 60 сек. привела к существенным искажениям полученного временного ряда. Процесс агрегирования производился в соответствии с методикой предложенной в [2] по формуле:

$$Y_k^{(m)} = \frac{1}{m} \sum_{i=(k-1)m+1}^{k \cdot m} Y_i \quad (1)$$

где Y_i — номер отсчета в полученном профиле, m — размер блока или интервал времени агрегирования, k — номер блока.

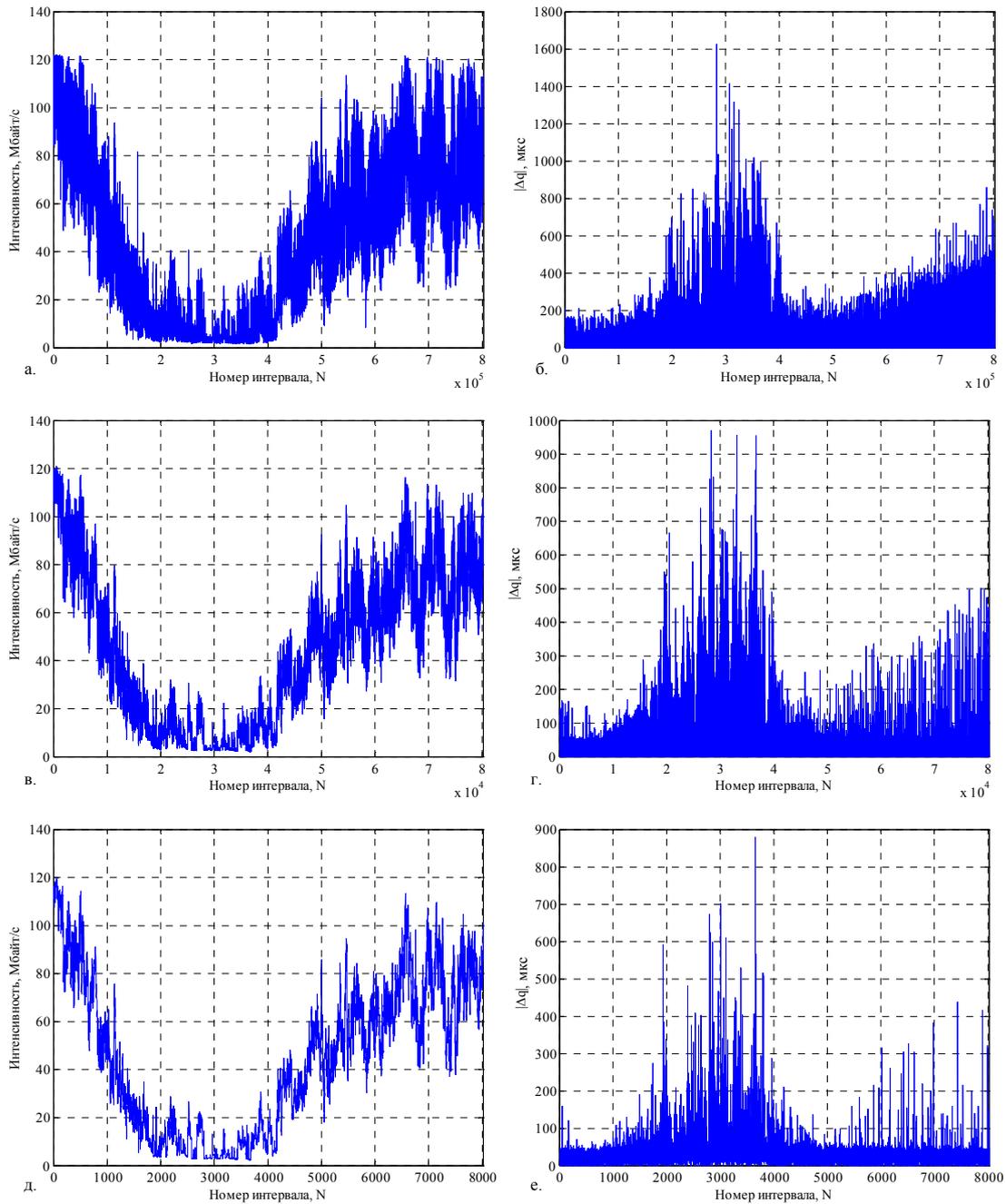


Рис. 2. Временные ряды с временем агрегации 0,1с, 1с, 10с (а, в, д) и ошибки, полученные в процессе агрегации (б, г, е).

На графиках (Рис. 2, б, г, е) видно, что при увеличении времени агрегирования величина ошибки, вносимая процессом агрегирования, уменьшается. Также, можно заметить, области, где интенсивность принимает минимальное значение, а ошибка, в свою очередь, принимает максимальное значение. Это можно объяснить тем, что интервалы времени между поступающими кадрами, на этих промежутках, возрастали. Область от 1000 до 5000 на Рис. 2, д. соответствует временному суточному интервалу от 0 ч. 54 мин. до 9 ч. 14 мин., а максимальная интенсивность была в 0 или в 22 ч.

Визуальный анализ гистограммы распределения длин кадров показывает (Рис. 4. Количество данных переданное кадрами соответствующей длины), что в сети преобладают кадры размером 1518 байт, т.е. максимальный размер кадра, поддерживаемый технологией Ethernet.

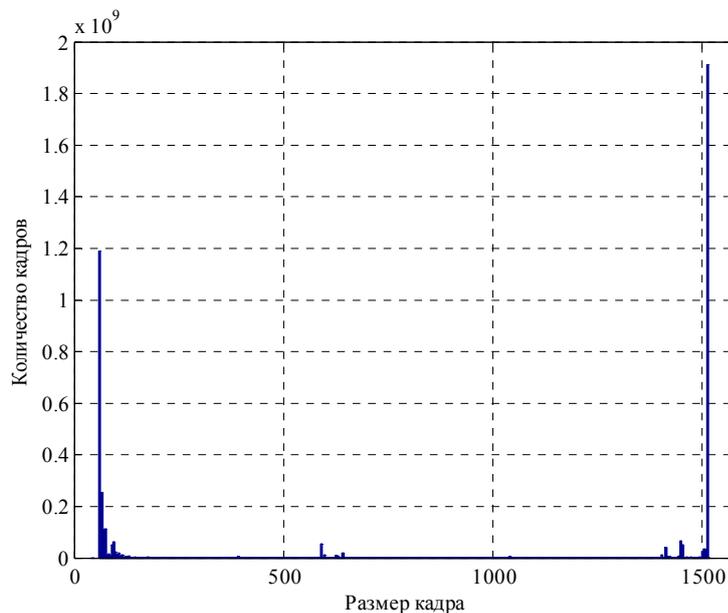


Рис. 3. Распределение длин кадров Ethernet

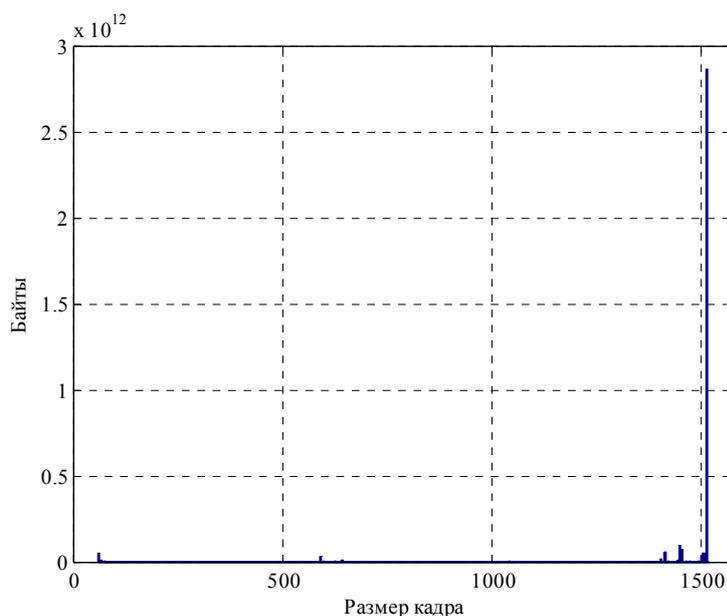


Рис. 4. Количество данных переданное кадрами соответствующей длины

В Таблица 1 приводится статистика полученной реализаций. Из таблицы видно, что за время создания профиля сетевого трафика было передано 3,54 Тбайт данных, 80,9% из которых были переданы кадром максимального размера и 0,015% данных было передано кадром минимального размера. Очевидно, что информация о распределении кадров может привести к заблуждению о количестве переданных данных кадрам соответствующей длины, поскольку в сети могут присутствовать множество небольших управляющих пакетов не переносящих полезных данных. Рис. 4 иллюстрирует распределение данных переданные кадрами различной длины.

Таблица 1. Статистика распределений длин кадров

Название реализации	Количество данных переданных кадром размером 64 байт, Гбайт	Количество данных переданных кадром размером 1518 байт, Гбайт	Всего переданных данных, Гбайт	Процент данных переданных кадром размером 64 байта от всех переданных данных, %	Процент данных переданных кадром размером 1518 байта от всех переданных данных, %
52.07.22.17.01.2009	5,4714311944	2867,844274500	3542,306493804	0,1545	80,9598

Изучив распределение интервалов времени между кадрами в полученной реализации сетевого трафика (Рис. 5), было найдено, что распределение межкадровых интервалов времени хорошо аппроксимируется экспоненциальной функцией плотности распределения вероятности, что соответствует выводам в [3], [4].

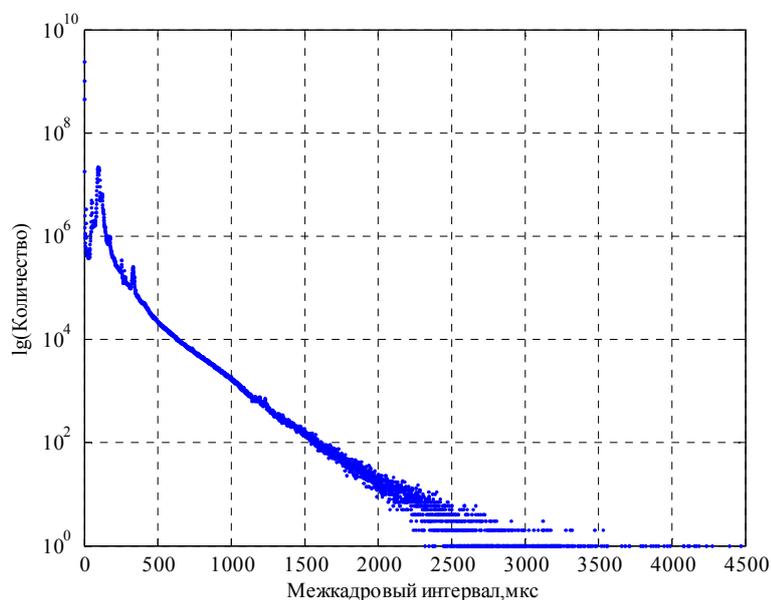


Рис. 5. График распределения времени между кадрами

Визуальный анализ гистограмм агрегированных реализаций с различным временем агрегации (Рис. 7) позволяет увидеть выступ вблизи 0, это объясняется ночным периодом, загрузка канала в это время была минимальна. В [5] говорится, что трафик в канале не затронутый перегрузками, проявляет повсеместную нестационарность.

Если же рассмотреть более короткие интервалы временного ряда, то можно заметить, что на большинстве таких интервалов ряд подчиняется законам распределения с «тяжелым хвостом» (лог-нормальное и Вейбулла [6]), которые имеют субэкспоненциально спадающие хвосты. При оценке параметров распределения на этих промежутках использовался метод максимального правдоподобия с 95% доверительным интервалом, а гипотеза о принадлежности ряда к логнормальному распределению или распределению Вейбулла выполнялась с помощью критерия Колмогорова-Смирнова.

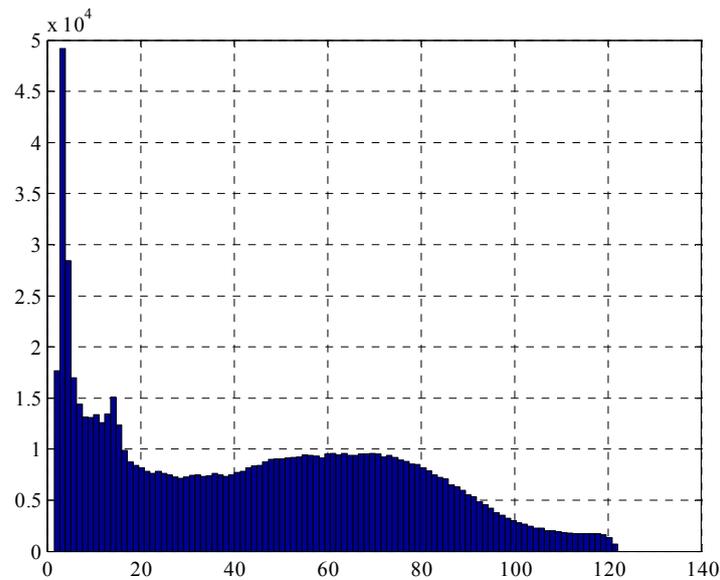


Рис. 7. Гистограмма изучаемой реализации, время агрегации 0,1 сек.

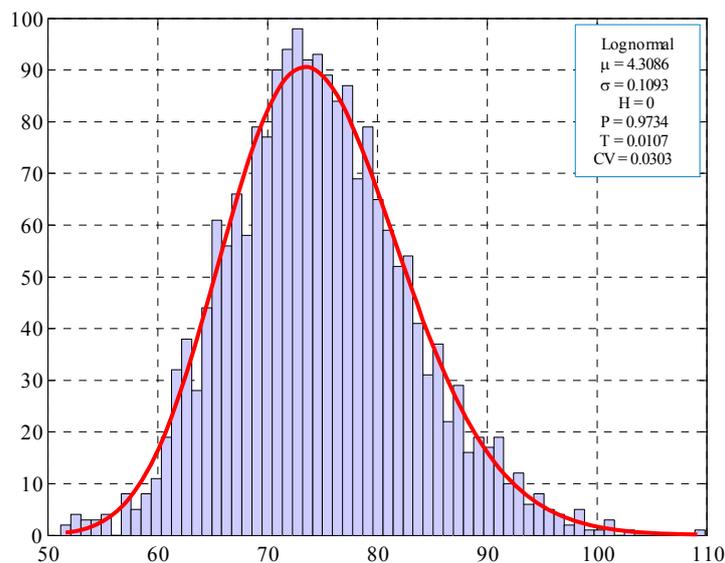


Рис. 6 - Гистограмма небольшого интервала исходного временного ряда

Дальнейший анализ производился с временным рядом полученным в процессе агрегации с интервалом агрегации равным 10 секунд. Это обусловлено тем, что при данном времени агрегации вносится минимальная ошибка, создаваемая процессом агрегации временного ряда.

Визуально выявить присутствие самоподобного процесса а также долговременной зависимости (ДВЗ) можно исследуя частотную область. Если рассматривать самоподобные процессы в частотной области, то явление ДВЗ приводит к степенному характеру спектральной плотности вблизи нуля.

$$S(\omega) = \omega^{-\gamma} L(\omega) \text{ при } \omega \rightarrow 0, \tag{2}$$

где $0 < \gamma < 1$, L — медленно меняющаяся функция в нуле, $S(\omega)$ — функция спектральной плотности.

Таким образом, спектральная плотность стремится к $+\infty$, когда ω приближается к нулю, подобное явление называется $1/f$ - шумом.

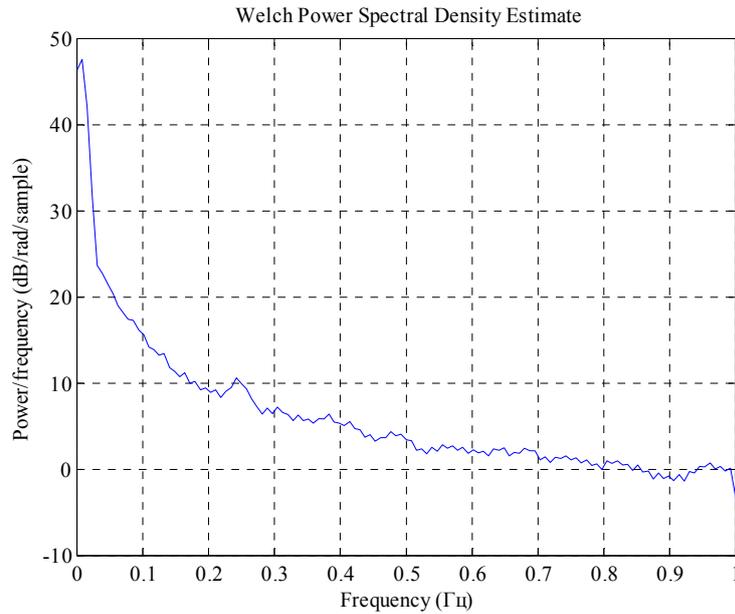


Рис. 8. График функции спектральной плотности

Из рис. 9 видно, что функция спектральной плотности достигает максимального значения в нуле, а ее форма напоминает гиперболическую функцию. Вейвлет анализ позволяет, также, получить визуальное подтверждение того, что рассматриваемый временной ряд проявляет самоподобные свойства (Рис. 9). Таким образом, визуальный анализ частотной области временного ряда показал присутствие ДВЗ процессов.

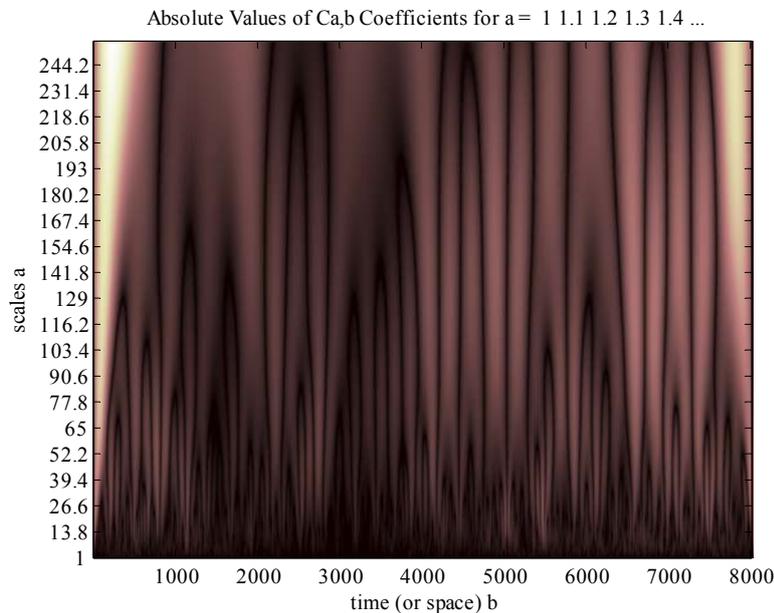


Рис. 9. График вейвлет-спектров

Для более точной оценки самоподобных свойств трафика вводят понятие коэффициента Херста H [7], [8]. Если данный коэффициент находится в пределах $0,5 < H < 1$, то исследуемый процесс проявляет самоподобные свойства. Приближение H к 1 говорит о высокой самоподобности данного процесса и о том, что поведение процесса является персистентным или процесс обладает длительной памятью. То есть, если на некотором временном промежутке в прошлом наблюдалось положительное

приращение процесса, другими словами, увеличение, то и в будущем в среднем будет происходить увеличение.

При $H = 0,5$ отклонение процесса от среднего является случайными и не зависит от предыдущих значений.

При $0 < H < 0,5$ процесс является переменчивым, т.е. увеличение относительно среднего в прошлом, в будущем сменится в противоположном направлении.

На практике проверка присутствия ДВЗ и оценка параметра Херста является не простой задачей. Главная проблема в том, что весьма трудно найти различие между стационарным и нестационарным ДВЗ процессом в следствии факта присутствия локальных трендов, циклов и т.д., что свойственно нестационарным процессам. Имея большую выборку, дать более точную оценку параметру Херста становится более легкой задачей, но можно быть уверенным, что в больших выборках нестационарные эффекты присутствуют благодаря дневным циклам трафика. Гипотеза о стационарности с высокой достоверностью может быть принята только в периоды высокой загрузки канала. Важной частью статистического анализа временных рядов является идентификация и удаление тренда.

Очевидным подходом для решения проблемы стационарности является выбор интервалов времени, где предположение о стационарности трафика справедливо (локальная стационарность).

Для пульсирующих данных, как измеренный трафик (Рис. 4. Количество данных переданное кадрами соответствующей длины) необходимо использовать инструментальное средство, основанное на методе определения точки измерения, которое заключается в том, чтобы перемещать окно по данным и сравнивать распределение выборок в двух половинах окна. Если два распределения существенно различны, то предположение о стационарности для окна отвергается. Сравнение распределений двух рядов равного размера выполняются по критерию Колмогорова-Смирнова.

На основании проведенных тестов на стационарность для анализа можно выбрать несколько подмножеств из всех измеренных данных, каждый из которых получает оценку Херста H . Более подробно данный процесс можно описать следующим образом.

Оценим показатель Херста для блоков данных D . Рассмотрим K сегментов ряда, каждый длиной N . Показатель Херста H оценивается в каждом сегменте $S_i, i = 1, 2, \dots, K = D/N$ с использованием, в нашем случае, R/S -анализа, анализ изменения дисперсий, периодограммный анализ, анализ абсолютных значений. В [9] подробно излагается преимущества и недостатки методов используемых в этой статье. Если оценки в i -м блоке, обозначены как \hat{H}_i , то для соответствующего N оценку показателя Херста можно найти в виде

$$\hat{H}_N = \frac{N}{D} \sum_{i=1}^{D/N} \hat{H}_i. \quad (3)$$

Таким образом, если выбрать N достаточно большим, то можно обеспечить приемлемую сходимость оценки так, чтобы для стационарного процесса оценка \hat{H}_N не зависела от N .

В итоге были получены следующие оценки показателя Херста H (Таблица 2).

Таблица 2 – Оценка показателя Херста различными методами

	R/S -статистика	Периодограммный	Изменения дисперсии	Абсолютных значений
\hat{H}	0,8512	0,8697	0,8784	0,863
σ^2	0,0049	0,0219	0,0009	0,0013
СКО	0,0701	0,1482	0,03	0,0363

Заключение

Современные сети — сети с широким набором всевозможных сервисов и услуг. Часть из них являются сервисами времязависимого трафика (сервисы IP-телефонии, потокового видео, различных видов диспетчеризации оборудования), другая часть, например, файлообменные сети FTP-серверов не критичны к текущей пропускной способности. Особого внимания заслуживают сервисы распределенного хранения данных P2P. Повсеместное распространение этого вида услуг, отсутствие «узкого горла» как элемента системы, ограничивающего общий объем транслирующихся данных позволяют рассматривать дальнейшее развитие сетей P2P как фактор, вносящий дисбаланс в существующую сетевую инфраструктуру. Построение систем on-line мониторинга с возможностью динамического управления выделенной пропускной способностью для определенных видов сервиса — важная перспективная задача.

Подводя итоги работы необходимо сделать следующий вывод — трафик гигабитного Ethernet канала проявляет самоподобные свойства с достаточно большим показателем Херста. Особое внимание заслуживают области, вызванные антропогенным характером, т.к. именно в это время возникает большая вероятность переполнения буферов устройств, что может привести к возникновению очередей в системе и как следствие к резкому ухудшению качества обслуживания всего спектра существующих сервисов, предоставляемые провайдером услуг связи.

Самоподобные свойства трафика позволяют с достаточной степенью достоверности прогнозировать появление на сегменте сети временных периодов с перегрузкой по производительности оборудования и линий связи, что в свою очередь, делает возможным построение системы с динамическим управлением возможной пропускной способности для отдельных видов трафика. Как известно, прогнозирование играет значимую роль в разработке алгоритмов, направленных на повышение качества обслуживания. Сегодня с появлением широкополосных сетевых сервисов, задачи повышения качества обслуживания, долгосрочного прогнозирования загрузки каналов связи, инженерия и управление сетью становятся все более актуальными.

Таким образом, проделанная работа позволяет приблизиться к следующим пунктам дальнейшего исследования:

Выявить какой из сетевых сервисов является доминирующим и оценить его влияние на сеть провайдера связи, особенно в местах наиболее высокой интенсивности сетевого трафика, вызванные антропогенным характером.

Разработать алгоритм и методы динамического управления пропускной способности доминирующего сетевого сервиса или группы сетевых сервисов.

Список литературы

1. Norris, M. Gigabit Ethernet / M. Norris // Technology and Applications. — Artech House, 2003.
2. Петров, В. В. Статистический анализ сетевого трафика / В. В. Петров. — Москва, 2003.
3. Karagiannis, T. Poisson View of Internet Traffic / T. Karagiannis, M. Molle, M. A. Faloutsos // Nonstationary INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, March 2004. — vol.3, issue, 7-11. — P. 1558 - 1569.
4. Афонцев, Э. В., Разработка методики выявления аномалий трафика в магистральных интернет-каналах: дис. ... канд. техн. наук / Э. В. Афонцев. — Екатеринбург, 2007.
5. Шелухин, О. И. Самоподобие и фракталы. Телекоммуникационные приложения / О. И. Шелухин, А. В. Осин, С. М. Смольский. — М.: Физматлит, 2008.
6. Крюков Ю. А. Разработка программно-ориентированной многоцелевой сети распределенных вычислений масштаба малого города: дис. ... канд. техн. наук / Ю. А. Крюков. — Дубна, 2004.
7. Hurst, H. E. Long-term storage capacity of reservoirs. Trans. Am. Soc. Civil Engineers. — 116:770-799, 1951.
8. Rose, O. Estimation of the Hurst Parameter Time Series. — 1996.