

ИССЛЕДОВАНИЕ САМОПОДОБИЯ ТРАФИКА ВЫСОКОСКОРОСТНОГО КАНАЛА ПЕРЕДАЧИ ПАКЕТНЫХ ДАННЫХ

Крюков Юрий Алексеевич¹, Чернягин Денис Викторович²

¹Международный университет природы, общества и человека, кафедра САУ
141980, г. Московская обл., ул. Университетская, 19, e-mail: kua@uni-dubna.ru

²Международный университет природы, общества и человека, кафедра САУ
141980, г. Московская обл., ул. Университетская, 19, e-mail: dancher2000@mail.ru

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 происходит радикальные изменения в различных типах сервисов и телекоммуникационных технологиях. Интегрирование гетерогенных сетевых сервисов, 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.

Методика 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 ста миллионов кадров снифер останавливался и производил обработку данных, приводя полученные реализации в эквидистантную форму с постоянным

шагом m при помощи процедуры агрегирования, а затем повторно производил захват трафика и анализ данных. Время начала захвата трафика 07:41:16.05.2009, а общее время измерения составило 15:08:54. Таким образом, было получено 18 профилей сетевого трафика. Логическая схема организации сети отображена на Рис. 1.

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

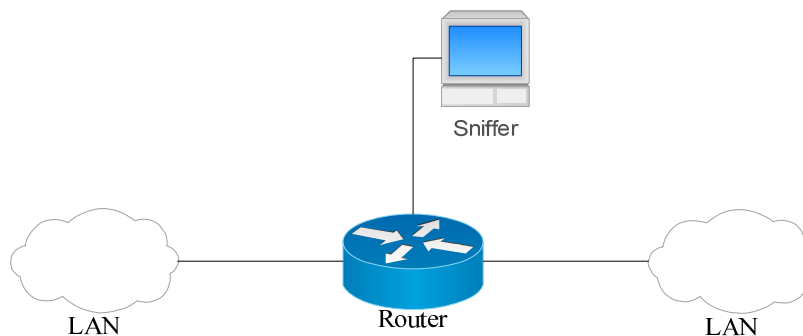


Рис. 1 – Логическая схема организации мониторинга сетевого трафика

контрольной последовательности кадра и поля расширения несущей. Размер преамбулы зафиксирован и составляет 64 бита, служебной части в 144 бита, полезной нагрузки в 368 бит, контрольной последовательности в 32 бита, 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.

Анализ полученных данных

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. Примеры профилей байтовой интенсивности, а также интенсивностей пакетов в единицу времени, приведены на (Рис. 2). 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.

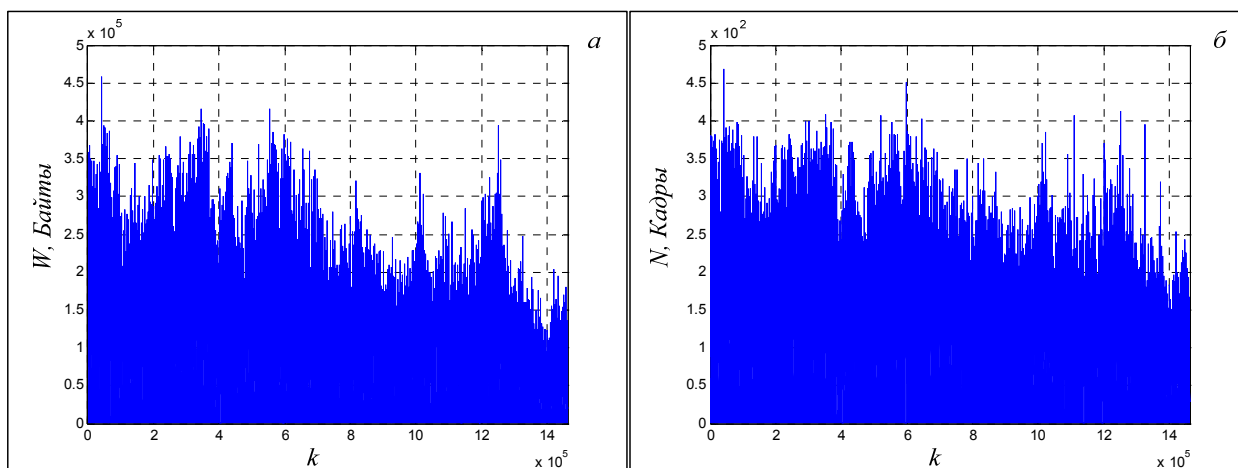


Рис. 2 - Измеренные реализации: а) байтовая интенсивность; б) интенсивность пакетов

Каждая точка на данном графике представляет собой количество байт информации, переданной в магистральном канале за интервал времени в 0,001 секунду. Длительность данных реализаций составляет 1461867 точек или 24 минуты и 21 секунда. Как результат, для того чтобы передать такой трафик без потерь, пропускная способность канала должна соответствовать уровню пиковых выбросов, т.е. в данном случае быть не менее 3668800 Бит/мс.

Визуальный анализ гистограмм распределения длин кадров показывает (Рис. 3), что в сети преобладают кадры размером 1518 байт, т.е. максимальный размер кадра, поддерживаемый технологией Ethernet.

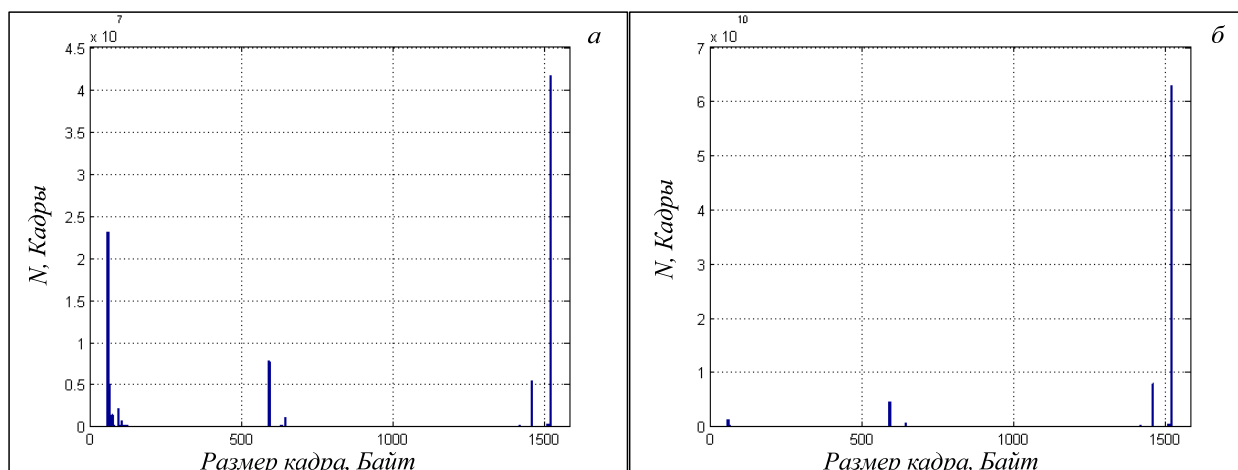


Рис. 3 – Гистограммы: а) распределение длин кадров, б) распределение количества данных переданное кадрами соответствующей длины

Очевидно, что для практических целей более полезной является информация о степени загруженности сети. Информация о количестве пакетов в единицу времени может визуально вводить в заблуждение, поскольку в сети может присутствовать много небольших управляющих пакетов, не несущих полезной информации, и пакетов минимальной длины, которые создают пики, несовпадающие с пиками байтовой скорости.

В Табл. 1 приводится статистика полученных реализаций. Из таблицы видно, что за время создания профилей сетевого трафика было передано 1,54 Тбайт данных, 75,33% в среднем из которых были переданы кадром максимального размера, и 0,0044% данных в среднем было передано кадром минимального размера. В приведенной таблице также может вызвать интерес профиль трафика под номером 16, который имеет минимальный процент данных переданных кадром размером 1518 байт. Это можно объяснить тем, что в этот временной период данные передавались преимущественно кадрами, длина которых отлична от рассматриваемых.

Табл. 1 – Статистика распределений длин кадров

№	Название реализации	Количество данных переданных кадром размером 64 байт, Байт	Количество данных переданных кадром размером 1518 байт, Байт	Всего переданных данных, Байт	Процент данных переданных кадром размером 64 байта от всех переданных данных, %	Процент данных переданных кадром размером 1518 байта от всех переданных данных, %
1	07.41.16.05.10.2009	2084352	80087938388	94182789518	0,0022%	85,0346%
2	39.17.17.05.10.2009	1596992	75781109522	96592015584	0,0017%	78,4548%
3	20.47.17.05.10.2009	1770816	62788527758	81889732017	0,0022%	76,6745%
4	33.20.18.05.10.2009	1813376	78568949758	94668673869	0,0019%	82,9936%
5	44.54.18.05.10.2009	2029504	71497815786	87749499857	0,0023%	81,4795%
6	35.25.19.05.10.2009	1653056	62512934338	81533283187	0,0020%	76,6717%
7	06.57.19.05.10.2009	1717504	64159153472	83324079817	0,0021%	76,9995%
8	30.30.20.05.10.2009	1506432	73557785382	88694541378	0,0017%	82,9338%
9	48.58.20.05.10.2009	1625088	65226962532	84078551606	0,0019%	77,5786%
10	51.30.21.05.10.2009	1518912	67350417176	87890062866	0,0017%	76,6303%
11	39.00.22.05.10.2009	1720128	66441331334	85141916508	0,0020%	78,0360%
12	42.31.22.05.10.2009	1854400	67204367652	87147616740	0,0021%	77,1156%
13	19.06.23.05.10.2009	2289792	64344943982	80737410877	0,0028%	79,6966%
14	13.45.23.05.10.2009	4077824	53103018586	77754997124	0,0052%	68,2953%
15	19.52.00.06.10.2009	6315328	49301555350	84050882091	0,0075%	58,6568%
16	49.34.02.06.10.2009	19381376	35877170188	73215120704	0,0265%	49,0024%
17	43.40.07.06.10.2009	7658432	74421216554	95619080236	0,0080%	77,8309%
18	58.03.10.06.10.2009	4934272	60459594548	84147015412	0,0059%	71,8500%
Итого:		65547584	1172684792306	1548417269391	0,0044%	75,3297%

Приведенная гистограмма байтовой интенсивности (Рис. 4, а) соответствует логнормальному распределению которое имеет субэкспоненциально спадающие хвосты, но обладают конечной дисперсией, т.е. относится к так называемым распределениям с «тяжелыми хвостами» [3], [4]. Гистограмма на (Рис. 4, б) имеет достаточно сложную структуру и свидетельствует о наличие в распределение «тяжелого хвоста», что соответствует выводам, сделанным в [3].

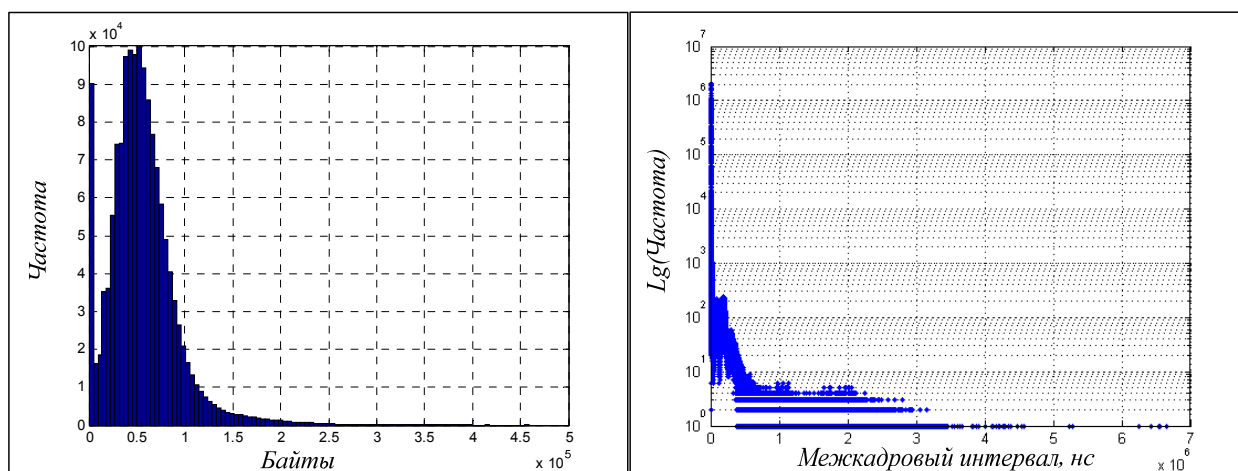


Рис. 4 – Гистограммы: а) распределение байтовой интенсивности; б) распределение межкадровых интервалов

Известно, что самоподобные процессы с H от 0,5 до 1 обладают свойством долговременной зависимостью (ДВЗ), т.е. их автокорреляционная функция (АКФ) на бесконечности не сходится к нулю. Оценив степень затухания автокорреляционной функции, можно оценить H . Таким образом, необходимо решить задачу регрессии и вычислить по экспериментальной АКФ методом наименьших квадратов параметры a_0 и b модели (1).

$$R(k) = a_0 \cdot k^{-b} \quad \text{при } k \rightarrow \infty, \quad (2)$$

При построении автокорреляционной функции рассматривались первые 200 задержек (Рис. 5, а).

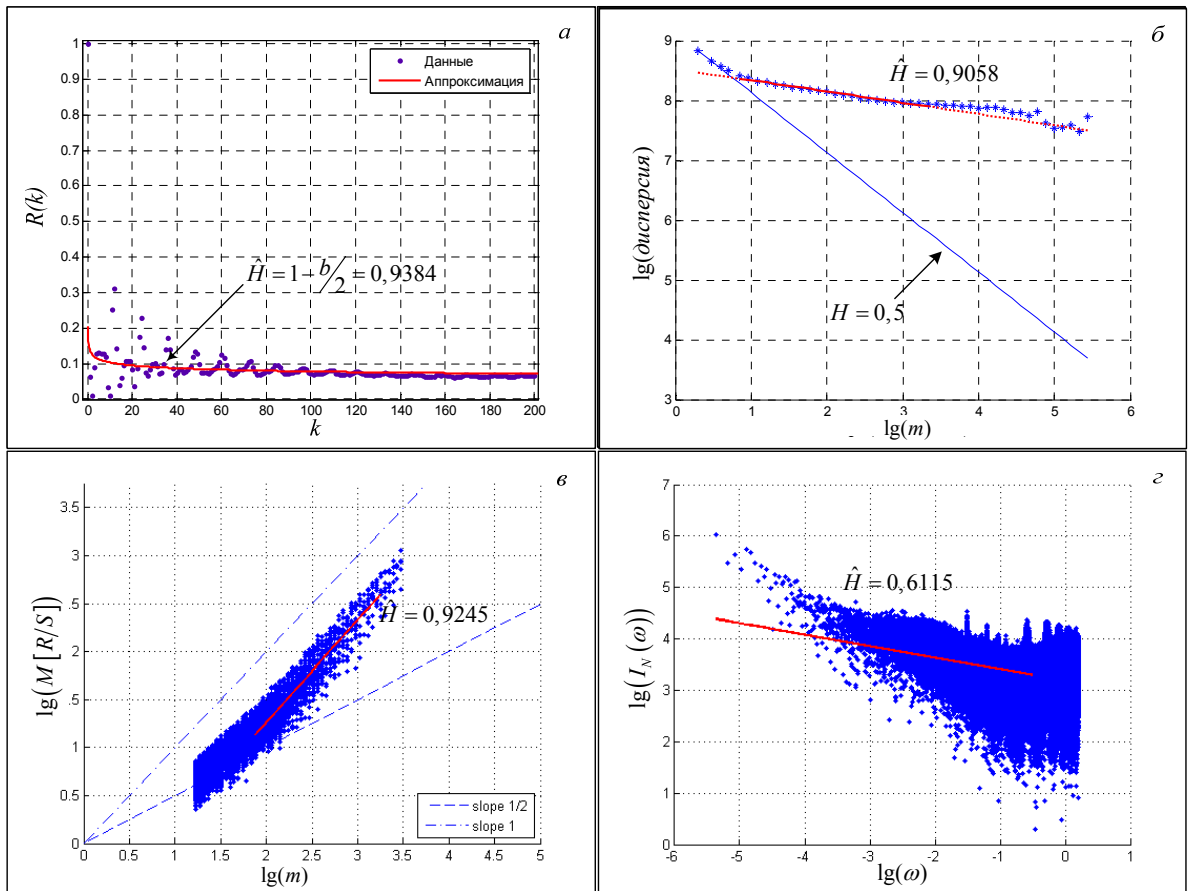


Рис. 5 – Графики оценки показателя Херста: а) коэффициент корреляции от задержки; б) изменение дисперсии; в) R/S – статистика; г) периодограммный метод.

Анализ изменения дисперсии (Рис. 5, б) основывается на свойстве медленно затухающей дисперсии самоподобных процессов при объединении (агрегировании) (1). В соответствии с этим дисперсия объединенного (точно или приблизительно) самоподобного процесса удовлетворяет зависимости

$$\sigma^2 \left[Y^{(m)} \right] \sim m^{-b}, \quad (3)$$

где b – параметр, связанный с H соотношением $b = 2 - 2H$.

При агрегировании процесса с различными уровнями m дисперсия обычно затухает очень быстро (при $H = 0,5$). Исключения составляют самоподобные процессы, для которых дисперсия затухает медленно, по степенному закону (при больших значениях H).

Прологарифмировав обе части соотношения (3) для объединенной дисперсии получим выражение

$$\log \left(\sigma^2 \left(Y^{(m)} \right) \right) \sim -b \log(m) + \log(a) \text{ при } m \rightarrow \infty$$

Как видно, оценку b можно получить вычислением $\sigma^2 \left[Y^{(m)} \right]$ для различных значений m , отображая результаты от $\log(m)$ и проведя прямую линию по методу наименьших квадратов через полученные точки. Оценка \hat{b} для b определяется как отрицательный наклон линии. Значение наклона в интервале $(-1; 0)$ подразумевает самоподобность.

Для выявления самподобия во временных рядах часто прибегают к R/S-статистике (Рис. 5, в).

На практике удобно пользоваться выражением $\log \left(M \left(\frac{R(n)}{S(n)} \right) \right) \sim H \log(n) + \log(c)$ при $n \rightarrow \infty$.

Воспользовавшись этим выражением, можно оценить H , изобразив график зависимости

$\log\left(M\left(\frac{R(n)}{S(n)}\right)\right)$ от $\log(n)$. Подбирая прямую по методу наименьших квадратов к точкам R/S-графика, по наклону линии регрессии находят оценку для H .

Рассмотренные методы оценки показателя Херста H не слишком точны и дают только оценку уровня самоподобности во временном ряде. Следовательно, этот метод может использоваться только для того, чтобы протестировать является ли временной ряд самоподобным, и, если является, получить грубую оценку H .

Оценка показателя Херста, основанная на графике спектральной плотности (Рис. 5, з), составляет суть метода, который обеспечивает большую статистическую строгость, чем оценки основанные на объединении. Однако ценой существования параметрического метода является требование, чтобы параметрическая модель процесса была известна заранее. Периодограмма $I_N(\omega)$ оценивает спектральную плотность дискретного стохастического процесса Y_t и может быть оценена рядом (4) на интервале времени N :

$$I_N(\omega) = \frac{1}{2\pi N} \left| \sum_{k=1}^N Y_k e^{jk\omega} \right|^2, \quad \omega \in [0; \pi] \quad (4)$$

где $\{Y_k\}$ - временной ряд; N – длина временного ряда.

Учитывая, что самоподобность влияет на характер спектра $S(\omega)$ при $\omega \rightarrow 0$, должен получиться график зависимости спектральной плотности вида

$$I_N(\omega) \sim [\omega]^{1-2H} \quad \text{при } \omega \rightarrow 0. \quad (5)$$

Начертив график $\log(I_N(\omega))$ от $\log(\omega)$ (только для низких частот), подбирают касательную прямую линию к кривой. Наклон линии будет приблизительно равен $1 - 2H$.

Важно отметить, что 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.

Практически 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.

Для того чтобы избежать сложных тестов на проверку стационарности можно воспользоваться следующим методом. 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, анализа изменения дисперсий, периодограмный анализ и анализ АКФ. В [9] подробно излагается преимущества и недостатки методов используемые 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

$$\hat{H}_N = \frac{N}{D} \sum_{i=1}^{D/N} \hat{H}_i. \quad (2)$$

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

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

Табл. 2 – Оценка показателя Херста различными методами

№	Название реализации	Метод оценки показателя Херста			
		R/S - статистика	Периодограммный	Изменения дисперсии	АКФ
1	07.41.16.05.10.2009	0,9062	0,6690	0,9597	0,9534
2	39.17.17.05.10.2009	0,9116	0,6744	0,9389	0,9211
3	20.47.17.05.10.2009	0,9245	0,7138	0,9560	0,9476
4	33.20.18.05.10.2009	0,9477	0,5874	0,9379	0,9348
5	44.54.18.05.10.2009	0,9031	0,7014	0,9496	0,9356
6	35.25.19.05.10.2009	0,8955	0,6358	0,9188	0,9255
7	06.57.19.05.10.2009	0,8731	0,6829	0,9058	0,9205
8	30.30.20.05.10.2009	0,8576	0,6706	0,9142	0,9110
9	48.58.20.05.10.2009	0,9643	0,6900	0,9466	0,9344
10	51.30.21.05.10.2009	0,8911	0,6794	0,9129	0,9083
11	39.00.22.05.10.2009	0,8987	0,6998	0,9062	0,9152
12	42.31.22.05.10.2009	0,9418	0,7108	0,9233	0,9284
13	19.06.23.05.10.2009	0,9369	0,7191	0,9545	0,9511
14	13.45.23.05.10.2009	1,0475	0,7050	0,9645	0,9816
15	19.52.00.06.10.2009	1,0284	0,7263	0,9820	0,9629
16	49.34.02.06.10.2009	1,0144	0,7286	0,9514	0,9797
17	43.40.07.06.10.2009	1,0609	0,7154	0,9761	0,9715
18	58.03.10.06.10.2009	0,9990	0,7308	0,9606	0,9506
\hat{H}		0,9446	0,6911	0,9422	0,9407
СКО		0,0614	0,0360	0,0237	0,0229

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 [9]. 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.

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. Шелухин, О. И. Самоподобие и фракталы. Телекоммуникационные приложения / О. И. Шелухин, А. В. Осин, С. М. Смольский. – М. : Физматлит, 2008.
4. Крюков Ю. А. Разработка программно-ориентированной многоцелевой сети распределенных вычислений масштаба малого города: дис. ... канд. техн. наук / Ю. А. Крюков. – Дубна, 2004.
5. 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.
6. Афонцев Э.В., Разработка методики выявления аномалий трафика в магистральных интернет-каналах. *Диссертация*, 2007.
7. Hurst, H. E. Long-term storage capacity of reservoirs. Trans. Am. Soc. Civil Engineers. — 116:770-799, 1951.
8. O Rose., Estimation of the Hurst Parameter Time Series. 1996.
9. Крюков, Ю. А. Исследование влияния P2P потоков на производительность магистрального канала сети «ЛАНПОЛИС» / Ю. А. Крюков, Д. В. Чернягин // sanse.ru: сайт электронного журнала [Электронный ресурс]. — Режим доступа: <http://sanse.ru/download/33> (дата обращения 09.10.2009).