

Работа сетевого сервера при самоподобной (self-similar) нагрузке

мгр. Ильницкий С.В.

Исследования различных типов сетевого трафика доказывают, что сетевой трафик является самоподобным (self-similar) или фрактальным (fractal) по своей природе, т.е. в нем присутствуют так называемые вспышки или пачки (burst) пакетов, наблюдаемые в различных временных интервалах (от миллисекунд до минут или даже часов) и корреляция между пакетами. Из этого следует, что широко используемые в настоящее время методы моделирования и расчета сетевых систем, основанные на использовании пуассоновских потоков, не дают полной и точной картины происходящего в сети.

В данной публикации проведен анализ реального трафика компьютерной сети, используя один из методов определения самоподобности. На основании данных, полученных в ходе экспериментов, доказана самоподобная природа реального сетевого трафика. Разработана модель, которая описывает самоподобный трафик с различными степенями самоподобности и проведено сравнение результатов с результатами, полученными для модели, обслуживающей системы с пуассоновским трафиком.

КОМПЬЮТЕРНЫЕ СЕТИ, ТРАФИК, МОДЕЛИРОВАНИЕ,
МЕТОДЫ РАСЧЕТА, САМОПОДОБНОСТЬ.

Keywords: network traffic, self-similar traffic, traffic analysis.

1. Введение

Исследования различных типов сетевого трафика доказывают, что сетевой трафик является самоподобным (self-similar) или фрактальным (fractal) по своей природе, т.е. в нем присутствуют так называемые вспышки или пачки (burst) пакетов, наблюдаемые в различных временных интервалах (от миллисекунд до минут или даже часов). Из этого следует, что широко используемые в настоящее время методы моделирования и расчета сетевых систем, основанные на использовании пуассоновских потоков, не дают полной и точной картины происходящего в сети.

Целью данной публикации является анализ реального трафика компьютерной сети, используя один из методов определения самоподобности, разработка модели, которая описывает самоподобный трафик с различными степенями самоподобности и

сравнение результатов с результатами традиционного моделирования.

2. Математическое описание самоподобного (фрактального) процесса.

Непрерывный стохастический процесс $X(t)$, считается статистически самоподобным с параметром H ($0.5 \leq H \leq 1$), если для любого положительного числа a , процессы $X(t)$ и $a^{-H} X(at)$ будут иметь идентичные распределения, т.е. иметь одинаковые статистические свойства для всех положительных целых n

$$\{X(t_1), X(t_2) \dots X(t_n)\} \stackrel{D}{\sim} \{a^{-H} X(at_1), a^{-H} X(at_2) \dots a^{-H} X(at_n)\} \quad (1)$$

Отношение $\stackrel{D}{\sim}$ обозначает асимптотическое равенство в смысле распределения. Практически статистическая самоподобность подразумевает, что выполняются следующие условия [8]:

- среднее $E[X(t)] = \frac{E[X(at)]}{a^H}$ (2)

- дисперсия $Var[X(t)] = \frac{Var[X(at)]}{a^{2H}}$ (3)

- функция автокорреляции $R(t, \tau) = \frac{R(at, a\tau)}{a^{2H}}$ (4)

H - параметр Харста (Hurst), показывает “степень” самоподобности. Значение $H = 0.5$ показывает отсутствие самоподобности, а большие значения H (близкие к 1) показывают большую степень самоподобности или длительно-зависимости (long-range dependent, LRD) в процессе. Это обозначает, что если LRD процесс имеет тенденцию к увеличению (или уменьшению) в прошлом, то с большой вероятностью он будет иметь тенденцию к увеличению (или уменьшению) в будущем.

3. Математическое описание дискретного самоподобного процесса.

Рассмотрим временной процесс $X = \{X_n, n \in Z^+\}$ и определим другой временной процесс (m-aggregated) $X^{(m)} = \{X_n^{(m)}, n \in Z^+\}$ путем усреднения оригинального временного процесса на непересекающихся соседствующих блоках длиной m :

$$X_n^{(m)} = \frac{1}{m} \sum_{i=nm-(m-1)}^{nm} X_i \quad (5)$$

$X^{(1)}$ представляет в этом случае наибольшее, возможное для процесса, разрешение. Последующие эволюции процесса $X^{(m)}$ могут быть получены путем m -усреднения процесса X_n , например

$$X_n^{(4)} = \frac{X_{4n-3} + X_{4n-2} + X_{4n-1} + X_{4n}}{4} \quad (6)$$

Процесс $X^{(m)}$ представляет собой менее детализированную копию процесса $X^{(1)}$. В случае, если статистические свойства (среднее, дисперсия) сохраняются при усреднении, тогда процесс является самоподобным.

Существует два класса самоподобных процессов, так называемые точно самоподобные и асимптотически самоподобные процессы.

Процесс X называется точно самоподобным с параметром β ($0 < \beta < 1$) если для $m \in Z^+$ выполняются следующие условия:

- дисперсия $Var[X^{(m)}] = \frac{Var[X]}{m^\beta}$ (7)

- функция автокорреляции $R(k, X^{(m)}) = R(k, X)$ (8)

Параметр β связан с параметром Харста H следующим соотношением:

$$\beta = 2(1-H) \quad (9)$$

Существует другой класс самоподобных процессов – так называемых асимптотически самоподобных. Процесс X называется асимптотически самоподобным если для больших k

- дисперсия $Var[X^{(m)}] = \frac{Var[X]}{m^\beta}$ (10)

- функция автокорреляции $R(k, X^{(m)}) \rightarrow R(k, X)$ (11)

при $m \rightarrow \infty$

Имеются наблюдения, что для обоих классов самоподобных процессов дисперсия $Var[X^{(m)}]$ уменьшается намного медленнее чем $\frac{1}{m}$ при $m \rightarrow \infty$ по сравнению со стохастическими процессами, где

дисперсия уменьшается пропорционально $\frac{1}{m}$ и приближается к 0 при $m \rightarrow \infty$.

Наиболее точным свойством самоподобных процессов является то, что функция автокорреляции не вырождается при $m \rightarrow \infty$, в отличие от стохастических процессов, где $R(k, X) \rightarrow 0$ при $m \rightarrow \infty$.

4. Описание модели самоподобного потока.

Существует несколько подходов в формировании самоподобного потока. Наиболее известным является метод, первоначально предложенный Мандельбротом [9].

Данный метод основан на суперпозиции нескольких (строго чередующихся) независимых и имеющих одинаковое распределение ON/OFF источников, интервалы между ON и OFF периодами которого обладают эффектом Ноа (Noah effect). Под строго чередующимися ON/OFF источниками мы подразумеваем модель, где ON и OFF периоды строго чередуются, длительности ON периодов независимы и имеют одинаковое распределение, длительности OFF периодов тоже независимы и имеют одинаковое распределение, и последовательности длительностей ON и OFF периодов не зависят друг от друга. При этом длительности ON и OFF периодов могут иметь разные распределения.

Причем, как показано в [10], именно эффект Ноа в распределении длительностей ON/OFF периодов является основной точкой при моделировании самоподобного трафика в отличие от моделей, когда используются стандартные экспоненциальное или геометрическое распределение. Эффект Ноа является синонимом синдрома бесконечной дисперсии, появившемуся благодаря эмпирическим наблюдениям того, что многие природные явления могут быть описаны распределением с бесконечной дисперсией. Математически для достижения эффекта Ноа можно использовать распределение Парэто или логарифмически-нормальное распределение, которые также часто называют *heavy-tailed distributions*. Наиболее популярным является распределение Парэто.

Распределение Парэто имеет функцию распределения

$$F(x) = 1 - \left(\frac{\beta}{x} \right)^\alpha \quad (12)$$

где α - параметр формы, характеризующий, будет ли распределение иметь конечное или бесконечное среднее и

дисперсию, и β - параметр нижней границы (минимальное значение случайной величины x).

Плотность распределения Парэто задается функцией

$$f(x) = \frac{\alpha}{\beta} \left(\frac{\beta}{x} \right)^{\alpha+1} \quad (13)$$

для $x > \beta$ и $\alpha > 0$,

$$f(x) = F(X) = 0 \quad (14)$$

для $x \leq \beta$.

Параметр α определяет среднее и дисперсию x следующим образом:

- для $\alpha \leq 1$ распределение имеет бесконечное среднее
- для $1 \leq \alpha \leq 2$ распределение имеет конечное среднее и бесконечную дисперсию
- для $\alpha \leq 2$ распределение имеет бесконечную дисперсию

Также существует отношение между параметром α и параметром Харста H :

$$H = \frac{3 - \alpha}{2} \quad (15)$$

5. Анализ трафика сетевого сервера

5.1. Описание исследуемой сети

Анализ производился в корпоративной сети компании «Х», которая является компанией, разрабатывающей программное обеспечение и которая широко использует для этого базы данных внутри и за пределами корпоративной сети. Максимальная скорость передачи данных с корпоративной сети равняется 10 Mbit/s. Из большого числа доступных серверов был выбран один, работавший большую часть времени с максимальной нагрузкой. Основными функциями данного сервера в сети являются: WEB-сервер, сервер баз данных, файловый сервер и сервер печати.

Результаты измерений трафика были получены с помощью Multi Router Traffic Grapher (MRTG). Данные собирались в течение месяца, 24 часа в сутки с интервалом 5 минут. Для дальнейшего анализа полученные данные были разделены на группы: трафик в

течение суток, трафик в течение недели, трафик в течение месяца. Трафик в течение суток дополнительно был разделен на подгруппы, в зависимости от интенсивности нагрузки:

- с 02:00 до 08:00 – низкая интенсивность трафика нагрузки;
- с 08:00 до 14:00 – средняя интенсивность трафика нагрузки;
- с 14:00 до 20:00 – высокая интенсивность трафика нагрузки;
- с 20:00 до 02:00 – максимальная интенсивность трафика нагрузки.

5.2. Результаты измерений

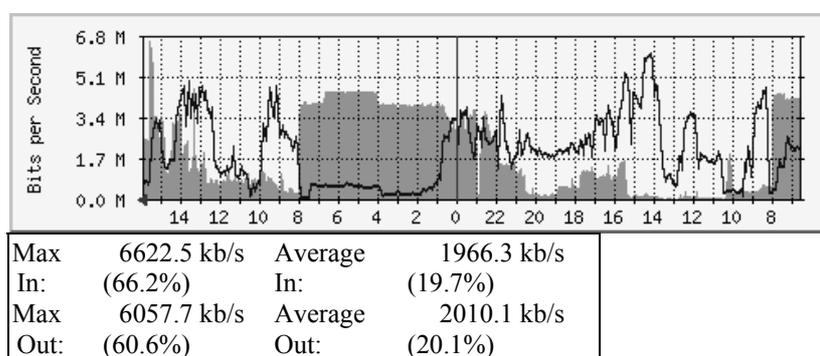


Рисунок 1: Трафик в течение суток

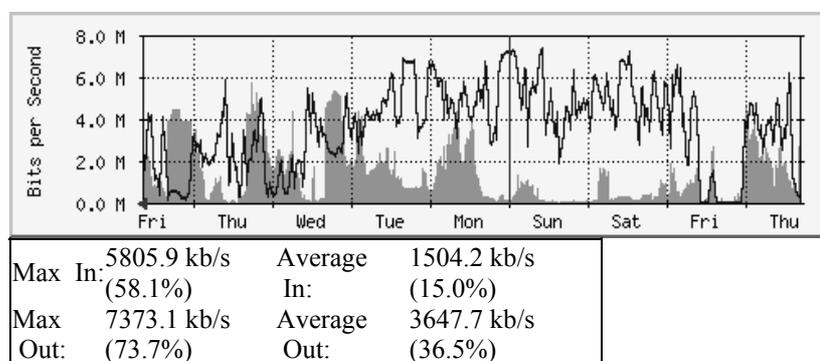


Рисунок 2: Трафик в течение недели

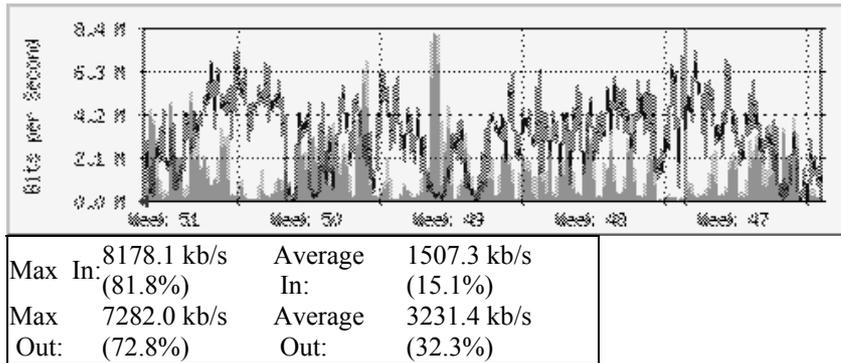


Рисунок 3: Трафик в течение месяца

Итак, предполагаем, что трафик является самоподобным. Наша цель доказать это утверждение. Для этого воспользуемся методом абсолютных моментов.

В данном методе исходная последовательность с длиной N разделяется на блоки с длиной m . На границах блока последовательность имеет среднее значение (дисперсия):

$$X^{(m)}(k) := \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad k=1,2,\dots,[N/m]. \quad (16)$$

Для каждого блока рассчитываем дисперсию $X^{(m)}$ и математическое ожидание \bar{X} для всей последовательности. После этого для каждого блока находим момент n :

$$AM_n^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k) - \bar{X}|^n. \quad (16)$$

В данном выражении $n=1$ (абсолютное среднее значение). Далее последовательность обозначим как m и построим график в логарифмическом масштабе – зависимость средних значений (дисперсий) абсолютных моментов для последовательности от m . С помощью полученных точек строим аппроксимирующую прямую по методу минимального среднеквадратического отклонения от экспериментальных данных. В результате, получившийся наклон линии будет равен β . С помощью β можем найти коэффициент самоподобности $H = 1 - |\beta|$.

Необходимо, чтобы длина каждого блока и число блоков были большими. Если последовательность – это не процесс с медленно изменяющейся зависимостью, то $H=0.5$ и наклон аппроксимационной линии будет равен $1/2$. Если процесс самоподобен, тогда $0.5 < H < 1.0$ и наклон линии будет меньше чем $1/2$.

Данный метод реализован с помощью программы, написанной на языке MatLab 5.0.

С помощью данной программы были построены log-log графики, графики зависимости дисперсии от среднего m и был определен параметр Харста для каждого дня недели, месяца, а также зависимость параметра Харста от нагрузки в течение суток.

Используя вышеописанную методику, были получены результаты, один из которых можно увидеть на **Рисунке 4**.

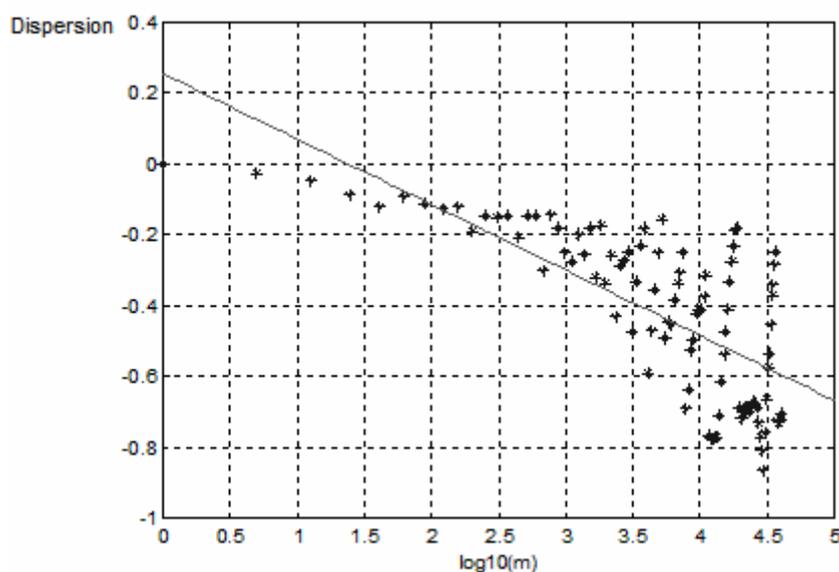


Рисунок 4. Дисперсионно-временной график для данных, полученных в понедельник

Значения коэффициента Харста, рассчитанные с помощью метода абсолютных моментов для различных интервалов времени, показаны в **Таблице 1** и **Таблице 2**.

День недели	Abs(β)	Коэф. Харста
Понедельник	0.1845	0.82
Вторник	0.2206	0.78
Среда	0.2537	0.75
Четверг	0.1620	0.84
Пятница	0.2120	0.79
Суббота	0.1518	0.85
Воскресенье	0.2441	0.76
Общий трафик за неделю	0.2131	0.79

Таблица 1: Коэффициент Харста для трафика за неделю

Интервал времени	Abs(β)	Коэф. Харста
02:00 - 08:00 Низкая интенсивность	0.1845	0.71
08:00 – 14:00 Средняя интенсивность	0.2206	0.73
14:00 – 20:00 Высокая интенсивность	0.2537	0.76
20:00 – 02:00 Максимальная интенсивность	0.1620	0.81
Трафик за сутки	0.2184	0.75

Таблица 2: Коэффициент Харста для трафика за сутки

Трафик за сутки был специально разделен на интервалы. Существует мнение, что коэффициент самоподобности зависит от интенсивности трафика, а также от типа передаваемой информации. На самом деле, интенсивность трафика очень мала в интервал времени 02:00 – 08:00 и это связано с особенностями поведения пользователей. В данном интервале времени параметр Харста минимален. Рабочий день начинается в 08:00 и интенсивность трафика невелика, но, несмотря на это, параметр Харста уже увеличивается. В интервале с 20:00 до 02:00, когда программисты начинают тестировать написанный программный код в автоматическом режиме, интенсивность потока заявок к серверу вырастает очень значительно и параметр Харста также возрастает до значения 0.81. Сетевой трафик в случае максимальной интенсивности (20:00 – 02:00) имеет очень высокую степень самоподобности.

После анализа результатов, можем сделать некоторые выводы:

- Трафик в локальной сети – это самоподобный процесс.
- Эффект самоподобности проявляется в широком диапазоне времени: от нескольких часов до нескольких месяцев.
- Коэффициент самоподобности трафика меняется в пределах $\approx 0.7 - 0.85$.
- В случае если возрастает интенсивность трафика, также возрастает коэффициент самоподобности.

- Для трафика в течение суток значение коэффициента Харста одинаково для любой интенсивности.
- Для трафика в течение месяца коэффициент Харста равен ≈ 0.81 .

5.3. Моделирование самоподобного трафика в среде GPSS

Для того чтобы смоделировать самоподобный входной трафик, для начала необходимо задать все главные параметры, характерные для трафика, использовавшегося в ходе экспериментов.

Из топологии исследуемой сети известно, что у нас имеется 350 рабочих станций, посылающих запросы к серверу, подключенному к маршрутизатору. Поэтому для моделирования нам потребуется 350 ON-OFF источников запросов, один маршрутизатор (для упрощения будем рассматривать его как концентратор) и один прибор массового обслуживания (сервер) (**Рисунок 5**).

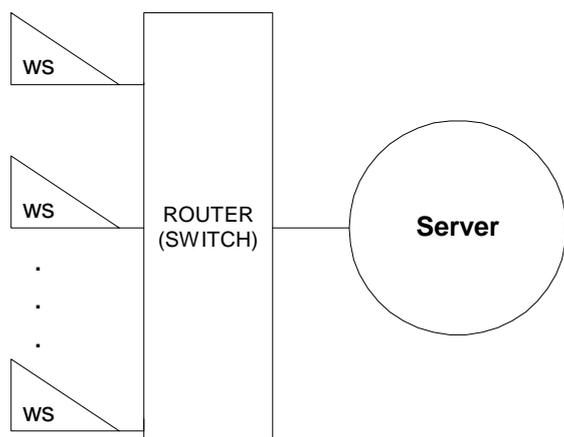


Рисунок 5. Упрощенная схема анализируемой сети.

Для примера возьмем трафик за сутки, полученный в ходе экспериментов (**Рисунок 6**):

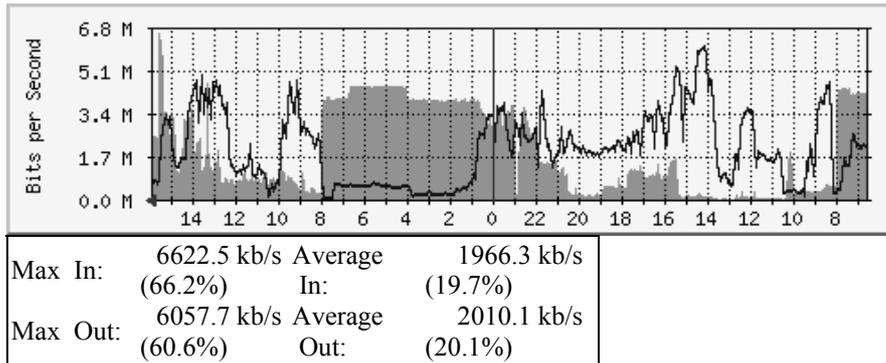


Рисунок 6. Трафик за сутки

Как видим, средний выходной трафик достигает значения 2 Mbit/s. Это означает, что канал загружен на 20%. Будем использовать это значение в качестве средней интенсивности работы ON-OFF источников. Следовательно, ON-период в 5 раз меньше чем OFF-период. Функция GPSS позволяет генерировать распределение Парето с параметром α . В нашем эксперименте параметр равен $\alpha=1.4$, что соответствует значению коэффициента Харста $H=0.8$. Усредняющий коэффициент m будем изменять от 1 до 10 000. Текст программы на языке GPSS приведен ниже:

```

ALFA      EQU 1.4
TL1       EQU 5000
TM1       EQU 1000
TL2       EQU 100
AGGM      EQU 1
USTR      STORAGE 100
PACK      FVARIABLE X$SUM1/X$I1
DPPARETO  FVARIABLE PARETO(1,1,ALFA)
DPPARETOL FVARIABLE TL1#PARETO(1,1,ALFA)
DPPARETOM FVARIABLE TM1#PARETO(1,1,ALFA)
TAB1      TABLE V$PACK,0,1,100
          INITIAL X$SUM1,0
          INITIAL X$I1,0
          INITIAL X$L3,0

          *** 1 blok

***
          GENERATE V$DPPARETO,, ,35
MET1      ADVANCE V$DPPARETOL
          ENTER USTR
          ADVANCE V$DPPARETOM
          LEAVE USTR
          TRANSFER ,MET1

          *** 2 blok ***
          GENERATE TL2
          SAVEVALUE SUM1+,S$USTR
          SAVEVALUE I1+,1

```

```

TEST E X$I1,AGGM,MET2
TABULATE TAB1
SAVEVALUE L3,V$PACK
SAVEVALUE SUM1,0
SAVEVALUE I1,0
MET2 SPLIT X$L3,MET3
TERMINATE

*** 3 blok
***
MET3 TERMINATE

*** 4 blok ***
GENERATE 10000000
TERMINATE 1

```

Проверим, является ли полученный таким образом поток самоподобным с параметром Харста $H = 0.8$ (значение получено ранее с помощью метода абсолютных моментов).

m	log(m)	Var(x)	log[Var]
1	0	2,371	0,374932
2	0,30103	2,334	0,368101
3	0,477121	2,305	0,362671
5	0,69897	2,251	0,352375
10	1	2,115	0,32531
20	1,30103	1,843	0,265525
50	1,69897	1,297	0,11294
100	2	1,018	0,007748
200	2,30103	0,826	-0,08302
500	2,69897	0,632	-0,19928
1000	3	0,5	-0,30103
2000	3,30103	0,396	-0,4023
5000	3,69897	0,281	-0,55129
10000	4	0,222	-0,65365

Таблица 3: Дисперсионно-временная зависимость

Используя данные из **Таблицы 3**, строим дисперсионно-временной график, который показан на **Рисунке 7**.

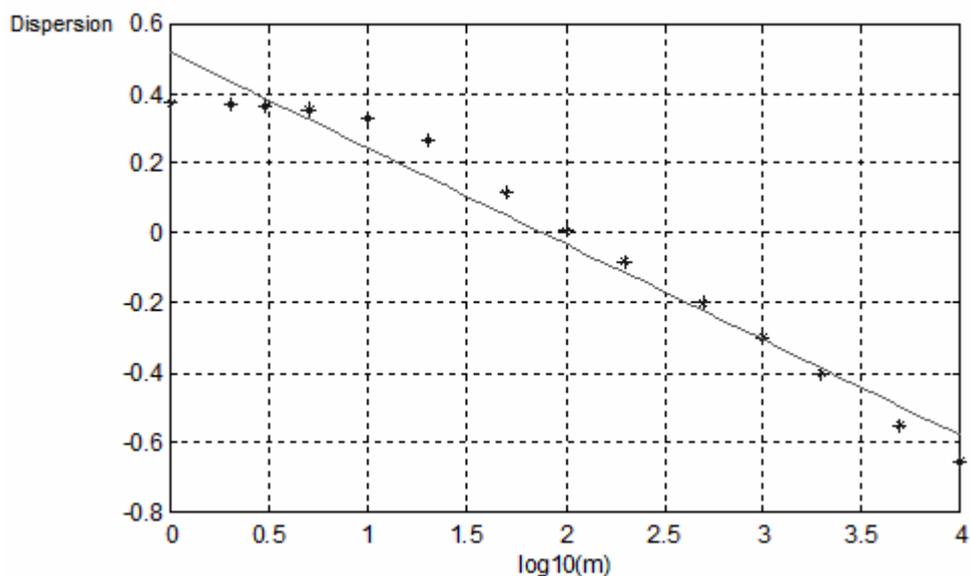


Рисунок 7. Дисперсионный график для сгенерированного трафика, $\beta=-0.2744$, $H=0.8628$

Угол наклона прямой составляет -0.2744 . Это означает что значение параметра Харста равно 0.8628 , что приблизительно соответствует значению параметра Харста, полученного в ходе измерения реального трафика сервера в локальной сети.

Это дает нам возможность использовать разработанную модель для исследования реальных сетей и сравнивать рабочие параметры потока в случаях, когда входной поток имеет самоподобный структуру или в случае Пуассоновского входного потока.

Библиография

[1] Leland W.E., Taqqu M.S., Willinger W. and Wilson D.V. On The Self-Similar Nature Of Ethernet Traffic.- Proc. ACM SIGCOMM'93, San-Fransisco, 1993, p 183-193.

[2] Walter Willinger, Murad S. Taqqu, Robert Sherman and Daniel V. Wilson, "Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level". IEEE/ACM Transactions on Networking, Vol. 5, No. 1, 1997

[3] Adrian Popescu, Traffic Self-Similarity, Department of Telecommunications and Signal Processing, University of Karlskrona/Ronneby

[4] Paxson V., Fast Approximation of Self-Similar Network Traffic, preprint Lawrence Berkely Laboratory and EECS Disvision, University of California, 1996

[5] P.Ulanovs, E.Petersons, Modeling of Self-similar traffic in high-performance Computer Networks, RTU, report collection for 42 International Conference, Riga, 2001

[6] Цыбаков Б.С. Модель телетрафика на основе самоподобного случайного процесса. "Радиотехника", 1999 г., № 5.

[7] Willinger W., Taqqu M.S., Erramilli A., A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for Modern High-Speed Networks, Stochastic Networks: Theory and Applications, Oxford University Press, 1996.

[8] Stalling W., High-Speed Networks TCP/IP and ATM Design Principles, Prentice-Hall Inc., 1998.

[9] Mandelbrot B.B., Long-Run Linearity, Locally Gaussian Processes, H-Spectra and Infinite Variances, International Economic Review, Vol.10, pp. 82-113, 1969.

[10] Willinger W., Taqqu M.S., Sherman R., Wilson D.V., Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level, IEEE/ACM Transactions on Networking, Vol. 5, No. 1, 1997.

Об авторе:

Ильницкий Сергей Васильевич, докторант Рижского Технического Университета, факультет „Транспортная Телематика“.

Служ. адрес: Латвия, Рига, Ломоносова 1, Рижский Технический Университет. Тел: (+371) 7089984,

Дом. адрес: Латвия, Олайне, LV-2114, ул. Земгалес, 34-48.

Моб. тел: (+371) 9441875

E-mail: Sergejs.Ilnickis@dati.lv