

УДК 621.3:004.732

А.В. Кораблєв,  
А.В. СеровГосударственное высшее учебное заведение „Националь-  
ный горный университет“, г. Днепропетровск, Украина,  
e-mail: artem.serov@gmail.com**СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОЦЕНКИ  
СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ТРАФИКА СЕТИ ETHERNET**A.V. Korablev,  
A.V. SerovState Higher Educational Institution “National Mining University”,  
Dnipropetrovsk, Ukraine, e-mail: artem.serov@gmail.com**COMPARATIVE ANALYSIS OF ETHERNET TRAFFIC STATISTICAL  
CHARACTERISTICS EVALUATION METHODS**

**Цель.** Установить характер статистического распределения размеров кадров самоподобного трафика локальной сети Ethernet и разработать рекомендации по применению методов оценки его вероятностных характеристик.

**Методика.** Представлена оригинальная методика исследования, первый этап которой заключается в выполнении процедуры агрегирования трафика с целью приведения исходных данных к шкале с постоянным шагом по оси времени. Для дальнейшего анализа распределения трафика и оценки его статистических характеристик процедура агрегирования проводится на временных интервалах различной длительности. На следующем шаге выполняется оценка параметра формы распределения размеров кадров исследуемого трафика тремя различными методами. На завершающем этапе исследования оценивается параметр Херста четырьмя различными способами.

**Результаты.** Представлены оценки параметра тяжести „хвоста“ распределения размеров кадров самоподобного трафика и значения параметра Херста.

**Научная новизна.** Заключается в анализе методов оценки характеристик самоподобного трафика в сетях с пакетной передачей, позволившем выбрать наиболее эффективный метод и сформулировать рекомендации по его применению.

**Практическая значимость.** Оценки полученных параметров позволяют не только сделать вывод о виде статистического распределения кадров трафика локальной сети Ethernet, но и дать рекомендации относительно выбора конкретного метода оценки.

**Ключевые слова:** самоподобный трафик, распределение с тяжелым хвостом, параметр Херста

**Введение.** В настоящее время, в связи с широким внедрением мультисервисных сетей, происходит объединение разнородных потоков информации, например таких, как передача речи, данных, сигналов аналогового и цифрового телевидения. Основой для организации таких систем является пакетная коммутация. Повышение эффективности работы сетей предполагает увеличение скорости передачи информации за счёт разработки высокопроизводительных алгоритмов обработки сетевого трафика, не прибегая к их аппаратной модернизации. В этой связи алгоритмы должны учитывать такие свойства трафика, как пачечность, самоподобие, приоритетность обслуживания и др.

**Постановка задачи. Цель работы.** В настоящее время доказано, что пакетный трафик обладает свойством самоподобия [1]. Для разработки высокоэффективных алгоритмов обработки сетевого трафика необходимо знать его статистические характеристики.

В связи с тем, что существуют различные методы оценки статистических характеристик, возникает задача выбора конкретного метода для конкретной сети передачи информации.

Целью настоящей работы является сравнительный анализ методов оценки статистических характе-

ристик трафика сети Ethernet и выработка рекомендаций для их применения.

**Основная часть.** В качестве экспериментальных данных в работе использовались результаты измерений трафика, полученные путём протоколирования процессов, происходящих в реальной Ethernet-сети [2]. Существует несколько версий формата кадров Ethernet, но поскольку анализ трафика проводился на основании стека протоколов TCP/IP, в данном случае формат кадра Ethernet имеет структуру, известную как EthernetII (Ethernet DIX).

Preamble	DA	SA	L	Data	FCS
8	6	6	2	46–1500	4

Рис. 1. Формат кадра Ethernet

В верхней строке рис.1 указаны названия полей кадра, а в нижней – их размеры в байтах.

В связи с тем, что размер кадров в сети Ethernet является переменным, выполним расчёт минимального размера кадра путём суммирования длин его полей (преамбула не учитывается).  $FRAME\_SIZE = length(DA) + length(SA) + length(L) + length(Data) + length(FCS) = 6 + 6 + 2 + 46 + 4 = 64$  (байт), где DA – адрес получателя, SA – адрес отправителя, L – размер кадра, Data – поле данных, FCS – контрольная сумма кадра.

Аналогічним образом вычисляется максимально возможный размер кадра, который равен 1518 байт. Таким образом, размер исследуемых кадров Ethernet находится в диапазоне от 64 до 1518 байт.

Экспериментальные данные представлены значениями моментов времени (в секундах) появления кадров в сети и их соответствующими размерами. Для дальнейшего анализа исходных данных выполним процедуру агрегирования, заключающуюся в ус-

реднении последовательных наблюдений на временных интервалах заданной длительности

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

где  $k$  – количество интервалов агрегирования,  $m$  – длительность интервала агрегирования;  $Y_k$  – номер отсчета в полученном ряде.

На рис. 2 приведены графики агрегирования на различных временных интервалах.

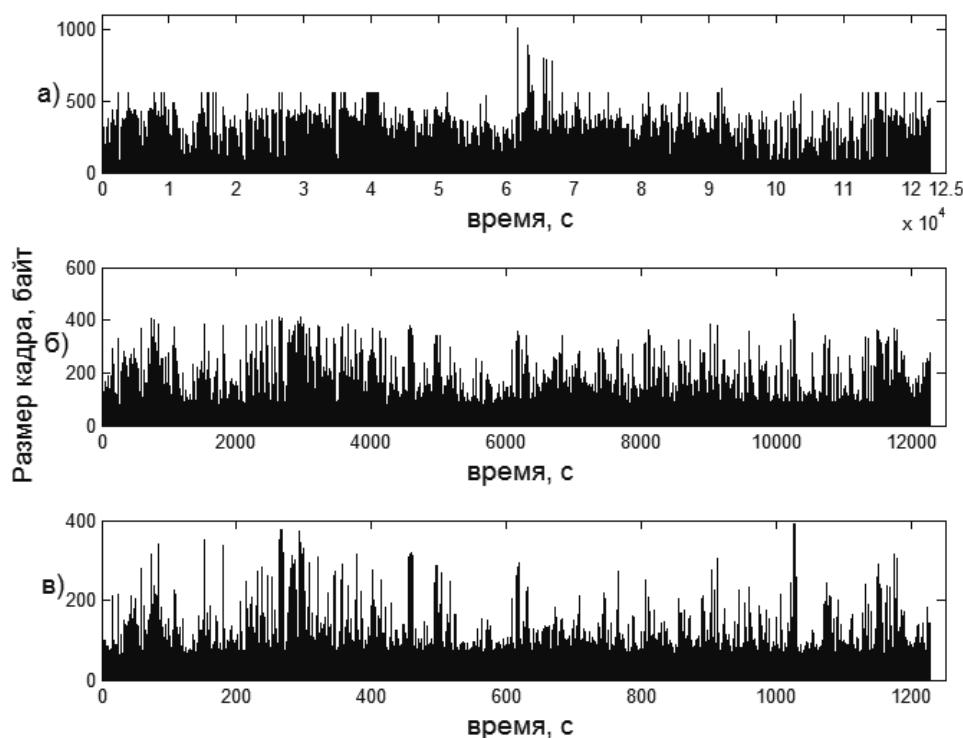


Рис. 2. Графики агрегирования исходных данных на временных интервалах: а) 1 с; б) 10 с; в) 100 с

Исходя из предположения о самоподобии трафика, проанализируем вид его распределения. Как известно, самоподобные процессы могут быть представлены распределениями с тяжелыми хвостами, которые имеют следующий вид [3]

$$P(X > x) \sim x^{-\alpha}; \quad 0 < \alpha < 2, x \rightarrow \infty.$$

Простейшим распределением с тяжелым хвостом является распределение Парето, для которого функция плотности распределения имеет вид  $f(x) = \frac{\alpha}{k} \left(\frac{k}{x}\right)^{\alpha+1}$ , а функция распределения –  $F(x) = 1 - \left(\frac{k}{x}\right)^{\alpha}, x > k, \alpha > 0$ .

Если параметр  $1 \leq \alpha \leq 2$ , то распределение имеет бесконечную дисперсию, а если  $0 < \alpha \leq 1$ , то – бесконечные дисперсию и математическое ожидание. Фактически, тяжелый хвост означает наличие бесконечной дисперсии, то есть случайная величина может принимать очень большие значения, но с очень малой вероятностью.

Существует целый ряд методов оценки тяжести хвоста, среди которых наиболее распространёнными являются следующие [4, 5]:

- метод Log-Log Complementary Distribution Plot (LLCD Plot);
- метод Хилла;
- метод Кровеллы.

Метод LLCDPlot заключается в построении графика комплементарной функции распределения  $\bar{F}(x) = 1 - F(x) = P[X > x]$  в логарифмическом масштабе. При этом график функции в области хвоста будет представлять собой прямую линию, такую, что для больших значений  $x$  справедливо выражение

$$\lim_{x \rightarrow \infty} \frac{d \ln \bar{F}(x)}{d \ln x} \sim -\alpha.$$

Присутствие на графике линейного участка свидетельствует о том, что данные имеют распределение с тяжелым хвостом. Тогда, построив линию регрессии к линейному участку графика и определив тан-

генс угла её наклона, становится возможным оценить значение параметра  $\alpha$  [5] (для реализации алгоритмов агрегирования, оценки статистических характеристик трафика и построения соответствующих графиков было разработано оригинальное программное обеспечение в среде MATLABR2011b).

На рис.3 приведен пример реализации метода LLCDDPlot для исследуемых данных с интервалом агрегирования 1 с. Угол наклона линии регрессии равен  $59^\circ$ , тангенс угла наклона – 1,7, следовательно, оценка параметра  $\alpha$  составляет 1,7.

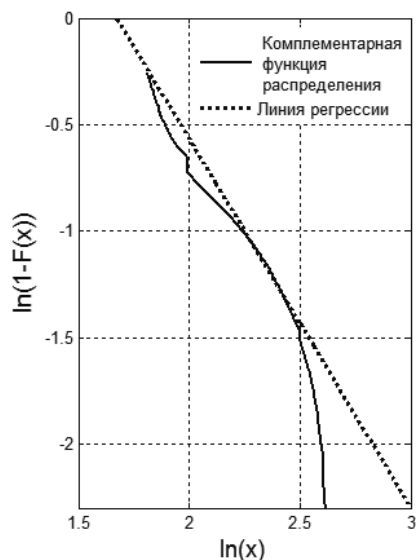


Рис. 3. Оценка параметра  $\alpha$  методом LLCDDPlot

Другим способом оценки показателя тяжести хвоста является использование метода Хилла, который позволяет оценить параметр  $\alpha$  на основании вычисления функции от  $k$  наибольших элементов набора данных

$$H_{k,n} = \left( \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i} - \log X_{n-k}) \right)^{-1},$$

где значения  $X_{(1)} \leq \dots \leq X_{(n)}$  определяют порядковые статистики ряда, т.е. элементы данных, расположенные по возрастанию. Оценку Хилла строят следующим образом: по оси абсцисс откладывают  $k$ , а по оси ординат – значение функции  $H_{k,n}$ ,  $1 \leq k \leq n$ . Если линия графика стабилизируется на уровне определенного значения, то именно такое значение и является оценкой параметра  $\alpha$ .

На рис.4 изображен график функции Хилла для ряда с интервалом агрегирования 1 с, по которому видно, что значение параметра  $\alpha$  приблизительно равно 1,8.

Несмотря на то, что методы Хилла и LLCDDPlot достаточно распространены для определения показателя тяжести хвоста, они имеют общий недостаток, заключающийся в необходимости визуального (а значит неточного и не всегда однозначного) определения точки, начиная с которой наблюдается линейный участок графика. Более того, если говорить об экспонен-

циальной функции, то определение точного начального значения становится еще более ответственной задачей, т.к. незначительное его увеличение существенно влияет на угол наклона линейного участка графика. При этом, чем более крутым является выбранный участок графика, тем больший существует риск пренебречь некоторой частью исходных данных, что может негативно отразиться на точности оценки.

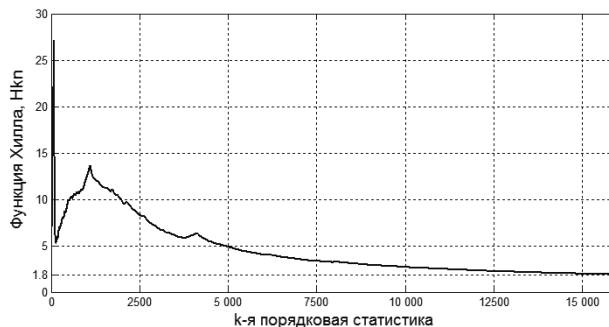


Рис. 4. Оценка параметра  $\alpha$  методом Хилла

Проведем дополнительную оценку методом Кровеллы, который устраняет вышеуказанный недостаток.

На рис. 5 приведен график семейства комплементарных функций распределения для исходных данных с разными степенями агрегирования, построенный по методу Кровеллы [4].

Нижняя кривая на графике соответствует исходным данным, а последующие – данным со степенью агрегирования  $m = 2^i$ , где  $i = 1..9$ . В связи с тем, что алгоритм расчета параметра  $\alpha$  по данному методу достаточно сложный, в настоящей работе приводятся только окончательные результаты в графической и численной форме (табл. 1).

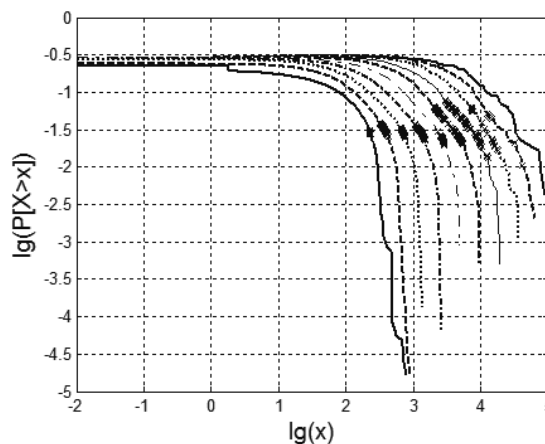


Рис. 5. Оценка параметра  $\alpha$  методом Кровеллы для данных с интервалом агрегирования 1 с

Оценки, полученные по вышеприведенным методам для трех агрегированных рядов с интервалами агрегирования 1, 10 и 100 секунд, представлены в табл. 1.

Таблиця 1

Оценка параметра  $\alpha$

Интервал агрегирования, с	Метод Кровеллы	Метод LLCDD Plot	Метод Хилла
1	1,18	1,7	1,8
10	1,29	1,9	2
100	1,69	2	2

Важным параметром, характеризующим степень самоподобия случайного процесса, является параметр Херста. Основываясь на результатах анализа трафика, при наличии хвоста распределения, найдем теоретическое значение параметра Херста для всех агрегированных рядов как  $H = \frac{3-\alpha}{2}$  [3]. Полученные показатели сравним с оценками, найденными с помощью известных методов определения параметра Херста, большинство из которых основано на нахождении углового коэффициента линии регрессии для различных вариантов агрегирования выборок данных [5]. Полученные результаты приведены в табл. 2, 3.

Таблиця 2

Теоретическое значение параметра Херста

Интервал агрегирования, с	Параметр Херста $H$		
	Метод Кровеллы	Метод LLCDD Plot	Метод Хилла
1	0,91	0,65	0,6
10	0,855	0,55	0,5
100	0,655	0,5	0,5

Таблиця 3

Оценка параметра Херста различными методами

Метод оценки параметра Херста	Интервал агрегирования, с		
	1	10	100
R/S-статистика	0,712	0,598	0,468
Метод агрегированных дисперсий	0,811	0,71	0,778
Метод разностей агрегированных дисперсий	0,978	0,908	0,902
Метод периодограмм	0,976	0,959	0,661

На рис. 6, 7 и 8 приведены графики оценки параметра Херста различными методами для данных с интервалом агрегирования 1 с.

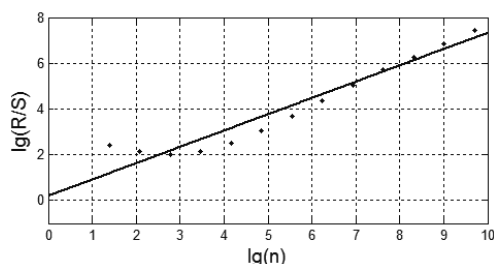


Рис. 6. Оценка параметра Херста методом R/S-статистик

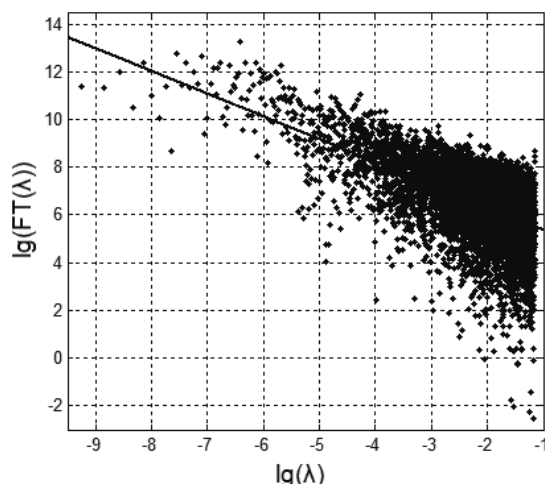


Рис. 7. Оценка параметра Херста методом периодограмм

Метод R/S-статистик основан на масштабных свойствах отсчетов временного ряда и определяется соотношением [6]

$$(R/S)_n = Cn^H,$$

где  $n$  – размер выборки;  $H$  – параметр Херста;  $R$  – размах выборки, т.е. разность между максимальной и минимальной суммой приращений, которые берутся с заданным шагом, зависящим от  $n$ ;  $S$  – эмпирическая дисперсия;  $C$  – некоторая константа.

Метод периодограмм заключается в построении графика в логарифмическом масштабе, где по горизонтальной оси откладываются значения частот, а на вертикальной – значения спектральной плотности, которая вычисляется по следующей формуле [7]

$$FT(\lambda) = \frac{1}{2\pi n} \left| \sum_{k=1}^n X_k e^{ik\lambda} \right|^2,$$

где  $\lambda$  – система частот.

В методе агрегированных дисперсий вычисляются значения выборочной дисперсии  $S_{(m)}^2$  в агрегированных блоках размера  $m$  из выборки  $n$  рассматриваемых значений [7]

$$S_{(m)}^2 = \frac{1}{(n/m) - 1} \sum_{k=1}^{n/m} (X_{(m)}(k) - \bar{X})^2,$$

где  $X_{(m)}$  – блок из  $m$  значений, взятый из выборки  $X = \{X_1, X_2, \dots, X_n\}$ . Найденные значения откладываются в логарифмическом масштабе по оси ординат напротив соответствующих значений  $\ln m$ . Тогда параметр Херста вычисляется по формуле  $H = 1 + \frac{\beta}{2}$ , где  $\beta$  – угловой коэффициент линии регрессии.

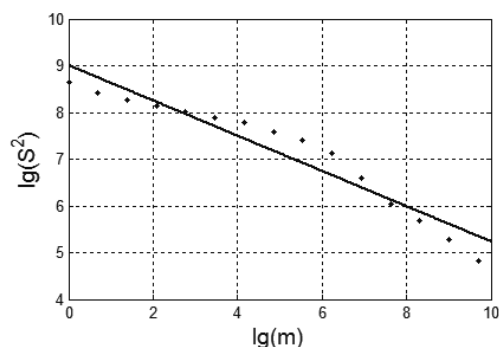


Рис. 8. Оценка параметра Херста методом агрегированных дисперсий

**Выводы.** Установлено, что при оценке тяжести хвоста распределения трафика сети Ethernet наиболее эффективным является метод Кровеллы, поскольку метод LLCPlot даёт граничное значение  $\alpha$  при временном интервале агрегирования 100 с, а метод Хилла – уже при 10 с (табл. 1).

При непосредственной оценке параметра Херста, на основании экспериментальных данных о трафике, наиболее эффективным является метод разностей агрегированных дисперсий, поскольку разностный алгоритм менее чувствителен к ошибкам измерений и вычислений (табл. 3).

#### Список литературы / References

1. Leland, W., Taqqu, M., Willinger, W. and Wilson, D. (1994), "On the self-similar nature of Ethernet traffic (extended version)", *IEEE/ACM Transactions on networking*, vol.2, no.1. pp. 1–15.
2. "The Internet Traffic Archive", available at: <http://ita.ee.lbl.gov/html/contrib/BC.html>
3. Adler, R.A., Feldman, R. and Taqqu, M. (1998), *Practical guide to heavy tails: statistical techniques and applications*, Birkhauser, Boston.
4. Crovella, M. and Taqqu, M. (1999), "Estimating the heavy-tail index from scaling properties", *Methodology and computing in applied probability*, vol.1, no.1, pp. 55–79.
5. Grimm, C. and Schluchtermann, G. (2008), *IP Traffic Theory and Performance*, Springer-Verlag Berlin and Heidelberg, Berlin.
6. Востров Г.Н. Сегментация и анализ временных рядов на основе стохастической фрактальной модели / Г.Н. Востров, М.В. Полякова, В.В. Любченко // Тр. Одес. политехн. ун-та. – Одесса, 2001. – Вып. 1. – С. 109–114. – Библиогр.: с.114.  
Vostrov, G.N., Polyakova, M.V. and Lyubchenko, V.V. (2001), "Segmentation and analysis of time series based on stochastic fractal model", *Proc. Odes. polytekh. un-ta*, Odessa, Issue 1. pp. 109–114.
7. Гуда А.Н. Модели оценки параметров телекоммуникационного трафика в автоматизированных информационно-управляющих системах / А.Н. Гуда, М.А. Бутакова, Н.А. Москат // Вопросы современной науки и практики. Ун-т им. В.И. Вернадского. – 2010. – №4–6(29). – С. 71–87. – Библиогр.: с.87.  
Guda, A.N., Butakova, M.A. and Moskat, N.A. (2010), "Telecommunication traffic parameters estimation models in automated information-management systems", *Voprosy sovremennoy nauki i praktiki*, Published by V.I. Vernadsky University, no.4–6(29), pp. 71–87.

tion models in automated information-management systems", *Voprosy sovremennoy nauki i praktiki*, Published by V.I. Vernadsky University, no.4–6(29), pp. 71–87.

**Мета.** Встановити характер статистичного розподілу розмірів кадрів самоподібного трафіку локальної мережі Ethernet і розробити рекомендації із застосування методів оцінки його ймовірнісних характеристик.

**Методика.** Представлено оригінальну методику дослідження, перший етап якої полягає у виконанні процедури агрегування трафіку з метою приведення вихідних даних до шкали з постійним кроком по осі часу. Для подальшого аналізу розподілу трафіку та оцінки його статистичних характеристик процедура агрегування проводиться на часових інтервалах різної тривалості. На наступному кроці виконується оцінка параметра форми розподілу розмірів кадрів досліджуваного трафіку трьома різними методами. На завершальному етапі дослідження оцінюється параметр Херста чотирма різними способами.

**Результати.** Представлено оцінки параметра тяжкості „хвоста“ розподілу розмірів кадрів самоподібного трафіку і значення параметра Херста.

**Наукова новизна.** Полягає в аналізі методів оцінки характеристик самоподібного трафіку в мережах з пакетною передачею, що дозволив обрати найбільш ефективний метод і сформулювати рекомендації щодо його застосування.

**Практична значущість.** Оцінки отриманих параметрів дозволяють не тільки зробити висновок про вид статистичного розподілу кадрів трафіку локальної мережі Ethernet, але й надати рекомендації щодо вибору конкретного методу оцінки.

**Ключові слова:** само подібний трафік, розподіл з важким хвостом, параметр Херста

**Purpose.** To find out Ethernet self-similar traffic frame size statistical distribution and give some recommendations concerning use of different methods of traffic probability characteristics estimation.

**Methodology.** An authentic research methodology has been presented. Aggregation procedure of initial data unification according to time scale constant step was put in its base. Traffic distribution further analysis and its characteristics estimation with different time interval lengths was carried out. Three ways of frame size distribution shape parameter evaluation were tested. Four approaches were applied to Hurst parameter estimation.

**Findings.** Tail index and Hurst parameter estimation results are presented.

**Originality** consists in self-similar traffic characteristics estimation methods analysis that allowed choosing the most effective method and giving recommendations concerning its usage.

**Practical value.** Recommendations on definite estimation method choice are given.

**Keywords:** self-similar traffic, heavy-tail distribution, Hurst parameter

Рекомендовано до публікації докт. техн. наук В.І. Корнієнком. Дата надходження рукопису 20.09.11