

УДК 621.396.67

## ФРАКТАЛЬНЫЕ СВОЙСТВА СЕТЕВОГО ТРАФИКА В КЛИЕНТ-СЕРВЕРНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЕ

*Треногин Н.Г., Соколов Д.Е.*

Сибирский государственный университет телекоммуникаций и информатики (СибГУТИ)

Последнее десятилетие ознаменовалось существенным достижением в области теории телетрафика – открытием самоподобия, или фрактальных свойств процессов, протекающих в сетях передачи данных. Появились исследования, в которых выявлены фрактальные свойства трафика, порождаемого самыми различными приложениями – видеоконференцсвязь, файл-серверные приложения, доступ к Web-ресурсам и т.д. В то же время при оптимизации вероятностно-временных характеристик клиент-серверных информационных систем (ИС) чаще всего исходят из предположения об экспоненциальности потоков событий на всех уровнях, в том числе в сети.

Проведенные исследования показывают наличие в сетевом трафике реальных SQL-серверных приложений всех важнейших свойств самоподобия, оказывающих существенное влияние на вероятностно-временные характеристики системы.

### Понятие и свойства самоподобного трафика

Неформально самоподобный (фрактальный) процесс можно определить как случайный процесс, статистические характеристики которого проявляют свойства масштабирования. Самоподобный процесс существенно не меняет вида при рассмотрении в различных масштабах по шкале времени. В частности, в отличие от процессов, не обладающих фрактальными свойствами, не происходит быстрого «сглаживания» процесса при усреднении по шкале времени – процесс сохраняет склонность к всплескам.

Пусть  $\{X_k; k=0;1;2;\dots\}$  – стационарный случайный процесс. Учитывая стационарность и предположение о существовании и конечности двух первых моментов, введем обозначения:

$m = E[X_t]$  - среднее значение, или математическое ожидание;

$\sigma^2 = E[X_t - m]^2$  - дисперсия;

$R(k) = E[(X_{t+k} - m)(X_t - m)]$  - корреляционная функция,

$r(k) = R(k) / R(0) = R(k) / \sigma^2$  - коэффициент корреляции.

Под усреднением по шкале времени будем понимать переход к процессу  $\{X^{(m)}\}$ , такому, что

$$X_k^{(m)} = \frac{1}{m} \sum_{i=km-m+1}^{km} X_i \quad (1)$$

При моделировании сетевого трафика значения  $X_k$  интерпретируются как число пакетов (реже – как суммарный объем данных в байтах), поступивших в канал или сеть в течение  $k$ -го интервала времени. Исходный процесс при этом уже является усредненным. В некоторых случаях, когда есть необходимость избежать такого начального усреднения, рассматривается точечный процесс, или поток событий, т.е. последовательность моментов поступления единичных пакетов в сеть.

Случайный процесс  $X(t)$  является самоподобным с параметром Херста  $H$ , если  $X(t)$  и  $a^{-H}Z(at)$  имеют идентичные конечномерные распределения вероятностей для всех  $a > 0$ . Отметим, что на практике обычно встречаются не строго самоподобные, а асимптотически самоподобные процессы. Исчерпывающий обзор подходов к определению самоподобия приведен в [2].

Параметр Херста  $H \in (0.5; 1)$  определяет степень самоподобия процесса. Чем ближе этот параметр к единице, тем более ярко проявляется фрактальные свойства. Напротив, равенство  $H=0.5$  свидетельствует об отсутствии самоподобия.

Самоподобные процессы, в том числе описывающие явления в сетях передачи данных, обладают рядом свойств, существенно отличающих их от потоков, рассматриваемых в классической теории телетрафика.

### 1. Долговременная зависимость.

Самоподобные процессы обладают гиперболически затухающим коэффициентом корреляции вида

$$r(k) = \frac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}), \quad (2)$$

или, для асимптотически самоподобных процессов, корреляционная функция

$$R(k) \approx k^{2H-2}L(k) \text{ при } k \rightarrow \infty \quad (3)$$

где  $L(t)$  – медленно меняющаяся функция на бесконечности (т.е.  $\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1$  для всех  $x > 0$ ). Следовательно, корреляционная функция является несуммируемой – ряд, образованный последовательными значениями корреляционной функции, расходится. Это свойство характеризует практически все самоподобные процессы и отличает их от процессов без долговременной зависимости, у которых корреляционная функция убывает по показательному закону и суммируема.

Долговременная зависимость является причиной ярко выраженных пульсаций процесса, однако позволяет говорить о некоторой предсказуемости в небольших пределах времени. С точки зрения теории очередей, важным следствием коррелированности потока является неприемлемость оценок параметров очереди, основанных на предположении об одинаковом и независимом распределении интервалов во входящем потоке.

## 2. Медленно убывающая дисперсия.

При усреднении процесса дисперсия выборочного среднего затухает медленнее, чем величина, обратная размеру выборки, по закону:

$$\sigma^2(X^{(m)}) \propto m^{2H-2}, \text{ при } m \rightarrow \infty \quad (4)$$

в то время как для традиционных стационарных случайных процессов

$$\sigma^2(X^{(m)}) = \frac{1}{m} \sigma^2(X), \quad (5)$$

т.е. уменьшается обратно пропорционально объему выборки.

Свойство медленно убывающей дисперсии говорит о возможности существенных, не сглаживаемых усреднением, «выбросов» в случайном процессе, и связывает самоподобие с таким понятием, как распределения с весомами хвостами. Важное следствие свойства медленно затухающей дисперсии состоит в том, что в случае классических статистических тестов (например, вычисление доверительных интервалов) общепринятая мера среднеквадратического отклонения  $\sigma$  является ошибочной.

С данным свойством связано и «нехарактерное» поведение индекса дисперсии, или индекса разброса, для отсчетов процесса (IDC), так же называемого фактором Фано. IDC определяется как отношение дисперсии числа событий на заданном временном интервале  $T$  к математическому ожиданию этой величины:

$$F(T) = \frac{\text{Var}[N(T)]}{E[N(T)]} \quad (6)$$

Здесь  $N(T)$  – число событий исследуемого потока, наступивших в интервале (окне)  $T$ .

Для самоподобных процессов логарифм индекса разброса  $F(T)$  линейно возрастает:

$$\ln[F(T)-1] = (2H-1) \ln T + y. \quad (7)$$

## Распределения с весомами хвостами

Случайная величина  $Z$  имеет распределение с весомым хвостом (PBX), если вероятность  $P[Z > x] \sim cx^{-\alpha}$  при  $x \rightarrow \infty$ , т.е. хвост распределения затухает по степенному закону. Примером распределения с весомым

хвостом является распределение Парето. При  $0 < \alpha < 2$  величина  $Z$  обладает бесконечной дисперсией, а при  $0 < \alpha < 1$  среднее значение также бесконечно.

Наиболее существенной особенностью случайной величины, обладающей распределением с весомым хвостом, является чрезвычайная изменчивость. С вероятностью, которая не является пренебрежимо малой, в выборке могут присутствовать некоторое число «очень больших» значений. Такие распределения существенно снижают точность статистических оценок; скажем, конечный объем выборки приводит к заниженной оценке среднего и дисперсии.

Наличие РВХ во внешних по отношению к рассматриваемым процессам явлениях является одной из причин возникновения самоподобия в соответствующих стохастических моделях.

Часто при рассмотрении самоподобных процессов говорят о комплексе взаимосвязанных понятий: самоподобии, масштабировании, долговременной зависимости, РВХ и степенных законах статистических характеристик. Этот комплекс свойств отличает процессы, называемые самоподобными, от классических случайных процессов, например, пуассоновского.

### **Исследуемый трафик**

Многочисленные измерения показывают наличие существенно самоподобных свойств трафика в клиент-серверных ИС различной архитектуры – от «классических» двухзвенных, до многоуровневых с Web-доступом и терминальных [8, 9].

Фрактальный характер может быть рассмотрен на примере среза трафика (рис.1), полученного при работе удаленного подразделения из нескольких рабочих мест с сервером СУБД. Параметры среза трафика приведены в таблице 1.

Данные были получены путем перехвата кадров на FastEthernet-интерфейсе сервера СУБД с помощью программы tcpdump. Исходя из предположения о дуплексности канала, к рассмотрению принят трафик одного направления – исходящий по отношению к рабочим станциям. Сервером СУБД является сервер Oracle<sup>®</sup>i; приложение (биллинговая система оператора связи) реализовано по классической двухзвенной клиент-серверной схеме, т.е. сетевое взаимодействие происходит на базе TNS / SQL\*net поверх TCP. Поскольку пропускная способность рассматриваемого канала (100 Мбит/с FastEthernet) существенно превышает суммарный трафик, а сторонняя нагрузка в день исследований пренебрежимо мала, срез можно считать «свободным трафиком», в терминологии, введенной И.Норросом [5] – т.е., трафиком, полностью определяемым своим источником и не испытывающим влияния сети.

Рисунок 1 наглядно демонстрирует сильно пульсирующий характер трафика с существенной дисперсией, наличием резких всплесков, группировкой в «пачки».

Таблица 1.

Дата	26.11.2003
Выходной файл tcpdump	s24-26nov.td
Длительность, с	21 936
Длительность, ч	6.093
Число пакетов	688108
Интенсивность $\lambda$ , $c^{-1}$	31.368
Средний объем пакета, байт	193.2
Рабочих мест	25
Параметр Херста $H$ (IDC)	0.729
Параметр Херста $H$ (автокорреляция)	0.724

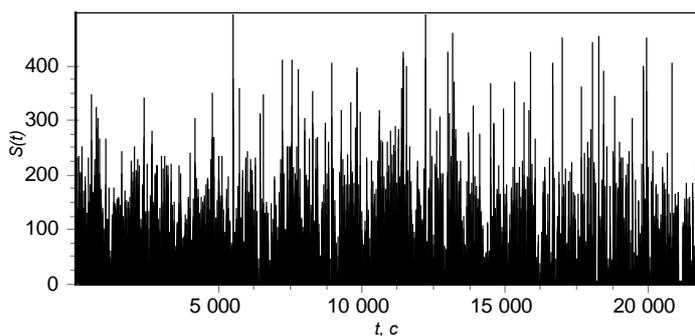


Рис. 1. Интенсивность трафика, пакетов в секунду.

### Корреляционная структура

Графическое представление коэффициента корреляции позволяет визуально убедиться в том, что исследуемый трафик обладает долгосрочной зависимостью.

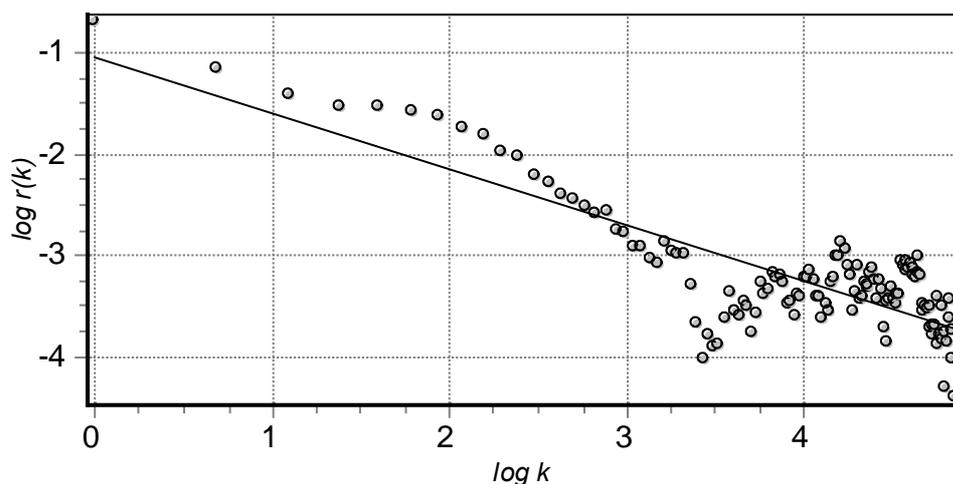


Рис.2 - Значения коэффициента корреляции (логарифмический масштаб)

На рис. 3. приведен график коэффициента корреляции для процесса, соответствующего исследуемому срезу трафика, а на рис. 2 – те же значения в логарифмическом масштабе. Очевидно, что точки на рис. 2 в целом группируются вокруг прямой, угловой коэффициент которой может быть определен путем линейной регрессии. Если процесс самоподобен, то в соответствии с (3) угловой коэффициент  $\beta = 2 - 2H$ . При полученном значении  $\beta = 0,55004$  параметр Херста оказывается равным 0.724.

На рис. 3а для сравнения приведена кривая, соответствующая значениям коэффициента корреляции строго самоподобного процесса (2). Очевидно, что выборочные оценки достаточно точно соответствуют «идеальной» кривой, особенно при увеличении аргумента  $k$ .

Существенную роль долговременной зависимости в исходном процессе можно выявить и на основе анализа так называемого «перемешанного» процесса. Такой процесс получается из исходного путем перестановки интервалов между поступлениями пакетов в случайном порядке. В таком процессе (рис. 3б) корреляция убывает существенно быстрее, достаточно быстро стремясь к нулю.

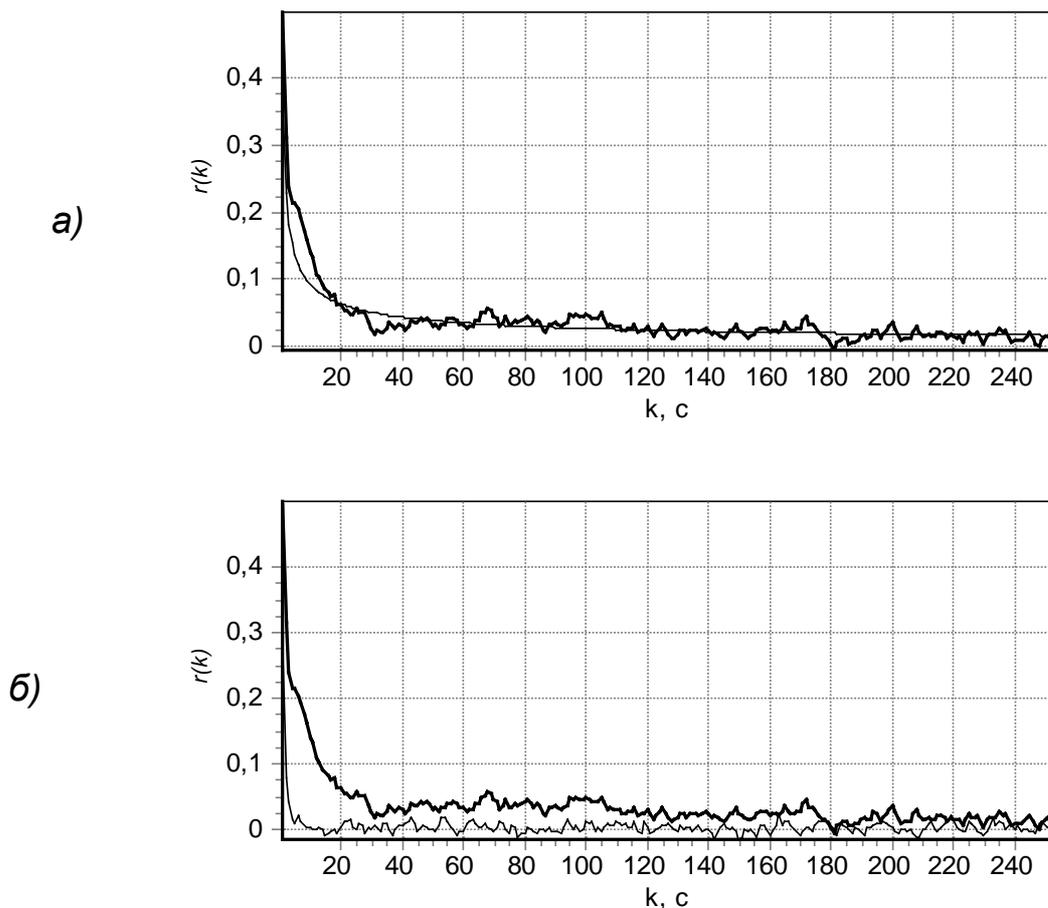


Рис. 3. - График коэффициента корреляции для трафика: а) в сравнении с кривой коэффициента корреляции строго самоподобного процесса (1) при  $\beta = 0.55$ ; б) в сравнении с кривой коэффициента корреляции с «перемешанными» случайным образом интервалами.

### Параметр Херста

Традиционно самоподобие в стохастическом процессе выявляется путем определения параметра Херста  $H$ . Тот факт, что  $0.5 < H < 1$ , т.е. значение параметра Херста отлично от 0.5, считается достаточным основанием для признания процесса самоподобным (по крайней мере, асимптотически). Следует отметить, что значение  $H$ , близкое к единице, может означать, что процесс является детерминированным, т.е. не случайным: для ряда строго детерминированных процессов структура строго повторяется на любом масштабе, что приводит к единичному значению параметра Херста.

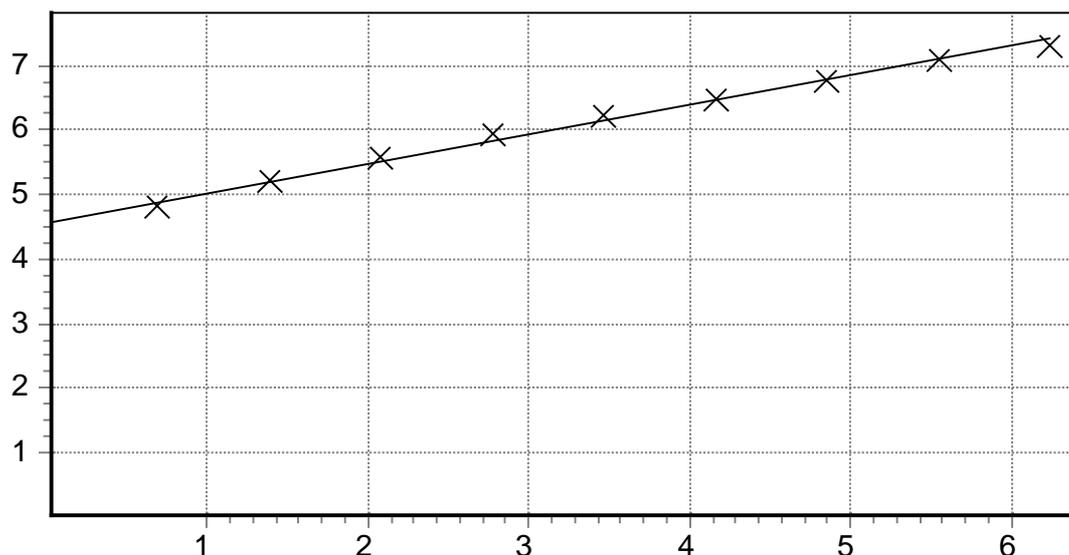


Рис. 4.- Определение параметра Херста  $H$  по значениям IDC, или фактора Фано

В рассматриваемом случае значения параметра Херста, определенные из вида кривой коэффициента корреляции и путем анализа IDC (0.724 и 0.729 соответственно) практически совпадают. При этом значение говорит о существенно выраженных фрактальных свойствах.

На рис. 4 приведены значения  $\ln[F(T)-I]$  в зависимости от  $\ln T$ . В соответствии с (7), точки образуют прямую, позволяющую однозначно определить параметр Херста. Прямая, полученная путем линейной регрессии, имеет угловой коэффициент 0.459, что соответствует  $H=0.729$ .

Для процесса с «перемешанными» случайным образом, значение получается существенно меньше:  $H=0.563$ . Этот факт, совместно с видом корреляционной структуры исходного и «перемешанного» процессов, позволяет утверждать, что для рассматриваемого трафика самоподобие заключено не столько в «тяжелом» распределении интервалов, сколько в долгосрочной зависимости – группировке коротких интервалов в пачки.

### Моделирование очереди

Самоподобная нагрузка в системах массового обслуживания (СМО) приводит к существенно большим задержкам при том же значении загрузки, чем «классическая» нагрузка [2, 7, 8, 9], за счет сильной группированности заявок и значительной дисперсии.

Имитационное моделирование СМО с неограниченным буфером и входным потоком, который определяется рассматриваемым трафиком, показывает характерное для фрактального трафика поведение очереди (рис. 4). При этом, как показывают результаты моделирования (рис. 4б), на возрастание очереди и времени задержки оказывает влияние не столько распределение интервалов, сколько корреляционная структура процесса, т.е.

долговременная зависимость. «Перемешанный» в случайном порядке трафик, при сохранении распределений интервалов между поступлениями и объемов пакетов, демонстрирует существенно менее резкое нарастание очереди.

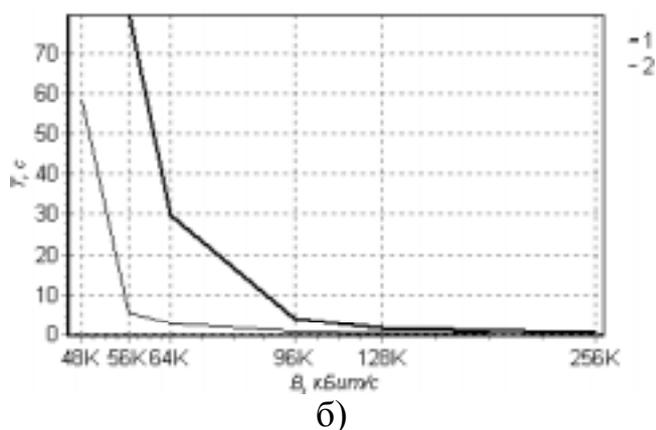
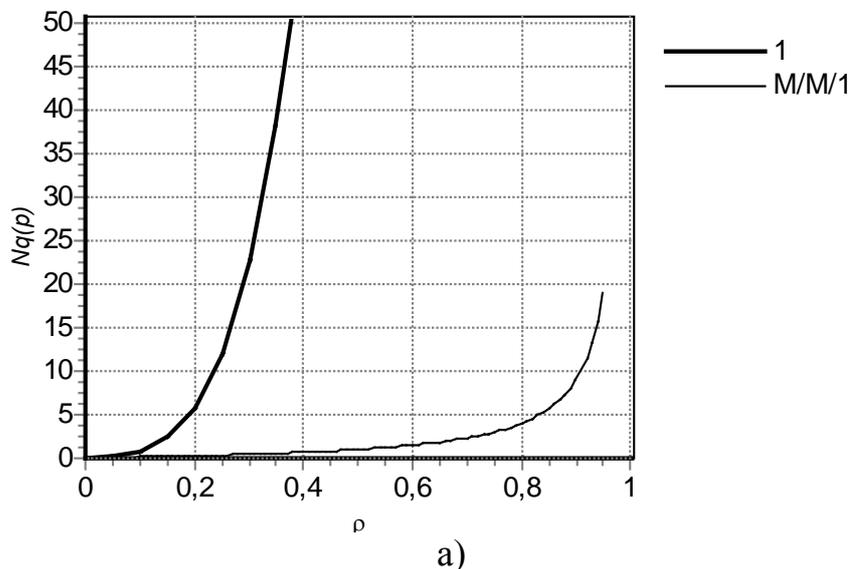


Рис.4.- Результаты имитационного моделирования для СМО с неограниченным буфером и исследуемым трафиком в качестве входного процесса: а) средняя длина очереди в зависимости от загрузки СМО; б) среднее время пребывания в СМО для исходного (1) и «перемешанного» (2) трафика.

В заключение следует отметить, что для полноценного описания трафика в клиент серверных ИС требуется не только анализ статистических характеристик, но и выбор адекватной математической модели – самоподобного процесса. В качестве такой модели могут выступать фрактальный дробовой процесс (FSNDP) [6, 8, 9] или  $\alpha$ -устойчивые фрактальные процессы.

## Литература

1. Leland W.E., Taqqu M.S., Willinger W. and Wilson D.W. On the self-similar nature of Ethernet traffic // *Computer Communications Review*, 1993:23, p.193-193.
2. Шелухин О.И., Тенякшев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях. Монография / Под ред. О.И.Шелухина - М.: Радиотехника, 2003.- 480 с.
3. Цыбаков Б.С. Модель телетрафика на основе самоподобного случайного процесса. // «Радиотехника», 1999, № 5, с. 24–31.
4. Нейман В.И. Новое направление в теории телетрафика. // «Электросвязь», 1998, № 7, с. 27–30.
5. Norros I. The Management of Large Flows of Connectionless Traffic on the Basis of Self-Similar Modeling // ICC '95, IEEE International Conference on Communications. – Seattle, 1995.
6. Ryu B.K. Fractal Network Traffic: From Understanding to Implications. Ph.D. thesis. - Columbia University, 1996. – 143 p.
7. М.Н.Петров, Д.Ю. Пономарев. Самоподобие в системах массового обслуживания с ограниченным буфером. // "Электросвязь", 2002, №2, с. 35-39
8. Соколов Д.Е., Треногин Н.Г. Характер сетевого трафика на клиентском участке распределенной клиент-серверной системы. // *Материалы Международной научно-технической конференции «Информатика и проблемы телекоммуникаций»*. – Новосибирск: СибГУТИ, 2001, с.34–35.
9. Соколов Д.Е. Моделирование нагрузки в клиент-серверных системах на основе фрактальных процессов. // *Материалы межвузовской научно-технической конференции «Управляющие и вычислительные системы. Новые технологии»*. – Вологда, 2001, с.59 - 60.
10. Терехов В.И., Соколов Д.Е., Треногин Н.Г. Мониторинг и анализ трафика в IP-сетях: комплексный подход. // *Материалы Международной научно-технической конференции «Информатика и проблемы телекоммуникаций»*. – Новосибирск: СибГУТИ, 2001, с.36.