

Анализ сетевого трафика корпоративной сети университета методами нелинейной динамики

08, август 2013

DOI: 10.7463/0813.0587054

Басараб М. А., Колесников А. В., Иванов И. П.

УДК 004.724: 004.728

Россия, МГТУ им. Н.Э. Баумана

avkolesnikov90@list.ru

ivanov@bmstu.ru

bmic@mail.ru

Введение

В процессе обработки и хранения информации неизбежно возникает необходимость в обмене данными между участниками этого процесса. С конца 70-х годов началось бурное развитие компьютерных сетей и сопутствующего сетевого оборудования. Локальные и глобальные сети продолжают развиваться, возникают новые протоколы передачи данных, расширяются аппаратные возможности сетевого оборудования, растет число подключенных абонентов и суммарный объем трафика.

Такое интенсивное развитие области влечет за собой ряд проблем. Одна из них заключается в том, что при растущем числе потребителей информационных услуг возрастают требования к сетевому и серверному оборудованию, необходимому для поддержания надлежащего уровня качества обслуживания. Развитие сетевого оборудования и транспортных протоколов должно опираться на адекватные математические модели параметров трафика и инструменты моделирования сетевых процессов [1, 2]. Характер сетевого трафика определяется рядом факторов – от поведения пользователей либо прикладного программного обеспечения, до протоколов передачи и используемого оборудования. Очевидно, что макропараметры сетевого трафика (на относительно больших временных интервалах) определяются человеком. Однако характер трафика на интервалах порядка микросекунд определяется большей частью транспортными протоколами, сетевым оборудованием и серверным программным

обеспечением. Таким образом, исследование основных характеристик сервера, таких как выделение оперативной памяти, загрузка центрального процессора, состояние процессов операционной системы на фоне интенсивного сетевого трафика, является актуальной задачей [3, 4].

Целью работы является рассмотрение различных моделей сетевого трафика и анализ наиболее перспективной модели, учитывающим самоподобные свойства трафика как временного ряда [5, 6]. На примере корпоративной сети университета рассчитаны динамические характеристики входящего и исходящего трафика, а также распределения аппаратных мощностей сервера. Для сетевого трафика найдены экспонента Ляпунова и показатель Хёрста, характеризующие его хаотические и фрактальные свойства. Новизна работы заключена в том, что фазовый анализ временных рядов впервые дал возможность выявить аттракторы, позволяющие провести более глубокое исследование влияния нагрузки на пропускную способность сети. Для всех процессов было установлено самоподобие, что подтверждает возможность применения фрактальных моделей для работы с данными, в частности, для прогнозирования поведения временных рядов.

1 Обзор моделей сетевого трафика

Ранние стохастические модели. Стохастические модели трафика, широко использовавшиеся в прошлом [7], в основном представляли марковские процессы, то есть обладали кратковременной зависимостью. Такие модели описывались распределением Пуассона с длиной сообщения, изменявшейся по экспоненциальному закону, и основывались на теории массового обслуживания. Эти модели сформировались в период ранних сетей ARPANET. Результаты моделирования трафика на основе теории массового обслуживания соответствовали распределению времени обработки звонка в телефонных сетях.

Однако, позже стали все чаще возникать работы, в которых указывалось о растущей тенденции межсетевого обмена данными и объема суммарного трафика. Также на характер трафика стали оказывать все большее влияние новые протоколы передачи данных в сети. На основе подобных исследований в 1986 году была разработана концепция «цепочки пакетов» (packet train). В рамках модели считается, что пакеты в сети передаются вместе, в то время как в пуассоновских моделях каждый пакет обрабатывался отдельно [8].

Не так давно также стала популярной модель, в рамках которой объем трафика обладает долговременной зависимостью. Также было установлено, что трафик самоподобен и характеризуется распределением с тяжелым хвостом [9].

Классическая модель. В рамках классической модели трафика считается, что источники данных работают перемененно [10]. То есть периоды высокой активности сменяются длительными задержками. Таким образом, было заключено, что время поступления сообщений и длина сообщений подчиняются экспоненциальному распределению, а процесс поступления сообщений от источников данных - пуассоновский процесс. Все процессы стационарные и независимые.

Пуассоновская модель не учитывает, что реальный сетевой трафик обладает периодами сильных всплесков активности. Для классической модели автокорреляционная функция стремится к нулю для больших отсчетов, в то время как наличие всплесков активности в реальном исследуемом трафике приводит к положительной автокорреляции.

Модель цепочки сообщений. Модель трафика была сформулирована и стала популярной в 80-е гг. XX в. [11]. В рамках модели подразумевается, что пакеты трафика передаются вместе и могут обрабатываться как одно целое. Сетевое оборудование в каждой точке сети может принимать решение о дальнейшей обработке цепочки по первому сообщению. Подобный алгоритм предохранил бы сеть от бесполезных операций по анализу кадров. Однако стоит отметить, что это модель источников сообщений. Модель применима только для сообщений, обладающим одним пунктом назначения. Очевидно, что реализация транспортных протоколов и сетевого оборудования для модели цепочки сообщений и классической модели будет кардинально отличаться.

Самоподобная модель. Во многих современных работах отмечается [9, 12], что объединение трафика от нескольких переменных источников приводит к тому, что трафик становится сильно автокоррелированным с долговременной зависимостью [13]. Это приводит к тому, что устойчивость корреляционных структур не исчезает даже для больших значений лага. Другими словами, совокупность множества источников данных, проявляющих синдром бесконечной дисперсии, в результате дает самоподобный объединенный сетевой трафик, стремящийся к фрактальному броуновскому движению. Кроме того, исследование различных источников трафика показывает, что сильно изменчивое поведение – это свойство, присущее архитектуре клиент/сервер.

Не существует единого причинного фактора, вызывающего самоподобность. Различные корреляции, существующие в самоподобном сетевом трафике, которые воздействуют на различных временных масштабах, могут возникать по различным причинам, проявляя себя в характеристиках на конкретных временных масштабах.

Причиной долговременной зависимости в сетевом трафике могут быть следующие факторы [5]:

- поведение пользователя и прикладного программного обеспечения;
- генерация, структура и поиск данных;
- объединение трафика;
- средства администрирования сети;
- механизмы оптимизации, основанные на обратной связи;
- усложнение структуры сети, увеличение числа абонентов.

Обычно степень самоподобности описывается с помощью параметра Хёрста H [14], который принимает значения от 0 до 1. Для белого шума параметр Хёрста составляет 0,5, что означает полное отсутствие долговременной или кратковременной зависимости. Если $H > 0,5$, то исследуемый процесс обладает долговременной зависимостью [15].

2 Сбор данных

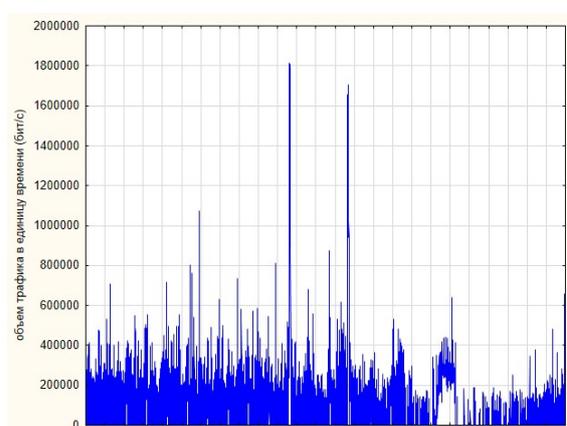
В исследовании производился мониторинг сервера корпоративной сети МГТУ им. Н.Э. Баумана. Физическая машина поделена на несколько виртуальных с ОС Linux, каждая из которых используется под ряд задач, таких как СУБД, web-серверы, файл-серверы и т.д. Большая часть трафика передается по HTTP, интерфейс 100Мбит Ethernet.

Каждый физический и виртуальный сервер сети университета использует для мониторинга Zabbix [16]. Zabbix – это клиент-серверное приложение, используемое для сбора, хранения и обработки информации о состоянии сети, сетевой нагрузке, а также состоянии операционной системы сервера в реальном времени. Приложение широко используется администраторами сети для мониторинга и своевременного оповещения о сбоях, перегрузках и отказах оборудования. Для дальнейшей обработки проводилось накопление данных следующих параметров:

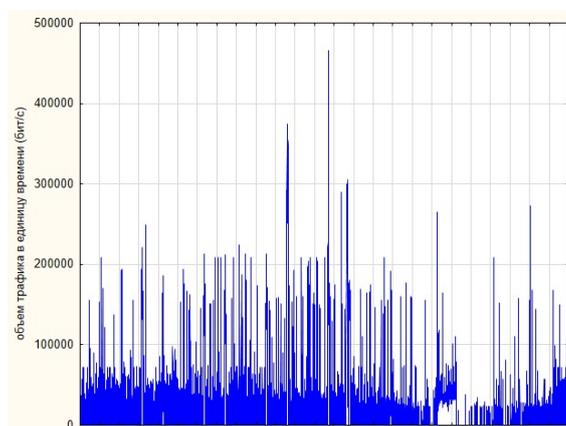
- объем кэшированной памяти;
- объем буферизированной памяти;
- время простоя процессора при операциях ввода/вывода;
- время процессора в режиме ожидания;
- время обработки процессором пользовательских процессов;

- время обработки процессором системных процессов;
- объем свободной памяти;
- входящий/исходящий трафик (бит/с);
- число процессов ОС;
- число процессов web-сервера Apache;
- суммарная загрузка процессора.

Данные снимались с различными временными промежутками для различных параметров, от 1 до 60 секунд в течение суток. Временная зависимость объема трафика в единицу времени (бит/с) приведена на рисунке 1 для входящего и исходящего трафика. На рисунке 2 показана степень загрузки ЦПУ (в %), а рисунок 3 иллюстрирует объем свободной памяти (бит).



а



б

Рисунок 1 – Временная зависимость объема трафика в единицу времени: (*а*) – входящий дневной трафик, (*б*) – исходящий дневной трафик

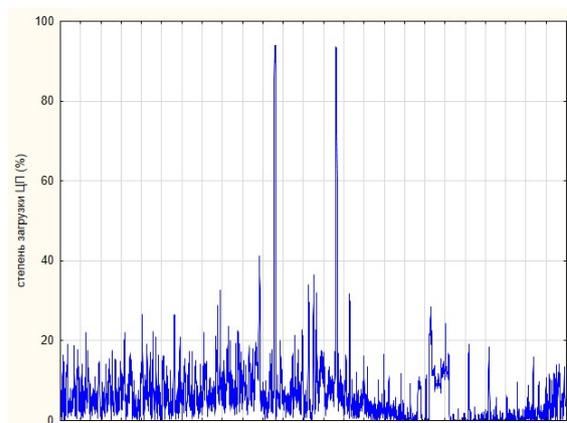


Рисунок 2 – Загрузка ЦПУ

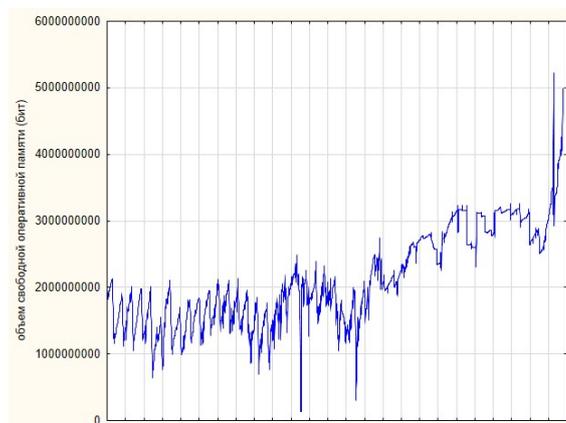


Рисунок 3 – Объем свободной памяти

3 Анализ автокорреляционной функции (АКФ)

Расчет автокорреляционной функции был произведен, чтобы оценить степень убывания зависимости процесса. Процесс с медленно убывающей зависимостью характеризуется АКФ, убывающей по степенному закону при увеличении временной задержки. Для самоподобных процессов функция автокорреляции агрегированного процесса $X^{(m)}$ при $m \rightarrow \infty$ не обращается в 0 (рисунок 4).

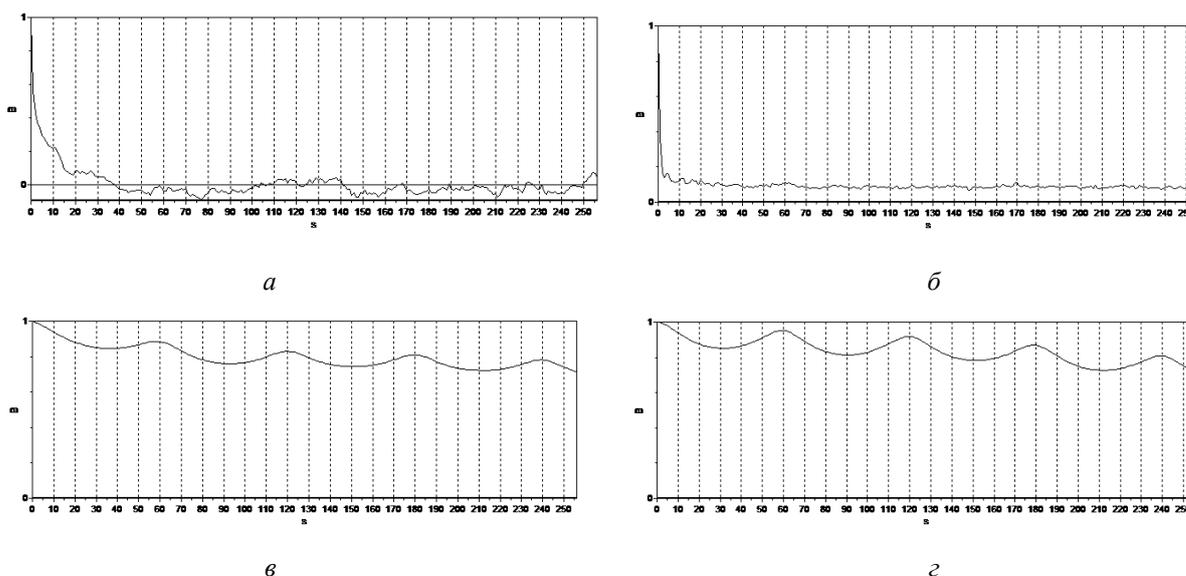


Рисунок 4 – Автокорреляционная функция входящего трафика (а), исходящего трафика (б), кэшированной памяти (в), свободного дискового пространства (г)

Из графиков АКФ можно заключить, что процесс передачи трафика обладает медленно убывающей зависимостью. Также стоит отметить, что АКФ для процесса выделения памяти обладает ярко выраженной периодичностью. Очевидно, что в данном случае процесс в первую очередь характеризуется логикой работы операционной системы, что вносит сильную периодичную составляющую.

В целом можно заключить, что исследование АКФ сетевого трафика, а также большинства аппаратных характеристик, позволяет в дальнейшем работать с процессами с учетом самоподобия и медленно убывающей зависимости.

4 Расчет параметра Хёрста

Как отмечалось выше, параметр Хёрста может быть мерой оценки долгосрочной зависимости временного ряда [17]. Оценка параметра не только может помочь сделать заключение о самоподобии процесса, но и позволит в дальнейшем применить к нему ряд

математических методов по прогнозированию фрактальных процессов [18]. Выполнение краткосрочных прогнозов нагрузки сервера может помочь в разработке методов аппаратной и программной оптимизации сети.

Параметр Хёрста H оценивался с помощью R/S анализа выборки данных:

$$\frac{R}{S} = \left(\frac{N}{2} \right)^H,$$

где R – размах временного ряда, S – среднеквадратичное отклонение, N – объем выборки.

Как и ожидалось, значение параметра указало на существование долгосрочной зависимости и самоподобия. Значения показателя Хёрста для различных характеристик трафика приведены в таблице 1.

Таблица 1 – Значения показателя Хёрста для основных характеристик трафика сети

Характеристика	Параметр Хёрста
Объем буферизированной памяти (бит/с)	0.9656
Объем кэшированной памяти (бит/с)	0.9868
Время процессора в режиме ожидания(%)	0.9575
Время обработки процессором системных задач(%)	0.9903
Объем свободной памяти (бит/с)	0.9336
Входящий трафик (бит/с)	0.9775
Число процессов ОС	0.8835
Число запущенных процессов web – сервера	0.8343
Исходящий трафик (бит/с)	0.9712

Была подтверждена гипотеза, указывающая на связь между параметром Хёрста и интенсивностью трафика (таблица 2). Так как в ряде работ [18] указывается на зависимость показателя Хёрста от числа отсчетов временного ряда, выборки брались на равных интервалах.

Таблица 2 – Зависимость показателя Хёрста от интенсивности трафика

	Значение параметра Хёрста			
	00:00 – 6:00	06:00 – 12:00	12:00 – 18:00	18:00 – 24:00
Входящий трафик (бит/с)	0.9516	0.9836	0.6154	0.7142
Исходящий трафик (бит/с)	0.9112	0.9766	0.7385	0.8180

Высокая утренняя интенсивность трафика объясняется работой ряда сервисов, в частности, документооборота университета, при котором автоматическая система занимается рассылкой документов и приказов в утренние часы.

5 Трафик как детерминированный хаос

В ряде работ по изучению свойств трафика процесс передачи данных по сети рассматривается с позиций теории хаоса [19]. Эта теория описывает поведение систем, чувствительных к начальным условиям. Выделение аттракторов на графике фазового пространства сигнализирует о фазовых переходах в процессах, что также может позволить сделать прогноз поведения системы (рисунок 5).

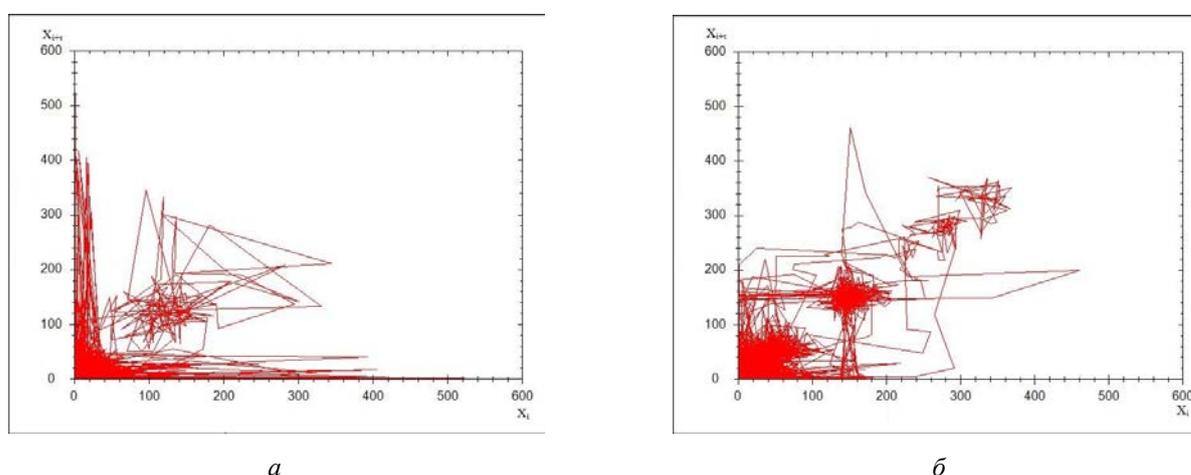


Рисунок 5 – Фазовые диаграммы входящего (а) и исходящего (б) сетевого трафика

В рамках исследования была произведена оценка экспоненты Ляпунова [20] как параметра, характеризующего хаотические системы:

$$\lambda_1(i) = \frac{1}{i\Delta t} \frac{1}{(M-i)} \sum_{j=1}^{M-i} \ln \frac{d_j(i)}{d_j(0)},$$

где Δt – период выборки, $d_j(i)$ – расстояние между j -й парой ближайших соседей после i дискретных шагов, M – число восстановленных точек.

Было разработано специальное программное обеспечение, с помощью которого на основе фазовой диаграммы производилась оценка старшего показателя Ляпунова.

Для всех типов трафика старший показатель Ляпунова λ_1 принимает значение от 1,2 до 2,5, что в дальнейшем позволит работать с трафиком в рамках методов нелинейной динамики. Стоит отметить, что показатель Ляпунова для процессов распределения памяти

сервера принимает значения ниже нуля, что говорит о высокой степени периодичности процессов.

Заключение

В работе были проведены исследования суммарного передаваемого сетевого трафика и соответствующих аппаратных процессов сервера сети университета. Для сетевого трафика был рассчитан показатель Ляпунова, характеризующий наличие хаоса в системе. Анализ экспоненты Ляпунова позволяет использовать методы нелинейной динамики в исследовании природы входящего и исходящего трафика. Для всех процессов было установлено самоподобие и рассчитан показатель Хёрста, характеризующий наличие долговременной памяти временного ряда. На основе этого можно сделать вывод о возможности в дальнейшем применения фрактальных моделей для работы с данными, в том числе использования методов прогнозирования.

Список литературы

1. Столлингс В. Современные компьютерные сети : пер. с англ. СПб.: Питер, 2003. 783 с. (Сер. "Классика computer science").
2. Таненбаум Э. Компьютерные сети : пер. с англ. СПб.: Питер, 2003. 992 с. (Сер. "Классика computer science").
3. Иванов И.П., Бойченко М.К. Мониторинг ресурсов узлов корпоративной сети // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2010. № 2. С. 114-120.
4. Бойченко М.К., Иванов И.П., Кондратьев А.Ю. Доступность ресурсов транспортных подсистем компьютерных сетей // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2010. № 3. С. 103-118.
5. Шелухин И.О., Тенякшев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях. М.: Радиотехника, 2003. 480 с.
6. Шелухин И.О., Осин А.В., Смольский С.М. Самоподобие и фракталы. Телекоммуникационные приложения. М.: Физмалит, 2008. 368 с.
7. Kelly F.P. Networks of Queues with Customers of Different Types // Journal of Applied Probability. 1975. Vol. 12, no. 3. P. 542-554.
8. Cheng Song. Packet Train Model: Optimizing Network Data Transfer Performance. Computer science technical report No. 867. University of Wisconsin-Madison. Aug. 1989.

9. 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. Feb. 1997. Vol. 5, no. 1. P. 71-86.
10. Heyman, D.P., Sobel M.J. Stochastic Models in Operations Research. Vol. I. Stochastic Processes and Operating Characteristics. New York: McGraw-Hill, 1982.
11. Jain R., Routhier S.A. Packet Trains – Measurement and a New Model for Computer Network Traffic // IEEE Journal on Selected Areas in Communications. Sep. 1986. Vol. 4, no. 6. P. 986-995.
12. Leland W.E., Taqqu M.S., Willinger W., Wilson D.V. On the Self-Similar Nature of Ethernet Traffic (Extended Version) // IEEE/ACM Transactions on Networking. February 1994. Vol. 2, no. 1. P. 1-15.
13. Ramakrishnan P. Self-Similar Traffic Models. Technical research report. 1999. Available at: http://www.isr.umd.edu/TechReports/ISR/1999/TR_99-12/TR_99-12.phtml , accessed 09.07.2013.
14. Hurst H., Black R. Long-Term Storage: An Experimental Study. London: Constable, 1965.
15. Hurst H.E. Long-Term Storage Capacity of Reservoirs // Transactions of the American Society of Civil Engineering. 1951. Vol. 116. P. 770-799.
16. User manual Zabbix. Available at: <http://www.zabbix.com/en/documentation.php> , accessed 09.07.2013.
17. Федер Е. Фракталы : пер. с англ. М.: Мир, 1991. 254 с.
18. Кириллов Д.С., Короб О.В., Митин Н.А., Орлов Ю.Н., Плешаков Р.В. Распределения показателя Херста нестационарного маркированного временного ряда. М.: ИПМ им. М.В. Келдыша. 2013. 16 с. (Препринт / ИПМ им. М.В. Келдыша; № 11). Режим доступа: <http://library.keldysh.ru/preprint.asp?id=2013-11> (дата обращения 09.07.3013).
19. Шредер М. Фракталы, хаос, степенные законы. Миниатюры из бесконечного рая : пер. с англ. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. 528 с.
20. Rosenstein M.T., Collins J.J., De Luca C.J. A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets // Physica D: Nonlinear Phenomena. 1993. Vol. 65, iss. 1-2, P. 117-134. [http://dx.doi.org/10.1016/0167-2789\(93\)90009-P](http://dx.doi.org/10.1016/0167-2789(93)90009-P)

Analysis of University's corporative network traffic by the methods of nonlinear dynamics

08, August 2013

DOI: 10.7463/0813.0587054

Basarab M.A., Kolesnikov A.V., Ivanov I.P.

Bauman Moscow State Technical University, 105005, Moscow, Russian Federation

avkolesnikov90@list.ru

ivanov@bmstu.ru

bmic@mail.ru

Results of calculations of incoming and outgoing traffic's dynamic characteristics were presented in this paper along with the distribution of hardware capacities by the example of the university's corporate network server. Lyapunov and Hurst exponents, which characterize chaotic and fractal properties of the processes, were evaluated for network traffic. Traffic's phase diagrams were investigated and, for the first time, appearance of attractors was revealed; these attractors allow one to conduct a more detailed analysis of the impact of load on the network's traffic capacity. A self-similarity was found for all the processes; this fact confirms the possibility of using fractal models for working with data, including solving urgent problems of predicting behavior of the time series.

Publications with keywords: [the Lyapunov exponent](#), [Hurst exponent](#), [network traffic](#), [time series](#), [chaotic dynamics](#)

Publications with words: [the Lyapunov exponent](#), [Hurst exponent](#), [network traffic](#), [time series](#), [chaotic dynamics](#)

References

1. William S. *High-Speed Networks and Internets: Performance and Quality of Service*. 2nd ed. Prentice Hall, 2001. 715 p. (Russ. ed.: Stollings V. *Sovremennye komp'yuternye seti*. St. Petersburg, Piter, 2003. 783 p.).
2. Tanenbaum Andrew S. *Computer Networks*. 4th ed. Prentice Hall, 2003. 891 p. (Russ. ed.: Tanenbaum E. *Komp'yuternye seti*. St. Petersburg, Piter, 2003. 992 p.).
3. Ivanov I.P., Boychenko M.K. Monitoring resursov uzlov korporativnoy seti [Monitoring of resources of corporate network junctions]. *Vestnik MGTU im. N.E. Baumana. Ser. Priborostroenie* [Herald of the Bauman MSTU. Ser. Instrument Engineering], 2010, no. 2, pp. 114-120.
4. Boychenko M.K., Ivanov I.P., Kondrat'ev A.Yu. Dostupnost' resursov transportnykh podsystem komp'yuternykh setey [Availability of resources of transport subsystems of corporative networks]. *Vestnik MGTU im. N.E. Baumana. Ser. Priborostroenie* [Herald of the Bauman MSTU. Ser. Instrument Engineering], 2010, no. 3, pp. 103-118.

5. Shelukhin I.O., Tenyakshev A.M., Osin A.V. *Fraktal'nye protsessy v telekommunikatsiyakh* [Fractal processes in telecommunications]. Moscow, Radiotekhnika, 2003. 480 p.
6. Shelukhin I.O., Osin A.V., Smol'skiy S.M. *Samopodobie i fraktaly. Telekommunikatsionnye prilozheniya* [Self-similarity and fractals. Telecommunication applications]. Moscow, Fizmatlit, 2008. 368 p.
7. Kelly F.P. Networks of Queues with Customers of Different Types. *Journal of Applied Probability*, 1975, vol. 12, no. 3, pp. 542-554.
8. Cheng Song. *Packet Train Model: Optimizing Network Data Transfer Performance*. Computer science technical report no. 867. University of Wisconsin-Madison, Aug. 1989.
9. 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*, Feb. 1997, vol. 5, no. 1, pp. 71-86.
10. Heyman, D.P., Sobel M.J. *Stochastic Models in Operations Research. Vol. I. Stochastic Processes and Operating Characteristics*. New York, McGraw-Hill, 1982.
11. Jain R., Routhier S.A. Packet Trains – Measurement and a New Model for Computer Network Traffic. *IEEE Journal on Selected Areas in Communications*, Sep. 1986, vol. 4, no. 6, pp. 986-995.
12. Leland W.E., Taqqu M.S., Willinger W., Wilson D.V. On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transactions on Networking*, February 1994, vol. 2, no. 1, pp. 1-15.
13. Ramakrishnan P. *Self-Similar Traffic Models. Technical research report*. 1999. Available at: http://www.isr.umd.edu/TechReports/ISR/1999/TR_99-12/TR_99-12.phtml , accessed 09.07.2013.
14. Hurst H., Black R. *Long-Term Storage: An Experimental Study*. London, Constable, 1965.
15. Hurst H.E. Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineering*, 1951, vol. 116, pp. 770-799.
16. *User manual Zabbix*. Available at: <http://www.zabbix.com/en/documentation.php> , accessed 09.07.2013.
17. Feder J. *Fractals*. Plenum Press, New York, 1988. (Russ. ed.: Feder E. *Fraktaly*. Moscow, Mir, 1991. 254 p.).
18. Kirillov D.S., Korob O.V., Mitin N.A., Orlov Yu.N., Pleshakov R.V. *Raspredeleniya pokazatelya Khersta nestatsionarnogo markirovannogo vremennogo ryada. Preprint no. 11* [On the stationary distributions of the Hurst indicator for the non-stationary marked time series. Preprint no. 11]. Moscow, Keldysh Institute of Applied Mathematics (Russian Academy of Sciences), 2013. 16 p. Available at: <http://library.keldysh.ru/preprint.asp?id=2013-11> , accessed 09.07.2013.
19. Schroeder M. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W.H. Freeman and Company, 1991. 429 p. (Russ. ed.: Shreder M. *Fraktaly, khaos, stepennyye zakony. Miniatyury iz beskonechnogo raya*. Izhevsk, NITs RKD Publ., 2001. 528 p.).
20. Rosenstein M.T., Collins J.J., De Luca C.J. A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets. *Physica D: Nonlinear Phenomena*, 1993, vol. 65, no. 1-2, pp. 117-134. [http://dx.doi.org/10.1016/0167-2789\(93\)90009-P](http://dx.doi.org/10.1016/0167-2789(93)90009-P)