які дозволяють враховувати витрати часу в моделі, наближаючи до поведінки реальної системи або процесу.

Networks Petri allow to model various systems and processes by mathematical representation, with an opportunity of reception of the information on behaviour of structure of model. The analysis of networks Petri should be spent for correct and adequate work. The analysis of correct work of models can be carried out by means of a tree of approachibilities and the matrix equations which allow to reveal errors of modelling. The analysis of adequate work of models can be lead by means of time networks of events which allows to consider expenses of time in model, approaching to behaviour of real system or process.

УДК 621.395.7

ИГНАТЕНКО Е.Г., аспирант (ДонНТУ),
ТУРУПАЛОВ В.В., к.т.н, доцент (ДонНТУ),
БЕССАРАБ В.И. к.т.н, доцент (ДонНТУ).

Исследование структуры потока http-запросов в телекоммуникационных сетях

Актуальность

В классической теории телетрафика принято считать, что информационные потоки в телекоммуникационных сетях, адекватно описываются Пуассоновским распределением. Такое допущение является верным для сетей небольшого размера и позволяет использовать известные методы теории массового обслуживания для расчета сетевых параметров. Но с ростом размера сетей, увеличением разнообразия сетевых приложений, появлением новых протоколов передачи данных в поведении трафика проявляются фрактальные свойства [1]. Для такого трафика методы расчета, которые традиционно базируются на пуассоновских моделях и формулах Эрланга, приводят к неадекватной оценке реальной нагрузки. Также известно, что фрактальным процессам присуще свойство самоподобия или масштабной инвариантности, которое для сетевых процессов заключается в том, что с увеличением интервала агрегирования временного ряда сохраняется структура нижележащих уровней, и интервал корреляции теоретически стремится к бесконечности [1-3].

К настоящему времени показано, то самоподобной структурой обладает трафик в телекоммуникационных сетях при работе протоколов Ethernet [4], VoIP [5], TCP [6] и ОКС7 [7]. При этом проводился анализ сведений об информации, полученной пользователями сети.

Постановка цели и задач исследования

Целью работы является разработка методики анализа временного ряда входящего потока http-запросов, обладающего свойством самоподобия. Для достижения поставленной цели необходимо решить следующие задачи:

- проанализировать существующие подходы к исследованию временных рядов трафика;
- проанализировать временной ряд

ABBYY PDF Transformer 2.0 Click here to buy www.ABBYY.com

входящего потока http-запросов с целью выявления свойств самоподобия для последующего построения на его основе прогнозных моделей.

Основная часть

Экспериментальный случайный процесс рассматривается как дискретная последовательность случайных величин, т.е. аргументом считается порядковый номер такой единицы времени:

$$X = \{X_t : t = 0,1,2...\}. \qquad (1)$$

Будем считать, что рассматриваемый случайный процесс имеет ограниченную ковариацию $B(x_i, x_{i+\tau}) < \infty, \forall \tau$ и следовательно, дисперсию $\sigma_{x_i}^2 = B(x_i, x_{i+\tau}) < \infty$. Случайный процесс будет обладать свойством самоподобия, если агрегированный процесс m-го порядка:

$$X^{(m)} = [x_1^{(m)}, x_2^{(m)}, \ldots, x_n^{(m)}] = \left\{\frac{1}{m}\sum_{i=1}^{m} x_{(k-1)m+1},\ k = 1,2,\ldots\right\} \qquad (2)$$

будет иметь корреляционную функцию $r^m(k)$, совпадающую с корреляционной функцией исходного случайного процесса $r(k)$ для любых m.

Важнейшим параметром, характеризующим степень самоподобия, является параметр Херста (H). С учетом показателя Херста выделяют три типа случайных процессов:

1. $0 \le H \le 0.5$ - случайный процесс, который не обладает самоподобием;
2. $H = 0.5$ - полностью случайный ряд, аналогичный случайным смещениям при классическом броуновском движении частицы;
3. $H > 0.5$ - самоподдерживающийся процесс, который обладает длительной памятью и является самоподобным.

Коэффициент Херста можно определять различными способами - с помощью R/S -анализа, анализа изменения дисперсий, периодограммного анализа и анализа АКФ. Все они отличаются друг от друга по сложности вычислений и точности получаемого результата.

Считается, что временной ряд, обладающий свойством самоподобия, подчиняется распределению с тяжелым хвостом. Распределение имеет тяжелый хвост, если

$$P(x > X) = x^{-\alpha},\ 0 < \alpha < 2. \qquad (3)$$

Простейшим распределением с тяжелым хвостом является распределение Парето, для которого функция распределения имеет вид:

$$F(x) = \left(1 - \frac{k}{x}\right)^{\alpha}. \qquad (4)$$

Распределение с тяжелым хвостом имеет ряд свойств, которые существенно отличают его от наиболее известных распределений, таких как экспоненциальное, нормальное или Пуассоновское. Существует несколько методов для оценки тяжести хвоста α по экспериментальным данным. Для оценки $-\alpha$ строят линию регрессии для дополнительного распределения $\overline{F}(x) = 1 - F(x)$ в логарифмическом масштабе и вычисляют тангенс угла наклона к горизонтальной оси. Параметр Херста связан с параметром тяжести хвоста следующим соотношением:

$$H = \frac{3 - \alpha}{2}. \qquad (5)$$

Самоподобные процессы, в отличие от пуассоновских, характеризуются наличием последействия, т.е. вероятность поступления следующего события зависит не только от времени, но и от предыдущих значений событий. Это означает, что число текущих событий может зависеть от предыдущих событий в отдаленные про-

ABBYY PDF Transformer 2.0 Click here to buy www.ABBYY.com

межутки времени. Именно поэтому одним из основных свойств самоподобного процесса является медленно убывающая зависимость (МУЗ). Считается, что процесс обладает МУЗ, если он характеризуется АКФ, которая убывает гиперболически (по степенному закону) при увеличении лага [8]. Понятие медленно убывающей зависимости имеет ключевое значение в теории самоподобных процессов. Процесс X обладает медленно убывающей зависимостью (МУЗ), если выполняется условие [9-11]:

$$r(k) \sim k^{-\beta} L_1(k),\ k \to \infty \qquad (6)$$

где $0 < \beta < 1$, L_1 - медленно меняющаяся на бесконечности функция, т.е. $\lim\limits_{t\to\infty} \frac{L_1(tx)}{L_1(t)} = 1$ для всех x>0.

Для того чтобы проверить гипотезу о присутствии МУЗ в некотором временном ряде, необходимо решить задачу регрессии, т.е. вычислить по экспериментальной АКФ методом наименьших квадратов параметры A_0 и β модели:

$$r(k) = A_0 \cdot k^{-\beta}, 0 < \beta < 1, k \to \infty, A_0 - const. (7)$$

Для оценки степени точности необходимо найти дисперсию $\sigma^2{}_{muz}$ разности экспериментальной АКФ и полученной ее аппроксимации (7). Одновременно проверяется гипотеза о присутствии быстро убывающей зависимости (БУЗ) в этом же ряде. Для этого аналогичным образом вычисляются параметры B_0 и ρ модели (8):

$$r(k) = B_0 \cdot \rho^k,\ k \to \infty, 0 < \rho < 1, B_0 - const. (8)$$

Для оценки степени точности необходимо найти дисперсию $\sigma^2{}_{buz}$ разности экспериментальной АКФ и полученной ее аппроксимации (8). Вывод о присутствии МУЗ в ряде делают, если $\sigma^2{}_{muz}$ достаточно мало и значение коэффициента LR<1.

$$LR = \frac{\sigma^2_{muz}}{\sigma^2_{buz}}. \qquad (9)$$

В рассматриваемой статье разрабатывается обоснованная методика, подтверждающая самоподобие http-трафика, основанная на измерениях характеристик реального сетевого ресурса. В качестве объекта исследования рассматривается процесс поступления http-запросов на реальный web-сервер за период длиной 1 неделя с 10 апреля 2010 года по 17 апреля 2010 года. Фиксируемые характеристики – моменты поступления http-запросов. Все действия сервера записываются в так называемый лог-файл. Необходимыми данными для исследований является количество запросов к серверу, которые получаются из анализа лог-файла.

Анализируемые данные, которые представляют собой количество запросов пользователей к серверу в течение суток, взяты непосредственно из лог-файла. Так же можно рассматривать размеры передаваемых файлов, интервалы передачи и другие реальные характеристики сетевого трафика.

Осуществим с имеющимся временным рядом следующий процесс агрегирования – определим количество http-запросов за 1 минуту $X = (X_t, t = 1,2,...,1440)$. Агрегированный по 1 минуте поток запросов приведен на рисунке 1.

Выполним уменьшение размера шкалы наблюдений в 5 раз. Для этого вычислим новый ряд согласно формуле (2). Агрегированный по 5 минут поток запросов приведен на рисунке 2. Далее проведем такую же процедуру с исходным рядом (рисунок 1) при m=10. Таким образом, одно деление будет содержать 10 единиц исходной реализации. Агрегированный по 10 минут поток запросов приведен на рисунке 3.

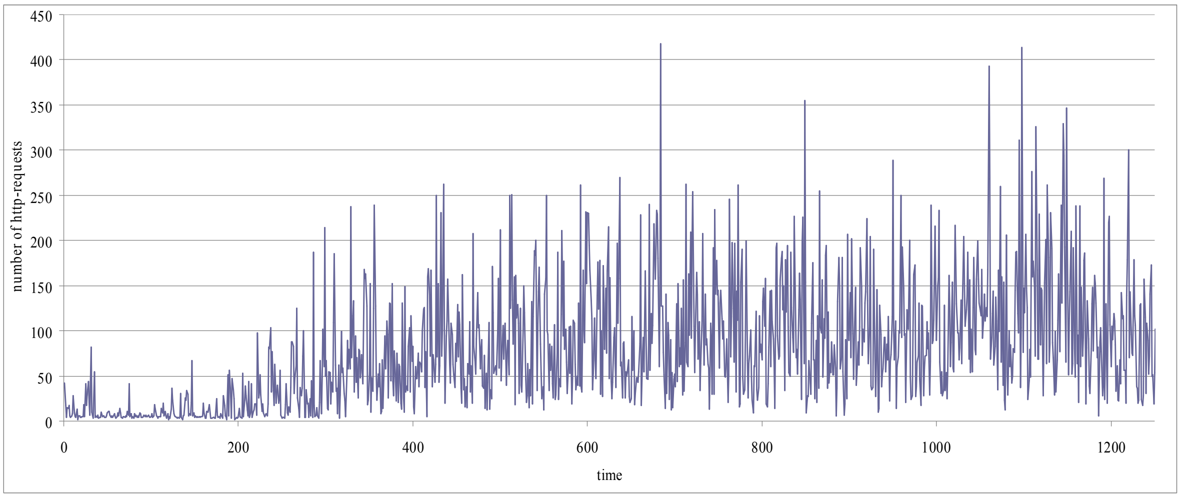

Рисунок 1. – Агрегированная за 1 минуту реализация входящего потока запросов

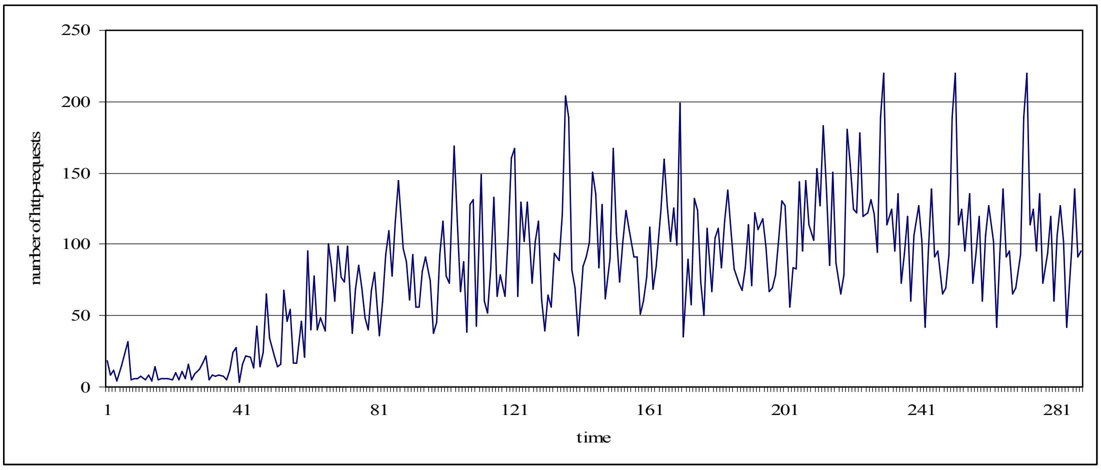

Рисунок 2. – Агрегирование входящего потока $X^{(m)}$, $m = 5$

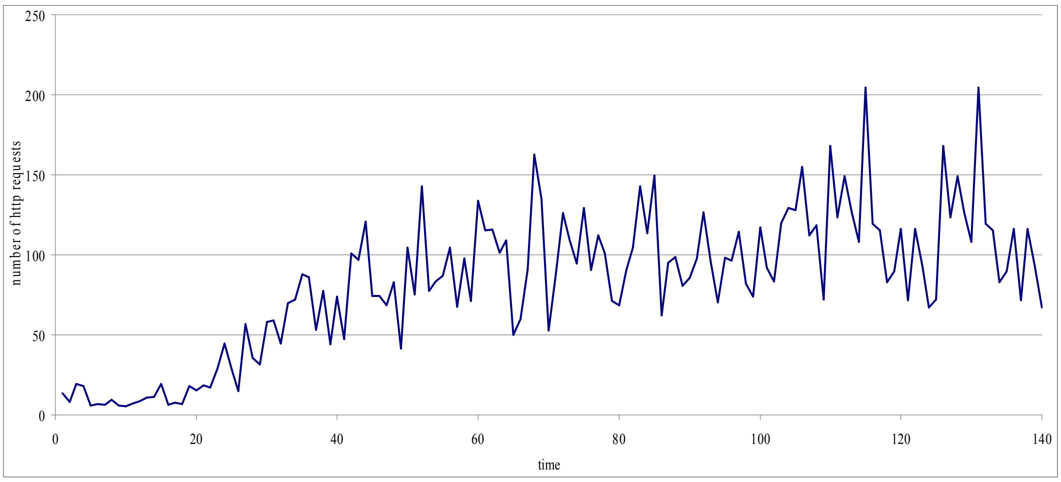

Рисунок 3. – Агрегирование входящего потока $X^{(m)}$, $m = 10$

Основываясь на рекомендациях [12-14] и проанализировав полученные результаты, можно сделать предположение о том, что рассматриваемый входящий поток имеет самоподобную структуру. Этот вывод можно сделать исходя из определения самоподобия, которое говорит о том, что структура ряда, полученного усреднением групп элементов, остается такой же, как и структура исходного. Этот эффект наблюдается на рисунке 1 - рисунке 3, иллюстрирующих изменение шкалы. Этот факт является предпосылкой для предположения о самоподобной структуре рассматриваемого потока и основанием для проведения детального анализа.

Проведем оценку тяжести хвоста α по экспериментальным данным:

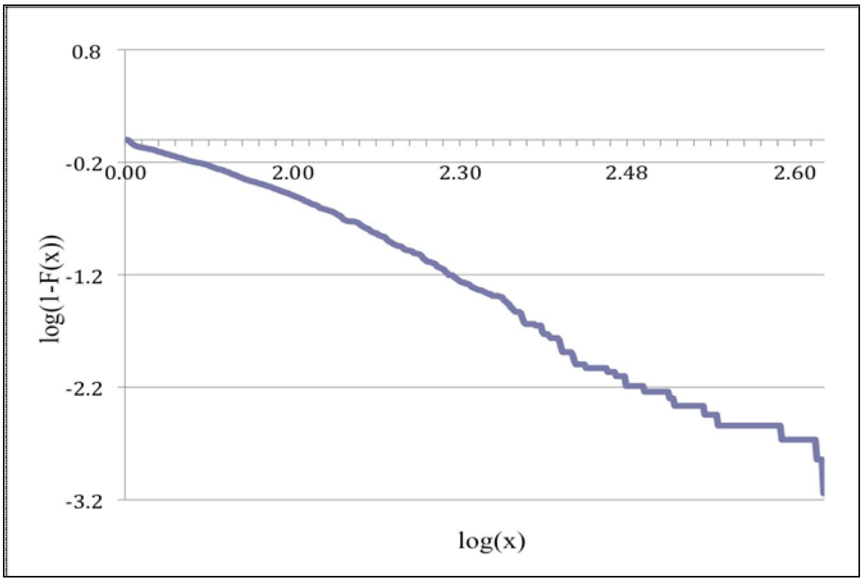

Рисунок 4. - График дополнительного распределения

Коэффициент $-\alpha$ - является оценкой тяжести хвоста. Построив линию регрессии для дополнительной функции распределения и вычислив тангенс угла наклона к горизонтальной оси, определяем значение параметра $\alpha = 1{,}36$. Значение попадает в промежуток от 0 до 2, откуда следует, что анализируемое распределение имеет свойство тяжелого хвоста. Тяжелые хвосты также являются причиной хорошей предсказуемости потока [15].

Используя найденное значение параметра α, по формуле (5) определяем значение параметра Херста H=0,82. Для сравнения в таблице приведены значения параметра Херста, рассчитанные различными способами.

Таблица 1. - Оценка показателя Херста различными методами

R/S - анализ	Периодограммный метод	Изменения дисперсии	АКФ
0,73062	0,6891	0,9097	0,8725

Построим АКФ для исходной реализации с H=0,82 и для сравнения АКФ с параметрами Херста H=0,9; 0,7 и 0,5

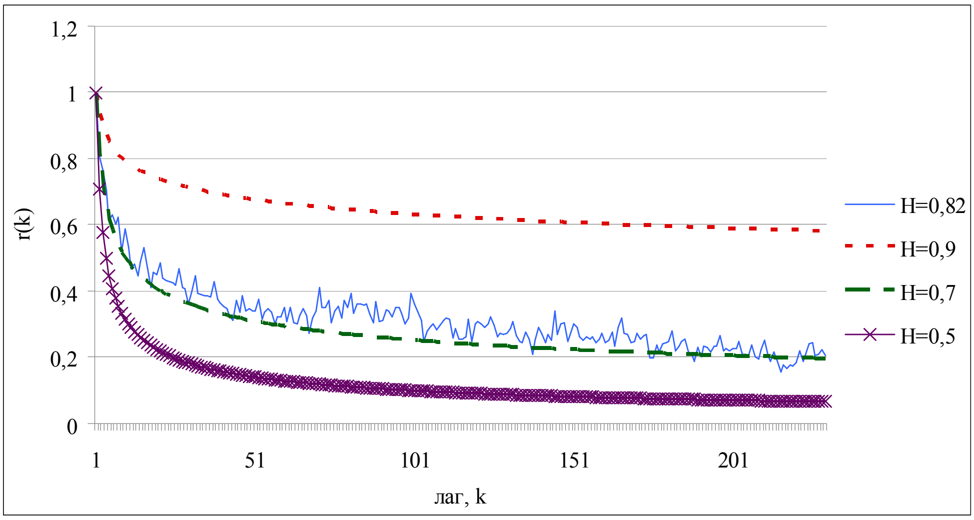

Рисунок 5. - АКФ исследуемого процесса H=0,82; АКФ при H=0,9; 0,7 и 0,5

Анализируя АКФ, можно заметить достаточно слабое ее убывание: при сдвиге 150 значений АКФ остается выше 0,3. В общем случае АКФ характеризует внутреннюю зависимость между временным рядом и тем же рядом, но сдвинутым на некоторый промежуток (сдвиг) времени, который называется лагом.

Для исходной реализации процесса построены регрессионные модели МУЗ и БУЗ, рассчитаны $\sigma^2{}_{muz}$ и $\sigma^2{}_{buz}$, оценен коэффициент LR=0,3868. Таким образом, результаты анализа АКФ реализации http трафика подтверждают присутствие медленно убывающей зависимости.

С понятием МУЗ связано важнейшее прогнозирующее свойство – продолжительная память, характеризующаяся зависимостью текущих параметров процесса от предыдущих.

Выводы

Проведен эксперимент по сбору и исследованию характеристик http трафика в телекоммуникационной сети. Проведенное исследование подтверждает наличие самоподобных свойств в http трафике современных телекоммуникационных сетей. Показано, что входящий поток подчиняется распределению с тяжелым хвостом. Результаты анализа АКФ реализации http трафика подтверждают присутствие медленно убывающей зависимости. С целью получения достоверных результатов, показатель Херста рассчитывался несколькими методами: анализа дисперсии, нормированного размаха (R/S), периодограмм. Для всех методов показатель Херста H>0.5, значит, трафик относится к классу персистентных процессов.

В процессе анализа выявлено присутствие регулярной детерминированной, циклической и случайной составляющих в агрегированном сетевом трафике, что может быть использовано при решении задачи прогнозирования трафика с целью управления телекоммуникационными сетями.

Список литературы

1.Городецкий А. Я., Заборовский В. С. Информатика. Фрактальные процессы в компьютерных сетях. Учебное пособие. - СПб.: Изд-во СПбГТУ, 2000 - 102 с.

2. Морозов А. Д. Введение в теорию фракталов. М.: Институт компьютерных исследований, 2002. - 159 с.

3. Кроновер Р. М. Фракталы и хаос в динамических системах. М.: Постмаркет, 2000. - 352 с.

4. W.E.Leland, M.S.Taqqu, W.Willinger, and D.V.Wilson. On the self-similar nature of Ethernet traffic (extended version).IEEE/ACM Transactions of Networking, 2(1):1-15,1994.

5. T. D. Dang, B. Sonkoly, S. Moln·r, Fractal Analysis and Modelling of VoIP Traffic, NETWORKS2004, Vienna, Austria, 2004.

6. Feng W., Tinnakornsrisuphap P. The Failure of TCP in High-Performance Computational Grids //SC2000: High-Performance Network and Computing Conference, Dallas, TX , November 2000.

7. Криштофович А.Ю. Самоподобие трафика сети ОКС №7. МКИССиТ, Санкт- Петербург, 2002 г.

8. Крылов В.В., Самохвалова С.С. Теория телетрафика и ее приложения.- СПб.: БХВ-Петербург, 2005.

9. Tsybakov B.S., Georganas N.D. Self-similar processes in communications networks // IEEE Trans. Inform. Theory, vol. 44. Sep.1998. P. 1713-1725.

10. Leland W.E., Taqqu M.S., Willinger W., and Wilson D.V. On the self-similar nature of ethernet traffic // IEEE/ACM Transactions of Networking, 2(1), 1994. P. 1-15.

11. Цыбаков Б.С. Модель телетрафика на основе самоподобного случайного процесса // Радиотехника. 1999. № 5. С. 24-31.

12. Mark E. Crovella and Azer Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes" in IEEE/ACM Transactions on Networking, 5(6):835--846, December 1997.

13. Mark E. Crovella, Murad Taqqu and Azer Bestavros, "Heavy Tailed-Probability distributions in the World Wide Web" 5(6):835--846, December 1997.

14. Mark E. Crovella, Azer Bestavros, Paul Baarfor, Adam Bradley "Change in Web Client Access Patterns. Characteristics and Caching Implications", Computer Science Department Boston University, 1999.

15. Tsybakov B.S., Georganas N.D. Self-similar processes in communications networks // IEEE Trans. Inform. Theory, vol. 44. Sep.1998. P. 1713-1725.

Аннотации:

Ключевые слова: http трафик, самоподобие, параметр Херста, АКФ.

Проведено исследование характеристик http трафика в телекоммуникационной сети. Проанализирована структура трафика и выявлены его составляющие с целью управления телекоммуникационными сетями. Разработана методика анализа временного ряда входящего потока http-запросов, обладающего свойством самоподобия.

Проведено дослідження характеристик http трафіку в телекомунікаційній мережі. Проаналізовано структуру трафіку та виявлено його складові з метою управління телекомунікаційними мережами. Розроблено методику аналізу самоподібного часового ряду вхідного потоку http-запитів.

There was an examination of HTTP traffic characteristics accomplished. The structure of traffic was considered and components identified in order to control internetworking. Method of incoming http requests stream time series analysys was developed. This applies for self-similar time series.