УДК 004.7

ИСПОЛЬЗОВАНИЕ ФРАКТАЛЬНЫХ СВОЙСТВ ТРАФИКА В ЦИФРОВЫХ СЕТЯХ СВЯЗИ ДЛЯ ДЕТЕКТИРОВАНИЯ СЕТЕВЫХ АНОМАЛИЙ

И. С. Барсуков, М. П. Ряполов

Воронежский государственный университет

Поступила в редакцию 20.08.2018 г.

Аннотация. В статье рассмотрены проблемы распознавания сетевых DoS атак. Предложен метод детектирования на основе оценки фрактальных (самоподобных) свойств сетевого трафика путем вычисления коэффициента Херста. Для его реализации использован алгоритм анализа R/S статистик. Проведено экспериментальное исследование образцов сетевого трафика с помощью этого метода, полученные результаты указывают на то, что по оценке изменения коэффициента Херста в реальном времени можно делать вывод о наличии аномальных выбросов в трафике.

Ключевые слова: самоподобие трафика, коэффициент Херста, сетевые аномалии, DoS атаки.

Annotation. Article describes problems of network DoS attacks recognition. Detecting method based on evaluation of traffic fractal properties by calculating of Hurst parameter has been proposed. R/S (rescaled range) analysis algorithm was used for its realization. Experimental research of network traffic samples through aforementioned algorithm has been conducted; obtained results reveal that it's possible to make decision about existence of network traffic abnormalities based on Hurst parameter evaluation.

Keywords: self-similarity, Hurst parameter, network abnormalities, DoS attack.

ВВЕДЕНИЕ

Наличие фрактальных свойств сетевого трафика было обнаружено несколько десятилетий назад, когда было установлено, что на больших масштабах он обладает свойством самоподобия, то есть выглядит качественно одинаково при достаточно больших масштабах временной оси и проявляет долговременную зависимость. Господствовавшие до этого модели трафика, основанные на марковских процессах, обладали кратковременной зависимостью. Они были заимствованы из телефонных сетей и, применительно к компьютерным сетям приводили к недооценке нагрузки. Открытие самоподобия трафика оказало существенное влияние на последующее развитие клиент-серверных информационных систем и позволило переосмыслить вероятностно-временные характеристики таких систем.

Одним из актуальных и практически значимых приложений фрактального анализа трафика является обнаружение атак и других сетевых аномалий. Злоумышленники постоянно совершенствуются и видоизменяют методы и стратегии атак. Анализ существующих решений показал, что пока не существует однозначно надежного и действенного инструмента по детектированию атак в реальном времени, поэтому это направление исследований продолжает оставаться высокоперспективным. Фрактальный анализ позволяет выявить несвойственные для обычного трафика структурные особенности, вызванные аномальными изменениями, что в свою очередь может послужить сигналом для своевременного блокирования атаки [2, 3, 6].

Методика детектирования аномалий через фрактальный анализ еще не имеют устоявшегося подхода, и работ, касающихся этой темы, не очень много. Среди имеющихся стоит выделить [5], в которой предлагается использовать вейвлет-анализ для оценки фрактальных

свойств трафика. В работе [11] описывается адаптивный метод, основанный на оценке АКФ трафика, а в [12] основное внимание отводится описанию лабораторного макета и модуля для детектирования аномалий, но не уточняется, как именно оценивались самоподобные свойства трафика.

В настоящей статье с помощью алгоритма оценки показателя Херста, характеризующего фрактальную размерность, исследовались нормальные реализации трафика и реализации, подверженные аномалиям. Алгоритм определения показателя Херста, реализованный с помощью программного пакета математического моделирования MATLAB, представляет собой видоизмененный метод анализа R/S статистик, в который были внесены поправки с целью повышения точности вычисления коэффициента.

Целью работы явилось обнаружение самоподобных свойств в обычном пользовательском трафике, а также проверка предположения, что самоподобные свойства аномального трафика имеют значительные отличия от нормального. В итоге предложен метод детектирования аномального трафика и рассмотрена его работа на примере корпоративной сети университета.

МАТЕРИАЛЫ И МЕТОДЫ

Фрактальный процесс можно определить как случайный процесс, статистические характеристики которого проявляют свойства самоподобия (масштабной инвариантности). 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.

Введем понятие автокорреляционной функции стационарного случайного процесса $X = (X_i : i = 0,1,2,\ldots)$ с постоянным средним $\mu < \infty$:

$$r(k) = E[(X_i - \mu)(X_{i+k} - \mu)] / E[(X_i - \mu)^2] \quad (1)$$

Положим, что $r(k) \sim a_1 k^{-\beta}$, где $0 < \beta < 1$ и $a_1 = \text{const}$. Далее обозначим через $X^{(m)} = (X_k^{(m)} : k = 1,2,3,\ldots)$ усредненный по блокам длины m процесс X, компоненты которого определяются равенством $X_k^{(m)} = \frac{1}{m}(X_{km-m+1} + \cdots + X_{km})$, где $m, k \in N$, и назовем его агрегированным. АКФ такого процесса будет равна $r^{(m)}$. Тогда процесс X будет называться строго самоподобным в широком смысле (аналогичное понятие в зарубежной литературе – exactly second-order self-similar) с параметром $H = 1 - \beta / 2$, если $r^{(m)}(k) = r(k)$, или другими словами, если агрегированный процесс $X^{(m)}$ эквивалентен исходному X, как минимум в отношении статистических характеристик 2-го порядка.

Введенный ранее параметр H называется показателем Херста и по его значениям можно судить о степени самоподобности процесса. Доказано, что для белого шума показатель Херста равен 0.5, что означает полное отсутствие какой-либо зависимости. При $0.5 < H < 1$ процесс является персистентным, то есть сохраняющим имеющуюся тенденцию. При $H < 0.5$ процесс характеризуется антиперсистентностью – любая тенденция стремится смениться противоположной.

Одним из способов вычисления коэффициента Херста является анализ т. н. R/S статистики (нормированного размаха):

$$E\left[\frac{R(n)}{S(n)}\right] = Cn^H, \quad n \to \infty, \; C = \text{const}, \quad (2)$$

где $R(n)$ – размах первых n значений ряда, $S(n)$ – стандартное отклонение, $E[x]$ – математическое ожидание.

Если прологарифмировать обе части (2):

$$\log E\left[\frac{R(n)}{S(n)}\right] = H \log(n) + \log(C) \quad (3)$$

и построить график зависимости $\log E\left[\frac{R(n)}{S(n)}\right]$ от $\log(n)$, то наклон прямой, аппроксимирующей эту зависимость и будет являться показателем Херста [8].

Как показывают исследования, трафик обладает свойством самоподобия как до, так и во время аномальных изменений. Причем во время появления аномальных выбросов, связанных с увеличением интенсивности тра-

фика, наблюдается скачок фрактальной размерности, что характеризуется ростом значений показателя Херста, при этом зачастую оказывается, что $H > 1$ [5, 8, 12]. Это свидетельствует о потенциале фрактального подхода для обнаружения вторжений в компьютерные сети в режиме реального времени.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Исследования трафика в рамках данной статьи носили двухэтапный характер.

На первом этапе исследовался реальный пользовательский трафик сети ВГУ, генерируемый при обращении к веб-серверу https://edu.vsu.ru/. Дампы трафика снимались десятиминутными интервалами в 5-ти временных отрезках (9.00, 12.00, 16.00, 18.00, 20.00 часов) в течение нескольких недель. Из полученных дампов выполнялась выборка определенных атрибутов трафика (статистических характеристик) для формирования числового ряда. Атрибутами трафика, подлежащими исследованию, были выбранные следующие характеристики:

– интервал времени между пакетами

– количество переданных байт трафика в секунду

– продолжительность TCP сессий

– число новых соединений в секунду (для TCP пакетов).

Далее для сформированных числовых рядов рассчитывался показатель Херста по методу R/S статистик.

В табл. 1 приведены усредненные значения показателя Херста в разные суточные временные отрезки (9.00, 12.00, 16.00, 18.00, 20.00 часов) для вышеупомянутых атрибутов трафика. Усреднение проведено по всему диапазону наблюдения (25 дней). Анализ приведенных в таблице данных показывает, что на всех анализируемых временных интервалах трафик обладает самоподобными свойствами, так как значение показателя Херста лежит в интервале от 0.5 до 1.

На втором этапе проводилось исследование нормального пользовательского трафика при обращении к веб-серверу https://edu.vsu.ru/ с примесью аномального трафика, генерируемого приложением для проведения нагрузочного тестирования JMeter. На рис. 1 представлена общая схема лабораторного стенда, собранного для выполнения исследования. При создании аномального трафика за основу брался нормальный трафик единичного пользователя, из которого далее с помощью тестового плана JMeter формировался многократно усиленный, лавинообразный трафик со значительными выбросами по сравнению со стандартным повседневным трафиком при обращении к данному ресурсу (узел А рис. 1). Параметры аномального трафика варьировались через следующие опции тестового плана – количество потоков (т. е. число одновременно симулируемых пользователей, “Thread Count”), время набега (в течение которого происходит симуляция всех пользователей, “Ramp-Up Time”), количество повторений (сколько раз будет запущен тестовый план, “Loop Count”). Исследуемые дампы (табл. 2) включали в себя два интервала нормального трафика в начале и в конце и участок ано-

Таблица 1

Средние значения показателя Херста нормального трафика в разные суточные временные отрезки для разных атрибутов трафика

Атрибут трафика				Временной отрезок, ч.
Время между пакетами	Кол-во байт в сек.	Продолжительность TCP сессий	Кол-во новых соединений в сек.	
0.824	0.709	0.613	0.690	9.00
0.798	0.627	0.599	0.716	12.00
0.791	0.717	0.662	0.739	16.00
0.774	0.615	0.620	0.697	18.00
0.778	0.609	0.595	0.721	20.00

Таблица 2

Характеристики исследуемых дампов трафика

Номер образца		1	2
Общая продолжительность, сек		720	1400
Продолжительность аномал. трафика, сек		120	440
Параметры тест. плана JMeter	Кол-во потоков	500	300
	Время набега, сек	30	300
	Кол-во повторений	4	4

Рис 1. Стенд для выполнения исследования

мального трафика между ними, сгенерированный JMeter'ом.

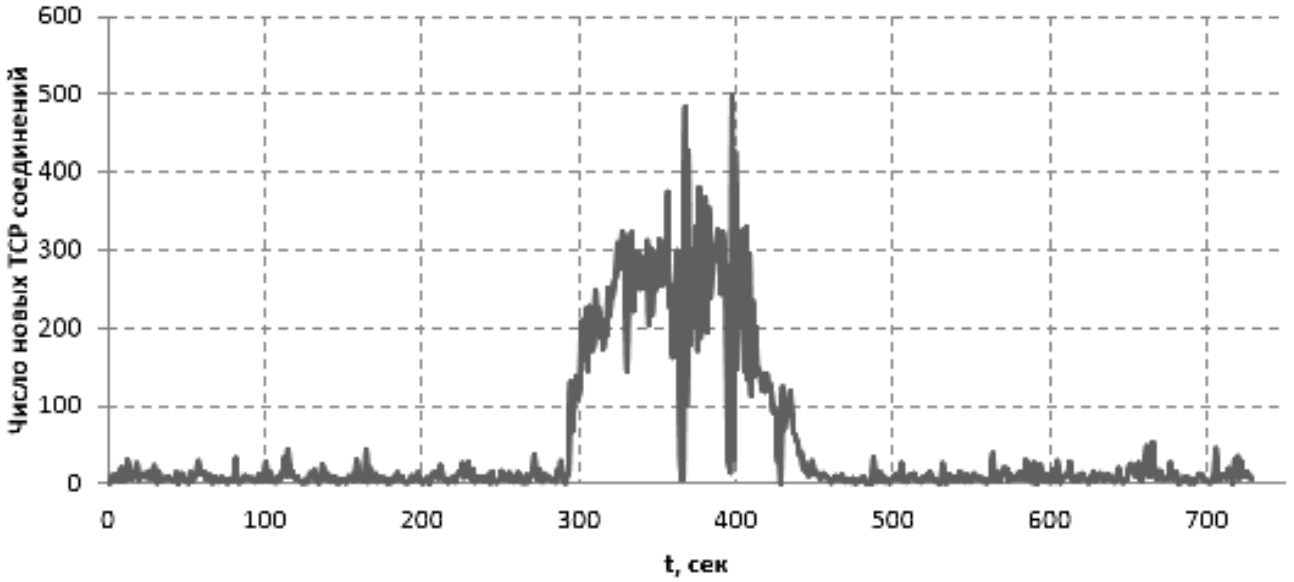
Далее полученный дамп трафика копировался на другой рабочий узел (узел Б на рис. 1) и из него формировалась выборка статистической характеристики, по которой будет проводиться оценка самоподобных свойств. Такой характеристикой стало количество новых TCP соединений, появившихся в трафике раз в 1 секунду и в 1 децисекунду (0.1 секунды). Для исходного дампа оценивалась эта характеристика и формировался массив данных. Коэффициент Херста рассчитывался для отсчетов этого массива, попавших в «окно» заданного размера, которое проходило по всему массиву. Таким образом формировалась выборка коэффициентов, соответствующих разным участкам трафика.

Следует заметить, что в виду особенностей алгоритма минимально допустимый размер окна составляет порядка 70–80 отчетов, и чем отчетов больше, тем точнее рассчитываемое значение показателя Херста. Однако с другой стороны детектор сетевых аномалий должен быть системой реального времени, поэтому желательно получать точные значения показателя Херста для как можно меньшего числа подряд идущих отчетов (собранных за как можно меньший интервал времени). В попытке преодолеть это противоречие, было решено формировать выборку исследуемой стат. характеристики также каждую децисекунду наряду с посекундным формированием. Это позволило применять большие размеры окна для одних и тех же временных интервалов.

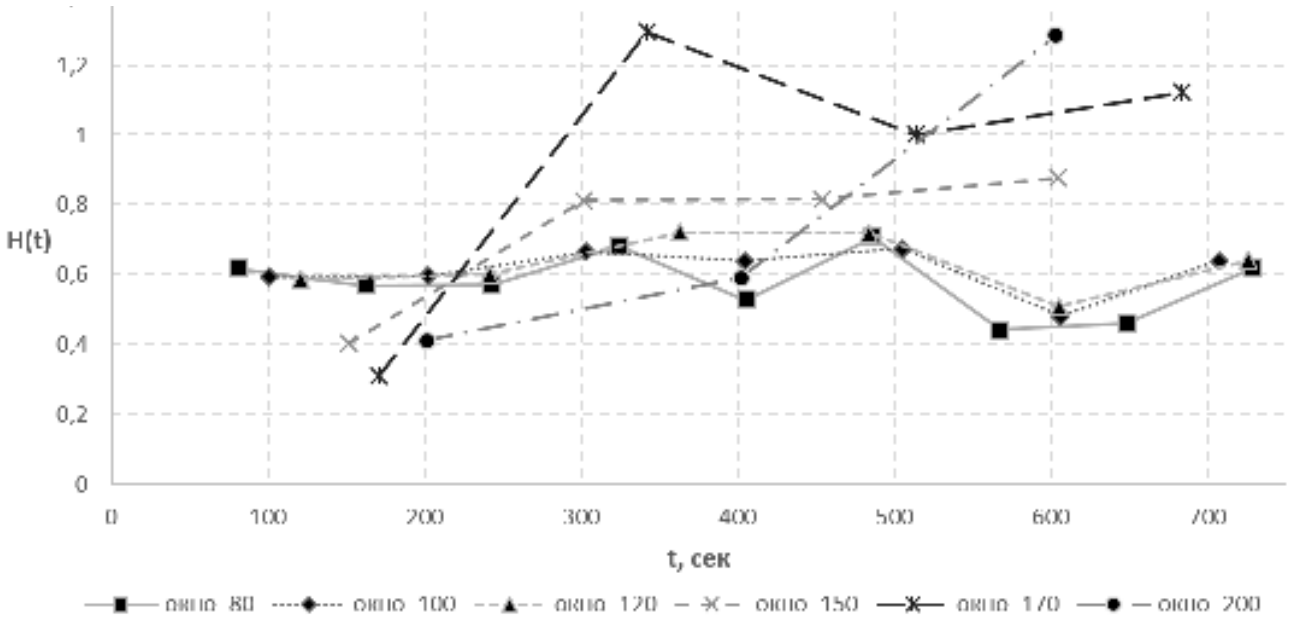
На рис. 2–5 представлена зависимость коэффициентов Херста от времени в зависимости от разных размеров окна для исследуемого дампа.

По представленным графикам можно сделать следующие выводы:

– для посекундной агрегации данных при малых размерах окна (менее 120 отчетов) коэффициент Херста (H) имеет значения меньше 1 на всем временном диапазоне, скачка фрактальной размерности не наблюдается, кривые выглядят сглаженными, без существенных всплесков в момент преобладания аномального трафика (рис. 2, 4, кривые для окон 80, 100, 120); оптимальный размер окна – порядка 150–170 отчетов, в этом случае графики H имеют выраженный всплеск в диапазоне преобладания аномального трафика, причем для образца трафика № 2 наблюдается скачок фрактальной размерности именно в моменты начала и конца атаки (рис. 2, 4, кривые для окон 150, 170); при больших размерах окна может возникнуть ситуация, когда оно больше диапазона аномального трафика (как



а)

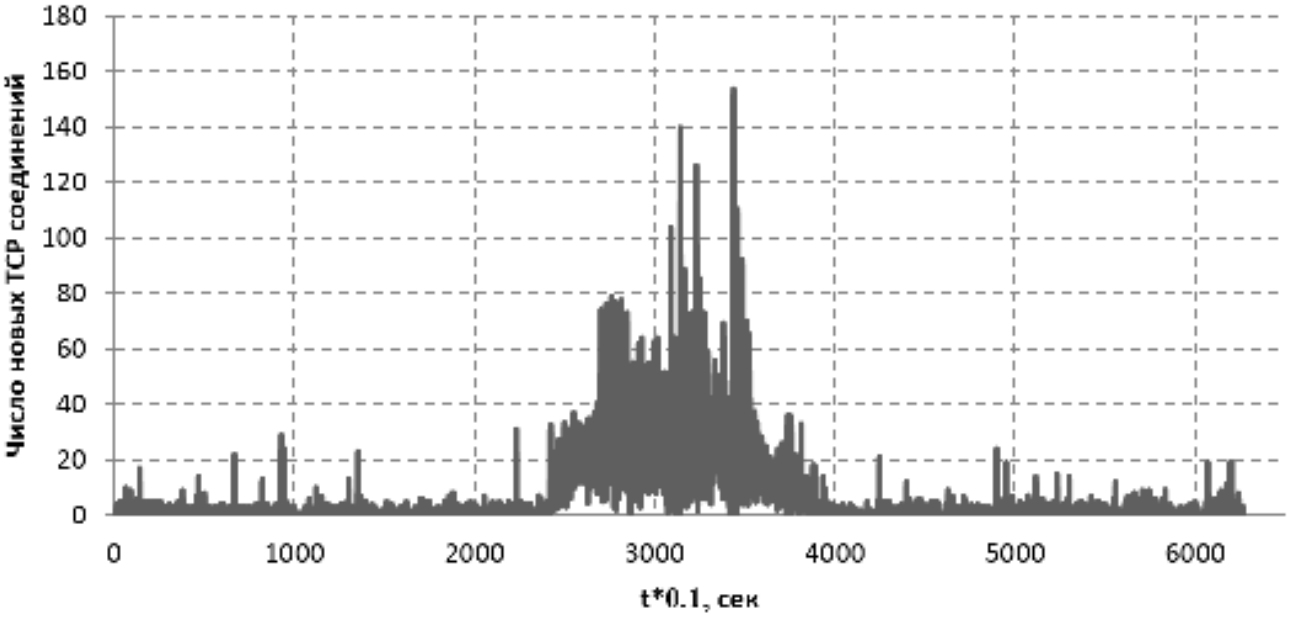


б)

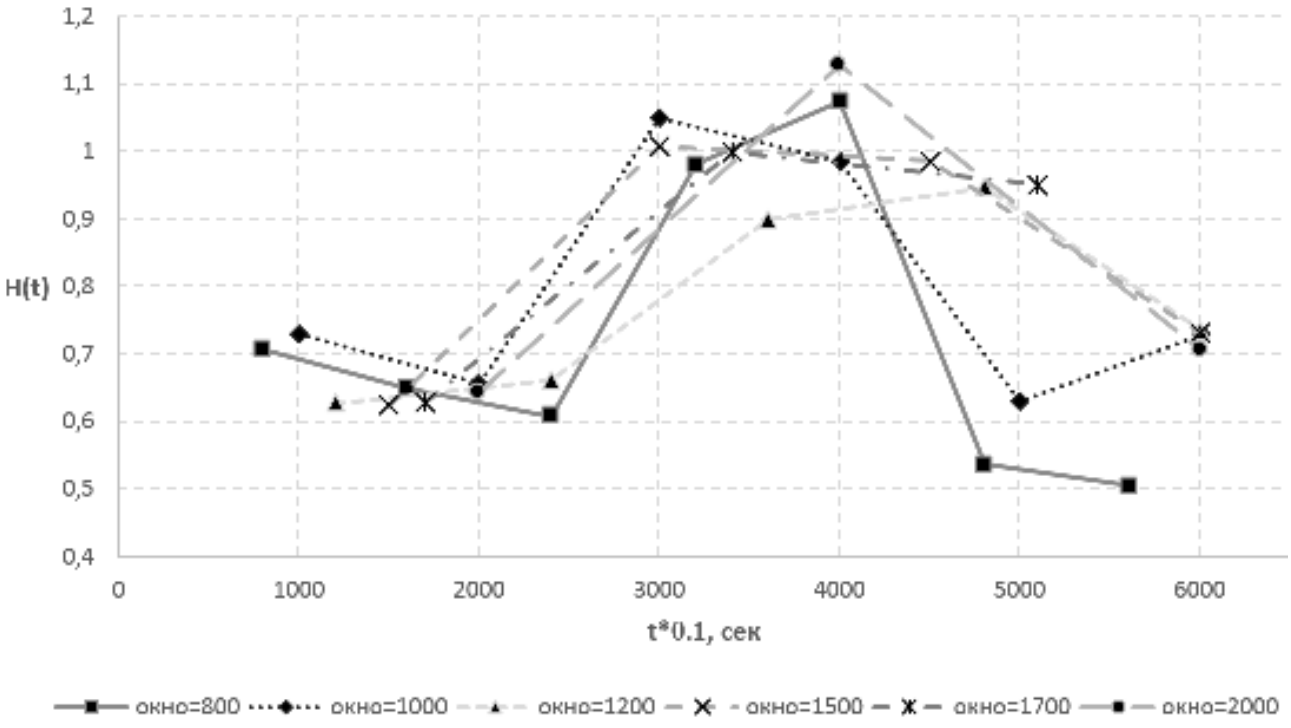
Рис. 2. Образец трафика № 1, агрегация данных раз в 1 секунду: а) Зависимость числа новых TCP соединений от времени; б) Динамика изменений коэффициента Херста с течением времени для разных размеров окна

на рис. 2, кривая для окна 200), в этом случае информативность графика невелика, и по нему нельзя достоверно определить, в какой момент была сетевая атака.

– для агрегации раз в 0.1 секунды при всех рассмотренных размерах окна прослеживаются всплески на графиках H, соответствующие диапазону аномального трафика; оптимальный размер окна – порядка 1200–1700 отчетов, в этом случае динамика изменения H передается наиболее достоверно (рис. 3, 5, кривые для окон 1200, 1500, 1700).

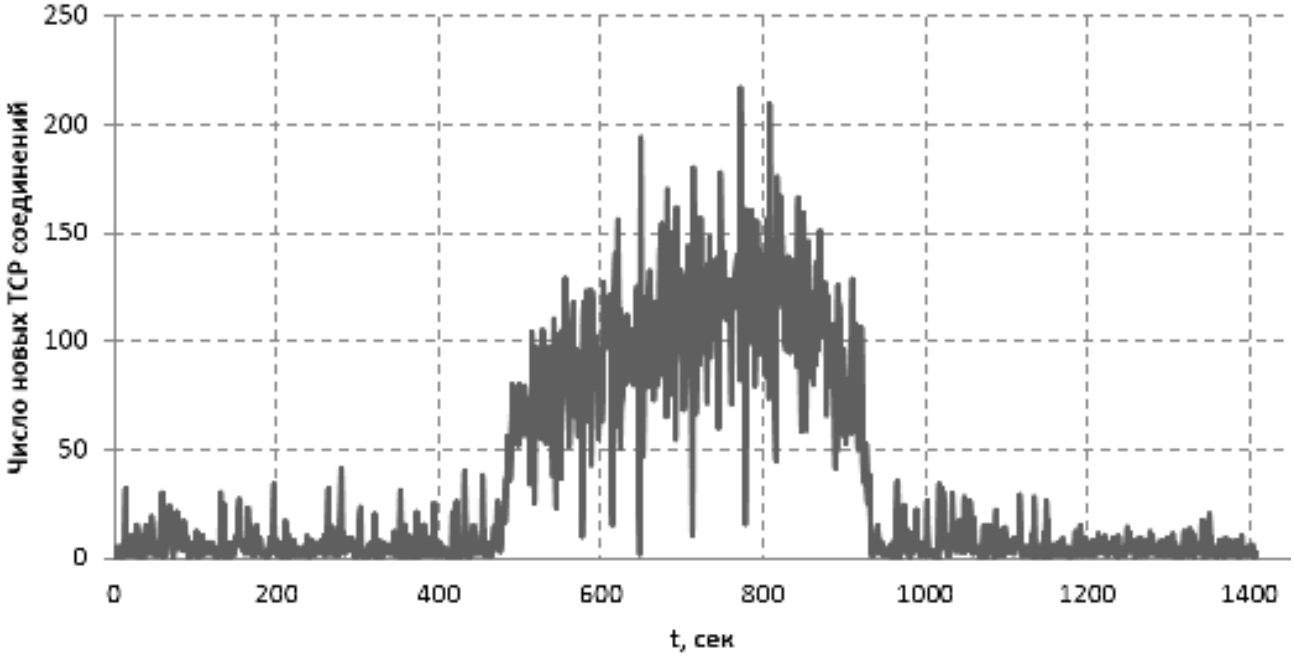


а)

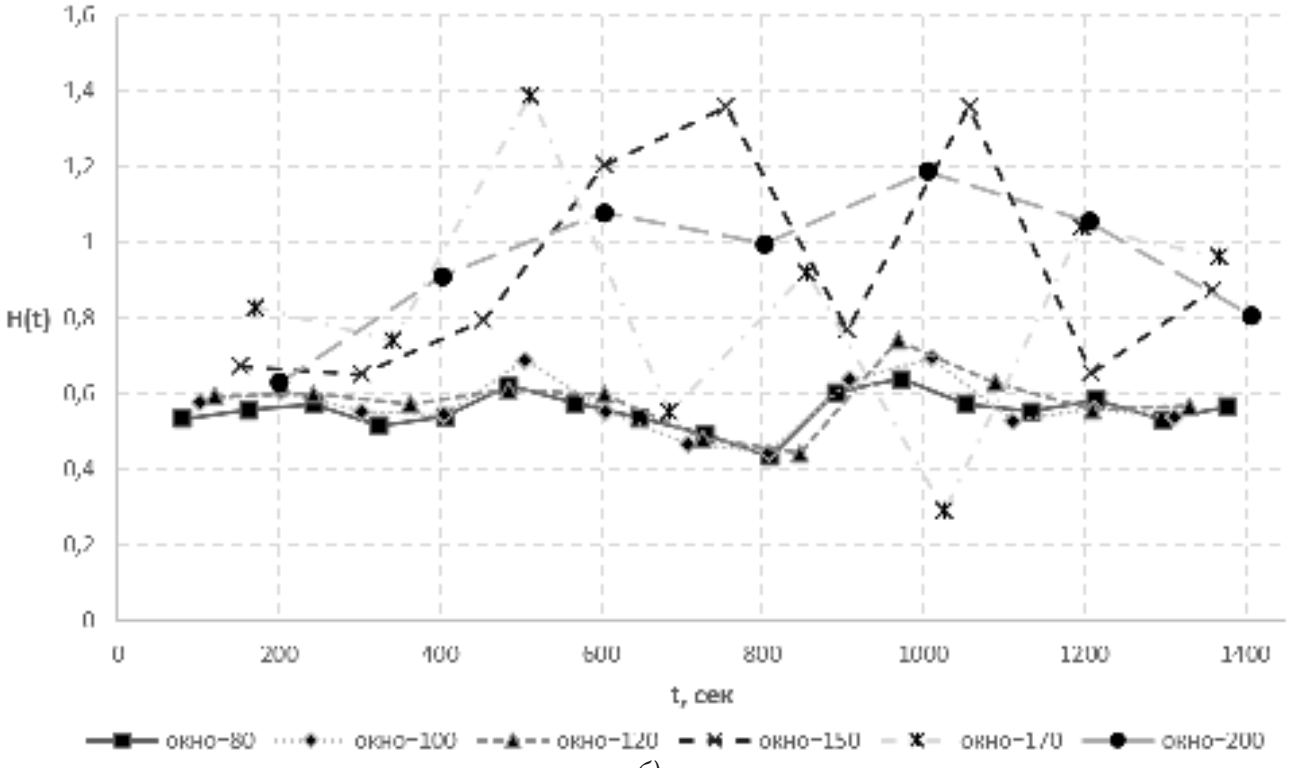


б)

Рис. 3. Образец трафика № 1, агрегация данных раз в 0.1 секунды: а) Зависимость числа новых TCP соединений от времени; б) Динамика изменений коэффициента Херста с течением времени для разных размеров окна

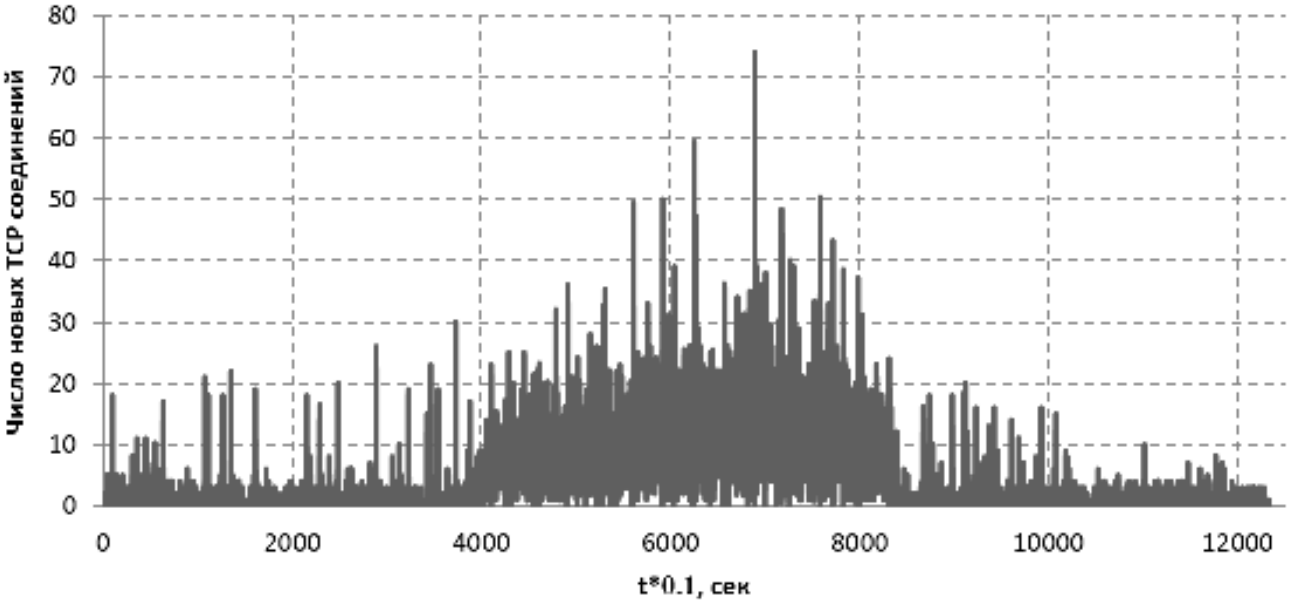


а)

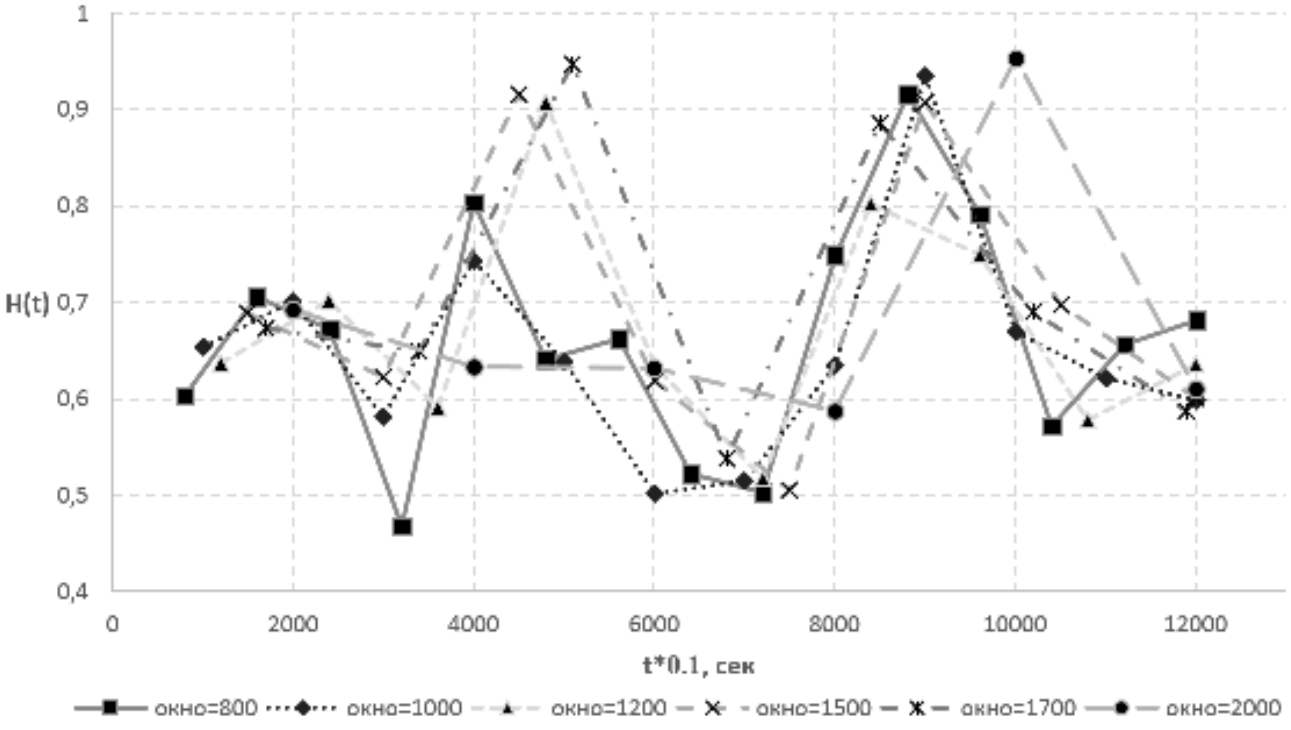


б)

Рис. 4. Образец трафика № 2, агрегация данных раз в 1 секунду: а) Зависимость числа новых TCP соединений от времени; б) Динамика изменений коэффициента Херста с течением времени для разных размеров окна



а)



б)

Рис. 5. Образец трафика № 2, агрегация данных раз в 0.1 секунды: а) Зависимость числа новых TCP соединений от времени; б) Динамика изменений коэффициента Херста с течением времени для разных размеров окна

ЗАКЛЮЧЕНИЕ

В рамках статьи было проведено исследование самоподобных (фрактальных) свойств нормального и аномального сетевого трафика путем оценки коэффициента Херста с помощью метода анализа R/S статистик. Полученные результаты говорят о том, что любому сетевому трафику присущи самоподобные свойства, причем при появлении сетевых аномалий (вызванных например DoS/DDoS атаками или сбоями оборудования) характер этих свойств начинает существенно отличаться от нормального трафика, происходит

скачок фрактальной размерности. Предложен алгоритм по детектированию сетевых аномалий и рассмотрена его работа в рамках реальной сети. В дальнейшем планируется проверка применимости данного алгоритма для детектирования разных типов сетевых DoS атак.

СПИСОК ЛИТЕРАТУРЫ

1. *Барсуков, И. С.* Исследование фрактальных свойств трафика веб-сервера / И. С. Барсуков, М. П. Ряполов // XVIII Международная конференция «Информатика: проблемы, методология, технологии» (Воронеж, 8-9 февраля, 2018). – Воронеж, 2018. – С. 7–12.

2. *Басараб, М. А.* Анализ сетевого трафика корпоративной сети университета методами нелинейной динамики / М. А. Басараб, А. В. Колесников, И. П. Иванов // Науч. издание МГТУ им. Н. Э. Баумана «Наука и образование». – 2013. – № 8. – С. 341–352.

3. *Петерс, Э. Э.* Фрактальный анализ финансовых рынков / Э. Э. Петерс. – М. : Интернет-трейдинг, 2004. – 304 с.

4. *Треногин, Н. Г.* Фрактальные свойства сетевого трафика в клиент-серверной информационной системе / Н. Г. Треногин, Д. Е. Соколов // Вестник НИИ СУВПТ. – 2006. – № 2. – С. 162–173.

5. *Шелухин, О. И.* Анализ изменений фрактальных свойств телекоммуникационного трафика вызванных аномальными вторжениями / О. И. Шелухин, А. А. Антонян // T-Comm. – 2014. – № 6. – С. 61–64.

6. *Шелухин, О. И.* Моделирование информационных систем / О. И. Шелухин, А. М. Тенякшев, А. В. Осин. – М. : Радиотехника, 2005. – 368 с.

7. *Chen, Y.* Filtering of Shrew DDoS Attacks in Frequency Domain / Y. Chen, K. Hwang, Y. Kwok // The IEEE Conference on Local Computer Networks 30th Anniversary (Sydney, Australia, 15–17 November, 2005). – Sydney, 2005. – P. 786–793.

8. *Deka, R. K.* Self-similarity based DDoS attack detection using Hurst parameter / R. K. Deka, D. K. Bhattacharyya // Security and Communication networks. – 2016. – № 5. – P. 4468–4481.

9. *Douligeris, C.* DDoS attacks and defense mechanisms: classification and state-of-the-art / C. Douligeris, A. Mitrokotsa // Computer Networks. – 2003. – №44. – P. 643–666.

10. *Gong, W.* Self-similarity and long range dependence on the internet: a second look at the evidence, origins and implications / W. Gong, Y. Liu, V. Misra, D. Towsley // Computer Networks. – 2005. – № 48. – P. 377–399.

11. *Li, M.* An approach to reliably identifying signs of DDOS flood attacks based on LRD traffic pattern recognition / M. Li // Computers & Security. – 2004. – № 23. – P. 549–558.

12. *Mazurek, M.* Network anomaly detection based on the statistical self-similarity factor for HTTP protocol / M. Mazurek, P. Dymora // Przegląd elektrotechniczny. – 2014. – № 1. – P. 127–130.

13. *Popa, S. M.* Using Traffic Self-Similarity for Network Anomalies Detection / S. M. Popa, G. M. Manea // 20th International Conference on Control Systems and Science (Bucharest, Romania, 27–29 May, 2015). – Bucharest, 2015. – P. 639–644.

Барсуков И. С. – аспирант кафедры электроники, физический факультет, Воронежский государственный университет.
E-mail: barsukov@phys.vsu.ru

Ряполов М. П. – к.ф.-м.н., доцент кафедры электроники, физический факультет, Воронежский государственный университет.
E-mail: ryapolov@phys.vsu.ru

Barsukov I. S. – Postgraduate Student, Department of Electronics, Physics Faculty, Voronezh State University.
E-mail: barsukov@phys.vsu.ru

Ryapolov M. P. – Ph.D. of Physics and Mathematics, Associate Professor at the Department of Electronics, Physics Faculty, Voronezh State University.
E-mail: ryapolov@phys.vsu.ru