АНАЛИЗ СЕТЕВОГО ТРАФИКА С ПОМОЩЬЮ УСТОЙЧИВОГО РАСПРЕДЕЛЕНИЯ

Т.В. Соболева

Белорусский государственный университет
Минск, Республика Беларусь
Soboleva@bsu.by

Рассматриваются вопросы использования устойчивых распределений при анализе сетевого трафика.

Ключевые слова: самоподобие; устойчивое распределение; показатель Херста; сетевой трафик.

NETWORK TRAFFIC ANALYSIS USING SUSTAINABLE DISTRIBUTION

T.V. Soboleva

Belarusian State University
Minsk, Republic of Belarus
E-mail: Soboleva@bsu.by

The issues of using stable distributions in the analysis of network traffic are considered.

Keywords: self-similarity; sustainable distribution; Hurst exponent; network traffic.

Введение

Как известно, при проектировании сетей используют каналы связи с пакетной коммутацией, что увеличивает эффективность использования каналов, однако снижает надёжность доставки информации. Постоянное увеличения объёма передаваемых данных привело к тому, что задача прогнозирования сетевого трафика стала достаточно актуальной. Кроме того, пульсирующий характер и структурные изменения в поведении сетевого трафика могут свидетельствовать о различных атаках на сеть. Таким образом, своевременное обнаружение аномального трафика, позволяет вовремя блокировать возможную атаку.

Последние исследования показывают, что сетевой трафик, имеющий пульсирующий характер, хорошо описывается с помощью устойчивых процессов. Самоподобная структура сетевого трафика оказывает сильное влияние на производительность сети и является характерной особенно-

стью современных телекоммуникационных сетей, что делает учет описанных свойств актуальной задачей.

В статье приведены результаты статистических исследований реального сетевого трафика с помощью устойчивых распределений.

1. Основная часть

Сетевым трафиком (англ. traffic - «движение») называется объем информации, передаваемой через компьютерную сеть за определенный промежуток времени. Количество трафика может быть измерено в пакетах или таких единицах измерения как биты, байты, и их производных. Трафик подразделяется на внешний и внутренний, исходящий и входящий. [1] Мы будем использовать понятие нормального и аномального трафика. Под аномальным трафиком будем подразумевать трафик, не характерный для обычной работы сети. Такой трафик может информировать об осуществлении сетевой атаки.

О самопомоподобной структуре сетевого трафика начали говорить в начале 90-х годов, когда резко увеличился объем передаваемых данных. Ряд статей указывают на то, что объединенный из нескольких источников трафик становится сильно автокоррелированным с долговременной зависимостью [2, 3]. Такое поведение может быть объяснено тем, что будущее процесса определяется его прошлым, причем с убывающей степенью влияния этого прошлого на процесс. Совокупность множества источников данных, проявляющих синдром бесконечной дисперсии, в результате дает самоподобный объединенный сетевой трафик.

Коэффициент Херста [4, 5] является важнейшим параметром, характеризующим степень самоподобия. Этот параметр был назван в честь Х.Е. Херста –британского гидролога, который посвятил себя изучению реки Нил, а также проблеме хранения воды. Оценка параметра помогает не только решить, является ли процесс самоподобным, но и позволяет применить к процессу ряд метод по прогнозированию фрактальных процессов. Коэффициент Херста принимает значения $0 < H < 1$.

• При значении коэффициента $0.5 < H < 1$ говорят персистентном (поддерживающемся) поведении процесса, либо о том, что процесс обладает длительной памятью, то есть является самоподобным. Персистентные стохастические процессы обнаруживают четко выраженные тенденции изменения при относительно малом “шуме”.

• В случае $H = 0.5$ говорят о полностью случайном ряде, аналогичном смещениям частицы при классическом броуновском движении.

• В случае $0 < H < 0.5$ говорят о антиперсистентности процесса. Такой ряд не обладает самоподобием.

Отметим, что существуют различные методы для оценки параметра *H* для того, чтобы выявить самоподобие или медленно убывающую зависимость.

Временные методы оценки параметра Херста:

• Метод R/S статистики (В основе R/S анализа лежит формула Альберта

Эйнштейна о броуновском движении частиц).

• Метод вариаций.

• Метод абсолютного момента.

• Метод отношения вариации остатков.

• Частотные методы оценки параметра Херста.

В работе использовался метод R/S статистики.

Многочисленные исследования сетевого трафика показали, что он лучше всего описывается так называемыми распределениями с “тяжелыми хвостами”. Рассмотрим понятие устойчивого распределения. Устойчивое распределение — это распределение, которое может быть получено как предел по распределению сумм независимых случайных величин [6].

Пусть $\xi, \xi_1, \xi_2, \ldots$ — независимые, одинаково распределённые случайные величины.

Определение 1. Случайная величина ξ называется устойчивой (устойчивой в широком смысле), если для каждого m, $m = 1, 2, \ldots$, существуют такие постоянные $c_m = m^{1/\alpha}$, $\alpha \in (0,2]$, и $\gamma_m \in R$, что

$$\sum_{i=1}^{m} \xi_i \overset{d}{=} c_m \xi + \gamma_m,$$

где «$\overset{d}{=}$» означает равенство по распределению, $R = (-\infty, +\infty)$.

При исследовании устойчивых распределений удобно использовать аппарат характеристических функций. Для того, чтобы случайная величина ξ была устойчивой, необходимо и достаточно, чтобы логарифм её характеристической функции $\varphi_\xi(t)$ имел представление

$$\ln \varphi_\xi(t) = i\mu t - \sigma^\alpha |t|^\alpha + i\sigma^\alpha t \omega(t, \alpha, \beta),$$

где $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma > 0$, $\mu \in R$, $t \in R$,

$$\omega(t, \alpha, \beta) = \begin{cases} |t|^{\alpha-1} \beta tg \dfrac{\pi}{2} \alpha, & \alpha \neq 1, \\ -\beta \dfrac{2}{\pi} \ln|t|, & \alpha = 1. \end{cases}$$

Таким образом, класс устойчивых распределений представляет собой четырёхпараметрическое семейство с параметрами: α — характеристическим показателем, β — параметром асимметрии, σ — параметром масштаба, μ — параметром положения.

Данные реального сетевого трафика при больших объемах обладают свойством самоподобия, поэтому классические модели не могут быть использованы для моделирования. Для моделирования такого трафика используется устойчивое распределение.

Для исследований сетевого трафика, использовались выборки из наблюдений за трафиком организации в течении одного дня. Файлы с данными были предобработаны и подготовлены для последующего моделирования и исследования.

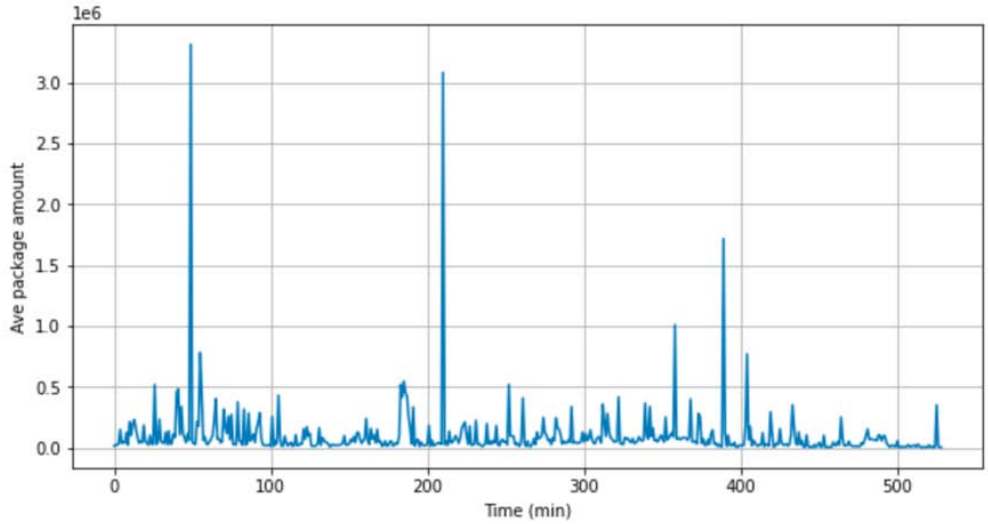
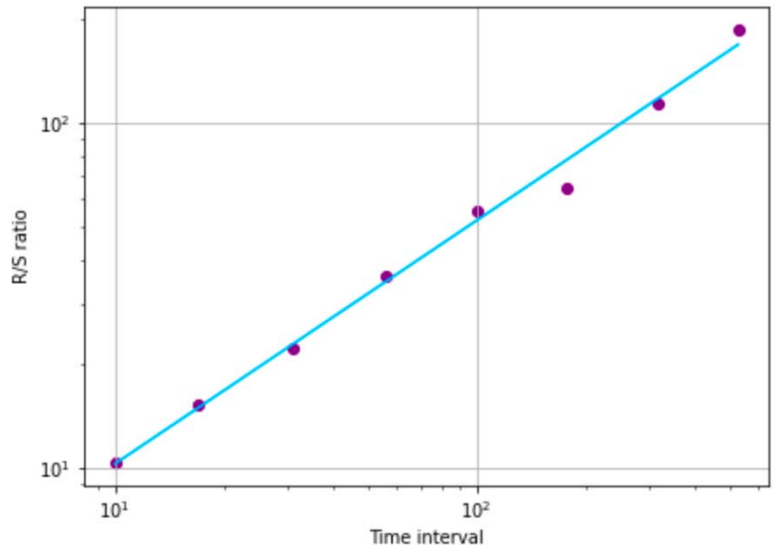


Рис.1. – График количества пакетов в зависимости от момента времени

Из рисунка 1 видно, что у нас присутствуют скачки активности, которые выходят за обычные значения.

Для оценки показателя Херста, использовался пакет «hurst» языка Python.



H=0.7046, c=2.0317

Рис.2. – Результаты R/S анализа исходного временного ряда

Как видно из рисунка 2, показатель Херста больше граничного значения 0.5, следовательно, временной ряд обладает свойством самоподобия.

Заключение

Итак, результаты исследований показали, что временной ряд, полученный в ходе преобразований исходных данных, собранных в течение одного дня обладает характеристикой самоподобия и может быть смоделирован с помощью устойчивого распределения.

Библиографические ссылки

1. Lucas M. W. Network flow analysis. N.-Y, 1990. Ch. 1. P. 9–11.
2. 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. № 6. P. 986–995.
3. 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 Transcriptions on Networking. Feb. 1997. Vol. 5. № 1. P. 71–86.
4. Anis A.A., Lloyd E.H. The expected value of the adjusted rescaled Hurst range of independent normal summands // Biometrica. 1976. № 63. P. 283–298.
5. Simmross-Wattenberg F., Anomaly Detection in Network Traffic Based on Statistical Inference and alpha-Stable Modeling // IEEE Transactions on Dependable and Secure Computing Jul. 2001. Vol. 8. № 4. P.494–509.
6. Труш Н. Н., Соболева Т.В. Статистический анализ оценок спектральных плотностей устойчивых случайных процессов. Минск: БГУ, 2008. 100 с.