

# СТАТИСТИЧЕСКИЙ АНАЛИЗ СЕТЕВОГО ТРАФИКА

**И. А. Лобан**

В последние годы исследования, связанные с анализом сетевого трафика показывают, что он обладает свойством масштабной инвариантности, то есть обладают свойством самоподобия [1–3].

В работе рассматриваются задачи исследования структуры сетевого трафика, направленные на изучение его характерных особенностей.

Одним из основных методов формирования самоподобного потока является метод, первоначально предложенный Мандельбротом. В данном методе предусматривается наличие нескольких независимых ON/OFF источников. Для каждого источника эти периоды строго чередуются. Длительности ON (как и OFF) периодов также независимы и одинаково распределены, причем распределение длительностей ON периодов может отличаться от распределения для OFF периодов. Каждый источник генерирует пакеты только в состоянии ON. Результирующим значением в каждый период времени является сумма величин генерируемых всеми источниками.

Появление самоподобия объясняется наличием эффекта Ноа (Noah effect) в распределении длительностей ON/OFF периодов. Для достижения этого эффекта можно использовать распределение Парето, имеющего функцию распределения,

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^\alpha$$

либо логарифмически-нормальное распределение (heavy-tailed distributions). Более популярно распределение Парето. Параметр  $\alpha$  – параметр формы, определяющий, конечность или бесконечность среднего значения и дисперсии для распределения, а параметр  $\beta$  задает минимальное значение величины  $x$ . Параметр  $\alpha$  задает среднее значение и дисперсию следующим образом:

1. для  $0 < \alpha \leq 1$  распределение имеет бесконечное математическое ожидание (МО) и дисперсию;
2. для  $1 < \alpha \leq 2$  распределение имеет конечное МО и бесконечную дисперсию;
3. для  $\alpha > 2$  распределение имеет конечное МО и дисперсию.

Существует связь между показателем Херста (Hurst exponential) и параметром  $\alpha$ :

$$H = \frac{3 - \alpha}{2}$$

Параметр  $\alpha$  называется фрактальным показателем временного ряда.

Благодаря простоте задания параметра Херста метод Мандельброта получил наибольшее распространение при моделировании самоподобные случайных процессов. Указанный метод позволяет получить реализации самоподобного процесса с любой фиксированной степенью самоподобия.

Для изучения фрактальных свойств сетевого трафика используются его реализации, полученные в университете города Наполи. Замеры проводились каждые 10 мс, их общее число составляет более 10000. Изменялись такие характеристики сетевого трафика как уровень временной задержки (delay) и джиттера (jitter) – это нежелательные фазовые или частотные случайные отклонения передаваемого сигнала. Пакеты различной длины (по 64, 128, 256, 512, 1024 байт) передавались по протоколам UDP и TCP в сетях на основе технологий ADSL и Wireless.

Автокорреляционная функция (АКФ) – это характеристика сигнала, которая помогает находить повторяющиеся участки сигнала или определять несущую частоту сигнала, скрытую из-за наложений шума и колебаний на других частотах.

Случайный процесс обладает медленно убывающей зависимостью (МУЗ) [1], если для АКФ выполняется условие:

$$r(k) \sim k^{-\beta} L(k) + C, k \rightarrow +\infty$$

где  $C = const$ ,  $0 < \beta < 1$ ,  $L(x)$  – медленно меняющаяся на бесконечности функция. Процессы с МУЗ характеризуются АКФ, которая убывает по степенному закону при увеличении временной задержки. Для процессов с быстро убывающей зависимостью (БУЗ) значения АКФ убывают по экспоненциальному закону.

Известно, что самоподобные процессы обладают медленно убывающей зависимостью, но в тоже время не каждый случайный процесс с МУЗ является самоподобным.

Для определения, какие ряды являются фрактальными, а какие случайными, проведем вычисление показателя Херста (Hurst exponent). Этот показатель является мерой МУЗ [3]: если показатель Херста близок к 0,5, то это говорит о том, что временной ряд является случайным, а не фрактальным; чем ближе  $H$  к единице, тем сильнее в ряду проявляются фрактальные свойства; такой ряд является трендоустойчивым и тенден-

ция его изменения может быть спрогнозирована; приближение показателя Херста к нулю говорит о том, что временной ряд состоит из частых спадов и подъемов, такие ряды называют антиперсистентными.

Для вычисления показателя Херста используем метод R/S анализа временных рядов [3].

1. В исходном ряду выделяем 20 отрезков (от 1 до  $\tau = \frac{iN}{20}$ ,  $i = \overline{1,20}$ ).

2. В каждом из них вычисляем среднее значение, накопленные отклонения для отрезков длины  $t$  по формуле:  $X_{\tau,t} = \sum_{i=1}^t (X_i - \bar{X}_\tau)$ ,  $t = \overline{1,\tau}$ . Затем вычисляем размах  $R$  между максимальным и минимальным накопленными отклонениями.

3. Также для каждого отрезка вычисляем стандартное отклонение  $S$  (квадратный корень из выборочной дисперсии).

4. Показатель Херста  $H$  вычисляется из эмпирической формулы  $\frac{R}{S} = \left(\frac{\tau}{2}\right)^H$ , путем логарифмирования обеих частей. По 20 точкам с координатами  $(\lg \frac{\tau}{2}, \lg \frac{R}{S})$  методом наименьших квадратов строим тренд с линейной зависимостью. Таким образом, определяем значение  $H$  для временного ряда.

В результате работы было разработано приложение, которое позволяет: строить графики реализаций сетевого трафика с поддержкой масштабируемости и возможностью сохранения в формате PNG. Для этого используется сторонняя свободная библиотека JFreeChart (см. рис. 1); строить графики автокорреляционной функции для временных рядов; вычислять показатель Херста и другие связанные с ним фрактальные параметры временных рядов, такие как фрактальная размерность  $D = 2 - H$ , корреляционный параметр  $\beta = 2(1 - H)$ , спектральный показатель  $b = 2H + 1$ , фрактальный показатель  $\alpha = 3 - 2H$  (см. рис. 2).

Приложение имеет модульную структуру. Каждый модуль выполняет отведенную ему функцию в отдельном окне.

Исходный код проекта Net Traffic Discovery доступен на одном из крупнейших веб-сервисов для хостинга проектов – GitHub (<https://github.com/igor-loban/net-traffic-discovery>).



Рис. 1. Масштабируемые графики

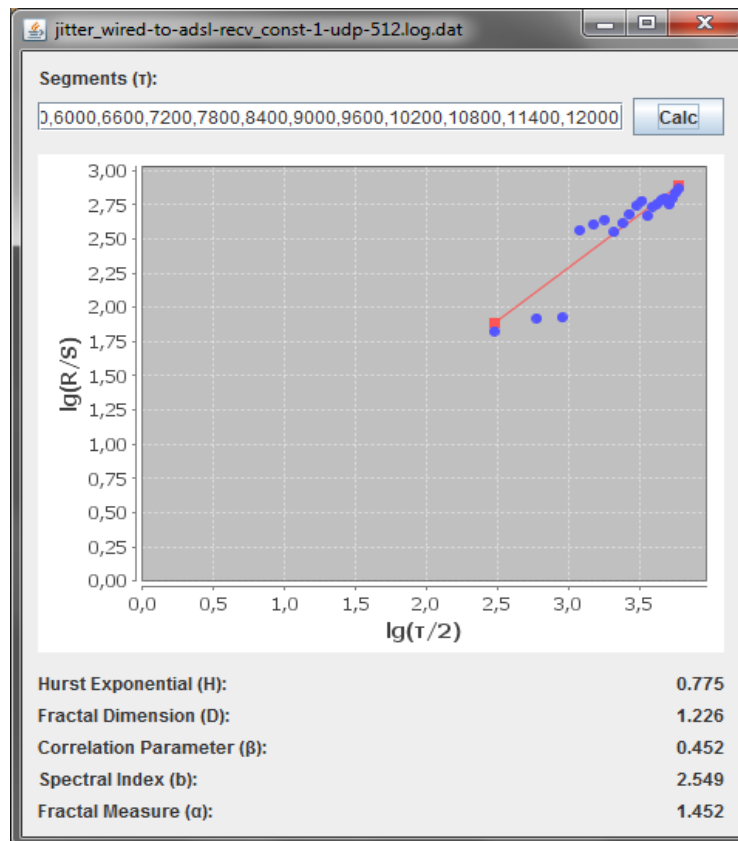


Рис. 2. Вычисление показателя Херста и других фрактальных показателей

### Литература

1. Бельков Д. В. Статистический анализ сетевого трафика / Д. В. Бельков, Е. Н. Едемская, Л. В. Незамова // Науч. тр. ДонНТУ / ДонНТУ. Донецк, 2011. Вып. 13: Информатика, кибернетика и вычислительная техника. С. 66–75.

2. Гребенников А. В. Моделирование сетевого трафика и прогнозирование с помощью модели ARIMA / А.В. Гребенников, Ю.А. Крюков, Д. В. Чернягин // Электронный журнал «Системный анализ в науке и образовании» 2011. № 1. С. 1–11.
3. Numerical Method [Electronic resource] Mode of access: <http://numericalmethod.com/blog/>. Date of access: 13.04.2013.

## **СИСТЕМА ПОИСКА ПЛАГИАТА**

**А. И. Логинов**

### **ВВЕДЕНИЕ**

В настоящее время большинство служб поиска плагиата ищут его исключительно в обычном художественном тексте. В то же время исходные тексты программ во многом остаются без внимания. Конечно, существуют и такие сервисы. Вот только практически все они являются либо платными, либо уже давно не поддерживаются, либо никогда не выкладывались в общий доступ.

Именно поэтому целью данной работы стало создание такой системы поиска плагиата, которая бы никогда не потеряла своей актуальности, системы, которую можно было бы использовать даже в случае прекращения поддержки. Эта универсальная программа должна искать плагиат в исходных кодах на разных языках программирования с возможностью расширения их количества.

### **1 ЯП. РАСШИРЯЕМОСТЬ**

#### **1.1 Универсальная система**

Одной из ключевых особенностей данной системы является токенизация исходного текста программ. Тем самым, задача расширяемости свелась к динамическому способу добавления токенизаторов для различных языков программирования.

#### **1.2 Генерация парсеров**

Генераторы парсеров позволяют программисту работать на более высоком уровне абстракции. Создание анализатора языка в этом случае включает в себя написание грамматики. Очевидно, что поддерживать несколько сотен строк грамматики намного проще, чем несколько тысяч строк парсера, написанного вручную.