ИССЛЕДОВАНИЕ СЕТЕВОГО ТРАФИКА

Бельков Д.В.

До недавнего времени теоретическую базу для проектирования систем распределения информации составляла теория массового обслуживания. Моделью потока вызовов (данных) в этой теории является простейший поток (стационарный ординарный поток без последействия). В 1993 году группа американских исследователей: W.Leland, M.Taqqu, W.Willinger и D.Wilson опубликовали результаты работы, которая в корне изменила представления о процессах, происходящих в телекоммуникационных сетях с коммутацией пакетов. Оказалось, что потоки в современных сетях нельзя аппроксимировать простейшими, поскольку они имеют иную структуру, чем принято в классической теории телетрафика. Было установлено, что трафик сети обладает свойством самоподобия (масштабной инвариантности), имеет память (последействие), а также обладает высокой пульсацией. По этой причине расчет параметров системы распределения информации, предназначенной для обработки сетевого трафика, по классическим формулам дает некорректные, неоправданно оптимистические результаты. Алгоритмы обработки трафика, созданные для работы с простейшим потоком неэффективны для фрактальных потоков с самоподобием. Статистические характеристики (среднее значение, спектральная плотность, автокорреляционная функция и др.) самоподобного трафика имеют характер спада сильно отличающийся от экспоненциального. Поэтому требуют корректировки исходные предпосылки, которые делались ранее при разработке многих сетевых устройств.

Несмотря на продолжительный период изучения проблемы самоподобия телетрафика, остается ряд нерешенных задач:
- фактически отсутствует строгая теоретическая база, которая пришла бы на смену классической теории массового обслуживания при проектировании современных систем распределения информации с самоподобным трафиком;
- нет единой общепризнанной модели самоподобного трафика;
- не существует достоверной и признанной методики расчета параметров и показателей качества систем распределения информации при влиянии эффекта самоподобия;
- отсутствуют алгоритмы и механизмы, обеспечивающие качество обслуживания в условиях самоподобного трафика [1].

Решение указанных задач имеет не только теоретическое, но и практическое значение. Большинство современных приложений являются синхронными и предъявляет высокие требования к качеству соединения. Сократить задержку передачи данных по сравнению с протоколом TCP, позволяет протокол без гарантированной доставки UDP. Однако, обеспечить повышенные требования к качеству соединения только с помощью транспортного протокола (UDP или TCP) затруднительно, поскольку причины, приводящие к большим задержкам, большей частью находятся на сетевом уровне [2].

Свойство масштабной инвариантности сетевого трафика позволяет разработать алгоритмы прогнозирования, которые смогут посредством анализа трафика на относительно небольшом отрезке времени предсказать его поведение на более длительных интервалах. Используя такие прогнозы, можно будет создавать более эффективные методы управления пропускной способностью, что позволит сократить задержки передачи данных по сети и потери пакетов.

Ситуация, сложившаяся в современных глобальных компьютерных сетях, наличие большого количества сетевых маршрутов на которых периодически возникают резкие колебания задержки в передаче данных и большой процент потерь пакетов, появление новых свойств сетевого трафика, необходимость обеспечения высокого качества обслуживания различных категорий приложений, делают актуальной задачу исследования фрактального трафика.

Целью данной работы является определение фрактальных характеристик (H и D) временного ряда методами агрегирования, которые могут быть использованы при исследовании сетевого трафика. Их применение показано на конкретном примере.

Задачи работы:

1. Сформулировать свойства фрактальных процессов;
2. Определить степень фрактальности трафика методом агрегирования;
3. Определить фрактальную размерность трафика методом агрегирования;
4. Определить вероятностное распределение трафика.

Математическое описание дискретного фрактального процесса

Существует два класса фрактальных процессов, так называемые точно самоподобные и асимптотически самоподобные процессы. Процесс X называется точно самоподобным с параметром β (0<β<1), если выполняются следующие условия:

1. $D_m = D/m^{\beta}$, D - дисперсия процесса X, D_m - дисперсия агрегированного процесса $X^{(m)}$, полученного уменьшением размера шкалы наблюдений X в m раз.
2. Автокорреляционная функция (АКФ) сохраняется на всех масштабах: $R(k,X^{(m)}) = R(k,X)$.

Процесс X называется асимптотически самоподобным если для больших k выполняются условия:

1. $D_m = D/m^{\beta}$, D - дисперсия процесса X, D_m - дисперсия агрегированного процесса $X^{(m)}$, полученного уменьшением размера шкалы наблюдений X в m раз. Параметр β связан с параметром Херста H соотношением $\beta = 2(1-H)$.
2. Автокорреляционная функция (АКФ) сохраняется при $m \to \infty$: $R(k,X^{(m)}) \to R(k,X)$.

Наиболее точным свойством самоподобных процессов является то, что АКФ не вырождается при $m \to \infty$, в отличие от стохастических процессов, где $R(k,X) \to 0$ при $m \to \infty$.

Метод агрегирования-1

Пусть исходный ряд показан на рисунке 1.

Рисунок 1.- Исходный ряд

Для него осуществлен следующий агрегационный процесс. Выполнено уменьшение размера шкалы наблюдений в 2 раза. Для этого сформирован новый ряд, полученный при помощи операции нахождения среднего каждых двух последовательных исходных наблюдений. Полученный ряд состоит из 15 событий. Произошло уменьшение рассматриваемой шкалы в 2 раза: каждое единичное деление новой шкалы содержит 2 единицы исходной. Затем аналогично выполнено уменьшение размера исходной шкалы наблюдений в m раз, для m=3, m=5, m=6 и m=10. Каждое деление новой шкалы содержит m единиц исходной. Структура полученных рядов подобна структуре исходного ряда.

Согласно определению самоподобного процесса, имеет место следующее соотношение дисперсий временных рядов:

$$D_{X^m} = \frac{D_X}{m^{\beta}} \qquad (1)$$

Логарифмируя выражение (1), получим:

$$ln(D_{X^m}) = ln(D_X) - \beta \cdot ln(m) \qquad (2)$$

Поскольку $ln(D_X)$ является константой, не зависящей от m, то график зависимости $ln(D_X)$ от $ln(m)$ представляет собой прямую с наклоном, равным $(-\beta)$. Построив график зависимости (2) и линию тренда, как показано на рисунке 2, определим аппроксимированное значение β: $\beta = 0{,}3813 \approx 0{,}38$. Учитывая, что параметр β связан с показателем Херста H как $H = 1 - \frac{\beta}{2}$, получим значение H: H=0,81. Фрактальная размерность D временного ряда в таком случае: $D = 2 - H = 2 - 0{,}81 = 1{,}19$. Поскольку H>0,5, то степень устойчивости долгосрочной зависимости исследуемого временного ряда выше среднего и ряд является самоподобным (фрактальным).

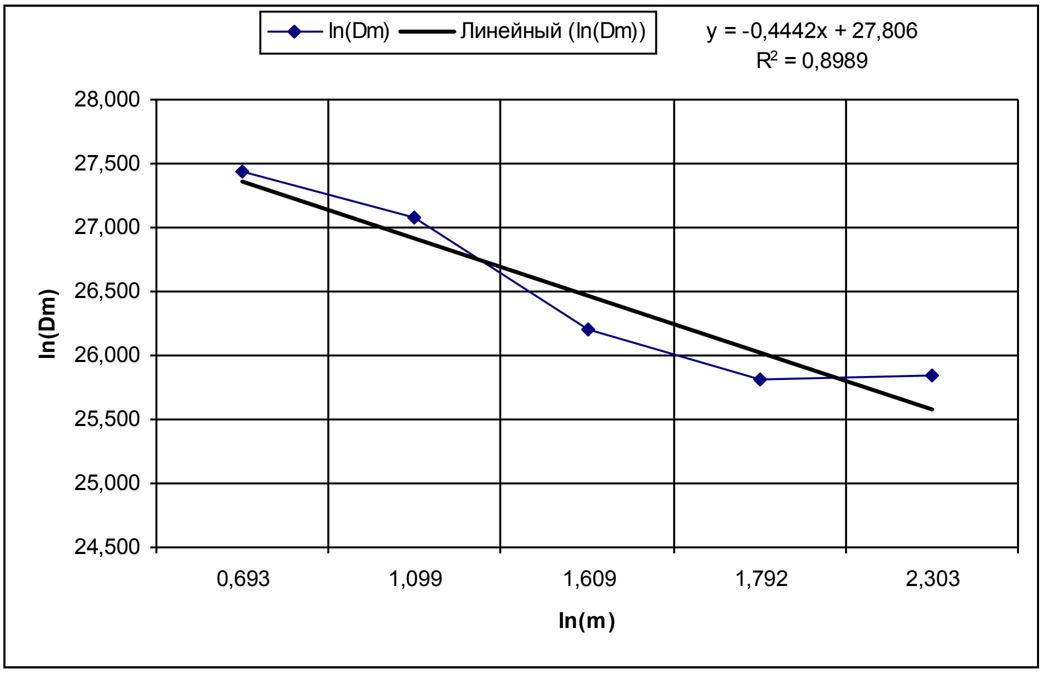

Рисунок 2.- Линия тренда для определения β

Метод агрегирования-2

Как и в предыдущем разделе, получим агрегированные временные ряды, в которых одно деление новой шкалы содержит m единиц исходной. Для каждого временного ряда вычислим коэффициенты вариации по формуле (3), где D_m - дисперсия, μ - среднее значение, одинаковое для всех рядов.

$$CV_m = \frac{(D_m)^{0,5}}{\mu} \qquad (3)$$

График зависимости $ln(CV_m)$ от $ln(m)$ представляет собой прямую с наклоном, равным $(-\gamma)$. Фрактальная размерность D временного ряда равна $D = 1 + \gamma$. Построив график зависимости и линию тренда, как показано на рисунке 3, определим аппроксимированное значение γ: $\gamma = 0,194$. Следовательно, $D = 1 + 0,194 = 1,194$ и H=2-D=0,806. Это примерно соответствует результатам, полученным в предыдущем разделе.

Оценка тяжести хвоста вероятностного распределения

Тяжелый хвост вероятностного распределения случайной величины может быть свидетельством фрактальности временного ряда. Распределение имеет тяжелый хвост, если выполняется условие (4):

$$P[X > x] \sim x^{-\alpha}, 0 < \alpha < 2 \qquad (4)$$

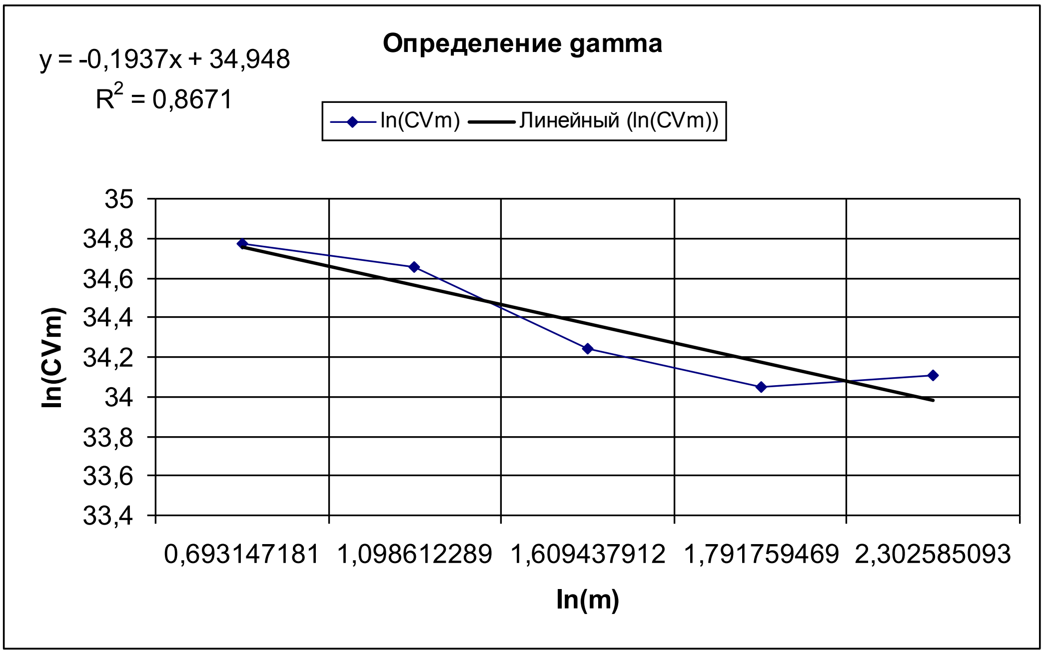

Рисунок 3.- Определение γ

Простейшим распределением с тяжелым хвостом является распределение Парето, для которого функция плотности распределения имеет вид $p(x) = \alpha k^{-\alpha-1}, \alpha, k > 0, x \ge k$ и функция распределения $F(x) = P[X \le x] = 1 - \left(\frac{k}{x}\right)^{\alpha}$.

Чтобы оценить тяжесть хвоста для имеющихся данных, разделим диапазон данных на 10 непересекающихся интервалов, вычислим частоты попадания в каждый интервал, вычислим функции распределения $F(x)$ и $1 - F(x)$. График дополнительной функции распределения $1 - F(x)$ в логарифмической шкале показан на рисунке 4.

Построив линию тренда, как показано на рисунке 5, получим тангенс угла ее наклона к горизонтальной оси. Он является оценкой тяжести хвоста распределения и равен $-\alpha$. В данном случае α принимает значение равное 1,354, попадает в промежуток от 0 до 2, следовательно, распределение имеет свойство тяжелого хвоста. Показатель Херста H связан с α по формуле $H = \frac{3-\alpha}{2}$. Вычислив H, получим $H = \frac{3-1,354}{2} = 0,823$, что примерно соответствует результатам, полученным в предыдущих разделах.

Выводы

Теория фракталов служит базой для количественного описания диссипативных структур, формирующихся в условиях далеких от равновесных. Такие структуры формируются в компьютерных сетях при передаче информационных потоков. Рассмотренные в статье методы определения фрактальных характеристик динамических рядов могут быть использованы для оценки фрактальности сетевого трафика.

Рисунок 4.-Тяжелый хвост

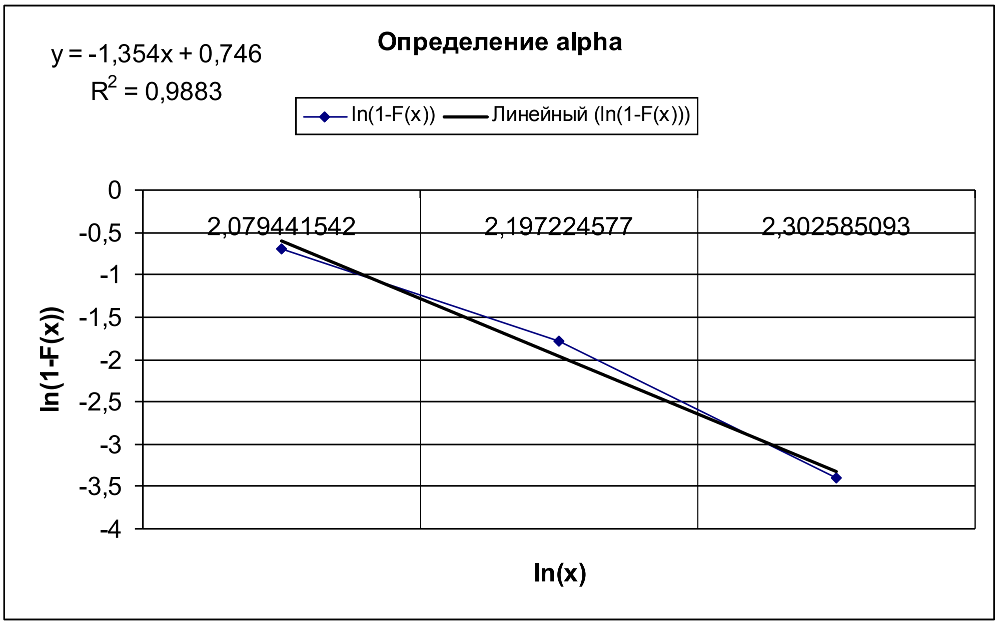

Рисунок 5.- Определение α

Литература

1. Петров В.В. Структура телетрафика и алгоритм обеспечения качества обслуживания при влиянии эффекта самоподобия. Автореферат диссертации. Москва. – 2005. – 20 с.
2. Иванов А.В. Разработка и исследование алгоритмов прогнозирования и управления очередями в компьютерных сетях. Автореферат диссертации. Санкт-Петербург. – 2001. – 18 с.