УДК 519.216

ОЦЕНКА ПАРАМЕТРОВ НЕОДНОРОДНОГО ТРАФИКА

Галактионова О.В., Хохлов Ю.С.
Кафедра математической статистики и системного анализа

Поступила в редакцию 01.07.2010, после переработки 10.09.2010.

Мы рассматриваем задачу оценки параметров входящего потока телекоммуникационной системы. Предполагается, что поток есть композиция нескольких независимых асимптотически самоподобных гауссовских процессов (например, дробных броуновских движений). На практике очень важно уметь оценивать параметры Херста компонент процесса (или, что эквивалентно, параметры долговременной зависимости). Предлагаемый метод заключается в использовании последовательности вейвлет-коэффициентов процесса и применении к ней регрессионного анализа.

We consider the problem of parameter estimation of input flow of telecommunication system. It is assumed that the flow is the composition of several independent asymptotically selfsimilar gaussian processes (for example, fractional brownian motions). Very important practical problem is the estimation of Hurst parameters of the components of the process (equivalently, the parameters of long-range dependence). The main idea is the using the set of wavelet coefficients of the process and then applying the regression analysis.

Ключевые слова: долговременная зависимость, дробный гауссовский шум, оценка параметров, телекоммуникационные сети.
Keywords: long-range dependence, fractional brownian noise, parameter estimation, telecommunications networks.

1. Введение

В 90-х годах XX века были проведены исследования свойств интернет-трафика корпорации Bellcore (Bellcore Morristown Reserch and Engineering Centre) при различных условиях. Установлено, что трафик обладает свойством масштабируемости. На основе этих результатов были сделаны заключения о том, что при описании трафика основную роль играют три параметра ([4]):
- скорость;
- параметр Херста;
- параметр пиковости.

Этого достаточно для решения таких технических вопросов, как, например, конструирование трафика, предотвращение возможных перегрузок в сети, оптимизация использования сетевых ресурсов, определение параметров качества обслуживания. Поэтому параметр Херста занимает центральное место в описании самоподобного трафика.

Необходимо отметить, что самоподобные процессы, как правило, обладают свойством долговременной зависимости (ДВЗ). Напомним, что если $X(t)$ является стационарным стохастическим процессом с автоковариационной функцией $\gamma_X(\tau)$ и спектральной плотностью $f_X(\nu)$, тогда это свойство означает что

$$\gamma_X(\tau) \sim c_\gamma |\tau|^{-(1-\alpha)}, \tau \to \infty. \tag{1}$$

В терминах спектральной плотности это эквивалентно тому, что

$$f_X(\nu) \sim c_f |\nu|^{-\alpha}, \nu \to \infty \tag{2}$$

при $\alpha \in (0,1)$. Функция $f_X(\nu)$ в случае процесса с дискретным временем удовлетворяет соотношению

$$\gamma_X(0) = \sigma_X^2 = \int_{-1/2}^{1/2} f_X(\nu) d\nu,$$

где σ_X^2 - дисперсия (или мощность) X_t.

Каждое из этих определений включает два параметра (α, c_γ) и (α, c_f), причем

$$c_f = 2(2\pi)^{-\alpha}\Gamma(\alpha)\sin(\frac{(1-\alpha)\pi}{2}) \cdot c_\gamma.$$

Параметр Херста описывает (асимптотически) самоподобие кумулятивной (агрегированной) величины трафика $\int_0^t X(s)ds$, в то время как α описывает ДВЗ процесса $X(t)$. Эти два параметра связаны между собой соотношением $H = (1 + \alpha)/2$.

Параметр α является основным, так как он определяет существование самого феномена и управляет поведением масштабной характеристики. Параметр c_f играет важную роль в проблеме оценки среднего. Для процессов с ДВЗ классическое асимптотическое выражение σ_X^2/n для дисперсии выборочного среднего как функции от размера выборки n заменяется на $2c_f n^{\alpha-1}/(1+\alpha)\alpha$. Так как дисперсия пропорциональна c_f, получаем, что доверительные интервалы для выборочного среднего, по существу, пропорциональны c_f. Кроме того, в такой области приложений как телекоммуникации этот параметр играет важную роль, поскольку, как показано в ряде исследований, его рост увеличивает задержку очереди.

Современные мультисервисные сети предоставляют самые различные интегральные услуги. Например, пользователь может запросить трансляцию телевизионного канала, затем доставку веб-документа или телефонное соединение. Таким образом, в одном физическом канале присутствует огромное количество информации различной по своей природе (аудио, видео, данные). В силу этого самоподобие трафика наблюдается лишь в определенном диапазоне временных шкал и есть основания полагать, что трафик обладает более сложной структурой. Такую структуру называют *мультифрактальной* (см. [1]). В рамках гауссовских процессов такие модели могут быть охарактеризованы переходом от дробного броуновского движения с постоянным параметром H к процессу, где этот параметр, в свою очередь, является некоторым случайным процессом $H = \eta(t)$. Хотя мультифрактальные модели являются более общими и, по-видимому, способны преодолеть недостатки самоподобных моделей, сложность таких моделей делает их практическое использование затруднительным.

Анализ современных технологий и сетей передачи информации показывает, что необходимо рассматривать трафик как совокупность нескольких компонент, каждая из которых отражает свойства трафика некоторых групп абонентов с одинаковыми характеристиками ([5]). Предполагается, что деление на компоненты осуществляется по "природе происхождения"трафика: голосовой трафик, трафик данных, видео в режиме реального времени и т.д. Компоненты такого процесса являются, как правило, статистически независимыми.

Таким образом, агрегированный трафик X при определенных условиях на характеристики источников асимптотически представляет собой сумму нескольких независимых самоподобных процессов:

$$X = \sum_{i=1}^{n} X_i,$$

где X_i - некоторый стандартный процесс, описывающий i-ую компоненту трафика.

За счет наличия нескольких аддитивных компонент такие модели

- способны более точно воспроизводить свойства реального трафика, аппроксимируя поведение неоднородного трафика,
- не используют громоздкий мультифрактальный формализм, что положительно сказывается на возможностях их практического применения.

Проведенные исследования показали, что для разных компонент телетрафика параметр Херста принимает различные значения. Поэтому при проектировании необходимо учитывать все компоненты трафика. В связи с этим актуальной является задача оценки параметров Херста каждой аддитивной компоненты в отдельности. Решению данной задачи и посвящена настоящая работа.

1. Оценка параметров однородного трафика

Для оценок параметра ДВЗ α и параметра Херста H для стационарных данных Patrice Abry и Darryl Veitch в 1996 году предложили использовать метод, основная идея которого заключается в использовании последовательности вейвлет-коэффициентов процесса ([3]). Предложенная оценка может быть применена как к сигналам непрерывного, так и дискретного времени. Она использовалась при анализе трафика LAN Ethernet и продемонстрировала хорошие свойства. Рассмотрим этот метод более подробно. Все приведенные в этой главе свойства и утверждения подробно рассмотрены и доказаны в статье [3].

Пусть $\psi_0(t) \in L^2(\mathbb{R})$ - так называемый материнский вейвлет. Ортогональный вейвлет-базис определяется следующим образом:

$$\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k), j, k \in \mathbb{Z},$$

где $j = 1, ..., J$; $k \in \mathbb{Z}$.

Всюду далее предполагается, что выполнены следующие предположения.

Предположение 1. *Базисная система функций строится с использованием оператора масштабного преобразования:*

$$\psi_{j,0} = 2^{-j/2}\psi_0(2^{-j}t).$$

Это означает, что анализирующее семейство проявляет масштабно-инвариантное свойство. Феномен ДВЗ состоит в том, что отсутствует какая-либо характеристическая частота в диапазоне частот близких к нулю. Таким образом, свойство ДВЗ может быть интерпретировано как характеристика инвариантная к масштабу, что эффективно анализируется вейвлетами.

Предположение 2. ψ_0 *имеет* N *нулевых моментов:*

$$\int t^k\psi_0(t)dt = 0,$$

$k = 0, ..., N-1$ *(но не для* $k \geq N$*!).*

В этом случае любая функция $g(t) \in L^2(\mathbb{R})$ разложима по этому базису

$$g(t) = \sum_{j\in\mathbf{Z}}\sum_{k\in\mathbf{Z}} c_{j,k}\psi_{j,k}(t),$$

где сходимость понимается в смысле сходимости по норме $L^2(\mathbb{R})$. Коэффициенты разложения $c_{j,k}$ определяются по правилу

$$c_{j,k} = \int \psi_{j,k}(t)g(t)dt.$$

На практике вейвлет-преобразование выполняется путем так называемого дискретного вейвлет-преобразования. Существуют очень эффективные алгоритмы реализации дискретного вейвлет-преобразования, по порядку сложности не превосходящие $O(n)$. Если исходные данные имеют длину N, то выходные данные $c_{j,k}$ разбиты на $J = \lfloor \log_2 N \rfloor$ октав. Число доступных вейвлет-коэффициентов для каждой октавы j равно $n_j = \lfloor 2^{-j}N \rfloor$. Данное выше определение можно перенести и на случайные функции, т.е. случайные процессы с непрерывным временем, используя понятие среднеквадратического интеграла. В работе используются гауссовские случайные процессы, что гарантирует корректность такого определения.

Коэффициенты вейвлет-преобразования определяются по правилу:

$$d_{j,k} = \int \psi_{j,k}(t)X(t)dt, \tag{3}$$

где интеграл понимается в среднеквадратическом смысле.

Вейвлет-коэффициенты обладают следующими свойствами ([3]):

Свойство 1. *В силу Предположения 1, имеет место точная масштабная инвариантность (степенное поведение):*

$$E(d_{j,k})^2 = 2^{j\alpha}c_f C\ , \tag{4}$$

где $C = \int |\nu|^{-\alpha}|\Psi_0(\nu)|^2 d\nu$,
Ψ_0 *- преобразование Фурье функции* ψ_0.

Свойство 2. *В силу Предположений 1 и 2, $d_X(j,k)$ есть совокупность случайных величин, которые являются квази-декоррелированными. В частности, долговременная зависимость, присутствующая в области временного представления, полностью отсутствует в плоскости (j,k) вейвлетных коэффициентов.*

В дальнейшем мы будем предполагать, что для процесса $X(t)$ выполнены следующие дополнительные условия:

Предположение 3. *Процесс $X(t)$ и, следовательно, $d_{j,k}$ являются гауссовскими.*

Предположение 4. *Для фиксированного j с.в. $\{d_{j,k}\}_{k\in\mathbb{Z}}$ независимы и одинаково распределены.*

Предположение 5. *Последовательности $\{d_{j,k}\}_{k\in\mathbb{Z}}$ при разных j независимы.*

Предположения являются достаточно "сильными однако в некоторых интересных нам случаях они заведомо выполняются (например, для дробного броуновского движения). В общем случае, Предположения 4 и 5 можно считать приемлемыми в силу того, что долговременная зависимость, присутствующая в исходном временном представлении, отсутствует в плоскости коэффициентов (j,k).

Для оценки $E(d_{j,k})^2$ будем использовать величину

$$\mu_j = \frac{1}{n_j}\sum_{k=1}^{n_j} d_{j,k}^2, \tag{5}$$

где n_j - число коэффициентов октавы j.

В силу Предположения 3 $d_{j,k}$ есть нормально распределенные случайные величины с нулевым средним (поскольку среднеквадратический интеграл есть линейное преобразование). В этом случае в силу Предположения 4 μ_j имеет χ^2-распределение. Или, более точно,

$$\mu_j \stackrel{d}{=} z_j \frac{X_{n_j}}{n_j}\ , \tag{6}$$

где X_{n_j} - с.в., имеющие χ^2-распределение с n_j степенями свободы.

Далее нам понадобится оценка величины $log_2 z_j$. Из-за нелинейности логарифмической функции $E\log_2\mu_j \neq \log_2 E\mu_j = \log_2 z_j$. Однако, при введенных выше дополнительных предположениях, смещение оценки можно найти явно, а значит и построить несмещенную оценку $\log_2 z_j$. Используя представление (6), имеем

$$\log_2\mu_j \stackrel{d}{=} \log_2 z_j - \log_2 n_j + \ln X_{n_j}/\ln 2\ .$$

Тогда, используя таблицу интегралов, можно показать (см. [3]), что

$$E\log_2\mu_j = \log_2 z_j + g_j\ ,$$

$$D\log_2\mu_j = \zeta(2, n_j/2)/\ln^2 2\ ,$$

где g_j - неслучайная функция аргумента n_j:

$$g_j = \psi(n_j/2)/\ln 2 - \log_2(n_j/2)\ . \tag{7}$$

Здесь $\psi(z) = \Gamma'(z)/\Gamma(z)$ есть ψ-функция и $\zeta(z,q) = \sum_{k=q}^{\infty} k^{-s}$ - ζ-функция Римана.

Другими словами, нам удалось найти случайную величину

$$y_j = log_2\mu_j - g_j \ , \tag{8}$$

такую, что

$$Ey_j = \log_2 z_j \ , \tag{9}$$

$$Dy_j = \zeta(2, n_j/2)/ln^2 2 \ . \tag{10}$$

Далее нам потребуются асимптотические выражения для g_j и Dy_j. Известно, что

$$\psi(z) \sim \ln z + \frac{1}{2z} + o(\frac{1}{z}) \ , \qquad z \to \infty \ ,$$

поэтому

$$g_j \sim \frac{-1}{n_j \ln 2} \ , \qquad n_j \to \infty \ . \tag{11}$$

Используя тот факт, что

$$\int_{q+1}^{\infty} x^{-s} dx \le \zeta(s,q) \le \int_{q}^{\infty} x^{-s} dx \ ,$$

легко видеть, что

$$Dy_j \sim \frac{2}{n_j \ln^2 2} \ , \qquad n_j \to \infty \ . \tag{12}$$

В силу определения процесса с ДВЗ в терминах спектральной плотности при выполнении описанных выше условий регулярности получаем:

$$z_j = 2^{j\alpha} c_f C \ , \tag{13}$$

где

$$C = \int |\Psi_0(\nu)|^2 |\nu|^{-\alpha} d\nu \ . \tag{14}$$

Вид уравнения (13) позволяет говорить о возможности оценки параметров с помощью линейной регрессии, т.к. при взятии логарифма от обоих частей равенства получим

$$log_2 z_j = j\alpha + \log_2(c_f C). \tag{15}$$

Полученная выше оценка y_j для выражения $\log_2 z_j$ позволяет воспользоваться взвешенной одномерной линейной регрессией для оценки параметров α и $\log_2(c_f C)$. Действительно, в силу (9) и (15),

$$Ey_j = j\alpha + \log_2(c_f C) \ ,$$

где в качестве фактора x_j выступает номер октавы $x_j = j$. Введем обозначения

$$S = \sum 1/\sigma_j^2, \qquad S_x = \sum x_j/\sigma_j^2, \qquad S_{xx} = \sum x_j^2/\sigma_j^2 \tag{16}$$

и положим $\sigma_j^2 = Dy_j$. Тогда оценки МНК параметров $b = \alpha$ и $a = \log_2(c_f C)$ равны соотвественно

$$\widehat{b} = \frac{\sum y_j (S x_j - S_x)/\sigma_j^2}{S S_{xx} - S_x^2} = \sum w_j y_j \ , \tag{17}$$

$$\widehat{a} = \frac{\sum y_j(S_{xx} - S_x x_j)/\sigma_j^2}{SS_{xx} - S_x^2} = \sum v_j y_j \ . \tag{18}$$

Легко проверить, что

$$\sum w_j = \sum j v_j = 0 \ , \sum v_j = \sum j w_j = 1 \ . \tag{19}$$

Полученные оценки $\widehat{a}$ и $\widehat{b}$ являются несмещенными и состоятельными для параметров $\log_2 c_f C$ и α. Рассмотрим подробнее оценку $c_f C$. Для этого естественно использовать оценку $2^{\widehat{\alpha}}$. Однако такая оценка будет смещенной из-за нелинейности показательной функции. Выделим это смещение. Сначала, используя определение (18) и обозначение $exp_2 x = 2^x$, получим

$$2^{\widehat{\alpha}} = exp_2\big(\sum v_j y_j\big) = exp_2\big(\sum v_j(\log_2 \mu_j - g_j)\big) = exp_2(-\sum v_j g_j)\prod \mu_j^{v_j} \ . \tag{20}$$

Преобразуем отдельно множители в правой части этого выражения:

$$\prod \mu_j^{v_j} \stackrel{d}{=} \prod (2^{j\alpha} c_f C X_{n_j}/n_j)^{v_j} \stackrel{d}{=} 2^{\alpha \sum j v_j}(c_f C)^{\sum v_J} \prod X_{n_j}^{v_j} \prod n_j^{-v_j} \stackrel{d}{=}$$

$$\stackrel{d}{=} c_f C \prod n_j^{-v_j} \prod X_{n_j}^{v_j},$$

где мы учли условия (19) и соотношение (6). Далее, используя определение (7), получаем

$$exp_2\big(-\sum v_j g_j\big) = exp_2\big(-\sum v_j[\psi(n_j/2)/\ln 2 - \log_2(n_j/2)]\big)$$

$$= exp_2\big(-\log_2 \exp \sum v_j \psi(n_j/2)\big) exp_2\big(\sum v_j \log_2(n_j/2)\big) =$$

$$= exp\big(-\sum v_j \psi_j(n_j/2)\big) \prod (n_j/2)_j^v \ .$$

Подставляя полученные формулы в исходное выражение, получаем

$$2^{\widehat{\alpha}} = \frac{c_f C}{exp(\sum v_j \psi(n_j/2))} \prod X_{n_j}^{v_j} \ . \tag{21}$$

Поскольку в силу предположения 5 случайные величины μ_j независимы в совокупности, то $E\prod X_{n_j}^{v_j} = \prod E X_{n_j}^{v_j}$. С помощью таблицы интегралов получаем:

$$E\prod X_{n_j}^{v_j} = \prod \frac{\Gamma(v_j + n_j/2)}{\Gamma(n_j/2)}.$$

Таким образом, величина обратная смещению равна

$$r = \prod \frac{exp(v_j \psi_j(n_j/2))\Gamma(n_j/2)}{\Gamma(v_j + n_j/2)} \ . \tag{22}$$

Определим оценку $c_f C$ как

$$S(\widehat{a}) = r 2^{\widehat{a}} \ . \tag{23}$$

Заметим, что оценка корректно определена только для тех j, для которых выполнено $2v_j/n_j > -1$. Введенная оценка очевидно является несмещенной. Ее состоятельность при $N \to \infty$ следует из состоятельности $\widehat{a}$ и непрерывности показательной функции.

2. Оценка параметров неоднородного трафика

В случае, когда результирующий процесс состоит из двух независимых стационарных компонент с нулевым средним, автоковариационная функция такого процесса есть сумма автоковариаций каждой из компонент. Заметим, что используемый метод может быть распространен на случай произвольного количества слагаемых.

В силу аддитивности преобразования Фурье рассматриваемая модель неоднородного трафика может быть определена следующим образом: при $|\nu| > 0$ спектральная плотность случайного процесса удовлетворяет соотношению:

$$f_X(\nu) \sim c_1|\nu|^{-\alpha_1} + c_2|\nu|^{-\alpha_2} \ , \tag{24}$$

$$0 < \alpha_2 < \alpha_1 < 1 \ .$$

Введем обозначения

$$g_i(\nu) = c_i|\nu|^{-\alpha_i}, \qquad i = 1, 2 \ ,$$

$$h(\nu) = f_X(\nu) - g_1(\nu) - g_2(\nu) \ .$$

При выполнении сформулированных выше условий получаем

$$z_j = 2^{j\alpha_1}C_1 + 2^{j\alpha_2}C_2 \ , \tag{25}$$

где

$$C_i = c_i \int |\Psi_0(\nu)|^2|\nu|^{-\alpha_i} d\nu \ .$$

В дальнейшем нам потребуется оценка для выражения $2^{l\alpha}c_fC$ (индекс j как и раньше будем использовать для суммирования/умножения по диапазону октав при построении линейной регрессии). Рассмотрим оценку $2^{l\alpha}c_fC$, построенную как произведение оценок $2^{\widehat{lb}}$ и $2^{\widehat{a}}$. Распределение последней величины нам уже известно (21). Найдем распределение $2^{\widehat{lb}}$.

Используя представление (17) получаем

$$2^{\widehat{b}} = exp_2\big(- \sum w_j y_j\big) \prod \mu_j^{w_j}.$$

Преобразуем по отдельности оба множителя:

$$\exp_2\left(w_j y_j\right) = \big(\sum w_j \psi_j(n_j/2)\big) \prod (n_j/2)^{w_j} \ .$$

Учитывая условия (19), имеем

$$\prod \mu_j^{w_j} \stackrel{d}{=} \prod (2^{j\alpha}c_f C X_{n_j}/n_j)^{w_j} \stackrel{d}{=} 2^{\alpha \sum w_j} \prod X_{n_j}^{w_j} \prod (n_j/2)^{w_j} \stackrel{d}{=}$$

$$\stackrel{d}{=} 2^{\alpha} \prod n_j^{-w_j} \prod X_{n_j} - w_j \ .$$

Возводя оба полученных выражения в степень l, получаем

$$2^{l\widehat{b}} = \frac{2^{l\alpha}}{exp(l \sum w_j \psi(n_j/2))} \prod X_{n_j}^{lw_j} \ .$$

Объединяя (21) и (15), имеем

$$2^{l\widehat{b}+\widehat{a}} = \frac{2^{l\alpha} c_f C}{exp(\sum lw_j + v_j)\psi(n_j/2))} \prod X_{n_j}^{lw_j+v_j} \ .$$

Пользуясь независимостью X_{n_j} и формулой для EX_v^d, легко видеть, что

$$E \prod X_{n_j}^{lw_j+v_j} = \prod \frac{\Gamma(lw_j + v_j + n_j/2)}{\Gamma(n_j/2)} \ .$$

Таким образом, корректирующий множитель равен

$$q = \prod \frac{exp[lw_j + v_j)\psi(n_j/2)]\Gamma(n_j/2)}{\Gamma(lw_j + v_j + n_j/2)} \ , \tag{26}$$

и несмещенная оценка выражения $2^{l\alpha} c_f C$ определяется как

$$Q_l(\widehat{b}, \widehat{a}) = q 2^{l\widehat{b}+\widehat{a}} \ . \tag{27}$$

Q_l определено только для тех j для которых $lw_j + v_j + n_j/2 > 0$. Состоятельность оценки при $N \to \infty$ следует из того факта, что если

$$\xi_n \xrightarrow{P} \xi, \qquad \eta_n \xrightarrow{P} \eta \ ,$$

то

$$a\xi_n + b\eta_n \xrightarrow{P} a\xi + b\eta,$$

состоятельности $\widehat{b}$, $\widehat{a}$ и непрерывности показательной функции.

Вернемся к формуле (25). Поскольку $\alpha_2 < \alpha_1$, то при больших j первое слагаемое в правой части (25) доминирует второе

$$z_j \simeq 2^{j\alpha_1} C_1 \ .$$

Поэтому при больших j для оценивания α_1 и C_1 можно применить описанный выше метод оценки параметров случайного процесса с долговременной зависимостью. С другой стороны, имея оценку для первого слагаемого, мы можем оценить второе, исходя из того, что при всех j

$$z_j - 2^{j\alpha_1} C_1 = 2^{j\alpha_2} C_2 \ .$$

Построим оценки для α_1 и $\log_2 C_1$ как описано в предыдущем разделе. Единственное различие состоит в том, что регрессия осуществляется не по всему диапазону доступных октав, а лишь по тем j, при которых второе слагаемое правой части

(25) достаточно мало. Например, можно использовать вторую половину всего диапазона $j \in J_1 = \{J/2, ..., J\}$:

$$\widehat{b_1} = \sum w_j y_j, \qquad \widehat{a_1} = \sum v_j y_j \ .$$

Полученные таким образом оценки $\widehat{b_1}$ и $\widehat{a_1}$ будут состоятельными.

Построим оценки для α_2 и $\log_2 C_2$. Для этого введем

$$\mu'_j = \mu_j - Q_j(\widehat{b_1}, \widehat{a_1}) \ , \tag{28}$$

где Q_l определено (для тех j, для которых $lw_j + v_j + n_j/2 > 0$) по формуле (27).

Второе слагаемое в (28) по определению есть несмещенная оценка $2^{j\alpha_1}C_1$, поэтому из (4) следует

$$E\mu'_j = 2^{j\alpha_2} C_2 \ .$$

Далее, определим

$$y'_j = \log_2 \mu'_j + g_j \ . \tag{29}$$

Пусть $\mu'_0 > 0$ - некоторое фиксированное число. Будем считать y_j определенным только для тех j, для которых $\mu'_j > \mu'_0$ и Q_j определено в (27). Обозначим это множество как J_2. Вообще говоря, целесообразным выглядит построение регрессии для параметров доминируемой части по значительно меньшему числу наблюдений, чем при построении регрессии для доминирующей компоненты ($|J_1| >> |J_2|$). Другими словами, перед началом второго шага, мы должны быть уверены в том, что на первом шаге мы оценили параметры достаточно точно.

Используя взвешенную линейную регрессию по $j \in J_2$

$$\widehat{b_2} = \sum_{J_2} w_j y'_j, \qquad \widehat{a_2} = \sum_{J_2} v_j y'_j \ , \tag{30}$$

получим несмещенные оценки $\widehat{b_2}$ и $\widehat{a_2}$ параметров α_2 и C_2.

Оценки $\widehat{b_2}$ и $\widehat{a_2}$ также являются состоятельными. Мы определяем y'_j только для тех j, для которых $\mu'_j > \mu'_0$. Иначе говоря, мы рассматриваем функцию

$$y'_j(v) = \log_2(v) - g_j$$

на множестве (μ'_0, ∞), на котором она непрерывна. Поскольку оценка $Q_j(\widehat{b_1}, \widehat{a_1})$ - состоятельна, то и оценка

$$y'_j(\mu'_j) = y'_j(\mu_j - Q_j(\widehat{b_1}, \widehat{a_1}))$$

при каждом фиксированном j состоятельна при $N \to \infty$. Состоятельность $\widehat{b_2}$ и $\widehat{a_2}$ следует из того, что они определяются как линейные комбинации y'_j (см. (30)) (линейная комбинация случайных величин, каждая из которых сходится по вероятности, тоже сходится по вероятности).

3. Численное моделирование

На основе предложенной нами методики, разработана численная процедура оценки параметров неоднородного трафика, а также ее программная реализация.

В качестве исходных данных используем смоделированный процесс, состоящий из двух независимых дробных броуновских движений с различными параметрами Херста.

Дробное Броуновское Движение (ДБД) - наиболее известный пример самоподобного случайного процесса. Это гауссовский стохастический процесс $(Y(t), t \geq 0$ со стационарными приращениями, который имеет нулевое среднее и ковариационную функцию вида:

$$R(t,s) = \frac{\sigma^2}{2}(t^{2H} + s^{2H} + |t-s|^{2H}) \ .$$

где $H \in (0,1)$ - параметр Херста. Этот процесс не является стационарным. Перейдем к приращениям процесса X. Зафиксируем некоторое $\delta > 0$ и рассмотрим новый процесс с дискретным временем:

$$X(n) := Y(n \cdot \delta) - Y((n-1) \cdot \delta) \ .$$

Этот процесс является стационарным и его автоковариационная функция обладает следующим свойством:

$$r(n) := Cov(X(m), X(m+n)) \sim H(2H-1) \cdot n^{2-2H} \ .$$

Если $H \in (1/2, 1)$, то он обладает свойством долговременной зависимости.

В работе [6] рассмотрено несколько способов генерации процесса ДБД. Для моделирования мы использовали один из них, а именно алгоритм случайного перемещения средней точки (RMA). Основной принцип данного алгоритма - рекурсивно расширять сгенерированную выборку, добавляя новые значения в средних точках относительно значений в оконечных точках.

Цель деления интервала между 0 и 1 - построение гауссовских приращений для X. Добавляя смещение к средним точкам, создают нормальное маргинальное распределение полученной последовательности. Теоретическая алгоритмическая сложность метода - $O(n)$.

В случае неоднородности трафика необходимо выделить различные компоненты трафика. С этой целью был проведен модельный эксперимент. На вход исходного алгоритма, предложенного P. Abry и D. Veitch, подавался неоднородный процесс. А именно, процесс, состоящий из суммы двух независимых дробных броуновских движений с различными параметрами Херста. Результаты моделирования представлены в таблице 1. В таблице принято следующее обозначение: ДБД(0.8)+ДБД(0.6) - сумма двух дробных броуновских движений с параметрами Херста 0.8 и 0.6 соответственно. Проведено по 150 измерений каждого процесса. Общее количество точек в каждом смоделированном процессе есть 2^{16}.

Был получен следующий результат: оценка параметра степени самоподобия (параметра Херста) является средним арифметическим параметров Херста обеих компонент. Заметим, что в данном случае оценка параметра проводится по всему диапазону октав вейвлет-коэффициентов.

Затем был проведен эксперимент с однородным и неоднородным трафиком, но расчеты проводились не по всему диапазону октав, а по его части, а именно, по второй половине октав, как это предлагается выполнять в предлагаемой нами методике. Результаты моделирования представлены в таблице 2.

Таблица 1: Оценка параметров неоднородного процесса с помощью алгоритма P. Abry и D. Veitch

Исходный процесс	Оценка параметра Херста
ДБД(0.8)+ДБД(0.6)	$\widehat{H} = 0.72$
ДБД(0.9)+ДБД(0.6)	$\widehat{H} = 0.75$
ДБД(0.9)+ДБД(0.7)	$\widehat{H} = 0.79$

Таблица 2: Оценка параметров неоднородного процесса с помощью алгоритма P. Abry и D. Veitch

Исходный процесс	Оценка параметра Херста (метод Abry и Veitch)	Оценки параметров Херста компонент процесса
ДБД(0.6)	$\widehat{H} = 0.63$	$\widehat{H_1} = 0.59$ $\widehat{H_2} = 0.40$
ДБД(0.7)	$\widehat{H} = 0.71$	$\widehat{H_1} = 0.69$ $\widehat{H_2} = 0.49$
ДБД(0.8)	$\widehat{H} = 0.80$	$\widehat{H_1} = 0.79$ $\widehat{H_2} = 0.53$
ДБД(0.9)	$\widehat{H} = 0.89$	$\widehat{H_1} = 0.89$ $\widehat{H_2} = 0.58$
ДБД(0.8)+ДБД(0.6)	$\widehat{H} = 0.72$	$\widehat{H_1} = 0.80$ $\widehat{H_2} = 0.61$
ДБД(0.9)+ДБД(0.6)	$\widehat{H} = 0.75$	$\widehat{H_1} = 0.89$ $\widehat{H_2} = 0.62$
ДБД(0.9)+ДБД(0.7)	$\widehat{H} = 0.79$	$\widehat{H_1} = 0.88$ $\widehat{H_2} = 0.68$

В результате было установлено следующее. В случае однородного трафика оценка, рассчитанная по всему диапазону октав и по его части совпадает. Если же мы имеем дело с неоднородным трафиком, то в результате расчетов, произведенных по второй половине октав вейвлет-коэффициентов мы получаем оценку для той компоненты, в которой степень самоподобия выше.

На основание вышесказанного можно сделать следующий вывод. Сравнив результат о рассчитанных оценках по всему диапазону октав и по второй половине октав, мы можем однозначно ответить на вопрос об однородности трафика. Если оценка, рассчитанная по исходной методике выше или равна оценке, рассчитанной по второй половине октав, то это говорит о том, что трафик однороден. В противном случае мы имеем дело с неоднородным трафиком.

Алгоритм для оценки неоднородного трафика можно представить в следующем виде:

- В качестве конкретного вейвлет-базиса можно использовать вейвлеты Добеши.
- Применяя дискретное вейвлет-преобразование к исследуемому процессу $X(t)$ получаем вейвлет-коээфициенты $d_{j,k}$.
- По соответствующим формулам (5) и (8)находим величины μ_j и y_j соответственно.

- Далее, используя МНК, рассчитываем оценки $(\widehat{b_1}, \widehat{a_1})$. Расчеты выполняются по второй половине октав исходного диапазона $[J/2, \ldots, J]$.
- По соответствующим формулам (28) и (29)находим величины μ'_j и y'_j соответственно.
- Далее, используя МНК, рассчитываем оценки $(\widehat{b_2}, \widehat{a_2})$ при условии, что величины y'_j определены только для $j \in J_2 = \{j : \mu'_j > \mu'_0, \mu'_0 > 0\}$.

4. Практическое применение

Предлагаемая методика была опробована на реальных данных. Данные для анализа взяты из открытых источников сети интернет [2]. Был проведен анализ трех различных реализаций телетрафика:

- BCpAug89 - трафик Интранет, измеренный в лаборатории Bellcore Morristown Research and Engineering facility. Архив содержит миллион измерений.
- DecPkt1TCP - трафик сообщений между компанией Digital Equipment Corporation и внешними абонентами глобальной сети Интернет. Архив содержит 3,3 миллиона измерений.
- LblConn7 - трафик сообщений между лабораторией Lawrence Berkeley Laboratory (LBL) и внешними абонентами глобальной сети Интернет. Период измерений - 30 дней.

Исходные данные представлены двумя колонками в ASCII-формате. Первая колонка содержит временные метки появления пакетов (в сек.), а вторая - общий размер пакета в байтах. Результаты оценки представлены в таблице 3.

Таблица 3: Оценка параметров реального трафика

Трафик	Вид процесса	Параметр Херста (в случае однородного процесса)	Параметры Херста (в случае неоднородного процесса)
BCpAug89	неоднородный	$H = 0.8035$	$H_1 = 0.8824$ $H_2 = 0.6326$
DecPkt1TCP	неоднородный	$H = 0.5974$	$H_1 = 0.7684$ $H_2 = 0.5371$
LblConn7	однородный	$H = 0.6250$	$H_1 = 0.6290$ $H_2 = 0.2321$

После проведенных расчетов можно сделать следующие выводы. Реализации трафиков BCpAug89 и DecPkt1TCP представляют собой неоднородные процессы,

Таблица 4: Оценка параметров спектр. плотности реального трафика

Трафик	Оценка параметров
BCpAug89	$\alpha_1 = 0.7648$; $c_f1 = 3.32 \cdot 10^4$ $\alpha_2 = 0.2652$; $c_f2 = 4.35 \cdot 10^4$
DecPkt1TCP	$\alpha_1 = 0.5368$; $c_f1 = 5.06 \cdot 10^6$ $\alpha_2 = 0.0742$; $c_f2 = 3.58 \cdot 10^7$
LblConn7	$\alpha = 0.25$ $c_f = 4.13 \cdot 10^{11}$

то есть можно выделить несколько компонент процесса за счет передачи информации разных видов и на основе этой информации смоделировать трафик для решения практических задач, таких как предотвращение возможных перегрузок сети, оптимизация использования сетевых каналов, решения других вопросов. Для реализации LblConn7 более адекватной является однокомпонентная модель трафика, то есть модель трафика, основанная на предположении, что весь трафик описывается процессом с фиксированным значением параметра Херста.

Заключение

В настоящей статье предложен метод оценки параметров процесса, являющегося суммой двух независимых случайных процессов с долговременной зависимостью. Фактически процедура оценивания состоит из двух этапов:

- построение вейвлет-преобразования исходных данных,
- двукратное применение взвешенной линейной регрессии,

Аппробация на реальных данных показала вычислительную эффективность предлагаемой схемы. На основе предложенной схемы разработана численная процедура оценки параметров процесса. Описан метод, позволяющий определять однородность-неоднородность трафика. Процедура тестировалась на смоделированных данных, а затем применялась к траекториям реального трафика, где показала свою эффективность.

Список литературы

[1] P. Flandrin Wavelet analysis and synthesis of fractional brownian motion // IEEE Transactions on Information Theory. — 1992.— Vol. IT- 38. — pp. 910-917.

[2] LBL "Internet traffic archive $http://ita.ee.lbl.gov/html/traces.html$

[3] Veitch D., Abry P., A wavelet based joint estimator of the parameter of long-range dependence. // *IEEE Transaction on Information Theory.*— 1999. — v. 45, n. 3. — pp. 878-897.

[4] Бестугин и др. Контроль и диагностирование телекоммуникационных сетей / А. Р. Бестугин, А. Ф. Богданова, Г. В. Стогов. —- СПб: Политехника, 2003. 174 с.

[5] Цитович И.И. Устойчивые модели трафика мультисервисных сетей/ И. И. Цитович. // 60-я Научная сессия, посвященная Дню радио, 17-19 мая 2005. — Санкт-Петербург. — с. 271-273.

[6] Шелухин О.И., Тенякшев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях. Монография. — М: Радиотехника, 1981. 199 с.