

Статистические методы исследования сетевого трафика

Т. М. Татарникова^а, доктор техн. наук, доцент, orcid.org/0000-0002-6419-0072, tm-tatarn@yandex.ru

^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ

Постановка проблемы: концепция единой мультисервисной сети, предполагающая интеграцию передачи речи, данных и мультимедиа, побудила интерес к изучению природы сетевого трафика. Исследования трасс трафика, записанного в крупных масштабах времени, показывают наличие в нем самоподобной структуры, что требует пересмотра результатов моделирования инфокоммуникационных сетей в предположении о пуассоновском потоке данных. **Цель исследования:** с целью получения инструментальных средств для генерации искусственного трафика, адекватно отражающего реальный трафик сети с учетом выявленных свойств самоподобия, изучить последовательность применения методов изучения природы сетевого трафика, выявляющих самоподобную природу трафика в виде статистических оценок и показателя Херста. **Результаты:** проверены свойства самоподобия рассматриваемого 3G-трафика на разных временных шкалах, полученных агрегацией по 5, 10, 15 и 20 мин на имеющихся суточных данных 3G-трафика. Получена оценка тяжести «хвоста» распределения самоподобного трафика методом построения линии регрессии для дополнительной функции распределения в логарифмическом масштабе. Значение параметра самоподобия, определяемое тяжестью «хвоста» распределения, позволило подтвердить предположение о самоподобии 3G-трафика. Выполнен обзор моделей, имитирующих реальный сетевой трафик с самоподобной структурой. Реализованы инструментальные средства для генерации искусственного трафика в соответствии с рассмотренными моделями. Выполнено сравнение генераторов искусственного сетевого трафика по критерию метода наименьших квадратов аппроксимации точечных значений искусственного трафика аппроксимирующей функцией 3G-трафика. Учтены качественные оценки генераторов искусственного сетевого трафика в виде трудоемкости их программной реализации, что, впрочем, может быть субъективной оценкой. Сравнительные характеристики позволяют выбрать генератор(ы), которые максимально правдоподобно имитируют реальный сетевой трафик. **Практическая значимость:** предложенная последовательность методов исследования свойств сетевого трафика необходима для понимания его природы и разработки соответствующих моделей, имитирующих реальный сетевой трафик.

Ключевые слова — сетевой трафик, самоподобный процесс, долговременная зависимость, медленно затухающая дисперсия, распределение с тяжелым «хвостом», параметр самоподобия, модели имитации самоподобного трафика.

Цитирование: Татарникова Т. М. Статистические методы исследования сетевого трафика. *Информационно-управляющие системы*, 2018, № 5, с. 35–43. doi:10.31799/1684-8853-2018-5-35-43

Citation: Tatarnikova T. M. Statistical methods for studying network traffic. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 5, pp. 35–43 (In Russian). doi:10.31799/1684-8853-2018-5-35-43

Введение

Моделирование оценивания характеристик обслуживания сетевого трафика остается в настоящее время актуальной научной задачей. Достоверные оценки сетевого трафика необходимы при планировании развития телекоммуникационных сетей, выборе политики дифференцированного обслуживания и характеристик вычислительных ресурсов, гарантирующих требуемое качество обслуживания при соответствующей сетевой нагрузке [1, 2].

Вспыхнувший интерес к изучению природы сетевого трафика объясняется результатами исследований, показавших наличие в трафике долговременных зависимостей или процесса самоподобия. Эти изменения в структуре трафика связаны с реализацией концепции единой мультисервисной сети, предполагающей интеграцию передачи речи, данных и мультимедиа [3, 4].

На сегодняшний день теория самоподобных стохастических процессов не так хорошо развита,

как теория пуассоновских процессов. Учитывая известные выводы о самоподобности сетевого трафика, актуальными задачами становятся методы его изучения и разработка инструментальных средств для генерации искусственного трафика, адекватно отражающего реальный неоднородный трафик сети [5, 6].

Свойства и характеристики самоподобных процессов

Самоподобность описывает явление, при котором некоторые статистические характеристики процесса сохраняются при масштабировании времени. При усреднении по шкале времени у самоподобного процесса не происходит быстрого «сглаживания», т. е. сохраняется склонность к всплескам [7, 8].

К свойствам, характеризующим самоподобность процесса, относят такие, как медленно затухающая дисперсия, долговременная зависи-

мость, наличие распределения с тяжелыми «хвостами» [9–11].

Свойство медленно затухающей дисперсии заключается в том, что дисперсия выборочного среднего затухает медленнее, чем величина, обратная размеру выборки, т. е.

$$D(X^{(n)}(t)) = \sigma^2 n^{2H-2}, \quad n \rightarrow \infty, \quad (1)$$

где σ^2 — дисперсия процесса $X(t)$; n — объем выборки; H — параметр Херста (параметр самоподобия), $0,5 < H < 1$.

Отметим, что для традиционных случайных процессов дисперсия выборочного среднего уменьшается обратно пропорционально объему

$$\text{выборки: } D(X^{(n)}(t)) = \frac{\sigma^2}{n}.$$

Наличие долговременной зависимости заключается в том, что самоподобный процесс обладает гиперболически затухающей корреляционной функцией

$$R(k) \cong k^{(2H-2)}L(k), \quad \forall k \geq 1, \quad k \rightarrow \infty, \quad (2)$$

где $L(k)$ — медленно меняющаяся функция на бесконечности, для которой

$$\lim_{k \rightarrow \infty} \frac{L(kx)}{L(k)} = 1 \quad \text{для всех } x > 0. \quad (3)$$

Свойство наличия распределения с тяжелым «хвостом» заключается в том, что случайная величина X имеет распределение с тяжелым «хвостом», если [12]

$$P(X > x) \sim cx^{-\alpha}, \quad x \rightarrow \infty, \quad (4)$$

где $0 < \alpha < 2$ — параметр формы распределения; c — некоторая положительная константа.

Методы исследования самоподобного процесса

Существует ряд приемов, позволяющих выполнить проверку свойства самоподобия исследуемого процесса [13–17].

Эффект самоподобия можно наблюдать на графиках (рис. 1–4), иллюстрирующих изменение временной шкалы, при котором структура ряда, полученного усреднением групп элементов, остается такой же, как и структура исходного. Этот факт является предпосылкой для предположения о самоподобной структуре рассматриваемого процесса и основанием для проведения дальнейшего более полного анализа [18].

Рассмотрим суточные данные 3G-трафика от 20.08.2018 г., предоставленные сотовым оператором МТС г. Санкт-Петербурга (рис. 1). Временной ряд состоит из 1440 наблюдений, каждое из которых представляет собой объем эксплуатируемой услуги X_t Гбайт, за время $t = 1, 1440$ мин.

Выполним агрегацию временного ряда, уменьшив размер шкалы наблюдений в 5 раз. Значения нового временного ряда получены в соответствии со следующим выражением:

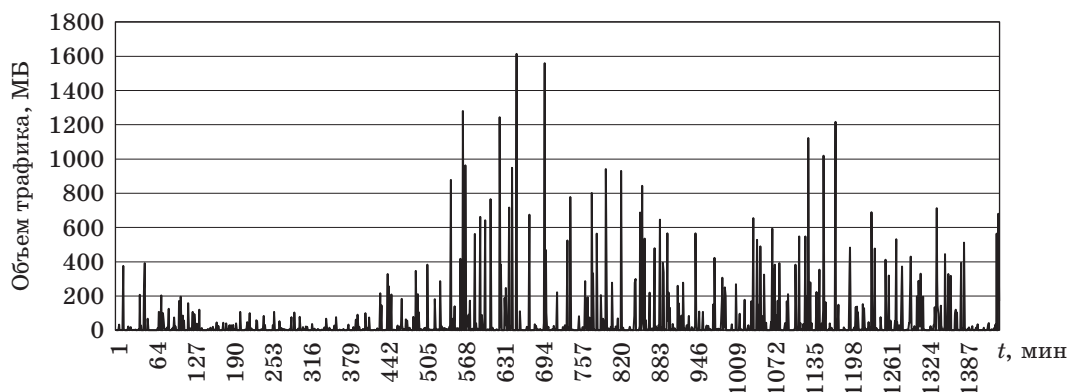
$$X_t = \frac{\sum_{i=(t-1)m+1}^{tm} X_i}{m}, \quad (5)$$

где m — число усредняемых последовательных членов ряда.

Новый ряд включает 288 событий и приведен на рис. 2.

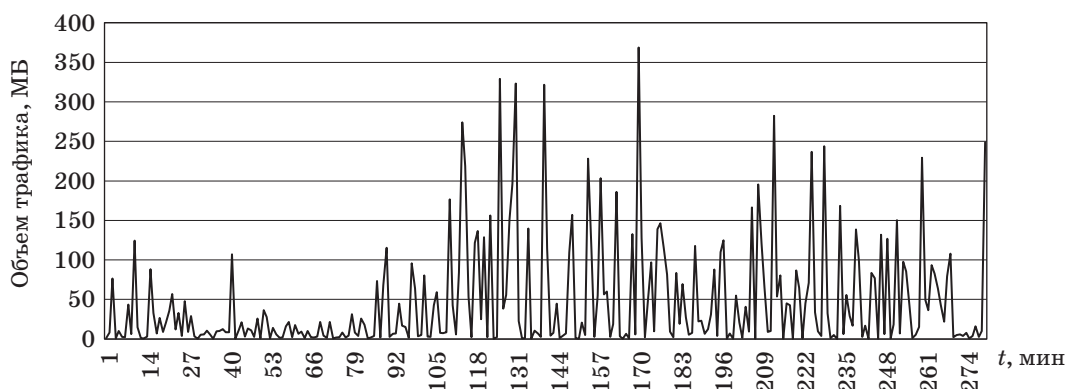
Та же процедура с уменьшением размера временной шкалы исходного ряда наблюдений в 10 раз приводит к результату, представленному на рис. 3.

Уменьшение размера временной шкалы исходного ряда наблюдений в 20 раз приводит к результату, представленному на рис. 4.

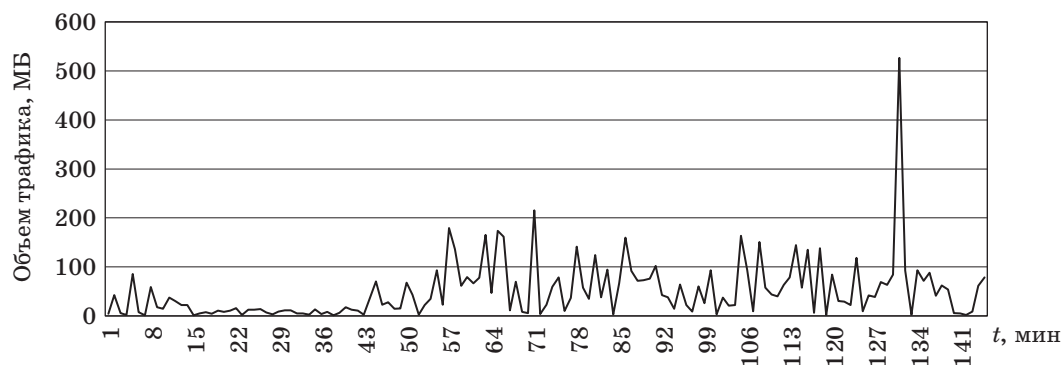


■ Рис. 1. Исходный 3G-трафик за период 1440 мин

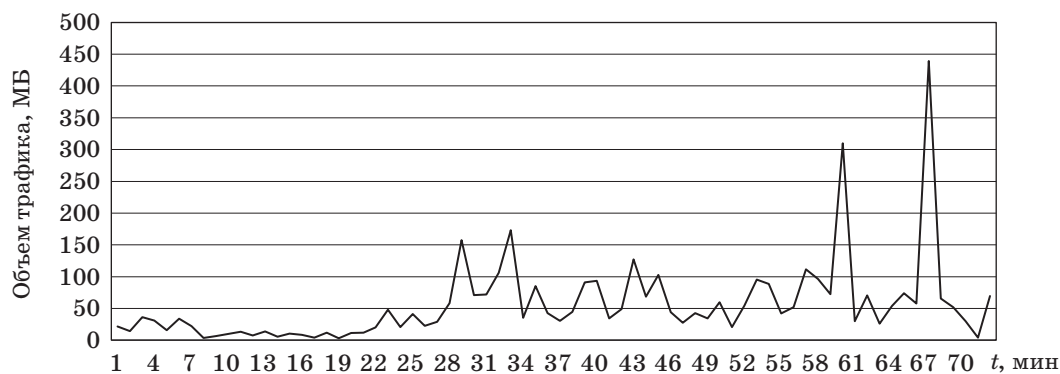
■ Fig. 1. Initial 3G-traffic period of 1440 min



■ **Рис. 2.** Агрегация 3G-трафика по 5 мин за период 1440 мин
 ■ **Fig. 2.** Aggregation 3G-traffic over 5 min for the period of 1440 min



■ **Рис. 3.** Агрегация 3G-трафика по 10 мин за период 1440 мин
 ■ **Fig. 3.** Aggregation 3G-traffic over 10 min for the period of 1440 min



■ **Рис. 4.** Агрегация 3G-трафика по 20 мин за период 1440 мин
 ■ **Fig. 4.** Aggregation 3G-traffic over 20 min for the period of 1440 min

Визуальное наблюдение агрегированных процессов (см. рис. 1–4) позволяет сделать вывод о сохранении структуры процесса.

Далее необходимо оценить тяжесть «хвоста» распределения — показатель α . С этой целью построим график дополнительного распределения

$\bar{F}(x) = 1 - F(x) = P(X > x)$ из непересекающихся групп, объединяющих события исходного процесса, и вычислим частоты попадания в каждый интервал. Построим новый ряд, вычислив среднее значение в каждом интервале, и соответствующую частоту. Полученная функция рас-

предела позволяет вычислить значения $\bar{F}(x)$ и построить ее график в логарифмической шкале (рис. 5).

Построим линию регрессии для дополнительной функции распределения (рис. 6) и, вычислив тангенс угла наклона к горизонтальной оси, получим значение параметра α

$$\alpha = \frac{\partial \log \bar{F}(x)}{\partial \log x}, \quad (6)$$

через который выражается параметр самоподобия $H = \frac{3 - \alpha}{2}$.

Показатель α может принимать значения на интервале $[0, 2]$. Чем меньше это значение, тем тяжелее «хвост» распределения [18].

Распределение с тяжелым «хвостом» имеет ряд свойств, которые существенно отличают его от наиболее известных распределений, таких как экспоненциальное, нормальное или пуассоновское [19].

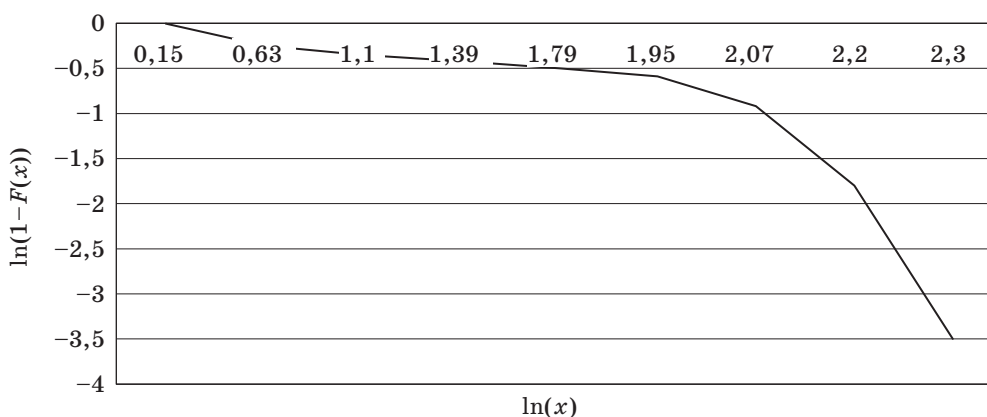
Если $\alpha \leq 2$, тогда распределение имеет бесконечную дисперсию.

Если $\alpha \leq 1$, то распределение имеет бесконечное среднее.

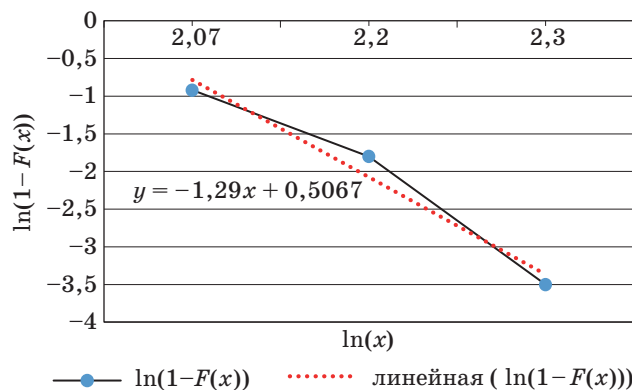
При уменьшении α произвольная большая порция плотности может быть представлена в «хвосте» распределения. Фактически тяжелый «хвост» означает наличие бесконечной дисперсии, иначе говоря, случайная величина может принимать очень большие значения, но с очень маленькой вероятностью.

Выведенное уравнение регрессии показывает, что α принимает значение, равное 1,29, и $\alpha \in [0; 2]$, откуда следует, что распределение данных 3G-трафика имеет свойство тяжелого «хвоста».

Вычислим значение параметра самоподобия $H = (3 - 1,29)/2 = 0,855$, что также подтверждает свойства самоподобия рассматриваемого процесса, так как $H \in [0,5; 1]$. Рассчитанный параметр H будем использовать при увеличении масштаба шкалы. В табл. 1 приведены результаты оценки дисперсии, среднего и коэффициента корреля-



■ Рис. 5. $\bar{F}(x)$ «хвоста» распределения для непересекающихся групп данных
 ■ Fig. 5. $\bar{F}(x)$ tail distribution for disjoint data groups



■ Рис. 6. Линия регрессии и вывод уравнения регрессии
 ■ Fig. 6. The regression line and regression equation

■ Таблица 1. Результаты оценки свойств 3G-трафика
 ■ Table 1. 3G-traffic properties evaluation results

Трафик	$D(X)$	$M(X)$	$R(k), k = 2$
3G исходный	4149,47	37,82	0,0233
3G агрегированный по 5 мин	4341,17	37,25	0,0712
3G агрегированный по 10 мин	3354,93	40,39	0,0678
3G агрегированный по 20 мин	3991,34	38,21	0,0349

ции для исходного и агрегированного трафиков, демонстрирующих свойства распределений с тяжелыми «хвостами».

Модели имитации самоподобного трафика

Свойства, полученные в результате измерений реального сетевого трафика, часто отличаются от свойств трафика, полученного традиционными моделями, в частности моделью пуассоновского потока [20]. Основной причиной такого различия является структура, лежащая в основе зависимости.

Пакетный трафик, представляющий собой интеграцию речи, данных и мультимедиа, охватывает сильно отличающиеся масштабы времени — от микросекунд до секунд и даже минут [1]. В каждом масштабе времени присутствует долговременная зависимость — наличие пульсаций, которые состоят из периодов активности, разделенных менее активными периодами. В традиционных моделях сетевого трафика пульсации сильно сглажены на больших масштабах времени, отчего свойство долговременной зависимости пропадает [21].

В результате оказывается, что традиционный анализ инфокоммуникационных систем, в основе которого лежит предположение о пуассоновском потоке, не может достаточно точно оценить объемы вычислительных ресурсов и производительность системы в условиях пульсирующего трафика.

Необходимы инструментальные средства для генерации искусственного трафика, соответствующего свойствам реального сетевого трафика, кото-

рые можно использовать при моделировании процессов передачи, сохранения и обработки сетевого трафика.

Известны всего несколько моделей, предназначенных для имитации самоподобного трафика. Анализ доступных публикаций по моделированию самоподобного трафика позволяет выделить модели [22], приведенные в табл. 2.

В работе реализованы инструментальные средства для генерации искусственного трафика по моделям, перечисленным в табл. 2. Сравнительные характеристики позволяют выбрать генераторы, которые максимально правдоподобно имитируют реальный сетевой трафик. При сравнении оценивался критерий S метода наименьших квадратов аппроксимации точечных значений искусственного трафика аппроксимирующей функцией 3G-трафика

$$S = \sum_{i=1}^N (F(x_i) - y_i)^2, \tag{7}$$

где $F(x_i)$ — значения аппроксимирующей функции в узловых точках x_i искусственного трафика; y_i — заданный массив исходного 3G-трафика в узловых точках x_i .

За узловую точку принята каждая 60-я минута, итого 24 часа — 24 узловые точки.

Кроме количественной оценки, в табл. 2 приведены и качественные оценки в виде трудоемкости реализации программного генератора (число настраиваемых параметров или необходимость обучения). Это субъективная оценка, которую труд-

■ **Таблица 2.** Модели генераторов самоподобного трафика

■ **Table 2.** Models of self-similar traffic generators

Модель	Математическое описание модели	Число настраиваемых параметров модели или обучения	S
Фрактальное броуновское движение — процесс движения частицы, совершающей хаотическое перемещение с шагом, задаваемым предысторией движения	$B_H(t) = \frac{1}{\Gamma(H-1/2)} \int_{-\infty}^t K(t-t') dB(t'),$ <p>где $\Gamma(\cdot)$ — гамма-функция; $dB(t')$ — независимые случайные перемещения частицы в моменты времени t'; $K(t-t')$ — функция памяти системы:</p> $K(t-t') = \begin{cases} (t-t')^{H-1/2}, & 0 \leq t' \leq t \\ (t-t') - (-t')^{H-1/2}, & t' \leq 0 \end{cases}$	Выбор и настройка функции $K(t-t')$	0,324
Фрактальный гауссовский шум — итерационный процесс последовательного деления отрезка единичной длины пополам	$y_c = \frac{(y_1 + y_2)}{2} + h,$ <p>где h — случайная величина, распределенная по нормальному гауссову закону с нулевым средним и дисперсией $\sigma_h^2 = r^H$, где $r = (x_1 - x_2)/2$ — расстояние от средней точки x_c рабочего отрезка $[x_1, x_2]$</p>	Настройка параметров генерации h , выбор числа итераций	0,355

■ Окончание табл. 2
 ■ Table 2

Модель	Математическое описание модели	Число настраиваемых параметров модели или обучения	S
Хаотические отображения — логистическое уравнение	$X_{n+1} = CX_n - C(X_n)^2 = CX_n(1 - X_n)$, где C — параметр скорости распространения случайной величины X , самоподобность проявляется при $3 < C < 3,57$	Настройка параметра C и параметров базовой генерации X	0,289
Динамическое моделирование Маркова — автоматы с конечным числом состояний, в которых реализован вероятностный переход из одного состояния в другое [23]	$P_{\text{пер } A \rightarrow B} = \frac{C_{\text{пер } A \rightarrow B}}{\sum_i C_i}$, где i — номер счетчика; C_i — значение i -го счетчика; $P_{\text{пер } A \rightarrow B}$ — вероятность перехода из состояния A в состояние B	Настройка вероятностей перехода	0,243
Модели нечеткой логики — уровни пачечности представлены в виде ступенчатой функции полезности f	$U = f(\lambda, P)$, где U — уровень пачечности трафика; λ — средняя интенсивность трафика на интервале самоподобия $t_i, i = \overline{1, n}$; $P \sim p_i, i = \overline{1, n}$ — вероятности разыгрывания значения $\lambda_i, i = \overline{1, n}$	Настройка ступенчатой функции полезности	0,379
Нейросетевые модели — аппроксимация функций нескольких переменных по обучающей выборке временного ряда [24]	$Y(t) = \hat{F}(Z(t))$, где $\hat{F}(\cdot)$ — оператор нейросетевого отображения; $Z(t) = \{t, X(t), X(t-1), \dots, X(t-n)\}$ — временной ряд; $X(t)$ — интенсивность трафика в момент времени t ; n — размер пачки	Обучение нейронной сети требует настройки $N = x^2y$ весов, где x — число нейронов в слое; y — число внутренних слоев	0,02
Авторегрессионные модели — модели временных рядов, в которых любые значения временного ряда линейно зависят от предыдущих значений этого же ряда	$X_t = c + \sum_{i=1}^p a_i X_{t-1} + \varepsilon_t$, где a — коэффициенты авторегрессии; c — постоянная; p — размер пачки трафика; ε_t — белый шум	Настройка a, c , параметров генератора белого шума ε_t	0,09
ON/OFF-модели — источник пакетов в ON-периоды в течение времени T_0 генерирует пакеты, в OFF-периоды в течение времени T_1 источник пассивен [25]	Периоды T_0, T_1 — случайные переменные с функцией плотности вероятности $w_0(t)$ и $w_1(t)$ соответственно; распределения $w_0(t), w_1(t)$ являются распределениями с тяжелыми «хвостами»	Обучение не требуется	0,05

но оценить, например, как время, затраченное на программирование генератора или сложность алгоритма, поскольку все зависит от размера пачки, времени, затраченного на настройку одного параметра и комплекса параметров, хорошего владения языком программирования и др. Например, несмотря на то, что нейросетевая модель показала лучший результат по критерию S , основное время при ее реализации было потрачено на выбор архитектуры нейронной сети, а затем на настройку модели (3 дня), в то время как модель фрактального гауссовского шума была реализована за 40 мин, но критерий S в 17,5 раз больше, чем у нейросетевой модели. Более того, для имитации трафика

с другим показателем Херста вновь потребуется процедура выбора архитектуры нейронной сети и ее обучение.

Анализ приведенных моделей позволяет сконцентрироваться на трех последних, представленных в табл. 2, и использовать их в решении задач моделирования инфокоммуникационных систем и сетей с вытекающими из этого глобальными проблемами: планирование развития инфокоммуникационных сетей, реализация дифференцированного обслуживания, оценка характеристик вычислительных ресурсов, гарантирующих требуемое качество обслуживания соответствующего трафика.

Заключение

В статье представлены результаты исследования 3G-трафика с целью выявления свойства самоподобия. Предположение о самоподобной структуре 3G-трафика основывается на рассмотрении имеющихся данных для различной временной шкалы. Используя метод построения дополнительной функции распределения для логарифмической шкалы, оценены тяжесть «хвоста» распределения и параметр самоподобия. Полученные результаты позволили проверить свойства самоподобия рассматриваемого трафика

согласно определению и таким образом подтвердить предположение о самоподобии 3G-трафика.

Подобные исследования необходимы для понимания поведения сетевого трафика и синтеза моделей, имитирующих процесс поступления реального трафика в сети.

Обзор и сравнение существующих моделей имитации самоподобного трафика по результатам аппроксимации 3G-трафика позволили сделать предположение, что настройка модели может быть выполнена по показателю Херста при наличии записанных трасс реального сетевого трафика.

Литература

1. Tanenbaum A., Wetherall D. Computer Networks. Prentice Hall, 2010. 960 p.
2. Кутузов О. И., Татарникова Т. М. Инфокоммуникационные сети. Моделирование и оценка вероятностно-временных характеристик. СПб.: ГУАП, 2015. 382 с.
3. Крылов В. В., Самохвалова С. С. Теория телетрафика и ее приложения. СПб.: БХВ-Петербург, 2005. 288 с.
4. Кутузов О. И., Татарникова Т. М. Моделирование систем и сетей телекоммуникаций. СПб.: РГМУ, 2012. 134 с.
5. Петров В. В. То, что вы хотели знать о самоподобных процессах, но стеснялись спросить. М.: Радиотехника, 2003. 112 с.
6. Addie R. G., Neame T. D., Zukerman M. Performance evaluation of a queue fed by a poisson pareto burst process. *Computer Networks*, 2002, vol. 40, no. 3, pp. 377–397.
7. Шелухин О. И., Осин А. В., Смольский С. М. Самоподобие и фракталы. Телекоммуникационные приложения. М.: Физматлит, 2008. 368 с.
8. Erramilli A., Narayan O., Willinger W. Experimental queueing analysis with long-range dependent traffic. *IEEE/ACM Trans. Networking*, 1996, vol. 4, pp. 209–223.
9. Leland W. E., Taqqu M. S., Willinger W., Wilson D. V. On the self-similar nature of ethernet traffic. *Proc. ACM SIGCOMM'93*. San-Fransisco, 1993, pp. 183–193.
10. Шелухин О. И. Мультифракталы. Инфокоммуникационные приложения. М.: Горячая линия — Телеком, 2011. 576 с.
11. 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. *EEE Trans. Networking*, 1997, vol. 5(1), pp. 71–86.
12. Zwart A. P. Queueing systems with heavy tails/ eindhoven university of technology, 2001. 227 p.
13. Костромицкий А. И., Волотка В. С. Подходы к моделированию самоподобного трафика. Восточно-Европейский журнал передовых технологий. 2010. № 46. С. 46–49.
14. Yeryomin R., Petersons E. Transient process relaxation time research under the condition of self-similar traffic input in wireless networks. *Automatic Control and Computer Sciences*, 2009, vol. 43, no. 3, pp. 138–147.
15. Choi B. D., Kim B., Wee I. Asymptotic behavior of loss probability in GI/M/1/K queue as K-tends to infinity. *Queueing Systems*, 2000, vol. 36, pp. 437–442.
16. Guanghui He, Yuan Gao I., Hou Jennifer H., Kihong Park. A case for exploiting self-similarity of network traffic in TCP congestion control. *The International Journal of Computer and Telecommunications Networking*, 2004, vol. 45, issue 6, pp. 743–766.
17. Осин А. В. Сравнительный анализ методик оценки самоподобности телекоммуникационного трафика. Радиотехника и информатика: сб. науч. трудов. М.: МГУС. 2002. С. 37–46.
18. Crovella M. E., Bestavros A. Self-similarity in world wide web traffic evidence and possible causes. Proc. of the 1996 ACM SIGMETRICS: Intern. conf. on Measurement and modeling of computer systems, May, 1996, *IEEE/ACM Trans on Networking*, 1997, vol. 5, no. 6, pp. 835–846.
19. Zadorozhnyi V. N. *Simulation modeling of fractal queues, in dynamics of systems, mechanisms and machines (dynamics)*, 2014, pp. 1–4. doi:10.1109 / Dynamics.2014.7005703
20. Kutuzov O., Tatarnikova T. Evaluation and comparison of classical and fractal queueing systems. *XV International Symposium "Problems of Redundancy in Information and Control Systems"*, September 26–29, 2016.
21. Kihong P., Willinger W. Self-Similar network traffic and performance evaluation. New York, John Wiley & Sons, Inc., 2000. 576 p.
22. Задорожный В. Н. О качестве программных генераторов случайных чисел. Омский научный вестник. 2009. № 2 (80). С. 199–205.
23. Богатырев В. А., Кармановский Н. С., Попцова Н. А., Паршутина С. А., Воронина Д. А., Богатырев С. В. Имитационная модель поддержки проектирования инфокоммуникационных резервированных

систем. Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 5(105). С. 831–838. doi:10.17586/2226-1494-2016-16-5-831-838

24. Тархов Д. А. Нейросетевые модели и алгоритмы. М.: Радиотехника, 2014. 349 с.

25. Привалов А. Ю., Благоев А. В. Об использовании некоторых моделей самоподобного сетевого трафика в имитационном моделировании // Математическое моделирование. 2011. Т. 23. № 7. С. 114–128.

UDC 004.732

doi:10.31799/1684-8853-2018-5-35-43

Statistical methods for studying network traffic

T. M. Tatarnikova^a, Dr. Sc., Tech., Associate Professor, orcid.org/0000-0002-6419-0072, tm-tatarn@yandex.ru

^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

Introduction: The single multiservice network concept which involves the integration of voice, data and multimedia communication has prompted interest in studying the nature of network traffic. Studies of traffic traces recorded on a large scale show the presence of a self-similar structure in it, which requires a revision of the results of modeling infocommunication networks under the assumption of a Poisson data flow. **Purpose:** Studying a sequence of application methods for studying the network traffic nature, identifying the self-similar traffic nature in the form of statistical estimates and the Hurst index. This should offer tools for generating artificial traffic which adequately reflects a real network traffic, taking into account the revealed self-similarity properties. **Results:** The self-similarity properties of the considered 3G traffic were checked on different time scales obtained by aggregation of 5, 10, 15 and 20 minutes on the available daily traffic of 3G data. An estimate of the tail severity for self-similar traffic distribution was obtained by constructing a regression line for the additional distribution function on a logarithmic scale. The self-similarity parameter value, determined by the severity of the distribution “tail”, made it possible to confirm the assumption of 3G traffic self-similarity. A review of models simulating real network traffic with a self-similar structure was performed. Tools were implemented for generating artificial traffic in accordance with the considered models. Various artificial network traffic generators were compared, according to the least squares method criterion, for approximating the artificial traffic point values by the approximation function of 3G traffic. Qualitative assessments of the traffic generators were taken into account, in the form of their software implementation complexity, which, however, can be a subjective assessment. Comparative characteristics allow you to choose a generator which most faithfully simulates real network traffic. **Practical relevance:** The proposed sequence of methods to study the network traffic properties is necessary for understanding its nature and for developing appropriate models which simulate real network traffic.

Keywords — network traffic, self-similar process, long-term dependence, slowly decaying dispersion, heavy tail distribution, self-similarity parameter, self-similar traffic simulation models.

Citation: Tatarnikova T. M. Statistical methods for studying network traffic. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2018, no. 5, pp. 35–43 (In Russian). doi:10.31799/1684-8853-2018-5-35-43

References

1. Tanenbaum A., Wetherall D. *Computer Networks*. Prentice Hall, 2010. 960 p.
2. Kutuzov O. I., Tatarnikova T. M. *Infokommunikacionnye seti. Modelirovanie i ocenka veroyatnostno-vremennykh kharakteristik* [Infocommunication Network. Simulation and Evaluation of Probability-time Characteristics]. SPb., GUAP Publ., 2013. 148 p. (In Russian).
3. Krylov V. V., Samohvalova S. S. *Teoriya teletrafika i ee prilozheniya* [Teletraffic theory and its applications]. SPb., BHV-Peterburg, 2005. 288 p. (In Russian).
4. Kutuzov O. I., Tatarnikova T. M. *Modelirovanie sistem i setei telekommunikatsii* [Simulation of Systems and Telecommunications Networks]. SPb., RGGMU Publ., 2012. 134 p. (In Russian).
5. Petrov V. V. *Chto vy hoteli znat' o samopodobnykh processah, no stesnjalis' sprositi'* [What You Wanted to Know about Self-similar Processes, but Hesitated to Ask]. Moscow, Radiotekhnika Publ., 2003. 112 p. (In Russian).
6. Addie R. G., Neame T. D., Zukerman M. Performance evaluation of a queue fed by a poisson pareto burst process. *Computer Networks*, 2002, vol. 40, no. 3, pp. 377–397.
7. Shelukhin O. I., Osin A. V., Smol'skij S. M. *Samopodobie i fraktaly. Telekommunikacionnye prilozheniya* [Self-similarity and Fractals. Telecommunication Applications]. Moscow, Fizmatlit Publ., 2008. 368 p. (In Russian).
8. Erramilli A., Narayan O., Willinger W. Experimental queueing analysis with long-range dependent traffic. *IEEE/ACM Trans. Networking*, 1996, vol. 4, pp. 209–223.
9. Leland W. E., Taqqu M. S., Willinger W., Wilson D. V. On the self-similar nature of ethernet traffic. *Proc. ACM SIGCOMM'93*. San-Fransisko, 1993, pp. 183–193.
10. Shelukhin O. I. *Mul'tifraktaly. Infokommunikacionnye prilozheniya* [Multifractals. Infocommunication applications]. Moscow, Goryachaya liniya — Telekom Publ., 2011. 576 p. (In Russian).
11. 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. *EEE Trans. Networking*, 1997, vol. 5(1), pp. 71–86.
12. Zwart A. P. *Queueing systems with heavy tails/ eindhoven university of technology*, 2001. 227 p.
13. Kostromickij A. I., Volotka V. S. Approaches to modeling self-similar traffic. *Vostochno-Europeiskii zhurnal perezodovyykh tekhnologii* [Eastern-European Journal of Enterprise Technologies], 2010, no. 46, pp. 46–49 (In Russian).
14. Yeryomin R., Petersons E. Transient process relaxation time research under the condition of self-similar traffic input in wireless networks. *Automatic Control and Computer Sciences*, 2009, vol. 43, no. 3, pp. 138–147.
15. Choi B. D., Kim B., Wee I. Asymptotic behavior of loss probability in GI/M/1/K queue as K-tends to infinity. *Queueing Systems*, 2000, vol. 36, pp. 437–442.
16. Guanghui He, Yuan Gao I., Hou Jennifer H., Kihong Park. A case for exploiting self-similarity of network traffic in TCP congestion control. *The International Journal of Computer and Telecommunications Networking*, 2004, vol. 45, issue 6, pp. 743–766.
17. Osin A. V. Comparative analysis of methods for assessing the self-similarity of telecommunication traffic. *Vestnik MGUS* [Bulletin of MGUS], 2002, pp. 37–46.
18. Crovella M. E., Bestavros A. Self-similarity in world wide web traffic evidence and possible causes. *Proc. of the 1996*

- ACM SIGMETRICS: Intern. conf. on Measurement and modeling of computer systems, May, 1996, *IEEE/ACM Trans on Networking*, 1997, vol. 5, no. 6, pp. 835–846.
19. Zadorozhnyi V. N. *Simulation modeling of fractal queues, in dynamics of systems, mechanisms and machines (dynamics)*, 2014, pp. 1–4. doi:10.1109/Dynamics.2014.7005703
 20. Kutuzov O., Tatarnikova T. Evaluation and comparison of classical and fractal queuing systems. *XV International Symposium "Problems of Redundancy in Information and Control Systems"*, September 26–29, 2016.
 21. Kihong P., Willinger W. Self-Similar network traffic and performance evaluation. New York, John Wiley & Sons, Inc., 2000. 576 p.
 22. Zadorozhnyi V. N. On the quality of program generators of random numbers. *Omskij nauchnyj vestnik* [Omsk Scientific Herald], 2009, no. 2 (80), pp. 199–205 (In Russian).
 23. Bogatyrev V. A., Karmanovsky N. S., Poptcova N. A., Parshutina S. A., Voronina D. A., Bogatyrev S. V. Simulation model for design support of infocomm redundant systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 5, pp. 831–838 (In Russian). doi:10.17586/2226-1494-2016-16-5-831-838/
 24. Tarhov D. A. *Nejrosetevye modeli i algoritmy. Spravochnik*. [Neural network models and algorithms. Directory]. Moscow, Radio Engineering Publ. 2014. 349 p. (In Russian).
 25. Privalov A. Yu., Blagov A. V. About some self-similar telecommunication traffic models used in traffic simulation. *Matematicheskoe modelirovanie* [Math modeling], 2011, vol. 23, no. 7, pp. 114–128 (In Russian).

Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, электронные адреса авторов, которые по требованию ВАК должны быть опубликованы на страницах журнала. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы. Предоставляйте подрисовочные подписи и названия таблиц на русском и английском языках.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно, в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени — эта информация будет опубликована в ссылке на первой странице.

Формулы набирайте в Word, не используя формульный редактор (Mathtype или Equation), при необходимости можно использовать формульный редактор; для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта никогда не пользуйтесь вкладкой Other..., используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; в формулах не отделяйте пробелами знаки: + = -.

Для набора формул в Word никогда не используйте Конструктор (на верхней панели: «Работа с формулами» — «Конструктор»), так как этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Иллюстрации предоставляются отдельными исходными файлами, поддающимися редактированию:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio (*.vsd, *.vsdx); Coreldraw (*.cdr); Excel (*.xls); Word (*.docx); Adobe Illustrator (*.ai); AutoCad (*.dxf); Matlab (*.ps, *.pdf или экспорт в формат *.ai);

— если редактор, в котором Вы изготавливаете рисунок, не позволяет сохранить в векторном формате, используйте функцию экспорта (только по отношению к рисунку), например, в формат *.ai, *.esp, *.wmf, *.emf, *.svg;

— фото и растровые — в формате *.tif, *.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подрисовочных подписей обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате *.tif, *.png с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40×55 мм;

— экспертное заключение.

Список литературы составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Список литературы оформляйте двумя отдельными блоками по образцам lit.dot на сайте журнала (<http://i-us.ru/paperrules>): Литература и References.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Оформление статей».

Контакты

Куда: 190000, Санкт-Петербург,
Б. Морская ул., д. 67, ГУАП, РИЦ
Кому: Редакция журнала «Информационно-управляющие системы»
Тел.: (812) 494-70-02
Эл. почта: i.us.spb@gmail.com
Сайт: www.i-us.ru