

УДК 004.942

АНАЛИЗ ВРЕМЕННЫХ ХАРАКТЕРИСТИК НЕПУАССОНОВСКОГО ТРАФИКА

Н. Ф. Бахарева, доцент кафедры «ПО и управление в технических системах» ПГУТИ, к. т.н.; vt@ist.psati.ru

И. В. Карташевский, ассистент кафедры «ПО и управление в технических системах» ПГУТИ

Ключевые слова: трафик мультисервисных сетей, корреляция временных интервалов, процессы восстановления.

Введение. Методы практического исследования трафика современных телекоммуникационных и вычислительных сетей базируются на применении таких средств, как TraffPro, NetUp UTM 6 — OSS/BSS, CommTraffic [1—3] и др. Использование этих методов, как правило, приводит к формированию трасс, являющихся реализациями целочисленного случайного процесса, каждый отчет которого есть количество событий, зафиксированных в выбранный измерительный интервал времени (Δ). Анализ таких трасс позволяет во многих случаях установить свойство самоподобия трафика [4], которое искажает аналитическую картину функционирования сети, полученную с помощью методов теории массового обслуживания, подразумевающих задание вероятностных характеристик интервалов времени между событиями и интервалов времени обслуживания.

В данной статье предпринята попытка установить некоторые статистические характеристики интервалов времени между событиями путем анализа свойств наблюдаемой трассы.

Расчет коэффициента корреляции. Связь трафика как целочисленного случайного процесса N_i (количество событий в интервале времени $[0..t]$) с последовательностью случайных интервалов времени $\{T_i\}$ между событиями устанавливается на основе соотношения [5]:

$$P\{N_i < n\} = P\{T_1 + T_2 + \dots + T_n > t\}, \quad (1)$$

где P — символ вероятности; t — момент времени ($t \neq 0$), в который регистрируется число событий N_i (предполагаемая стационарность последовательности $\{T_i\}$).

Будем считать, что случайные величины $\{T_i\}$ имеют плотность распределения вероятности $w(\tau)$ со следующими параметрами: σ_τ^2 — дисперсия; $E(T)$ — среднее значение. Используя (1), спектральную плотность мощности последовательности $\{T_i\}$ можно представить в виде [5]:

$$G(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \rho_k e^{-i\omega k} = \frac{1}{\pi} \cdot \frac{\int_0^{\infty} [\varphi(e^{i\omega}, t) + \varphi(e^{-i\omega}, t)] dt}{2 \int_0^{\infty} \varphi(0, t) dt - E(T)},$$

где $\varphi(z, t)$ — производящая функция последовательности N_i :

$$\varphi(z, t) = \sum_{i=0}^{\infty} z^i P\{N_i = i\};$$

ρ_k — коэффициент корреляции случайной последовательности $\{T_i\}$:

$$\rho_k = \frac{E(T)}{\sigma_\tau^2} \left[\int_0^{\infty} p(k, t) dt - E(T) \right], \quad k = 1, 2, \dots \quad (2)$$

При этом $E(T)$ можно найти из соотношения

$$\frac{dE(N_i)}{dt} = \frac{1}{E(T)},$$

а σ_τ^2 как $\sigma_\tau^2 = E(T)\{2\varphi^*(0, 0) - E(T)\}$, где $\varphi^*(z, \xi)$ — преобразование Лапласа $\varphi(z, t)$, а

$$p(k, t) = \frac{1}{k!} \cdot \left. \frac{d^k \varphi(z, t)}{dz^k} \right|_{z=0}.$$

Представленные соотношения связывают статистические свойства последовательностей N_i и $\{T_i\}$. В частности, коэффициенты корреляции ρ_k , характеризующие в определенной степени самоподобие трафика, могут быть рассчитаны аналитически или численно, если получена хорошая аппроксимация $P\{N_i\}$.

Расчет коэффициентов корреляции ρ_k — довольно сложная задача, так как после регистрации трассы, т. е. последовательности N_i , ни одна из составляющих формулы (2) не может быть получена непосредственно.

Обратимся к выражению $\int_0^{\infty} p(k, t) dt$. Следует заметить, что при анализе реальной конечной трассы интегрирование будет осуществляться на конечном интервале $[0, T_a]$. Физический смысл величины $p(k, t)$ раскрывает соотношение [5]:

$$p(i; t) = P\{N_i = i\} = P(W + T_1 + \dots + T_i > t) - P(W + T_1 + \dots + T_{i-1} > t), \quad (i = 1, 2, \dots), \quad (3)$$

где W — некоторый случайный интервал времени от начала отсчета до появления первого события. (Учет W в рассматриваемой последовательности наделяет ее нестационарностью, пренебречь которой можно уже после получения нескольких первых значений i .) Из (3) следует, что $p(i; t) = P\{N_i = i\}$ есть вероятность наступления i событий на интервале времени $[0..t]$.

Таким образом, если, например, $i = 1$, то для нахождения коэффициента ρ_1 (коэффициент корреляции соседних интервалов $\{T_i\}$) интервал $[0, T_a]$ можно разбить на последовательность пересекающихся интервалов: $[0..t_1], [0..t_2], \dots, [0..t_k], \dots, [0..T_a]$, $t_1 < t_2 < \dots < t_k < \dots < T_a$, и на каждом интервале найти вероятность наступления одного события $p(1, t_k)$ и вычислить интеграл $\int_0^{T_a} p(1, t) dt$. Данные интервалы целесообразно находить из соотношения $t_{k+1} - t_k = \frac{T_a}{K}$, где K — некоторое целое число, выбираемое исходя из условия, при котором размер интервала позволяет накопить статистику для N_i , достаточную для достоверного вычисления величины $p(1, t_k)$. При этом измерительных интервалов внутри каждого интервала $[0..t_k]$ должно быть достаточно для нахождения вида распределения последовательности N_i , а интеграл $\int_0^{T_a} p(1, t) dt$ придется вычислять приближенно с использованием квадратурных формул [6].

Оставшиеся в формуле (2) для коэффициента корреляции величины $E(T)$ и σ_τ^2 , очевидно, не могут быть найде-

ны по реализации N_i , поскольку неизвестна плотность $w(\tau)$. В данной ситуации можно воспользоваться рекомендациями [7], согласно которым при достаточно долгой реализации N_i для ненаблюдаемой последовательности $\{T_i\}$ справедливы условия применимости центральной предельной теоремы, что позволяет для $E(T)$ и σ_τ^2 использовать приближенные выражения:

$$E(T) = \frac{T_a}{N}; \tag{4}$$

$$\sigma_\tau^2 = \frac{\sigma_N^2 (E(T))^3}{T_a}, \tag{5}$$

где N — общее число событий на интервале T_a ; σ_N^2 — выборочная дисперсия реализации N_i .

Очевидно, что по данной методике могут быть оценены и другие коэффициенты корреляции ($\rho_i, i = 2, 3, \dots$) с учетом общего соотношения (3).

Проверка применимости методики оценки коэффициента корреляции интервалов времени между событиями была проведена на искусственно смоделированной трассе с показательным законом распределения $\{T_i\}$ и заданным коэффициентом корреляции ρ_i :

$$w(\tau) = \frac{1}{2\sigma_0^2} \exp\left(-\frac{\tau}{2\sigma_0^2}\right); \tag{6}$$

$$E(\tau) = 2\sigma_0^2; \tag{7}$$

$$\sigma_\tau^2 = 4\sigma_0^4. \tag{8}$$

Моделирование коррелированной последовательности $\tau(n)$ с показательной плотностью (6) осуществлялось по алгоритму [8]:

$$\tau(n) = \xi_1^2(n) + \xi_2^2(n) = \sigma_0^2 \left[(\sqrt{1-r^2}x_1(n) + r\xi_1(n-1))^2 + (\sqrt{1-r^2}x_2(n) + r\xi_2(n-1))^2 \right],$$

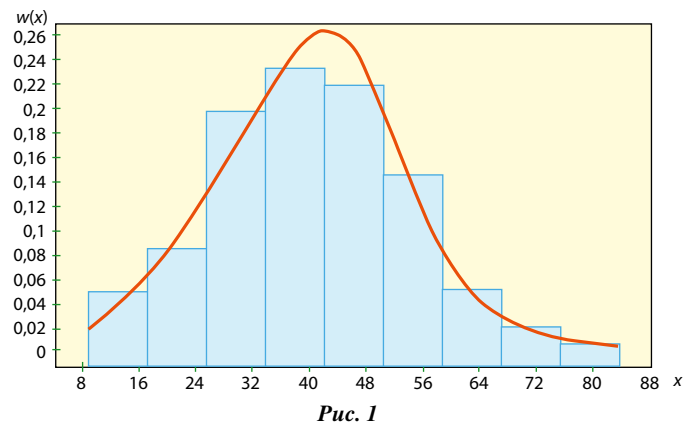
где $x_1(n)$ и $x_2(n)$ — последовательности независимых нормальных случайных чисел с нулевым средним и единичной дисперсией, а $r^2 = \rho_1$.

При параметрах $\sigma_0^2 = 0,25$ и $\rho_1 = 0,64$ была смоделирована трасса общим количеством 250 тыс. интервалов $\tau(n)$ на интервале времени $T_a = 125512$ условных единиц времени; измерительный интервал выбран в виде $\Delta = 20$ и $K = 7$. Выбор нечетного значения K обусловлен тем, что численное интегрирование по методу Ньютона-Котеса, используемое в (2), обладает большей точностью именно при нечетном числе интервалов.

Анализ распределений для реализаций N_i на интервалах $[0..t_k]$, $k = \overline{1,7}$ проводился с использованием программного продукта EasyFit [9]. Иллюстративный материал для расчета вероятности наступления одного события, например в пятом интервале ($k = 5$) $p(1;t_5) = P\{N_5 = 1\}$, представлен на рис. 1, где приведены гистограмма распределения N_i и аппроксимирующая ее плотность $w(x)$.

Ситуация, изображенная на рис. 1, соответствует случаю, когда гистограмма аппроксимируется распределением Dagum (каппа-распределение) с параметрами $k = 0,31682$, $\alpha = 9,1205$, $\beta = 51,17$. Плотность вероятности данного распределения имеет вид

$$w(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha k - 1}}{\beta \left[1 + \left(\frac{\alpha}{\beta}\right)^\alpha\right]^{k+1}}, \quad 0 \leq x < \infty.$$



Соответственно интегральная функция записывается как

$$F(x) = \left[1 + \left(\frac{\alpha}{\beta}\right)^{-\alpha}\right]^{-k}.$$

Принятие гипотезы об истинности данного распределения осуществляется по критерию Колмогорова–Смирнова с вероятностью 0,77. В данной ситуации распределение Dagum оказалось наиболее подходящим по данному критерию из возможных 59, анализируемых программой EasyFit. Вероятность наступления одного события при этом вычислялась как

$$p(1;t_5) = F(1) - F(0). \tag{9}$$

Согласно (9) определялись вероятности $p(1;t_k)$ для всех k — от 1 до 7. Вычисление интеграла $\int_0^{T_a} p(1,t)dt$ проводилось с предварительной нормировкой временных переменных из условия $T_a = 1$.

В рассматриваемом примере для вычислений по формуле (2) значения $E(T)$ и σ_τ^2 определялись не по (4) и (5), а согласно (7) и (8), проверенным по смоделированной последовательности τ_i . Это позволило исключить влияние приближенных формул (4) и (5) на результаты основного статистического анализа.

Вычисления по формуле (2) для коэффициента корреляции ρ_1 дали $\rho_1 = 0,68$, что в сравнении с моделируемым значением $\rho_1 = 0,64$ следует признать вполне удовлетворительным результатом.

Далее по предлагаемой методике определялся коэффициент корреляции агрегированного трафика абонентов сети IP-телефонии одного из операторов связи [10]. Трафик имел вид последовательности N_i , т. е. являлся целочисленным случайным процессом при $\Delta = 1$ мин. Из всей трассы (рис. 2) были выделены периоды активности пользователей, длящиеся приблизительно 12 часов, с 8.00 до 20.00. Как и в случае моделирования трассы, для удобства последующего интегрирования

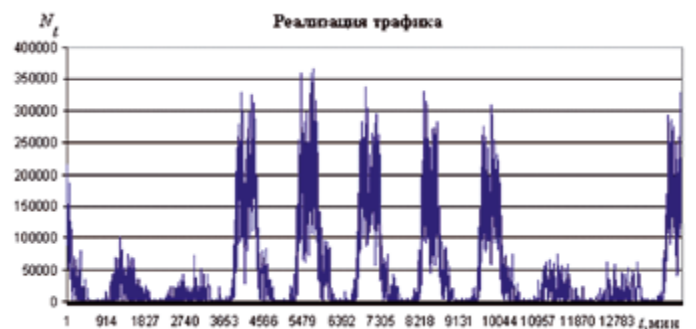


Рис. 2

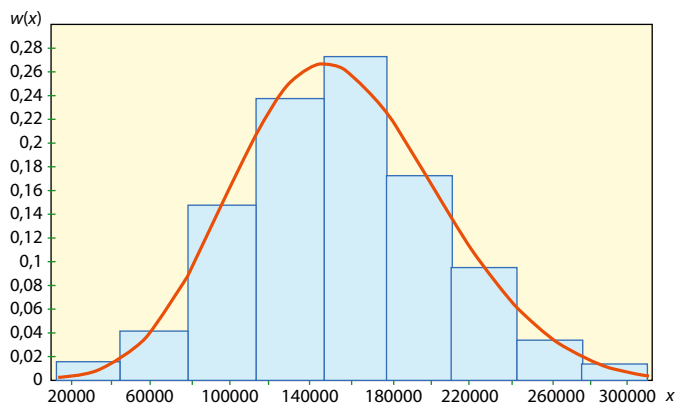


Рис. 3

было выбрано нечетное число K , равное 7. Таким образом, длина интервала анализа составила 700 минут.

При анализе пересекающихся интервалов $[0...t_1], [0...t_2]...[0...t_k]...[0...T_a]$, $t_1 < t_2 < \dots < t_k < \dots < T_a$ наиболее вероятным оказалось распределение Generalized extreme value с функцией плотности вероятности

$$w(x) = \begin{cases} \frac{1}{\sigma} \exp\left(-\left(1+k\frac{x-\mu}{\sigma}\right)^{-1/k}\right) \left(1+k\frac{x-\mu}{\sigma}\right)^{-1-1/k}, & k \neq 0; \\ \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma} - \exp\left(-\frac{x-\mu}{\sigma}\right)\right), & k = 0; \end{cases}$$

$$\begin{cases} 1+k\frac{x-\mu}{\sigma} > 0, & k \neq 0; \\ -\infty < x < +\infty, & k = 0 \end{cases}$$

и интегральной функции

$$F(x) = \begin{cases} \exp\left(-\left(1+k\frac{x-\mu}{\sigma}\right)^{-1/k}\right), & k \neq 0; \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), & k = 0. \end{cases}$$

На рис. 3 изображена гистограмма и аппроксимирующее ее распределение Generalized extreme value с параметрами $k = -0,2119$; $\sigma = 47602,5$; $\mu = 135307,1$ для случая $p(1; t_5) = P\{N_5 = 1\}$.

Принятие гипотезы об истинности данного распределения осуществлялось по критерию Колмогорова–Смирнова с вероятностью 0,9. Далее, согласно (9), определялись вероятности $p(1; t_k)$ для всех k от 1 до 7, а значения $E(T)$ и σ_T^2 вычислялись по формулам (4) и (5). Вычисления по формуле (2) для коэффициента корреляции ρ_1 дали $\rho_1 = 0,00008$, что говорит о практическом отсутствии корреляции интервалов времени, т. е. последовательность этих интервалов представляет собой процесс восстановления [4].

Аналогичным образом, с учетом рекомендации Международного союза электросвязи (ITU-T Rec. E.502 (02/2001)), был проанализирован трехчасовой интервал. Число K также было выбрано нечетным, но уже равным 5, длина интервала составила 180 минут.

При анализе пересекающихся интервалов $[0...t_1], [0...t_2]...[0...t_k]...[0...T_a]$, $t_1 < t_2 < \dots < t_k < \dots < T_a$ наиболее вероятным оказалось трехпараметрическое распределение Вейбулла с функцией плотности вероятности

$$w(x) = \frac{\alpha}{\beta} \left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x-\gamma}{\beta}\right)^\alpha}, \quad \gamma \leq x < \infty.$$

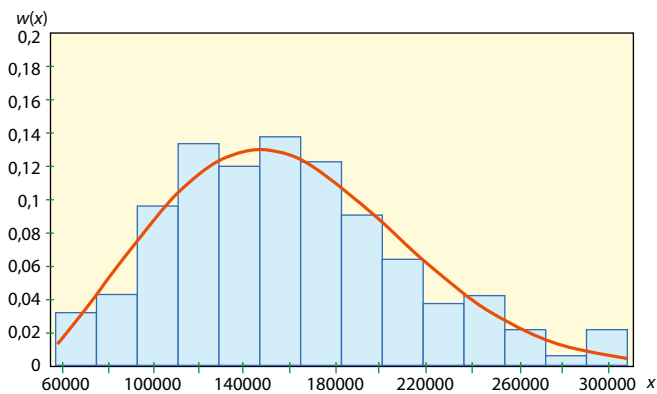


Рис. 4

На рис. 4 изображено трехпараметрическое распределение Вейбулла с параметрами $\alpha = 2,3158$, $\beta = 127520$, $\gamma = 46676$ для случая $p(1; t_5) = P\{N_5 = 1\}$.

Вычисления по формуле (2) привели к результату $\rho_1 = 0,00012$. Несмотря на полуторакратное увеличение коэффициента корреляции, его значение остается достаточно малым, поэтому последовательность интервалов времени можно отнести к процессам восстановления.

Заключение. Разработанная методика анализа сетевого трафика как целочисленного случайного процесса позволяет оценить корреляцию интервалов времени между событиями (пакетами). В случае обнаружения корреляционных связей для последовательности временных интервалов появляется возможность расчета энергетического спектра последовательности и далее, вплоть до оценки плотности вероятности временных интервалов, соответствующих анализируемому трафику [11].

ЛИТЕРАТУРА

1. <http://www.traffpro.ru/4—josjonn-traffpro.html> (дата обращения 27.05.2010).
2. <http://www.netup.ru/UTM6> (дата обращения 12.06.2010).
3. <http://www.tamos.ru/htmlhelp/commtraffic> (дата обращения 27.05.2010).
4. Шелухин О.И., Тенякшев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях.— М.: Радиотехника, 2003.— 480 с.
5. Кокс Д., Льюис П. Статистический анализ последовательностей событий.— М.: Мир, 1969.— 312 с.
6. Демидович Б.П., Марон И.А. Основы вычислительной математики.— М.: Наука, 1966.— 664 с.
7. Клейнрок Л. Вычислительные системы с очередями.— М.: Мир, 1979.— 575 с.
8. Быков В.В. Цифровое моделирование в статистической радиотехнике.— М.: Советское радио, 1971.— 328 с.
9. http://www.mathwave.com/products/easyfit_desc.html (дата обращения 16.12.2009).
10. Карташевский И.В., Криштофович А.Ю., Фомин В.В. Анализ и исследование трафика протокола SIP//Труды 12-й международной конференции «Цифровая обработка сигналов и ее применение».— Москва, 2010.— Вып. 12-1.— С. 258—259.
11. Тарасов В.Н., Карташевский И.В. Анализ характеристик трафика спектральным методом//Труды 65-й научной сессии, посвященной Дню радио.— Москва, 2008.— С. 127—129.

Получено 22.10.10