УДК 681.324

П.Е. Пустовойтов, Н.И. Ящук

Национальный технический университет «ХПИ», Харьков

ФОРМИРОВАНИЕ САМОПОДОБНОГО СЛУЧАЙНОГО ПРОЦЕССА С ЗАДАННЫМИ СВОЙСТВАМИ

Разработана методика получения самоподобного потока из входящего потока Эрланга. Характер поведения самоподобного потока задается параметром Херста. Методика была апробирована в имитационной модели.

Ключевые слова: *самоподобный случайный процесс, имитационное моделирование.*

Введение

Традиционные технологии анализа и оценки эффективности функционирования компьютерных сетей основаны на использовании тех или иных математических моделей систем массового обслуживания (СМО), отличающихся друг от друга аналитическим описанием входящего потока требований, структурой системы, дисциплиной обслуживания и характером законов распределения продолжительностей обслуживания и ожидания начала обслуживания. При этом наиболее существенные результаты теории массового обслуживания связаны с использованием марковских моделей случайного процесса, описывающего входящий поток, а также процессы обслуживания и ожидания. Однако сравнительно недавно (в 1993г) появилась статья [1], разрушившая иллюзию о том, что предположение о пуассоновском характере входящего потока обеспечивает адекватную модель сетевого трафика. Появившиеся основные претензии к пуассоновской модели потока связаны не только (и не столько) с тем, что реальный поток не является стационарным. Гораздо более серьезное отличие реального потока от идеализированной пуассоновской модели состоит в том, что этот поток – не марковский, он обладает последействием. Это принципиальное свойство потока данных, поступающих на вход компьютерной сети, является следствием нового качества потока – самоподобия.

Постановка задачи исследования

Исходный процесс является самоподобным, если его статистические характеристики (математическое ожидание, дисперсия, корреляционная функция) близки к соответствующим характеристикам агрегированного процесса. Кроме того, свойство самоподобия проявляется в том, что среднее поведение такого процесса на коротком промежутком времени не отличается от поведения на длительном промежутке времени, то есть самоподобный процесс инвариантен относительно изменений масштаба.

При рассмотрении самоподобных процессов с формальных позиций вводится так называемый параметр Херста – $H \in (0,5 < H < 1)$ [2]. С учетом этого параметра случайный процесс $X(t)$ называется статистически самоподобным с параметром H, если для любого вещественного $a > 0$ процесс $a^{-H}X(at)$ обладает теми же статистическими характеристиками, что и процесс $X(t)$, то есть имеют место равенства средних

$$M[X(t)] = a^{-H}M[X(at)],$$

дисперсий

$$D[X(t)] = a^{-2H}D[X(at)],$$

автокорреляционных функций

$$R_X(t,s) = a^{-2H}R_X(at,as),$$

где t и s – произвольные моменты времени в пределах интервала наблюдений.

Характерной чертой самоподобного процесса является долгосрочная зависимость процесса, наблюдаемого на каком-либо интервале, от значений этого процесса на предыдущих интервалах. Эта зависимость проявляется в более медленном убывании значений автокорреляционной функции такого процесса по сравнению с процессами, для которых имеет место краткосрочная зависимость.

Основные результаты

Нормированная автокорреляционная функция для самоподобного процесса описывается выражением

$$r_X(\tau) = (1+\tau)^{-\frac{\beta}{2}}, \qquad (1)$$

где показатель степени β связан параметром Херста H следующим соотношением

$$\beta = 2(1-H). \qquad (2)$$

Тогда

$$r(\tau) = (1+\tau)^{-(1-H)}. \qquad (3)$$

Для несамоподобного процесса $H = 0,5$ и $\beta = 1$. При наличии самоподобия параметр $H > 0,5$. Тогда нормированная автокорреляционная функция убывает медленнее, чем это имеет место для пуассоновского процесса, то есть появляется долгосрочная зависимость. Эта долгосрочная зависимость применительно к процессам передачи данных проявляет себя не только в медленном убывании автокорреляционной функции, но и в характере распределения случайной величины интервалов между поступлениями пакетов.

Считается, что распределение переменной X медленно затухает, если

$$\text{Вер}(X > x) = 1 - F(x) \sim x^{-\alpha}, \quad \alpha > 0.$$

Самое простое аналитическое описание медленно затухающего распределения (Парето) имеет вид [3]:

$$F(x) = 1 - \left(\frac{k}{x}\right)^{\alpha}, \quad x > k, \ \alpha > 0. \qquad (4)$$

Соответствующая плотность распределения описывается выражением

$$f(x) = \frac{\alpha}{k}\left(\frac{k}{x}\right)^{\alpha+1}. \qquad (5)$$

Основная претензия к Парето-распределению, как к модели самоподобного процесса состоит в следующем. Для этого распределения вероятность появления больших случайных величин тем меньше, чем больше их значение, что для самоподобного процесса не характерно. Приемлемой альтернативой является поток Эрланга надлежащего порядка. Плотность распределения длины интервала между событиями для потока Эрланга n -порядка имеет вид

$$f_n(x) = \frac{\lambda^n x^{n-1}}{(n-1)!}e^{-\lambda x}, \ x > 0. \qquad (6)$$

Наличие степенного сомножителя в описании плотности (6) приводит к замедлению ее убывания и появлению типичного для самоподобных процессов «тяжелого» хвоста.

В [3] отмечается, что наличие тяжелого хвоста является основной причиной долгосрочной зависимости и самоподобия сетевого трафика.

Специфические особенности трафика с медленно затухающим распределением длин пакетов и интервалов между ними негативно сказывается на качестве функционирования узла компьютерной сети. Высокая дисперсия такого распределения приводит к крайней изменчивости и непостоянству нагрузки. Кроме того, как отмечено в [3], суперпозиция самоподобных потоков не приводит к пуассоновости результирующего потока, что является прямым следствием наличия последействия в составляющих суммарного потока. Это, в частности, приводит к тому, что традиционно используемые соотношения теории массового обслуживания, полученные на основе допущения об экспоненциальности входящего потока, оказываются неверными для самоподобных потоков. Это обстоятельство приводит к необходимости использования для оценки эффективности систем, на вход которых поступает самоподобный поток, имитационных моделей. В связи с этим рассмотрим процедуру формирования самоподобного потока на основе распределения Эрланга n-го порядка. Характеристическое свойство Эрланговского потока состоит в том, что он представляет собой просеянный пуассоновский поток. Это значит, что для получения потока Эрланга, например, 2-го порядка интенсивностью λ достаточно сформировать пуассоновский поток интенсивностью 2λ и извлечь из этого потока каждое четное событие.

Вернемся теперь к описанию второго принципиального свойства самоподобного потока, появившегося в наличии последействия.

В соответствии с (1) для заданного значения параметра Херста H могут быть рассчитаны коэффициенты корреляции между случайными длинами межсобытийных интервалов. Тогда задача формирования потока с заданным показателем самоподобия H заключается в следующем: получить последовательность случайных величин с плотностью распределения Эрланга заданного порядка, коррелированных в соответствии с корреляционной матрицей $K=\{k_{ij}\}$, элементы которой определяются корреляционной функцией (1).

Пусть последовательность случайных величин $\xi_1,\xi_2,\ldots,\xi_n$ имеет плотность распределения Эрланга заданного (например, второго) порядка. Теперь искомую последовательность $x_1,x_2,\ldots,x_n$ случайных величин с тем же математическим ожиданием (например, равным a), но коррелированных в соответствии с матрицей $\{k_{ij}\}$ получим с использованием линейного преобразования

$$\begin{aligned}
&x_1=c_{11}(\xi_1-a)+a;\\
&x_2=c_{12}(\xi_1-a)+c_{22}(\xi_2-a)+a;\\
&x_3=c_{13}(\xi_1-a)+c_{23}(\xi_2-a)+c_{33}(\xi_3-a)+a;\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&x_n=c_{1n}(\xi_1-a)+c_{2n}(\xi_2-a)+\ldots+c_{nn}(\xi_n-a)+a.
\end{aligned} \tag{7}$$

Коэффициенты преобразования (7) найдем из условий:

$$M[(x_1-a)^2]=M[c_{11}(\xi_1-a)^2]=$$
$$=c^2{}_{11}M[(\xi_1-a)^2]=c^2{}_{11}a^2=k_{11};$$
$$M[(x_1-a)(x_2-a)]=$$
$$=M[c_{11}(\xi_1-a)(c_{12}(\xi_1-a)+c_{22}(\xi_2-a))]=$$
$$=c_{11}c_{12}M[(\xi_1-a)^2]+$$
$$+c_{11}c_{22}M[(\xi_1-a)(\xi_2-a)]=c_{11}c_{12}a^2=k_{12};$$
$$M[(x_2-a)^2]=M[(c_{12}(\xi_1-a)+c_{22}(\xi_2-a))^2]=$$
$$=c^2{}_{12}M[(\xi_1-a)^2]+$$
$$+2c_{12}c_{22}M[(\xi_1-a)(\xi_2-a)]+c^2{}_{22}M[(\xi_2-a)^2]=$$
$$=c^2{}_{12}a^2+c^2{}_{22}a^2=a^2(c^2{}_{12}+c^2{}_{22})=k_{22};$$
$$M[(x_1-a)(x_3-a)]=$$
$$=M[c_{11}(\xi_1-a)(c_{13}(\xi_1-a)+c_{23}(\xi_2-a)+c_{33}(\xi_3-a))]=$$
$$=c_{11}c_{13}M[(\xi_1-a)^2]+c_{11}c_{23}M[(\xi_1-a)(\xi_2-a)]+$$
$$+c_{11}c_{33}M[(\xi_1-a)(\xi_3-a)]=c_{11}c_{13}a^2=k_{13};,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$M[(x_i-a)(x_j-a)]=M[(c_{1i}(\xi_1-a)+c_{2i}(\xi_2-a)+\ldots+$$
$$+c_{ii}(\xi_i-a))\times\times(c_{1j}(\xi_1-a)+c_{2j}(\xi_2-a)+\ldots+c_{ij}(\xi_i-a)+\ldots+$$
$$+c_{jj}(\xi_j-a))]=(c_{1i}c_{1j}+c_{2i}c_{2j}+\ldots+c_{ii}c_{ij})a^2=k_{ij},\ \ j>i;$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$M[(x_n-a)^2]=$$
$$=M[\left(c_{1n}(\xi_1-a)+c_{2n}(\xi_2-a)+\ldots+c_{nn}(\xi_i-a)\right)^2]=$$
$$=(c^2{}_{1n}+c^2{}_{2n}+\ldots+c^2{}_{nn})a^2=k_{nn}.$$

Полученная система уравнений относительно неизвестных $(c_{11},c_{12},\ldots,c_{1n},c_{21},\ldots,c_{2n},\ldots,c_{1n},\ldots,c_{nn})$ решается рекуррентно. Из первого уравнения системы имеем $c_{11}=\frac{1}{a}k_{11}^{1/2}$. Далее, из второго определим

$$c_{12}=\frac{1}{c_{11}a^2}k_{12}=\frac{1}{a}\frac{k_{12}}{k_{11}^{1/2}}.$$

Из третьего уравнения вычислим

$$c_{22}=\left(k_{22}/a^2-c_{12}^2\right)^{1/2}=\left(k_{22}/a^2-\frac{k_{12}^2}{a^2k_{11}}\right)^{1/2}=$$
$$=\frac{1}{a}\left(\frac{k_{11}k_{12}-k_{12}^2}{k_{11}}\right)^{1/2}.$$

Дальнейшие соотношения приведем без пояснений.

$$c_{13}=\frac{1}{c_{11}a^2}k_{13}=\frac{1}{a}\frac{k_{13}}{k_{11}^{\frac{1}{2}}};\ \ c_{23}=\frac{k_{23}/a^2-c_{12}c_{13}}{c_{22}};$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$c_{ij}=\frac{\left(k_{ij}/a^2-c_{1i}c_{1j}-c_{2i}c_{2j}-\ldots-c_{i-1,i}\,c_{i-1,j}\right)}{c_{ii}},\ \ j>i;$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$c_{jj} = \left(\frac{k_{jj}}{a^2} - c_{1j}^2 - c_{2j}^2 - \ldots - c_{j-1,j}^2 \right)^{1/2};$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$c_{nn}^2 = \left(\frac{k_{nn}}{a^2} - c_{1n}^2 - c_{2n}^2 - \ldots - c_{n-1,n}^2 \right)^{1/2}.$$

Подстановка полученных значений c_{ij}, $i = 1, 2, ..., n$ $j \ge i$, в (7) позволяет рассчитать искомый набор коррелированных случайных величин.

Более простая схема получения коррелированной последовательности случайных величин соответствует ситуации, когда учитывается только корреляция между соседними элементами последовательности. Пусть по-прежнему, в качестве базовой используется некоррелированная последовательность случайных величин $\xi_1, \xi_2, ..., \xi_n$, имеющая плотность распределения Эрланга заданного порядка. Эта последовательность формируется путем просеивания последовательности случайных величин с пуассоновским распределением.

Пусть $M\left[\xi_j\right] = a$, $M\left[\left(\xi_j - a\right)^2\right] = D$. Зададим требуемое значение коэффициента корреляции k между соседними элементами конструируемой последовательности. Искомую коррелированную последовательность получим по формулам

$$\begin{aligned} &\zeta_1 = \xi_1; \\ &\zeta_2 = c_{12}\left(\xi_1 - a\right) + c_{22}\left(\xi_2 - a\right) + a; \\ &\ldots\ldots\ldots\ldots\ldots \\ &\zeta_j = c_{j-1,j}\left(\xi_{j-1} - a\right) + c_{jj}\left(\xi_j - a\right) + a; \\ &\zeta_{j+1} = c_{j,j+1}\left(\xi_j - a\right) + c_{j+1j+1}\left(\xi_{j+1} - a\right) + a; \\ &\ldots\ldots\ldots\ldots\ldots \end{aligned} \tag{8}$$

Запишем систему уравнений для отыскания коэффициентов преобразования.

$$M\left[\frac{1}{D}(\zeta_1 - a)(\zeta_2 - a)\right] =$$

$$= M\left[\frac{1}{D}(\xi_1 - a)\left(C_{12}(\xi_1 - a) + C_{22}(\xi_2 - a)\right)\right] =$$

$$= \frac{C_{12}}{D} M\left[(\xi_1 - a)^2\right] + \frac{C_{22}}{D} M\left[(\xi_1 - a)(\xi_2 - a)\right] = C_{12} = k;$$

$$M\left[\frac{1}{D}(\zeta_2 - a)^2\right] = M\left[\frac{1}{D}\left(C_{12}(\xi_1 - a) + C_{22}(\xi_2 - a)\right)^2\right] =$$

$$= \frac{1}{D} C_{12}^2 M\left[(\xi_1 - a)^2\right] + \frac{1}{D} C_{22}^2 M\left[(\xi_{22} - a)^2\right] +$$

$$+ \frac{2}{D} C_{12} C_{22} M\left[(\xi_1 - a)(\xi_2 - a)\right] = = C_{12}^2 + C^2{}_{22} = 1;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$M\left[\frac{1}{D}\left(\zeta_j - a\right)\left(\zeta_{j+1} - a\right)\right] =$$

$$= M\left[\frac{1}{D}\left[C_{j-1,j}\left(\xi_{j-1} - a\right) + C_{jj}\left(\xi_j - a\right)\right]\right] \times$$

$$\times \left[C_{j,j+1}\left(\xi_j - a\right) + C_{j+1,j+1}\left(\xi_{j+1} - a\right)\right] = C_{jj} C_{j,j+1} = k;$$

$$M\left[\frac{1}{D}\left(\zeta_{j+1} - a\right)^2\right] =$$

$$= M\left[\frac{1}{D}\left[C_{j,j+1}\left(\xi_j - a\right) + C_{j+1,j+1}\left(\xi_{j+1} - a\right)\right]^2\right] =$$

$$= C_{j,j+1}^2 + C_{j+1,j+1}^2 = 1;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

Полученная система уравнений легко решается.

$$\begin{aligned} &C_{12} = k; \qquad C_{22} = \left(1 - C_{12}^2\right)^{1/2}; \\ &C_{23} = \frac{k}{C_{22}}; \quad C_{33} = \left(1 - C_{23}^2\right)^{1/2}; \\ &------------------------------ \\ &C_{j,j+1} = \frac{k}{C_{jj}}; \quad C_{j+1,j+1} = \left(1 - C_{j,j+1}^2\right)^{1/2}; \\ &------------------------------ \end{aligned} \tag{9}$$

С использованием этих соотношений построим последовательность случайных величин $\zeta_1, \zeta_2, \ldots$.

Проверка качества получаемой коррелированной последовательности случайных величин осуществляется следующим образом. Пусть задан набор элементов последовательности $x_1, x_2, ..., x_n$.

Рассчитаем эмпирические значения для среднего $m = \frac{1}{n}\sum_{j=1}^{n} x_j$ и дисперсии $S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - m)^2$. Кроме того, для каждого $j = 1, 2, ..., n$ определяем интегральное отклонение $D_j = \sum_{k=1}^{j} x_k - jm$.

Теперь оценим уровень изменчивости заданной случайной последовательности длиной n по формуле

$$R_n = \max_j D_j - \min_j D_j.$$

В [2] было показано, что для самоподобных процессов имеет место соотношение

$$R_n / S_n \cong (n/2)^H, \tag{10}$$

где H – параметр самоподобия.

Из (10) следует

$$\log\left(R_n / S_n\right) \cong H \log(n/2). \tag{11}$$

Соотношение (11) позволяет оценить значение параметра Херста для любой последовательности случайных величин. С этой целью для возрастающей последовательности $n_1, n_2, ..., n_p$ наборов случайных величин рассчитаем соответствующие значения $\frac{R_{n_1}}{S_{n_1}}, \frac{R_{n_2}}{S_{n_2}}, ..., \frac{R_{n_p}}{S_{n_p}}$. Тогда наилучшая в смысле наименьших квадратов оценка параметра H определяется путем минимизации

$$J = \sum_{\ell=1}^{p}\left(\log\frac{R_{n_\ell}}{S_{n_\ell}} - H\log\frac{n_\ell}{2}\right)^2,$$

откуда $\frac{dJ}{dH} = -2\sum_{\ell=1}^{P}\left(\log\frac{R_{n_\ell}}{S_{n_\ell}} - H\log\frac{n_\ell}{2}\right) = 0$ и

$$\hat{H} = \sum_{\ell=1}^{P}\log\frac{R_{n_\ell}}{S_{n_\ell}} \Bigg/ \sum_{\ell}^{P}\log\frac{n_\ell}{2} = \log\prod_{\ell=1}^{P}\frac{R_{n_\ell}}{S_{n_\ell}} \Bigg/ \log\prod_{\ell=1}^{P}\frac{n_\ell}{2}. \quad (12)$$

В конкретном эксперименте формировались последовательности коррелированных случайных величин с заданными значениями $H_1 = 0,5$, $H_2 = 0,65$, $H_3 = 0,8$.

При этом вначале путем просеивания пуассоновского потока с заданной интенсивностью формировался набор случайных величин $\xi_1, \xi_2, \xi_3, \dots$, распределенных в соответствии с законом Эрланга второго порядка. Даже с использованием (8), (9) осуществляется построение самоподобных потоков с коэффициентом корреляции, которые рассчитывались по формуле (3). Результаты эксперимента: $\hat{H}_1 = 0,47$, $\hat{H}_2 = 0,62$, $\hat{H}_3 = 0,78$.

Выводы

В работе рассмотрена задача математического обеспечения процедуры построения имитационных моделей реальных систем, функционирующих в условиях самоподобного входящего потока.

Предложена методика формирования самоподобного потока событий с заданным значением параметра Херста. Вычислительная процедура основана на использовании базовой последовательности случайных величин, имеющих распределение Эрланга надлежащего порядка с «тяжелым хвостом». Требуемое значение корреляции между длинами случайных интервалов между событиями в потоке обеспечивается специальным линейным преобразованием базовой последовательности.

Проведенное исследование качества формируемой последовательности показало высокую эффективность предложенной процедуры.

Список литературы

1. On the Self-similar Nature of Ethernet Traffic / W. Leland, M. Taqqu, W. Willinger, D. Wilson // IEEE/ACM Transactions on Networking – 1994. – Vol. 12. – P. 2-15.

2. Hurst H. Long-term Storage. An Experimental Study / H. Hurst, R. Black, Y. Simaika. – London: Coustable, 1965. – 184 p.

3. Park K. Self-similar Network Traffic and Performance Evaluation / K. Park, W. Willinger. – New York : Wiley, 2000. – 196 p.

Поступила в редколлегию 23.04.2009

Рецензент: д-р техн. наук, проф. Л.Г. Раскин, Национальный технический университет «ХПИ», Харьков.

ФОРМУВАННЯ САМОПОДІБНОГО ВИПАДКОВОГО ПРОЦЕСУ ІЗ ЗАДАНИМИ ВЛАСТИВОСТЯМИ

П.Є. Пустовойтов, Н.І. Ящук

Розроблено методику отримання самоподібного потоку з потоку Ерлангу, який надається. Характер поведінки самоподібного потоку задається параметром Херста. Методику було апробовано у імітаційній моделі.

Ключові слова: *самоподібний випадковий процес, імітаційне моделювання.*

STOCHASTIC SELF-SIMILARITY PROCESS GENERATION WITH GIVEN PARAMETERS

P.E. Pustovoitov, N.I. Yaschuk

The methodic of self-similarity flow receiving from Erlang flow was developed. The behavior of self-similarity flow is set by Hurst-parameter. The methodic was approved in imitation model.

Keywords: *selfsimilar casual process, imitation design.*