

МОДЕЛИРОВАНИЕ САМОПОДОБНОГО ТРАФИКА

© 2006 А.Ю. Привалов, М.В. Баева

Самарский государственный аэрокосмический университет

Предложен метод имитационного моделирования сетевого телекоммуникационного трафика, обладающего самоподобными свойствами. Для генерируемого трафика может быть задан параметр Херста, характеризующий самоподобные свойства и одномерное распределение вероятностей.

Введение

Целью статьи является исследование одного из возможных подходов к моделированию самоподобного трафика в сетях передачи данных. Поскольку самоподобная природа сетевого трафика затрудняет его анализ и моделирование классическими методами, последнее десятилетие специалисты уделяли большое внимание разработке новых моделей, подходов и методов решения такого рода задач. На сегодняшний день получен ряд важных и интересных результатов в этой области (см., например, [1-3]), некоторые из которых послужили отправной точкой данной работы.

Одним из популярных способов моделирования трафика, обладающего заданными самоподобными свойствами, является использование модели "Input M/G/∞". Поскольку русская терминология в данной области ещё не сформировалась, будем использовать пословный перевод, и называть ее "моделью типа входная M/G/Γ". Популярность этой модели обусловлена рядом её достоинств. Во-первых, она относительно проста и ее легко использовать для генерации искусственных трасс трафика произвольной длины, во-вторых она дает простое физическое обоснование самоподобных свойств трафика (такое, например, как наличие долговременных зависимостей и т.д.).

В данной работе мы рассмотрим один из возможных подходов к моделированию самоподобного трафика с заданными статистическими характеристиками с использованием модели "входная M/G/Γ".

Основные определения

В дальнейшем изложении нам понадобится определение самоподобия и достаточное условие самоподобия в широком смысле. В данном параграфе мы приведём необходимые определения, а также подробное описание исходной модели "входная M/G/Γ".

Пусть $X = \{X_t\} = (\dots, X_1, X_2, \dots)$ стационарный в широком смысле случайный процесс дискретного времени, $t \in I_{-\infty} = \{\dots, -1, 0, 1, \dots\}$.

Пусть $\mu = EX_t < \infty$, $x_t = X_t - \mu$ (центрированный процесс), дисперсия и автокорреляционная функция процессов X и x есть:

$$\text{var } X_t = \text{var } x_t = \sigma^2 = EX_t^2 < \infty, \quad (1)$$

$$r_X(k) = \frac{E(x_t x_{t+k})}{\sigma_X^2}, \quad (2)$$

$\mu, \sigma_X, r_X(k)$ не зависят от t вследствие стационарности X_t и x_t .

Пусть

$$X_t^{(m)} = \frac{1}{m}(X_{tm-m+1} + \dots + X_{tm}), t \in I_1. \quad (3)$$

Процесс $X^{(m)} = (\dots, X_1^{(m)}, X_2^{(m)}, \dots)$ - стационарный в широком смысле случайный процесс с дискретным временем, обозначим через $r_X^{(m)}(k)$ автокорреляционную функцию процесса $X_t^{(m)}$.

Определение 1: Процесс X называется асимптотически самоподобным второго порядка с параметром Херста $H = 1 - \beta/2$,

$0 < \beta < 1$, если для всех $k \in I_1 = \{1, 2, \dots\}$

$$\lim_{m \rightarrow \infty} r_X^{(m)}(k) = \frac{1}{2}((k+1)^{2-\beta} - 2k^{2-\beta} + (k-1)^{2-\beta}) = g(k). \quad (4)$$

Свойство: (доказано в [1]) Если

$\lim_{m \rightarrow \infty} r_X^{(m)}(k) / k^{-\beta} = const$, тогда

- 1) $\lim_{m \rightarrow \infty} var X^{(m)} / m^{-\beta} = const$;
- 2) $\lim_{m \rightarrow \infty} r_X^{(m)}(k) = g(k), \quad k \in I_1.$ (5)

Именно асимптотическое самоподобие второго порядка мы исследуем в данной работе, и далее будем называть его просто самоподобием.

Опишем модель трафика “входная M/G/Γ”, обладающую самоподобными свойствами. Пусть имеется некоторый пуассоновский процесс с интенсивностью λ , и пусть A_t – количество пуассоновских событий в интервале $[t, t+1)$, $t \in I_{-\infty}$. Обозначим через $A = (\dots A_{-1}, A_0, A_1, \dots)$ последовательность независимых одинаково распределённых случайных величин A_t с пуассоновским законом распределения $Pr\{A_t = k\} = e^{-\lambda} \lambda^k / k!$. Предположим, что A_t – количество новых активных источников информационного трафика, появившихся в некоторой сети связи во временном окне t . Каждый такой источник имеет период активности в сети, после окончания которого он исчезает. Длина периода активности (“время жизни в сети”) источника i , возникшего во временном окне t ($i = 1, 2, \dots, A_t$) – это случайная величина $\tau_{t,i} \in I_1$. Кроме того, каждый источник имеет свою скорость генерации информации в течение активного периода (измеряемую в условных единицах, называемых ячейками), которая описывается как случайный процесс $S_{t,i}(n)$, где n – время от начала этого периода, $S_{t,i}(n) \in I_0, n \in I_0$. Пусть случайные пары $(\tau_{t,i}, S_{t,i}(n))$ независимы, одинаково распределены и не зависят A .

Определим сетевой трафик как случайный процесс $Y = (\dots, Y_{-1}, Y_0, Y_1, \dots)$,

$$Y_t = \sum_{k=-\infty}^t \sum_{i=1}^{A_t} S_{k,i}(t-k) I(\tau_{k,i} > t-k), \quad (6)$$

где $I(\cdot)$ индикатор события:

$$I(\tau_{k,i} > t-k) = \begin{cases} 1, & \text{если } \tau_{k,i} > t-k \\ 0, & \text{иначе} \end{cases},$$

то есть значение случайного процесса Y_t в момент времени t является суммарной интенсивностью генерации ячеек от всех активных источников. Если $EY_t < \infty$, то процесс Y_t стационарный и эргодический.

Одним из самых простых, и в тоже время интересных для практики случаев является случай постоянной скорости, одинаковой для всех источников, когда $S_{t,i}(n) = R = const$, а времена жизни источников являются независимыми одинаково распределёнными случайными величинами.

Теорема: (доказана в [1]). Если для указанного выше частного случая для распределения времени жизни источника выполняется условие

$$\lim_{m \rightarrow \infty} Pr\{\tau > n\} = \frac{const}{n^\alpha}, \quad 1 < \alpha < 2, \quad (7)$$

то процесс Y_t будет асимптотически самоподобным второго порядка процессом с параметром Херста

$$H = (3 - \alpha) / 2. \quad (8)$$

Даже такой простой случай представляет богатые возможности для моделирования трафика, обладающего нужными статистическими характеристиками. Однако, он имеет и определённые ограничения. В частности, одномерное распределение процесса Y_t всегда является Пуассоновским с параметром

$$\Lambda = \lambda E\tau. \quad (9)$$

Модель трафика с источниками случайной скорости

Основное уравнение

Рассмотрим ситуацию, когда скорости источников не являются одинаковыми. Пусть $S_{t,i}(n) = S_{t,i}$, где все $S_{t,i}$ – независимые и одинаково распределенные случайные переменные, принимающие значения в I_1 . Как ранее упоминалось, $\tau_{t,i}$ тоже независимые и одинаково распределенные случайные величины, и $S_{t,i}$ и $\tau_{t,i}$ не зависят и друг от друга. Тогда не трудно показать, что:

$$EY_t = \lambda ES \sum_{n=0}^{\infty} \Pr\{\tau > n\} = \lambda ES E\tau,$$

$$\sigma_Y^2 = \lambda ES^2 \sum_{n=0}^{\infty} \Pr\{\tau > n\} = \lambda ES^2 E\tau, \quad (10)$$

$$\sigma_Y^2 r_Y(k) = \lambda ES^2 \sum_{n=0}^{\infty} \Pr\{\tau > n+k\}.$$

Для этого случая нетрудно доказать теорему, аналогичную теореме из предыдущего параграфа. Используя эти результаты, можно получить процесс с заданной автокорреляционной функцией и одномерным распределением. Для этого необходимо решить следующую задачу – по заданному одномерному распределению трафика найти распределение вероятностей для скорости одного источника.

Для вывода основного уравнения воспользуемся аппаратом производящих функций. Пусть скорость отдельного источника имеет производящую функцию

$$\Phi_S(z) = \sum_{k=1}^{\infty} \Pr\{S = k\} \cdot z^k. \quad (11)$$

Обозначим количество активных источников в момент времени t за N_t . Как ранее отмечалось, согласно (13) одномерное распределение случайного процесса является пуассоновским с параметром Λ , поэтому производящую функцию процесса Y_t можно записать как:

$$\Phi_Y(z) = \sum_{k=0}^{\infty} \Pr\{N_t = k\} (\Phi_S(z))^k =$$

$$= \sum_{k=0}^{\infty} e^{-\Lambda} \frac{\Lambda^k (\Phi_S(z))^k}{k!} = \exp(\Lambda(\Phi_S(z) - 1)), \quad (12)$$

$$\text{откуда } \Phi_S(z) = 1 + \frac{1}{\Lambda} \ln(\Phi_Y(z)). \quad (13)$$

Это и есть основное уравнение, связывающее распределение вероятностей для общего трафика с распределением вероятностей скорости одного источника в рассматриваемой модели.

Нахождение распределения вероятности источника

Использовать полученное уравнение можно различными способами. Здесь мы представим некоторые результаты для случая, когда распределение вероятностей трафика задано в численной форме. Вид распределения неизвестен. Такой случай будет иметь место в ситуации, когда одномерное распределение процесса можно оценить по экспериментальным данным, например, построению гистограммы.

Для получения точных результатов нам необходимо сделать следующее ограничение. А именно, скорость источника не может быть нулевой, т.е. $\Pr\{S = 0\} = 0$. Это условие имеет вполне понятный физический смысл – источник не считается активным, если в процессе передачи пакеты от этого источника отсутствуют.

Пусть нам заданы параметр Херста H и набор вероятностей $\Pr\{Y_t = k\}$, $k \in I_0$. Введём обозначение:

$$p_k = \exp(-\Lambda) \frac{\Lambda^k}{k!}, \quad (14)$$

Нетрудно видеть, что, учитывая условие ненулевой скорости источника, можно записать $\Pr\{Y_t = 0\} = p_0$.

$$\text{Отсюда } \Lambda = -\log(\Pr\{Y_t = 0\}). \quad (15)$$

Кроме того, $\Pr\{Y_t = 1\} = p_1 \Pr\{S = 1\}$, тогда

$$\Pr\{S=1\} = \frac{\Pr\{Y_t = 1\}}{p_1}. \quad (16)$$

Продолжая подобные рассуждения далее, получаем

$$\Pr\{Y_t = 2\} = p_1 \Pr\{S = 2\} + p_2 \Pr\{S_1 + S_2 = 2\}$$

где S_1 и S_2 – независимые одинаково распределённые случайные переменные с минимальным значением 1. Значит,

$$\Pr\{S_1 + S_2 = 2\} = (\Pr\{S = 1\})^2 \text{ и}$$

$$\Pr\{S = 2\} = \frac{\Pr\{Y_t = 2\} - p_2 (\Pr\{S = 1\})^2}{p_1}. \quad (17)$$

Используя такой подход легко получить общее выражение для $\Pr\{S = k\}$:

$$\Pr\{S = k\} = \frac{\Pr\{Y_t = k\} - \sum_{m=2}^k p_m \Pr\{\sum_{n=1}^m S_n = k\}}{p_1}, \quad (18)$$

где $\Pr\{\sum_{n=1}^m S_n = k\}$ может быть найдена, как m -

кратная свёртка моментов конечной последовательности $\Pr\{S=1\}, \Pr\{S=2\}, \dots, \Pr\{S=k-m\}$, где все эти вероятности известны из предыдущих итераций. Такая итерационная процедура является одним из возможных подходов к численному решению основного уравнения при условии ненулевой скорости источника.

Имитационное моделирование

Применим полученные формулы для имитационного моделирования самоподобного трафика с заданным одномерным распределением. Для экспериментов были выбраны трассы известной коллекции трасс трафика Ethernet фирмы “Bellcore”. Приведем результаты экспериментов с трассами из файлов “BC-pOct89.TL” и “BC-pAug89.TL”. Файлы типа TL состоят из записей, каждая из которых включает два поля данных, где первое поле – время появления пакета (в секундах, с 6 цифрами после запятой), а второе – размер пакета в байтах. В каждом из использованных файлов содержится 1000000 таких записей.

Чтобы преобразовать полученные данные в случайный процесс дискретного времени,

мы провели дискретизацию с шагом 15 миллисекунд. Значение процесса равно сумме размеров пакетов, прибывающих в систему в течение соответствующего временного интервала, делённой на размер условной ячейки (использованный размер ячейки 100 байт).

Чтобы применить рассматриваемый подход, для каждого из полученных процессов дискретного времени находим эмпирические оценки математического ожидания и дисперсии, а также оцениваем параметр Херста методом частичных дисперсий. По параметру Херста из (8) можно определить необходимое α , и учитывая, что

$$\Pr\{\tau = n\} = \frac{const}{n^{\alpha+1}}, \quad n = 1, 2, \dots$$

обеспечить выполнение (7).

Затем из (15) находим LL, а из (9) II. Распределение скорости источника вычисляется итерациями (18). Соответствующие вычисления были проделаны для каждой из исследуемых трасс.

Необходимо заметить, что итерационная процедура (18), применённая к данным, полученным из гистограммы реального процесса, может на некотором шаге выдать отрицательное значение для очередной вероятности. Это означает, что для имеющихся данных основное уравнение не может быть решено точно. Для нахождения некоторого приближения, нами использовалась следующая эвристическая процедура: каждый раз, когда итерация (18) давала отрицательный результат, соответствующая вероятность принималась равной 0, и вычисления продолжались. Итерации останавливались, когда сумма всех найденных вероятностей становилась больше 1. Последняя вероятность корректировалась так, чтобы сумма равнялась 1.

Описанная выше процедура использовалась для получения распределения скорости источника, которое использовалось в программе имитационного моделирования для генерации искусственных трасс трафика.

На рис.1 и 2 изображены реальные одномерные распределения реального и сгенерированного с помощью программы имитационного моделирования трафиков для исследуемых трасс. В табл. 1 и 3 приведены

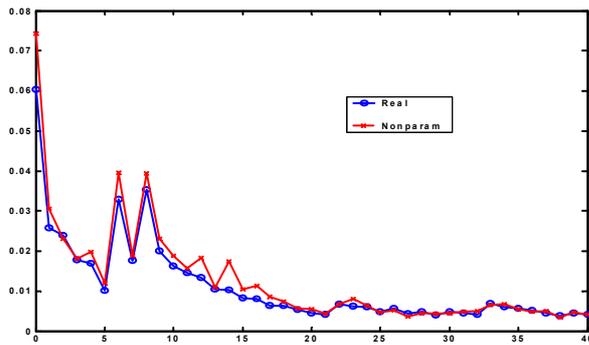


Рис. 1. Начальная часть одномерных распределений реального трафика файла “BC-pOct89.TL” и смоделированного трафика

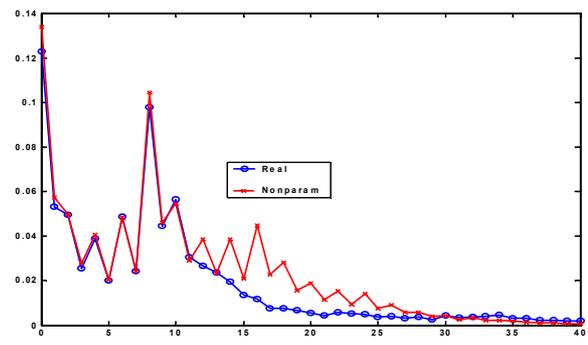


Рис. 2. Начальная часть одномерных распределений реального трафика файла “BC-pAug89.TL” и смоделированного трафика

Таблица 1. Значения статистических характеристик для трафика файла “BC-pOct89.TL” и смоделированного трафика

	EY_t	$\text{var } Y_t$	H
Real	57,10	2555,03	0,91
Simul	50,89	2796,9	0,74

Таблица 2. Начальная часть распределения скорости источника для файла “BC-pAug89.TL”, использованная при имитационном моделировании

k	1	2	3	4
$\text{Pr}\{S=k\}$	0,2062	0,1477	0,0292	0,1005
k	5	6	7	8
$\text{Pr}\{S=k\}$	0,0109	0,1339	0,0067	0,2987

Таблица 3: Значения статистических характеристик для трафика файла “BC-pAug89.TL” и смоделированного трафика

	EY_t	$\text{var } Y_t$	H
Real	22,47	1004,03	0,83
Simul	10,04	706,76	0,70

моменты процессов и значения их параметра Херста. В табл. 2 представлена начальная часть распределения скорости источника, вычисленная в соответствии с рассмотренной процедурой.

Из приведённых данных видно, что предложенный метод хорошо аппроксимирует одномерные распределения процесса (в их начальной части), но первые два момента смоделированного процесса могут весьма сильно отличаться от исходных.

Выводы

Представленная модель “входная M/G/∞” и метод имитационного моделирования яв-

ляются удобным средством для генерации самоподобного трафика в ситуации, когда важно одномерное распределение на начальном участке. Повышение точности аппроксимации первых моментов может являться предметом дальнейшего изучения.

Благодарности

Данная работа выполнена при частичной финансовой поддержке Гранта президента Российской Федерации для молодых докторов наук МД-1639.2005.9 и при частичной финансовой поддержке Гранта Научно-образовательного центра “Математические основы дифракционной оптики и

обработки изображений” Самарского государственного аэрокосмического университета для молодых учёных, студентов и аспирантов.

СПИСОК ЛИТЕРАТУРЫ

1. Self-Similar Network Traffic and Performance Evaluation. / Ed. by K. Park and W. Willinger.

New York: Wiley, 2000.

2. *V. Paxson and S. Floyd*. Wide Area Traffic: The Failure of Poisson Modeling. // IEEE/ACM Trans. Networking, 1995, 3 (3).

3. *N. Likhonov, B. Tsybakov and N. D. Georganas*. Analysis of an ATM buffer with self-similar (“fractal”) input traffic. // IEEE INFOCOM’95, Boston, pp. 985-992, April, 1995.

SELF-SIMILAR TRAFFIC MODELING AND SIMULATION

© 2006 A. Yu. Privalov, M. V. Baeva

Samara State Aerospace University

Method for self-similar telecommunication traffic simulation is presented. It allows to simulate traffic with given Hurst parameter and given marginal probability distribution.