

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТЕОРЕТИЧЕСКИХ ПОДХОДОВ МОДЕЛИРОВАНИЯ ТРАФИКА С ТОЧКИ ЗРЕНИЯ СООТВЕТСТВИЯ СЕТЯМ НОВОГО ПОКОЛЕНИЯ

О.А. Сиропятов, В.Я. Чечельницкий

Одесский национальный политехнический университет,
пр. Шевченко, 1, Одесса, 65044, Украина; e-mail: saa5511@ukr.net

В работе проведен анализ теоретических подходов к моделированию трафика. В результате проведенных исследований показаны принципы и основные этапы технологии моделирования высокоскоростного трафика коммерческих сетей. Показана необходимость применения комбинированного подхода для статистического описания трафика современных сетей. Вероятностное распределение анализируемых сигналов может иметь вид аддитивной смеси двух или более распределений с различными наборами параметров.

Ключевые слова: глобальные сети, сетевой трафик, моделирование трафика, матрица трафика

Введение

Современные теоретические методы и подходы описания процессов в информационных системах и сетях разнообразны и требуют научного осмысления для применения на практике. Все большее значение приобретают теоретико-вероятностные методы исследований, основанные на вероятностной трактовке протекающих в информационных системах процессов. Статистический подход позволяет более полно учесть состояние динамической системы, характер управляющих и возмущающих воздействий, результирующее поведение информационных потоков в многофункциональных сетях и во многих случаях более адекватен для решения различных практических задач [1].

Круг вытекающих из указанного подхода проблем достаточно широк: описание математических моделей случайных процессов в информационных системах, формирование на их базе статистических методов проверки гипотез и обнаружения, оценивания и фильтрации, интерполяции (сглаживания) и экстраполяции (прогнозирования), а также разработка алгоритмов оптимального управления стохастическими системами. Основными критериями при выборе теоретического подхода моделирования сети являются следующие:

- модель должна быть пригодной для описания современных высокоскоростных коммерческих сетей;
- модель должна быть универсальной, т.е. адаптируемой к различным типам сетей или их фрагментам;
- точность соответствия реальным потокам данных должна быть не более 10% при приемлемой вычислительной громоздкости;
- модель должна быть пригодной для прогнозирования поведения сети при более интенсивном трафике;

- модель должна отображать маршруты трафика и их изменение.

Для того чтобы определить, какой из теоретико-вероятностных методов наиболее близок к поведению высокоскоростных телекоммуникационных систем, и необходимо выполнить их сравнительный анализ. Задача усложняется тем, что большинство авторов осуществляли оценку точности модели при использовании данных локальных сетей, что не гарантирует пригодность подхода для больших территориально-разнесенных систем. Другими словами, необходима дополнительная проверка предлагаемых подходов, основанная на данных измерения трафика реальных высокоскоростных сетей.

Цель статьи и постановка задач исследований

Целью работы является проведение сравнительного анализа теоретических подходов моделирования сетевого трафика для разработки основных этапов моделирования высокоскоростного трафика коммерческих сетей.

Для достижения цели необходимо решить следующие *задачи*:

1. Провести анализ теоретических подходов к моделированию трафика.
2. Показать принципы и основные этапы технологии моделирования высокоскоростного трафика коммерческих сетей.
3. Обосновать необходимость применения комбинированного подхода для статистического описания трафика современных сетей.

Анализ теоретических подходов моделирования трафика

Поскольку основной математической моделью процессов, описывающих трафик в телекоммуникационных системах, является случайный поток данных, вполне оправданны попытки создания теоретической модели трафика на основании статистической теории. Случайный поток в рассматриваемом практическом приложении обладает следующими основными свойствами:

- независимость вероятностных характеристик от времени (стационарность);
- зависимости вероятностей событий (случайные процессы с памятью);
- бесконечно малая вероятность более одного события за бесконечно малый интервал времени.

Поток как случайный процесс характеризуется своими статистическими свойствами. Чаще всего используются: плотность вероятности поступления данных за период, функция вероятности потока и автокорреляционная функция.

Классической моделью трафика в информационных сетях является Пуассоновский поток. Он характеризуется набором вероятностей $P(k)$ поступления k сообщений за временной интервал t :

$$P(k) = \frac{(\lambda t)^k}{k!} \cdot e^{-\lambda t} \quad (1)$$

где: $k = 0, 1, 2, \dots$ — число сообщений; λ — интенсивность потока.

Заметим, что интервал времени измерения количества сообщений t и интенсивность потока λ являются постоянными величинами.

Для семейства Пуассоновских распределений (1) большее значение λ соответствует более широкому и симметричному графику плотности и большему объему информационных потоков. Зная вероятность поступления данных за период, можно получить распределение интервала τ между соседними событиями:

$P(\tau) = \lambda \cdot e^{-\lambda t}$. Основным свойством пуассоновского потока, обуславливающим его широкое применение при моделировании, является аддитивность: результирующий поток суммы пуассоновских потоков тоже является пуассоновским с суммарной интенсивностью

$$\lambda_{\Sigma} = \sum_{n=1}^N \lambda_n.$$

Однако, применение статистики (1) с учетом многофункциональности современных сетей возможно исключительно для описания очередей пакетов. Динамические процессы, происходящие в современных сетях, имеют сложную природу и относятся к стохастическим процессам. Такие свойства трафика возникают из-за недетерминированности системы в целом. Другими словами, долгосрочное прогнозирование действий, осуществляемых обрабатывающими трафик алгоритмами, невозможно предсказать, как впрочем, и массу других воздействующих на трафик факторов. На практике трафик обрабатывается алгоритмами, используемыми в различных реализациях протоколов семейства TCP/IP: генерация трафика протоколами транспортного уровня, управление трафиком на промежуточных сетевых устройствах, динамическая маршрутизация и т.д. В результате, процессы в компьютерных сетях находятся под постоянным влиянием регулирующих и возбуждающих стохастических воздействий, обуславливающих сложные флуктуации исследуемых процессов. Другими словами, требуется модель, являющаяся случайным процессом, управляемым другим случайным процессом.

В научных исследованиях последних лет отдается предпочтение описанию статистических свойств сети на основе применения Марковских скрытых цепей управляемых пуассоновской статистикой для описания очередей потоков (рис. 1).

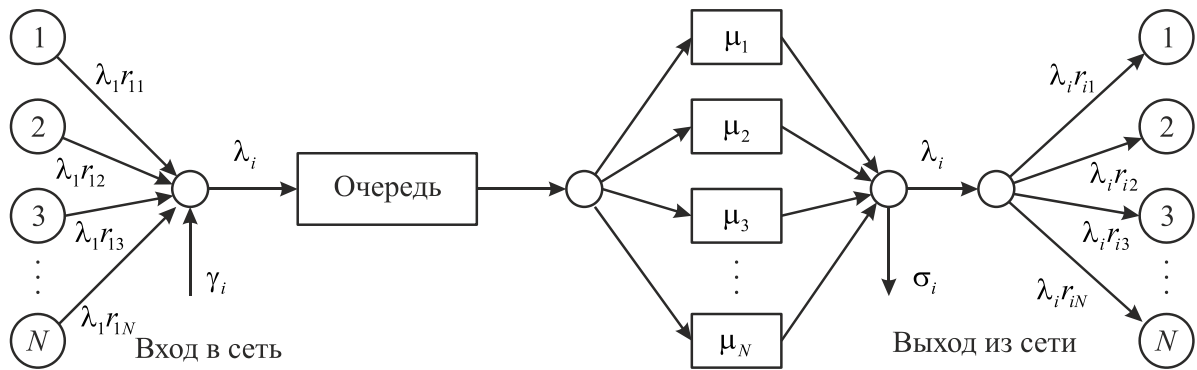


Рис. 1. Структура моделирования очередей трафика: r_{ij} — вероятность входа/выхода; λ_i — полная интенсивность потока; γ_i — параметр управления

Дискретный случайный Марковский поток характеризуется следующими свойствами

$$P(X_{n+1} = j | X_n = i_n, X_{n-1} = i_{n-1}, \dots) = P(X_{n+1} = j | X_n = i_n) \tag{2}$$

$$P(X_{m+1} = j | X_n = i) = p_{ij}$$

где: p_{ij} - вероятность перехода из состояния i в состояние j , $p_{ij} \geq 0$ для $\forall i, j$ и $\sum_j p_{ij} = 1$ для $\forall i$.

Процесс перехода некоторого дискретного случайного процесса из одного состояния в другое описывается матрицей переходов:

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \dots & p_{0j} & \dots & p_{0N} \\ p_{10} & p_{11} & p_{12} & \dots & p_{1j} & \dots & p_{1N} \\ p_{20} & p_{21} & p_{22} & \dots & p_{2j} & \dots & p_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ p_{i0} & p_{i1} & p_{i2} & \dots & p_{ij} & \dots & p_{iN} \\ \vdots & \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ p_{N0} & p_{N2} & p_{N3} & \dots & p_{Nj} & \dots & p_{NN} \end{pmatrix}$$

Наиболее адекватно отобразить стохастические переходы сети из одного состояния в другое позволяет модель скрытых цепей Маркова (СМЦ) (рис. 2).

Для определения модели на основе СМЦ необходимо задать число возможных состояний N , объем символов используемого алфавита, M . Для некоторого конечного алфавита можно определить конечный набор вероятностей перехода $\Lambda = (a_{ij})$:

$$a_{ij} = P(s_{t+1} = j | s_t = i), 1 \leq i, j \leq N, \tag{3}$$

где s_t — описывает текущее состояние.

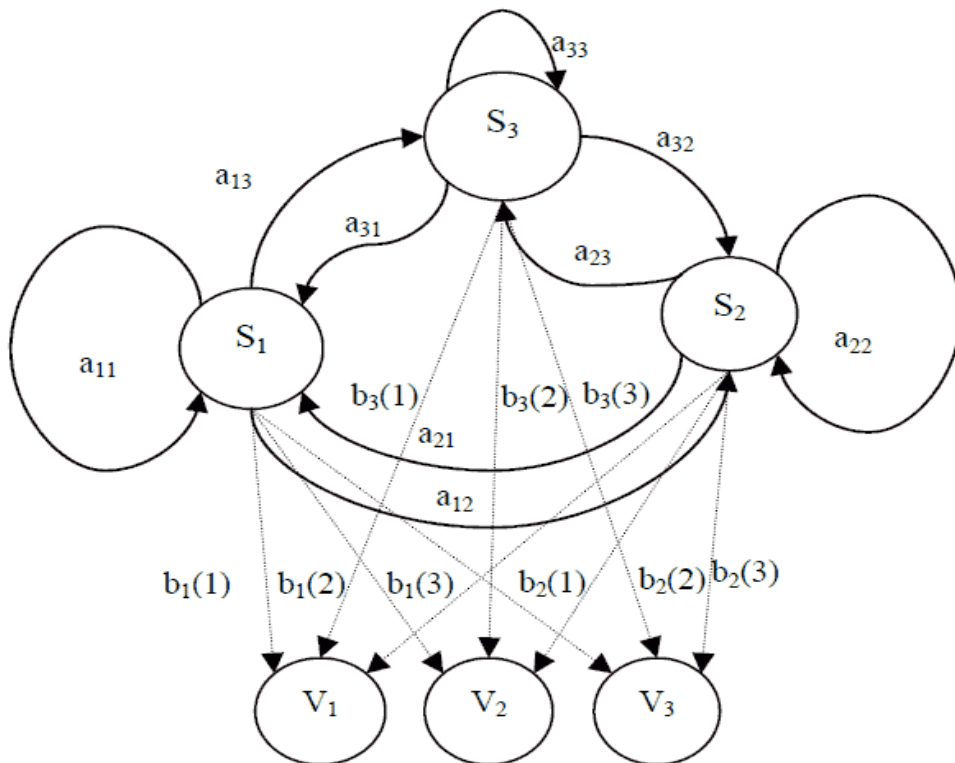


Рис. 2. Граф скрытой цепи Маркова

Для вероятностей перехода

$$\sum_{j=1}^N a_{ij} = 1, a_{ij} \geq 0, 1 \leq i, j \leq N.$$

Функция вероятности возможных состояний $B = \{b_i(k)\}$

$$b_j(k) = P(o_t = v_k | s_j = j), 1 \leq j \leq N, 1 \leq k \leq M,$$

$$b_j(k) = P(o_t = v_k | q_t = j), 1 \leq j \leq N, 1 \leq k \leq M. \tag{4}$$

где v_k — определяет k^{th} наблюдаемый символ алфавита, o_t - отображает текущее состояние.

Для описанной стохастической модели:

$$b_j(k) \geq 0; 1 \leq j \leq N, 1 \leq k \leq M,$$

$$\sum_{k=1}^M b_j(k) = 1, 1 \leq j \leq N. \tag{5}$$

Если на основе, например, пуассоновской статистики определяется случайный процесс инициализации, $\pi = (\pi_i)$, где $\pi_i = P(s_1 = i)$, $1 \leq i \leq N$, то для определения СМЦ необходимо задание $\lambda = (\Lambda, B, \pi)$.

Таким образом, для рассматриваемого практического приложения при заданном λ и наблюдениях $Y = y_1, y_2, \dots, y_T$, модель (1)–(5) позволяет оценить условную вероятность $P\{Y|\lambda\}$. Проблема моделирования будет заключаться в нахождении таких параметров модели $\{\Lambda, B, \pi\}$, которые позволят максимизировать вероятность $P\{Y|\lambda\}$.

Однако, несмотря на вычислительную сложность реализации данного подхода при моделировании даже локальных сетей проверка точности модели на данных реальных измерений высокоскоростных сетей дает большие расхождения, т.е. демонстрирует несостоятельность классических методов оценки вероятностно-временных характеристик сетей пакетной коммутации для больших территориально-разнесенных систем [2–4]. Усложнение модели посредством дополнения итерационных процедур для коррекции оценки параметров позволяет улучшить точность. Но открытыми остаются вопросы инициализации алгоритмов.

Основная проблема заключается в игнорировании стохастической динамики реальных сетей (рис. 3). Широкие перспективы применения адаптивных транспортных протоколов для мультимедиа приложений стимулируют разработку новой технологии моделирования трафика современных сетей.

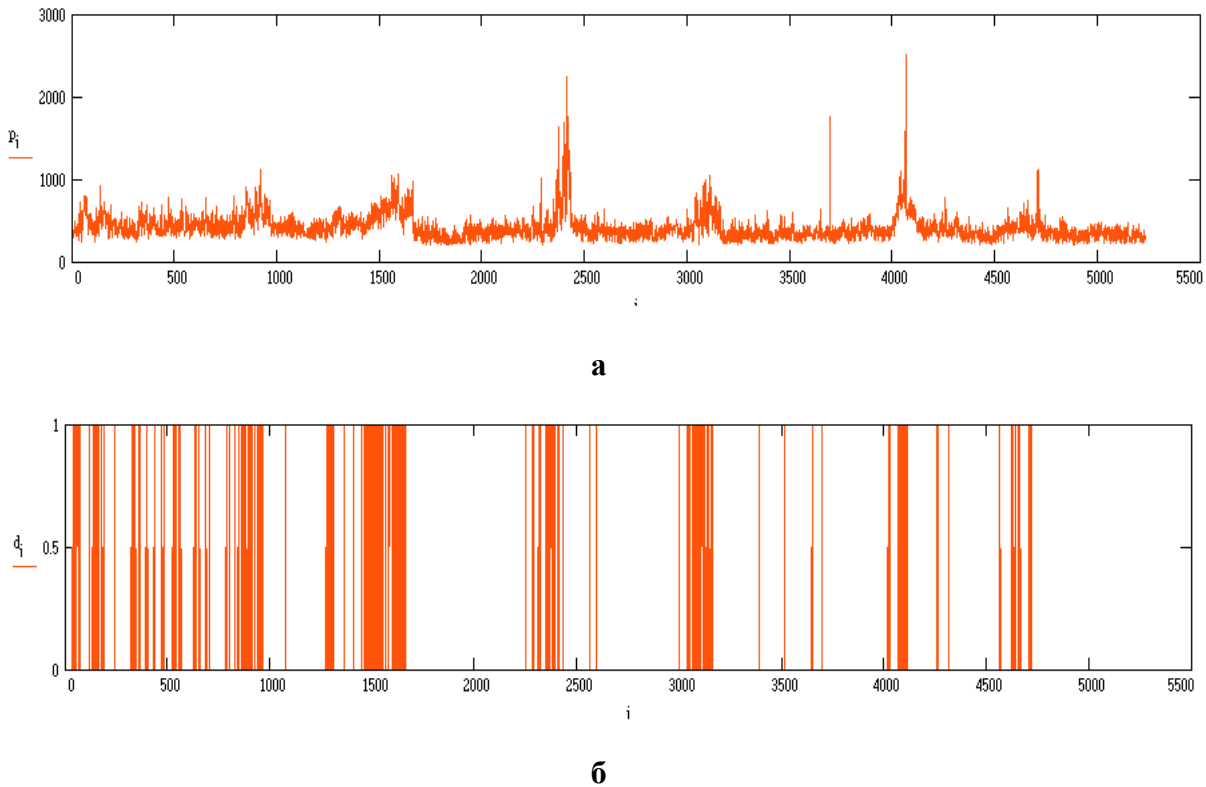


Рис 3. Характер изменения задержек при передаче данных: а – абсолютные значения интервалов времени между отправкой пакета и получением подтверждения, мс; б – структура потока, превышающего пороговое значение 600 мс

С другой стороны, если протоколы маршрутизации используют только локальную информацию о связанности узлов пакетной коммутации, то для них индикатором разрыва соединения является увеличение объема выходной очереди и ее возможное переполнение, если время перекоммутации виртуального соединения превышает определенное пороговое значение. В силу случайного характера рассматриваемых событий их вероятностное моделирование должно отражать возможность появления подобных резких отклонений. Таким образом, возникает задача формализации основных свойств современного трафика.

Статистические свойства современного трафика

Многие особенности процессов в сети Интернет в настоящее время еще слабо изучены. На текущем этапе развития сети в связи с расширением мультисервисных услуг, происходит объединение разнородных потоков информации, например таких, как передача речи, данных, сигналов аналогового и цифрового телевидения, сигналов радиовещания. Основой для организации таких систем является пакетная коммутация. Однако такое объединение приводит не только к удобному использованию информационных потоков со стороны потребителя телекоммуникационных услуг, но и к некоторым техническим сложностям, например: недостаточно проработан вопрос о приоритетных потоках, необходимо учитывать пачечность трафика и т.д. Очевидно, что в агрегированном трафике современных сетей отчетливо выделяется случайная составляющая резких отклонений, корреляционная связанность которых сохраняется на весьма продолжительном интервале наблюдения. При этом статистические характеристики трафика обладают свойством временной масштабной инвариантности,

которое в современной научной литературе принято называть самоподобием [4]. Данное свойство является следствием протяженных статистических зависимостей, проявляющихся в степенном характере затухания автокорреляционных функций временных отсчетов. В зависимости от прогнозируемых изменений сетевых ресурсов можно регулировать сетевую производительность. Уменьшение объема буферной памяти относительно интенсивности потоков требует значительно более точного прогнозирования состояния сетевых ресурсов. В противном случае возникновение локальных перегрузок приводит к фактическому разрыву виртуальных соединений.

На транспортном уровне существенное влияние на характер сетевого трафика оказывает распределение времени подтверждения прихода пакетов. На рис. 4 показана зависимость логарифма экспериментально полученной функции распределения времен подтверждения передачи для трафика состоящего из 5000 пакетов.

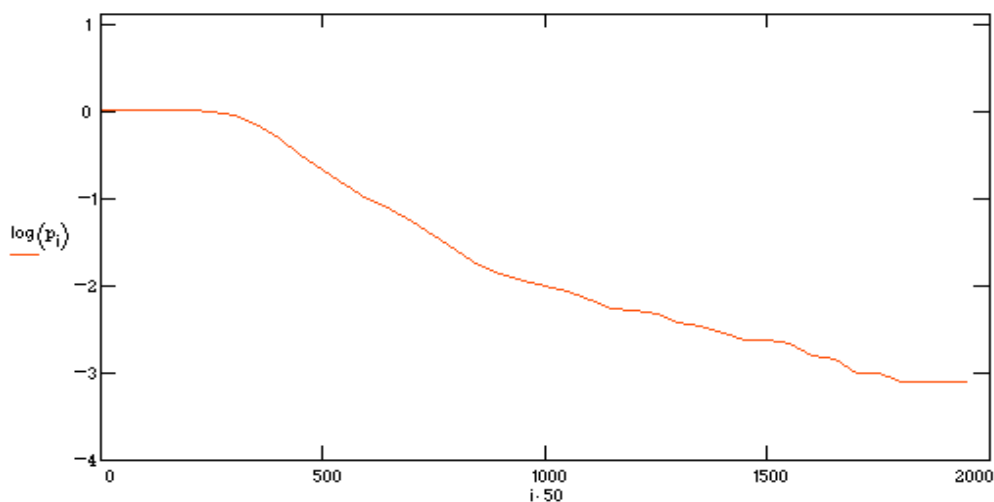


Рис. 4. Экспериментально полученное распределение времени прихода пакетов подтверждения

Фрактальные свойства нагрузки сети обусловили появление ряда моделей трафика на основе сильнокоррелированных самоподобных, или фрактальных стохастических процессов [4]. Такие функции распределения вероятности изучаются в теории статистики экстремальных значений, для применения которой необходимо выполнение следующих условий:

1. Наблюдаемый процесс обладает свойством цикличности, то есть определенной повторяемостью своего поведения через некоторый фиксированный интервал времени (интервал цикличности). Этот интервал задается, как правило, естественным путем, определяясь самой природой процесса, а не вводится искусственно.

2. Наибольший интерес с точки зрения исследователя (пользователя) представляют именно экстремальные значения, имеющие место в каждом цикле, поскольку они могут привести к серьезным последствиям, скачкообразно возрастающим потерям, к утрате ее работоспособности и неизбежным негативным последствиям различного рода.

Однако в рамках ограниченности измерений трудно говорить о самоподобии трафика современных коммерческих сетей в целом. Скорее данное свойство отображается для фрагмента сети и на заданном временном интервале. В общем случае интервал естественной цикличности может быть равен одним суткам, неделе, году.

Необходимо проанализировать и обосновать выбор аналитической зависимости функций распределения для описания наблюдаемых статистических закономерностей с учетом статистических свойств трафика сетей на основе собранных данных.

Функция распределения самоподобной случайной величины X характеризуется наличием тяжелого хвоста, если $P[X > x] \sim x^{-\alpha-1}$, $x \rightarrow \infty$, $0 \leq \alpha < 2$, соответственно, плотность распределения вероятности $p(x) \sim \alpha x^{-\alpha-1}$, $\alpha > 0$.

Не существует точных рекомендаций по выбору той или иной плотности вероятности со свойством самоподобия или в соответствие с другим определением — с тяжелым хвостом.

Наиболее распространенной является плотность распределения вероятности Парето

$$p(x) = \alpha k^\alpha x^{-\alpha-1}, \alpha < 2, k > 0, x \geq k, \quad (6)$$

и функция вероятности

$$F(x) - P[X \leq x] = 1 - (k/x)^\alpha. \quad (7)$$

При $\alpha \leq 2$ случайный процесс (6), (7) характеризуется бесконечной дисперсией и при $\alpha \leq 1$ - бесконечным средним.

Плотность распределения вероятности с тяжелым хвостом Вейбула является двухпараметрическим распределением

$$p(x) = \alpha \lambda^{-\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\lambda}\right)^\alpha}, \alpha, \lambda > 0, x \geq 0. \quad (8)$$

Для сравнения статистик (6) и (8) было выполнено моделирование случайных процессов (рис. 5, рис. 6).

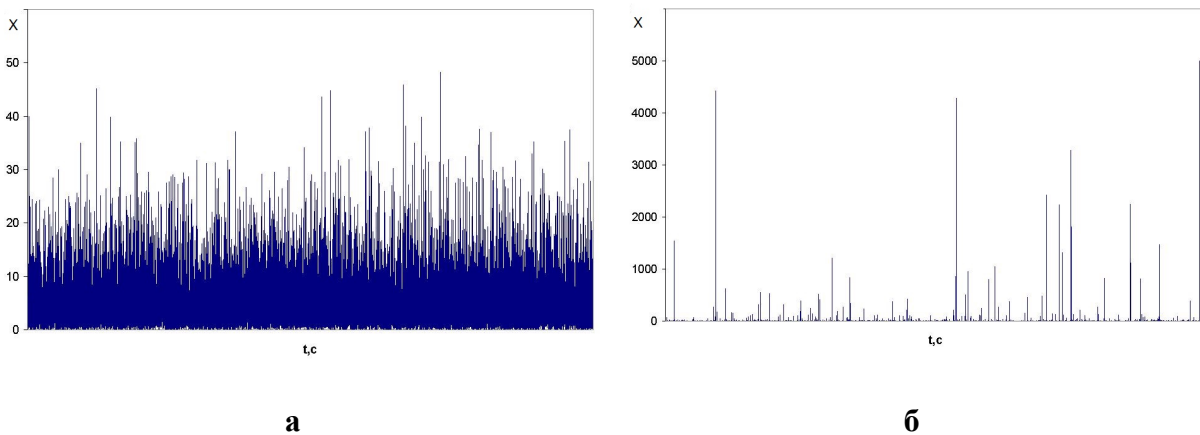


Рис. 5. Результаты имитационного моделирования потока с распределением Вейбулла: а - $\alpha = 2$, б - $\alpha = 1$

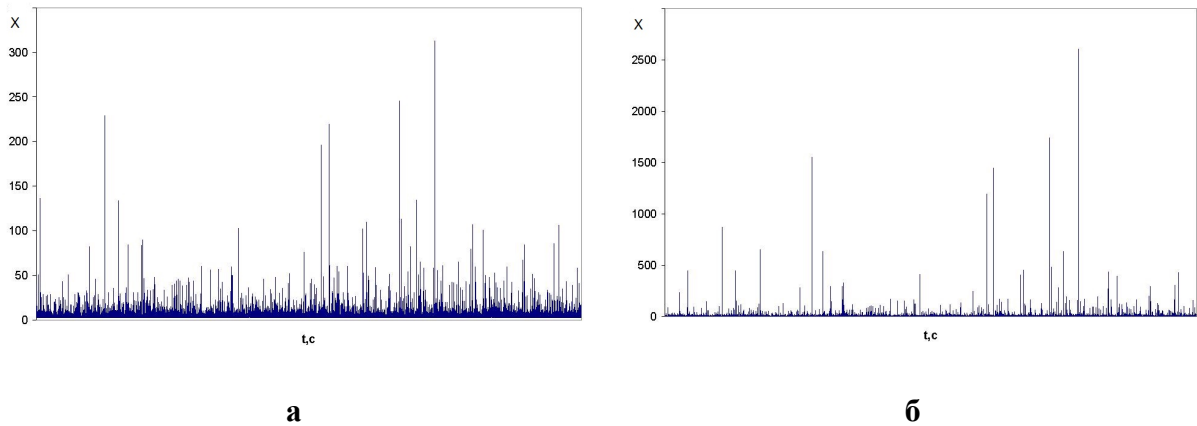


Рис. 6. Результаты имитационного моделирования потока с распределением Парето: а - $\alpha = 1$; б - $\alpha = 0.1$

Наибольшую взрывоподобность проявляет поток с распределением Вейбулла. Для распределения Вейбулла оказалось достаточно сложно при небольшом среднем значении ($\mu = 5$) получить достоверные результаты для малого значения α . Это связано с тем, что значение коэффициента имеет величину такого порядка малости, что происходит влияние на процесс точности моделирования. Определенные вычислительные трудности характерны и для распределения Парето из-за большого значения дисперсии. Другими словами, для получения достоверных оценок требуется рассматривать большие объемы статистической информации. В результате эксперимента было установлено свойство масштабной инвариантности статистических характеристик. В отличие от процессов, не обладающих фрактальными свойствами, не происходит быстрого «сглаживания» процесса при усреднении по шкале времени — процесс сохраняет склонность к всплескам.

Важной задачей является определение, какая из моделей самоподобных случайных процессов приемлема для рассматриваемого приложения. Самоподобные процессы, в том числе описывающие явления в сетях передачи данных, обладают рядом свойств, существенно отличающих их от потоков, рассматриваемых в классической теории телетрафика.

Традиционно самоподобие в стохастическом процессе выявляется путем определения параметра Херста

$$H_{k,n} = \frac{1}{k} \sum_{i=n-k+1}^n (\log X_i - \log X_{i+1}). \quad (9)$$

Тот факт, что $0.5 < H < 1$ считается достаточным основанием для признания процесса самоподобным (по крайней мере, асимптотически).

Имитационное моделирование и проверка условий (9) продемонстрировало неприемлемость использования логнормального распределения для рассматриваемого приложения. Данные реальных измерений трафика ОП пары фрагмента сети, соответствующей wi-fi технологии, наблюдались в течение месяца, затем был выполнен анализ статистических свойств полученного массива, был построен график хвоста эмпирической функции распределения $1 - F(x)$ в логарифмической шкале (рис. 7).

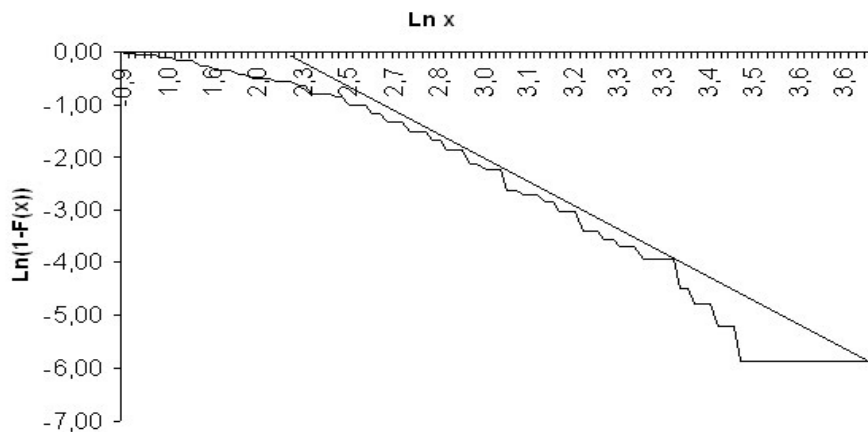


Рис. 7. График хвоста эмпирической функции распределения трафика локальной сети

Эксперимент продемонстрировал зависимость самоподобности от дня недели. Коэффициент самоподобности зависит от интенсивности трафика, а также от типа передаваемой информации. Аналогично параметр самоподобности меняется в течение суток, и это связано с особенностями поведения пользователей. Интенсивность трафика очень мала в интервале времени 02:00–08:00. В данном интервале времени параметр Хорста минимален. Рабочий день начинается в 08:00 и интенсивность трафика невелика, но, несмотря на это, параметр Хорста уже увеличивается. В интервале с 20:00 до 02:00, когда администраторы сети начинают тестировать написанный программный код в автоматическом режиме, интенсивность потока заявок к серверу вырастает очень значительно, и параметр Хорста также возрастает до значения 0.84 во все рабочие дни. Сетевой трафик в случае максимальной интенсивности (20:00–02:00) имеет более высокую степень самоподобности.

Однако данный подход применим для фрагмента сети и при ограничении временного интервала. Даже в этом случае для достижения приемлемой точности необходимо использовать комбинацию нескольких статистических составляющих, например фрактальный дробовой процесс.

Выводы

В результате проведенных исследований становятся ясны принципы и основные этапы технологии моделирования высокоскоростного трафика коммерческих сетей.

Важным выводом из проведенного анализа является необходимость применения комбинированного подхода для статистического описания трафика современных сетей.

Вероятностное распределение анализируемых сигналов, например, времени подтверждения, может иметь вид аддитивной смеси двух или более распределений с различными наборами параметров

$$P(x) = \sum_{j=1}^m p(x|j)p(j), \tag{10}$$

где: $p(x|j)$ — плотность вероятности j -ой компоненты; $p(j)$ - весовой коэффициент учета j -ой компоненты в общей суперпозиции.

Открытым остается вопрос о том, какое семейство плотности вероятности нужно применять в качестве составляющих (10) для достижения наилучшей точности модели.

Очевидно, что ответ на этот вопрос зависит от того, можно ли получить данные только на основании изучения статистических свойств постоянно изменяющегося реального трафика коммерческих сетей.

Список литературы

1. Cao, J. Time-Varying Network Tomography: Router Link Data / J. Cao, D. Davis, S. Vander Wiel, B. Yu // Journal of the American Statistical Association. - vol. 95, no. 452, Dec. 2000.
2. Waldbusser, S. Remote Network Monitoring Management Information Base / S. Waldbusser // IETF RFC 2819, May 2002.
3. Van der Merwe. mmdump: A tool for monitoring Internet multimedia traffic / Van der Merwe, R. Caceres, Y. Chu, C. J. Sreenan // ACM Computer Communication Review, vol. 30, October 2000. - pp. 48-59.
4. Lakhina, A. Characterization of Network-Wide Anomalies in Traffic Flows (2004) / A. Lakhina, M. Crovella, C. Diot. May 19, 2004.

ПОРІВНЯЛЬНИЙ АНАЛІЗ ТЕОРЕТИЧНИХ ПІДХОДІВ МОДЕЛЮВАННЯ ТРАФІКА З ПОГЛЯДУ ВІДПОВІДНОСТІ МЕРЕЖАМ НОВОГО ПОКОЛІННЯ

О.А. Сиропятов, В.Я. Чечельницький

Одеський національний політехнічний університет,
просп. Шевченка, 1, Одеса, 65044, Україна; e-mail: saa5511@ukr.net

У роботі проведено аналіз теоретичних підходів до моделювання трафіку. В результаті проведених досліджень показані принципи та основні етапи технології моделювання високошвидкісного трафіку комерційних мереж. Показано необхідність застосування комбінованого підходу для статистичного опису трафіку сучасних мереж. Імовірнісний розподіл аналізованих сигналів може мати вигляд адитивної суміші двох або більше розподілів з різними наборами параметрів.

Ключові слова: глобальні мережі, мережений трафік, моделювання трафіку, матриця трафіку

COMPARATIVE ANALYSIS OF THEORETICAL APPROACHES OF TRAFFIC SIMULATION IN TERMS OF CORRESPONDENCE TO NEW GENERATION NETWORKS

O.A. Siropyatov, V.J. Chechelnitsky

Odessa National Polytechnic University,
1 Shevchenko Ave., Odessa, 65044, Ukraine; e-mail: saa5511@ukr.net

An analysis of theoretical approaches to traffic simulation was performed in this work. Principles and main stages of high-speed traffic of commercial networks simulation were demonstrated as a result of investigations carried out. The need in applying combined approach for statistical description of modern networks traffic was demonstrated. Probabilistic distribution of analyzed signals can have the form of an additive mixture having two or more distributions with different sets of parameters.

Keywords: global networks, network traffic, traffic simulation, traffic matrix