

МЕТОДЫ ПРОГНОЗИРОВАНИЯ ТРАФИКА В КОМПЬЮТЕРНЫХ СЕТЯХ

Батыр С.С., Ступак Г.В.

Донецкий национальный технический университет, г. Донецк,

Кафедра «Автоматики и телекоммуникаций»

e-mail: ibatyr@gmail.com, stupakgv@gmail.com

Abstract

Batyr S.S. Stupak G.V. Methods of traffic prediction in computer networks. The authors analyzes two methods of traffic prediction in computer networks. The Puasson process and the self-similar process were described. The traffic prediction models were tested and compared. The self-similar model is more adequate then the Puasson-based model.

Общая постановка проблемы. На сегодняшний день существует два основных подхода к описанию и прогнозированию трафика в компьютерных сетях, каждый из которых имеет свои плюсы и минусы. Первый способ базируется на использовании вероятностных характеристик сетевого трафика, второй – фрактальности процессов в сетях, при этом используется в корне отличный математический аппарат описания процессов.

Постановка задач исследования. Для разработки оптимального средства прогнозирования и моделирования сетей, сетевой нагрузки необходимо более детально рассмотреть каждый из подходов, и определить наиболее оптимальный.

В случае использования аппарата теории массового обслуживания, телетрафик определяется как движение информационных потоков в информационных системах [1]. Поток как случайный процесс характеризуется своими статистическими свойствами. Чаще всего используются: плотность вероятности поступления данных за период, плотность вероятности интервалов между поступлениями данных и автокорреляционная функция. При этом, случайный поток обладает следующими свойствами:

стационарность – независимость вероятностных характеристик от времени;

степень последствия – уровень зависимости вероятности поступления событий от предыдущих;

ординарность – бесконечно малая вероятность поступления более одного события за бесконечно малый интервал времени.

В расчетах используют классическую модель трафика, а именно – Пуассоновский поток. Он характеризуется набором вероятностей $P(k)$ поступления k сообщений за временной интервал t :

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

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

Заметим, что интервал времени измерения количества сообщений t и интенсивность потока λ являются постоянными величинами. Математическое ожидание и дисперсия для Пуассоновского потока равны λ . Зная вероятность поступления данных за период, можно получить распределение интервала τ между соседними событиями:

$$P(\tau) = \lambda \cdot e^{-\lambda \tau} \quad (2)$$

Отсюда вывод: пуассоновский поток характеризуется экспоненциальным распределением интервалов между событиями.

Основным свойством пуассоновского потока, обуславливающим его широкое применение при моделировании, является аддитивность: результирующий поток суммы пуассоновских потоков тоже является пуассоновским с суммарной интенсивностью:

$$\lambda_{\Sigma} = \sum \lambda \quad (3)$$

При моделировании, Пуассоновский поток можно получить мультиплексированием совокупности ON/OFF источников, которые называются Марковскими процессами (рисунок 1).

Считается, что случайная величина имеет распределение с тяжелым хвостом (РТХ или Heavy Tailed), если:

$$P(x) \approx x^{-\alpha} \text{ при } x \rightarrow \infty \quad (4)$$

То есть хвост распределения затухает по степенному закону.

Считается, что сетевой трафик во многих случаях лучше всего описывается именно тяжелохвостым распределением, например распределением Парето [1].

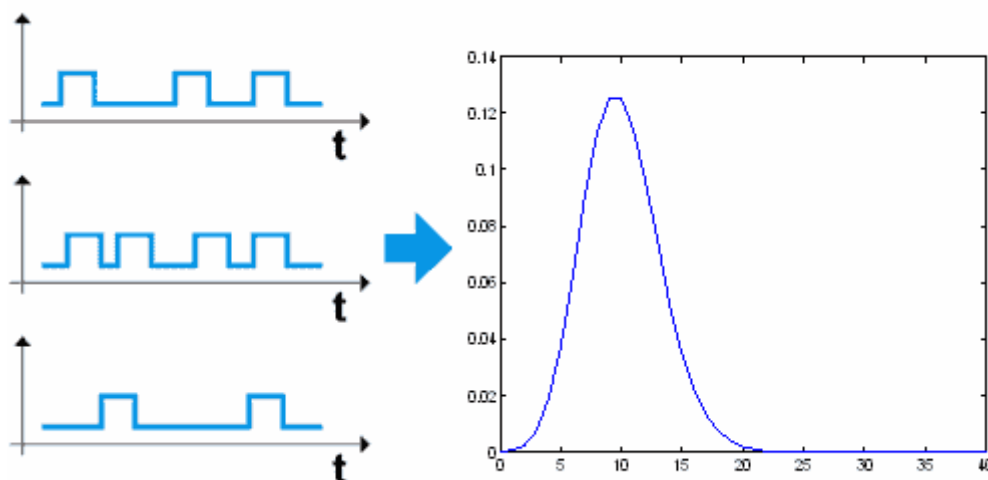


Рисунок 1 – Получение Пуассоновского потока

При моделировании сетевого трафика необходимо использование автокорреляционной функции (АКФ). В общем случае АКФ характеризует внутреннюю зависимость между временным рядом и тем же рядом, но сдвинутым на некоторый промежуток времени, который называется лагом. Вычисления АКФ проводятся по классической формуле:

$$r(k) = \frac{\sum_i (X_i - \bar{X}) \cdot (X_{i+k} - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad (5)$$

где \bar{X} с чертой - выборочное среднее, $k=0,1,2, \dots$ - лаг.

Также при использовании классической модели считается, что процесс обладает медленно убывающей зависимостью (МУЗ), если он характеризуется АКФ, которая убывает гиперболически (по степенному закону) при увеличении лага [1]. В противоположность МУЗ существует понятие быстро убывающей зависимости (БУЗ).

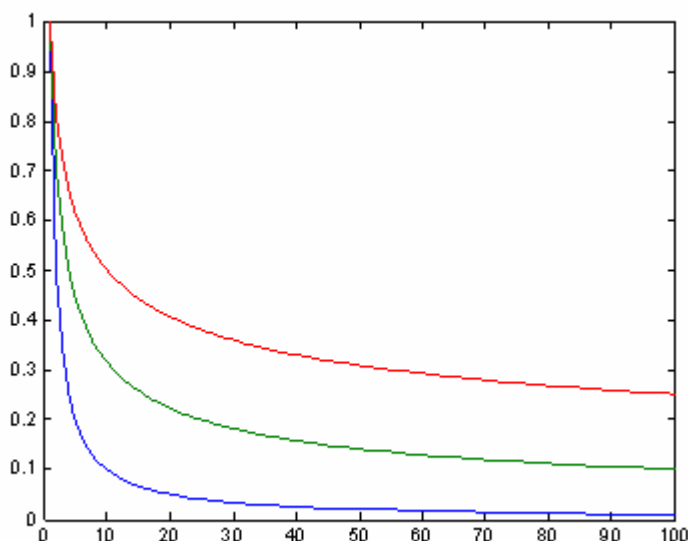


Рисунок 2 – АКФ БУЗ (первая снизу) и МУЗ (вторая и третья)

С понятием МУЗ связано важнейшее прогнозирующее свойство – продолжительная память, характеризующаяся зависимостью текущих параметров процесса от предыдущих.

Исходя из вышеприведенной методики рассчитывают параметры сетевого трафика и на их основании производят прогнозирование, но в последнее время, методика основанная на теории массового обслуживания была частично опровергнута. В первую очередь это связано с возросшими величинами потоков в сети и скорости в каналах связи.

Решение задач и результаты исследований. Установлено [2], что трафик данных высокоскоростных сетей, обладает самоподобными, или фрактальными, свойствами, поэтому расчеты при использовании моделей марковских процессов и формул Эрланга дают явно завышенные результаты, что приводит к нерациональной загрузке каналов например. «Самоподобие» представляет собой свойство процесса сохранять свое поведение и внешние признаки при рассмотрении в разном масштабе. Кроме того, самоподобный трафик имеет особую структуру, сохраняющуюся при многократном масштабировании - в реализации, как правило, присутствует некоторое количество выбросов при относительно небольшом среднем уровне трафика. Данное явление ухудшает характеристики (увеличивает потери, задержки, джиттер пакетов) при прохождении самоподобного трафика через узлы сети. На практике это проявляется в том, что пакеты, при высокой скорости их движения по сети, поступают на узел не по отдельности, а целой пачкой, что может приводить к их потерям из-за ограниченности буфера, рассчитанного по классическим методикам.

В реальных условиях случайные процессы сохраняют свойство самоподобия только до определенного предела. Этот предел или мера статистической устойчивости процесса при многократном масштабировании определяется так называемым параметром Херста или параметром самоподобия [3]. Случайный процесс $x(t)$ является статистически самоподобным с параметром Херста H ($0,5 \leq H \leq 1$), если для любого вещественного значения $a > 0$ процесс $x(at)/a^H$ обладает теми же статистическими характеристиками, что и сам процесс $x(t)$:

математическое ожидание –

$$M[x(t)] = \frac{M[x(at)]}{a^H} \quad (6)$$

дисперсия –

$$D[x(t)] = \frac{D[x(at)]}{a^{2H}} \quad (7)$$

корреляционная функция –

$$R(t, \tau) = \frac{R(at, a\tau)}{a^{2H}} \quad (8)$$

Чем больше H , тем дольше сохраняется свойство самоподобия при многократном масштабировании. При $H = 0,5$ это свойство практически отсутствует.

Для временных последовательностей масштабируемой величиной является время. Исходя из определения самоподобия, можно утверждать, что временные и спектральные характеристики трафика при изменении масштаба усреднения будут описываться одними и теми же уравнениями, функциями, но с соответствующими масштабными коэффициентами.

Как и при использовании аппарата теории массового обслуживания в качестве источника сообщений выступает модель ON/OFF источника, который может находиться в одном из двух состояний: ON, в котором он активно передает кадры (при этом время между появлением кадров имеет экспоненциальное распределение), и OFF, в течение которого он бездействует, но для описания временных периодов ON и OFF используется распределение с бесконечной дисперсией – распределение Парето. Параметры распределения Парето определяют степень самоподобия трафика. Они выбираются с таким расчетом, что бы степень самоподобия искусственного трафика была равна степени самоподобия реального трафика [4]. Оценка степени самоподобия реального трафика происходит на основании проведенного анализа по формулам 6-8, и нахождении параметра Херста.

В работе [4] исследовался сегмент сети (Ethernet), содержащий пять рабочих станций, подключенных к концентратору. Концентратор, в свою очередь, был подключен к коммутатору второго уровня. В течение 2 часов была зафиксирована передача 20001 кадра. Затем были построены зависимости (рисунок 3) среднего времени передачи и числа отказов при передаче от интенсивности появления кадров в сети.

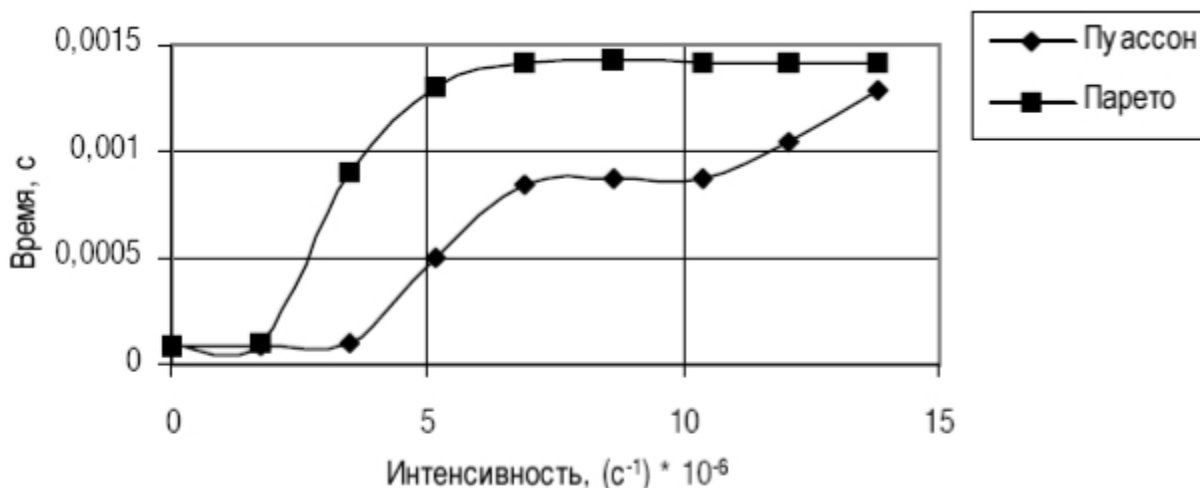


Рисунок 3 – Зависимости среднего времени передачи от интенсивности

Анализ результатов показал, что использование аппарата теории массового обслуживания дает заведомо ложные результаты относительно времени передачи пакетов и интенсивности. Данное следствие может привести к переполнению буферной памяти коммутационных устройств сети и потере пакетов, росте вероятности отказов в сети в целом. Следовательно, для проведения дальнейших исследований, модель сетевого трафика построенная с учетом самоподобия трафика, будет наиболее оптимальной.

Выводы.

1. Использование аппарата теории массового обслуживания при моделировании потоков данных компьютерных сетей дает неадекватные результаты. Характеристики трафика генерируемого современными сетевыми приложениями не соответствуют характеристикам потоков генерируемых с помощью простейшего Пуассоновского распределения.

2. При проектировании сетевого коммуникационного оборудования и разработке телекоммуникационных программных пакетов учитывать самоподобную структуру передаваемых потоков данных.

3. В результаты имитационного моделирования работы сетей средствами, которые используют для генерации потоков данных распределения Пуассона, необходимо вносить коррекцию в сторону увеличения при интенсивностях потока более $2 \cdot 10^{-6}$ (c^{-1}).

4. При разработке новых средств имитационного моделирования компьютерных сетей использовать методы генерации трафика с учетом эффекта самоподобия. При этом должна существовать возможность задания уставочного значения и текущей оценки параметра Херста для изменения степени самоподобия генерируемого трафика.

Литература

1. Крылов В.В., Самохвалова С.С. Теория телеграфика и ее приложения.- СПб.: БХВ-Петербург, 2005.

2. M. Garrett and W. Willinger. Analysis, Modeling, and Generation of Self-Similar VBR Video Traffic / Proceedings, SIGCOMM'94, August 1994.

3. Столлингс В. Современные компьютерные сети. 2-е издание. – С.Пб.: Питер, 2003. – 783 с.

4. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СЕТИ С УЧЕТОМ САМОПОДОБИЯ Лукьянов В.С., Мокров А.Ю., / Информационные технологии в науке, образовании, телекоммуникации и бизнесе / IT+S&E`05, майская сессия, с141-142