Информационные процессы, Том 11, № 3, 2011, стр. 369–377
© *2011 Цитович, Титов.*

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ТЕХНИЧЕСКИХ И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМАХ

О временно́м масштабе в математической модели источника нагрузки с бесконечной дисперсией времени обслуживания

И.И. Цитович[*], **И.Н. Титов**[**]

[*] *Институт проблем передачи информации им. А.А. Харкевича РАН, Москва, Россия*
[**] *Московский технический университет связи и информатики, Министерство российской федерации по связи и информатизации, Москва, Россия*

Поступила в редколлегию 25.09.2011

Аннотация—В настоящей статье рассмотрена модель трафика сервера данных, в которой суммарный трафик разделён на несколько потоков. Требованиям из этих потоков соответствует свой объём передаваемых данных, различающихся на порядки для различных потоков. Рассматриваемая модель предполагает, что потоки описываются одинаковыми процессами, но каждому соответствует свой масштаб времени. Показано, что для анализа трафика сервера, а также для разработки наиболее эффективных методов управления этим трафиком, необходимо правильно определить масштаб времени для каждого из потоков, а также масштаб времени, в котором колебания суммарного трафика оказывают существенное влияние на QoS.

1. ВВЕДЕНИЕ

В последнее время было получено и проанализировано большое число наблюдений за мультимедийным трафиком в высокоскоростных сетях, таким как потоки пакетов в локальных сетях, потоки ячеек при передаче видео с переменной скоростью в ATM сетях и т.д. Эти измерения показали, что трафик в этих сетях является самоподобным по своей структуре [2]. Для самоподобного трафика свойственно, что корреляция не обращается в ноль в широких масштабах времени. Эти фрактальные свойства приводят к пульсирующему характеру трафика. Но математический анализ моделей, основанных на самоподобных процессах, является очень сложным. С другой стороны, традиционные модели, такие как пуассоновский поступающий поток, марковский поступающий поток и т.д., являются хорошо изученными, но они не могут обеспечить достаточное соответствие реальному трафику в современных сетях, включая долговременные зависимости. Обобщение на марковские потоки с групповым поступлением заявок на обслуживание [1], [3] не даёт существенных улучшений; но при этом такие модели оказываются более сложным для определения их параметров и исследования свойств моделей.

В настоящей статье предполагается, что математическая модель трафика основана на классических пуассоновских потоках, но каждому из потоков соответствует свой временной масштаб. Данное предположение основывается на том, что трафик, порождаемый сервером данных, можно разделить на группы в зависимости от объёма запрашиваемых данных. Следовательно, общий трафик может рассматриваться, как линейная комбинация однородного трафика, но с различной шкалой времени для каждой компоненты. Такой подход даёт нам возможность анализировать долговременные зависимости, используя компоненты, которым соответствует большой временной масштаб. В отличие от самоподобного трафика этот подход даёт возможность использовать классические модели телетрафика для исследования процессов с долговременными зависимостями.

В разделе 2 представлен анализ реального трафика Web сервера, который генерирует четыре потока, для которых объём запрошенных данных отличается от 10 до 30 раз. В разделе 3

приведена математическая модель трафика, генерируемого источником с «бесконечной» дисперсией времени обслуживания. В разделе 4 исследуется зависимость дисперсии создаваемой нагрузки от времени агрегирования, определяющего масштаб времени, в котором рассматриваются колебания трафика. В разделе 5 приведены результаты имитационного моделирования рассмотренной модели и некоторых методов управления скоростью передачи данных.

2. АНАЛИЗ СВОЙСТВ ТРАФИКА WEB СЕРВЕРА

Рассмотрим трафик, поступающий от музыкального ресурса. Для этого ресурса характерно предоставление пользователям доступа к файлам различного типа. Большинство запросов поступает на передачу файлов небольшого размера (HTML страницы с изображениями) при поиске и просмотре дополнительной информации. Требований пользователей на передачу конкретного mp3-файла значительно меньше, однако для обработки таких запросов необходимо намного больше времени. Максимальная длительность обслуживания характерна для требований на передачу архивных файлов, содержащих музыкальные альбомы. Таким образом, можно сказать, что на вход сервера поступает несколько типов потоков требований пользователей, отличающихся интенсивностью и объемом запрошенных данных.

Необходимо отметить, что размеры файлов одного типа могут существенно различаться. Например, mp3-файлы имеют различное качество (bitrate) и длительность, а размер архивных файлов отличается значительно в зависимости от количества музыкальных композиций в альбоме, что не позволяет, основываясь на размере передаваемого объекта, провести четкую границу между файлами различного типа.

Обычно для загрузки каждого файла открывается отдельная TCP сессия, однако, протокол HTTP/1.1 позволяет передавать несколько объектов в рамках одной сессии (persistent connection) и, напротив, передавать за одну сессию только фрагмент файла (partial content). Это также приводит к тому, что граница между объемом передаваемых данных для файлов различного типа становится менее явной.

При проведении измерений трафик фиксировался на границе сети крупного оператора связи, предоставляющего пользователям доступ в Интернет. Проводился анализ заголовков сетевого и транспортного уровня. Критерием отбора был пул IP адресов, принадлежащих рассматриваемому ресурсу (primary и non-primary серверы), а также TCP порт источника 80 (HTTP). Для исследования распределения объема переданных данных была произведена оценка количества информации, поступающей от сервера к клиенту в рамках отдельной TCP сессии. Трафик анализировался непрерывно в течение 76 дней (с сентября по ноябрь 2010). За период наблюдений было зафиксировано 161000 TCP сессий, было передано более 45 Гбайт данных.

На рис. 1 изображён фрагмент гистограммы объёма переданных данных. По оси абсцисс отложен объем передаваемых данных в байтах, по оси ординат – частота попадания в соответствующий 100 байтный интервал.

Из представленного рисунка хорошо видно, что распределение вероятностей убывает немонотонно. Для различных значений объёма переданных данных наблюдаются локальные максимумы, соответствующие передаче большого числа близких по размеру объектов.

Для получения более детального представления о распределении количества переданной информации построим эмпирическую функцию распределения. Для исследований свойств хвоста распределения на рис. 2 представлена зависимость $1 - F_n(x)$ от x в log-log масштабе.

Из данной зависимости хорошо видно, что для трафика сервера данных характерны периоды с медленным убыванием $1-F_n(x)$ (т.е. за данный период было зафиксировано сравнительно небольшое количество сессий соответствующего размера), чередующиеся с периодами с высокой скоростью убывания $1 - F_n(x)$ (т.е. на данном периоде наблюдался локальный максимум

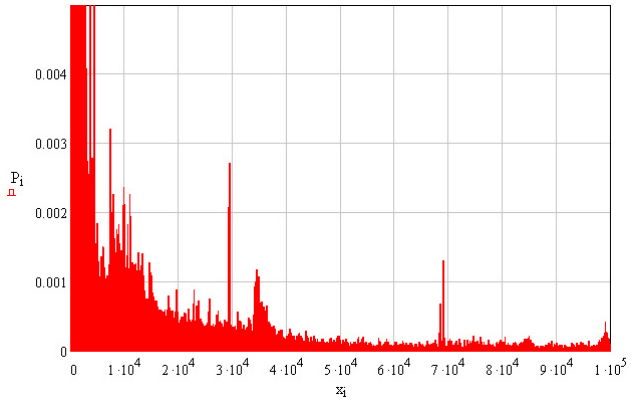

Рис. 1. Фрагмент гистограммы объёма переданных данных

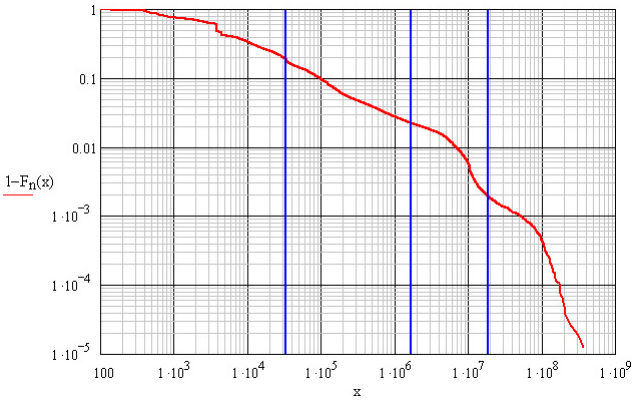

Рис. 2. Зависимость $1 - F_n\,(x)$ от объёма переданных данных

на гистограмме). Таким образом, мы можем отделить различные потоки требований в зависимости от скорости изменения эмпирической функции распределения, а также основываясь на характере изменения гистограммы. На рис. 2 изображены три вертикальные линии, соответствующие значениям объёма переданных данных: $34 \cdot 10^3$, $1{,}7 \cdot 10^6$ и $19 \cdot 10^6$ байт, отделяющие различные потоки запросов пользователей.

В табл. представлены основные характеристики потоков данных, генерируемых сервером по запросу пользователей: количество сессий — $\lambda \cdot T$, зафиксированное за время наблюдения T, средний объем переданных данных в рамках отдельной TCP сессии — $V_{ср}$, а также суммарный объем переданной информации для каждого потока — $\lambda \cdot T \cdot V_{ср}$.

Таблица

Характеристики потоков данных

№	1	2	3	4
$\lambda \cdot T$	131314	26152	3224	307
$V_{ср}$, Кбайт	5,6	189,2	6426	69549
$\lambda \cdot T \cdot V_{ср}$, Гбайт	0,701	4,72	19,76	20,36

Первому потоку соответствует загрузка HTML страниц, содержащих изображения в форматах jpeg и gif различного размера, а также скриптов языка Flash и JavaScript. Наибольшее количество TCP сессий (82%) открываются для загрузки именно этих типов файлов, однако объём трафика, создаваемого этим потоком, составляет всего 1,5% от общего трафика. Мак-

симальную нагрузку (43% и 45%) создают третий и четвёртый потоки требований, то есть загрузка отдельных mp3-файлов и заархивированных альбомов, соответственно.

Таким образом, анализ реального трафика Web сервера показал, что, основываясь на скорости изменения эмпирической функции распределения объёма переданных данных, трафик, генерируемый сервером, может быть разделён на 4 потока со средней длительностью обслуживания запросов отличающейся более чем на порядок. Следовательно, каждому из потоков будет соответствовать свой масштаб времени, отличающийся существенно для различных потоков.

3. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Для простоты исследований будем считать, что источник нагрузки порождает n пуассоновских потоков с интенсивностями λ_i, $\Lambda = \sum_{i=1}^{n} \lambda_i$ — суммарная интенсивность поступающих требований, $p_i = \lambda_i/\Lambda$ — вероятность поступления требования из i-го потока. Требование i-го потока обслуживается в течение a_i единиц времени. Тогда среднее время обслуживание требования $A = \sum_{i=1}^{n} p_i \cdot a_i$ единиц времени. В качестве характеристики случайности источника нагрузки рассматривается среднеквадратическое отклонение длительности обслуживания $\sigma = \sqrt{D}$, $D = \sum_{i=1}^{n} p_i \cdot a_i^2 - A^2$.

Мы рассматриваем ситуацию, когда $\sigma/A \gg 1$, что с инженерной точки зрения означает, что источник нагрузки не обладает конечной дисперсией и такой источник является объектом наших исследований.

Для простоты положим, что длительность обслуживания требования i-го потока $a_i = k^{i-1}$, где k — некоторое достаточно большое число. Введение параметра k позволяет нам регулировать соотношение между длительностями обслуживания требований, принадлежащих различным потокам.

Для данной модели, трафик, порождаемый сервером данных, можно разделить на n стационарных независимых потоков, которые имеют одинаковую структуру и отличаются только масштабом времени. Пусть X_t — стационарный случайный процесс, который описывает процесс обслуживания требований в системе. Он соответствует разности нагрузки, создаваемой потоком в момент времени t, и средней нагрузкой. Пусть $K(t,s)$ — корреляционная функция процесса X_t, тогда известно, что для рассматриваемой системы $K(t,s) = \sigma_0^2 \cdot (1 - |t-s|/a)$ при $|t-s| \le a$ и $K(t,s) = 0$ при $|t-s| \ge a$, где a — длительность обслуживания требований.

Отклонение от среднего значения суммарной нагрузки, порождаемой n потоками, может быть найдено как

$$Y_t = \alpha_1 \cdot X_t^{(1)} + \alpha_2 \cdot X_{t/k}^{(2)} + ... + \alpha_n \cdot X_{t/k^{n-1}}^{(n)}, \tag{1}$$

где $X_t^{(i)}$ — независимые копии процесса X_t, а α_i задают долю соответствующей компоненты в общей нагрузке системы.

Поскольку потоки независимы и стационарны, максимальное отклонение создаваемой нагрузки от среднего можно определить как

$$\max_{t\in[0;\infty)}(Y_t) = \sum_{i=1}^{n}\left(\alpha_i \cdot \max_{t\in[0;\infty)}\left(X_{t/k^{i-1}}^{(i)}\right)\right) \tag{2}$$

В этом случае отклонение исходного процесса можно получить как линейную комбинацию независимых одинаково распределенных случайных величин. Но с практической точки зрения это выражение даёт слишком завышенную оценку из-за медленной сходимости по k, т.к. в реальных системах этот параметр всегда ограничен. Также необходимо учесть, что в

современных сетях для компенсации колебаний трафика и, соответственно, уменьшения потерь на коммутационном устройстве, обслуживающем трафик сервера данных, используется буфер, куда попадают пакеты, когда создаваемая нагрузка превышает ёмкость выходного канала. Вероятность переполнения буфера является основной характеристикой QoS, поскольку файлы обладают высокой чувствительностью к потере отдельных, даже небольших, фрагментов данных. Поэтому максимальное отклонение нас интересует за период времени $[0, T]$, где T — время, сопоставимое со временем заполнения буфера коммутационного устройства. Это время не настолько большое, чтобы применение (2) давало достаточную точность. Поэтому необходимо использовать более тонкие методы оценки отклонения.

4. АНАЛИЗ ХАРАКТЕРИСТИК ТРАФИКА ПРИ РАЗЛИЧНЫХ МАСШТАБАХ ВРЕМЕНИ

Рассмотрим отношение отклонения от среднего значения объёма данных, поступивших от сервера за время T, к величине промежутка времени T:

$$\overline{Y} = \tfrac{1}{T} \cdot \int_0^T Y_t dt = \tfrac{1}{T} \cdot \int_0^T \left(\alpha_1 \cdot X_t^{(1)} + \alpha_2 \cdot X_{t/k}^{(2)} + ... \ ... + \alpha_n \cdot X_{t/k^{n-1}}^{(n)} \right) dt = \tfrac{\alpha_1}{T} \cdot \int_0^T X_t^{(1)} dt +$$

$$+ \tfrac{\alpha_2 \cdot k}{T} \cdot \int_0^{\mathrm{T/k}} X_t^{(2)} dt + ... + \tfrac{\alpha_n \cdot k^{n-1}}{T} \cdot \cdot \int_0^{\mathrm{T/k^{n-1}}} X_t^{(n)} dt. \qquad (3)$$

Поскольку потоки независимы и $E\overline{Y} = 0$, то дисперсию суммарного процесса вычислим как сумму дисперсий:

$$D\left[\overline{Y}\right] = E\left[\left(\tfrac{\alpha_1}{T} \cdot \int_0^T (X_t)\, dt\right)^2\right] + E\left[\left(\tfrac{\alpha_2 \cdot k}{T} \cdot \int_0^{T/k} (X_t)\, dt\right)^2\right] + ...$$

$$+ E\left[\left(\tfrac{\alpha_n \cdot k^{n-1}}{T} \cdot \int_0^{T/k^{n-1}} (X_t)\, dt\right)^2\right] =$$

$$= \tfrac{\alpha_1^2}{T^2} \cdot \int_0^T \int_0^T K(t,s)\, dtds + \tfrac{\alpha_n^2 \cdot k^2}{T^2} \cdot \int_0^{T/k} \int_0^{T/k} K(t,s)\, dtds + ...$$

$$... + \tfrac{\alpha_n^2 \cdot k^{2(n-1)}}{T^2} \cdot \int_0^{T/k^{n-1}} \int_0^{T/k^{n-1}} K(t,s)\, dtds. \qquad (4)$$

При $T \le a$ дисперсия для процесса $X_t^{(1)}$равна

$$D_1 = \frac{2 \cdot \alpha_1^2}{T^2} \cdot \int_0^{\mathrm{T}/\sqrt{2}} \sqrt{2} \cdot \left(T - \sqrt{2} \cdot t\right) \cdot \sigma_0^2 \cdot \left(1 - \sqrt{2} \cdot t \Big/ a\right) \mathrm{dt} \Bigg|_{T \le a} =$$

$$= \alpha_1^2 \cdot \sigma_0^2 \cdot (1 - T/(3 \cdot a))\big|_{T \le a}. \qquad (5)$$

При $T \ge a$ дисперсия для процесса $X_t^{(1)}$равна

$$D_1 = \frac{2 \cdot \alpha_1^2}{T^2} \cdot \int_0^{\mathrm{a}/\sqrt{2}} \sqrt{2} \cdot \left(T - \sqrt{2} \cdot t\right) \cdot \sigma_0^2 \cdot \left(1 - \sqrt{2} \cdot t \Big/ a\right) dt \Bigg|_{T \ge a} =$$

$$= \left(\alpha_1^2 \cdot \sigma_0^2 \cdot a/T\right) \cdot (1 - a/(3 \cdot T))\big|_{T \ge a}. \qquad (6)$$

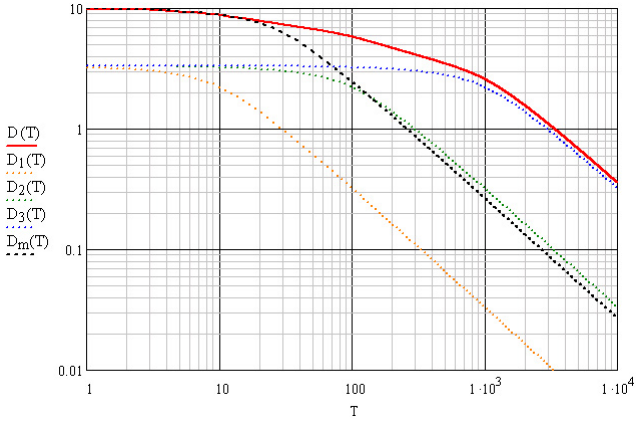

Рис. 3. График изменения дисперсии для систем с тремя и одним потоком

Подставив в (4) формулы (5) и (6), получим дисперсию суммарного процесса

$$D\left[Y\right] = \sum_{i=0}^{n-1}\left(\alpha_{i+1}^2 \cdot \sigma_0^2 \cdot \left(1 - \frac{T}{3 \cdot a \cdot k^i}\right) I(T \le a \cdot k^{n-1}) + \right.$$
$$\left. + \frac{\alpha_n^2 \cdot \sigma_0^2 \cdot a \cdot k^{n-1}}{T} \cdot \left(1 - \frac{a \cdot k^{n-1}}{3 \cdot T}\right) I(T \ge a \cdot k^{n-1})\right), \qquad (7)$$

где $I(\cdot)$ — индикаторная функция.

Из (7) следует, что при $T \gg a$ дисперсия отклонения объёма данных, генерируемого одним потоком, убывает пропорционально росту T, однако при $T \ll a$ дисперсия уменьшается со значительно меньшей скоростью. Это связано с тем, что при $T \ll a$ изменение объёма поступивших данных от сервера происходит чаще, чем меняется состояние системы (т.е. поступают новые запросы или оканчивается обслуживание требований поступившие ранее). Поскольку, АКФ для двух последовательных значений будет близка к 1, дальнейшее уменьшение T не приводит к росту дисперсии.

На рис. 3 изображена зависимость дисперсии $D\left[Y\right]$ от T, соответствующей сумме дисперсий каждого потока, при $n = 3$, $k = 10$, $a = 10$, $\sigma_0^2 = 10/3$ и $\alpha_1 = \alpha_2 = \alpha_3 = 1$. Для сравнения на рис. 3 также представлена зависимость дисперсии от T для системы, на вход которой поступает один пуассоновский поток с интенсивностью Λ, и постоянной длительностью обслуживания A — $D_m(T)$, то есть системы, для которой не учитывается структура источника нагрузки, а используются только усреднённые характеристики.

Из рис. 3, видно, что при $T < a$ дисперсия примерно одинакова для обеих систем, поэтому, например, для систем с явными потерями (без буфера) вероятность потерь будет одинаковой. Однако убывание дисперсии с высокой скоростью для второй системы начинается при значительно меньших значениях T. Например, при $T = a \cdot k^2$ дисперсия отличается в 10 раз. Следовательно, для систем, в которых среднее время заполнения буфера превышает A, необходимо учитывать структуру источника нагрузки, иначе предположения о вероятности переполнения буфера окажутся слишком оптимистичными.

Также на рис. 3 изображена зависимость дисперсии от T для каждого из потоков в отдельности — $D_i(T)$. Из этой зависимости и формулы (7) следует, что при малых значениях T дисперсия для всех потоков одинакова и равна σ_0^2, однако линейный спад дисперсии для различных потоков начинается при значениях T, отличающихся в k раз.

Поэтому, например, для 1-го потока буфера размером B может быть достаточно, чтобы обеспечить низкий уровень потерь. Пакеты будут часто попадать в буфер, но будут находить-

ся там недолго, т.к. система быстро меняет своё состояние, следовательно, объем задействованного ресурса буфера будет редко достигать максимального значения.

В тоже время для n-го потока такой размер буфера окажется недостаточным, поскольку n-й поток находится в другом масштабе времени. Нагрузка, создаваемая этим потоком, будет значительно реже достигать максимальных значений, но находиться в таких состояниях система будет дольше, что будет приводить к переполнению буфера. Обслуживание трафика этого потока будет приводить к длительным периодам с высокой средней нагрузкой, а, следовательно, высокому уровню потерь.

Следовательно, в зависимости от соотношения между масштабом времени, определяющим скорость заполнения буфера, и масштабами времени, соответствующими различным потокам, вклад, вносимый каждым из потоков в вероятность потерь, будет различным. Например, при малых значениях T все потоки будут вносить одинаковый вклад в вероятность переполнения буфера. При $T > a \cdot k^{n-1}$, напротив, вероятность потерь будет определять только n-м потоком. При $a < T < a \cdot k^{n-1}$ вклад, вносимый i-м потоком будет зависеть от T, и, следовательно, от T будет зависеть эффективность методов управления, применяемых к различным потокам.

5. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Для численных исследований была использована математическая модель, рассмотренная в разделе 3, при $n = 3$, $k = 10$ и $\lambda_1 = 33{,}3$. Для того чтобы каждый из потоков создавал одинаковую нагрузку на систему (поскольку при этом эффекты, вызванные сосуществованием потоков различной природы, проявляются более явно), интенсивность i-го потока должна определяться, как $\lambda_{\text{i}} = \lambda_1 \cdot k^{-(i-1)}$. Тогда $\Lambda = \lambda_1 \cdot (k^{-n} - 1) \big/ \left(k^{-1} - 1\right)$, $A = \lambda_1 \cdot n/\Lambda$, а общая нагрузка $A \cdot \Lambda = n \cdot \lambda_1$. Дисперсия такого источника $D = \sum_{i=1}^{n} \frac{\lambda_1 \cdot k^{i-1}}{\Lambda} - \frac{\lambda_1^2 \cdot n^2}{\Lambda^2} \approx k^{n-1} - n^2$ при больших k или n. Поэтому $\sigma/A \approx k^{0.5 \cdot (n-1)}/n \to \infty$ при больших k или n. Следовательно, предлагаемая модель обладает свойством, позволяющим исследовать эффекты, соответствующие неограниченной дисперсии времени обслуживания.

Все требования, поступающие в систему, получают одинаковую скорость Cm, следовательно, длительность обслуживания определяется только объёмом запрошенного файла.

Также были рассмотрены четыре системы с управлением: когда скорость передачи данных уменьшалась до Cm/d для всех требований (2-я система); когда скорость уменьшалась только для запросов на передачу файлов максимального размера (3-я система); когда одинаковое управление применялось к требованиям 2-го и 3-го потоков (4-я система); и когда скорость уменьшалась только для запросов на передачу файлов минимального размера (5-я система). Для 3-й, 4-й и 5-й систем время обслуживания требований для потоков, к которым применялось управление, выбиралось таким образом, чтобы среднее время обслуживания было одинаковым для всех четырёх систем и, соответственно, равным $A \cdot d$.

Рассмотрим зависимость убывания дисперсии от T для суммарного трафика 1-й системы, а также для каждого из потоков в отдельности, представленную на рис. 4. Кроме этого на рис.4 представлена аналогичная зависимость для 3-й системы обслуживания при $d = 1.5$.

Из рис. 3 и 4 видно, что оценка изменения дисперсии для различных потоков, полученная на основании анализа результатов моделирования трафика сервера, соответствует аналогичной характеристике, полученной в аналитическом виде в разделе 4. Из рис. 4 также можно сделать вывод, что применение управления только к требованиям 3-го потока позволяет увеличить скорость спада дисперсии уже при $T > a \cdot k$.

На рис. 5 представлена зависимость изменения дисперсии для 5-ти систем при $d = 1.5$.

Из данной зависимости видно, что для различных методов управления минимальные значения дисперсии соответствуют различным диапазонам T. Следовательно, в зависимости от

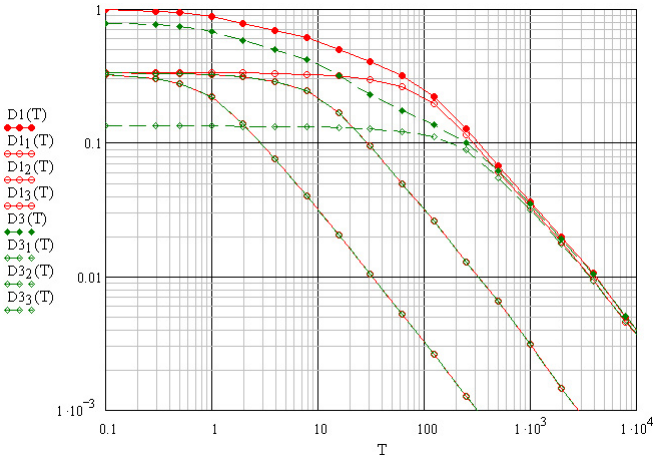

Рис. 4. Графики изменения дисперсии для 1-й и 3-й систем

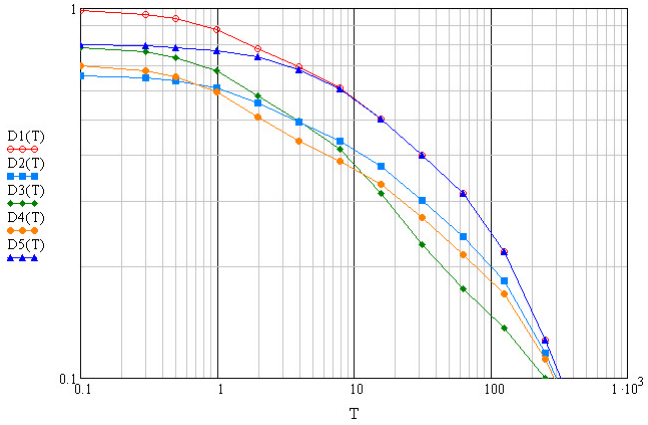

Рис. 5. Графики изменения дисперсии для 5-ти систем при $d = 1,5$

размера буфера и ёмкости канала эффективность рассмотренных методов управления будет различной. Например, на рис. 6 представлена зависимость вероятности потерь от размера буфера при ёмкости канала $C = 120 \cdot Cm.$ и $d = 1{,}5$. Из данного графика видно, что с ростом размера буфера растёт эффективность уменьшения скорости передачи только для запросов максимального объёма ресурсов (3-я система) относительно одинакового уменьшения скорости для всех потоков (2-я система). Метод управления, реализованный в 4-й системе, оказывается наиболее эффективным и позволяет снизить вероятность переполнения буфера в среднем в 15 раз.

6. ЗАКЛЮЧЕНИЕ

Трафик сервера данных, предоставляющего пользователям доступ к файлам различного типа, может быть разделен на потоки в зависимости от объёма запрашиваемых данных. Анализ реального трафика Web сервера показывает, что объемы запрашиваемого ресурса для этих потоков отличаются существенно. Следовательно, необходимо правильно определить масштаб времени, соответствующий каждому из потоков, поскольку от этого зависит, какие компоненты трафика и каким образом должны быть учтены.

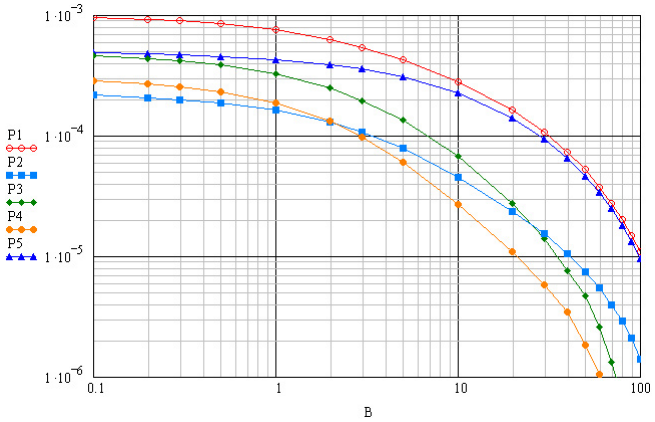

Рис. 6. Зависимость вероятности потерь от размера буфера

Также необходимо определить масштаб времени, соответствующий времени заполнения буфера коммутационного устройства, обслуживающего трафик сервера данных. Он зависит от размера буфера и ёмкости выходного канала.

Анализ предложенной модели показал, что вероятность потерь и, соответственно, эффективность различных методов управления будут определяться на основании соотношения между масштабом времени, на котором мы фиксируем колебания трафика, и реальными масштабами времени, соответствующими различным потокам.

СПИСОК ЛИТЕРАТУРЫ

1. Anderson A.T., Nielsen B.F. A Markovian approach for modeling packet traffic with long-range dependence // IEEE Journal on Selected Areas in Communications. 1998. V. 16, № 5. P. 719–732.

2. Dudin A. N., Klimenok V. I., Tsarenkov G. V. A Single-Server Queueing System with Batch Markov Arrivals, Semi-Markov Service, and Finite Buffer: Its Characteristics // Automation and Remote Control. V. 63, № 8. P. 1285–1297.

3. Leland W., Taqqu M., Willinger W., Wilson D. On the self-similar nature of ethernet traffic // IEEE/ACM Transactions on Networking. 1994. V. 2, № 1. P. 1–15.

4. Титов И.Н. Исследование характеристик потоков данных, генерируемых Web-сервером // T-Comm: телекоммуникации и транспорт. 2010. № 5. С. 30–34.

5. Цитович И.И., Титов И.Н. Исследование вероятности переполнения буфера при обслуживании трафика сервера, предоставляющего данные различного объёма // Сборник трудов 33-й конференции молодых учёных и специалистов ИППИ РАН: Информационные технологии и системы ИТиС'10. М.: ИППИ. 2010. С. 104–107.