

Модели нагрузки в веб-ориентированных компьютерных сетях

Анопrienко А.Я., Аль Абабнех Хасан
Кафедра ЭВМ ДонНТУ
anoprien@cs.dgtu.donetsk.ua

Аннотация

Анопrienко А.Я., Аль Абабнех Хасан, Модели нагрузки в веб-ориентированных компьютерных сетях. Рассмотрены особенности изменения параметров нагрузки в веб-ориентированных компьютерных сетях за последнее десятилетие. На основе экспериментального исследования структуры нагрузки на примере наиболее типичных веб-ресурсов предложены существенно уточненные модели нагрузки.

Ключевые слова: модели нагрузки, компьютерные сети

Анотація

Анопrienко О.Я., Аль Абабнех Хасан, Моделі навантаження в веб-орієнтованих комп'ютерних мережах. Розглянуто особливості зміни параметрів навантаження у веб-орієнтованих комп'ютерних мережах за останнє десятиліття. На основі експериментального дослідження структури навантаження на прикладі найбільш типових веб-ресурсів запропоновано істотно уточнені моделі навантаження.

Ключові слова: моделі навантаження, комп'ютерні мережі

Abstract

Anoprienko A., Ababneh H. Loading models in the web focused computer networks. Main variants of Internet-focused network infrastructure efficiency providing on the base of simulation and Excel-tools are described. Some loading models for the web focused computer networks are proposed.

Keywords: load model, computer networks

Введение

Большинство современных компьютерных сетей в той или иной степени имеют веб-ориентированный характер, что обусловлено как глобальным ростом популярности веб-ресурсов, так и расширяющимся использованием веб-интерфейсов практически во всех современных компьютерных приложениях. В связи с этим вопросы эффективности использования сетевой инфраструктуры при работе с веб-ресурсами приобретают особую актуальность.

В контексте обеспечения эффективности вопросы производительности всегда играли ключевую роль в процессе развития компьютерной техники, в связи с чем их исследованию на всех этапах развития компьютерных

технологий уделялось повышенное внимание. Но до 90-х годов XX века речь шла преимущественно о производительности локальных компьютерных систем (см., например, [1]). В 90-е годы в связи с интенсивным формированием глобальной инфраструктуры Интернет резко начало возрастать использование удаленных ресурсов, что в области сетевых технологий привело к формированию образа «перевернутой пирамиды» при описании изменений, которые произошли в 90-е годы с сетевым трафиком: если к началу этого периода межкомпьютерное взаимодействие в основном замыкалось на уровне локальных сетей, выход за пределы которых был вызван в основном довольно ограниченным кругом информационных потребностей, обусловленных преимущественно обменом отдельными файлами и электронной почтой, то к концу 90-х годов основные источники сетевого трафика находились, как правило, за пределами локальных сетей пользователей.

Параллельно происходило и смещение акцентов с использования преимущественно локальных вычислительных и информационных ресурсов к использованию глобальных ресурсов, что концептуально представлено на рисунке 1, где ось G соответствует «степени глобализации» используемых компьютерных ресурсов. При этом уже на рубеже тысячелетий можно было констатировать примерное равновесие в использовании локальных и глобальных ресурсов, которое уже в последующие годы начало заметно смещаться в сторону использования глобальных ресурсов за счет широкого распространения таких технологий как Интернет-2, Web 2.0, Grid, Cloud computing и др.

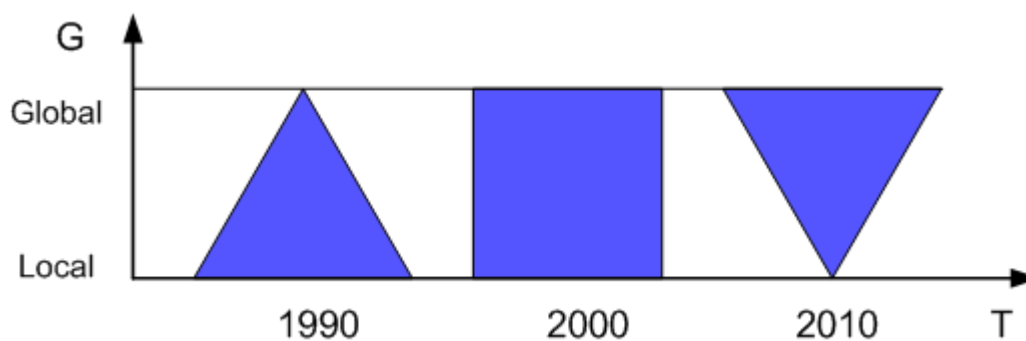


Рис. 1. «Глобализация» использования компьютерных ресурсов: модели соотношения локальных и глобальных ресурсов, используемых типичным пользователем в 1990, 2000 и 2010 гг.

Вопросы производительности при этом, несмотря на постоянный и значительный рост всех технических характеристик, отнюдь не стали менее актуальными. Об этом, в частности, свидетельствуют исследования пользовательских предпочтений, проведенные в 2008 году и показавшие, что для большинства пользователей быстродействие по-прежнему является главным фактором, существенно более важным, чем удобство

использования, гибкость интерфейса и другие показатели, значение которых зачастую преувеличивается (рис. 2).

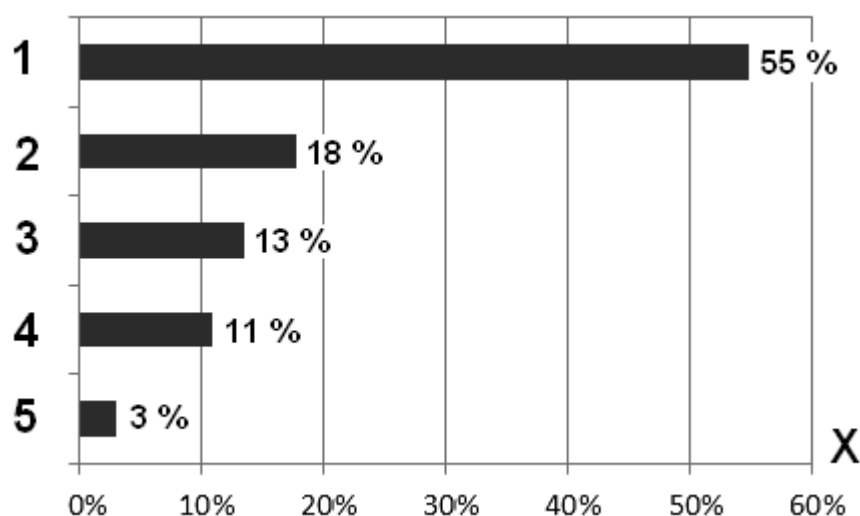


Рис. 2. Сравнительная значимость различных параметров при использовании веб-ресурсов: X – относительная значимость в процентах, 1 – скорость, 2 – минимизация затрат памяти, 3 – адаптивность поведения, 4 – плотность контента, 5 – возможность настройки [2].

В работе [3], опубликованной в 2007 году, были рассмотрены различные подходы к решению проблемы обеспечения заданной производительности веб-ориентированных компьютерных приложений и был предложен подход, основанный на реализации численных моделей инфраструктуры на базе электронных таблиц, что явилось развитием идей, предложенных в работе [4]. При этом было отмечено, что для успешной и продуктивной реализации таких подходов требуются достаточно адекватные модели возможных рабочих нагрузок.

Основные тенденции в эволюции структуры нагрузки

В связи с тем, что большинство приложений приобретает в настоящее время веб-ориентированный характер, существенную роль в формировании итоговых показателей производительности начинают играть не только показатели производительности отдельных компьютерных подсистем, задействованных в веб-ориентированных приложениях, но и состав той сопутствующей информационной нагрузки, которая хранится и используется на серверах при работе с клиентскими приложениями, а также передается им по сетевым каналам связи.

Все более существенную роль, в частности, играют характеристики веб-сайтов как основных интерфейсов для веб-ориентированных приложений. За 12 лет, в период с 1995 по 2008 год, в связи с перманентным усложнением структуры и наполнения сайтов объем хранимой, обрабатываемой и передаваемой информации в расчете на один сайт вырос почти на порядок (рис. 3). При этом переход на технологии web 2.0

(примерно после 2003 года) привел к существенному ускорению этого роста. А все более интенсивное использование в последние годы видеоинформации не только в сугубо информационных и развлекательных целях, но и как эффективного средства научной и образовательной визуализации, ведет к тому, что в последующие годы сугубо информационная и интерфейсная нагрузка значительно вырастет практически для всех веб-приложений.

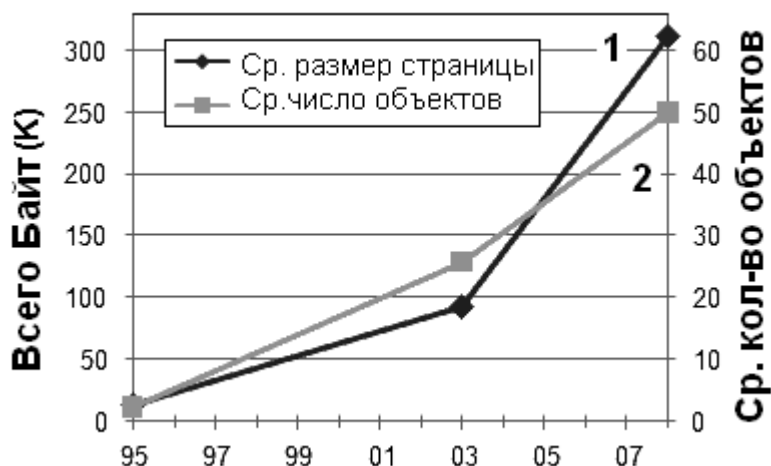


Рис. 3. Рост сложности и средних размеров веб-страниц за период с 1995 по 2008 гг. [5]

Интенсивность использования видеоинформации в Интернет существенно повысилась благодаря проекту YouTube, который был основан в феврале 2005 года на базе использования технологии Flash Video (flv), позволяющей получить хорошее качество записи при относительно небольшом объеме передаваемых данных. Благодаря простоте и удобству использования, YouTube породил явление видеоблоггинга, стимулировавшее резкий рост видеоконтента силами миллионов пользователей, обладающих цифровой видеотехникой. В октябре 2006 года YouTube был куплен компанией Google более чем за 1,5 миллиарда долларов, что стало своеобразным символом перехода к новому этапу развития видеосервисов.

В последующие годы сервис YouTube вышел на немыслимые прежде масштабы использования видеоинформации, обеспечивая параллельный одновременный показ более 100 тысяч различных видеозаписей, что означало просмотр пользователями данного сервиса порядка 100 млн. видеороликов в сутки. Суммарный исходящий трафик YouTube превысил при этом 200 терабайт в сутки. Этого могло бы хватить для трансляции в виде цифрового видеопотока всех существующих на планете телевизионных станций. При этом надо иметь ввиду, что YouTube использует технологию показа роликов, не позволяющую сохранять на диске видеофайлы, что стимулирует дальнейший быстрый рост трафика. В перспективе Google планирует стать самым крупным видеопровайдером на планете, в связи с чем компания строит собственную скоростную оптоволоконную сеть для осуществления этих планов.

В целом с 2006 года человечество ежегодно создается более 100 экзабайт (100 000 000 000 ГБ) цифровой информации. Это примерно эквивалентно 30 млрд. DVD-дисков или 10 стопкам книг высотой от Земли до Солнца. К 2010 году ожидается достижение рубежа в 1 зеттабайт (1000 экзабайт) глобальных информационных ресурсов, содержащихся в совокупности на всех серверах глобального Интернет.

При этом по прогнозам аналитиков, с 2010 года основным генератором трафика в Интернет станет YouTube и подобные ему видеосервисы, которые в совокупности станут источником не менее 50-70 % общего трафика [6]. Каждый месяц трафик YouTube увеличивается примерно на 20 %, и для поддержания такого роста приходится постоянно добавлять новые 10-гигабитные каналы [7]. Кроме увеличения количества видеоматериалов существенную роль играет и постоянный рост размера используемых видеофайлов (рис. 4), который в дальнейшем может заметно ускориться ввиду перехода на цифровое телевидение высокого разрешения. При этом надо иметь ввиду, что стандартный телевизионный трафик должен передаваться с быстродействием 3,5 Мбит/с, в то время как видеотрафик с высоким разрешением — с быстродействием 19 Мбит/с. По преобладающему мнению экспертов, современный Интернет не рассчитан на доставку телепередач с высоким качеством изображения — часовой видеоролик требует передачи по сети такого объема данных, который занимают все электронные письма в мире примерно за год [8].

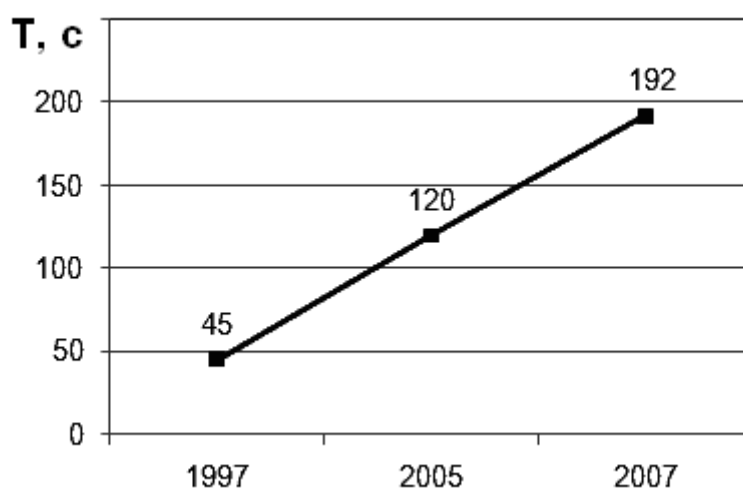


Рис. 4. Рост средней продолжительности видеороликов, доступных в качестве информационных веб-ресурсов [5]

Хотя резкий рост удельного веса видеоинформации и является одним из наиболее заметных факторов изменения структуры нагрузки в современных компьютерных сетях, существенные изменения связаны также и с другими типами файлов, что для периода с 2000 по 2007 представлено в таблицах 1-3.

Таблица 1. Основные статистические показатели изменения размеров файлов, составляющих веб-трафик, в период с 2000 по 2007 гг. [9]

	2000	2007
Минимальный размер, байт	17	85
Максимальный размер, Гбайт	0.23	2.15
Среднее значение, байт	12294	68275
Среднеквадратичное отклонение, байт	321	3425

Таблица 2. Изменения в структуре веб-трафика по количеству запросов различных типов файлов в период с 2000 (слева) по 2007 (справа) гг. [9]

Тип файлов	Удельный вес количества запросов, %	Тип файлов	Удельный вес количества запросов, %
image/gif	53	image/jpeg	33
image/jpeg	25	image/gif	29
text/html	18	text/html	16
application/x-javascript	1	application/x-javascript	7
text/plain	1	image/png	4

Таблица 3. Изменения в структуре веб-трафика по объему различных типов файлов в период с 2000 (слева) по 2007 (справа) гг. [9]

Тип файлов	Удельный объем в общем трафике, %	Средний размер, Кбайт	Тип файлов	Удельный объем в общем трафике, %	Средний размер, Кбайт
image/jpeg	22	10	application/octet-stream	35	1766
image/gif	16	4	image/jpeg	7	13
text/html	15	9	application/x-otrkey	7	240610
application/msword	9	4147	text/plain	6	231
application/octet-stream	8	672	video/x-msvideo	6	109533
application/zip	8	1322	video/x-flv	6	10954
video/mpeg	7	861	video/flv	5	6730
application/ms-excel	3	3637	video/x-ms-wmv	3	42636

text/plain	2	49
audio/mpeg	2	3360

text/html	3	13
application/ zip	2	9632

В частности, следует отметить, что примерно пропорционально росту среднего размера веб-страниц в этот период выросли и средние размеры соответствующих файлов (почти в 5,5 раз) и более чем на порядок выросло разнообразие размеров файлов (табл. 1). Последнее произошло в основном за счет существенного роста использования файлов больших размеров (вплоть до нескольких Гбайт). Количественные характеристики, связанные с частотой использования файлов различных типов, также претерпели заметные изменения (табл. 2). В частности, больше 65 % всех передаваемых по сети файлов по-прежнему составляют изображения, но безусловно лидировавший в 2000 году формат GIF к 2007 году уже уступил первенство формату JPEG (в основном за счет резкого роста масштабов использования цифровых фотографий).

Но наиболее значительные изменения произошли в структуре объемов файлов различных типов (табл. 3). На фоне роста размеров, характерного в указанный период для файлов практически всех типов, наиболее примечательным является следующий факт: если в 2000 году более трети общего объема трафика составляли изображения, то уже в 2007 году лидерство перешло к различным битовым файлам, являющимся преимущественно программным кодом, что является отражением тенденций, связанных с формированием инфраструктуры глобального компьютеринга. Резко уменьшился удельный вес и таких нехарактерных для веб-технологий форматов как текстовые файлы в формате MS Word, электронные таблицы в формате Excel и архивированные файлы в формате zip. Для распространения видео вместо основного в 2000 году формата MPEG преимущественно стал использоваться более эффективный формат FLV. В целом удельный вес видеоинформации в общем трафике вырос с 7 до 20 %. При этом надо отметить, что именно удельный вес видеоинформации наиболее интенсивно растет в последние годы.

Приведенная выше информация в большинстве случаев является достаточной для формирования относительно грубых моделей нагрузки. Однако, в целом ряде случаев для расчета параметров сетевой инфраструктуры требуются существенно более точные модели, учитывающие специфику определенных видов ресурсов, характер распределения размеров файлов и другие параметры. Актуальные данные о такого рода моделях в открытом информационном пространстве выявить не удалось в связи с чем было проведено экспериментальное исследование структуры сетевой нагрузки, формируемой при обращении к веб-серверам различного типа.

Экспериментальное уточнение параметров для типичных моделей рабочей нагрузки

Для исследования были выбраны наиболее типичные виды ресурсов: текстовые, тексто-графические (в том числе на основе формата pdf), графические (преимущественно на основе jpg), аудио (mp3 и wma), видео (преимущественно на основе flv). В качестве инструмента для проведения экспериментов был выбран Интернет браузер Opera v 9.64, позволяющий достаточно удобно анализировать состояние кеш-памяти браузера в процессе исследования с целью сбора необходимой статистической информации.

Сбор статистики велся по двум основным параметрам:

1. Объем целевого информационного файла (носитель информации).
2. Общий объем целевого информационного сайта (в совокупности со всеми сопутствующими файлами)

Объектом для исследований текстового формата был выбран известный сайт Lib.ru (Библиотека Машкова). В ходе исследований оценивался объем отдельно текстового файла и совокупный объем страницы с ресурсом. Получено 1596 значений. Результаты представлены на рис. 5.

Объектом для исследований графического формата был выбран типичный сайт содержащий широкоформатные изображения высокого разрешения, предназначенные для использования в качестве заставок рабочего стола ОС Windows XP/Vista – bestwallpapers.net.ru. В ходе исследования оценивался объем графического файла и совокупный объем страницы с ресурсом. Получено 1000 значений по каждому параметру. Результаты представлены на рис. 6.

Объектом для исследований аудио-ресурсов был выбран сайт 6rb2d.com как яркий представитель соответствующих ресурсов сети Интернет. В ходе исследования объем ресурса и аудиофайла оценивался отдельно. Получено 77 значений. Результаты представлены на рис. 7.

Объектом для исследований видео формата был выбран сайт YouTube.com. В ходе исследования оценивался объем видеофайла и совокупный объем страницы с ресурсом. Получено 152 значения. Результаты представлены на рис. 8.

Исходя из специфики экспериментов и большого массива данных выбор метода статистического анализа был сделан в пользу «Подгонки вероятностных распределений к реальным данным». Статистический анализ данных велся с помощью программы STATISTICA 7.0. Под подгонкой (английский термин fitting) при этом имелись ввиду аналитические процедуры, позволяющие подобрать распределение, которое с достаточной степенью точности описывает наблюдаемые данные. Другими словами, имея значения переменной X , мы проверяли гипотезы, согласно которым распределение X описывается вероятностным законом F .

Наиболее показательным критерием согласия наблюдаемых данных с гипотезой является критерий хи-квадрат Пирсона. Для применения этого критерия область значений переменной X вначале была разбита на некоторое число интервалов N . Затем подсчитывалось число наблюдений, попавших в i -й интервал, что обозначалось как $n(i)$. Полученное значение сравнивалось со средней или ожидаемой при гипотезе частотой, обозначенной как $\underline{n}(i)$.

Статистика хи-квадрат вычислялась следующим образом:

$$\text{Хи - квадрат} = \sum ((n(i) - \underline{n}(i))^2) / \underline{n}(i).$$

В этой формуле суммирование распространяется на все интервалы, на которые разбита область значений переменной. При этом сравнивались наблюдаемые и ожидаемые частоты. Статистика принимает значения от нуля до бесконечности. Чем меньше значение статистики хи-квадрат, тем более вероятно, что гипотеза верна, чем больше значение статистики хи-квадрат, тем меньше вероятность того, что гипотеза соответствует данным.

Исходя из вышеизложенного статистика хи-квадрат – это разумная мера согласия (соответствия) данных с гипотезой. В ходе анализа данных в определенных случаях были внесены поправки в степень свободы – параметр (df). По уровню значимости $p = 10^{-7}$ можно сделать вывод о том, что данные подчиняются законам распределения, указанным на графиках. Вероятность ошибиться при отклонении данных гипотез составляет 99,99%. Т.е. уровень доверительной вероятности данной гипотезе составляет 99,99%. Граничный уровень значения p составляет 0,05 – это означает, что любой вывод считается статистически значимым на уровне $p < 0,05$. В ходе анализа, на основе вышеизложенных критериев, сделан вывод о том, что распределение большинства исследованных величин подчиняется нормальному закону.

Для текстовых ресурсов дополнительно был исследован вариант, когда текстовые файлы представлены в архивированном виде. В этом случае было получено гамма-распределение (рис. 9).

Объектами для исследований тексто-графического формата был выбран сайты books.google.com. Но в данном случае не удалось достаточно полно оценить распределение объемов целевых файлов ввиду некоторых технических особенностей данных ресурсов. Из-за специфической политики безопасности ресурса Google.com не удалось получить достаточно большой для статистического анализа объем выборки. Оценивался совокупный объем страницы с ресурсом. Получено 15 значений, характер распределения которых в диапазоне от 10-ти до 32-х Мбайт близок к хаотичному (рис. 10). Предполагается, что при получении статистически достаточного объема выборки для указанного диапазона может быть выявлен нормальный закон распределения. Теоретической основой нормального закона распределения вероятностей является центральная предельная теорема Ляпунова, утверждающая, что распределение суммы независимых случайных величин с любым исходным распределением будет нормальным, если число слагаемых достаточно велико, а вклад каждого в сумму мал [10, с. 148]. Именно такая ситуация возникает при оценке объемов отдельных целевых файлов.

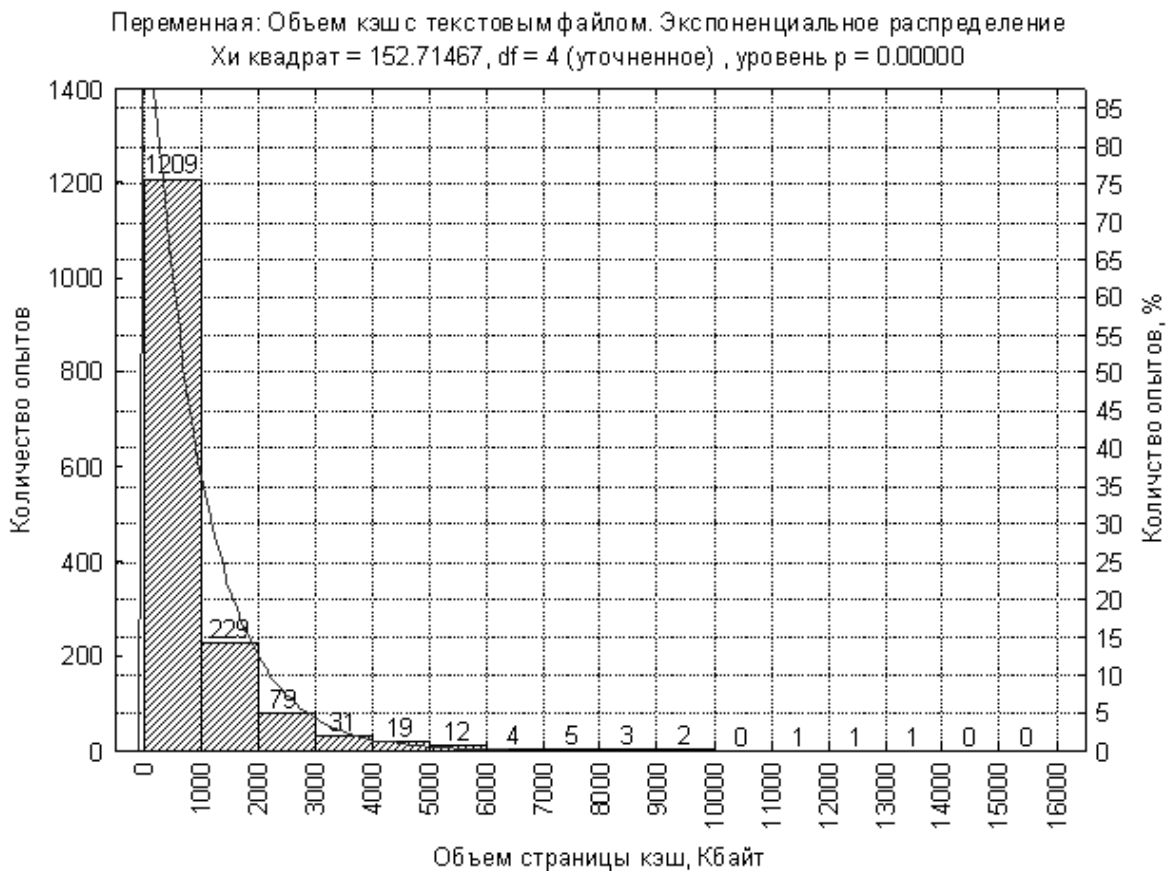
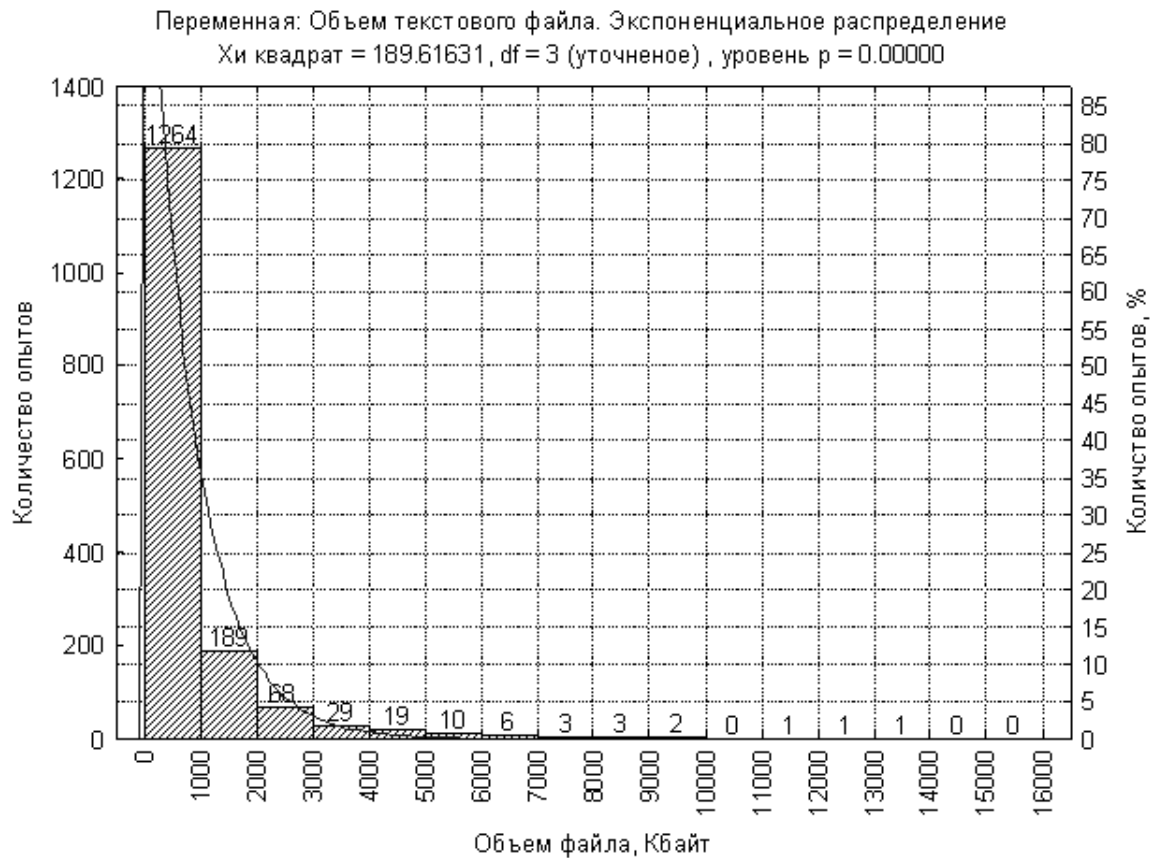


Рис. 5. Экспоненциальное распределение размеров целевых текстовых файлов (вверху) и размеров соответствующих сайтов (внизу) для ресурсов текстового типа (на примере сайта Lib.ru)

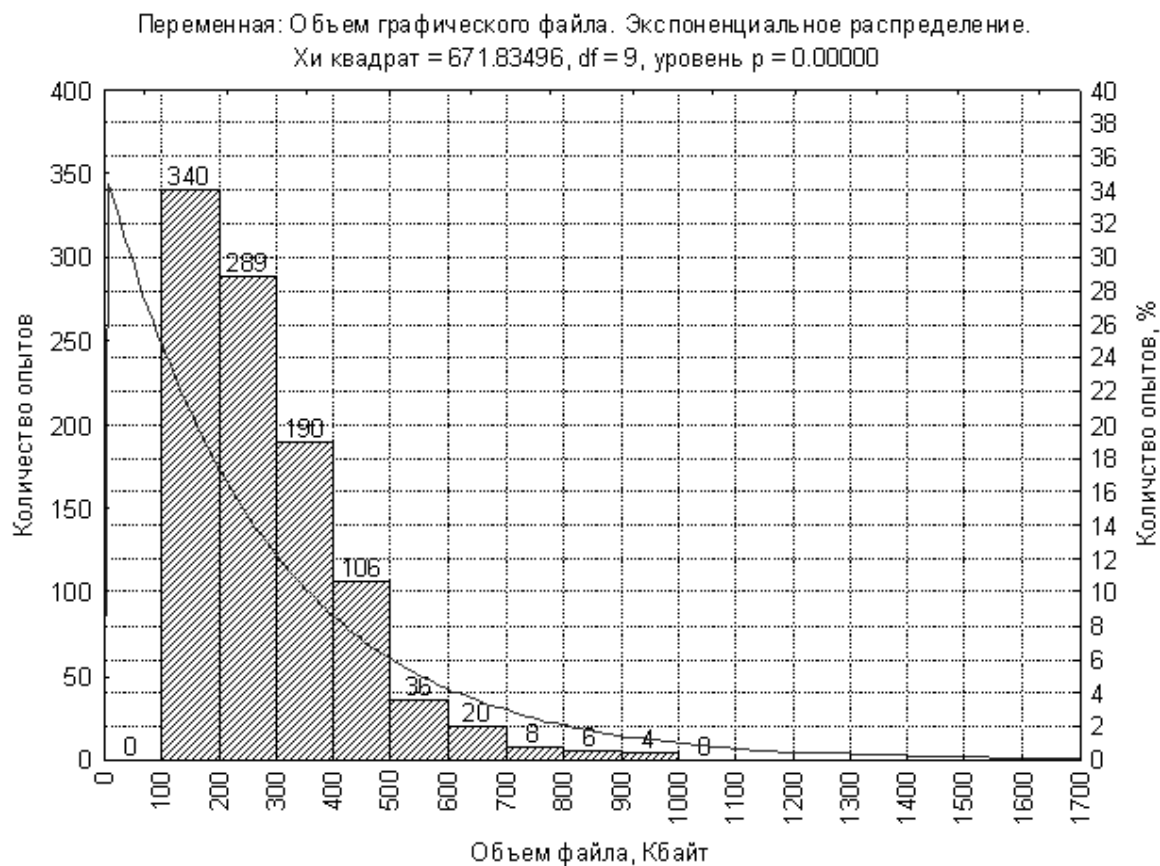


Рис. 6. Экспоненциальное распределение размеров целевых графических файлов (вверху) и нормальное распределение размеров соответствующих сайтов (внизу) для ресурсов графического типа (на примере сайта bestwallpapers.net.ru)

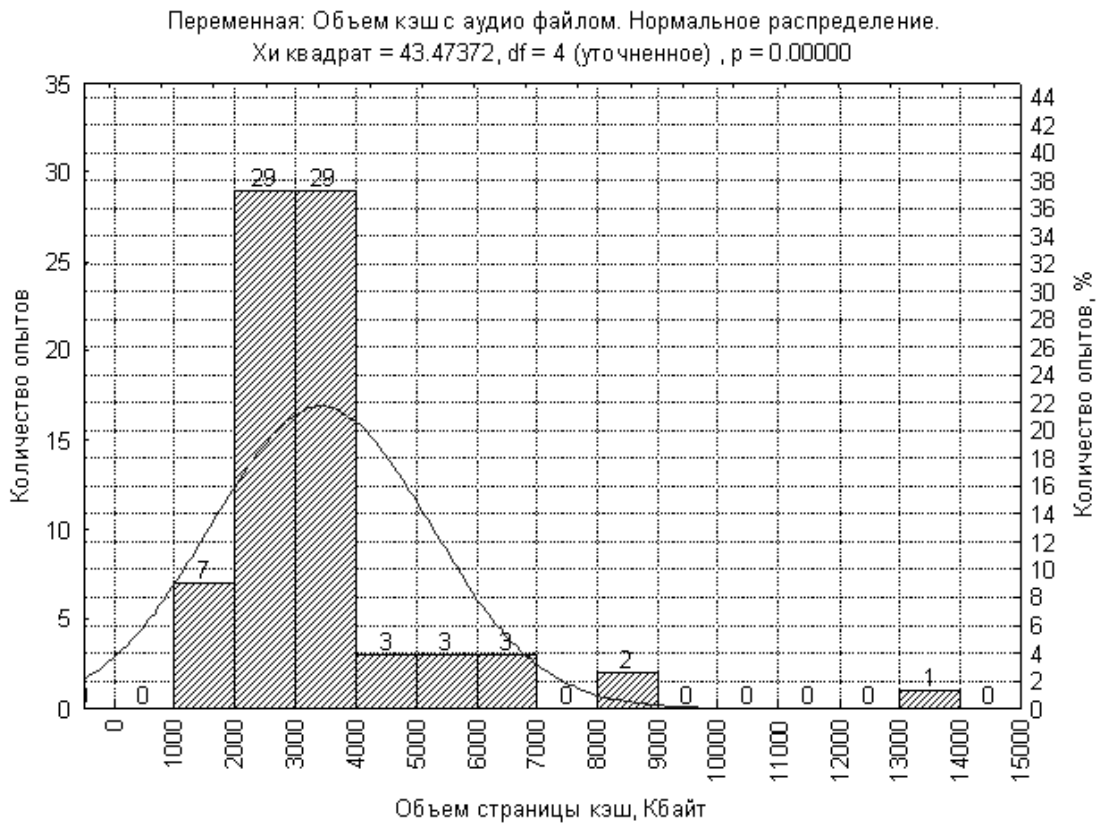
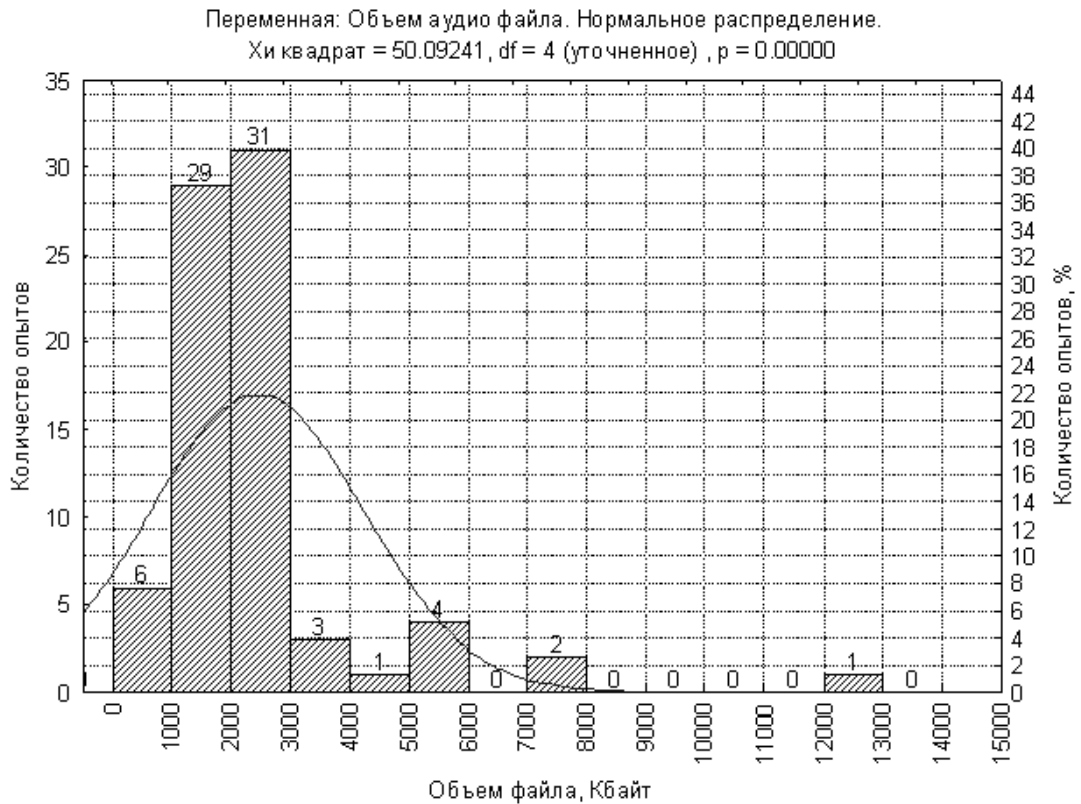


Рис. 7. Нормальное распределение размеров целевых аудиофайлов (вверху) и нормальное распределение размеров соответствующих сайтов (внизу) для аудиоресурсов (на примере сайта 6rb2d.com)



Рис. 8. Нормальное распределение размеров целевых видеофайлов (вверху) и нормальное распределение размеров соответствующих сайтов (внизу) для видеоресурсов (на примере сайта YouTube.com)

Полученные результаты вполне согласуются с тем, что нормальное распределение является краеугольным камнем математической статистики в силу следующих причин:

- схема его возникновения соответствует многим реальным физическим процессам, порождающим результаты обрабатываемых наблюдений;
- при возрастании объема выборки предельное распределение для большинства распределений является нормальным и с успехом может использоваться для аппроксимации последних;
- нормальное распределение обладает рядом благоприятных математико-статистических свойств (легко нормируется и аппроксимируется, обладает свойством аддитивности).

Соответствующая функция распределения имеет следующий вид:

$$F(x; \mu; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} dt$$

где σ – стандартное отклонение.
 μ – среднее значение.

В целом можно сделать следующие выводы о массивах полученных экспериментальных данных, подчиняющихся нормальному закону распределения:

- имеется сильная тенденция группироваться вокруг центра (в нашем случае таким центром является среднее значение каждой выборки);
- положительные и отрицательные отклонения от центра равновероятны;
- частота отклонений быстро падает, когда отклонения от центра становятся большими.

В ряде случаев (текстовые ресурсы) было получено экспоненциальное распределение, часто встречающееся в теории надежности и в теории массового обслуживания. Например, наработка на отказ большой многокомпонентной системы может быть описана экспоненциальным распределением при любом распределении наработки на отказ компонентов системы. Функция распределения при этом имеет следующий вид:

$$F(x; b) = 1 - \exp\left(-\frac{x}{b}\right), \quad x \geq 0$$

где b – среднее.

Экспоненциальное распределение является частным случаем гамма-распределения, которое также было выявлено в случае исследования ресурса с архивированными текстовыми файлами (рис. 9).

Заключение и перспективы исследований

Проведенные исследования показали эффективность предложенного подхода на фоне относительной простоты его реализации, что позволяет признать целесообразной реализацию целого комплекса моделей, подобных описанной в данной работе, позволяющих производить достаточно точную оценку адекватности различных элементов сетевой инфраструктуры реальной и прогнозируемой нагрузке. Полученные в результате проведенных исследований результаты планируется использовать при построении уточненных моделей нагрузки для расчета параметров веб-ориентированных компьютерных сетей.

Литература

1. Феррари Д. Оценка производительности вычислительных систем. – М.: Мир, 1981. – 576 с.
2. Study: Web Users Prefer Speed Over Customization // Website optimization, <http://www.websiteoptimization.com/speed/tweak/design-factors>.
3. Аноприенко А.Я., Аль Абабнех Хасан. Повышение эффективности Интернет-ориентированной сетевой инфраструктуры: Методы, задачи и инструменты // Научные труды Донецкого национального технического университета. Серия "Проблемы моделирования и автоматизации проектирования динамических систем" (МАП-2007). Выпуск 6 (127): Донецк: ДонНТУ, 2007. С. 228-233.
4. Менаске Д., Алмейда В. Производительность Web-служб. Анализ, оценка и планирование: Пер. с англ. – СПб: ООО «ДиаСофтЮП», 2003. – 480 с.
5. Average Web Page Size Triples Since 2003 // Website optimization, <http://www.websiteoptimization.com/speed/tweak/average-web-page>.
6. YouTube съест 70% трафика на планете / CNews.ru: Главные новости, 07.03.2007, <http://www.cnews.ru/news/top/index.shtml?2007/03/07/239221>.
7. Рыбаков А. YouTube убьет Интернет? / Вебпланета, 30/01/2007, <http://webplanet.ru/>.
8. Инфраструктура Интернета не справляется с онлайн-телевидением // NBSР. Журнал для вебмастеров, Февраль 12, 2007, <http://www.nbsp.ru/news/index.html>.
9. Evolution of the Web from 2000 to 2007 // Website optimization, <http://www.websiteoptimization.com/speed/tweak/evolution-web>
10. Вероятностные методы в вычислительной технике. – М.: Высшая школа, 1986. – 312 с.



Аноприенко Александр Яковлевич.

К. т. н., доцент, декан факультета компьютерных наук и технологий ДонНТУ.

Научные интересы: Интернет и веб-технологии, компьютерное моделирование и компьютерная графика.



Аль Абабнех Хасан.

Стажер кафедры ЭВМ ДонНТУ. Закончил аспирантуру при кафедре под руководством Аноприенко А.Я.

Научные интересы: Методы и средства расчета и обеспечения эффективности компьютерной инфраструктуры.

Дата надходження до редакції 22.12.2008 р.

Как правильно ссылаться на эту статью:

Аноприенко А.Я., Аль Абабнех Хасан, Модели нагрузки в веб-ориентированных компьютерных сетях // Научные труды Донецкого национального технического университета. Серия «Проблемы моделирования и автоматизации проектирования динамических систем» (МАП-2008). Выпуск 7 (150): Донецк: ДонНТУ, 2008. С. 258-274.