

Исследование аналитического метода оценки эффективности сетей доставки контента

П. А. Приставка, О. А. Ключникова, И. В. Климова

В данной статье рассматривается решение задачи аналитической оценки эффективности сетей передачи данных на основе метода, предложенного Б. Я. Рябко. Показывается и анализируется зависимость итоговой оценки от различных параметров: объема кэширующих серверов, а также скорости канала передачи данных между клиентом и оригинальным сервером. На основе полученных расчетных данных делается вывод о логическом соответствии результатов вычислений характеру изменений, вносимых в параметры сети, показывается адекватность аналитической оценки.

Ключевые слова: сети передачи данных, оценка эффективности.

1. Введение

Сеть Интернет создавалась как средство связи между двумя узлами и в настоящее время эволюционировала из сети коммуникации в глобальную платформу для бизнеса, развлечения, средств массовой информации. Ежедневно в сеть выходят миллиарды различных устройств, а количество хранящейся в сети информации исчисляются уже в сотнях эксабайтов. В связи с тем, что Интернет первоначально не был предназначен для широкого распространения самой различной цифровой информации и для обеспечения надежного и быстрого доступа к ней пользователей, за последние десятилетия появилось множество решений для адаптации сети Интернет к современным потребностям.

Одним из таких решений являются различные кэширующие механизмы. Основным смыслом таких решений заключается в создании буфера с быстрым доступом, но сравнительно небольшим объемом, который содержит именно ту информацию, которая с большей вероятностью будет востребована. В качестве примера кэширующей сети в данной статье будут рассмотрены сети доставки контента (CDN).

В настоящее время для оценки эффективности таких сетей широко используются экспериментальные показатели, такие как, например, попадание в кэш (наличие запрашиваемых данных в кэше буфера). Указанный подход позволяет вычислить характеристику уже созданной сети, однако не позволяет получить необходимую оценку на этапе проектирования сети. В работе Б. Я. Рябко «Using Information Theory to Study the Efficiency and Capacity of Caching in the Computer Networks» [1] представлен метод для аналитической оценки эффективности создаваемой сети. В указанной работе описан общий подход к решению задачи, а также рассмотрены базовые примеры. Данная статья посвящена исследованию рассматриваемого метода. Целью работы является анализ зависимости итоговой оценки эффективности моделируемой сети от значений таких ее параметров, как объем кэширующих серверов и скорость канала передачи данных между оригинальным сервером и клиентом. Кроме того, на основе полученных данных делается вывод о логическом соответствии характера изменений вычисляемой оценки устанавливаемым значениям параметров сети.

2. Описание CDN-инфраструктуры

CDN (Content Delivery Network) – это географически распределенная программно-аппаратная платформа, которая оптимизирует доставку данных до конечных пользователей Интернета. В начале 2000-х CDN-сети получили широкую известность благодаря своей эффективности и до сих пор остаются одним из самых востребованных решений для быстрого получения контента в сети Интернет.

В основе CDN-инфраструктуры лежит подход, согласно которому часто запрашиваемые данные распределяются по географически удаленным кэширующим серверам. Таким образом, пользователи получают информацию не с оригинального сервера, а с ближайшего граничного сервера.

Клиентом CDN-услуги обычно является компания, которая заинтересована в том, чтобы информационные услуги её пользователям предоставлялись эффективным способом. CDN-провайдер предоставляет свои граничные серверы и различные инструменты для управления и мониторинга клиентом своих данных.

Большинство современных CDN имеют схожую структуру и состоят из оригинального сервера (кластера серверов), граничных серверов, а также системы управления компонентами. На рис. 1 изображена инфраструктура CDN компании Akamai [2], которая на сегодняшний день является лидером в области предоставления данной услуги. Кратко указанная инфраструктура может быть описана следующим образом.

Система отображения (Mapping System) переводит вводимый пользователем URL в IP-адрес граничного сервера, выбранного согласно состоянию сети. Граничные серверы (Edge Server Platform) отвечают за обработку и удовлетворение запросов пользователей. Иногда для этого требуется запрос данных с оригинального сервера (origin), который осуществляется через транспортную систему (Transport System). Оригинальный сервер принадлежит и контролируется зачастую клиентом CDN-сети, а не оператором. Система контроля и коммуникации (Communications and Control System) используется для распространения управляющих сообщений, сообщений о статусе информации и обновлениях конфигураций.

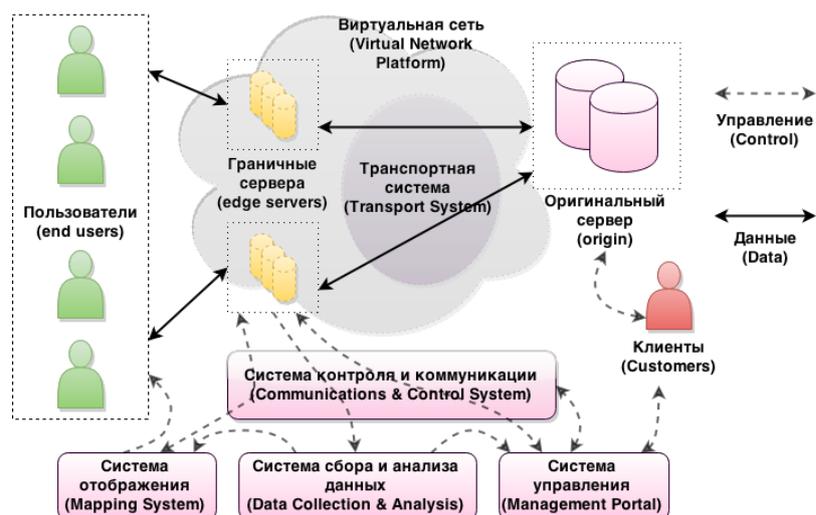


Рис. 1. Инфраструктура CDN-сети

Система сбора и анализа данных (Data Collection and Analysis System) ответственна за сбор информации из журналов серверов и клиентов и данных о состоянии сети. Полученные данные используются для мониторинга, анализа, составления отчетов и счетов. Система управления (Management Portal) предоставляет клиенту, во-первых, инструменты контроля

приложений и информации, доступных пользователям; во-вторых, информацию об аудитории, использующей эти ресурсы.

3. Оценка эффективности

Эффективность CDN-сети заключается в способности сети предоставить пользователям запрашиваемую услугу и/или информацию. Оценка эффективности сети в настоящее время характеризуется комбинацией различных экспериментальных данных, таких как: попадание в кэш (cache hit ratio), то есть отношение количества информации, хранящейся в кэше, к общему количеству затребованной информации; резервная пропускная способность, то есть пропускная способность оригинального сервера; время отклика сети, осознаваемое пользователем; количество потерянных данных при передаче; время, в течение которого граничный сервер остается занятым после запроса.

Таким образом, для оценки эффективности сети используются оценки эффективности различных её компонентов. Предлагаемый метод, напротив, позволяет оценить производительность кэширующей сети, такой как CDN, в целом.

3.1. Описание алгоритма

В работе Б. Я. Рябко «Using Information Theory to Study the Efficiency and Capacity of Caching in the Computer Networks» [1] предлагается основанный на теории информации подход для оценки эффективности сетей доставки контента. Ниже приведем теоретическое описание рассматриваемого метода.

Рассмотрим сеть Ω с количеством узлов n $w_1 \dots w_n$ и некоторым набором файлов $F_i, i = 1, 2, \dots, n$ на каждом узле. Обозначим время загрузки файла $f \in F_j, j = 1, 2, \dots, n$ узлом w_i из узла w_j через $T_{ij}(f)$. Будем считать, что для двух различных узлов ($w_{j_1} \neq w_{j_2}$) $T_{ij_1}(f) \neq T_{ij_2}(f)$ (т.е. время загрузки узлом файла с двух различных, обладающих необходимым файлом, узлов различно) и требуемый файл будет загружаться с узла, позволяющего достичь минимального времени передачи (скачивания), т.е.

$$T_{w_i}(f) = \min_j T_{ij}(f).$$

Задачей f^{\wedge} будем считать последовательность файлов $f_1 \dots f_s, s \geq 1$, загружаемую узлом $w_i, i = 1, 2, \dots, n$. Время загрузки такой последовательности будет вычисляться как сумма времен загрузки каждого файла из этой последовательности:

$$T_{w_i}(f^{\wedge}) = \sum_{k=1}^s T_{w_i}(f_k).$$

Положим, что v_i – количество задач, которые могут быть выполнены узлом w_i за время

$$T: v_i(T) = |\{f^{\wedge}: T_{w_i}(f^{\wedge}) = T\}|.$$

Обратим внимание, что при больших T количество задач, которые могут быть выполнены за время $2T$, т.е. $v_i(2T)$, равняется $|v_i(T)|^2$, потому что задачи (последовательности файлов), которые могут быть выполнены за время $2T$, представляют собой все комбинации множества задач, выполняемых за время T (первая половина промежутка) с таким же множеством (задачи, выполняемые за вторую половину промежутка). Будем считать это наблюдение нестрогим объяснением экспоненциального роста количества задач, выполняемых узлом задач в зависимости от времени.

В оригинальной англоязычной статье Б. Я. Рябко [1] для обозначения показателя, характеризующего эффективность сети доставки контента, использовался термин *capacity*, кото-

рый мы, во избежание двусмысленности, предлагаем переводить как «загрузочная способность». Следует особенно отметить, что данный термин обозначает не пропускную способность, а некоторую вводимую характеристику эффективности, большее значение которой означает лучший результат.

Итак, так как v_i растёт экспоненциально, то определим загрузочную способность узла w_i как:

$$C(w_i) = \limsup_{T \rightarrow \infty} \frac{\log v_i(T)}{T},$$

бит за единицу времени.

Загрузочной способностью сети будет являться сумма загрузочных способностей всех её узлов:

$$C(\Omega) = \sum_{w \in \Omega} C(w).$$

Одним из способов нахождения $C(w_i)$ является теоретико-информационный метод, предложенный Шенноном в [3], согласно которому для нахождения искомого значения требуется найти логарифм наибольшего положительного корня уравнения:

$$\sum_{f \in F} X^{-T w_i(f)} = 1.$$

Отметим здесь, что при вычислении степеней характеристического уравнения необходимо использовать данные о скорости передачи информации между узлами, таким образом, указанная характеристика оказывает важное, а иногда и ключевое, влияние на итоговую оценку эффективности. Характеристическое уравнение может быть решено с помощью метода деления пополам [4]. Очевидно, что за один и тот же промежуток времени пользователем с одного узла может быть скачано различное количество информации, поэтому отметим, что в наших расчетах под скоростью передачи данных подразумевается наибольшая возможная скорость передачи пользовательских данных. Соответственно, получаемая загрузочная способность сети – это оценка верхней границы эффективности CDN.

Более подробное описание рассматриваемого метода, а также примеры расчета оценки загрузочной способности сетей с простейшими конфигурациями могут быть найдены в [1]. Кроме того, в данной работе ниже, в разделе, посвященном вычислительной части исследования, может быть найден подробный пример практических расчетов для заданной моделируемой сети.

3.2. Исследование аналитического метода на примере модели сети доставки контента

В качестве основных целей выполненного исследования можно выделить: вычисление оценки эффективности CDN для более сложных, чем рассматриваемые в оригинальной статье Б. Я. Рябко, примеров конфигураций сетей, анализ зависимости итоговой оценки эффективности моделируемой сети от значений таких ее параметров, как объем кэширующих серверов и скорость канала передачи данных между оригинальным сервером и клиентом, анализ логического соответствия характера изменений вычисляемой оценки устанавливаемым значениям параметров сети.

Для оптимизации вычислений на языке C++ был разработан программный комплекс, состоящий из двух модулей: модуля, предназначенного для моделирования состояния сети по заданной конфигурации, и модуля, производящего расчеты загрузочной способности на основе данных, передаваемых первым модулем. На основе текущего состояния сети первым модулем производится вычисление коэффициентов характеристического уравнения описываемого метода, после чего они передаются в качестве входных данных второму модулю для нахождения корней характеристического уравнения методом половинного деления и расчета загрузочной способности сети.

Содержание проводимых вычислений может быть описано следующим образом. На вход программного комплекса подается файл, определяющий конфигурацию моделируемой сети, а именно, следующие параметры:

- количество кэширующих серверов;
- размер каждого кэширующего сервера (в данной работе он одинаков для всех кэширующих серверов);
- структура связей между пользователями, кэширующими серверами и оригинальным сервером (определяет наличие или отсутствие канала передачи данных между конкретными узлами сети);
- скорости передачи данных между конечными пользователями и кэширующими серверами;
- скорость передачи данных между конечными пользователями и оригинальным сервером;
- количество групп файлов, находящихся на оригинальном сервере;
- количество файлов внутри группы (одинаковый для всех групп);
- размер файла группы (одинаковый для всех групп).

Далее в ходе первого шага серии вычислений часть групп файлов оригинального сервера случайным образом (с помощью линейного конгруэнтного генератора псевдослучайных чисел) распределяется по кэширующим серверам. Процесс распределения групп файлов считается завершенным, если на кэширующих серверах не осталось места, доступного для хранения одной группы. Для текущего состояния сети программным комплексом на основе метода, описанного в п. 3.1, оценивается загрузочная способность. После этого выполняется второй шаг серии вычислений, в рамках которого текущая конфигурация сохраняется прежней, но группы файлов оригинального сервера вновь распределяются по кэширующим серверам произвольным образом, задавая другое состояние сети передачи данных. Для обновленного состояния вычисляется оценка загрузочной способности сети. Всего в рамках серии вычислений над заданной конфигурацией проводится десять расчетов (шагов), после чего вычисляется итоговая оценка загрузочной способности сети как усредненная сумма десяти оценок. Ввиду того, что конфигурация моделируемой сети оставалась неизменной, а менялось лишь ее состояние, полученную итоговую оценку предлагается считать итоговой оценкой загрузочной способности для заданной конфигурации. Аналогичные серии вычислений повторяются для других конфигураций сети передачи данных.

Как отмечалось выше, одной из целей исследования стал анализ зависимости итоговой оценки эффективности моделируемой сети от значений конкретных ее параметров, поэтому при выполнении серий вычислений оставались неизменными все параметры конфигурации сети, кроме:

- объема кэширующих серверов;
- скорости канала передачи данных между главным сервером и пользователем (в случае обращения к главному серверу при отсутствии файла на кэширующем сервере).

Таким образом, на данном этапе исследований была произведена серия вычислений, в ходе каждого из которых для заданной конфигурации производилось вычисление оценки загрузочной способности при различных состояниях сети (различное распределение групп файлов по кэширующим серверам) и формировалась итоговая оценка эффективности путем усреднения. Изменение лишь строго определенных параметров конфигурации и анализ соответствующей оценки, получаемой на основе метода, позволили установить характер зависимости между значением загрузочной способности как предлагаемого показателя эффективности и такими параметрами, как объем кэширующих серверов и скорость передачи данных между оригинальным сервером и пользователем.

Для наглядности приведем пример расчета загрузочной способности моделируемой сети в определенном состоянии, которое представлено на рис. 2.

Рассматриваемая сеть состоит из оригинального сервера, содержащего 25 групп файлов. Каждая группа состоит из 10^6 файлов размером 1 Гб каждый. В сети доступны три гранич-

ных сервера объемом $3 \cdot 10^6$ Гб каждый, на которых заэшированы соответствующие группы файлов. Например, кэширующий сервер 1 содержит 1, 2 и 10 группы файлов. В качестве конечных потребителей контента в сети выступают три пользователя, возможность скачивания файлов показана направлениями стрелок. Скорость скачивания файлов с соответствующих граничных серверов указана на рис. 2 рядом со стрелками. Например, пользователь 3 может скачивать файлы со второго граничного сервера со скоростью 110 Мбит/с, а с третьего – со скоростью 130 Мбит/с. Скорость скачивания файлов с оригинального сервера составляет 20 Мбит/с. Отметим еще раз, что при выполнении серий вычислений исследования оставались неизменными все параметры CDN, кроме объема кэширующих серверов и скорости передачи данных между пользователем и оригинальным сервером.

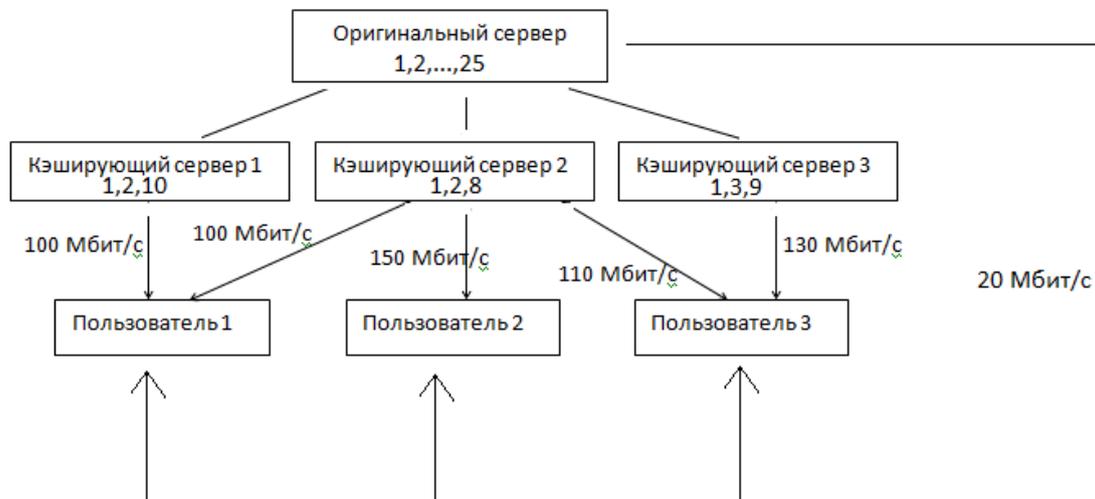


Рис. 2. Состояние моделируемой сети

Как было указано в п. 3.1, загрузочная способность сети определяется как сумма пропускных способностей ее узлов. Из определения загрузочной способности узла следует, что если узел не может скачивать файлы с других узлов сети, то его пропускная способность равна нулю. Следовательно, для нахождения искомой характеристики необходимо определить только пропускные способности узлов пользователей.

Рассмотрим подробно составление уравнения для первого пользователя. Первый файл первой группы файлов расположен на первом и втором кэширующих серверах. Поскольку скорость скачивания с этих узлов первым пользователем одинаковая, то будем считать, скачивание производилось с первого граничного сервера. Так как доступная скорость составляет 100 Мбит/с, то время скачивания первого файла группы составит 81.92 с ($1 \text{ Гб} : 100 \text{ Мбит/с}$). Очевидно, такое же время займет скачивание каждого из оставшихся файлов первой группы. Следовательно, первое слагаемое характеристического уравнения для первого пользователя примет вид:

$$\frac{10^6}{x^{81.92}}.$$

Аналогично получаем второе слагаемое характеристического уравнения (для файлов второй группы):

$$\frac{10^6}{x^{81.92}}.$$

Очевидно, что слагаемые для восьмой и десятой группы будут иметь идентичный вид.

Третья группа файлов не является заэкшированной ни на одном из доступных первому пользователю граничных серверов, поэтому каждый файл этой группы может быть скачан только с оригинального сервера за время 409.6 с (1 Гб : 20 Мбит/с). Следовательно, третье слагаемое характеристического уравнения примет вид:

$$\frac{10^6}{x^{409.6}}.$$

Из аналогичных соображений слагаемые для групп 4–7, а также 11–25 будут такими же, как и для третьей группы.

Таким образом, полное характеристическое уравнение для первого пользователя примет вид:

$$\begin{aligned} \frac{10^6}{x^{81.92}} + \frac{10^6}{x^{81.92}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{81.92}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{81.92}} + \frac{10^6}{x^{409.6}} \\ + \frac{10^6}{x^{409.6}} \\ + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{409.6}} + \frac{10^6}{x^{409.6}} = 1 \end{aligned}$$

или

$$\frac{4 * 10^6}{x^{81.92}} + \frac{21 * 10^6}{x^{409.6}} = 1.$$

Решая полученное уравнение методом половинного деления, получаем оценку пропускной способности для узла первого пользователя:

$$C(w_1) = \log_2 1.203903 = 0.268,$$

бит/с.

Аналогично для узла второго пользователя характеристическое уравнение будет выглядеть как

$$\frac{3 * 10^6}{x^{54.61}} + \frac{22 * 10^6}{x^{409.6}} = 1.$$

Таким образом, $C(w_2) = \log_2 1.314035 = 0.394$, бит/с.

Характеристическим уравнением узла третьего пользователя является равенство вида

$$\frac{3 * 10^6}{x^{63.02}} + \frac{2 * 10^6}{x^{74.47}} + \frac{20 * 10^6}{x^{409.6}} = 1.$$

При составлении данного уравнения следует обратить внимание, что заэкшированная на втором и третьем доступных граничных серверах первая группа файлов будет скачиваться с последнего, так как для третьего пользователя скорость скачивания с этого узла выше.

После решения уравнения получим $C(w_3) = \log_2 1.267873 = 0.342$, бит/с.

Совокупная пропускная способность всей рассматриваемой CDN-инфраструктуры определяется следующим образом:

$$C(\Omega) = C(w_1) + C(w_2) + C(w_3) = 0.268 + 0.394 + 0.342 = 1.004,$$

бит/с.

Итоговые результаты исследования аналитического метода оценки эффективности CDN представлены в табл. 1.

Таблица 1. Результаты вычислений

Объем кэширующих серверов, 10 ⁶ Гб	Загрузочная способность сети, бит/с	
	Скорость передачи данных между ориг. сервером и пользователем	
	20 Мбит/с	90 Мбит/с
3	1.010690	1.01834
5	1.043080	1.04821
7	1.064820	1.06808
10	1.087410	1.08883
12	1.098980	1.10014
15	1.111890	1.11267
18	1.123230	1.12296
20	1.128240	1.12878
22	1.133300	1.13334
25	1.139970	1.13997

На основе полученных данных были построены представленные на рис. 3 графики зависимости загрузочной способности сети от объема кэширующих серверов и скорости канала между пользователем и главным сервером.

На рис. 3 по оси абсцисс расположены значения объема кэширующих серверов, а по оси ординат – соответствующие значения оценки загрузочной способности сети. Синим цветом изображен график зависимости исследуемой характеристики от объема кэширующих серверов для канала передачи данных между конечным пользователем и оригинальным сервером со скоростью 20 Мбит/с, а красным – аналогичный график для скорости 90 Мбит/с. Графики демонстрируют, что оценка загрузочной способности улучшается при увеличении скорости канала между конечным пользователем и оригинальным сервером при одинаковом значении объема кэширующих серверов. Кроме того, увеличение объема кэширующих серверов также приводит к улучшению исследуемой характеристики. Очевидно, что полученные результаты логически соответствуют ожиданиям при внесении указанных изменений в конфигурацию сети.

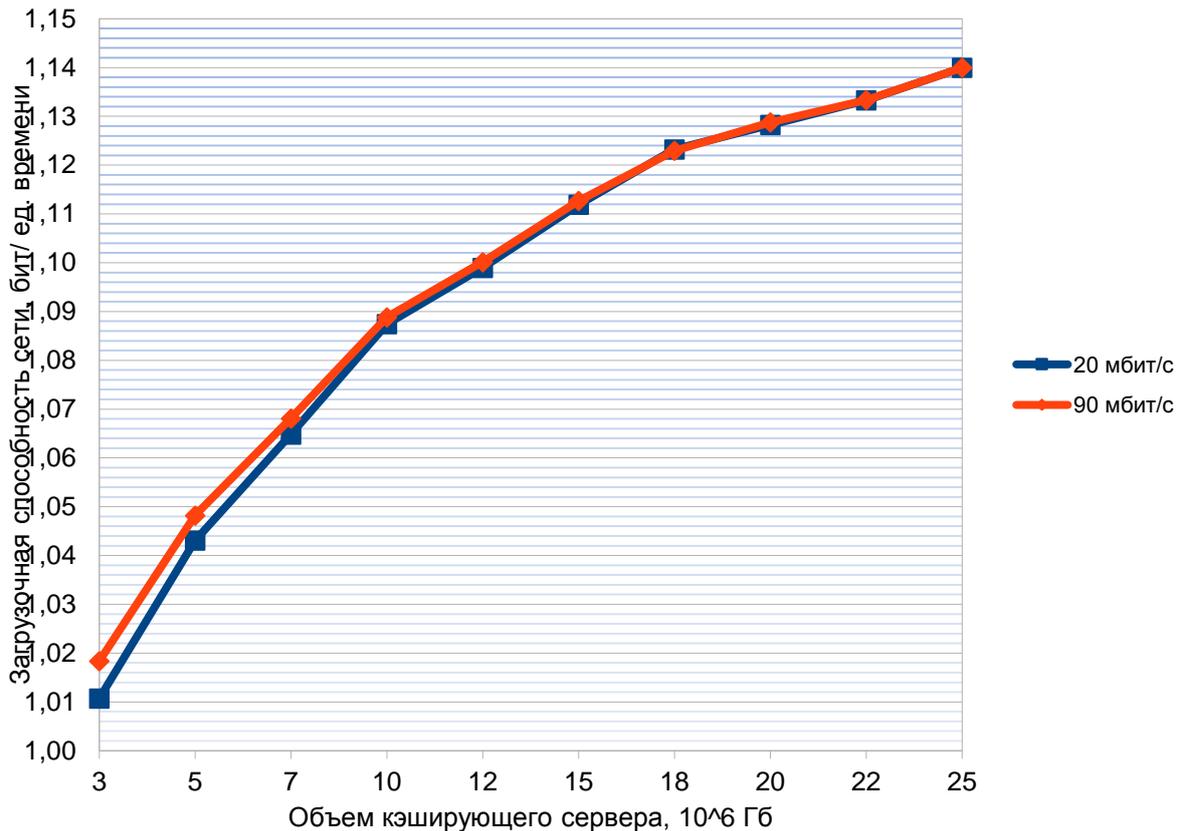


Рис. 3. Графики зависимости пропускной способности сети от объема кэширующих серверов и скорости канала между пользователем и главным сервером

4. Заключение

В рамках данной работы был разработан программный комплекс, который на основе таких входных данных, как ёмкость и количество кэш серверов, количество конечных пользователей, скорость передачи данных от каждого сервера и структура сети, вычисляет пропускную способность сети, используя предложенный Б. Я. Рябко теоретико-информационный подход для оценки эффективности CDN-сетей.

Результаты исследований показывают, что применение предложенного метода на практике позволяет не только оценить загрузочную способность сети аналитически, но и заранее, т.е. на этапе проектирования, сформировать конфигурацию, обеспечивающую наилучшую эффективность.

Благодарность

Работа выполнена при частичной финансовой поддержке РФФИ, грант № 15-07-01851.

Литература

1. *Ryabko B.* An information-theoretic approach to estimate the capacity of processing units // Performance Evaluation. 2012. V. 69. P. 267–273.
2. *Nygren E., Sitaraman R. K., Sun J.* The Akamai Network: A Platform for High-Performance Internet Applications. SIGOPS Oper. Syst. Rev. 2010. 44 (3). P. 2–19.

3. *Shannon C. E.* A Mathematical Theory of Communication // The Bell System Technical Journal. 1948. V. 27. P. 379–423, 623–656.
4. *Knuth D. E.* The art of computer programming, volume 1, Fascicle 1, MMIX: A RISC Computer for the New Millennium, 2005.

*Статья поступила в редакцию 14.11.2015;
переработанный вариант – 01.03.2016.*

Приставка Павел Анатольевич

к.т.н., доцент кафедры прикладной математики и кибернетики СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: ppa@ngs.ru.

Ключникова Ольга Александровна

ассистент кафедры вычислительных систем СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: olgauleaf@gmail.com.

Климова Ирина Вячеславовна

ассистент кафедры прикладной математики и кибернетики СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: irin.klimova@gmail.com.

Analytical method research for assessing the effectiveness of content delivery networks

P. Pristavka, O. Klyuchnikova, I. Klimova

In this paper, problem solving of analytic estimation of data-transmission network efficiency based on the method proposed by B. Ya. Ryabko is considered. Dependence of the channels speed estimation, caching servers volume, and network configuration is demonstrated and analyzed. A conclusion about logical correspondence between computational results and the character of network alterations based on the obtained results is drawn and so adequateness of the analytic estimation method is experimentally demonstrated.

Keywords: data-transmission network efficiency, estimation efficiency.