

УДК 621.391

О.Ю. Евсева, М.Б. Кадер

Харьковский национальный университет радиоэлектроники, Харьков

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ МАРШРУТИЗАЦИИ ЗАПРОСОВ В СЕТЯХ ДОСТАВКИ КОНТЕНТА

Популярность и постоянный рост объемов мультимедийного и других типов контента привели к возникновению и развитию так называемых сетей доставки контента CDN. Одной из важных задач, решаемых в CDN, является маршрутизация запросов пользователей, которая объединяет в себе процедуру выбора контент-сервера с традиционной маршрутизацией трафика от выбранного сервера до пользователя, запросившего услугу. В статье предложена математическая модель, которая позволяет формализовать задачу маршрутизации запросов в CDN как оптимизационную и решить ее совместно с задачей традиционной многопутьевой маршрутизации в транспортной телекоммуникационной сети.

Ключевые слова: *сети доставки контента, маршрутизация запросов, математическая модель.*

Введение

Одной из разновидностей мультисервисных сетей являются сети доставки контента (Content Delivery Network, CDN), основное назначение которых заключается в хранении контента и организации эффективной его доставки интернет-пользователям. При этом сам контент может быть различным: от текстовых файлов и статичных изображений, например, книги, журналы, новостные ленты, фото-галереи, до потокового видео и вещания в реальном масштабе времени, например, фильмы с высоким разрешением или цифровое телевидение [1, 2]. Как правило, подобные сети строятся на основе существующей сетевой инфраструктуры, фактически используя Internet как транспортную основу для передачи любого доступного в сети данного контент-провайдера контента.

Основная идея построения сетей CDN заключается в дублировании (или зеркалировании) контента и/или сервисов основного сервера (origin server) несколькими периферийными серверами, называемыми репликами (replica) или суррогатными (surrogate), которые географически расположены на удалении от основного сервера и друг от друга, но вблизи от потребителя услуг [1, 3]. Тогда запросы, поступающие от интернет-пользователей из различных локальных сетей, направляются на обслуживание на ближайший незагруженный сервер (основной или периферийный), обеспечивая тем самым повышение пропускной способности, масштабируемости, надежности и эффективности всей системы доставки контента в конечном итоге.

В целом функционирование CDN связано с решением следующих основных задач [1 – 3]:

- 1) выбор числа и точек размещения периферийных серверов;
- 2) обновление контента периферийных серверов и его синхронизация с контентом основного сервера;

- 3) маршрутизация запроса (request-routing), которая заключается в выборе ближайшего в определенной метрике незагруженного сервера (основного или периферийного), способного предоставить запрашиваемый контент;

- 4) управление доступом к контенту: аутентификация, авторизация, учет пользователей;

- 5) обеспечение надлежащей информационной безопасности как периферийных, так и основных серверов.

Каждая из перечисленных задач вносит свой вклад в результирующее качество доставки контента, при этом одну из ключевых ролей играет маршрутизация запросов пользователей, поскольку определяет порядок использования доступных сетевых и вычислительных ресурсов CDN. В свою очередь, эффективность использования ресурсов представляет собой ни что иное, как эффективность функционирования всей системы – в данном случае, сети доставки контента.

Анализ известных подходов к решению задачи маршрутизации запросов в CDN

Задача маршрутизации запросов в CDN заключается в перенаправлении запроса пользователя на «ближайший» в рамках некоторой заранее выбранной метрики сервер, основной или периферийный. Как правило, это географически наименее удаленный сервер, в идеале принадлежащий той же локальной сети, что и запрашивающий услугу пользователь, либо это сервер, обеспечивающий в данный момент наименьшее время доставки контента. В более широком смысле маршрутизация запросов нацелена на снижение задержек доставки контента путем балансировки нагрузки на серверы CDN и телекоммуникационную сеть, непосредственно транспортирующую данный контент [1 – 3]. Механизм маршрутизации запросов

заключается в следующем. Пользователь, желая получить определенный контент, например, аудиозапись в формате mp3, обращается к сайту audio.com. При этом согласно URL, указанному пользователем в Web-браузере, DNS-сервер локальной сети пользователя определяет IP-адрес основного сервера, с которым и устанавливается соединение (шаг 1 на рис. 1). Основной сервер выполняет авторизацию и аутентификацию, в результате чего запрос либо принимается к обслуживанию, либо получает отказ. В случае положительного решения основной сервер генерирует html-страницу с указанием доступного пользователю контента (index.html, шаг 2 на рис. 1) и передает запрос на сервер CDN-провайдера для дальнейшего обслуживания (шаг 3). Сервер CDN-провайдера на основании выбранного пользователем контента, текущей сетевой топологии, загруженности сети и периферийных серверов обеспечивает выбор одного из них (шаг 4), который и будет выступать в качестве источника запрашиваемого контента (шаг 5).

Таким образом, маршрутизация запросов путем динамического выбора источника контента и создаваемого при этом трафика реализует распределение сетевой нагрузки и в конечном итоге определяет качество обслуживания пользователей. Как результат задача поиска наиболее эффективного алгоритма маршрутизации запросов в CDN является актуальной и ее решению посвящен ряд научных работ [1 – 3].

В целом известные подходы к решению задачи маршрутизации запросов можно разделить на адаптивные, в рамках которых решение о перенаправлении запроса принимается на основании оценки текущего состояния телекоммуникационной сети и загруженности серверов, и неадаптивные, базирующиеся на некоторых эвристиках и постоянных метриках, таких как

географическая удаленность и производительность серверов, реже пропускная способность сетевых трактов передачи. Основное преимущество неадаптивных методов маршрутизации запросов заключается в простоте их реализации, как, например, в случае циклического (round-robin) или случайного (random) выбора периферийного сервера [1]. Однако, учитывая, что серверы могут находиться на существенном удалении друг от друга и обладать разной производительностью, данные подходы не обеспечивают ожидаемой эффективности в использовании вычислительных ресурсов. Тем не менее, главный их недостаток заключается в отсутствии учета не только загруженности серверов, но и текущей загруженности транспортной телекоммуникационной сети, что может привести к неоправданным отказам или низкому качеству обслуживания пользователей, прежде всего за счет большой задержки доставки контента. Таким образом, с целью обеспечения максимальной производительности сети доставки контента реализуемый в ней алгоритм маршрутизации запросов должен быть адаптивным, т.е. учитывать текущую загруженность сетевых и вычислительных ресурсов и обеспечивать сбалансированное их использование. Подобный алгоритм адаптивной маршрутизации был разработан для сети Akamai [4], но будучи коммерческим продуктом, является полностью закрытым. В этой связи продолжение исследований в области адаптивной маршрутизации запросов в CDN является актуальным.

Как показывает анализ, известные технологические решения преимущественно базируются на эвристических алгоритмах, которые маршрутизируют запрос на сервер, обладающий по результатам активных измерений или пассивного мониторинга наилучшей метрикой.

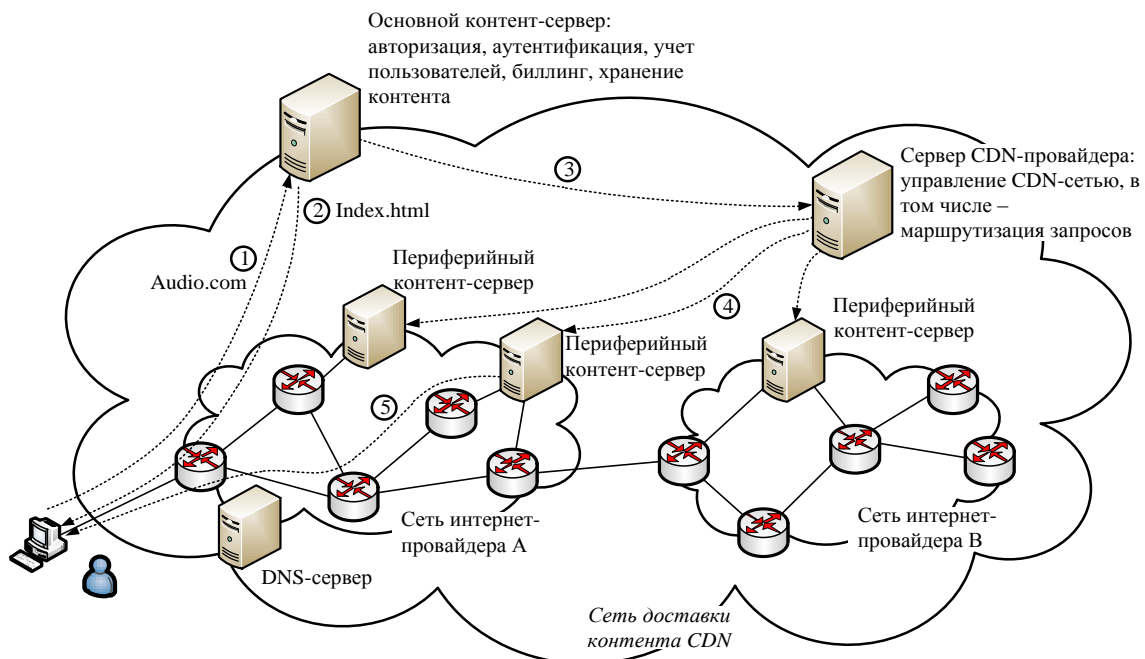


Рис. 1. Процесс маршрутизации запросов в CDN

В качестве метрик, как правило, выступают задержка, включая время передачи в сети и время ответа сервера, потери пакетов, скорость доставки контента, загруженность сервера, оцениваемая через загруженность центрального процессора и интерфейсов ввода-вывода [1]. При этом основной недостаток известных решений заключается в том, что они не имеют достаточного теоретического обоснования. Максимальная эффективность процесса маршрутизации может быть достигнута только лишь при условии, во-первых, наличия адекватной математической модели сети доставки контента, во-вторых, формулировки маршрутизации как оптимизационной задачи с задействованием для ее решения методов оптимального управления.

Математическая модель маршрутизации запросов в CDN

При построении математической модели CDN будем исходить из того, что в процессе передачи трафика участвуют два компонента: транспортная телекоммуникационная сеть и периферийные серверы, выступающие, фактически, в роли источников трафика. В этой связи введем следующие множества: N – множество узлов (маршрутизаторов) транспортной телекоммуникационной сети, G_1 – множество периферийных серверов, на которых доступен контент 1-го типа, $G_1 \in G$, G – множество всех серверов CDN. Для описания транспортной телекоммуникационной сети введем переменные x_{ij}^{kl} , которые указывают на долю трафика, передаваемого по тракту передачи (i, j) от i -го маршрутизатора к j -му, где индекс l отражает тип передаваемого контента, а индекс k связан с парой отправитель-получатель $\{s_k, t_k\}$, $s_k \in G_1$, $t_k \in N$. Выражение $t_k \in N$ означает, что в качестве получателя в рамках предлагаемой математической модели будем рассматривать не непосредственно конечного пользователя, а узел транспортной телекоммуникационной сети, к которой он подключен.

С целью достижения максимальной эффективности использования сетевых ресурсов будем ориентироваться на многопутевой способ доставки контента. В этой связи на введенные переменные накладываются ограничения

$$0 \leq x_{ij}^{kl} \leq 1, k \in K, l \in L, (i, j) \in E, \quad (1)$$

где K – множество пар отправитель-получатель в обслуживаемой сети доставки контента; L – множество возможных типов контента, распространяемых данной сетью; E – множество трактов передачи транспортной телекоммуникационной сети.

Потоковый характер разрабатываемой модели и переменных x_{ij}^{kl} обуславливает необходимость

введения в модель закона сохранения потока в узлах транспортной телекоммуникационной сети [5, 6] с адаптацией его для решения задачи маршрутизации запросов в CDN:

$$\sum_{j \in N} x_{ij}^{kl} - \sum_{j \in N} x_{ji}^{kl} - \sum_{g \in G_1} y_g^{kl} x_{gi}^{kl} = \begin{cases} 0, & \text{если } i \neq s_k, t_k; \\ -1, & \text{если } i = t_k; \end{cases} \quad (2)$$

где y_g^{kl} – переменная использования g -го сервера, $g \in G_1$, в качестве источника контента l -го типа в k -й паре отправитель-получатель

$$y_g^{kl} = \begin{cases} 0, & \text{если } g \neq s_k, \\ 1, & \text{если } g = s_k. \end{cases} \quad (3)$$

Поскольку ни один из узлов транспортной телекоммуникационной сети не может выступать в качестве отправителя контента в рамках CDN, условие (2) предусматривает только два возможных значения для потока: 0 для промежуточных узлов и -1 для узла-получателя. Для отправителя контента $s_k \in G_1$ условие сохранения потока принимает вид

$$\sum_{j \in N} x_{gj}^{kl} = \begin{cases} 0, & \text{если } g \neq s_k; \\ 1, & \text{если } g = s_k; \end{cases} \quad g \in G_1. \quad (4)$$

Ограниченность сетевых и вычислительных ресурсов CDN в рамках введенных обозначений формализуется следующим образом

$$\sum_{k \in K} \sum_{l \in L} r^{kl} x_{ij}^{kl} \leq c_{ij}, \quad (5)$$

$$\sum_{k \in K} y_g^{kl} \leq S_g^1, \quad (6)$$

где r^{kl} – интенсивность трафика, порождаемого при передаче контента l -го типа между k -й парой сервер-пользователь; S_g^1 – максимальное число сессий l -го типа, которое способен обслужить g -й сервер (производительность сервера).

При этом, выражение (5) ограничивает суммарный трафик, передаваемый вдоль тракта передачи (i, j) , а условие (6) – число одновременно поддерживаемых сервером g сессий l -го типа, т.е. сессий, связанных с передачей контента l -го типа. В общем случае производительность сервера является интегральным показателем, который зависит от производительности центрального процессора, быстродействия дисковых накопителей, объемов оперативной памяти, пропускной способности интерфейсов и пр. Если «узким» местом в данной системе является сетевой интерфейс, то ограничение по производительности сервера (6) может быть представлено как ограничение по пропускной способности интерфейса:

$$\sum_{k \in K} \sum_{l \in L} r^{kl} x_{gj}^{kl} \leq c'_{gj}, \quad g \in G_1, \quad (7)$$

где c'_{gj} – пропускная способность интерфейса, обеспечивающего физического соединения g -го сервера

с j -м маршрутизатором транспортной телекоммуникационной сети.

Таким образом, в рамках модели (1) – (6) процесс маршрутизации запросов в CDN связан с поиском переменных y_g^{kl} , а наличие неизвестных переменных x_{ij}^{kl} позволяет решить эту задачу совместно с традиционной задачей многопутевой маршрутизации в транспортной телекоммуникационной сети. При этом в качестве целевой функции может выступать стоимостная функция

$$W = Q_x \bar{x} + Q_y \bar{y} \rightarrow \min, \quad (8)$$

где \bar{x} – вектор, объединяющий в себе переменные x_{ij}^{kl} ; \bar{y} – вектор, объединяющий переменные y_g^{kl} ; Q_x , Q_y – векторы весовых коэффициентов, определяющие стоимость использования сетевых и вычислительных ресурсов соответственно.

Как показано в литературе, наиболее распространенным критерием выбора сервера в рамках маршрутизации запросов является сбалансированное их использование, что объясняется стремлением уменьшить среднюю задержку доставки контента [7]. Такой подход требует представления ограниченных (5) и (6) в следующем виде

$$\sum_{k \in K} \sum_{l \in L} x_{ij}^{kl} \leq \alpha \leq 1, \quad (9)$$

$$\sum_{k \in K} y_g^{kl} / S_g^l \leq \beta \leq 1, \quad (10)$$

где α – управляемый порог использования канальных ресурсов транспортной телекоммуникационной сети, $0 \leq \alpha \leq 1$; β – управляемый порог использования вычислительных ресурсов всех серверов CDN, $0 \leq \beta \leq 1$.

Тогда в качестве целевой функции, обеспечивающей одновременно сбалансированное использование всех ресурсов (и сетевых (α), и вычислительных (β)), выступает

$$W = Q_\alpha \alpha + Q_\beta \beta \rightarrow \min, \quad (11)$$

где Q_α , Q_β – весовые коэффициенты, определяющие важность каждого слагаемого.

В случае (11), как и в случае (8) задача маршрутизации запросов в CDN формализуется как оптимизационная задача смешанного линейного программирования, для решения которой известен ряд методов.

Выводы

Таким образом, математическую модель (1) – (11) следует рассматривать как базовую модель, позволяющую формализовать задачу маршрутизации запросов в CDN как оптимизационную, при этом ее основным преимуществом является воз-

можность совместного (а значит, в высокой степени согласованного) решения задачи маршрутизации запросов и многопутевой маршрутизации в транспортной телекоммуникационной сети.

Приведенные в статье формулировки целевых функций (8) и (11) представляют собой разные точки зрения на протекающие в сети процессы, и выбор одной из них (или формулировка новых) представляет собой отдельную задачу исследований. Так критерий (8) обусловлен стремлением минимизировать затраты на обслуживание запросов, а критерий (11) направлен на улучшение качества обслуживания пользователей через сбалансированное использование сетевых и вычислительных ресурсов CDN. И хотя сбалансированное использование ресурсов за счет равномерной загрузки сетевых элементов и минимизации перегрузок действительно способствует улучшению качества обслуживания в среднем, оно не гарантирует заданных требований к качеству доставки контента, что указывает на направление для дальнейшего развития предложенной модели.

Список литературы

1. *Content Delivery Networks* / Buyya, Rajkumar; Pathan, Mukaddim; Vakali, Athena (Eds.). – Springer, 2008. – 418 p. – (Series: Lecture Notes in Electrical Engineering).
2. Bartolini N. A Walk through Content Delivery Networks / Novella Bartolini, Emiliano Casalicchio, Salvatore Tucci // *Performance Tools and Applications to Networked Systems: Revised Tutorial Lectures*; Edited by Maria Carla Calzarossa, Erol Gelenbe – Springer, 2004. – P. 1 – 25.
3. Kabir M. H. Request-routing Trends and Techniques in Content Distribution Networks / Md Humayun Kabir, Eric G Manning, Gholamali C Shoja, // *International Conference on Computer and Information Technology (ICCIT 2002): proc. of conference.* – Dhaka, Bangladesh, 2002. – P. 315 – 320.
4. Nygren E. The Akamai Network: A Platform for High-Performance Internet Applications/ Nygren E.; Sitaraman R. K., Sun, J. // *ACM SIGOPS Operating Systems Review.* – 2010. – Vol. 44, Issue 3. – P. 2 – 19.
5. Seok Yo. Dynamic constrained multipath routing for MPLS networks / Seok Yo., Lee Yo., Choi Ya // *IEEE International Conference on Computer Communications and Networks, 2001: proceedings of the conference.* – Vol. 2., Issue 1. – P. 348 – 353.
6. Wang Y. Explicit routing algorithms for Internet Traffic Engineering / Wang Y., Wang Z. // *8th International Conference on Computer Communications and Networks: proceedings of the conference.* – Paris, 1999. – P. 582 – 588.
7. Лемешко А.В. Усовершенствование потоковой модели многопутевой маршрутизации на основе балансировки нагрузки [Электронный ресурс] / А.В. Лемешко, Т.В. Вавенко // *Проблемы телекоммуникации.* – 2012. – № 1 (6). – С. 12 – 29. – Режим доступа до статьи: http://pt.journal.kh.ua/2012/1/1/121_lemeshko_multipath.pdf.

Поступила в редколлегию 30.10.2012

Рецензент: д-р техн. наук, проф. А.В. Лемешко, Харьковский национальный университет радиоэлектроники, Харьков.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ МАРШРУТИЗАЦИИ ЗАПРОСОВ В СЕТЯХ ДОСТАВКИ КОНТЕНТА

О.Ю. Євсєєва, М.Б. Кадер

Популярність і постійне зростання обсягів мультимедійного та іншого типу контенту призвели до появи та розвитку так званих мереж доставки контенту CDN. Однією з важливих задач, що розв'язуються в CDN, є маршрутизація запитів користувачів, яка об'єднує в собі вибір контент-сервера з традиційною маршрутизацією трафіка від вибраного сервера до користувача, який сформував запит. У статті запропоновано математичну модель, яка дозволяє формалізувати задачу маршрутизації запитів в CDN як оптимізаційну і розв'язати її спільно із задачею традиційної багатошляхової маршрутизації в транспортній телекомунікаційній мережі.

Ключові слова: мережі доставки контенту, маршрутизація запитів, математична модель.

MATHEMATICAL MODEL OF REQUEST-ROUTING IN CONTENT DELIVERY NETWORKS

O. Yu. Yevsyeyeva, M.B. Khader

According increasing popularity and volume of multimedia and other contents, content delivery networks (CDN) have been proposed and deployed. One of the important problems that are solved in the CDN is request routing which combines the content server choosing and traditional routing from the chosen server to the user. The paper proposes a mathematical model which allows to formulate the problem of request routing in CDN as optimization problem and to resolve it in concordance with the traditional task of multipath routing in transport telecommunications network.

Keywords: content delivery network, request-routing, mathematical model.