

УДК 621.391

МОДЕЛЬ И МЕТОД ИЕРАРХИЧЕСКОГО УПРАВЛЕНИЯ ЗАПРОСАМИ В ГИБРИДНОЙ СЕТИ ДОСТАВКИ КОНТЕНТА



[О.Ю. ЕВСЕЕВА](#), [М.Б. КАДЕР](#)

Харьковский национальный
университет радиоэлектроники

Abstract – A key challenge for infocommunication networks including Internet is delivering large-scale content of different types to growing number of users. Content Delivery Networks (CDN) and Peer-to-Peer (P2P) networks are two dominant architectures that can be implemented. But CDN-P2P-Hybrid architecture (HCDN), that is integrating P2P and CDN features, promises to combine their advantages and to achieve maximal effectiveness of content distribution with guaranteed quality of service and relatively low cost. At same time HCDN needs appropriate management practices for solving different problems where request routing is one of. In order to maximize the effectiveness of content delivering to the user, to provide guaranteed quality of services and to use of resources in effective way method for request routing in HCDN should be based on a rigorous mathematical justification. This article offers two flow-based mathematical models which allow servicing the request by using different number of CDN-servers for content sourcing and turn the request routing task into optimization problem. By downloading from several CDN-servers use of computing and network resources becomes balanced that contribute to the result quality of service improving. Mathematically load-balancing is achieved by choice of the optimality criterion. A distinctive feature of the proposed approach is related to taking into account state of the underlying transport infrastructure. It allows to solve the problem of request routing on HCDN and traffic routing problem in transport telecommunication network jointly. The models contain tensor QoS-restrictions satisfaction of which guarantees on the quality of delivery at the network level within transport network. According to offered models hierarchical two layers control method for request routing on HCDN is proposed.

Анотація – Перехід до гібридних мереж доставки контенту (HCDN) дозволяє поєднати низьку вартість послуг з високою їх якістю, однак вимагає відповідних методів управління, у тому числі і управління запитами. У роботі пропонуються два варіанти потокової математичної моделі управління запитами, що використовують для їх обслуговування різну кількість CDN-серверів. Особливістю запропонованого підходу полягає в урахуванні стану транспортної інфраструктури при розв'язанні задач управління запитами. За рахунок введення в модель QoS-обмежень вдається забезпечити гарантії за якістю доставки на мережному рівні, а за рахунок вибору критерію оптимальності всі типи ресурсів (обчислювальні ресурси серверів і каналні ресурси трактів передачі транспортної мережі) використовуються збалансовано.

Аннотация – Переход к гибридным сетям доставки контента (HCDN) позволяет сочетать низкую стоимость услуг с высоким их качеством, однако требует соответствующих методов управления, в том числе и управления запросами. В работе предлагаются два варианта потоковой математической модели управления запросами, предполагающие использование для их обслуживания разного количества CDN-серверов. Особенность предлагаемого подхода состоит в учете состояния нижележащей транспортной инфраструктуры при решении задач управления запросами. За счет введения в модель QoS-ограничений удастся обеспечить гарантии по качеству доставки на сетевом уровне, а за счет выбора критерия оптимальности все типы ресурсов (вычислительные ресурсы серверов и каналные ресурсы трактов передачи транспортной сети) используются сбалансированно.

Введение

Своевременная и эффективная доставка контента любого вида является одной из основных задач, возлагаемых на инфокоммуникационные сети, в том числе Internet. Растущие объемы данных, подлежащие хранению и передаче, наряду с ростом числа пользователей, их запрашивающих, обуславливают перманентную актуальность задачи повышения эффективности создаваемых специально с этой целью сетей доставки контента (Content Delivery Network, CDN) [1]. Организационно качество их функционирования может быть повышено за счет интеграции различных принципов хранения и доставки контента в рамках так называемых гибридных CDN (Hybrid CND, HCDN) [2, 3].

I. Принципы построения гибридных сетей доставки контента

Создание гибридных CDN обусловлено стремлением объединить преимущества и компенсировать недостатки двух принципиально разных подходов, представленных традиционными сетями доставки контента CDN, например, Akamai, Limelight, и пиринговыми (одноранговыми) сетями (Peer-to-Peer, P2P), функционирующими, например, на основе протоколов BitTorrent и eDonkey. Сравнительная характеристика принципов построения и вытекающих из них свойств сетей CDN и P2P представлена в табл. 1 [2, 3].

Хотя сети CDN весьма эффективны в плане организации доступа и доставки контента, основная причина, препятствующая их доминированию в данном сегменте рынка, связана с высокой стоимостью услуг и сложностью управления. Функционирование CDN связано с решением ряда задач, среди которых репликация контента на периферийных серверах и их синхронизация, авторизация и маршрутизация запросов пользователей на тот или иной сервер, что ведет к усложнению системы. С другой стороны, растущая популярность P2P-сетей обусловлена их высокой масштабируемостью и низкой себестоимостью услуг по доставке контента, поскольку каждый хост (клиент) одновременно является как потребителем услуги, так и ее поставщиком. Тем не менее, основным недостатком P2P-сетей связан с отсутствием гарантий относительно качества предоставляемой услуги. В этой связи P2P-сети не могут рассматриваться как альтернатива сетей CDN, а, напротив, с целью повышения качества обслуживания пользователей, построения масштабируемой и при этом контролируемой и управляемой системы хранения и доставки контента P2P-сети должны дополнять серверную архитектуру CDN, что и происходит в рамках гибридных сетей доставки контента.

В гибридной сети пользователь может загрузить контент с сервера CDN, с выполняющих роль серверов хостов P2P или использовать оба типа источника. С точки зрения организации логических связей в рамках гибридной сети доставки контента, как показывает анализ, наиболее перспективной видится модель взаимодействия, в которой ресурсы одной CDN доступны для нескольких P2P-сетей, а интеграция технологий CDN и P2P осуществляется по принципу CAP (CDN-aided P2P), где основная роль отводится P2P-сети. P2P-сеть дополняется CDN-серверами, которые будут задействованы только в том случае, если запрашиваемый контент не может быть предоставлен P2P-сетью [2]. Причем в качестве источника запроса к CDN будет выступать сервер P2P-сети.

Гибридизация и усложнение архитектуры CDN неразрывно влечет за собой усложнение задач управления нею, к которым среди прочего относятся задачи маршрутизации запросов. В ходе их решения следует исходить из следующих факторов. HCDN имеет четко выраженную иерархическую архитектуру, а потому управление запросами должно реализовывать принципы иерархического управления.

Таблица 1. Сравнительная характеристика различных сетей доставки контента

Критерии сравнения	Традиционные сети доставки контента CDN	Пиринговые (одноранговые) сети P2P
Принцип построения	Четкое разделение функций клиента и сервера: контент-серверы обеспечивают хранение и доставку контента, узел-клиент может лишь запрашивать тот или иной контент	Каждый узел сети наделен функциями и клиента, и сервера: обладая тем или иным контентом, узел может предоставлять сервис по его доставке, в то же время по отношению к другому контенту этот же узел может выступать в качестве клиента
Масштабируемость	Масштабируемость достигается за счет использования совокупности периферийных серверов, называемыми репликами (replica), которые полностью или частично зеркалируют содержимое основного сервера, но географически расположены вблизи потребителя услуги	Высокая масштабируемость достигается за счет объединения в рамках единой сети пиров, территориально удаленных друг от друга
Пропускная способность системы доставки контента в целом. Стоимость услуг.	Пропускная способность фиксирована и определяется числом периферийных серверов. Стоимость услуг высокая	С ростом числа пиров пропускная способность растет, а удельная стоимость услуги падает
Возможность предоставления услуг гарантированного качества	Высокая надежность, структура сети постоянна. Возможно обслуживание с гарантированными показателями качества	Количество пиров в сети является случайной неконтролируемой величиной, потому предоставление услуг с гарантированным качеством затруднено
Соблюдение авторских прав. Безопасность	Контент полностью контролируется. Соблюдение авторских прав гарантируется. Функции безопасности реализованы централизованно	Присутствует возможность нарушения авторских прав
Особенности трафика, возникающего в процессе доставки контента	Так называемый ISP-дружественный (ISP-friendly) трафик: трафик в процессе доставки контента передается от контент-сервера к клиенту, его запрашивающего, по кратчайшему маршруту, с минимальным участием сетей других провайдеров (Internet Service Provider, ISP)	Поскольку пиры, обладающие запрашиваемым контентом (особенно, если речь идет о наиболее популярных файлах), могут быть подключены к сетям разных провайдеров, в процессе доставки контента соответствующий трафик будет передан через сети нескольких ISP, создавая для них дополнительную нагрузку (ISP-unfriendly traffic)

С целью предоставления услуг гарантированного качества (Quality of Service, QoS) управление запросами в HCDN, включая выбор сервера или совокупности серверов и непосредственно доставку контента, должно осуществляться исходя из типа услуги и запрашиваемого качества ее предоставления. Кроме того, по своей сути сети

доставки контента являются оверлейными сетями, в которых система хранения контента, выполняющая задачи выделения емкостей запоминающих устройств серверов, организации доступа, кэширования, биллинга и пр., накладывается на телекоммуникационную сеть (ТКС), обеспечивающую транспортные функции по доставке контента. Поскольку каждый из компонентов вносит существенный вклад в результирующее качество предоставления услуги, в процессе маршрутизации запросов необходимо учитывать состояние каждого из них. Другими словами, в процессе принятия управляющего решения необходимо учитывать состояние нижележащей инфраструктуры, т.е. управление вычислительными ресурсами серверов CDN и P2P должно быть скоординировано с распределением канальных и буферных ресурсов телекоммуникационной сети. В дополнение, с целью эффективного использования всех типов ресурсов распределение нагрузки между CDN и P2P-сетями, равно как использование ресурсов внутри каждой из них, должно быть сбалансированным.

Как показывает анализ, методы маршрутизации запросов, реализованные на практике и предлагаемые в литературе, зачастую не отвечают перечисленным требованиям в полной мере [1 – 3]. Традиционно задача маршрутизации запросов сводится к выбору сервера, на который будет перенаправлен запрос для дальнейшего обслуживания, и в такой постановке относится к задачам управления сервисами, которые решаются отдельно от задач маршрутизации потоков трафика, обеспечивающих доставку контента на сетевом уровне (в транспортной телекоммуникационной сети). Кроме того, при решении задачи маршрутизации запросов преобладают эвристические подходы, где решение сводится к выбору «лучшего» по результатам пассивного или активного мониторинга контент-сервера, на который и направляется запрос.

Таким образом, актуальной является задача разработки математических моделей и методов маршрутизации запросов в гибридной сети доставки контента, учитывающие состояние телекоммуникационного компонента, нацеленные на доставку контента с гарантированным качеством и обеспечивающие сбалансированное использование ресурсов различных типов.

II. Математическая модель управления запросами в гибридной сети доставки контента с гарантированным качеством обслуживания

С целью описания структурных свойств HCDN введем граф $G = \{N, E\}$, множество вершин которого N моделирует серверы P2P-сетей (N^{P2P}), серверы CDN (N^{CDN}) и маршрутизаторы транспортной телекоммуникационной сети (N^{TTN}), т.е. $N = N^{P2P} \cup N^{TTN} \cup N^{CDN}$. Множество ветвей E графа G отображает множество трактов передачи как между маршрутизаторами ТКС, так и между маршрутизаторами и серверами различных типов (рис. 1). Поскольку в соответствии с архитектурой построения HCDN в качестве источников контента и соответствующих потоков трафика выступают периферийные серверы CDN, а в качестве источников запросов и получателей контента – серверы P2P-сетей, исток s_k^l и сток t_k^l данной сети ограничены

множествами N^{CDN} и N^{P2P} , $s_k^l \in N^{CDN}$, $t_k^l \in N^{P2P}$, где индекс k указывает на номер полюсной пары (индекс запроса) в рамках множества K , $k \in K$, l – на запрашиваемый контент, $l \in L$.

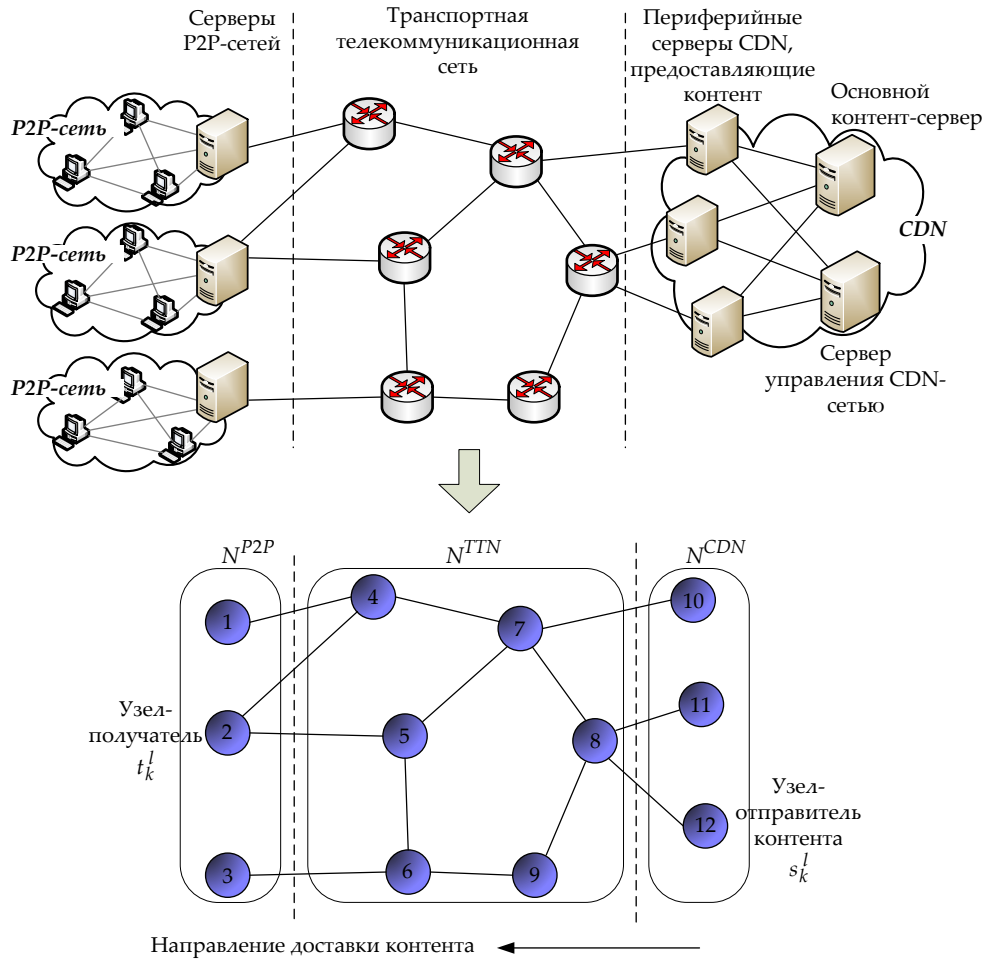


Рис. 1. Структурная модель гибридной сети доставки контента

В процессе маршрутизации запросов необходимо решить задачу выбора источника контента и сформировать маршруты для его доставки. В этой связи введем переменные управления двух типов: переменные выбора сервера y_g^{kl} , отражающие использование g -го сервера, $g \in N^{CDN}$, в качестве источника контента l -го типа в k -й паре полюсов, и маршрутные переменные x_{ij}^{kl} , указывающие на долю потока, передаваемого по тракту передачи $(i, j) \in E$ от i -го маршрутизатора (сервера) к j -му в процессе доставки данного контента.

Маршрутные переменные x_{ij}^{kl} определяют порядок маршрутизации потоков в транспортной телекоммуникационной сети, и с целью реализации более эффективного многопутевого способа потребуем, чтобы

$$0 \leq x_{ij}^{kl} \leq 1. \quad (1)$$

По определению переменные выбора сервера являются булевыми:

$$y_g^{kl} = \begin{cases} 1, & \text{если } g\text{-й сервер используется в качестве источника} \\ & l\text{-го контента в } k\text{-й паре,} \\ 0, & \text{в противном случае.} \end{cases} \quad (2)$$

Поскольку модель ориентирована на гибридные сети доставки контента, предполагается, что запросы поступают к серверам CDN от серверов нижележащих P2P-подсетей и фактически представляют собой не единичные, а агрегированные запросы. В этом случае целесообразно обращаться к нескольким CDN-серверам одновременно, тогда переменные y_g^{kl} будут указывать на долю контента, загружаемого с g -го CDN-сервера, т.е.

$$0 \leq y_g^{kl} \leq 1, \quad (3)$$

а с целью обеспечения целостности запрашиваемого контента необходимо выполнить условия

$$\sum_{g \in N^{CDN}} y_g^{kl} = 1. \quad (4)$$

В рамках введенных переменных закон сохранения потока для маршрутизаторов транспортной телекоммуникационной сети и серверов P2P-сетей, выступающих в роли получателей контента, примет вид

$$\sum_{j \in N^{P2P} \cup N^{TTN}} x_{ij}^{kl} - \sum_{j \in N^{TTN}} x_{ji}^{kl} - \sum_{g \in N^{CDN}} x_{gi}^{kl} = \begin{cases} 0, & \text{если } i \neq s_k^l, t_k^l, \\ -1, & \text{если } i = t_k^l, \end{cases} \quad i \in N^{P2P} \cup N^{TTN}. \quad (5)$$

В выражении (5) третье слагаемое указывает на долю потока, поступающего от g -го CDN-сервера. Тогда условие (5) может быть дополнено выражением

$$\sum_{j \in N^{P2P} \cup N^{TTN}} x_{gj}^{kl} = y_g^{kl}, \quad g \in N^{CDN}. \quad (6)$$

Ввиду ограниченности пропускных способностей трактов передачи и вычислительных возможностей серверов CDN в модель вводятся условия вида:

$$\sum_{k \in K} \sum_{l \in L} r^{kl} x_{ij}^{kl} \leq c_{ij}, \quad (7)$$

$$\sum_{k \in K} y_g^{kl} \leq S_g^l, \quad (8)$$

где r^{kl} – интенсивность потока, создаваемого при передаче контента l -го типа в рамках k -й пары; c_{ij} – пропускная способность тракта передачи $(i, j) \in E$; S_g^l – максимальное число сессий l -го типа, которое способен обслужить g -й CDN-сервер (производительность сервера).

Одно из преимуществ сетей CDN заключается в высоком качестве предоставляемых услуг, а потому важным требованием к модели управления запросами в HCDN является обеспечение доставки контента с гарантированными показателями качества. Результирующее качество, воспринимаемое конечным пользователем, является интегральной характеристикой, зависящей от множества факторов, где определяющую роль играют QoS-показатели сетевого уровня в рамках транспортной телекоммуникационной сети. С целью обеспечения гарантированного качества доставки контента на сетевом уровне модель управления запросами должна быть дополнена соответствующими условиями. При этом целесообразно воспользоваться результатами тензорного моделирования ТКС, однако аналитический вид условий обеспечения QoS зависит от принятой модели обслуживания запросов [4, 5]. Так, в случае задействия для обслуживания запроса одного источника контента (модель (1), (2), (5) – (8)) условие обеспечения качества обслуживания для каждого потока, создаваемого при доставке контента l -го типа в рамках k -й полюсной пары, имеет вид [4]:

$$\lambda^{(mp\bar{b})} \leq \left(G_{\pi\eta}^{(4,1)} - G_{\pi\eta}^{(4,2)} \left[G_{\pi\eta}^{(4,4)} \right]^{-1} G_{\pi\eta}^{(4,3)} \right) \tau_{(don)}, \quad (9)$$

а в случае задействия множества источников в рамках модели (1), (3) – (8) условие гарантированного качества обслуживания потока формулируется как [5]

$$\Lambda_{\eta}^{(g)} \leq \left(G_{\pi\eta}^{(g;4,1)} - G_{\pi\eta}^{(g;4,2)} \left[G_{\pi\eta}^{(g;4,4)} \right]^{-1} G_{\pi\eta}^{(g;4,3)} \right) \Gamma_{\eta}^{(g)}, \quad (10)$$

где $\lambda^{(mp\bar{b})}$ – требуемая скорость передачи, с которой должен быть передан поток пакетов, переносящих l -й контент в рамках k -й полюсной пары; $\tau_{(don)}$ – допустимое в рамках транспортной ТКС значение межконцевой средней задержки; $\Lambda_{\eta}^{(g)}$ – вектор размера g , элементы которого указывают на интенсивности потоков трафика, поступающих к пользователю от различных источников (серверов); g – число задействованных в передаче контента серверов; $\Gamma_{\eta}^{(g)}$ – вектор размера g , элементы которого

содержат значения $\tau_{(don)}$; $\left\| \begin{array}{c|c} G_{\pi\eta}^{(g;4,1)} & G_{\pi\eta}^{(g;4,2)} \\ \hline \text{---} & \text{---} \end{array} \right\| + \left\| \begin{array}{c|c} G_{\pi\eta}^{(g;4,3)} & G_{\pi\eta}^{(g;4,4)} \\ \hline \text{---} & \text{---} \end{array} \right\| = \left\| \begin{array}{c|c} G_{\pi\eta}^{(4,1)} & G_{\pi\eta}^{(4,2)} \\ \hline G_{\pi\eta}^{(4,3)} & G_{\pi\eta}^{(4,4)} \end{array} \right\| = G_{\pi\eta}^{(4)}$,

$\left\| \begin{array}{c|c} G_{\pi\eta}^{(1)} & G_{\pi\eta}^{(2)} \\ \hline \text{---} & \text{---} \end{array} \right\| + \left\| \begin{array}{c|c} G_{\pi\eta}^{(3)} & G_{\pi\eta}^{(4)} \\ \hline \text{---} & \text{---} \end{array} \right\| = G_{\pi\eta}$, причем $G_{\pi\eta}^{(4,1)}$ – первый элемент матрицы $G_{\pi\eta}^{(4)}$, $G_{\pi\eta}^{(g;4,1)}$ – квад-

ратная матрица размера $g \times g$, $G_{\pi\eta}^{(4)}$ – квадратная подматрицы размера $\phi \times \phi$, $\phi = m - 1$, m – общее число узлов в сети доставки контента, т.е. мощность множества N ; $G_{\pi\eta}$ – матрица размера $n \times n$, являющаяся проекцией метрического тензора сети доставки контента в базисе узловых пар и контуров; n – число трактов передачи в сети доставки контента, т.е. мощность множества E .

На рис. 2 приведен пример сети, где для загрузки контента задействованы три CDN-сервера, образующие с узлом, сформировавшим запрос (узел-получатель на рис. 2), три узловые пары ($\mathcal{G}=3$) из семи возможных ($\phi=7$) [5]. Каждая z -я узловая пара характеризуется двумя QoS-показателями: скоростью $\lambda_{(\eta)}^z$, с которой в ее рамках осуществляется доставка контента, и претерпеваемой при этом задержкой $\tau_z^{(\eta)}$.

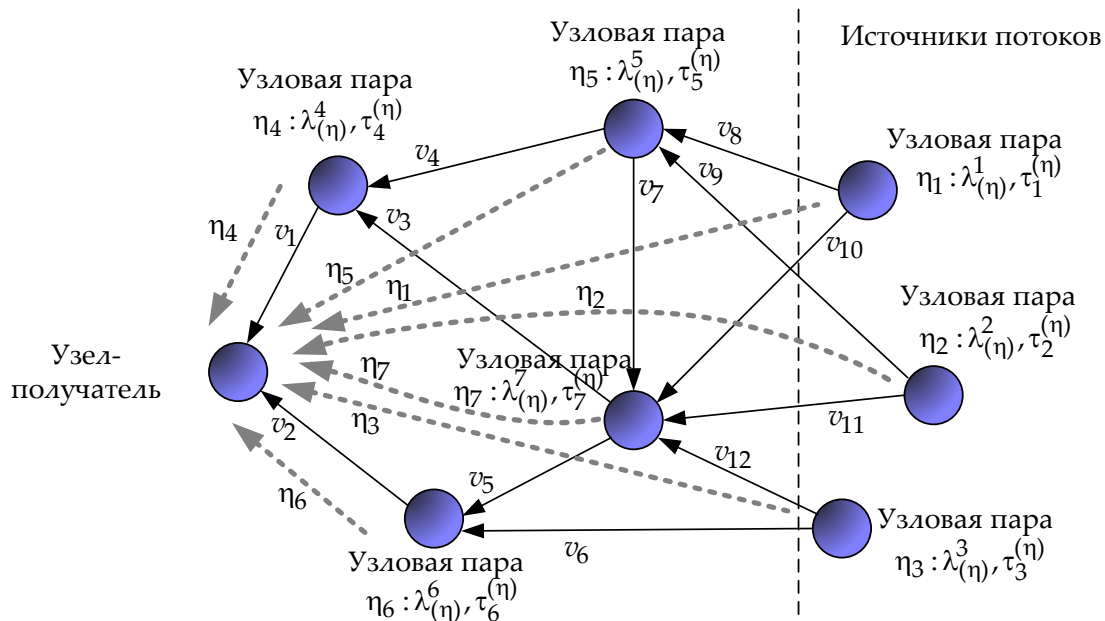


Рис. 2. Пример сети доставки контента и ее базисные узловые пары

Поскольку каждый запрос имеет свои требования относительно $\lambda_{(\eta)}^z$ и $\tau_z^{(\eta)}$, условие (9) (или (10)) должно быть записано для каждого из них в отдельности. Для упрощения дальнейшего изложения опустим индексы l и k , указывающие на тип контента и запрашивающего его пользователя. Тогда элементы вектора $\Lambda_{\eta}^{(\mathcal{G})}$ связаны с ранее введенными переменными следующим образом:

$$\lambda_{(\eta)}^z = y_g r, \quad (11)$$

где все переменные относятся к одной и той же паре l и k ; r – скорость загрузки контента на сетевом уровне; z – номер узловой пары, создаваемой стоком k -й полюсной пары и g -м CDN-сервером, $z = \overline{1, \mathcal{G}}$.

Причем в соответствии с исходными данными суммарная интенсивность потока r , поступающего от множества источников, должна отвечать скоростным требованиям к качеству обслуживания, т.е.

$$\sum_{z=1}^{\mathcal{G}} \lambda_{(\eta)}^z = r \geq \lambda^{(mp\bar{o})}. \quad (12)$$

В выражениях (9) и (10) матрица проекции метрического тензора вычисляется следующим образом

$$G_{\pi\eta} = A^t G_v A, \quad (13)$$

где A – матрица ковариантного координатного преобразования, определяемая на основании структуры сети, $G_v = \|g_v^{ii}\|$ – диагональная матрица размера $n \times n$, элементы главной диагонали которой относятся к трактам передачи сети и отражают количественную связь между интенсивностью потока в них с задержкой их передачи.

Особенностью тензорного моделирования является то, что на правила формирования элементов g_v^{ii} не накладывается дополнительных ограничений. С этой целью могут быть использованы различные подходы и математические аппараты. Например, при условии описания каждого сетевого интерфейса системой массового обслуживания $M/M/1/R$ g_v^{ii} рассчитываются согласно выражениям [4]:

$$g_{(v)}^{ii} = \frac{(1 - (\rho_{(v)}^i)^{R_i + 1})(1 - \rho_{(v)}^i)(\lambda_{(v)}^i)^2}{\rho_{(v)}^i - (\rho_{(v)}^i)^{R_i + 2} - (N_i + 1)(\rho_{(v)}^i)^{R_i + 1}(1 - \rho_{(v)}^i)}, \quad (14)$$

где R_i – емкость буфера на интерфейсе i -го тракта передачи; $\rho_{(v)}^i = \frac{\lambda_{(v)}^i}{c_i}$ – коэффициент загрузки i -го тракта передачи; $\lambda_{(v)}^i$ и c_i – интенсивность потока пакетов и пакетная пропускная способность i -го тракта передачи соответственно, причем $\lambda_{(v)}^i = r x_{ab}^i$, при условии, что i -й тракт передачи телекоммуникационной сети соответствует тракту (a, b) в двухиндексной нумерации.

Таким образом, в рамках предлагаемой модели (1), (2), (5) – (9), (13), (14), ориентированной на использование одного источника контента, и модели (1), (3) – (8), (10) – (14), допускающей множество источников, процесс управления запросом связан с поиском переменных двух типов: переменных y_g^{kl} , за счет которых реализуется управление ресурсами серверов CDN, а также переменных x_{ij}^{kl} , обеспечивающих решение задачи многопутевой маршрутизации в транспортной телекоммуникационной сети. При этом в соответствии с требованием о сбалансированном использовании всех типов ресурсов сети доставки контента в качестве целевой функции целесообразно использовать [6]:

$$W = Q_\alpha \alpha + Q_\beta \beta \rightarrow \min, \quad (15)$$

$$\sum_{k \in K} \sum_{l \in L} x_{ij}^{kl} \leq \alpha \leq 1, \quad \frac{\sum_{k \in K} y_g^{kl}}{S_g^l} \leq \beta \leq 1, \quad (16)$$

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

III. Метод иерархического управления запросами гибридной сети доставки контента

На основании изложенных моделей в соответствии с иерархическим принципом построением гибридной сети доставки контента может быть предложен следующий метод управления запросами в ней. В рамках метода предполагается два уровня, где нижний связан с управлением в рамках P2P-подсетей, а верхний уровень отвечает за взаимодействие разных подсистем. Фактически верхний уровень является координатором между P2P-подсетями, контент-ресурсами CDN и нижележащей транспортной телекоммуникационной сетью.

В рамках нижнего уровня, т.е. в пределах P2P-сети все функции мониторинга, контроля и принятия решений возлагаются на единый центр управления, который кроме этого выполняет традиционные функции индекс-сервера (трекера). На основании оценки текущего состояния сети (в первую очередь числа активных пиров) и анализа поступающих запросов сервер управления P2P-сетью принимает решение о возможности обслужить новый запрос без обращения к внешним серверам CDN. На рис. 3 схематически показаны этапы поступления запроса от пользователя (шаг 1) и его обработка сервером P2P-сети (шаг 2). Если загрузка контента с требуемым качеством невозможна, например, по причине недостаточного числа активных пиров с запрашиваемым контентом или низкого качества соединений между ними, P2P-сервер формирует запрос к CDN-серверу, выполняющему роль центра управления верхнего уровня (шаг 3 на рис.3). Заметим, что в рамках CDN функции авторизации, аутентификации, хранения учетных записей, биллинга, хранения контента и управления запросами могут быть реализованы как на разных серверах, так и на виртуальных серверах одного физического сервера.

После выполнения процедур авторизации и аутентификации в случае положительного решения основной сервер передает запрос на сервер CDN-провайдера для дальнейшего обслуживания (шаг 4). Сервер CDN-провайдера принимает решение о маршрутизации запроса на один или несколько зеркалирующих серверов (шаги 5 и 6). Это решение формируется в ходе решения оптимизационной задачи (1), (2), (5) – (9), (13), (14) или (1), (3) – (8), (10) – (14) с учетом текущей загруженности как серверов CDN, так и трактов передачи телекоммуникационной сети, а также запрашиваемого качества обслуживания. При этом речь идет о маршрутизации агрегированных запросов, поступающих не от одиночных пользователей, а от серверов управления P2P-сетью, параллельно выполняющих роль агрегации запросов. Загрузка от нескольких CDN-серверов (шаг 7) позволяет сбалансированно использовать их ресурсы, повысить результирующую скорость загрузки контента, сбалансированно использовать ресурсы телекоммуникационной сети.

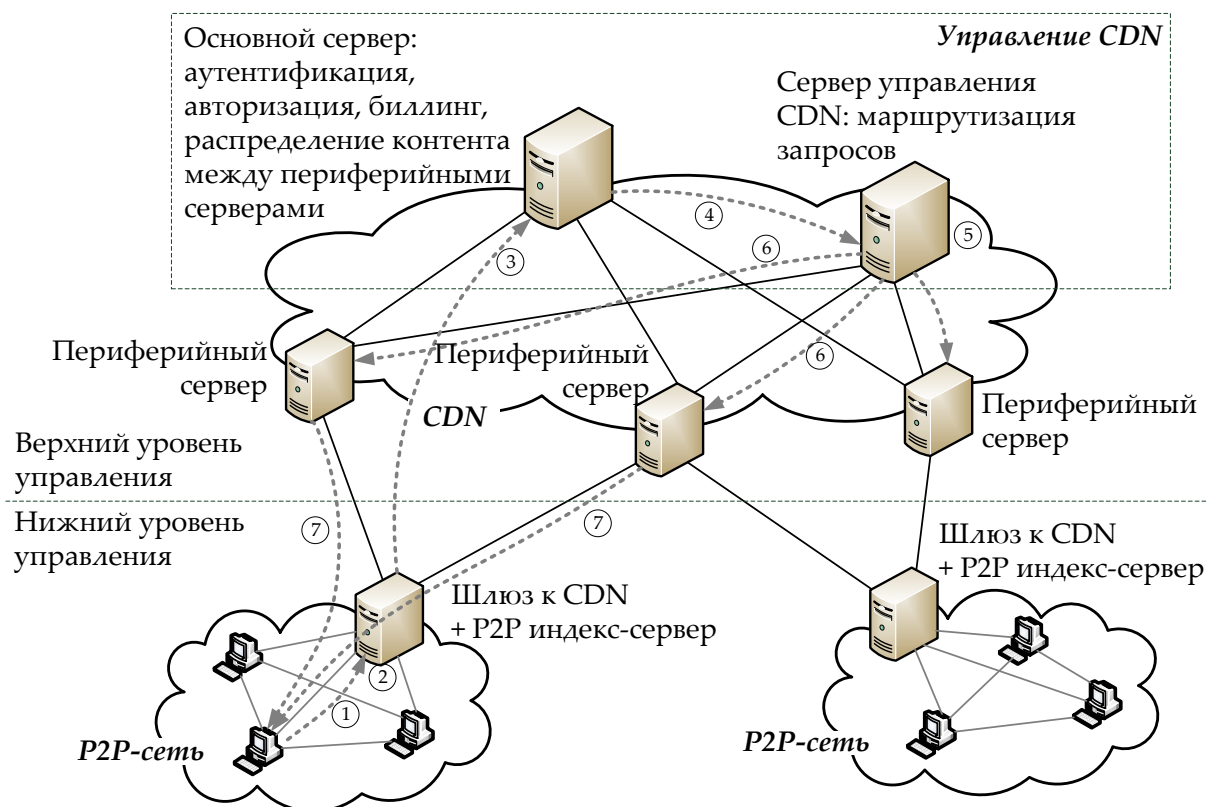


Рис.3. Метод иерархического управления запросами гибридной сети доставки контента

Выводы

Переход к гибридным сетям доставки контента позволяет сочетать низкую стоимость услуг с высоким их качеством, однако требует соответствующих методов управления, в том числе и управления запросами. С целью достижения максимальной эффективности доставки контента пользователю, гарантированного качества предоставления таких услуг и при этом рационального использования ресурсов решение задачи управления запросами HCDN должно иметь под собой строгое математическое обоснование. В данной статье предлагаются два варианта потоковой математической модели, предполагающие для обслуживания запроса использование разного количества CDN-серверов в качестве источников контента. Отличительная особенность предлагаемого подхода состоит в учете при решении задач управления запросами, относящихся к уровню управления сервисами, состояния нижележащей транспортной инфраструктуры. За счет введения в модель QoS-ограничений удастся обеспечить гарантии по качеству доставки на сетевом уровне, а за счет выбора критерия оптимальности все типы ресурсов HCDN (вычислительные ресурсы серверов и каналные ресурсы трактов передачи транспортной сети) используются сбалансированно.

Список литературы:

1. *Content Delivery Networks* / Buyya, Rajkumar; Pathan, Mukaddim; Vakali, Athena (Eds.). – Springer, 2008. – 418 p. (Series: Lecture Notes in Electrical Engineering).
2. *Zhi Hui Lu, Ye Wang, Yang Richard Yang. An Analysis and Comparison of CDN-P2P-hybrid Content Delivery System and Model* // Journal of Communications – 2012. – Vol. 7, N. 3. – P. 232 – 245.
3. *Efficient large-scale content distribution with combination of CDN and P2P networks* / [Hai Jiang, Jun Li, Zhongcheng Li, Xiangyu Bai] // International Journal of Hybrid Infomational technology. – 2009. – Vol. 2, N. 2. – P. 13 – 24.
4. *Лемешко А.В., Евсеева О.Ю. Тензорная модель многопутевой маршрутизации с гарантиями качества обслуживания одновременно по множеству разнородных показателей [Электронный ресурс] // Проблемы телекоммуникаций. – 2012. – № 4 (9). – С. 16 - 31. – Режим доступа: http://pt.journal.kh.ua/2012/4/1/124_lemeshko_tensor.pdf.*
5. *Евсеева О.Ю. Тензорная модель многополюсной телекоммуникационной сети // Радиотехника: Всеукр. межвед. науч.-техн. сб. – 2013. – Вып. 175. – С. 154 – 159.*
6. *Лемешко А.В., Вавенко Т.В. Усовершенствование потоковой модели многопутевой маршрутизации на основе балансировки нагрузки [Электронный ресурс] // Проблемы телекоммуникаций. – 2012. – № 1 (6). – С. 12 – 29. – Режим доступа: http://pt.journal.kh.ua/2012/1/1/121_lemeshko_multipath.pdf.*