УДК 004.7 DOI 10.25559/SITITO.2017.3.618

Ромасевич П.В.[1, 2]

[1] Компания D-Link, г. Волгоград, Россия
[2] Волгоградский государственный университет, г. Волгоград, Россия

ОЦЕНКА СРЕДНИХ ЗАДЕРЖЕК ТЕЛЕКОММУНИКАЦИОННЫХ ИНФРАСТРУКТУР И КАНАЛОВ МЕЖСОЕДИНЕНИЯ ЦЕНТРОВ ОБРАБОТКИ ДАННЫХ В АРХИТЕКТУРЕ WEB SCALE

Аннотация

В данной работе рассмотрены типовые архитектуры Центров обработки данных (ЦОД) и тенденции их объединения в рамках концепции Web Scale в связи развитием технологий виртуализации, «облачных» услуг и требованиями все более быстрой доставки возрастающих объемов контента с надлежащим качетвом обслуживания (QoS). Предложены формулы оценок средней сетевой задержки канала межсоединения Центров обработки данных в архитектуре Web Scale и средней задержки телекоммуникационной инфраструктуры ЦОД произвольной архитектуры с учетом эффектов самоподобия трафика и потери пакетов. Дана характеристика специального программного обеспечения Data Center Bridging (DCB) и аппаратного функционала активного сетевого оборудования, необходимого для работы в специфических условиях эксплуатации Центров обработки данных и представлены соответствующие модели коммутаторов D-Link.

Ключевые слова

ЦОД, межсоединение, сетевая задержка, Web Scale, виртуализация, коэффициент использования, самоподобие трафика, время задержки, архитектура, leaf-spine, 10G, 40G, QoS, контент, «облачные» услуги, DCB.

Romasevich P.V.[1, 2]

[1] D-Link, Volgograd, Russia
[2] Volgograd State University, Volgograd, Russia

AVERAGE DELAYS ASSESSMENT OF TELECOMMUNICATION INFRASTRUCTURES AND INTERCONNECTING CHANNELS OF DATA PROCESSING CENTERS IN WEB SCALE ARCHITECTURE

Abstract

In this article are considered standard architecture of the Data Centers (DC) and a tendency of their combining within the concept of Web Scale in communication by development of technologies of virtualization, "cloudy" services and requirements of more and more fast delivery of the increasing content volumes with appropriate quality of service (QoS). The proposed formulas estimates the average cross-network delay of channel interconnects for Data Centers in architecture of Web Scale and average delay of DC telecommunication infrastructure of arbitrary architecture taking into account effects of self-similarity of a traffic and loss of packets. The characteristic of the special software of Data Center Bridging (DCB) and hardware functionality of the active network equipment necessary for operation in specific operating conditions of Data Centers is also this and the appropriate models D-Link switches are provided.

Keywords

Data processing center, DPC, interconnecting, cross-network delay, Web Scale, virtualization, utilization coefficient, self-similarity of a traffic, delay period, architecture, leaf-spine, 10G, 40G, QoS, content, "cloudy" services, DCB.

Введение

Распространение облачных услуг и виртуализации привело к появлению множества Центров Обработки Данных (ЦОД), результатом чего стал эффект «масштаба Web» (Web Scale) (Рис. 1). Это стало стимулом для поставщиков интернет-контента и облачных услуг предложить своим клиентам разнообразные приложения и предоставить доступ огромному множеству пользователей [1].

Фундаментом этих преобразований является сеть. Доступность ЦОД в мире Web Scale напрямую зависит от масштабируемости сети и эффективной бизнес-модели предоставления услуг.

Сегодня ЦОД подвергаются радикальной трансформации, нацеленной на обеспечение эффективной работы в условиях постоянно меняющихся объемов и природы трафика, растущей потребности в пропускной способности в связи с лавинообразным появлением новых приложений. Получение максимально быстрого доступа к интернет-контенту становится для пользователей принципиальным для возможности его использования [1].

Ввиду массового распространения мобильных устройств и появления в связи с этим разнообразных современных услуг и приложений пользовательский контент должен доставляться максимально быстро с соблюдением соответствующего для приложения качества обслуживания (QoS), что стимулирует строительство дополнительных ЦОД вне крупных городов с тенденцией их объединения.

Уже сегодня корпоративные ИТ-подразделения все активнее переходят от собственных заказных ИТ-приложений к стандартным «облачным» решениям, контент которых хранится в ЦОД, географическое положение которого клиент может и не знать. А с учётом роста количества центров обработки данных и объемов запрашиваемой информации, а также всё большего распространения виртуализации серверов и вычислительных устройств, межсоединение ЦОД – DCI (Data Center Interconnet) становится одним из ключевых факторов обеспечения успешной реализации приложений верхнего уровня.

Виртуализация затронула все компоненты ЦОД и современные технологии теперь позволяют использовать два или несколько ЦОД как один — с разделением нагрузки и задач с целью сведения к минимуму эксплуатационных расходов и достижения максимальной производительности [1].

Поэтому сетевой обмен данными является фундаментом для основных функций ЦОД независимо от его типа и решает следующие задачи:

Организация соединений между серверами для компьютеров и вычислительных устройств с высоким быстродействием.

Доступ к устройствам хранения

Соединения между различными ЦОД, а также между ЦОД и внешним миром.

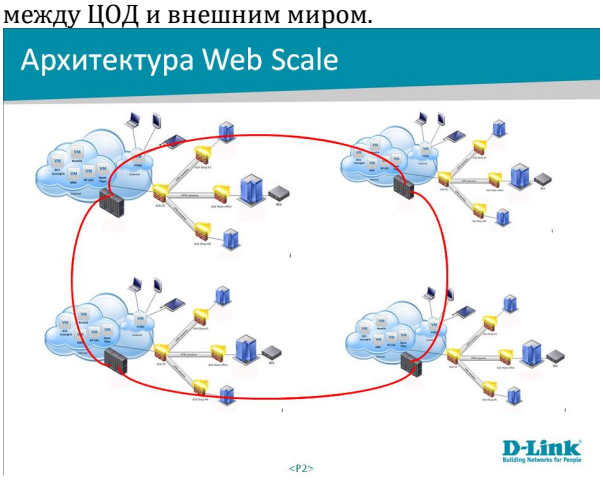

Рис. 1

В этой связи превентивная оценка средних задержек, вносимых каналами межсоединений и самими коммутационными инфраструктурами ЦОД при обмене трафиком между ними в результате миграции виртуальных машин, уже на этапе проектирования или модернизации позволяет определить необходимые параметры сетевой инфраструктуры для обеспечения необходимого качества обслуживания (QoS) соответствующего «специализации» конкретного центра обработки данных.

Оценка средней задержки канала межсоединения ЦОД

Учитывая постоянно растущую долю мультимедийного трафика, обладающего выраженной самоподобной природой, в оценках принципиально важно учитывать данный эффект, отрицательно влияющий на сетевую задержку и повышающий требования к аппаратным возможностям активного сетевого оборудования [4].

При этом в качестве исходных данных целесообразно оперировать получаемыми из систем мониторинга сети средними за исследуемый период параметрами телекоммуникационных систем, которые в случае межсоединения ЦОДов являются одноканальными.

С точки зрения многолетнего опыта автора по созданию и модернизации

телекоммуникационных сетей, на практике важна именно оценка сетевых задержек в конкретных условиях эксплуатации.

Поэтому в данном случае можно применить формулу средней задержки, полученную в [2], учитывающую пропускную способность канала C и коэффициент его использования ρ, а также входящую скорость λ в пакетах/с и параметр Херста H, отражающий степень самоподобия трафика:

$$T = \frac{\rho}{\lambda(1-\rho)}\left[\sqrt{\frac{\rho}{C}}\cdot\frac{1}{1-\rho}\right]^{\frac{2H-1}{1-H}}. \quad (1)$$

Из формулы видно, что положительно влияют на задержку в сторону её уменьшения увеличение пропускной способности канала и скорости трафика в нём. Напротив, увеличение коэффициента использования канала и параметра Херста способствуют её увеличению, что в определенных случаях может оказаться неприемлемым для приложений и потребует масштабирования параметров телекоммуникационной системы.

Уже сейчас на рынке доступны для использования сетевые коммутаторы с максимальной скоростью интерфейса 40G, сделав еще недавно фантастические скорости 10G обычным техническим решением.

Поскольку подключение к «внешнему миру» делается на граничном устройстве, порты которого, как правило, также задействованы и под «внутренние» подключения в ЦОД, имеет практический смысл ввести понятие среднего коэффициента одновременной загрузки портов, что при наличии интерфейсов 10G описывается формулой, полученной в [3] с условием, что пакеты не должны теряться из-за переполнения буферной памяти интерфейсов коммутатора:

$$\rho < \frac{8(64V+W)}{1000(K_{1G}+10K_{10G})} \quad (2)$$

где:

V – скорость передачи узлом 64-байтного пакета (если не оговорен иной размер пакета), 10^6 пакетов в секунду (Mpps);

K_{10G} – общее число портов с пропускной способностью до 10 Гбит/с;

K_{1G} – общее число портов с пропускной способностью до 1 Гбит/с;

W – общий объем памяти ввода/вывода, который в современных архитектурах коммутаторов не делится жестко между всеми портами коммутатора, а может быть динамически распределен между реально работающими.

Продолжая преобразования, подставляя (2) в (1) можно получить формулу оценки средней задержки канала межсоединения ЦОД в зависимости от документированных характеристик коммутатора, степени самоподобия трафика, а также параметров телекоммуникационной системы – максимальной пропускной способности подведенного канала и среднего значения пакетной скорости трафика в нём.

Полученное выражение может быть легко обобщено для случая уже имеющихся на рынке интерфейсов 40G и использовано для инженерной оценки среднего значения сетевой задержки на основании документированных характеристик применяемого активного оборудования в телекоммуникационной системе межсоединения ЦОД на этапе проектирования или модернизации.

Оценка средней задержки телекоммуникационной инфраструктуры ЦОД

Общая сетевая задержка зависит также и от задержки самой телекоммуникационной инфраструктуры центра обработки данных.

Доминирующая в последние десятилетия иерархическая трехуровневая сетевая архитектура, вполне подходящая для распределенных инфраструктур и представляющая собой те или иные комбинации древовидной и кольцевой топологий, в ЦОД сменилась на двухуровневую инфраструктуру сетевой фабрики, традиционно применяющейся в системах хранения данных (СХД), на основе топологии «каждый с каждым» для обеспечения максимально быстрого сетевого обмена за счет минимизации «длины пути» между узлами.

Архитектура ЦОД эволюционирует вместе с развитием технологий виртуализации и облачных моделей. Для перехода к «облакам» каналы передачи должны обеспечивать максимально быстрый обмен данными между серверами и как раз с этой целью была разработана двухуровневая архитектура leaf-spine (Рис.2), благодаря которой приложения, размещенные на любом вычислительном устройстве, и все хранилища данных могут работать и масштабироваться согласованно, вне зависимости от того, где они физически расположены в инфраструктуре ЦОД [4].

При этом должно быть обеспечено соответствующее качество обслуживания, что делает актуальным превентивную оценку одного из основных критериев QoS – средней сетевой задержки телекоммуникационной инфраструктуры ЦОДа для понимания степени адекватности её архитектуры задачам выполняемых приложений.

Одним из вариантов реализации данной

концепции является сетевой фабрики по топологии leaf-spine с установкой коммутаторов в каждой стойке (ToR). Объем медной кабельной инфраструктуры минимален: серверы подключаются к коммутаторам соответствующими шнурами или короткими твинкоаксиальными кабелями непосредственно в стойках с ИТ-оборудованием через интерфейсы 10G (Рис.3). При этом основная часть кабельной инфраструктуры является оптической [5].

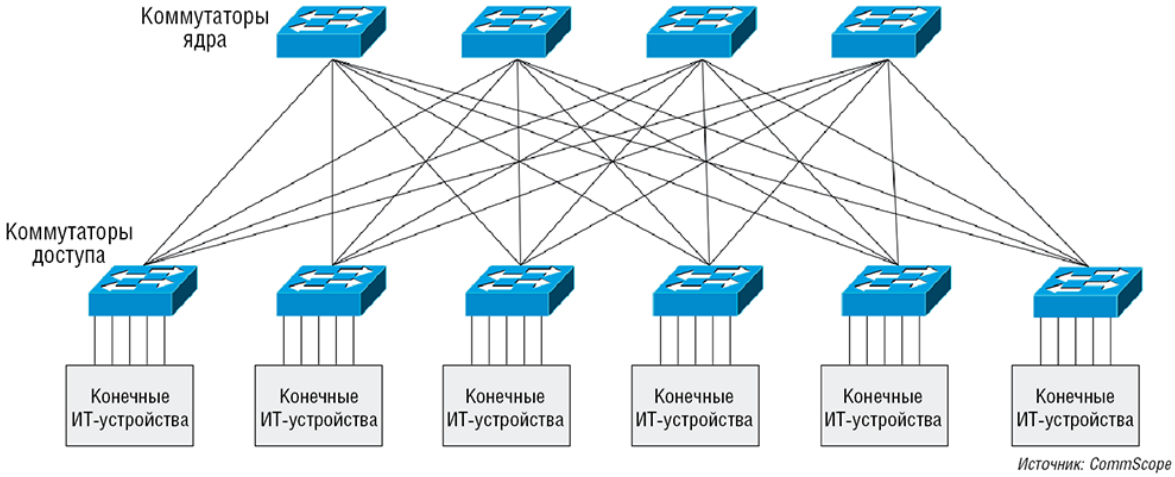

Рис. 2

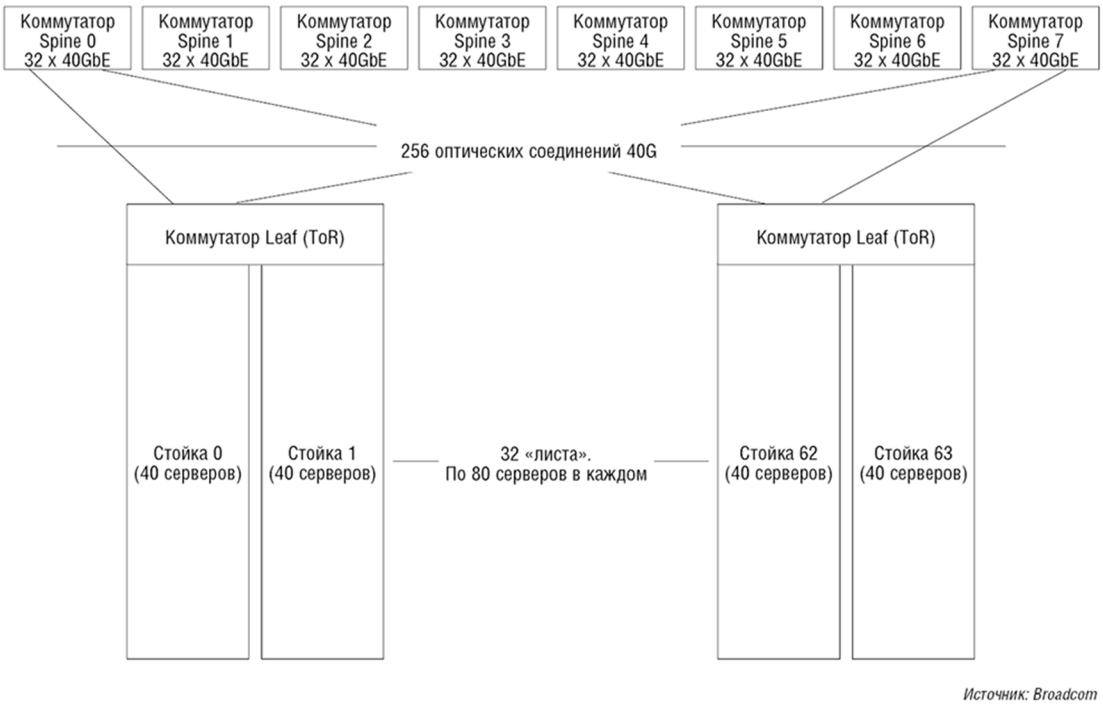

Рис. 3

Другой подход состоит в концентрации коммутаторов в середине (MoR) или в конце (EoR) ряда стоек и подключении к ним серверов каналами 10G (Рис. 4). Это позволяет существенно снизить стоимость активного оборудования, за счёт использования коммутаторов с медными портами 40GBase-T ввиду сокращения длин каналов 40G, реализуемые посредством медных кабелей прямого подключения (DAC), и возможности размещения коммутаторов доступа и ядра рядом, а в ряде случаев даже в одной стойке. И несмотря на рост расходов на ту часть СКС, которая используется для подключения серверов к коммутаторам доступа, по оценкам Broadcom, стоимость данного варианта оказывается примерно в три раза ниже [5].

В [6] была предложена формула оценки

средней задержки для сети произвольной архитектуры с учетом самоподобия трафика

$$T = \frac{\overline{n}}{\lambda}\sum_i \left(\frac{af^{-1}(\varepsilon)^{2H}\rho}{C^{2H-1}(1-\rho)^{2H}} \right)_i^{\frac{1}{2(1-H)}}, \qquad (3)$$

где средняя задержка телекоммуникационной сети зависит от вариации трафика a, скорости канала C и обратной функции от распределения вероятности потери пакетов $f^{-1}(\varepsilon)$, средней длины пути в телекоммуникационной сети $\overline{n}$, параметра Херста H и полной средней скорости трафика внутри сети λ.

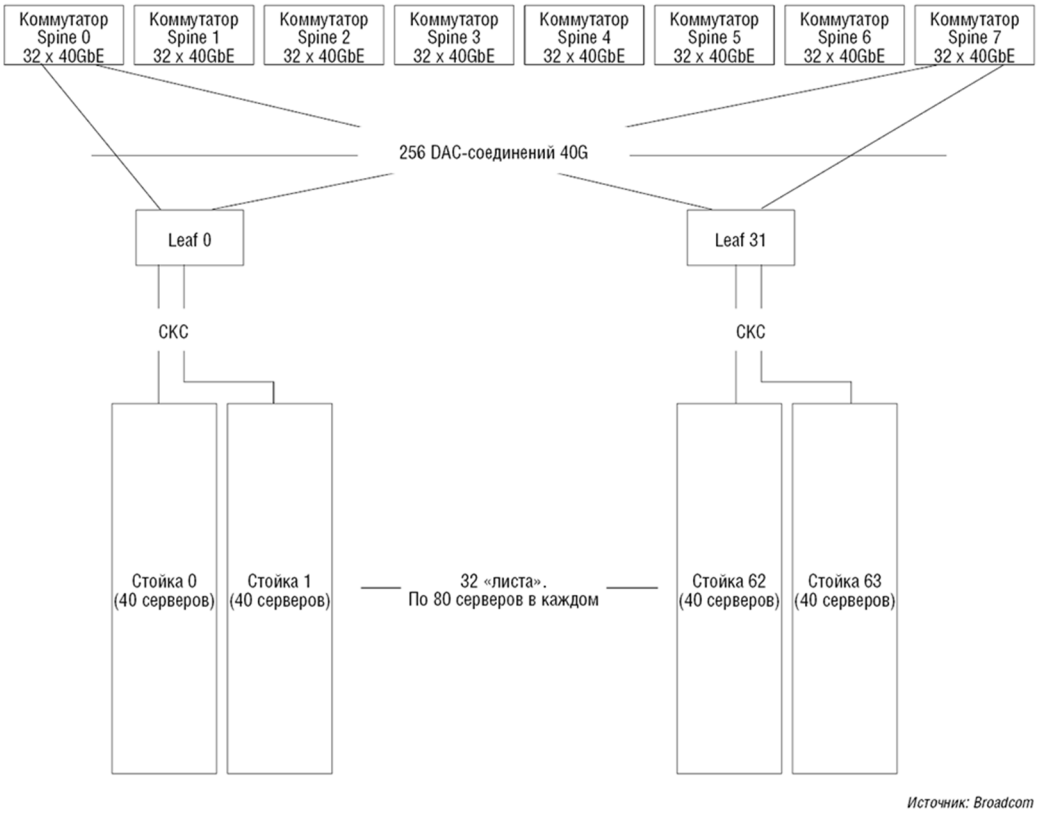

Рис. 4

Средняя длина пути телекоммуникационной системы произвольной архитектуры с внешними каналами может быть получена по формуле теоремы Клейнрока $\overline{n} = \frac{\lambda}{\gamma}$ [7], где γ – средняя величина общего внешнего трафика телекоммуникационной сети, которую легко определить, т.к. к ЦОДу подходит ограниченное количество внешних каналов. Таким образом в формуле (1) мы избавляемся от трудоемкого расчета λ и $\overline{n}$.

Если кабельная система ЦОД выполнена по соответствующим стандартам, то вклад кабельной системы в потери пакетов ничтожен и основной причиной роста задержки будет являться переполнение буферной памяти интерфейсов за счет превышения суммы среднего значения *m* и вариации трафика *a* над C при отсутствии шейпинга, а также эффект самоподобия трафика, который увеличивает требования к размеру буферной памяти при прочих равных условиях.

Для этого случая формула (3) модифицирована в [8] и имеет вид

$$T = \frac{1}{\gamma}\sum_i \left[\frac{a\rho}{C^{2H-1}} \cdot \left(\frac{-\ln\left(\frac{C^{2H-1}}{2\rho a}\left(\frac{1-\rho}{H}\right)^{2H}\left(\frac{x}{1-H}\right)^{2(1-H)}\right)}{1-\rho} \right)^{2H} \right]^{\frac{1}{2(1-H)}} \qquad (4)$$

Итак, общая задержка при транзите трафика через ЦОД в архитектуре Web Scale будет выражаться суммой задержек, выражаемых формулами (1) с учетом (2) и (4).

Заключение

Полученный результат может быть использован специалистами системных интеграторов для предварительной оценки необходимых параметров телекоммуникационной инфраструктуры ЦОД в архитектуре Web Scale на этапах проектирования и модернизации.

Необходимо отметить, что специфика работы коммутаторов в ЦОД отличается от иных условий эксплуатации и предполагает поддержку определенного функционала – Data Center Bridging (DCB), который является обязательной установкой специальных расширений технологии Ethernet для сетевой работы в дата-центрах. Главными DCB являются IEEE 802.1Qbb, IEEE 802.1Qaz и IEEE 802.1Qau. IEEE 802.1Qbb отвечает за контроль потока на основе приоритетов, отвечающий контроль потока для нивелирования потерь данных во время сетевой перегрузки. IEEE 802.1Qaz обеспечивает выбор расширенной передачи, задача которого состоит в управлении распределением ширины полосы пропускания среди различных классов трафика. IEEE 802.1Qau инициирует уведомление о перегрузке, обеспечивающий управление перегрузкой для потоков данных внутри сетевых доменов в целях предотвращения перегрузки.

При этом высокая плотность размещения большого количества оборудования в аппаратных стойках делает актуальными задачи экономии электроэнергии и поддержание соответствующего температурного баланса. В коммутаторах D-Link это достигается наличием автоматического выбора направления вентиляции от задней панели к передней или наоборот, что обеспечивает максимальное кондиционирование воздуха для более эффективного охлаждения всех систем, монтируемых в стойках ЦОД. Коммутаторы также оснащены встроенными интеллектуальными вентиляторами, внутренними термодатчиками, контролирующими изменение температуры и реагирующими соответственно на использование различной скорости вентиляторов при разных температурах. При низких температурах скорость вентиляторов снижается, что сокращает потребление энергии и снижает уровень шума.

Компания D-Link предлагает линейку коммутаторов, обладающих этими качествами и специально предназначенных для работы в ЦОД – DGS-3400-24SC [9], DGS-3400-24TC [10], DGS-3600-16S [11] и DGS-3600-32S [12].

В частности, стек коммутаторов, построенных на модели DGS-3400-24TC, уже около года успешно эксплуатируется в качестве ядра корпоративной сети в главном телекоммуникационном узле Волгоградского государственного университета, где имеет честь также работать автор.

Литература

1. Open System Publications [электронный ресурс] // URL: https://www.osp.ru/lan/2016/04/13049078/ (дата обращения 24.08.2017)
2. П.В.Ромасевич, Оценка некоторых параметров одноканальной телекоммуникационной системы при самоподобном трафике, Сборник избранных трудов V Международной научно-практической конференции «Современные информационные технологии и ИТ-образование», Доиздание, Москва, 2011, с.87-91.
3. П.В. Ромасевич, Метод оценки возможности применения коммутаторов на уровнях иерархии пакетных телекоммуникационных сетей. – Тезисы IV Региональной научно-практической конференции «Проблемы передачи информации в телекоммуникационных системах», 2012, Волгоград, с. 51-54.
4. Open System Publications [электронный ресурс] // URL: https://www.osp.ru/lan/2016/10/13050614/ (дата обращения 24.08.2017)
5. Open System Publications [электронный ресурс] // URL: https://www.osp.ru/lan/2017/01-02/13051371/ (дата обращения 24.08.2017)
6. П.В.Ромасевич Оценка влияния параметров телекоммуникационной системы на среднее время задержки в условиях самоподобного трафика // Инфокоммуникационные технологии. – 2005. – №3. – С. 21-26.
7. Л.Клейнрок, Коммуникационные сети, М., Наука, 1970, 255 с.
8. Ромасевич П.В., Оценка необходимой канальной емкости телекоммуникационной ситсемы с ограниченной буферной памятью в условиях самоподобного трафика – Сборник избранных трудов IX Международной научно-практической конференции «Современные информационные технологии и ИТ-образование», Москва, 2014, с.456-461, ISBN 978-5-9556-0165-6.
9. D-Link [электронный ресурс] // URL: http://www.dlink.ru/ru/products/1/2117.html (дата обращения 24.08.2017)
10. D-Link [электронный ресурс] // URL: http://www.dlink.ru/ru/products/1/2116.html (дата обращения 24.08.2017)
11. D-Link [электронный ресурс] // URL: http://www.dlink.ru/ru/products/1/1698.html (дата обращения 24.08.2017)
12. D-Link [электронный ресурс] // URL: http://www.dlink.ru/ru/products/1/1529.html (дата обращения 24.08.2017)

References

1. Open System Publications [электронный ресурс] // URL: https://www.osp.ru/lan/2016/04/13049078/ (дата обращения 24.08.2017)
2. P.V. Romasevich, Ocenka nekotoryh parametrov odnokanal'noj telekommunikacionnoj sistemy pri samopodobnom trafike, Sbornik izbrannyh trudov V Mezhdunarodnoj nauchno-prakticheskoj konferencii «Sovremennye informacionnye tehnologii i IT-obrazovanie», Doizdanie, Moskva, 2011, s.87-91.
3. P.V. Romasevich, Metod ocenki vozmozhnosti primenenija kommutatorov na urovnjah ierarhii paketnyh telekommunikacionnyh setej. – Tezisy IV Regional'noj nauchno-prakticheskoj konferencii «Problemy peredachi informacii v telekommunikacionnyh sistemah», 2012, Volgograd, s. 51-54.

4. Open System Publications [электронный ресурс] // URL: https://www.osp.ru/lan/2016/10/13050614/ (дата обращения 24.08.2017)
5. Open System Publications [электронный ресурс] // URL: https://www.osp.ru/lan/2017/01-02/13051371/ (дата обращения 24.08.2017)
6. P.V. Romasevich Ocenka vlijanija parametrov telekommunikacionnoj sistemy na srednee vremja zaderzhki v uslovijah samopodobnogo trafika // Infokommunikacionnye tehnologii. – 2005. – №3. – S. 21-26.
7. L.Klejnrok, Kommunikacionnye seti, M., Nauka, 1970, 255 s.
8. Romasevich P.V., Ocenka neobhodimoj kanal'noj emkosti telekommunikacionnoj sitsemy s ogranichennoj bufernoj pamjat'ju v uslovijah samopodobnogo trafika – Sbornik izbrannyh trudov IX Mezhdunarodnoj nauchno-prakticheskoj konferencii «Sovremennye informacionnye tehnologii i IT-obrazovanie», Moskva, 2014, s.456-461, ISBN 978-5-9556-0165-6.
9. D-Link [электронный ресурс] // URL: http://www.dlink.ru/ru/products/1/2117.html (дата обращения 24.08.2017)
10. D-Link [электронный ресурс] // URL: http://www.dlink.ru/ru/products/1/2116.html (дата обращения 24.08.2017)
11. D-Link [электронный ресурс] // URL: http://www.dlink.ru/ru/products/1/1698.html (дата обращения 24.08.2017)
12. D-Link [электронный ресурс] // URL: http://www.dlink.ru/ru/products/1/1529.html (дата обращения 24.08.2017)

Поступила: 10.09.2017

Об авторе:

Ромасевич Павел Владимирович, кандидат технических наук, региональный менеджер компании D-Link по Волгоградской, Астраханской областям и республике Калмыкия; доцент кафедры «Телекоммуникационных систем», Волгоградский государственный университет, promasevich@dlink.ru

Note on the author:

Romasevich Pavel V., Candidate of Technical Sciences, Head of the D-Link company on the Volgograd, Astrakhan regions and the Republic of Kalmykia; Associate professor of «Telecommunication systems», Volgograd State University, promasevich@dlink.ru