№9
июль 2018

MSK IX

Интернет изнутри

Интернет в цифрах
Глобальный рост IP-трафика дата-центров и облачных хранилищ
с.16

Сегментная маршрутизация
Отбрасываем шелуху и находим золотой самородок в предложении IETF
с.18

Сервисы глобальной Сети через призму BGP
В современной глобальной Сети всё большее значение приобретает устойчивость к атакам
с. 36

Календарь событий
Лучшие события 2018 года
с. 48

Виртуализация

Виртуализация центров обработки данных

Рецепт успешного развития современных центров обработки и хранения данных (ЦОДов), похоже, ясен – виртуализация

с. 30

Содержание:

Передовица
с. 4

Оптимизация инфраструктуры центра данных
Концепция виртуализации высокой плотности

Интернет в цифрах
с. 16

IP-трафик центров данных
Глобальный рост IP-трафика

Технология в деталях
с. 18

Сегментная маршрутизация
отбрасываем шелуху и находим золотой самородок в предложении IETF

Технология в деталях
с. 24

А вы готовы к 5G-слайсингу?
Один из интереснейших аспектов мобильной архитектуры 5G

Стандарты Интернета
с. 30

Виртуализация центров обработки данных
Технологии виртуализации стремительно развиваются

Исследования
с. 36

Сервисы глобальной Сети через призму BGP
География сервисов и трансграничные переходы

Новости науки и техники
с. 44

Новости интернет-отрасли
Новости доменной индустрии

Календарь событий
с. 48

2018 год
Журнал «Интернет изнутри» рекомендует

Журнал «Интернет изнутри»
По всем вопросам пишите на info@internetinside.ru

Порядковый номер выпуска и дата его выхода в свет:
Выпуск №9, дата выхода: июль 2018 г.

Свидетельство о регистрации СМИ в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций.
Регистрационный номер:
ПИ № ФС77-71202 от 27.09.2017

Публикуется при поддержке АНО «ЦВКС «МСК-IX»

Главный редактор:
Андрей Робачевский

Зам. главного редактора:
Новикова Татьяна

Редакционная коллегия:
Воронина Елена
Платонов Алексей

Дизайн:
Чернега Наталья

Корректор:
Рябова Наталья

Технологии виртуализации

главный редактор,
Андрей Робачевский

Дорогой читатель!

Виртуализация - популярный термин. Виртуальное пространство все глубже проникает в нашу жизнь - виртуальная реальность и искусственный интеллект начинают трансформировать нашу материальную жизнь. А что говорить об интернете вещей! Сенсоры и актюаторы создают вполне материальные связи с виртуальным пространством, постепенно стирая границы между двумя мирами.

Хотя тема эта захватывающая, в сегодняшнем номере мы предлагаем вашему вниманию другой, хотя и не менее важный тип виртуализации - виртуализацию технологическую, позволяющую отделить услуги, приложения и саму инфраструктуру от физических компонентов. Эти технологии позволяют строить чрезвычайно гибкие и масштабируемые сетевые системы без привязки к конкретному оборудованию или сетевой технологии, используемой в опорной инфраструктуре. В свою очередь, эти гибкие системы являются технологической платформой облачных элементов, на которых создаются приложения виртуальной реальности, искусственного интеллекта, больших данных и интернета вещей.

Центры хранения и обработки данных являются одним из объектов интенсивного применения технологий виртуализации. Вопросы масштабирования, оптимизации загрузки, поддержки определенных параметров качества и обслуживания множества арендаторов решаются с помощью этих технологий. По существу, для каждого из клиентов создается собственный виртуальный дата-центр! О практических шагах по оптимизации инфраструктуры дата-центра подробно рассказывает Иван Пепельняк в своей статье. Другая статья, "Виртуализация центров обработки данных", более подробно познакомит читателя с технологиями виртуализации на сетевом уровне.

Говоря о сети, новый стандарт мобильной связи 5G с самого начала поддерживает функции виртуализации. Одной из таких функций является так называемый слайсинг, или сегментация сети. Что это такое и как его можно применять обсуждает в своей статье Ченгиз Алаэттиноглу.

В этом номере мы создали новый раздел - "Исследования", посвященный работам по изучению Интернета и измерениям различных его аспектов. Раздел открывает статья Александра Венедюхина "Сервисы глобальной Сети через призму BGP", анализирующая влияние маршрутизации на надежность и доступность интернет-услуг. Надеюсь, это станет нашей постоянной рубрикой, так что будем рады интересным исследованиям Интернета.

Как всегда, нам очень интересно и важно знать ваше мнение. Что понравилось и что можно улучшить? Какие темы вы хотели бы увидеть в следующих выпусках?

Пишите нам по адресу info@internetinside.ru.

Оптимизация инфраструктуры центра данных

Иван Пепельняк (Ivan Pepelnjak)

Мало какие корпоративные центры обработки и хранения данных (ЦОДы или ЦОХДы), существующие на сегодняшний момент, требуют более двух коммутаторов, поэтому обсуждение коммутационных матриц и сравнение характеристик матриц от различных поставщиков не имеет большого смысла.

Но, как всегда, вы можете оказаться исключением. Вдруг у вас десятки тысяч виртуальных машин? Или огромные кластеры Hadoop? Или что-то другое, но столь же масштабное: например, базы данных SAP HANA или гигантские кластеры БД Oracle? Короче, вам вполне может оказаться нужна крупная коммутационная матрица. Однако сразу скажу: за последние пять с лишним лет я очень редко встречал заказчиков, которым не хватило бы двух коммутаторов последнего поколения для ЦОДа (правда, сами клиенты обычно так не думали).

Краткий исторический экскурс

Года где-то до 2010 мы рекомендовали сетевую инфраструктуру центров данных, похожую на ту, что изображена на рисунке 1.

Коммутаторы ЦОДов в те времена были на порядок слабее нынешних. У них также было меньше портов, из-за чего нам приходилось строить сети в три уровня: ядра, агрегирования и доступа. Просто потому, что подключить все сразу к центральному коммутатору было нельзя.

Мало того, обычно нам приходилось еще обсуждать границу между пересылкой уровня 2 (сетевые мосты) и уровня 3 (маршрутизация), чтобы решить, надо ли нам распространять одну VLAN на весь ЦОД, ограничившись функционалом уровня 3 только в ядре, или же использовать мосты только на уровне доступа, а провести границу L2/L3 на уровне агрегирования.

В существующих крупных ЦОДах обычно встречаются три решения этой проблемы:

- маршрутизация в ядре (включая аплинки к коммутаторам агрегации) и мосты внутри уровня агрегации; или
- маршрутизация на коммутаторах ядра и мосты через ядро;
- сети ЦОД на основе мостов с маршрутизацией на сетевых приставках (брандмауэры или балансировщики нагрузки).

Независимо от того, где проходит граница между уровнями 2 и 3, в традиционных ЦОДах обычно было три уровня и множество коммутаторов. Конечно, попадались и ЦОДы, где все подключалось к двум Catalyst 6500, но как правило, в небольших фирмах.

Использование современных технологий ЦОД

Используя современные технологии и агрессивную виртуализацию, мы можем избавиться от большей части инфраструктуры ЦОД и построить нужную вам физическую инфраструктуру всего на двух коммутаторах за несколько простых для понимания шагов:

- виртуализовать серверы;
- отказаться от устаревших технологий;
- сократить число аплинков на каждый сервер;
- использовать распределенную файловую систему;
- виртуализовать сетевые сервисы;
- построить оптимизированную коммутационную матрицу из двух коммутаторов.

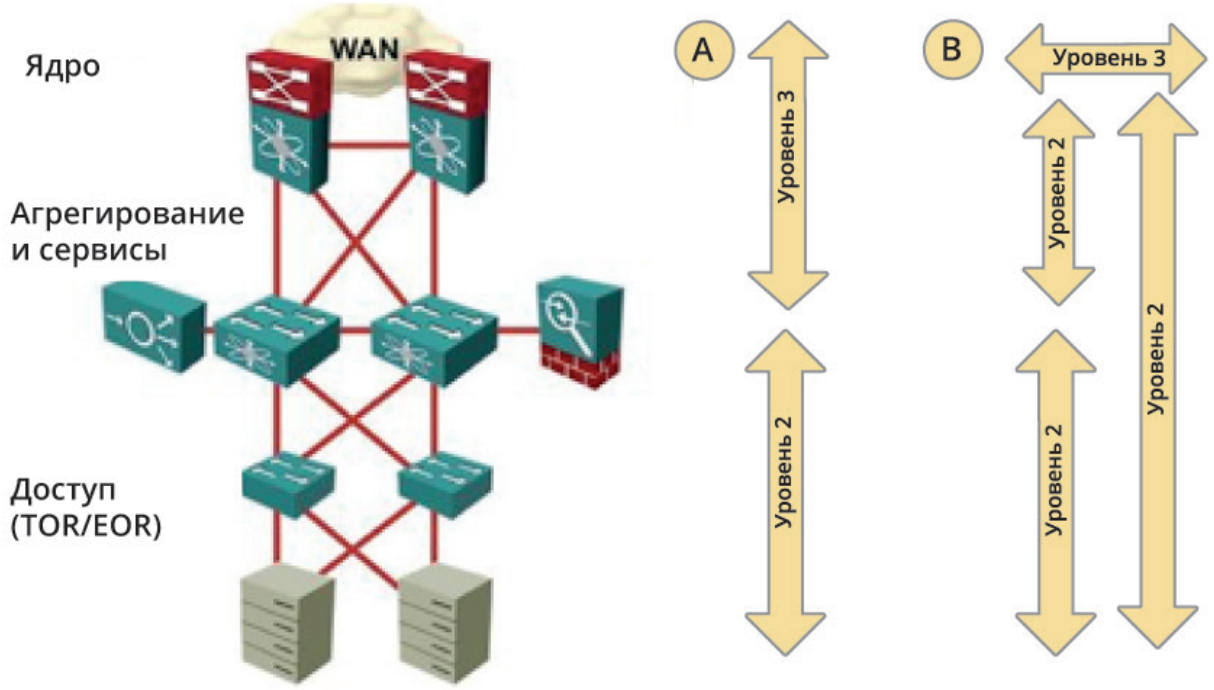
Виртуализация серверов

Первый шаг на вашем пути к оптимизированной инфраструктуре центра данных – это беспощадная виртуализация. Однако меня не перестает удивлять, сколько народу так и не внедрило виртуализацию всерьез.

С другой стороны, множество предприятий виртуализовали не менее 95% своей рабочей нагрузки, и все, что у них сейчас работает на «чисто железных» серверах, – это большие БД или приложения, которые иначе как на физическом сервере запустить невозможно.

Виртуализация большей части нагрузки целесообразна практически в любой среде. Например, мы много лет назад виртуализовали менеджер

Рис. 1. Сеть центра данных (ок. 2010 г.).



вызовов Cisco, хотя это тогда официально не поддерживалось. Я знаю людей, у которых на виртуальных машинах работают базы данных Oracle, хотя сама Oracle поддерживает такое решение лишь ограниченно. У других моих знакомых на виртуальных машинах работают базы данных SAP HANA, ворочая более терабайта оперативной памяти.

Как только вы виртуализуете нагрузку и обретете гибкость, можно переходить к выбору подходящего размера серверов. Современные серверы поддерживают CPU с большим числом ядер на каждый процессор и большим числом сокетов CPU на каждый сервер. Поэтому часто можно увидеть более 50 виртуальных машин (ВМ) приличного размера (4 ГБ оперативной памяти, 1 vCPU) на одном типовом сервере.

Чтобы оптимизировать размеры серверов под вашу среду, используйте следующую последовательность действий (в предположении, что вы знаете типовые требования к ВМ):

- выберите тип сервера, который неплохо вам подойдет;
- определите, сколько ВМ уместится на таком сервере – по числу ядер CPU, которые поддерживает выбранный сервер, типовому размеру ВМ и желательному уровню переиспользования vCPU;
- вычислите необходимый объем ОЗУ для сервера – по числу ВМ и типовому расходу ОЗУ на каждую виртуальную машину;
- найдите ограничивающий фактор, будь то число ядер или объем ОЗУ, и повторяйте процедуру до тех пор, пока не получите на выходе модель сервера с оптимальной стоимостью в пересчете на ядро.

Обычно выгоднее всего в пересчете на ядро оказываются серверы среднего уровня, но так как производители серверов непрерывно обновляют свои линейки, процедуру придется повторять каждый раз при покупке новой партии серверов.

Реальна ли виртуализация высокой плотности?

Когда я рассказываю о концепции виртуализации высокой плотности, я как правило получаю два типа ответов: а) *«да, мы так и делаем»* и б) *«какой дурак так делает?»*

Как всегда, лучше основывать свои решения на данных, чем на эмоциях. Когда Фрэнк Деннеман (Frank Denneman) работал в PernixData (теперь в составе Nutanix), он вел корпоративный блог. В одном из постов он описал анонимную статистику, собранную по тысячам серверов ESXi (vSphere) у заказчиков. Вот что выяснилось (http://frankdenneman.nl/2015/12/23/insights-into-cpu-and-memory-configuration-of-esxi-hosts/):

- У большинства серверов два сокета (84%) и от 8 до 24 ядер (у 25% было по 16 ядер). Получается, что самая распространенная конфигурация сервера – это два сокета с 8 ядрами на сокет.
- У большинства хостов ESXi от 192 до 512 ГБ оперативной памяти, причем почти 50% приходится на интервал от 256 до 384 ГБ.

Правда, PernixData – стартап, а потому его клиенты могут быть более продвинутыми, чем средний

корпоративный заказчик. Следует ожидать, что типовой корпоративный ЦОД окажется более «отсталым».

В предположении, что типичная небольшая ВМ потребует 4 ГБ памяти, на сервере, описанном в исследовании PernixData, можно будет запустить 60-70 виртуальных машин. Даже для более производительных ВМ и при небольшом уровне переиспользования цифра в 50+ виртуальных машин на сервер выглядит реальной.

В другом посте Фрэнка обсуждались данные по плотности ВМ (http://frankdenneman.nl/2016/02/15/insights-into-vm-density/). Тут мы видим еще более интересную картину, так как данные о плотности оказались совершенно разнородными: от 10 и менее виртуальных машин на хост до 250 и более. Однако, если исключить из рассмотрения экстремально мелкие случаи, на 70% из обследованных хостов оказалось более 20 ВМ, а на 34% из обследованных хостов их было более 50. Напрашивается вывод, что многие заказчики ставят по много ВМ на свои хосты ESXi.

В тему: знакомый инженер как-то рассказывал, что регулярно видит хосты ESXi с малым количеством ВМ, но при этом на них впустую пропадает чудовищное количество ресурсов. «Можно было втиснуть в эти хосты больше, – говорил мне коллега, – но почему-то они так не делают».

Еще один источник данных – индекс Cisco Global Cloud (см. рис.2), который показывает плотность ВМ гораздо ниже обсуждавшейся. Не знаю, почему так происходит. Может, облачные провайдеры используют маломощные серверы? Или участники опроса только начали процесс виртуализации нагрузки?

Краткий итог: на один физический сервер среднего уровня без проблем влезет несколько десятков ВМ. В качестве первого этапа консолидации избавьтесь от «чисто железных» серверов и втисните свою нагрузку в как можно меньшее число виртуализационных хостов оптимального размера.

Вопросы для обсуждения

Вы сами видели гигантские ВМ, такие как базы данных, особенно БД с размещением в оперативной памяти, или для этого все-таки используются «чисто железные» серверы?

Я видел гигантские ВМ (30+ ядер, 1 ТБ ОЗУ) и у общедоступных облачных провайдеров, и в корпоративных центрах данных. Как правило, они конфигурируются по одной ВМ на сервер.

Огромные рабочие нагрузки виртуализуются для упрощения обслуживания: размещение сервера БД на виртуальной машине позволяет разнести обслуживание железа и обслуживание сервера.

Например, если потребуется апгрейд сервера, замена оборудования или обновление ядра гипервизора, можно перенести ВМ заказчика на другой эквивалентный сервер и спокойно выполнить техобслуживание оборудования или гипервизора.

Либо, если вам нужен апгрейд виртуализованного сервера (требуется добавить больше процессоров или памяти), провайдер может перенести вашу ВМ на более мощный физический сервер, а затем просто сменить

Рис. 2. Уровень виртуализации все еще слишком низок. Источник: Cisco Global Cloud Index 2014-2019.

настройки виртуальной памяти и виртуальных процессоров, не прерывая работу ВМ. В худшем случае придется перезапустить ВМ, что на порядок быстрее, чем копание в физическом сервере и установка дополнительных процессоров или памяти.

У большинства заказчиков со сверхбольшими ВМ выполняется одна ВМ на физический сервер, но они все же используют виртуализацию по описанным выше причинам. Однако будьте осторожны при создании ВМ с очень большим числом ядер CPU. Например, если ВМ вашей базы данных требует 32 ядер, она не запустится до тех пор, пока не станут доступны как минимум 32 ядра. Поэтому выполнение 32-ядерной ВМ на 32-ядерном сервере приведет к падению быстродействия: каждый раз, когда гипервизору потребуется запустить новую задачу, он остановит ВМ, и в это время большинство ядер будет простаивать. Обязательно держите на физическом сервере больше ядер, чем требуется для ВМ.

Имеет ли смысл виртуализовать кластеры Hadoop?

Ответ на этот вопрос зависит от двух аспектов:

- Какой уровень гибкости вам нужен?
- Сколько ресурсов может потребовать сервер?

Если ваши серверы (или задания, которые на них выполняются) долгоживущи, имеет смысл их виртуализовать. Если вы планируете независимые краткосрочные задания, которые не используют постоянных данных на серверах, то неважно, виртуализованы они или нет – их в любой момент можно будет отключить.

Еще одно соображение: если ваши серверы Hadoop могут захватить все ядра CPU (или все ОЗУ) на физическом сервере (я предполагаю, что вы выполнили описанную выше процедуру и нашли оптимальную конфигурацию физического сервера), то виртуализация кластера Hadoop может и не иметь смысла, помимо преимуществ при техобслуживании/модернизации.

Если же рабочая нагрузка на сервере не может захватить все его ресурсы ОЗУ и процессоров, то опять же виртуализация имеет смысл, чтобы выжать максимум из купленных физических ресурсов.

Универсального ответа тут нет. Единственный способ понять, можно ли выжать больше из имеющегося оборудования – мониторинг использования ресурсов на физических серверах.

Долой старье

Если вы хотите продолжать оптимизацию вашей сети центра данных, необходимо отказаться от устаревших сетевых технологий, в первую очередь от Gigabit Ethernet в качестве основного средства связи между серверами и сетью.

Сегодня применение Gigabit Ethernet имеет смысл для внеполосного управления или в сетях iLO/KVM.

Краткий экскурс в прошлое

Хотя в начале 2010-х годов некоторые прогрессивные заказчики и думали о 10 Gigabit Ethernet (10GE), мы все еще настоятельно рекомендовали использовать для серверных соединений Gigabit Ethernet (GE). В основном потому, что GE был хорошо известной и проверенной на деле технологией, которая уже имелась на материнских платах большинства серверов. Кроме того, GE поддерживал медные кабели, которые тогда не поддерживались 10GE.

Основным недостатком соединений GE было большое количество необходимых интерфейсов на одном сервере. По правилам дизайна VMware рекомендовалось иметь от четырех до десяти интерфейсных карт GE на один хост ESX (две для пользовательских данных, две для хранения, две для vMotion...). Также было невозможно консолидировать хранение и сеть, реализовать передачу данных без потерь по сетям GE без ущерба для обычного сетевого трафика.

В те дни ЦОДы, устанавливавшие каналы 10GE, уже использовали быстрый vMotion, а также объединяли инфраструктуры хранения и сети.

Установка 10GE NIC также позволяла сократить число физических интерфейсов, хотя приходилось использовать аппаратные «фишки» вроде сетевой платы Cisco Palo, которая разделяла физическую сетевую плату на несколько виртуальных NIC, на которых можно было реализовать правила дизайна VMware – у самих ESXi в те дни не было встроенного QOS.

Нынешние рекомендации

Если вы строите сети для нового центра данных, я вас буквально умоляю не использовать GE в качестве основной технологии подключения серверов: только для сети внеполосного управления - и нигде больше. Для основных каналов подключения серверов используйте 10GE и выбирайте коммутаторы, которые уже сейчас поддерживают каналы 25GE: через несколько лет на серверных материнских платах появятся сетевые карты 25GE.

В наши дни 10GE открывает массу возможностей для подключения:

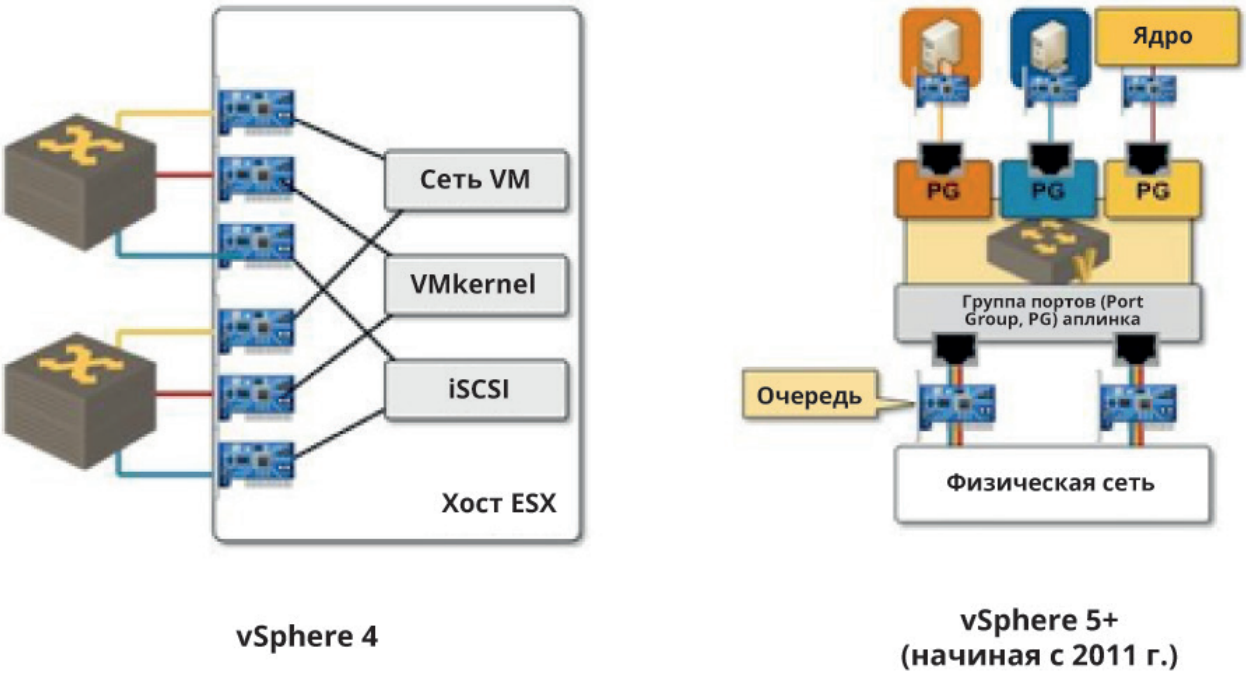
- Можно использовать 10GBASE-T, если вам нужны медные кабели в рамках стойки или небольшого кластера. Коммутаторы с портами 10GBASE-T предлагают практически все поставщики коммутаторов, и нетрудно купить серверы со встроенными интерфейсами 10GBASE-T.
- Если вам нужно использовать интерфейсы SFP+ или QSFP, возьмите твинаксиальный кабель для коротких расстояний и оптоволокно для длинных.

История из реальной жизни

В конце 2014 года меня пригласили проверить дизайн ЦОДа для финансовой организации, которая модернизировала сеть в своем дата-центре.

Заказчик менял старые коммутаторы Catalyst 6500 на новые коммутаторы от другого поставщика. У новых коммутаторов были только серверные соединения 1GE и аплинки 10GE, поэтому я спросил: *«Почему? На дворе 2014 год, почему вы не покупаете 10-гигабитные линки?»*. И получил ожидаемый ответ: *«у нас все серверы в центре данных с гигабитными NIC»*.

Рис. 3. Прочтите новейшее руководство по дизайну vSphere.



Следующий вопрос: *«О'кей, вы сейчас купите коммутаторы с интерфейсами GE, которые через несколько лет устареют. И как вы будете к ним подключать серверы нового поколения с 10GE NIC, которые вы рано или поздно все равно купите?»*

К сожалению, тут сделать ничего было нельзя. Мои собеседники прекрасно понимали, что делают не лучший выбор, но серверы покупались вообще из другого бюджета, поэтому модернизация серверов никогда не синхронизировалась с модернизацией сети. Кроме того, раз большинство серверов требовало нескольких соединений GE, покупать для них коммутаторы с портами 10GE было бы неоправданно дорого.

Вывод: если вы хотите консолидировать свой центр данных и оптимизировать затраты, учитывайте сразу все компоненты центра данных (серверы, системы хранения и сеть), а при необходимости модернизируйте все компоненты одновременно.

Вопросы для обсуждения

Мы рассматриваем возможность перехода на 10GE, но нас отпугивает стоимость подлинных SFP, а вендоры не хотят поддерживать твинаксиальный кабель в смешанной среде – например, при использовании коммутаторов Cisco с серверами IBM, поэтому мы предпочли бы 10GE оптоволокно.

Понятно, что поставщики сетевых решений просто вынуждены работать с высокой маржой, чем можно объяснить тот факт, что модули RAM и SFP у них стоят дороже; но цифры, приведенные в статье, уже не лезут ни в какие ворота.

Похоже, что крупные поставщики решений для центров данных закладывают в цену SFP-модуля скрытую лицензионную плату за порт, чтобы снизить стоимость базового оборудования и заработать побольше на SFP, которые вы вынуждены покупать у них же.

Возможно, ситуация улучшится по мере того, как появятся небрендовые коммутаторы, а инженеры центров данных поймут, что их тоже можно использовать в некоторых средах, но пройдет еще немало времени, пока корпоративные центры данных решатся покупать noname-коммутаторы под управлением ПО независимых разработчиков.

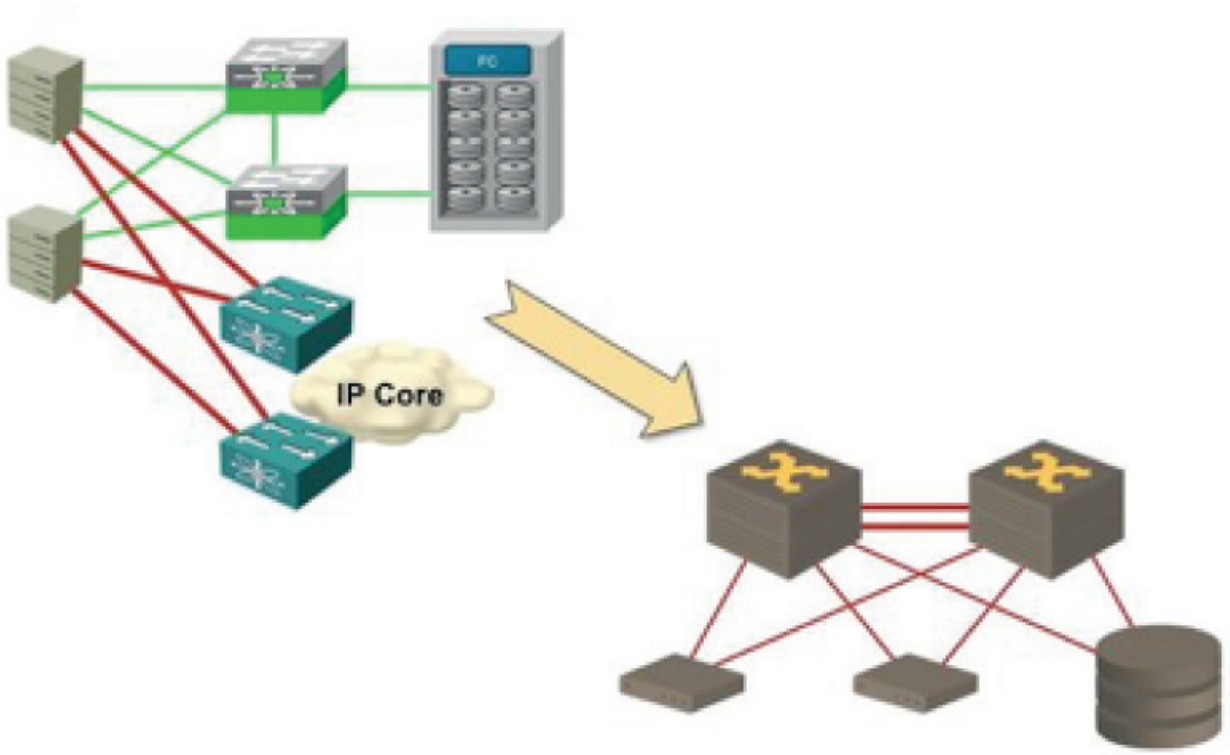
Сокращение числа аплинков «сервер-сеть» в вашем центре данных

Простой способ свести к минимуму число трансиверов, кабелей и сетевого оборудования в центре данных – минимизировать число аплинков на каждый сервер. Вроде бы, просто, но путь к этой цели редко бывает тривиальным.

Первое препятствие на вашем пути – то, что архитекторы виртуализации часто используют устаревшие руководства по дизайну vSphere. На рисунке 3 приведено сравнение рекомендаций VMware для ESXi версии 4 и более новых рекомендаций для ESXi версии 5 и выше.

В vSphere версии 4 не было QoS, а потому VMware рекомендовала использовать отдельные интерфейсы для трафика ВМ, для трафика ядра (например, VMotion) и для трафика хранения. В результате с минимальным резервированием получалось не менее шести сетевых интерфейсов на сервер.

Рис. 4. Замените FC/FCoE на iSCSI или NFS.



Даже в vSphere 4 можно было проектировать более оптимальные сети, используя функции QoS из Cisco Nexus 1000V либо виртуальные сетевые интерфейсы из лезвийных шасси UCS.

В vSphere 5 компания VMware реализовала зачатки QoS под названием NIOC (Network I/O Control) и начала рекомендовать по два интерфейса 10GE на сервер с настройкой QoS на серверных аплинках для выделения полосы трафику ВМ, хранения и vMotion.

Напомню, что сейчас используется vSphere 6, а vSphere 5 появилась в 2011 году. К сожалению, я слышал про шесть сетевых карт на каждый сервер долгие годы после выхода vSphere 5.

Как эти изменения рекомендаций по дизайну влияют на вас, если вам требуется перейти от старого сервера под управлением vSphere 4 на новый, более мощный, под управлением vSphere 5 или 6? Можно повысить уровень виртуализации (например, от десяти ВМ на сервер до 50 ВМ на сервер), в то же время сократив число интерфейсов.

Если почти десять лет назад у вас был коммутатор Gigabit Ethernet на 48 портов, к нему можно было подключить шесть серверов и 80 ВМ. Сегодня к одному коммутатору 10GE на 48 портов можно подключить 24 сервера, на которых будет работать до тысячи виртуальных машин – таким образом, размер сети и ее сложность сократятся радикально.

После консолидации сетевых интерфейсов на серверах пришло время пересмотреть дизайн сети хранения данных. На традиционных серверах были интерфейсы Ethernet, выделенные для ВМ или трафика управления, и отдельные интерфейсы Fibre Channel для хранилищ. В современном дизайне весь трафик консолидируется в одном наборе аплинков и используется хранилище на базе IP или FCoE (см. рис.4).

Замена Fibre Channel на IP-хранилища (NFS или iSCSI) позволяет значительно снизить число портов и сложность:

- число серверных аплинков (вместе с оптикой и кабелями) сокращается вдвое;
- число коммутаторов доступа (листьев) сокращается вдвое;
- убирается целый технологический стек.

Инженеры систем хранения данных старого закала не доверяют IP-хранению, потому что не имели дела с технологиями хранения на основе IP. С другой стороны, немало крупных предприятий уже не первый год эксплуатируют крупные системы на базе iSCSI или NFS.

Если вам по-прежнему нужна поддержка Fibre Channel в вашем центре данных (например, для подключения устройств резервного копирования на ленту), постарайтесь консолидировать серверные аплинки. Замените раздельные интерфейсы LAN и SAN на каждом сервере (это минимум четыре порта) на два порта 10GE с поддержкой FCoE.

Консолидация Fibre Channel и Ethernet между серверами и коммутаторами доступа вдвое снижает число точек управления, так как вам больше не нужны отдельные коммутаторы доступа для Fibre Channel и Ethernet.

Если вы хотите использовать FCoE между серверами и коммутаторами доступа, разделите трафик Fibre Channel и Ethernet на коммутаторах доступа и организуйте отдельные ядра Fibre Channel и Ethernet. Multihop FCoE привносит в дизайн ненужный уровень сложности.

Некоторые архитекторы систем хранения предпочитают отделять сети iSCSI от сетей данных или как минимум использовать отдельные каналы «доступ-ядро» и выделенные коммутаторы в ядре для iSCSI-компонента сети.

Однако, если вы уже выполнили остальные действия по консолидации, вам будет нетрудно настроить серверы на пересылку данных ВМ в основном через один коммутатор доступа, а другой можно задействовать прежде всего для iSCSI или NFS (при этом каждый из коммутаторов будет реализовывать все функции резервирования для другого).

Конечный результат: до тех пор, пока все порты работают, данные ВМ и данные хранения смешиваться не будут – они относятся к двум разным VLAN и используют отдельные коммутаторы доступа. Но стоит одному из коммутаторов доступа или линии от сервера к коммутатору отказать, как весь трафик автоматически переключается на второй коммутатор или линию.

Использование распределенной файловой системы

Предыдущие шаги по оптимизации инфраструктуры центра данных (от массивной виртуализации до перехода к двум аплинкам 10GE на сервер) уже нашли одобрение в индустрии. А вокруг использования распределенной файловой системы (VMware VSAN, Nutanix, Ceph или GlusterFS) все еще ломают копья.

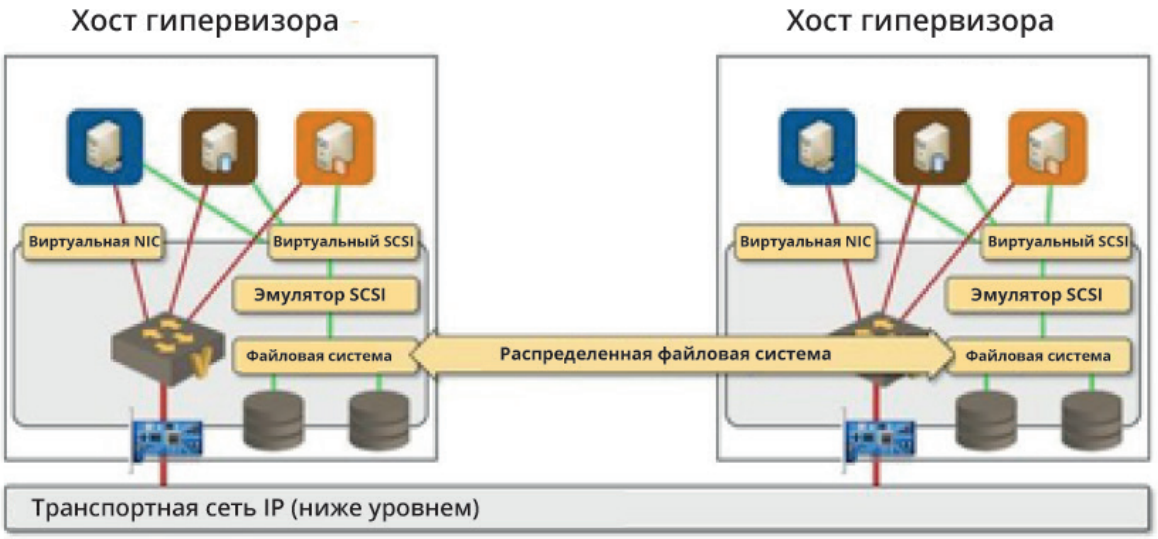
Основополагающая идея очень проста. Вместо традиционных массивов хранения данных можно организовать локальное хранение на хостах гипервизора (на жестких дисках или SSD) в распределенной глобальной файловой системе с репликацией между узлами гипервизора по IP-сети центра данных (см. рис.5).

Преимущества такого подхода очевидны:

- нужно меньше различных аппаратных компонентов;
- повышается устойчивость: централизованные компоненты высокой доступности (массивы хранения) заменяются на распределенную сеть дешевых устройств;
- сокращается потенциал для сбоев – отказ одного узла в распределенной файловой системе меньше влияет на общую систему, чем отказ массива хранения;
- повышается общее быстродействие... если, конечно, гипервизоры в первую очередь обращаются к собственной локальной файловой системе;
- линейное масштабирование быстродействия – чем больше узлов в распределенной файловой системе, тем больше ее общее быстродействие.

Недостатки у распределенных файловых систем тоже имеются, как очевидные, так и не очень:

Рис. 5. Распределенные файловые системы систем хранения DAS (Direct attached storage).



- На каждом гипервизоре используются локальные диски (DAS) и/или SSD
- Глобальная файловая система или хранилище объектов с репликацией файлов между узлами гипервизора
- Примеры: Ceph, GlusterFS, VMware VSAN, OpenStack Swift

- репликация данных между узлами требует высокопроизводительной сети центра данных;
- репликация в трех направлениях, используемая многими распределенными файловыми системами, дает перегрузку в 200% по сравнению с 20-25% для RAID-6 (если используется один массив хранения);
- сетевая инфраструктура становится критическим компонентом – отказ сети быстро приводит к полному коллапсу хранения;
- распределенные файловые системы еще не достигли такой зрелости, как дисковые массивы, а администраторы систем хранения данных – народ пугливый и консервативный.

Несколько лет назад у распределенных файловых систем не было шансов в большинстве корпоративных сред, даже несмотря на то, что Hyper-V и Linux поддерживали распределенные файловые системы уже довольно давно.

Сегодня распределенные файловые системы успешно работают в ряде очень крупных инсталляций. Например, общедоступные облака, построенные на базе OpenStack, часто используют Ceph или GlusterFS, или даже Swift (хранилище объектов OpenStack) для хранения образов ВМ.

Среды vSphere были последним бастионом традиционного подхода к хранению. Первая распределенная файловая система под vSphere была создана Nutanix, потом через несколько лет появилась VMware VSAN (VSAN вышла в конце 2013 года).

Очевидно, вряд ли удастся убедить кого-либо хранить базу данных Oracle в распределенной файловой системе, но можно найти распределенную файловую систему, достаточно хорошую для виртуальных дисков ВМ, что позволит снизить требования к дисковым массивам.

Виртуализация сетевых сервисов

После виртуализации вычислительных ресурсов и инфраструктуры хранения пришло время виртуализовать сетевые сервисы – а в последние несколько лет предложения виртуальных сетевых устройств от различных поставщиков посыпались на рынок одно за другим. К сожалению, эти устройства обычно представляют собой всего лишь традиционные продукты в «обертке» для ВМ-формата, с низкоуровневым кодом пересылки пакетов, переписанным для паравиртуальных драйверов.

Начнем с виртуальных маршрутизаторов. Vyatta (теперь Brocade, часть Broadcom – прим. ред.) был, возможно, первым коммерческим виртуальным маршрутизатором, за ним последовал Juniper со своей виртуальной SRX (сейчас мы не будем касаться вопроса о том, правильнее считать vSRX маршрутизатором или брандмауэром). Несколько лет назад Cisco выпустила Cloud Services Router (IOS-XE), а сегодня каждый крупный поставщик сетевых решений (включая HP и Alcatel-Lucent, теперь в составе Nokia) предлагает виртуальный маршрутизатор. Также имеются продукты для провайдеров коммуникационных услуг в виртуальном формате, такие как Cisco IOS XR или Juniper vMX.

Каждый крупный поставщик брандмауэров предлагает виртуальный брандмауэр. Началось это уже давно, с Juniper vSRX, через несколько лет Cisco выпустила vASA и ASAv, затем последовали Palo Alto и Checkpoint. На рынке балансировщиков нагрузки ситуация примерно та же. Сейчас в виртуальном формате предлагаются уже практически любые сетевые устройства.

Примеры продуктов

- Маршрутизаторы: Brocade Vyatta, Cisco CSR, Juniper vMX.
- Брандмауэры: pfSense, Juniper vSRX, Palo Alto, Vyatta, vShield Edge (VMware), vASA (Cisco).
- Балансировщики нагрузки: BIGNIP VTM (F5), Zeus Traffic Manager (теперь Riverbed), vShield Edge (VMware), Embrane, LineRate Systems (теперь F5), Citrix NetScaler.

Зачем может потребоваться замена физических сетевых устройств виртуальными? Вы немедленно обретаете большую гибкость, так как развертывать виртуальные устройства можно по мере надобности. Нет задержки на покупку новых стоек и кластеров, а иногда можно использовать временные лицензии, чтобы упростить процедуру приобретения и развертывать сетевые сервисы по требованию.

Интересно, что как только что-то начинает работать в производстве по временной лицензии, бюджет обычно перестает быть проблемой.

Виртуальные приставки также упрощают дизайн для высокой доступности и аварийного восстановления:

- Часто оказывается, что вам не нужен резервный брандмауэр или парные балансировщики нагрузки (либо их кластеры), потому что можно перезапустить виртуальное устройство за считанные секунды.
- Виртуальное устройство всегда можно перезапустить (это же обыкновенная виртуальная машина, ничем не отличающаяся от прочих) в центре данных для аварийного восстановления, поэтому вам не нужно запасное оборудование, которое будет пылиться на запасном пункте в ожидании своего часа.

В ЦОДе аварийного восстановления, построенном по виртуальным технологиям, вам нужно лишь то оборудование, которое обеспечит достаточно вычислительной мощности для обработки дополнительной нагрузки при отказе основного центра.

Экономия при использовании виртуальных устройств стала настолько очевидна, что у F5 виртуальный балансировщик нагрузки стоит дороже физического: они знают, что вы купите только один, в то время как физических нужно четыре.

Мы уже говорили о том, как виртуальные устройства сокращают

время развертывания. Они также минимизируют требования к запасному оборудованию: вам больше не нужен запасной маршрутизатор, брандмауэр, балансировщик нагрузки, IDS или дорогостоящий контракт на техобслуживание всего этого железа – все ваши сетевые сервисы работают на унифицированной компьютерной инфраструктуре.

И, наконец, виртуальные устройства упрощают инфраструктуру физической сети. Вместо спутанного клубка из аппаратных устройств, подключенных к коммутаторам в ЦОДе (см. левую часть рисунка ниже), мы получим серверы, подключенные по схеме «ствол и листья».

Можно сказать, что мы виртуализовали этот клубок – и это будет правильно (см. рис.6).

Пойдем дальше: если вы захотите использовать распределенную файловую систему и виртуальные сетевые устройства в вашем центре данных, то вам потребуется всего два аппаратных компонента: коммутаторы и серверы. Представьте, насколько сократятся расходы на техобслуживание!

Некоторые инженеры дошли до того, что подключили свое закрытое облако к общедоступному Интернету прямо через виртуальные SRX на своих серверах.

Создание оптимизированной коммутационной матрицы

Подведем некоторые итоги. Мы:

- виртуализовали все серверы;
- перешли на 10GE (или 25GE);
- сократили число серверных аплинков (и портов коммутаторов);
- распределили хранение по хостам гипервизора; и
- виртуализовали сетевые сервисы.

В результате каждый вычислительный ресурс (включая сетевые сервисы) оказался виртуализован, и вы можете разместить все это на гораздо меньшем количестве хостов гипервизора, чем раньше.

Создание кластера сетевых сервисов

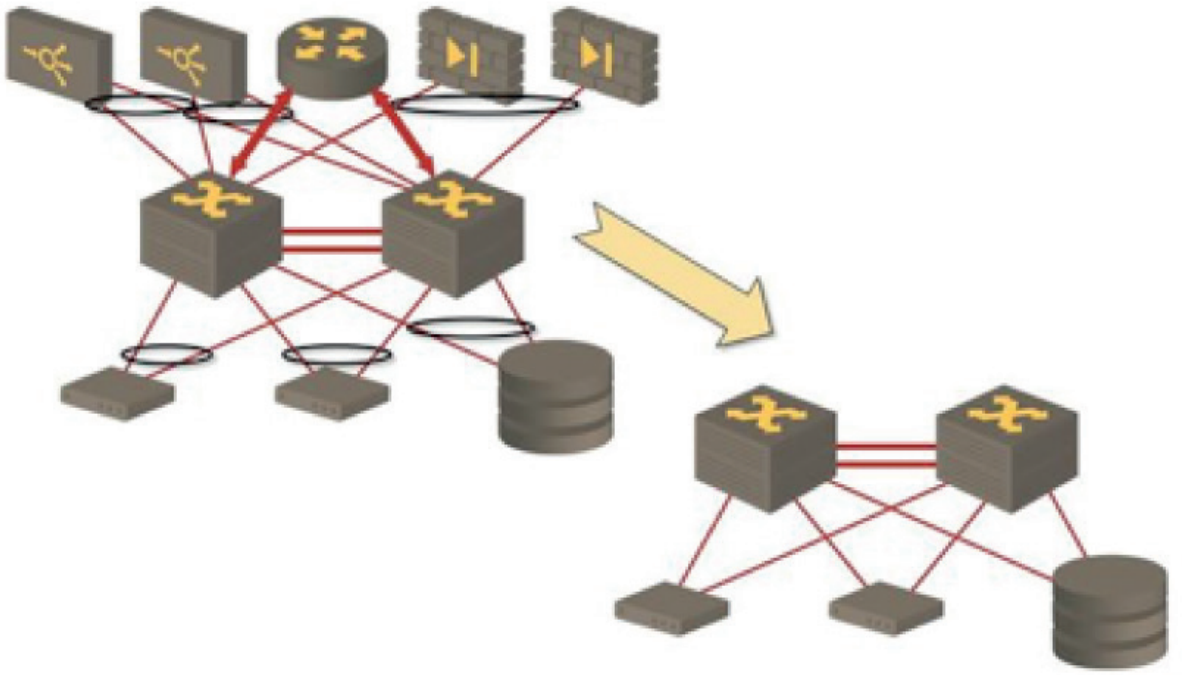
Всегда, за исключением очень маленьких сред, где просто нельзя оправдать дополнительные капиталовложения, запускайте виртуальные сетевые устройства на отдельных физических серверах – из соображений безопасности и быстродействия.

Если ваши виртуальные устройства подключаются непосредственно к общедоступной сети, подумайте о выделенных интерфейсах Ethernet для подключения к общедоступной сети, чтобы полностью отделить внутреннюю матрицу ЦОД от внешней сети, как показано на диаграмме, представленной на рисунке 7.

Создание физической инфраструктуры коммутации

Сколько коммутаторов вам нужно в оптимизированном дата-центре? По самой приблизительной оценке, двух коммутаторов вам хватит примерно на две тысячи виртуальных машин. Иван Раабок (Iwan Rahabok) проделал более тщательный анализ (http://virtual-red-dot.info/1000-vm-per-rack-is-the-new-minimum/) и вычислил, что можно уместить примерно тысячу

Рис. 6. Виртуализация устройств.



ВМ в одной стойке (он принял для расчетов весьма консервативный коэффициент виртуализации 30:1), а все хосты ESXi, необходимые для этого, подключить к двум коммутаторам ToR.

Абсолютно все изготовители коммутаторов для центров данных предлагают 64-портовые коммутаторы 10GE или 25GE. Иногда в спецификациях говорится, что у коммутатора 48 портов 10GE и 4-6 портов 40GE, но порты 40GE обычно можно развести на четыре канала по 10GE (а порты 100GE – на четыре канала 10GE или 25GE).

Cisco часто является исключением из этого «правила». Многие из ее коммутаторов используют тот же самый кремний Broadcom, что и конкуренты, но дают вам только 48 портов, потому что чипсеты Broadcom (в частности, чипсеты Trident-II) не могут пересылать небольшие пакеты со скоростью линии.

Также имеется ряд поставщиков (включая Cisco), предлагающих коммутаторы высотой 1 RU или 2 RU с более чем 64 портами на коммутатор – начиная с 96 портов 10GE у Cisco 93120 до 128 портов 10GE в некоторых коммутаторах Dell, Arista или Juniper.

Подытожим: если в вашем ЦОДе не более 2000 серверов (физических серверов или ВМ) и вы можете их виртуализовать, то вам скорее всего не потребуется больше двух коммутаторов. Большинство корпоративных дата-центров с запасом попадают в эту категорию.

Вопросы для обсуждения

Как реализовать обходные пути (multipathing) для уровня 2 в таком простом центре данных, не используя TRILL (или FabricPath) или VXLAN?

В простой сети на два коммутатора один из них используется в качестве основного для всего трафика ВМ, а второй в качестве основного для всего трафика хранения, поэтому необходимости в балансировке нагрузки или обходных путях просто не возникает.

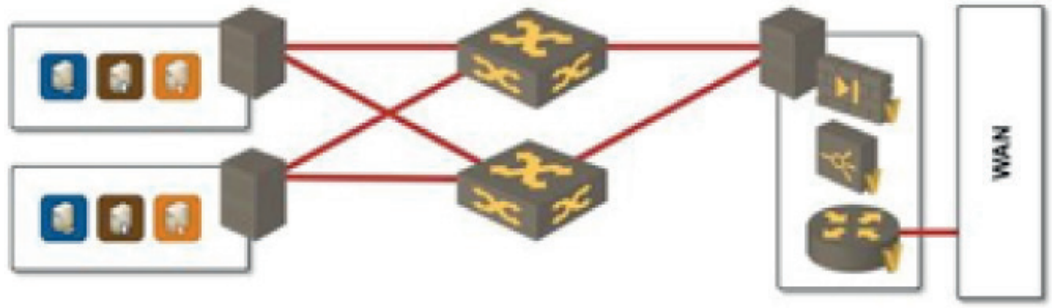
Если коммутаторов больше двух, то, очевидно, потребуются обходные пути, так как нельзя ограничивать себя остовным деревом. В таких случаях я бы настоятельно рекомендовал VXLAN на коммутаторах ToR (Top of the Rack) или (если ваши коммутаторы не поддерживают VXLAN) MLAG между границей сети и коммутаторами ядра.

Если вы разделяете трафик iSCI и ВМ между двумя коммутаторами, то какой метод лучше подойдет для отказоустойчивости на случай отказа одного коммутатора?

Очевидно, оба коммутатора должны находиться в одной VLAN, т.е. ваша сеть из двух коммутаторов должна быть чистой сетью уровня 2 с мостами через канал между маршрутизаторами.

Дальше настройте простую отказоустойчивость на ваших хостах vSphere или KVM. Для каждого типа трафика создайте один основной аплинк и один резервный.

Рис. 7. Будет ли это работать?



- 1500 ВМ на стоечном пространстве в 28 RU (56 блейд-хостов) http://virtual-red-dot.info/1000-vm-per-rack-is-the-new-minimum/
- У Cisco Nexus 93120 насчитывается 96 портов 10GE

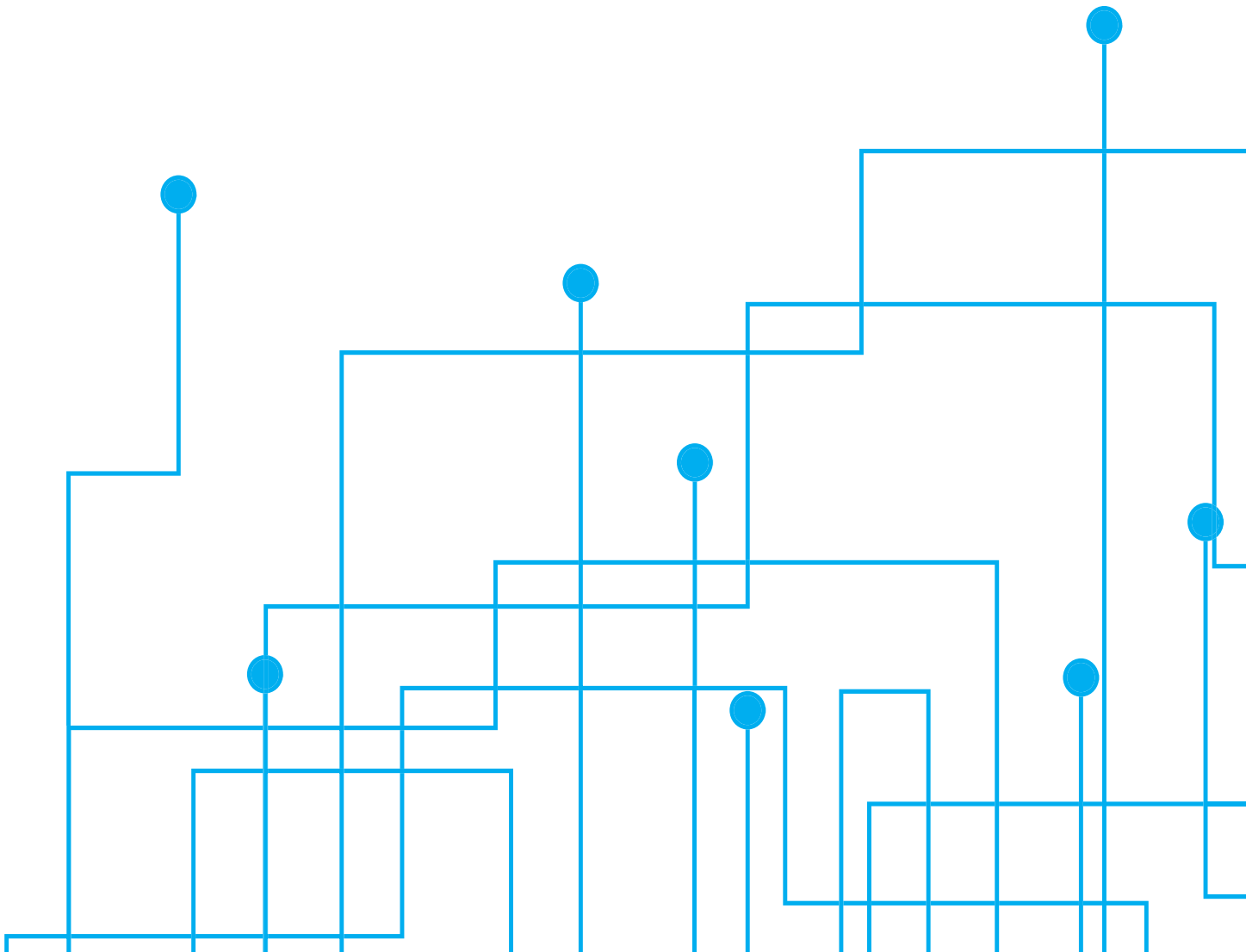
Вывод: все, что нужно для создания центра обработки данных, это два коммутатора

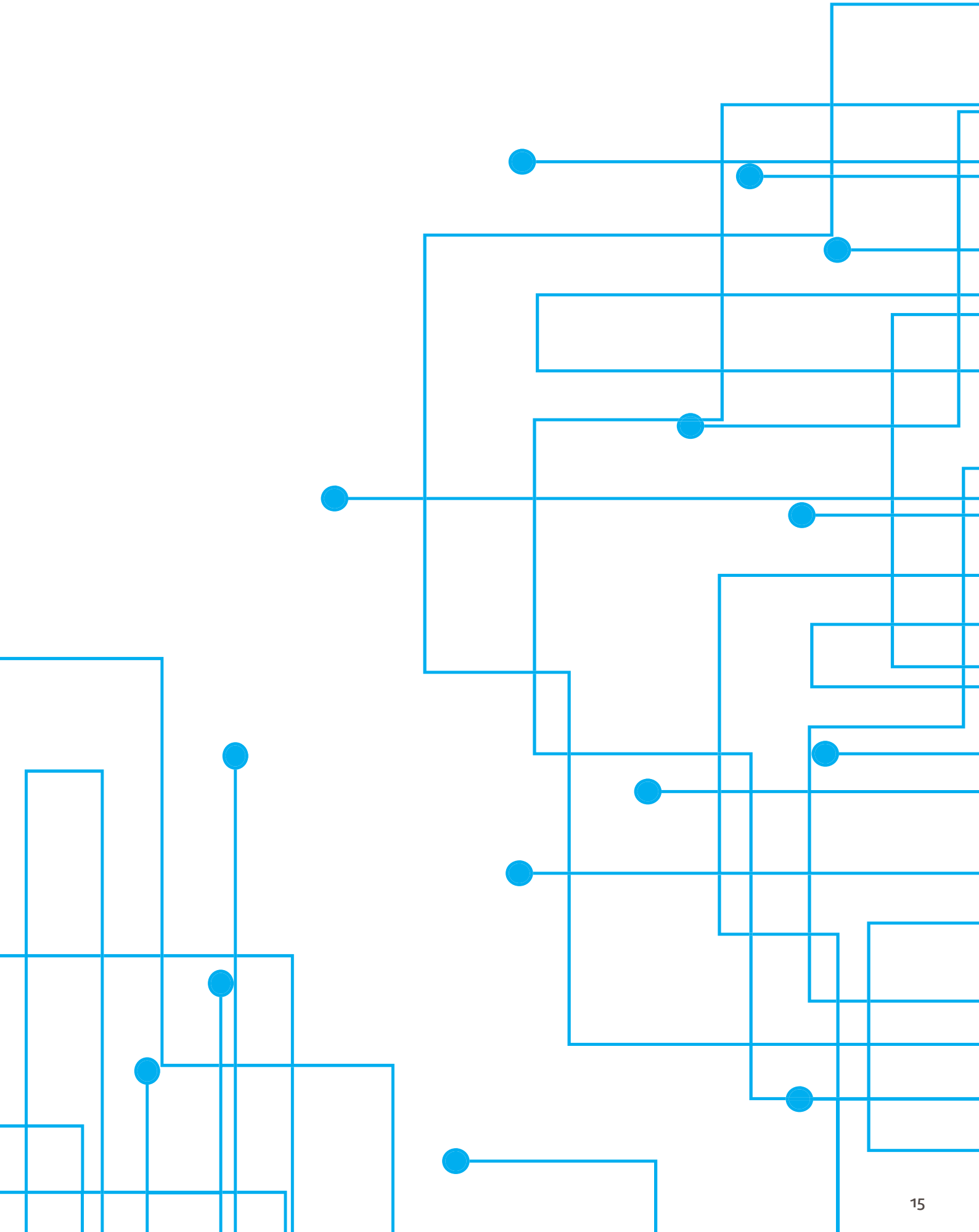
Об авторе

Иван Пепельняк, почетный член CCIE (CCIE#1354), занимается разработкой и внедрением крупномасштабных сетей поставщиков услуг и корпоративных сетей, а также преподаванием и написанием книг о передовых технологиях с 1990 года.

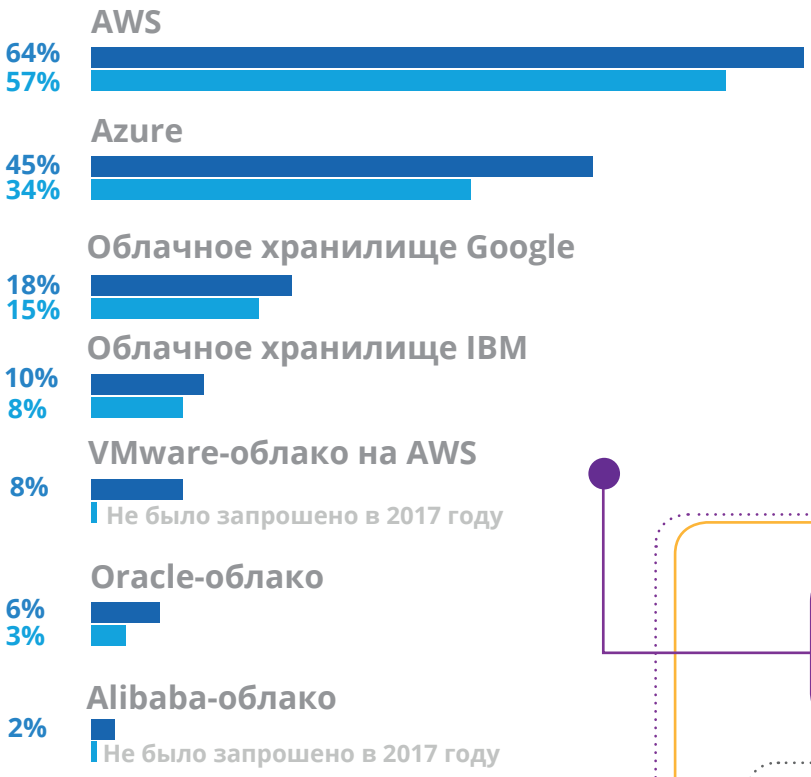
Он автор нескольких книг Cisco Press, а также плодовитый блогер и писатель, консультант и автор серии очень успешных вебинаров.

В настоящее время в центре его интересов крупномасштабные облачные системы, программируемые сети (SDN) и центры обработки данных (SDDC), а также виртуализация сетевых функций (NFV).





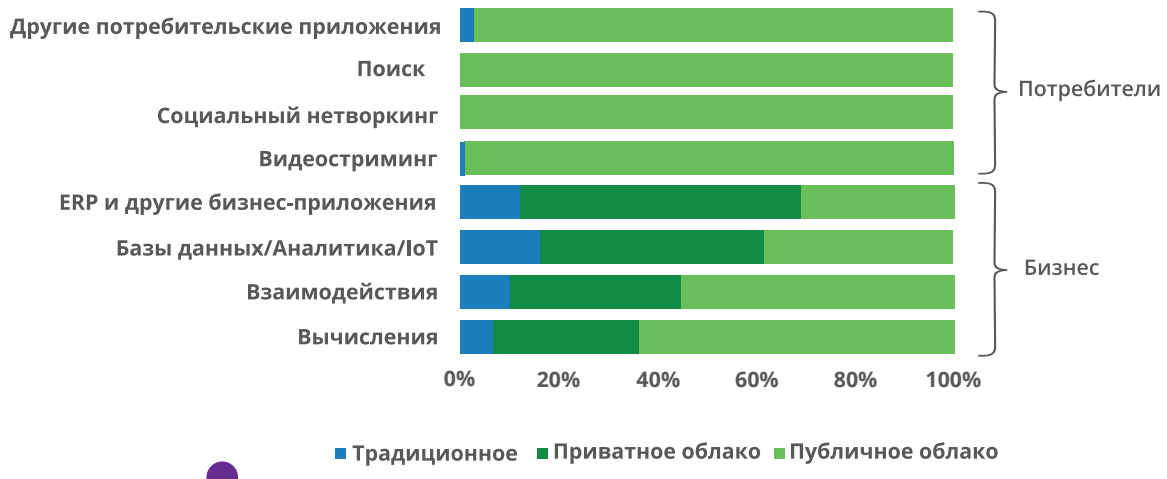
РОСТ ГЛОБАЛЬНОГО IP-ТРАФИКА ДАТА-ЦЕНТРОВ И ОБЛАЧНЫХ ХРАНИЛИЩ



Источник:
RightScale 2018 State of the Cloud Report

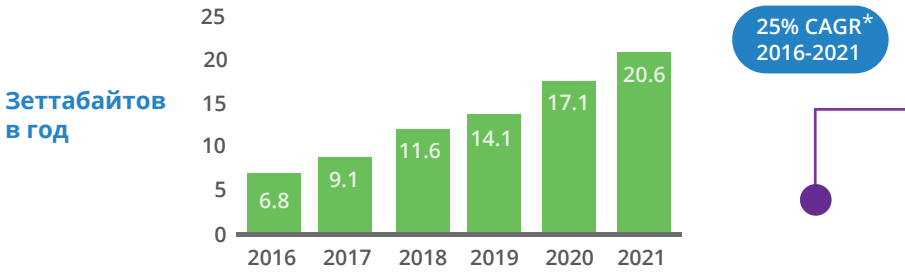
Используемые платформы и приложения в 2017-2018

Источник:
Cisco Global Cloud Index, 2016-2021



Перспективы использования вычислительных платформ (традиционных и облачных) для различных приложений в дата-центрах, 2021 год

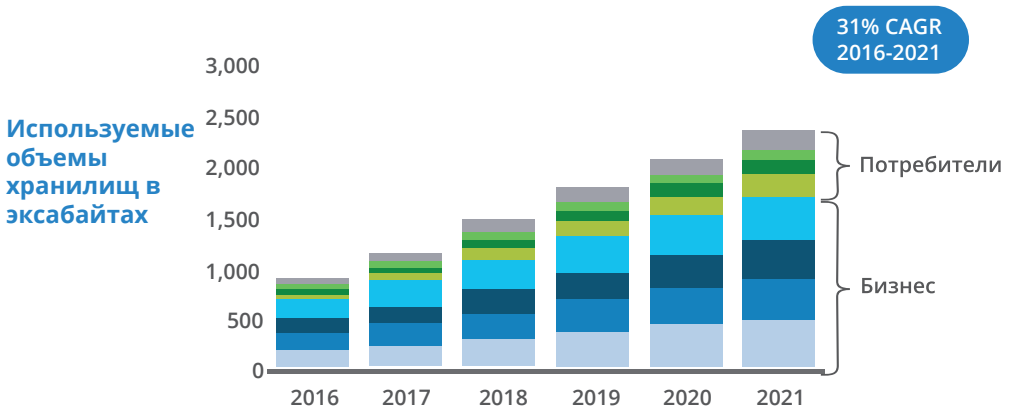
Глобальный рост IP трафика дата-центров



Источник:
Cisco Global Cloud Index, 2016-2021

*Compound Annual Growth Rate (CAGR) – совокупный годовой прирост

Источник:
Cisco Global Cloud Index, 2016-2021



Глобальный объем хранилищ дата-центров

A В дата-центре (71.5%)

Хранилища, продакшн и разработческая среда, аутентификация

B Дата-центр к дата-центру (13,6%)

Репликации, CDN, межоблачные ссылки

C Дата-центр к пользователю (14,9%)

Web, email, внутренний VoD, WebEx...

Источник:
Cisco Global Cloud Index, 2016-2021

Глобальный трафик дата центров по назначению в 2021 году

Сегментная маршрутизация:

отбрасываем шелуху и находим золотой самородок в предложении IETF

Эдриан Фаррел (Adrian Farrel),
Рон Боника (Ron Bonica)

Сегментная маршрутизация (Segment Routing, SR) – это новая технология инжиниринга трафика, разрабатываемая рабочей группой SPRING в составе IETF. Для SR определяется две инкапсуляции плоскости передачи (forwarding plane): MLPS (Multiprotocol Label Switching) и IPv6 с заголовком расширения сегментной маршрутизации (Segment Routing Extension Header). В этой статье мы обсудим исторический контекст, описав протоколы плоскости передачи и плоскости управления MPLS, объясним, как работает сегментная маршрутизация, введем концепцию плоскости передачи MPLS-SR и покажем, как работает плоскость управления SR. И, наконец, мы сравним SR с традиционными системами MPLS и перечислим ее уникальные преимущества.

Передача MPLS

Технологии MPLS уже почти 20 лет. Домен MPLS представляет собой непрерывный набор маршрутизаторов LSR (Label Switching Router). Пакеты поступают в домен MPLS через входной LSR и покидают его через выходной LSR. Один и тот же LSR может служить входным для одних пакетов и выходным для других.

LSP-маршрут (Label Switched Path) осуществляет соединение между входным и выходным LSR. LSP может проходить по пути наименьших затрат или по пути, определяемому путем инжиниринга трафика.

Когда входной LSR получает пакет, он назначает ему класс эквивалентности передачи (Forwarding Equivalence Class, FEC) и инкапсулирует пакет с помощью стека меток MPLS. Затем он передает пакет следующему транзитному участку, связанному с данным FEC.

Стек меток MPLS содержит одну или несколько записей. Каждая запись из стека меток содержит метку, индикатор времени жизни (TTL), индикатор класса трафика (TC) и нижнюю часть индикатора стека. Эти данные определяют то, как транзитный LSR будет обрабатывать пакет. В этом смысле каждая запись из стека меток представляет собой инструкцию для LSR.

Когда LSR получает пакет, он считывает верхнюю запись в стеке меток и уменьшает TTL. Если TTL еще не равно нулю, то LSR выполняет поиск в своей базе информации о передаче (Forwarding Information Base, FIB) записи, которая соответствует входящей метке.

Если LSR найдет запись FIB, которая соответствует входящей метке, то эта запись FIB будет содержать следующие данные:

- действие с меткой;
- интерфейс следующего транзитного участка.

Действия с меткой следующие:

- добавить в стек меток одну или несколько новых записей;
- снять верхнюю запись со стека меток;
- заменить метку в верхней записи.

Найдя подходящую запись FIB, LSR выполняет действие с меткой и пересылает пакет через интерфейс следующего транзитного участка. Это может быть внутренний или внешний интерфейс. Если интерфейс внутренний, то LSR пересылает пакет сам себе и обрабатывает его так, как если бы только что его получил, начиная с самого внешнего заголовка протокола. Если интерфейс внешний, то LSR пересылает пакет, куда следует.

Когда пакет достигает предпоследнего транзитного участка на одном LSP, этот LSR может снять со стека меток последнюю запись и переслать пакет полезной нагрузки вообще без инкапсуляции.

Плоскость управления MPLS

Протоколы маршрутизации

В сетях MPLS активно используются внутренние протоколы маршрутизации IGP (Interior Gateway Protocol) — OSPF или IS-IS. Они служат для изучения топологии сети, выработки путей наименьших затрат и сбора информации для вычисления путей инжиниринга трафика. Для рассылки сведений о соединениях и метриках используются обычные объявления IGP, и эти сообщения дополняются информацией, описывающей линки (такой как пропускная способность).

Протокол LDP (Label Distribution Protocol)

LDP – это протокол на основе TCP, который может работать между соседними LSR в сети MPLS. Каждый LSR с помощью этого протокола рассылает метку, которую необходимо использовать при посылке на этот LSR пакетов с инкапсуляцией MPLS для окончательной доставки на некий IP-префикс. По мере того, как каждый LSR получает объявления от других LSR, он создает в FIB соответствующие записи, которые указывают, как отобразить метку полученного пакета (метку, которую этот LSP рассылал) в метку пакета, который он пересылает дальше (т.е. которую он узнал от других LSP).

В результате использования LDP трафик пересылается по пути наименьших затрат, а инжиниринг трафика не поддерживается.

Протокол RSVP-TE (Resource Reservation Protocol with TE Extensions)

При использовании RSVP-TE операторы сети административно назначают атрибуты TE интерфейсам. К числу атрибутов TE относятся (но не ограничиваются ими): доступная полоса пропускания, зарезервированная полоса пропускания и административный цвет. Эти атрибуты TE массово рассылаются с помощью IGP, так что каждый узел в рамках домена IGP содержит идентичную копию базы данных состояния связей (Link State Database, LSDB) и базы данных инжиниринга трафика (Traffic Engineering Database, TED). LSDB описывает топологию IGP, а TED дополняет LSDB атрибутами связей TE.

Сетевым операторам нужны LSP, которые соответствовали бы определенным требованиям. Например, сетевой оператор может потребовать, чтобы LSP начинался с узла A, заканчивался в узле Z, резервировал 100 мегабит в секунду и проходил только по синим интерфейсам. Модуль расчета путей, расположенный в центральном контроллере – например, PCE (Path Computation Element) – или входном LSR, вычисляет путь, который соответствует всем ограничениям. Чтобы создать такой SR-путь, функция расчета путей обращается к LSDB и TED.

RSVP-TE – это протокол сигнализации, работающий непосредственно поверх IP. Он использует сообщение Path, чтобы сигнализировать о пути LSP, а сообщение Resv сообщает о резервировании сетевых ресурсов и подтверждает создание LSP. Сообщение Path содержит данные о запрошенном LSP (полоса пропускания и т.п.), а также объект ERO (Explicit Route Object), в котором перечисляются все узлы и связи, по которым должен пройти LSP. Сообщение Resv сообщает о зарезервированных ресурсах (полоса пропускания и т.п.) и содержит объект RRO (Record Route Object), который подтверждает путь LSP.

Каждый LSR выбирает метку, которую будет использовать для получения трафика по LSP. Он включает эту метку в отправляемое сообщение Resv. Поэтому каждый LSR может создать запись FIB для LSP, которая отображает метку, которую он рассылал, на метку, которую он получил.

RSVP-TE требует, чтобы в сети для каждого LSP поддерживалось состояние, а протокол является «протоколом мягких состояний». Это означает, что для поддержки LSP в активном состоянии необходим периодический обмен сообщениями Path и Resv.

Сегментная маршрутизация

Терминология

Домен SR – это непрерывный набор маршрутизаторов с поддержкой SR. Путь SR (т.е. LSP, просигнализированный по SR) осуществляет соединение в рамках домена SR. Путь SR может проходить по пути наименьших затрат IGP между его начальной и конечной точками. Также он может проходить по пути, вычисленном с помощью инжиниринга трафика.

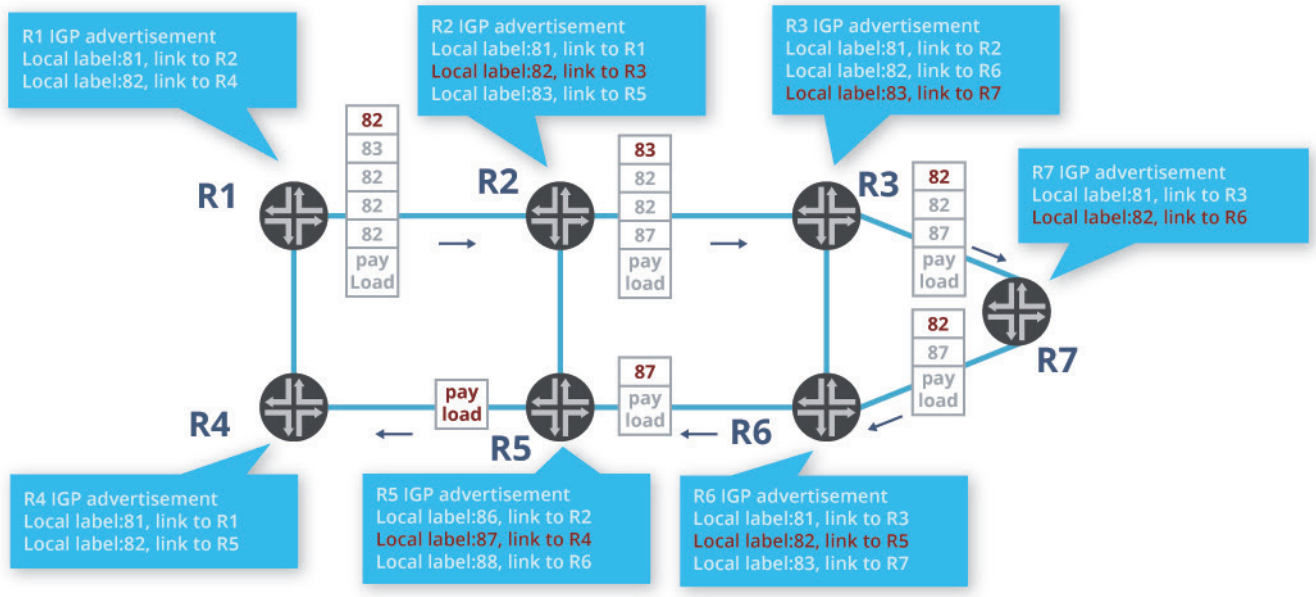
Путь SR содержит один или несколько сегментов, а сегмент содержит один или несколько транзитных участков маршрутизации. Рабочая группа SPRING предложила множество типов сегментов. Однако наиболее распространенными являются следующие из них:

- соседство (Adjacency);
- префикс (Prefix);
- массовая рассылка (Anycast);
- привязка (Binding).

Сегменты соседства отражают соседские отношения IGP между двумя маршрутизаторами. Как правило, такой сегмент состоит из одного транзитного участка, но их может быть и больше. Префиксные сегменты отражают путь наименьших затрат IGP между любым маршрутизатором и указанным префиксом. Префиксный сегмент содержит один или несколько транзитных участков маршрутизации. Сегменты массовой рассылки похожи на префиксные: они также отражают путь наименьших затрат IGP между любым роутером и указанным префиксом. Разница в том, что в этом случае префикс может объявляться из нескольких точек в сети. Префиксы привязки отражают туннели в домене SR. Такой туннель может быть другим путем SR, LSP, объявленным по LDP, LSP, объявленным по RSVP-TE, либо использовать любую другую инкапсуляцию.

Каждый сегмент идентифицируется идентификатором сегмента (Segment Identifier, SID). Идентификаторы SID для префиксных сегментов и сегментов массовой рассылки

Рис. 1. Сегменты соседства.



используются в рамках всего домена. Поэтому сетевые операторы назначают их с помощью процедуры, похожей на процедуру выделения частных IP-адресов (см. RFC 1918). Напротив, SID сегментов соседства и привязки имеют лишь локальное значение. Эти SID назначаются маршрутизаторами с поддержкой SR автоматически, без необходимости координации по всему домену.

Каждому SID сопоставлена метка MPLS. Как уже говорилось, метки MPLS имеют лишь локальное значение. Поэтому SID локального значения можно сопоставлять меткам MPLS напрямую. А вот SID доменного значения требуют специальной обработки.

Каждый SR-маршрутизатор резервирует ряд меток MPLS, называемый глобальным блоком SR (SR Global Block, SRGB). Например, маршрутизатор A может зарезервировать метки с 10 000 по 20 000, в то время как маршрутизатор B резервирует метки с 20 000 по 40 000. Оба маршрутизатора сопоставляют SID меткам MPLS, добавляя SID к наименьшему значению SRGB. Поэтому маршрутизатор A сопоставляет SID 1 метке MPLS 10 001, а маршрутизатор B сопоставляет тот же самый SID метке MPLS 20 001.

Передача SR

Когда входной SR-маршрутизатор получает пакет, он назначает ему класс FEC и инкапсулирует пакет с помощью стека меток MPLS. Затем он передает пакет следующему транзитному участку, связанному с данным FEC.

Стек меток MPLS отражает путь SR, связанный с данным FEC. Каждая запись в стеке меток соответствует сегменту пути SR.

На рисунке 1 маршрутизатор R1 поддерживает путь SR до R4. Этот путь SR содержит пять сегментов соседства, начинающихся на маршрутизаторах R2, R3, R7, R6 и R5. Входной LSR (R1) создает стек меток с одной записью на каждый сегмент соседства. Потом R1 передает пакет R2, где начинается первый сегмент соседства. R2 обрабатывает внешнюю запись стека меток, снимает ее и пересылает пакет R3. Каждый следующий LSR вниз по потоку повторяет процедуру до тех пор, пока пакет не достигнет R4.

На рисунке 2 маршрутизаторы R1-R6 поддерживают путь SR до R7. Этот путь SR содержит один префиксный сегмент, которому соответствует SID 7. Рассмотрим путь от R4 до R7.

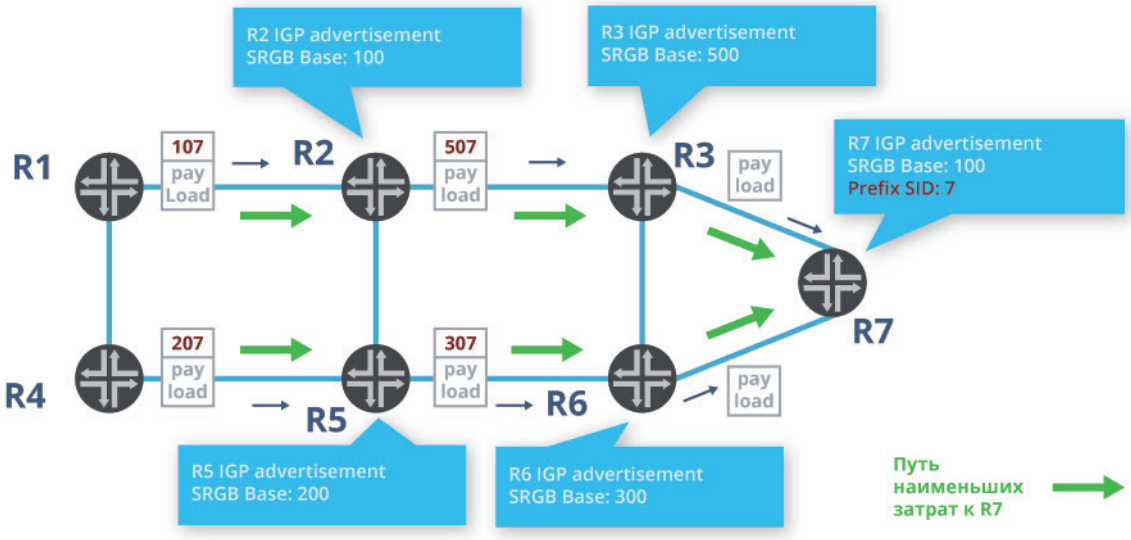
Входной маршрутизатор (R4) создает стек меток, содержащий ровно одну запись, которая представляет префиксный сегмент (т.е. путь наименьших затрат IGP) между R4 и R7. Эта запись в стеке меток содержит метку, соответствующую SID 7. Чтобы вычислить эту метку, R4 добавляет значение SID (7) к базе SRGB, сообщенной следующим транзитным участком R5 (на нашем рисунке это 200). В результате мы получим 207. И, наконец, R4 передает пакет R5.

R5 обрабатывает метку. Для этого он определяет маршрутизатор на пути наименьших затрат IGP, ведущий к R7 (в нашем примере это R6). Затем R5 заменяет метку на значение, которое в R6 соответствует SID 7 (т.е. 307). После этого он передает пакет R6. R6 повторяет процедуру, и пакет прибывает на R7.

На рисунке 3 маршрутизатор R1 поддерживает рассчитанный с помощью инжиниринга трафика путь SR до R4 через R7. Этот путь SR содержит два префиксных сегмента. Один префиксный сегмент соответствует пути наименьших затрат IGP от R1 до R7, а второй соответствует пути наименьших затрат IGP от R7 до R4.

Входной LSR (R1) создает стек меток с одной записью на каждый префиксный сегмент. Он вычисляет внутреннее значение метки, прибавляя SID R4 (4) к базе SRGB R7 (300). Он вычисляет внешнее значение метки, прибавляя SID R7 (7) к базе SRGB R2. После этого R1 передает пакет R2. Все

Рис. 2. Один префиксный сегмент.



маршрутизаторы ниже по течению обрабатывают пакет так, как описано в предыдущем примере, и пакет прибывает на R4.

Расширения IGP для сегментной маршрутизации

Каждый SR-маршрутизатор выделяет SID и метку:

- каждому префиксному сегменту или сегменту массовой рассылки, который заканчивается на этом маршрутизаторе;
- каждому сегменту соседства или привязки, который начинается на этом маршрутизаторе.

Сделав это, он создает запись RIB (Routing Information Base – прим. ред.) для каждого из вышеперечисленных сегментов и вносит записи RIB в FIB.

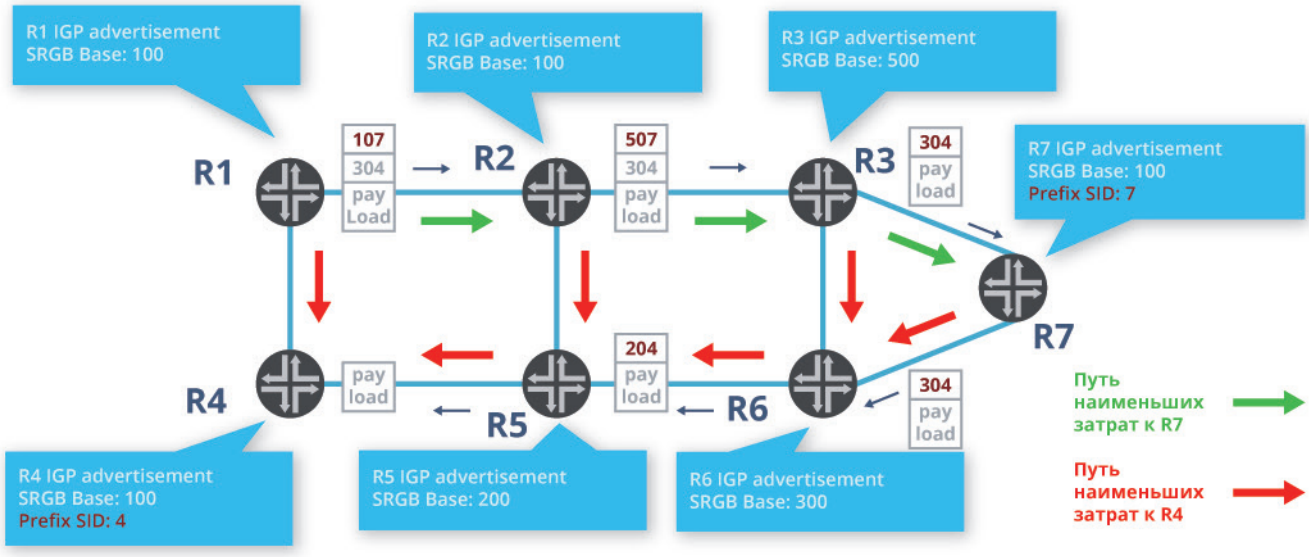
Далее SR-маршрутизатор рассылает с помощью IGP следующую информацию:

- свои характеристики SRGB;
- все префиксные сегменты и сегменты массовой рассылки, которые заканчиваются на этом маршрутизаторе;
- все сегменты соседства или привязки, которые начинаются на этом маршрутизаторе.

IGP рассылает эти данные, в дополнение к ранее упоминавшимся атрибутам связи TE, по всему домену IGP. Поэтому каждый узел в рамках домена IGP содержит идентичную копию базы данных состояния связей (LSDB) и базы данных инжиниринга трафика (TED). LSDB описывает топологию IGP, включая SID и данные SRGB, а TED дополняет LSDB атрибутами связей TE.

Когда рассылка завершается, каждый узел в рамках

Рис. 3. Инжиниринг трафика с использованием префиксных сегментов.



домена IGP создает по две записи RIB для каждого префиксного сегмента и сегмента массовой рассылки, который не заканчивается на этом узле. Первая запись RIB указывает локальному устройству обрабатывать весь входящий трафик IP, направляемый на этот префикс, следующим образом:

- добавить на самый верх стека MPLS запись, чья метка соответствует данному SID;
- переслать пакет на следующий транзитный участок по пути наименьших затрат IGP, ведущему к конечной точке сегмента.

Вторая запись RIB указывает локальному устройству обрабатывать весь входящий трафик MPLS, чья самая внешняя метка соответствует сегменту, следующим образом:

- заменить внешнюю метку, пересчитав ее для SRGB следующего транзитного участка;
- переслать пакет на следующий транзитный участок по пути наименьших затрат IGP, ведущему к конечной точке сегмента.

Расчет путей

Функция расчета путей рассчитывает пути SR. Получив набор ограничений TE, функция расчета путей выдает на выходе стек меток MPLS, образующих путь SR, который соответствует ограничениям. Чтобы создать такой SR-путь, функция расчета путей обращается к LSDB и TED.

Функция расчета путей может располагаться в центральном контроллере или, наоборот, распределяться по входным LSR.

Анализ

LDP и RSVP-TE – сквозные протоколы сигнализации, которые создают состояния передачи для каждого LSP в LSR. Поскольку LDP и RSVP-TE поддерживают все необходимые состояния передачи в LSR, то LSP, обработанный по LDP или RSVP-TE, может быть представлен одной записью стека MPLS.

Напротив, методика SR перемещает часть состояний передачи (хотя только часть) из сети в пакет. Путь SR представлен стеком меток, где каждая запись соответствует сегменту на пути SR. Поэтому сеть поддерживает достаточно состояний, чтобы пересылать пакеты от входного сегмента к выходному, а пакет – достаточно, чтобы пересылаться от сегмента к сегменту.

Перенеся информацию о состояниях из сети в пакеты, SR снижает требования к памяти для LSR и объем вычислений, необходимый для поддержания состояний. А ведь проблема памяти и объема вычислений остается актуальной, несмотря на то, что в последнее время маршрутизаторы стали оснащаться большим количеством вычислительных ресурсов и памяти, а также на доработки протокола RSVP-TE и его реализаций.

Еще более важно то, что перенос информации о состояниях из сети в пакеты устранило потребность в сквозном протоколе сигнализации. Да, для работы SR требуется IGP и модуль расчета путей, но протокол сигнализации, подобный LDP или RSVP-TE, больше не нужен.

Однако ряд дополнительных функций RSVP-TE зависит от сквозной сигнализации и состояний LSP в сети. Это и резервирование полосы пропускания, и обнаружение ошибок, и быстрая перемаршрутизация.

В RSVP-TE функция расчета путей может быть распределена по входным LSR, даже для случая, когда ограничения TE содержат резервирование полосы пропускания. Это возможно, так как в RSVP-TE каждый LSR поддерживает состояние для каждого своего LSP. По этому состоянию можно вычислить остаток пропускной способности на каждом интерфейсе с поддержкой RSVP и разослать эту информацию по IGP. Таким образом, каждый узел в IGP поддерживает LSDB и TED с достаточным количеством информации для работы функции расчета путей.

В SR такого механизма нет. Поэтому на случай, когда ограничения TE включают в себя полосу пропускания, функция расчета путей должна быть централизованной и располагаться в контроллере, обладающем глобальной информацией о распределении полосы пропускания.

В RSVP-TE сквозные механизмы сигнализации также реализуют функциональность OAM (Operation, Administration, Maintenance – процессы, необходимые для управления, администрирования и обслуживания системы – прим. ред.). При отказе соседского сеанса RSVP-TE маршрутизатор LSR выше по течению от места сбоя сигнализирует входному LSR, заставляя запустить процесс восстановления. Если LSR выше по течению от места сбоя настроен соответствующим образом, он может также запустить локальные процедуры восстановления.

В SR восстановление выглядит сложнее. Если сбой происходит на входном сегменте, то какой-либо механизм OAM вне SR обнаруживает сбой и информирует модуль расчета путей. Последний запускает процедуры восстановления и пересчитывает путь SR между входным SR и выходным. Локальные процедуры восстановления для SR теоретически возможны, но стандарта на этот счет пока нет.

Если сбой происходит не в конечной точке сегмента, SR вынужден полагаться на внешние механизмы восстановления. Например, если сбой произошел посреди префиксного сегмента, то SR должен использовать IGP, чтобы обнаружить сбой, разослать изменения топологии и вычислить новый путь наименьших затрат IGP до конечной точки сегмента. В этом примере можно развернуть TI-LFA, чтобы снизить зависимость от сходимости IGP.

Источник: IETF Journal (https://www.ietfjournal.org/segment-routing-cutting-through-the-hype-and-finding-the-ietfs-innovative-nugget-of-gold/)

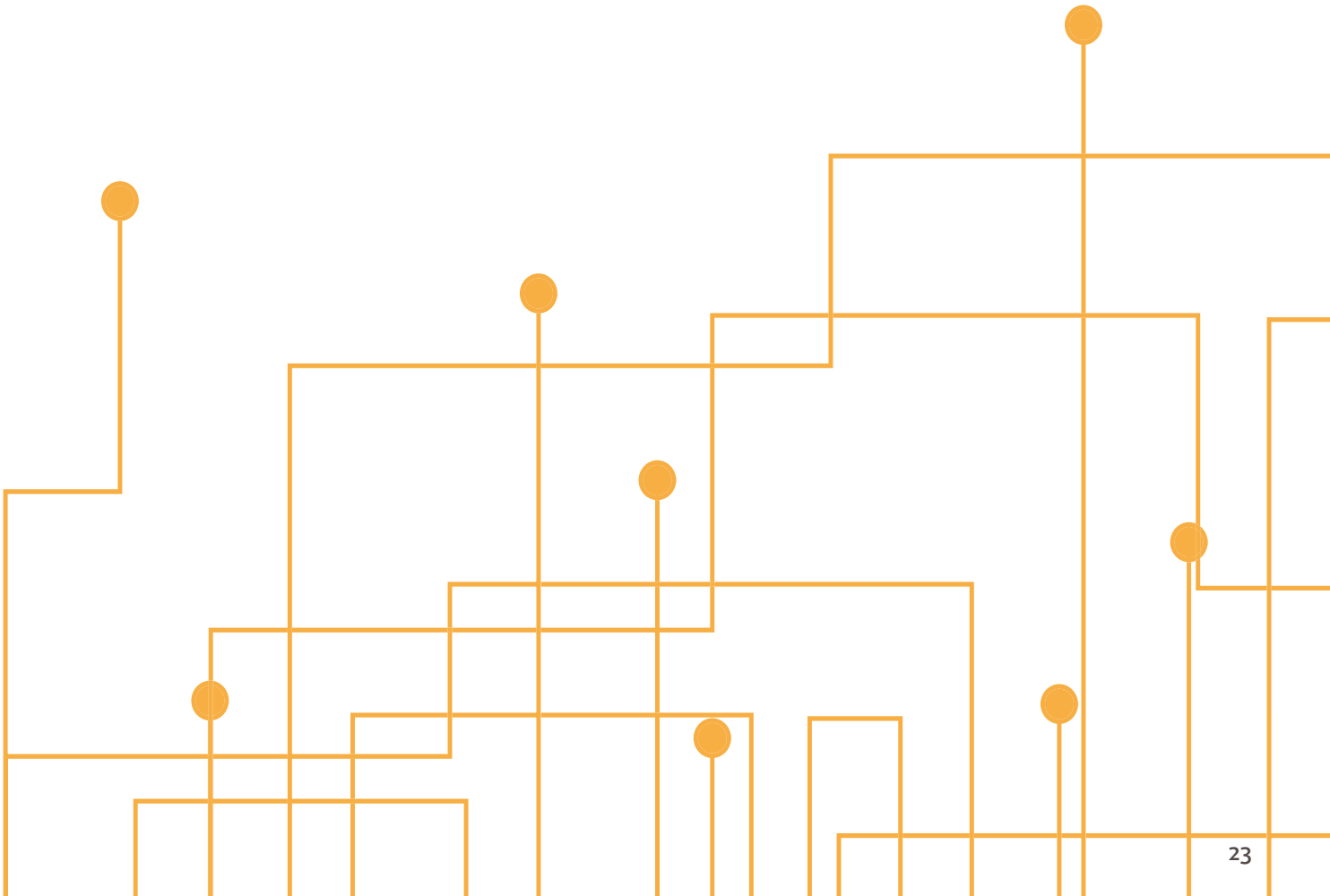
Выводы

SR поддерживает инжиниринг трафика, в то же время снижая количество состояний, которые должна обрабатывать сеть. Во многих случаях SR делает ненужными протоколы сигнализации MPLS (такие как LDP и RSVP-TE). Потому IETF следует продолжить разработку SR.

В частности, IETF следует продолжить стандартизацию расширений IGP для SR, а также расширений BGP, которые могут потребоваться для того, чтобы SR могла выйти за пределы IGP. Крайне необходимо проработать ключевые функции сети, такие как OAM и возможности передачи энтропии для разрешения выборов ECMP. Кроме того, поставщики сетевого оборудования и сетевые операторы должны совместными усилиями разработать прототипы и поэкспериментировать с SR для того, чтобы IETF получила обратную связь, на основе которой можно будет усовершенствовать SR и подготовить ее к крупномасштабному внедрению.

Весьма вероятно, что сетевые операторы начнут постепенно развертывать SR в ближайшие несколько лет. По мере внедрения сообщество SR будет набирать практический опыт, стандарты SR будут дорабатываться, а практика внедрения – совершенствоваться. Кроме того, сетевые операторы выявят сценарии, для которых SR хорошо подойдет, а также ситуации, где лучше использовать LDP и RSVP-TE.

По этим причинам, а также для поддержки большой базы инсталляций, IETF и производители сетевого оборудования должны продолжать доработку и поддержку LDP и RSVP-TE на том же уровне усилий, на котором они сейчас развивают SR.



А вы готовы к 5G-слайсингу?

Ченгиз Алаэттиноглу (Cengiz Alaettinoglu), технический директор, Packet Design

Один из интереснейших аспектов мобильной архитектуры 5G – это так называемый *слайсинг*, или сегментация сети (англ. network slicing), который наверняка найдет применение за пределами мобильных сетей, в сетях фиксированной связи либо за пределами мобильных и стационарных сетей. В этой статье мы поговорим о том, что такое 5G-слайсинг, о его особенностях и преимуществах для провайдеров услуг, базовых технологиях и быстродействии.

Что такое 5G-слайсинг?

Сетевой сегмент, или слайс (англ. network slice, буквально – «ломтик сети») – это динамически созданная логическая сквозная сеть с оптимизированной топологией, созданная под определенную задачу – для конкретного класса обслуживания или конкретного клиента. Оператор мобильной сети сможет «нарезать на ломтики» сетевые ресурсы (маршрутизаторы и каналы) вместе с вычислительными ресурсами и ресурсами хранения (для NFV и облачных приложений) и выделять их сервисам. Эта технология появилась в рамках ориентированного на сотовую связь проекта 3GPP (3rd Generation Partnership Project (http://www.3gpp.org/)).

Оркестровка и быстродействие сетевого слайсинга

Две привлекательных особенности сетевого слайсинга – это оркестровка (orchestration) и изолированные гарантии производительности. Оркестратор может «нарезать» сеть на ломтики (или слайсы, англ. *slice*) вместе с вычислительными ресурсами и ресурсами хранения и запустить сетевой сервис в отдельном ломтике. Изолированные гарантии производительности обеспечивают, что один слайс не может влиять на производительность другого слайса. При большом числе ломтиков это будет трудной задачей. Один сетевой слайс может обеспечивать критически важные сервисы (такие как реагирование в чрезвычайных ситуациях), другой – обслуживать традиционных пользователей сотовой связи, третий – быть выделен для устройств интернета вещей, а четвертый, скажем, будет выделен для пользователя MVNO (Mobile Virtual Network Operator)... и так далее.

Пользовательское оборудование (например, телефоны) будет способно подключаться к нескольким (до восьми) слайсам сети одновременно для доступа к различным сервисам. Эти слайсы могут управляться и оперироваться независимо друг от друга с помощью оркестратора, который настраивает ресурсы сети, вычислительные ресурсы и ресурсы хранения под потребности каждого слайса.

Еще один вариант применения слайсинга, который рассматривают провайдеры услуг фиксированной связи, представляет собой портал, где предприятия смогут заказывать и автоматически оркестровать полную сквозную сеть с гарантиями уровня сервиса. Такая сеть использует SD-WAN для подключения корпоративных мобильных пользователей и филиалов к центрам данных, эксплуатируемым предприятием, провайдером услуг или высокоуровневым провайдером, таким как AWS или Azure. Сегодня предприятия должны заказывать каждый фрагмент такой сети отдельно, сами придумывать способ создания объемлющей сети, которая будет их объединять, и надеяться на то, что гарантии производительности окажутся подходящими для используемых приложений.

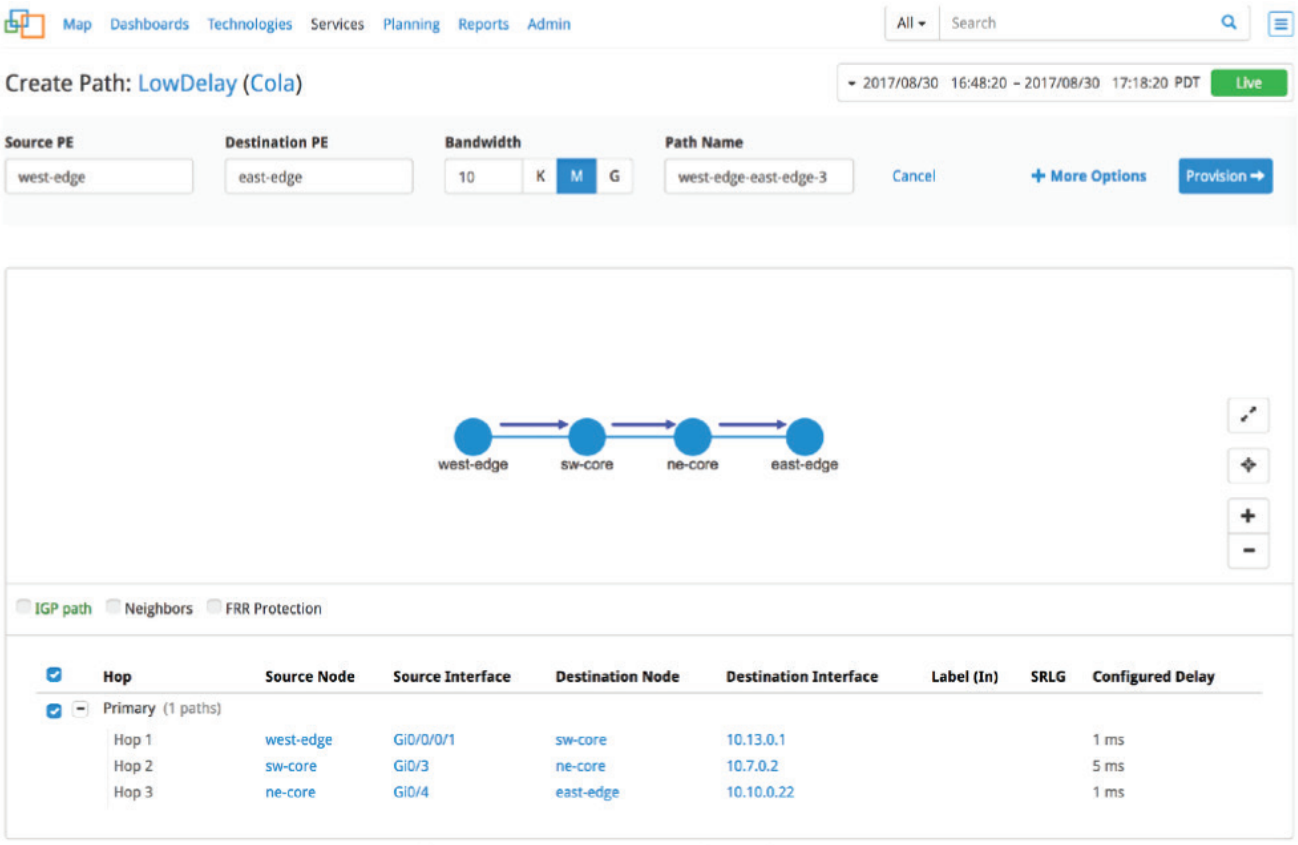
Гарантии производительности

Из последнего примера видно, что слайсинг действительно очень похож на облачное приложение с наложенной поверх него виртуальной сетью. Разница «только» в изолированных гарантиях производительности. Они подразделяются на две группы: жесткие и мягкие гарантии. Жесткие гарантии производительности было бы легко обеспечить с помощью TDM-сетей прошлых времен. Но те отошли в прошлое, вытесненные сетями IP/MPLS, поскольку статистическое мультиплексирование в технологии коммутации пакетов позволяет превышать лимит трафика, не жертвуя производительностью, и значительно устойчивее ведет себя при отказах узлов и линий.

Чтобы вообще можно было давать какие-то гарантии производительности, сетевая архитектура IP/MPLS использует

Сетевой сегмент, или слайс (англ. network slice, буквально – «ломтик сети») – это динамически созданная логическая сквозная сеть с оптимизированной топологией, созданная под определенную задачу – для конкретного класса обслуживания или конкретного клиента.

Рис. 1. Приложение SDN Path Provisioning App позволяет автоматизировать создание сервисных путей.



дифференцированные сервисные биты в составе пакетов, дисциплину пересылки пакетов через маршрутизаторы, а также механизм политик на краю сети, чтобы обеспечить соответствие трафика профилю (например, чтобы сервис с высоким приоритетом не захватывал больше ресурсов, чем ему положено). Для жестких гарантий производительности, возможно, самым многообещающим решением будет политика ускоренной пересылки в сочетании с механизмом политик на краю сети. Однако ограничение трафика на краю сети означает отсутствие возможности превышения лимитов трафика. Для критически важных сервисов, где необходим своевременный ответ с жесткими гарантиями, это, возможно, единственный вариант.

Сценарий применения слайсинга

Рассмотрим потенциальное применение слайсинга в реальном мире. У Packet Design есть европейский заказчик, занимающийся передачей электроэнергии. Эта компания управляет сетью электропередачи национального масштаба, соединяя поставщиков энергии с потребителями путем включения и отключения различных «переключателей» в сети. Компания фактически содержит рынок, на котором цена электроэнергии колеблется в зависимости от спроса и предложения. Неудивительно, что безопасность для этого заказчика очень важна: территорию компании от возможных террористических атак охраняют военные.

Всякий раз, когда возникает сбой передачи электроэнергии, в том числе из-за плохой погоды, необходимо сразу же найти новый источник энергии, чтобы избежать обесточивания целых городов. Если альтернативный источник найти не удается, можно отрубить питание сталелитейного завода (по предварительному соглашению) и направить ток в города. Принятие решения и переключение сети должно произойти в течение 600 миллисекунд во избежание перебоев. Сегодня этот заказчик содержит собственную вычислительную сеть, так как ее работа критически важна. Если реализовать правильный набор гарантий производительности, можно будет переключиться на использование сетевого слайса. Это дало бы сетевому оператору новый источник дохода, а заказчику позволило бы сократить затраты – т.е. обе стороны оказались бы в выигрыше.

На пути к широкому внедрению слайсинга

Хотя формального определения еще нет, мы ожидаем, что мягкие гарантии производительности будут более совместимы со статистическим мультиплексированием пакетов и позволят хотя бы немного превышать лимиты по трафику. Возможно, это удастся реализовать с помощью более мягких механизмов политик на краю. Превышение лимитов оборачивается лучшими экономическими показателями для сетевых операторов, так как очень мало сервисов/слайсов работает все время с пиковой нагрузкой, и уж точно не одновременно в разных слайсах.

Что касается оптимальной топологии для сетевых ломтиков, лучшими кандидатами будут VPN уровней 2 и 3. Судя по всему, для реализации слайсинга новые сетевые протоколы не понадобятся. Разумеется, это не помешает инженерам разрабатывать новые решения для того, чтобы облегчить развертывание, мониторинг, обнаружение сетевых слайсов и управление ими. Очень интересным нам представляется VPN+ (https://tools.ietf.org/html/draft-bryant-rtgwg-enhanced-vpn-00), предложение, которое распространяет VPN уровня 3 на 5G-слайсинг. Кстати говоря, приложение SDN Path Provisioning App (https://www.packetdesign.com/products/explorer-sdn-path-provisioning/) в составе Packet Design Explorer Suite, совместно с оркестратором от одного из наших партнеров (см. Packet Design partners with Ciena BluePlanet (https://www.packetdesign.com/press/packet-design-partners-cienas-blue-planet-division-advance-sdn-management-orchestration-network-operators/) и NEC/Netcracker (https://www.packetdesign.com/press/packet-design-necnetcracker-partner-advance-sdn-service-providers/)), может оказывать услуги VPN+ с путями низкой загрузки с сегментной маршрутизацией между динамически создаваемыми VRF для VPN уровня 3.

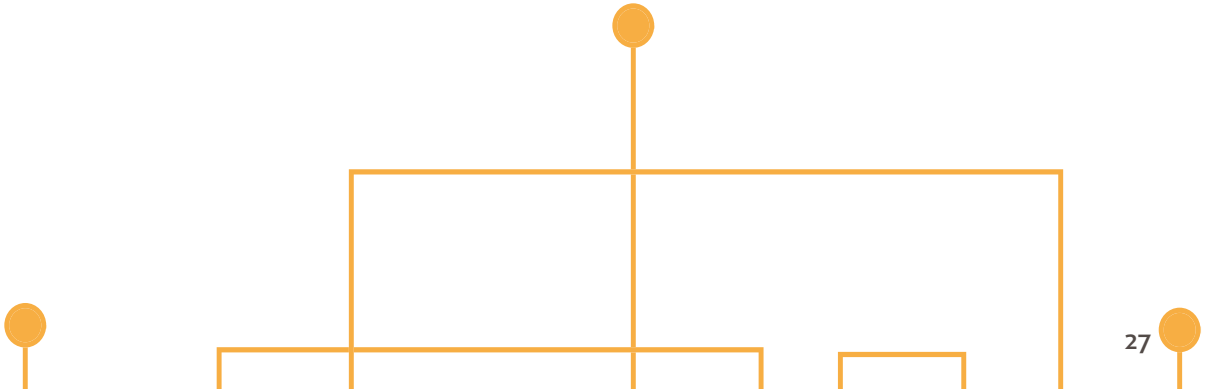
Еще одна технология, о которой часто говорят в связи со слайсингом, это Flex-Ethernet. Flex-Ethernet может создать субканал по тому же самому физическому носителю, но с собственной дисциплиной времени работы. Некоторые проприетарные расширения Flex-Ethernet убирают коммутацию пакетов при хранении и пересылке, так что получается в каком-то смысле аналог TDM, гарантирующий высокие уровни сервиса. Однако, как уже упоминалось, такая система не масштабируется и непроизводительно расходует ресурсы сети.

Рабочая группа IETF Deterministic Networking (DetNet) также работает над детерминистскими путями данных, которые позволяют лимитировать уровень потери данных, задержки и ее вариации (jitter), а также повысить надежность.

Однако эта группа слишком ориентирована на коммутацию пакетов и мыслит в основном в направлении очередей, политик на краю сети, временнОго планирования и предотвращения сбоев. DetNet может резервировать ресурсы по потокам и обеспечивает, возможно, лучшие жесткие гарантии по сравнению с простой дифференциацией IP-сервисов.

Роль аналитики в слайсинге

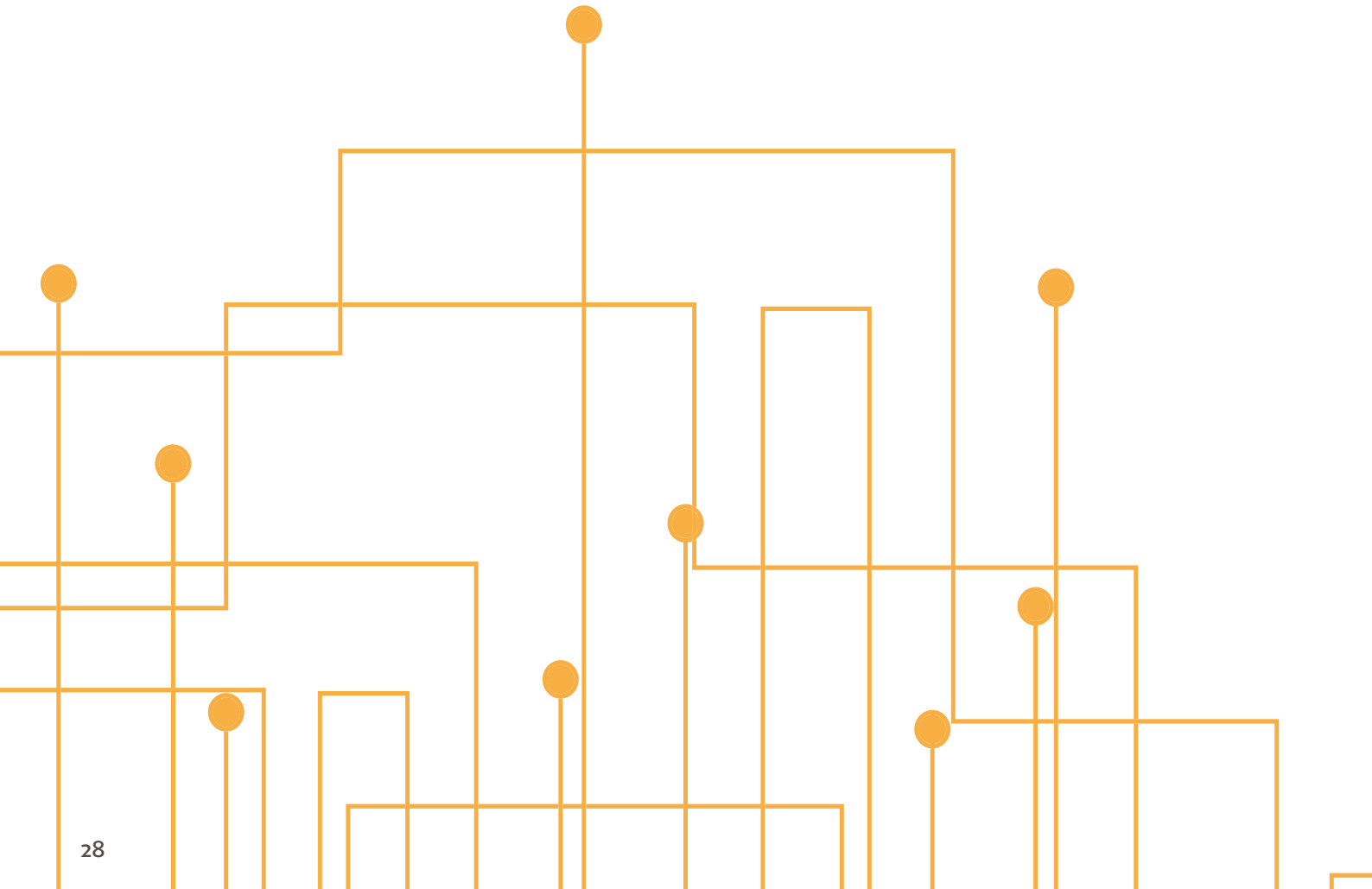
Зачем же нам нужна аналитика? Для того, чтобы управлять сетью с жесткими, мягкими гарантиями и «лучшим из возможного» (best-effort guarantees), сетевым операторам необходимо отслеживать топологию, трафик и производительность каждого слайса, а также агрегированной базовой сети. Например, если какой-то слайс перегружен, сетевой оператор должен как-то узнать, можно ли нарастить ему ресурсы



(и насколько), не жертвуя работой других слайсов. Для этого оператору нужно знать топологию слайса, его пути и объемы трафика по каждому классу пересылки, показатели задержки/вариации/потерь, а также то, какие еще слайсы используют те же самые ресурсы базовой сети и на какую величину. Возможно, единственным вариантом решения проблемы станет перенаправление части путей слайса на недогруженные участки базовой сети.

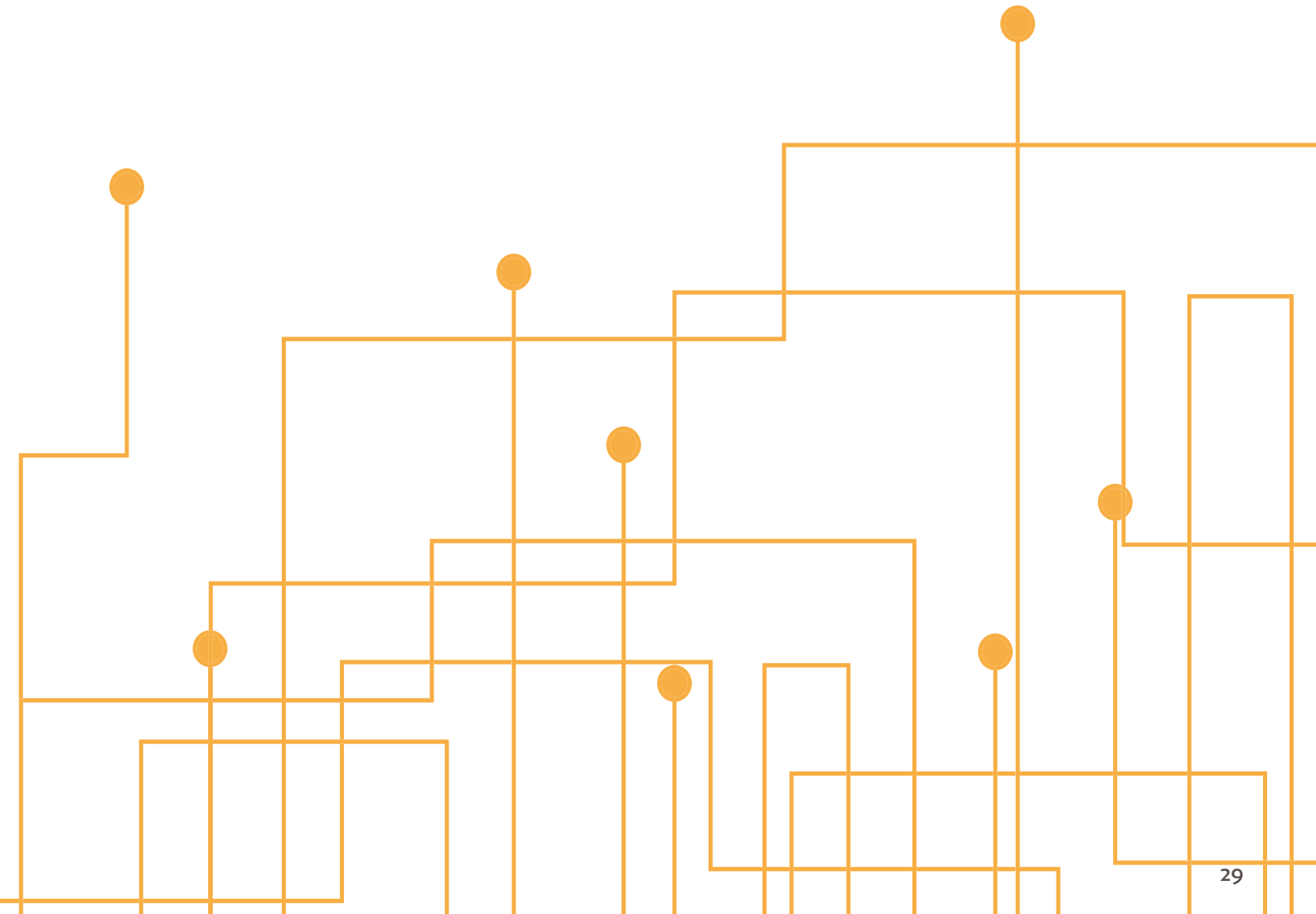
Хорошая система аналитики должна предоставлять все эти данные и содержать SDN-приложение для создания путей, которое будет прокладывать такие пути автоматически, но так, чтобы при этом критически важные слайсы не выходили за границы «безопасной/собственной» сети как в IP/MPLS, так и в нижележащей топологии оптических каналов. Чтобы упростить эту задачу, уже ведутся работы по стандартизации управления и оркестровки слайсинга (Network Slicing Management and Orchestration, MANO).

Еще одна важная проблема касается развертывания технологии в зрелых системах. Нетрудно вообразить себе развертывание с нуля, где каждый сервис будет локализован в том или ином слайсе сети. Но сегодня сети уже передают трафик, относящийся к нескольким классам обслуживания, с различными политиками пересылки и политиками на краю. Как нарезать на слайсы зрелую сеть? Поскольку вопрос этот прозвучал почти как риторический, ответим в том же стиле: очень, очень аккуратно. Разумеется, с помощью аналитики топологии/трафика/производительности иногда возможно оценить сетевые потребности старых приложений, перенести их в «старый» слайс и добавить к нему новые слайсы.



Заключение

Слайсинг – это новая, увлекательная технология, которая может принести провайдерам услуг новые доходы. В то же время создание слайсов и управление ими немыслимо без серьезной аналитики. Поэтому вопрос стоит так: готовы ли вы и ваша сеть к слайсингу?



Виртуализация центров обработки данных

Андрей Робачевский

Рецепт успешного развития современных центров обработки и хранения данных (ЦОДов), похоже, ясен – виртуализация. Но какая из многочисленных технологий наиболее эффективна в конкретной ситуации? Это и вопрос инвестиций, и обучения персонала – какая из них будет доминировать на рынке через пять лет?

Постоянно растущие требования к сокращению операционных затрат, времени развертывания услуг и бизнес-приложений, а также к их масштабируемости определяют тенденцию развития IT-инфраструктуры предприятий. Достижение этих требований возможно путем виртуализации - абстрагирования вычислительных ресурсов, инфраструктуры хранения данных и сети от физической инфраструктуры. При этом виртуализация серверов, хранения данных и сети могут быть обеспечены на различных уровнях компьютерных систем или сети.

Такой подход, совместно с тенденцией аутсорсинга IT-услуг, определяет архитектуру и развитие современных центров хранения и обработки данных (ЦОД/ЦХОД): виртуализация и концентрация.

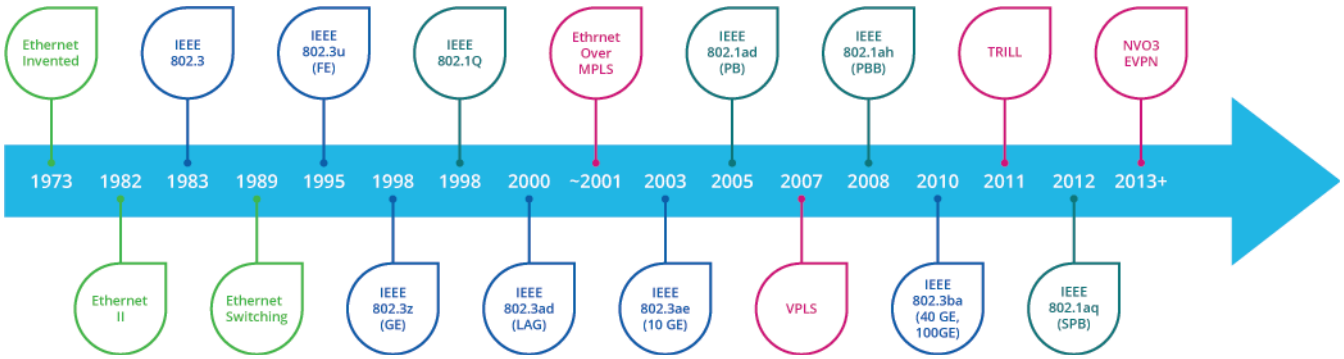
С сетевой точки зрения, ключевым требованием к архитектуре многоклиентного ЦОДа – ЦОДа, обслуживающего сотни, а то и тысячи арендаторов ресурсов, - является изоляция трафика, так что трафик одного арендатора недоступен никакому другому арендатору. Другим требованием является изоляция адресного пространства, благодаря чему разные арендаторы могут использовать одно и то же адресное пространство в разных виртуальных сетях. Изоляция трафика и адресного пространства достигается путем присвоения каждому арендатору одной или нескольких виртуальных сетей; при этом трафик из одной виртуальной сети может обмениваться с другой виртуальной сетью только через ограниченные контролируемые точки - например, через маршрутизатор или шлюз безопасности.

Однако для решения вопроса масштабируемости в современных ЦОДах необходимо удовлетворить дополнительное требование – возможность быстрого размещения (и перемещения) вычислительных ресурсов внутри дата-центра. Вычислительные ресурсы в этом случае представлены виртуальными машинами (ВМ). Задача состоит в том, чтобы иметь возможность перемещать виртуальные машины без привязки их адресации к адресной структуре сети самого дата-центра и такими образом, - без необходимости переадресации.

Более того, виртуальная машина может быть перенесена с одного сервера на другой в режиме реального времени - без необходимости остановки и перезапуска в новом местоположении и прерывания вычислительной задачи. Это может быть необходимо для оптимизации нагрузки и определенных параметров качества/производительности.

Ключевым требованием для живой миграции является то, что ВМ сохраняет параметры состояния сети в своем

Рис. 1. Эволюция технологии Ethernet[1].



новом местоположении, включая его IP- и MAC-адреса. Сохранение MAC-адресов может потребоваться, поскольку любое изменение MAC-адресов виртуальных машин, возникающих в результате перемещения, будет видимым для виртуальной машины и, таким образом, может привести к неожиданным последствиям. Сохранение IP-адресов после перемещения необходимо для предотвращения обрыва существующих транспортных соединений (например, TCP) и их повторного запуска.

Одним из основных элементов достижения требований современных ЦОДов является сетевая инфраструктура. Эта инфраструктура также виртуализирована, предоставляя каждому клиенту-арендатору собственную сетевую инфраструктуру второго или третьего уровня, полностью изолированную от инфраструктуры других клиентов и самого ЦОДа. Обмен данными между этими сетями возможен, но происходит через контрольные узлы, например, маршрутизаторы или шлюзы, в соответствии с политикой, определенной клиентом.

Технологии виртуализации

Технологии виртуализации стремительно развиваются, но суть их примерно одинакова и сводится к созданию виртуальных сетей поверх существующей инфраструктуры, так называемых сетевых оверлеев. Опорная инфраструктура может использовать различные технологии передачи данных, пример - IP или MPLS. В контрольной плоскости для создания топологии и построения маршрутов могут использоваться различные протоколы маршрутизации, например, IS-IS, OSPF или BGP.

Оверлеи также могут существенно различаться – это могут быть виртуальные сети уровня 3 – IP VPN, или же уровня 2, эмулирующие Ethernet.

Помимо удовлетворения требований, о которых я рассказал в начале статьи, технологии виртуализации решают две основные задачи: в плоскости контроля - как осуществить обмен маршрутизационной информацией и привязать топологию оверлея к опорной инфраструктуре, а в плоскости передачи данных – вопрос инкапсуляции изначального трафика (пакетов IP или фреймов Ethernet) в данные для передачи в опорной сети.

Мы начнем разговор с оверлейных IP-сетей, а затем перейдем к сетям уровня 2. Применительно к ЦОДам виртуализация сетей Ethernet является наиболее популярной, так как позволяет упростить архитектуру клиентских сетей и эффективно обеспечить мобильность ВМ. С момента появления технология Ethernet не перестает развиваться, охватывая все более широкий класс задач, как с точки зрения скоростей передачи, так и с точки зрения виртуализации. Живучесть и эволюция этой технологии впечатляют, см. рис 1.

BGP/MPLS IP VPNs

VPN-сети, основанные на технологии BGP/MPLS уровня 3, позволяют поддерживать сложные топологии и решают проблему масштабируемости обычных VPN точка-точка, таких как IPSec VPN. Для добавления нового сайта в BGP/MPLS VPN требуется одно изменение на граничном маршрутизаторе провайдера (PE), к которому подключен клиентский маршрутизатор. Маршрутизаторы клиента (CE) обмениваются маршрутной информацией с маршрутизатором PE с использованием BGP и не знают о существовании VPN. VPN создаются между PE-маршрутизаторами, которые используют BGP с мультипротокольным расширением (MP-BGP, RFC 4760, https://datatracker.ietf.org/doc/rfc4760/) для обмена префиксами клиентов и связанных с ними меток VPN.

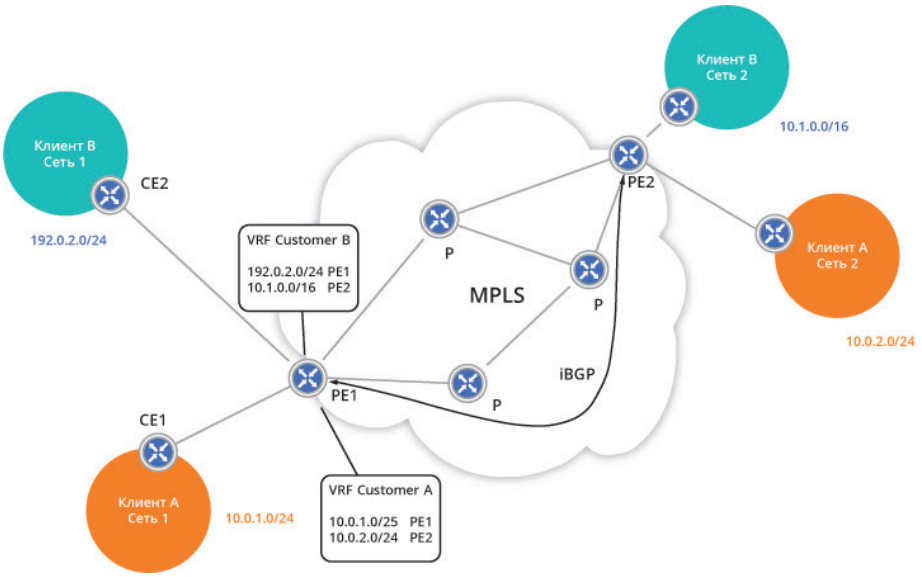
На граничных маршрутизаторах провайдера PE для каждого подключенного клиента создается виртуальная таблица маршрутизации с использованием технологии VRF (VPN Routing and Forwarding). VRF позволяет одновременно использовать несколько экземпляров таблицы маршрутизации на одном и том же маршрутизаторе. Поскольку экземпляры таблицы маршрутизации независимы, одинаковые или перекрывающиеся IP-адреса могут использоваться без конфликта друг с другом. Благодаря этому достигается полная изоляция и независимость виртуальных сетей клиентов друг от друга.

На рис. 2 представлена схема реализации виртуальных сетей для двух клиентов - A и B.

Граничный маршрутизатор PE1 обслуживает две виртуальные таблицы VRF для двух клиентов. Обмен маршрутной информацией между граничными маршрутизаторами осуществляется с помощью iBGP. Однако возникает проблема - каким образом различить адресные пространства различных VPN, которые к тому же могут пересекаться?

Использование MP-BGP позволяет обмениваться маршрутами из различных «семейств адресов». Для IPv4-адресов VPN вводится понятие «семейство адресов VPN-IPv4».

Рис. 2. Реализация IP-оверлея с помощью технологии BGP/MPLS.



Адрес VPN-IPv4 представляет собой 12-байтовое поле, состоящее из 8-байтового селектора маршрута (Route Distinguisher, RD) и 4-байтного адреса IPv4. Если несколько VPN используют один и тот же префикс адреса IPv4, PE преобразуют их в уникальные префиксы адресов VPN-IPv4. Это гарантирует, что если даже один и тот же адрес используется в разных VPN, маршрутизаторы могут обмениваться совершенно разными маршрутами для этого адреса, по одному для каждого VPN.

Для передачи трафика используется технология мультипротокольной коммутации меток (MPLS). Трафик следует через предопределенный путь коммутации меток (LSP), который является однонаправленным туннелем между двумя маршрутизаторами PE. Каждый из PE является т.н. граничным маршрутизатором меток (Label Edge Router, LER), который инкапсулирует IP-пакеты VPN в пакеты MPLS с соответствующими метками.

Коммутацию меток осуществляют опорные маршрутизаторы провайдера (P). Эти маршрутизаторы не подключены к каким-либо CE-маршрутизаторам и также не содержат маршруты VPN-IPv4, а только внутренние маршруты к другим маршрутизаторам P и PE. P-маршрутизаторы проверяют только самую верхнюю (внешнюю) LSP-метку и заменяют ее новой меткой LSP перед пересылкой пакета.

VPN-сети уровня L3 масштабируются до тысяч VPN и миллионов префиксов. BGP/MPLS IP VPN применяются в крупных корпоративных центрах обработки данных. Потенциальным ограничением для применения этой технологии в ЦОДах является практичность использования, особенно для доступа к серверам или гипервизорам. По мере роста ЦОДа трудно решаемыми становятся вопросы масштабируемости (поддержка полносвязной топологии MPLS), а также сходимости и согласованности BGP.

Оверлеи EVPN

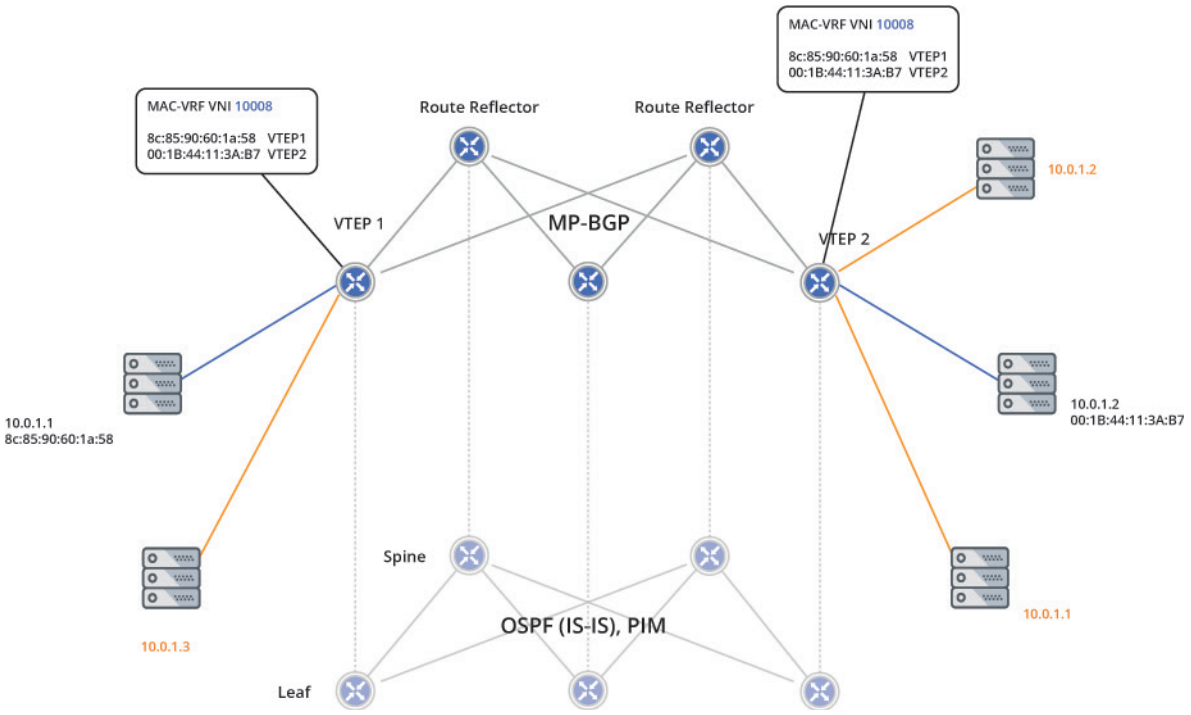
Виртуальные сети Ethernet - EVPN (RFC7348, https://datatracker.ietf.org/doc/rfc7348) - представляют собой эмулированную услугу L2, в которой каждый арендатор имеет свою собственную сеть - Ethernet-оверлей через общую IP-инфраструктуру.

Так же, как в случае BGP IP VPN, для обмена информацией о MAC-адресах и идентификаторов виртуальных сетей участвующих устройств (ими могут быть хосты, маршрутизаторы или коммутаторы) используется MP-BGP. Граничный маршрутизатор PE обслуживает отдельную таблицу MAC-VRF для каждого клиента, обеспечивая требуемую изоляцию.

Поскольку EVPN обеспечивает эмуляцию сети уровня L2, возникает вопрос обработки широковещательного трафика типа бродкаст, малтикаст и «неизвестный юникаст» - так называемого трафика BUM (Broadcast, multicast, unknown unicast). Одним из вариантов является репликация такого трафика на входном PE. Более оптимальным подходом, однако, является использование малтикаста. В этом случае для каждого широковещательного домена создаются соответствующие малтикаст-группы.

Технология EVPN позволяет значительно увеличить утилизацию инфраструктуры. Дело в том, что в натуральных L2-сетях используются два механизма, которые плохо масштабируются. Первый, о котором мы только что говорили, связан с обслуживанием BUM-трафика. В сетях Ethernet для этого используется механизм «наводнения» (flooding), когда пакеты рассылаются по всей связующей инфраструктуре. Второй – механизм исключения «петель» - циклического обмена маршрутной информацией L2. Для этого в сетях Ethernet используется протокол STP (Spanning Tree Protocol), с помощью которого в сети отключаются отдельные каналы связи для предотвращения петель. В

Рис. 3. Поддерживающая IP-инфраструктура и оверлей VXLAN.



результате часть инфраструктуры не используется. Применение L3 поддерживающей инфраструктуры позволяет избежать этой проблемы, а использование технологии ECMP (https://en.wikipedia.org/wiki/Equal-cost_multi-path_routing) обеспечивает максимальную утилизацию.

Хотя изначально архитектура EVPN была разработана с использованием MPLS для передачи данных, на сегодня различные типы инкапсуляции данных нашли свое применение.

Наиболее используемыми являются технологии Virtual Extensible LAN (VXLAN, RFC7348, https://datatracker.ietf.org/doc/rfc7348/), Network Virtualization using Generic Routing Encapsulation (NVGRE, RFC7637, https://datatracker.ietf.org/doc/rfc7637/) и MPLS over Generic Routing Encapsulation (GRE, RFC4023, https://datatracker.ietf.org/doc/rfc4023/). В настоящее время разрабатывается более общий и гибкий подход к построению оверлеев (туннелей), так называемый GENEVE (Generic Network Virtualization Encapsulation, https://datatracker.ietf.org/doc/draft-ietf-nvo3-geneve/). Давайте рассмотрим некоторые из них.

Virtual eXtensible Local Area Network (VXLAN)

VXLAN является методом построения туннелей уровня L2 на опорной IP-инфраструктуре (L3) на основе EVPN. Эта технология документирована в RFC7348 (https://datatracker.ietf.org/doc/rfc7348/). VXLAN использует инкапсуляцию IP/UDP. 24-разрядный идентификатор сегмента VXLAN, или VNI, обеспечивает до 16 миллионов сегментов VXLAN для изоляции и сегментации трафика, в отличие от 4000 сегментов, доступных в VLAN. Каждый из этих сегментов представляет собой уникальный широковещательный домен уровня 2, изолированный от сегментов других клиентов. Таким образом, использование клиентами пересекающегося адресного пространства, как IP, так и MAC, не представляет проблемы.

Опорная IP-инфраструктура использует внутренний протокол маршрутизации, такой как OSPF или IS-IS, для определения внутренней топологии и обеспечения связности. Также используется протокол PIM для создания малтикаст-инфраструктуры, необходимой для обработки BUM-трафика.

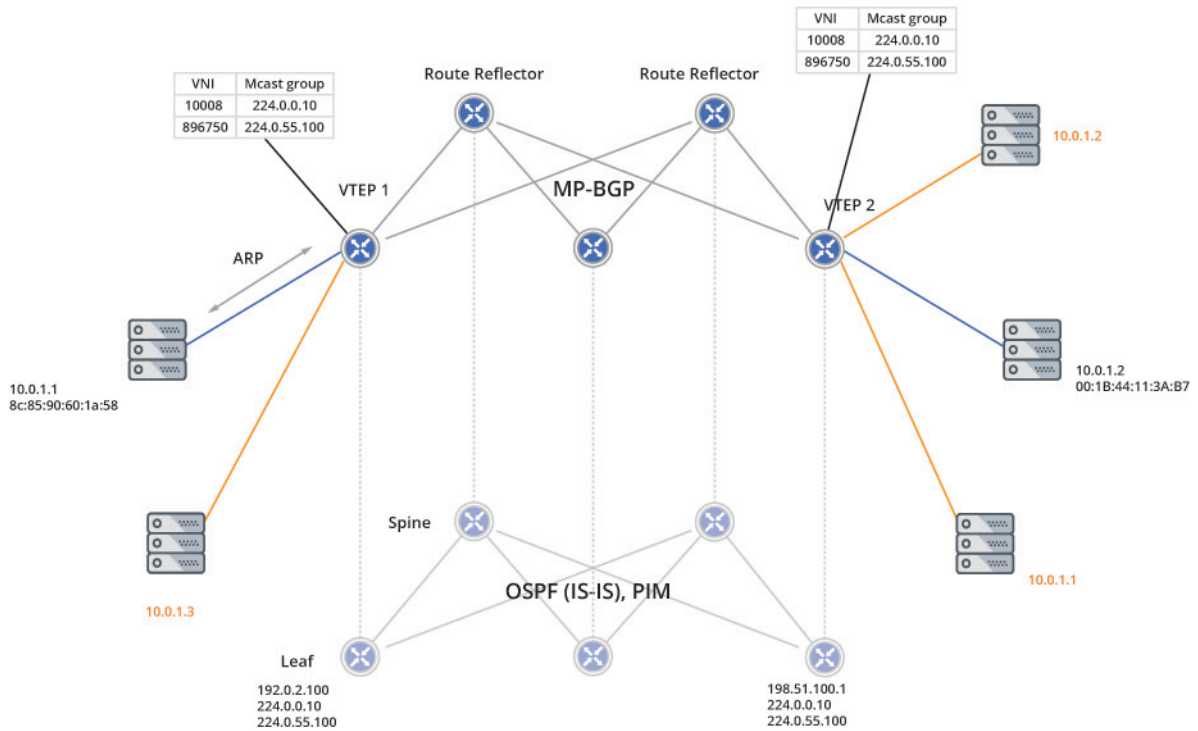
В плоскости передачи данных используется инкапсуляция IP/UDP для туннелирования трафика через опорную инфраструктуру. Граничные маршрутизаторы обычно также выполняют роль т.н. VTEP (VXLAN tunnel endpoint), обеспечивая инкапсуляцию и декапсуляцию трафика.

Каждое из устройств VTEP обменивается информацией о MAC-адресах подключенных к ним устройств и связанных с ними VNI с другими VTEP с помощью MP-BGP. При этом для каждого сегмента с уникальным VNI создается собственная таблица MAC-VRF, благодаря чему обеспечивается изоляция адресных пространств различных VXLAN.

Схема поддерживающей IP-инфраструктуры и оверлея VXLAN представлена на рис 3. Голубым цветом показана VXLAN VNI 10008. Хотя различные виртуальные сети используют ту же инфраструктуру, их адресное пространство и топология полностью независимы. Например, устройства «оранжевой» VXLAN используют то же адресное пространство, что и «голубой».

Поддержка BUM-трафика показана на рис 4. Хотя устройства виртуальной сети для получения информации о MAC-адресах других устройств традиционно используют такие методы, как ARP, эти запросы не проходят дальше устройства VTEP. Как мы только что видели, все устройства

Рис. 4. Реализация обслуживания BUM-трафика в VXLAN.



VTEP виртуальной сети используют MP-BGP для получения этой информации. Поэтому, когда приходит запрос ARP, VTEP уже знает ответ, который он и отправляет запрашивающему устройству.

Однако BUM-трафик не исчерпывается запросами ARP. Для его обработки для каждого широковещательного домена (другими словами, для каждой VXLAN) создается отдельная малтикаст-группа. Этот процесс невидим для устройств VXLAN и реализуется с помощью таких протоколов, как PIM в опорной IP-инфраструктуре.

Network Virtualization using Generic Routing Encapsulation (NVGRE)

Сетевая виртуализация с использованием универсальной инкапсуляции маршрутизации, или NVGRE, позволяет создавать виртуальные сети уровня 2 поверх опорной инфраструктуры уровня 3. Это достигается путем туннелирования Ethernet-фреймов внутри IP-пакета по физической сети. Подобно VXLAN, NVGRE поддерживает идентификатор 24-битного сегмента или идентификатор виртуальной подсети (VSID), предоставляя до 16 миллионов виртуальных сегментов, которые могут однозначно идентифицировать сетевой сегмент и соответствующее ему адресное пространство отдельного клиента.

Технология NVGRE очень похожа на VXLAN. Основное различие заключается в том, что заголовок NVGRE содержит необязательное поле идентификатора потока. В многосвязных опорных сетях маршрутизаторы и коммутаторы могут анализировать заголовок пакетов и использовать это поле вместе с VSID для трафик-инжиниринга, хотя для этой функции требуются дополнительные возможности аппаратного обеспечения.

Как и в случае с VXLAN, в стандарте NVGRE не указывается метод обнаружения достижимости конечной точки. Для этого применяются решения типа EVPN.

Программно-конфигурируемая сеть

Рассмотренные нами технологии построения оверлеев решают основные задачи, стоящие перед современным многоклиентным ЦОДом. Однако их использование связанно с дополнительными накладными расходами. Например, затраты, связанные с инкапсуляцией, использованием различных протоколов в опорной инфраструктуре и оверлее. Сосуществование функций управления и обработки и передачи данных, а также их распределённый характер усложняют архитектуру и обслуживание сети.

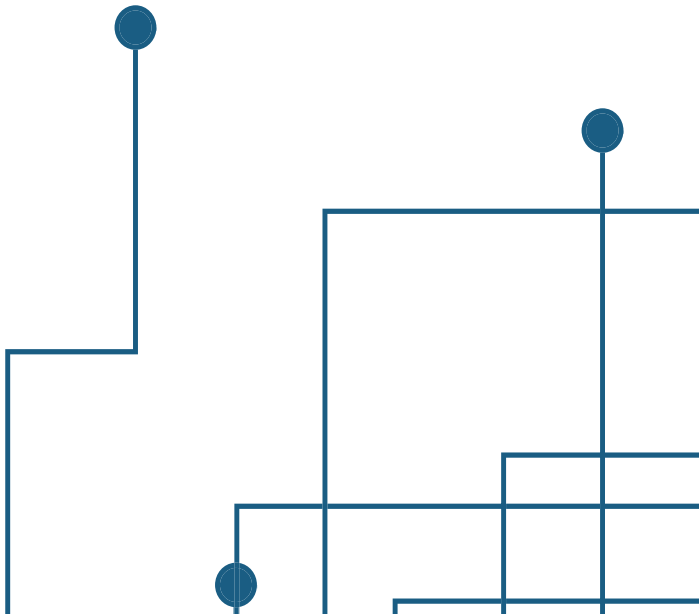
Концепция SDN (Software Defined Network, программно-конфигурируемая сеть) предусматривает передачу управляющих функций центральному устройству - т.н. контроллеру, таким образом заменяя традиционную распределенную модель маршрутизации на централизованную. Соответственно, и процесс управления сетью, включающий создание маршрутов, является не чем иным, как программированием сети в целом.

Концепция SDN (Software Defined Network, программно-конфигурируемая сеть) предусматривает передачу управляющих функций центральному устройству - т.н. контроллеру, таким образом заменяя традиционную распределенную модель маршрутизации на централизованную. Соответственно, и процесс управления сетью, включающий создание маршрутов, является не чем иным, как программированием сети в целом.

С использованием SDN можно существенно упростить процесс распознавания топологии и создания маршрутов. В отличие от традиционной сети, в том числе и оверлейной, где маршрутизация - это распределенный итеративный процесс, при котором рабочая топология сети «вычисляется» совместно всеми устройствами, в SDN - это не что иное, как программа моделирования сети с заданными параметрами.

В случае использования управляющего центра расчет новой топологии производится исходя из знания о всей сети в целом. Мы также можем задать необходимую топологию следующего состояния. Наконец, поскольку создание новой топологии - это чисто вычислительный процесс, он может быть выполнен значительно быстрее.

На первый взгляд, SDN является привлекательной технологией в достижении целей виртуализации, стоящими перед современным ЦОДом, по крайней мере, в архитектурном плане. Однако успех внедрения той или иной технологии зависит от многих факторов: знание и опыт использования сотрудниками, существующая инфра-



структура, доступность решений и оборудования, их цена и производительность и т.п.

Увидим ли мы дальнейшее развитие оверлеев на базе традиционных технологий и существующей инфраструктуры или централизованные программные решения станут более популярными – покажет будущее.

Ссылки

1. Диаграмма любезно предоставлена Грегом Хэнкинсом (Greg Hankins), Nokia.

Сервисы глобальной Сети через призму BGP

Александр Венедюхин,
Фонд развития сетевых технологий «ИнДата»

Введение

В современной глобальной Сети всё большее значение приобретает устойчивость к атакам, использующим особенности инфраструктуры, в том числе административные. Несмотря на то, что Сеть всё ещё остаётся единой, регулярно возникают вопросы, которые касаются взаимодействия больших групп автономных систем, объединённых государственной принадлежностью, между собой, то есть проекции географии в виртуальное пространство.

С точки зрения типичного пользователя, глобальная сеть Интернет состоит из наборов сервисов: веб, электронная почта, приложения в смартфоне и так далее. Надёжность и доступность этих сервисов напрямую зависит от того, как доставляются пакеты данных, то есть от организации маршрутов в Сети. Рассматривая маршрутную информацию в разрезе сервисов, можно увидеть особенности, которые не ясны из анализа только таблиц BGP. Например, транзитные автономные системы (AS) можно ранжировать по количеству сервисов, для которых они являются транзитными. Также содержательные результаты даёт анализ маршрутов и транзитных AS применительно к трафику защищённых протоколов уровня приложений, прежде всего, TLS.

В настоящем исследовании рассматриваются основные сервисы, к которым мы относим веб (в том числе HTTPS/TLS) и электронную почту. Из распространённых пользовательских сервисов не рассматриваются мессенджеры и приложения социальных сетей, предназначенные для смартфонов.

Административная принадлежность автономных систем, определяющих структуру маршрутов в Сети, позволяет говорить о географических особенностях маршрутов, основная принадлежность которых - российский сегмент Интернета (Рунет).

Методика

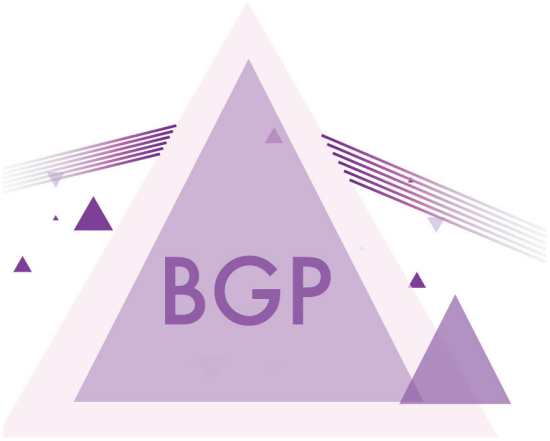
Для обозначения сервисов используются доменные имена. Доступ непосредственно по IP-адресам является довольно редким решением, тем более среди пользователей веб-сайтов и электронной почты. Поэтому исходными данными для построения списков сервисов послужили доменные имена (второго уровня) в российских доменных зонах (.ru, .su, .рф). Второй характеристикой, обозначающей наличие на заданном узле нужного сервиса, является доступность по хорошо известному (для данного приложения) номеру порта.

Рассмотрим доменное имя ididb.ru в качестве примера. Получим для этого имени из DNS значение A-записи: 62.76.121.136 (это IP-адрес). Установив с данным IP-адресом соединение по 80/tcp и отправив HTTP-запрос, мы можем определить, что здесь отвечает *веб-сервер* (то есть размещён веб-сайт). Снова воспользуемся DNS и определим имя почтового сервера для ididb.ru, запросив MX-запись: mail.ididb.ru. Определим для mail.ididb.ru значение A-записи (62.76.121.133) и проверим, что на обращения по 25/tcp по этому адресу отвечает *почтовый сервер* (то есть размещён сервис электронной почты). Итак, использовав в качестве отправной точки имя ididb.ru, мы нашли связанные с этим именем два сервиса (веб и почта), имеющие различные IP-адреса.

Таблица 1. Общие показатели по числу IP-адресов (2017, 2018 гг.)

Показатель/зона	.ru	.su	.рф
Число уникальных **IP веб-узлов**			
Сентябрь 2017	275 539	19 310	38 416
Март 2018	275 200	18 915	38 173
Число уникальных **IP TLS-узлов (HTTPS)**			
Сентябрь 2017	185 171	15 217	29 910
Март 2018	198 137	16 261	32 795
Число уникальных **IP MX-узлов**, по всем зонам			
Октябрь 2017	100 522		
Март 2018	104 878		

При помощи анализа маршрутной информации в соответствие обнаруженным IP-адресам поставим автономную систему. Это будет та автономная система, которая анонсирует префикс, содержащий нужный IP-адрес, и при этом является оконечной AS (то есть представляет собой *назначение маршрута*; подробнее понятие оконечной AS определено ниже). Таким образом, каждому сервису сопоставляется автономная система и для такой автономной системы изучаются маршруты в глобальных таблицах BGP (Full View), полученных от нескольких источников. Кроме того, мы анализируем маршруты на доступных route-серверах крупнейших российских точек обмена трафиком (MSK-IX и DATAIX). Сервис и маршруты связаны через систему доменных имён (DNS).



Для веб-сайтов мы принимаем, что адресам с префиксом www. соответствует тот же адрес, что и доменному имени второго уровня (как показывает практика, это не вносит больших искажений в данные). Под «**веб-узлом**» далее мы понимаем узел, подключенный к глобальной Сети, который доступен для TCP-соединений на номер порта 80 и отвечает непустым результатом на HTTP-запрос GET. Веб-узлу соответствуют IP-адреса и имена доменов, связанные между собой DNS-записями. Отдельно рассматриваются веб-узлы, доступные по защищённому *протоколу TLS через 443/tcp*.

Таким образом, веб-узел характеризуется парой «доменное имя, IP-адрес». Нередки варианты, когда одному имени соответствует несколько IP-адресов. Чрезвычайно распространён и «обратный» случай, когда один IP-адрес соответствует большому количеству имён *различных* веб-сервисов. Эти особенности учитываются отдельно.

С **электронной почтой** связаны DNS-записи MX, в которых содержится имя сервера, принимающего сообщения для данного домена. (Запись также содержит приоритет данного сервера, но в рамках нашего анализа сервисов мы опрашиваем все указанные узлы вне зависимости от приоритета.) Предположим, что для test.ru указана MX-запись mail.test.ru. Это означает, что другие серверы, осуществляющие доставку, при обнаружении сообщения с адресом внутри test.ru (например, box@test.ru) будут пытаться доставить это сообщение серверу, на который указывает имя mail.test.ru. В отличие от веб-узлов, при отображении DNS-имён в IP-адресное пространство, соответствующее электронной почте, необходима «двухэтапная» схема: на первом этапе определяется имя MX-сервера, на втором - значение A-записи для этого имени (а не для исходного имени второго уровня). Рекомендациями (RFC 5321) допускается возможность доставки почты непосредственно по значению A-записи для базового домена (вершины зоны) при отсутствии в ней MX-записей. Однако такой вариант в нашем исследовании не рассматривается.

Итак, сервер электронной почты характеризуется IP-адресом и символьным именем, указанным в MX-записи. Один и тот же сервер может иметь не только множество различных IP-адресов и различных имён, он может обслуживать большое число доменных зон, в каждой из которых одно из его имён указано в значении MX-записи. Типичными для Рунета показателями является обслуживание десятков тысяч различных зон одной группой почтовых серверов. Мы считаем почтовым узлом подключенный к глобальной Сети сервер, принимающий TCP-соединения по номеру порта 25 и содержательно отвечающий на SMTP-команду EHLO.

Поддержка TLS (защищённый протокол) исследуемыми узлами определялась только в разрезе HTTPS (443/tcp), путём проведения начальной фазы установления TLS-соединения, в рамках которой от узла, в частности, принимался TLS-сертификат. TLS-сертификаты проверялись в рамках процедуры валидации, сходной с аналогичной процедурой веб-браузеров. Валидация позволяет оценивать потенциальную доступность защищённых веб-сервисов и сравнивать статистику по узлам с валидными/не валидными сертификатами.

Итак, список IP-адресов формируется в результате опроса DNS. Каждому IP-адресу из полученного списка при помощи анализа таблиц BGP сопоставляется содержащий его префикс (блок IP-адресов). По префиксу определяется автономная система. Анонсирующая данный префикс автономная система, которая является последней точкой маршрута, называется **оконечной автономной системой**. В рамках настоящего исследования предполагается, что оконечная автономная система, к которой принадлежит префикс, содержащий IP-адрес сервиса, содержит и этот сервис. Такой подход отражает реальное положение дел лишь с сетевой точки зрения: административно сервис может относиться к организации, не связанной напрямую с данной оконечной автономной системой. Типичный пример - размещение оборудования в дата-центре, где арендуется канал доступа. Тем не менее, сетевая принадлежность сервиса к той или иной автономной системе в большинстве случаев означает, что, как минимум, между администратором сервиса и администратором автономной системы существуют некоторые договорные отношения. Общие показатели по IP-адресам приведены в таблице 1, с разбивкой по зонам.

Мы выделяем российские автономные системы (и сервисы в них). Автономная система относится к российским на основании регистрационных данных, полученных из регистратуры RIPE NCC. При выделении номера автономной системы (и/или блоков IP-адресов) RIPE NCC требует предоставления документов, идентифицирующих юридическое или (в редких случаях) физическое лицо, выступающее в качестве администратора AS. На основании

Таблица 2. Число уникальных оконечных AS и соответствующих им маршрутов (октябрь 2017)

Период	Уникальные оконечные AS	Маршруты Full View	Маршруты с route-серверов
Октябрь 2017	3854 (Full View)	36 009	9330
Март 2018	3901 (Full View)	34 958	9224

этих документов определяются адреса юрисдикции администратора, которые позволяют говорить о географической принадлежности AS.

В таблице 2 указано общее число автономных систем, являющихся оконечными, то есть автономных систем, в которых находится узел с тем или иным интернет-сервисом. Соответствующие маршруты - это различающиеся хотя бы в одном «хопе» маршруты, которые ведут к автономным системам с сервисами. Так, разнообразие маршрутов, полученных с route-серверов, существенно меньше (в три с лишним раза), чем в глобальных таблицах BGP. Это объясняется достаточно очевидным соображением: маршрутизация в Интернете в целом устроена сложнее, чем маршруты между участниками пиринга на точках обмена трафиком. Таким образом, число различных AS, размещающих сервисы Рунета, измеряется тысячами. При этом существует концентрация ресурсов по AS (см. таблицу 3 ниже), соответствующим крупным хостинг-провайдерам, что, конечно, находится в полном соответствии с эвристическими представлениями об устройстве Сети.

В таблице 3 представлен рейтинг оконечных автономных систем, составленный по числу уникальных IP-адресов (IPv4), соответствующих узлу с тем или иным сервисом из исследуемых (веб, почта, TLS). Кодом RU обозначены российские AS. Так как это рейтинг по IP-адресам, он отражает политику распределения IP-адресного пространства между ресурсами конкретным провайдером: так, большое число веб-ресурсов, размещённых на узле с одним IP-адресом, будут в данном рейтинге учитываться как *один* узел.

Верхушка рейтинга из таблицы 3 – это иностранные AS, в частности, один из крупнейших мировых провайдеров Cloudflare. Причина такого положения дел в том, что многие сервисы, адресуемые российскими доменными именами, размещены на иностранных площадках. Данные иностранные AS нами не рассматриваются при анализе маршрутов: как указано выше, в таком случае мы не относим ресурс к полностью российским.

Необходимо отметить, что географическая принадлежность автономных систем - понятие в некотором роде условное: проекция политической карты мира в Интернет имеет свои особенности. Например, автономная система, зарегистрированная российским юридическим лицом, может фактически являться иностранной, в том смысле, что основные эксплуатируемые сети физически находятся за пределами России. Напротив, AS, отмеченная на основании регистрационных данных как иностранная, может действовать только на территории России, в интересах того или иного российского провайдера, являясь, таким образом, российской в сетевом смысле.

Возможны и более сложные случаи. Практический анализ маршрутной информации в разрезе сервисов полностью подтверждает такое положение дел. Тем не менее, наблюдение над данными географической привязки, проведённой описанным способом, и выборочная ручная проверка показывают, что способ этот достаточно точный, а главное - единственный универсальный. Других методов относительно незатратного массового и точного определения принадлежности автономных систем к определённой стране пока не существует.

Анализ маршрутной информации

Понятие автономной системы является базовым для определения маршрутов доставки пакетов данных в Интернете. Если рассматривать это понятие в разрезе протокола IP, то автономная система – это, фактически, набор маршрутизаторов, формирующих видимую для Интернета IP-сеть, находящуюся под единым управлением. Последний момент является определяющим: именно при помощи единого управления определяется то, как пакеты доставляются внутри сети. Это называется политикой маршрутизации. «Автономной» систему делает не то, что у неё есть собственная политика маршрутизации, а то, что данный «набор маршрутизаторов» взаимодействует с другими автономными системами, составляющими Интернет. Это ключевой момент - определение автономной системы является рекурсивным, т.е. нельзя корректно определить понятие AS, если нет других AS. Это выглядит контринтуитивно. Тем не менее, для глобальной Сети такое определение логично и обосновано. Почему? Потому что «единство управления маршрутизацией» возможно определить только с внешней точки зрения. Так, другие автономные системы, взаимодействующие с данной через пограничный узел, видят внутреннюю политику маршрутизации как *единую*. Данная оговорка очень важна: политика маршрутизации может не являться единой *внутри* AS, но в рамках исследования *внешних таблиц* BGP выявить это в общем случае невозможно. Другими словами: определение автономной системы – не столько техническое, сколько административное.

Основу взаимодействия автономных систем через протокол BGP составляет распространение информации о желании соседних AS доставлять пакеты в адрес того или иного IP-префикса. Под “соседними” здесь подразуме-

Таблица 3. Рейтинг оконечных AS, по числу уникальных IP с сервисами (октябрь 2017)

1	AS13335 (CLOUDFLARENET-AS)
2	AS29182 (ISPSYSTEM-AS)
3	AS24940 (HETZNER-AS)
4	AS197695 (AS-REGRU) RU
5	AS25535 (ASN-RUCENTER-HOSTING) RU
6	AS14061 (DOSFO - DigitalOcean)
7	AS49505 (SELECTEL) RU
8	AS16276 (OVH)
9	AS198068 (FASTNET)
10	AS9123 (TimeWeb-AS) RU

ваются такие AS, пограничные маршрутизаторы которых взаимодействуют непосредственно в смысле обмена маршрутами (это не означает, что маршрутизаторы «соединены напрямую»). Именно эту информацию, в конечном итоге, можно видеть в глобальных таблицах BGP. В рамках настоящего исследования мы используем таблицы BGP Full View, состав которых зависит от точки получения, а конкретно, от пограничного маршрутизатора, послужившего источником таблицы. Второй набор маршрутной информации - это данные, полученные с route-серверов точек обмена трафиком. Данный набор также имеет свои ограничения, определяющиеся, в основном, собственными пиринговыми политиками участников обмена. Маршрутизация в Интернете динамическая, поэтому на состав таблиц маршрутов влияет не только точка получения, но и время.

Всякое исследование BGP как *внешнего протокола обмена маршрутами* - это не более чем исследование BGP. Реальная связность Интернета существенно сложнее, кроме того, она постоянно меняется. Распространена ситуация, когда некоторые операторы организуют между собой канал, который не анонсируют наружу, то есть соответствующие автономные системы при взгляде через BGP Full View или route-серверы *не будут* учитываться как соседние, несмотря на то, что они обмениваются IP-пакетами, адресованными друг другу, напрямую. Тем не менее, на основании анализа таблиц BGP можно строить оценки, отражающие распределение адресного пространства между сервисами, а также свойства инфраструктуры, обеспечивающей доставку пакетов. И в этом ключе дополнение BGP-информации сведениями о сервисах, сформированными на основе проверки реальной связности (то есть под данным IP-адресом действительно доступен некоторый узел), оказывается весьма полезным, так как предоставляет ещё одну точку опоры при анализе связности и доступности.

Исследование интернет-сервисов на основании маршрутной информации подразумевает использование некоторого мгновенного слепка маршрутов. Предположим, что в физической связности Сети произошли какие-то существенные изменения, например, обрушился какой-то значимый канал связи. Это достаточно быстро приведёт к перестройке маршрутов, в результате, изменится и картина для интернет-сервисов. При этом выстроить статистику, точно учитывающую подобные эффекты, невозможно из-за непредсказуемости результата изменения глобальных таблиц.

Второй временной аспект связан с получением информации о соответствии имён и IP-адресов исследуемым сервисам. Построение исходного набора данных связано с опросом DNS и обнаружением сервисов на узлах: этот процесс, как отмечено выше, сам основан на характеристиках связности, которую предполагается исследовать. Например, отмечены следующие проблемы: некоторые авторитативные DNS-серверы (то есть серверы, на которых размещена доменная зона) имеют географические настройки и, соответственно, отвечают на запросы по-разному, в зависимости от того, к какому региону принадлежит источник запроса (определяется по IP-адресу); также DNS-сервер может отвечать разным составом значений A-записей в зависимости от текущей загрузки сервиса в целом. Крупные сервис-провайдеры используют anycast-адреса, что также вносит сложно устранимые искажения и в интерпретацию маршрутов, и в данные по соответствию имён и сервисов. Тем не менее, эти проблемы имеют локальный характер, и хоть и влияют на точность анализа, но не являются препятствием для построения статистики, учитывающей миллионы имён и сервисов. Эвристическая оценка погрешности, вносимой только что описанными факторами, составляет около 5%.

Транзит интернет-трафика и сервисы

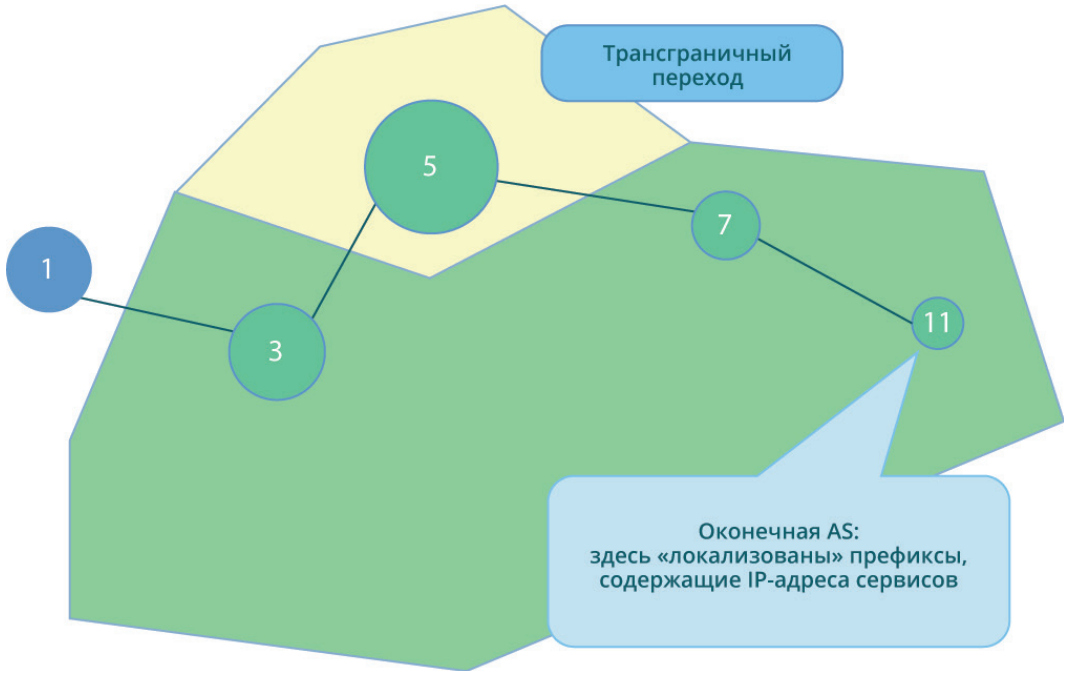
Неотъемлемой частью маршрутизации в Сети являются транзитные автономные системы, обеспечивающие доставку пакетов между различными AS. Без транзитных систем маршрутизация в Интернете невозможна. С точки зрения сервисов, важна концентрация маршрутов, ведущих к тем или иным сервисам, по транзитным автономным системам. В качестве иллюстрации в таблице 4 приведён топ-10 рейтинга российских транзитных AS по числу уникальных *имён* ресурсов, для которых данная AS является транзитной. Данный рейтинг построен не по IP-адресам, а по уникальным именам сервисов.

Данные из таблицы 4 демонстрируют некоторые особенности распределения транзита трафика по различным типам сервисов. Рейтинги в разных колонках существенно различаются. Этому есть несколько причин. Во-первых,

Таблица 4. Транзитные AS в разрезе концентрации сервисов (октябрь 2017)

	Веб		Почта		TLS
1	AS3216 (SOVAM-AS)	1	AS43048 (scrubbing-center, IBANK2.RU)	1	AS48287 (RU-SERVICE-AS)
2	AS12389 (ROSTELECOM-AS)	2	AS3216 (SOVAM-AS)	2	AS8920 (VTC-AS)
3	AS20485 (TRANSTELECOM)	3	AS20485 (TRANSTELECOM)	3	AS35000 (PROMETEY)
4	AS8359 (MTS)	4	AS39741 (ZRA-AS)	4	AS41066 (RTCOMM-SIBIR-AS)
5	AS31133 (MF-MGSM-AS)	5	AS12389 (ROSTELECOM-AS)	5	AS56534 (PIRIX-INET-AS)
6	AS20764 (RASCOM-AS)	6	AS197068 (QRATOR)	6	AS56694 (DHUB)
7	AS199599 (CIREX)	7	AS13238 (Yandex)	7	AS42244 (ESERVER)
8	AS28917 (Fiord-AS)	8	AS20764 (RASCOM-AS)	8	AS5537 (RU-CENTER-AS)
9	AS3267 (RUNNET)	9	AS206977 (CONTMP-AS)	9	AS5563 (URAL)
10	AS57724 (DDOS-GUARD)	10	AS8359 (MTS)	10	AS3216 (SOVAM-AS)

Рис. 1. Схема маршрута с трансграничным переходом.



разные сервисы, соответствующие одному имени, нередко находятся у разных провайдеров. Обычный пример: веб и почта. Веб-сервер может быть размещён у одного хостинг-провайдера, а почтовый сервер - принадлежать тому или иному массовому почтовому провайдеру. Так, существенный вклад вносит практика размещения сервиса почты на серверах «Яндекса» или Google. По состоянию на август 2017 года MX с именем mx.yandex.ru был указан для 227 тысяч имён в зоне .ru. Более того, существенное число имён вообще не адресуют веб-серверов, но используются для почты. По данным проекта Statdom.ru, таких имён в августе 2017 года было около 525 тысяч в зоне .ru. Что касается TLS, то этот протокол нами рассматривается в разрезе HTTPS, а безопасная версия используется далеко не для всех веб-узлов. Соответственно, выборка по TLS фактически означает сужение выборки всех веб-узлов на те хостинги, где поддерживается HTTPS.

Таблица 5. Число имён узлов с трансграничными переходами*

Период	Узлы, которым соответствует хотя бы один маршрут с трансграничным переходом		
	Веб	TLS	MX
	Все источники BGP:		
Сентябрь/октябрь 2017	544 326	484 750	296 438
	Route-серверы (MSK-IX, DATA-IX):		
Сентябрь/октябрь 2017	398 847	353 503	218 006
	Backbone Full View:		
Сентябрь/октябрь 2017	539 725	481 087	293 491
Март 2018	584 704	444 419	323 332

*— Приводится число уникальных доменных имён, которые указывают на узел, оказавшийся в маршруте с трансграничным переходом. Одному IP-адресу может соответствовать несколько имён.

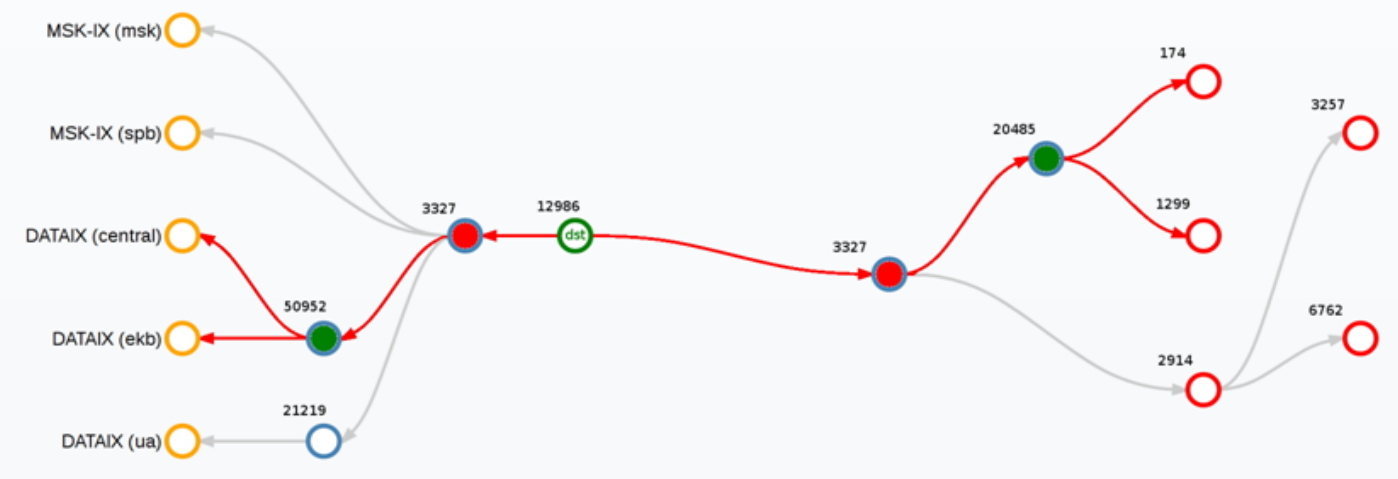
География сервисов и трансграничные переходы

Для российских сервисов и автономных систем мы рассматриваем трансграничные переходы маршрутов. Трансграничные переходы представляют интерес по следующей причине: содержащие их маршруты соответствуют сервисам, расположенным в российских автономных системах и адресуемых российскими доменными именами, однако при этом включают в себя неоправданные «двойные переходы» между разными странами.

Достаточно большая доля (около 30%) интернет-сервисов, адресуемых доменными именами в российских зонах, размещена на зарубежных площадках, которые используют адреса из иностранных оконечных AS. Так как в данном случае сам сервис находится в зарубежных сетях, обнаружение дополнительных трансграничных переходов не имеет смысла. Поэтому мы рассматриваем маршруты, у которых оконечная AS (с сервисом) является российской. В случае анализа маршрутов из BGP Full View, полученных от мировых backbone-операторов, первой AS в маршруте является иностранная - это источник маршрутной информации. Тем не менее, так как оконечная AS российская, маршрут должен прийти к одной из российских автономных систем.

Следуя типовому способу записи маршрутов, примем, что оконечная AS, содержащая сервис, находится в записи AS-PATH крайней справа. Тран-

Рис. 2. Пример практических маршрутов с трансграничным переходом (выделены красным).



зитными мы считаем все автономные системы, у которых в записи маршрута есть две соседние AS. Маршрут, ведущий к *российской оконечной* AS, относится к маршрутам с трансграничным переходом, если он содержит хотя бы одну иностранную (не российскую) транзитную автономную систему, с которой *слева* соседствует российская.

Трансграничные переходы определяются при последовательном разборе маршрута с сопоставлением каждой автономной системе флага «российская/не российская». В результате выявляются маршруты, в которых трафик дополнительно пересекает «виртуальную границу» - уже после того, как маршрут «пришёл» в российский сегмент.

В таблице 5 приведено распределение уникальных имён узлов с сервисами по типам сервисов, для маршрутов с трансграничными переходами. Максимальное число относится к веб-узлам: свыше 500 тысяч веб-узлов видны через маршруты с трансграничными переходами. При этом для маршрутов, полученных с route-серверов, число сервисов с трансграничными переходами в маршрутах несколько меньше.

На рис. 2 показана типичная ситуация, наблюдающаяся на практике: AS12976 является оконечной для нескольких веб-узлов, данная AS - российская, однако и на route-серверах (левая часть диаграммы), и в Full View backbone-операторов оконечной AS предшествует AS3227, которая не является российской. На данной диаграмме не все маршруты являются маршрутами с трансграничным переходом. Так, трансграничный переход (в нашей терминологии) *содержат* маршруты Full View, проходящие через российскую AS20485 (это российская AS), а другие маршруты, представленные справа, трансграничных переходов не содержат, так как все входящие в них AS являются иностранными (кроме оконечной). Аналогично, среди маршрутов route-серверов только один маршрут, проходящий через AS50952, содержит трансграничный переход (AS21219 является иностранной).

В таблице 6 представлены два рейтинга (TOP-10) автономных систем, которые «реализуют» трансграничный переход (см. описание ниже). Рейтинги отражают вариативность маршрутов, в которых данная AS является транзитной. О чём идет речь? Предположим, что у нас есть оконечные автономные системы AS65536 и AS65547, содержащие веб-узлы. Рассмотрим следующие маршруты:

1. AS65539, AS65538, AS65550, AS65536
2. AS65539, AS65538, AS65550, AS65547

(То есть два маршрута, различающихся только последней AS.)

Таблица 6. Рейтинг (топ-10) AS, встретившихся в маршрутах с трансграничным переходом, по числу уникальных маршрутов (для веб-узлов)

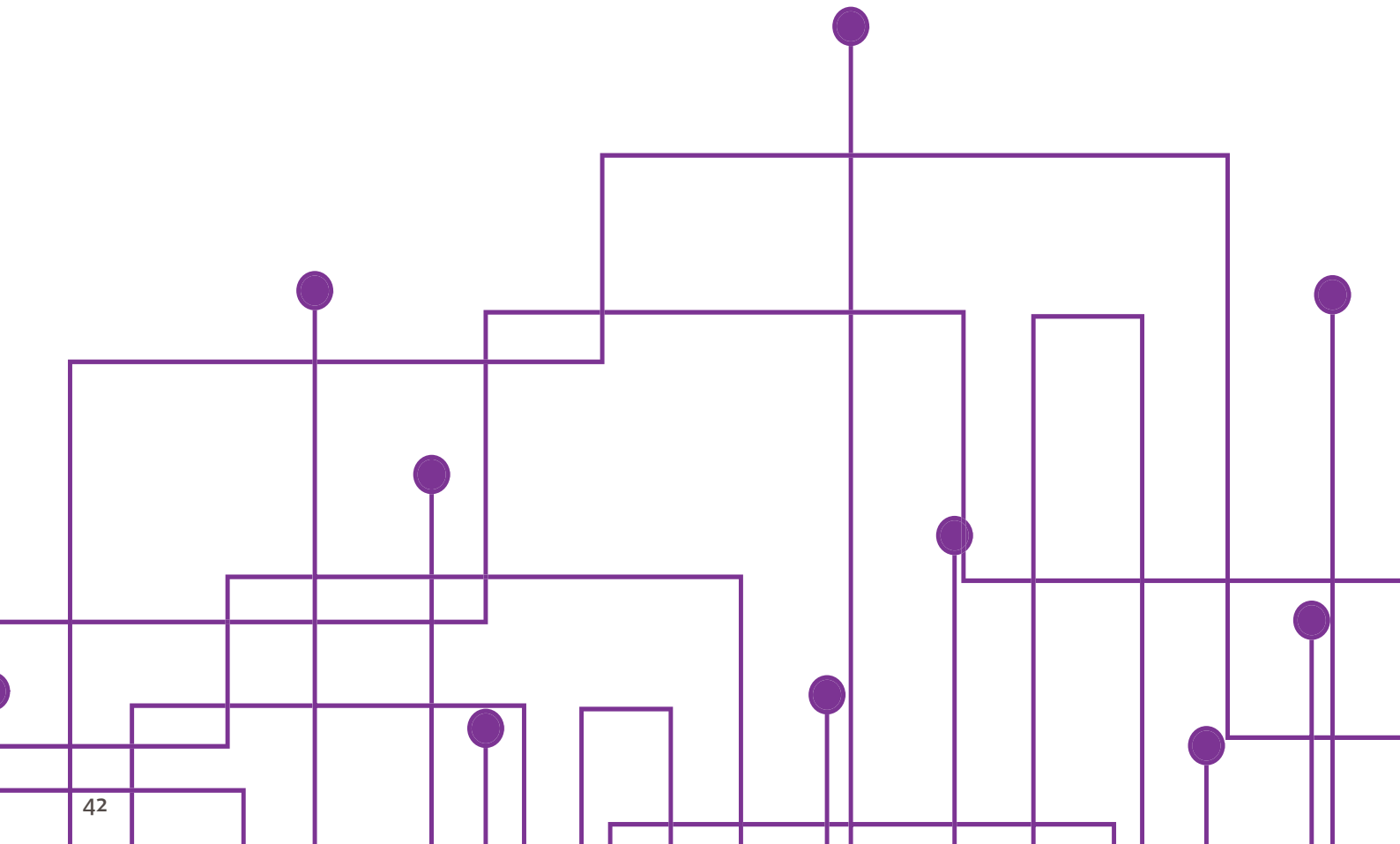
	все источники BGP / число маршрутов (октябрь 2017)		глобальные Full View / число маршрутов (октябрь 2017)
1	AS28761 (CrimeaCom LTD.) / 141	1	AS28761 (CrimeaCom LTD.) / 120
2	AS50384 (W-IX LTD) / 55	2	AS47379 (SITS-AS) / 33
3	AS6679 (SKADI-AS) / 41	3	AS6679 (SKADI-AS) / 26
4	AS47379 (SITS-AS) / 38	4	AS56630 (Melbikomas UAB) / 25
5	AS13178 (DIGCOMM) / 35	5	AS3327 (CITIC Telecom CPC Netherlands B.V.) / 23
6	AS3327 (CITIC Telecom CPC Netherlands B.V.) / 31	6	AS13178 (DIGCOMM) / 20
7	AS56630 (Melbikomas UAB) / 28	7	**AS2683 (RADIO-MSU) / 19**
8	AS2683 (RADIO-MSU) / 25	8	AS50241 (UNITTEL-AS) / 16
9	AS50241 (UNITTEL-AS) / 19	9	AS49320 (KOMTEKS TRC FIORD) / 13
10	AS197556 (TNS) / 17	10	AS197556 (TNS) / 12

Будем считать, что AS65550 является иностранной, а остальные AS, присутствующие в маршруте, российские. Оба указанных маршрута содержат трансграничный переход. Соответственно, такие маршруты будут подсчитаны как *два* уникальных маршрута и отнесены в приведённых в таблице 6 рейтингах к AS65550, которая получит показатель 2 (если других подходящих маршрутов не будет найдено). В случае, если маршрут содержит несколько последовательных иностранных AS, учитывается только та из них, которая является соседней слева к российской.

Другими словами, маршруты, ведущие к разным оконечным AS, но проходящие через общую транзитную, являющуюся «пограничной», будут учитываться как различные трансграничные маршруты, в которых участвует данная AS. Рейтинг, соответственно, отражает разнообразие маршрутов, трафик которых иностранная AS может наблюдать. Это только маршруты, по которым нельзя судить, насколько большой поток трафика передаётся между AS. Так, из того, что какая-то AS «видит» сто различных маршрутов с трансграничным переходом, прямо не следует, что эта AS «видит» больше или меньше трафика, чем AS, которая является «пограничной» всего лишь для трёх маршрутов. Более того, так как трансграничный переход определяется нами административно (по принадлежности AS), его наличие не является ни необходимым, ни достаточным признаком того, что трафик, следующий по данному маршруту, действительно покидает пределы России. Так, возможны конфигурации, когда провайдер, оператор российской автономной системы, использует виртуальные или физические каналы, проходящие через иностранные узлы связи, но при этом данные переходы относятся к уровню ниже IP, соответственно, в BGP наблюдаться не могут. С другой стороны, оператор иностранной AS может для работы внутри России использовать только российские каналы связи и, тем не менее, соответствующие маршруты будут учитываться как маршруты с трансграничным переходом.

Отметим, что таблица 6 в части рейтинга Full View хорошо показывает ограничения географической привязки по данным RIPE NCC. Так, в таблице указана AS2683 (**RADIO-MSU**), которая в базе данных RIPE NCC принадлежит к региону EU (Европейский Союз). Однако данная AS фактически является российской.

Анализ маршрутов сам по себе не позволяет оценить ни объёмы трафика, ни набор ресурсов, трафик которых проходит через данную транзитную AS. Однако если маршрутную информацию сопоставить с именами ресурсов, то можно определить «населённость» тех или иных маршрутов сервисами.



Заключительные замечания

Сервисы являются ключевым элементом современного Интернета. Сопоставление маршрутной информации, получаемой при помощи сбора данных BGP, с распределением имён и адресов сервисов, позволяет обнаружить некоторые закономерности и рассчитать показатели связности, которые не видны только из BGP-данных. Например, можно выделить транзитные AS, связанные с большим числом веб-узлов или почтовых серверов; обнаружить кластеры веб-узлов, маршруты в направлении которых содержат трансграничные переходы, что потенциально может привести к дополнительным проблемам с доставкой данных в случае нарушения глобальной связности Сети. Наличие трансграничного перехода в конфигурации BGP для конкретного момента времени само по себе не означает, что между узлами, которые «разделены» данным переходом, невозможны другие маршруты: в случае нарушения связности одного маршрута, достаточно быстро может быть выбран другой, в том числе, уже не содержащий трансграничного перехода.

Очередной раз отметим, что речь идёт только о тех маршрутах, в которых оконечная AS - российская, а переход через (виртуальную) границу происходит уже после того, как маршрут «пришёл» в российский сегмент. То есть такой переход часто выглядит лишним, хоть и может быть обоснован с сетевой точки зрения. Для узлов Рунета доля маршрутов, содержащих трансграничный переход, уже не велика. Возникновение трансграничного перехода часто объясняется не столько техническими, сколько административными причинами (к которым относится, например, стратегия пиринга, используемая операторами). Дальнейшему снижению количества подобных маршрутов может способствовать развитие инфраструктуры российских точек обмена трафиком (IXP), так как они служат средой, позволяющей нивелировать как административные, так и технические причины возникновения излишних трансграничных переходов.

При этом сопоставление данных по именам/адресам сервисов с BGP характеризуется определёнными трудностями. Так, обход узлов роботами с целью определения доступности сервисов занимает заметное время, в течение которого BGP-таблицы могут измениться. BGP-таблицы оказываются самой динамичной частью исходных данных. Значительные изменения в них могут произойти в течение суток. Адресация веб-узлов и почтовых сервисов изменяется существенно медленнее, типичным интервалом времени здесь является месяц. Таким образом, при анализе миллионов имён узлов возможные скачкообразные изменения маршрутов (например, утечки маршрутов или кратковременный «угон» префиксов) часто сглаживаются, поэтому оказываются не видны.

Новости Доменной индустрии

Важные события 2018

RIGF 2018: КИБЕРБЕЗОПАСНОСТЬ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МОЛОДЕЖНАЯ ТЕМАТИКА

06.04.2018

В Санкт-Петербурге состоялось официальное открытие Девятого российского форума по управлению Интернетом. Форум уже во второй раз проходит не в Москве – на этот раз городом проведения была выбрана «вторая российская столица» Санкт-Петербург. Не забыт и молодежный аспект: 5 апреля спикеры и эксперты RIGF 2018 выступили с лекциями перед студентами вузов города.

Ежегодно на Российском форуме по управлению Интернетом вручается специальная награда – орден «За заслуги в сфере Интернета» (Virtuti Interneti). Лауреатами награды становятся представители интернет-сообщества, бизнеса, науки и государства, внесшие значимый вклад в развитие Рунета и глобального Интернета. В этом году орден был вручен Алексею Платонову – директору Технического центра Интернет.

Программа Девятого российского форума по управлению Интернетом практически полностью была посвящена вопросам кибербезопасности. На четырех секциях форума участники обсудили информационную безопасность с разных сторон, выбрав наиболее острые проблемы: фейковые новости, международное сотрудничество в области кибербезопасности, чрезвычайные ситуации в киберпространстве, преимущества и опасности искусственного интеллекта.

НА РИФ 2018 ПРЕДСТАВЛЕНО ИССЛЕДОВАНИЕ «ТЕНДЕНЦИИ РАЗВИТИЯ ИНТЕРНЕТА В РОССИИ»

18.04.2018

18 апреля в Подмосковье открылся XXII Российский интернет-форум РИФ 2018. Координационный центр доменов .ru/.рф, как и в прошлые годы, выступил партнером форума. Директор КЦ Андрей Воробьев выступил на секции РИФ «Бизнес на данных. Баланс интересов пользователя, компаний и государства», где рассказал об исследовании «Тенденции развития Интернета в России», которое Координационный центр провел совместно с Высшей школой экономики.

Главное отличие этого исследования от похожих работ состоит в том, что оно опирается не на экспертную оценку, а на статистические данные, получаемые из официальных источников. Среди источников - Росстат, Минкомсвязь России, Минобрнауки России, Минкультуры России, Банк России. Методика исследования базируется на лучших мировых практиках, которые используются такими организациями как ОЭСР, Евростат, Международный союз электросвязи. «Мы предполагаем, что текущее исследование будет востребовано различными государственными органами РФ. Кроме того, оно сможет служить своеобразным «камертоном», по которому можно будет проверять другие исследовательские инструменты. Выводы, основанные на статистике, покажут, насколько адекватны и близки к реальности другие типы исследований, основанные на экспертном мнении», – рассказал Андрей Воробьев.

РИФ/2018
www.rif.ru

Согласно данным исследования, за последние пять лет в Российской Федерации на четверть выросла доля граждан, владеющих навыками использования Интернета – с 66% в 2012 году до 80,8% в 2016. Российские граждане все активнее взаимодействуют с органами государственной власти и местного самоуправления по Интернету и обращаются за госуслугами в электронной форме: в 2014–2016 гг. доля таких граждан выросла почти в 1,5 раза (до 51,3%). За самой популярной электронной услугой – записью на прием к врачу – обращалась треть (32,4%) респондентов, использующих официальные сайты и порталы госуслуг.

.RU и .РФ

В 2017 ГОДУ С ДЕЛЕГИРОВАНИЯ БЫЛО СНЯТО 5108 ДОМЕННЫХ ИМЕН .RU И .РФ

25.04.2018

Координационный центр доменов .ru/.рф подготовил расширенный годовой отчет о деятельности компетентных организаций за 2017 год. В 2017 году аккредитованные регистраторы получили от компетентных организаций 5487 обращений о прекращении делегирования доменных имён-нарушителей, из них в период до 31 декабря 2017 года было рассмотрено 5481 обращение, 6 находились на рассмотрении.

В результате за год с делегирования было снято 5108 доменных имен, что составляет 93,2% от общего числа обращений. Еще 45 доменных имен (0,8%) были заблокированы хостинг-провайдерами. 328 доменных имен (5,9%) остались делегированными, т.к. 139 из них были разблокированы вследствие устранения причин блокировки администраторами доменных имен, а 189 доменных имен не были заблокированы вследствие оперативного устранения причин блокировки.

Наибольшая доля обнаруженных доменов-нарушителей приходится на фишинг – 2698 имен или 49,2% от общего числа. Чуть меньшее количество пришлось на ресурсы, с помощью которых происходило распространение вредоносного ПО – 2520 имен или 45,9%. 222 домена-нарушителя являются ботнет-контроллерами, еще 35 представляют собой финансовые пирамиды, а 12 – мошеннические сайты.

АЗИАТСКО-ТИХООКЕАНСКИЙ ФОРУМ ПО УПРАВЛЕНИЮ ИНТЕРНЕТОМ ВПЕРВЫЕ ПРОЙДЕТ В РОССИИ

26.04.2018

25 апреля Мультистейкхолдерная группа (MSG), действующая в Азиатско-Тихоокеанском регионе и занимающаяся вопросами организации региональных форумов, приняла решение о месте проведения Азиатско-Тихоокеанского форума по управлению Интернетом (APrIGF) в 2019 году. Всего рассматривались две заявки на проведение форума – от России и от Китая. Российскую заявку подавал Координационный центр доменов .ru/.рф, и именно она стала победителем. При голосовании в Мультистейкхолдерной группе российская заявка набрала 27 голосов против 9 голосов за заявку от Китая.

APrIGF 2019 пройдет в конце июля 2019 года во Владивостоке на острове Русский, в Дальневосточном федеральном университете. Точная дата проведения форума будет названа дополнительно.

Азиатско-Тихоокеанский региональный форум по управлению Интернетом (APrIGF) проводится с 2010 года. Он служит платформой для обсуждения проблем региона, обмена и сотрудничества на региональном уровне, а также для объединения повесток национальных форумов и, в конечном счете, для развития управления интернетом в Азиатско-Тихоокеанском регионе.

КООРДИНАЦИОННЫЙ ЦЕНТР ДОМЕНОВ .RU/.РФ СОЗДАЕТ МОЛОДЕЖНУЮ ПАЛАТУ

27.04.2018

26 апреля в ЦВК «Экспоцентр» на площадке Большого медиа-коммуникационного форума 2018 (БМКФ 2018) прошел шестой молодёжный форум «Этические, культурологические и цивилизационные аспекты работы в сети Интернет», посвященный вопросам формирования культуры информационной безопасности и изучению доверия и безопасности при использовании ИКТ. Основная аудитория форума – студенты и преподаватели профильных вузов. Форум организован общественно-государственным объединением «Ассоциация документальной электросвязи» при поддержке Координационного центра доменов .ru/.рф.

Перед участниками форума выступил директор Координационного центра доменов .ru/.рф Андрей Воробьев. Он сообщил участникам, что при Координационном центре создается молодежная палата – консультативный орган, при помощи которого представители молодого поколения будут вовлекаться в процесс управления Интернетом. «Российский форум по управлению Интернетом (RIGF) уже три года активно формирует молодежную повестку и привлекает студентов к участию. Мы видим большой интерес молодежи к этому процессу, желание учиться, развиваться, вносить свой вклад в управление Интернетом. Для нас включение молодежи в эту работу также очень важно – благодаря этому мы получаем новый и свежий взгляд на процесс управления Интернетом сегодня. Мы планируем сохранить эту традицию и для наших будущих мероприятий, и, конечно же, молодежная палата при Координационном центре будет принимать в их подготовке самое активное участие», – рассказал Андрей Воробьев.

ICANN ВЫСОКО ОЦЕНИЛА РОССИЙСКИЙ ОПЫТ БОРЬБЫ С КИБЕРУГРОЗАМИ

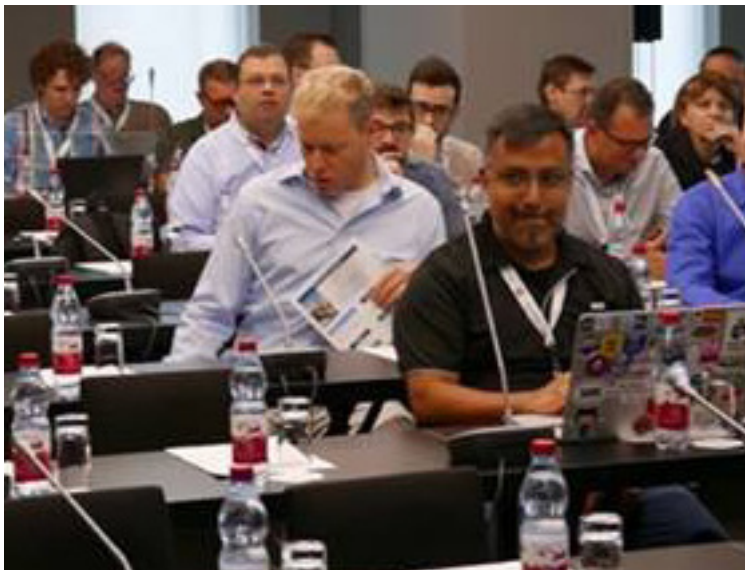
23.05.2018

23 мая 2018 года прошла открытая встреча CEO и президента ICANN Йорана Марби и главного технического директора ICANN Дэвида Конрада со специалистами Координационного центра доменов .ru/.рф и представителями российского рынка регистрации доменов. Во время встречи обсуждались основные проблемы и последние тенденции в области развития мировой системы интернет-адресации. Встречи с руководством ICANN в Москве уже проводились, но Йоран Марби и Дэвид Конрад приехали в Россию впервые.

Одной из самых важных тем, поднятых на встрече, стала борьба с киберугрозами. Руководители ICANN высказали свое восхищение тем, как в Рунете построена эта работа. «В России делаются очень интересные шаги в области обеспечения кибербезопасности – например, созданы «горячие линии». И мы в ICANN можем перенять этот опыт и распространить его на другие страны», - сказал Марби. Дэвид Конрад также рассказал о разработке функционала для правоохранительных органов, которой сейчас занимается ICANN. «Одна из причин, по которой мы приехали в Москву, – это наше желание начать совместную работу в этом направлении», – отметил Дэвид Конрад. Эти темы поднимались и на закрытой встрече руководства Координационного центра доменов .ru/.рф и ICANN, которая прошла перед началом открытого мероприятия. На ней также обсуждались вопросы управления доменным пространством в регионе, позиция российского интернет-сообщества по ключевым вопросам развития Сети, обмен опытом и организация совместных мероприятий.

В МОСКВУ ПРИЕХАЛИ ЭКСПЕРТЫ ЕВРОПЕЙСКИХ НАЦИОНАЛЬНЫХ РЕГИСТРАТУР

30.05.2018



30 мая открылась конференция CENTR Jamboree 2018, которая впервые в своей истории проводилась в Москве. В конференции участвовали более 150 экспертов и специалистов, представляющих регистратуры национальных доменов верхнего уровня, входящих в ассоциацию CENTR (Council of European National TLD Registries). В течение трех дней участники обменивались опытом и новыми идеями на семинарах по техническим и юридическим вопросам, безопасности, маркетингу, познакомились с выступлениями приглашенных экспертов.

CENTR - это ассоциация европейских регистратур национальных доменов верхнего уровня. В настоящее время CENTR насчитывает 54 полных и 9 ассоциированных членов - вместе они отвечают за более чем 80% всех зарегистрированных в национальных доменах доменных имен во всем мире. В Совет директоров CENTR входит заместитель директора Координационного центра доменов .ru/.рф Ирина Данелия. Цели CENTR заключаются в содействии и участии в разработке стандартов и передовой практики среди реестров национальных доменов верхнего уровня (ccTLDs), организации мероприятий для обмена опытом и экспертизой, проведении совместных исследований по различным вопросам и представлении интересов регистратур доменов на общеевропейском уровне.

Календарь событий: 2018 год

Международные события

14-20 июля 2018
IETF 102,
Монреаль, Канада

IETF (Internet Engineering Task Force) является одной из основных организаций по разработке стандартов Интернета. В основном работа в IETF проходит в многочисленных списках рассылки, соответствующих различным рабочим группам (этих групп более 100). Три раза в год IETF проводит недельные совещания, на которые приезжают разработчики протоколов, инженеры и операторы со всего мира (в среднем около 1200 участников из более 50 стран мира). Совещания IETF - это хорошая возможность познакомиться с новейшими тенденциями в области сетевых технологий и принять участие в их разработке. В выходные перед началом совещаний пройдет IETF Hackathon, посвященный практическому воплощению стандартов IETF, и IETF Codesprint по доработке приложения datatracker - важного инструментария IETF.
https://www.ietf.org/how/meetings/102/

1-3 октября 2018
NANOG 74,
Ванкувер, Канада

Североамериканская группа сетевых операторов (The North American Network Operators Group, NANOG) является одной из самых активных профессиональных ассоциаций в области сетевой архитектуры, конфигурации и технического администрирования сетей в Интернете. Основной фокус NANOG на технологиях и системах, обеспечивающих работу Интернета: систему глобальной маршрутизации, DNS, пиринг и связность. NANOG имеет активный список рассылки и проводит конференции три раза в год. Поскольку инженерные вопросы NANOG имеют глобальный характер, участие в списке рассылки и конференциях может быть полезно широкому кругу технических специалистов в области сетевых технологий Интернета. https://nanog.org/meetings/future

12-14 октября 2018
29 Семинар DNS-OARC,
Амстердам, Нидерланды

DNS-OARC - некоммерческая организация, целью которой является улучшение безопасности и стабильности инфраструктуры DNS, а также исследование работы этой глобальной системы. Семинары DNS-OARC открыты для членов OARC и для всех других участников, заинтересованных в работе и исследованиях DNS. В этот раз семинар проводится совместно со встречей CENTR-Tech 39.
Прием докладов заканчивается 13 июля.
https://indico.dns-oarc.net/event/29/

15-19 октября 2018
RIPE 77,
Амстердам, Нидерланды

Встреча RIPE проводятся два раза в год и собирают более 700 участников для обсуждения вопросов политики распределения номерных ресурсов (IP-адресов и номеров автономных систем) в зоне обслуживания RIPE NCC, сотрудничества, а также технических вопросов, связанных с маршрутизацией, DNS, связностью, измерениями и инструментарием. Встреча длится 5 дней и начинается с двухдневной пленарной программы, за которой следуют несколько параллельных сессий заседаний рабочих групп.
Прием докладов заканчивается 26 августа. https://ripe77.ripe.net/

20-26 октября 2018
ICANN 63,
Барселона, Испания

Встречи ICANN проводятся три раза в год в различных регионах земного шара для того, чтобы предоставить возможность активным членам сообщества ICANN лично поучаствовать в обсуждении насущных проблем. Общей темой, конечно, является DNS - глобальная система трансляции имен. Здесь обсуждаются как технические вопросы обслуживания услуг DNS, так и юридические и бизнес-аспекты предоставления регистрационных услуг. Участие во встречах ICANN бесплатно.
https://meetings.icann.org/en/barcelona63

24-25 октября 2018
Peering Asia 2.0,
Гонконг

Так же, как Глобальный пиринговый форум (GPF) и Европейский пиринговый форум (EPF), Peering Asia 2.0 является открытой пиринговой встречей в Азиатско-Тихоокеанском регионе.
http://www.peeringasia.com/

3-9 ноября 2018
IETF 103,
Бангкок, Таиланд

Осенняя встреча IETF. IETF (Internet Engineering Task Force) является одной из основных организаций по разработке стандартов Интернета. В основном работа в IETF проходит в многочисленных списках рассылки, соответствующих различным рабочим группам (этих групп более 100).
https://www.ietf.org/how/meetings/103/

В России

16 августа 2018,
Якутск
и другие города

Код информационной безопасности
Серия конференций, проходящих через 27 крупных городов России, Казахстана, Белоруссии, Грузии, Азербайджана и Армении. Исчерпывающая информация о новинках и трендах в современных IT-угрозах и достижениях в борьбе с ними. https://conf.codeib.ru/
О мероприятиях, проводимых в других городах, также см. по ссылке.

6 сентября 2018,
Екатеринбург
и другие города

BIT-2018
Серия конференций «Вокруг Облака. Вокруг ЦОД. Вокруг Данных. Вокруг IoT. Вокруг IP...» и пр.
BIT – это событие №1 на отечественном рынке высоких технологий. Мероприятие покрывает все вопросы, связанные с центрами обработки данных, современными коммуникационными сервисами, облачными вычислениями, Интернетом вещей, хранением данных и бизнес-аналитикой.
https://ciseventsgroup.com/events.html
О мероприятиях, проводимых в других городах, также см. по ссылке.

18 сентября 2018,
Санкт-Петербург

Ipv6 day 2018
«Российский день IPv6 MSK-IX» – ежегодное место обсуждения последних тенденций в области связи и передачи данных, а также технических стандартов Интернета.
Организатор - MSK-IX, https://www.msk-ix.ru/

19-21 сентября 2018,
Сочи

Промышленная кибербезопасность: цифровая трансформация — вызовы и возможности
Двухдневная деловая программа, организованная «Лабораторий Касперского» для специалистов, работающих с системами управления технологическими процессами и критически важными инфраструктурами. https://ics.kaspersky.ru/conference/

В Москве

25 сентября 2018

Интернет вещей 2018
Представители технологических стартапов и госструктур, разработчики, операторы связи, интеграторы, инженеры и робототехники, специалисты по кибербезопасности, а также предприниматели и инвесторы обсудят главные задачи IoT-рынка в России и поделятся практическим опытом по внедрению Интернета вещей в бизнес. https://iotconf.ru/ru

3 октября 2018

Cloud Day 2018
Преимущества от миграции в облако, сложности и риски в процессе такого «переезда» и как они решаются, критерии выбора поставщика облачных сервисов, эти и другие вопросы будут обсуждаться на конференции TAdviser Cloud Day 2018.
http://www.tadviser.ru/index.php/Статья:Конференция_Cloud_Day_2018

8-9 ноября 2018

Highload++ 2018
Крупнейшая профессиональная конференция для разработчиков высоконагруженных систем. http://www.highload.ru/

15-16 ноября 2018

ZeroNights – 2018
Международная конференция, посвященная практическим аспектам информационной безопасности. https://2018.zeronights.ru/

6 декабря 2018

Пиринговый форум MSK-IX
Крупнейшая ежегодная встреча участников интернет- и телеком-рынка. Форум проводится с 2005 года, с каждым годом собирая все большее число профессионалов. За это время из внутрисетевого мероприятия среди участников MSK-IX форум превратился в открытую деловую площадку, на которой обсуждаются самые актуальные вопросы развития сети и обмена трафиком.
Организатор – MSK-IX, https://www.msk-ix.ru/

MSK IX

10
ГОРОДОВ

500+
УЧАСТНИКОВ

42
ПЛОЩАДКИ

21
УЗЛЕЛ DNS-СЕТИ

ПОДКЛЮЧЕНИЯ ДО
100G

ТРАФИК
2,8Тбит/с

MSK-IX ускоряет коммуникации между интернет-компаниями, предоставляя нейтральную платформу Internet eXchange для обмена IP-трафиком между сетями и глобальную распределенную сеть DNS-серверов для поддержки корневых доменных зон.

Более 500 организаций используют сервисы MSK-IX для развития сетевого присутствия в 10 городах. К MSK-IX подключены операторы связи, социальные сети, поисковые системы, видеопорталы, провайдеры облачных сервисов, корпоративные и научно-образовательные сети.

127083, г. Москва, ул. 8 Марта, д. 1, стр. 12
тел.: +7 495 737-92-95
www.msk-ix.ru

+7 495 737-92-95

WWW.MSK-IX.RU

Интернет изнутри

2018