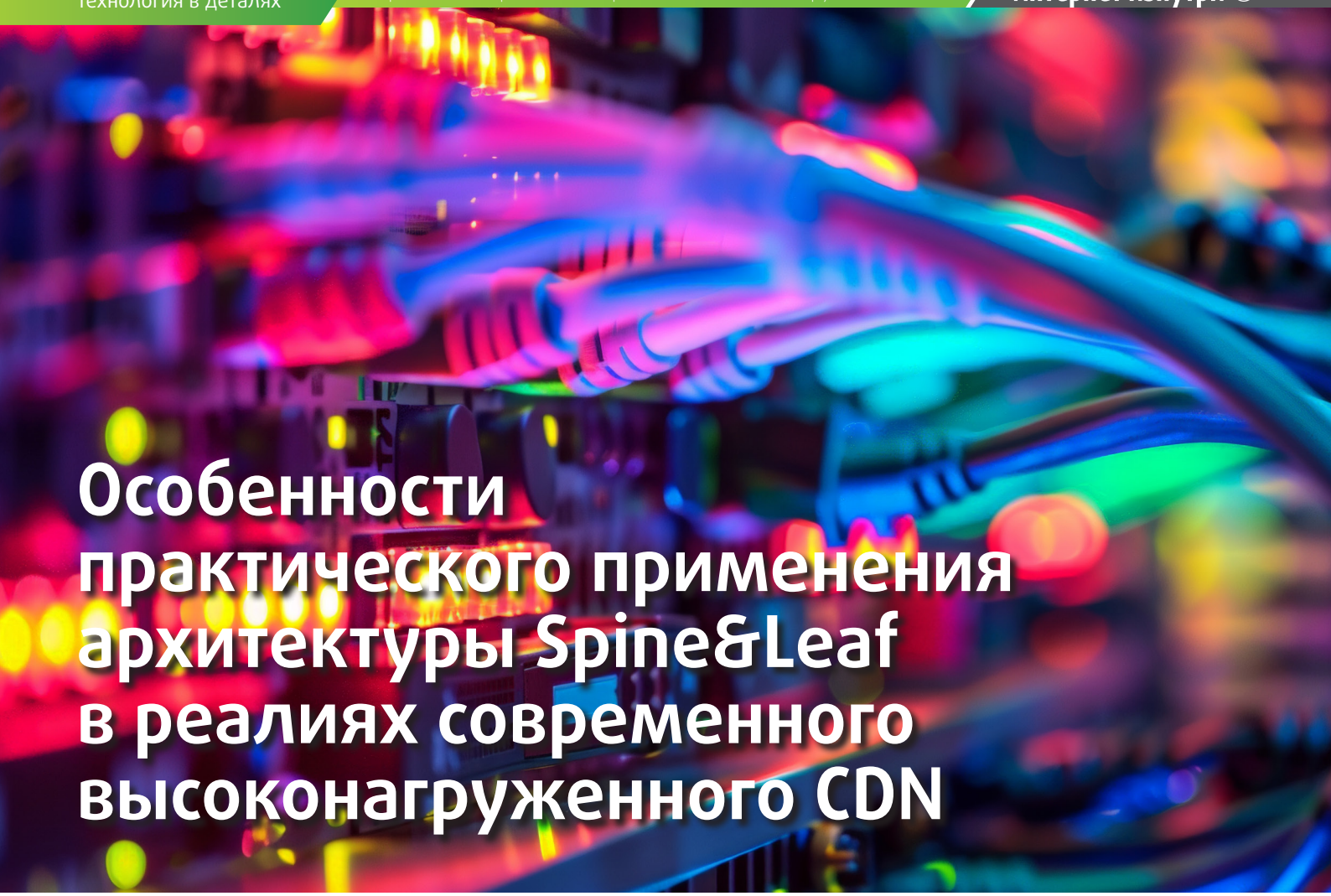

Особенности практического применения архитектуры Spine&Leaf в реалиях современного высоконагруженного CDN

Арсений Великород

Аннотация

В статье рассматриваются вопросы оптимизации использования архитектуры Spine&Leaf на высоконагруженной сети оператора сети доставки контента. Производится анализ проблемы оптимального использования портовой ёмкости в рамках архитектуры Spine&Leaf

Ключевые слова:

дата-центр, центр обработки данных, CDN, high-load, spine&leaf, datacenters.

Проблематика

Целью данной статьи является попытка практической рационализации архитектурной модели S&L (Spine&Leaf), которая сегодня широко применяется при построении сетевой инфраструктуры в дата-центрах. Первым принципиальным отличием её от классической трёхуровневой модели является то, что S&L — двухуровневая. В ней spine совмещает функции агрегации и ядра (поэтому spine является ещё и маршрутизатором), в него включаются leaf-узлы. По своему функционалу это L2+ коммутаторы. Именно к ним подключается конечное оборудование (серверы) в дата-центрах. Основным фокусом статьи являются проблемы практического применения данного типа архитектуры на высоконагруженных сервисах, в частности, на инфраструктуре оператора сети доставки контента (Content Delivery Network, CDN).

Выбор столь узкой проблематики продиктован, с одной стороны, опытом проектирования, строительства и эксплуатации автором статьи реальной сетевой инфраструктуры оператора сети доставки контента, с другой стороны — возникшими фундаментальными вопросами к реализации теоретической модели в реальном дата-центре и на реальном оборудовании.

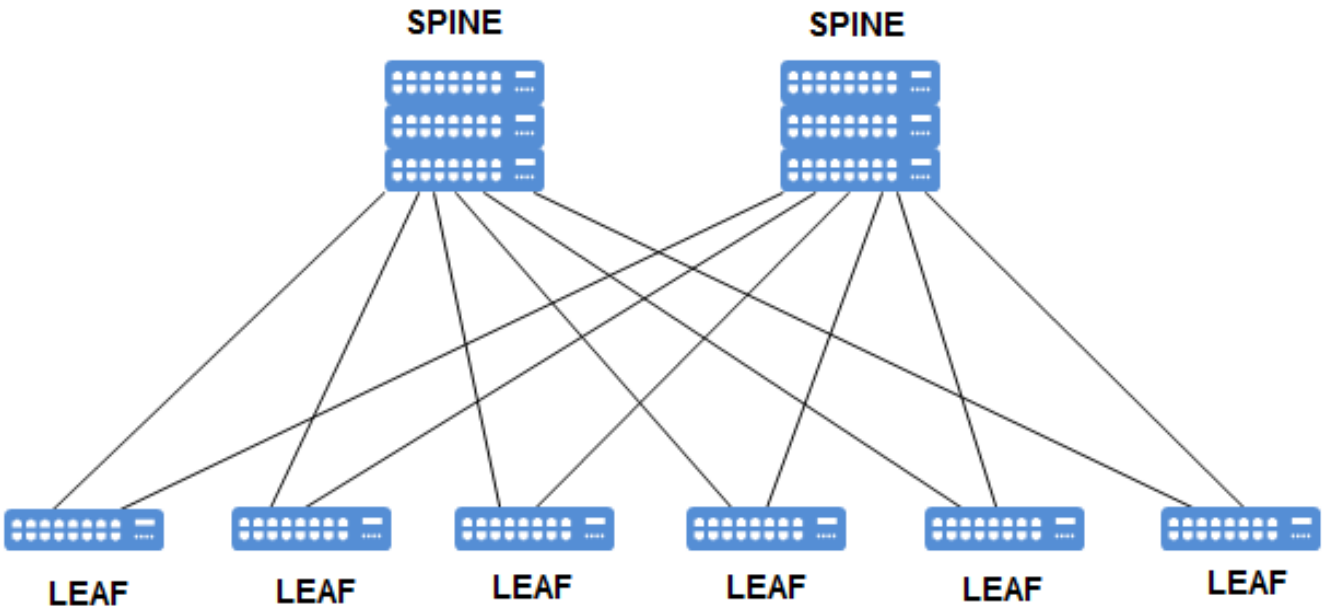

Рис. 1. Взаимодействие между leaf- и spine-узлами.

В частности, основная проблема, с которой приходится сталкиваться, — это проблема портовой ёмкости. В чистом S&L все spine соединяются со всеми leaf так, как показано на рисунке 1. Но в своей практике мы редуцировали эту схему за счёт размещения нескольких локаций[1] в одном дата-центре. Ведь каждый spine-коммутатор должен быть связан со всеми leaf-коммутаторами каналами достаточной ёмкости, но если на сети приходится иметь дело с тяжелым трафиком (VOD (Video On Demand), онлайн-трансляции), то каждый из узлов CDN может генерировать трафик, трёх- или четырёхкратный объём которого может полностью утилизировать полосу 100 Гбит/с. В этой ситуации существует вполне разумное решение — увеличить ширину канала между spine и таким leaf, куда включены подобные серверы сети. Но данное решение в условиях использования коммутаторов класса datacentre может серьёзно уменьшить количество портов, которые используются для включения соединений всех уровней (spine-spine, spine-leaf, leaf-leaf). Если же следовать экстенсивному пути развития этой архитектуры и наложить серьёзные структурные ограничения на возможности соединений типа leaf-leaf с попутным масштабированием структуры по размеру, при этом стараясь сохранить потребную для каждого типа трафика необходимую полосу, то можно очень быстро прийти к вопросам экономической целесообразности, а также, что более важно с точки зрения высокой доступности сервиса, возможности резервирования как самого сетевого оборудования, так и доступности внешних каналов.

Условия применимости рассматриваемых моделей

Прежде всего оговоримся, что далее мы будем рассматривать работу сетевой инфраструктуры в рамках резервированной на программном уровне сети доставки контента. Здесь подразумевается такая программная модель CDN, когда падение одного физического сервера (ноды) не приводит к остановке сервиса, а приводит к переводу или перераспределению нагрузки между другими нодами кластера.

Также отдельно отметим, что рассматриваемая ситуация и модели применимы к дата-центрам, которые не относятся к топ-классу. Сегодня подавляющее большинство коммерческих дата-центров в России (Москва, Санкт-Петербург, Екатеринбург, Новосибирск, Иркутск, Хабаровск) предлагают чаще всего электрическую ёмкость 5 кВт (с резервированием двумя «лучами» или без. Данные сведения были собраны за период с января 2024 по сентябрь 2024 года). Так как основная задача CDN — это обеспечить раздачу контента из места, находящегося как можно ближе (в сетевом смысле) к конечному потребителю, что продиктовано соображениями минимизации задержки (RTT), то мы столкнёмся либо с ограничениями по предлагаемым в дата-центре условиям, либо с необходимостью применения модели, предусматривающей возможности самой гибкой адаптации к местным условиям. Поэтому мы исходим из условий, когда на стойку 48U выделяется средняя стандартная электрическая мощность 5 кВт. Тот факт, что на большой локации сети необходимо размещать большое количество нагруженных сетевым трафиком серверов с большими объёмами дисковых массивов, ставит серьёзные ограничения по энергетике для каждой стойки.

Вышесказанное, в свою очередь, влияет на форм-фактор выбираемого сетевого оборудования (в подавляющем случае — маршрутизирующих коммутаторов), его портовую ёмкость и энергетические эксплуатационные параметры. В частности, исходя из баланса технических условий и финансово-экономических показателей, первым выбором чаще всего являются коммутаторы серий, адаптированных под дата-центры форм-фактора 1–2U.

[1] Локацией мы называем инстанс дерева архитектуры S-L, объединённый одной политикой маршрутизации. У этого инстанса существуют свои каналы вовне, свой домен BGP-маршрутизации, своё адресное пространство. Технически это самостоятельная стойка (группа стоек с нодами, leaf-коммутаторами и spine-маршрутизаторами, к которым имеется доступ из публичной сети Интернет).

Проблема портовой ёмкости как основная

Проблема распределения и оптимального использования портовой ёмкости на узле CDN с высокой нагрузкой — это основная проблема, с которой нам пришлось столкнуться в процессе проектирования новых и модернизации старых узлов (локаций). Структурно каждая локация представляет из себя от одной до трёх-четырёх стоек. Нами изначально была применена архитектура Spine&Leaf, но в силу естественных причин роста и развития наша реальная инфраструктура не была «чиста» в плане архитектурном — горизонтальные связи типа L-L (leaf-leaf), пиринговые присоединения на leaf'ах, «последовательная» группировка leaf'ов — всё это существовало и со временем, из-за значительного роста трафика, стало представлять проблему из-за возникновения многих точек на сети, склонных к перегрузкам на большом трафике.

Вышесказанное относится к подавляющему большинству сетевых инфраструктур, которые выросли из маленьких проектов. И это нормально, но самым главным этапом на пути этого роста является верный выбор архитектурной модели сети, которая должна не только учитывать конкретные особенности бизнеса и трафика, но и обеспечивать масштабируемость без больших капитальных затрат.

Сегодня сети доставки контента обслуживают практически весь спектр медиаданных, начиная от мелкой статики и заканчивая VOD и онлайн-трансляциями. Данная ситуация навязывает оператору сети в том числе и требования по классификации трафика (traffic affinity), что в свою очередь приводит и к раздельному обслуживанию трафика. Таким образом, мы приходим к тому, что разные типы трафика обслуживаются на разных серверах локации (см. объяснение термина в сноске выше). Каждая категория трафика обладает своим уникальным профилем — это касается как временных характеристик (волнообразность трафика или его равномерность в течение суток), так и качественно-количественных характеристик (проектируемая мощность сервера на отдачу трафика, соотношение входящего и исходящего трафика на порту сервера, его способность по распределению между каналами в Интернете и пирами локации, или же наоборот — его поляризация в какой-то из них).

В целом, если говорить о больших локациях, то на сети приходится иметь дело с широким спектром трафика, который, в идеальной ситуации, равномерно и пропорционально заполняет внешние каналы локации.

Однако, как было ранее отмечено, разные серверы раздают по-разному, и поэтому здесь возникает вопрос о планировании портовой ёмкости под все типы трафика, учитывая все неравномерности и колебания трафика, запроектированную пиковую нагрузку, а также требования к каналам по работе без перегрузки, как внешним, так и внутренним по отношению к локации.

В этой задаче мы рассмотрим несколько возможных моделей планирования портовой ёмкости с целью определить самые оптимальные, но стараясь не выйти за пределы архитектуры S&L, а также за условия, обозначенные ранее.

Также следует сделать важное замечание. В рамках внутренних соединений на локации мы считаем важным переход на порты с линейными скоростями 25/40/100 Гбит/с. Это продиктовано несколькими обстоятельствами: объёмом нагрузки на серверы (в особенности на те, которые обслуживают тяжёлые виды трафика (Live video или VOD)), распространённостью сегодня как сетевых карт на 25 Гбит/с, так и коммутаторов с комбинированными портами 10/25G, неэффективностью использования портовой ёмкости коммутатора при обслуживании скоростей раздачи более 30 Гбит/с (для таких соединений нужно включать серверы агрегатами из 10G портов, одновременно появляются издержки на расходные материалы, кроме того, такие агрегаты требуют большего внимания при эксплуатации и особенно тонкой настройки со стороны сервера).

Архитектура Spine&Leaf предоставляет широкие возможности по резервированию оборудования. На теоретическом уровне сегодня в дата-центрах мы видим разумным и более прогрессивным по сравнению с технологией стекирования (stack) [1] использование технологии MC-LAG [2] для резервирования каждого уровня (spine, leaf). Прежде всего, это связано с её большей гибкостью по сравнению со стекированием (априорное сохранение независимой конфигурации для каждого шасси). В рассматриваемых ниже ситуациях, если у нас есть MC-LAG-объединённые шасси для узла spine (S-S') и такие же для leaf (L-L'), то межуровневые соединения будут выполнены соответственно: S-L и S'-L'.

Однако мы намеренно редуцируем эти моменты в нашем последующем описании, т.к. на уровне архитектуры они ничего не меняют, а на практическом уровне нам не пришлось применять такое в силу определённой самостоятельности каждой локации и нашего изначального программного дизайна, который позволяет практически безболезненно вывести из работы любую локацию. Отдельные моменты особенностей применения технологий Stack и MC-LAG, а точнее, нашего не-выбора в их сторону в нашем частном случае будут оговорены далее

Равномерное распределение

Прежде всего, мы рассмотрим такой способ использования портовой ёмкости на локации, когда все серверы включаются равномерно как в leaf'ы, так и частично в spine. И если для тех из них, которые оказываются подключёнными непосредственно к spine, не возникает каких-либо ограничений по скорости раздачи, кроме как по суммарной ёмкости аплинков, то для серверов, включённых в leaf'ы, возникает ограничение по пропускной способности канала S-L (spine-leaf). Эксплуатация последовательно соединенных leaf'ов в таком случае становится ещё более затруднительной в силу полного исчерпания портовой ёмкости под соединения между коммутаторами.

Поясним этот момент. В рассматриваемой модели мы стараемся включать серверы равномерно во все коммутаторы локации. В этом случае в каждом свитче должно оказаться примерно поровну серверов, которые обслуживают разные

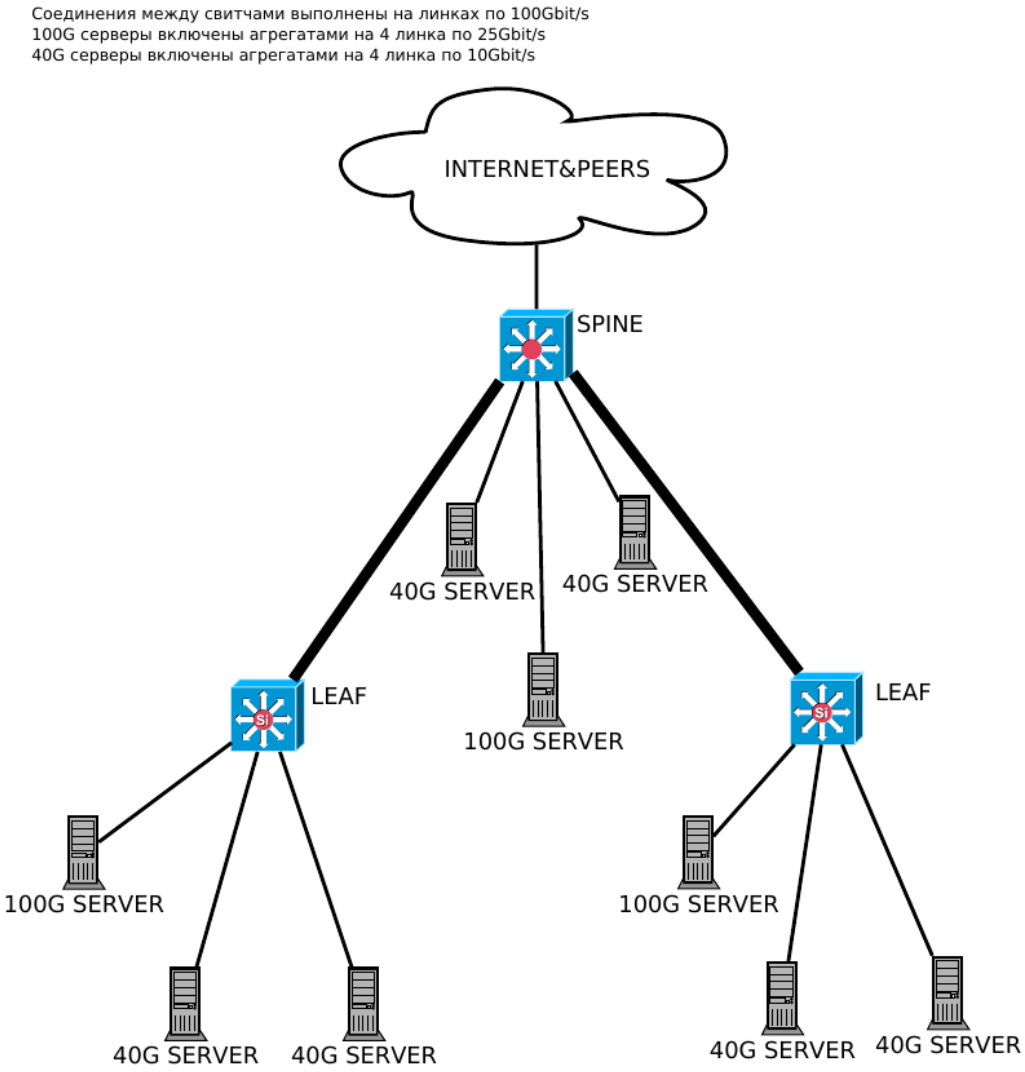

Рис. 2. Схема включения серверов при равномерном распределении.

типы трафика. Соответственно, объём суммарного трафика с каждого leaf планируется иметь примерно одинаковым по объёму. При включении только по одному leaf в порт spine проблем не возникает, так как не нужно планировать ёмкость для второго leaf'а. Однако, нарушив архитектуру и включив leaf в другой leaf, необходимо будет предусмотреть удвоенную ёмкость соединений как минимум между spine и первым leaf в этой ветке. Поэтому более разумно будет его включить отдельным портом. Здесь мы косвенно находим первое существенное ограничение этой модели — ограничение полосы между spine и leaf, которое, как мы увидим далее, является фундаментальным.

Но картина остаётся стабильной до тех пор, пока мы имеем дело с относительно умеренным трафиком с leaf и серверы, подключённые в него в час наибольшей нагрузки (ЧНН), отдают объём трафика не более чем 75% линейной скорости порта на коммутаторе (на самом деле, это очень мало, учитывая тот факт, что каждый сервер мы включаем как минимум двумя линками, чтобы, если даже один линк упал, то второй работал и сервер был доступен).

В этом случае максимальная эмпирическая нагрузка для 48-портового leaf-коммутатора с шестью 100-гигабитными портами будет выглядеть как:

(48 x (10 Гбит/с x 0,75) x 0,5) x 0,55 = 99 Гбит/с.

Коэффициент 0,55 был получен нами практическим путём как действующая величина поправки от неравномерности нагрузки. Применение коэффициента 0,5 связано с необходимостью включать сервер как минимум двумя портами в агрегате.

Как только мы захотим отдавать с сервера более чем 100% линейной скорости одного порта (обычного, не в сторону другого коммутатора) на leaf'е, мы очень быстро придём к существенным ограничениям по связи «наверх» — к spine. Прежде всего это связано с тем, что при использовании портов 25G необходимость раздавать с сервера 30–40 Гбит/с в ЧНН ведёт нас к быстрому исчерпанию 100 Gbit соединения со spine.

В этом случае максимальная эмпирическая нагрузка для такого же leaf-коммутатора будет выглядеть как:

(48 x (10 Гбит/с x 1,3) x 0,5) x 0,55 = 171,6 Гбит/с.

Конечно, такой трафик возможно пропустить только через два порта 100 Гбит/с.

В этом случае можно было бы, конечно, перейти на горизонтальное расширение — принципиально ограничиться одним «тяжёлым» сервером на leaf, а при необходимости добавления еще одного — организовывать еще один leaf и к нему присоединять этот сервер. Однако это приведёт к необходимости закупки в качестве spine коммутаторов с большим количеством 100G интерфейсов. В таком случае все leaf будут подключены к нему одним-двумя соединениями на 100G и проблема, казалось бы, исчерпана.

Тем не менее, проблема исчерпывается только на теоретическом уровне. Такие схемы могут быть эффективны только в развитых дата-центах с большим количеством свободных стойко-мест и адекватным запасом по энергетике. Это касается в основном коммерческих дата-центров в Москве и Санкт-Петербурге. В регионах же, во-первых, мало где можно найти достаточно места под такие локации, а во-вторых, до сих пор существует проблема с наличием портов с линейной скоростью более 10G у операторов связи.

Отдельным ограничивающим условием становится тот факт, что не в каждом дата-центре присутствует достаточное с точки зрения CDN количество операторов. Конечно, существует практика «партнёрских локаций», которые размещаются у заинтересованных региональных операторов, объём трафика на которые достаточно высок (обычно это крупные региональные операторы), но такие локации редко разрастаются на более чем одну стойку 48U. Поэтому кроме «партнёрских локаций» важно иметь по одной крупной локации как минимум на ФО (федеральный округ). В этом случае все партнёрские локации, расположенные в том же ФО, будут снабжены не только качественным трафиком того типа, под который они планировались, но и всеми остальными. В таких крупных региональных локациях всё равно приходится проектировать достаточные ёмкости на раздачу трафика. Сегодня запросы на услуги CDN растут, на рынок приходят новые клиенты, и всё чаще это клиенты с тяжёлыми видами трафика. И здесь сеть доставки контента оказывается между потребностью обеспечить необходимую под все типы трафика ёмкость под раздачу в регионе и реальным наличием такой ёмкости у операторов — как портовой, так и суммарной в регионе. Именно последним продиктована рекомендация располагать локации там, где имеется максимальное количество операторов — и федеральных, и региональных, — чтобы не концентрировать весь трафик раздачи в одном операторе. У такого оператора, возможно, существуют ограничения по ширине стыков с другими операторами в этом регионе или, что хуже, их в регионе нет вообще, а они находятся в соседнем. Здесь не стоит забывать о первоочередной задаче CDN — на раздаче быть ближе всего к конечному потребителю.

Таким образом, в реальных условиях предложенная выше модель становится трудноосуществимой в условиях необходимости множественного размещения. Выбор дата-центра под локацию зажат в узких рамках технических и экономических требований. Это и умеренная стоимость размещения, и отсутствие проблем по энергетике, наличие стойко-мест, наличие нескольких операторов, у которых должны быть технические возможности на включение запрошенной ёмкости. В этих условиях, если наш spine в локации будет иметь только 100G порты, то для присоединения к оператору понадобится, как минимум, один коммутатор с портами 10/25G. Также не стоит забывать, что, скорее всего, в этот коммутатор будут включаться несколько других операторов, и его соединение со spine будет одной из точек отказа. Вопрос планирования портовой ёмкости под каналы в сеть Интернет и под точки обмена трафиком останется за пределами этой статьи, однако стоит сказать, что решение и этого вопроса на практике не всегда оказывается простым делом.

Подводя итог, можно сказать, что модель равномерного распределения серверов между коммутаторами в локации видится излишне утяжелённой, как по капитальным затратам, так и по спектру вопросов, которые нужно будет решить попутно, если такую модель выбрать при включении серверов, раздающих на скоростях, сопоставимых с 40 Гбит/с. При этом такая модель организации локации несёт в себе значительное количество ограничений и необходимость организации множественных точек контроля. Также её возможности по масштабируемости напрямую связаны с капитальными затратами и расширением списка используемого сетевого оборудования, к которому нужно иметь ЗИП (и/или возможность быстро купить и заменить вышедший из строя модуль или весь коммутатор целиком).

Распределение с учётом максимальной нагрузки

Преодолеть проблемы предыдущей модели распределения нагрузки по портовой ёмкости мы можем, если при выборе места включения сервера будем принимать во внимание его планируемую способность к раздаче трафика. В частности, как ранее было указано, серверы, раздающие в ЧНН на скорости большей, чем линейная скорость одного порта (10 или 25 Гбит/с), являются основной причиной неоптимального использования портовой ёмкости. Однако, если такие серверы включать непосредственно в spine-коммутатор портами по 25 Гбит/с, а все остальные серверы, не создающие столь тяжёлый трафик, включать в leaf-коммутаторы, то можно достичь существенной оптимизации использования как портовой ёмкости, так и соединений между spine и его leaf'ами.

Это достигается тем, что самые загруженные серверы локации оказываются ближе всего к внешним каналам, и на этот трафик не требуется тратить ёмкость соединений между коммутаторами. А нагрузка от серверов, величина которой более адаптирована к линейной скорости портов, оказывается включенной в leaf'ы, соединения к которым хоть и выполнены чаще всего соединениями на линейной скорости 40 или 100 Гбит/с, но всё же обладают меньшей пропускной способностью, чем внутренняя шина коммутатора. В этом случае максимальная эмпирически рассчитанная нагрузка от leaf-коммутатора поддаётся расчёту по первой формуле из предыдущего раздела.

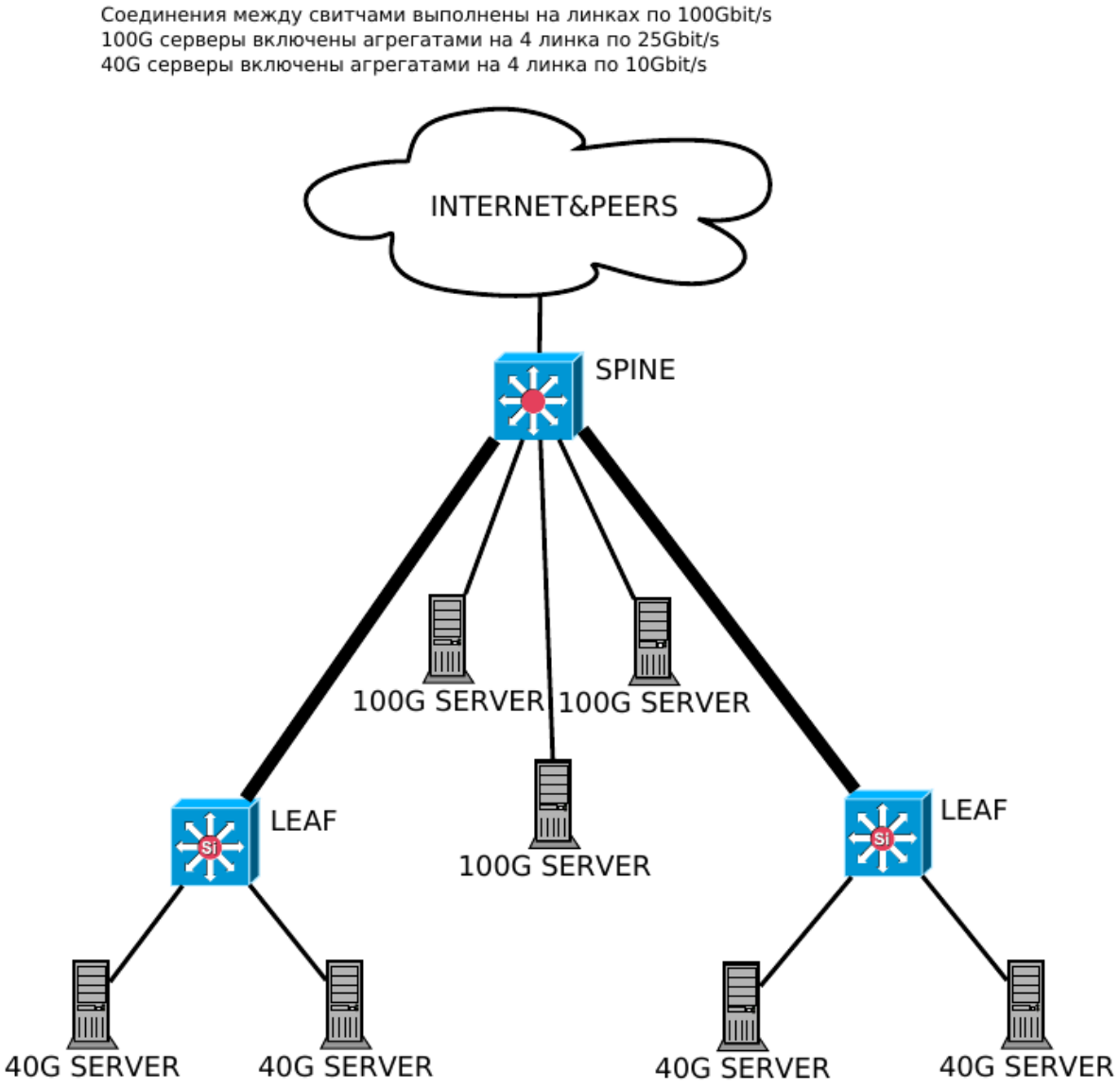

Рис. 3. Схема включения серверов с учетом нагрузки от них по трафику.

В целом, за счёт включения большой нагрузки как можно ближе к внешним каналам, можно добиться сбалансированности «дерева» по величине трафика. В особых случаях, если это позволяют объёмы трафика, допускается подключение к одному leaf-коммутатору другого leaf-коммутатора. Несмотря на некоторый отход от архитектуры, эта схема включения позволяет сэкономить порты по направлению «вниз» на spine-коммутаторе, к которым обычно подключаются leaf'ы. Таким образом, возможно выделять места включения сервисных систем, которые не обслуживают напрямую продуктовый трафик, но необходимы для функционирования внутренних систем (статистика, базы данных, хранение конфигурации и её резервирование, различные репозитарии и пр.)

Можно, конечно, возразить, что ничто не мешает кратно уменьшить объём отдачи с серверов, которые обрабатывают VOD и LIVE-трафик, а также другие его виды, которые требуют большой полосы на одну сессию, или же длительность сессии составляет более 5 секунд. Но здесь действуют соображения энергетической эффективности, которые не могут быть проигнорированы ни в каком дата-центре. Кратное уменьшение максимальной отдачи с сервера влечёт за собой необходимость кратного увеличения количества серверов, а значит, многократного увеличения потребления электроэнергии.

В своей практике мы пришли к пониманию, что все серверы, включённые в spine, должны иметь линки 25G. Это оправданно экономически и не создаёт накладных расходов по переключению портов 25G в режим 10G, что на коммутаторах Huawei и Juniper возможно только группами портов (обычно по четыре порта в группе).

Также ощутимым преимуществом модели с учётом нагрузки от включаемых серверов является то, что мы не оказываемся перед необходимостью ставить высокопроизводительные leaf-коммутаторы с комбинацией портов на 25/100 Гбит/с. Вполне возможно обойтись и более старыми линейками с комбинацией линейных портов на 10/40 Гбит/с. Также существуют переходные линейки с комбинацией портов 10/100 Гбит/с.

Отдельно отметим, что схемы расширения портовой ёмкости за счёт технологий MC-LAG и уж тем более стекирования (Stack) перестали нами рассматриваться в процессе эксплуатации из-за необходимости соблюдения баланса трафика

между всеми физическими шасси, включёнными в стек или объединёнными через mc-lag, что на практике означало бы наличие линка в каждое физическое шасси из каждого физического сервера. Учитывая практику установки по одному коммутатору в стойку, все возможные удобства от таких технологий нивелируются необходимостью межстоечных соединений.

В целом, рассмотренная модель распределения нагрузки по локации позволяет добиться оптимального использования портовой ёмкости и мощностей сетевого оборудования. Самый тяжёлый трафик оказывается ближе к внешним каналам, серверы, обслуживающие его, включаются в spine-коммутатор, а лёгкий трафик приходит из leaf-коммутаторов. Кроме того, такая модель оказывается экономически эффективной, когда есть необходимость в активном наращивании ёмкости локации по раздаваемому трафику, но нет одномоментной возможности провести комплексную модернизацию сетевой инфраструктуры локации. В этом случае цикл ротации оборудования обеспечивает наибольший срок его эксплуатации.

Гибридизация архитектуры Spine&Leaf

Тем не менее, существуют ситуации, где и учёт нагрузки не может покрыть требования к оптимальному использованию портовой ёмкости и внешних каналов.

Представим себе ситуацию, что в дата-центре А, где у нас уже стоит локация, закончилось место, и какой-то оператор предлагает разместиться у себя, а также предлагает трафик по хорошей цене.

В этой ситуации возможно рассмотреть вариант с организацией отдельной локации в ДЦ этого оператора. Однако в то же время не хочется и останавливать развитие существующей локации. В дата-центре Б, который, по сути, является дата-центром оператора, есть физическое размещение, но нет предположительно того спектра других операторов связи, который представлен в дата-центре А. Развитие отдельной локации в дата-центре Б может существенно нарушить баланс раздачи в том смысле, что раздача контента из новой локации будет вестись только через одного оператора. В случае проблем у оператора на сети, особенно если он предлагает нам канал в сеть Интернет полосой не менее 40 Гбит/с, мы можем существенно потерять в региональной ёмкости раздачи.

Если всё же мы решаемся развить существующую локацию в дата-центре А за счёт размещения в дата-центре Б и дополнительного включения там предлагаемого канала в Интернет, мы сталкиваемся с нестандартной задачей с точки зрения архитектуры.

Сделать коммутатор, установленный в дата-центре Б, просто leaf'ом того spin'а, который стоит в дата-центре А, видится самым простым решением. Но в таком случае мы не сможем полноценно решить обе задачи: и эффективно использовать новый канал в сеть Интернет, и размещать в дата-центре Б серверы для раздачи. В противном случае придётся соединять дата-центры А и Б весьма широким каналом. Но даже при наличии такого канала, при его падении мы получим не только падение одного внешнего канала на локации, но и недоступность тех серверов, которые находятся в дата-центре Б.

Для того чтобы решить эту проблему, мы решили воспользоваться гибридизацией, где коммутатор в дата-центре Б становится маршрутизатором. В него включается доступ в Интернет от оператора в дата-центре Б, устанавливается BGP-сессия с этим оператором. Но кроме того, через канал между дата-центрами устанавливается BGP-сессия со spine-коммутатором в дата-центре А. Подробно осветить выбор в этой схеме между IBGP и EBGP, к сожалению, мы здесь не можем т.к. каждый раз это определяется местными условиями, предпочтениями, конфигурацией сети и возможностями сетевого оборудования. Но общий смысл сводится к тому, что в итоге обе стойки (в дата-центрах А и Б) начинают пользоваться внешними каналами друг друга, они обмениваются не только внутренними маршрутами к L3-интерфейсам в серверные сегменты, но и маршрутами вовне.

Как результат, мы получим схему, резервированную лучше, чем та, которую бы мы получили, если бы сделали leaf в стойке дата-центра Б и включили бы в него канал доступа в Интернет от оператора связи. В нашей схеме отсутствие связанности по соединению между дата-центрами не приведёт к остановке сервиса ни в одной стоек. Возможно, потребуется поправить балансировку нагрузки, но это не приведёт к полному отсутствию сервиса. Такое включение даёт более гибкие возможности по расширению существующей инфраструктуры оператора CDN за счёт использования новых площадок, географически разделённых с основным дата-центром.

Заключение

В данной статье мы постарались изложить наш опыт и решённые проблемы реализации архитектуры Spine&Leaf на сетевой подсистеме высоконагруженного CDN. Наш опыт является скорее рационализаторством, чем инновацией. Описанный опыт — это попытка сократить издержки за счёт продуманного подхода к проектированию, развитию и эксплуатации, при этом не потеряв в надёжности сервиса.

В целом, из всего описанного можно сделать вывод, что вполне допустимо отходить от «чистой архитектуры» во имя результатов. Однако это не должно пониматься как полное пренебрежение принципами выбранной архитектурной модели и «изобретательство велосипеда». Наоборот, важен сбалансированный подход и долгосрочное планирование. В частности, это позволяет прийти к более гибким решениям, которые будет возможно развивать без особых проблем. Основные проблемы на сетевой ин-

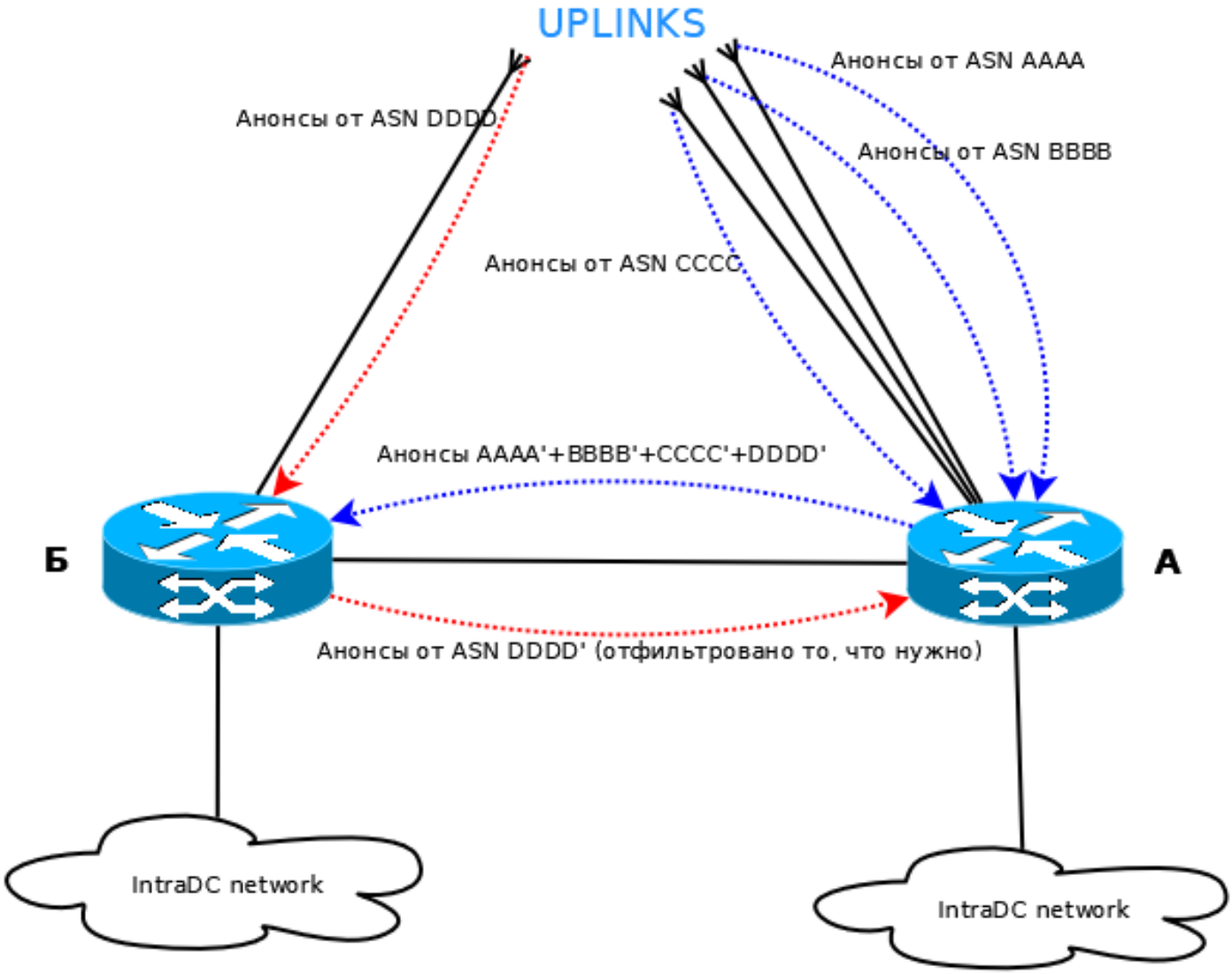

Рис. 4. Принципиальная схема гибридизации (одна из возможных схем) с иллюстрацией возможного обмена маршрутной информацией внутри локации.

фраструктуре узла CDN возникают с приходом реально большой нагрузки. В этой статье мы предложили читателю некоторые практические приёмы, позволяющие лучше адаптировать сетевую инфраструктуру под всю гамму нагрузок в т.ч. в случае, если серверы раздают на больших скоростях в довольно узком коридоре технических требований, реальных условий и финансово-экономических показателей.

Мы постарались совместить и теорию, и практику. Но основным вопросом всё же остаётся метод решения этих задач в реальных условиях реального дата-центра. Сегодня именно операторы сетей доставки трафика сталкиваются с огромными объёмами трафика, которые необходимо эффективно и оптимально распределить по подчас скромным ёмкостям региональных операторов связи. Эта задача требует особого подхода, один из аспектов которого мы попытались изложить в данной статье, совершенно не претендуя на исчерпывающее решение.

Список литературы:

[1] https://www.osp.ru/lan/2000/12/131492

[2] https://en.wikipedia.org/wiki/Multi-chassis_link_aggregation_group

Об авторе

Арсений Валентинович Великород,
старший сетевой инженер CDN-Video, Россия, Москва