Networking
For everyone

Virtual Port-Channel

Темы модуля

- Архитектура vPC
- vPC и смежные технологии
- Вопросы по дизайну
- Восстановление после аварий
- vPC и маршрутизация

Networking
For everyone

Архитектура vPC

Основные компоненты

- Два физических коммутатора
  - их называют “vPC peer”
- vPC домен
  - domain ID
- Интерфейс peer-keepalive
  - udp ping между пирами
- Интерфейс peer-link
  - синхронизация параметров поверх CFSoE
- vPC интерфейса

vPC домен

- Совокупность из vPC-peer устройств, интерфейса vPC peer-keepalive и всех портов vPC member
- Основной параметр – domain ID
  - используется для формирования LACP ID

Peer-keepalive

- Используется коммутаторами для периодической отсылки keepalive’ов и мониторинга состояния соседа
- Обычный L3 интерфейс
  - между коммутаторами должна быть L3 связность (любая)
- Есть два таймаута
  - hold-timeout
  - timeout
    - стартует в момент, когда счетчик hold-timeout достигает нуля
    - при обнулении secondary берёт на себя роль active

Peer-link

- Коммутаторы договариваются кто будет primary, кто secondary
  - выбор на основе MAC + vPC приоритет
  - по сути неважно кто кем будет
    - за редким исключением ☺
- Обязательно port-channel, минимум 10 Гбит/с

Cisco Fabric Services

- Протокол, используемые для синхронизации различных состояний между vPC-peer коммутаторами
- CFSoverEthernet включается автоматически после включения feature vpc
- *show cfs application*

Compatibility параметры

- Большинство настроек двух коммутаторов в vPC домене должны быть идентичны
- Конфигурация сравнивается посредством CFSoE
- Проверка реализована с помощью двух типов
  - тип 1
  - тип 2

vPC проверка первого типа

- В случае несоответствия настроек первого типа принимаются радикальные меры
- Для глобальных настроек все vPC интерфейсы переходят в состояние down
- Для настроек интерфейса, соответствующий vPC интерфейс переходит в состояние down

Глобальные параметры первого типа

- Режим STP
  - RPVSTP или MST
- Включение STP для VLAN
- Настройка MST региона
  - имя региона
  - номер ревизии
  - соответствие VLAN - регион
- Глобальные настройки STP
  - loop guard
  - bridge assurance
  - BPDU filter
  - и пр.

Интерфейсные параметры первого типа

- Режим агрегации
  - on/off/active/passive
- link speed/duplex
- Режим trunk
- Режим STP
- STP region в случае использования MSTP
- STP port type
- loop/root guard
- MTU

Параметры второго типа

- MAC aging timer
- ACL config
- QoS
- Port Security
- CTS
- DHCP snooping (trust/untrust)
- DAI
- IPSG
- HSRP/GLBP
- PIM

Graceful проверка настроек vPC

- При разнице конфигураций, vPC переходит в состояние suspend
- При включении graceful проверки, только интерфейсы на secondary коммутаторе перейдут в состояние suspend

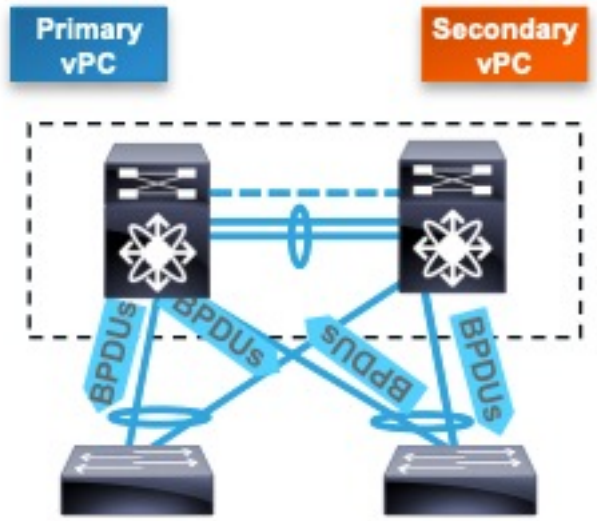
vPC и смежные технологии

vPC Peer-Gateway

- Позволяет свитчу обрабатывать пакеты, которые адресованы на MAC-адрес vPC соседа
- Оптимизация использования peer-link
- Отключается ip redirect на всех SVI, которые относятся к vPC домену
- MAC соседа устанавливается с флагом G
- Основное предназначение – балансировщики
  - F5 Auto Last Hop

vPC Peer-Switch

- В целом vPC не отменяет необходимости в STP
- Один из 2-ух коммутаторов должен быть корневым
  - При его отказе – конвергенция STP
- Peer-switch позволяет оба коммутатора видеть как один
- Используется общий Virtual Bridge ID
  - vPC system MAC + Priority



vPC Pseudo Node

- А что если устройства подключаются не только по vPC, но и по классическому STP?
- В случае Peer-switch оба коммутатора становятся корневыми
  - нет возможности настроить балансировку ☺
- STP pseudo-information позволяет независимо настраивать STP priority для устройств, подключенных по STP

vPC и LACP

- Чаще всего устройства к vPC подключаются по LACP
  - необходим один LACP ID
- Используется vPC system MAC + domain ID

vPC и ARP

- Посредством CFSoE, vPC устройства могут синхронизировать между собой ARP записи

vPC и маршрутизация

- С т.з. L3 два vPC устройства представляют собой два независимых маршрутизатора
- Чаще всего требуется OSPF/BGP соседство между ними
- С т.з. FHRP всё остается как и было – Active/Standby на control plane

vPC и Orphan интерфейсы

- Orphan – интерфейс, подключенный не по vPC
- Необходимо дополнительное внимание к данным интерфейсам, т.к. есть нюансы при поломках vPC домена

Вопросы по дизайну

Общие рекомендации

- Рекомендации могут меняться в зависимости от платформы (напр. N9K vs N7K)
- Всё что можно сконфигурировать явно – сконфигурировать
- Не забывайте про таймеры

Выбор Peer-Keepalive

Модульные коммутаторы

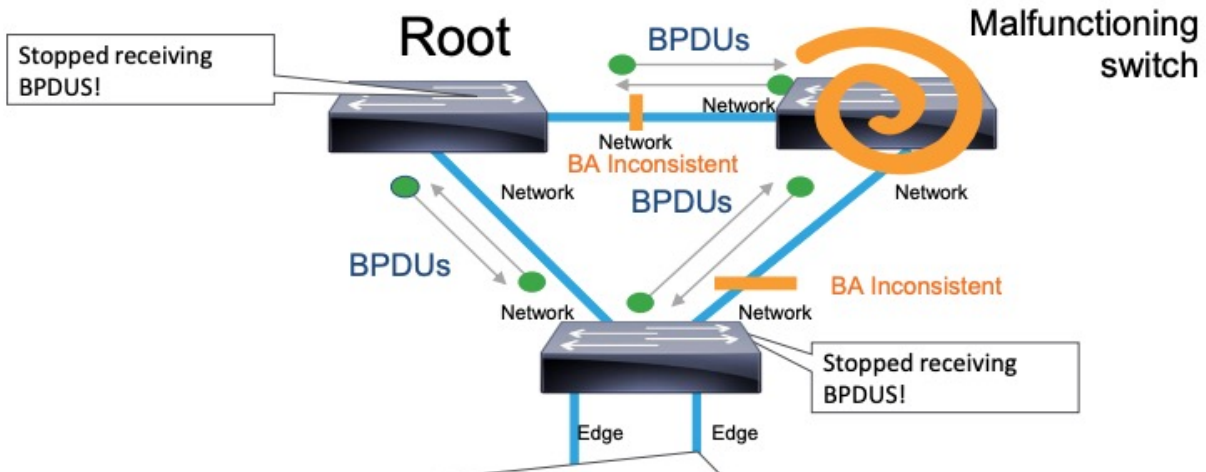
- Выделенный интерфейс
- mgmt0
  - не подключать back-to-back
- L3 инфраструктура

Фиксированные коммутаторы

- mgmt0
- Выделенный интерфейс
- L3 инфраструктура

Bridge Assurance

- STP BPDU – однонаправленная передача пакетов
- Включение BA позволяет коммутаторам обмениваться списком VLAN на интерфейсе
  - проверка двунаправленной связности
- Включать только на peer-link



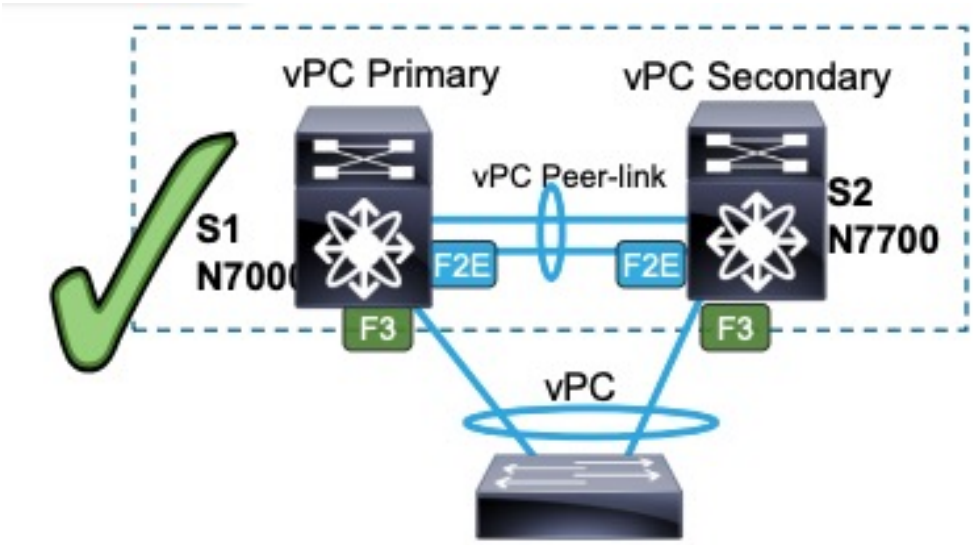

```
%STP-2-BRIDGE_ASSURANCE_BLOCK: Bridge Assurance blocking port Ethernet2/48
VLAN0700
switch# show spanning vl 700 | in -i bkn
Eth2/48          Altn BKN*4         128.304  Network P2p *BA_Inc
```

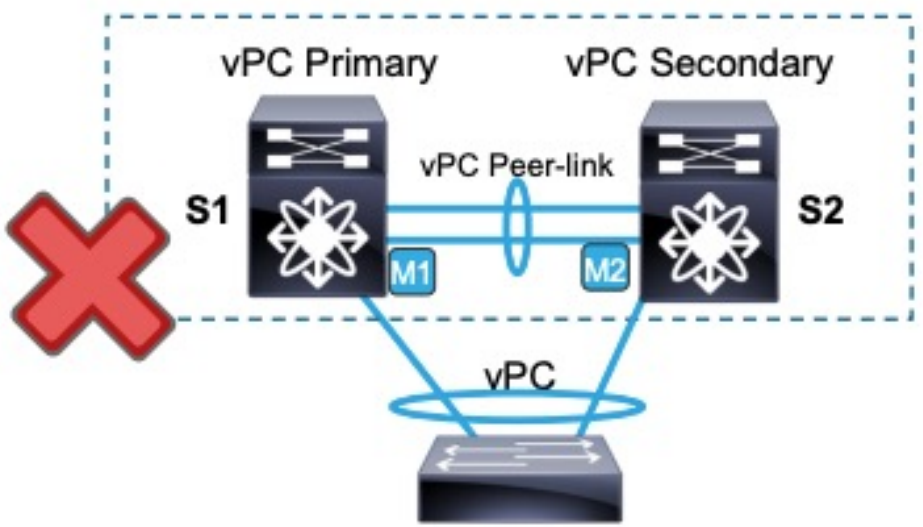
UDLD

- UDLD – отличная технология сама по себе
- Однако
  - нет смысла включать совместно с LACP
  - нет смысла включать на peer-link совместно с BA
  - можно включать на пользовательских портах
    - normal mode

Типы шасси

- vPC соседи должны быть из одной линейки устройств
  - N7000 и N7700 – Ок
  - N5500 и N5600 не поддерживается
- Идентичные линейные карты

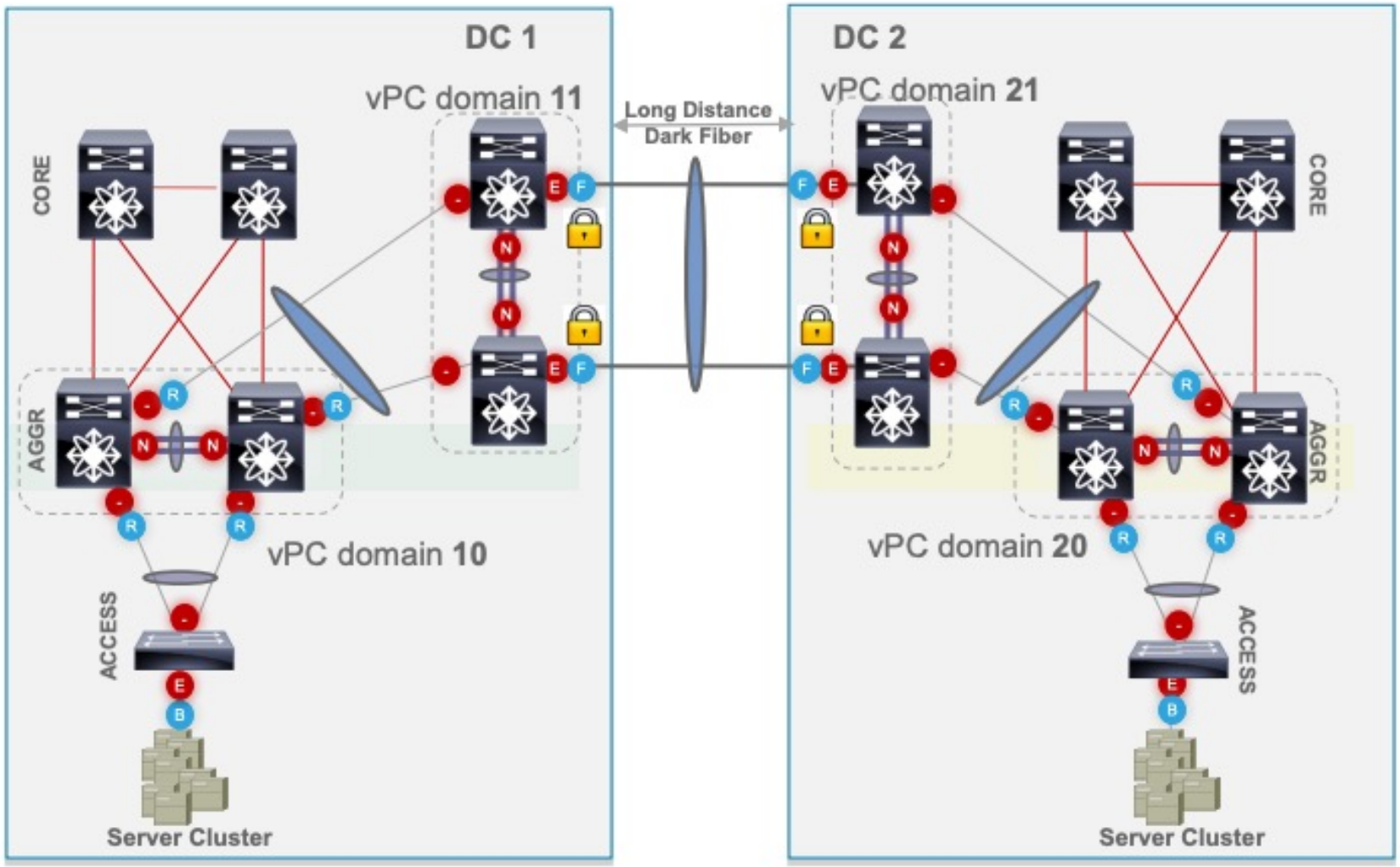




Networking
For everyone

Организация DCI

DCI с помощью vPC



Основные моменты

- Domain ID должны быть разные
- Включить BPDU Filter на DCI интерфейсах
  - изоляция STP
- Включить Portfast на DCI интерфейса
  - STP type Edge Trunk
- Изоляция HSRP
  - оптимизация исходящего трафика
  - независимые, активные HSRP пары с каждой стороны
- Что делать с входящим трафиком?

Networking
For everyone

Неполадки в vPC

Отказ peer-link

- Отказ peer-link может привести к ситуации split-brain
  - ситуация может привести к L2 петле
- Для того, чтобы не было единой точки отказа, устройства проверяют доступность друг друга по peer-link и peer-keepalive
- Поведение vPC домена на отказ peer-link зависит от того, работает peer-keepalive или нет

Отказ peer-link

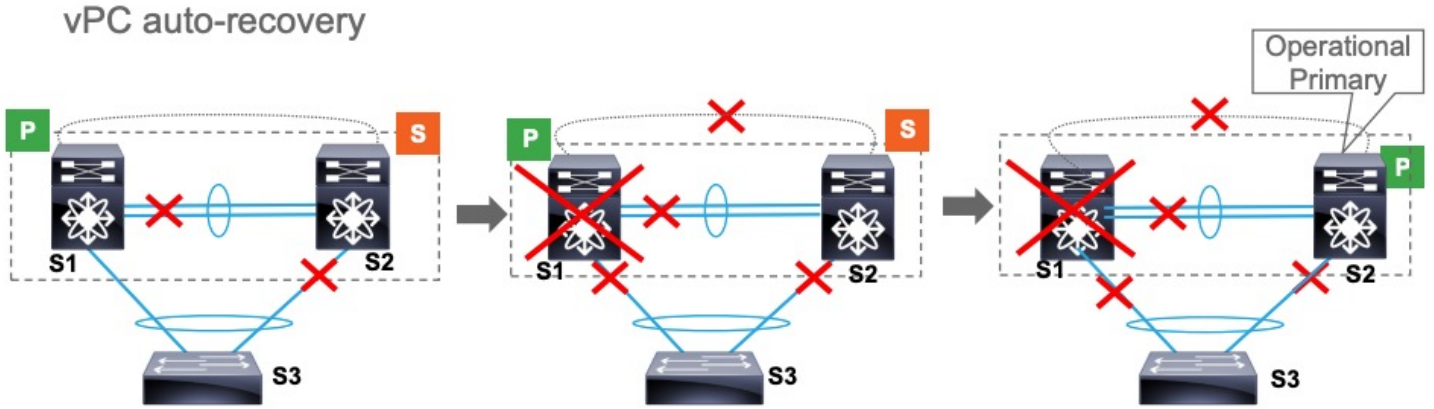
- vPC secondary постоянно проверяет доступность vPC Primary через PKL
  - vPC primary доступен
    - secondary устройство отключает все свои vPC интерфейсы
    - secondary устройство отключает все свои SVI интерфейсы
  - vPC primary недоступен
    - secondary забирает на себя роль Primary
- Никогда не используйте peer-link как транспорт для PKL

Восстановление после сбоя

- Чаще всего оба vPC устройства устанавливаются в одну или соседние стойки
- Предположим, что в комнате ЦОД пропало питание и один из коммутаторов «умер»
  - оставшийся коммутатор включается
  - PKL находится в статусе “never come up”
  - peer-link не может быть инициализирован
  - vPC member интерфейсы не могут быть инициализированы
  - серверы не могут получить доступ к инфраструктуре
- Решение auto-recovery
  - если peer-link не инициализируется в течение таймаута, коммутатор берёт на себя роль Primary

Восстановление после сбоя

- Auto-recovery помогает и в случае сбоя peer-link
  - peer-link переходит в состояние Down, однако PKL живой
  - vPC secondary переводит свои vPC интерфейсы в состояние Down
  - vPC primary полностью выходит из строя
  - vPC secondary не активирует свои vPC интерфейсы
- При включении auto-recovery
  - после падения PKL, Secondary забирает на себя роль Primary
  - vPC secondary активирует свои vPC интерфейсы



Изоляция Orphan портов

- Возможная проблематика:
  - vPC домен выступает в качестве шлюза для серверов (SVI)
  - orphan порты подключены к vPC secondary
  - падает peer-link, PKL жив
  - vPC secondary гасит свои vPC интерфейсы и SVI
  - orphan порт остаётся в состоянии UP
    - изолирован от SVI

Изоляция Orphan портов

- Можно рассмотреть несколько решений проблемы
  - не использовать orphan подключения
    - все сервера подключаются к двум vPC коммутаторам
    - идеальное решение
  - подключение через промежуточный коммутатор, который подключен к двум vPC коммутаторам
  - использовать не vPC VLANы
  - не отключать SVI на vPC secondary

Изоляция Orphan портов

- Иногда подключаемые устройства используют свои интерфейсы по принципу Active/Standby
- Решение проблемы
  - использовать Active/Active ☺
  - на vPC коммутаторах настроить выключение orphan портов

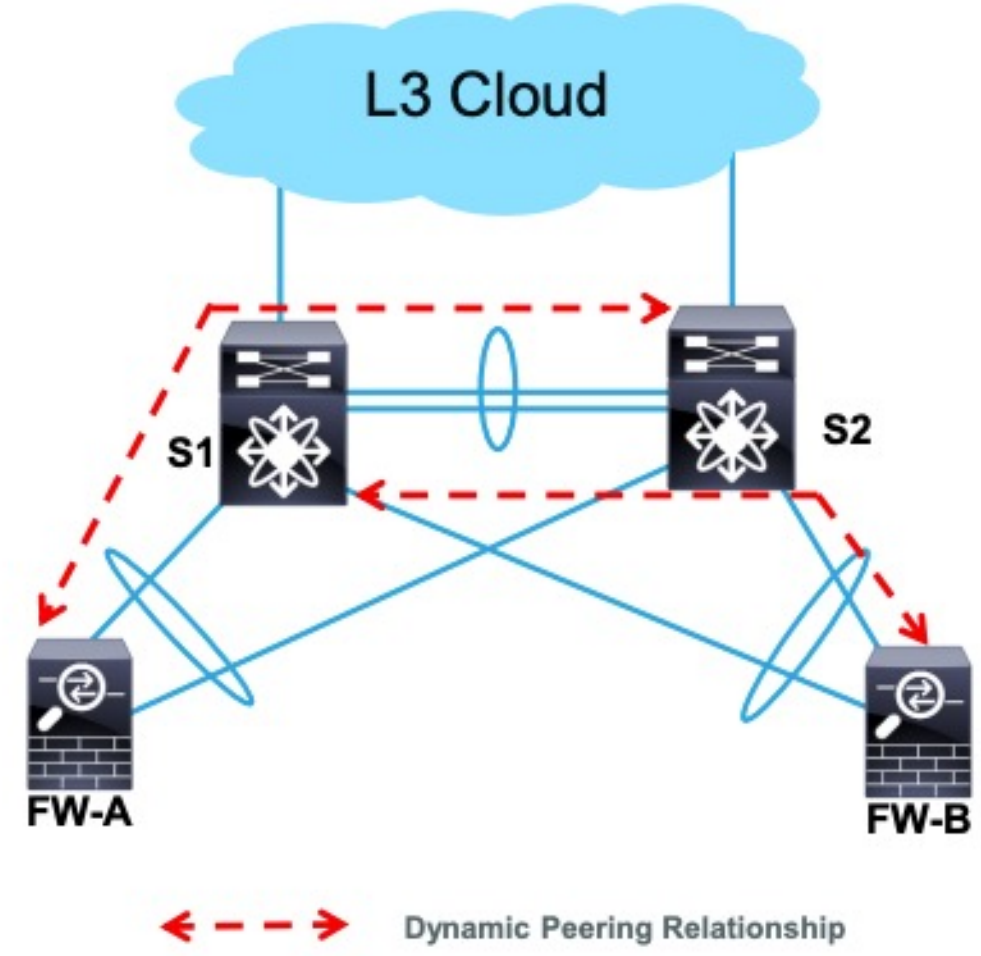
vPC и маршрутизация

vPC и маршрутизация

- С т.з. L3, vPC коммутаторы представляют собой два абсолютно независимых устройства
- Технически, можно не смешивать L2 и L3
- Использовать для L3 отдельные интерфейсы

Маршрутизация поверх vPC

- Иногда есть необходимость настроить маршрутизацию поверх vPC
- При прохождении через peer—link уменьшается TTL
  - как результат – невозможно построение OSPF соседства



Маршрутизация поверх vPC

- Решение – отключить уменьшение TTL
- В зависимости от платформы, поддерживаются различные сценарии
- Подробнее: https://www.cisco.com/c/en/us/support/docs/ip/ip-routing/118997-technote-nexus-00.html

Networking For everyone