



Отказоустойчивый ЦОД за 5 дней

День 4: сеть и сервисные устройства

Илья Дрей
idrey@cisco.com

Хаванкин Максим
mkhavank@cisco.com

4 февраля 2021

Обсуждение – в Telegram канале: <https://t.me/CiscoDR>

Содержание

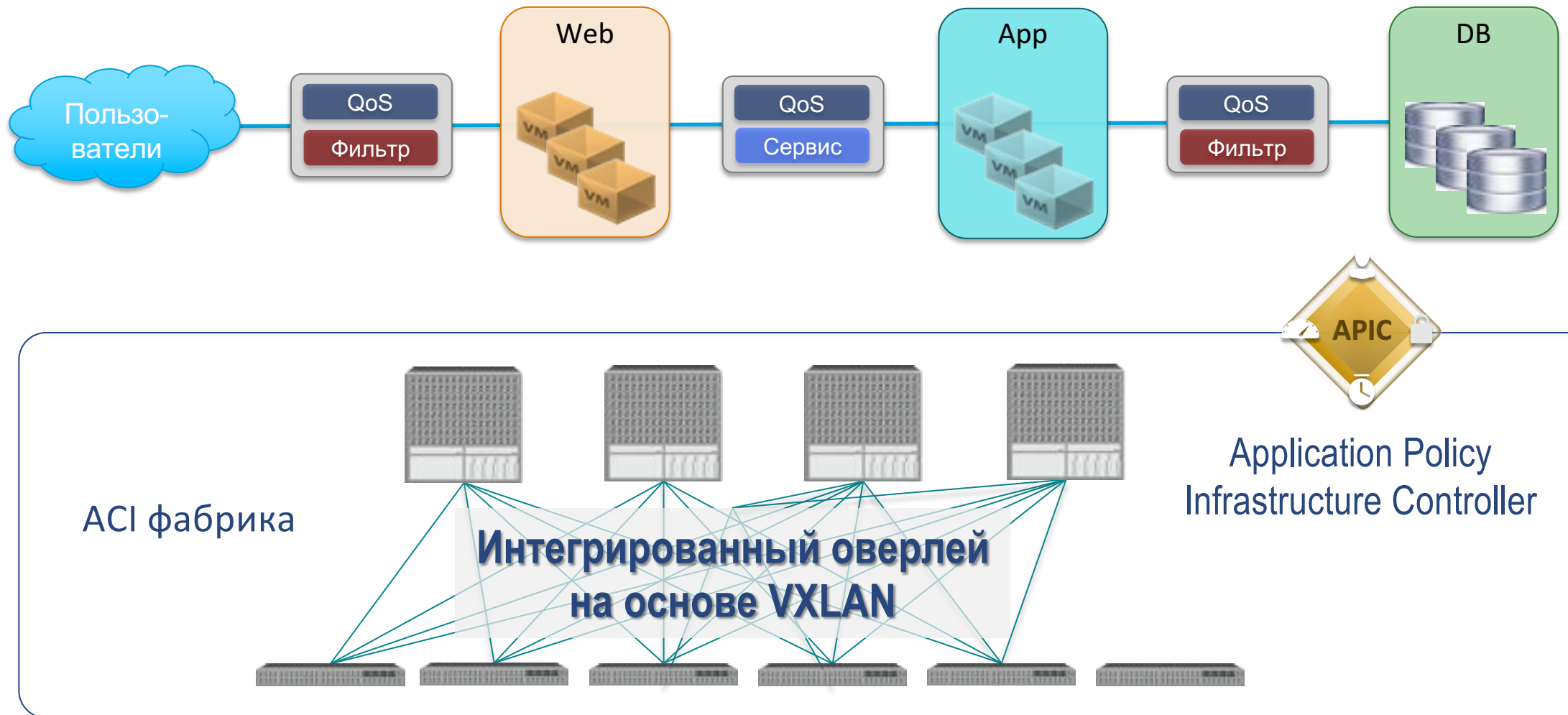
- Часть 1 - Распределенные фабрики на базе Cisco ACI
- Часть 2 – Распределенные фабрики на базе VXLAN & EVPN
- Часть 3 – Организация сетевых подключений для традиционных ЦОД

Часть 1 - Распределенные фабрики на базе Cisco ACI

АСИ – краткое введение

Application Centric Infrastructure (ACI)

Базовые элементы архитектуры



Cisco ACI

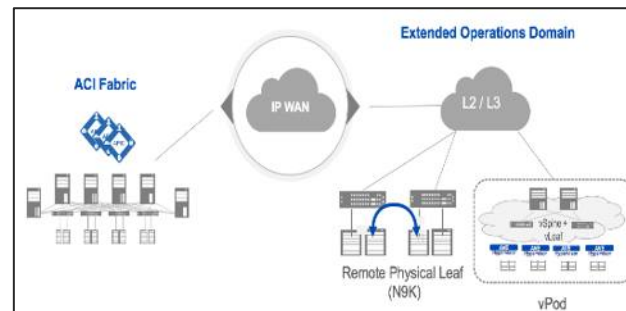
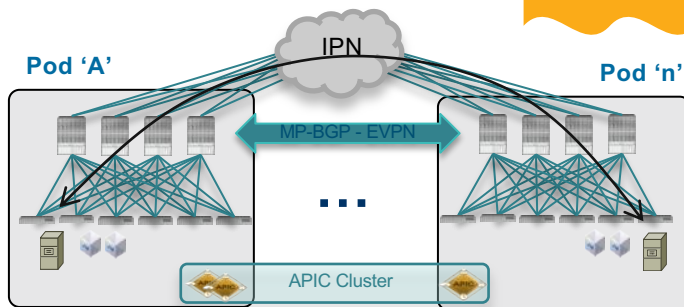
Основные этапы развития решений для объединения ЦОД

ACI фабрика на один Pod



ACI Multi-Pod Fabric

HA



ACI 4.1-5.1 – Зоны доступности в публичных облаках AWS/Azure/GCP*

ACI 1.0 - Leaf/Spine фабрика на один Pod

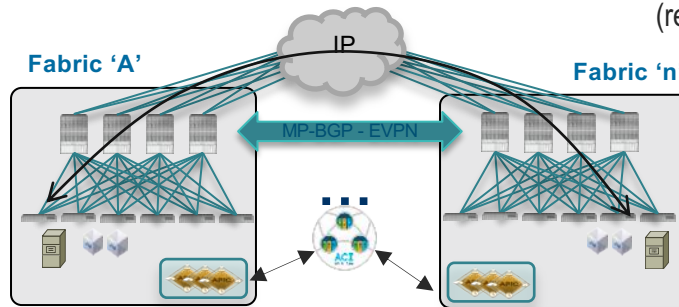
ACI Stretched Fabric



ACI 1.1 – Один географический растянутый Pod

ACI 2.0 – несколько сетей (Pod-ов) в одном домене сбоя – (Availability Zone)

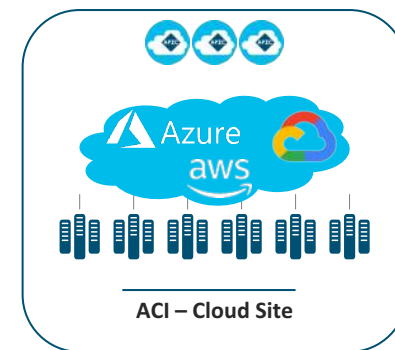
ACI 3.0 – Несколько доменов сбоя (Fabrics) в одном регионе и управление политиками в неск. регионах



ACI Multi-Site

DR

ACI 3.1/3.2 - Remote Leaf и vPod расширяют зоны доступности (Fabric) на удаленные площадки (remote locations)



Материалы Cisco ACI Sprint 2020



Сессия

Материалы

Cisco ACI: первоначальный запуск и базовые операции

[Запись](#), [слайды](#)

ACI для информационной безопасности

[Запись](#), [слайды](#)

Дополнительная часть «Создание сервисных цепочек в ACI»

[Запись](#), [слайды](#)

Интеграция ACI с системами виртуализации и контейнерной оркестрации

[Запись](#), слайды (часть [1](#), [2](#))

Решения для связи ЦОД на основе Cisco ACI

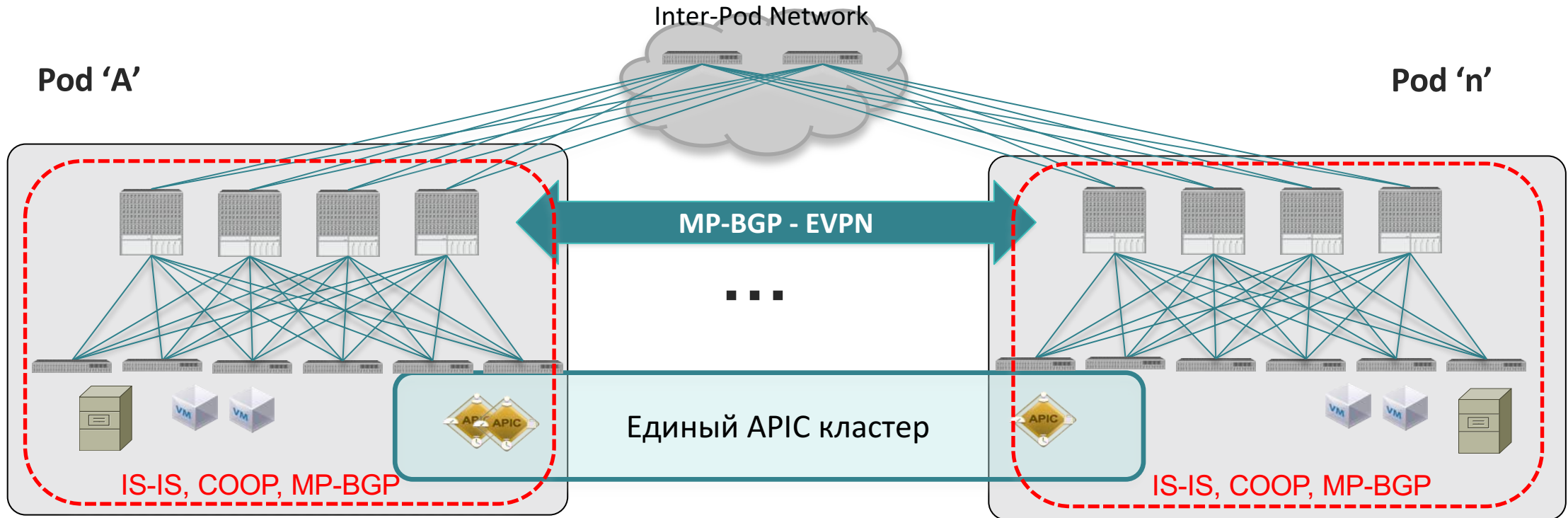
[Запись](#), слайды (часть [1](#), [2](#), [3](#))

Миграция на Cisco ACI. Автоматизация настройки и эксплуатации ACI фабрики

[Запись](#), слайды (часть [1](#), [2](#))

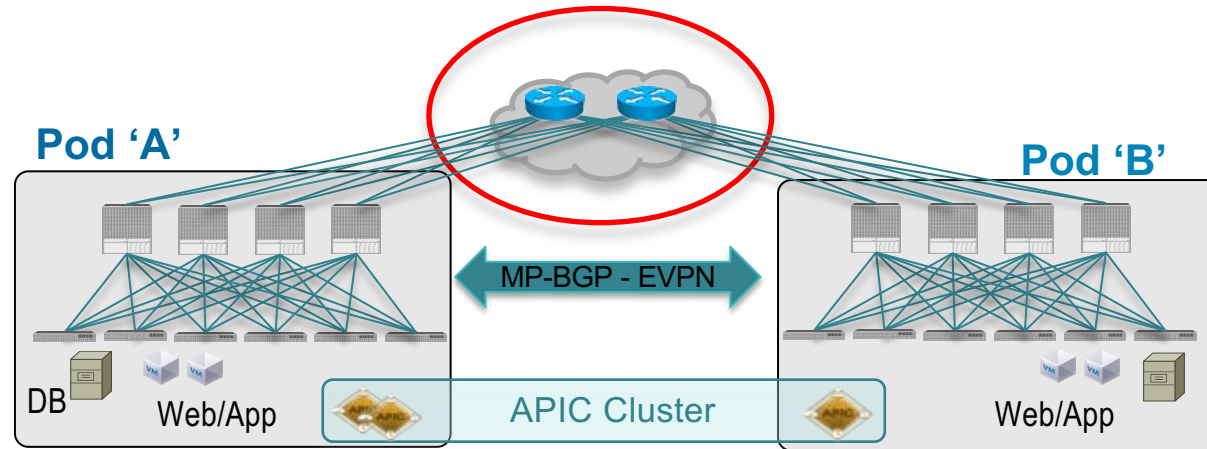
Распределенная сетевая фабрика на основе Cisco ACI Multi-Pod

ACI Multi-Pod



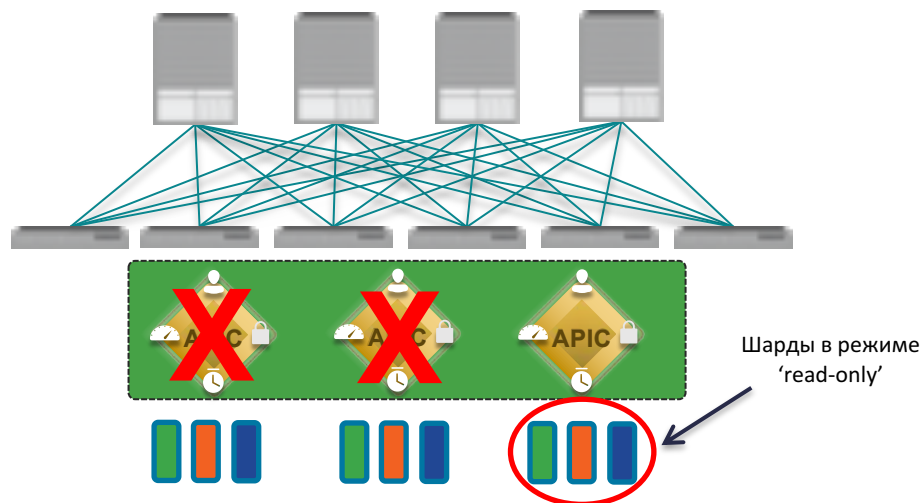
- Несколько модулей (ACI Pod) связанных IP сетью (Inter-Pod network), каждый состоит из набора leaf и spine
- Управлением единым кластером APIC
- Единый домен управления и политик
- Изоляция доменов отказов протоколов control plane (IS-IS, COOP)
- Инкапсуляция VXLAN между модулями
- Сквозное применение политик

ACI Multi-Pod: Требования к Inter-Pod Network (IPN)

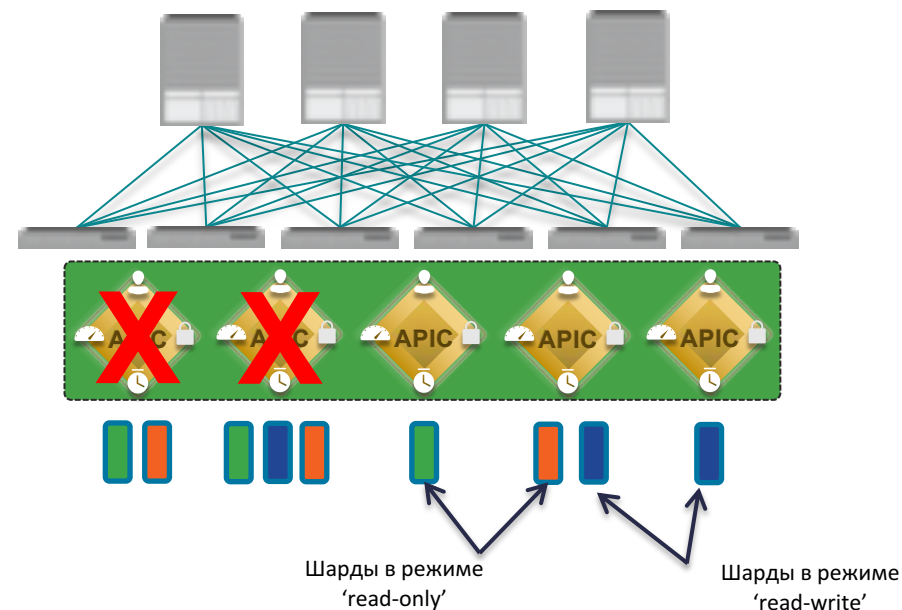


- Не управляется при помощи APIC, требуется отдельная настройка (day-0)
- Топология IPN произвольная, все spine узлы к IPN подключать не обязательно
- Основные требования:
 - ✓ Multicast BiDir PIM для передачи L2 BUM* трафика
 - ✓ OSPF для установления соседских отношений между spine и IPN для обеспечения связности между VTEP
 - ✓ Поддержка Jumbo MTU (9150B) для передачи VXLAN инкапсулированного трафика
 - ✓ DHCP-Relay
 - ✓ noswitchport с поддержкой 802.1Q subinterface

APIC – работа кластера в случае с одним Pod

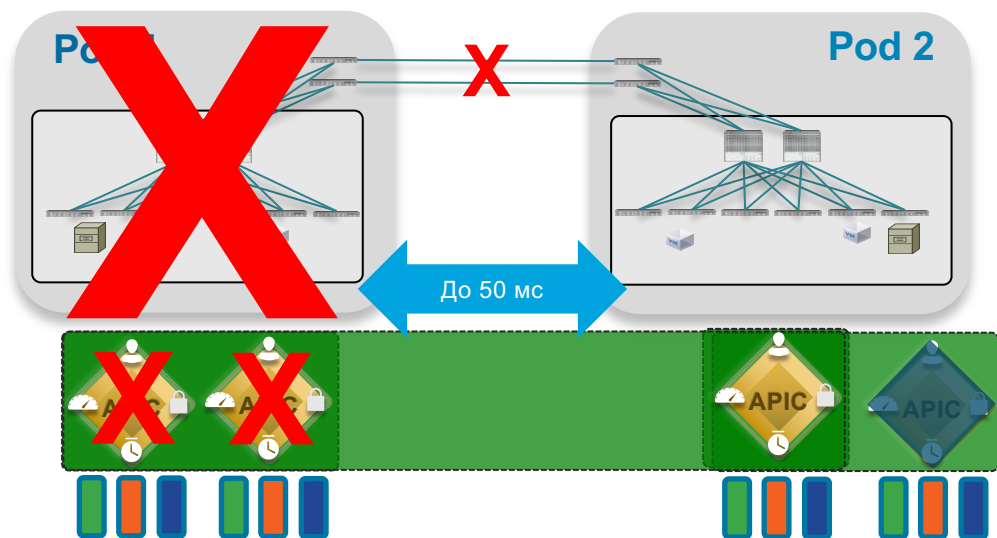


- APIC отключает доступ на запись к Базе Данных, когда только один узел кластер остается активным (standard DB quorum)
- **Аппаратная ошибка** на двух узлах приводит к тому что все шарды (shards) переключаются в режим 'read-only' (по мере восстановления узлов кластер переключается в исходный режим)

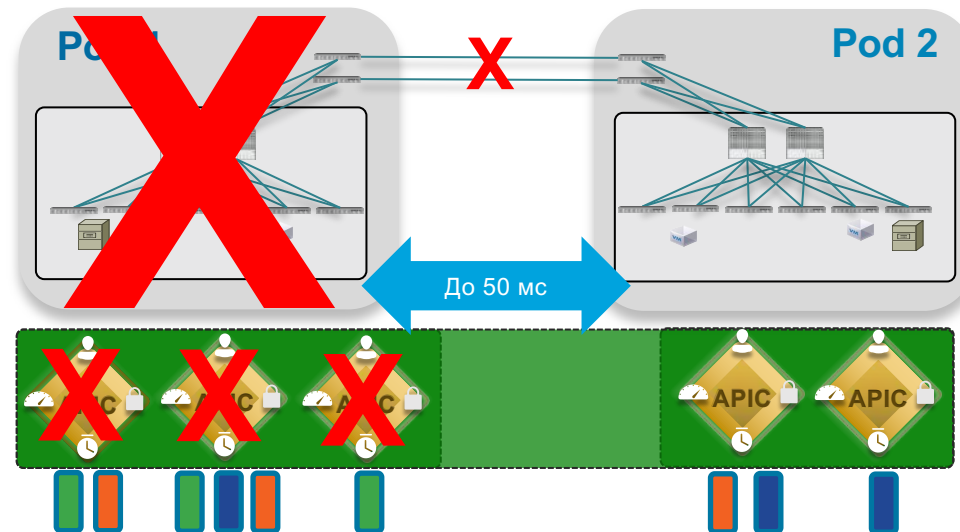


- Дополнительные узлы APIC кластера увеличивают масштабируемость, но не добавляют отказоустойчивости
- **Аппаратная ошибка** на двух серверах приводит к неконсистентному поведению для некоторых шард (некоторые окажутся в режиме 'read-only' некоторые в режиме 'read-write')

APIC – работа кластера в случае с 2-мя Pod



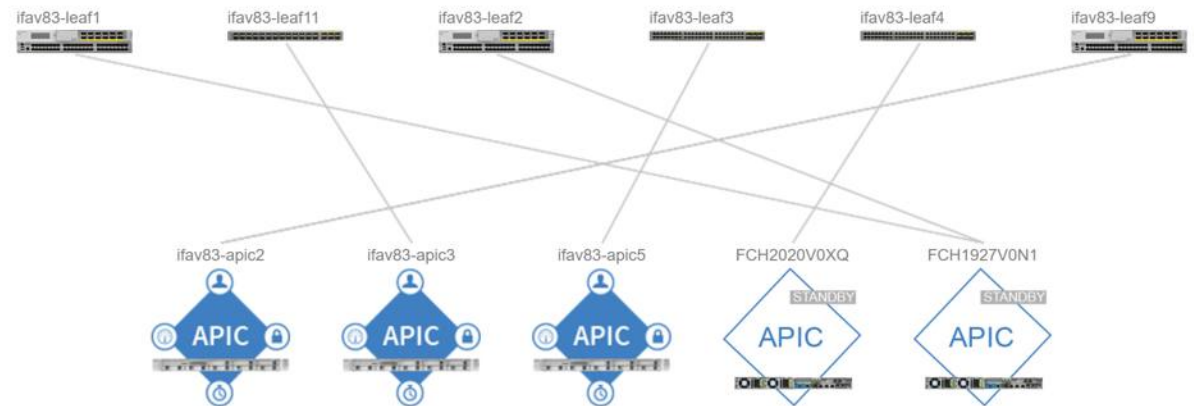
- **Сценарий изоляции Pod:** изменения возможны на узлах в Pod1, но не в Pod2. Кластер восстанавливается после того как Pod-ы восстановят связанность
- **Полный выход из строя одного из Pod:** при полном выходе из строя одного из Pod рекомендуется активировать Standby узел чтобы вернуть возможность вносить изменения в конфигурацию



- **Сценарий изоляции Pod:** такое же поведение как и в случае с кластером на 3 ноды (некоторые шарды могут перейти в режим "read-only")
- **Полный выход из строя одного из Pod:** полный выход из строя Pod1 может привести к потере информации
- Возможно восстановление состояния при наличии снапшота конфигурации ('ID Recovery' procedure – требуется вовлечение BU и TAC)

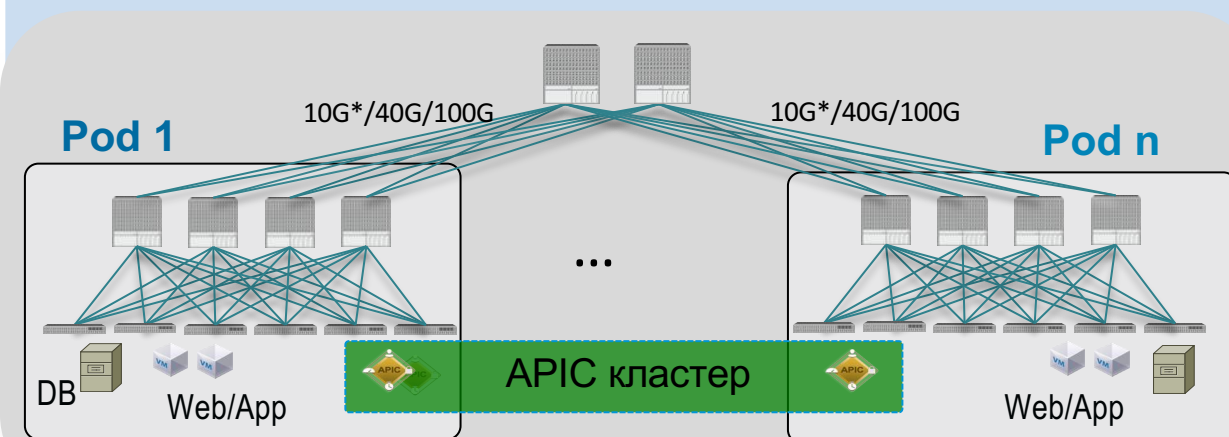
Функция Standby APIC

- Упрощает замену вышедших из строя APIC контроллеров
- Standby APIC используется для замены любого активного узла в кластере
- Не принимает участия в конфигурации или управлении фабрикой до момента активации
- Никакие данные на него не реплицируются, даже admin credentials (Rescue-user логин работает)
- Минимальный рекомендуемый размер кластера - 3
- Поддерживается в Multi-Pod

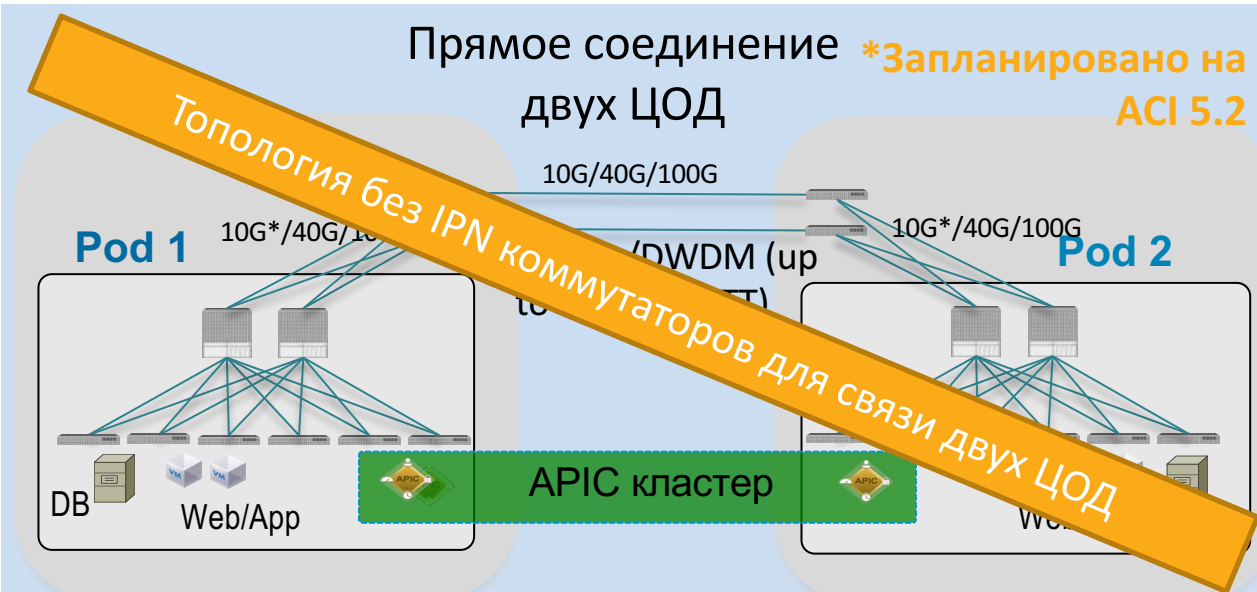


Cisco ACI Multi-Pod: Поддерживаемые топологии

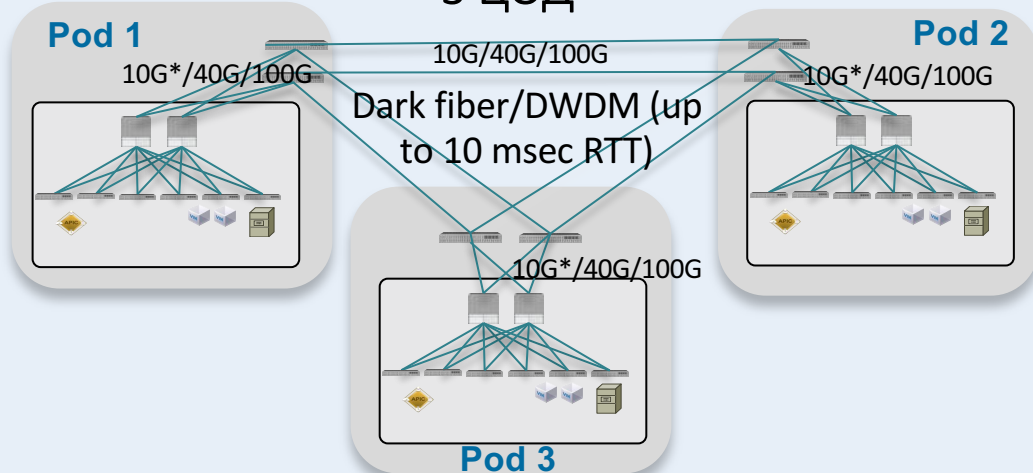
Внутри ЦОД



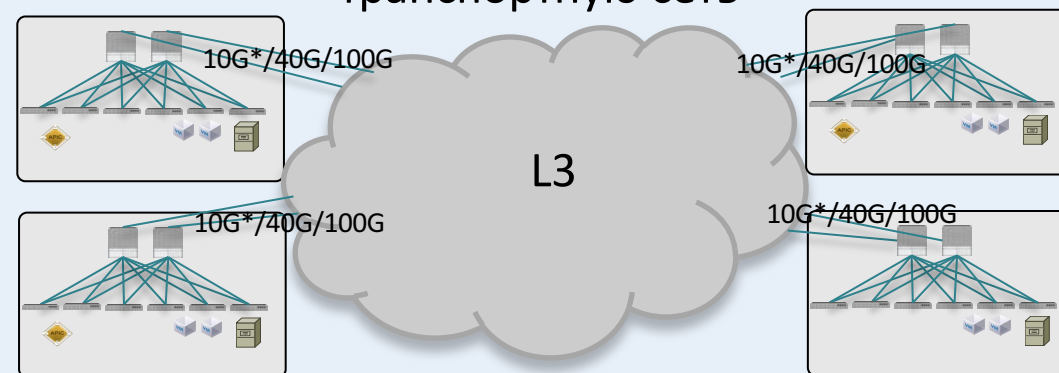
Прямое соединение двух ЦОД *Запланировано на ACI 5.2



3 ЦОД

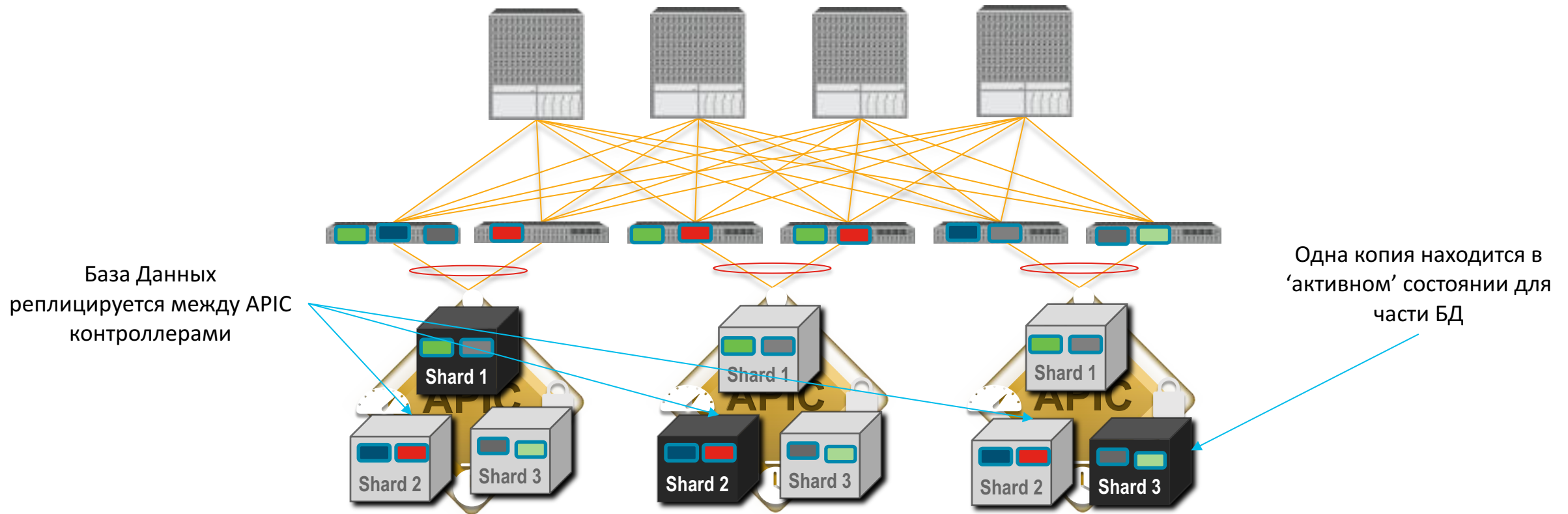


Много ЦОД через транспортную сеть



* 10G только с QSA адаптерами на Spine коммутаторах -EX

APIC – распределенная Multi-Active база данных



- Процессы активны на всех узлах (не active/standby)
- Данные распределяются как активные + 2 резервные копии (shards) для каждого атрибута/объекта/политики

APIC – рекомендации по развертыванию кластера при большом масштабе

- Главное правило гласит: Избегать размещения более двух APIC контроллеров в одном ЦОД
- В случае с двумя Pod и 81-200 коммутаторами рекомендуется использовать кластер 4+1

В случае с фабрикой суммарным размером 200+ Leaf рекомендуется:

	Pod1	Pod2	Pod3	Pod4	Pod5	Pod6
2 Pods*	  	  Standby				
3 Pods	 	 				
4 Pods	 					
5 Pods						
6+ Pods						

Наличие APIC контроллера в каждом Pod не является обязательным



Cisco ACI Multi-Pod для НА?

- Единый домен управления и политик для всех POD
 - Возможность использовать единый VMM домен во всех Pod (один vDS на стороне среды виртуализации, одна ACI CNI интеграция)
- Доступность самых последних функций ACI

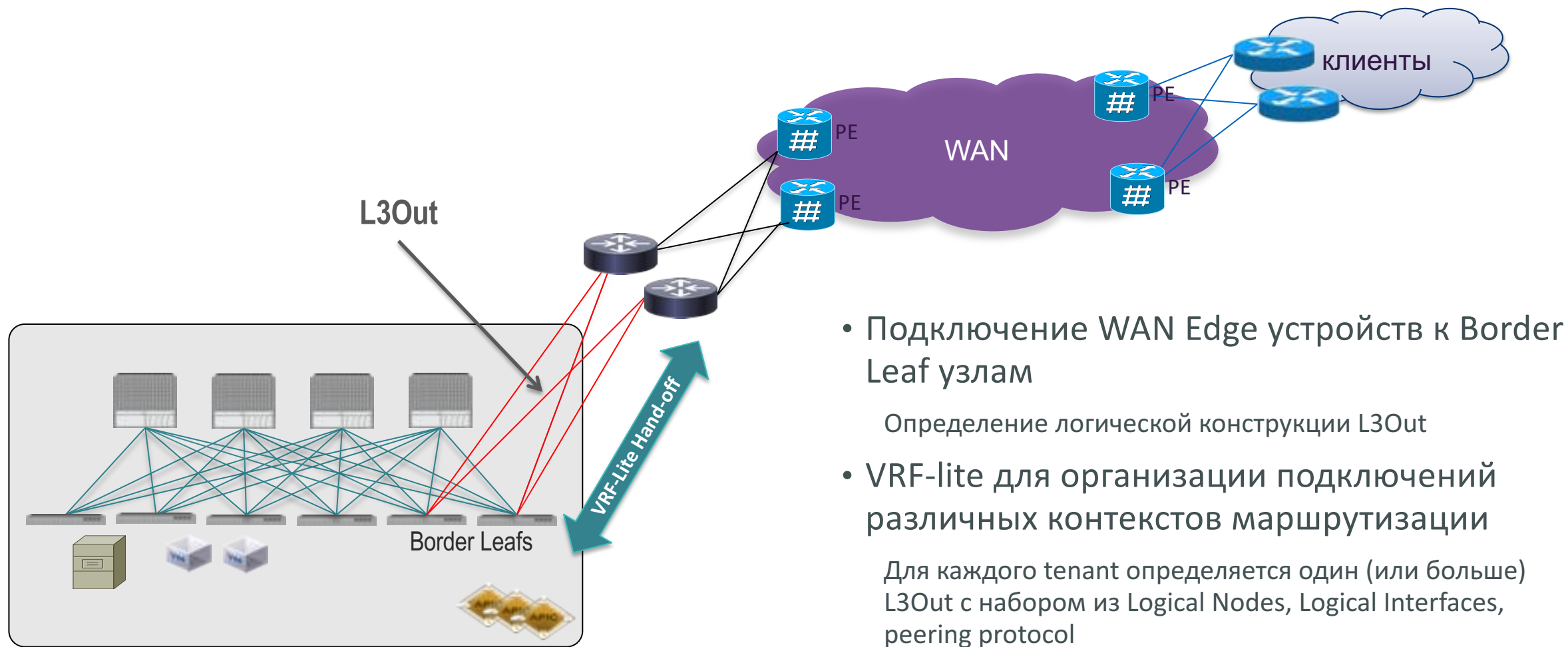
Cisco ACI Multi-Pod для DR?



- Единый домен управления и политик может являться причиной сбоя во всех ЦОД
- Отсутствует возможность гибко управлять распространением BUM трафика внутри ЦОД (POD) и между ЦОД (POD)

ACI Multi-Pod: подключение ко внешним сетям

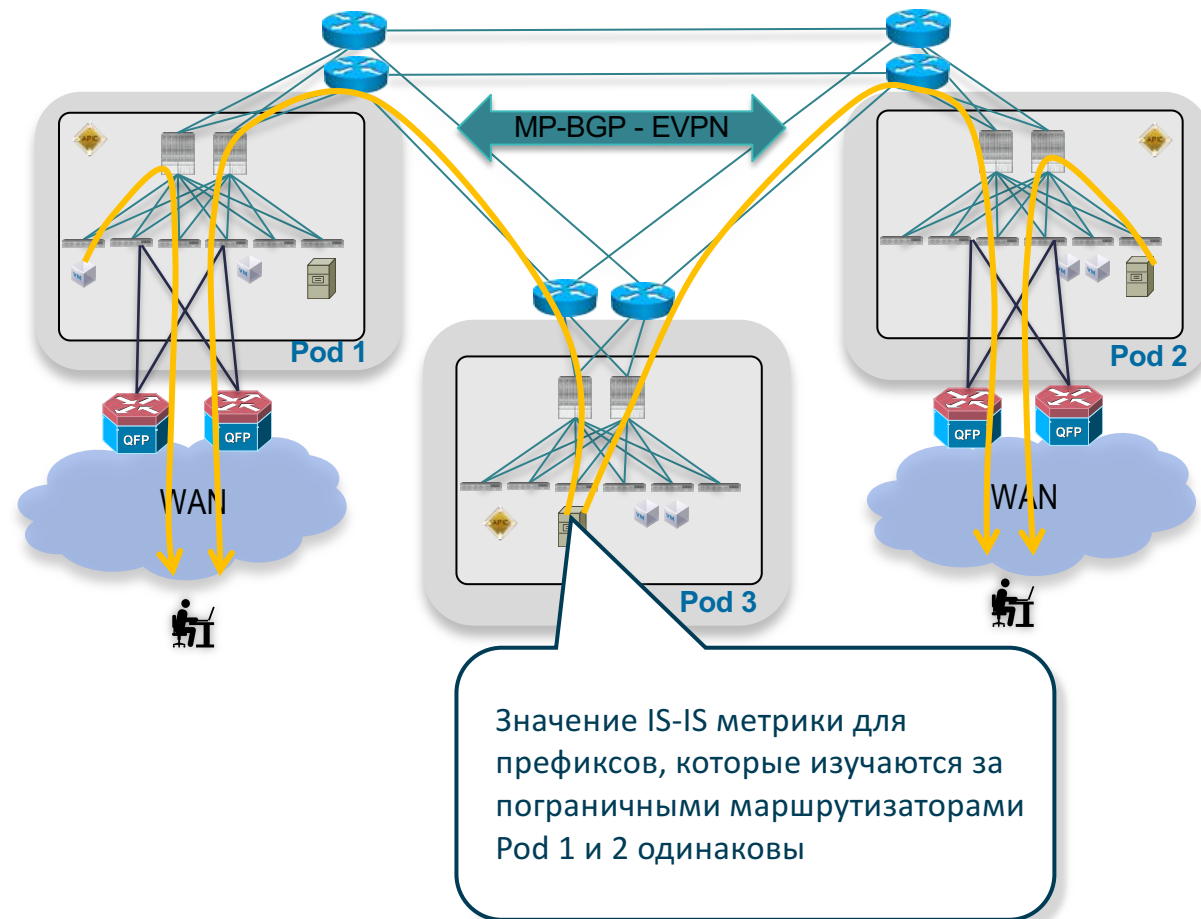
Подключение ACI Multi-Pod к L3 домену 'Традиционный' L3Out на пограничных узлах



Подключение ACI Multi-Pod к L3 домену

Подключение транзитного POD – балансировка по умолчанию

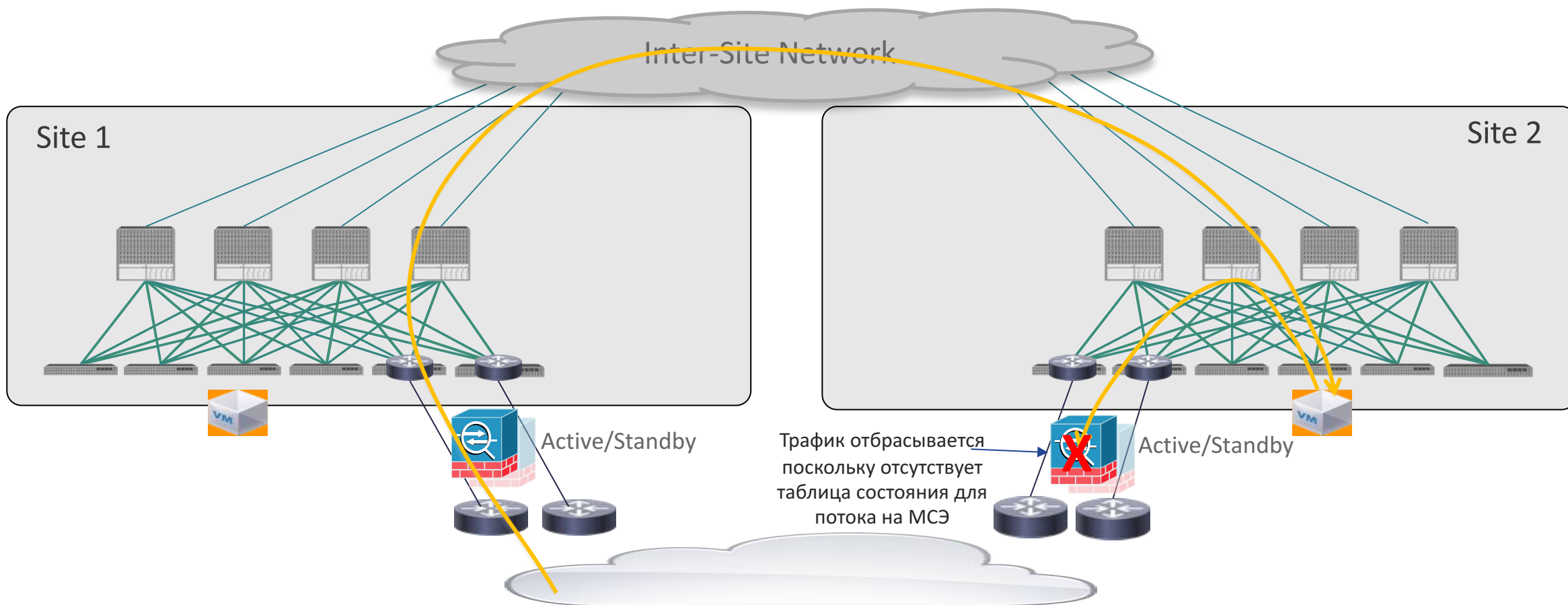
- Не обязательно для каждого Pod иметь выделенное подключение к WAN сегментам (поддерживается сценарий с транзитным Pod)
- Несколько WAN соединений может быть организовано для разных Pod
- Исходящий трафик: по умолчанию все VTEP выбирают WAN подключение в локальном Pod на основании метрики
- Leaf коммутаторы в Pod-е без локального L3Out будут балансировать трафик между различными L3Out в других Pod



АСІ: Проблема асимметрии
для разных типов трафика

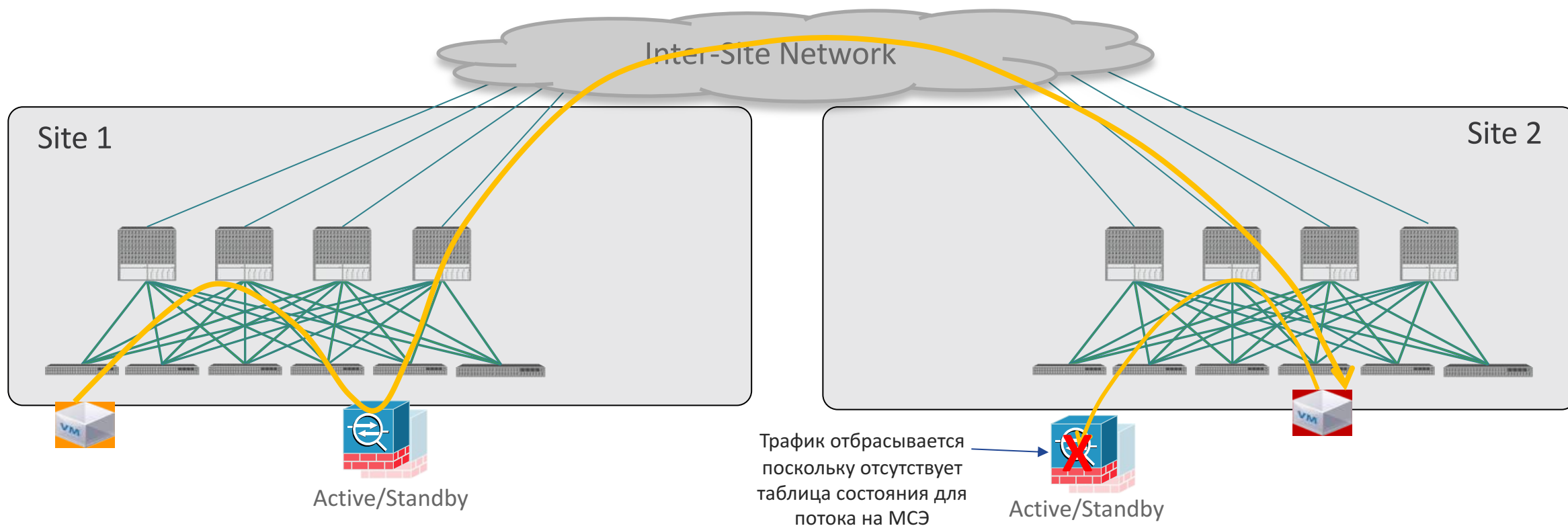
Проблемы асимметрии для трафика «Север-Юг»

Для выхода всегда используется локальный L3out



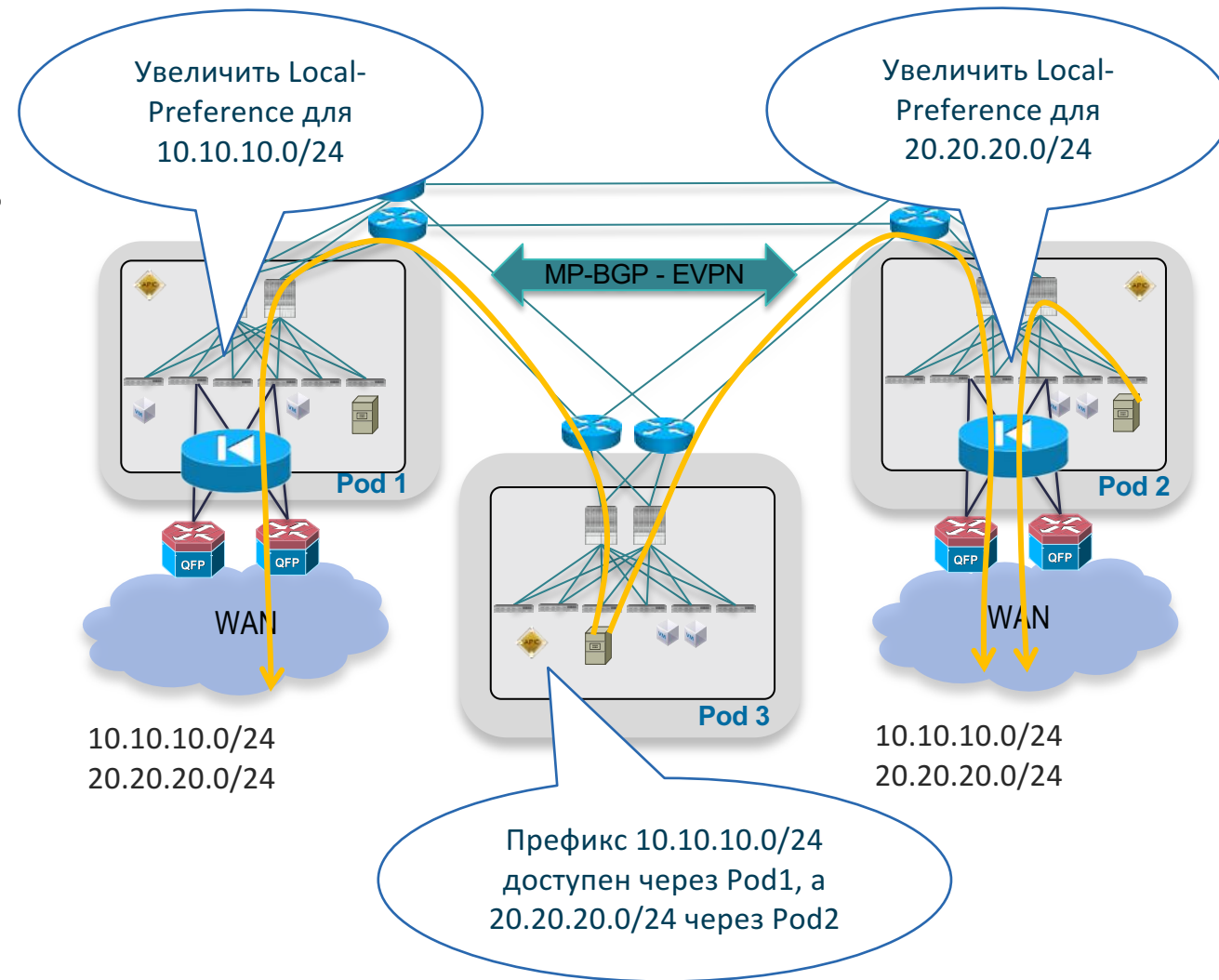
Проблемы асимметрии для трафика «Запад-Восток»

Требуется избегать асимметричных путей при передаче



Управление исходящим трафиком при помощи BGP

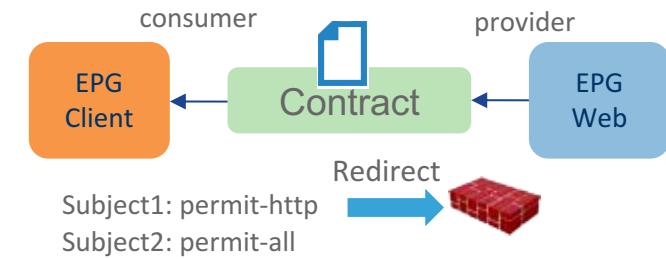
- При организации подключения L4/L7 устройств может понадобиться организовать использование определенного L3Out для доступа ко внешнему сегменту из всех Pod
- Тюнинг нужно производить или на уровне VRF или для специфических префиксов из внешней сети
- Можно использовать route-map для настройки атрибута Local-Preference для префиксов, которые BL получает от WAN Edge маршрутизаторов



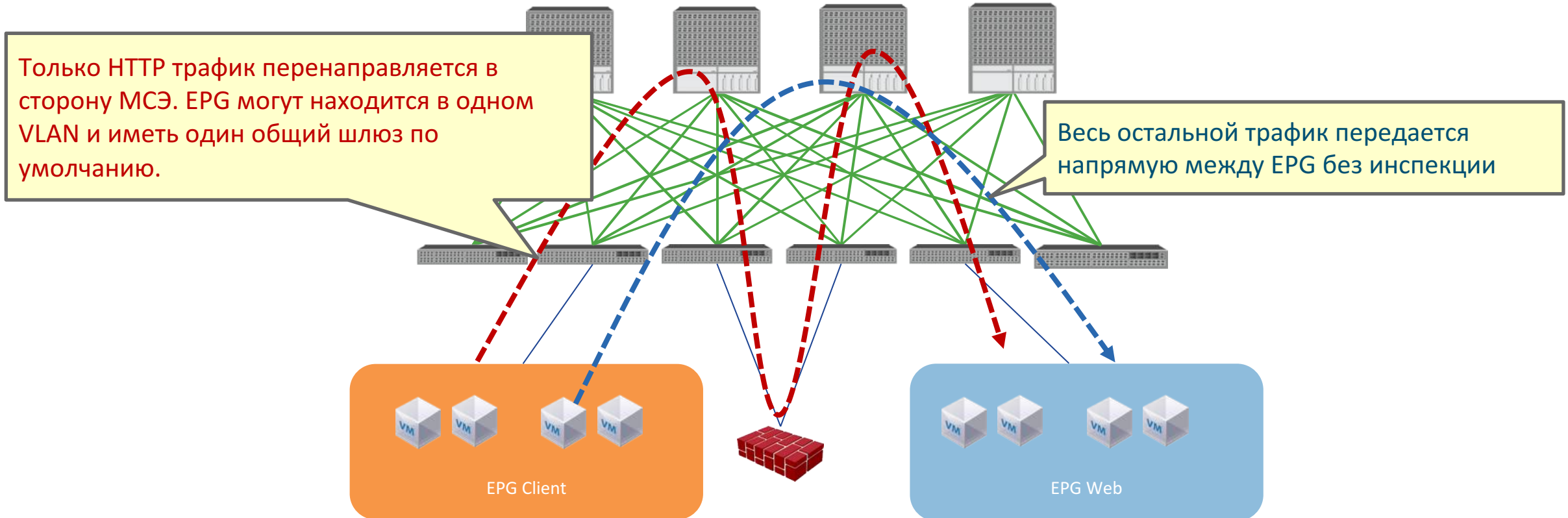
Сервисная цепочка

Policy Based Redirect (PBR)

Выборочная инспекция трафика



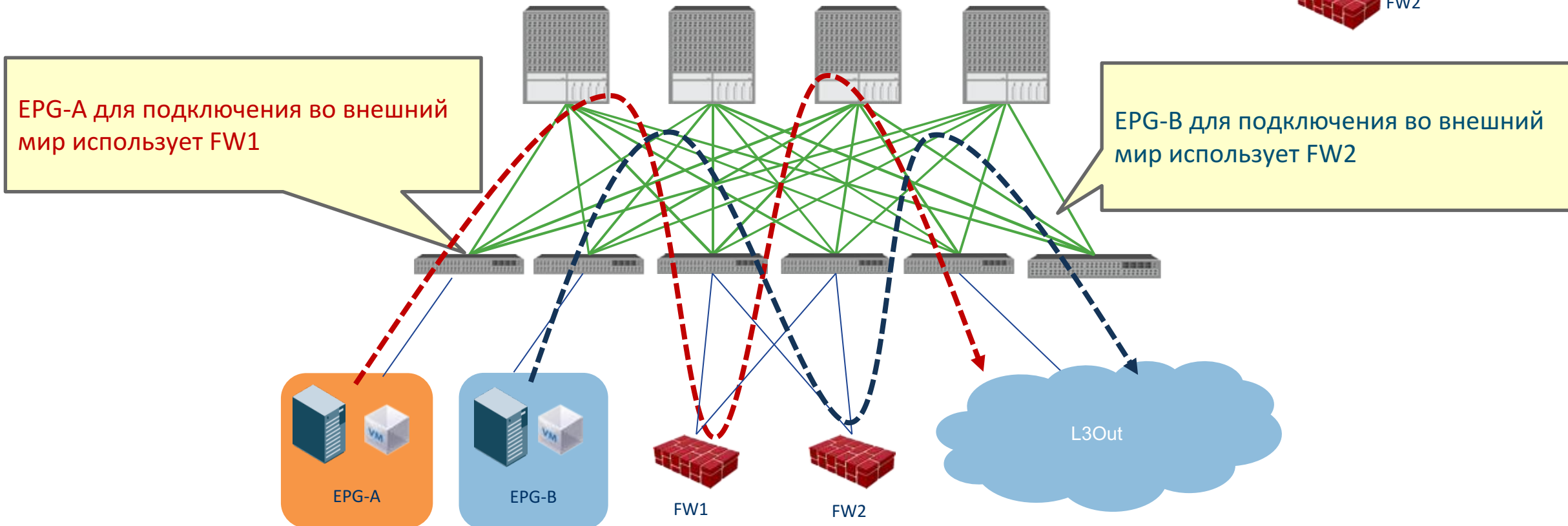
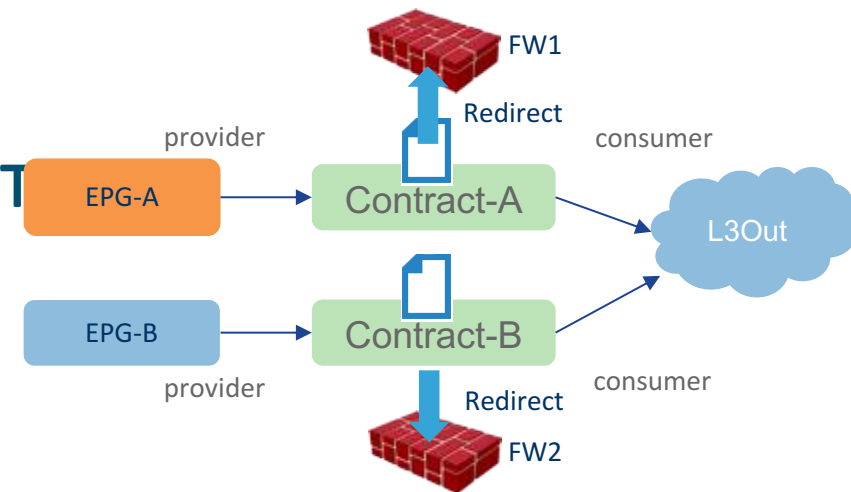
- **Сценарий № 1:** инспекция специфичного трафика при помощи МСЭ между EPG в одном сегменте или в разных сегментах



Policy Based Redirect (PBR)

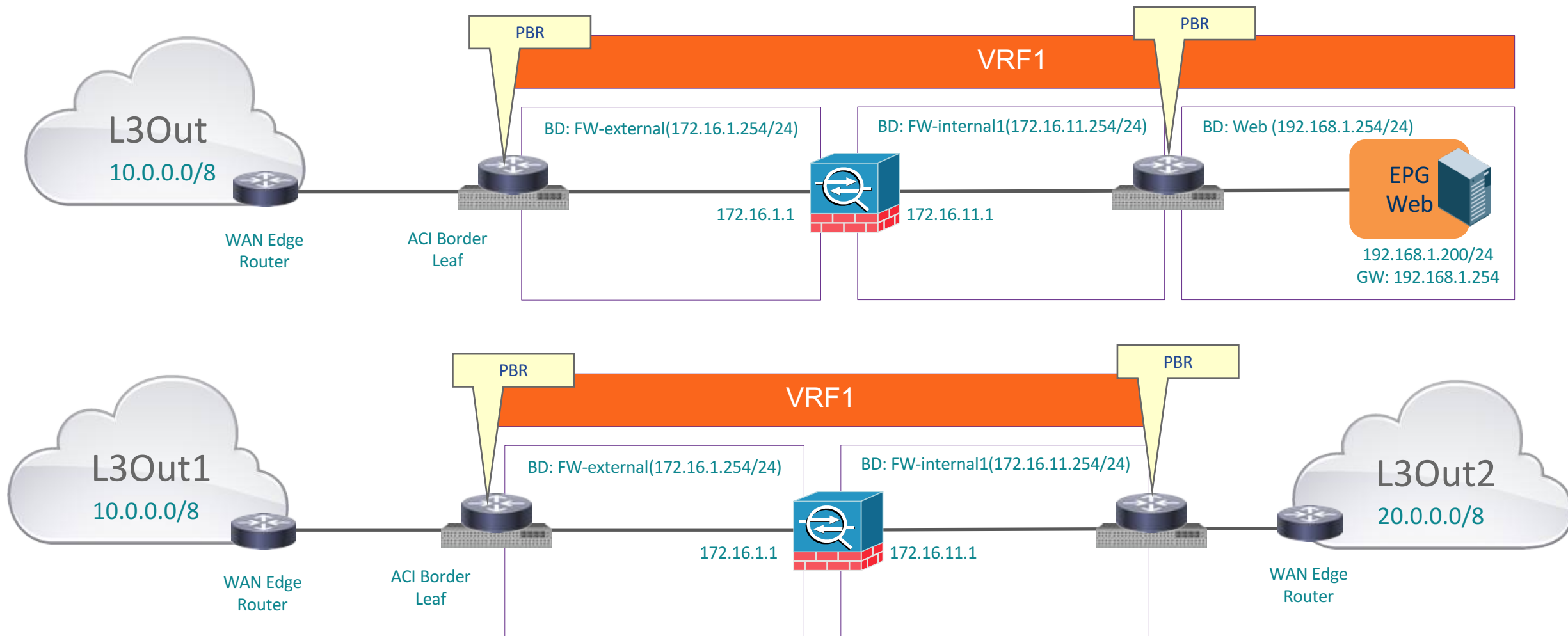
Несколько «выходов» из одного сегмент

- Сценарий № 2: различные МСЭ выбираются в зависимости от EPG, при этом EPG находятся в одном сегменте

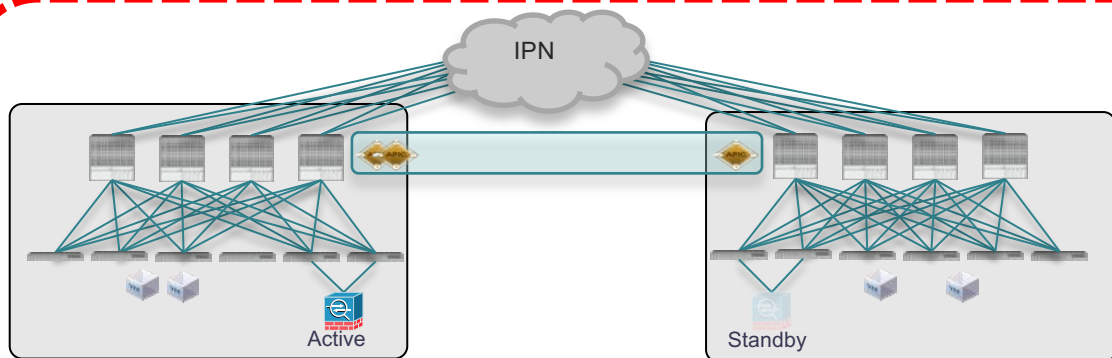


PBR может использоваться для L3out EPG

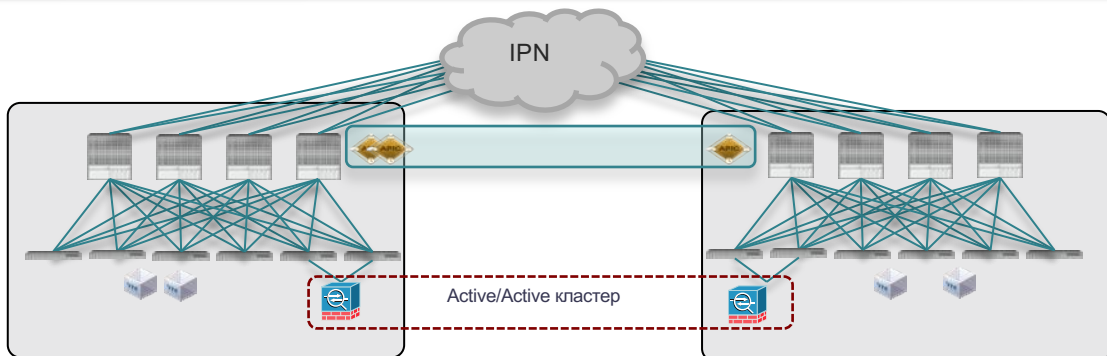
Упрощение традиционных «сэндвич» дизайнов – отказ от множества VRF



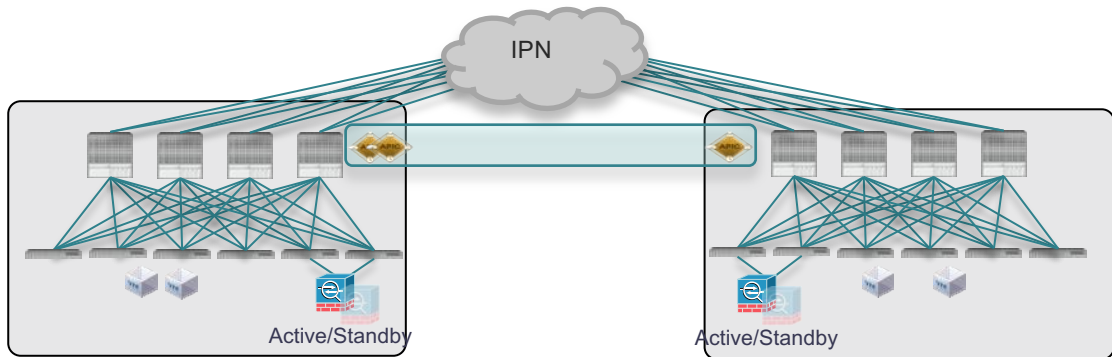
ACI Multi-Pod: Режимы интеграции МСЭ



- Одна Active/Standby пара «растянутая» между Pod
- Отсутствуют проблемы с асимметричными потоками



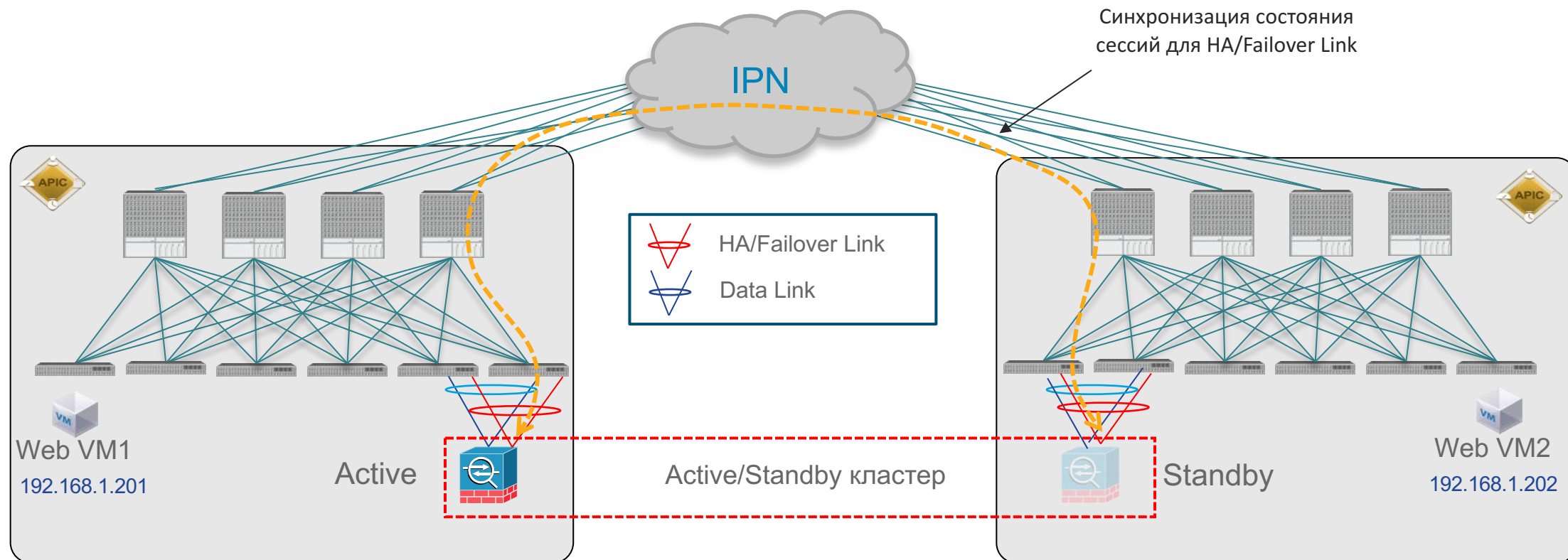
- Active/Active кластер «растянутый» между Pod (один логический FW)
- Требуется возможность подключения к фабрике одного и того же MAC/IP в нескольких POD одновременно
- Поддерживается начиная с версии ACI 3.2(4d) с обязательным использованием сервисного графа и PBR



- Независимые пары Active/Standby развернутые в каждом Pod
- Требуется использовать сервисный граф и симметричный PBR

Одна Active/Standby пара «растянутая» между POD

Растягивание HA/Failover Link при помощи ACI

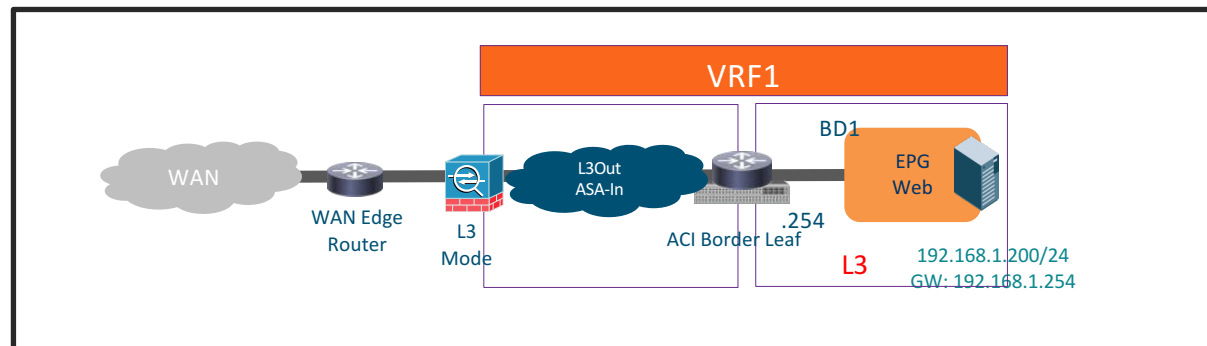


- Рекомендуется использовать отдельные интерфейсы для передачи продуктивного трафика и синхронизации состояния сессий между устройствами
- Связанность для HA/Failover link организуется за счет растягивания выделенных для них EPG/BD между Pod-ами

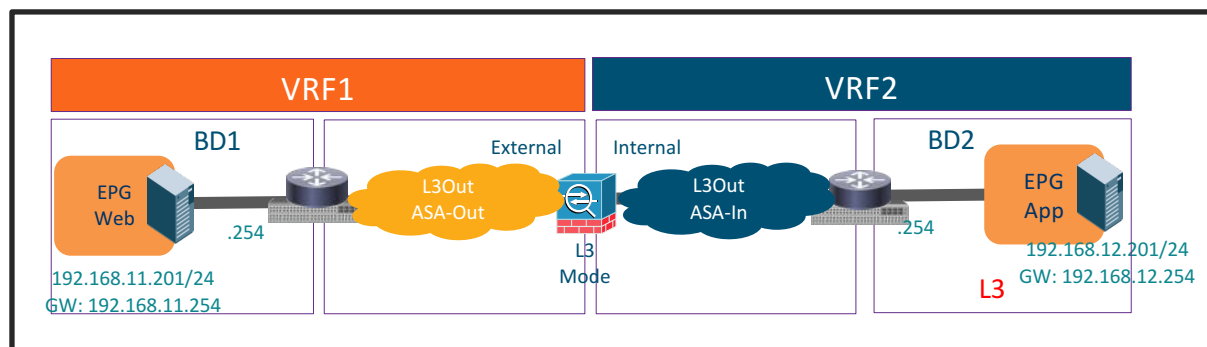
Одна Active/Standby пара «растянутая» между POD

Вариант 1: FW в L3 режиме и L3Out для его подключения к фабрике

North-South



East-West

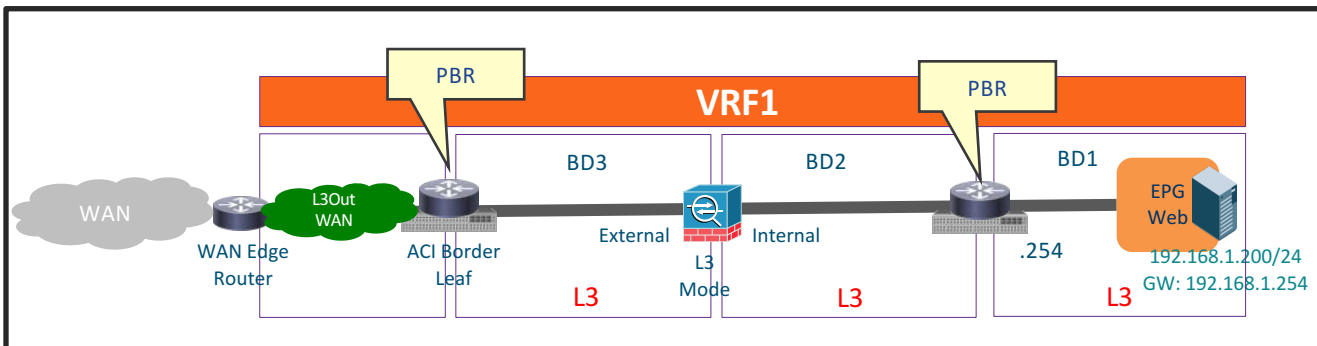


- Шлюз по умолчанию для нагрузки всегда настраивается на ACI фабрике (Anycast gateway)
- Передача трафика через L3Out подключение Active FW обеспечивается за счет использования динамической или статической маршрутизации (VRF-sandwich for East-West)
- Поддерживаются все режимы подключения сервисного устройства (service-graph managed и un-managed, без service-graph)

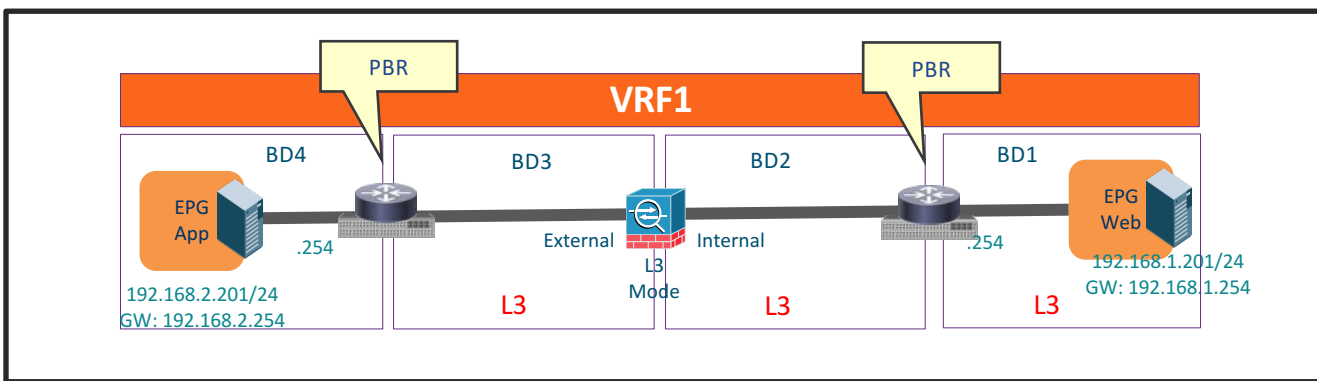
Одна Active/Standby пара «растянутая» между POD

Вариант 2: FW в L3 режиме и PBR для подключения к фабрике

North-South

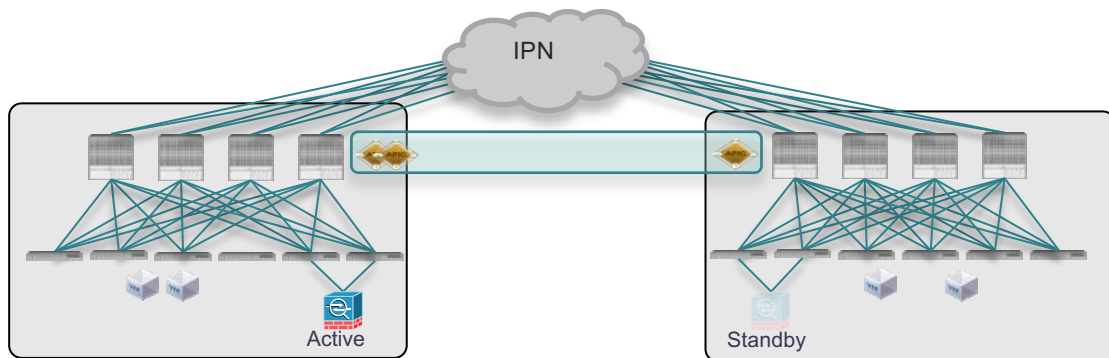


East-West

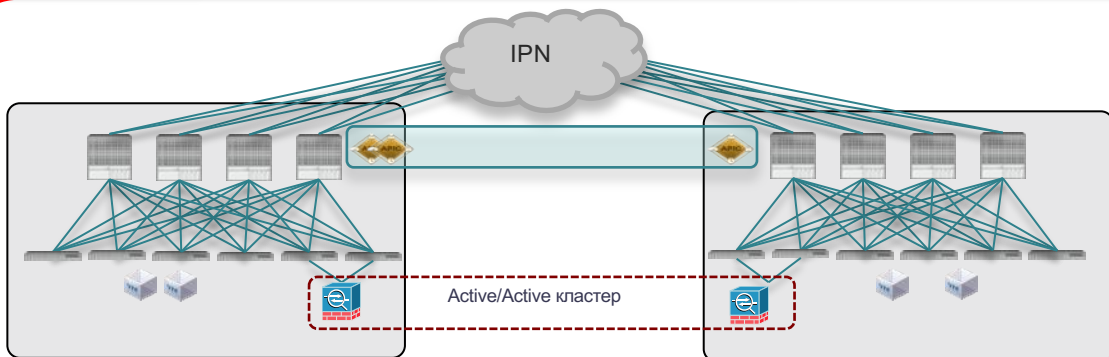


- Шлюз по умолчанию для нагрузки всегда настраивается на ACI фабрике (Anycast gateway)
- Трафик перенаправляется через активный FW при помощи PBR политики
- Использование сервисного графа обязательно

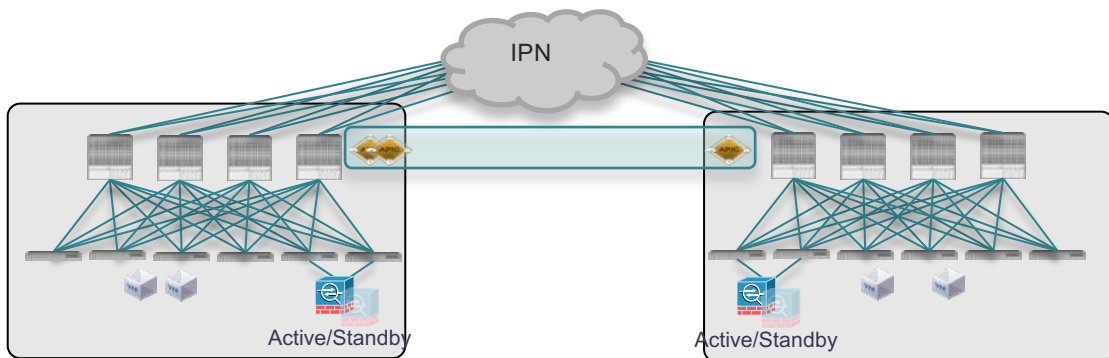
ACI Multi-Pod: Режимы интеграции МСЭ



- Одна Active/Standby пара «растянутая» между Pod
- Отсутствуют проблемы с асимметричными потоками



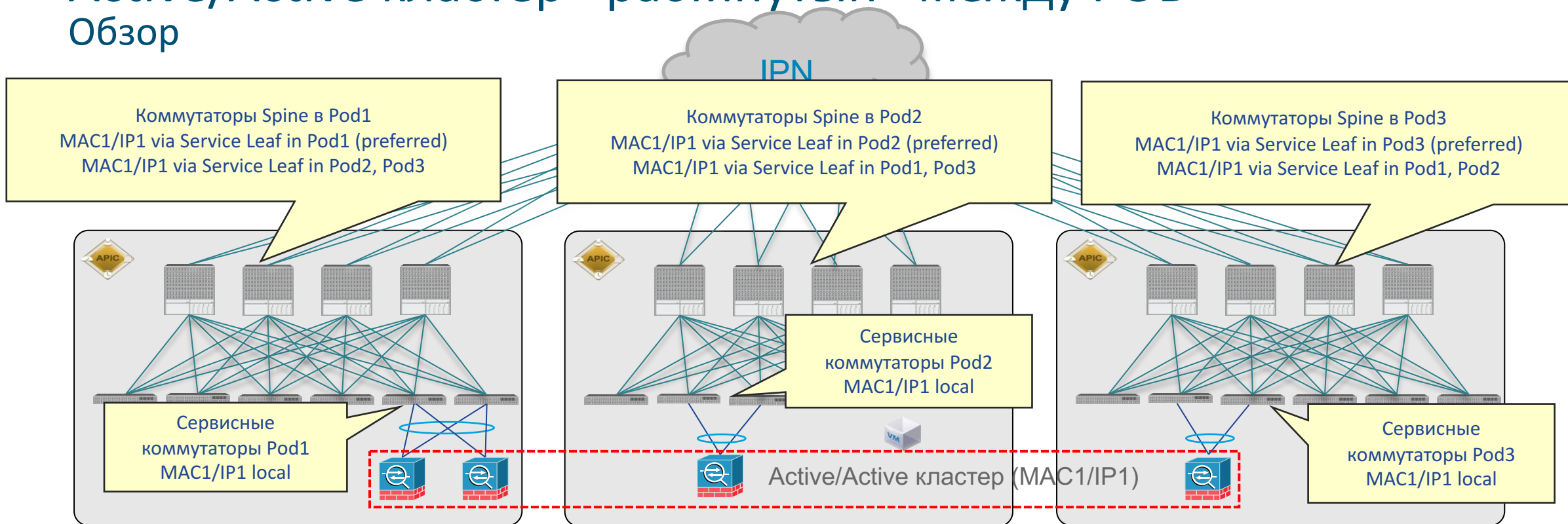
- Active/Active кластер «растянутый» между Pod (один логический FW)
- Требуется возможность подключения к фабрике одного и того же MAC/IP в нескольких POD одновременно
- Поддерживается начиная с версии ACI 3.2(4d) с обязательным использованием сервисного графа и PBR



- Независимые пары Active/Standby развернутые в каждом Pod
- Требуется использовать сервисный граф и симметричный PBR

Active/Active кластер «растянутый» между POD

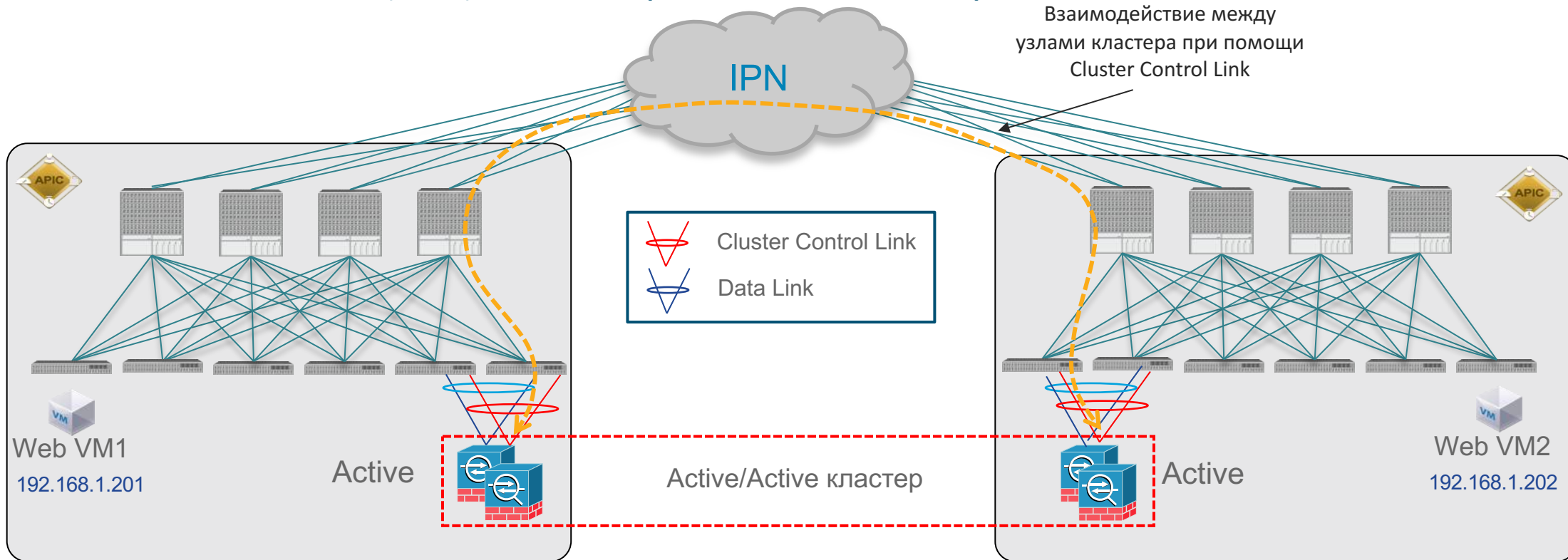
Обзор



- Поддерживается Cisco Firepower NGFW в режиме “Split Spanned EtherChannel”
 - Все ноды кластеризуются и с точки зрения сети ЦОД выглядят как одна сущность
- Внутри POD все ноды одного кластера FW должны быть подключены к одной и той же паре коммутаторов leaf

Active/Active кластер «растянутый» между POD

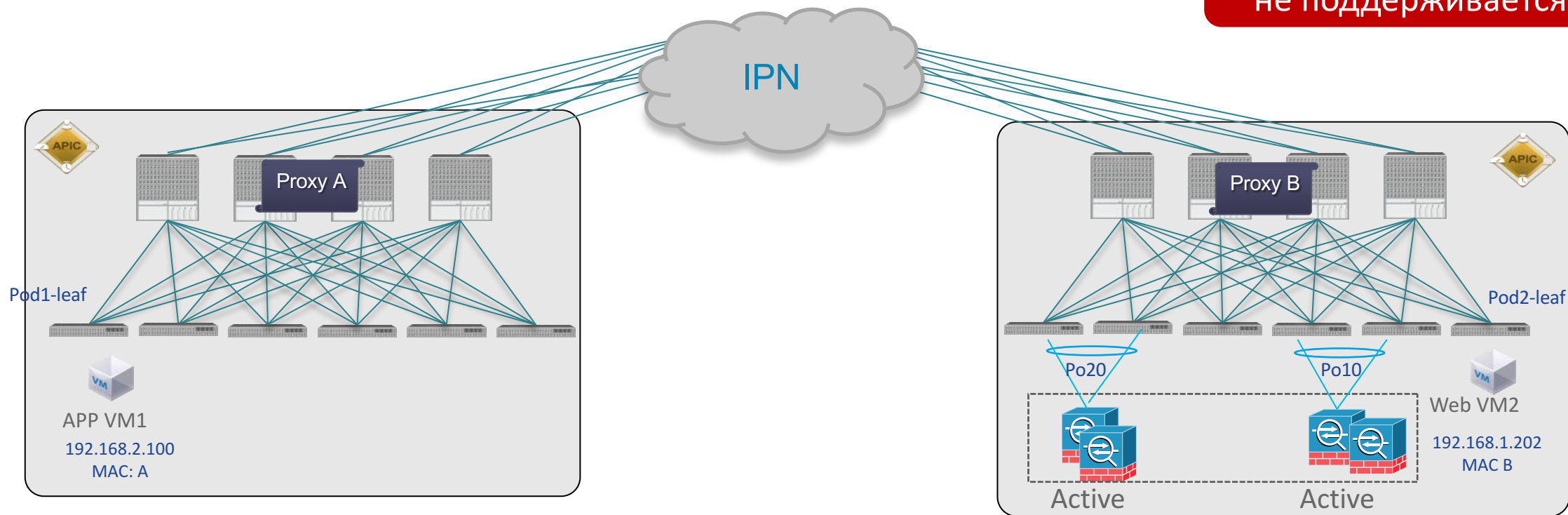
Cluster Control Link (CCL) так же «растягивается» при помощи ACI



- Рекомендуется использовать выделенные физические интерфейсы для CCL
 - нужно помнить, что CCL так же используется для передачи ассиметричных потоков данных, а не только для синхронизации узлов кластера друг с другом
- Для CCL требуется использовать выделенные EPG/BD, которые растянуты между Pod

Все ноды кластера подключены в один POD, но к разным парам Leaf-коммутаторов

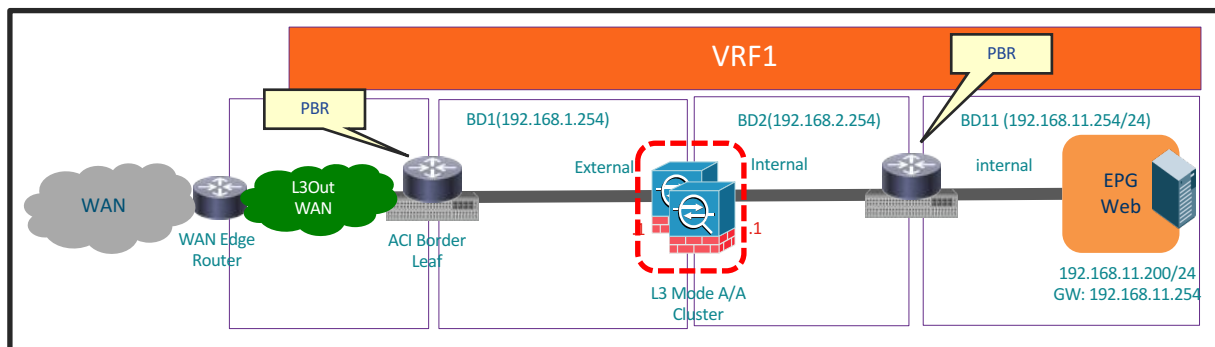
Дизайн
не поддерживается



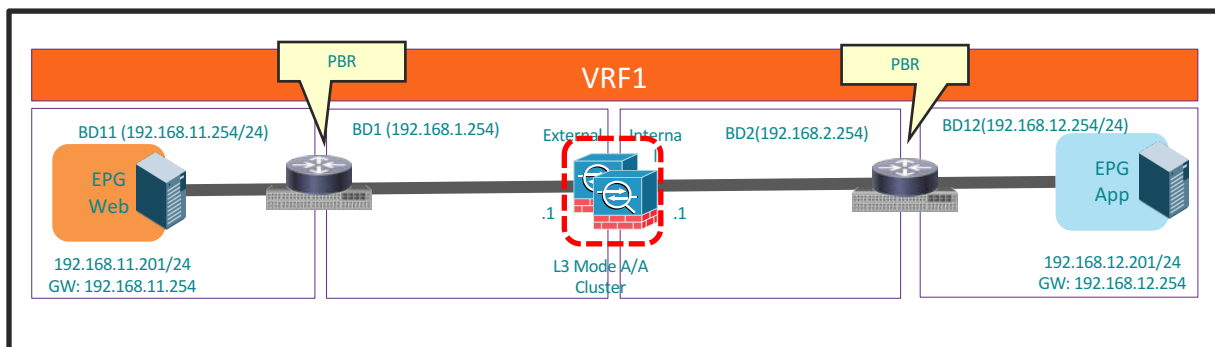
Active/Active кластер «растянутый» между POD

FW кластер в L3 режиме с сервисной цепочкой + PBR

North-South

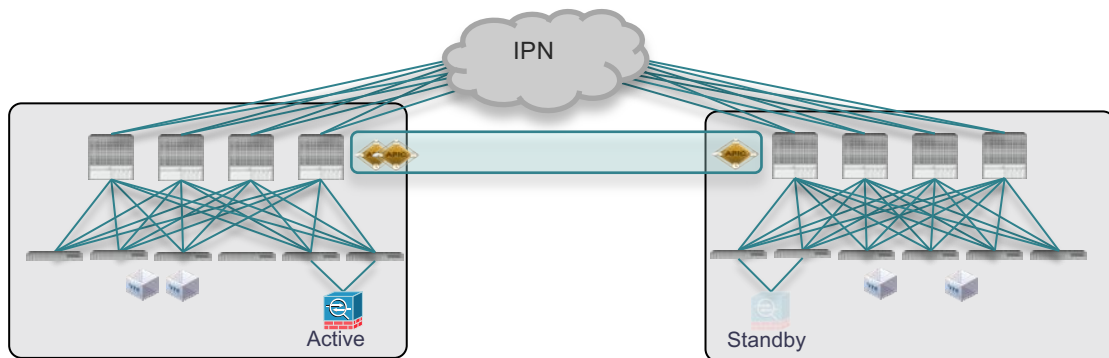


East-West

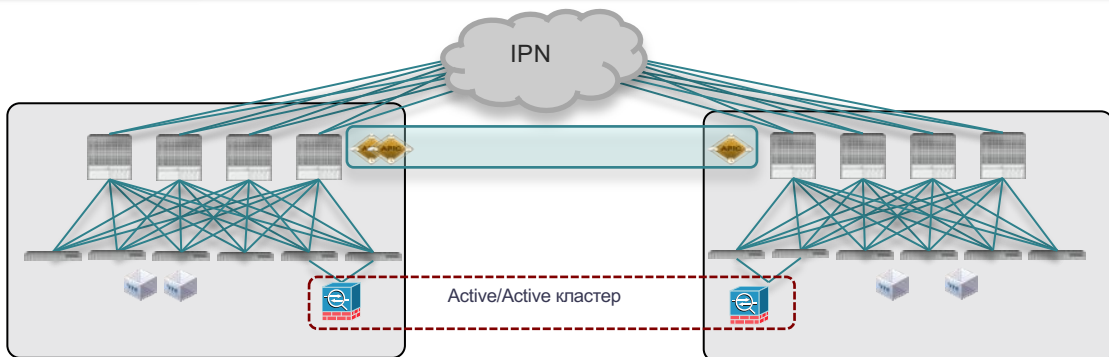


- Шлюз по умолчанию для нагрузки всегда настраивается на ACI фабрике (Anycast gateway)
- Трафик перенаправляется через кластер FW при помощи PBR политики
- Обязательно использовать сервисный граф (в режимах managed или un-managed mode)

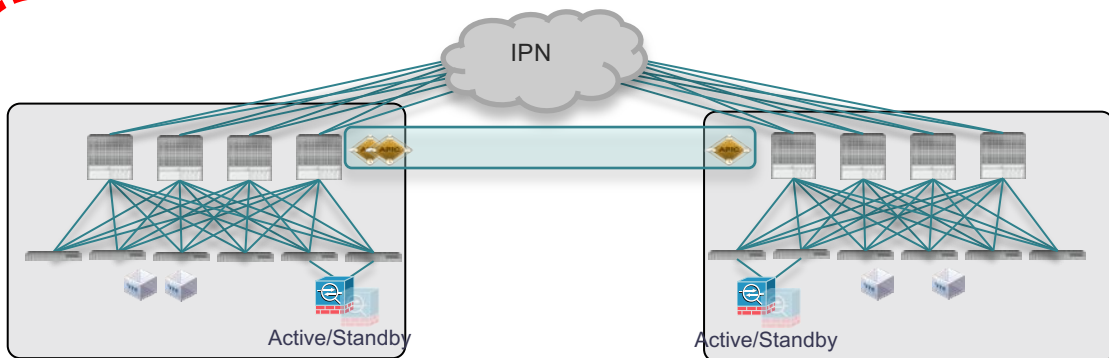
ACI Multi-Pod: Режимы интеграции МСЭ



- Одна Active/Standby пара «растянутая» между Pod
- Отсутствуют проблемы с асимметричными потоками



- Active/Active кластер «растянутый» между Pod (один логический FW)
- Требуется возможность подключения к фабрике одного и того же MAC/IP в нескольких POD одновременно
- Поддерживается начиная с версии ACI 3.2(4d) с обязательным использованием сервисного графа и PBR

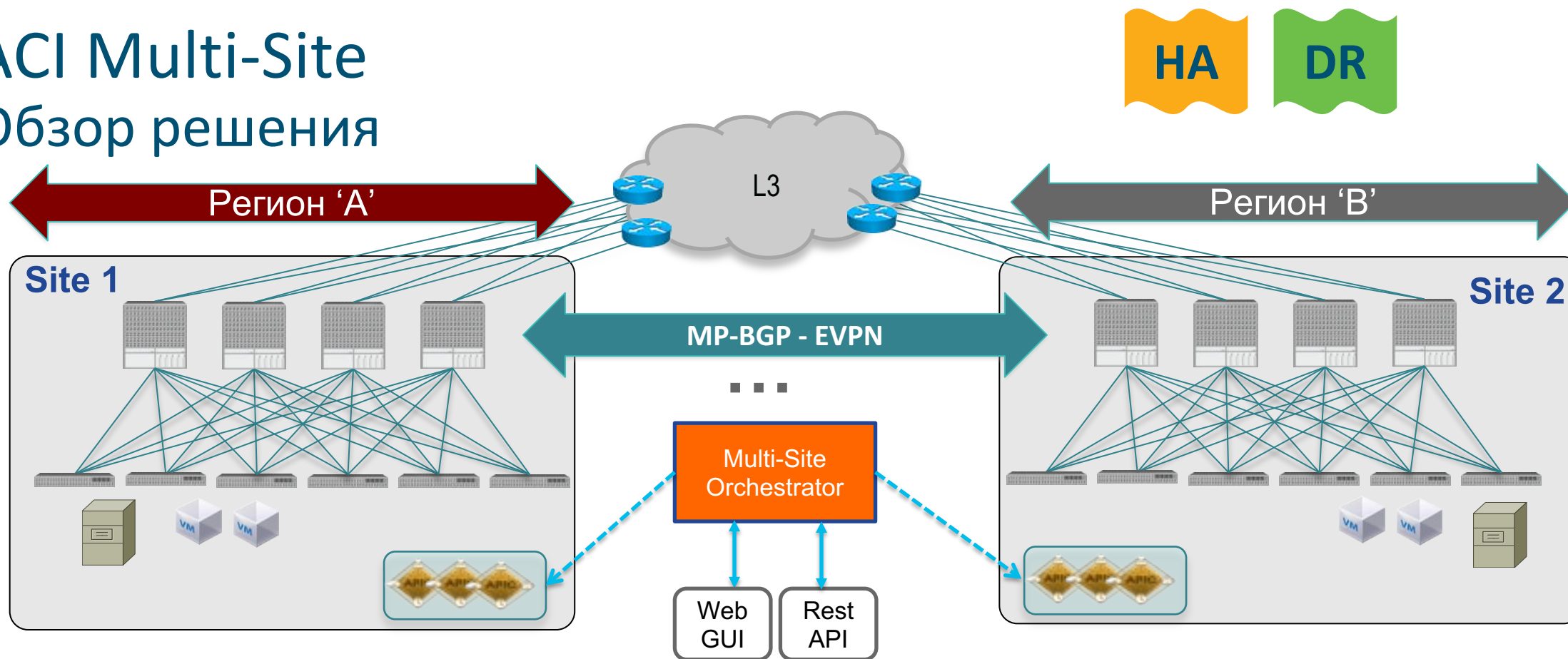


- Независимые пары Active/Standby развернутые в каждом Pod
- Требуется использовать сервисный граф и симметричный PBR

Катастрофоустойчивая сетевая фабрика на основе Cisco ACI Multi-Site

ACI Multi-Site

Обзор решения



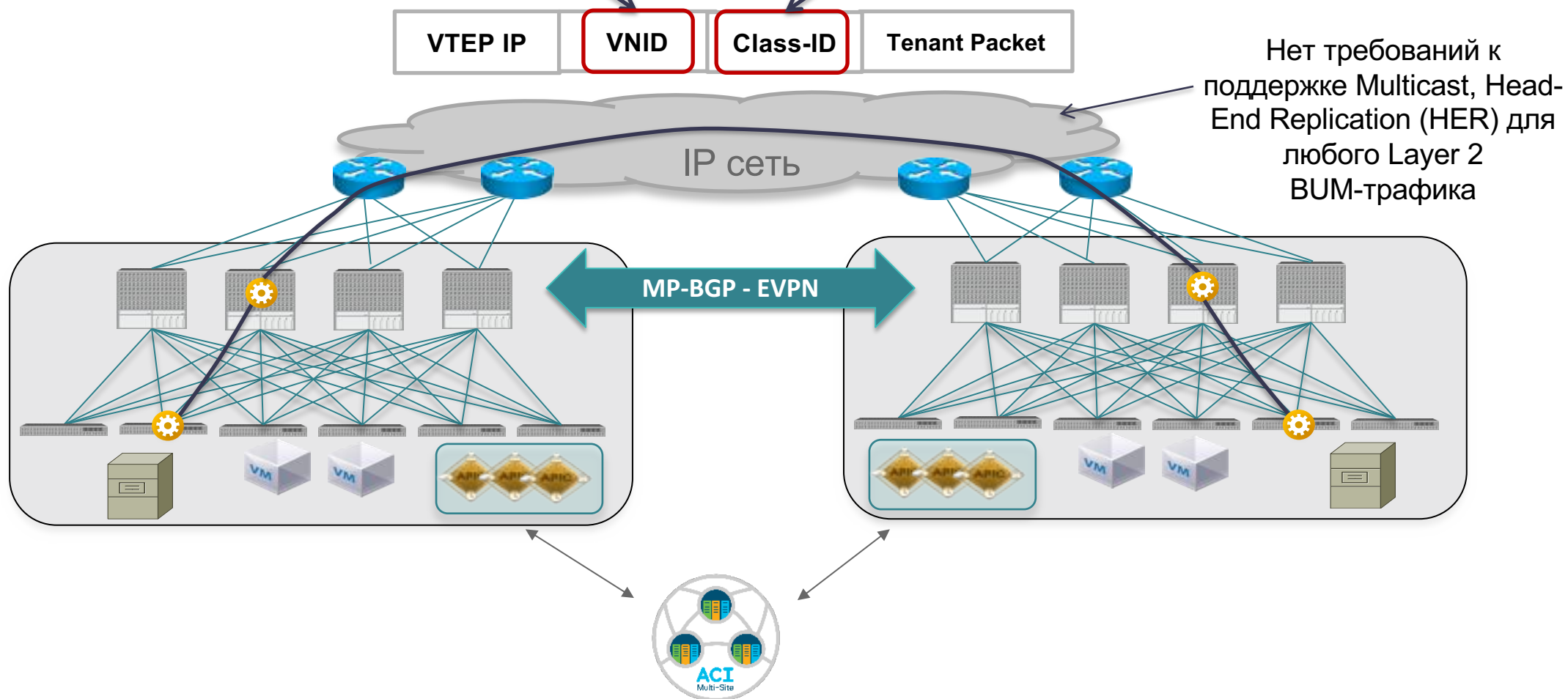
- Отдельные ACI фабрики с независимыми кластерами APIC
- Каждая фабрика рассматривается как отдельный «регион» и «зона доступности»
- Ограничение зоны вносимых изменений
- Использование для сценариев DR и Active/Active
- «Нет ограничений» по расстоянию
- Multi-Site контроллер настраивает сквозные конфигурации в нескольких кластерах APIC
- MP-BGP EVPN с VXLAN инкапсуляцией между сайтами
- Сквозное применение политик
- Поддержка Multi-Pod сайтов

ACI Multi-Site

Обмен сетевыми и Identity-атрибутами между фабриками

Сетевые атрибуты передаются между фабриками (Availability Zones)

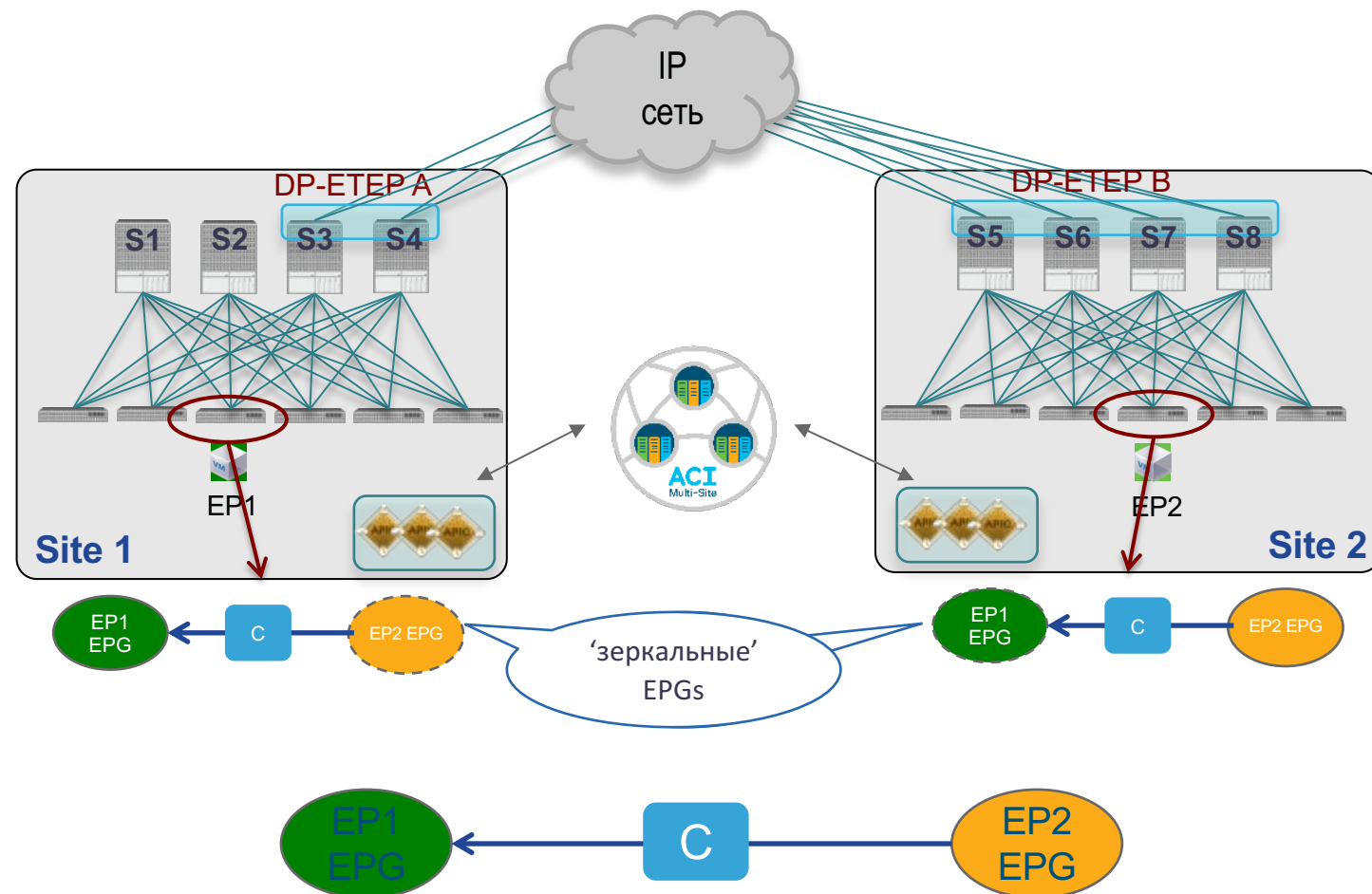
Identity атрибуты так же передаются между фабриками (Availability Zones)



ACI Multi-Site

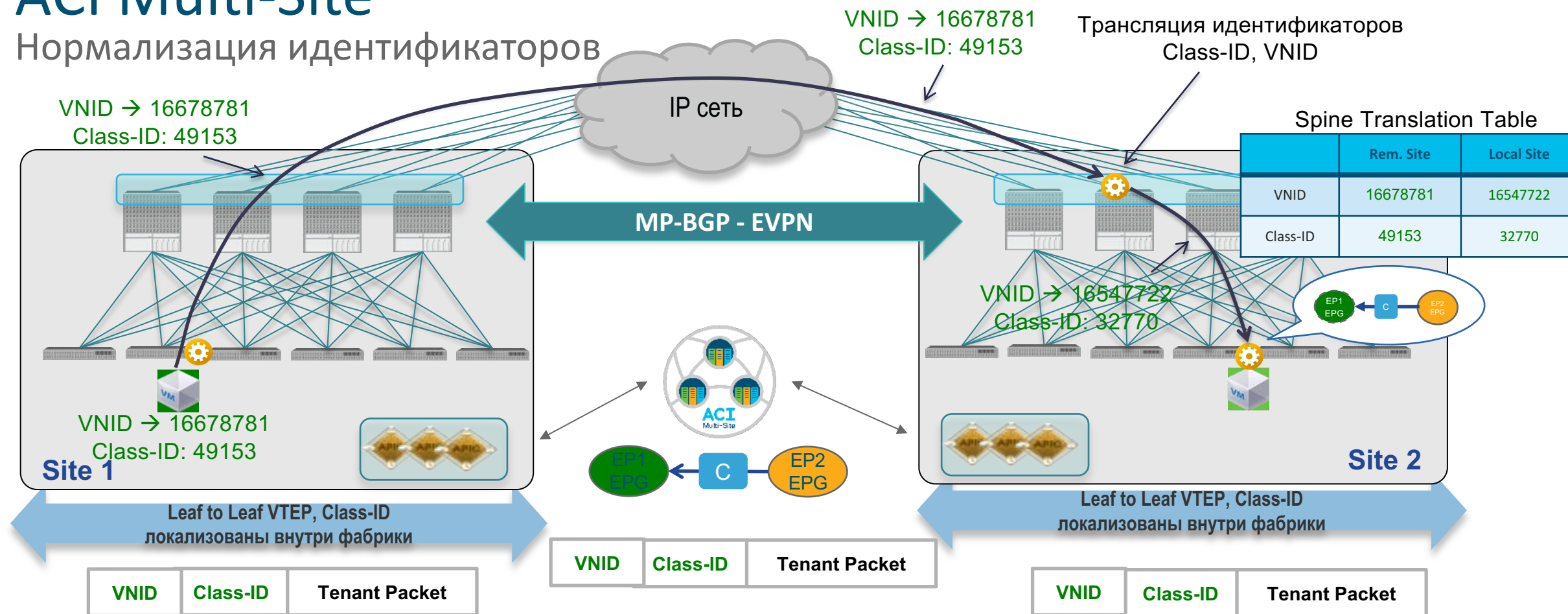
Inter-Site политики и 'зеркальные' EPG

- Inter-Site политики определяются на ACI Multi-Site Orchestrator и программируются в соответствующих APIC доменах
 - Сквозная консистентность политики
 - Создание 'зеркальных' EPG для локальной реализации политик
- Политики применяются на входе в leaf коммутатор в установившемся состоянии передачи данных



ACI Multi-Site

Нормализация идентификаторов



- Трансляция идентификаторов выполняется на spine-коммутаторах
- Требуется специализированная аппаратура для выполнения этих функций
- Multi-Site Orchestrator инструктирует APIC-контроллеры о том как настроить таблицу трансляций на spine-коммутаторах

Multi-Site

Проверяем автоматическую настройку нормализации

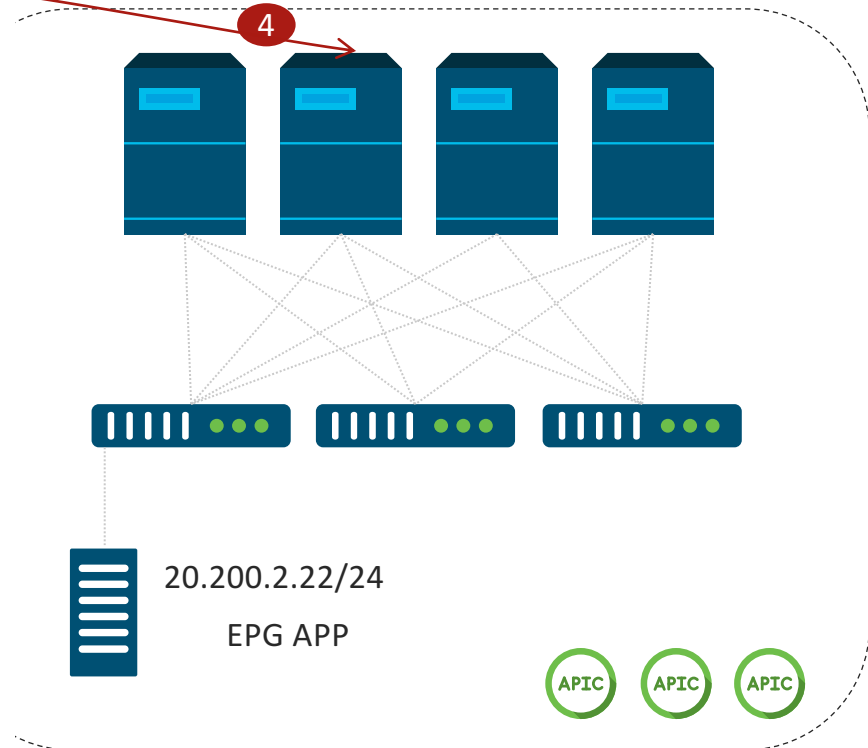
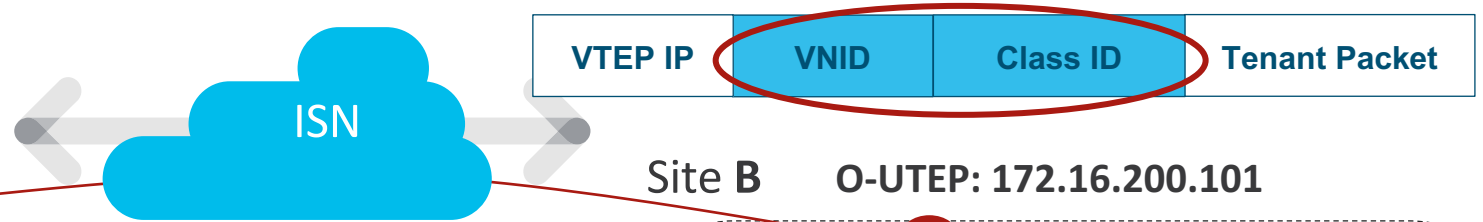
Site A O-UTEP: 172.16.100.101

```
S2P1-Spine201# show dcimgr repo vnid-maps
```

Remote			Local		
site	Vrf	Bd	Vrf	Bd	Rel-state
1	2129921		2949120		[formed]
	Site A		Site B		

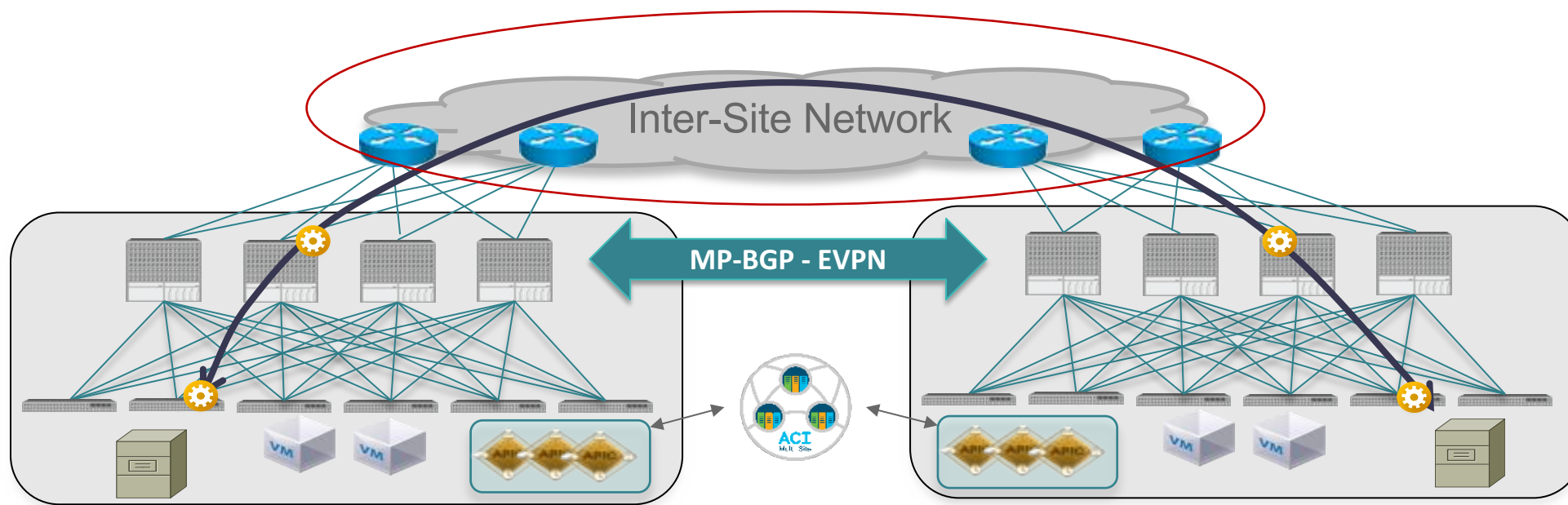
```
S2P1-Spine201# show dcimgr repo sclass-maps
```

Remote			Local		
site	Vrf	PcTag	Vrf	PcTag	Rel-state
1	2260993	16386	2424832	16386	[formed]
1	2129921	32772	2949120	16386	[formed]
1	2129921	32770	2949120	32771	[formed]
1	2129921	32771	2949120	16387	[formed]



ACI Multi-Site

Требования к Inter-Site Network (ISN)



- Не управляется при помощи APIC, должна быть настроена отдельно
- IP топология может быть гибкой (arbitrary), не обязательно подключать все коммутаторы spine, сайты могут быть разнесены на существенное расстояние друг от друга
- Основные требования:
 - ✓ OSPF на стыке spine коммутатора и следующего узла для распространения информации о префиксах E-TEP
 - ✓ Поддержка увеличенного MTU (как минимум +100B) для передачи VXLAN трафика между сайтами

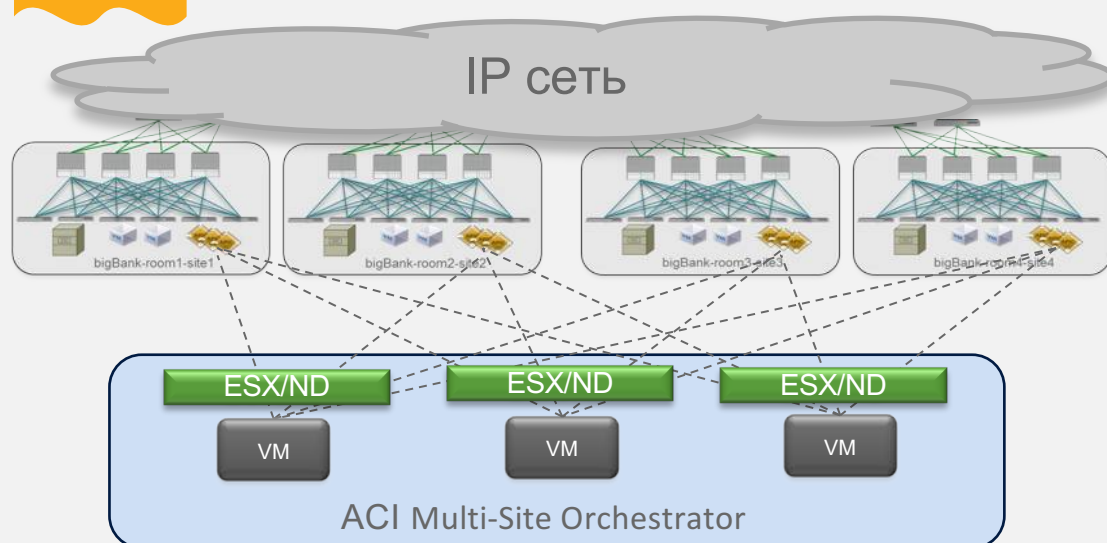
ACI Multi-Site

MSO - соображения по развертыванию



Развертывание внутри ЦОД

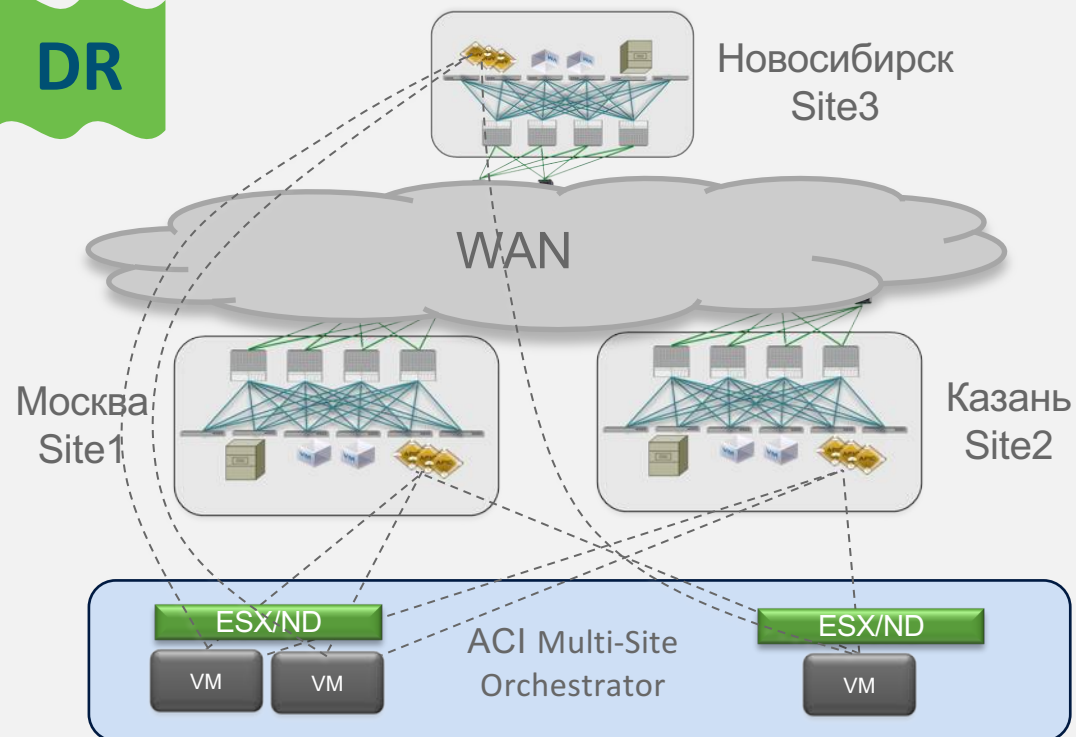
HA



- Гипервизоры могут быть подключены к сети DC OOB
- Каждый MSO имеет уникальный маршрутизируемый IP
- Асинхронные вызовы от MSO к APIC

Объединение ЦОД при помощи WAN

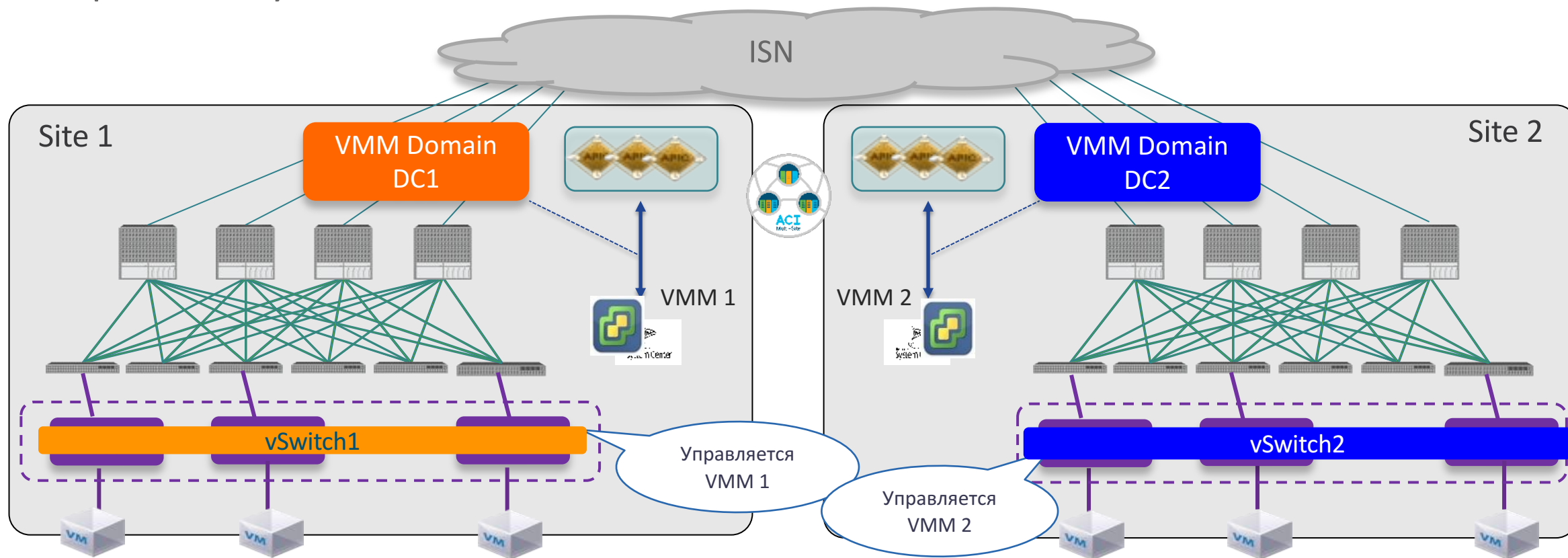
DR



- До 150 msec RTT поддерживается между узлами MSO
- Большая задержка (500 msec до 1 sec RTT) между MSO и управляемыми APIC кластерами
- По возможности MSO узлы разворачиваются на разных сайтах

ACI Multi-Site и VMM интеграция

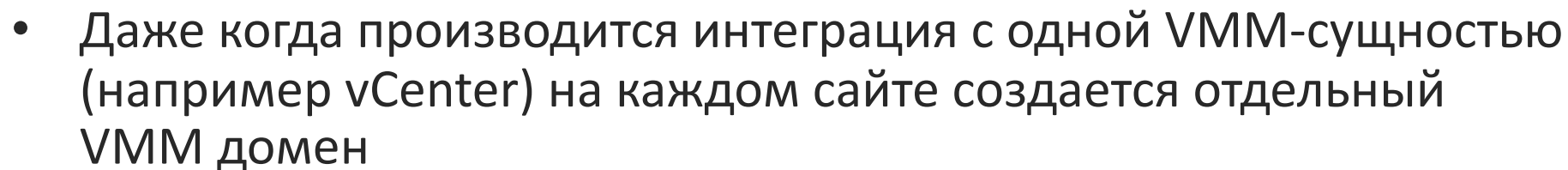
Вариант 1 – уникальный VMM на сайт



- Типичное внедрение для ACI Multi-Site
- На каждом сайте создаются отдельные VMM домены, которые потом делаются доступными Multi-Site Orchestrator

Вариант 2 – Один vCenter для управления хостами на двух сайтах

Вариант 2 – Один vCenter для управления хостами на двух сайтах

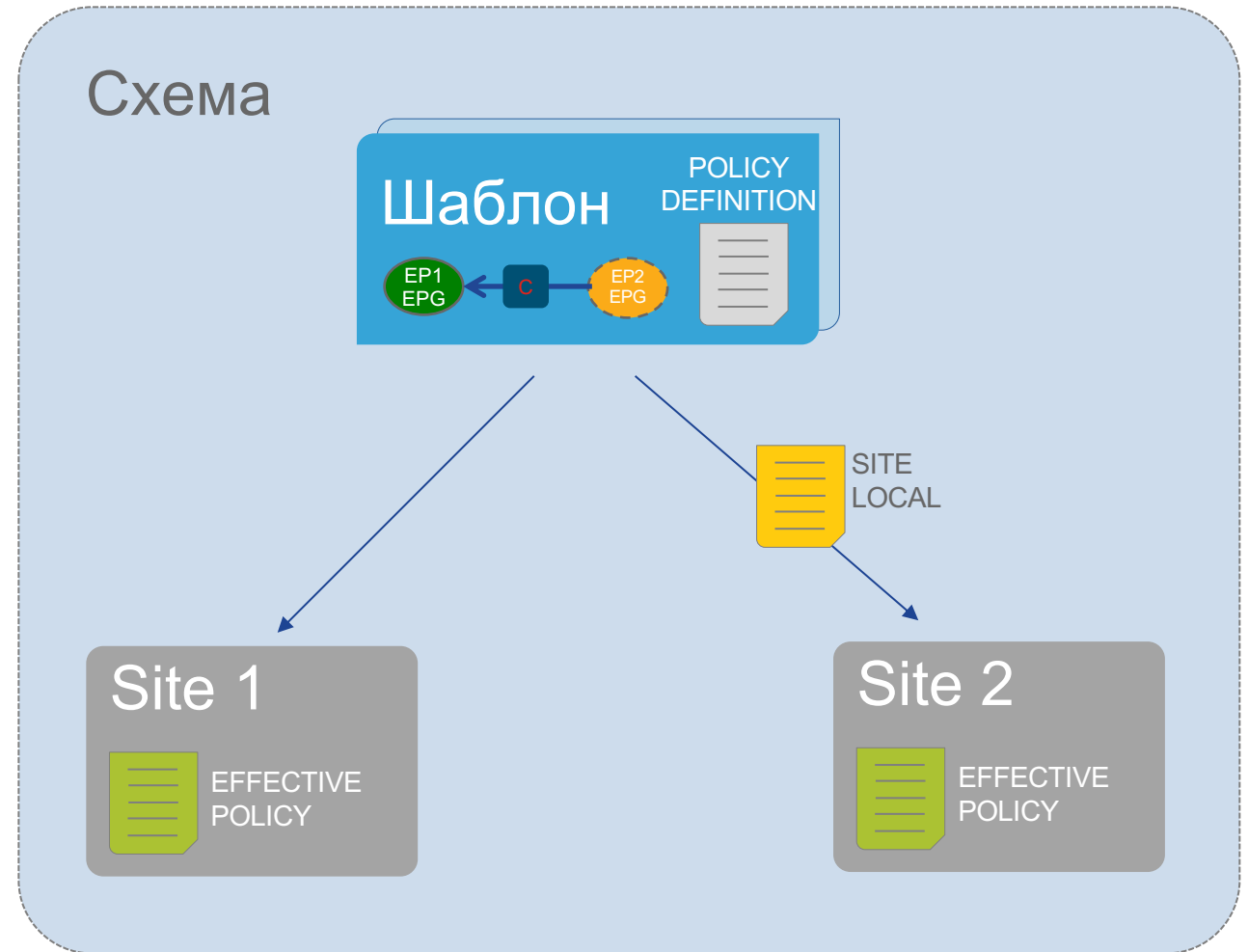


- Даже когда производится интеграция с одной VMM-сущностью (например vCenter) на каждом сайте создается отдельный VMM домен

ACI Multi-Site

MSO схема и шаблоны

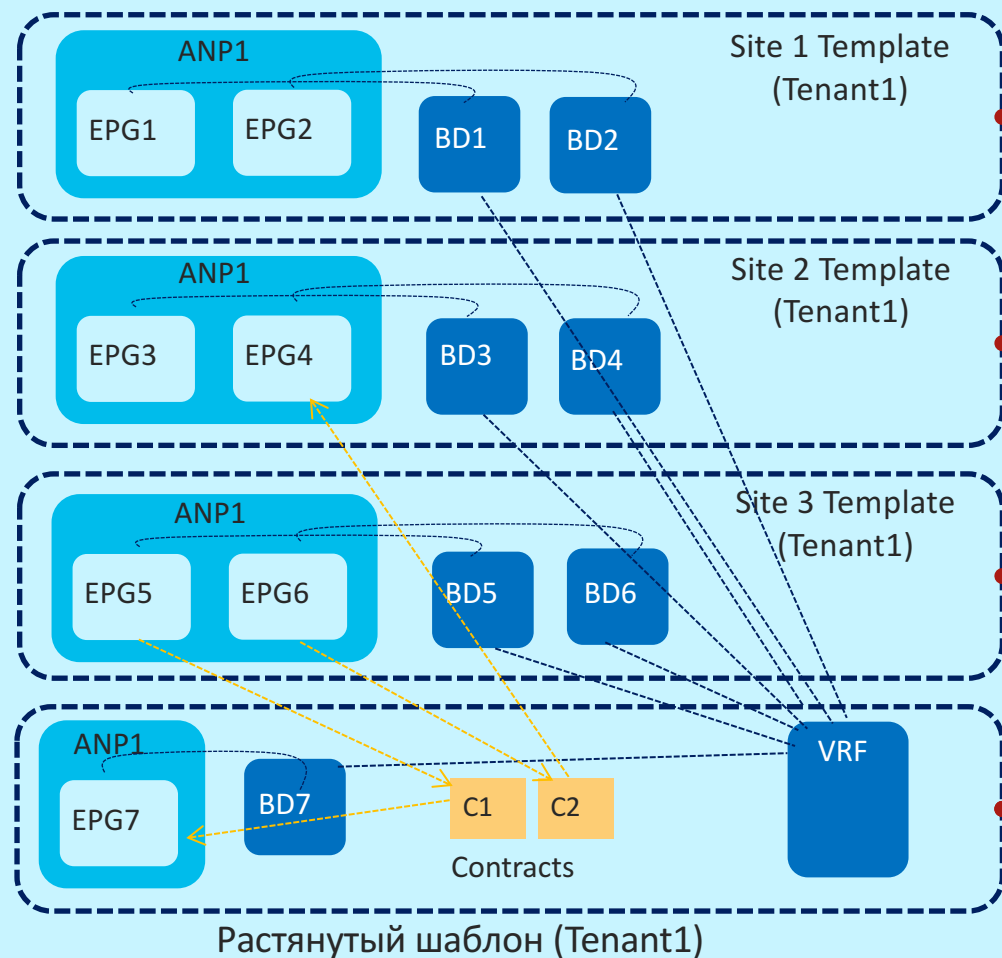
- Шаблон (template) = определение APIC политики (App & Network)
- Шаблоны локализуются на сайтах в виде конкретных настроек
- Шаблоны ассоциируются со всеми сайтами или только с частью
- Схема = контейнер для шаблонов которые создаются для реализации общего сценария
- Область изменений (scope of change): политики в разных шаблонах могут быть локализованы на разных сайтах в разное время



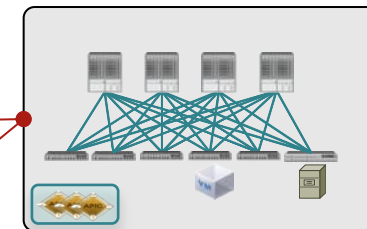
Дизайн схемы

Один шаблон на каждый сайт и растянутый шаблон на каждый из сайтов

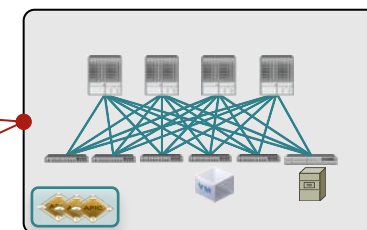
Схема



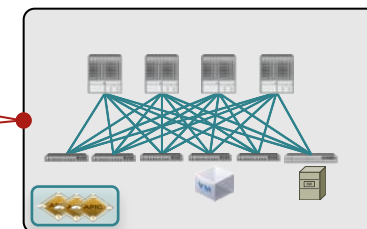
Сайт 1



Сайт 2



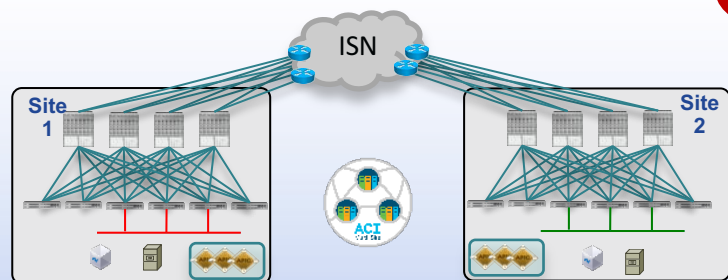
Сайт 3



ACI Multi-Site - организации сетевой связанности

Варианты настройки Bridge Domain (L2-сегмент)

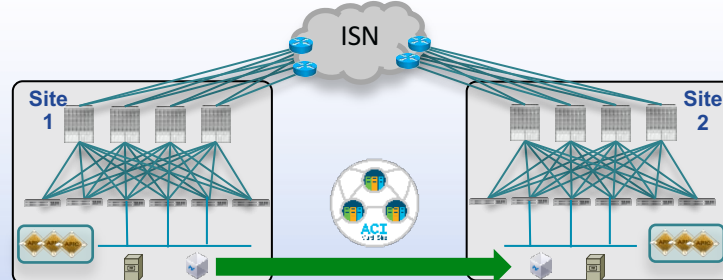
Только L3 связанность



- Bridge Domain и подсети не растягиваются между сайтами
- Layer 3 Intra-VRF или Inter-VRF взаимодействие (разделяемые сервисы между VRF/Tenant-ами)



IP мобильность без передачи BUM трафика

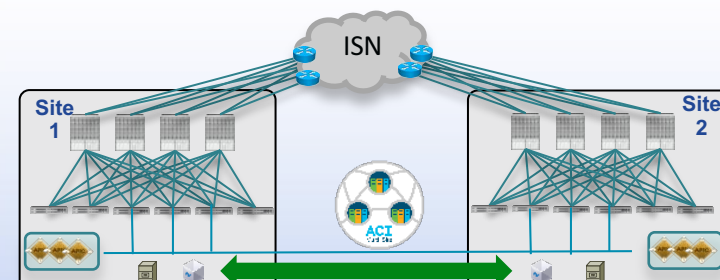


- Одна и та же IP подсеть определена на разных сайтах
- Поддерживается IP Mobility ('холодная' и 'живая'* миграция VM) и взаимодействие между хостами внутри одной подсети
- Нет Layer 2 BUM флдинга между сайтами

MSO GUI (BD)



Классическая L2 связанность



- Взаимодействие между сайтами с возможностями изоляции доменов
- L2 растягивается между сайтами, 'живая'* VM миграция и кластеризация приложений
- Есть Layer 2 BUM флдинг между сайтами

MSO GUI (BD)

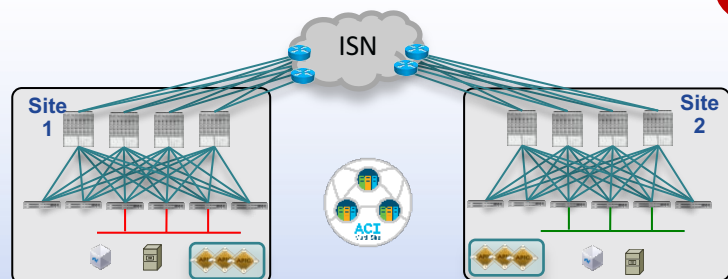


*'Живая' миграция поддерживается начиная с релиза ACI 3.2

ACI Multi-Site - организации сетевой связанности

Варианты настройки Bridge Domain (L2-сегмент)

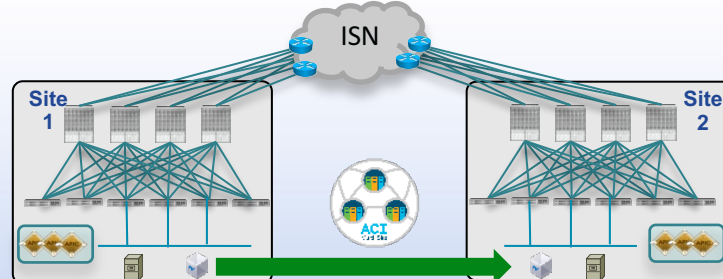
Только L3 связанность



- Bridge Domain и подсети не растягиваются между сайтами
- Layer 3 Intra-VRF или Inter-VRF взаимодействие (разделяемые сервисы между VRF/Tenant-ами)



IP мобильность без передачи BUM трафика

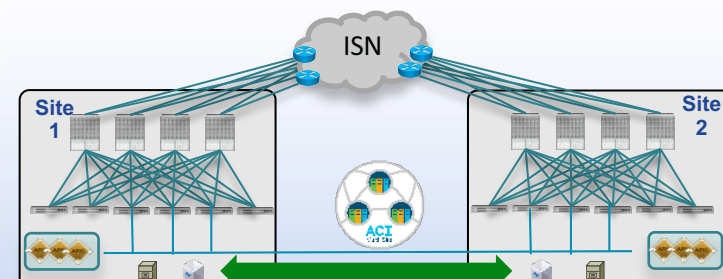


- Одна и так же IP подсеть определена на разных сайтах
- Поддерживается IP Mobility ('холодная' и 'живая'* миграция VM) и взаимодействие между хостами внутри одной подсети
- Нет Layer 2 BUM флдинга между сайтами

MSO GUI (BD)



Классическая L2 связанность



- Взаимодействие между сайтами с возможностями изоляции доменов
- L2 растягивается между сайтами, 'живая'* VM миграция и кластеризация приложений
- Есть Layer 2 BUM флдинг между сайтами

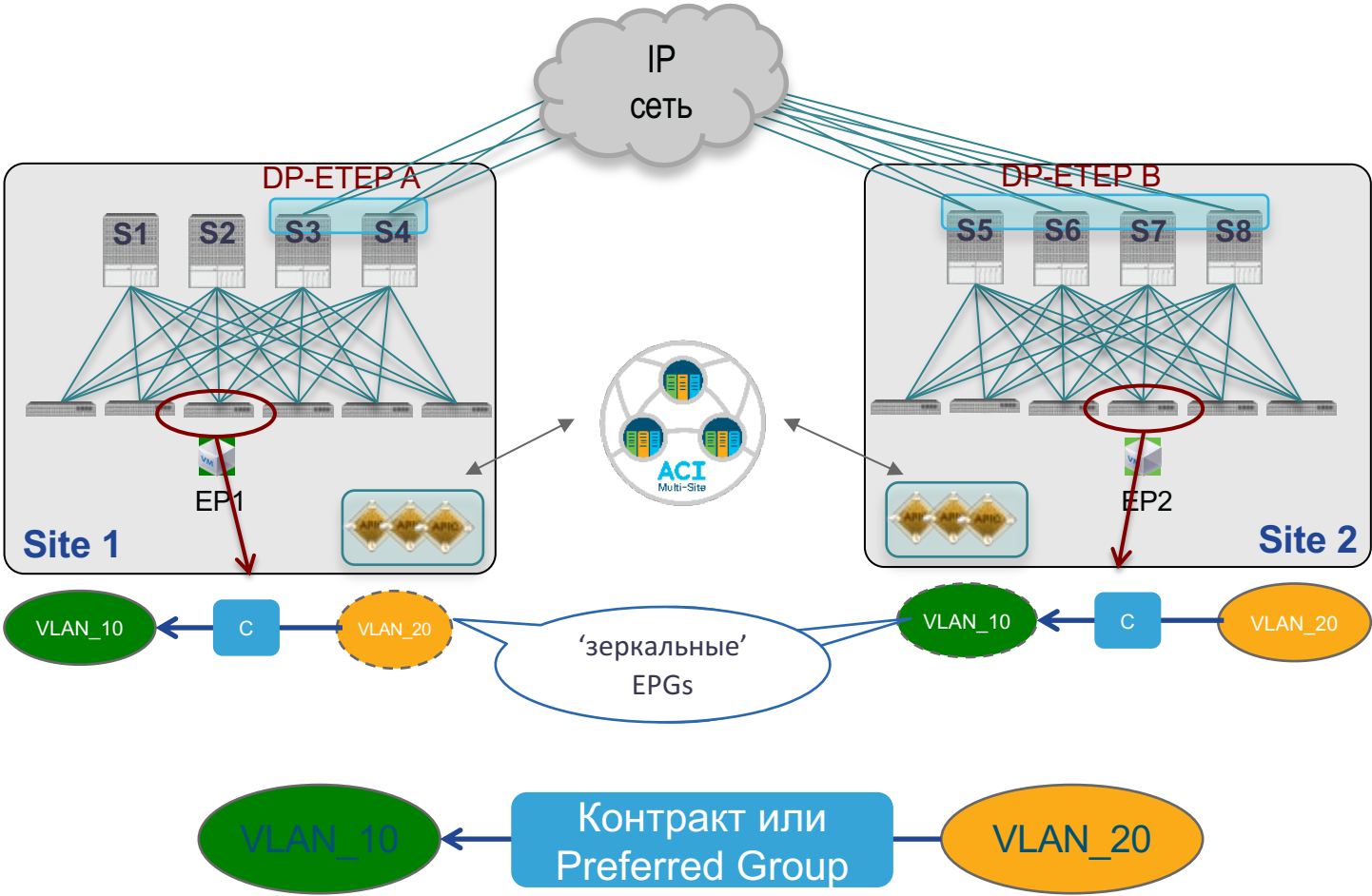
MSO GUI (BD)



*'Живая' миграция поддерживается начиная с релиза ACI 3.2

ACI Multi-Site

Демонстрация – организация L3 связанности между сайтами



ACI Multi-Site

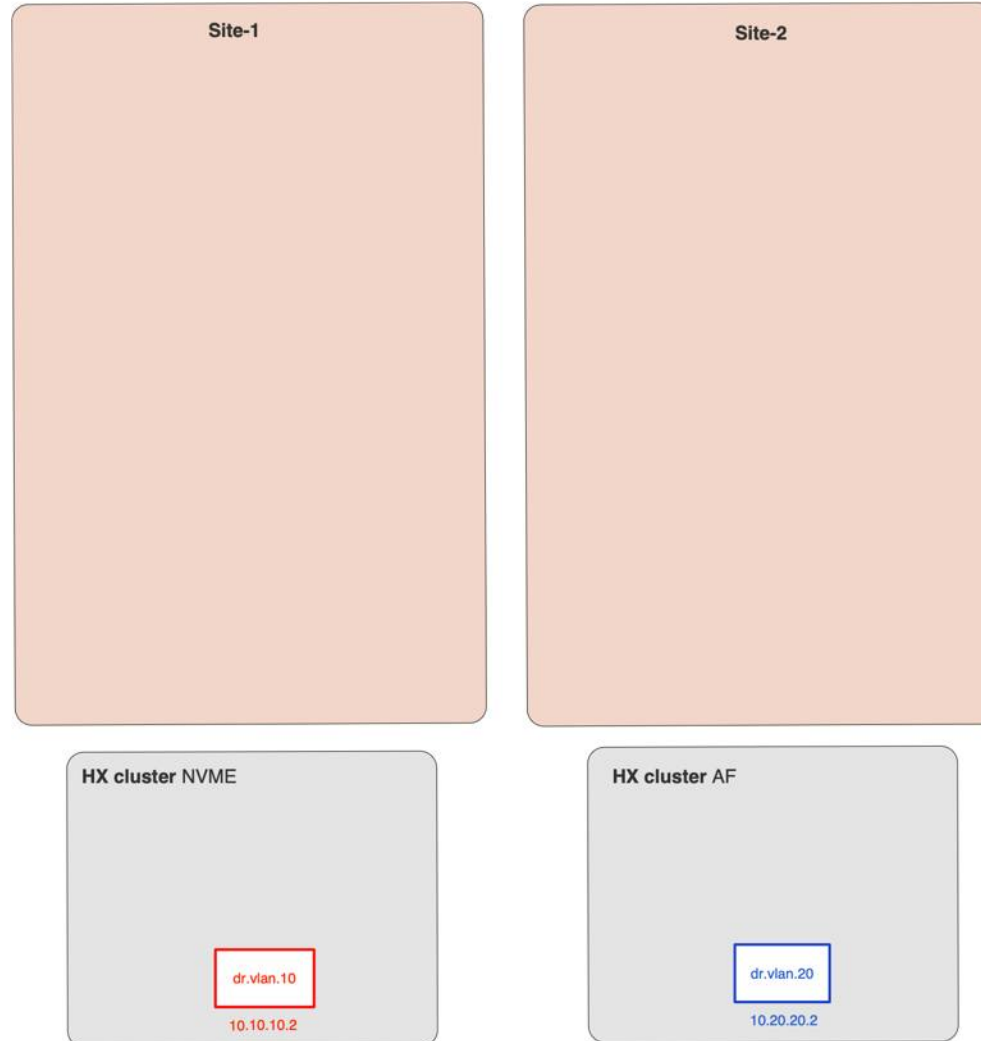
Демонстрация – организация L3 связанности между сайтами

0

Создание tenant

1

Исходное состояние



ACI Multi-Site

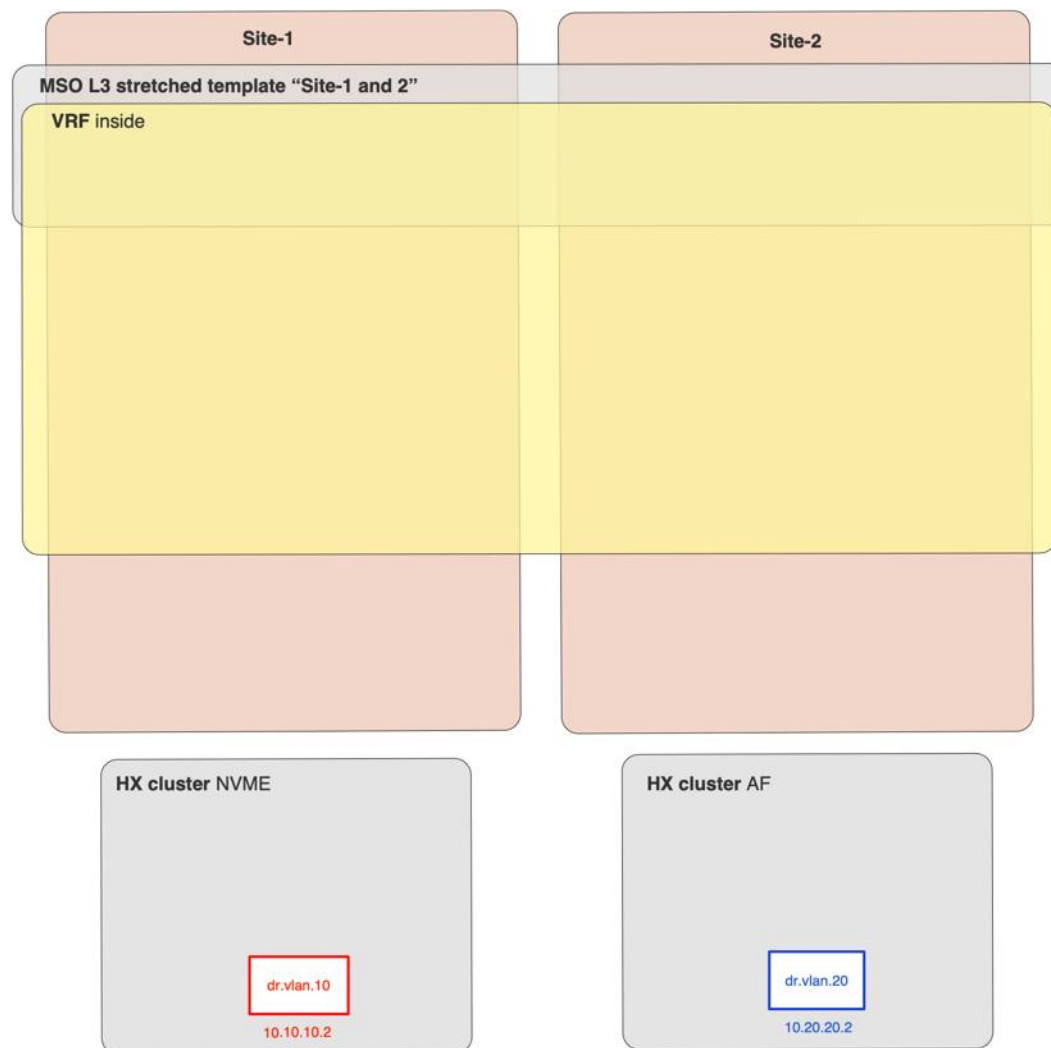
Демонстрация – организация L3 связанности между сайтами

2

Создаем общий шаблон настроек для 2-х сайтов

3

Создаем VRF, который автоматически растягивается между сайтами



ACI Multi-Site

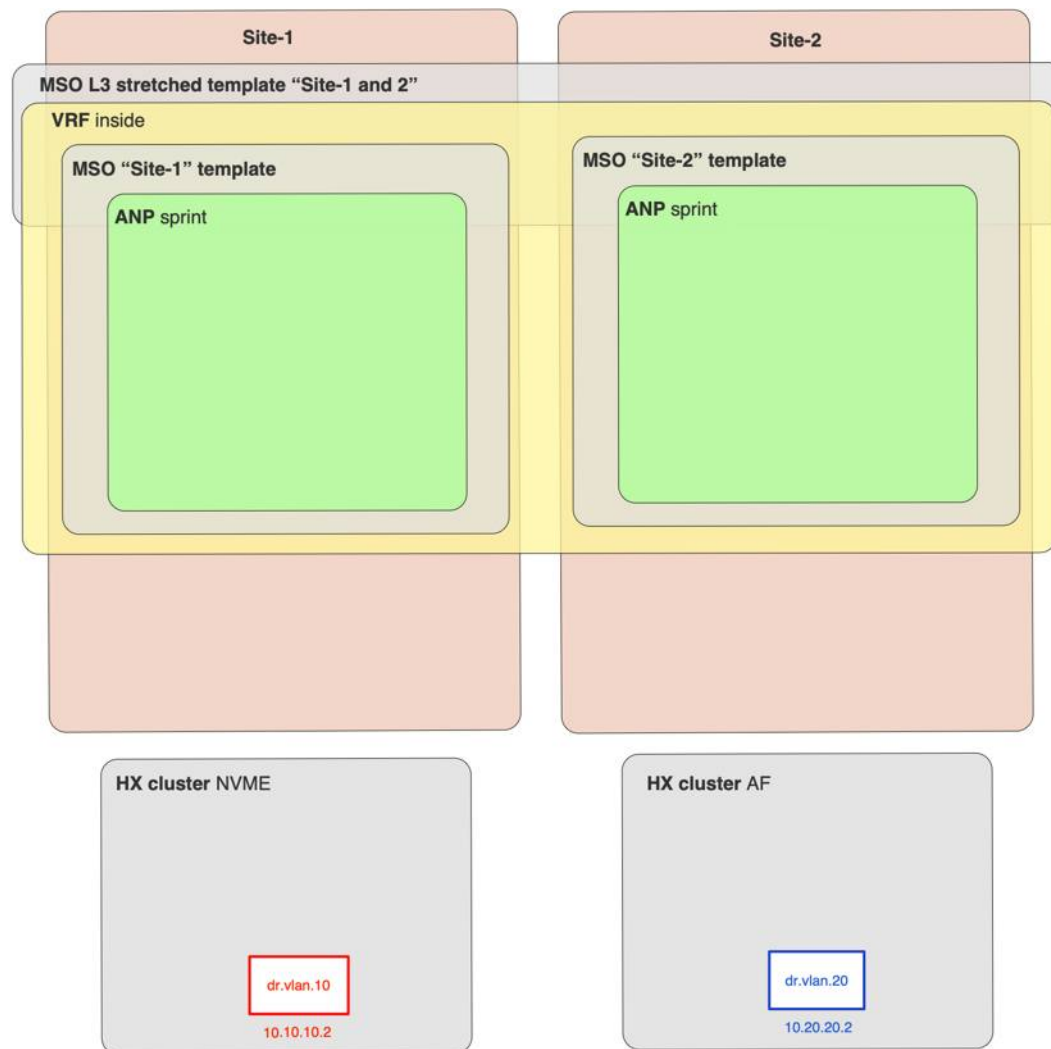
Демонстрация – организация L3 связанности между сайтами

4

Создаем индивидуальные шаблоны настроек для каждого сайта

5

Создаем профиль приложения (контейнер для зон безопасности = EPG)



ACI Multi-Site

Демонстрация – организация L3 связанности между сайтами

6

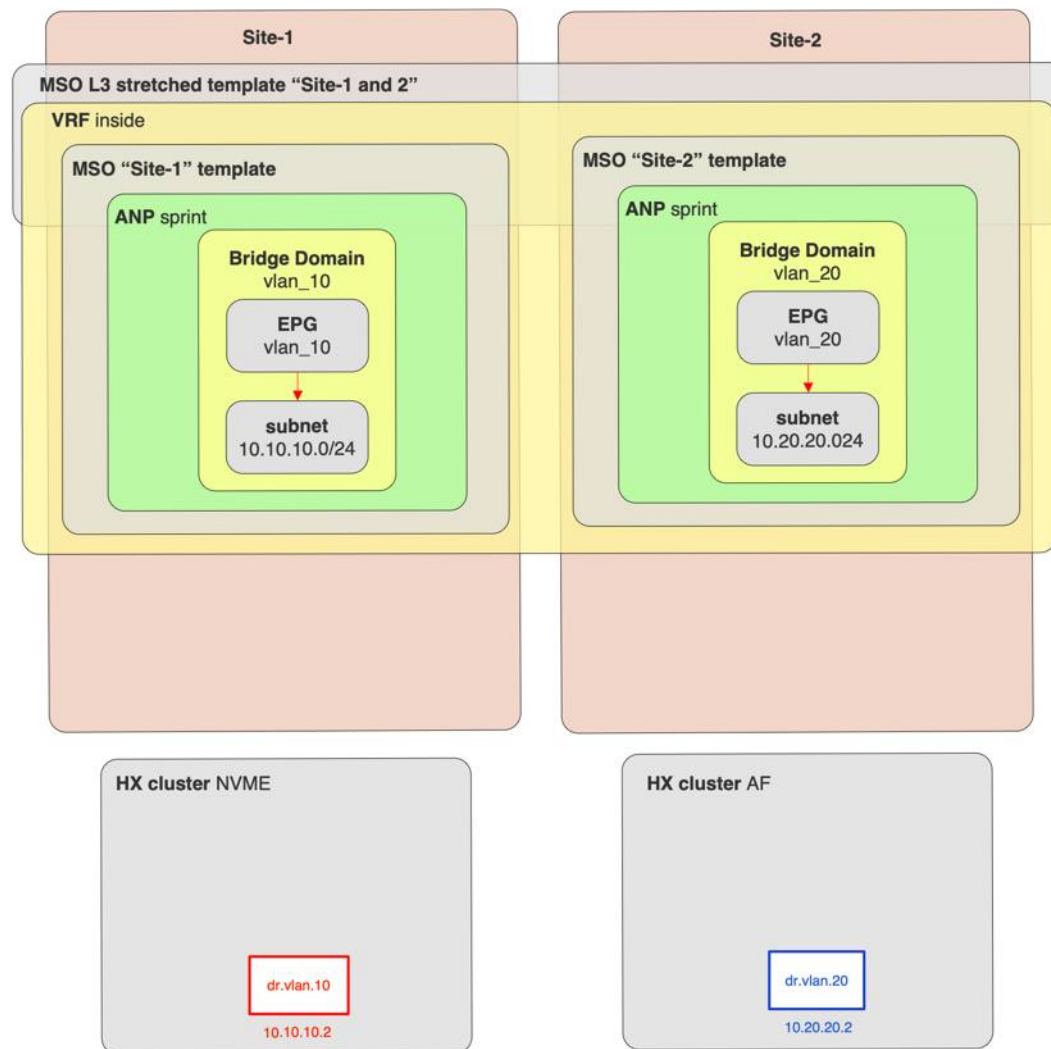
Создаем L2 сегменты =
Bridge Domain

7

Настраиваем L3 subnet

8

Настраиваем зону
безопасности = EPG

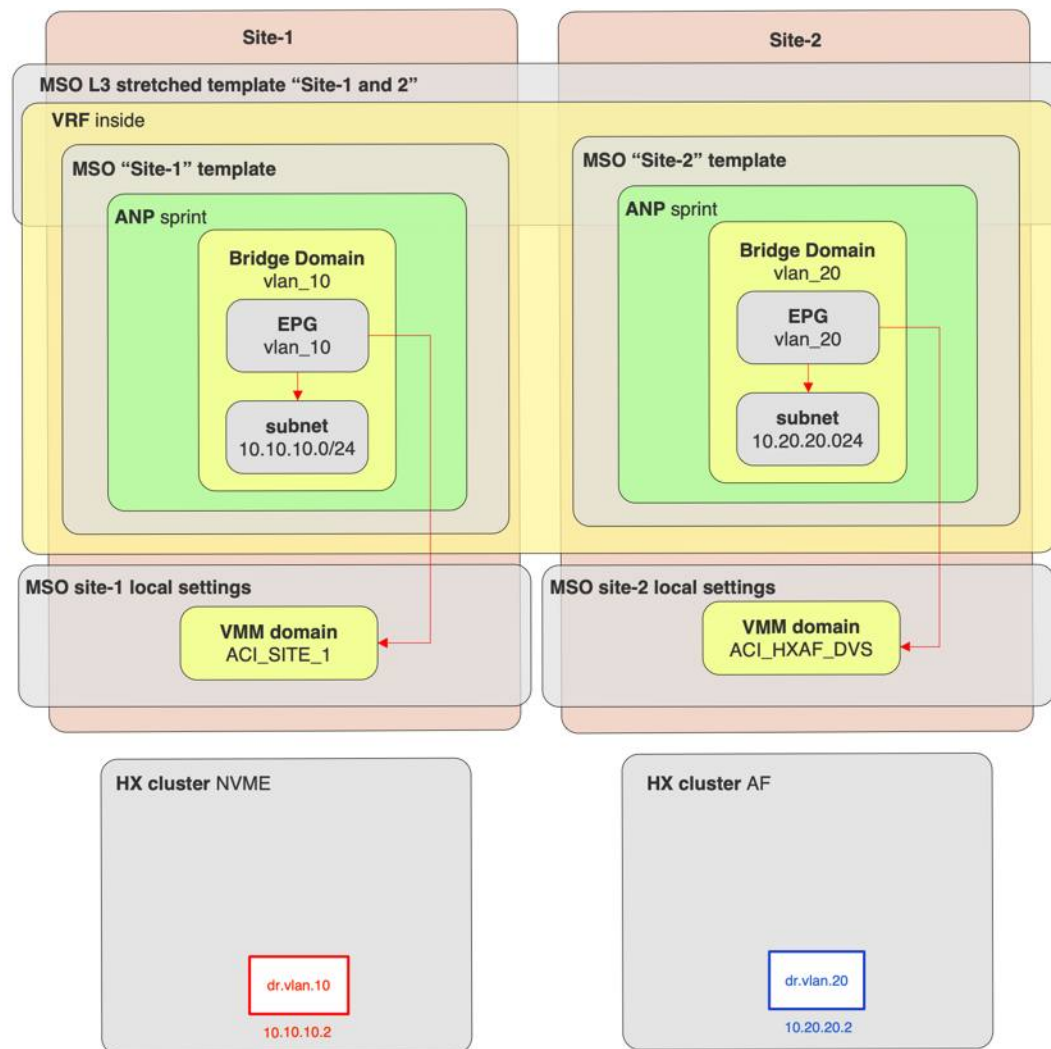


ACI Multi-Site

Демонстрация – организация L3 связанности между сайтами

9

Выполняем ассоциацию
с VMM-доменом на
уровне настроек EPG
внутри сайта

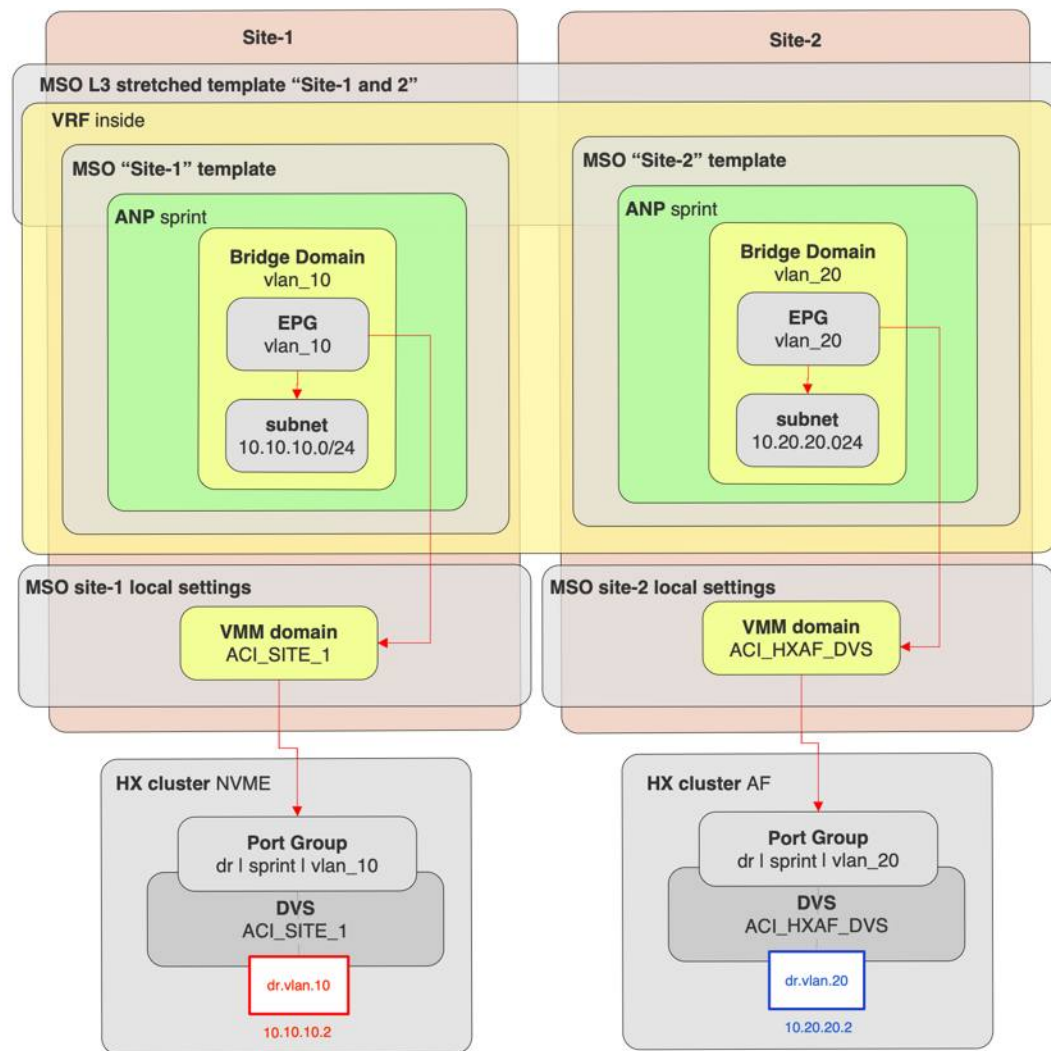


ACI Multi-Site

Демонстрация – организация L3 связанности между сайтами

10

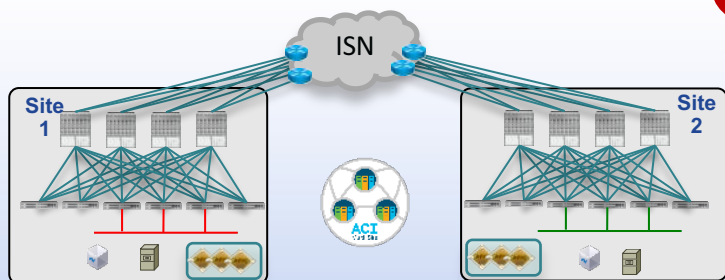
APIC автоматизирует
создание портовых
группы на DVS и VLAN
на системе
гиперконвергенции
Cisco Hyperflex



ACI Multi-Site - организации сетевой связанности

Варианты настройки Bridge Domain (L2-сегмент)

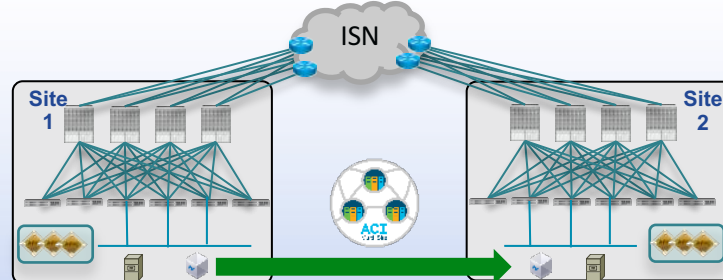
Только L3 связанность



- Bridge Domain и подсети не растягиваются между сайтами
- Layer 3 Intra-VRF или Inter-VRF взаимодействие (разделяемые сервисы между VRF/Tenant-ами)



IP мобильность без передачи BUM трафика

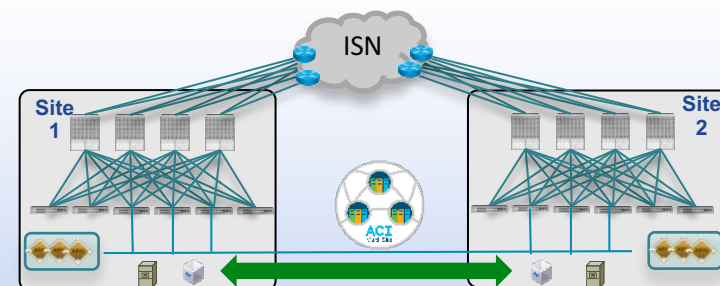


- Одна и так же IP подсеть определена на разных сайтах
- Поддерживается IP Mobility ('холодная' и 'живая'* миграция VM) и взаимодействие между хостами внутри одной подсети
- Нет Layer 2 BUM флдинга между сайтами**

MSO GUI (BD)



Классическая L2 связанность



- Взаимодействие между сайтами с возможностями изоляции доменов
- L2 растягивается между сайтами, 'живая'* VM миграция и кластеризация приложений
- Есть Layer 2 BUM флдинг между сайтами**

MSO GUI (BD)



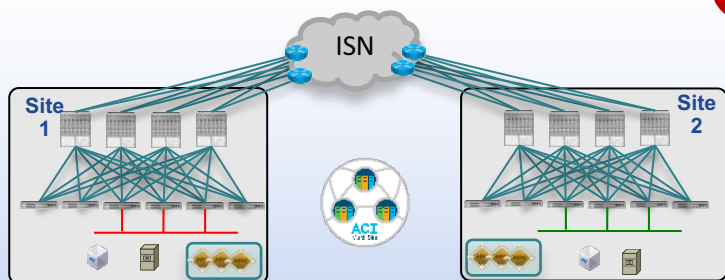
*'Живая' миграция поддерживается начиная с релиза ACI 3.2

ACI Multi-Site - организации сетевой связанности

Варианты настройки Bridge Domain (L2-сегмент)

Только L3 связанность

1

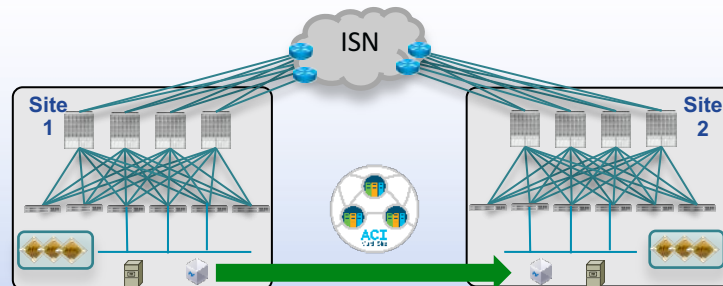


- Bridge Domain и подсети не растягиваются между сайтами
- Layer 3 Intra-VRF или Inter-VRF взаимодействие (разделяемые сервисы между VRF/Tenant-ами)



IP мобильность без передачи BUM трафика

2



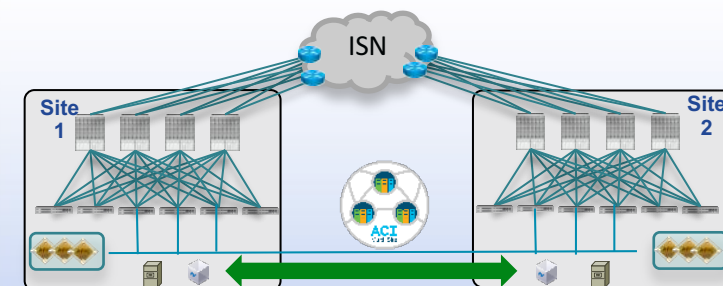
- Одна и так же IP подсеть определена на разных сайтах
- Поддерживается IP Mobility ('холодная' и 'живая'* миграция VM) и взаимодействие между хостами внутри одной подсети
- Нет Layer 2 BUM флдинга между сайтами

MSO GUI (BD)



Классическая L2 связанность

3



- Взаимодействие между сайтами с возможностями изоляции доменов
- L2 растягивается между сайтами, 'живая'* VM миграция и кластеризация приложений
- Есть Layer 2 BUM флдинг между сайтами

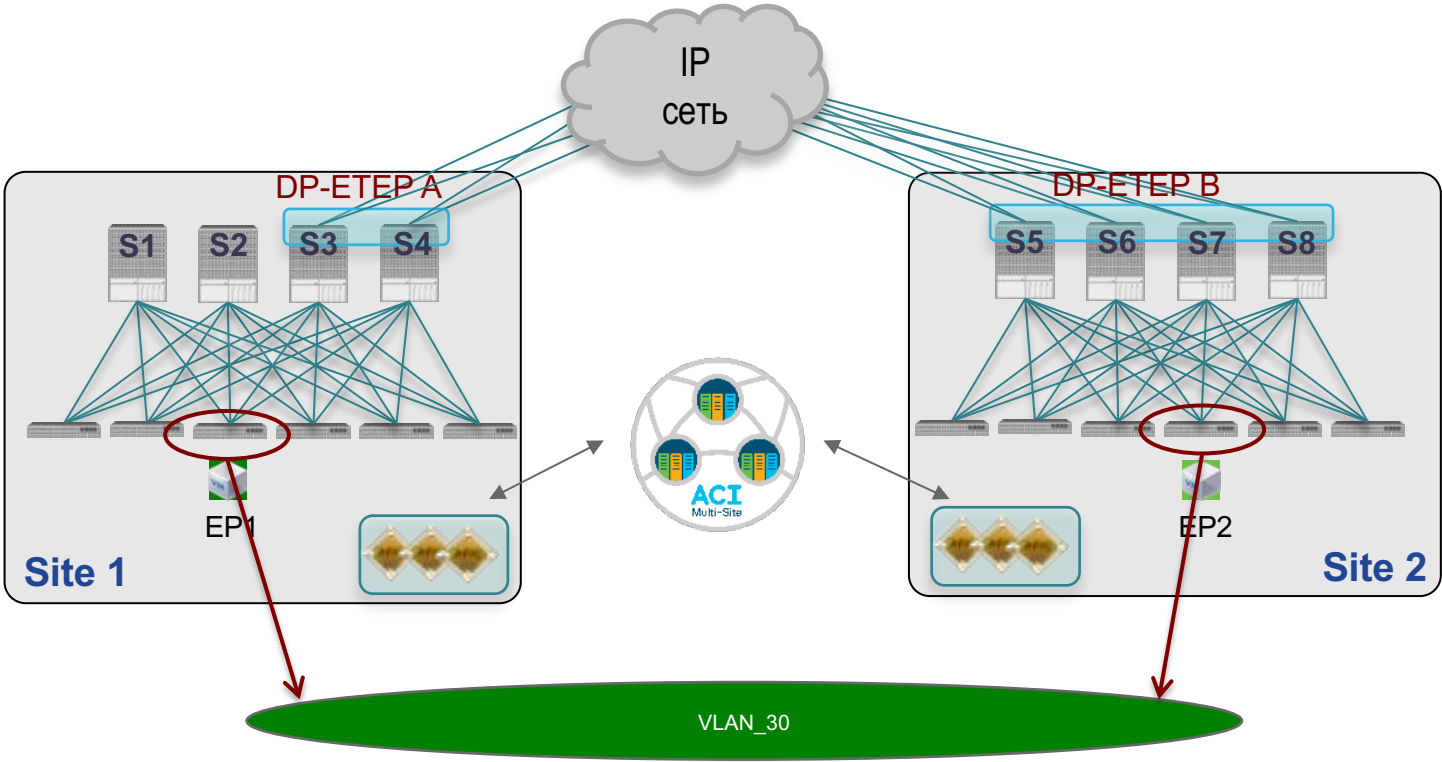
MSO GUI (BD)



*'Живая' миграция поддерживается начиная с релиза ACI 3.2

ACI Multi-Site

Демонстрация – организация L2 связанности между сайтами



ACI Multi-Site

Демонстрация – организация L2 связанности между сайтами

0

Создание tenant

1

Исходное состояние



ACI Multi-Site

Демонстрация – организация L2 связанности между сайтами

2

Создаем **общий** шаблон настроек для 2-х сайтов

3

Создаем VRF, который автоматически растягивается между сайтами



ACI Multi-Site

Демонстрация – организация L2 связанности между сайтами

4

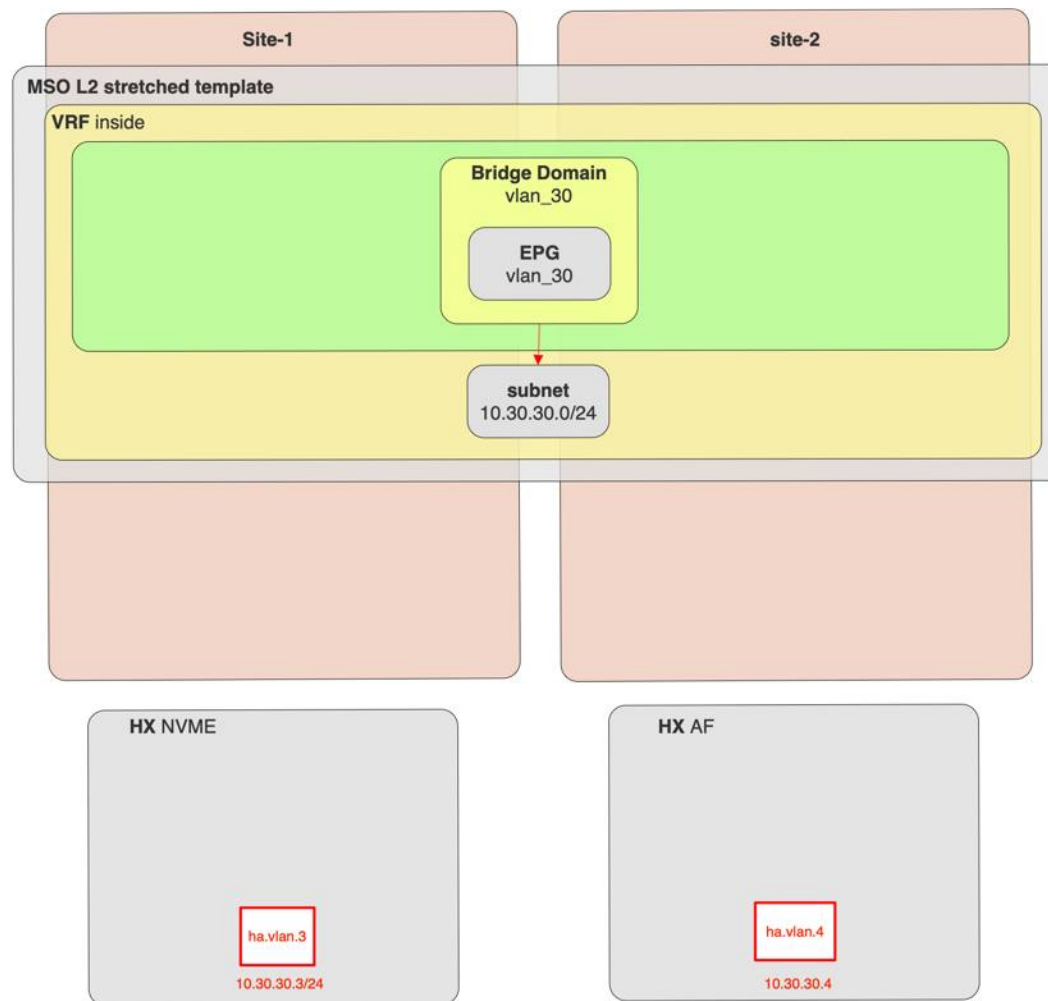
Создаем профиль приложения (контейнер для зон безопасности = EPG)

5

Создаем растянутый L2 сегмент = Bridge Domain

6

Настраиваем L3 subnet



ACI Multi-Site

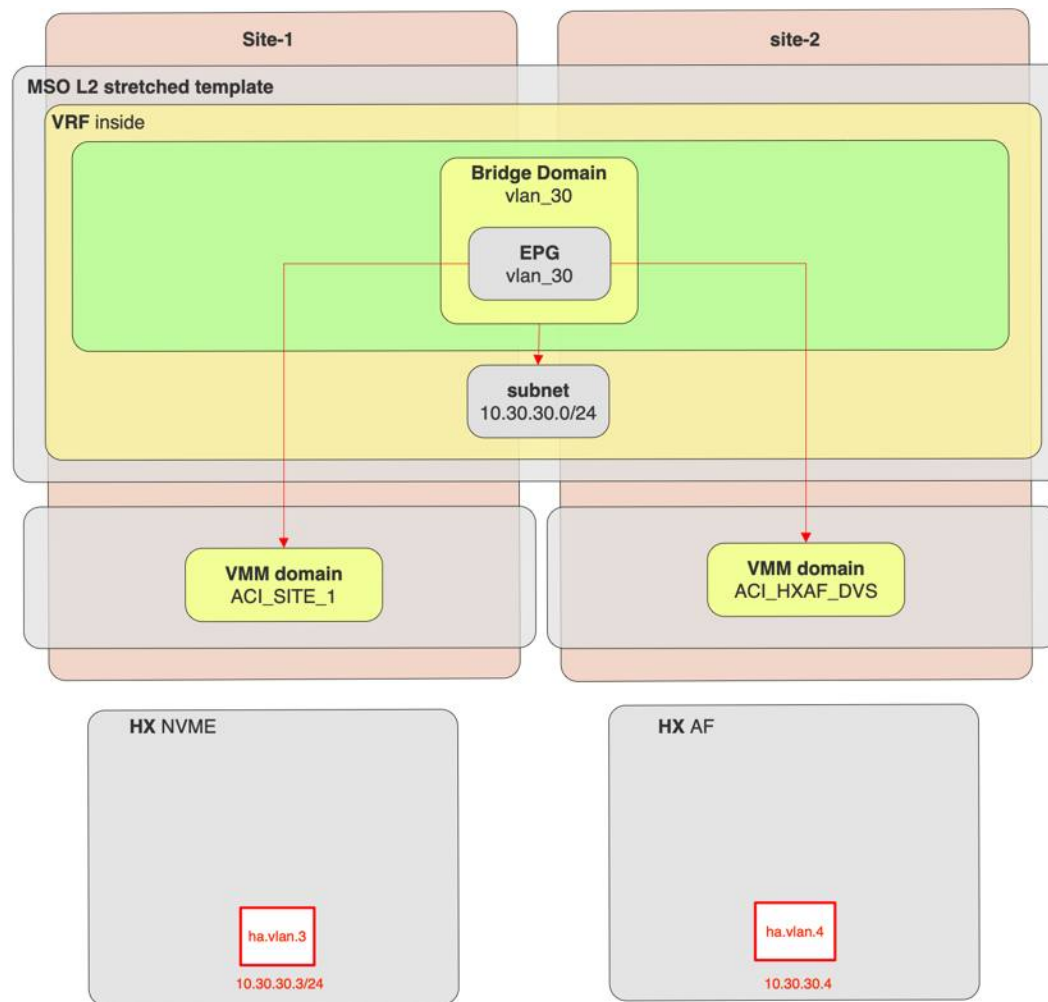
Демонстрация – организация L2 связанности между сайтами

7

Настраиваем зону безопасности = EPG

8

Выполняем ассоциацию с VMM-доменами на уровне настроек EPG внутри сайта

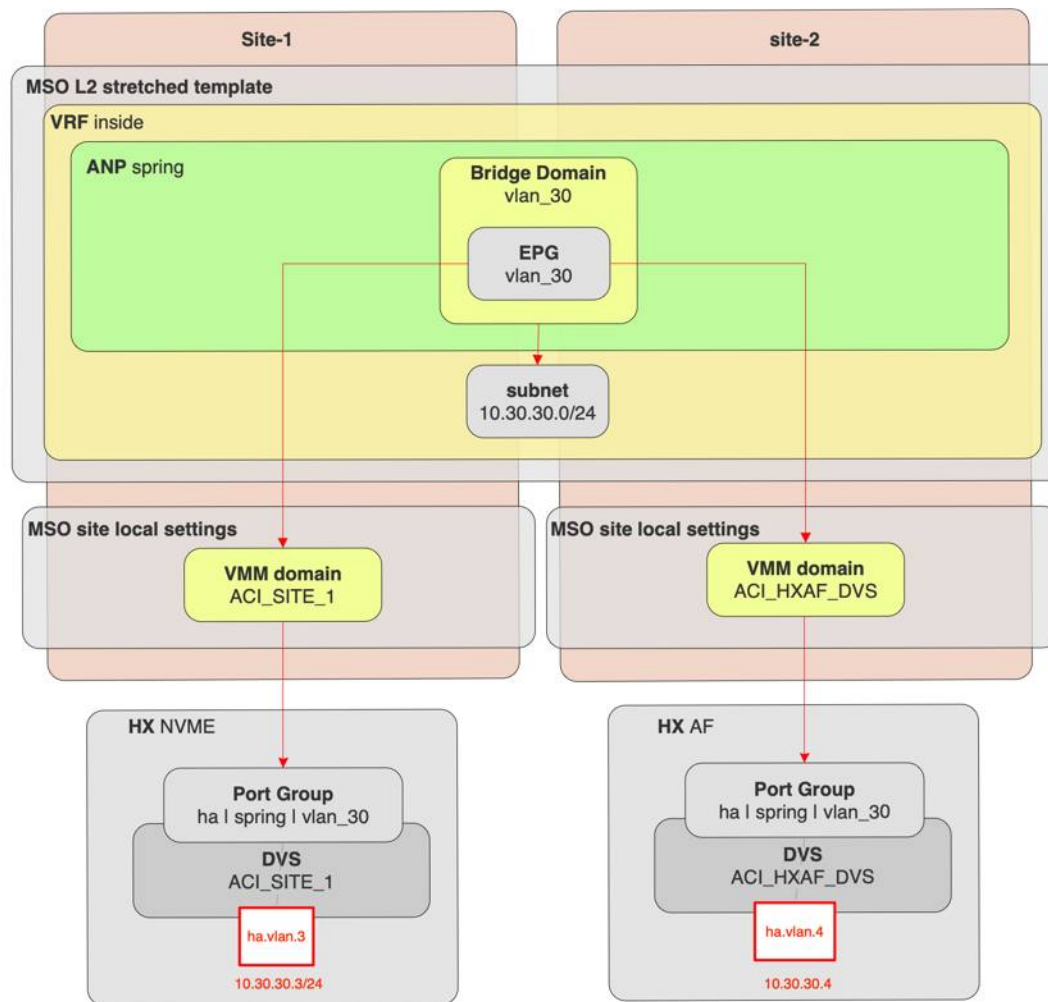


ACI Multi-Site

Демонстрация – организация L2 связанности между сайтами

9

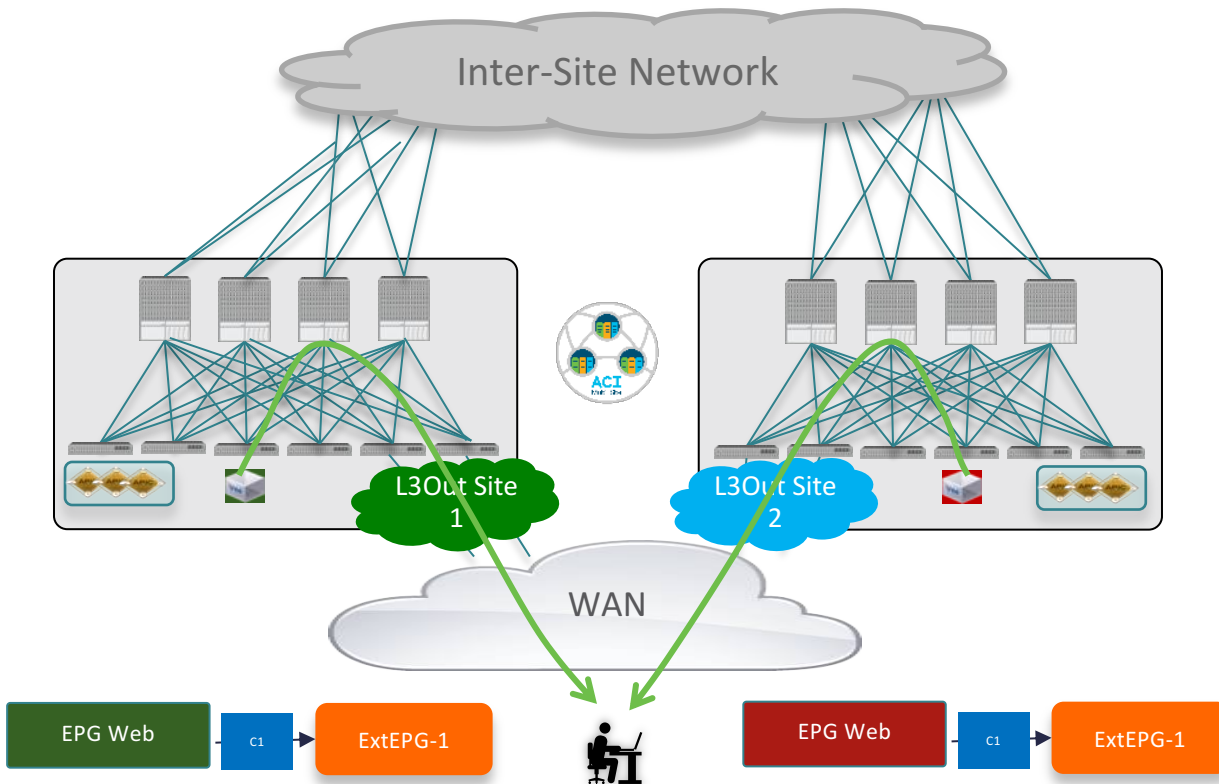
APIC автоматизирует
создание портовых
групп на DVS и VLAN
на системе
гиперконвергенции
Cisco Hyperflex



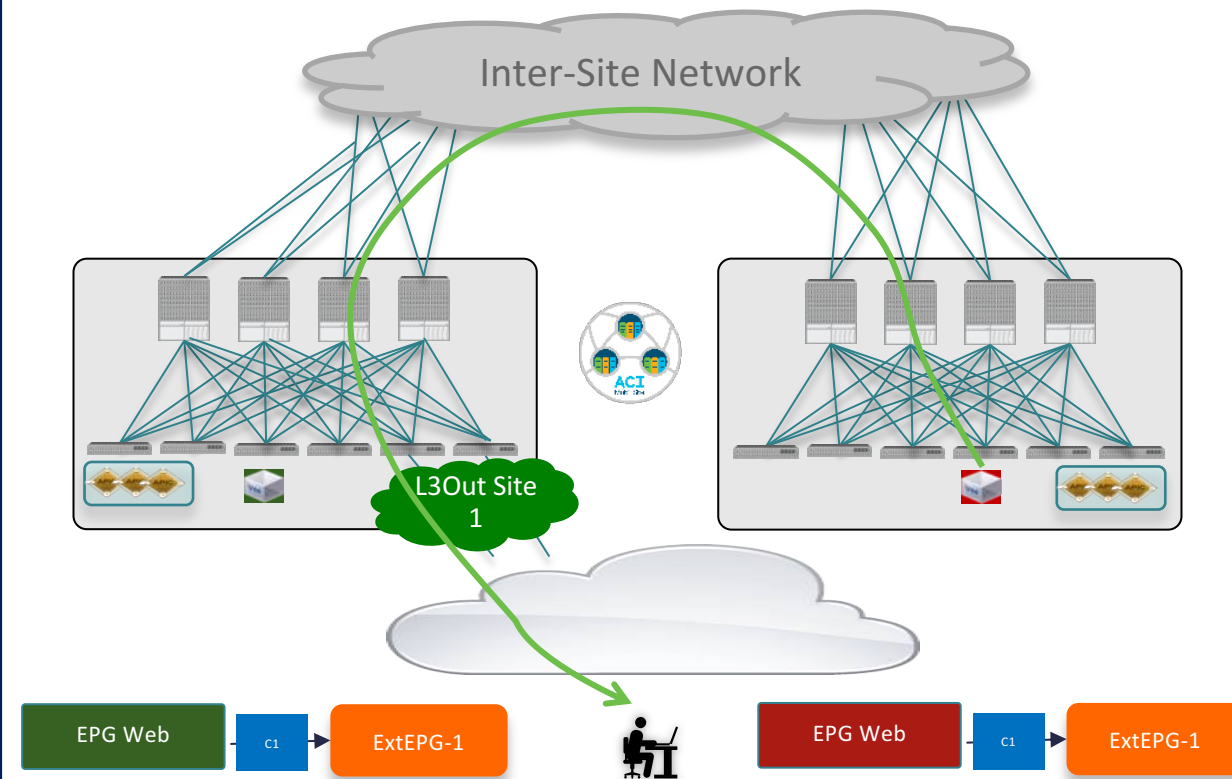
Multi-Site и L3Out

Различные возможности по организации внешних подключений

Независимые L3out

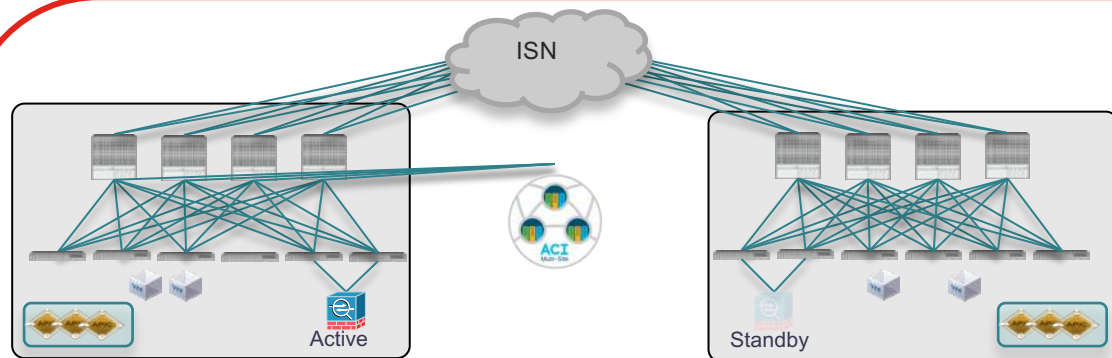


Использование L3out другого сайта

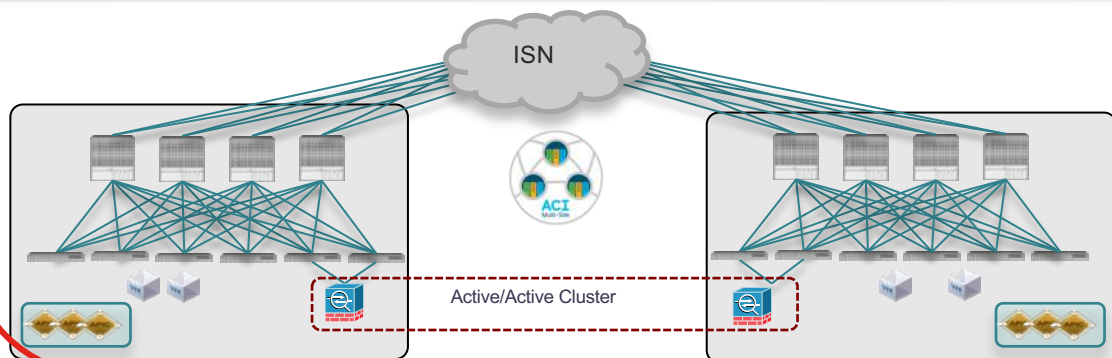


Multi-Site и режимы интеграции L4/L7 устройств

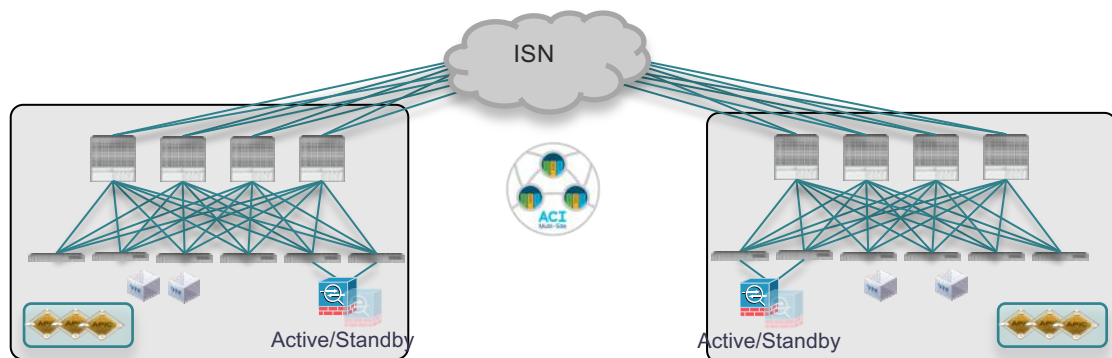
Подробности завтра в пятницу – 5 февраля



- Одна пара Active/Standby устройств «растянутая» между сайтами
- **Поддерживается если MCЭ работают в L2 режиме или в L3 режиме, но используются как шлюзы по умолчанию для нагрузки**



- Active/Active кластер MCЭ растянутый между сайтами (один логический FW)
- Requires the ability of discovering the same MAC/IP info in separate sites at the same time
- **Не поддерживается и не рекомендуется в Multi-Site**



- **Рекомендуется для режима ACI Multi-Site**
- Вариант 1: с версии ACI 3.0 для режима N-S если FW подключается в L3 режиме к фабрике → обязательно использовать оптимизацию для вх. трафика
- Вариант 2: с версии ACI 3.2 с использованием сервисного графа и Policy Based Redirection (PBR)

Nexus Dashboard

Единая платформа для хостинга MSO и приложений Day2ops

Просто внедрить
Просто использовать

Multi-Site
Orchestrator



Nexus
Insights



Nexus
Assurance

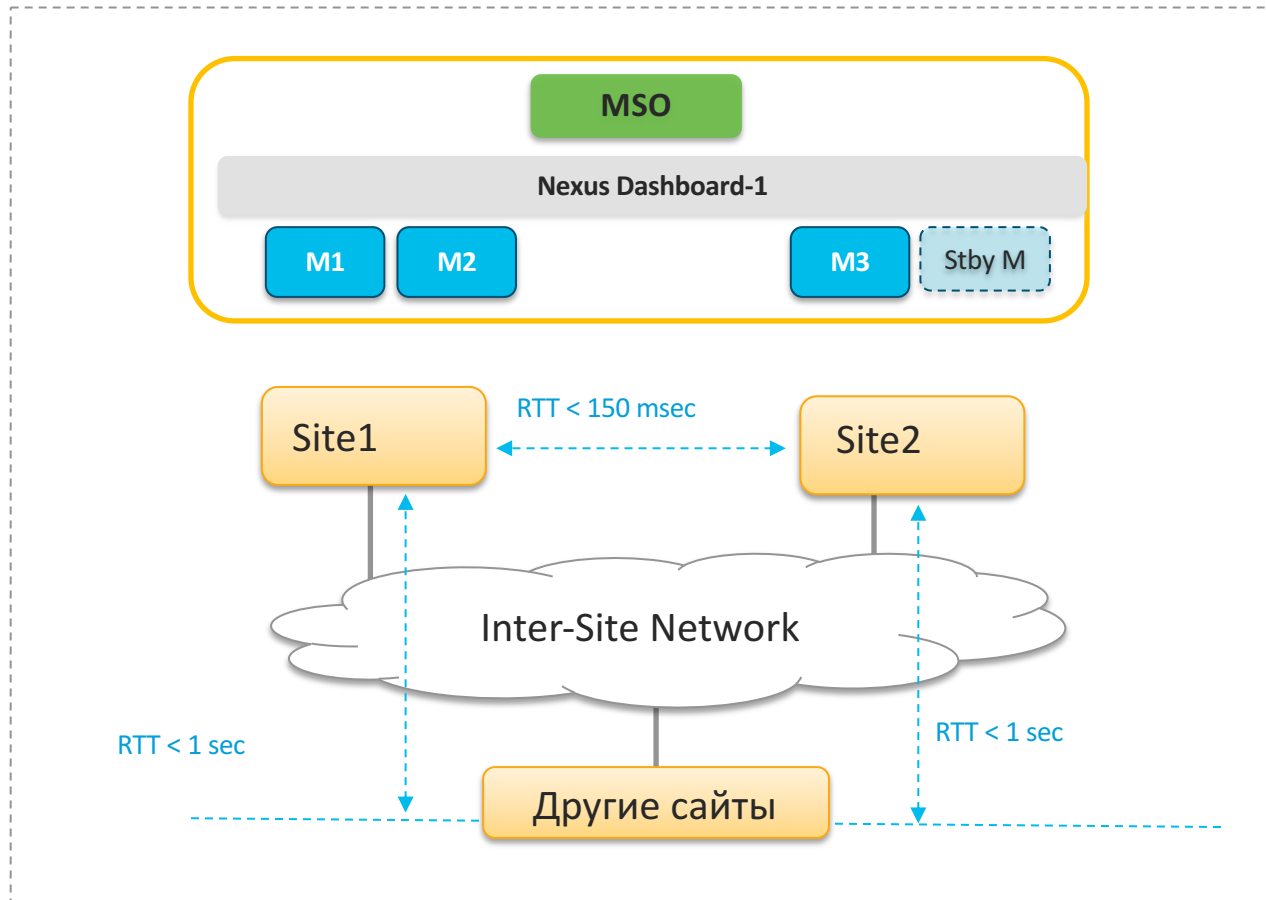


3rd Party
Tools



Nexus Dashboard: единое окно в эксплуатацию ACI и NX-OS фабрик

MSO на Nexus Dashboard



Задержка между ND узлами: < 150 msec

Latency between DC network to MSO nodes: RTT < 1sec

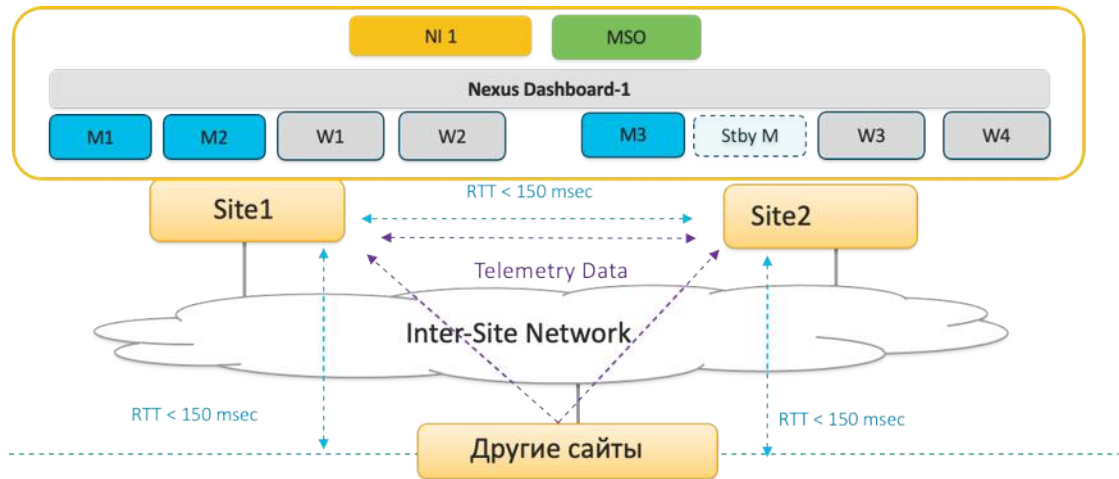
Все сервисы MSO работают на master nod-ах Все три master узла активны

Некоторые сервисы реплицируются на worker nod-ы

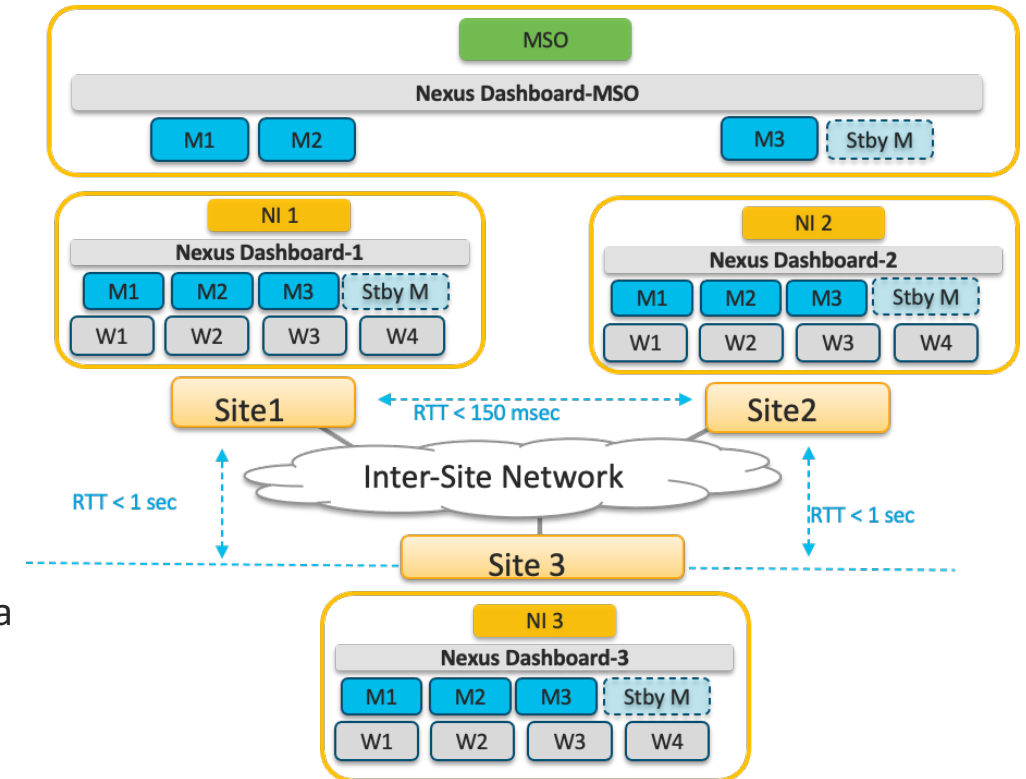
Для отказоустойчивости MSO узлы ND master должны быть развернуты на разных сайтах

MSO + NI на Nexus Dashboard

Вариант 1: MSO & NI размещаются на одном ND кластере



Вариант 2: MSO & NI размещаются на разных ND кластерах



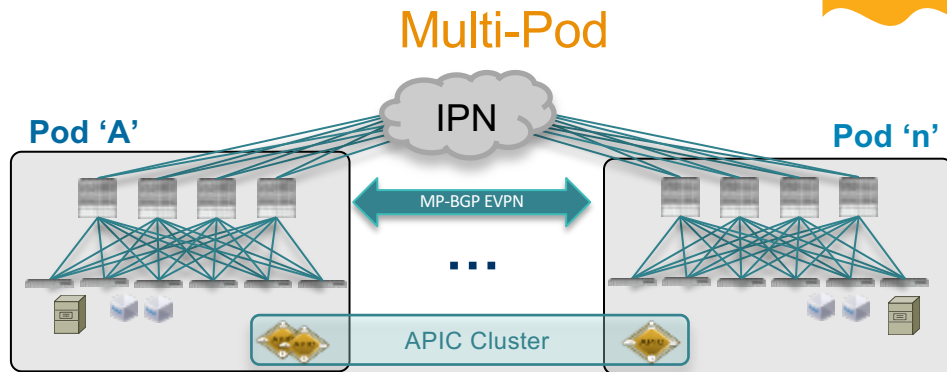
- Отказоустойчивость для приложения MSO и Day2ops продукта Tychus Insights
- Единый комплекс на несколько сайтов
- Телеметрия должна передаваться поверх сети ISN
- Масштабирование ограничено максимальным размером кластера ND

- Два приложения разворачиваются независимо
- MSO разворачивается на распределенном ND кластере
- Приложение NI может быть развернуто на разных ND кластерах
- Телеметрия локализована на каждом их сайтов

ACI Multi-Pod и ACI Multi-Site

Основные отличия

HA



Простота
эксплуатации

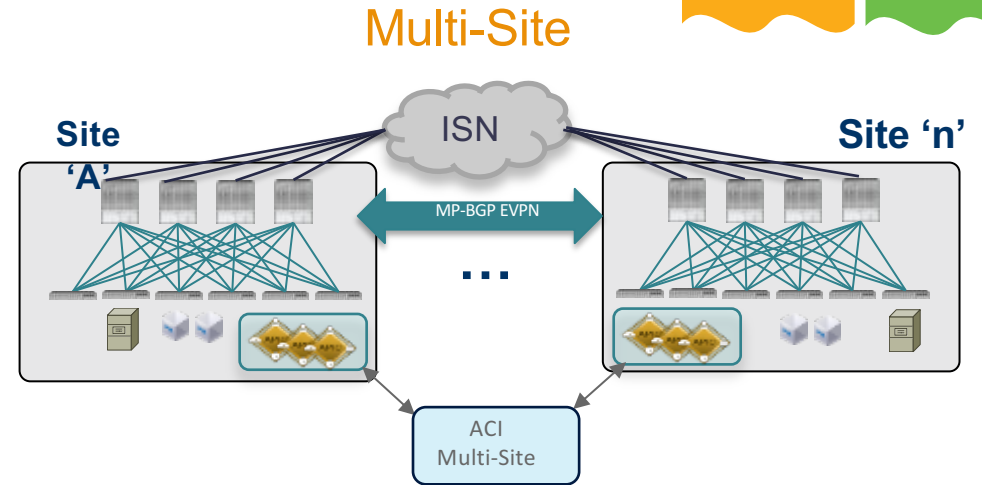
Меньше число
APIC

Сквозные
функции между
модулями

Единые VMM на
все модули

HA

DR



Изоляция доменов
настройки

Более высокая
допустимая задержка
на транспорте

Выше число
узлов

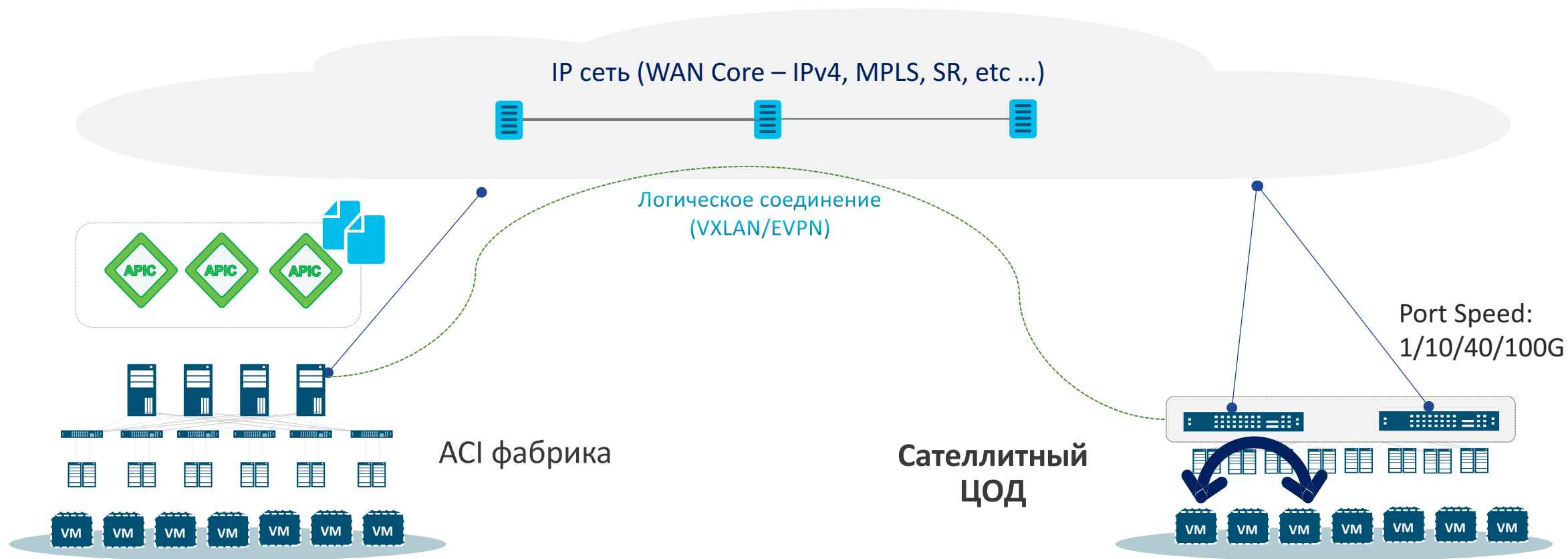
Не нужен
multicast на
транспорте

Требуется MSO и
лицензии Advantage

Решения для гибридных
сценариев на основе ACI
Virtual PoD, ACI Remote Leaf,
ACI Cloud APIC

«Сателлитные ЦОД»: ACI Remote Leaf

Вариант подключения традиционных сетей в Multi-Pod архитектуре



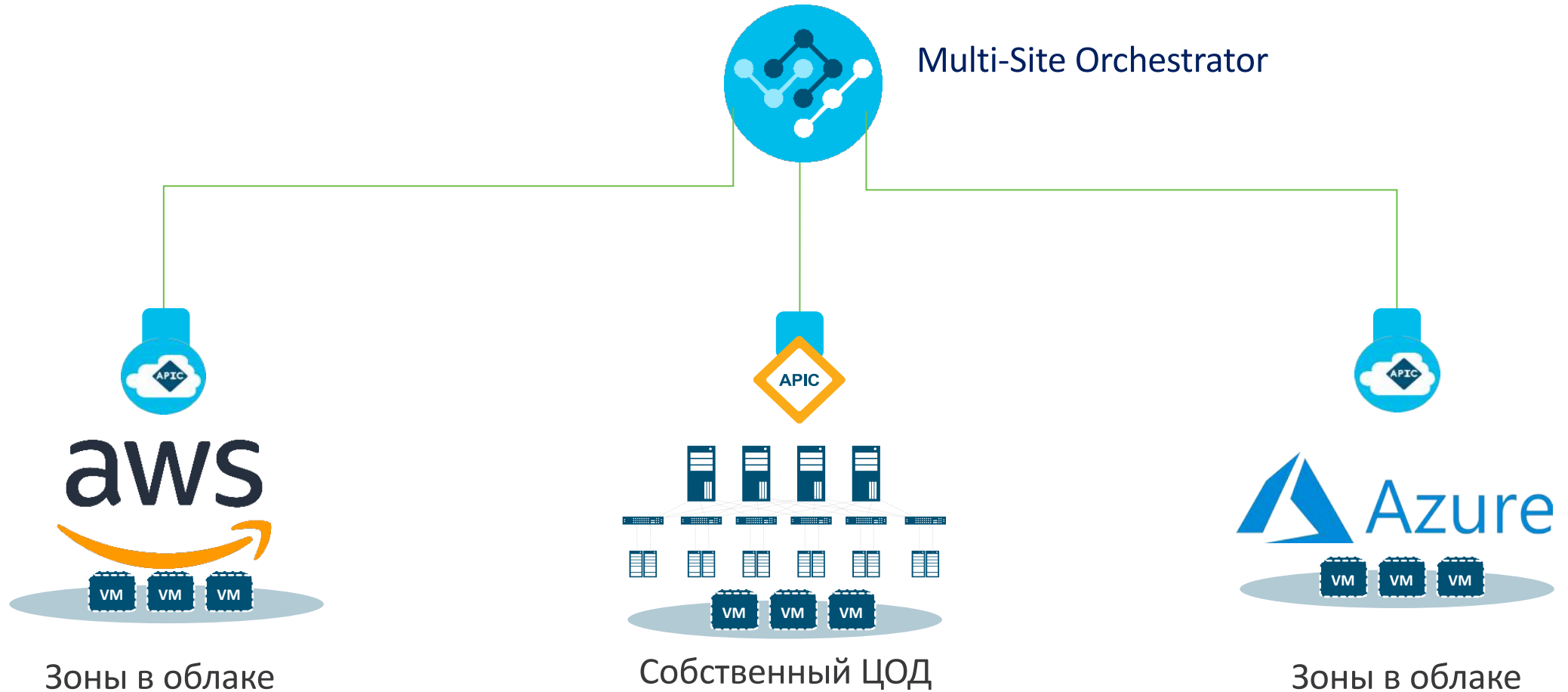
Автоматическое обнаружение
вынесенных коммутаторов

До 64 сателлитных ЦОД
(пар Remote Leaf)

Продолжение EPG, BD, VRF, тенантов,
контрактов

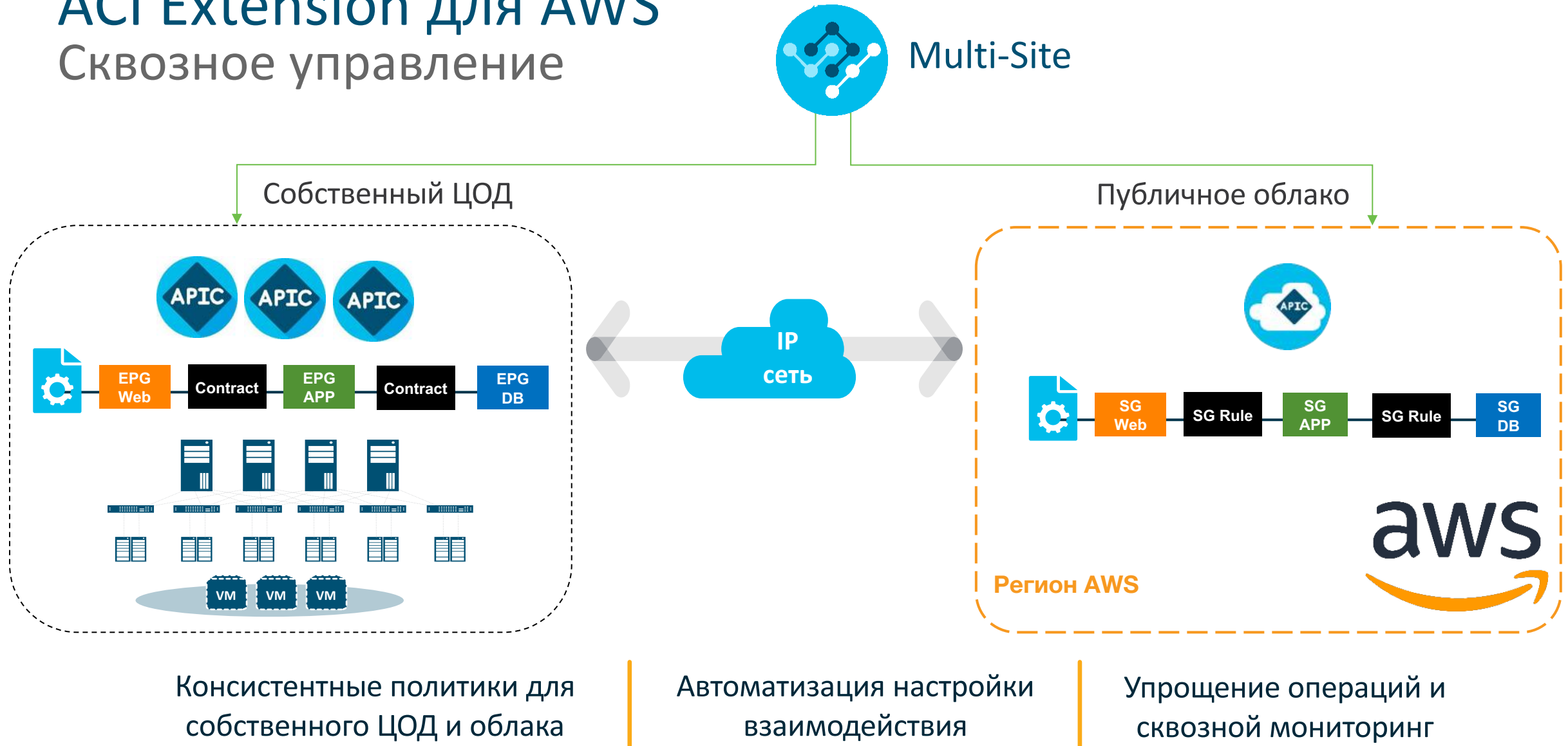
Диагностика и статистика

ACI Extensions для Multi-Cloud



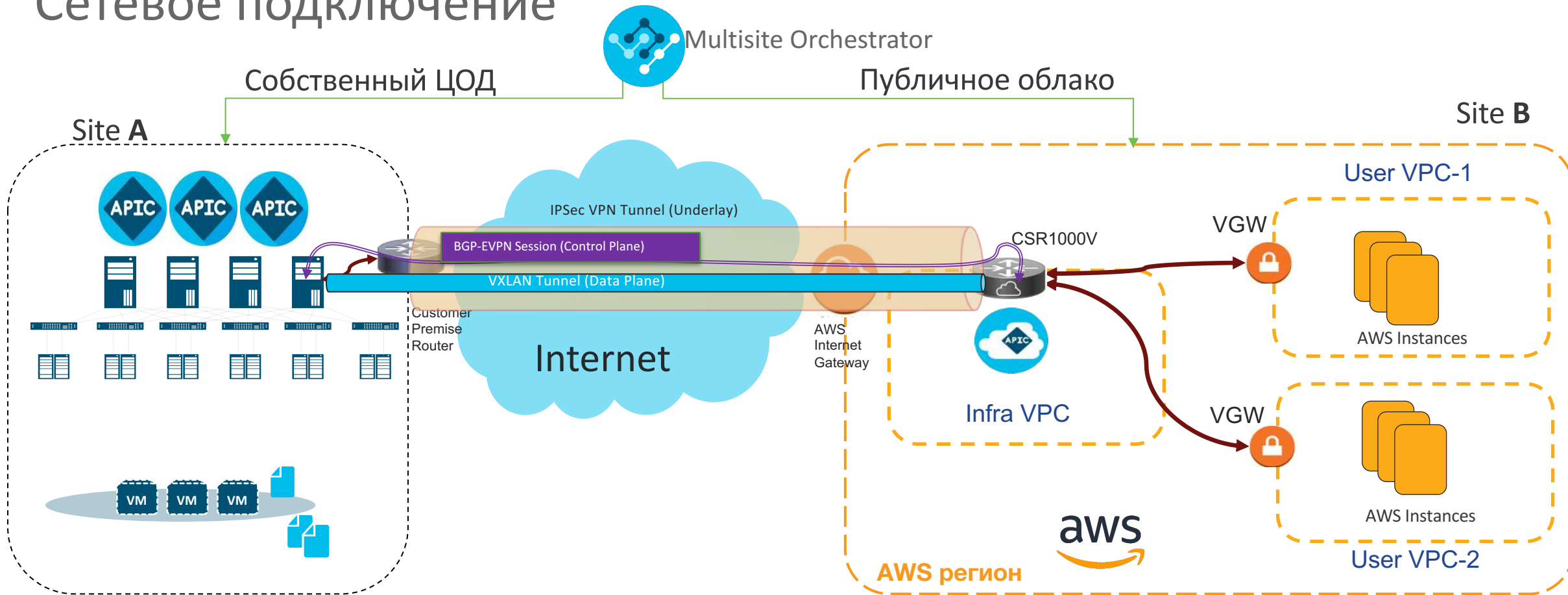
ACI Extension для AWS

Сквозное управление



ACI Extension для AWS

Сетевое подключение

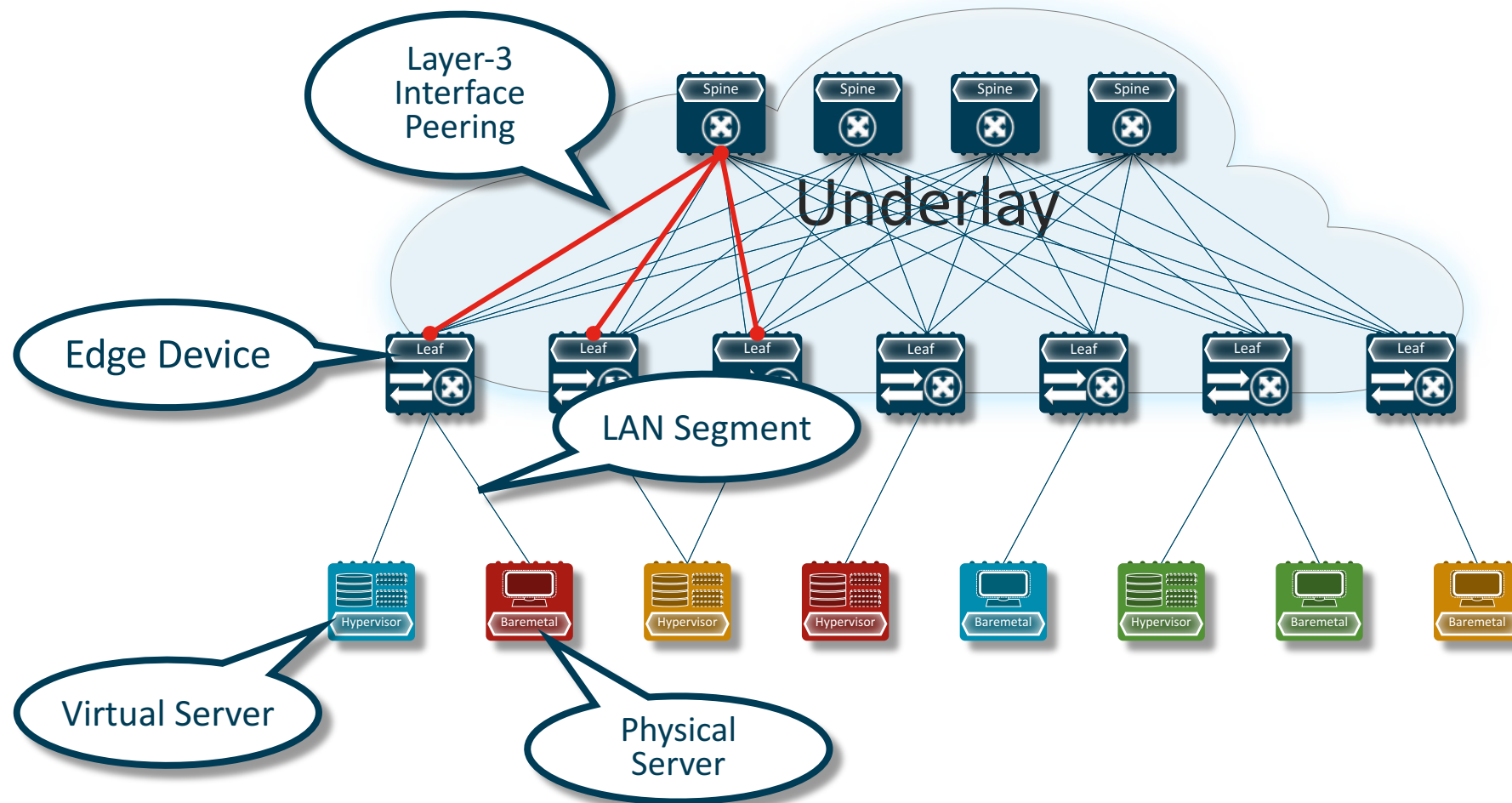


- Организация передачи данных при помощи VXLAN
- Обеспечение связанности при помощи BGP-EVPN
- Безопасное IPSec VPN соединение между маршрутизатором подключенном к ACI фабрике и CSR1kv

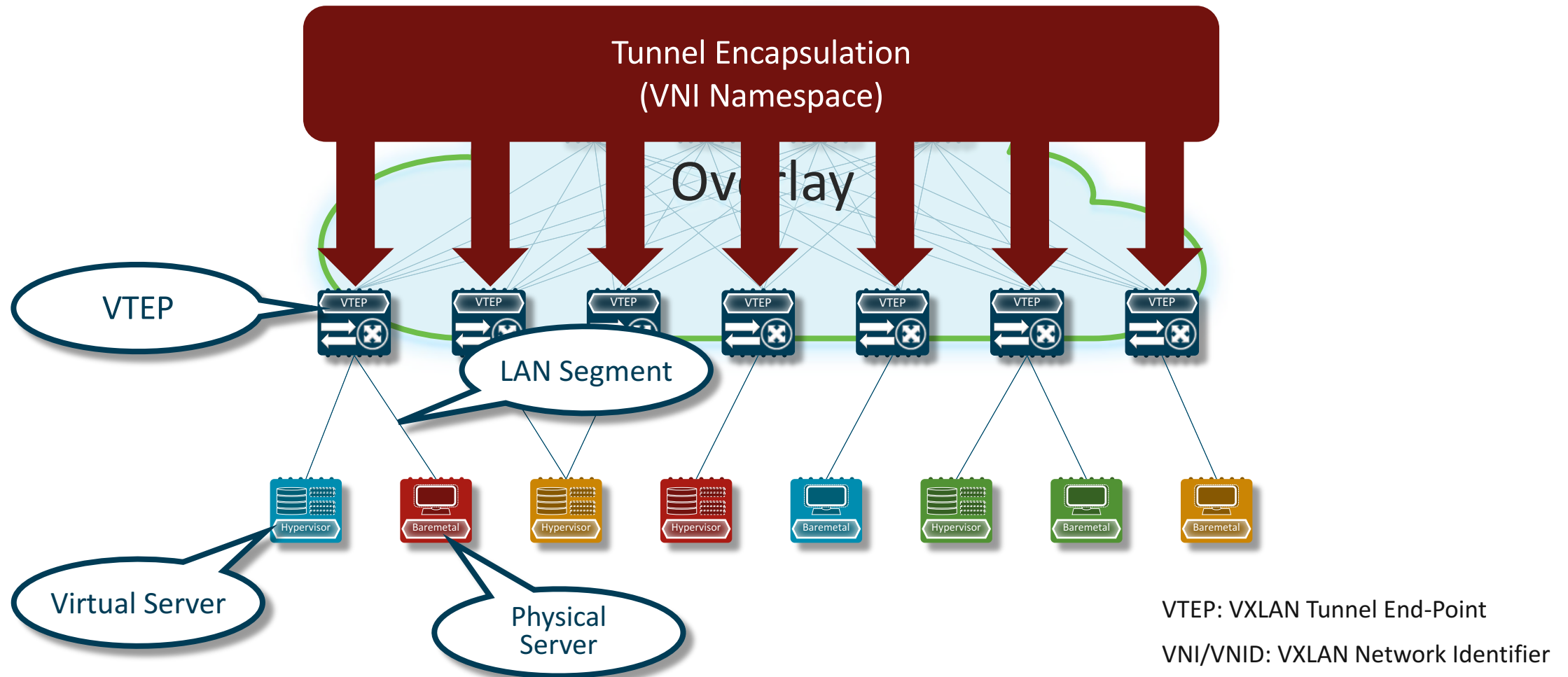
Часть 2 – Распределенные фабрики на базе VXLAN & EVPN

VXLAN&EVPN – краткое введение

Underlay – опорная сеть



Overlay – наложенная сеть предоставляющая сервисы



Архитектура VXLAN EVPN фабрики

Базовые составляющие сервиса

Оверлеи

- Layer-2
- Layer-3
- Layer-2 и Layer-3

VXLAN инкапсуляция

Underlay транспортная сеть

Плоскость управления на основе BGP

- Peer-Discovery
- Route Learning and Distribution
 - Local Learning
 - Remote Learning

Плоскость передачи данных

- Overlay Layer-2/Layer-3 Unicast Traffic
- Overlay Broadcast, Unknown Unicast Multicast traffic (BUM traffic) forwarding
- Ingress Replication (Unicast)
- Multicast

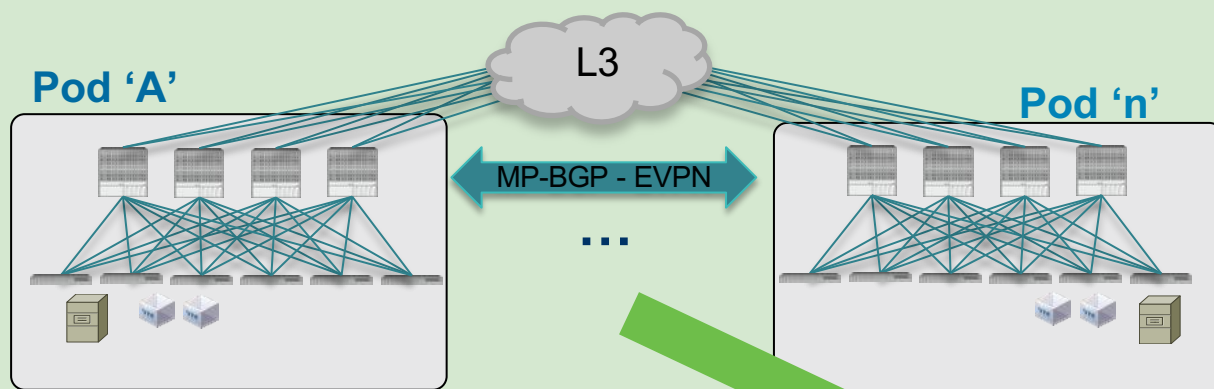
Распределенная сетевая фабрика на основе Cisco VXLAN & EVPN Multi-Site

Связь сетевых фабрик ЦОД на основе VXLAN/EVPN

VXLAN/EVPN Multi-Site – рекомендованный подход

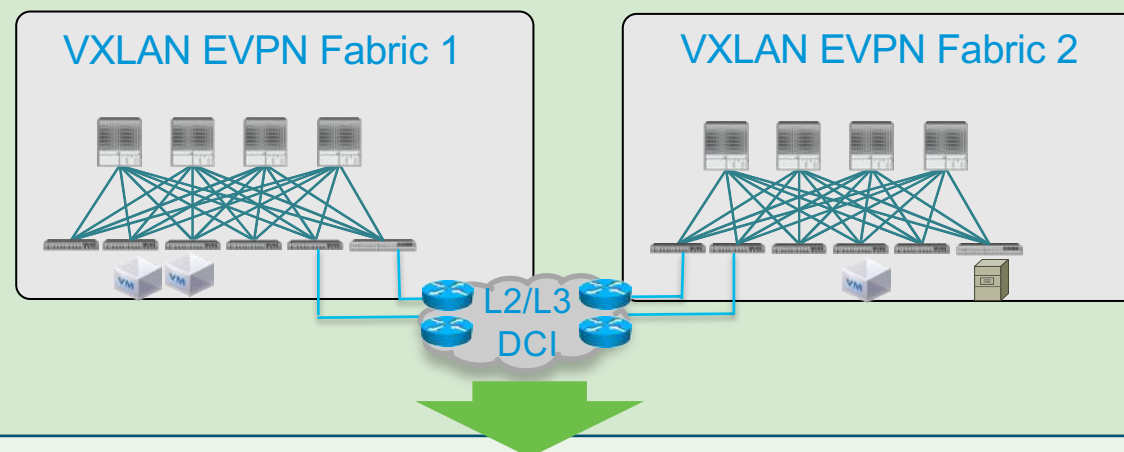
Растянутая фабрика

Multi-Pod – единая фабрика

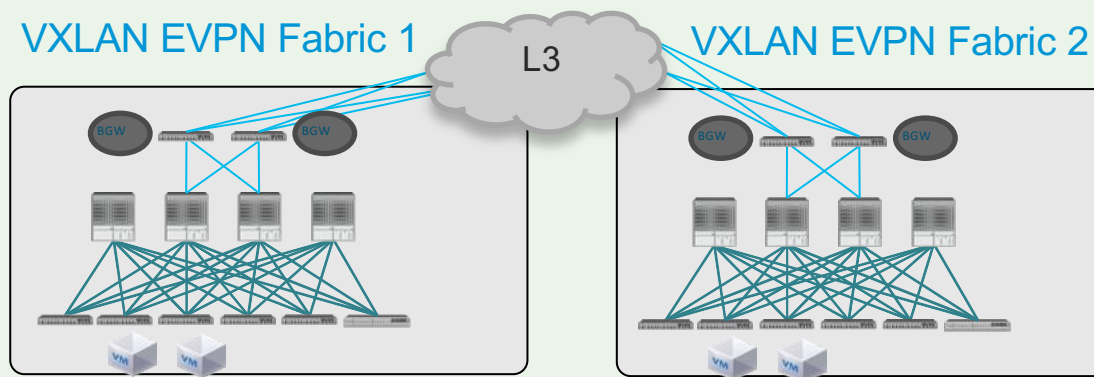


Изолированные фабрики

Multi-Fabric (L2 и L3 DCI)



VXLAN/EVPN Multi-Site

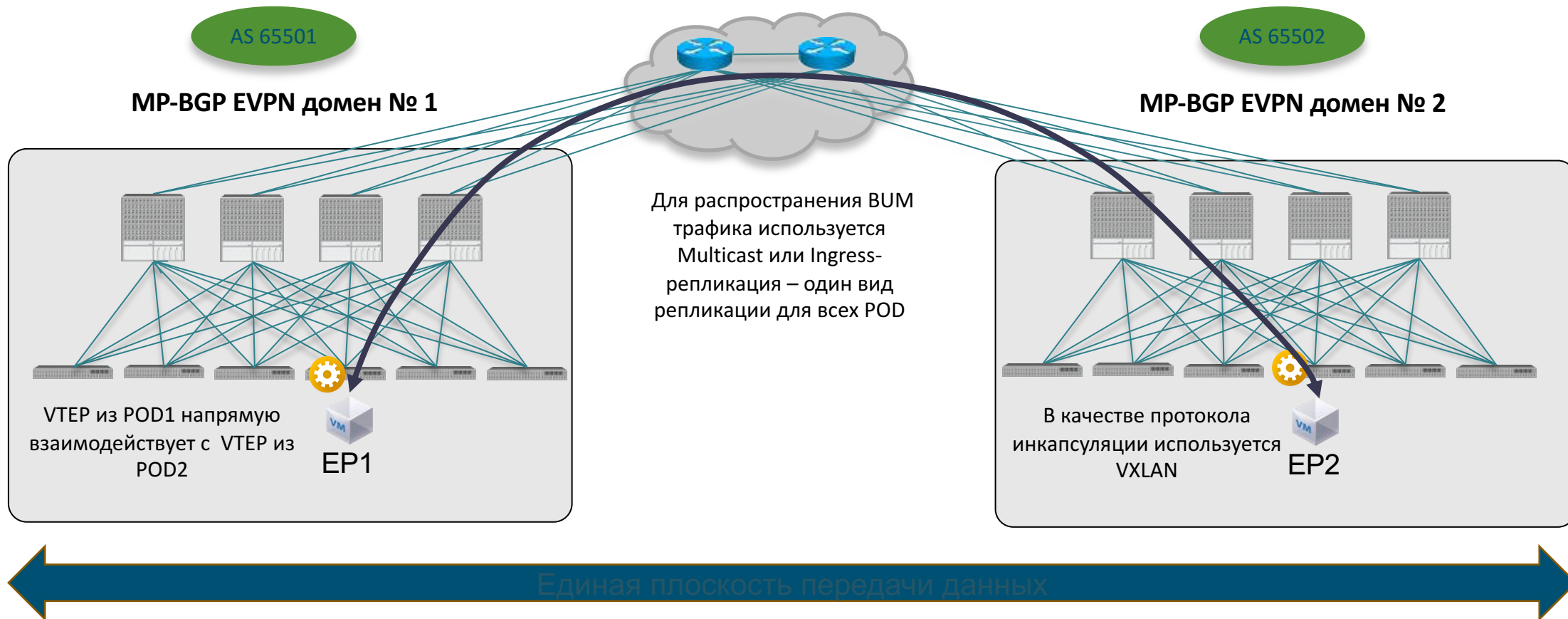


VXLAN EVPN Multi-Pod

Основные свойства архитектуры

⚙️ = VXLAN Encap/Decap

IPN сеть представляет собой
L3 транспорт с поддержкой
multicast (если требуется)



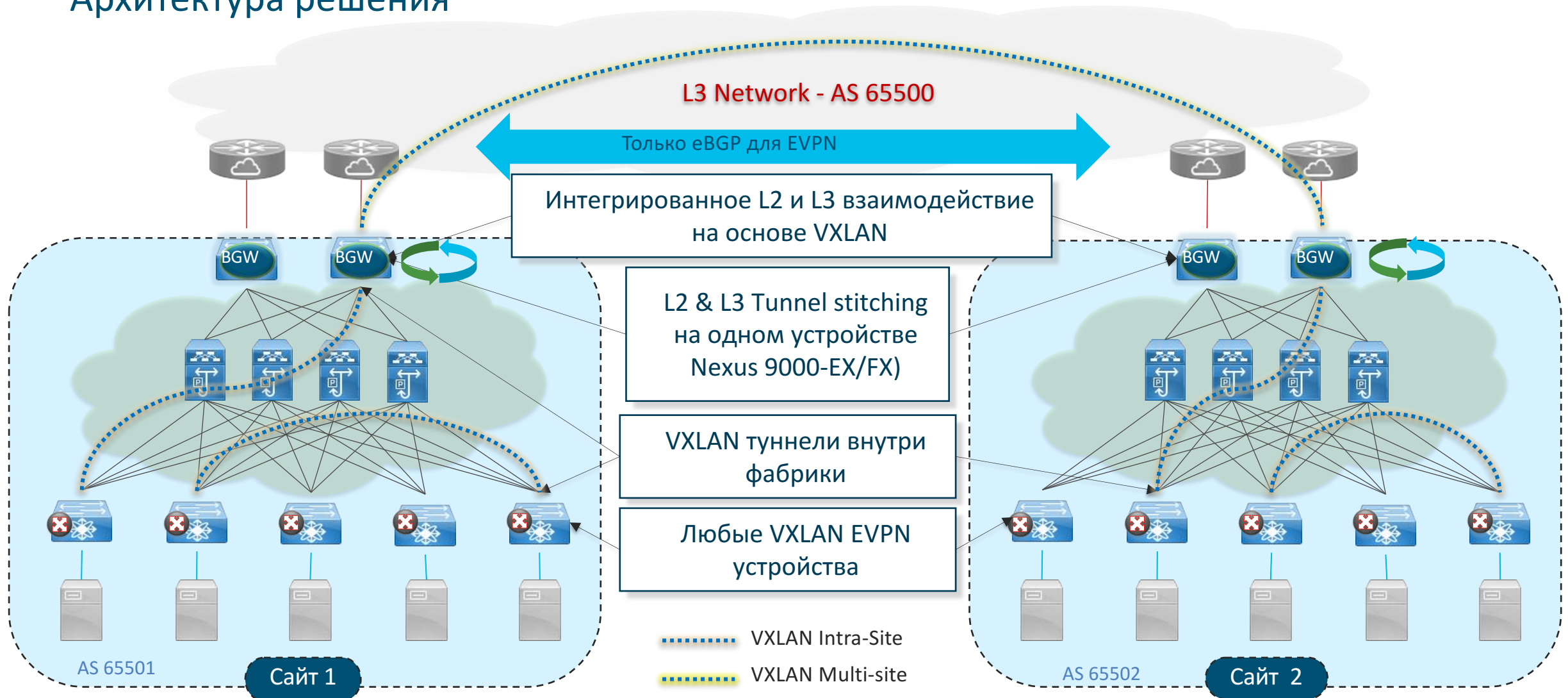
Multi-Site основные характеристики

- **Несколько** Overlay доменов – Interconnected & Controlled
- **Несколько** Overlay Control-Plane Domains – Interconnected & Controlled
- **Несколько** Underlay доменов - изолированных
- **Несколько** Replication доменов для BUM – Interconnected & Controlled

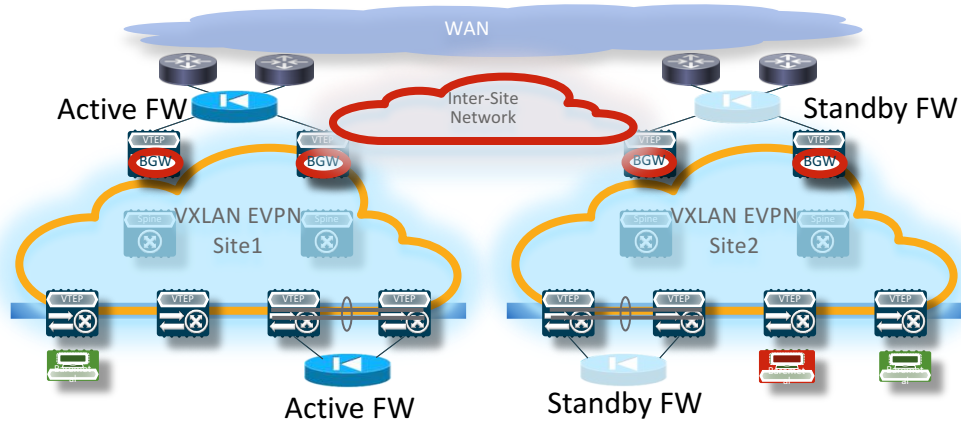
Изоляция на уровне Underlay и иерархия Overlay

VXLAN EVPN Multi-Site

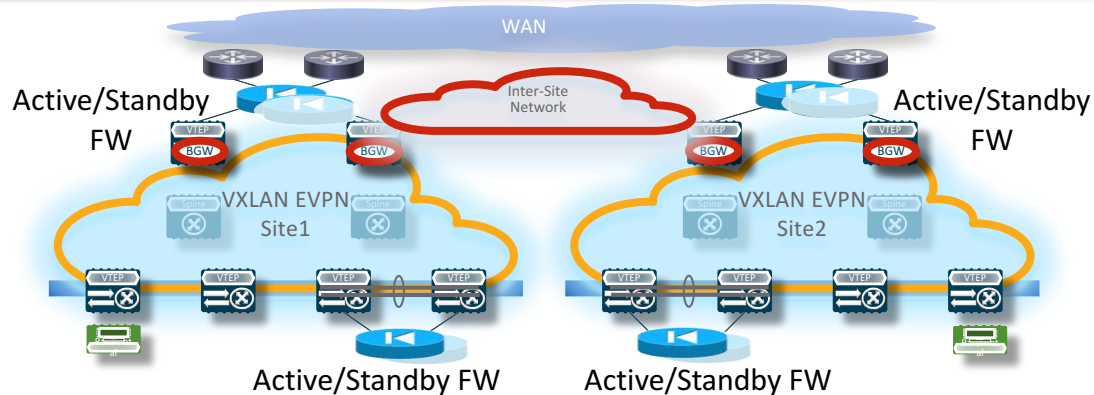
Архитектура решения



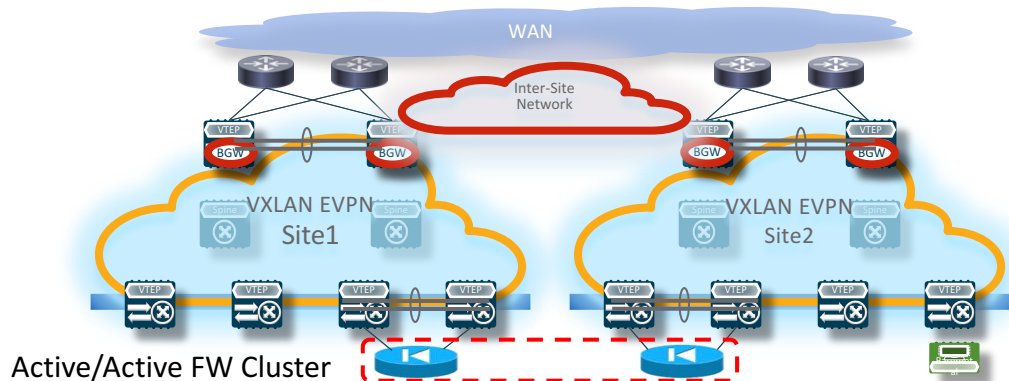
VXLAN EVPN Multi-Site и интеграция МСЭ



- Active и Standby устройства разворачиваются каждое на своем сайте, контроль для N-S и E-W потоков
- Проблемы с асимметрией отсутствуют
- Разные варианты МСЭ как шлюз по умолчанию или фабрика как шлюз по умолчанию для хостов



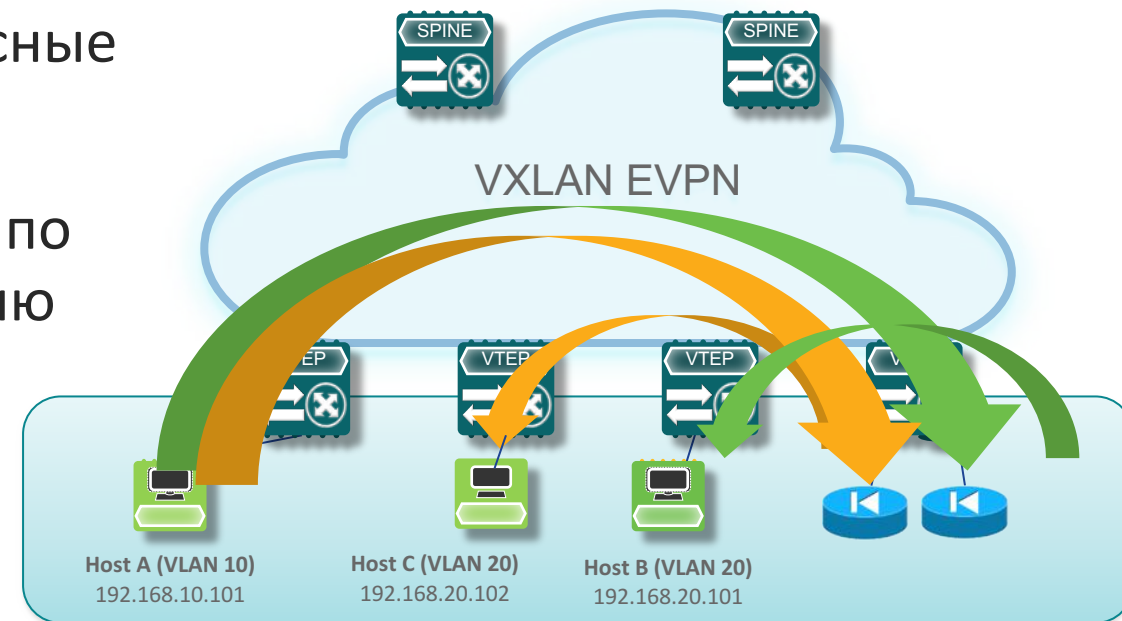
- Независимые пары Active/Standby на каждом сайте
- Требуется избегать асимметрии при организации передачи данных через МСЭ
 - Поддерживаются N-S потоки для этого нужно организовать анонсирование host routes advertisement или настроить PBR



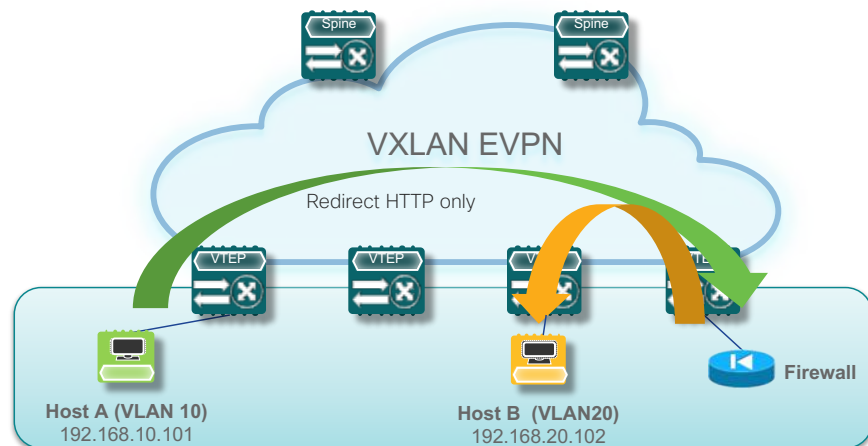
- Active/Active МСЭ кластер «растянутый» между сайтами
 - **Режим Split spanned ether-channel : не поддерживается,**
 - Режим Individual: поддерживается только для ПО Cisco ASA для N-S и E-W потоков

Интеграция сервисных устройств с помощью Enhanced PBR

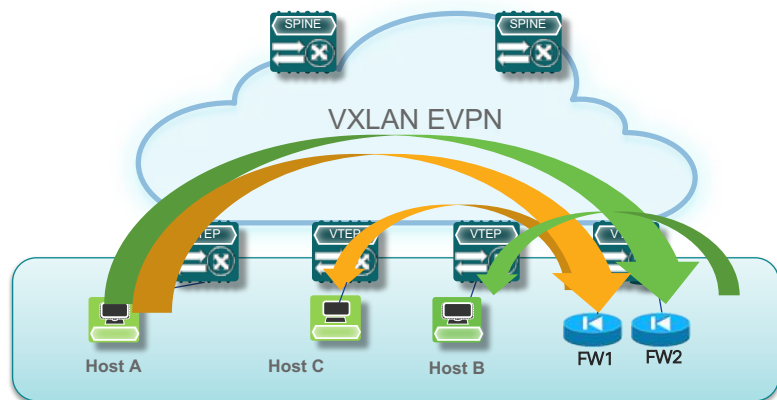
- EPBR позволяет гибко интегрировать сервисные устройства
- Поддерживается выборочная возможность по полному, либо частичному перенаправлению трафика на одно или несколько устройств
- Интегрированные возможности для балансировки
- Возможность мониторинга доступности сервисного устройства
- Line-rate!



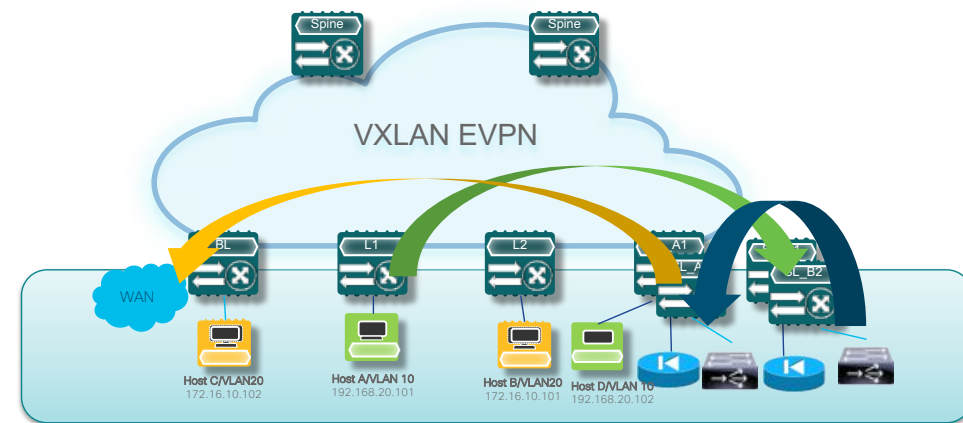
Примеры сценариев



Выборочное перенаправление трафика



Перенаправление трафика и
балансировка нагрузки



Возможность сформировать
сервисную цепочку

И многое другое..

Больше деталей?

Подпишись на наш канал в Telegram и оставайся на связи!

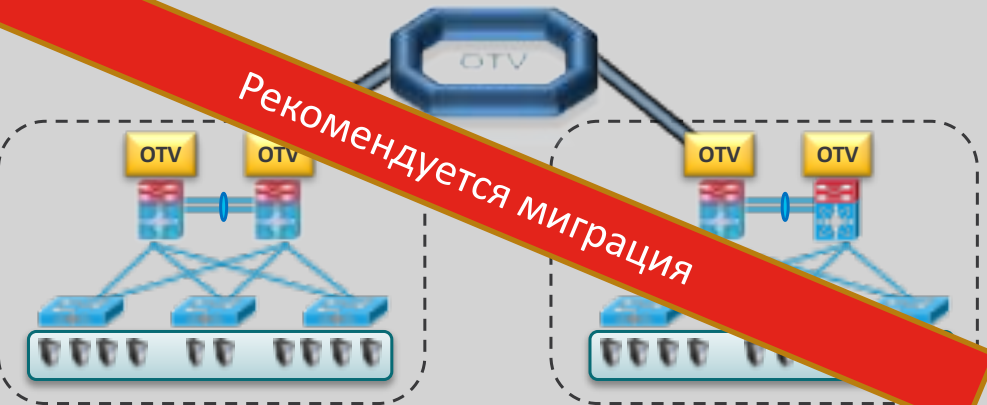
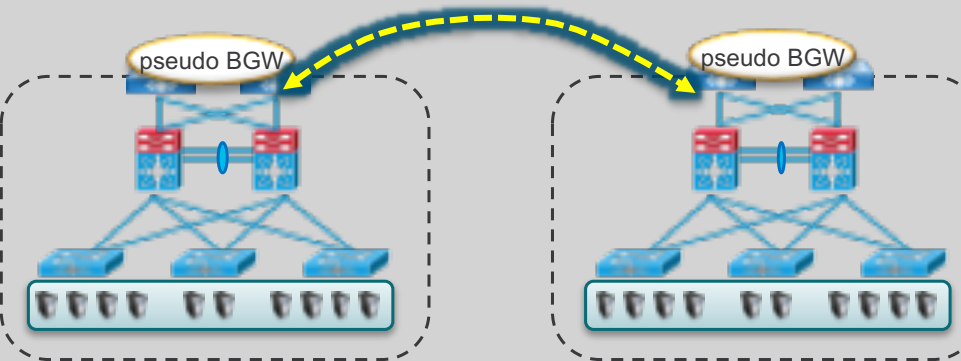


NX-OS марафон, весна 2021

- Основы VXLAN & EVPN
- Детальный разбор архитектуры Cisco VXLAN & EVPN Multi-Site
- Расширенная интеграция L4/L7
- Tenant Routed Multicast
- Управление с помощью DCNM и MSO
- Day2 Operations

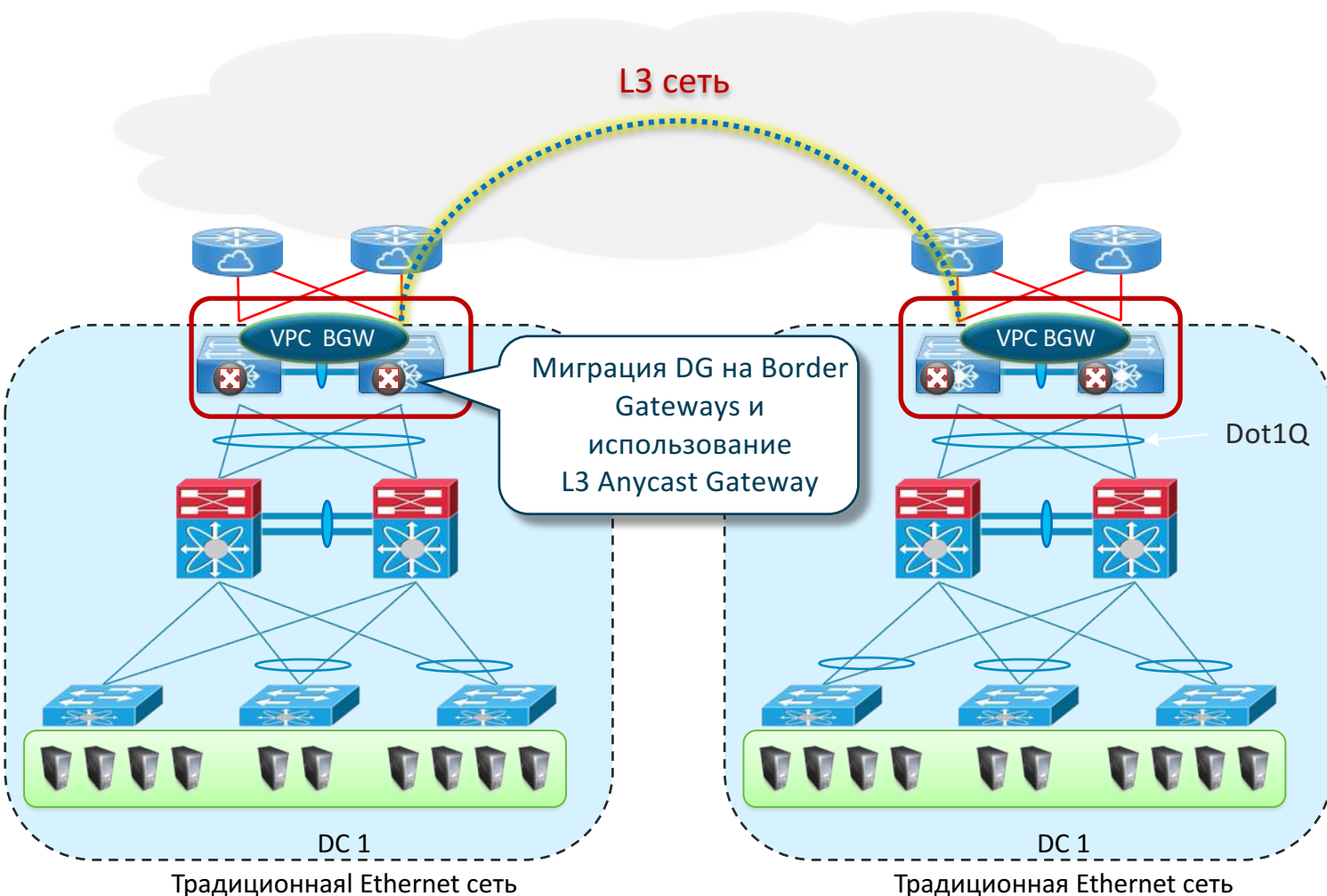
Часть 3 – Организация сетевых подключений для традиционных ЦОД

Объединение традиционных сетей ЦОД

Использование OTV	Использование VXLAN EVPN
Классический OTV дизайн  Рекомендуется миграция	VXLAN EVPN Multi-Site 
OTV как часть фабрики  Рекомендуется миграция	

VXLAN EVPN Multi-Site для связи традиционных ЦОД

VXLAN EVPN VPC BGW



- Пара VPC BGW устройств на каждом сайте
- Миграция с классических архитектур на архитектуру VXLAN EVPN Multi-Site
- Поддержка функций Anycast Gateway
- Интегрированная возможность Multi-Tenancy (L3VPN)
- Возможность постепенного перехода на Spine & Leaf топологию
- Контроль за BUM-трафиком (блокирование или ограничение)
- Control Plane with Selective Advertisement (L2 & L3)

