



Сеть ЦОД на основе VXLAN/EVPN за 5 дней

День 3: VXLAN/EVPN фабрика – распределенные топологии

Максим Хаванкин
mkhavank@cisco.com

Обсуждение – в Telegram канале:



Программа спринта

День	Тема
5 апреля, понедельник	VXLAN/EVPN фабрика – основы
6 апреля, вторник	VXLAN/EVPN фабрика – клиенты и внешние сегменты, расширенные функции NX-OS
7 апреля, среда	VXLAN/EVPN фабрика – распределенные топологии
8 апреля, четверг	VXLAN/EVPN фабрика – интеграция L4-L7 сервисов
9 апреля, пятница	VXLAN/EVPN фабрика – эксплуатация и поддержка

Содержание

- Развитие решений для распределенных топологий
 - Основные свойства архитектуры
 - Поддерживаемые топологии
 - Режимы работы BGW
 - Развитие решений по управлению распределенными фабриками
 - Регионы и зоны доступности
 - Интеграция DCNM и MSO
 - Демонстрация
-
- Плоскости передачи данных и управления
 - Плоскость протоколов управления
 - Растягивание BUM-трафика между сайтами
 - Растягивание L3 между сайтами
 - Демонстрация
 - Растягивание L2 между сайтами
 - Демонстрация
 - Подключение ко внешним L3-сетям
 - Оптимизация входящего и исходящего трафика

Часть 1

Перерыв – 10 мин

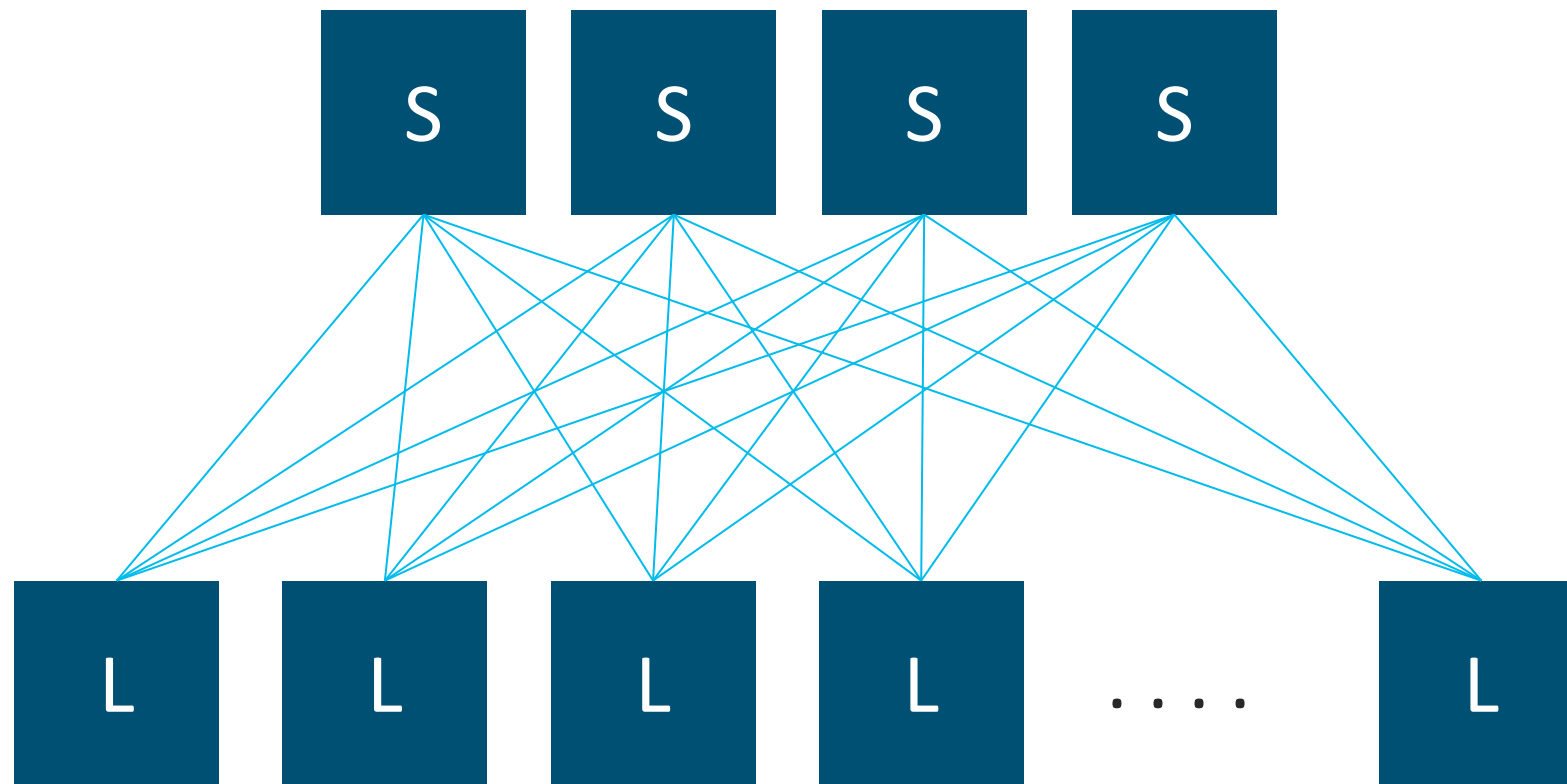
Часть 2

Что не вошло...

- Разбор сценариев сбоя
 - Интеграция с TRM
 - Миграция
-
- <https://www.ciscolive.com/global/on-demand-library.html?search=DGTL-BRKDCN-2035#/session/1573153544426001Jy2f>

Развитие решений для распределенных топологий

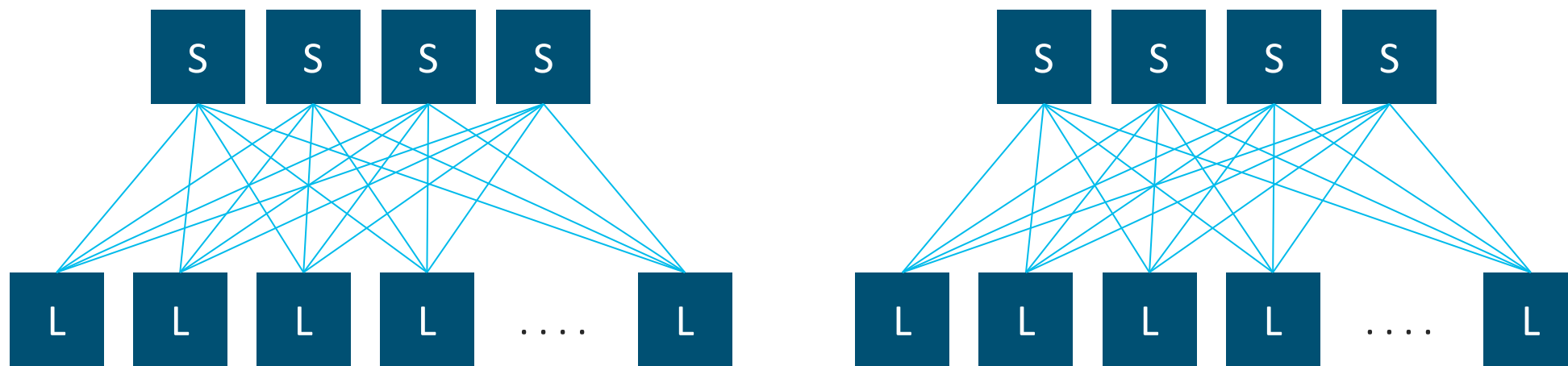
Leaf Spine топология



Один логический ЦОД

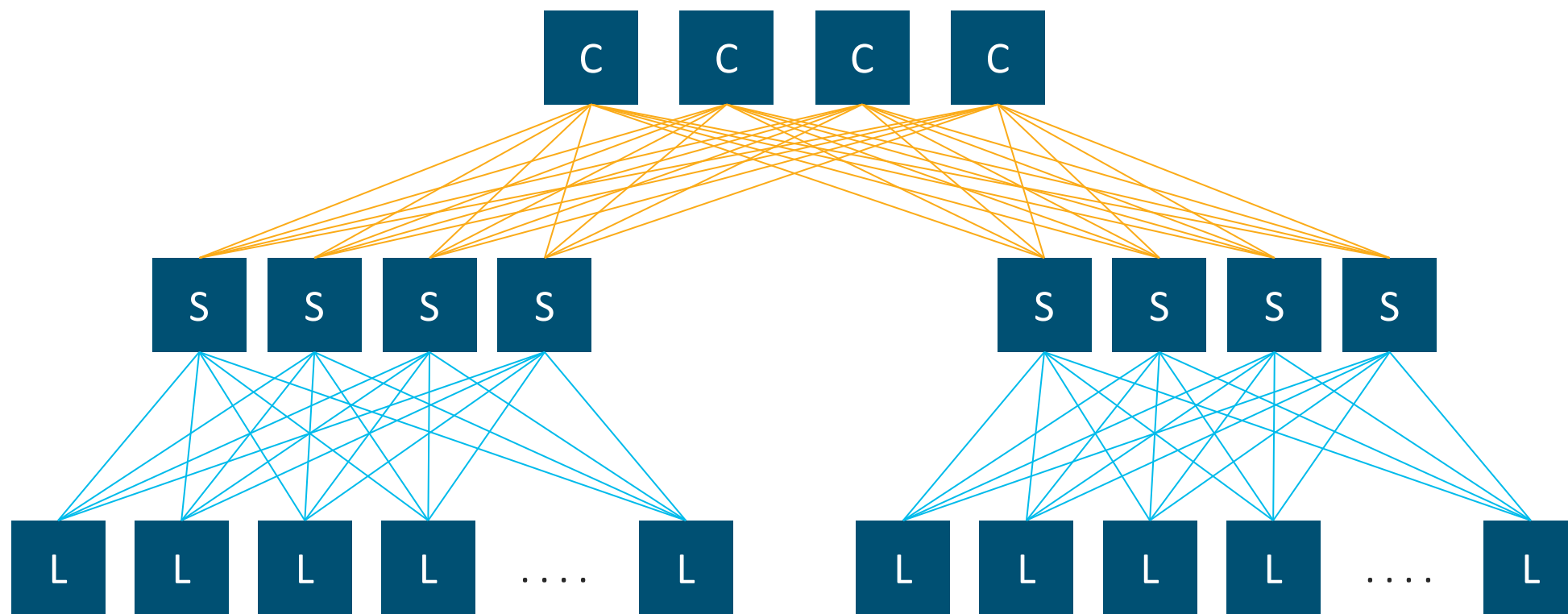
Как соединить?

- Leaf Spine топология



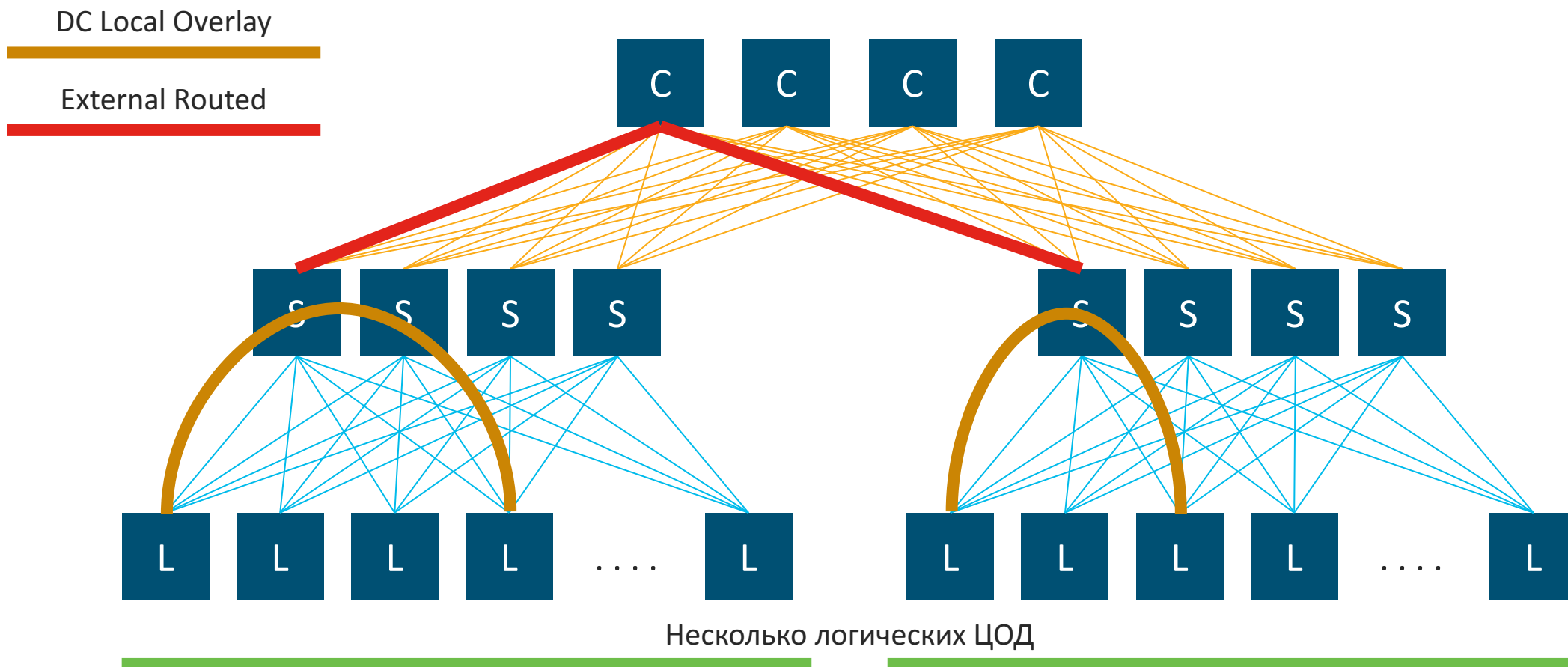
Несколько логических ЦОД

Использование DC Core

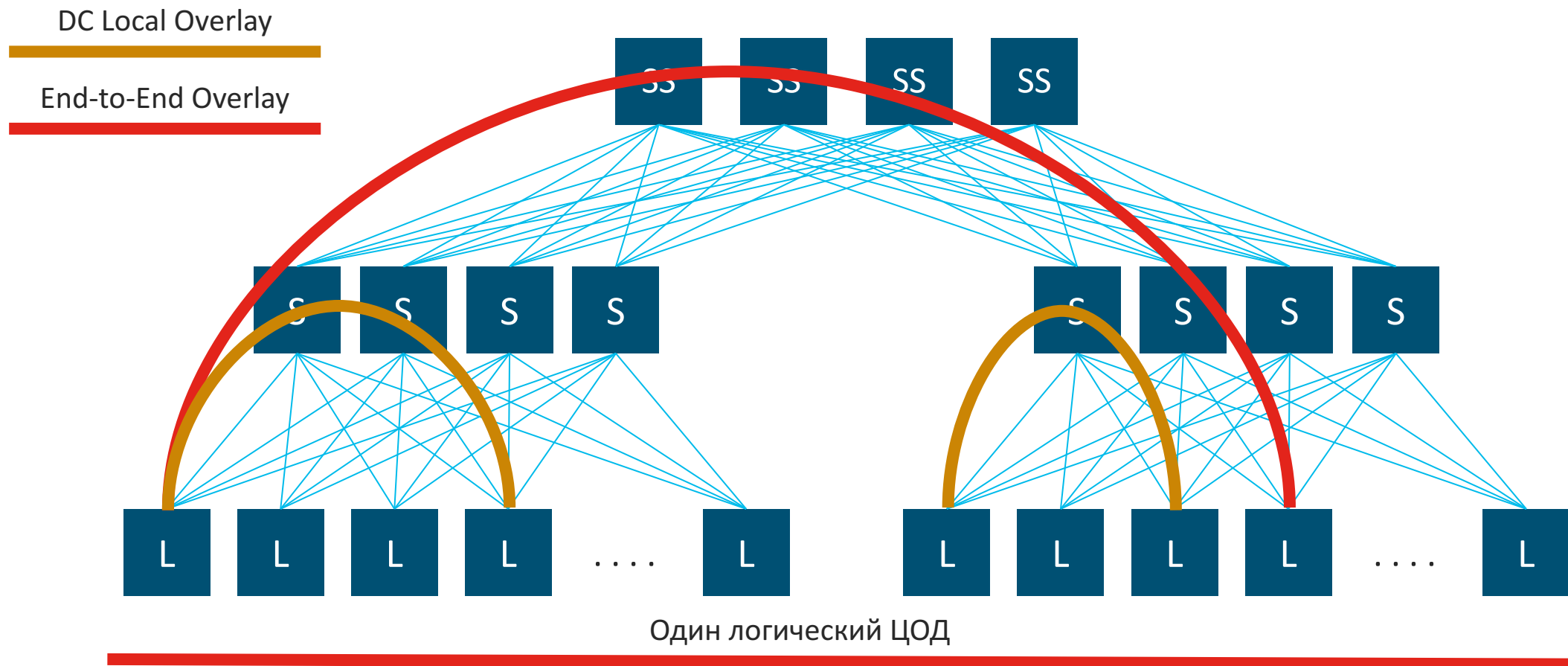


Несколько логических ЦОД

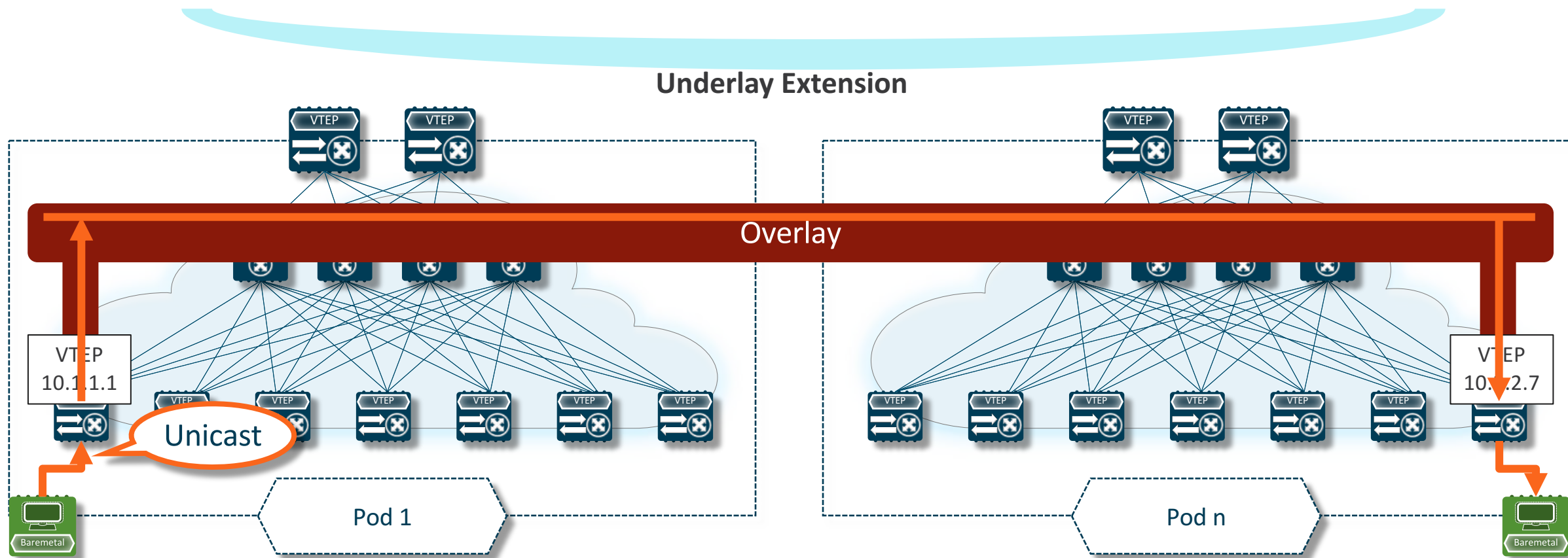
Взаимодействие между ЦОД: только Layer-3



Multi-Pod: растянутый оверлей



Multi-Pod сквозная передача VXLAN-инкапсулированного трафика



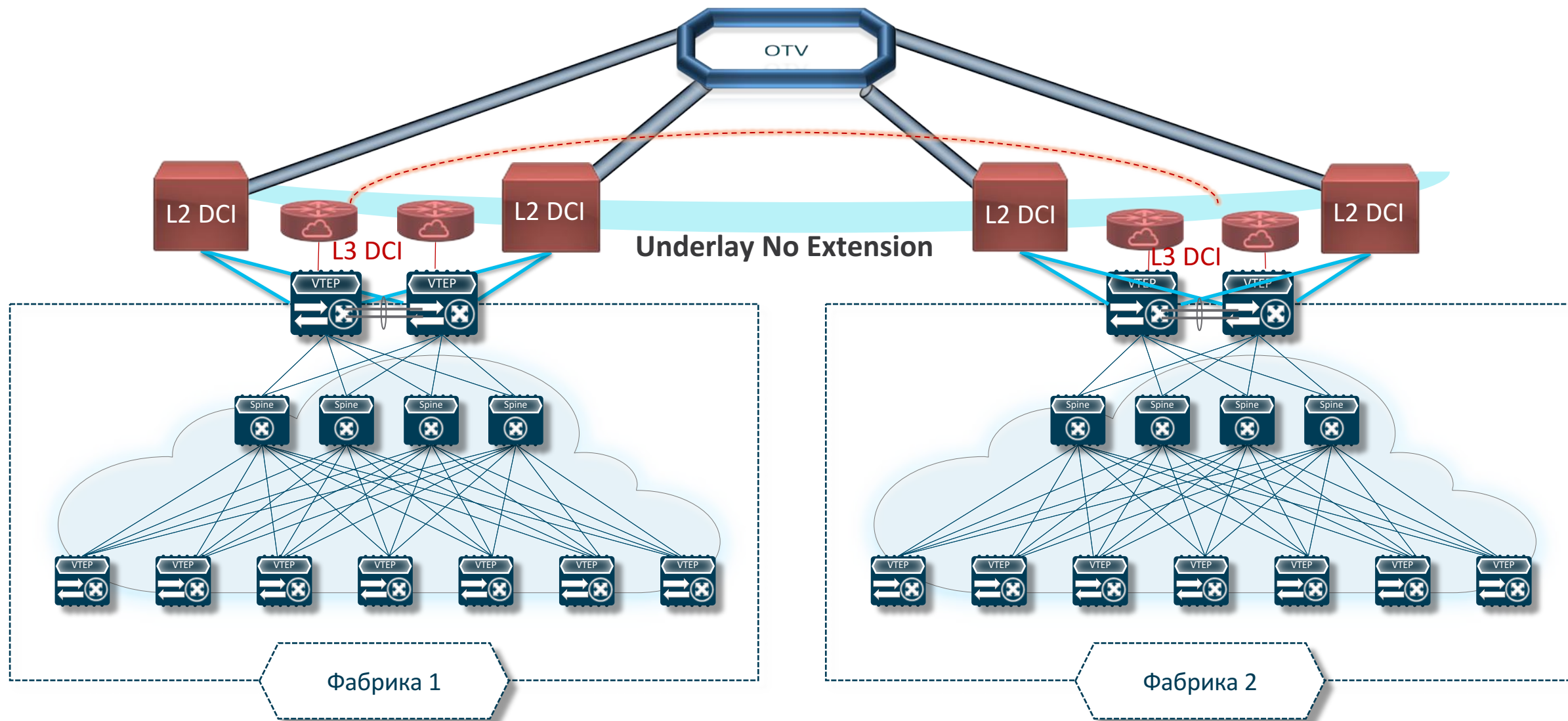


Свойства Multi-Pod – “Единый”

- **Единый** Overlay домен – End-to-End Encapsulation
- **Единый** Overlay Control-Plane домен – End-to-End EVPN апдейты
- **Единый** сквозной Underlay домен
- **Единый** домен репликации для BUM
- **Единый** VNI административный домен

Иерархия в оверлее отсутствует

VXLAN EVPN – Multi-Fabric



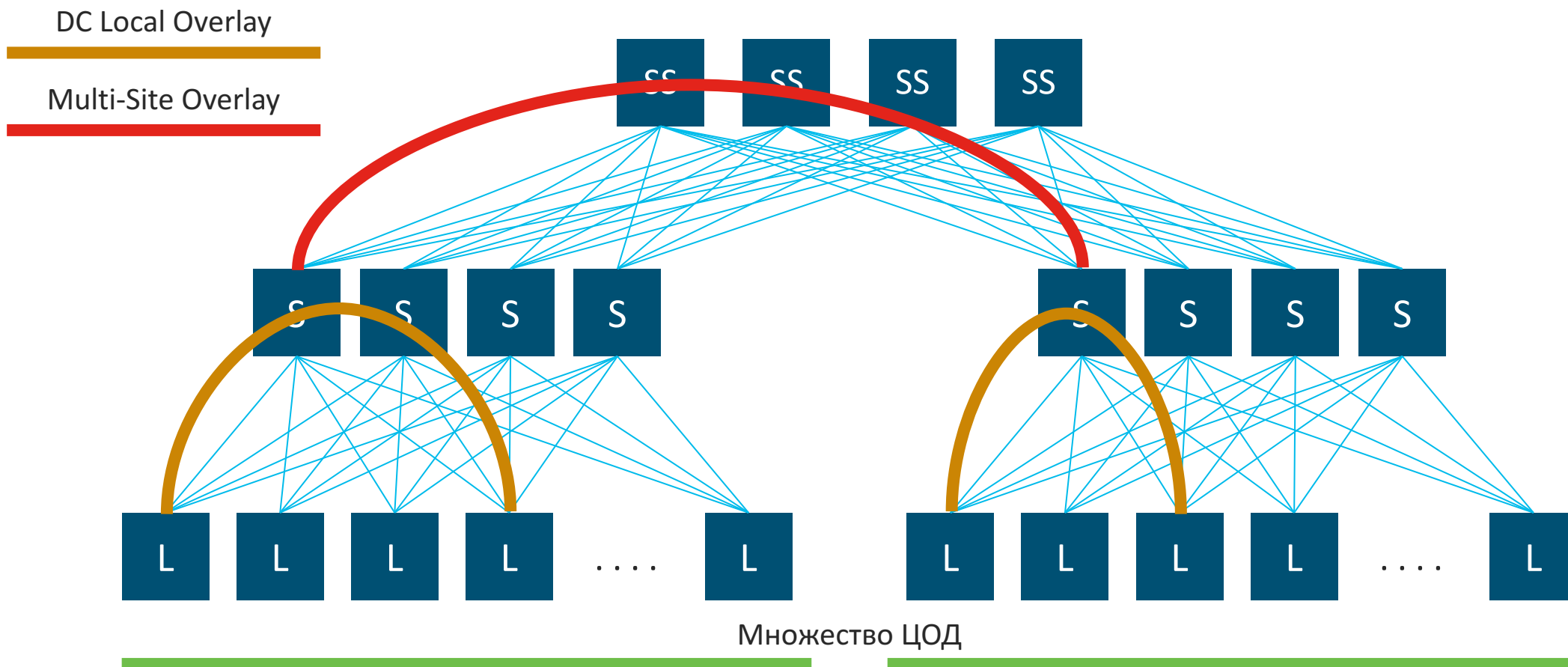


Свойства Multi-Fabric – “Раздельный”

- **Раздельные** Overlay домены – независимые L2 и L3 DCI (сложность)
- **Раздельные** Overlay Control-Plane домены – ручная настройка
- **Раздельные** Underlay домены - Isolated
- **Раздельные** домены репликации для BUM – Independent BUM transport/DCI
- **Выделенный Border Leaf** – нельзя подключать локальные End-Point

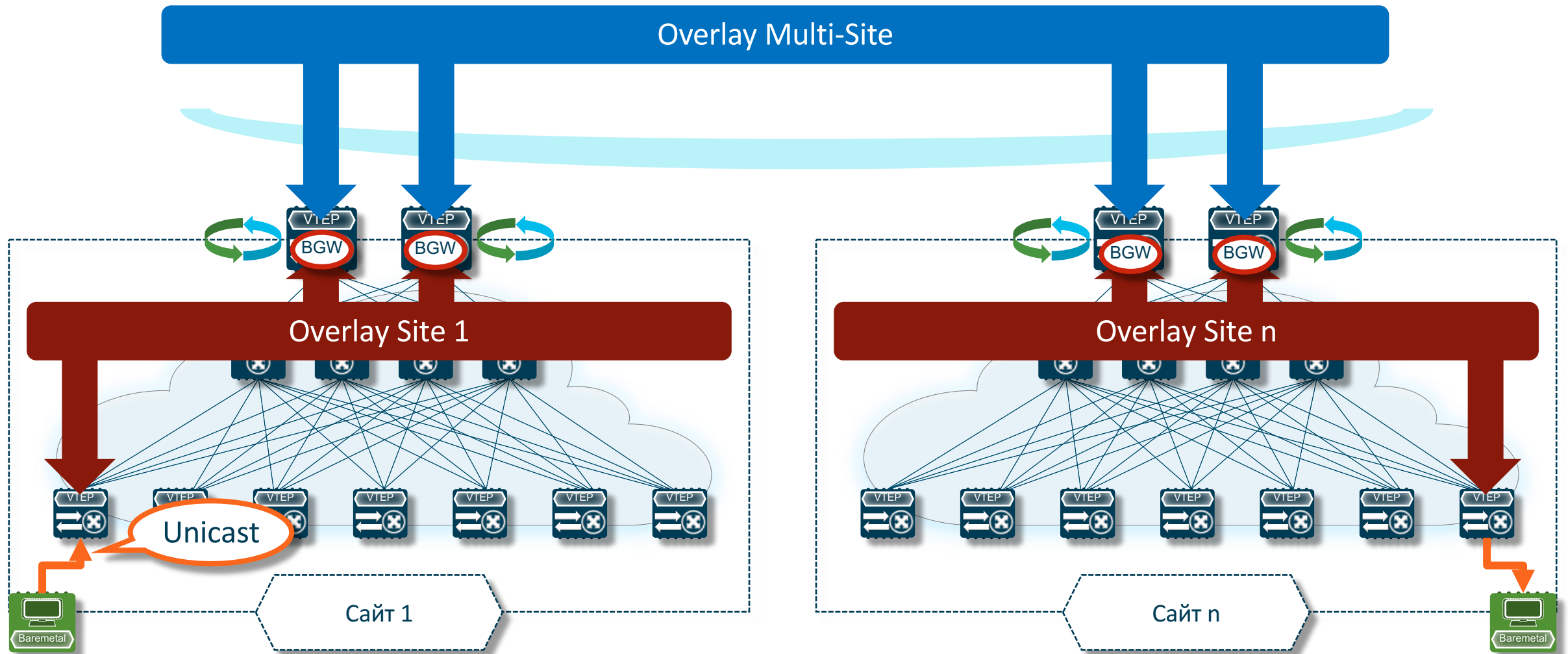
Underlay изоляция – выделенные технологии для DCI

Multi-Site: парадигма иерархических оверлеев



VXLAN Multi-Site

- Иерархия доменов с оверлеями





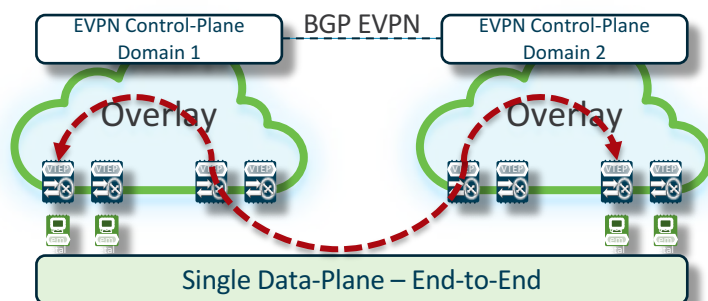
Свойства VXLAN Multi-Site

- **Множество** Overlay доменов – Interconnected & Controlled
- **Множество** Overlay Control-Plane доменов – Interconnected & Controlled
- **Множество** Underlay доменов - изолированные
- **Множество** доменов репликации BUM – Interconnected & Controlled
- **Множество** VNI административных доменов – Downstream VNI

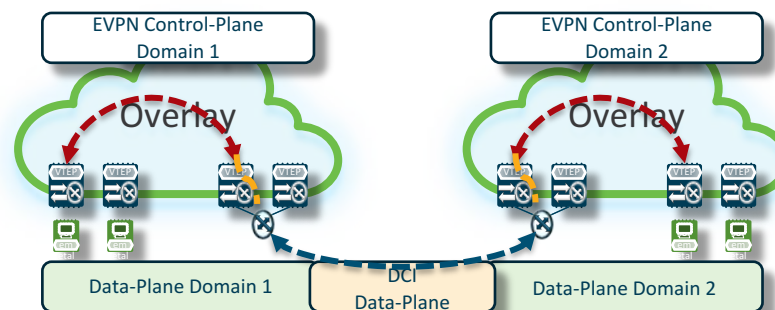
Underlay Isolation – Overlay Hierarchies

Развитие технологий для объединения ЦОД

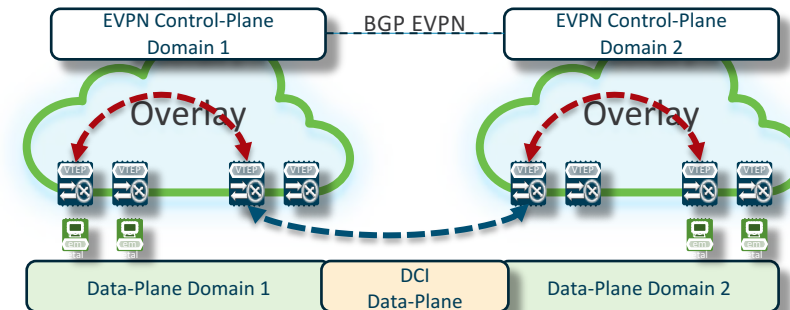
VXLAN Multi-Pod



VXLAN Multi-Fabric



VXLAN Multi-Site



- Единая фабрика со сквозной инкапсуляцией
- Иерархия на уровне Underlay – плоская архитектура внутри overlay

- Multiple Fabrics – нормализация через Ethernet-стык
- Multiple Fabrics объединяются путем DCI (Layer 2 и Layer 3)

- Многие фабрики с интегрированным DCI (DCI²)
- Integrated DCI – масштабирование внутри и между фабриками

	Multi-Pod	Multi-Fabric	Multi-Site
Underlay Control Plane	Unified Underlay Domain	Separated Underlay Domains	Separated Underlay Domains
Overlay Control Plane	Separated Overlay Control-Plane Domains		
Overlay Data Plane	Single Data-Plane	Separated Data-Planes	Separated Data-Planes
BUM Replication in DCI	Unified Underlay Domain (All Multicast or All Ingress Replication)	Dependency on DCI	Choice (Unicast/Multicast)
ARP Flood Suppression (DCI)	yes	yes	yes
Unknown Unicast Flood Suppression (DCI)	no	yes	yes
Broadcast Suppression/Limit (DCI)	no	yes	yes
Layer-2 Loop Prevention	Loop mitigation (Edge Protection)	VPC at Border	Loop mitigation (At DCI)

Основные свойства архитектуры

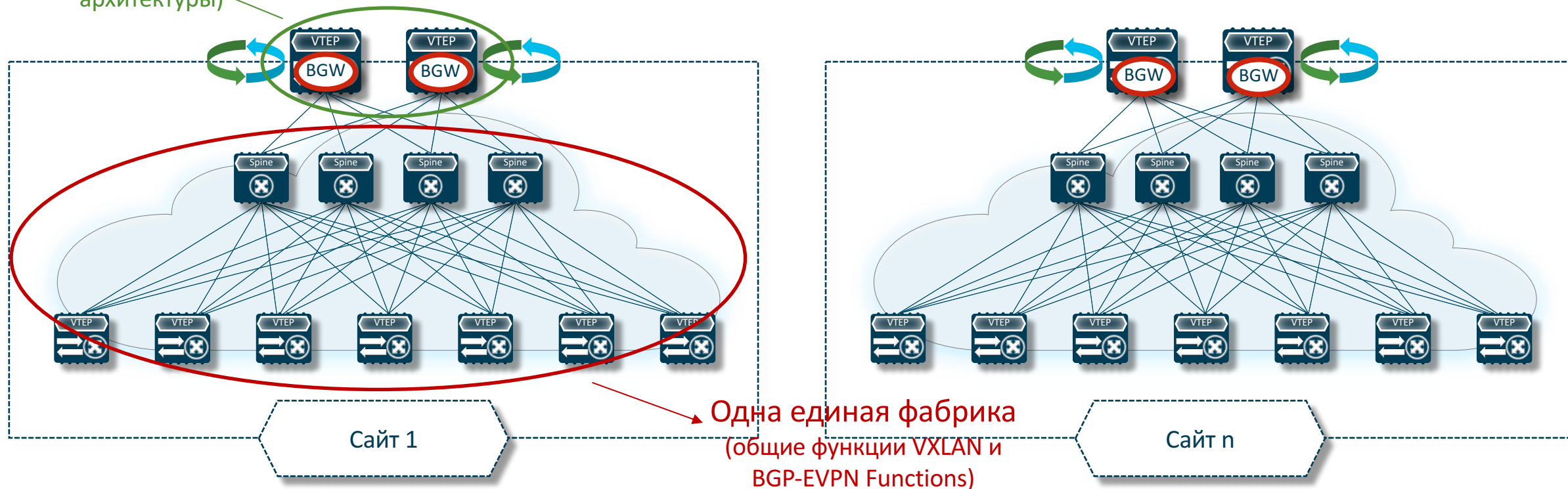
VXLAN Multi-Site

<https://tools.ietf.org/html/draft-sharma-multi-site-evpn>

- Функциональные компоненты

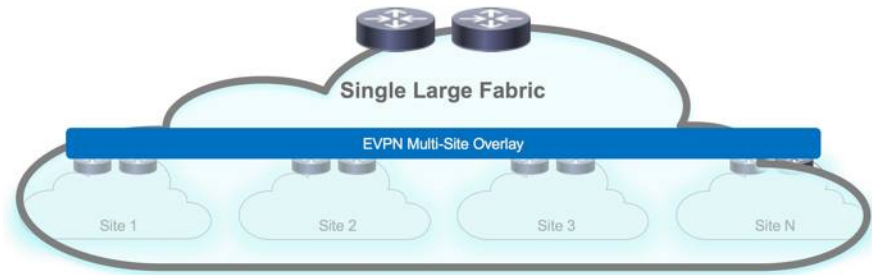
Border Gateways
(ключевой функциональный
компоненты VXLAN Multi-Site
архитектуры)

DCI между сайтами
(Поддержка IP маршрутизации
и увеличенного MTU)



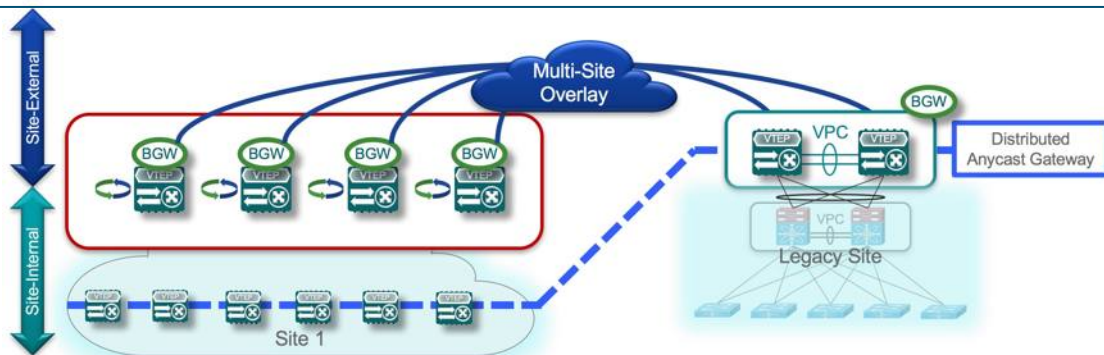
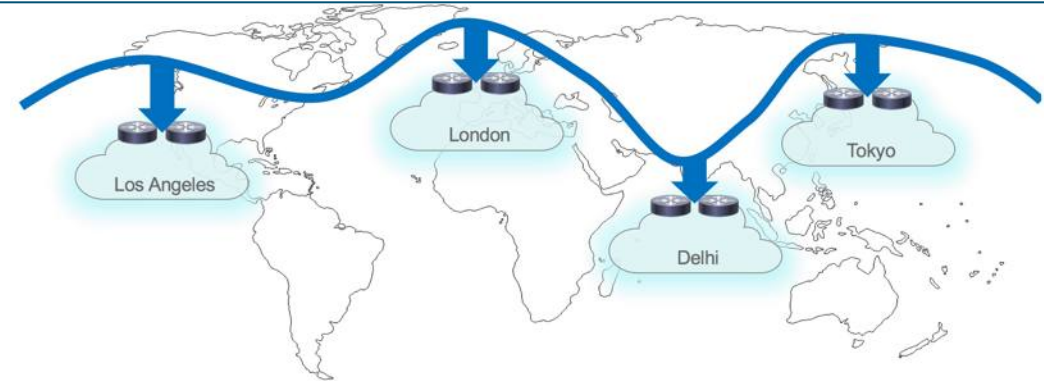
VXLAN Multi-Site

- Основные сценарии



Scale-Up модель для построения
гигантского ЦОД

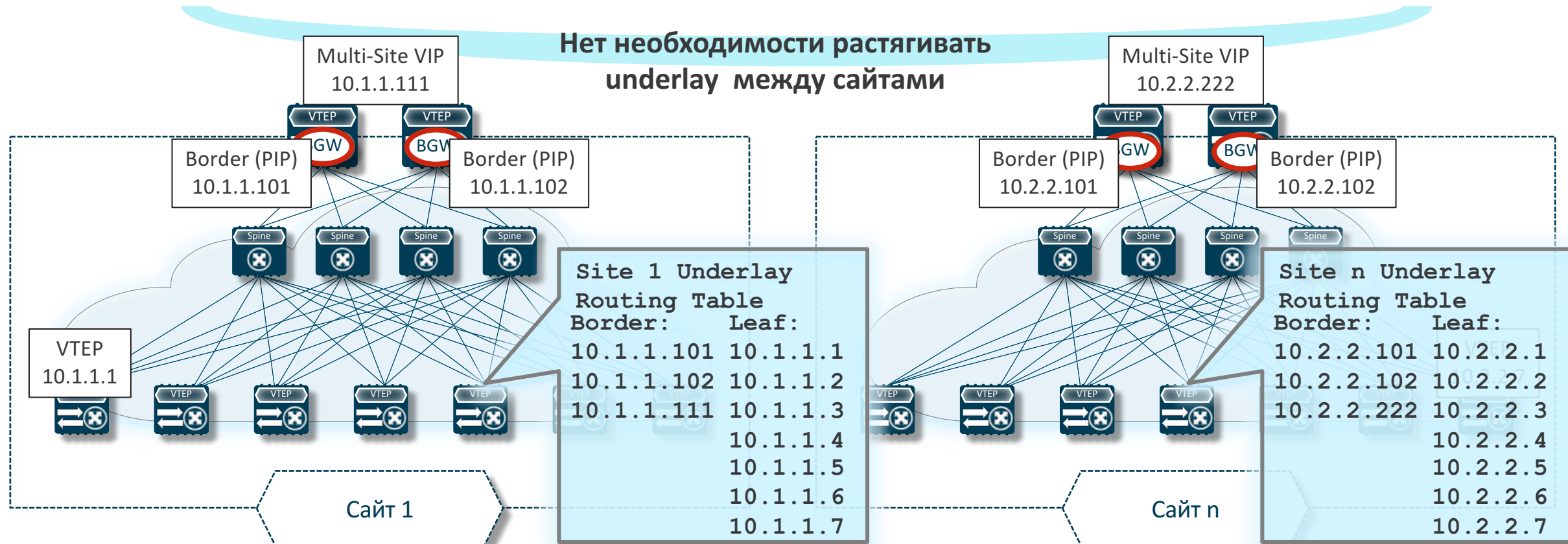
Data Center Interconnect (DCI)



Интеграция с существующим
оборудованием (со-существование
и миграция)

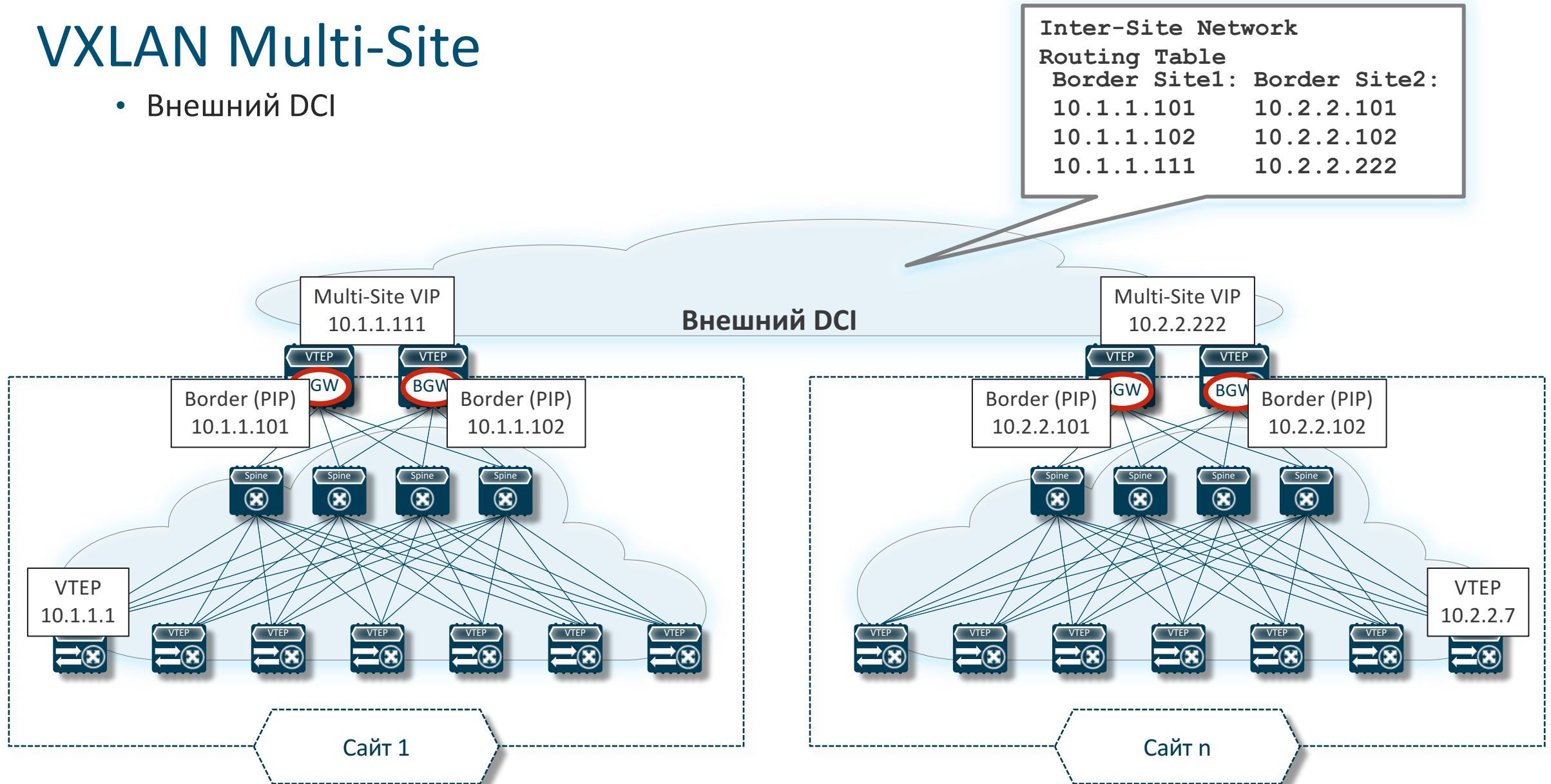
VXLAN Multi-Site

- Изоляция underlay



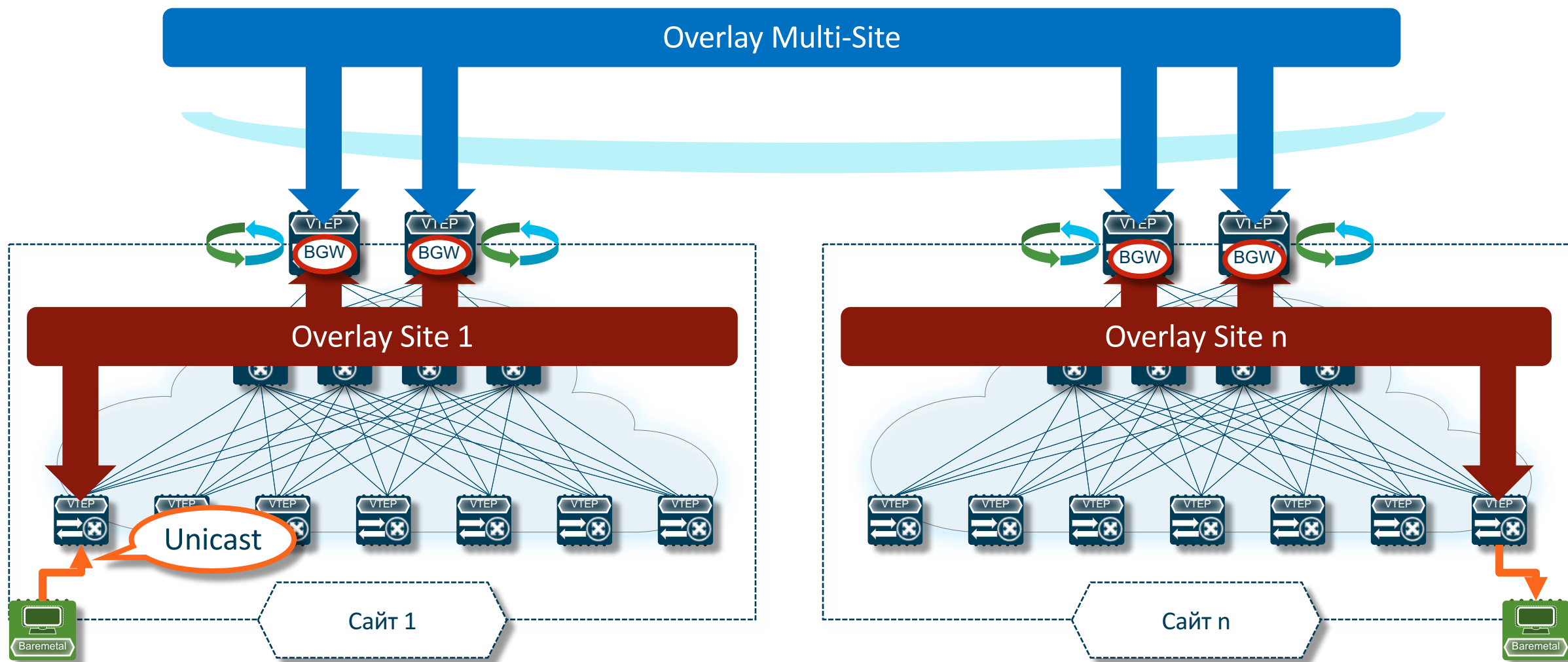
VXLAN Multi-Site

- Внешний DCI



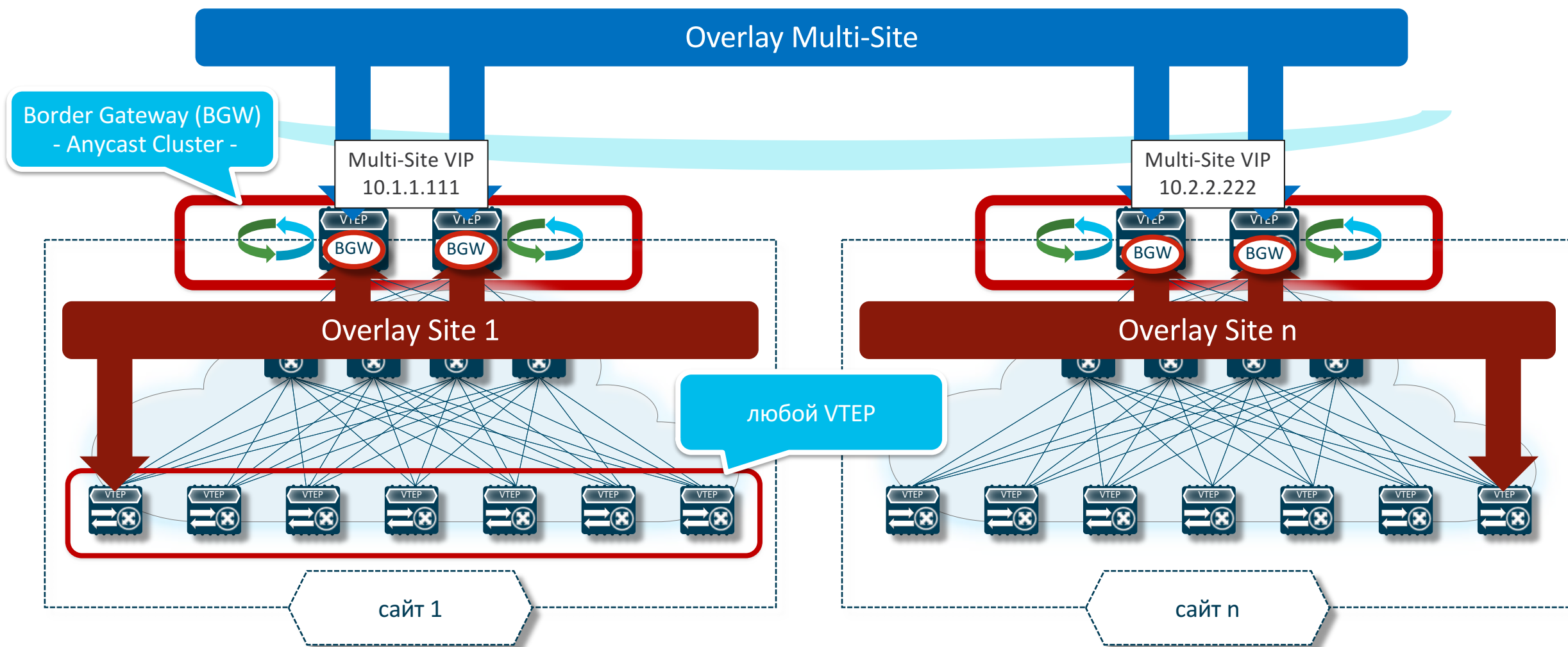
VXLAN Multi-Site

- Иерархия оверлейных доменов



VXLAN Multi-Site

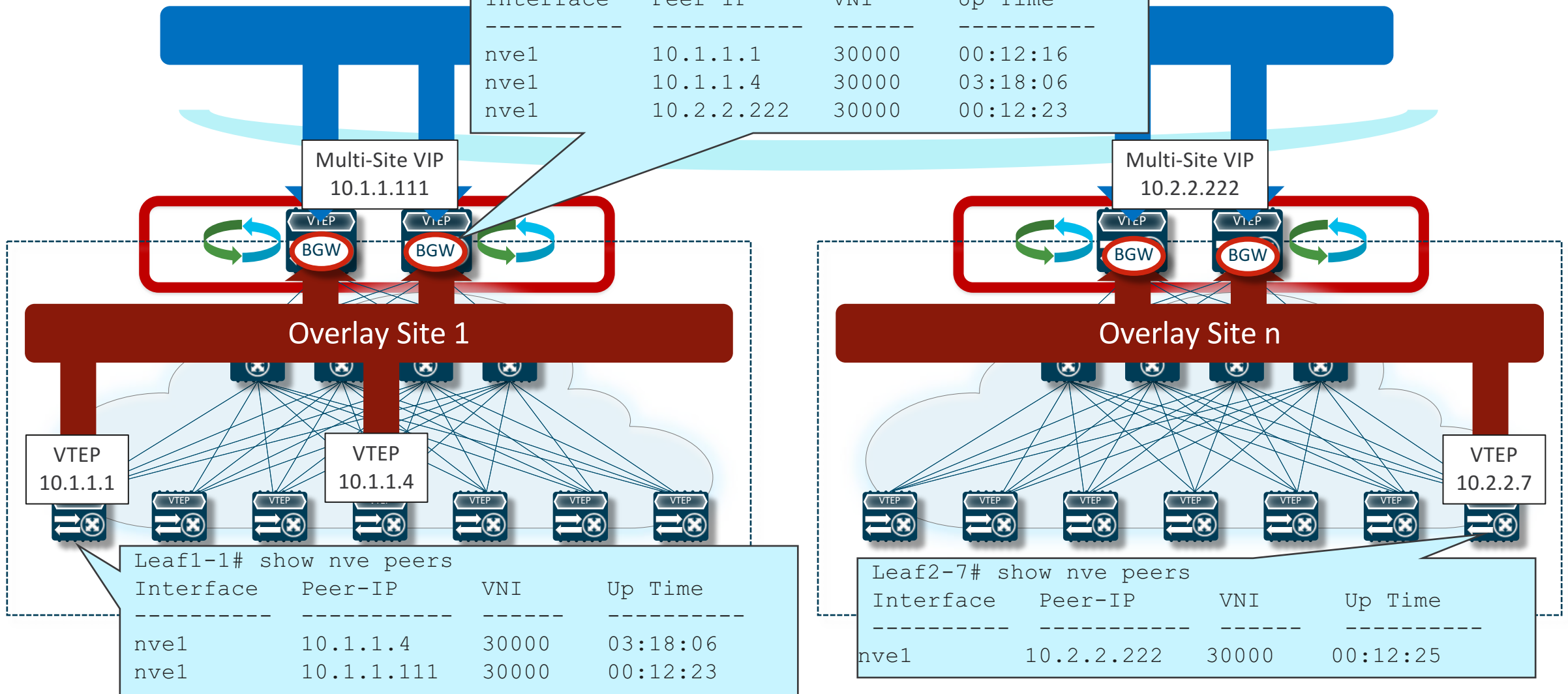
- Border Gateway – ключевой элемент архитектуры



Multi-Site – VXLAN Tunnel Adjacencies

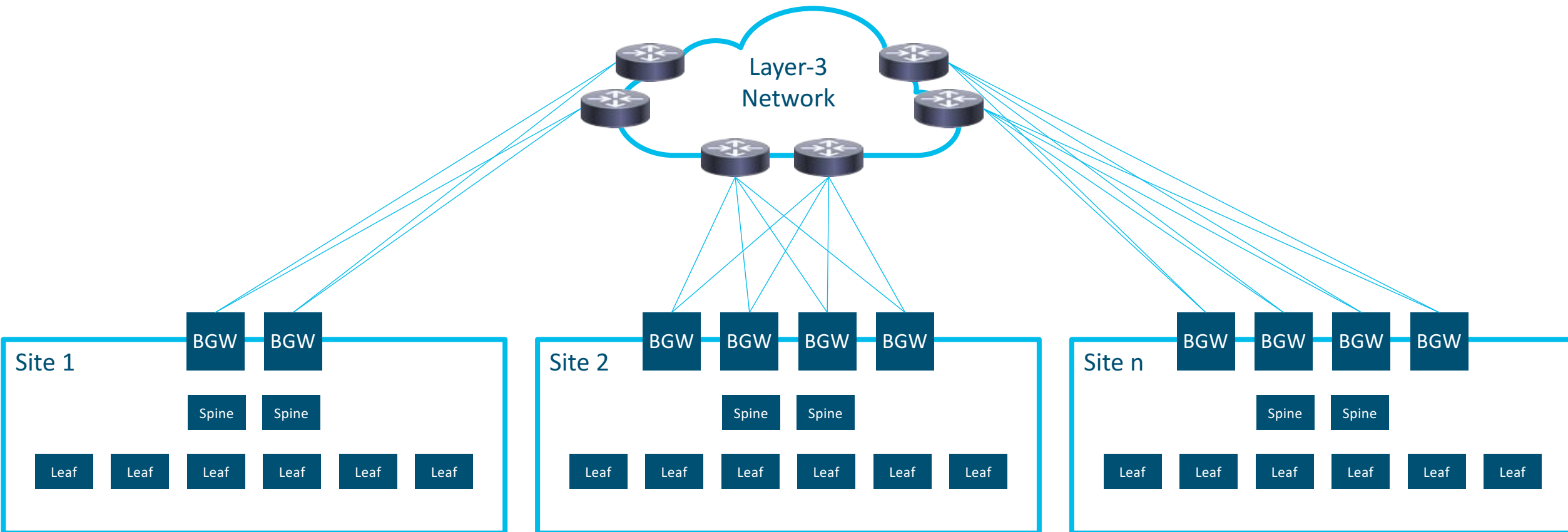
```
BG102# show nve peers
```

Interface	Peer-IP	VNI	Up Time
nve1	10.1.1.1	30000	00:12:16
nve1	10.1.1.4	30000	03:18:06
nve1	10.2.2.222	30000	00:12:23

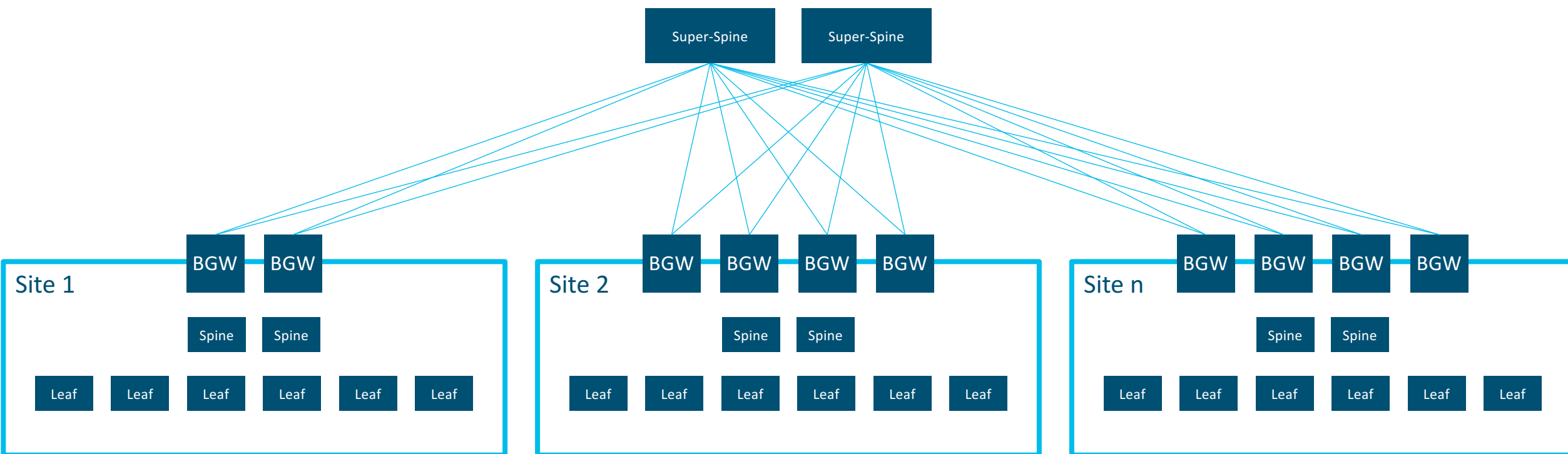


Поддерживаемые топологии

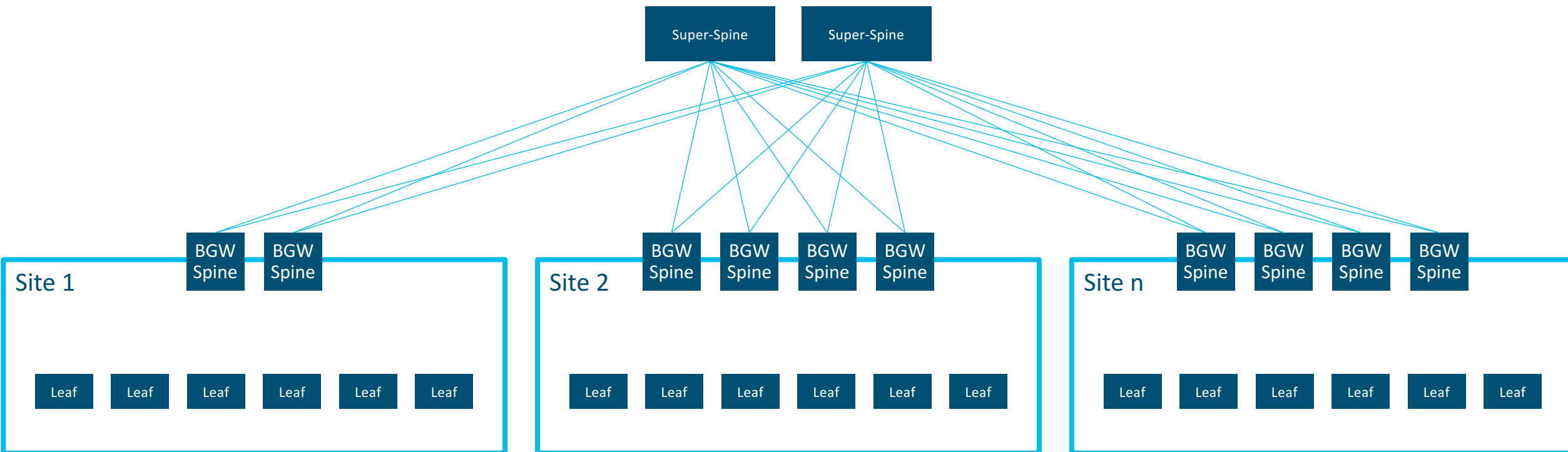
BGW-to-Cloud



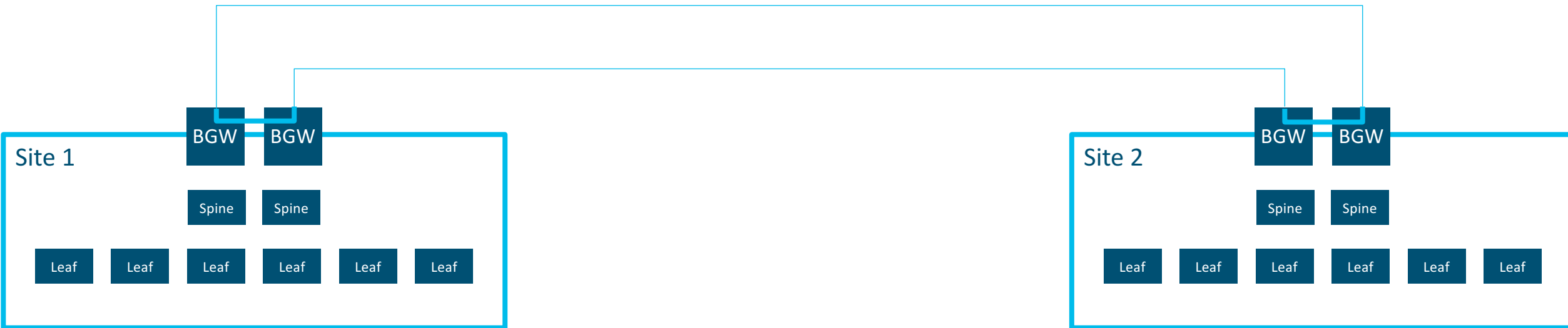
BGW между Spine и Super-Spine



BGW совмещенный со Spine



BGWs Back-to-Back

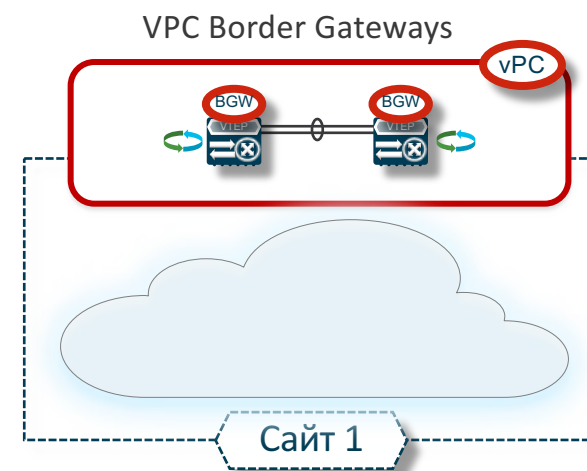
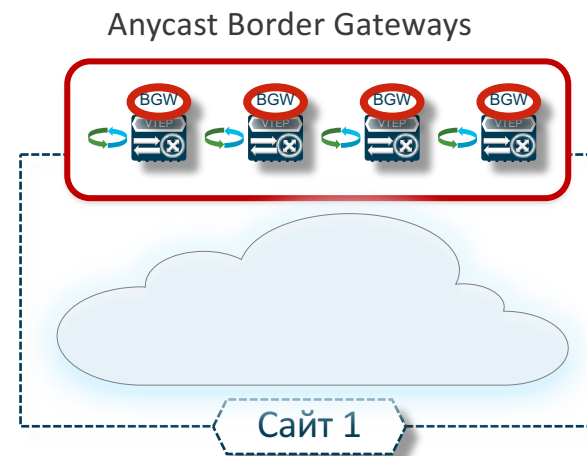


Режимы работы BGW

VXLAN Multi-Site

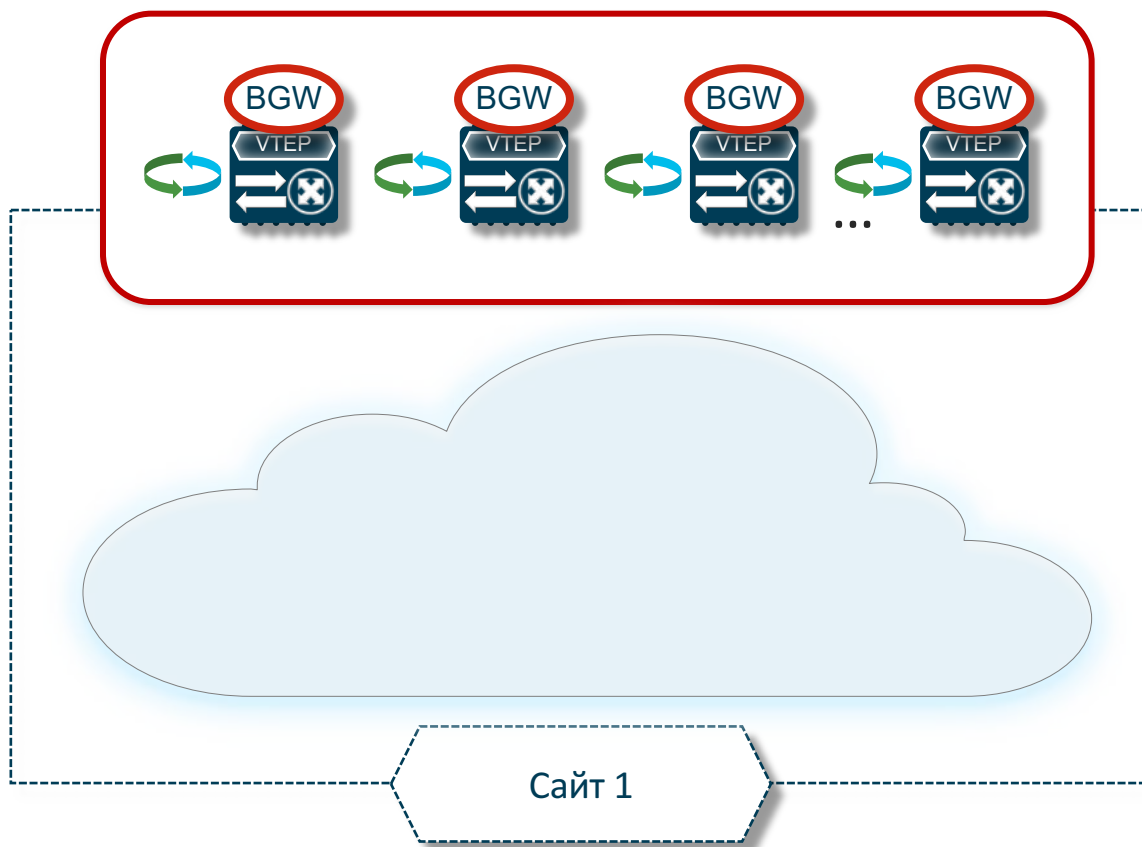
Соображения по развертыванию Border Gateways

- Border Gateway для двух основных функций:
 - Подключение каждого сайта к Inter-Site сети (для потоков трафика East-West)
 - Подключение каждого сайта ко внешнему Layer 3 домену (для потоков трафика North-South)
 - Может так же использоваться для подключения конечных хостов и/или сетевых сервисных устройств (FWs, ADCs)
- Возможные режимы развертывания:
 - Anycast Border Gateways
 - VPC Border Gateways
- BGW функции доступные внутри VXLAN EVPN :
 - BGW совмещенный с leaf
 - BGW совмещенный со spine (Border-Spines)



VXLAN Multi-Site

Anycast Border Gateway (1)

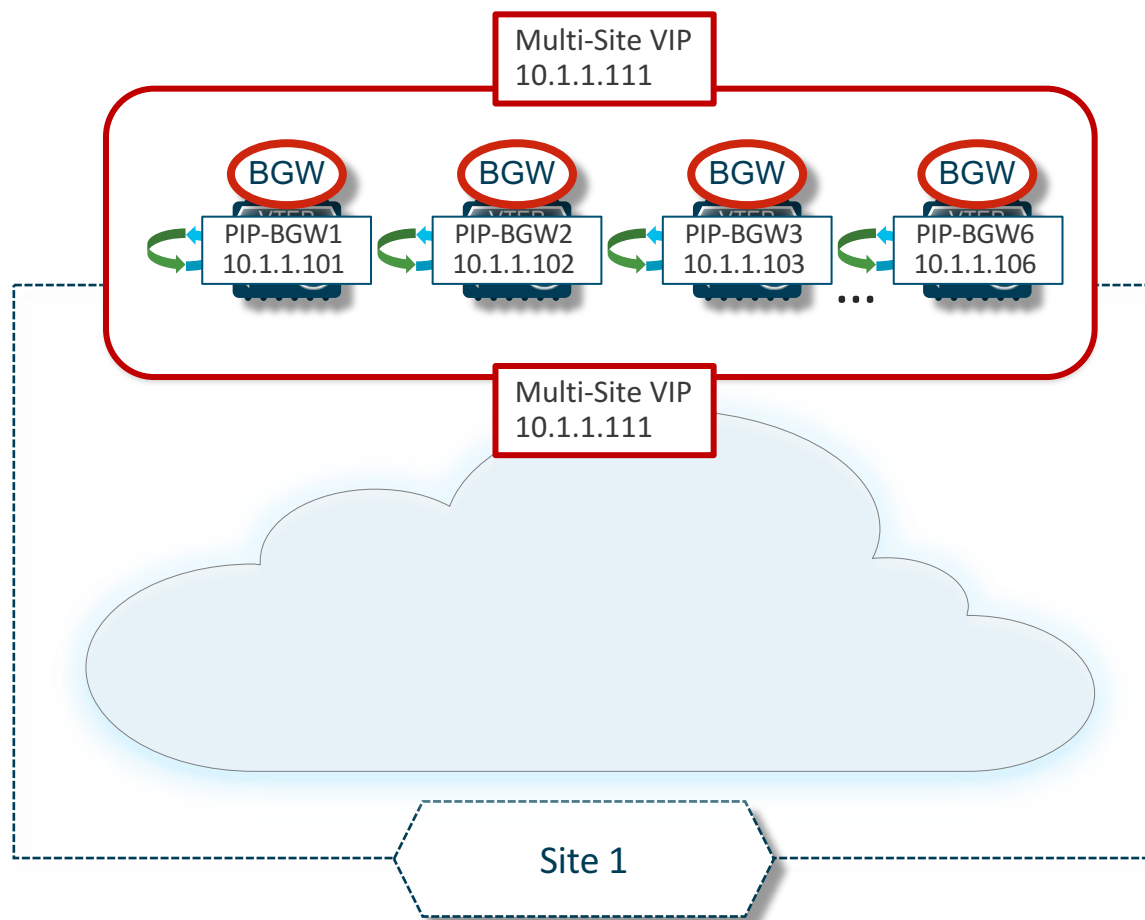


Anycast Border Gateway

- До 6 Border Gateways
- Border Gateway
 - Поддержка роли Leaf – с релиза 7.0(3)I7(1)
 - Поддержка роли Spine – с релиза 7.0(3)I7(2)

VXLAN Multi-Site

Anycast Border Gateway (2)

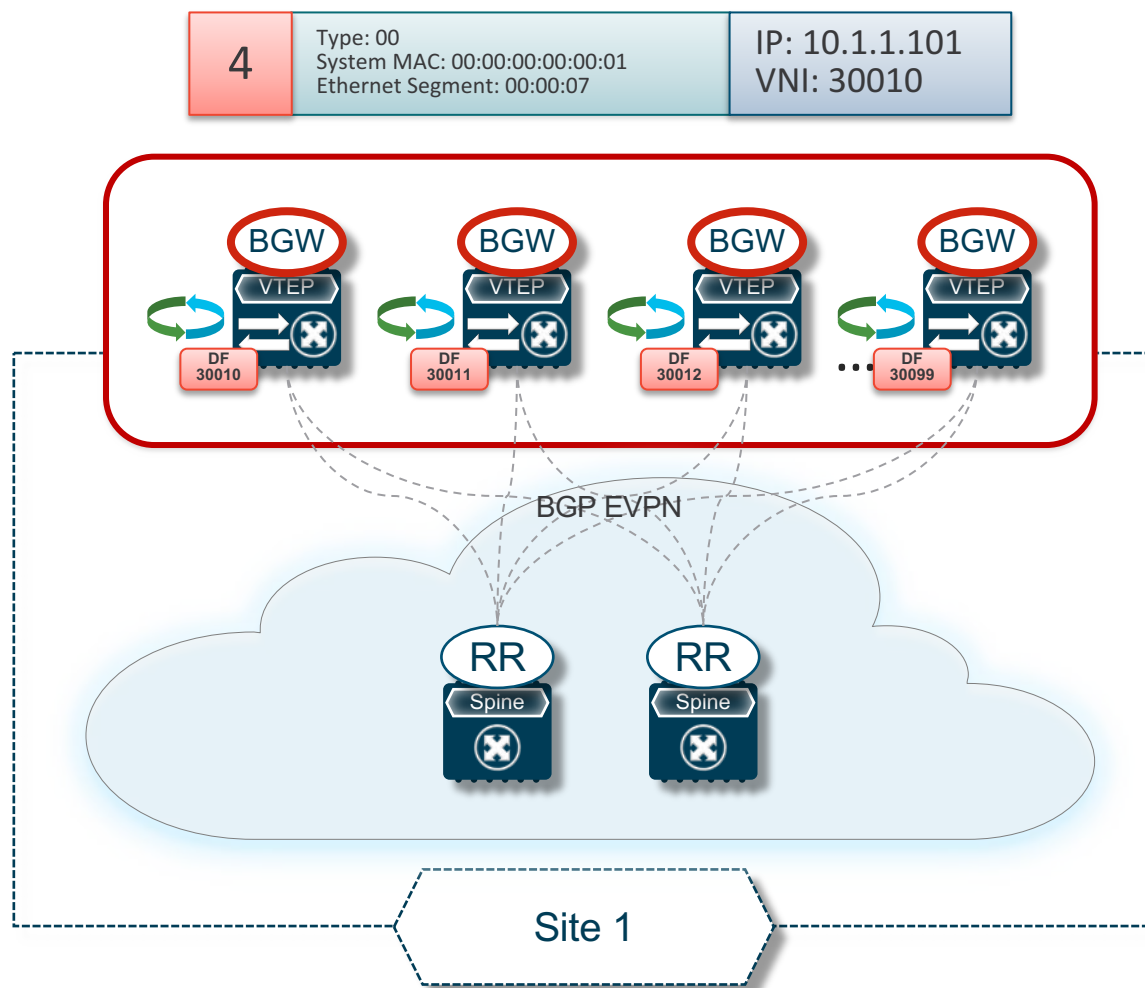


Anycast Border Gateway

- Общий Multi-Site Virtual IP (Multi-Site VIP) между BGW
 - Multi-Site VIP используется для взаимодействия между Border Gateways расположенных **на разных сайтах**
 - Multi-Site VIP используется для взаимодействий между Border Gateways и Leaf коммутаторами **внутри сайта**
- Individual Primary IP (PIP) на BGW
 - Используется для репликации Broadcast, Unknown Unicast and Multicast (BUM)
 - PIP используется для подключение Single-Homed хостов (routed only) внутри и между сайтами

VXLAN Multi-Site

Anycast Border Gateway (3)

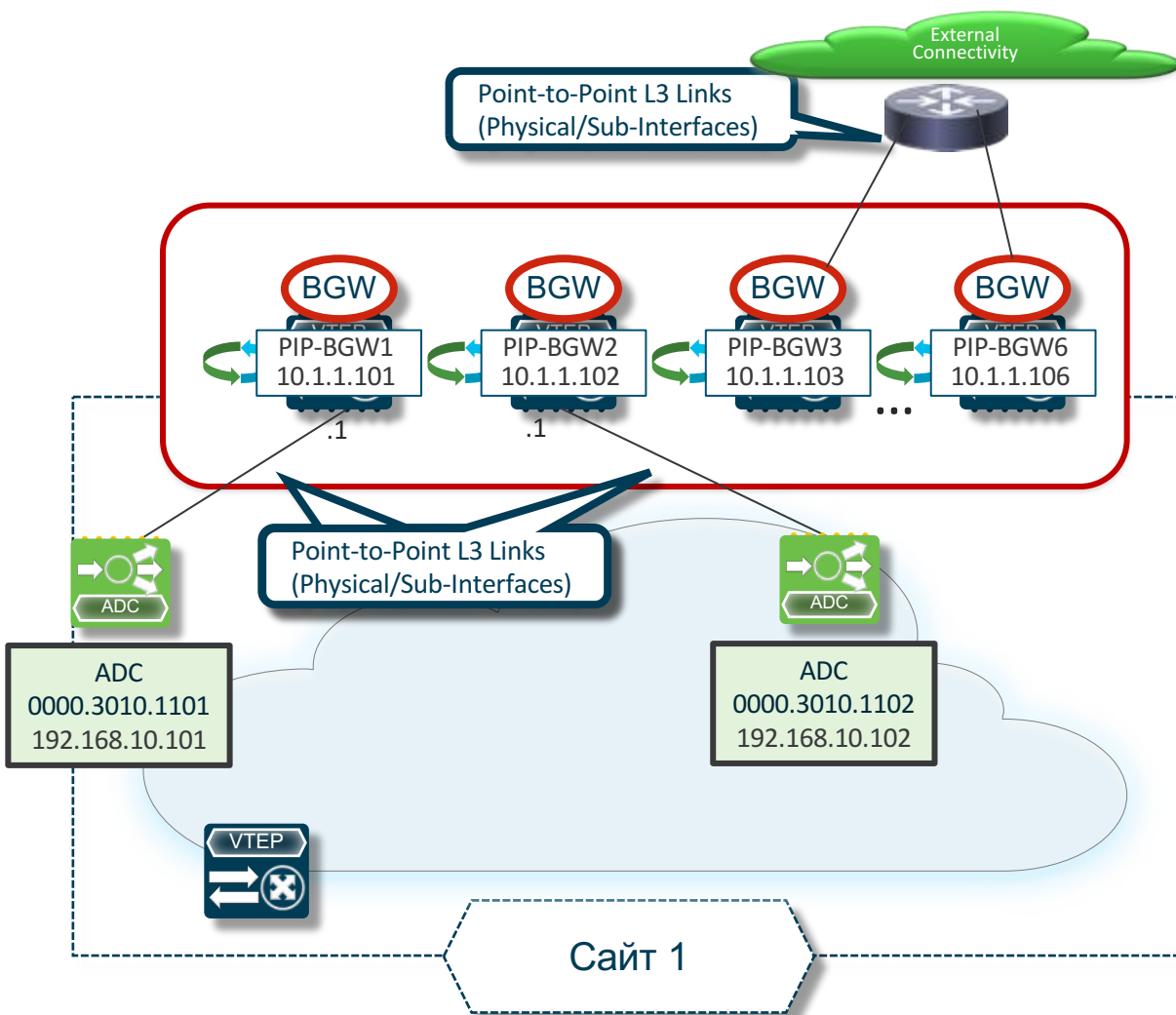


Anycast Border Gateway

- Per-VNI Designated Forwarder (DF) election
 - Каждый BGW может выступать как DF для одного или нескольких Layer-2 VNI
 - DF election автоматический процесс
- BGP EVPN Route Type 4 используются для DF election
 - Operator Managed Assignment (Type: 00)
 - Six Octet Site Identifier (System MAC: 00:00:00:00:00:01)
 - Multi-Site Discriminator (Ethernet-Segment: 00:00:07)
 - Originators IP Address (PIP): 10.1.1.101
 - Layer-2 VNI: 30010

VXLAN Multi-Site

Anycast Border Gateway (4)



Anycast Border Gateway

- Single-Homed подключаются только при помощи L3 линков
 - Сервисные устройства (например Firewall, ADC)
 - Внешние маршрутизаторы
 - SVI на BGW узлах не поддерживается
- Анонсирование и достижимость при помощи Individual Primary IP Address (PIP)
 - Внутри сайта: Leaf коммутаторы используют PIP для того чтобы достигать до таких устройств, подключенных к Border Gateway
 - Между сайтами: Удаленные Border Gateways используют PIP чтобы достигать до таких устройств, подключенных к Border Gateways

Anycast BGW и VPC Border Gateway

Anycast Border Gateway

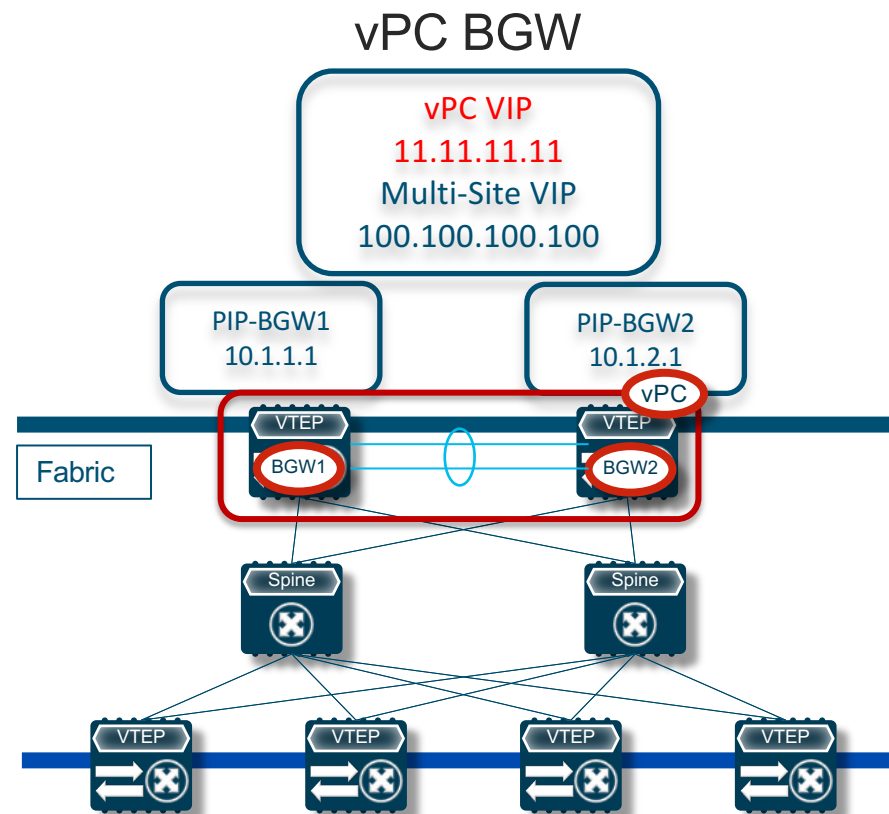
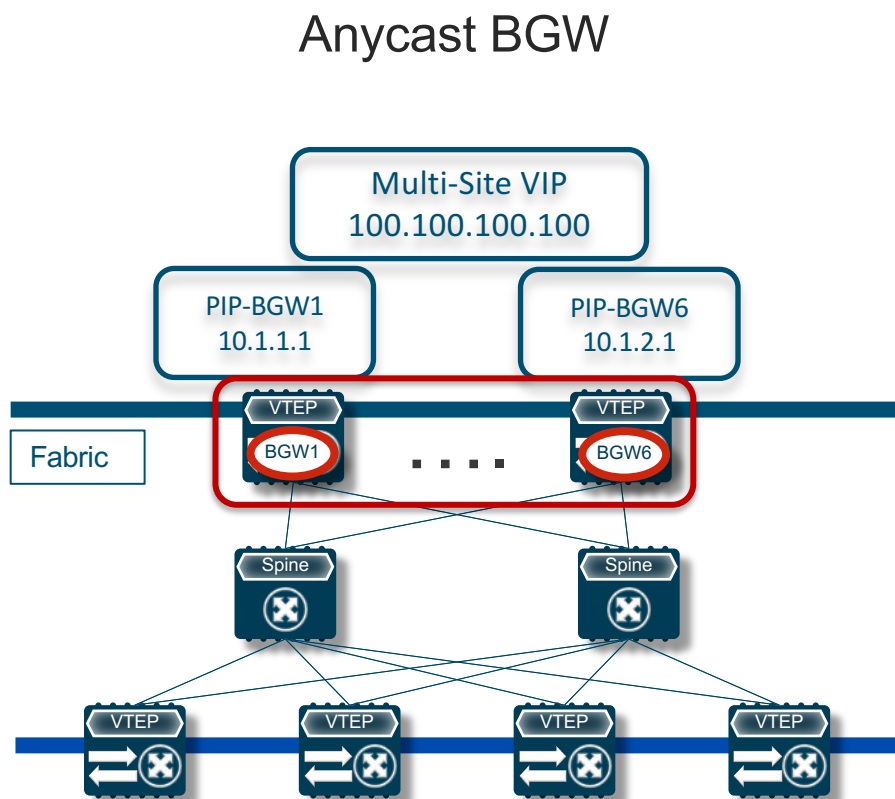
- До 6 BGW
 - Shared Nothing
 - Простые сценарии сбоя
- Любой сценарий внедрения
 - End-Point или Network Services не подключаются к BGW
- Greenfield Deployments

VPC Border Gateway

- 2 BGW с физическим VPC Peer-Link
- Небольшие внедрения
 - End-Point или Network Services подключаются к BGW
- Сценарии миграции (Brownfield)
 - Pseudo-BGW в BGW
 - Классический Ethernet/FabricPath в VXLAN EVPN

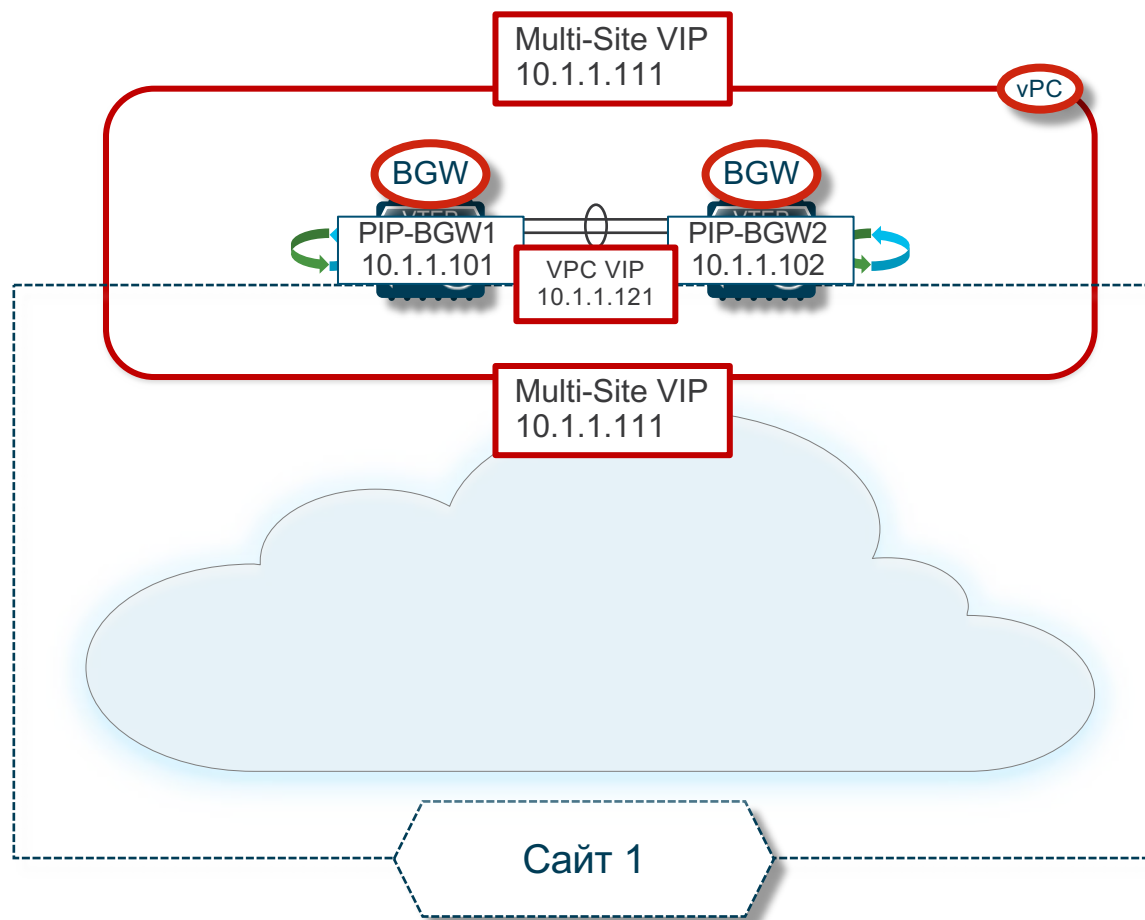
Multi-Site Border Gateway – Anycast vs. vPC

- Оба и Anycast и vPC Border Gateway требуют настройки и Multi-Site VIP адреса и individual Primary IP (PIP) адреса
- vPC Border Gateways требует настройки еще одного secondary IP адреса, который используется как vPC virtual IP (vPC VIP)



VXLAN Multi-Site

VPC Border Gateway и транзитный трафик

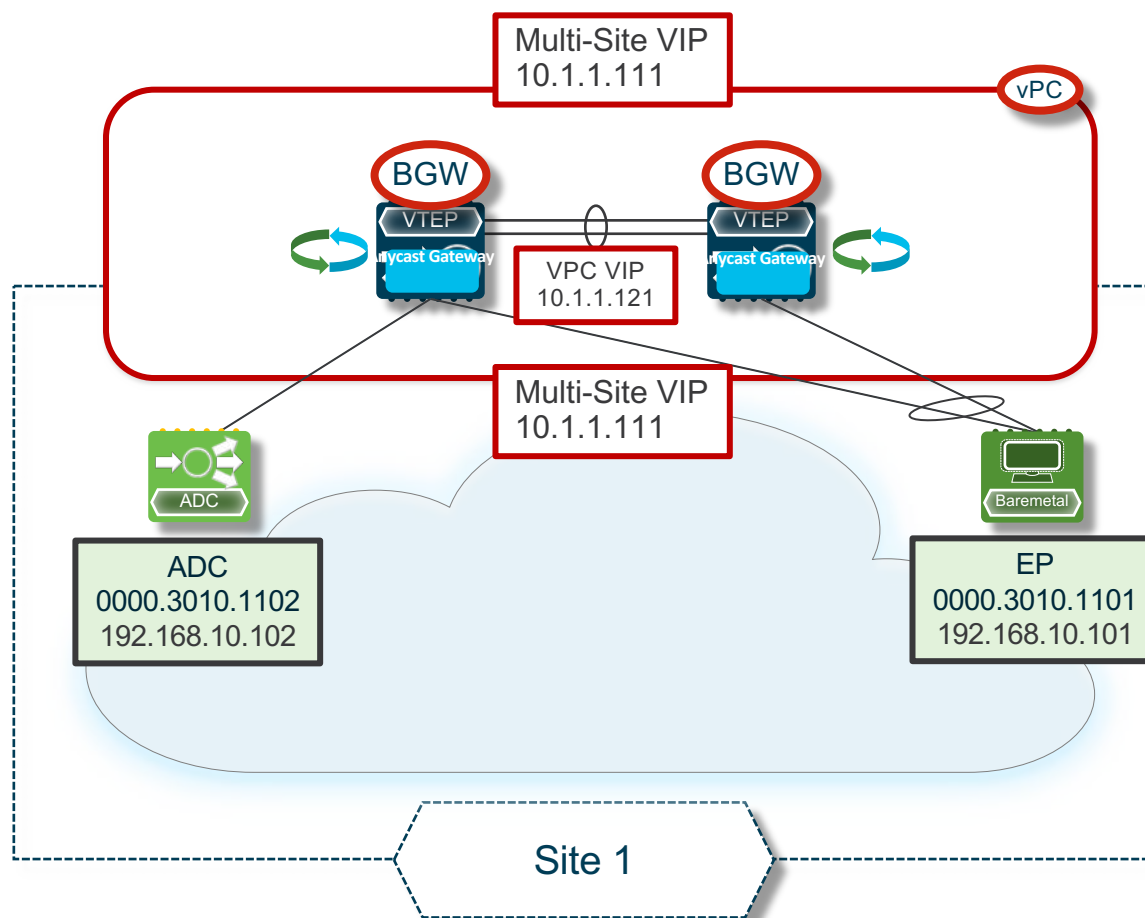


VPC Border Gateway

- Общий адрес Multi-Site Virtual IP (Multi-Site VIP) для BGW
 - Multi-Site VIP используется для транзитного трафика между сайтами (transit)
- Общий VPC Virtual IP (VPC VIP) между BGW
 - Используется по умолчанию для взаимодействия со внешними сетями (external networks)
 - Используется для репликации Broadcast, Unknown Unicast and Multicast (BUM)
- Individual Primary IP (PIP) на BGW
 - Используется для взаимодействия со внешними сетями (external networks) когда настроена функция “advertised-pip”

VXLAN Multi-Site

VPC Border Gateway и локально подключенные хосты

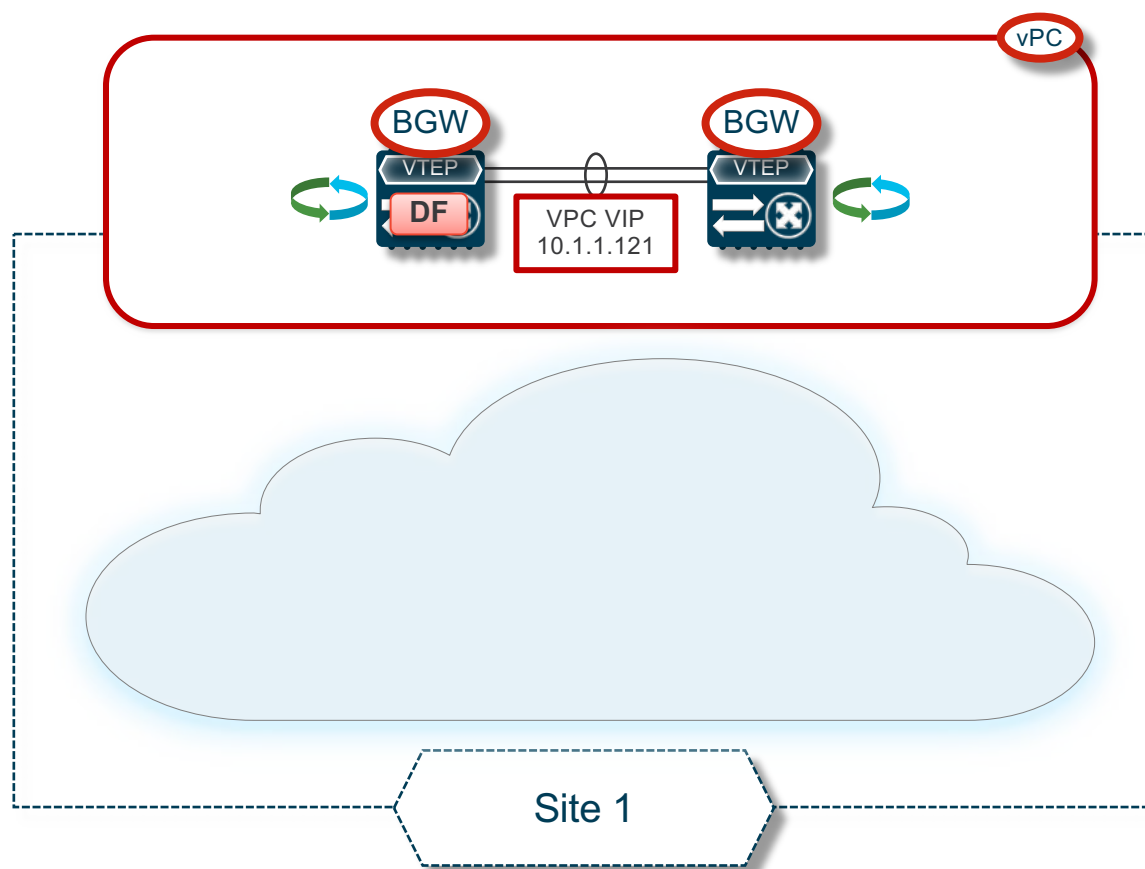


VPC Border Gateway

- Single- или Dual-Homed конечные хосты
 - Сервисные устройства (i.e. Firewall, ADC etc.)
 - Физические или виртуализированные серверы
 - Anycast Gateway функция работает для конечных хостов
- Анонсирование и достижимость при помощи VPC Virtual IP Address (VPC VIP)
 - Внутри сайта: Leaf узлы используют VPC VIP для взаимодействия с хостами, которые подключены к Border Gateway
 - Между сайтами: Удаленные Border Gateways используют VPC VIP для доступа к хостам, которые подключены к Border Gateway
 - Трафик потенциально может передаваться по VPC Peer-Link

VXLAN Multi-Site

VPC Border Gateway и Designated BUM Forwarder

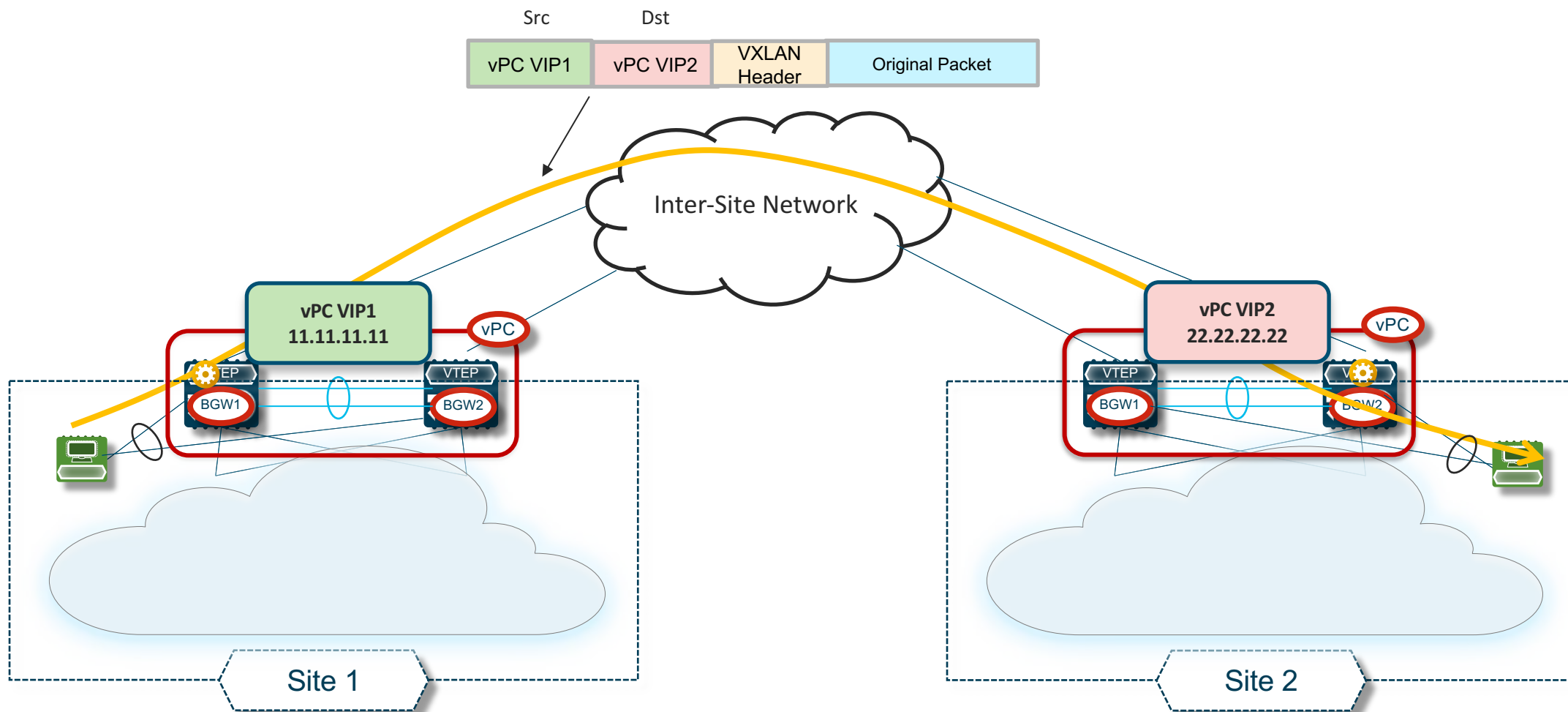


VPC Border Gateway

- VPC-based Designated Forwarder Election
- Per-Site Designated Forwarder (DF) election
 - Такой же подход как и в VPC
 - Best Path to Rendezvous-Point или VPC Primary Node
 - Один и тот же VPC узел выбирается как DF для **всех** растягиваемых Layer-2 VNI

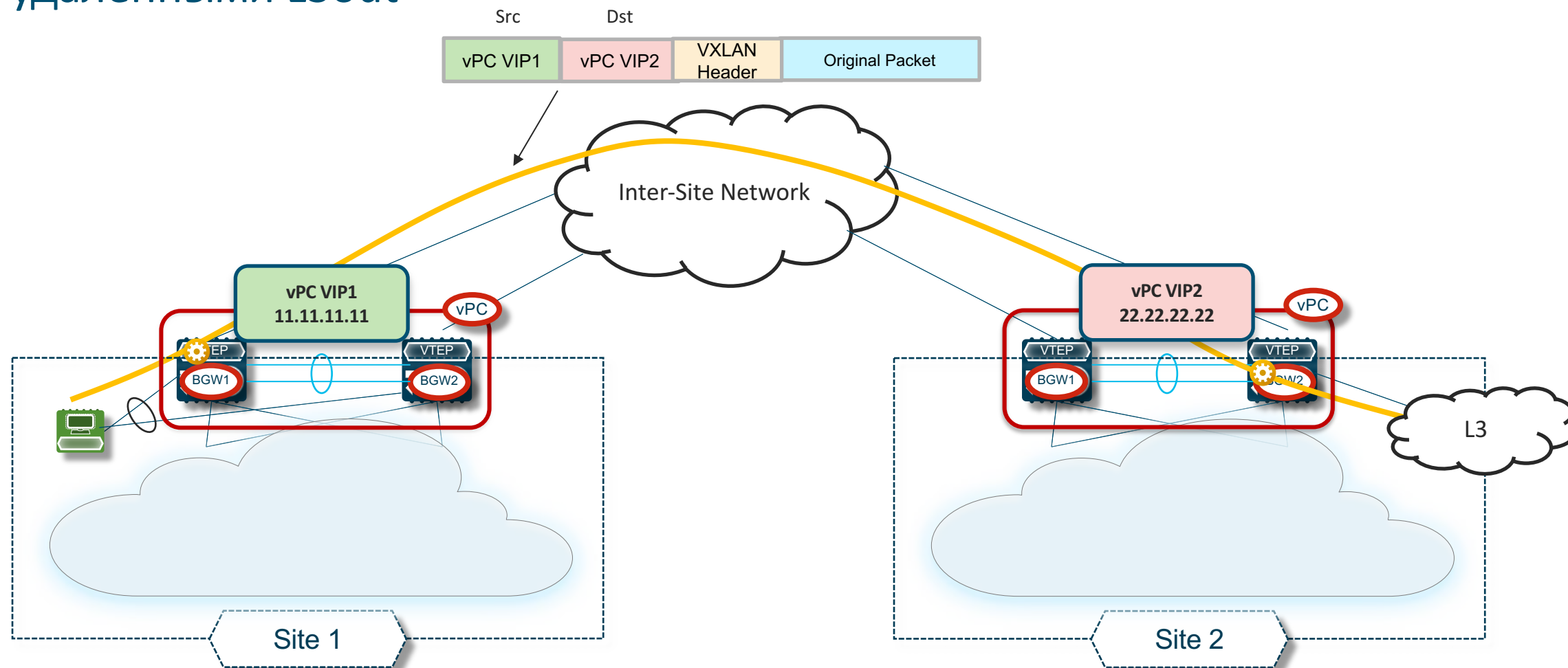
VPC Border Gateways

Передача данных между системами, подключенными напрямую к vPC



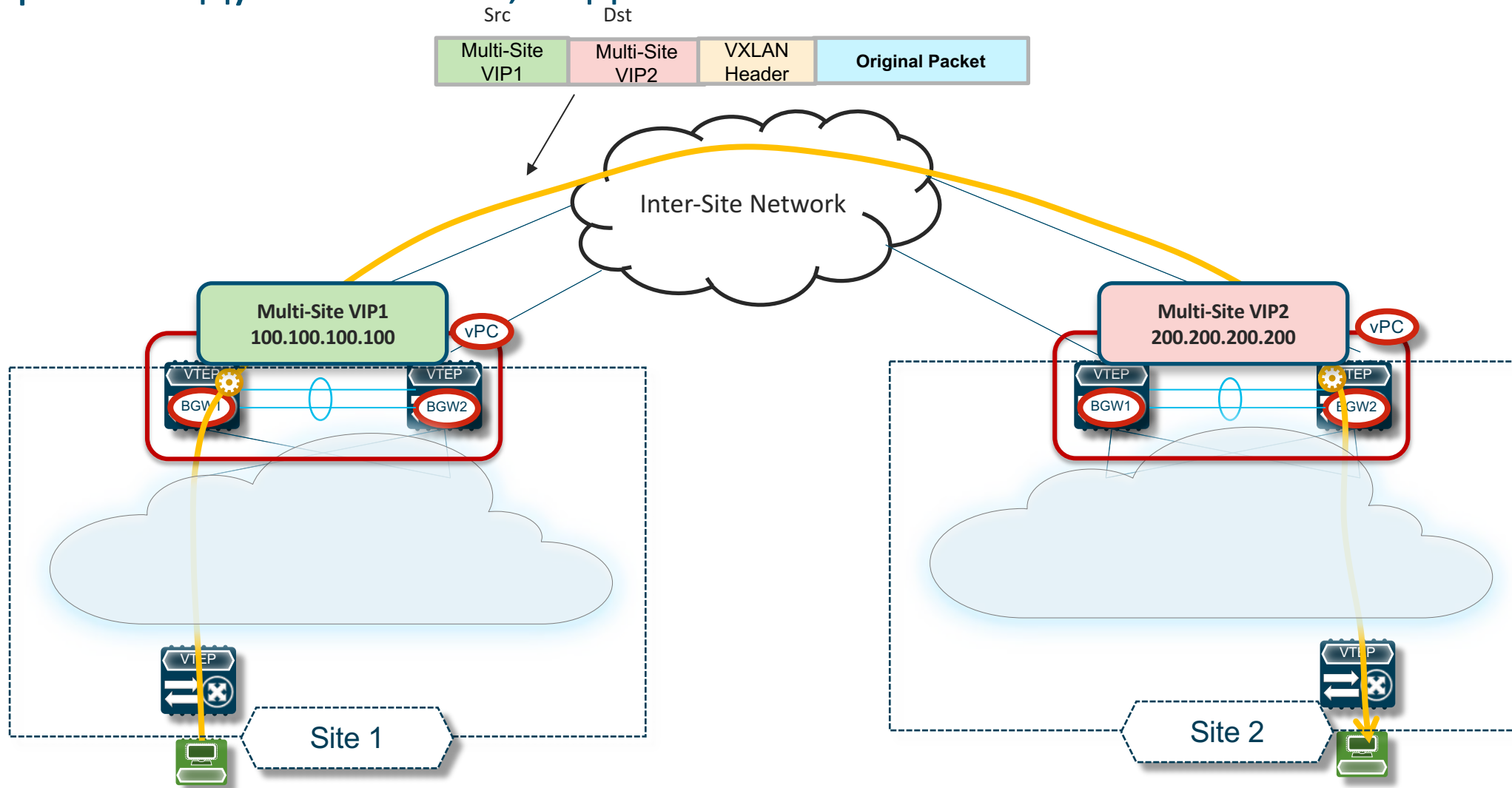
VPC Border Gateways

Передача данных между системой, подключенной напрямую к vPC и удаленными L3out



VPC Border Gateways

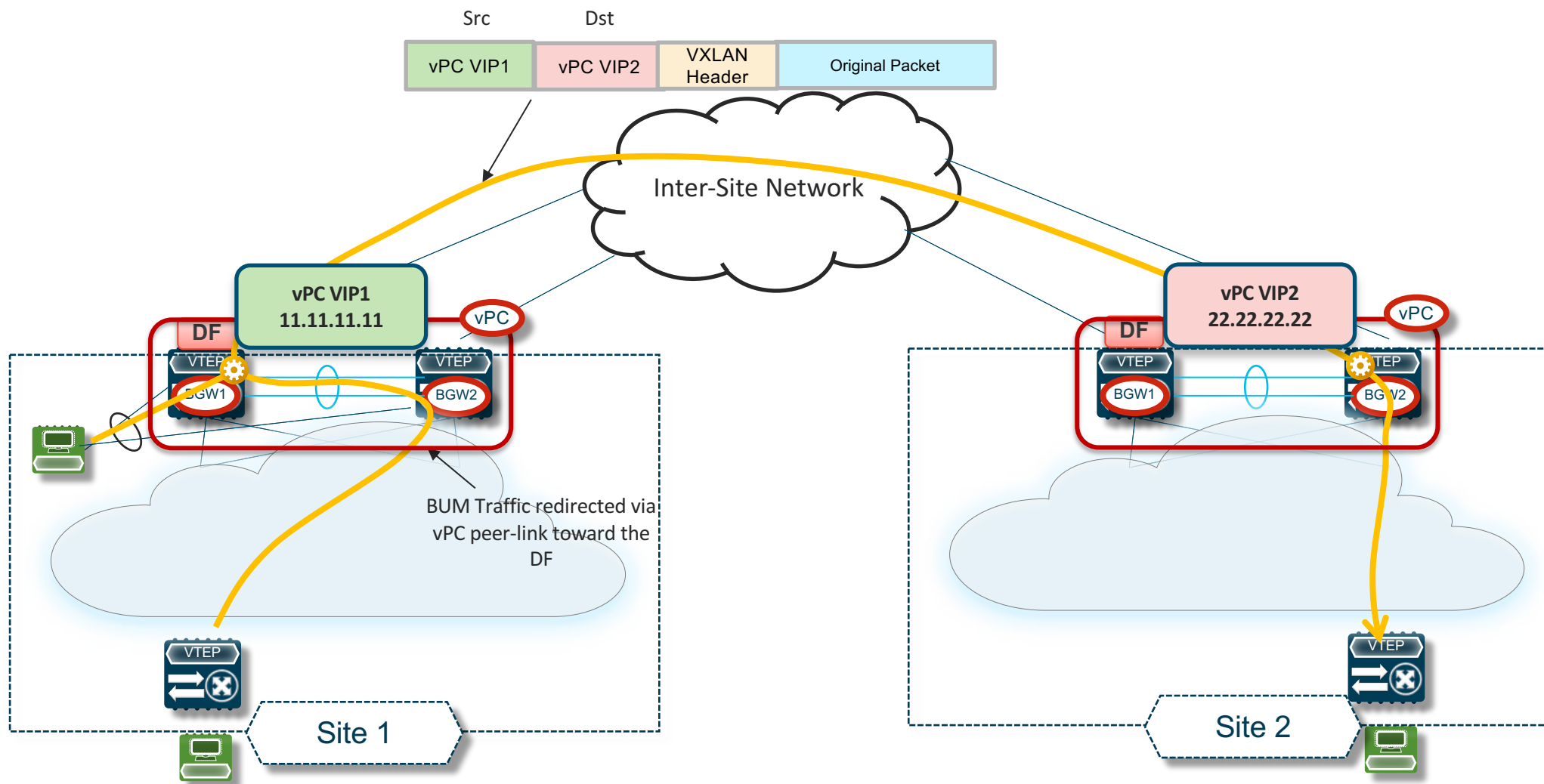
Трафик между системами, подключенными к сайтам



VPC Border Gateways

BUM трафик между сайтами

NX-OS Release
9.2(1)

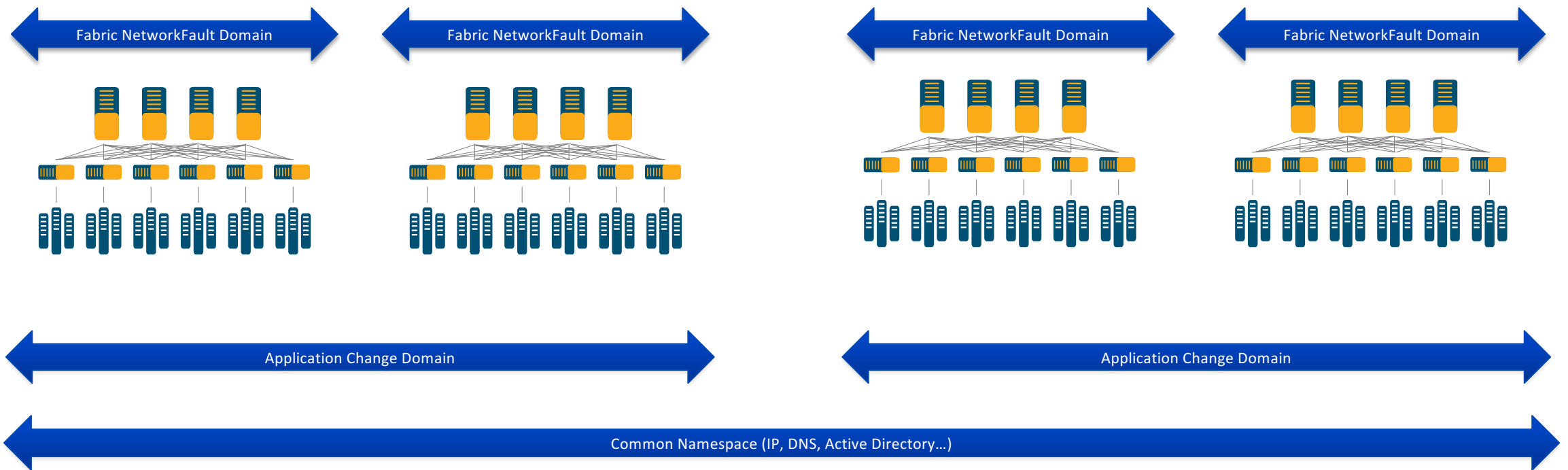


Развитие решений по управлению
распределенными фабриками

Регионы и зоны доступности

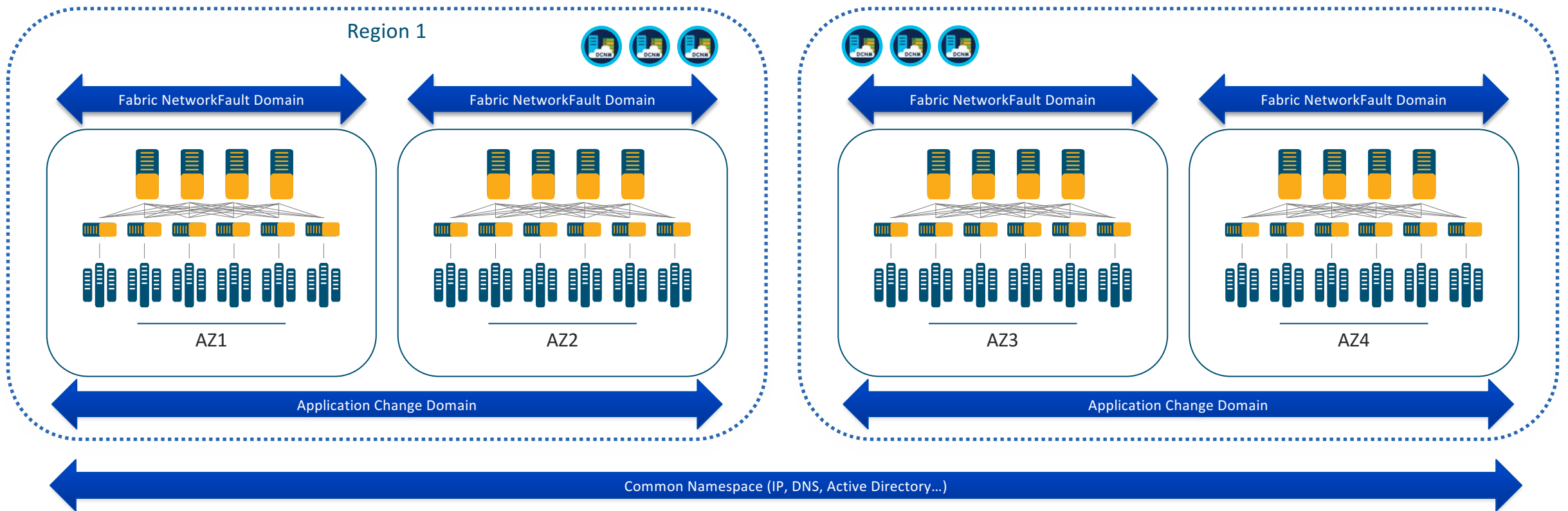
Зоны доступности и регионы

Изоляция изменений и сетевых доменов сбоя



Зоны доступности и регионы

Изоляция изменений и сетевых доменов сбоя



Зоны доступности и регионы

Изоляция изменений и сетевых доменов сбоя

Active Workloads
Layer 2 & Layer 3

Region 1

Fabric NetworkFault Domain

Fabric NetworkFault Domain

AZ1

AZ2

Application Change Domain

Active Workloads
Layer 2 & Layer 3

Fabric NetworkFault Domain

Fabric NetworkFault Domain

AZ3

AZ4

Application Change Domain

Common Namespace (IP, DNS, Active Directory...)

Зоны доступности и регионы

Изоляция изменений и сетевых доменов сбоя

Active Workloads
Layer 2 & Layer 3

Region 1

Fabric NetworkFault Domain

Fabric NetworkFault Domain

AZ1

Application 1

Workload deployed across AZs

Application Change Domain

Active Workloads
Layer 2 & Layer 3

Fabric NetworkFault Domain

Fabric NetworkFault Domain

AZ3

Application 2

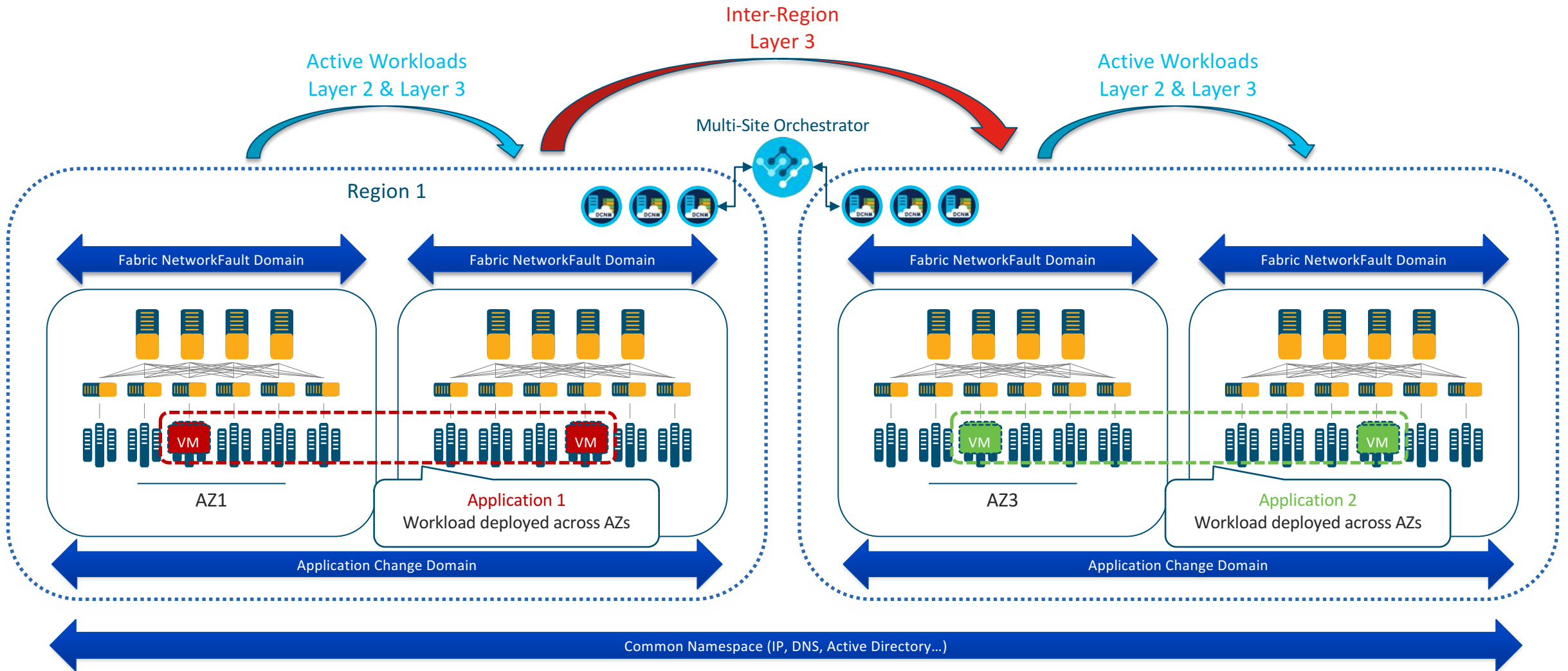
Workload deployed across AZs

Application Change Domain

Common Namespace (IP, DNS, Active Directory...)

Зоны доступности и регионы

Изоляция изменений и сетевых доменов сбоя



Интеграция DCNM и MSO

MSO и DCNM интеграция

Требуемые компоненты

MSO-DCNM интеграция требует 3 компоненты:

1. Nexus Dashboard

Вычислительная платформа для хостинга приложений day-2 operations и оркестрации

2. MSO

Единая платформа управления и оркестрации для всех подключенных сайтов. Этот оркестратор настраивает политики для взаимодействия между сайтами

3. DCNM MSDs (два или более)

Единая точка управления одной или несколькими фабриками VXLAN EVPN

Минимальные версии ПО

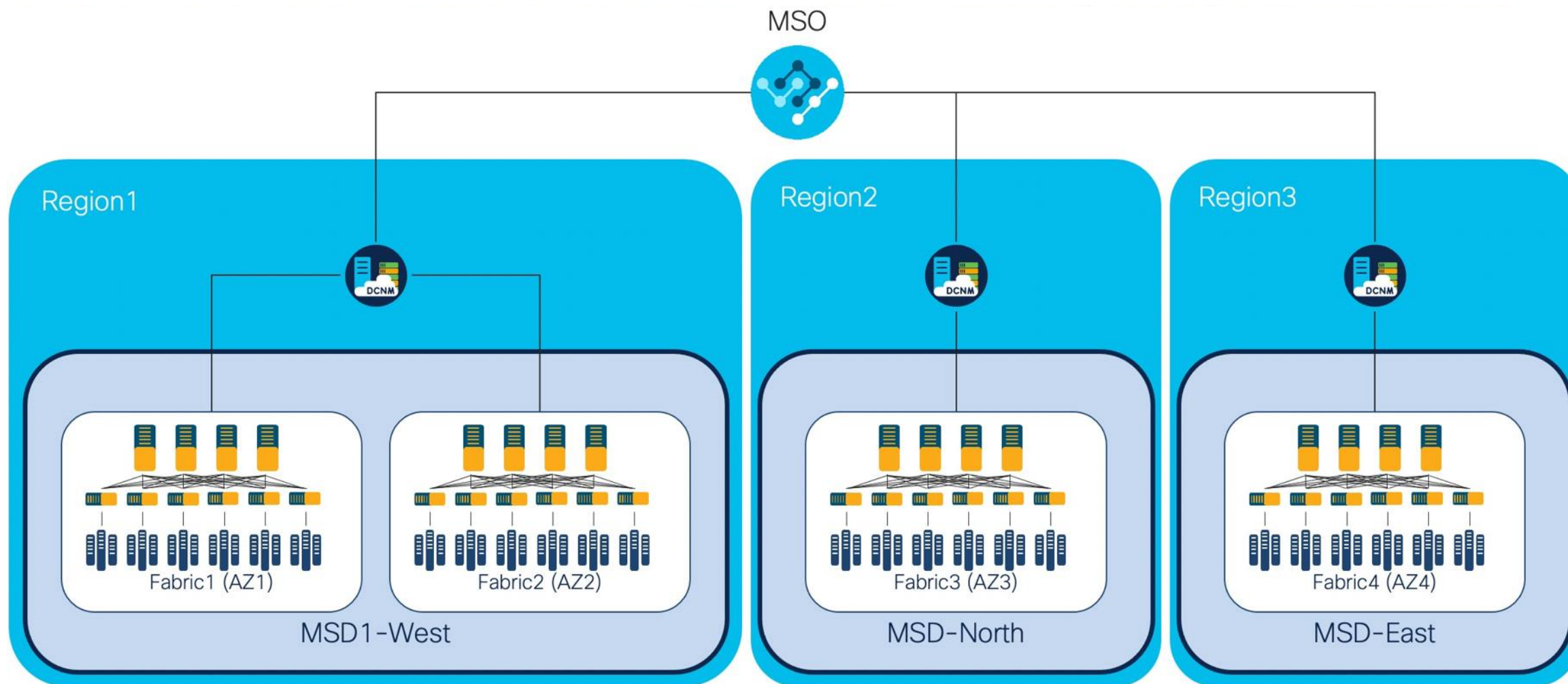
ND - 2.0(1)

MSO - 3.2(1)

DCNM - 11.5(1)

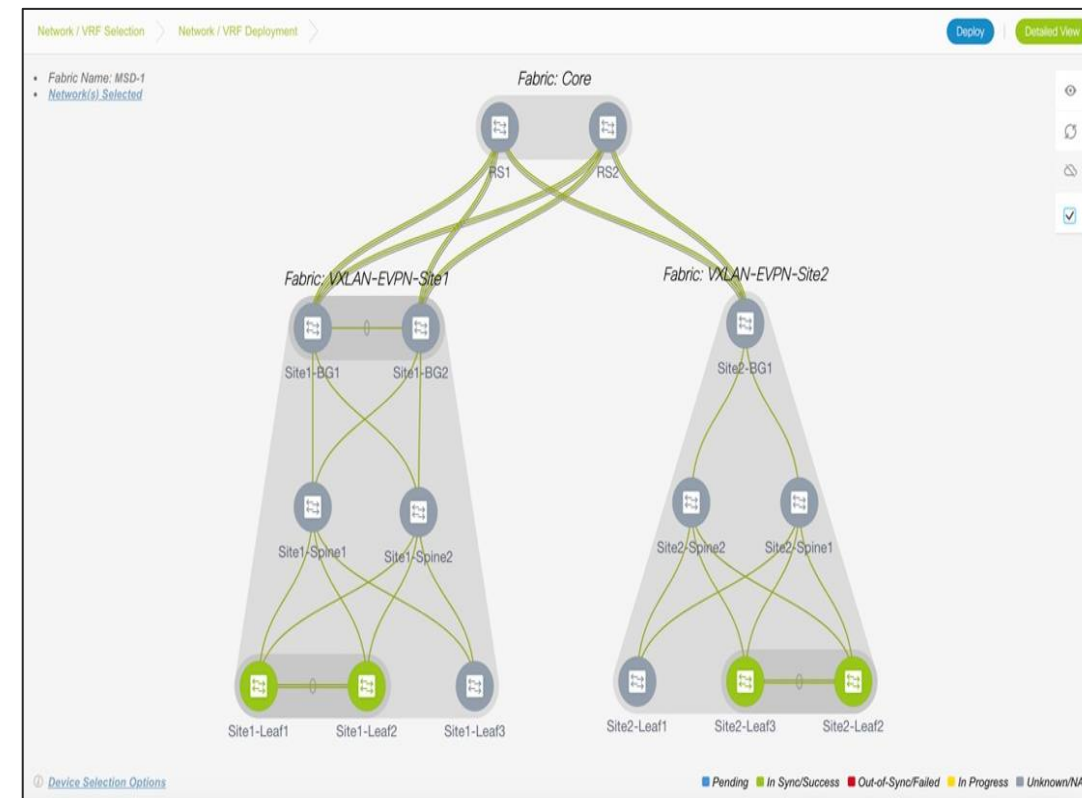
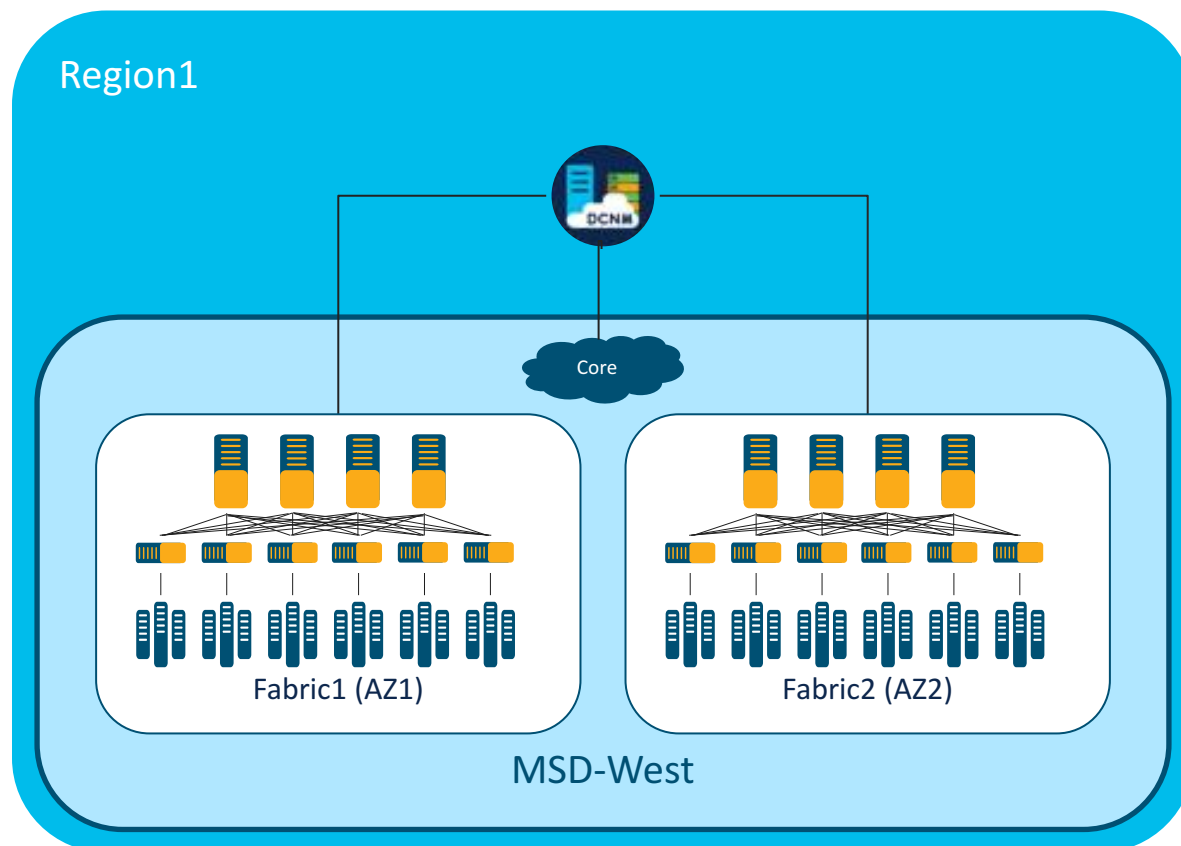
MSO и DCNM интеграция

Отношения между фабриками и MSD



MSO и DCNM интеграция

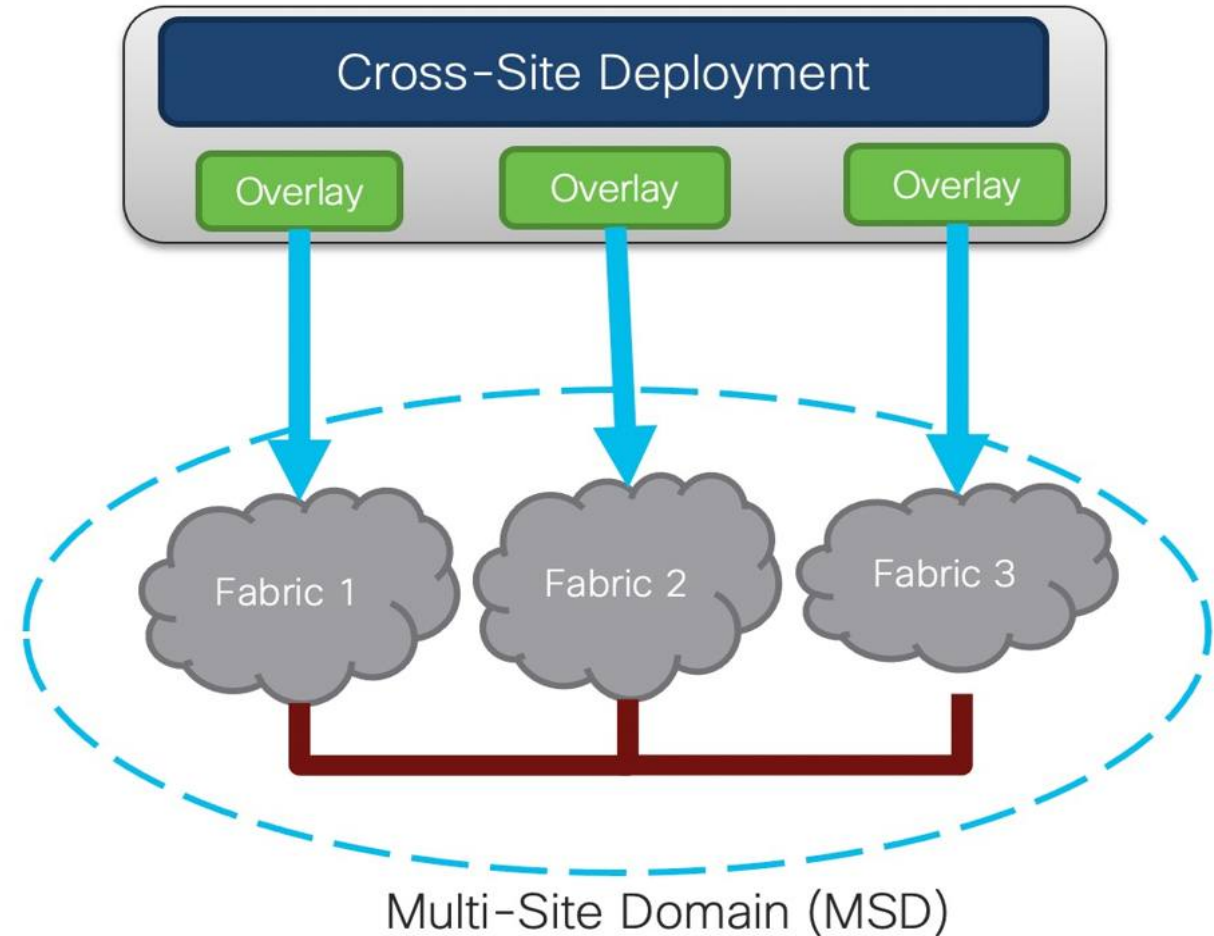
Определение MSD



DCNM Multi-Site Domain внутри DCNM

Создание зон доступностей

- Настройка Border Gateway
- Настройка Multi-Site Underlay и Overlay peering
- Multi-Site Domain с участниками (сайтами)
- Единое определение объектов Networks & VRFs
 - Определение автоматически наследуется на участников
- Расширение VRFs/Networks по требованию при помощи Multi-Site



Интеграция MSO и DCNM

Объединение регионов

Атрибуты и проблемы:

- Гибкость гео-резервирования DCNM
 - Один DCNM экземпляр на регион
 - Stretched HA не поддерживается
- MSO организует и синхронизирует зоны доступности между регионами
 - DCNM управляет многими зонами доступности внутри одного региона
- Масштабирование опирается на масштабирование DCNM умноженное на масштабирование MSO
- Несколько DCNM экземпляров
 - Standalone или HA
 - Синхронизация при помощи MSO
- MSO управляет фабриками на нескольких DCNM экземплярах
 - Несколько фабрик может быть организовано в группу
 - Множество доменов контроля изменений, один такой домен на один экземпляр DCNM

Управление фабрикой при помощи DCNM и MSO



Управление фабрикой при помощи DCNM и MSO

- 1 Добавление сайтов DCNM в интерфейс Nexus Dashboard
- 2 Перевод сайтов DCNM под управление MSO
- 3 Настройка скорости на интерфейсах в ISN-сети*
- 4 Настройка связанности между BGW и ISN

* Опциональный шаг, связан с особенностями работы устройств в лаборатории

Демонстрация

Добавление сайтов DCNM в MSO

Nexus Dashboard и MSO
развернуты

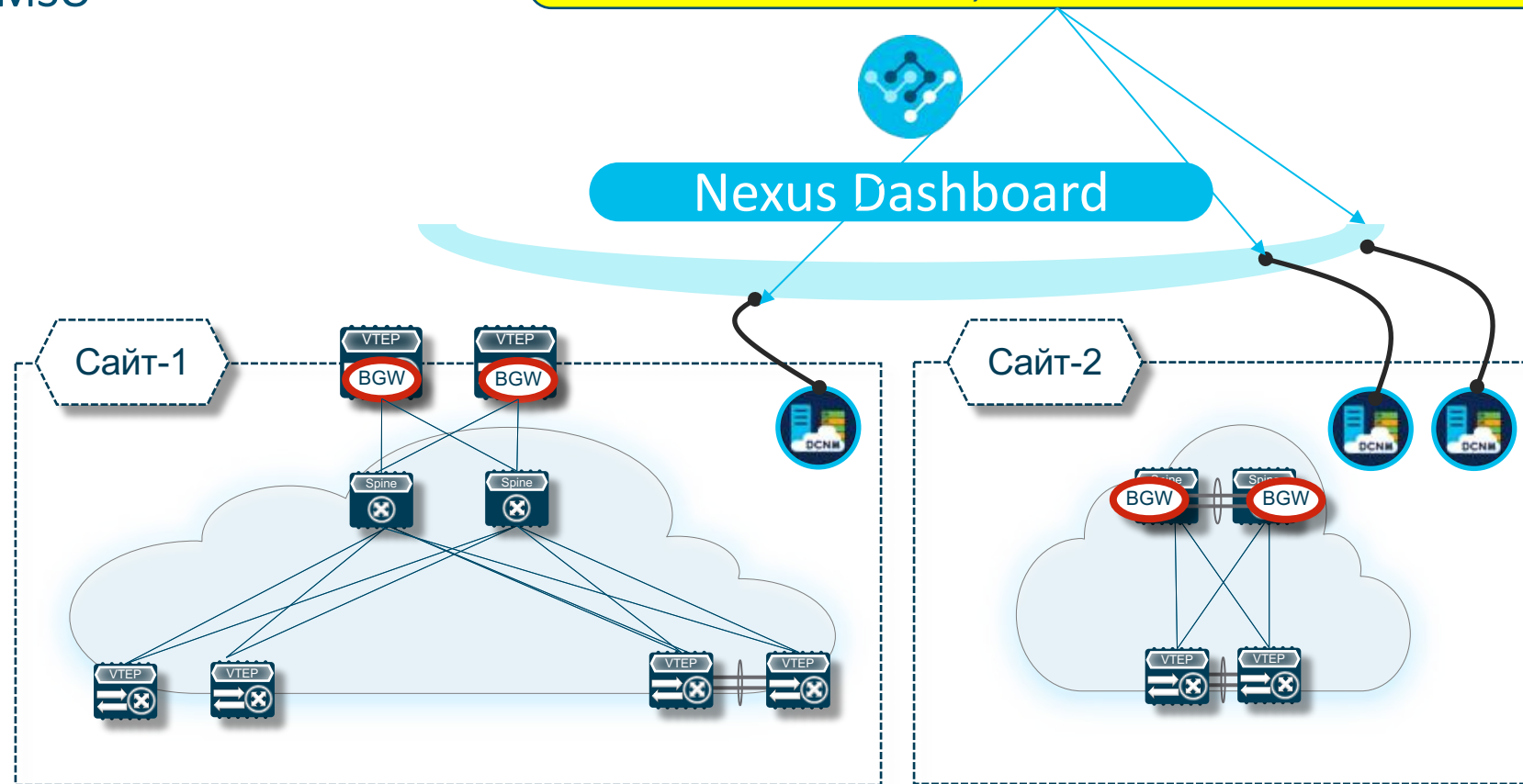
0

DCNM в каждом из ЦОД
развернут и управляет
устройствами
локальными для сайта

1

Добавляем сайты
DCNM в MSO,
используем in-band
интерфейс

Для подключения DCNM к MSO используются in-band интерфейсы
Если они не были настроены изначально, то это можно сделать и
после установки DCNM



ISN устройства имеет смысл добавлять к одному из сайтов
DCNM умеет управлять NX-OS устройствами, которые не являются частью фабрики

Nexus Dashboard

Version 2.0(1d)



Username

Password

Login

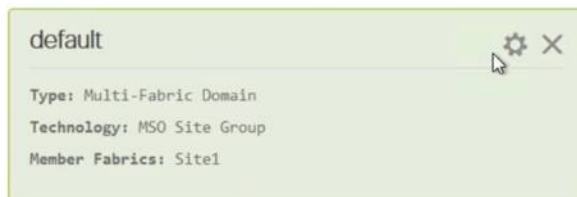
Демонстрация

Перевод сайтов DCNM под управление MSO

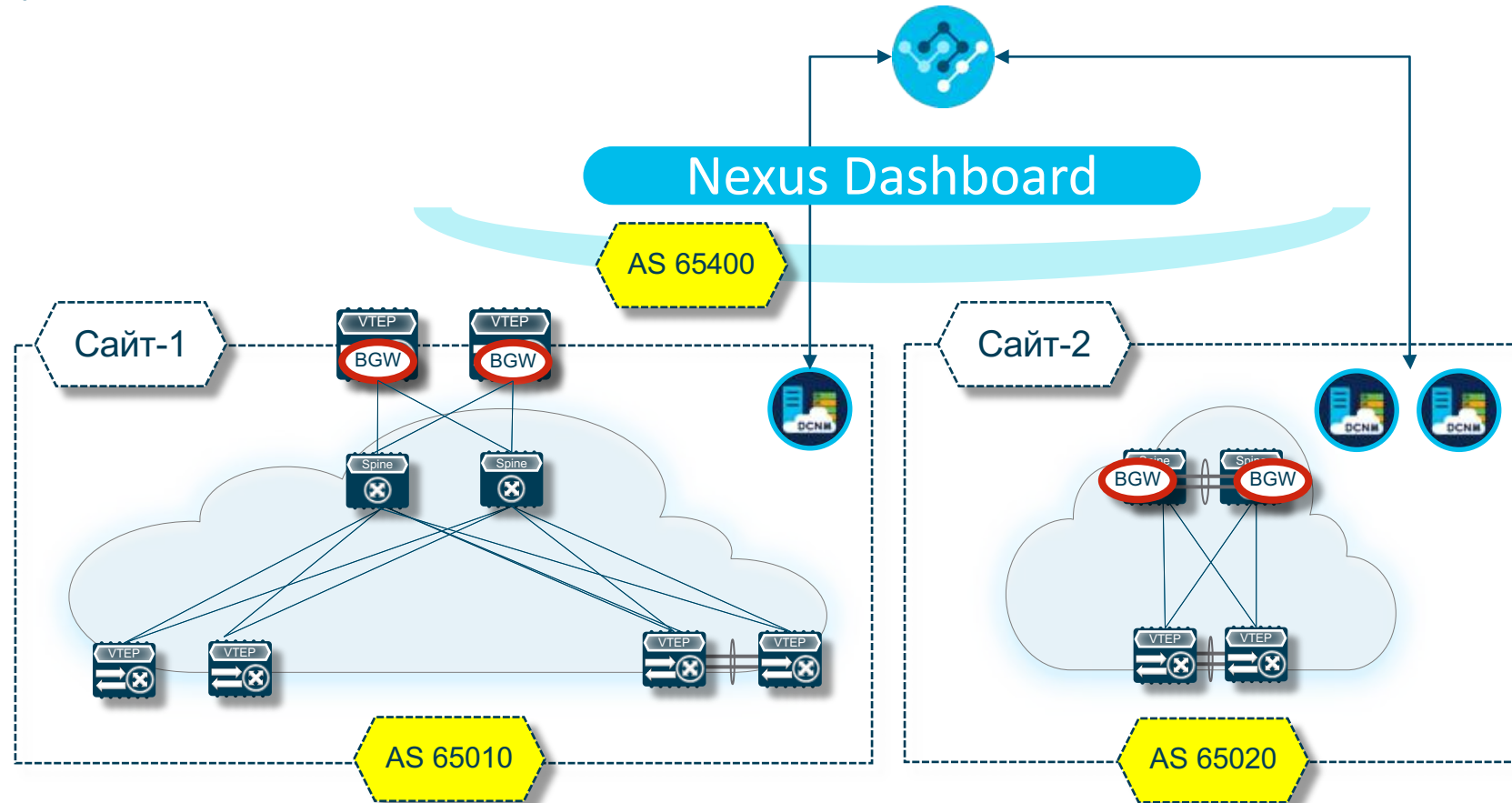
2

Изменяем режим управления DCNM в MSO с unmanaged на managed

Задаем номера AS для сайтов и ISN сегмента



Объект Multi-Fabric Domain создается автоматически внутри каждого DCNM



Filter by attributes

Actions

© 2021 Cisco and/or its affiliates. All rights reserved. Cisco Public

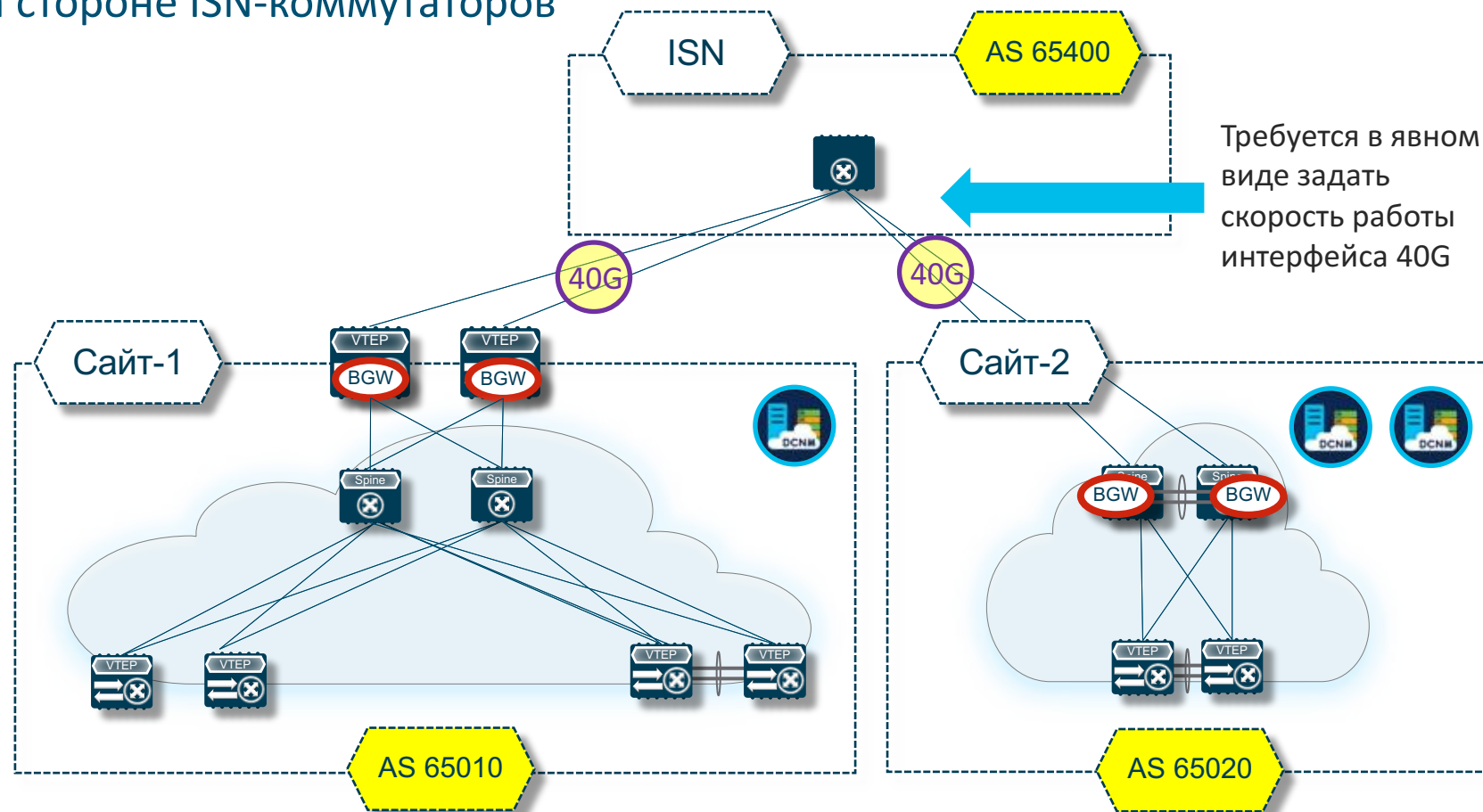
Демонстрация

Настраиваем интерфейсы на стороне ISN-коммутаторов

3

Создаем шаблон настроек для интерфейса (device level)

Применяем шаблон настроек интерфейса со скоростью 40G



Dashboard

Topology

Control

Monitor

Administration

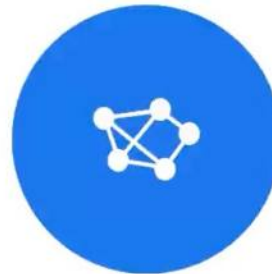
Applications

Good evening, admin!
Let's get started.



DCNM Licenses

License this copy of DCNM for each managed switch to unlock Performance Collection.



Fabric Builder

Creates a managed and controlled SDN fabric.



Networks & VRFs

Simple network overlay provisioning for N9K VXLAN EVPN Fabrics.



Documentation

Access cisco.com from documentation on configuration, maintenance and operation.

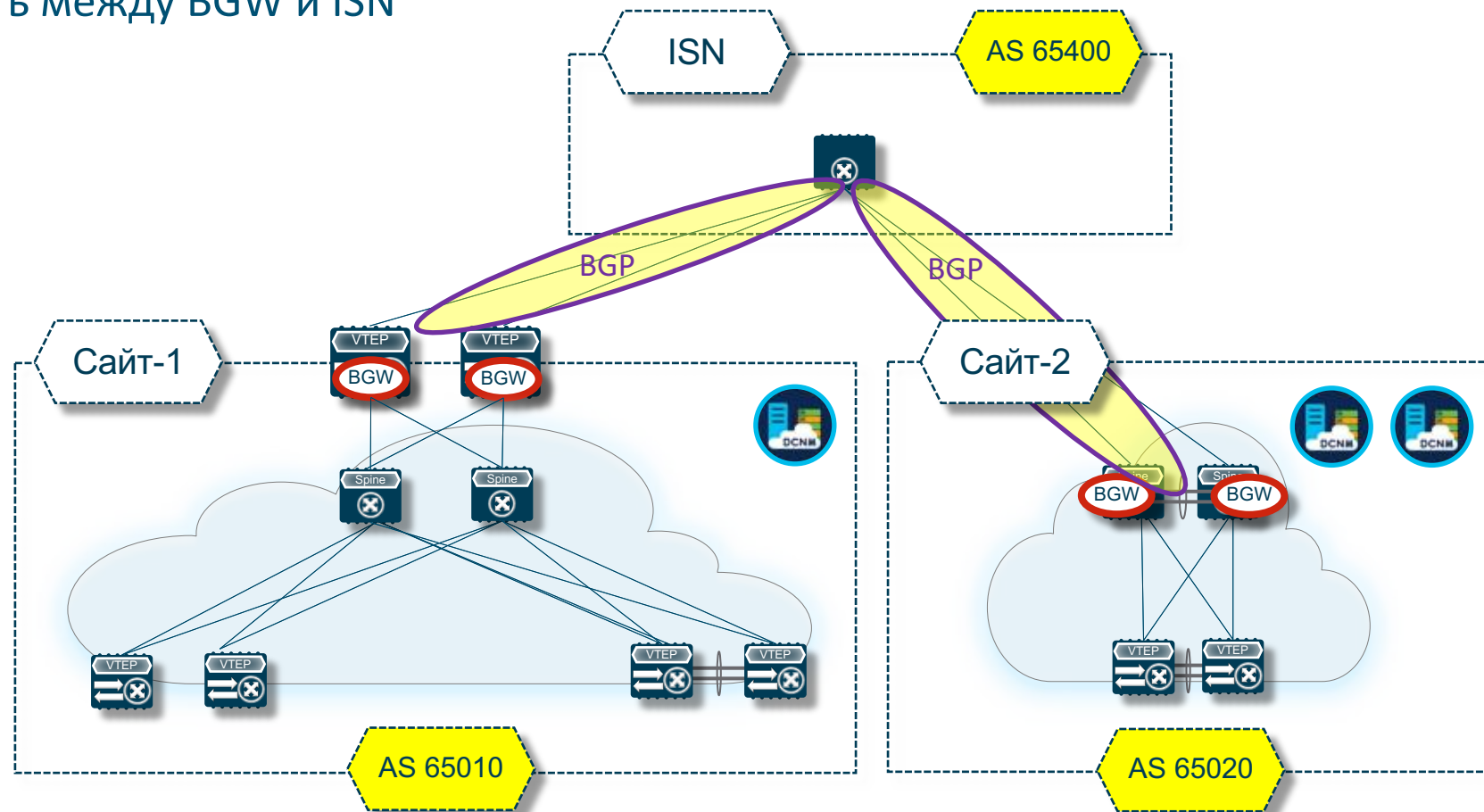
Демонстрация

MSO настраивает связанность между BGW и ISN

4

MSO настраивает BGP между BGW и ISN устройствами – underlay между сайтами

Адресация для каждого интерфейса между BGW и ISN указывается руками



Nexus Dashboard

Version 2.0(1d)



Username

admin

Password

••••••••

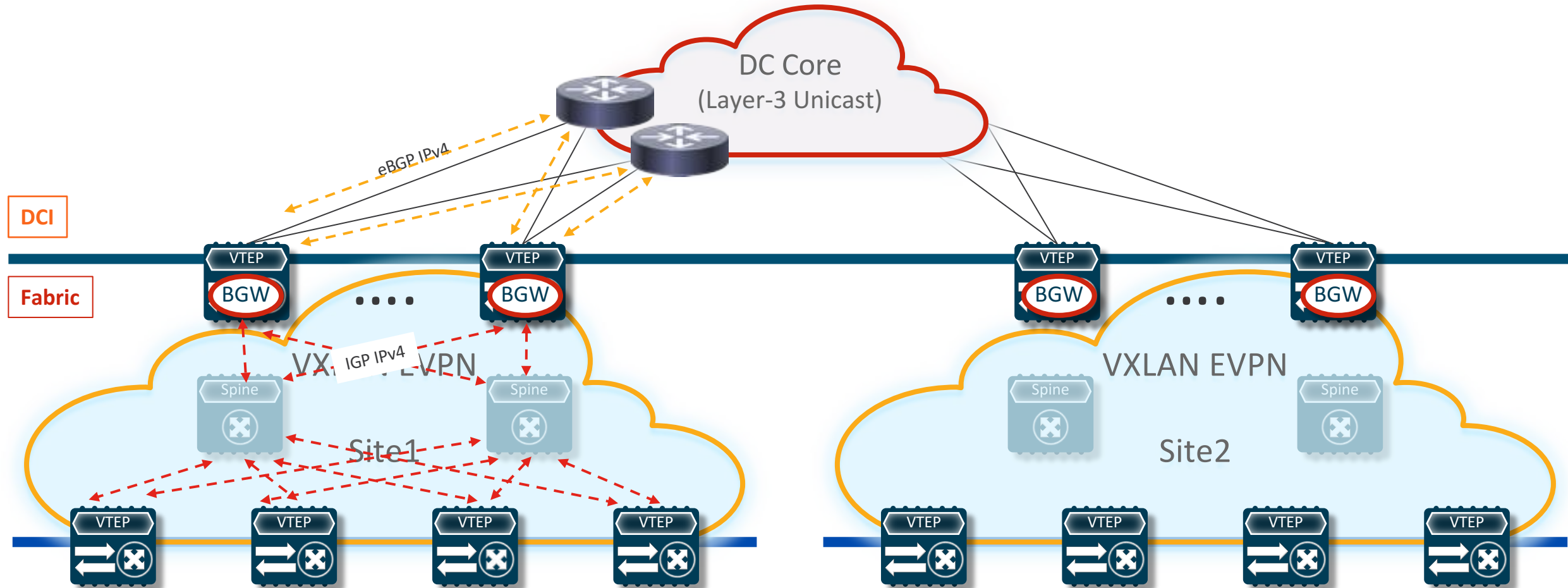
Login

Перерыв

Плоскость протоколов управления

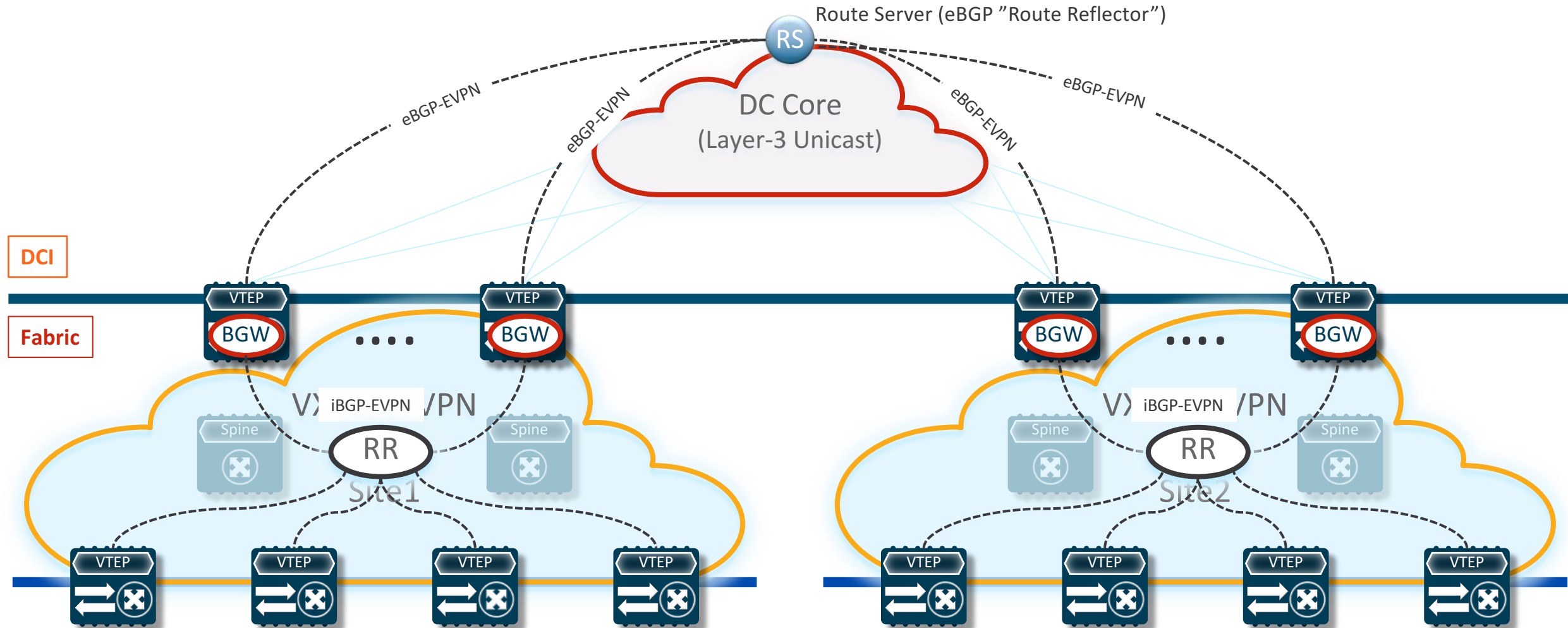
VXLAN Multi-Site

Underlay Control Plane

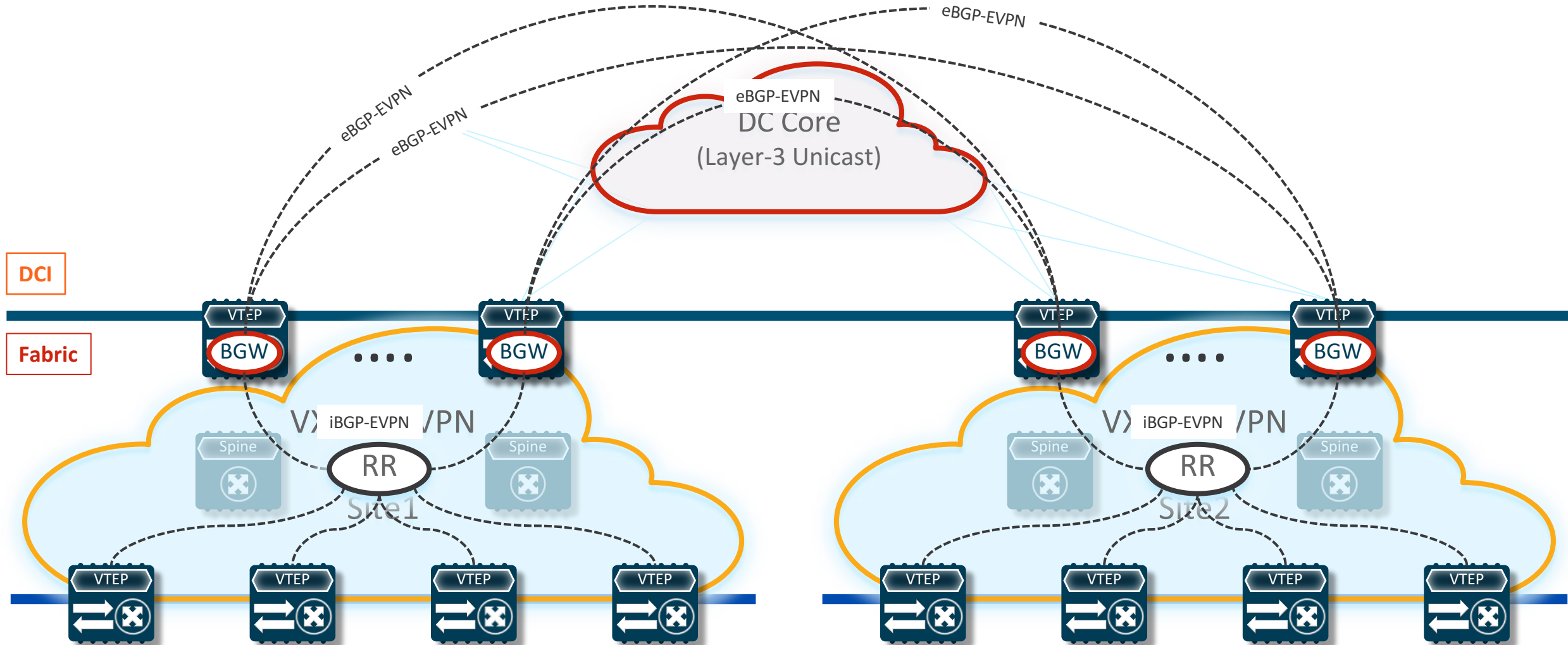


VXLAN Multi-Site

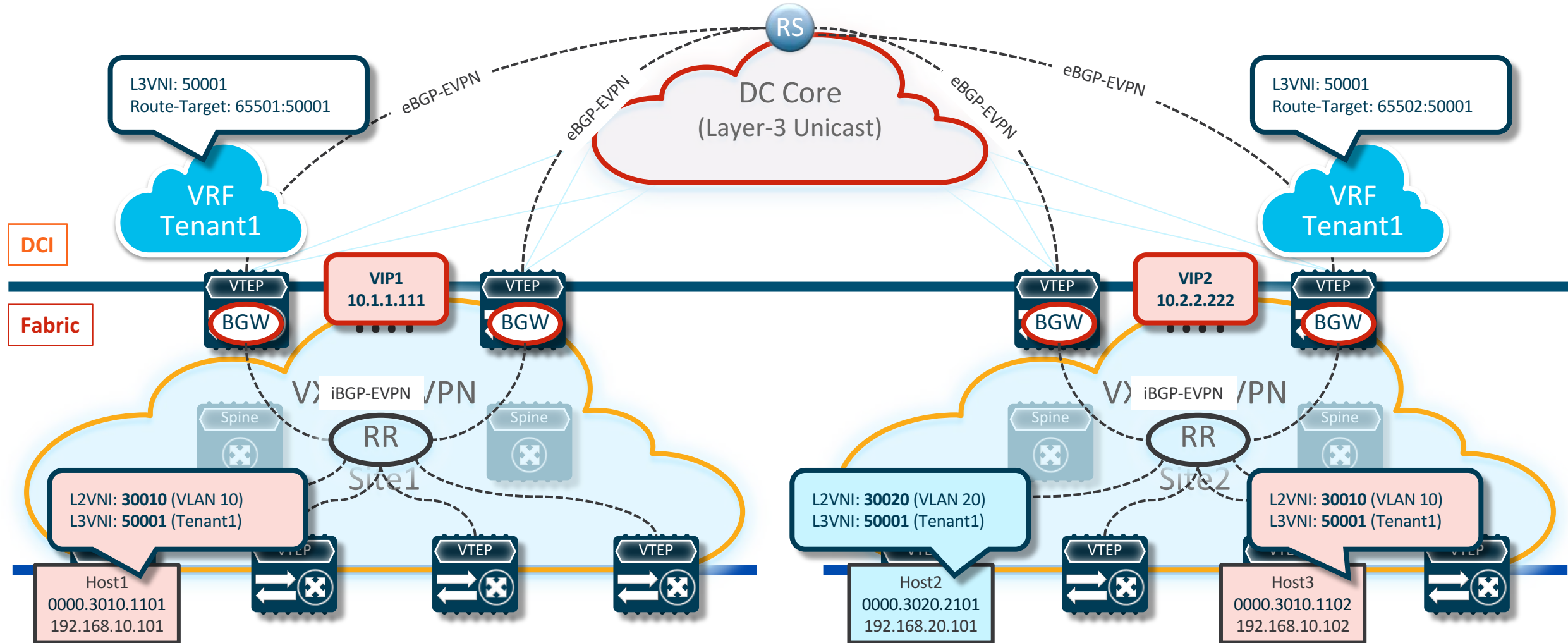
Overlay Control Plane (L3 Core)



Overlay Control Plane (L3 Core, no RS)

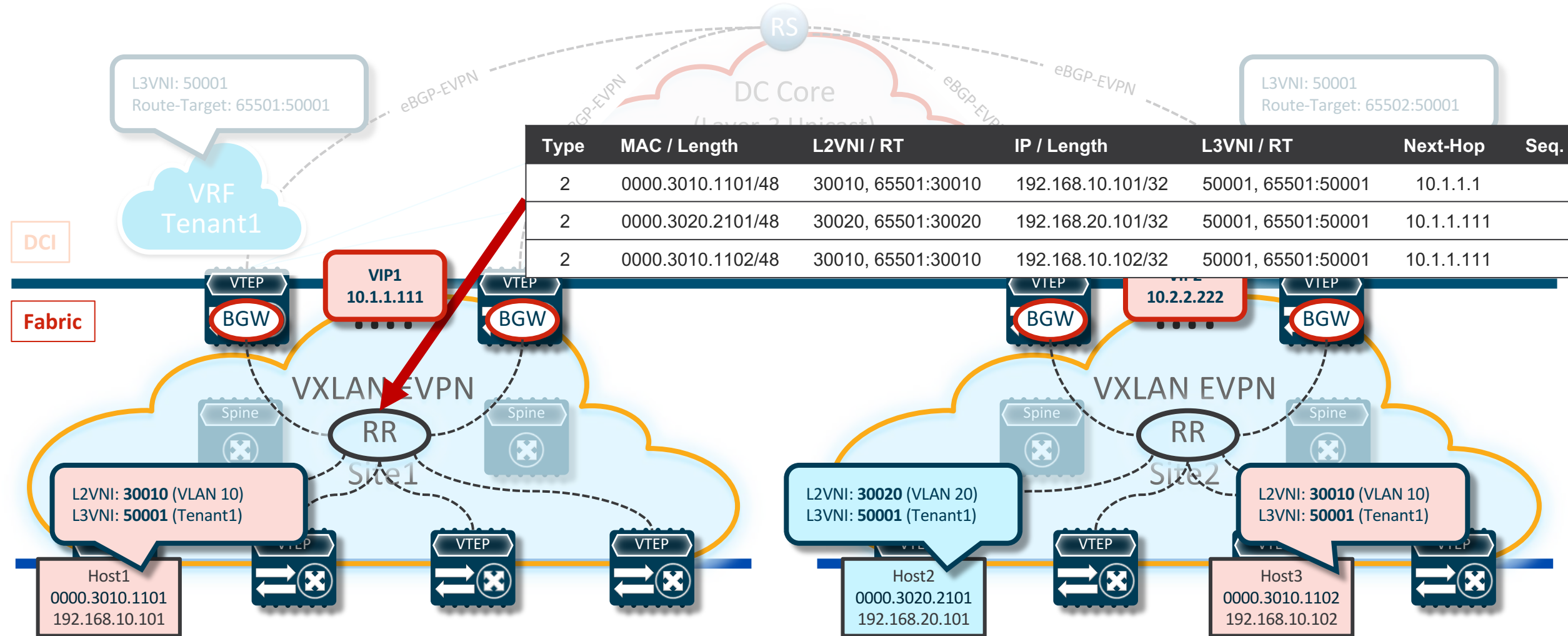


VXLAN Multi-Site Overlay Control Plane



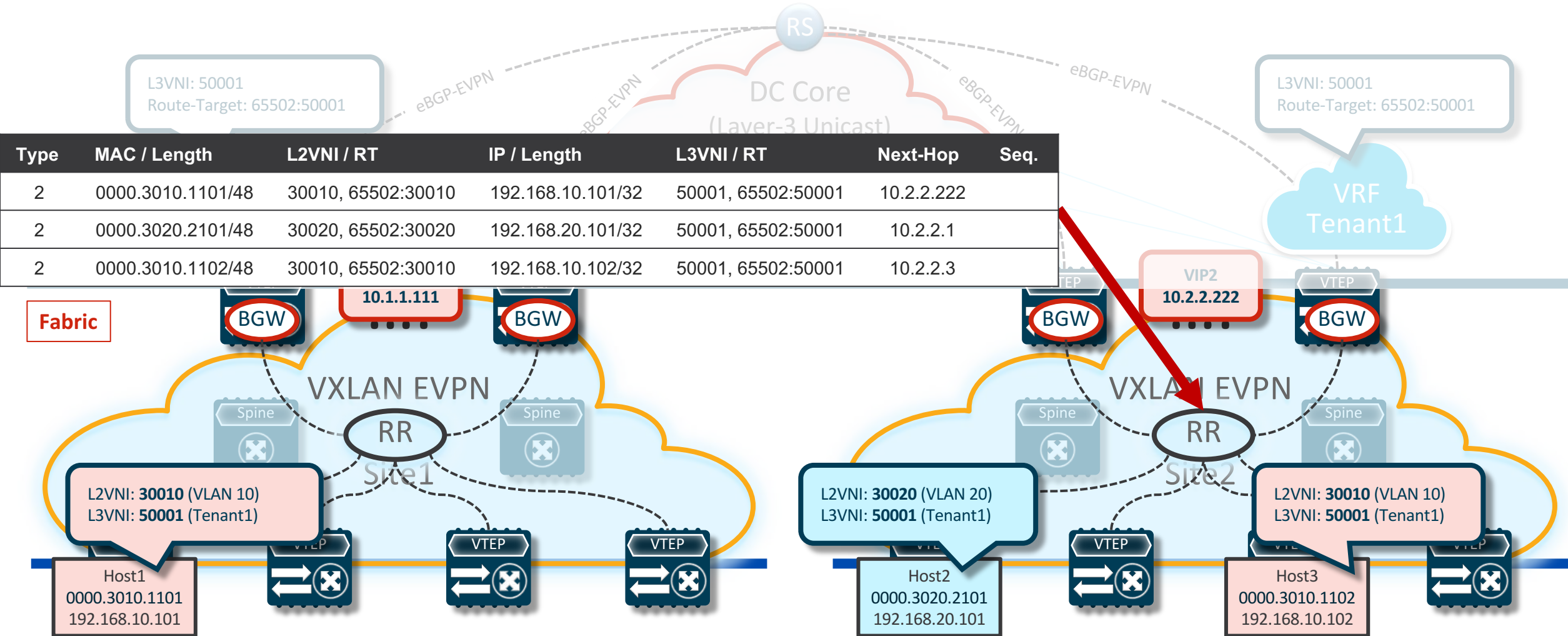
VXLAN Multi-Site

Overlay Control Plane (Site 1)



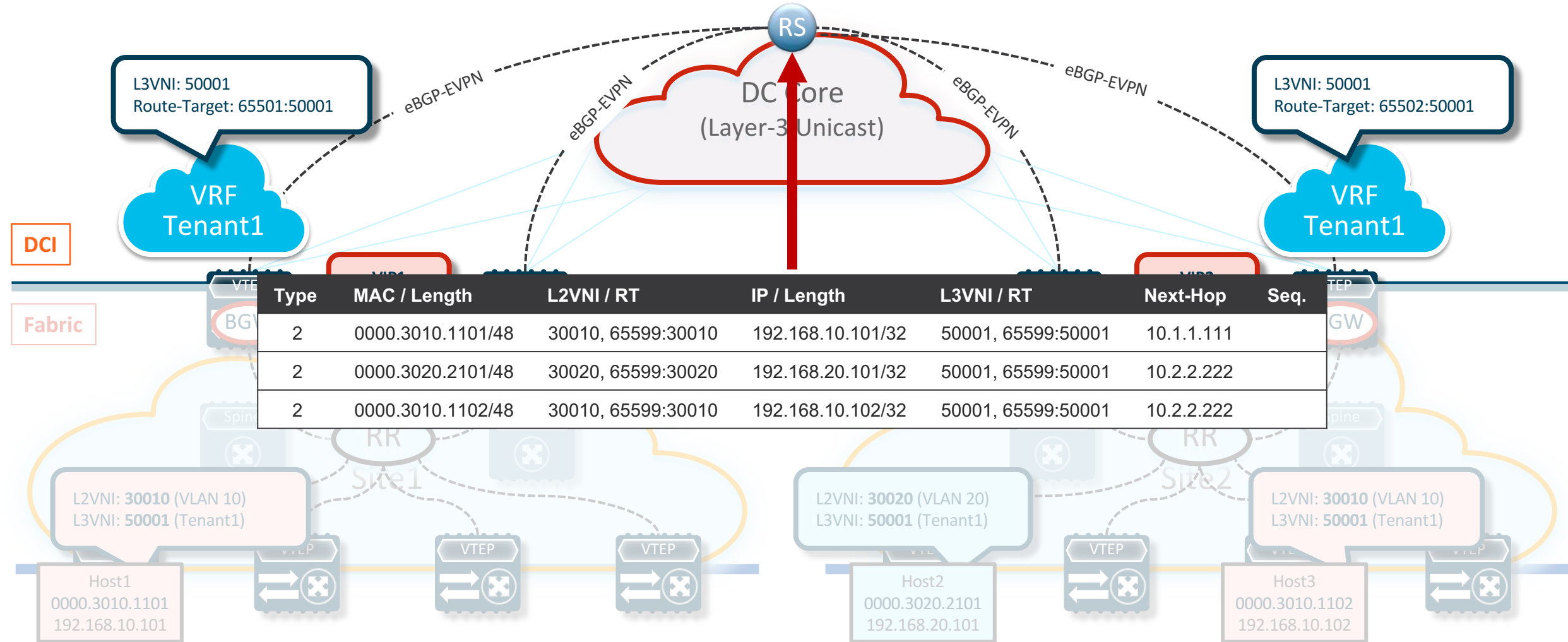
VXLAN Multi-Site

Overlay Control Plane (Site 2)



VXLAN Multi-Site

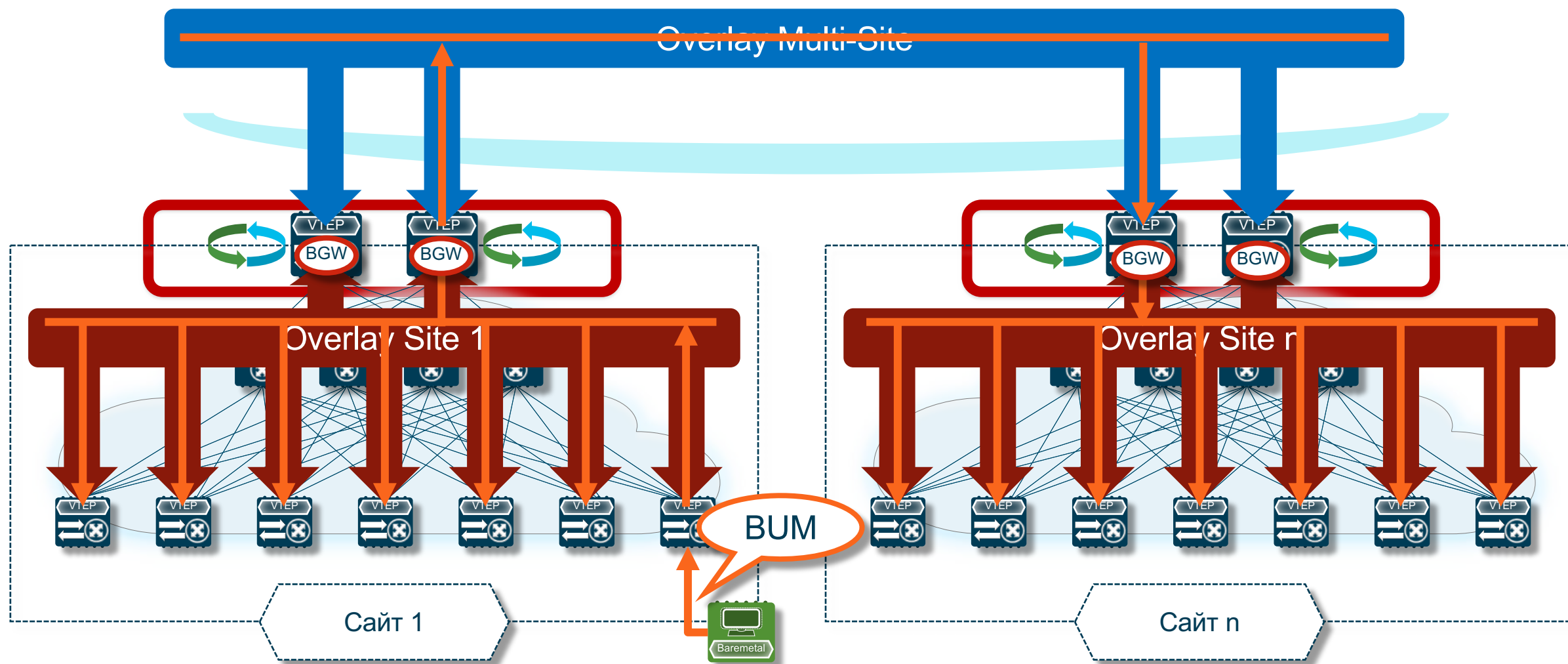
Overlay Control Plane (DCI)



Растягивание BUM-трафика
между сайтами

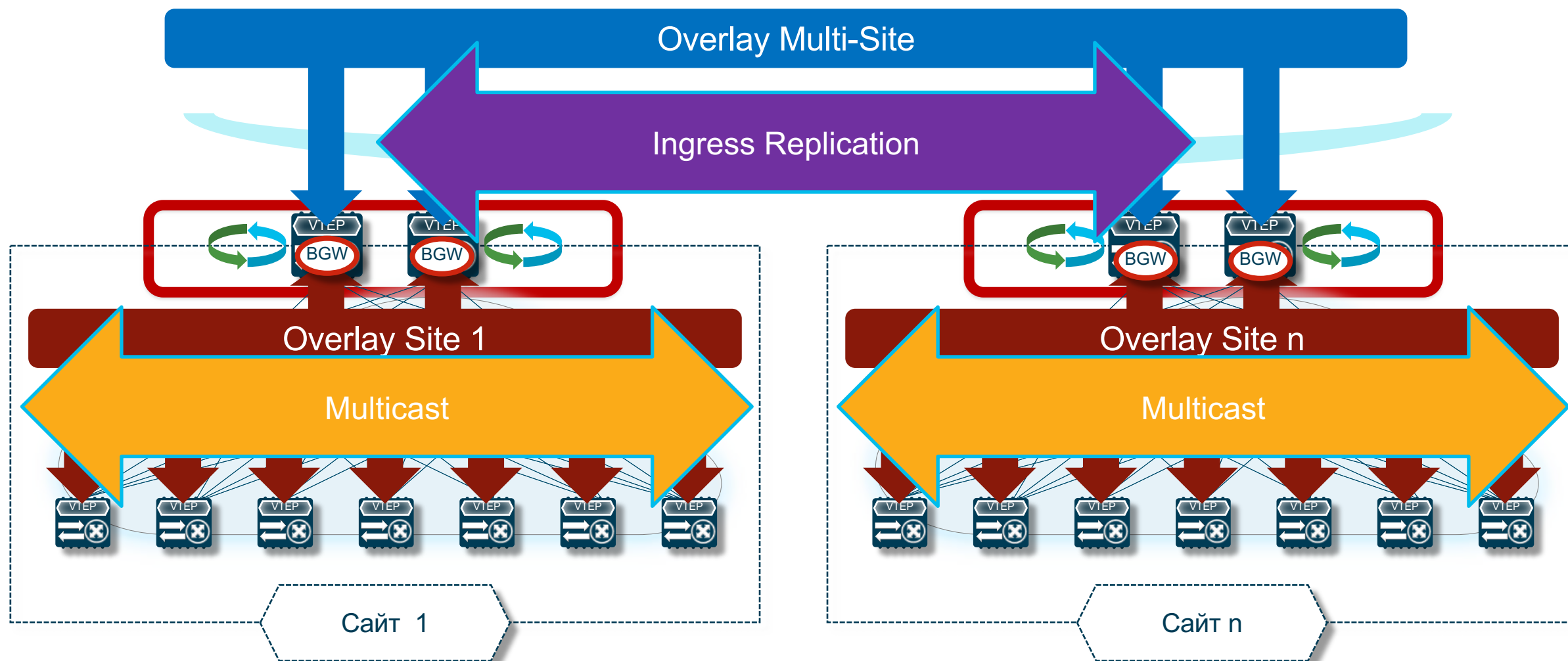
VXLAN Multi-Site

Передача BUM трафика



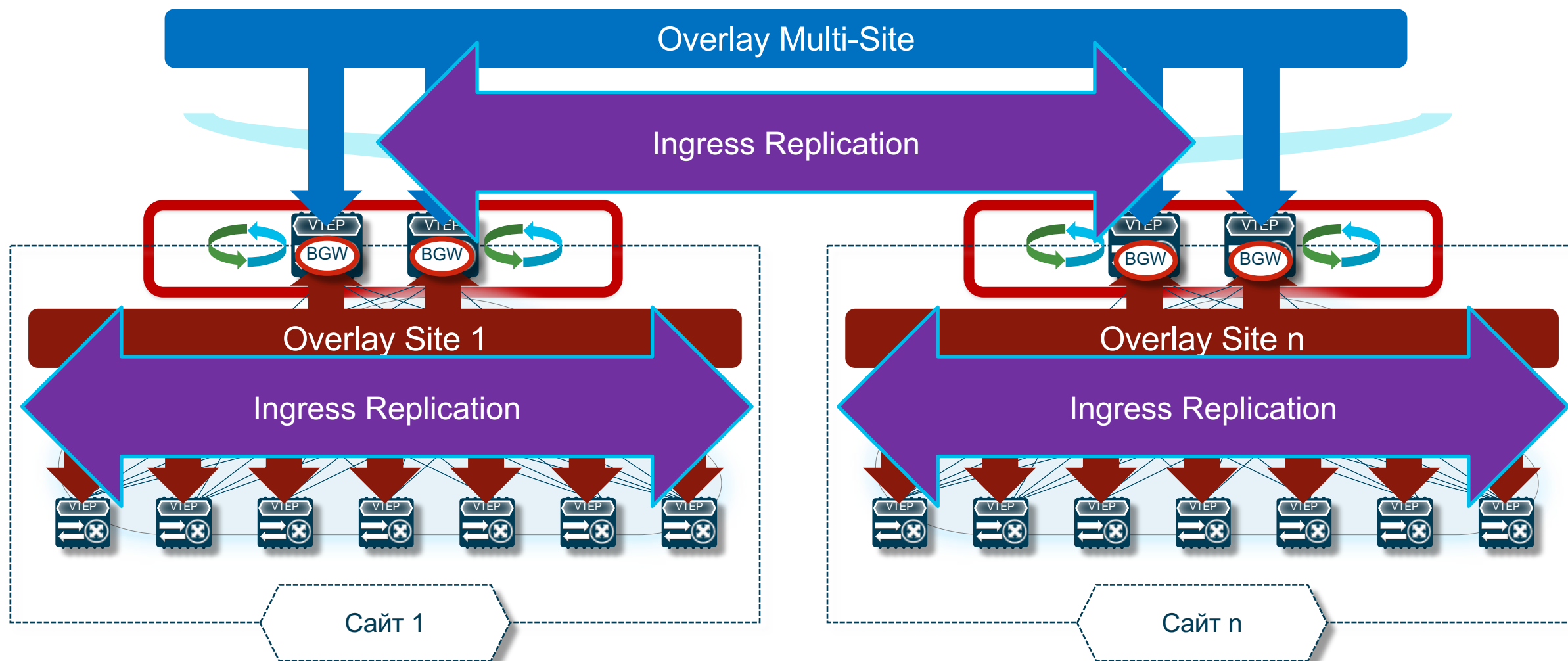
VXLAN Multi-Site

BUM режимы репликации (Multicast Intra-Site)



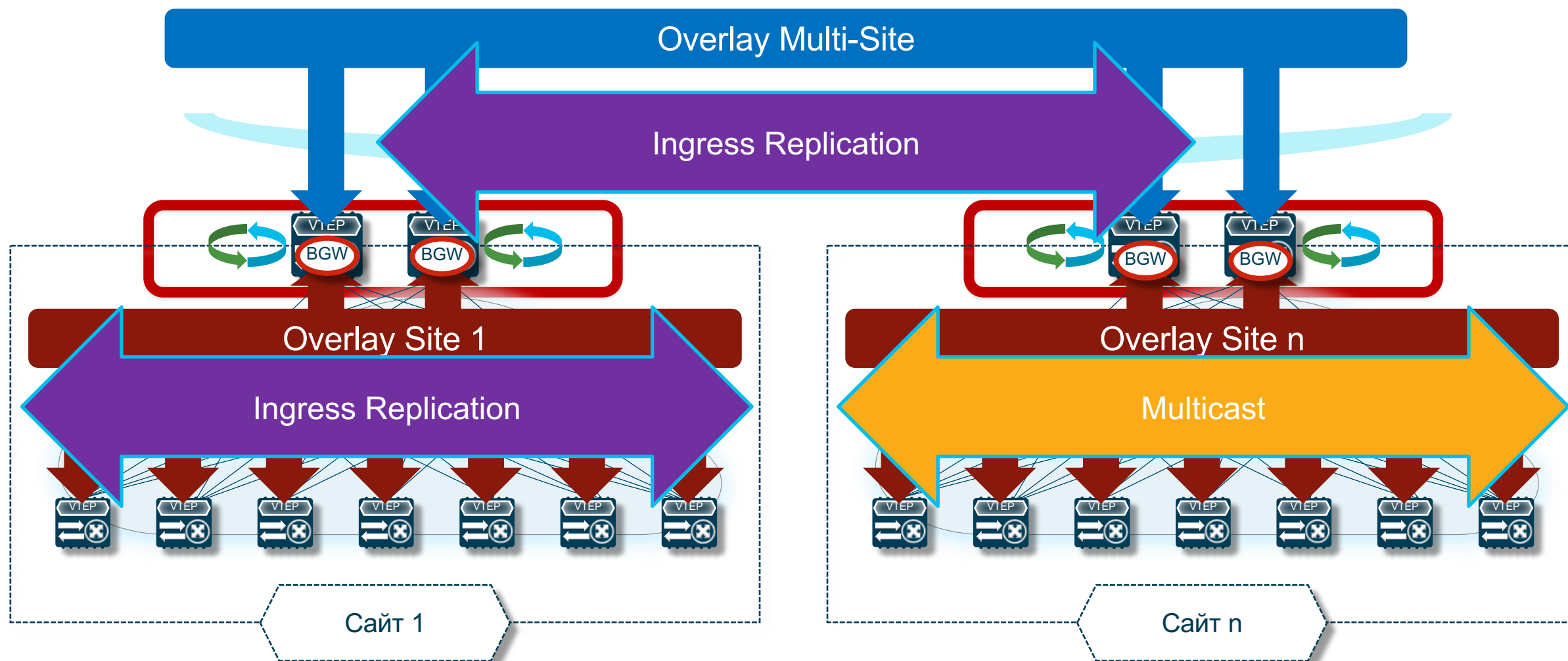
VXLAN Multi-Site

BUM режимы репликации (Ingress Replication Only)



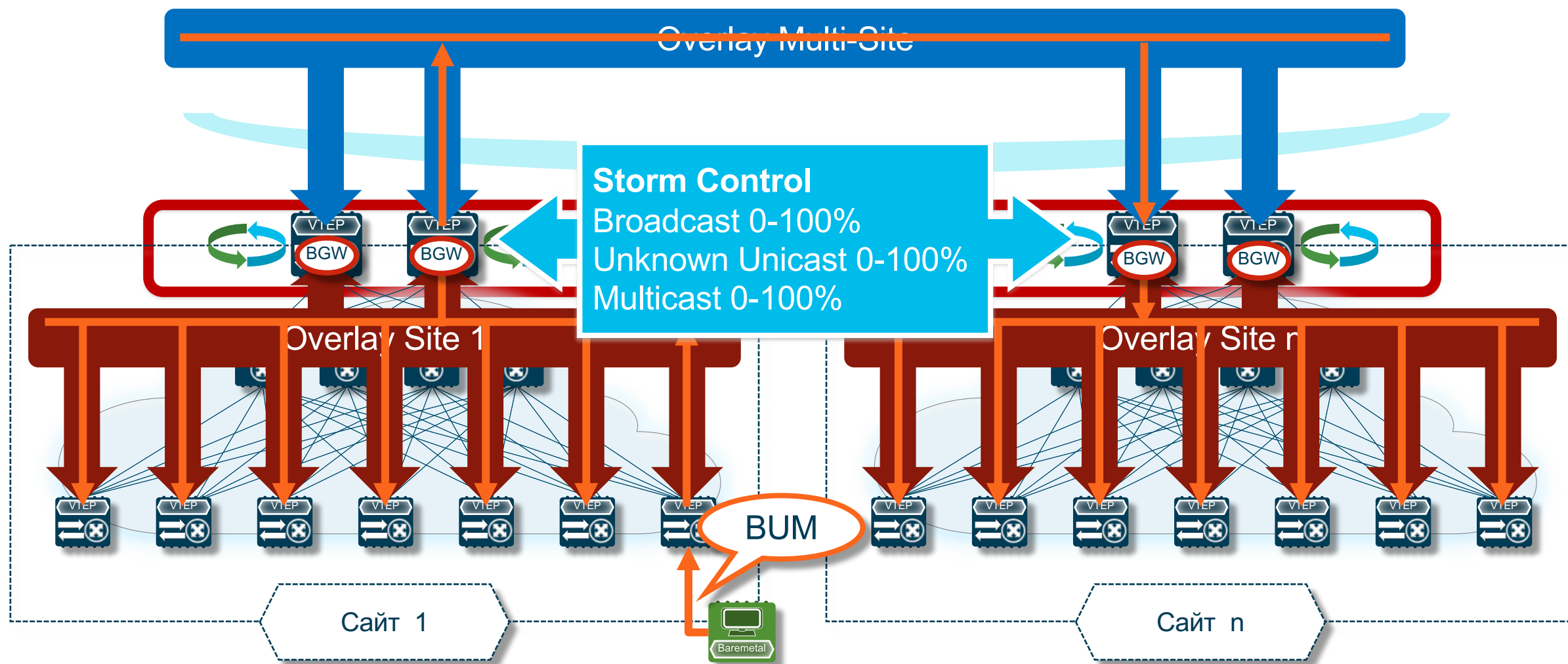
VXLAN Multi-Site

BUM режимы репликации (Mixed Mode Intra-Site)



VXLAN Multi-Site

BUM Traffic Policing

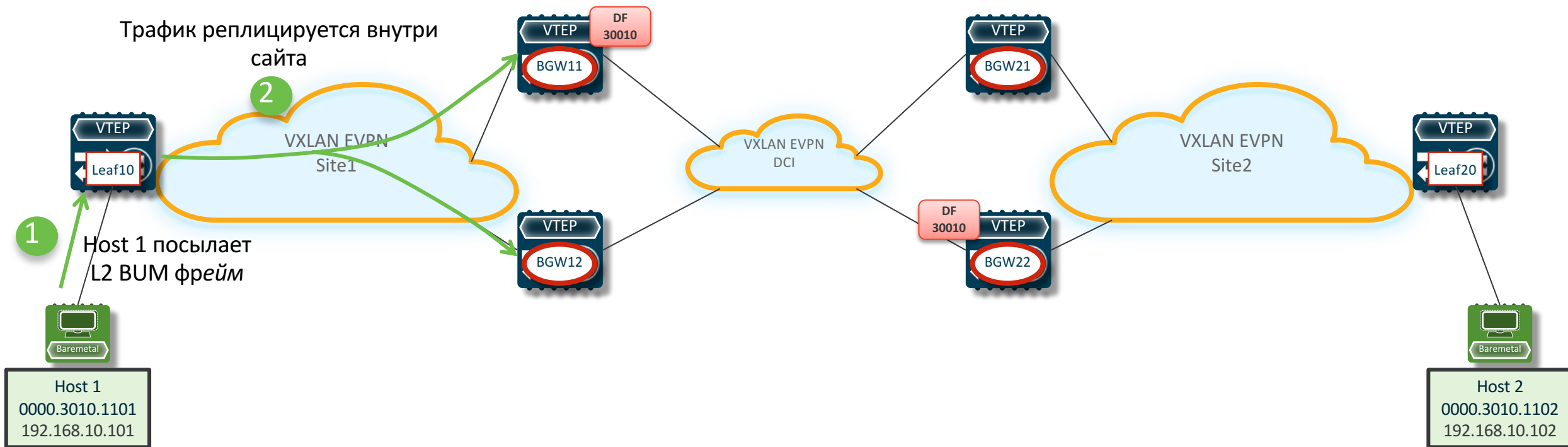


Передача пакетов в VXLAN Multi-Site

- Layer 2 (BUM) – Site 1

Передача BUM трафика

SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
L10	DGROUP	30010	H1-MAC	ALL-F	H1-IP	ALL-255	

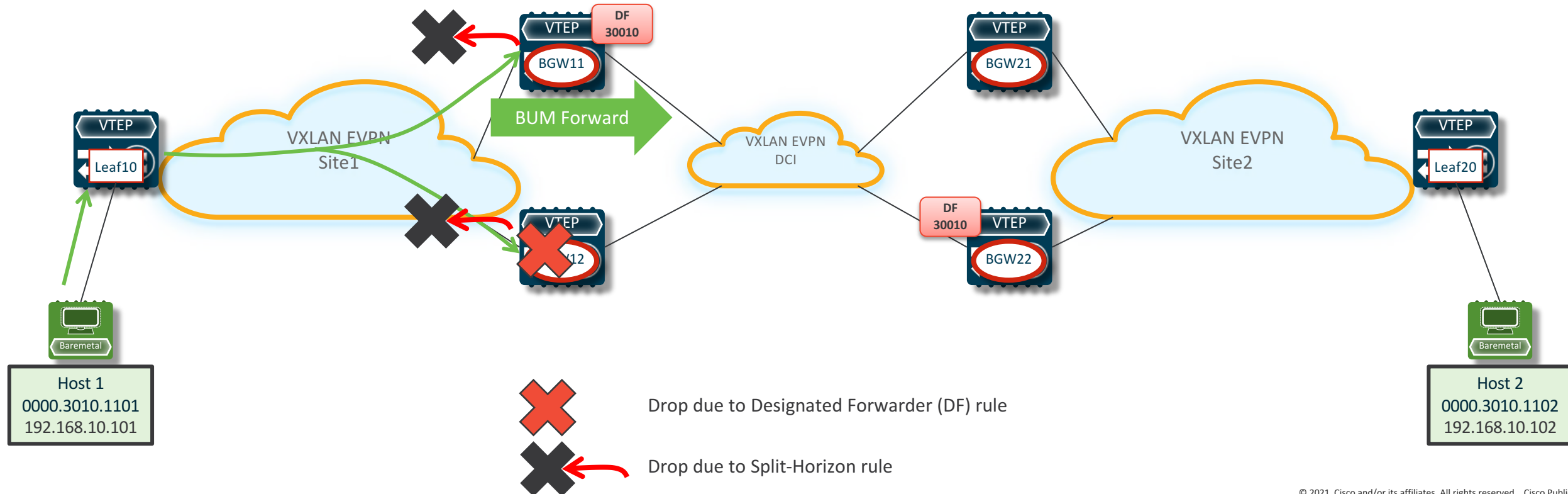


Передача пакетов в VXLAN Multi-Site

- Layer 2 (DF and Split Horizon) – Site 1

Передача BUM трафика

SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
L10	DGROUP	30010	H1-MAC	ALL-F	H1-IP	ALL-255	

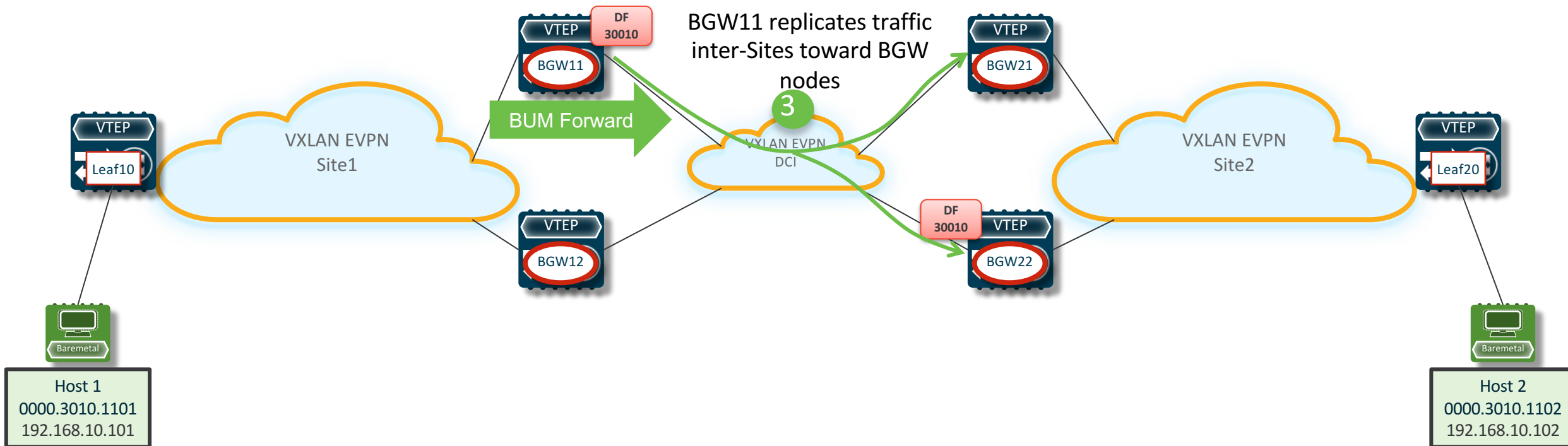


Передача пакетов в VXLAN Multi-Site

- Layer 2 (BUM) – DCI

Передача BUM трафика

SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
BGW11-PIP	BGW21	30010	H1-MAC	ALL-F	H1-IP	ALL-255	
BGW11-PIP	BGW22	30010	H1-MAC	ALL-F	H1-IP	ALL-255	

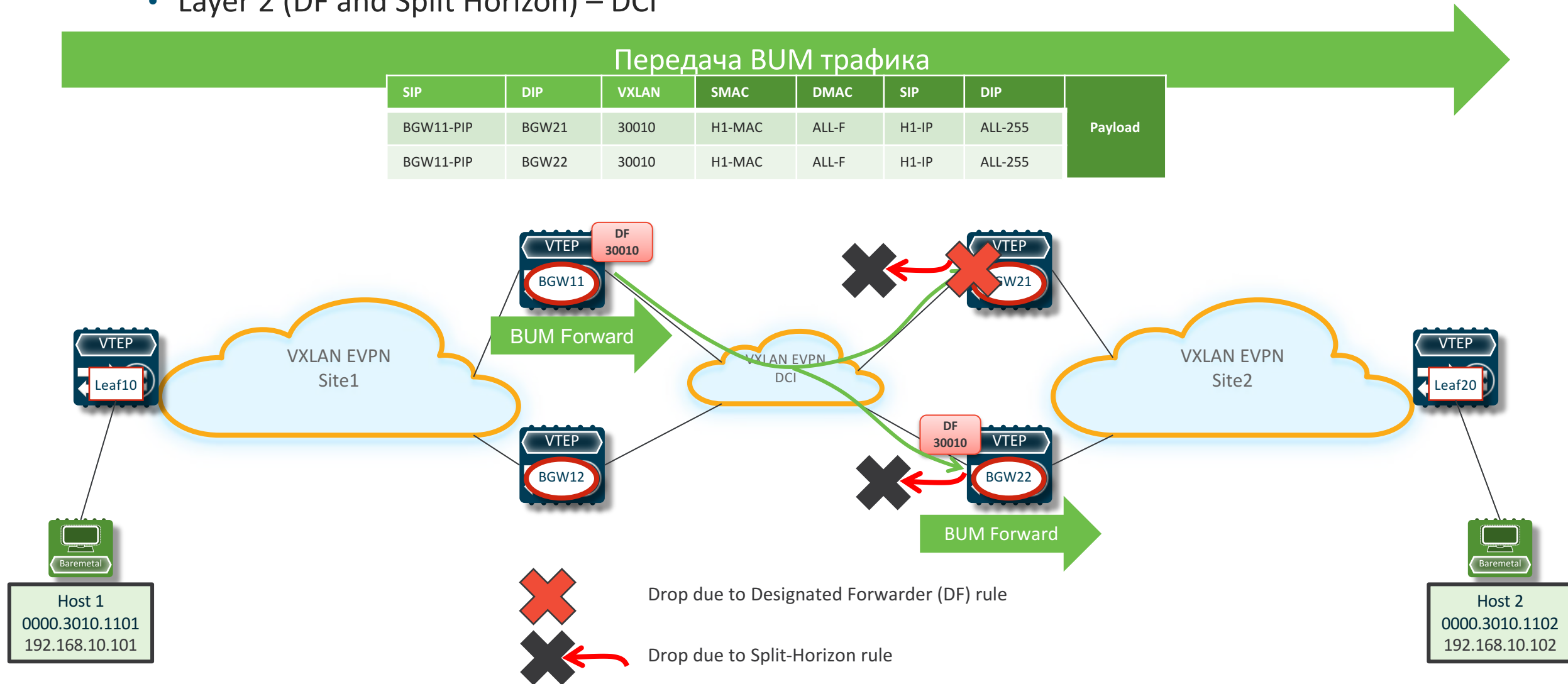


Передача пакетов в VXLAN Multi-Site

- Layer 2 (DF and Split Horizon) – DCI

Передача BUM трафика

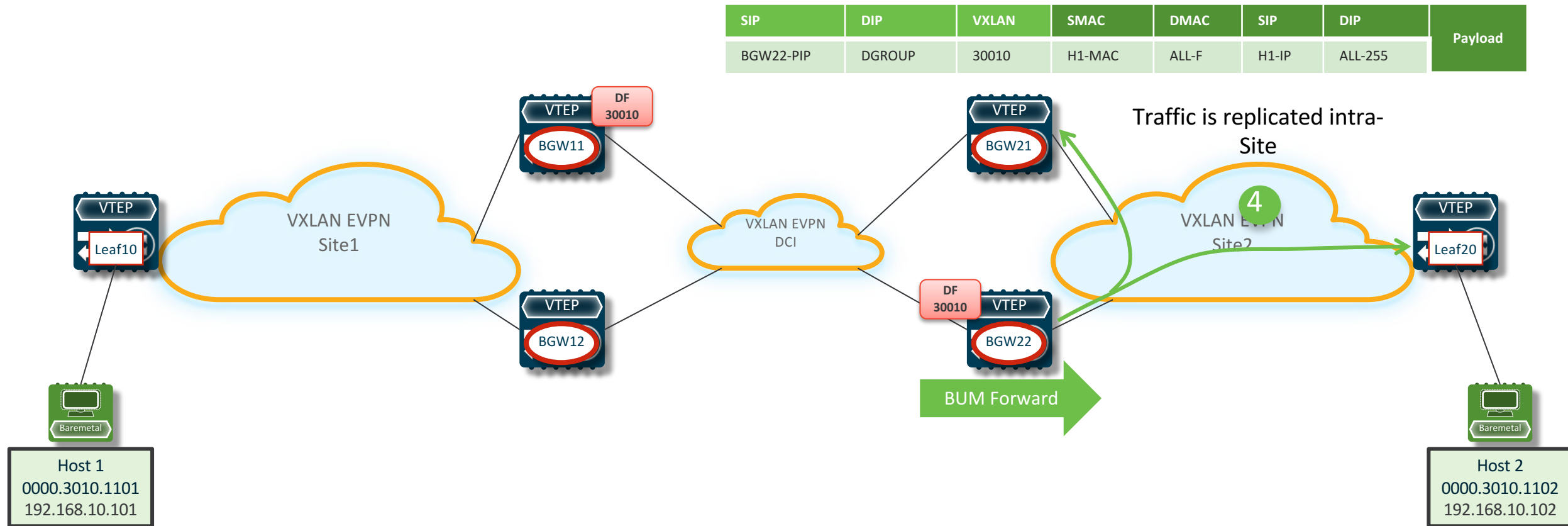
SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
BGW11-PIP	BGW21	30010	H1-MAC	ALL-F	H1-IP	ALL-255	
BGW11-PIP	BGW22	30010	H1-MAC	ALL-F	H1-IP	ALL-255	



Передача пакетов в VXLAN Multi-Site

- Layer 2 (BUM) – Site 2

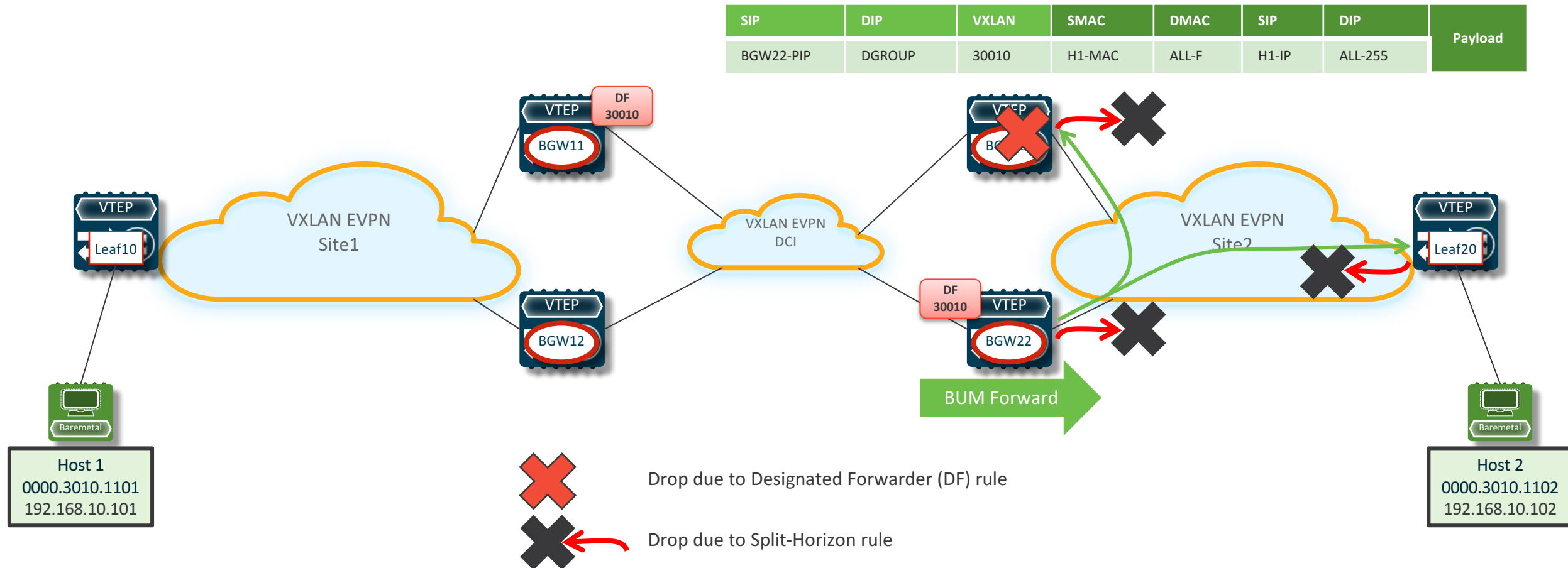
Передача BUM трафика



Передача пакетов в VXLAN Multi-Site

- Layer 2 (DF and Split Horizon) – Site 2

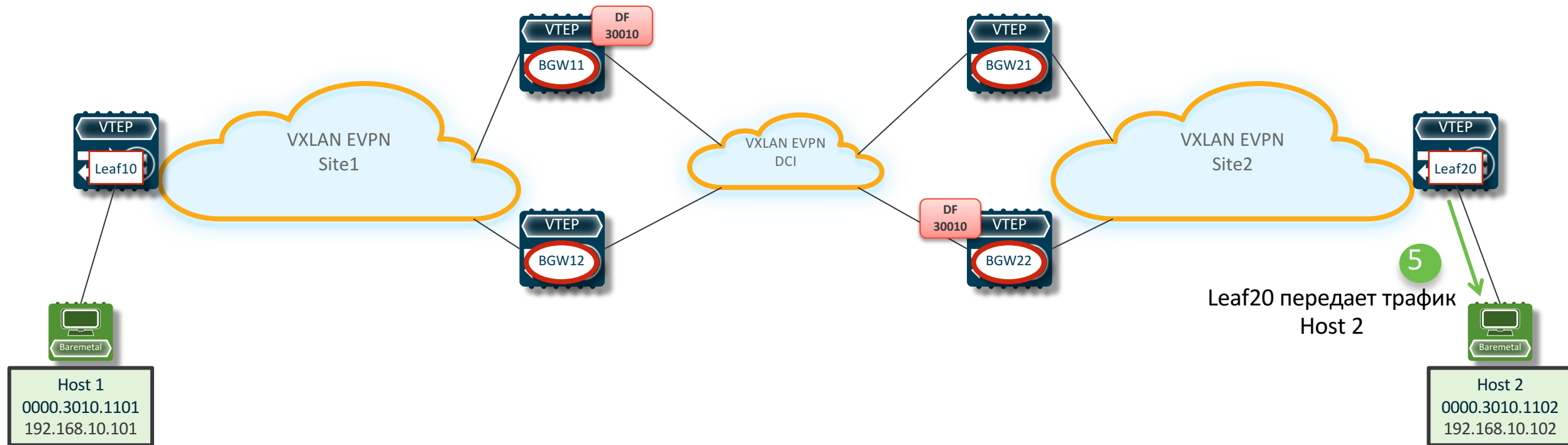
Передача BUM трафика



Передача пакетов в VXLAN Multi-Site

- Layer 2 (BUM) – Site 2

Передача BUM трафика



Растягивание L3 между
сайтами

Настройка L3-сервисов в Multi-Site архитектуре при помощи DCNM и MSO



Настройка L3-сервисов в Multi-Site архитектуре при помощи DCNM и MSO

- 0 Сайт № 1 - проверка L3-настроек, выполненных на DCNM
- 0 Сайт № 1 – проверка взаимодействия внутри VRF между сегментами
- 1 Создание схемы в MSO
- 2 Создание шаблона в MSO
- 3 Добавление сайтов в шаблон
- 4 Импорт настроек vrf из DCNM в MSO
- 5 Сайт № 1 – растягивание vrf на BGW
- 6 Сайт №2 – растягивание vrf на все устройства сайта
- 7 Сайт № 2 - проверка настройки vrf на DCNM

Настройка L3-сервисов в Multi-Site архитектуре при помощи DCNM и MSO

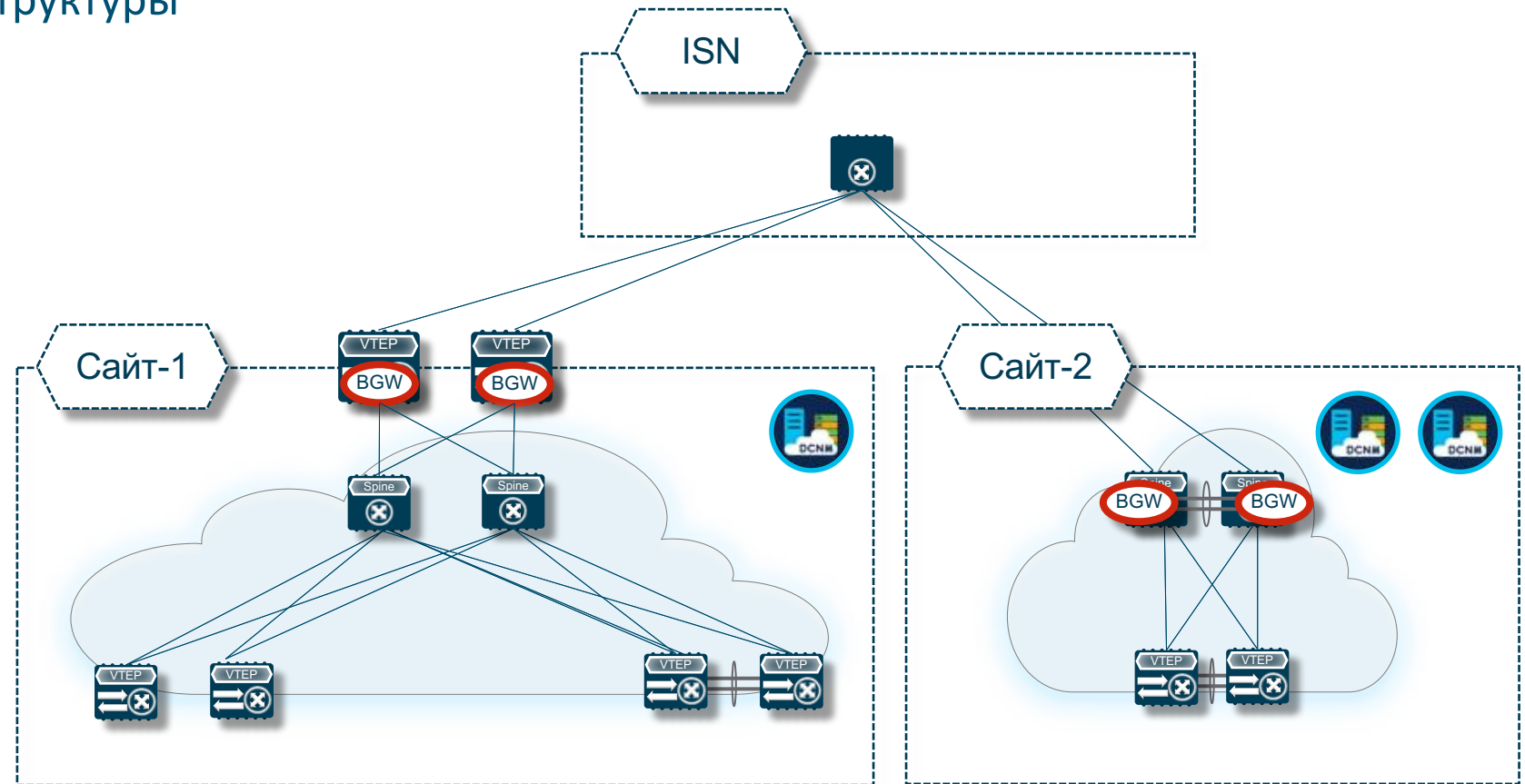
- 8 Сайт № 2 - создание локального сегмента
- 9 Создание loopback на всех BGW при помощи шаблона настроек DCNM
- 10 Проверка связанности внутри растянутого vrf между сайтами

Демонстрация

Исходное состояние инфраструктуры

0

На сайте № 1 настроены VRF и сетевые сегменты



vrf: s1_inside

VLAN 2200



10.100.0.2/24

VLAN 2201



10.100.1.2/24

- Dashboard
- Topology
- Control
- Monitor
- Administration
- Applications

Hello, admin!
Let's get started.



DCNM Licenses

License this copy of DCNM for each managed switch to unlock Performance Collection.



Fabric Builder

Creates a managed and controlled SDN fabric.



Networks & VRFs

Simple network overlay provisioning for N9K VXLAN EVPN Fabrics.



Documentation

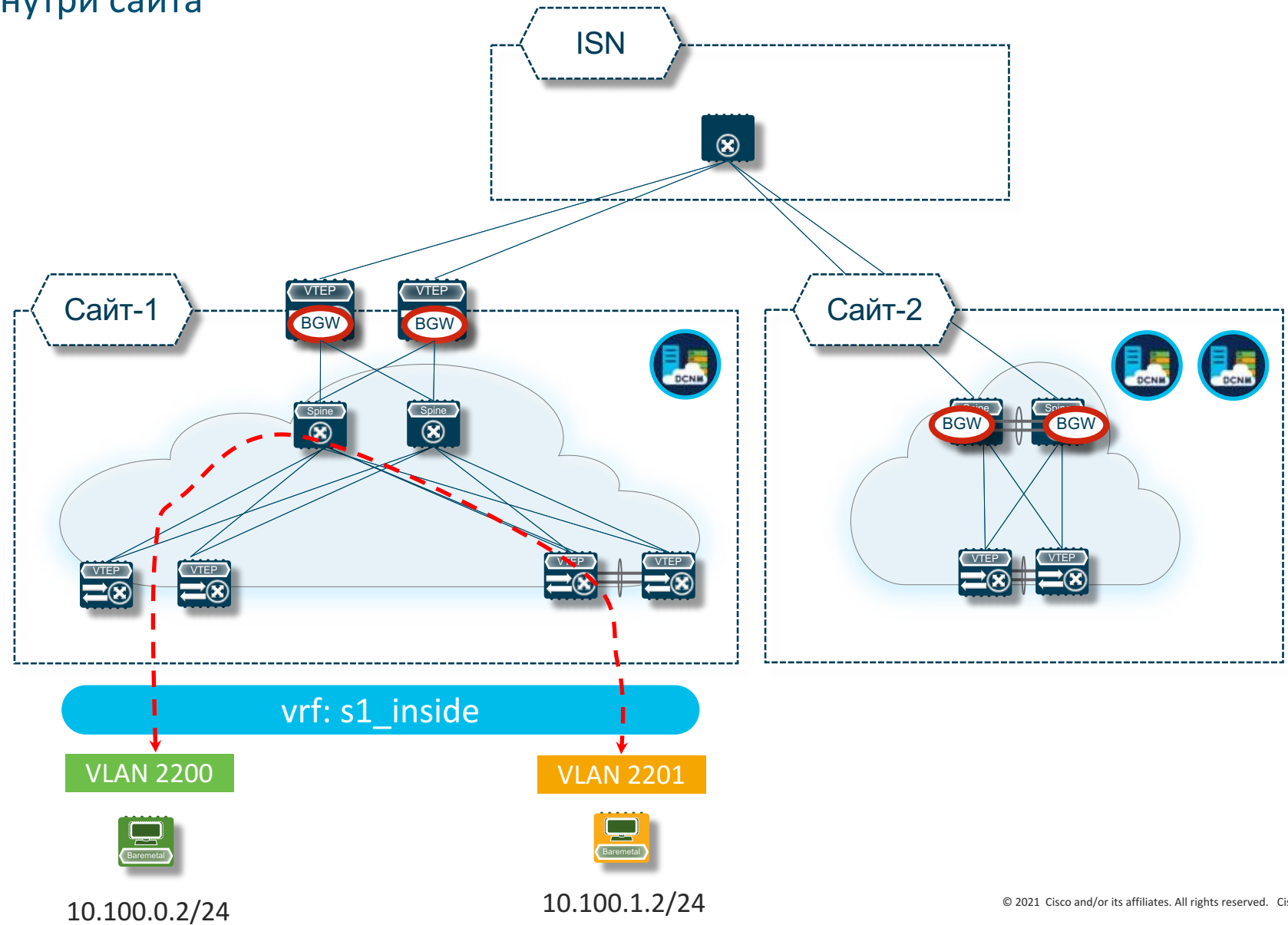
Access cisco.com from documentation on configuration, maintenance and operation.

Демонстрация

Взаимодействие устройств внутри сайта

0

Устройства внутри сайта взаимодействуют друг с другом



Dashboard

Topology

Control

Monitor

Administration

Applications

SCOPE: Site1

Network / VRF Selection > Network / VRF Deployment >

Network View | Continue

Fabric Selected: Site1

Выбрано 1 / Всего 2

Показать Все

☐	VRF Name	VRF ID	Status
☐	S1_DMZ	51002	NA
☒	s1_inside	51001	DEPLOYED

Демонстрация

Настройка схемы, шаблона и импорт vrf из DCNM

1

Создание схемы в MSO

2

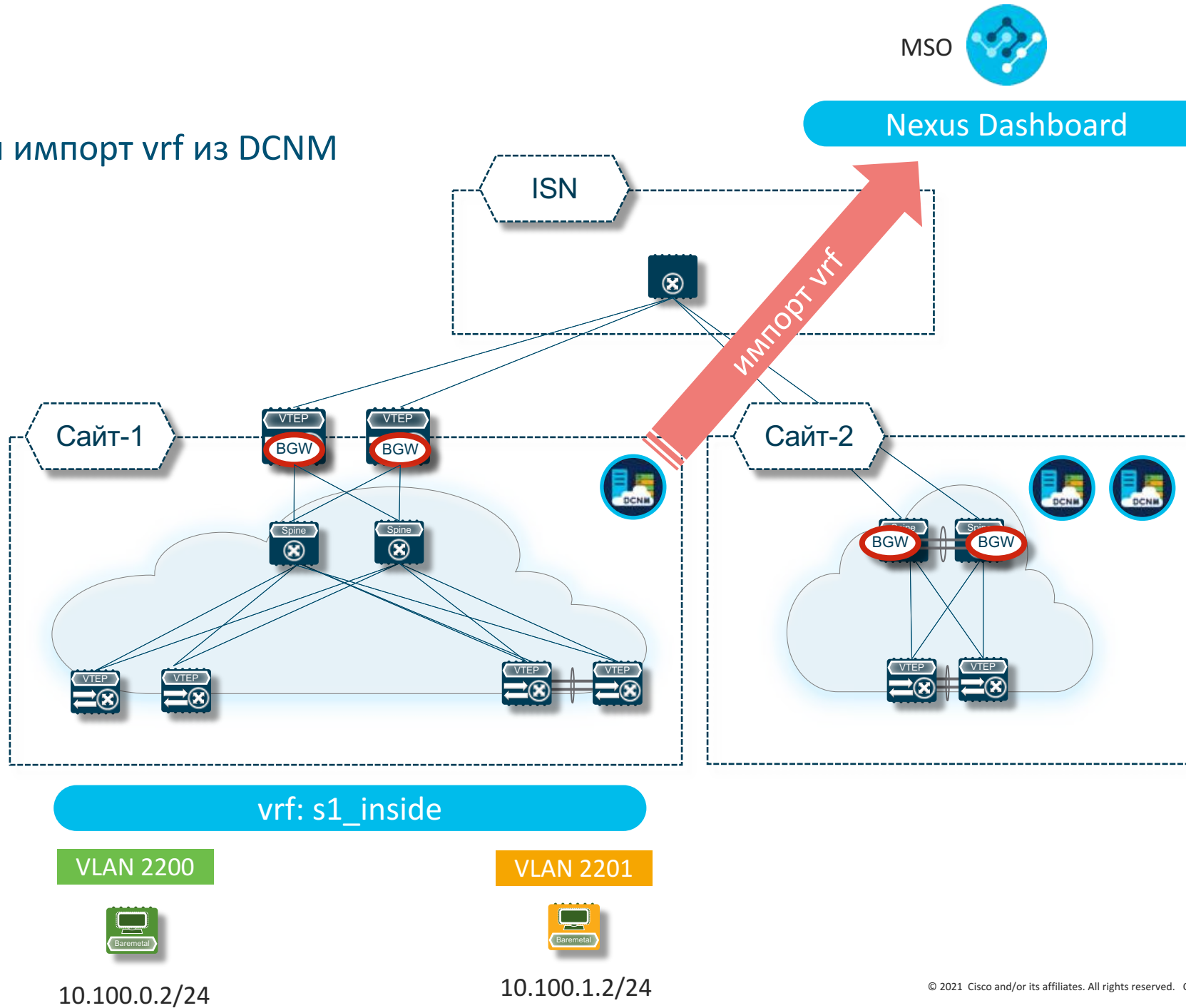
Создание шаблона в MSO

3

Добавление сайтов в шаблон

4

Импорт настроек vrf из DCNM в MSO



Nexus Dashboard

Version 2.0(1d)



Username

admin

Password

••••••••

Login

Демонстрация

Растягивание VRF с одного сайта на другой

5

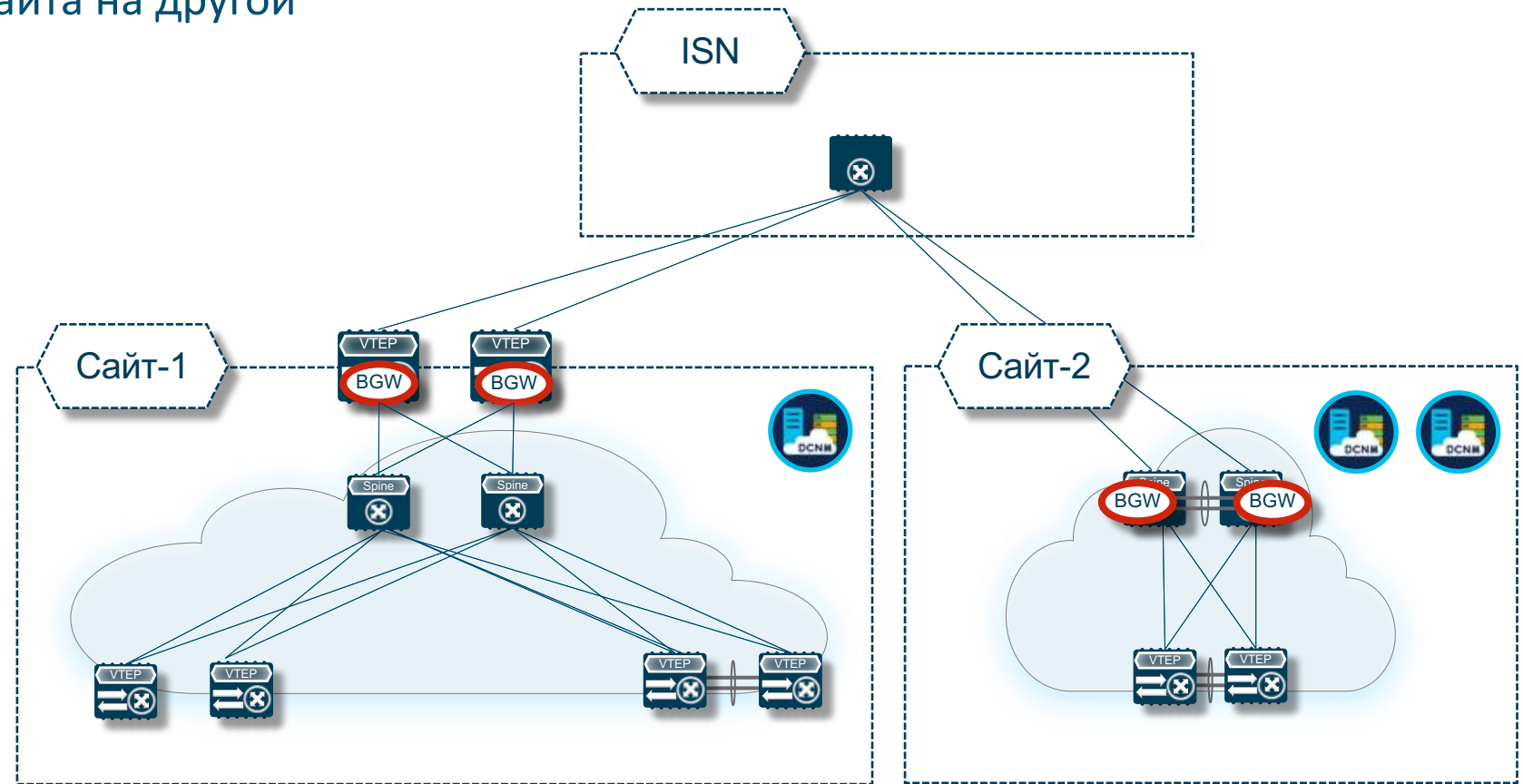
Сайт № 1 - “растягиваем”
vrf на BGW*

6

Сайт № 2 - “растягиваем”
vrf все устройства

7

Сайт № 2 – проверяем
настройку vrf на DCNM



vrf: s1_inside

VLAN 2200



10.100.0.2/24

VLAN 2201



10.100.1.2/24

растягивание vrf

*Растягивание VRF на BGW не работает в DCNM пока не создан объект описывающий подключение к другому сайту

L3 Import

TEMPLATES

L3 import

SITES

Site1 (DCNM) 11.5(1)

L3 import

Site2 (DCNM) 11.5(1)

L3 import

Site1
L3 Import
Tenant: dcnm-default-tn

FILTERS

VRFs

s1_inside

Networks

VRF s1_inside

0 0 0 0

Template Properties

* Display Name

s1_inside

Deployed Name: s1_inside

VRF ID ⓘ

51001

Site Local Properties

Tenant Routed Multicast

☐

RP External

☐

Static Leaf Nodes

Node/Switch

ServerLeaf1_Site1 - ServerLeaf2_Site1

VLAN: 2002

BorderLeaf3_Site1 - BorderLeaf4_Site1

VLAN: 2002

+ Add Static Leaf

Демонстрация

Создание сегмента на 2-м сайте

8

Создаем L2 сегмент на 2-м сайте

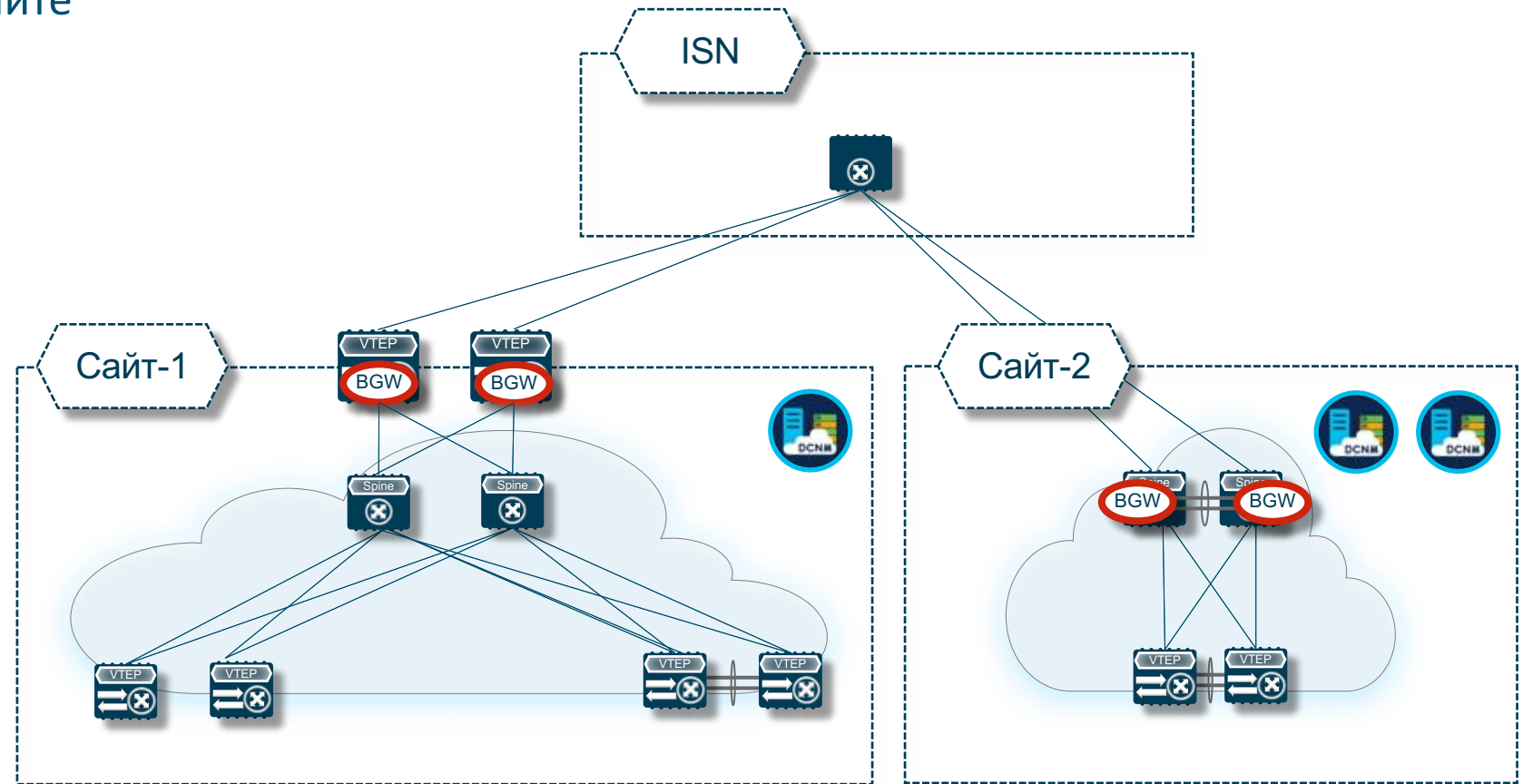
Шаблон

Привязка к сайту

Создание сегмента с
Anycast GW

Настройка статичных
портов с локальным
VLAN

Растягивание сегмента
на BGW



vrf: s1_inside

VLAN 2200



10.100.0.2/24

VLAN 2201



10.100.1.2/24

VLAN 2301



10.200.1.2/24

L3 Import

TEMPLATES

L3 Import

SITES

Site1 (DCNM) 11.5(1)

L3 import

Site2 (DCNM) 11.5(1)

L3 import



L3 import

Applied to 2 sites
Tenant: dcnm-default-tn

Last Deployed: Apr 6, 2021 12:46 am

Deploy to sites

FILTERS

IMPORT ▾

SELECT

+ CREATE OBJECT



VRFs ▾

s1_inside



Networks

TEMPLATE
L3 import

Template Settings

* Display Name

L3 import

Template Type
Networking

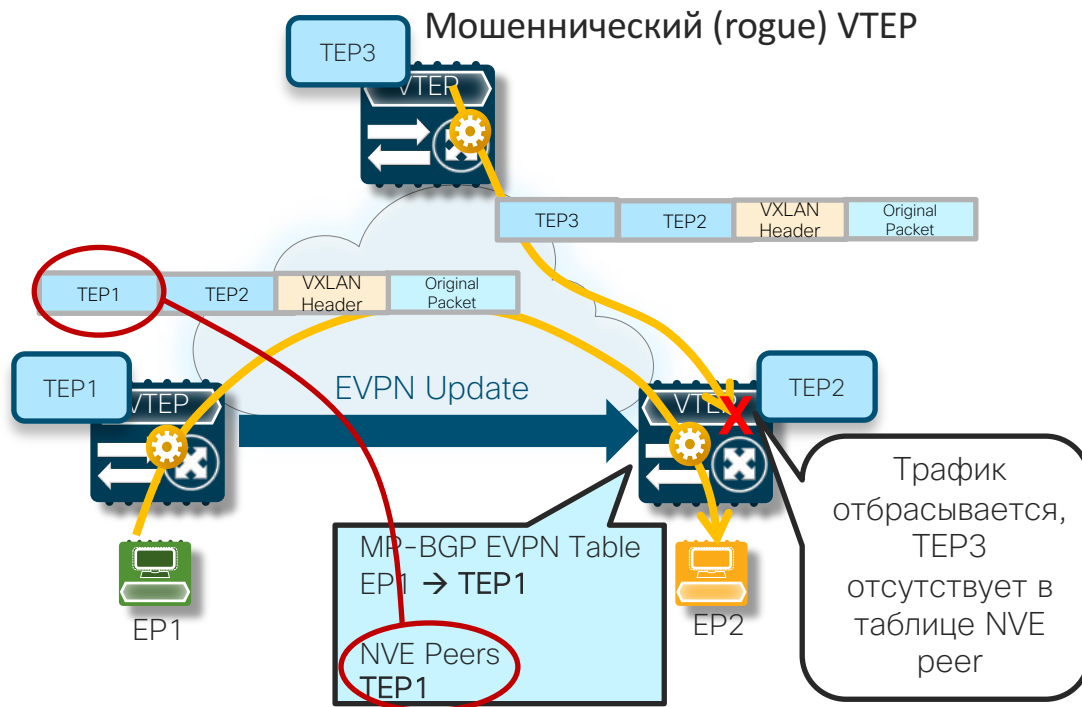
Tenant Settings

Display name
dcnm-default-tnName
dcnm-default-tnDescription
Default tenant for DCNM sites

Отдельные соображения по
растягиванию L3 между
фабриками

VXLAN EVPN

Функции безопасности, встроенные в плоскость передачи данных



- VXLAN EVPN предлагает функцию безопасности встроенную в плоскость передачи данных
 - VXLAN трафик проходящий от удаленного VTEP принимается и обрабатывается только если они пришел с адреса, который присутствует в списке "NVE peer"
 - Адрес добавляется в таблицу NVE peer's после получения MP-BGP EVPN апдейта, в котором этот адрес используется как next-hop¹¹⁴
- Предотвращает добавление мошеннических (rogue) VTEP в фабрику VXLAN EVPN

Почему эта функция безопасности
может быть проблемой при
организации растягивания L3 VRF
между сайтами?

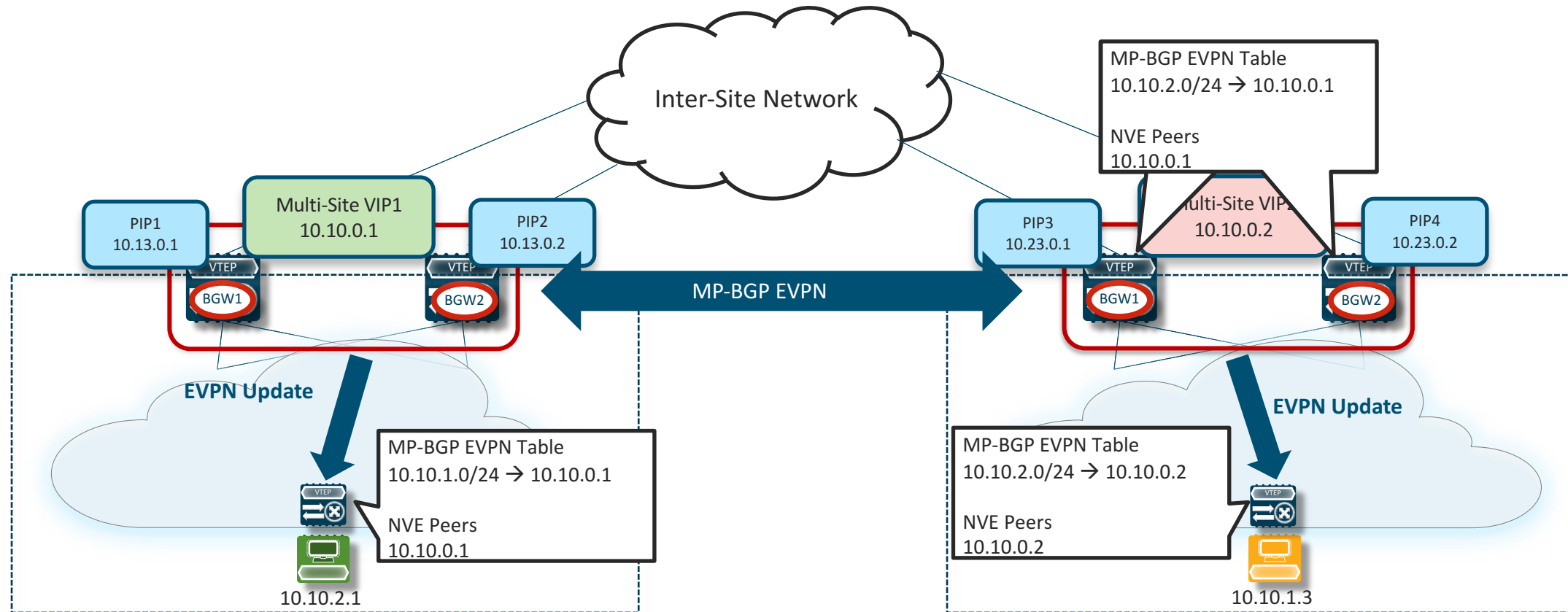
115

VXLAN Multi-Site

Inter-Site Layer 3 трафик – плоскость управления

- Inter-Site Type-2 и Type-5 EVPN апдейты всегда используют адрес локального Multi-Site VIP как адрес next-hop

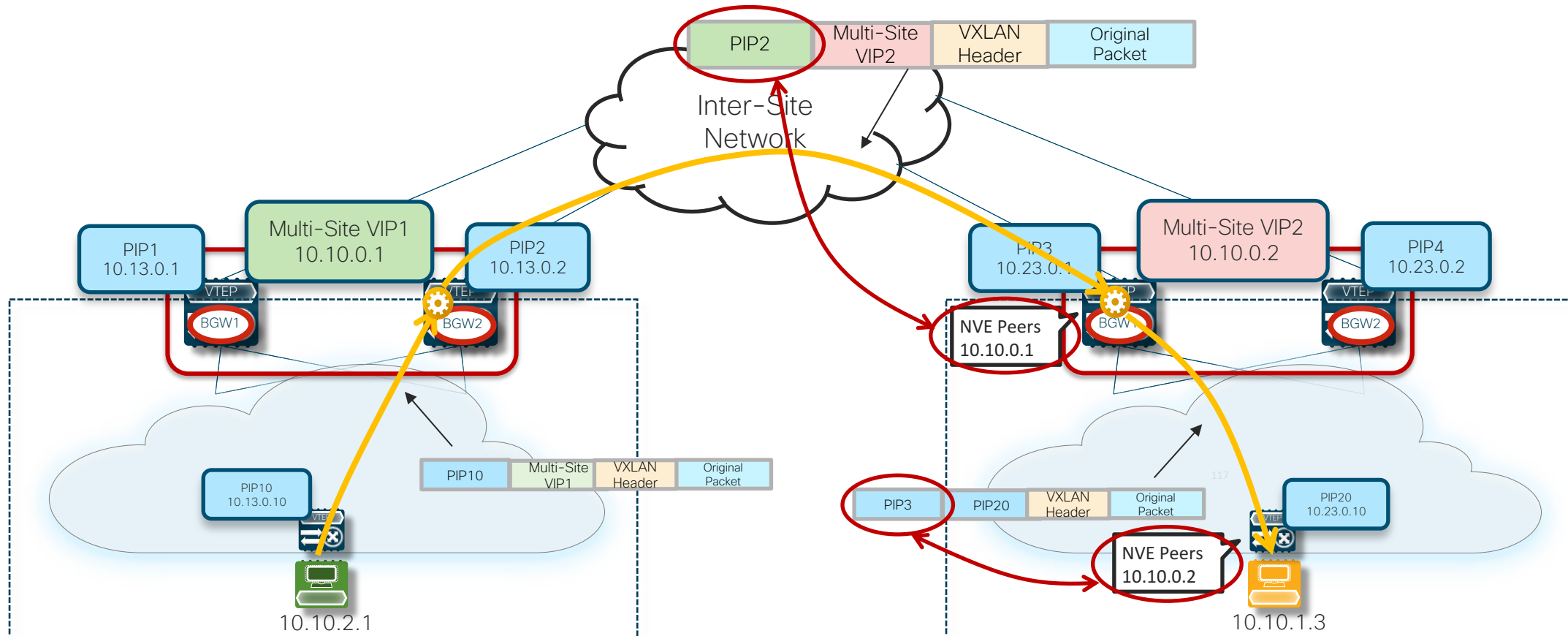
Исключение составляют Type-5 апдейты для L3 сетей, которые подключены напрямую к BGW



VXLAN Multi-Site

Inter-Site Layer 3 Traffic – плоскость передачи данных

- Для передачи Inter-Site Layer 3 трафика в качестве адреса источника всегда используется локальный адрес BGW PIP
- Это же справедливо и для трафика внутри сайта между локальным BGW и leaf коммутатором



Как решить эту проблему?

VXLAN Multi-Site

Решение проблемы с дропами Inter-Site Layer 3 трафика

- Вариант 1: растянуть хотя бы один L2VNI End-to-End
L2VNI должен быть растянут между leaf и BGW
L2VNI должен быть растянут между сайтами (i.e. defined on the local and remote BGW nodes)
Это не обязательно должен быть один и тот же L2VNI

- Вариант 2: включить функцию “advertise-pip” на всех BGW узлах

```
router bgp 65001
address-family l2vpn evpn
advertise-pip
```

на всех BGW узлах

- Вариант 3: создать на всех BGW узлах loopback interface (в специфическом VRF) и анонсировать информацию о нем между сайтами при помощи type-5 EVPN update

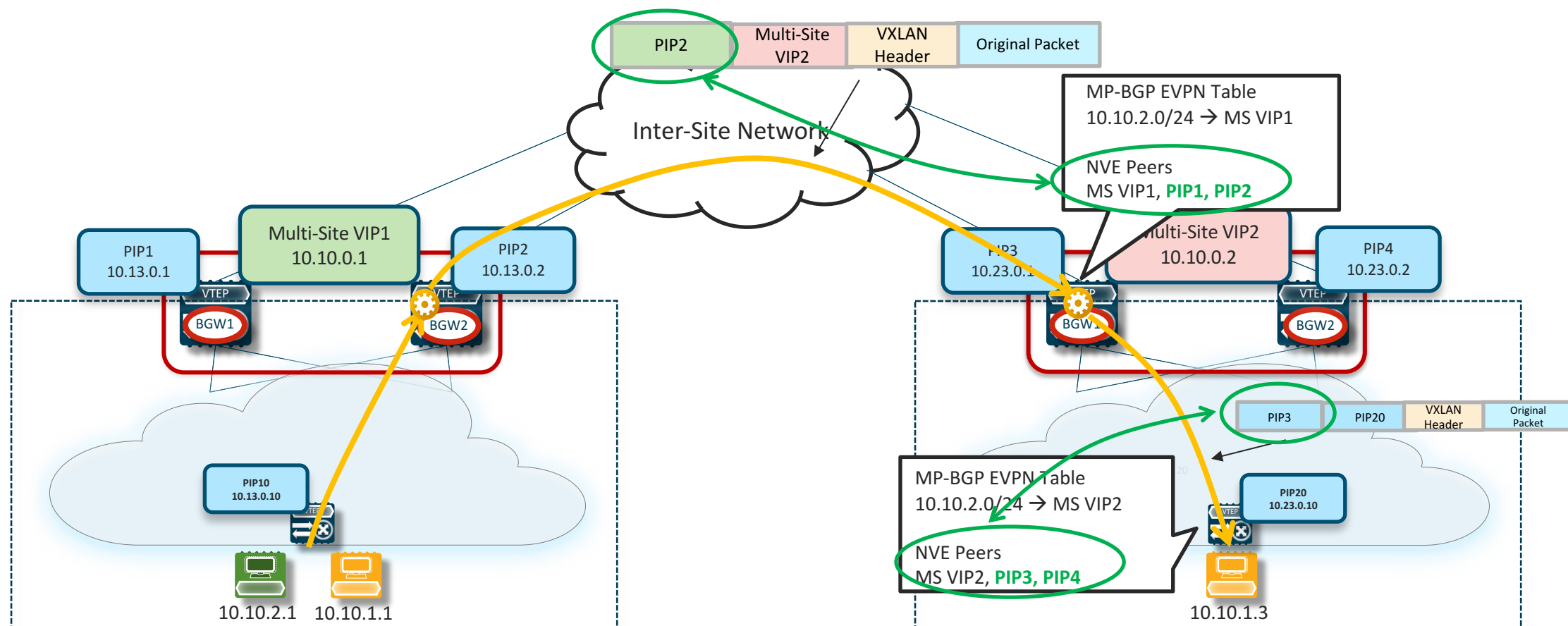
```
route-map fabric-rmap-redist-subnet permit 10
match tag 12345
!
interface loopback2
vrf member t1-vrf1
ip address 33.33.33.1/32 tag 12345
!
router bgp 65001
vrf t1-vrf1
address-family ipv4 unicast
redistribute direct route-map fabric-rmap-redist-subnet
```

на всех BGW узлах

Решение проблемы с дропами Inter-Site Layer 3 трафика

Установка адреса PIP в таблицу NVE Peers

- Применение любого из 3-х приведенных решений обеспечивает установку адреса PIP на BGW в таблицу NVE как на локальных так и на удаленных BGW узлах

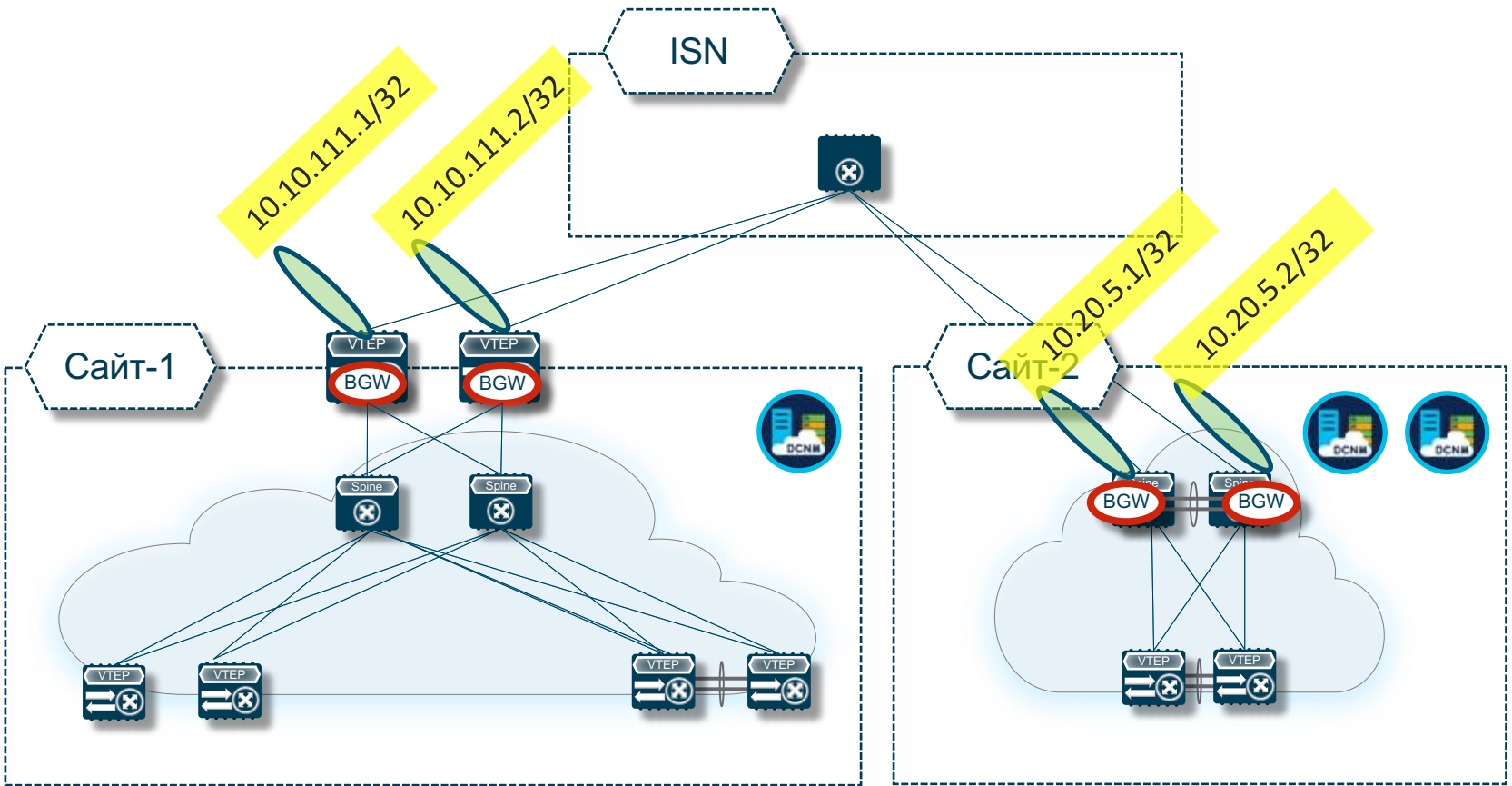


Демонстрация

Создание loopback-ов

9

Создаем loopback интерфейсы при помощи шаблона настроек



vrf: s1_inside

VLAN 2200



10.100.0.2/24

VLAN 2201



10.100.1.2/24

VLAN 2301



10.200.1.2/24

Network Manager

admin

Dashboard

Topology

Control

Monitor

Administration

Applications

Hello, admin!

Let's get started.



DCNM Licenses

License this copy of DCNM for each managed switch to unlock Performance Collection.



Fabric Builder

Creates a managed and controlled SDN fabric.



Networks & VRFs

Simple network overlay provisioning for N9K VXLAN EVPN Fabrics.



Documentation

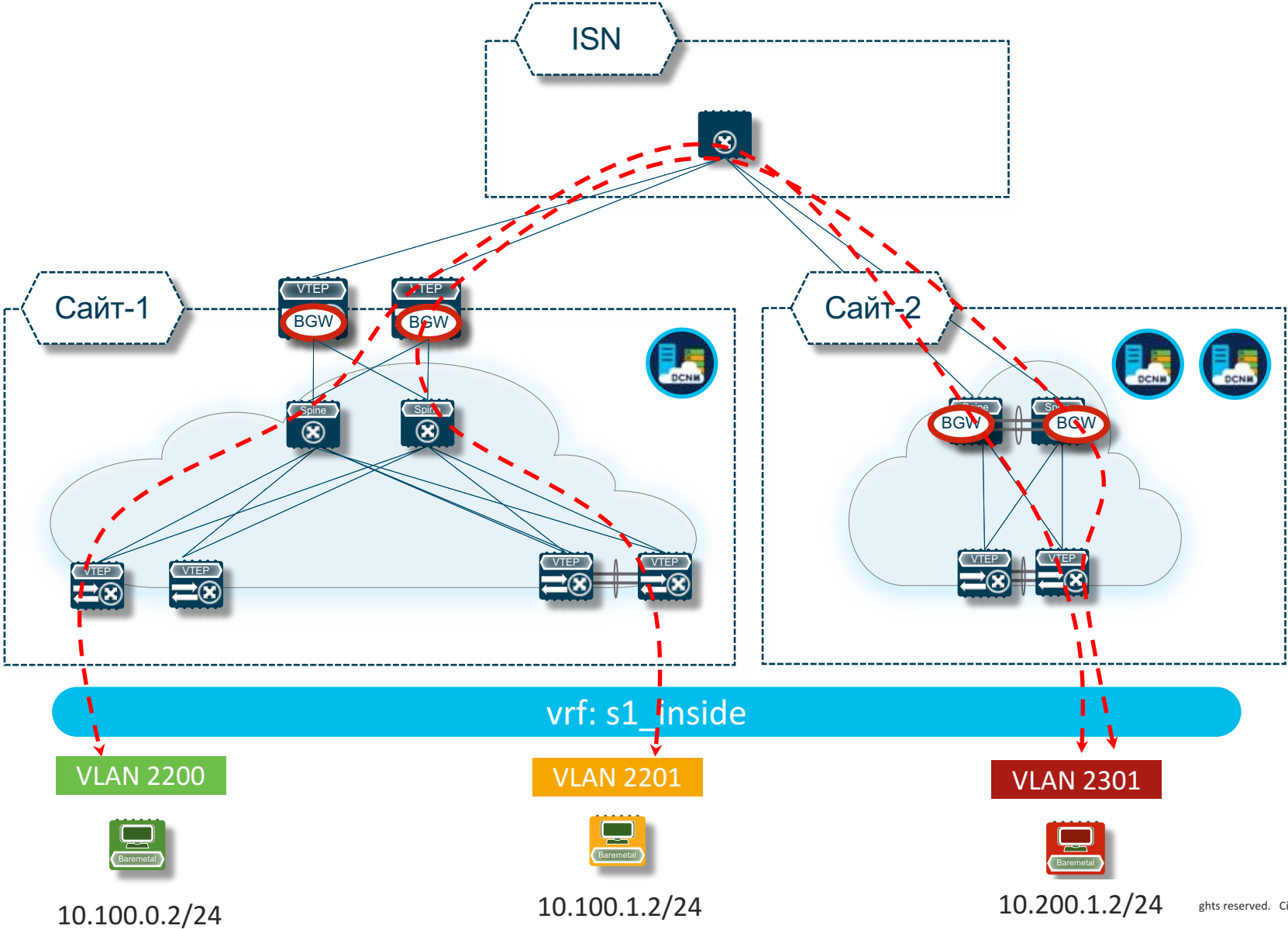
Access cisco.com from documentation on configuration, maintenance and operation.

Демонстрация

Проверка связанности

10

Проверка связанности
внутри растянутого VRF
между сайтами



vm vSphere Client Menu Search in all environments Administrator@VSPHERE.LOCAL

testhx
UTM_0.3
vCenter Server Appliance
vCSA 6.7 U3
vm-DEF-SR3
vrf.dmz.vlan.2202.2
vrf.inside.vlan.2200.2
vrf.inside.vlan.2201.2
vrf.shared.vlan.2203.2
vrf.svc.d.2251.2
vrf.svc.s.2250.2
W2K-TS
vcsa-hxnvme.moslab.cisco.com
Main DC
hxnvme-1
10.5.32.197
10.5.32.198
10.5.32.199
dc1-border-router
dc1-loadbalancer
sprint.vlan.2300
stCtVM-WZP24210ARA
stCtVM-WZP24210ARC
stCtVM-WZP24210ASA
vrf.inside.vlan.2301.2
vrf.shared.vlan.2302.2

vrf.inside.vlan.2301.2

Summary Monitor Configure Permissions Datastores Networks Updates

Powered On

Launch Web Console
Launch Remote Console

Guest OS: CentOS 7 (64-bit)
Compatibility: ESXi 6.5 and later (VM version 13)
VMware Tools: Not running, not installed
More info

DNS Name:
IP Addresses:
Host: 10.5.32.197

CPU USAGE
0 Hz
MEMORY USAGE
20 MB
STORAGE USAGE
313.96 MB

VMware Tools is not installed on this virtual machine. Install VMware Tools...

VM Hardware

CPU	1 CPU(s)
Memory	2 GB, 0.02 GB memory active
Hard disk 1	16 GB
Network adapter 1	NX-OS-2301 (connected)
CD/DVD drive 1	Disconnected
Video card	8 MB
VMCI device	Device on the virtual machine PCI bus that provides support for the virtual machine communication interface
Other	Additional Hardware

Notes
Edit Notes...

Custom Attributes

Attribute	Value
No items to display	

Recent Tasks Alarms

Task Name	Target	Status	Details	Initiator	Queued For	Start Time	Completion Time	Server
Running								

More Tasks

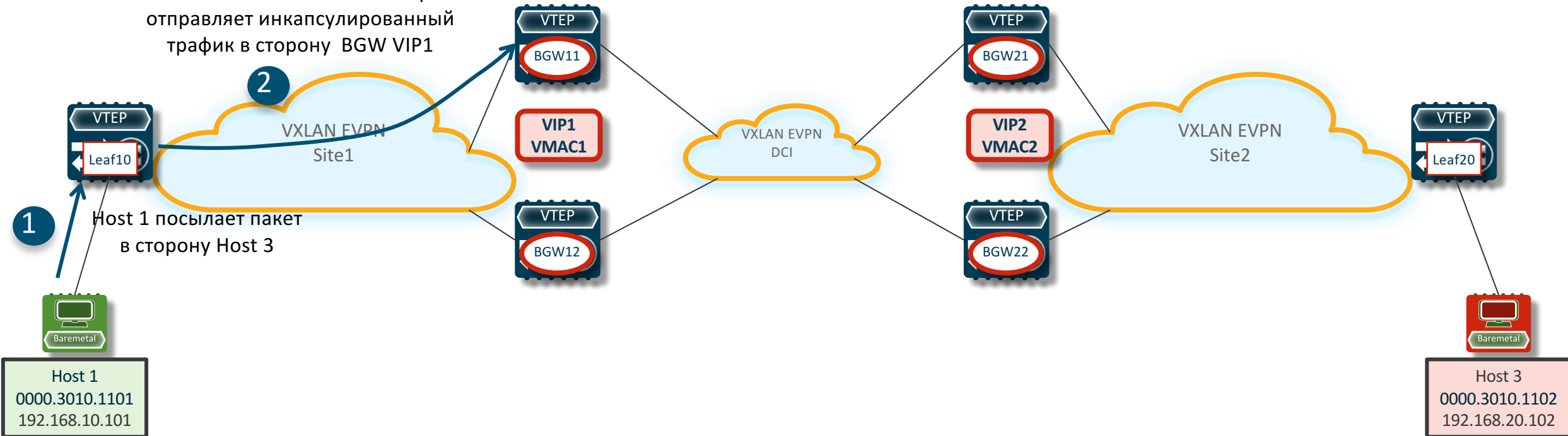
Передача пакетов между
сайтами (маршрутизация)

Передача пакетов в VXLAN Multi-Site Layer 3 (Host 1 в сторону Host 3) – сайт 1

маршрутизация

SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
L10	BGW-VIP1	50001	L10-MAC	BGW-VMAC1	H1-IP	H3-IP	

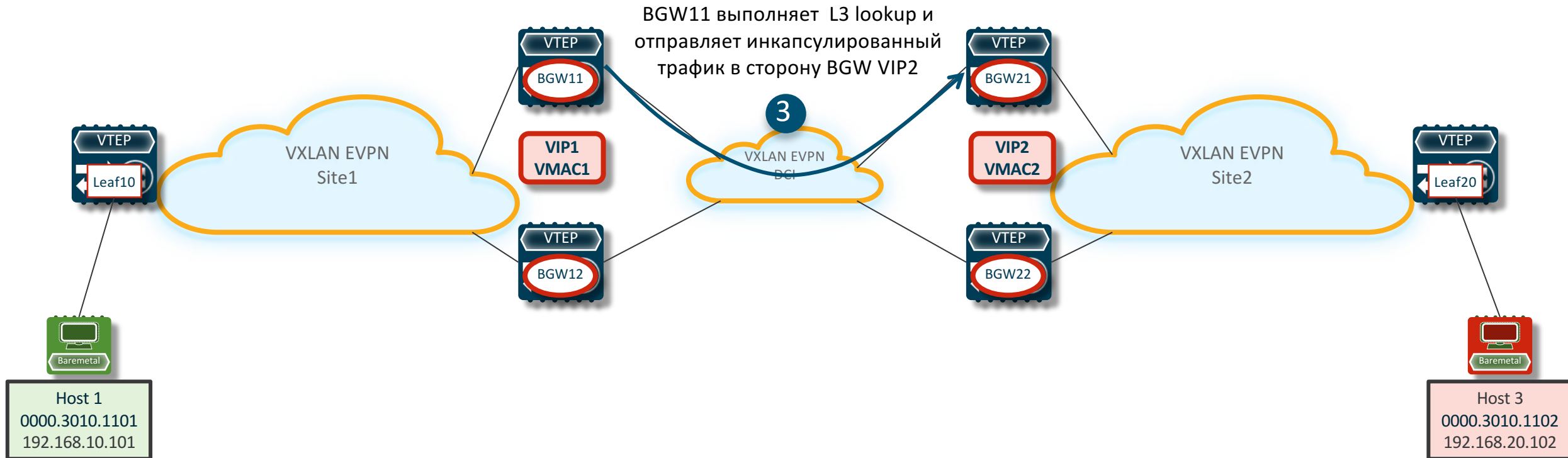
Leaf10 выполняет L3 lookup и отправляет инкапсулированный трафик в сторону BGW VIP1



Передача пакетов в VXLAN Multi-Site Layer 3 (Host 1 в сторону Host 3) – DCI

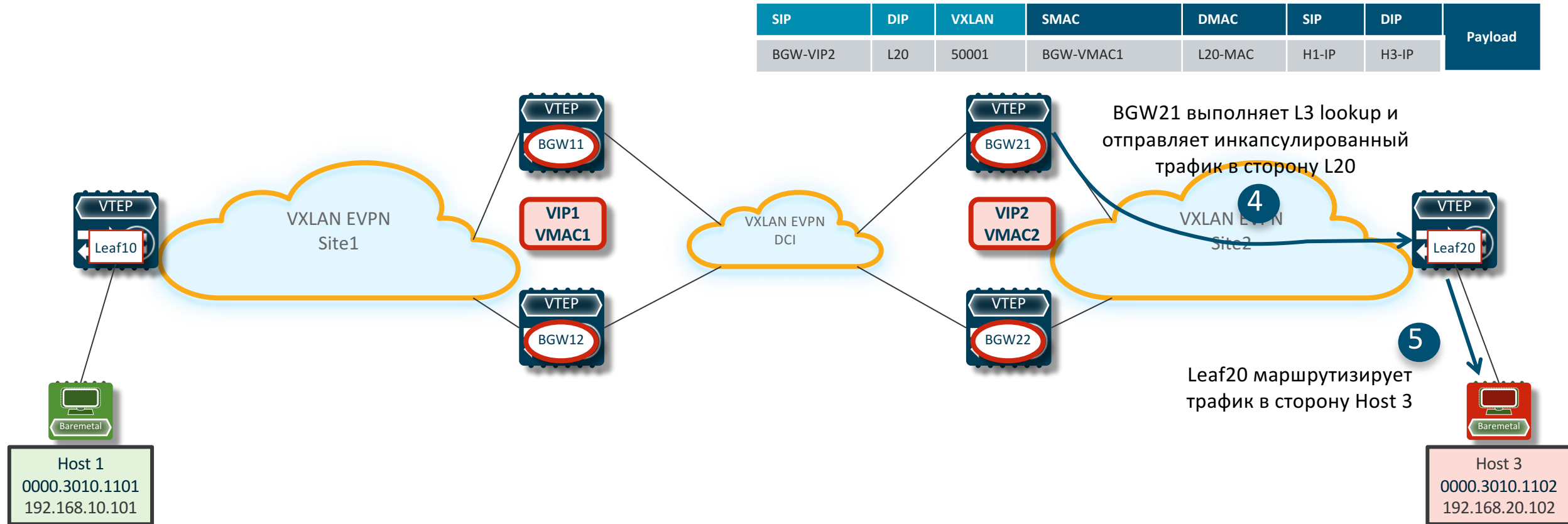
Маршрутизация

SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
BGW-VIP1	BGW-VIP2	50001	BGW-VMAC1	BGW-VMAC2	H1-IP	H3-IP	



Передача пакетов в VXLAN Multi-Site Layer 3 (Host 1 в сторону Host 3) – сайт 2

Маршрутизация



Растягивание L2 между
сайтами

Настройка L2-сервисов в Multi-Site архитектуре при помощи DCNM и MSO



Демонстрация

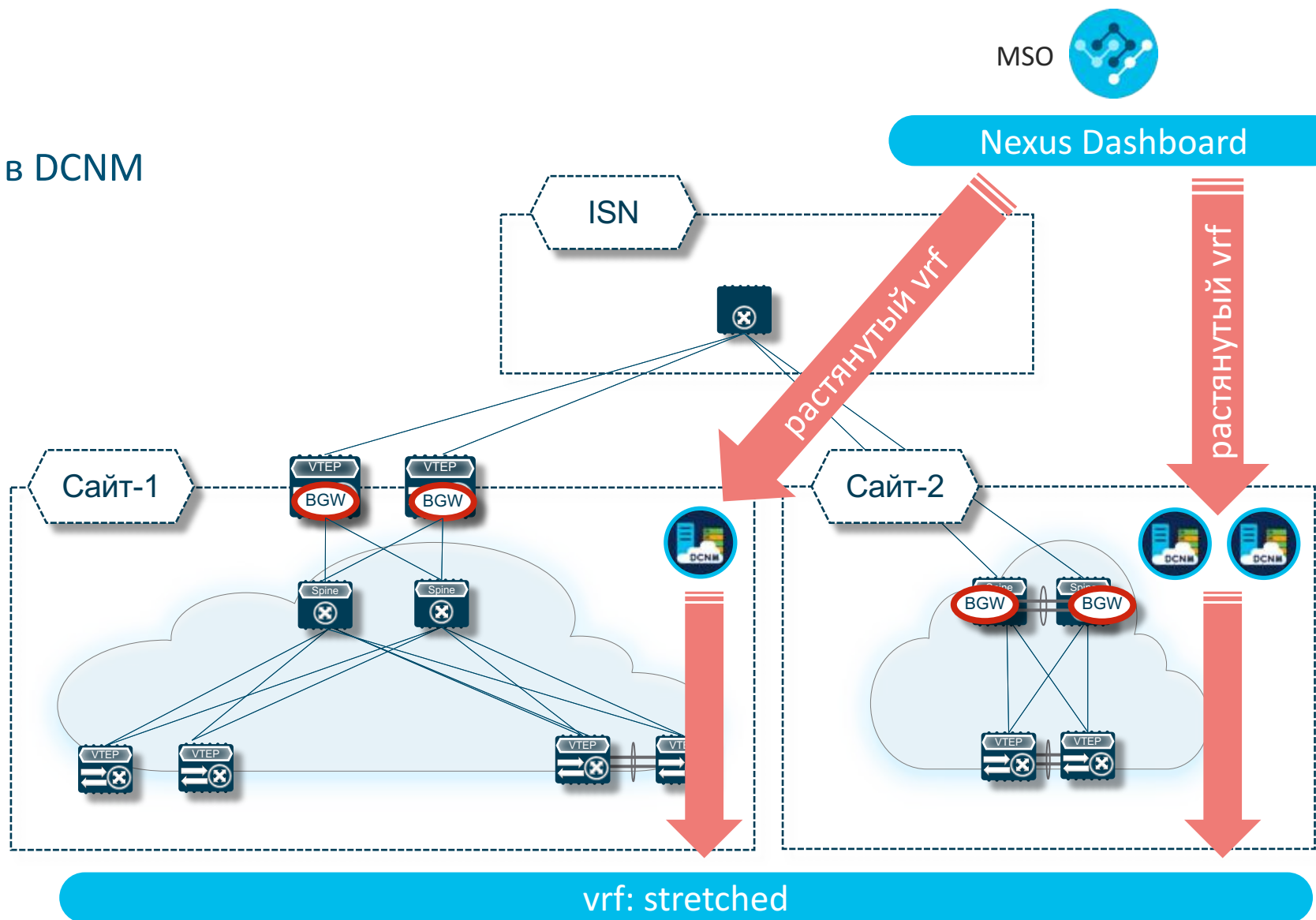
Настройка схемы и шаблона в DCNM

1

Создание схемы в MSO

Создание шаблона в MSO

Добавление сайтов в шаблон



Nexus Dashboard

Version 2.0(1d)



Username

admin

Password

••••••••

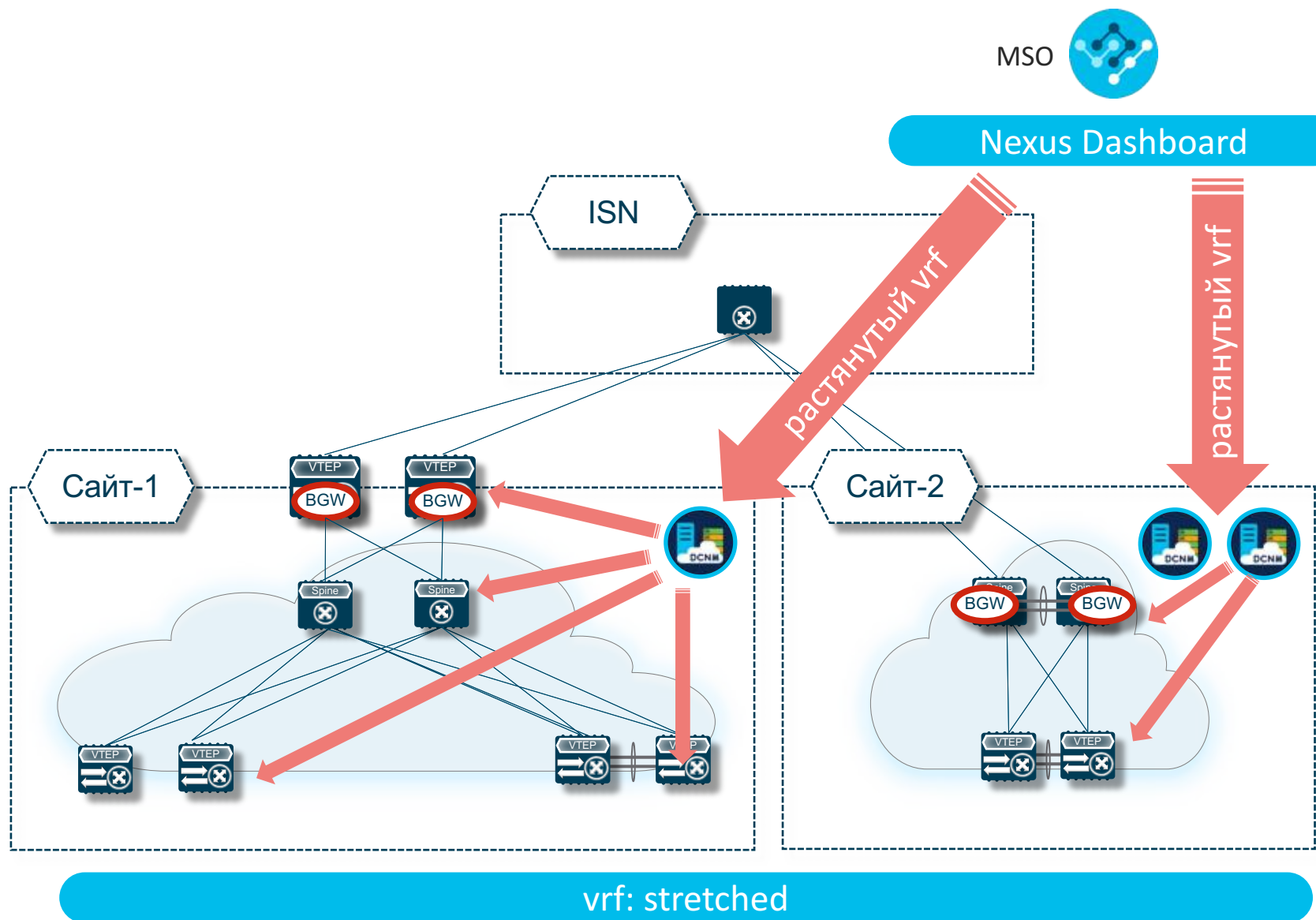
Login

Демонстрация

Растягивание VRF

2

Растягивание vrf по всем устройствам каждой из фабрик



L2 Stretched

0 Policies

☒ Autosave

Save

☆

↺

↻

✕

L2 Stretched

TEMPLATES

+

S1, S2

...

SITES

+

Site1 (DCNM) 11.5(1)

^

S1, S2

⚠

Site2 (DCNM) 11.5(1)

^

S1, S2

⚠

T

S1, S2

Applied to 2 sites

Tenant: dcnm-default-tn

Deploy to sites

FILTERS

IMPORT ▾

SELECT

+ CREATE OBJECT

VRF

VRFs

N

Networks

T

TEMPLATE

S1, S2

✕

Template Settings

^

* Display Name

S1, S2

Template Type

Networking

Tenant Settings

^

Display name

dcnm-default-tn

Name

dcnm-default-tn

Description

Default tenant for DCNM sites

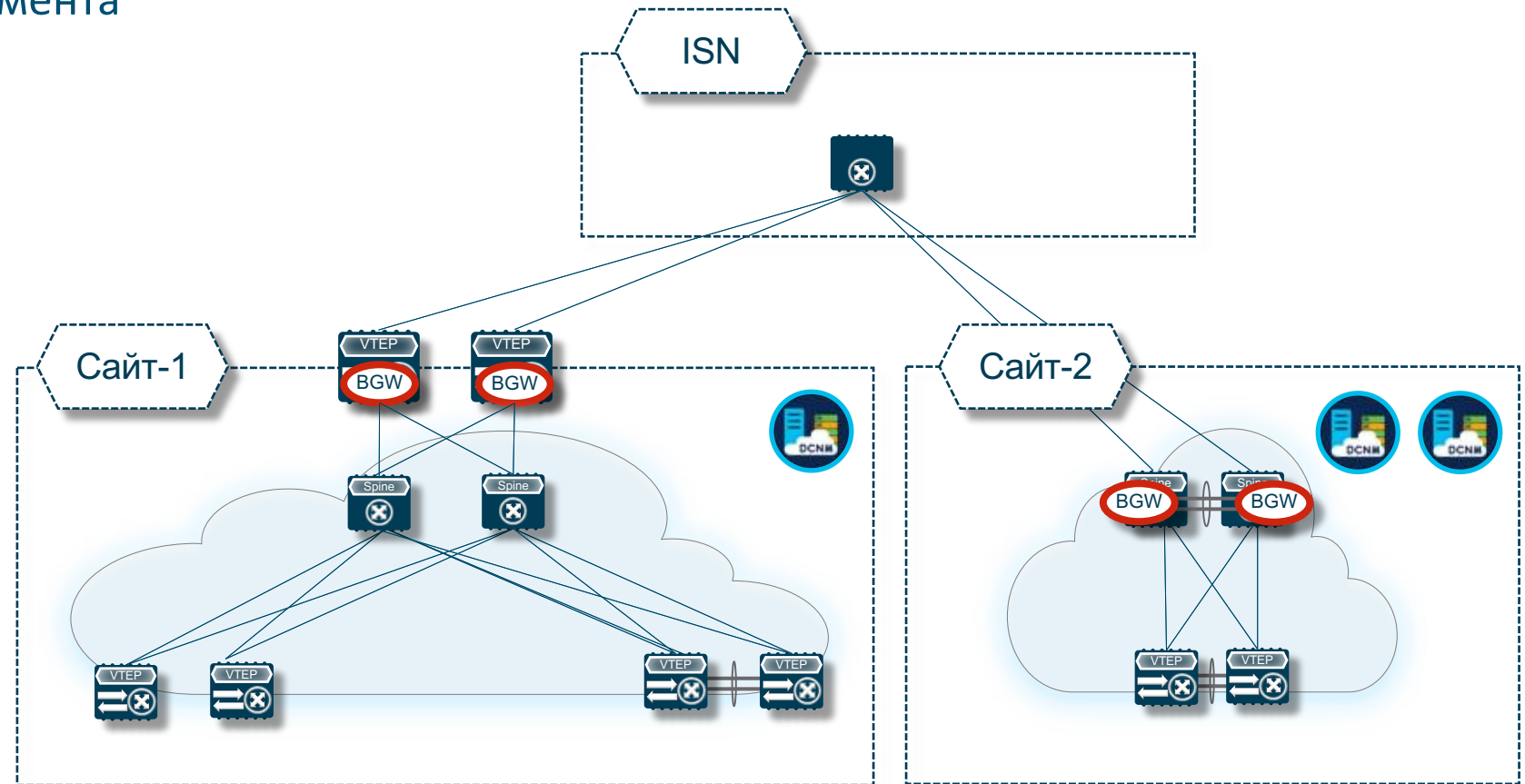
Демонстрация

Создание растянутого L2-сегмента

2

Создание растянутого L2 сегмента

Настройка уникальных для сайта encaps-VLAN



VLAN 2203



10.12.12.2

VLAN 2302



10.12.12.2

L2 Stretched

1 Policy ☒ Autosave Save ☆ ↺ ↻ ✕

L2 Stretched

TEMPLATES +

S1, S2 ...

SITES +

Site1 (DCNM) 11.5(1) ^

S1, S2 ⚠

Site2 (DCNM) 11.5(1) ^

S1, S2 ⚠

T S1, S2

Applied to 2 sites
Tenant: dcnm-default-tn

Deploy to sites

FILTERS

IMPORT ▼ SELECT + CREATE OBJECT

VRF VRFs ▼

Stretched

N Networks

T TEMPLATE S1, S2 ✕

Template Settings ^

* Display Name

S1, S2

Template Type

Networking

Tenant Settings ^

Display name

dcnm-default-tn

Name

dcnm-default-tn

Description

Default tenant for DCNM sites

© 2021 Cisco and/or its affiliates. All rights reserved. Cisco Public

Проверка растянутого L2-сегмента на DCNM

3

Проверка созданного L2 сегмента на каждом DCNM

Nexus Dashboard



VLAN 2203

10.12.12.2

VLAN 2302

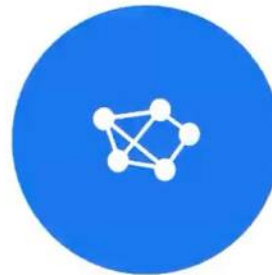
10.12.12.2



Hello, admin!
Let's get started.



DCNM Licenses
License this copy of DCNM for each managed switch to unlock Performance Collection.



Fabric Builder
Creates a managed and controlled SDN fabric.



Networks & VRFs
Simple network overlay provisioning for N9K VXLAN EVPN Fabrics.



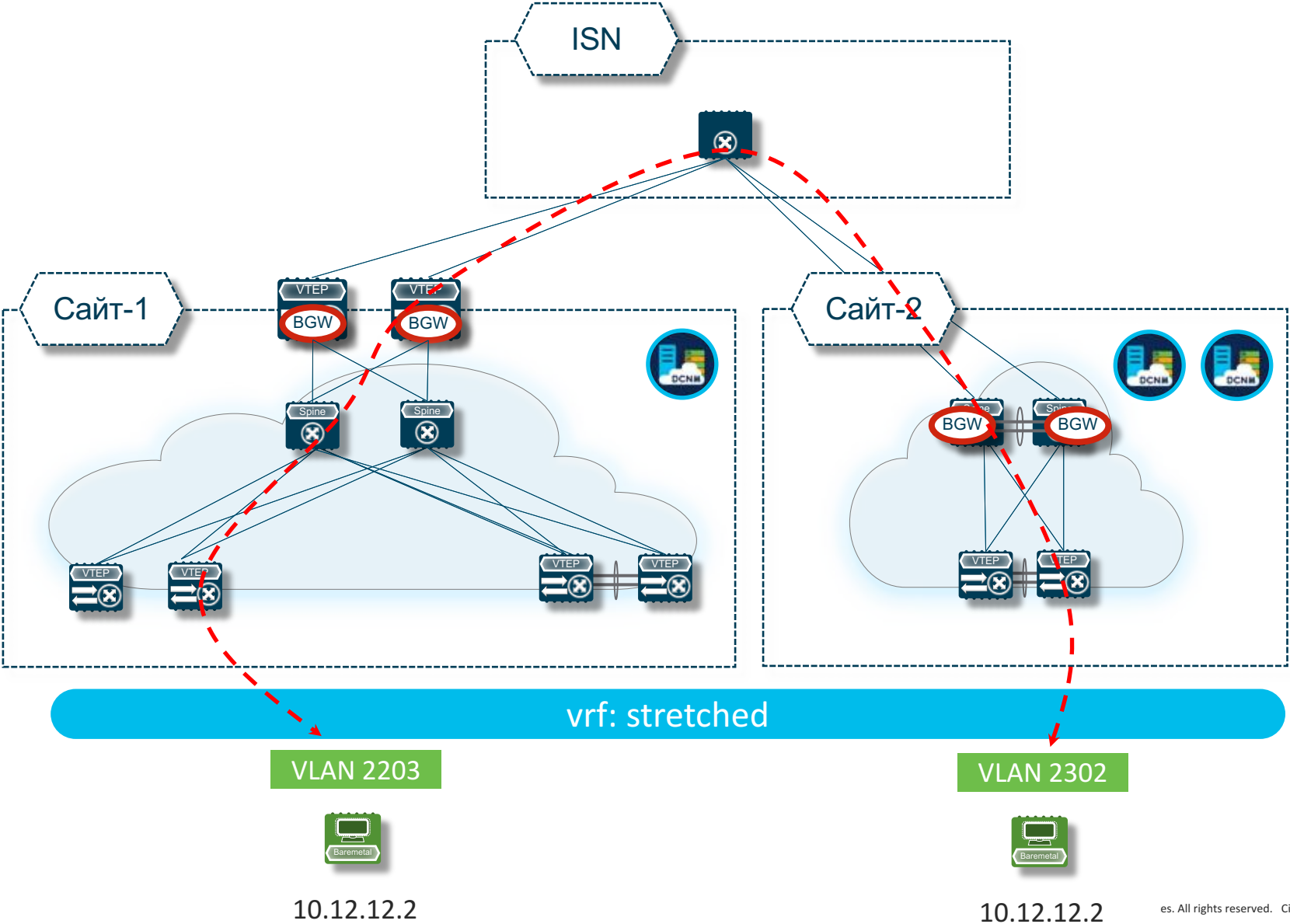
Documentation
Access cisco.com from documentation on configuration, maintenance and operation.

Демонстрация

Проверка связанности

4

Проверка связанности в
растянутом L2-сегменте



vmvSphere ClientMenuSearch in all environments

testhxUTM_0.3vCenter Server AppliancevCSA 6.7 U3vm-DEF-SR3vrf.dmz.vlan.2202.2vrf.inside.vlan.2200.2vrf.inside.vlan.2201.2vrf.shared.vlan.2203.2vrf.svc.d.2251.2vrf.svc.s.2250.2W2K-TS

vcsa-hxnvme.moslab.cisco.comMain DC

hxnvme-110.5.32.19710.5.32.19810.5.32.199dc1-border-routerdc1-loadbalancersprint.vlan.2300stCtIVM-WZP24210ARASTCtIVM-WZP24210ARCstCtIVM-WZP24210ASAvrf.inside.vlan.2301.2vrf.shared.vlan.2302.2

vrf.shared.vlan.2203.2

SummaryMonitorConfigurePermissionsDatastoresNetworksUpdates

Powered On

Launch Web ConsoleLaunch Remote Console

Guest OS:CentOS 7 (64-bit)

Compatibility:ESXi 6.5 and later (VM version 13)

VMware Tools:Not running, not installed

DNS Name:

IP Addresses:

Host:10.5.30.204

CPU USAGE0 Hz

MEMORY USAGE20 MB

STORAGE USAGE895.67 MB

VMware Tools is not installed on this virtual machine.

VM Hardware

CPU	1 CPU(s)
Memory	2 GB, 0.02 GB memory active
Hard disk 1	16 GB
Network adapter 1	NX-OS-2203 (connected)
CD/DVD drive 1	Disconnected
Video card	8 MB
VMCI device	Device on the virtual machine PCI bus that provides support for the virtual machine communication interface
Other	Additional Hardware

Notes

Custom Attributes

Recent TasksAlarms

Task Name	Target	Status	Details	Initiator	Queued For	Start Time	Completion Time	Server
Running								

More Tasks

© 2021 Cisco and/or its affiliates. All rights reserved. Cisco Public

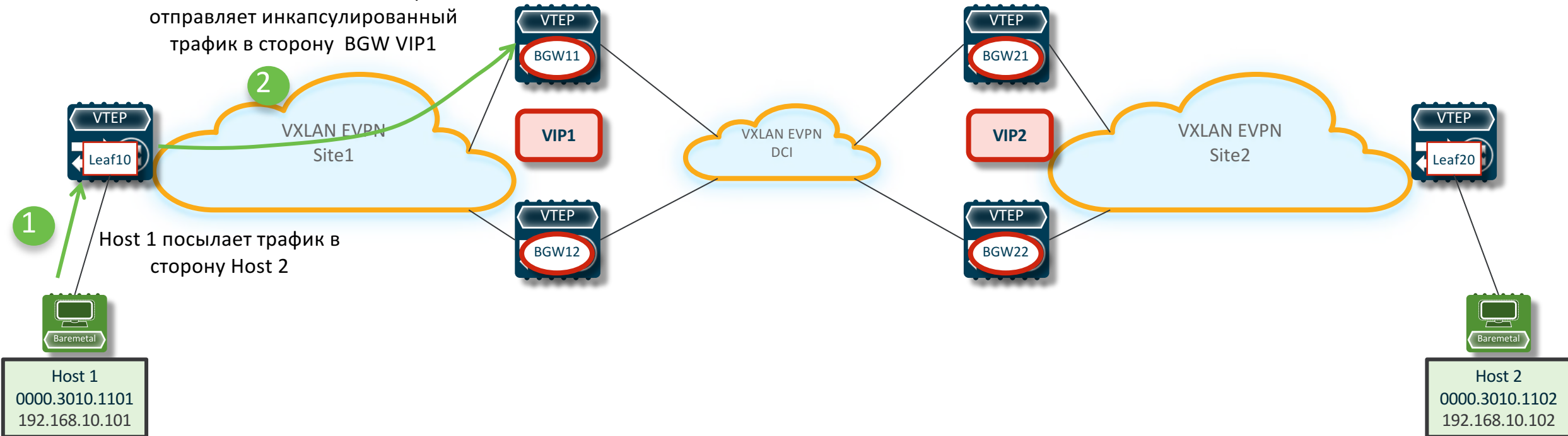
Передача пакетов между
сайтами (коммутация)

Передача пакетов в VXLAN Multi-Site Layer 2 (Host 1 в сторону Host 2) – сайт 1

Коммутация

SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
L10	BGW-VIP1	30010	H1-MAC	H2-MAC	H1-IP	H2-IP	

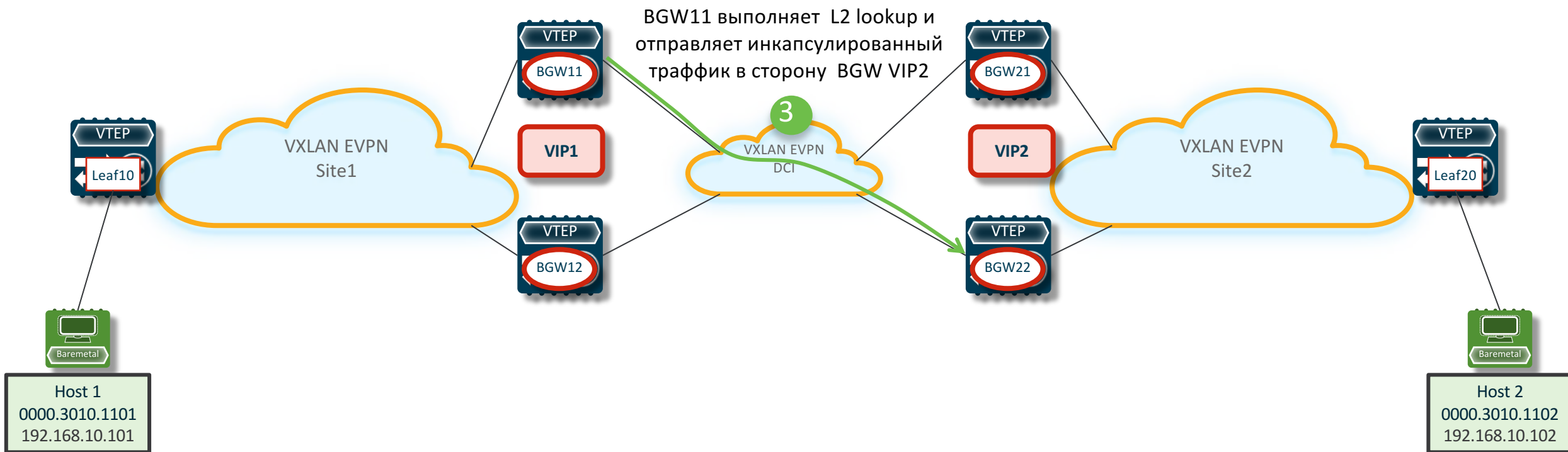
Leaf10 выполняет L2 lookup и отправляет инкапсулированный трафик в сторону BGW VIP1



Передача пакетов в VXLAN Multi-Site Layer 2 (Host 1 в сторону Host 2) – DCI

Коммутация

SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
BGW-VIP1	BGW-VIP2	30010	H1-MAC	H2-MAC	H1-IP	H2-IP	

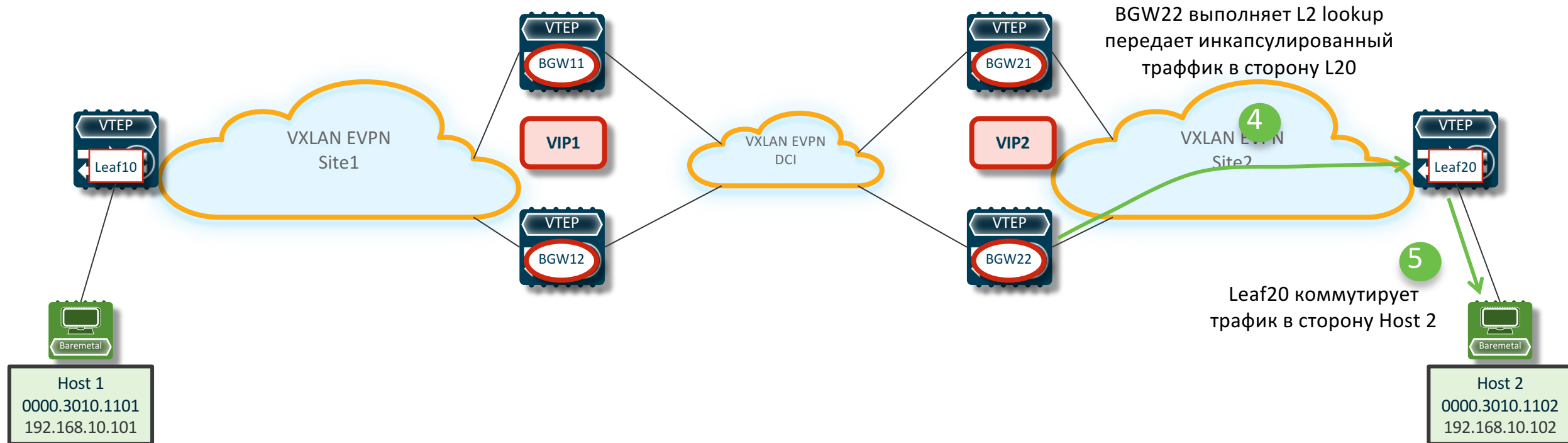


Передача пакетов в VXLAN Multi-Site

Layer 2 (Host 1 в сторону Host 2) – сайт 2

Коммутация

SIP	DIP	VXLAN	SMAC	DMAC	SIP	DIP	Payload
BGW-VIP2	L20	30010	H1-MAC	H2-MAC	H1-IP	H2-IP	



Подключение ко внешним L3
сетям (L3out)

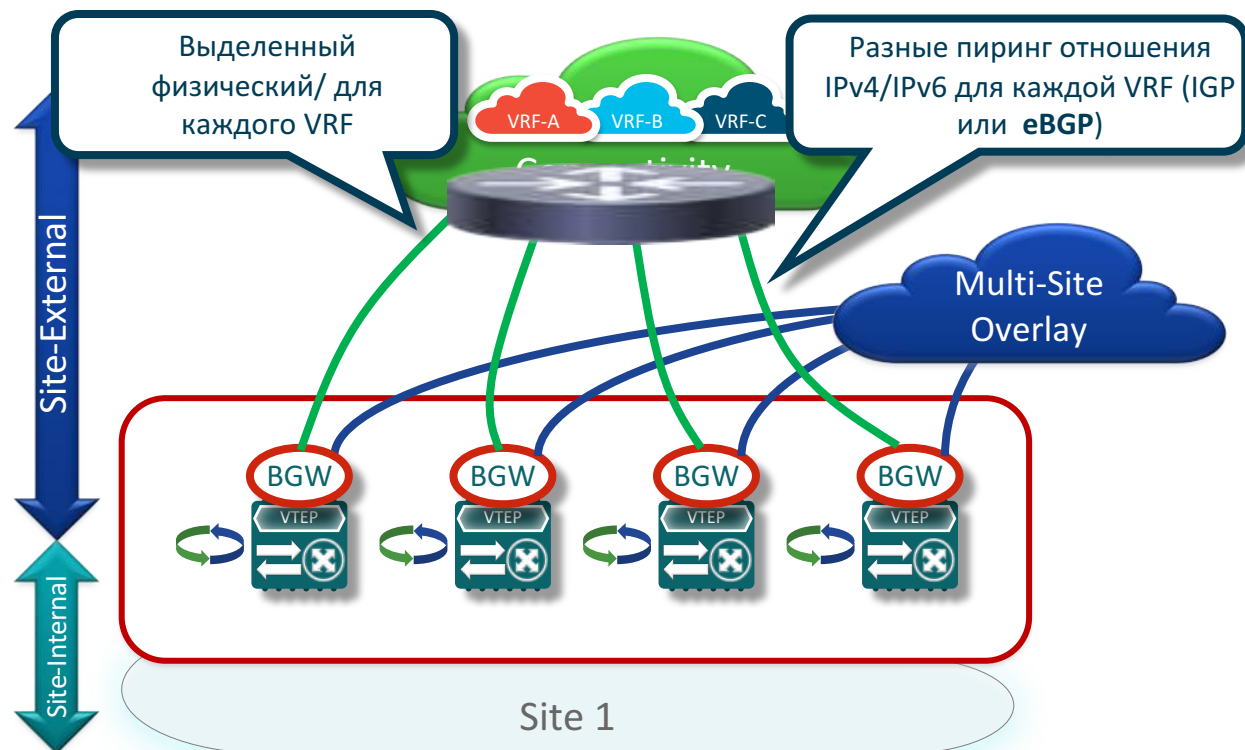
VXLAN Multi-Site

Подключение ко внешним L3 доменам

- BGW узлы могут использоваться для подключения ко внешним сетям на каждом сайте
 - Альтернатива – использовать Border Leaf
- Поддерживаются различные режимы
 - VRF-Lite пиринг с пограничными WAN Edge маршрутизаторами
 - MP-BGP EVPN пиринг с пограничными WAN Edge маршрутизаторами (Режим Shared Border deployment)
 - Выделенная или общая для нескольких сайтов пара WAN Edge устройств
- Подключение внешних Layer-3 сетей может выполняться отдельно от DCI сети, которая используется для inter-site взаимодействия

VXLAN Multi-Site

Border Gateways и VRF-Lite на внешних маршрутизаторах

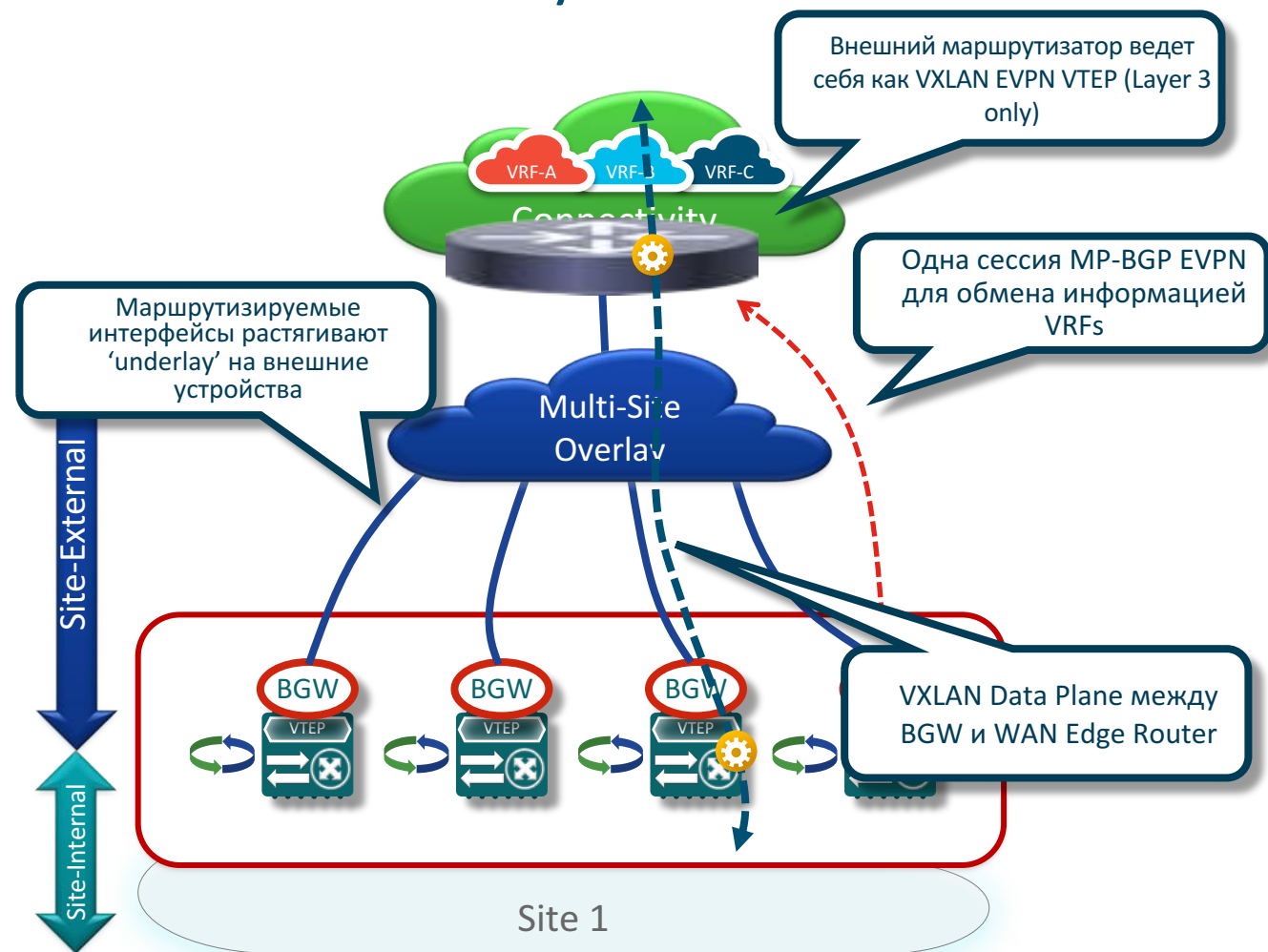


- Выделенный IPv4/IPv6 пиринг для каждой VRF с использованием выделенных physical interfaces/sub-interfaces
- Требуется использовать выделенные интерфейсы для inter-site взаимодействия
 - No support for VXLAN encapsulated traffic on sub-interfaces*

*VXLAN on the main interface and VRF-Lite on sub-interfaces planned for a future NXOS release

VXLAN Multi-Site

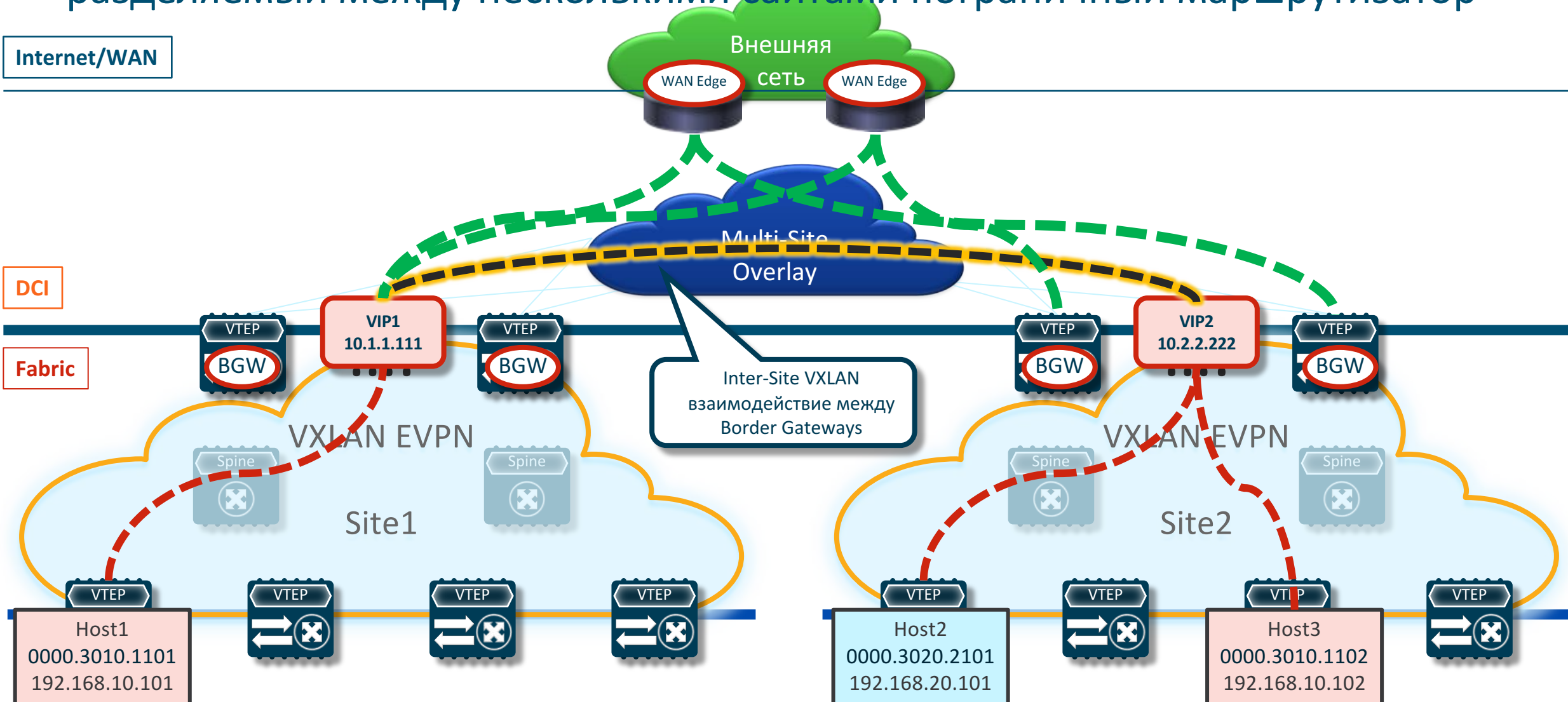
Border Gateway и Shared Border



- Один MP-BGP EVPN пиринг для обмена маршрутной информацией для всех VRF
- VXLAN Data-Plane между BGW и внешними маршрутизаторами
- Одни и те же аплинки на spine используются для всего VXLAN-трафика North-South и East-West)
 - Required because of the use of DCI link tracking
- Разные опции hand-off в зависимости по поддержки аппаратурой: VRF-Lite, MPLS-VPN, LISP

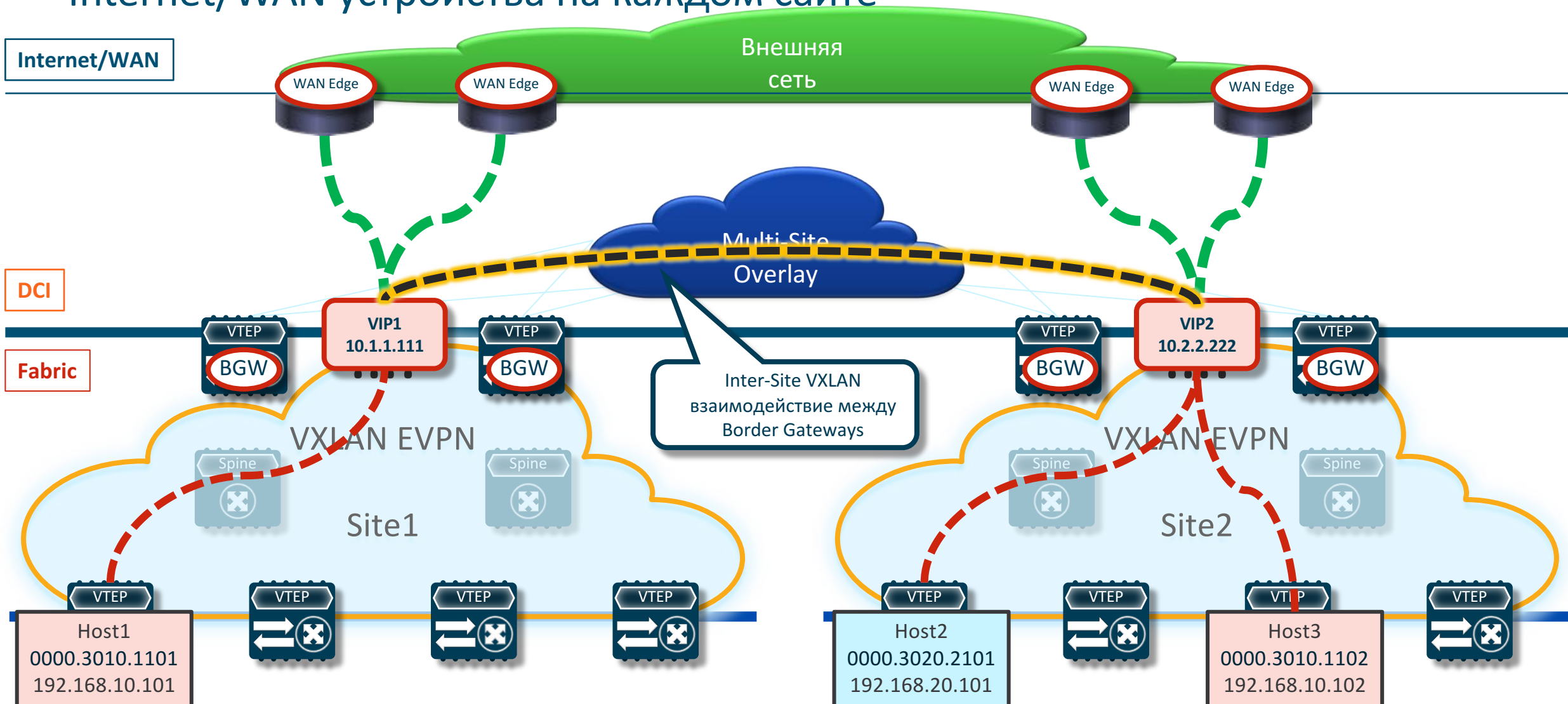
VXLAN Multi-Site

разделяемый между несколькими сайтами пограничный маршрутизатор



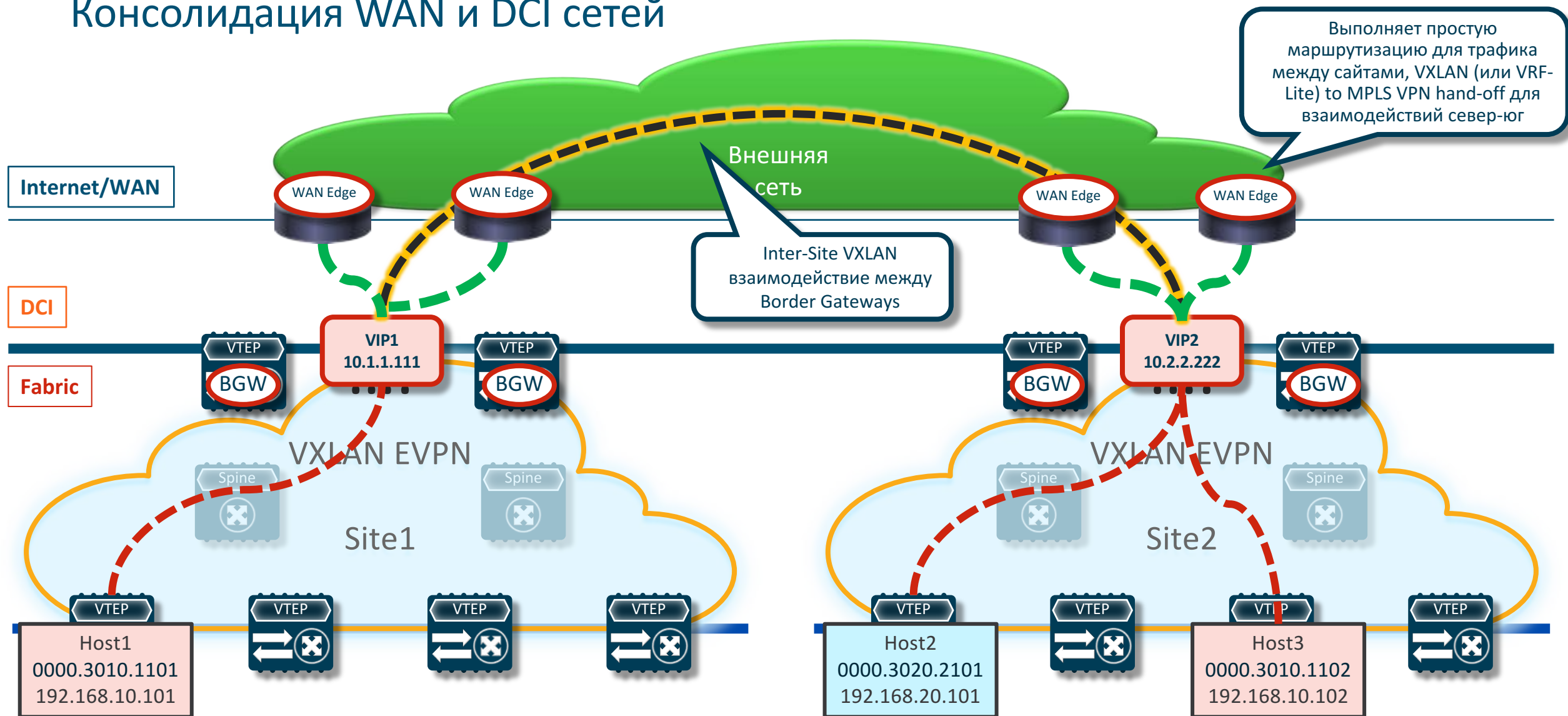
VXLAN Multi-Site

Internet/WAN устройства на каждом сайте



VXLAN Multi-Site

Консолидация WAN и DCI сетей

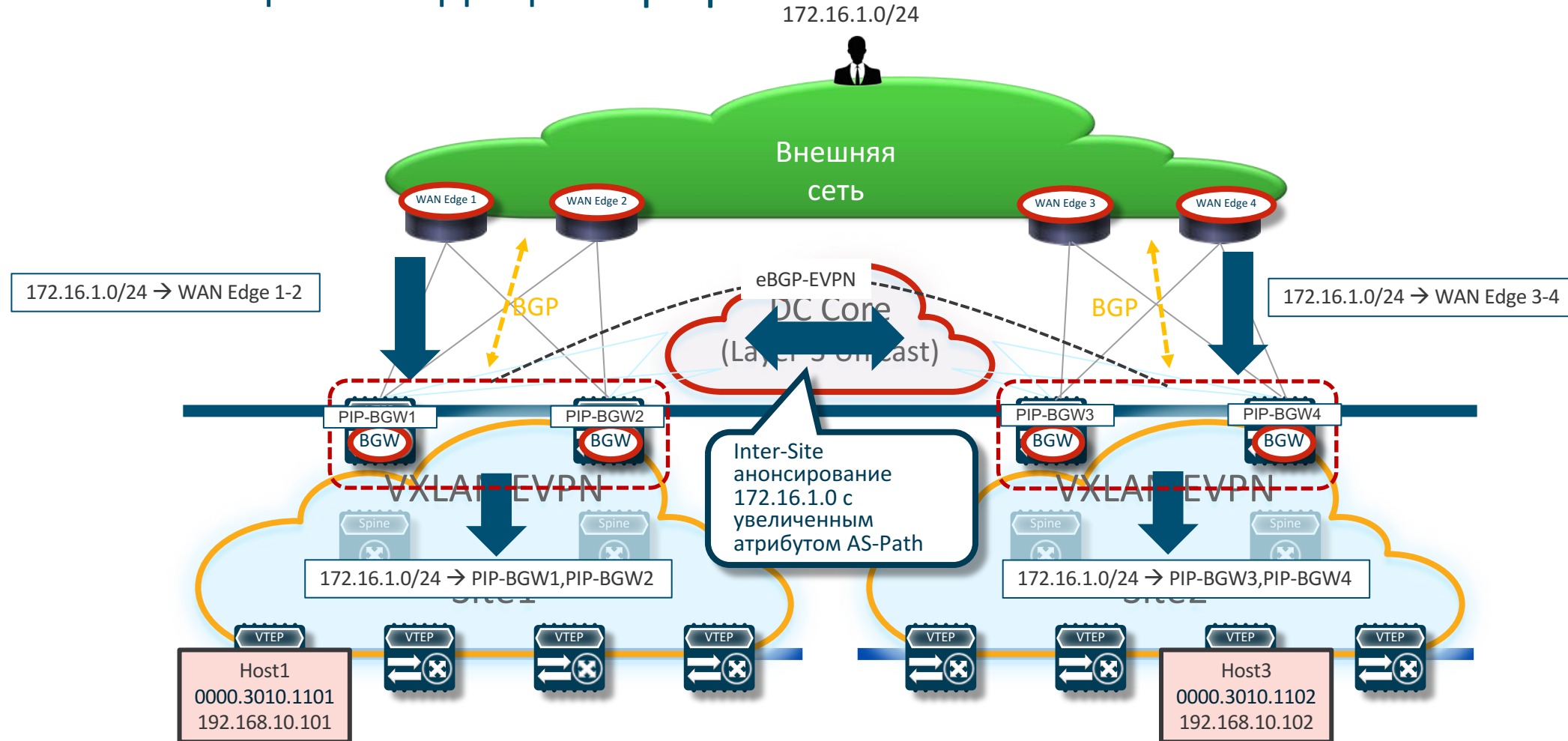


Оптимизация входящего и исходящего трафика

Internet/WAN шлюзы

Оптимизация исходящего трафика

172.16.1.0/24



Internet/WAN шлюзы

Оптимизация исходящего трафика

172.16.1.0/24

Внешняя сеть

Оптимальный
исходящий путь

Оптимальный
исходящий путь

еBGP-EVPN
DC Core
(Layer-3 Unicast)

IP Prefix	Next-Hop
172.16.1.0/24	WAN Edge 1,2

IP Prefix	Next-Hop
172.16.1.0/24	WAN Edge 3,4

Outbound flow toward
172.16.1.0/24

Outbound flow toward
172.16.1.0/24

Site1

Site2

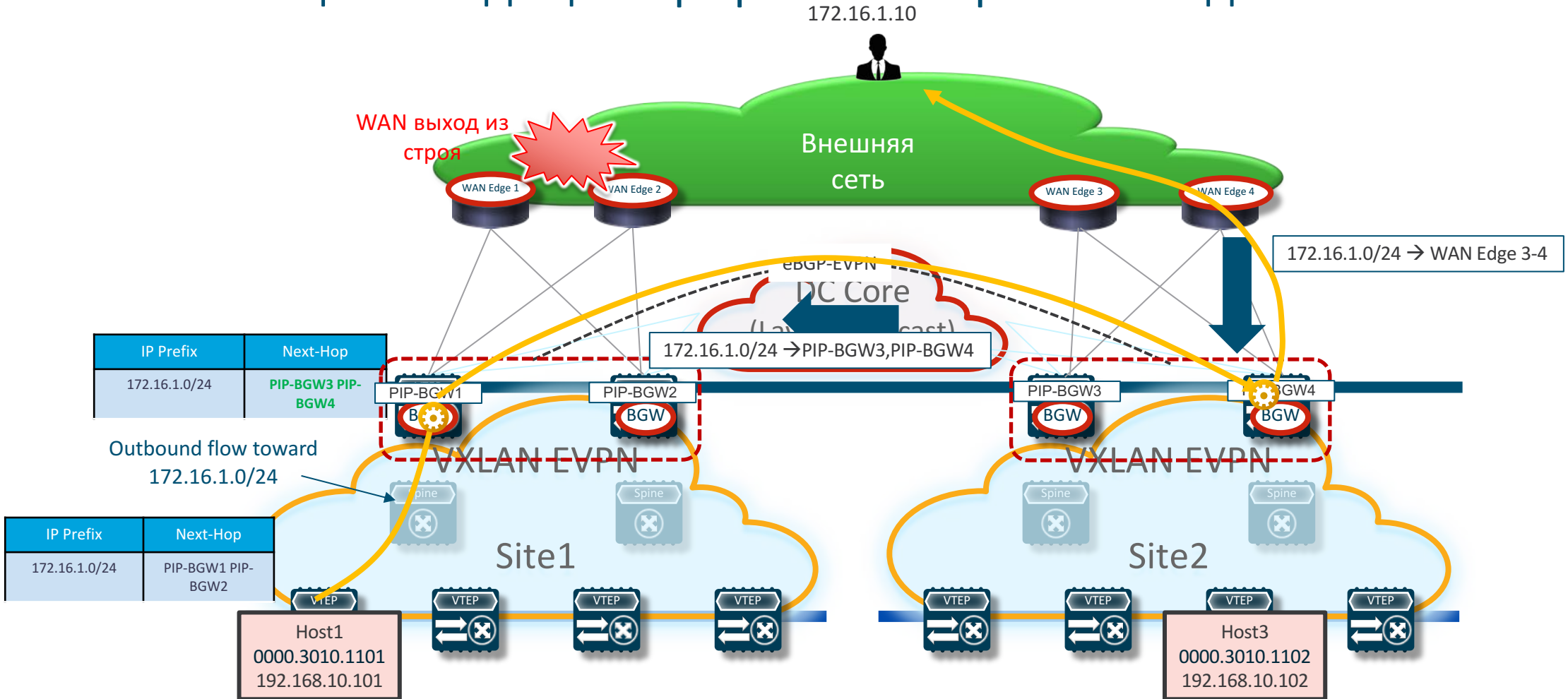
Host1
0000.3010.1101
192.168.10.101

Host3
0000.3010.1102
192.168.10.102

IP Prefix	Next-Hop
172.16.1.0/24	PIP-BGW3 PIP-BGW4

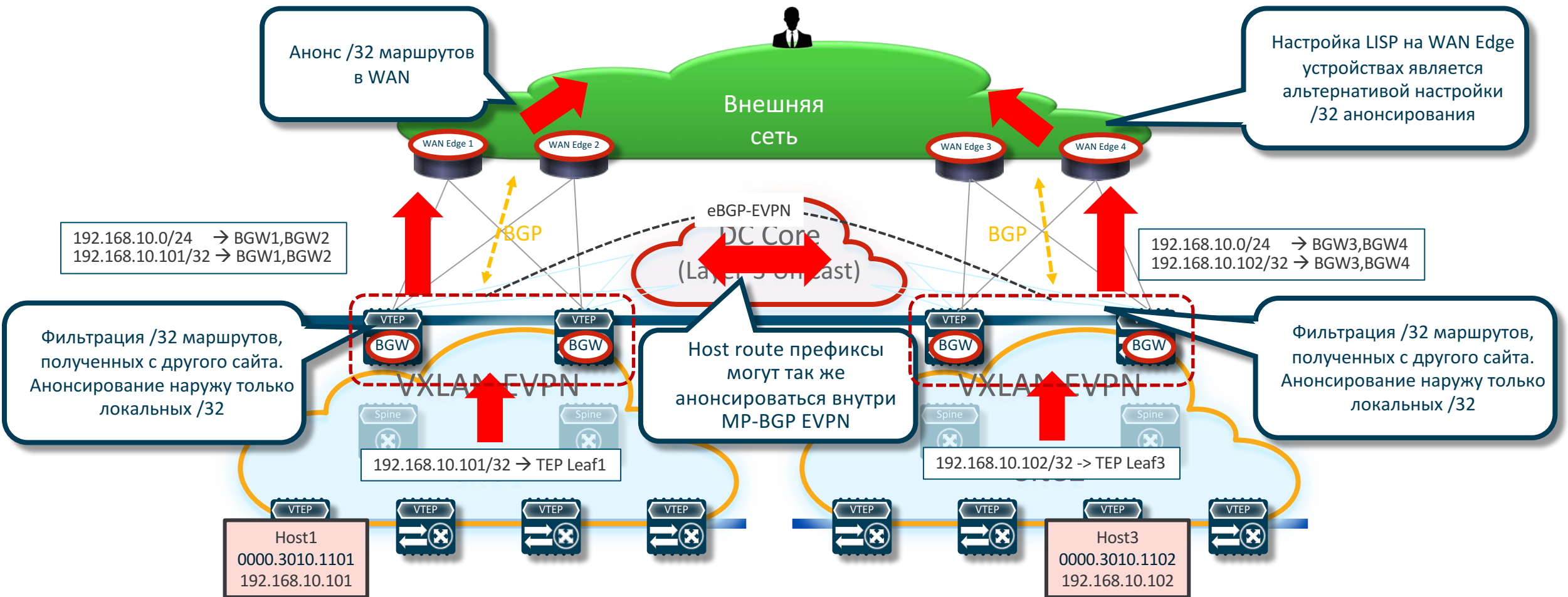
Internet/WAN шлюзы

Оптимизация исходящего трафика – потеря WAN подключения на сайте



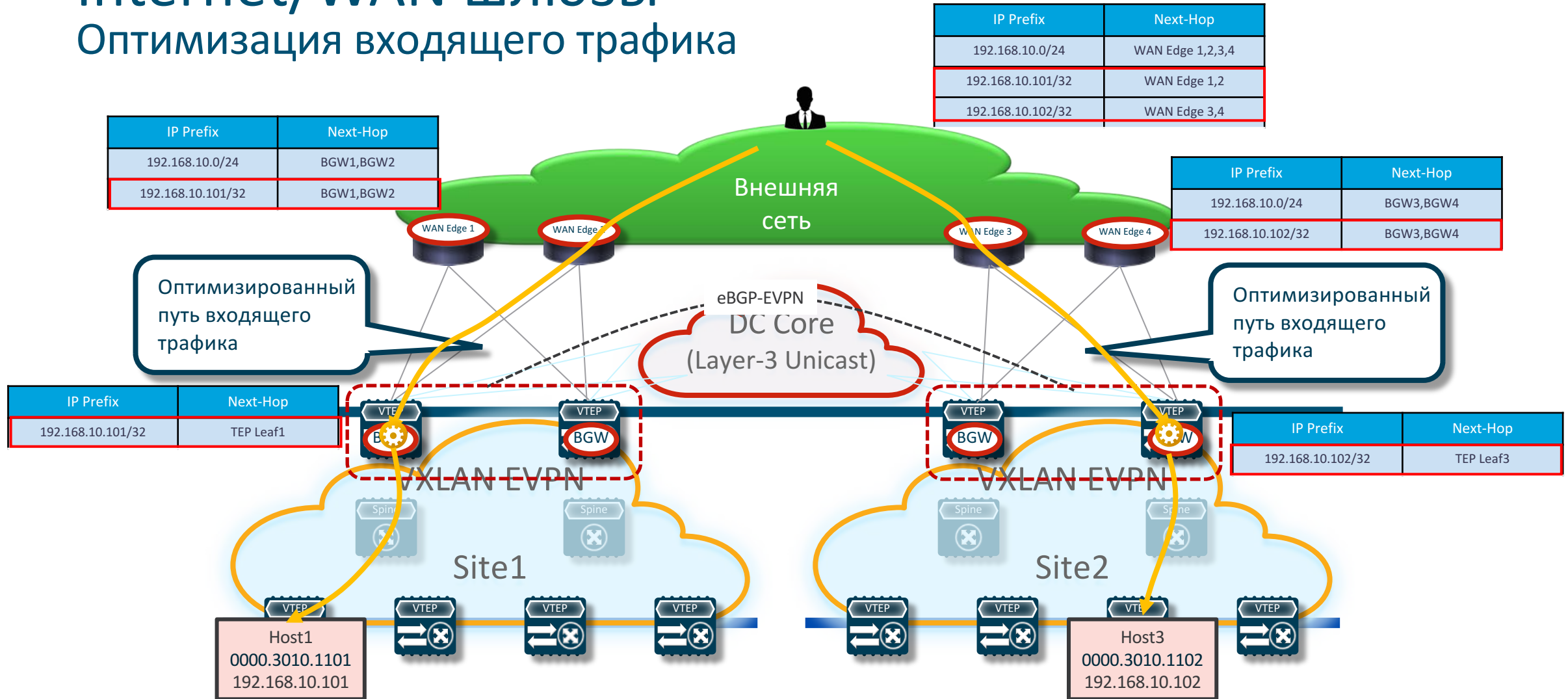
Internet/WAN шлюзы

Оптимизация входящего трафика



Internet/WAN шлюзы

Оптимизация входящего трафика



Internet/WAN шлюзы

Оптимизация входящего трафика – потеря WAN подключения на сайте

