Сеть ЦОД на основе VXLAN/EVPN за 5 дней

День 2: VXLAN/EVPN фабрика – клиенты, внешние сегменты, расширенные возможности NX-OS

Илья Дрей
Системный архитектор
06/04/2021

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

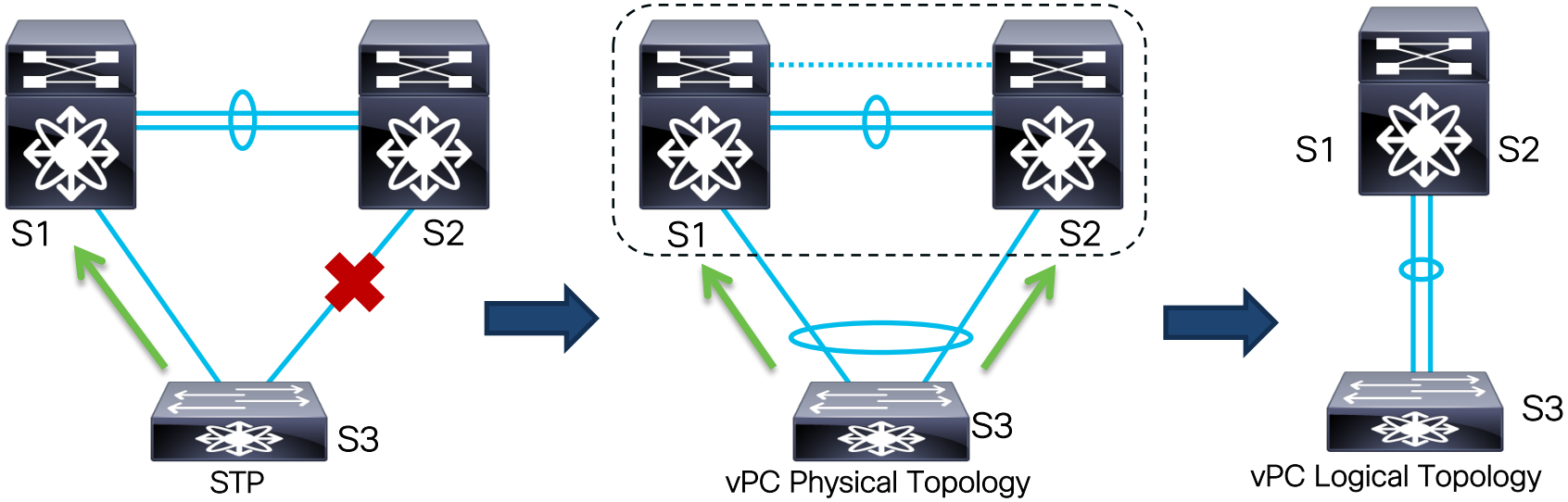
Содержание

- Часть 1: Организация отказоустойчивых подключений клиентских устройств
- Часть 2: Внешние подключения
- Демо: Конфигурация сервисов и внешних подключений с помощью Cisco DCNM
- Часть 3: Расширенные функции NX-OS

Часть 1: Организация отказоустойчивых подключений клиентских устройств

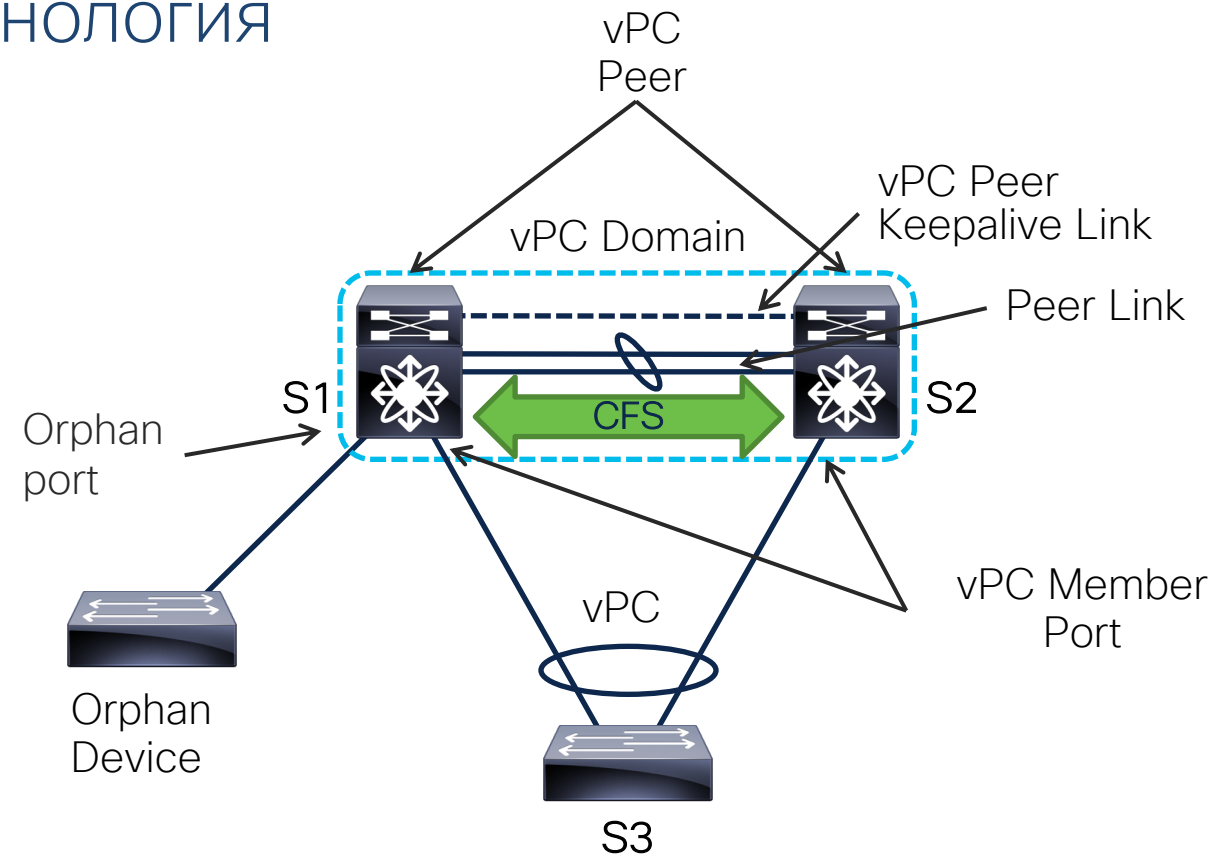
Virtual Port-Channel (vPC)

Для чего нужен vPC?



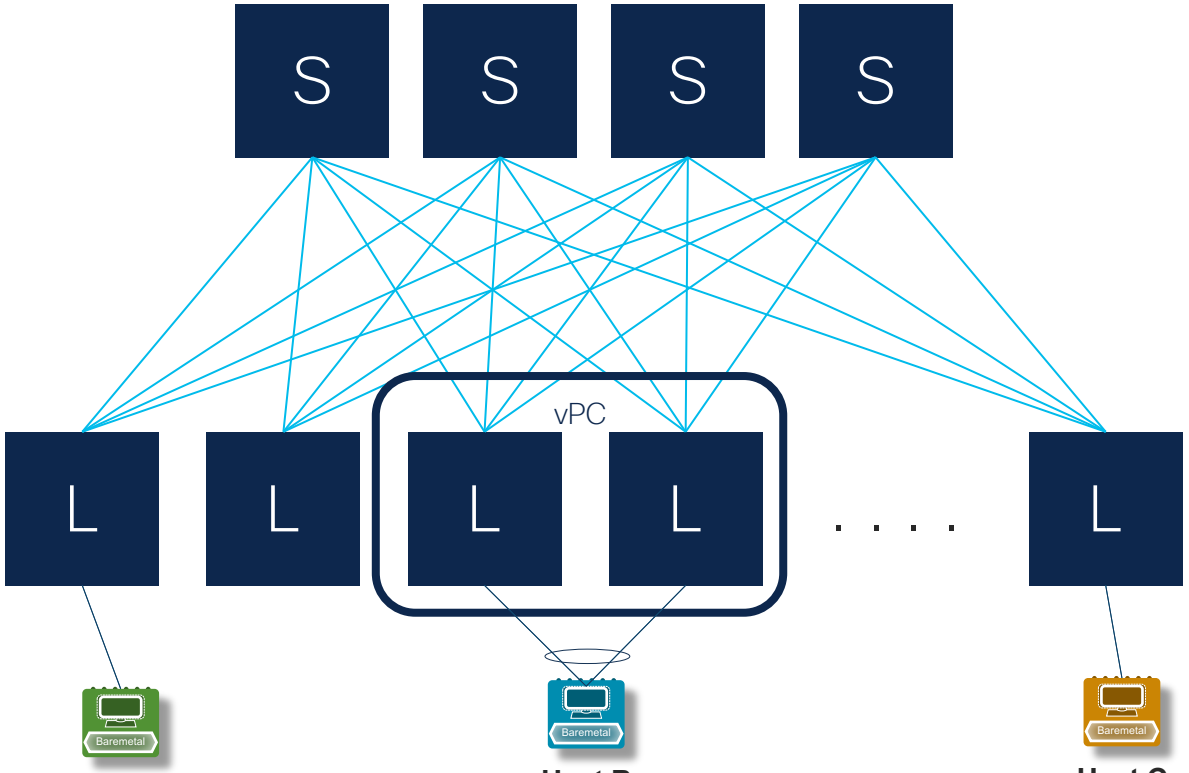
- Отсутствие заблокированных портов
- Оптимальное использование пропускной полосы

Терминология



VXLAN + vPC = Anycast VTEP

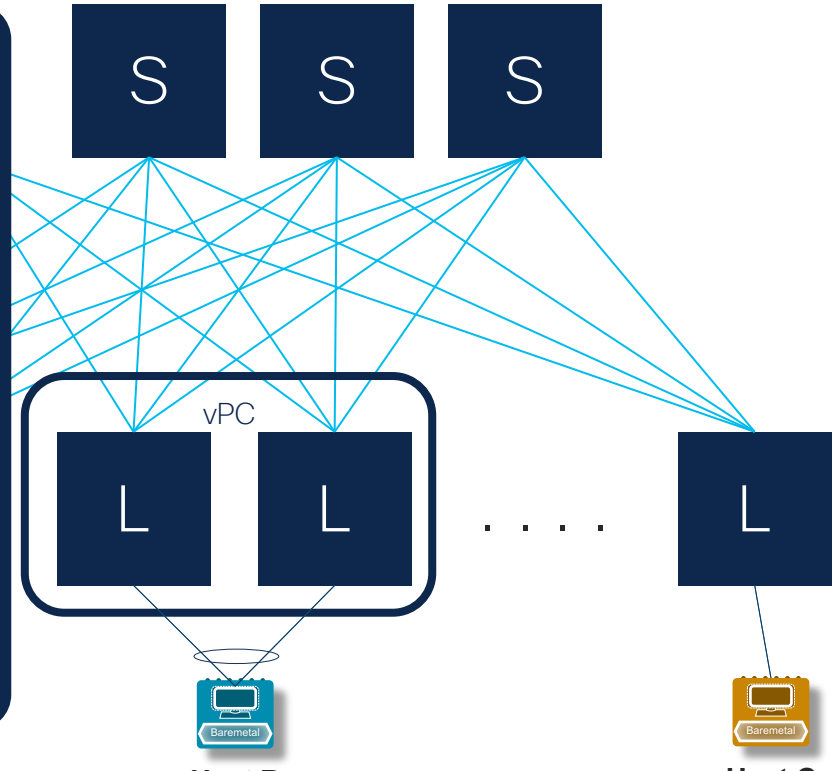
Anycast VTEP



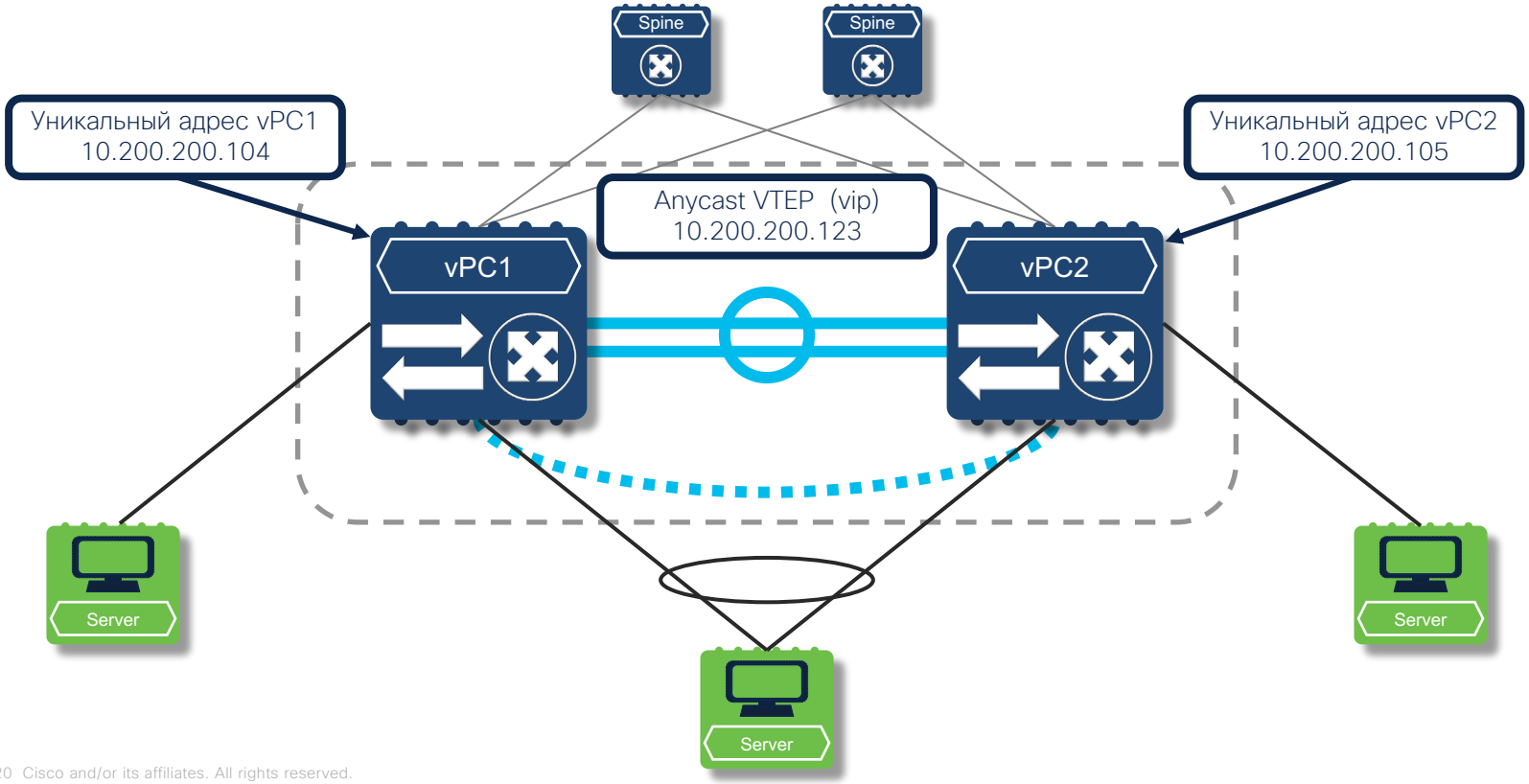
9

Anycast VTEP

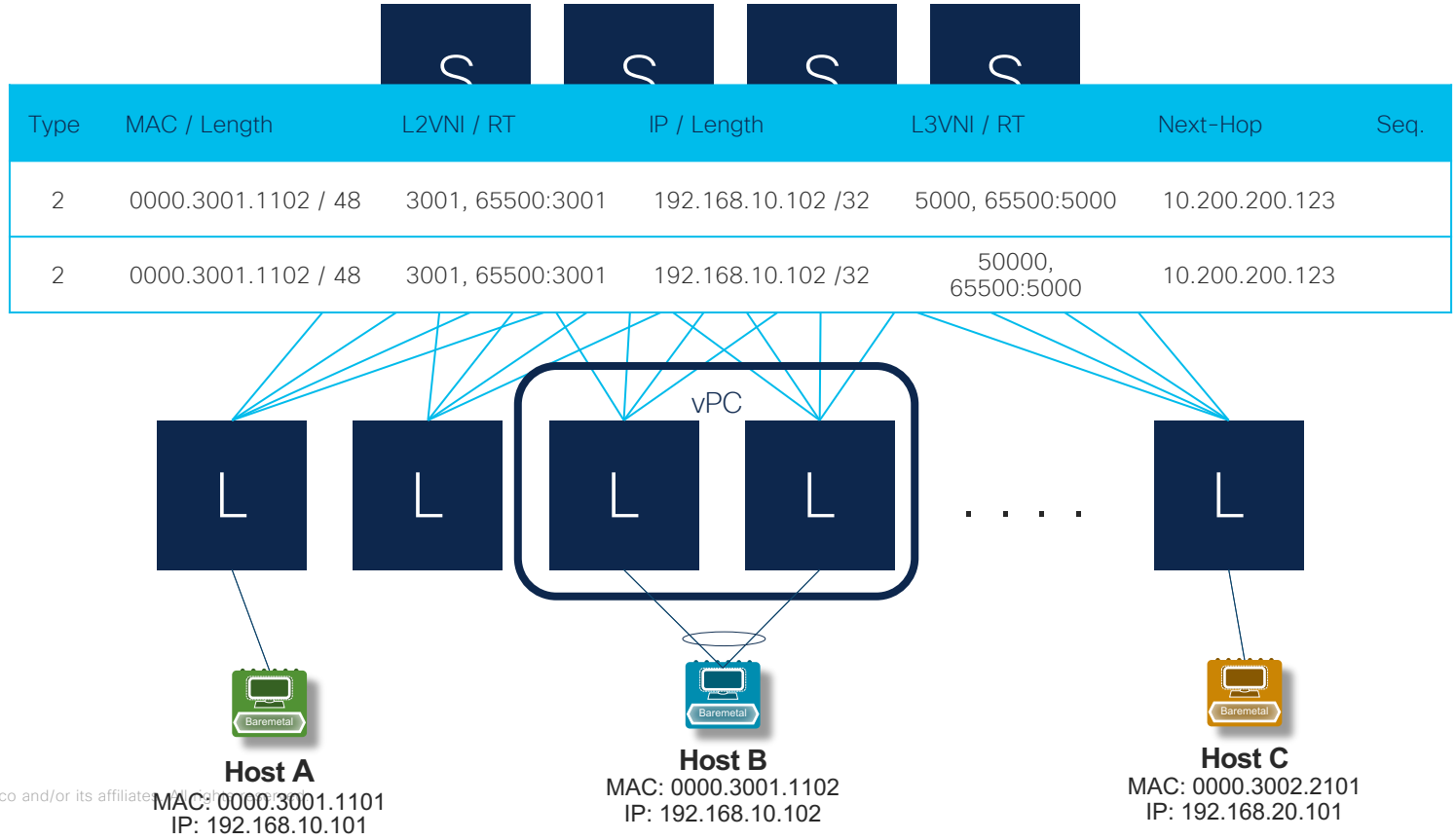
- Каждый из членов vPC домена является самостоятельным участником EVPN control-plane
  - Собственный сессии Underlay и Overlay
  - Уникальные RD
- Единый VTEP с точки зрения VXLAN data-plane
  - Next-Hop = Anycast VTEP IP
  - Балансировка за счет ECMP в Underlay



Anycast VTEP – Адресация

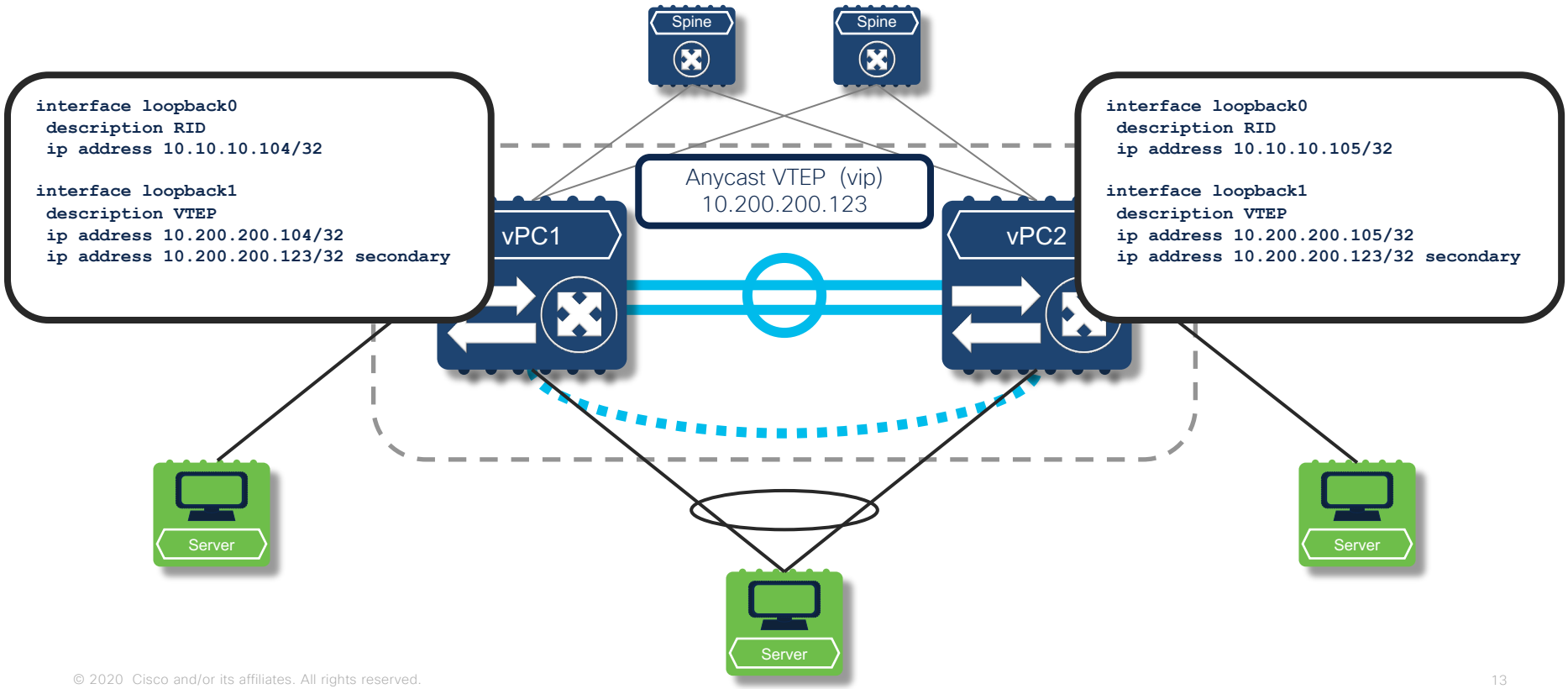


Anycast VTEP – Пример маршрута

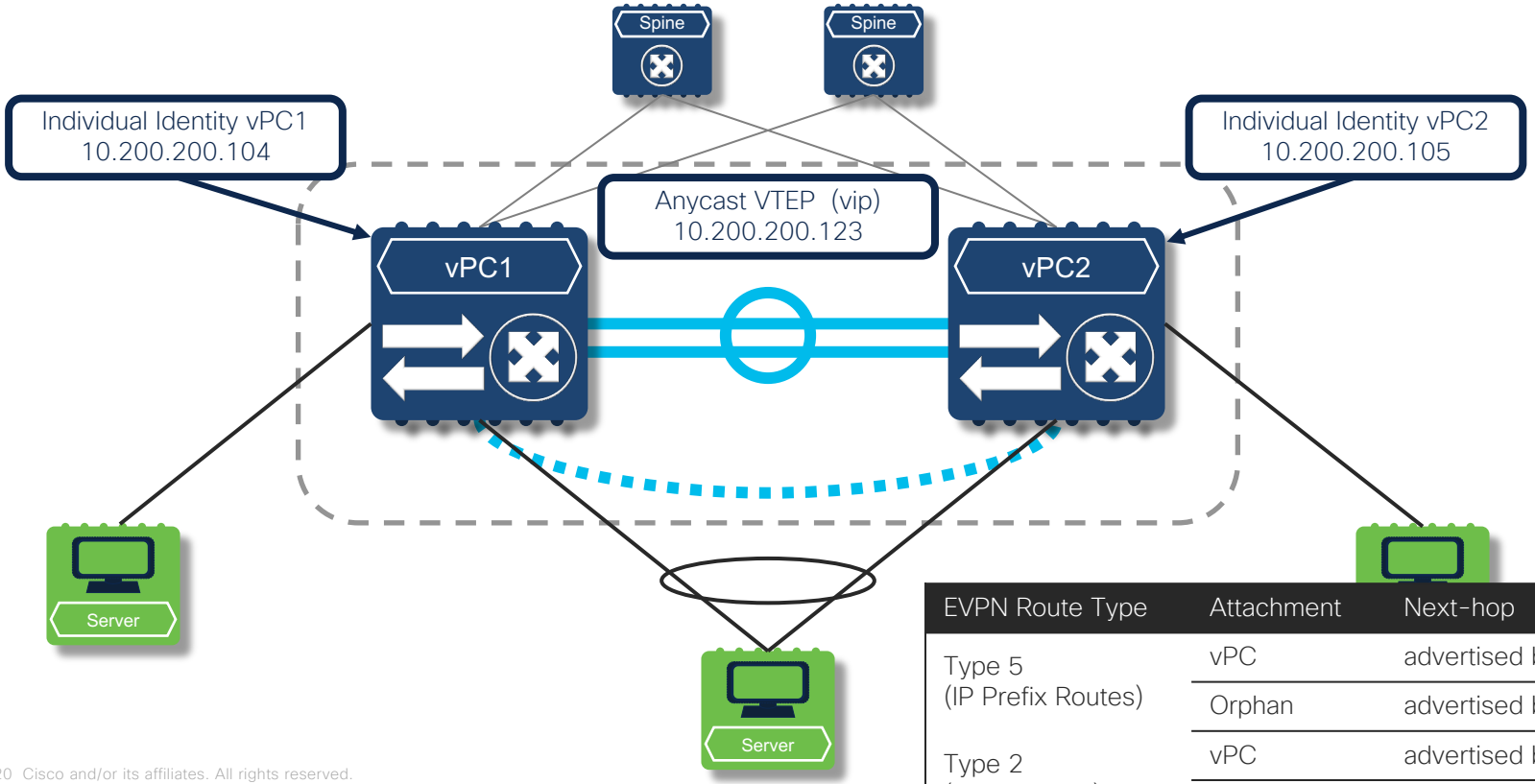


12

Anycast VTEP – Адресация



Anycast VTEP – Адресация

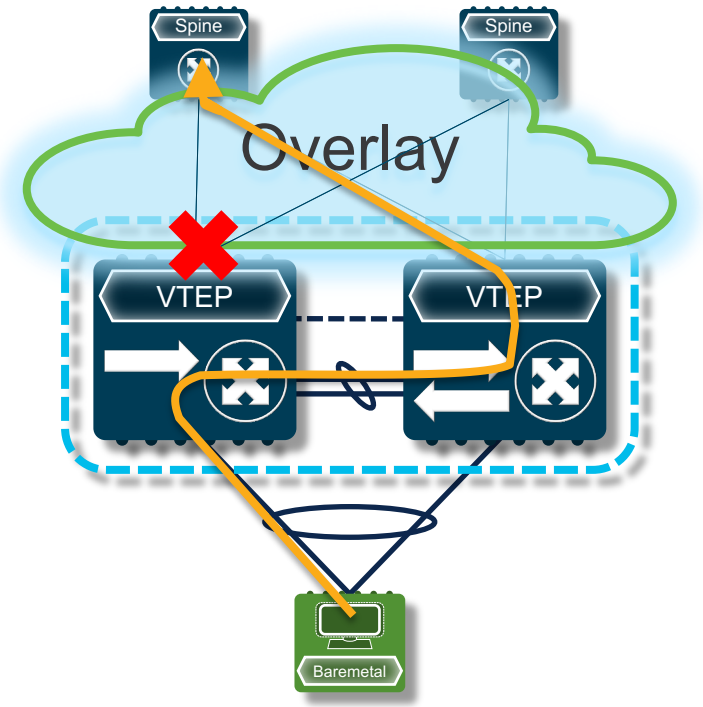


EVPN Route Type	Attachment	Next-hop
Type 5 (IP Prefix Routes)	vPC	advertised by vip*
	Orphan	advertised by vip*
Type 2 (Host Routes)	vPC	advertised by vip
	Orphan Port	advertised by vip

*В случае использования опции “advertise-pip” будет использоваться PIP

Маршрутизация через vPC peer-link

- В случае использования vPC важно обеспечить резервную маршрутизацию через vPC peer-link в underlay
- В случае отсутствия резервного маршрута и потери связи со Spine коммутаторами будет происходить частичная потеря трафика

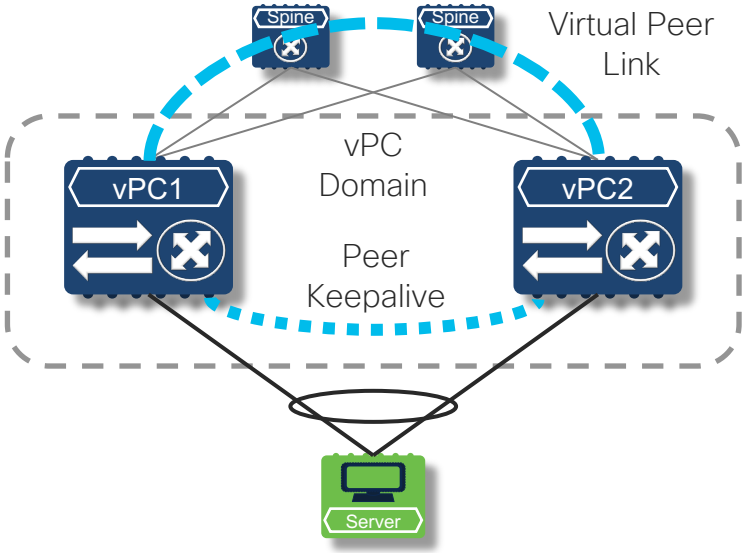


vPC Fabric Peering

NX-OS 9.2(3)

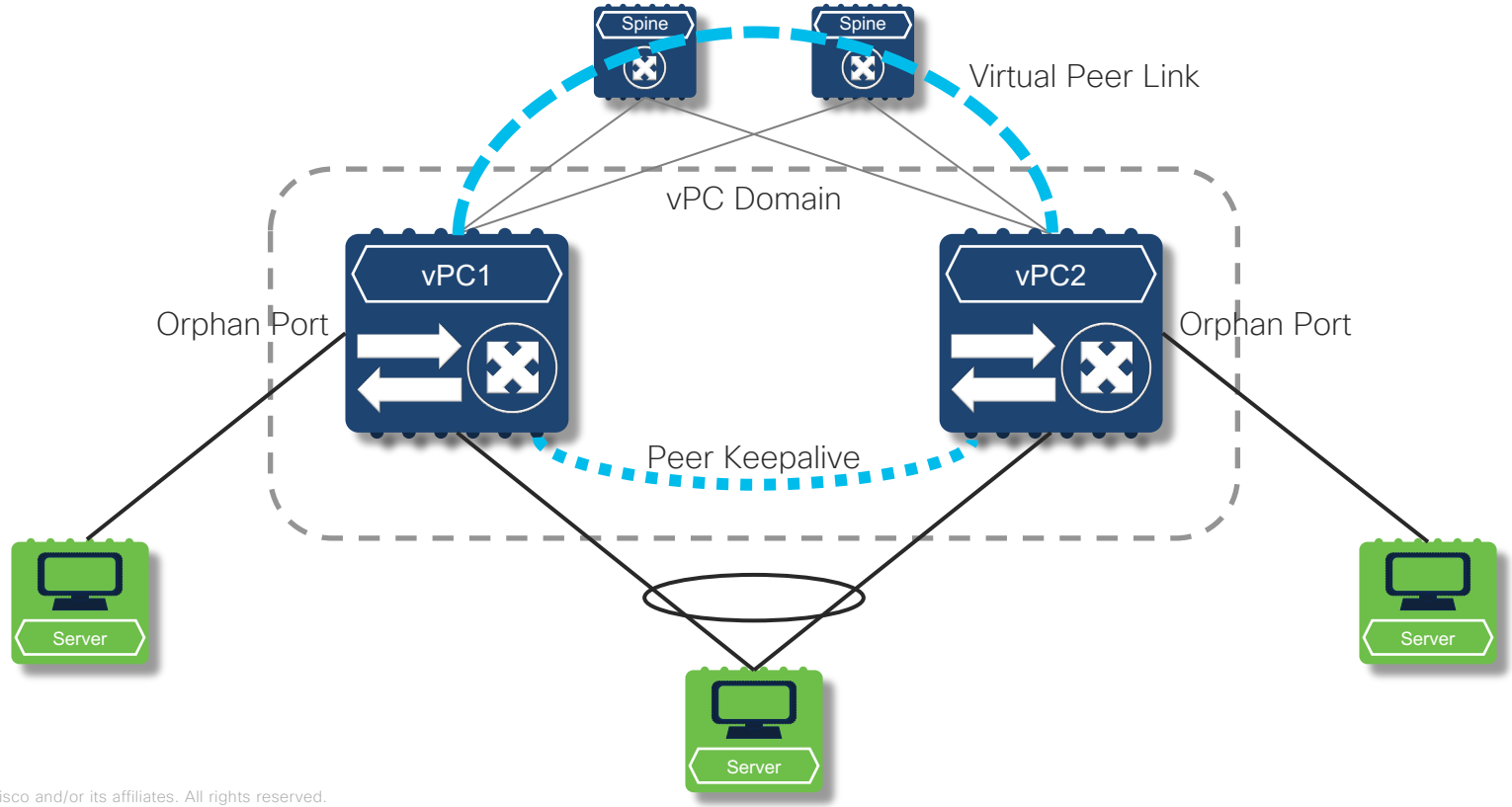
vPC Fabric Peering

- Позволяет сформировать vPC домен без наличия физического peer-link между коммутаторами
- Технология основывается на формировании виртуального vPC peer-link через underlay
- Наследует общие правила работы vPC домена



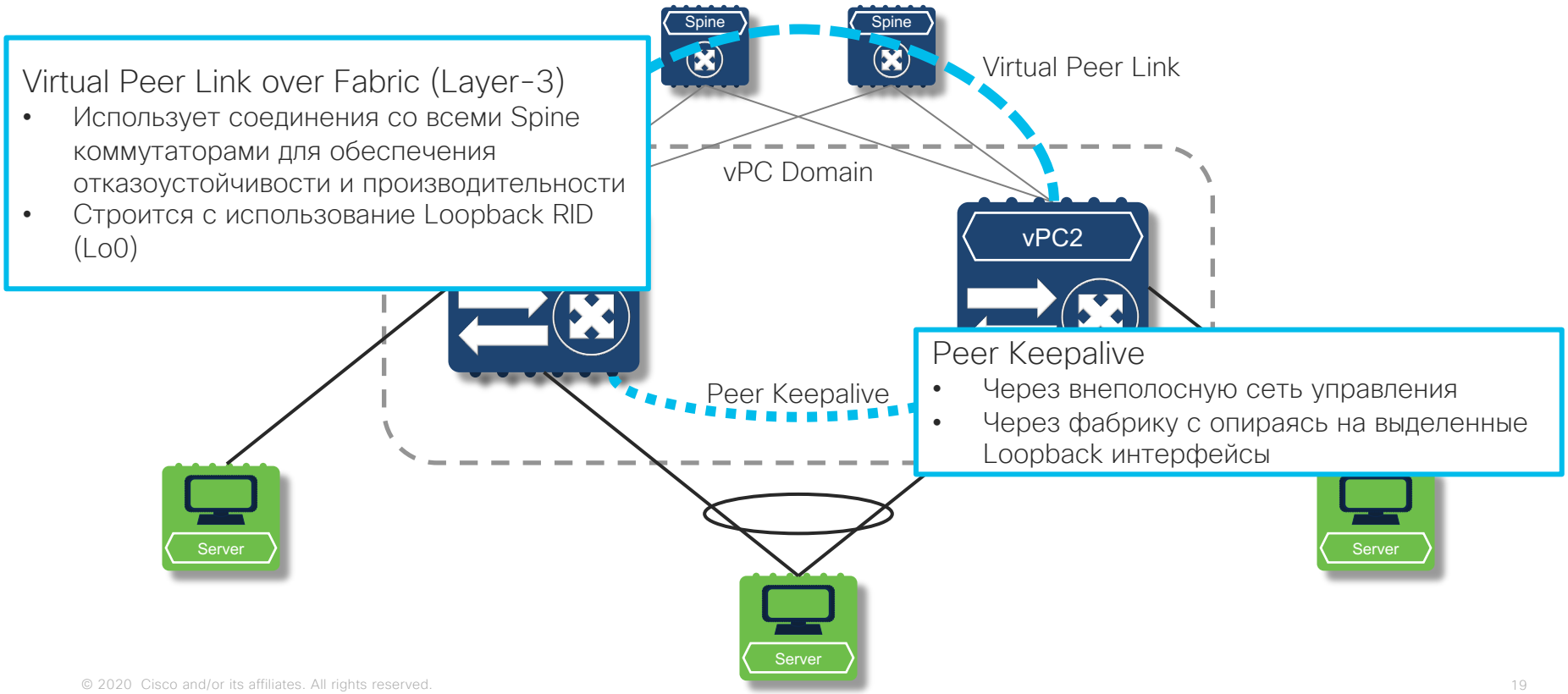
NX-OS 9.2(3)

vPC with Fabric Peering



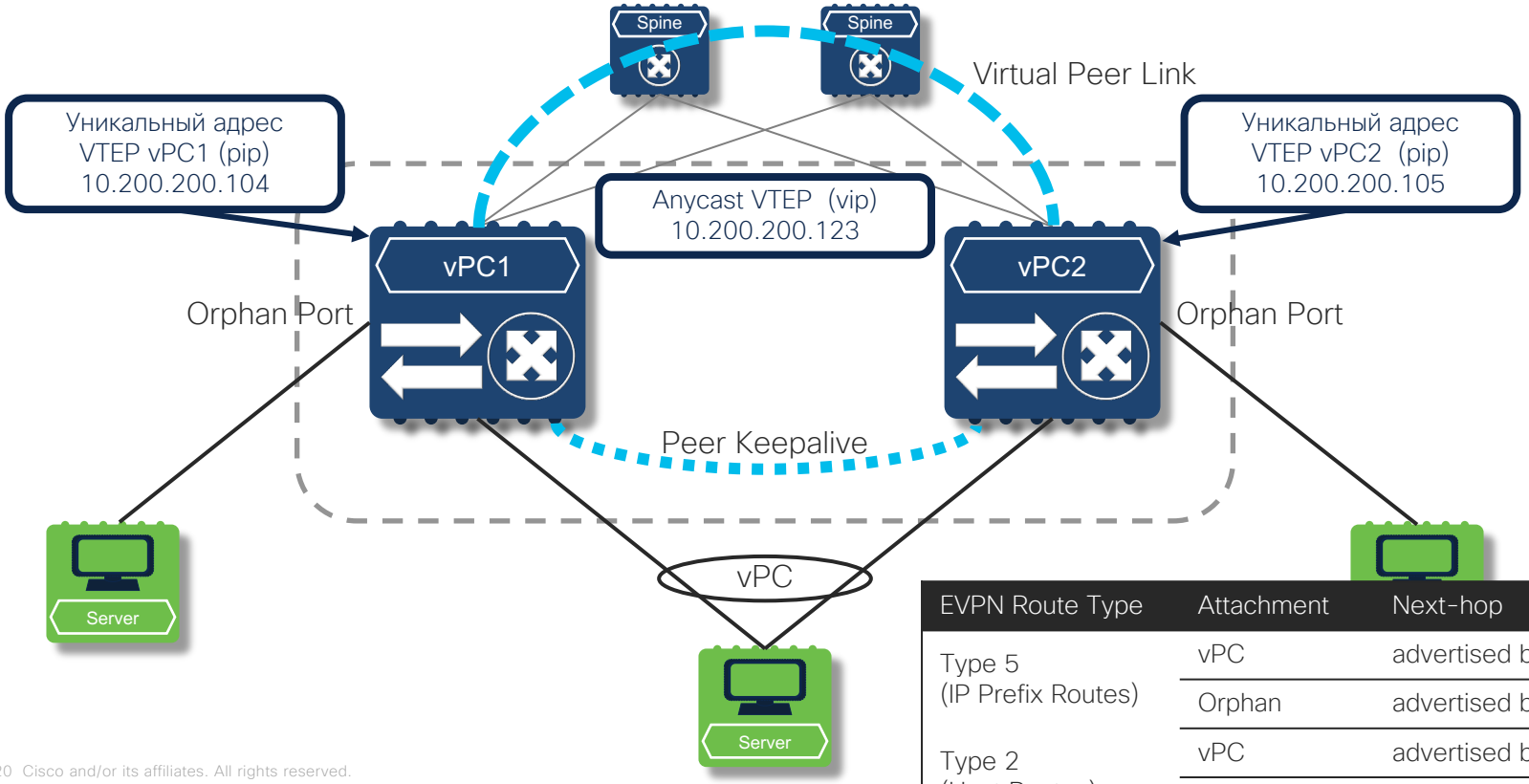
NX-OS 9.2(3)

vPC with Fabric Peering



NX-OS 9.2(3)

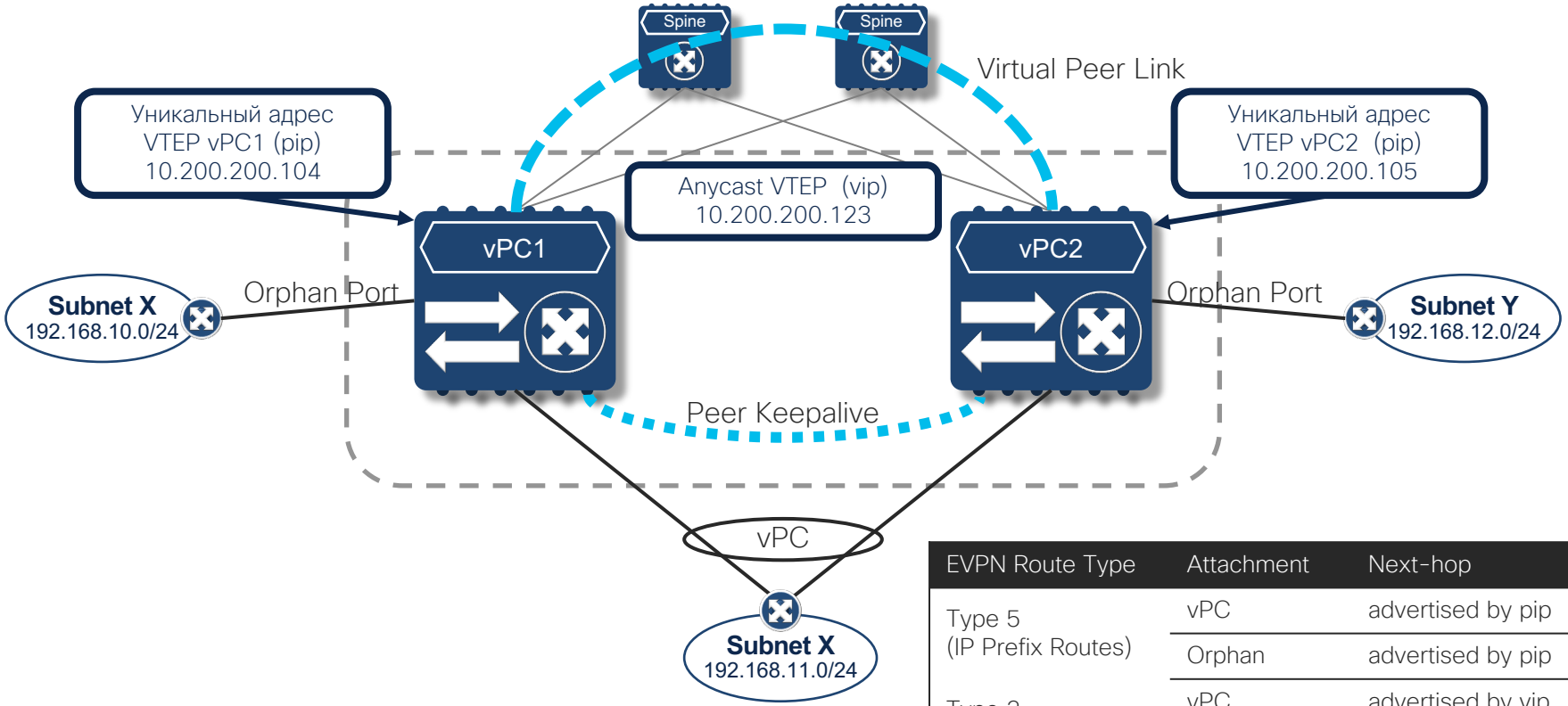
Подключение клиентов



EVPN Route Type	Attachment	Next-hop
Type 5 (IP Prefix Routes)	vPC	advertised by pip
	Orphan	advertised by pip
Type 2 (Host Routes)	vPC	advertised by vip
	Orphan Port	advertised by pip

NX-OS 9.2(3)

Подключение клиентов



EVPN Route Type	Attachment	Next-hop
Type 5 (IP Prefix Routes)	vPC	advertised by pip
	Orphan	advertised by pip
Type 2 (Host Routes)	vPC	advertised by vip
	Orphan Port	advertised by pip

Пример конфигурации vPC Fabric Peering (1/2)

```
vpc domain 1
  peer-switch
  peer-keepalive destination 10.10.10.82 source 10.10.10.81
  virtual peer-link destination 10.44.0.4 source 10.44.0.3 dscp 56
  delay restore 150
  peer-gateway
  auto-recovery reload-delay 360
  ipv6 nd synchronize
  ip arp synchronize

interface port-channel500
  description "vpc-peer-link"
  switchport
  switchport mode trunk
  spanning-tree port type network
  vpc peer-link
```

vPC Domain

Virtual peer-link

vPC Peer-Link Port-Channel
(не должен иметь физических интерфейсов)

Активируем роль порта peer-link

Пример конфигурации vPC Fabric Peering (2/2)

```
interface Ethernet1/49
  mtu 9216
  port-type fabric
  ip address 10.144.0.41/30
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
  no shutdown

interface Ethernet1/50
  mtu 9216
  port-type fabric
  ip address 10.144.0.29/30
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
  no shutdown
```

Задаем тип порта!

Интерфейсы подключения к Spine

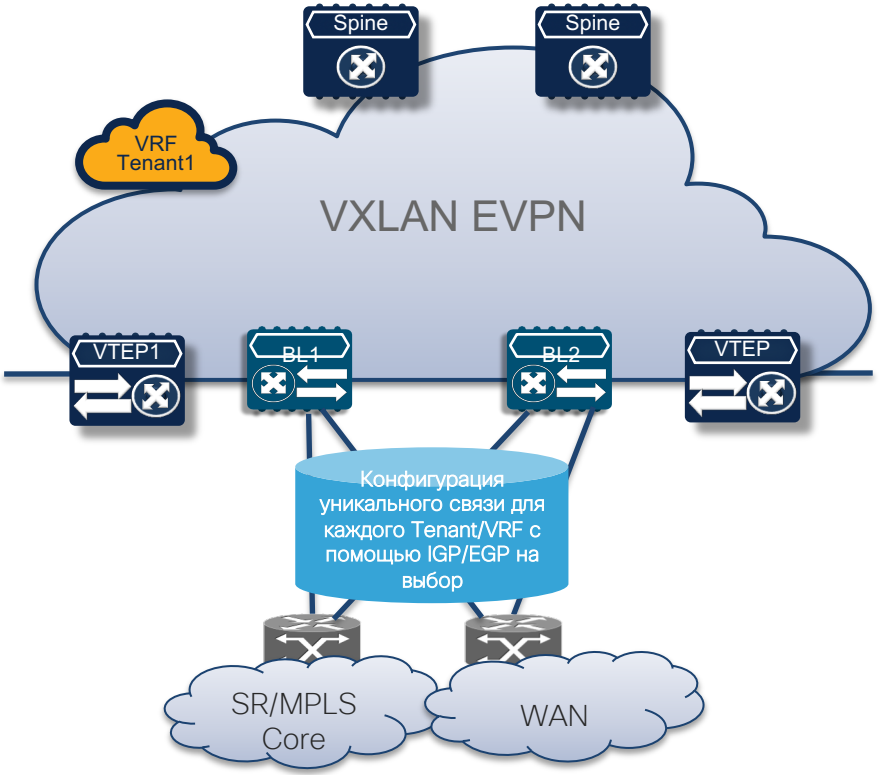
Часть 2: Внешние подключения

Аспекты подключения к внешним сетям

- В рамках VXLAN EVPN фабрики 1x Tenant = 1x VRF/L3VNI
- Существует три варианта маршрутизируемых подключений к внешним сетям:
  1. Организация уникальной связи для каждого Tenant с помощью VRF-Lite на стыке с граничным оборудованием (маршрутизаторы, МСЭ и т.д.)
  2. Использование Centralized Route Leaking или Downstream VNI* и формирование единого стыка с граничным оборудованием
  3. Организация тесной интеграции с внешней мультитранспортной сетью с помощью MPLS hand-off

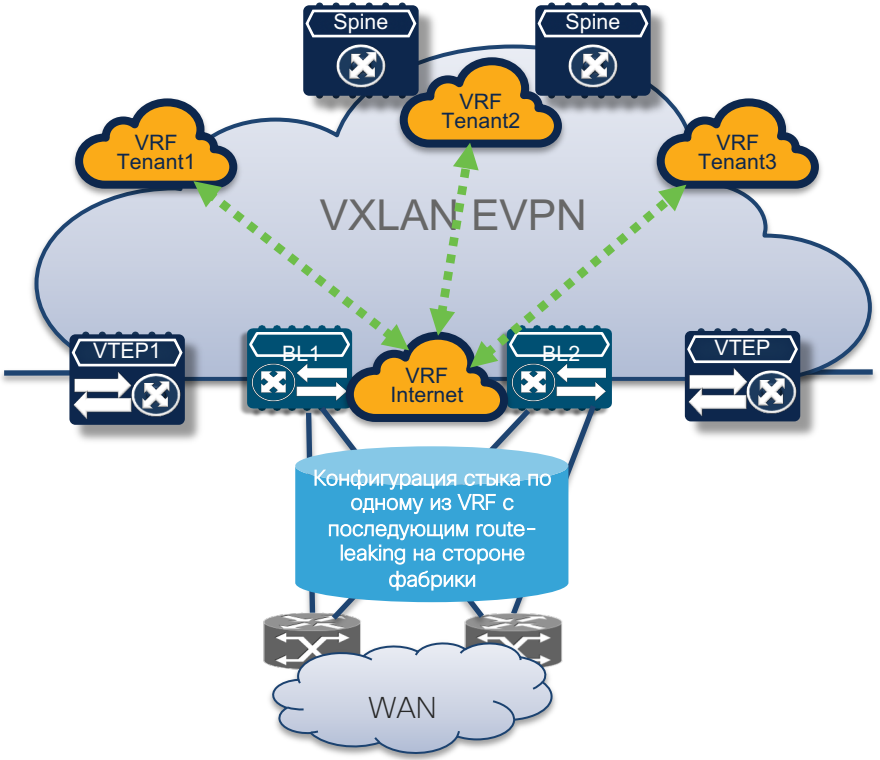
Внешние подключение с помощью VRF-Lite

- Достоинства:
  - Отсутствие специальных требований к граничному оборудованию
  - Простота настройки
- Недостатки
  - Разрастание конфигурации при большом количестве Tenant/VRF



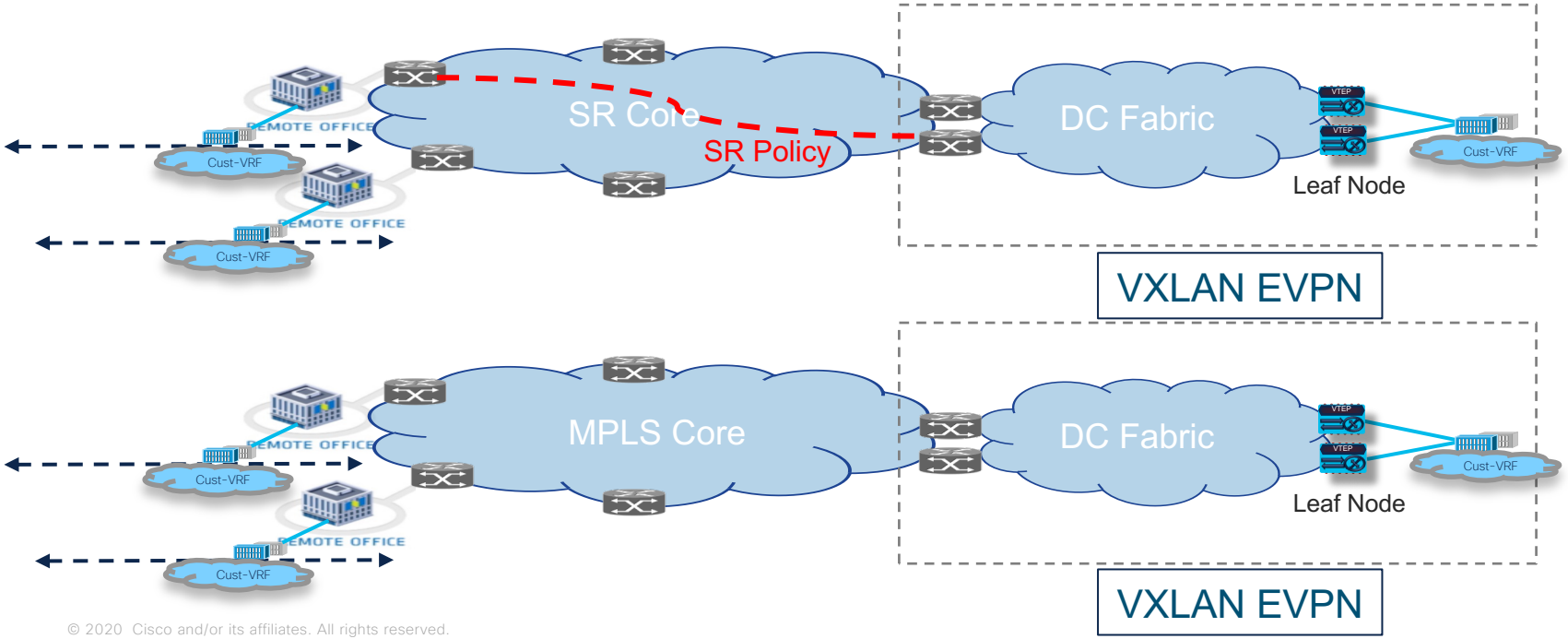
Route-Leaking / Downstream VNI*

- Достоинства:
  - Отсутствие специальных требований к граничному оборудованию
  - Существенно сокращает размер конфигурации
- Недостатки
  - Подходит не для всех сценариев и типов сервисов



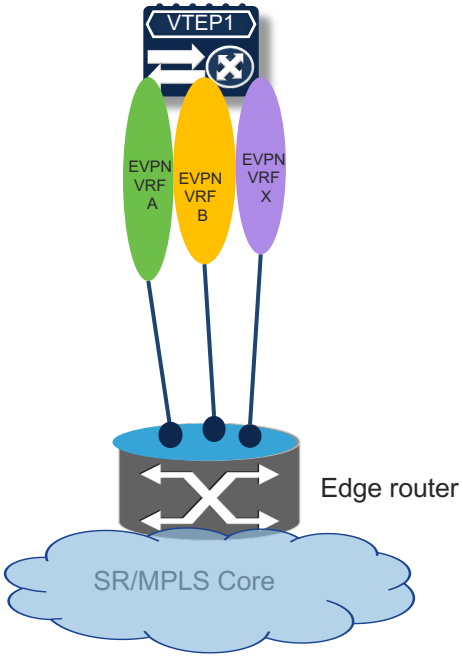
Глубокая интеграция с транспортной сетью

Расширенные возможности при интеграции с мультитранспортными сетями



В чем качественное отличие?

- VXLAN EVPN Border Leaf выполняет две функции:
  1. Граничное устройство на стороне VXLAN EVPN фабрики
  2. Edge устройство в составе мультсервисной сети MPLS
- Сервис настраивается один раз на стороне VXLAN EVPN Border Leaf



Интеграция с SR-MPLS/MPLS LDP

VXLAN EVPN Border Leaf / WAN Edge

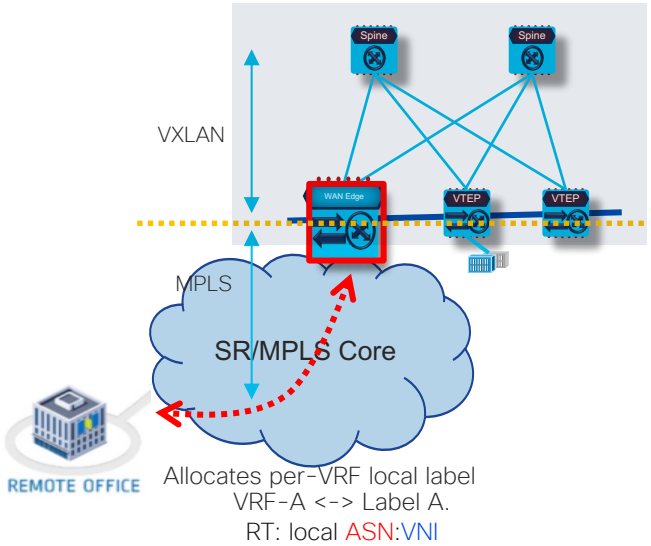
- Отвечает за подключение VXLAN EVPN фабрики к SR/MPLS – L3VPN
- Отвечает за распространение маршрутов (re-originate) и смену data-plane инкапсуляции

Управление Route Target

- Изменяет VXLAN EVPN RT на значения согласованные с мультитранспортной сетью SR/MPLS L3VPN

Управление маршрутами

- Хост-маршруты со стороны VXLAN EVPN фабрики могут быть предварительно агрегированы перед отправкой в L3VPN
- Все L3VPN маршруты полученные из мультитранспортной сети программируются в FIB



Пример конфигурации

Конфигурация BGP

```
router bgp 65535
address-family vpnv4 unicast
 address-family vpnv6 unicast
address-family l2vpn evpn
 neighbor 10.1.1.2
  remote-as 65535
  update-source loopback2
  address-family ipv4 unicast
  address-family l2vpn evpn
   send-community extended
   import vpn unicast reoriginate
 neighbor 20.1.1.2
  remote-as 65535
  update-source loopback2
  address-family ipv4 unicast
  address-family l2vpn evpn
   send-community extended
   import vpn unicast reoriginate
```

VXLAN Spine

VXLAN Spine

Импорт маршрутов полученных от VPNv4/VPNv6 соседей в VXLAN BGP EVPN

Конфигурация BGP (продолжение)

```
neighbor 40.2.1.3
  remote-as 65001
  update-source loopback3
  ebgp-multihop 10
  address-family vpnv4 unicast
   send-community extended
   import l2vpn evpn reoriginate
  address-family vpnv6 unicast
   send-community extended
   import l2vpn evpn reoriginate
vrf vxlan-900001
  address-family ipv4 unicast
   advertise l2vpn evpn
   redistribute direct route-map permitall
   aggregate-address 4.1.1.0/24 summary-only
   aggregate-address 4.2.2.0/24 summary-only
  address-family ipv6 unicast
   advertise l2vpn evpn
   redistribute direct route-map permitall

route-map permitall permit 10
```

PE за пределами ЦОД

Импортирование VXLAN BGP EVPN маршрутов в VPNv4/VPNv6

Настройка агрегации маршрутов

Поддержка на Cisco Nexus 9000

	SR MPLS	MPLS LDP
L3VPN	N9300-FX2/FX3/GX N9K R-Series (Jericho+) N3600 R-Series (Jericho+)	N9K R-Series (Jericho+) N3600 R-Series (Jericho+)

Конфигурация клиентских сервисов с помощью Cisco DCNM

Часть 3: Расширенные функции NX-OS

Предотвращение петель с помощью механизма Loop Prevention

VXLAN EVPN и L2 петли
Определение проблемы

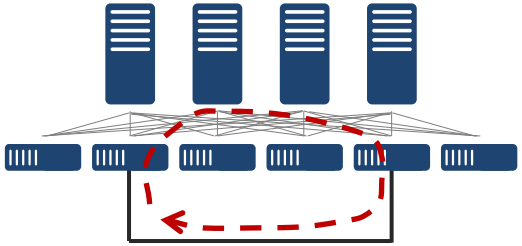
https://blogs.cisco.com/datacenter/detecting-and-mitigating-loops-in-vxlan-networks

VXLAN EVPN и L2 петли
Определение проблемы

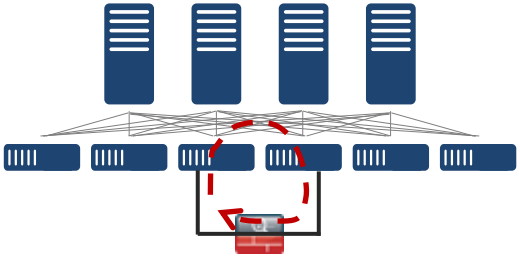
- Современные Layer-2 оверлеи опираются на loop-free топологии
  - Петли могут создаваться из-за ошибок коммутации или подключения внешних устройств
- Петли существенно влияют на производительность сети распространяя broadcast storm
  - Могут приводить к высокой утилизации CPU и сетевой полосы
- Петли могут сделать управление сетевым устройством через mgmt. Interface недоступным
- В сценариях DCI петли могут распространяться между сайтами
  - В рекомендованном к использованию решение на базе VXLAN Multi-Site рекомендуется задействовать функцию BUM Rate-Limiting на устройствах BGW

VXLAN EVPN и L2 петли

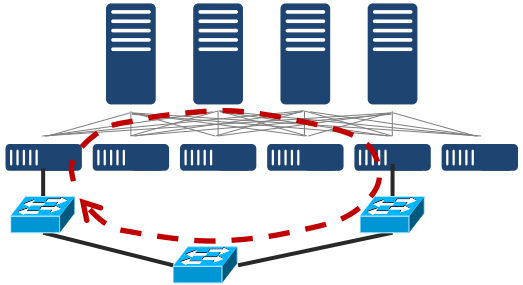
Сценарии приводящие к возникновению сквозных L2 петель



Ошибки в физической коммутации

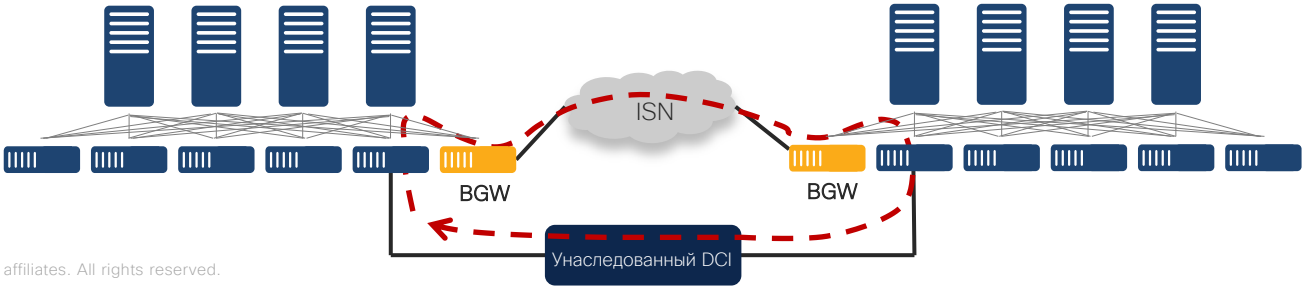


Коммутация через внешние устройства

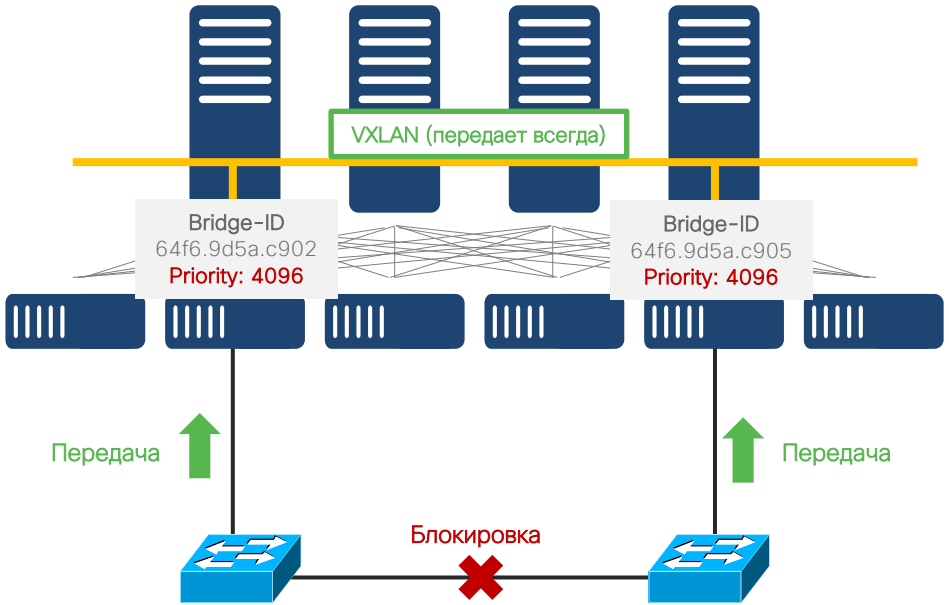


Петли через унаследованные L2 **сети**

Backdoor L2 extension между VXLAN EVPN фабриками



VXLAN EVPN и L2 петли
Как протокол Spanning-Tree может помочь?



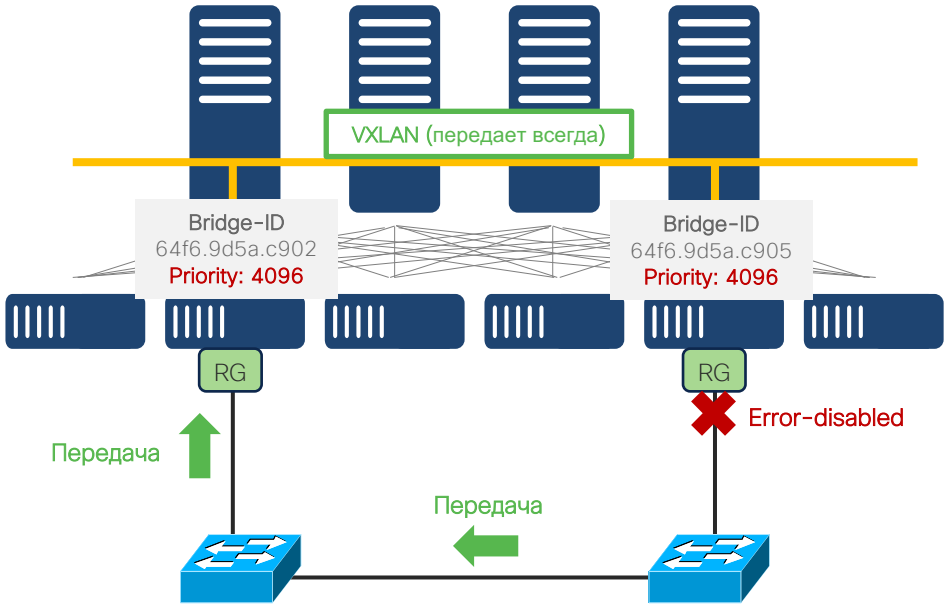
- STP Root перемещается из старой сети на VTEP-ы внутри фабрики
- VTEP-ы настраиваются как STP Root с уникальными Bridge-IDs

  Decision Order for Forwarding:
  1. Priority
  2. Path Length/Cost
  3. Bridge-ID (lowest)
- VXLAN туннель **передает всегда**

  STP BPDU через фабрику не передаются
- Каналы в сторону VTEP не блокируются
- Канал между унаследованными коммутаторами заблокирован STP

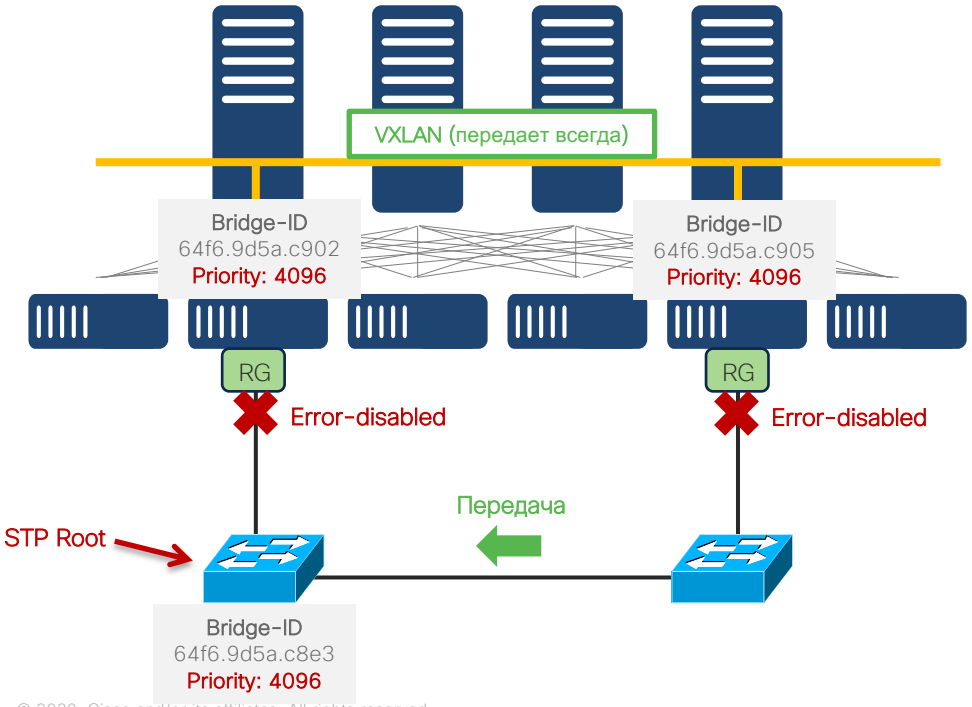
  Трафик между унаследованными коммутаторами передается через VXLAN

VXLAN EVPN и L2 петли
Как протокол Spanning-Tree может помочь?



- Рекомендуется настройка Root-Guard для защиты STP root
- Root-Guard переведет порт в состояние error-disable на том VTEP который менее предпочтительный по сравнению с Root
  - Single Root (lower Bridge-ID)
  - Recovery is Manual/Timer based
- Канал по направлению к VTEP выбирается как STP Root и не блокируется
- Каналы между унаследованными системами не блокируются
  - Путь в сторону единственного Root используется в унаследованной сети для передачи

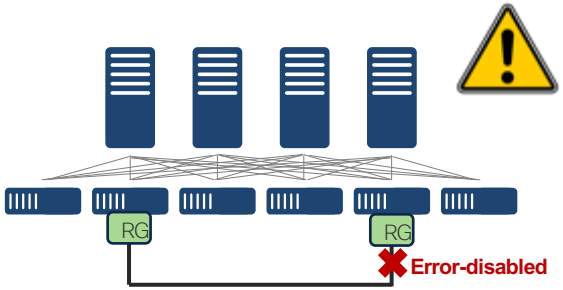
VXLAN EVPN и L2 петли
Как протокол Spanning-Tree может помочь?



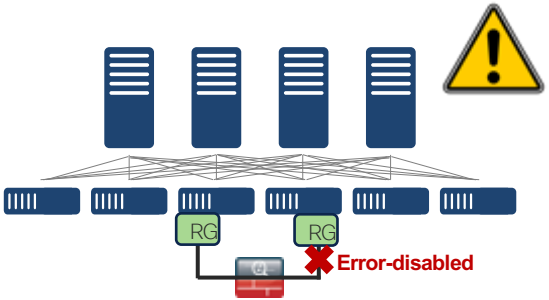
- Рекомендуется настройка Root-Guard для защиты STP root
- Если к унаследованной сети подключается коммутатор с более лучшим Bridge-ID, фнукиця Root-Guard переведет интерфейсы на обоих VTEP в состояние error-disable

  Унаследованная сеть и VXLAN фабрика в таком случае оказываются полностью изолированными
- Такие действия на унаследованной сети не рекомендуются

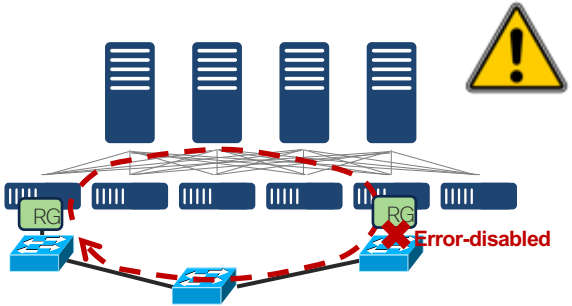
VXLAN EVPN и L2 петли
Как протокол Spanning-Tree может помочь?



Требуется настройка Root-Guard

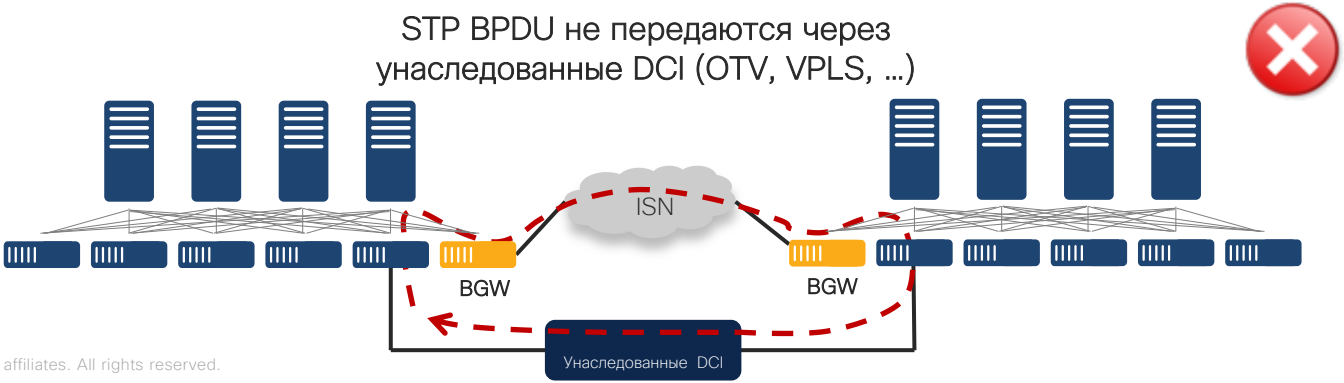


Требуется передача BPDU внешним устройством и настройка Root-Guard



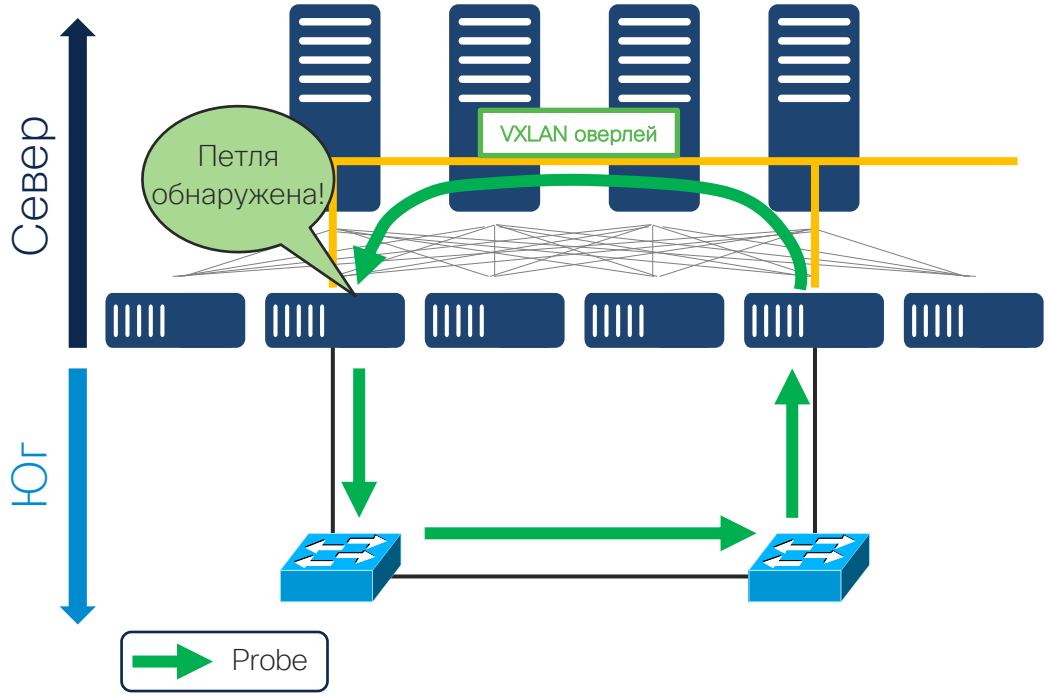
Перемещение STP root на фабрику и настройка Root-Guard

STP BPDU не передаются через унаследованные DCI (OTV, VPLS, …)



Обнаружение и устранение петель
Обзор решения

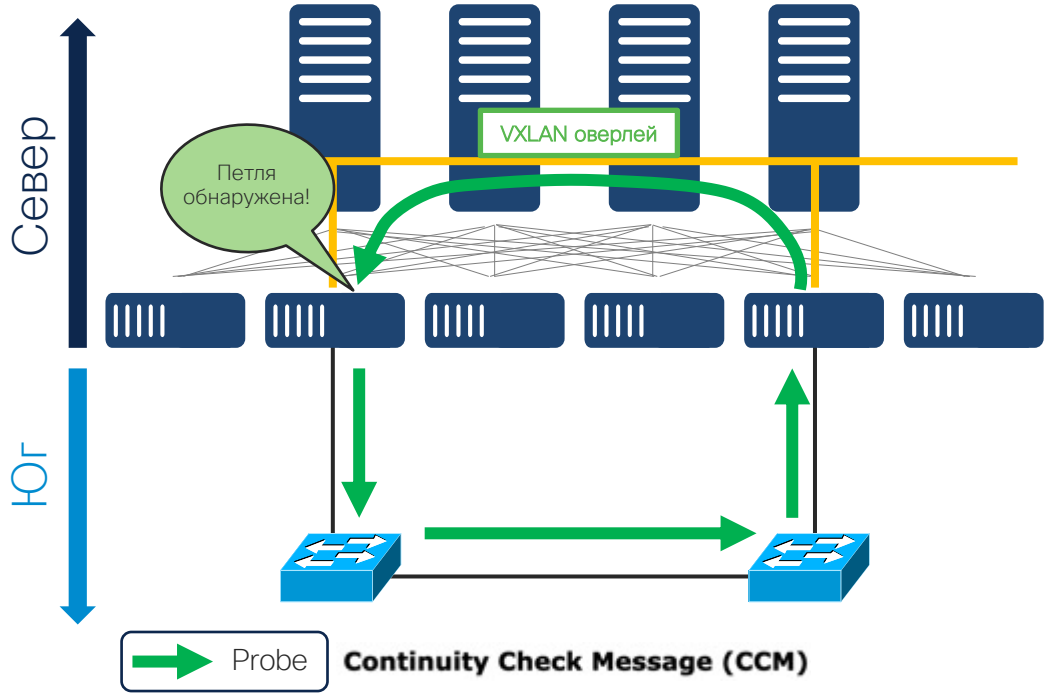
Обнаружение и устранение петель
Обзор решения



- Функция обнаружения и устранения петель использует Ethernet CFM probes для обнаружения петель с южной стороны Leaf коммутаторов
- Функция использует 3 фазы для своей работы
  1. Detection
  2. Mitigation
  3. Recovery

Обнаружение и устранение петель
Фаза 1: Detection

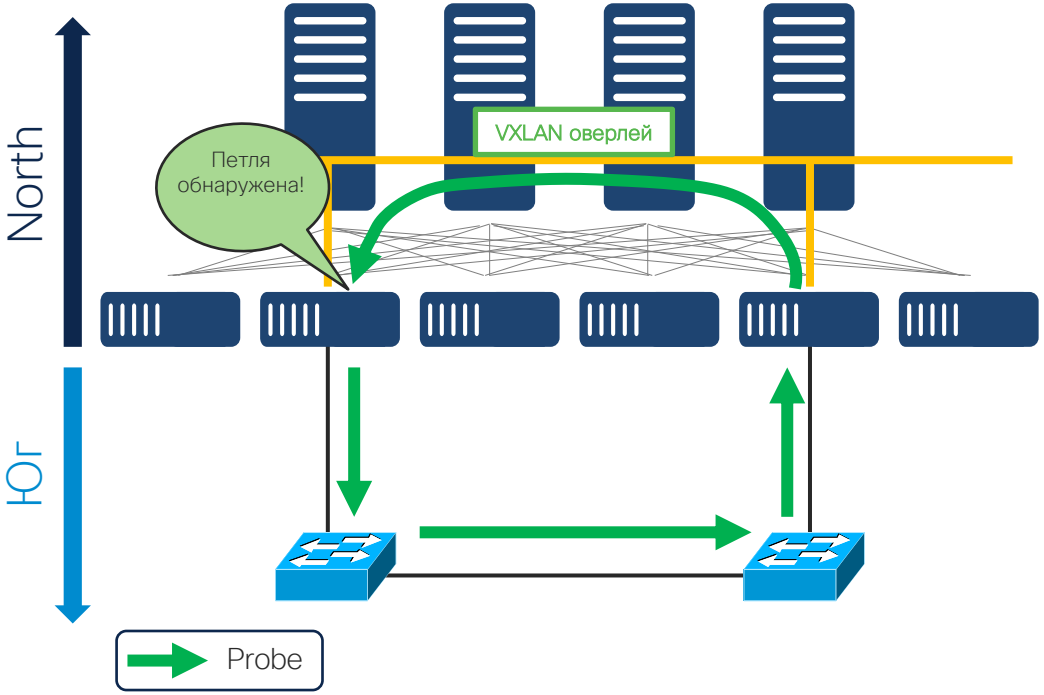
1



- Обнаружение петель срабатывает на Leaf-коммутаторе
- Leaf посылает сообщение CCM Loopback
  - CCM messages это L2 multicast, поэтому они распространяются по южной части топологии и попадают на leaf
  - Это подтверждает наличие петли
- Коммутатор источник CCM пакета сравнивает src-address и VPC role
  - На основе результатов сравнения, коммутатор источник CCM пакета может принять решение, что на сети появилась петля

Обнаружение и устранение петель
Фаза 1: Detection (продолжение)

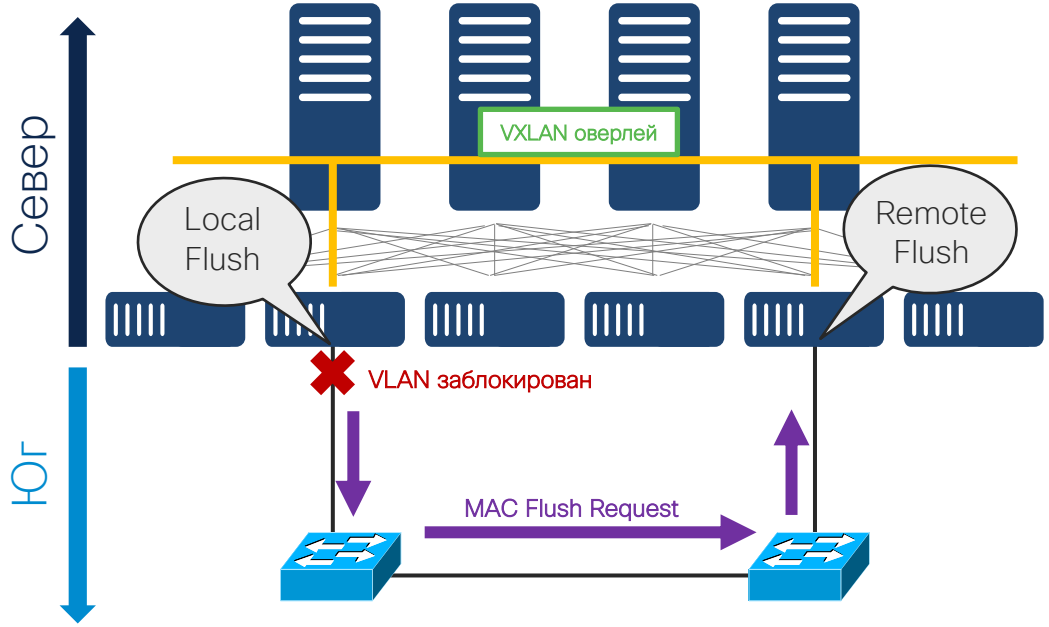
1



- Обнаружение петли срабатывает в следующих случаях:
  - Periodic-Probes: с интервалом (настройка через CLI)
  - Срабатывание по событиям:
    - Состояние порта доступа меняется с down на up
    - Когда VLAN настраивается на access порту
    - Когда дублирующийся MAC обнаруживается EVPN
  - через CLI или REST

Обнаружение и устранение петель
Фаза 2: Mitigation

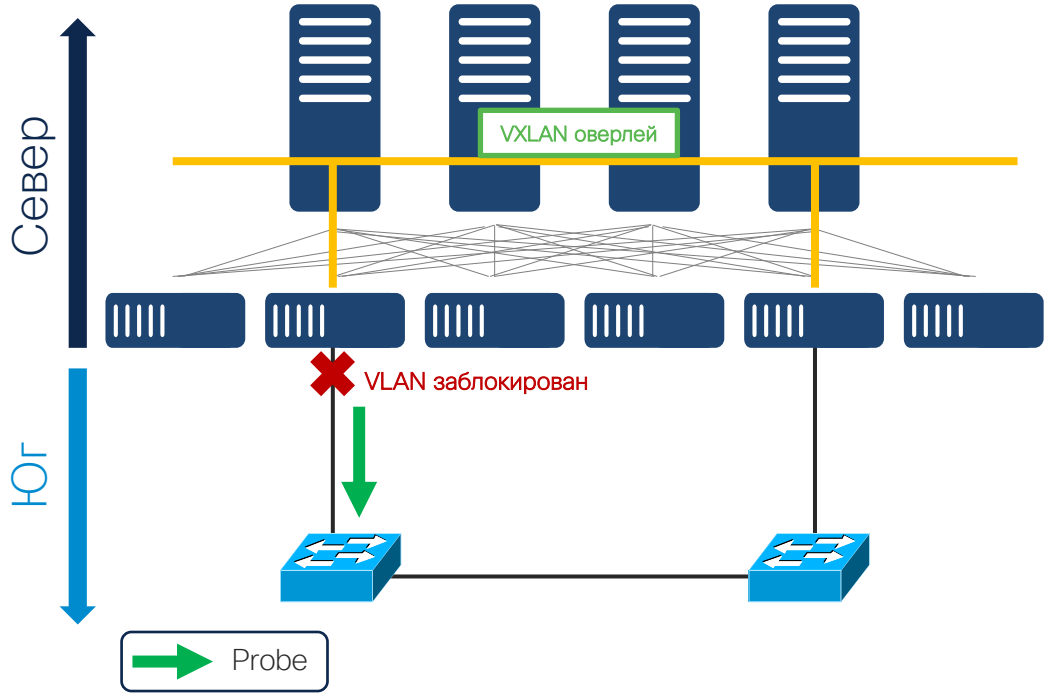
2



- Когда для VLAN обнаруживается петля, то VLAN ассоциированный с портом блокируется
- Очистка ошибочно выученных, из-за появления петли, MAC адресов
  - Local Flush: на коммутаторе Leaf MAC адреса выученные для VLAN очищаются
  - Remote Flush: запрос MAC Flush посылается в южный сегмент сети
  - При получении запроса MAC Flush, коммутатор VTEP на удаленной стороне очистит записи в таблице MAC адресов, которые ассоциированы с исходным VLAN

Обнаружение и устранение петель
Фаза 3: Recovery

3

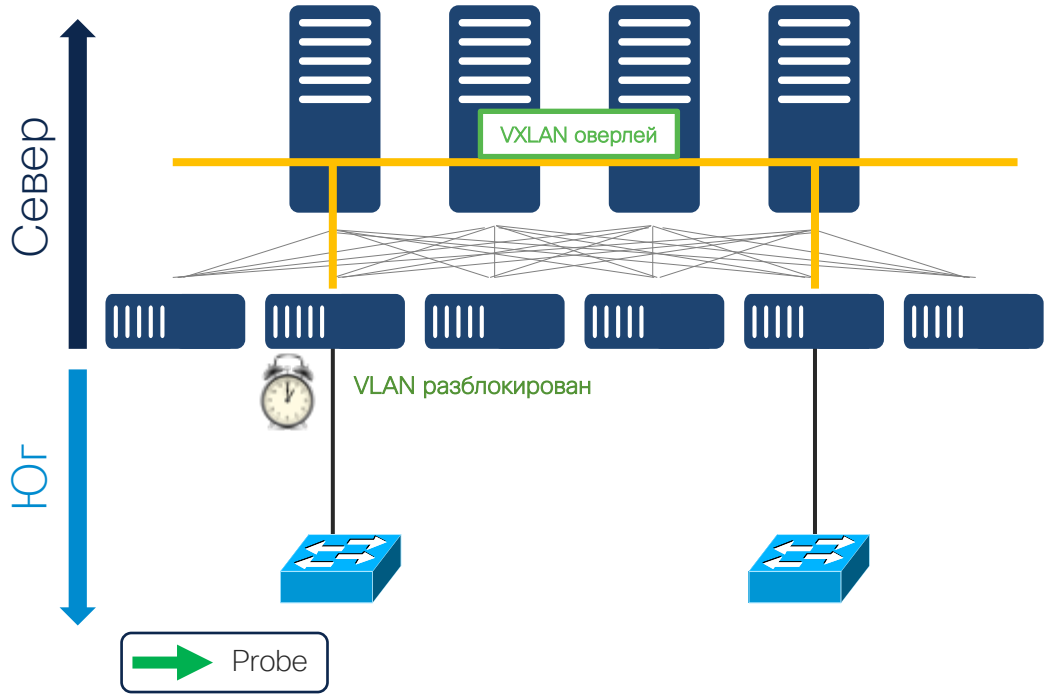


- Поддерживается два способа восстановления:
  - Automatic Recovery

    По окончании настраиваемого recovery time interval, коммутатор Leaf посылает новое Probe-сообщение

    Если probe-сообщение не возвращается назад, то Leaf восстанавливает конфигурацию VLAN на порту после еще одного recovery interval
  - CLI: может вручную восстановить VLAN на интерфейсе

Обнаружение и устранение петель
Фаза 3: Recovery (продолжение)

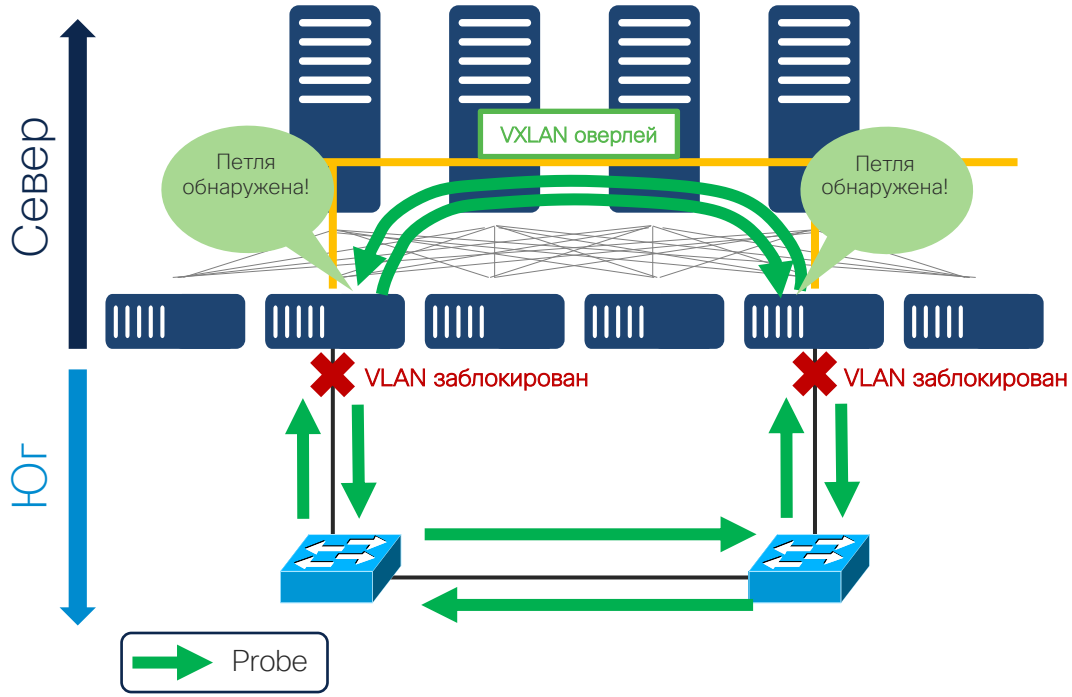


- Поддерживается два способа восстановления:
  - Automatic Recovery

    По окончании настраиваемого recovery time interval, коммутатор Leaf посылает еще одно Probe-сообщение

    Если probe-сообщение не возвращается назад, то Leaf восстанавливает конфигурацию VLAN на порту после еще одного recovery interval
  - CLI: может вручную восстановить VLAN на интерфейсе

Обнаружение и устранение петель
Loop Detection Tie Breaker



- Поскольку коммутаторы шлют probes пакеты независимо, возможна ситуация когда они это сделают одновременно
  - Оба probe пакета вернутся к своим отправителям, поскольку оба они решат что на сети имеется петлям, VLAN будет заблокирован на обоих портах
- Чтобы уменьшить вероятность возникновения этого события алгоритм использует рандомизацию при создании этих probe пакетов

Обнаружение и устранение петель
Ограничения и рекомендации

- Функция VXLAN EVPN loop detection and mitigation не поддерживается совместно со следующими функциями:
  - Private VLAN
  - VLAN translation
  - VXLAN Cross Connect
  - Q-in-VNI
- Порты или VLAN(ы) использующие эти функции должны быть исключены из работы функции VXLAN EVPN loop detection and mitigation
  - Можно воспользоваться командой **`disable {vlan <vlan-range>} [port <port-range>]`**

Обнаружение и устранение петель
Настройка

Обнаружение и устранение петель
Поддерживаемые платформы

- Следующие платформы поддерживают VXLAN EVPN loop detection and mitigation функцию:
  - Cisco Nexus 9332C и 9364C
  - Cisco Nexus 9300-EX
  - Cisco Nexus 9300-FX/FX2/FXP/FX3
  - Cisco Nexus 9300-GX
  - Cisco Nexus 9500 коммутаторы с линейными картами -EX/FX

Обнаружение и устранение петель
Feature Prerequisites

- Включить NGOAM

```
feature ngoam
```

- Выделить память TCAM ing-sup region

```
hardware access-list tcam region ing-racl* <default value – 256>
hardware access-list tcam region ing-sup 768
```

*Требуется 768 для ing-sup (по умолчанию 512), ing-racl (или egr-racl) примеры того где можно взять ресурсы

Обнаружение и устранение петель
Настройка

(config)# **ngoam loop-detection**
Enables/disables loop-detection for all VLANs/Ports. The loop detection would be disabled by default.

(config-ng-oam-loop-detection)# **disable {vlan <vlan-range>} [port <port-range>]**
Disables the loop detection and brings any loop-detected ports up. “no disable” brings the active monitoring of these VLANs/ports back

(config-ng-oam-loop-detection)# **periodic-probe-interval <val>**
This is the time-gap between sending two periodic probes. The range is from 60 seconds to 3600 seconds (60 minutes). The default value is 300 seconds (5 minutes)

(config-ng-oam-loop-detection)# **port-recovery-interval <val>**
Once the port/vlan is down, run this timer before sending recovery probes. The range is from 300 seconds to 3600 seconds (60 minutes). The default value is 600 seconds (10 minutes).

(exec) **ngoam loop-detection probe {vlan <vlan-range>}[port <port-range>]**
Sends a probe on the specified vlan/port and notifies whether the probe was successfully sent out or not back to the user

(exec) **ngoam loop-detection bringup {vlan <vlan-range>}[port <port-range>]**
Brings up the specified vlan/port that was blocked earlier. No action is taken if the port/vlan was not blocked earlier

Обнаружение и устранение петель
Show команды

show ngoam loop-detection status [history][vlan <vlan-range>] [port <port-range>]

- Display the status of the Blocked/Forwarding port/vlans
- Without history option this command displays only blocked ports
- History option will display blocked and forwarding ports
- Display the status on a given vlan/port (number of CCMs sent, loops detected, ports brought down, …)
- Sample output:

```
show ngoam loop-detection status
VlanId  Port       Status         NumLoops      DetectionTime              ClearedTime
=============================================================================================
1000    Eth1/40    BLOCKED          13          Thu May 14 03:33:57 2020  Thu May 14 03:33:50 2020
show ngoam loop-detection status history
VlanId  Port       Status         NumLoops      DetectionTime              ClearedTime
=============================================================================================
1000    Eth1/40    BLOCKED          13          Thu May 14 03:33:57 2020  Thu May 14 03:33:50 2020
1000    Eth1/41    FORWARDING       1           Thu May 14 00:36:29 2020  Thu May 14 01:45:28 2020
```

Обнаружение и устранение петель
Show команды

Display the summary of parameters/stats:

show ngoam loop-detection summary

```
        Loop detection:enabled
        Periodic probe interval: 3600
        Port recovery interval: 3500
        Number of vlans: 1
        Number of ports: 1
        Number of loops: 1
        Number of ports blocked: 1
        Number of vlans disabled: 0
        Number of ports disabled: 0
        Total number of probes sent: 214
        Total number of probes received: 102
        Next probe window start: Thu May 14 15:14:23 2020 (0 seconds)
        Next recovery window start: Thu May 14 15:54:23 2020 (126 seconds)
```

Clear probe statistics in summary output and the ports already in forwarding state from the status output:

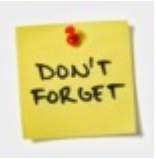
clear ngoam loop-detection statistics

Downstream VNI

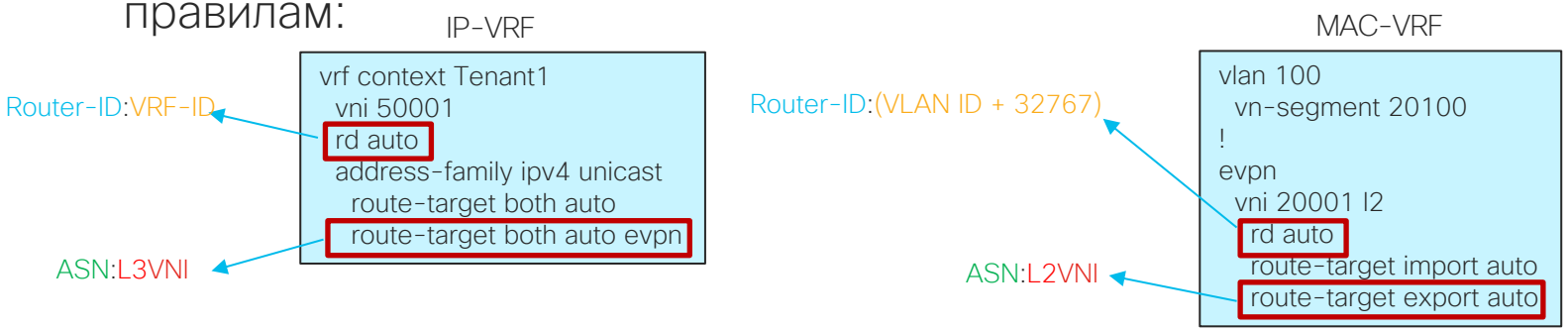
VXLAN EVPN и Downstream VNI
Обзор

- В более ранних версиях NXOS, настройки VNI должны быть консистентны на всех узлах VXLAN EVPN сети для обеспечения Layer 2 и Layer 3 взаимодействия между хостами
- Cisco NX-OS Release 9.3(5) реализует функцию VXLAN EVPN и Downstream VNI, который предоставляет альтернативный способ обмена информацией о MAC/IP (и ассоциированных с ними L2 и L3 VNI) для VXLAN фабрик (intra-site и inter-site)
- Корректная настройка значений Route-Target (для IP-VRF и MAC-VRF) разных VTEP узлах является ключевым фактором успеха “ассиметричных” взаимодействий

Route Distinguisher и Route Target
Автоматическая генерация значений с 2-byte ASN



- Рекомендуемым способом является использование auto generated значений route-distinguisher и route-target
- Эти значения автоматически создаются в EVPN по следующим правилам:



Примеры:

- IP-VRF с BGP Router ID 192.0.2.1 и VRF ID 6 → RD 192.0.2.1:6
- MAC-VRF с BGP Router ID 192.0.2.1 и VLAN 20 → RD 192.0.2.1:32787
- IP-VRF с ASN 65001 и L3VNI 50001 → RT 65001:50001
- MAC-VRF с ASN 65001 и L2VNI 30001 → RT 65001:30001

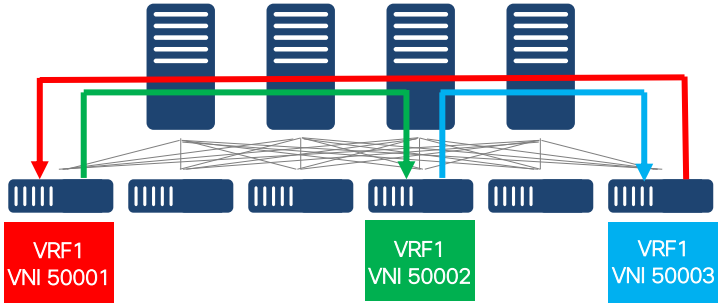
VXLAN EVPN и Downstream VNI
Настройка

- Никаких новых команд для функции не вводится
- Используются существующие route-target команды
- Для внедрения ассиметричных VNI когда используется route-target auto, route targets между двумя VTEP больше не совпадают из-за различных значений VNI

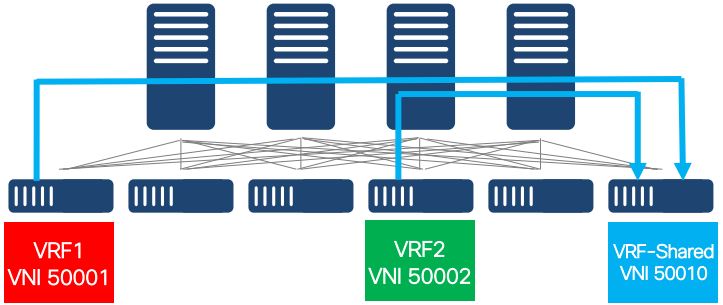
  Требуется ручная настройка route-target для поддержки «склеивания» VNI
- Выполнение таких настроек через DCNM планируется в CY21 (можно использовать free-form шаблоны пока функция не вышла)

VXLAN EVPN и Downstream VNI
Сценарии использования

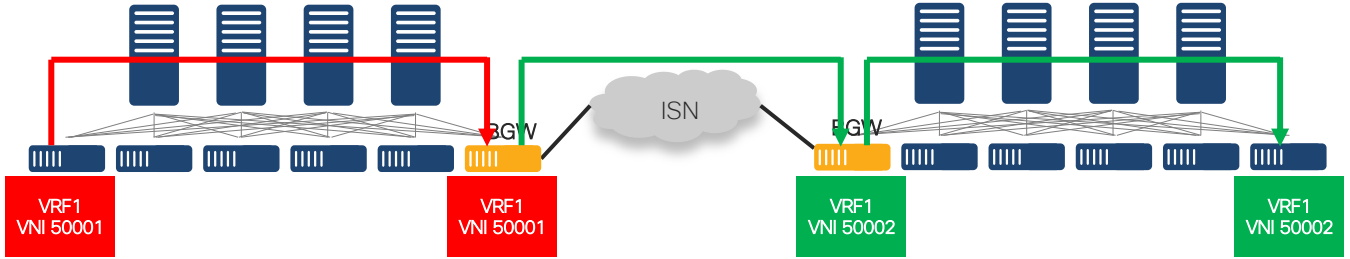
Asymmetric VNI взаимодействие (Intra-Fabric)



Shared Services VRFs (Route-Leaking)



Multi-Site и Asymmetric VNI



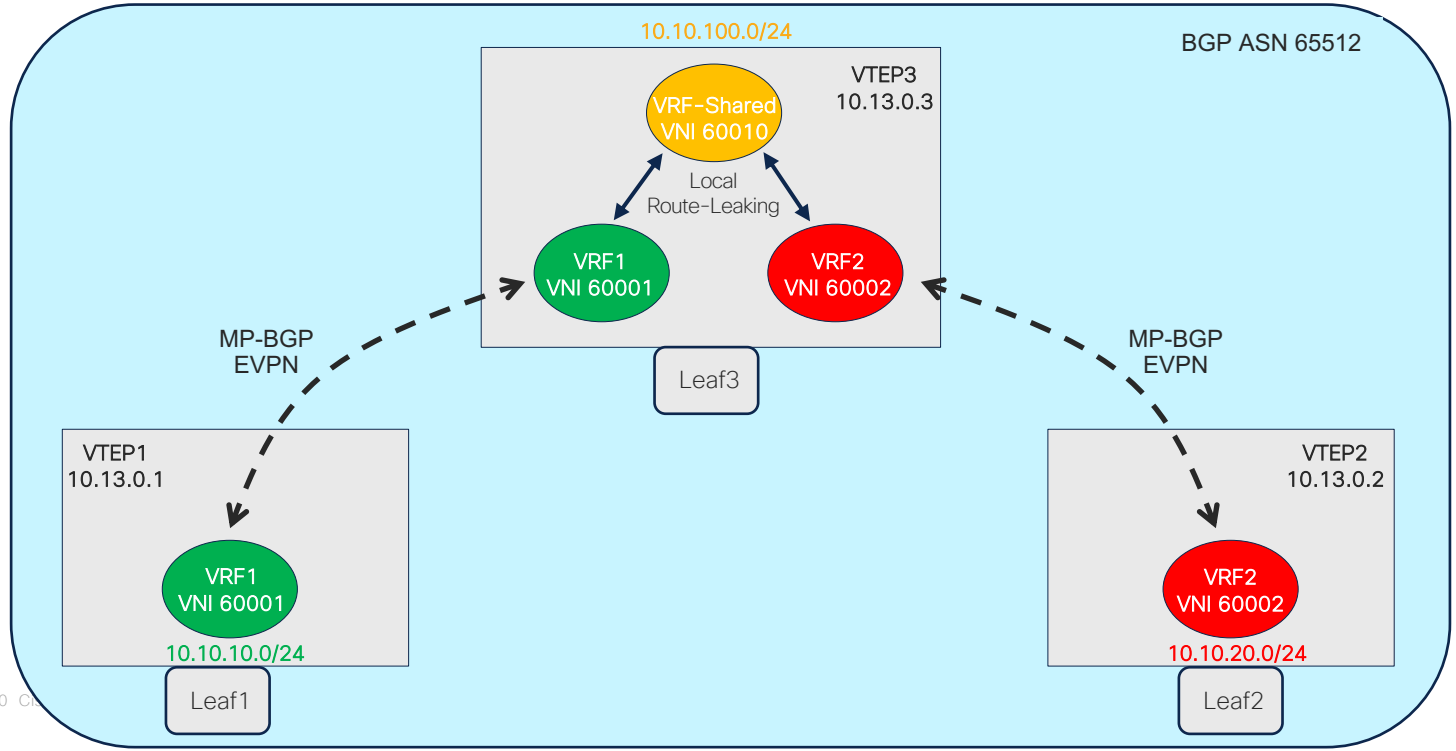
Организация доступа к инфраструктурным (shared) сервисам с помощью Downstram VNI

VXLAN EVPN и Downstream VNI
Сценарий 2: Shared Services VRFs

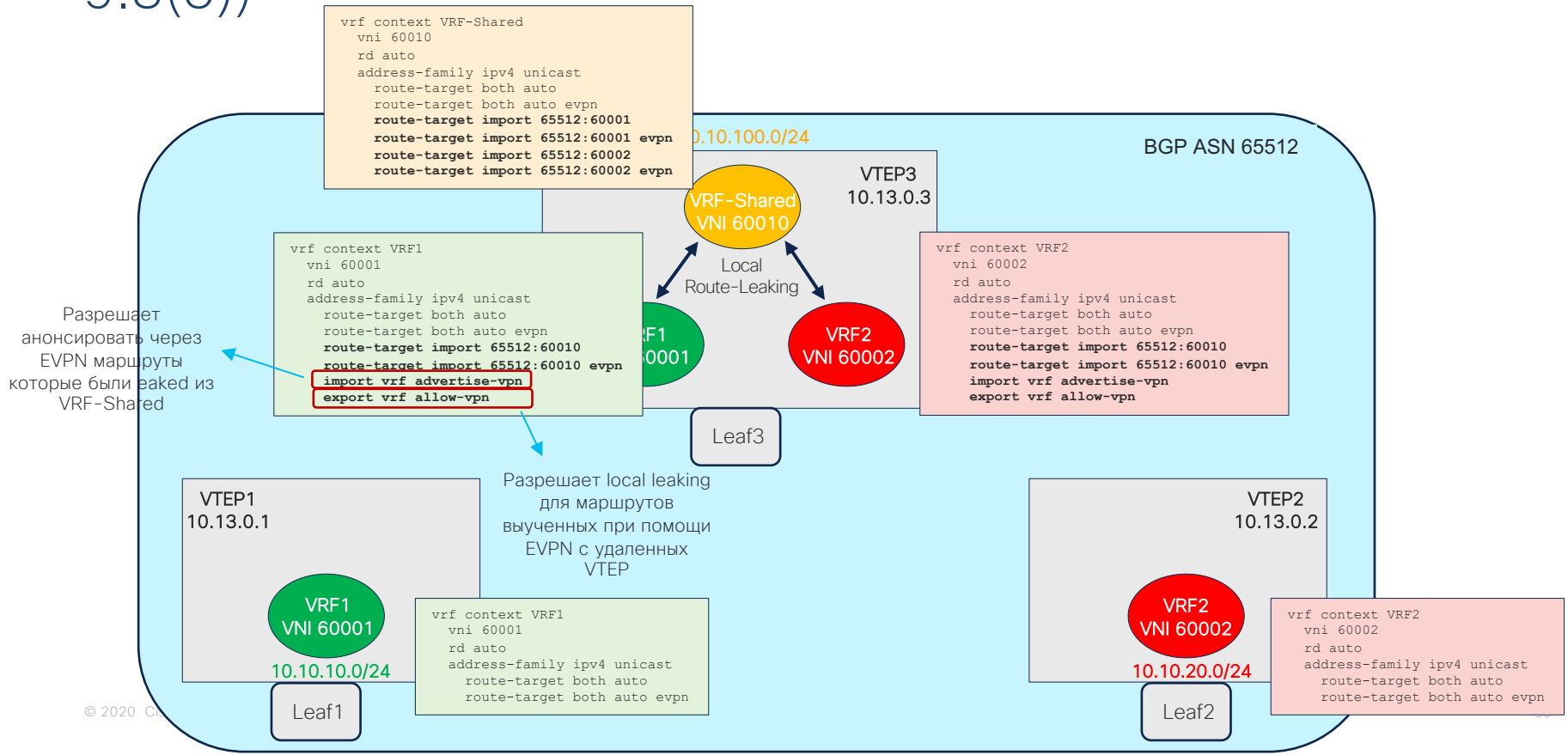
- VXLAN EVPN и downstream VNI поддерживает shared services VRF
- Реализуется при помощи импорта префиксов из разных VRF в один shared VRF и поддержки соответствующих значений для каждого нисходящего L3VNI
- Сценарии для Route-Leaking
  1. До версии NXOS 9.3(5): Централизованный Route-Leaking
  2. NXOS 9.3(5): Централизованный Route-Leaking и Downstream VNI
  3. NXOS 9.3(5): “Распределенный” Route-Leaking и Downstream VNI

До релиза NXOS 9.3(5)
Централизованный Route-Leaking

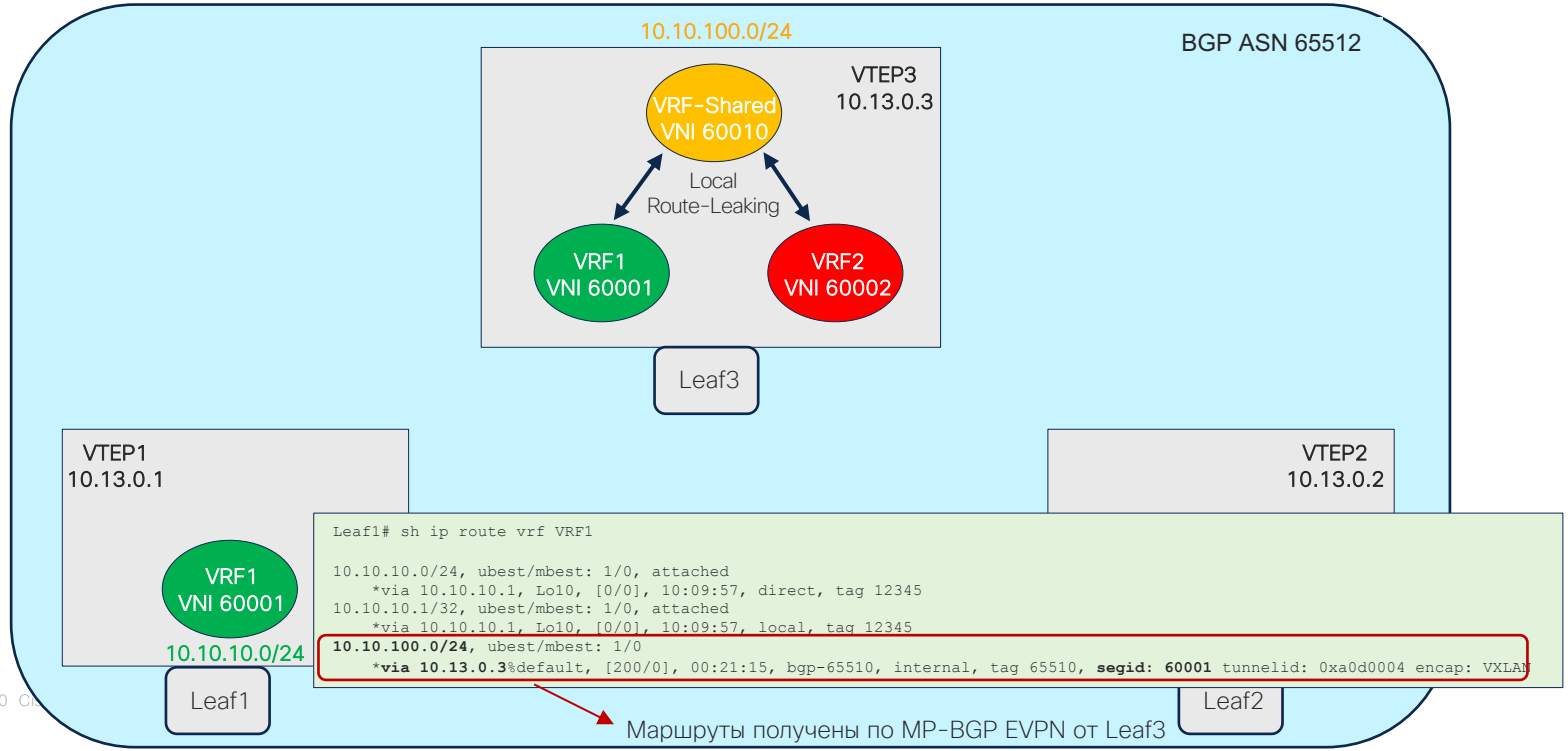
Централизованный Route-Leaking (До версии NXOS 9.3(5))



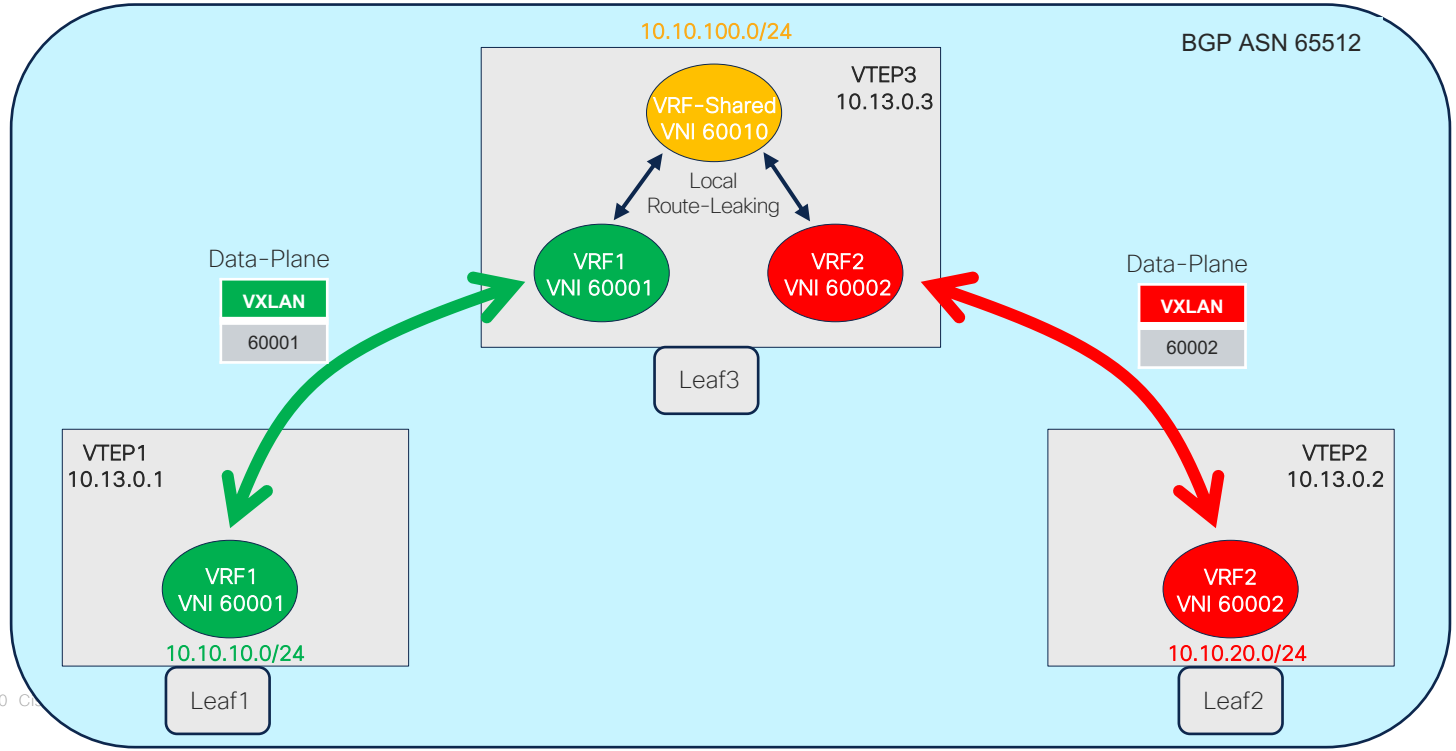
Централизованный Route-Leaking (До версии NXOS 9.3(5))



Централизованный Route-Leaking (До версии NXOS 9.3(5))



Централизованный Route-Leaking (До версии NXOS 9.3(5))

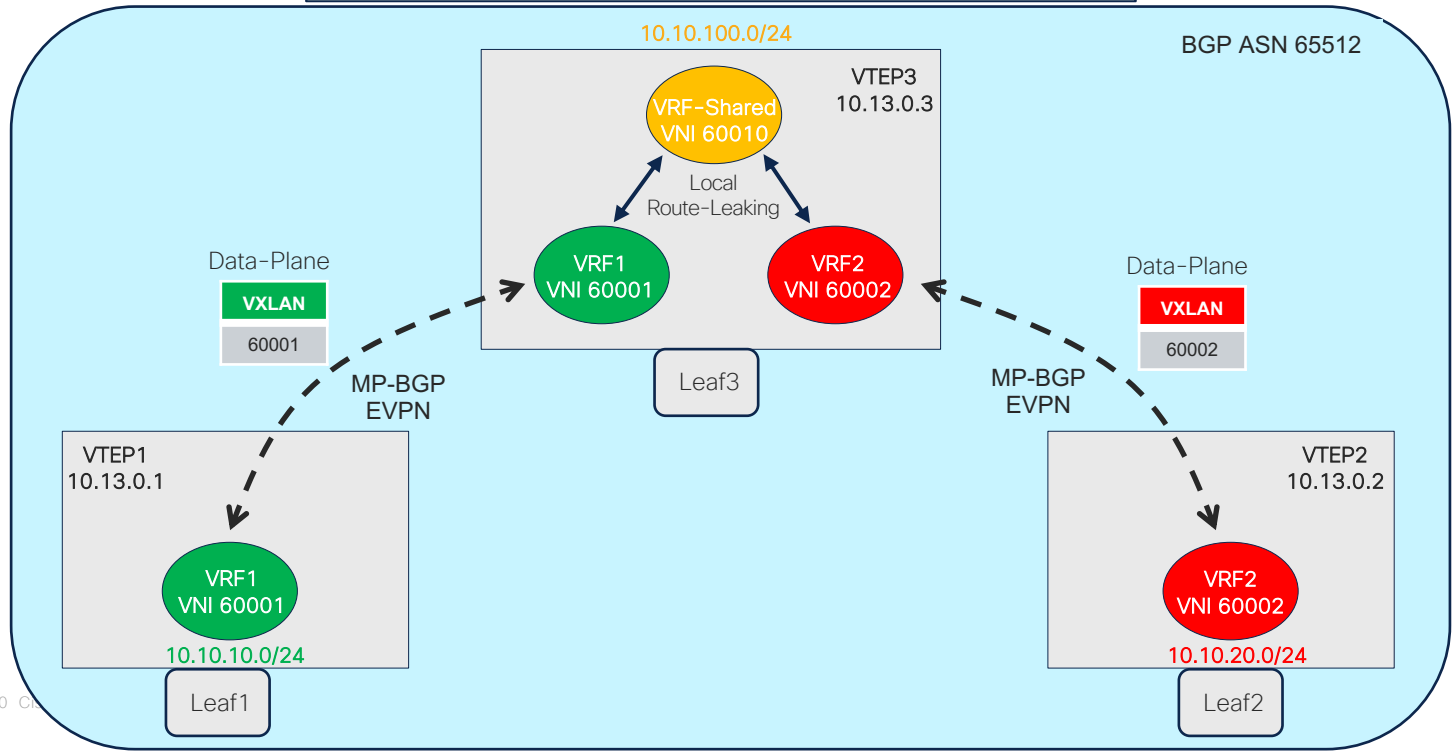


NXOS 9.3(5)
Централизованный Route-Leaking и Downstream VNI

Централизованный Route-Leaking и Downstream VNI

9.3(5)

После обновления конфигурацию можно оставить, поведение системы не изменится

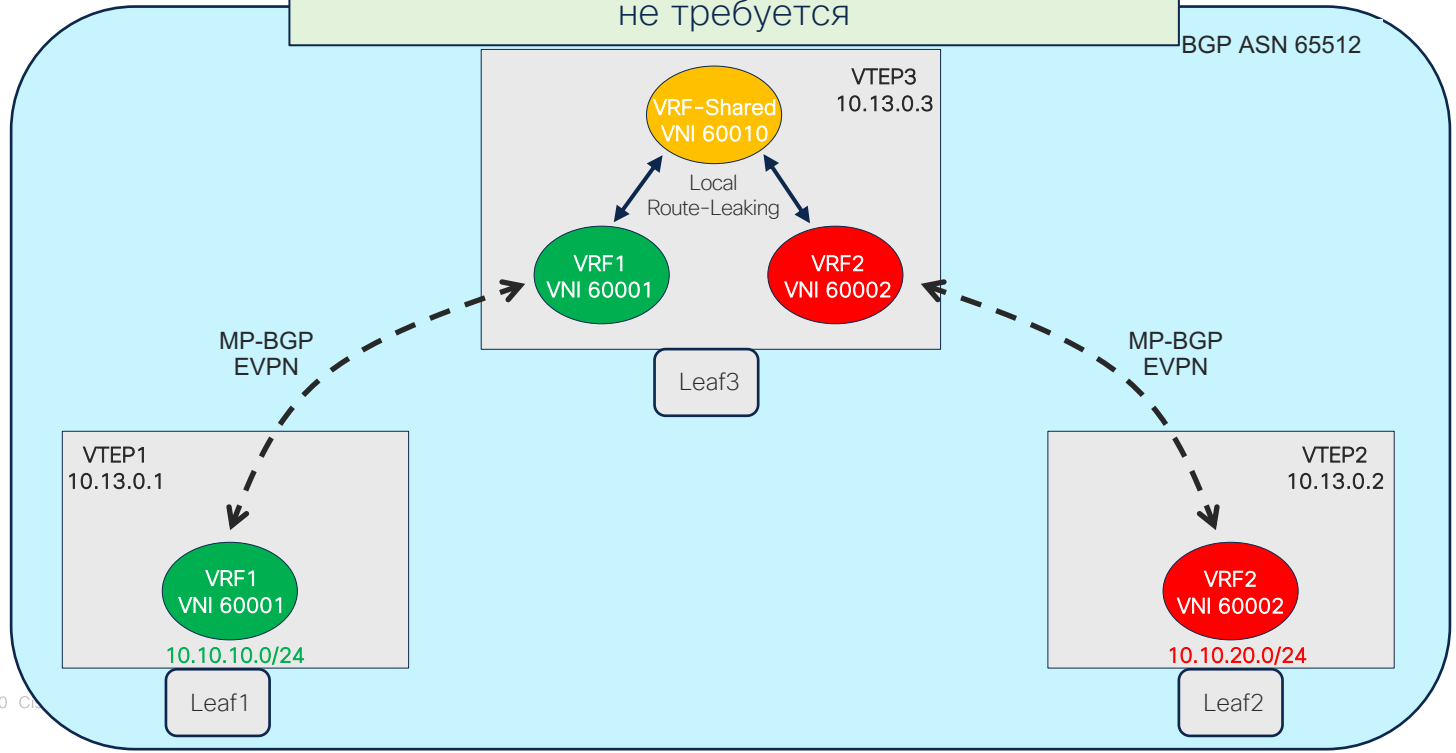


NXOS 9.3(5)
“Распределенный” Route-Leaking и Downstream VNI

“Распределенный” Route-Leaking и Downstream VNI

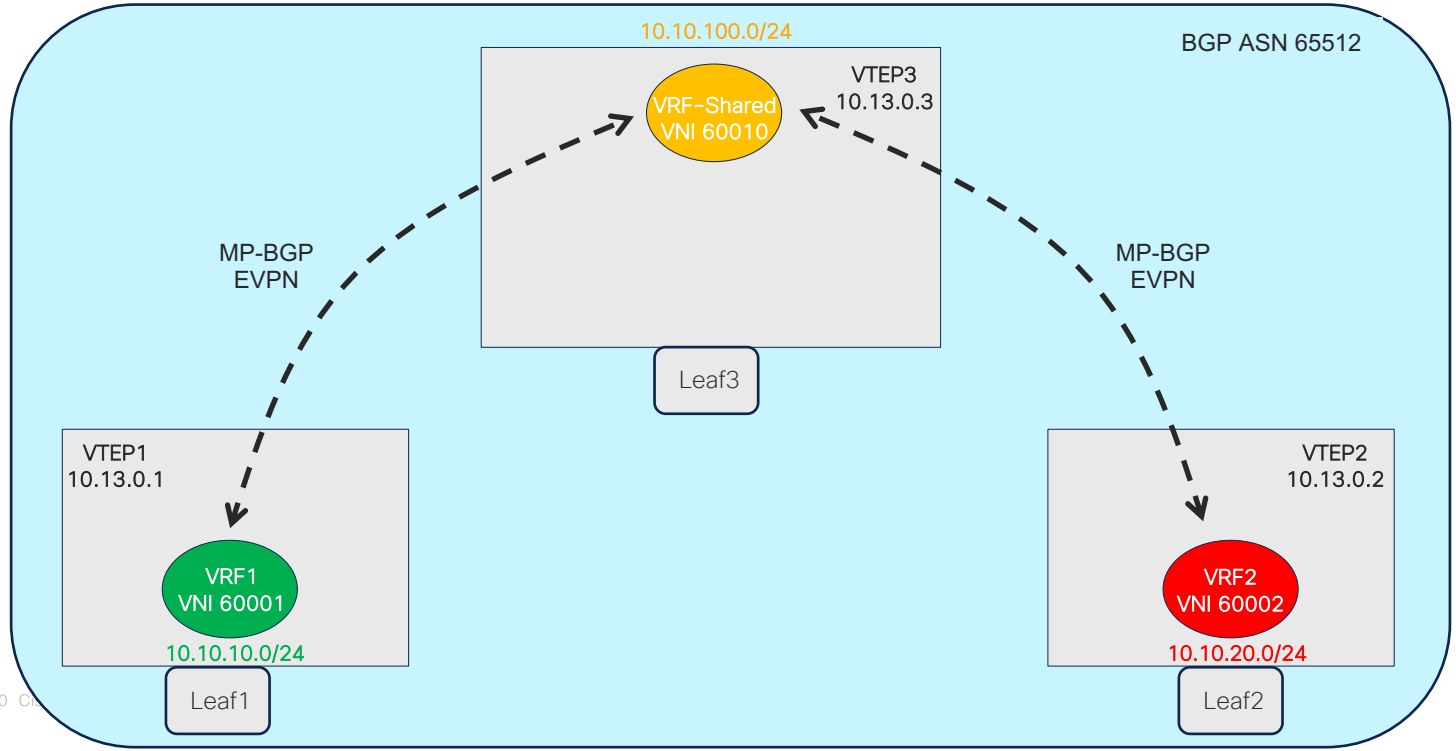
9.3(5)

Основная цель “упростить” конфигурацию при помощи удаления VRF с VTEP-ов на которых он не требуется



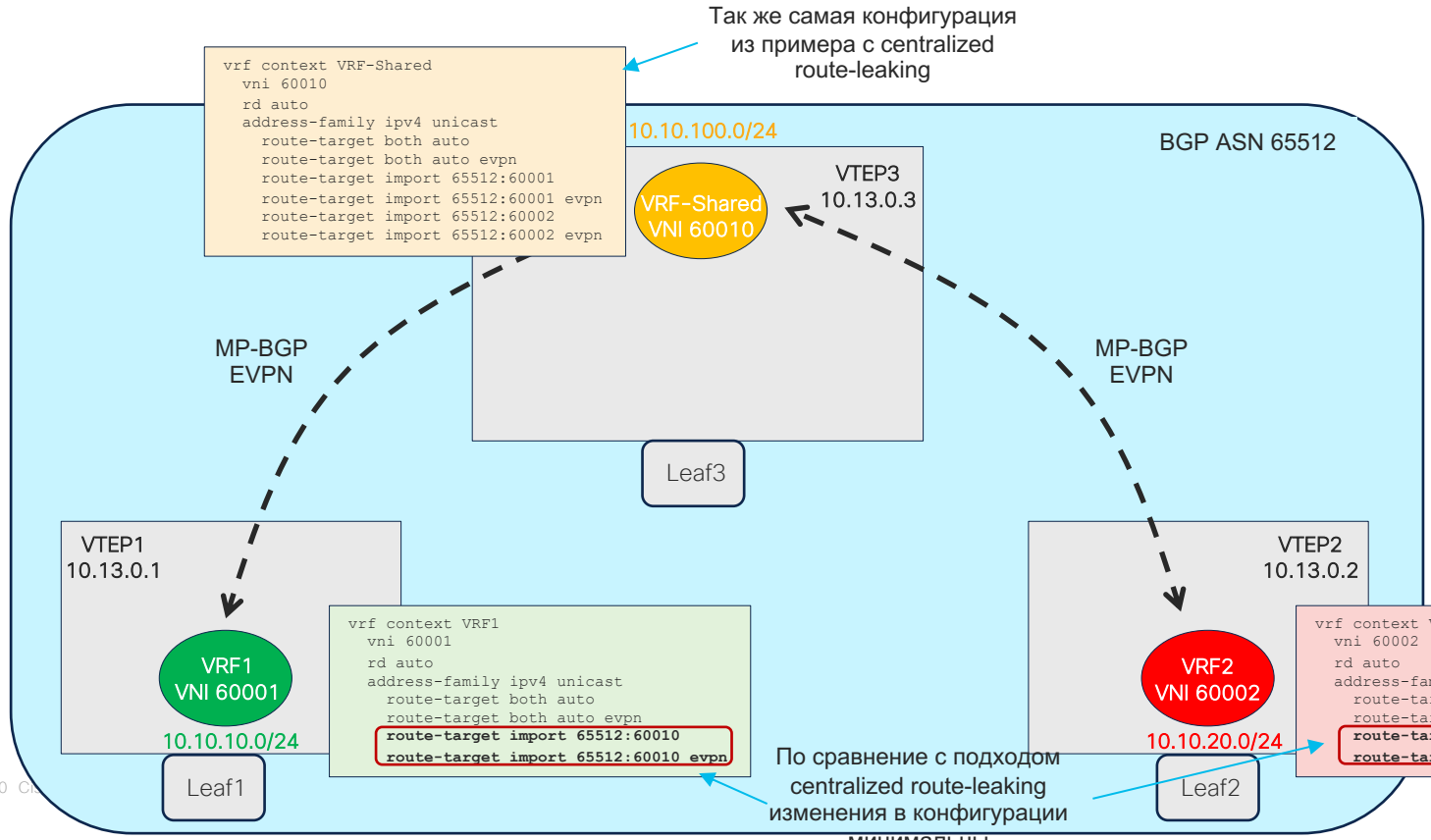
“Распределенный” Route-Leaking и Downstream VNI

9.3(5)



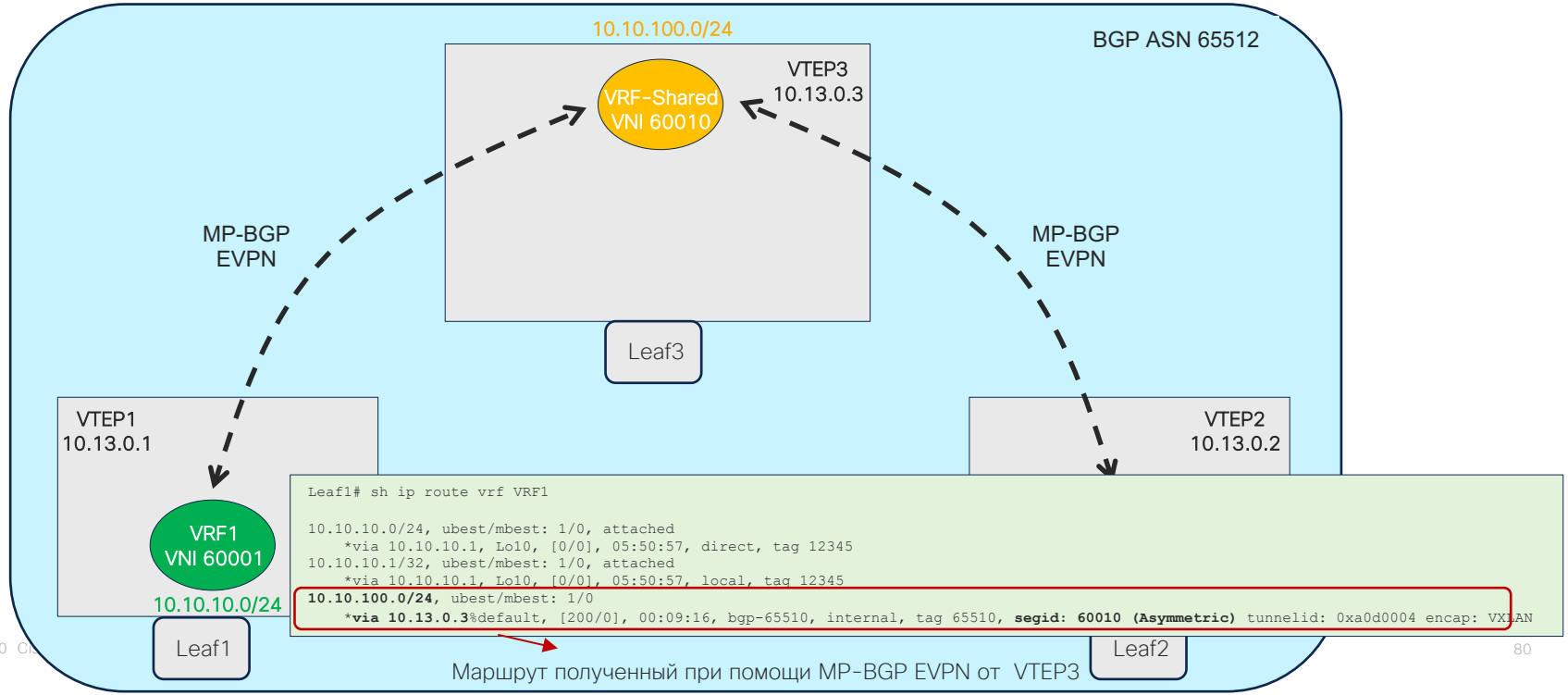
“Распределенный” Route-Leaking и Downstream VNI

9.3(5)



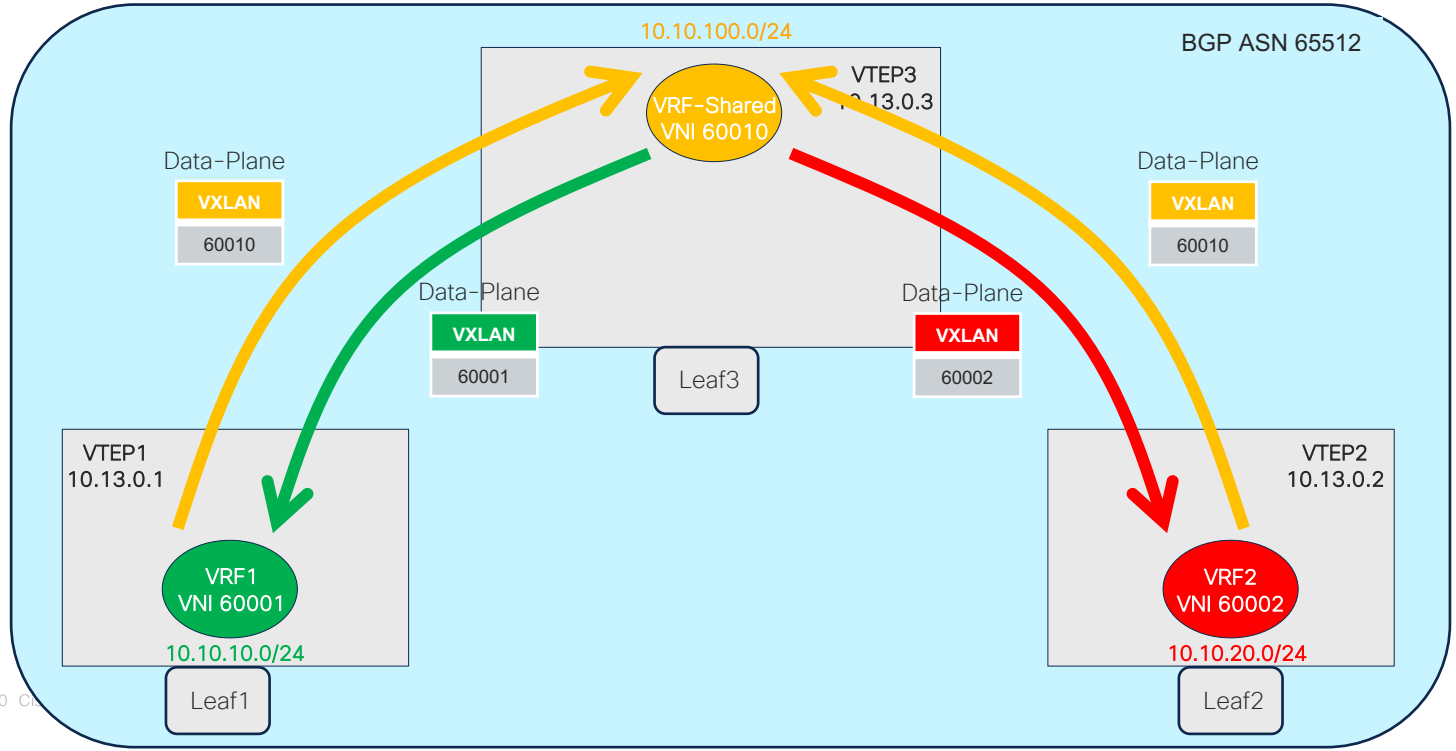
“Распределенный” Route-Leaking и Downstream VNI

9.3(5)



“Распределенный” Route-Leaking и Downstream VNI

9.3(5)



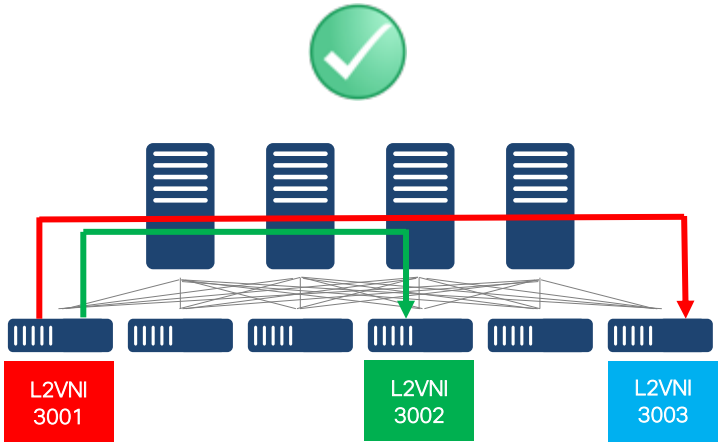
VXLAN EVPN и Downstream VNI
Рекомендации/Ограничения

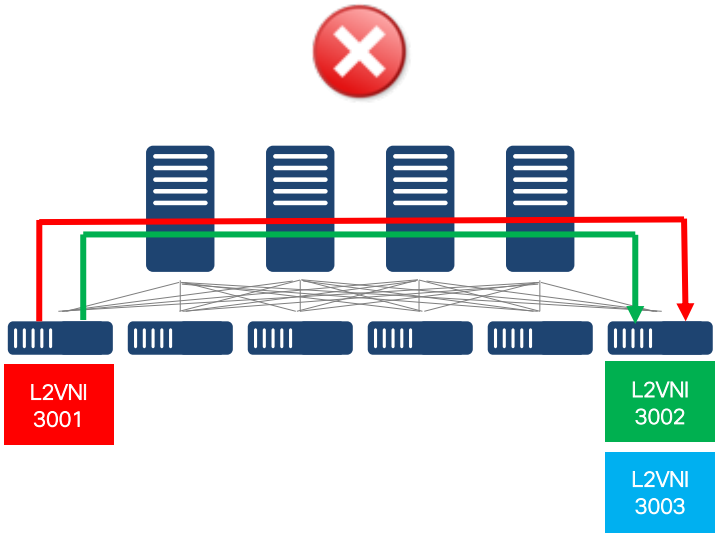
VXLAN EVPN и Downstream VNI
Рекомендации и ограничения

- VXLAN EVPN и Downstream VNI поддерживается на Cisco Nexus 9332C, 9364C, 9300-EX, 9300-FX/FX2/FXP/FX3 и 9500 с линейными картами -EX/FX
- Функция поддерживается только для внедрений с IPv4 underlay
- Для ассиметричных Layer 2 взаимодействий (например между сегментами с разными L2 VNI), в релизе NXOS 9.3(5) поддерживается только ingress replication для распространения BUM трафика
- Настройки должны быть консистентными между vPC узлами (внутри фабрики) а так же между border gateways которые принадлежат одному сайту
- Унаследованные сайты (например VXLAN EVPN фабрики с ПО NXOS 9.3(4) или более ранние) должны взаимодействовать с фабриками на основе VXLAN EVPN версии NXOS 9.3(5) с использованием симметричных VNI

VXLAN EVPN и Downstream VNI
Рекомендации и ограничения (продолжение)

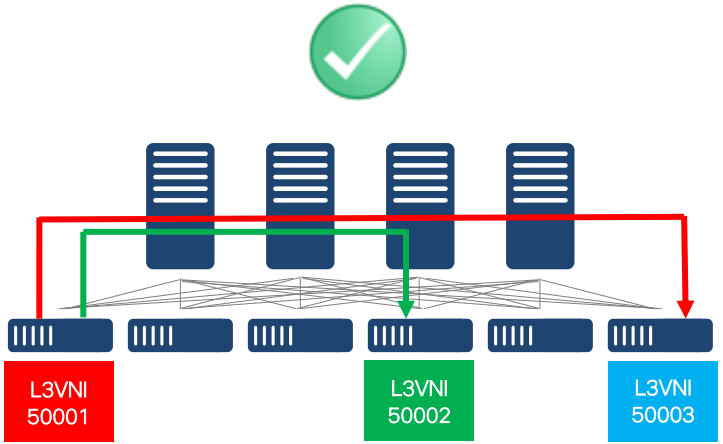
- Локальный L2VNI может быть ассоциирован только с одним L2VNI на одном VTEP peer

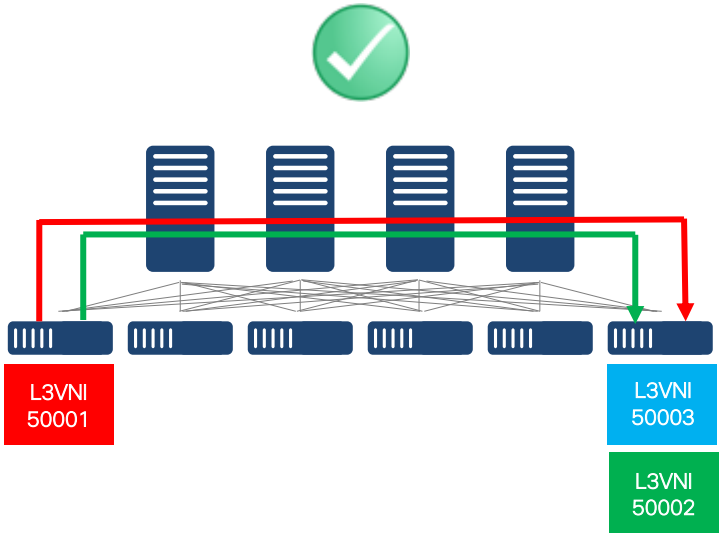




VXLAN EVPN и Downstream VNI
Рекомендации и ограничения (продолжение)

- Локальный VRF может быть ассоциирован с одним и более L3VNI на VTEP peer

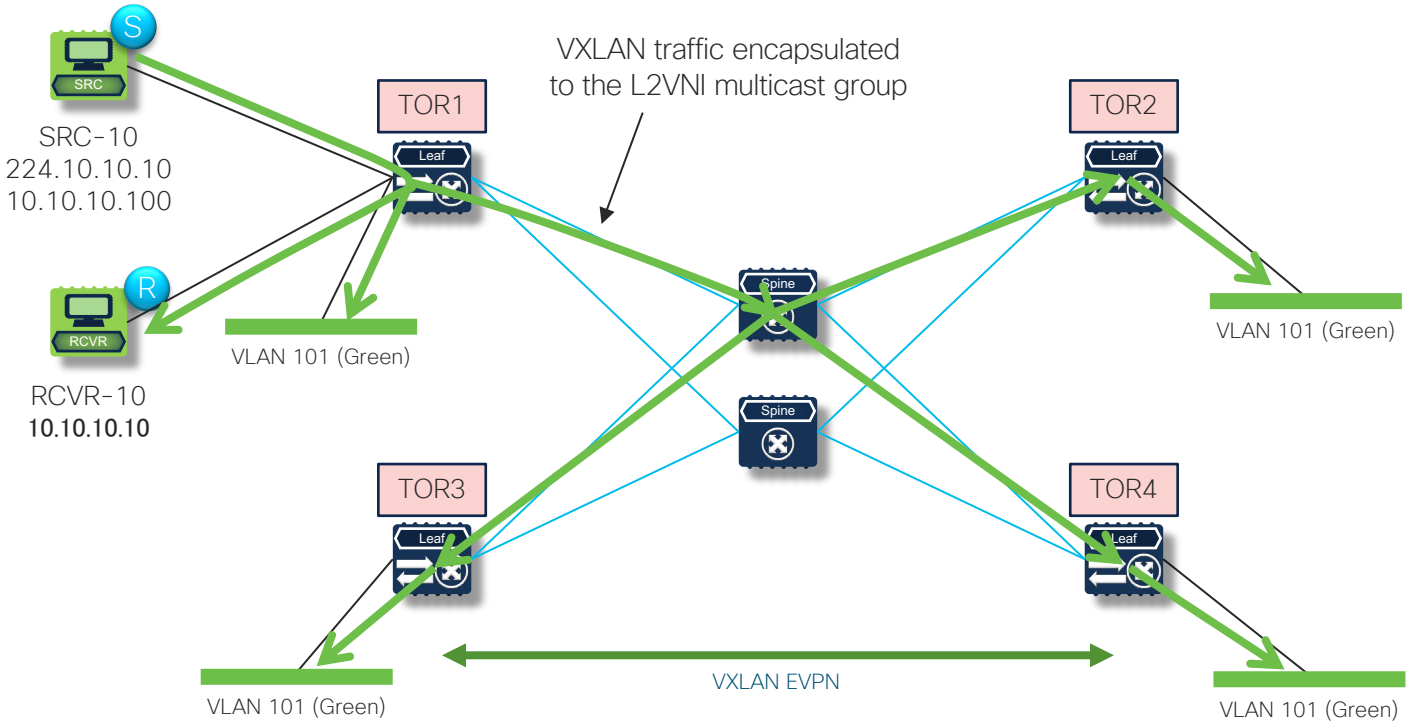




Маршрутизация multicast трафика в Overlay с помощью TRM

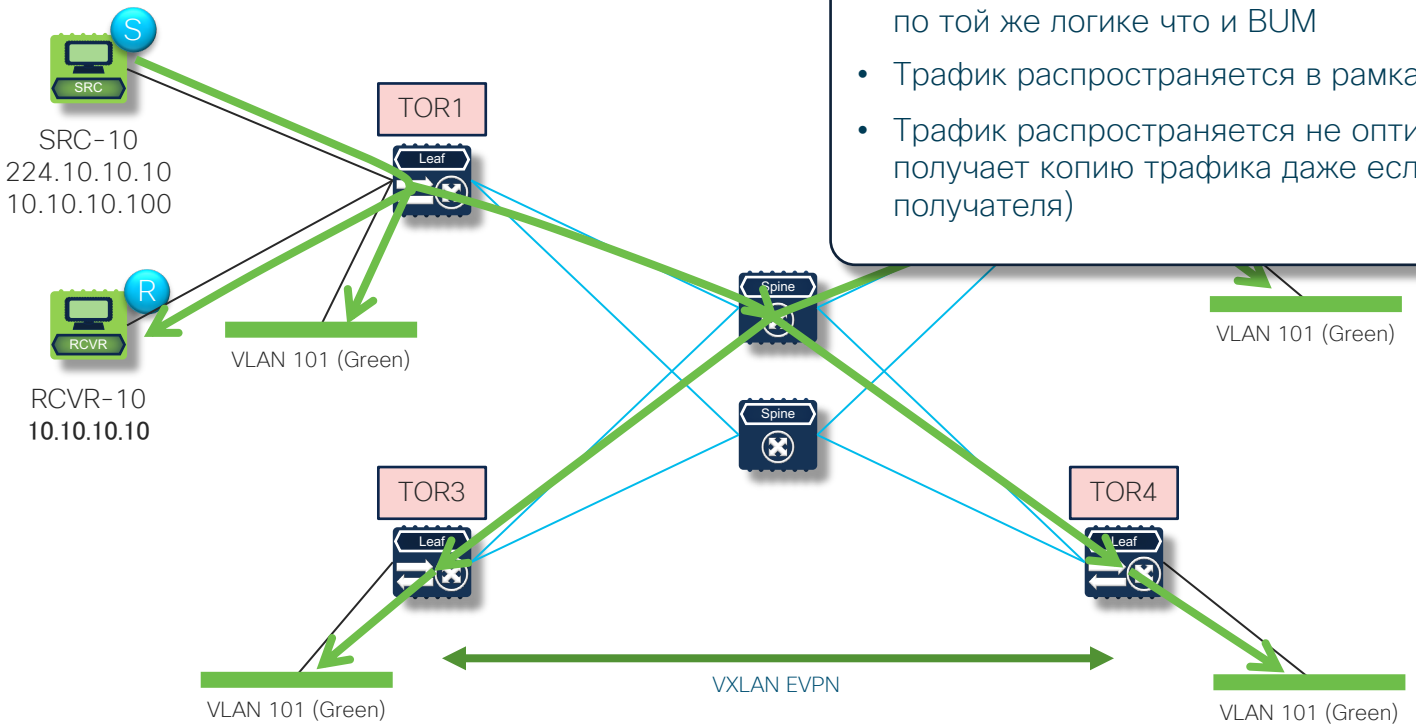
Коммутация multicast в overlay

IGMP Snooping отключен



Коммутация multicast в overlay

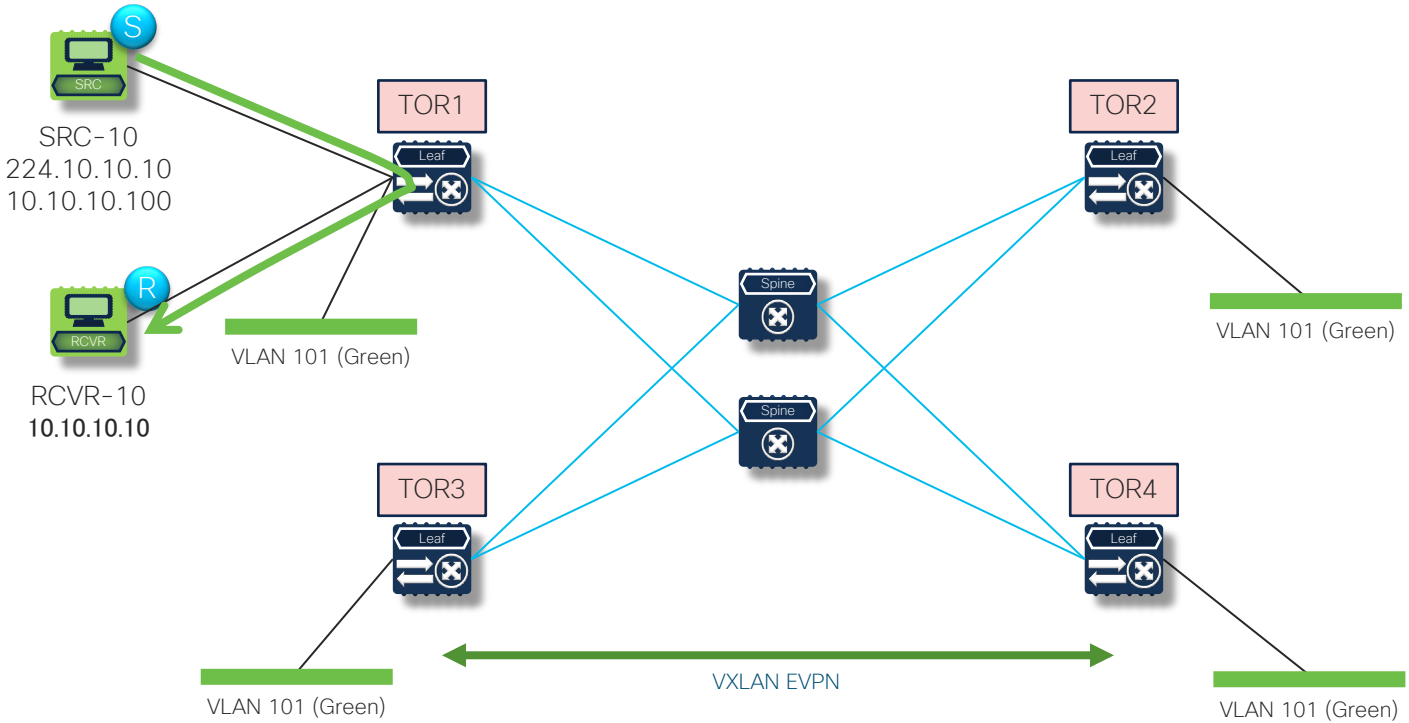
IGMP Snooping включен



- Продуктивный multicast трафик распространяется по той же логике что и BUM
- Трафик распространяется в рамках одной сети
- Трафик распространяется не оптимально (VTEP получает копию трафика даже если нет активного получателя)

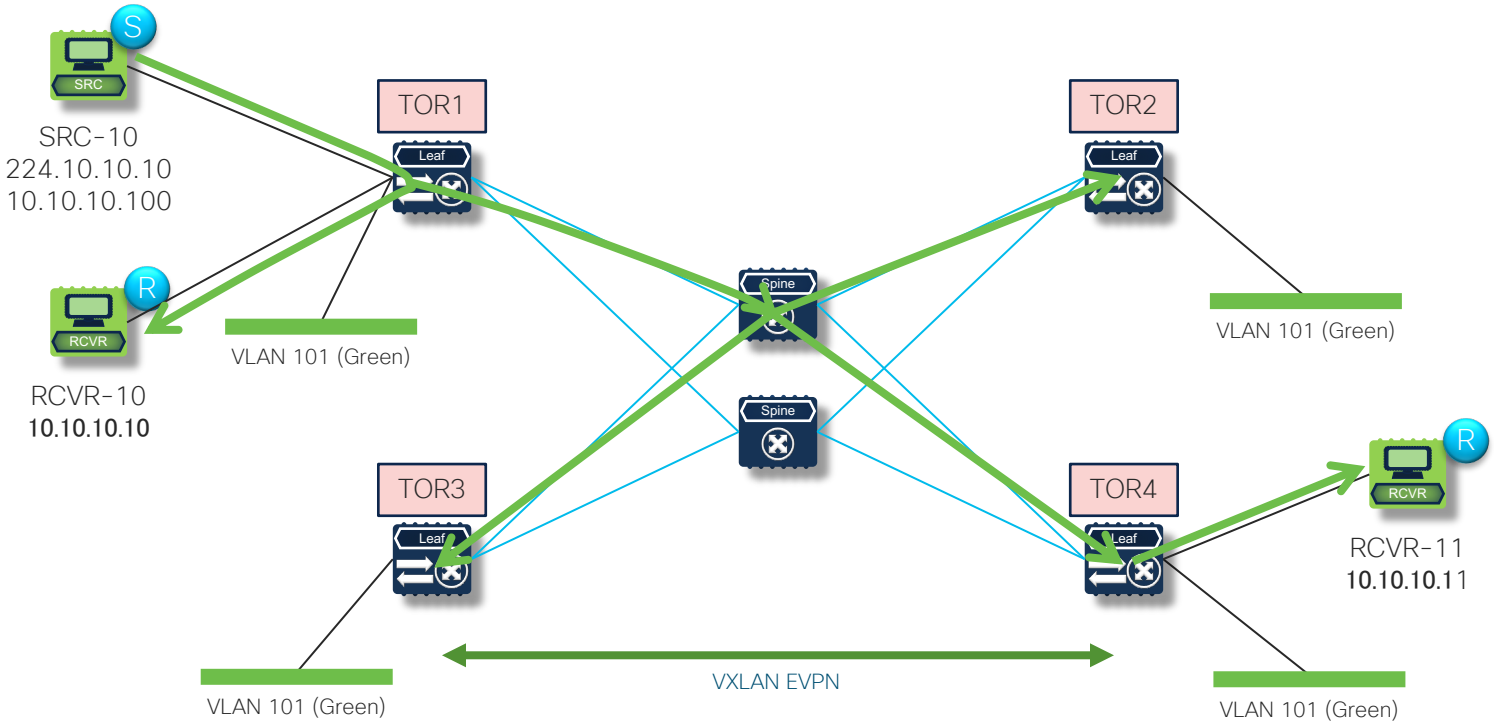
Коммутация multicast в overlay

IGMP Snooping включен



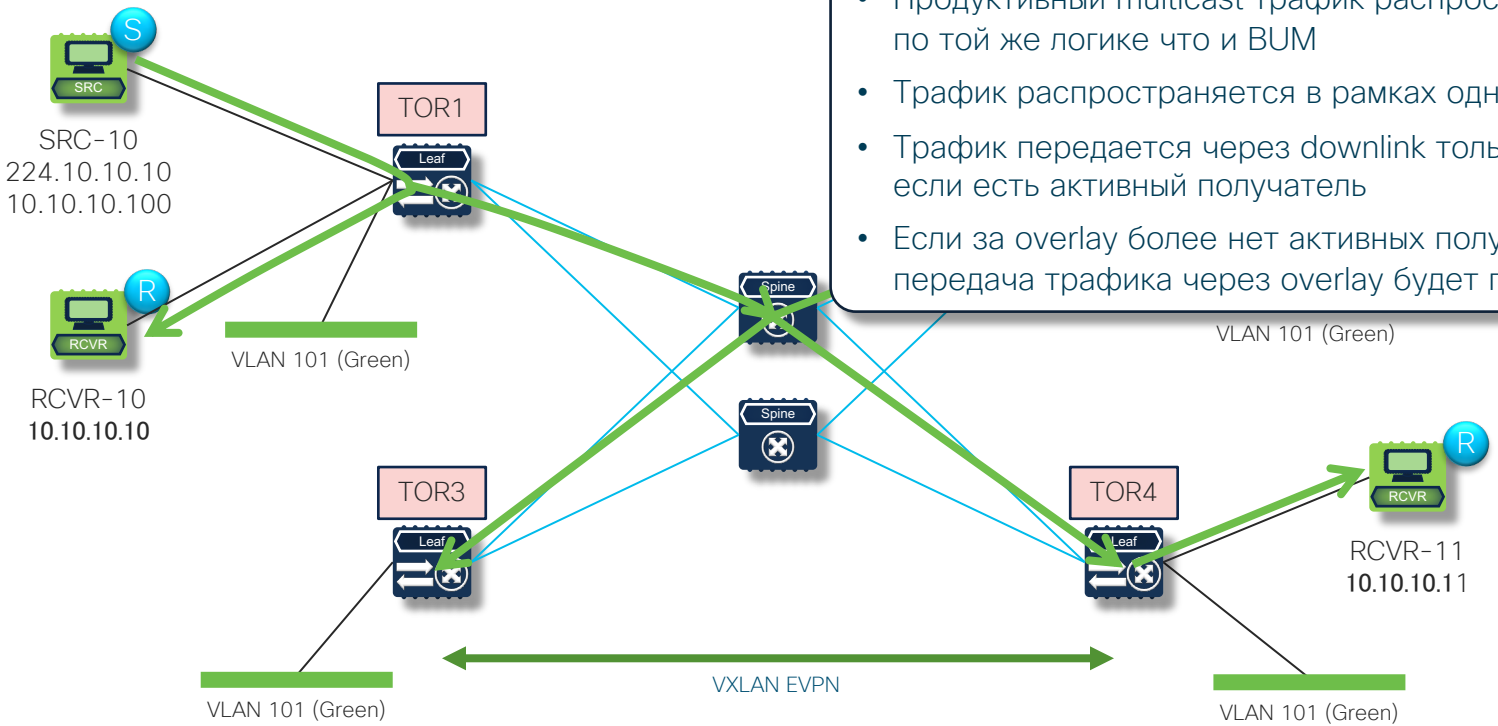
Коммутация multicast в overlay

IGMP Snooping включен



Same Subnet Forwarding with IGMP Snooping

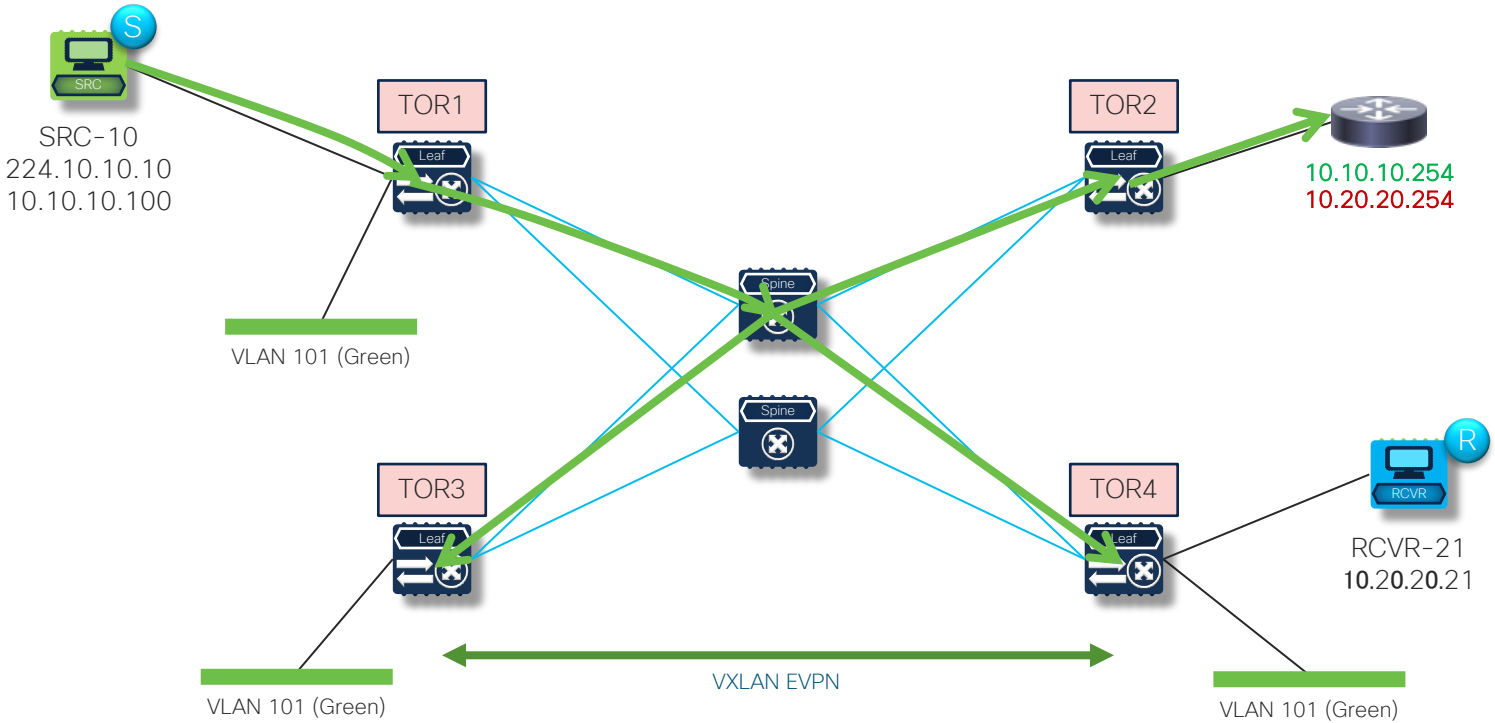
Traditional Forwarding in VXLAN Overlays



- Продуктивный multicast трафик распространяется по той же логике что и BUM
- Трафик распространяется в рамках одной сети
- Трафик передается через downlink только в случае если есть активный получатель
- Если за overlay более нет активных получателей передача трафика через overlay будет прекращена

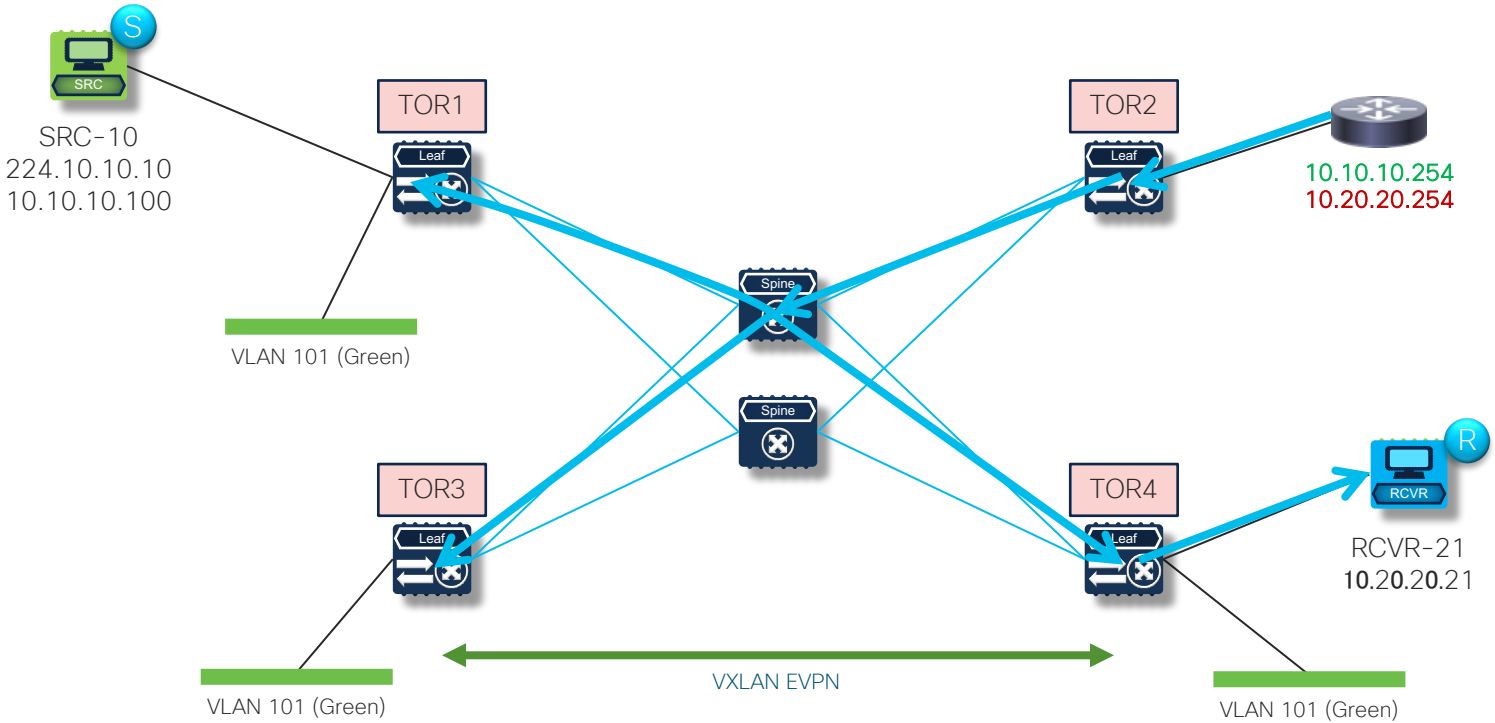
Маршрутизация multicast трафика в overlay

Router on-a-Stick



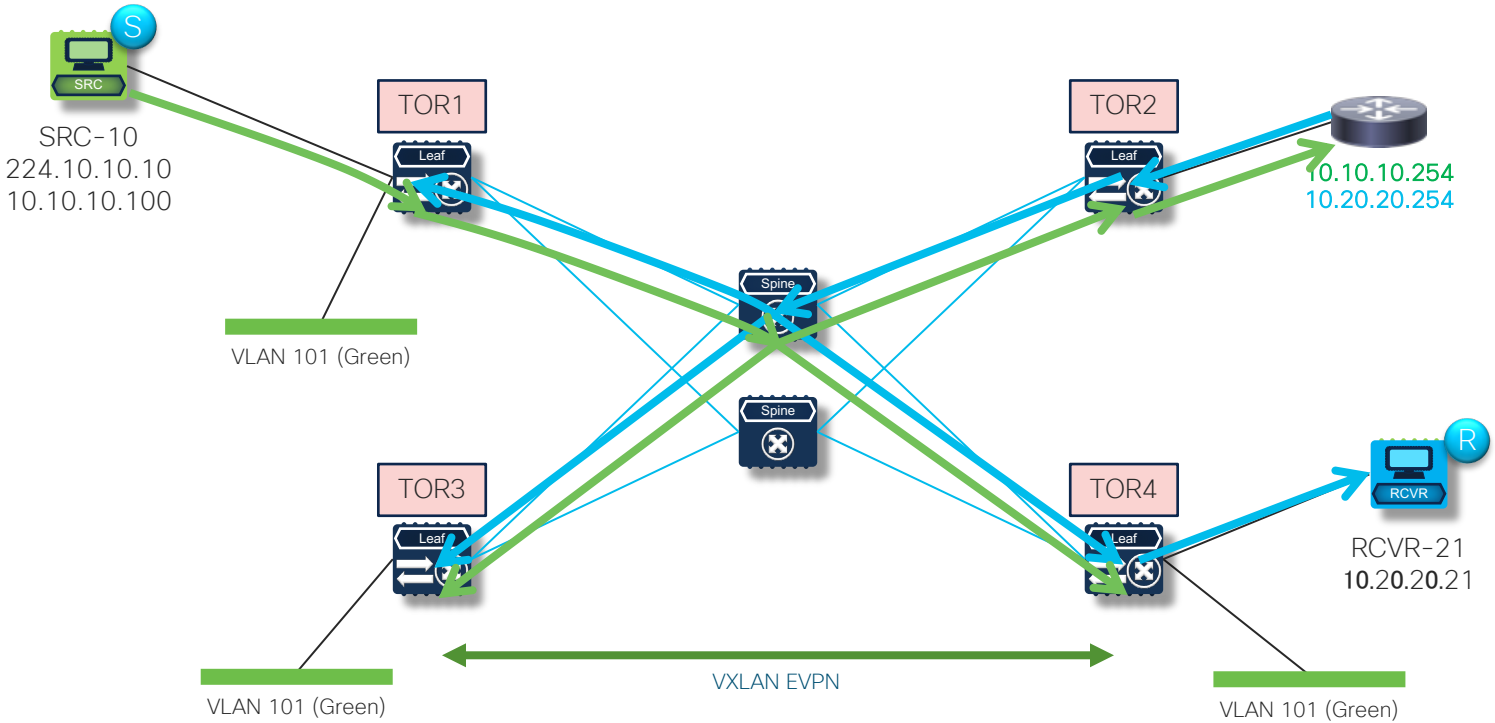
Маршрутизация multicast трафика в overlay

Router on-a-Stick



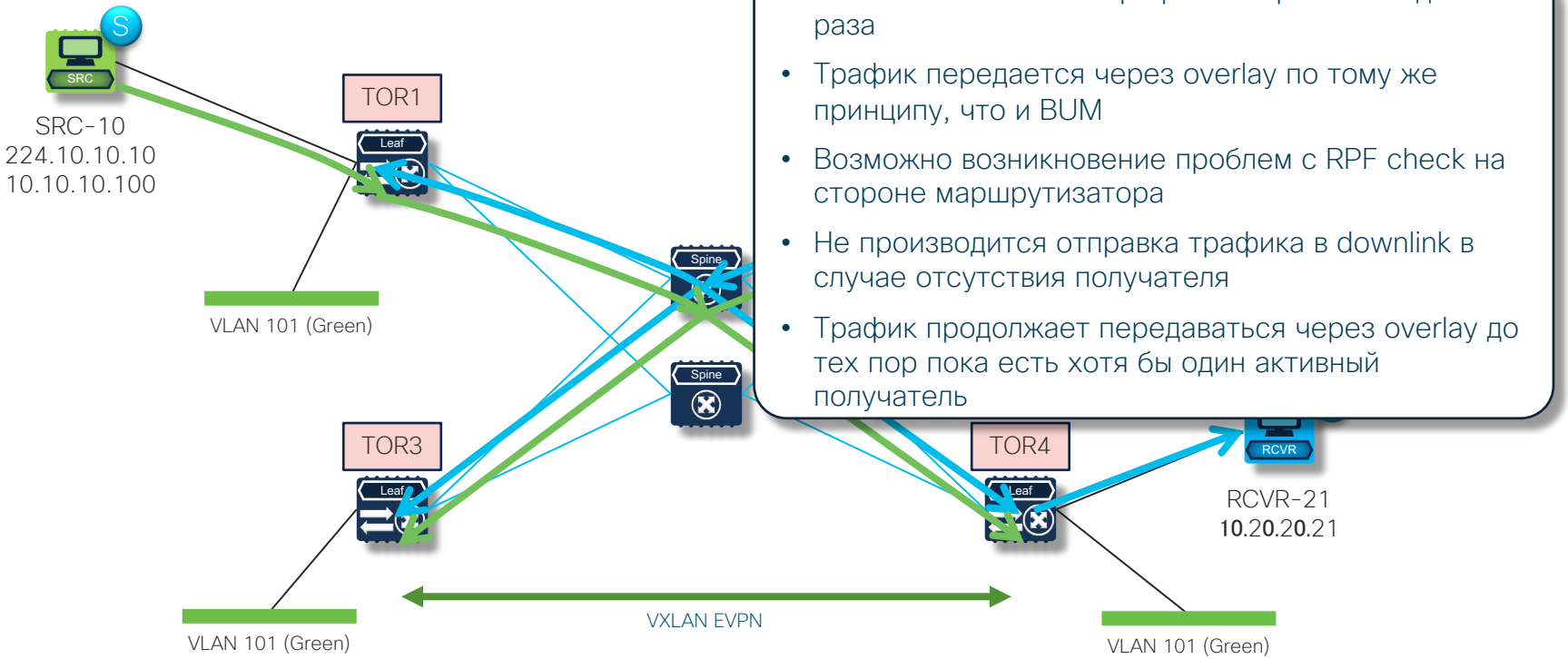
Маршрутизация multicast трафика в overlay

Router on-a-Stick



Маршрутизация multicast трафика в overlay

Router on-a-Stick



- Количество multicast трафика возрастает в два раза
- Трафик передается через overlay по тому же принципу, что и BUM
- Возможно возникновение проблем с RPF check на стороне маршрутизатора
- Не производится отправка трафика в downlink в случае отсутствия получателя
- Трафик продолжает передаваться через overlay до тех пор пока есть хотя бы один активный получатель

Tenant Routed Multicast

107

Tenant Routed Multicast

Рекомендации и ограничения

- TRM требует наличия IPv4 PIM ASM underlay
- TRM поддерживает overlay маршрутизацию только для PIM ASM и PIM SSM only
- На данный момент PIM BiDir не поддерживается в overlay
- TRM supports IPv4 multicast only
- Multicast local scope multicast (224.0.0.0/24) не маршрутизируется с помощью TRM
- TRM имеет следующие аппаратные требования:
  - Cisco Nexus 9200, 9300-EX, 9300-FX, and 9300-FX2
  - Cisco Nexus 9500 с установленными интерфейсными картами 9700-EX/-FX или их комбинацией
- TRM поддерживается начиная с NX-OS 7.0(3)I7(1)
- Подробное описание и рекомендации приведены в документе по ссылке:

https://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus9000/sw/92x/vxlan-92x/configuration/guide/b-cisco-nexus-9000-series-nx-os-vxlan-configuration-guide-92x/b_Cisco_Nexus_9000_Series_NX-OS_VXLAN_Configuration_Guide_9x_chapter_011.html#reference_lbw_qx1_zfb

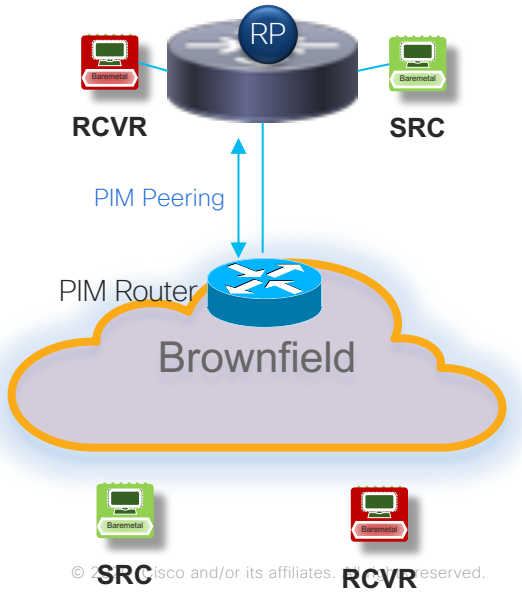
© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public 108

Tenant Routed Multicast

Интеграция с существующей инфраструктурой

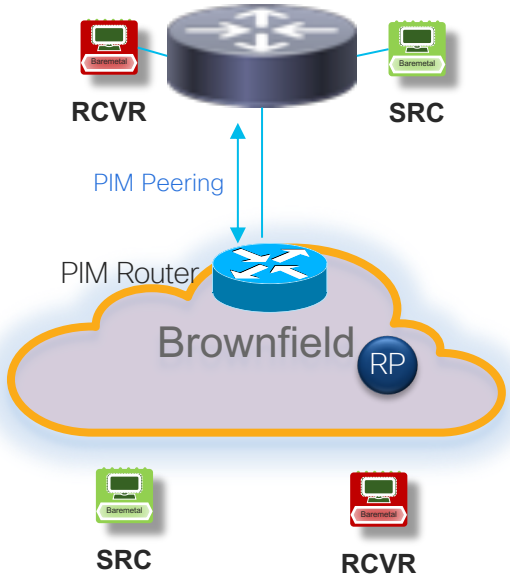
RP-Less

VXLAN EVPN Fabric является распределенным PIM маршрутизатором и так же выполняет роль RP



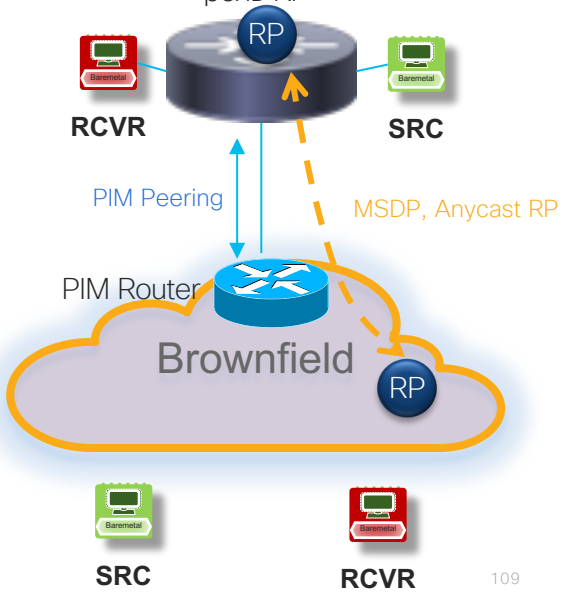
External RP

VXLAN EVPN Fabric является распределенным PIM маршрутизатором и так же выполняет роль RP



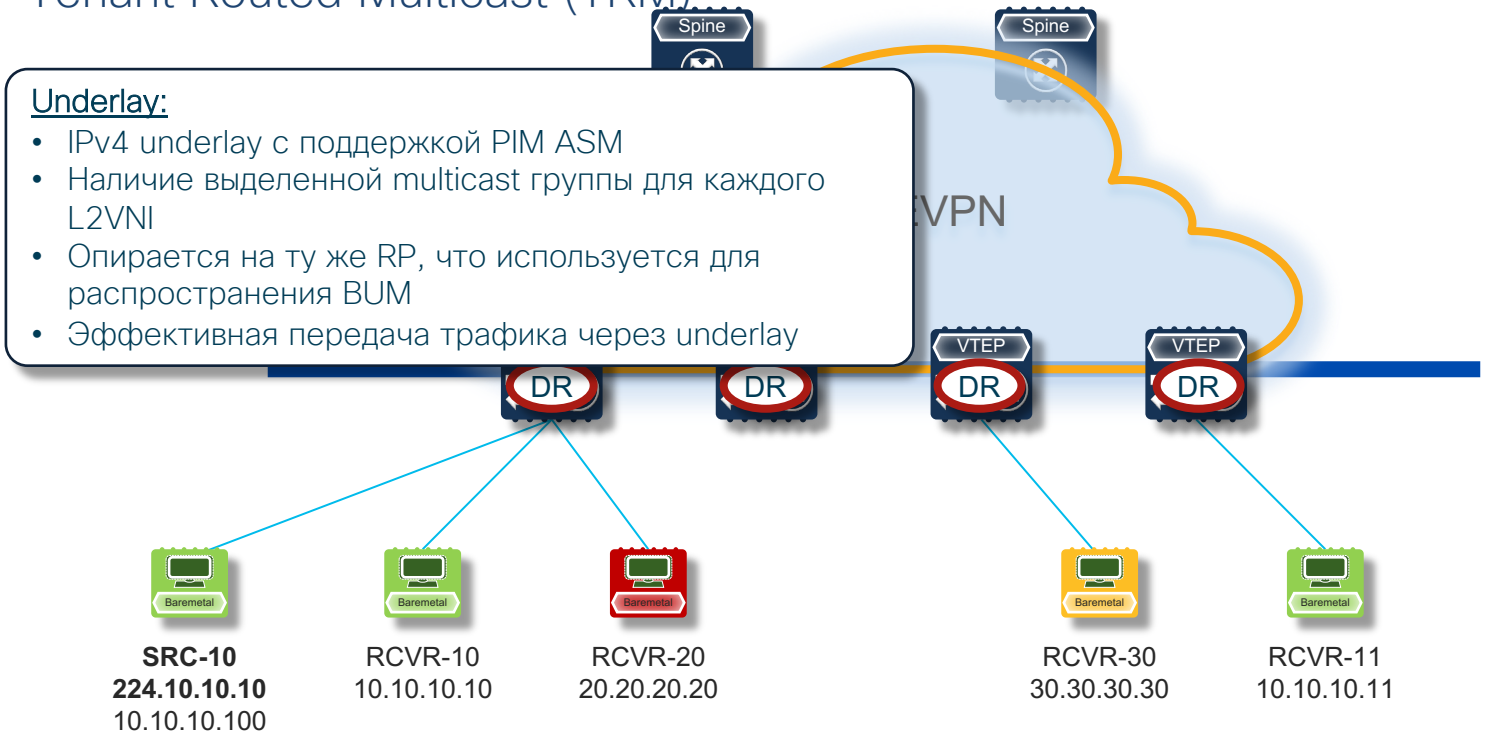
RP Anywhere

VXLAN EVPN Fabric является распределенным PIM маршрутизатором и так же выполняет роль RP



Основные элементы решения

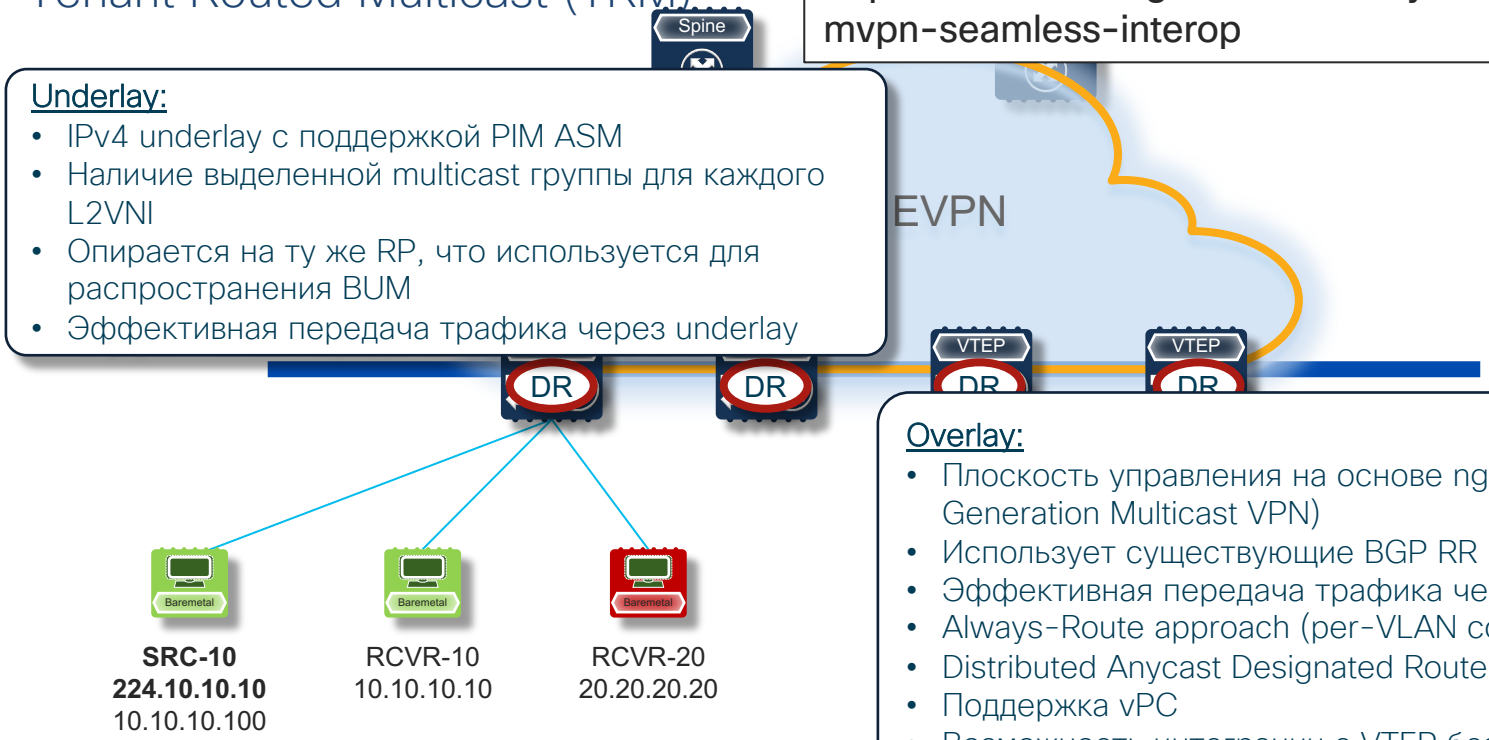
Tenant Routed Multicast (TRM)



Основные элементы решения

Tenant Routed Multicast (TRM)

https://tools.ietf.org/html/draft-sajassi-bess-evpn-mvpn-seamless-interop



Underlay:

- IPv4 underlay с поддержкой PIM ASM
- Наличие выделенной multicast группы для каждого L2VNI
- Опирается на ту же RP, что используется для распространения BUM
- Эффективная передача трафика через underlay

Overlay:

- Плоскость управления на основе ngMVPN (Next-Generation Multicast VPN)
- Использует существующие BGP RR
- Эффективная передача трафика через underlay
- Always-Route approach (per-VLAN config)
- Distributed Anycast Designated Router (DR)
- Поддержка vPC
- Возможность интеграции с VTEP без поддержки TRM

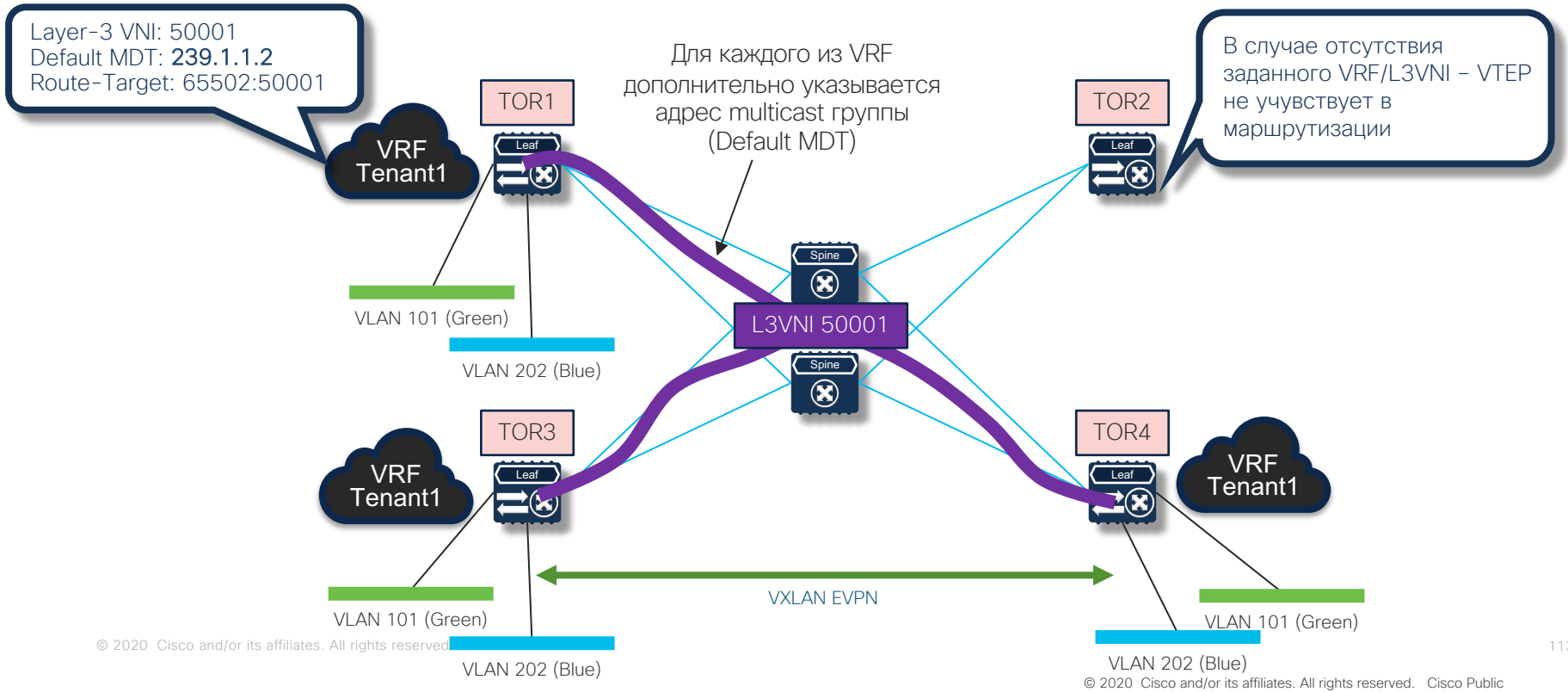
Tenant Routed Multicast

Плоскость передачи данных

112

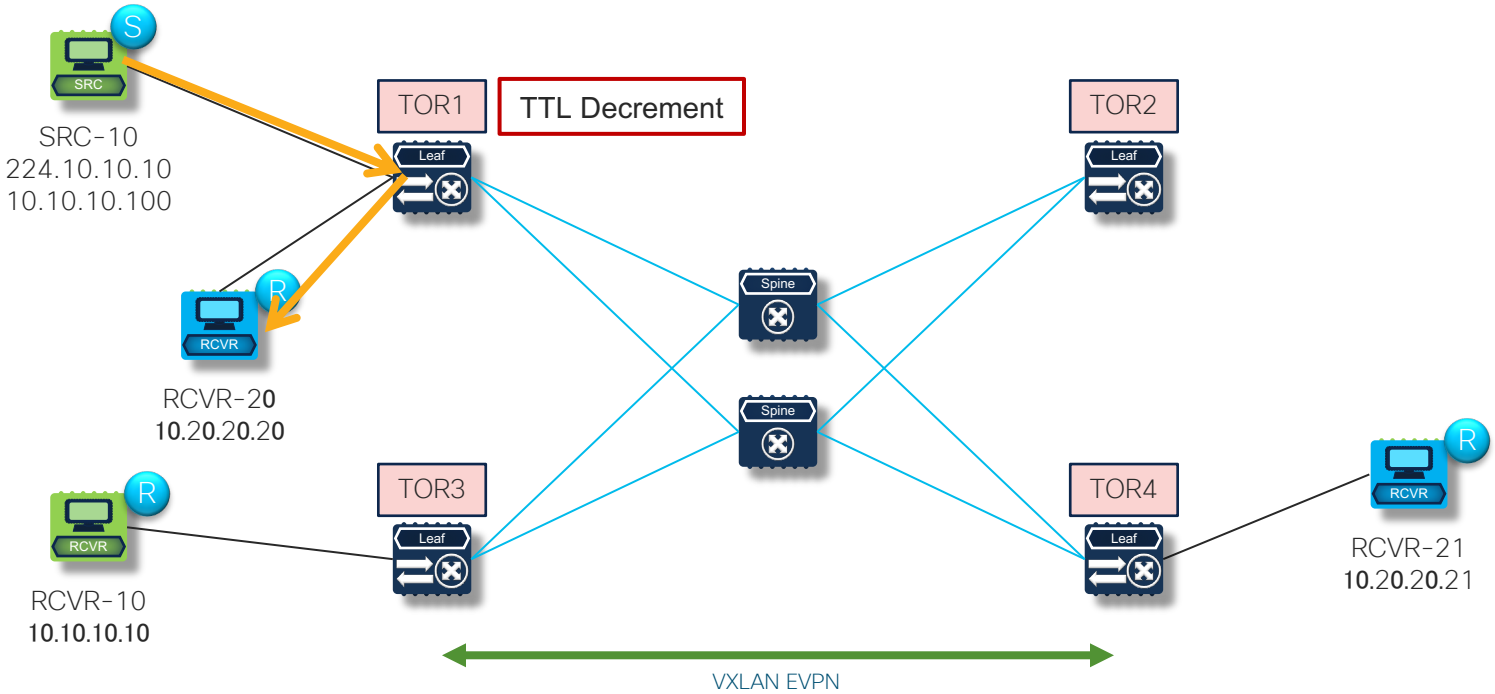
Маршрутизация и коммутация трафика с помощью TRM (L3 mode)

Layer-3 Mode



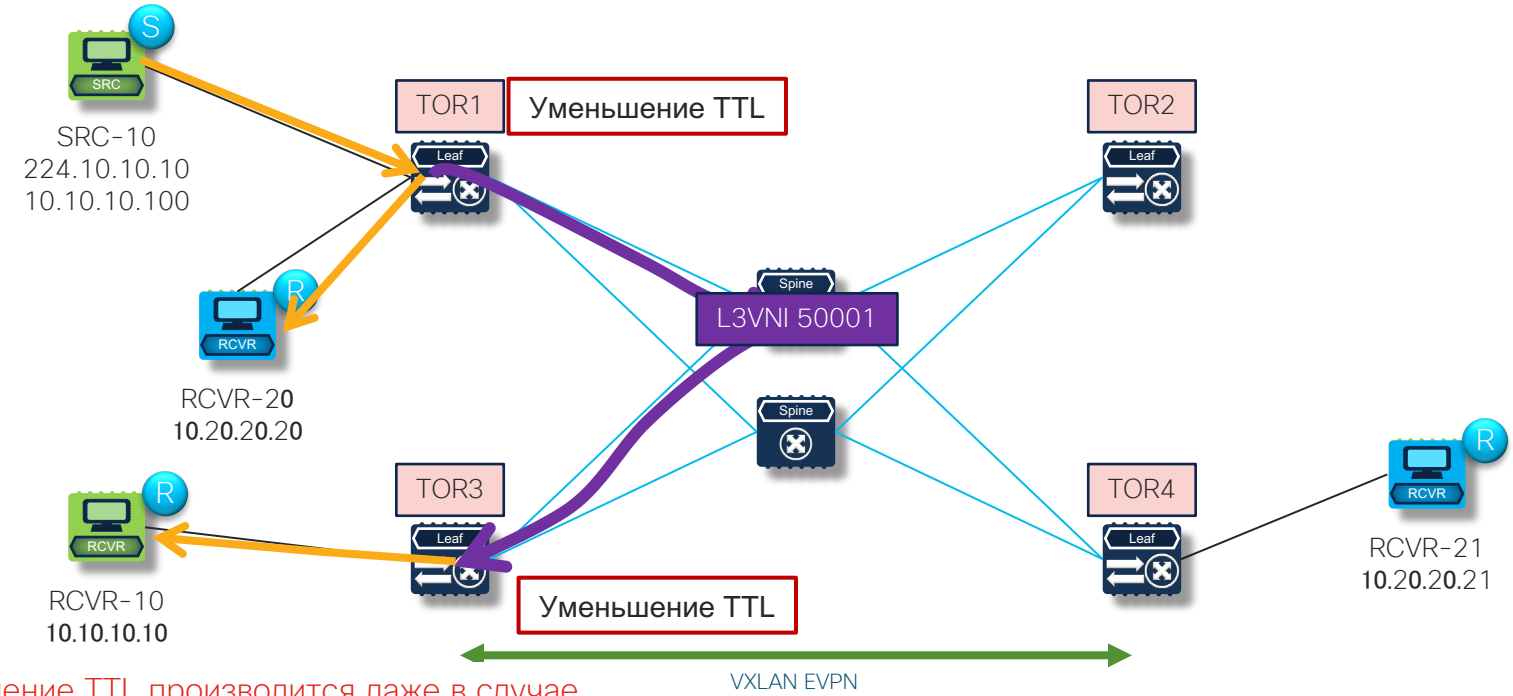
Маршрутизация и коммутация трафика с помощью TRM

Layer-3 Mode



Маршрутизация и коммутация трафика с помощью TRM

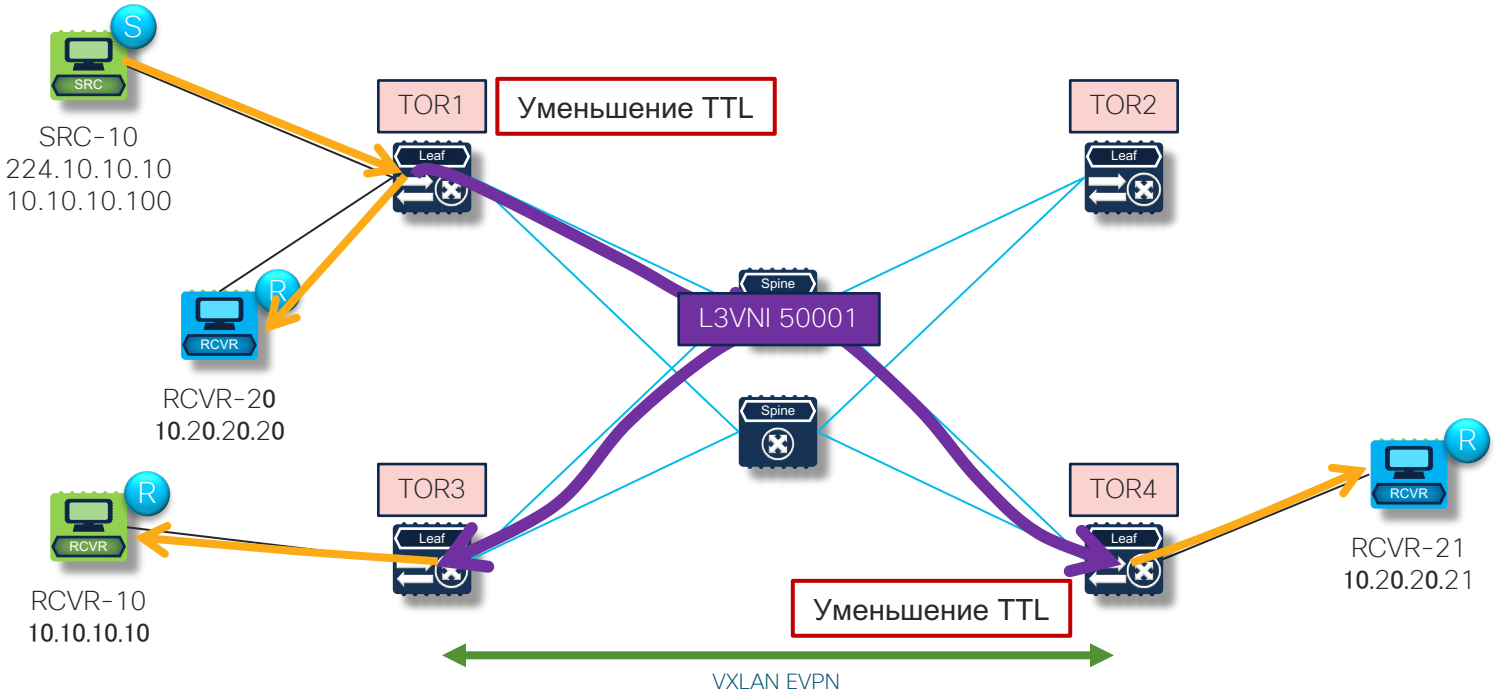
Layer-3 Mode



Уменьшение TTL производится даже в случае передачи multicast трафика в рамках одной подсети!

Маршрутизация и коммутация трафика с помощью TRM

Layer-3 Mode

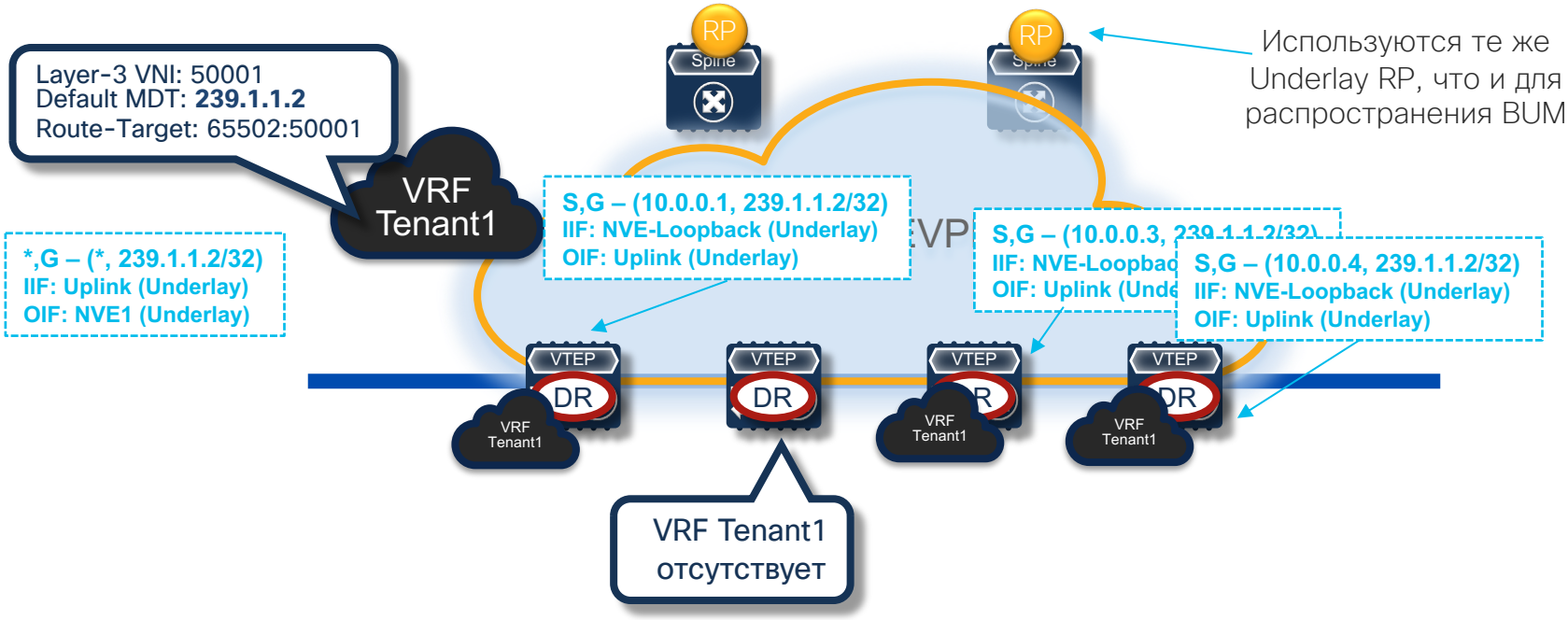


TRM плоскость управления и передачи данных

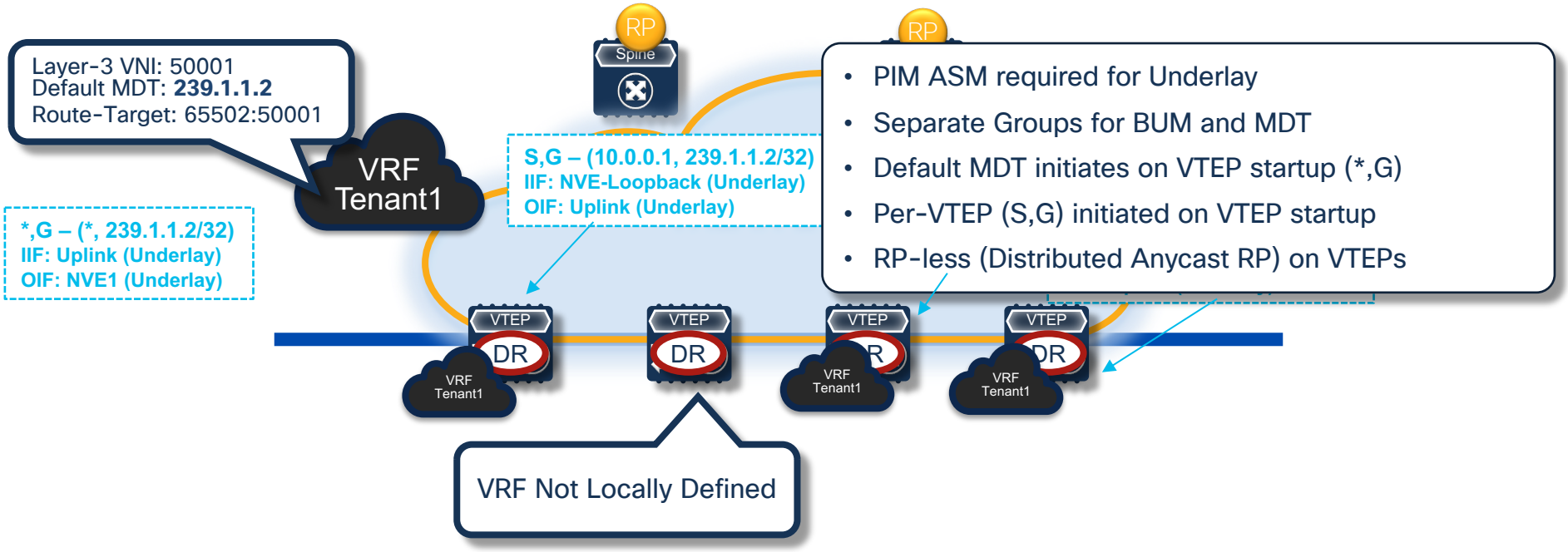
RP-Less

117

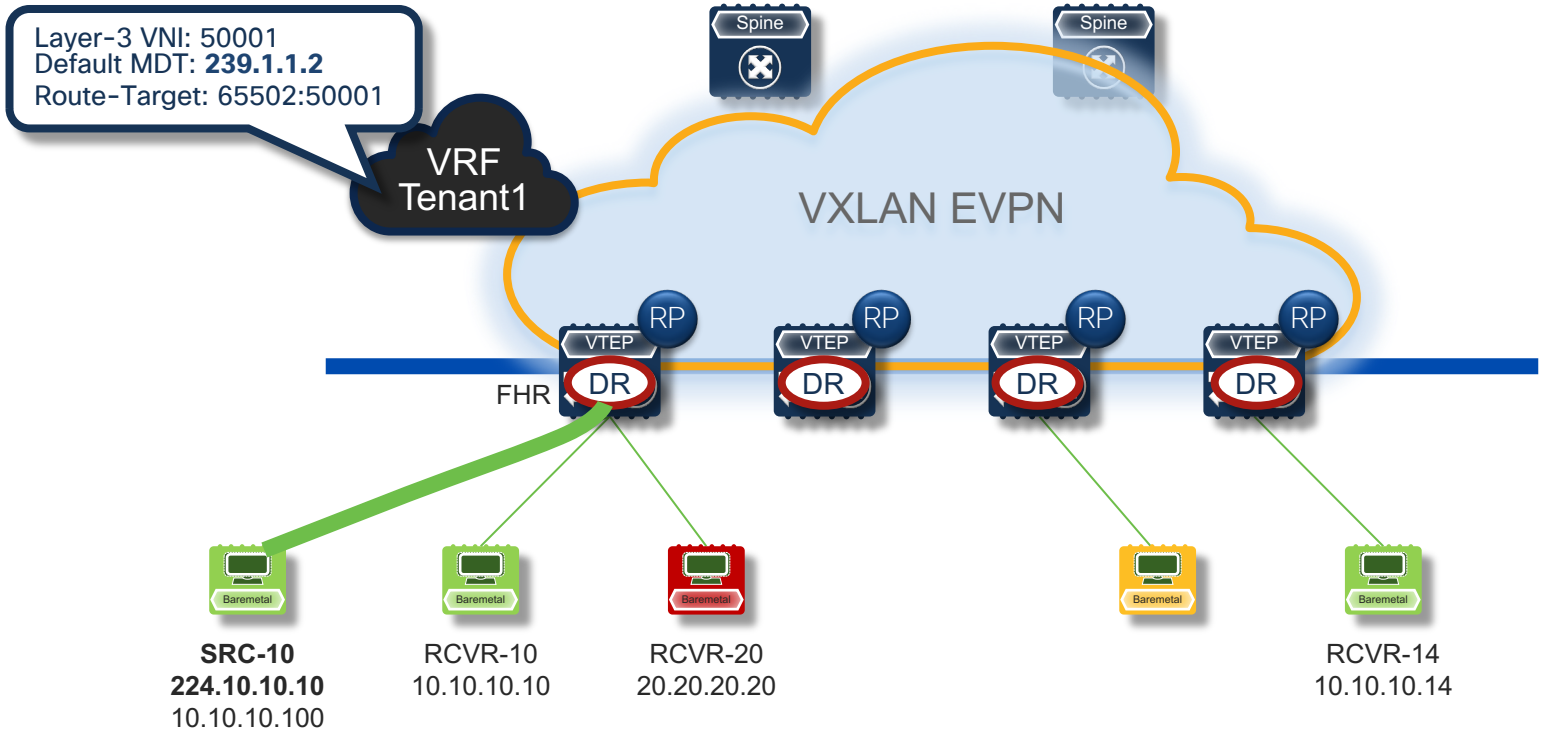
Состояние PIM в underlay



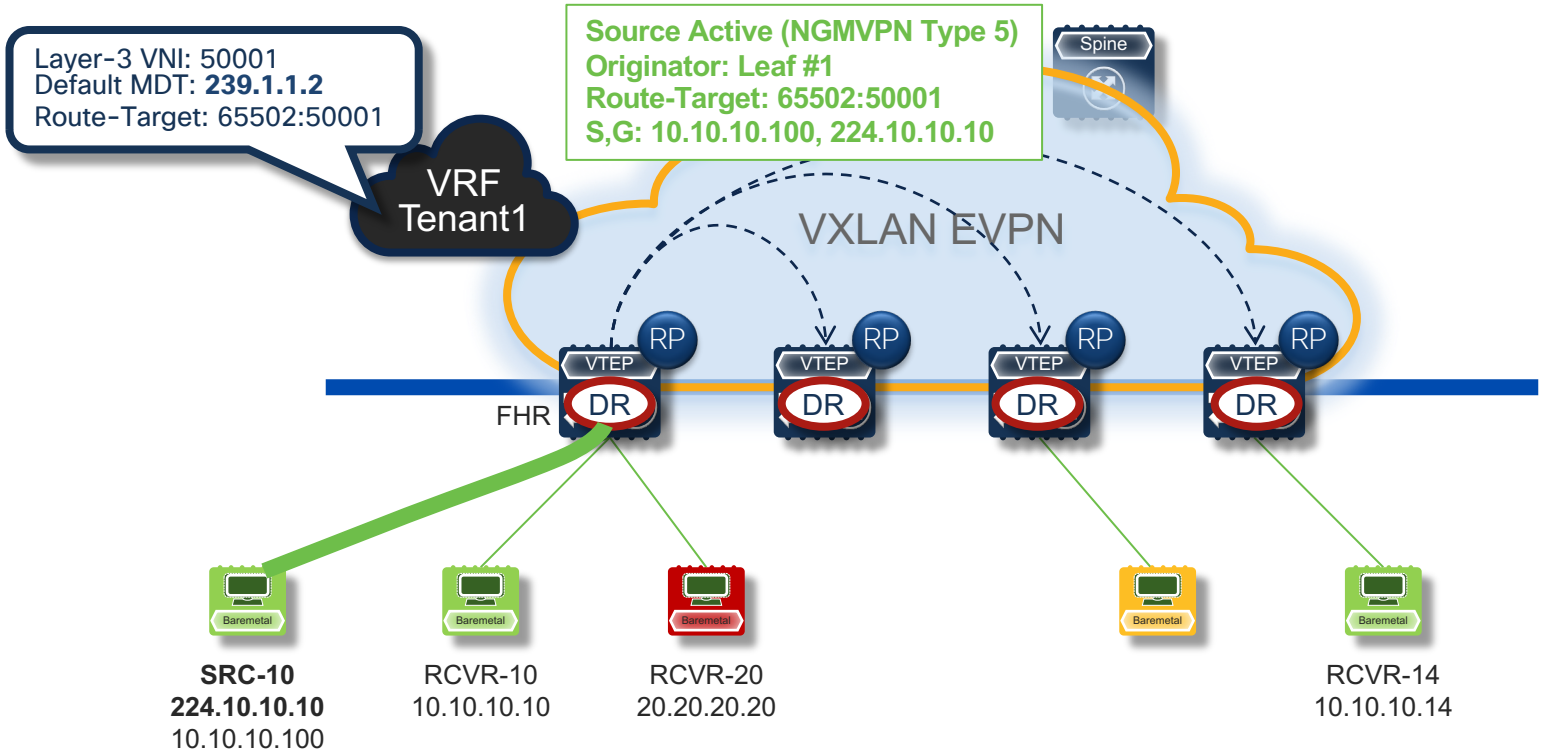
Состояние PIM в underlay



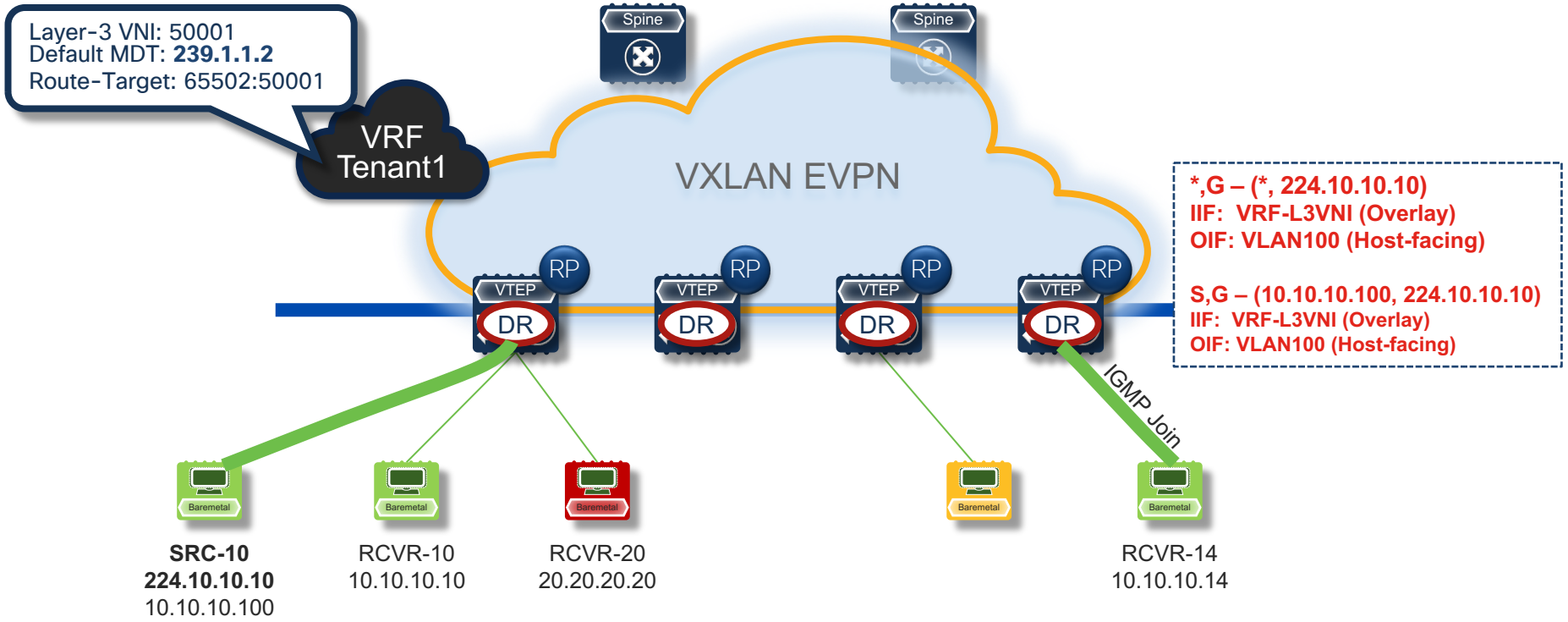
Шаг 1: Источник начинает отправлять трафик



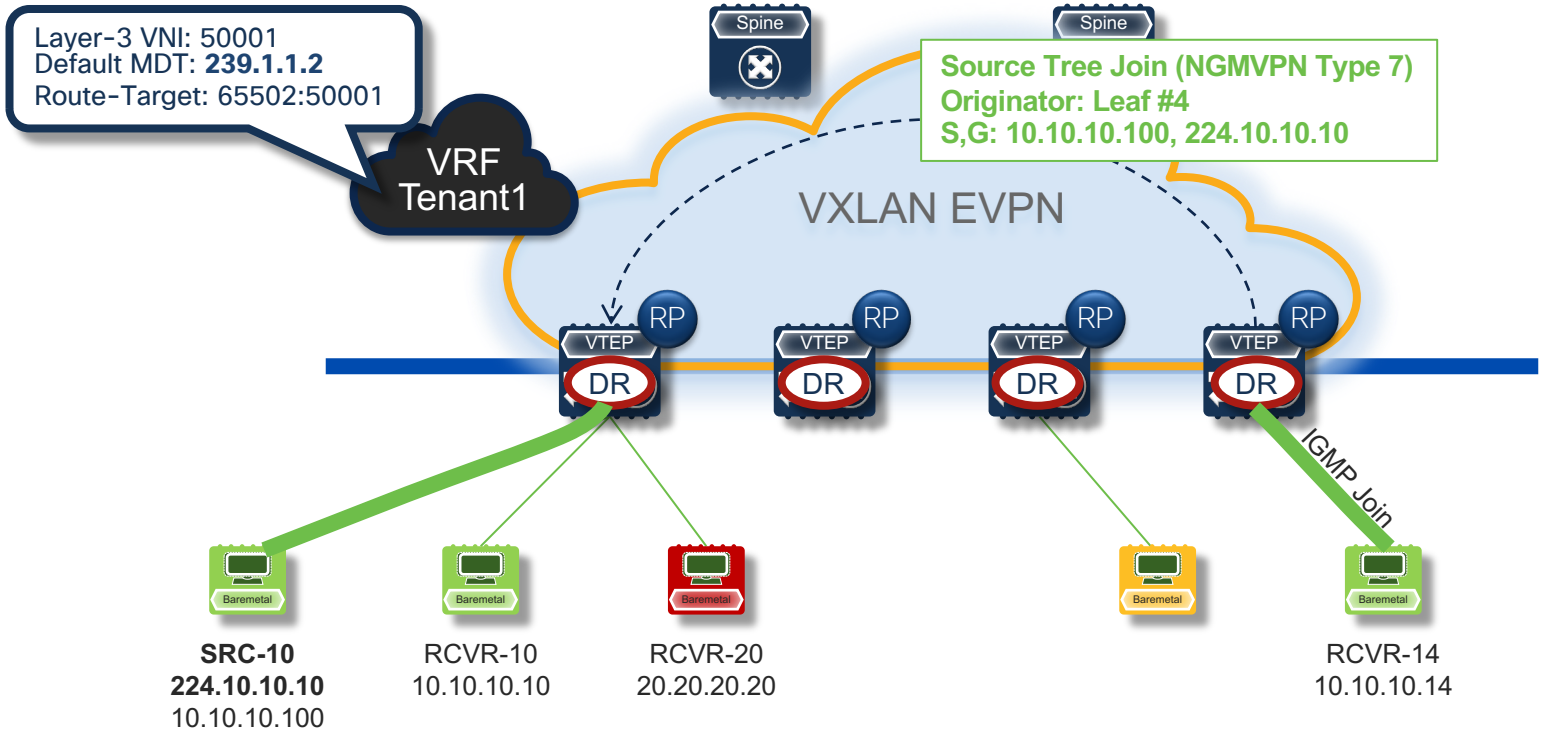
Шаг 2: Распространение NGMVPN Source Active Advertisement (MVPN Type 5)



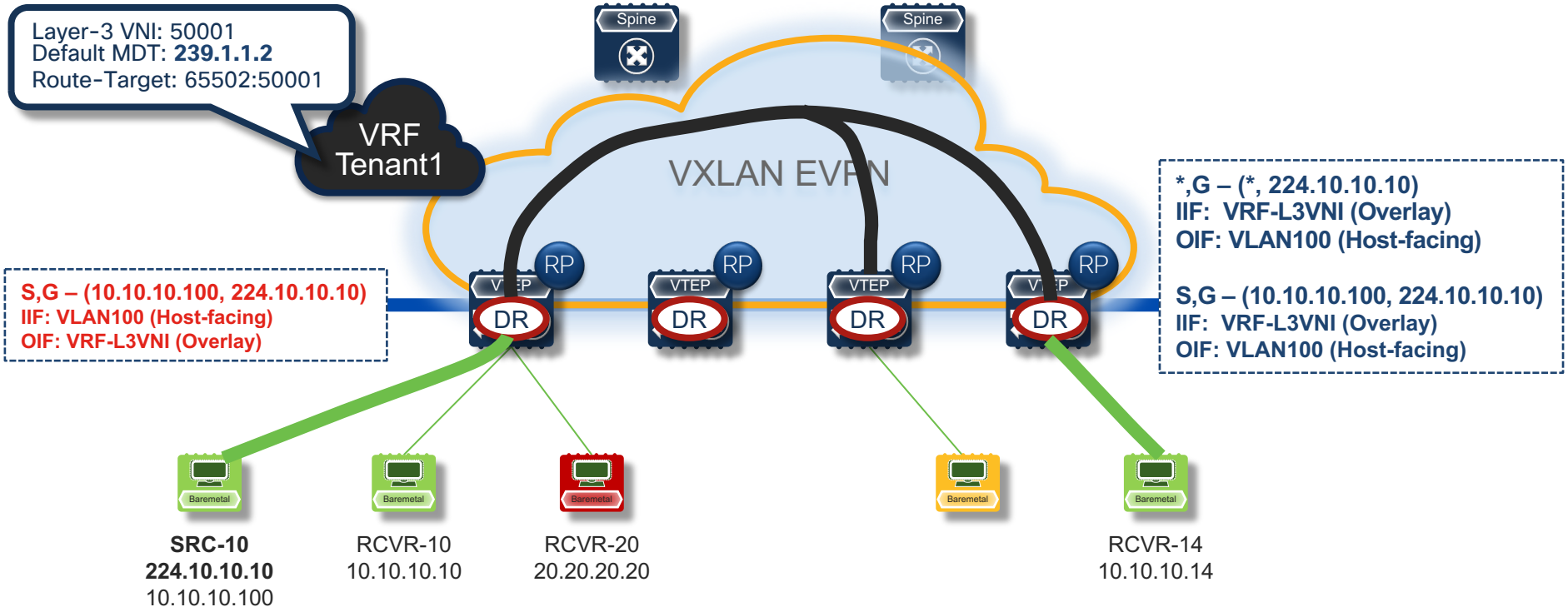
Шаг 3: Получатель отправляет IGMP Join



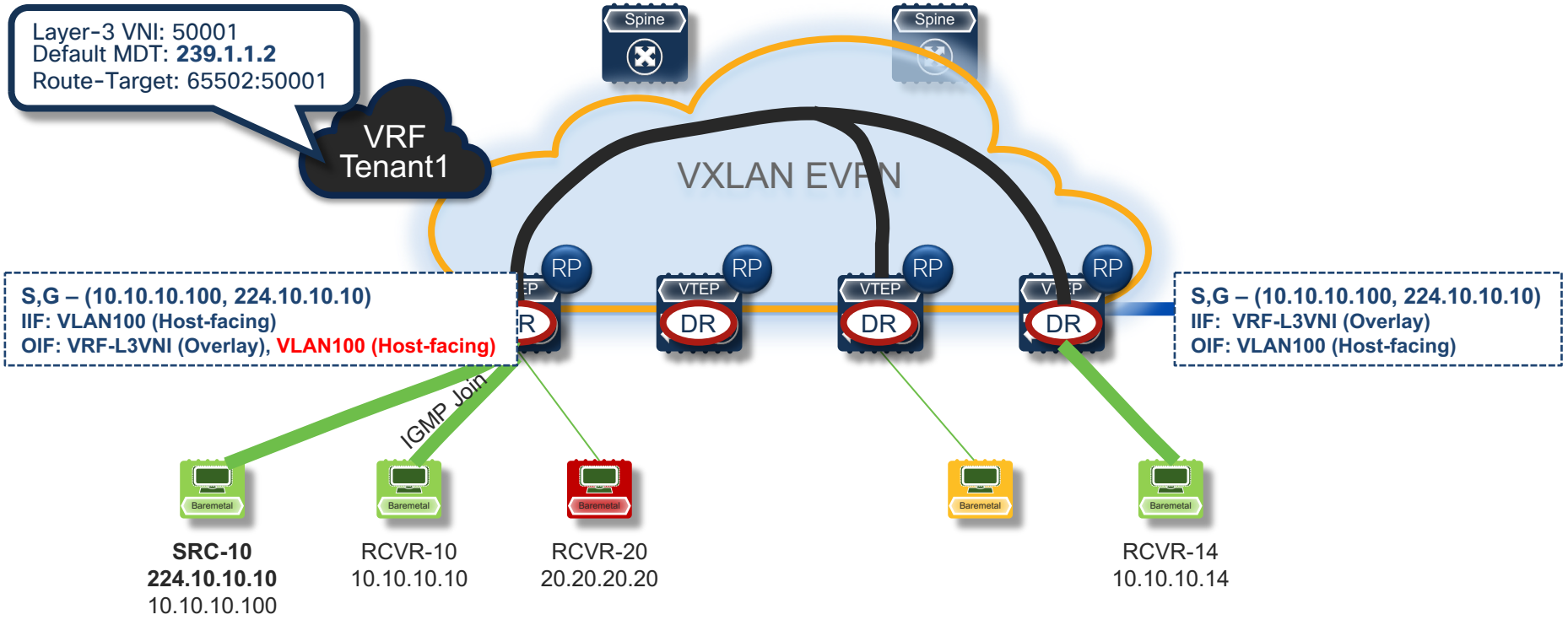
Шаг 4: Распространение NGMVPN: Source Tree Join (MVPN Type 7)



Шаг 5: Начало передачи multicast трафика через overlay



Добавление локального получателя трафика



Фильтрация EVPN NLRI

EVPN Filtering

- Гибкое управление EVPN BGP NLRI и изменение их параметров
- Были добавлены новые критерии отбора (match) и возможность изменения параметров (set) в политиках route-map

Примеры критериев и изменения параметров

- Отбор опираясь на тип EVPN маршрута (route-type)
- Отбор EVPN type-2 маршрутов по MAC адресу
- Отбор EVPN маршрутов по RMAC extended community
- Установка RMAC Extended Community
- Установка EVPN Next-hop IP
- Установка Gateway IP для EVPN маршрутов type-5

BGP EVPN Filtering – Примеры использования

Use-Case 1

- Фильтрация VMAC/VIP изученных со стороны legacy сети

Use-Case 2

- Фильтрация VMAC/VIP между сайтами на Multi-Site BGW

Use-Case 3

- Смена Next-Hop и перезапись RMAC Extended Community для перенаправления трафика


CISCO