Sommario

FIGURE

VXLAN EVPN

VXLAN è una tecnologia che permette di encapsulare frame layer 2 dentro UDP header con l'obiettivo di estendere il dominio di switching attraverso una rete layer 3 IP.

All'interno dell'header UDP abbiamo l'header VXLAN con il suo VNI (VXLAN Network Identifier) costituito da 24 byte per un massimo di estensione vlans pari ad oltre 16 milioni di segmenti logici.

Vantaggi VXLAN:

- Estensione del range di vlans da 4096 ad oltre 16 milioni
- L2 extensions Data Centers
- Alta scalabilità, alta affidabilità e migliori performance di rete con un basso consumo di risorse
- Ottimizzazione relative allo spanning tree protocol
- MP-BGP EVPN integrato alla prima versione di VXLAN (quest'ultimo prevede multicast per la conoscenza di VTEP e raggiungiblità di hosts/servers) come control-plane (piano di controllo) prevede un nuovo address-family chiamato **L2VPN EVPN**

MP-BGP EVPN

Con il MP-BGP EVPN control-plane abbiamo:

- Informazioni layer 2 (MAC address) e layer 3 (host IP address) imparate localmente da ogni VTEP sono propagate ad altri VTEP permettendo funzionalità di switching e routing all'interno della stessa fabbrica
- Le routes sono annunciate tra VTEP attraverso route-targe policy
- Utilizzo di VRF e route-distinguisher per routes/subnet
- Le informazioni layer 2 sono distribuite tra VTEP con la funzionalità di ARP cache per minimizzare il flooding
- Le sessioni L2VPN EVPN tra VTEP possono essere autenticate via MD5 per mitigare problematiche di sicurezza (Rogue VTEP)

MP-BGP EVPN utilizza due routing advertisement:

- **Route type 2:** usato per annunciare host MAC ed IP address information per gli endpoint direttamente collegati alla VXLAN EVPN Fabric, ed anche trasportare extended community attribute, come route-target, router MAC address e sequence number;

- **Route type 5:** annuncio di IP Prefix oppure host routes (loopback interface) ed anche trasporto di extended community attribute, come route-target, router MAC address e sequence number

Distributed Anycast Protocol Gateway

Protocolli FHRP quali HSRP, VRRP e GLBP hanno funzionalità di alta affidabilità layer 3 attraverso meccanismi active-standby routers e VIP address gateway condiviso.

Distributed Anycast Protocol, supera la limitazione di avere solo due routers peers HSRP/VRRP in ambienti Data Centers, costruendo una VXLAN EVPN VTEP Fabric con una architettura di tipo Spine-Leaf.

Distributed Anycast Proocol offre i seguenti vantaggi:

- Stesso IP address gateway per tutti gli Edge Switch; ogni endpoint ha come gateway il proprio local VTEP il quale ruota poi il traffico esternamente ad altri VTEP attraverso una rete IP core (questo vale sia per VXLAN EVPN costruito come Fabric locale che geograficamente distribuito);
- La funzionalità di ARP suppression permette di ridurre il flooding all'interno del proprio dominio di switching (Leaf to Edge Switch);
- Permette il moving di host/server continuando a mantenere lo stesso IP address gateway configurato nel local VTEP, all'interno di ciascuna VXLAN EVPN Fabric locale o geograficamente distribuita
- No FHRP Filtering tra VXLAN EVPN Fabrics
- Permette
    - VLAN and VRF-Lite hand-off to DCI
    - MAN/WAN connectivity to external Layer 3 network domain
    - Connectivity to network services

Learning process EndPoint information

Il processo di learning Endpoint avviene a livello Edge Switch Leaf Node di una VXLAN EVPN Fabric, dove l'endpoint è direttamente connesso; le informazioni MAC address a livello locale sono calcolate attraverso la tabella di forwarding locale (data-plane table) mentre l'IP address è imparato attraverso meccanismi di ARP, GARP (Gratitous ARP) oppure IPv6 neighbor discovery message.

I comandi:

show l2route evpn mac mostra il contenuto della VPN table (L2 RIB routing information base) popolato via BGP updates

show l2route evpn mac-ip mostra le informazioni contenute nella L2RIB insieme alle host route information ricevute via route-type 2 EVPN updates.

Una volta avvenuto il processo di apprendimento MAC + IP a livello locale, queste informazioni vengono annunciate dai rispettivi VTEP attraverso il MP-BGP EVPN control-plane utilizzando le EVPN route-type 2 advertisement trasmette a tutti i VTEP Edge Devices che appartengono alla stessa VXLAN EVPN Fabric.

Di conseguenza, tutti gli Edge Devices imparano le informazioni EndPoint che appartengono ai rispettivi VNI (VXLAN segment Network Identifier) ed essere importate all'interno della propria forwarding table.

Intra-Subnet communication via Fabrics EVPN

La comunicazione tra due EndPoint ubicati su EVPN Fabric differenti è stabilito attraverso la combinazione di creare un bridge domain L2 VXLAN (all'interno di ogni Fabric) e un L2 extension segment di rete IP address tra Fabrics (via OTV o altra tecnologia)

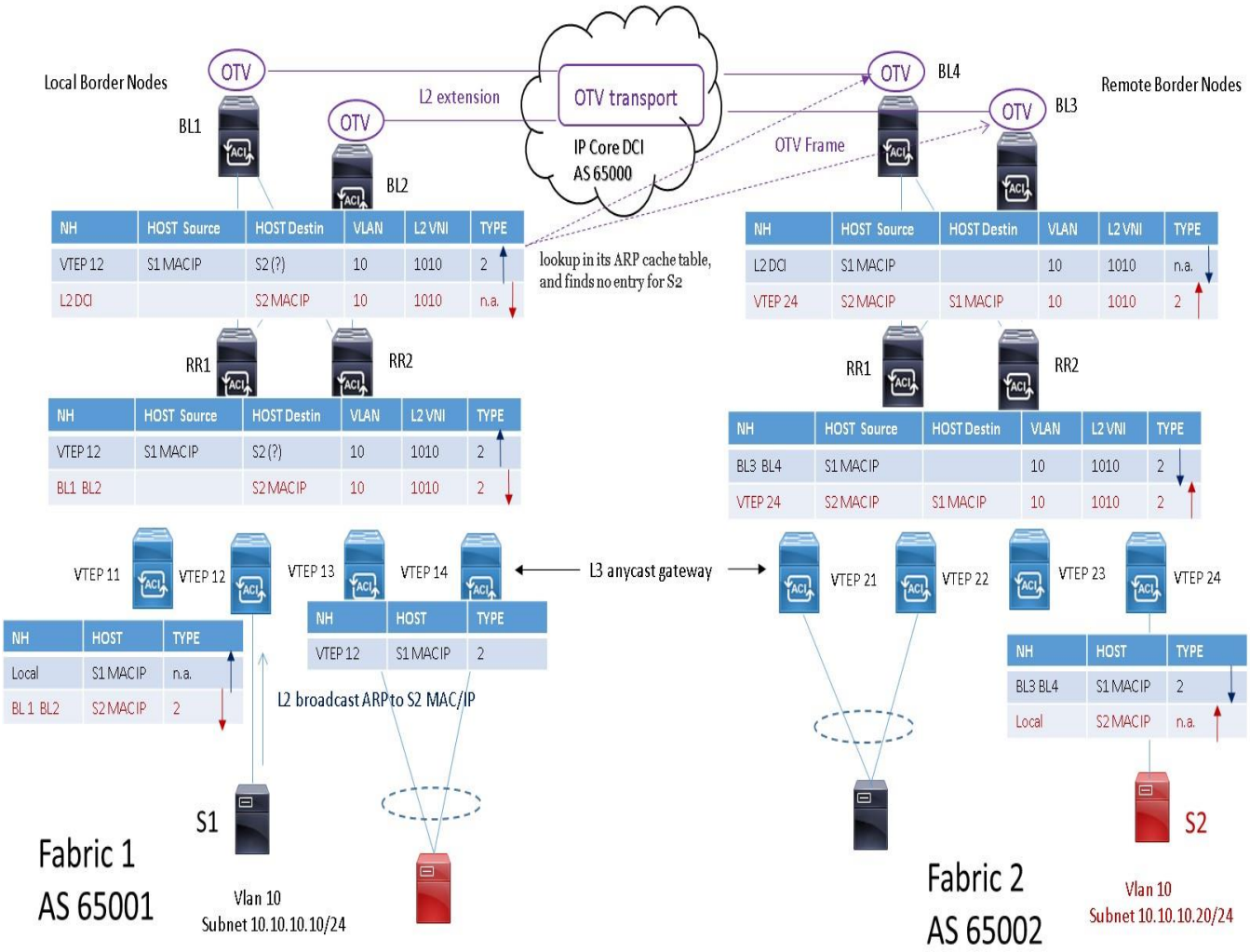

Figura 1: *esempio di learning process MAC IP host intra-subnet IP*

Inter-Subnet communication via Fabrics EVPN

In questo caso abbiamo sempre la comunicazione tra due endpoint EVPN ubicati in differenti Fabrics, ma con due differenti subnets IP default gateway.

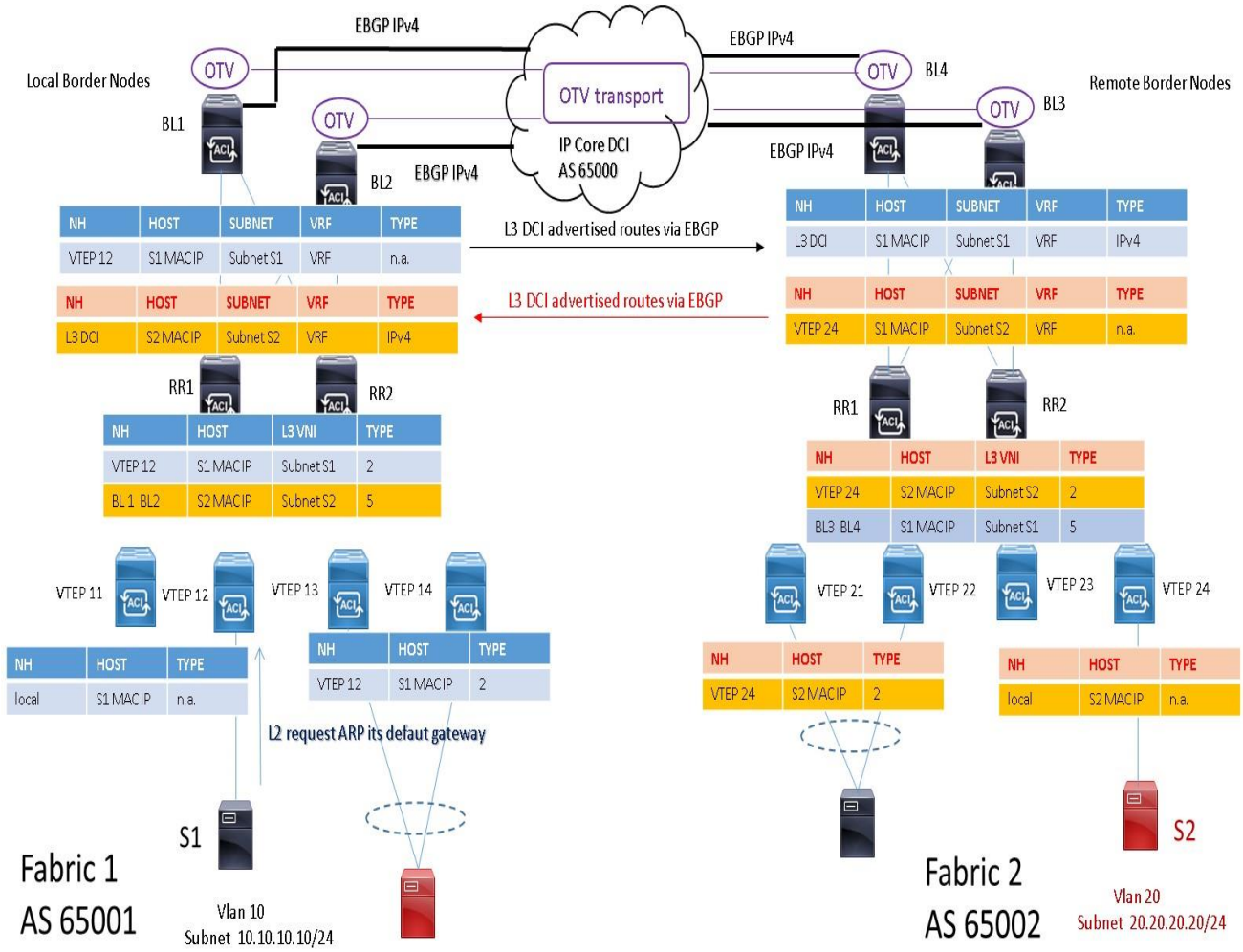

Figura 2: *esempio di communication endpoint with different subnet IP between Fabrics EVPN*

Nota:

Quando si configura un anycast gateway vMAC address attraverso Fabrics VXLAN, gli OTV devices ad ogni sites continueranno ad aggiornare le loro tabelle di forwarding (L2 table) affinchè possano continuare a ricevere sulle loro internal interface le richieste ARP trasmesse dagli endpoint connessi localmente.

E' una buona pratica applicare una route-map al piano di controllo OTV per evitare comunicazioni anycast gateway MAC address information tra OTV devices tra siti remoti; è possibile applicare una route-map via OTV IS-IS control-plane come nel seguente esempio:

```
mac-list anycast_GW_MAC_deny seq 10 deny 0001.0001.0001 ffff.ffff.ffff
mac-list anycast_GW_MAC_deny seq 20 permit 0000.0000.0000 0000.0000.0000
route-map anycast_GW_MAC_filter permit 10
  match mac-list anycast_GW_MAC_deny
!
otv-isis default
  vpn Overlay0
    redistribute filter route-map anycast_GW_MAC_filter
```

Diagramma di rete IP di TEST

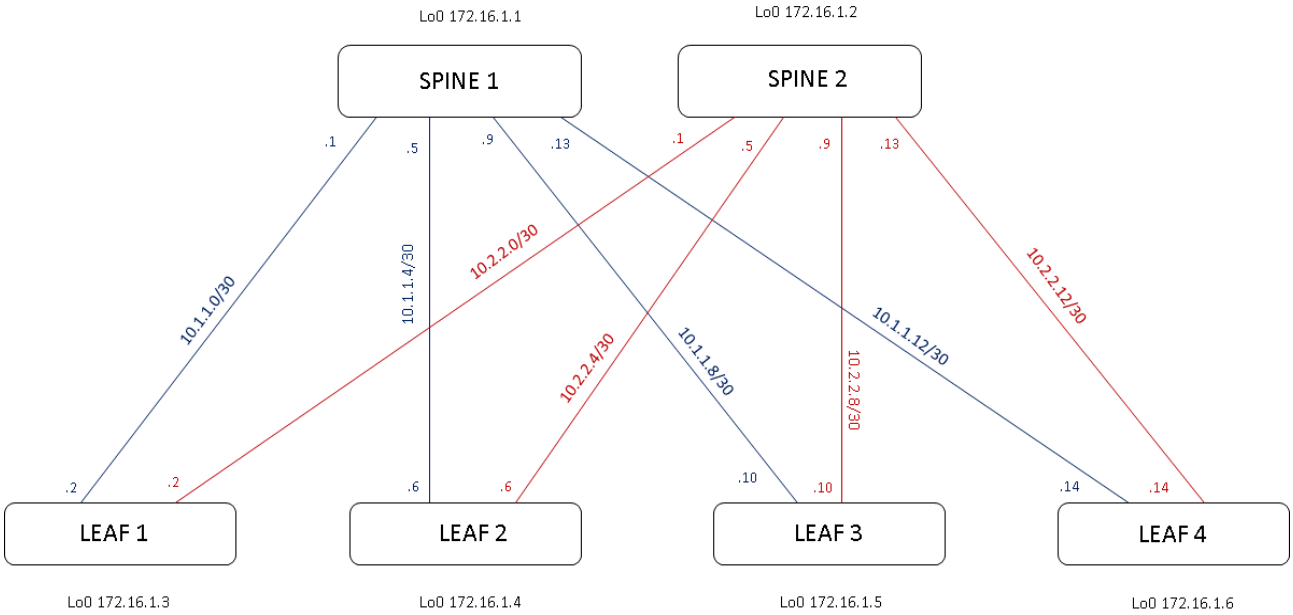

Figura 3: *architettura Data Center Spine Leaf di test*

Leaf Nodes stabiliscono sessioni IBGP EVPN con i Spine Nodes; quest'ultimi hanno ruolo di Router Reflector EVPN e scambiano informazioni layer 2 e layer 3 tra VTEP (Leaf Nodes)

Diagramma di rete MP-BGP EVPN

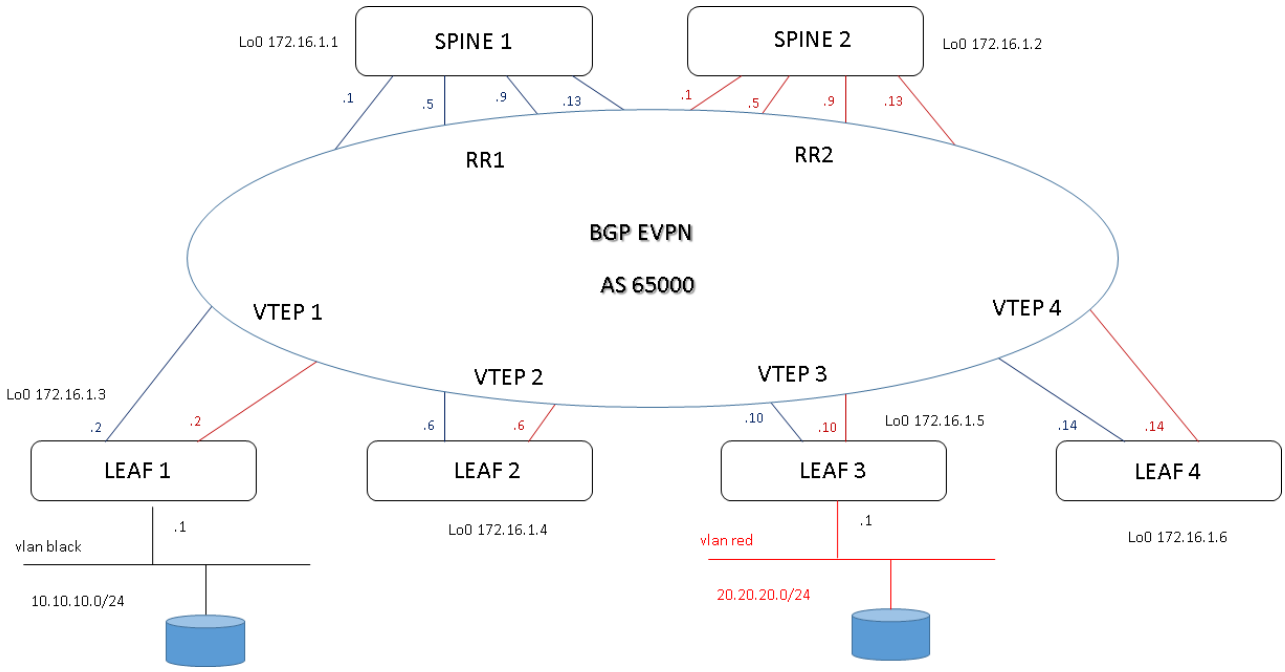

Figura 4: *architettura Data Center Spine MP-BGP overlay di test*

Le sessioni IBGP sono stabilite su base loopback; pertanto è necessario un protocollo IGP (OSPF, ISIS) per la redistribuzione delle looopback VTEPs

Considerazioni BGP EVPN design (EBGP or IBGP)

In genere un data centers IaaS costruito su una architettura Spine-Leaf utilizza per migliorare le sue performance di raggiungibilità layer 2 e 3, un processo ECMP (Equal Cost Multi Path) via IGP.

In caso di crescita della Fabric con la separazione multi-tenant, si può pensare a meccanismi di scalabilità come il protocollo BGP e scegliere se utilizzare Internal-BGP oppure external in considerazione anche di meccanismi ECMP molto utili in ambienti datacenters

IBGP richiede sessioni tra tutti i PE VTEP e l'impiego di Router Reflector aiuta molto in termini di scalabilità delle sessioni configurati a livello Spine; questo tipo standard di soluzione, in ogni caso, riflette solo il best-single-prefix verso i loro client ed nella soluzione di utilizzare ECMP bisogna configurare un BGP addpath feature per aggiungere ECMP all'interno degli annuncia da parte dei RRs

EBGP, invece, supporta ECMP senza addpath ed è semplice nella sua tradizionale configurazione; con EBGP ogni devices della Fabric utilizza un proprio AS (Autonomous System)

Di seguito una rappresentazione grafica delle due soluzioni BGP:

EVPN E-BGP and ASN underlay design

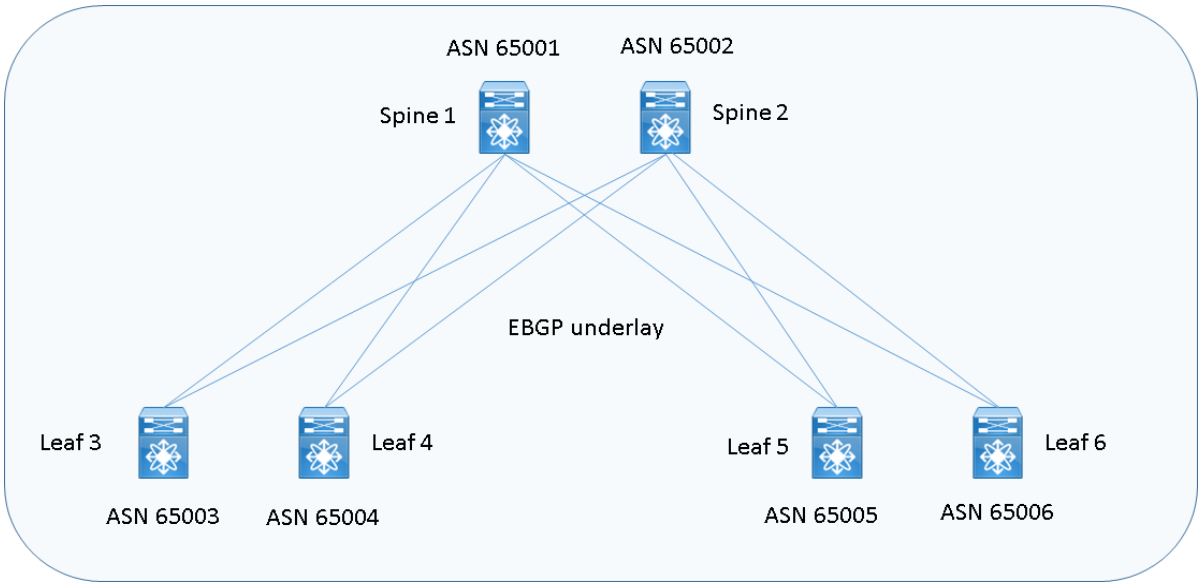

Figura 5: *architettura Data Center Spine E-BGP one-Fabric underlay design*

EVPN I-BGP with Router Reflector one-Fabric overlay design

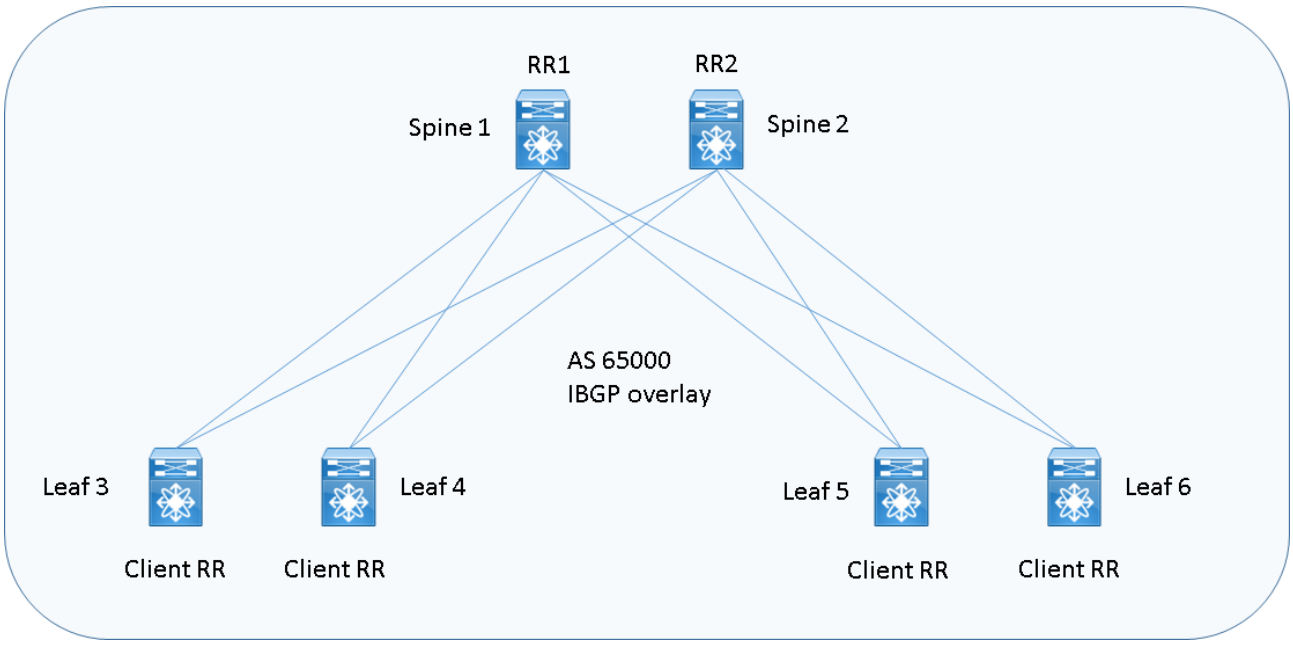

Figura 6: *architettura Data Center Spine I-BGP one-Fabric overlay design*

CONFIGURATION EVPN Fabric with Router Reflector IBGP

Enable Feature config

feature bgp #activate bgp protocol that will be used for L2VPN EVPN address-family

feature vn-segment-VLAN-based #this feature allow you to map a VNI to a VLAN

feature nv overlay #this is VXLAN Feature

feauture nv overlay evpn

Other features need to be activated for your underlay infrastructure like:

feature ospf

feature pim

feature interface-VLAN

VTEP 1 config

Vlans database, fabric forwarding anycast-gateway-mac and pim multicast configuration parameters

vlan 1,10,20,30

!

```
fabric forwarding anycast-gateway-mac 0001.0001.0001
```

!

```
ip pim rp-address 172.16.1.1 group-list 224.0.0.0/4
ip pim ssm range 232.0.0.0/8
```

Vlans Black ed associarla ad un segmento VXLAN VNI and L3-VNI intervlan routing

VLAN 10 # vlan 10 is used as Layer 3 VNI to route inter-vlan routing

name L3-VNI

vn-segment 1000010

!

VLAN 20

name BLACK

vn-segment 2000020

EVPN configuration permit the exchange of L2 reachability between VTEPs

```
evpn
vni 2000020 l2
 rd auto                    #  RD is default calculated as VNI:BGP Router ID
route-target import auto          # RT is default calculated as BGP AS:VNI
route-target export auto
```

Definizione layer 3 VRF per inter-VNI traffic

```
vrf context EVPN
vni 1000010
rd auto
address-family ipv4 unicast
route-target both auto
route-target both auto evpn
!
interface VLAN 20
description BLACK
vrf member EVPN
ip address 10.10.10.1/24
no shutdown
fabric forwarding mode anycast-gateway
!
interface VLAN 10                  # Layer 3 VNI associated interface vlan does not have an ip address.
vrf member EVPN
no shutdown
```

Definizione NVE tunnel logical interface where VXLAN packets are encapsulated and decapsulated

```
interface nve1
no shutdown
source-interface loopback0
host-reachability protocol bgp
member vni 1000010 associate-vrf
member vni 2000020
   mcast-group 239.1.1.1
suppress-arp
```

suppress arp permit to VTEP to cache host-reachability information for remote VTEPs and behave later like a proxy-arp when it receives an ARP request from end host and the information is already in his cache table.

Configurazione physical interface and ospf underlay

```
interface Ethernet1/2                # ospf with PIM is used as Underlay.
 description "to Spine"
 no switchport
 ip address 10.1.1.2/30
 ip router ospf UNDERLAY area 0.0.0.0
 ip pim sparse-mode
 no shutdown
!
interface Ethernet1/10               # Port to Host A.
 switchport mode trunk
!
interface loopback01                 # Loopback for BGP Peering.
 description "Loopback for "BGP"
 ip address 172.16.1.3/32
 ip router ospf UNDERLAY area 0.0.0.0
 ip pim sparse-mode
!
router ospf UNDERLAY
```

VTEP 3 config

Vlans database, fabric forwarding anycast-gateway-mac and pim multicast configuration parameters

```
vlan 1,10,20,30
!
fabric forwarding anycast-gateway-mac 0001.0001.0001
!
ip pim rp-address 172.16.1.1 group-list 224.0.0.0/4
ip pim ssm range 232.0.0.0/8
```

Vlans Red ed associarla ad un segmento VXLAN VNI and L3-VNI intervlan routing

```
VLAN 10                    # vlan 10 is used as Layer 3 VNI to route inter-vlan routing
name L3-VNI
vn-segment 1000010
!
VLAN 30
name RED
vn-segment 3000030
```

EVPN configuration permit the exchange of L2 reachability between VTEPs

```
evpn
vni 3000030 l2
 rd auto                  #  RD is default calculated as VNI:BGP Router ID
route-target import auto          # RT is default calculated as BGP AS:VNI
route-target export auto
```

Definizione layer 3 VRF per inter-VNI traffic

```
vrf context EVPN
vni 1000010
rd auto
address-family ipv4 unicast
route-target both auto
route-target both auto evpn
!
```

```
interface VLAN 30
description RED
vrf member EVPN
Ip address 20.20.20.1/24
no shutdown
fabric forwarding mode anycast-gateway
!
interface VLAN 10                    # Layer 3 VNI associated interface vlan does not have an ip address.
vrf member EVPN
no shutdown
```

Definizione NVE tunnel logical interface where VXLAN packets are encapsulated and decapsulated

```
 interface nve1
no shutdown
source-interface loopback0
host-reachability protocol bgp
member vni 10000 associate-vrf

member vni 3000030
  mcast-group 239.1.1.2
  suppress-arp
```

Nota: la configurazione delle interfacce fisiche viste per il VTEP 1 e l'ospf underlay è medesimo al paragrafo precedente.

BGP RR config

```
router bgp 65000
address-family ipv4 unicast
address-family l2vpn evpn
  retain route-target all
template peer IBGP-EVPN
  remote-as 65000
  update-source loopback0
  address-family ipv4 unicast
```

```
    send-community extended
    route-reflector-client
  address-family l2vpn evpn
    send-community extended
    route-reflector-client
neighbor 172.16.1.3
  inherit peer IBGP-EVPN
neighbor 172.16.1.4
  inherit peer IBGP-EVPN
neighbor 172.16.1.5
  inherit peer IBGP-EVPN
neighbor 172.16.1.6
  inherit peer IBGP-EVPN
```

BGP VTEP config

```
router bgp 65000
address-family ipv4 unicast
address-family l2vpn evpn
neighbor 172.16.1.1
  remote-as 65000
  update-source loopback0
  address-family ipv4 unicast
  address-family l2vpn evpn
    send-community extended
neighbor 172.16.1.2
  remote-as 65000
  update-source loopback0
  address-family ipv4 unicast
  address-family l2vpn evpn
    send-community extended
vrf EVPN
 address-family ipv4 unicast
 advertise l2vpn evpn
```

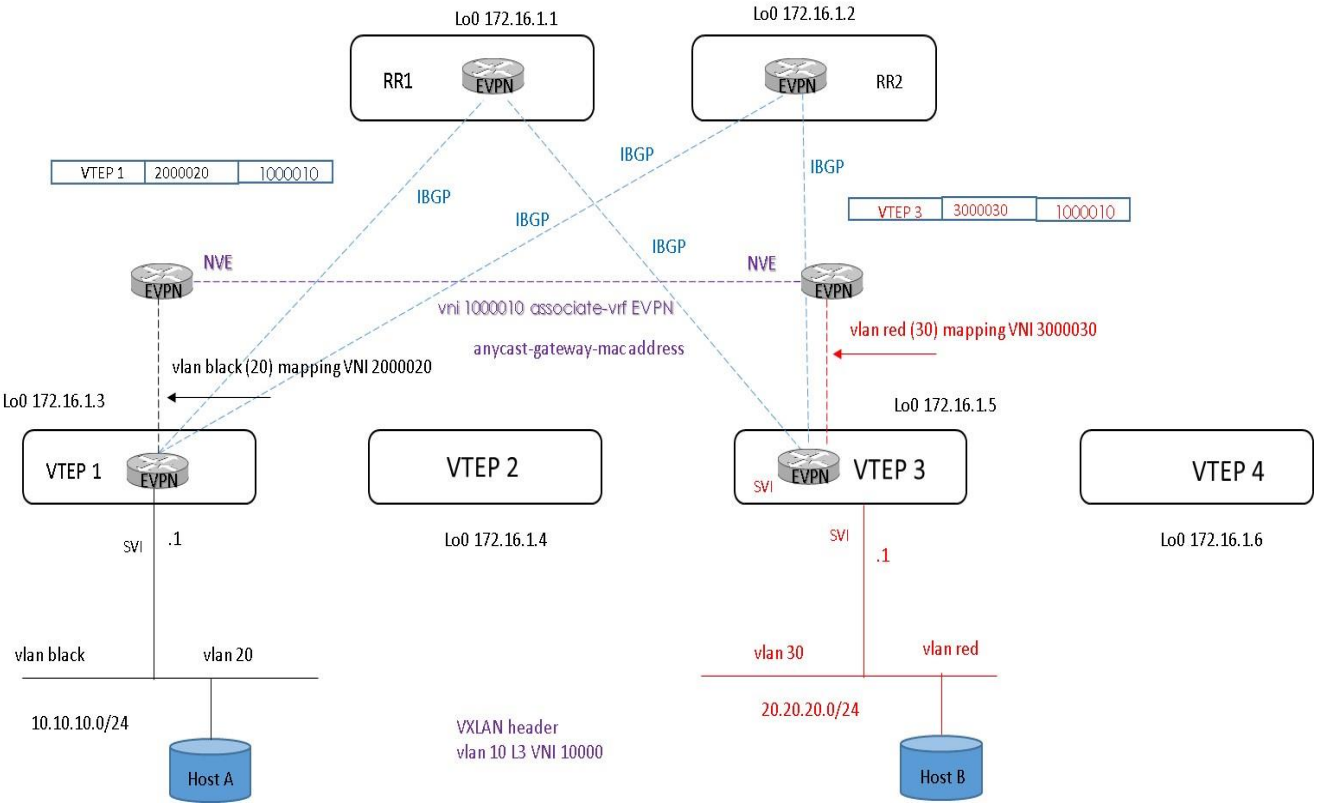

Figura 7: *architettura Data Centers CLOS MP-BGP L2VNI L3VNI*