Configurazione e risoluzione dei problemi di VXLAN vPC Fabric Peering per NXOS

Sommario

Introduzione

In questo documento viene descritto come configurare il peering della struttura vPC per NXOS e il flusso del traffico BUM e verificarlo.

Prerequisiti

Requisiti

Cisco raccomanda la conoscenza dei seguenti argomenti:

- vPC (canale porta virtuale)
- VXLAN (Virtual Extensible LAN)

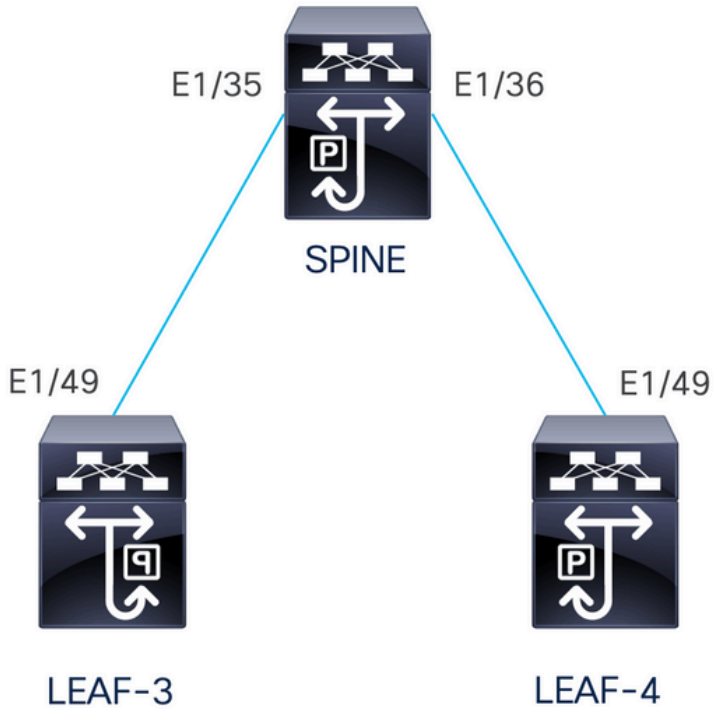
Componenti usati

Le informazioni fornite in questo documento si basano sulle seguenti versioni software e hardware:

- N9K-C93240YC-FX2 per switch Leaf versione: 10.3(3)
- N9K-C9336C-FX2 per switch dorso versione: 10.3(3)

Le informazioni discusse in questo documento fanno riferimento a dispositivi usati in uno specifico ambiente di emulazione. Su tutti i dispositivi menzionati nel documento la configurazione è stata ripristinata ai valori predefiniti. Se la rete è operativa, valutare attentamente eventuali conseguenze derivanti dall'uso dei comandi.

Esempio di rete



vPC Fabric Peering offre una soluzione avanzata di accesso dual-homing senza il sovraccarico dello spreco di porte fisiche per vPC Peer Link. Questa caratteristica conserva tutte le caratteristiche di un vPC tradizionale.

In questa installazione sono presenti Leaf-3 e Leaf-4 configurati come vPC con peer fabric.

Configurazione

Configurazione TCAM

Prima della configurazione, è presente un controllo sulla memoria TCAM:

```
LEAF-4(config-if)# sh hardware access-list tcam region
                                NAT ACL[nat] size = 0
                     Ingress PACL [ing-ifacl] size = 0
                                  VACL [vacl] size = 0
                      Ingress RACL [ing-racl] size = 2304
                   Ingress L2 QOS [ing-l2-qos] size = 256
       Ingress L3/VLAN QOS [ing-l3-vlan-qos] size = 512
                       Ingress SUP [ing-sup] size = 512
 Ingress L2 SPAN filter [ing-l2-span-filter] size = 256
 Ingress L3 SPAN filter [ing-l3-span-filter] size = 256
                    Ingress FSTAT [ing-fstat] size = 0
                                  span [span] size = 512
                      Egress RACL [egr-racl] size = 1792
                        Egress SUP [egr-sup] size = 256
             Ingress Redirect [ing-redirect] size = 0
                  Egress L2 QOS [egr-l2-qos] size = 0
        Egress L3/VLAN QOS [egr-l3-vlan-qos] size = 0
     Ingress Netflow/Analytics [ing-netflow] size = 512
                       Ingress NBM [ing-nbm] size = 0
                        TCP NAT ACL[tcp-nat] size = 0
          Egress sup control plane[egr-copp] size = 0
   Ingress Flow Redirect [ing-flow-redirect] size = 0 <<<<<<<<
Ingress PACL IPv4 Lite [ing-ifacl-ipv4-lite] size = 0
Ingress PACL IPv6 Lite [ing-ifacl-ipv6-lite] size = 0
                  Ingress CNTACL [ing-cntacl] size = 0
                   Egress CNTACL [egr-cntacl] size = 0
                     MCAST NAT ACL[mcast-nat] size = 0
                      Ingress DACL [ing-dacl] size = 0
     Ingress PACL Super Bridge [ing-pacl-sb] size = 0
   Ingress Storm Control [ing-storm-control] size = 0
         Ingress VACL redirect [ing-vacl-nh] size = 0
                      Egress PACL [egr-ifacl] size = 0
                Egress Netflow [egr-netflow] size = 0
```

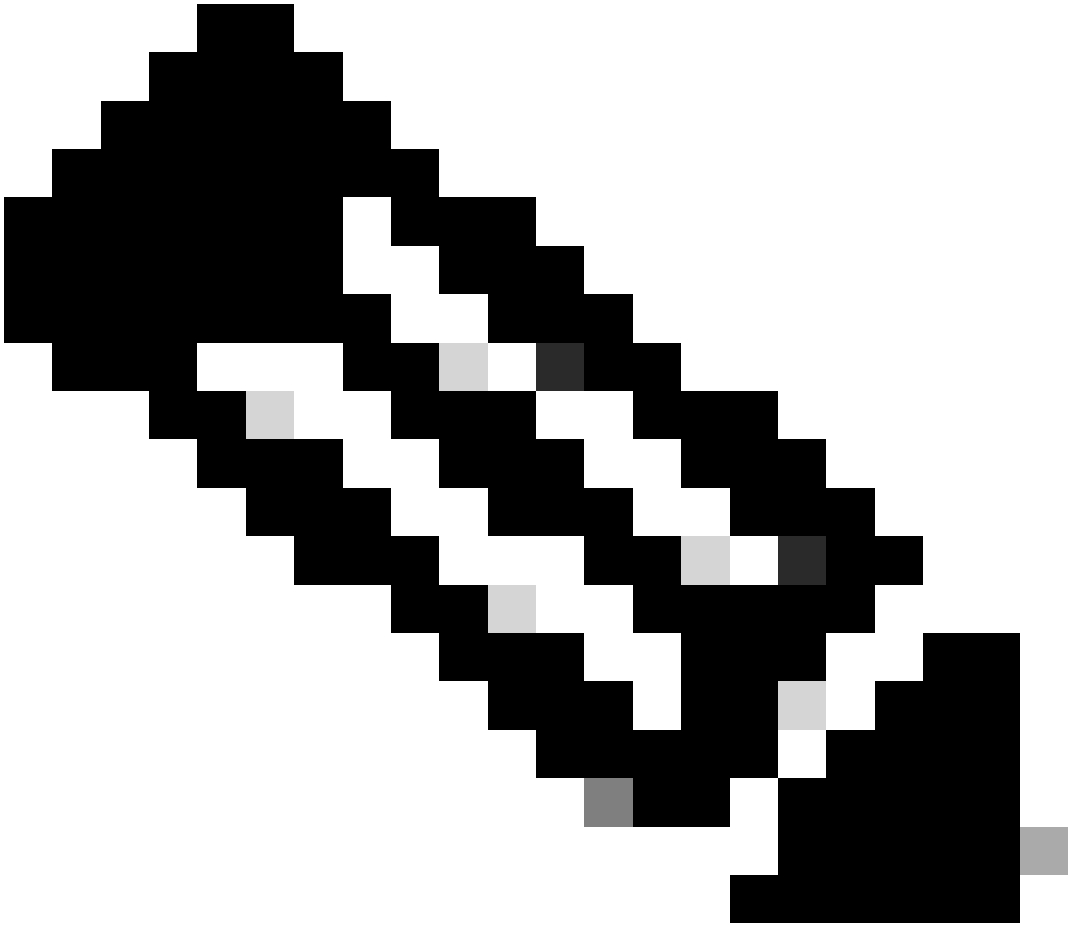
vPC Fabric Peering richiede l'applicazione di sculture TCAM della regione ing-flow-redirect. Per eseguire il taglio TCAM, è necessario salvare la configurazione e ricaricare lo switch prima di utilizzare la funzione.

Questo spazio sulla TCAM è di larghezza doppia, quindi il minimo che possiamo assegnare è 512.

TCAM Scolpendo

In questo scenario, ing-racl ha abbastanza spazio per prendere il 512 e assegnarlo al 512 per il reindirizzamento del flusso di ing.

```
LEAF-4(config-if)# hardware access-list tcam region ing-racl 1792
Please save config and reload the system for the configuration to take effect

LEAF-4(config)# hardware access-list tcam region ing-flow-redirect 512
Please save config and reload the system for the configuration to take effect
```

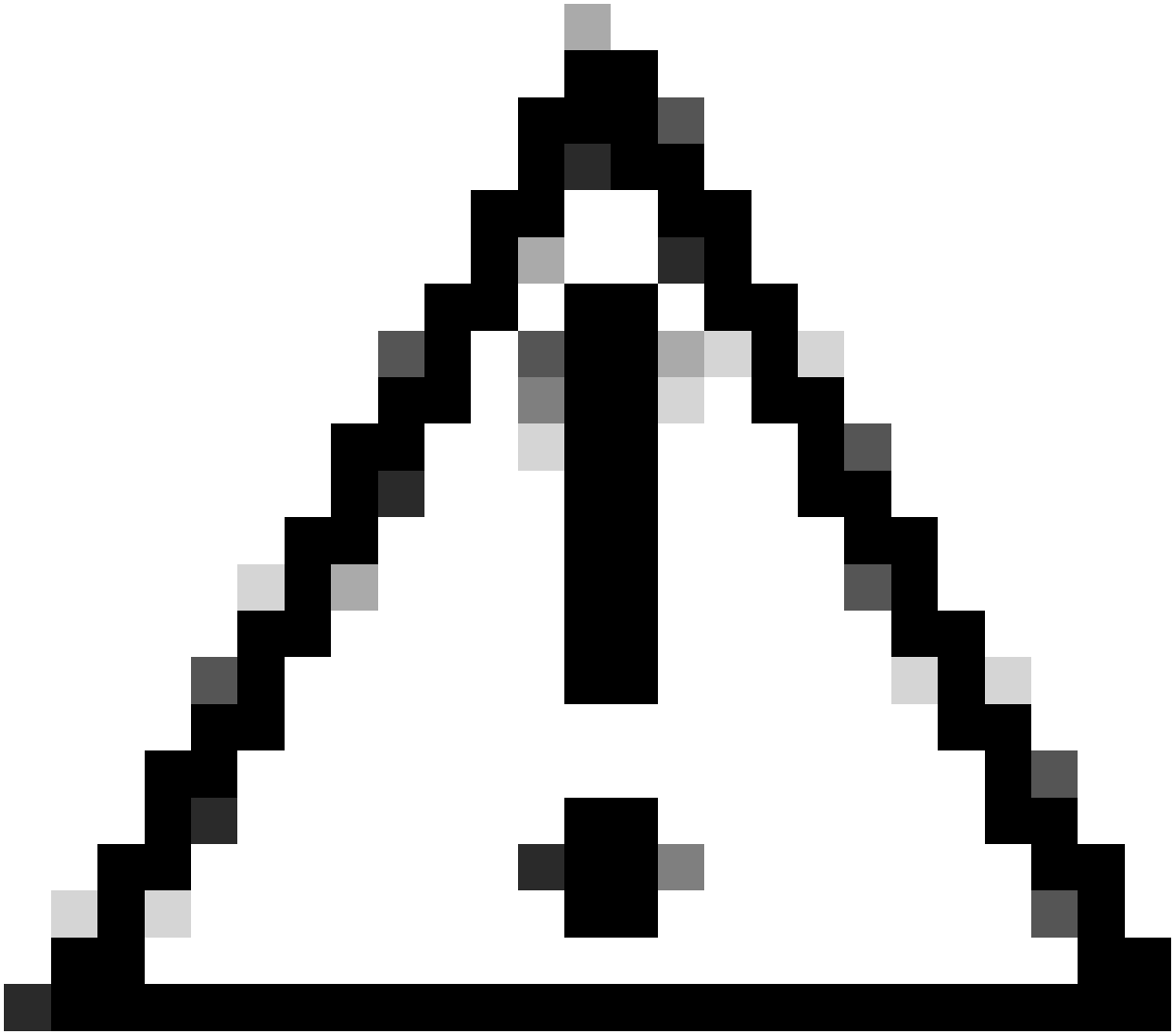


Nota: quando si configura il peering del fabric vPC tramite DCNM, l'intaglio TCAM verrà eseguito, ma è necessario un ricaricamento per avere effetto

Una volta apportata la modifica, questa si rifletterà sul comando:

```
513E-B-11-N9K-C93240YC-FX2-4# sh hardware access-list tcam region
                                       NAT ACL[nat] size = 0
                              Ingress PACL [ing-ifacl] size = 0
                                        VACL [vacl] size = 0
                                Ingress RACL [ing-racl] size = 2304
                           Ingress L2 QOS [ing-l2-qos] size = 256
                 Ingress L3/VLAN QOS [ing-l3-vlan-qos] size = 512
                                Ingress SUP [ing-sup] size = 512
          Ingress L2 SPAN filter [ing-l2-span-filter] size = 256
          Ingress L3 SPAN filter [ing-l3-span-filter] size = 256
                           Ingress FSTAT [ing-fstat] size = 0
                                        span [span] size = 512
                              Egress RACL [egr-racl] size = 1792
                                Egress SUP [egr-sup] size = 256
```

```
            Ingress Redirect [ing-redirect] size = 0
                 Egress L2 QOS [egr-l2-qos] size = 0
       Egress L3/VLAN QOS [egr-l3-vlan-qos] size = 0
     Ingress Netflow/Analytics [ing-netflow] size = 512 <<<<<
                    Ingress NBM [ing-nbm] size = 0
                     TCP NAT ACL[tcp-nat] size = 0
         Egress sup control plane[egr-copp] size = 0
   Ingress Flow Redirect [ing-flow-redirect] size = 0
Ingress PACL IPv4 Lite [ing-ifacl-ipv4-lite] size = 0
Ingress PACL IPv6 Lite [ing-ifacl-ipv6-lite] size = 0
                Ingress CNTACL [ing-cntacl] size = 0
                 Egress CNTACL [egr-cntacl] size = 0
                   MCAST NAT ACL[mcast-nat] size = 0
                    Ingress DACL [ing-dacl] size = 0
     Ingress PACL Super Bridge [ing-pacl-sb] size = 0
   Ingress Storm Control [ing-storm-control] size = 0
         Ingress VACL redirect [ing-vacl-nh] size = 0
                    Egress PACL [egr-ifacl] size = 0
```



Attenzione: assicurarsi che il dispositivo venga ricaricato dopo le modifiche apportate al TCAM, altrimenti il VPC non si accenderà a causa di modifiche non applicate al TCAM.

Configurazione per vPC

Dominio VPC

Su LEAF-3 e LEAF-4 nel dominio VPC la configurazione è quella di specificare gli indirizzi IP per il keep-alive e il collegamento peer virtuale

```
vpc domain 1
  peer-keepalive destination 192.168.1.1 source 192.168.1.2 vrf management
  virtual peer-link destination 10.10.10.2 source 10.10.10.1 dscp 56

interface port-channel1
  vpc peer-link
```

Keep-alive

Qualsiasi collegamento diretto di layer 3 tra peer vPC deve essere utilizzato solo per il mantenimento in attività dei peer. Deve trovarsi in un VRF separato dedicato solo al keep-alive. In questo scenario, viene usata la gestione dell'interfaccia dello switch.

```
LEAF-3
interface mgmt0
  vrf member management
  ip address 192.168.1.1/24

LEAF-4
interface mgmt0
  vrf member management
  ip address 192.168.1.2/24
```

Interfaccia di layer 3 per il collegamento peer virtuale

L'interfaccia di layer 3 utilizzata per il collegamento peer virtuale non deve essere la stessa utilizzata per il keep-alive, è possibile utilizzare lo stesso loopback utilizzato per l'underlay o può essere un loopback dedicato sul Nexus

Qui il loopback0 è per l'underlay e il loopback2 è un loopback dedicato per il peer-link virtuale, mentre loopback1 è l'interfaccia associata alla nostra interfaccia NVE.

```
LEAF-3
interface loopback0
  ip address 10.1.1.1/32
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
```

```
interface loopback1
  ip address 172.16.1.2/32
  ip address 172.16.1.1/32 secondary
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface loopback2
  ip address 10.10.10.2/32
  ip router ospf 1 area 0.0.0.0


LEAF-4
interface loopback0
  ip address 10.1.1.2/32
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface loopback1
  ip address 172.16.1.3/32
  ip address 172.16.1.1/32 secondary
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface loopback2
  ip address 10.10.10.1/32
  ip router ospf 1 area 0.0.0.0
```

VPC Peer-link

Al collegamento peer deve essere assegnato un canale della porta anche se non si intende assegnare un'interfaccia fisica al canale della porta.

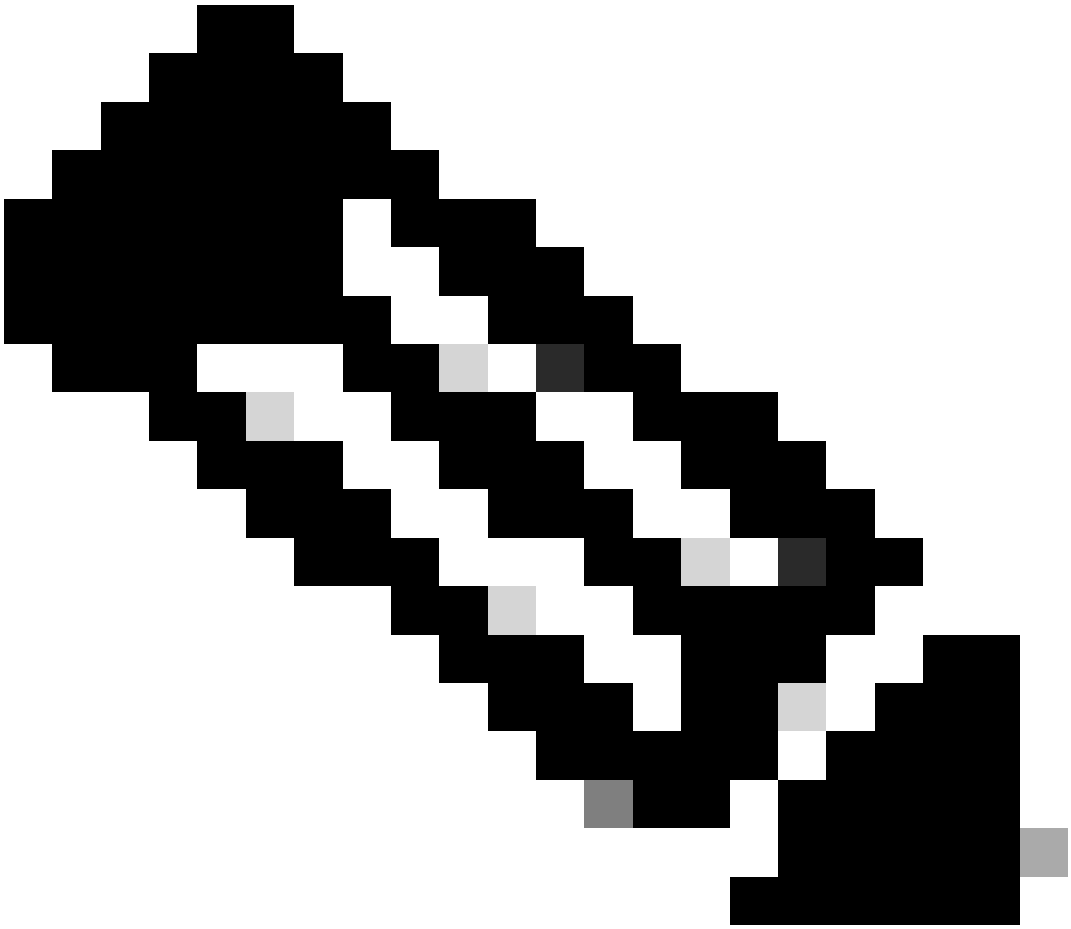
```
LEAF-3(config-if)# sh run interface port-channel 1 membership

interface port-channel1
  switchport
  switchport mode trunk
  spanning-tree port type network
  vpc peer-link
```

Collegamenti crescenti

L'ultima parte della configurazione consiste nel configurare i collegamenti su entrambi i lati verso la SPINE con il comando port-type fabric.

```
interface Ethernet1/49
  port-type fabric <<<<<<<<
  medium p2p
  ip unnumbered loopback0
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
```

```
no shutdown
```



> Nota: se non si configura la struttura di tipo porta, non è possibile visualizzare il keep-alive generato da Nexus

---

Configurazione SPINES

Sugli spine si consiglia di impostare QoS in modo che corrisponda al valore DSCP configurato sul dominio VPC, poiché il collegamento peer del fabric vPC è stabilito sulla rete di trasporto.

Le informazioni sul control plane CFS utilizzate per sincronizzare le informazioni sullo stato delle porte, le informazioni sulla VLAN, il mapping tra VLAN e VNI, gli indirizzi MAC dell'host e i gruppi di snooping IGMP vengono trasmesse sulla struttura. I messaggi CFS sono contrassegnati con il valore DSCP appropriato, che deve essere protetto nella rete di trasporto.

```
class-map type qos match-all CFS
  match dscp 56

policy-map type qos CFS
  class CFS
    Set qos-group 7 <<< Depending on the platform it can be 4

interface Ethernet 1/35-36
  service-policy type qos input CFS
```

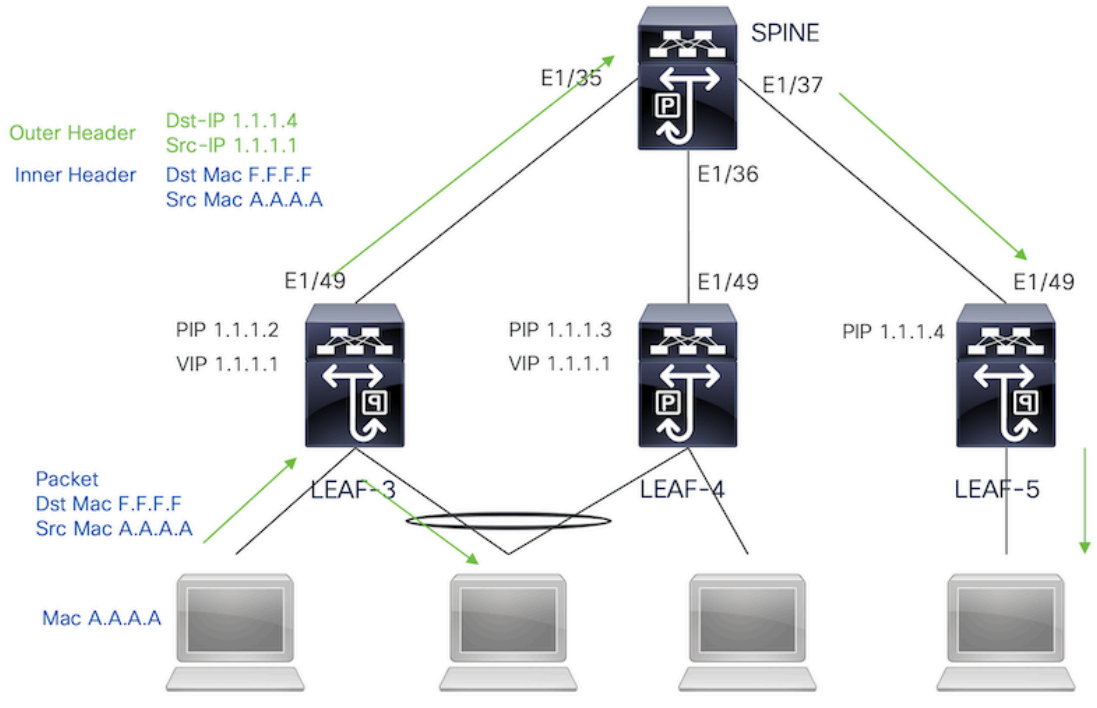
Traffico broadcast, unicast sconosciuto e multicast con incapsulamento della replica in ingresso

Quando il nexus riceve un pacchetto che deve essere trasmesso genera 2 copie del pacchetto.

1. A tutti i VTEPS remoti nell'elenco di inondazioni per il VNI, comprese le porte di accesso locale

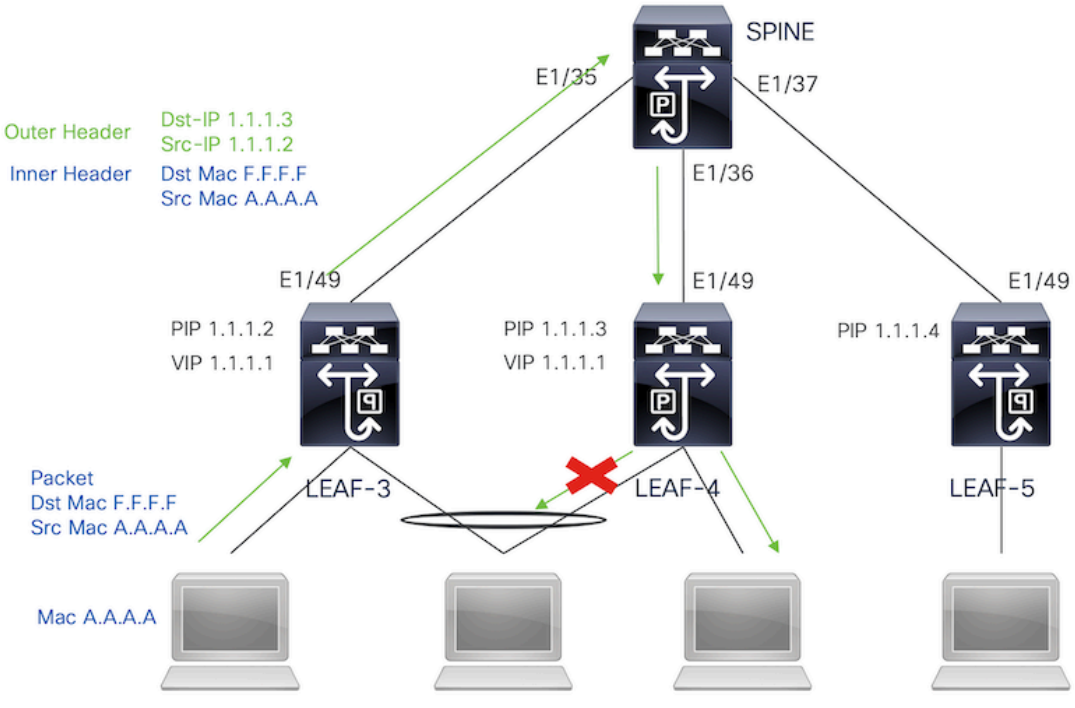
2. Al peer VPC remoto

Per la prima copia, il Nexus incapsulava il traffico usando l'IP di origine dell'indirizzo IP secondario e l'IP di destinazione del VTEP remoto e anche le porte di accesso locale.



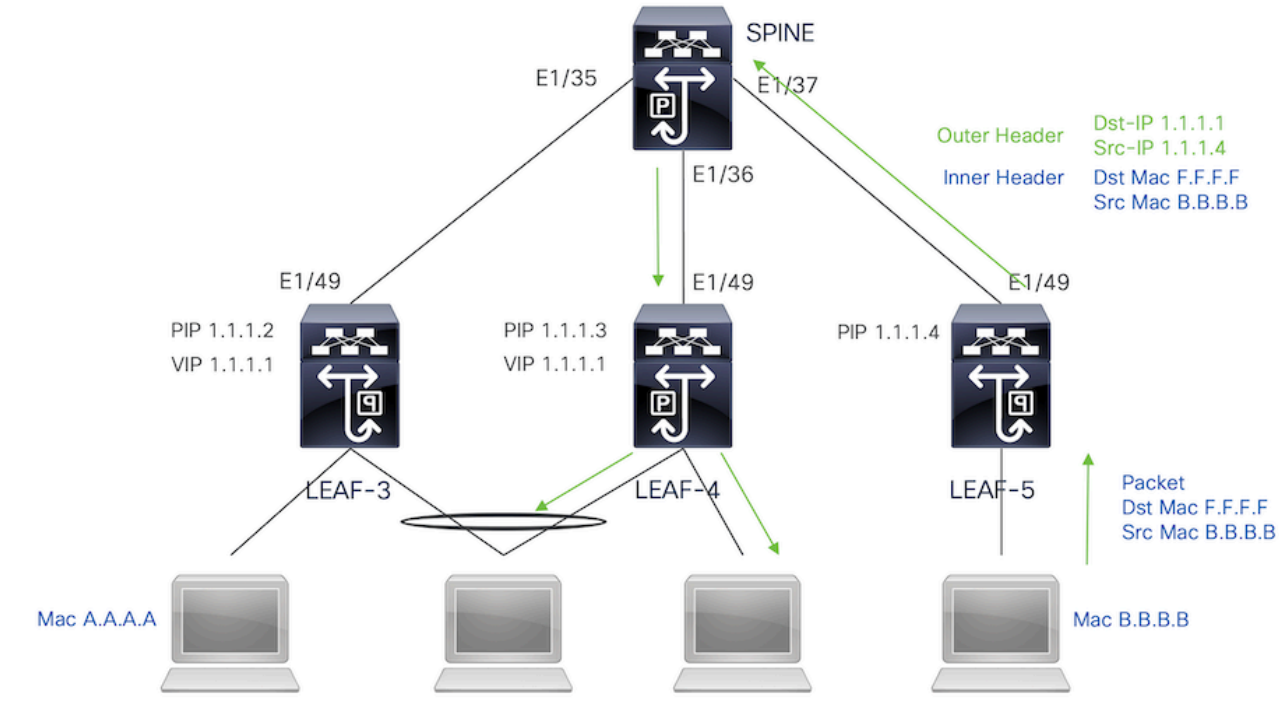
Per la seconda copia verrà inviata al peer VPC remoto. L'IP di origine sarà l'indirizzo primario del loopback e l'IP di destinazione sarà il PIP del peer VPC remoto.

Una volta ricevuto il pacchetto dalla spine, il VTEP remoto inoltrerà il pacchetto solo alle porte orfane.



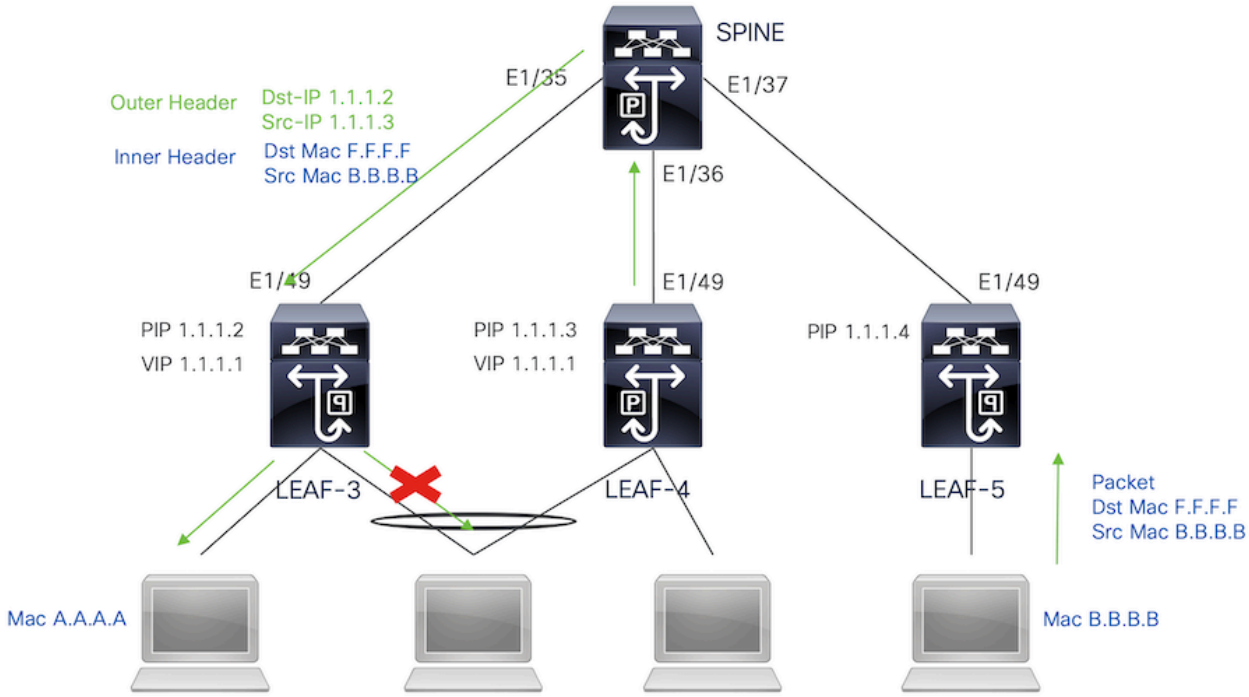
Traffico broadcast, unicast sconosciuto e multicast con decapsulamento replica in ingresso

Poiché l'IP di destinazione per il traffico BUM ricevuto da un altro VTEP è il VIP, il traffico hash su uno dei dispositivi VPC, decapsula il pacchetto e lo invia alle porte di accesso.



Per consentire al traffico di raggiungere le porte orfane connesse sul peer VPC remoto, il nexus genera una copia del pacchetto e lo invia solo al VPC remoto utilizzando l'indirizzo IP primario come IP di origine/destinazione.

Una volta ricevuto sul peer vpc remoto, il nexus decapsula il traffico e lo inoltra solo alle porte orfane.



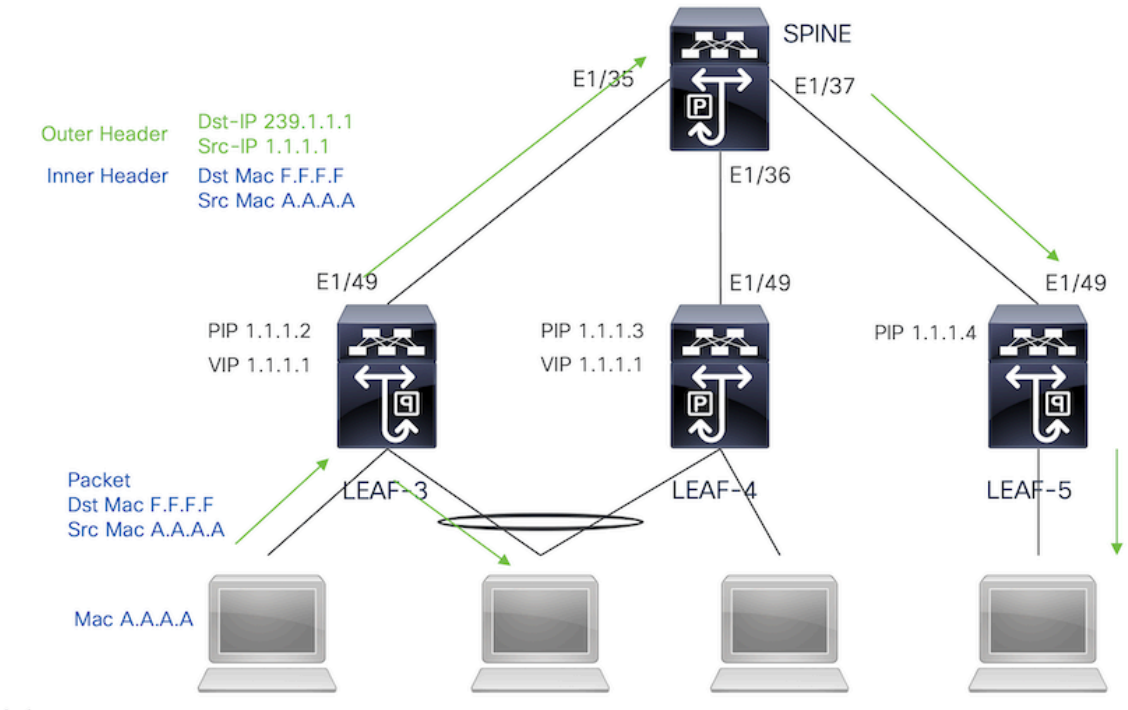
Traffico broadcast, unicast sconosciuto e multicast con incapsulamento multicast

Quando il nexus riceve un pacchetto che deve essere trasmesso genera 2 copie del pacchetto.

1. Il pacchetto verrà inviato a tutti gli OIF nella voce multicast S,G, incluse le porte di accesso locale

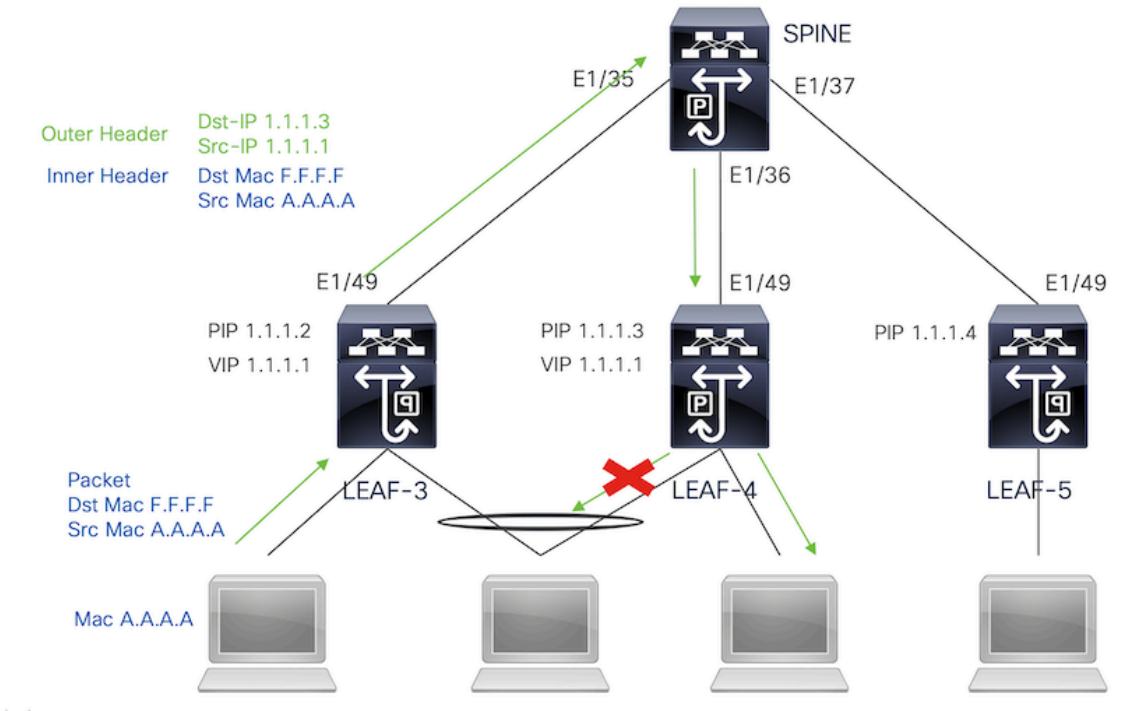
2. Al peer VPC remoto

Per la prima copia, il Nexus ha incapsulato il traffico utilizzando l'IP di origine dell'indirizzo IP secondario e l'IP di destinazione del gruppo multicast configurato.



Per la seconda copia verrà inviata al peer VPC remoto, l'IP di origine sarà il secondo del loopback e l'IP di destinazione sarà il PIP del peer VPC remoto.

Una volta ricevuto il pacchetto dalla spine, il VTEP remoto inoltra il pacchetto solo alle porte orfane.

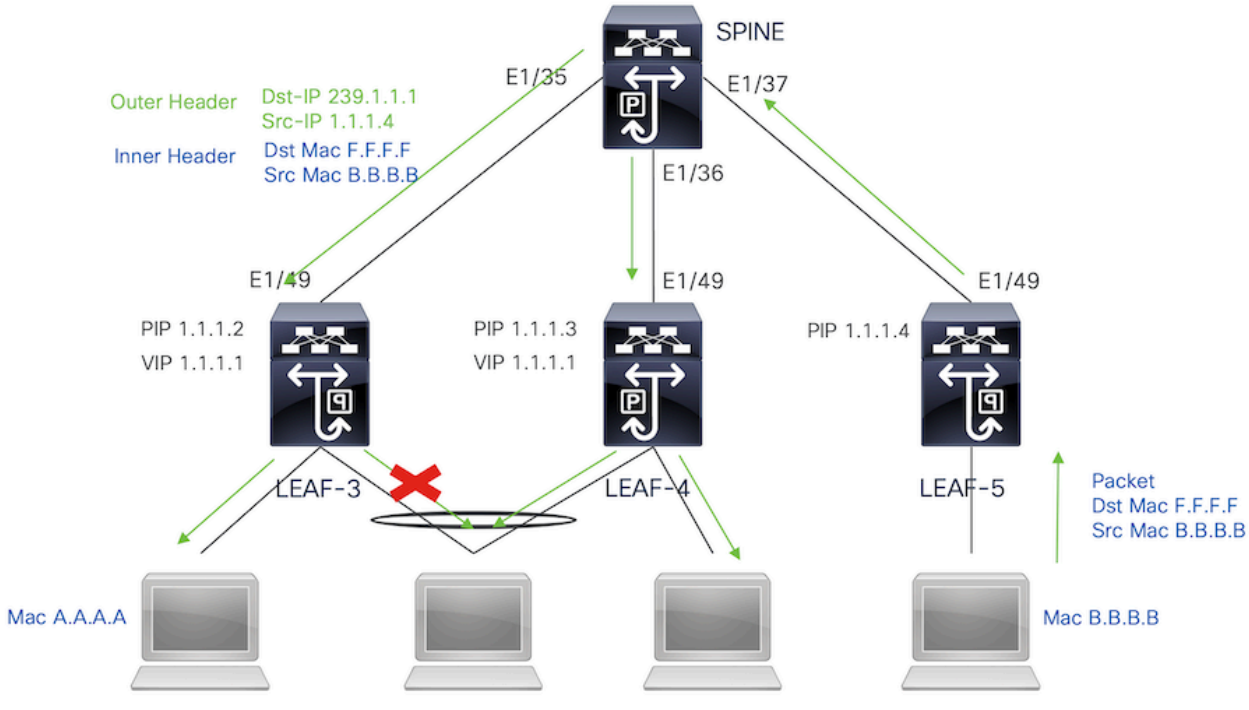


Traffico broadcast, unicast sconosciuto e multicast con decapsulamento multicast

Per il processo di decapsulamento, il pacchetto arriverà a entrambi i peer VPC. Solo un dispositivo

VPC inoltrerà il traffico attraverso i canali delle porte VPC. La decisione verrà presa dal server d'inoltro visualizzato nel comando.

```
module-1# show forwarding internal vpc-df-hash

VPC DF: FORWARDER
```



Verifica

Per verificare che il VPC sia attivo, eseguire i comandi seguenti:

Verificare la raggiungibilità degli indirizzi IP utilizzati per il collegamento peer virtuale.

```
LEAF-3# sh ip route 10.10.10.1
IP Route Table for VRF "default"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

10.10.10.1/32, ubest/mbest: 1/0
    *via 192.168.120.1, Eth1/49, [110/3], 01:15:01, ospf-1, intra

LEAF-3# ping 10.10.10.1
PING 10.10.10.1 (10.10.10.1): 56 data bytes
64 bytes from 10.10.10.1: icmp_seq=0 ttl=253 time=0.898 ms
64 bytes from 10.10.10.1: icmp_seq=1 ttl=253 time=0.505 ms
64 bytes from 10.10.10.1: icmp_seq=2 ttl=253 time=0.433 ms
64 bytes from 10.10.10.1: icmp_seq=3 ttl=253 time=0.465 ms
```

```
64 bytes from 10.10.10.1: icmp_seq=4 ttl=253 time=0.558 ms
```

```
LEAF-3(config-if)# show vpc brief
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link

vPC domain id                          : 1
Peer status                            : peer adjacency formed ok <<<<
vPC keep-alive status                  : peer is alive <<<<
Configuration consistency status       : success
Per-vlan consistency status            : success
Type-2 consistency status              : success
vPC role                               : secondary
Number of vPCs configured              : 0
Peer Gateway                           : Disabled
Dual-active excluded VLANs             : -
Graceful Consistency Check             : Enabled
Auto-recovery status                   : Disabled
Delay-restore status                   : Timer is off.(timeout = 30s)
Delay-restore SVI status               : Timer is off.(timeout = 10s)
Delay-restore Orphan-port status       : Timer is off.(timeout = 0s)
Operational Layer3 Peer-router         : Disabled
Virtual-peerlink mode                  : Enabled <<<<<<<

vPC Peer-link status
---------------------------------------------------------------------
id   Port   Status  Active vlans
--   ----   ------  --------------------------------------------------
1    Po1    up      1,10,50,600-604,608,610-611,614-618,638-639,
                    662-663,701-704
```

Per controllare i ruoli per il VPC, eseguire il comando:

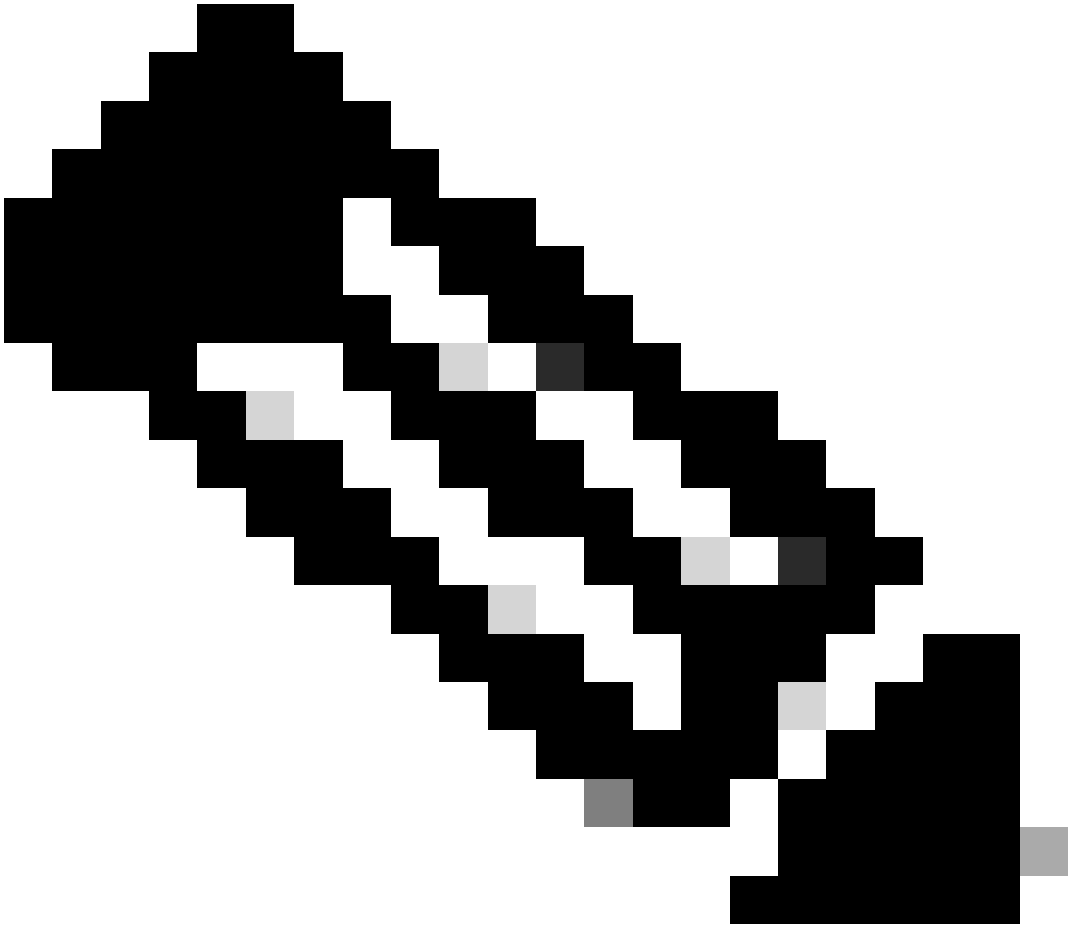
```
LEAF-3(config-if)# sh vpc role

vPC Role status
----------------------------------------------------
vPC role                              : secondary <<<<
Dual Active Detection Status          : 0
vPC system-mac                        : 00:23:04:ee:be:01
vPC system-priority                   : 32667
vPC local system-mac                  : d0:e0:42:e2:09:6f
vPC local role-priority               : 32667
vPC local config role-priority        : 32667
vPC peer system-mac                   : 2c:4f:52:3f:46:df
vPC peer role-priority                : 32667
vPC peer config role-priority         : 32667
```

Tutte le vlan consentite nel canale della porta del collegamento peer devono essere mappate su un VNI, nel caso in cui non vengano visualizzate come incoerenti

```
LEAF-3(config-if)# show vpc virtual-peerlink vlan consistency
Following vlans are inconsistent
1 608 610 611 614 615 616 617 618 638 639 701 702 703 704
```

Per verificare che la configurazione sui collegamenti attivi sia programmata correttamente, eseguire il comando:

```
LEAF-3(config-if)# show vpc fabric-ports
Number of Fabric port : 1
Number of Fabric port active : 1

Fabric   Ports      State
------------------------------------
Ethernet 1/49       UP
```



Nota: il NVE e l'interfaccia di loopback associata verranno visualizzati a meno che il VPC

---

non sia attivo.

Informazioni correlate

- Supporto tecnico Cisco e download

Informazioni su questa traduzione

Cisco ha tradotto questo documento utilizzando una combinazione di tecnologie automatiche e umane per offrire ai nostri utenti in tutto il mondo contenuti di supporto nella propria lingua. Si noti che anche la migliore traduzione automatica non sarà mai accurata come quella fornita da un traduttore professionista. Cisco Systems, Inc. non si assume alcuna responsabilità per l’accuratezza di queste traduzioni e consiglia di consultare sempre il documento originale in inglese (disponibile al link fornito).