Indice

Indice delle figure

1 CLOS FABRIC & tecnology

1.1 VXLAN

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.
VXLAN viene supportato da una infrastruttura con le seguenti caratteristiche:

- Multicast
  - IGMPv2 e IGMPv3
  - PIM sparse mode, bidirectional pim
- IGP
  - OSPF
  - ISIS
  - BGP
- Gateway IP
  - VTEP (Vlan Tunnel End-Point) e provvede ad incapsulare e decapsulare servizi layer 2 in VXLAN

1.2 MP-BGP EVPN

EVPN (Ethernet Virtual Private Network) collega un gruppo di users sites usando un virtual bridge layer 2

Tratta indirizzi MAC come address ruotabili e distribuisce queste informazioni via MP-BGP.

Utilizzato in ambienti Data Centers multi-tenancy con end-point virtualizzati; supporta encapsulamento VXLAN e lo scambio di indirizzi IP host e IP-Prefix.

1.3 MP-BGP EVPN

- 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-target 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)

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-BGP 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 add-path feature per aggiungere ECMP all'interno degli annuncia da parte dei RRs
EBGP, invece, supporta ECMP senza add-path ed è semplice nella sua tradizionale configurazione; con EBGP ogni devices della Fabric utilizza un proprio AS (Autonomous System)

1.4 MP-BGP EVPN type route

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;

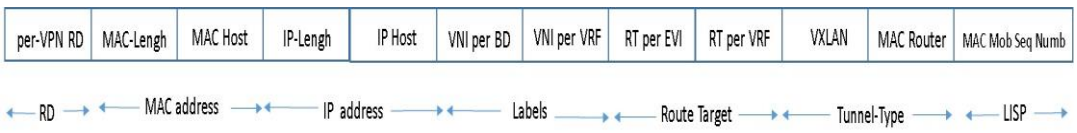

Figura 1: BGP EVPN MAC Route type 2

- ✓ **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

VRF RD	ETH segment	ETH Tag	IP-Lengh	IP Prefix	IP Gateway	L3 VNI	RT per VRF	VXLAN	MAC Router

Figura 2: BGP EVPN IP Prefix Route type 5

1.5 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

1.6 Learning Process End-Point

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.

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 end-point che appartengono ai rispettivi VNI (VXLAN segment Network Identifier) ed essere importate all'interno della propria forwarding table.

1.7 Intra-Subnet communication via Fabric

La comunicazione tra due end-point intra-subnet (stessa subnet IP) 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

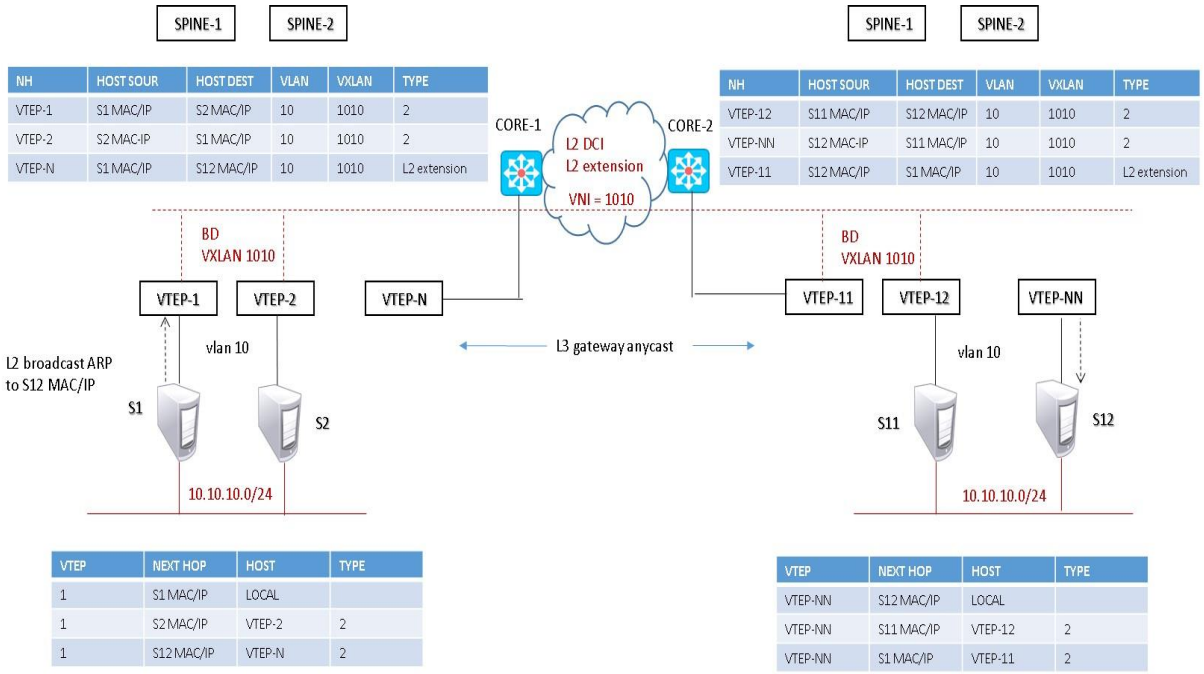

Figura 3: esempio di comunicazione intra-subnet tra CLOS Fabric L2 extension

1.8 Inter-Subnet communication via Fabric

La comunicazione tra due end-point inter-subnet (differente subnet IP) avviene sempre tra due endpoint EVPN ubicati in differenti Fabrics, ma con due differenti subnets IP default gateway.

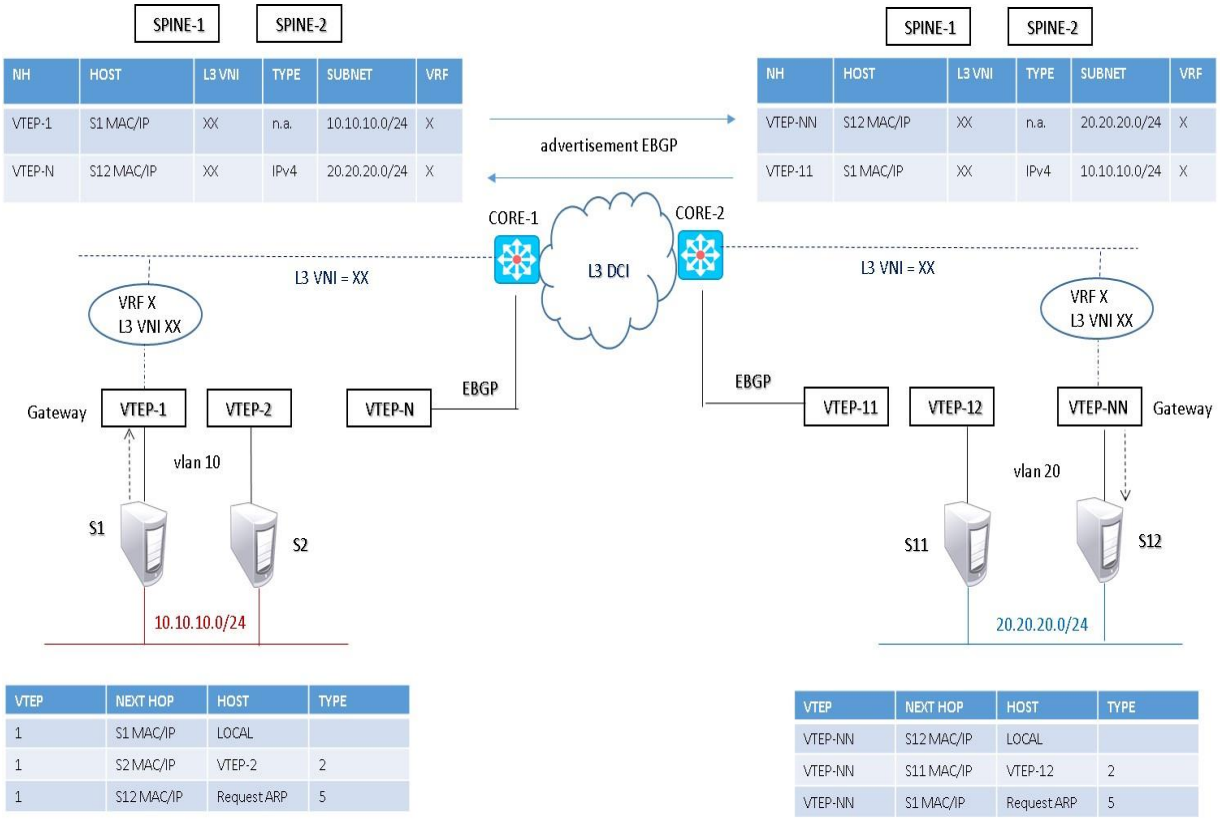

Figura 4: esempio di comunicazione inter-subnet tra CLOS Fabric L3 DCI

2 Architettura CLOS & benefit

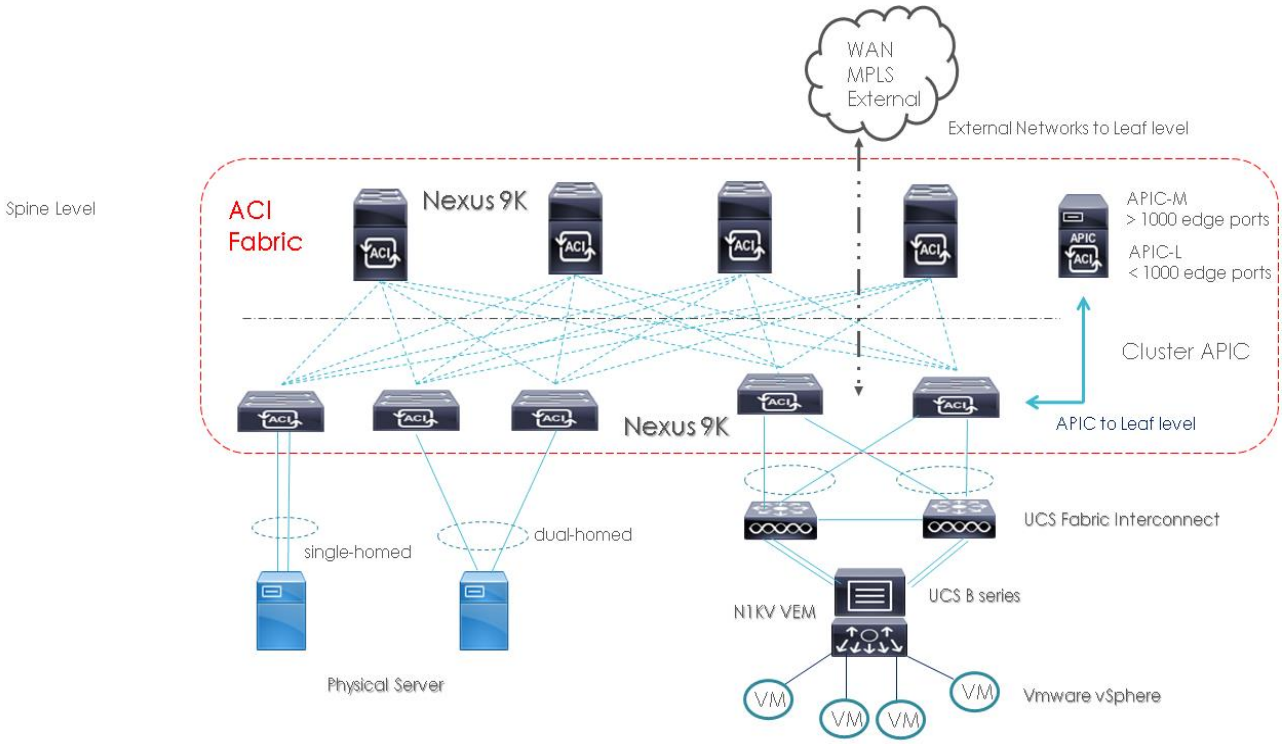

Figura 5: esempio di architettura Nexus N9KCLOS Fabric CISCO

I vantaggi di questa configurazione si possono riassumere:

- ✓ Architettura a due livelli in configurazione FABRIC unico dominio;
- ✓ Alta scalabilità con la possibilità di inserimento di nuovi elementi ed una grande capacità in numero di porte;
- ✓ Riduzione OpEx, ossia la possibilità di avere un numero minore di apparati rispetto a quelli previsti in una tradizionale architettura a tre livelli;
- ✓ Riduzione CapEx, ossia risparmio energetico;
- ✓ STP Free, assenza di Spanning Tree Protocol;
- ✓ L3 ECMP (Equal Cost Multi Path);
- ✓ L2 services (switching) attraverso L3 capability (IPv4 or IPv6);
- ✓ Capacità di funzionalità quali VXLAN, FCoE, VMware integration;
- ✓ LISP (Locator Identifier Separation Protocol) dove previsto un moving di un end-point ed i suoi parametri di rete (addressing) non cambiano.

2.1 SDN Controller & Policies

Il pushing delle configurazioni fa leva su questi items:

- ✓ Policy-based:
  - networks policies sulla base di sistemi applicativi;
  - numero di end-point groups costituiti da servers all'interno di uno stesso segmento di rete (vlans);
  - sistema di comunicazione tra end-points

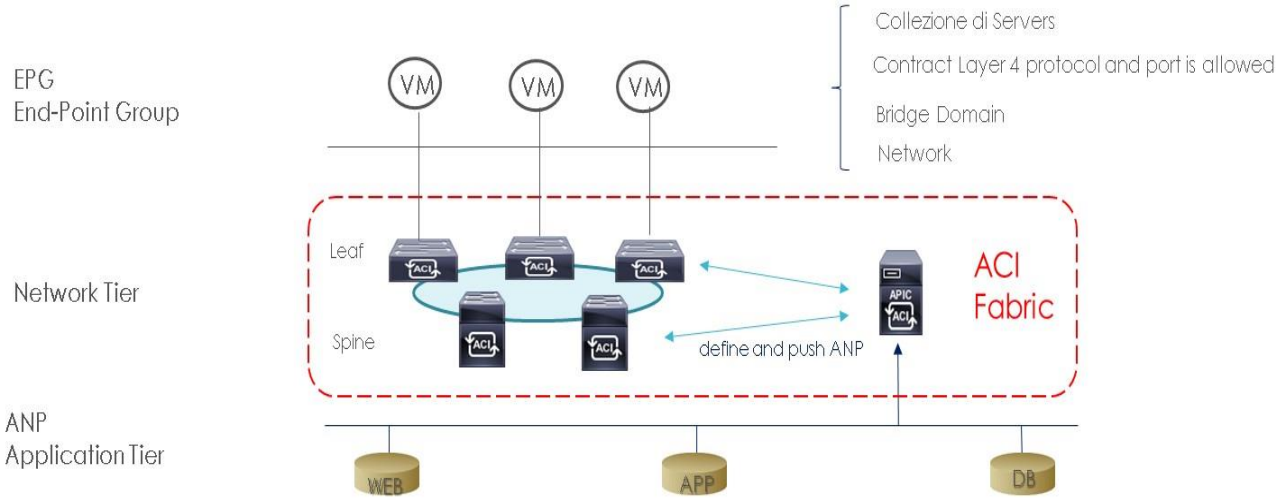

Figura 6: esempio architettura EPG + ANP + Network Tier

✓ Access-policies:

- Vlan-Pool per la definizione di un singolo segmento di rete oppure un pool di segmenti;
- Physical Domain per definire domini di scopo dove è creato il vlans pool;
- Access Entity Profile per definire un modo di raggruppare multipli domini applicabili ad un profilo su base interfaccia;
- Interface Policy e Profile per definire parametri di rete quali LLDP, LACP; etc e contiene policies su base interfaccia/porta
- Switch Profile per applicare il profilo su base interfaccia con la policy associata ad uno o più access node Leaf.

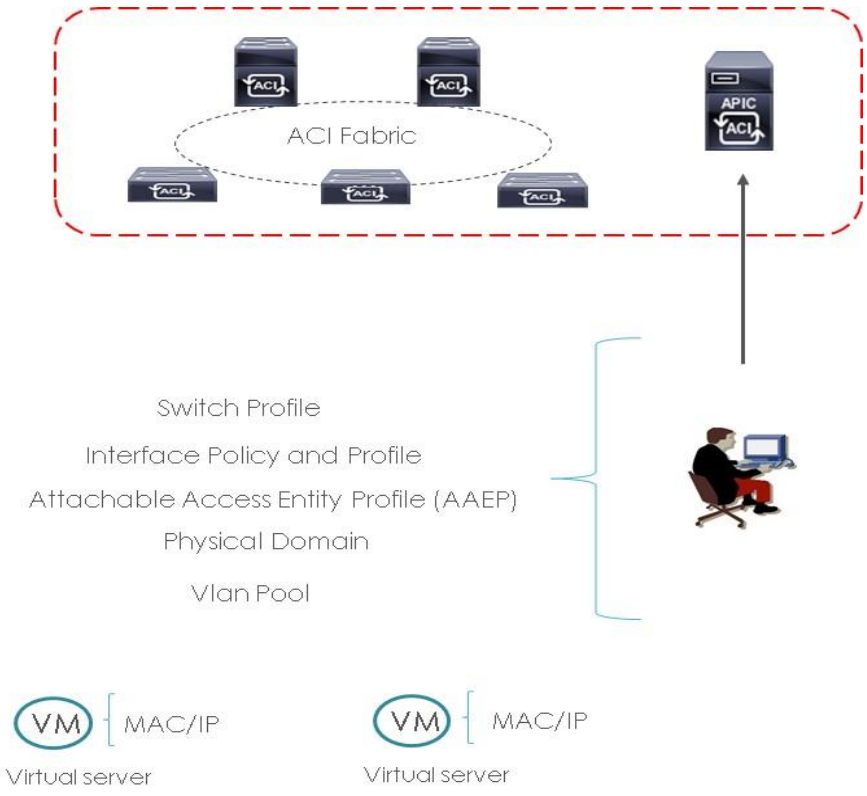

Figura 7: esempio di pushing access policies

✓ Steps di configurazione layer 2:

- BD (Bridge Domain) associato alla VRF instanza (senza L3 IP address);
- Configurazione BD per ottimizzare funzioni di switching con mapping-database oppure via flood-and-learn;
- Razionalizzazione di end-point associati allo stesso bridge domain;
- Policies di comunicazione tra end-points;
- Creazione di access policies switch e port profile assegnando i parametri richiesti ed associarli agli access node Leaf di pertinenza
- Abilitazione di flooding layer 2 unknow unicast packets
- Arp flooding

- L2 extensions

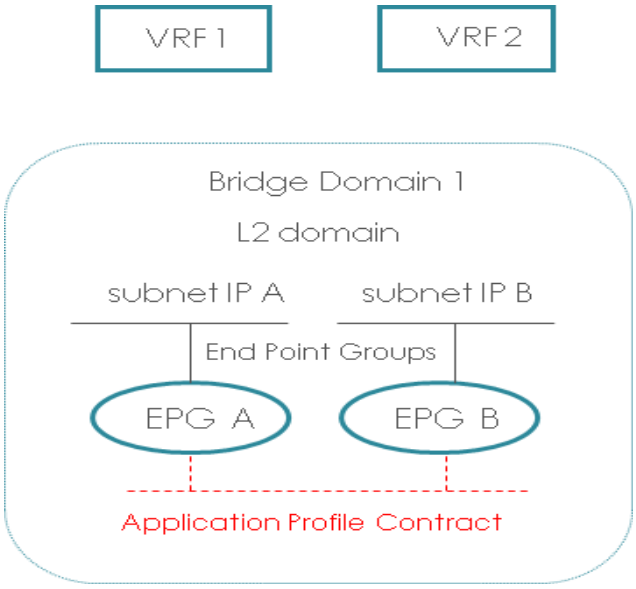

Figura 8: esempio di L2 Bridge Domain EPG

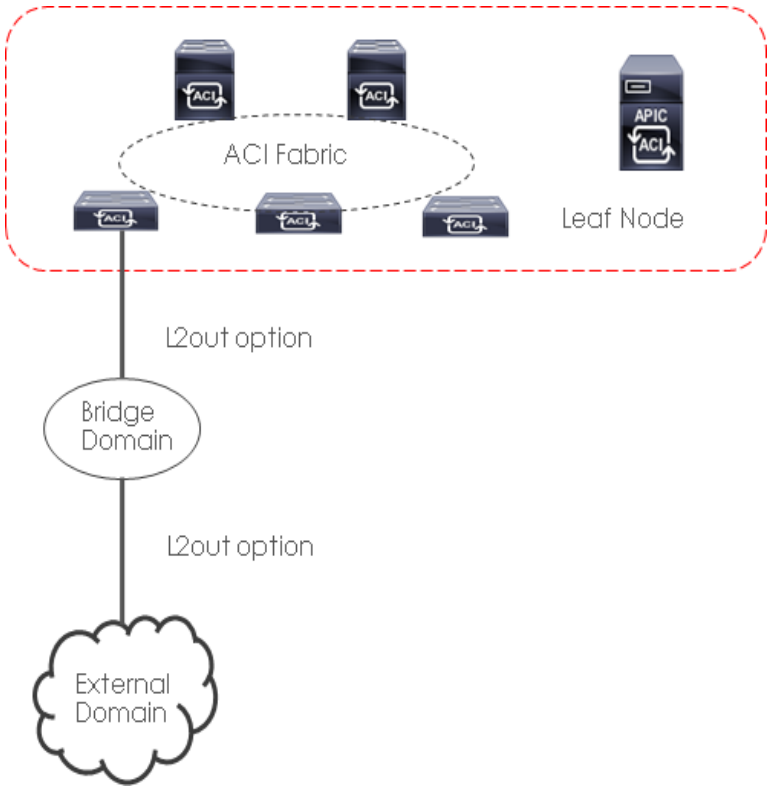

Figura 9: esempio di L2out option Bridge Domain

✓ Steps di configurazione layer 3:

- Layer 3 interface routed, usata quando si connette un determinato external devices per-tenant o VRF
- 802.1q tagging utilizzata per connessioni condivise ad un determinato external devices attraverso tenant o VRF;
- L3 virtual interface usata quando si condivide un layer 2 ed un layer 3 sulla stessa interfaccia;
- Configurazione layer 3 verso external networks domains

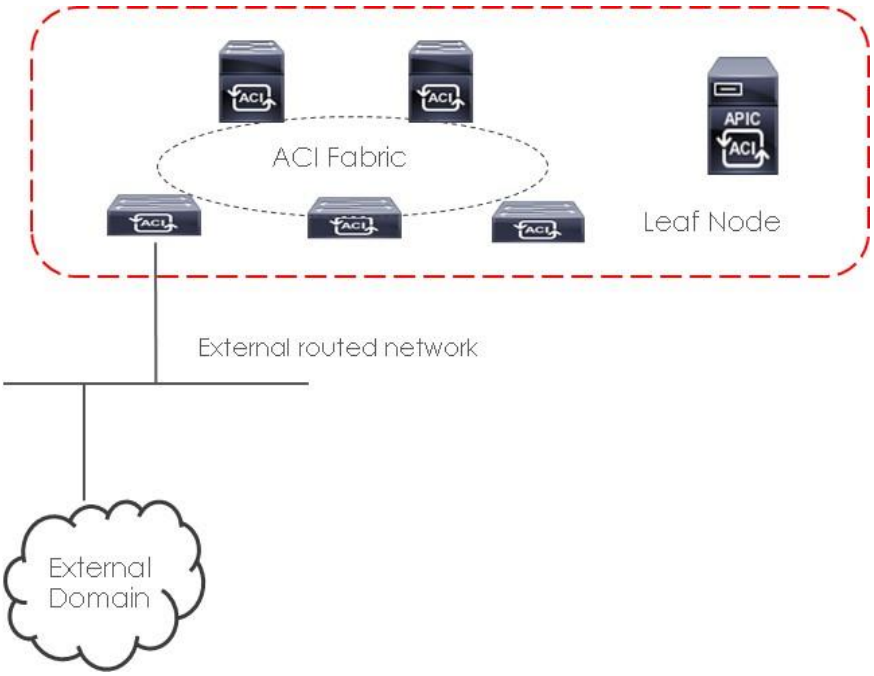

Figura 10: esempio external routed network domain