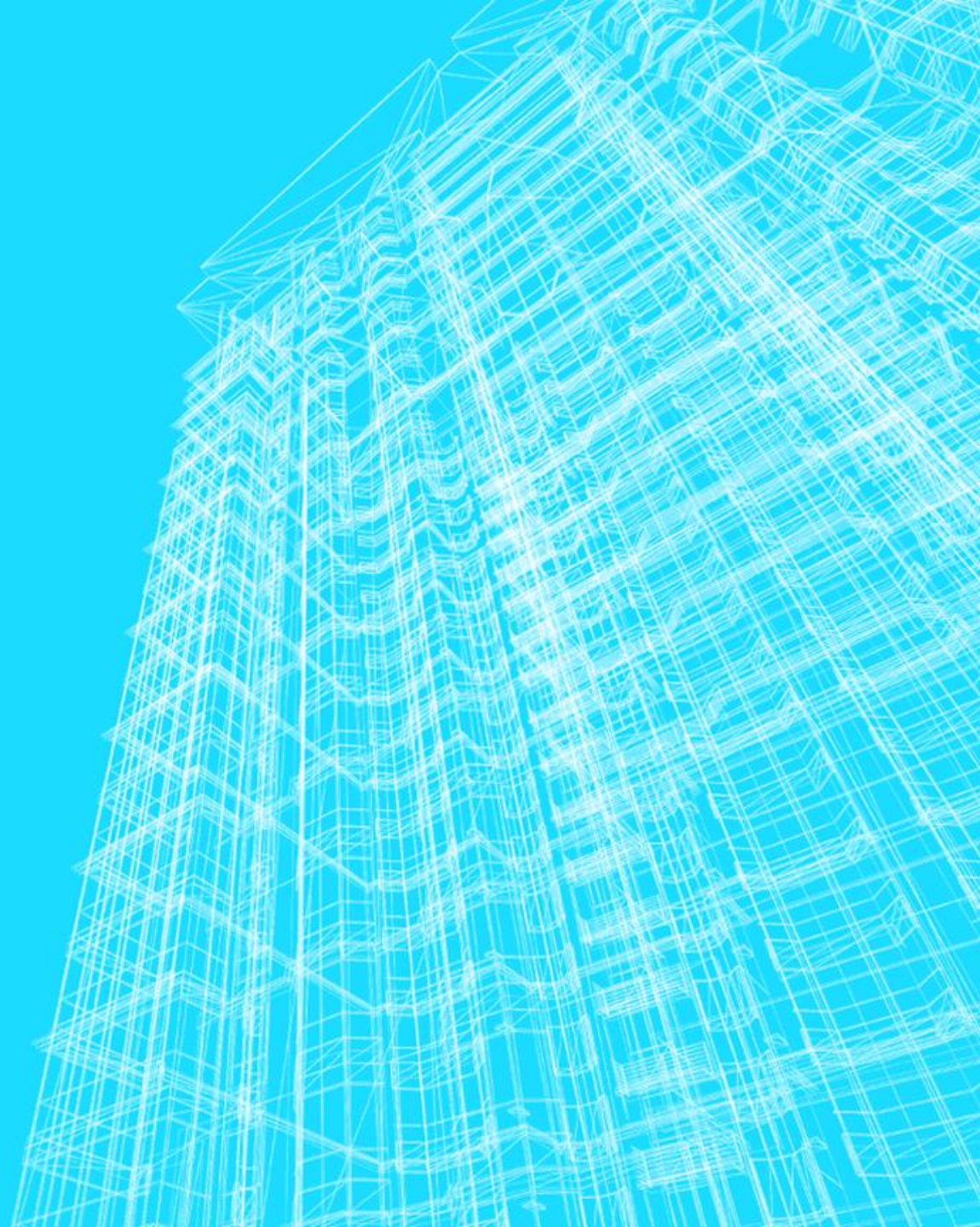
OTV OVERLAY TRANSPORT VIRTUALIZATION

Massimiliano Sbaraglia



OTV CONCEPTS

OTV è una infrastruttura di collegamento inter-datacenters IP-based che provvede a L2-extensions; l'infrastruttura di trasporto tra datacenters può essere DMDW, MPLS, IP routed, WAN, ATM, Frame Relay, etc…

Il requisito principale che deve esserci connettività IP tra i due Data Centers

OTV è una tecnologia multipunto per servizi L2 extensions e indipendenti layer 2 domini tra data centers, preservando fault-isolation, resilienza e load-balancing.

OTV introduce il concetto di Layer 2 MAC routing che abilita il piano di controllo (control-plane) di annunciare la raggiungibilità MAC addresses; con il piano di controllo MAC address learning, OTV non trasmette (flood) unknown unicast traffic ed ARP traffic è trasmesso solo in modo controllato.

OTV non propaga BPDU STP attraverso l'infrastruttura di trasporto overlay

OTV utilizza Nexus Cisco con VDC (Virtual Context Domain) ed è mandatorio avere vlans extended con layer 3 SVI (switched virtual interface) per una data vlan

OTV CONCEPTS AND INTERFACES

OTV Edge Device: performa le funzionalità e le operazioni OTV; riceve le frame ethernet traffic per tutte le vlans soggette ad L2-extensions tra data centers OTV peers e dinamicamente le incapsula dentro IP packets che sono trasmessi via overlay transport infrastructure.

OTV internal interface: sono le interfacce di un edge device che connette il datacenter locale con una configurazione generalmente in trunk trasportando multiple vlans. Non prevedono nessuna configurazione OTV compliant.

OTV join interface: sono le interfacce uplink di un edge device che si affacciano alla rete core overlay IP; questo tipo di interfacce sono point-to-point layer 3 routed, subinterface, port-channel oppure port-channel subinterface (No loopback) ed hanno lo scopo di essere le sorgenti di traffico OTV incapsulato e trasmesso verso l'infrastrutura overlay.

La sua configurazione prevede:

associazione ad una determinata rete IP overlay oppure ad multiple reti overlay;
IGMP client che si unisce ad un determinato gruppo multcast configurato a livello overlay IP interface con lo scopo di trovare altri e remoti OTV Edge Devices;
una volta terminato il discovery e le adiancenze a livello piano di controllo, i peers possono trasmetter e ricevere MAC reachability information e trasmettere e ricevere unicast e multicast traffic

OTV CONCEPTS AND INTERFACES

OTV overlay interface: sono interfacce logiche virtuali dove risiede tutta la configurazione OTV; incapsula le frame layer 2 in IP unicast o multicast packets che sono trasmesse verso altri datacenters. Questo pemette agli edge device di performare un dinamico encapsulations.

OTV site vlan: è una funzionalità utilizzata per scoprire altri Edge Devices in una topologia multi-homed.

La sua configurazione prevede:

OTV edge device elegge un Authorative Edge Devices (AED) per ogni extended vlan;

OTV site vlan necessità di un trunk con la OTV internal interface dell'edge device;

OTV site vlan è unico per location e non deve essere in overlap con nessuna altra vlans all'interno della topologia multi-homed (significa che la vlan scelta come

OTV vlan site è utilizzata solo per questo scopo).

OTV site ID: sappiamo che le adiancenze OTV sono costruite via le join interface attraverso la rete IP overlay; ogni edge device all'interno dello stesso site hanno lo stesso site-id configurato; dalla release NX-OS 5.2.1 una seconda OTV adiancenza è mantenuta con lo scopo di protezione in caso di partizionamento di site-vlan tra edge devices all'interno dello stesso site.

OTV CONCEPTS AED

AED authoritatve edge device: è responsabile della trasmissione di layer 2 traffic incluso unicast, multicast e broadcast; è responsabile di annunciare la raggiungibilità dei mac-addresses verso i datacenters remoti;

La funzionalità site-vlan è utilizzata per la scoperta di edge devices remoti in una topologia multi-homed: in aggiunta al site-vlan, l'edge devices mantiene una seconda OTV adiacenza con gli altri edge devices appartenenti allo stesso datacenter secondo questa configurazione:

site adjacency: OTV edge devices continua ad usare il site-vlan per scoprire altri edge devices all'interno dello stesso datacenter:
overlay adjacency; stabilisce adiacenze attraverso la join interface via IP overlay network

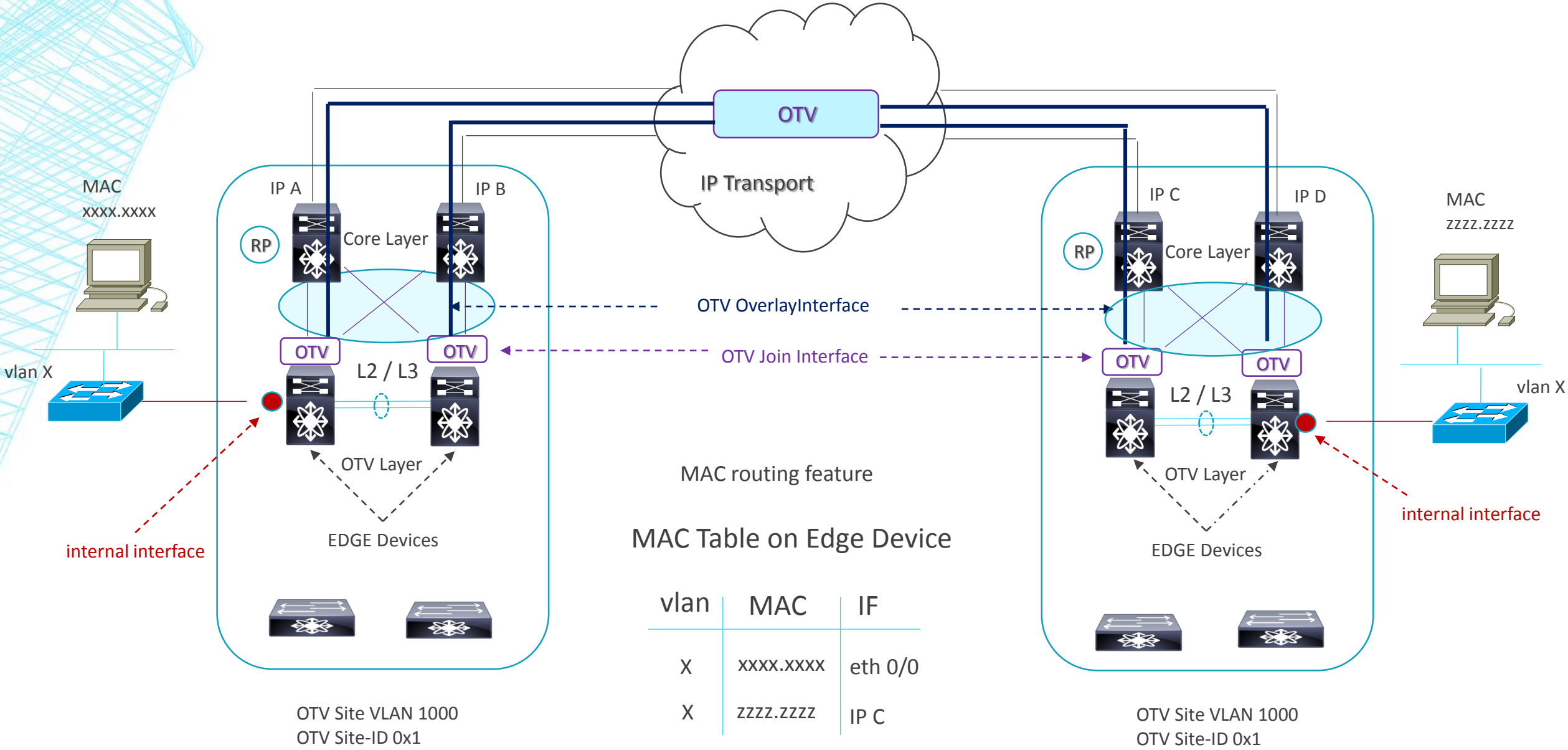
Note:

Il site-id per la overlay adjancency è mandatoria; valore range = 0x1 to 0xffffffff (formato esadecimale o mac-address

Se il site-id non è configurato, la overlay interface non si attiva

Tutti gli edge devices devono essere configurato con lo stesso site-id; il site-id è annunciato in ISIS hello packets e la combinazione di un site-id con ISIS system-ID è usato per identificare un neighbor edge device all'interno dello stesso datacenters

OTV DESIGN



MAC routing feature

MAC Table on Edge Device

vlan	MAC	IF
X	xxxx.xxxx	eth 0/0
X	zzzz.zzzz	IP C

OTV CONFIGURATION EXAMPLE INTERFACES

OTV internal interface:

```
interface port-channel 200
switchport
switchport mode trunk
switchport trunk native vlan 100
switchport trunk allowed vlan 10,12,14,20-30,40-50,70-99,1000
spanning-tree port type normal
mac packet-classify
!
interface  ethernet 3/23
switchport
switchport mode trunk
switchport trunk native vlan 100
switchport trunk allowed vlan 10,12,14,20-30,40-50,70-99,1000
spanning-tree port type normal
channel-group 200 mode active
no shut
!
interface  ethernet 7/23
switchport
switchport mode trunk
switchport trunk native vlan 100
switchport trunk allowed vlan 10,12,14,20-30,40-50,70-99,1000
spanning-tree port type normal
channel-group 200 mode active
no shut
!
```

OTV CONFIGURATION EXAMPLE INTERFACES

OTV join interface:

```
interface port-channel 300
mtu 1600
ip address 172.16.1.1/30
ip ospf network point-to-point
ip router ospf 10 area 0.0.0.0
ip igmp version 3
no shut
!
interface  ethernet 4/16
mtu 1600
channel-group 300 mode active
no shut
!
interface  ethernet 5/18
mtu 1600
channel-group 300 mode active
no shut
!
```

OTV CONFIGURATION EXAMPLE INTERFACES

OTV overlay interface:

```
interface overlay 1
otv join-interface port-channel 300
otv control-group 239.1.1.1
otv data-group 232.0.0.0/24
otv extend-vlan 10,12,14,20-30,40-50,70-99
no shut
!
```

OTV CONFIGURATION EXAMPLE SITE-VLAN AND SITE-ID

OTV site vlan:

```
vlan 1000
name otv-site-vlan
otv site-vlan 1000
!
!
otv site-identifier 0x1
```

OTV MULTICAST ENABLED TRANSPORT OVERLAY

OTV Edge Devices sono configurati per unirsi ad uno specifico ASM (Any Source Multicast) group; in questo modo ogni OTV edge devices diventa receiver e source multicast traffic;

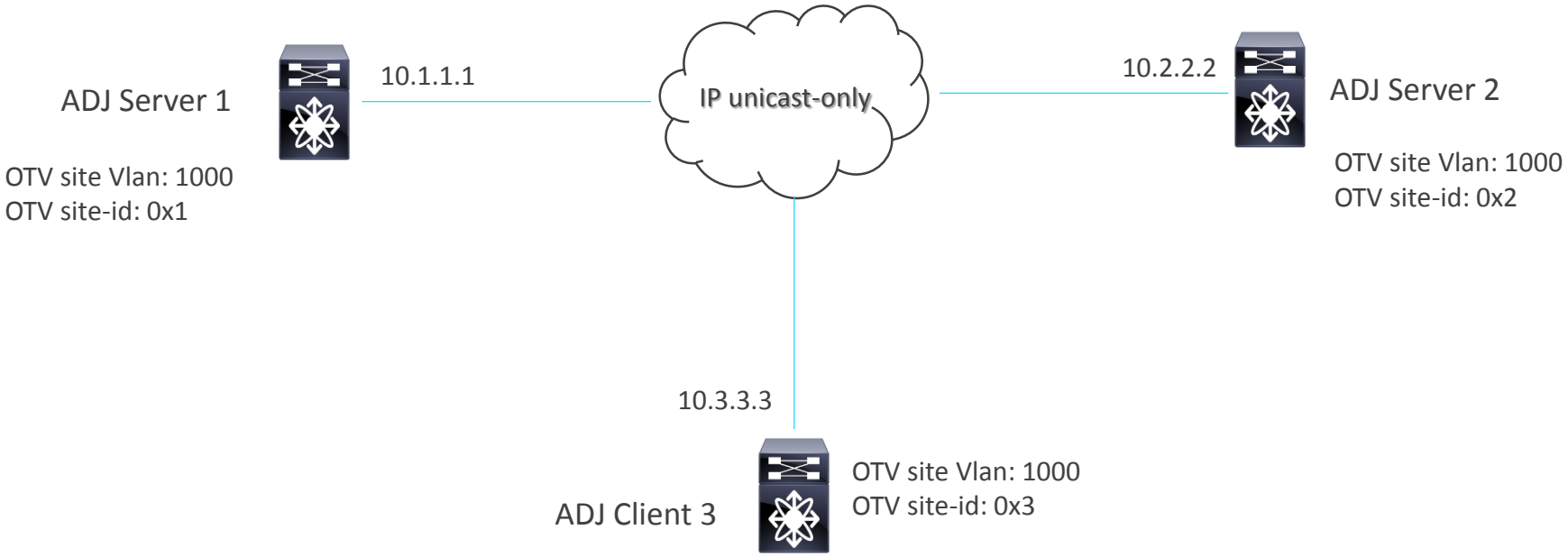
Le interfacce in upstream layer 3 debbono essere configurate in PIM sparse-mode ed ogni device deve specificare il SSM group da usare;

Un RP (Rendezvou Point) router deve essere definito (due RP per ridondanza, dove quest'ultima può essere ottenuta usando Anycast RP);

OTV UNICAST ENABLED TRANSPORT OVERLAY

Nella situazione dove non è possibile avere un Multicast Overlay Transport, è possibile utilizzare un trasporto di tipo unicast-only; la differenza sta che ogni Edge Device deve creare multiple copie di ogni control-plane packet relativo ad ogni edge devices remoto facente parte dello stesso logical overlay interface.

Un nuovo concetto di adiancenza è introdotto: **OTV adjacency server**; ogni OTV device cerca di unirsi ad una specifica logical overlay interface avendo il bisogno di registro verso il server inviando hello message; questi messaggi servono al server per costruire una lista di tutti gli OTV devices che dovranno far parte dello stesso dominio overlay (unicast-replication-list).



OTV UNICAST ENABLED TRANSPORT OVERLAY CONFIGURATION

ADJ Server 1

```
otv side-identifier 0x1
otv site-vlan 1000
interface overlay 1
 otv join-interface port-channel 300
 otv adjacency-server unicast-only
 otv extend-vlan 10,12,14,20-30,40-50,70-99
```

ADJ Server 2

```
otv side-identifier 0x2
otv site-vlan 1000
interface overlay 1
 otv join-interface port-channel 300
 otv adjacency-server unicast-only
 otv use-adjacency-server 10.1.1.1 unicast-only
 otv extend-vlan 10,12,14,20-30,40-50,70-99
```

OTV UNICAST ENABLED TRANSPORT OVERLAY CONFIGURATION

ADJ Client 1

```
otv side-identifier 0x3
otv site-vlan 1000
interface overlay 1
 otv join-interface port-channel 300
 otv use-adjacency-server 10.1.1.1  20.2.2.2 unicast-only
 otv extend-vlan 10,12,14,20-30,40-50,70-99
!
!
show otv overlay 1
```

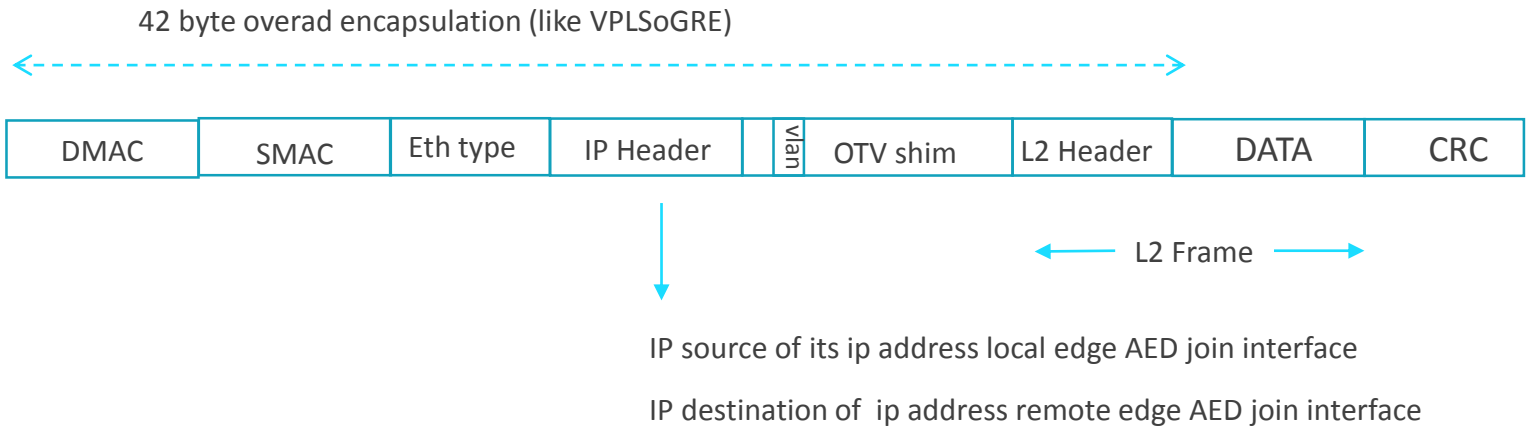
OTV DATA PLANE

OTV necessità di avere una completa tabella di adiancenze con i propri Edge Peers OTV per scambiare il MAC-address reachability information.

Ci sono due differenti scenari:

1. Local layer 2 switching for devices che comunicano con gli edge devices OTV all'interno della stessa VLAN ed all'interno dello stesso datacenters (intrasite-datacenter)

2. Remote layer 2 switching tra devices ubicati tra differenti datacenters sempre all'interno della stessa VLAN (intersite-datacenters); questo prevede un tipo di encapsulation a livello data-plane con un overhead di 42 byte.



OTV AND QOS

OTV distingue la QoS in riferimento al control-plane e data-plane traffic

1. Control-Plane: le frames sono sempre originate da un OTV Edge Device e staticamente marcati con CoS = 6 (DSCP = 48)
2. Data-Plane: le frame layer 2 ricevute da un OTV Edge Devices ed incapsulate sono già state marcate da una prospettiva CoS e DSCP

In fase di encapsulation, il QoS marking agisce:

Con 802.1p tagged, L2 CoS è copiato verso l'outer DSCP (l'inner DSCP è riservato per l'infrastruttura overlay)

Con 802.1p untagged l'outer DSCP è zero (TOS = 0x00) e l'inner DSCP è riservato attraverso il tunnel

OTV BENEFITS

Spanning Tree Protocol Isolation: BPDU non sono trasmessi attraverso la rete overlay

Soppressione di traffico unicastdi tipo unknown

Broadcast policy control: con Multicast enabled transport le broadcast frame sono trasmesse a tutti gli OTV edge devices all'interno dello stesso ASM group

ARP optimization

FHRP (First-Hop Routing Protocol): Virtual default-gateway IP address (vMAC addresses); questa funzionalità comunque prevede di flitrare FHRP message transitanti per la overlay network che abilitano la coesistenza di un default gateway. Il filtering permette quindi di droppare i messaggi FHRP a livello AED e considerare il default gateway VIP in modalità active active ad ogni sito datacenter (questo elimina l'hair pinning di traffico destinato al default gateway tra siti)

OTV CONFIGURATION EXAMPLE DISABLING ARP-ND CACHE

OTV edge device:

```
interface overlay 1
 otv join-interace port-channel 30
 otv conrol-group 239.1.1.1
 otv data-group 232.0.0.0/24
 no otv suppress-arp-nd
 otv extend-vlan 10,12,14,20-30,40-50,70-99
!
!
show otv arp-nd-cache
```

OTV ARP/ND L3→L2 address mapping cache

OTV CONFIGURATION EXAMPLE DISABLING HSRP MESSAGE BETWEEN EDGE DEVICES AND PERFORM HSRP LOCALIZATION (1)

OTV edge device:

```
ip access-list ALL-IP
 10 permit ip any any
!
mac access-list ALL-MAC
 10 permit any any
!
ip access-list HSRP-IP
 10 permit udp any 224.0.0.2/32 eq 1985
 20 permit udp any 224.0.0.102/32 eq 1985
!
!
vlan access-map HSRP-LOCAL 10
 match ip address HSRP-IP
 action drop

vlan access-map HSRP-LOCAL 20
 match ip address ALL-IP
 action forward

vlan filter HSRP-LOCAL vlan-list a,b
```

OTV CONFIGURATION EXAMPLE FILTER LEARNING HSRP VMAC ACROSS OTV (2)

OTV edge device (block the HSRP vMAC from being advertised to other routers)

```
mac access-list HSRP-VMAC
 10 permit 0000.0c07.ac00  0000.0000.00ff  any
 20 permit 0000.0c9f.f0000  0000.0000.00ff any
!
vlan access-map HSRP-LOCAL 10
 match mac address HSRP-VMAC
 match ip address HSRP-IP
 action drop
!
vlan access-map HSRP-LOCAL 20
match mac address  ALL-MAC
match ip address ALL-IP
 action forward
!
vlan filter HSRP-LOCAL vlan-list a,b
!
mac-list HSRP-VMAC-DENY seq 5   deny 0000.0c07.ac00  ffff.ffff.ff00
mac-list HSRP-VMAC-DENY seq 10 deny 0000.0c9f.f000  0000.0000.0fff
mac-list HSRP-VMAC-DENY seq 15 permit 0000.0000.0000  0000.0000.0000
!
rout-map stop-HSRP permit 10
match mac-list HSRP-VMAC-DENY
!
otv-isis default
vpn overlay 1
redistribution filter route-map stop-HSRP
```

OTV CONFIGURATION EXAMPLE PREVENT DUPLICATE HSRP GRATUITOUS ARP FROM HSRP VIP

OTV edge device:

```
arp access-list HSRP-VMAC-ARP
 10 deny ip any mac 0000.0c07.ac00  ffff.ffff.ff00
 20 deny ip any mac 0000.0c9f.f000    ffff.ffff.ff00
 30 permit ip any mac any
!
feauture dhcp
ip arp inspection filter HSRP-VMAC-ARP 10,12,500,501,800,801
```