Nexus 9000: Configurar y verificar VXLAN Xconnect

Contenido

Introducción

El documento describe una referencia rápida sobre cómo configurar y verificar VXLAN Xconnect en switches Nexus 9000.

Prerequisites

Requirements

Cisco recomienda que tenga conocimiento de VXLAN EVPN.

Componentes Utilizados

La información que contiene este documento se basa en las siguientes versiones de software y hardware.

- N9K-C93180YC-EX
- NXOS 9.2(1)

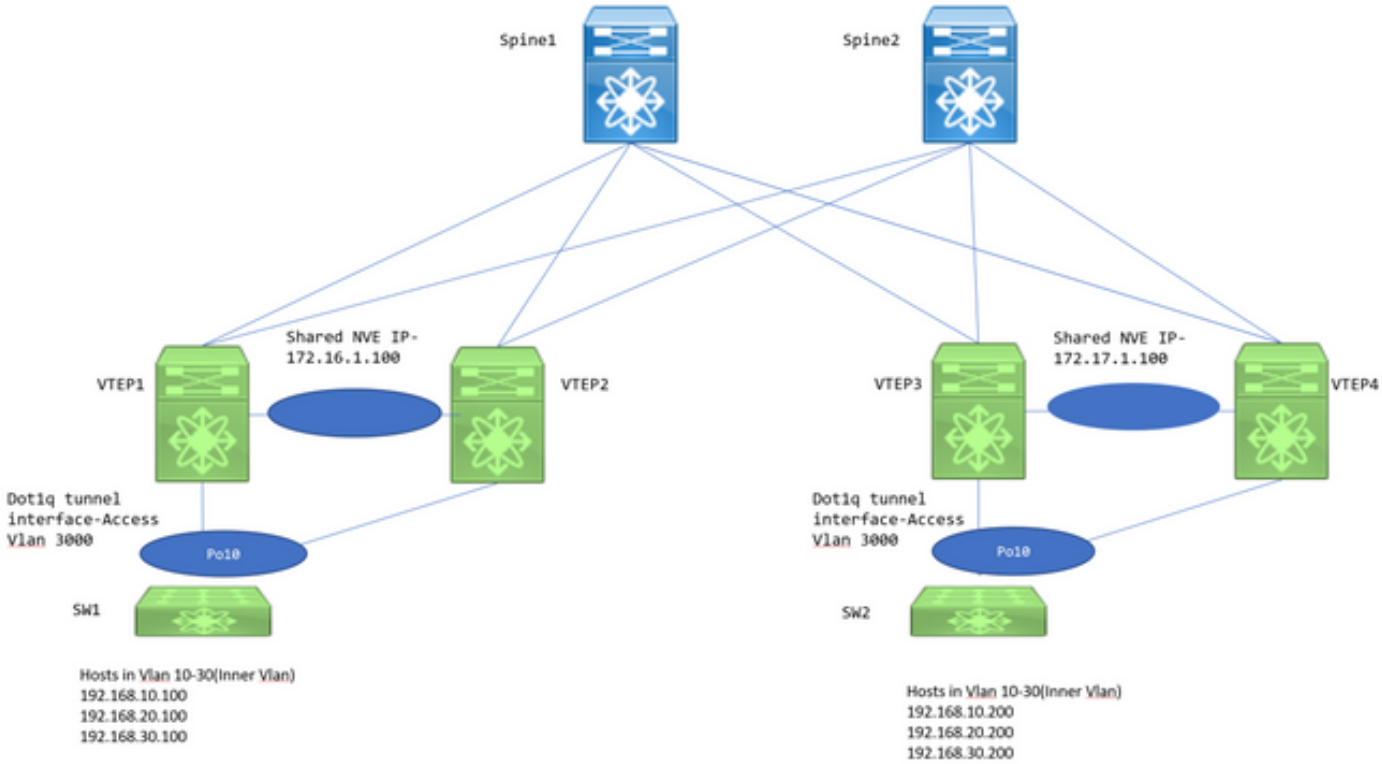
The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

Overview

VXLAN Xconnect es un mecanismo para un túnel punto a punto para paquetes de datos y control de una hoja a otra. Las etiquetas Dot1q internas se conservan y VXLAN se encapsula dentro del VNID externo especificado como VNID Xconnect. Las tramas de control de capa 2, como el

protocolo de descubrimiento de la capa de enlace (LLDP), el protocolo de detección de Cisco (CDP) y el protocolo de árbol de extensión (STP), se encapsulan en VXLAN y se envían a otros extremos del túnel.

Topología



VTEP1, VTEP2, VTEP3 y VTEP4 son dos pares VTEP de vPC configurados de tal manera que se preserven las etiquetas dot1q internas de los switches descendentes y, cuando VXLAN se encapsula, utilice VXLAN VNID de ID de VLAN externa para enviar al VTEP remoto. Todos los VTEP son N9K-C93180YC-EX.

Los switches descendentes son Nexus 3ks que se configuran con la interfaz virtual del switch (SVI) en las VLAN respectivas para imitar los hosts.

Configurar

1. La VLAN exterior utilizada en esta topología Xconnect es 3000. Éste sería el que tendría la configuración VNID y Xconnect.

```
VTEP1# sh run vlan 3000

vlan 3000
  vn-segment 1003000
  xconnect
```

2. La función NGOAM debe estar habilitada y necesita esta configuración.

```
VTEP1# sh run ngoam
```

```
feature ngoam

ngoam install acl
ngoam xconnect hb-interval 5000
```

3. Configuración del túnel Dot1q hacia el switch descendente.

```
VTEP1# sh run int po10

interface port-channel10
  switchport
  switchport mode dot1q-tunnel
  switchport access vlan 3000
  speed 40000
  no negotiate auto
  vpc 10
```

Las configuraciones de vPC sólo se requieren cuando los VTEP se implementan como vPC. De lo contrario, omita las configuraciones de vPC mencionadas en este documento. VXLAN Xconnect también se puede configurar en un VTEP independiente.

4. El grupo de multidifusión debe definirse en la interfaz NVE para encargarse del reenvío. Tenga en cuenta para habilitar **ip pim sparse-mode** en los uplinks relevantes y definir PIM RP, así como para que el ruteo multicast y los mensajes PIM se intercambien apropiadamente. Normalmente, el RP PIM se define en la Capa de la columna.

```
VTEP1# sh run int nve1

  no shutdown
  host-reachability protocol bgp
  source-interface loopback1
  member vni 1003000 mcast-group 239.30.30.30
```

5. La VLAN Infra debe ser especificada y permitida como la VLAN nativa dentro del link de peer. Este paso es necesario para vPC VTEP.

```
VTEP1# sh run span|infra
no spanning-tree vlan 3000
system nve infra-vlans 999

VTEP1# sh run int po1

interface port-channel1
  switchport
  switchport mode trunk
  switchport trunk native vlan 999
  spanning-tree port type network
  vpc peer-link
```

6. Configuración de BGP/EVPN: Se necesitan vecindarios L2VPN EVPN entre la hoja/columna para intercambiar las rutas de tipo 3 necesarias para establecer VXLAN Xconnect.

- Aquí, las direcciones IP 192.168.100.1 y 192.168.100.2 son las columnas en la topología. Por lo general, los vecinos L2VPN EVPN se forman en las columnas. Las columnas configuran todos los switches de hoja como clientes de reflector de ruta en un escenario iBGP.

- Se recomienda utilizar loopbacks separados para propósitos BGP/OSPF y NVE.

```
feature bgp

router bgp 65000
  router-id 192.168.100.3
  neighbor 192.168.100.1
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community
      send-community extended
  neighbor 192.168.100.2
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
send-community
send-community extended evpn vni 1003000 l2 rd auto route-target import auto route-target export
auto
```

> **Nota:** El STP debe ser inhabilitado dentro de la VLAN Xconnect. El aprendizaje de MAC no ocurrirá dentro de la VLAN Xconnect, lo que significa esencialmente que no hay actualizaciones de VPN bgp l2vpn de tipo 2 para las direcciones MAC. Debido a esto, el tráfico de una vtep se encapsulará con la dirección IP de destino externa establecida en el grupo de multidifusión(239.30.30.30) definido para la VLAN Xconnect.

Verificación

Use esta sección para confirmar que su configuración funciona correctamente.

1. vecindad BGP.

```
 VTEP1# sh bgp l2vpn evpn sum
BGP summary information for VRF default, address family L2VPN EVPN
BGP router identifier 192.168.100.3, local AS number 65000
BGP table version is 14, L2VPN EVPN config peers 2, capable peers 1
4 network entries and 5 paths using 756 bytes of memory
BGP attribute entries [3/492], BGP AS path entries [0/0]
BGP community entries [0/0], BGP clusterlist entries [2/8]

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
192.168.100.1   4 65000      92      90       14    0    0 01:21:41 2
```

2. Prefijos de tipo 3 de recepción.

```
VTEP1# sh bgp l2vpn evpn
BGP routing table information for VRF default, address family L2VPN EVPN
BGP table version is 14, Local Router ID is 192.168.100.3
Status: s-suppressed, x-deleted, S-stale, d-dampened, h-history, *-valid, >-best
Path type: i-internal, e-external, c-confed, l-local, a-aggregate, r-redist, I-injected
Origin codes: i - IGP, e - EGP, ? - incomplete, | - multipath, & - backup

   Network            Next Hop            Metric     LocPrf     Weight Path
Route Distinguisher: 192.168.100.3:35767    (L2VNI 1003000)
*>l[3]:[0]:[32]:[172.16.1.100]/88
                      172.16.1.100                       100      32768 i
```

```
* i[3]:[0]:[32]:[172.17.1.100]/88<<< bgp type 3
                      172.17.1.100                       100          0 i
*>i                   172.17.1.100                       100          0 i

Route Distinguisher: 192.168.100.5:35767
*>i[3]:[0]:[32]:[172.17.1.100]/88
                      172.17.1.100                       100          0 i

Route Distinguisher: 192.168.100.6:35767
*>i[3]:[0]:[32]:[172.17.1.100]/88
                      172.17.1.100                       100          0 i
```

3. NVE Peering.

```
 VTEP1# sh nve peer
Interface Peer-IP          State LearnType Uptime   Router-Mac
--------- ---------------  ----- --------- -------- -----------------
nve1      172.17.1.100     Up    CP        00:58:06 n/a


VTEP1# show nve vni
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       SU - Suppress Unknown Unicast

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ------------------ -----
nve1      1003000  239.30.30.30      Up    CP   L2 [3000]                  Xconn  <<<
```

4. Comprobaciones de NGOAM.

```
VTEP1# show ngoam xconnect sess all

States: LD = Local interface down, RD = Remote interface Down
          HB = Heartbeat lost, DB = Database/Routes not present
          * - Showing Vpc-peer interface info
Vlan          Peer-ip/vni      XC-State       Local-if/State    Rmt-if/State
================================================================================
3000  172.17.1.100 / 1003000      Active          Po10 / UP          Po10 / UP


VTEP1# show ngoam xconnect sess 3000
Vlan ID: 3000
Peer IP: 172.17.1.100  VNI : 1003000
State: Active                                    <<< State should be active
Last state update: 12/10/2018 17:13:45.337
Local interface: Po10  State:  UP
Local vpc interface Po10  State:  UP
Remote interface: Po10  State:  UP
Remote vpc interface: Po10  State:  UP
```

Una vez que la sesión de NGOAM está activa, los N3k se verían en CDP. Las BPDU STP también se tunelizan para que los switches también acuerden la ubicación del puente raíz.

5. Verificaciones en los switches de flujo descendente.

```
SW1(config)# sh span vl 10

VLAN0010
  Spanning tree enabled protocol rstp
  Root ID    Priority    32778
```

```
             Address     7079.b348.6cb7
             This bridge is the root
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    32778  (priority 32768 sys-id-ext 10)
             Address     7079.b348.6cb7
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Po10             Desg FWD 1         128.4105 P2p

SW2(config)# sh span vl 10

VLAN0010
  Spanning tree enabled protocol rstp
  Root ID    Priority    32778
             Address     7079.b348.6cb7
             Cost        1
             Port        4105 (port-channel10)
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    32778  (priority 32768 sys-id-ext 10)
             Address     707d.b964.9441
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Po10             Root FWD 1         128.4105 P2p

SW1(config)# show ip int b
IP Interface Status for VRF "default"(1)
Interface            IP Address      Interface Status
Vlan10               192.168.10.100  protocol-up/link-up/admin-up
Vlan20               192.168.20.100  protocol-up/link-up/admin-up
Vlan30               192.168.30.100  protocol-up/link-up/admin-up

SW2(config)# show ip int b
IP Interface Status for VRF "default"(1)
Interface            IP Address      Interface Status
Vlan10               192.168.10.200  protocol-up/link-up/admin-up
Vlan20               192.168.20.200  protocol-up/link-up/admin-up
Vlan30               192.168.30.200  protocol-up/link-up/admin-up

SW1(config)# ping 192.168.10.200
PING 192.168.10.200 (192.168.10.200): 56 data bytes
64 bytes from 192.168.10.200: icmp_seq=0 ttl=254 time=0.826 ms
64 bytes from 192.168.10.200: icmp_seq=1 ttl=254 time=0.531 ms
64 bytes from 192.168.10.200: icmp_seq=2 ttl=254 time=0.54 ms
64 bytes from 192.168.10.200: icmp_seq=3 ttl=254 time=0.522 ms
64 bytes from 192.168.10.200: icmp_seq=4 ttl=254 time=0.571 ms
```

Troubleshoot

Actualmente, no hay información específica de troubleshooting disponible para esta configuración.

Advertencias

1. Las interfaces de túnel dot1q se atascarán en el **estado de error disabled** en una configuración Xconnect VXLAN si las configuraciones dentro de los switches vPC no son consistentes. A continuación se muestran algunos de los casos en los que la interfaz se encontrará en estado de error inhabilitada;

- Si la VLAN al segmento VN no está definida en ambos switches vPC.
- Si el grupo NVE a multicast no está definido en ambos switches vPC.
- Si no se reciben los latidos NGOAM (utilice ethanalyzer con filter**=cfm** para capturar los paquetes de latidos de NGOAM).
- Incluso si la interfaz de túnel dot1q está conectada de forma huérfana en una configuración vPC, todavía es necesario configurar el grupo multicast bajo la interfaz NVE para el segmento VN que forma parte de Xconnect en ambos switches.
- Los latidos de NGOAM son procesados/enviados por el switch principal vPC. Los mensajes de latidos que llegan a vPC secundarios se sincronizarán con el principal

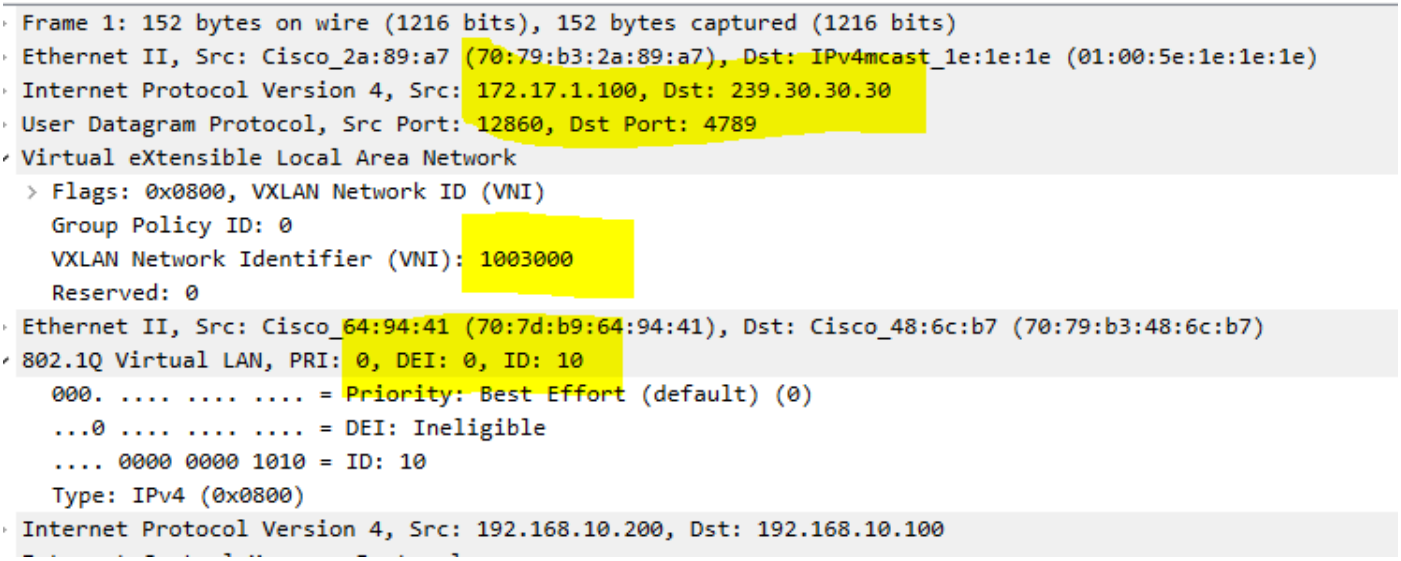
2. Cuando Xconnect se configura en una VLAN, el tráfico de un sitio a otro se encapsula con la dirección de destino externa=dirección multicast definida en la interfaz NVE para ese segmento vn en particular. Se recomienda utilizar un grupo multicast único para las VLAN Xconnect. La multidifusión en el núcleo/columna debe ser funcional.

3. El tráfico de multidifusión podría llegar a ambos equipos vPC en el lado remoto de Xconnect; Sin embargo, el ganador de Decap (el cuadro que puede desencapsular la BUM) será sólo un switch en un par vPC. Esto se puede verificar con el comando **show forwarding multicast route group <Group address> source <SRC IP>**. Si el indicador que se muestra aquí es un **v** en minúscula, significa que el cuadro es un perdedor de la captura y si es un **V en** mayúscula**, el cuadro es el ganador de la decapitación, por lo que puede desencapsular el tráfico multicast y reenviarlo hacia abajo.**

4. En las plataformas basadas en 93180YC, cuando el host está huérfano conectado a 9k1 y si no hay OIL para S, G en 9k1, se envía una copia del paquete multicast al par vPC usando una encapsulación especial de IP de origen**>** 127.0.0.1 y IP de destino**->** IP de NVE compartida y si el 9k2 tiene OIL para S, Entrada G, el reenvío de tráfico será atendido por el 9k2 a los sitios remotos.

Captura de paquete

Esta es una captura de pantalla de una captura de paquetes tomada en el switch de columna:



- Encabezado dot1q interno=10 se conserva
- El VNI utilizado es 1003000 (que es el VNID de la VLAN externa)
- La dirección IP de destino sería el grupo de multidifusión definido en la nueva interfaz