Nexus 9000: Configurar e verificar o VXLAN Xconnect

Contents

Introduction

O documento descreve uma referência rápida sobre como configurar e verificar o VXLAN Xconnect em Nexus 9000 Switches.

Prerequisites

Requirements

A Cisco recomenda que você tenha conhecimento do VXLAN EVPN.

Componentes Utilizados

As informações neste documento são baseadas nestas versões de software e hardware:

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

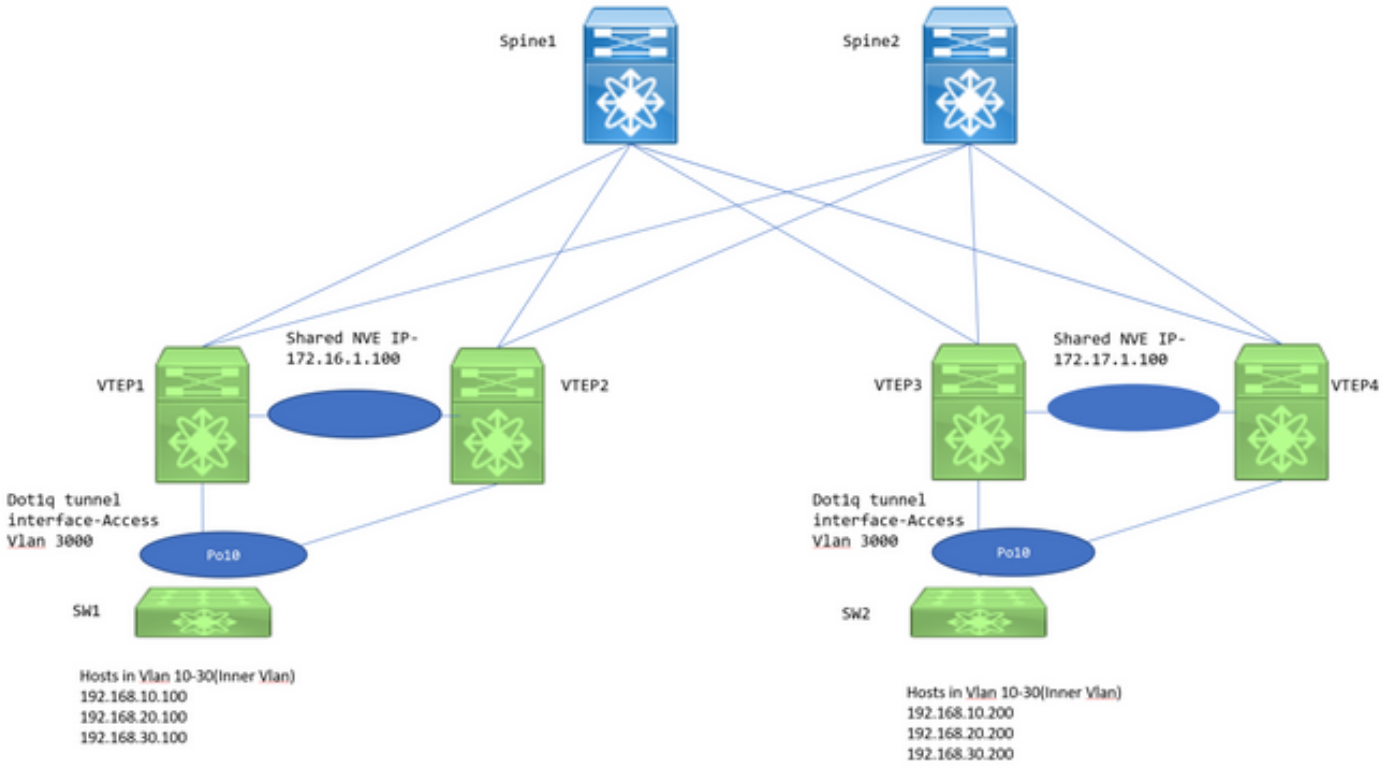
The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

Overview

O VXLAN Xconnect é um mecanismo para um túnel ponto a ponto para dados e pacotes de controle de uma folha para outra. As marcas Dot1q internas são preservadas e a VXLAN encapsulada no VNID externo especificado como o VNID do Xconnect. Os quadros de controle da camada 2, como o Link Layer Discovery Protocol (LLDP), Cisco Discovery Protocol (CDP),

Spanning Tree Protocol (STP), são encapsulados em VXLAN e enviados para outras extremidades do túnel.

Topologia



VTEP1, VTEP2, VTEP3 e VTEP4 são dois pares VTEP vPC configurados de tal forma que as marcas internas dot1q dos switches downstream sejam preservadas e quando VXLAN encapsulada, use VXLAN VNID do ID da VLAN externa para enviar para o VTEP remoto. Todos os VTEPs são N9K-C93180YC-EX.

Os switches downstream são Nexus 3ks configurados com SVIs (Switch Virtual Interface, interface virtual do switch) nas respectivas VLANs para imitar os hosts.

Configurar

1. A VLAN externa usada nesta topologia Xconnect é 3000. Essa seria a configuração com VNID e Xconnect.

```
VTEP1# sh run vlan 3000

vlan 3000
  vn-segment 1003000
  xconnect
```

2. O recurso NGOAM deve ser ativado e precisa dessa configuração.

```
VTEP1# sh run ngoam
```

```
feature ngoam

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

3. Configuração de túnel Dot1q em direção ao switch downstream.

```
VTEP1# sh run int po10

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

As configurações vPC são necessárias somente quando VTEPs são implantados como vPC. Caso contrário, ignore as configurações vPC mencionadas neste documento. O VXLAN Xconnect também é configurável em um VTEP autônomo.

4. O grupo multicast precisa ser definido na interface NVE para cuidar do encaminhamento. Observação para ativar o **modo esparso do pim ip** em uplinks relevantes e definir o PIM RP, assim como para que o roteamento multicast e as mensagens PIM sejam trocadas apropriadamente. Geralmente, o PIM RP é definido na camada spine.

```
VTEP1# sh run int nve1

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

5. A VLAN de infravermelho precisa ser especificada e permitida como a VLAN nativa no link de peer. Essa etapa é necessária para VTEPs vPC.

```
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. Configuração BGP/EVPN: Os vizinhos L2VPN EVPN são necessários entre leaf/spine para trocar as rotas Tipo 3 necessárias para estabelecer o VXLAN Xconnect.

- Aqui, os endereços IP - 192.168.100.1 e 192.168.100.2 são os Spines na topologia. Normalmente, os vizinhos L2VPN EVPN são formados para os Spines. Os spines configuram todos os switches Leaf como clientes refletores de rota em um cenário iBGP.
- Recomenda-se usar loopbacks separados para fins de BGP/OSPF e 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
```

> **Note**: O STP deve ser desabilitado na VLAN Xconnect. O aprendizado MAC não acontecerá dentro da VLAN Xconnect, o que significa que não há atualizações de vpn de bgp l2vpn tipo 2 para endereços MAC Devido a isso, o tráfego de um vtep será encapsulado com o endereço IP de destino externo definido para o grupo Mcast (239.30.30.30) definido para a VLAN Xconnect.

Verificar

Use esta seção para confirmar se a sua configuração funciona corretamente.

1. BGP vizinho.

```
 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. Receber prefixos tipo 3.

```
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. Verificações 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
```

Quando a sessão NGOAM estiver ativa, os N3k's se veriam no CDP. As BPDUs do STP também são encapsuladas para que os switches concordem com o posicionamento da bridge raiz também.

5. Verificações nos switches downstream.

```
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

Atualmente, não existem informações disponíveis específicas sobre Troubleshooting para esta configuração.

Caveats

1. As interfaces de túnel dot1q serão travadas no **estado desativado por erro** em uma configuração Xconnect VXLAN se as configurações nos switches vPC não forem consistentes. Abaixo estão alguns dos casos em que a interface será desabilitada por erro.

- Se a VLAN para o segmento VN não estiver definida em ambos os switches vPC.
- Se o grupo NVE para multicast não estiver definido em ambos os switches vPC.
- Se os batimentos cardíacos NGOAM não forem recebidos (use etanalyzer com filter=**cfm** para capturar os pacotes de pulsação NGOAM).
- Mesmo que a interface de túnel dot1q seja órfã conectada em uma configuração vPC, ainda é necessário configurar o grupo multicast na Interface NVE para o segmento VN que faz parte do Xconnect em ambos os switches.
- Os batimentos cardíacos NGOAM são processados/enviados pelo switch principal do vPC. Mensagens de pulsação que chegam ao secundário do vPC serão sincronizadas com o principal

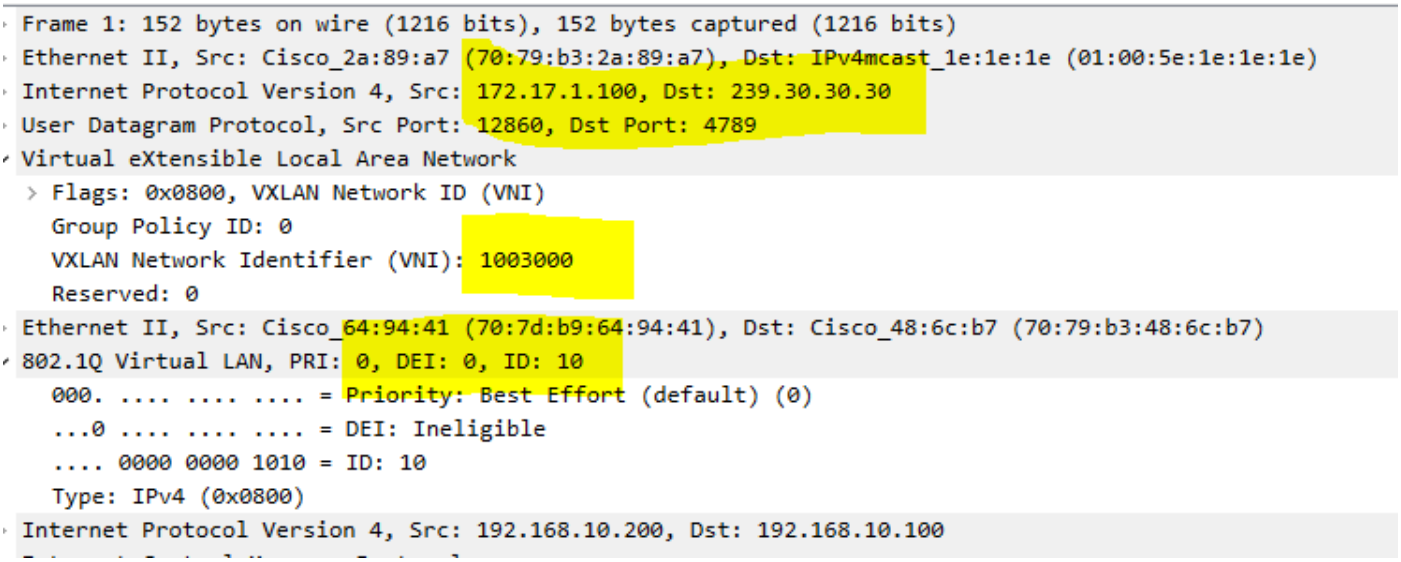
2. Quando Xconnect é configurado em uma VLAN, o tráfego de um local para outro é encapsulado com o endereço de destino externo=endereço multicast definido na interface NVE para esse segmento vn específico. É recomendável usar um grupo multicast exclusivo para as VLANs Xconnect. O multicast no núcleo/spine deve estar funcionando.

3. O tráfego multicast pode atingir ambas as caixas do vPC no lado remoto do Xconnect; No entanto, o vencedor do Decap (a caixa que pode desencapsular o BUM) será apenas um switch em um par vPC. Isso pode ser verificado usando o comando **show forwarding multicast route group <Group address> source <SRC IP>**. Se o Flag mostrado aqui for um **v** minúsculo, significa que a caixa é um perdedor de decantação e, se for um **V** maiúsculo, a caixa é o vencedor do decap e, portanto, pode desencapsular o tráfego multicast e encaminhá-lo mais para baixo.

4. Nas plataformas baseadas em 93180YC, quando o Host é órfão conectado a 9k1 e se não há OIL para S, G em 9k1, uma cópia do pacote multicast é enviada ao peer vPC usando um encapsulamento especial de IP de origem-> 127.0.0.1 e IP de destino-> NVE IP compartilhado e se o 9k2 tem OIL para S, G entry (entrada S, G), e o encaminhamento de tráfego será feito com cuidado pelo 9k2 para os locais remotos.

Captura do pacote

Aqui está uma captura de tela de uma captura de pacote feita no switch spine:



- Cabeçalho dot1q interno=10 preservado
- O VNI usado é 1003000 (que é o VNID da VLAN externa)
- O endereço IP destino seria o grupo multicast definido na interface nve