Nexus 9000:設定和驗證VXLAN Xconnect

目錄

簡介

本文檔介紹如何在Nexus 9000交換機上配置和驗證VXLAN Xconnect的快速參考。

必要條件

需求

思科建議您瞭解VXLAN EVPN。

採用元件

本文中的資訊係根據以下軟體和硬體版本：

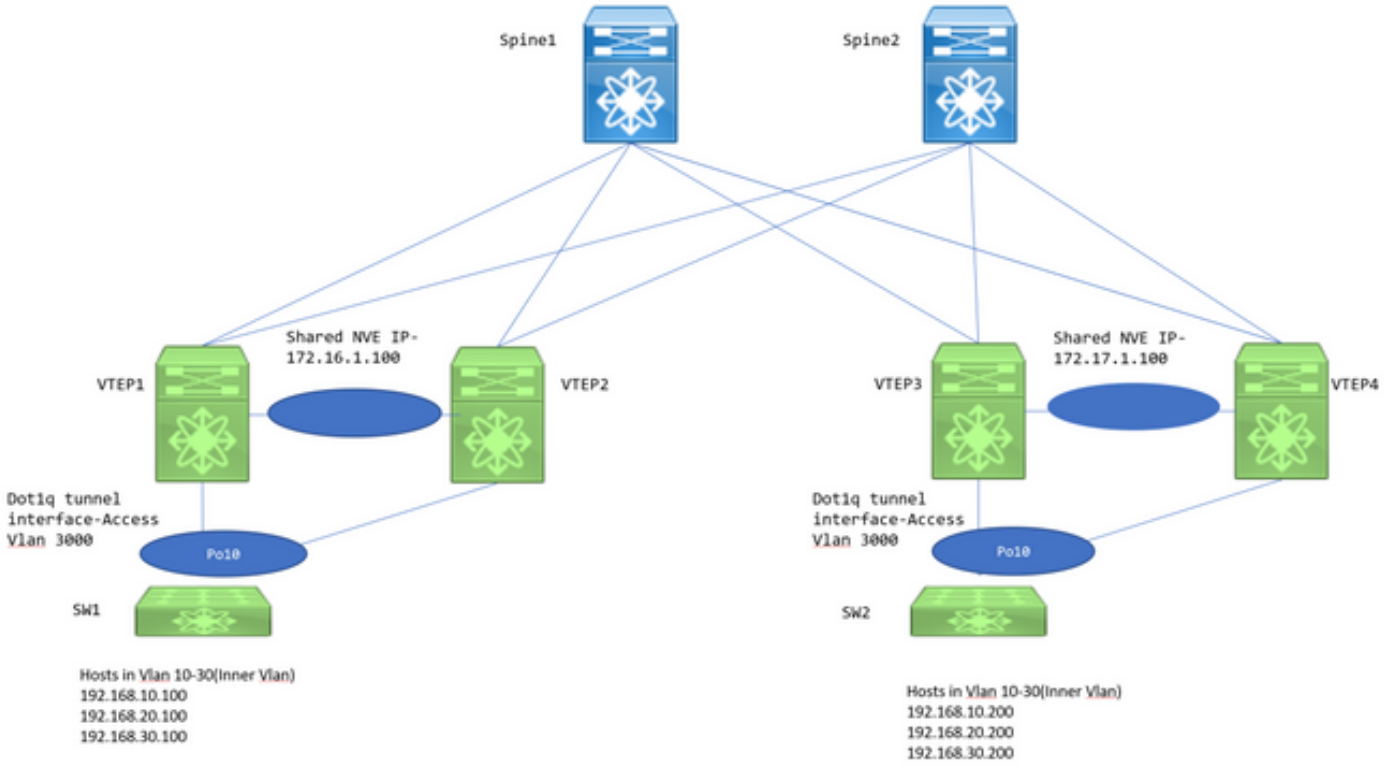
- N9K-C93180YC-EX
- NXOS 9.2(1)

本文中的資訊是根據特定實驗室環境內的裝置所建立。文中使用到的所有裝置皆從已清除（預設）的組態來啟動。如果您的網路正在作用，請確保您已瞭解任何指令可能造成的影響。

概觀

VXLAN Xconnect是一種點對點隧道的機制，用於從一個枝葉到另一個枝葉傳輸資料和控制資料包。保留內部Dot1q標籤，並將VXLAN封裝在指定為Xconnect VNID的外部VNID中。第2層控制幀(例如鏈路層發現協定(LLDP)、Cisco Discovery Protocol(CDP)、生成樹協定(STP))採用VXLAN封裝，並傳送到隧道的其它端。

拓撲



VTEP1、VTEP2、VTEP3和VTEP4是兩個vPC VTEP對，其設定方式可保留下游交換器的內部dot1q標籤，且封裝VXLAN時，使用外部VLAN ID的VXLAN VNID來傳送至遠端VTEP。 所有VTEP都是N9K-C93180YC-EX。

下游交換機是Nexus 3ks，在各VLAN中配置交換機虛擬介面(SVI)以模擬主機。

設定

1.此Xconnect拓撲中使用的外部VLAN是3000。這是使用VNID和Xconnect配置的路由器。

```
VTEP1# sh run vlan 3000

vlan 3000
  vn-segment 1003000
  xconnect
```

2.必須啟用功能NGOAM，並且需要此配置。

```
VTEP1# sh run ngoam

feature ngoam

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

3.通向下游交換機的Dot1q隧道配置。

```
VTEP1# sh run int po10

interface port-channel10
```

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

僅當將VTEP部署為vPC時，才需要vPC配置。否則，請跳過本文檔中提到的vPC配置。VXLAN Xconnect也可以在獨立VTEP上配置。

4.組播組必須在NVE介面下定義才能處理轉發。請注意，應在相關上行鏈路上啟用**ip pim sparse-mode**，並定義PIM RP，以便正確交換組播路由和PIM消息。通常，PIM RP是在主幹層定義的。

```
VTEP1# sh run int nve1

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

5.需要指定並允許在對等鏈路內將子網VLAN指定為本地VLAN。vPC VTEP需要此步驟。

```
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. BGP/EVPN配置：枝葉/主幹之間需要L2VPN EVPN鄰居來交換建立VXLAN Xconnect所需的3類路由。

- 這裡，IP地址192.168.100.1和192.168.100.2是拓撲中的主幹。通常L2VPN EVPN鄰居關係會形成到Spines。在iBGP方案中，主幹將所有枝葉交換機配置為路由反射器客戶端。
- 建議為BGP/OSPF和NVE使用單獨的環回。

```
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
```

附註：必須在Xconnect VLAN中禁用STP。MAC學習不會在Xconnect VLAN中發生，這基本上意味著沒有用於MAC地址的第2類bgp l2vpn evpn更新因此，來自一個vtep的流量將進行封裝，並將外部目標IP地址設定為為Xconnect VLAN定義的Mcast組(239.30.30.30)。

驗證

使用本節內容，確認您的組態是否正常運作。

1. BGP鄰居關係。

```
 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.接收型別3字首。

```
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對等。

```
 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. NGOAM檢查。

```
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
```

一旦NGOAM會話啟動，N3k將在CDP中看到對方。STP BPDU也通過隧道傳輸，因此交換機也同意根網橋的位置。

5.在下游交換機上進行驗證。

```
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
```

疑難排解

目前尚無適用於此組態的具體疑難排解資訊。

注意事項

1.如果vPC交換機中的配置不一致，在Xconnect VXLAN設定中，dot1q隧道介面將停滯在**error disabled**狀態。以下是介面將因為錯誤而停用的一些情況；

- 如果未在兩台vPC交換機上定義VLAN到VN段。
- 如果未在兩台vPC交換機上定義NVE到組播組。
- 如果未收到NGOAM心跳(請使用ethanalyzer和filter=**cfm**捕獲NGOAM心跳資料包)。
- 即使dot1q通道介面在vPC設定中是孤立連線的，仍需要在NVE介面下為兩台交換器上的Xconnect一部分的VN區段設定多點傳送群組。
- vPC主交換機處理/傳送NGOAM心跳。停在vPC輔助節點上的心跳消息將同步到主節點

2.在VLAN中設定Xconnect時，從一站點到另一站點的流量會使用在NVE介面下為該特定VN網段定義的外部目的地址=組播地址進行封裝。建議為Xconnect VLAN使用唯一組播組。核心/骨幹中的組播必須正常運行。

3.多點傳播流量可能同時命中Xconnect遠端的vPC盒；但是，Decap winner（可解封BUM的盒）將只是vPC對中的一個交換機。可以使用**show forwarding multicast route group <Group address> source <SRC IP>**命令驗證這一點。如果此處顯示的Flag是小寫**v**，則表示該框是丟包者，如果是**大**

寫V，則表示該框是丟包者，因此可以解除對組播流量的封裝並向下轉發它。

4.在基於93180YC的平台上，當主機孤立連線到9k1且在9k1上沒有S、G的OIL時，將使用源IP->127.0.0.1和目標IP->共用NVE IP的特殊封裝將組播資料包的副本傳送到vPC對等體，如果9k2具有S、G條目的OIL，則9k2將處理到遠端站點的流量轉發。

封包擷取

以下是在主幹交換器上擷取封包擷取的截圖：

```
Frame 1: 152 bytes on wire (1216 bits), 152 bytes captured (1216 bits)
Ethernet II, Src: Cisco_2a:89:a7 (70:79:b3:2a:89:a7), Dst: IPv4mcast_1e:1e:1e (01:00:5e:1e:1e:1e)
Internet Protocol Version 4, Src: 172.17.1.100, Dst: 239.30.30.30
User Datagram Protocol, Src Port: 12860, Dst Port: 4789
Virtual eXtensible Local Area Network
  > Flags: 0x0800, VXLAN Network ID (VNI)
    Group Policy ID: 0
    VXLAN Network Identifier (VNI): 1003000
    Reserved: 0
Ethernet II, Src: Cisco_64:94:41 (70:7d:b9:64:94:41), Dst: Cisco_48:6c:b7 (70:79:b3:48:6c:b7)
802.1Q Virtual LAN, PRI: 0, DEI: 0, ID: 10
    000. .... .... .... = Priority: Best Effort (default) (0)
    ...0 .... .... .... = DEI: Ineligible
    .... 0000 0000 1010 = ID: 10
    Type: IPv4 (0x0800)
Internet Protocol Version 4, Src: 192.168.10.200, Dst: 192.168.10.100
```

- 保留內部dot1q header=10
- 使用的VNI 1003000（外部VLAN的VNID）
- 目的IP地址是在NVE介面下定義的組播組