對方形拓撲中採用CloudSec的多站點VXLAN進行故障排除

目錄

簡介

本文檔介紹以方形拓撲連線的邊界網關之間的VXLAN多站點配置和使用CloudSec進行故障排除。

必要條件

需求

思科建議您熟悉以下主題：

- Nexus NXOS ©軟體。
- VXLAN EVPN技術。
- BGP和OSPF路由協定。

採用元件

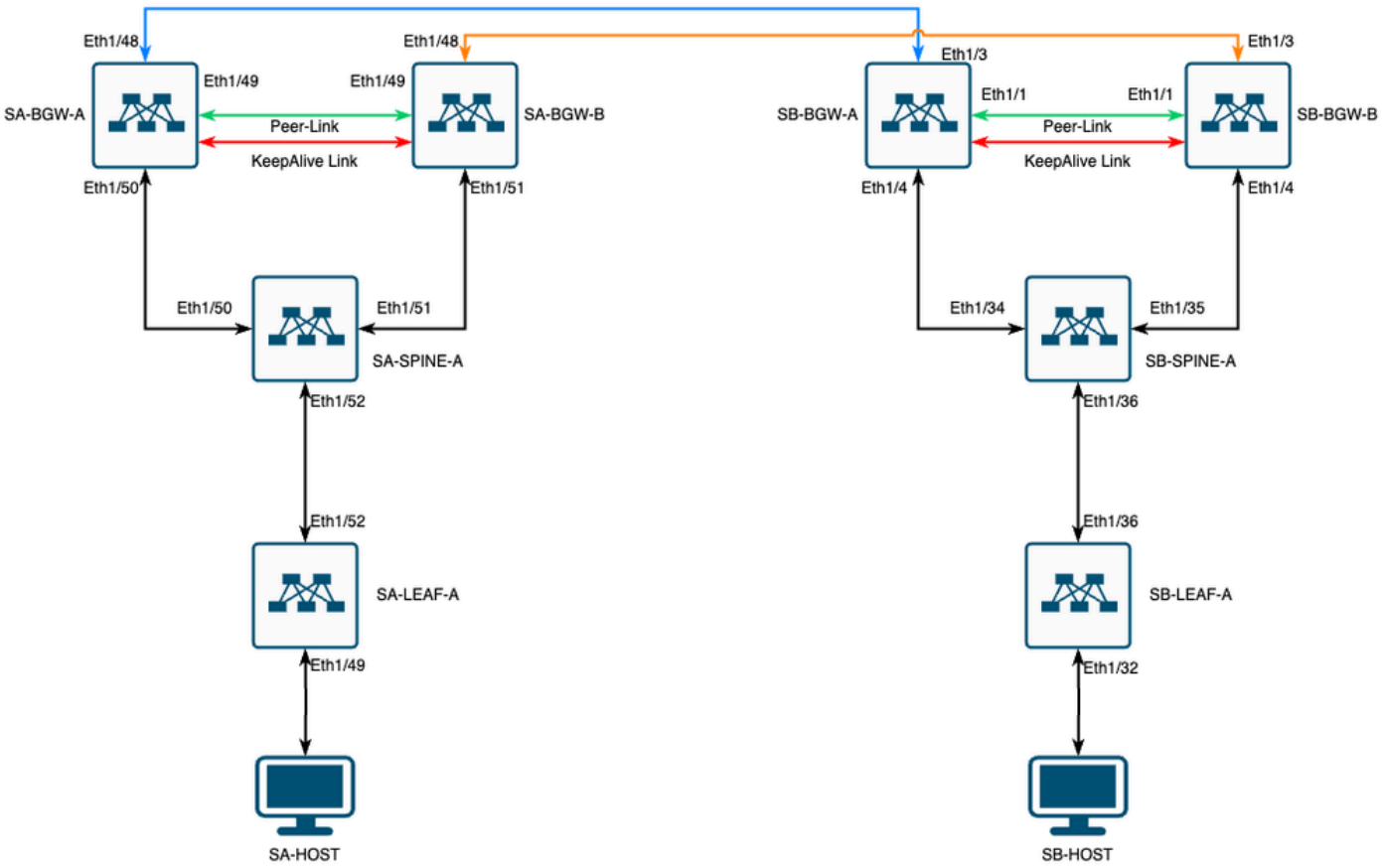
本檔案中的資訊是根據以下軟體和硬體版本：

- Cisco Nexus 9000。
- NXOS 10.3(4a)版。

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。如果您的網路運作中，請確保您瞭解任何指令可能造成的影響。

設定

網路圖表



方形拓撲中具有CloudSec的VXLAN MultiSite

拓撲的詳細資訊

- 雙站點多站點VXLAN EVPN交換矩陣。
- 兩個站點都配置了vPC邊界網關。
- 終端託管在VLAN 1100中。
- 每個站點上的邊界網關透過SVI介面Vlan3600相互之間具有IPv4 iBGP鄰居關係。
- 一個站點上的邊界網關僅與另一個站點上的直接連線的邊界網關具有eBGP IPv4鄰居關係。
- 站點A的邊界網關與站點B的邊界網關具有eBGP L2VPN EVPN鄰居關係。

編址計畫

表中的IP地址在配置期間使用：

	站點A	站點B				
裝置角色	介面ID	物理整合IP	RID環路IP	NVE回圈IP	MSITE-VIP	備份SVI IP

分葉	Eth1/52	192.168.1.1/30	192.168.2.1/32	192.168.3.1/32	不適用	不適用
骨幹	Eth1/52	192.168.1.2/30			不適用	
Eth1/50	192.168.1.5/30	192.168.2.2/32	不適用	不適用	不適用	Eth1/34
Eth1/51	192.168.1.9/30			不適用		Eth1/35
BGW-A	Eth1/51	192.168.1.6/30	192.168.2.3/32	192.168.3.2/32	192.168.100.1/32	192.168.4.1/3
Eth1/48	10.12.10.1/30		192.168.3.254/32			Eth1/3
BGW-B	Eth1/51	192.168.1.10/30	192.168.2.4/32	192.168.3.3/32	192.168.100.1/32	192.168.4.2/3
Eth1/48	10.12.10.5/30		192.168.3.254/32			Eth1/3

組態

- 請注意，本指南中僅顯示與多站點相關的配置。對於完整配置，您可以使用Cisco官方文檔VXLAN指南Cisco Nexus 9000系列NX-OS VXLAN配置指南，版本10.3(x)

要啟用CloudSec，必須在evpn multisite border-gateway下配置dci-advertise-pip 命令：

SA-BGW-A和SA-BGW-B	SB-BGW-A和SB-BGW-B
evpn multisite border-gateway 65001 dci-advertise-pip	evpn multisite border-gateway 65002 dci-advertise-pip

BGP配置

此配置特定於站點。

SA-BGW-A和SA-BGW-B	SB-BGW-A和SB-BGW-B
router bgp 65001 address-family ipv4 unicast maximum-paths 64 address-family l2vpn evpn maximum-paths 64 additional-paths send additional-paths receive	router bgp 65002 address-family ipv4 unicast maximum-paths 64 address-family l2vpn evpn maximum-paths 64 additional-paths send additional-paths receive

- **maximum-path** 命令允許從鄰居接收多個eBGP L2VPN EVPN路徑。
- **additional-path** 命令指示BGP進程通告裝置有能力傳送/接收其他路徑

對於邊界網關上的所有L3VNI VRF，還必須配置多路徑：

SA-BGW-A和SA-BGW-B	SB-BGW-A和SB-BGW-B

router bgp 65001 vrf tenant-1 address-family ipv4 unicast maximum-paths 64 address-family ipv6 unicast maximum-paths 64	router bgp 65002 vrf tenant-1 address-family ipv4 unicast maximum-paths 64 address-family ipv6 unicast maximum-paths 64

隧道加密配置

所有邊界網關上的此配置必須相同：

key chain CloudSec_Key_Chain1 tunnel-encryption key 1000 key-octet-string Cl0udSec! cryptographic-algorithm AES_128_CMAC feature tunnel-encryp

此配置特定於站點。tunnel-encryption命令必須僅應用於具有evpn multisite dci-tracking命令的介面。

SA-BGW-A和SA-BGW-B	SB-BGW-A和SB-BGW-B
tunnel-encryption peer-ip 192.168.13.2 keychain CloudSec_Key_Chain1 policy CloudSec_Policy1 tunnel-encryption peer-ip 192.168.13.3 keychain CloudSec_Key_Chain1 policy CloudSec_Policy1 interface Ethernet1/48 tunnel-encryption	tunnel-encryption peer-ip 192.168.3.2 keychain CloudSec_Key_Chain1 policy CloudSec_Policy1 tunnel-encryption peer-ip 192.168.3.3 keychain CloudSec_Key_Chain1 policy CloudSec_Policy1 interface Ethernet1/3 tunnel-encryption

啟用通道加密後，向鄰居通告路由時會將附加屬性增加到本地環回，並且所有eBGP IPv4單播鄰居必須看到此屬性：

<#root>

SA-BGW-A# show ip bgp 192.168.2.3 BGP routing table information for VRF default, address family IPv4 Unicast BGP routing table entry for 192.168.2

!---

This is a new attribute

Path type: redist, path is valid, not best reason: Locally originated, no labeled nexthop AS-Path: NON

對於Route Type-2，還有新屬性：

<#root>

SA-BGW-A# show bgp l2vpn evpn 00ea.bd27.86ef BGP routing table information for VRF default, address family L2VPN EVPN Route Distinguisher: 65

!---

Ethernet Segment Identifier (ESI) is also new attribute

```
 Path-id 1 (dual) advertised to peers: 192.168.2.2 SA-BGW-A#
```

驗證

啟用cloudsec之前，最好先檢查一下沒有它的安裝程式是否運作正常：

SA-BGW-A(config)# show clock Warning: No NTP peer/server configured. Time may be out of sync. 10:02:01.016 UTC Fri Jul 19 2024 Time source is N

配置完cloudsec後，SA上的終端必須成功ping站點B上的終端。但是，在某些情況下，ping操作可能會失敗。這取決於本地裝置選擇用於傳送cloudsec加密流量的cloudsec對等體。

SA-HOST-A# ping 10.100.20.10 PING 10.100.20.10 (10.100.20.10): 56 data bytes Request 0 timed out Request 1 timed out Request 2 timed out Request 3

疑難排解

檢查源終端上的本地ARP表：

SA-HOST-A# ping 10.100.20.10 count unlimited interval 1 Request 352 timed out Request 353 timed out Request 354 timed out 356 packets transmitted, 0

此輸出可證明BUM流量正在傳遞且控制平面正在運作。下一步是檢查通道加密狀態：

SA-BGW-A# show tunnel-encryption session Tunnel-Encryption Peer Policy Keychain RxStatus TxStatus ---------- ---------- ----------

此輸出顯示CloudSec會話已建立。下一步可以在SA-HOST-A上運行無限制的ping：

SA-HOST-A# ping 10.100.20.10 count unlimited interval 1

從現在起，您必須檢查站點A上的裝置，檢視流量是否到達此裝置。您可以在站點A上沿途的所有裝置上使用ELAM完成此任務。將預設值6更改 in-select 為9可以根據內部報頭進行匹配。您可以在此連結上閱讀有關ELAM的詳細資訊：Nexus 9000雲規模ASIC (Tahoe) NX-OS ELAM。

SA-LEAF-A上的ELAM

在生產網路中，存在多個SPINE裝置。要瞭解流量被傳送到哪個主幹，您必須先在LEAF上獲取ELAM。儘管使用 in-select 9 了此功能，但在連線到源的枝葉上，必須使用外部ipv4報頭，因為到達此LEAF的流量未進行VXLAN加密。在真實網路中，可能很難捕獲到您生成的確切資料包。在這種情況下，您可以使用特定長度運行ping，並使用Pkt len報頭來辨識資料包。預設情況下，icmp資料包長度為64位元組。加上20位元組的IP標頭，總結起來可得出84位元組的PKT Len：

```
<#root>

SA-LEAF-A# debug platform internal tah elam SA-LEAF-A(TAH-elam)# trigger init in-select 9 Slot 1: param values: start asic 0, start slice 0, lu-a2d 1, in

!---Note dpid value

 Dst Idx : 0xcd, Dst BD : 1100 Packet Type: IPv4 Outer Dst IPv4 address: 10.100.20.10 Outer Src IPv4 ad

Pkt len = 84

, Checksum = 0xb4ae

!---64 byte + 20 byte IP header Pkt len = 84

 Inner Payload Type: CE L4 Protocol : 1 L4 info not available Drop Info: ---------- LUA: LUB: LUC: LUD:

!---

Put dpid value here

 IF_STATIC_INFO: port_name=Ethernet1/52,if_index:0x1a006600,ltl=5940,slot=0, nxos_port=204,dmod=1,dpid=
```

從該輸出中，您可以看到流量到達SA-LEAF-A並從介面Ethernet1/52轉發，該介面從拓撲連線到SA-SPINE-A。

SA-SPINE-A上的伊蘭

在SPINE上，由於50位元組的VXLAN報頭也增加了，Pkt Len值將更高。預設情況下，沒有 vxlan-parse 或 feature nv overlay ，SPINE在內部報頭上不匹配。因此，必須對SPINE使用 vxlan-parse enable 命令：

```
<#root>

SA-SPINE-A(config-if)# debug platform internal tah elam SA-SPINE-A(TAH-elam)# trigger init in-select 9 Slot 1: param values: start asic 0, start slice 0,

!---

84 bytes + 50 bytes VXLAN header Pkt len = 134

 Inner Payload Type: IPv4 Inner Dst IPv4 address: 10.100.20.10 Inner Src IPv4 address: 10.100.10.10 L4
```

SA-SPINE-A根據輸出向SA-BGW-A傳送流量。

SA-BGW-A上的ELAM

SA-BGW-A(TAH-elam-insel9)# set inner ipv4 src_ip 10.100.10.10 dst_ip 10.100.20.10 SA-BGW-A(TAH-elam-insel9)# start SA-BGW-A(TAH-elam-ins

根據SA-BGW-A的輸出，流量從Ethernet1/48流向SB-BGW-A。下一步是檢查SB-BGW-A：

<#root>

SB-BGW-A# debug platform internal tah elam SB-BGW-A(TAH-elam)# trigger init in-select 9 Slot 1: param values: start asic 0, start slice 0, lu-a2d 1, in-

!---Reset the previous filter and start again just in case if packet was not captured.

SB-BGW-A(TAH-elam-insel9)# reset SB-BGW-A(TAH-elam-insel9)# set inner ipv4 src_ip 10.100.10.10 dst_ip

根據SB-BGW-A的輸出，ELAM甚至沒有被觸發。這意味著SB-BGW-B正在接收資料包，並且無法正確解密和解析這些資料包，或者根本無法接收這些資料包。要瞭解cloudsec流量發生的情況，可以在SB-BGW-A上再次運行ELAM，但觸發過濾器必須設定為用於cloudsec的外部IP地址，因為沒有辦法檢視cloudsec加密傳輸資料包的內部報頭。從先前的輸出中，您知道SA-BGW-A處理了流量，這意味著SA-BGW-A使用cloudsec加密流量。因此，您可以使用SA-BGW-A的NVE IP作為ELAM的觸發過濾器。根據前面的輸出，VXLAN加密的ICMP資料包長度為134位元組。加上摘要中的32位元組cloudsec標頭，可得出166位元組：

<#root>

SB-BGW-A(TAH-elam-insel9)# reset SB-BGW-A(TAH-elam-insel9)# set outer ipv4 src_ip 192.168.3.2 SB-BGW-A(TAH-elam-insel9)# start SB-BGW-

192.168.13.3 ***!---NVE IP address of SB-BGW-B***

Outer Src IPv4 address: 192.168.3.2 Ver = 4, DSCP = 0, Don't Fragment = 0 Proto = 17, TTL = 254, More

!---134 byte VXLAN packet + 32 byte cloudsec header Pkt len = 166

Inner Payload Type: CE L4 Protocol : 17 L4 info not available Drop Info: ---------- LUA: LUB: LUC: LUD

!---To reach SB-BGW-B NVE IP traffic was sent out of Ethernet1/4 which is connected to SB-SPINE-A

SB-BGW-A(TAH-elam-insel9)# show system internal ethpm info all | i i "dpid=130" IF_STATIC_INFO: port_n
SB-BGW-A(TAH-elam-insel9)# show cdp neighbors interface ethernet 1/4 Capability Codes: R - Router, T - Trans-Bridge, B - Source-Route-Bridge S - S

192.168.13.3/32

, ubest/mbest: 1/0 *via 192.168.11.5,

Eth1/4

, [110/6], 00:56:13, ospf-UNDERLAY, intra via

192.168.14.2

, [200/0], 01:13:46, bgp-65002, internal, tag 65002

!---The device still have a route for SB-BGW-B NVE IP via SVI

SB-BGW-A(TAH-elam-insel9)# show ip route 192.168.14.2 IP Route Table for VRF "default" '*' denotes best

***via 192.168.14.2, Vlan3600**

, [250/0], 01:15:05, am SB-BGW-A(TAH-elam-insel9)# show ip arp 192.168.14.2 Flags: * - Adjacencies lear

ecce.1324.c803

Vlan3600

SB-BGW-A(TAH-elam-insel9)# show mac address-table address ecce.1324.c803 Legend: * - primary entry, G

3600

ecce.1324.c803

static - F F

vPC Peer-Link(R)

SB-BGW-A(TAH-elam-insel9)#

從該輸出中，您可以看到，根據路由表，Cloudsec流量透過介面Ethernet1/4轉發到SB-BGW-B。根據Cisco Nexus 9000系列NX-OS VXLAN配置指南，版本10.3(x)指南和限制：

- 發往交換機的CloudSec流量必須透過DCI上行鏈路進入交換機。

根據同一指南的vPC Border Gateway Support for Cloudsec部分，如果vPC BGW獲知對等vPC BGW的PIP地址並在DCI端進行通告，則來自兩個vPC BGW的BGP路徑屬性將相同。因此，DCI中間節點最終可以從不擁有PIP地址的vPC BGW中選擇路徑。在此場景中，MCT鏈路用於來自遠端站點的加密流量。但在本例中，使用的是指向SPINE的介面，儘管如此，BGW也透過BackUp SVI具有OSPF鄰接關係。

SB-BGW-A(TAH-elam-insel9)# show ip ospf neighbors OSPF Process ID UNDERLAY VRF default Total number of neighbors: 2 Neighbor ID Pri State

問題的原因和修復

原因是SVI介面的OSPF開銷。預設情況下，在NXOS上，自動開銷參考頻寬為40G。SVI介面的頻寬為1Gbps，而物理介面的頻寬為10Gbps：

<#root>

SB-BGW-A(TAH-elam-insel9)# show ip ospf interface brief OSPF Process ID UNDERLAY VRF default Total number of interface: 5 Interface ID Area C

<Output omitted>

Eth1/4 5 0.0.0.0 1 P2P 1 up

在這種情況下，對SVI成本進行管理更改即可解決問題。必須在所有邊界網關上完成調整。

<#root>

SB-BGW-A(config)# int vlan 3600 SB-BGW-A(config-if)# ip ospf cost 1 SB-BGW-A(config-if)# sh ip route 192.168.13.3 IP Route Table for VRF "defau

via 192.168.14.2

, Vlan3600, [110/2], 00:00:08, ospf-UNDERLAY, intra via 192.168.14.2, [200/0], 01:34:07, bgp-65002, int

!---The ping is started to work immediately

Request 1204 timed out Request 1205 timed out Request 1206 timed out 64 bytes from 10.100.20.10: icmp_seq=1207 ttl=254 time=1.476 ms 64 bytes from

關於此翻譯

思科已使用電腦和人工技術翻譯本文件，讓全世界的使用者能夠以自己的語言理解支援內容。請注意，即使是最佳機器翻譯，也不如專業譯者翻譯的內容準確。Cisco Systems, Inc. 對這些翻譯的準確度概不負責，並建議一律查看原始英文文件（提供連結）。