Risoluzione dei problemi della VXLAN multisito con CloudSec nella topologia quadrata

Sommario

Introduzione

Questo documento descrive la configurazione e la risoluzione dei problemi di VXLAN Multisite con CloudSec tra i gateway di confine connessi nella topologia quadrata.

Prerequisiti

Requisiti

Cisco raccomanda la conoscenza dei seguenti argomenti:

- Nexus NXOS © Software
- Tecnologia VXLAN VPN.
- protocolli di routing BGP e OSPF.

Componenti usati

Le informazioni di questo documento si basano sulle seguenti versioni software e hardware:

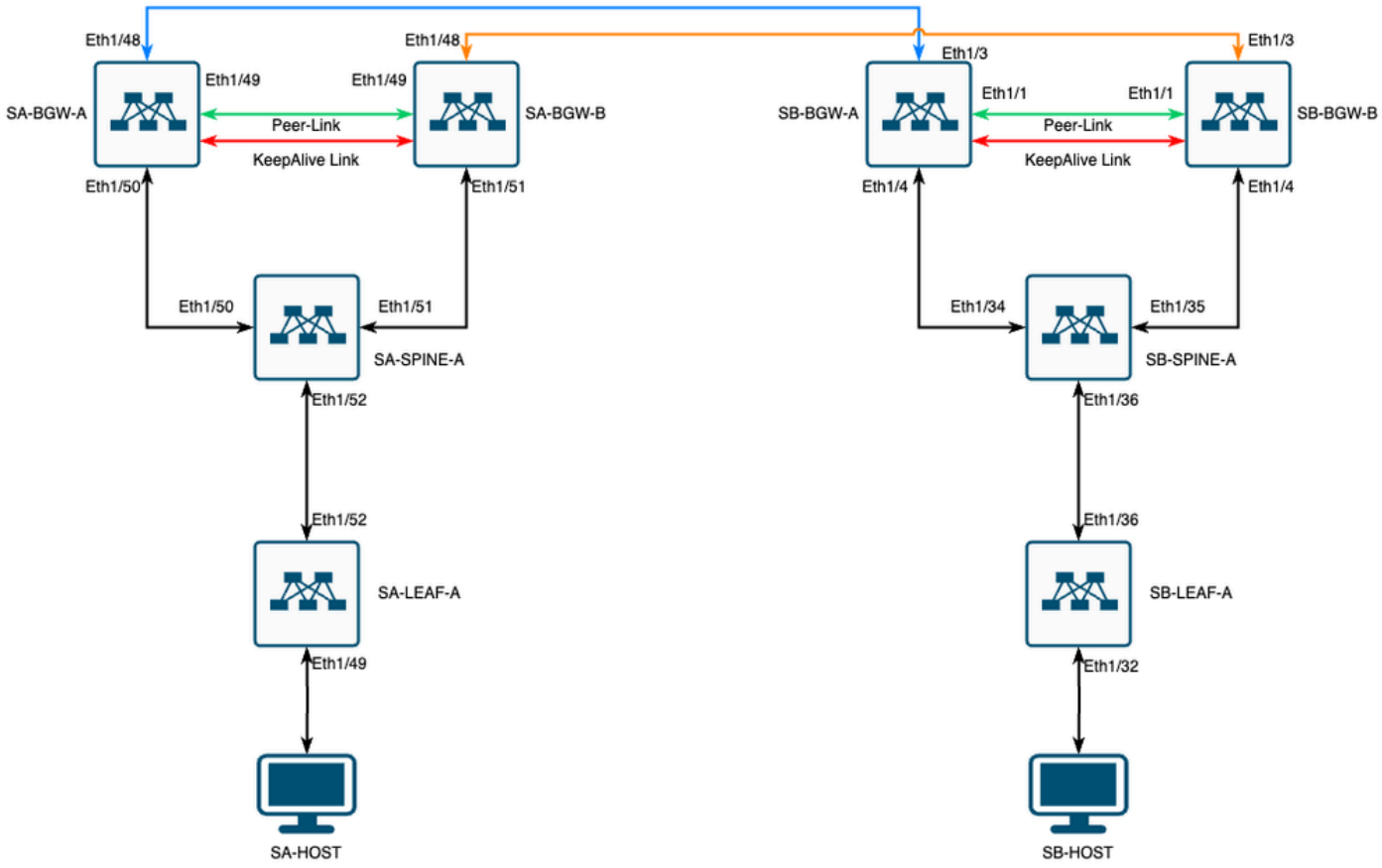
- Cisco Nexus 9000

- NXOS versione 10.3(4a).

Le informazioni discusse in questo documento fanno riferimento a dispositivi usati in uno specifico ambiente di emulazione. Su tutti i dispositivi menzionati nel documento la configurazione è stata ripristinata ai valori predefiniti. Se la rete è operativa, valutare attentamente eventuali conseguenze derivanti dall'uso dei comandi.

Configurazione

Esempio di rete



VXLAN MultiSite con CloudSec nella topologia quadrata

Dettagli della topologia

- Due fabric VPN VXLAN multisito.
- Entrambi i siti sono configurati con vPC Border Gateway.
- Gli endpoint sono ospitati sulla VLAN 1100.
- I gateway di confine di ciascun sito dispongono di una connessione IPv4 iBGP tra di loro sull'interfaccia SVI Vlan3600.
- I gateway di confine su un sito hanno un vicinato eBGP IPv4 solo con gateway di confine direttamente connesso sull'altro sito.
- I gateway di confine sul sito A hanno un vicinato eBGP L2VPN con gateway di confine sul sito B.

Piano di indirizzamento

Gli indirizzi IP riportati nella tabella vengono utilizzati durante la configurazione.

	SITO A	SITO B				
Ruolo dispositivo	ID interfaccia	IP interno fisico	IP loop RID	NVE Loop IP	MSITE-VIP	Backup su IP
FOGLIA	Eth 1/52	192.168.1.1/30	192.168.2.1/32	192.168.3.1/32	N/D	N/D
DORSO	Eth 1/52	192.168.1.2/30			N/D	
Eth 1/50	192.168.1.5/30	192.168.2.2/32	N/D	N/D	N/D	Eth 1/3
Eth 1/51	192.168.1.9/30			N/D		Eth 1/3
BGW-A	Eth 1/51	192.168.1.6/30	192.168.2.3/32	192.168.3.2/32	192.168.100.1/32	192.168.4
Eth 1/48	10.12.10.1/30		192.168.3.254/32			Eth 1/3
BGW-B	Eth 1/51	192.168.1.10/30	192.168.2.4/32	192.168.3.3/32	192.168.100.1/32	192.168.4
Eth 1/48	10.12.10.5/30		192.168.3.254/32			Eth 1/3

Configurazioni

- In questa guida viene mostrata solo la configurazione relativa a più siti. Per la configurazione completa, è possibile utilizzare la guida ufficiale Cisco per la documentazione della VXLAN [Cisco Nexus serie 9000 NX-OS VXLAN Configuration Guide, versione 10.3(x)]

Per abilitare CloudSec, il dci-advertise-pip comando deve essere configurato in evpn multisito border-gateway:

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

configurazione BGP

Questa configurazione è specifica del sito.

SA-BGW-A e SA-BGW-B	SB-BGW-A e 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

- Il comando **maximum-path** consente di ricevere più percorsi VPN eBGP L2VPN dal router adiacente.
- Il comando **additional-path** indica al processo BGP di indicare che il dispositivo è in grado di inviare/ricevere percorsi aggiuntivi

Per tutti i VRF L3VNI sui gateway di confine, anche il multipath deve essere configurato:

SA-BGW-A e SA-BGW-B	SB-BGW-A e 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

Configurazione della crittografia del tunnel

Questa configurazione deve essere la stessa su tutti i gateway di confine:

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

Questa configurazione è specifica del sito. Il comandotunnel-encryption deve essere applicato solo all'interfaccia che dispone del evpn multisite dci-trackingcomando.

SA-BGW-A e SA-BGW-B	SB-BGW-A e 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

Dopo aver abilitato la crittografia del tunnel, vengono aggiunti attributi aggiuntivi al loopback locale mentre vengono annunciate le route verso il router adiacente e tutti i router adiacenti unicast eBGP IPv4 devono visualizzare questo attributo:

<#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
```

Per il tipo di instradamento 2 è disponibile anche un nuovo attributo:

<#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#
```

Verifica

Prima di abilitare cloudsec, è consigliabile verificare se l'installazione funziona correttamente senza di esso:

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

Anche dopo la configurazione cloud-sec, l'endpoint nell'associazione di protezione deve eseguire correttamente il ping dell'endpoint nel sito B. Tuttavia, in alcuni casi il ping può avere esito negativo. Dipende dal peer cloudsec selezionato dal dispositivo locale per inviare il traffico crittografato cloudsec.

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

Risoluzione dei problemi

Controllare la tabella ARP locale sull'endpoint di origine:

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

Questo output dimostra che il traffico BUM è in transito e il Control-Plane funziona. Il passaggio successivo è verificare lo stato della crittografia del tunnel:

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

Questo output mostra che la sessione CloudSec è stata stabilita. Nel passaggio successivo, è possibile eseguire il ping illimitato su SA-HOST-A:

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

Da questo punto è necessario controllare i dispositivi sul sito A e verificare se il traffico sta raggiungendo questi dispositivi. È possibile eseguire questa operazione con ELAM su tutti i dispositivi lungo il percorso nel sito A. Il passaggio in-select dal valore predefinito 6 a 9 consente di ottenere una corrispondenza in base alle intestazioni interne. Per ulteriori informazioni su ELAM, fare clic sul seguente collegamento: Nexus 9000 Cloud Scale ASIC (Tahoe) NX-OS ELAM.

ELAM su SA-LEAF-A

Nella rete di produzione sono presenti più dispositivi SPINE. Per capire a quale dorso è stato inviato il traffico, devi prima prendere un ELAM su FOGLIA. Nonostante ciò, in-select 9 sul router LEAF collegato all'origine, deve essere usata l'intestazione ipv4 esterna, in quanto il traffico raggiunto da questa foglia non è criptato da VXLAN. Nella rete reale, può essere difficile intercettare il pacchetto esattamente che è stato generato. In questi casi, è possibile eseguire il ping con una lunghezza specifica e usare l'intestazione Pkt Len per identificare il pacchetto. Per impostazione predefinita, il pacchetto icmp è lungo 64 byte. Più 20 byte di intestazione IP, che in riepilogo ha dato 84 byte 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=

Da questo output si può vedere che il traffico raggiunge SA-LEAF-A e viene inoltrato all'interfaccia Ethernet1/52, che è collegata a SA-SPINE-A dalla topologia.

ELAM su SA-SPINE-A

Su SPINE il valore Pkt Len sarà maggiore, poiché è stata aggiunta anche l'intestazione VXLAN da 50 byte. Per impostazione predefinita, SPINE non può corrispondere nelle intestazioni interne senza vxlan-parse o feature nv overlay . È quindi necessario utilizzare il vxlan-parse

enable comando su SPINE:

<#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 invia il traffico verso SA-BGW-A in base all'output.

ELAM su SA-BGW-A

<#root>

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

Secondo l'output di SA-BGW-A, il traffico è stato indirizzato su Ethernet1/48 verso SB-BGW-A. Il passo successivo è controllare su 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

Secondo l'output di SB-BGW-A, ELAM non è stato nemmeno attivato. Ciò significa che o SB-BGW-B riceve i pacchetti e non è in grado di decriptarli e analizzarli correttamente, o non li riceve affatto. Per capire cosa è successo con il traffico cloudsec, è possibile eseguire un ELAM su SB-BGW-A di nuovo, ma il filtro trigger deve essere impostato sull'indirizzo IP esterno che viene usato per cloudsec, in quanto non c'è modo di vedere l'intestazione interna del pacchetto di transito crittografato cloudsec. Dall'output precedente si sa che la SA-BGW-A ha gestito il traffico, il che significa che la SA-BGW-A cripta il traffico con cloudsec. Quindi, è possibile utilizzare NVE IP di SA-BGW-A come filtro trigger per ELAM. Dagli output precedenti, la lunghezza del pacchetto ICMP crittografato con VXLAN è 134 byte. Più 32 byte cloudsec intestazione in riepilogo fornisce 166 byte:

<#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)#

Da questo output, è possibile vedere che il traffico del cloudsec viene inoltrato verso l'SB-BGW-B tramite l'interfaccia Ethernet1/4, in base alla tabella di routing. In base alla guida alla configurazione della VXLAN per Cisco Nexus serie 9000 NX-OS, le linee guida e le limitazioni della release 10.3(x):

- Il traffico CloudSec destinato allo switch deve entrare nello switch tramite gli uplink DCI.

In base alla sezione Supporto vPC Border Gateway per Cloudsec della stessa guida, se vPC BGW apprende l'indirizzo PIP dei BGW vPC peer e pubblicizza sul lato DCI, gli attributi del percorso BGP di entrambi i BGW vPC saranno gli stessi. Quindi i nodi intermedi DCI possono finire per scegliere il percorso da vPC BGW che non possiede l'indirizzo PIP. In questo scenario, il collegamento MCT viene utilizzato per il traffico

crittografato proveniente dal sito remoto. Tuttavia, in questo caso, viene utilizzata l'interfaccia verso la SPINE, nonostante questo, i BGW hanno anche un'adiacenza OSPF tramite la SVI di BackUp.

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

Motivo del problema e correzione

Il motivo è il costo OSPF dell'interfaccia SVI. Per impostazione predefinita, la larghezza di banda di riferimento del costo automatico di NXOS è di 40 GB. Le interfacce SVI hanno una larghezza di banda di 1 Gbps, mentre l'interfaccia fisica ha una larghezza di banda di 10 Gbps:

<#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

In tal caso, la modifica amministrativa del costo per SVI può risolvere il problema. La messa a punto deve essere effettuata su tutti i gateway di confine.

<#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

Informazioni su questa traduzione

Cisco ha tradotto questo documento utilizzando una combinazione di tecnologie automatiche e umane per offrire ai nostri utenti in tutto il mondo contenuti di supporto nella propria lingua. Si noti che anche la migliore traduzione automatica non sarà mai accurata come quella fornita da un traduttore professionista. Cisco Systems, Inc. non si assume alcuna responsabilità per l’accuratezza di queste traduzioni e consiglia di consultare sempre il documento originale in inglese (disponibile al link fornito).