Dépannage d'un VXLAN multisite avec CloudSec dans une topologie carrée

Table des matières

Introduction

Ce document décrit la configuration et le dépannage VXLAN multisite avec CloudSec entre les passerelles de périphérie connectées en topologie carrée.

Conditions préalables

Exigences

Cisco vous recommande de vous familiariser avec les sujets suivants :

- Logiciel Nexus NXOS ©.
- Technologie EVPN VXLAN.
- Protocoles de routage BGP et OSPF.

Composants utilisés

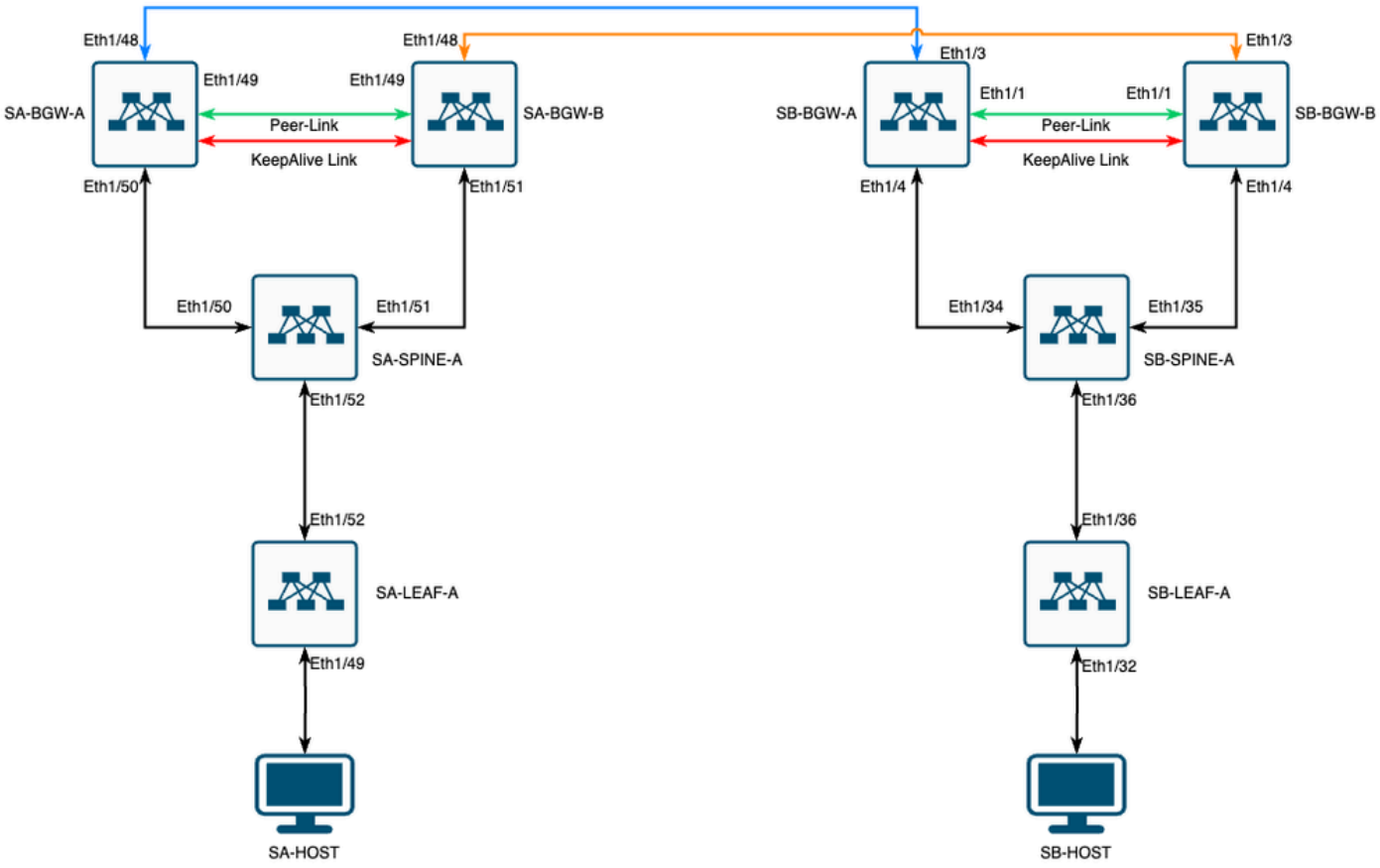
Les informations contenues dans ce document sont basées sur les versions logicielles et matérielles suivantes :

- Cisco Nexus 9000.
- NXOS version 10.3(4a).

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. Si votre réseau est en ligne, assurez-vous de bien comprendre l'incidence possible des commandes.

Configurer

Diagramme du réseau



VXLAN MultiSite avec CloudSec en topologie carrée

Détails de la topologie

- Fabric EVPN VXLAN multisite à deux sites.
- Les deux sites sont configurés avec des passerelles de périphérie vPC.
- Les terminaux sont hébergés dans le VLAN 1100.
- Les passerelles de périphérie sur chaque site ont un voisinage IPv4 iBGP entre elles sur l'interface SVI Vlan3600.
- Les passerelles en limite sur un site ont un voisinage eBGP IPv4 uniquement avec une passerelle en limite connectée directement sur l'autre site.
- Les passerelles en limite sur le site A ont un voisinage EVPN L2VPN eBGP avec des passerelles en limite sur le site B.

Plan d'adressage

Les adresses IP de la table sont utilisées lors de la configuration :

	SITE A	SITE B				
Rôle du périphérique	ID interface	IP Int physique	Adresse IP de boucle RID	IP de boucle NVE	VIP DU SITE	IP SV sauveg
FEUILLE	Eth1/52	192.168.1.1/30	192.168.2.1/32	192.168.3.1/32	S/O	S/
DOS	Eth1/52	192.168.1.2/30			S/O	
Eth1/50	192.168.1.5/30	192.168.2.2/32	S/O	S/O	S/O	Eth1
Eth1/51	192.168.1.9/30			S/O		Eth1
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
Eth1/48	10.12.10.1/30		192.168.3.254/32			Eth1
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
Eth1/48	10.12.10.5/30		192.168.3.254/32			Eth1

Configurations

- Notez que dans ce guide, seule la configuration multisite est affichée. Pour une configuration complète, vous pouvez utiliser le guide de documentation officiel Cisco pour VXLAN Guide de configuration VXLAN de la gamme Cisco Nexus 9000 NX-OS, version 10.3(x)

Afin d'activer CloudSec, la dci-advertise-pip commande doit être configurée sous la passerelle evpn multisite border-gateway :

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

Configuration BGP

Cette configuration est spécifique au site.

SA-BGW-A et SA-BGW-B	SB-BGW-A et 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

- La commande **maximum-path** permet de recevoir plusieurs chemins EVPN L2VPN eBGP du voisin.

- La commande **additional-path** indique au processus BGP d'annoncer que le périphérique est capable d'envoyer/recevoir des chemins supplémentaires

Pour tous les VRF L3VNI sur les passerelles périphériques, le multichemin doit également être configuré :

SA-BGW-A et SA-BGW-B	SB-BGW-A et 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

Configuration du chiffrement de tunnel

Cette configuration doit être identique sur toutes les passerelles de périphérie :

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

Cette configuration est spécifique au site. La tunnel-encryption commande doit être appliquée uniquement à l'interface qui a la evpn multisite dci-trackingcommande.

SA-BGW-A et SA-BGW-B	SB-BGW-A et 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

Après l'activation du cryptage de tunnel, des attributs supplémentaires sont ajoutés au bouclage local lors de l'annonce des routes au voisin et tous les voisins de monodiffusion IPv4 eBGP doivent voir cet attribut :

<#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
```

Pour le type de route 2, il existe également un nouvel attribut :

<#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#
```

Vérifier

Avant d'activer cloudsec, il est bon de vérifier si la configuration fonctionne correctement sans :

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

Après la configuration de cloudsec également, le point de terminaison sur SA doit envoyer une requête ping au point de terminaison sur le site B. Mais, dans certains cas, la requête ping peut échouer. Cela dépend de l'homologue cloudsec sélectionné par le périphérique local pour envoyer le trafic chiffré 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

Dépannage

Vérifiez la table ARP locale sur le point d'extrémité source :

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

Ce résultat prouve que le trafic de la carte BUM est en cours de transfert et que le plan de contrôle fonctionne. L'étape suivante consiste à vérifier l'état de cryptage du tunnel :

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

Ce résultat montre que la session CloudSec est établie. L'étape suivante consiste à exécuter une commande ping illimitée sur SA-HOST-A :

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

À partir de ce point, vous devez vérifier les périphériques sur le site A et voir si le trafic atteint ces périphériques. Vous pouvez effectuer cette tâche avec ELAM sur tous les périphériques situés sur le chemin du site A. La modification in-select de la valeur par défaut de 6 à 9 permet de faire correspondre les en-têtes internes. Pour en savoir plus sur ELAM, cliquez sur ce lien : Nexus 9000 Cloud Scale ASIC (Tahoe) NX-OS ELAM.

ELAM sur SA-LEAF-A

Dans le réseau de production, il existe plusieurs périphériques SPINE. Pour savoir à quelle colonne vertébrale le trafic a été envoyé, vous devez d'abord prendre un ELAM sur LEAF. Malgré cela, au niveau du LEAF connecté à la source, l'en-tête ipv4 du routeur doit être in-select 9 utilisé, car le trafic atteint par ce LEAF n'est pas chiffré VXLAN. Dans un réseau réel, il peut être difficile d'attraper le paquet exact que vous avez généré. Dans de tels cas, vous pouvez exécuter une requête ping avec une longueur spécifique et utiliser l'en-tête Pkt len pour identifier votre paquet. Par défaut, le paquet icmp a une longueur de 64 octets. Plus 20 octets d'en-tête IP, qui en résumé vous a donné 84 octets 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=

À partir de ce résultat, vous pouvez voir que le trafic est atteint par SA-LEAF-A et transféré à l'interface Ethernet1/52, qui est connectée à SA-SPINE-A à partir de la topologie.

ELAM sur SA-SPINE-A

Sur SPINE, la valeur Pkt Len sera plus élevée, puisque l'en-tête VXLAN de 50 octets a également été ajouté. Par défaut, SPINE ne peut pas correspondre sur les en-têtes internes sans vxlan-parse ou feature nv overlay . Vous devez donc utiliser la vxlan-parse enable commande sur

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 envoie le trafic vers SA-BGW-A conformément à la sortie.

ELAM sur SA-BGW-A

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

Selon le résultat de SA-BGW-A, le trafic a été acheminé par Ethernet1/48 vers SB-BGW-A. L'étape suivante consiste à vérifier sur 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

D'après le résultat de SB-BGW-A, ELAM n'a même pas été déclenché. Cela signifie que soit SB-BGW-B reçoit les paquets et ne peut pas les décrypter et les analyser correctement, soit il ne les reçoit pas du tout. Pour comprendre ce qui s'est passé avec le trafic cloudsec, vous pouvez exécuter un ELAM sur SB-BGW-A à nouveau, mais le filtre de déclenchement doit être défini sur l'adresse IP externe qui est utilisée pour cloudsec, car il n'y a aucun moyen de voir l'en-tête interne du paquet de transit chiffré cloudsec. D'après le résultat précédent, vous savez que la SA-BGW-A a géré le trafic, ce qui signifie que la SA-BGW-A chiffre le trafic avec cloudsec. Ainsi, vous pouvez utiliser l'IP NVE de SA-BGW-A comme filtre de déclenchement pour ELAM. D'après les résultats précédents, la longueur du paquet ICMP chiffré VXLAN est de 134 octets. Plus 32 octets en-tête cloudsec dans le résumé vous donne 166 octets :

<#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)#

À partir de ce résultat, vous pouvez voir que le trafic cloudsec est transféré vers SB-BGW-B via l'interface Ethernet1/4, en fonction de la table de routage. Selon le Guide de configuration du VXLAN NX-OS de la gamme Cisco Nexus 9000, version 10.3(x), directives et limites :

- Le trafic CloudSec destiné au commutateur doit entrer dans le commutateur via les liaisons ascendantes DCI.

Selon la section Prise en charge de la passerelle frontière vPC pour Cloudsec du même guide, si vPC BGW apprend l'adresse PIP des homologues vPC BGW et annonce côté DCI, les attributs de chemin BGP des deux vPC BGW seront identiques. Par conséquent, les noeuds intermédiaires DCI peuvent finir par choisir le chemin à partir de vPC BGW qui ne possède pas l'adresse PIP. Dans ce scénario, la liaison MCT est utilisée pour le trafic chiffré provenant du site distant. Mais dans ce cas, l'interface vers le SPINE est utilisée, malgré cela, les BGW ont

également une contiguïté OSPF via l'interface SVI de sauvegarde.

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

Raison du problème et résolution

La raison est le coût OSPF de l'interface SVI. Par défaut, sur NXOS, la bande passante de référence de coût automatique est de 40 G. Les interfaces SVI ont une bande passante de 1 Gbit/s, tandis que l'interface physique a une bande passante de 10 Gbit/s :

<#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

Dans ce cas, la modification administrative du coût de l'interface SVI peut résoudre le problème. Le réglage doit être effectué sur toutes les passerelles de périphérie.

<#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

À propos de cette traduction
Cisco a traduit ce document en traduction automatisée vérifiée par une personne dans le cadre d'un service mondial permettant à nos utilisateurs d'obtenir le contenu d'assistance dans leur propre langue.

Il convient cependant de noter que même la meilleure traduction automatisée ne sera pas aussi précise que celle fournie par un traducteur professionnel.