Fehlerbehebung bei VXLAN mit mehreren Standorten und CloudSec in Square-Topologie

Inhalt

Einleitung

In diesem Dokument werden die Konfiguration und Fehlerbehebung für mehrere VXLANs mit CloudSec zwischen in Vierkanttopologie verbundenen Grenz-Gateways beschrieben.

Voraussetzungen

Anforderungen

Cisco empfiehlt, dass Sie mit den folgenden Themen vertraut sind:

- Nexus NX-OS © Software.
- VXLAN-EVPN-Technologie.
- BGP- und OSPF-Routing-Protokolle.

Verwendete Komponenten

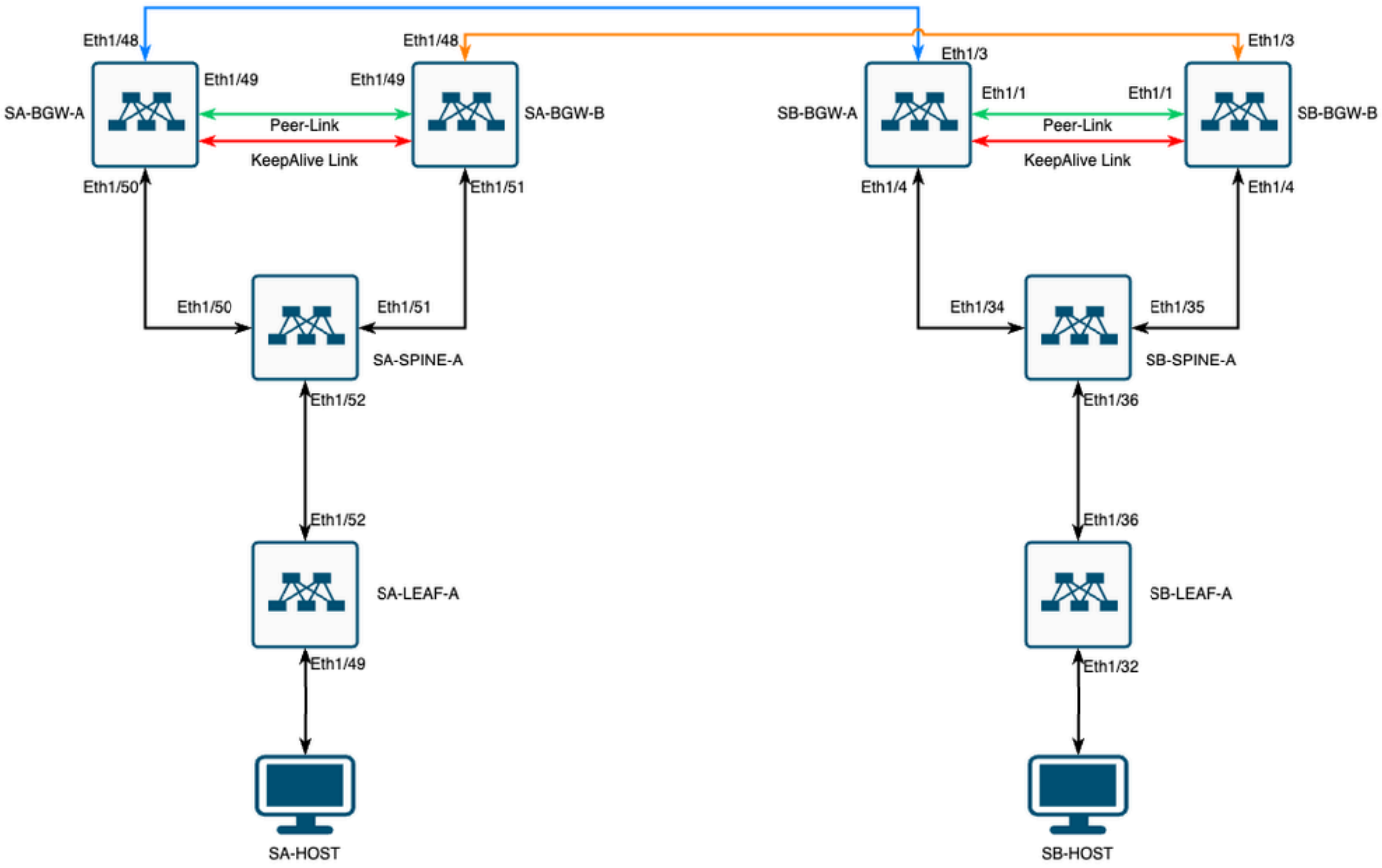
Die Informationen in diesem Dokument basieren auf den folgenden Software- und Hardwareversionen:

- Cisco Nexus 9000
- NX-OS-Version 10.3(4a):

Die Informationen in diesem Dokument beziehen sich auf Geräte in einer speziell eingerichteten Testumgebung. Alle Geräte, die in diesem Dokument benutzt wurden, begannen mit einer gelöschten (Nichterfüllungs) Konfiguration. Wenn Ihr Netzwerk in Betrieb ist, stellen Sie sicher, dass Sie die möglichen Auswirkungen aller Befehle kennen.

Konfigurieren

Netzwerkdiagramm



VXLAN MultiSite mit CloudSec in quadratischer Topologie

Einzelheiten der Topologie

- VXLAN-EVPN-Fabric mit zwei Standorten
- Beide Standorte sind mit vPC Border Gateways konfiguriert.
- Endpunkte werden in VLAN 1100 gehostet.
- Die Grenz-Gateways der einzelnen Standorte weisen über die SVI-Schnittstelle Vlan3600 eine IPv4-iBGP-Nachbarschaft zueinander auf.
- Grenz-Gateways an einem Standort haben eine eBGP-IPv4-Nachbarschaft und verfügen am anderen Standort nur über ein direkt verbundenes Grenz-Gateway.
- Die Grenz-Gateways von Standort A verfügen über eine eBGP L2VPN-EVPN-Nachbarschaft mit Grenz-Gateways von Standort B.

Adressierungsplan

Die IP-Adressen in der Tabelle werden während der Konfiguration verwendet:

	STANDORT A	STANDORT B				
Geräterolle	Schnittstellen-ID	Physische Int-IP	RID-Schleife IP	NVE-Loop-IP	MSITE-VIP	Bac
BLATT	Eth1/52	192.168.1.1/30	192.168.2.1/32	192.168.3.1/32	–	
WIRBELSÄULE	Eth1/52	192.168.1.2/30			–	
Eth1/50	192.168.1.5/30	192.168.2.2/32	–	–	–	E
Eth1/51	192.168.1.9/30			–		E
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.
Eth1/48	10.12.10.1/30		192.168.3.254/32			E
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.
Eth1/48	10.12.10.5/30		192.168.3.254/32			E

Konfigurationen

- Beachten Sie, dass in diesem Leitfaden nur standortübergreifende Konfigurationen dargestellt werden. Die vollständige Konfiguration können Sie der offiziellen Dokumentation von Cisco für VXLAN Cisco Nexus Serie 9000 NX-OS VXLAN-Konfigurationsanleitung, Version 10.3(x), entnehmen.

Um CloudSec zu aktivieren, muss der dci-advertise-pip Befehl unter dem Border-Gateway "evpn multisite" konfiguriert werden:

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

BGP-Konfiguration

Diese Konfiguration ist standortspezifisch.

SA-BGW-A und SA-BGW-B	SB-BGW-A und 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

- Mit dem Befehl **maximum-path** können mehrere eBGP L2VPN-EVPN-Pfade vom Nachbarn empfangen werden.
- Der Befehl **additional-path** weist den BGP-Prozess an, anzukündigen, dass das Gerät zusätzliche Pfade senden/empfangen kann.

Für alle L3VNI-VRFs auf Grenz-Gateways muss Multipath ebenfalls konfiguriert werden:

SA-BGW-A und SA-BGW-B	SB-BGW-A und 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

Konfiguration der Tunnelverschlüsselung

Diese Konfiguration muss für alle Border Gateways identisch sein:

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

Diese Konfiguration ist standortspezifisch. Der tunnel-encryption Befehl darf nur auf die Schnittstelle angewendet werden, die über den Befehl evpn multisite dci-trackingverfügt.

SA-BGW-A und SA-BGW-B	SB-BGW-A und 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

Nach der Aktivierung der Tunnelverschlüsselung werden dem lokalen Loopback weitere Attribute hinzugefügt, während Routen an den Nachbarn gemeldet werden. Allen eBGP-IPv4-Unicast-Nachbarn muss dieses Attribut angezeigt werden:

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

Für den Routing-Typ 2 gibt es auch ein neues 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#

Überprüfung

Vor dem Aktivieren von CloudSec sollten Sie überprüfen, ob das Setup ohne CloudSec ordnungsgemäß funktioniert:

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

Nach der CloudSec-Konfiguration muss auch der Endpunkt auf der SA erfolgreich einen Ping an den Endpunkt auf Standort B senden. In einigen Fällen kann der Ping-Test jedoch erfolglos sein. Es hängt davon ab, welcher Cloudsec-Peer vom lokalen Gerät zum Senden von Cloudsec-verschlüsseltem Datenverkehr ausgewählt wurde.

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

Fehlerbehebung

Überprüfen Sie die lokale ARP-Tabelle auf dem Quellendpunkt:

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

Diese Ausgabe bestätigt, dass der BUM-Datenverkehr weitergeleitet wird und die Kontrollebene funktioniert. Der nächste Schritt ist die Überprüfung des Tunnel-Verschlüsselungsstatus:

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

Diese Ausgabe zeigt, dass die CloudSec-Sitzung eingerichtet ist. Als Nächstes können Sie unbegrenzte Pings auf SA-HOST-A ausführen:

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

Von diesem Punkt an müssen Sie Geräte an Standort A überprüfen und feststellen, ob der Datenverkehr diese Geräte erreicht. Sie können diese Aufgabe mit ELAM auf allen Geräten entlang des Pfads am Standort A durchführen. Wenn Sie in-select den Standardwert 6 bis 9 ändern, können Sie die Übereinstimmung anhand der inneren Header festlegen. Weitere Informationen zu ELAM finden Sie unter diesem Link: Nexus 9000 Cloud Scale ASIC (Tahoe) NX-OS ELAM.

ELAM auf SA-LEAF-A

Im Produktionsnetzwerk sind mehr als ein SPINE-Gerät vorhanden. Um zu verstehen, an welches Spine der Datenverkehr gesendet wurde, müssen Sie zuerst ein ELAM auf LEAF starten. Trotz der in-select 9 Verwendung muss auf dem mit der Quelle verbundenen LEAF der äußere IPv4-Header verwendet werden, da der Datenverkehr, der diesen LEAF erreicht hat, nicht VXLAN-verschlüsselt ist. In realen Netzwerken kann es schwierig sein, das von Ihnen erzeugte Paket genau zu erfassen. In solchen Fällen können Sie Ping mit einer bestimmten Länge ausführen und den Pkt-len-Header verwenden, um Ihr Paket zu identifizieren. Standardmäßig hat ein icmp-Paket eine Länge von 64 Byte. Plus 20 Byte IP-Header, was zusammenfassend 84 Byte PKT Len ergab:

<#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=

An dieser Ausgabe können Sie sehen, dass der Datenverkehr SA-LEAF-A erreicht und über die Schnittstelle Ethernet1/52 weitergeleitet wird, die von der Topologie aus mit SA-SPINE-A verbunden ist.

ELAM auf SA-SPINE-A

Auf SPINE wird der Pkt-Len-Wert größer sein, da der 50-Byte-VXLAN-Header ebenfalls hinzugefügt wurde. Standardmäßig kann SPINE auf internen Headern ohne vxlan-parse oder feature nv overlay nicht übereinstimmen. Daher müssen Sie den vxlan-parse enable Befehl auf SPINE verwenden:

<#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 sendet den Datenverkehr je nach Ausgabe an SA-BGW-A.

ELAM auf 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

Laut Ausgabe von SA-BGW-A wurde der Datenverkehr über Ethernet 1/48 an SB-BGW-A weitergeleitet. Im nächsten Schritt überprüfen Sie 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

Laut Ausgabe von SB-BGW-A wurde ELAM nicht einmal ausgelöst. Das bedeutet, dass entweder der SB-BGW-B die Pakete empfängt und sie nicht richtig entschlüsseln und analysieren kann oder gar nicht. Um zu verstehen, was mit dem Cloudsec-Datenverkehr passiert ist, können Sie erneut ein ELAM auf SB-BGW-A ausführen. Der Triggerfilter muss jedoch auf die äußere IP-Adresse gesetzt werden, die für Cloudsec verwendet wird, da der innere Header des verschlüsselten Cloudsec-Datenpakets nicht sichtbar ist. Aus der vorherigen Ausgabe wissen Sie, dass der SA-BGW-A den Datenverkehr behandelt hat, was bedeutet, dass SA-BGW-A den Datenverkehr mit Cloudsec verschlüsselt. Sie können also NVE IP von SA-BGW-A als Triggerfilter für ELAM verwenden. Die Länge der VXLAN-verschlüsselten ICMP-Pakete der vorherigen Ausgänge beträgt 134 Byte. Plus 32 Byte CloudSec-Header in der Zusammenfassung ergibt 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)#

Aus dieser Ausgabe können Sie sehen, dass der CloudSec-Datenverkehr basierend auf der Routing-Tabelle über die Schnittstelle Ethernet1/4 an den SB-BGW-B weitergeleitet wird. Gemäß Cisco Nexus Serie 9000 NX-OS VXLAN Configuration Guide, Release 10.3(x) Richtlinien und Einschränkungen:

- CloudSec-Datenverkehr, der für den Switch bestimmt ist, muss über die DCI-Uplinks in den Switch gelangen.

Laut dem Abschnitt zur Unterstützung von CloudSec durch vPC Border Gateway im selben Leitfaden sind die BGP-Pfadattribute beider vPC

BGWs identisch, wenn der vPC BGW die PIP-Adresse des vPC BGWs erlernt und auf der DCI-Seite ankündigt. Daher können die zwischengeschalteten DCI-Knoten den Pfad vom vPC-BGW wählen, der nicht die PIP-Adresse besitzt. In diesem Szenario wird die MCT-Verbindung für verschlüsselten Datenverkehr vom Remote-Standort verwendet. In diesem Fall wird jedoch die Schnittstelle zum SPINE verwendet. Dennoch verfügen die BGWs über eine OSPF-Adjacency über die BackUp-SVI.

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

Grund des Problems und Behebung

Der Grund hierfür sind die OSPF-Kosten der SVI-Schnittstelle. Standardmäßig beträgt die Referenzbandbreite für automatische Kosten in NX-OS 40 G. SVI-Schnittstellen haben eine Bandbreite von 1 Gbit/s, während die physische Schnittstelle eine Bandbreite von 10 Gbit/s aufweist:

<#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 diesem Fall kann das Problem durch eine administrative Änderung der Kosten für SVI gelöst werden. Die Feinabstimmung muss an allen Border Gateways erfolgen.

<#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

Informationen zu dieser Übersetzung

Cisco hat dieses Dokument maschinell übersetzen und von einem menschlichen Übersetzer editieren und korrigieren lassen, um unseren Benutzern auf der ganzen Welt Support-Inhalte in ihrer eigenen Sprache zu bieten. Bitte beachten Sie, dass selbst die beste maschinelle Übersetzung nicht so genau ist wie eine von einem professionellen Übersetzer angefertigte. Cisco Systems, Inc. übernimmt keine Haftung für die Richtigkeit dieser Übersetzungen und empfiehlt, immer das englische Originaldokument (siehe bereitgestellter Link) heranzuziehen.