Solución de problemas de VXLAN multisitio con CloudSec en topología cuadrada

Contenido

Introducción

Este documento describe la configuración y solución de problemas de VXLAN Multisite con CloudSec entre gateways de borde conectados en topología cuadrada.

Prerequisites

Requirements

Cisco recomienda que esté familiarizado con estos temas:

- Software Nexus NXOS ©.
- Tecnología VXLAN EVPN.
- Protocolos de ruteo BGP y OSPF.

Componentes Utilizados

La información de este documento se basa en las siguientes versiones de software y hardware:

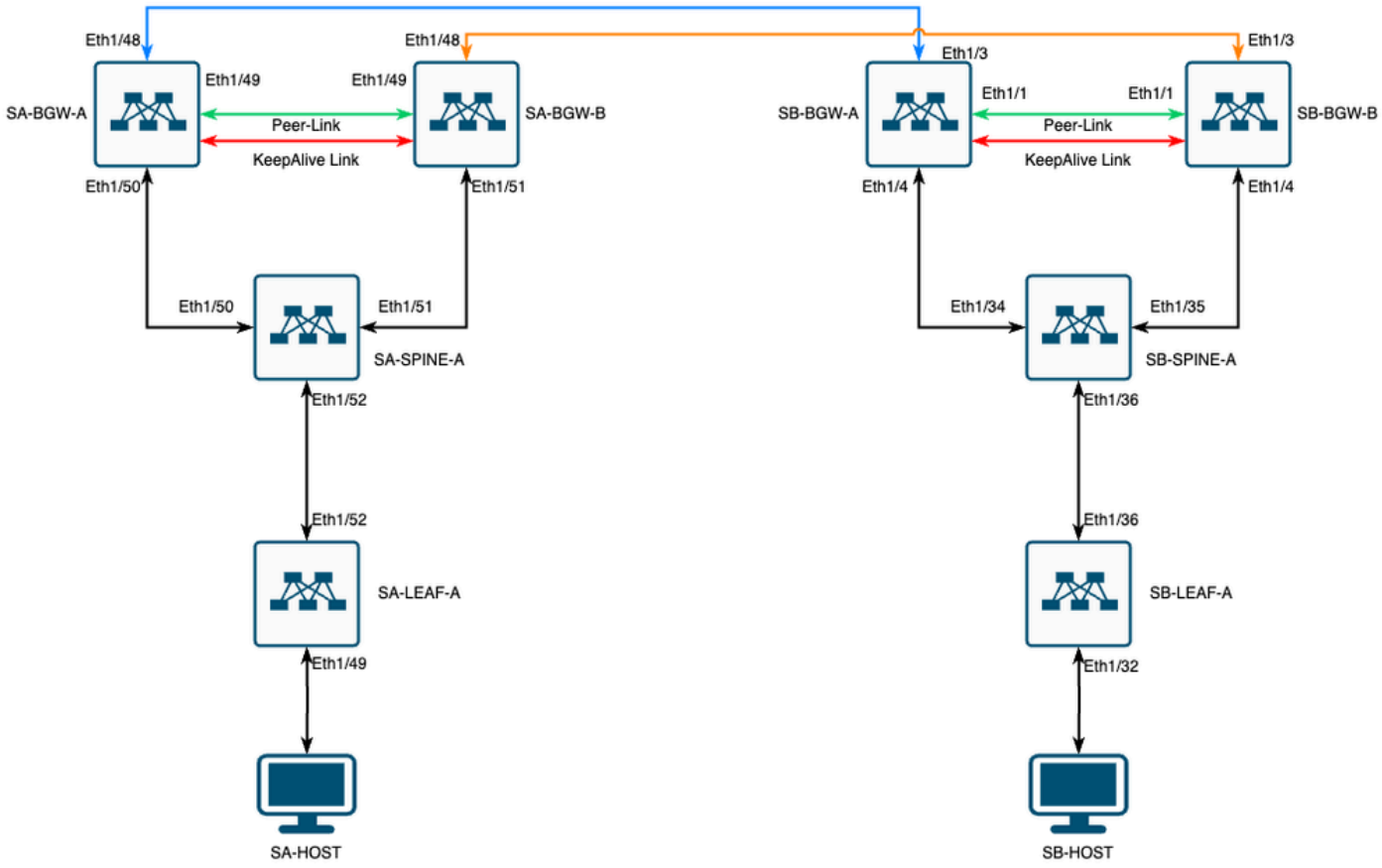
- Cisco Nexus 9000.

- NXOS versión 10.3(4a).

La información que contiene este documento se creó a partir de los dispositivos en un ambiente de laboratorio específico. Todos los dispositivos que se utilizan en este documento se pusieron en funcionamiento con una configuración verificada (predeterminada). Si tiene una red en vivo, asegúrese de entender el posible impacto de cualquier comando.

Configurar

Diagrama de la red



VXLAN MultiSite con CloudSec en topología cuadrada

Detalles de la topología

- Fabric VPN VXLAN multisitio de dos sitios.
- Ambos sitios se configuran con gateways de frontera vPC.
- Los terminales están alojados en VLAN 1100.
- Los gateways de borde en cada sitio tienen una vecindad iBGP IPv4 entre sí a través de la interfaz SVI Vlan3600.
- Los gateways de borde en un sitio tienen vecindad IPv4 eBGP sólo con gateway de borde conectado directamente en el otro sitio.
- Las gateways de borde en el sitio A tienen vecindad de EVPN L2VPN eBGP con gateways de borde en el sitio B.

Plan de direccionamiento

Las direcciones IP de la tabla se utilizan durante la configuración:

	SITIO A	SITIO B				
Función de dispositivo	ID de interfaz	IP de Int física	IP de bucle RID	NVE Loop IP	MSITE-VIP	IP de S resp
HOJA	Eth1/52	192.168.1.1/30	192.168.2.1/32	192.168.3.1/32	N/A	N/
COLUMNA VERTEBRAL	Eth1/52	192.168.1.2/30			N/A	
Eth1/50	192.168.1.5/30	192.168.2.2/32	N/A	N/A	N/A	Eth1
Eth1/51	192.168.1.9/30			N/A		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			Eth
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			Eth

Configuraciones

- Tenga en cuenta que en esta guía solo se muestra la configuración relacionada con varios sitios. Para la configuración completa, puede utilizar la guía de documentación oficial de Cisco para la Guía de configuración de VXLAN NX-OS Nexus de Cisco serie 9000, versión 10.3(x)

Para habilitar CloudSec, el dci-advertise-pip comando debe configurarse en la puerta de enlace de frontera multisitio de evpn:

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

configuración de BGP

Esta configuración es específica del sitio.

SA-BGW-A y SA-BGW-B	SB-BGW-A y 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

- El comando **maximum-path** permite recibir varias trayectorias EVPN L2VPN eBGP del vecino.
- El comando **additional-path** indica al proceso BGP que anuncie que el dispositivo es capaz de enviar/recibir trayectos adicionales

Para todos los L3VNI VRFs en gateways de borde, también se debe configurar el trayecto múltiple:

SA-BGW-A y SA-BGW-B	SB-BGW-A y 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

Configuración de encripción del túnel

Esta configuración debe ser la misma en todos los gateways de borde:

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

Esta configuración es específica del sitio. El tunnel-encryption comando debe aplicarse solamente a la interfaz que tiene el evpn multisite dci-trackingcomando.

SA-BGW-A y SA-BGW-B	SB-BGW-A y 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

Después de habilitar el encripción de túnel, se agregan atributos adicionales al loopback local mientras se anuncian las rutas al vecino y todos los vecinos de unidifusión IPv4 de eBGP deben ver este atributo:

<#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

Para el tipo de ruta 2 también hay un nuevo atributo:

<#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#

Verificación

Antes de habilitar cloudsec, es bueno comprobar si la configuración funciona correctamente sin él:

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

Después de la configuración de cloudsec, el terminal de SA debe hacer ping correctamente al terminal del sitio B. Pero, en algunos casos, el ping puede ser infructuoso. Depende del par de cloudsec seleccionado por el dispositivo local para enviar el tráfico cifrado de 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

Troubleshoot

Verifique la tabla ARP local en el punto final de origen:

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

Este resultado demuestra que, el tráfico BUM está pasando y el plano de control está funcionando. El siguiente paso es verificar el estado de encripción del túnel:

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

Este resultado muestra que se ha establecido la sesión de CloudSec. Como siguiente paso puede ejecutar un ping ilimitado en SA-HOST-A:

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

A partir de este punto, debe comprobar los dispositivos del sitio A y ver si el tráfico está llegando a estos dispositivos. Puede realizar esta tarea con ELAM en todos los dispositivos a lo largo de la ruta en el sitio A. El cambio in-select del valor predeterminado de 6 a 9 permite que coincidan en función de los encabezados internos. Puede obtener más información sobre ELAM en este enlace: ASIC de ampliación de nube Nexus 9000 (Tahoe) NX-OS ELAM.

ELAM en SA-LEAF-A

En la red de producción existen más de un dispositivo SPINE. Para comprender a qué columna se envió el tráfico, primero debe tomar un ELAM en LEAF. A pesar de que se in-select 9 utiliza, en el LEAF conectado con el origen, se debe utilizar el encabezado ipv4 externo, ya que el tráfico que llega a este LEAF no está cifrado con VXLAN. En una red real, puede ser difícil capturar el paquete exacto que generó. En estos casos, puede ejecutar ping con una longitud específica y utilizar el encabezado Pkt len para identificar su paquete. De forma predeterminada, el paquete icmp tiene una longitud de 64 bytes. Más 20 bytes de encabezado IP, que en resumen le dio 84 bytes 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=

A partir de esta salida, puede ver que el tráfico se alcanza con SA-LEAF-A y se reenvía fuera de la interfaz Ethernet1/52, que está conectada con SA-SPINE-A desde la topología.

ELAM en SA-SPINE-A

En SPINE el valor de Pkt Len va a ser mayor, ya que el encabezado VXLAN de 50 bytes también se agregó. De forma predeterminada, SPINE

no puede coincidir en encabezados internos sin vxlan-parse o feature nv overlay . Por lo tanto, debe utilizar el vxlan-parse enable comando en 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 envía el tráfico hacia SA-BGW-A según la salida.

ELAM en 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

Según la salida de SA-BGW-A, el tráfico se desvió por Ethernet1/48 hacia SB-BGW-A. El siguiente paso es comprobar en 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

Según el resultado de SB-BGW-A, ELAM ni siquiera se activó. Esto significa que SB-BGW-B recibe los paquetes y no puede descifrarlos y analizarlos correctamente, o no los recibe en absoluto. Para comprender lo que sucedió con el tráfico de cloudsec, puede ejecutar un ELAM en SB-BGW-A nuevamente, pero el filtro del disparador debe configurarse en la dirección IP externa que se utiliza para cloudsec, ya que no hay manera de ver el encabezado interno del paquete de tránsito cifrado de cloudsec. De la salida anterior que conoce, se desprende que SA-BGW-A manejó el tráfico, lo que significa que SA-BGW-A cifra el tráfico con cloudsec. Por lo tanto, puede utilizar la IP NVE de SA-BGW-A como un filtro de activación para ELAM. De las salidas anteriores, la longitud del paquete ICMP cifrado VXLAN es de 134 bytes. Además, el encabezado cloudsec de 32 bytes en resumen le proporciona 166 bytes:

<#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)#

En esta salida, puede ver que el tráfico de cloudsec se reenvía hacia SB-BGW-B a través de la interfaz Ethernet1/4, según la tabla de ruteo.

Según la Guía de configuración de NX-OS VXLAN para Nexus de Cisco serie 9000, las directrices y limitaciones de la versión 10.3(x):

- El tráfico de CloudSec destinado al switch debe entrar en el switch a través de los enlaces ascendentes de DCI.

Según la sección Compatibilidad con vPC Border Gateway para Cloudsec de la misma guía, si vPC BGW aprende la dirección PIP de vPC BGW del par y se anuncia en el lado DCI, los atributos de ruta BGP de ambos vPC BGW serán iguales. Por lo tanto, los nodos intermedios DCI pueden terminar eligiendo la trayectoria de vPC BGW que no posee la dirección PIP. En esta situación, el enlace MCT se utiliza para el tráfico cifrado procedente del sitio remoto. Pero en este caso, se utiliza la interfaz hacia la COLUMNA, a pesar de eso, los BGW también tienen una

adyacencia OSPF a través de la SVI de respaldo.

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 y solución

La razón es el costo OSPF de la interfaz SVI. De forma predeterminada, en NXOS el ancho de banda de referencia de coste automático es de 40 G. Las interfaces SVI tienen un ancho de banda de 1 Gbps, mientras que la interfaz física tiene un ancho de banda de 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

En tal caso, el cambio administrativo del costo de SVI puede resolver el problema. El ajuste debe realizarse en todos los gateways de borde.

<#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

Acerca de esta traducción
Cisco ha traducido este documento combinando la traducción automática y los recursos humanos a fin de ofrecer a nuestros usuarios en todo el mundo contenido en su propio idioma.

Tenga en cuenta que incluso la mejor traducción automática podría no ser tan precisa como la proporcionada por un traductor profesional.

Cisco Systems, Inc. no asume ninguna responsabilidad por la precisión de estas traducciones y recomienda remitirse siempre al documento original escrito en inglés (insertar vínculo URL).