Identificar e Solucionar Problemas de VXLAN Multisite com CloudSec na Topologia Quadrada

Contents

Introdução

Este documento descreve a configuração de vários locais VXLAN e a solução de problemas com o CloudSec entre gateways de borda conectados na topologia quadrada.

Pré-requisitos

Requisitos

A Cisco recomenda que você esteja familiarizado com estes tópicos:

- Software Nexus NXOS ©.
- Tecnologia VXLAN EVPN.
- BGP e protocolos de roteamento OSPF.

Componentes Utilizados

As informações neste documento são baseadas nas seguintes versões de software e hardware:

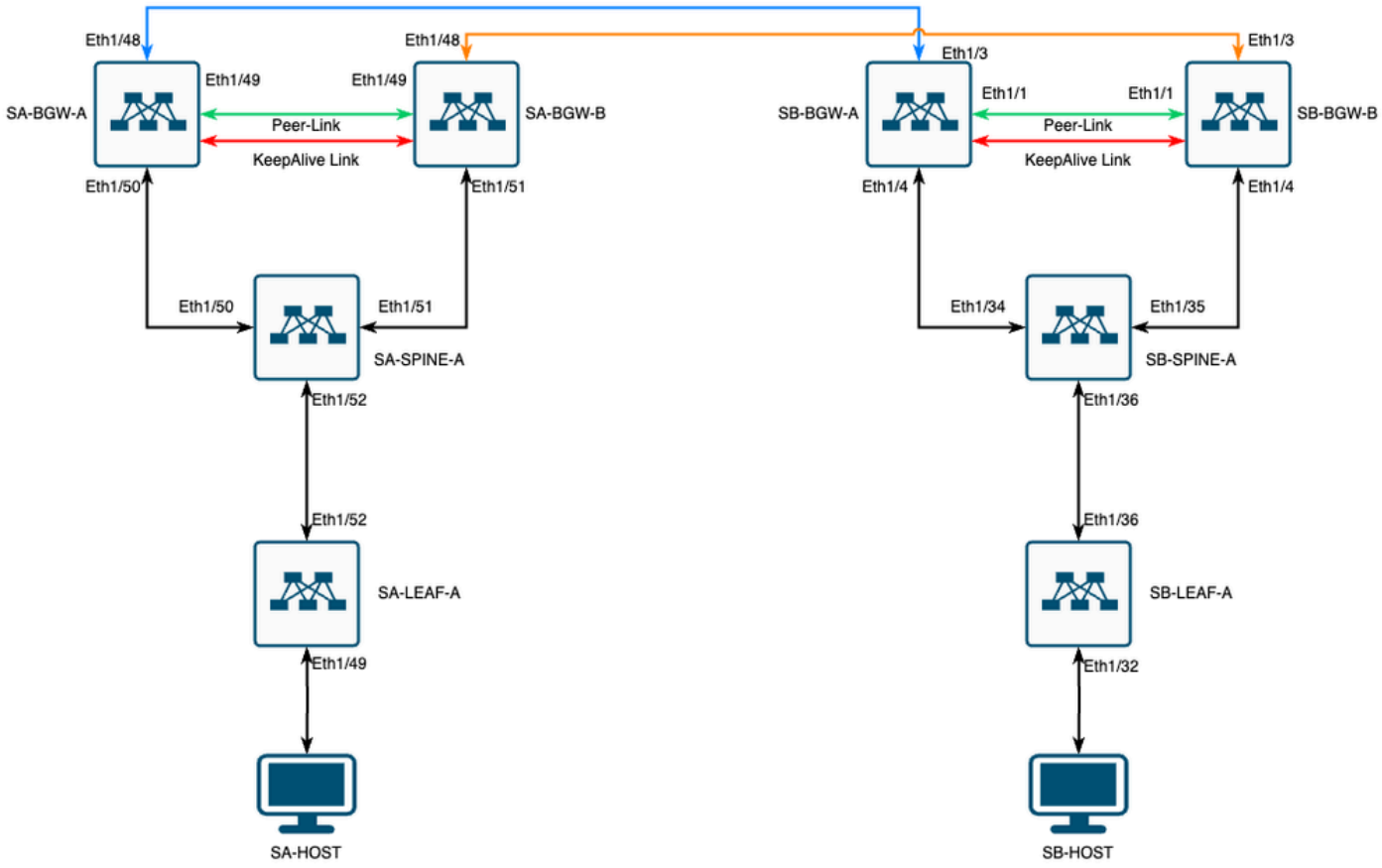
- Cisco Nexus 9000

- NXOS versão 10.3(4a).

As informações neste documento foram criadas a partir de dispositivos em um ambiente de laboratório específico. Todos os dispositivos utilizados neste documento foram iniciados com uma configuração (padrão) inicial. Se a rede estiver ativa, certifique-se de que você entenda o impacto potencial de qualquer comando.

Configurar

Diagrama de Rede



VXLAN MultiSite com CloudSec na topologia quadrada

Detalhes da topologia

- Estrutura EVPN VXLAN de vários locais de dois locais.
- Ambos os locais são configurados com gateways de borda vPC.
- Os endpoints são hospedados na VLAN 1100.
- Os gateways de borda em cada site têm vizinhança IPv4 iBGP entre si na interface Vlan3600 do SVI.
- Os gateways de borda em um site têm vizinhança IPv4 eBGP somente com gateway de borda diretamente conectado no outro site.
- Os gateways de borda no site A têm vizinhança EVPN L2VPN eBGP com gateways de borda no site B.

Plano de endereçamento

Os endereços IP na tabela são usados durante a configuração:

	LOCAL A	LOCAL B				
Função do dispositivo	ID da interface	IP Int Físico	IP de loop RID	IP de loop NVE	MSITE-VIP	IP do SV backu
FOLHA	Eth1/52	192.168.1.1/30	192.168.2.1/32	192.168.3.1/32	N/A	N/A
COLUNA	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	Eth1/3
Eth1/51	192.168.1.9/30			N/A		Eth1/3
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
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
Eth1/48	10.12.10.5/30		192.168.3.254/32			Eth1/3

Configurações

- Observe que neste guia somente a configuração relacionada a vários sites é mostrada. Para obter a configuração completa, você pode usar o guia de documentação oficial da Cisco para o Guia de configuração de VXLAN do Cisco Nexus 9000 Series NX-OS VXLAN, versão 10.3(x)

Para habilitar o CloudSec, o dci-advertise-pip comando deve ser configurado no gateway de borda de vários sites da vpn:

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

configuração de BGP

Essa configuração é específica do site.

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

- O comando **maximum-path** permite receber vários caminhos EVPN L2VPN eBGP do vizinho.
- O comando **additional-path** instrui o processo BGP para anunciar que o dispositivo é capaz de enviar/receber caminhos adicionais

Para todos os VRFs L3VNI nos gateways de borda, o multipath também deve ser configurado:

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

Configuração de criptografia de túnel

Essa configuração deve ser a mesma em todos os gateways de borda:

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

Essa configuração é específica do site. O comandotunnel-encryption deve ser aplicado somente à interface que tem o 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

Depois de ativar a criptografia de túnel, atributos adicionais são adicionados ao loopback local ao anunciar rotas para o vizinho e todos os vizinhos unicast IPv4 do eBGP devem 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 o tipo de rota 2, também há um novo 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#

Verificar

Antes de habilitar o cloudsec, é bom verificar se a configuração está funcionando bem sem ele:

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

Após a configuração de cloudsec também, o endpoint em SA deve efetuar ping com êxito no endpoint no site B. Mas, em alguns casos, o ping pode ser malsucedido. Depende de qual peer cloudsec foi selecionado pelo dispositivo local para enviar o tráfego criptografado 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

Troubleshooting

Verifique a tabela ARP local no ponto final de origem:

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,

Essa saída comprova que o tráfego de BUM está passando e o plano de controle está funcionando. A próxima etapa é verificar o status da criptografia do túnel:

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

Esta saída mostra que a sessão do CloudSec foi estabelecida. Como próxima etapa, você pode executar ping ilimitado em SA-HOST-A:

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

A partir desse ponto, você deve verificar os dispositivos no site A e ver se o tráfego está chegando a esses dispositivos. Você pode realizar essa tarefa com o ELAM em todos os dispositivos ao longo do caminho no site A. Alterar in-select do valor padrão de 6 para 9 permite que o faça a correspondência com base nos cabeçalhos internos. Leia mais sobre o ELAM neste link: Nexus 9000 Cloud Scale ASIC (Tahoe) NX-OS ELAM.

ELAM no SA-LEAF-A

Na rede de produção, existem mais de um dispositivo SPINE. Para entender para qual coluna o tráfego foi enviado, você deve primeiro obter um ELAM no LEAF. Apesar do in-select 9 usado, na FOLHA conectada à origem, o cabeçalho ipv4 externo deve ser usado, já que o tráfego que chega a essa FOLHA não é criptografado por VXLAN. Em uma rede real, pode ser difícil capturar o pacote exato gerado. Nesses casos, você pode executar o ping com um comprimento específico e usar o cabeçalho Pkt len para identificar seu pacote. Por padrão, o pacote icmp tem 64 bytes de comprimento. Mais 20 bytes de cabeçalho IP, que em resumo lhe deu 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 dessa saída, você pode ver que o tráfego é alcançado pelo SA-LEAF-A e encaminhado pela interface Ethernet1/52, que está conectada ao SA-SPINE-A a partir da topologia.

ELAM em SA-SPINE-A

Em SPINE, o valor de Pkt Len será maior, já que o cabeçalho VXLAN de 50 bytes também foi adicionado. Por padrão, SPINE não pode corresponder em cabeçalhos internos sem vxlan-parse ou feature nv overlay . Portanto, você deve usar o vxlan-parse enable comando no 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 envia tráfego para o SA-BGW-A de acordo com a saída.

ELAM no 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

De acordo com a saída do SA-BGW-A, o tráfego saiu da Ethernet1/48 em direção ao SB-BGW-A. A próxima etapa é verificar o 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

De acordo com a saída do SB-BGW-A, o ELAM nem sequer foi acionado. Isso significa que o SB-BGW-B está recebendo os pacotes e não está sendo capaz de descriptografá-los e analisá-los corretamente, ou não os recebe de forma alguma. Para entender o que aconteceu com o tráfego de cloudsec, você pode executar um ELAM no SB-BGW-A novamente, mas o filtro de gatilho deve ser definido como endereço IP externo que é usado para cloudsec, pois não há como ver o cabeçalho interno do pacote de trânsito criptografado do cloudsec. A partir da saída anterior, você sabe, que o SA-BGW-A tratou o tráfego, o que significa que o SA-BGW-A criptografa o tráfego com o cloudsec. Assim, você pode usar o IP NVE de SA-BGW-A como um filtro acionador para ELAM. A partir das saídas anteriores, o comprimento do pacote ICMP criptografado da VXLAN é de 134 bytes. Além disso, o cabeçalho cloudsec de 32 bytes em resumo fornece 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)#

A partir dessa saída, você pode ver que o tráfego de cloudsec é encaminhado para o SB-BGW-B através da interface Ethernet1/4, com base na tabela de roteamento. De acordo com o Guia de configuração do Cisco Nexus 9000 Series NX-OS VXLAN, versão 10.3(x), diretrizes e limitações:

- O tráfego do CloudSec destinado ao switch deve entrar no switch através dos uplinks DCI.

De acordo com a seção Suporte do vPC Border Gateway para Cloudsec do mesmo guia, se o BGW do vPC aprender o endereço IP do BGW do vPC e anunciar no lado do DCI, os atributos de caminho do BGP de ambos os BGW do vPC serão os mesmos. Portanto, os nós intermediários de DCI podem acabar escolhendo o caminho do vPC BGW que não possui o endereço PIP. Neste cenário, o link MCT é usado para tráfego criptografado proveniente do site remoto. Mas, nesse caso, a interface em direção ao SPINE é usada, apesar disso, os BGWs também têm uma adjacência OSPF através do SVI de 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 do problema e correção

O motivo é o custo OSPF da interface SVI. Por padrão, a largura de banda de referência de custo automático do NXOS é 40G. As interfaces SVI têm uma largura de banda de 1 Gbps, enquanto a interface física tem uma largura 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

Nesse caso, a alteração administrativa do custo do SVI pode resolver o problema. O ajuste deve ser feito em todos os gateways de borda.

<#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

Sobre esta tradução
A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês (link fornecido) seja sempre consultado.