정방형 토폴로지의 CloudSec으로 멀티사이트 VXLAN 문제 해결

목차

소개

이 문서에서는 정사각형 토폴로지에 연결된 경계 게이트웨이 간에 CloudSec을 사용한 VXLAN 멀티사이트 컨피그레이션 및 트러블슈팅에 대해 설명합니다.

사전 요구 사항

요구 사항

Cisco에서는 다음 항목에 대해 숙지할 것을 권장합니다.

- Nexus NXOS © Software.
- VXLAN EVPN 기술.
- BGP 및 OSPF 라우팅 프로토콜.

사용되는 구성 요소

이 문서의 정보는 다음 소프트웨어 및 하드웨어 버전을 기반으로 합니다.

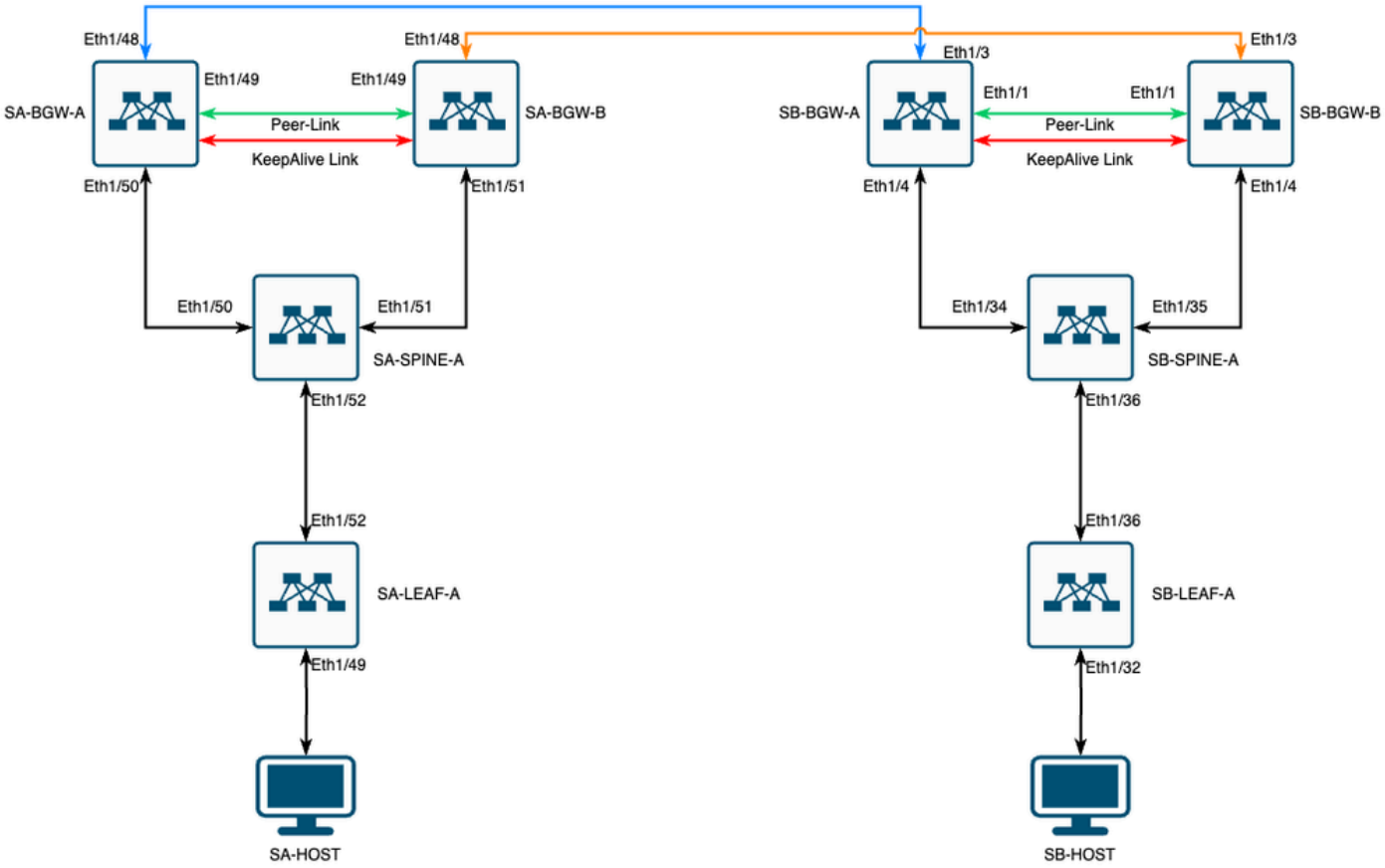
- Cisco Nexus 9000.

- NXOS 버전 10.3(4a).

이 문서의 정보는 특정 랩 환경의 디바이스를 토대로 작성되었습니다. 이 문서에 사용된 모든 디바이스는 초기화된(기본) 컨피그레이션으로 시작되었습니다. 현재 네트워크가 작동 중인 경우 모든 명령의 잠재적인 영향을 미리 숙지하시기 바랍니다.

구성

네트워크 다이어그램



정사각형 토폴로지의 CloudSec을 사용하는 VXLAN MultiSite

토폴로지 세부 정보

- 2개 사이트 멀티사이트 VXLAN EVPN 패브릭.
- 두 사이트 모두 vPC Border Gateway로 구성됩니다.
- 엔드포인트는 VLAN 1100에서 호스팅됩니다.
- 각 사이트의 보더 게이트웨이는 SVI 인터페이스 Vlan3600을 통해 서로 간에 IPv4 iBGP 인접 관계가 있습니다.
- 한 사이트의 보더 게이트웨이는 다른 사이트의 직접 연결된 보더 게이트웨이와만 eBGP IPv4 네이버를 가집니다.
- 사이트 A의 보더 게이트웨이는 사이트 B의 보더 게이트웨이와 eBGP L2VPN EVPN 네이버를 갖습니다.

처리 계획

테이블의 IP 주소는 컨피그레이션 중에 사용됩니다.

	사이트 A	사이트 B				
디바이스 역할	인터페이스 ID	물리적 내부 IP	RID 루프 IP	NVE 루프 IP	사이트 VIP	SVI IP 백업
리프	Eth1/52	192.168.1.1/30	192.168.2.1/32	192.168.3.1/32	해당 없음	해당 없음
등뼈	Eth1/52	192.168.1.2/30			해당 없음	
Eth1/50	192.168.1.5/30	192.168.2.2/32	해당 없음	해당 없음	해당 없음	Eth1/34
Eth1/51	192.168.1.9/30			해당 없음		Eth1/35
BGW-A	Eth1/51	192.168.1.6/30	192.168.2.3/32	192.168.3.2/32	192.168.100.1/32	192.168.4.1/3
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	192.168.4.2/3
Eth1/48	10.12.10.5/30		192.168.3.254/32			Eth1/3

설정

- 이 가이드에서는 멀티사이트 관련 컨피그레이션만 표시됩니다. 전체 구성을 위해 VXLAN용 Cisco 공식 설명서 가이드, Cisco Nexus 9000 Series NX-OS VXLAN 구성 설명서, 릴리스 10.3(x)을 사용할 수 있습니다

CloudSec을 활성화하려면 evpn dci-advertise-pip multisite border-gateway 아래에서 명령을 구성해야 합니다.

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

BGP 컨피그레이션

이 컨피그레이션은 사이트별로 다릅니다.

SA-BGW-A 및 SA-BGW-B	SB-BGW-A 및 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

- **maximum-path** 명령을 사용하면 네이버에서 여러 eBGP L2VPN EVPN 경로를 수신할 수 있습니다.

- additional-**path 명령**은 BGP 프로세스에 디바이스에서 추가 경로를 전송/수신할 수 있음을 알립니다

경계 게이트웨이의 모든 L3VNI VRF에 대해 다중 경로도 구성해야 합니다.

SA-BGW-A 및 SA-BGW-B	SB-BGW-A 및 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

터널 암호화 컨피그레이션

이 컨피그레이션은 모든 보더 게이트웨이에서 동일해야 합니다.

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

이 컨피그레이션은 사이트별로 다릅니다. 명령tunnel-encryption은 명령이 있는 인터페이스에만 적용해야 evpn multisite dci-tracking합니다.

SA-BGW-A 및 SA-BGW-B	SB-BGW-A 및 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

터널 암호화를 활성화한 후 인접 디바이스 및 모든 eBGP IPv4 유니캐스트 인접 디바이스에 경로를 광고하는 동안 추가 특성이 로컬 루프백에 추가됩니다.

<#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
```

Route Type-2에는 새로운 특성도 있습니다.

```
<#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#
```

다음을 확인합니다.

cloudsec을 활성화하기 전에, 설정이 제대로 작동하는지 확인하는 것이 좋습니다.

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

Cloudsec 컨피그레이션도 마친 후에는 SA의 엔드포인트가 사이트 B의 엔드포인트에 성공적으로 ping해야 합니다. 그러나 경우에 따라 ping이 실패할 수 있습니다. 로컬 디바이스에서 cloudsec 암호화 트래픽을 전송하도록 선택한 cloudsec 피어에 따라 달라집니다.

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

문제 해결

소스 엔드포인트에서 로컬 ARP 테이블을 확인합니다.

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

이 출력은 BUM 트래픽이 통과하고 있으며 Control-Plane이 작동하고 있음을 입증합니다. 다음 단계는 tunnel-encryption 상태를 확인하는 것입니다.

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

이 출력은 CloudSec 세션이 설정되었음을 보여줍니다. 다음 단계로 SA-HOST-A에서 무제한 ping을 실행할 수 있습니다.

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

이 시점부터 사이트 A의 디바이스를 확인하고 트래픽이 이 디바이스에 도달하는지 확인해야 합니다. 사이트 A의 경로를 따라 모든 디바이스에서 ELAM을 사용하여 이 작업을 수행할 수 있습니다. 기본값 6 in-select 에서 9로 변경하면 내부 헤더를 기준으로 일치할 수 있습니다. ELAM에 대한 자세한 내용은 Nexus 9000 Cloud Scale ASIC(Tahoe) NX-OS ELAM 링크를 참조하십시오.

SA-LEAF-A의 ELAM

프로덕션 네트워크에는 둘 이상의 SPINE 디바이스가 있습니다. 트래픽이 전송된 스파인을 파악하려면 먼저 LEAF에서 ELAM을 사용해야 합니다. 이 in-select 9 를 사용하더라도 소스에 연결된 LEAF에서는 외부 ipv4 헤더를 사용해야 합니다. 이 LEAF에 도달하는 트래픽은 VXLAN으로 암호화되지 않기 때문입니다. 실제 네트워크에서는 사용자가 생성한 정확한 패킷을 포착하기 어려울 수 있습니다. 이러한 경우 특정 길이로 ping을 실행하고 패킷 헤더(Pkt len header)를 사용하여 패킷을 식별할 수 있습니다. 기본적으로 icmp 패킷의 길이는 64바이트입니다. IP 헤더의 20바이트를 더한 것으로 요약하면 84바이트 PKT Len입니다.

<#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=

이 출력에서 트래픽이 SA-LEAF-A에 도달하고 토폴로지에서 SA-SPINE-A에 연결된 인터페이스 Ethernet1/52로 전달됨을 확인할 수 있습니다.

SA-SPINE-A의 ELAM

SPINE에서는 50바이트 VXLAN 헤더도 추가되었으므로 Pkt Len 값이 더 커집니다. 기본적으로 SPINE은 또는 이 없는 내부 헤더에서 일치할 수 vxlan-parse 없습니다 feature nv overlay . 따라서 SPINE에서 다음 명령 vxlan-parse enable 을 사용해야 합니다.

<#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-A는 출력에 따라 SA-BGW-A로 트래픽을 전송합니다.

SA-BGW-A의 ELAM

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

SA-BGW-A의 출력에 따르면 트래픽은 Ethernet1/48에서 SB-BGW-A로 이동했습니다. 다음 단계는 SB-BGW-A를 확인하는 것입니다.

<#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

SB-BGW-A의 출력에 따르면 ELAM은 트리거조차 되지 않았다. 즉, SB-BGW-B가 패킷을 수신하고 있으며 패킷을 올바르게 해독하고 구문 분석할 수 없거나 패킷을 전혀 수신하지 않습니다. cloudsec 트래픽에 발생한 상황을 이해하려면 SB-BGW-A에서 ELAM을 다시 실행할 수 있지만, cloudsec에 사용되는 외부 IP 주소로 트리거 필터를 설정해야 합니다. cloudsec 암호화 트랜짓 패킷의 내부 헤더를 볼 수 있는 방법은 없기 때문입니다. 이전 출력에서 SA-BGW-A가 트래픽을 처리했음을 알 수 있습니다. 즉, SA-BGW-A가 cloudsec으로 트래픽을 암호화합니다. 따라서 SA-BGW-A의 NVE IP를 ELAM의 트리거 필터로 사용할 수 있습니다. 위의 출력에서 VXLAN 암호화 ICMP 패킷 길이는 134바이트입니다. 또한 요약에서 32바이트 cloudsec 헤더는 166바이트를 제공합니다.

<#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)#

이 출력에서 라우팅 테이블을 기반으로 cloudsec 트래픽이 인터페이스 Ethernet1/4를 통해 SB-BGW-B로 전달됨을 확인할 수 있습니다. Cisco Nexus 9000 Series NX-OS VXLAN 컨피그레이션 가이드, 릴리스 10.3(x) 가이드 및 제한 사항:

- 스위치로 향하는 CloudSec 트래픽은 DCI 업링크를 통해 스위치에 들어가야 합니다.

동일한 가이드의 vPC Border Gateway Support for Cloudsec 섹션에 따르면, vPC BGW가 피어 vPC BGW의 PIP 주소를 학습하고 DCI에 광고를 하는 경우 두 vPC BGW의 BGP 경로 특성이 모두 동일합니다. 따라서 DCI 중간 노드는 PIP 주소를 소유하지 않는 vPC BGW에서 경로를 선택할 수 있습니다. 이 시나리오에서는 MCT 링크가 원격 사이트에서 들어오는 암호화된 트래픽에 사용됩니다. 그러나 이 경우 SPINE으로의 인터페이스가 사용되지만, BGW도 백업 SVI를 통해 OSPF 인접성을 갖습니다.

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

문제 및 해결 이유

그 이유는 SVI 인터페이스의 OSPF 비용입니다. 기본적으로 NXOS 자동 비용 참조 대역폭은 40G입니다. SVI 인터페이스의 대역폭은 1Gbps이지만 물리적 인터페이스의 대역폭은 10Gbps입니다.

<#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

이러한 경우 SVI에 대한 비용의 관리 변경을 통해 문제를 해결할 수 있습니다. 모든 보더 게이트웨이에서 튜닝이 수행되어야 합니다.

<#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

이 번역에 관하여

Cisco는 전 세계 사용자에게 다양한 언어로 지원 콘텐츠를 제공하기 위해 기계 번역 기술과 수작업 번역을 병행하여 이 문서를 번역했습니다. 아무리 품질이 높은 기계 번역이라도 전문 번역가의 번역 결과물만큼 정확하지는 않습니다. Cisco Systems, Inc.는 이 같은 번역에 대해 어떠한 책임도 지지 않으며 항상 원본 영문 문서(링크 제공됨)를 참조할 것을 권장합니다.