SquareトポロジのCloudSecを使用したマルチサイトVXLANのトラブルシューティング

内容

はじめに

このドキュメントでは、Squareトポロジで接続されたボーダーゲートウェイ間のCloudSecを使用したVXLANマルチサイトの設定およびトラブルシューティングについて説明します。

前提条件

要件

次の項目について十分に理解しておくことをお勧めします。

- Nexus NXOS ©ソフトウェア
- VXLAN EVPNテクノロジー。
- BGPおよびOSPFルーティングプロトコル。

使用するコンポーネント

このドキュメントの情報は、次のソフトウェアとハードウェアのバージョンに基づくものです。

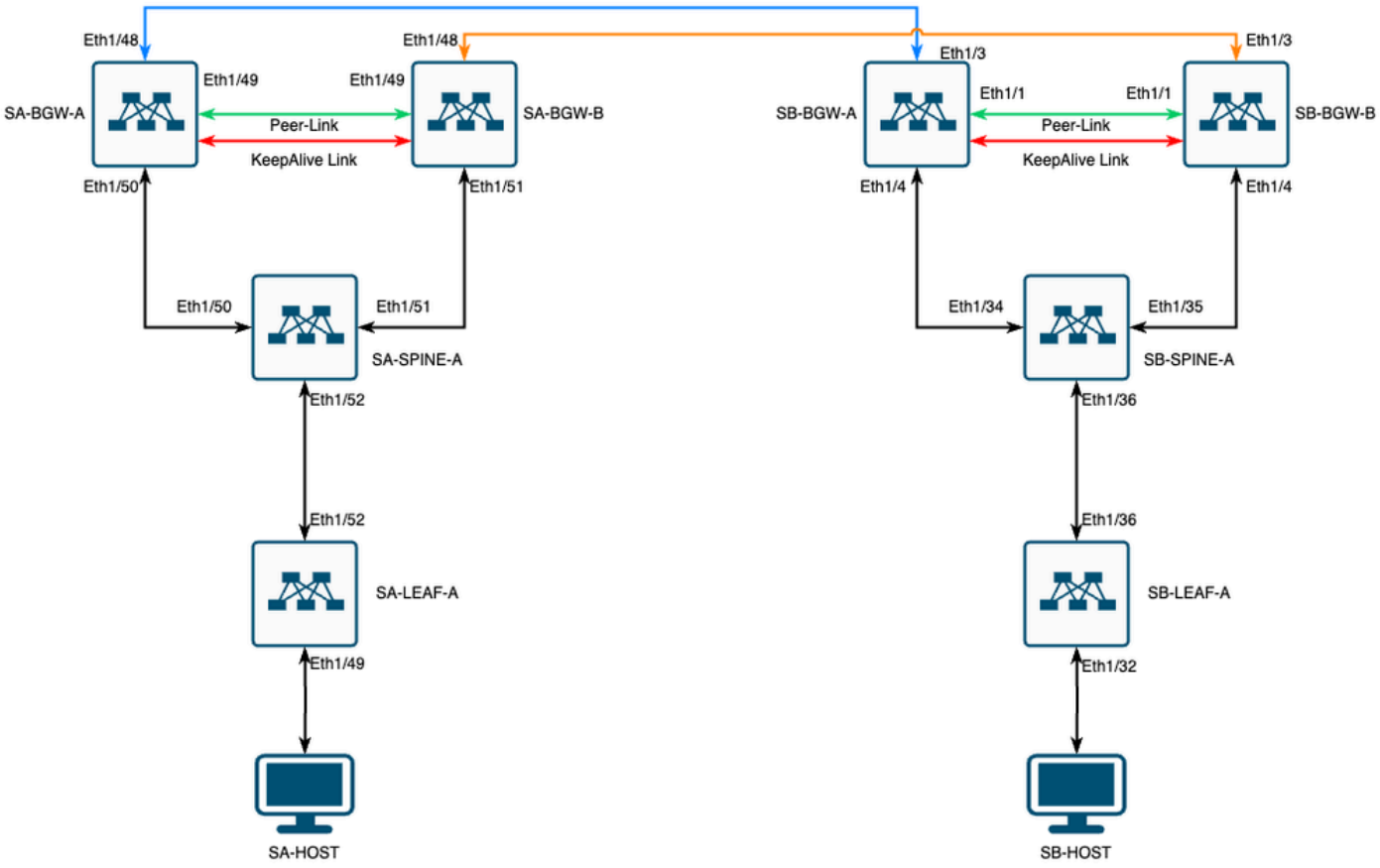
- Cisco Nexus 9000.

- NXOSバージョン10.3(4a)。

このドキュメントの情報は、特定のラボ環境にあるデバイスに基づいて作成されました。このドキュメントで使用するすべてのデバイスは、クリアな（デフォルト）設定で作業を開始しています。本稼働中のネットワークでは、各コマンドによって起こる可能性がある影響を十分確認してください。

設定

ネットワーク図



SquareトポロジのCloudSecを使用したVXLAN MultiSite

トポロジの詳細

- 2サイトマルチサイトVXLAN EVPNファブリック。
- 両方のサイトがvPCボーダーゲートウェイで設定されています。
- エンドポイントはVLAN 1100でホストされます。
- 各サイトのボーダーゲートウェイは、SVIインターフェイスVlan3600を介して相互にIPv4 iBGPネイバーシップを持ちます。
- 一方のサイトのボーダーゲートウェイは、他方のサイトの直接接続ボーダーゲートウェイとのみeBGP IPv4ネイバーシップを持ちます。
- サイトAのボーダーゲートウェイは、サイトBのボーダーゲートウェイとeBGP L2VPN EVPNネイバーシップを持ちます。

アドレス指定計画

テーブル内のIPアドレスは、設定時に使用されます。

	サイトA	サイトB				
デバイスロール	インターフェイスID	物理インターフェイスIP	RIDループIP	NVEループIP	マルチサイトVIP	バックアップSVI IP
リーフ	Eth1/52	192.168.1.1/30	192.168.2.1/32	192.168.3.1/32	N/A	N/A
スパイン	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	Eth1/34
Eth1/51	192.168.1.9/30			N/A		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

コンフィギュレーション

- このガイドでは、マルチサイト関連の設定だけが示されています。完全な設定については、VXLANに関するシスコの公式ドキュメントガイド『Cisco Nexus 9000シリーズNX-OS VXLANコンフィギュレーションガイド、リリース10.3(x)

CloudSecを有効にするには、evpn multisite border-gatewayの下でdci-advertise-pip コマンドを設定する必要があります。

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

BGPの設定

この設定はサイトに固有です。

SA-BGW-AおよびSA-BGW-B	SB-BGW-Aおよび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**コマンドにより、はネイバーから複数のeBGP L2VPN EVPNパスを受信できます。
- **additional-path**コマンドは、デバイスが追加パスを送受信できることをアドバタイズするようにBGPプロセスに指示します

ボーダーゲートウェイ上のすべてのL3VNI VRFでは、マルチパスも設定する必要があります。

SA-BGW-AおよびSA-BGW-B	SB-BGW-Aおよび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

トンネル暗号化設定

次の設定は、すべてのボーダーゲートウェイで同じである必要があります。

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

この設定はサイトに固有です。tunnel-encryptionコマンドは、evpn multisite dci-trackingコマンドが設定されているインターフェイスにのみ適用する必要があります。

SA-BGW-AおよびSA-BGW-B	SB-BGW-Aおよび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

tunnel-encryptionを有効にすると、ルートをネイバーおよびすべてのeBGP IPv4ユニキャストネイバーにアドバタイズする際に、ローカルループバックに追加のアトリビュートが追加されます。

<#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

ルートタイプ2には、新しい属性もあります。

<#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#

確認

cloudsecを有効にする前に、設定が正しく動作しているかどうかを確認することをお勧めします。

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の設定後も、SA上のエンドポイントはサイトBのエンドポイントに対してpingを正常に実行する必要があります。ただし、場合によってはpingが失敗することがあります。これは、ローカルデバイスがcloudsec暗号化トラフィックを送信するために選択したcloudsecピアによって異なります。

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

トラブルシュート

送信元エンドポイントのローカルARPテーブルを確認します。

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,

この出力は、BUMトラフィックが通過し、コントロールプレーンが機能していることを示しています。次のステップは、トンネ

ル暗号化のステータスを確認することです。

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

この出力は、CloudSecセッションが確立されていることを示しています。次の手順では、SA-HOST-Aに対して無制限のpingを実行できます。

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

この時点から、サイトAのデバイスをチェックして、トラフィックがこのデバイスに到達しているかどうかを確認する必要があります。サイトAのパスに沿ったすべてのデバイスでELAMを使用してこのタスクを実行できます。デフォルト値の6から9に変更す in-select ると、内側のヘッダーに基づいてが照合されます。ELAMの詳細については、Nexus 9000 Cloud Scale ASIC(Tahoe)NX-OS ELAMを参照してください。

SA-LEAF-AのELAM

実稼働ネットワークには、複数のSPINEデバイスが存在します。トラフィックがどのスパインに送信されたかを理解するには、まずLEAFでELAMを取得する必要があります。 in-select 9 が使用されているにもかかわらず、送信元に接続されているリーフでは、このリーフに到達するトラフィックはVXLAN暗号化されないため、外部ipv4ヘッダーを使用する必要があります。実際のネットワークでは、生成したパケットを正確に把握することは困難です。このような場合は、特定の長さのpingを実行し、Pkt lenヘッダーを使用してパケットを識別できます。デフォルトでは、icmpパケットの長さは64バイトです。さらに20バイトのIPヘッダーがあり、要約すると84バイトの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=

この出力から、トラフィックがSA-LEAF-Aに到達し、トポロジからSA-SPINE-Aに接続されているインターフェイスEthernet1/52か

ら転送されていることがわかります。

SA-SPINE-AのELAM

50バイトのVXLANヘッダーも追加されたため、SPINEではPkt Lenの値が大きくなります。デフォルトでは、 vxlan-parse または feature nv overlay を使用しないと、SPINEを内部ヘッダーで照合できません。したがって、SPINEでは vxlan-parse enable コマンドを使用する必要があります。

<#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は、出力に従ってSA-BGW-Aにトラフィックを送信します。

SA-BGW-AのELAM

<#root>

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-Aからの出力によると、トラフィックはEthernet1/48からSB-BGW-Aに向かっていました。次に、SB-BGW-Aで確認します。

<#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-Aからの出力によると、ELAMはトリガーさえされていません。これは、SB-BGW-Bがパケットを受信していて、正しく復号化および解析できないか、パケットをまったく受信しないことを意味します。cloudsecトラフィックで何が起こったかを理解するには、SB-BGW-Aで再度ELAMを実行できますが、cloudsec暗号化された中継パケットの内側のヘッダーを確認する方法がないため、トリガーフィルタはcloudsecで使用される外側のIPアドレスに設定する必要があります。前述の出力から、SA-BGW-Aがトラフィックを処理したことが分かります。つまり、SA-BGW-Aはトラフィックをcloudsecで暗号化します。したがって、SA-BGW-AのNVE IPをELAMのトリガーフィルタとして使用できます。前述の出力から、VXLAN暗号化ICMPパケット長は134バイトです。要約すると、32バイトのcloudsecヘッダーによって166バイトになります。

<#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-A

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 learn

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)#

この出力から、ルーティングテーブルに基づいて、インターフェイスEthernet1/4を介してCloudsecトラフィックがSB-BGW-Bに転送されていることがわかります。『Cisco Nexus 9000 Series NX-OS VXLAN Configuration Guide, Release 10.3(x)』のガイドラインと制限事項によれば、次のとおりです。

- スイッチ宛てのCloudSecトラフィックは、DCIアップリンクを介してスイッチに入る必要があります。

同じガイドの「vPC Border Gateway for Cloudsec」セクションによると、vPC BGWがピアvPC BGWのPIPアドレスを学習してDCI側でアドバタイズすると、両方のvPC BGWからのBGPパス属性が同じになります。したがって、DCI中間ノードは、PIPアドレスを持たないvPC BGWからのパスを選択することになります。このシナリオでは、リモートサイトからの暗号化トラフィックにMCTリンクが使用されます。ただし、この場合はスパインへのインターフェイスが使用されますが、BGWにもバックアップSVI経由のOSPFアジャセンシー関係があります。

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

問題と修正の理由

これは、SVIインターフェイスのOSPFコストが原因です。デフォルトでは、NXOSの自動コスト参照帯域幅は40Gです。SVIインターフェイスの帯域幅は1 Gbpsですが、物理インターフェイスの帯域幅は10 Gbpsです。

<#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

この場合、SVIのコストの管理上の変更によって問題を解決できます。すべてのボーダーゲートウェイで調整を行う必要があります。

<#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

翻訳について
シスコは世界中のユーザにそれぞれの言語でサポート コンテンツを提供するために、機械と人による翻訳を組み合わせて、本ドキュメントを翻訳しています。ただし、最高度の機械翻訳であっても、専門家による翻訳のような正確性は確保されません。シスコは、これら翻訳の正確性について法的責任を負いません。原典である英語版（リンクからアクセス可能）もあわせて参照することを推奨します。