マルチホーミングの設定

この章は、次の項で構成されています。

VXLAN EVPN マルチホーミングの概要

マルチホーミングの概要

Cisco Nexus プラットフォームは、vPC ベースのマルチホーミングをサポートします。このマルチホーミングでは、スイッチのペアが冗長性のために単一のデバイスとして機能し、両方のスイッチがアクティブ モードで機能します。VXLAN BGP EVPN 環境の Cisco Nexus 9000 シリーズ スイッチでは、レイヤ 2 マルチホーミングをサポートする 2 つのソリューションがあります。ソリューションは、従来の vPC（エミュレートまたは仮想 IP アドレス）と BGP EVPN 技術に基づいています。

従来型の vPC は、設定の交換および互換性の確認をするため vPC ペアとして設定された 2 つのスイッチが使用するメカニズムである整合性チェックを利用します。BGP EVPN 技術には整合性チェック メカニズムはありませんが、LACP を使用して設定ミスを検出します。また、vPC で従来使用されていた MCT リンクも不要になり、各 VTEP を 1 つ以上の冗長グループに含めることができるため、柔軟性が向上します。特定のグループ内の多数の VTEP を潜在的にサポートできます。

BGP EVPN マルチホーミング

BGP EVPN コントロール プレーンを使用する場合、各スイッチは自身のローカル IP アドレスを VTEP IP アドレスとして使用でき、アクティブ/アクティブ冗長性を提供します。さらに、

BGP EVPNベースのマルチホーミングは、特定の障害シナリオで高速コンバージェンスを提供します。

BGP EVPN マルチホーミングの用語

BGP EVPN マルチホーミングで使用される用語については、次の項を参照してください。

- EVI：VNI で表される EVPN インスタンス。
- MAC-VRF：MAC アドレスの仮想転送テーブルを格納するコンテナ。MAC-VRF ごとに、一意のルート識別子とインポート/エクスポートターゲットを設定できます。
- ES：バンドル リンクのセットを構成できるイーサネット セグメント。
- ESI：ネットワーク全体にわたってイーサネット セグメントを一意に表すイーサネット セグメント識別子。

EVPN マルチホーミングの実装

EVPN オーバーレイでは、次のような場合に、BGP MPLS ベースの EVPN ソリューションを VXLAN のカプセル化を伴うネットワーク仮想化オーバーレイとして適用できるようにするアダプテーションを指定します。BGP MPLS EVPN のプロバイダー エッジ（PE）ノードの役割はVTEP/ネットワーク仮想化エッジデバイス（NVE）に相当し、VTEP はデータプレーンの学習ではなく、BGPを介したコントロール プレーンの学習と配信を使用します。

現在定義されている 5 つの異なるルート タイプがあります。

- イーサネット自動検出（EAD）ルート
- MAC アドバタイズメント ルート
- 包括的なマルチキャスト ルート
- イーサネット セグメント ルート
- IP プレフィックス ルート

Cisco NX-OS で実行されている BGP EVPN は、ルート タイプ 2 を使用して MAC および IP（ホスト）情報をアドバタイズし、ルートタイプ 3 を使用して VTEP 情報を伝送し（特に入力複製用）、EVPN ルート タイプ 5 はルートキーに MAC アドレスがないネットワーク層到達可能性情報（NLRI）の IPv4 または IPv6 プレフィックスをアドバタイズすることを許可します。

EVPN マルチホーミングの導入により、Cisco NX-OS ソフトウェアは、イーサネット セグメント識別子とイーサネットタグ ID が NLRI のプレフィックスの一部と見なされるイーサネット自動検出（EAD）ルートを利用します。エンドポイントの到達可能性は BGP コントロール プレーンを介して学習されるため、ネットワーク コンバージェンス時間は、障害シナリオの場合に VTEP によって取り消される必要がある MAC/IP ルートの数の関数です。このような状況に対処するために、各VTEPは、ローカルに接続された各イーサネットセグメントに対して、ES ルートごとに1つ以上のイーサネット自動検出のセットをアドバタイズし、接続状態のセグメ

ントに障害が発生すると、ES ルートごとに対応するイーサネット自動検出のセットを取り消します。

イーサネット セグメント ルートは、EVPN マルチホーミングを備えた Cisco NX-OS ソフトウェアで主に BUM トラフィックの Designated Forwarder（DF）選定に使用されるもう 1 つのルートタイプです。イーサネット セグメントがマルチ ホームの場合、複数の DF が存在すると、パケットの重複に加えてループが転送される可能性があります。そのため、イーサネット セグメント ルート（タイプ 4）を使用して、指定フォワーダを選択し、スプリット ホライズン フィルタリングを適用します。イーサネット セグメントが設定されているすべての VTEP/PE がこのルートを発信します。

EVPN マルチホーミングの新しい実装概念を要約すると、次のようになります。

- EAD/ES：ES ごとのイーサネット自動検出ルートはタイプ 1 ルートとも呼ばれます。このルートは、アクセス失敗のシナリオ時にトラフィックを早急に収束するために使用されます。このルートにはイーサネット タグ 0xFFFFFFFF が使用されます。
- EAD/EVI：EVI ごとのイーサネット自動検出ルートはタイプ 1 ルートとも呼ばれます。このルートは、トラフィックはスイッチの 1 つにのみハッシュされるときのエイリアシングとロードバランシングに使用されます。EAD/ES ルートと区別するため、このルートにはイーサネット タグ値 0xFFFFFF を使用できません。
- イーサネット セグメント ルートはタイプ 4 ルートとも呼びます。このルートは、BUM トラフィックの指定フォワーダ（DF）の選択に使用されます。
- エイリアシング：タイプ 1 ルートの EAD/EVI ルートを使用する所定のイーサネット セグメントで接続されているすべてのスイッチへのトラフィックのロードバランシングに使用されます。これはホストを実際に学習するスイッチとは関係なく実行されます。
- 大量撤回：タイプ 1 EAD/ES ルートを使用し、アクセス障害シナリオ時に早急に収束するために使用されます。
- DF 選択：1 つのスイッチだけが特定のイーサネット セグメントのトラフィックをカプセル化解除および転送できるため、ループおよび重複の転送を防止します。
- スプリット ホライズン：BUM トラフィックのループおよび重複の転送を防ぐために使用されます。リモート サイトから発信された BUM トラフィックのみがローカル サイトに転送されます。

EVPN マルチホーミング冗長グループ

スイッチ L1 および L2 が Integrated Routing and Bridging（IRB）を実行する分散型エニーキャスト VXLAN ゲートウェイであるデュアル ホーム トポロジを考えます。ホスト H2 は、L1 と L2 の両方にデュアル ホーム接続されたアクセス スイッチに接続されています。

アクセス スイッチは、バンドルされた物理リンクのペアを介して L1 および L2 に接続されます。スイッチは、バンドルが反対側の 2 つの異なるデバイスで設定されていることを認識しません。ただし、L1 と L2 の両方が同じバンドルの一部であることを認識している必要があります。

L1 スイッチと L2 スイッチの間にマルチシャーシ EtherChannel トランク（MCT）リンクがなく、各スイッチが同じネイバーのセットと共有される同様の複数のバンドル リンクを持つことができることに注意してください。

スイッチ L1 と L2 が同じバンドル リンクの一部であることを認識させるために、NX-OS ソフトウェアは、イーサネットセグメント識別子（ESI）とインターフェイス（PO）で設定されたシステム MAC アドレス（system-mac）を利用します。

イーサネット セグメント識別子

EVPN には、イーサネット セグメント識別子（ESI）の概念が導入されています。各スイッチは、マルチホーム ネイバーと共有するバンドル リンクの下に 10 バイトの ESI 値を設定します。ESI 値は、手動で設定することも、自動で取得することもできます。

LACP バンドリング

LACPをオンにして、マルチホームポートチャネルバンドルでESIの設定ミスを検出します。これは、LACP が ESI で設定された MAC アドレス値をアクセス スイッチに送信するためです。LACP は ESI とともに必須ではありません。特定の ESI インターフェイス（PO）は、グループ内の VTEP 間で同じ ESI ID を共有します。

アクセス スイッチは、両方のスイッチ（L1 および L2）から同じ設定済み MAC 値を受信します。したがって、バンドルされたリンクは UP 状態になります。ES MAC はスイッチ上のすべてのイーサネット セグメントで共有できるため、LACP PDU はシステム MAC アドレスとして ES MAC を使用し、admin_key は ES ID を伝送します。

LACP PDU には、誤って設定された ES ID を検出して処理するメカニズムがあるため、スイッチとアクセス デバイス間で LACP を実行することを推奨します。同じ PO で設定された ES ID に不一致がある場合、LACP はリンクの 1 つをダウンさせます（オンラインになる最初のリンクはアップのままです）。デフォルトで、ほとんどの Cisco Nexus プラットフォームでは、LACP は、ピアから LACP PDU を受信しない場合、ポートを一時停止状態に設定します。**lacp suspend-individual** コマンドは、デフォルトで有効になっています。このコマンドは、ESI 設定の不一致が原因で発生するループの防止に役立ちます。したがって、アクセス スイッチとサーバのポートチャネルでこのコマンドを有効にすることをお勧めします。

これによって、サーバの中には起動に失敗するものがあります。そのようなサーバは、LACP が論理的にポートを稼働状態にしていることを必要とするからです。静的ポートチャネルを使用していて、ES ID が一致していない場合、MAC アドレスは L1 スイッチと L2 スイッチの両方から学習されます。そのため、両方のスイッチが、異なる ES ID に属する同じ MAC アドレスをアドバタイズして、MAC アドレス移動シナリオをトリガーします。最終的に、L1 スイッチと L2 スイッチの両方で学習された MAC アドレスのトラフィックは、そのノードに転送されません。

ESI を使用した EVPN マルチホーミングとの相互運用性

Cisco NX-OS リリース 10.2(2)F以降、予約されていない ESI（0 または MAX-ESI）値と予約されている ESI（0 または MAX-ESI）値を持つ EVPN MAC/IP ルート（タイプ 2）は、転送（ESI RX）のために評価されます。EVPN MAC/IP ルート解決の定義は、RFC 7432 Section 9.2.2 で定義されています。

EVPN MAC/IP ルート（タイプ 2）-

- 予約されている ESI 値（0 または MAX-ESI）は、MAC/IP ルート単独（タイプ 2 内の BGP ネクストホップ）によって単独で解決されます。
- 予約されていない ESI 値は、適合する ES イーサネット自動検出ルート（タイプ 1、ES EAD ごと）が存在する場合、単独で解決されます。

異なる ESI 値を使用した EVPN MAC/IP ルート解決は、Cisco Nexus 9300-EX、-FX、-FX2、-FX3、および -GX プラットフォーム スイッチでサポートされています。

Cisco NX-OS Release 10.2(3) 以降、VXLAN から MPLS-SR へのゲートウェイは、Cisco Nexus 9300-GX2 プラットフォーム スイッチでサポートされます。

VXLAN EVPN マルチホーミングの注意事項と制限事項

VXLAN EVPN マルチホーミングの設定については、次の制限事項を参照してください。

- Cisco NX-OS リリース 10.2(2)F 以降、ND-ISSU はすべての ESI-RX ノードでサポートされています。
- Cisco NX-OS リリース 9.2(3) 以降では、ピアリンク レス vPC/vPC^2 を使用するVXLAN VLAN 上の FEX メンバー ポートはサポートされません。
- VXLAN EVPN マルチホーミングは、iBGP または eBGP コントロール プレーンで動作します。 iBGP が推奨されます。
- iBGP を VXLAN EVPN マルチホーミングで使用する場合、ローカルで学習されたエンドポイントのアドミニストレーティブ ディスタンスの値は、iBGP の値よりも小さくする必要があります。

（注） ローカル学習エンドポイントのデフォルト値は 190、eBGP のデフォルト値は 20、iBGP のデフォルト値は 200です。

- eBGP を VXLAN EVPN マルチホーミングで使用する場合、ローカルで学習したエンドポイントのアドミニストレーティブ ディスタンスは、eBGP の値よりも小さくする必要があります。アドミニストレーティブ ディスタンスは、 **fabric forwarding admin-distance** *distance* コマンドを入力して変更できます。

（注） ローカル学習エンドポイントのデフォルト値は 190、eBGP のデフォルト値は 20、iBGP のデフォルト値は 200です。

- EVPN マルチホーミングは、Cisco Nexus 9300 プラットフォーム スイッチでのみサポートされます。Cisco Nexus 9300-EX/FX/FXP/FX2/FX3、9300-GX、および 9500 プラットフォーム スイッチではサポートされません。Cisco Nexus 9500 プラットフォーム スイッチはスパイン スイッチとして使用できますが、VTEP として使用することはできません。
- EVPN マルチホーミングでは、特定のネットワーク内のすべてのスイッチが EVPN マルチホーミングに対応している必要があります。EVPN マルチホーミングの有無にかかわらずプラットフォームを混在させることはサポートされていません。
- EVPN マルチホーミングは FEX ではサポートされていません。
- ARP 抑制は EVPN マルチホーミングでサポートされます。
- EVPN マルチホーミングは、2 つのスイッチへのマルチホーミングでのみサポートされます。
- EVPN マルチホーミングを有効にするには、Cisco NX-OS リリース 7.0(3)I5(2) 以降のソフトウェア バージョンを実行している必要があります。
- スイッチポート トランク ネイティブ VLANは、トランク インターフェイスではサポートされません。
- ES PO で LACP を有効にすることを推奨します。
- IPv6 は現時点でサポートされていません。
- ISSU は ESI が Cisco Nexus 9300 シリーズ スイッチで設定されている場合には、サポートされません。

VXLAN EVPN マルチホーミングの設定

EVPN マルチホーミングを有効にする

Cisco NX-OS では、vPC ベースの EVPN マルチホーミングまたは ESI ベースの EVPN マルチホーミングが可能です。両方の機能を同時に有効にすることはできません。ESI ベースのマルチホーミングは、**evpn esi multihoming** CLI コマンドを使用して有効にします。ESI マルチホーミングのコマンドを使用すると、イーサネット セグメント設定とスイッチでのイーサネット セグメント ルートの生成が可能になることに注意してください。

有効なESIを持つタイプ1およびタイプ2ルートの受信とパスリスト解決は、**evpn esi multihoming** コマンドに関連付けられません。スイッチが有効なESIを持つMAC/MAC-IPルートを受信し、コマンドが有効になっていない場合でも、ES ベースのパス解決ロジックはこれらのリモート

ルートに適用されます。これは、vPC 対応スイッチと ESI 対応スイッチ間の相互運用性のために必要です。

EVPN マルチホーミングを設定するには、次の手順を実行します。

始める前に

EVPN ESI マルチホーミングを有効にする前に、VXLAN を BGP-EVPN で設定する必要があります。

手順の概要

1. **evpn esi multihoming**
2. **address-family l2vpn evpn maximum-paths <>maximum-paths ibgp <>**
3. **evpn multihoming core-tracking**
4. **interface port-channel Ethernet-segment <>System-mac <>**
5. **hardware access-list tcam region vpc-convergence 256**

手順の詳細

	コマンドまたはアクション	目的
ステップ 1	**evpn esi multihoming**	EVPN マルチホーミングをグローバルに有効にします。

ステップ 2	**address-family l2vpn evpn maximum-paths <>maximum-paths ibgp <>**
	例：
	`address-family l2vpn evpn`
	`maximum-paths 64`
	`maximum-paths ibgp 64`
		BGP 最大パスを有効にして、MAC ルートの ECMP を有効にします。それ以外の場合、MAC ルートにはネクストホップとして 1 つの VTEP しかありません。この設定は、グローバル レベルの BGP で必要です。

ステップ 3	**evpn multihoming core-tracking**	EVPN マルチホーミング コアリンクを有効にします。コアへのアップリンク インターフェイスを追跡します。すべてのアップリンクがダウンしている場合、POに基づくローカルESはシャットダウン/一時停止されます。これは主に、アップリンクが使用できない場合の South-to-North へのトラフィックのブラックホール化を回避するために使用されます。

ステップ 4	**interface port-channel Ethernet-segment <>System-mac <>**
	例：
	`ethernet-segment 11`
	`system-mac 0000.0000.0011`
		ローカル イーサネット セグメント ID を設定します。ES ID は、PO がマルチホームである VTEP で一致する必要があります。イーサネット セグメント ID は PO ごとに一意である必要があります。
		PO がマルチホーム接続されている VTEP で一致する必要があるローカル システム MAC ID を設定します。システム MAC アドレスは、複数の PO 間で共有できます。

	コマンドまたはアクション	目的
ステップ 5	**hardware access-list tcam region vpc-convergence 256** 例： `hardware access-list tcam region vpc-convergence 256`	TCAM を設定します。このコマンドは、ハードウェアでスプリット ホライズン ACL を設定するために使用されます。このコマンドは、共有 ES PO での BUM トラフィックの重複を回避します。

VXLAN EVPN マルチホーミングの設定例

スイッチのサンプル VXLAN EVPN マルチホーミング設定を参照してください。

```
Switch 1 (L1)

evpn esi multihoming

router bgp 1001
   address-family l2vpn evpn
   maximum-paths ibgp 2

interface Ethernet2/1
  no switchport
  evpn multihoming core-tracking
  mtu 9216
  ip address 10.1.1.1/30
  ip pim sparse-mode
  no shutdown

interface Ethernet2/2
  no switchport
  evpn multihoming core-tracking
  mtu 9216
  ip address 10.1.1.5/30
  ip pim sparse-mode
  no shutdown

interface port-channel11
  switchport mode trunk
  switchport trunk allowed vlan 901-902,1001-1050
  ethernet-segment 2011
    system-mac 0000.0000.2011
  mtu 9216


Switch 2 (L2)

evpn esi multihoming

router bgp 1001
   address-family l2vpn evpn
   maximum-paths ibgp 2

interface Ethernet2/1
  no switchport
  evpn multihoming core-tracking
  mtu 9216
  ip address 10.1.1.2/30
```

```
  ip pim sparse-mode
  no shutdown

interface Ethernet2/2
  no switchport
  evpn multihoming core-tracking
  mtu 9216
  ip address 10.1.1.6/30
  ip pim sparse-mode
  no shutdown

interface port-channel11
  switchport mode trunk
  switchport access vlan 1001
  switchport trunk allowed vlan 901-902,1001-1050
  ethernet-segment 2011
    system-mac 0000.0000.2011
  mtu 9216
```

レイヤ 2 ゲートウェイ STP の設定

レイヤ 2 ゲートウェイ STP の概要

EVPN マルチホーミングは、レイヤ 2 ゲートウェイ スパニング ツリー プロトコル（L2G-STP）でサポートされます。レイヤ 2 ゲートウェイ スパニング ツリー プロトコル（L2G-STP）はループフリー ツリー トポロジを構築します。ただし、スパニング ツリー プロトコルのルートは常に VXLAN ファブリック内にある必要があります。スパニング ツリー プロトコルのブリッジ ID は、MAC アドレスおよびブリッジ優先順位で構成されます。システムが VXLAN ファブリックで実行中、システムは自動的に、予約済みの MAC アドレスのプールから MAC アドレス c84c.75fa.6000 で VTEP を割り当てます。その結果、各スイッチは単一の論理疑似ルートをエミュレートするブリッジ ID に同じ MAC アドレスを使用します。

レイヤ 2 ゲートウェイ スパニング ツリー プロトコル（L2G-STP）は、EVPN ESI マルチホーミング VLAN ではデフォルトで無効になっています。**spanning-tree domain enable** CLI コマンドを使用して、すべての VTEP で L2G-STP を有効にします。L2G-STP を有効にすると、VXLAN ファブリック（すべての VTEP）は、カスタマー アクセス スイッチの単一の擬似ルート スイッチをエミュレートします。L2G-STP はブート時にデフォルトですべての VXLAN VLAN で実行するように開始され、ルートはオーバーレイで固定されます。L2G-STP では、すべてのアクセス ポートでルート ガードがデフォルトで有効になります。さらに、スパニング ツリー トポロジ変更通知（STP-TCN）をファブリック全体でトンネリングできるようにするには、**spanning-tree domain** <*id*> を使用します。

カスタマー アクセス スイッチに接続する VTEP からのすべてのアクセス ポートは、デフォルトで *desg* フォワーディング ステートになっています。VTEP に接続するカスタマー アクセス スイッチ上のすべてのポートは、ルート ポートフォワーディングまたは代替ポート ブロッキング ステートのいずれかです。優れたまたは優れた STP 情報がカスタマー アクセススイッチから受信されると、ルート ガードが起動し、ポートを *blk l2g_inc* 状態にして、オーバーレイ ファブリックのルートを保護し、ループを防止します。

レイヤ 2 ゲートウェイ STP への移行に関するガイドライン

レイヤ 2 ゲートウェイ STP に移行するには、次の手順を実行します。

- レイヤ 2 ゲートウェイ STP では、ルート ガードはすべてのアクセス ポートでデフォルトで有効になっています。
- レイヤ 2 ゲートウェイ STP を有効にすると、VXLAN ファブリック（すべての VTEP）がカスタマー アクセス スイッチの単一の疑似ルート スイッチをエミュレートします。
- カスタマーアクセス スイッチに接続する VTEP からのすべてのアクセス ポートは、デフォルトで **Desg FWD** 状態になっています。
- VTEP に接続しているカスタマー アクセス スイッチのすべてのポートは、ルートポート FWDまたは Altn BLK 状態のいずれかです。
- ルート ガードは、カスタマー アクセス スイッチから上位スパニング ツリー情報を受信した場合にアクティブになります。このプロセスでは、ポートを **BLK L2GW_Inc** 状態にして、VXLAN ファブリックのルートを保護し、ループを防止します。
- ファブリック全体でスパニング ツリー BPDU トンネリングを有効にするには、明示的なドメイン ID 設定が必要です。
- ベスト プラクティスとして、接続されているスパニングツリー ドメインのすべてのすべてのスイッチの中で最も低いスパニングツリーの優先順位で、すべての VTEPを設定する必要があります。すべての VTEP をルート ブリッジとして設定すると、VXLAN ファブリック全体が 1 つの仮想ブリッジのように見えます。
- スパニング ツリー エッジ モードでは、レイヤ 2 ゲートウェイ STP を VTEP およびアクセス レイヤで実行できるようにするため、ESI インターフェイスを有効にしないでください。
- スパニング ツリー プロトコルを実行しておらず、エンド ホストであるホストまたはサーバに直接接続している場合は、スパニング ツリー エッジ モードで ESI またはオーファン（シングル ホーム ホスト）を引き続き使用できます。
- 同じレイヤ 2 ゲートウェイ STP ドメイン内の共通のカスタマー アクセス レイヤによって接続されているすべての VTEP を設定します。理想的には、ホストが存在し、ホストが移動できるファブリック上のすべての VTEP。
- レイヤ 2 ゲートウェイ STP ドメイン スコープはグローバルであり、特定の VTEP 上のすべての ESI は 1 つのドメインにのみ参加できます。
- 複数のスパニング ツリー（MST）インスタンスと VLAN 間のマッピングは、特定のレイヤ 2 ゲートウェイ STP ドメイン内の VTEP 間で一貫している必要があります。
- 非レイヤ 2 ゲートウェイ STP 対応 VTEP は、レイヤ 2 ゲートウェイ STP 対応 VTEP に直接接続できません。このアクションを実行すると、非レイヤ 2 ゲートウェイ STP VTEP が BPDU を送信し続け、ルートを外部に誘導できるため、競合と紛争が発生します。

- VXLAN ファブリックに対してローカルな STP ドメインのルートが VTEP であるか、ファブリック内に配置されていることを確認します。
- 最新のビルドにアップグレードした後、Cisco Nexus スイッチとアクセス スイッチの両方で現在のエッジと BPDU フィルタの設定を保持します。
- 推奨される優先順位および必要に応じて *mst* インスタンスマッピングを使用して、すべてのスイッチでレイヤ 2 ゲートウェイ STP を有効にします。 **spanning-tree domain enable** コマンドおよび **spanning-tree mst <instance-id's> priority 8192** コマンドを使用します。
- 最初にスイッチ側の BPDU フィルタ設定を削除します。
- BPDU フィルタ設定とカスタマー アクセス スイッチのエッジを取り外します。

  これで、トポロジはレイヤ 2 ゲートウェイ STP に収束し、冗長接続のブロッキングはアクセス スイッチ レイヤにプッシュされます。

スイッチでのレイヤ 2 ゲートウェイ STP の有効化

スイッチでレイヤ 2 ゲートウェイ STP を有効にするには、次の手順を実行します。

手順の概要

1. **spanning-tree mode <rapid-pvst, mst>**
2. **spanning-tree domain enable**
3. **spanning-tree domain 1**
4. **spanning-tree mst <id> priority 8192**
5. **spanning-tree vlan <id> priority 8192**
6. **spanning-tree domain disable**

手順の詳細

	コマンドまたはアクション	目的
ステップ 1	**spanning-tree mode <rapid-pvst, mst>**	スパニング ツリー プロトコル モードを有効にします。
ステップ 2	**spanning-tree domain enable**	スイッチでレイヤ 2 ゲートウェイ STP を有効にします。すべての EVPN ESI マルチホーミング VLAN でレイヤ 2 ゲートウェイ STP を無効にします。
ステップ 3	**spanning-tree domain 1**	明示的なドメイン ID は、エンコードされた BPDU をコアおよびコアから受信したプロセスにトンネリングするために必要です。
ステップ 4	**spanning-tree mst <id> priority 8192**	スパニング ツリー プロトコルの優先順位の設定
ステップ 5	**spanning-tree vlan <id> priority 8192**	スパニング ツリー プロトコルの優先順位を設定します。

	コマンドまたはアクション	目的
ステップ 6	**spanning-tree domain disable**	VTEP でレイヤ 2 ゲートウェイ STP を無効にします。

例

すべてのレイヤ 2 ゲートウェイ STP VLAN は、カスタマーエッジ（CE）トポロジよりも低いスパニング ツリーの優先順位に設定し、VTEP がこの VLAN のスパニング ツリールートであることを確認する必要があります。アクセス スイッチの優先順位が高い場合は、レイヤ 2 ゲートウェイ STP の優先順位を0に設定して、VXLAN ファブリックにレイヤ 2 ゲートウェイ STP ルートを保持できます。次の設定例を参照してください。

```
switch# show spanning-tree summary
Switch is in mst mode (IEEE Standard)
Root bridge for: MST0000
L2 Gateway STP bridge for: MST0000
L2 Gateway Domain ID: 1
Port Type Default                        is disable
Edge Port [PortFast] BPDU Guard Default  is disabled
Edge Port [PortFast] BPDU Filter Default is disabled
Bridge Assurance                         is enabled
Loopguard Default                        is disabled
Pathcost method used                     is long
PVST Simulation                          is enabled
STP-Lite                                 is disabled

Name                   Blocking Listening Learning Forwarding STP Active
---------------------- -------- --------- -------- ---------- ----------
MST0000                     0         0        0         12         12
---------------------- -------- --------- -------- ---------- ----------
1 mst                       0         0        0         12         12


switch# show spanning-tree vlan 1001

MST0000
  Spanning tree enabled protocol mstp

  Root ID    Priority    8192
             Address     c84c.75fa.6001   L2G-STP reserved mac+ domain id
             This bridge is the root
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    8192   (priority 8192 sys-id-ext 0)
             Address     c84c.75fa.6001
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec
```

出力には、スパニング ツリーの優先順位が8192（デフォルトは 32768）に設定されていることが示されます。スパニング ツリーの優先順位は4096の倍数で設定されます。個々のインスタンスの優先順位は、プライオリティと Instance_ID として計算されま

す。この場合、優先順位は 8192 + 0 = 8192 として計算されます。レイヤ 2 ゲートウェイ STP では、アクセス ポート（アクセス スイッチに接続された VTEP ポート）でルート ガードが有効になっています。上位 BPDU がエッジ ポートで受信されると、次の例で示されるように、条件がクリアされるまでポートはレイヤ 2 ゲートウェイの一貫性のないステートのままになります。

```
2016 Aug 29 19:14:19 TOR9-leaf4 %$ VDC-1 %$ %STP-2-L2GW_BACKBONE_BLOCK: L2 Gateway
Backbone port inconsistency blocking port Ethernet1/1 on MST0000.
2016 Aug 29 19:14:19 TOR9-leaf4 %$ VDC-1 %$ %STP-2-L2GW_BACKBONE_BLOCK: L2 Gateway
Backbone port inconsistency blocking port port-channel13 on MST0000.

switch# show spanning-tree

MST0000
  Spanning tree enabled protocol mstp
  Root ID    Priority    8192
             Address     c84c.75fa.6001
             This bridge is the root
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    8192   (priority 8192 sys-id-ext 0)
             Address     c84c.75fa.6001
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------------
Po1              Desg FWD 20000     128.4096 Edge P2p
Po2              Desg FWD 20000     128.4097 Edge P2p
Po3              Desg FWD 20000     128.4098 Edge P2p
Po12             Desg BKN*2000      128.4107 P2p *L2GW_Inc
Po13             Desg BKN*1000      128.4108 P2p *L2GW_Inc
Eth1/1        Desg BKN*2000       128.1    P2p *L2GW_Inc
```

VTEP でレイヤ 2 ゲートウェイ STP を無効にするには、**spanning-tree domain disable** CLIコマンドを入力します。このコマンドは、すべての EVPN ESI マルチホーム VLAN でレイヤ 2 ゲートウェイ STP を無効にします。ブリッジの MAC アドレスがシステムの MAC アドレスに復元され、VTEP が必ずしもルートになるとは限りません。次の場合、レイヤ 2 ゲートウェイ STP が無効になっているため、アクセス スイッチがルートの役割を引き受けます。

```
switch(config)# spanning-tree domain disable

switch# show spanning-tree summary
Switch is in mst mode (IEEE Standard)
Root bridge for: none
L2 Gateway STP                          is disabled
Port Type Default                       is disable
Edge Port [PortFast] BPDU Guard Default  is disabled
Edge Port [PortFast] BPDU Filter Default is disabled
Bridge Assurance                        is enabled
Loopguard Default                       is disabled
Pathcost method used                    is long
PVST Simulation                         is enabled
STP-Lite                                is disabled

Name                   Blocking Listening Learning Forwarding STP Active
```

```
---------------------- -------- --------- -------- ---------- ----------
MST0000                     4         0        0          8         12
---------------------- -------- --------- -------- ---------- ----------
1 mst                       4         0        0          8         12


switch# show spanning-tree vlan 1001

MST0000
  Spanning tree enabled protocol mstp
  Root ID    Priority    4096
             Address     00c8.8ba6.5073
             Cost        0
             Port        4108 (port-channel13)
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID  Priority    8192   (priority 8192 sys-id-ext 0)
             Address     5897.bd1d.db95
             Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec
```

レイヤ 2 ゲートウェイ STP では、VTEP のアクセス ポートは、通常のスパニング ツリー ポートのように動作し、アクセス スイッチから BPDU を受信するため、エッジ ポートには配置できません。この場合、VTEP のアクセス ポートは高速伝送の利点を失い、代わりにイーサネット セグメント リンク フラップで転送されます。（FWD-Desg の役割を引き受ける前に、提案と合意のハンドシェイクを行う必要があります）。

VXLAN EVPN マルチホーミング トラフィック フローの設定

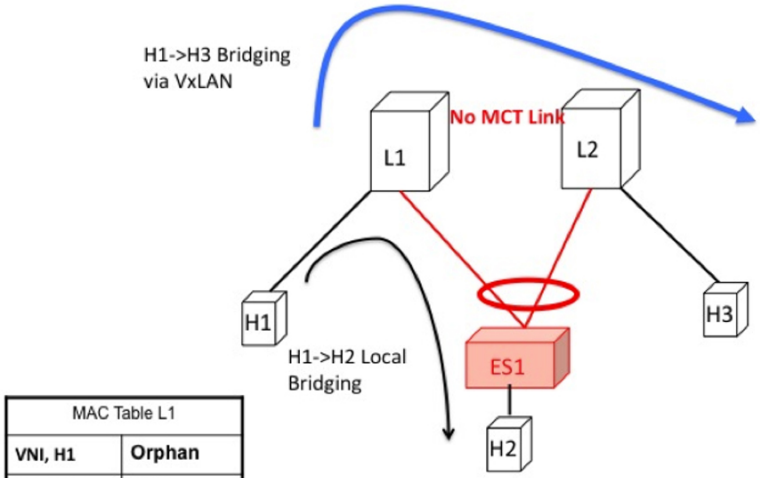
EVPN マルチホーミング ローカル トラフィック フロー

（ESI で定義されている）同じ冗長グループに属するすべてのスイッチは、アクセス スイッチ/ホストに対して単一の仮想スイッチとして機能します。ただし、ローカル アクセス用にトラフィックをブリッジおよびルーティングする MCT リンクはありません。

ローカル ブリッジング トラフィック

ホスト H2 はデュアル ホームであるのに対し、ホスト H1 と H3 はシングル ホームです（孤立とも呼ばれる）。トラフィックは L1 を介して H1 から H2 にローカルにブリッジされます。ただし、孤立した H1 と H3 の間でパケットをブリッジする必要がある場合は、パケットを VXLAN オーバーレイ経由でブリッジする必要があります。

図 ***1 : L1*** でのローカル ブリッジング。***VXLAN*** 経由の ***H1*** から ***H3*** へのブリッジング。***vPC*** では、***H1*** から ***H3*** へは ***MCT*** リンク経由です。



MAC Table L1	
VNI, H1	Orphan
VNI, H2	ESI1
VNI, H3	Peer-L2

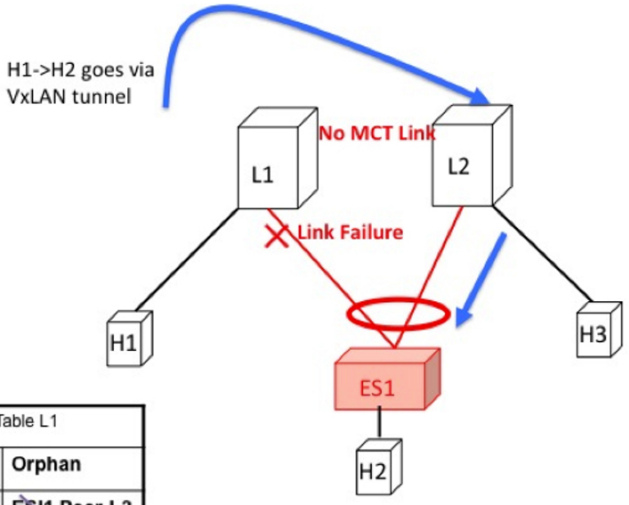
ローカルにブリッジされたトラフィックのアクセス障害

L1 の ESI リンクに障害が発生した場合、ブリッジされたトラフィックが H1 から H2 に到達するためのパスはありません。したがって、ローカル ブリッジ トラフィックは、H1 から H3 への孤立フローと同様に、最適ではないパスを使用します。

（注） このような状況が発生すると、H2 の MAC テーブル エントリは、ポート チャネル インターフェイスを指すローカル ルートから、L2 のピア ID を指すリモート オーバーレイ ルートに変わります。変更は BGP からシステムに浸透します。

マルチホーミングの設定

図 *2 : L1* での *ES1* 障害。*H1-> H2* が *VXLAN* トンネル経由でブリッジされます。

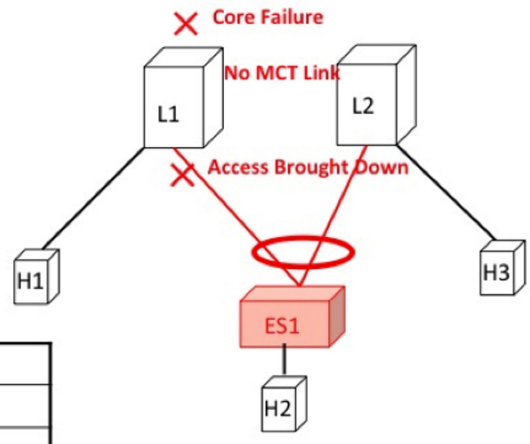


MAC Table L1	
VNI, H1	Orphan
VNI, H2	ESI1 Peer-L2
VNI, H3	Peer-L2

ローカル ブリッジング トラフィックのコア障害

スイッチ L1 がコアから分離された場合、アクセス トラフィックを引き付け続けることはできません。これは、スイッチ L1 がカプセル化してオーバーレイ上で送信できないためです。これは、L1 がコア到達可能性を失った場合、アクセス リンクを L1 でダウンさせる必要があることを意味します。このシナリオでは、専用 MCT リンクがないため、孤立した H1 はリモート ホストとローカルに接続されたホストの両方へのすべての接続を失います。

図 *3 : L1* のコア障害。*MCT* がないため、*H1-> H2* はすべての接続を失います。



MAC Table L1	
VNI, H1	Orphan
VNI, H2	Loss
VNI, H3	Loss

ローカル ルーティング トラフィック

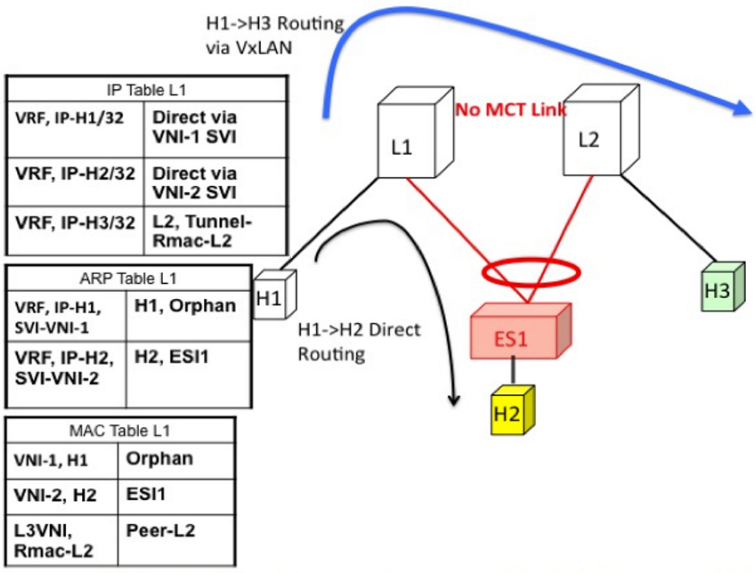
H1、H2、およびH3が異なるサブネットにあり、L1/L2が分散型エニーキャスト ゲートウェイであるとします。

H1 から H2 にルーティングされるパケットは、ネイティブ ルーティングを介してL1から直接送信されます。

ただし、ホスト H3 はローカルに接続された隣接関係ではありません。これは、ARP エントリがローカルに接続された隣接関係として L1 に同期する vPC の場合とは異なります。代わりに、H3は、L3 VNI のコンテキストでインストールされた L1 の IP テーブルにリモート ホストとして表示されます。このパケットは、L2 のルータ MAC にカプセル化され、VXLAN オーバーレイを介して L2 にルーティングされる必要があります。

したがって、H1からH3へのルーティングされたトラフィックは、異なるサブネットの真のリモート ホスト間のルーティングされたトラフィックとまったく同じ方法で発生します。

図 *4 : L1* は分散型エニーキャスト ゲートウェイです。*H1*、*H2*、および *H3* は異なる *VLAN* にあります。*H1* から *H3* へのルーティングは、*VXLAN* トンネル カプセル化によって行われます。*vPC* では、*H3 ARP* は *MCT* とダイレクト ルーティングを介して同期されます。

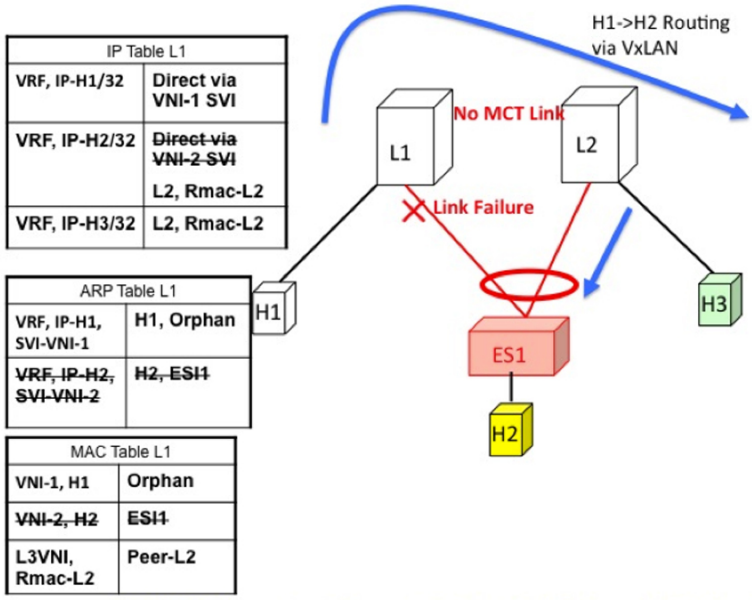


ローカルにルーティングされたトラフィックのアクセス障害

スイッチ L1 の ESI リンクに障害が発生した場合、ルーティングされたトラフィックが H1 から H2 に到達するためのパスはありません。したがって、ローカルにルーティングされたトラフィックは、H1 から H3 への孤立フローと同様に、最適ではないパスを使用します。

マルチホーミングの設定

図 *5 : H1*、*H2*、および *H3* は異なる *VLAN* にあります。*L1* で *ESI* が失敗します。*H1* から *H2* へのルーティングは、*VXLAN* トンネル カプセル化によって行われます。

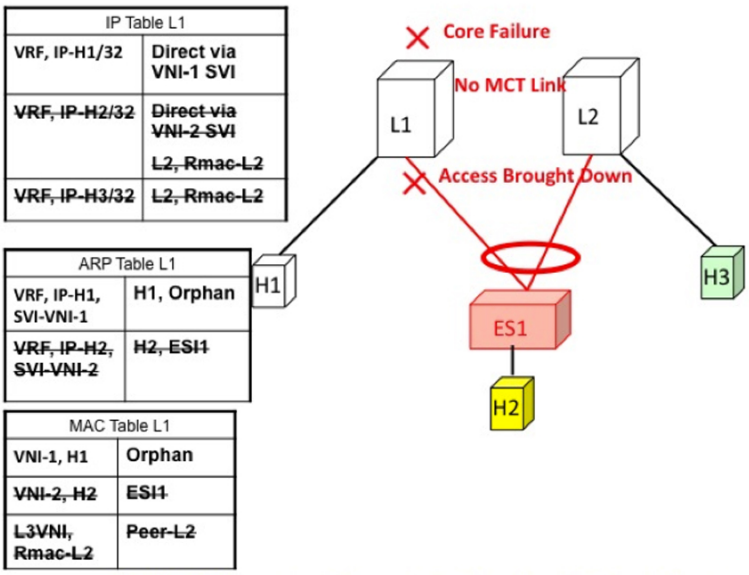


ローカルにルーティングされたトラフィックのコア障害

スイッチ L1 がコアから分離された場合、アクセス トラフィックを引き付け続けることはできません。これは、スイッチ L1 がカプセル化してオーバーレイ上で送信できないためです。これは、L1 がコア到達可能性を失った場合、アクセスリンクを L1 でダウンさせる必要があることを意味します。

このシナリオでは、専用の MCT リンクがないため、リモート H1 とローカルに接続されたホストの両方へのすべての接続が失われます。

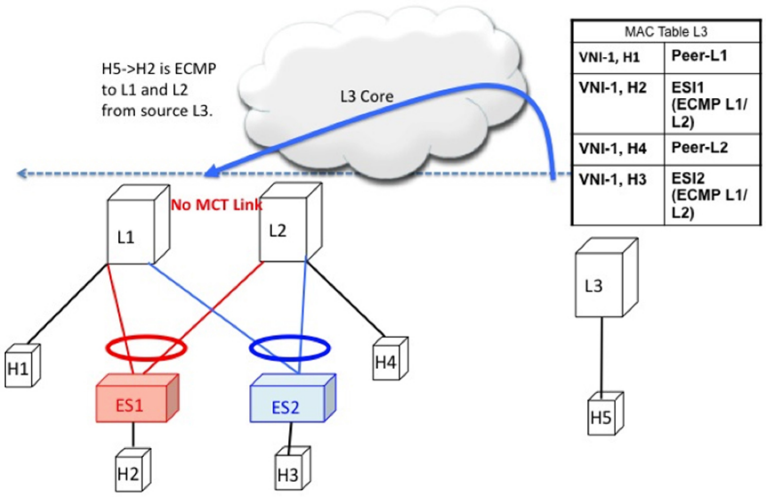
図 *6 : H1*、*H2*、および *H3* は異なる *VLAN* にあります。コアが *L1* で失敗します。アクセスがダウンします。*H1* はすべての接続を失います。



EVPN マルチホーミングのリモート トラフィック フロー

スイッチ L1 と L2 で構成されるマルチホーム コンプレックスにブリッジングおよびルーティングされたトラフィックを送信するリモート スイッチ L3 を考えます。この MH コンプレックスを表す仮想またはエミュレートされた IP がないため、L3 はブリッジド トラフィックとルーテッド トラフィックの両方の送信元で ECMP を実行する必要があります。この項では、ブリッジおよびルーテッドの両方のケースでスイッチ L3 で ECMP がどのように達成されるか、およびシステムがコア障害とアクセス障害と相互作用する方法について説明します。

図 *7 :* レイヤ *2 VXLAN* ゲートウェイ。*L3* は *MAC ECMP* を *L1/L2* に実行します。



リモート ブリッジド トラフィック

EVPN MH コンプレックス（L1、L2）の背後にあるホスト H2 にトラフィックをブリッジするリモート ホスト H5 を考えます。ホスト H2 は、RFC 7432 で定義されているルールに従って ECMP リストを作成します。スイッチ L3 の MAC テーブルは、H2 の MAC エントリが IP-L1 と IP-L2 で構成される ECMP PathList を指していることを示しています。H5 から H2 に向かうブリッジされたトラフィックはすべて VXLAN カプセル化され、スイッチ L1 および L2 にロードバランシングされます。ECMP リストを作成する場合は、次の構成要素に留意する必要があります。

- 一括撤回：PathList 修正の原因となる障害は、MAC の規模に依存しません。
- エイリアシング：PathList の挿入は、MAC の規模に依存しない場合があります（オプションのルートのサポートに基づく）。

次に、この MAC ECMP PathList を作成するために必要な主な構成要素を示します。

ES ごとのイーサネット自動検出ルート（タイプ 1）

EVPN は、イーサネット セグメントへの接続に障害が発生したときに、転送テーブルを更新する必要性を効率的かつ迅速に通知するメカニズムを定義します。各PEに、ローカルに接続された各イーサネット セグメントの ES ルートごとに 1 つ以上のイーサネット AD のセットをアドバタイズさせることで、これを行います。

ES ごとのイーサネット自動検出ルート（ルート タイプ 1）		
NLRI	Route Type（ルート タイプ）	イーサネット セグメント
	ルート識別子	ルータ ID：セグメント ID（VNID << 8）
	ESI	<Type: 1B><MAC: 6B><LD: 3B>
	イーサネット タグ	MAX-ET
	MPLS Label	0
ATTRS	拡張コミュニティ ESI ラベル = 0	Single Active = False
	Next-Hop	NVE ループバック IP
	Route Target	ES でアクティブなすべての EVI に関連付けられている MAC-VRF の RT リストのサブセット

MAC-IP ルート（タイプ 2）

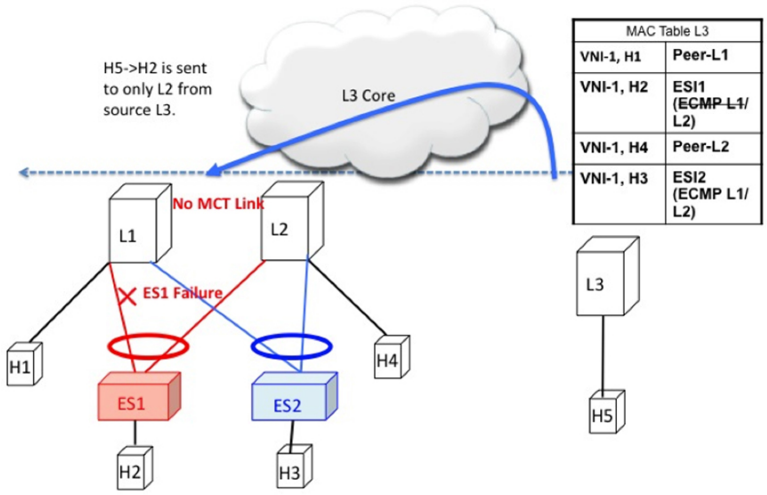
MAC-IP ルートは、現在の vPC マルチホーミングおよびスタンドアロン シングル ホーミング ソリューションで使用されているものと同じです。ただし、現在はマルチホーム ホストであり、ECMP パス解決の候補であることを示すゼロ以外の ESI フィールドがあります。

MAC IP ルート（ルート タイプ 2）		
NLRI	Route Type（ルート タイプ）	MAC IP ルート（タイプ 2）
	ルート識別子	ホストに関連付けられた MAC-VRF の RD
	ESI	<Type: 1B><MAC: 6B><LD: 3B>
	イーサネット タグ	MAX-ET
	MAC Addr	ホストの MAC アドレス
	IP Addr	ホストの IP アドレス
	ラベル	MAC-VRF に関連付けられた L2VNI L3-VRF に関連付けられた L3VNI
ATTRS	Next-Hop	NVE のループバック
	RT のエクスポート	ホストに関連付けられた MAC-VRF（AND/OR）L3-VRF で設定された RT

リモート ブリッジド トラフィックのアクセス障害

ESI リンクに障害が発生した場合は、一括で取り消されます。EAD/ES ルートが取り消され、リモート デバイスが特定の ES の ECMP リストからスイッチをリモートにします。

図 *8 :* レイヤ *2 VXLAN* ゲートウェイ。*L1* の *ESI* 障害。*L3* は *MAC ECMP* リストから *L1* を削除します。これは、*L1*からの *EAD / ES*の一括回収が原因で発生します。

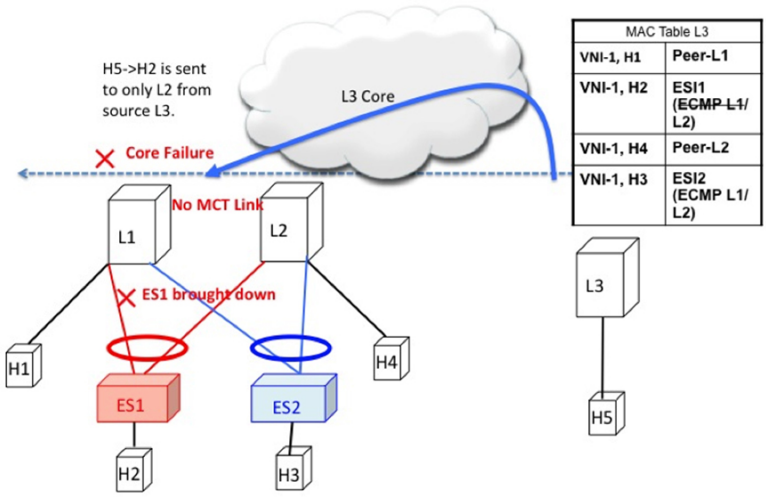


マルチホーミングの設定

リモート ブリッジド トラフィックのコア障害

スイッチ L1 がコアから分離された場合、アクセス トラフィックを引き付け続けることはできません。これは、スイッチ L1 がカプセル化してオーバーレイで送信できないためです。これは、L1 がコア到達可能性を失った場合、アクセスリンクを L1 でダウンさせる必要があることを意味します。

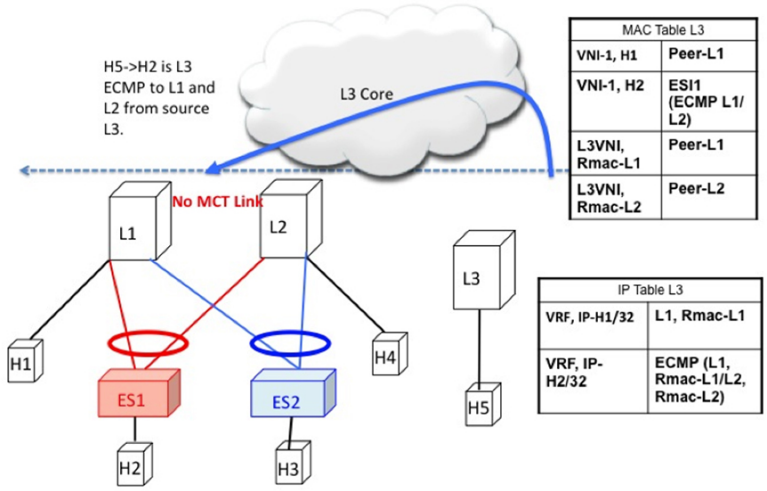
図 *9 :* レイヤ *2 VXLAN* ゲートウェイ。*L1* でコア障害が発生しました。*L3* は *MAC ECMP* リストから *L1* を削除します。これは、*L1* へのルート到達可能性が *L3* でなくなるために発生します。



リモート ルーテッド トラフィック

L3 がレイヤ 3 VXLAN ゲートウェイであり、H5 と H2 が異なるサブネットに属しているとします。この場合、L3 から L1/L2 に向かうサブネット間トラフィックは、分散エニーキャスト ゲートウェイである L3 でルーティングされます。L1 と L2 の両方が、ホスト H2 の MAC-IP ルートをアドバタイズします。これらのルートの受信により、L3 は L1 と L2 で構成される L3 ECMP リストを作成します。

図 *10 :* レイヤ *3 VXLAN* ゲートウェイ。*L3* は、サブネット間トラフィックの *L1/L2* に対して *IP ECMP* を実行します。



リモート ルーテッド トラフィックのアクセス障害

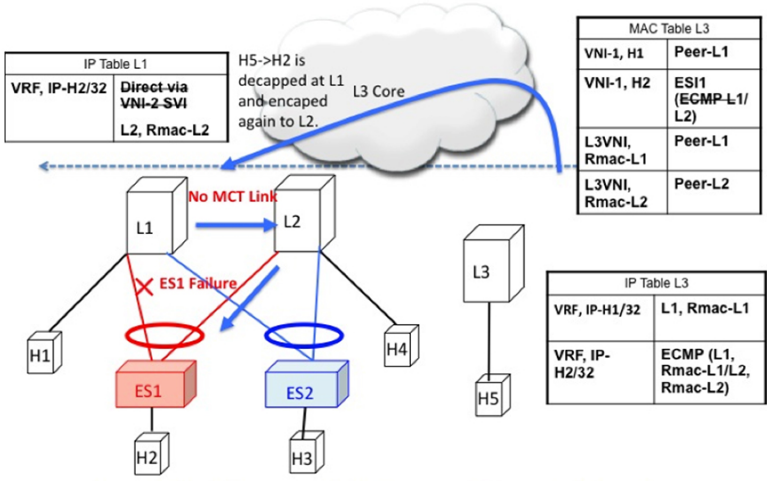
ES1 を指すアクセス リンクが L1 でダウンすると、大量の撤回ルートが EAD/ES の形式で送信されるため、L3 は MAC ECMP PathList から L1 を削除し、イントラ サブネット（L2）トラフィックを迅速に収束させます。L1 は E2 リンクを介して直接接続されていないため、H2 を VxLAN オーバーレイ経由で到達可能なリモート ルートとして扱います。これにより、H2 宛てのトラフィックは次善のパス L3-> L1-> L2 になります。

サブネット間トラフィック H5-> H2 は次のパスに従います。

- パケットは H5 によって L3 のゲートウェイに送信されます。
- L3 は対称 IRB を実行し、VXLAN オーバーレイを介してパケットを L1 にルーティングします。
- L1 はパケットのカプセル化を解除し、H2 の内部 IP ルックアップを実行します。
- H2 はリモート ルートです。したがって、L1 は VXLAN オーバーレイを介して L2 にパケットをルーティングします。
- L2 はパケットのカプセル化を解除し、IP ルックアップを実行して、直接接続された SVI にルーティングします。

したがって、ルーティングは L3、L1、および L2 でそれぞれ 1 回ずつ行われます。この次善の動作は、タイプ 2 ルートが BGP によって L1 によって取り消されるまで続きます。

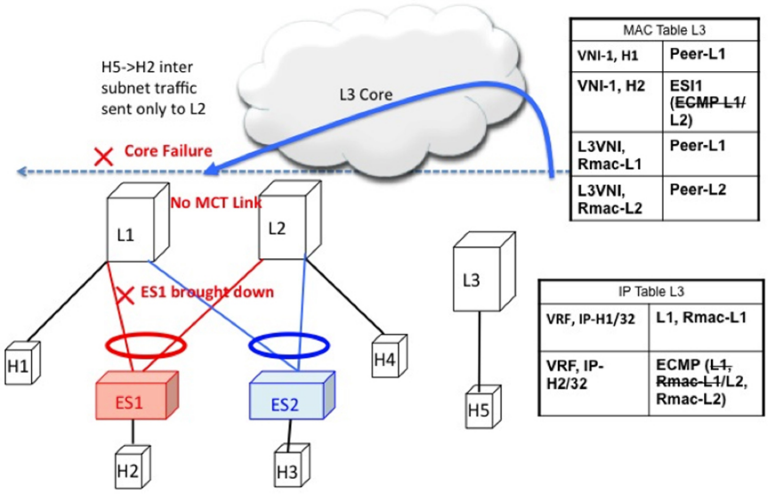
図 *11* : レイヤ *3 VXLAN* ゲートウェイ。*ESI* の障害により、*L2 ECMP* にのみ影響を与える *ES* の一括撤回が発生します。*L2 ECMP* は、*Type2* が取り消されるまで続行されます。*L3* トラフィックは、それまで最適ではないパス *L3-> L1-> L2* を介して *H2* に到達します。



リモート ルーテッド トラフィックのコア障害

リモート ルーテッド トラフィックのコア障害は、リモート ブリッジド トラフィックのコア障害と同じように動作します。アンダーレイ ルーティング プロトコルはすべてのリモート スイッチから L1 のループバック到達可能性を取り消すため、L1 は MAC ECMP と IP ECMP の両方のリストから削除されます。

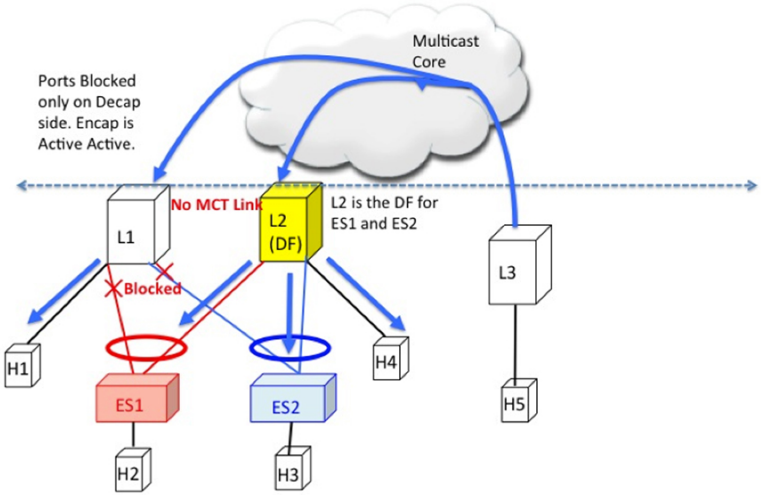
図 *12 : レイヤ 3 VXLAN* ゲートウェイ。コア障害。*L1* へのすべての *L3 ECMP* パスは、ルートの到達可能性がなくなるため、*L3* で取り消されます。



EVPN マルチホーミング BUM フロー

NX-OS は、ESI でアンダーレイのマルチキャスト コアをサポートします。H5 から発信される BUM トラフィックを検討します。BUM パケットは、VNI にマッピングされたマルチキャスト グループにカプセル化されます。L1 と L2 の両方が L2VNI マッピングに基づいてアンダーレイ グループの共有ツリー（*、G）に参加しているため、両方が BUM トラフィックのコピーを受信します。

図 *13 : L3* で発信される *BUM* トラフィック。*L2* は *ES1* および *ES2* の *DF* です。*L2* はカプセル化を解除し、*ES1*、*ES2*、およびオーファンに転送します。*L1* はカプセル化を解除し、オーファンにのみ転送します。



指定フォワーダ

冗長グループのスイッチの 1 つだけが ESI リンクを介して BUM トラフィックをカプセル化解除して転送することが重要です。この目的のために、イーサネットセグメントごとに一意の指

定フォワーダ（DF）が選択されます。DFの役割は、リモートセグメントから発信されたBUMトラフィックをカプセル化解除し、デバイスがDFである宛先ローカルセグメントに転送することです。DF 選択の主な側面は次のとおりです。

- DF 選択は（ES、VLAN）単位です。特定の VLAN の ES1 と ES2 に異なる DF を設定できます。
- DF の選択結果は、受信側の RX 側の BUM トラフィックにのみ適用されます。
- すべてのスイッチは、単一のホーム リンクまたはオーファン リンクに転送するために BUM トラフィックをカプセル化解除する必要があります。
- DFロールが重複すると、DHNでパケットまたはループが重複します。したがって、（ES、VLAN）ごとに一意の DF が必要です。

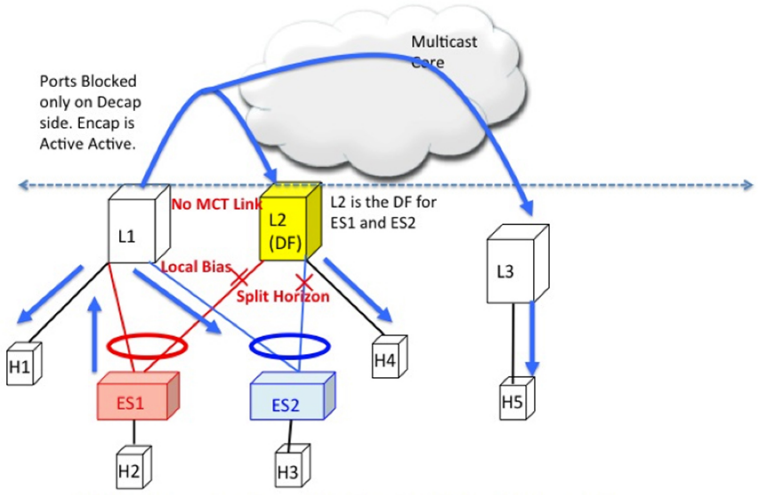
スプリット ホライズンとローカル バイアス

H2 から発信される BUM トラフィックを検討します。このトラフィックはL1 でハッシュされると考えてください。L1 は、このトラフィックをオーバーレイ マルチキャスト グループにカプセル化し、パケットをコアに送信します。同じマルチキャスト グループに参加し、同じ L2VNI を持つすべてのスイッチがこのパケットを受信します。また、L1 は、直接接続されているすべてのオーファン ポートおよびESI ポートでBUMパケットをローカルに複製します。たとえば、BUMパケットがES1から発信された場合、L1はそれをES2およびオーファン ポートにローカルに複製します。ローカルに接続されたすべてのリンクに複製するこの手法は、ローカル バイアスと呼ばれます。

リモート スイッチはDF の状態に基づいて、それをカプセル化解除し、ESI およびオーファンリンクに転送します。ただし、このパケットは、発信側スイッチ L1 と同じ冗長グループに属するL2でも受信されます。L2は、オーファン ポートに送信するためにパケットのカプセル化を解除する必要があります。ただし、L2 が ES1 のDFであっても、L2 はこのパケットを ES1 リンクに転送してはなりません。このパケットは、L1 が local-bias を行ったため、ES1 を L1 と共有しているピアから受信されたものであり、ES2では重複コピーは受信されません。したがって、L2（DF）は、L1 と共有する ES1 および ES2 の L1-IP にスプリット ホライズン フィルタを適用します。このフィルタは、VLAN のコンテキストで適用されます。

マルチホーミングの設定

図 *14 : L1* で発信される *BUM* トラフィック。*L2* は *ES1* および *ES2* の *DF* です。ただし、*L1* は *ES1* と *ES2* を *L1* と共有するため、ここでスプリット ホライズン チェックを実行する必要があります。ただし、*L2*



イーサネット セグメント ルート (タイプ 4)

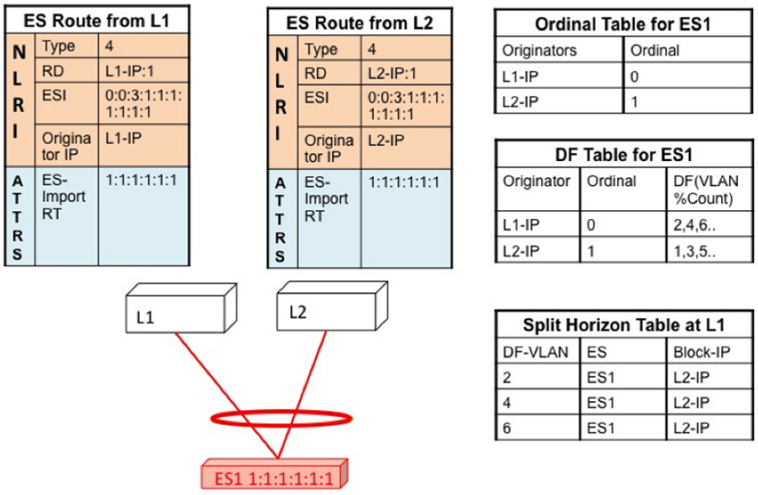
イーサネット セグメント ルートは、指定フォワーダを選択し、スプリット ホライズン フィルタリングを適用するために使用されます。イーサネットセグメントが設定されているすべてのスイッチは、このルートから発信されます。イーサネット セグメント ルートは、ESI が PC でローカルに設定されている場合にエクスポートおよびインポートされます。

イーサネット セグメント ルート (ルート タイプ 4)		
NLRI	Route Type（ルート タイプ）	イーサネットセグメント ルート (タイプ 4)
	RD	ルータ ID：Base＋ポート チャネル番号
	ESI	<Type: 1B><MAC: 6B><LD: 3B>
	発信側 IP	NVE ループバック IP
ATTRS	ES-Import RT	ESI から派生した 6 バイトの MAC

DF の選択と VLAN カービング

ESI の設定時に、L1 と L2 の両方が ES ルートをアドバタイズします。ESI MAC は L1 と L2 で共通であり、ネットワーク内で一意です。したがって、L1 と L2 だけが互いの ES ルートをインポートします。

図 *15 : VLAN%* カウントが順序と等しい場合は、*DF* ロールを使用します。



BUM トラフィックのコアおよびサイトの障害

ES1 に関連するアクセス リンクが L1 で失敗した場合、L1 は ES1 の ES ルートを取り消します。これは、DF の再計算をトリガーする変更につながります。L2 は順序テーブルに残っている唯一の TOR であるため、すべての VLAN の DF ロールを引き継ぎます。

Cisco Nexus 9000 シリーズ スイッチでの BGP EVPN マルチホーミングは、最小限の運用コストと配線コスト、プロビジョニングのシンプルさ、フローベースのロードバランシング、マルチパス、およびフェール セーフ冗長性を提供します。

ESI ARP 抑制の設定

ESI ARP 抑制の概要

イーサネット セグメント識別子（ESI）ARP 抑制は、VXLAN EVPN の ARP 抑制ソリューションを拡張したものです。データセンターの ARP ブロードキャストを大幅に削減することで、ESI マルチホーミング機能を最適化します。

ホストは通常、ARP 要求で VLAN をフラッディングします。ARP キャッシュをリーフ スイッチでローカルに維持することで、このフラッディングを最小限に抑えることができます。ARP キャッシュは次によって構築されます。

- すべての ARP パケットをスヌーピングし、要求からの送信元 IP アドレスと MAC バインディングを ARP キャッシュに入力する
- BGP EVPN IP または MAC ルート アドバタイズメントによる IP ホストまたは MAC アドレス情報の学習

ESI ARP 抑制では、最初の ARP 要求がすべてのサイトにブロードキャストされます。ただし、後続の ARP 要求は最初のホップ リーフ スイッチで抑制され、可能な場合はローカルに応答さ

れます。このように、ESI ARP 抑制により、オーバーレイ全体の ARP トラフィックが大幅に削減されます。キャッシュ ルックアップが失敗し、応答をローカルに生成できない場合は、ARP 要求をフラッディングできます。これは、サイレント ホストの検出に役立ちます。

ESI ARP 抑制は VNI（L2 VNI）単位の機能であり、VXLAN EVPN（分散ゲートウェイ）でのみサポートされます。この機能は L3 モードでのみサポートされています。

ESI ARP 抑制の制限事項

ESI ARP 抑制については、次の制限事項を参照してください。

- ESI マルチホーミング ソリューションは、リーフの Cisco Nexus 9300 シリーズ スイッチでのみサポートされます。
- ESI ARP 抑制は、L3 [SVI] モードでのみサポートされます。
- ESI ARP 抑制キャッシュの制限は 64K で、ローカルとリモートの両方のエントリが含まれます。

ESI ARP 抑制の設定

ARP 抑制 VACL を機能させるには、**hardware access-list tcam region arp-ether 256** CLI コマンドを使用して TCAM カービングを設定します。

```
Interface nve1
   no shutdown
   source-interface loopback1
   host-reachability protocol bgp
   member vni 10000
       suppress-arp
   mcast-group 224.1.1.10
```

ESI ARP 抑制の Show コマンドの表示

ESI ARP 抑制については、次の Show コマンドの出力を参照してください。

```
switch# show ip arp suppression-cache ?
 detail        Show details
   local          Show local entries
   remote     Show remote entries
   statistics  Show statistics
   summary Show summary
   vlan           L2vlan

switch# show ip arp suppression-cache local

Flags: + - Adjacencies synced via CFSoE
       L - Local Adjacency
       R - Remote Adjacency
       L2 - Learnt over L2 interface
      PS - Added via L2RIB, Peer Sync
      RO - Dervied from L2RIB Peer Sync Entry
```

```
Ip Address      Age      Mac Address    Vlan Physical-ifindex    Flags
Remote Vtep Addrs

61.1.1.20       00:07:54 0000.0610.0020  610 port-channel20     L
61.1.1.30       00:07:54 0000.0610.0030  610 port-channel2       L[PS RO]
61.1.1.10       00:07:54 0000.0610.0010  610 Ethernet1/96        L


switch# show ip arp suppression-cache remote
 Flags: + - Adjacencies synced via CFSoE
        L - Local Adjacency
        R - Remote Adjacency
        L2 - Learnt over L2 interface
          PS - Added via L2RIB, Peer Sync
          RO - Dervied from L2RIB Peer Sync Entry
          Ip Address      Age      Mac Address         Vlan    Physical-ifindex    Flags
      Remote Vtep Addrs
 61.1.1.40       00:48:37 0000.0610.0040   610     (null)                      R
    VTEP1, VTEP2.. VTEPn


switch# show ip arp suppression-cache detail
       Flags: + - Adjacencies synced via CFSoE
       L - Local Adjacency
       R - Remote Adjacency
       L2 - Learnt over L2 interface
       PS - Added via L2RIB, Peer Sync
       RO - Derived from L2RIB Peer Sync Entry

             Ip Address      Age      Mac Address    Vlan Physical-ifindex    Flags
        Remote Vtep Addrs
 61.1.1.20       00:00:07 0000.0610.0020  610 port-channel20     L
 61.1.1.30       00:00:07 0000.0610.0030  610 port-channel2       L[PS RO]
 61.1.1.10       00:00:07 0000.0610.0010  610 Ethernet1/96        L
 61.1.1.40       00:00:07 0000.0610.0040  610 (null)                    R
    VTEP1, VTEP2.. VTEPn


switch# show ip arp suppression-cache summary
 IP ARP suppression-cache Summary
 Remote          :1
 Local              :3
 Total               :4
switch# show ip arp suppression-cache statistics
 ARP packet statistics for suppression-cache
 Suppressed:
 Total 0, Requests 0, Requests on L2 0, Gratuitous 0, Gratuitous on L2 0
 Forwarded :
 Total: 364
  L3 mode :      Requests 364, Replies 0
  Request on core port 364, Reply on core port 0
                 Dropped 0
  L2 mode :      Requests 0, Replies 0
                 Request on core port 0, Reply on core port 0
                 Dropped 0
 Received:
 Total: 3016
 L3 mode:       Requests 376, Replies 2640
  Local Request 12, Local Responses 2640
                 Gratuitous 0, Dropped 0
  L2 mode :      Requests 0, Replies 0
                 Gratuitous 0, Dropped 0
```

```
switch# sh ip arp multihoming-statistics vrf all
ARP Multihoming statistics for all contexts
Route Stats
============
 Receieved ADD from L2RIB        :1756 | 1756:Processed ADD from L2RIB Receieved DEL
from L2RIB        :88 | 87:Processed DEL from L2RIB Receieved PC shut from L2RIB    :0
| 1755:Processed PC shut from L2RIB Receieved remote UPD from L2RIB :5004 | 0:Processed
 remote UPD from L2RIB
ERRORS
=======
Multihoming ADD error invalid flag         :0
Multihoming DEL error invalid flag         :0
Multihoming ADD error invalid current state:0
Multihoming DEL error invalid current state:0
Peer sync DEL error MAC mismatch           :0
Peer sync DEL error second delete          :0
Peer sync DEL error deleteing TL route     :0
True local DEL error deleteing PS RO route :0

switch#
```

VLAN 整合性検査の設定

VLAN 整合性チェックの概要

一般的なマルチホーミング展開シナリオでは、VLAN X に属するホスト 1 はアクセス スイッチにトラフィックを送信し、アクセス スイッチは VTEP1 と VTEP2 の両方のアップリンクにトラフィックを送信します。アクセス スイッチには、VTEP1 および VTEP2 の VLAN X 設定に関する情報はありません。VTEP1 または VTEP2 で VLAN X の設定が一致しないと、ホスト 1 のトラフィックが部分的に失われます。VLAN の整合性チェックは、このような設定の不一致の検出に役立ちます。

VLAN の整合性チェックには、CFSoIP が使用されます。Cisco Fabric Services（CFS）は、同じネットワーク内のスイッチ間でデータを交換するための共通インフラストラクチャを提供します。CFS にはネットワーク内の CFS 対応スイッチを検出し、すべての CFS 対応スイッチの能力を検出する機能が備わっています。IP を介した CFS（CFSoIP）を使用して、1 台のシスコデバイスまたはネットワークの他のすべてのシスコデバイスにコンフィギュレーションを配信し、同期させることができます。

CFSoIP は、管理 IP ネットワークのすべてのピアを検出します。EVPN マルチホーミング VLAN の整合性チェックでは、デフォルトの CFS マルチキャスト アドレスを **cfs ipv4 mcast-address** <*mcast address*> CLI コマンドの mcast address で上書きすることを推奨します。CFSoIP を有効にするには、**cfs ipv4 distribute** CLI コマンドを使用する必要があります。

マルチホーミング ピアの 1 つでトリガー（たとえば、デバイスの起動、VLAN 設定の変更、VLAN の管理状態の変更）が発行されると、イーサネット セグメント（ES）がすべての CFS ピアに送信されます。

ブロードキャスト要求を受信すると、リクエスタと同じ ES を共有するすべての CFS ピアが、その VLAN リスト（ES ごとに設定され、管理上 VLAN リスト）を返します。VLAN 整合性チェックは、ブロードキャスト要求または応答を受信すると実行されます。

ブロードキャスト要求を送信する前に、15 秒のタイマーが開始されます。ブロードキャスト要求または応答を受信すると、ローカル VLAN リストが ES ピアのリストと比較されます。一致しない VLAN は中断されます。新しく一致した VLAN が一時停止されなくなります。

VLAN 整合性チェックは、次のイベントに対して実行されます。

- グローバル VLAN 設定：Add、Delete、Shut、または no shut イベント。

  ポート チャネル VLAN の設定：トランクの許可または削除、または VLAN の変更。

- CFS イベント：CFS ピアが追加または削除されるか、CFSoIP 設定が削除されます。
- ES ピア イベント：ES ピアが追加または削除されました。

応答が受信されない場合、ブロードキャスト要求は再送信されます。3 回の再送信後に応答が受信されない場合、VLAN 整合性チェックは失敗します。

VLAN の整合性チェックの注意事項と制限事項

VLAN の整合性チェックについては、次の注意事項と制限事項を参照してください。

- VLAN の整合性チェックは CFSoIP を使用します。管理インターフェイスを介したアウトオブバンド アクセスは、ネットワーク内のすべてのマルチホーミング スイッチで必須です。
- デフォルトの CFS マルチキャスト アドレスをCLI **cfs ipv4 mcast-address** <*mcast address*> コマンドで上書きすることを推奨します。
- VLAN 整合性チェックは、**switchport trunk native vlan** 設定の不一致を検出できません。
- CFSoIP と CFSoE は同じデバイスで使用しないでください。
- CFSoIP は、VLAN 整合性チェックに使用されないデバイスでは使用しないでください。
- VLAN の整合性チェックに参加しないデバイスで CFSoIP が必要な場合は、CLI **cfs ipv4 mcast-address** <*mcast address*> コマンドを使用して、VLAN の整合性に参加するデバイスに別のマルチキャスト グループを設定する必要があります。

VLAN 整合性検査の設定

デフォルトの CFS マルチキャストアドレスを上書きするには、CLI コマンドの **cfs ipv4 mcast-address** <*mcast address*> を使用します。**cfs ipv4 distribute** CLI コマンドを使用して、CFSoIP を有効にします。

VLAN 整合性チェックを有効または無効にするには、**evpn esi multihoming** モードで追加された新しい **vlan-consistency-check** CLI コマンドを使用します。

```
switch (config)# sh running-config | in cfs
cfs ipv4 mcast-address 239.255.200.200
cfs ipv4 distribute

switch# sh run | i vlan-consistency
evpn esi multihoming
    vlan-consistency-check
```

VLAN 整合性チェックの show コマンド出力の表示

VLAN の整合性チェックについては、次の show コマンドの出力を参照してください。

CFS ピアを一覧表示するには、**sh cfs peers name nve** CLI コマンドを使用します。

```
switch# sh cfs peers name nve

Scope      : Physical-ip
--------------------------------------------------------------------------
 Switch WWN              IP Address
--------------------------------------------------------------------------
 20:00:f8:c2:88:23:19:47 172.31.202.228                          [Local]
                         Switch
 20:00:f8:c2:88:90:c6:21 172.31.201.172                          [Not Merged]
 20:00:f8:c2:88:23:22:8f 172.31.203.38                           [Not Merged]
 20:00:f8:c2:88:23:1d:e1 172.31.150.132                          [Not Merged]
 20:00:f8:c2:88:23:1b:37 172.31.202.233                          [Not Merged]
 20:00:f8:c2:88:23:05:1d 172.31.150.134                          [Not Merged]
```

show nve ethernet-segment コマンドを発行すると、次の詳細が表示されます。

- 整合性チェックに失敗した VLAN のリスト。
- グローバル VLAN CC タイマーの残りの値（秒単位）。

```
switch# sh nve ethernet-segment
ESI Database
----------------------------------------
ESI: 03aa.aaaa.aaaa.aa00.0001,
   Parent interface: port-channel2,
  ES State: Up
  Port-channel state: Up
  NVE Interface: nve1
   NVE State: Up
   Host Learning Mode: control-plane
  Active Vlans: 3001-3002
   DF Vlans: 3002
   Active VNIs: 30001-30002
  CC failed VLANs: 0-3000,3003-4095
  CC timer status: 10 seconds left
  Number of ES  members: 2
  My ordinal: 0
  DF timer start time: 00:00:00
  Config State: config-applied
  DF List: 201.1.1.1 202.1.1.1
  ES route added to L2RIB: True
```

```
  EAD routes added to L2RIB: True
```

次の Syslog 出力を参照してください。

```
switch(config)# 2017 Jan ?7 19:44:35 Switch %ETHPORT-3-IF_ERROR_VLANS_SUSPENDED: VLANs
2999-3000 on Interface port-channel40 are being suspended.
(Reason: SUCCESS)

After Fixing configuration
2017 Jan ?7 19:50:55 Switch %ETHPORT-3-IF_ERROR_VLANS_REMOVED: VLANs 2999-3000 on Interface
 port-channel40 are removed from suspended state.
```

マルチホーミングの設定

翻訳について
このドキュメントは、米国シスコ発行ドキュメントの参考和訳です。リンク情報につきましては、日本語版掲載時点で、英語版にアップデートがあり、リンク先のページが移動/変更されている場合がありますことをご了承ください。あくまでも参考和訳となりますので、正式な内容については米国サイトのドキュメントを参照ください。