ホワイトペーパー
Cisco public

Cisco ACI マルチサイト アーキテクチャ

目次

はじめに

Cisco Application Centric Infrastructure（Cisco ACI）がファブリックテクノロジーとして普及するのに伴い、個別の Cisco ACI ファブリックを相互接続することが企業やサービスプロバイダーの共通の課題となっています。事業の継続や災害の回避といったビジネス要件によってデータセンターファブリックが個別に展開されているため、それらを相互接続する必要が生じています。このドキュメントで説明するように、これらのファブリックは、使用される展開オプションの違いにより、ポッド、ファブリック、サイトと呼ばれます。

注： このドキュメントに示されている設計を十分理解するには、Cisco ACI とその機能、単一のサイトまたは単一のポッドで運用する場合の設計方法について、少なくとも基本的な知識を持っている必要があります。詳細は、https://www.cisco.com/c/ja_jp/solutions/data-center-virtualization/application-centric-infrastructure/white-paper-listing.html にある Cisco ACI のホワイトペーパーを参照してください。

図 1 は、異なる ACI ネットワーク間で接続とポリシー適用を拡張するためのアーキテクチャオプションについて、Cisco ACI の発売から現在までの変遷を示しています。

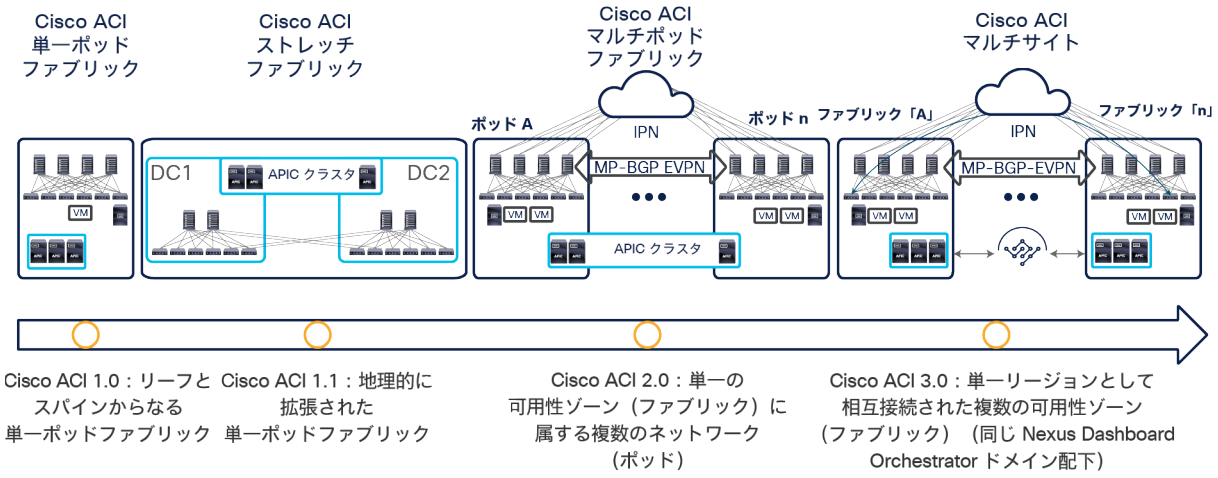


図 1.
Cisco ACI の接続オプションとポリシードメインの進化

- 最初のオプションは、Cisco ACI リリース 1.0 以降利用可能で、リーフとスパインからなる従来型の 2 階層ファブリック（単一のポッド）から構成されています。この構成では、展開されたすべてのリーフノードと展開されたすべてのスパインノードがフルメッシュで接続されている構造となっています。Cisco ACI コントロール プレーン プロトコルの単一のインスタンスが、ポッド内のすべてのネットワークデバイス間で実行されます。ポッド全体が単一の Cisco Application Policy Infrastructure Controller（APIC）クラスタの管理下にあり、このクラスタがポリシー定義の単一のポイントでもあります。
- Cisco ACI の進化の第 2 段階では、ポッドが複数の物理データセンターのロケーションに地理的に拡張されます（ストレッチファブリック）。ただし、通常は、同一の都市圏に展開されます。これらのロケーション間で十分な光ファイバ接続が確保できない場合、ストレッチファブリックでは部分的なメッシュトポロジが使用されます。このトポロジでは、一部のリーフノード（トランジットリーフノードと呼ばれます）がローカルとリモートの両方のスパインノードに接続され、残りのリーフノードはローカルのスパインノードにのみ接続されます。部分的なメッシュ接続を使用しているにもかかわらず、ストレッチファブリックは機能的には依然として単一のポッド展開となっていて、すべての Cisco ACI コントロール プレーン プロトコルの単

一のインスタンスが、相互接続されたすべてのデータセンターサイトで実行されるため、単一の障害ドメインが作成されます。

注：Cisco ACI ストレッチファブリックの展開オプションの詳細は、https://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/apic/sw/kb/b_kb-aci-stretched-fabric.html を参照してください。

- 単一のネットワーク障害ドメインがストレッチ ファブリック トポロジ全体に拡張されることに対する懸念に対処するために、Cisco ACI リリース 2.0 では、Cisco ACI マルチポッドアーキテクチャが導入されました。このモデルでは、別々の Cisco ACI ポッドを展開する必要があります。各ポッドでは、コントロール プレーン プロトコルの別々のインスタンスが実行され、ポッド同士は、外部 IP ルーテッドネットワーク（またはポッド間ネットワーク（IPN））を介して相互接続されます。Cisco ACI マルチポッド設計では、ポッド全体に展開されたノードは、すべて同じ APIC クラスタの管理下にあり、機能的には単一のファブリックです。それでも、ポッドにまたがるネットワークレベルで完全な復元力が実現します。したがって、各ポッドは別々の可用性ゾーンと見なすことができます。同じ APIC クラスタの管理下にあるすべてのポッドは、同じファブリック（リージョン）に属します。

  Cisco ACI マルチポッド設計の主なメリットは、運用が簡単なことです。個別のポッドが複数あっても、論理的に単一のエンティティであるかのように管理されます。このアプローチでは、単一のポッド展開において使用可能なすべての Cisco ACI 機能（ネットワーク サービス チェーン、マイクロセグメンテーション、Virtual Machine Manager（VMM）ドメイン統合など）をポッドにまたがってシームレスに展開できます。これは、このアーキテクチャが実現する固有の価値です。ただし、Cisco ACI マルチポッドアーキテクチャは単一のファブリック（APIC ドメイン）として管理されるため、単一のテナントの変更ドメインとなります。あるテナントのコンテキストで適用された構成やポリシーの変更は、すべてのポットにわたってただちに適用されることに注意が必要です。この動作がマルチポッド設計の運用が簡単であることの理由の 1 つですが、構成エラーが伝播する懸念も引き起こします。

  注：変更はすべてのポッドにすぐに適用されますが、所定のテナントのコンテキストのみにとどまります。Cisco ACI ファブリックは暗黙のマルチテナントとなっているため、個別のテナントに展開されたすべてのリソースは完全に分離され、エラーや中断といったイベントから保護されます。Cisco ACI マルチポッド設計の詳細は、https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-737855.html を参照してください。

  さらに、Cisco ACI リリース 2.3(1) 以降では、ポッド間の遅延が最大 50 ミリ秒 RTT まで許容されます。それより前の Cisco ACI リリースでは、10 ミリ秒 RTT が上限です。

- 個別の Cisco ACI ネットワーク間で完全な分離（ネットワークレベルとテナントの変更ドメインレベルの両方）が求められたことから、Cisco ACI マルチサイトアーキテクチャが生まれ、Cisco ACI リリース 3.0(1) で導入されました。このアーキテクチャがこのドキュメントの主なテーマであり、以降のセクションで詳しく説明します。

- ACI マルチサイトの同じアーキテクチャアプローチが拡張され、オンプレミスの ACI ファブリックとパブリッククラウドリソースを接続し、ポリシーを拡張することも可能になっています（このドキュメントの執筆時点で、AWS および Azure との統合が可能です）。

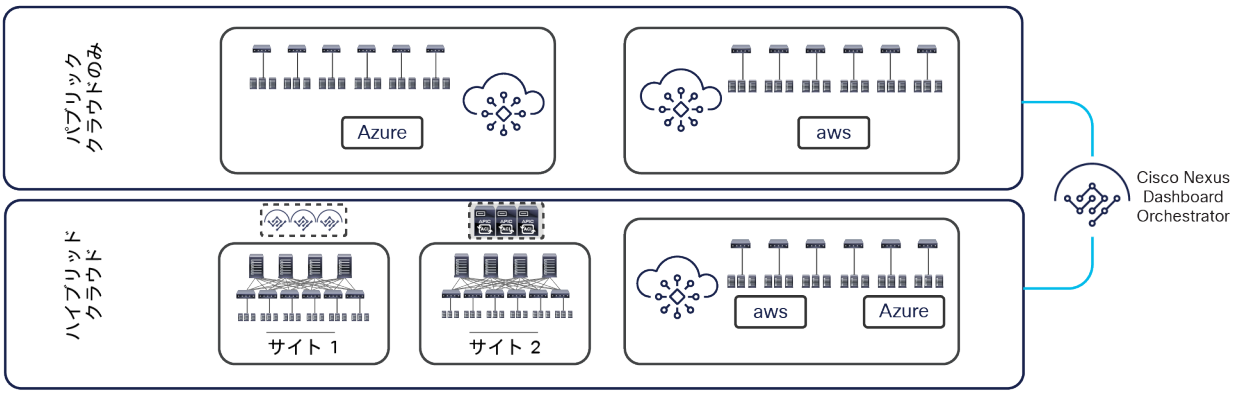


図 2.
ハイブリッドクラウドへの展開とパブリッククラウドのみへの展開のオプションをサポート

上図のように、ハイブリッドクラウド（つまり、パブリッククラウドリソースに接続するオンプレミスの ACI ファブリック）のシナリオとパブリッククラウドのみのシナリオの両方が現在サポートされています。これらの展開オプションについて詳しく説明することは、このホワイトペーパーの範囲外です。詳細は、以下のホワイトペーパーを参照してください。

https://www.cisco.com/c/ja_jp/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-741998.html

https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-742844.html

Cisco ACI マルチサイト設計の詳細に入る前に、シスコがマルチポッドアーキテクチャとマルチサイトアーキテクチャの両方を使用する理由と、それらが相互に補完しあってさまざまなビジネス要件を満たすように配置する方法を理解する必要があります。まず、このドキュメントで使用されている主な用語と、AWS パブリッククラウドの導入で頻繁に使用される命名規則を理解する必要があります。

- **ポッド：**ポッドとは、共通のコントロールプレーン（Intermediate System-to-Intermediate System（ISIS）プロトコル、Border Gateway Protocol（BGP）、Council of Oracle Protocol（COOP）など）を共有するリーフとスパインからなるネットワークを意味します。したがって、ポッドは AWS でいう可用性ゾーンに相当する単一のネットワーク障害ドメインです。
- **ファブリック：**ファブリックとは、同じ APIC ドメインの管理下にあるリーフノードとスパインノードのセットを意味します。各ファブリックが、それぞれテナントの変更ドメインになります。APIC で適用されたすべての構成とポリシーの変更は、ファブリック全体に広がる所定のテナントに適用されるためです。したがって、ファブリックは AWS でいうリージョンに相当します。
- **マルチポッド：**マルチポッド設計とは、リーフとスパインからなるネットワーク（ポッド）が複数相互に接続された単一の APIC ドメインからなるアーキテクチャを意味します。結果的に、マルチポッド設計は機能的にはファブリック（可用性ゾーンの相互接続）になりますが、単一のネットワーク障害ドメインになるわけではありません。各ポッドが、コントロール プレーン プロトコルの別々のインスタンスを実行するためです。したがって、マルチポッドファブリックは、異なる AWS 可用性ゾーンを相互接続する AWS リージョンに相当します。
- **マルチサイト：**マルチサイト設計とは、複数の APIC クラスタドメインがそれに関連するポッドとともに相互接続されるアーキテクチャを意味します。マルチサイト設計は、それぞれが単一のポッドまたは複数のポッ

ド（マルチポッド設計）として展開された個別のリージョン（ファブリック）を相互接続するため、マルチファブリック設計と呼ぶこともできます。

注：「マルチファブリック設計」を「デュアルファブリック設計」と混同しないでください。前者はこのドキュメントで説明しているマルチサイトアーキテクチャを指していて、後者はマルチサイト設計の前身を指しています。複数の ACI ファブリックを運用する場合は、リーフスイッチを介して個々の ACI ファブリックを相互接続する（デュアルファブリック設計）代わりに、マルチサイトを展開することを強くお勧めします。後者の方法は、マルチサイト機能が登場する前には選択の余地がなかったと思われますが、現在は公式にはサポートされていません。特に、個別に導入された、レイヤ 2 ドメインをサイトにまたがって拡張する Data Center Interconnect（DCI）テクノロジー（OTV、VPLS など）とこのトポロジを併用する場合の検証と品質保証検査が実施されていないためです。Cisco ACI には ACI ファブリックを相互接続するための機能として、マルチサイトに先立つ共通パーベイシブゲートウェイと呼ばれる機能があります。しかし、レイヤ 2 を APIC ドメインにまたがって拡張する必要がある場合は、上記の理由からマルチサイトを用いた新しい ACI マルチファブリック展開を設計することを強くお勧めします。

可用性ゾーンやリージョンといった AWS の構造を理解することは、シスコが Cisco ACI マルチポッド設計をすでに提供した後でマルチサイトアーキテクチャへの投資を決定した理由を理解するために不可欠です。通常、組織は、別々のリージョンを構成するデータセンターファブリックにまたがってアプリケーションのさまざまなインスタンスを展開する必要があります。1 つのリージョンで発生するネットワークレベルの障害、構成のエラー、ポリシー定義のエラーが、別のリージョンで実行されているアプリケーションのワークロードに伝播しないようにするためには、このような構成が不可欠です。これによって、災害回避とディザスタリカバリの両方の機能が強化されます。

Cisco ACI マルチポッドおよびマルチサイトのアーキテクチャを組み合わせて展開することで、2 つの異なる要件を満たすことができます。柔軟な Cisco ACI の独立したネットワークのグループを作成することができます。これらのネットワークは単一の論理構成体（ファブリックまたはリージョン）として監視と運用が可能で、これを用いてアプリケーションの機能コンポーネントを従来型のアクティブ/アクティブモデルで展開できます（言い換えると、アプリケーション階層を構成する異なるエンドポイントを、同一のファブリックに属する可用性ゾーンにまたがって自由に展開できます。次に、これらのファブリックは信頼性の高い相互接続と拡張が可能なため、異なるアプリケーション インスタンスを別々のリージョンに展開することも、リージョン間で完全なアプリケーションリカバリ機能を実装することもできます。前者は、災害回避の要件を充足するために用いられるアプリケーションごとのアクティブ/アクティブ導入モデルであり、後者はディザスタリカバリのユースケースです。図 3 を参照してください。

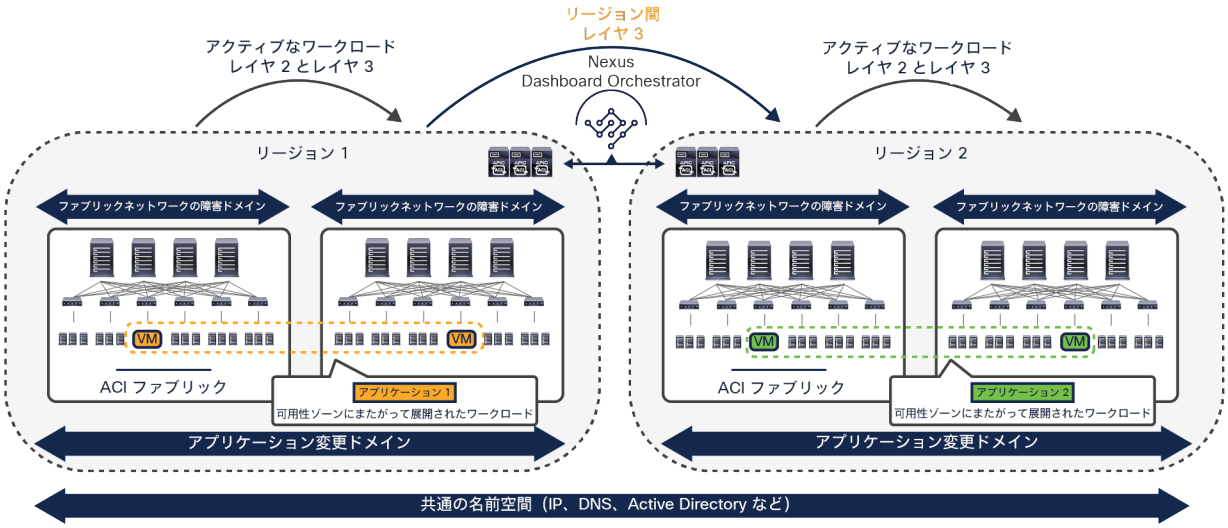


図 3.
変更ドメインとネットワーク障害ドメインの分離

注： 上図の Cisco ACI マルチポッドおよびマルチサイトのアーキテクチャを併用した展開は、Cisco ACI リリース 3.2(1) 以降でサポートされています。

小規模な展開では、データセンターの同じロケーションを 2 つ用意し、この 2 つを災害回避とディザスタリカバリの両方の目的に使用することも非常に一般的です。ACI マルチポッドと ACI マルチサイトを組み合わせると、このような要件にも対処でき、サイトにまたがる従来型のアクティブ/アクティブモデルによるアプリケーションの展開と、ディザスタリカバリのシナリオで必要な一般的なアプリケーション リカバリ メカニズムを同時に実現します（図 4）。

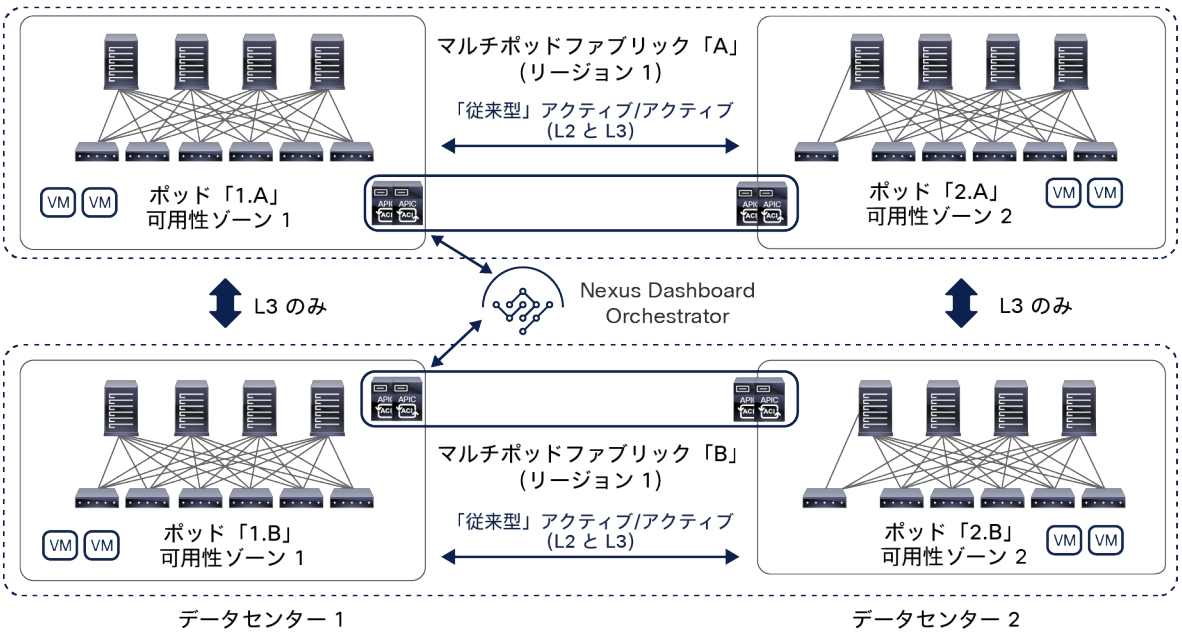


図 4.
Cisco ACI マルチポッドおよびマルチサイトのアーキテクチャを組み合わせた展開

前の 2 つの図に示した展開では、個別のファブリック（リージョン）を レイヤ 3 で相互接続する手段としてマルチサイトを使用し、レイヤ 2 を拡張するサービスの提供はマルチポッドに委ねています。これが望ましい推奨モデルです。とはいえ、このホワイトペーパーで後述するように、マルチサイトにもネイティブのレイヤ 2 拡張機能があります。通常ならマルチポッドの採用を検討する特定のユースケースに対処するために、このアーキテクチャを適用できます。その際、機能面で制約が生じる可能性があることを考慮に入れることが重要です（FW または SLB クラスタをマルチサイトに統合する場合など）。

このドキュメントの残りの部分では、Cisco ACI マルチサイトアーキテクチャに焦点を当て、初めに機能コンポーネントの概要を説明します。

Cisco ACI マルチサイトアーキテクチャ

Cisco ACI マルチサイトアーキテクチャの全体図を図 5 に示します。

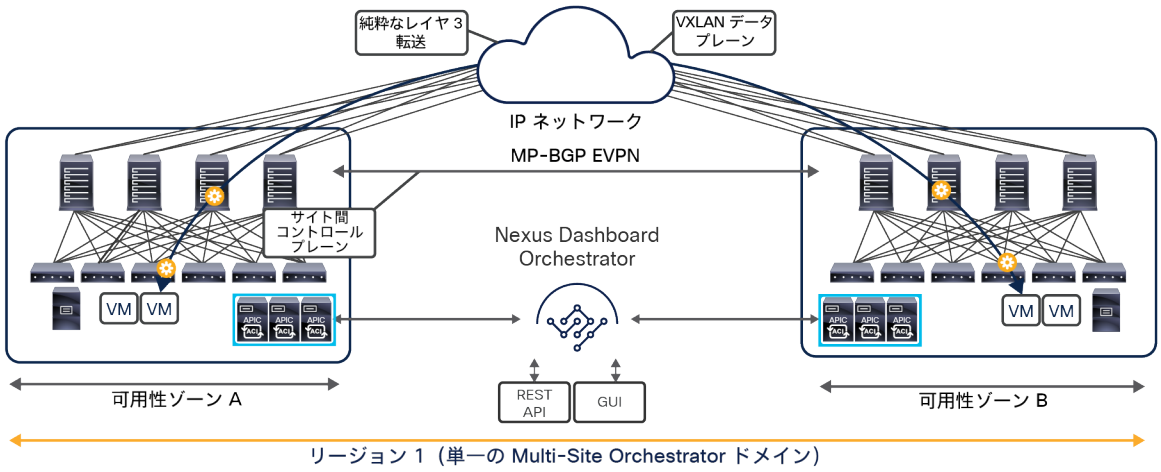


図 5.
Cisco ACI マルチサイトアーキテクチャ

Cisco ACI マルチサイトは、個別の Cisco APIC クラスタドメイン（ファブリック）を相互接続できるアーキテクチャです。これによって、異なるリージョンを構成する各ドメインがすべて、同じ Cisco ACI マルチサイトドメインに属することになります。その結果、マルチテナントのレイヤ 2 とレイヤ 3 のネットワークがサイト間で接続され、ポリシードメインがシステム全体にエンドツーエンドで拡張されます。

注：

この設計は、以下の機能コンポーネントによって実現されます。

- Cisco Nexus Dashboard Orchestrator（NDO）：このコンポーネントは、サイトにまたがってポリシーを管理します。このコンポーネントが提供する一元的な管理によって、相互接続されたすべてのサイトの正常性スコアをモニタリングできます。また、すべてのサイト間ポリシーを 1 か所で定義した後、異なる APIC ドメインにプッシュして、そのファブリックを構成する物理スイッチにレンダリングできます。これらのポリシーをいつどこにプッシュするかを高度に制御できるため、変更ドメインの分離が可能になります。これは、Cisco ACI マルチサイトアーキテクチャ独自の特徴です。

  Orchestrator ソフトウェアリリース 3.2(1) より前は、このコンポーネントは Multi-Site Orchestrator（MSO）という名前でした。新しいリリースでは、Cisco Nexus Dashboard（ND）コンピューティング プラット

フォームで実行されるアプリケーションとしてのみサポートされるため、Nexus Dashboard Orchestrator（NDO）に改名されました。ただし、シスコのドキュメントでは、「Multi-Site Orchestrator（MSO）」、「Nexus Dashboard Orchestrator（NDO）」、または単に「Orchestrator サービス」と、区別せずに呼ばれている場合があります。これらの名前はすべて、Cisco Multi-Site アーキテクチャの同じ機能コンポーネントを指します。

Cisco Nexus Dashboard Orchestrator の詳細については、「Cisco Nexus Dashboard Orchestrator」のセクションを参照してください。

- サイト間コントロールプレーン：エンドポイントの到達可能性情報は、マルチプロトコル BGP（MP-BGP）イーサネット VPN（EVPN）コントロールプレーンを使用してサイト間で交換されます。このアプローチにより、サイト間で通信するエンドポイントの MAC および IP アドレス情報が交換されます。MP-BGP EVPN セッションは、Cisco Nexus Dashboard Orchestrator の同じインスタンスによって管理される別々のファブリックに展開されたスパインノード間で確立されます。詳細については、「Cisco ACI マルチサイトのオーバーレイ コントロール プレーン」セクションを参照してください。
- サイト間データプレーン：異なるサイトに接続されているエンドポイント間のすべての通信（レイヤ 2 またはレイヤ 3）は、さまざまなサイトを相互接続する汎用 IP ネットワークを通過するサイト間仮想拡張 LAN（VXLAN）トンネルを確立することによって実現されます。「サイト間ネットワーク（ISN）の展開に関する考慮事項」セクションで説明するように、この IP ネットワークには、ルーティングのサポートと最大伝送ユニット（MTU）サイズの拡大（VXLAN カプセル化によるオーバーヘッドを考慮）以外に、特定の機能要件はありません。

注： Nexus Dashboard Orchestrator ソフトウェアリリース 4.0(1) 以降、新しい導入モデルがサポートされ、NDO を導入して最大 100 の「自律型ファブリック」を管理できます。このようなユースケースでは、マルチサイトドメインに属するファブリック間に VXLAN EVPN によるサイト間接続が存在せず、基本的に Orchestrator がこれらすべてのサイトに構成を一元的にプロビジョニングします。L3Out データパスを利用することで、ファブリック間ではレイヤ 3 のみの通信が可能です。NDO の導入と「自律型ファブリック」の詳細については、「レイヤ 3 のみのサイト間接続」セクションを参照してください。

サイト間 VXLAN カプセル化を使用すると、サイト間 IP ネットワークに必要な構成と機能が大幅に簡素化されます。また、ネットワークとポリシーの情報（メタデータ）をサイト間で伝送することもできます（図 6）。

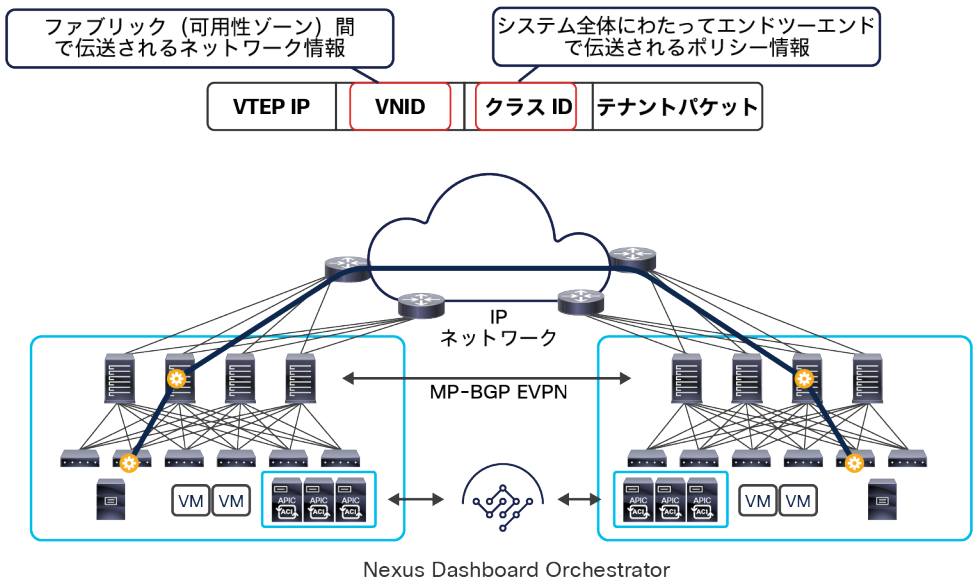


図 6.
ネットワークとポリシーの情報をサイト間で伝送

VXLAN ネットワーク識別子（VNID）は、レイヤ 2 の通信の場合、ブリッジ ドメイン（BD）を識別します。レイヤ 3 のトラフィックの場合、仮想ルーティングおよび転送（VRF）内通信のトラフィックを送信するエンドポイントの VRF インスタンスを識別します。クラス ID は、送信元エンドポイントグループ（EPG）の一意の識別子です。ただし、これらの値はファブリック内でのみ有効です。完全に独立した APIC ドメインおよびファブリックが宛先サイトに展開されているため、宛先サイト内でトラフィックを転送する前に、変換機能（「名前空間の正規化」とも呼ばれます）を適用する必要があります。これによって、宛先サイト内で有効で、送信元の EPG、ブリッジドメイン、VRF インスタンスを識別できる値が使用できるようになります。

この名前空間の正規化機能の必要性について理解を深めるには、別々のサイトに展開された 2 つの EPG 間でコントラクトが定義されるときの動作を明確にすることが重要です。このコントラクトは、それぞれの EPG に属するエンドポイント間でサイト間通信を行うために必要です。図 7 に示すように、目的の構成（意図）が Cisco Nexus Dashboard Orchestrator で定義され、異なる APIC ドメインにプッシュされると、シャドウ EPG と呼ばれる EPG のコピーが各 APIC ドメインに自動的に作成されます。これにより、NDO で一元的に定義された設定全体を各サイトでローカルにインスタンス化でき、各 EPG がローカルでのみ定義されサイトにまたがって拡張されていない場合でも、セキュリティポリシーが適切に適用されます（特定の VNID とクラス ID が、各 APIC ドメインのシャドウオブジェクトに割り当てられます）。

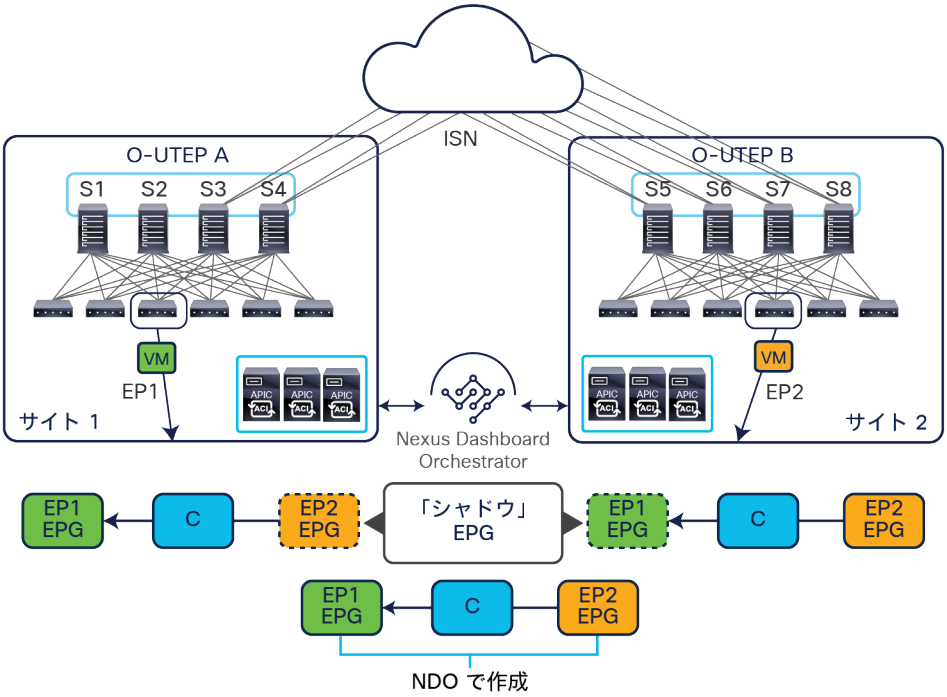


図 7.
シャドウ EPG の作成

上図の例では、黄色の EP2 EPG（およびそれに関連付けられた BD）が APIC ドメイン 1 に「シャドウ」EPG として作成され、緑色の EP1 EPG（およびそれに関連付けられた BD）が APIC ドメイン 2 に「シャドウ」EPG として作成されます。Cisco ACI リリース 5.0(1) までは、シャドウ EPG および BD を APIC GUI 上で簡単に区別できないため、それらの存在と役割を認識することが非常に重要です。

両方の APIC ドメインが互いに完全に独立しているため、サイト全体で見たとき、特定の EPG に異なる VNID とクラス ID の値が割り当てられる（「実際の」コピーと「シャドウ」コピー）と想定されます。そのため、下の図 8 に示すように、ローカルサイトにトラフィックを送る前に、リモートサイトからデータプレーントラフィックを受信するスパインでこれらの値を変換する必要があります。

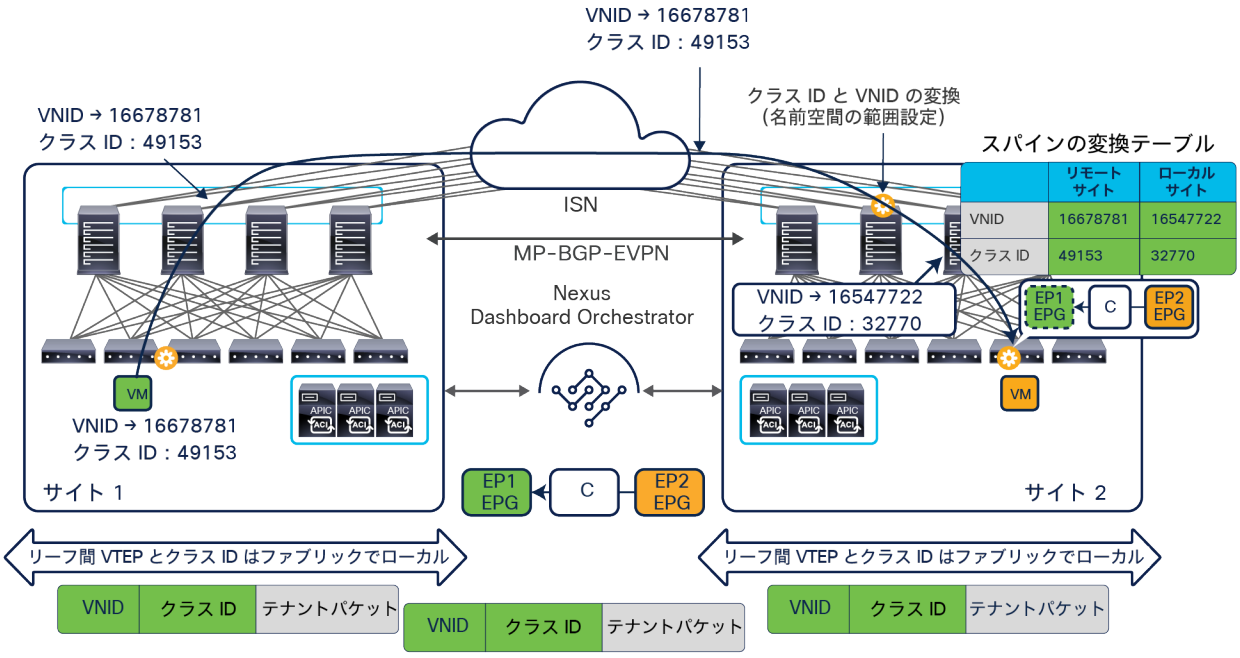


図 8.
受信スパインの名前空間の変換機能

「EP1 EPG」と「EP2 EPG」の通信が必要となるポリシーが Cisco Nexus Dashboard Orchestrator で作成されると、Nexus Dashboard Orchestrator は各 APIC コントローラから、ローカルオブジェクトとシャドウオブジェクトに割り当てられた識別子（pcTag、L2VNI、L3VNI）を受け取ります。さらに、ローカルスパインで適切な変換ルールをプログラミングするように APIC コントローラに指示します。その結果、トラフィックを宛先エンドポイントに送信する前に、構成されたポリシーがリーフノードに正しく適用されます。さらに、別々のサイトにローカルに展開された EPG 間でコントラクトを作成すると、通常、これらの EPG に関連付けられた BD サブネットもリモートサイトのリーフノード上に構成されます（ローカルスパインノードを介したプロキシパスを有効にするため）。詳細については、「Cisco ACI マルチサイトのオーバーレイデータプレーン」セクションを参照してください。

注： 図 8 の例は、宛先サイトのリーフノードでポリシーが適用される状況を示しています。通常、この状況は、送信元サイトのリーフノードがデータプレーンを介してリモートエンドポイントのロケーション情報を学習するまで続きます。学習が終わると、ポリシーが入力リーフで直接適用できるようになります。ポリシーベースリダイレクト（PBR）でのサービスグラフの使用は、これが常に当てはまるとは限らないことを示すケースです。詳細については、「ネットワークサービスの統合」セクションを参照してください。

この名前空間の変換機能は、サイト間通信のパフォーマンスに悪影響を与えないようにラインレートで実行する必要があります。これを実現するには、Cisco ACI マルチサイトアーキテクチャに展開されたスパインノードに特定のハードウェアを使用する必要があります。マルチサイト展開では、Cisco Nexus EX プラットフォーム（およびそれ以降）の世代のスパインスイッチのみがサポートされます。第 1 世代のスパインスイッチはスパインスイッチの新しいモデルと共存できることに注意してください。ただし、後者のみが外部 IP ネットワークに接続され、サイト間通信に使用されるようになっている必要があります（図 9）。

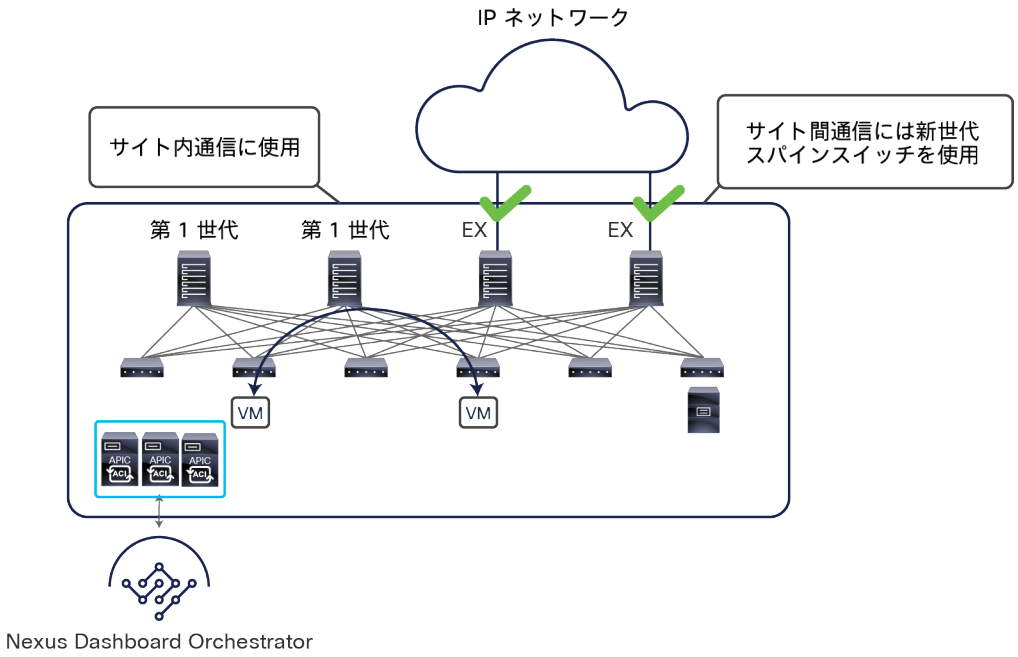


図 9.
第 1 世代スパインスイッチと EX プラットフォーム（およびそれ以降）のスパインスイッチの共存

図 9 の共存シナリオは、展開されたすべてのスパインを外部レイヤ 3 ドメインに接続する必要がないことも示しています。使用できる具体的なハードウェアと、復元力とスループットの要求レベルに基づいて、外部 IP ネットワークへの接続に使用するスパインとリンクの数を決定します。

注： 9332C プラットフォームと 9364C プラットフォームのようなモジュラ型でないスパインモデルの場合、ネイティブ 10G インターフェイス（SFP ベース）を使用してサイト間ネットワーク（ISN）デバイスに接続することもできます。

Cisco ACI マルチサイトアーキテクチャの導入により、相互接続されたファブリック全体に展開されるリーフノードとスパインノードの総数、さらにはエンドポイントの総数を拡大することもできます。これが可能なことが、Cisco ACI マルチサイトおよびマルチポッドの設計の主要な相違点の 1 つです。後者のオプションは、単一のファブリックとしての設計が抱える拡張性の制約に依然として縛られているためです。

注： Cisco ACI の展開を計画するときは、https://www.cisco.com/c/ja_jp/partners/partner-with-cisco/integrator/levels.html で提供されているスケーラビリティガイドを必ず参照してください。たとえば、Cisco ACI リリース 4.0(1) に適用されるスケーラビリティガイドについては、https://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/apic/sw/4-x/verified-scalability/Cisco-ACI-Verified-Scalability-Guide-401.html を参照してください。

EPG がサイトにまたがって拡張されている場合、スパインノードの変換テーブルにエントリが常に追加されます。これは、デフォルトで許可されている EPG 内通信を可能にするために必要です。一方、別々のサイトでローカルに定義された EPG に属するエンドポイント間でサイト間通信を確立する必要がある場合、それぞれの変換テーブルに正しくエントリを追加するために、EPG 間でコントラクトを定義する必要があります（この例は、上の図 7 と図 8 を参照してください）。したがって、サイト間ネットワーク（VXLAN データパス）を介してサイト間通信を確立する必要がある場合は、コントラクトと EPG を定義し、NDO でコントラクトを EPG 間に直接適用することが必須で

す。any-to-any 通信を可能にすることが目的である場合、非常にシンプルなコントラクト（「すべて許可」に相当）を作成して、異なるすべての EPG ペアに適用できることに注意してください。

2 つの追加機能（優先グループと vzAny）が Nexus Dashboard Orchestrator で利用できるようになりました。これを用いることで、EPG 間のポリシー定義の簡素化と、サイト間接続に必要なスパイン上の変換テーブルの適切なプログラミングが可能になります。これらの機能については、以下の 2 つのセクションで説明します。

注： 変換テーブルの内容を確認するには、スパインノードに接続して、CLI コマンド「show dcimgr repo {eteps | sclass-maps | vnid-maps}」を実行します。

Cisco ACI マルチサイトと優先グループのサポート

Cisco ACI リリース 4.0(2) では、Cisco ACI マルチサイトで EPG 優先グループが利用できます。優先グループの構造は VRF レベルで有効であり、その VRF で定義されたすべての（または一部の）EPG をグループ化できます。図 10 に示すように、優先グループに属する EPG は、コントラクトを作成しなくても相互に通信できます。優先グループから除外されている EPG が他の除外された EPG や優先グループに含まれる EPG と通信するためには、コントラクトの定義が必要です。

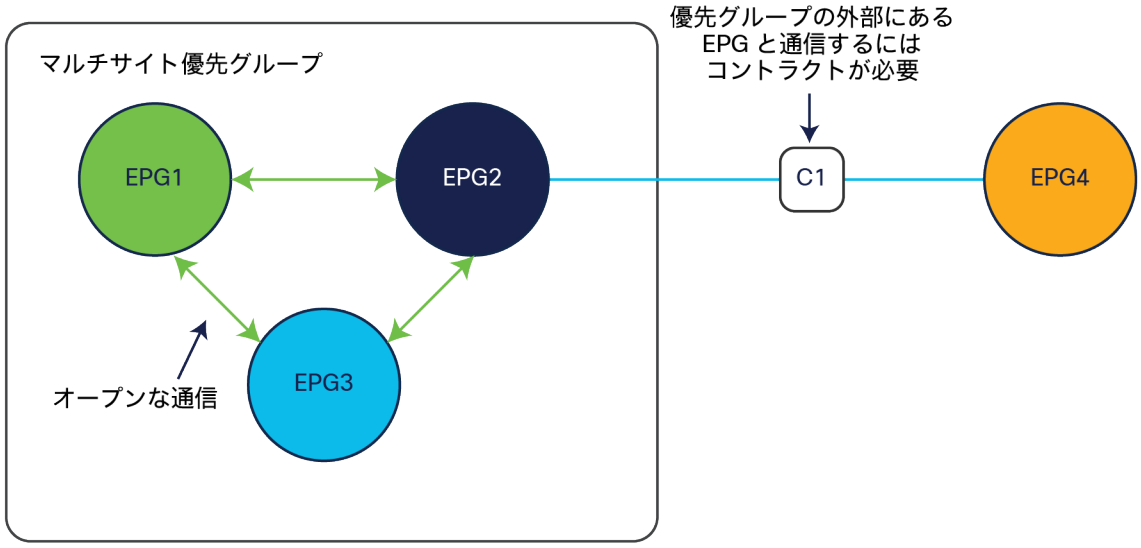


図 10.
EPG と優先グループ

注： 優先グループの外部の EPG と通信するには、優先グループに属する個々の EPG すべてにコントラクトを適用する必要があります。

この Cisco ACI マルチサイトシナリオの場合、直接 NDO で EPG を優先グループに属させる必要があります。そうすることで、スパインに正しい変換エントリが自動的に作成され、これらの EPG に属するエンドポイント間でのサイト間通信、さらには外部ネットワークドメインとの垂直方向通信が可能になります（図 11）。

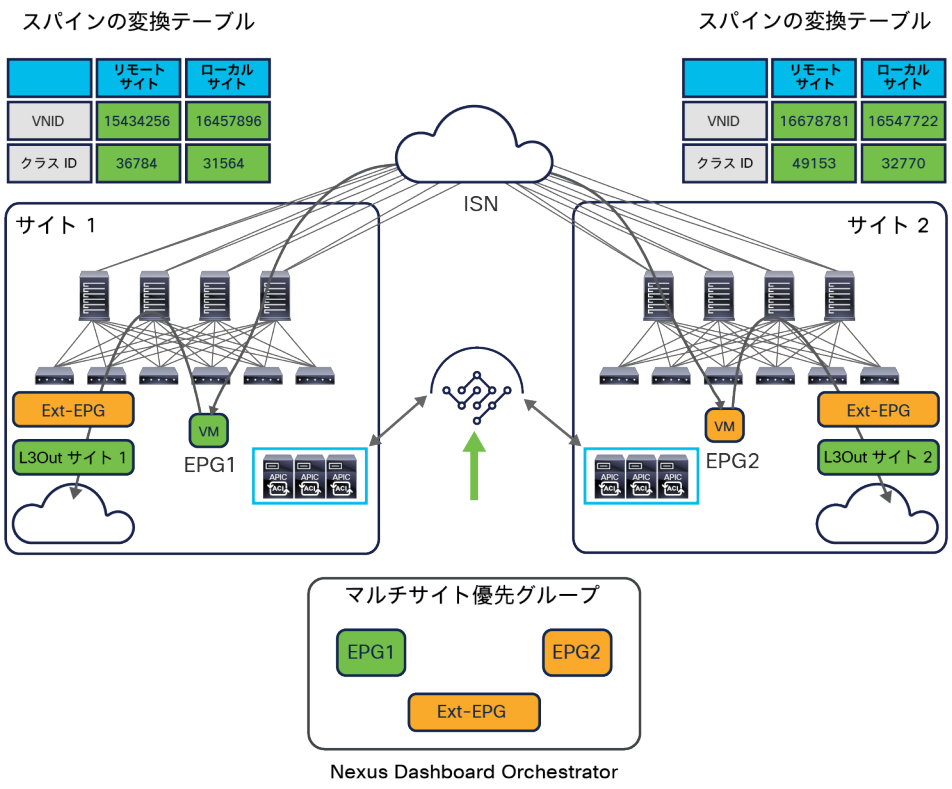


図 11.
優先グループを使用した**垂直方向**と**水平方向**の通信

そのため、拡張性の全体的な上限値を考慮したうえで、優先グループに入れる EPG の数を決めることが重要です。前述のように、この情報は、Web サイト Cisco.com にある Verified Scalability Guide で入手できます。

もう 1 つの重要な考慮事項は、L3Out に関連付けられた外部 EPG（Ext-EPG）に対する優先グループの設定です。これを行う場合、分類のために Ext-EPG にプレフィックス 0.0.0.0/0 を設定することはできません。これは、トラフィックが L3Out で受信され、このプレフィックスに基づいて分類される場合、Ext-EPG 固有のクラス ID ではなく、VRF のクラス ID が着信パケットに割り当てられるためです。その結果、VRF クラス ID から特定の EPG クラス ID へのトラフィックを許可するセキュリティルールが作成されていないため、その VRF の優先グループに属する他の EPG との通信が許可されません。回避策として、分類のために 2 つの個別のプレフィックス（0.0.0.0/1 と 128.0.0.0/1）を作成して、アドレス空間全体をカバーするようにすることもできます。

最後に、Orchestrator のリリース 3.4(1) と APIC の 5.2(1) では、優先グループを使用して、サイト間 L3Out 接続を有効にすることはできません。サイト間 L3Out の詳細については、「サイト間 L3Out 機能の導入（Cisco ACI リリース 4.2(1)/MSO リリース 2.2(1) 以降）」セクションを参照してください。

Cisco ACI マルチサイトと vzAny のサポート

Cisco Multi-Site Orchestrator リリース 2.2(4) では、マルチサイトで vzAny 機能が利用できます。vzAny は、1 つの VRF に属するすべての EPG（内部と外部）を表す論理構造です。これらすべての項目を 1 つのオブジェクトで表せるため、図 12 に示す 2 つの主要なケースで展開が簡素化されます。

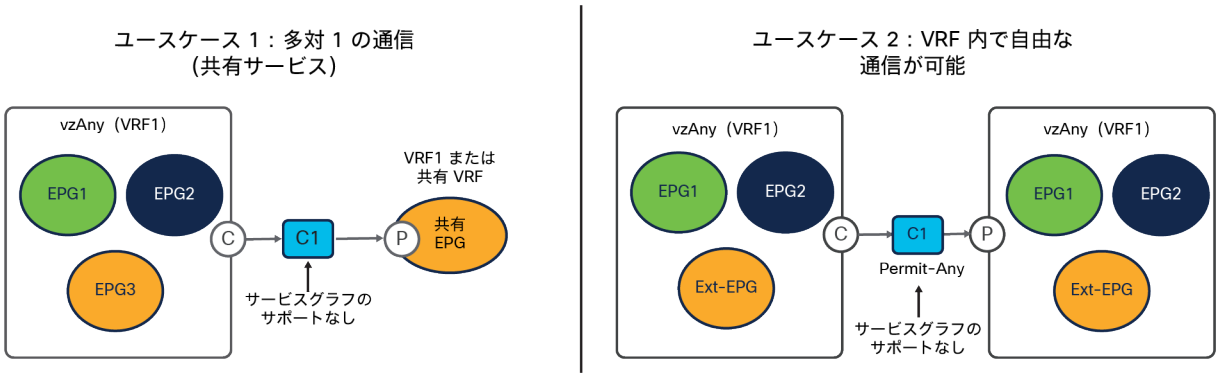


図 12.
vzAny が利用できる主なユースケース

- 最初のユースケースでは、1 つの VRF に属するすべての EPG と、同じ VRF または別の VRF に属する共有リソースの間で「多対 1」の通信を確立しています。個々の EPG と共有 EPG の間にコントラクトを適用する代わりに、共有 EPG によって提供されたコントラクトのコンシューマとして「vzAny」を設定することができます。これにより、各 EPG と共有 EPG の間の通信のみが確保され、同じ VRF に属する EPG 間の VRF 内通信は確保されないことに注意してください。

  リリース 2.2(4) 以降の Orchestrator で vzAny 構造を公開すると、エンドポイントが同じ ACI ファブリックに属しているか、異なるサイトに展開されているかに関係なく、この「多対 1」の通信パラダイムを展開できます。外部ネットワークドメインの共有リソースへのアクセスも可能になります（図 13）。

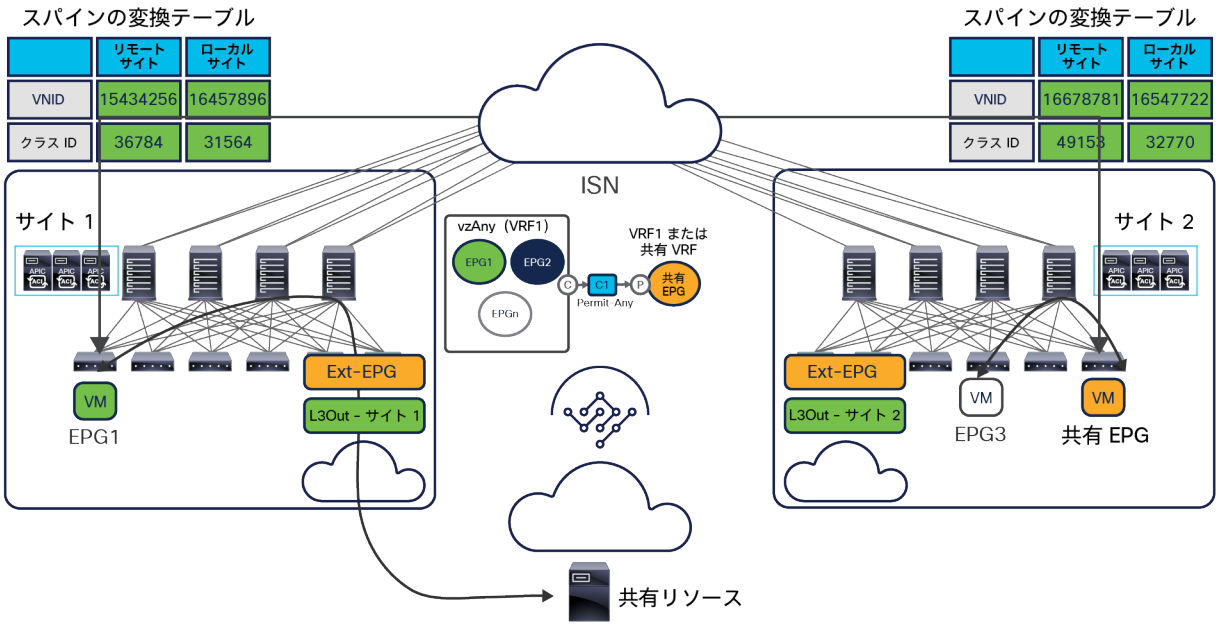


図 13.
サイト間および外部ネットワークとの多対 1 の通信

- 一方、同じ VRF に属するすべての EPG が相互に通信できるようにすることが目的である場合（優先グループの代替となる構成オプション）、「permit any」フィルタルールで定義された単一のコントラクトのコンシューマとプロバイダーとして vzAny を設定できます。これにより、機能的には「VRF unenforced」オプ

ション（NDO ではサポートされません）と同じ目的が達成できます。セキュリティポリシーを式に組み入れる必要がなくなり、ネットワーク接続のためだけに ACI マルチサイト展開を使用できます。図 14 に示すように、この構成オプションを使用すると、水平方向と垂直方向の両方の通信を確立できます。さらに、必要な変換エントリがスパインに適切にプログラミングされます。

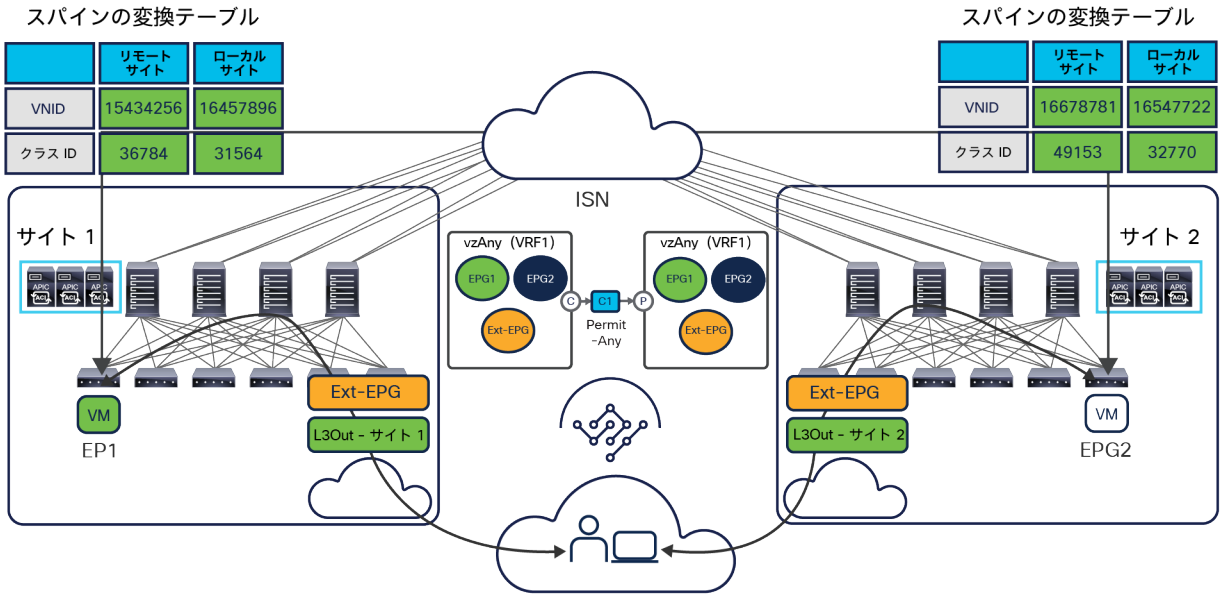


図 14.
VRF 内での any-to-any 通信の確立

このシナリオで vzAny を使用すると、EPG 間のフルメッシュコントラクトを作成する必要がなくなり、構成が簡素化されるだけでなく、TCAM の使用率も大幅に削減されます。

マルチサイトアーキテクチャで vzAny を利用するための要件について指摘しておくことが重要です。Cisco Multi-Site Orchestrator リリース 2.2(4)（または最新の NDO バージョン）を実行することが唯一の条件であり、同じマルチサイトドメインに属する他のファブリックに導入された ACI ソフトウェアのバージョンには依存しません。NDO と APIC のソフトウェアリリース間の依存関係に関するその他の考慮事項は、「バージョン間サポート（Cisco Multi-Site Orchestrator リリース 2.2(1) 以降）」セクションを参照してください。

また、Cisco Nexus Dashboard Orchestrator リリース 4.0(1) では、vzAny によって使用されるコントラクトにサービスグラフを追加することはできません（多対 1 と any-to-any の両方のユースケース）。この機能は、将来のソフトウェアリリースでサポートされる予定です。

Cisco Nexus Dashboard Orchestrator

Cisco Nexus Dashboard Orchestrator（NDO）は、プロビジョニングやヘルスモニタリングを行うだけでなく、世界に展開された Cisco ACI サイト全体で Cisco ACI のネットワーク、ファブリック、テナントポリシーをライフサイクル全体を通じて管理する役割を担った製品です。NDO は基本的に信頼できる唯一の情報源であり、別々の ACI ファブリックに展開されたエンドポイント間でサイト間（つまり、水平方向）通信を確立するために必要なすべてのテナントポリシーを提供します。

最新の実装では、Cisco Nexus Dashboard Orchestrator は、Nexus Dashboard と呼ばれるシスコのコンピューティングクラスタ上でサービスとして実行できるようになっています。以前の実装では、この製品は「Cisco Multi-Site Orchestrator（MSO）」と呼ばれていました。そのため、このドキュメントの以降の記述では、特に旧版の

Orchestrator で導入されていた機能について説明する際に「MSO」の名称を使用する場合があります。Multi-Site Orchestrator の以前の実装に関する詳細は、付録 B を参照してください。

Cisco Nexus Dashboard Orchestrator には、いくつかの主要な機能があります。注目すべき点としては、ユーザープロファイルと RBAC ルールの作成と管理がサイトオンボーディングのプロセスとともに Orchestrator から取り除かれ、Nexus Dashboard コンピューティング プラットフォームに追加されたことが挙げられます。これらのサービスが、ND コンピューティングクラスタで実行できるさまざまなアプリケーションすべてに共通したものであるためです。

NDO で現在利用できる機能は以下のとおりです。

- ヘルスダッシュボードを使用して、サイト間ポリシーの正常性、障害、ログをモニタリングできます。Cisco Multi-Site ドメインに属するすべての Cisco ACI ファブリックが対象です。正常性スコアに関する情報は各 APIC ドメインから取得され、図 15 に示すように一元化された方法で表示されます。

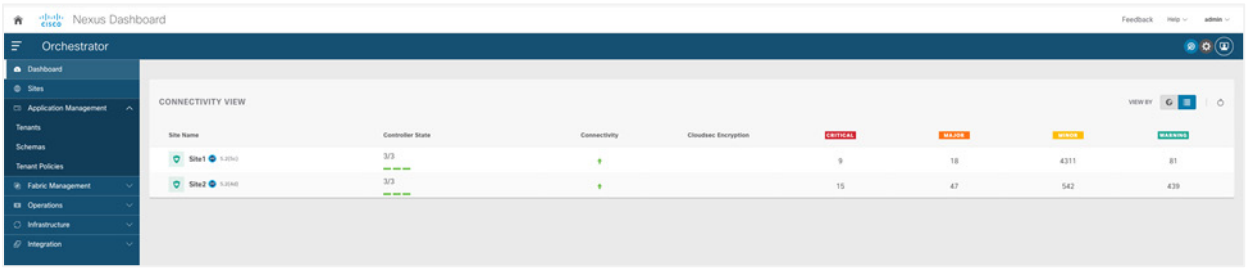


図 15.
Cisco Nexus Dashboard Orchestrator のダッシュボード

- Day-0 インフラストラクチャのプロビジョニングによって、すべての Cisco ACI サイトのスパインスイッチを直接接続されたサイト間ネットワーク（ISN）デバイスとピアリングします。マルチサイトドメインに属する各ファブリックで ISN とのピアリングが完了すると、すべてのスパインに対して NDO が MP-BGP EVPN を自動的に構成します。これによって、スパインが相互接続できるようになります。その結果、MP-BGP EVPN コントロールプレーンにおける到達可能性が確立され、サイト間でエンドポイントのホスト情報（MAC および IPv4/IPv6 アドレス）を交換できるようになります。
- 新しいテナントを作成し、接続されているすべてのサイト（またはその一部）に展開します。
- アプリケーション テンプレートを定義します。図 16 に示すように、各アプリケーション テンプレートをファブリックの特定のセットに関連付けて、そのセットにプッシュすることができます。

  注：1 つ以上のテンプレートをスキーマの要素としてグループ化できます。スキーマは、ポリシーの「コンテナ」と見なすことができます。ただし、テナントへのポリシーの関連付けは、常にテンプレートレベルで行われます。スキーマレベルではありません。

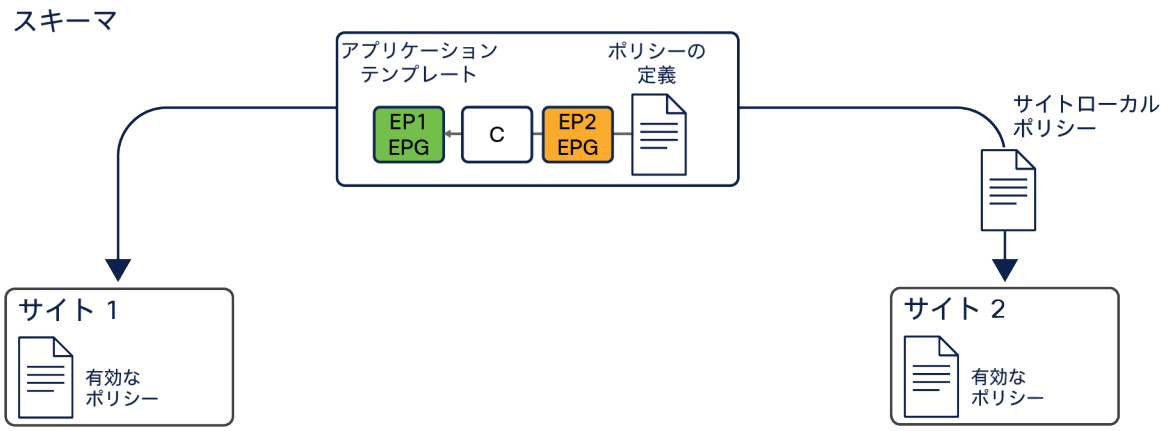


図 16.
スキーマ、テンプレート、サイト

この機能は、変更管理のために範囲を設定したポリシーを定義してプロビジョニングする機能とともに、Cisco Nexus Dashboard Orchestrator が提供する最も重要な機能の 1 つです。サイト間ポリシーを定義すると、Cisco Nexus Dashboard Orchestrator が、サイト全体にわたって、マルチサイト対応スパインスイッチで必要な名前空間変換ルールも適切にプログラミングします。以前のセクションで述べたように、サイト間通信を行うには、マルチサイトドメインに属する各ファブリックにあるスパインノードに変換エントリを作成する必要があります。これは、サイト間通信を許可するポリシーが Nexus Dashboard Orchestrator で定義され、ファブリックを管理する別の APIC クラスタにプッシュされた場合にのみ実行されます。そのため、すべてのテナントオブジェクト（EPG、BD など）の構成を NDO で直接管理することがベストプラクティスの推奨事項です。これらのオブジェクトが複数のサイトにまたがって拡張されているか、特定のサイトでローカルに定義されているかは関係ありません。NDO スキーマとアプリケーション テンプレートの展開方法の詳細については、このホワイトペーパーの「NDO のスキーマとテンプレートの展開」セクションを参照してください。

- ソフトウェアリリース 4.0(1) から、上記のアプリケーション テンプレートに加えて、他のタイプのテンプレートが NDO に追加されました。これらを用いて、テンプレートポリシー（テナント ポリシー テンプレート）、ファブリックポリシー（ファブリック ポリシー テンプレートとファブリック リソース ポリシー テンプレート）、SPAN モニタリングポリシー（モニタリング ポリシー テンプレート）をプロビジョニングできます。これらの新しいタイプのテンプレートの詳細については、このホワイトペーパーの「自律型アプリケーション テンプレート（NDO リリース 4.0(1)）」セクションを参照してください。
- すでに展開されている Cisco ACI ファブリック（ブラウンフィールド展開）からポリシーをインポートし、それらを別の新しく展開されたサイト（グリーンフィールド展開）に拡張できます。詳細については、「ブラウンフィールド統合シナリオ」セクションを参照してください。

Cisco Nexus Dashboard Orchestrator はマイクロサービス アーキテクチャに基づいて設計されており、NDO サービスはアクティブ/アクティブ方式で連携する Nexus Dashboard のクラスタノードにまたがって導入されます。Cisco Nexus Dashboard Orchestrator サービスは、異なるサイトに展開された各 APIC ノードと通信する必要があります。NDO および APIC クラスタ間の通信は、アウトオブバンド（OOB）インターフェイス、インバンド（IB）インターフェイス、またはその両方に対して確立できます。導入に関するより具体的な情報は、「Cisco Nexus Dashboard の導入に関する考慮事項」セクションを参照してください。Orchestrator は、Representational State Transfer（REST）

API または GUI（HTTPS）を介したノースバウンドアクセスも提供します。これによって、サイト全体に展開する必要があるネットワークとテナントのポリシーをライフサイクル全体にわたって管理できます（図 17）。

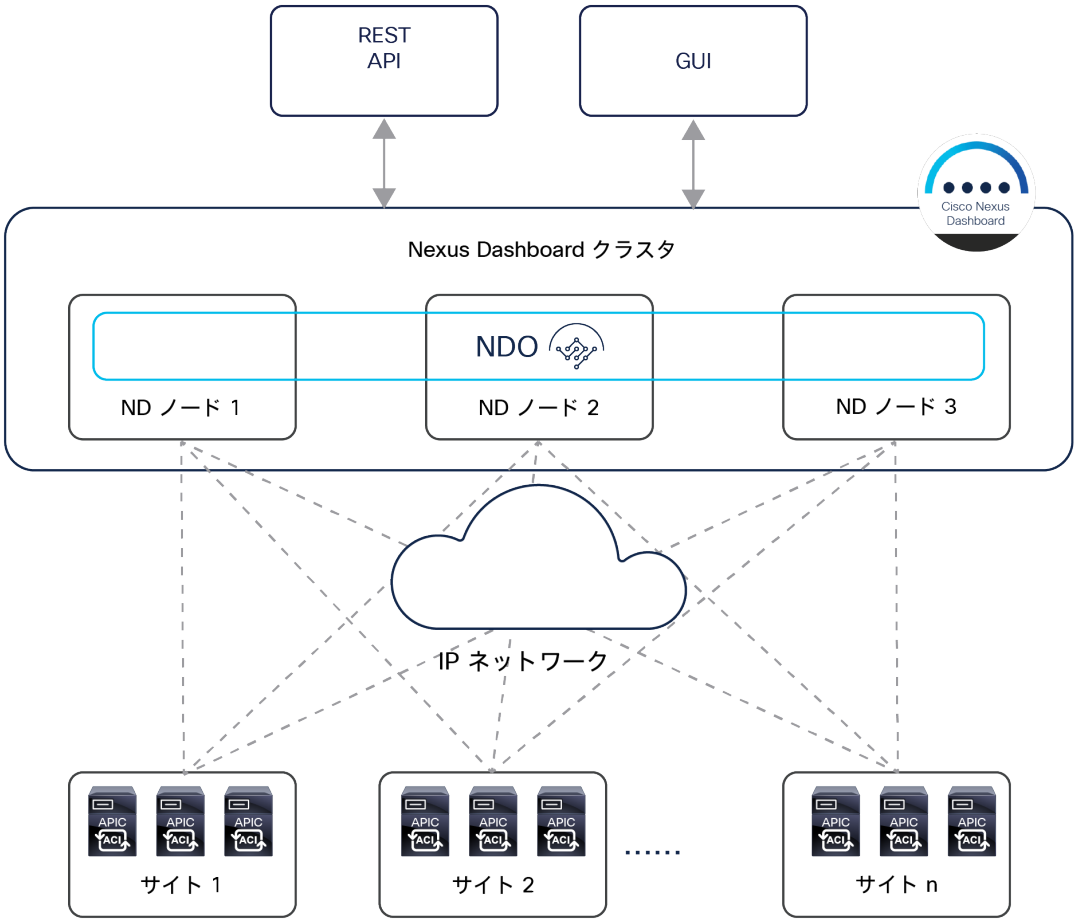


図 17.
Nexus Dashboard クラスタで実行されている Cisco Nexus Dashboard Orchestrator

Cisco Nexus Dashboard Orchestrator クラスタには、セキュリティを強化する仕組みが組み込まれています。Cisco Nexus Dashboard Orchestrator のクラスタ設計は、Nessus、WhiteHat、Chaos Corona、Norad など、先進の業界ベンチマークに基づくすべての脆弱性テストに合格し、セキュリティの脆弱性は検出されていません。

さらに、異なる物理（または仮想）ノードで実行されている Orchestrator サービス間のすべてのトラフィックは常に保護されています。Nexus Dashboard クラスタの異なるノードで実行されている NDO サービスの場合、それぞれのサービスがトラフィックの暗号化を担っています。たとえば、Mongo DB の情報を配布するために TLS が使用され、APIC への接続は HTTPS を介して行われます。Kafka も TLS を使用します。したがって、Orchestrator サービスは、サービスの相互接続に利用されるネットワーク インフラストラクチャが何であっても、セキュアに導入することができます。

Cisco Nexus Dashboard Orchestrator の典型的なユースケース

「はじめに」で説明したように、Cisco Nexus Dashboard Orchestrator によって管理される Cisco ACI マルチサイトアーキテクチャの展開には、2 つの一般的なユースケースがあります。

- 集中型（ローカル）データセンター：拡張性の要件または障害ドメイン分離の要件のために、同じ DC のロケーションに別々のファブリックを作成する必要がある場合
- 都市、国、大陸にまたがって地理的に分散したデータセンター：各データセンターを「リージョン」として扱い、リージョンにまたがって拡張されたポリシーを展開するために、プロビジョニング、モニタリング、管理を一元化する必要がある場合

以下の 2 つのセクションでは、これらの導入モデルについてさらに詳しく説明します。以下のセクションで明らかになるように、Orchestrator をホストする ND クラスタをパブリッククラウド（つまり、AWS または Microsoft Azure の特定のリージョン）に導入し、マルチサイトドメインに属するすべての ACI ファブリックをクラウドから管理することもできます。このアプローチは、以下で説明する両方のユースケースに適用できます。

リーフノードの拡張性を高めることを目的としたローカルデータセンターへの Cisco ACI マルチサイト展開

集中型展開のユースケースは、金融部門や政府部門でよく見られ、大規模なサービスプロバイダーでも採用されています。これらのシナリオでは、ベアメタルサーバー、仮想マシン、コンテナを接続するために非常に多くのポート数が必要な建物やローカルキャンパスに Cisco ACI マルチサイト設計が展開されます。多数のリーフノードを別々の Cisco ACI ファブリックに展開することによって、展開をスケールアウトしながら障害ドメインの範囲を制限し、すべてを一元管理できます。

図 18 に示す例では、4 つの Cisco ACI ファブリックが 1 つのホールにある 4 つの部屋ごとに展開され、各 Cisco ACI ファブリックは最大 500 のリーフスイッチで構成されます（マルチポッドファブリックを展開する場合）。Cisco Nexus Dashboard Orchestrator サービスは、Nexus Dashboard クラスタ（この例では 3 つの仮想マシン。3 つの物理ノードも可能）に導入されます。ND の各仮想ノードをそれぞれ別のハイパーバイザ（ESXi または KVM）に導入すれば、シングルポイント障害を避けることができます。Cisco Nexus Dashboard Orchestrator インターフェイスを介して、すべてのテナントポリシーを 4 つの Cisco ACI サイトすべてに拡張できます。追加の ND スタンバイプライマリノードを導入すれば、障害が発生した ND アクティブプライマリノードを置き換えることができます。

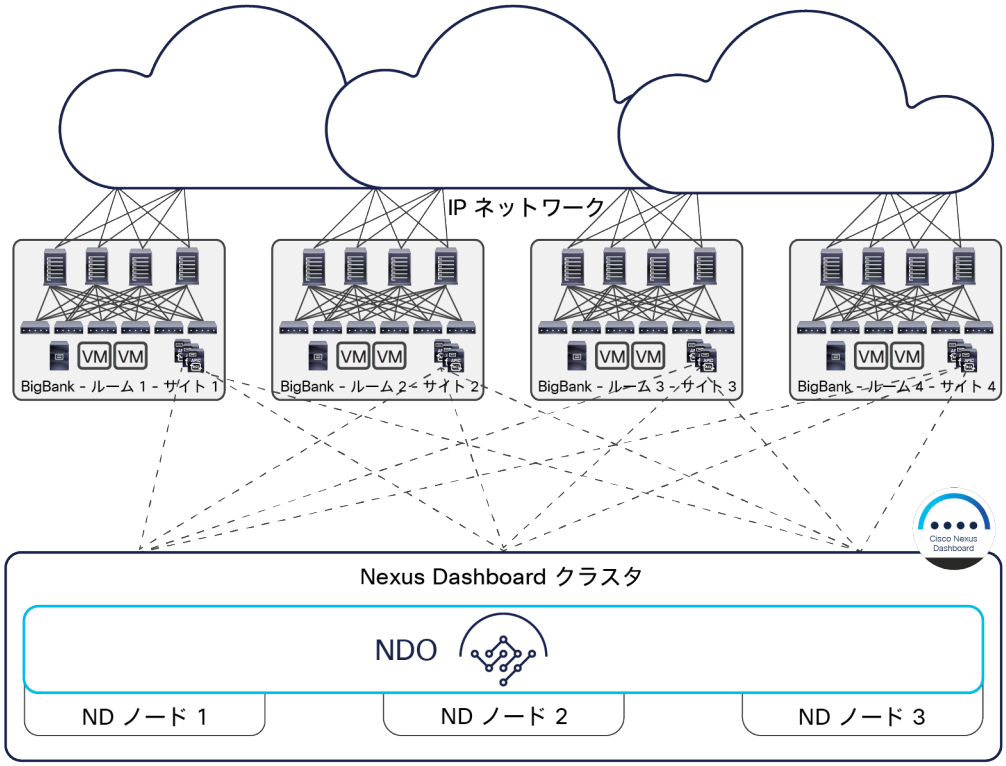


図 18.
データセンター内に導入された Cisco Nexus Dashboard クラスタ

Orchestrator サービスをホストするベアメタルサーバーまたは仮想サーバーを ACI ファブリックに直接接続する必要はありません（通常は推奨されません）。ACI ファブリック接続に問題が発生して NDO が APIC クラスタと通信できなくなる事態を避けるためです。後のセクションで明らかになるように、APIC クラスタの OOB、IB、または両方のインターフェイスとの通信が可能なため（NDO 導入モデルによる）、これらのクラスタをどこに接続するかについては、完全な柔軟性があります。たとえば、データセンターの外部に導入することも可能です。

WAN 経由で相互接続されたデータセンターでの Cisco Nexus Dashboard Orchestrator の導入

WAN のユースケースは、企業やサービスプロバイダーで広く採用されています。このシナリオでは、地理的に離れたデータセンターが、異なる国または大陸の都市間で相互接続されます。

図 19 に示す例では、3 つの Cisco ACI ファブリックがそれぞれローマ、ミラノ、ニューヨークに展開されていて、3 つすべてが、ローマとミラノにまたがって拡張された仮想（または物理）ND クラスタで実行されている Cisco Nexus Dashboard Orchestrator サービスから管理されています。注目すべき興味深い点は、イタリアに導入された Nexus Dashboard クラスタで実行されている NDO がニューヨークのサイトをリモートで管理できることです。これは、ND ノードとそれが管理する APIC コントローラクラスタとの間で最大 500 ミリ秒 RTT の遅延が許容されているためです。

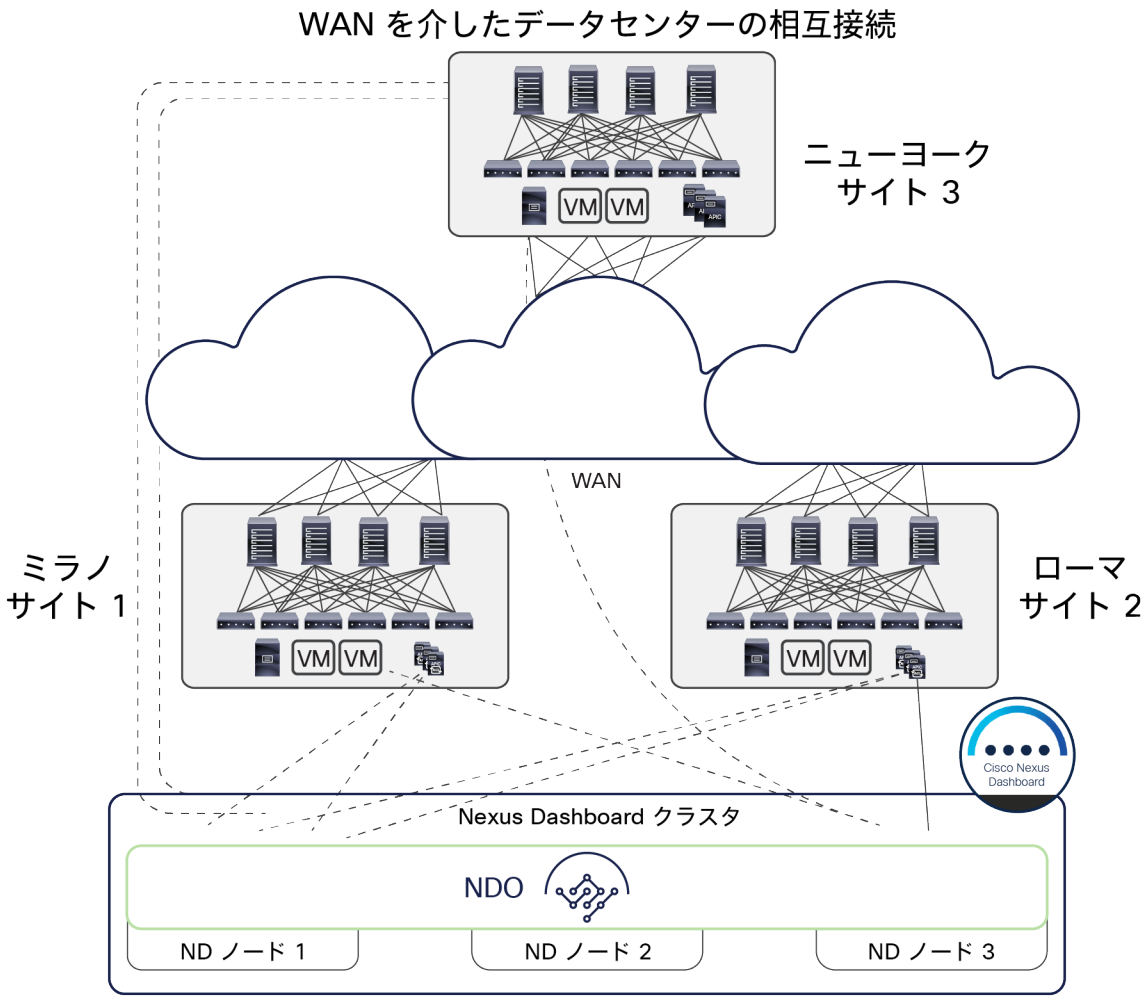


図 19.
WAN 経由で相互接続されたデータセンターにまたがって導入された Cisco Nexus Dashboard クラスタ

世界に広がる Cisco ACI ファブリックを管理する場合でも、Orchestrator をホストする Nexus Dashboard ノードを常に同じ地理的リージョン（米国、ヨーロッパ、アジアなど）に導入することをベストプラクティスとしてお勧めします。これは、同じ ND クラスタに属する ND ノード間の通信で許容される遅延は 150 ミリ秒 RTT であるためです。

Cisco Nexus Dashboard の同じクラスタに属するノードが WAN にまたがって導入されている場合、これらの間に安定したデータプレーン接続が存在する必要があります。Cisco Nexus Dashboard クラスタのノードは TCP 接続を介して相互に通信するため、WAN でドロップが発生すると、ドロップされたパケットが再送信されます。ND クラスタの各ノードには一意の IP アドレスが割り当てられます。ノード間の通信をルーティングできるため、これらの IP アドレスが同じ IP サブネットに属する必要はありません。

Cisco Nexus Dashboard Orchestrator クラスタのノード間の推奨される接続帯域幅は、300 Mbps から 1 Gbps です。この数値は、非常に大規模な構成の追加と削除を頻繁に行った内部ストレステストの結果に基づいています。

Cisco Nexus Dashboard の導入に関する考慮事項

Orchestrator（NDO）は、Cisco Nexus Dashboard（ND）と呼ばれるコンピューティングリソースのクラスタで実行されるアプリケーションとして導入されます。Nexus Dashboard の最初のリリース 2.0(1) は、3 つの物理 ND コンピューティングノードからなるクラスタのみをサポートしていました。Nexus Dashboard リリース 2.0(2) 以降は、オンプレミスまたはパブリッククラウドに導入する仮想フォームファクタとして使用できます（図 20）。

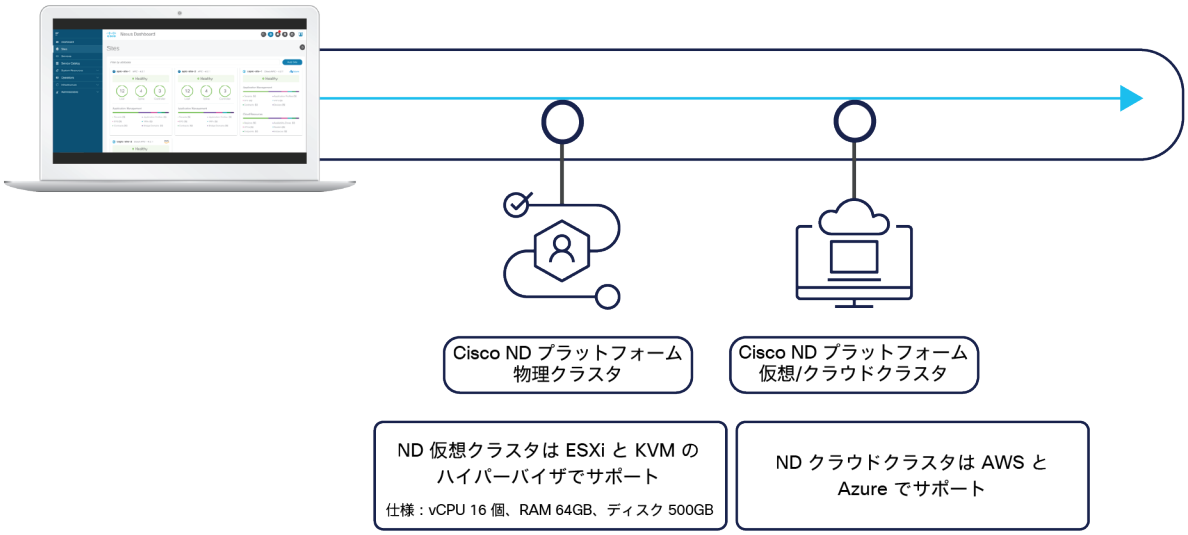


図 20.
NDO をサポートする Nexus Dashboard クラスタのフォームファクタ

注： 同じクラスタ内でこれらのフォームファクタを「混在させる」ことはできず、同種の導入モデルのみがサポートされます。言い換えると、すべてが物理ノード、すべてがオンプレミス上で同じハイパーバイザのフレーバに導入された仮想ノード、すべてが同じクラウドサービス プロバイダーのクラウドに導入された仮想ノードのいずれかである必要があります。

付録 B で説明されている以前の MSO 導入モデルと比較すると、Nexus Dashboard で NDO をサービスとして実行する場合、2 つの違いがあります。

- ND ノードと APIC クラスタ間の最大遅延は、500 ミリ秒 RTT に短縮されています（以前 MSO で許容されていた 1 秒 RTT ではありません）。
- ND ノードは、OOB アドレス、IB アドレス、またはその両方を使用して APIC コントローラと通信できます（以前の MSO オプションでは OOB 接続のみがサポートされていました）。

最後の点については、重要な考慮事項があります。Nexus Dashboard の各コンピューティングノードには、フォームファクタにかかわらず、管理インターフェイスとデータインターフェイスの 2 種類のインターフェイスがあります（図 21）。

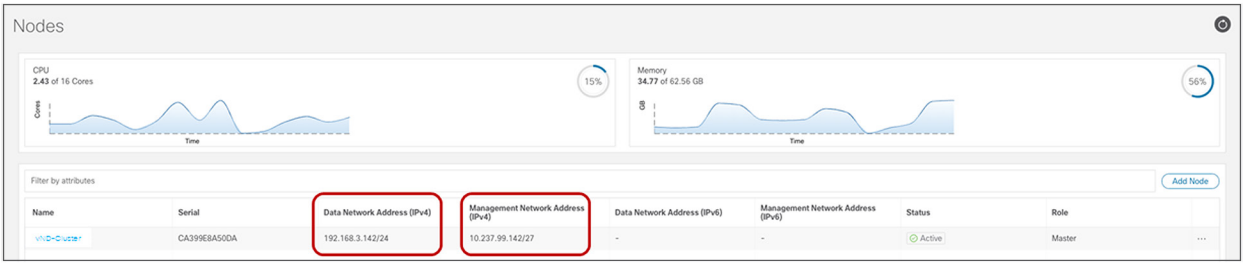


図 21.
Nexus Dashboard のコンピューティングノードが持つインターフェイス

注： 上の図は、ラボで使用される単一ノードの仮想 ND クラスタのものです。実際に導入される ND では、同じクラスタに 3 つのノードが属し、それぞれに個別のインターフェイスがあります。

これらのインターフェイスのそれぞれに異なるルーティングテーブルが関連付けられ、ネクストホップデバイスを指すデフォルトルートがそれぞれのテーブルに追加されます。以下のコマンドを使用して、2 つのインターフェイスに関連付けられたルーティングテーブルを表示できます（SSH で Nexus Dashboard ノードに「rescue-user」として接続した後）。

管理インターフェイス（bond1br）

```
rescue-user@vND-Cluster:~$ ip route show
default via 10.237.99.129 dev bond1br
10.237.99.128/27 dev bond1br proto kernel scope link src 10.237.99.142
100.80.0.0/16 dev bond0br scope link
169.254.0.0/25 dev k8br1 proto kernel scope link src 169.254.0.44
169.254.0.0/16 dev bond0br scope link metric 1006
169.254.0.0/16 dev bond1br scope link metric 1009
172.17.0.0/16 dev k8br0 proto kernel scope link src 172.17.222.0
192.168.3.0/24 dev bond0br proto kernel scope link src 192.168.3.142
```

データインターフェイス（bond0br）

```
rescue-user@vND-Cluster:~$ ip route show table 100
default via 192.168.3.150 dev bond0br
10.237.99.128/27 dev bond1br scope link
172.17.0.0/16 dev k8br0 scope link
192.168.3.0/24 dev bond0br scope link
```

各 ND クラスタノードをネットワーク インフラストラクチャに接続する方法を決定するには、ND で実行されるさまざまな機能とサービスが、（デフォルトで）到達可能性情報を活用するように設計されていることを必ず理解する必要があります。この情報は、さまざまな外部のサービスと通信するために記述された 2 つのルーティングテーブルのいずれかに含まれています。ND がこのように実装されていることから、以下を考慮する必要があります。

たとえば、ND が NTP サーバーまたはプロキシサーバーに接続しようとする場合、その接続先に到達するためのルックアップは第 1 のルーティングテーブルで実行されます。一方、APIC の OOB または IB の IP アドレスを指定することによって、ND で APIC コントローラのオンボーディングを実行しようとする場合、その接続先に到達するためのルックアップは第 2 のルーティングテーブルで実行されます。

- ND 管理インターフェイスは、ND クラスタの管理のみに用いられます。これは、このインターフェイスに関連付けられたルーティングテーブルが ND によって NTP サーバーと DC プロキシサーバー、Cisco Intersight クラスタ、DNS サーバーに接続するために用いられるからです。ND は、この接続を用いて ND（および ND アプリ）への UI アクセスを提供し、ND とそこで実行されているアプリケーションに対してファームウェアのアップグレードを実施します。上記のすべてのサービスが ND 管理インターフェイスに割り当てられたものとは異なる IP サブネットに導入されている場合、そのルーティングテーブルで定義されたデフォルトルートを用いて、それらすべてのサービスと通信します（上記の例では、ネクストホップのデバイス 10.237.99.129 を指しています）。
- ND データインターフェイスは、ND クラスタ（ノード間通信）を稼働させるために使用されます。また、ND で実行されている特定のサービス（NDO、NDI、NDFC など）によって、コントローラおよびスイッチと通信するために使用されます。コントローラとスイッチが ND データインターフェイスに割り当てられたものとは異なる IP サブネットに属している場合、ND データインターフェイスのルーティングテーブルで定義されたデフォルトルートを用いて、それらのデバイスと通信します（上記の例では、ネクストホップのデバイス 192.168.3.150 を指しています）。
- 上記のデフォルトの動作は、ND 管理インターフェイスまたは ND データインターフェイスに、ND が通信する必要のある外部のサービスおよびデバイスと同じ IP サブネットを割り当てることによって変更できます。たとえば、ND 管理インターフェイスが APIC コントローラと同じ IP サブネットに導入されている場合、上の例に示す第 2 のルーティングテーブルにエントリ 10.237.99.128/27 が関連付けられているため、APIC との接続には必ず管理インターフェイスが使用されます。または、特定の ND インターフェイスに関連付けられたスタティックルートを追加することにより、そのインターフェイスの使用を強制することもできます。たとえば、APIC が IP サブネット 192.168.1.0/24 に属している場合、そのスタティックルートを ND 管理インターフェイスに関連付けると、エントリ 192.168.1.0/24 が第 2 のルーティングテーブルにインストールされ、これが ND 管理インターフェイスを指します。
- 上記の箇条書きで説明したように、接続に関して ND プラットフォームには大きな柔軟性がありますが、ND コンピューティングクラスタで NDO を実行する場合のベストプラクティスの推奨事項は以下のとおりです。
    - ND クラスタで NDO サービスのみを実行している場合は、ND の管理インターフェイスとデータインターフェイスの両方を同じ IP サブネットに割り当てることができます。しかし、同じクラスタで追加のサービス（NDI など）を有効化した際に問題が発生する可能性があるため、これら 2 つのインターフェイスは異なる IP サブネットに割り当てることを強くお勧めします。
    - また、上記の最初の 2 点で説明した特定のデフォルト通信に 2 つの ND インターフェイスを使い分けるデフォルトの動作を維持することを強くお勧めします。これは、ND の管理インターフェイスとデータインターフェイスを ND が接続する必要のある外部エンティティが使用する IP サブネットとは異なる IP サブ

ネットに割り当てることを意味します。このようにすることで、各ルーティングテーブル（および関連付けられたインターフェイス）のデフォルトルートが、必要な通信に応じて使い分けられます。

- 同じ ND クラスタに属するノードの ND 管理インターフェイスを同じ IP サブネットに割り当てることも、異なる IP サブネットに割り当てることもできます。前者は通常、ND クラスタが同じ DC のロケーションに導入されている場合で、後者は ND クラスタが異なる DC のロケーションにまたがって拡張されている場合です。ND クラスタノードの ND データインターフェイスについても同じことが言えます。

図 22 は、NDO が実行されている Nexus Dashboard クラスタの一般的な導入シナリオをいくつか示しています。

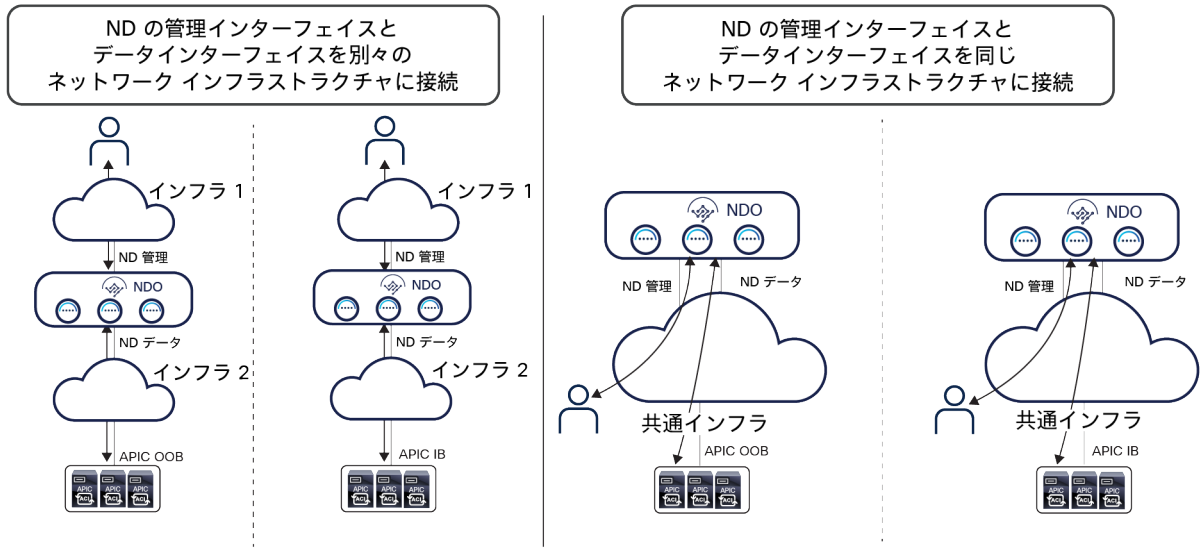


図 22.
Nexus Dashboard のコンピューティングノードが持つインターフェイス

左側の 2 つのシナリオでは、ND の管理インターフェイスとデータインターフェイスが、別々のネットワーク インフラストラクチャに接続されています。これにより、上記で説明した特定の接続に関する各インターフェイスの役割分担が明確になります。

逆に、右側の 2 つのユースケースでは、共通のネットワーク インフラストラクチャを使用して、ND の管理インターフェイスとデータインターフェイスの両方を接続しています。このようなシナリオでは、ネットワーク インフラストラクチャの中で異なる VRF を使用して、さまざまなタイプの必要な通信を分離し続けるのが非常に一般的と思われます。

注： Nexus Dashboard クラスタで NDO のみを実行する場合、図 22 に示すユースケースすべてで、APIC アウトオブバンド（OOB）インターフェイス、インバンド（IB）インターフェイス、またはその両方との接続を制約なく確立することができます。サイトの Nexus Dashboard でのオンボーディング（たとえば、その ACI ファブリックを管理する APIC のオンボーディング）は、いずれかの APIC ノードの IP アドレスから 1 つ（OOB または IB）を指定することで実行できます。ND が、指定されたアドレスに接続できれば（前述のベストプラクティスの推奨事項に従う場合、データインターフェイスを使用）、同じクラスタに属する他の APIC ノードすべての IP アドレスも自動的に検出されます。

Cisco Nexus Dashboard と、そのアプリケーションをホストする機能の詳細は、以下のリンクにあるドキュメントを参照してください。

https://www.cisco.com/c/ja_jp/support/data-center-analytics/nexus-dashboard/products-installation-guides-list.html

https://www.cisco.com/c/ja_jp/support/cloud-systems-management/multi-site-orchestrator/products-installation-guides-list.html

Nexus Dashboard の物理的なコンピューティングクラスタは複数のアプリケーションとサービスをホストできますが、現時点では、単一の NDO インスタンスに関連付けられたサービスを別々の Nexus Dashboard クラスタにインストールすることはできません。同じ Nexus Dashboard クラスタのノードにまたがってインストールすることだけが可能です。Nexus Dashboard Insights（NDI）と Nexus Dashboard Orchestrator の両方のサービスを活用しようと考えた場合、この制約は興味深い留意点です。NDI の導入にかかわる考慮事項はこのホワイトペーパーの範囲外ですが、NDI をホストする ND クラスタを地理的に離れた場所に分散させないようにすることが、常に基本的な原則となります（主に NDI のテレメトリデータの取り込みに関する要件のため）。そのため、たとえば、2 サイトに導入するシナリオの場合、NDI をホストする別の物理 ND クラスタをそれぞれの DC のロケーションに導入する必要があります（図 23）。

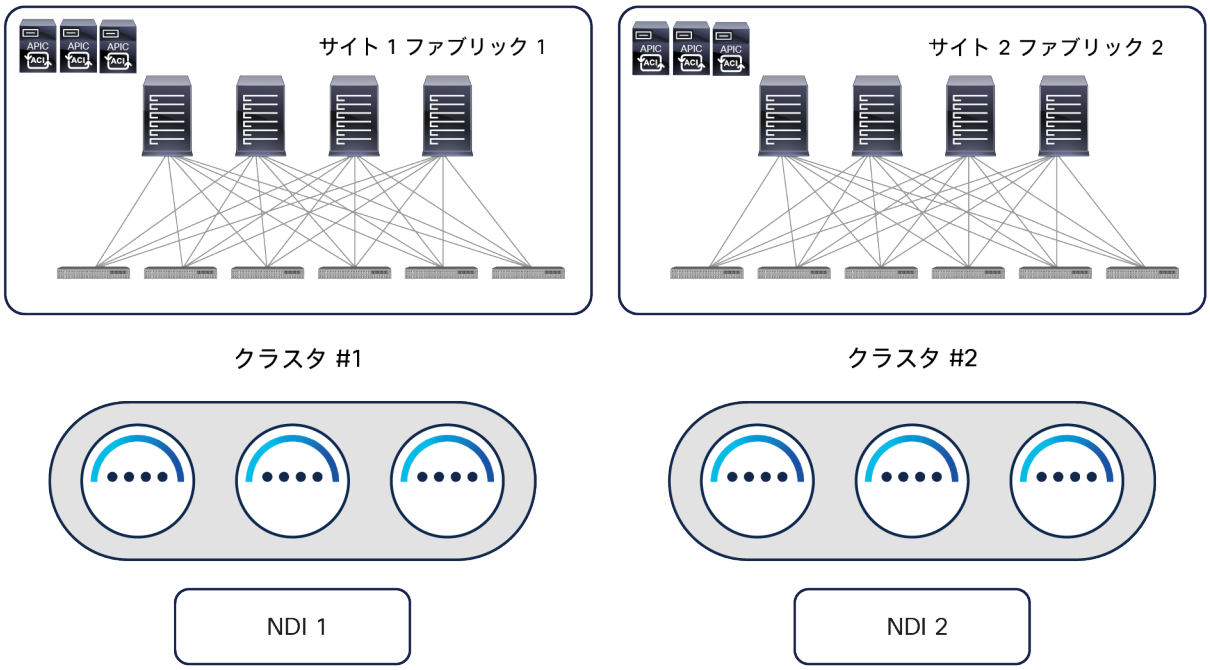


図 23.
地理的に分散したデータセンターにおける NDI の一般的な導入方法

このようなシナリオで NDO も導入する必要がある場合、現状では、単一の NDO インスタンスを異なる ND コンピューティングクラスタに導入できないため、NDO の他の導入オプションがいくつか考えられます。

1. DC の複数のロケーションに分散できる専用の仮想 ND（vND）クラスタに Orchestrator サービスを導入します。図 24 に示すように、この導入の推奨モデルでは、オンプレミスでホストされ異なるサイトに導入された 3 つの仮想マシン（ND プライマリノード）を使用して、NDO 専用の vND クラスタを構築できます。4 つ目の仮想 ND プライマリノードをスタンバイとして導入すれば、障害が発生したアクティブなプライマリノードを代替できます。

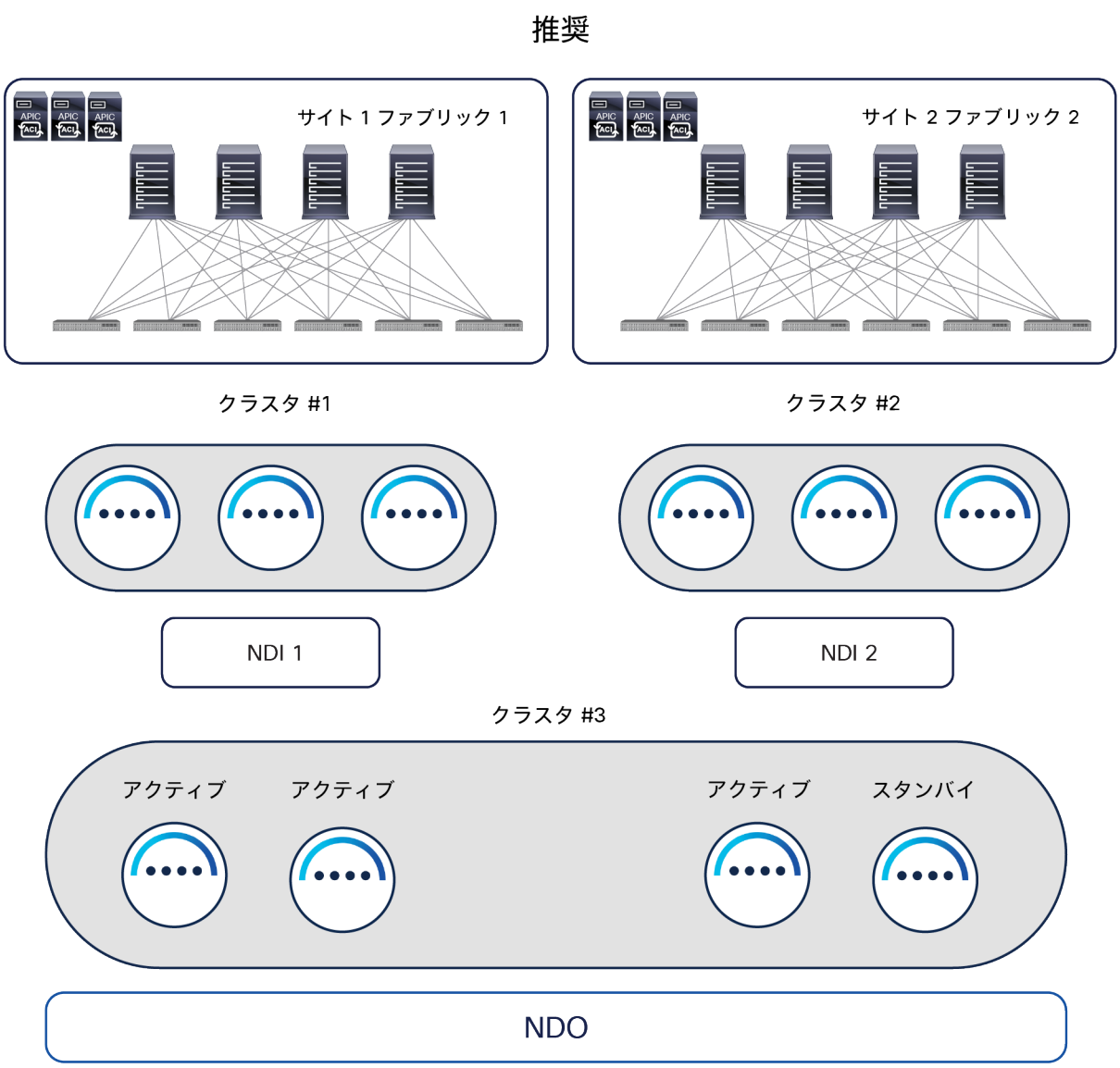


図 24.
地理的に分散したデータセンターに推奨される NDO と NDI の導入オプション

専用の仮想 ND クラスタで Orchestrator サービスを実行する主な優位性は以下のとおりです。

- 最も柔軟性の高い導入オプションです。Orchestrator サービスをホストする vND ノードは、遅延が最大 150 ミリ秒 RTT までであれば地理的に分散できます。地理的に分散したファブリックが同じマルチサイトドメインに属しているシナリオでは、このオプションが理想的です。これらの分散したデータセンターのロケーションに vND ノードを導入できるためです（同時に複数の vND ノードが失われる可能性が低くなります）。図 24 に示すシナリオでは、DC1 の機能が完全に失われた場合、DC2 の vND スタンバイノードをアクティブに昇格させ、この vND クラスタをマジョリティの状態にします。これによって、機能しているデータセンターへのポリシーのプロビジョニングが可能になります。
- このシナリオでは、APIC クラスタとの通信方法を選択できます。これは、vND ノードと APIC の間で Orchestration サービスに必要な通信チャネルとして、IB、OOB、または両方が利用できるためです（他のサービスが ND クラスタでともにホストされている場合は、できない可能性があります）。

- サイトの数とサイトあたりのリーフノードの数（サポートされる最大値あり）に関係なく、わずか 3 つの vND プライマリノードで実行できます。他のサービスをホストする ND クラスタでは、サイトやリーフの拡張性要件によって、追加の「ワーカー」ノードの導入が必要になる場合があります。サービスのさまざまな組み合わせをサポートするために必要な ND リソースの詳細は、https://www.cisco.com/c/dam/en/us/td/docs/dcn/tools/nd-sizing/index.html にある Nexus Dashboard Capacity Planning ツールを参照してください。
- vND クラスタは AWS や Azure のパブリッククラウドで直接ホストすることが可能で、そのうえで Orchestrator サービスを実行できます。

2. 図 25 に示すように、各データセンターに導入され NDI インスタンスをホストしている 2 つの ND クラスタに、2 つの別々の NDO インスタンスをインストールします。

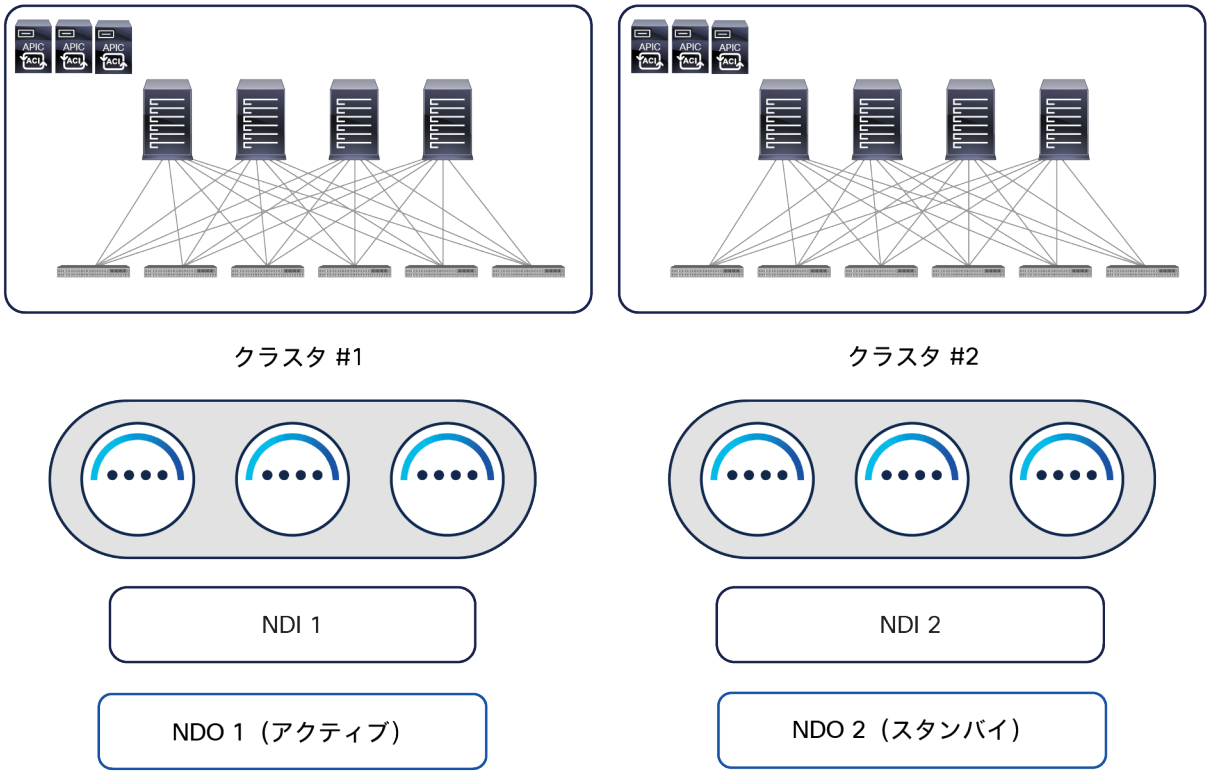


図 25.
地理的に分散したデータセンターに適用できる NDO と NDI の他の導入オプション

1 つ目の NDO インスタンスが「アクティブ」として実行され、マルチサイトドメインに属するすべてのファブリック（上図の例では、DC1 と DC2 の 2 つのファブリック）を管理します。2 つ目の NDO インスタンスはインストールされていますが、「アクティブ」としては使用されていません（一種の「スタンバイ」モード）。NDO 構成の定期的なバックアップを NDO のアクティブインスタンスで取得し、安全なリモートロケーションに保存できます。DC1 で大規模な災害が発生し、ローカルリソース（ND クラスタを含む）が失われた場合、DR 手順の 1 つのステップとして、DC2 で実行されている 2 つ目の NDO インスタンスに利用可能な最新のバックアップをインポートし、その構成にロールバックできます。その時点で、2 つ目の NDO インスタンスは実質的に「アクティブ」になり、これを用いて、マルチサイトドメインに属する残りすべてのファブリックの管理を開始できます。

注： アクティブな NDO インスタンスから頻繁にバックアップを取得すると、ディザスタリカバリのシナリオでの目標復旧時点（RPO）を最短にすることができます。

さらに、利用可能な NDO ソフトウェアの最新バージョンを常に導入することを強くお勧めします。異なる NDO ソフトウェアリリース間のアップグレードは非常に簡単なプロセスであり、Nexus Dashboard UI から直接処理できます。VM ベースまたは CASE ベースで導入された MSO クラスタと NDO の間の移行手順に関しては、異なる考慮事項があります。このケースでは、図 26 に示す手順を実行できます。

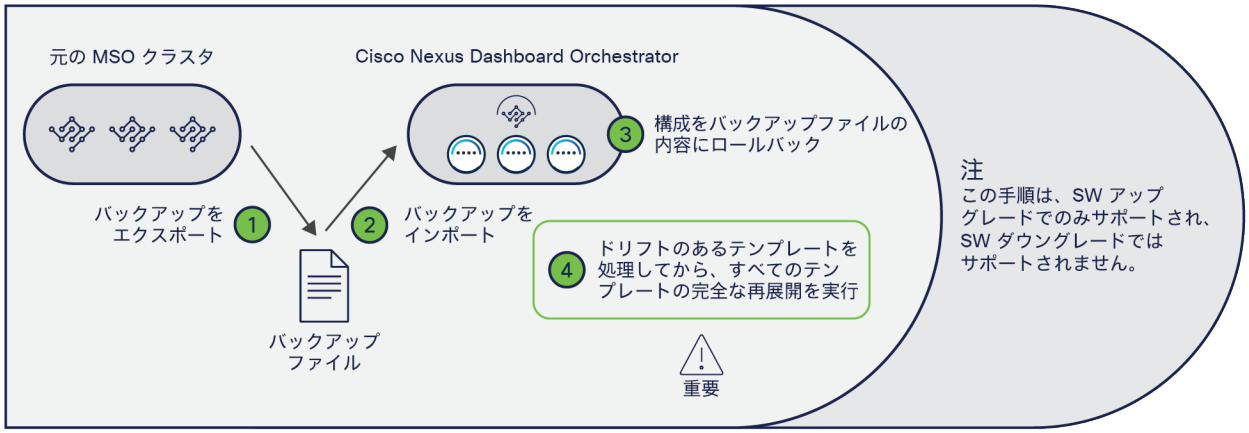


図 26.
MSO と NDO の間でアップグレードや移行を行う手順

注： 以下で説明する手順は、3.x NDO リリースへの移行に適用されます。NDO 4.0(1) への移行手順には、異なる考慮事項があります。詳細は、https://www.cisco.com/c/ja_jp/td/docs/dcn/ndo/4x/deployment/cisco-nexus-dashboard-orchestrator-deployment-guide-401/ndo-deploy-migrate-40x.html にあるドキュメントを参照してください。

- NDO アプリケーションを新しい ND クラスタにインストールします。この時点で、または以下の手順の後に、このクラスタをネットワークに接続できます。このとき、古い MSO クラスタを切断します（新旧クラスタで同じ IP アドレスを使用したい場合は、これが必要な可能性があります）。
- 古い VM ベースの MSO クラスタで構成ファイルのバックアップを作成し、そのファイルを簡単に見つけられる場所にダウンロードします。
- 古い VM ベースの MSO クラスタをシャットダウン（または単に切断）します。
- このファイルを ND クラスタで実行されている新しい NDO アプリケーションにインポートします（最初のステップでネットワークに接続していない場合は、接続してから実行します）。
- ND クラスタで実行されている新しい NDO アプリケーションの構成を、先ほどインポートした構成ファイルに含まれている構成にロールバックします。これにより、NDO インフラストラクチャの構成とテナント固有のポリシーの両方が新しいクラスタにインポートされます。

注： ACI サイトのオンボーディングは Nexus Dashboard で直接管理されるため、NDO の構成を正常にロールバックできるようにするためには、ND にオンボーディングされた ACI ファブリックに割り当てられた名前を、元々 MSO クラスタにオンボーディングされていた ACI ファブリックの名前と確実に一致させる必要があります。

- 構成のロールバックが完了すると、一部のテンプレートにドリフトが発生する場合があります。ドリフトとは、1 つ（または複数）のオブジェクトの APIC 構成と NDO 構成の間に不一致があることを意味します。管理できる ACI オブジェクト（またはオブジェクトのプロパティ）が異なる、2 つの Orchestrator リリース間で構成をロールバックした後、ドリフトが発生することがあります。これは、たとえば、MSO 2.2 から NDO 3.7 に移行する場合に当てはまります。NDO 3.7 は MSO 2.2 よりも多くのオブジェクトを管理できるためです。そのため、ロールバックが完了すると、これまで MSO 2.2 によって管理されていなかったオブジェクトすべてに、NDO 3.7 がデフォルト値を割り当てます。マルチサイトドメインに属するファブリックを管理する APIC でそれらのオブジェクトが持つ実際の値と、割り当てられたデフォルト値が異なることがあります。MSO 2.2 がオブジェクトを管理できなかったために、そのオブジェクトの値を APIC 上で直接変更した場合などに、このような差異が生じます。このようなドリフトは、NDO ソフトウェアリリース 3.4(1) 以降に導入されたドリフト調整ワークフローを利用して解決できます。詳細は、「NDO の運用面での機能強化」セクションを参照してください。
- 3.8(1) より前の NDO リリースに移行する場合は、すべてのドリフトを解決した後に、もう 1 つ手順を実行する必要があります。この手順では、定義されたすべてのアプリケーション テンプレートを再展開します。これは、すべてのテンプレートの構成情報が NDO データベースに正しく保存されるようにするために必要な手順です（このデータベースの、NDO で実装されたフォーマットが、MSO で使われたフォーマットから更新されているため）。NDO リリース 3.8(1) 以降、この「再展開」は移行手順の一部として自動的に処理されます。したがって、ロールバック後に対処する必要があるのは、構成のドリフトがあった場合にこれを解決することだけです。

注： この MSO から NDO への移行手順に関する詳細は、https://www.cisco.com/c/ja_jp/td/docs/dcn/ndo/3x/deployment/cisco-nexus-dashboard-orchestrator-deployment-guide-371/ndo-deploy-migrate-37x.html にあるドキュメントを参照してください。

NDO のスキーマとテンプレートの展開

マルチサイトのアプリケーション テンプレート

テナント固有のポリシー（EPG、BD、VRF、コントラクトなど）は、NDO において 1 つのアプリケーション テンプレートの中に作成されます。これは、それぞれのテンプレートが常に 1 つの（唯一の）テナントに関連付けられているためです。複数のアプリケーション テンプレートをスキーマに入れてグループ化することができます。スキーマは、基本的にアプリケーション テンプレートのコンテナになります。

スキーマは、特定のテナントに直接関連付けられることはありません。それでも、特定のテナントに関連付けられたすべてのアプリケーション テンプレートをスキーマに入れてグループ化することは、展開オプションとして非常に一般的であり、ベストプラクティスでもあります。このグループ化の目的は、NDO GUI からテナントポリシーの可視化と変更を容易に行えるようにすることです。

次に、定義された各アプリケーション テンプレートを、同じマルチサイトドメインに属する 1 つ（または複数）のサイトにマッピングする必要があります。

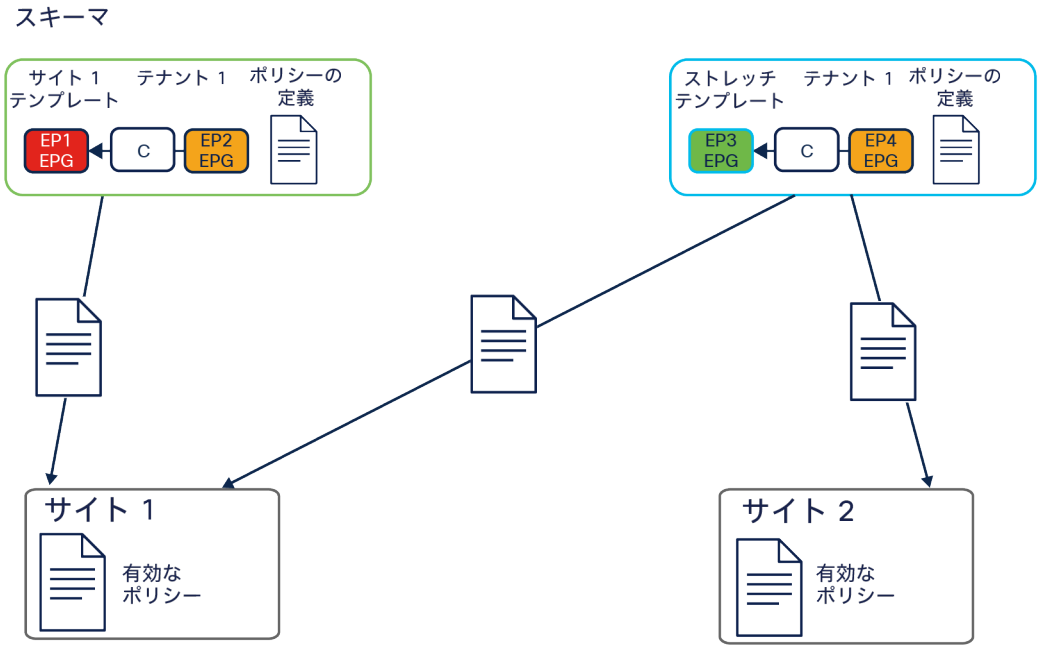


図 27.
アプリケーション テンプレートの ACI サイトへのマッピング

図 27 では、アプリケーション テンプレートを 2 つ作成し、同じ Tenant-1 に関連付けています。

- 「サイト 1 テンプレート」は ACI サイト 1 にマッピングされています。これによって、このアプリケーション テンプレートで作成されたすべてのテナントポリシーが、この ACI ファブリックを管理する APIC クラスタにのみプッシュされ、展開（「レンダリング」）できるようになります。
- 「ストレッチテンプレート」は ACI サイト 1 と 2 の両方にマッピングされています。これによって、テンプレートで定義されたすべてのテナントポリシーが両方のサイトに展開され、「ストレッチ」オブジェクト（複数のサイトでレンダリングされるオブジェクト）が作成されます。たとえば、BD をサイトにまたがって拡張できるようにするには、ストレッチ アプリケーション テンプレートを使用する必要があります（「ACI マルチサイトのユースケース」で説明します）。

現在の NDO の実装では、アプリケーション テンプレートがポリシー変更の最小単位になります。つまり、そのテンプレートに適用されたすべての変更が、テンプレートにマッピングされたすべてのサイトに常にただちにプッシュされます。特定のサイトにのみマッピングされたアプリケーション テンプレートを変更した場合、その変更はそのサイトにのみプッシュされます。

ポリシーオブジェクトは、アプリケーション テンプレートとスキーマで非常に柔軟に編成することができます。ネットワーク固有のオブジェクト（BD、VRF）とポリシー固有のオブジェクト（EPG、コントラクトなど）を別々のアプリケーション テンプレートまたはスキーマに入れると、便利な場合があります。ポリシー固有のオブジェクトはネットワーク固有のオブジェクトよりも頻繁に変更されると想定されるからです。

図 28 は、オブジェクトが、同じアプリケーション テンプレート、同じスキーマに属する他のプリケーション テンプレート、他のスキーマに属するアプリケーション テンプレートのいずれで定義されていても、容易に参照関係を作成できることを示しています。このような参照関係はいずれも、アプリケーション テンプレートが、同じテナントに関連付けられていても、別々のテナントに関連付けられていても作成できます（「共通」テナントにネットワーク固有のオブジェクトを定義する場合が典型例です）。

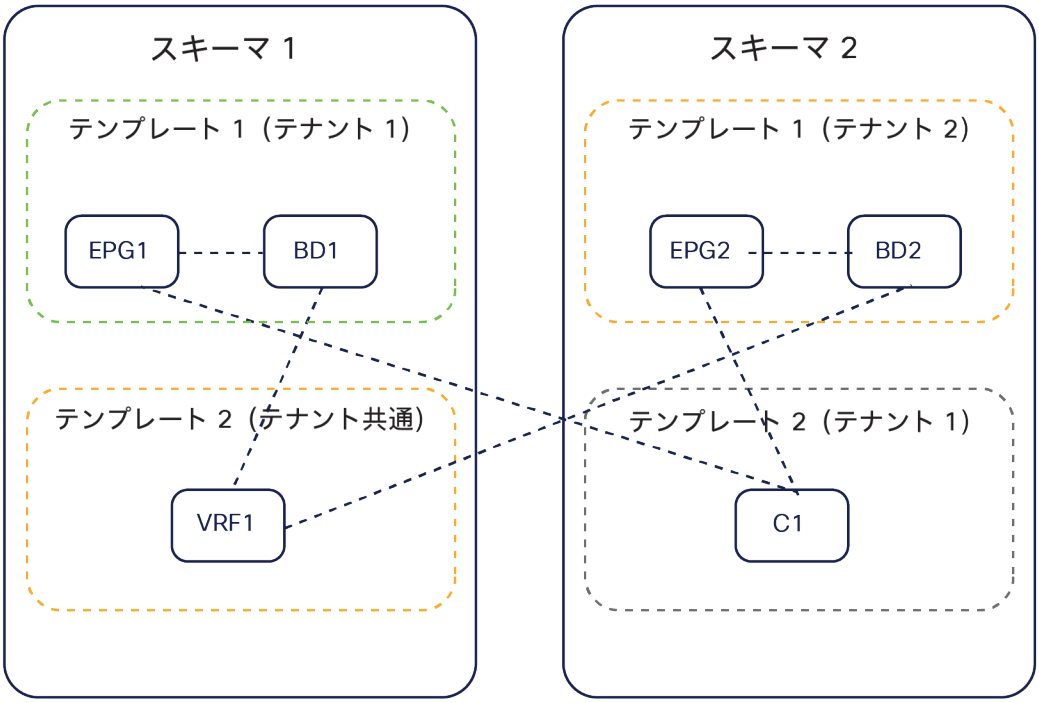


図 28.
アプリケーション テンプレート間とスキーマ間でのオブジェクトの参照

技術的には、同じアプリケーション テナントの構成オブジェクトを複数のテンプレートとスキーマに分散させることが可能です。しかし、図 29 に示すように、同じテナントに関連付けられたアプリケーション テンプレートはすべて特定のスキーマ（「テナントスキーマ」）に集約することを強くお勧めします。

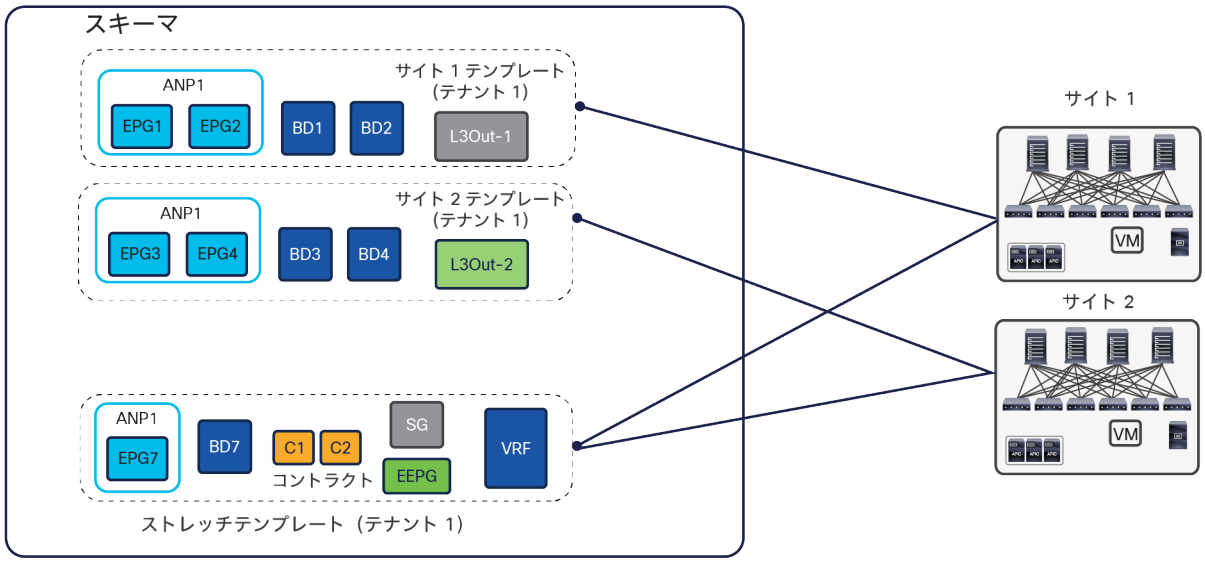


図 29.
NDO 3.x リリースでスキーマ内にアプリケーション テンプレートを定義する場合のベストプラクティス

専用アプリケーション テンプレートとは、マルチサイトドメインの各サイトに 1 対 1 でマッピングされたテンプレートを指します。ストレッチ アプリケーション テンプレートとは、すべてのサイトにマッピングされたテンプレートを指します。VRF とコントラクトは、通常、すべてのサイトで使用できる必要があるため、ストレッチ アプリケーション テンプレートで定義されます。BD や EPG は、ローカルであるか、サイトにまたがって拡張されているかに応じて、サイトの専用アプリケーション テンプレートまたはストレッチ アプリケーション テンプレートで定義されます。

このアプローチに従う場合、スキーマでサポートされるアプリケーション テンプレートの数とオブジェクトの総数の上限について、導入しているソフトウェアリリースの ACI Verified Scalability Guide（VSG）を確認することが重要です。大規模な展開では、各スキーマの検証済みでサポートの対象となる拡張範囲に収めるため、同じテナントに関連付けられたテンプレートを複数のスキーマに分けて展開することが必要になる場合があります。

図 29 に示すアプローチは、NDO 3.x ソフトウェアリリースのベストプラクティスの導入モデルを表しています。リリース 4.0(1) 以降の Nexus Dashboard Orchestrator では、テンプレートの設計と展開の際に、以下のようないくつかのベストプラクティスについて検証が行われ、これに従う必要があります。

- すべてのポリシーオブジェクトは、依存関係に従って正しい順序で展開する必要があります。

  たとえば、ブリッジドメイン（BD）を作成するときは、それを VRF に関連付ける必要があります。この場合、BD が VRF に依存するため、VRF を BD より前または一緒にファブリックに展開する必要があります。これらの 2 つのオブジェクトが同じテンプレートで定義されていれば、展開時に Orchestrator が VRF を最初に作成し、ブリッジドメインをそれに関連付けます。しかし、この 2 つのオブジェクトを別々のテンプレートで定義し、BD を定義するテンプレートを先に展開しようとすると、関連する VRF がまだ展開されていないため、Orchestrator が検証エラーを返します。この場合、最初に VRF を定義するテンプレートを展開してから、BD を定義するテンプレートを展開する必要があります。

- すべてのポリシーオブジェクトは、依存関係に従って正しい順序で展開を解除する必要があります。言い換えると、展開された順序と逆の順序で展開を解除する必要があります。

  同じ理由で、テンプレートの展開を解除するときは、他のオブジェクトが依存しているオブジェクトの展開を解除することはできません。たとえば、VRF に関連付けられている BD の展開を解除する前に、VRF の展開を解除することはできません。

- 複数のテンプレートの間で循環的な依存関係を作ることはできません。

  ブリッジドメイン BD1 が VRF1 に関連付けられ、EPG1 が BD1 に関連付けられている場合を考えてみます。VRF1 をテンプレート 1 に作成してこのテンプレートを展開し、次に BD1 をテンプレート 2 に作成してこのテンプレートを展開した場合、正しい順序でオブジェクトが展開されるため、検証エラーは発生しません。

  しかし、その後 EPG1 をテンプレート 1 に作成しようとすると、2 つのテンプレートの間に循環的な依存関係が作成されるため、EPG が新しく追加されたテンプレート 1 の保存を Orchestrator が許可しません。

このような追加のルールと要件を導入したことによる主な影響は以下の 2 つです。

1. NDO 4.0(1) で直接作成されたグリーンフィールド構成の場合、上の図 29 に示したベストプラクティスの推奨事項が、図 30 に示すようにわずかに変更されています。

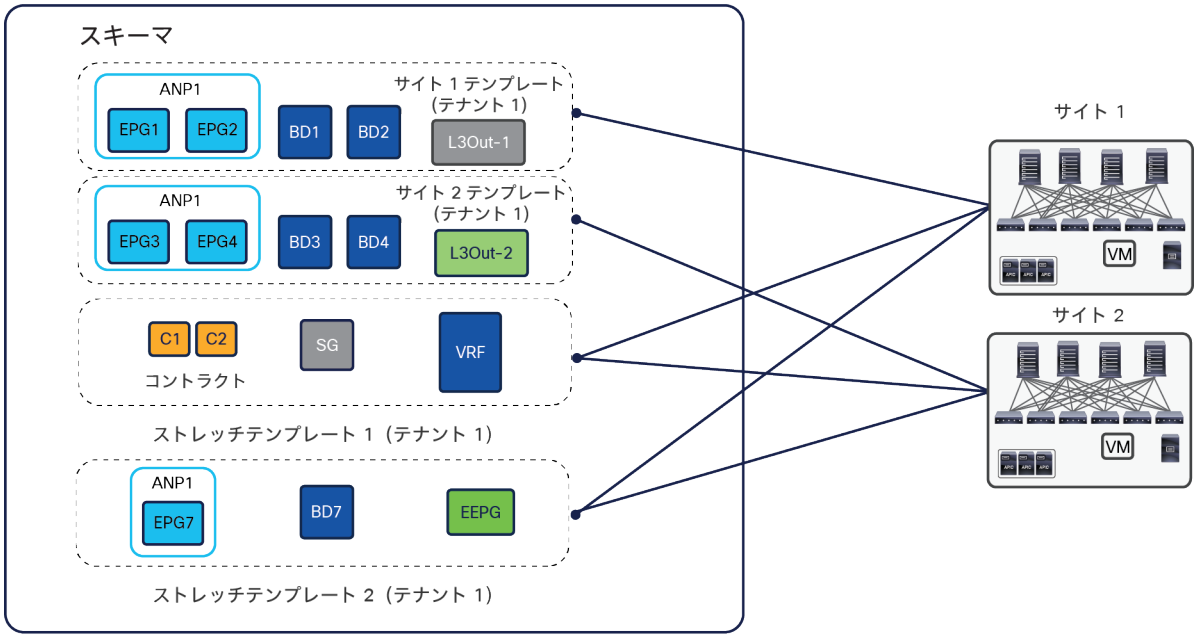


図 30.
NDO 4.x 以降のリリースでスキーマ内にアプリケーション テンプレートを定義する場合のベストプラクティス

主な違いは、NDO 4.0(x) 以降では、ストレッチテンプレートが 2 つ必要になる点です。このようにすると、たとえば、VRF と外部 EPG（その VRF を参照）を 2 つの別々のテンプレートで定義でき、これらのオブジェクト間の循環的な依存関係を回避できます。この場合、以前使っていた単一のストレッチテンプレートを展開しようとすると、エラーが発生します。

2. 以前のリリース 3.x からリリース 4.0(1) 以降にアップグレードするには、既存のすべてのテンプレートを分析し、上で説明した新しい要件を満たさないテンプレートを変換する必要があります。この分析は移行プロセス中に自動的に実行され、新しいベストプラクティスに準拠させるために既存のテンプレートを変更する必要がある場合、必要なすべての変更について詳細なレポートが出力されます。図 29 に示すベストプラクティスの導入モデルから始める場合、システムがオブジェクトを自動的に再編成して、図 30 に示す新しいベストプラクティスの導入モデルを再現します。

注： NDO 3.x リリースから NDO 4.x リリースにアップグレードするために必要な手順の詳細は、https://www.cisco.com/c/ja_jp/td/docs/dcn/ndo/4x/deployment/cisco-nexus-dashboard-orchestrator-deployment-guide-401/ndo-deploy-migrate-40x.html にあるドキュメントを参照してください。

ポリシーオブジェクトを編成する際の柔軟性をさらに高めるために、Cisco Nexus Dashboard Orchestrator では、同じテナントに関連付けられているアプリケーション テンプレートの間（同じスキーマ内またはスキーマ間）で EPG と BD を移行できるようにしています。この機能の一般的なユースケースは、サイト内でローカルに定義されていた EPG/BD ペアのサイトにまたがった拡張やその逆を開始できるようにする場合です。

自律型アプリケーション テンプレート（NDO リリース 4.0(1)）

NDO ソフトウェアリリース 4.0(1) では、「自律型テンプレート」と呼ばれる新しいタイプのアプリケーション テンプレートが導入されています。このタイプのアプリケーション テンプレートを使用すると、従来のアプリケーション テンプレート（「マルチサイトテンプレート」という名前に変更）と同じオブジェクト（EPG、BD、VRF など）をプロビジョニングできます。このテンプレートを単一のサイトまたは複数のサイトに関連付けることもできます。ただし、両タイプのアプリケーション テンプレートの基本的な違いは、「自律型テンプレート」を複数のサイトに展開しても、「ストレッチ」オブジェクト（およびそれに関連付けられた変換エントリ）が作成されないことです。変換エントリの使用については、このドキュメントの後のセクションで説明します。

図 31 に示すように、「自律型アプリケーション テンプレート」の展開のユースケースは、互いに独立して展開され運用される（言い換えると、異なる ACI ファブリックのスパインを接続する ISN インフラストラクチャがない）複数のファブリック間で同じ構成を複製する必要がある場合です。

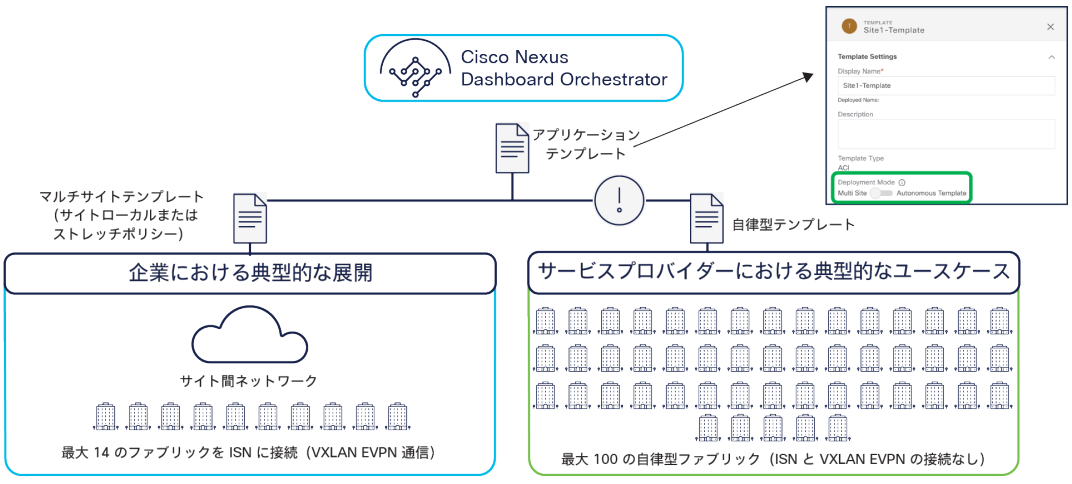


図 31.
マルチサイトテンプレートと自律型アプリケーション テンプレート

重要なので繰り返しますが、自律型テンプレートを複数のサイトに関連付けると、最終的に「ストレッチオブジェクト」が作成されるのではなく、単に同じ名前が付いた独立したオブジェクトが作成されます。たとえば、サイト 1 とサイト 2 に関連付けられた自律型アプリケーション テンプレートで VRF1 を定義した場合、両方の APIC ドメインで同じ VRF1 という名のオブジェクトが作成されます。ただし、機能の点では、L3Out データパスを介してのみ相互接続できる 2 つの完全に独立した VRF です。同じ名前で作成される EPG についても同様です。

これらのことを考慮すると、マルチサイトテンプレートと自律型アプリケーション テンプレートの展開にあたっては、具体的なガイドラインに従うことが重要です。

- マルチサイトテンプレートは、「マルチサイトが有効になっている」複数のサイトにのみ関連付ける必要があります。「マルチサイトが有効になっている」とは、各サイトに関連付けられた「マルチサイト」ノブがオンになっているだけでなく、各サイトのインフラストラクチャ構成が完全にプロビジョニングされ、サイト間ネットワーク（ISN）に接続できるようになっていることを意味します。すでに述べたように、マルチサイトテンプレートを展開すると、ストレッチオブジェクトのプロビジョニングが開始されますが、これには ISN を介した VXLAN サイト間通信が必要なためです。

注： NDO リリース 4.0(3) 以降では、Orchestrator がこのガイドラインを適用し、マルチサイトテンプレートが複数の「自律」サイトに関連付けられることを防止します。

- 自律型テンプレートは、マルチサイトが有効になっていて（すなわち、対応する「マルチサイト」フラグがそれらのサイトに対してチェックされていて）ISN を介して相互接続されているサイトにも関連付けることもできます。そのようにしても、サイトにまたがるストレッチオブジェクトが作成されないことを理解することが重要です。

NDO リリース 4.0(1) で導入された新しいテンプレートタイプ

NDO ソフトウェアリリース 4.0(1) 以降、Nexus Dashboard Orchestrator のプロビジョニング機能を拡張するために、新しいテンプレートタイプが導入されました。これらの新しいテンプレートを図 32 に示し、以下で簡単に説明します。

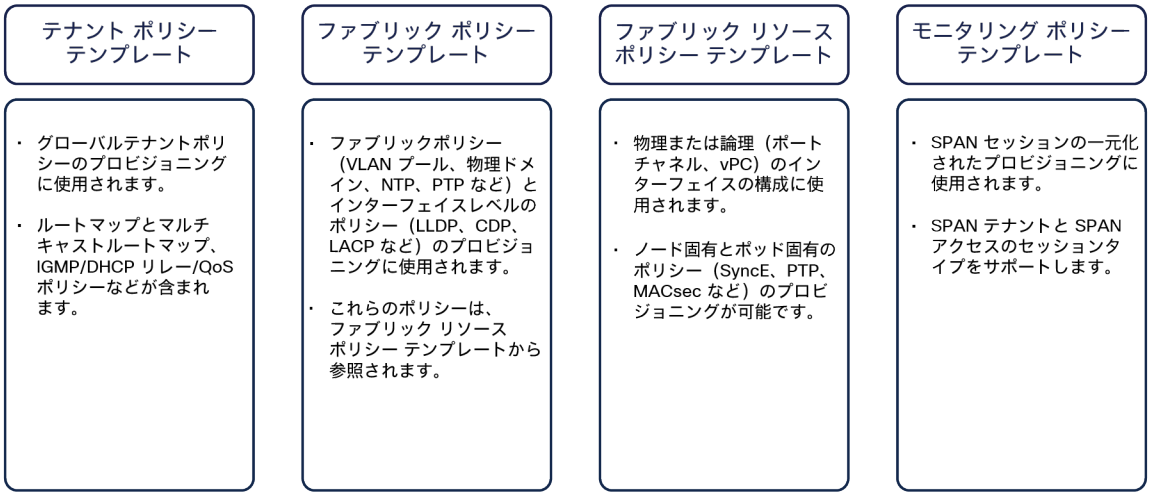


図 32.
NDO 4.0(1) で導入されたテンプレートタイプ

- テナント ポリシー テンプレート：このテンプレートは、テナントにポリシーをプロビジョニングするためのもので、さまざまな目的に使用することができます。たとえば、テナント ルーテッド マルチキャスト設定に使用されるルートマップ、または SR-MPLS L3Out やカスタム QoS ポリシーなどでアドバタイズされるルートを制御するルートマップが挙げられます。定義されたそれぞれのテナント ポリシー テンプレートは、1 つまたは複数のサイトに関連付けることができます。これば、ポリシーをサイトごとに固有にするか、サイトのグループに共通的に適用するかに依存します。
- ファブリック ポリシー テンプレートとファブリック リソース ポリシー テンプレート：これら 2 種類のテンプレートは、「ファブリック管理」テンプレートの例であり、ファブリック固有のポリシー（インターフェイスとインターフェイスのプロパティ、物理ドメインおよび関連する VLAN プールなど）のプロビジョニングに使用できます。NDO 4.0(1) より前は、このような構成は、APIC レベルごとに個別にプロビジョニングのみする必要がありました。テナント ポリシー テンプレートの場合と同様に、これらのファブリック管理テンプレートは 1 つまたは複数のサイトに関連付けることができます。これば、ポリシーをサイトごとに固有にするか、サイトのグループに共通的に適用するかに依存します。
- モニタリング ポリシー テンプレート：このタイプのテンプレートは、SPAN セッションをプロビジョニングするために使用できます。このセッションは、モニタリング用にトラフィックを複製して外部コレクタに送ります。テナント SPAN とアクセス SPAN の 2 種類の SPAN 設定がサポートされています。

注： これらの新しいテンプレートの構成に関する詳細な説明は、このホワイトペーパーの範囲外です。詳細なプロビジョニング情報は、https://www.cisco.com/c/en/us/td/docs/dcn/ndo/4x/configuration/cisco-nexus-dashboard-orchestrator-configuration-guide-aci-401/ndo-configuration-aci-fabric-management-40x.html にあるドキュメントを参照してください。

NDO の運用面での機能強化

NDO のソフトウェアリリースごとに、Cisco ACI マルチサイトアーキテクチャの運用面での簡素化と改善を目的とするいくつかの重要な機能強化が行われました。図 33 は、これらのテンプレートレベルの拡張機能を示しています。

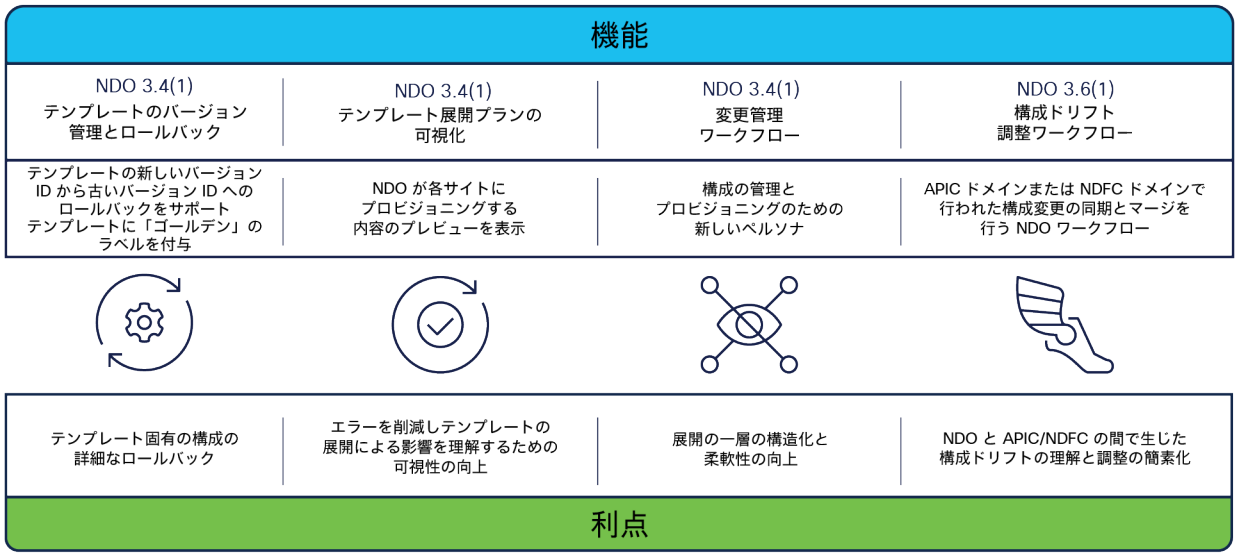


図 33.
NDO の運用面での機能強化

注： これらの機能を網羅した具体的な設定情報は、NDO 設定ガイド https://www.cisco.com/c/ja_jp/td/docs/dcn/ndo/3x/configuration/cisco-nexus-dashboard-orchestrator-configuration-guide-aci-371/ndo-configuration-aci-managing-schemas-37x.html?bookSearch=true にあります。

- テンプレートのバージョン管理とロールバック：Orchestrator の以前のリリースでは、構成のバックアップとロールバックはグローバルシステムレベルに限られていました。この機能は非常に便利で、ソフトウェアの最新バージョンでも引き続き使用できますが、より細かいレベルのアプローチが求められていました。NDO からのプロビジョニングの最小単位はテンプレート（タイプによらず）であるため、この要件に応えるためにテンプレートレベルでのバックアップとロールバックの機能が導入されました。

  NDO は、テンプレートの異なるバージョンを最大 20 まで追跡できます。また、特定のバージョンを「ゴールデン」として設定し、システムから自動的に削除されないようにすることもできます。古いバージョンのテンプレートを選択して最新のバージョンとの詳細な違いをグラフィック表示し（図 34 を参照）、選択した古いバージョンにテンプレートの構成をいつでもロールバックできます。

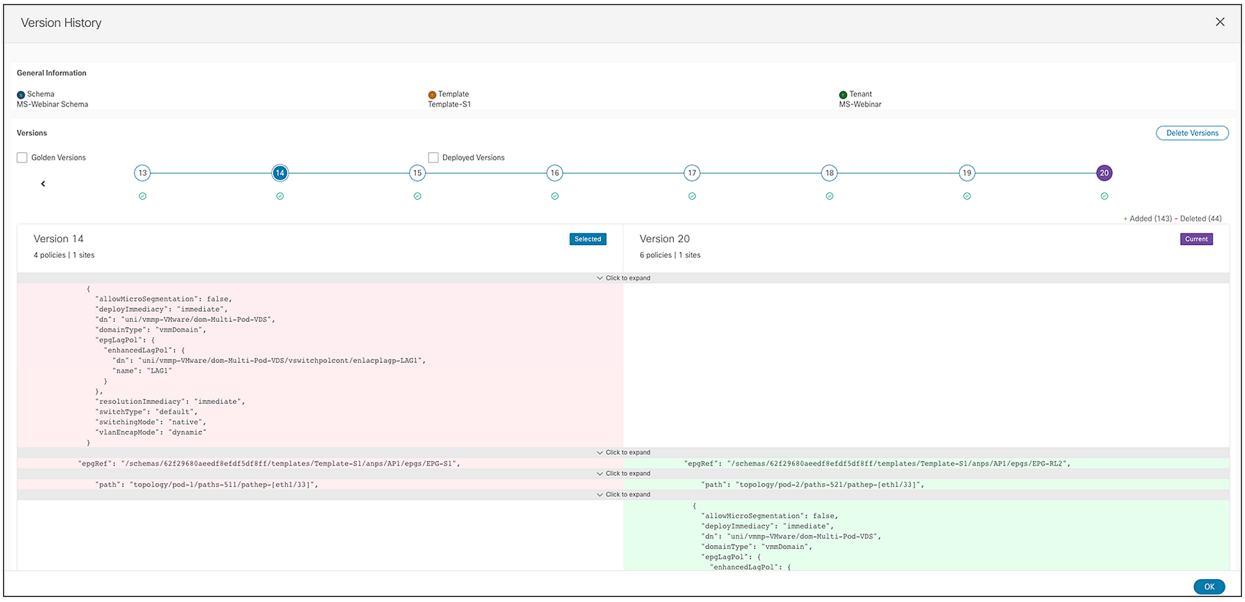


図 34.
テンプレートの改訂履歴

この最新の機能が特に興味深いのは、いわゆる「元に戻す」機能がシステムに組み込まれているためです。テンプレートを展開したタイミングで機能や接続に何らかの問題が発生した場合、正常に動作していたことがわかっている以前の展開に構成を速やかにロールバックできます。

テンプレートのバージョン管理とロールバックの詳細とライブデモについては、https://video.cisco.com/video/6277140235001 を参照してください。

- テンプレート展開プランの可視化：NDO を活用しているお客様が日々直面している主な課題の 1 つは、可視化の向上によって、テンプレートを展開するときに NDO がどの構成をどこ（APIC ドメイン）にプッシュしているかを把握できるようにすることです。テンプレートに作成された構成によっては、展開の際にテンプレートが関連付けられているサイトとは異なるサイトにオブジェクトがプロビジョニングされることが実際にあります。サイト間の VXLAN データパスを有効にするために作成されるシャドウオブジェクトがその一例です（シャドウオブジェクトの使用については、このドキュメントで詳しく後述します）。テンプレート展

開プランの可視化が導入された目的は、テンプレートの展開によって、どのような変更がどのサイトに適用されるかをグラフ形式と XML 形式の両方でユーザーに明確に示すことです。これにより、ユーザーは予期しない動作（設定不備、またはシステムのバグによる）を事前に把握することで、テンプレートの展開を中断して機能停止の可能性を防ぐことができます。図 35 は、テンプレート展開プランのグラフと XML による出力を示しています。

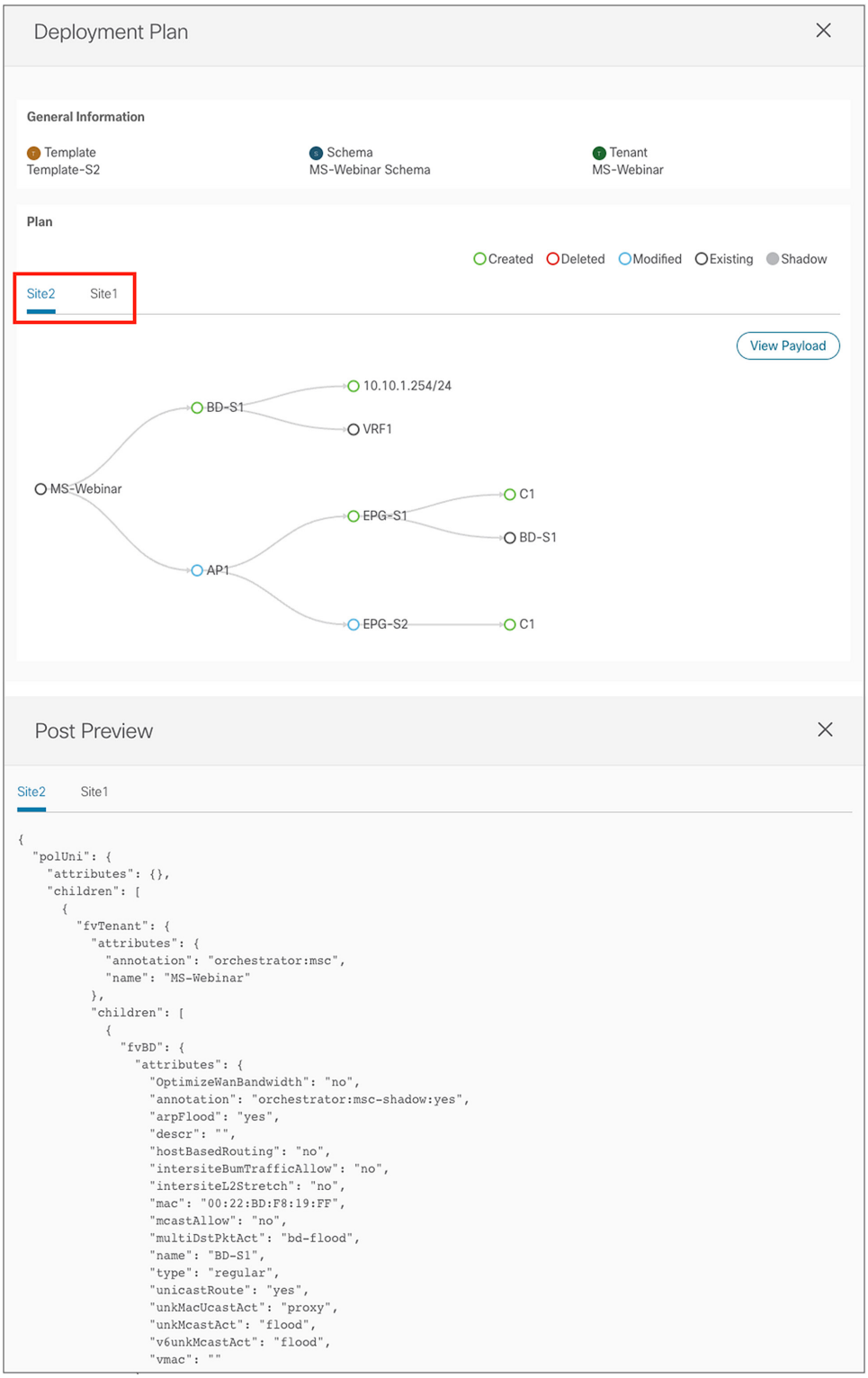


図 35.
テンプレート展開プランのグラフィカルビューと XML ビュー

テンプレート展開プランの可視化に関する詳細とライブデモについては、https://video.cisco.com/video/6277137504001 を参照してください。

- 変更管理ワークフロー：組織によって NDO の運用方法が異なります。ほとんどの場合、テンプレートの構成のプロビジョニングにはさまざまなタイプのユーザーが必要で、それぞれが異なる部分を担当します。そのため、遵守すべき具体的で厳格なルールがないと、変更をシステムに適用することができません。変更管理ワークフローが NDO に導入され、3 つの異なるタイプのユーザーロールを定義することが可能になりました。
  - 設計者：テンプレートの構成の作成または変更を担当します。
  - 承認者：設計者によって提案された構成変更を確認し、承認または否認を行います。また、テンプレートの展開に複数の承認者からの承認を必要とすることもできます。承認者がテンプレートの展開を否認した場合、否認の理由を説明するメッセージが設計者に送信されます。
  - 展開者：テンプレートの展開を担当します。展開者は、テンプレートの展開を拒否し、設計者にメッセージを返すこともできます。このとき、設計者は必要な修正アクションを実行できます。

  上記のロールは柔軟に定義でき、要件に応じてさまざまなロールを組み合わせることができます。また、この変更管理ワークフローは NDO に組み込まれていますが、当初から拡張可能な設計になっています。将来的には、外部の変更管理システムと統合できるようになる予定です。

  変更管理ワークフローの詳細とライブデモについては、https://video.cisco.com/video/6277140011001 を参照してください。

- 構成ドリフト調整ワークフロー：ACI マルチサイト展開では、常に NDO のみから構成をプロビジョニングすることをお勧めします。一方で、NDO によって APIC にプッシュされる（および APIC によって物理ネットワークデバイスにレンダリングされる）オブジェクトはロックされておらず、変更や削除が APIC から直接行われる可能性があります。
- したがって、ユーザーの意図に対する APIC と NDO の「ビュー」が同期しているかどうかを任意の時点で明確に把握できることが重要です。APIC と NDO でオブジェクトの構成が異なると「ドリフト」が発生します。APIC と NDO の間には通知チャネルがあるため、NDO は 2 つのシステムの「ビュー」を常に比較し、検出されたドリフト状態をユーザーに通知できます。これによって、ユーザーは適切な是正措置を講じることができます。

  NDO 構成ドリフト調整ワークフローは、以下のようにしてこの問題を解決します。まず、NDO が管理するオブジェクトに対して APIC で変更が適用された場合（またはその逆）、ユーザーにタイムリーに警告します。次に、グラフィカルで直感的なワークフローを通じて、ユーザーにそのドリフトを調整する選択肢を提供します。ユーザーは APIC にある構成を NDO にインポートするか、APIC の構成を NDO の構成で置き換えることができます。

バージョン間サポート

Cisco Multi-Site Orchestrator リリース 2.2(1) 以降、バージョン間サポートが導入され、異なる ACI ソフトウェアリリースを実行している APIC ドメインを MSO が管理できるようになっています（図 36）。

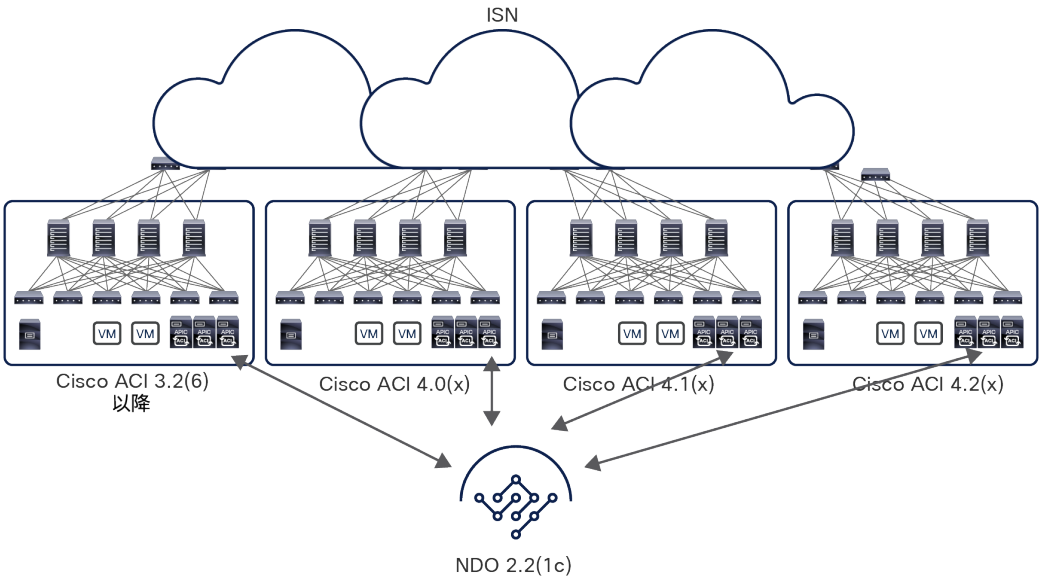


図 36.
Cisco Multi-Site Orchestrator リリース 2.2(1) 以降のバージョン間サポート

図 37 に示すように、Nexus Dashboard Orchestrator でもすべてのバージョンで同じ機能を使用できます。

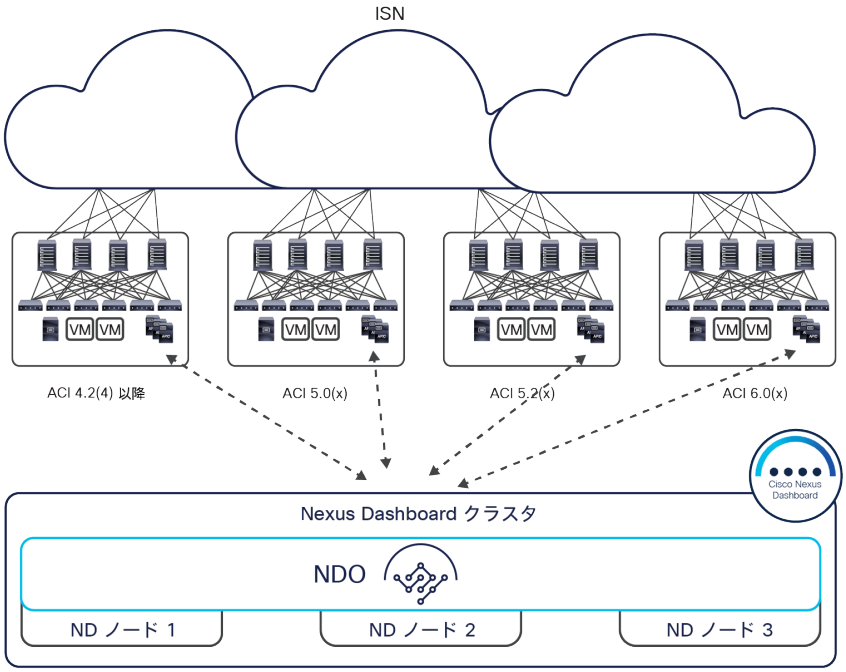


図 37.
Cisco Nexus Dashboard Orchestrator リリースのバージョン間サポート

NDO によって管理されるマルチサイトドメインで利用できる APIC ソフトウェアの最小バージョンは、Cisco ACI リリース 4.2(4) であることに注意してください。これは、Nexus Dashboard のコンピューティング プラットフォームでは、これより古いバージョンを実行しているファブリックのオンボーディングができないことによる制約です。

異なるバージョンにまたがる機能をサポートするために、NDO は接続された APIC ドメインの ACI バージョンを認識する必要があります。そのため、NDO と、NDO によって管理される各 APIC との間を WebSocket で接続し、この情報を取得します。この接続を使用する目的は、APIC がダウンしたときにこれを検出することで、復帰したときに NDO が APIC バージョンを照会できるようにすることです。たとえば、APIC をアップグレードしているときに、これが必要になります。

このようにして、所定のサイトに関連付けられたテンプレート上で構成された機能がそのファブリックで有効にサポートされているかどうかを、NDO が ACI リリースに基づいてチェックできます。下の表 1 では、主な機能とその機能がサポートされる ACI の最小リリースを一覧にしています。

表 1. ACI の機能とその機能がサポートされる APIC の最小バージョン

機能	APIC の最小バージョン
Cisco ACI マルチポッドのサポート	リリース 4.2(4)
サービスグラフ（L4 - L7 のサービス）	リリース 4.2(4)
外部 EPG	リリース 4.2(4)
Cisco ACI Virtual Edge VMM のサポート	リリース 4.2(4)
DHCP のサポート	リリース 4.2(4)
整合性チェッカー	リリース 4.2(4)
CloudSec 暗号化	リリース 4.2(4)
レイヤ 3 マルチキャスト	リリース 4.2(4)
OSPF の MD5 認証	リリース 4.2(4)
ホストベースのルーティング	リリース 4.2(4)
サイト間 L3Out	リリース 4.2(4)
vzAny	リリース 4.2(4)
SR-MPLS ハンドオフ	リリース 5.0(1)

NDO では、以下の 2 つの方法で APIC のバージョンがチェックされます。

- テンプレートの「保存」操作中：このチェックは、テンプレートの設計者が作成した構成が現在の APIC ソフトウェアリリースでサポートされていないことを早い段階で通知できる点で重要です。

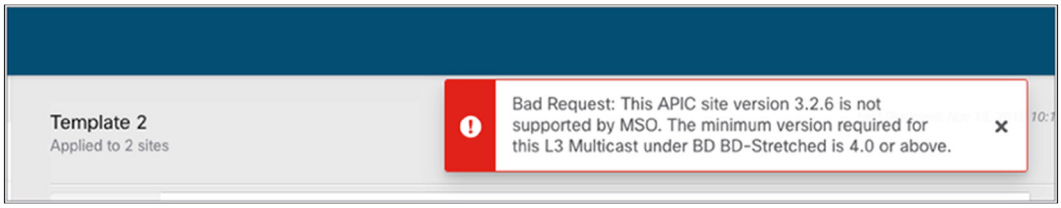


図 38.
テンプレートの「保存」操作中のバージョンチェック

- テンプレートの「展開」操作中：このチェックは、ユーザーが最初にテンプレートを保存せずに展開しようとした場合に必要になります。また、保存操作中のバージョンチェックはすべて通過したものの、実際にそのテンプレートを展開する前に APIC がダウングレードされた場合にも対応できます。

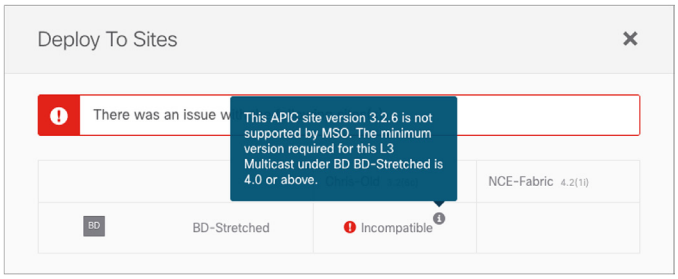


図 39.
テンプレートの展開時の APIC バージョンチェック

ブリッジドメインでの動作の観点から見た Cisco ACI マルチサイト

Cisco ACI マルチサイトアーキテクチャは、災害回避やディザスタリカバリといったさまざまなビジネス要件を満たすためのものとして位置づけられます。基本的に、それぞれのユースケースごとにブリッジドメインの接続シナリオが異なります。それぞれを以下のセクションで説明します。

注： 各ユースケースでの詳細な構成情報は、以下のリンクにある ACI ファブリック向け Cisco マルチサイト導入ガイドを参照してください。

https://www.cisco.com/c/ja_jp/td/docs/dcn/whitepapers/cisco-multi-site-deployment-guide-for-aci-fabrics.html

レイヤ 3 のみのサイト間接続

図 40 に示すように、展開シナリオの多くで、サイト間の通信をルーティングされた通信に限定することが基本的要件となっています。

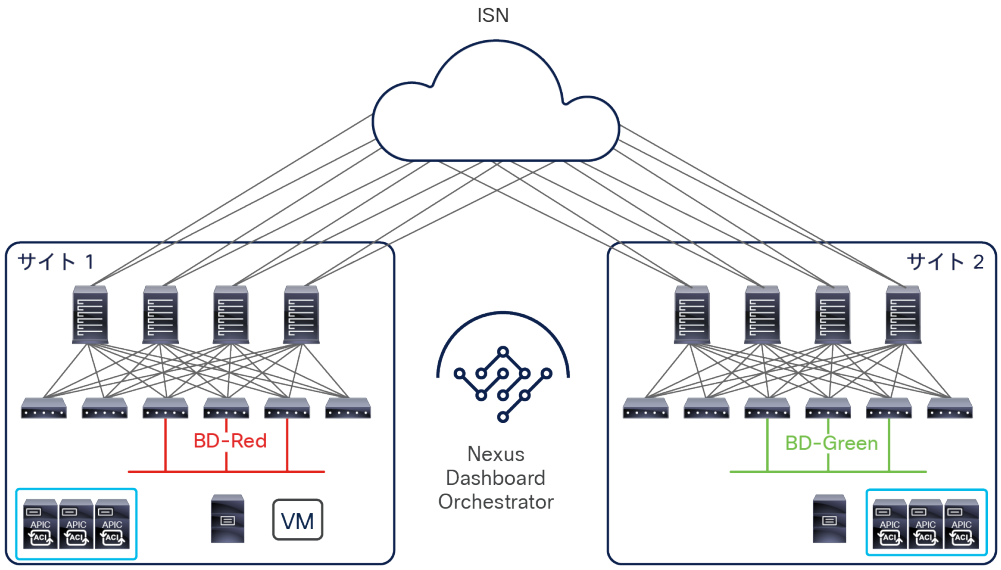


図 40.
レイヤ 3 のみのサイト間接続

このユースケースでは、レイヤ 2 の拡張やフラッディングは使用できません。また、それぞれのサイトで異なるブリッジドメインと IP サブネットが定義されます。Cisco ACI の原則として、EPG 間の通信は、適切なセキュリティポリシーが両者の間のコントラクトとして適用されない限り、確立されません。ただし、まずは接続することにのみ重点を置くために、ポリシーコンポーネントを削除することは可能です（たとえば、Cisco ACI リリース 4.0(2) 以降利用できる EPG の優先グループ機能、または Cisco Multi-Site Orchestrator リリース 2.2(4) で導入された vzAny の活用）。以下のようにさまざまなタイプのレイヤ 3 接続がサイト間で確立できます。

- VRF 内通信：これは、送信元 EPG と宛先 EPG が同じ VRF インスタンス（同じテナント）にマッピングされた異なるブリッジドメインに属している、よくあるケースのシナリオです。テナントと VRF インスタンスはサイト全体に拡張されます。さらに、MP-BGP EVPN を使用してホストルーティング情報が交換されるため、サイト間通信が可能になります（図 41）。

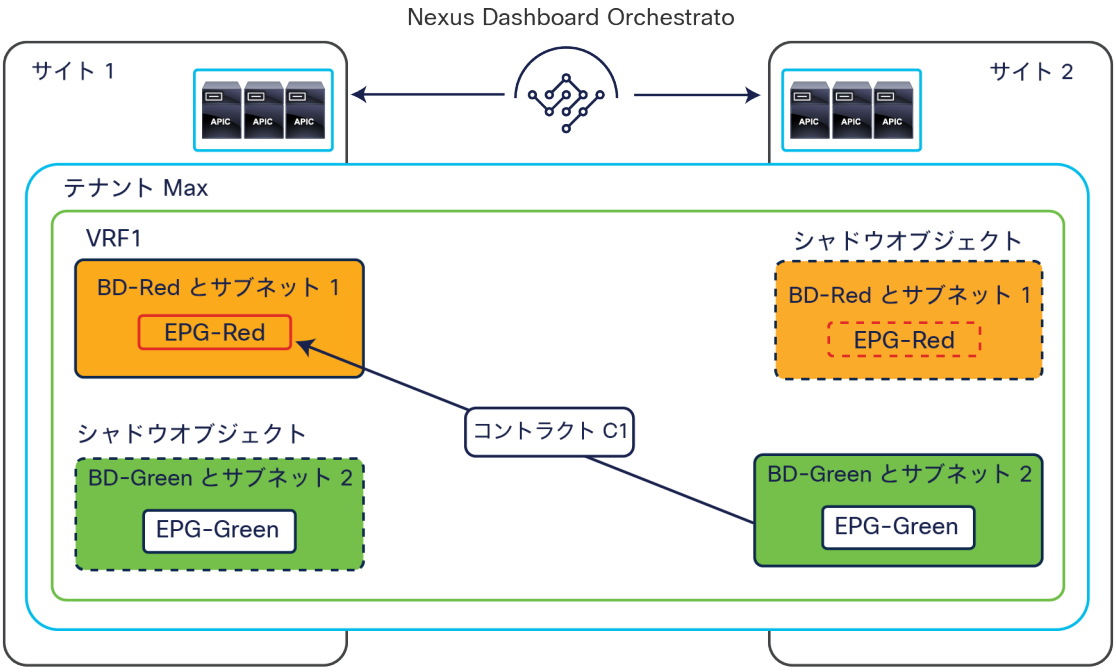


図 41.
VRF 内レイヤ 3 通信を用いたサイト間接続

EPG-Red と EPG-Green の間でコントラクトが確立すると、それぞれに対応するシャドウオブジェクトがリモートサイトに作成されます。これが作成されることで、スパインに変換エントリを正しくプログラミングできるようになります。

VRF 内レイヤ 3 通信を使用してサイト間を接続する具体的なユースケースを下の図 42 に示します。

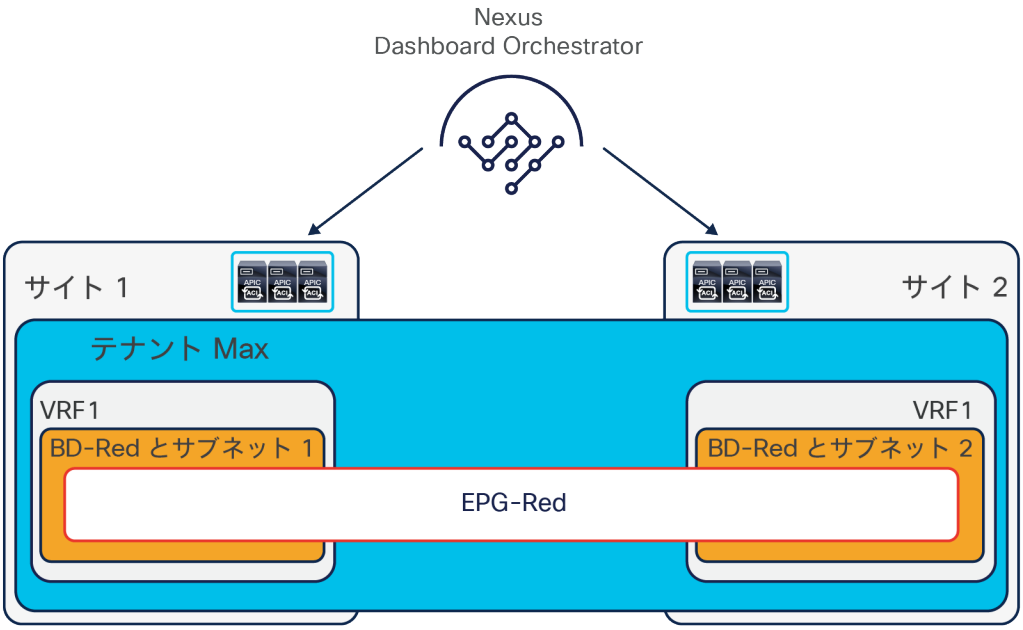


図 42.
VRF 内レイヤ 3 通信を用いた、拡張された EPG を共有するサイト間接続

このユースケースでは、EPG がサイト間で拡張されていますが、この EPG に関連付けられている BD は、レイヤ 2 ストレッチオブジェクトとして構成されていません。そのため、サイトごとに異なる IP サブネットを BD に割り当てることができます。その結果、この EPG に属するエンドポイント間のサイト間通信がルーティングされます。一方で、EPG 内通信はデフォルトで許可されているため、コントラクトを作成する必要はありません。

注：Orchestrator でこのユースケースを設定するには、サイト 1 とサイト 2 の両方に関連付けられた 1 つのテンプレートで EPG と BD の両方を定義する必要があります。次に、「L2 Stretched」フラグを無効にして BD を構成する必要があります。テンプレートの定義と ACI サイトへの関連付けの詳細は、「NDO のスキーマとテンプレートの展開」セクションを参照してください。

- VRF 間通信：これは、送信元と宛先のブリッジドメインが、異なる VRF インスタンス（同じまたは異なるテナントに属する）に属しているケースのシナリオです。この通信に、必要なルートリーク機能を動作させるには、送信元 EPG と宛先 EPG の間のコントラクトの作成とプロバイダー EPG でのサブネットの構成を行えば十分です。これは、単一サイトを展開するときに APIC で必要になる手順と同一です。図 43 に示すように、EPG 間でコントラクトが確立すると、リモートサイトにシャドウオブジェクトが作成されます。

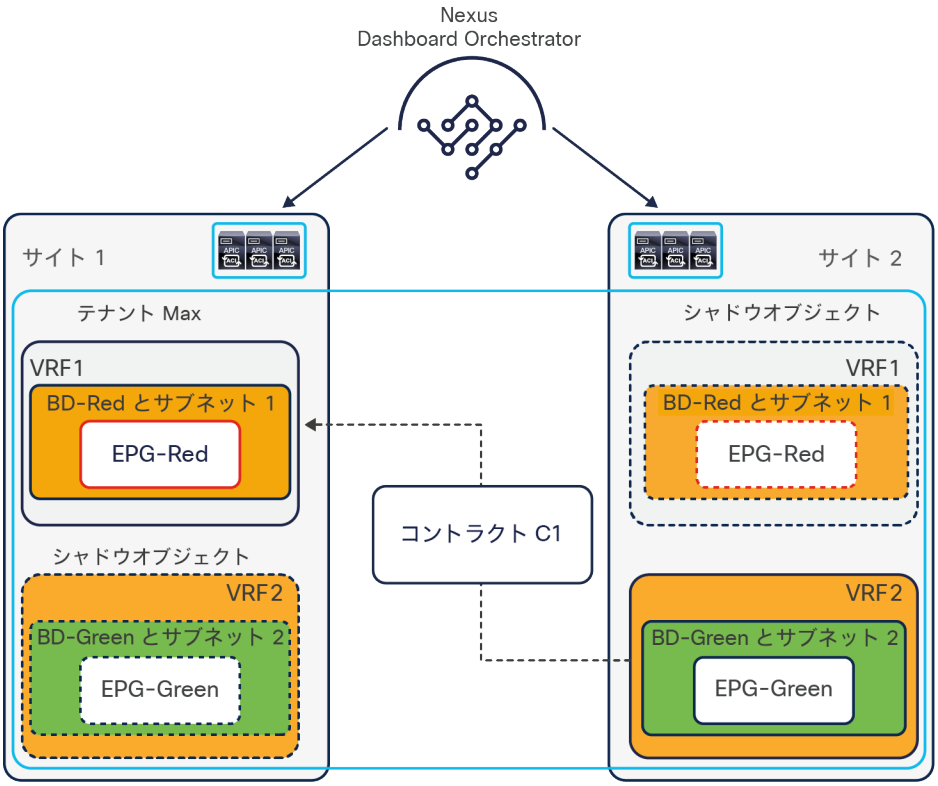


図 43.
VRF 間レイヤ 3 通信を用いたサイト間接続

- 共有サービス：上で説明した VRF 間通信のシナリオの一例として、共有サービスが個別の VRF インスタンスまたはテナントで提供され、これにアクセスする必要があるそれぞれの VRF インスタンスまたはテナントに複数の送信元 IP サブネットが設定されるケースがあります（図 44）。多対 1 の接続が必要な典型例ですが、必要なルーティング情報を交換できるようにするには、送信元と宛先の EPG 間で適切なセキュリティポリシーを確立すれば十分です。

注：下の図では、わかりやすくするためにシャドウオブジェクトを省略しています。

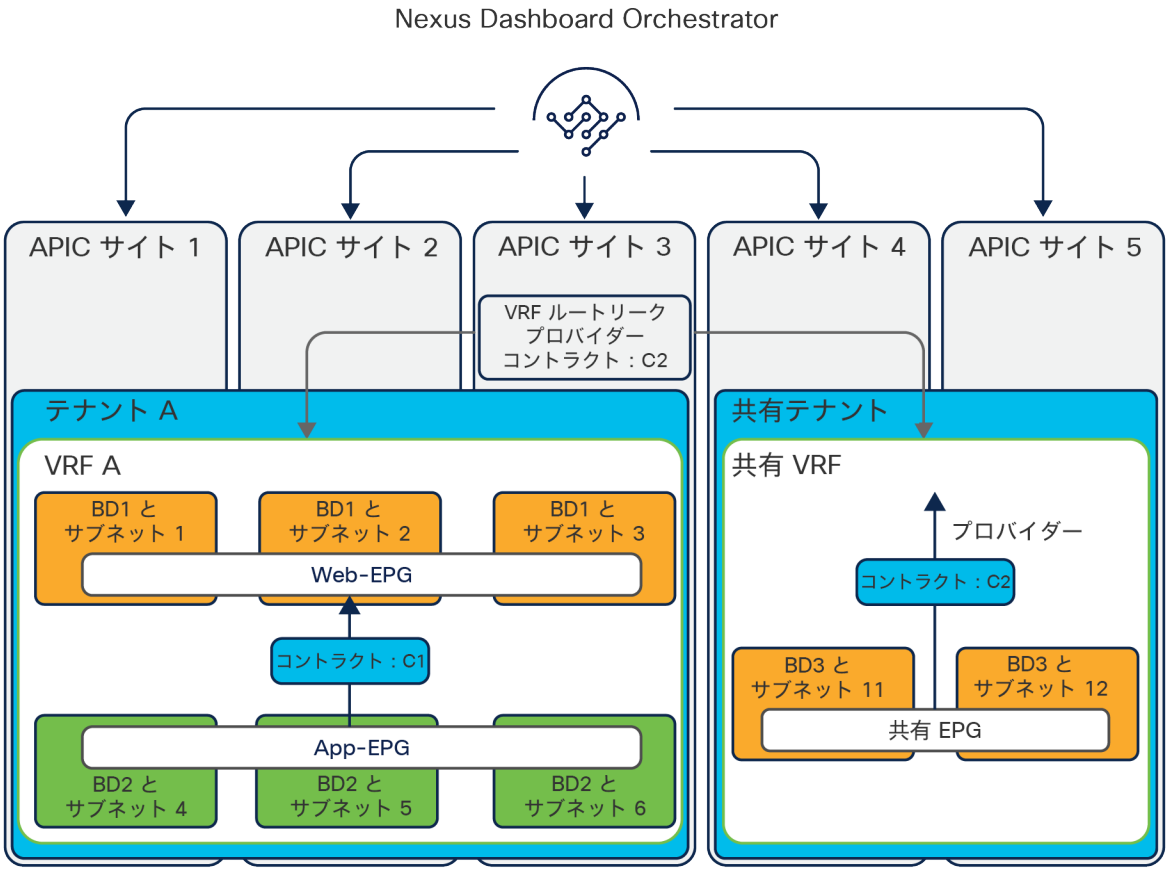


図 44.
VRF 間通信を用いたサイト間接続（共有サービス）

レイヤ 3 のみでサイト間を接続する Cisco ACI マルチサイト展開を検討していると、よく浮かぶ疑問として、このユースケースでサイト間ネットワークを接続するためになぜ VXLAN データパスを使うのだろうか、ボーダーリーフ（BL）ノードから L3Out の論理接続を介してそれぞれの Cisco ACI ファブリックを相互接続するだけではなぜいけないのか、などがあります。

図 45 は、この展開モデルと、これを展開しようとすると考慮しなければならない主な点を表しています。

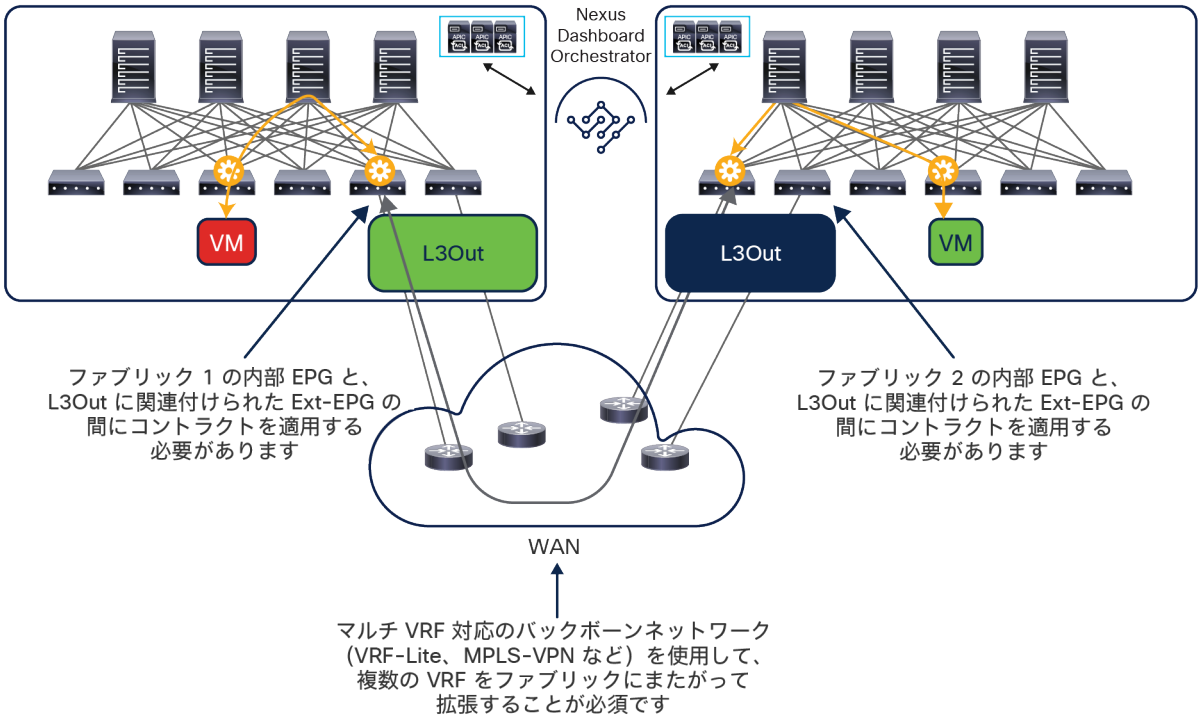


図 45.
L3Out を介したレイヤ 3 通信を実現した Cisco ACI ファブリックの相互接続

- 接続の観点から見ると、マルチ VRF 通信（マルチテナント展開）をサポートするには、論理的に分離されたルーテッドドメイン（VRF または L3VPN）を提供できる外部ネットワークを用いて、ファブリックを相互接続する必要があります。そのためには、さらにコストがかかるオプション（たとえば、サービスプロバイダーから複数の L3VPN サービスを購入）またはさらに複雑なオプション（ほんの数例を挙げると、VRF-Lite エンドツーエンドや MPLSoGRE など）を採用する必要があります。一方、「ボーダーリーフノードでの SR-MPLS/MPLS ハンドオフ」セクションで説明するように、サイト間フローを WAN 全体で包括的に可視化することが望まれる場合があります。これが実現すれば、フローを区分して優先順位を付けたり、トランスポートチームが既存のモニタリングツールを使用してデータセンター間フローをモニタリングしたりといったことが可能になります。前述のように、マルチサイトネイティブの VXLAN データパスを使用すると、外部ネットワークはルーティングが可能なインフラストラクチャを提供するだけで済みます。このインフラストラクチャを活用してサイト間 VXLAN トンネルが確立され、マルチテナントのルーティングが可能になります。
- ポリシーの観点から見ると、図 45 に示した相互接続のオプションでは、ポリシードメインが切り離されたままになっています。これは、トラフィックを外部ルータに送る前に VXLAN カプセル化がボーダーリーフノードで終了してしまい、データプレーントラフィックでポリシー情報が配信されないためです。そのため、トラフィックがリモート Cisco ACI ファブリックに入る前に適切に再分類される必要があり、別々のファブリックにある EPG 間で通信するには、ファブリックごとに個別のポリシーを作成することが必須になります。その結果、運用上のエラーによってトラフィックがドロップされる可能性が高まります。ネイティブのマルチサイト機能を使用すれば、単一のポリシードメインがそれぞれの Cisco ACI ファブリックにまたがって拡張されます。Cisco Nexus Dashboard Orchestrator で単一のシンプルなポリシーを定義することで、異なる EPG に属するエンドポイントがサイト間でセキュアに通信できるようになります。

- 最後に、非常に多い例として、レイヤ 3 のみでサイト間を接続するという当初の要件が、さまざまな IP モビリティのユースケースに対応するために IP サブネットを拡張するという要件に発展することがあります。マルチサイトを使用しない場合、これには外部ネットワークに別のレイヤ 2 データセンター相互接続（DCI）テクノロジーを展開する必要があります。この展開オプションは内部で検証されておらず、推奨もされていません。以下の 2 つのセクションで詳しく説明するように、Cisco Nexus Dashboard Orchestrator でシンプルな構成を行うことで、ブリッジドメインをサイトにまたがって拡張できます。

重要：図 45 に示す展開モデルには、サイト間接続にネイティブ VXLAN データパスを使用する場合と比較して欠点がありますが、レイヤ 3 のみでサイト間を接続するという要件を満たす完全にサポートされた実行可能なオプションです。このような設計に Nexus Dashboard Orchestrator を導入した場合でも、そのメリットが変わらず実現します。つまり、各ファブリックでの構成のプロビジョニングや Day-2 運用に関するアクティビティを一元的に行えます。これは特に、このドキュメントで前述した「自律型ファブリック」を管理するために NDO を利用する場合にも当てはまります。ただし、どちらかのオプションのみを選択することを強くお勧めします。「マルチサイト」の別々のサイトに展開された EPG 間の水平方向接続にネイティブ VXLAN データパスと L3Out パスを混在させて使用することは避ける必要があります。その理由を理解するために、下の図 46 に示す例を参照してください。

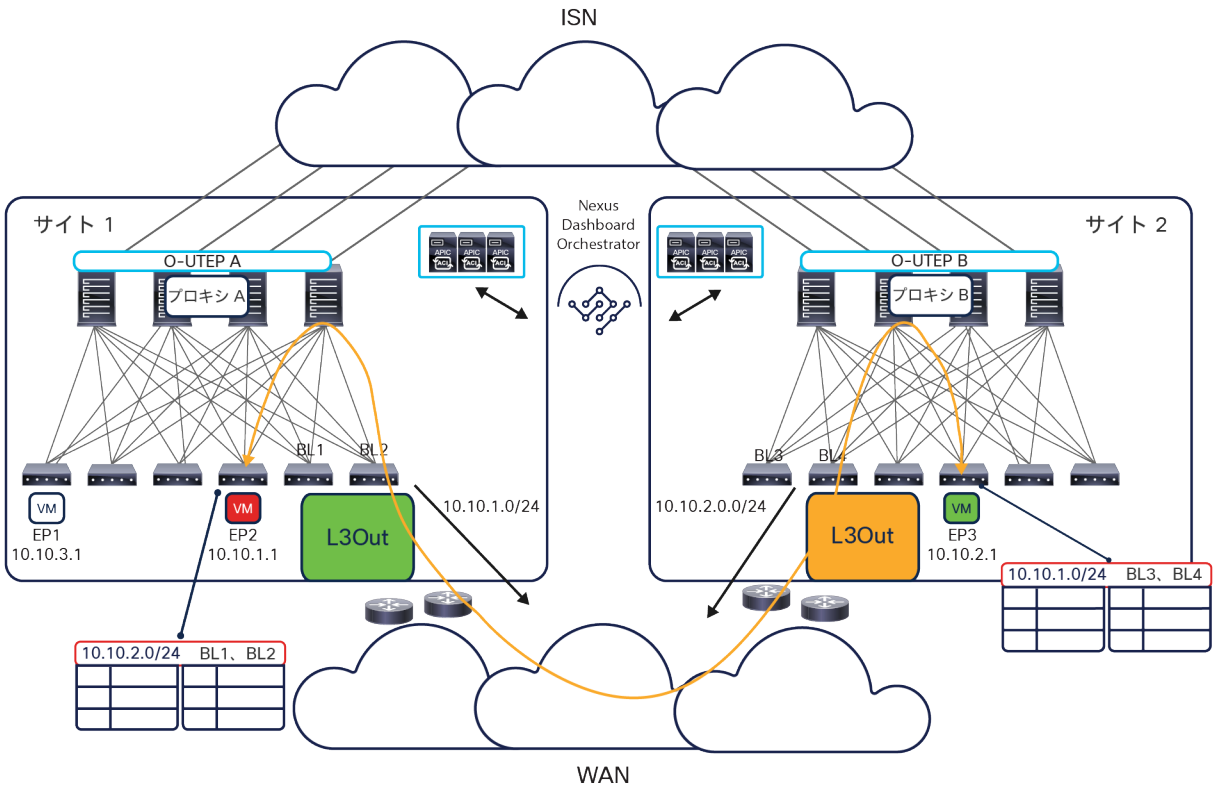


図 46.
初期状態：L3Out パスを使用したレイヤ 3 のサイト間通信

サイト 1 でローカルに定義された Red EPG/BD に属する EP2 が、サイト 2 でローカルに定義された Green EPG/BD に属する EP3 と通信しています。Green BD（10.10.2.0/24）の IP サブネットがサイト 1 の BL ノードの L3Out に到達し、Red BD（10.10.1.0/24）の IP サブネットについても逆が成立するため、このルーティングされた通信は L3Out パスを介して確立されています。

次に、サイト 1 でローカルに定義された Blue EPG/BD に属する EP1 がネットワークに接続されたとします。Blue EPG と Green EPG の間のコントラクトが NDO で作成されると、EP1 は ISN を通過する VXLAN を使用して通信するようになります（図 47）。

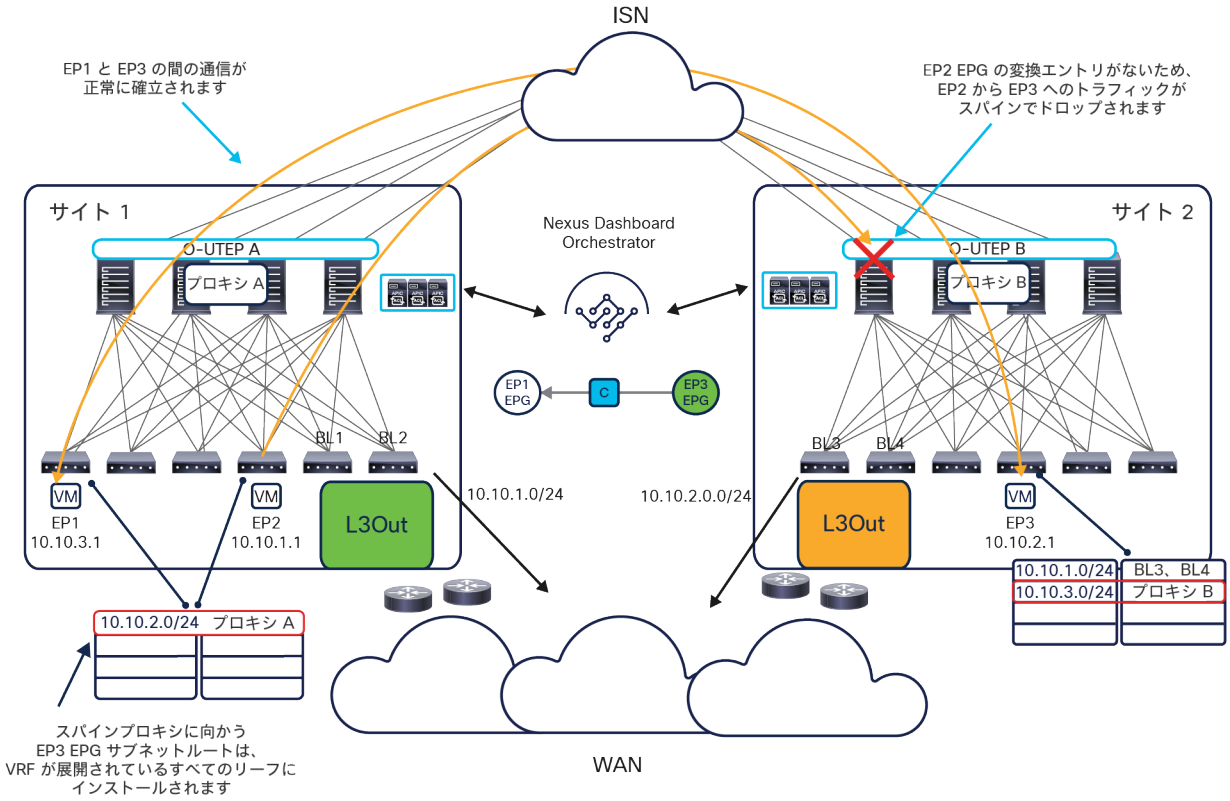


図 47.
VXLAN と L3Out のトラフィックパスを混在させる場合の問題

Blue EPG と Green EPG の間でコントラクトを作成すると、その直接の結果として、両方のサイトにシャドウオブジェクトが作成されます。特に、サイト 1 に Green BD のシャドウオブジェクトが作成されると、対応する VRF がインストールされているすべてのリーフノードのルーティングテーブルに Green BD サブネット（10.10.2.0/24）へのルートもインストールされます。このルートはプロキシスパイン VTEP アドレスを指しているため、ISN を介した通信が行われます。図 47 に示すように、このルート情報は以前に L3Out から学習し、Red EPG と Green EPG の間の通信を確立するために使用していた情報を置き換えます（直接接続されたルートとしてインストールされるため）。

その結果、Red EPG と Green EPG の間のトラフィックも、ISN を通過する VXLAN データパスを介して送信され始めます。しかし、Red EPG と Green EPG 間のコントラクトが NDO で構成されていないため、Red EPG/BD のシャドウオブジェクトがサイト 2 に作成されていません。その結果、サイト 2 のスパインがトラフィックを受信したときに、対応する変換エントリがないため、そのトラフィックがドロップされます。

図 47 に示すのは一例にすぎず、予期しないトラフィックのドロップにつながる唯一のシナリオではありません。したがって、問題を回避するには、レイヤ 3 接続の確立に L3Out データパスを用いるか VXLAN データパスを用いるかを事前に明確に決定することが重要です。

フラッディングを使用しないレイヤ 2 のサイト間接続

Cisco ACI マルチサイトアーキテクチャは、サイト間の IP モビリティのサポートにレイヤ 2 フラッディングを必要としていません。これは、重要で非常に独特な機能です（図 48）。

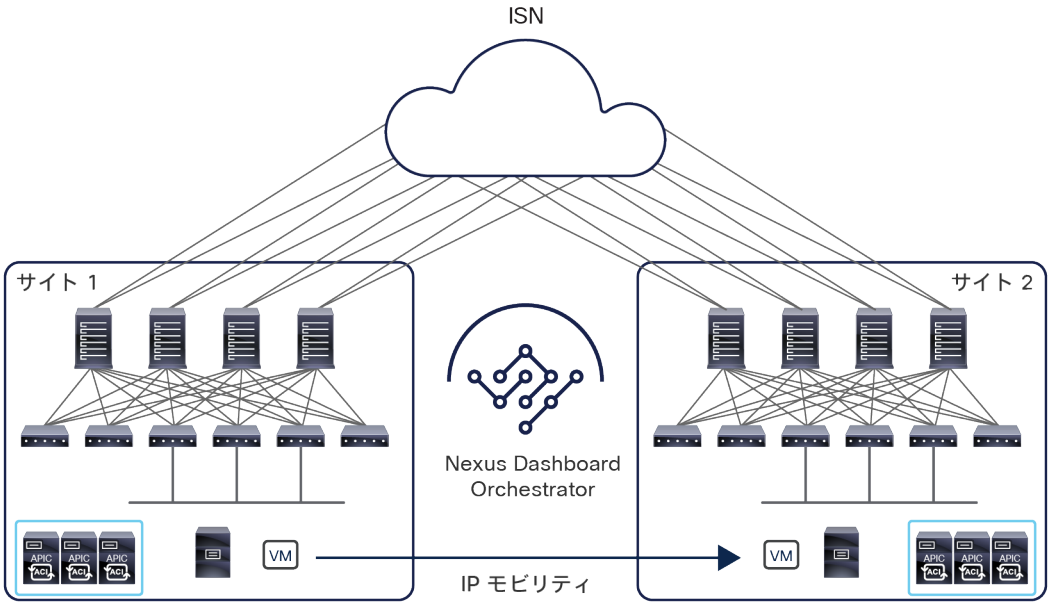


図 48.
フラッディングを使用しないレイヤ 2 のサイト間接続

IP モビリティのサポートが必要とされる主なシナリオには以下の 2 つがあります。

- ディザスタリカバリのシナリオ（コールドマイグレーション）では、当初ファブリック 1 で実行されていたアプリケーションが別のサイトに移動されます。このアプリケーションにアクセスするために使用されていた IP アドレスを変更し、DNS ベースのメカニズムを利用してクライアントをその新しい IP アドレスに向けることでこれを実現できますが、元のサイトで実行されていたときにアプリケーションが持っていた同じ IP アドレスを維持することが望ましい場合も多くあります。
- 事業継続のシナリオ（ライブマイグレーション）では、移行中のサービスへのアクセスを中断させることなく、ワークロードを一時的に他のサイトに再配置することが望まれます。この目的に使用できる機能の典型的な例は、vSphere vMotion です。これについては、「仮想マシンマネージャ統合モデル」セクションの中で詳しく説明します。

注： サイト間のライブマイグレーション（ブロードキャスト、宛先不明のユニキャスト、マルチキャスト（BUM）フラッディングの有効化の状況にかかわらず）は、Cisco ACI リリース 3.2(1) 以降で正式にサポートされています。

図 49 の論理構成図に示すように、このユースケースでは、すべてのオブジェクト（テナント、VRF インスタンス、ブリッジドメイン、EPG）をサイトにまたがって拡張する必要があります。ただし、それぞれのブリッジドメイン構成では、サイトにまたがるブロードキャスト、宛先不明のユニキャスト、マルチキャスト（BUM）のフラッディングが許可されないように指定する必要があります。

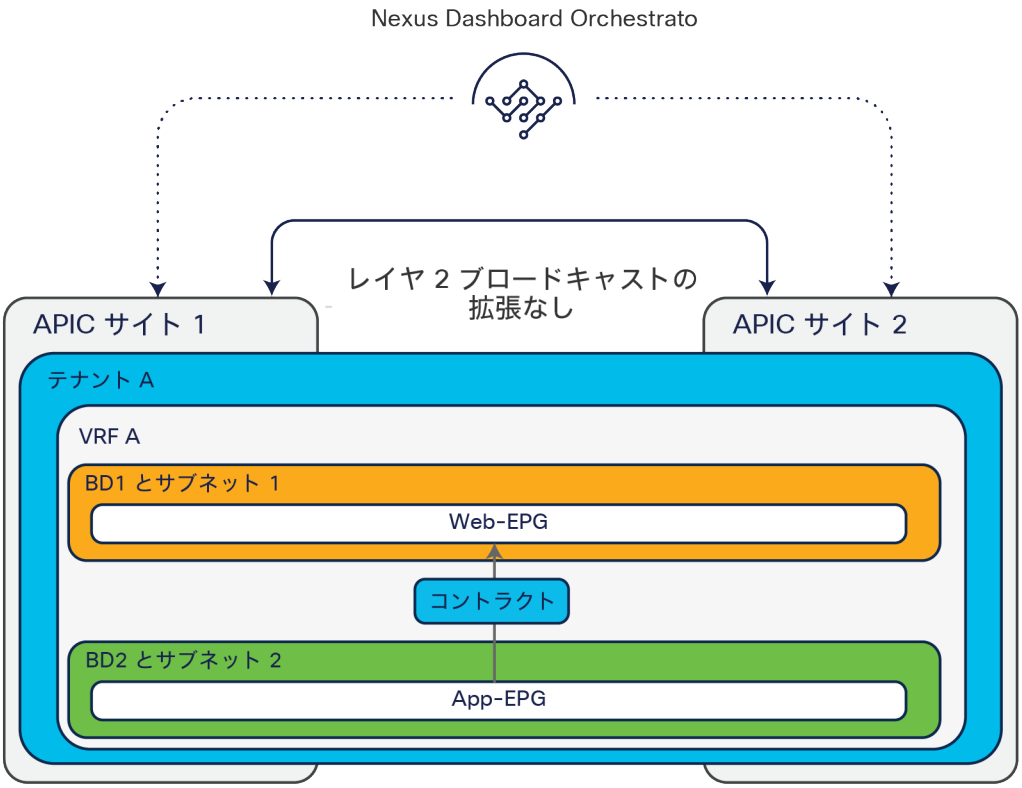


図 49.
フラッディングを使用しないレイヤ 2 のサイト間接続（論理構成図）

BUM フラッディングを使用しない IP モビリティのサポートは、Cisco ACI マルチサイトアーキテクチャによって提供される独自の機能であり、アーキテクチャの全体としての復元力を損なうことなくサイト間を柔軟に接続するために不可欠です。あるサイトのブリッジドメインで発生した問題（ブロードキャストストームなど）は、実際にそのサイト内に限定され、接続された他のファブリックに影響を与えることはありません。

IP モビリティのユースケースをサポートするために、Cisco ACI マルチサイト ソリューションには以下のような重要な機能がいくつか用意されています。

- 再配置されたエンドポイントは、同じ（または異なる）IP サブネットに属する他のエンドポイントと通信できる必要がありますが、このサブネットが依然として元のサイトに接続されている可能性があります。この要件に対処するには、元のサイトから、再配置されたエンドポイントに対して ARP 要求が送信された場合、Cisco ACI マルチサイトアーキテクチャが、その ARP 要求をリモートサイトのエンドポイントに配信する必要があります。移行されたエンドポイントの IP アドレスがファブリックによって検出されていた場合（通常はコールドマイグレーションのシナリオの場合）、ARP 要求をユニキャストモードでサイトにまたがって配信できます（VXLAN カプセル化を行い、宛先エンドポイントが現在接続されている新しいサイトを識別するエニーキャスト TEP アドレスを宛先として、そのスパインに直接配信します）。

  移行されたエンドポイントの IP アドレスが最初に検出されない場合（ホットマイグレーションで、移行されたエンドポイントが移動後も「サイレント」のままである場合）、元のサイトのスパインが「ARP Glean」機能を実行してサイレントホストに強制的に ARP 応答を発信させ、このホストを検出します。その時点で、前の段落で説明したように、新しく発信された ARP 要求がユニキャストモードで配信されます。

注：ARP Glean 機能は、1 つ以上の IP アドレスが定義されているブリッジドメインでのみサポートされ、レイヤ 2 のみのブリッジドメインではサポートされません。

- 外部レイヤ 3 ドメインから発信されたトラフィックを、再配置されたエンドポイントに配信する必要があります。外部ネットワークとの間でレイヤ 3 外部（L3Out）接続が確立される方法に応じて、いくつかの異なるシナリオを考慮する必要があります。

  ボーダーリーフ（BL）ノードでの従来型の L3Out 接続：両方のサイトに同じ IP サブネットが展開されているため、通常、同じ IP プレフィックス情報が 2 つのサイトから WAN に送信されます。この動作によって、デフォルトでは、着信トラフィックがサイト 1 またはサイト 2 に無差別に配信される可能性があります。しかし一般的には、2 つのサイトのいずれかをその IP サブネットのホームサイト（定常状態で、そのサブネットのほとんどのエンドポイントが接続されるサイト）に指定します。この場合、WAN に送信されるルート更新を適切に調整することで、図 50 に示すように、すべての着信トラフィックをホームサイトにステアリングすることができます。

  注：この「ホームサイト」のユースケースに関する詳細は、このホワイトペーパーの「Cisco ACI マルチサイトとボーダーリーフノードでの L3Out 接続」セクションを参照してください。

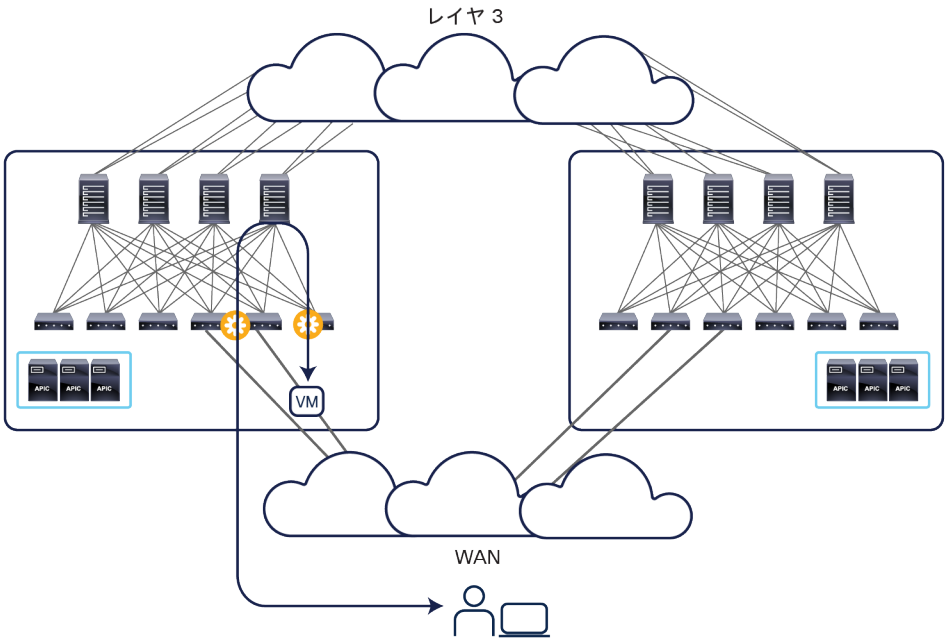


図 50.
ホームサイトにステアリングされた入力トラフィック

エンドポイントをサイト 2 に移行しても、上記の動作は変更されない可能性が高く、入力トラフィックは引き続きその IP サブネットのホームサイトにステアリングされます。したがって、Cisco ACI マルチサイトアーキテクチャには、図 51 に示すように、トラフィックフローをサイト間 IP ネットワークを通して宛先のエンドポイントに到達させる機能が必要です。

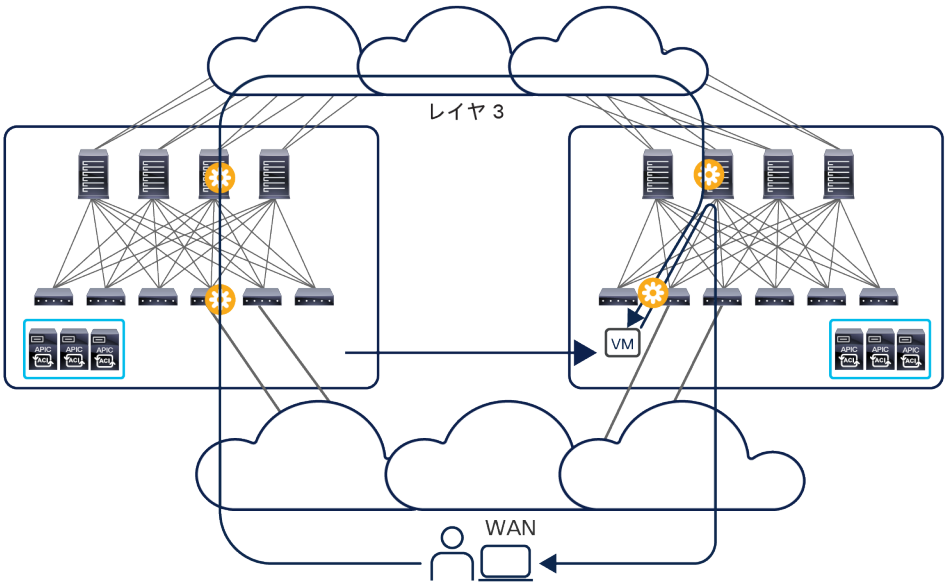


図 51.
移行されたエンドポイントへ向かうトラフィックのサイト間でのバウンス

繰り返しますが、この動作が可能なのは、移行されたエンドポイントがサイト 2 で動的に検出され、サイト 2 のスパインノードからサイト 1 のスパインノードへの EVPN 更新がトリガーされるためです。この更新の基本的な役割は、検出されたエンドポイントのロケーション情報をサイト 1 に提供することです。図 51 に示すようなトラフィックフローのリダイレクトを行うには、この情報が必要です。

移行後、移行されたエンドポイントからリモートクライアントへのリターントラフィックが発生すると、ファブリック 2 のローカル L3Out 接続の使用が開始されます。これによって、非対称のトラフィックパスが作成されることに注意してください。ファブリックと WAN の間に、独立したステートフル ファイアウォールが展開されている設計では、この動作によってトラフィックがドロップされる可能性があります。この問題は、最も限定性の高いホストルート情報を WAN にアドバタイズする新しい機能を活用することで解決できます。この機能は Cisco ACI リリース 4.0(1) 以降、ボーダーリーフ L3Out でサポートされています。

図 52 は、ホームサイトから移動されたエンドポイントに限ってホストルートのアドバタイズを行う方法を示しています（WAN に注入されるホストルート情報の量を減らすことが目的です）。マルチサイトとボーダーリーフ L3Out の統合に関するその他の考慮事項については、セクション「Cisco ACI マルチサイトとボーダーリーフノードでの L3Out 接続」を参照してください。

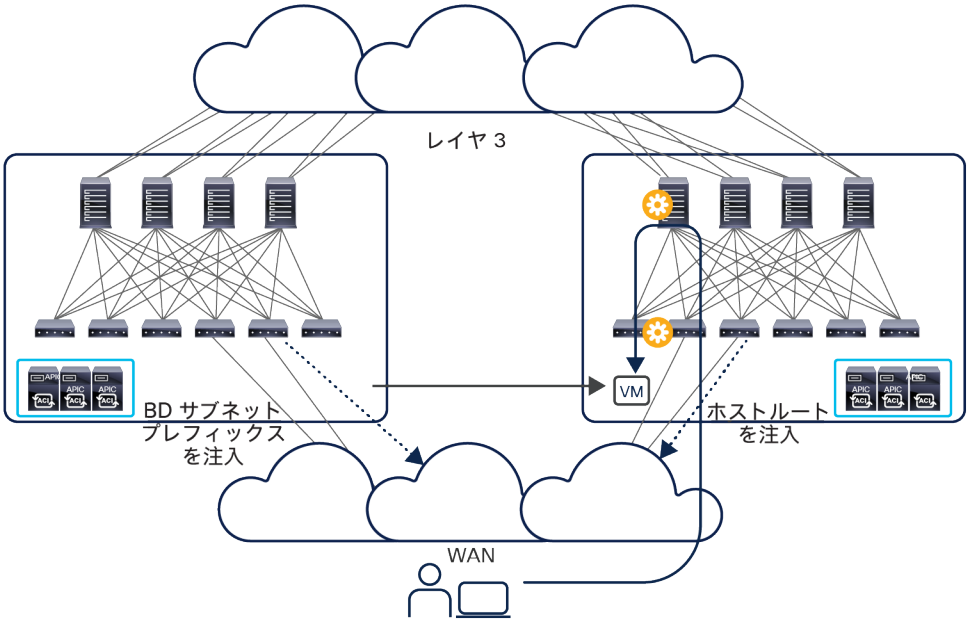


図 52.
ボーダーリーフ L3Out でのホストルートのアドバタイズ

b) GOLF L3Out：Cisco ACI リリース 3.0(1) 以降は、GOLF L3Out 接続の展開が可能です。この機能は、IP サブネットに関する情報に加え、そのサブネットに接続されたエンドポイントのホストルートに関する情報を WAN エッジデバイスに送信することを目的に導入されました。この場合も、入力トラフィックがそのエンドポイントがあるサイトに常にステアリングされるようにすることができます（図 53）。

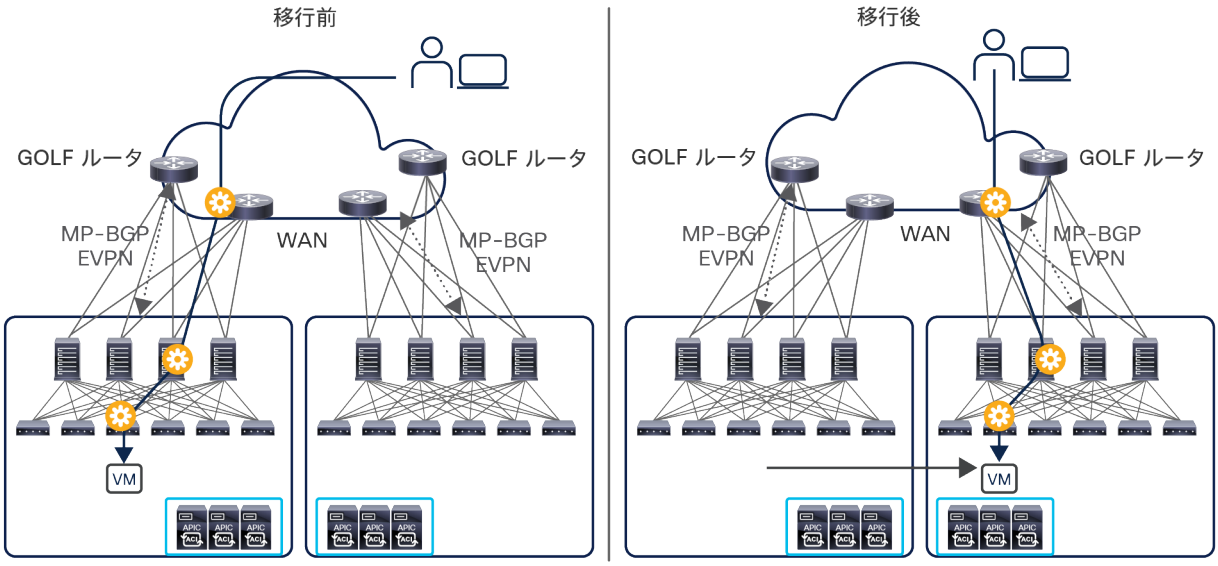


図 53.
GOLF L3Out による入力トラフィックの最適化

マルチサイトと GOLF の統合に関するその他の考慮事項については、付録 C を参照してください。

フラッディングを使用したレイヤ 2 のサイト間接続

次のユースケースの Cisco ACI マルチサイト設計では、ブリッジドメインをサイトにまたがって拡張する従来型のレイヤ 2 ストレッチを実現しています。これには、レイヤ 2 BUM フレームをフラッディングする機能が含まれます（図 54）。

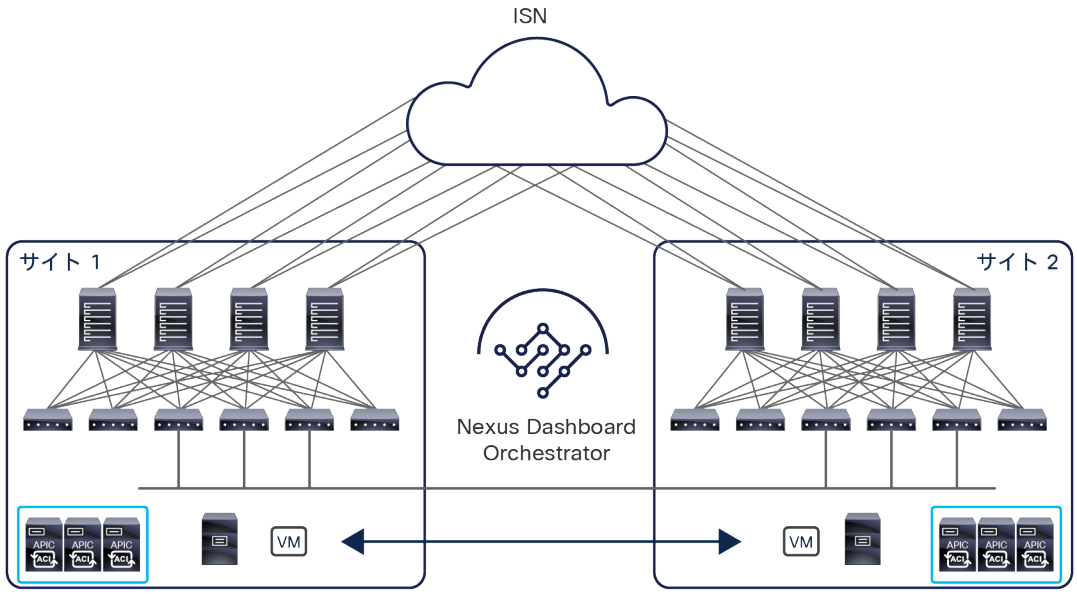


図 54.
フラッディングを使用したレイヤ 2 のサイト間接続

特定の要件がある場合に BUM トラフィックのフラッディングが必要になります。その一例がアプリケーション クラスタリングです（この従来型のクラスタリングでは、異なるアプリケーション クラスタ ノードの間でレイヤ 2 マルチキャスト通信が必要です）。

図 55 は、このユースケースの論理構成図を示しています。サイトにまたがる BUM フラッディングが有効になっている点を除いて、以前に図 49 で示したケースとほぼ同じです。

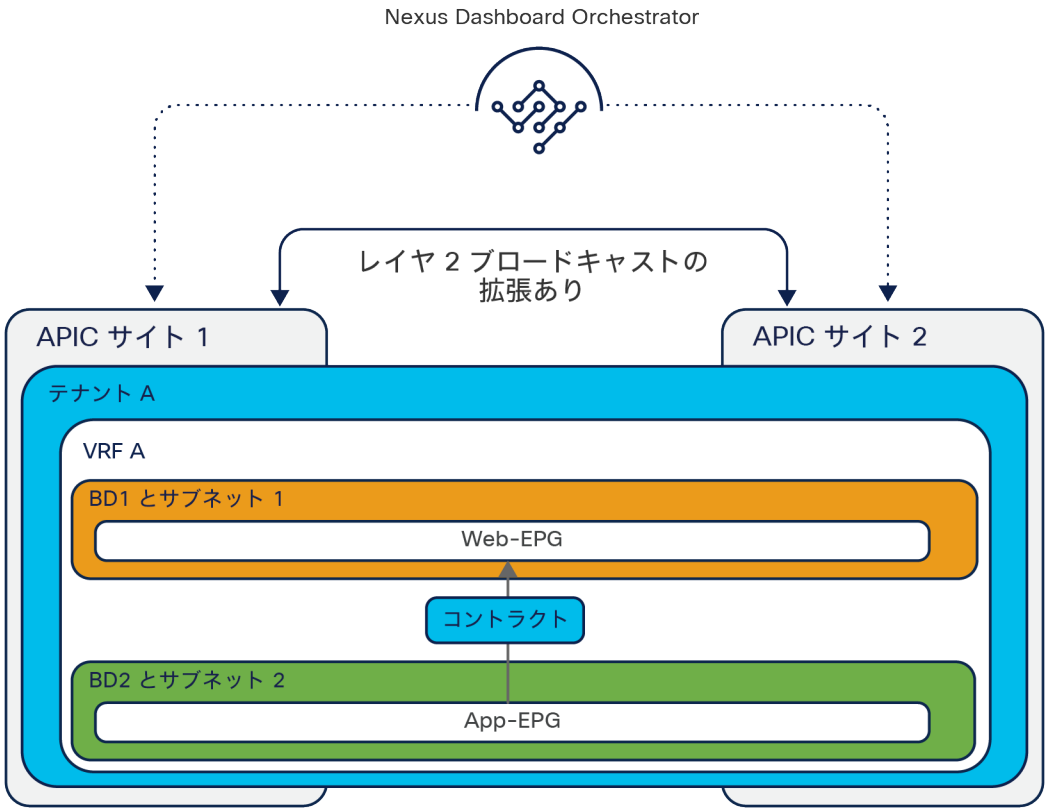


図 55.
フラッディングを使用したレイヤ 2 のサイト間接続（論理構成図）

Cisco ACI の独立したネットワーク間で完全なレイヤ 2 フラッディングを必要とする展開の場合、通常は Cisco ACI マルチポッドがアーキテクチャの選択肢となります。しかし、この構成のマルチサイト設計から得られるメリットの 1 つは、どのブリッジドメインを拡張し（BUM 転送を有効化するかどうかにかかわらず）、どのドメインをローカルに保持するかを厳密に制御できることです。実際、Cisco Nexus Dashboard Orchestrator を使用すると、ブリッジドメインレベルでフラッディングの動作を差別化できます。実際のシナリオでは、一部のストレッチブリッジドメインでのみフラッディングの動作のサポートが必要となることがあります。差別化はこのような場合に役立ちます。

サイト間ネットワーク（ISN）の展開に関する考慮事項

前述のように、一般的にサイト間ネットワーク（ISN）と呼ばれる汎用的なレイヤ 3 インフラストラクチャが、異なる APIC ドメインを相互接続しています。図 56 は、ISN を通して確立されたサイト間 VXLAN トンネルを示しています。

注： このドキュメントの残りの部分では、「ISN」、「IP WAN」、および「IP ネットワーク」という用語を区別なく使用します（図の中でも同様です）。いずれも、同じマルチサイトドメインに属する ACI ファブリック間のネットワーク接続に使用されるルーテッド ネットワーク インフラストラクチャを意味します。

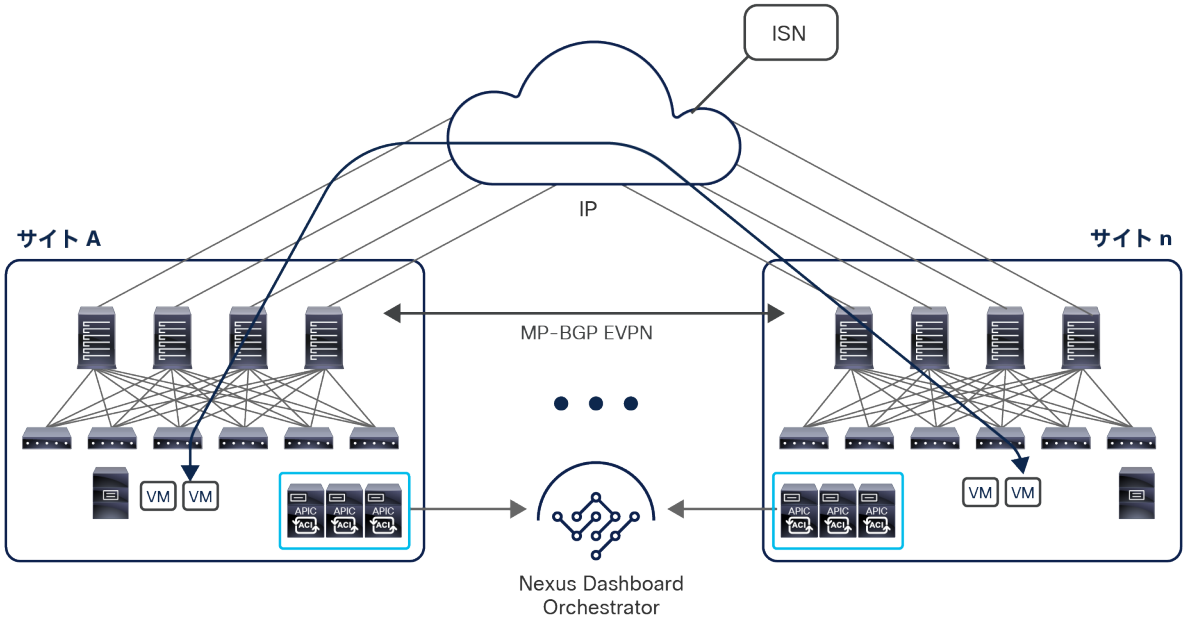


図 56.
サイト間ネットワーク（ISN）

ISN では、プレーンな IP ルーティングがサポートされていれば、サイト間 VXLAN トンネルを確立できます。要件がこれだけのため、単一のシンプルなルータデバイス（冗長性のために 2 つ使用することが常に推奨されます）から、世界中に広がる複雑なネットワーク インフラストラクチャまで、好みの方法で ISN を構築できます。

スパインインターフェイスは、ポイントツーポイントのルーテッドインターフェイスを介して ISN デバイスに接続されます。ただし、スパインインターフェイスから発信されるトラフィックには常に 802.1q VLAN 4 の値がタグ付けされます。そのため、スパインとそれに直接接続された IPN デバイスの両方でレイヤ 3 サブインターフェイスを定義してサポートする必要があります。したがって、同じ VLAN 4 タグを使用しながら別々のポイントツーポイント L3 リンクとして機能する複数のサブインターフェイスを同じデバイス上に定義できる IPN ルータを選択することが重要です。

注： ISN デバイスでサブインターフェイスを使用する必要があるのは、スパインに接続する場合のみです。

ISN にマルチサイト専用のネットワーク インフラストラクチャを使用することも可能ですが、他の接続サービスを併せて提供しているネットワークをこの目的に使用することも、多くの実際のシナリオではきわめて一般的です。この後者の場合、専用の VRF ルーテッドドメインを展開して、マルチサイトのコントロールプレーンとデータプレーンのトラフィックを転送することを強くお勧めします。これを推奨する理由は以下の 2 つです。

- 専用のルーティングドメインを用いると、サイト間接続の問題のトラブルシューティングが必要になったときに運用上のメリットが得られます（ルーティングテーブルが小さくなるなど）。
- 不要なルーティングプレフィックスで ACI「overlay-1」VRF が過負荷になるのを防ぐ必要があります。前述のように、サイト間接続を確立するには同じマルチサイトドメインに属するファブリック間でプレフィックスを交換する必要がありますが、その数はごくわずかです。実際、ACI「overlay-1」ルーティングスペースにはルート数の上限があり、現在のところ、上限は 1,000 プレフィックスです。この基盤となるルーティングドメインに注入されるプレフィックスが増えると、ACI スパインノードの安定性と機能が損なわれる可能性があります。

ISN では、デフォルトの 1500 バイト値を超える MTU のサポートも必要です。VXLAN データプレーントラフィックでは 50 バイトのオーバーヘッドが追加されます（元のフレームの IEEE 802.1q ヘッダーを保持する場合は 54 バイト）。したがって、ISN インフラストラクチャのすべてのレイヤ 3 インターフェイスがこの MTU サイズの増加をサポートできることを確認する必要があります（一般的な推奨事項は、ISN デバイスのすべてのインターフェイスで、エンドツーエンドで必要な MTU の下限より少なくとも 100 バイト多い MTU をサポートすることです）。これは、ACI スパインノードが、フラグメント化された VXLAN フレームをハードウェア内で再構築できないためです。たとえば、送信元サイト 1 のスパインが MTU の大きなパケットを ISN に送信し、ISN デバイスがそのインターフェイスの 1 つからそのパケットを送信する前にフラグメント化せざるを得なかったとします（サポートされる MTU が小さいため）。この場合、宛先サイト 2 でそれを受信したスパインはフラグメントを再構築できず、応答することなくフラグメントを破棄するため、サイト間接続が切断されます。

注： フレームが ISN インフラストラクチャ内でフラグメント化され再構築された場合（たとえば、送信元と宛先の ISN デバイスが相互に IPsec 接続を確立することによって）、宛先スパインがフルサイズのフラグメント化されていないフレームを処理するため、サイト間の正常な接続が維持されます。ただし、このような回避策を採用する場合は、スループットの観点からパフォーマンスへの影響を考慮することが重要です。これは主に、フラグメンテーションと再構築を行う ISN プラットフォームに依存します。

ISN で設定する MTU の最小値は、以下の 2 つの要因によって決まります。

- ファブリックに接続されたエンドポイントによって生成されるフレームの MTU の最大値：エンドポイントがジャンボ フレーム（9,000 バイト）をサポートするように構成されている場合、ISN は少なくとも 9,050 バイトの MTU 値になるよう構成する必要があります。エンドポイントがデフォルトの 1,500 バイト値になるよう構成されている場合は、ISN の MTU サイズを 1,550 バイトに減らすことができます。これは、他のカプセル化のオーバーヘッドが各フレームに追加される場合を除きます。たとえば、IPsec または CloudSec で暗号化する場合が該当します。
- 異なるサイトにあるスパインノード間の MP-BGP コントロールプレーン通信の MTU：デフォルトでは、エンドポイントルーティング情報を交換するためにスパインノードが 9,000 バイトのパケットを生成します。そのデフォルト値が変更されていない場合、ISN は少なくとも 9,000 バイトの MTU サイズをサポートする必要があります。そうでない場合、サイト間でのコントロールプレーン情報の MP-BGP 交換は成功しません（MP-BGP 隣接関係を確立できるにもかかわらず）。デフォルト値は、APIC ドメインごとに、対応するシステム設定を変更することで調整できます（図 57）。

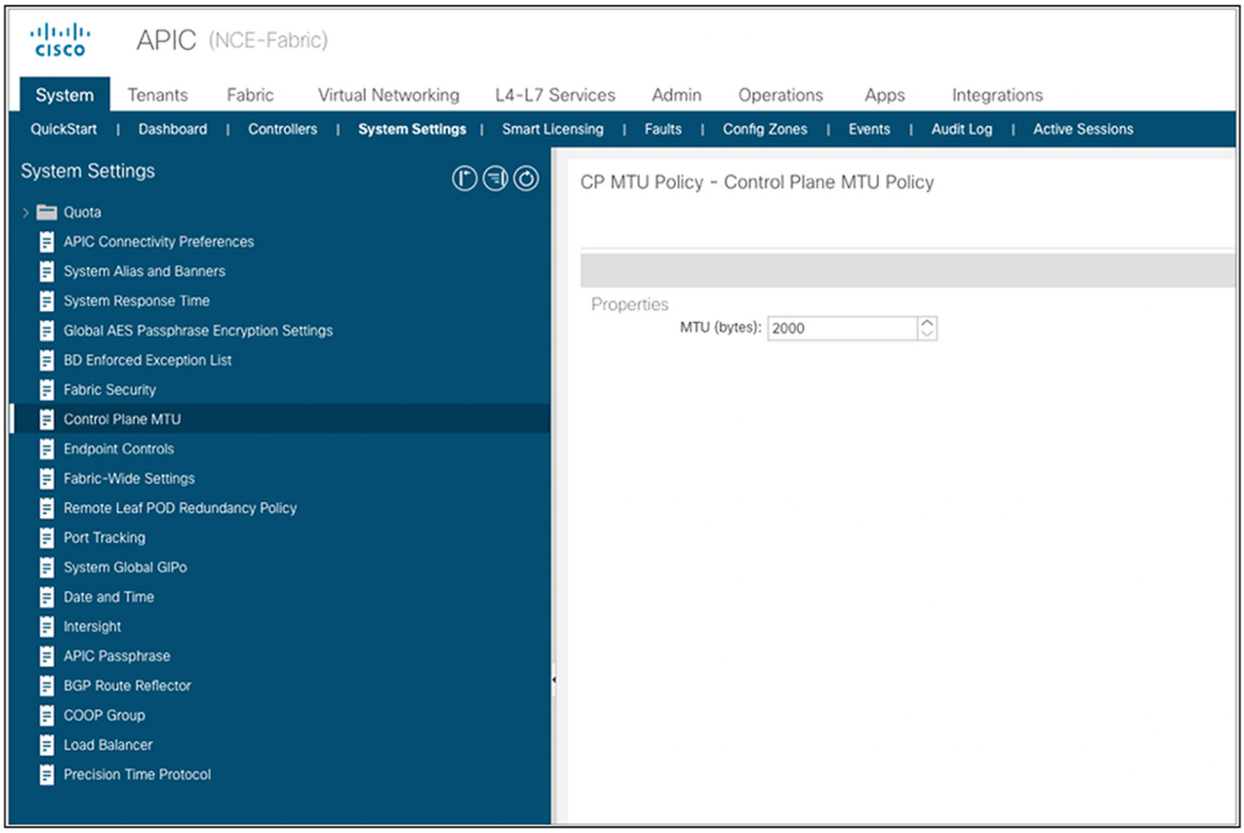


図 57.
Cisco APIC でのコントロールプレーン MTU の設定

注： 上の図に示されているグローバル MTU 設定は、ACI ファブリック内で使用されるコントロール プレーン プロトコル（IS-IS、COOP、MP-BGP VPNv4）、およびスパインと ISN デバイスの間で使用される OSPF プロトコルにも適用されます。ただし、これらのコントロール プレーン フレームの最大サイズを減らした場合でも、パフォーマンスに対する大きな影響は確認されていません。

ISN と QoS の展開に関する考慮事項

展開シナリオによっては、サイト間の異なるトラフィックフローで QoS の動作を差別化することが望ましい場合があります。マルチサイト設計でこの観点から提供されている機能を理解するには、ACI ファブリック内での QoS の動作について手短に振り返ることが役立ちます。

CoS 値	DEI ビット	サービスクラス
2	0	レベル 1 ユーザーデータ
1	0	レベル 2 ユーザーデータ
0	0	レベル 3 ユーザーデータ
2	1	レベル 4 ユーザーデータ
3	1	レベル 5 ユーザーデータ
5	1	レベル 6 ユーザーデータ
3	0	APIC コントローラ トラフィック
4	0	SPAN トラフィック
5	0	コントロール プレーン トラフィック
6	0	traceroute
7	0/1	予約済み

図 58.
ACI ファブリックの QoS クラス

図 58 に示すように、ACI ファブリックは、ファブリック内のトラフィックを処理するために複数のサービスクラスをサポートします。

- ユーザーデータ（テナントなど）トラフィックのためのユーザーが構成可能な 6 つのサービスクラス：外部コネクテッドデバイス（エンドポイント、ルータ、サービスノードなど）からリーフノードが受信するすべてのユーザートラフィックに使用されます。デフォルトでは、これらのデバイスから受信したトラフィックはすべてレベル 3 クラスに割り当てられます。この割り当ては、外部デバイスが属する EPG（または L3Out の場合は Ext-EPG）、または EPG 間のコントラクトによって確立された関係に基づいて変更できます。

  ファブリック管理者は、（ファブリック アクセス ポリシー構成の一部として）これらのユーザーレベルの QoS クラスのそれぞれに関連付けられた、最小バッファ、輻輳アルゴリズム、帯域幅割り当て率などのプロパティを調整できます。

  注：Cisco ACI リリース 4.0(1) までは、3 つのユーザーデータクラスのみが使用できます。
- 予約済みの 4 つのサービスクラス：APIC コントローラノード間のトラフィック、リーフノードとスパインノードによって生成されるコントロール プレーン トラフィック、SPAN と traceroute のトラフィックに使用されます。

ACI ファブリック内のさまざまなトラフィックフローは、VXLAN カプセル化パケットの 802.1Q ヘッダーに含まれる CoS と DEI のビットに割り当てられた値に基づいて、異なる QoS クラスに区分されます。

VXLAN カプセル化トラフィックが同じマルチサイトドメインに属するファブリック間の通信のためにサイト間ネットワークに送られる場合、この動作に問題が生じます。そもそも 802.1Q ヘッダーが ISN 内で存在するか、あったとしてエンドツーエンドでこれらのビットが保存されるかが保証できないためです。したがって、サイト間のトラフィックフローをそれらが属すると想定される QoS クラスに基づいて差別化できるようにするには、別のメカニズムが必要です。

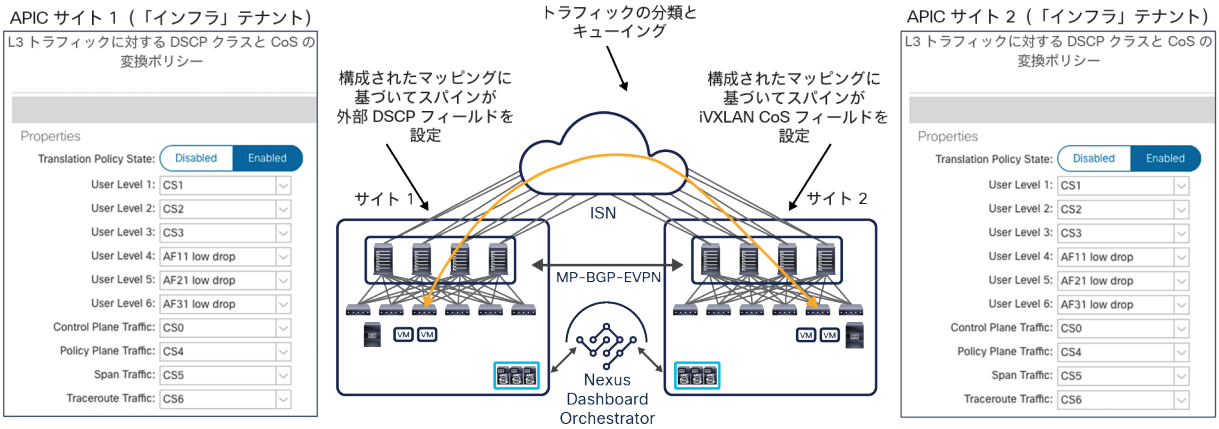


図 59.
ACI サイト間のエンドツーエンドで整合性のとれた QoS の動作

図 59 は、マルチサイトアーキテクチャでこの目的を達成する方法を示しています。それぞれの APIC ドメインにマッピングテーブルが構成し、各 QoS クラスを Differentiated Services Code Point（DSCP）の特定の値に一貫性がとれるようにマッピングします。この値は、ISN に注入される VXLAN カプセル化パケットの外部 IP ヘッダーに設定されます。この機能により、以下の 2 つの目的を達成できます。

- パケットで伝送される DSCP 値に基づいて、リモート ACI ファブリックが受信したトラフィックを適切な QoS クラスに関連付けることができます。これにより、同じマルチサイトドメインに属する異なる ACI ファブリックで、一貫した QoS の処理を実現できます。
- ファブリックを相互接続する ISN インフラストラクチャで、トラフィックの分類と優先順位付けができます。たとえば、異なるサイトのスパインノード間の MP-BGP コントロール プレーン トラフィックに常に優先順位を付けて、データトラフィックがマルチサイト展開の全体的な安定性を損なわないようにすることができます。これは基本的かつ一般的な推奨事項です。6 つのユーザーデータクラスを使用すると、同じテナントまたは異なるテナントに属するトラフィックフローに提供されるサービスを差別化することもできます。

  注：さまざまなタイプのトラフィックを分類して優先順位を付けるためには ISN デバイスでの構成が必要です。この構成は展開されている特定のハードウェア プラットフォームに依存するため、このホワイトペーパーの範囲外です。それぞれの製品の添付資料を参照してください。

上記の 2 つの目的を達成するための基本的な前提は、ISN 内のデバイスによって DSCP 値が変更されないことです。変更されると、トラフィックがリモートサイトで受信されたときにトラフィックを適切な QoS クラスに関連付けることができなくなるためです。

また、QoS グループと、対応する DSCP 値の間のマッピングは、NDO が直接実行する必要があります。そうすることで、マルチサイトドメインのすべてのサイトに、一貫したマッピングが展開されます。この設定はすべての NDO リリースでサポートされています。また、Cisco Nexus Dashboard Orchestrator リリース 4.0(1) 以降では、ファブリック ポリシー テンプレートの一部になっています。

Cisco ACI マルチサイトのアンダーレイ コントロール プレーン

ISN デバイスは、ルーティングプロトコルを使用して、各サイトに展開されたスパインノードとの隣接関係を確立し、サイト間のコントロールプレーンとデータプレーンを確立するために必要なプレフィックスをサイト間で交換できるようにする必要があります（図 60）。

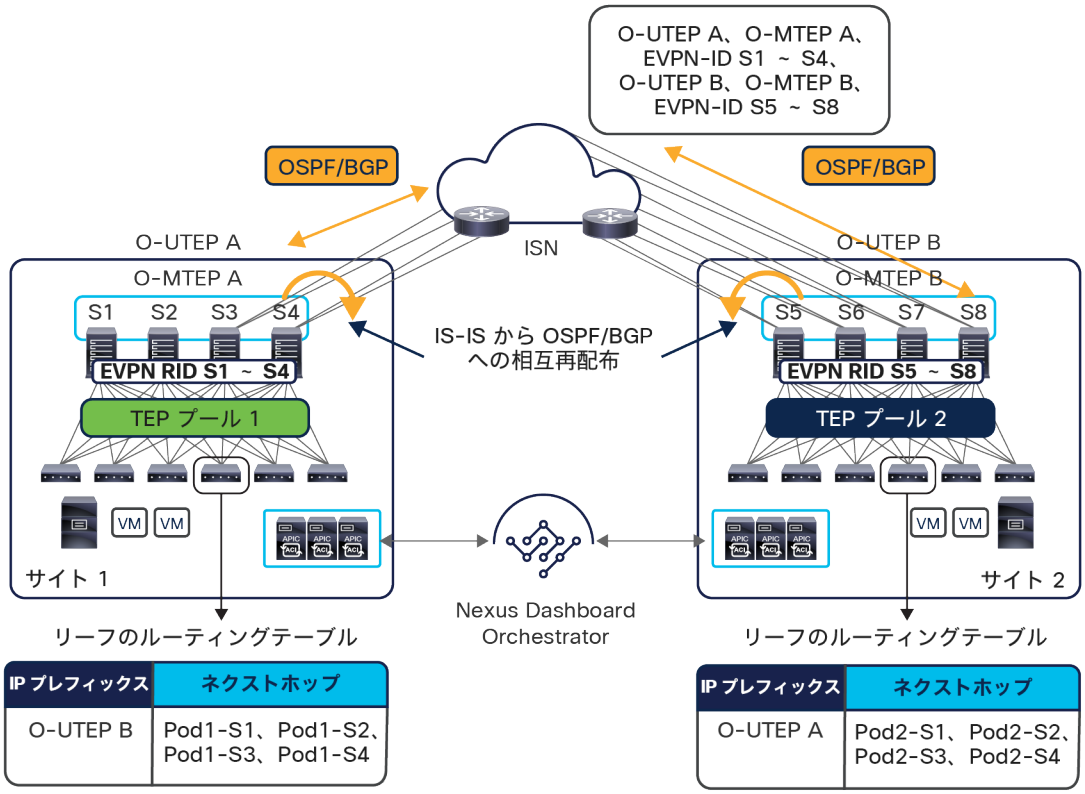


図 60.
各サイトのスパインノードと ISN の間の OSPF/BGP ピアリング

当初は OSPF が、サポートされる唯一のコントロールプレーンでしたが、NDO リリース 3.5(1) 以降は、BGP も有効なオプションとして追加されました。

注： BGP を使用するには、Cisco ACI リリース 5.2(1) の展開も必要です。

OSPF と BGP は必要に応じて併用でき、スパインノードで定義された以下の IP アドレスのルーティング情報をサイト間で交換するために使用されます。

- BGP-EVPN ルータ ID（EVPN-RID）：この一意の IP アドレスは、ファブリックに属する各スパインノードで定義され、リモートサイトのスパインノードと MP-BGP EVPN および VPNv4 の隣接関係を確立するために使用されます。
- オーバーレイユニキャスト TEP（O-UTEP）：この共通エニーキャストアドレスは、同じポッド内のすべてのスパインノードによって共有され、ユニキャスト VXLAN データプレーントラフィックの送受信に使用されます。各ポッドは、基本的に（単一ポッドファブリックが展開された場合）サイトを一意に識別する O-UTEP アドレスによって特徴付けられます。

  注：マルチポッドファブリックを展開すると、各ポッドに一意の O-UTEP アドレスが割り当てられます。詳細は、「Cisco ACI のマルチポッドとマルチサイトの統合」セクションを参照してください。

- オーバーレイマルチキャスト TEP（O-MTEP）：この共通エニーキャストアドレスは、同じサイト内のすべてのスパインノードによって共有され、BUM トラフィックのヘッドエンド レプリケーションの実行に使用されます。BUM トラフィックは、ローカルスパインノードで定義された O-UTEP アドレスから発信され、所定のブリッジドメインが拡張されているリモートサイトの O-MTEP に配信されます。

O-UTEP と O-MTEP を定義することで、特定のサイトに接続されているすべてのリソースが、この 2 つの IP アドレスのいずれかを介してリモートサイトから到達可能であると常に見なされるようになります（どちらのアドレスを使うかは、ユニキャストトラフィックか BUM/マルチキャスト通信かによって異なります）。その結果、2 つのサイト間で交換される ISN 内の水平方向 VXLAN トラフィックでは、すべてのパケットが常に送信元サイトを識別する O-UTEP アドレスから送信され、宛先サイトの O-UTEP（または O-MTEP）に向かうようになります。これは、複数の ECMP リンクを活用して 2 つのサイトを相互接続することを妨げるものではないことに注意してください。VXLAN カプセル化パケットを構築するとき、エンドポイントが生成した元のフレームの L2/L3/L4 ヘッダーをハッシュした結果が外側のヘッダーの UDP 送信元ポートとして使用されます。このようにしてパケットに「エントロピー」が形成されます。同じサイト間であっても（およびエンドポイントの同じペア間であっても）アプリケーションフローが異なれば、生成される UDP 送信元ポートの値が異なります。すなわち、ISN 内のネットワークデバイスが L4 ポート情報を見てトラフィックを転送するリンクを選択する限り、複数のパス間で VXLAN パケットの負荷が分散されます。

図 60 に示すように、EVPN-RID、O-UTEP、O-MTEP の 3 つのアドレスが、サイト間の EVPN コントロールプレーンと VXLAN データプレーンを有効化するためにサイト間で交換する必要があるプレフィックスのすべてです。したがって、ISN ルーティングドメインで学習する必要があるのは、この 3 つのプレフィックスのみです。つまり、この 3 つの IP アドレスは ISN 全体でグローバルにルーティング可能になっている必要がありますが、通常これは問題にならないはずです。この 3 つは各ファブリックに関連付けられた元の TEP プールから独立し、マルチサイトを展開したときに Cisco Nexus Dashboard Orchestrator で個別に割り当てられているためです。

専用の IP の範囲を設けてそこからこの少数の IP アドレスを割り当て、サイトにまたがってこの /32 プレフィックスをすべてアドバタイズできるようにすることが、ベストプラクティスの推奨事項となります。必要に応じて、1 つのサイトで使用されるすべての /32 プレフィックスを集約し、集約ルートのみをマルチサイトドメインに属するリモートファブリックに送信することもできます。その場合、受信した集約ルートをファブリック内部の IS-IS コントロールプレーンに再配布できるようにするには、すべてのリモート APIC ドメインで追加の構成手順が必要になります。

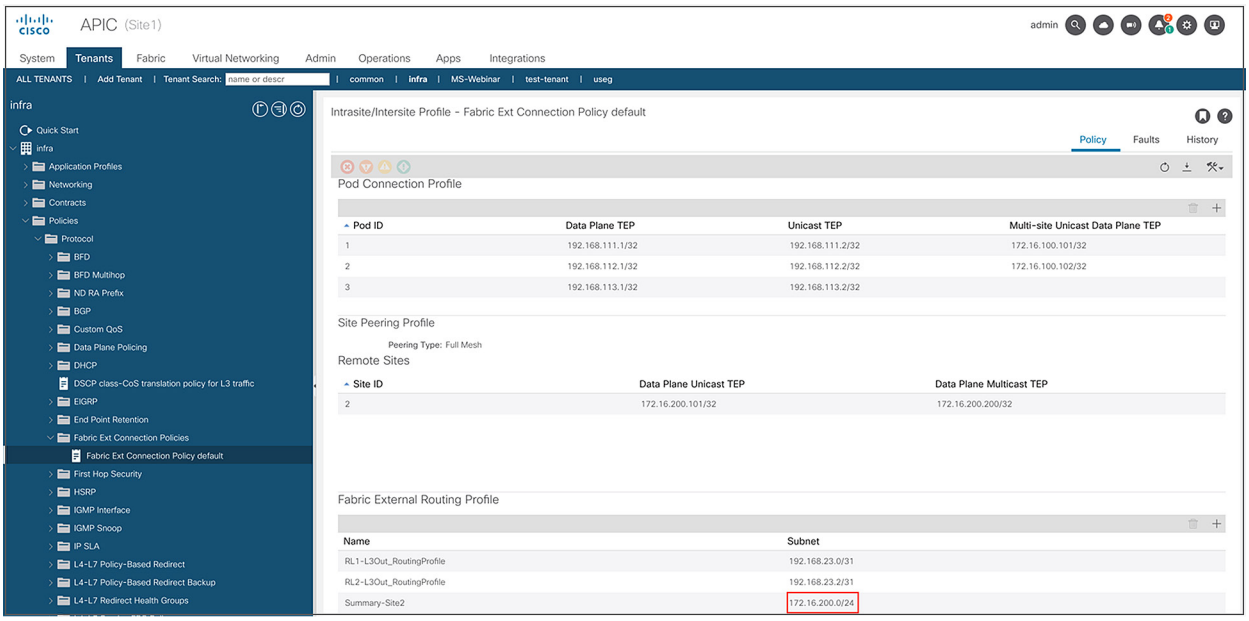


図 61.
サイト 1 の APIC で定義されたリモートサイト 2 の集約ルート

図 61 に示すように、サマリープレフィックスは、[インフラ（infra）] テナントに対して定義された [ファブリック外部ルーティングプロファイル（Fabric External Routing Profile）] ポリシーの一部として APIC で設定する必要があります。

注： 各サイト内で使用され、ファブリックの起動時に割り当てられる内部 TEP プールのプレフィックスは、サイト間通信を可能にするためにサイト間で交換する必要はありません。したがって、これらのプールに対する割り当て方に技術的な制約はなく、重複する内部 TEP プールを使用する ACI ファブリックが同じマルチサイトドメインに属している場合があります。ただし、内部 TEP プールのサマリープレフィックスは常にスパインから ISN に向けて送信されます。これは、Cisco ACI のマルチポッドとマルチサイトのアーキテクチャを統合する場合に必要になるためです。したがって、ベストプラクティスとしては、内部 TEP プールのプレフィックスが最初の ISN デバイスでフィルタリングされ、ISN ネットワークに注入されないようにすることが必要です（ネットワークのバックボーンまたはリモートファブリックにすでに展開されているアドレス空間と重複する可能性があるため）。詳細は、「Cisco ACI のマルチポッドとマルチサイトの統合」セクションを参照してください。

フラッディングが有効なストレッチブリッジドメインのユースケースでは、ポッド間でレイヤ 2 マルチ宛先（BUM）トラフィックを交換できるようにする場合でも、ISN ルーティングドメイン内でマルチキャストをサポートする必要がないことに注意してください。Cisco ACI マルチサイト設計では、送信元サイトのスパインノードの入力レプリケーション機能によって、そのブリッジドメインが拡張されているすべてのリモートサイトに BUM トラフィックが複製されるためです。この機能の詳細は、「サイトにまたがるレイヤ 2 BUM トラフィックの処理」セクションを参照してください。

Cisco ACI マルチサイトスパインのバックツーバック接続

Cisco ACI リリース 3.2(1) 以降、図 62 に示すように、別々のファブリックに属するスパイン間のバックツーバックトポロジもサポートされます。

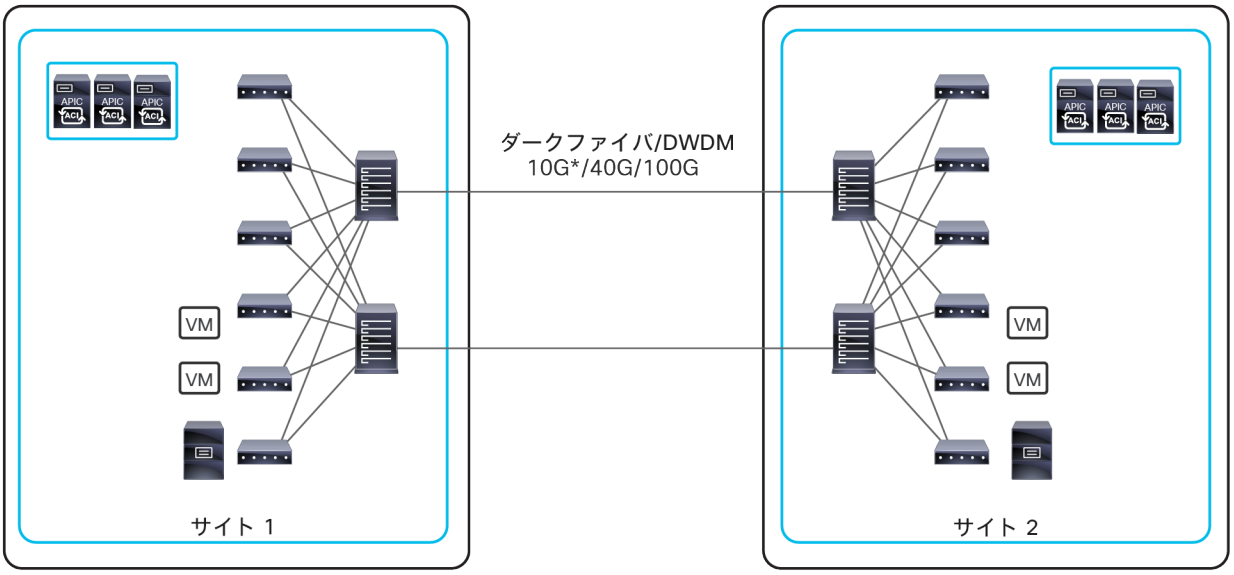


図 62.
Cisco ACI マルチサイトスパインのバックツーバック接続（Cisco ACI リリース 3.2 以降）

このトポロジの設計に関する重要な考慮事項は以下のとおりです。

- あるサイトに展開されたスパインすべてがリモートサイトのスパインに接続している必要はありません。スパインの一部のみがサイトにまたがってバックツーバックで接続される場合もあります。
- 使用するスパインやリンクの数は、主にそれらの専用リンクを展開するためのコストと、接続に求められる帯域幅や復元力とのバランスを考慮して決定されます。
- 別々の DC サイト間のすべての通信を暗号化する必要がある場合は、MACsec 暗号化を有効にできます。暗号化をサポートするためのハードウェア要件の詳細は、「Cisco ACI マルチサイトとサイト間トラフィックの暗号化（CloudSec）」セクションを参照してください。
- 現在、バックツーバックで接続できるサイト（ファブリック）は 2 つに制限されています。3 つ以上のサイトがダイレクトリンクで接続されるトポロジはサポートされていません。これは、あるサイトのスパインがそのサイトとは異なるサイトのペア間で行われる VXLAN カプセル化通信を転送できないためです。同じ理由で、2 つのサイトがバックツーバックで接続され、他のリモートサイトとは汎用のサイト間ネットワーク経由で接続される「ハイブリッド」トポロジの展開はサポートされていません（図 63）。

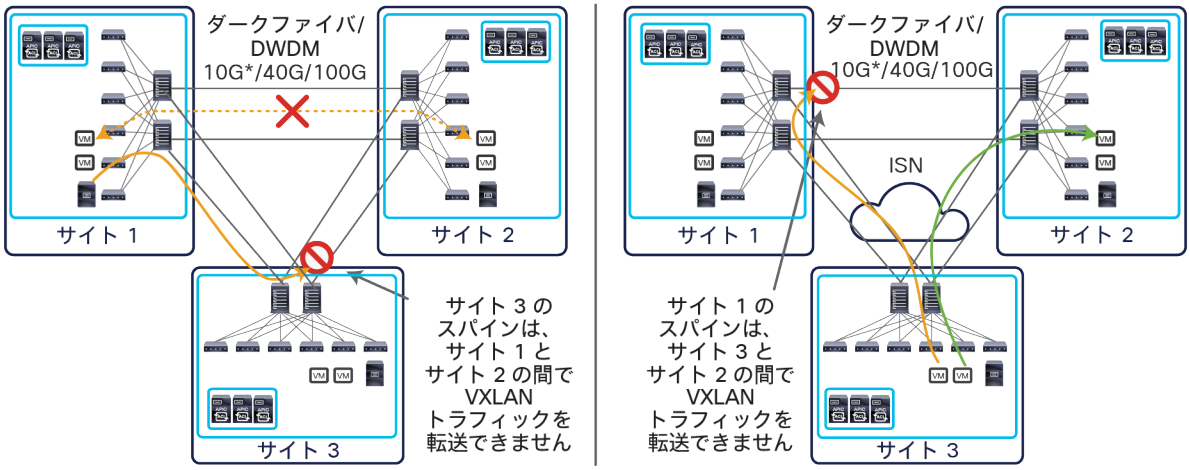


図 63.
サポートされていないバックツーバックトポロジ

上図の左側のシナリオは、サイト 1 とサイト 2 の間のダイレクトリンクに障害が発生した場合に、VXLAN トラフィックがサイト 3 のスパイン経由にステアリングされる様子を示しています。右側の「ハイブリッド」シナリオは、トラフィックが ISN を介してサイト 2 のスパインに直接配信されるよう適切にステアリングされ、サイト 3 とサイト 2 の間で通信が成功している様子を示しています。ただし、同じフローがサイト 1 のスパインを経由するようステアリングされた場合（たとえば、ISN での再ルーティングが原因で）、サイト 3 とサイト 2 の間の通信は失敗します。

サイトのペア間でより高い帯域幅の接続が利用できるために「ハイブリッド」シナリオを採用する場合、図 64 に示すように、この接続を使用してサイト間ネットワークのファーストホップのデバイス間でダイレクトリンクを確立することをお勧めします（スパインノード間ではダイレクトリンクを使用しません）。

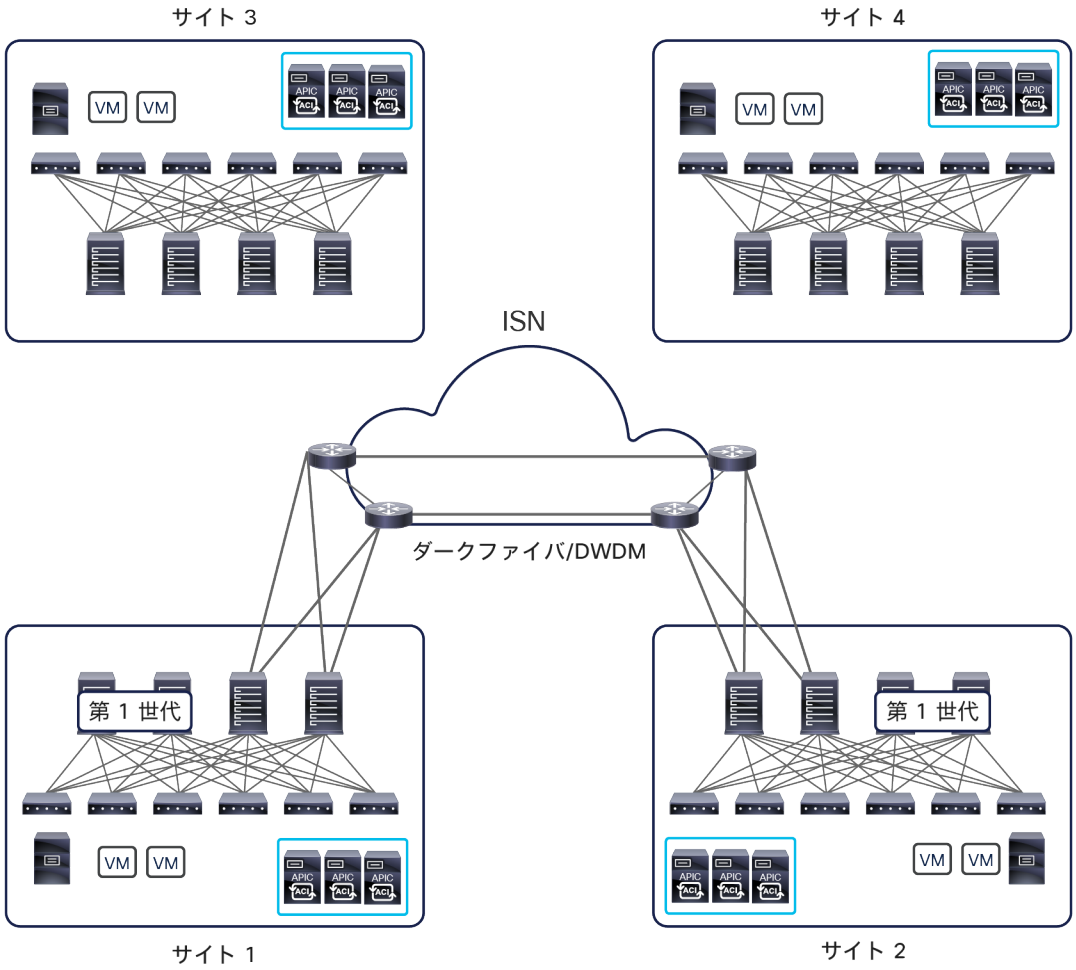


図 64.
ファーストホップ ISN ルータのバックツーバック接続

- 上の図 62 に示すように、スパインはスクエアトポロジで接続できます。このトポロジでも、スパインはリモートサイトのスパインとフルメッシュの BGP ピアリングを確立しますが（ルートリフレクタは構成されていないと仮定します）、各スパインから発信されるデータプレーントラフィックがリモートスパインに接続する単一のリンクを流れることは明らかです。サイト間のすべての BUM トラフィックは、常に、そのブリッジドメインの指定フォワーダとして選択されたスパインによって発信され、O-MTEP アドレス宛てに送信されます。これを受信したスパインは、ローカルサイト内で BUM トラフィックを転送します（受信サイトでこの機能を担う指定フォワーダを選択する必要はありません）。
- 上記の説明にもかかわらず、トポロジのベストプラクティスは、別々のサイトに属するスパイン間にフルメッシュ接続を作成することです。このトポロジを採用することで、図 65 に示すリモートスパインで障害が発生した場合のような障害シナリオにおいて、直接的なメリットが 2 つ得られます。まず、サイト間のレイヤ 2/レイヤ 3 ユニキャスト通信の場合、それぞれのローカルスパインで利用できる残された接続にローカルで VXLAN トラフィックを振り分けるだけでトラフィックを回復できます。次に、サイト間のレイヤ 2 BUM トラフィックにも同じことが言えます。この場合、ローカルの指定フォワーダを選択し直す必要はありません（別のリモートスパインの少なくとも 1 つと接続された状態にあるため）。

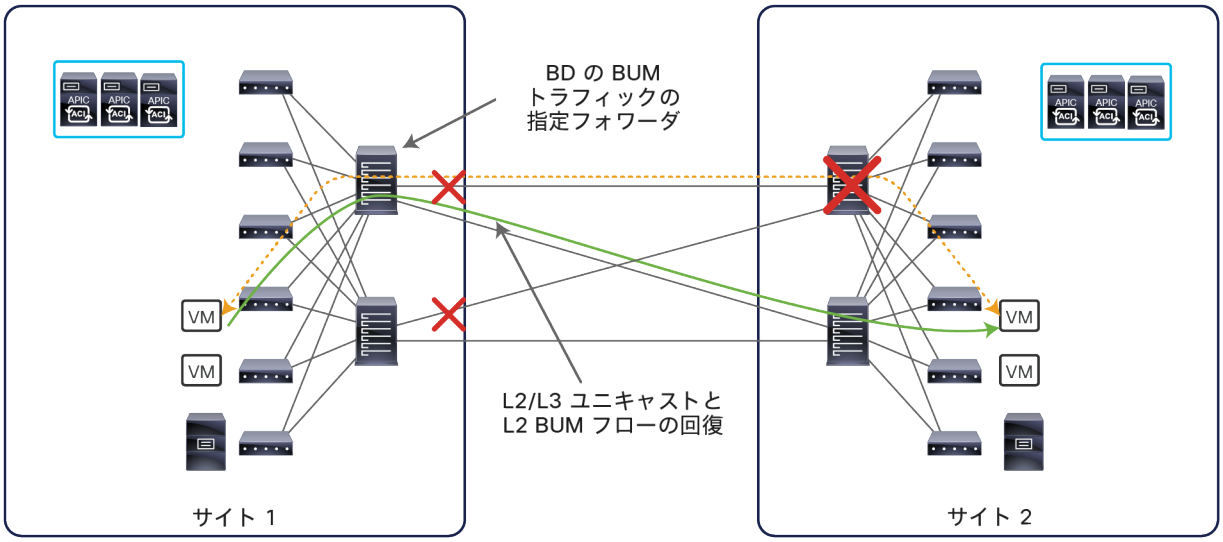


図 65.
リモートスパインの障害シナリオでのフルメッシュ接続によるトラフィックの回復

注： 上図に示したトラフィックの回復は、リモートスパインで障害が発生するまで、すべての L2/L3 ユニキャストトラフィックと BUM トラフィックが点線に沿って流れていた場合を想定しています。

- 別々のサイトに展開されたスパインは、（論理的または物理的に）直接接続する必要があります。つまり、スパイン間にレイヤ 2 インフラストラクチャを展開できず、ダークファイバまたは DWDM 回線で接続する必要があります。
- EoMPLS 疑似回線も使用できます。別々のサイトにあるスパイン間のポイントツーポイント接続を MPLS コアネットワークにまたがって論理的に拡張できるためです。

Cisco ACI マルチサイトとサイト間トラフィックの暗号化（CloudSec）

異なるデータセンターにまたがってアプリケーション（またはアプリケーション コンポーネント）を展開する場合、サイト間で確立された通信のプライバシーと機密性を確保するために、データセンターのロケーションから送信されるトラフィックをすべて暗号化する必要に迫られることがよくあります。

従来の方法を採用する場合は、データパスにアドホック暗号化デバイスを展開するか、IPsec や MACsec などのネットワークベースのソリューションを有効にすることで、これが達成できます。どのシナリオを採用しても、データセンター間の通信を保護するために、必要な機能を備えた追加のハードウェアを展開する必要があります。

Cisco ACI リリース 4.0(1) では、「CloudSec」と呼ばれるセキュリティソリューションがサポートされています。CloudSec を一種の「マルチホップ MACsec」と考えると、その機能が簡単に理解できます。CloudSec を用いると、汎用的なレイヤ 3 ネットワークを介して接続された 2 つの VTEP デバイス間の通信を暗号化できます。

CloudSec を Cisco ACI マルチサイトアーキテクチャのコンテキストに挿入すると、ローカルスパインを通ってローカルサイトを出るトラフィックおよびローカルスパインからリモートスパインに入るトラフィックをすべて暗号化できます（図 66）。

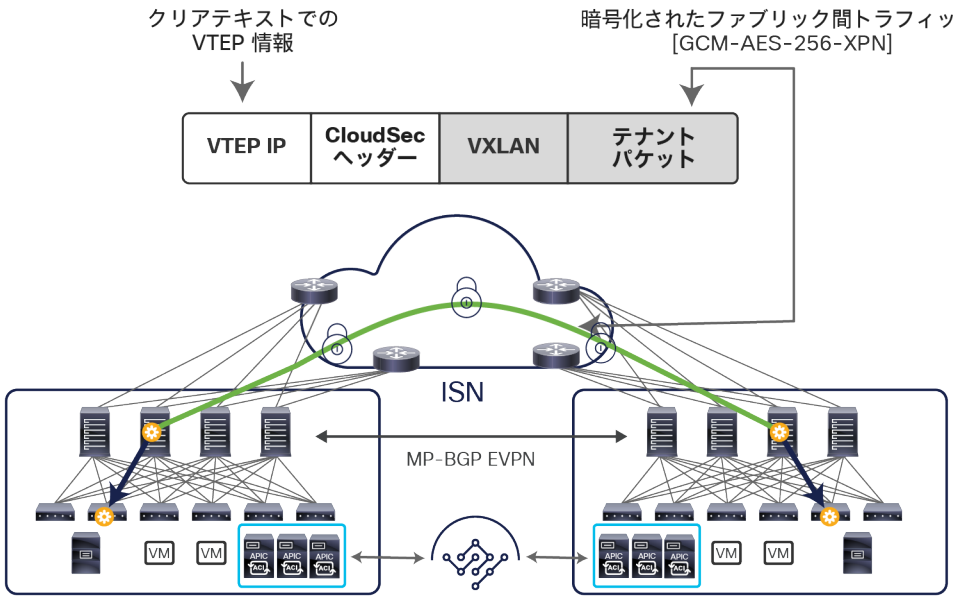


図 66.
CloudSec を使用したサイト間通信の暗号化

CloudSec を使用すると、VXLAN ヘッダーを含む元のパケットを暗号化できます。暗号化によってサイト間で送信される各パケットの全体的な MTU サイズが、VXLAN カプセル化による 50 バイトに加えて、CloudSec ヘッダー用にさらに 40 バイト増加します。そのため、ISN の MTU 設定では、この増加も考慮する必要があります。

重要：CloudSec は、Nexus 9000 サイト間ネットワーク（ISN）インフラストラクチャを使用して内部的に検証されています。ISN インフラストラクチャがさまざまなデバイスで構成されている場合、またはデバイスが不明な場合（回線をサービスプロバイダーから購入した場合など）、各ファブリックのスパインとそれらの ISN デバイスとの間に Cisco 1000 シリーズ アグリゲーション サービス ルータ（ASR）を挿入する必要があります。パディングフィックスアップが有効になっている 1000 シリーズ ASR ルータにより、CloudSec トラフィックがサイト間の任意の IP ネットワークを通過できるようになります。

事前共有キーは最初にサポートされた唯一のモデルであり、これを用いることでサイトにまたがって展開されたスパインがトラフィックの暗号化と復号を適切にできるようになります。

この機能は、スパインノードに以下のハードウェアモデルを使用すればラインレートで（つまり、パフォーマンスに影響を与えることなく）実行されます（このホワイトペーパーの執筆時点では、CloudSec をサポートするモデルはこれがすべてです）。

- 9736C-FX シリーズ ラインカードを装着した Cisco Nexus 9500 モジュラスイッチのポート 29 ~ 36
- Cisco Nexus 9364C 非モジュラスイッチのポート 49 ~ 64
- Cisco Nexus 9332C 非モジュラスイッチのポート 25 ~ 32

CloudSec 機能は、スパイン HW モデルとして上に列挙した暗号化対応インターフェイスのいずれかでのみ有効化できます。したがって、サイトにまたがって CloudSec 暗号化が必要な場合は、これらのインターフェイスを使用してスパインを ISN に接続する必要があります。

現在の実装では、暗号スイート GCM-AES-256-XPN（256 ビットキーを使用）がデフォルトとして CloudSec で使用されています。このオプションの構成は不要です。Cisco Nexus 9000 ハードウェアでサポートされている最も自動化された安全なオプションです。

注： Cisco ACI リリース 5.1(1) より前は、サイト間の CloudSec 暗号化は、外部（ルーティング可能）TEP プールと併用できません。この TEP プールは、サイト間 L3Out 機能を有効にするとき、またはリモートリーフノードをマルチサイトドメインに属する ACI ファブリックに接続するときに必要となります。Cisco ACI リリース 5.1(1) 以降はこの制限がなくなったため、外部 TEP プールを展開し、同時に CloudSec 暗号化をオンにすることができます。ただし、別々の ACI ファブリックに接続されている内部エンドポイント間の通信のみが暗号化されます。ファブリック 1 に接続されたエンドポイントがリモートサイトに展開された L3Out 接続を介してこのファブリックの外部のリソースと通信するトラフィック（サイト間 L3Out 機能）の暗号化、または異なるサイトで定義された L3Out に接続されたリソース間のトラフィック（サイト間トランジットルーティング）の暗号化は、Cisco ACI リリース 5.2(4) 以降でサポートされます。

Cisco ACI マルチサイト展開で CloudSec 暗号化を有効にする方法の詳細は、https://www.cisco.com/c/ja_jp/td/docs/dcn/ndo/3x/configuration/cisco-nexus-dashboard-orchestrator-configuration-guide-aci-371/ndo-configuration-aci-infra-cloudsec-37x.html を参照してください。

Cisco ACI マルチサイトのオーバーレイ コントロール プレーン

Cisco ACI ファブリックでは、リーフノードに接続されているすべてのエンドポイントに関する情報は、スパインノードに用意された COOP データベースに保存されます。エンドポイントがリーフノードにローカルで接続されたことが検出されるたびに、そのリーフノードが COOP コントロール プレーン メッセージを発信して、エンドポイント情報（IPv4/IPv6 および MAC アドレス）をスパインノードと共有します。スパインはさらに、COOP を用いてスパイン間でこの情報を同期します。

Cisco ACI マルチサイト展開では、エンドポイント間の水平方向通信を可能にするために、検出されたエンドポイントのホスト情報を別々のファブリックに属するスパインノード間で交換する必要があります。ホスト情報をサイト間で交換する必要があるのは、そのエンドポイントが実際に通信する必要がある場合のみです。サイトにまたがって拡張された EPG に属するエンドポイントの場合（EPG 内の通信はコントラクトの定義がなくてもデフォルトで許可されているため）と、拡張されていない EPG に接続されたエンドポイントが相互の通信を定義済みコントラクトによって許可されている場合が該当します。このように動作を制御することは重要です。これにより、一部のエンドポイントの情報のみがサイトにまたがって同期されるため、サイト全体でサポートできるエンドポイントの総数を増やすことができます。

注： ブリッジドメインレベルでエンドポイント情報の交換を制御できます。したがって、複数の EPG が同じブリッジドメインに属している場合、ポリシーがそれらの EPG の 1 つについてルート情報の交換を指示すると、他のすべての EPG にもエンドポイント情報が送信されます。また、前述のように、優先グループのメンバーとして EPG を展開するか、vzAny を使用して「permit-all」コントラクトのプロバイダーやコンシューマとして設定すると、それらの EPG のメンバーとして検出されたすべてのエンドポイントとホストルート情報を交換できるようになります。

図 67 は、オーバーレイ コントロール プレーンでのイベントのシーケンスについて詳細を示しています。

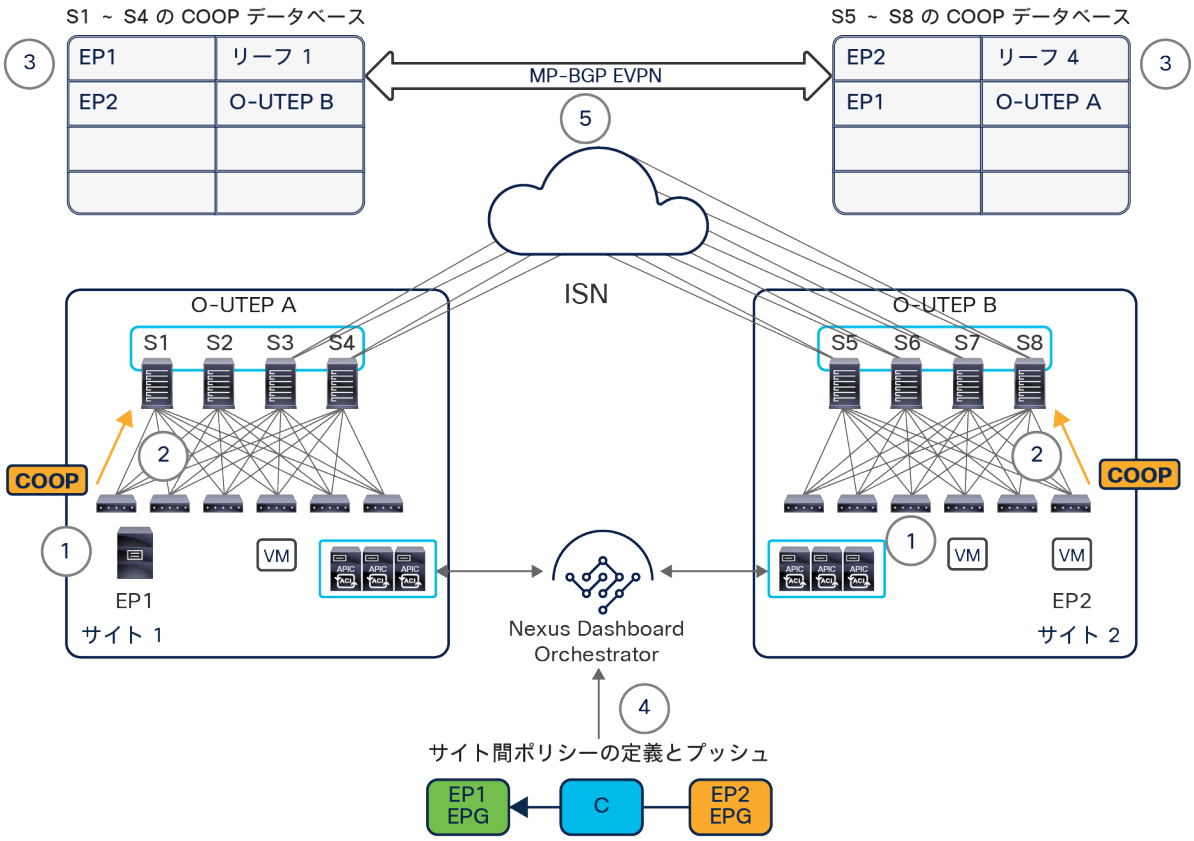


図 67.
Cisco ACI マルチサイトのオーバーレイ コントロール プレーン

1. エンドポイント EP1 と エンドポイント EP2 が、それぞれサイト 1 とサイト 2 に接続されます。
2. 各ファブリック内で COOP 通知が生成されます。この通知は、EP1 と EP2 が検出されたリーフノードからローカルスパインノードに送信されます。
3. このエンドポイント情報が、ローカルの COOP データベースに保存されます。サイト 1 のスパインノードは、ローカルに接続されたエンドポイントを認識します。サイト 2 のスパインノードも同様です。この時点では、EP1 EPG と EP2 EPG の情報がサイト間で交換されないことに注意してください。これらのエンドポイントの通信が必要となるポリシーがまだないためです。
4. サイト間ポリシーが、Cisco Nexus Dashboard Orchestrator で定義され、2 つのサイトにプッシュされレンダリングされます。
5. サイト間ポリシーが作成されると、EVPN タイプ 2 更新がサイト間でトリガーされ、EP1 と EP2 のホストルート情報が交換されます。エンドポイント情報は常に O-UTEP アドレスに関連付けられていて、各エンドポイントが検出されたサイトを一義的に識別することに注意してください。したがって、エンドポイントが同じファブリックに属する異なるリーフノード間を移動する場合、追加の EVPN 更新は必要ありません。エンドポイントが別のサイトに移行する際に必要になります。

注：エンドポイントのホストルート情報を交換するために、常に EVPN タイプ 2 更新のみがサイト間で送信されることに注意してください。Cisco ACI マルチサイトの現在の実装では、ブリッジドメインに関連付けられた IP サブネットプレフィックスのサイト間アドバタイズに EVPN タイプ 5 ルート更新を使用しません。接続されたエンドポイントが属する BD がレイヤ 2 で拡張されているか、各サイトでローカルに定義されているかにかかわらず、EVPN タイプ 2 更新が常に使用されます。

前述のように、EVPN-RID アドレスを使用して、MP-BGP EVPN 隣接関係が、異なるファブリックに属するスパインノード間に確立されます。各サイトが属する BGP 自律システムに応じて、MP 内部 BGP（MP-iBGP）セッションと MP 外部 BGP（MP-eBGP）セッションの両方がサポートされます。eBGP セッションをサイト間に展開すると、隣接関係のフルメッシュが NDO によって自動的に作成されます。すなわち、外部 IP ネットワークに接続されている各サイトのスパインが、すべてのリモートスパインスイッチとの間で EVPN ピアリングを確立します（図 68）。

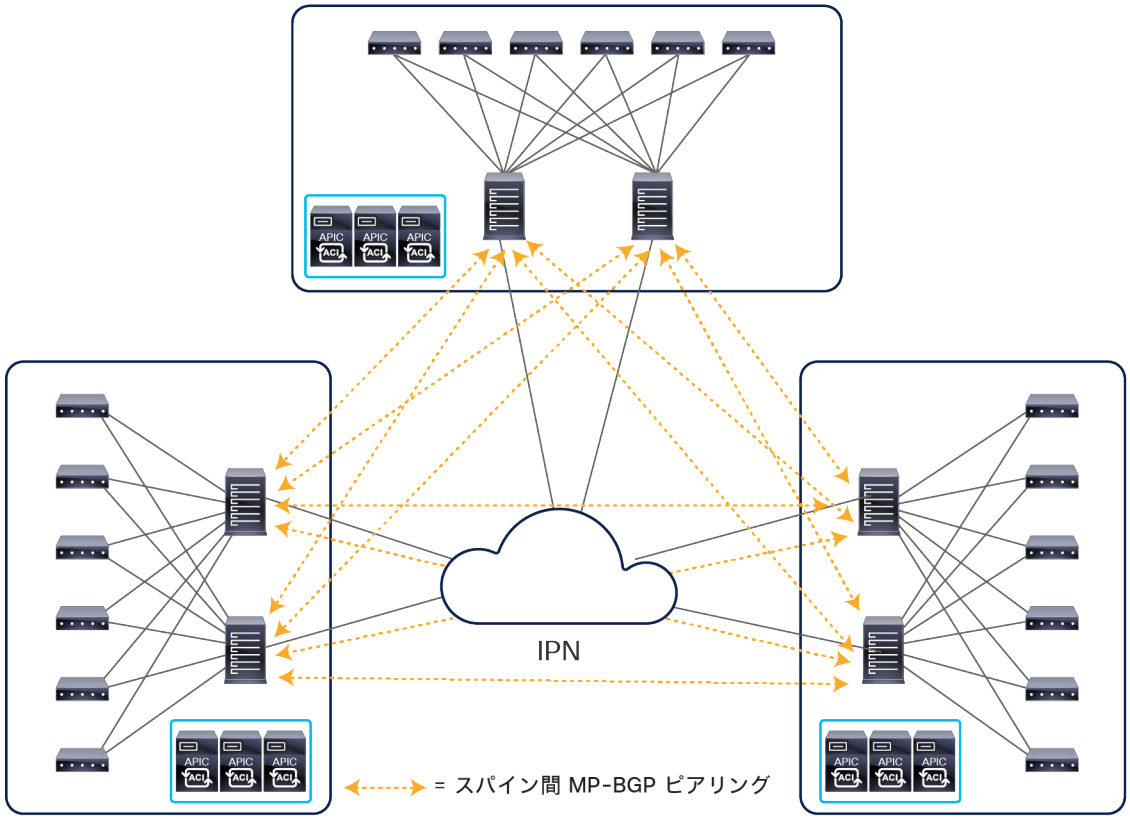


図 68.
サイト間に確立されたフルメッシュの EVPN セッション

サイト間で iBGP を使用する場合は、フルメッシュを使用するかルートリフレクタノードを導入するかを選択できます。このノードは、一般的に外部 RR（Ext-RR）と呼ばれます。iBGP を展開する場合でも、デフォルトの動作であるフルメッシュのピアリングを使用することをお勧めします。限られた数のサイト（つまり、20 未満）が相互接続されるのが常であり、拡張性に懸念が生じることはないと予想されるためです。

ただし、ルートリフレクタを導入したい場合は、Ext-RR ノードをいくつか展開し、それぞれのノードを別のサイトに配置することで復元力を確保する必要があります。図 69 に示すように、外部ルートリフレクタノードは相互にピアリングし、すべてのリモートスパインノードともピアリングします。

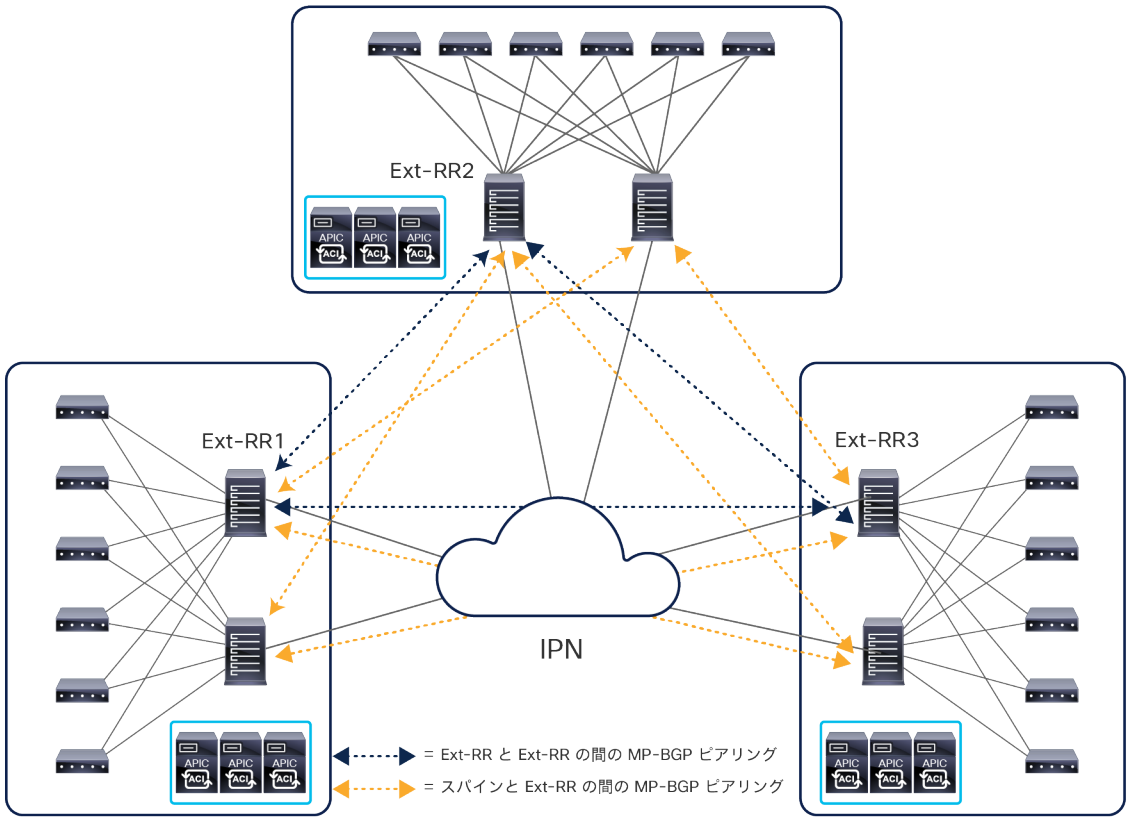


図 69.
サイトにまたがる MP-iBGP EVPN 展開での外部ルートリフレクタの使用

内部の実装方法に起因する Ext-RR ノードの展開に関するベストプラクティスの推奨事項は以下のとおりです。

- Ext-RR として構成されていないスパインノードは、常に少なくとも 1 つのリモート Ext-RR ノードとピアリングしている必要があります。スパインは、サイト内で EVPN 隣接関係を確立しません。つまり、Ext-RR ノードとして構成されていないスパインは、常に 2 つのリモート Ext-RR とピアリングしている必要があります（リモート Ext-RR ノードに障害が発生した場合に動作を継続するため）。その結果、サイト間で確立する必要がある EVPN 隣接関係の全体的な数を実質的に減らすことができないため、2 サイトの展開で Ext-RR を構成することはほとんど意味がありません。
- 3 サイト以上の展開の場合、上の図 69 に示すように、最初の 3 つのサイトのみに 1 つの Ext-RR ノードを定義します（合計 3 つの Ext-RR ノード）。

注： 上記で説明した Ext-RR ノードは、別々のサイトに展開されたスパインノード間に MP-BGP EVPN ピアリングを確立するために使用されます。これらのノードの機能は、内部 RR ノードの機能とは異なります。内部 RR ノードは、常に L3Out 論理接続で学習された外部 IPv4/IPv6 プレフィックスを同じファブリックに属するすべてのリーフノードに配布するために展開されます。

Cisco ACI マルチサイトのオーバーレイデータプレーン

サイト間でエンドポイント情報が交換されると、VXLAN データプレーンを使用したレイヤ 2 とレイヤ 3 のサイト間通信が可能になります。この通信を確立する方法を詳しく見ていく前に、レイヤ 2 マルチ宛先トラフィック（通常は BUM と呼ばれます）がサイト間で処理される方法を理解する必要があります。

サイトにまたがるレイヤ 2 BUM トラフィックの処理

VXLAN が展開されている場合、複数のレイヤ 3 ホップで隔てられたエンドポイントが、同じ論理レイヤ 2 ドメインに属しているかのように通信できる、論理的抽象化が可能です。したがって、これらのエンドポイントには、レイヤ 2 マルチ宛先フレームを送信する機能が必要です。このフレームは、実際の物理的な場所に関係なく、同じレイヤ 2 セグメントに接続されている他のすべてのエンドポイントが受信できます。

この機能を実現するには、次の 2 つの方法があります。1 つ目はエンドポイントを相互接続するレイヤ 3 インフラストラクチャによって提供されるネイティブ マルチキャスト レプリケーション機能を使用する方法、2 つ目は送信元 VXLAN TEP（VTEP）デバイスで入力レプリケーション機能を有効にする方法です。後者の方法では、各 BUM フレームのユニキャストのコピーが複数作成され、同じレイヤ 2 ドメインに属するエンドポイントが接続されているすべてのリモート VTEP に送信されます。

Cisco ACI マルチサイトの設計では、サイト間 BUM 転送にこの 2 つ目のアプローチを採用しており、マルチサイト対応のスパインスイッチが入力レプリケーション機能を実行します。これを採用した理由は、相互接続されたファブリックが世界中に展開される可能性があり、1 つ目のアプローチでは相互接続ネットワーク インフラストラクチャ全体でマルチキャストを適切にサポートすることが困難になるおそれがあるからです（このアプローチは、Cisco ACI マルチポッドアーキテクチャで採用されています）。

サイト間でのレイヤ 2 BUM フレームの送信は、フラッディングが有効な（つまり、[サイト間BUMトラフィック許可（Intersite BUM Traffic Allow）] フラグが設定された）ストレッチブリッジドメインに対してのみ必要です。レイヤ 2 BUM トラフィックには、3 つの異なるタイプがあります。以下では、ブリッジドメインのサイト間で BUM が許可されている場合のサイト間転送動作についてタイプごとに説明します。ここでは、ブリッジドメインの特定の構成ノブが APIC レベルで変更されることはなく、ブリッジドメインの構成は Cisco Nexus Dashboard Orchestrator レベルでのみ制御されると仮定しています。

- レイヤ 2 ブロードキャストフレーム（B）：このフレームは、常にサイト間で転送されます。レイヤ 2 ブロードキャスト トラフィックの特殊なタイプとして ARP があります。これについての考慮事項は、「サイト間のサブネット内ユニキャスト通信」セクションを参照してください。
- レイヤ 2 の宛先不明のユニキャストフレーム（U）：このフレームは、デフォルトではサイト間でフラッディングされず、ユニキャストモードで転送されます。宛先 MAC がローカルスパインの COOP データベースに登録されていることが前提です（そうでない場合、トラフィックは受信したスパインによってドロップされます）。ただし、Cisco Nexus Dashboard Orchestrator のブリッジドメイン固有の設定で、[宛先不明のL2ユニキャスト（L2 UNKNOWN UNICAST）] トラフィックに関連付けられた [フラッド（flood）] オプションを選択することにより、この動作を変更できます。
- レイヤ 2 マルチキャストフレーム（M）：同じ転送動作が、ブリッジドメイン内の（つまり、送信元と受信者が同じ IP サブネットにあるか、違う IP サブネットにあるが同じブリッジドメインに属する）レイヤ 3 マルチキャストフレーム、または「本当」の（つまり、宛先 MAC アドレスがマルチキャストであり、パケットに IP ヘッダーがない）レイヤ 2 マルチキャストフレームに当てはまります。どちらの場合も、ブリッジドメ

インが拡張されているサイト間でトラフィックを転送するには、そのブリッジドメインで BUM 転送が有効になっている必要があります。

注：上記の説明は、マルチキャストルーティングが有効になっていない場合を想定しています。Cisco ACI マルチサイトでのマルチキャストルーティングのサポートに関するその他の考慮事項については、「マルチサイトにおけるレイヤ 3 マルチキャスト」セクションを参照してください。

図 70 は、サイト間でレイヤ 2 BUM フレームを送信するために必要なイベントのシーケンスを示しています。

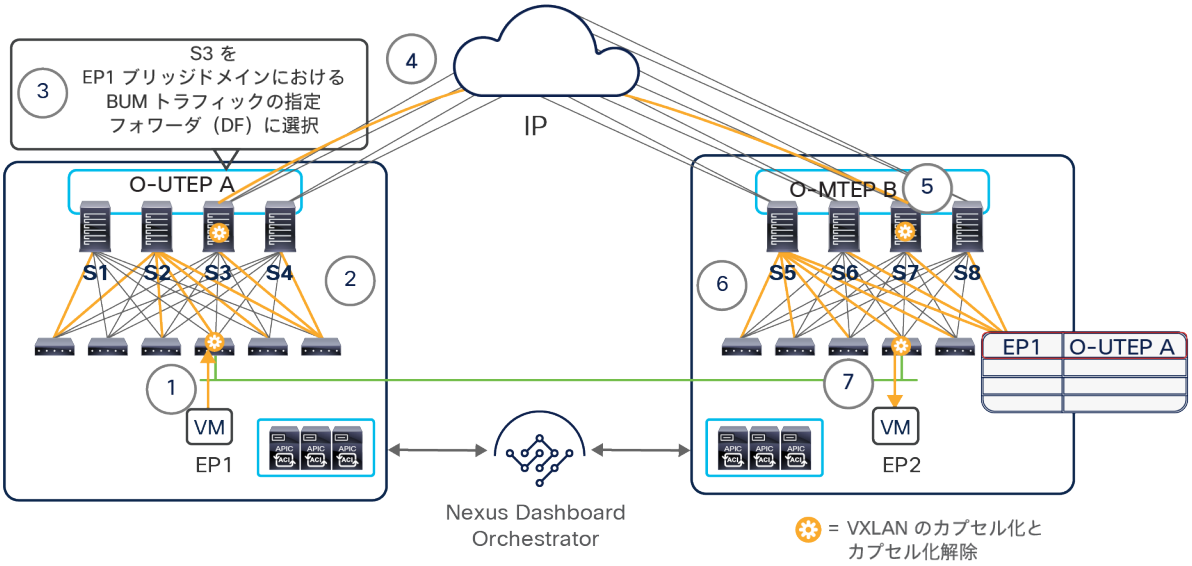


図 70.
サイトにまたがるレイヤ 2 BUM トラフィック

1. あるブリッジドメインに属する EP1 が、レイヤ 2 BUM フレームを生成します。
2. フレームのタイプと、対応するブリッジドメインの設定（上記で説明）によっては、リーフがそのレイヤ 2 ドメインでトラフィックをフラッディングする必要があります。このフレームは、VXLAN カプセル化の後、そのファブリック内でブリッジドメインに関連付けられた GIPo と呼ばれるマルチキャストグループを宛先として、GIPo に関連付けられたマルチ宛先ツリーの 1 つに従って送信されます。これによって、フレームが他のすべてのリーフおよびスパインノードに到達するようになります。
3. 外部のサイト間ネットワークに接続されているスパインノードの 1 つが、そのブリッジドメインの指定フォワーダとして選択されます（この選択は、スパインノード間で IS-IS プロトコル交換を使用して行われます）。指定フォワーダは、そのブリッジドメインの各 BUM フレームを、同じストレッチブリッジドメインを持つすべてのリモートサイトに複製する役割を担っています。
4. 指定フォワーダが、BUM フレームのコピーを作成しリモートサイトに送信します。パケットの VXLAN カプセル化の際に使用される宛先 IP アドレスは、各リモートサイトを識別する特別な IP アドレス（O-MTEP）です。この IP アドレスは、特にサイト間で BUM トラフィックを送信する際に使用されます。O-MTEP は、サイト間ネットワークに接続されているすべてのリモートスパインノードで定義された別のエニーキャスト IP アドレスです（各サイトが、一意の O-MTEP アドレスを使用します）。パケットの VXLAN カプセル化の際に使用される送信元 IP アドレスは、サイト間ネットワークに接続されているすべてのローカルスパインノードに展開されたエニーキャスト O-UTEP アドレスです。

注：O-MTEP（Cisco Nexus Dashboard Orchestrator の GUI では [オーバーレイマルチキャストTEP（Overlay Multicast TEP）] と表示されます）は、さらに別の IP アドレスであり、ファブリックを接続するレイヤ 3 ネットワークに送信する必要があります。

5. リモートスパインノードの 1 つがパケットを受信すると、そのヘッダーに含まれる VNID 値を同じブリッジドメインに関連付けられたローカルで有効な VNID 値に変換し、そのブリッジドメインに対して定義されたローカルのマルチ宛先ツリーの 1 つに従ってトラフィックをサイトに送信します（VXLAN ヘッダー内の宛先として、その BD にローカルで割り当てられたマルチキャストグループが使用されます）。
6. トラフィックがサイト内で転送され、すべてのスパインノードと、このブリッジドメインにアクティブに接続されているエンドポイントを持つリーフノードに到達します。
7. 受信側のリーフノードが、VXLAN ヘッダーに含まれる情報を使用して、BUM フレームを送信したエンドポイント EP1 のサイトのロケーションを学習します。また、このリーフノードは、ブリッジドメインに関連付けられたすべて（または一部）のローカルインターフェイスに BUM フレームを送信し、エンドポイント EP2（この例では）がフレームを受信できるようにします。

前述のように、定義されたブリッジドメインはすべて、マルチキャスト グループ アドレス（またはマルチキャストアドレスのセット）に関連付けられます。これらは通常、GIPo アドレスと呼ばれます。構成されたブリッジドメインの数によっては、同じ GIPo アドレスが、異なるブリッジドメインに関連付けられることがあります。この場合、これらのブリッジドメインの 1 つでフラッディングがサイトにまたがって有効になっていると、同じ GIPo アドレスを使用する他のブリッジドメインの BUM トラフィックもサイト間で送信され、受信したスパインノードでドロップされます。この動作により、サイト間ネットワークの帯域幅使用率が増加する可能性があります。

これを避けるため、ブリッジドメインが Cisco Nexus Dashboard Orchestrator GUI で BUM フラッディングが有効なストレッチドメインとして構成された場合、デフォルトでは、GIPo アドレスがマルチキャストアドレスの別の範囲から割り当てられます。GUI では、[WAN帯域幅の最適化（OPTIMIZE WAN BANDWIDTH）] フラグで設定されます。図 71 に示すように、NDO で直接作成される BD に対してはデフォルトで有効になっています。

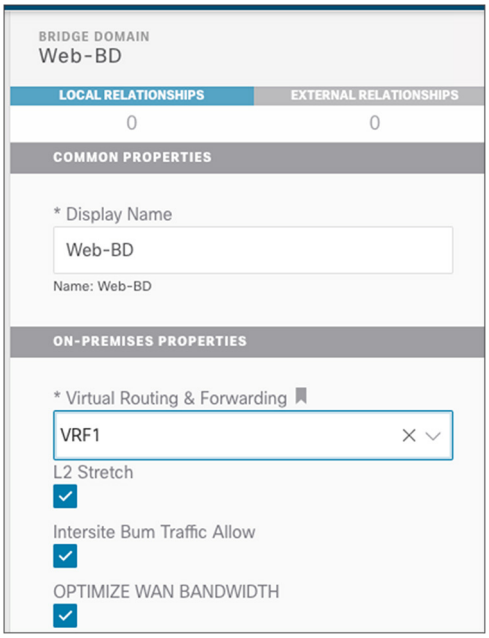


図 71.
サイト間の BUM フラッディングの最適化

ただし、ブリッジドメイン構成が APIC ドメインからインポートされた場合、フラグはデフォルトで無効になっています。手動でブリッジドメインを構成し、すでに関連付けられている GIPo アドレスを変更する必要があります。これを行うと、そのブリッジドメインが展開されているすべてのリーフノードで GIPo アドレスが更新されます。その数秒間、ブリッジドメインのファブリック内 BUM トラフィックが停止することに注意してください。

サイト間のサブネット内ユニキャスト通信

サイト間のサブネット内 IP 通信を実現するためには、まず送信元エンドポイントと宛先エンドポイント間の ARP 交換を完了する必要があります。前述のように、ARP は特殊なタイプのレイヤ 2 ブロードキャスト トラフィックであり、サイト間の転送はブリッジドメインレベルで制御できます。

注： 以下の説明では、ブリッジドメインが単一の IP サブネットに関連付けられていることを想定としていますが、同じブリッジドメインに対して複数のサブネットが定義されているシナリオもあり得ます。

考慮すべきシナリオは以下の 2 つです。

- ブリッジドメインで ARP フラッディングが有効：Cisco Nexus Dashboard Orchestrator で BUM 転送が**有効**なストレッチブリッジドメインを作成した場合のデフォルト設定です。この場合、前のセクションで説明した動作（図 70）と同一の動作になります。ARP 要求がリモートサイトの宛先エンドポイントに到達し、リモートリーフノードが送信元エンドポイントのサイトのロケーションを学習します。その結果、ARP ユニキャスト応答が VXLAN カプセル化されて、EP1 サイトを識別する O-UTEP アドレスに直接送信されます。これを受信したスパインノードの 1 つが VNID とクラス ID の変換を実行し、EP1 が接続されているローカルリーフノードにそのフレームを送信します（図 72）。

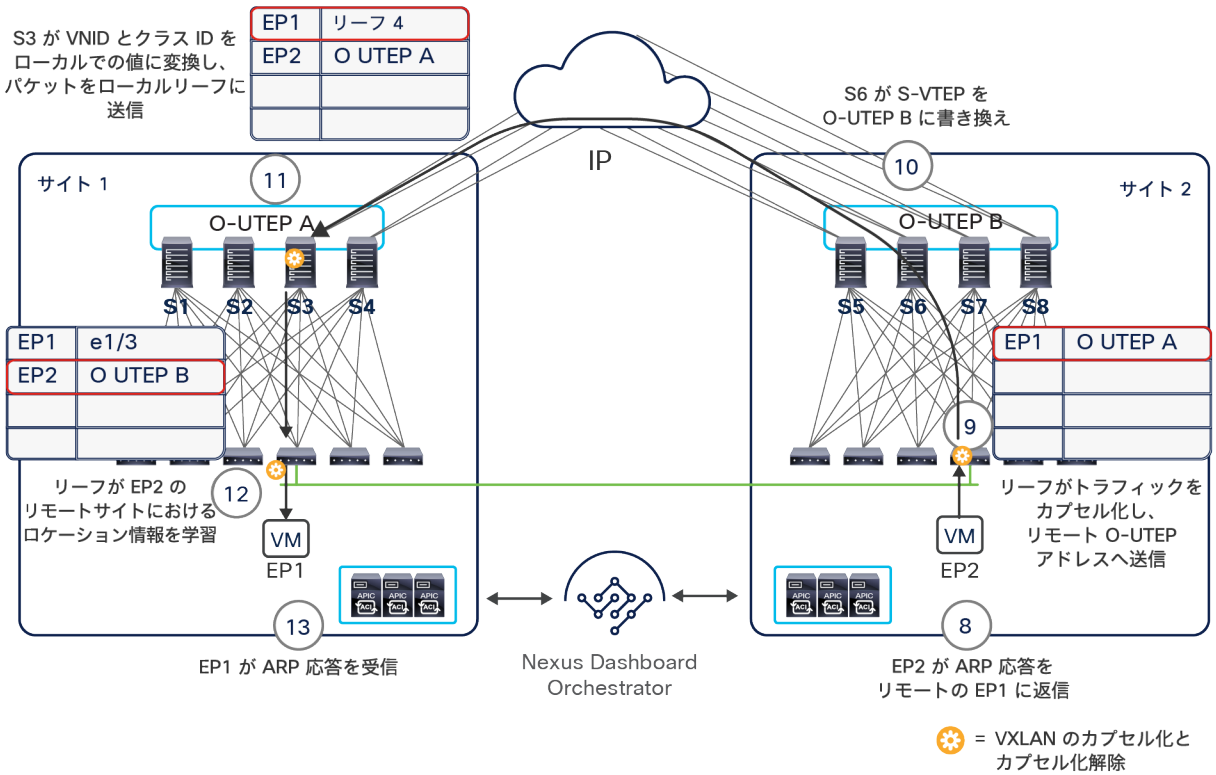


図 72.
サイト 1 の EP1 に配信される ARP 応答

ARP 応答を受信することで、サイト 1 のリーフノードが EP2 のサイトの情報を学習できることに注意してください（EP2 は、EP2 サイトのスパインノードを識別する O-UTEP B アドレスに関連付けられています）。

注：ブリッジドメインの ARP フラッディングを有効にし（APIC レベルで）、対応する BUM 転送を有効にしないのは設定不備です。この設定では、サイト間の ARP 交換が完了できず、サイト間のブリッジドメイン内接続が切断されます。このような問題を防ぐために、Nexus Dashboard Orchestrator でのみ BD 転送特性を設定することを強くお勧めします。

- ブリッジドメインで ARP フラッディングが無効：Cisco Nexus Dashboard Orchestrator で BUM 転送が**無効**なストレッチブリッジドメインを作成した場合のデフォルト設定です。この場合、ローカルスパインが受信した ARP 要求はサイト間でフラッディングできないため、VXLAN ユニキャストパケットへのカプセル化が必要です（図 73）。

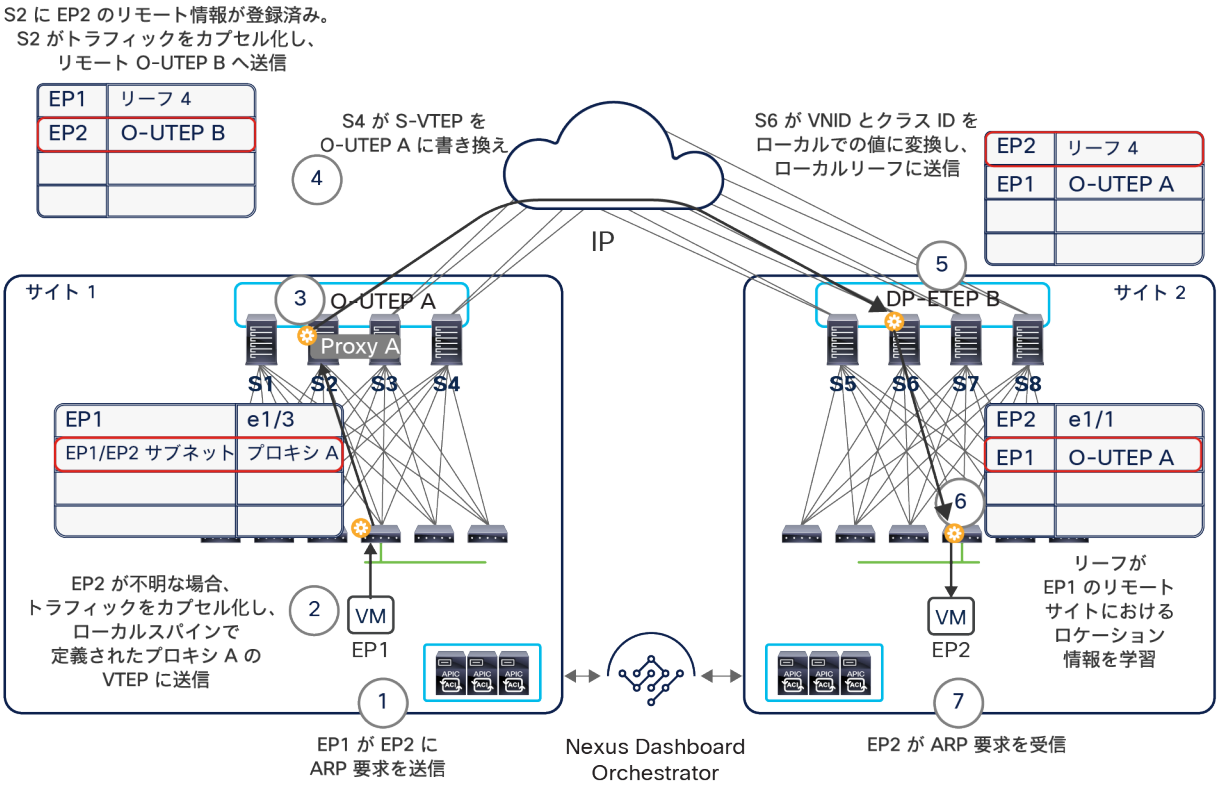


図 73.
フラッディングを用いない場合のサイト間の ARP 要求

以下は、このユースケースでサイト 2 に ARP 要求を送信するために必要な一連のステップです。

1. EP1 が EP2 の IP アドレスの ARP 要求を生成します。
2. ローカルリーフノードが ARP ペイロードを検査し、宛先である EP2 の IP アドレスを決定します。最初に EP2 の IP 情報がローカルリーフで見つからなかった場合、COOP データベースでルックアップを実行するために、ARP 要求がカプセル化されてすべてのローカルスパインで定義されたプロキシ A のエニーキャスト VTEP アドレスに送信されます（ローカル ルーティング テーブルにインストールされているパーベイシブ EP1/EP2 IP サブネット情報が使用されます）。
3. ローカルスパインノードの 1 つが、ローカルリーフノードから ARP 要求を受信します。

4. サイト間で「ユニキャストモード」を用いて ARP 要求を転送できるかどうかは、主にリモートエンドポイントの IP アドレスに関する情報（リモートスパインから MP-BGP EVPN コントロールプレーン経由で受信）が COOP データベースにあるかどうかに依存します。まず、リモートエンドポイントの IP アドレスがわかっている（つまり、EP2 が「サイレントホスト」ではない）場合は、EP2 が接続されているサイトを識別するリモート O-UTEP アドレスをローカルスパインノードが知っているため、パケットをカプセル化して ISN を介してリモートサイトに送信できます。スパインが VXLAN カプセル化パケットの送信元 IP アドレスも書き換えることに注意してください。リーフノードの VTEP アドレスがローカルサイトを識別するローカル O-UTEP A アドレスに置き換えられます。この動作は非常に重要です。サイト間の EVPN コントロールプレーンでのやり取りで説明したように、外部 IP ネットワークが認識するスパインノードの IP アドレスを、EVPN-RID、O-UTEP、O-MTEP に限定する必要があるためです。
5. この VXLAN フレームがリモートスパインノードの 1 つで受信されます。このノードが元の VNID とクラス ID をローカルで有効な値に変換し、ARP 要求をカプセル化したうえで、EP2 が接続されているローカルリーフノードに送信します。
6. リーフノードがフレームを受信してカプセル化を解除し、リモートエンドポイント EP1 のクラス ID とサイトのロケーションの情報を学習します。
7. 次に、EP2 を学習しているインターフェイスからこのフレームが送信され、エンドポイントに到達します。

この時点で、EP2 がユニキャスト ARP 応答を返信できるようになります。この応答は、図 72 で説明したのと同じ一連のステップで EP1 に配信されます（唯一の違いは、サイト間でフラッディングが有効になっていないことです）。

前のステップ 4 でリモートエンドポイントの IP アドレスがサイト 1 の COOP データベースになかった場合、Cisco ACI リリース 3.2(1) 以降導入された新しい「ARP Glean」機能が、リモートの「サイレントホスト」に ARP 要求の受信と応答を促し、リモートサイトで検出できるようします（下の図 74）。

注： 「ARP Glean」メッセージは、ブリッジドメインに関連付けられたエニーキャストゲートウェイ IP アドレスから発信されます。つまり、ARP 応答は EP2 が接続されているリーフノードによって常にローカルで消費されます。しかし、このプロセスによって EP2 が検出できるようになります（この時点で「サイレント」ではなくなっています）。

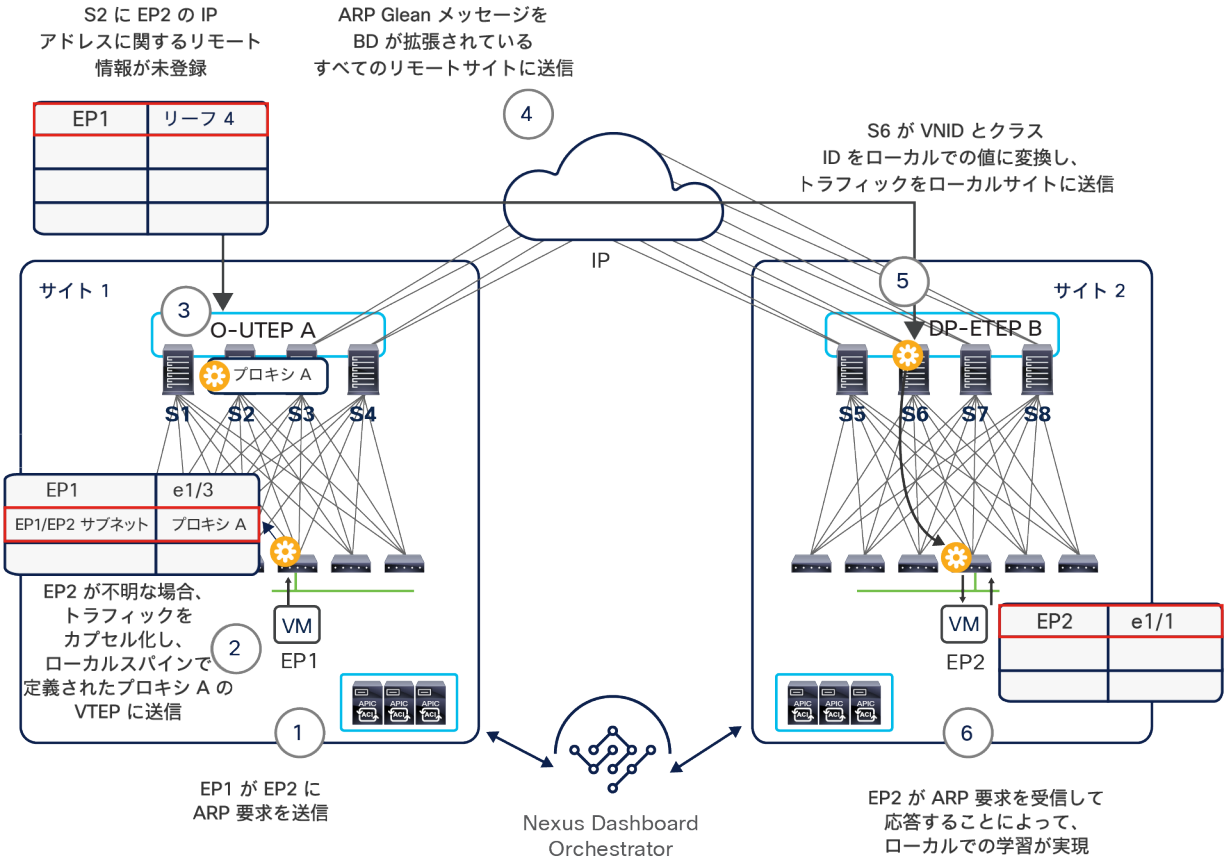


図 74.
サブネット内通信のユースケースにおける ARP Glean 機能

リモートエンドポイントの IP アドレスが検出され、EVPN コントロールプレーンを介してサイト間で共有されると、図 73 に示したように、EP1 によって発信された ARP 要求が、EP2 に向けてユニキャストモードで送信できるようになります。

ARP を交換する上記のプロセスが完了すると、各リーフノードは、通信しようとしているリモートエンドポイントのクラス ID とロケーションを完全に学習しています。その結果、この時点からトラフィックが常に双方向に流れるようになります。図 75 に示すように、サイトのリーフノードが常にトラフィックをカプセル化し、それを宛先エンドポイントが接続されたサイトを識別する O-UTEP アドレスに向けて送信します。

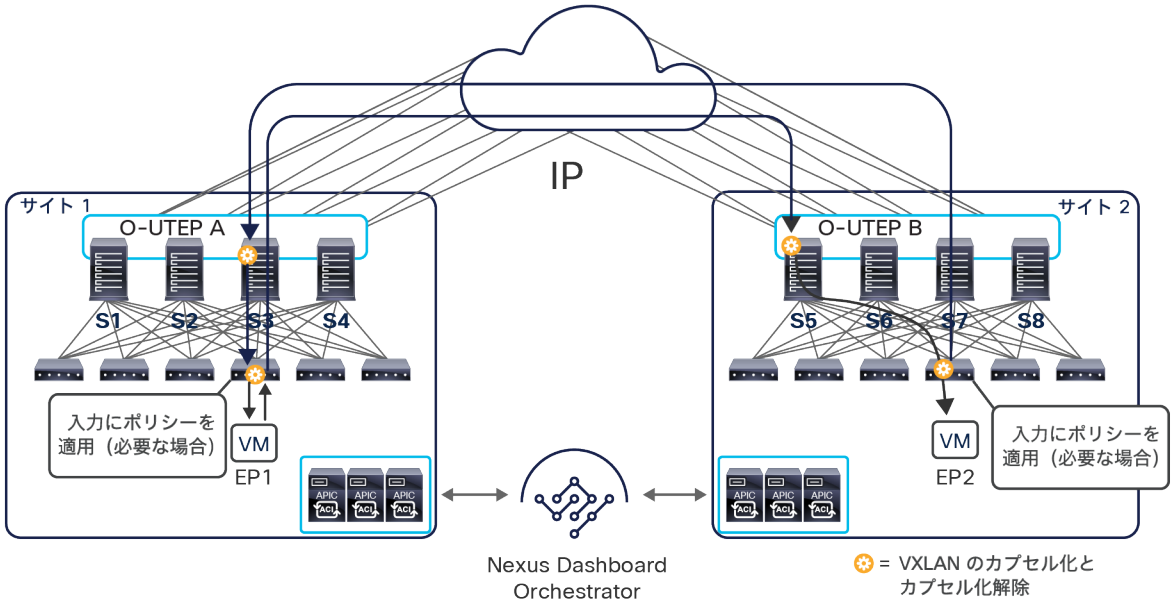


図 75.
サイト間のサブネット内通信

セキュリティポリシーの適用という点から見ると、サブネット内通信のユースケースについて考慮すべき主なシナリオは以下の 3 つです。

- EP1 と EP2 が同じ EPG と同じブリッジドメインにあり、マイクロセグメンテーションが構成されていない（かつ EPG が分離されるように構成されていない）場合：何のポリシーも適用されず、EP1 は EP2 と自由に通信できます。
- EP1 と EP2 が同じベース EPG と同じブリッジドメインにあり、コントラクトを持つ 2 つのマイクロ EPG に関連付けられている場合：定常状態では、発信元リーフノードの入力にポリシーが常に適用されます。
- **注：**データプレーン学習を無効にすると、この動作が変更され、出力リーフノードにポリシーが常に適用されます。
- EP1 と EP2 が同じブリッジドメインと同じ IP サブネットに属する異なる 2 つの EPG にある場合：EPG 間で定義されたコントラクトが通信を規定します。前の場合と同様、定常状態では通常、発信元リーフノードの入力にポリシーが適用されます。これに対する例外は、コントラクトがサービスグラフをポリシーベースリダイレクト（PBR）に関連付けている場合です。その場合、「ネットワークサービスの統合」セクションで詳しく説明するように、ポリシーの適用は、コンシューマエンドポイントとプロバイダーエンドポイントが接続されている場所に依存します。

サイト間のサブネット間ユニキャスト通信

サイト間でサブネット間通信を有効にする場合の考慮事項と機能は、サブネット内通信について前のセクションで説明したものと同様です。送信元エンドポイントと宛先エンドポイントが異なる EPG に属する場合（異なるブリッジドメイン、または複数の IP サブネットが構成されている同じブリッジドメインに属する場合）、EPG 間の水平方向通信を有効にするためには、Cisco Nexus Dashboard Orchestrator で EPG 間にコントラクトを作成する必要があります。一方、送信元エンドポイントと宛先エンドポイントが同じ EPG に属する場合、たとえ異なる IP サブネットにある場合でも、以下に説明するようにルーティングが行われ、コントラクトを構成する必要はありません。また、送信元エンドポイントは、常に接続先のローカルリーフノードを使用して ARP 情報を解決することでデフォルトゲートウェイを知り、リモートエンドポイント宛てのデータパケットをリーフノードに送信します。次に、リーフノードはトラフィックを宛先エンドポイントに配信する必要があります。宛先ブリッジドメインがサイトにまたがって拡張されていない場合、以下の 2 つのシナリオが考えられます。

- EP2 の IP アドレスがサイト 2 でまだ検出されておらず、その結果、送信元のサイト 1 で認識されていない場合：Cisco ACI リリース 3.2(1) より前では、リーフからトラフィックを受信するローカルスパインがルックアップを実行します。EP2 が不明であるため、トラフィックはドロップされます。この場合でも、Cisco ACI リリース 3.2(1) で導入された「ARP Glean」機能を呼び出すと、図 76 に示すように、EP2 の IP アドレスが検出できるようになります。
- EP2 が検出されると、以下の箇条書きで説明するように、EP2 へのデータプレーン通信を確立できます。

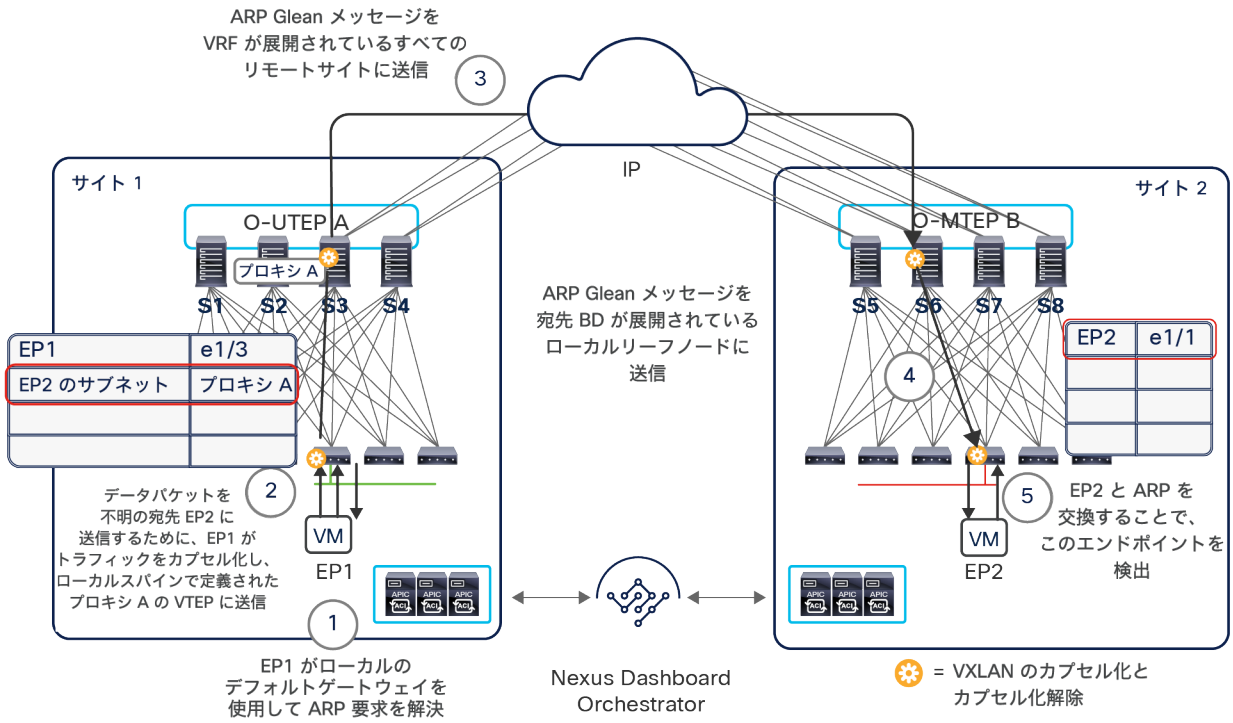


図 76.
サブネット間通信のユースケースにおける ARP Glean 機能

- EP2 の IP アドレスが送信元のサイト 1 に属するスパインの COOP データベースに保存されている場合：図 77 に示すように、ローカルスパインノードがトラフィックをカプセル化して EP2 が属するリモートサイトを識別する O-UTEP アドレスに送信し、最終的に宛先エンドポイントがそのパケットを受信します。

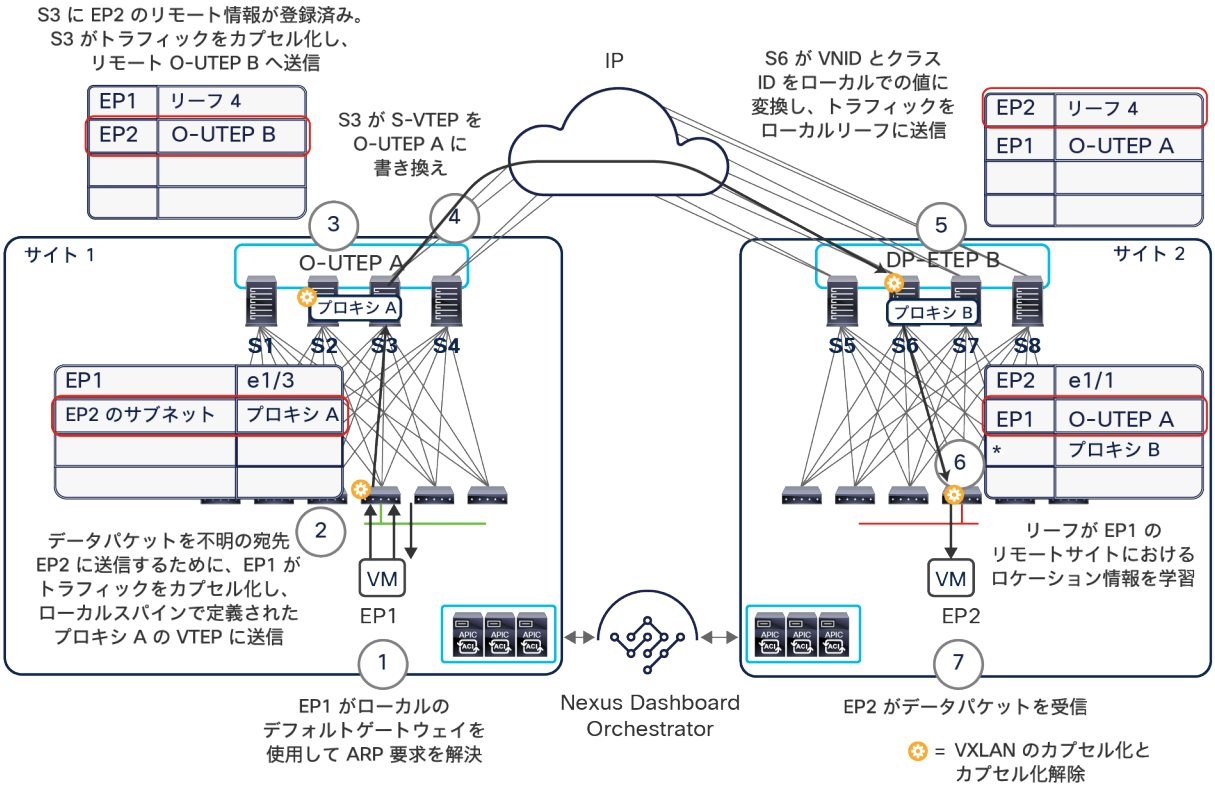


図 77.
サイト間のサブネット間トラフィックの配信

マルチサイトにおけるレイヤ 3 マルチキャスト（テナント ルーテッド マルチキャスト – TRM）

Cisco ACI でのレイヤ 3 マルチキャスト通信は、リリース 2.0(1) からサポートされています。ただし、Cisco ACI リリース 4.0(1) より前は、単一ファブリック（単一ポッドまたはマルチポッド）の展開のみに対応しています。

注： Cisco ACI でのレイヤ 3 マルチキャストのサポートに関する詳細は、https://www.cisco.com/c/ja_jp/td/docs/switches/datacenter/aci/apic/sw/2-x/L3_config/b_Cisco_APIC_Layer_3_Configuration_Guide/b_Cisco_APIC_Layer_3_Configuration_Guide_chapter_01111.html を参照してください。

そのため、Cisco ACI リリース 4.0(1) より前の場合、別々の Cisco ACI ファブリックに接続された送信元と受信者の間で L3 マルチキャスト通信を拡張するには、これらのファブリックを個別に展開し、外部 L3 ネットワークを利用してファブリック間でマルチキャストフローを伝送する必要があります。このモデルを下の図 78 に示します。

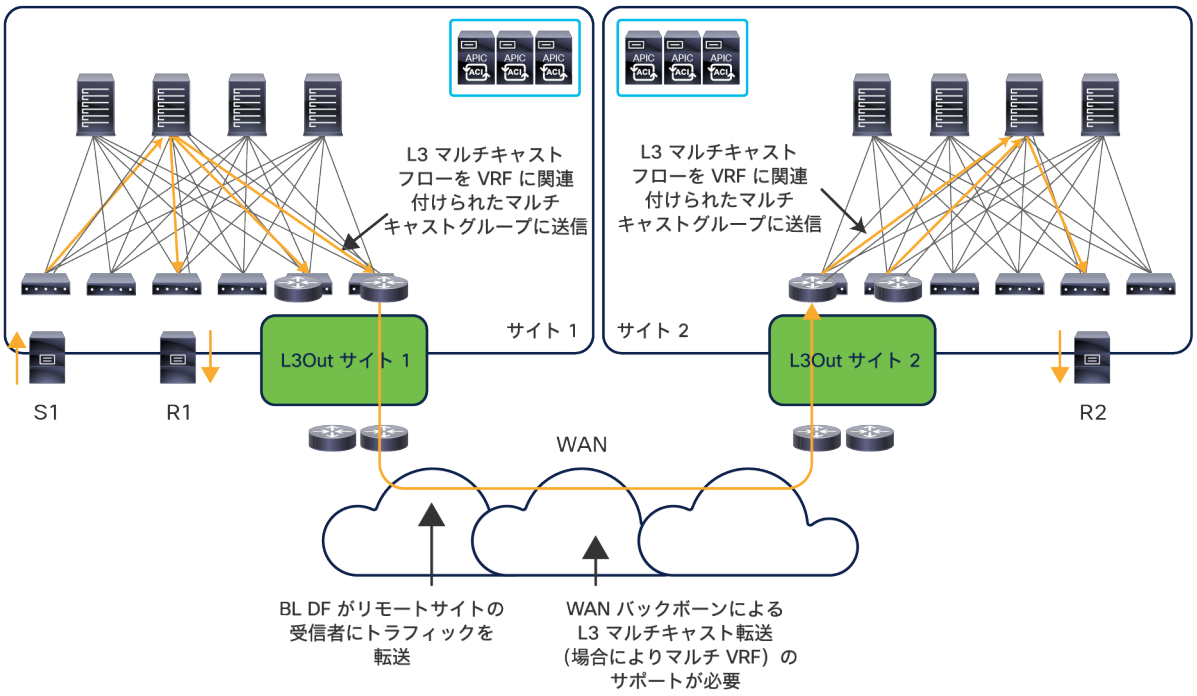


図 78.
異なる Cisco ACI ファブリックにまたがるレイヤ 3 マルチキャストの転送

Cisco ACI リリース 4.0(2) と Cisco Multi-Site Orchestrator リリース 2.0(2) では、同じ Cisco ACI マルチサイトドメインに属する Cisco ACI ファブリック間で TRM が利用できます。つまり、異なる Cisco ACI ファブリックに接続された送信元と受信者の間のレイヤ 3 マルチキャストフローが、VXLAN カプセル化トラフィックとして ISN を通して転送できるようになります。これは、レイヤ 2 とレイヤ 3 のユニキャスト通信について前のセクションで説明した方法と同様で、マルチキャスト対応のバックボーンネットワークを展開する必要がなくなります（図 79）。

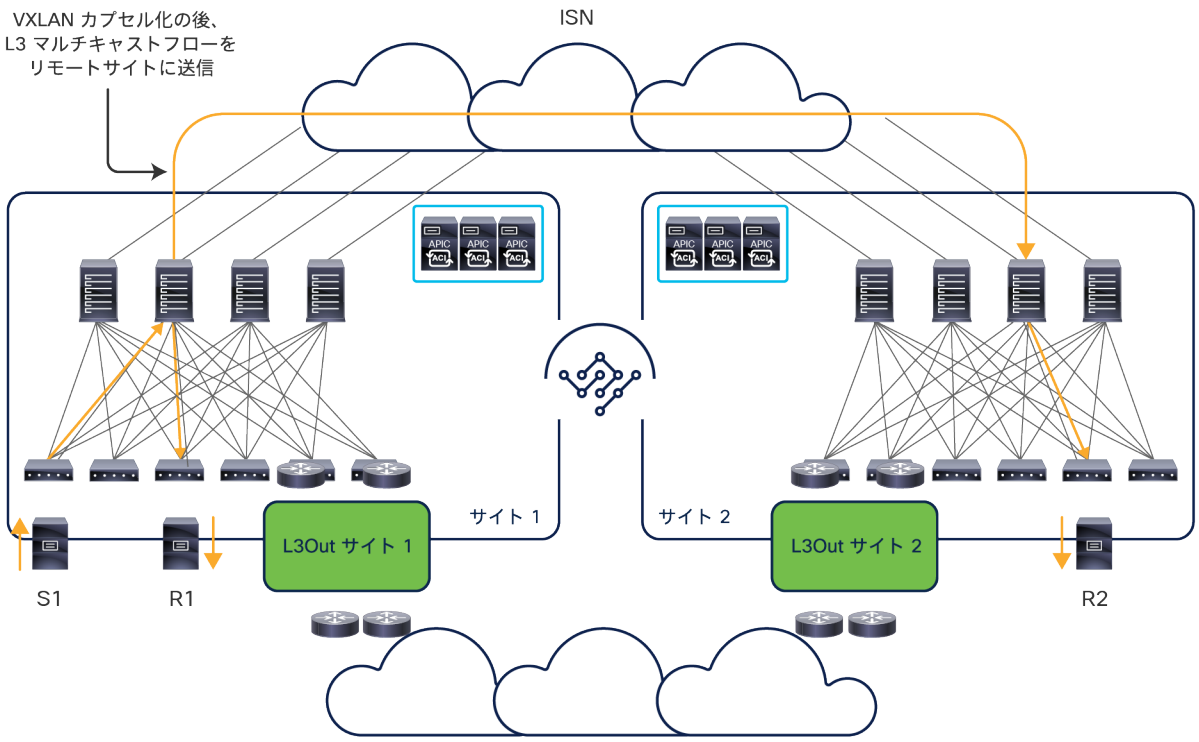


図 79.
Cisco ACI マルチサイトにおけるレイヤ 3 マルチキャストのサポート（Cisco ACI リリース 4.0(2) 以降）

コントロールプレーンレベルとデータプレーンレベルの両方でサイト間の転送を有効にする方法を説明する前に、このモデルに関連する設計と展開の考慮事項を確認しておくことが重要です。

- 単一ファブリックの展開と同様、レイヤ 3 マルチキャストの転送がサポートされるのは、第 2 世代のリーフデバイス（Cisco Nexus 9300 EX モデル以降）が展開されている場合のみです。
- 単一サイトの場合と同様、TRM は PIM ASM と PIM SSM の両方に対応しています。マルチキャストの送信元と受信者は、同じサイト、異なるサイト、および Cisco ACI ファブリックの外部にも展開できます（すべての組み合わせに完全に対応しています）。PIM ASM の展開に必要なランデブーポイント（RP）の使用に関しては、Cisco ACI リリース 5.0(1) と Cisco Multi-Site Orchestrator リリース 3.0(1) 以降、Cisco ACI ファブリックの内部にある RP がマルチサイトでサポートされるようになりました。これより前の ACI リリースでは、マルチサイトでレイヤ 3 マルチキャストを展開するために、Cisco ACI ファブリックの外部にある RP を使用する必要があります。
- Cisco ACI を使用したマルチキャストルーティングは、「常にルーティング」アプローチでサポートされます。すなわち、送信元と受信者が同じ IP サブネットに属していても、そのトラフィックの TTL が 2 回減算されます（入力と出力のリーフノードで）。また、Cisco ACI では、マルチキャストルーティングを有効にすることなく、レイヤ 2 マルチキャスト通信としてこれらのフローを処理するだけで、サブネット内でマルチキャストを転送できることに注意してください。ただし、これら 2 つの動作は共存できず、ブリッジドメインでマルチキャストルーティングが有効になると、「常にルーティング」アプローチが使用されます。
- マルチキャストルーティングのトラフィックに対しては、サイト内ポリシーもサイト間ポリシーも適用されません。したがって、トラフィックの転送に送信元の EPG と受信者の EPG の間のコントラクトは不要です。

- 構成の観点から見ると、マルチキャストルーティングは Cisco Nexus Dashboard Orchestrator を用いて VRF レベルで有効にする必要があります。これにより、GIPo マルチキャストアドレスが VRF に割り当てられます。ブリッジドメインに関連付けられる（ファブリック内で BUM トラフィックを転送するために使用される）GIPo アドレスとは異なる GIPo アドレスのセットが VRF 用に予約されています。
- ある VRF に対してマルチキャストルーティングを有効にしたら、レイヤ 3 マルチキャストの送信元または受信者が属する個々のブリッジドメインに対しても有効にする必要があります。これらのブリッジドメインがサイトにまたがって拡張されているかどうかは関係ないことに注意してください。このソリューションはどちらの場合でも機能します（拡張されているかどうかにかかわらず、送信元ブリッジドメインと受信者ブリッジドメインが別のサイトにある場合）
- トラフィックがあるサイトの送信元から発信され、ISN を通してリモートサイトに転送される場合、スパインに適切な（VRF VNID と EPG クラス ID の）変換エントリが作成されている必要があります。この点は、ユニキャスト通信のユースケースですでに説明したとおりです。さらに、送信元ブリッジドメインがサイトにまたがって拡張されていない場合は、リモートサイトのリーフノードに送信元 IP サブネットを構成して、マルチキャストトラフィックを受信したときにリバース パス フォワーディング（RPF）のチェックが正常に行われるようにする必要もあります。レイヤ 3 マルチキャストが有効になっている EPG やブリッジドメインすべてではなく、送信元が接続されているものに限定してこの構成を作成するには、Cisco Nexus Dashboard Orchestrator でマルチキャストの送信元を含む EPG を明示的に指定する必要があります。

マルチサイトドメインでのファブリック RP のサポート

TRM のコントロールプレーンとデータプレーンの動作に関する詳細に入る前に、Cisco ACI リリース 5.0(1) と Cisco Multi-Site Orchestrator リリース 3.0(1) で導入された新機能について説明する必要があります。この機能を用いると、同じマルチサイトドメインに属するファブリックの内部に、複数のエニーキャスト RP ノードを構成することができます。

これより前のソフトウェアリリースでは、マルチサイトドメインに属するファブリックに冗長 RP 機能を提供するには、図 80 に示すように、外部ネットワークドメインにエニーキャスト RP ノードを展開する必要がありました。

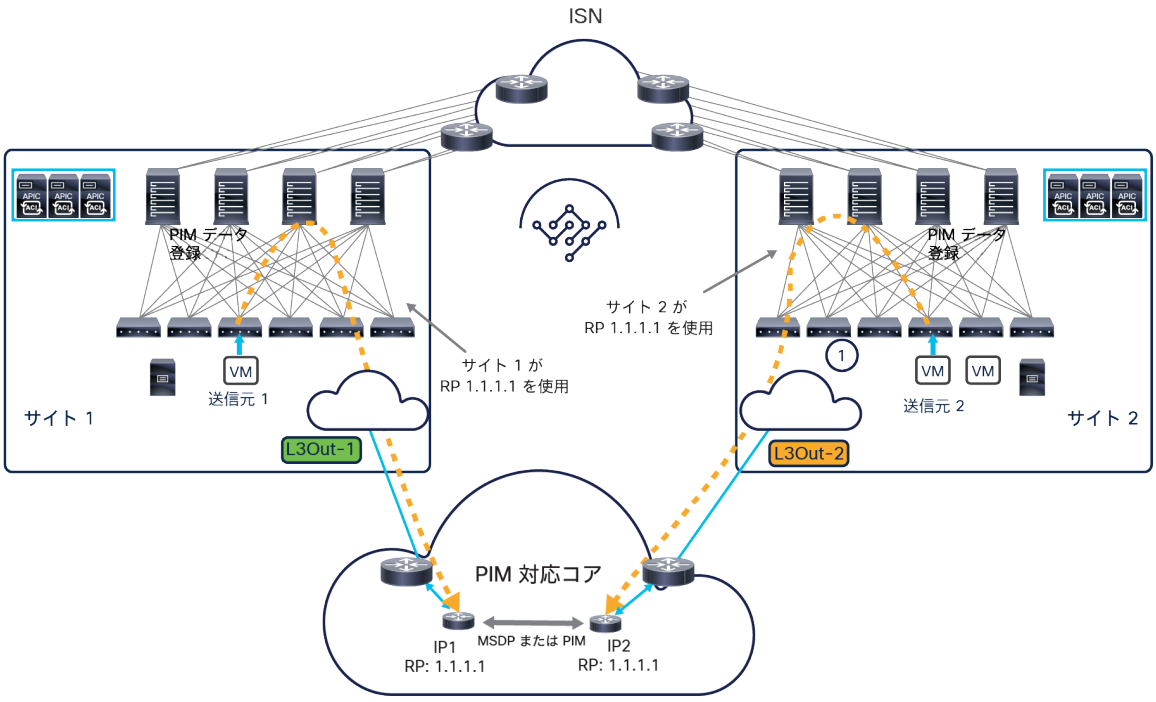


図 80.
外部レイヤ 3 ネットワークに展開されたエニーキャスト RP

異なる ACI ファブリックに接続された送信元がマルチキャストストリームの生成を開始すると、送信元が接続されたファーストホップルータ（FHR）が、外部レイヤ 3 ネットワークに展開された RP ノードに向けて PIM データ登録メッセージを送信します。送信元が属するサイトによって、この PIM メッセージを受信する RP ノードが異なる可能性があります。これは、すべての RP が同じエニーキャスト RP アドレスを共有するためです（冗長 RP 展開を実現するエニーキャスト RP の場合と同様の基本概念です）。したがって、さまざまな RP ノード間に追加のコントロールプレーンを展開して、アクティブな送信元に関する情報を RP ノード間で同期する必要があります。通常、このコントロールプレーンには、マルチキャストソース検出プロトコル（MSDP）またはエニーキャスト RP PIM（RFC 4610）が導入されます。

マルチサイト展開でファブリック RP がサポートされているため、異なるファブリックに属する複数の ACI リーフノードに冗長 RP 機能を容易に展開できます（図 81）。

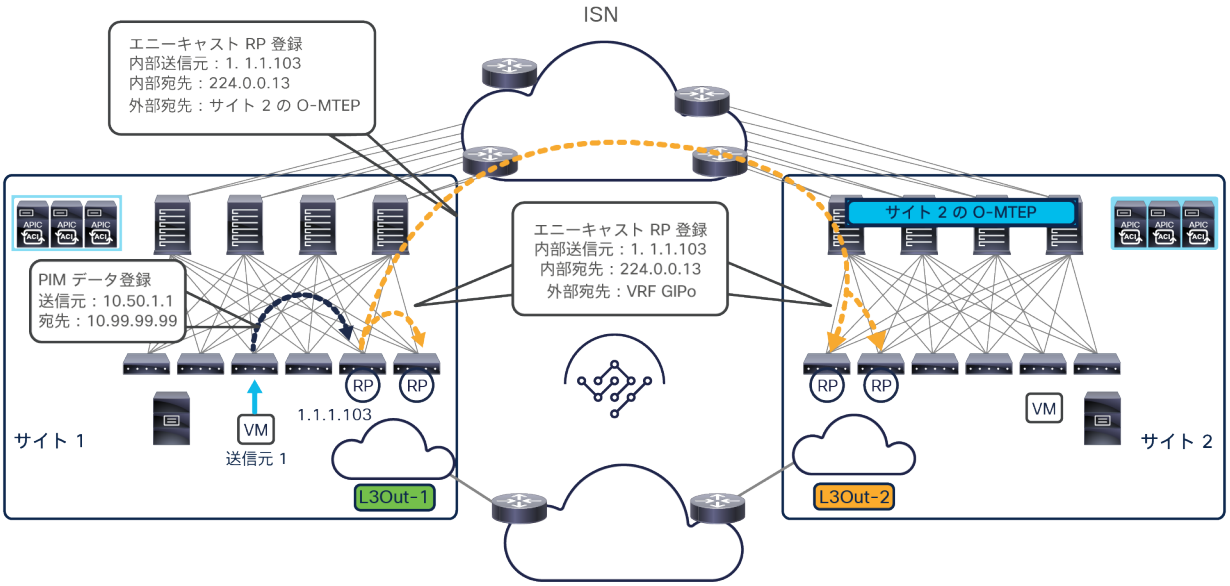


図 81.
マルチサイト展開でのファブリック RP の使用

エニーキャスト RP アドレスは、各ファブリック内に展開された一連のボーダーリーフノードで有効になっています。そのため、外部レイヤ 3 ネットワークからは、各サイトに展開された L3Out 接続を介して複数のエニーキャスト RP ノードが到達可能であると見なされます。上図の例では、サイト 1 の ACI リーフノードに接続された送信元がマルチキャストストリームを生成すると、FHR デバイスが PIM データ登録メッセージを生成し、ローカルサイトで使用可能な BL ノードの 1 つに転送します。

このメッセージを受信した BL ノードは、エニーキャスト RP 登録メッセージ を（マルチキャストグループ 224.0.0.13 宛てに）生成します。このメッセージは、VRF GIPo マルチキャストアドレスを宛先とする VXLAN トラフィックとしてローカルファブリック内で転送されます（各 VRF には通常、専用の GIPo アドレスが割り当てられます）。これにより、RP として構成された他のすべてのローカル BL ノードが情報を受信できるようになります。さらに、その VRF GIPo に関連付けられたトラフィックの指定フォワーダとして選択されたスパインが、VRF が拡張されているすべてのリモートサイトに向けてエニーキャスト RP 登録メッセージを複製します。さらに、リモート RP ノードは、アクティブ化された送信元に関する情報を受信することもできます。

構成方法としては、NDO でエニーキャスト RP アドレスを設定し VRF に関連付けます。次に、サイトごとに少なくとも 1 つの（その同じ VRF に関連付けられている）L3Out で PIM を有効にする必要があります。これによって、これらの L3Out の一部である BL ノードで RP アドレスがアクティブ化されます。

注： ファブリック RP 機能を有効にするには、（対応する VRF が定義されている）各サイトで L3Out を定義する必要があります。マルチキャストの送信元と受信者がそのファブリックのみに接続されていて、その L3Out に接続された外部ネットワークとマルチキャストストリームを送受信する必要がない構成であっても、これが必要です。

次に、ファブリック RP を使用してレイヤ 3 マルチキャストストリームを送信元から宛先に配信する方法を説明します。送信元と受信者は、マルチサイトドメインに属する ACI ファブリック内部で接続することも、外部レイヤ 3 インフラストラクチャで接続することもできます。

TRM のコントロールプレーンとデータプレーンに関する考慮事項

一連の機能（IGMP スヌーピング、COOP、PIM）が Cisco ACI 内で連携して、ACI リーフノードに適切な (*,G) ステートと (S,G) ステートを作成します。図 82 は、サイトにまたがって展開されたエニーキャスト RP ノードを使用する PIM-ASM シナリオにおけるこれらの機能の動作を示しています。この例では、送信元がサイト 1 でアクティブ化され、受信者がサイト 2 と外部ネットワークに接続されています。

注： 運用が簡単になるため、ファブリック RP 構成を使用することをお勧めします。ACI ファブリックの外部に RP を展開する場合のコントロールプレーンとデータプレーンの動作に関する詳細は、付録 A を参照してください。

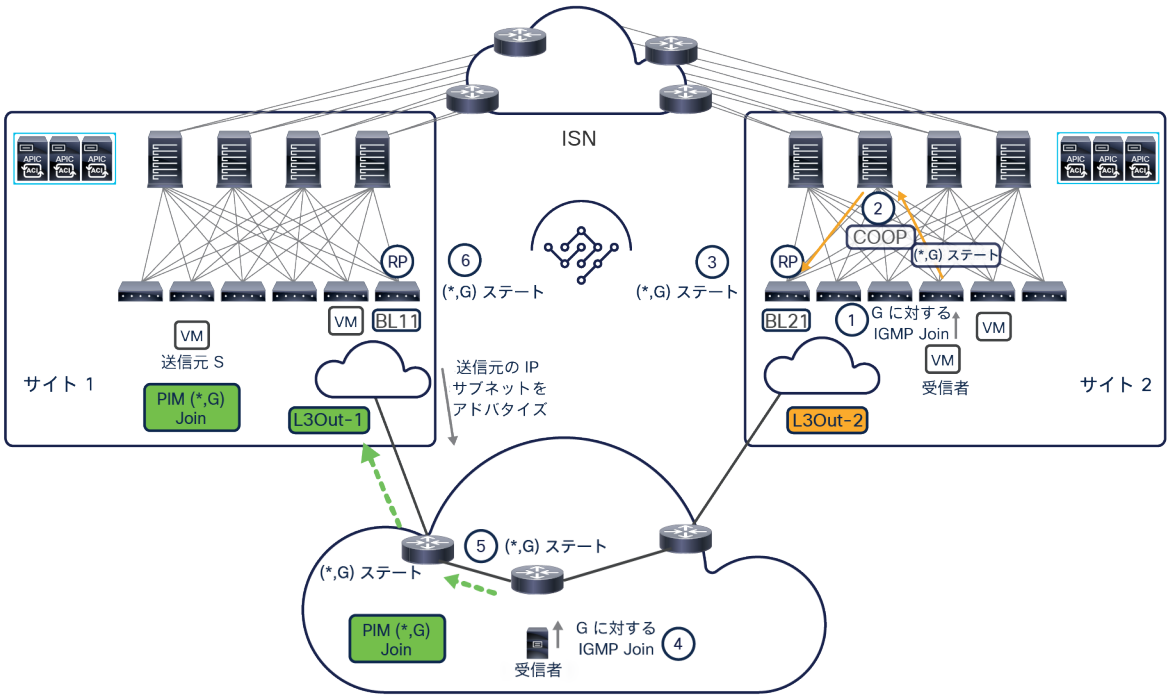


図 82.
内部と外部の受信者がマルチキャストグループ G に参加するときのコントロールプレーンでのアクティビティ

- サイト 2 の受信者が、グループ G に向けられたマルチキャストトラフィックに関する受信要求を宣言するために、IGMP-Join を発信します。
- 受信者が接続されている Cisco ACI リーフノード（ラストホップルータ（LHR））が、この IGMP-Join を受信します。LHR は、ローカルに接続された受信者の受信要求を登録し（(*,G) ローカルエントリを作成します）、COOP メッセージを生成してスパインに同じ情報を提供します。
- スパインは、グループ G に関する受信者の受信要求を登録し、COOP 通知を生成して、ファブリック RP として構成されているローカルボーダーリーフ（BL）ノードにこの情報を伝えます。(*,G) ローカル エントリが BL ノードに作成されます。
- 受信者が外部レイヤ 3 ネットワークに接続され、同じマルチキャストグループ G に関する IGMP-Join メッセージを送信します。

- LHR がメッセージを受信すると、ローカル (*,G) ステートを作成し、RP に向けて (*, G) PIM-Join メッセージを送信します。RP へのルートは両方の ACI ファブリックから外部ネットワークにアドバタイズされているため、ルーティング情報のみに基づいて RP へのベストパスが選択されます。
- 図 82 の例では、サイト 1 の RP BL ノードが直接接続された外部ルータから (*,G) PIM-Join メッセージを受信し、ローカル (*,G) ステートを作成します。

図 83 は、送信元がサイト 1 に接続され、マルチキャストグループ G に向けてトラフィックのストリーミングを開始した後に発生するコントロールプレーンでのアクティビティを示しています。

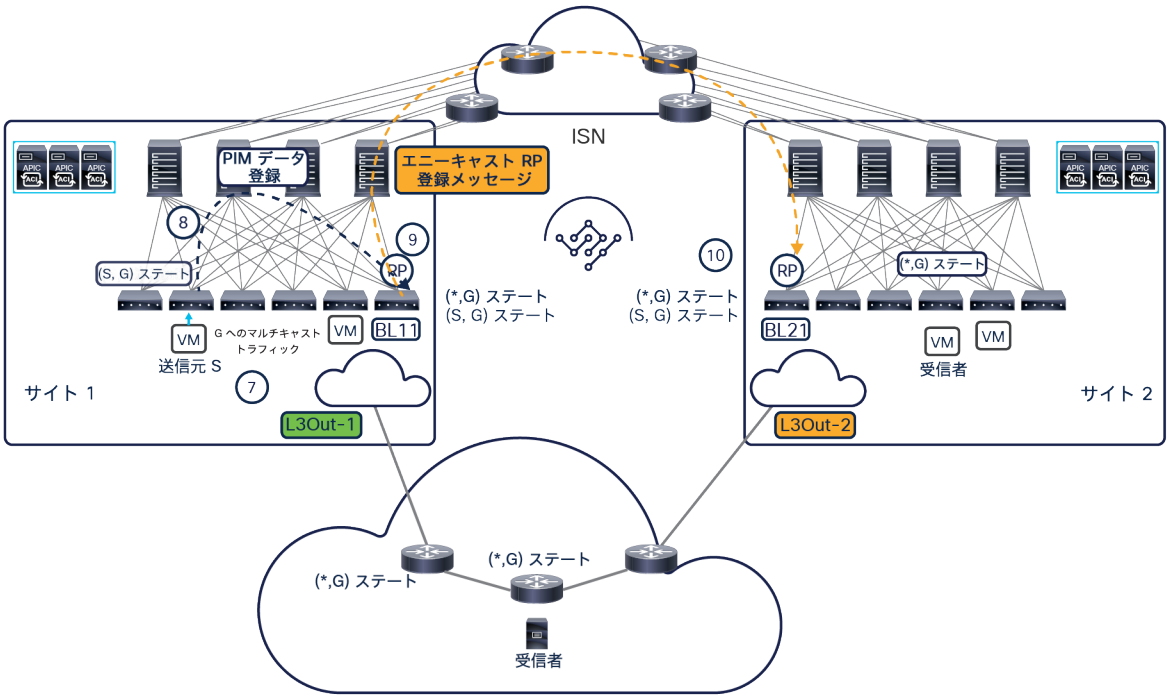


図 83.
内部送信元がマルチキャストグループ G に向けてトラフィックのストリーミングを開始するときのコントロールプレーンでのアクティビティ

- サイト 1 に接続されているマルチキャストの送信元 S が、グループ G 宛てのトラフィックのストリーミングを開始します。
- 送信元が接続されているファーストホップのリーフノード（FHR）が、(S,G) ローカルステートを作成し、RP に向けて PIM データ登録メッセージを送信します。PIM データ登録パケットは、IP ヘッダーに PIM プロトコル番号 103 が設定されたユニキャストパケットであり、RP として構成されたローカル ボーダー リーフ ノードに向けて Cisco ACI ファブリックを介して転送されます。
- 上の図 83 に示すように、RP が (S,G) ローカルステートを作成した後、エニーキャスト RP 登録メッセージを生成してサイトにまたがって転送します。
- リモートサイトの RP がこのメッセージを受信し、(S,G) ローカルステートを作成します。

この時点で、下の図 84 に示すように、マルチキャストストリームのデータプレーン転送が可能になります。

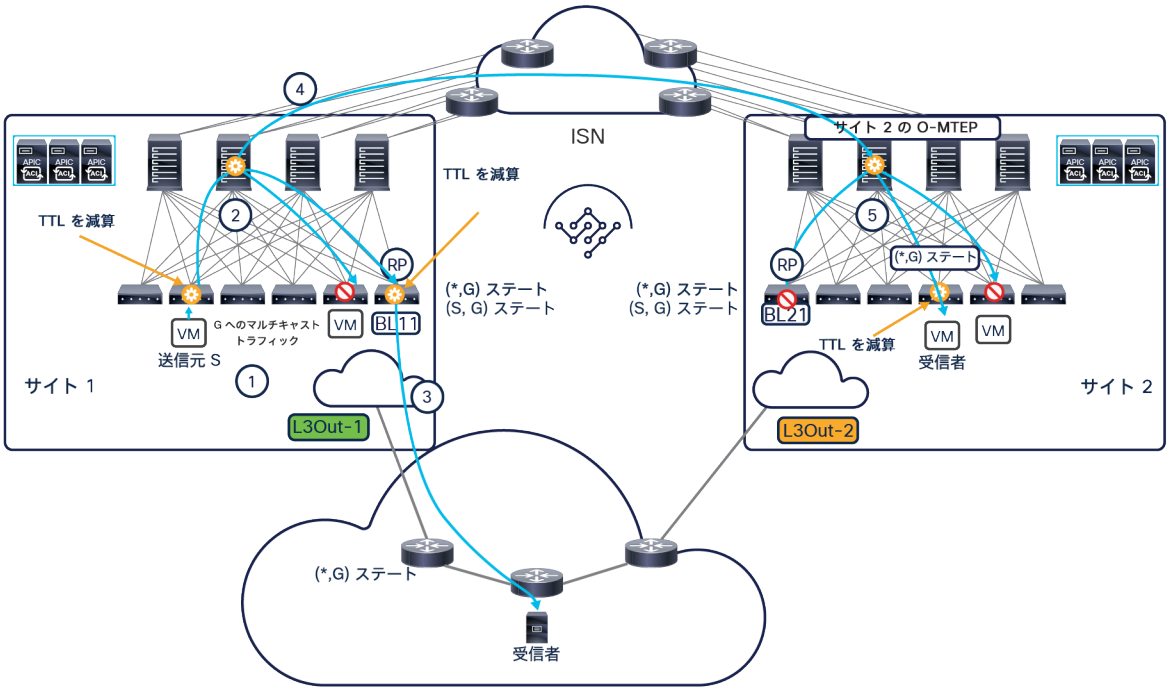


図 84.
リモートと外部の受信者に向けたマルチキャストストリームのデータプレーン転送

- 送信元 S が、グループ G に向けたマルチキャストストリームの送信を開始します。
- FHR がパケットの TTL を減算し、VXLAN がそのパケットをカプセル化します。外側の IP ヘッダーで使用される宛先アドレスは、VRF に関連付けられたマルチキャストグループ（GIPo）です。このパケットはファブリック内で複製され、BL11 ノードを含む、VRF が展開されているすべてのスパインとすべてのリーフノードに到達します。
- BL11 がトラフィックのカプセル化を解除して TTL を減算した後、このストリームを外部ネットワークに転送します。その結果、すでにグループに参加している外部の受信者にストリームが到達します。
- 上記のステップ 3 で説明したアクティビティと並行して、VRF の指定フォワーダとして選択されたスパインノードが、VRF が拡張されているすべてのリモートサイトに向けてストリームの入力レプリケーションを開始します。VXLAN カプセル化パケットの外側の宛先 IP アドレスは、常にリモートサイトの O-MTEP アドレスです。
- リモートサイトのスパインの 1 つがそのパケットを受信し、宛先 IP アドレスをローカル VRF の GIPo に変更した後、ファブリック内に転送します。このパケットは、VRF が展開されているすべてのリーフノードに到達し、すでにそのグループに参加している直接接続された受信者に転送されます。
- 注：ある VRF に属する送信元から発信され、特定のグループに送信されたマルチキャストトラフィックは、その VRF が拡張されているすべてのリモートサイトに複製されます。前述のように、そのグループからの受信を要求している受信者がある場合、受信側のスパインがファブリック内にトラフィックを転送します。受信者が検出されない場合は、スパインがそのトラフィックをドロップします。

図 85 に示すシナリオには、固有の考慮事項があります。このシナリオでは、送信元が外部ネットワークに接続され、受信者がマルチサイトドメインに属する異なる ACI ファブリックに接続されています。

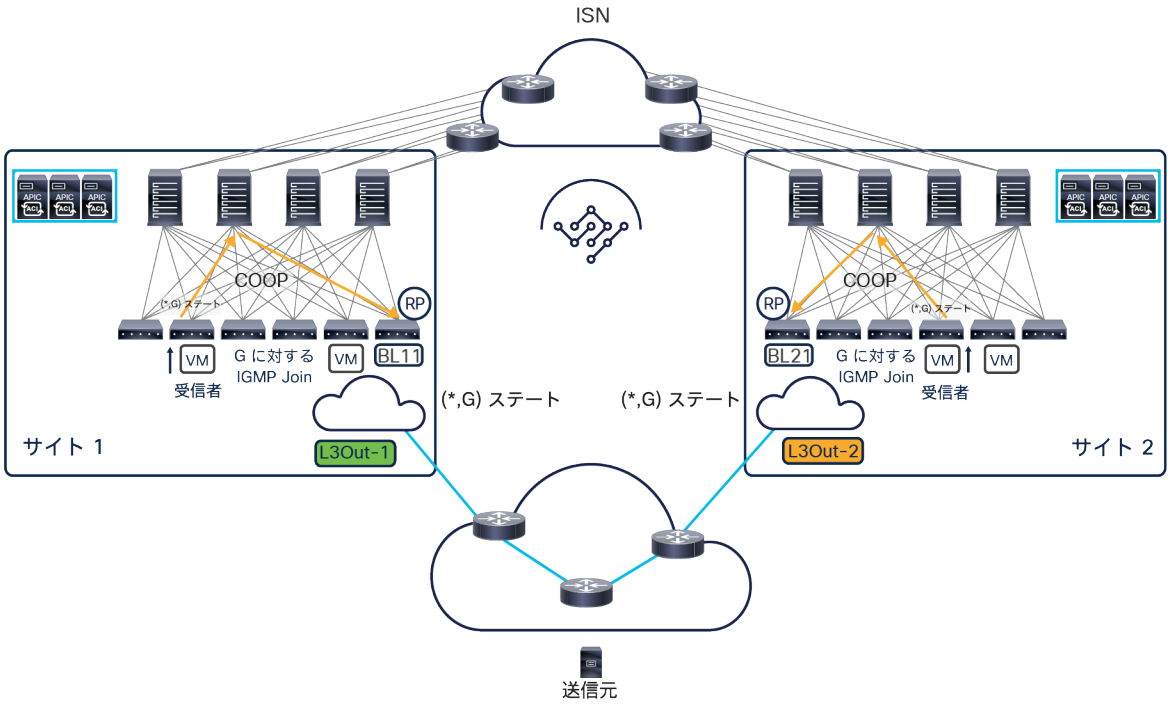


図 85.
内部受信者がマルチキャストグループ G に参加するときのコントロールプレーンでのアクティビティ

図 85 に示す、内部受信者が特定のマルチキャストグループ G に参加するときのコントロールプレーンでのアクティビティは、以前に図 81 に示したものと非常に類似しています。これにより、各ファブリックに展開された RP がローカル (*,G) ステートを適切に作成できます。

図 86 は、外部送信元がアクティブ化され、グループ G に向けてトラフィックのストリーミングを開始するときのコントロールプレーンでのアクティビティを示しています。

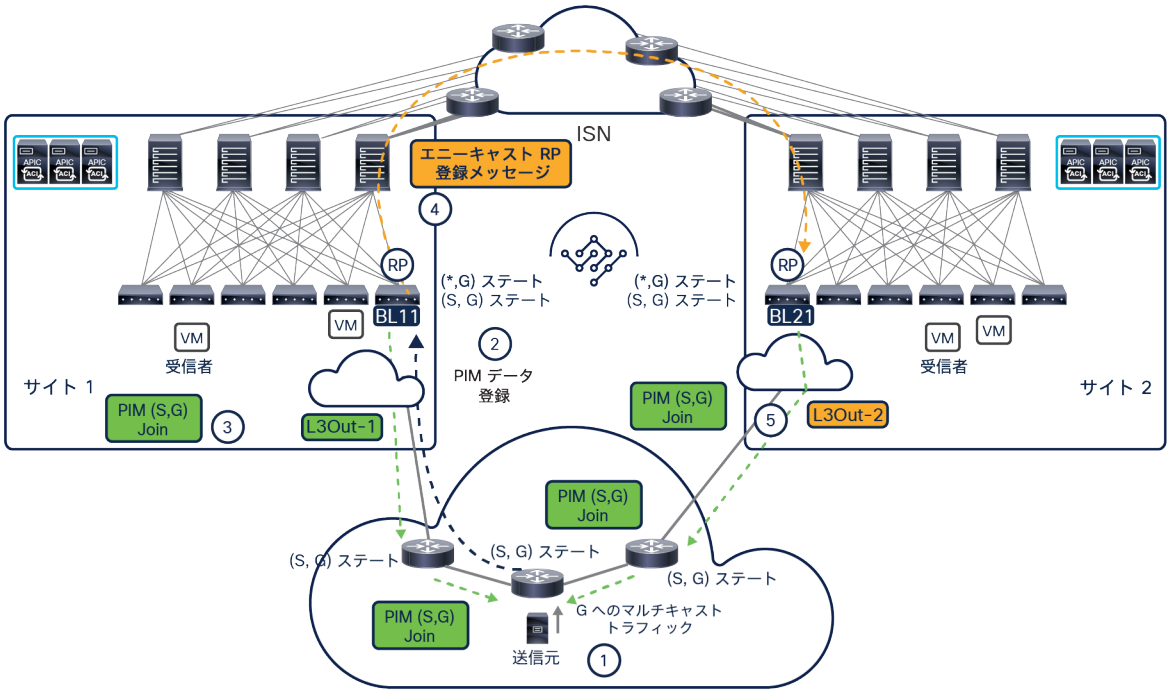


図 86.
外部送信元がマルチキャストグループ G に向けてトラフィックのストリーミングを開始するときのコントロールプレーンでのアクティビティ

- 外部送信元がマルチキャストグループ G に向けてトラフィックのストリーミングを開始します。
- FHR が PIM データ登録メッセージを RP に送信します。前述のように、ルーティング情報によって、このメッセージが特定のファブリック（この例ではサイト 1）の RP にステアリングされます。
- RP として機能しているサイト 1 の BL11 がメッセージを受信すると、ローカル (S,G) ステートを作成し、FHR に向けて (S,G) PIM-Join を送信します。これにより、BL11 と FHR の間にあるすべての L3 ルータで (S,G) ステートが構築されます。
- ステップ 3 のアクティビティと並行して、BL 11 が、ファブリック内の他のローカル RP（存在する場合）と、リモートサイトの RP に送信される、エニーキャスト RP 登録メッセージも生成します。
- サイト 2 の BL21 の RP がそのメッセージを受信すると、ローカル (S,G) ステートを作成し、送信元が接続されている FHR に (S,G) PIM-Join メッセージを送信します。このメッセージは、送信元から受信した G ストリームをサイト 1 とサイト 2 の両方向に複製する必要があることを FHR デバイスに伝えています。

コントロールプレーンでのアクティビティが完了すると、図 87 に示すように、マルチキャストストリームのデータ転送が可能になります。

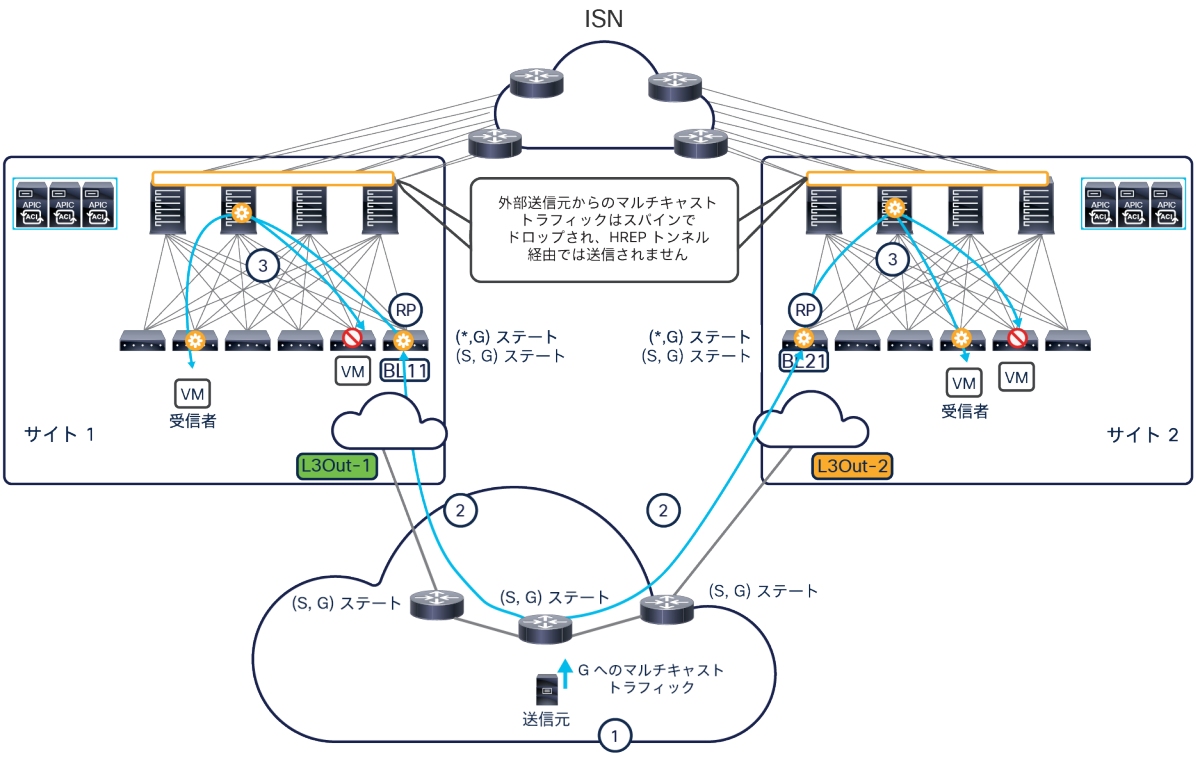


図 87.
内部受信者に向けたマルチキャストストリームのデータプレーン転送

- 送信元がマルチキャストグループ G に向けてトラフィックのストリーミングを開始します。
- FHR が両方のファブリックに向けてストリームを複製します。これは以前に (S,G) PIM-Join メッセージを受信しているためです。
- 各ファブリックの BL ノードがそのストリームを受信すると、VRF GIPo 宛ての VXLAN フレームにカプセル化します。これにより、トラフィックが各ファブリック内に配信され、内部受信者に到達します。

図 87 に示すように、スパインノードは、マルチキャストストリームをリモートサイトに転送しないようにプログラミングされています。これは、リモートサイトにある内部受信者が、重複したストリームを受信する可能性を回避するためです。このような動作の結果、外部送信元が発信したマルチキャストストリームをファブリック内の受信者が受信できるようにするには、アクティブなローカル L3Out 接続が不可欠となります。

注： PIM SSM を展開する場合も、図 84 と図 87 に示すようなデータパスの動作になります。唯一の違いは、SSM の場合、外部 RP を定義する必要がないことです。実際、SSM シナリオでは、受信者が IGMPv3 を使用して特定の送信元からのマルチキャストストリームに関する受信要求を宣言します。これにより、受信者と送信元の間に直接マルチキャストツリーが作成されます。

マルチキャストトラフィックに対するデータ プレーン フィルタリング

Cisco ACI リリース 5.0(1) より前は、マルチキャストトラフィックに対するコントロール プレーン フィルタリングのみがサポートされ、IGMP レポートフィルタ、PIM Join プルーニングフィルタ、RP フィルタの設定が可能でした。

Cisco ACI リリース 5.0(1) では、マルチキャストトラフィックに対するデータ プレーン フィルタリングの機能がサポートされています。このデータ プレーン フィルタリングは、ユーザー定義のルートマップ設定によって制御されます。この設定は直接、APIC（単一ファブリック展開の場合）で行うことも、MSO（マルチサイト展開の場合、Cisco Multi-Site Orchestrator リリース 3.0(1) 以降）で行うことも可能です。

データ プレーン フィルタリングはブリッジドメインレベルで適用され、フィルタリングの対象は以下のとおりです。

- 特定の送信元（または一連の送信元）からすべてのマルチキャストグループまたは特定の範囲の受信者のみに送られるトラフィック。一連の送信元とマルチキャストの範囲は、ブリッジドメインに適用されるルートマップで設定されます。
- 特定の受信者（または一連の受信者）が受信するトラフィック。これらのストリームを発信する送信元（または一連の送信元）とマルチキャストグループの範囲を指定することもできます。この場合も、受信者が接続されているブリッジドメインに適用されるルートマップで、該当するフィールドを設定します。

図 88 は、送信元を用いたフィルタリングの動作を示しています。

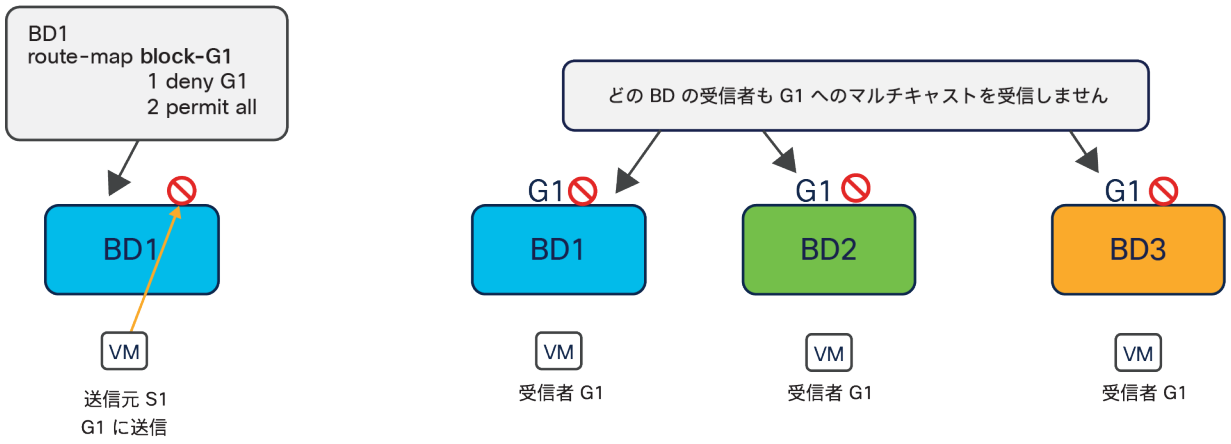


図 88.
マルチキャストに対する送信元を用いたデータ プレーン フィルタリング

このケースでは、送信元 S1 によって生成されグループ G1 に送られるマルチキャストトラフィックをドロップするよう設定されたルートマップが、（送信元 S1 が接続されている）BD1 に適用されています。その結果、（前のセクションで説明したコントロールプレーンでのアクティビティを実行することによって）グループ G1 に参加した受信者は、マルチキャストストリームを受信できなくなります。このフィルタリングが送信元と同じブリッジドメインに接続されている受信者にも適用されることに注意してください。これは、送信元と受信者が同じ ACI リーフに接続されているか、異なるリーフノードに接続されているかに依存しません。

図 89 は、データ プレーン フィルタリングが受信者のブリッジドメインに適用されている場合の機能的な動作を示しています。

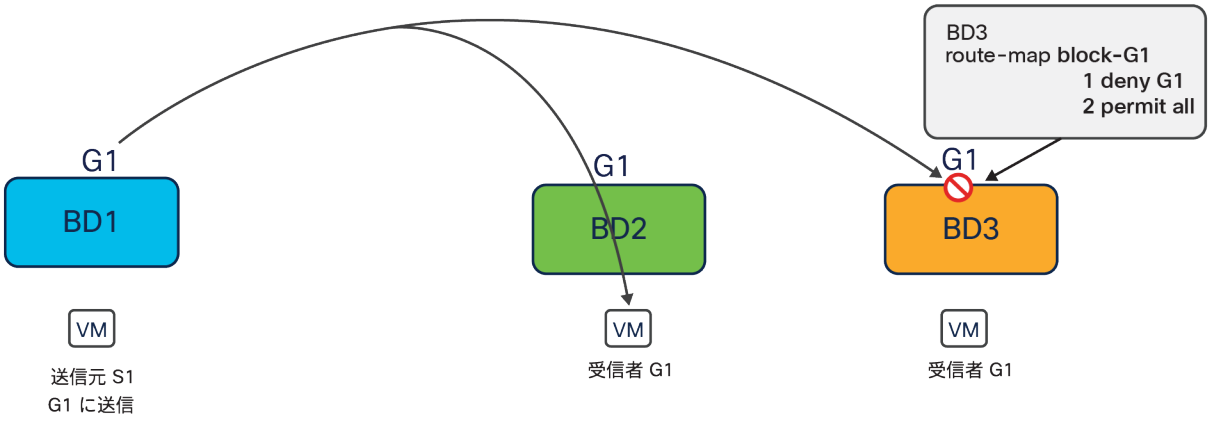


図 89.
マルチキャストに対する受信者を用いたデータ プレーン フィルタリング

この例では、S1 から G1 宛てのトラフィックが（ファブリック内またはサイト間で）転送され、受信者が接続されているリーフノードに到達しています。その時点で、データ プレーン フィルタリングを BD3 に適用すると、このブリッジドメインに接続された受信者はストリームの受信ができなくなります。一方、異なるブリッジドメインに接続された受信者は影響を受けません。

ルートマップを使用すると、コントロール プレーン フィルタリングを使用する場合よりもマルチキャストトラフィックのフィルタリングを柔軟に定義できるため、より複雑なユースケースに対応できます。ただし、Cisco ACI リリース 5.0(1) では、マルチキャストに対するデータ プレーン フィルタリングの機能に制約があります。以下はその一部です。

- IPv4 マルチキャストトラフィックのみがサポートの対象です。
- マルチキャスト フィルタリングは BD レベルで実行され、BD 内のすべての EPG に適用されます。そのため、同じ BD 内の異なる EPG に対して、異なるフィルタリングポリシーを設定することはできません。EPG レベルでさらに詳細にフィルタリングを適用する必要がある場合は、EPG を個別の BD に構成する必要があります。
- マルチキャスト フィルタリングは、Any-Source Multicast（ASM）の範囲に限定して使用することを目的としています。Source-Specific Multicast（SSM）は送信元フィルタリングではサポートされず、受信者フィルタリングでのみサポートされます。

APIC と NDO でマルチキャストに対するデータ プレーン フィルタリングを設定する方法の詳細は、以下のドキュメントを参照してください。

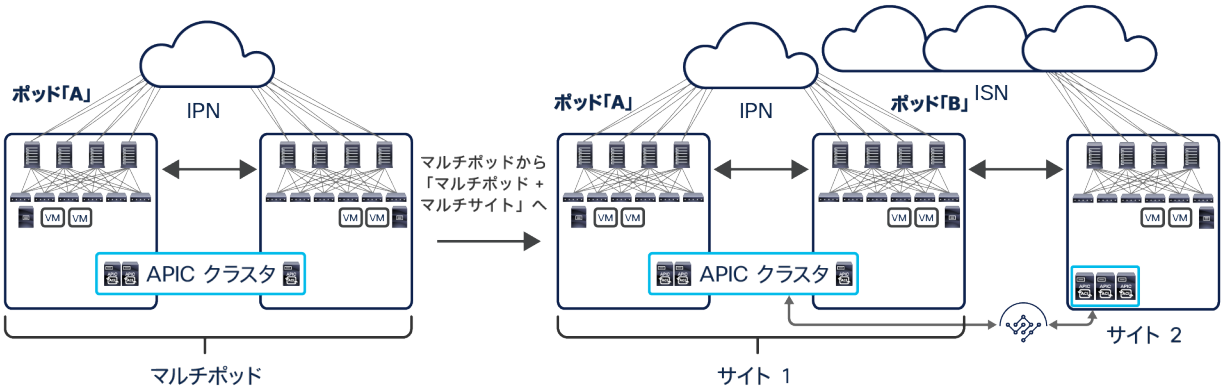
https://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/apic/sw/5-x/l2-configuration/cisco-apic-layer-2-networking-configuration-guide-50x/m_bridge.html#Cisco_Concept.dita_3254c6b4-f5a0-4542-bbe1-6868d1f8353b

https://www.cisco.com/c/ja_jp/td/docs/dcn/ndo/3x/configuration/cisco-nexus-dashboard-orchestrator-configuration-guide-aci-371/ndo-configuration-aci-use-case-multicast-37x.html#concept_i4j_lrx_5lb

Cisco ACI のマルチポッドとマルチサイトの統合

Cisco ACI リリース 3.2(1) では、Cisco ACI マルチポッドアーキテクチャと Cisco ACI マルチサイトアーキテクチャの組み合わせを可能にする「階層型」設計がサポートされます。この統合が役立つ主なユースケースは以下の 2 つです（図 90）。

ユースケース 1：Cisco Nexus Dashboard Orchestrator（NDO）で「サイト」としてマルチポッドファブリックを追加



ユースケース 2：単一ポッドファブリック（NDO に追加済み）をマルチポッドに変換

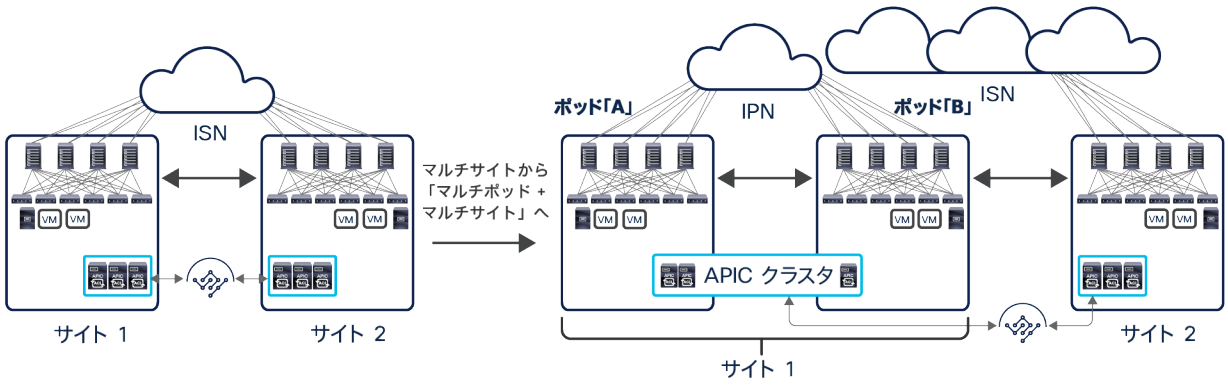


図 90.
Cisco ACI マルチポッドと Cisco ACI マルチサイトの統合に関するユースケース

1. 最初のシナリオは、Cisco ACI マルチポッドファブリックがすでに実稼働環境に展開されている状況で、1 つ（または 1 つ以上）の追加のファブリックを Cisco ACI マルチサイトを利用して既存のファブリックに接続する場合です。これは、Cisco ACI マルチポッドを（多くの場合、同じ場所または近接した場所に）展開してアクティブ/アクティブデータセンターを相互接続している状況で、ディザスタリカバリサイトとして機能するリモートデータセンターにそのマルチポッドファブリックを接続したい場合によく見られるユースケースです。

   注：この最初のユースケースでは、すでに展開されているマルチポッドファブリックが「サイト」として Cisco Nexus Dashboard Orchestrator に追加されると、NDO はすでに展開された構成に関する情報

を APIC から自動的に取得します。「インフラ」テナントにあるスパインを IPN に接続するための情報（インターフェイス、IP アドレス、OSPF 構成など）が対象です。

2. 2 番目のシナリオは、Cisco ACI マルチサイトが実稼働環境に展開され単一ポッドファブリックを相互接続している状況で、1 つ以上の単一ポッドファブリックをマルチポッドに変換する必要が生じた場合（たとえば、サポートされるリーフノードの総数の点から拡張が必要になった場合）です。

上図でわかるとおり、検討中のユースケースがどちらであっても、最終的な結果は Cisco ACI のマルチポッドとマルチサイトを組み合わせた「階層型」アーキテクチャです。以下のセクションでは、この展開オプションに関する技術上と設計上の考慮事項についてさらに詳しく説明します。

ポッドとサイトの間の接続

ポッド間ネットワーク（IPN）は、特定の要件（PIM-Bidir のサポート、DHCP リレーのサポート、MTU の拡大）を満たすことで、同じマルチポッドファブリックに属するさまざまなポッドを接続しています。サイト間ネットワーク（ISN）は、それより単純なルーテッド インフラストラクチャであり、さまざまなファブリックを相互接続するために必要なものです（必要なサポートは MTU の拡大のみ）。Cisco ACI マルチポッドを Cisco ACI マルチサイトと統合する場合、主要な展開オプションが 2 つあり、それぞれに独自の設計上の考慮事項があります。

- 単一のネットワーク インフラストラクチャで IPN 接続と ISN 接続の両方を実現（図 91）

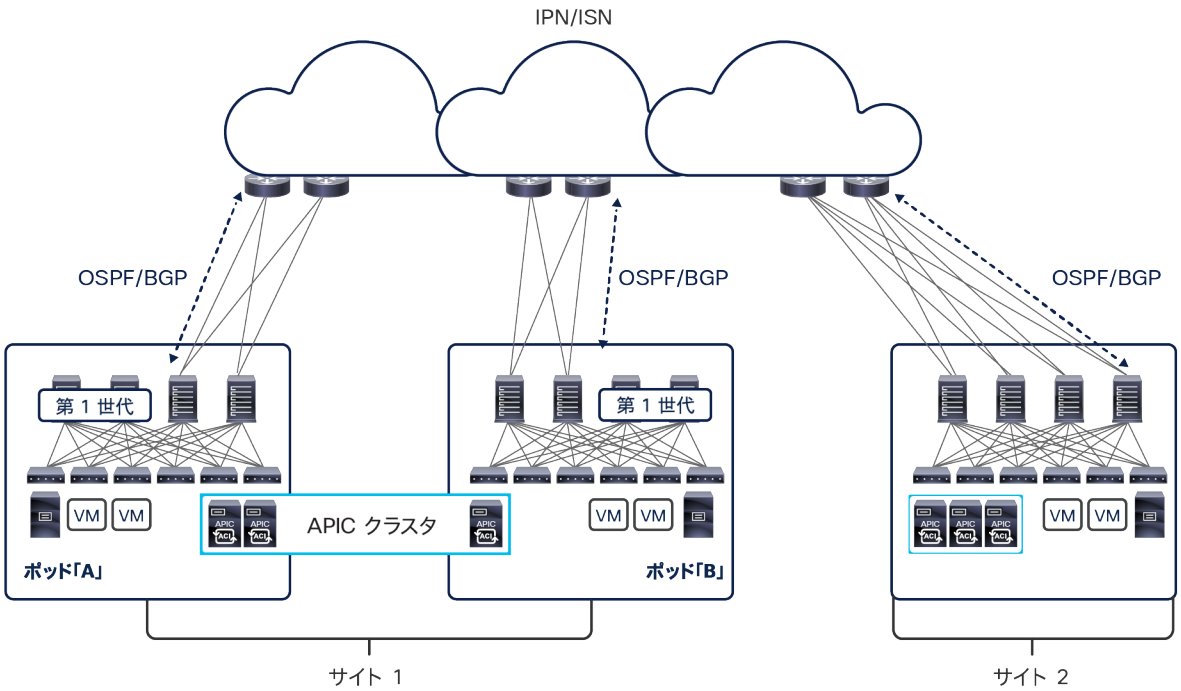


図 91.
単一のネットワーク インフラストラクチャで IPN 接続と ISN 接続の両方を実現

上図で最初に注目すべき重要なことは、第 1 世代のスパインモデル（Cisco Nexus 9336PQ、または第 1 世代のラインカードを装着したモジュラシャーシなど）は、Cisco ACI マルチポッドファブリック設計でサポートされていますが、Cisco ACI のマルチポッドとマルチサイトを組み合わせる場合には使用できないことです。第 2 世代の Cisco Nexus 9000 スパインスイッチ（Cisco Nexus 9332C/9364C、EX/FX 以降のラインカードを装着した Cisco Nexus 9500 モジュラモデル）のみを、リモートポッドやリモートサイトと通信する外部ネットワークに接続する必要があ

ります。また、マルチポッドファブリックの各ポッドをマルチサイトドメインのサイトとして追加する場合は、それぞれに少なくとも 1 つの第 2 世代スパイン（冗長性のためには 2 つ）を展開する必要があります。

注： 9332C および 9364C プラットフォームのようなモジュラ型でないスパインモデルの場合、ネイティブ 10G インターフェイス（SFP ベース）を使用して ISN デバイスに接続することもできます。

IPN 接続と ISN 接続のサービスの両方に同じネットワークを使用する場合、当然ながら、両方のタイプの水平方向トラフィックでスパインとファーストホップ IPN/ISN ルータの間の一連のリンクを共有することになります。物理リンクの数やその容量は、必要な水平方向通信の推定量に応じてスケールアップまたはスケールダウンが可能です。

各ポッドで、スパインを IPN/ISN に接続するすべての物理リンクを同じ「インフラ」L3Out 論理接続の一部として構成します。そうすることで、スパインと IPN/ISN ファーストホップルータの間で OSPF ピアリングを確立できます。

注： 「マルチサイトと GOLF L3Out 接続」セクションで詳しく説明するように、垂直方向通信のために GOLF L3Out を展開する場合は、さらに考慮が必要です。

IPN/ISN ファーストホップルータで学習された「アンダーレイ」ルートは、IPN/ISN ネットワーク内の異なるコントロール プレーン プロトコルに再配布できます。

- 個別のネットワーク インフラストラクチャで IPN 接続と ISN 接続を実現（図 92）

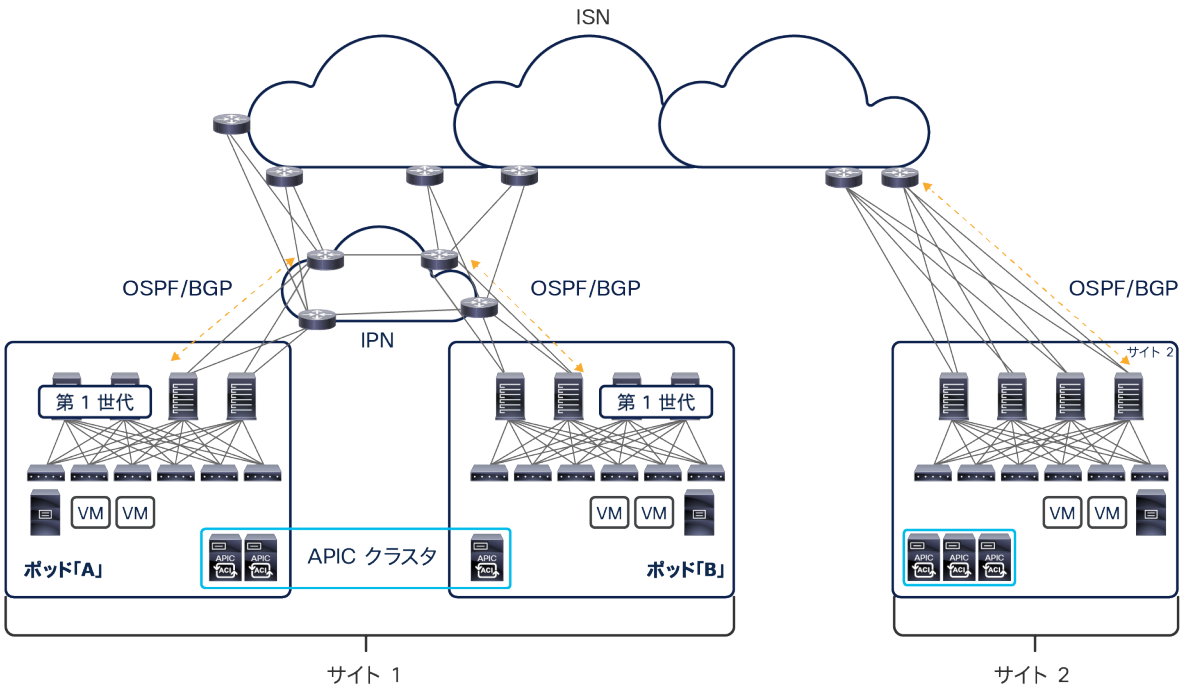


図 92.
個別のネットワーク インフラストラクチャで IPN 接続と ISN 接続を実現

Cisco ACI のマルチポッドとマルチサイトの通信に個別のネットワーク インフラストラクチャを展開し使用することはきわめて一般的です（つまり、IPN と ISN が物理的に分かれた 2 つのネットワークになります）。これは、マルチポッドが、同じ物理データセンターのロケーション（ポッドが、同じデータセンターの部屋またはホールに対応）または近接したロケーション（ポッドが、同じキャンパスまたは同じ大都市圏の個別のデータセンターに対応）に展開された Cisco ACI ネットワークを相互接続するために展開されることが多いためです。そのため、同じマルチポッドファブリックに属するポッドを相互接続するために、専用のダークファイバまたは高密度波長分割多重（DWDM）回線が頻繁に使用されます。一方、マルチサイトは、多くの場合、地理的に離れた場所に展開された Cisco ACI ファブリック間を接続するための手段です。したがって、ISN サービスを提供するために個別の WAN ネットワークが使用されます。

IPN と ISN に個別のインフラストラクチャを使用しているにもかかわらず、スパインは、共通の一連のリンクを使用して外部ルータと OSPF または BGP のピアリングを確立し、マルチポッドとマルチサイトの両方の水平方向トラフィックフローを処理する必要があります。そのため、IPN と ISN のネットワークを相互に接続する必要があります。スパインと IPN/ISN の間で個別の接続を使用して 2 つのタイプの通信を処理することはできません（図 93）。

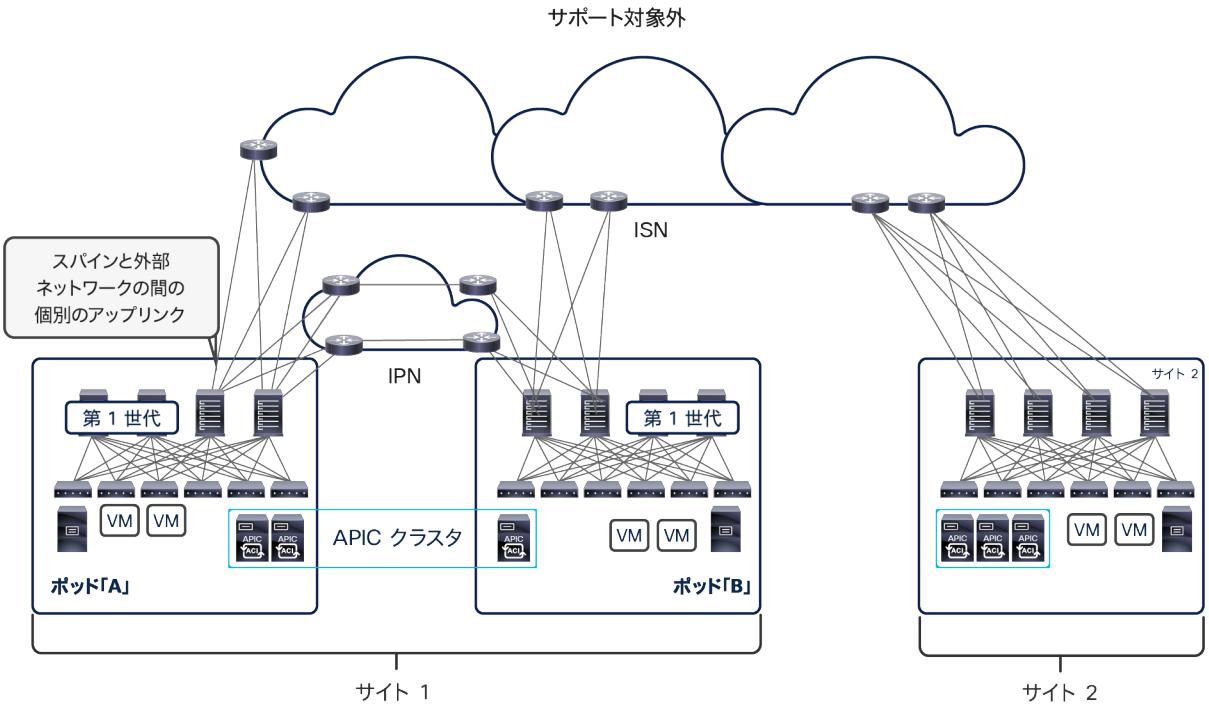


図 93.
スパインと外部ネットワークの間で個別のアップリンクを使用

コントロールプレーンに関する考慮事項

MP-BGP EVPN は、Cisco ACI のマルチポッドとマルチサイトのアーキテクチャで共通して使用される機能です。2 つの設計を組み合わせる場合、階層型ピアリングモデルを導入し、別々のサイトに属するスパインノード間に作成された EVPN 隣接関係を最小限に抑えます。

この階層型ピアリングモデルをサポートするために、各スパインノードが以下の 2 つのロールのいずれかを担います。

- MP-BGP EVPN スピーカー：スピーカーとして構成されたすべてのスパインノードが、リモートサイトに展開されたスピーカーと EVPN 隣接関係を確立します。スピーカーのロールは、Cisco Nexus Dashboard Orchestrator を利用して（GUI または REST API 呼び出しを介して）スパインに明示的に割り当てる必要があることに注意してください。
- MP-BGP EVPN フォワーダ：スピーカーとして明示的に構成されていないすべてのスパインがフォワーダになります。フォワーダは、同じファブリックに展開されたすべてのスピーカーと EVPN 隣接関係を確立します。

注： BGP スピーカーのロールは、マルチサイト対応スパイン（つまり、第 2 世代のハードウェアモデルである EX 以降）に対してのみ有効にできます。一方、暗黙のフォワーダロールは、第 1 世代を含むすべてのスパインに割り当てることができます。

下の図 94 は、Cisco ACI のマルチポッドとマルチサイトを統合した導入モデルで確立された EVPN 隣接関係を示しています。

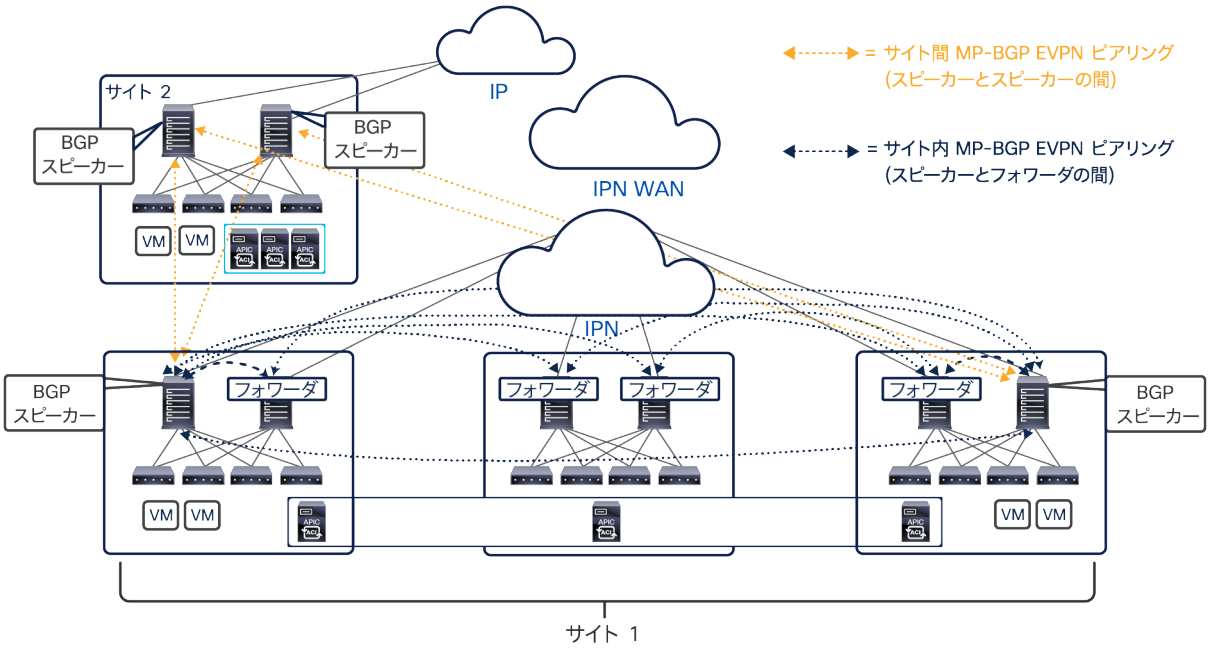


図 94.
階層型 MP-BGP EVPN ピアリング

注： MP-BGP EVPN 隣接関係は、同じマルチポッドファブリックの別々のポッドに展開された 2 つのスピーカー間でも確立されます。

上図のように、冗長性を確保するために、各サイトに 2 つの MP-BGP EVPN スピーカーを展開することを常にお勧めします。サイトが単一ポッドとして展開されているかマルチポッドとして展開されているかによって、2 つのスピーカーを同じポッドに展開することも（たとえば、図 94 のサイト 2）、ポッドにまたがって展開することも（たとえば、同じ図 94 のサイト 1）できます。

スピーカーとフォワーダの間の EVPN 隣接関係により、ローカルスピーカーがリモートサイトのスピーカースパインから受信したエンドポイント情報をフォワーダが確実に学習できるようになり、地理的な EVPN 隣接関係の必要総数を低く抑えられます。

注： 各スパインノードで同じループバック インターフェイスを構成して、同じマルチポッドファブリックに属する他のポッドのスパインとリモートサイトのスパインの両方と EVPN ピアリングを確立できます。これは、前述の両方のユースケース、すなわち、既存のマルチポッドファブリックをマルチサイトドメインに追加する場合と、すでにマルチサイトドメインに属している単一ポッドファブリックをマルチポッドファブリックに拡張する場合の両方に適用できます。TEP 構成（すなわち、VXLAN カプセル化トラフィックを受信するために使用される IP アドレス）の観点では、一意のアドレスを展開してマルチポッドとマルチサイトのトラフィックフローに使用することをお勧めします。

データプレーンに関する考慮事項

「Cisco ACI マルチサイトのオーバーレイデータプレーン」セクションで説明したように、サイト間でさまざまなタイプの水平方向通信を確立できます。いずれの場合も、VXLAN データプレーンカプセル化を使用して、サイト間ネットワークに必要な構成と機能を簡素化します。これは、Cisco ACI のマルチポッドとマルチサイトを統合する場合にも有効ですが、以下のセクションで説明するように、設計に関してさらに考慮が必要な点がいくつかあります。

レイヤ 2 とレイヤ 3 のユニキャスト通信

エンドポイントの到達可能性情報がサイト間で交換され、ARP 交換も正常に完了したと仮定すると、リーフノードとスパインノードでは、レイヤ 2、レイヤ 3、COOP の各テーブルが適切に設定され、サイト間ユニキャスト トラフィック フローが確立できるようになっているはずです（図 95）。

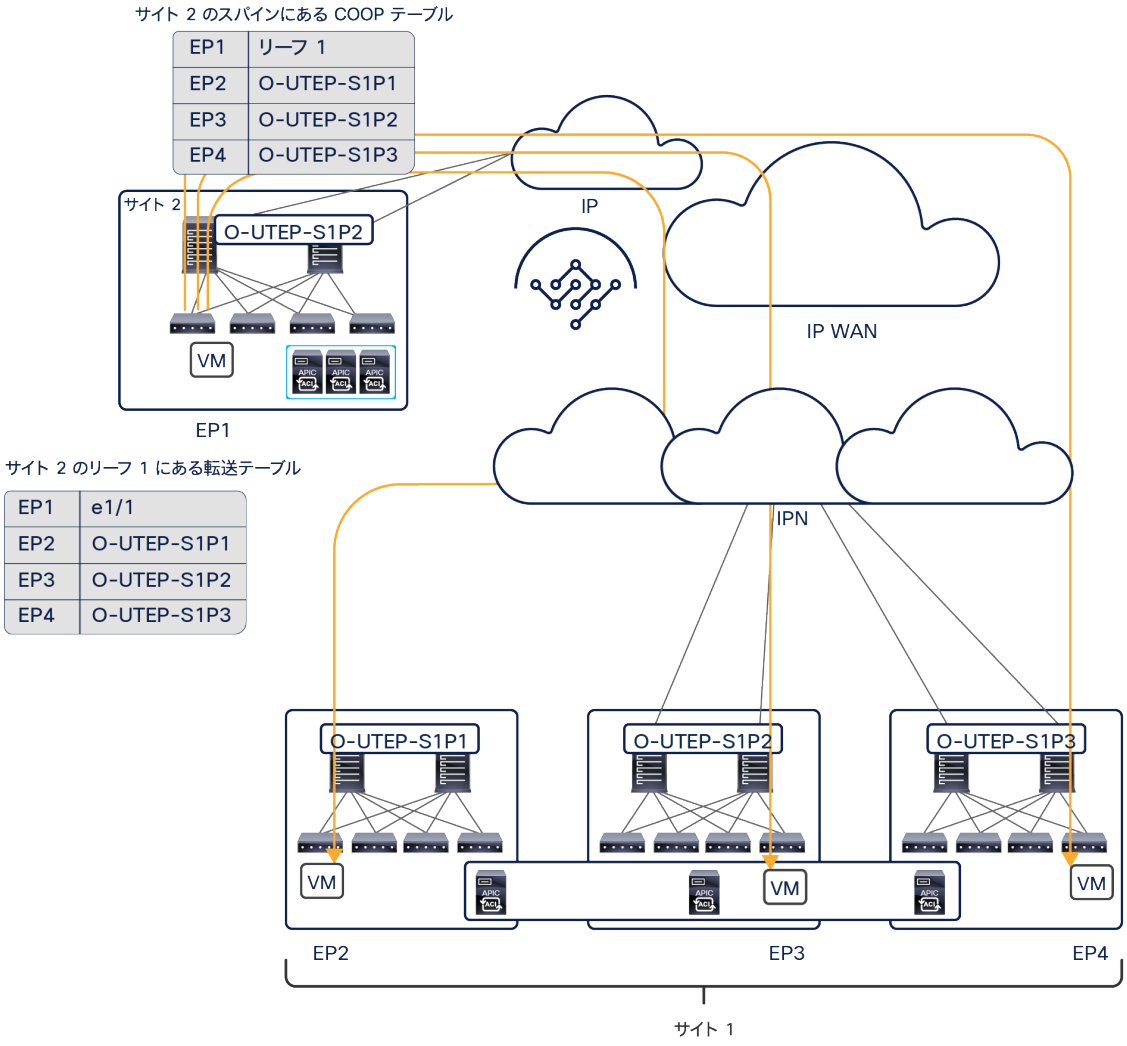


図 95.
Cisco ACI マルチサイトに属するマルチポッドファブリックに入力されるユニキャストトラフィック

Cisco ACI マルチポッドファブリックがマルチサイトアーキテクチャに統合されると、各ポッドには一意のオーバーレイユニキャスト TEP アドレス（O-UTEP）が割り当てられ、そのポッドに展開されたすべてのマルチサイト対応スパインに関連付けられます。これにより、リモートサイトから受信したレイヤ 2 またはレイヤ 3 のトラフィックを、宛先エンドポイントが接続されているポッドに直接ステアリングできるようになります。

注： ポッド間通信に使用されるデータプレーン TEP は、ファブリック間通信に使用される O-UTEP アドレスとは異なる必要があります。

一方、図 96 は、レイヤ 2 BUM トラフィックを、サイトとして展開されたマルチポッドファブリックにリモートサイトから送信する方法を示しています。

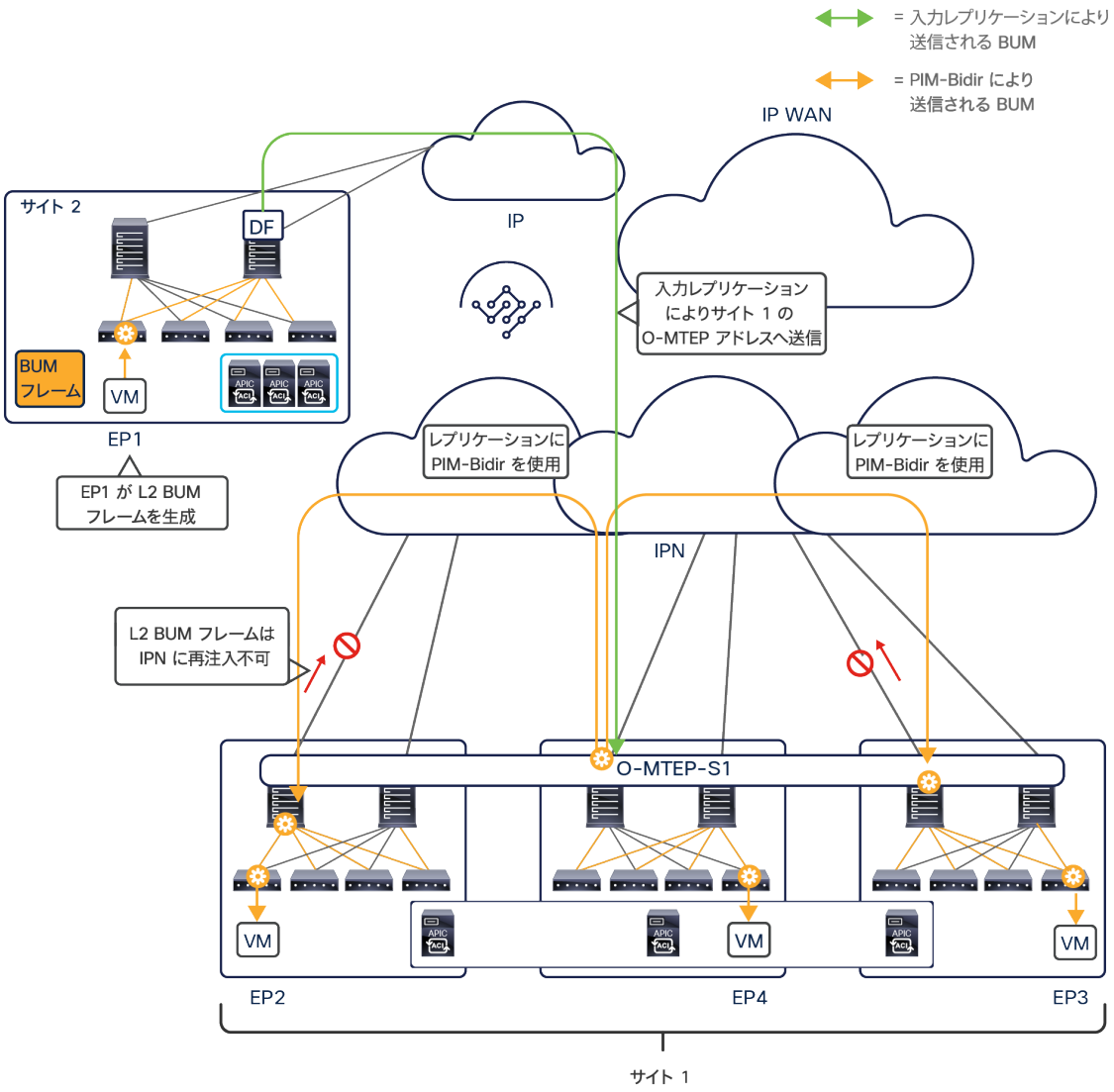


図 96.
Cisco ACI マルチサイトに属するマルチポッドファブリックに入力されるレイヤ 2 BUM トラフィック

ユニキャストのシナリオとは異なり、同じマルチポッドファブリックに属するすべてのマルチサイト対応スパインに（つまり、展開されたすべてのポッドにまたがって）単一のオーバーレイマルチキャスト TEP アドレスが関連付けられます。そのため、外部バックボーンネットワークで利用可能なルーティング情報のみに基づいて、L2 BUM の入

力トラフィックを任意のポッドのスパインに配信できます。受信したスパインは、以下の 2 つの操作を実行する必要があります。

- トラフィックが属するブリッジドメインに関連付けられた FTAG ツリーに従って、ポッド内で BUM トラフィックを転送
- IPN を通して BUM トラフィックを他のポッドに転送

他のポッドのスパインが BUM トラフィックを受信する場合も、ポッド内でローカルに転送することができます。ただし、トラフィックの重複を避けるため、トラフィックを IPN に再注入することは許可されません。

サイト間での TEP プールプレフィックスのフィルタリング

これまでに説明したことから、すべてのサイト間データプレーン通信がオーバーレイ TEP アドレス（ユニキャストとマルチキャスト）を宛先として行われていることが明確になったと思われます。そのため、ファブリックを起動するために APIC に最初に割り当てられた TEP プールの範囲は、このサイト間通信で何の役割も果たしません。それでも、内部 TEP プールはスパインから外部ネットワークにアドバタイズされます。図 97 に示すように、これは主に Cisco ACI のマルチポッドとマルチサイトの統合を可能にするためです。

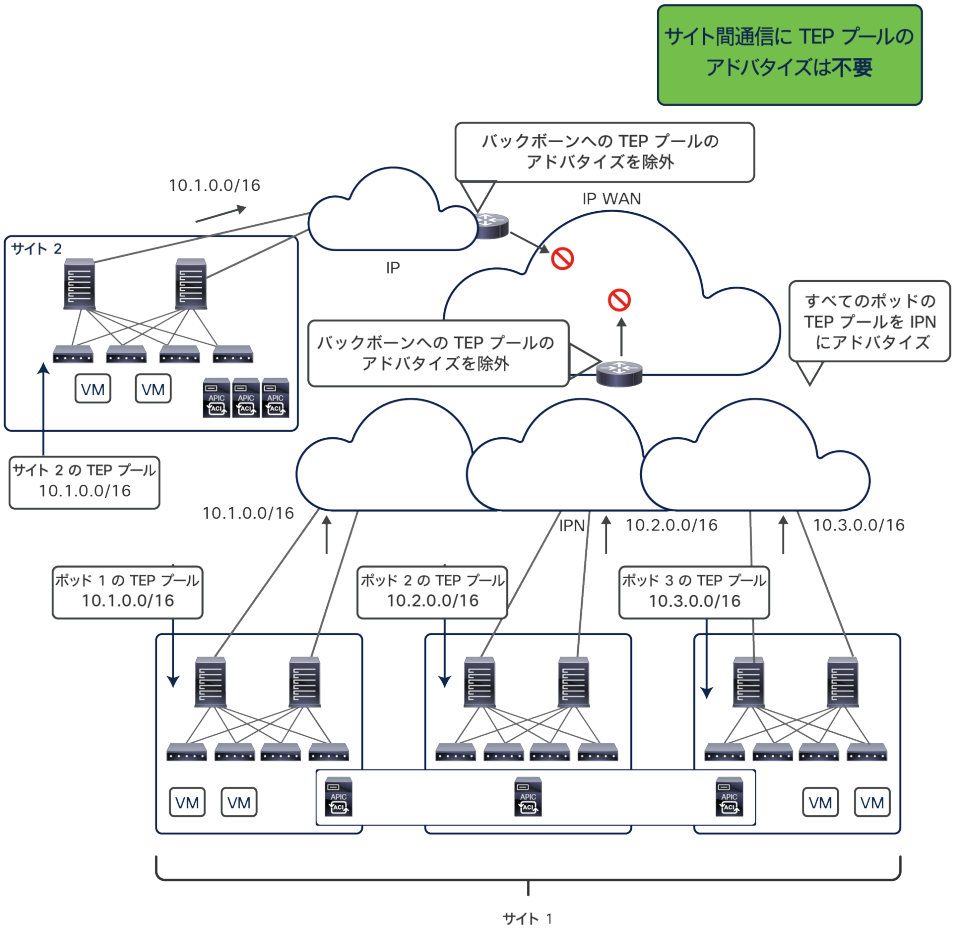


図 97.
各ポッドからアドバタイズされる TEP プールとそのバックボーンでのフィルタリング

同じマルチポッドファブリックに属する異なるポッドに所属のエンドポイント間の通信は、エンドポイントが接続されたリーフノード間に直接 VXLAN トンネルを作成することによって確立されます。そのため、同じファブリックに属する異なるポッド間で TEP プールを交換する必要があります。

異なるファブリック間の TEP プール情報の交換は必要ないだけでなく、複数のサイトで同じ TEP プール（10.1.0.0/16）が定義されている上図のようなシナリオでは、実際に問題が発生する可能性があります。したがって、TEP プール情報をフィルタリングしてネットワークのバックボーンに注入されないようにすることが、ベストプラクティスの推奨事項となります。

同様に、ポッド間でのレイヤ 2 BUM トラフィックの転送には、PIM Bidir が使用されます。サイト間 BUM 転送には、入力レプリケーションが使用されます。そのため、異なるサイトを相互接続するネットワークのバックボーンで、これらのマルチキャストグループに対して PIM を有効にしないことをお勧めします。これは、同じ Cisco ACI マルチサイトアーキテクチャに複数のマルチポッドファブリックが属しているシナリオで、転送に問題が生じることを防ぐためです。

外部レイヤ 3 ドメインへの接続

Cisco ACI ファブリック内で定義された VRF インスタンスを外部レイヤ 3 ネットワークドメインに接続するには、L3Out 接続と呼ばれる論理接続を作成する必要があります。L3Out 接続の構成には、以下の 2 つの設計オプションが用意されています。

- ボーダーリーフノードで定義された L3Out 接続
- EVPN ベースの L3Out 接続（「GOLF」設計オプションとも呼ばれます）

図 98 は、ボーダーリーフノードで定義された L3Out 接続を示しています。このオプションは、初期の ACI からサポートされています。

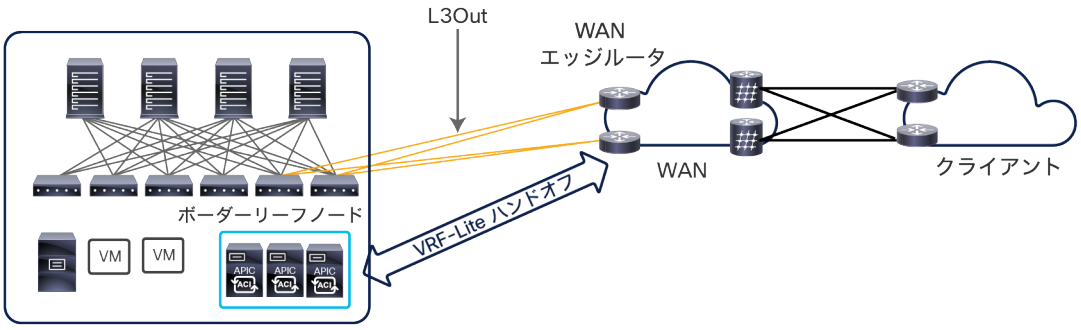


図 98.
ボーダーリーフノードでの L3Out 接続

このアプローチでは、VRF-Lite を構成することで、各 VRF インスタンスの接続を WAN エッジルータに拡張します。Cisco ACI ファブリック内で使用されるデータプレーン VXLAN カプセル化は、トラフィックが WAN エッジルータに送信される前にボーダーリーフノードで終了します。このアプローチでは「共有 L3Out」モデルもサポートされます。単一の L3Out（通常は「共通」テナント構成の一部として定義されます）を使用して、ファブリック内で定義され異なる VRF やテナントに属するブリッジドメインを外部に接続することができます。

注： ボーダーリーフ L3Out 接続の詳細は、https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/guide-c07-743150.html にあるドキュメントを参照してください。

図 99 は、Cisco ACI リリース 5.0(1) 以降で利用できる、ボーダーリーフノードでの SR-MPLS/MPLS ハンドオフ機能を示しています。

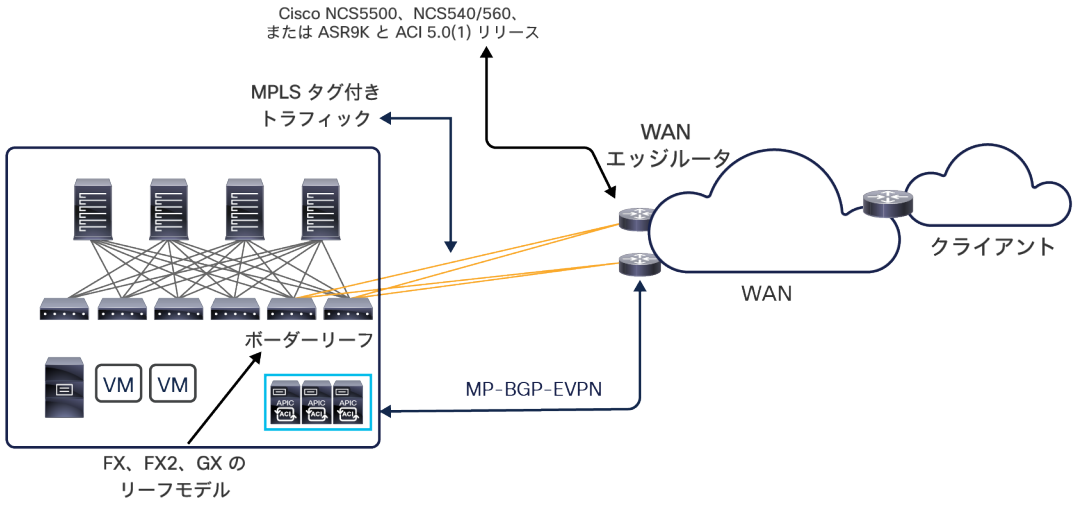


図 99.
ボーダーリーフノードでの SR-MPLS/MPLS ハンドオフ

このアプローチでは、従来の VRF-lite モデルの特徴であった VRF ごとのルーティングピアリングの代わりに、ボーダーリーフノードと DC-PE ルータの間で単一の MP-BGP EVPN コントロールプレーンが確立されます。この単一のコントロールプレーンにより、外部ネットワークドメインとの接続を必要とする、ファブリック内で定義されたすべての VRF でコントロールプレーン情報の交換が可能になります。データプレーンの観点から見ると、802.1q タグを MPLS タグに置き換えることで、異なる VRF に属するレイヤ 3 通信を論理的に分離しています。

BL ノードの SR-MPLS/MPLS ハンドオフの構成は、単一ファブリック展開の場合、APIC でプロビジョニングできます。マルチサイトドメインに属するファブリックの場合は、さらに Cisco Nexus Dashboard Orchestrator で直接公開されます。これを利用すれば、マルチサイトドメインに属するすべてのファブリックに対して垂直方向接続を一元的にプロビジョニングできます。また、それだけではなく、サイト間レイヤ 3 通信を外部 MPLS 対応コアを介して処理できるようになり、ネイティブのマルチサイト VXLAN データパスが利用されなくなります。サイト間レイヤ 3 通信について説明した際に述べたように、SR-MPLS/MPLS ハンドオフ機能を備えた L3Out パスにはメリットがあります。すなわち、SR-MPLS TE ポリシーを ACI ネットワーク間の水平方向トラフィックフローに適用できます。また、WAN チームがデータセンター間通信のモニタリングに、使い慣れたツールを活用できます。

SR-MPLS ハンドオフを展開して外部ネットワークとのピアリングを簡素化することで、マルチ VRF ルーティング情報を交換（つまり、VRF-lite の展開でよく見られる VRF モデルごとのルーティングプロトコルではなく、単一の MO-BGP EVPN コントロールプレーンを展開）し、垂直方向接続を確立しながら、VXLAN を活用した ISN 経由の水平方向通信も維持したい場合があります。そのために、NDO 4.0(2) では、SR-MPLS L3Out ハンドオフを従来の IP ベースの L3Out ハンドオフと同じ方法で処理できる機能が導入されています。これについては、以下のセクションで説明します。

注： SR-MPLS/MPLS ハンドオフの展開についての詳細は、https://www.cisco.com/c/ja_jp/td/docs/dcn/ndo/3x/configuration/cisco-nexus-dashboard-orchestrator-configuration-guide-aci-371/ndo-configuration-aci-use-case-sr-mpls-37x.html にあるドキュメントを参照してください。

図 100 は、EVPN ベースの L3Out（GOLF）を使用する設計オプションを示しています。GOLF アプローチは当初、外部レイヤ 3 ネットワークドメインに接続する VRF インスタンスの数を拡大するために導入されました（GOLF の Cisco ACI への統合以降、1,000 個の VRF インスタンスがサポートされています）。

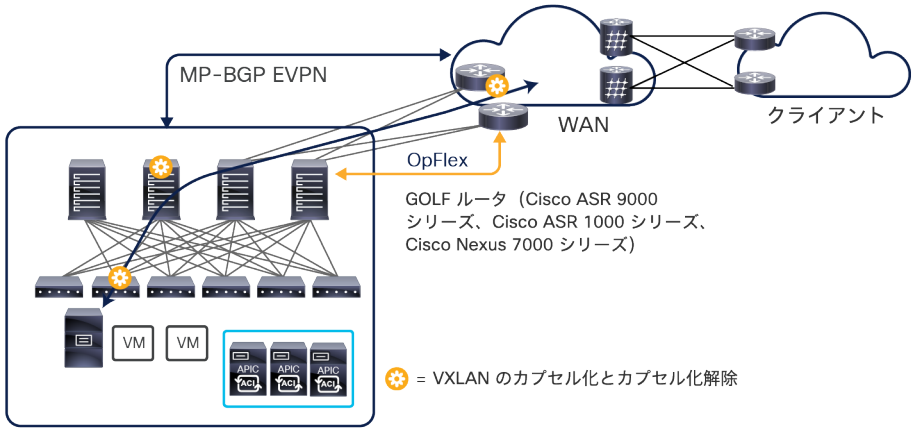


図 100.
GOLF L3Out 接続

GOLF の統合によって、ボーダーリーフノードを介した WAN エッジルータへの接続はできなくなりました。これらのルータは現在、スパインノードに（直接的または間接的に）接続されます。先ほど説明した SR-MPLS ハンドオフと同様に、MP-BGP EVPN コントロールプレーンによって、外部接続を必要とするすべての ACI VRF でルート情報の交換が可能になります。また、OpFlex コントロールプレーンによって GOLF ルータでファブリックにかかわる VRF の構成が自動化され、最終的に VXLAN データプレーンによる垂直方向通信が可能になります。

注： GOLF についての詳細は、https://www.cisco.com/site/jp/ja/products/networking/cloud-networking/application-centric-infrastructure/index.html にあるドキュメントを参照してください。

https://www.cisco.com/c/ja_jp/td/docs/switches/datacenter/aci/apic/sw/2-x/L3_config/b_Cisco_APIC_Layer_3_Configuration_Guide/b_Cisco_APIC_Layer_3_Configuration_Guide_chapter_010010.html

以下の 2 つのセクションでは、ボーダーリーフ L3Out（IP ベースおよび SR-MPLS ハンドオフ）と Cisco ACI マルチサイトアーキテクチャの統合について詳しく説明します。GOLF L3Out の展開は、新しい実稼働のユースケースでは推奨されなくなったため（既存の展開では引き続き完全にサポートされます）、関連する内容はこのホワイトペーパーの付録 C に移動されました。

Cisco ACI マルチサイトとボーダーリーフノードでの L3Out 接続

ボーダーリーフノードに IP ベースの L3Out 接続を展開する場合、図 101 に示す 2 つのシナリオが利用できます。これらは、Cisco ACI マルチサイトアーキテクチャの初期のリリースから完全にサポートされています。

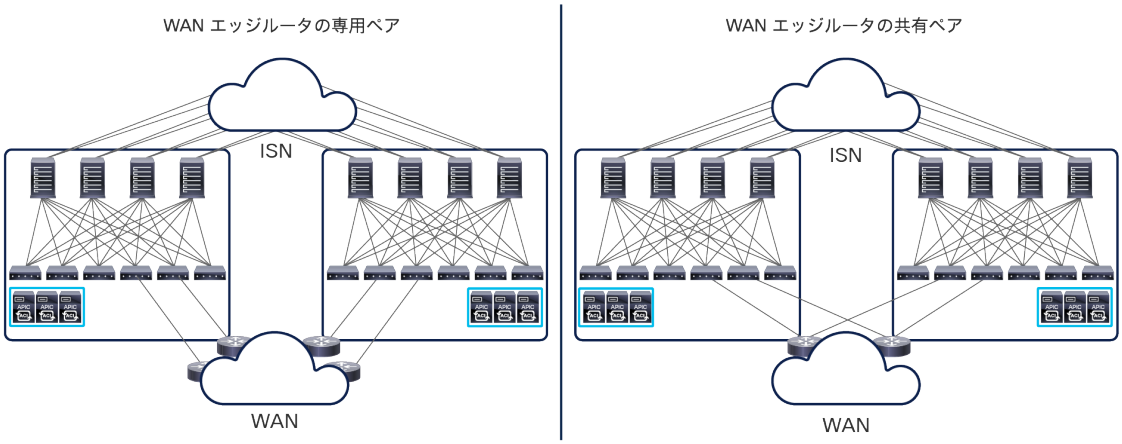


図 101.
WAN エッジルータの専用ペアまたは共有ペア

注： NDO 4.0(2) 以降は、SR-MPLS L3Out ハンドオフでも同じモデルがサポートされています。

左側のシナリオでは、異なる Cisco ACI ファブリックが別々の物理データセンターのロケーションにマッピングされます。これは、大部分のマルチサイト展開で最も一般的なシナリオです。この場合、通常は、WAN エッジルータの専用ペアを使用して各ファブリックを外部 WAN ネットワークに接続します。右側のシナリオでは、Cisco ACI マルチサイト設計を使用して、同じ地理的ロケーションに展開されたファブリックを相互接続します。この場合、一般的には、WAN エッジルータの共有ペアを使用して WAN に接続します。どちらの場合も、Cisco ACI リリース 4.2(1) より前では、常に各サイト（つまり、各 APIC クラスタ）に別々の L3Out 論理接続を展開する必要があります。言い換えると、下の図に示すように、サイトに展開されたエンドポイントは、ローカル L3Out 接続を介してのみ外部ネットワークドメインと通信できます。

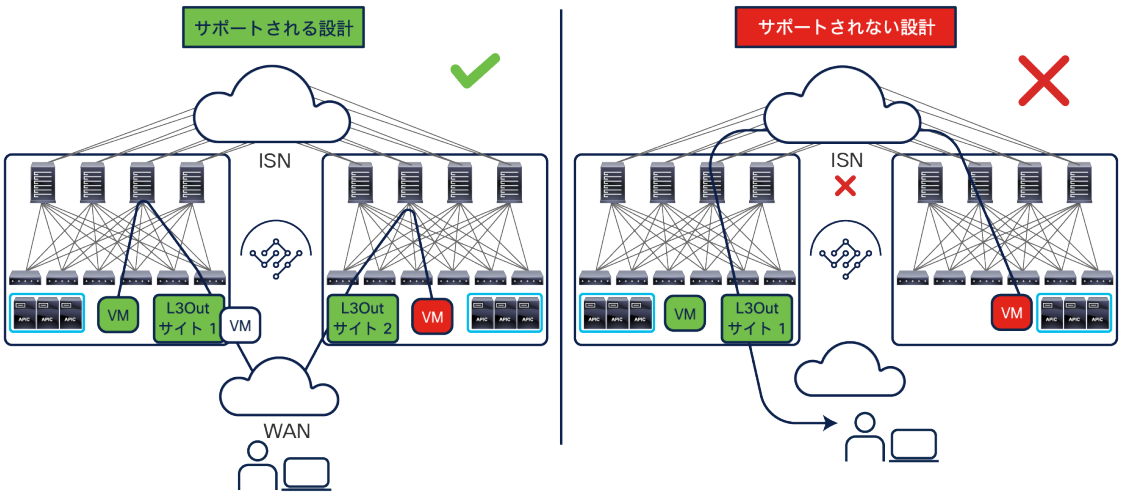


図 102.
Cisco ACI マルチサイトと L3Out 接続（リリース 4.2(1) より前）

Cisco ACI リリース 4.2(1) では、「サイト間」L3Out と呼ばれる新機能が導入され、上記の制限が取り除かれています。サイト間 L3Out の詳細とユースケースについては、「サイト間 L3Out 機能の導入（Cisco ACI リリース 4.2(1)/MSO リリース 2.2(1) 以降）」セクションを参照してください。このセクションの残りの部分の説明では、サイトごとにローカル L3Out が少なくとも 1 つ展開されていることを想定しています。

このセクションで後述するように、各 APIC ドメインで定義された L3Out オブジェクトは Cisco Nexus Dashboard Orchestrator に公開され、NDO で直接定義できる外部 EPG に関連付けられます。

上の図 101 では、2 つの別々のネットワーク インフラストラクチャが使用されています。1 つは異なるファブリックに接続されたエンドポイント間のすべての水平方向通信に使用されるレイヤ 3 サイト間ネットワーク（ISN）、もう 1 つはリモートクライアントへの垂直方向接続を確立するために使用される WAN ネットワークです。これが一般的な導入モデルです。ただし、図 103 に示すように、両方の目的に同じ WAN インフラストラクチャを使用することもできます。

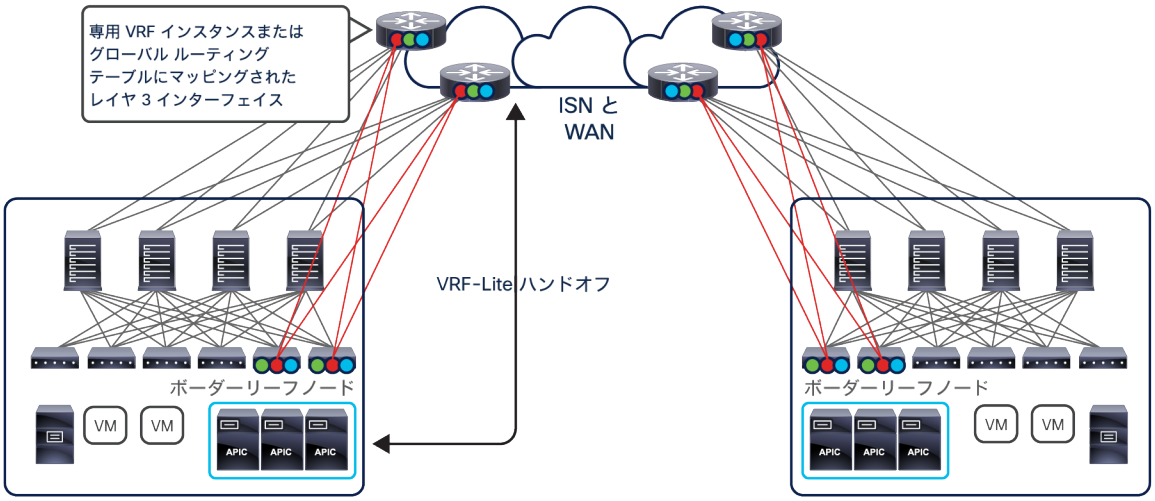


図 103.
水平方向と垂直方向のトラフィックに共通のレイヤ 3 インフラストラクチャを使用

このシナリオでは、WAN エッジルータの同じペアが以下の 2 つの目的を果たします。

- ボーダーリーフノードに接続して、各 Cisco ACI ファブリック内に展開された VRF インスタンスに外部接続を提供（垂直方向のトラフィックフロー）：IP ベースの L3Out を利用する場合、各 VRF インスタンスのルーティング情報を交換するために、VRF-lite が使用されます。一方、SR-MPLS L3Out を利用する場合、MP-BGP EVPN が使用されます。どちらの場合も、通常、異なる VRF インスタンスが WAN エッジルータにも展開されます。

  注：例外は、異なる Cisco ACI VRF インスタンスが外部 WAN に向けられた共通のルーティングドメインに接続するユースケースです。このシナリオでは、WAN エッジルータで定義する必要があるのは共有 VRF（グローバル ルーティング テーブルの場合もあります）のみです（共有 L3Out の展開オプション）。この「共有 L3Out」モデルは、Cisco ACI リリース 4.0(1) 以降、Cisco ACI マルチサイトでサポートされています。

- スパインノードに接続して、サイト間で交換される VXLAN トラフィックのルーティングをサポート（水平方向のトラフィックフロー）：この通信は、WAN ネットワークの内部の専用 VRF インスタンス（ベストプラクティスの推奨事項）またはグローバル ルーティング テーブルのいずれかで発生します。

図 104 は、IP ベースの L3Out を展開するシナリオで、別々の Cisco ACI ファブリックに接続されたエンドポイントに外部接続を提供するために必要な一連のステップを示しています。この例では、Web エンドポイントが異なるブリッジドメインに属し、そのブリッジドメインが各サイトでローカルにのみ定義されていることに注意してください（つまり、Web EPG は別々のアプリケーションスタックに属していると見なすことができます）。

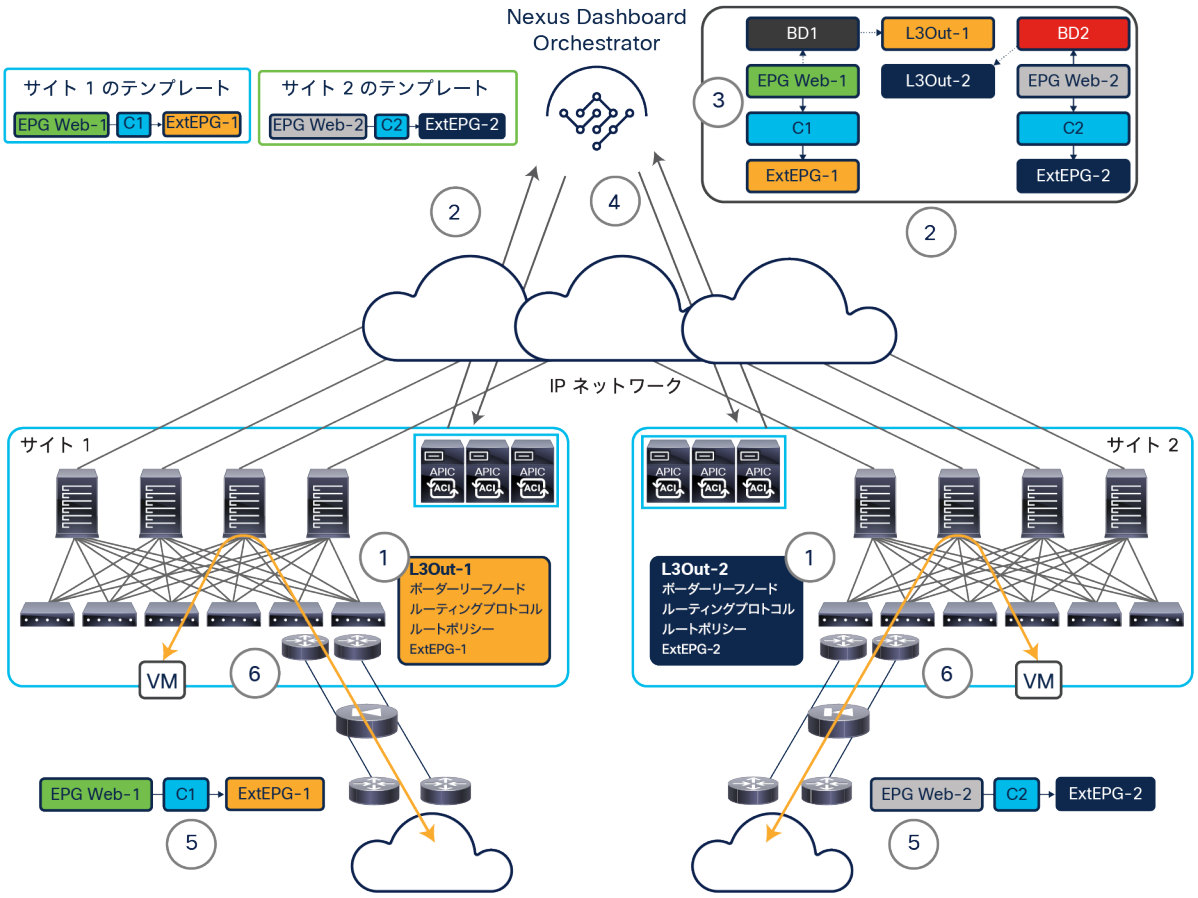


図 104.
マルチサイトとボーダーリーフノードでの IP ベースの L3Out 接続

- 最初に、Web ブリッジドメインを外部レイヤ 3 ドメインに接続するために使用する L3Out 構成を作成します。このステップでは、ボーダーリーフノードとそのインターフェイス、および外部デバイスとのピアリングに使用するルーティングプロトコルを、Cisco Multi-Site Orchestrator ではなく APIC レベルで（したがって、サイトごとに個別に）定義します。Cisco Multi-Site Orchestrator リリース 2.2(1) 以降では、L3Out オブジェクトを Orchestrator に直接作成し、L3Out が定義されたテンプレートに関連付けられた 1 つ以上のサイトに展開することもできます。ただしその場合でも、論理ノードや論理インターフェイスなどの構成は、各ファブリックの APIC レベルで行います。また、ストレッチテンプレートで L3Out を構成するのではなく、サイト固有のテンプレートで個別の L3Out オブジェクトを定義することをお勧めします。異なるサイトの L3Out を明確に識別して差別化できるようになるため、たとえば、内部 BD サブネットをアドバタイズする範囲の制御を強化できます。この点は、「サイト間 L3Out 機能の導入（Cisco ACI リリース 4.2(1)/MSO リリース 2.2(1) 以降）」セクションで説明します。

注：L3Out 接続は、常に VRF インスタンスに関連付けられます。したがって、この例で説明するシナリオでは、サイトにまたがって拡張されたこのような VRF を Orchestrator で作成した後、ローカルでの L3Out の作成が可能になります。

- 次に、NDO で外部 EPG オブジェクトを定義し、それらを各 L3Out に関連付けます。こうすることで、内部 EPG に対して固有の接続要件を構成できるようになります。
- Cisco Nexus Dashboard Orchestrator でテンプレートを 2 つ定義し、それぞれに、外部ネットワークドメインにアクセスする必要がある Web EPG を規定するアプリケーション ネットワーク プロファイルを記述します。アクセスには、すでに各サイトにプロビジョニングされた L3Out 接続が使用されます。それぞれのテンプレートで、Web EPG と、ローカル L3Out に関連付けられた外部 EPG の間のコントラクトを作成すると、定義が完了します。作成されたテンプレートを、プッシュ先のサイトに個別に関連付けます（図 104 では、1 つ目のテンプレートはサイト 1 に関連付け、2 つ目のテンプレートはサイト 2 に関連付けます）。

  注：別のアプローチとして、両方のサイトにマッピングされたテンプレートで単一の外部 EPG（Ext-EPG）を定義することもできます。定義したストレッチ Ext-EPG は、サイトレベルで各 L3Out にマッピングできます。このようなアプローチは、両方のサイトの L3Out が同じ外部ネットワークリソースへのアクセスを提供する場合に効果的です。Ext-EPG 構成（および関連するセキュリティポリシー）の定義がたった 1 回で済むためです（図 105）。

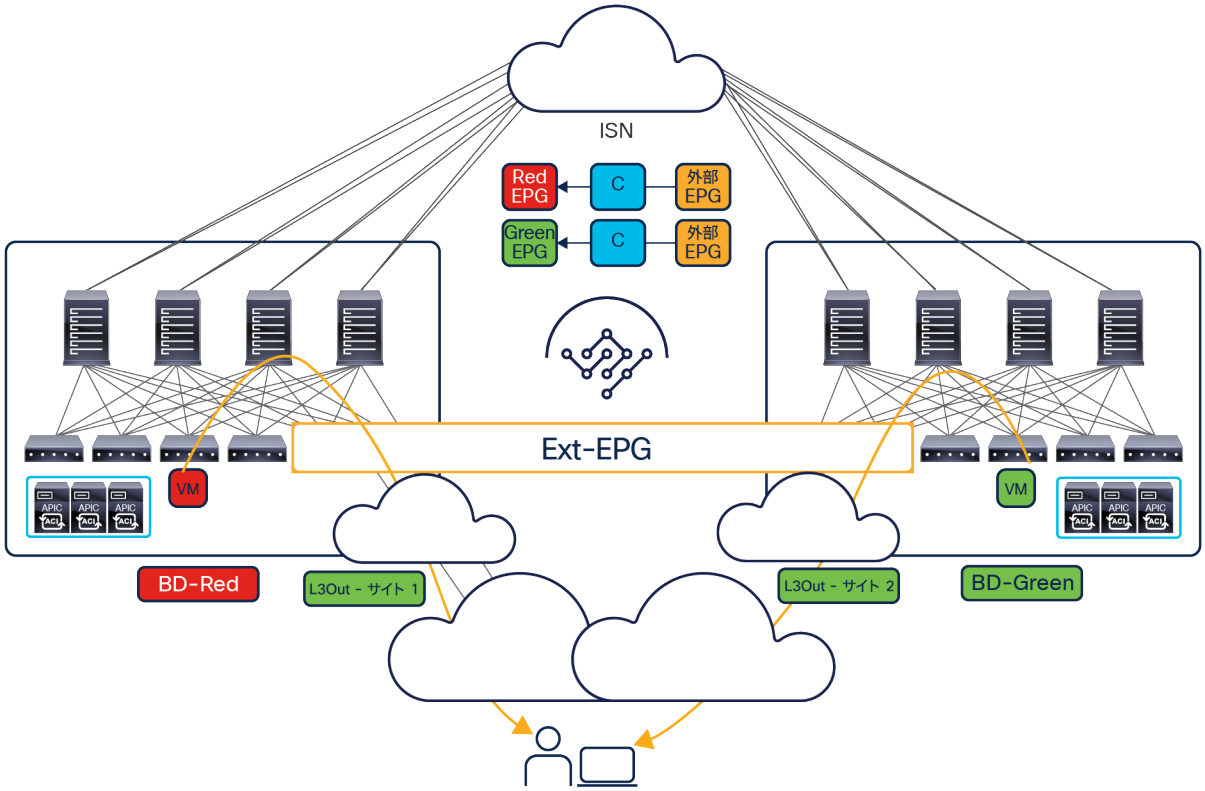


図 105.
ストレッチ外部 EPG（Ext-EPG）の使用

- テンプレートの構成を、対応するサイトにプッシュします。
- その結果、構成が各 APIC ドメインに適用され、各サイトで定義された Web EPG に属するエンドポイントへの外部接続が可能になります。

- 各サイトのローカル L3Out 接続は、インバウンドとアウトバウンドの接続に使用されます。

ここで説明したシナリオが単純なのは、各 Web EPG（関連付けられたブリッジドメインと IP サブネットを持つ）が各サイトで一意に定義されているからです。その結果、WAN から発信されるインバウンドトラフィックは、常に Web の宛先エンドポイントがあるサイトにステアリングされます。WAN と各 Cisco ACI サイトの間に境界ファイアウォールが展開されている場合、フローはこれを常に対称的に通過します。

図 106 に示すように、Web EPG とそれに関連付けられたブリッジドメインがサイトにまたがって拡張されると、異なる動作が発生する場合があります。

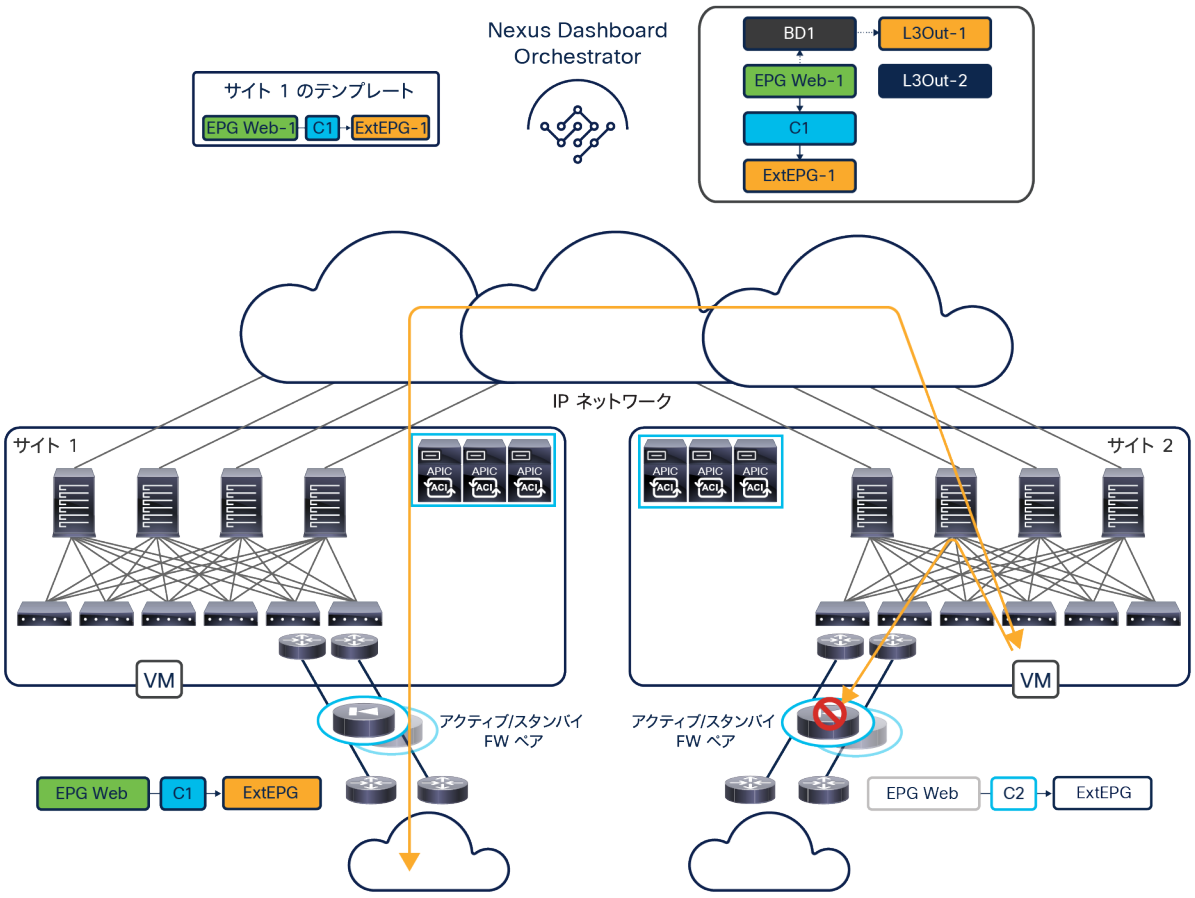


図 106.
マルチサイトとストレッチブリッジドメインを使用した従来型の L3Out 接続

この場合、サイトごとに異なる L3Out 接続が定義されますが、Cisco Nexus Dashboard Orchestrator で作成されたテンプレートが、同じ Web EPG を両方に関連付けます。これは、Web エンドポイントがサイトにまたがって拡張されたブリッジドメインに接続されているためです。Cisco ACI リリース 4.2(1) より前では、両方の L3Out に関連付けられた外部 EPG が、両方のサイトに関連付けられた同じテンプレートに属する「ストレッチ」オブジェクトでなければならないことに注意してください。任意のサイト（図 106 のサイト 1）に入ったトラフィックが、宛先エンドポイントが接続されているリモートサイトに VXLAN トンネル経由で転送されるようにするためには、ストレッチオブジェクトであることが必要です。外部 EPG が拡張されることによって、正しい変換エントリがスパインに作成されるからです。Cisco ACI リリース 4.2(1) 以降では、Ext-EPG をサイトでローカルにのみ展開することが可能になります。リモートサイトの内部（または外部）EPG とコントラクトを確立すると、スパイン内に必要な変換エ

ントリが作成されるためです（詳細は、「サイト間 L3Out 機能の導入（Cisco ACI リリース 4.2(1)/MSO リリース 2.2(1) 以降）」セクションを参照してください）。

結果として、デフォルトでは、同じ Web IP プレフィックスが両方のサイトから外部レイヤ 3 ネットワークに向けてアドバタイズされ、宛先エンドポイントがサイト 2 に接続されているにもかかわらず、インバウンドトラフィックがサイト 1 のボーダーリーフノードに配信される可能性があります。この場合、Cisco ACI マルチサイトがトラフィックを ISN 経由でエンドポイントに配信しますが、WAN へのリターンフローはサイト 2 のローカル L3Out 接続を使用します。このアプローチでは、サイト 2 の境界を保護するためにステートフル ファイアウォール デバイスが（サイト 1 に展開されたものとは独立して）展開されている場合、トラフィックがドロップされます。

Cisco ACI リリース 4.0(1) 以降、ボーダーリーフノードの L3Out 接続からより詳細なホストルートをアドバタイズする機能がサポートされ、この問題に対する「ネットワーク中心型」ソリューションが可能になります。

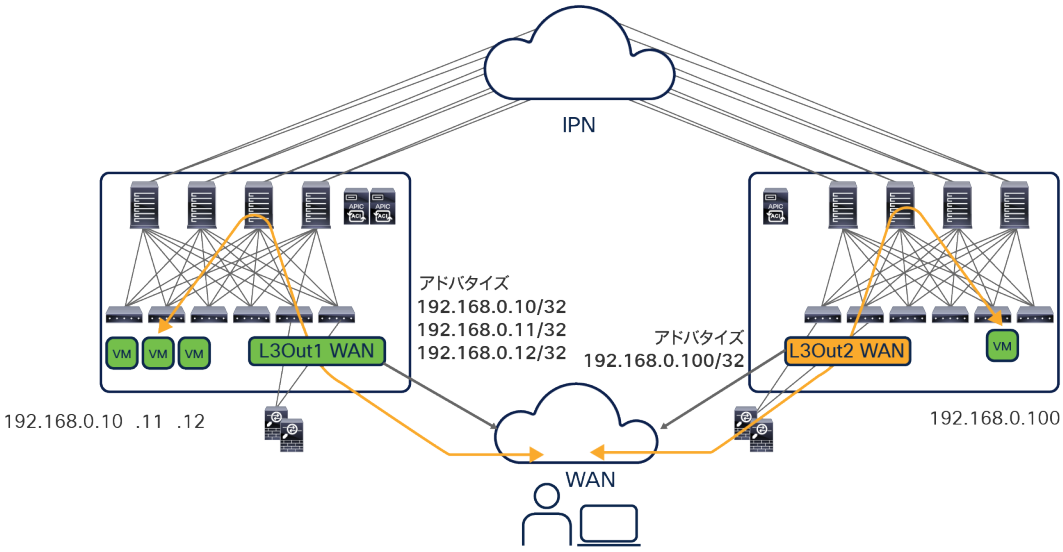


図 107.
ホストルートアドバタイズによるインバウンド/アウトバウンド トラフィック フローの最適化

注： Cisco ACI リリース 5.2(5) と NDO リリース 4.0(2) の時点では、SR-MPLS L3Out ハンドオフでのホストルートアドバタイズはサポートされていません。

図 107 に示すように、ローカル L3Out 接続でホストルートアドバタイズを有効にすると、入力トラフィックフローと出力トラフィックフローが対称になり、外部ネットワークと内部 EPG の間の通信が必ず同じステートフルサービスを通過するようになります。ホストルートアドバタイズはブリッジドメインレベルで有効にできるため、ファブリックの外部にアドバタイズするホストルートを厳格に制御し、拡張性の問題を軽減できます。

注： ホストルートを外部ネットワークに注入する場合、そのホストルートがリモートサイトの L3Out 接続で受信されないようにすること、およびファブリック内に再アドバタイズされないようにすることが重要です（VXLAN データプレーン経由で確立されたサイト間のネイティブな水平方向通信を妨げるおそれがあるため）。各ファブリックの L3Out とバックボーンで使用されるルーティングプロトコルによっては、受信側 L3Out でこのホストルートフィルタリングが自動的に適用されます。これに該当しない場合は、入力ルートフィルタリングを適用して受信したホストルートを明示的にドロップすることをお勧めします。

ホストルートアドバタイズのサポートにより、1 つの ACI サイトをある IP サブネットの「ホームサイト」とする設計オプションの展開が可能になります。ホームサイトの本質的な目的は、その IP サブネットに属するエンドポイントの大部分を、普段はそのサイトに接続しておき、特定の理由（事業継続性やディザスタリカバリのシナリオなど）があった場合のみ別のサイトに移行できるようにすることです。

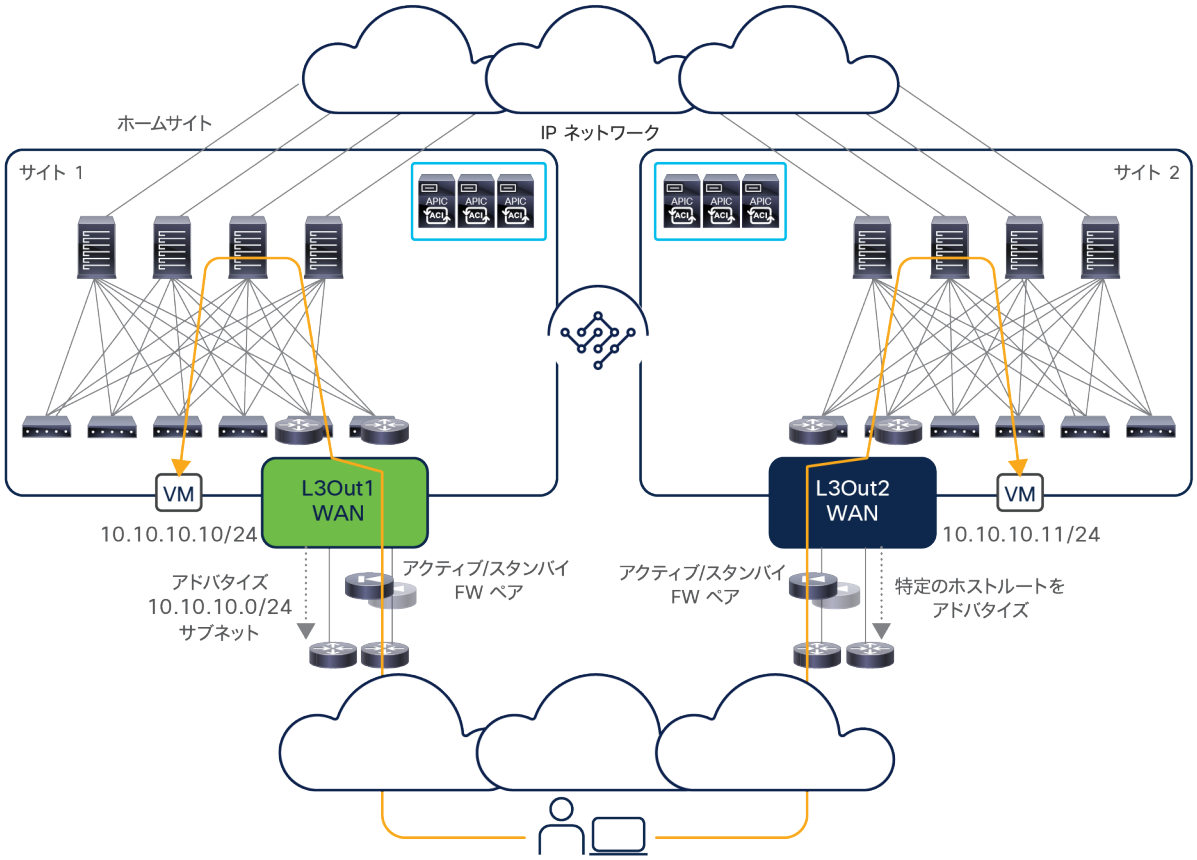


図 108.
ホームサイトの展開

図 108 に示す動作は、各サイトの L3Out に関連付けられたルートマップを定義し、これを使用して外部ネットワークドメインにアドバタイズする必要のある IP プレフィックスを指定することで、簡単に実現できます。サイト 1 のルートマップは IP サブネットのプレフィックスのみに一致し、サイト 2 のルートマップは特定のホストルートと IP サブネットのプレフィックスに一致するように定義します。ただし、後者の IP サブネットのプレフィックスは、サイト 1 より低い優先度を付けてアドバタイズする必要があります。IP サブネット情報があればインバウンドトラフィックをホームサイトにステアリングできるため、結果として注入する必要のあるホストルートの総数が削減されます。

注： L3Out に関連付けられたルートマップの構成は、Cisco Nexus Dashboard Orchestrator では実行されず、ローカル APIC レベルで実行する必要があります。この構成の具体的な詳細は、このホワイトペーパーの範囲外です。詳細は、https://www.cisco.com にある Cisco ACI 構成ガイドを参照してください。

ホストルート情報を WAN でアドバタイズできない（拡張性に懸念があるか、ホストルートが WAN ネットワークで受け入れられないため）シナリオでは、ホストルートを外部ルータにエクスポートすることを検討してください。こうすることで、以下のような設計オプションが利用できるようになります。

- 外部ルータをリモートルータに直接接続するオーバーレイトンネルを確立するか、別々のデータセンターのロケーションに展開された外部ルータ間を直接接続するオーバーレイトンネルを確立することで（Generic Routing Encapsulation（GRE）を通したマルチプロトコル ラベル スイッチング（MPLS）（MPLSoGRE）またはプレーンな GRE トンネリングを使用）、ホストルートが WAN に公開されることを防止します。

  注：外部ルータでの GRE サポートについては、https://www.cisco.com/c/ja_jp/partners/partner-with-cisco/integrator/levels.html にある特定のプラットフォームのドキュメントで確認してください。
- WAN ネットワークにまたがって一種のオーバーレイ コントロール プレーンを確立できる SD-WAN ソリューション（Cisco SD-WAN など）を統合することで、詳細なルーティング情報をリモート WAN エッジルータに直接提供できるようにします。
- 外部ルータで Locator ID Separation Protocol（LISP）を使用することで、リモート LISP ルータが、LISP マッピングサーバーに登録されたエンドポイント情報に基づいて、トラフィックをカプセル化して正しいサイトに送信できるようにします。

サイト間 L3Out 機能の導入（Cisco ACI リリース 4.2(1)/MSO リリース 2.2(1) 以降）

上の図 108 に示す動作では、あるサイトに展開されたエンドポイントが、別のサイトに展開された L3Out 接続を介してファブリックの外部にあるリソースと通信できないため、重要なユースケースが実現できない場合があります。

L3Out 接続は、通常、外部ネットワークドメイン（WAN またはインターネット）への接続に使用されますが、メインフレーム、ファイアウォール、ロードバランサなどのリソースとの接続を確立するために展開されることもまれではありません。サイト間通信が必要になる場合が多いためです。

これが、Cisco ACI リリース 4.2(1)/MSO リリース 2.2(1) で新しいサイト間 L3Out 機能を導入した主な理由です。これによって、下の図 109 に示すユースケースが可能になります。

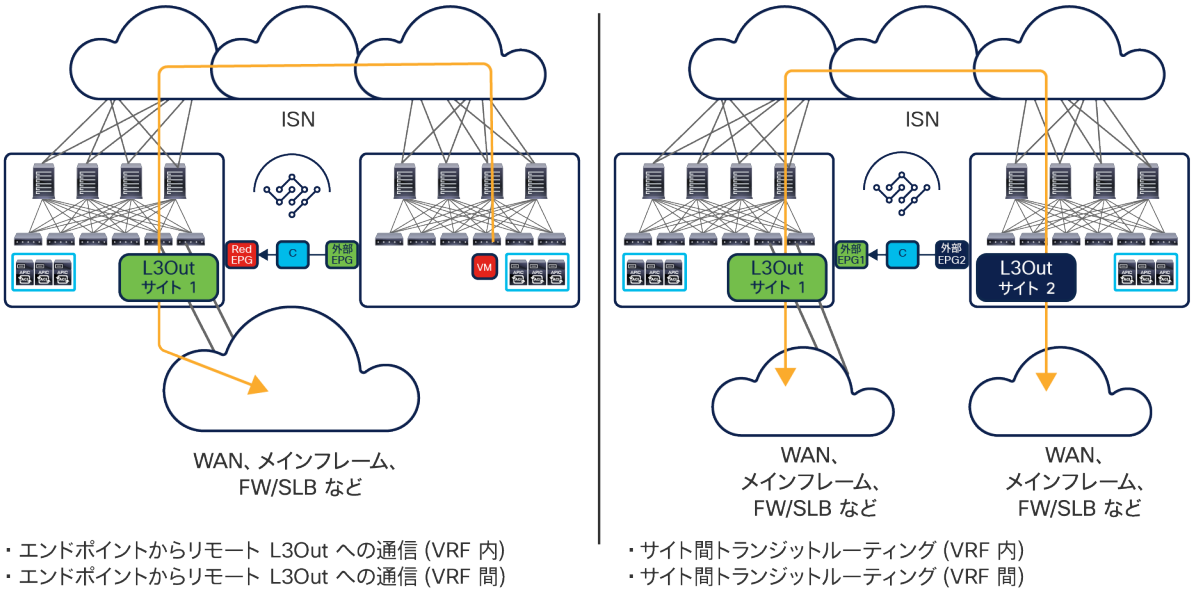


図 109.
サイト間 L3Out 機能でサポートされるユースケース

上図のように、サイト間 L3Out によって、あるサイトに接続されたエンドポイントが、別のサイトに展開されそのリモート L3Out 接続（VRF 内と VRF 間の両方）を経由して到達できる「エンティティ」（WAN、メインフレーム、サービスノードなど）と通信できるようになります。さらに、サイトにまたがるトランジットルーティング（VRF 内と VRF 間）も可能になります。これは、さまざまなエッジネットワークドメインを相互接続するコアネットワークとしてマルチサイトドメインを展開するシナリオで重要な役割を果たします。

注： 前述のように、図 109 に示す動作は、SR-MPLS L3Out ハンドオフでも実現できますが、NDO リリース 4.0(2) 以降が必要です。

サイト間 L3Out のガイドラインと制約事項

このドキュメントの執筆時点で（つまり、Cisco ACI ソフトウェアリリース 5.0(1) で）サイト間 L3Out について考慮が必要な制約事項を以下に挙げます。検討しているソフトウェアバージョンのリリースノートを常にチェックして、完全にサポートされている機能を確認してください。

- サイト間 L3Out 通信を確立するためには、ACI サイトで少なくとも Cisco ACI リリース 4.2(1) を実行する必要があります。同じマルチサイトドメインに以前のソフトウェアバージョンを実行しているファブリックを展開しておくこともできますが、そこに接続されているローカルエンドポイントにサイト間 L3Out 接続の要件がない場合に限られます。
- 図 109 に示すユースケースをサポートするためには、サイト間 L3Out 接続を必要とするすべてのサイトで Cisco ACI リリース 4.2(2) 以降を実行することを強くお勧めします。4.2(1) ソフトウェアイメージでいくつかの不具合が検出されたためです。
- サイト間 L3Out はボーダーリーフ L3Out を展開する場合に限ってサポートされ、GOLF L3Out ではサポートされません。

- Cisco ACI マルチサイトとリモートリーフを展開する場合、サイト間 L3Out 接続はサポートされません。つまり、リモートリーフのロケーションに接続されたエンドポイントが別のサイトに展開された L3Out と通信することはできません（その逆も同様です）。
- 図 109 に示すサイト間 L3Out のユースケースで CloudSec トラフィック暗号化がサポートされますが、Cisco ACI リリース 5.2(4) 以降に限られます。

サイト間 L3Out のコントロールプレーンとデータプレーンに関する考慮事項

サイト間 L3Out をサポートするには、マルチサイトドメインに属する各サイトに別の TEP プール（「外部 TEP プール」または「ルーティング可能 TEP プール」と呼ばれます）を展開する必要があります。各サイトの外部 TEP プールの構成は、Nexus Dashboard Orchestrator から直接管理されます（インフラストラクチャ構成タスクの一環として）。この新しいプールの導入により、追加の TEP アドレスが BL ノードに割り当てられます。図 110 に示すように、あるサイトのエンドポイントと、リモートサイトにある L3Out 接続を介してアクセスできる外部リソースの間の通信が、リーフ間を直接接続する VXLAN トンネルを作成することで確立されるためです。

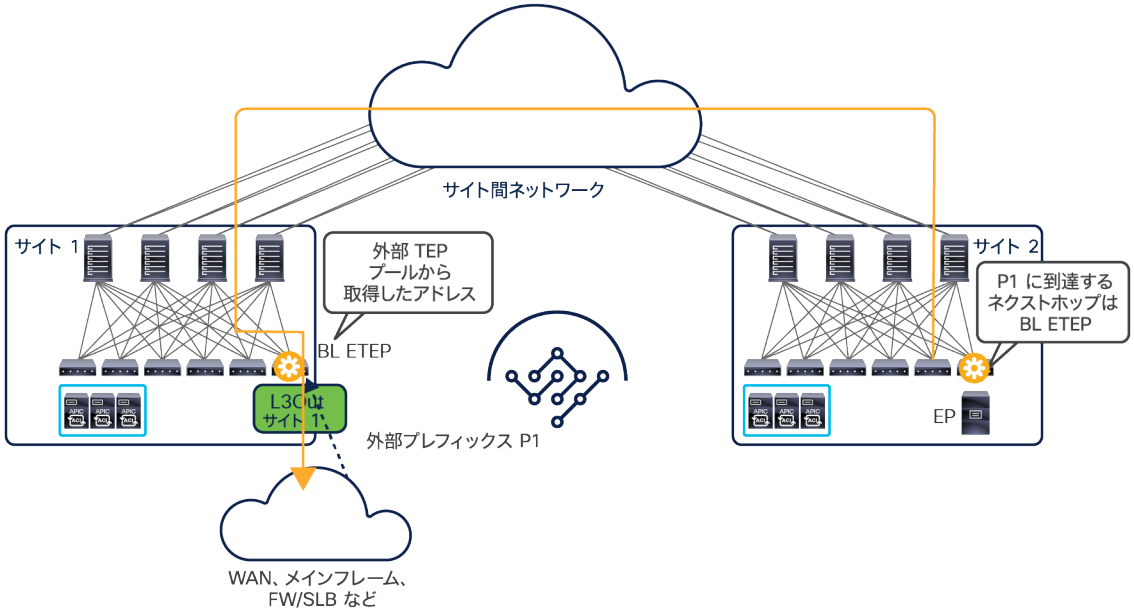


図 110.
サイトにまたがるリーフ間 VXLAN トンネルの確立

外部 TEP プールがない場合、上に示した VXLAN トンネルを確立するには、各ファブリックに割り当てられた元の TEP プールが一意になっていて、ISN 経由でルーティングできることが必須です。この要件が厳しすぎると受け取られることも少なくありません。さらに、同じマルチサイトドメインに属するファブリック間で元の TEP プールが重複している場合があります。このシナリオは、上に示したリーフ間トンネルが確立できない原因になります。

NDO で外部 TEP プールを構成し、内部 EPG と外部 EPG の間のコントラクトを作成すると、別々のサイトに展開されたスパイン間に MP-BGP VPNv4（または VPNv6）の隣接関係が確立されます。これが、ローカル L3Out が最初に構成されていない場合でも、外部 TEP プールを同じマルチサイトドメインに属するすべてのサイトに割り当てる必要がある理由です。

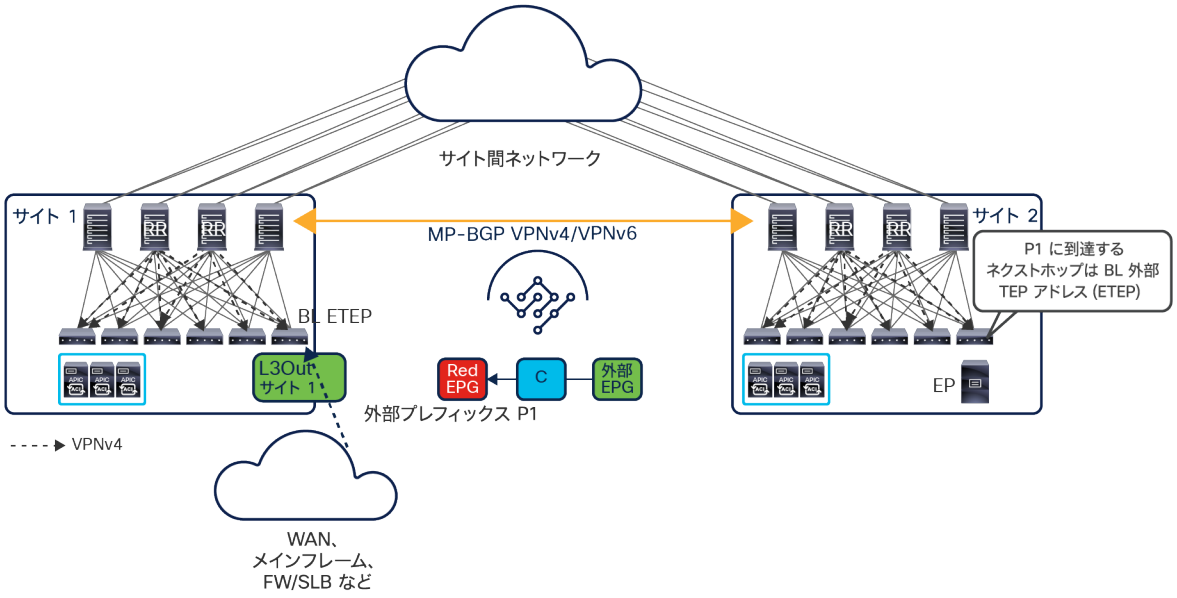


図 111.
サイト間の VPNv4/VPNv6 BGP の隣接関係

これらの隣接関係は、すべてのテナントの外部プレフィックスについて L3Out 接続から学習される到達可能性情報を交換するために使用され、エンドポイントの到達可能性情報を交換するために使われる既存の MP-BGP EVPN セッションとは別に確立されます。上の図 111 の例では、最初に、サイト 1 の L3Out で受信された外部プレフィックス P1 が、ローカルスパインで構成されたルートリフレクタ機能を利用して、MP-BGP VPNv4 コントロールプレーン経由ですべてのローカルリーフノードに配布されます。その後、サイト 1 のスパインからサイト 2 のスパインにアドバタイズされ、そのスパインからローカルリーフノードに情報が伝達されます。最終的に、Red EPG エンドポイントが接続されているサイト 2 のリーフが、サイト 1 の BL ノードの外部 TEP アドレスを介してプレフィックス P1 へ到達できることを示すエントリをルーティングテーブルに追加します。

重要なのは、エンドポイント間の通常のサイト間（水平方向）通信とは異なり、受信側のサイト 1 のスパインが、外部プレフィックスを宛先とするサイト間通信で VNID/クラス ID 変換を実行する役割を負わない点です（図 112）。

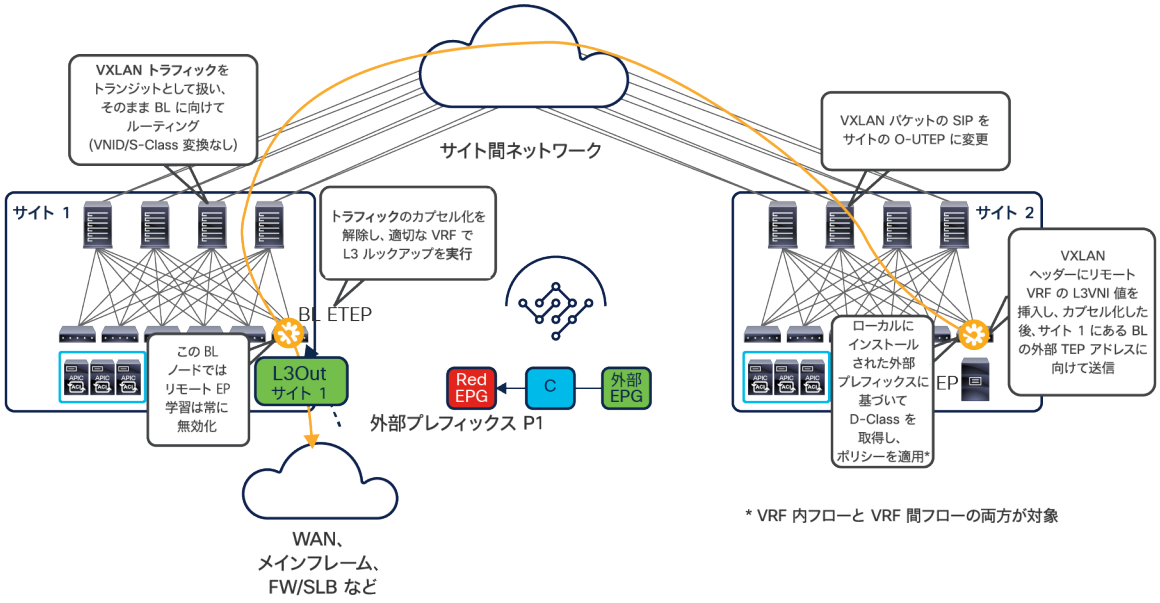


図 112.
受信側のスパインでの VNID/S-Class 変換は不要

これは、BL ノードに直接接続される VXLAN トンネルが確立されることによって、サイト 1 のスパイン（およびサイト 2 のスパイン）がアンダーレイ インフラストラクチャのルーテッドホップそのものになり、VXLAN 通信が可能になるためです。これから、以下の 3 つのことが言えます。

- サイト 1 の L3Out の L3VNI 値に関する情報を、MP-BGP VPNv4 コントロールプレーンを介してサイト 1 からサイト 2 に伝達する必要があります。これにより、サイト 2 のコンピューティングリーフがこの情報をパケットの VXLAN ヘッダーに追加できるようになります。また、サイト 1 の BL ノードがこのトラフィックを受信するとき、L3Out 接続から外部に送信する前に、正しい VRF に対してレイヤ 3 ルックアップを実行できるようになります。ユースケースによって、VRF がサイト 2 のエンドポイントの VRF と同じ場合も異なる場合もあります。
- セキュリティポリシーは常にサイト 2 のコンピューティングリーフで適用されます。サイト 1 に展開された外部 EPG のクラス ID がそこでローカルにプログラミングされているためです。
- Red EPG エンドポイント情報が、（通常のように）データプレーンでのアクティビティに基づいてサイト 1 の BL ノードで学習されることはありません。これは、前述のように、サイト 1 のスパインでクラス ID/VNI 変換が行われないためです（トラフィックが BL ノードの VTEP に直接送信されるため）。したがって、Red EPG を識別する、サイト 2 での送信元クラス ID は、サイト 1 の APIC ドメインでは意味がない（または完全に異なる意味になる）かもしれません。

サイト 1 からサイト 2 へのリターンフローを以下の図 113 に示します。

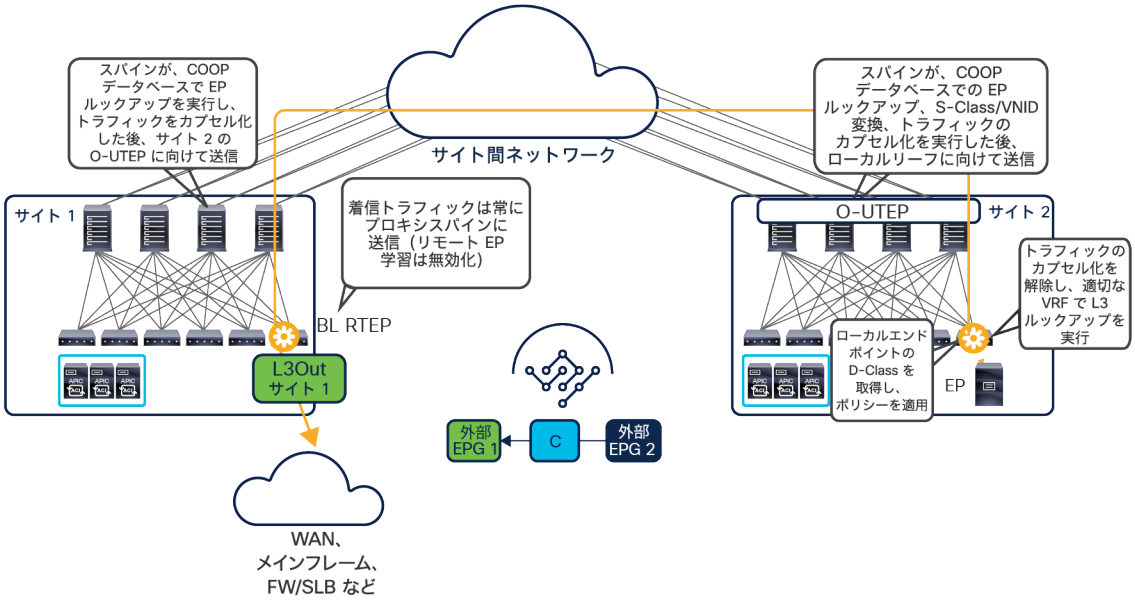


図 113.
L3Out からリモートエンドポイントへのリターントラフィック

注目すべきポイントを以下に示します。

- L3Out 接続から受信され、Red EPG に属するリモートエンドポイントに向かうトラフィックは、常に BL ノードでカプセル化され、ローカルのプロキシスパインサービスに送信されます。前述のように、BL ノードには学習されたエンドポイント情報がないためです。BL ノードは、ローカルサイトで割り当てられた宛先エンドポイントの VRF を識別する L3VNI を VXLAN ヘッダーに挿入します。
- 次に、ローカルスパインがこれをカプセル化し、宛先エンドポイントが検出されたリモートサイトのスパインノードに送信します。これは、「Cisco ACI マルチサイトのオーバーレイデータプレーン」セクションですでに説明したサイト間通信の通常の動作です。したがって、受信側のスパインは、エンドポイント間の水平方向通信で通常行われているとおり、S-Class/VNID 変換サービスを実行します。
- 次に、受信側のコンピューティング リーフ ノードが、正しい VRF でレイヤ 3 ルックアップを実行します。さらに、ポリシーを適用し、許可された場合はトラフィックを宛先エンドポイントに転送します。

一方、図 114 は、別々のサイトに展開された L3Out 接続間でトラフィックがルーティングされるトランジットルーティングのユースケースを示しています。

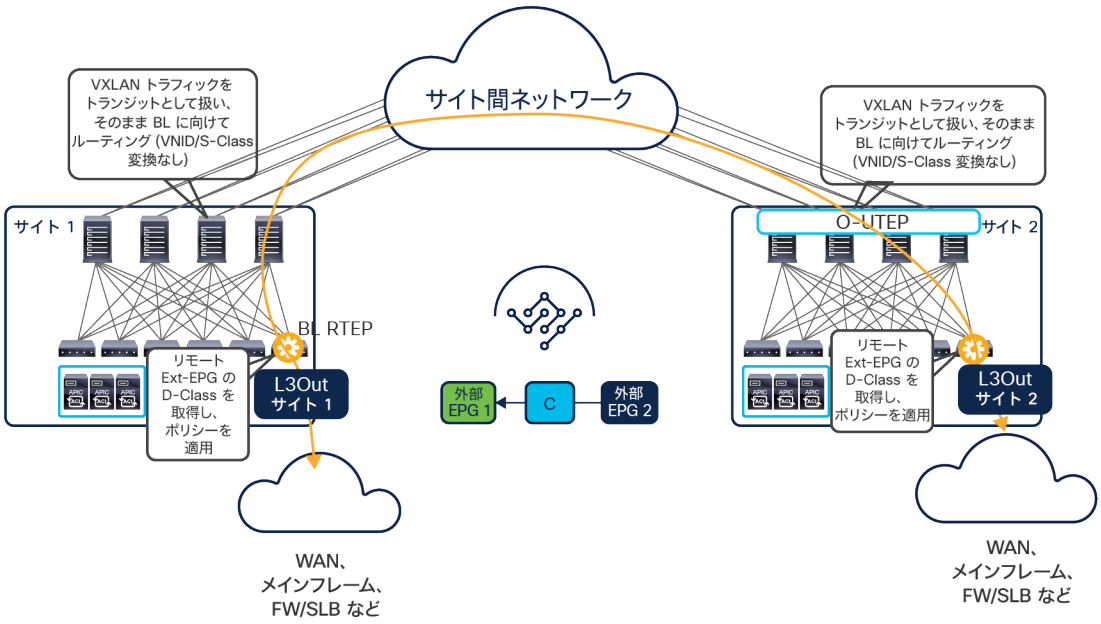


図 114.
ACI サイト間のトランジットルーティング

トランジットルーティングが、同じ VRF 内で発生した場合も、異なる VRF（共有サービスのユースケース）間で発生した場合も、常に、それぞれのサイトに展開された BL ノード間を直接接続する VXLAN トンネルが確立されます。したがって、スパインは VXLAN カプセル化トラフィックをルーティングするだけです。

ポリシーの観点から見ると、コントラクトが適用されるのは常にコンピューティングノードです。これは、リモートの外部 EPG のクラス ID がローカルで常に取得できるためです。

サイト間 L3Out の展開に関する考慮事項

サイト間 L3Out を展開する場合、L3Out の構成とそれに関連する外部 EPG を展開する方法を最初に決定する必要があります。Cisco Multi-Site Orchestrator リリース 2.2(1) 以降、実際には、Orchestrator テンプレートで直接 L3Out オブジェクトを構成することができます（外部 EPG の場合は、Orchestrator の最初のリリースから対応しています）。

重要な点を指摘しておくと、その場合でも、L3Out の特定の構成（論理ノード、論理インターフェイス、ルーティングプロトコル、ルートマップなど）は必ず APIC レベルで実行する必要があります。Orchestrator で L3Out 「コンテナ」を公開する必要があるのは、通常、あるサイトに展開された BD サブネットを、他のサイトに展開された L3Out 接続から外部にアドバタイズできるようにする場合です。この点は、以下で明らかになります。

L3Out と外部 EPG の両方を NDO で作成できるため、それらを（複数のサイトにマッピングされた 1 つのテンプレートで定義される）ストレッチオブジェクトとして構成する必要があるかを最初に検討します。

- L3Out オブジェクトに関しては、各サイトの L3Out に一意の名前を付けるほうが運用上簡単になると思われます。そのためには、1 つのサイトだけにマッピングされたテンプレートを用意し、テンプレートごとに個別

の L3Out を作成する必要があります。そうすることで、外部ネットワークドメインに実際に接続されるサイトにのみ L3Out オブジェクトを作成することもできます。

- L3Out 接続が 1 つ以上のサイトにすでに展開されているブラウンフィールドのシナリオでは、各 APIC ドメインから、そのサイトにマッピングされたテンプレートに L3Out オブジェクトをインポートすることをお勧めします。
- 前述のように、外部 EPG をストレッチオブジェクトとして作成するかどうかは、多くの場合 L3Out によって接続されるリソースで決まります。WAN の場合、L3Out 接続が展開されたすべてのサイトから同じ外部リソースにアクセスするのが非常に一般的です。したがって、外部 EPG としてストレッチ EPG を使用すると、ポリシー定義が簡素化されます。他方で、特定のサイトに接続されたメインフレームサーバーに L3Out を介してアクセスする場合、ローカル Ext-EPG を使用するほうが妥当です。
- 外部 EPG をローカルオブジェクトとして展開するかストレッチオブジェクトとして展開するかにかかわらず、展開された L3Out 接続にそれらを常にリンクする必要があります。

L3Out と外部 EPG が作成されると、内部リソースと外部ネットワークの間で接続を確立できるようになります。また、ACI マルチサイトアーキテクチャを経由する外部ネットワーク間のルーティングが可能になります（L3Out-to-L3Out 通信またはトランジットルーティング）

サイト間 L3Out が導入される前は、前述のように、内部エンドポイントと外部ネットワークの間のトラフィックパスはかなり限定的でした。

- サイトでローカルに定義された（すなわち、拡張されていない）EPG/BD に属するエンドポイント、またはサイトにまたがって拡張されたエンドポイントの場合、アウトバウンド通信はローカル L3Out 接続経由に限定されていました。
- サイトでローカルに定義された（すなわち、拡張されていない）EPG/BD に属するエンドポイントの場合、リモート L3Out 接続から BD サブネットをアドバタイズできなかったため、インバウンド通信もローカル L3Out 経由に限定されていました。
- サイトにまたがって拡張された EPG/BD に属するエンドポイントの場合、リモートサイトで L3Out を使用すると、インバウンドフローが ISN を経由する最適ではないパスに固定される可能性がありました。インバウンドフローが常に最適なパスを通るようにするために、ホストベースのルーティングアドバタイズを可能にするオプションが Cisco ACI リリース 4.0(1) 以降提供されています。

サイト間 L3Out を有効にすると、上記の動作が変更される可能性があるため、そうすることによる機能面での具体的な影響を理解することが非常に重要です。

図 115 は、常にローカル L3Out 接続を優先オプションとして使用する最適なアウトバウンド トラフィック パスを示しています（この例では、Red EPG/BD と Green EPG/BD は拡張されておらず、各サイトでローカルに定義されています）。

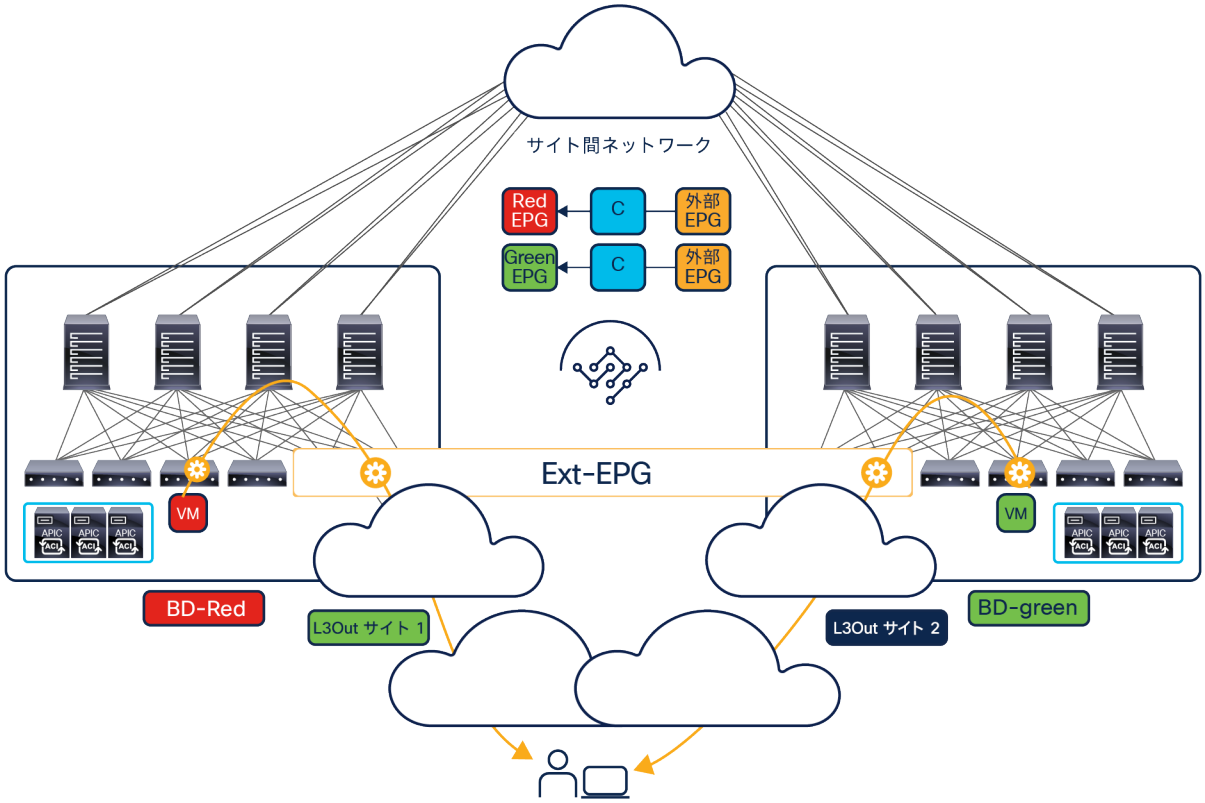


図 115.
最適なアウトバウンド トラフィック パス

これは、サイト間 L3Out が構成されていなくてもサポートされる唯一の動作ですが、ACI ファブリックを外部ネットワークに接続するために使用されるルーティングプロトコルによっては、この機能を有効にする場合に状況が変わる可能性があります。

- 外部ネットワークで OSPF を使用している場合、サイト 1 とサイト 2 の L3Out を介して同じ外部プレフィックスを受信すると、デフォルトでは、各サイトのすべてのリーフノードがローカルアウトバウンドパスを優先します。これは、受信された外部プレフィックスが各サイトのボーダーリーフノードによって ACI VPNv4 コントロールプレーンに注入され、スパイン間で確立された VPNv4 セッションを介してサイト間で交換されるためです（上の図 115 を参照）。ボーダーリーフノードでルートマップを適用してそのプレフィックスの BGP 属性を変更しない限り、各リーフは IS-IS メトリックの観点からトポロジ的に近いボーダーリーフノード（したがって、ローカルサイトのボーダーリーフノード）から受信した情報を常に優先します。
- 外部ネットワークで EIGRP を使用している場合、L3Out で受信したプレフィックスに関連付けられた EIGRP メトリックが、ファブリック内で実行されている ACI VPNv4 BGP プロセスで MED として伝達されます。その結果、サイト 1 とサイト 2 の L3Out で同じプレフィックスを受信した場合、EIGRP メトリックが同じ場合にのみ、ローカル L3Out がデフォルトでアウトバウンドフローに使用されます。あるサイトの L3Out で受信されたプレフィックスの EIGRP メトリックが「良」（すなわち、最小値）の場合、プレフィックスは「良」

（より小さい）MED 値で BGP に注入されるため、（ローカルおよびリモートサイトからの）すべてのアウトバウンドフローは、その L3Out 接続を介して送信されます。

- 外部ネットワークで BGP を使用している場合（通常は EBGP が一般的なオプションです）、ACI VPNv4 コントロールプレーンにアドバタイズされるプレフィックスの BGP 属性は、デフォルトで外部 IPv4/IPv6 ピアリングによって伝送されます。その結果、たとえば、サイト 1 で受信したルートの AS-Path 属性がサイト 2 で受信した同じルートの AS-Path 属性よりも悪かった（すなわち、長かった）場合、同じマルチサイトドメインに属するすべてのリーフノードからのアウトバウンド通信は、常にサイト 2 の L3Out を優先します（図 116）。

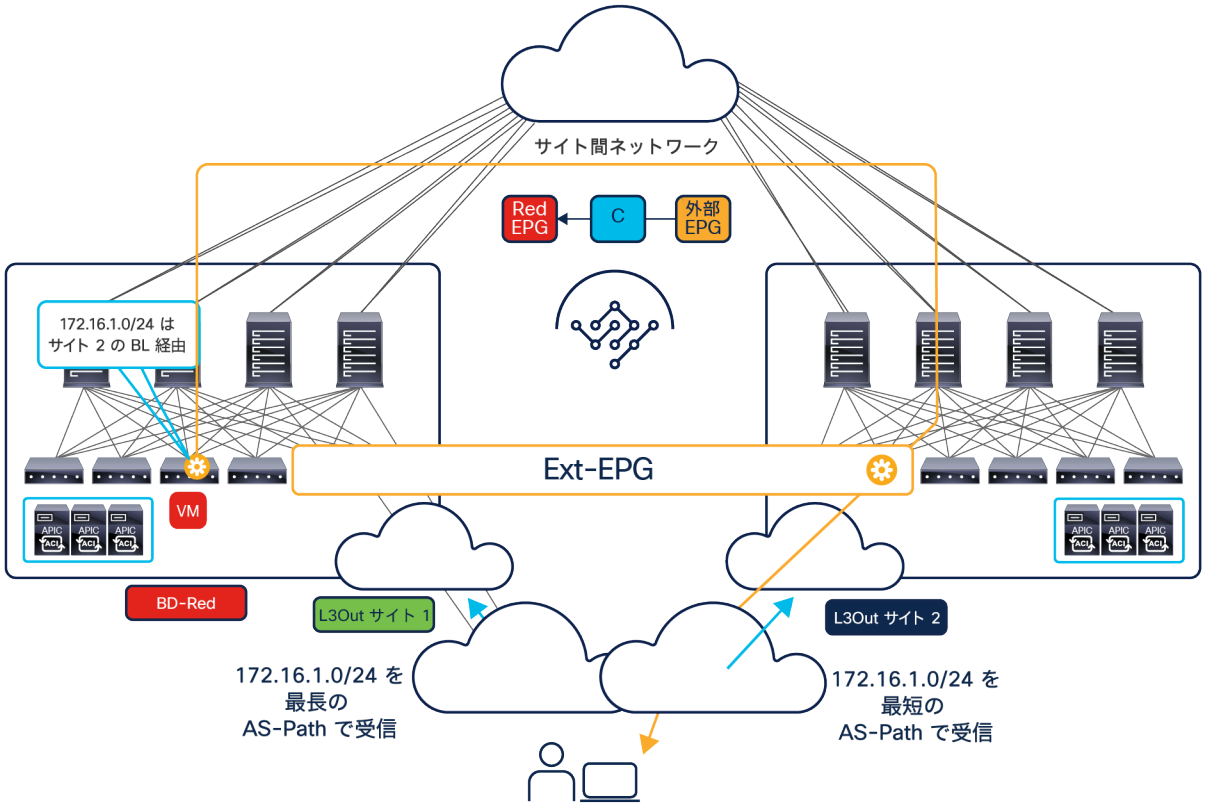


図 116.
最適でないアウトバウンド トラフィック パス

重要な点を 1 つ指摘しておくと、現在の実装では、サイト間 L3Out を有効にした場合、あるサイトの任意のローカル L3Out 接続で学習されたプレフィックスが、常にすべてのリモートサイトにアドバタイズされます。これらのプレフィックスは、対応する VRF が展開されているリモートサイトのリーフノードにインストールされます。これにより、図 117 に示す例のように、予期しないトラフィックパスのシナリオが作成される可能性があります。

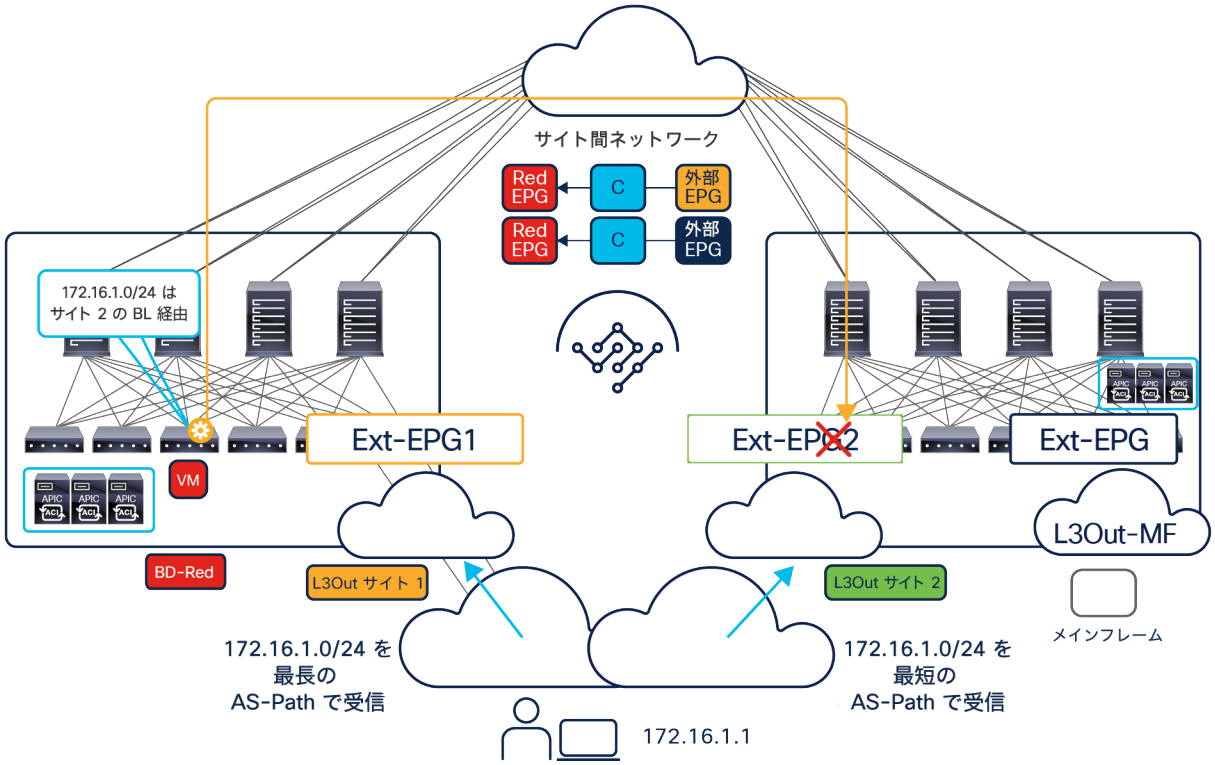


図 117.
別々の外部 EPG を使用している場合に想定されるアウトバウンドトラフィックのドロップ

このケースでは、サイト間 L3Out を有効にして、サイト 1 で接続された Red EPG のエンドポイントと、サイト 2 の L3Out 接続の背後に展開されたメインフレームサーバーの間の通信を許可しています（したがって、Red EPG と、L3Out-MF に関連付けられた Ext-EPG の間にコントラクトが構成されています）。したがって、サイト 2 に展開された両方の L3Out で学習された外部プレフィックスがサイト 1 に伝達され、この例では、172.16.1.0/24 プレフィックスに関連付けられた AS-Path 属性により、サイト 1 のエンドポイントがサイト 2 の L3Out を経由するパスを優先するようになります。外部 EPG をストレッチ EPG として展開する場合、前の図 116 に示した最適ではないアウトバウンドパスが発生します。図 117 の例のように、サイトごとに別々の外部 EPG が展開されている場合、サイト 1 の Red EPG とサイト 2 の外部 EPG の間でコントラクトが作成されていないため、Red EPG と外部クライアントの間の通信はドロップされます。したがって、サイト間 L3Out を有効にする前に、既存の通信パターンへの想定される影響を慎重に検討することが重要です。

- 外部ルーテッドドメインで使用されるルーティングプロトコルとは無関係に、あるサイト内の特定の L3Out をアウトバウンドフローが使用するように強制することも可能です。そのためには、サイト 2 の BL ノードにインバウンドルートマップを適用して、VPNv4 コントロールプレーンにプレフィックスを注入する前に特定の BGP 属性を調整する必要があります。たとえば、2 つのファブリックが、同じ BGP ASN に属している場合、local-preference の調整が、推奨されるアプローチです（値が大きいほど優先され、デフォルト値は 100 です）。一方、異なる ASN に属している場合は、local-preference 属性が EBGP セッションにまたがって伝送されないため、AS-Path の調整が推奨されます。図 118 は、特定の VRF のすべてのアウトバウンド通信がサイト 2 の L3Out 接続を経由するように強制されるシナリオを示しています。これを実現する方

法として、サイト 1 の L3Out で受信したすべてのプレフィックスに対して、より長い AS-Path 値（AS-Path プリペンドを使用）を設定しています。

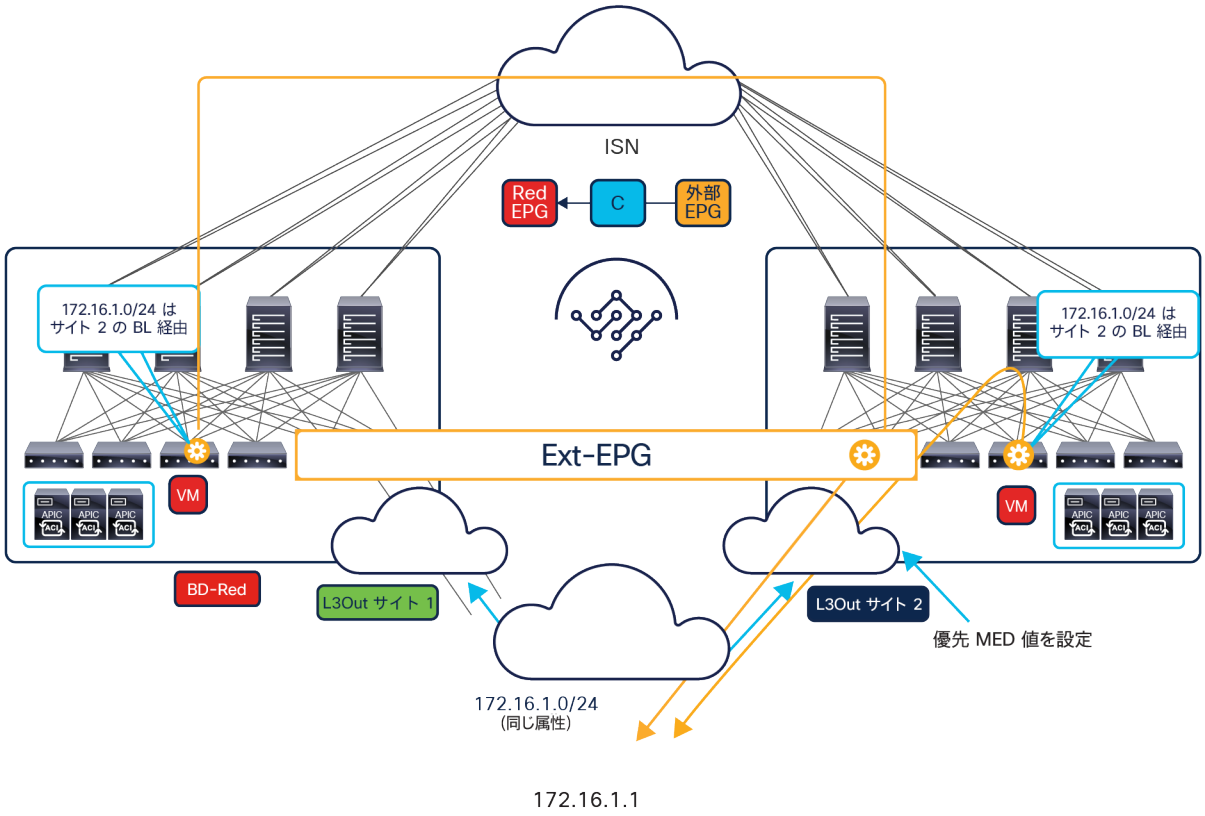


図 118.
特定の L3Out を経由するように強制されたアウトバウンドトラフィック

注： BL ノードで受信した外部プレフィックスの AS-Path を設定するには、BL ノードと外部ルータの間のルーティングプロトコルとして BGP を実行している必要があります。

一方で、インバウンド トラフィック フローに関しては、サイト間 L3Out を有効にすると、下の図 119 に示すような動作が発生する可能性があります。

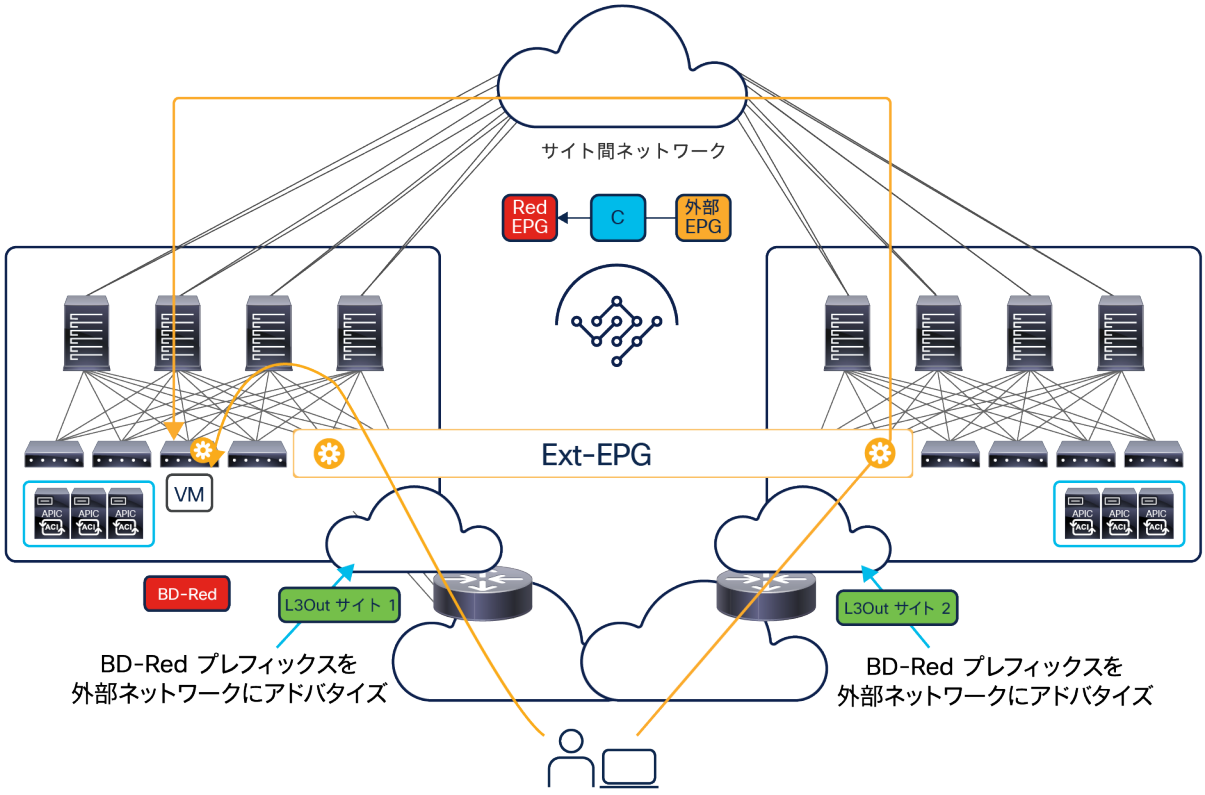


図 119.
最適でないインバウンド トラフィック パスの生成

この例では、サイト間 L3Out が有効になっていて、サイト 1 のローカル L3Out に障害が発生した場合でも、Red エンドポイントはサイト 2 に展開された L3Out 接続を介して引き続き外部ネットワークと通信できます。これを可能にするためには、（サイト 1 でローカルに展開された）BD-Red に関連付けられたサブネットのプレフィックスを、サイト 1 の L3Out からだけでなく、サイト 2 の L3Out からもアナウンスすることが必要なのは明らかです。

注： あるサイトで定義された BD をリモートサイトの L3Out からアナウンスする動作は、直接 Orchestrator で BD を L3Out オブジェクトにマッピングすることによって制御できます。これが、Cisco Multi-Site Orchestrator リリース 2.2(1) 以降、L3Out オブジェクトを GUI で公開した理由の 1 つです。このアプローチが望ましくない場合は、直接 APIC レベルで L3Out に適用されるルートマップを使用することも可能です。

その結果、外部ネットワークのルーティング設計によっては、外部クライアントから発信されて Red エンドポイントに向かうトラフィックが、BD-Red がサイトにまたがって拡張されておらず、サイト 1 にとどまっているにもかかわらず、サイト 2 のボーダーリーフノードに向けてステアリングされる可能性があります。

4.2(1) より前の ACI ソフトウェアリリースでは、サイトでローカルに定義された BD の IP プレフィックスがローカル L3Out 接続からしかアナウンスされなかったことを考えると、この結果は意外と思われるかもしれません。したがって、サイト間 L3Out 機能を有効にすることによって起こり得るこのような影響を常に考慮することをお勧めします。解決策としては、直接 APIC レベルで L3Out にルートマップを適用して、外部にアドバタイズされるプレフィックスのプロパティを変更することが考えられます。たとえば、EBGP を外部ルータとピアリングする場合、AS-Path プリペンド構成を実行して、BD が元々展開されていないサイトでインバウンドパスの優先度を下げることが可能です。

ネットワークサービスの統合

Cisco ACI マルチサイトアーキテクチャを検討する際には、さまざまなネットワークサービス統合モデルが考えられます。

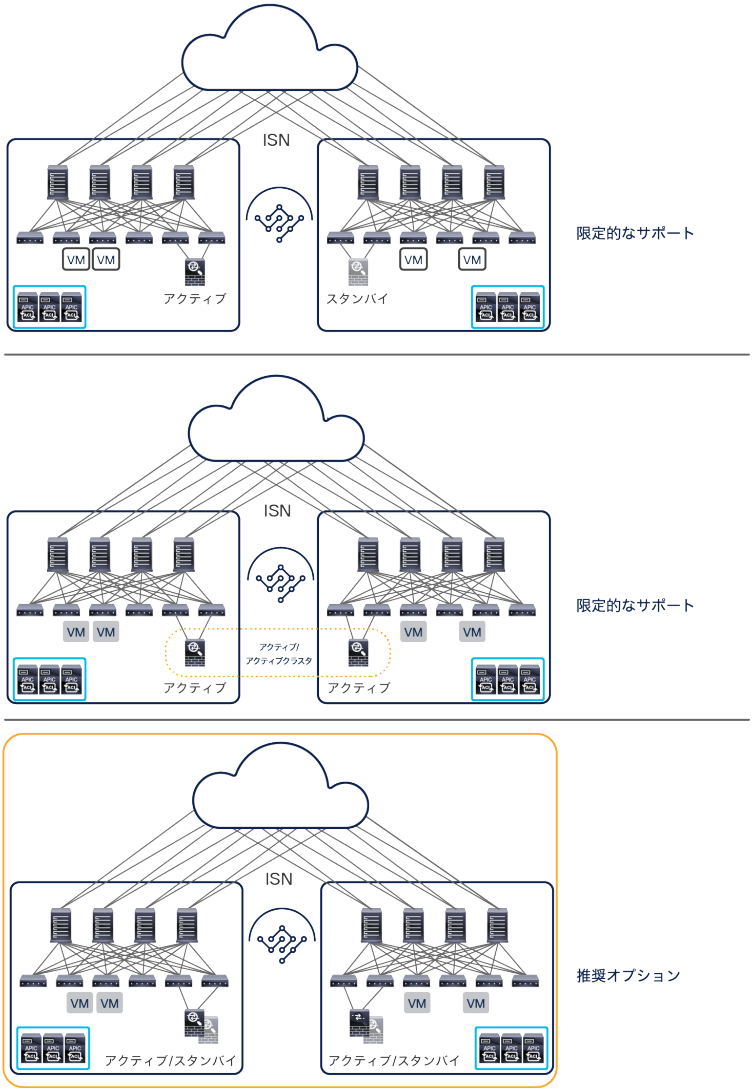


図 120.
Cisco ACI マルチサイトとネットワークサービス統合モデル

最初の 2 つのモデルでは、サイトにまたがるクラスタ化されたサービスの展開が必要です。Cisco ACI リリース 5.1(1) より前は、サイトにまたがって展開されたサービスノードのアクティブ/スタンバイクラスタに対するサポートが非常に限定的で、Cisco ACI がレイヤ 2 転送のみを実行するシナリオに制約されています（ファイアウォールを、エンドポイントのデフォルトゲートウェイとするか、透過モードで使用）。

Cisco ACI リリース 4.2(1) 以降は、サイト間 L3Out 機能の導入により、各サイトの L3Out 接続に接続された境界 FW ノードのアクティブ/スタンバイクラスタを展開できます。ただし、別々のサイトに展開されたサービスノード間でデータ VLAN 上の L2 接続を確立する必要がない場合に限られます（L3Out に関連付けられた BD をサイトにまたがって拡張できないため）。

サイトにまたがって展開されたサービスノードのアクティブ/アクティブクラスタに対するサポートも限定的で、「アクティブ/アクティブ」をどのように定義するかに大きく依存します。同じクラスタに属するすべてのファイアウォールノードが、同じ MAC/IP アドレスを持っている場合、たとえば、Cisco FTD ファイアウォールをクラスタリングする際に、ノードを別々のファブリックに展開することはできません。一方、クラスタ内の各ノードが、異なる MAC/IP アドレスを使用していれば、それらを異なるサイトに接続することができます。

注： 一般的に言えば、クラスタ化されたサービスをデータセンターにまたがって展開する場合は、Cisco ACI マルチポッドが、推奨されるアーキテクチャです。詳細は、
https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-739571.html にあるホワイトペーパーを参照してください。

Cisco ACI マルチサイトアーキテクチャは、ネットワーク障害ドメインと管理の両方のレベルで、個別の ACI ファブリックを相互接続するように設計されています。これを踏まえると、サービス統合の推奨オプションで、クラスタ化（アクティブ/スタンバイまたはアクティブ/アクティブ）されたサービスを各ファブリックに独立して展開する必要があるのは妥当と言えます。

ACI マルチサイトアーキテクチャにおけるサービスノードの統合についての詳細な情報は、
https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-743107.html にあるホワイトペーパーを参照してください。

仮想マシンマネージャ統合モデル

Virtual Machine Manager（VMM）ドメインを Cisco ACI マルチサイトアーキテクチャに統合できます。サイトごとに APIC クラスタがあるため、サイトごとに別々の VMM ドメインが作成されます。その後、これらの VMM ドメインを Cisco Nexus Dashboard Orchestrator に公開して、そこで定義された EPG に関連付けることができます。これについては、このセクションの後半で説明します。

以下の 2 つの導入モデルが可能です。

- 複数の VMM インスタンス（vCenter Server、SCVMM など）を各サイトで使用できます。各インスタンスはローカル APIC クラスタとペアリングされます。
- 単一の VMM インスタンスを使用して、サイトにまたがって展開されたハイパーバイザを管理できます。このインスタンスは、異なるローカル APIC クラスタとペアリングされます。この導入モデルは、VMware vCenter と統合する場合に限られます。

以下の 2 つのセクションでは、これらのモデルについて詳しく説明します。サポートされている VMM（VMware vCenter Server、Microsoft System Center VMM（SCVMM）、OpenStack コントローラ）で VMM ドメインを構成

する方法についての詳細な情報は、https://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/apic/sw/2-x/virtualization/b_ACI_Virtualization_Guide_2_3_1.html を参照してください。

また、マルチサイトと VMM の統合は Cisco AVE を仮想スイッチとして展開する場合もサポートされますが、Cisco ACI リリース 4.2(1) 以降が必要である点に注意してください。

各サイトに展開された仮想マシンマネージャ

マルチサイト展開では、通常、VMM が各サイトに展開されハイパーバイザのローカルクラスタを管理します。このシナリオを図 121 に示します。

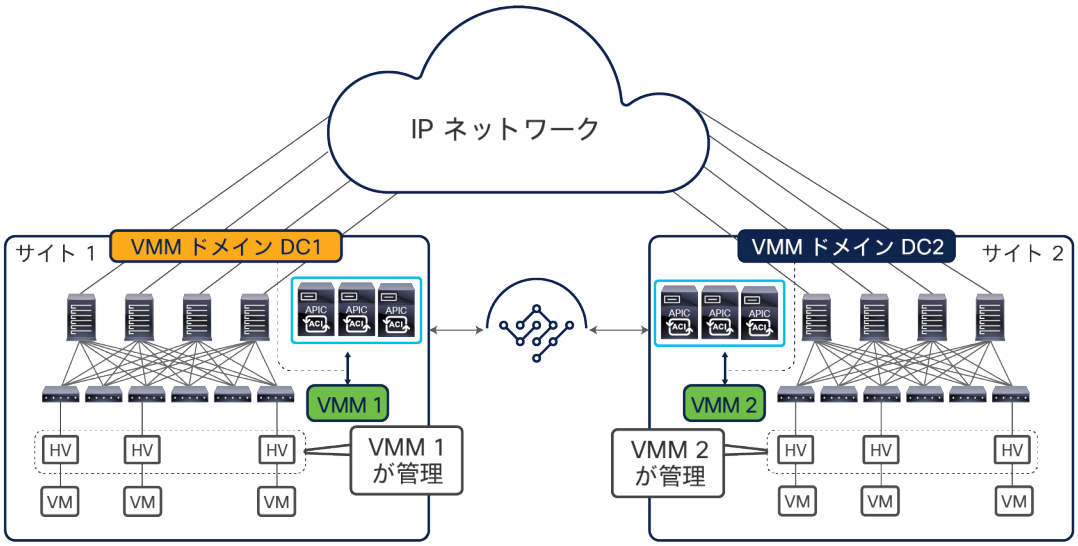


図 121.
各サイトに展開された仮想マシンマネージャ

各サイトの VMM は、ローカルホストを管理し、ローカル APIC ドメインとピアリングすることでローカル VMM ドメインを作成します。図 121 に示すモデルは、Cisco ACI で使用可能なすべての VMM オプション（VMware vCenter Server、Microsoft SCVMM、OpenStack コントローラ）でサポートされています。

VMM ドメインの構成は、ローカル APIC レベルで実行されます。その後、作成された VMM ドメインを Cisco Nexus Dashboard Orchestrator にインポートし、一元的に作成されたテンプレートで指定された EPG に関連付けることができます。たとえば、EPG 1 がマルチサイトレベルで作成される場合、EPG 1 を VMM ドメイン DC1 と VMM ドメイン DC2 に関連付けた後、そのポリシーをサイト 1 とサイト 2 にプッシュすることで、ローカルに導入することができます。

サイトごとに別々の VMM ドメインを作成すると、通常、サイト間での仮想マシンの移動がコールドマイグレーションのシナリオに制限されます。ただし、VMware vSphere 6.0 以降を使用する設計では、別々の vCenter Server によって管理されるハイパーバイザのクラスタ間でホットマイグレーションを実行できます。図 122 は、そのような構成を作成するために必要な手順を示しています。

注： このドキュメントの執筆時点では、異なる Cisco ACI ファブリック間でのライブマイグレーションが可能な VMM オプションは、vCenter Server リリース 6.0 以降に限られています。他の VMM（6.0 より前の vCenter リリースや SCVMM など）でライブマイグレーションを実行したい場合は、単一の Cisco ACI ファブリック（単一のポッドまたはマルチポッド）に VMM を展開する必要があります。

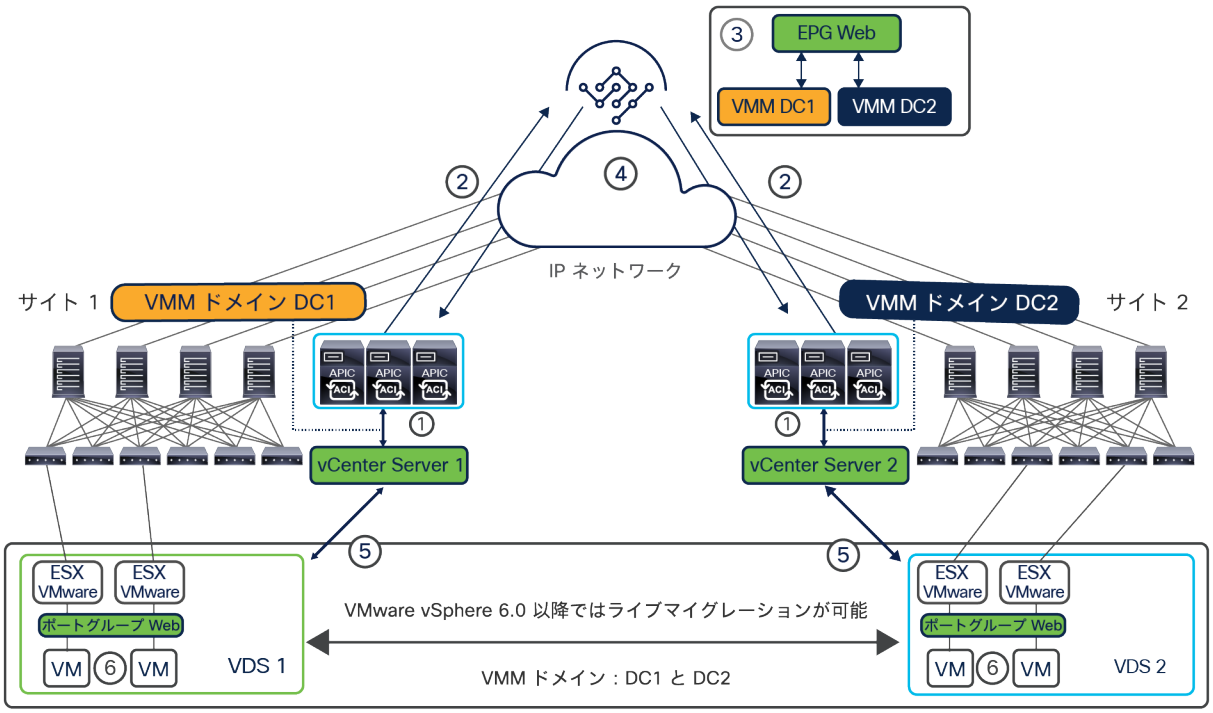


図 122.
VMware vCenter 6.0 以降を使用した場合の VMM ドメイン間でのライブマイグレーション

1. ローカル vCenter Server と APIC をピアリングして、各ファブリックに VMM ドメインを作成します。このピアリングにより、ESXi クラスタにローカル VMware 分散スイッチ（サイト 1 の VDS 1 とサイト 2 の VDS 2）が作成されます。
2. 作成された VMM ドメインは、Cisco Nexus Dashboard Orchestrator に公開できます。
3. サイト 1 とサイト 2 の両方に関連付けられた 1 つのテンプレートに新しい Web EPG を 1 つ定義します。この EPG を、対応する Web ブリッジドメインにマッピングします。このブリッジドメインは、サイト間のストレッチドメインとして構成されている必要があります（BUM 転送はオプションです）。次に、各サイトで、先ほど作成したローカル VMM ドメインにこの EPG を関連付けます。
4. テンプレートポリシーをサイト 1 とサイト 2 にプッシュします。
5. この EPG が各ファブリックに作成されます。また、EPG が VMM ドメインに関連付けられているため、各 APIC がローカル vCenter Server と通信し、このサーバーが関連付けられた Web ポートグループを各 VDS にプッシュします。
6. その後、新しく作成された Web ポートグループにサーバー管理者が Web 仮想マシンを接続できます。この時点で、サイト間でライブマイグレーションが実行できるようになります。

上記のライブマイグレーションは手動でトリガーする必要があることに注意してください。VDS にまたがった vSphere Distributed Resource Scheduler（DRS）の動作がサポートされていないため、そのダイナミックトリガーを利用することはできません。また、vSphere High Availability（HA）や vSphere Fault Tolerance（FT）などの機能も VDS 内での動作のみがサポートされているため、ファブリックにまたがって利用することはできません。

サイトにまたがる単一の仮想マシンマネージャ

図 123 は、単一の VMM がサイトにまたがって使用されるシナリオを示しています。

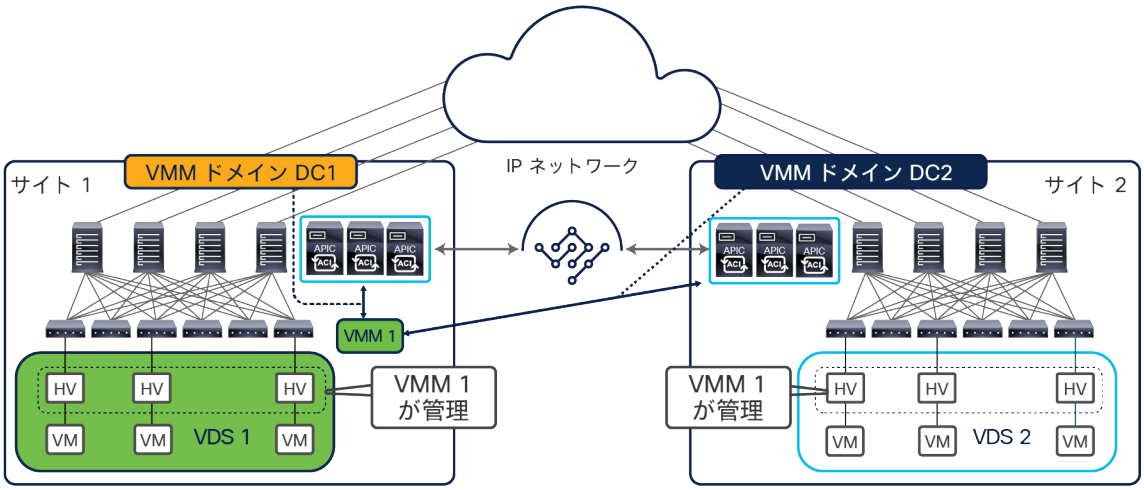


図 123.
サイトごとに個別の VMM ドメインを持つ単一の VMM

このシナリオは、APIC が VMware vCenter Server に統合されている場合に限ってサポートされます。このシナリオではサイト 1 に展開された単一の VMM が、同じファブリック内と別のファブリック内に展開されたハイパーバイザのクラスタを管理します。この構成の場合も、各ファブリックに異なる VMM ドメインが作成され、その結果、ローカルに展開された ESXi ホストに異なる VDS スイッチがプッシュされることに注意してください。このシナリオの場合も、ファブリックにまたがる仮想マシンのコールドマイグレーションとホットマイグレーションに関するサポートについての考慮事項は、本質的に前のセクションで説明したとおりです。

ブラウンフィールド統合シナリオ

実稼働環境にすでに ACI ファブリックが展開されていて、これを管理するために Nexus Dashboard Orchestrator の導入が必要になる場合も多くあります。この場合、そのファブリックの構成が、APIC を利用してすでにプロビジョニングされています。したがって、主な論点は、このファブリックをマルチサイトドメインに追加する方法と、追加した後に NDO から構成を引き続き管理できるようにする方法です。

実際の展開で見られる代表的な 2 つのシナリオを図 124 に示します。いずれも、既存の ACI ファブリックを管理するために NDO の導入が必要になるケースです。

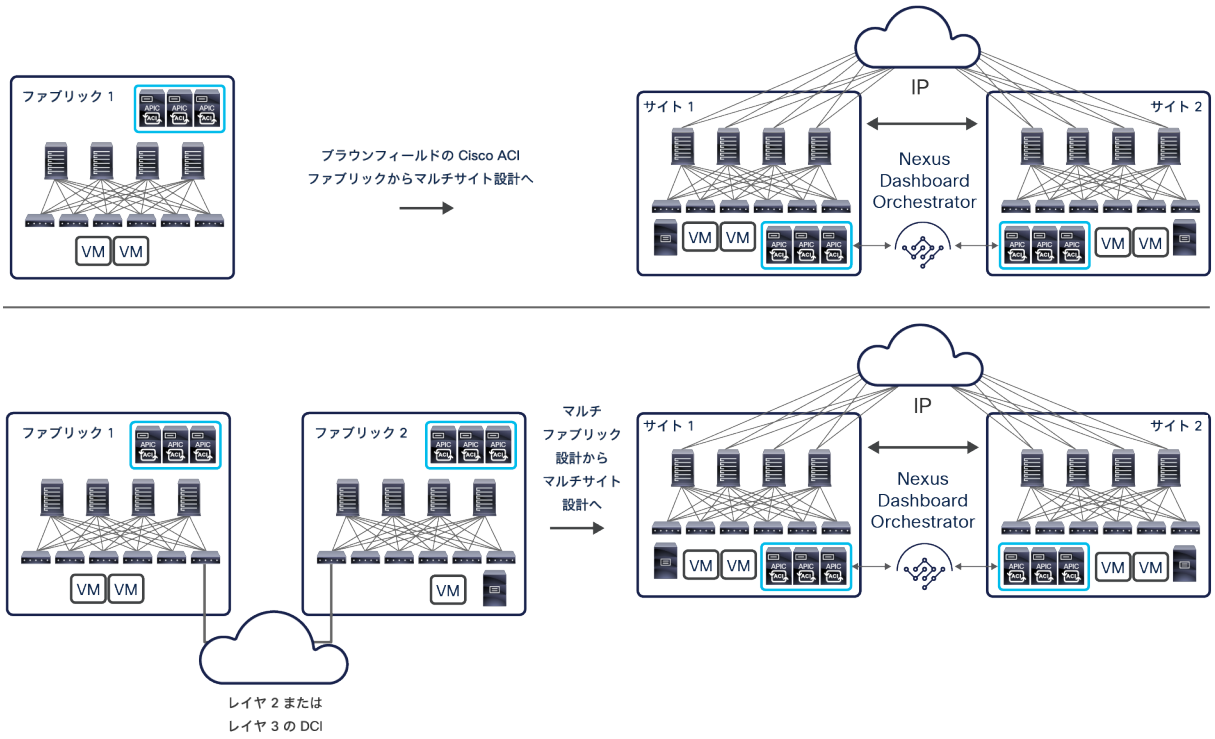


図 124.
Cisco ACI マルチサイトの移行シナリオ

図 124 の上側にあるユースケースは非常に単純で、1 つ（または複数）の Cisco ACI ファブリックを既存のファブリックに追加する場合です。「Cisco ACI のマルチポッドとマルチサイトの統合」セクションですでに説明したように、このユースケースは、2 つの DC のロケーション（単一のマルチポッドファブリックを展開）でアクティブ/アクティブ設計を実行していて、この環境を新しい DR のサイトに接続する必要がある場合に該当します。

図 124 の下側にあるユースケースは、既存のマルチファブリック設計を Cisco ACI マルチサイト設計に変換する場合です。「レイヤ 3 のみのサイト間接続」セクションで説明したように、ACI ファブリックを引き続き「自律型ファブリック」として実行し、L3Out データパスを介してそれらの間に限ったレイヤ 3 接続を確立することは可能です。しかし、一元管理の手段として NDO を導入し、サイト間の水平方向接続に VXLAN データパスを使用することには、いくつかの優位性があります。さらに、独立した ACI ファブリック間にレイヤ 2 接続を拡張するために外部 DCI テクノロジー（OTV、VPLS など）を活用する展開が、過去にいくつか本稼働しています。この「デュアルファブリック」設計は現在推奨されていません。そのため、マルチサイトへの移行パスを用意する必要があります。

インフラストラクチャの観点から見ると、上記の両方のシナリオにおける最大の変更点は、異なるファブリックのスパイン間がサイト間ネットワーク（ISN）を経由して接続されるようになることです。プロビジョニングの観点から見ると、APIC からではなく NDO からポリシーの構成を処理するようになることが大きな変化です。自動化の観点から見ても、この変化によって影響を受けることは明らかです。NDO とやり取りするための API が、APIC とやり取りする API と異なるためです。また、さらに重要なこととして、APIC で元々定義されていたポリシーを NDO に効率的にインポートする方法が課題となります。

Cisco APIC から Cisco Nexus Dashboard Orchestrator への既存のポリシーのインポート

Cisco APIC から NDO に既存のポリシーをインポートする 2 つの一般的なシナリオを図 125 に示します。

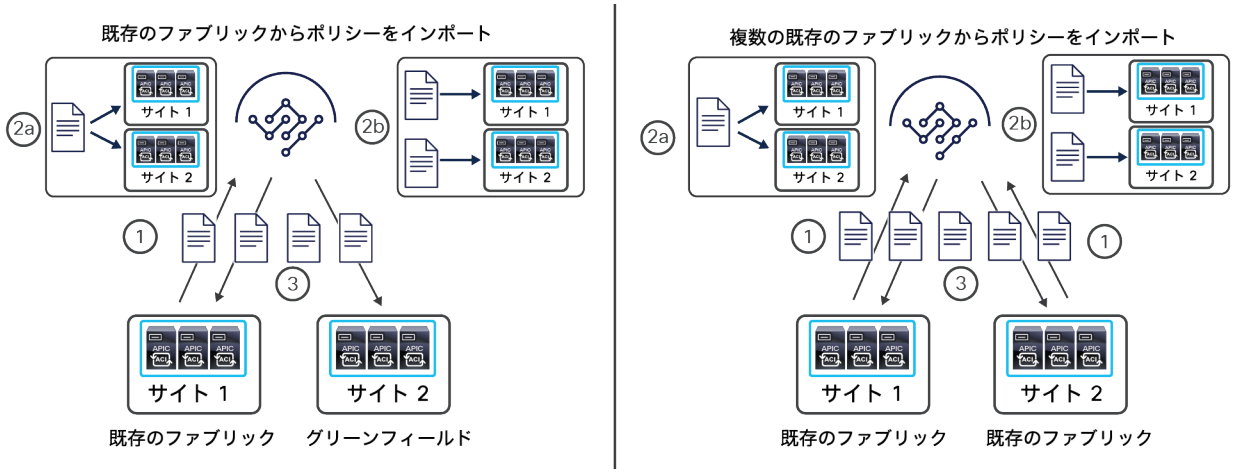


図 125.
Cisco Nexus Dashboard Orchestrator へのポリシーのインポート

注： 特にテナントポリシーのインポートに関して以下の点を考慮する必要があります。「NDO リリース 4.0(1) で導入された新しいテンプレートタイプ」セクションで説明したように、リリース 4.0(2) 以降、NDO でファブリックポリシーとモニタリングポリシーの管理とプロビジョニングが可能になったため、これらのテンプレートタイプに関しても同じ考慮が必要です。

テナントポリシーをインポートできる前提として、そのテナントが NDO に存在している必要があることは明らかです。これは、サイト 1 にすでに展開されているテナントとまったく同じ名前のテナントを NDO で直接作成するか、サイト 1 からそのテナントを「インポート」することによって実現できます。この時点でテナントの「インポート」を実行すると、NDO にテナントのみが作成され、テナント関連の構成はインポートされないことに注意してください。

図 125 の左側にあるシナリオは非常に単純で、必要なインポート手順は以下のとおりです。

1. 最初に、テナントの既存のポリシーを、展開済みの Cisco ACI ファブリックから NDO にインポートする必要があります。
2. インポートされたポリシーは、「NDO のスキーマとテンプレートの展開」セクションで説明したように、さまざまなテンプレートに編成する必要があります。たとえば、既存のファブリックのみにローカルにプロビジョニングされたテナント構成をそのままにしておきたい場合は、そのサイトにのみ関連付けられているテンプレートにインポートする必要があります。VRF の接続をサイトにまたがって拡張する計画であれば、その VRF を両方のサイトに関連付けられているテンプレートにインポートする必要があります。サイト 1 にすでに展開されているその他のオブジェクト、たとえば、BD、EPG、コントラクトをサイト 2 に拡張する必要がある場合も同様です。さらに、サイト 2 に関連付けられた専用のテンプレートを使用してそのサイトの新しいローカルポリシーを作成できます。

   BD をストレッチテンプレートにインポートする場合、BD の構成を変更して [L2ストレッチ（L2 Stretch）] オプションを有効にする必要があることに注意してください。その際、下に示す警告メッセージが表示されます。

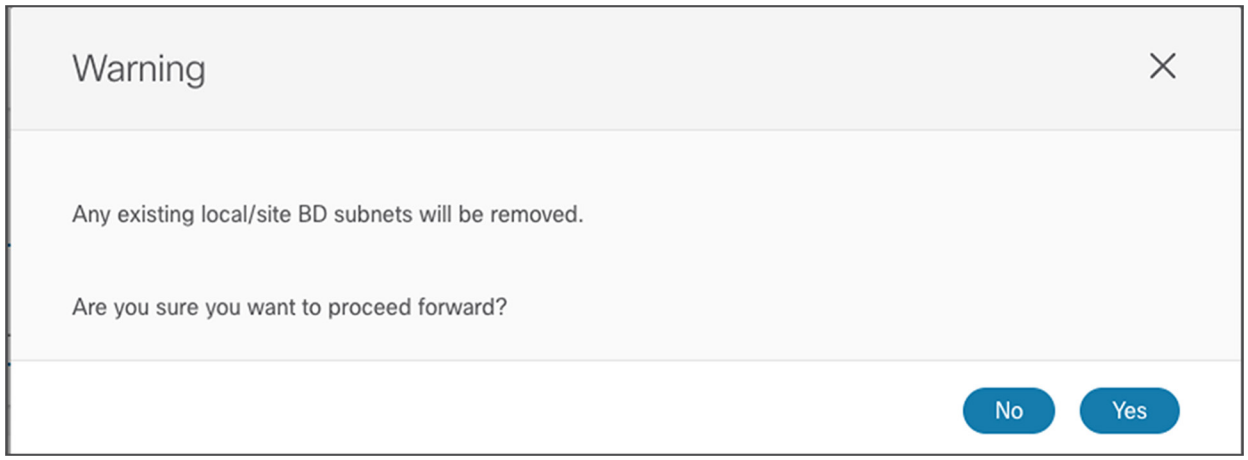


図 126.
BD の [L2ストレッチ（L2 Stretch）] オプションを有効にする際の警告メッセージ

このメッセージは単に、BD がサイトにまたがって L2 で拡張されると、BD サブネットがすべてのサイトに共通のグローバルプロパティになるため（これは分散エニーキャストゲートウェイ機能のインスタンスです）、サイトローカルのプロパティではなく、グローバルテンプレートの構成の一部として構成する必要があることを示しています。ただし、これを実行してテンプレートを展開しても、元のサイト（BD サブネットがすでに展開されているサイト）で接続に問題が発生することはありません。サイト 2 でデフォルトゲートウェイ機能が有効になるだけです。

注：BD/EPG ペアを最初に元のサイト 1 にのみ関連付けられたテンプレートにインポートする場合（これらはローカルのみのオブジェクトであるため）でも、後にこの BD/EPG ペアを拡張する必要が生じる場合、Cisco Nexus Dashboard Orchestrator でこれらのオブジェクトをストレッチテンプレートに移行できます。逆の操作も可能です。この場合は、拡張されている BD/EPG ペアをリモートサイトから削除し、サイトのローカルオブジェクトにする必要があります。NDO リリース 4.0(2) の時点では、テンプレート間でオブジェクトを移行するこの機能は、BD と EPG のみが対象で、テンプレートが同じテナントに関連付けられている場合に限って利用できます。

その後、この構成を NDO から APIC ドメインにプッシュバックします。これにより、サイト 1 の既存のオブジェクトには注釈属性のみが追加され、このオブジェクトが NDO に管理されていることを示す情報がグラフィックで表示されます。同時に、サイト 2 に新しいオブジェクトが作成されます。

注： インポートされたポリシーをサイト 1 に再度プッシュしても、ファブリック内やファブリックと外部ネットワークドメインの間にすでに確立されている既存の通信が中断されることはありません。

図 125 の右側にあるユースケースについても同様の考慮が必要です。さらに次の点を考慮することが重要で、これはマルチサイト展開で欠かせない基本的なルールの 1 つです。つまり、あるオブジェクトが、異なるサイトに存在しながら同じ「もの」を表す必要がある場合（たとえば、ストレッチ EPG やストレッチ BD）、NDO では単一のオブジェクトと見なされ管理される必要があります。そのため、既存のポリシーを異なる APIC ドメインから NDO にインポートする必要がある場合、特定のテナントのそれぞれのサイトですでに定義されているポリシーに（テナント名自体も含めて）一貫した名前を与えておく必要があります。これは、現在、NDO に、異なる名前のオブジェクトを「マージ」する機能がないためです。

OTV が 2 つの ACI ファブリック間に展開され、EPG/BD がサイトにまたがって拡張されているシナリオを例に説明します。各サイトで EPG と BD に一貫した名前が与えられている場合、それらのオブジェクトは、両方の APIC ド

メインからストレッチテンプレートに簡単にインポートできます（サイト固有のプロパティをインポートする目的も達成できます）。しかし、EPG-Site1/BD-Site1 および EPG-Site2/BD-Site2 という名前が初めに与えられていると、NDO にインポートされたときに異なるオブジェクトであると見なされます。

上で説明した重要な点を考慮すると、図 125 の右側にあるシナリオの場合、ブラウンフィールドをインポートする手順は、以下のように整理できます。

1. 最初に、展開済みの両方の Cisco ACI ファブリックからテナントの既存のポリシーを NDO にインポートする必要があります。
2. インポートされたポリシーは、前のシナリオですでに説明したように、さまざまなテンプレートに編成する必要があります。また、ストレッチテンプレートに追加する必要がある共通オブジェクトに命名するときには、上記の点を考慮するのを忘れないでください。
3. 次に、構成を NDO から APIC ドメインにプッシュバックします。注釈がすべてのオブジェクトに追加され、この時点から、テナントのプロビジョニングの管理が NDO に完全に引き継がれます。

展開のベストプラクティス

このセクションでは、Cisco ACI マルチサイト設計を簡単かつスムーズに展開するのに役立つベストプラクティスとヒントについてまとめています。これらの推奨事項は、コンセプト実証段階から実稼働環境への展開にいたるまで、お客様のネットワークに実際の Cisco ACI マルチサイト設計を展開したときに、シスコ自身が学び経験したことをベースにしています。

Cisco Nexus Dashboard Orchestrator クラスタの展開

Cisco Nexus Dashboard Orchestrator クラスタを展開する際には、以下の推奨事項に注意してください。

- MSO リリース 3.1(1) 以前を実行している場合、アウトオブバンド（OOB）管理ネットワークを使用して Cisco Multi-Site Orchestrator クラスタを APIC に接続してください。これが公式にサポートされている唯一のオプションであるためです。Nexus Dashboard で NDO を実行している場合は、アウトオブバンド（OOB）アドレス、インバンド（IB）アドレス、またはその両方を使用して APIC に柔軟に通信できます。
- Orchestrator サービスの実行に使用する Cisco クラスタノードは、そのサービスがサイトとして管理することになる Cisco ACI ファブリック内に展開しないでください（ファブリック内接続が影響を受けたときに構成を変更できないようになることを防ぐため）。Cisco ACI ファブリックの外部に展開するのが望ましい方法です。使用される Orchestrator リリースに応じて、APIC の OOB/IB（または両方の）インターフェイスにアクセスできるインフラストラクチャに接続してください。
- Cisco Multi-Site Orchestrator の各クラスタノード（または、ND で Orchestrator をサービスとして実行している場合は Nexus Dashboard ノード）には、ルーティング可能な IP アドレスが必要です。さらに、3 つのノードすべてで相互に ping が成功する必要があります。Multi-Site Orchestrator（または Nexus Dashboard）のクラスタノードに別々の IP サブネットから IP アドレスを割り当てることができます（つまり、ノード間の L2 隣接関係は必要ありません）。
- Multi-Site Orchestrator の Docker ベースのバージョンを展開する場合、クラスタの可用性を高めるために、すべての MSO ノードが別々の ESXi ホストに展開されるようにしてください。
- Nexus Dashboard のサービスとして NDO を導入する場合、Orchestrator サービス専用の仮想 ND クラスタを用意することをお勧めします（つまり、このクラスタに他のアプリケーションをホストしないでください）。

- クラスタ内の NDO ノード（または ND ノード）間の最大遅延時間は 150 ミリ秒 RTT 未満である必要があります。
- Docker ベースの MSO クラスタを実行している場合、Cisco Multi-Site Orchestrator クラスタノードと Cisco ACI APIC ノードの間の最大遅延時間は 1 秒 RTT まで許容されます。Nexus Dashboard コンピューティングクラスタで Orchestrator サービスを実行する場合、許容される最大遅延時間が 500 ミリ秒 RTT に減少します。
- Cisco Multi-Site Orchestrator クラスタは、内部のコントロールプレーンとデータプレーンに以下のポートを使用します。したがって、アンダーレイネットワークでこれらのポートが常に開いている必要があります（ネットワーク内にファイアウォールを展開して ACL を設定する場合）。
  - TCP ポート 2377（クラスタを管理するための通信に使用）
  - TCP ポート 7946 および UDP ポート 7946（ノード間の通信に使用）
  - UDP ポート 4789（オーバーレイ ネットワーク トラフィックに使用）
  - TCP ポート 443（Cisco Multi-Site Orchestrator ユーザーインターフェイス（UI）に使用）
  - IP 50（カプセル化セキュリティプロトコル（ESP）による暗号化に使用）
- セキュリティを確保するため、Multi-Site Orchestrator クラスタ内でのコントロールプレーンとデータプレーンの通信はすべて IPsec を用いて暗号化されます。MSO ノードは 150 ミリ秒 RTT まで離れて配置することができ、その結果、クラスタ内の通信がセキュアでないネットワーク インフラストラクチャを通過する場合が想定されるためです。
- Docker ベースのインストールの場合、Cisco Multi-Site Orchestrator 仮想マシンの最小仕様は、以下に示すように導入されたソフトウェアリリースによって異なります。
  - Cisco Multi-Site Orchestrator リリース 1.0(x) の場合：

    VMware ESXi 5.5 以降

    最小要件：仮想 CPU（vCPU）4 個、メモリ 8 Gbps、ディスク容量 50 GB
  - Cisco Multi-Site Orchestrator リリース 1.1(x) の場合：

    VMware ESXi 6.0 以降

    最小要件：仮想 CPU（vCPU）4 個、メモリ 8 Gbps、ディスク容量 50 GB
  - Cisco Multi-Site Orchestrator リリース 1.2(x) 以降の場合：

    VMware ESXi 6.0 以降

    最小要件：仮想 CPU（vCPU）8 個、メモリ 24 Gbps、ディスク容量 100 GB
  - Cisco Multi-Site Orchestrator リリース 2.2(x) 以降の場合：

    VMware ESXi 6.0 以降

    最小要件：仮想 CPU（vCPU）8 個、メモリ 48 Gbps、ディスク容量 64 GB

- NDO を Cisco Service Engine または Cisco Nexus Dashboard コンピューティングクラスタ上にアプリケーションとして展開する場合の物理サーバーや仮想マシンの要件については、Service Engine と Nexus Dashboard のドキュメントを参照してください。

マルチサイト インフラストラクチャの Day-0 構成

個別のファブリック間にマルチサイト接続を展開する際の推奨されるベストプラクティスは以下のとおりです。

- スパインノードをサイト間ネットワークに接続する物理インターフェイスは、すべて Cisco Nexus EX プラットフォーム（またはそれ以降）のラインカード上にある必要があります。第 1 世代のスパインノードはサポートされていないため、ファブリック内の通信にのみ使用できます。
- 一般的に言えば、マルチサイト展開における APIC ファブリック ID は、マルチサイトドメインに属するファブリックごとに一意にすることも、複数のサイトで同じ値を使用することもできます（これは、同じマルチサイトドメインにファブリック ID が重複するブラウンフィールド ファブリックを複数追加するのに便利で、欠かせない場合もあります）。ただし、考慮が必要なシナリオが 2 つあります。これらは、ファブリック ID の定義方法の選択に影響を与える可能性があります。

  a. 共有 GOLF（Cisco ACI リリース 3.1(1) 以降サポート）の展開で、具体的には自動 RT が有効で、すべてのサイトが、同じ BGP ASN に属している場合、同じマルチサイトドメインに属する各サイトに異なるファブリック ID を展開することが必須です。前述の 2 つの条件のいずれかが成立しない場合（つまり、自動 RT が無効か、ファブリックが、異なる BGP ASN に属している場合）、共有 GOLF の設計であっても、すべてのサイトに同じファブリック ID を展開しても何ら問題ありません。

  b. マルチサイトドメインに属するファブリックが、異なるファブリック ID を使用し、ACI BD 上で IGMP スヌーピングクエリアを構成する必要がある場合、CSCvd59276 で報告された問題が発生する可能性があります。上記の DDTS に記述されているように、同じファブリック ID を使用すると問題が解決します。または、マルチサイトドメインのファブリックの 1 つで他より大きなクエリア IP アドレスが設定されている限り、ファブリック ID を異なるままにしておくことも可能です。

  注：IGMP スヌーピングクエリアの構成が必要になるのは、ブリッジドメインで PIM が有効になっていない環境でエンドポイント間の L2 マルチキャスト トラフィック フローをサポートする場合に限られます。ブリッジドメインで PIM を有効にすると、SVI で IGMP スヌーピングクエリア機能が自動的にオンになるため、クエリアの明示的な構成は必要ありません。また、外部ネットワーク インフラストラクチャにクエリアがすでに展開されている場合は、IGMP スヌーピングクエリアの構成は必要ありません。

- Cisco ACI マルチサイトのサイト ID またはサイト名は、すべてのサイトで一意になっている必要があります。このパラメータは Cisco Nexus Dashboard Orchestrator で直接構成され、割り当て後に変更することはできません。サイト ID を再構成するには、クリーンワイプを実行して工場出荷時のデフォルトに戻す必要があります。

  注：Cisco ACI マルチサイト ID は、APIC で割り当てられた Cisco ACI ファブリック ID とは異なります。

- 重複する TEP プールが展開された ACI ファブリック間にサイト間接続を確立することは可能ですが、グリーンフィールド展開でのベストプラクティスの推奨事項は、同じマルチサイトドメインに属する各ファブリックに、独立した TEP プールを割り当てることです。いずれの場合も、ISN への TEP プールのアドバタイズを除外するよう、各サイトのスパインに接続された ISN ルータの最初のレイヤを構成することを強くお勧めします。
- サイト間 L3Out 機能を有効にするために Cisco Nexus Dashboard Orchestrator で外部 TEP プールを割り当てることができます。この TEP プールのマスクは、/22 から /29 までの範囲にする必要があります。（必要な場合は）複数の外部 TEP プールを定義できます。それらの間に隣接関係は必要ありません。

外部 TEP プールが APIC から特定の ACI ファブリックにすでに割り当てられている場合（たとえば、リモートリーフを展開した結果）、ACI ファブリックがマルチサイトドメインに追加されるときに、そのプールが NDO に自動的にインポートされることに注意してください。

- マルチサイトのコントロールプレーンとデータプレーンのすべての TEP アドレスは、ISN を介した外部ルーティングが可能になっている必要があります。
- 以下の 4 種類の TEP アドレスを設定する必要があります。
  - マルチポッドデータプレーン TEP：Cisco ACI リリース 3.0(1) の各 APIC クラスタで直接このアドレスを設定します。Cisco ACI マルチサイトアーキテクチャでマルチポッドファブリックが最初にサポートされていない場合でも、リリース 3.2(1) より前のすべての ACI ソフトウェアリリースで、この構成が必須です。Cisco ACI リリース 3.2(1) 以降、単一ポッドファブリックをマルチサイトドメインに接続する場合は、この構成が必要なくなりました。
  - オーバーレイマルチキャスト TEP（O-MTEP）：このアドレスは、ファブリック全体に入力レプリケーションモードで送信される BUM トラフィックの宛先 IP アドレスとして使用されます。
  - オーバーレイユニキャスト TEP（O-UTEP）：このアドレスは、各ポッドのエニーキャスト VTEP アドレスとして使用されます。マルチポッドファブリックのポッドごとに 1 つのアドレスを設定します。このアドレスは、ローカルスパインノードで、リモートエンドポイントに到達するためのネクストホップとして使用されます。
  - MP-BGP EVPN Router-ID（EVPN-RID）：このアドレスは、MP-BGP のサイト間セッションを形成するために使用されます。スパインノードごとに一意の IP アドレスを 1 つ定義します。
- マルチポッドとマルチサイトを一緒に展開する場合、同じ EVPN-RID アドレスを使用して、異なるポッド（同じファブリックに属する）のスパイン間、および異なるサイト（異なるファブリックに属する）のスパイン間に EVPN 隣接関係を確立できます。O-UTEP、O-MTEP、マルチポッドデータプレーン TEP の各アドレスは、常に一意である必要があります。
- 専用の IP 範囲を設定してこれらの IP アドレスを割り当て、それらの /32 プレフィックスすべてを全サイトにアドバタイズすることが、ベストプラクティスの推奨事項です。必要に応じて、1 つのサイトで使用されるすべての /32 プレフィックスを集約し、集約ルートのみをマルチサイトドメインに属するリモートファブリックに送信することもできます。その場合、受信した集約ルートをファブリック内部の IS-IS コントロールプレーンに再配布できるようにするには、すべてのリモート APIC ドメインで追加の構成手順が必要になります。
- 1 つの Cisco ACI ポッドで少なくとも 2 つのスパインノードをマルチサイト BGP-EVPN ピアリングに割り当ててください（冗長性のため）。すべてのスパインノードが BGP-EVPN ピアである必要はないことに注意してください。
- サイトがマルチポッドファブリックとして展開されている場合、2 つのスパイン（異なるポッド内）を BGP スピーカーとして定義してください（つまり、NDO からこれらのスパインで BGP を有効にします）。残りのスパインは BGP フォワーダのままにしておきます。BGP スピーカーだけが、リモートサイトのスピーカーと BGP EVPN 隣接関係を確立します。
- ルートリフレクタの代わりにフルメッシュの BGP-EVPN ピアリングを使用することをお勧めします。このほうが単純なアプローチであるためです。BGP-EVPN は、フルメッシュの iBGP ピアリングと eBGP ピアリングを自動的に形成します。

- ルートリフレクタを使用する場合、同じ自律システム内の iBGP セッションのみにこれが適用され、eBGP では、異なる自律システム間で引き続きフルメッシュが使用されます。スパインノードは、両方のタイプのピアリングを同時にサポートできます。
- N サイトの高可用性を備えたルートリフレクタ（RR）を展開する場合、すべてのサイトに RR インスタンスを 1 つずつ展開するのではなく、サイトの 1 つのスパインノードに RR インスタンスを 1 つ展開し、これを 3 つのサイトで展開することで、N サイトの高可用性の要件を満たすことができます。
- マルチサイト ISN では、BGP と OSPF の一般設定にデフォルトを使用してください。
- サイト間 BGP-EVPN セッションの送信元インターフェイスに MP-BGP EVPN ルータ ID が設定され、インフラ L3Out 接続が、正しいループバックコネクタに割り当てられていることを確認してください。
- セキュア BGP パスワードが、さまざまなサイトで一致することを確認してください（設定されている場合）。
- BGP コミュニティのフォーマットの例は **extended:as2-nn4:4:15** です。

Cisco ACI マルチサイト設計の一般的なベストプラクティス

以下は Cisco ACI マルチサイト設計の試行と実際の展開から学んだ教訓ですので留意してください。

- WAN 接続が GOLF を介している場合は、以下の 2 つのシナリオを考慮する必要があります。
  - シナリオ 1：サイト 1 に個別の非ストレッチ BD1、サブネット 1、GOLF L3Out-1 接続があり、サイト 2 に個別の非ストレッチ BD2、サブネット 2、GOLF L3Out-2 接続がある場合、各 GOLF L3Out 接続がそのサイトのブリッジドメインサブネットをそのサイトの GOLF ルータにアドバタイズするため、ホストルーティングは必要ありません。
  - シナリオ 2
    - BD1 とサブネット 1 がサイト 1 と 2 に拡張されています。
    - レイヤ 2 ストレッチフラグが BD1 に対して有効になっています。
    - サイト間 BUM トラフィック転送は有効でも無効でも構いません。
    - さらに、各サイトには個別のローカル GOLF L3Out 接続があり、この GOLF L3Out 接続を介してサブネットがそのサイトの GOLF ルータにアドバタイズされます。
    - WAN のサブネットでは、2 つの GOLF ルータに向かう等コストマルチパス（ECMP）が形成されます。
    - IPN を通過する GOLF トラフィックの最適でないルーティングはサポートされません。つまり、トラフィックを GOLF ルータから特定のサイトに配信し、さらに別のサイトにリダイレクトして、リモートの宛先エンドポイントに到達させることはできません。したがって、すべてのストレッチブリッジドメインに対してホストルートアドバタイズを有効にすることが必須です。
- サイト間 VXLAN トンネルは ISN を通過する必要があり、別のサイトを通過することはできません。したがって、ノードまたはリンクで障害が発生したシナリオでも 2 つのサイトが常に ISN を介して接続されるように、ISN に十分な冗長性を持たせる必要があります。
- Cisco ACI リリース 4.2(1) より前では、各サイトにローカル L3Out 接続を展開する必要があります。
  - サイトが、別のサイトに展開されたエンドポイントに L3Out ルーティングサービスを提供することはできません。
  - WAN エッジルータのペアをサイト間で共有できます（ボーダーリーフノードでの従来型の L3Out 接続）。

    - サイト間の共有 WAN エッジルータは、GOLF L3Out 接続が展開されている場合、Cisco ACI マルチサイトリリース 3.1(1) からサポートされます。
- Cisco ACI リリース 4.2(1) 以降、サイト間 L3Out 機能により、サイトが、別のサイトにあるエンドポイントに L3Out サービスを提供することが可能になりました。
    - サイト間 L3Out の有効化によって既存のインバウンドとアウトバウンドのトラフィックフローが受ける可能性のある影響を十分に理解してください。
- テナントや VRF インスタンスは、それぞれ個別の L3Out 接続を持つことができます。1 つの L3Out 接続が複数の VRF インスタンス間で共有される共有 L3Out 接続がマルチサイトでサポートされますが、Cisco ACI リリース 4.0(1) 以降が必須です。
- マルチポッドファブリックは、Cisco ACI リリース 3.2(1) 以降、Cisco ACI マルチサイトアーキテクチャのサイトとしてサポートされます。
- ドメイン（VMM と物理）の定義と関連付けは、サイトレベルで実行されます。
- Cisco Nexus Dashboard Orchestrator からサイトにプッシュされたポリシーは、APIC でローカルに変更できます。サイトに導入されたポリシーが Nexus Dashboard Orchestrator テンプレートで指定されたポリシーと異なる場合、Nexus Dashboard Orchestrator に警告が表示されます。
- EPG 内分離またはマイクロセグメンテーションが構成されている場合、WAN でのサービス品質（QoS）マーキングはサポートされません。
- サイトで QoS ポリシーが構成されていない場合、ISN を通過する VXLAN パケットの外側の IP ヘッダーにある DSCP のデフォルト値が 0（ゼロ）に設定されます。ISN で QoS が適切に処理されるようにするには、スパインノードで QoS DSCP マーキングポリシーを構成する必要があります。
- マルチサイト展開を有効にするには、少なくとも 1 つのスパインインターフェイスを ISN に接続することが必須です。
- スパインポートが ISN に接続されていて、ピアリングが無効になっている場合、そのスパインノードではデータプレーンのみが有効になります。
- スパインポートが ISN に接続されていて、ピアリングが有効になっている場合（つまり、スパインが BGP スピーカーとして構成されている場合）、コントロールプレーン MP-BGP EVPN セッションがスパインノード間で形成されます。さらに、ピアリングが有効になっているサイト間にも、MP-BGP EVPN ルータ ID を介して形成されます。
- iBGP、eBGP、またはハイブリッド（iBGP と eBGP）を使用したルートリフレクタまたはフルメッシュのシナリオにおける MP-BGP EVPN コンバージェンスは、中規模展開を模したラボ環境では通常 1 秒以下です。ただし、実際の中規模展開では、コンバージェンスが 5 秒以下になるのが一般的です。これは、WAN などの外部要因によるものです。
- Cisco Nexus Dashboard Orchestrator は、APIC からポッドとサポートされているスパインのラインカード情報を検出し、インフラ構成を更新します。
- 管理対象オブジェクトから作成されているサイトの APIC から、入手可能なインフラ構成がすべて取得され、Cisco Nexus Dashboard Orchestrator の構成に自動入力されます。
- APIC では、図 127 に示すように、マルチサイトのインフラ L3Out 接続は、[インフラ（infra）] テナントの下に作成され、[インターサイト（intersite）] という名前が付けられます。

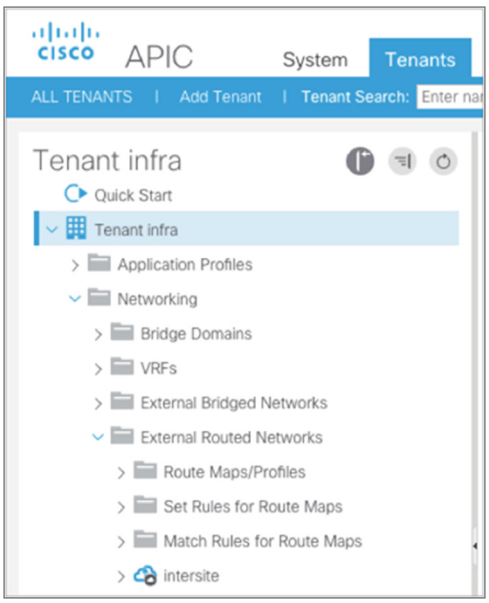


図 127.
Cisco APIC でのサイト間 L3Out の作成

- サイトがマルチサイトドメインに追加されると、BGP スピーカーとして構成されているスパインで BGP RID が変更されます。これにより、スパイン（ファブリックの RR として構成されている）とリーフノードの間のファブリック内 VPNv4 隣接関係が再確立され、その結果、垂直方向通信に影響が及ぶ可能性があります。したがって、マルチサイトドメインへのサイトの追加はメンテナンス期間中に実行することをお勧めします。
- Cisco ACI マルチサイト展開からサイトを削除すると、そのコントロールプレーンとデータプレーンの機能が無効になります。ただし、Cisco Nexus Dashboard Orchestrator はインフラ構成を保持し、そのサイトが再度追加される場合、構成を自動的に入力します。必要に応じて、APIC コントローラから直接インフラ構成を削除する必要があります。

まとめ

新しい Cisco ACI マルチサイトアーキテクチャを使用すると、個別の Cisco ACI ファブリックを相互接続できます。各ファブリックは、それぞれの APIC クラスタによって管理され、AWS のリージョンに相当します。

MP-BGP EVPN オーバーレイ コントロール プレーンと VXLAN データプレーンのカプセル化を使用すると、ファブリックにまたがるレイヤ 2 とレイヤ 3 のマルチテナント通信を簡単に確立できます。ファブリックを相互接続するネットワーク インフラストラクチャでは、アンダーレイ ルーティング サービスのみが必要です。サイト間 VXLAN データプレーンでは、ネットワークとポリシーの情報（メタデータ）も伝送されます。これを使用することで、エンドツーエンドでポリシードメインを拡張できます。

Cisco Nexus Dashboard Orchestrator の導入により、管理が一元化され、相互接続されたファブリックの正常性のモニタリング、MP-BGP EVPN コントロールプレーンにおける隣接関係の確立に必要な Day-0 構成タスクの実行、さまざまな APIC ドメインに導入するサイト間ポリシーテンプレートの定義が可能になります。

Cisco ACI マルチサイトの設計は、既存の Cisco ACI マルチポッドアーキテクチャを補完するものであり、置き換えるものではありません。ユースケースやビジネス要件によっては、両方のオプションの展開が必要です。図 128 に示すように、これらのオプションには基本的な相違点があります。

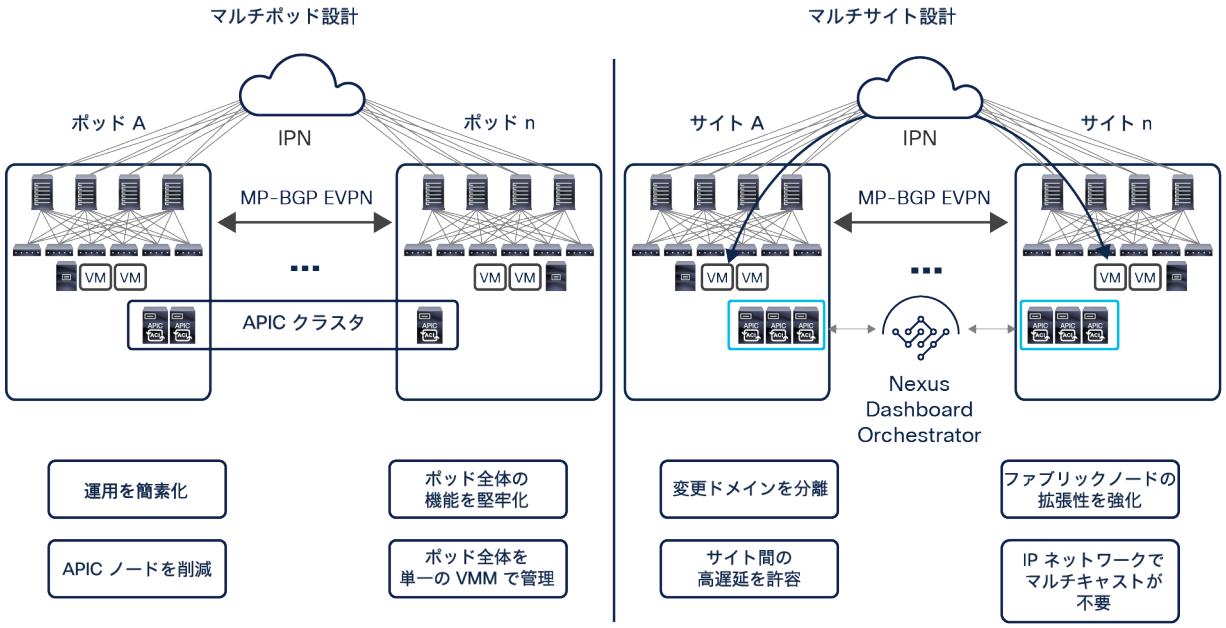


図 128.
Cisco ACI マルチポッドと Cisco ACI マルチサイトのアーキテクチャにおける相違点の概要

マルチポッド設計では、単一の APIC クラスタが相互接続されたすべてのポッドを管理するため、運用がよりシンプルになります。また、この設計により、単一ポッドファブリックでサポートされているすべての Cisco ACI 機能がマルチポッドファブリックでも使用できるようになります（サービスグラフの定義、共有 L3Out 接続、複数のポッドにまたがる単一の VMM ドメインの作成など）。

Cisco ACI マルチサイトアーキテクチャは、相互接続されたファブリック間で、完全な障害ドメイン分離と変更ドメイン分離を実現します。さらに、個々のサイトに接続できる Cisco ACI のノードとエンドポイントの数の合計で見たアーキテクチャの全体規模を拡大できるようになります。最後に、Cisco ACI マルチサイトの設計により、ファブリックを相互接続するレイヤ 3 インフラストラクチャにマルチキャストを展開する必要もなくなります。スパインスイッチで実行されるヘッドエンド レプリケーションによって、BUM トラフィックを要求元のストレッチブリッジドメインすべてに向けて、ファブリックにまたがって送信できるようになるためです。

Cisco ACI リリース 3.2(1) 以降、1 つ以上の Cisco ACI マルチポッドファブリックをマルチサイトアーキテクチャの「サイト」として展開できます。この 2 つのアーキテクチャオプションを組み合わせることで、データセンターネットワークを相互接続するための要件を満たす柔軟性と豊富な機能が実現します。

詳細情報

このホワイトペーパーで説明した Cisco ACI マルチサイトアーキテクチャやその他のアーキテクチャの詳細は、次のリンクにあるドキュメントを参照してください。

- ACI マルチポッド ホワイトペーパー
  https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-737855.html
- ACI マルチポッド構成ホワイトペーパー
  https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-739714.html
- ACI マルチポッドおよびサービスノード統合ホワイトペーパー
  https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-739571.html
- ACI ファブリック向け Cisco マルチサイト導入ガイド
  https://www.cisco.com/c/ja_jp/td/docs/dcn/whitepapers/cisco-multi-site-deployment-guide-for-aci-fabrics.html
- ACI マルチサイトおよびサービスノード統合ホワイトペーパー
  https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-743107.html
- ACI マルチサイト トレーニング ビデオ
  https://www.cisco.com/c/en/us/solutions/data-center/learning.html#~nexus-dashboard
- ACI リモート リーフ アーキテクチャ ホワイトペーパー
  https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-740861.html

付録 A：外部 RP を使用したマルチサイトのレイヤ 3 マルチキャスト

一連の機能（IGMP スヌーピング、COOP、PIM）が Cisco ACI 内で連携して、ACI リーフノードに適切な (*,G) ステートと (S,G) ステートを作成します。図 129 は、外部 RP の使用を必要とする PIM-ASM シナリオでのこれらの機能の動作を示しています。

注： エニーキャスト RP ノードを外部ネットワークに展開し、RP 機能を冗長化できます。その場合、送信元情報を同期するために、異なる RP 間で MSDP または PIM を使用できます。

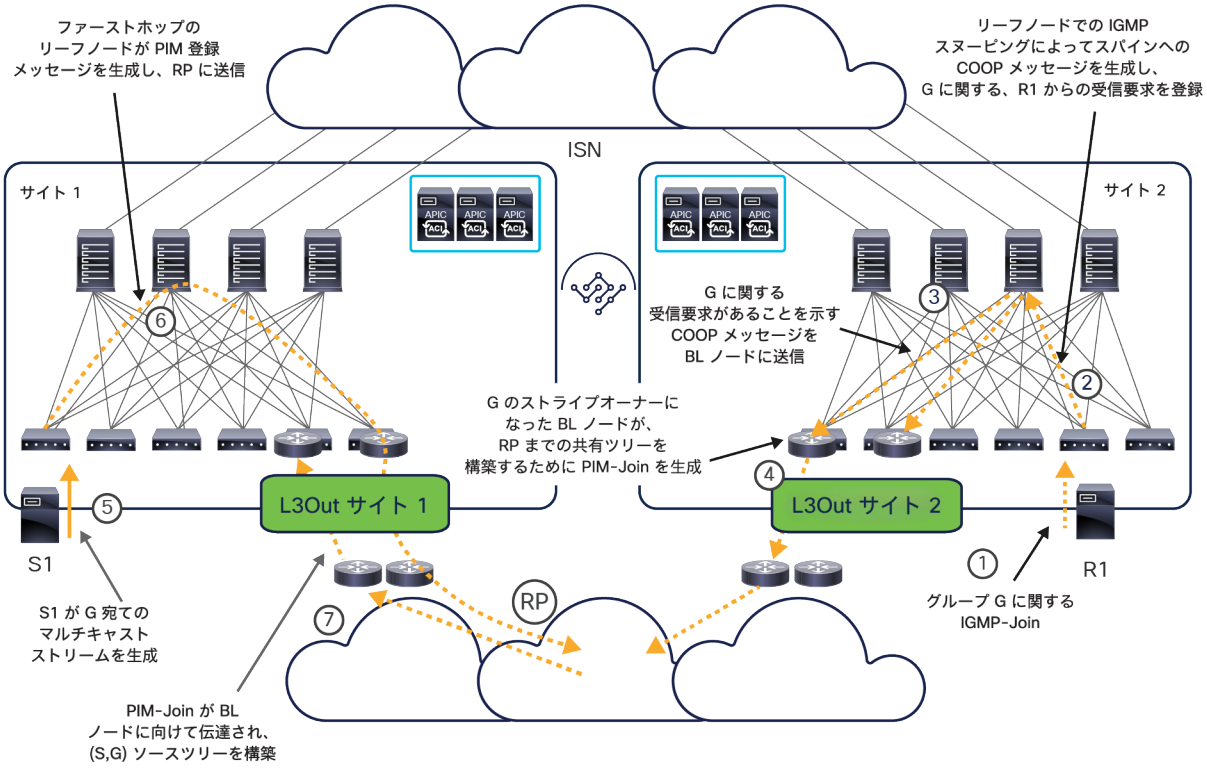


図 129.
PIM-ASM ユースケースにおけるコントロールプレーンでの動作

受信者が Cisco ACI ファブリックに接続されているとき：

1. 受信者が、グループ G に向けられたマルチキャストトラフィックに関する受信要求を宣言するために、IGMP-Join を発信します。
2. 受信者が接続されている Cisco ACI リーフノードが、この IGMP-Join を受信します。リーフは、ローカルに接続された受信者の受信要求を登録し（(*,G) ローカルエントリを作成します）、COOP メッセージを生成してスパインに同じ情報を提供します。
3. スパインは、グループ G に関する受信者の受信要求を登録し、COOP 通知を生成してローカル ボーダー リーフ ノードにこの情報を伝えます。
4. ボーダーリーフノードの 1 つがマルチキャストグループ G の「ストライプオーナー」として選択され、PIM-Join を生成して RP に向かうパス上にある外部ルータへ送信します。外部 PIM ルータがこの動作を繰り返し、RP までの共有ツリーを構築します。

5. この時点で、マルチキャストの送信元 S1 が接続され、グループ G 宛てのトラフィックのストリーミングを開始します。
6. 送信元が接続されているファーストホップのリーフノードが、(S1,G) ローカルステートを作成し、RP に向けて PIM 登録メッセージを送信します。PIM 登録パケットは、IP ヘッダーに PIM プロトコル番号 103 が設定されたユニキャストパケットであり、Cisco ACI ファブリックを介してローカル ボーダー リーフ ノードに転送されます。現在の実装では、PIM パケットを許可するデフォルトルールが Cisco ACI ノードに構成されているため、このコントロールプレーンでのやり取りを成功させるためにコントラクトは必要ありません。
7. 次に、RP がサイト 1 のボーダーリーフノードに向けて PIM-Join メッセージの送信を開始し、(S1,G) 送信元ツリーが構築されます。これは、S1 の BD が拡張されていない場合です。S1 の BD が拡張されている場合、Join メッセージはサイト 1 に送られることもサイト 2 に送られることもあります。

重要な点を指摘しておくと、外部 RP を使用するのは、異なるサイトに接続された送信元と受信者の間でレイヤ 3 マルチキャストが転送されるときに RP 自体を経由する必要があるからではありません。このレイヤ 3 マルチキャストの水平方向通信は、ISN を通過する VXLAN データプレーンを使用して処理されます。ただし、アクティブな L3Out 接続の存在は、コントロールプレーンでの上記のやり取りを可能にするために必須です。

図 130 は、マルチキャストの送信元がサイト 1 に接続され、受信者がローカルサイト、リモートサイト、外部ネットワークに接続されていると仮定したときのデータプレーンでの転送動作を網羅しています。

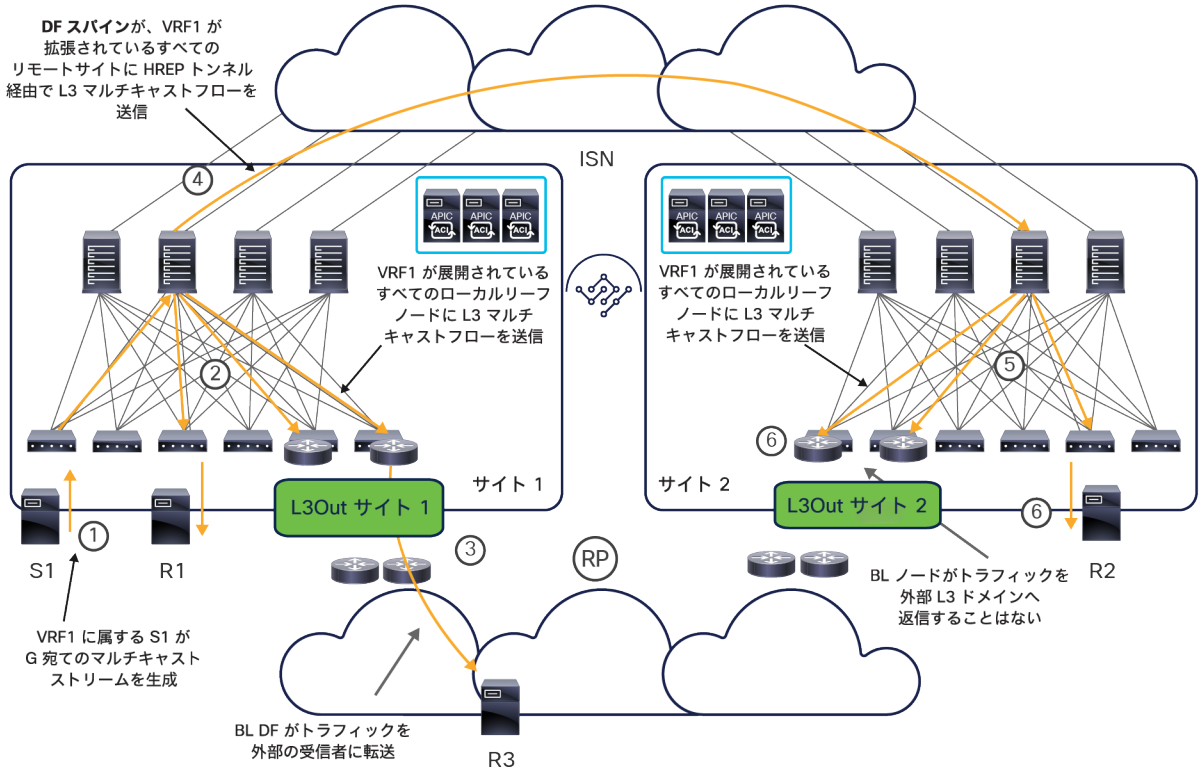


図 130.
内部送信元と内部/外部受信者の間で転送されるレイヤ 3 マルチキャスト

- VRF1 に属するマルチキャストの送信元がリーフノードに接続され、グループ G 宛てのマルチキャストストリームを生成します。
- 送信元が接続されているリーフノードがこのマルチキャストストリームを受信します。リーフはトラフィックをカプセル化し、VRF1 に関連付けられた GIPo に送信します。次に、マルチキャストトラフィックがサイト内で転送され、VRF1 が展開されているすべてのリーフノード（ボーダーリーフノードを含む）に到達します。ASM の場合、この動作には、このグループのための有効な RP が外部ネットワークに展開されている必要があることに注意してください。リーフは最初に登録メッセージを RP に送信し、次に VRF GIPo に向かうマルチキャストトラフィックを転送します。利用可能な RP がない場合、FHR リーフはマルチキャストトラフィックを転送できません。
- 外部ルータから PIM-Join を受信したボーダーリーフノードが、マルチキャストストリームを外部ネットワークに転送します。この時点で、トラフィックは、最短パスツリー（SPT）のスイッチオーバー切り替えが発生したかどうかに応じて、ダイレクトパスをたどるか、または RP を介してマルチキャストの外部の受信者 H3 に到達します。
- 同時に、サイト 1 の VRF1 の 指定フォワーダ（DF）として選択されたスパインが、VRF1 が拡張されているすべてのリモートサイトに向けて、VXLAN カプセル化が行われたマルチキャストストリームを転送します。この転送には、入力レプリケーション機能（HREP トンネル）が利用されます。このトラフィックに使用される VXLAN 宛先アドレスは、各サイトに関連付けられたオーバーレイマルチキャスト TEP アドレスです。これは、サイト間の L2 BUM トラフィック転送にもすでに使用されています。
- 各リモートサイト内のスパインの 1 つがマルチキャストトラフィックを受信し、VRF1 GIPo に関連付けられたツリーに従ってローカルサイト内に転送します。VRF1 に関連付けられた GIPo アドレスは、別の APIC コントローラによって割り当てられるため、サイト 1 で使用されるものとは異なる可能性があることに注意してください。GIPo はそれが定義されているローカルファブリック内に限って使用されるため、これは実際には重要ではありません。
- ボーダーリーフノードを含む、VRF1 が展開されているすべてのリーフノードがこのストリームを受信します。ボーダーリーフノードは、マルチキャストストリームを受信することで、ストリームがマルチサイトデータパスを介して配信されたことを知ります。そこで、グループ G の指定フォワーダになっているボーダーリーフが、（図 129 に示す）コントロールプレーンでのアクティビティに基づいて構築された、外部 RP に向かう共有ツリーをプルーニングし、重複したトラフィックが外部受信者に送信されないようにします。

一方、図 131 は、マルチキャストの送信元が外部レイヤ 3 ネットワークに接続され、受信者が Cisco ACI ファブリック内に展開されているシナリオを示しています。

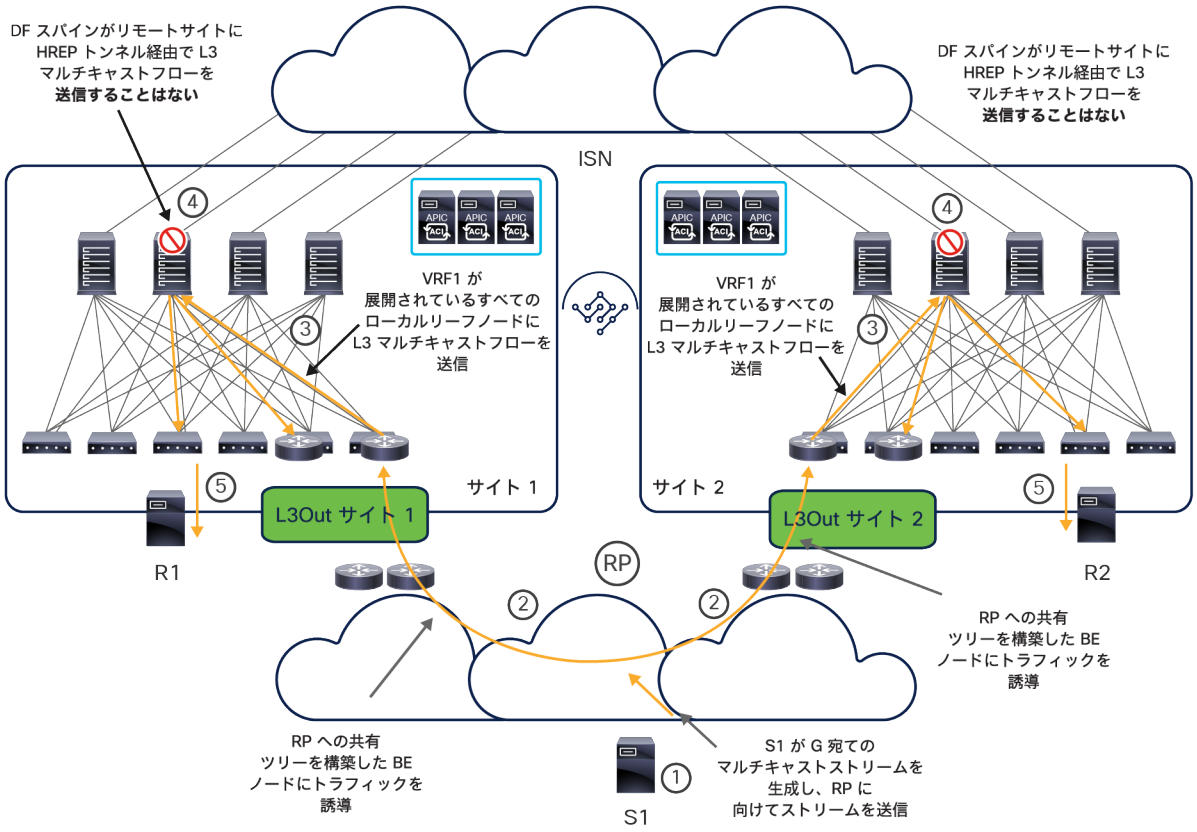


図 131.
外部送信元と内部/外部受信者の間で転送されるレイヤ 3 マルチキャスト

この場合、受信者が接続されているすべてのファブリックのボーダーリーフノードがマルチキャストストリームを誘導し、ローカルの Cisco ACI ファブリック内に転送して、受信者が受信できるようにします。このとき、ストリームを取得した DF スパインがこれを複製してリモートサイトに送信することはありません。これは、リモートの受信者がトラフィックを重複して受信しないようにするためです。

注： PIM SSM を展開する場合も、上図に示すようなデータパスの動作になります。唯一の違いは、SSM の場合、外部 RP を定義する必要がないことです。実際、SSM シナリオでは、受信者が IGMPv3 を使用して特定の送信元からのマルチキャストストリームに関する受信要求を宣言します。これにより、受信者と送信元の間に直接マルチキャストツリーが作成されます。

付録 B：マルチ DC オーケストレーション サービスの以前の展開オプション

このホワイトペーパーですでに述べたように、オーケストレーション サービスの現在推奨されている導入モデルは、Nexus Dashboard コンピューティングクラスタ上でこれを実行することです（したがって、Nexus Dashboard Orchestrator（NDO）になります）。以下のセクションでは、以前に提供されていた Cisco Multi-Site Orchestrator（MSO）の導入モデルについて説明します。これらは、ユーザーの実稼働展開から徐々に消えていくと思われます。

VM ベースの MSO クラスタを直接 VMware ESXi 仮想マシンに展開

これは、Cisco Multi-Site Orchestrator の最初のリリースからサポートされている導入モデルです（図 132）。ただし、3.1(1) が、この展開オプションがサポートされる最後の MSO リリースです。

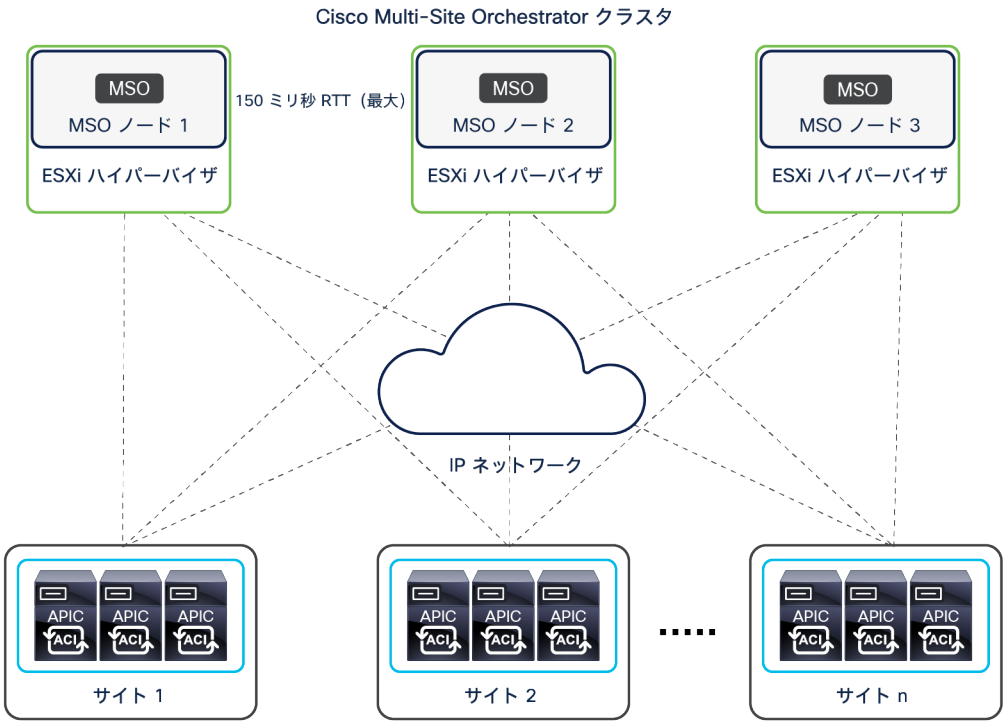


図 132.
VMware ESXi ホストで実行される VM ベースの MSO クラスタ

この展開オプションでは、Cisco Multi-Site Orchestrator の各ノードが VMware vSphere 仮想アプライアンスにパッケージ化されます。高可用性を実現するには、Cisco Multi-Site Orchestrator の各仮想マシンを個別の VMware ESXi ホストに展開する必要があります。3 つの仮想マシンすべてが 3 つの異なる ESXi ホスト上にあってクラスタを形成することで、シングルポイント障害が解消されます。

Multi-Site Orchestrator クラスタの作成に 3 つの仮想マシンを使用する構成では、仮想マシンを 1 つ失っても、クラスタは引き続き完全に機能することができます。しかし、仮想マシンを 2 つ失うと、クラスタは非アクティブになります。各仮想マシンを個別の ESXi ホストに展開することが推奨されるのはそのためです。

クラスタ内の Multi-Site Orchestrator ノード間の遅延の許容されるラウンドトリップ時間（RTT）は最大 150 ミリ秒（ms）です。そのため、必要に応じて仮想マシンを別々の物理的なロケーションに配置し、地理的に分散させることができます。Multi-Site Orchestrator クラスタは、セキュアな TCP 接続を介して各サイトの APIC クラスタと

通信します。すべての API 呼び出しは非同期です。現在許容されている最大 RTT 距離は、Multi-Site クラスタと各サイトの APIC クラスタの間で 1 秒です。

各仮想マシンに求められる VMware vSphere 仮想アプライアンスの要件は、以下に示すように、展開された Cisco Multi-Site Orchestrator のリリースによって異なります。

Cisco Multi-Site Orchestrator リリース 1.0(x) の場合：

- VMware ESXi 5.5 以降
- 最小要件：仮想 CPU（vCPU）4 個、メモリ 8 Gbps、ディスク容量 50 GB

Cisco Multi-Site Orchestrator リリース 1.1(x) の場合：

- VMware ESXi 6.0 以降
- 最小要件：仮想 CPU（vCPU）4 個、メモリ 8 Gbps、ディスク容量 50 GB

Cisco Multi-Site Orchestrator リリース 1.2(x) 以降の場合：

- VMware ESXi 6.0 以降
- 最小要件：仮想 CPU（vCPU）8 個、メモリ 24 Gbps、ディスク容量 100 GB

MSO をアプリケーションとして Cisco Application Services Engine（CASE）クラスタに展開

このオプションは、Cisco Multi-Site Orchestrator リリース 2.2(3) から利用可能で、アプリケーション（.aci 形式）を Cisco Application Services Engine（CASE）にインストールします。ただし、3.1(1) が、この展開オプションがサポートされる最後の MSO リリースです。オーケストレーションの新しいリリースを実行したい場合、Nexus Dashboard コンピューティングクラスタで実行される Orchestrator サービスに移行する必要があります。Nexus Dashboard は CASE が進化したものです。単純なソフトウェアアップグレードにより、CASE コンピューティングクラスタを Nexus Dashboard コンピューティングクラスタに変換できることに注意してください。これにより、すでに購入した物理サーバーを再利用できます。

図 133 に示すように、Multi-Site Orchestrator は、Cisco Application Services Engine の 3 つの異なるフォームファクタにインストールできます。

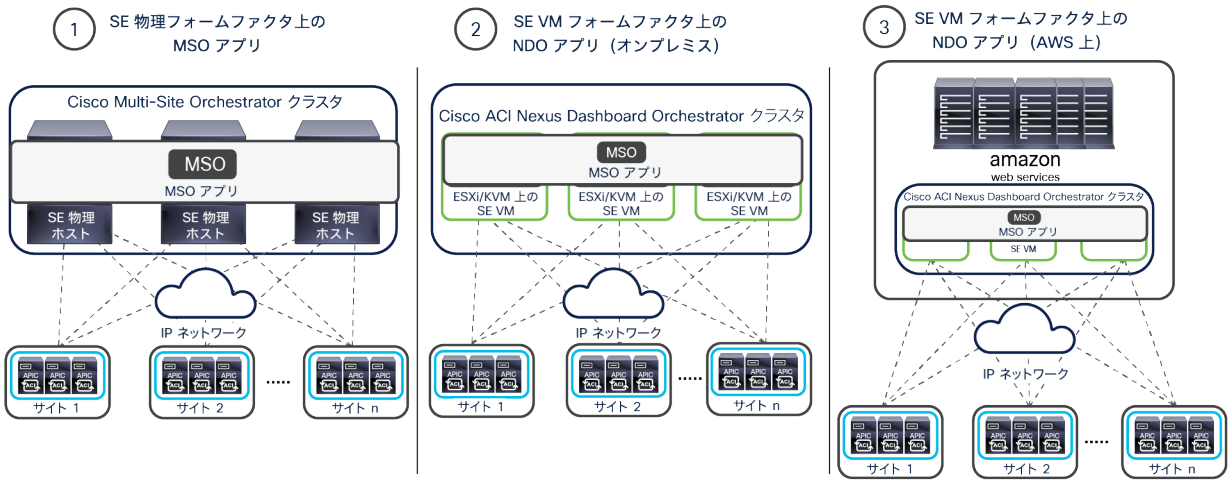


図 133.
Cisco Application Services Engine の異なるフォームファクタに展開された MSO クラスタ

- 物理フォームファクタ：基本的に、サーバーノードが 3 つある物理クラスタを構築し、それらに MSO マイクロサービスを展開します。Cisco Application Services Engine の特定の .iso ファイルを Cisco.com からダウンロードし、クラスタ化されたこれらのベアメタルサーバーにインストールできます。
- 仮想マシンフォームファクタ（オンプレミス展開）：仮想アプライアンスの 2 つのフレーバが CASE 用に用意されています。1 つは VMware ESXi ホストで実行され、もう 1 つは Linux KVM ハイパーバイザで実行されます。
- 仮想マシンフォームファクタ（AWS パブリッククラウド展開）：AWS 用の CloudFormation テンプレート（CFT）を使用して Cisco Application Services Engine をパブリッククラウドに展開できます（CASE 仮想アプライアンス用の .ami ファイルが利用できます）。これにより、3 つの CASE VM からなるクラスタを特定の AWS リージョンに直接展開できます。

注： いずれの CASE フォームファクタを選択した場合でも、MSO アプリケーションは 3 ノードの CASE クラスタ上に同じ方法でインストールされます。したがって、VM ベースの MSO インストールの説明で述べたクラスタの復元力に関する考慮事項はここでも有効です。また、Cisco Application Services Engine リリース 1.1.3 以降の CASE に MSO アプリケーションを展開することをお勧めします。これには、Cisco Multi-Site Orchestrator リリース 3.0(2) 以降も必要です。

図 132 に示す遅延に関する考慮事項は、CASE クラスタで実行されるアプリケーションとして MSO を展開する場合にも当てはまります。

Cisco Application Services Engine クラスタとその上で実行される MSO アプリケーションのインストールについての詳細は、以下のドキュメントを参照してください。

https://www.cisco.com/c/en/us/td/docs/data-center-analytics/service-engine/APIC/1-1-3/getting-started-guide/b_cisco_application_services_engine_getting_started_guide_release_1-1-3_x.html

https://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/aci_multi-site/sw/2x/installation/Cisco-ACI-Multi-Site-Installation-Upgrade-Guide-221.html

付録 C：マルチサイトと GOLF L3Out 接続

外部レイヤ 3 ドメインへの接続に GOLF が使用されている場合でも、図 134 に示すように、GOLF ルータの専用ペアまたは共有ペアを展開して、異なるファブリックにサービスを提供できます。

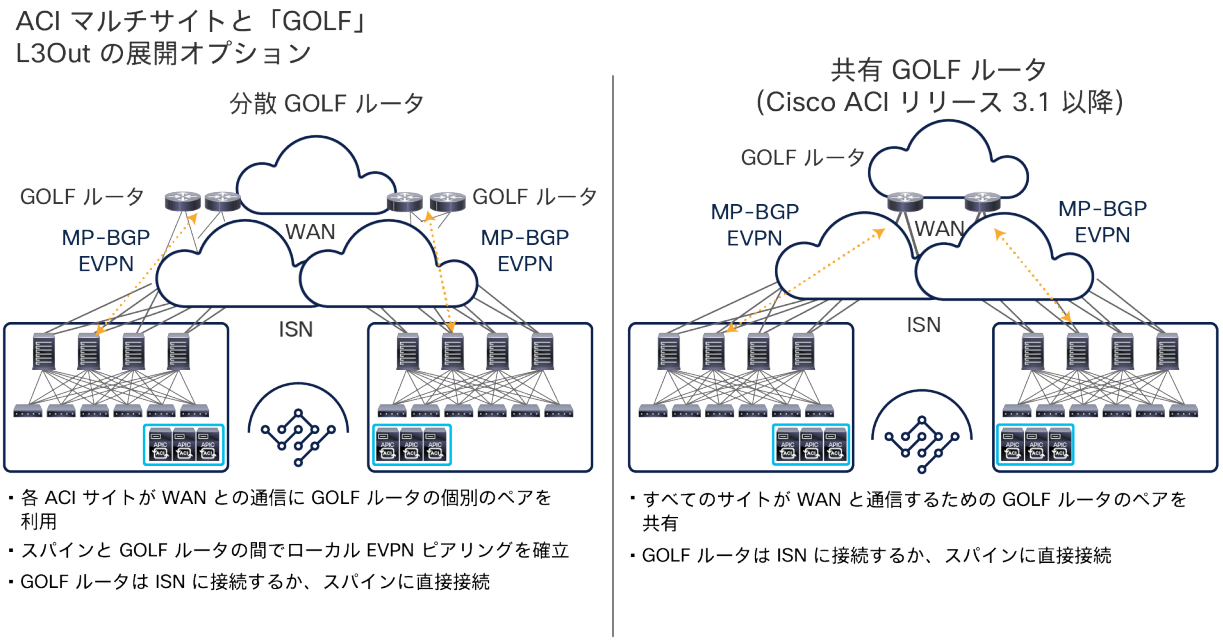


図 134.
GOLF ルータの専用ペアまたは共有ペア

注： 右側に示す共有 GOLF デバイスを使用する設計は、Cisco ACI リリース 3.1(1) 以降、サポートされています。

専用 GOLF デバイスを使用するユースケースでは、各ファブリックに展開されたスパインノードがローカル GOLF ルータと MP-BGP EVPN 隣接関係を確立します。これによって、到達可能性情報の交換と、データプレーンにおける垂直方向通信に必要な VXLAN トンネルの構築が可能になります。

専用 GOLF デバイスを使用するシナリオでは、垂直方向と水平方向の通信に共通のレイヤ 3 インフラストラクチャを使用する場合、展開の際に以下の点を考慮する必要があります。

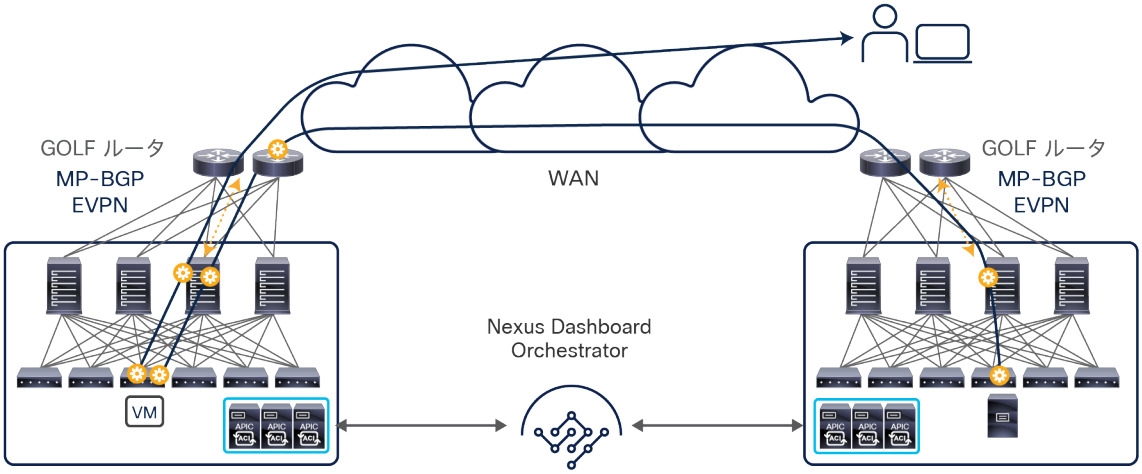


図 135.
水平方向と垂直方向のトラフィックに共通のレイヤ 3 インフラストラクチャを使用

図 135 に示すように、この場合、GOLF ルータは 2 つの役割を果たします。Cisco ACI ファブリックと WAN の間の垂直方向通信では、VXLAN カプセル化およびカプセル化解除を実行する VTEP になり、サイト間の水平方向通信では、標準のレイヤ 3 ルーティングを実行する手段になります。図 136 に示すように、2 つのアプローチが可能です。スパインノードと GOLF ルータの間に設けた同じ物理接続のセットを共有して両方のタイプのトラフィックを伝送する方法と、それぞれのタイプの通信に専用の接続を使用する方法です。

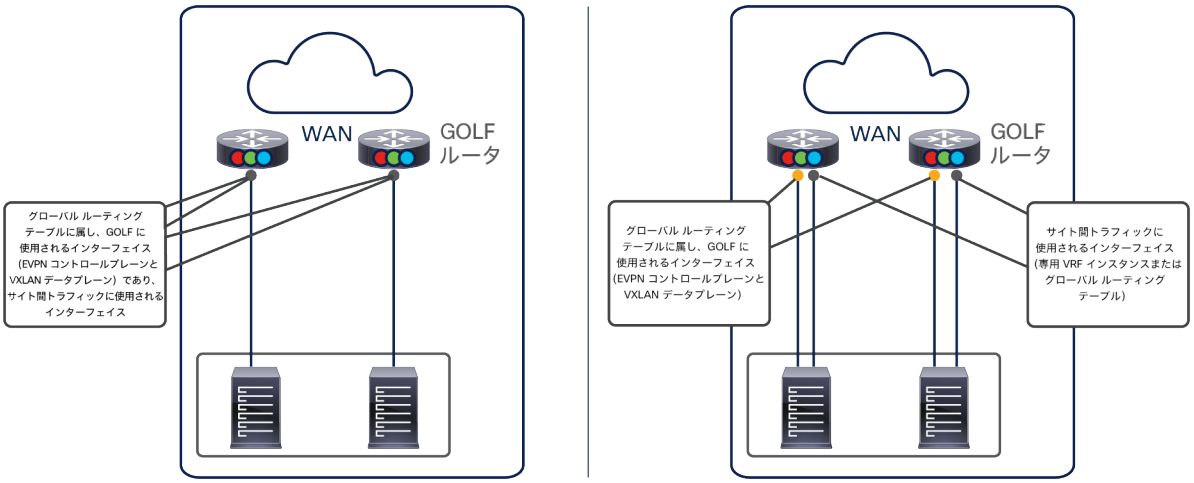


図 136.
GOLF とマルチサイトの展開における共有または専用の物理接続

スパインノードと GOLF ルータの間に MP-BGP EVPN 隣接関係を確立するには、GOLF インターフェイスがグローバル ルーティング テーブルのルーティングドメインに属している必要があることに注意してください。したがって、図 136 の左側に示すように物理接続を 1 セットのみ使用すると、マルチサイトのトラフィックがそのグローバル テーブルのルーティングドメインでルーティングされ、異なる VRF インスタンス内で転送できません。

注： マルチサイトの水平方向トラフィックを専用ルーティングドメインで伝送したい場合、グローバルテーブルから選択したルートを専用 VRF にリークすることは技術的に可能です。ただし、これを行うと、構成が複雑になり、構成を誤って問題が発生する可能性があります。

専用 VRF インスタンスでマルチサイトのトラフィックを伝送する唯一の安全な方法は、図 136 の右側に示すように、個別の物理インターフェイスを使用することです。このアプローチは、マルチプロトコル ラベル スイッチング（MPLS）VPN WAN サービスを使用する際にしばしば重要になります。これは、サイト間トラフィックがグローバル ルーティング テーブルで伝送されるのを避ける必要があるためです。

GOLF との通信やサイト間通信に共有の物理インターフェイスを使用するか専用のものを使用するかに関係なく、Cisco ACI リリース 3.2(1) より前の Cisco ACI マルチサイト設計では、GOLF（垂直方向）とマルチサイト（水平方向）の各タイプの通信に使用される 2 つの個別の L3Out 接続をインフラテナント内に定義する必要があります。両方のタイプのトラフィックに同じ物理インターフェイスを使用する場合、以下の設定ガイドラインに従う必要があります。

- 両方のインフラ L3Out 接続でスパインノードに同じルータ ID を定義します。
- 論理インターフェイスと、関連する IP アドレスの同じセットを定義します。
- 両方の L3Out 接続を同じ overlay-1 VRF インスタンスに関連付けます。
- 両方の L3Out 接続で同じ OSPF エリアを定義します。

Cisco ACI リリース 4.2(1) 以降では、GOLF とマルチサイトのトラフィックに同じインフラ L3Out を定義して使用できます。このアプローチを使用すると、構成が簡素化されます。両方のタイプの通信に物理インターフェイスの同じセットを使用する場合に適しています。

最初の Cisco ACI マルチサイトリリース 3.0(1) 以降、GOLF L3Out を展開することで、外部ネットワークにホストルートをアドバタイズできます。図 137 に示すように、また「ボーダーリーフ L3Out でのホストルートアドバタイズ」セクションで説明したように、ホストルートアドバタイズは、ブリッジドメインが別々のファブリックにまたがって拡張されている展開で役立ちます。入力トラフィックを宛先エンドポイントがあるサイトに正しくステアリングできるようになるからです。

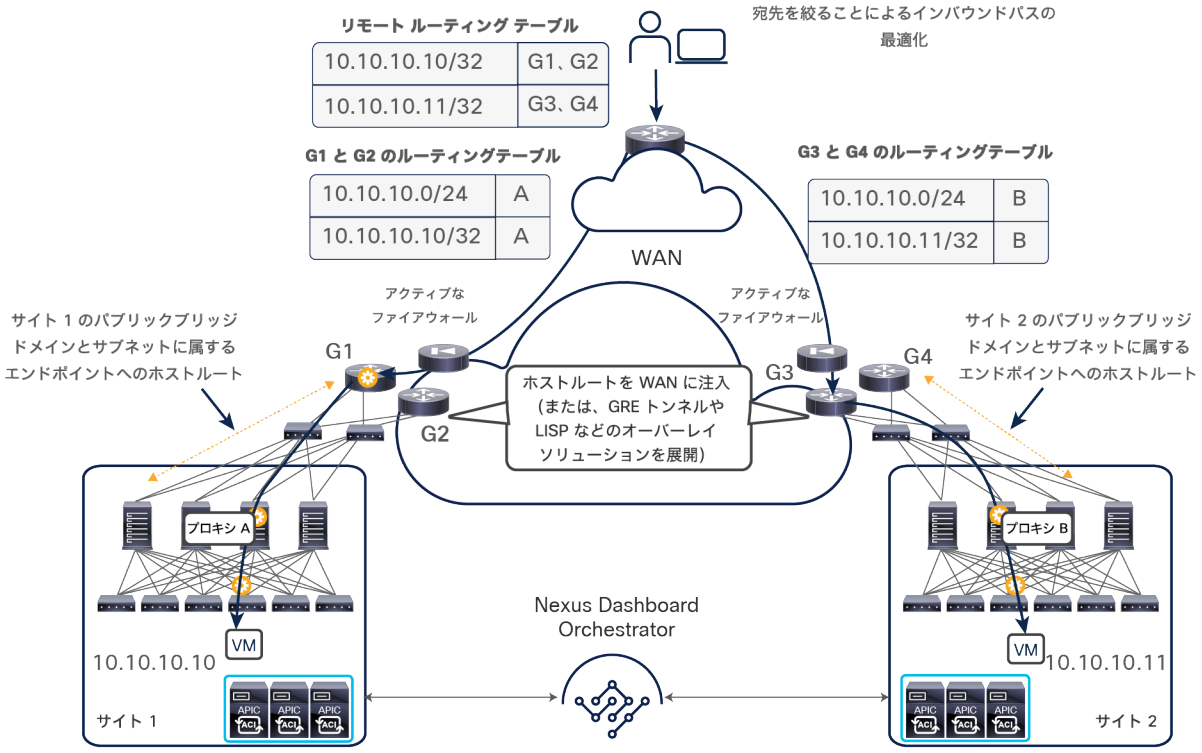


図 137.
ホストルートアドバタイズを使用した入力パスの最適化

このアプローチにより、エンドポイントと同じサイトでアクティブなローカル ファイアウォール ノードを介してトラフィックが常に送信されるようになります。

重要な点としては、ストレッチブリッジドメインがある場合にホストルートアドバタイズを有効にすることは、最適化のためだけではありません。ACI マルチサイトと組み合わせて GOLF L3Out を展開する際には必須の設定になります。

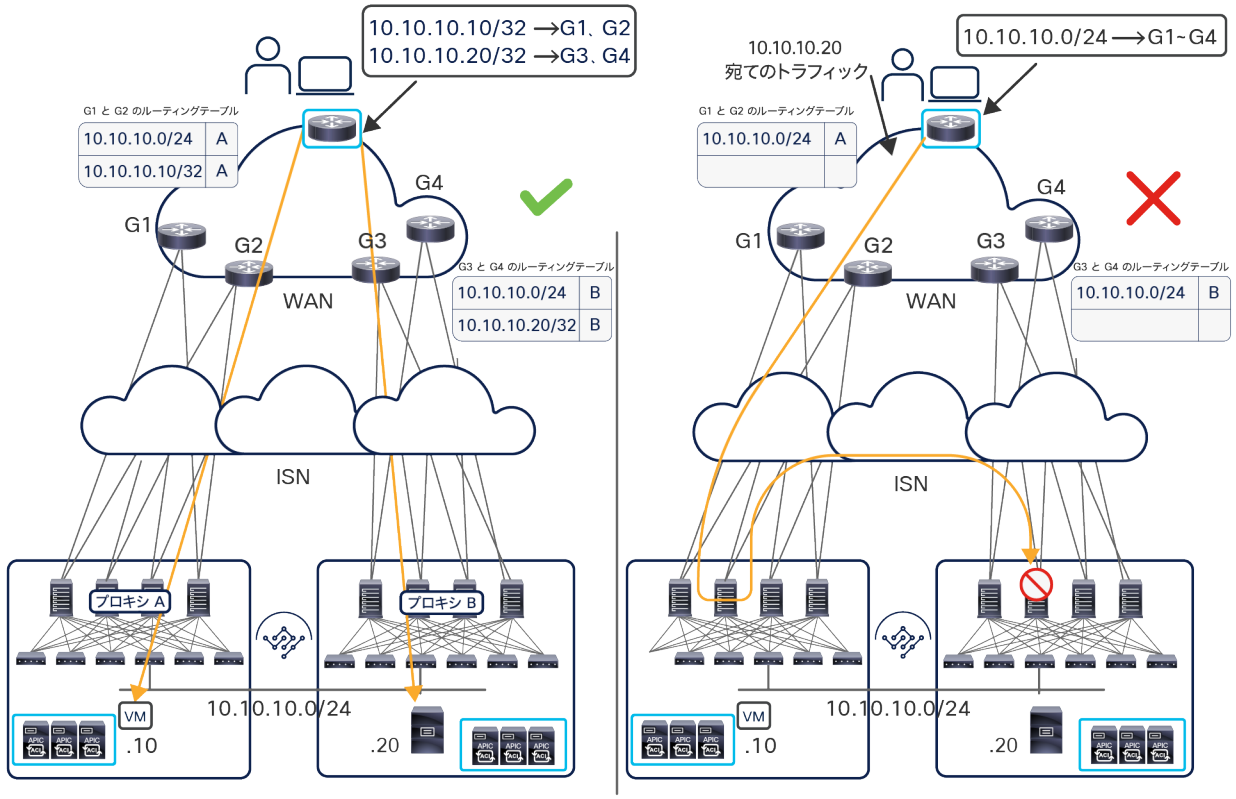


図 138.
GOLF L3Out でホストルートをアドバタイズしない場合に発生する入力トラフィックのドロップ

図 138 の右側に示すように、ホストルートをアドバタイズしない場合、リモート WAN ルータがストレッチブリッジドメインのルーティング情報をすべての GOLF ルータから受信する可能性があります。その結果、宛先エンドポイントがサイト 2 にあるにもかかわらず、入力トラフィックがサイト 1 にステアリングされる可能性があります。ボーダーリーフ L3Out を展開している場合は、この最適ではない動作がサポートされます。ただし、GOLF L3Out を使用する Cisco ACI マルチサイトの現在の実装では、変換テーブルに適切なエントリがないため、最適ではない入力フローが宛先サイトのスパインによってドロップされます。すなわち、GOLF L3Out を展開するときに機能する唯一のオプションは、ホストルートアドバタイズを活用して、（図 138 の左側のシナリオに示すように）入力トラフィックが常に最適に配信されるようにすることです。

注： BD が拡張されていない場合は、このような考慮は必要ありません。拡張されていない各 BD の IP サブネットは、常に、BD が展開されているファブリックのスパインノードによってのみアドバタイズされることが前提になっているためです。

マニュアルの変更履歴

新規トピックまたは改訂されたトピック	説明	日付
ドキュメント全体のコンテンツを更新して、MSO を NDO に修正		2022 年 11 月 9 日
Nexus Dashboard の導入に関する考慮事項を説明する新しいセクションを追加	「Cisco Nexus Dashboard の導入に関する考慮事項」セクション	2022 年 11 月 9 日
NDO リリース 4.0(1) で導入された新しいテンプレートタイプの説明を追加	「NDO リリース 4.0(1) で導入された新しいテンプレートタイプ」セクション	2022 年 11 月 9 日
NDO の運用面での機能強化（テンプレートレベルの機能）に関するセクションを追加	「NDO の運用面での機能強化」セクション	2022 年 11 月 9 日
ブラウンフィールド統合シナリオに関連するコンテンツを更新	「ブラウンフィールド統合シナリオ」セクション	2022 年 11 月 9 日
古い MSO クラスタの展開に関する考慮事項を付録 B に移動	付録 B	2022 年 11 月 9 日
GOLF L3Out 接続に関連するコンテンツを付録 C に移動	付録 C	2022 年 11 月 9 日

シスコシステムズ合同会社
〒107-6227 東京都港区赤坂9-7-1 ミッドタウン・タワー
cisco.com/jp

0713-2023-000-X