



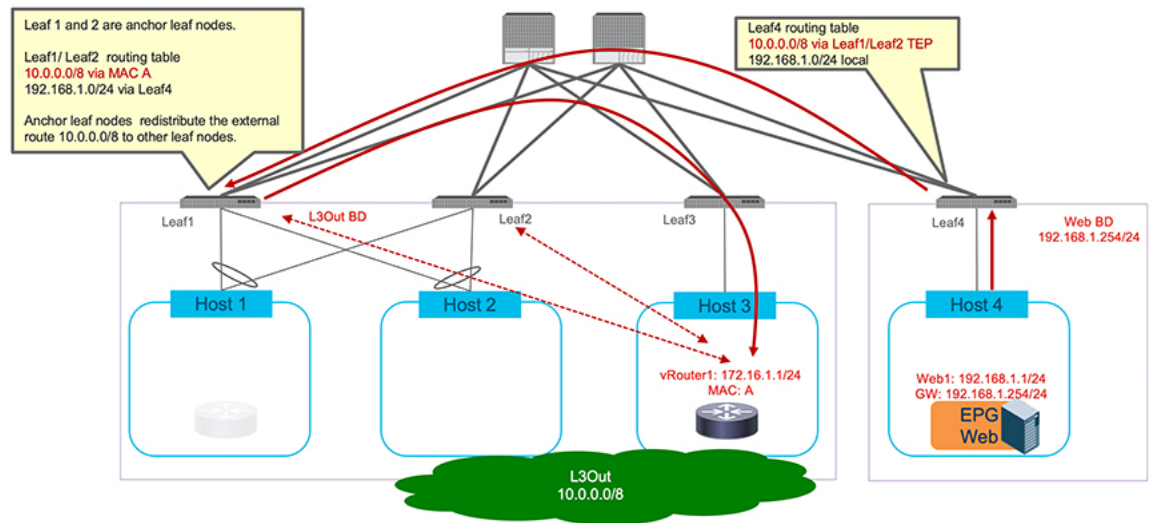
Cisco ACI 内部エンドポイントからフローティング L3Out への最適ではないトラフィックの回避

- [Cisco ACI 内部エンドポイントからフローティング L3Out への最適ではないトラフィックの回避 \(1 ページ\)](#)
- [マルチプロトコル再帰ルート解決のサポート \(5 ページ\)](#)
- [外部ルート アドバタイズメント専用デバイスを使用した代替導入モデル \(9 ページ\)](#)

Cisco ACI 内部エンドポイントからフローティング L3Out への最適ではないトラフィックの回避

Cisco Application Centric Infrastructure (ACI) リリース 5.0(1) 以前は、外部ルータが非アンカーリーフ ノードの下に接続されている場合でも、外部宛先に送信された Cisco ACI 内部エンドポイントからのトラフィックはアンカーリーフ ノードに移動し、非アンカーリーフ ノードを経由して外部ルータにリダイレクトされるため、トラフィック パスが最適化されていませんでした。

図 1: 内部 EPG と外部ネットワーク間の最適ではないトラフィック フローの例



Cisco ACI リリース5.0(1)以降では、次の2つの機能を構成することで、このアウトバウンドの最適ではないトラフィック パスの動作を回避できます。

- **ネクストホップ伝達**：この構成は、アンカー リーフ ノードにのみ適用され、これらの外部プレフィックスをアナウンスする外部ルータのネクストホップ IP アドレスを使用して、Cisco ACI ファブリック内の外部プレフィックスを再配布できるようにします。このようにして、コンピューティングリーフ ノード（図 2: [ネクストホップ伝播が有効な外部プレフィックスのアドバタイズメント](#)（3 ページ）の例の Leaf4）は、外部ルータの IP アドレスをネクストホップとして外部プレフィックス（下の例では 172.16.1.1 を使用して到達可能な 10.0.0.0/8）を受信し、転送テーブルにインストールします。
- **直接ホスト アドバタイズメントルート制御プロファイル**：この構成は、外部ルータが接続されているすべてのアンカー リーフ ノードと非アンカー リーフ ノードに適用されます。これにより、これらのリーフ ノードは、Cisco ACI ファブリック（次の例では Leaf3 TEP を使用する 172.16.1.1）内に直接接続されたホストルート（外部ルータの IP を表す）を再配布できます。これは、コンピューティングノードが再帰ルックアップを実行し、アンカーまたは非アンカー リーフ ノードに関係なく、外部ルータが接続されているリーフ ノードにアウトバウンドフローを直接送信できるようにするために重要です。



(注) 上記の機能は、VMM ドメインではなく、物理ドメインを持つフローティング L3Outs でのみサポートされます。

図 2: ネクストホップ伝播が有効な外部プレフィックスのアドバタイズメント

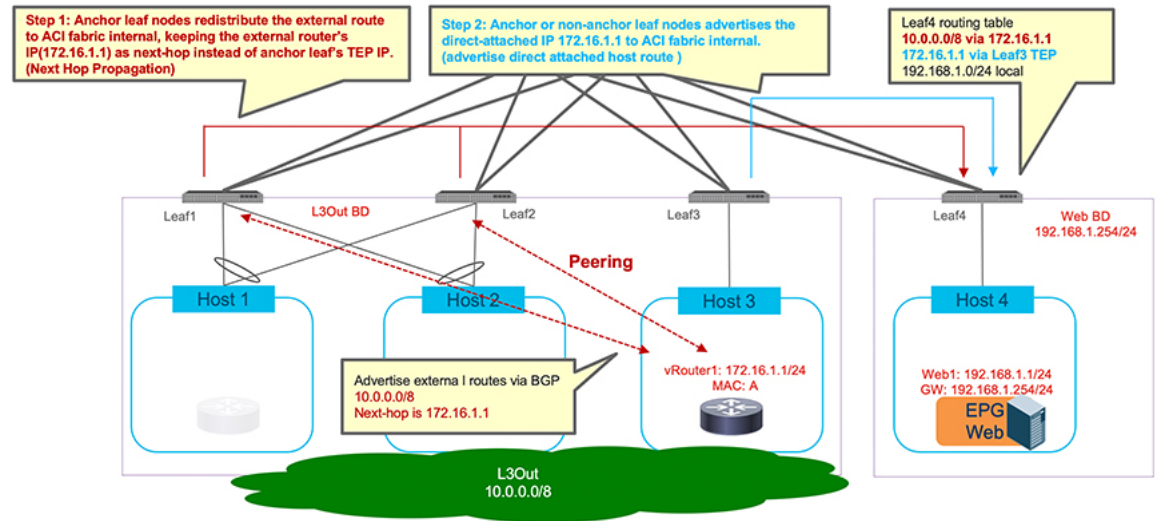
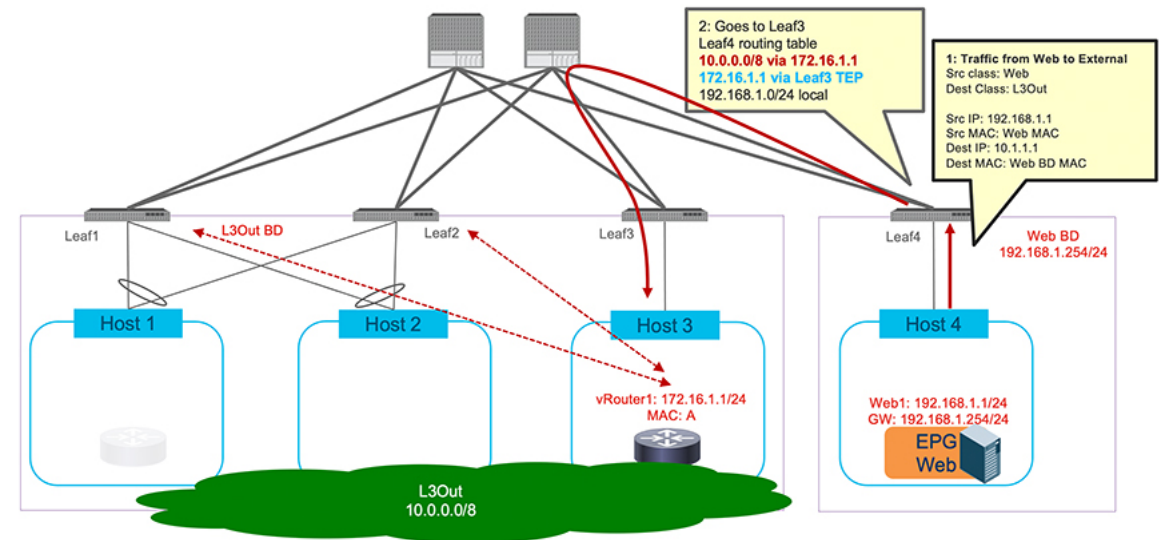


図 3: 内部 EPG と外部ネットワーク間の最適化されたトラフィック フロー



Cisco ACI リリース 5.0(1) から 5.2(1)の間では、上記の機能を使用して、外部デバイスとアンカーノード間のピアリングに eBGP を使用する場合に、この準最適パスを回避できます。Cisco ACI リリース 5.2(1)以降、この準最適パスの回避は、ピアリングに OSPF を使用する場合、または静的ルーティングを使用する場合でもサポートされます。

Cisco ACI リリース 6.0(2)以降、上記のアウトバウンドトラフィック最適化機能は、VRF 内トラフィックに対してのみサポートされます。コンシューマとプロバイダが異なる VRF にある VRF 間トラフィックはサポートされません。この考慮事項は、EPG から外部 EPG および外部 EPG から外部 EPG の両方のコントラクトに適用されます。

上記の例では、アンカーリーフノードとのピアリングと外部ネットワークドメインとの間でトラフィックを転送するために単一の外部デバイスを使用していますが、異なる外部デバイス

を使用することもできます。図 4: OSPF または静的ルートを使用して複数の外部ルータを展開する場合の外部プレフィックスの ECMP (4 ページ) は OSPF または静的ルートを使用した例を示し、図 5: BGP を使用して複数の外部ルータを展開する場合の外部プレフィックスの ECMP の欠如 (4 ページ) は BGP を使用した例を示しています。各外部デバイスは、アンカーリーフノードとのルーティングピアリングを確立して、外部プレフィックス情報をファブリックに伝播できます。

図 4: OSPF または静的ルートを使用して複数の外部ルータを展開する場合の外部プレフィックスの ECMP

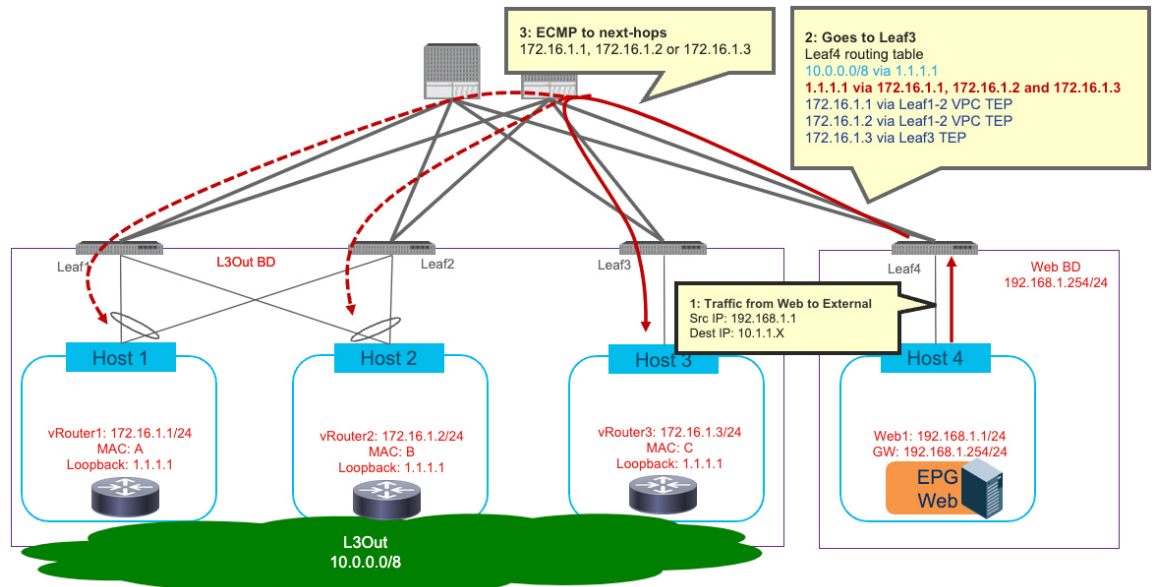
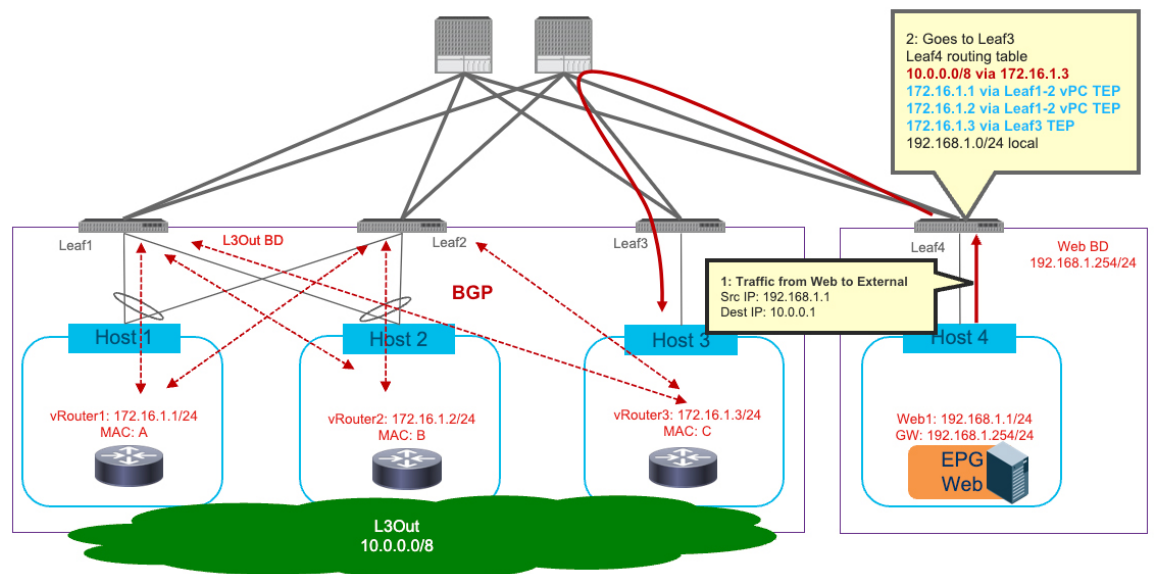


図 5: BGP を使用して複数の外部ルータを展開する場合の外部プレフィックスの ECMP の欠如



BGPを使用して同じ外部プレフィックスを学習する場合、各外部ルータが同じ外部プレフィックス（この例では 10.0.0.0/8）をアドバタイズしても、それらを受信するコンピューティングリーフノードはプレフィックスごとに1つのネクストホップのみをインストールします。つまり、同じ外部プレフィックスに到達するためにECMPを利用することはできません。この制限は、Cisco ACI リリース6.0(1) および Cisco ACI フローティング L3Out 展開の特定のケースでは、BGP を使用して学習されたプレフィックスに到達するための外部ネクストホップとして Cisco ACI リーフ ノードに 1つの IPアドレスのみをインストールできるためです。この考慮事項は、IPv4 と IPv6 の両方に適用されます。



- (注) 上の図に示すように、外部プレフィックスへの単一の経路がコンピューティングリーフノードにインストールされます。これは、外部ルータが受信した外部プレフィックスへのすべてのパスを正常にインストールできるアンカーリーフノードには当てはまりません。

外部プレフィックスに到達するためにECMPを活用するためのソリューションは、次のことが発生するように、すべてのフォワーダノードに同じループバック IP アドレスを展開することです。

- すべての外部ルータは、プレフィックスのネクストホップとして同じアドレス（ループバックインターフェイスの IP アドレス）を使用して、BGP を介してアンカーリーフノードに特定の外部プレフィックスをアドバタイズすることができます。
- アンカーリーフノードは外部プレフィックスを受信し、次の 2 つの機能を実行します。
 1. 外部プレフィックス情報を ACI ファブリックに再配布します。各外部ルータで構成された共通のループバック IP アドレスで表される単一のネクストホップを使用します。
 2. 直接接続された L3Out SVI サブネットの外部ルータ部分の IP アドレスをネクストホップとして、共通ループバック IP アドレスを ACI ファブリックに再配布します。この情報は、外部ルータで確立された OSPF 隣接関係を介してアンカーノードで学習するか、静的ルートを構成する必要があります。

アンカーノードから上記のコントロールプレーン情報を受信するコンピューティングリーフノードは、Cisco ACI リリース5.2(1) で導入された「再帰ルート解決」機能を活用して、外部プレフィックスへのアウトバウンドフローの ECMP の利点を活用できます。この動作については、次のセクション [マルチプロトコル再帰ルート解決のサポート（5 ページ）](#) で詳しく説明します。

マルチプロトコル再帰ルート解決のサポート

Cisco Application Centric Infrastructure (ACI) リリース5.2(1) 以降では、BGP を使用して受信され、ファブリックに再配布された外部プレフィックスの再帰ルート解決度を有効にするためのサポートを使用できます。これを実現するには、BGP を介してアンカーリーフノードで学習された外部プレフィックスのネクストホップとして使用されるループバック IP アドレスを外部ルータに定義する必要があります。

図 6: マルチパスが有効になっているネクストホップ伝達による外部プレフィックスのアドバタイズメント (6 ページ) では、このような構成の例を示します。外部ルート 10.0.0.0/8 は、外部ルータから BGP を使用してアドバタイズされます。外部プレフィックスのネクストホップは、L3Out SVI サブネット (172.16.1.1) に接続されている外部ルータの IP アドレスではなく、外部ルータで定義され、OSPF を使用して (または静的ルーティングを介して) アドバタイズされるループバックアドレスです。これにより、マルチレベルの再帰が発生し、BGP ルートのネクストホップが OSPF ルートを介して解決され、OSPF ルートが最終的に Cisco ACI ファブリック内で形成された直接的な隣接を介して解決されます。

このような場合、前のセクションで説明したネクストホップ伝達およびダイレクトホストアドバタイズメントルート制御プロファイルに加えて、追加の機能を有効にする必要があります。

- マルチパスを使用したネクストホップ伝達：これは、アンカーリーフノードが追加の再帰ルックアップに使用される、OSPF (または静的ルーティング) で学習した外部ルータのループバックアドレスをネクストホップ IP アドレスとして、外部プレフィックスをファブリックに再分配することを可能にします。コンピューティングリーフノード (次の例では Leaf4) は、外部ルータの直接接続された IP アドレスをネクストホップとして使用する外部ルータのループバック IP アドレス (次の例では 172.16.1.1 を介して 1.1.1.1/32) を受信します。これは、外部デバイスの背後にある外部ルート (次の例では 1.1.1.1 を介して 10.0.0.0/8) に到達します。

図 6: マルチパスが有効になっているネクストホップ伝達による外部プレフィックスのアドバタイズメント

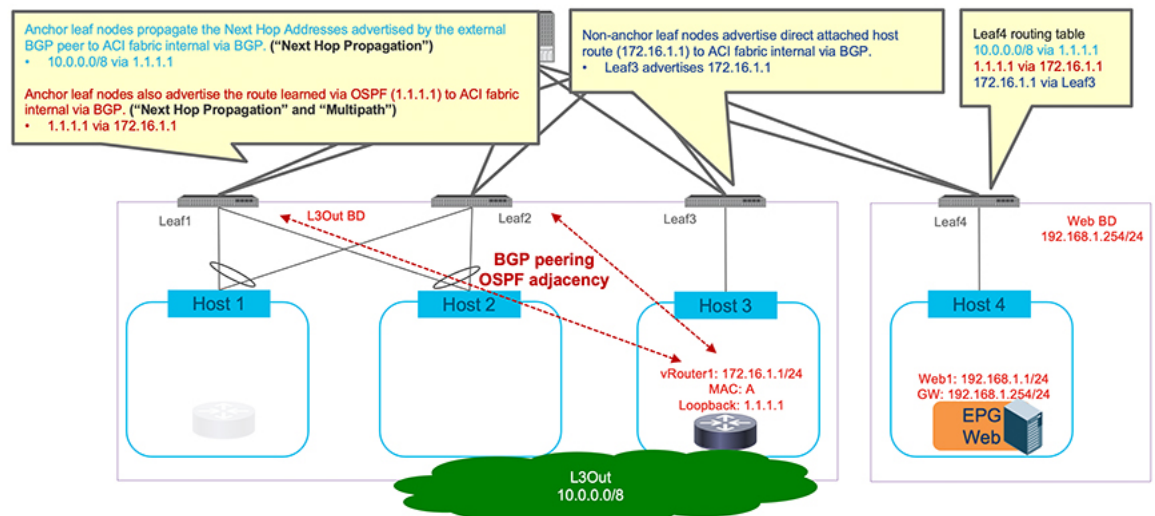
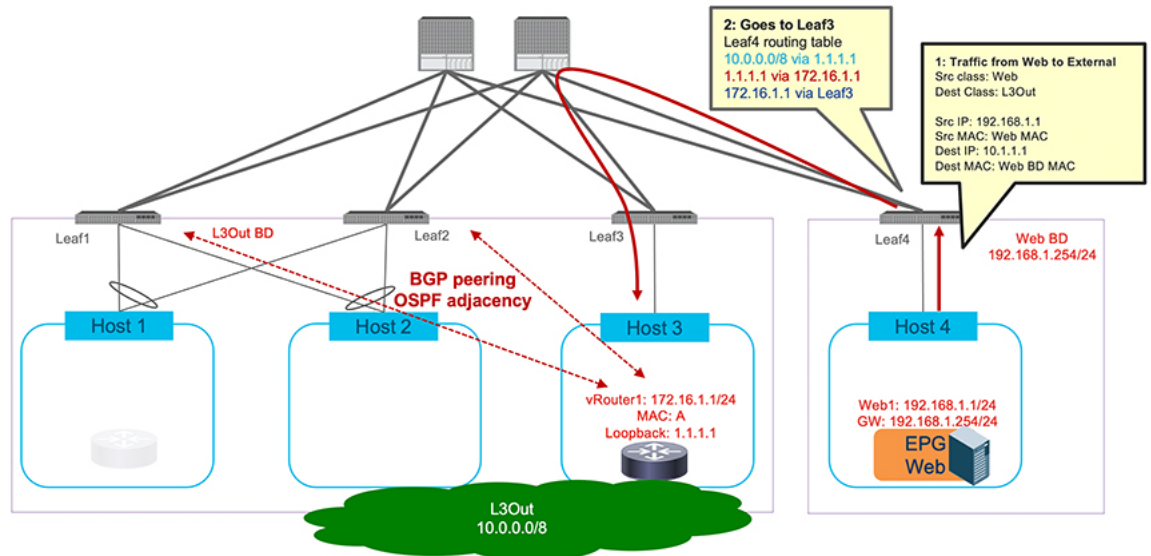


図 7: 内部 EPG と外部ネットワーク間のトラフィック フローの最適化



(注) 同じ外部デバイスを使用して、BGP を使用して外部プレフィックスをアドバタイズし、OSPF を使用してネクストホップ IP アドレスをアドバタイズできますが、異なるルーティングプロトコル (OSPF と BGP) が使用されるため、上記の設計には 2 つの L3Out が必要です。たとえば、L3Out-OSPF と L3Out-BGP は、「VRF」カプセル化範囲を使用して、同じ VLAN カプセル化を持つ同じフローティング SVI を持つことができます。適切な分類サブネット (たとえば、この特定の例では 10.0.0.0/8) を持つ外部 EPG は、2 つの L3Out のいずれかに対してのみ構成する必要があります。「VRF」カプセル化範囲の詳細については、「[ACI ファブリック L3Out ガイド](#)」を参照してください。

この例では OSPF を使用していますが、OSPF の代わりに静的ルートを使用することもできます。外部ルートの BGP とネクストホップループバック IP アドレスの静的ルートの組み合わせである場合は、同じ L3Out で構成できます。

外部プレフィックスに到達するためのネクストホップとしてループバック IP アドレスを使用することは、同じ外部ネットワークに接続する外部デバイスをさらに追加する可能性がある場合に、同じ外部プレフィックスに到達するために ECMP 機能を活用して転送規模を拡大できるため、有用となります。

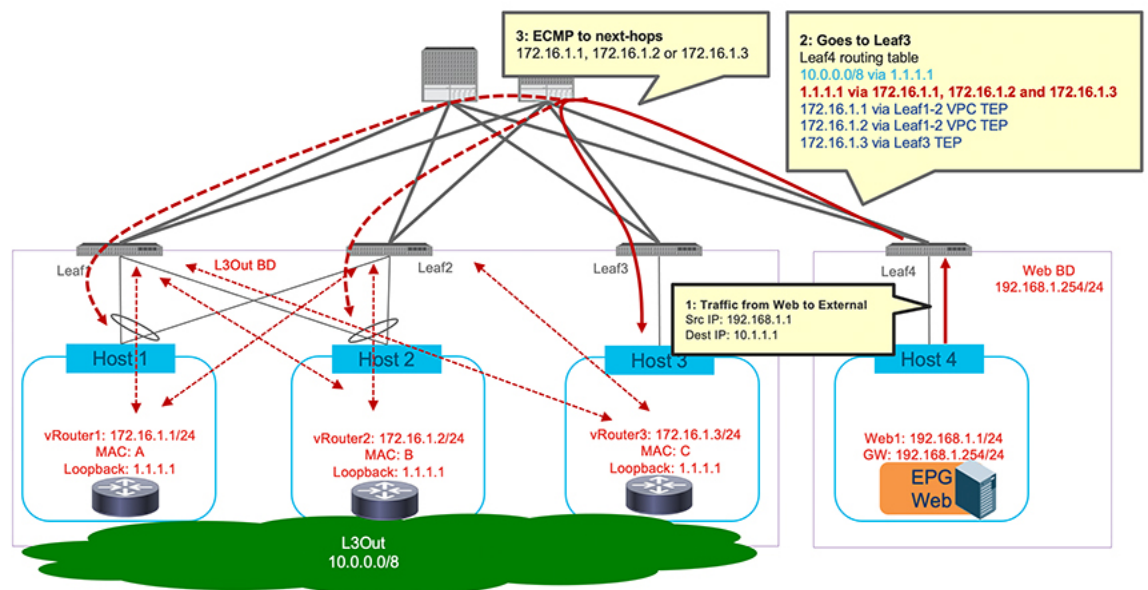
前述のように、この記事の執筆時点で利用可能な最新の Cisco ACI リリース 6.0(1) では、Cisco ACI コンピューティング リーフ ノードにインストールできるのは、BGP を介してフローティング L3Out で受信した外部プレフィックスのネクストホップ 1 つだけです。外部プレフィックスに到達するために ECMP が必要な場合は、次の操作を実行できます。

- BGP は、すべての外部ルータで構成されているのと同じループバック IP アドレスで表される、プレフィックスのプライマリ ネクストホップを 1 つ送信します。[図 8: 外部プレフィックスに到達するためのコンピューティングリーフノードでの複数の再帰ルックアップ](#)

プ (8 ページ) の例では、外部プレフィックス 10.0.0.0/8 は、ネクストホップ 1.1.1.1 を使用して BGP ビアで学習されます。

- 外部プレフィックスの 1.1.1.1 ネクストホップは、アンカーリーフ ノードが MP-BGP VPNv4 ファブリック コントロール プレインでこの情報を再配布するときに保持されます。
- 同時に、アンカーリーフ ノードは、外部ルータと確立された OSPF 隣接関係を使用して 1.1.1.1 ループバックアドレスを学習します。各ルータは、このようなプレフィックスのネクストホップとして、L3Out の SVI サブネットに接続されているローカルインターフェイスの IP アドレス (172.16.1.1、172.16.1.2、および 172.16.1.3) をアドバタイズします。
- 一般的なコンピューティングリーフノード (次の図の例の Leaf4) のルーティングテーブルを見ると、最終的に複数の再帰ルーティングルックアップが使用されます。
 - 外部プレフィックス (10.0.0.0/8) に到達するために、単一のループバック ネクストホップアドレス (1.1.1.1) がインストールされています。
 - BGP の代わりに OSPF または静的ルートがループバック ルートに使用されている場合、複数のパス (異なる外部ルータを識別する 172.16.1.1、172.16.1.2、および 172.16.1.3 アドレスを使用) を使用して、ループバック ネクストホップアドレスに到達できます。
 - 異なるリーフ ノードの VTEP アドレスは、L3Out の SVI サブネットに直接接続されている外部ルータの IP アドレスに到達するために使用されます。

図 8: 外部プレフィックスに到達するためのコンピューティングリーフノードでの複数の再帰ルックアップ



外部ルートアドバタイズメント専用デバイスを使用した代替導入モデル

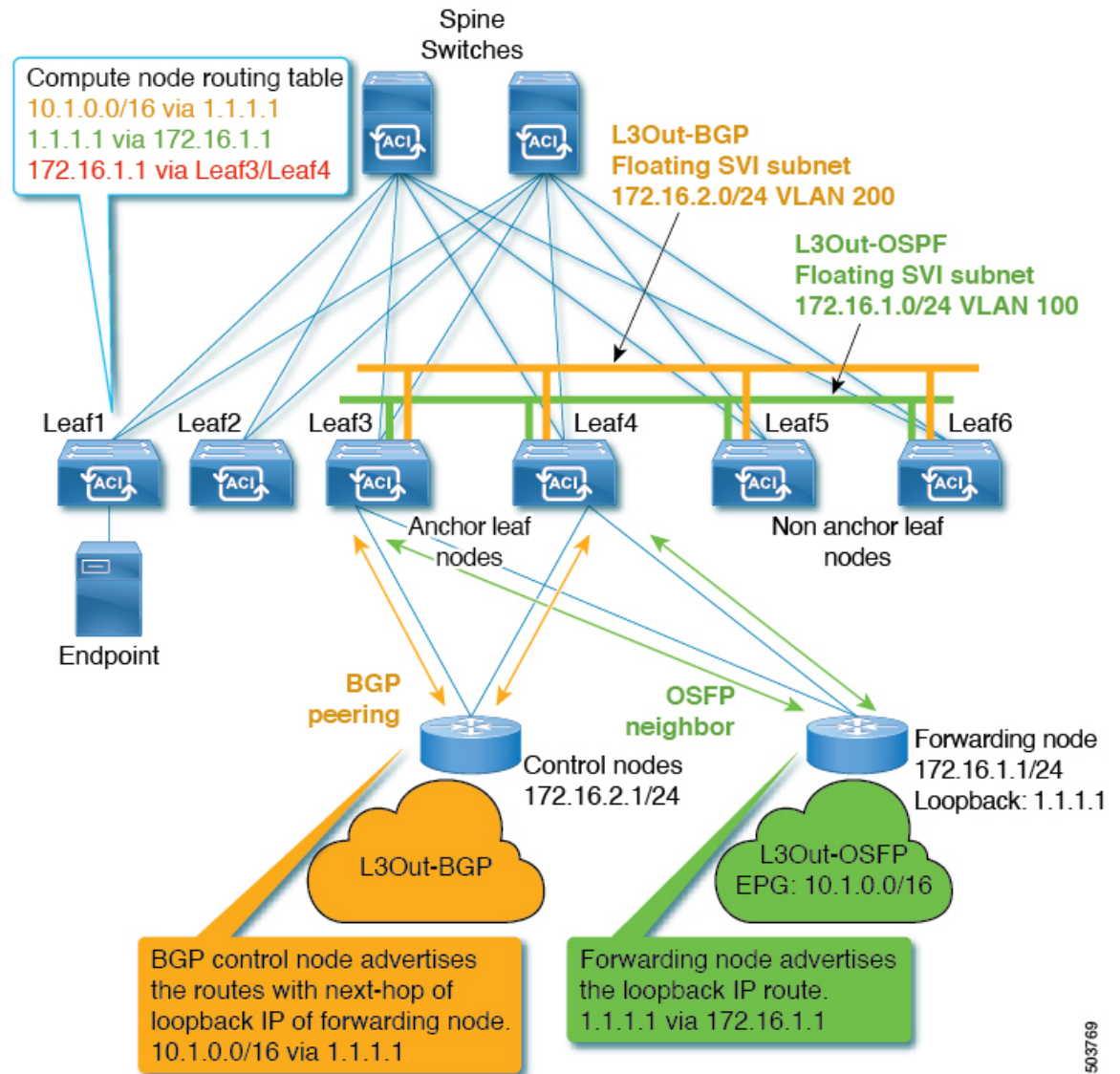
前の [図 8: 外部プレフィックスに到達するためのコンピューティングリーフノードでの複数の再帰ルックアップ \(8 ページ\)](#) に示したトポロジでは、外部デバイスを使用して、アンカーリーフノードとのコントロールプレーンの隣接関係を確立し、Cisco Application Centric Infrastructure (ACI) ファブリックと外部ネットワークの間でトラフィックを転送します。代替の展開モデルとして、Cisco ACI ファブリックに外部プレフィックスをアドバタイズするために専用の制御ノードデバイスを展開し (BGP ルーティングプロトコルを使用して)、実際のデータパス転送には個別の転送ノードを使用できます。この設計では、展開されている転送ノードの数に関係なく、外部プレフィックスの追加のコントロールプレーンピアリング (この例では BGP) を確立することなく、ノードの数を増やして転送キャパシティをスケールアップできます。

[図 9: ループバックアドレスを使用して準最適トラフィックを回避するための個別の外部デバイスの例 \(第 1 段階\) \(10 ページ\)](#) および [図 10: ループバックアドレスを使用して準最適トラフィックを回避するための個別の外部デバイスの例 \(第 2 段階\) \(11 ページ\)](#) は、OSPF を使用してアドバタイズされるループバックネクストホップアドレスの例を示しています。制御ノードは BGP ルートアドバタイズメント用で、転送ノードは OSPF ルートアドバタイズメント用であり、実際のデータパス転送に使用されます。

[図 9: ループバックアドレスを使用して準最適トラフィックを回避するための個別の外部デバイスの例 \(第 1 段階\) \(10 ページ\)](#) では、構成の準備段階が示されています。その場合、次のようになります。

- Leaf3 と Leaf4 は、L3Out-BGP と L3Out-OSPF の両方のアンカーリーフノードです。
- Leaf5 と Leaf6 は、非アンカーリーフノードです。
- 4 つのリーフノード (Leaf3 ~ Leaf6) すべてにまたがるオレンジと緑の線は、L3Outs の SVI サブネットの到達可能性を示します (つまり、各 L3Out に関連付けられた外部ブリッジドメインを表します)。
- BGP および OSPF セッションは、外部ルータ (コントロールおよび転送ノード) とアンカーリーフノード (Leaf3 および Leaf4) の間にあります。

図 9: ループバック アドレスを使用して準最適トラフィックを回避するための個別の外部デバイスの例 (第 1 段階)

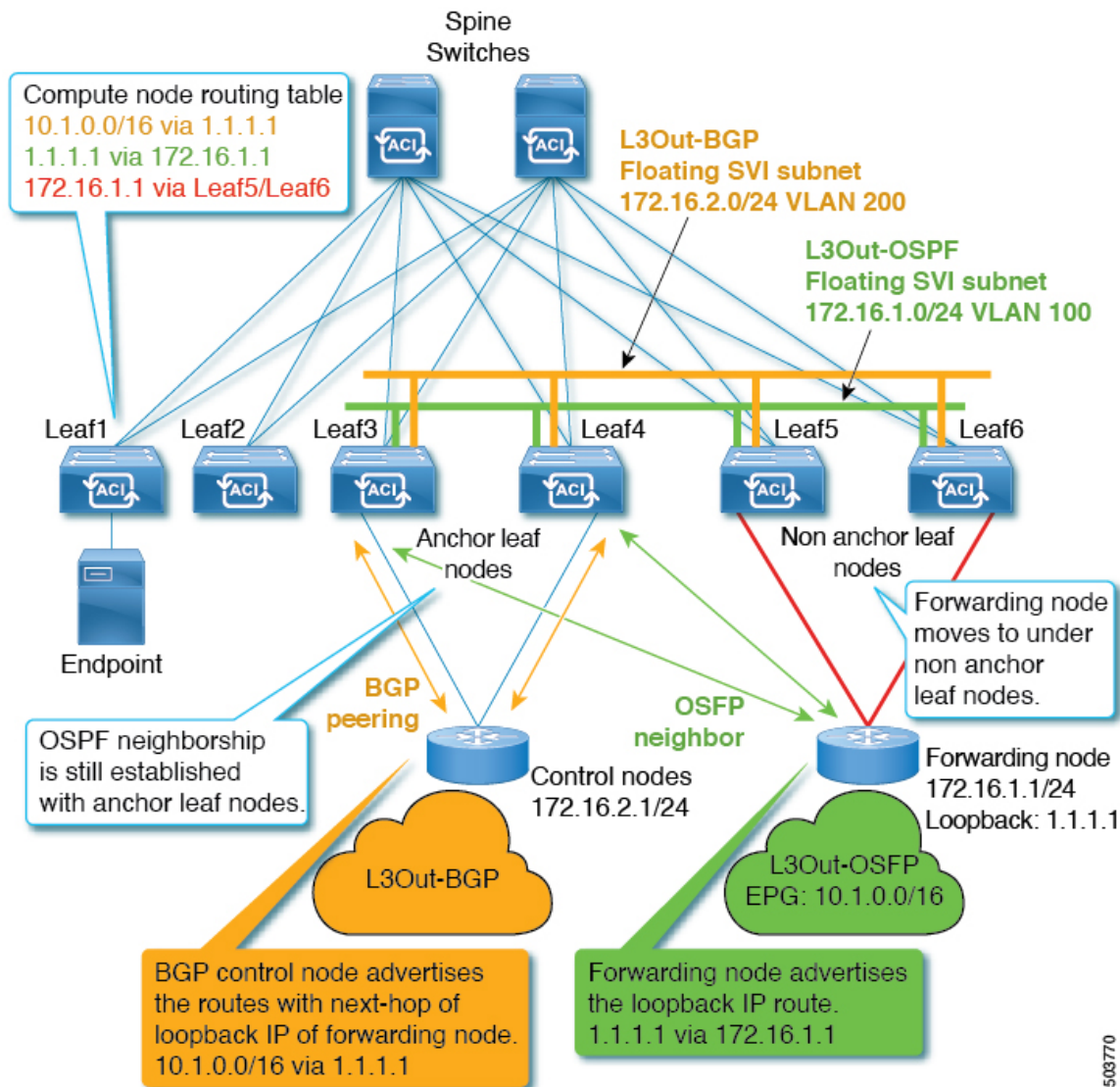


503769

次の 図 10: ループバック アドレスを使用して準最適トラフィックを回避するための個別の外部デバイスの例 (第 2 段階) (11 ページ) では、プロセスの次の段階が示されています。この段階のプロセス:

- 転送ノードは、フローティング SVI 動作を使用して、非アンカーリーフノードペア (Leaf5 と Leaf6) に移動しました。
- BGP および OSPF プロトコルセッションは、転送ノードの移動後も引き続き外部ルータとアンカーリーフノード (Leaf3 および Leaf4) の間で確立されます。
- コンピューティングリーフノードは、転送ノードが接続されている非アンカーリーフノードペア (Leaf5 と Leaf6) を使用して 172.16.1.1 アドレス (ループバックアドレス 1.1.1.1 に到達するネクストホップ) を指し示すようになりました。

図 10: ループバック アドレスを使用して準最適トラフィックを回避するための個別の外部デバイスの例 (第 2 段階)



503770

この設計には、次の考慮事項があります。

- 転送ノードは、転送ノード IP（つまり、L3Out の SVI サブネットに直接接続されている転送ノードの IP アドレス）をネクストホップとして、ループバック IP アドレスをアドバタイズします。上記の例では OSPF が使用されていますが、静的ルートを使用して、アンカーリーフノードにループバック IP アドレスのルートを追加できます。
- 制御ノードは、ネクストホップとしてループバック IP アドレスを使用して外部プレフィックスをアドバタイズする必要があります。通常、eBGP はこのために使用されるオプションです。
- 制御ノードは、フローティング L3Out または通常の L3Out に接続できます。これらの制御ノードがファブリック内を移動しない物理デバイスである場合は、フローティング L3Out の代わりに通常の L3Out を使用できます。

- BGP と OSPF を使用する場合は、同じフローティング SVI サブネットが BGP ピアリング と OSPF ネイバーに使用されている場合でも、2 つの異なる L3Out (1 つは BGP 用、もう 1 つは OSPF 用) を構成する必要があります。これは、フローティング L3Outs に固有のものではなく、L3Outs の一般的な考慮事項です。
- Cisco ACI リリース 6.0(1) の時点で、Cisco ACI フローティング L3Out の場合、BGP を使用して学習したプレフィックスに到達するための外部ネクストホップとして Cisco ACI リーフ ノードに 1 つの IP アドレスのみをインストールできます。この考慮事項は、IPv4 と IPv6 の両方に適用されます。すべての転送ノードに同じループバック IP アドレスを展開することで、外部プレフィックスに到達するために ECMP を活用できます。例：
 - 制御ノードで実行されている BGP は、外部プレフィックスの 1 つのプライマリ ネクストホップをアンカー リーフ ノードに送信します (たとえば、1.1.1.1 を介して 10.1.0.0/16)。
 - アンカー リーフ ノードと非アンカー リーフ ノード (Leaf1、Leaf2、および Leaf3 など) は、そのループバック ネクストホップ アドレスに到達するために使用できる直接ルートをファブリックに再配布します (たとえば、1.1.1.1 は 172.16.1.1, 172.16.1.2 および 172.16.1.3 を使用して到達可能です)。
- ネクストホップ伝達が有効になっている場合、同じ ECMP パスを持つ同じ外部プレフィックスをアドバタイジングする複数の制御ノードの展開は、CSCwd28918 の影響を受けます。詳細は [フローティング L3Out の概要の考察事項と制限事項](#) を参照してください。

翻訳について

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