



最適でないトラフィックと ECMP を回避したトポロジ例

- [最適でないトラフィックと ECMP を回避したトポロジ例 \(1 ページ\)](#)

最適でないトラフィックと ECMP を回避したトポロジ例

このセクションでは、最適ではないトラフィックと ECMP を回避するフローティング L3Out の構成要件の概要を示します。構成ステップと各構成オプションについては、「[Cisco APIC GUI を使用した 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の構成](#)」を参照してください。

このセクションでは、次の例について説明します。

- 外部プレフィックスは BGP をピア学習されます。BGP をピア学習されたルートのネクストホップは1つしか構成できないため、ループバックを使用した再帰ルックアップが必要です。例：
 - 10.0.0.0/8 は、外部ルータのループバック IP アドレス 1.1.1.1 であるネクストホップを使用して BGP をピア学習された外部プレフィックスです。このループバックアドレスは、OSPF をピア学習されるか、静的ルーティングを介して構成されます。
 - 172.16.1.1 および 172.16.1.3 経由で 01.1.1.1/32。OSPF または静的ルート経由で学習されます。
- 外部プレフィックスは、OSPF または静的ルートを介して学習されます。例：
 - 10.0.0.0/8 は、外部ルータの接続 IP アドレス 172.16.1.1 であるネクストホップを使用して、OSPF をピア学習された（または静的ルーティングをピアで構成された）外部プレフィックスです。172.16.1.2 および 172.16.1.3。
- 外部プレフィックスは、複数のネクストホップを使用して BGP を介して学習されます。このオプションには、Cisco Application Centric Infrastructure (ACI) リリース 6.0(2)以降が必要です。
 - BGP をピアで学習される 1.1.1.1、1.1.1.2、および 1.1.1.3 を介した 10.0.0.0/8

- 172.16.1.1 経由の 1.1.1.1、172.16.1.2 経由の 1.1.1.2、および 172.16.1.3 経由の 1.1.1.3。OSPF または静的ルートを介して学習されます。

例 1：外部プレフィックスが BGP をピア学習される

このオプションには、Cisco ACI リリース 5.2 以降が必要です。次の図に例を示します。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、172.16.1.2、および 172.16.1.3) 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。

必要な構成は次のとおりです。

- BGP アドレス ファミリ コンテキスト ポリシーの「Local Max ECMP」を増やす。
- BGP の L3Out で次のオプションを使用してルートマップを構成します。
 - 外部プレフィックスのネクストホップ伝播 (10.0.0.0/8)。
- OSPF の L3Out で次のオプションを使用してルートマップを構成します。
 - ループバック IP (1.1.1.1) のネクストホップ伝播とマルチパス。
 - 直接接続されたネクストホップ IP (172.16.1.1、172.16.1.2、および 172.16.1.3) の直接接続ホストをアドバタイズします

OSPF の代わりに静的ルートを使用する場合は、BGP の L3Out と OSPF の L3Out を 1 つの L3Out に結合できます。

図 1: 外部プレフィックスは **BGP** をピア学習されます。

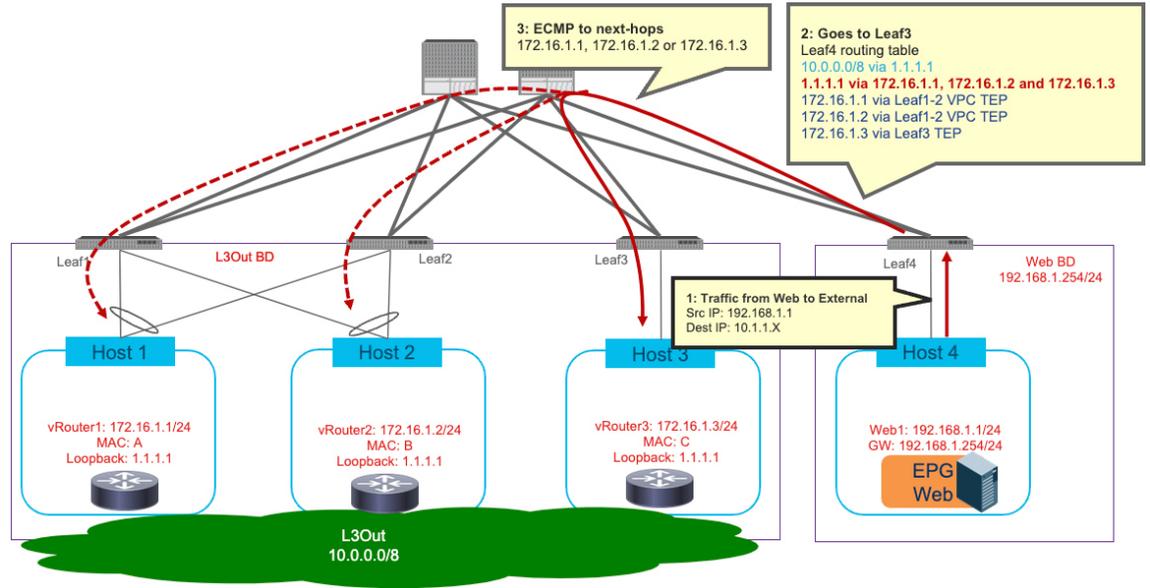
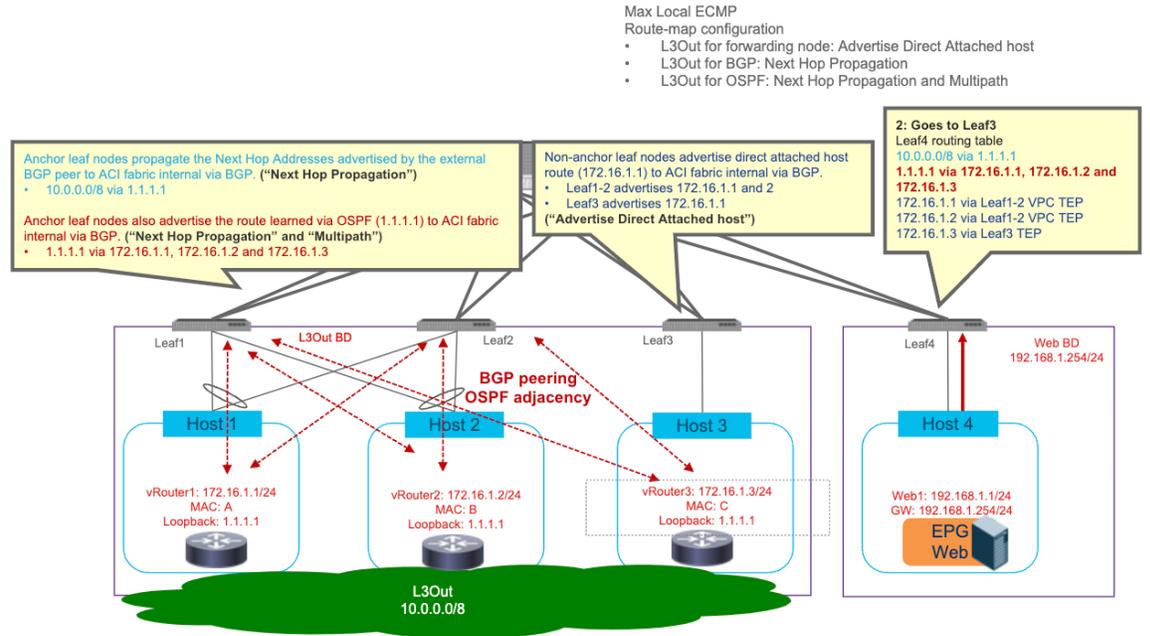


図 2: 外部プレフィックスは **BGP** をピア学習されます (ルートマップ構成)



例 2: 外部プレフィックスが **OSPF** または静的ルートを介して学習されます

このオプションには、Cisco ACI リリース 5.2 以降が必要です。次の図に例を示します。外部ルート 10.0.0.0/8 は、外部ルータから 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 サブネットに接続されている外部ルータの IP アドレスです（172.16.1.1、172.16.1.2、および 172.16.1.3）。

必要な構成は次のとおりです。

- BGP アドレス ファミリ コンテキスト ポリシーの「Local Max ECMP」を増やす。
- OSPF の L3Out で次のオプションを使用してルートマップを構成します。
 - 外部プレフィックス（10.0.0.0/8）のネクストホップ伝播とマルチパス。
 - 直接接続されたネクストホップ IP（172.16.1.1、172.16.1.2、および172.16.1.3）の直接接続ホストをアドバタイズします。

図 3: 外部プレフィックスは、**OSPF**または静的ルートをピア学習されます。

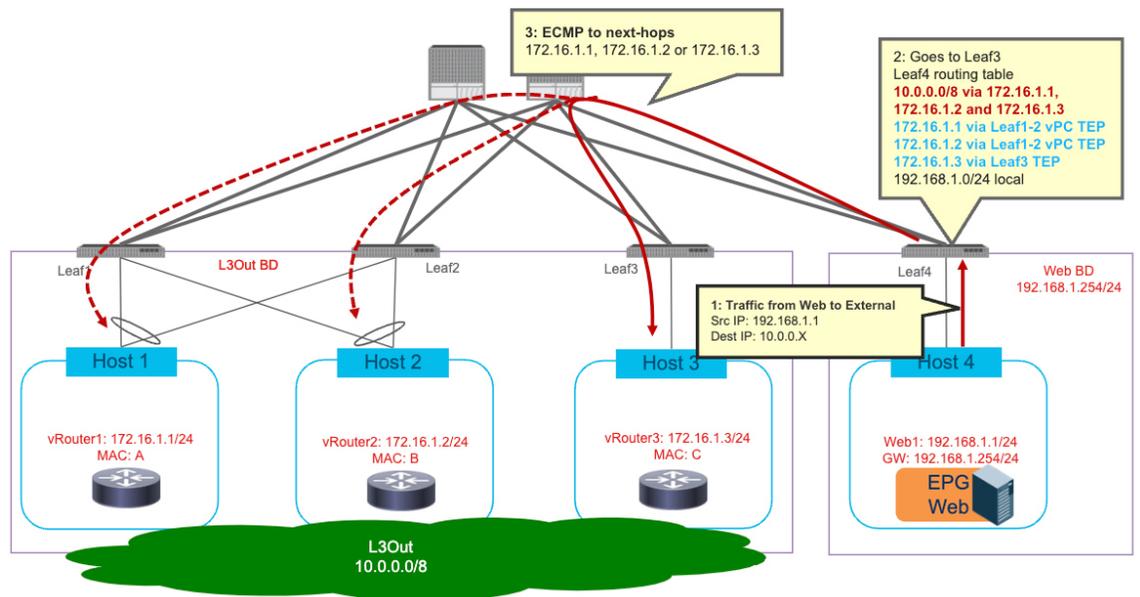
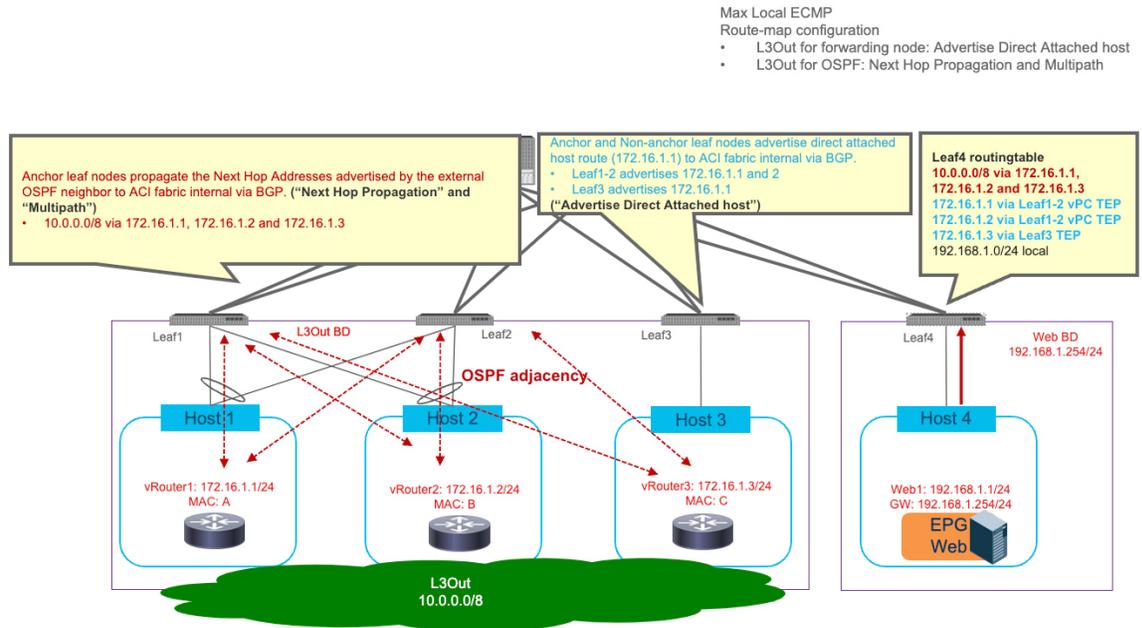


図 4:外部プレフィックスは、OSPFまたは静的ルート（ルートマップ構成）を介して学習されます。



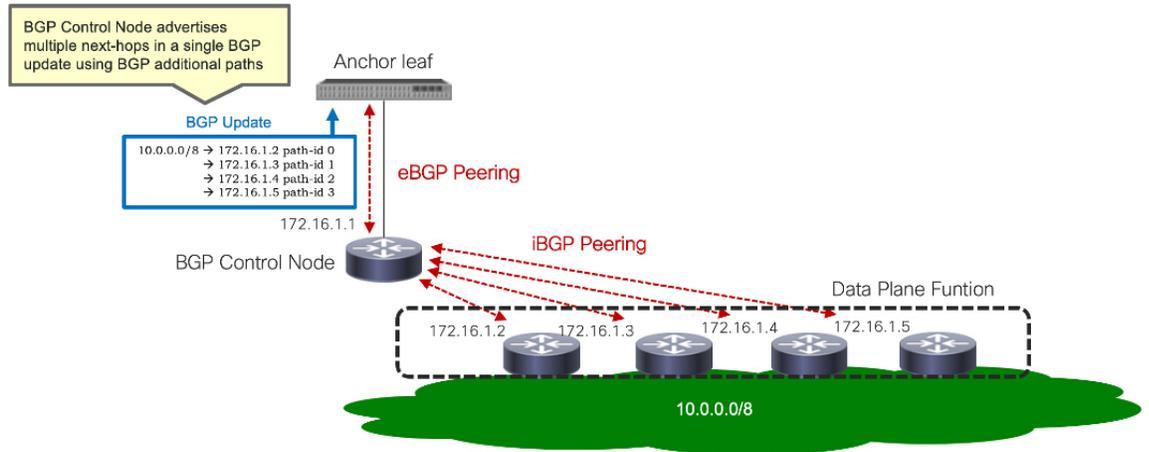
例 3 : BGP 追加パス機能と直接接続されたネクストホップを使用して BGP をピア学習された外部プレフィックス

このオプションには、Cisco ACI リリース 6.0(2) と、BGP 追加パス受信機能構成の使用が必要です。

BGP 追加パス機能は、BGP スピーカーが単一の BGP 更新で複数のパス（ネクストホップ）を送信できるようにする BGP 拡張機能です。この機能がない場合、同じピアからの異なるパスを持つプレフィックスのアドバタイズメントは、そのプレフィックスの以前のパスを置き換えます。BGP スピーカーは、追加パスの送信、受信、または送受信の両方を行うように構成できます。Cisco ACI 6.0(2) は受信機能のみをサポートします。

BGP 追加パス機能は、コントロールプレーンとデータプレーンの機能が異なるノードまたは仮想ルータで実行されるトポロジで役立ちます。制御ノードは、ネクストホップアドレスがデータプレーン機能を実行する別の仮想ルータであるプレフィックスをアドバタイズする BGP スピーカーです。BGP 追加パスを使用すると、BGP 制御ノードは、単一の BGP 更新で単一のプレフィックスの複数のネクストホップをアドバタイズできます。これには、制御ノードで BGP 追加パス送信機能が必要です。BGP 追加パスを使用すると、パスごとに個別の BGP セッションを必要とせず、複数の ECMP パスでデータプレーントラフィックのロードバランシングが可能になります。次の図に、BGP 追加パス機能を使用できるトポロジの例を示します。BGP 制御ノードは、追加パス機能を使用して、単一の BGP 更新で内部 BGP ピアから学習したルートアドバタイズするように構成されます。

図 5: BGP 追加パス機能を使用して複数のパスをアドバタイジングする BGP 制御ノード。



次の図は、この機能がフローティング L3Out で使用され、ネクストホップを伝播する例を示しています。外部ルート 10.0.0.0/8 は、それぞれが異なるネクストホップアドレスを持つ 3 つのパス ID を使用して制御ノードによっ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アドレスは、L3Out SVI サブネットに接続されている外部ルータの IP アドレスです (172.16.1.2、172.16.1.3、および 172.16.1.4)。プレフィックス 10.0.0.0/8 の ECMP パスは、3 つのネクストホップすべてを使用して、アンカーノード (Leaf1 および Leaf2) のルーティングテーブルにインストールされます。この ECMP パスは、BGP を介して非境界リーフスイッチにも伝播され、3 つの伝播されたネクストホップすべてを使用して ECMP パスがインストールされます。



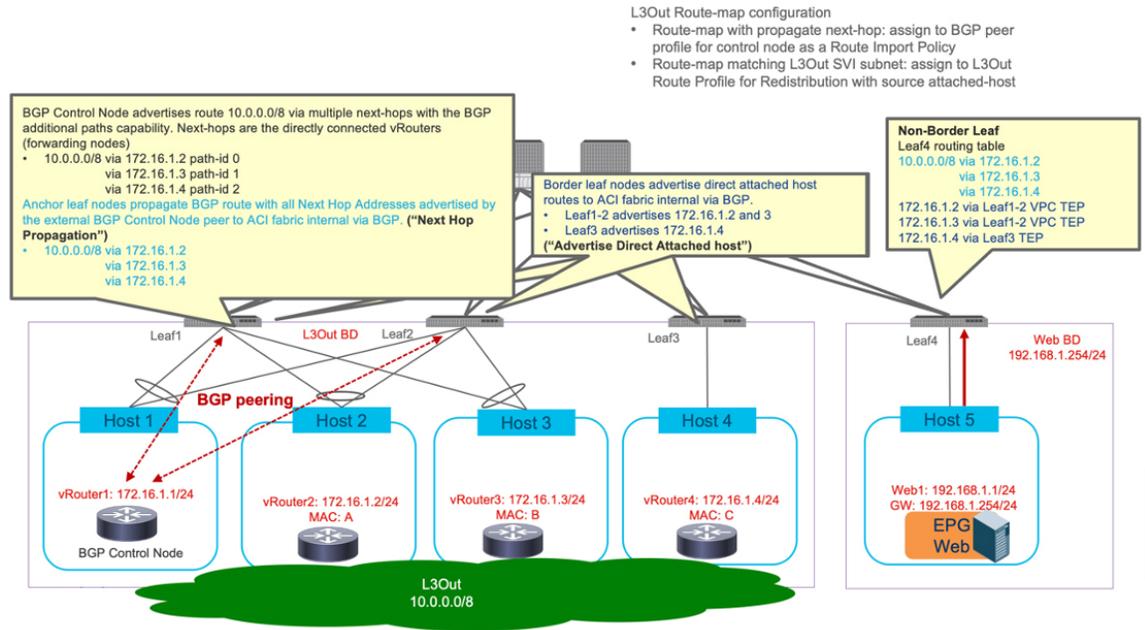
- (注) アンカーリーフから非境界リーフスイッチへの ECMP パスの伝達には、BGP 追加パス機能は必要ありません。Cisco ACI リリース 6.0.2 より前では、アンカーリーフに BGP をピア学習された ECMP パスがある場合、アンカーリーフは非境界リーフスイッチに単一のパスのみをアドバタイズします。この動作は、Cisco ACI リリース 6.0(2) で変更され、新しく導入された追加パス機能を最大限に活用できるようになりました。アンカーリーフは、BGP で学習されたルートの ECMP パスをアドバタイズします。アンカーリーフの ECMP パスは、BGP 追加パス機能を使用するか、直接接続された BGP ピアからの個々の更新から学習できます。

必要な構成は次のとおりです (この構成例では、1 つの L3Out のみが必要です)。

- 外部 BGP スピーカー (制御ノード) は、BGP 追加パス送信機能を使用して構成されており (これは Cisco ACI ファブリックでは構成されていません)、アンカーノード (Leaf1 および Leaf2) との BGP 隣接関係を作成しています。
- Cisco ACI で BGP 追加パス受信機能を有効にします。これは、BGP ピア プロファイル レベルまたは VRF レベルで有効にできます。
- 次のオプションを使用してルートマップを構成し、ルートインポートポリシーとして制御ノードの BGP ピアに適用します。

- 外部プレフィックスのネクスト ホップ伝播 (10.0.0.0/8)。
- 直接接続されたネクストホップの次のオプションを使用してルートマップを構成します。
 - 直接接続されたネクストホップ IP (172.16.1.2、172.16.1.3、および172.16.1.4) の直接接続ホストをアダバタイズします

図 6: 追加パス機能を備えた BGP 制御ノードをピア学習された外部プレフィックス ECMP パス

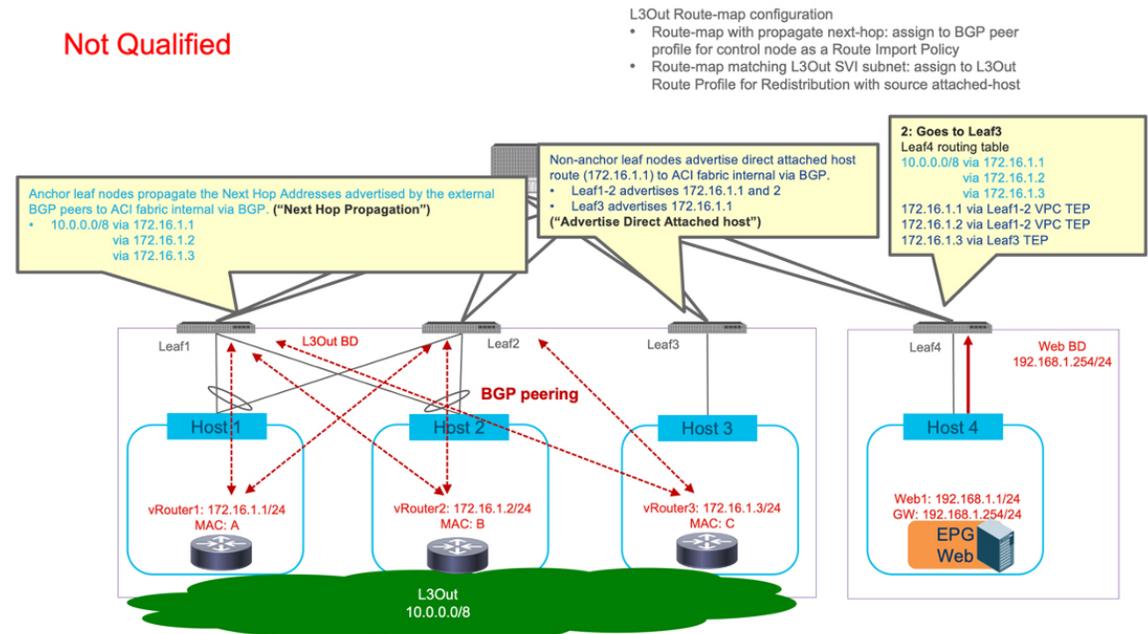


次の図は、制御ノードを使用せずに、直接接続された BGP ピアを ECMP パスが学習される例を示しています。このトポロジでは、BGP 追加パス機能は必要ありません。



(注) この構成は現時点では認定されていません。

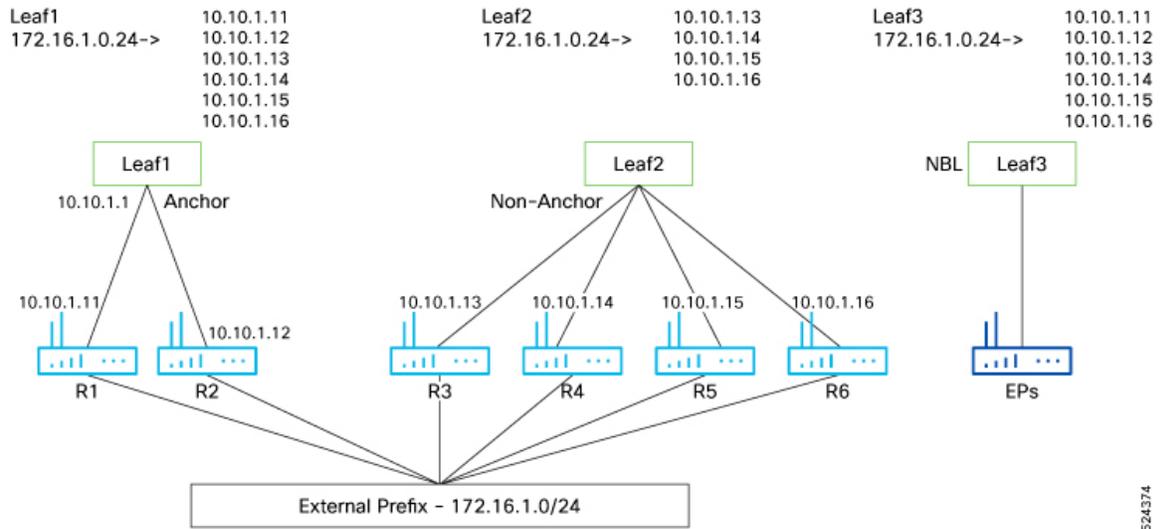
図 7: 直接接続された BGP スピーカーをピア学習された外部プレフィックス ECMP パス



ネクストホップ伝達を伴う等コストマルチパス（ECMP）を持つ外部プレフィックスでは、トラフィック流通が不均一になる可能性があります。これは、ECMP パスのピアデバイスがアンカーノードと非アンカーノードにまたがって接続されている場合に発生します。アンカーノードの背後にあるピアデバイスは、受信するトラフィックが少なくなる可能性があります。

たとえば、次のトポロジプレフィックス 172.16.1.0/24 には6個のネクストホップがあります。ネクストホップ伝達では、6個のパスすべてが Leaf3 で使用可能になります。Leaf3 の背後にある EP からプレフィックス 172.168.1.0 へのトラフィックは、6つのパスすべてで Leaf3 でロードバランシングされます。Leaf3 から Leaf2 に流れるパケットは、4つのローカルパスにハッシュされ、R3 から R6 に到達します。

Leaf3 から Leaf1 へのトラフィックは、Leaf1 で6つのパスに再ハッシュされます。これにより、Leaf1 から Leaf2 へのトラフィックのヘアピンが発生します。また、R1 と R2 は、R3 から R6 と比較して受信するトラフィックが少なくなります。



524374

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。