



ACI ファブリック内での転送

この章は、次の内容で構成されています。

- [ACI ファブリック内の転送について \(1 ページ\)](#)
- [ACI ファブリックは現代のデータセンタートラフィックフローを最適化する \(2 ページ\)](#)
- [ACI で VXLAN \(3 ページ\)](#)
- [サブネット間のテナントトラフィックの転送を促進するレイヤ 3 VNID \(5 ページ\)](#)
- [ポリシー ID と適用 \(7 ページ\)](#)
- [ACI ファブリック ネットワーク アクセス セキュリティ ポリシー モデル \(契約\) \(8 ページ\)](#)
- [マルチキャスト ツリー トポロジ \(14 ページ\)](#)
- [トラフィック ストーム制御について \(16 ページ\)](#)
- [ストーム制御の注意事項と制約事項 \(16 ページ\)](#)
- [ファブリック ロード バランシング \(19 ページ\)](#)
- [エンドポイントの保持 \(22 ページ\)](#)
- [IP エンドポイントの学習動作 \(24 ページ\)](#)
- [プロキシ ARP について \(25 ページ\)](#)
- [ループ検出 \(32 ページ\)](#)
- [不正なエンドポイントの検出 \(35 ページ\)](#)
- [TCP MSS調整について \(36 ページ\)](#)

ACI ファブリック内の転送について

ACI ファブリックは、64,000 以上の専用テナントネットワークをサポートしています。単一のファブリックは、100 万以上の IPv4/IPv6 エンドポイント、64,000 以上のテナント、および 200,000 以上の 10G ポートをサポートできます。ACI ファブリックにより、物理サービスと仮想サービス間を接続する追加のソフトウェアやハードウェアゲートウェイを必要とすることなくサービス（物理または仮想）がどこでも可能になり、Virtual Extensible Local Area Network (VXLAN) /VLAN/Network Virtualization using Generic Routing Encapsulation (NVGRE) のカプセル化が正規化されます。

ACI ファブリックは、基盤となる転送グラフからエンドポイント ID ポリシーおよび関連するポリシーを分離します。また、最適なレイヤ3およびレイヤ2フォワーディングを保証する分散レイヤ3ゲートウェイが提供されます。ファブリックは、一般的な場所の制約（あらゆる場所の IP アドレス）なしで標準のブリッジングおよびルーティングのセマンティックをサポートし、IP コントロールプレーンの Address Resolution Protocol (ARP) /Gratuitous Address Resolution Protocol (GARP) に関するフラッド要件を削除します。ファブリック内のすべてのトラフィックは、VXLAN 内にカプセル化されます。

ACI ファブリックは現代のデータセンタートラフィックフローを最適化する

Cisco ACI アーキテクチャは、従来のデータセンター設計から来る制限を解放して、最新のデータセンターで増大する East-West トラフィックの需要に対応します。

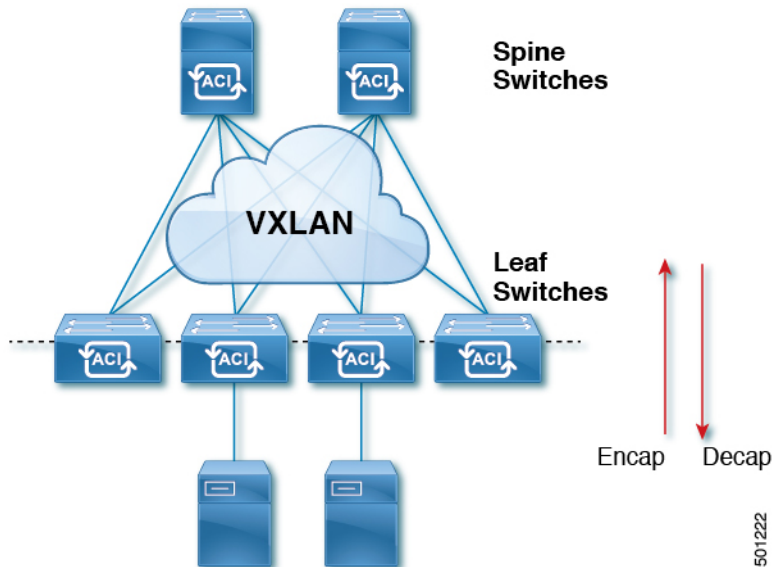
今日のアプリケーション設計は、データセンターのアクセスレイヤを通る、サーバ間の East-West トラフィックを増大させています。このシフトを促進しているアプリケーションには、Hadoop のようなビッグデータの分散処理の設計、VMware vMotion のようなライブの仮想マシンまたはワークロードの移行、サーバのクラスタリング、および多層アプリケーションなどが含まれます。

North-South トラフィックは、コア、集約、およびアクセスレイヤ、またはコラプストコアとアクセスレイヤが重要となる、従来型のデータセンター設計を推進します。クライアントデータはWAN またはインターネットで受信され、サーバの処理を受けた後、データセンターを出ます。このような方式のため、WAN またはインターネットの帯域幅の制限により、データセンターのハードウェアは過剰設備になりがちです。ただし、スパンニングツリープロトコルが、ループをブロックするために要求されます。これは、ブロックされたリンクにより利用可能な帯域幅を制限し、トラフィックが準最適なパスを通るように強制する可能性があります。

従来のデータセンター設計においては、IEEE 802.1Q VLAN がレイヤ2境界の論理セグメンテーションまたはブロードキャストドメインを提供します。ただし、ネットワークリンクのVLANの使用は効率的ではありません。データセンターネットワークでデバイスの配置要件は柔軟性に欠け、VLANの最大値である4094のVLANが制限となり得ます。IT部門とクラウドプロバイダが大規模なマルチテナントデータセンターを構築するようになるにつれ、VLANの制限は問題となりつつあります。

スパインリーフアーキテクチャは、これらの制限に対処します。ACI ファブリックは、外界からは、ブリッジングとルーティングが可能な単一のスイッチに見えます。レイヤ3のルーティングをアクセスレイヤに移動すると、最新のアプリケーションが必要としている、レイヤ2の到達可能性が制限されます。仮想マシンワークロードモビリティや一部のクラスタリングのソフトウェアのようなアプリケーションは、送信元と宛先のサーバ間がレイヤ2で隣接していることを必要とします。アクセスレイヤでルーティングを行えば、トランクダウンされた同じVLANの同じアクセススイッチに接続したサーバだけが、レイヤ2で隣接します。ACIでは、VXLANが、基盤となるレイヤ3ネットワークインフラストラクチャからレイヤ2のドメインを切り離すことにより、このジレンマを解決します。

図 1: ACI ファブリック



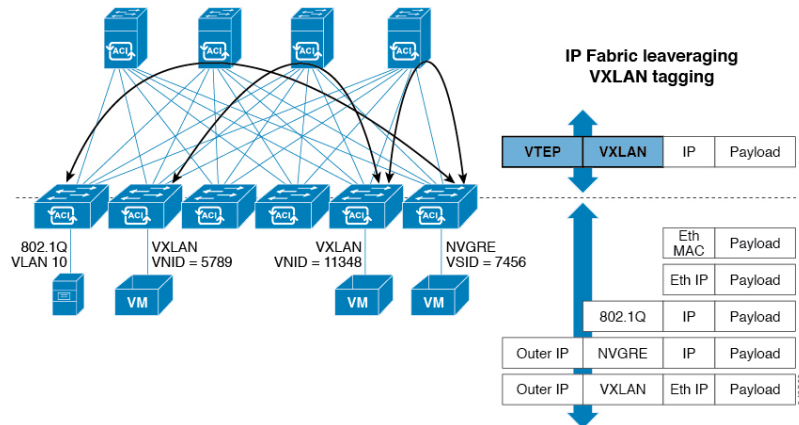
トラフィックがファブリックに入ると、ACIがカプセル化してポリシーを適用し、必要に応じてスパインスイッチ (最大 2 ホップ) によってファブリックを通過させ、ファブリックを出るときにカプセル化を解除します。ファブリック内では、ACIはエンドポイント間通信でのすべての転送について、Intermediate System-to-Intermediate System プロトコル (IS-IS) および Council of Oracle Protocol (COOP) を使用します。これにより、すべての ACI リンクがアクティブで、ファブリック内での等コストマルチパス (ECMP) 転送と高速再コンバージョンが可能になります。ファブリック内と、ファブリックの外部のルータ内でのソフトウェア定義ネットワーク間のルーティング情報を伝播するために、ACIはマルチプロトコル Border Gateway Protocol (MP-BGP) を使用します。

ACI で VXLAN

VXLAN は、レイヤ 2 オーバーレイの論理ネットワークを構築するレイヤ 3 のインフラストラクチャ上でレイヤ 2 のセグメントを拡張する業界標準プロトコルです。ACIインフラストラクチャレイヤ 2 ドメインが隔離ブロードキャストと障害ブリッジドメインをオーバーレイ内に存在します。このアプローチは大きすぎる、障害ドメインの作成のリスクなしで大きくなるデータセンターネットワークを使用できます。

すべてのトラフィック、ACIファブリックはVXLANパケットとして正規化されます。入力でACI VXLANパケットで外部VLAN、VXLAN、およびNVGREパケットをカプセル化します。次の図は、ACIカプセル化の正規化を示します。

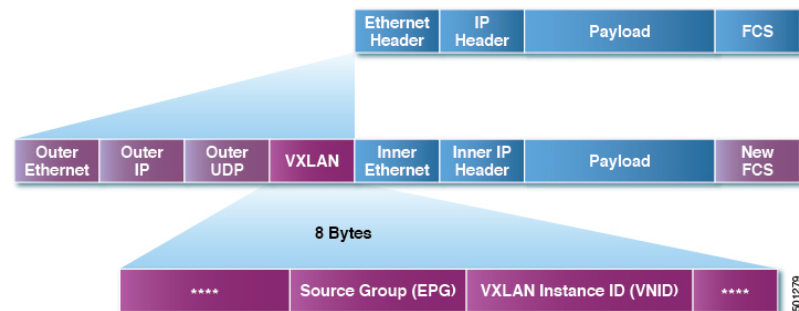
図 2: ACI カプセル化の正規化



ACI ファブリックでの転送は、カプセル化のタイプまたはカプセル化のオーバーレイ ネットワークによって制限または制約されません。ACI ブリッジドメインのフォワーディング ポリシーは、必要な場合に標準の VLAN 動作を提供するために定義できます。

ファブリック内のすべてのパケットに ACI ポリシー属性が含まれているため、ACI は完全に分散された方法でポリシーを一貫して適用できます。ACI により、アプリケーションポリシーの EPGID が転送から分離されます。次の図に示すように、ACI VXLAN ヘッダーは、ファブリック内のアプリケーション ポリシーを特定します。

図 3: ACI VXLAN のパケット形式



ACI VXLAN パケットには、レイヤ 2 の MAC アドレスとレイヤ 3 IP アドレスの送信元と宛先フィールド、ファブリック内の効率的な拡張性の転送を有効にします。ACI VXLAN パケットヘッダーの送信元グループフィールドは、パケットが属するアプリケーションポリシーエンドポイントグループ (EPG) を特定します。VXLAN インスタンス ID (VNID) は、テナントの仮想ルーティングおよび転送 (VRF) ドメインファブリック内で、パケットの転送を有効にします。VXLAN ヘッダーで 24 ビット VNID フィールドでは、同じネットワークで一意的なレイヤ 2 のセグメントを最大 16 個の拡張アドレス空間を提供します。この拡張アドレス空間は、大規模なマルチテナントデータセンターを構築する柔軟性 IT 部門とクラウドプロバイダーを提供します。

VXLAN を有効に ACI ファブリック全体にわたってスケールでの仮想ネットワークインフラストラクチャのレイヤ 3 のアンダーレイ レイヤ 2 を展開します。アプリケーションエンドポイントホスト柔軟に配置できます、アンダーレイ インフラストラクチャのレイヤ 3 バウンダリ

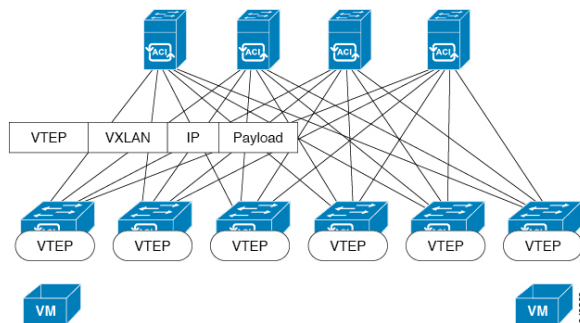
のリスクなしでデータセンターネットワーク間をオーバーレイネットワーク、VXLANでレイヤ2の隣接関係を維持します。

サブネット間のテナントトラフィックの転送を促進するレイヤ3 VNID

ACI ファブリックは、ACI ファブリック VXLAN ネットワーク間のルーティングを実行するテナントのデフォルトゲートウェイ機能を備えています。各テナントに対して、ファブリックはテナントに割り当てられたすべてのリーフスイッチにまたがる仮想デフォルトゲートウェイを提供します。これは、エンドポイントに接続された最初のリーフスイッチの入力インターフェイスで提供されます。各入力インターフェイスはデフォルトゲートウェイインターフェイスをサポートします。ファブリック全体のすべての入力インターフェイスは、特定のテナントサブネットに対して同一のルータのIPアドレスとMACアドレスを共有します。

ACI ファブリックは、エンドポイントのロケータまたは VXLAN トンネルエンドポイント (VTEP) アドレスで定義された場所から、テナントエンドポイントアドレスとその識別子を切り離します。ファブリック内の転送はVTEP間で行われます。次の図は、ACIで切り離されたIDと場所を示します。

図 4: ACIによって切り離された ID と場所



VXLAN は VTEP デバイスを使用してテナントのエンドデバイスを VXLAN セグメントにマッピングし、VXLAN のカプセル化およびカプセル化解除を実行します。各 VTEP 機能には、次の 2 つのインターフェイスがあります。

- ブリッジングを介したローカルエンドポイント通信をサポートするローカル LAN セグメントのスイッチインターフェイス
- 転送 IP ネットワークへの IP インターフェイス

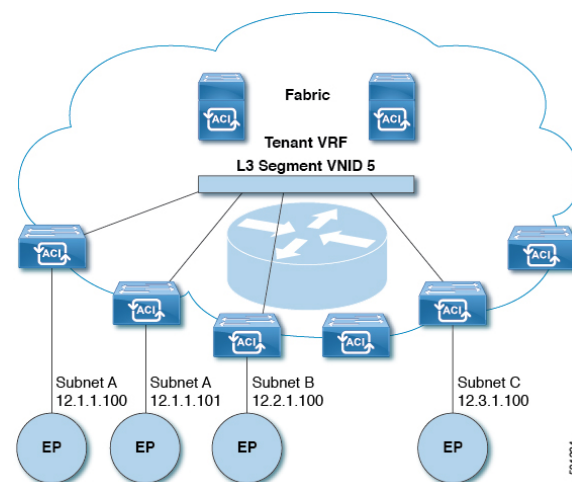
IP インターフェイスには一意の IP アドレスがあります。これは、インフラストラクチャ VLAN として知られる、転送 IP ネットワーク上の VTEP を識別します。VTEP デバイスはこの IP アドレスを使用してイーサネットフレームをカプセル化し、カプセル化されたパケットを、IP インターフェイスを介して転送ネットワークへ送信します。また、VTEP デバイスはリモート VTEP で VXLAN セグメントを検出し、IP インターフェイスを介してリモートの MAC Address-to-VTEP マッピングについて学習します。

ACI の VTEP は分散マッピングデータベースを使用して、内部テナントの MAC アドレスまたは IP アドレスを特定の場所にマッピングします。VTEP はルックアップの完了後に、宛先リーフスイッチ上の VTEP を宛先アドレスとして、VXLAN 内でカプセル化された元のデータパケットを送信します。宛先リーフスイッチはパケットをカプセル化解除して受信ホストに送信します。このモデルにより、ACI はスパニングツリープロトコルを使用することなく、フルメッシュでシングルホップのループフリートポロジを使用してループを回避します。

VXLAN セグメントは基盤となるネットワークトポロジに依存しません。逆に、VTEP 間の基盤となる IP ネットワークは、VXLAN オーバーレイに依存しません。これは送信元 IP アドレスとして開始 VTEP を持ち、宛先 IP アドレスとして終端 VTEP を持っており、外部 IP アドレスヘッダーに基づいてパケットをカプセル化します。

次の図は、テナント内のルーティングがどのように行われるかを示します。

図 5: ACI のサブネット間のテナントトラフィックを転送するレイヤ3 VNID



ACI はファブリックの各テナント VRF に単一の L3 VNID を割り当てます。ACI は、L3 VNID に従ってファブリック全体にトラフィックを転送します。出力リーフスイッチでは、ACI によって L3 VNID からのパケットが出力サブネットの VNID にルーティングされます。

ACI のファブリック デフォルト ゲートウェイに送信されてファブリック入力に到達したトラフィックは、レイヤ3 VNID にルーティングされます。これにより、テナント内でルーティングされるトラフィックはファブリックで非常に効率的に転送されます。このモデルを使用すると、たとえば同じ物理ホスト上の同じテナントに属し、サブネットが異なる 2 つの VM 間では、トラフィックが (最小パスコストを使用して) 正しい宛先にルーティングされる際に経由する必要があるは入力スイッチインターフェイスのみです。

ACI ルート リフレクタは、ファブリック内での外部ルートの配布にマルチプロトコル BGP (MP-BGP) を使用します。ファブリック管理者は自律システム (AS) 番号を提供し、ルートリフレクタにするスパインスイッチを指定します。



(注) Cisco ACI は IP フラグメンテーションをサポートしていません。したがって、外部ルータへのレイヤ 3 Outside (L3Out) 接続、または Inter-Pod Network (IPN) を介したマルチポッド接続を設定する場合は、インターフェイス MTU がリンクの両端で適切に設定することを推奨します。

IGP プロトコルパケット (EIGRP、OSPFv3) は、インターフェイス MTU サイズに基づいてコンポーネントによって構築されます。Cisco ACI では、CPU MTU サイズがインターフェイス MTU サイズよりも小さく、構築されたパケットサイズが CPU MTU より大きい場合、パケットはカーネルによってドロップされます (特に IPv6)。このような制御パケットのドロップを回避するには、コントロールプレーンとインターフェイスの両方で常に同じ MTU 値を設定します。

Cisco ACI、Cisco NX-OS、および Cisco IOS などの一部のプラットフォームでは、設定可能な MTU 値はイーサネットヘッダー (一致する IP MTU、14-18 イーサネットヘッダーサイズを除く) を考慮していません。また、IOS XR などの他のプラットフォームには、設定された MTU 値にイーサネットヘッダーが含まれています。設定された値が 9000 の場合、Cisco ACI、Cisco NX-OS および Cisco IOS の最大 IP パケットサイズは 9000 バイトになりますが、IOS-XR のタグなしインターフェイスの最大 IP パケットサイズは 8986 バイトになります。

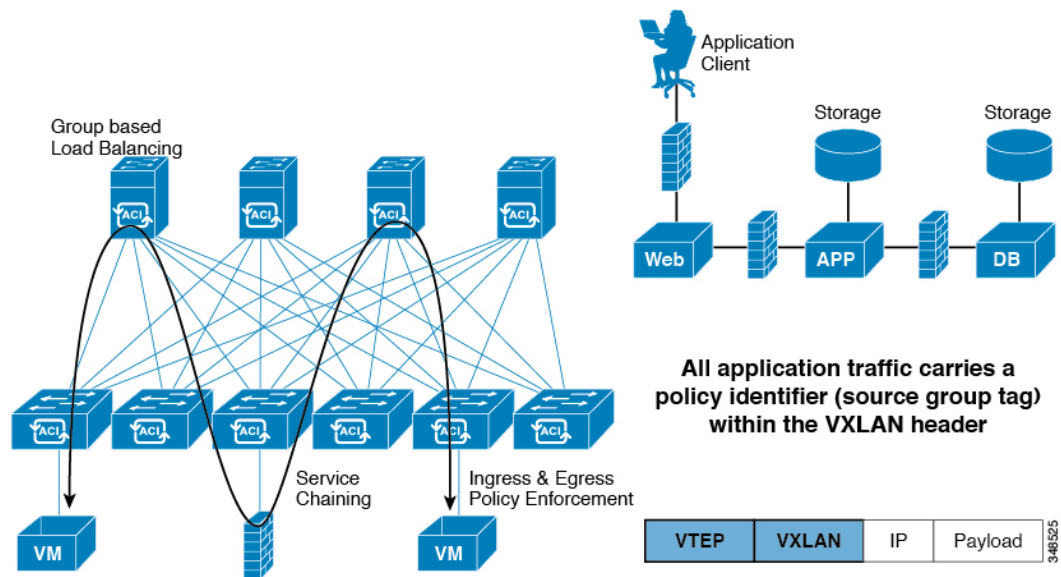
各プラットフォームの適切な MTU 値については、それぞれの設定ガイドを参照してください。

CLI ベースのコマンドを使用して MTU をテストすることを強く推奨します。たとえば、Cisco NX-OS CLI で、コマンド、`ping 1.1.1.1 df-bit packet-size 9000 source-interface ethernet 1/1` を使用してください。

ポリシー ID と適用

アプリケーションポリシーは、VXLAN パケットで送信される個別のタギング属性を使用して転送から分離されます。ポリシー ID は、ACI ファブリック内のすべてのパケットで送信され、完全に分散した形でポリシーの一貫した適用を行うことができます。次の図は、ポリシー ID を示します。

図 6: ポリシー ID と適用



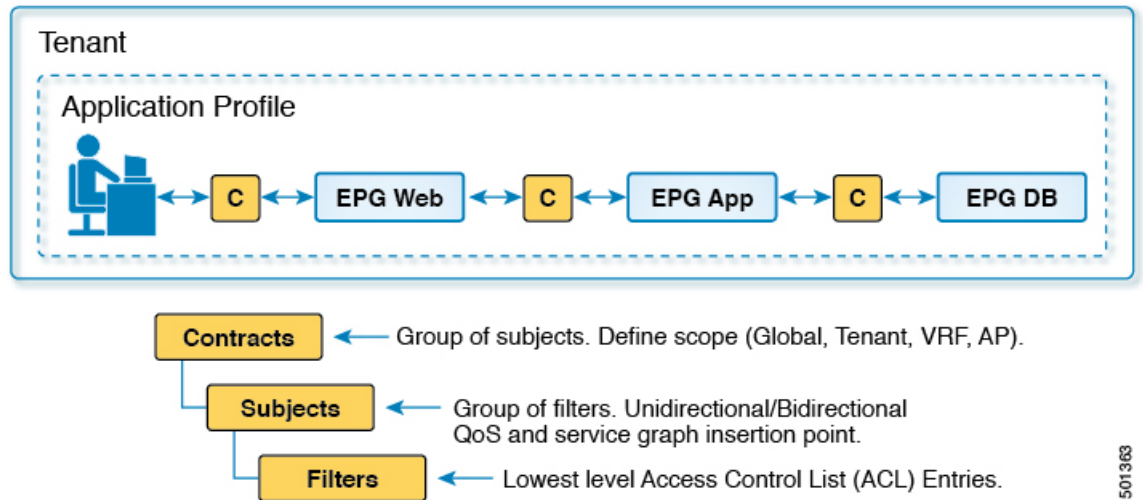
ファブリックおよびアクセスポリシーは、内部のファブリック インターフェイスおよび外部のアクセス インターフェイスの動作を管理します。システムは、デフォルトのファブリックおよびアクセスポリシーを自動的に作成します。ファブリックの管理者（ファブリック全体へのアクセス権がある者）は、要件に応じてデフォルトのポリシーを変更したり、新しいポリシーを作成できます。ファブリックおよびアクセスポリシーにより、さまざまな機能やプロトコルを有効にできます。APICのセレクタにより、ファブリックの管理者は、ポリシーを適用するノードおよびインターフェイスを選択できます。

ACI ファブリック ネットワーク アクセス セキュリティ ポリシー モデル (契約)

ACIのファブリック セキュリティ ポリシー モデルはコントラクトに基づいています。このアプローチにより、従来のアクセス コントロール リスト (ACL) の制限に対応できます。コントラクトには、エンドポイントグループ間のトラフィックで適用されるセキュリティポリシーの仕様が含まれます。

次の図は、契約のコンポーネントを示しています。

図 7: 契約のコンポーネント



501363

EPG 通信にはコントラクトが必要です。EPG/EPG 通信はコントラクトなしでは許可されません。APIC は、コントラクトや関連する EPG などのポリシーモデル全体を各スイッチの具象モデルにレンダリングします。入力時に、ファブリックに入るパケットはすべて、必要なポリシーの詳細でマークされます。EPG の間を通過できるトラフィックの種類を選択するためにコントラクトが必要とされるので、コントラクトはセキュリティポリシーを適用します。コントラクトは、従来のネットワーク設定でのアクセス コントロール リスト (ACL) によって扱われるセキュリティ要件を満たす一方で、柔軟性が高く、管理が容易な、包括的なセキュリティポリシー ソリューションです。

アクセス コントロール リストの制限

従来のアクセス コントロール リスト (ACL) には、ACI ファブリック セキュリティ モデルが対応する多数の制限があります。従来の ACL は、ネットワーク トポロジと非常に強固に結合されています。それらは通常、ルータまたはスイッチの入力および出力インターフェイスごとに設定され、そのインターフェイス、およびそれらのインターフェイスを流れることが予想されるトラフィックに合わせてカスタマイズされます。このカスタマイズにより、それらは多くの場合インターフェイス間で再利用できません。もちろんこれはルータまたはスイッチ間にも当てはまります。

従来の ACL は、非常に複雑で曖昧です。なぜなら、そのリストには、許可された特定の IP アドレス、サブネット、およびプロトコルのリストと、明確に許可されていない多くのものが含まれているためです。この複雑さは、問題が生じるのを管理者が懸念して ACL ルールを削除するのを躊躇するため、維持が困難で、多くの場合は増大するだけということを意味します。複雑さは、それらが通常 WAN と企業間または WAN とデータセンター間の境界などのネットワーク内の特定の境界ポイントでのみ配置されていることを意味します。この場合、ACL のセキュリティのメリットは、エンタープライズ内またはデータセンターに含まれるトラフィック向けには生かされません。

別の問題として、1 つの ACL 内のエントリ数の大幅増加が考えられます。ユーザは多くの場合、一連の送信元が一連のプロトコルを使用して一連の宛先と通信するのを許可する ACL を

作成します。最悪の場合、 N の送信元が K のプロトコルを使用して M の宛先と対話する場合、ACLに $N * M * K$ の行が存在する場合があります。ACLは、プロトコルごとに各宛先と通信する各送信元を一覧表示する必要があります。また、ACLが非常に大きくなる前に多くのデバイスやプロトコルを取得することはありません。

ACI ファブリック セキュリティ モデルは、これらの ACL の問題に処理します。ACI ファブリックセキュリティモデルは、管理者の意図を直接表します。管理者は、連絡先、フィルタ、およびラベルの管理対象オブジェクトを使用してエンドポイントのグループがどのように通信するかを指定します。これらの管理対象オブジェクトは、ネットワークのトポロジに関連していません。なぜなら、それらは特定のインターフェイスに適用されないためです。それらは、エンドポイントのこれらのグループの接続場所に関係なく、ネットワークが強要しなければならない簡易なルールです。このトポロジの独立性は、これらの管理対象オブジェクトが特定の境界ポイントとしてだけでなくデータセンター全体にわたって容易に配置して再利用できることを意味します。

ACI ファブリック セキュリティ モデルは、エンドポイントのグループ化コンストラクトを直接使用するため、サーバのグループが相互に通信できるようにするための概念はシンプルです。1つのルールにより、任意の数の送信元が同様に任意の数の宛先と通信することを可能にできます。このような簡略化により、そのスケールと保守性が大幅に向上します。つまり、データセンター全体でより簡単に使用できることにもつながります。

セキュリティ ポリシー仕様を含むコントラクト

ACI セキュリティ モデルでは、コントラクトに EPG 間の通信を管理するポリシーが含まれます。コントラクトは通信内容を指定し、EPGは通信の送信元と宛先を指定します。コントラクトは次のように EPG をリンクします。

EPG 1 ----- コントラクト ----- EPG 2

コントラクトで許可されていれば、EPG 1 のエンドポイントは EPG 2 のエンドポイントと通信でき、またその逆も可能です。このポリシーの構造には非常に柔軟性があります。たとえば、EPG 1 と EPG 2 間には多くのコントラクトが存在でき、1つのコントラクトを使用する EPG が 3つ以上存在でき、コントラクトは複数の EPG のセットで再利用できます。

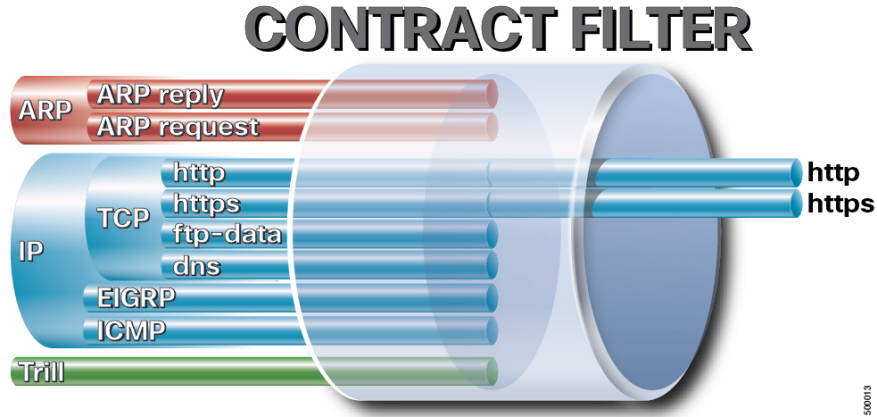
また EPG とコントラクトの関係には方向性があります。EPG はコントラクトを提供または消費できます。コントラクトを提供する EPG は通常、一連のクライアントデバイスにサービスを提供する一連のエンドポイントです。そのサービスによって使用されるプロトコルはコントラクトで定義されます。コントラクトを消費する EPG は通常、そのサービスのクライアントである一連のエンドポイントです。クライアント エンドポイント (コンシューマ) がサーバ エンドポイント (プロバイダー) に接続しようとする時、コントラクトはその接続が許可されるかどうかを確認します。特に指定のない限り、そのコントラクトは、サーバがクライアントへの接続を開始することを許可しません。ただし、EPG 間の別のコントラクトが、その方向の接続を簡単に許可する場合があります。

この提供/消費の関係は通常、EPG とコントラクト間を矢印を使って図で表されます。次に示す矢印の方向に注目してください。

EPG 1 <----- 消費 ----- コントラクト <----- 提供 ----- EPG 2

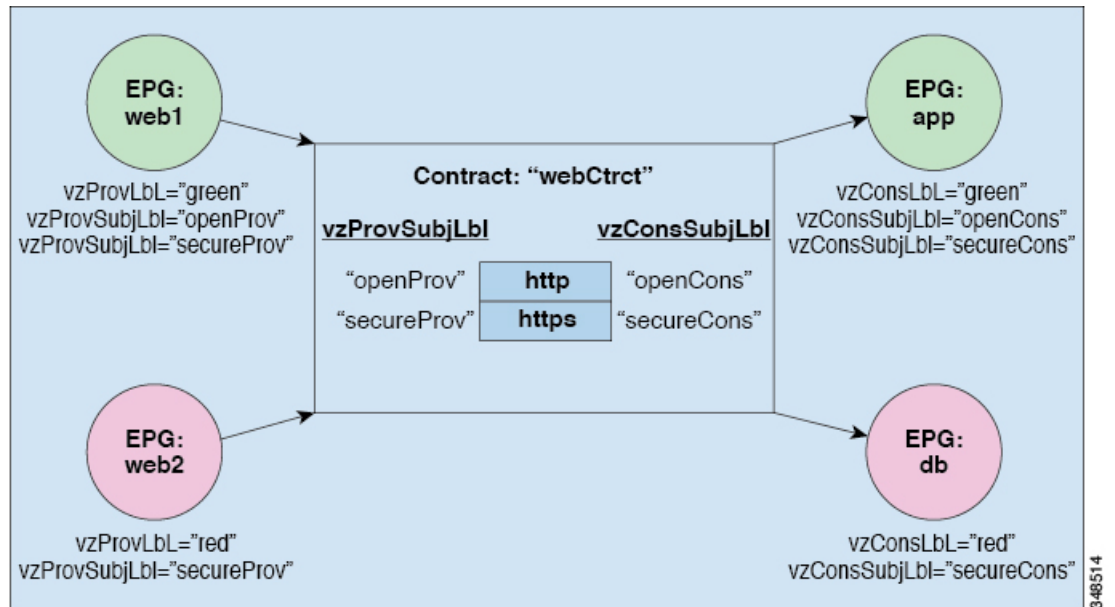
コントラクトは階層的に構築されます。1つ以上のサブジェクトで構成され、各サブジェクトには1つ以上のフィルタが含まれ、各フィルタは1つ以上のプロトコルを定義できます。

図 8: コントラクトフィルタ



次の図は、コントラクトが EPG の通信をどのように管理するかを示します。

図 9: EPG/EPG 通信を決定するコントラクト



たとえば、TCP ポート 80 とポート 8080 を指定する HTTP と呼ばれるフィルタと、TCP ポート 443 を指定する HTTPS と呼ばれる別のフィルタを定義できます。その後、2セットの情報カテゴリを持つ webCtrct と呼ばれるコントラクトを作成できます。openProv と openCons が HTTP フィルタが含まれる情報カテゴリです。secureProv と secureCons は HTTPS フィルタが含まれる情報カテゴリです。この webCtrct コントラクトは、Web サービスを提供する EPG とそのサービスを消費するエンドポイントを含む EPG 間のセキュアな Web トラフィックと非セキュアな Web トラフィックの両方を可能にするために使用できます。

これらの同じ構造は、仮想マシンのハイパーバイザを管理するポリシーにも適用されます。EPG が Virtual Machine Manager (VMM) のドメイン内に配置されると、APIC は EPG に関連付けられたすべてのポリシーを VMM ドメインに接続するインターフェイスを持つリーフスイッチにダウンロードします。VMM ドメインの完全な説明については、『*Application Centric Infrastructure Fundamentals*』の「*Virtual Machine Manager Domains*」の章を参照してください。このポリシーが作成されると、APIC は EPG のエンドポイントへの接続を可能にするスイッチを指定する VMM ドメインにそれをプッシュ（あらかじめ入力）します。VMM ドメインは、EPG 内のエンドポイントが接続できるスイッチとポートのセットを定義します。エンドポイントがオンラインになると、適切な EPG に関連付けられます。パケットが送信されると、送信元 EPG および宛先 EPG がパケットから取得され、対応するコントラクトで定義されたポリシーでパケットが許可されたかどうかを確認されます。許可された場合は、パケットが転送されます。許可されない場合は、パケットはドロップされます。

コントラクトは1つ以上のサブジェクトで構成されます。各サブジェクトには1つ以上のフィルタが含まれます。各フィルタには1つ以上のエントリが含まれます。各エントリは、アクセスコントロールリスト (ACL) の1行に相当し、エンドポイントグループ内のエンドポイントが接続されているリーフスイッチで適用されます。

詳細には、コントラクトは次の項目で構成されます。

- 名前：テナントによって消費されるすべてのコントラクト (**common** テナントまたはテナント自体で作成されたコントラクトを含む) にそれぞれ異なる名前が必要です。
- サブジェクト：特定のアプリケーションまたはサービス用のフィルタのグループ。
- フィルタ：レイヤ2～レイヤ4の属性(イーサネットタイプ、プロトコルタイプ、TCPフラグ、ポートなど)に基づいてトラフィックを分類するために使用します。
- アクション：フィルタリングされたトラフィックで実行されるアクション。次のアクションがサポートされます。
 - トラフィックの許可 (通常のコントラクトのみ)
 - トラフィックのマーク (DSCP/CoS) (通常のコントラクトのみ)
 - トラフィックのリダイレクト (サービス グラフによる通常のコントラクトのみ)
 - トラフィックのコピー (サービスグラフまたはSPANによる通常のコントラクトのみ)
 - トラフィックのブロック (禁止コントラクトのみ)

Cisco APIC リリース 3.2(x) および名前が EX または FX で終わるスイッチでは、標準コントラクトで代わりに件名 [拒否] アクションまたは [コントラクトまたは件名の除外] を使用して、指定のパターンを持つトラフィックをブロックできます。

 - トラフィックのログ (禁止コントラクトと通常のコントラクト)
- エイリアス：(任意)変更可能なオブジェクト名。オブジェクト名は作成後に変更できませんが、エイリアスは変更できるプロパティです。

このように、コントラクトによって許可や拒否よりも複雑なアクションが可能になります。コントラクトは、所定のサブジェクトに一致するトラフィックをサービスにリダイレクトした

り、コピーしたり、その QoS レベルを変更したりできることを指定可能です。具象モデルでアクセス ポリシーをあらかじめ入力すると、APIC がオフラインまたはアクセスできない場合でも、エンドポイントは移動でき、新しいエンドポイントをオンラインにでき、通信を行うことができます。APIC は、ネットワークの単一の障害発生時点から除外されます。ACI ファブリックにパケットが入力されると同時に、セキュリティポリシーがスイッチで実行している具象モデルによって適用されます。

セキュリティ ポリシーの適用

トラフィックは前面パネルのインターフェイスからリーフスイッチに入り、パケットは送信元 EPG の EPG でマーキングされます。リーフスイッチはその後、テナントエリア内のパケットの宛先 IP アドレスでフォワーディングルックアップを実行します。ヒットすると、次のシナリオのいずれかが発生する可能性があります。

1. ユニキャスト (/32) ヒットでは、宛先エンドポイントの EPG と宛先エンドポイントが存在するローカルインターフェイスまたはリモートリーフスイッチの VTEP IP アドレスが提供されます。
2. サブネットプレフィクス (/32 以外) のユニキャストヒットでは、宛先サブネットプレフィクスの EPG と宛先サブネットプレフィクスが存在するローカルインターフェイスまたはリモートリーフスイッチの VTEP IP アドレスが提供されます。
3. マルチキャストヒットでは、ファブリック全体の VXLAN カプセル化とマルチキャストグループの EPG で使用するローカルレシーバのローカルインターフェイスと外側の宛先 IP アドレスが提供されます。



- (注) マルチキャストと外部ルータのサブネットは、入力リーフスイッチでのヒットを常にもたらし、セキュリティポリシーの適用は、宛先 EPG が入力リーフスイッチによって認識されるとすぐに発生します。

転送テーブルの誤りにより、パケットがスパインスイッチの転送プロキシに送信されます。転送プロキシはその後、転送テーブル検索を実行します。これが誤りである場合、パケットはドロップされます。これがヒットの場合、パケットは宛先エンドポイントを含む出力リーフスイッチに送信されます。出力リーフスイッチが宛先の EPG を認識するため、セキュリティポリシーの適用が実行されます。出力リーフスイッチは、パケット送信元の EPG を認識する必要があります。ファブリックヘッダーは、入力リーフスイッチから出力リーフスイッチに EPG を伝送するため、このプロセスをイネーブルにします。スパインスイッチは、転送プロキシ機能を実行するときに、パケット内の元の EPG を保存します。

出力リーフスイッチでは、送信元 IP アドレス、送信元 VTEP、および送信元 EPG 情報は、学習によってローカルの転送テーブルに保存されます。ほとんどのフローが双方向であるため、応答パケットがフローの両側で転送テーブルに入力し、トラフィックが両方向で入力フィルタリングされます。

マルチキャストおよび EPG セキュリティ

マルチキャストトラフィックでは、興味深い問題が起こります。ユニキャストトラフィックでは、宛先 EPG はパケットの宛先の検査からはっきり知られています。ただし、マルチキャストトラフィックでは、宛先は抽象的なエンティティ、マルチキャストグループです。パケットの送信元はマルチキャストアドレスではないため、送信元 EPG は以前のユニキャストの例と同様に決定されます。宛先グループの起源はマルチキャストが異なる場所です。

マルチキャストグループが、ネットワークトポロジから若干独立しているため、グループバインディングへの (S, G) および (*, G) の静的設定は受け入れ可能です。マルチキャストグループが転送テーブルにある場合、マルチキャストグループに対応する EPG は、転送テーブルにも配置されます。



(注) このマニュアルでは、マルチキャストグループとしてマルチキャストストリームを参照しません。

リーフスイッチは、マルチキャストストリームに対応するグループを常に宛先 EPG と見なし、送信元 EPG と見なすことはありません。前述のアクセスコントロールマトリクスでは、マルチキャスト EPG が送信元の場合は行の内容は無効です。トラフィックは、マルチキャストストリームの送信元またはマルチキャストストリームに加わりたい宛先からマルチキャストストリームに送信されます。マルチキャストストリームが転送テーブルにある必要があり、ストリーム内に階層型アドレッシングがないため、マルチキャストトラフィックは、入力ファブリックの端でアクセスが制御されます。その結果、IPv4 マルチキャストは入力フィルタリングとして常に適用されます。

マルチキャストストリームの受信側は、トラフィックを受信する前にマルチキャストストリームに最初に加わる必要があります。IGMP Join 要求を送信すると、マルチキャストレシーバは実際に IGMP パケットの送信元になります。宛先はマルチキャストグループとして定義され、宛先 EPG は転送テーブルから取得されます。ルータが IGMP Join 要求を受信する入力点で、アクセス制御が適用されます。Join 要求が拒否された場合、レシーバはその特定のマルチキャストストリームからトラフィックを受信しません。

マルチキャスト EPG へのポリシーの適用は、前述のようにコントラクトのルールに従ってリーフスイッチにより入力時に発生します。また、EPG バインディングに対するマルチキャストグループは、APIC によって特定のテナント (VRF) を含むすべてのリーフスイッチにプッシュされます。

マルチキャスト ツリー トポロジ

ACI ファブリックは、アクセスポートからのユニキャスト、マルチキャスト、およびブロードキャストトラフィックの転送をサポートします。エンドポイントホストからのすべてのマルチキャストエネーショントラフィックは、ファブリックにマルチキャストトラフィックとして伝送されます。

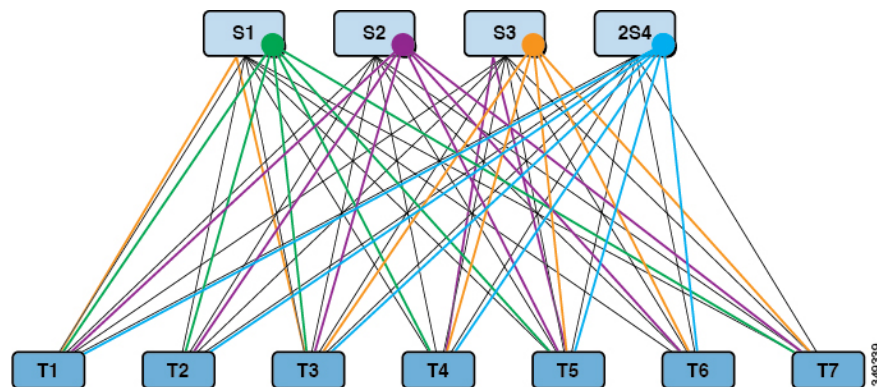
ACI ファブリックは、入力インターフェイスに入るトラフィックを使用可能な中間ステージのスパインスイッチを介して関連する出力スイッチにルーテッドできる Clos トポロジ (Charles Clos にちなんで名付けられた) に接続されるスパインおよびリーフスイッチで構成されます。リーフスイッチには次の2種類のポートがあります。スパインスイッチに接続するためのファブリックポートと、サーバー、サービスアプライアンス、ルータ、Fabric Extender (FEX; ファブリックエクステンダ)などを接続するアクセスポートです。

リーフスイッチ (top-of-rack (ToR; トップオブラック) スイッチとも呼ばれます) は、スパインスイッチ (「エンドオブロウ」または「EoR」スイッチとも呼ばれます) に接続されます。リーフスイッチは互いに接続されず、スパインスイッチはリーフスイッチのみに接続します。この Clos トポロジでは、すべての下位層のスイッチがフルメッシュ トポロジの最上位層のスイッチにそれぞれ接続されます。スパインスイッチに不具合があると、ACI ファブリック全体のパフォーマンスだけがわずかに低下します。データパスは、トラフィック負荷がスパインスイッチ間で均等に分散されるように選択されます。

ACI ファブリックは、Forwarding Tag (FTAG) ツリーを使用してバランスマルチデスティネーショントラフィックをロードします。すべてのマルチデスティネーショントラフィックは、ファブリック内でカプセル化された IP マルチキャストトラフィックの形式で転送されます。入力リーフは、FTAG をスパインに転送するときにトラフィックに割り当てます。FTAG は接続先マルチキャストアドレスの一部としてパケットに割り当てられます。ファブリックでは、トラフィックは指定された FTAG ツリーに沿って転送されます。スパインおよび中間リーフスイッチは、FTAG ID に基づいてトラフィックを転送します。転送ツリーは、FTAG ID 1 つにつき 1 つ構築されます。任意の 2 つのノード間で、FTAG 1 つにつきリンク 1 つだけが転送されます。複数の FTAG を使用することで、転送に異なるリンクを使用している各 FTAG でパラレルリンクを使用できます。ファブリック内の FTAG ツリーの数が多いほど、ロードバランシングの効果が大きい可能性があるということになります。ACI ファブリックは、最大 12 個の FTAG をサポートします。

次の図は、4 つの FTAG によるトポロジを示します。ファブリック内のすべてのリーフスイッチは、各 FTAG に直接または中継ノードを介して接続されます。1 つの FTAG が各スパインノードに根付いています。

図 10: マルチキャスト ツリー トポロジ



リーフスイッチはスパインへの直接接続性がある場合、直接パスを使用して FTAG ツリーに接続します。直接リンクがない場合、リーフスイッチは上記の図に示すように FTAG ツリーに接

続されている中継ノードを使用します。図には、各スパインが1つのFTAGツリーのルートとして示されていますが、複数のFTAGツリールートを1つのノード上に置くことができます。

ACI ファブリック 起動検出プロセスの一環として、FTAG ルートはスパインスイッチに配置されます。APIC は、各スパインスイッチをスパインがアンカーする FTAG で構成します。ルートの ID と FTAG の数は構成から取得されます。APIC は、使用される FTAG ツリーの数と各ツリーに対するルートを指定します。FTAG ツリーは、ファブリックでトポロジの変更があるたびに再計算されます。

ルートの配置は誘導される構成で、スパインスイッチの障害などのランタイム イベントで動的に再度ルート付けされることはありません。通常、FTAG 構成は静的です。スパインスイッチの追加または削除時は、管理者がスパインスイッチの残りのセットまたは拡張セット間で FTAG を再配布することを決める可能性があるため、FTAG はあるスパインから別のスパインへ再アンカーできます。

トラフィック ストーム制御について

トラフィック ストームは、パケットが LAN でフラッドする場合に発生するもので、過剰なトラフィックを生成し、ネットワークのパフォーマンスを低下させます。トラフィック ストーム制御ポリシーを使用すると、物理インターフェイス上におけるブロードキャスト、未知のマルチキャスト、または未知のユニキャストのトラフィック ストームによって、レイヤ 2 ポート経由の通信が妨害されるのを防ぐことができます。

デフォルトでは、ストーム制御は ACI ファブリックでは有効になっていません。ACI ブリッジドメイン (BD) レイヤ 2 の未知のユニキャストのフラッドは BD 内でデフォルトで有効になっていますが、管理者が無効にすることができます。その場合、ストーム制御ポリシーはブロードキャストと未知のマルチキャストのトラフィックにのみ適用されます。レイヤ 2 の未知のユニキャストのフラッドが BD で有効になっている場合、ストーム制御ポリシーは、ブロードキャストと未知のマルチキャストのトラフィックに加えて、レイヤ 2 の未知のユニキャストのフラッドに適用されます。

トラフィック ストーム制御 (トラフィック抑制ともいいます) を使用すると、着信するブロードキャスト、マルチキャスト、未知のユニキャストのトラフィックのレベルを 1 秒間隔でモニタできます。この間に、トラフィック レベル (ポートで使用可能な合計帯域幅のパーセンテージ、または特定のポートで許可される 1 秒あたりの最大パケット数として表されます) が、設定したトラフィック ストーム制御レベルと比較されます。入力トラフィックが、ポートに設定したトラフィック ストーム制御レベルに到達すると、トラフィック ストーム制御機能によってそのインターバルが終了するまでトラフィックがドロップされます。管理者は、ストーム制御しきい値を超えたときにエラーを発生させるようにモニタリングポリシーを設定できます。

ストーム制御の注意事項と制約事項

以下のガイドラインと制約事項に従って、トラフィック ストーム制御レベルを設定してください。

- 通常、ファブリック管理者は以下のインターフェイスのファブリック アクセス ポリシーでストーム制御を設定します。
 - 標準トランク インターフェイス。
 - 単一リーフ スイッチ上のダイレクト ポート チャネル。
 - バーチャル ポート チャネル (2つのリーフ スイッチ上のポート チャネル)。
- リリース 4.2(1)以降では、ストーム制御のしきい値に達した場合に、次の制約事項に従って、SNMP トラップをCisco Application Centric Infrastructure (ACI) からトリガーできるようになりました。
 - ストーム制御に関連するアクションには、ドロップとシャットダウンの2つがあります。シャットダウンアクションでは、インターフェイストラップが発生しますが、ストームがアクティブまたはクリアであることを示すためのストーム制御トラップは、シャットダウンアクションによっては決定されません。したがって、ポリシーでシャットダウンアクションが設定されているストーム制御トラップは無視する必要があります。
 - ストーム制御ポリシーがオンの状態でポートがフラップすると、統計情報の収集時にクリアトラップとアクティブトラップが一緒に表示されます。通常、クリアトラップとアクティブトラップは一緒に表示されませんが、この場合は予期される動作です。
- ポートチャネルおよびバーチャルポートチャネルでは、ストーム制御値 (1秒あたりのパケット数またはパーセンテージ) はポートチャネルのすべての個別メンバーに適用されます。



(注) Cisco Application Policy Infrastructure Controller (APIC) リリース 1.3(1)およびスイッチリリース 11.3(1)以降のスイッチハードウェアの場合、ポートチャネル設定では、集約ポートのトラフィック抑制は設定値の最大2倍になることがあります。新しいハードウェアポートは slice-0 と slice-1 の2つのグループに内部的にさらに分割されています。スライスマップを確認するには、vsh_lc コマンドの show platform internal hal 12 port gpd を使用して、s1 カラムで slice 0 または slice 1 を探します。ポートチャネルメンバーがスライス 0 とスライス 1 の両方に該当する場合、式は各スライスに基づいて計算されるため、許可されるストーム制御トラフィックが設定値の2倍になることがあります。

- 使用可能な帯域幅のパーセンテージで設定する場合、値 100 はトラフィックストーム制御を行わないことを意味し、値 0.01 はすべてのトラフィックを抑制します。
- ハードウェアの制限およびさまざまなサイズのパケットのカウンタ方式が原因で、レベルのパーセンテージは概数になります。着信トラフィックを構成するフレームのサイズに応

じて、実際に適用されるパーセンテージ レベルと設定したパーセンテージ レベルの間には、数パーセントの誤差がある可能性があります。1 秒あたりのパケット数 (PPS) の値は、256 バイトに基づいてパーセンテージに変換されます。

- 最大バーストは、通過するトラフィックがないときに許可されるレートの最大累積です。トラフィックが開始されると、最初の間隔では累積レートまでのすべてのトラフィックが許可されます。後続の間隔では、トラフィックは設定されたレートまでのみ許可されます。サポートされる最大数は65535 KBです。設定されたレートがこの値を超えると、PPS とパーセンテージの両方についてこの値で制限されます。
- 累積可能な最大バーストは 512 MB です。
- 最適化されたマルチキャストフラディング (OMF) モードの出力リーフスイッチでは、トラフィック ストーム制御は適用されません。
- OMF モードではない出力リーフスイッチでは、トラフィック ストーム制御が適用されません。
- FEX のリーフスイッチでは、ホスト側インターフェイスにはトラフィック ストーム制御を使用できません。
- Cisco Nexus C93128TX、C9396PX、C9396TX、C93120TX、C9332PQ、C9372PX、C9372TX、C9372PX-E、C9372TX-E の各スイッチでは、トラフィック ストーム制御のユニキャスト/マルチキャストの差別化がサポートされていません。
- Cisco Nexus C93128TX、C9396PX、C9396TX、C93120TX、C9332PQ、C9372PX、C9372TX、C9372PX-E、C9372TX-E の各スイッチでは、トラフィック ストーム制御の SNMP トラップがサポートされていません。
- Cisco Nexus C93128TX、C9396PX、C9396TX、C93120TX、C9332PQ、C9372PX、C9372TX、C9372PX-E、C9372TX-E の各スイッチでは、トラフィック ストーム制御トラップがサポートされていません。
- ストーム制御アクションは、物理イーサネット インターフェイスおよびポート チャネル インターフェイスでのみサポートされます。

リリース 4.1(1)以降では、ストーム制御シャットダウンオプションがサポートされています。デフォルトの Soak Instance Count を持つインターフェイスに対してシャットダウンアクションが選択されると、しきい値を超えるパケットは 3 秒間ドロップされ、ポートは 3 秒間シャットダウンされます。デフォルトのアクションは、ドロップです。シャットダウンアクションを選択すると、ユーザーはソーキング間隔を指定するオプションを使用できます。デフォルトのソーキング間隔は 3 秒です。設定可能な範囲は 3 ~ 10 秒です。

- インターフェイスに設定されたデータプレーンポリシング (DPP) ポリサーの値がストームポリサーの値よりも低い場合、DPP ポリサーが優先されます。DPP ポリサーとストームポリサーの間に設定されている低い方の値が、設定されたインターフェイスで適用されます。
- リリース 4.2(6)以降、ストームポリサーは、DHCP、ARP、ND、HSRP、PIM、IGMP、および EIGRP プロトコルに対応する、リーフスイッチのすべての転送制御トラフィックに強制されます。このことは、ブリッジドメインが **BD** でのフラディングまたはカプセル

化でのフラッディングのどちらに設定されているかには関係しません。この動作の変更は、EX 以降のリーフスイッチにのみ適用されます。

- EX スイッチでは、プロトコルの1つに対し、スーパーバイザ ポリサーとストーム ポリサーの両方を設定できます。この場合、サーバーが設定されたスーパーバイザ ポリサー レート（制御プレーン ポリシング、CoPP）よりも高いレートでトラフィックを送信すると、ストーム ポリサーはストーム ポリサー レートとして設定されているよりも多くのトラフィックを許可します。着信トラフィック レートがスーパーバイザ ポリサー レート以下の場合、ストーム ポリサーは設定されたストーム トラフィック レートを正しく許可します。この動作は、設定されたスーパーバイザポリサーおよびストーム ポリサーのレートに関係なく適用されます。
- ストーム ポリサーが、指定されたプロトコルのリーフ スイッチで転送されるすべての制御トラフィックに適用されるようになった結果、リーフスイッチで転送される制御トラフィックがストーム ポリサー ドロップの対象になります。以前のリリースでは、この動作の変更の影響を受けるプロトコルでは、このようなストーム ポリサーのドロップは発生しません。
- トラフィック ストーム制御は、PIM が有効になっているブリッジ ドメインまたは VRF インスタンスのマルチキャスト トラフィックをポリシングできません。
- ストーム コントロール ポリサーがポートチャネルインターフェイスに適用されている場合、許可されるレートが設定されているレートを超えることがあります。ポートチャネルのメンバーリンクが複数のスライスにまたがる場合、許可されるトラフィックレートは、構成されたレートにメンバーリンクがまたがるスライスの数を掛けたものに等しくなります。

ポートからスライスへのマッピングは、スイッチ モデルによって異なります。

例として、ストーム ポリサー レートが 10Mbps のメンバー リンク port1、port2、および port3 を持つポートチャネルがあるとします。

- port1、port2、port3 が slice1 に属している場合、トラフィックは 10Mbps にポリシングされます。
- port1 と port2 が slice1 に属し、port3 が slice2 に属している場合、トラフィックは 20Mbps にポリシングされます。
- port1 が slice1 に属し、port2が slice2 に属し、port3 が slice3 に属している場合、トラフィックは 30Mbps にポリシングされます。

ファブリック ロード バランシング

Cisco Application Centric Infrastructure (ACI) ファブリックでは、利用可能なアップリンク リンク間のトラフィックを平衡化するためのロード バランシング オプションがいくつか提供されます。ここでは、リーフからスパインへのスイッチ トラフィックのロード バランシングについて説明します。

スタティックハッシュロードバランシングは、各フローが5タプルのハッシュに基づいてアップリンクに割り当てられるネットワークで使用される従来のロードバランシング機構です。このロードバランシングにより、使用可能なリンクにほぼ均等な流量が分配されます。通常、流量が多いと、流量の均等な分配により帯域幅も均等に分配されます。ただし、いくつかのフローが残りよりも多いと、スタティックロードバランシングにより完全に最適ではない結果がもたらされる場合があります。

Cisco ACI ファブリック ダイナミックロードバランシング (DLB) は、輻輳レベルに従ってトラフィック割り当てを調整します。DLBでは、使用可能なパス間の輻輳が測定され、輻輳状態が最も少ないパスにフローが配置されるので、データが最適またはほぼ最適に配置されます。

DLBは、フローまたはフローレットの粒度を使用して使用可能なアップリンクにトラフィックを配置するように設定できます。フローレットは、時間の大きなギャップによって適切に区切られるフローからのパケットのバーストです。パケットの2つのバースト間のアイドル間隔が使用可能なパス間の遅延の最大差より大きい場合、2番目のバースト（またはフローレット）を1つ目とは異なるパスに沿ってパケットのリオーダーなしで送信できます。このアイドル間隔は、フローレットタイマーと呼ばれるタイマーによって測定されます。フローレットにより、パケットリオーダーを引き起こすことなくロードバランシングに対する粒度の高いフローの代替が提供されます。

DLB 動作モードは積極的または保守的です。これらのモードは、フローレットタイマーに使用するタイムアウト値に関係します。アグレッシブモードのフローレットタイムアウトは比較的小さい値です。この非常に精密なロードバランシングはトラフィックの分配に最適ですが、パケットリオーダーが発生する場合があります。ただし、アプリケーションのパフォーマンスに対する包括的なメリットは、保守的なモードと同等かそれよりも優れています。保守的なモードのフローレットタイムアウトは、パケットが並び替えられないことを保証する大きな値です。新しいフローレットの機会の頻度が少ないので、トレードオフは精度が低いロードバランシングです。DLBは常に最も最適なロードバランシングを提供できるわけではありませんが、スタティックハッシュロードバランシングより劣るということはありません。



- (注) すべてのNexus 9000シリーズスイッチにはDLBのハードウェアサポートがありますが、DLB機能は、第2世代プラットフォーム (EX、FX、およびFX2 サフィックスを持つスイッチ) の現在のソフトウェアリリースでは有効になっていません。

Cisco ACI ファブリックは、リンクがオフラインまたはオンラインになったことで使用可能なリンク数が変化すると、トラフィックを調整します。ファブリックは、リンクの新しいセットでトラフィックを再分配します。

スタティックまたはダイナミックのロードバランシングのすべてのモードでは、トラフィックは、Equal Cost Multipath (ECMP) の基準を満たすアップリンクまたはパス上でのみ送信され、これらのパスはルーティングの観点から同等で最もコストがかかりません。

ロードバランシング技術ではありませんが、Dynamic Packet Prioritization (DPP) は、スイッチでDLBと同じメカニズムをいくつか使用します。DPPの設定はDLB専用です。DPPは、長いフローよりも短いフローを優先します。短いフローは約15パケット未満です。短いフローは

長いフローよりも遅延の影響を受けやすいため、DPPはアプリケーション全体のパフォーマンスを向上させることができます。

6.0 (1) および6.0 (2) リリースでは、イントラリーフスイッチトラフィックの場合、カスタム QoS 構成に関係なく、すべての DPP 優先トラフィックが CoS 0 とマークされます。リーフ間スイッチトラフィックの場合、カスタム QoS 構成に関係なく、すべての DPP 優先トラフィックが CoS 3 とマークされます。6.0 (3) リリース以降、イントラリーフスイッチおよびインターリーフスイッチのトラフィックでは、カスタム QoS 構成に関係なく、すべての DPP 優先トラフィックが CoS 0 とマークされます。

GPRS トンネリングプロトコル (GTP) は、主にワイヤレスネットワークでデータを配信するために使用されます。Cisco Nexus スイッチは Telcom データセンター内の場所です。パケットがデータセンターの Cisco Nexus 9000 スイッチを介して送信される場合、トラフィックは GTP ヘッダーに基づいてロード バランシングされる必要があります。ファブリックがリンク バンドルを介して外部ルータに接続されている場合、トラフィックはすべてのバンドル メンバー (たとえば、レイヤ 2 ポート チャネル、レイヤ 3 ECMP リンク、レイヤ 3 ポート チャネル、およびポートチャネル上の L3Out) に均等に分散される必要があります。)。GTP トラフィックのロード バランシングは、ファブリック内でも実行されます。

GTP ロード バランシングを実現するために、Cisco Nexus 9000 シリーズ スイッチは 5 タプルのロード バランシング メカニズムを使用します。ロード バランシング メカニズムでは、パケットの送信元 IP、宛先 IP、プロトコル、レイヤ 4 リソース、および宛先ポート (トラフィックが TCP または UDP の場合) フィールドが考慮されます。GTP トラフィックの場合は、これらのフィールドへの一意の値の数が限られていると、トンネルでのトラフィック ロードの均等分散が制限されます。

ロード バランシングにおける GTP トラフィックの極性を回避するために、GTP ヘッダーのトンネル エンドポイント ID (TEID) が UDP ポート番号の代わりに使用されます。TEID がトンネルごとに異なるため、トラフィックをバンドルの複数のリンク間で均等にロード バランシングすることができます。

GTP ロード バランシングは、GTPU パケットに存在する 32 ビット TEID 値で送信元および宛先ポート情報を上書きします。

GTP トンネルのロード バランシング機能により、次のサポートが追加されます。

- 物理インターフェイスでの IPv4/IPv6 トランスポート ヘッダーによる GTP
- UDP ポート 2152 を使用した GTPU

Cisco ACI ファブリックのデフォルト構成では、従来の静的なハッシュが使用されます。スタティックなハッシュ機能により、アップリンク間のトラフィックがリーフスイッチからスパインスイッチに分配されます。リンクがダウンまたは起動すると、すべてのリンクのトラフィックが新しいアップリンク数に基づいて再分配されます。

リーフ/スパインスイッチ ダイナミック ロード バランシング アルゴリズム

次の表に、リーフ/スパインスイッチ ダイナミック ロード バランシングで使用されるデフォルトの設定不可能なアルゴリズムを示します。

表 1: Cisco ACIリーフ/スパインスイッチ ダイナミック ロード バランシング

Traffic Type	データ ポイントのハッシュ
リーフ/スパイン IP ユニキャスト	<ul style="list-style-type: none"> • 送信元 MAC アドレス • 宛先 MAC アドレス • 送信元 IP アドレス • 宛先 IP アドレス • プロトコル タイプ • 送信元レイヤ 4 ポート • 宛先レイヤ 4 ポート • セグメント ID (VXLAN VNID) または VLAN ID
リーフ/スパイン レイヤ 2	<ul style="list-style-type: none"> • 送信元 MAC アドレス • 宛先 MAC アドレス • セグメント ID (VXLAN VNID) または VLAN ID

エンドポイントの保持

スイッチでキャッシュ エンドポイントの MAC アドレスと IP アドレスを保持することで、パフォーマンスが向上します。スイッチは、アクティブになるときにエンドポイントについて学習します。ローカル エンドポイントはローカル スイッチにあります。リモート エンドポイントは他のスイッチにあります。ローカルでキャッシュされます。リーフスイッチは、直接（または直接接続されたレイヤ 2 スイッチまたはファブリックエクステンダを通じて）接続されたエンドポイント、ローカル エンドポイント、およびファブリックの他のリーフスイッチに接続されたエンドポイント（ハードウェアのリモート エンドポイント）に関する場所とポリシーの情報を保存します。スイッチは、ローカル エンドポイントには 32 Kb エントリ キャッシュを、リモート エンドポイントには 64 Kb エントリ キャッシュを使用します。

リーフスイッチで稼働するソフトウェアは、これらのテーブルを能動的に管理します。ローカル的に接続されたエンドポイントでは、ソフトウェアは各エントリの保持タイマーの期限切れ後にエントリをエージングアウトします。エンドポイント エントリは、エンドポイントのアクティビティが終了するとスイッチキャッシュからプルーニングされ、エンドポイントの場所が他のスイッチに移動するか、またはライフサイクルの状態がオフラインに変わります。ローカル保持タイマーのデフォルト値は15分です。非アクティブのエントリを削除する前に、リーフスイッチはエンドポイントに3つの ARP 要求を送信し、実際になくなっているかを確認します。スイッチが ARP 応答を受信しない場合、エントリはプルーニングされます。リモート

で接続されたエンドポイントの場合、スイッチは非アクティブになってから5分後にエントリをエージングアウトします。リモートエンドポイントは、再度アクティブになるとテーブルにすぐに再入力されます。



(注) バージョン 1.3(1g) では、仮想ホストおよびローカルホストに対してトリガーされるサイレントホストトラッキングが追加されています。

エンドポイントが再度キャッシュされるまでリモートリーフスイッチで適用されるポリシー以外にテーブルにリモートエンドポイントがなくても、パフォーマンスのペナルティはありません。

ブリッジドメインのサブネットが *enforced* に構成されている場合、エンドポイント保持ポリシーは次のように動作します。

- ブリッジドメインのサブネットに含まれていない IP アドレスを持つ新しいエンドポイントは学習されません。
- デバイスが追跡に応答しない場合、学習済みのエンドポイントはエンドポイント保持キャッシュからエージアウトします。

この実施プロセスは、サブネットがブリッジドメインで定義されているかどうか、またはサブネットが EPG で定義されているかどうかに関係なく、同じように動作します。

エンドポイントの保持タイマーポリシーは変更できます。静的エンドポイントの MAC および IP アドレスを設定すると、保持タイマーをゼロに設定することで、スイッチキャッシュに永久的に保存できます。エントリの保持タイマーをゼロに設定することは、それが自動的に削除されないことを意味します。この操作は慎重に行う必要があります。エンドポイントが移動したりポリシーが変化する場合、APIC を介してエントリを手動で最新情報に更新する必要があります。保持タイマーがゼロ以外の場合、この情報は APIC の介入なしで各パケットで確認されれば瞬時に更新されます。

エンドポイントの保持ポリシーは、プルーニングがどのように行われるかを決定します。ほとんどの場合、デフォルトのポリシーアルゴリズムが使用されます。エンドポイントの保持ポリシーを変更すると、システムパフォーマンスに影響を与える場合があります。何千ものエンドポイントと通信するスイッチの場合、エージング間隔を短くすると、多数のアクティブなエンドポイントをサポートするのに使用可能なキャッシュウィンドウの数が増えます。エンドポイントの数が 10,000 を超える場合は、複数のスイッチにエンドポイントを分散させることを推奨します。

デフォルトのエンドポイント保持ポリシーの変更に関しては、次のガイドラインに従ってください。

- リモート バウンス間隔 = (リモート エージ * 2) + 30 秒
 - 推奨されるデフォルト値：
 - ローカル エイジ = 900 秒
 - リモート エージ = 300 秒

- バウンス エージ = 630 秒
- アップグレードに関する考慮事項：リリース 1.0(1k) より前の ACI バージョンにアップグレードする場合は、テナント共通のエンドポイント保持ポリシー（epRetPol）のデフォルト値が次のようになっていることを確認してください: バウンス期間 = 660 秒。

IP エンドポイントの学習動作

ACI ブリッジ ドメインがユニキャスト ルーティングを有効にして構成されている場合、MAC アドレスを学習するだけでなく、MAC アドレスに関連付けられた IP アドレスも学習します。

ACI は MAC アドレスを追跡し、ブリッジ ドメインごとに一意である必要があります。ACI では、エンドポイントは単一の MAC アドレスに基づいていますが、任意の数の IP アドレスをブリッジ ドメインの単一の MAC アドレスに関連付けることができます。ACI は、これらの IP アドレスを MAC アドレスにリンクします。MAC アドレスが、IP アドレスのみを持つエンドポイントを表す場合があります。

したがって、ACI は次のようにローカルエンドポイントを学習して保存する場合があります。

- MAC アドレスのみ
- 単一の IP アドレスを持つ MAC アドレス
- 複数の IP アドレスを持つ MAC アドレス

3 番目のケースは、サーバーがプライマリおよびセカンダリ IP アドレスなど、同じ MAC アドレスに複数の IP アドレスを持っている場合に発生します。また、ACI ファブリックがファブリック上のサーバーの MAC アドレスと IP アドレスを学習したが、サーバーの IP アドレスがその後変更された場合にも発生する可能性があります。これが発生すると、ACI は MAC アドレスを保存し、古い IP アドレスと新しい IP アドレスの両方にリンクします。ACI ファブリックがエンドポイントをベース MAC アドレスでフラッシュするまで、古い IP アドレスは削除されません。

ACI でのローカルエンドポイントの移動には、主に 2 つのタイプがあります。

- MAC アドレスが別のインターフェイスに移動する場所
- IP アドレスが別の MAC アドレスに移動する場所

MAC アドレスが別のインターフェイスに移動すると、ブリッジドメインの MAC アドレスにリンクされているすべての IP アドレスも一緒に移動します。ACI ファブリックは、IP アドレスのみが移動した場合（および新しい MAC アドレスを受信した場合）も移動を追跡します。これは、たとえば、仮想サーバーの MAC アドレスが変更され、新しい ESXI サーバー（ポート）に移動された場合に発生する可能性があります。

VRF 内の複数の MAC アドレスに IP アドレスが存在する場合、これは IP フラップが発生したことを示しています（これは、ファブリック転送の決定に悪影響を与える可能性があります）。

これは、レガシー ネットワークの2つの個別のインターフェイスでの MAC フラッピング、またはブリッジドメインでの MAC フラップに似ています。

IP フラップが発生する可能性のあるシナリオの1つは、サーバーのネットワーク情報カード (NIC) ペアがアクティブ/アクティブに設定されているが、その2つが単一の論理リンク (ポートチャネルや仮想ポートチャネルなど) で接続されていない場合です。このタイプのセットアップにより、単一の IP アドレス (仮想マシンの IP アドレスなど) が、ファブリック内の2つの MAC アドレス間を常に移動する可能性があります。

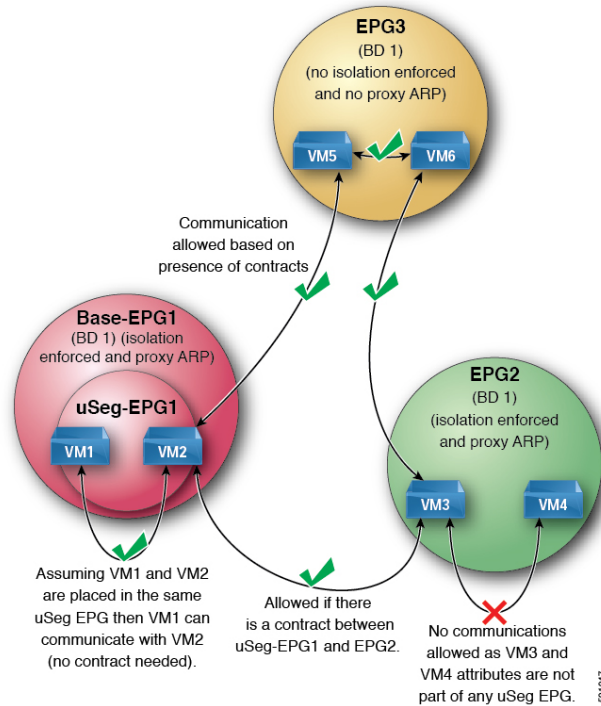
この種の動作に対処するには、NIC ペアを VPC の2つのレッグとして構成して、アクティブ/アクティブセットアップを実現することをお勧めします。サーバーハードウェアがアクティブ/アクティブ構成 (ブレードシャーシなど) をサポートしていない場合、アクティブ/スタンバイタイプの NIC ペア構成も IP フラッピングの発生を防ぎます。

プロキシ ARP について

Cisco ACI のプロキシ ARP は、ネットワークまたはサブネット内のエンドポイントが、別のエンドポイントの MAC アドレスを知らなくても、そのエンドポイントと通信できるようにします。プロキシ ARP はトラフィックの宛先場所を知っており、代わりに、最終的な宛先として自身の MAC アドレスを提供します。

プロキシ ARP を有効にするには、EPG 内エンドポイント分離を EPG で有効にする必要があります。詳細については、次の図を参照してください。EPG 内エンドポイント分離と Cisco ACI の詳細については、「Cisco ACI 仮想化ガイド」を参照してください。

図 11: プロキシ ARP および Cisco APIC



Cisco ACI ファブリック内のプロキシ ARP は従来のプロキシ ARP とは異なります。通信プロセスの例として、プロキシ ARP が EPG で有効になっているとき、エンドポイント A が ARP 要求をエンドポイント B に送信し、エンドポイント B がファブリック内で学習される場合、エンドポイント A はブリッジドメイン (BD) MAC からプロキシ ARP 応答を受信します。エンドポイント A が B、エンドポイントの ARP 要求を送信し、エンドポイント B はすでに ACI ファブリック内で学習しない場合は、ファブリックはプロキシ ARP の BD 内で要求を送信します。エンドポイント B は、ファブリックに戻る要求、このプロキシ ARP に応答します。この時点では、ファブリックはプロキシ ARP エンドポイント A への応答を送信しませんが、エンドポイント B は、ファブリック内で学習します。エンドポイント A は、エンドポイント B に別の ARP 要求を送信する場合、ファブリックはプロキシ ARP 応答から送信 BD mac です。次の例ではプロキシ ARP 解像度がクライアント VM1 と VM2 間の通信の手順します。

1. VM2 通信を VM1 が必要です。

図 12: VM2 通信を VM1 が必要です。

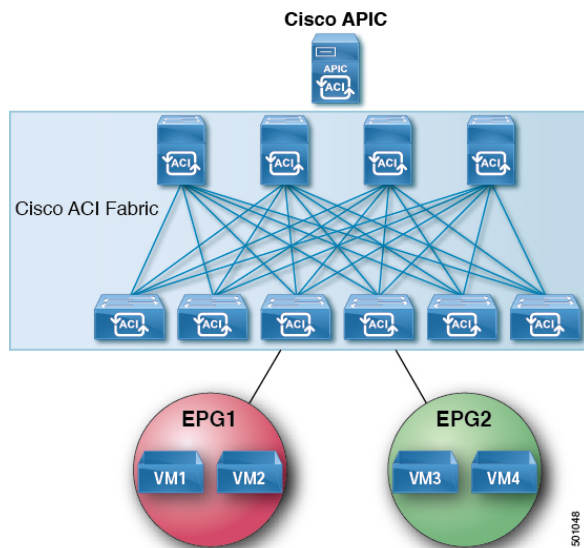


表 2: ARP 表の説明

デバイス	状態
VM1	IP = * MAC = *
ACI ファブリック	IP = * MAC = *
VM2	IP = * MAC = *

- VM1 は、ブロードキャスト MAC アドレスとともに ARP 要求を VM2 に送信します。

図 13: VM1 はブロードキャスト MAC アドレスとともに ARP 要求を VM2 に送信します

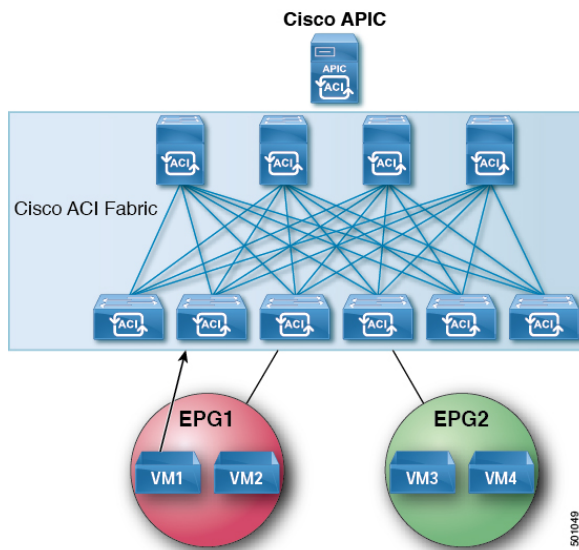


表 3: ARP 表の説明

デバイス	状態
VM1	IP = VM2 IP; MAC = ?
ACI ファブリック	IP = VM1 IP; MAC = VM1 MAC
VM2	IP = * MAC = *

- ACI ファブリックは、ブリッジドメイン (BD) 内のプロキシ ARP 要求をフラッディングします。

図 14: ACI ファブリックは BD 内のプロキシ ARP 要求をフラッディングします

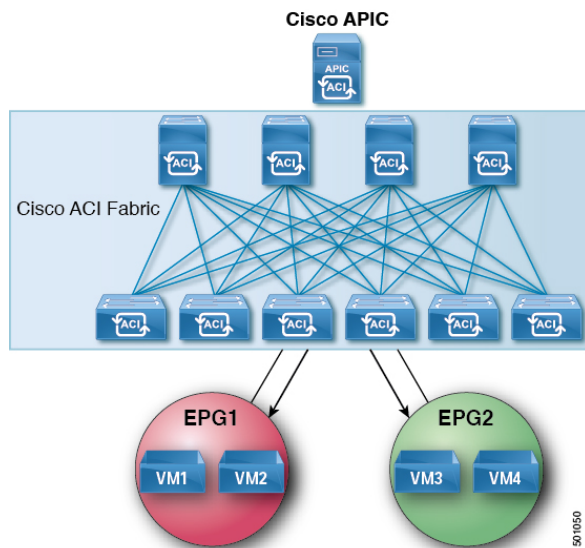


表 4: ARP 表の説明

デバイス	状態
VM1	IP = VM2 IP; MAC = ?
ACI ファブリック	IP = VM1 IP; MAC = VM1 MAC
VM2	IP = VM1 IP; MAC = BD MAC

4. VM2 は、ARP 応答を ACI ファブリックに送信します。

図 15: VM2 は ARP 応答を ACI ファブリックに送信します

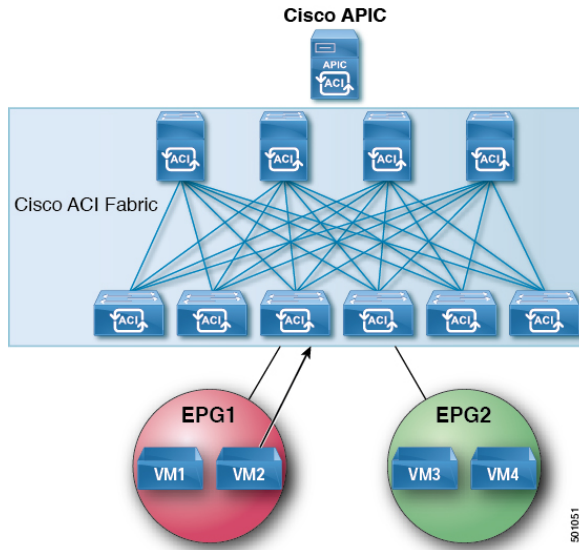


表 5: ARP 表の説明

デバイス	状態
VM1	IP = VM2 IP; MAC = ?
ACI ファブリック	IP = VM1 IP; MAC = VM1 MAC
VM2	IP = VM1 IP; MAC = BD MAC

5. VM2 が学習されます。

図 16: VM2 が学習されます

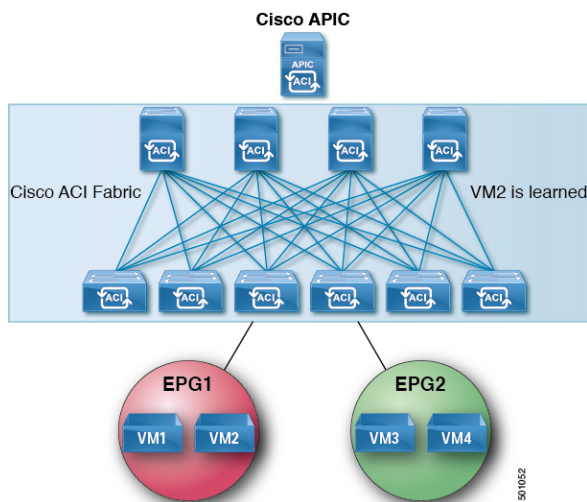


表 6: ARP 表の説明

デバイス	状態
VM1	IP = VM2 IP; MAC = ?
ACI ファブリック	IP = VM1 IP; MAC = VM1 MAC IP = VM2 IP; MAC = VM2 MAC
VM2	IP = VM1 IP; MAC = BD MAC

6. VM1 は、ブロードキャスト MAC アドレスとともに ARP 要求を VM2 に送信します。

図 17: VM1 はブロードキャスト MAC アドレスとともに ARP 要求を VM2 に送信します

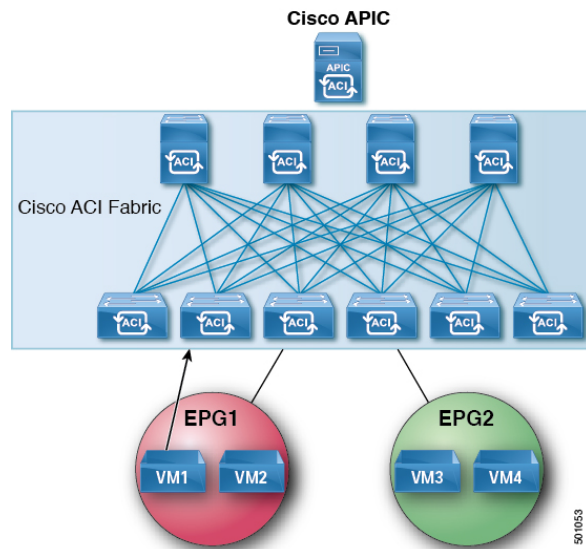


表 7: ARP 表の説明

デバイス	状態
VM1	IP = VM2 IP; MAC = ?
ACI ファブリック	IP = VM1 IP; MAC = VM1 MAC IP = VM2 IP; MAC = VM2 MAC
VM2	IP = VM1 IP; MAC = BD MAC

7. ACI ファブリックは、プロキシ ARP VM1 への応答を送信します。

図 18: ACI ファブリック VM1 にプロキシ ARP 応答を送信します。

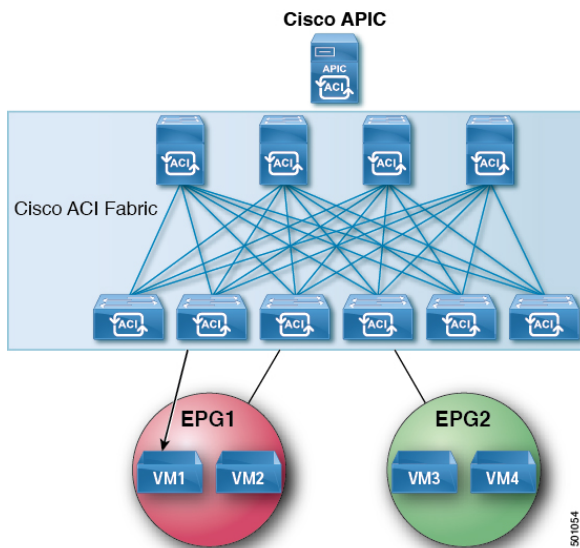


表 8: ARP 表の説明

デバイス	状態
VM1	IP = VM2 IP; MAC = BD MAC
ACI ファブリック	IP = VM1 IP; MAC = VM1 MAC IP = VM2 IP; MAC = VM2 MAC
VM2	IP = VM1 IP; MAC = BD MAC

ループ検出

Cisco Application Centric Infrastructure (ACI) ファブリックは、Cisco ACI アクセスポートに接続されているレイヤ2ネットワークセグメントのループを検出できるグローバルなデフォルトループ検出ポリシーを提供します。これらのグローバルポリシーはデフォルトで無効になっていますが、ポートレベルのポリシーはデフォルトで有効になっています。グローバルポリシーを有効にすると、個々のポートレベルで無効にされていない限り、すべてのアクセスポート、仮想ポート、および仮想ポートチャネル (VPC) でポリシーが有効になります。

Cisco ACI ファブリックは、スパニングツリープロトコル (STP) に参加していません。代わりに、ループを検出するために、ミスケールブルプロトコル (MCP) を実装します。MCP は、外部レイヤ2ネットワークで実行されている STP と補完的に機能します。



- (注) スパニングツリーを実行し、Cisco ACI ファブリックに接続されている外部スイッチからのインターフェイスは、`loop_inc` ステータスになる可能性があります。外部スイッチからのポートチャンネルをフラッピングすると、問題が解決します。外部スイッチでBDPUフィルタを有効にするか、ループガードを無効にすると、問題を回避できます。

ファブリック管理者は、Cisco ACI ファブリックによって開始されたMCPパケットを識別するためにMCPが使用するキーを提供します。管理者は、MCPポリシーがループを識別する方法と、ループに対処する方法（syslogのみ、またはポートを無効にする）を選択できます。

VMの移動などのエンドポイントの移動は正常ですが、頻度が高く、移動の間隔が短い場合は、ループの兆候である可能性があります。個別のグローバルなデフォルトエンドポイント移動ループ検出ポリシーを使用できますが、デフォルトでは無効になっています。管理者は、移動検出ループに対処する方法を選択できます。

また、エラー無効化の回復ポリシーは、管理者が構成できる間隔の後に、検出をループするポートを有効にし、BPDUポリシーを無効にすることができます。

MCPはネイティブVLANモードで実行され、デフォルトでは、送信されるMCP BPDUにVLANタグが付けられません。MCPは、ネイティブVLANで送信されたパケットがファブリックによって受信された場合、ケーブル接続の誤りによるループを検出できますが、EPGVLANの非ネイティブVLANにループがある場合は検出されません。リリース2.0(2)以降、Cisco Application Policy Infrastructure Controller (APIC)は構成されたEPG内のすべてのVLANでMCP BPDUの送信をサポートしているため、それらのVLAN内のループが検出されます。新しいMCP構成モードでは、送信されるPDUに各EPG VLAN IDを持つ802.1Qヘッダーを追加することにより、物理ポートが属するすべてのEPG VLANでMCP PDUが送信されるモードで動作するようにMCPを構成できます。

3.2(1) リリース以降、Cisco ACI ファブリックは100ミリ秒から300秒の送信頻度でより高速なループ検出を提供します。

5.2(3) リリース以降では、いずれかのインターフェイスで256を超えるVLANを動作させようとすると、フォールトF4268が生成されます。いずれかのリーフスイッチで2000を超える論理ポート（ポート x VLAN）動作させようとすると、フォールトF4269が生成されます。



- (注) VLANごとのMCPは、インターフェイスごとに256のVLANでのみ実行されます。256を超えるVLANがある場合、最初の数値の256 VLANが選択されます。

MCPは、Fabrix Extender (FEX) ホストインターフェイス (HIF) ポートではサポートされていません。

Mis-cabling プロトコルのモード

MCPは2つのモードで動作できます。

- 非厳格モード：MCP対応ポートがUPの場合、データトラフィックとコントロールプレーントラフィック（STP、MCPプロトコルパケットなど）が受け入れられます。MCPはリンクのループを監視し、ループが検出されると、リンクはエラー ディセーブルになります。このモードでは、グローバルMCPインスタンスポリシーに従って、パケットが2秒間隔で送信されます。このモードでのループ検出のデフォルト時間は7秒です。
- 厳格モード：リリース 5.2(4)以降、MCPは厳格モードをサポートします。MCPが有効になっているポートが起動するとすぐに、ループをチェックするためにMCPパケットが短期間、アグレッシブな間隔で送信されます。これは早期ループ検出フェーズと呼ばれ、データトラフィックを受け入れる前に、リンク（ポートに接続されている）にループがないかどうかチェックされます。ループが検出されると、ポートはエラーディセーブルになり、シャットダウンされます。ループが検出されない場合、ポートはデータトラフィックの転送を開始します。MCPは、グローバルMCPインスタンスポリシーに従って、非アグレッシブタイマーを使用してパケットの送信を開始します。

すでにUP状態のポートでMCP厳格モードが構成されている場合、MCPはこのポートで早期ループ検出を実行しません。MCP厳格モード構成のポートをフラップして、すぐに有効にします。

MCP厳格モードのイベントシーケンス：

1. MCP対応ポートが起動すると、初期遅延タイマーが開始されます。リンクレベルの制御パケット（LLDP、CDP、STPなど）のみが受け入れられ、転送されます。この期間中、データトラフィックは受け入れられません。初期遅延タイマーは、外部L2ネットワークでSTPがコンバージするための時間です。デフォルト値は0ですが、トポロジと外部ネットワークでのスパニングツリーの構成方法によっては、STPが収束してループを切断するまでの時間を確保するために、初期遅延を45～60秒に設定することもできます（必要な場合）。外部L2ネットワークでSTPが有効になっていない場合は、初期遅延タイマーを0に設定する必要があります。
2. 猶予期間タイマーは、初期遅延タイマーが期限切れになった後に開始されます。この間ポートは、ループ検出に使用されるMCPパケットをアグレッシブに送信します。この間に、早期のループ検出が行われます。ループが検出されると、ポートはエラーディセーブルになります。デフォルト値は3秒です。この期間中、データトラフィックは受け入れられません。

猶予タイマー期間中、MCPは次のグローバルMCPインスタンスポリシー構成を上書きします。

- ループが検出されると、GUIで[ポート無効化 (Port Disable)] チェックボックスが選択されていなくても、MCPはポートをエラー ディセーブルにします。
 - MCPの倍率が1より大きい値に設定されている場合でも、単一のMCPフレームを受信するとループと見なされます。
3. 猶予期間タイマーが期限切れになり、ループが検出されなくなると、ポートは転送ステートに移行し、データトラフィックが受け入れられます。MCPパケットは、グローバルな送信頻度構成に従って、非アグレッシブな間隔で送信されます。

MCP 厳格モードのガイドラインおよび制約事項

次のガイドラインおよび制約事項に従って、厳格モード MCP を構成します。

- MCP 厳格モードは FEX ポートではサポートされません。
- MCP 厳格モードは QinQ エッジ ポートではサポートされません。
- ポートで MCP 厳格モードが有効になっている場合、vPC 高速コンバージェンスはサポートされません。vPC トラフィックは、収束に時間がかかります。
- 厳格モードの MCP 制御パケットは、APIC リリース 5.2(4) より前のバージョンを実行しているリーフスイッチではデコードできません。したがって、厳格ループ検出を機能させるには、MCP に参加するすべてのリーフスイッチに 5.2(4) の最小バージョンが必要です。
- APIC バージョン 5.2(4) より前のリリースにファブリックをダウングレードする前に、MCP 対応ポートで厳格モードを無効にします。厳格モードが無効になっていない場合、厳格ループ検出は以前のバージョンのスイッチでは機能しません。
- リーフスイッチのポートで厳格モードが有効になっている場合、そのポートで特定の VLAN が有効になっていない場合でも、STP BPDU は受け入れられます。リモート側で設定不備があると、外部 L2 スイッチで意図しない STP 状態が発生する可能性があります。
- 厳格モードが有効になっているか、送信頻度が 2 秒未満の場合は、MCP CoPP バーストと CoPP レートを 5000 に設定します。
- 2 つの MCP 厳格対応ポートが同時に起動すると、どちらかの側でループが検出されたときに両方のポートがエラー無効になる可能性があります。

不正なエンドポイントの検出

不正なエンドポイントの制御ポリシーについて

不正なエンドポイントは、リーフスイッチを頻繁に攻撃し、異なるリーフスイッチポートにパケットを繰り返し挿入し、802.1Q タグを変更する（エンドポイントの移動をエミュレートする）ことで、学習されたクラスと EPG ポートを変更します。誤設定により頻繁に IP アドレスと MAC アドレスが変更（移動する）されることとなります。

ファブリックの急速な移動などで、大きなネットワークの不安定状態、高い CPU 使用率、まれなケースでは、大量かつ長期のメッセージおよびトランザクションサービス (MTS) バッファ消費のため、エンドポイント マッパー (EPM) および EPM クライアント (EPMC) がクラッシュすることとなります。また、このような頻繁な移動により、EPM および EPMC ログが非常にすばやくロールオーバーされ、無関係なエンドポイントのデバッグを妨害する可能性があります。

不正なエンドポイントの制御機能は脆弱性にすばやく対処します。

- 急速に移動する MAC および IP エンドポイントの特定。

- エンドポイントを一時的に静的にして、エンドポイントを隔離することによって移動を停止します。
- 3.2(6) リリースより前：**不正 EP 検出間隔**のエンドポイントを静的に維持し、不正エンドポイントとの間のトラフィックをドロップします。この時間が経過すると、不正な MAC アドレスまたは IP アドレスが削除されます。
- 3.2(6) リリース以降：**不正な EP 検出間隔**のエンドポイントを静的に維持（この機能はトラフィックをドロップしなくなりました）。この時間が経過すると、不正な MAC アドレスまたは IP アドレスが削除されます。
- ホスト トラッキング パケットを生成して、影響を受ける MAC または IP アドレスをシステムが再学習できるようにします。
- 修正アクションを有効にするための障害の発生。

不正なエンドポイント制御ポリシーはグローバルに設定されており、他のループ防止方法とは異なり、個々のエンドポイント レベルの機能です (IP および MAC アドレス)。ローカルまたはリモートの移動を区別していません。いかなる種類のインターフェイスの変更も、エンドポイントを隔離する必要があるかどうかを決定する際に移動と見なされます。

不正なエンドポイント制御機能は、デフォルトで無効になっています。

不正/COOP 例外リストについて

不正/COOP 例外リストを使用すると、エンドポイントが不正としてマークされる前に、不正エンドポイント制御によるエンドポイント移動の許容度を高くするエンドポイントの MAC アドレスを指定できます。不正/COOP 例外リストのエンドポイントは、10 分以内に 3000 回以上移動した場合にのみ不正としてマークされます。エンドポイントが不正としてマークされた後、学習を防ぐためにエンドポイントは静的なままになります。不正エンドポイントは 30 秒後に削除されます。

Cisco Application Policy Infrastructure Controller (APIC) 6.0 (3) リリース以降では、グローバルな不正/COOP 例外リストを作成できます。このリストは、MAC アドレスが見つかるブリッジドメイン上の全ての不正エンドポイント コントロールから MAC アドレスを除外します。そして、L3Outs の不正/COOP 例外リストを作成できます。ブリッジドメインまたは L3Out のすべての MAC アドレスを除外することもできます。これにより、すべての MAC アドレスに対して例外を作成する場合に例外リストを簡単に作成できます。各アドレスを個別に入力する必要はありません。

TCP MSS調整について

Cisco Application Policy Infrastructure Controller (APIC) 6.0(3) リリース以降、伝送制御プロトコル (TCP) 最大セグメントサイズ (MSS) 調整機能を使用すると、リーフスイッチによって転送される一時パケットの最大セグメントサイズを構成できます。ファブリック VXLAN トンネルを介して送受信される一時的なパケットのみが調整されます。2つのデバイスが TCP セッションを確立すると、最初の SYN パケットで MSS をネゴシエートします。この MSS 値は、

パケットのフラグメンテーションやドロップを回避するために、基盤となるネットワークでサポートされている MTU よりも小さくする必要があります。Cisco ACI ファブリックは、9216 バイトのレイヤ 2 MTU サイズをサポートします。Cisco ACI ファブリックに接続されたデバイスは、最大 9104 バイトの MTU をネゴシエートできます。ただし、Cisco ACI マルチポッド、Cisco ACI マルチサイト、および Cisco ACI ファブリックのリモートリーフスイッチなどの機能には、サポートされる MTU が Cisco ACI ファブリックでサポートされるものよりも小さい外部ネットワークを介したトラフィックフローが含まれます。

TCP MSS 調整機能を使用すると、TCP SYN/SYN-ACK パケット内の TCP MSS 値を事前定義された値に調整できます。この調整では、TCP SYN/SYN-ACK パケットのみが CPU にパントされます。ユーザーは、IPv4 と IPv6 に個別の MSS 値を指定できます。MSS 調整は、リーフスイッチごとに 1 秒あたり最大 2000 の TCP 接続に対して実行できます。

TCP MSS 調整には、さまざまなトポロジに対応するための 4 つのノブがあります。

1. **[無効 (Disabled)]** : 機能は無効です。
2. **[RL のみ (RL Only)]** : リモートリーフスイッチとの間で送受信されるトラフィックが調整されます。バックツーバックリンクを使用する 2 つのリモートリーフスイッチ間のトラフィックは調整されません。
3. **[RL およびマルチサイト (RL and Multi-Site)]** : Cisco ACI マルチサイト内のサイト間で送信されるトラフィックが、リモートリーフスイッチトラフィックに加えて調整されます。
4. **[グローバル (Global)]** : 異なる TEP プール内のリーフスイッチとの間で送受信されるトラフィック。

例 :

- 異なるリモートリーフスイッチ TEP プール内のリモートリーフスイッチとの間で送受信されるトラフィック
- リモートリーフスイッチとローカルリーフスイッチとの間で送受信されるトラフィック
- Cisco ACI マルチポッドの異なるポッドのリーフスイッチとの間で送受信されるトラフィック
- Cisco ACI マルチサイトの異なるサイトのスイッチとの間で送受信されるトラフィック

同じリモートリーフスイッチ TEP プール内のリモートリーフスイッチとの間で送受信されるトラフィックは調整されません。

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。