Campus 3.0 Virtual Switching System デザイン ガイド

Campus 3.0 Virtual Switching System Design Guide

Cisco Validated Design

2009 年 9 月 14 日

Text Part Number: OL-19829-01-J

【注意】シスコ製品をご使用になる前に、安全上の注意（www.cisco.com/jp/go/safety_warning/）をご確認ください。

本書は、米国シスコシステムズ発行ドキュメントの参考和訳です。米国サイト掲載ドキュメントとの差異が生じる場合があるため、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。また、契約等の記述については、弊社販売パートナー、または、弊社担当者にご確認ください。

Cisco Validated Design

Cisco Validated Design プログラムは、お客様による信頼性の高い、確実かつ速やかな展開を容易にするために、デザイン、テスト、および文書化されたシステムおよびソリューションで構成されています。詳細については、www.cisco.com/go/validateddesigns（英語）、www.cisco.com/jp/go/validateddesigns（日本語）にアクセスしてください。

このマニュアルに記載されているデザイン、仕様、表現、情報、および推奨事項（総称して「デザイン」）は、障害も含めて本マニュアル作成時点のものです。シスコシステムズおよびそのサプライヤは、商品性の保証、特定目的への準拠の保証、および権利を侵害しないことに関する保証、あるいは取引過程、使用、取引慣行によって発生する保証をはじめとする、一切の保証の責任を負わないものとします。いかなる場合においても、シスコシステムズおよびそのサプライヤは、このデザインの使用または使用できないことによって発生する利益の損失やデータの損傷をはじめとする、間接的、派生的、偶発的、あるいは特殊な損害について、あらゆる可能性がシスコシステムズまたはそのサプライヤに知らされていても、それらに対する責任を一切負わないものとします。

デザインは予告なしに変更されることがあります。このマニュアルに記載されているデザインの使用は、すべてユーザ側の責任になります。これらのデザインは、シスコシステムズ、そのサプライヤ、パートナーの技術的な助言や他の専門的な助言に相当するものではありません。ユーザは、デザインを実装する前に技術アドバイザーに相談してください。シスコによるテストの対象外となった要因によって、結果が異なることがあります。

Campus 3.0 Virtual Switching System デザイン ガイド

CONTENTS

はじめに

このマニュアルについて

対象読者

このデザイン ガイドは、Virtual Switching Systems 1440 を使用してキャンパス ネットワークのデザインを担当するシスコシステムズのエンジニアまたは企業のエンジニアを対象としています。

このマニュアルの目的

このマニュアルは、Cisco Catalyst 6500 シリーズ Virtual Switching System（VSS）1440 を階層型キャンパス アーキテクチャ内に実装するためのデザイン ガイダンスを提供します。第 1 章「Virtual Switching Systems のデザインの概要」は、従来型のキャンパスのデザイン手法とアーキテクチャの範囲について説明します。第 2 章「Virtual Switching System 1440 アーキテクチャ」は、VSS の重要なコンポーネントを紹介し、キャンパスに VSS を設定する場合に固有のベスト プラクティス デザイン オプションと推奨事項を示します。第 3 章「VSS 対応キャンパス デザイン」では、キャンパスにおける VSS のアプリケーションについて議論し、トラフィック フローとキャンパス特有のベスト プラクティス推奨デザインについて説明します。第 4 章「コンバージェンス」は、検証されたデザイン（Validated Design）の環境について説明します。VSS に対応したエンドツーエンドのキャンパスのコンバージェンス特性についても説明します。

（注） このマニュアルでは、今後、Cisco Catalyst 6500 シリーズ VSS を VSS と称します。

このデザイン ガイドは、蓄積されたベスト プラクティスのナレッジを文書化した各種の資料を参照および使用しました。その一覧を付録 B「参考資料」に示しています。ただし、このデザイン ガイドの執筆期間中に、多数のデザイン オプションが、VSS 固有の展開で推奨される特定のオプションに置き換えられたり更新されたりしています。このデザイン ガイドでは、シスコのベスト プラクティスを例示して説明するか、再確認するにとどめる説明のいずれかまたは両方を行っています。

マニュアルの構成

このデザイン ガイドは、次の各章と付録で構成されています。

章	説明
この章	このキャンパス 3.0 Virtual Switching Systems（VSS）ソリューションのマニュアルに記載される内容の簡潔な説明
第 1 章「Virtual Switching Systems のデザインの概要」	このマニュアルで説明する VSS デザインの概要
第 2 章「Virtual Switching System 1440 アーキテクチャ」	Cisco Catalyst 6500 シリーズ VSS 1440 のアーキテクチャとコンポーネントを中心とした説明
第 3 章「VSS 対応キャンパス デザイン」	VSS キャンパス デザインにおける、次の 3 つの主要な部分に関する説明 • EtherChannel の最適化、トラフィック フロー、VSL キャパシティ プランニング • マルチレイヤ デザインのベスト プラクティス • VSS を使ったルーティング
第 4 章「コンバージェンス」	VSS を使用するエンドツーエンド キャンパスに対応するネットワークのコンバージェンス特性に関する説明。
付録 A「VSS 対応キャンパスのベスト プラクティス設定例」	VSS 対応キャンパスのベスト プラクティス設定の例
付録 B「参考資料」	参考資料と関連マニュアルへのリンクを掲載

著者について

シスコシステムズ、CMO Enterprise Systems Engineering（ESE）、テクニカル リーダー、Nimish Desai

Nimish は現在、ESE 内データセンター アプリケーション グループのテクニカル リーダーとして活躍しています。ESE では、シスコ キャンパス ネットワーク用ベスト プラクティス デザインにおける Virtual Switching System ソリューションの開発と検証に携わるリード アーキテクトでした。ESE キャンパス ソリューション チームに参加する前は、Nimish は Cisco Advanced Services 部門で、デザイン コンサルテーションとテクニカル エスカレーション（サポート）を大企業のお客様に提供する業務を担当していました。

Nimish はインターネットワーキング テクノロジー業務にこの 17 年間従事しています。シスコに入社する以前、Nimish は、大手金融機関での取引場のサポート業務、物流および保険業界での大規模なエンタープライズ ネットワーク デザイン業務、IBM 社での製品開発の経験などを通じて専門的な実績を積んできました。Nimish は米国ニュージャージー工科大学の電子工学専門科学修士号（MSEE）を取得しています。趣味は釣りとアウトドアで、キャンピングカーでの国立公園めぐりも楽しんでいます。

CHAPTER 1

Virtual Switching Systems のデザインの概要

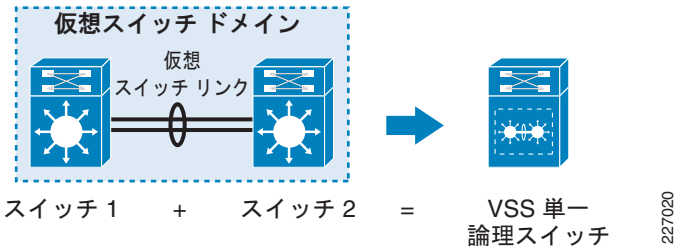
要約

ネットワークとシステムの冗長性を単一ノードに統合する VSS は、キャンパス ネットワークの機能性とアベイラビリティを無類のレベルに高めることができます。VSS の機能に対応したエンドツーエンドのキャンパス ネットワークは、このデザイン ガイドで説明する柔軟性とアベイラビリティが実現します。

この単一の論理ノードは、キャンパス ネットワークにおけるサービス統合を、大きく妥協することなくこれまで不可能だった範囲に拡張します。ワイヤレス、ファイアウォール サービス モジュール（FWSM）、侵入防御システム（IPS）、その他のサービス ブレードを VSS 内に統合することで、サービスの即時提供が可能なキャンパスをデザインする一連の機能が採用できます。たとえば、VSS を実装することで、先進のインターネット デザイン（対称フォワーディング）、データセンターの相互接続（ループなし障害リカバリ）、その他の多くの適用が可能になります。このマニュアルでは、キャンパスへの VSS の適用をディストリビューション レイヤに限定して説明していますが、ここで解説した原則を応用した新たな適用を創出することはネットワークのデザイン担当者に任されています。また、VSS はキャンパス環境での利用に限定されるものでもありません。

基本となる VSS の主な機能は、2 つの物理シャーシを単一の論理エンティティにクラスタ化できる機能です。図 1-1 を参照してください。

図 1-1　　VSS の概念図



この 2 つの物理シャーシを単一の論理スイッチに*仮想化*する技術は、キャンパス トポロジのデザインを根底から一新します。VSS で実現する最も大きい変化の一つが*ループのない*トポロジです。これに加えて、VSS には Stateful Switch Over（SSO; ステートフル スイッチオーバー）や Multi-chassis EtherChannel（MEC）といったシスコ独自の革新的機能が数多く組み込まれているため、帯域幅が拡張されて無停止通信が可能になり、アプリケーションの応答時間が大幅に短縮します。VSS の主なビジネス上の利点は次のとおりです。

- ループ型トポロジに関連したリスクの削減

- SSO 対応型スーパーバイザを持つ冗長シャーシの活用を通じた無停止によるビジネス通信
- 拡張された帯域幅で形成されるアクセス レイヤを利用した、既存の投資に対する回収率の改善
- ネットワークの仮想化、Network Admission Control（NAC）、キャンパス ネットワークでのファイアウォールやワイヤレス サービスといった新しいサービスを、単一の論理ノードで展開、管理する柔軟性が向上したことによる、運用経費（OPEX）の削減
- 構成に関するエラーの削減と、Hot Standby Routing Protocol（HSRP; ホット スタンバイ ルーティング プロトコル）、GLBP、および VRRP などの First Hop Redundancy Protocols（FHRP; ファースト ホップ冗長プロトコル）の排除
- 単一構成による管理の簡素化と、運用上の障害点の減少

加えて、VSS のサービス モジュールを統合する機能により、シスコのキャンパス ファブリックをサービス指向キャンパス アーキテクチャの中心として適切に認識できます。

Virtual Switching System（VSS）のデザイン

キャンパス ネットワークへの VSS の適用に関する理解を深めるには、既存のシスコのアーキテクチャと代替デザインを順守することが重要です。次のセクションで、シスコのキャンパス デザイン オプションの範囲とフレームワークを示し、また、ハイ アベイラビリティ、スケーラビリティ、復元力、柔軟性に関する課題がこれによってどのように解決されるかを説明します。一部のデザイン モデルに特有の非効率性についても説明します。

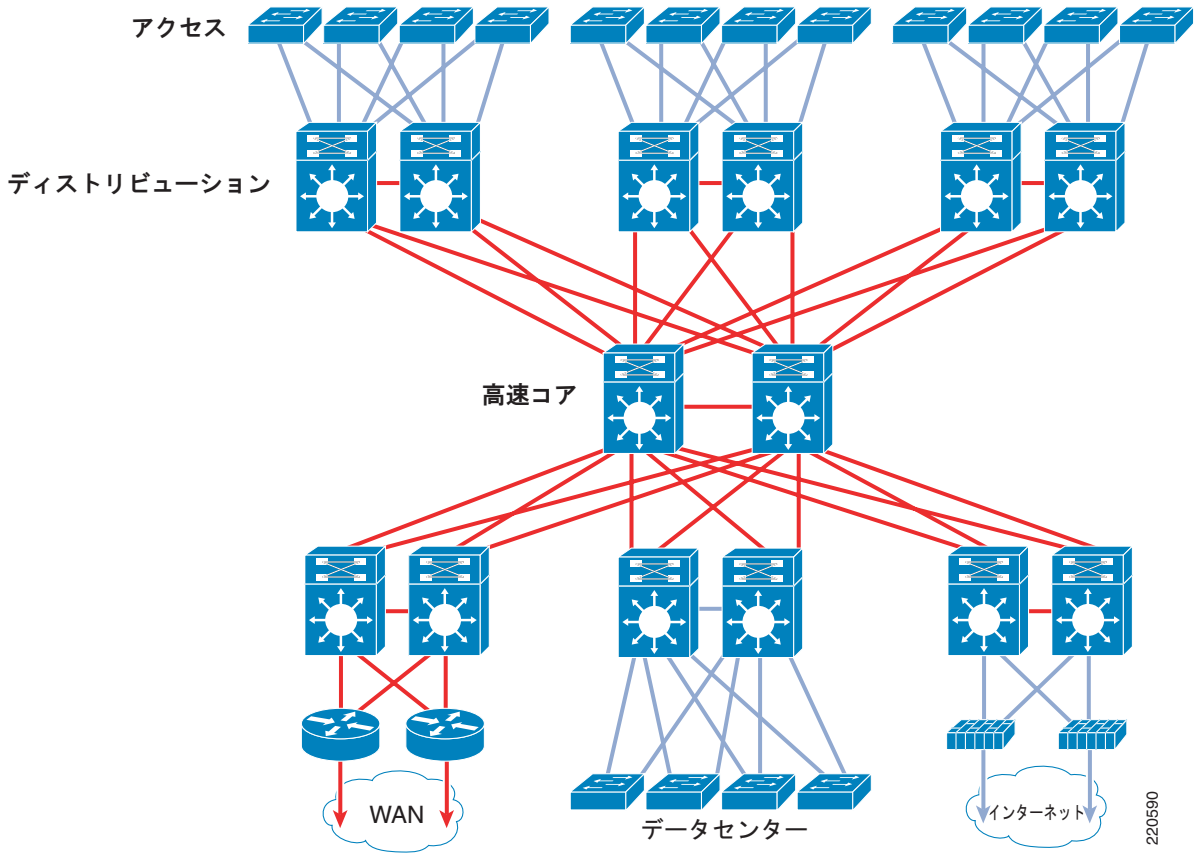
キャンパス アーキテクチャとデザイン

キャンパス アーキテクチャのデザイン プロセスには、新たなビジネス要件という課題があります。無停止通信のニーズはほとんどのキャンパス ネットワークで基本的な出発点になりつつあります。現代のキャンパス デザインに影響を及ぼすビジネス ケースや要因については、デザイン上のフレームワークに関する次のマニュアルで解説しています。

『*Enterprise Campus 3.0 Architecture: Overview and Framework*』
http://www.cisco.com/en/US/docs/solutions/Enterprise/Campus/campover.html

階層型デザインの原則を利用することにより、これらの要件を満たすキャンパス ネットワークを実装する基盤が提供されます。階層型デザインで採用されているビルディング ブロック手法では、複数の独立したディストリビューション ブロックが接続する高速ルーテッド コア ネットワーク レイヤを使用します。ディストリビューション ブロックは、ビルディング、フロア、およびセクション用のアグレゲータとして機能する実際のディストリビューション ノードと、ワイヤリング クローゼット アクセス スイッチという 2 種類のスイッチ レイヤで構成されています。図 1-2 を参照してください。

図 1-2　階層キャンパス ビルディング ブロック



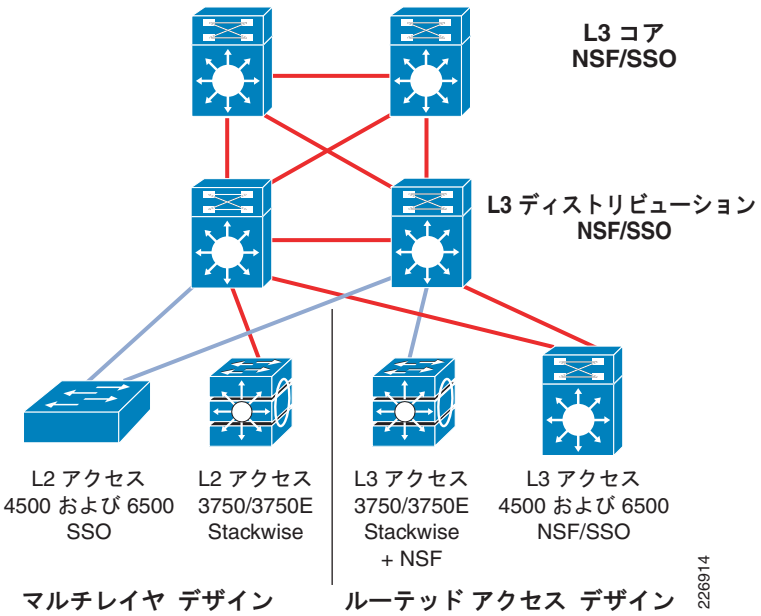
ディストリビューション ブロックの VSS

キャンパス 3.0 デザインのフレームワークには、ネットワークにおける階層の機能的利用が含まれます。ここでは、ディストリビューション ブロック アーキテクチャ（アクセスディストリビューション ブロックとも呼ばれる）がキャンパスのデザインの中心と機能性の重要な部分を管理しています。アクセスディストリビューション ブロックは、マルチティア キャンパス アーキテクチャ内の 3 つの階層ティアのうち、アクセス レイヤとディストリビューション レイヤの 2 つで構成されています。これら 2 つのレイヤはそれぞれ特定のサービス要件と機能要件がありますが、ディストリビューション ブロック同士の結合の方法や、これらをアーキテクチャ全体にどのように適合させるかを決定する際に要となるのが、ネットワーク トポロジ コントロール プレーン デザインの選択です（ルーティング プロトコルやスパニング ツリー プロトコルの選択）。アクセスディストリビューション ブロックと関連するコントロール プレーンの設定方法には、次の 2 種類の基本デザイン オプションがあります。

- マルチレイヤまたはマルチティア（アクセス ブロックのレイヤ 2）
- ルーテッド アクセス（アクセス ブロックのレイヤ 3）

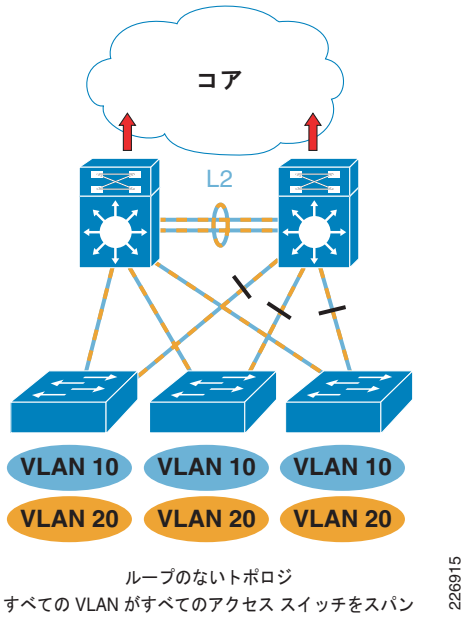
これらのデザインは同じ基本物理トポロジとケーブリング設備を使用しますが相違点もあります。レイヤ 2 とレイヤ 3 の境界が存在すること、ネットワーク トポロジ冗長性の実装方法、ロード バランシングを機能させる方法などです。また、各デザイン オプション間にも数々の主要な相違点が存在します。図 1-3 に、既存のデザインで利用可能な選択肢を示します。

図 1-3　従来のデザインにおける選択肢



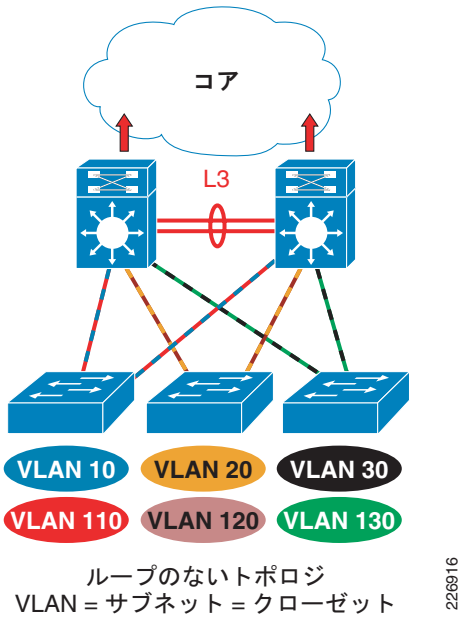
マルチレイヤ デザインは、お客様のネットワークで最古の、最も普及しているデザインです。一方のルーテッド アクセスは、比較的新しいデザインです。最も一般的なマルチレイヤ デザインは、レイヤ 2 の隣接関係が要求される（ルーティング不可能なプロトコルのブリッジ）アプリケーションへ柔軟性を与えるための複数のアクセス レイヤ スイッチにまたがった vlan と IPX や IP のような主要なプロトコルのルーティングにより構成されます。このデザイン方式には、不安定さ、非効率的なリソース利用、応答時間の遅さ、困難なエンド ホストの動作管理など、さまざまな問題があります。図 1-4 を参照してください。

図 1-4　マルチレイヤ デザイン：ループ型トポロジ



マルチレイヤ デザインの第 2 のタイプでは、VLAN は複数のクローゼットに渡りません。言い換えると、VLAN = サブネット = クローゼットという関係が成り立ちます。このデザインは、ベスト プラクティスのマルチレイヤ デザインの基本です。ここでは、VLAN をクローゼットに閉じ込めてスパニング ツリー ループの発生を防いでいます。図 1-5 を参照してください。ただし、このデザインは VLAN のスパニングを許可しません。その間接的結果として、ほとんどのレガシー ネットワークがループ型スパニング ツリー プロトコル（STP; スパニング ツリー プロトコル）ベースのトポロジを持ち続けることになります。ただし、テクノロジーにより強制的に別のネットワークトポロジが採用するか、Voice over IP（VoIP）の実装などの、より高い安定性が要求されるビジネス イベントがある場合を除きます。

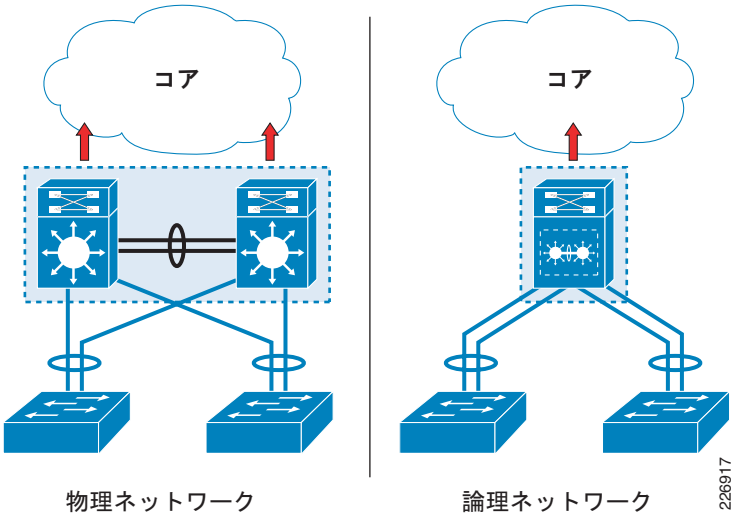
図 1-5　　マルチレイヤ デザイン：ループのないトポロジ



ループのないトポロジ
VLAN = サブネット = クローゼット

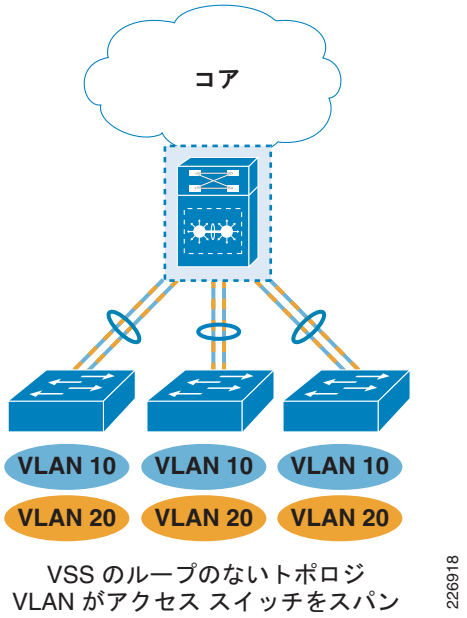
VSS をマルチレイヤ デザインのディストリビューション ブロックで使用すると、複数のクローゼットに VLAN を渡す機能がループを発生させずに実現します。図 1-6 に、VSS ペアとの物理接続と論理接続を示します。

図 1-6　ディストリビューション レイヤでの仮想スイッチ



ディストリビューション ブロックに VSS を導入すると、図 1-7 に示すように、両方のマルチレイヤ デザインが 1 つのデザイン オプションに変わります。ここでアクセス レイヤは単一の論理接続を通じて単一の論理ボックスに接続します。これはループのないトポロジの複数のクローゼットに VLAN を渡すことができるというかつてないオプションを可能にするトポロジです。

図 1-7　VSS を有効にしたループのないトポロジ



VSS の適用方法は多岐にわたります。VSS の適用は、コア、ディストリビューション、アクセスという階層型キャンパスの 3 種類すべてのティアと、マルチレイヤおよびルーテッドアクセスの両デザインのサービス ブロックで可能です。ただし、このデザイン ガイドでは、マルチレイヤ デザインにおけるディストリビューション レイヤへの VSS の適用に範囲を限定して説明します。また、この機能で行う、コアとの対話についても取り挙げます。多くのデザイン上の選択や観測が、ルーテッドアクセス デザインにおける VSS の使用時に適用可能です。というのは、これがレイヤ 3 のエンドツーエンド デザインであるためですが、VSS はマルチレイヤに導入した場合に最も効果が大きくなります。なぜなら、この場合にコントロール プレーンの簡素化とハイ アベイラビリティを備えたループのないトポロジが実現するからです。

VSS のアプリケーション

マルチレイヤ デザインにおける VSS のアプリケーションは、レイヤ 2 の隣接関係が必要なすべての場合に実施できます。これは、アプリケーションという観点だけでなく、柔軟性やネットワーク リソースの実践的な活用の点からも可能です。次のような適用事例があります。

- レイヤ 2 の隣接関係が必要なアプリケーション：複数のアクセス レイヤ スイッチに渡るデータ VLAN
- ビルディングまたはロケーションごとに VLAN を渡すことによるユーザの接続性の簡素化
- ネットワークの仮想化（一時的な接続をサポートするゲスト VLAN、社内接続、企業合併など）
- 複数の施設に渡る VLAN を使用した会議、メディア ルーム、公共アクセス
- Network Admission Control（NAC）VLAN（隔離、パスチャライズ、およびパッチ適用）
- またがった VLAN を必要とする外注グループ、およびエージェンシー間リソース
- 集中管理コントローラのないワイヤレス VLAN
- ネットワーク管理およびモニタリング（SNMP、SPAN）

Virtual Switching Systems（VSS）推奨のベスト プラクティス：概要

このデザイン ガイド全体を通じて、シスコでは推奨されるベスト プラクティスを提供しています。重要なものについては、該当するトピックの各セクションに「メモ」という目印が付いています。次の表に、シスコが推奨する重要なすべてのベスト プラクティスをわかりやすいように一覧します。

推奨される VSS のベスト プラクティス	トピック
複数の VSS ドメインを接続していない場合でも、ベスト プラクティスとして一意のドメイン ID を使用することを推奨します。	詳細については、「仮想ドメイン」を参照してください。
シスコではデフォルトの LMP（VSLP）タイマーを変更しないことを強くお勧めしています。	詳細については、「タイマーを変更すべきでない理由」を参照してください。
VSL リンクの負荷分散ハッシュ方式をデフォルト（Adaptive）のままにすることをお勧めします。この方式はリンク障害からフローを復元する場合に効率的です。	詳細については、「ハッシュ方式：Fixed vs Adaptive」を参照してください。
トラフィック フローの負荷分散を最適化するため、VSL ポート チャネルのリンク数は常に 2 の倍数（2、4、および 8）にバンドルします。	詳細については、「ハッシュ方式：Fixed vs Adaptive」を参照してください。

推奨される VSS のベスト プラクティス	トピック
シスコでは、次の理由から、スイッチのプリエンプションを*設定しない*ことをお勧めします。 • 設定すると、複数のスイッチがリセットされ、フォワーディング機能が低下し、予期しないネットワークの停止につながる。 • VSS は単一の論理スイッチまたはルータである。両方のスイッチ メンバは、アクティブなロールを担う機能があるという点で等価です。これは、企業ポリシーで必要とされない限り、どちらがアクティブであるかは関係ないためです。	詳細については、「スイッチのプリエンプション」を参照してください。
この場合のベスト プラクティスは、PAgP タイマーの設定をデフォルト値のままにし、通常の UDLD を使用してリンクの完全性をモニタすることです。	詳細については、「PAgP の hello 値をデフォルトに設定したままにする理由」を参照してください。
この場合のベスト プラクティスは、LACP タイマーの設定をデフォルト値のままにし、通常の UDLD を使用してリンクの完全性をモニタすることです。	詳細については、「LACP の hello 値をデフォルトに設定したままにする理由」を参照してください。
シスコでは、*switch virtual domain* コマンドを使用して VSS ドメインの仮想 MAC アドレスを設定することをお勧めします。	詳細については、「MAC アドレス」を参照してください。
シスコでは、デフォルト MAC OOB 同期化アクティビティ間隔の 160 秒（設定可能最小値）と、デフォルト MAC OOB 同期化アクティビティ間隔の 3 倍の値のアイドル MAC エージング タイマー（480 秒）をイネーブルにして、保持することをお勧めします。	詳細については、「Out-Of-Band Synchronization コンフィギュレーションの推奨」を参照してください。
シスコでは、VSS 対応デザインにおいて、インターフェイスの両端のトランクについて、desirable-desirable または auto-desirable オプションを使用して構成することをお勧めします。	詳細については、「トランキング構成のベスト プラクティス」を参照してください。
シスコでは、必要な VLAN についてはトランク経由で転送するように明示的に設定することをお勧めします。	詳細については、「トランク上の VLAN 構成」を参照してください。
アグレッシブ UDLD は、リンク整合性チェックとして使用するべきではありません。ケーブル配線の障害の検出や、リンク整合性のチェックには、通常モードの UDLD を使用してください。	詳細については、「Unidirectional Link Detection (UDLD; 単方向リンク検出)」を参照してください。
シスコでは、ループを回避し、リンクまたはノードの障害時に最良のコンバージェンスを実現するために、VSS に接続する各デバイスは MEC（レイヤ 2 およびレイヤ 3）機能を備えたスター型トポロジを常に使用することをお勧めします。	詳細については、VSS でのトポロジの考慮事項を参照してください。
シスコでは、VSS 対応キャンパス ネットワークでは、Loop Guard を*イネーブルにしない*ことをお勧めします。	詳細については、「Loop Guard」を参照してください。
VSS 対応ネットワークでは、エッジ ポートを STP に関与させないようにすることが非常に重要です。シスコでは、エッジ ポートで PortFast と BPDU Guard をイネーブルにすることを強くお勧めします。	詳細については、「PortFast および BPDU Guard」を参照してください。
シスコでは、表 3-5 に記載された値未満にチューニング*しない*ことを強くお勧めします。その他の NSF 関連のルート タイマーはすべてデフォルト値のままにし、変更しないでください。	詳細については、「NSF の復旧、および IGP の相互作用」を参照してください。
シスコでは、レイヤ 3 の MEC ベースのトポロジを使用して、VSL バンドルでのマルチキャスト トラフィックの複製を回避し、VSL リンクでのトラフィックの再ルーティングで発生する遅延を防ぐことをお勧めします。	詳細については、「ECMP と MEC のトラフィック フローの比較」を参照してください。

推奨される VSS のベスト プラクティス	トピック
レイヤ 2 環境では、2 つの VSS 間に単一の論理リンク（オプション 5）を配置する方法が*唯一の*推奨トポロジです。他の接続シナリオではトポロジがループします。	詳細については、「コアの VSS」を参照してください。
シスコでは、VSS 対応環境ではデュアルアクティブ検出を有効にすることをお勧めします。	詳細については、「VSS デュアルアクティブ スーパーバイザを使用したキャンパス復旧」を参照してください。
ベスト プラクティスの推奨事項は、VSS 環境でデュアルアクティブ イベントが発生している間は、コンフィギュレーション モードを開始*しない*ことです。しかし、VSL リンクの偶発的なシャットダウンに対して必要なコンフィギュレーションの変更や、VSL を適切に復旧するために必要なコンフィギュレーション変更は回避することができません。	詳細については、「VSL リンク関連のコンフィギュレーションの変更」を参照してください。
デフォルトの hello タイマーとホールド タイマーを実行するには、ルーティング プロトコルを使用することをお勧めします。	詳細については、「デュアルアクティブ状態によるコンバージェンスとユーザ データ トラフィックへの影響」を参照してください。

CHAPTER 2

Virtual Switching System 1440 アーキテクチャ

この章では、Cisco Catalyst 6500 シリーズ Virtual Switching System（VSS）1440 のアーキテクチャとコンポーネントについて扱います。このデザイン ガイドは VSS の（テクノロジー自体ではなく）展開仕様に焦点を当てていますが、キャンパス設計に影響のあるすべての必須 VSS コンポーネントについての重要な詳細が含まれています。これには、動作特性、設計のトレードオフ、推奨ベスト プラクティス設定が含まれています。VSS テクノロジーの詳細については、次のマニュアルを参照してください。

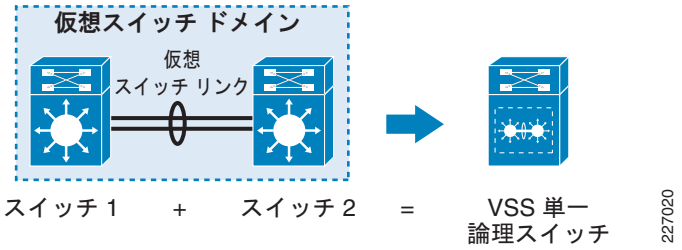
Cisco Catalyst 6500 Series Virtual Switching System (VSS) 1440 White Paper on VSS technology:

http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps9336/white_paper_c11_429338.pdf

VSS アーキテクチャおよび動作

VSS は、2 つの Cisco Catalyst 6500 スイッチで単独の論理ネットワーク エンティティ形成します。2 つのシャーシを組み合わせて、単一の*仮想スイッチ ドメイン*を形成し、単一の論理スイッチおよび/またはルータとして他のネットワークと対話します。図 2-1 を参照してください。

図 2-1　　VSS の概念図



VSS ドメインは、2 つのスーパーバイザで構成されていて、各メンバ シャーシ内にあるスーパーバイザは *Virtual Switch Link*（VSL; 仮想スイッチ リンク）を介して接続されています。VSL は、2 つのスイッチ間の通信を行っています。VSS 内では、1 つのシャーシ スーパーバイザが*アクティブ*、もう一方が*ホットスタンバイ*に指定されています。いずれも *Stateful Switch Over（SSO; ステートフル スイッチオーバー）*テクノロジーを使用しています。アクティブ スーパーバイザを含むスイッチは*アクティブ スイッチ*と呼ばれ、ホットスタンバイ スーパーバイザを含むスイッチは*ホットスタンバイ スイッチ*と呼ばれます。VSS は、分散フォワーディング アーキテクチャを持つ統合コントロール プレーン上で動作します。これは、アクティブ スーパーバイザ（またはスイッチ）がアクティブに残りのネットワークに関与し、コントロール プレーン情報を管理し維持する役割があります。アクティブ スイッチは、VSL を介して、システムの状態とネットワーク プロトコルに関する最新情報でホットスタンバイ

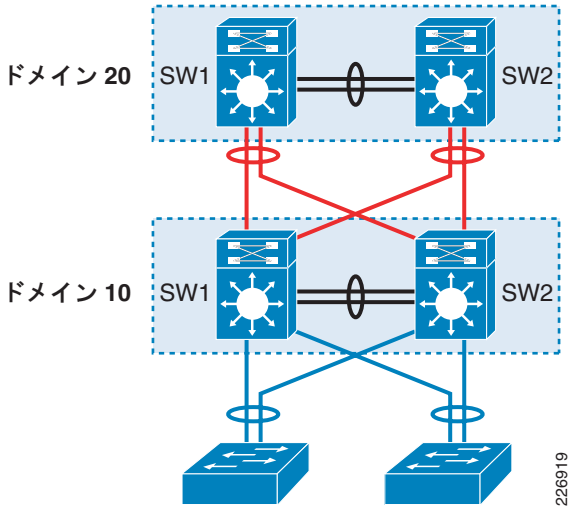
スーパーバイザを維持し更新します。アクティブ スーパーバイザに障害が発生した場合、ホットスタンバイ スーパーバイザがコントロール プレーンの管理でアクティブ ロールを担います。両物理シャーシがデータ フォワーディングを行い、データ フォワーディングは分散型で実行されます。VSS に隣接したデバイスは、*Multichassis EtherChannel（MEC）*を介して接続され、単一論理接続を形成します。単一論理スイッチは、MEC と組み合わせて、高アベイラビリティとループのないトポロジの基礎となります。これ以降、このセクションでは、キャンパス分散ブロックで VSS の展開に影響する、VSS のコンポーネントに関する動作とベスト プラクティス設定について詳細に説明します。このデザイン ガイドでは、既存ネットワークから VSS ベースの環境への移行に関する詳細な手順については説明せず、またすべての準備作業についても取り上げませんが、キャンパスにおける VSS の展開に影響する可能性のある重要な手順については説明しています。

Virtual Switch Domain（VSD; 仮想スイッチ ドメイン）およびスイッチ ID

仮想ドメイン

ドメイン ID を定義することは、2 つの物理シャーシから VSS を作成するための最初のステップです。固有のドメイン ID により、VSS ドメインを定義する同一 VSS ペアの一部と見なされている 2 つのスイッチが識別されます。ドメイン ID の割り当てにより、複数の仮想スイッチ ペアを階層型に接続することができるようになります。特定のドメインに関与できる VSS ペアは 1 つだけです。ドメイン ID には、1 ～ 255 の範囲の値を使用でき、複数の VSS ペアが接続されている場合、一意である必要があります。図 2-2 を参照してください。

図 2-2　　VSS ドメイン ID



ドメイン ID は、次の例で示すように両方の物理スイッチに定義されています。

スタンドアロン スイッチ 1：

```
VSS-SW1# config t
VSS-SW1(config)# switch virtual domain 10
```

スタンドアロン スイッチ 2：

```
VSS-SW2# config t
VSS-SW2(config)# switch virtual domain 10
```

ドメイン ID の使用は、図 2-2 で示すような形で VSS が複数レイヤで展開されているネットワークで重要です。この一意 ID は、仮想 Media Access Control（MAC; メディア アクセス制御）、Port Aggregation Protocol（PAgP; ポート集約プロトコル）、および Link Aggregate Control Protocol（LACP）制御パケットなど、多くの異なるプロトコルおよびシステム設定で使用されています。2 つの接続済み VSS ドメインに同じドメイン ID が含まれている場合、この競合が VSS の動作に影響します。

次の例は、LACP システム ID で使用されているドメイン ID のコマンド出力を示したものです。最後のオクテットの 0a（16 進数）は、ドメイン ID 10（10 進数）から派生したものです。

```
6500-VSS# sh lacp sys-id
32768,0200.0000.000a <--
```

ヒント　複数の VSS ドメインを接続していない場合でも、ベスト プラクティスとして一意のドメイン ID を使用することを推奨します。

スイッチ ID

VSS は、物理スイッチのペアで構成されていて、各シャーシを一意の番号で識別するためにスイッチ ID が必要です。スイッチ ID は 1 または 2 のいずれかで、各メンバ シャーシ内で一意である必要があります。この番号は、仮想スイッチのロール（アクティブ スイッチかホットスタンバイ スイッチか）に関係なくインターフェイス名が同じままであることを保証するために、インターフェイスの名前付けの一部として使用されます。通常、スイッチ ID は、次の例で示すように、コンフィギュレーション モードで設定する必要があります。

スタンドアロン スイッチ 1：

```
VSS-SW1# config t
VSS-SW1(config-vs-domain)# switch 1
```

スタンドアロン スイッチ 2：

```
VSS-SW2# config t
VSS-SW2(config-vs-domain)# switch 2
```

ただし、あるスーパーバイザに発生したハードウェア上の問題により新規スーパーバイザが必要になった場合、次の設定例に示しているように、command-line interface（CLI; コマンドライン インターフェイス）を介してスイッチ ID をイネーブル モードに設定できます。

スタンドアロン スイッチ 1：

```
6500-VSS# switch set switch_num
```

スタンドアロン スイッチ 2：

```
6500-VSS# switch set switch_num 2
```

いずれの方法も、スイッチ ID を各メンバ シャーシ ROMMON に書き込んでいます。ドメイン ID とスイッチ番号は、次の CLI 例を確認できます。

```
6500-VSS# show switch virtual
Switch mode                  : Virtual Switch
Virtual switch domain number : 10
Local switch number          : 1
Local switch operational role: Virtual Switch Active
Peer switch number           : 2
Peer switch operational role : Virtual Switch Standby
```

注意　コマンド **write erase** を使用して新規スタートアップ コンフィギュレーションをコピーすることは避けてください。このコマンドは、ROMMON に格納されているスイッチ番号を消去するため、両方のスイッチがスタンドアロン モードで起動することになります。スイッチ番号を設定する CLI は VSS モードで使用できないため、両方のスイッチがリブートした後でだけ **switch set switch_num** *1/2* コマンドを使用します。

仮想スイッチ リンク（VSL）

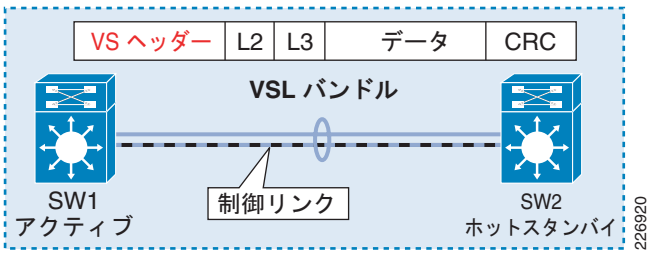
VSS は、単一の論理エンティティを構成するために 2 つの物理シャーシを組み合わせて構成されています。このコントロール プレーンの単一化は、システム制御、シグナリング、およびバックプレーンが両シャーシ内で単一エンティティとして存在している場合だけ可能です。このシステム コントロール プレーンの拡張は、専用 EtherChannel バンドルを介して実現されています。EtherChannel に属しているリンクは仮想スイッチ リンク（VSL）と呼ばれます。VSL は、ホットスタンバイ スーパーバイザ プログラミング、ラインカード ステータス、Distributed Forwarding Card（DFC; 分散型フォワーディング カード）カード プログラミング、システム管理、診断などの重要なシステム制御情報を搬送する論理接続として機能します。さらに、VSL は必要時にユーザ データ トラフィックを伝送することもできます。したがって、VSL には、システム制御の同期とデータ リンクのサポートという 2 つの目的があります。

VSL リンクはシステム制御リンクとして取り扱われ、すべてのトラフィックを Virtual Switch Header（VSH; 仮想スイッチ ヘッダー）と呼ばれる特別システム ヘッダーにカプセル化します。このカプセル化は、専用ハードウェア リソースを使用して実行され、VSL は次のハードウェア ポートの 10 ギガビット インターフェイス上でだけ設定可能です。

- Sup720-10G、10-Gbps ポート
- WS-X6708
- WS-X6716（パフォーマンス モードだけ、12.2(33)SXI が必要）

VSH のサイズは、Cisco Catalyst 6500 で使用されている内部コンパクト ヘッダーのサイズ（32 バイト長）と同じです。このヘッダーは、イーサネット プリアンブルの後ろでレイヤ 2 ヘッダーの直前に置かれます。図 2-3 を参照してください。

図 2-3　VSL ヘッダーおよび制御リンク



VSL リンクの初期化と動作特性

VSS は、分散フォワーディング アーキテクチャを備えた単一コントロール プレーンを装備しています（「ステートフル スイッチオーバー テクノロジー」（P.2-23）を参照）。1 つのスーパーバイザだけがコントロール プレーンを管理していても、両方のスイッチがコントロール プレーン情報の学習に関与しています。ホットスタンバイ スイッチを介して取得されたネットワークおよびシステム コントロール プ

レーン情報は、アクティブ スーパーバイザに送信され、結果としてホットスタンバイ スーパーバイザが更新されます。スイッチ間の学習および更新に関するこの双方向プロセスは、VSL リンクを通じて実行されます。

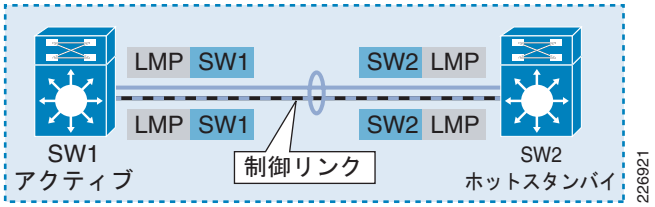
VSL はシステム初期化中に 2 つの独立したシステムが互いを認識することのできる唯一のパスであるため、VSS を構成するコンポーネントです。このシステム初期化中の相互認識は、アクティブまたはホットスタンバイ仮想スイッチのいずれかとなる際に、各物理シャーシの各ロールを決定するために必要です。このため、VSL リンクは、起動プロセスの非常に早い時期で、その他の主要なサービスやコンポーネントよりも前に始動します。このタスクを実行するために、各スイッチには、VSL リンクと関連ラインカードの起動シーケンスをアクティブ化するために、スイッチ番号と ROMMON で必要なその他の情報が格納されています。VSL 初期化中に、システムに対して仮想スイッチを形成するために必要なさまざまな互換性チェックが行われます。VSS では、両方のシャーシで同じスーパーバイザ タイプと Cisco IOS ソフトウェア バージョンが必要です。仮想スイッチを形成するために必要な追加のシステム要件については、該当するリリース ノートを参照してください。VSL リンク初期およびメンテナンスは、*VSL Protocol*（VSLP）フレームワークで実行されます。これは、*Link Management Protocol*（LMP; リンク管理プロトコル）と *Role Resolution Protocol*（RRP; ロール解決プロトコル）の 2 つのプロトコルで構成されています。LMP はリンクの完全性を管理し、RRP は仮想スイッチ ドメイン内の各スイッチ メンバのロールを決定します。RRP については、「ステートフル スイッチオーバー テクノロジー」（P.2-23）で説明されています。

リンク管理プロトコル（LMP）

LMP は、初期化される最初のプロトコルで、VSL ラインカード診断が完了し、VSL リンクがオンラインになるときに初期化されます。図 2-4 を参照してください。LMP は、次の重要な機能を実行するために設計されています。

- 始動時および通常動作中の双方向通信の確立と確認。
- 別の仮想スイッチに接続されている重複スイッチ ID またはメンバを検出するためのスイッチ ID の交換。これは、スイッチのロール（アクティブまたはスタンバイ）を判別するために RRP で使用されます。
- VSL およびピア スイッチの健全性を監視するための LMP hello タイマーの独立した送受信。

図 2-4　　リンク管理プロトコル（LMP）

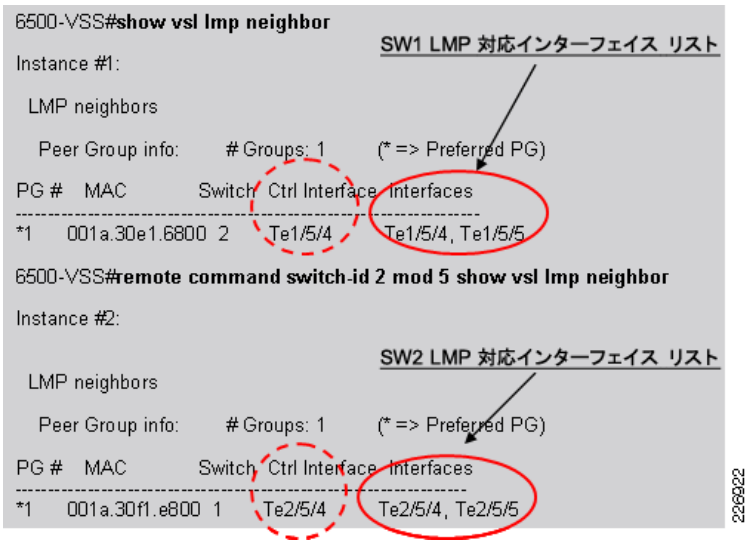


LMP は、同一仮想スイッチ ドメイン（VSD）内の各メンバ スイッチで独立して動作しています。PAgP、LACP、Interior Gateway Protocol（IGP）hello などの、アクティブ スイッチがプロトコルの発信と終端を行う単一コントロール プレーンを使用する他のプロトコルとは異なり、VSD 内の両スイッチは、Switch Processor（SP; スイッチ プロセッサ）上の LMP コントロール プレーン パケットを独立して発信および終端します。図 2-5 の赤い点線で囲まれた部分を参照してください。LMP は、異なるポートにまたがる同一ピアの複数のステート マシンを維持するために、各 VSL メンバ リンクで実行するように設計されています。ポートの単一方向状態が検出された場合、LMP はこれをダウンとマークして、ポートを *err-disable* 状態にするのではなく VSLP ネゴシエーションを再開しようとします。VSD の各メンバ スイッチでは、LMP が VSL リンクの共通セットで内部的に単一の一意 Peer Group（PG; ピア グループ）ID を作成します。すべての VSL インターフェイスがダウンする場合、LMP がピア グループを破棄して、RRP に適切なアクションを行うように通知します。アクティブ ス

イッチは、ホットスタンバイ スイッチに関連づけられたすべてのインターフェイスを切り離します。同時に、ホットスタンバイ スイッチが切り替えを実行して、アクティブ ロールを引き受けて、前のアクティブ スイッチに関連づけられていたすべてのインターフェイスを切り離します。

LMP hello パケットは、Cisco Discovery Protocol（CDP; シスコ検出プロトコル）01.00.0C.CC.CC.CC と一致する宛先 MAC アドレスでカプセル化された、Logical Link Control（LLC; 論理リンク制御）および Sub-network Access Protocol（SNAP; サブネットワーク アクセス プロトコル）です。LMP や hello パケットを含むすべてのシャーシ内部コントロール プレーン トラフィックは、Bridge Protocol Data Unit（BDPU; ブリッジ プロトコル データ ユニット）パケットとして分類され、自動的に送信プライオリティ キューに配置されます。

図 2-5　　LMP が有効なインターフェイスのリストを示す出力



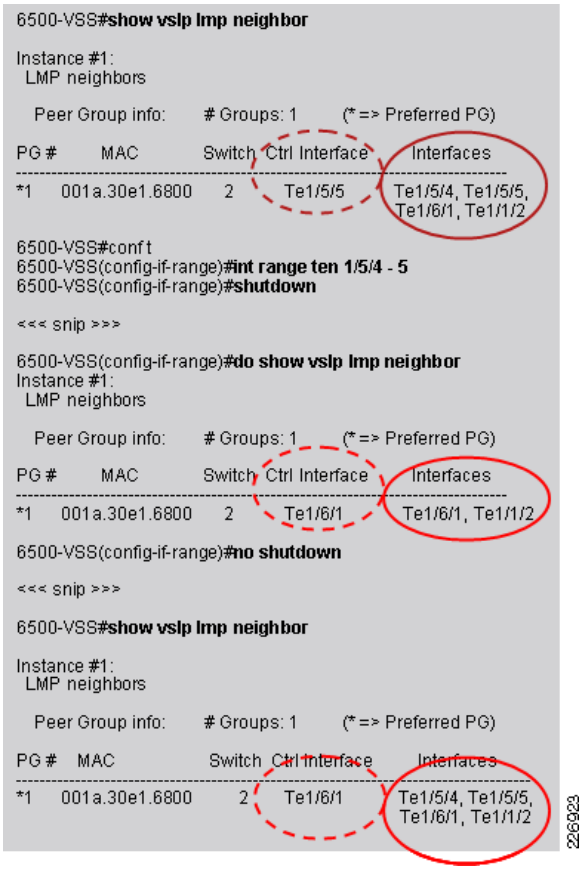
制御リンクとシャーシ間コントロール プレーン

VSL バンドルは、最大 8 メンバまで持つことのできる専用の EtherChannel です。設定メンバの中から 1 リンクだけが制御リンクとして選択され、その制御リンクはシャーシ間コントロール プレーンを搬送できる唯一のリンクです。制御リンクは、ラインカード通信用の Control Packet（SCP）、Inter-process Communication（IPC; プロセス間通信）パケット、およびプロトコル データベースとステートを通信する Inter-Card Communication（ICC）を含む、スイッチ間 External Out-of-Band Channel（EOBC; 外部アウトオブバンド チャネル）制御トラフィックを搬送し、ホットスタンバイ スーパーバイザを更新します。さらに、（発信元および宛先 MAC アドレスおよび/または IP アドレスに基づく）トラフィックのハッシュ方法に応じて、同じリンクでユーザと他のネットワーク制御トラフィックを搬送できます。残りのバンドルされたリンクは、ネットワーク コントロール プレーンおよびユーザ データ トラフィックを搬送しますが、シャーシ間コントロール プレーン トラフィックは搬送しません（「トラフィックの優先順位付けと VSL での負荷分散」（P.2-15）を参照）。制御リンクを図 2-4 に示します。

制御リンク選択手順は、VSS システムによって決定され、ユーザは管理できません。起動プロセス中に、LMP 関係（ステートマシン）を確立する最初の VSL リンクが制御リンクとして選択されます。Cisco Catalyst 6500 アーキテクチャに基づいて、スイッチに取り付けられたその他のモジュールよりも前にスーパーバイザ モジュールが動作可能になります。Sup720-10G モジュールの 10-Gbps ポートが VSL EtherChannel にバンドルされている場合、このポートが両方のスイッチがブート プロセスを実行するたびに制御リンク インターフェイスとして選択されます。

図 2-6 で説明されている **show vslp lmp neighbor** コマンド出力は、(赤い点線で囲まれている) 現在の制御リンク インターフェイスと（現在の制御リンク パスに障害が発生した場合に使用できる）バックアップ VSL インターフェイスのリストを示します。バックアップ インターフェイスは、ローカル仮想スイッチの VSL EtherChannel のメンバ リンクです。高度に冗長な VSL ネットワーク設計を行う場合、VSL EtherChannel はスイッチ メンバ間の複数の VSL 対応 10-Gbps ポートとバンドルされている必要があります。この場合、**show vslp lmp neighbor** コマンドの *Interfaces* 列に最初にリストされているインターフェイスは、現在の制御リンク インターフェイスに障害が発生するとすぐに制御リンク インターフェイスとなります。Sup720-10G モジュールの 10-Gbps ポートが復元されて再び VSL EtherChannel に加わると、現在の制御リンク インターフェイスに影響することなく、次の使用可能な制御リンク バックアップ パスとなります（制御リンクが復元された後の図 2-6 に示している出力の 2 番目の部分を参照）。

図 2-6　　制御リンク インターフェイス選択



LMP ハートビート

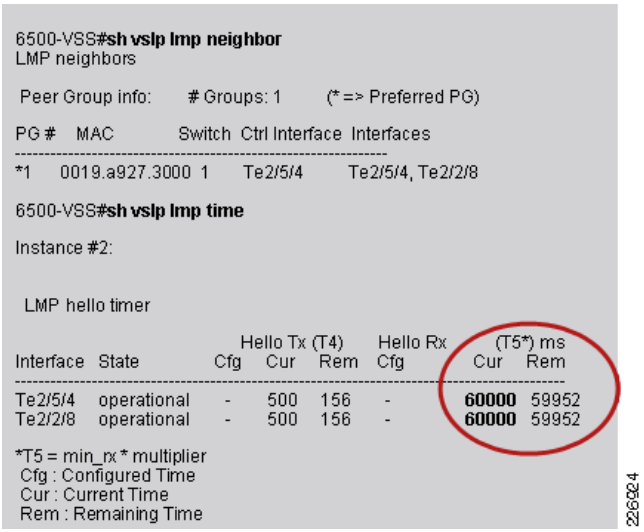
LMP ハートビート（LMP hello タイマー）は、ピア スイッチのアベイラビリティと接続性をチェックすることにより、VSS の完全性を維持するうえで重要な役割を担っています。両方の VSS メンバは、最後にバンドルされた VSL リンク上の設定済み保持タイマー設定の範囲で LMP hello メッセージを検出できない場合、独立した決定論的 SSO 切り替え動作を実行します。LMP タイマーのセットは、VSL リンクの健全なステータスを維持するために適用される hello 伝送の間隔を決定するために、組み合わせて使用されます。3 つのタイマーは次のとおりです。

- hello 伝送タイマー（T4）

- 最小受信タイマー（min_rx）
- T5 タイマー（min_rx * マルチプレクサ（multiplexer））

図 2-7 は、VSL リンク メンバごとのタイマー値を表している CLI 出力の例です。

図 2-7　VSL リンク メンバごとのタイマー値



デフォルトで、LMP hello 伝送タイマー（T4）および受信タイマー（min_rx）はそれぞれ 500 ミリ秒の値が割り当てられています。保持タイマー（T5 タイマー）は min_rx から取得され、デフォルトのマルチプレクサは 120 です（CLI はデフォルト マルチプレクサを示しません）。デフォルトで、VSL メンバ リンク タイム アウトは 60,000 ミリ秒（60 秒）で検出されます。T5 の期限は、リモート ピア（アクティブ スイッチまたはホットスタンバイ スイッチ）で想定される不安定性を示しています。各スイッチ メンバは、T5 タイマーが Expire（期限切れ）になると、それぞれ独立したアクションを実行します。このアクションには、次のものが含まれています（これに限定されません）。

- 制御リンクである VSL ポートで Expire が発生すると、問題を検出したスイッチで強制的に新たな制御リンクの選択が行われます。T5 タイマーがリモート ピアで Expire になっていなくても、リモート ピア スイッチは要求を尊重して、内部ロジックを再プログラミングし、コントロール プレーン トラフィックを、制御ポートとして新規に選択された VSL ポートに送信できます。
- 制御リンク ポート以外で Expire が発生した場合、障害を検出したスイッチがユーザ データ トラフィックに使用できるポートを選択します。最終的に、リモート ピアが変更を検出し、リンクをバンドルから削除します。
- 最後の VSL ポート（制御リンクとユーザ データ ポートの組合せ）で Expire が発生し、アクティブ スイッチでタイムアウトが検出された場合、アクティブ スイッチがすべてのピア スイッチ インターフェイスを削除して、これらのインターフェイス（レイヤ 2 またはレイヤ 3 プロトコル）の設定に応じて、残りのネットワークに対して変更をアナウンスします。最終的に、ホットスタンバイ モードにあるピア スイッチが T5 タイマーの Expire を検出し、LMP が RRP に通知して、ホットスタンバイ スイッチを強制的にアクティブ スイッチにします。これにより、*デュアルアクティブ*と呼ばれる状態が発生し、両方のスイッチがアクティブ ロールを宣言するという不安定な状態になります（このような状態を回避する方法については、「VSS デュアルアクティブ スーパーバイザを使用したキャンパス復旧」（P.4-19）を参照してください）。

タイマーを変更すべきでない理由

LMP タイマーは、主に（CPU の使用率が高い場合や異常なソフトウェアの動作中に）VSS の完全性を保証するためのものです。通常のハードウェア障害の検出や（ユーザやシステムが開始する）スイッチオーバーは、*Fast Link Notification*（FLN）と呼ばれるハードウェア メカニズムを通じて呼び出されます。適用可能なイベントが発生する場合、FLN が、WS-X6708 のファームウェアや Sup720-10G ポートに対して、必要なアクションを取るように（通常は 50 ～ 100 ミリ秒の速度で）通知します。FLN は、リモート スイッチの障害を検出するように設計されていません。ほとんどの場合、デフォルトの LMP タイマーをより短い値に変更してもコンバージェンスは改善されません。これは、LMP タイマーの前にシャーシ間のコントロール プレーン プロトコル タイマー（IPC タイマー）の期限が切れて、これによってアクションが取って代われるためです。

VSLP タイマーが積極的に変更される場合、予期しない結果となる可能性があります。たとえば、VSLP タイマーの値を低い値に変更すると、不安定さが大幅に増加することになります。シナリオ例の説明は、次のとおりです。

> VSLP タイマー値を低く設定すると、通常 VSL がメンバ スイッチ間のネイバー関係を構築できなくなり、スイッチのリブートが繰り返されます（Cisco IOS 側からはクラッシュとなります）。起動プロセス中に、VSS スイッチは複数の高優先順位アクティビティを処理する必要があります。アグレッシブ VSLP タイマーを有効にすると、リソース不足により各メンバが VSLP セッションを維持することができなくなる可能性があります。LMP hello がいずれかの側でタイムアウトすると、LMP が VSL EtherChannel から VSL メンバ リンクを削除します。結果的に、アクティブ スイッチとホットスタンバイ スイッチの間のすべてのリンクに障害が発生する可能性があります。ホットスタンバイ スイッチの場合、これが最悪のエラーと見なされ、これを回復する方法は、即座にリセット信号（crash）を送信してはじめからやり直す方法以外ありません。これは、少なくとも 1 つの LMP セッションが確立されるまで続くこともあるため、ネットワークが大幅に不安定になる可能性があります。

さらに、VSL リンク輻輳または CPU の高使用率によって、VSS メンバも設定された T5 タイマーの制限範囲内で LMP hello メッセージを送受信できなくなる可能性があります。そのような状態では、VSS システムが不安定になり、デュアルアクティブ状態になることもあります。

ヒント　**シスコではデフォルトの LMP（VSLP）タイマーを変更しないことを強くお勧めしています。**

ロール解決プロトコル（RRP）

RRP は、各 VSS スイッチ メンバの動作ステータスを決定する役割があります。設定されたパラメータに基づいて、メンバ スイッチはアクティブ、ホットスタンバイ、または Route Process Redundancy（RPR）のロールを担うことができます。RRP は、Cisco IOS ソフトウェア互換性のチェックを実行します。ソフトウェア バージョンに互換性がない場合、RRP が 1 つのスイッチを RPR モードにして、すべてのラインカードの電源をオフにします。アプリケーションがシャーシ冗長性と SSO 動作に関連している度合いが深いため、「仮想スイッチのロール、プライオリティ、およびスイッチ プリエンプション」（P.2-28）に RRP の詳細があります。

VSL バンドルの設定

VSL の設定は、（ドメイン ID を定義した後の）仮想スイッチを作成する 2 番目の手順です。VSL バンドルは、専用のポート チャネルです。各スタンドアロン スイッチは、ポートチャネル番号を割り当てる前に、一意のポートチャネル インターフェイス番号を使用して設定される必要があり、ポートチャネル インターフェイス番号はいずれかのスイッチの既存のスタンドアロン設定で使用されていないことを確認する必要があります。次の設定ステップは、スタンドアロン スイッチを VSS に変換するため

に各スイッチで必要です。詳細な変換プロセスについては、このマニュアルで取り扱う範囲を超えています。VSS 変換の取り扱いについては、次の URL にある『*Migrate Standalone Cisco Catalyst 6500 Switch to Cisco Catalyst 6500 Virtual Switching System*』の文書を参照してください。

http://www.cisco.com/en/US/products/ps9336/products_tech_note09186a0080a7c74c.shtml

スタンドアロン スイッチ 1：

```
VSS-SW1(config)# interface Port-Channel1
VSS-SW1(config-if) #switch virtual link 1

VSS-SW1(config-if)# interface range Ten5/4 - 5
VSS-SW1(config-if)# channel-group 1 mode on
```

スタンドアロン スイッチ 2：

```
VSS-SW2(config-if)# interface Port-Channel2
VSS-SW2(config-if)# switch virtual link 2

VSS-SW2(config-if)# interface range Ten5/4 - 5
VSS-SW2(config-if)# channel-group 2 mode on
```

VSL EtherChannel がメンバ リンク単位で LMP を使用するため、PAgP や LACP などのリンクアグリゲーション プロトコルは不要で、各メンバ リンクは、**channel-group group-***number* **mode on** コマンドを使用して無条件の EtherChannel モードで設定される必要があります。VSL 設定が完了すると、イネーブル プロンプトで **switch convert mode virtual** CLI コマンドを使用すると、変換プロセスが開始されます。変換プロセスには、*slot/interface* から *switch_number/slot/interface* へのインターフェイスの命名規則の変更、設定の保存、およびリブートが含まれています。スイッチのリブート中に、システムで VSL 設定が認識されて、関連 VSL ポート初期化プロセスに移行します。2 つのスイッチは、互いに通信し、どちらがアクティブ ロールまたはホットスタンバイ ロールを持つのかを判別します。この情報交換は、次のコンソール メッセージで明らかになります。

スタンドアロン スイッチ 1 コンソール：

```
System detected Virtual Switch
configuration...
Interface TenGigabitEthernet
1/5/4 is member of PortChannel 1
Interface TenGigabitEthernet
1/5/5 is member of PortChannel 1
<snip>
00:00:26: %VSL_BRINGUP-6-MODULE_UP: VSL module in slot 5 switch 1 brought up
Initializing as Virtual Switch active
```

スタンドアロン スイッチ 2 コンソール：

```
System detected Virtual Switch configuration...
Interface TenGigabitEthernet
2/5/4 is member of PortChannel 2
Interface TenGigabitEthernet
2/5/5 is member of PortChannel 2
<snip>
00:00:26: %VSL_BRINGUP-6-MODULE_UP: VSL module in slot 5 switch 2 brought up
Initializing as Virtual Switch standby
```

最初に、VSS 変換では、統合スイッチの仮想モード動作を受け入れる最終ステップとして次のコマンドを実行する**必要**があります。スイッチが変換される場合、または部分的に変換される場合、このコマンドを使用できません。

```
6500-VSS# switch accept mode virtual
```

このコマンドは、ホットスタンバイ スイッチからのすべての VSL リンク関連設定を強制的に統合して、実行コンフィギュレーションをこれらのコマンドに読み込みます。さらに、スタートアップ コンフィギュレーションが新しく結合されたコンフィギュレーションに更新されます。次のプロンプトが表示されます。

```
Do you want proceed? [yes/no]: yes
Merging the standby VSL configuration. . .
Building configuration...
[OK]
```

（注）　VSL 関連コンフィギュレーションだけが変換ステップで結合されます。その他のすべてのコンフィギュレーションは、ネットワーク サイト要件ごとに管理する必要があります。詳細については、関連シスコ製品のマニュアルを参照してください。

VSL 特性

VSL ポート チャネルは、内部システム リンクとして処理されます。結果として、その設定、復元力、動作モード、QoS（Quality of Service）、およびトラフィック負荷分散は、VSL 固有の規則に従います。このセクションでは、これらの規則に関連した設定要件について説明します。論理ポート チャネルおよびそのメンバ リンクには、それぞれ固有の制約事項があります。

VSL ポートチャネル論理インターフェイス設定は、VSL 関連設定に制限され、その他すべての Cisco IOS 機能はディセーブルです。次の出力は、使用可能なオプションについて説明したものです。

```
6500-VSS(config)# int po 1
6500-VSS(config-if)# ?
virtual link interface commands (restricted):
  default        Set a command to its defaults
  description    Interface specific description
  exit           Exit from virtual link interface configuration mode
  load-interval  Specify interval for load calculation for an interface
  logging        Configure logging for interface
  mls            mls sub/interface commands
  no             Negate a command or set its defaults
  port-channel   Port Channel interface subcommands
  shutdown       Shutdown the selected interface
  switch         Configure switch link
  vslp           VSLP interface configuration commands
```

VSL メンバ リンクは、VSL 設定が適用されるたびに制限されたコンフィギュレーション モードになります。以下を除くすべての Cisco IOS 設定オプションはディセーブルになります。

- EtherChannel
- Netflow 設定
- デフォルトの QoS 設定

次の出力は、使用可能なオプションについて説明したものです。

```
6500-VSS(config)# int ten 1/5/4
6500-VSS(config-if)# ?
virtual link interface commands (restricted):
  channel-group   Etherchannel/port bundling configuration
  default         Set a command to its defaults
  description     Interface specific description
  exit            Exit from virtual link interface configuration mode
  load-interval   Specify interval for load calculation for an interface
  logging         Configure logging for interface
  mls             mls sub/interface commands
  no              Negate a command or set its defaults
```

```
  priority-queue  Configure priority scheduling
  rcv-queue       Configure receive queue(s)
  shutdown        Shutdown the selected interface
  wrr-queue       Configure weighted round-robin xmt queues
```

VSL インターフェイスの設定時に、仮想スイッチごとに 1 つの VSL EtherChannel 設定だけが可能です。追加 VSL EtherChannel を設定するとエラーになります。生成されるエラー メッセージの例は次のとおりです。

```
6500-VSS(config)# interface port-channel 3
6500-VSS(config-if)# switch virtual link 1
% Can not configure this as switch 1 VSL Portchannel since it already had VSL Portchannel
1 configured
6500-VSS(config-if)#
6500-VSS(config-if)# switch virtual link 2
% Can not configure this as switch 2 VSL Portchannel since it already had VSL Portchannel
2 configured
```

さらに、VSS 変換後に、ネイバー スイッチ ポートをローカル スイッチの VSL EtherChannel にバンドルできません。ここで使用している例では、Switch 1 VSL EtherChannel は Switch 2 からの物理ポートにバンドルできません。ただし、通常の EtherChannel にはそのような制限がありません。次の出力例は、この未サポート モードを設定しようとしたときに生成されるエラー メッセージを示したものです。

```
6500-VSS(config-if)# int te2/1/1
6500-VSS(config-if)# channel-group 1 mode on
VSL bundle across chassis not allowed TenGigabitEthernet2/5/5 is not added to port channel
1
```

以降、このセクションでは、VSL リンク上のトラフィックの優先順位付け、レジリエンシー（復元力）、および VSL で使用可能なトラフィック負荷分散オプションに関するデザインの考慮事項について説明します。

VSL QoS およびトラフィックの優先順位付け

今日のエンタープライズ アプリケーションの要件はさまざまです。多くのタイムクリティカルなサービス（音声、映像、マルチキャスト）や、Customer Relationship Management（CRM; カスタマー リレーションシップ マネージメント）、SAP、Oracle などのエンタープライズ アプリケーションでは、キャンパス ネットワークで特定の優先順位が必要です。QoS デザイン ガイド（次の URL で使用可能）では、エッジでのユーザ トラフィックを分類し、Differentiated Services Code Point（DSCP; DiffServ コード ポイント）トラスト モデルを介してインテリジェントにこれらの優先順位を使用する手法について説明しています。このデザイン ガイドには、VSS での一般的な QoS 要件と動作についての説明がありませんが、このセクションでは、デフォルト動作に関して VSL リンクに適用される場合の QoS、コントロール プレーンの保護方法、およびネットワーク設計者が考慮すべき実装オプションについて取り扱います。

http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND_40/QoSCampus_40.html

ハードウェア構成の依存関係

設定の容易性、柔軟性、および帯域幅要件により、これまでディストリビューション レイヤではコアおよびアクセス レイヤへの接続にラインカードを使用していたため、スーパーバイザ ポートの使用が制限されていました。アクセス レイヤは、スーパーバイザまたは非モジュラ スイッチの専用ポートからのアップリンクを使用します。Sup720-10G スーパーバイザには、コアへの接続を提供するための、スーパーバイザで利用可能な 10-Gbps ポートを使用する新規オプションが搭載されています。これには、VSL ポートおよび/または残りのネットワークへのアップリンクとして使用される場合に、QoS に関連して、Sup720-10G アップリンク ポートで選択された現在の設計を理解する必要があります。VSL リンクは、10-Gbps ポートでだけ設定可能です。スーパーバイザ ポートまたはラインカードで

VSL リンク設定を選択すると、スーパーバイザの未使用ポートでの QoS 機能に影響します。Sup720-10G スーパーバイザには 2 つの 10-Gbps ポートと 3 つの 1-Gbps ポートがあります。Sup720-10G アップリンク ポートは次の 2 モードのいずれに設定できます。

- *デフォルト：非 10 Gbps-only モード*

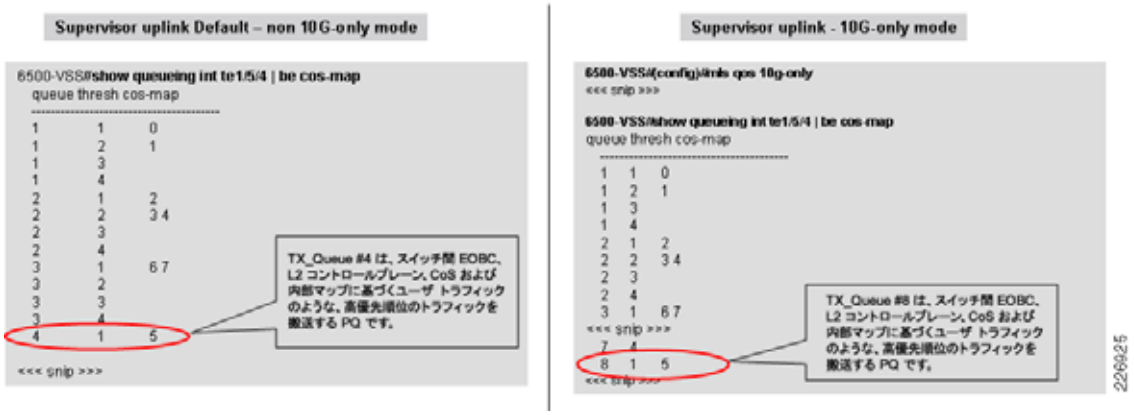
  このモードでは、すべてのポートが単一のキューイング モードに従う必要があります。いずれかの 10-Gbps ポートが VSL リンクで使用されている場合、VSL はサービス クラス（CoS）ベースのキューイングだけを許可するため、残りのポート（10 Gbps または 1Gbps）は他の非 VSL 接続用キューイングと同じ CoS モードに従います。

- *非ブロッキング：10Gbps-only モード*

  このモードでは、モジュール全体が非ブロッキング モードで動作しているため、すべての 1-Gbps ポートがディセーブルです。1 つの 10G ポートだけが VSL リンクとして使用されていた場合でも、両方の 10-Gbps ポートは CoS ベースのトラスト モデルに限定されます。12.2(33)SXI では、DSCP ベース キューイングを含むユーザ プリファレンスに基づいて未使用（非 VSL 設定）10-Gbps ポートを設定できるようになったことで、この制限が撤廃されました。

図 2-8 は、10-Gbps-only モードが、デフォルトの *1p3q4t* から *1p7q4t* 設定へ送信キューを増やすことを示しています。また、WS-X6708 モジュールと同じように受信キュー サイズを *2p4t* から *8q4t* へ増やします。ただし、デフォルトの Tx および Rx キュー マッピングは、10-Gbps-only モードのあるまたはない Cos ベースのままです。結果として、デフォルトの Cos とキューのマッピングは変更できないため、改良されたキュー構造を使用して各トラフィック クラスを個別のキューでマッピングすることに、実際上の利点はありません。

図 2-8　　デフォルトおよび非ブロッキング モードの比較



WS-X6708 ラインカード ポートの 1 つが VSL リンクとして使用されている場合、そのポートのポート キューイングが CoS ベースに制限されますが、残りのポートは独立した QoS 設定のセットを持つことができます。

レジリエント VSL デザインは、このことをデザイン要素として使用します。表 2-1 は、Sup720-10G アップリンク ポートのオプションと制限を要約したものです。「レジリエント VSL デザインの考慮事項」（P.2-19）では、高い復元力を持ち、柔軟な VSL デザインをデプロイする際のこれらのデザイン要素を採用することについて説明しています。

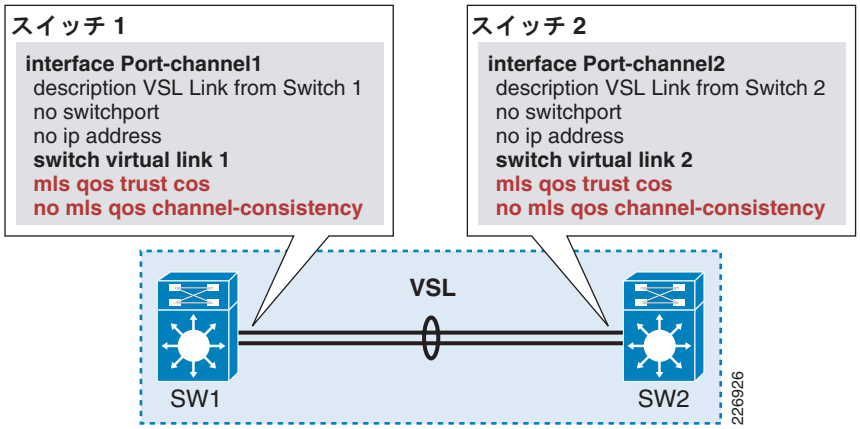
表 2-1　Sup720-10G アップリンク ポートのオプションと制限

Sup720-10G アップリンク ポート	10g-only モード	非 10g-only モード
キュー構造	Tx– 1p7q4t Rx – 8q4t	Tx –1p3q4t Rx – 2q4t
非 VSS モード（スタンドアロン モード） または VSS モード（VSL リンクとして使用されるスーパーバイザ ポートなし）	10-Gbps ポートだけが使用可能で、すべての 1-Gbps ポートはディセーブルです。 両方の 10-Gbps ポートには、トラスト モデルとキューイング モードを含む独立した QoS 設定のセットがあります。 DSCP ベースのキューイングが許可されます。	すべてのポートが使用可能です。すべてのアップリンクが単一 QoS 設定*だけ*を使用できます（マップ、しきい値、キュー）。 デフォルトのキューイング モードは CoS だけです。 DSCP ベースのキューイングは許可されません。
VSL リンクに使用される 10-Gbps ポート	両方の 10-Gbps ポートが VSL リンクに使用される場合、CoS ベースの信頼モードだけがサポートされます。 Cisco IOS 12.2(33) SXH では、単一の 10-Gbps ポートだけが VSL リンクに使用されている場合でも、両方の 10-Gbps ポートが CoS ベースのトラスト モデルに制限されます。Cisco IOS 12.2(33)SXI では、DSCP ベース キューイングを含むユーザ プリファレンスに基づいて残りの 10-Gbps ポートを設定できるようになったことで、この制限が撤廃されました。 残りの 1-Gbps ポートはディセーブルです。	両方の 10-Gbps ポートまたは単一の 10-Gbps ポートが VSL アップリンクとして使用され、1 つの QoS 設定だけを定義することができます。VSL は CoS ベースのキューイングだけを許可するために、残りの非 VSL ポートは、非 VSL 接続用の CoS ベースのキューイングに従います。

VSL 用のデフォルト QoS 設定

（デフォルトの）VSL は CoS ベースのトラスト モデルを使用します。このデフォルトの VSL EtherChannel の CoS-trust セッティングは、変更または削除ができません。グローバル モードでの QoS 設定に関係なく、VSL EtherChannel は CoS ベースのキューイング モードに設定されています。図 2-9 は、SW1 および SW 2 における VSL 設定を示しています。スタンドアロン スイッチと同様に、VSS は、ユーザ トラフィックを出力キューに配置するためのさまざまなマッピング テーブルを使用した内部 CoS ベースのキューイングを使用しています。VSL リンクを経由するトラフィックは、CoS ベースのトラスト モデルと同じファシリティを使用します。Weighted Round Robin（WRR）キューイング スケジューラは、各 VSL メンバ リンクにおいてデフォルトでイネーブルです。トラフィックが VSL リンクを経由する際に、VSL リンクの QoS 設定はユーザ トラフィックに設定された QoS マーキングを変更しません。

図 2-9　VSL CoS 設定例の比較



VSL リンク レジリエンシーの推奨されるベスト プラクティスは、別のソースから 2 つの 10-Gbps ポートをバンドルすることです。これを行うには、スーパーバイザから 1 ポート、Cisco 6708 ラインカードから 1 ポート必要になります。デフォルトでは、Sup720-10G スーパーバイザ 10-Gbps ポートには非分散型フォワーディング カード（DFC）1p3q4t Tx-queue 構造があり、WS-X6708 ラインカードには DFC- ベースの 1p7q4t Tx-queue 構造があります。EtherChannel の従来の設定では、一致しないキュー構造間のバンドルで障害が発生します。ただし、VSL バンドルは、EtherChannel を形成できる構成においてデフォルトでイネーブルです。これは、**no mls qos channel-consistency** コマンドを介してイネーブルになっています。

次の QoS コンフィギュレーションの制限は、ポートが VSL EtherChannel 内にバンドルされるたびに適用されます。

- CoS モード セッティングは変更できず、VSL ポート チャネルから削除できません。
- VSL ポート チャネルへのリンクのバンドル後、DSCP ベースのトラスト、受信キュー帯域幅、またはしきい値制限などのあらゆる QoS 設定は、Sup720-10GE モジュールの 1-Gbps または 10-Gbps ポートで変更することができません。このような制限のある QoS コンフィギュレーションの適用は、エラーとなります。
- ユーザ定義のモジュラ QoS CLI（MQC）サービスポリシーは、VSL EtherChannel に接続できません。
- VSL ポート チャネルの 10-Gbps ポートのバンドルは、未サポートの QoS 機能が事前設定されている場合に失敗します。すべての QoS 設定は、VSL EtherChannel へポートをバンドルする前に削除される必要があります。

（注）　WS-X6708 モジュールは、10-Gbps ポートの 1 つが VSL EtherChannel にバンドルされている場合でも、非 VSL 構成ポートの独立した QoS ポリシーをサポートしています。

トラフィックの優先順位付けと VSL での負荷分散

ここでの内容は、次のとおりです。

- 「VSL との負荷分散」（P.2-17）

特定のトラフィック タイプに対する優先順位付け

VSL リンクは、潜在的に 3 種類のトラフィックを搬送することが可能で、これらを識別するために QoS マッピングを使用できます。VSL リンクで搬送されるトラフィックのタイプは次のとおりです。

- ユーザ データ トラフィック
- ネットワーク コントロール プレーン トラフィック
- VSS スイッチ間通信およびリンク単位のレイヤ 2 制御トラフィック

それぞれについては、以降のセクションで簡単に説明します。

ユーザ データ トラフィック

推奨されるベストプラクティス コンフィギュレーション実装では、すべてのデバイスが MEC ベース接続を介して VSS に接続されます（「MEC の設定」（P.2-44）を参照）。デュアルホームの MEC 接続では、パススルーのユーザ データ トラフィックは VSL リンクを経由しません。ただし、特定の条件の元では、ユーザ データは VSL リンクを経由する必要があり、適用可能な条件については、次のとおりです（がこれに限定されません）。

- ダウンストリーム トラフィックが VSL リンクを流れることになる、アクセス レイヤから VSS へのアップリンク障害
- ある VSS スイッチ メンバから別のメンバへのリモート Switched Port Analyzer（SPAN; スイッチド ポート アナライザ）トラフィック
- FWSM、Wireless Services Module（WiSM; ワイヤレス サービス モジュール）、Intrusion Detection System（IDS; 侵入検知システム）、およびその他のモジュールからのサービスモジュール トラフィック フロー

VSL 自体は、ユーザ データ トラフィックの QoS マーキングを変更しません。CoS ベースのキューイングを使用して前述のトラフィック タイプを分類しているだけです。結果として、CoS に基づいていない入トラフィック QoS マーキングは、インターナル QoS マッピング テーブルを使用して、CoS ベースのキューイングの変換を行う必要があります。802.1p CoS 5、DSCP 46、または IPP 5 でマーキングされているエンドユーザ アプリケーション トラフィックは、プライオリティ キューに配置されます。

ネットワーク コントロール プレーン トラフィック

アクティブ スイッチは常に関与している隣接デバイスからのネットワーク コントロール プレーン トラフィックを発信し、終端します。アクティブ スイッチは、常にローカルに接続されたインターフェイス（リンク）を使用してコントロール プレーン トラフィックを転送します。アクティブ スイッチに接続されたローカル リンクの障害またはホットスタンバイ スイッチによって発信されたトラフィックによって、VSL を経由する必要のあるネットワーク コントロール プレーン トラフィックには、次のものがあります。

- *Layer-3 Equal Cost Multipath（ECMP; 等コスト マルチパス）リンクまたはホットスタンバイ スイッチのレイヤ 3 プロトコル トラフィック*：hello、更新、データベースなどのルーティング プロトコル制御トラフィック
- *VSS スーパーバイザを対象としたトラフィック*：他のデバイスからの Internet Control Message Protocol（ICMP; インターネット制御メッセージ プロトコル）応答、ホップ カウント単位で値 1 の Time-to-Live（TTL; 存続可能時間）（アクティブ スイッチで終端する必要あり）、SNMP、Telnet/SSH など

VSS スイッチ間通信および Layer-2 リンク単位の制御トラフィック

VSS メンバ スイッチ間のすべての通信は、スイッチ間トラフィックとして定義されます。VSS システムは、次のスイッチ間のコントロール プレーン プロトコルを自動的に Bridge Protocol Data Unit（BPDU; ブリッジ プロトコル データ ユニット）タイプ トラフィックとして分類します。

- スイッチ間通信
  - *シャーシ間 Ethernet Out Band Channel（EOBC; イーサネット アウト バウンド チャネル）トラフィック*：Serial Communication Protocol（SCP; シリアル通信プロトコル）、IPC、および ICC
  - *Virtual Switch Link Protocol（VSLP; 仮想スイッチ リンク プロトコル）*：LMP および RRP 制御リンク パケット
- *リンク単位のレイヤ 2 プロトコル*：Spanning Tree Protocol（STP; スパニング ツリー プロトコル）、BPDU、ポート集約プロトコル（PagP）+、LACP、CDP、および Unidirectional Link Detection（UDLD; 単方向リンク検出）、Link Layer Discovery Protocol（LLDP）、Root Link Query（RLQ）、イーサネット Operations, Administration, and Maintenance（OAM; 運用管理および保守）、802.1x、Dynamic Trunking Protocol（DTP; ダイナミック トランキング プロトコル）など

これらの BPDU パケットは、他のトラフィックの前にある送信プライオリティ キューに自動的に配置されます。

（注） ネットワーク コントロール プレーン トラフィック（レイヤ 2 およびレイヤ 3）は常にアクティブ スイッチに接続されたリンクに送信されます。ローカル リンクが使用できない、またはプロトコル フレームがホットスタンバイ ポートから発信される必要がある場合（PAgP、LACP、または UDLD など）だけ、VSL を通過します。

（注） プライオリティ キューは、コントロール プレーン トラフィックとともに高優先順位のユーザ データ トラフィック（緊急フォワーディング（expedited forwarding）[EF] としてマーク）によって共有されます。ただし、内部メカニズムでは常にコントロール プレーン トラフィックが他のプライオリティ キューイングされたどのトラフィックよりも優先されるようになっています。これにより、ユーザ データが不用意に VSS ドメイン全体の動作の安定性に影響を与えないことが保証されます。

VSL との負荷分散

前のセクションで説明したように、VSL は複数のタイプのトラフィックを VSL バンドルで搬送します。トラフィックの負荷分散の観点から、VSL バンドルはちょうど他の EtherChannel と類似しています。指定の Cisco IOS ソフトウェアで使用可能な EtherChannel ハッシュ アルゴリズムと同じルールに従います。スタンドアロンまたは仮想スイッチ モードでは、単一の EtherChannel ロード バランス ハッシュがシステム全体に適用可能です。つまり、VSL バンドルでは、VSS 内のすべての EtherChannel グループに適用されるのと同じ設定負荷分散モードが使用されています。次の出力例は、ロード バランス オプションを示しています。

```
6500-VSS(config)# port-channel load-balance ?
  dst-ipDst IP Addr
  dst-mac                 Dst Mac Addr
  dst-mixed-ip-port       Dst IP Addr and TCP/UDP Port
  dst-port                Dst TCP/UDP Port
  mpls                    Load Balancing for MPLS packets
  src-dst-ip              Src XOR Dst IP Addr
  src-dst-mac             Src XOR Dst Mac Addr
  src-dst-mixed-ip-port   Src XOR Dst IP Addr and TCP/UDP Port
```

```
src-dst-port            Src XOR Dst TCP/UDP Port
src-ip                  Src IP Addr
src-mac                 Src Mac Addr
src-mixed-ip-port       Src IP Addr and TCP/UDP Port
src-port                Src TCP/UDP Port
```

（注）　このセクションでは、VSL に関連した EtherChannel の特性だけを取り上げています。一般的な EtherChannel の設計と推奨事項については、「MEC の設定」（P.2-44）を参照してください。

実装されているロード バランス手法は、すべてのネットワーク制御トラフィックおよびユーザ データ トラフィックに適用されます。このルールの例外は、スイッチ間通信（常に制御リンクで搬送）および LMP hello（すべての VSL リンクで送信）だけです。ネットワーク コントロール パネルおよびユーザ データ トラフィックは、発信元および/または宛先 MAC アドレスおよび/または IP アドレスをトラフィックの負荷をするのに使用されるハッシュ計算の入力として使用します。VSL リンク間の負荷分散を行うことのできるネットワーク コントロール プレーン トラフィックには、次のトラフィックがありますが、これに限定されません（トラフィックの QoS 分類の違いに注意してください）。

- *レイヤ 2 プロトコル トラフィック*：ブロードキャスト、STP BPDU、PAgP+、LACP、CDP および UDLD、LLDP、RLQ、Ethernet OAM、802.1x、および DTP
- *レイヤ 3 ECMP リンクまたはホットスタンバイ スイッチのレイヤ 3 プロトコル トラフィック*：ルーティング プロトコル制御トラフィック（hello、更新、データベースなど）および ICMP 応答
- *VSS スーパーバイザ指定トラフィック*：他のデバイスからの ICMP 応答、TTL with 1 など

VSL リンクを通過するユーザ データ トラフィックのタイプは、「特定のトラフィック タイプに対する優先順位付け」（P.2-16）セクションで説明しています。

ハッシュ方式：Fixed vs Adaptive

ポートチャネルを通過するトラフィックは、各ポートチャネルのメンバ リンクに対して Result Based Hash （RBH）計算に基づいて分散されます。ポートチャネル メンバ リンクがグループに追加またはグループから削除されるたびに、グループ内のすべてのリンクに対して RBH が再計算される必要があります。短期間の場合、すべてのフローが再ハッシュされるため、トラフィックが切断されてしまいます。このハッシュ実装のことを *Fixed* と呼びます。

Cisco IOS Release 12.2(33) SXH より、Cisco Catalyst 6500 は各ポートチャネル メンバ リンクのハッシュをあらかじめ計算する拡張型ハッシュ アルゴリズムをサポートしています。リンク メンバに障害が発生した場合、ハッシュの動的な事前計算により、新規フローを既存リンクに追加することができるようになるため、すでにリンクにハッシュされているフローのパケット消失を減らすことができます。この拡張されたハッシュ実装は *Adaptive* と呼ばれます。次の例は、ハッシュ分散オプションを示したものです。

```
6500-VSS(config-if)# port-channel port hash-distribution ?
  adaptive  selective distribution of the bndl_hash among port-channel members
  fixed     fixed distribution of the bndl_hash among port-channel members

VSS(config-if)# port-channel port hash-distribution fixed
This command will take effect upon a member link UP/DOWN/ADDITION/DELETION event.
Please do a shut/no shut to take immediate effect
```

デフォルトでは、すべての非 VSL EtherChannel における負荷分散ハッシュ方式は Fixed です。対照的に、重要なスイッチ間コントロール プレーン トラフィックを搬送するため、VSL バンドルのデフォルトのハッシュ アルゴリズムは Adaptive です。デフォルトのハッシュ方式を変更することはできますが、ハッシュ アルゴリズムを有効にする唯一の方法は、リンクをリセットすることです。これを VSL

に適用すると、VSL リンクがバウンスすると両方のシャーシが互いに切断されるため、デュアルアクティブ状態がトリガされます（「VSS デュアルアクティブ スーパーバイザを使用したキャンパス復旧」（P.4-19）を参照）。

ヒント　VSL リンクの負荷分散ハッシュ方式をデフォルト（Adaptive）のままにすることをお勧めします。この方式はリンク障害からフローを復元する場合に効率的です。

現在の EtherChannel ハードウェアは、3 つの固有のバイナリ バケットだけで負荷分散できるため、すべてのバケットを満たすことのできる EtherChannel バンドルの任意の組み合せは、バンドル内のすべてのリンクを最適に使用することになります。これは、最適な負荷分散に対して、2 の倍数の数式を使用して、バンドル内のリンク数に変換されます。

ヒント　トラフィック フローの負荷分散を最適化するため、VSL ポートチャネルのリンク数は常に 2 の倍数（2、4、および 8）にバンドルします。

レジリエント VSL デザインの考慮事項

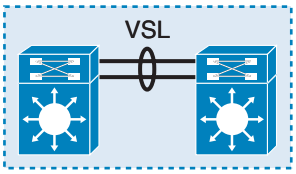
VSL は単一メンバ EtherChannel として設定できます。レジリエント VSL リンクの設定は、レジリエント EtherChannel 接続デバイスのデプロイに適用される設計原則と同じ原則に従っています。レジリエント EtherChannel デザインは、ラインカードまたはポートに関するシングル ポイント障害の回避で構成されています。デュアルアクティブ状態と VSS の不安定性を避けるために VSL の冗長性は重要です。スーパーバイザに障害が発生したかどうかに関係なく、VSL 冗長性は有益です。アクティブ スーパーバイザの障害の場合、VSL レジリエンシーに関係なく、ホットスタンバイ スイッチ（スーパーバイザ）が引き継ぐ準備ができています。スーパーバイザで障害が発生していない場合にシステムで VSS リンクの消失が発生し、デュアルアクティブ状態となるような場合に、VSL レジリエンシーが重要です。

レジリエント VSL を設計する場合次の重要な要素を考慮する必要があります。

- バンドル内の複数のメンバがポートチャネルを使用して、潜在的なシングル ポイント障害（ポート、ラインカード）を減らすようにします。
- 冗長ハードウェアを使用します（ポート、ラインカード、およびポートに接続している内部リソース）。
- 各 VSL リンクに対して多様なファイバ パスを使用します（別のコンジット（導管）、ファイバ終端、および物理パス）。
- VSL リンクで転送されるトラフィックを管理して、シングル ホーム デバイスを回避します。これについては、「VSS 対応キャンパスのトラフィック フロー」（P.3-5）で取り上げています。
- VSL は 10-Gbps ポートでだけ設定できるため、VSL バンドルのデプロイは Sup720-10G、WS-X6708、または WS-X6716 ハードウェアに限定されています。冗長性の他に、キャパシティ プランニングも VSL バンドルごとの VSL メンバ数に影響します。キャパシティ プランニングについては、「VSL バンドルのキャパシティ プランニング」（P.3-12）で説明しています。シングル ポイント障害の回避について、3 つの設計オプションがあります。
  - *Sup720-10G スーパーバイザで 2 つの 10-Gbps ポートを使用する*

    設計オプション 1（図 2-10）は、最も一般的で最も直感的な選択です。ここでは、スーパーバイザ上の両方の 10-Gbps ポートを使用します。しかし、両方のポートが単一の内部ファブリック接続に接続されるため、このオプションは最適なハードウェアの多様性を提供しません。両方のポートをファブリック バックプレーンに接続した場合のハードウェア エラーの確率は低くなりますが、エラーが発生しないわけではありません。

図 2-10　スーパーバイザ ポートを介した VSL

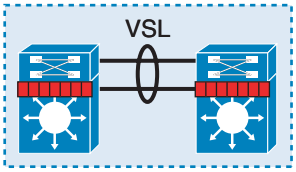


設計 1：スーパーバイザ ポートを介した VSL バンドル　226927

- *Sup720-10G スーパーバイザから 1 つの 10-Gbps ポートを使用し、もう 1 つのポートは VSL 対応ラインカード（WS-X6708 または WS-X6716）のポートを使用する*

  設計オプション 2（図 2-11）は、VSL リンクを 2 つの別のハードウェア ラインカードに分散させて、Sup720-10G アップリンクから 1 ポート、WS-X6708 ラインカードからもう 1 つのポートを使用します。これが最良のベースラインで、コストと冗長性が分散する最も実用的なオプションです。この設計は、CoS ベース キューイングを使用する Sup720-10G 上の未使用ポートに制限されます。Cisco IOS 12.2(33) SXI では、この制限が撤廃されました。

図 2-11　バンドルと 67xx ラインカード間の分散 VSL バンドル

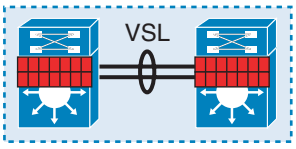


設計 2：スーパーバイザと 67xx ラインカード間の分散 VSL バンドル　226928

- 両方の 10-Gbps ポートを VSL 対応ラインカード（WS-X6708 または WS-X6716）から使用する

  設計オプション 3（図 2-12）は、VSL 接続用に別のラインカードを使用します。これは、スーパーバイザ上のアップリンク ポートの QoS 設定の点から最良の柔軟性を備えていますが、設計オプション 2 と同等の費用対効果はありません。

図 2-12　デュアル 67xx カード間の分散 VSL バンドル



設計 3：デュアル 67xx ラインカード間の分散 VSL バンドル　226929

シングル ポイント障害を回避することの他に、適切な設計を選択することは、ハードウェアの能力とスーパーバイザ上のアップリンク ポートの使用状況によって左右されます（「VSL QoS およびトラフィックの優先順位付け」（P.2-12）を参照）。1 つの 10-Gbps アップリンク ポートが VSL リンクとして使用される場合、もう 1 つの 10-Gbps ポートには CoS ベースのキューイングだけがあります。Cisco IOS 12.2(33) SXI より Cisco IOS ソフトウェアは CoS ベースのキューイング制限を撤廃し、他の非 VSL 10-Gbps ポートを DSCP ベース キューイングに設定できるようになりました。Sup720-10G のいずれかのポートが、QoS 要件に柔軟性が必要なコアまたは他のネットワーク レイヤの接続に使用される場合、設計オプション 2 および 3 のいずれかを選択できます。

最高のレジリエンシーを得るためには、1 つの VSL リンクをスーパーバイザ上に、もう 1 つをラインカード上にするのが最良のオプションです。この設計オプションは、ラインカードの突発的な切断の可能性を避けて、今後の拡大に備えて VSL バンドル内により多くのポートを追加することができるようになります。表 2-2 は、VSL リンクの設計選択を詳細に説明したものです。あらゆる実装について次のオプションから明確な選択を行うことができるわけでありませんが、表 2-2 では、可能な選択についていくつかのリスクで分類していることに注意してください。

表 2-2　VSL リンクの設計選択

	設計 1-1 デフォルトの非 10g-only	設計 1-2 10g-only（mls qos 10g-only）	設計 2-1 デフォルトの非 10g-only	設計 2-2 10g-only（mls qos 10g-only）	設計 –3
ハードウェア構成	両方の Sup 720-10G アップリンクに VSL リンク。すべてのアップリンク ポートが使用可能。	両方の Sup 720-10G アップリンクに VSL リンク。10-Gbps ポートだけが使用可能。	Sup 720 上の 1 ポートと 10-Gbps ラインカード上の 1 ポート。すべてのアップリンク ポートが使用可能。	Sup 720 上の 1 ポートと他のラインカード上の 1 ポート。10-Gbps ポートだけが使用可能。	別の 10-Gbps ラインカード上に両方の VSL リンク。
ポートの使用状況とパフォーマンス モード	すべてのアップリンク ポートが使用可能。10-Gbps および 1 Gbps が合計 20 Gbps の帯域幅を共有。	10-Gbps ポートだけが使用可能。10-Gbps ポートが非ブロック。	すべてのアップリンク ポートが使用可能。10-Gbps および 1 Gbps が合計 20 Gbps の帯域幅を共有。	10-Gbps ポートだけが使用可能。10-Gbps ポートが非ブロック。	すべてのスーパーバイザ アップリンク ポートが使用可能。 WS-X6708–2:1 オーバーサブスクリプション。
QoS[1] 設定の柔軟性	最小：すべてのポートは CoS ベース キューイングだけ可能。	すべての 10-Gbps ポートが VSL リンクとして使用されているため、CoS ベースだけ。ただし 3 つの 1-Gbps ポートが失われます。	制限されるものの、実用的。すべてのポートは CoS ベース キューイングだけ可能。10-Gbps ラインカードは独立 QoS を持てます。	オプション 2-1 よりも制限。残りの 10-Gbps ポートは VSL ポートと同じ QoS 設定に従います。12.2(33) SXI でこの制限は撤廃されました。 10-Gbps ラインカードは独立 QoS を持てます。	最も柔軟 すべてのスーパーバイザ アップリンクを使用可能。10 Gbps の全ポートが独立した QoS を 10g-only モードで持つことが可能。
VSL シングル ポイント障害	可能性あり。ファブリック相互接続の障害のため。障害のリスク：非常に低い。	可能性あり。ファブリック相互接続の障害のため。障害のリスク：非常に低い。	別個のハードウェアで両ポートが失われる可能性はまれ。	別個のハードウェアで両ポートが失われる可能性はリモート。障害のリスク：非常にまれ。	可能性あり。過酷な状況でラインカードの接続が失われるため。リスク：きわめて低い。

表 2-2　VSL リンクの設計選択 （続き）

VSS ブート動作[2]	最適	最適	最適	最適	遅い
全体的な選択基準	費用対効果が高い、効率的、1 Gbps アップリンクで QoS 設定の柔軟性が欠如 ハードウェアの多様性に欠けるため推奨せず。	費用対効果が高い、パフォーマンスの保証。ハードウェアの多様性に欠けるため推奨せず。	実用的 今後の拡張により、コスト、効率、およびリンクの冗長性が全体的に最良。ただし QoS の柔軟性が欠如。	実用的 今後の拡張により、コスト、効率、およびリンクの冗長性が全体的に最良。	最も効率的なポート使用方法と柔軟な QoS。費用対効果は高くなく、設計オプション 1-1 および 1-2 に最適化。

1. マッピングおよび割り当てが同じままであるため、キュー構造と深度が VSL の要因ではありません。「VSL QoS およびトラフィックの優先順位付け」（P.2-12）を参照してください。
2. ラインカードでイメージのダウンロードや初期化に時間がかかるため、Sup720-10G ポートの VSL リンクはラインカードの VSL よりも早く立ち上がります。VSS は、スーパーバイザ ポートで VSL リンクの迅速なブートに最適化されています。

VSL 動作モニタリング

このデザイン ガイドでは、VSS の動作モニタリングおよびトラブルシューティングについては取り上げませんが、帯域幅の使用状況を管理するニーズと、VSL ポートチャネルおよびそのメンバ リンクの健全性を強調するための重要な情報は含まれています。このセクション全体を通じて、関連 CLI 出力の例を示します。

VSS および VSL ポートチャネル インターフェイスのトラブルシューティングには、ネットワーク パケット デコーダが接続されているポートにポートチャネルをスパンする必要がある場合もあります。VSL ポートチャネルをスパンすることは可能です。ただし、ローカルのポートチャネルをローカルの宛先にスパンすることだけが可能です。次の CLI コマンド出力を参照してください。

```
6500-VSS# show inteface vsl

VSL Port-channel: Po1
 Port: Te1/5/4
 Port: Te1/5/5
VSL Port-channel: Po2
 Port: Te2/5/4
 Port: Te2/5/5

6500-VSS(config)# monitor session 2 source int po1
6500-VSS(config)# monitor session 2 destination int gi1/4/10

6500-VSS# show monitor session 2
Session 2
---------
Type                   : Local Session
Source Ports           :
    Both               : Po1
Destination Ports      : Gi1/4/10

Egress SPAN Replication State:
Operational mode       : Centralized
```

```
Configured mode        : Centralized (default)
```

この出力で示しているように、ポートチャネルがローカルであるスイッチにスパニングすることにより VSL ポートチャネルを監視することができます。ピア（リモート）スイッチに属する宛先を使用してポート モニタリングを作成しようとしていることを示す次の出力例を参照してください。この制限は、トラフィック ループの可能性を取り除き、VSL リンクの過剰な使用状況を回避します。ポートチャネル インターフェイス番号は、通常スイッチ番号 ID と一致するように作成されるため、ポートチャネル インターフェイスとスイッチ番号とのアフィニティを簡単に特定することができます。

```
6500-VSS# show interface vsl
VSL Port-channel: Po1
 Port: Te1/5/4
 Port: Te1/5/5

VSL Port-channel: Po2
 Port: Te2/5/4
 Port: Te2/5/5

6500-VSS(config)# monitor sess 1 source int po2
6500-VSS(config)# monitor sess 1 destination int gi1/4/10
% VSL cannot be monitor source with destination on different core
```

VSL インターフェイスの *ジャイアント* カウンタ（次の出力を参照）が、何か問題があるのではないかという想定につながる可能性があることに注意してください。実際には、これは通常の出力です。インターフェイス カウンタがジャイアントを通知する理由は、アクティブ スイッチとホットスタンバイ スイッチとの間で VSL スイッチ間制御フレーム パケットが 1518 バイト + 32 バイトの DBUS ヘッダーで送信されるという事実からです。このようなサイズ超過パケットは、VSL EtherChannel ではジャイアントとして見られます。

```
6500-VSS# show switch virtual link counters
.
.
! <snip>

Port          Single-Col Multi-Col  Late-Col Excess-Col Carri-Sen     Runs      Giants
Po1                    0         0         0          0         0        0    19788377
Te1/2/8                0         0         0          0         0        0          34
Te1/5/4                0         0         0          0         0        0    19788414
<snip>

Port          Single-Col Multi-Col  Late-Col Excess-Col Carri-Sen     Runs      Giants
Po2                    0         0         0          0         0        0      693910
Te2/2/8                0         0         0          0         0        0          89
Te2/5/4                0         0         0          0         0        0      693821
```

ステートフル スイッチオーバー：統合コントロール プレーンおよび分散データ転送

ステートフル スイッチオーバー テクノロジー

ステートフル スイッチオーバー（SSO）テクノロジーにより、スタンドアロン Cisco Catalyst 6000 シリーズ プラットフォームでスーパーバイザ冗長性が使用可能になります。SSO は、必要なコントロール プレーンとバックアップ スーパーバイザに複製されるプロトコルの状態を保持します。結果として、アクティブ スーパーバイザに障害が発生した場合でも、システムおよびネットワークがパケットのフォワーディングを継続し、ネットワーク プロトコルで残りのネットワーク デバイスに関与し続けるための十分な情報が、ホットスタンバイ スーパーバイザに存在することになります。デュアル スーパーバイザ対応システムは、起動時にさまざまな状態になります。初期化中に、Cisco IOS がシステム

にデュアル スーパーバイザがあるかどうかを判別し、ハードウェア モード（*シンプレックス*（単一スーパーバイザ）または*デュプレックス*（デュアル スーパーバイザ））を判別し、どのスーパーバイザがアクティブ ロールまたはホットスタンバイ ロールを担っているのかを特定します。また Cisco IOS ソフトウェアは、各スーパーバイザのソフトウェア バージョンをチェックして、スーパーバイザの状態を SSO または RPR モードにします。各スーパーバイザは、採用されたロール（アクティブまたはスタンバイ）に応じて、表 2-3 で説明しているような Redundancy Facility（RF）状態に従います。*プライマリ*として選択されたスーパーバイザの場合、SSO 起動に成功後、スーパーバイザは最低（ディセーブル）から最高（アクティブ）モードへ移行します。ホットスタンバイ スーパーバイザでは、表 2-3 の「*スタンバイホットになる場合の状態*」で説明しているように、別の状態遷移が進行します。

表 2-3　　RF 状態

RF 状態とコード	RF 状態アクティビティ
両スーパーバイザの共通状態	
RF_UNKNOWN = 0,	不明な冗長状態（たとえばスーパーバイザのブーティング）
RF_DISABLED = 1,	冗長がディセーブル（たとえばデュアル スーパーバイザが存在しない）
RF_INITIALIZATION = 2,	スーパーバイザ間の同期の第 1 段階
RF_NEGOTIATION = 3,	ディスカバリ モードと、誰がアクティブまたはホットスタンバイになるか
スタンバイホットになる際の状態	
RF_STANDBY_COLD = 4,	非アクティブ スーパーバイザの状態、ピアがアクティブ、RPR 状態
RF_STANDBY_CONFIG = 5,	アクティブからホットスタンバイへの同期設定
RF_STANDBY_FILESYS = 6,	アクティブからホットスタンバイへのファイル システム同期
RF_STANDBY_BULK = 7,	プロトコル（クライアント）状態：アクティブからホットスタンバイへのバルク同期
RF_STANDBY_HOT = 8,	アクティブへの準備が完了したスタンバイと、アクティブからの更新の取得
ACTIVE になる際の状態	
RF_ACTIVE_FAST = 9,	ホットスタンバイがアクティブになることの即時通知
RF_ACTIVE_DRAIN = 10,	クライアント クリーンアップ：ピアからのドレイン キュー メッセージ
RF_ACTIVE_PRECONFIG = 11,	事前処理設定、ブート環境
RF_ACTIVE_POSTCONFIG = 12	設定の事後処理
RF_ACTIVE = 13,	コントロール プレーンおよびデータ プレーンがアクティブで、ネットワークに関与

これらの 13 の状態の中で、*13-Active* および *8-Standby-Hot* が動作冗長性を決定するのに重要です。これらについては、次の概説で要約します。

- *状態 13-ACTIVE*：このアクティブ状態では、スーパーバイザは、パケットのフォワーディングとコントロール プレーンの管理を担います。コントロール プレーン機能には、レイヤ 3 ルーティング プロトコル、レイヤ 2 プロトコル（STP、BPDU）、管理（Telnet、Simple Network Management Protocol（SNMP; 簡易ネットワーク管理プロトコル）、Secure Shell（SSH; セキュアシェル）など）、リンクおよびデバイス管理（SPAN および CDP）などの処理が含まれています。アクティブ スーパーバイザは、セカンダリ スーパーバイザと設定を同期します。最後に、ホット

スタンバイ スーパーバイザが *Standby-HOT*（ホットスタンバイ）の状態になるたびに、アクティブ スーパーバイザが、プロトコルの状態とデータベースについて、セカンダリ スーパーバイザと同期します。

- *状態 8-Standby-Hot*：このホットスタンバイ状態で、スーパーバイザがアクティブ スーパーバイザと完全に同期し、必要な場合にアクティブ ロールを担うことができます。これが、ホットスタンバイ スーパーバイザの最後の状態です。この状態では、関連イベント（インターフェイスの状態変更、MAC 更新、変更、アップ、およびダウンなど）に基づく各 SSO 対応プロトコルが、アクティブ スーパーバイザからホットスタンバイ スーパーバイザへのメッセージをトリガーします。何らかの理由でプライマリ アクティブ スーパーバイザに障害が発生するたびに、ホットスタンバイ スーパーバイザのプロトコル状態が実行（動作）状態になります。たとえば、Cisco Express Forwarding（CEF）が SSO 対応クライアントです。CEF のテーブルで変更が発生するたびに、ホットスタンバイ スーパーバイザが更新を受信します。これは、ホットスタンバイ ユニットがアクティブになる際に、Forwarding Information Base（FIB; 転送情報ベース）の更新されたコピーがハードウェア内のデータ パケットを転送することができ、コントロール プレーンは回復プロセスを実行します。SSO の詳細については、次の URL を参照してください。
http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SXF/native/configuration/guide/nsfsso.html

VSS での SSO 動作

SSO は、VSS ハイ アベイラビリティを有効にするためのコア能力です。SSO 動作および機能のサポートは、デュアル スーパーバイザで動作するスタンドアロン ノードに類似しています。主な違いは次の 2 つです。

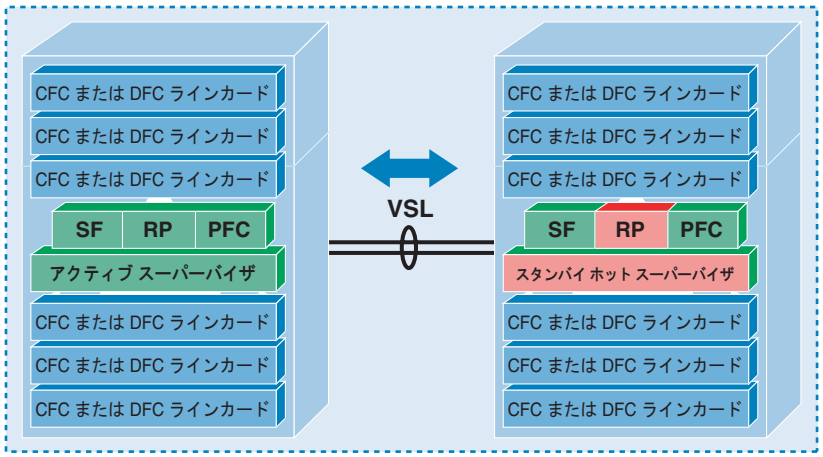
- SSO 動作は 2 つのシャーシに拡大しますが、ここで 1 つのスーパーバイザがアクティブ スーパーバイザとして選択され、他のシャーシはホットスタンバイとして指定されます。この機能は、*シャーシ間 SSO* として定義されます。図 2-13 を参照してください。
- 両方のシャーシとスーパーバイザでパケット転送が発生するため、VSS は*デュアル フォワーディング*ソリューションとなりますが、コントロール プレーンは 1 つのスーパーバイザだけで制御されています。

VSS の SSO 動作には、スタンドアロン環境と同じ依存性があります。シャーシ間 SSO モードでは、両方のメンバ シャーシで同じハードウェアと Cisco IOS ソフトウェアが必要です。SSO 冗長モードの依存性の詳細なリストについては、次の URL を参照してください。

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/configuration/guide/vss.html#wp1059586

（注） Cisco IOS 122.(33)SXI では、Cisco IOS ソフトウェア バージョニングの依存性を除去しています。詳細については、該当するリリース ノートを参照してください。

図 2-13　シャーシ間 SSO 動作



統合コントロール プレーン

図 2-13 で示すとおりに、1 つのスーパーバイザがアクティブに統合コントロール プレーンを提供します。アクティブ スーパーバイザを搭載するシャーシは、*アクティブ仮想スイッチ*と呼ばれます。アクティブ スイッチでは、Switch Fabric（SF; スイッチ ファブリック）、Route Processor（RP; ルート プロセッサ）、および Policy Forwarding Card（PFC; ポリシー フォワーディング カード）の 3 つすべてのコンポーネントがアクティブ（緑）になります。アクティブ スーパーバイザは、VSS 全体を通じて、すべての分散型フォワーディング カード（DFC）にハードウェアのフォワーディング情報をプログラミングする機能も果たしています。また、ホットスタンバイ仮想スイッチ スーパーバイザ エンジンでの policy feature card（PFC; ポリシー フィーチャ カード）のプログラミングも担当します。ユニファイド コントロール プレーンは、「ネットワーク コントロール プレーン トラフィック」（P.2-16）で説明したトラフィック タイプの発信と終端を担当し、また、「VSS スイッチ間通信および Layer-2 リンク単位の制御トラフィック」（P.2-17）で説明したように、スイッチ間通信の維持と管理を担当する唯一のコントロールポイントとしても機能します。

（注）　VSS の最初のリリースでは、物理シャーシごとに 1 スーパーバイザだけがサポートされます。このため、シャーシ間スーパーバイザ エンジンの冗長性はありません。今後のソフトウェア リリースで、2 番目のスーパーバイザ エンジンを各シャーシに追加する機能が提供される可能性があります。

分散型データ転送

図 2-13 で示したように、スーパーバイザ リソース（SF および PFC）は両方とも、ユーザデータ フォワーディングについてアクティブです。両スーパーバイザのポリシー フィーチャ カード（PFC）およびスイッチング ファブリック（ファブリック対応モジュールのバックプレーンの接続性）は、ユーザデータをアクティブに転送して、アクセス コントロール リスト（ACL）の適用やハードウェアでの QoS の実施といったポリシー機能を実行します。これに加えて、すべての分散型フォワーディング カード（DFC）は、VSS 全体でパケットの同時ルックアップを実行できます。両スイッチのスイッチング ファブリックもまたアクティブ ステートのため、Cisco VSS は、1440（720 Mbps x 2）Gbps または全体で 1.44 Tbps のスイッチ ファブリック機能を持ちます。

同期モードで実行するアクティブまたはホットスタンバイのスーパーバイザでは、次のシステム情報が VSL リンクで同期されます。

- ブート環境。
- 実行コンフィギュレーションの同期。
- プロトコル ステートおよびデータベース テーブル：SSO の冗長性をサポートできる（SSO 対応型）プロトコルだけが SSO ベースのリカバリを十分にサポートできる機能を備えています。
- ラインカードのステータス（インターフェイス ステート テーブルとその機能）。

初期化フェーズ中、ホットスタンバイ スーパーバイザは、アクティブなスーパーバイザとともに設定の同期（RF_STANDBY_CONFIG = 5）を行います（表 2-3 を参照してください）。この設定の同期について理解しておくことは、「VSS デュアルアクティブ スーパーバイザを使用したキャンパス復旧」（P.4-19）で詳細を説明した、スイッチ障害を検討するうえで重要です。

アクティブ スイッチとホットスタンバイ スイッチはいずれもアドレスを同時に学習できますが、アクティブな仮想スイッチは、隣接するデバイスから取得したネットワーク情報（MAC、STP、または CEF など）を管理します。いくつかのプロトコルは、アクティブ スイッチがプロトコル情報（データベース、プロトコル ステート）をホットスタンバイ スーパーバイザと同期する、SSO に対応した機能を備えています。さらに、アクティブ スーパーバイザは、両シャーシのインターフェイス情報とラインカードのステータスを管理および更新します。

スーパーバイザのコントロール プレーンのステート（アクティブ、ホットスタンバイ、その他）は、下記の CLI コマンドを使用してチェックできます。両シャーシでファブリック ステートがアクティブであり、デュアル フォワーディング ステートを示している点に注目してください。

```
6500-VSS# show switch virtual
Switch mode : Virtual Switch
Virtual switch domain number : 200
Local switch number : 1
Local switch operational role: Virtual Switch Active
Peer switch number : 2
Peer switch operational role : Virtual Switch Standby

6500-VSS# show switch virtual redundancy
My Switch Id = 1
Peer Switch Id = 2
Last switchover reason = none
Configured Redundancy Mode = sso
Operating Redundancy Mode = sso
Switch 1 Slot 5 Processor Information :
------------------------------------------------
Current Software state = ACTIVE
Uptime in current state = 3 weeks, 4 days, 9 minutes
Image Version = Cisco IOS Software, s72033_rp
Software (s72033_rp-ADVENTERPRISEK9_WAN_DBG-M), Version
12.2(SIERRA_INTEG_070502) INTERIM SOFTWARE
Synced to V122_32_8_11, 12.2(32.8.11)SR on rainier, Weekly
12.2(32.8.11)SX76
Technical Support: http://www.cisco.com/techsupport
Copyright (c) 1986-2007 by Cisco Systems, Inc.
Compiled Thu 03-May-07 09:46 by kchristi
BOOT = sup-bootdisk:s72033-
adventerprisek9_wan_dbg-mz.SIERRA_INTEG_070502,1;
CONFIG_FILE =
BOOTLDR =
Configuration register = 0x2102
Fabric State = ACTIVE
Control Plane State = ACTIVE
Switch 2 Slot 5 Processor Information :
------------------------------------------------
Current Software state = STANDBY HOT (switchover target)
```

```
Uptime in current state = 3 weeks, 4 days, 8 minutes
Image Version = Cisco IOS Software, s72033_rp
Software (s72033_rp-ADVENTERPRISEK9_WAN_DBG-M), Version
12.2(SIERRA_INTEG_070502) INTERIM SOFTWARE
Synced to V122_32_8_11, 12.2(32.8.11)SR on rainier, Weekly
12.2(32.8.11)SX76
Technical Support: http://www.cisco.com/techsupport
Copyright (c) 1986-2007 by Cisco Systems, Inc.
Compiled Thu 03-May-07 09:46 by kchristi
BOOT = sup-bootdisk:s72033-
adventerprisek9_wan_dbg-mz.SIERRA_INTEG_070502,1;
CONFIG_FILE =
BOOTLDR =
Configuration register = 0x2102
Fabric State = ACTIVE
Control Plane State = STANDBY
```

（注）　Cisco IOS ソフトウェアのイメージは、両スーパーバイザで一致する必要があります。そうでないと、スタンバイ スーパーバイザが RPR モードでブートし、ラインカードがそのシャーシでアクティブになりません。RPR の詳細については、『http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SXF/native/configuration/guide/redund.html』のマニュアルを参照してください。

仮想スイッチのロール、プライオリティ、およびスイッチ プリエンプション

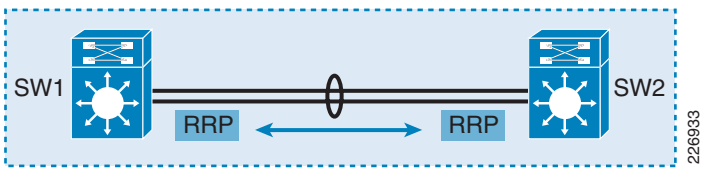
ロール解決プロトコル（RRP）

ユニファイド コントロール プレーンを使用し、複数のスイッチでの分散フォワーディングをサポートする場合、一定の形式のプロトコルが必要になります。このプロトコルによって、いずれのスイッチをアクティブにする必要があるか（アクティブにできるか）、デフォルト設定をどのように変更するか、さらに、いずれのスイッチ メンバを確定的にアクティブに設定するか、などを決定します。VSS には、ロール解決プロトコル（RRP）という専用のプロトコルがあり、これを使用してこのような動作を定義します。図 2-14 を参照してください。

RRP プロトコルは、SSO のロール（アクティブ、ホットスタンバイ、または RPR）の決定、スイッチのプライオリティのネゴシエート、さらには仮想スイッチのプリエンプションに使用します。また、VSS を形成するために同一にする必要がある、ソフトウェア バージョンのチェックも各スイッチで行います。

少なくとも 1 つの VSL ポートでリンク管理プロトコル（LMP）が十分に確立されると、RRP プロトコルはいったん初期化されます。SSO のロールとスイッチのプライオリティをネゴシエートするために、RRP プロトコルによって LMP コントロール リンクが選択されます。それぞれのスイッチ メンバはローカルの RRP ピア グループ インスタンスを形成して、すべての VSL リンクで実行するのではなく VSL バンドルのコントロール リンクを介して通信します。RRP ネゴシエーション パケットは、LMP プロトコルと同一の形式でカプセル化されます。このため、RRP パケットはこれをデータ トラフィックで優先処理するために、転送プライオリティ キューに配置されます。

図 2-14　　ロール選択のための RRP ネゴシエーション



RRP プロトコル ステータスは、図 2-15 に示すコマンドを使用すると確認できます。出力例に示すように、ホットスタンバイ スイッチの RRP ステートを確認するには、**remote command** *switch-id* コマンドが必要です。スイッチ ID、プライオリティ、プリエンプション、または現在の SSO ロールにかかわらず、ローカル スイッチの Peer Group は常に 0 にし、ネイバー スイッチでは常に 1 にします。

図 2-15　RRP プロトコルのステータス情報

```
6500-VSS#show vsl rrp summary
 RRP Summary:
-------------------------------------------------------------------
RRP information for Instance 1

-------------------------------------------------------------------
Valid Flags  Peer    Preferred Reserved
             Count   Peer      Peer

-------------------------------------------------------------------
TRUE  V      1       1         1

        Peer                Switch          Preempt      Priority                 Local  Remote
Switch  Group  Valid  Status Number          Oper(Conf)   Oper(Conf)  Role         SID    SID
-------------------------------------------------------------------------------------------
Local   0      TRUE   UP     1               N (N)        100(100)    ACTIVE       0      0
Remote  1      TRUE   UP     2               N (N)        100(100)    STANDBY      7418   3697

Peer 0 represents the local switch

Flags : V - Valid

6500-VSS#remote command switch-id 2 module 5 show vsl rrp summary

 RRP Summary:
-------------------------------------------------------------------
RRP information for Instance 2

-------------------------------------------------------------------
Valid Flags  Peer    Preferred Reserved
             Count   Peer      Peer

-------------------------------------------------------------------
TRUE  V      1       1         1

        Peer                Switch          Preempt      Priority                 Local  Remote
Switch  Group  Valid  Status Number          Oper(Conf)   Oper(Conf)  Role         SID    SID
-------------------------------------------------------------------------------------------
Local   0      TRUE   UP     2               N (N)        100(100)    STANDBY      0      0
Remote  1      TRUE   UP     1               N (N)        100(100)    ACTIVE       3697   7418

Peer 0 represents the local switch

Flags : V - Valid
```

226931

仮想スイッチ間の RRP セッションは、次の条件についてネゴシエートします。

- 仮想スイッチ ドメイン内の両仮想スイッチがいつ bootup モードになるか
- アクティブ スイッチが使用可能な間、ホットスタンバイ スイッチをいつ bootup プロセスにするか
- デュアルアクティブ リカバリ フェーズで VSL リンクが仮想スイッチ メンバ間でいつ復旧するか

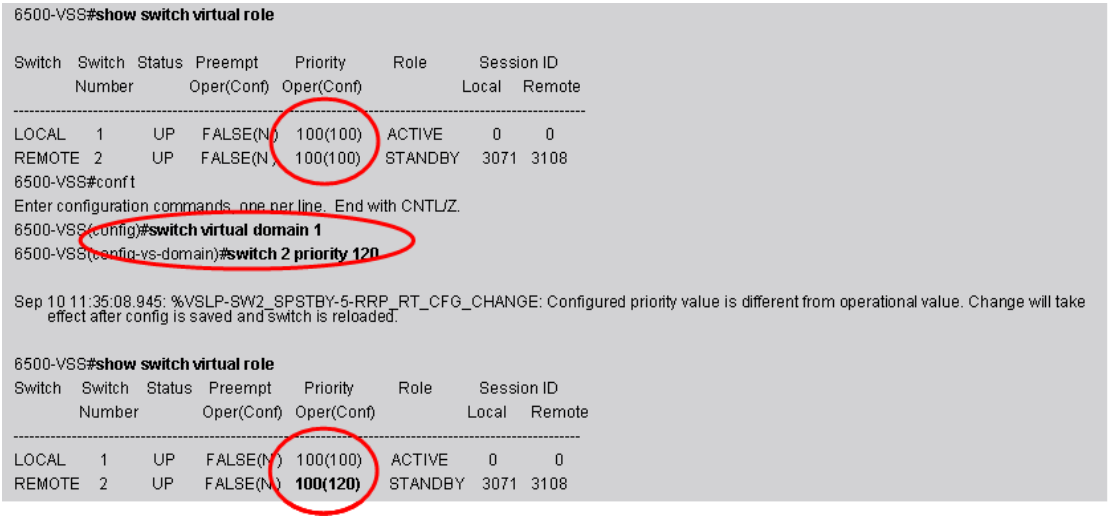
上記の条件が満たされると RRP セッションは一時的に確立されます。ただし、RRP は 2 つのスイッチ メンバ間でセッションを維持せず、RRP は、各スイッチ メンバの LMP からの内部通知に依存します。すべての VSL メンバのリンクに障害が発生した場合、RRP にはピア間の通信を行うコントロールリンク パスがなくなり、各仮想スイッチ上の LMP プロトコルはピア グループを削除して、SSO ベースのスイッチオーバー処理を引き継ぐように RRP プロセスに通知します。VSL リンクがないと、アクティブ スイッチが使用可能であってもこのスイッチで RRP 処理は行われません。ただし、ホットスタンバイ スイッチはアクティブ ロールに移行します。これは、リモートのピアのステータスを判別する方法がないためです。

仮想スイッチのプライオリティ

アクティブまたはホットスタンバイにするスイッチ メンバの選択は、このスイッチが初期化される順序や方法によって異なります。複数のスイッチ メンバが同時にブートした場合、最も下位のスイッチ ID を持つスイッチがアクティブな仮想スイッチになります。各スイッチ メンバが別々にブートした場合は、スイッチ ID にかかわらず最初に初期化したスイッチ メンバがアクティブな仮想スイッチになります。VSS でもスイッチのプライオリティを設定できます。通常、デフォルトのスイッチのプライオリティ（100）を変更する必要はありませんが、次の 2 つの要件が予期される場合は、スイッチのプライオリティを設定できます。

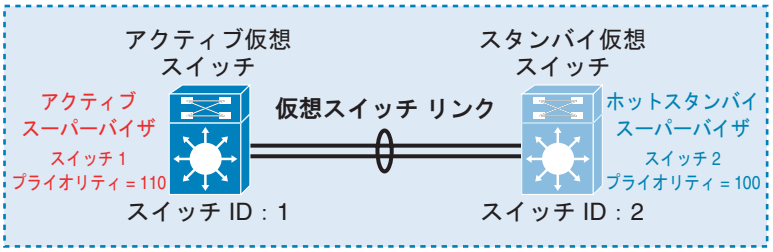
- 初期設定後にスイッチ ロールのデフォルトの選択を上書きする必要がある。これは、柔軟な動作上のニーズまたはモニタリング アクセス要件による場合があります。ただし、ソフトウェアは、設定されたスイッチのプライオリティをすぐには適用しません。この変更が反映されるようにするには、両スイッチをリロードして、RRP が指定のスイッチをアクティブにするための有効化を行う必要があります。図 2-16 に、スイッチのプライオリティの変更に関するコマンドと、新たに指定したプライオリティを適用するにはリロードが必要であることを通知する syslog メッセージを示します。下位のプライオリティのスイッチが最初に来た場合、ソフトウェアは、プリエンプションでスイッチが設定されている場合を除き、より上位のプライオリティのスイッチが後にブートしたときにアクティブ ロールの引き継ぎを強制しません。

図 2-16　スイッチのプライオリティの変更



- スイッチの ID に一致するようにスイッチのプライオリティを定義すると設定に表示されるため、変更を管理する場合にわかりやすくなります。このオプションは、図 2-17 に示すように、スイッチのプライオリティがスイッチの ID に一致するように設定されています（高いプライオリティと最下位のスイッチ ID とを一致させています）。

図 2-17　スイッチのプライオリティをスイッチの ID に一致させる設定



スイッチのプリエンプション

ある一つのスイッチをすべての条件でアクティブなロールに選択する目的がある場合、スイッチのプライオリティを高くするだけではこの目標を達成できません。ロールの選択を確定的に行うには、ブートの順序にかかわらず、より上位のスイッチ ID とスイッチ プリエンプションを目的のアクティブ スイッチに設定する必要があります。図 2-18 を参照してください。この CLI は、プリエンプション機能が下位のプライオリティのスイッチに設定されるようにしています。しかし、スイッチのプリエンプション設定は有効になりません。

図 2-18　プリエンプションの設定

```
6500-VSS#conf t
6500-VSS#(config)#switch virtual domain 1
6500-VSS#(config-vs-domain)#switch 2 priority 120

Sep 15 17:03:24.468: %VSLP-SW2_SPSTBY-5-RRP_RT_CFG_CHANGE: Configured
priority value is different from operational value. Change will take effect after config is
saved and switch is reloaded.

6500-VSS#(config-vs-domain)#switch 2 preempt

Please note that Preempt configuration will make the ACTIVE switch with lower priority to
reload forcefully when preempt timer expires

Sep 15 17:03:30.864: %VSLP-SW2_SPSTBY-5-RRP_RT_CFG_CHANGE: Configured
preempt value is different from operational value(s).
Change will take effect after config is saved and switch is reloaded.

6500-VSS#show vsl rrp summary
 RRP Summary:
-------------------------------------------------------------
RRP information for Instance 1

-------------------------------------------------------------
Valid   Flags  Peer    Preferred    Reserved
               Count   Peer         Peer
-------------------------------------------------------------
TRUE    V      1       1            1

        Peer  Valid    Switch   Status   Preempt      Priority      Role     Local  Remote
Switch  Group          Number            Oper(Conf)   Oper(Conf)            SID    SID
-------------------------------------------------------------------------------------------
Local   0     TRUE     1        UP       N (N )       100(100)      ACTIVE   0      0
Remote  1     TRUE     2        UP       N (Y*)       100(120)      STANDBY  3790   7230

Peer 0 represents the local switch

Flags : V - Valid
```

226935

ネットワークの可用性とスイッチのプリエンプションとの関連は深く、展開の前に評価する必要があります。スイッチのプリエンプションの運用上の影響は、広く知られている HSRP/GLBP プロトコルの動作と比較すべきではありません。このプロトコルでプリエンプションが可能なのは、アクティブな HSRP/GLBP フォワーダとしてのロールを、ネットワークへの影響が少ない（シャーシのリロードやリセットがない）、スタンバイ モードのロールに委任することだけです。

スイッチのプリエンプションにより、複数の VSS メンバが強制的にリブートされ、これによって、数箇所でネットワークが停止し、フォワーディング機能が低下することになり、一方でスイッチはいずれのスーパーバイザがアクティブ ロールを引き継ぐべきかを決定します。たとえば、プリエンプションを設定したスイッチに障害が発生（または障害に至らなくとも接続性が損失）した場合、ピア スイッチがアクティブなロールを一時的に引き継ぎます。プリエンプティブなスイッチがブートし、ピア スイッチがアクティブであることを検出した場合、このスイッチは、新しいアクティブなピア スイッチに対してアクティブなロールを強制的に放棄させます。ピア スイッチにアクティブなロールを放棄させる唯一の方法は、リセットしてホットスタンバイ ロールに移行することです。同様に、非プリエンプティブ（ホットスタンバイが指定されている）スイッチが何らかの形で最初に来た場合に（電源障害またはユーザ処理の遅延のいずれか）、アクティブなロールを引き継ぐときは、プリエンプティブ スイッチがオンラインになると強制的にリセットされます。

ヒント　**シスコでは、次の理由から、スイッチのプリエンプションを*設定しない*ことをお勧めします。**

- 設定すると、複数のスイッチがリセットされ、フォワーディング機能が低下し、予期しないネットワークの停止につながる。
- VSS は単一の論理スイッチまたはルータである。両方のスイッチ メンバは、アクティブなロールを担う機能があるという点で等価です。これは、企業ポリシーで必要とされない限り、どちらがアクティブであるかは関係ないためです。

仮想スイッチ メンバの起動の動作

通常の VSS の起動プロセスは、診断、VSL リンクの初期化、LMP の確立、および RRP を介したスイッチ ロールのネゴシエーションで構成されます。RRP は、SSO ステートになるように各スイッチのロールを決定しますが、ここではいずれかのスイッチがアクティブ、ピアがホットスタンバイ ステートになります。ただし、各スイッチ ロールが割り当てられる RRP のネゴシエーション フェーズ後に、VSL インターフェイスが非アクティブ、無効化、および障害発生につながる問題またはイベントがある場合は、この動作と最終的な結果は異なります。VSL リンクは、主に次のいずれかの理由で無効になるのが一般的です。

- VSL インターフェイスかラインカードのいずれかまたは両方が使用できない。
- アクティブなロールを引き継ぐスイッチに、スイッチのリセットにつながる問題があるために VSL インターフェイスが使用できない。

いずれの場合も、ホットスタンバイのロールを引き継いでいるピア スイッチは、ピア スイッチのステートを判別（または交換）する方法がないことから（VSL が停止しているため）、ブートを続行できません。Cisco IOS ソフトウェアは、強制的なリセットを実行し（Cisco IOS 用語では*クラッシュ*と言います）、ピア スイッチはブート プロセスを再開します。ピア（ホットスタンバイ）スイッチで VSL リンクがまだ停止していることが検出されると、ブートを続行します。RRP がない場合、アクティブなロールを引き継ぎます。この状況は、いずれのスイッチも互いを認識しないため、デュアルアクティブ条件になる場合があります。この理由から、VSS をキャンパス ネットワークに展開する場合、デュアルアクティブの検出は欠かせません。詳細については、「VSS デュアルアクティブ スーパーバイザを使用したキャンパス復旧」（P.4-19）を参照してください。

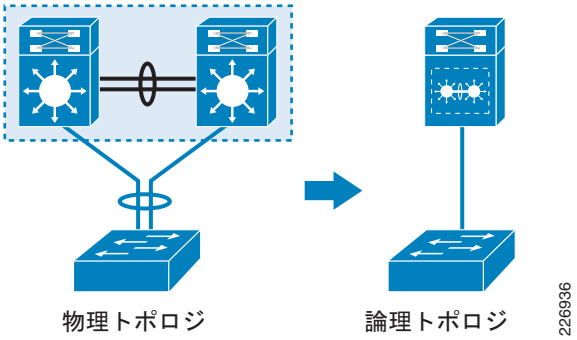
Multi-chassis Etherchannel（MEC）

従来の EtherChannel は、2 つのスイッチ間の複数の物理リンクを集約します。MEC は高度な EtherChannel テクノロジーで、リンク集約を 2 つの個別のスイッチにまたがるように拡張します。VSS で分散フォワーディングとユニファイド コントロール プレーンが可能になることから、MEC はアクティブ スイッチとスタンバイ スイッチの両方に存在する単一のポートチャネル インターフェイスのように見えます。アクセス レイヤが 2 つの物理リンクを介して別個の物理シャーシに接続するにもかかわらず、アクセス レイヤ スイッチ側からは、このポートチャネルの接続で単一の論理スイッチに

接続する単一の論理リンクが有効になります（*MEC 付き VSS* と呼ばれます）。図 2-19 は、物理から論理への変換を示します。ここでは論理トポロジが簡素化され（スパニング ツリー）、MEC 付きの VSS によってループのないトポロジが実現しています。

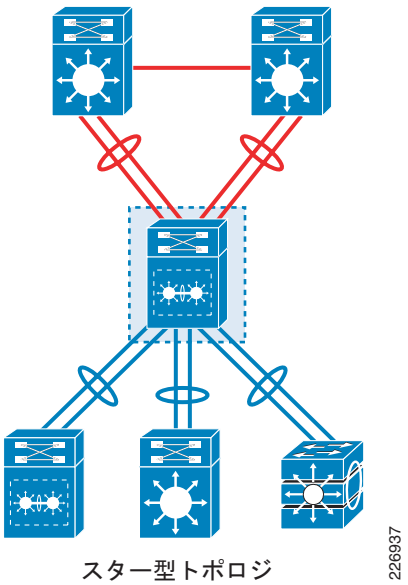
（注）　MEC は VSS でしか設定できませんが、VSS への接続性を必要とするアクセス レイヤ スイッチは、従来の EtherChannel インターフェイスで設定します。

図 2-19　　MEC：物理トポロジと 論理トポロジ



EtherChannel を仮想の単一の論理スイッチとして複数のスイッチにスパニングするこの機能は、すべてのデバイスが MEC を介して VSS に接続されるトポロジを形成し、これはスター型トポロジとして表示されます。図 2-20 を参照してください。

図 2-20　　スター型トポロジ

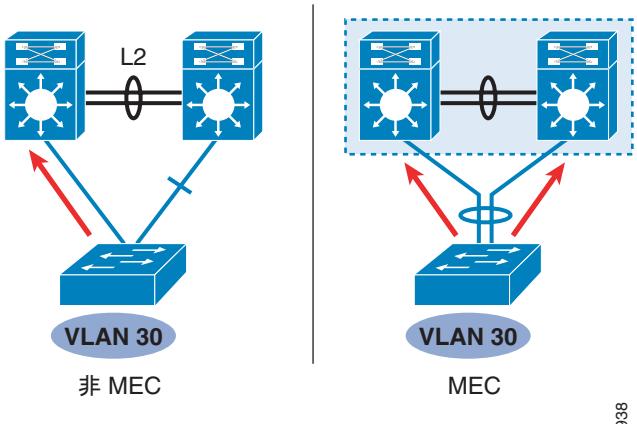


MEC と VSS は、強力で非常に効果的な変革をキャンパス トポロジにもたらします。鍵となる 2 つの利点は次のとおりです。

- *マルチレイヤ設計のループを排除*：複数のクローゼットにスパニングする VLAN では、これまで STP ループ型トポロジが形成されていました。これは、いずれかのアップリンクが STP でブロックされるためです（詳細については図 2-21 と図 1-4 を参照してください）。VSS に MEC の機能が加わると、キャンパス トポロジからループが排除されます。これは、STP が EtherChannel 論理

ポートで動作するようになったことで、各物理スイッチが単一の論理リンクを介して単一の論理スイッチに接続しているように見えるためです。STP の側から見ると、このスター型トポロジにはループがありません。STP にとって、ブロックしなければならない代替パスは存在しません。図 2-21 に、2 種類のトポロジを示します。1 つは、MEC のないアクセス レイヤから VSS までのトポロジで、ここではアップリンクがブロックされています。もう 1 つは、MEC がある VSS トポロジで、両リンクともループなしで転送されます。ループのないネットワークの利点については、第 3 章「VSS 対応キャンパス デザイン」を参照してください。

図 2-21　非 MEC トポロジと MEC トポロジの帯域幅容量



非 MEC トポロジと MEC トポロジの帯域幅容量

- *フォワーディングに利用できる帯域幅が 2 倍*：MEC は、リンクの冗長性を提供する手段としてスパニング ツリーを置き換えます。これは、MEC 下のすべての物理リンクがトラフィックのフォワーディングに利用できることを意味します。STP はもはや個々のリンクをブロックできません。自身のデータベースがループのないパスを計算できるリンクを持たないからです。ループ型トポロジのネットワークにとって、フォワーディングの総容量は、物理リンクで利用可能な帯域幅の半分です。MEC を備えた VSS は、すべてのリンクをフォワーディングに充てることができるため、利用可能な帯域幅が 2 倍になります。これによって既存ネットワークからの効果的な変革がもたらされます。既存の 10 ギガバイト少ないインフラストラクチャの場合、利用可能なリンクを効果的に活用するには、設計上の次善策（各アップリンクでの EtherChannel、ルーテッド アクセス、複数の HSRP グループの使用など）を選択する必要があります。MEC を備えた VSS の場合、すべての新設計でトポロジを簡素化しつつこれらの利点を享受できます。

VSS 対応のキャンパス デザインで MEC が不可欠な理由

MEC の機能性、運用方法、トラフィック フロー、およびキャンパス設計に与える影響の詳細を検討する前に、MEC が VSS 設計にとって不可欠な理由を理解することは重要です。MEC がなくても VSS 対応型ネットワークを構築することはできます。ただし、トポロジは結果的に単一障害点（隣接ネットワーク デバイスまでのリンクが 1 つしかない状態）か、ループ型トポロジのいずれかを抱えることになります。これは、指定のネットワーキング デバイスからの両リンクが、非 EtherChannel が設定された単一の論理 VSS スイッチに接続することが原因です。いずれの状況も、あらゆる所定のネットワークにおける VSS の利点を損ないます。第 3 章「VSS 対応キャンパス デザイン」では、MEC 対応の設計の重要性を正当化するために、いくつかの設計上の根拠を挙げて説明します。MEC 対応の設計の主な利点は、次のとおりです。

- ループのないトポロジの実現。
- 利用可能なフォワーディング帯域幅が 2 倍に拡大し、アプリケーションの応答時間の短縮、ネットワークの輻輳の減少、オペレーション コストの削減につながる。

- 単一リンクの障害（ノード障害またはメンバ リンク障害）に関連したコントロール プレーンのアクティビティが削減または排除される。
- トラフィック フローを高速に収束するためのハードウェアでの障害検出が可能。
- 1 つのリンクに障害が発生してもレイヤ 2 コントロール プレーンのコンバージェンスにならないため、MAC 学習が軽減する。
- ルーティング プロトコルの送出の減少。レイヤ 3 ネットワークの効率が向上します（集約の必要性とコントロール プレーン アクティビティが減少）。
- 1 つのリンクに障害が発生してもレイヤ 3 コントロール プレーンのコンバージェンスにならないため、マルチキャスト コントロール プレーンの変更を回避できる（マルチキャスト着信インターフェイスの変更の回避）。
- MEC 対応トポロジの VSL におけるマルチキャストの複製の回避。
- デュアルアクティブ検出方法を柔軟に展開できる（詳細については、「VSS デュアルアクティブ スーパーバイザを使用したキャンパス復旧」（P.4-19）を参照してください）。

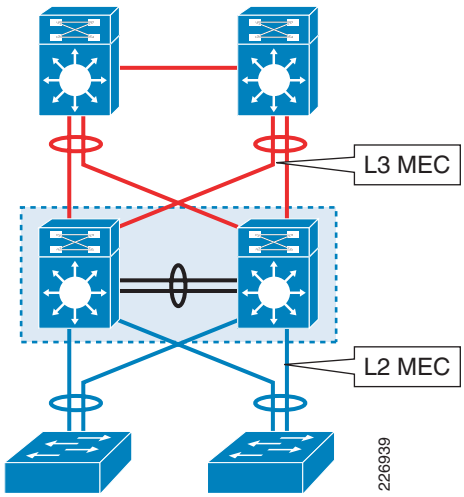
MEC の種類とリンク集約プロトコル

MEC は分散型 EtherChannel 環境です。ただし、従来型ネットワークで使用されるすべての EtherChannel の特性を受け継いでいます。このセクション以降は、MEC に適用される EtherChannel テクノロジーの側面について、その一部を説明します。ここでは、他のマニュアルや資料では必ずしも詳しく紹介されていない EtherChannel の機能を中心に解説します。特に、VSS の実装に必要と考えられている機能について説明します。一部の機能図やプロトコル設定を繰り返し掲載していますが、これは VSS 環境という文脈での MEC 展開に対する読者の理解を助けるためのものです。

MEC の種類

ネットワークの接続要件に応じて、MEC 設定にはレイヤ 2 とレイヤ 3 の 2 種類のモードがあります。図 2-22 を参照してください。

図 2-22　MEC の種類



レイヤ 2 MEC

図 2-22 に示すように、3 つの層からなる階層構造のネットワークでは、レイヤ 2 MEC は、アクセス レイヤとディストリビューション レイヤとの接続に適用されます（VSS 対応の場合）。このモードでは、レイヤ 2 の MEC は、STP、MAC アドレス学習、および他のレイヤ 2 オペレーションのホスティングに関与します。レイヤ 2 MEC 対応の接続は、ネットワーク全体がループのない大規模な階層構造のレイヤ 2 トポロジを形成するために拡張できます。このマニュアルではこのような設計については説明しません。

レイヤ 3 MEC

レイヤ 3 MEC は、ルーテッド ポートチャネル インターフェイスで構成されています。レイヤ 3 MEC では、ポートチャネル インターフェイスは、CEF でルーティングやフォワーディングなどのレイヤ 3 の機能を実施する IP アドレスを持つルーテッド インターフェイスとして設定されます。この種類の接続の典型的な拡張としては、複数のレイヤ 3 の VSS ペアを互いに接続して一つのルーテッド設計の基盤とする方法があります。レイヤ 3 MEC のルーティングの適用については、「VSS を使ったルーティング」（P.3-46）を参照してください。

リンク集約プロトコル

EtherChannel は論理インターフェイスです。物理メンバ リンクの残りのネットワークに対する動作（および運用上の影響）を管理するには、ある種のコントロール プレーンが必要になります。EtherChannel グループの基本となるリンクのコントロール プレーンを管理する方法として、次の 2 種類が利用できます。

- ポート集約プロトコル（PAgP）
- Link Aggregation Control Protocol（LACP）または IEEE 802.1ad

これらのプロトコルはいずれも次の一般的な機能があります。

- VSS とネイバー スイッチとの間でリンク集約パラメータの一貫性と 互換性を確保
- 集約の要件との準拠性を確保
- ローカルおよびリモートの EtherChannel 設定におけるランタイムの変更および障害に動的に対処する
- 単方向リンク接続を検出して EtherChannel バンドルから削除する

EtherChannel は VSS 対応のキャンパス設計における基本的なビルディング ブロックです。MEC の適切な導入には、予期しない動作を引き起さないトポロジを形成する、運用上の一貫性と複数のアクセス レイヤ スイッチとの対話が必要です。この理由により、前述した機能の利点を活用するには、MEC インターフェイスは PAgP または LACP のいずれかに対応する必要があります。従来型の EtherChannel と同様、PAgP または LACP に対応した MEC は、システム設定と基本となる物理リンクの運用ステートの一貫性チェックを行います。PAgP および LACP は不一致リンクをバンドルから削除するため、syslog メッセージを通じて、不一致の設定と運用検証に対する保護機能を追加します。次の設定は、MEC バンドルに関与する、すべての基本の物理リンク間で一致する必要があります。

- メンバ リンクでの VLAN の設定
- メンバ リンクのトランクの種類と設定
- ポート ステータス（フルまたは半デュプレックス）および基本のハードウェアがサポートする QoS はすべてのリンクで同一の必要がある

EtherChannel の形成における要件の全リストについては、www.cisco.com から入手できる個々の製品リリース ノートおよび関連するマニュアルを参照してください。ネットワーク設計に影響を与える PAgP および LACP 設計の考慮事項（VSS に適用される事項）については、「PAgP」と後続の「LACP（IEEE 802.1ad）」の各セクションを参照してください。

PAgP

PAgP は最も広く使用されているリンク集約プロトコルです。PAgP は大部分のシスコのデバイスおよび他のサードパーティ ネットワーク インターフェイス カード（NIC）でサポートされています。VSS のコンテキストでは、PAgP は、前述のセクションで概要を示した機能に加えて、デュアルアクティブなリカバリでの支援機能という付加価値を提供しています。PAgP はそれぞれの MEC リンク メンバで動作し、VSS とレイヤ 2 またはレイヤ 3 のパートナーとの間でネイバーとの隣接関係を確立および管理します。PAgP は、メンバ リンク ベースごとにピアスイッチ ステート、デバイス名、パートナー ポート、ポート集約の機能、ポートチャネル バンドルのグループ インデックスといった複数の属性を決定します。

デバイス ID

アクティブ スイッチは、PAgP コントロールプレーン トラフィックの発信と終端の役割を担っています。PAgP は、接続の各終端を固有のデバイス ステータスで識別するためにデバイス ID をアドバタイズします。VSS の場合、これは MAC アドレス形式による 48 ビットの数字です（02.00.00.00.00.*xx*）。デバイス ID は、先頭から 5 つのオクテット（02.00.00.00.00）と最終オクテットが可変値（*xx*）の固定プレフィクスを組み合わせて構成されます。可変の部分は、システムに設定されている仮想スイッチ ドメインの識別名です。この PAgP デバイス ID は、VSS ドメインを構成する 2 つのスイッチで終端する 2 つのリンクによって送信されます。デバイス ID は同一のため、リモート デバイスは、単一論理デバイス（VSS）からデバイス ID が送られてくると想定します。VSS でのロールのスイッチオーバー中であっても、再ネゴシエーション プロセスを防ぐために、デバイス ID は各 PAgP ネイバーとの一貫性を保ちます。このように PAgP および LACP で仮想スイッチ ドメインの識別名を使用するには、MEC を通じて相互接続されているすべての VSS ドメインでこの識別名が固有である必要があります。次のコマンド出力例は、このセクションで説明したデバイス ID の値で固定と可変のコンポーネントを示しています。

```
6500-VSS# sh run | inc virtual
switch virtual domain 10 ! <-- Device ID
6500-VSS# show pagp neighbor
Flags:S - Device is sending Slow hello.C - Device is in Consistent state.
        A - Device is in Auto mode.        P - Device learns on physical port.

Channel group 10 neighbors
          Partner       Partner          Partner           Partner     Group
Port      Name          Device ID        Port         Age  Flags       Cap.
Gi1/1     6500-VSS      0200.0000.000a   Gi1/4/1      7s   SC          A0001
Gi1/2     6500-VSS      0200.0000.000a   Gi2/4/1      8s   SC          A0001
```

PAgP の動作モード

PAgP には多数の設定オプションがあります。EtherChannel 用に PAgP を設定する場合のベスト プラクティスは、次の URL から入手可能なマニュアルを参照してください。

http://www.cisco.com/en/US/products/hw/switches/ps700/products_white_paper09186a00801b49a4.shtml

このデザイン ガイドでは、これらのマニュアルから詳細を選択し、VSS 対応のキャンパス環境に適した推奨事項を統合しています。

表 2-4 に示すのは、ベスト プラクティスベースの設定オプションだけです。表示される 3 つの選択肢のうち、PAgP ネイバーに推奨されるモードは、*desirable-desirable* です。この設定オプションは、両側で PAgP を使用可能にし、このセクションで前述した一貫性チェックを強制的に行います。desirable-desirable オプションは、MEC 接続するデバイスの動作の安全性と信頼性を確保するために最適なオプションです。LACP とは異なり、PAgP は、ポートをバンドルする前に、厳密なチャネル設定とコンフィギュレーション チェックを提供します。MEC バンドルは、PAgP+ がコンフィギュレー

ションの不一致を検出すると無効のステートのままとなり、エラーが修正されるまで使用できません。この追加の措置により EtherChannel の一貫性が欠如するのを回避できます。一貫性が保たれないと、大規模な EtherChannel 展開を管理する際に運用上の問題が発生します。

表 2-4　PAgP のベスト プラクティスベースの設定オプション

チャネル モード：レイヤ 2 およびレイヤ 3 の両方の MEC に設定	VSS	リモート ノード	MEC のステート
	desirable	desirable	operational
	desirable	auto	
	auto	desirable	

ネットワークの要件に応じて、追加の設定オプションの *silent* または *non-silent* が使用できます。ほとんどのネットワーク設計で望ましいオプションは silent モードです。silent モードは、データの有無にかかわらずリンクをバンドルに統合します。non-silent オプションは、リンクの完全性の間接的な測定に使用されます。ベスト プラクティスの設計で、リンクの完全性をチェックする場合に望ましい方法は UDLD です。結果的に、ほとんどのネットワーク実装では、silent オプションの方がより適切です。

PAgP の hello 値をデフォルトに設定したままにする理由

デフォルトでは、non-silent モードの PAgP は、PAgP の hello メッセージを MEC の各メンバ リンクで 30 秒間隔で 1 つずつ独自に送信します。これは、リモートのパートナーのアベイラビリティを検出するのに 105 秒（hello 間隔 3.5 回分）かかることから、*slow-hello* と呼ばれます。このタイマーは 1 秒に 1 回ずつの送信に変更できます。この場合は *fast-hello* と呼ばれます。UDLD が fast-hello より時間がかかるため（1 秒の 3 倍）、ネットワーク設計の多くがこのオプションを使用してリンク検出を高速化する傾向にあります。ただし、次の 2 つの理由から、VSS の展開では fast-hello の設定を避ける必要があります。

- VSS のコントロール プレーンは、3 秒以内に復旧しない場合があるため（SSO スイッチオーバー中）、リモート エンドで VSS が応答不能と宣言される前に、VSS が PAgP hello を送信する可能性があります。これによって false positive につながる可能性があります。
- fast-hello はリンクベースで送信されます。大規模な展開では、fast-hello の送信によってスイッチの CPU がオーバーランする場合があります。

（注） EtherChannel の一方の側が fast-hello を設定し、もう一方の側（またはデバイス側）が slow-hello を設定している場合、操作上は fast-hello がデバイス間で送受信されます。これはつまり、VSS に slow-hello が設定されていても、リモート デバイスで不適切な設定がされている場合、hello メッセージを送受信する操作モードが変更される可能性があることになります。

ヒント **この場合のベスト プラクティスは、PAgP タイマーの設定をデフォルト値のままにし、通常の UDLD を使用してリンクの完全性をモニタすることです。**

LACP（IEEE 802.1ad）

LACP は、マルチベンダーの相互運用性を可能にする業界標準のポート集約プロトコルです。LACP は機能性と動作において PAgP に非常によく似ています。VSS では、レイヤ 2 とレイヤ 3 の両方の MEC インターフェイスで動作します。LACP の動作と設定オプションの詳細については、次の URL を参照してください。

http://www.cisco.com/en/US/products/hw/switches/ps700/products_white_paper09186a00801b49a4.shtml

デバイス ID

LACP を VSS に実装する際は、事前に計算されたシステム ID が使用されます。LACP のシステム ID は、先頭から 5 つのオクテット（02.00.00.00.00）と最終オクテットが可変値（*xx*）の固定プレフィクスを組み合わせた、48 ビットのアドレスで構成されています。この可変部分は、システムに設定されている仮想スイッチ ドメインの識別名です。次の出力例は、システム ID の最終オクテットに使用された仮想スイッチ ドメイン番号でシステム ID を識別しています。

```
6500-VSS# sh lacp sys-id
32768,0200.0000.000a ! Device ID info
6500-VSS# sh run | inc virtual
switch virtual domain 10 ! Device ID info
```

LACP の動作モード

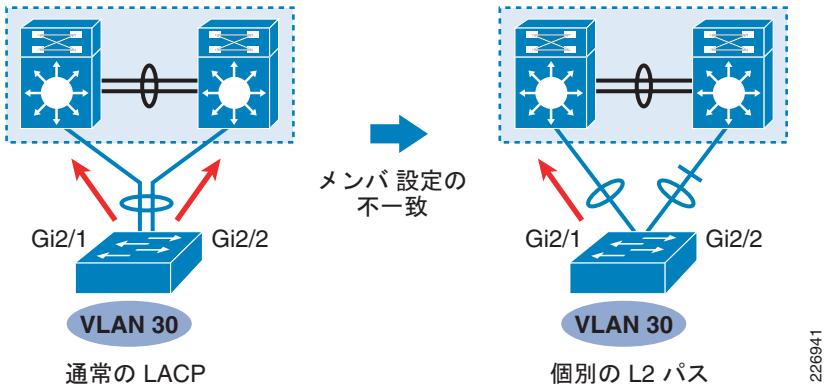
前述の PAgP と同じ理由で、表 2-5 には、ベスト プラクティスベースの設定オプションだけを示します。表示される 3 つの選択肢のうち、LACP ネイバーに推奨されるモードは、*アクティブ- アクティブ* です。

表 2-5　　LACP のベスト プラクティスベースの設定オプション

チャネル モード：レイヤ 2 およびレイヤ 3 の両方の MEC に設定	VSS	リモート ノード	MEC のステート
	active	active	operational
	active	passive	
	passive	passive	

LACP にアクティブ - アクティブ オプションを指定して EtherChannel を設定すると、PAgP と同様にメンバ リンクの設定の一貫性を保つことができますが、最終的な結果は異なります。EtherChannel をバンドルするプロセス中、LACP は、ポートチャネルのメンバになろうとしている物理リンクごとに設定の一貫性をチェックします。設定チェックが失敗した場合は、syslog メッセージが生成されます。さらに、システムから特殊な EtherChannel インターフェイスが生成され、これに固有の英字の ID が割り当てられます。システムで生成された LACP MEC は、すべての物理ポートを設定チェックに失敗した MEC にバンドルします。図 2-23 を参照してください。

図 2-23　　LACP 設定の不一致の図解



次に示す CLI 出力と設定プロセスは、この動作を表したものです。次の出力例で、この LACP リンク メンバは不整合な設定で再設定されています。不正な設定のインターフェイスがポートチャネル インターフェイスに参加を試みると、固有のシステム生成のポートチャネル インターフェイスが設定チェックからトリガーされます。

```
6500-VSS# show etherchannel 20 summary | inc Gi
```

```
Po20(SU)         LACP      Gi2/1(P)       Gi2/2(P)
6500-VSS# show spanning-tree | inc Po20
Po20               Root FWD 3         128.1667 P2p
6500-VSS# config t
6500-VSS(config)# int gi2/2
6500-VSS(config-if)# switchport nonegotiate
6500-VSS(config-if) # shut
6500-VSS(config-if)# no shut
%EC-SPSTBY-5-CANNOT_BUNDLE_LACP: Gi2/2 is not compatible with aggregators in channel 20
and cannot attach to them (trunk mode of Gi2/2 is trunk, Gi2/1 is dynamic)
%EC-SP-5-BUNDLE: Interface Gi2/2 joined port-channel Po20B ! A system generated
port-channel
6500-VSS# show etherchannel 20 summary | inc Gi
Po20(SU)        LACP      Gi2/1(P)
Po20B(SU)       LACP      Gi2/2(P) ! Bundled in separate system-generated port-channel
                                   ! interface

6500-VSS# show spanning-tree | inc Po20
Po20        Root FWD   4         128.1667     P2p
Po20B       Altn BLK   4         128.1668     P2p ! Individual port running STP is blocked
```

これによって次の 2 つのバンドルが作成されます。

- 最初のバンドルは、設定チェックが正常に行われたポートに関連しています。
- 2 番目のバンドルは、システムで生成され、設定が一致しなかったポートが含まれます。
- 結果的に、両方のポートチャネル インターフェイスでコントロール プレーンがアクティブになります（それぞれが 1 メンバを保有）。最終的なトポロジは、図 2-23 で示すように、アクセススイッチと VSS の間に作成された 2 つの固有のレイヤ 2 パスで構成されます（これはレイヤ 3 MEC でも同様になりますがこの例には示していません）。STP トポロジは、このようなネットワークをループ型と見なし、より上位の STP のプライオリティでポートをブロックします。これが PAgP と比較したときに LACP が抱える、ネットワーク トポロジにおける主な動作上の考慮事項の一つです。PAgP では、ポートのバンドル前に、より厳密なチャネル設定と設定チェックを行います。PAgP では、PAgP+ で設定の不整合が検出された場合、エラーが修正されるまでの間、MEC は無効のままとなります（表示されるステートは *down* です）。

LACP の hello 値をデフォルトに設定したままにする理由

PAgP では、LACP は hello の間隔をデフォルトの 30 秒（slow-hello）から 1 秒（fast-hello）までの間に設定できます。LACP のネイバー デバイス接続の両端で同一の設定がされていない限り、接続の相手側とは非対称のレートで LACP hello を送信できます。これはつまり、VSS に接続するリモート デバイスから LACP に slow-hello 送信を行い、VSS からは LACP に fast-hello 送信ができることになります。リンク障害の検出に fast-hello 方式を使用した場合は、設定に基づいて検出の時間間隔も変更できます（接続の一方で 30 秒、他方で 3 秒に設定するなど）。これは PAgP とは異なります。PAgP では、hello の送信レートは両端とも fast-hello がデフォルトです（fast-hello 要求が行われた場合）。VSS 展開で fast-hello の設定を回避する必要があるのは、次の 2 つの理由からです。

- （SSO スイッチオーバー中）VSS のコントロール プレーンが 3 秒以内に復旧しない場合があるため（fast-hello のタイムアウト）、リモート エンドで VSS が応答不能と宣言される前に、VSS が LACP hello を送信する可能性があります。これによって false positive につながる可能性があります。
- fast-hello はリンクベースで送信されます。大規模な展開では、fast-hello の送信によってスイッチの CPU がオーバーランする場合があります。
- VSS に接続するサーバが fast-hello の LACP-MEC に基づいて設定されている場合、VSS が fast-hello 送信を行うのを停止させる固有の方法はありません。このため、このようなサービスを必要とするサーバが増えると、CPU の使用状況が過大になる可能性があります。

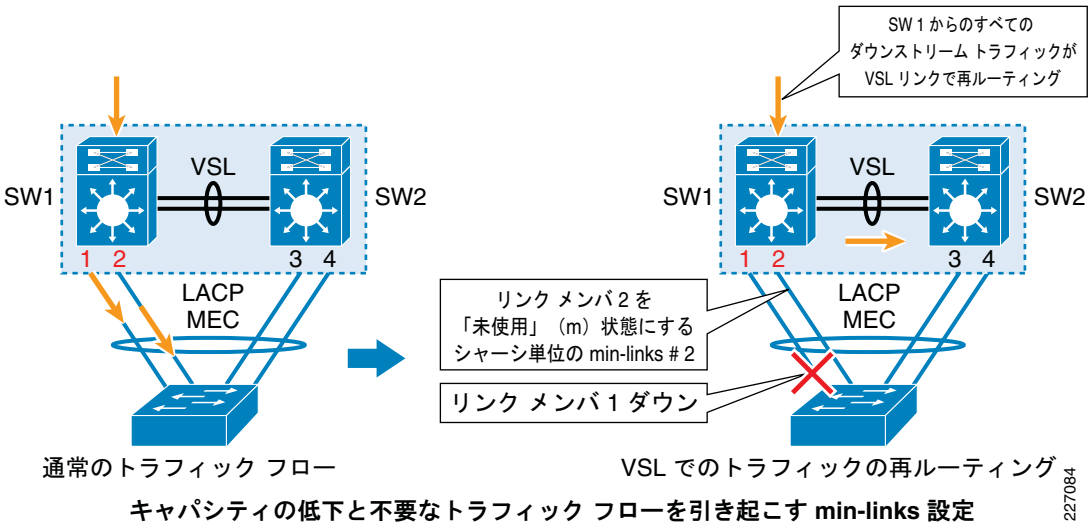
ヒント　**この場合のベスト プラクティスは、LACP タイマーの設定をデフォルト値のままにし、通常の UDLD を使用してリンクの完全性をモニタすることです。**

VSS における LACP の最小のリンク動作

指定のアプリケーションまたはネットワーキング要件で必要とされる帯域幅のレベルを維持するには、LACP の最小のリンク機能が使用されます（*up/up* ステートにあるインターフェイスの数を単位として）。一定の帯域幅をサポートするリンクの数が最小の要件未満になると、このバンドルがサービスから除外されます。スタンドアロン設計では、min-links 機能は個別の 2 ノード間に展開され、必要に応じてトラフィックのフォワーディングに代替ノードが使用可能になります。VSS では、min-links 機能の動作が異なります。VSS の場合、LACP EtherChannel の min-links は、実施が物理シャーシごとであるのにもかかわらず、設定がポートチャネルごとに維持されます。次の設定例および図 2-24 に、この機能の適用方法を示しています。

```
6500-VSS# show etherchannel 10 summary | inc Gi
10     Po10(SU)         LACP      Gi1/4/1(P)     Gi1/4/2(P)     Gi2/4/1(P)     Gi2/4/2(P)

6500-VSS# conf t
6500-VSS(config)# int po10
6500-VSS(config-if)# port-channel min-links 2
6500-VSS(config-if)# int gi1/4/1
6500-VSS(config-if)# shutdown
%LINK-5-CHANGED: Interface GigabitEthernet1/4/1, changed state to administratively down
%LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet1/4/1, changed state to
down
%LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet1/4/2, changed state to
down
%EC-SW2_SPSTBY-5-MINLINKS_NOTMET: Port-channel Po10 is down bundled ports (1) doesn't meet
min-links
%EC-SW1_SP-5-MINLINKS_NOTMET: Port-channel Po10 is down bundled ports (1) doesn't meet
min-links
—%LINEPROTO-SW1_SP-5-UPDOWN: Line protocol on Interface GigabitEthernet1/4/2, changed
state to down
%LINK-SW1_SP-5-CHANGED: Interface GigabitEthernet1/4/1, changed state to administratively
down
%LINEPROTO-SW1_SP-5-UPDOWN: Line protocol on Interface GigabitEthernet1/4/1, changed state
to down

6500-VSS# show etherchannel 10 summary
Flags:  D - down        P - bundled in port-channel
        M - not in use, no aggregation due to minimum links not met
        m - not in use, port not aggregated due to minimum links not met

10     Po10(SU)         LACP      Gi1/4/1(D)     Gi1/4/2(m)     Gi2/4/1(P)   Gi2/4/2(P)
```

図 2-24　キャパシティの低下を引き起こす min-links 設定



キャパシティの低下と不要なトラフィック フローを引き起こす min-links 設定

LACP コントロール プロトコルを使用する MEC の場合、*minlinks* を使用して MEC の各シャーシで動作可能にする物理リンクの最小数を定義します。下記の例では、MEC に対する **port-channel min-links 2** という設定により、それぞれの仮想スイッチ メンバは、MEC に関連付けるために、少なくとも 2 つの動作可能なローカル メンバ リンク ポートと一致する必要があることを示しています。メンバ リンクのいずれかがダウンすると、シャーシ内の他のメンバが *not in use* ステートに置かれます。このように実施されると、1 つのリンク障害のために 1 スイッチ メンバの接続性が完全に失われる結果となり、これは、前述の syslog 出力例と図 2-24 に示したように、他のリンクがトラフィックのフォワーディングに利用可能である場合も同様になります。これによって、VSL リンクのトラフィックの再ルーティングが強制的に行われ、他のシャーシのバンドル内の 2 つのリンクの輻輳がさらに深刻になります。

2 つのリンクのポートチャネル設定（アクセス スイッチから VSS までの一般的なネットワーク トポロジ）には min-links 機能は適用できません。これは、物理シャーシごとに少なくとも 2 つのリンクが検索され、すでに設定されていると MEC が動作可能にならないためです。次の出力例では、各物理チャネル シャーシにメンバ リンクが 1 つ存在する VSS にメンバ 2 つの EtherChannel を接続する設定で、LACP に min-link を設定した場合の動作を示します。

```
6500-VSS# show etherchannel 150 sum
Flags:  D - down        P - bundled in port-channel
Group  Port-channel  Protocol    Ports
------+-------------+-----------+-----------------------------------------------
150    Po150(SU)        LACP      Gi1/7/1(P)     Gi2/7/1(P)

6500-VSS# sh spanning-tree int po 150

Vlan                 Role Sts Cost      Prio.Nbr Type
-------------------- ---- --- --------- -------- --------------------------------
VLAN0050             Desg FWD 3         128.1667 P2p
VLAN0150             Desg FWD 3         128.1667 P2p

6500-VSS# sh int po 150
Port-channel150 is up, line protocol is up (connected)
  <snip>
  input flow-control is off, output flow-control is off
  Members in this channel: Gi1/7/1 Gi2/7/1
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input never, output never, output hang never
  <<snip>>
```

```
6500-VSS# conf t
Enter configuration commands, one per line.  End with CNTL/Z.
6500-VSS(config)# int po 150
6500-VSS(config-if)# port-channel min-links 2
6500-VSS(config-if)#
%LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet1/7/1, changed state to
down
%LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet2/7/1, changed state to
down
%LINEPROTO-5-UPDOWN: Line protocol on Interface Port-channel150, changed state to down
%LINK-3-UPDOWN: Interface Port-channel150, changed state to down
%EC-SW1_SP-5-MINLINKS_NOTMET: Port-channel Po150 is down bundled ports (1) doesn't meet
min-links
%SPANTREE-SW1_SP-6-PORT_STATE: Port Po150 instance 50 moving from forwarding to disabled
%SPANTREE-SW1_SP-6-PORT_STATE: Port Po150 instance 150 moving from forwarding to disabled
%EC-SW1_SP-5-MINLINKS_NOTMET: Port-channel Po150 is down bundled ports (1) doesn't meet
min-links
%LINEPROTO-SW1_SP-5-UPDOWN: Line protocol on Interface GigabitEthernet1/7/1, changed state
to down
%LINEPROTO-SW1_SP-5-UPDOWN: Line protocol on Interface GigabitEthernet2/7/1, changed state
to down
%EC-SW2_SPSTBY-5-MINLINKS_NOTMET: Port-channel Po150 is down bundled ports (0) doesn't
meet min-links
%EC-SW2_SPSTBY-5-MINLINKS_NOTMET: Port-channel Po150 is down bundled ports (1) doesn't
meet min-links
```

LACP コントロール プロトコルを使用する MEC では、min-links を設定すると、MEC を動作可能にする各シャーシの物理リンクの最小数が定義されます。各物理シャーシに単一のメンバが接続する場合、この設定によって min-links の要件が満たされなくなります。次の出力例に、ポートチャネル インターフェイスが無効化され、各メンバ リンクの LACP のステートが待機ステートに置かれる様子を示します。min-link 機能を使用すると、関連する接続で MEC が無効化されます。

```
6500-VSS# sh int po 150
Port-channel150 is down, line protocol is down (notconnect)
! <<snip>>
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input never, output never, output hang never
  Last clearing of "show interface" counters never
  Input queue: 0/2000/0/0 (size/max/drops/flushes); Total output drops: 0
  Queueing strategy: fifo
  Output queue: 0/2000 (size/max)

6500-VSS# sh etherchannel 150 su
Flags:  D - down        P - bundled in port-channel
        w - waiting to be aggregated

Group  Port-channel  Protocol    Ports
------+-------------+-----------+-----------------------------------------------
150    Po150(SM)       LACP      Gi1/7/1(w)     Gi2/7/1(w)

Last applied Hash Distribution Algorithm: Adaptive

6500-VSS# sh spanning-tree int po 150
no spanning tree info available for Port-channel150
```

ヒント　**キャンパス環境で min-links 機能を使用する効果はあまりありません。キャンパスで VSS を使用する場合、どのような実践的な展開においても（隣接するネットワーク デバイスごとに 2 つのアップリンクがあるキャンパス環境）、min-links 機能は使用しないでください。**

MEC の設定

MEC の設定要件は、標準の EtherChannel インターフェイスの要件とほぼ同じです。ここでは、MEC の設定における基本的な注意事項、QoS サポート、および syslog ガイドラインについて説明します。

レイヤ 2 EtherChannel の設定手順は、レイヤ 3 EtherChannel の手順とは異なります。主に次の点について考慮する必要があります。

- CLI による定義でレイヤ 2 MEC を明示的に作成しない。その代わりに、各メンバ インターフェイス以下のポートチャネル グループとの関連付けを Cisco IOS システムで行ってレイヤ 2 MEC インターフェイスを生成するようにします。
- CLI の実行と各メンバ インターフェイス以下のポートチャネル グループとの関連付けにより、レイヤ 3 MEC インターフェイスを明示的に作成する。

詳細については、次の URL を参照してください。

http://cco.cisco.com/en/US/partner/docs/switches/lan/catalyst6500/ios/12.2SX/configuration/guide/channel.html#wp1020478

MEC を備えた QoS

MEC での QoS は、あらゆる標準の EtherChannel 設定の手順と同様です。一般的な QoS サポートまたは VSS に関する制約事項については、次の URL のホワイト ペーパーの QoS に関する章を参照してください。

http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps9336/white_paper_c11_429338.pdf

モニタリング

EtherChannel は VSS 対応のキャンパス トポロジで広く使用されているため、これをモニタリングする重要性は非 VSS の EtherChannel 環境より高くなります。標準の EtherChannel インターフェイスのモニタリングで利用可能な一般的なツールは、MEC インターフェイスに十分適用できます。ただし、ここでは、EtherChannel の接続性に関する操作上の理解を深めるために有効にする必要がある、新しいコマンドの **show** と、特定の **logging** コマンドに関する詳細な追加情報を示します。

Cisco Catalyst 6500 プラットフォーム用 Cisco IOS リリース 12.2(33)SXH1 では、MEC または EtherChannel の特定のリンク メンバ上のトラフィック フローをモニタリングする、これまで非表示だったコマンドをサポートするようになりました。次の **remote** コマンドは、インターフェイスとポートチャネル インターフェイスが指定の発信元と宛先用に選択されていることを表す出力を生成します。

次のコマンドは非表示です。

```
Catalyst6500# remote command switch test EtherChannel load-balance interface po 1 ip
1.1.1.1 2.2.2.2
Would select Gi4/1 of Po1
```

次は、一般的な EtherChannel の実装のモニタリングに使用できる Cisco IOS コマンドです。

```
6500-VSS# show EtherChannel load-balance hash-result interface port-channel 2 205 ip
10.120.7.65 vlan 5 10.121.100.49
Computed RBH: 0x4
Would select Gi1/9/19 of Po205
```

MEC インターフェイスを備えた VSS では、次の syslog 設定（**logging** コマンド）が推奨されます。これらのコマンドは、VSS に接続するデバイスが対応する syslog の機能をサポートする限り、これらのデバイスにも適用できます。

ポートチャネル インターフェイスの設定は次のようになります。

```
interface Port-channel20
 logging event link-status
```

```
 logging event spanning-tree status

logging event link-status
%LINK-5-CHANGED: Interface Port-channel220, changed state to administratively down
%LINK-SW1_SP-5-CHANGED: Interface Port-channel220, changed state to administratively down

logging  event spanning-tree status
%SPANTREE-SW1_SP-6-PORT_STATE: Port Po220 instance 999 moving from learning to forwarding
```

メンバ リンクの設定は次のとおりです。

```
interface GigabitEthernet1/8/1
 description Link member to port-channel
 logging event link-status
 logging event trunk-status
 logging event bundle-status

logging event link-status
Mar 25 11:43:54.574: %LINK-3-UPDOWN: Interface GigabitEthernet1/8/1, changed state to down
Mar 25 11:43:54.990: %LINK-3-UPDOWN: Interface GigabitEthernet2/8/1, changed state to down

logging event trunk-status
%DTP-SW2_SPSTBY-5-NONTRUNKPORTON: Port Gi2/8/1 has become non-trunk
%DTP-SW1_SP-5-NONTRUNKPORTON: Port Gi2/8/1 has become non-trunk

logging event bundle-status
%EC-SW2_SPSTBY-5-BUNDLE: Interface Gi1/8/1 joined port-channel Po220
%EC-SW2_SPSTBY-5-BUNDLE: Interface Gi2/8/1 joined port-channel Po220
```

MEC の負荷分散、トラフィック フロー、および障害

負荷分散、トラフィック フローの動作、障害条件については、第 3 章「VSS 対応キャンパス デザイン」を参照してください。MEC の負荷分散の動作や影響は、キャンパス ネットワーク全体の設計とより密接に関連するためです。

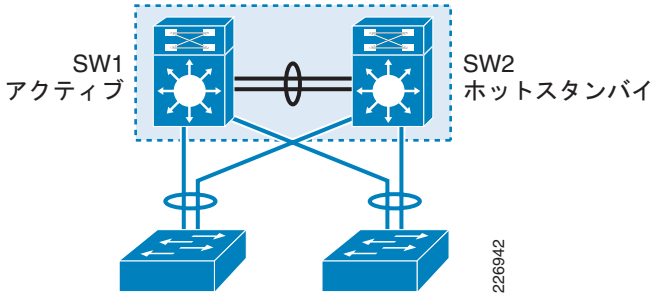
MEC のキャパシティ計画

VSS でサポートする EtherChannel の最大数は、Cisco IOS のバージョンによって異なります。Cisco IOS 12.2(33) SXH でサポートされる EtherChannel の最大数は 128 です。Cisco IOS 12.2(33) SXI ソフトウェア リリースではこの上限が 512 にまで増加します。VSS のディストリビューション ブロックに適用される範囲とスケーラビリティについては、「VSS を使用したマルチレイヤ デザインのベスト プラクティス」（P.3-14）を参照してください。

MAC アドレス

スタンドアロンの Cisco Catalyst 6500 では、各インターフェイスとコントロール プレーンで使用する MAC アドレスはバック プレーン EEPROM から取得します。VSS は、2 つのシャーシで構成されています（図 2-25 を参照してください）。VSS ペアの各物理メンバは、バックプレーン EEPROM に保存されている MAC アドレスのプールで構成されます。

図 2-25　VSS のロール



VSS MAC アドレス プールは、ロール解決のネゴシエーション中に決定されます。MAC アドレスのアクティブ シャーシ プールは、レイヤ 3 MEC インターフェイスを含む、レイヤ 2 の SVI とレイヤ 3 のルーテッド インターフェイスで使用されます。レイヤ 2 MEC インターフェイスは、リンクメンバの MAC アドレスのいずれかを使用します。次の CLI 出力例は、MAC アドレスのアクティブ プールを表したものです。

```
6500-VSS# show switch virtual role

Switch  Switch Status  Preempt    Priority  Role
-------------------------------------------------------------------
LOCAL    1     UP   FALSE(N )   110(110)  ACTIVE
REMOTE 2     UP   FALSE(N )   100(100)  STANDBY

6500-VSS# show catalyst6000 chassis-mac-addresses
  chassis MAC addresses: 1024 addresses from 0019.a927.3000 to 0019.a927.33ff
```

ホットスタンバイ スイッチがアクティブ スイッチを引き継ぐ場合、インターフェイスへの MAC アドレスの割り当てがスイッチオーバー イベント中に変更されることはありません。これは、VSS に接続するデバイスから gratuitous ARP が更新される（同じ IP アドレスの MAC アドレスが変更される）のを防ぎます。ただし、両方のシャーシが同時にリブートしてアクティブ スイッチの順序が変わる（古いホットスタンバイ スイッチが最初にアクティブになる）と、VSS ドメイン全体が初めのスイッチの MAC アドレス プールを使用します。これは、インターフェイスは新しい MAC アドレスを継承し、これによって gratuitous ARP がレイヤ 2 とレイヤ 3 のすべてのインターフェイスで更新されることを意味します。VSS から 1 ホップ先で接続するすべてのネットワーキング デバイス（および gratuitous ARP をサポートしないすべてのネットワーキング デバイス）は、デフォルトのゲートウェイおよびインターフェイスの MAC アドレスがリフレッシュされるかタイムアウトになるまで、トラフィックが中断されます。このような中断を回避するために、シスコでは、VSS のレイヤ 2 とレイヤ 3 インターフェイスの MAC アドレスを予約プールから取得する設定オプションを使用する方法をお勧めします。これで仮想スイッチ ドメインの識別名の利点を活かして MAC アドレスを形成できます。VSS ドメインの MAC アドレスは、ブートの順序にかかわらず、仮想 MAC アドレスの使用中は常に不変です。具体的な方法については、www.cisco.com で VSS のコマンドと設定に関する章を参照してください。

ヒント　**シスコでは、*switch virtual domain* コマンドを使用して VSS ドメインの仮想 MAC アドレスを設定することをお勧めします。**

```
6500-VSS(config)# switch virtual domain 10
6500-VSS(config-vs-domain)# mac-address use-virtual
```

各シャーシの EEPROM に常駐する個々の MAC アドレスは、デュアルアクティブ検出プロセスを補足するのに役立ちます。次の **show** コマンドは、各シャーシのベース アドレスの検索方法を示しています。

```
6500-VSS# sh idprom switch 1 backplane detail | inc mac
  mac base = 0019.A927.3000
```

```
6500-VSS# sh idprom switch 2 backplane detail | inc mac
  mac base = 0019.A924.E800
```

各シャーシに位置する上記のベース アドレスは、「VSS デュアルアクティブ スーパーバイザを使用したキャンパス復旧」（P.4-19）で説明するデュアルアクティブ検出方法で使用されます。

MAC アドレスおよび MEC

VSS では、レイヤ 2 の MEC インターフェイスの MAC アドレスは、いずれかのリンクメンバのバーンイン（bia）インターフェイスの MAC アドレスから取得します。通常は、ポートチャネル インターフェイスに追加された最初のインターフェイスの MAC アドレスは、レイヤ 2 のポートチャネル（MEC）インターフェイス用に選択されます。ポートチャネル用に MAC アドレスを使用していたインターフェイスが無効になると、そのポートチャネル インターフェイスは残りのメンバ インターフェイスの MAC アドレスを使用し始めます。ただし、無効になったばかりのインターフェイスが再びアクティブになった場合、レイヤ 2 MEC はそのインターフェイスの MAC アドレスを再利用しません。

レイヤ 2 ポートチャネルの MAC アドレスを継承するプロセスは下記のとおりです。ポートチャネルのバーンイン アドレス（bia）は、ポートチャネルに追加された最初のインターフェイスから取得されます。

```
6500-VSS#show etherchannel summary  | inc 220
220    Po220(SU)        LACP      Gi1/8/1(P)     Gi2/8/1(P)

6500-VSS#show interface gig 1/8/1 | inc bia
  Hardware is C6k 1000Mb 802.3, address is 0014.a922.598c (bia 0014.a922.598c)

6500-VSS#show interface gig 2/8/1 | inc bia
  Hardware is C6k 1000Mb 802.3, address is 0014.a92f.14d4 (bia 0014.a92f.14d4)
```

この出力でポートチャネル インターフェイスの MAC アドレスをギガビット インターフェイス 1/8/1 から取得している箇所に注目してください。

```
6500-VSS#show interface port-channel 220 | inc bia
  Hardware is EtherChannel, address is 0014.a922.598c (bia 0014.a922.598c)


6500-VSS# conf t
6500-VSS(config)# interface gi1/8/1
6500-VSS(config-if)# shutdown
```

レイヤ 2 MEC で自身の MAC アドレスが使用された、リンクメンバのギガビット インターフェイス 1/8/1 が無効になると、このポートチャネルは残りのメンバ（ギガビット インターフェイス 2/8/1）のバーンイン アドレスを使用し始めます。

```
6500-VSS#show interface port-channel 220 | inc bia
  Hardware is EtherChannel, address is 0014.a92f.14d4 (bia 0014.a92f.14d4)

6500-VSS#show interface gig 2/8/1 | inc bia
  Hardware is C6k 1000Mb 802.3, address is 0014.a92f.14d4 (bia 0014.a92f.14d4)
```

このインターフェイスがポートチャネル バンドルに再度追加されてもポートチャネルの MAC アドレスは変更しません。下記の CLI 出力でこの動作を表しています。

```
6500-VSS(config)#interface gig 1/8/1
6500-VSS(config-if)#no shutdown
```

```
6500-VSS#show interface port-channel 220 | inc bia
  Hardware is EtherChannel, address is 0014.a92f.14d4 (bia 0014.a92f.14d4)

6500-VSS#show interface g 2/8/1 | inc bia
  Hardware is C6k 1000Mb 802.3, address is 0014.a92f.14d4 (bia 0014.a92f.14d4)

6500-VSS#show interface g 1/8/1 | inc bia
  Hardware is C6k 1000Mb 802.3, address is 0014.a922.598c (bia 0014.a922.598c)
```

通常、レイヤ 2 MAC アドレスは、BPDU フレームの送信元として使用されます（リンクによる MEC インターフェイスのアクティブ化とそのポートでの STP オペレーションが追加されます）。インターフェイスのメンバ（自身の MAC アドレスをレイヤ 2 MEC が使用）が無効になると、VSS に接続するスイッチで BPDU フレームの送信元 MAC の変更が検出されますが、STP のルート MAC は変更されません。これは STP トポロジに変化がなかったことを示しています。詳細については、「VSS を使用した STP 操作」（P.3-38）を参照してください。

CHAPTER 3

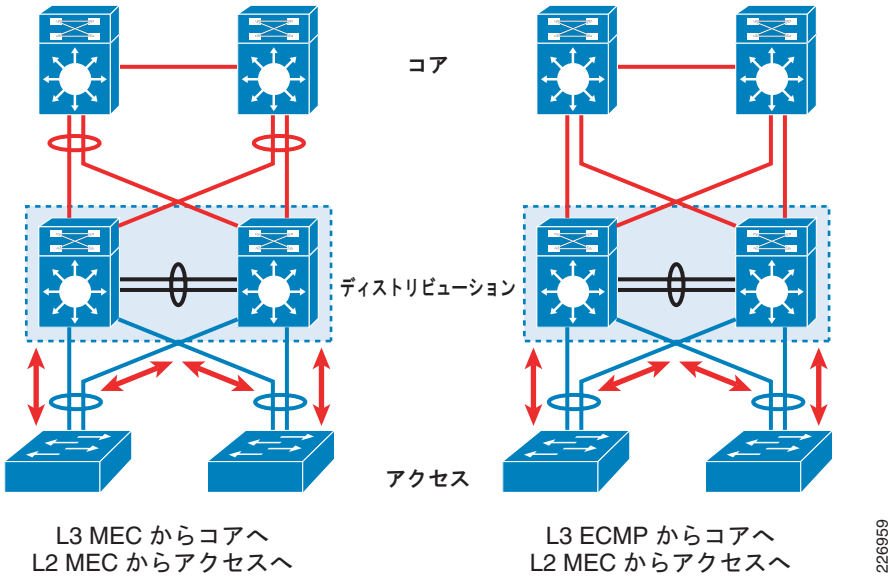
VSS 対応キャンパス デザイン

VSS 対応キャンパス デザインは、このデザイン ガイドの 第 1 章「Virtual Switching Systems のデザインの概要」で説明されている 3 階層のアーキテクチャ モデルと機能設計に準拠します。この章では、特にすべての関連するコンフィギュレーション詳細、トラフィック フロー、障害分析、推奨されるベスト プラクティスに対応するディストリビューション レイヤのキャンパス デザインにおける VSS 実装について説明します。この章は次の主要な項に分けられます。

EtherChannel 最適化、トラフィック フロー、キャンパスにおける VSS を使用した VSL キャパシティ プランニング

伝統的に、複数階層のキャンパス設計コンバージェンス、トラフィック負荷分散、および障害特性は、STP、FHRP、およびトポロジ（ループ型および非ループ型）の 3 つの主要技術要因が中心となっています。VSS 対応キャンパスにおいては、EtherChannel が 3 つの要因すべてに代わり、基本的なビルディング ブロックになっています。レイヤ 2 およびレイヤ 3 では、安定した状態から障害状態時のユーザ データ トラフィックの処理において、EtherChannel アプリケーションが中心的な役割を担います。ディストリビューション レイヤにおける VSS 配置によって、物理トポロジや、コア、ディストリビューション、アクセスといった階層レイヤ間の接続性が変わることはありません。図 3-1 で示されているように、ベスト プラクティスのネットワークはフルメッシュ型のトポロジであり、冗長システムとリンクを保持しています。アクセス レイヤと VSS 間の接続においては、レイヤ 2 MEC が必要であり、キャンパス デザインに不可欠な要素です。ディストリビューション レイヤ（一般的なレイヤ 2 およびレイヤ 3 の境界）の VSS からレイヤ 3 ドメインの接続オプションには、ECMP またはレイヤ 3 MEC の 2 つがあります。コンバージェンス、マルチキャスト フロー、デュアルアクティブ イベントの考慮というコンテキストにおいて、レイヤ 3 MEC オプションは ECMP オプションと比較されます。レイヤ 2 オプションとレイヤ 3 オプションの両方において、VSS ベース環境のあらゆる場所に MEC が存在します。したがって、いずれのシナリオの場合でも、トラフィック フローと障害動作の理解が VSS 対応デザインの MEC に関係してくるため、この理解が非常に重要となります。この項では、VSL バンドルと VSS キャンパス内のトラフィック フローに関連づけられたキャパシティ プランニングについても説明します。これに続くマルチレイヤの項およびルーティングの項では、さまざまな障害状況における推奨ベスト プラクティスを展開するために、これらの情報を使用します。

図 3-1　冗長 VSS 環境



EtherChannel と MEC を使用したトラフィック最適化

MEC は VSS 対応キャンパスの基本です。EtherChannel が作成する論理トポロジは、VSS 環境において、大部分のコンバージェンスとトラフィックの負荷分散を制御します。EtherChannel の負荷分散は、固有性の高いトポロジ、アプリケーション フロー、およびユーザ プロファイルから構成されます。EtherChannel ベースの環境でのトラフィック最適化における主要な概念は、*ハッシュ アルゴリズム*です。一般に、ハッシュ ベースのメカニズムは、数学的な関数に基づいて、トラフィック フローがさまざまなパスの間で統計学的に分配されるように開発されました。次の環境とその環境がハッシュ ベースの最適化の効果に対して与える影響を考えてみます。

- コア デバイスはさまざまなユーザとアプリケーション エンド ポイントからかなりの量のアプリケーション フローを伝送します。これらのフローは一意の送信元 IP アドレス、宛先 IP アドレス、およびポート番号を伝送します。このような*多対多*のフローはハッシュ アルゴリズムに対して有用な入力を提供し、デフォルト設定を使用した負荷分散の効率が向上する可能性があります。
- 一般的にアクセス レイヤとコア間のトラフィック パターンは、*少数対少数*トラフィック パターンから構成されています。これはエンド ホストがデフォルト ゲートウェイと通信するためです。結果として、アクセス スイッチ上のホストからのすべてのトラフィック フローの宛先 IP アドレスが同じになります。これによって、ハッシュ計算における入力のバリエーションの可能性が削減され、最適な負荷分散ができなくなる場合があります。また、トラフィック負荷は非対称です（ダウンストリーム フロー トラフィックがアップストリーム フローよりも高くなります）。

アプリケーション配置におけるバリエーションと使用パターンのため、ハッシュ チューニングによる負荷分散の最適化における*万能*ソリューションはありません。ネットワークを分析し、特定の企業の要件に基づいて最適化ツールをチューニングしなければならない可能性があります。次のガイドラインが適用されます。

- ハッシュ計算において入力として使用する値が多くなるほど、リンク選択においてハッシュ結果の偏りがなくなる可能性が高くなります。
- レイヤ 4 のハッシュはレイヤ 3 のハッシュよりもランダムになる傾向があります。レイヤ 4 アプリケーション ポート番号における多様性のため、入力のバリエーションが多くなることで、より効率的な負荷分散の可能性が高くなる傾向があります。

- すべてのユーザが 1 つのデフォルト ゲートウェイと通信している場合は、レイヤ 2 のハッシュは効率的ではありません。デフォルト ゲートウェイと通信するホストは、宛先と同じ MEC を使用します。結果として、レイヤ 2 ベースのハッシュ入力バリエーションだけが最適化されません。

このデザイン ガイドでは、前述の考慮事項により、あるハッシュ チューニング ソリューションを他のソリューションに対して検証しません。ただし、VSS 対応キャンパス デザインをデプロイするときに、EtherChannel トラフィック最適化に関連する最近の開発を理解して考慮する価値はあります。

Cisco Catalyst 6500 EtherChannel オプション

シスコはさまざまな EthernetChannel 対応システムを提供します。次のオプションはスタンドアロンおよび VSS 対応 Cisco Catalyst 6500 の両方に適用されます。

Adaptive と Fixed

Cisco IOS Release 12.2(33) SXH の時点では、Cisco Catalyst 6500 は、各ポート チャネル メンバ リンクに対して、ハッシュ値をあらかじめ計算する強化されたハッシュ アルゴリズムをサポートします。アダプティブ ハッシュでは、障害が発生したリンクのフローを再ハッシュ計算するために、メンバ リンクを更新する必要が*ありません*。このようにして、パケット損失を低減します。フローはバンドルで利用可能なリンクに対しては動的に再ハッシュ計算されません。この強化されたハッシュ実装は、*アダプティブ ハッシュ*と呼ばれます。次の出力例は、利用できるオプションを示します。

```
6500-VSS(config-if)# port-channel port hash-distribution ?
  adaptive  selective distribution of the bndl_hash among port-channel members
  fixed     fixed distribution of the bndl_hash among port-channel members

6500-VSS(config-if)# port-channel port hash-distribution fixed

This command takes effect when a member link UP/DOWN/ADDITION/DELETION event occurs.
Perform a shutdown and no shutdown command sequences to take immediate effect.
```

デフォルトでは、すべての非 VSL EtherChannel における負荷分散ハッシュ方式は *Fixed* です。リンク メンバ障害時にアップストリーム トラフィック損失を低減するアクセス レイヤにおけるスイッチでは、Adaptive アルゴリズムが効果的です。ただし、VSS 上のアプリケーションまたはコンフィギュレーションは、メンバ シャーシごとに 3 つ以上のリンクがある場合にだけ有効です。これは、2 つのリンクの場合、アルゴリズムによって障害が発生したリンクから、残りのローカル接続されたリンクにフローを回復できる可能性があるためです。各シャーシにはリンクが 1 つある（一般的な構成）があるため、障害が発生したリンクは、同じシャーシ内でメンバ リンクとは見なされない VSL 上のトラフィックを強制します。

ハッシュ変数としての VLAN ID

Sup720-3C および Sup720-3CX 対応スイッチの場合、Cisco Catalyst はハッシュに VLAN 情報を含む混合モード環境をサポートします。キーワードの **enhanced** を **show EtherChannel load-balance** コマンド出力で使用すると、VLAN がハッシュに含まれるかどうかを示します。次の出力例を参照してください。

```
6500-VSS# show platform hardware pfc mode
PFC operating mode : PFC3CXL ! Indicates supervisor capable of VLAN id used as a hash
Configured PFC operating mode : None
```

```
6500-VSS# sh EtherChannel load-balance
EtherChannel Load-Balancing Configuration:
        src-dst-ip enhanced ! Indicates VLAN id used as a hash
EtherChannel Load-Balancing Addresses Used Per-Protocol:
Non-IP: Source XOR Destination MAC address
  IPv4: Source XOR Destination IP address and TCP/UDP (layer-4) port number
! << snip >>
```

VLAN ID は特に次の 2 つの場合において、トラフィックの最適化を向上させるうえで有用です。

- VSS を使用すると、closet-switch ごとの VLAN 数を増やせるため、ハッシュ入力の追加変数とトラフィックを効率的に共有できます。
- たとえば、フロー データの類似性のために、トラフィックで適切なハッシュ計算が行われない場合は、共通のマルチキャスト トラフィックが同じバンドル メンバに対してハッシュ計算されることが多くあります。VLAN ID にはさらに別の差別化要因があります。

ただし、VLAN ハッシュは、VSS の各物理シャーシがアクセス レイヤに対して複数のリンクを持つ場合にだけ有効であることを理解しておくことが重要です。アクセス レイヤに対するシャーシ単位のシングル リンクを使用すると、各メンバ スイッチからの負荷分散は行われません。

ハッシュ チューニングのための任意のレイヤ 3 およびレイヤ 4 演算子

Cisco IOS Release 12.2 (33)SXH では、Cisco Catalyst スイッチはハッシュ計算にレイヤ 3 およびレイヤ 4 情報の両方を含む混合モードをサポートしています。デフォルト オプションは次のリストに太字で示されています。推奨オプションはポインタとしてリストに示されています。

```
VSS(config)# port-channel load-balance ?
  dst-ip                  Dst IP Addr
  dst-mac                 Dst Mac Addr
  dst-mixed-ip-port       Dst IP Addr and TCP/UDP Port <-
  dst-port                Dst TCP/UDP Port
  mpls                    Load Balancing for MPLS packets
  src-dst-ip              Src XOR Dst IP Addr
  src-dst-mac             Src XOR Dst Mac Addr
  src-dst-mixed-ip-port   Src XOR Dst IP Addr and TCP/UDP Port <-
  src-dst-port            Src XOR Dst TCP/UDP Port
  src-ip                  Src IP Addr
  src-mac                 Src Mac Addr
  src-mixed-ip-port       Src IP Addr and TCP/UDP Port <-
  src-port                Src TCP/UDP Port
```

MEC および標準 EtherChannel インターフェイスに対してフロー ハッシュを表示する CLI

特定のフローを監視するためのスイッチ CLI の使用を示す例については、「モニタリング」（P.2-44）を参照してください。CLI コマンドは Cisco Catalyst 6500 プラットフォームでだけ提供されています。

Catalyst 4500 および 3xxx プラットフォーム

EtherChannel ハッシュ チューニング オプションは各プラットフォームによって異なります。Cisco Catalyst 4500 は Cisco Catalyst 6500 と同等の柔軟性を提供し、レイヤ 3 およびレイヤ 4 の演算子を許可します。Cisco Catalyst 36xx および Cisco Catalyst 29xx シリーズ スイッチには、デフォルトのハッシュ設定があり、EtherChannel に対して同等の負荷分散を実現するには十分ではない可能性がある送信元 MAC アドレスを統合します。次のコンフィギュレーション例は、太字でデフォルト値を示し、推奨オプションを矢印で強調表示しています。

Cisco Catalyst 4500 :

```
Catalyst4500(config)# port-channel load-balance ?
  dst-ip        Dst IP Addr
  dst-mac       Dst Mac Addr
```

```
  dst-port        Dst TCP/UDP Port
  src-dst-ip      Src XOR Dst IP Addr
  src-dst-mac     Src XOR Dst Mac Addr
  src-dst-port    Src XOR Dst TCP/UDP Port <-
  src-ip          Src IP Addr
  src-mac         Src Mac Addr
  src-port        Src TCP/UDP Port
```

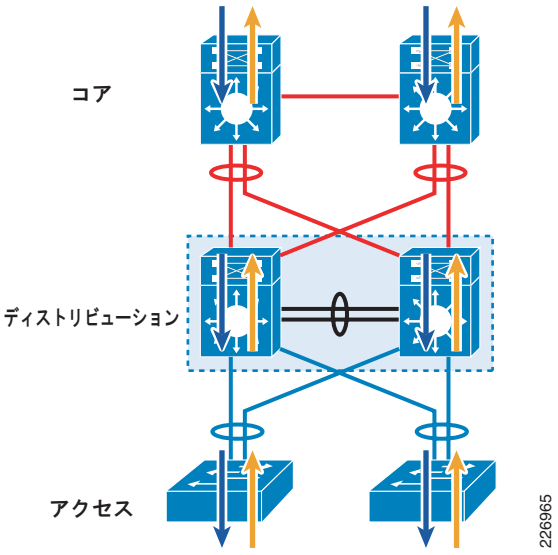
Cisco Catalyst 36xx、Cisco Catalyst 37xx スタック、および Cisco Catalyst 29xx：

```
Catalyst3700(config)# port-channel load-balance ?
  dst-ip        Dst IP Addr
  dst-mac       Dst Mac Addr
  src-dst-ip    Src XOR Dst IP Addr <-
  src-dst-mac   Src XOR Dst Mac Addr
  src-ip        Src IP Addr
  src-mac       Src Mac Addr
```

VSS 対応キャンパスのトラフィック フロー

VSS 環境は、データ転送が常にメンバ シャーシ内で行われるように設計されています。図 3-2 で示されているように、VSS は必ずローカルで利用できるリンク上でトラフィックを転送しようとします。これは、レイヤ 2 リンクとレイヤ 3 リンクの両方に当てはまります。ローカル フォワーディングの主な理由は、VSL リンク上における不要なデータ トラフィック送信を回避し、遅延（VSL 上での余分なホップ）と輻輳を低減することです。図 3-2 には、VSS 環境における通常のトラフィック フローを示します。この環境では、コアとアクセス レイヤの VSS 接続がフルメッシュ型の MEC でイネーブルになっています。このトポロジでは、アップストリーム トラフィック フロー負荷分散決定がアクセス レイヤのレイヤ 2 EtherChannel で制御されます。ダウンストリームの場合は、レイヤ 3 EtherChannel 経由で接続しているコア デバイスで制御されます。この双方向トラフィックの負荷は 2 つの VSS メンバ間で共有されます。ただし、各 VSS メンバという点では、受信および送信のトラフィック フォワーディングは、MEC の一部であるローカル接続されたリンクに基づいています。VSS 対応キャンパス ネットワークにおけるコンバージェンスと障害状態を理解するうえで、このローカル フォワーディングは主要な概念です。

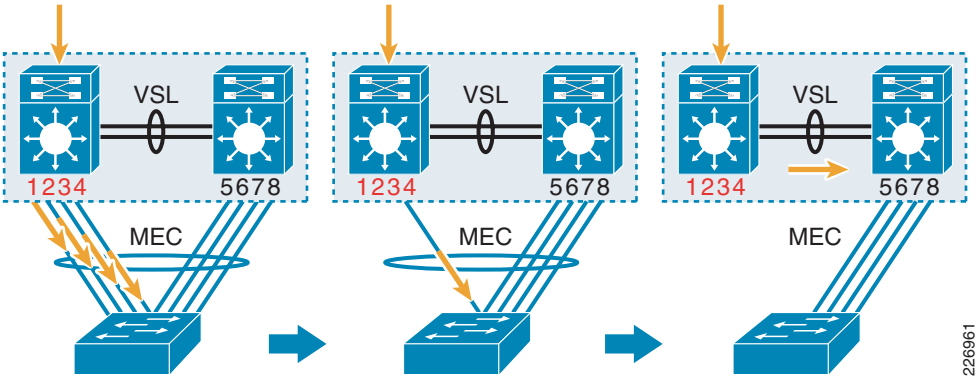
図 3-2　VSS トラフィック フローの概要



レイヤ 2 MEC トラフィック フロー

前述のとおり、通常モードでは、VSS は必ずローカル接続されたリンクを優先します。この点については、図 3-3 のレイヤ 2 MEC で詳細に説明しています。この場合は、トラフィック フロー動作がネットワーク接続の 3 つの異なるステートとして示されています。この例では、4 つのリンクのうち 3 つのリンクが動作していない障害状態（図 3-3 の中央を参照）を示しています。1 つのリンクはまだ VSS メンバに対して動作しているため、他の VSS メンバ スイッチの同じ VSL 経由でアクセスできる EtherChannel グループ内に、別の 4 つのリンクがあるにも関わらず、ダウンストリーム トラフィックはまだこのリンクを選択します。すべてのリンク（1、2、3、および 4）に障害が発生した場合は、VSS システムではこの状態がオーファン接続として検出され、コントロール プレーンが VSL リンク上のすべてのトラフィック フローを再プログラミングし、利用可能な MEC リンク経由でアクセス レイヤに転送します。

図 3-3　　レイヤ 2 MEC メンバ リンク障害時のレイヤ 2 MEC トラフィック フロー



シングル リンクがすべてのトラフィックを送信している障害を図 3-3 の中央に示します。この場合、リンクがオーバーサブスクリプション状態になる可能性があります。ただし、この種類の接続環境は一般的なトポロジではありません。通常は、アクセス レイヤ スイッチは 2 つのアップリンク経由で VSS に接続しています。この場合、シングル リンク障害によって、トラフィックが VSL リンクを横断します。コントロール プレーン トラフィック フローをユーザ データ トラフィック フローと区別することが重要です。コントロール プレーン トラフィックは、いずれかのスイッチを使用して、トラフィックを開始できます。たとえば、UDLD、CDP、あるいはその他のリンク単位で開始する必要があるリンク固有のプロトコルが VSL を横断します。ただし、リモート デバイスに対する *ping* 応答では、必ずピアに接続しているリンクからのローカル パスが選択されます。これは、要求が VSL 上で送信された可能性がある場合でも、リモート ノードがローカル ハッシュ結果に基づいてそのリンクを選択した可能性があるためです。

レイヤ 3 MEC トラフィック フロー

VSS からコアへのレイヤ 3 MEC 接続は、2 つのポート チャネルから構成されます。各ポート チャネルにはリンクが 2 つあり、それぞれが別々の物理シャーシ上にあります。ポート チャネルのリンク メンバのいずれかで障害が発生すると、VSS は別のポート チャネル インターフェイスにある別のローカルで利用可能なリンクを選択して、トラフィックを再ルーティングします。これは、ローカル システム リンクの利用可能性に基づいてパスが選択される ECMP 障害と類似しています。この種類の接続は、ルーティング プロトコル コンフィギュレーションへの依存関係があります。詳細については、「VSS を使ったルーティング」（P.3-46）で説明します。

レイヤ 3 ECMP トラフィック フロー

フルメッシュ型の ECMP トポロジは、VSS 全体の特定の宛先に対して、4 つの異なるルーティング パス（各リンクに 1 つ）から構成されています。ただし、各メンバ VSS は 2 つの一意の Cisco Express Forwarding（CEF）ハードウェア パスに変換される 2 つのパス（2 つのリンク）でプログラムされています。通常の状態の場合は、各メンバ シャーシのアクセス レイヤからコアへのトラフィックでは、2 つのローカルで利用可能なリンク（ハードウェア パス）を使用します。トラフィック フロー動作を示すために、図 3-4 を 3 つのステージに分割します。第 1 ステージ（図 3-4 - (A) を参照）では、受信トラフィックは 2 つの等コスト パス間での負荷分散です。1 つのリンクで障害（図 3-4 – (B) を参照）が発生すると、SW1 の受信トラフィックは残りのリンクを選択します。すべてのローカル リンクで障害（図 3-4 – (C) を参照）が発生する場合は、VSL リンク全体のすべてのフローを別のメンバに転送するように FIB が再プログラムされます。フォワーディング テーブルの出力は、図 3-5 に示され、図 3-4 の障害ステータスに対応しています。

図 3-4　ユニキャスト ECMP トラフィック フローの例

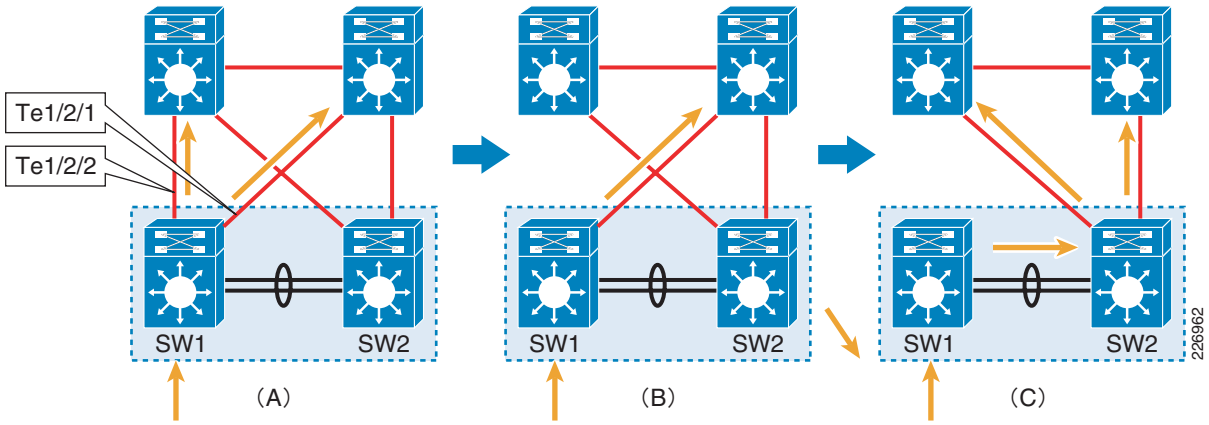
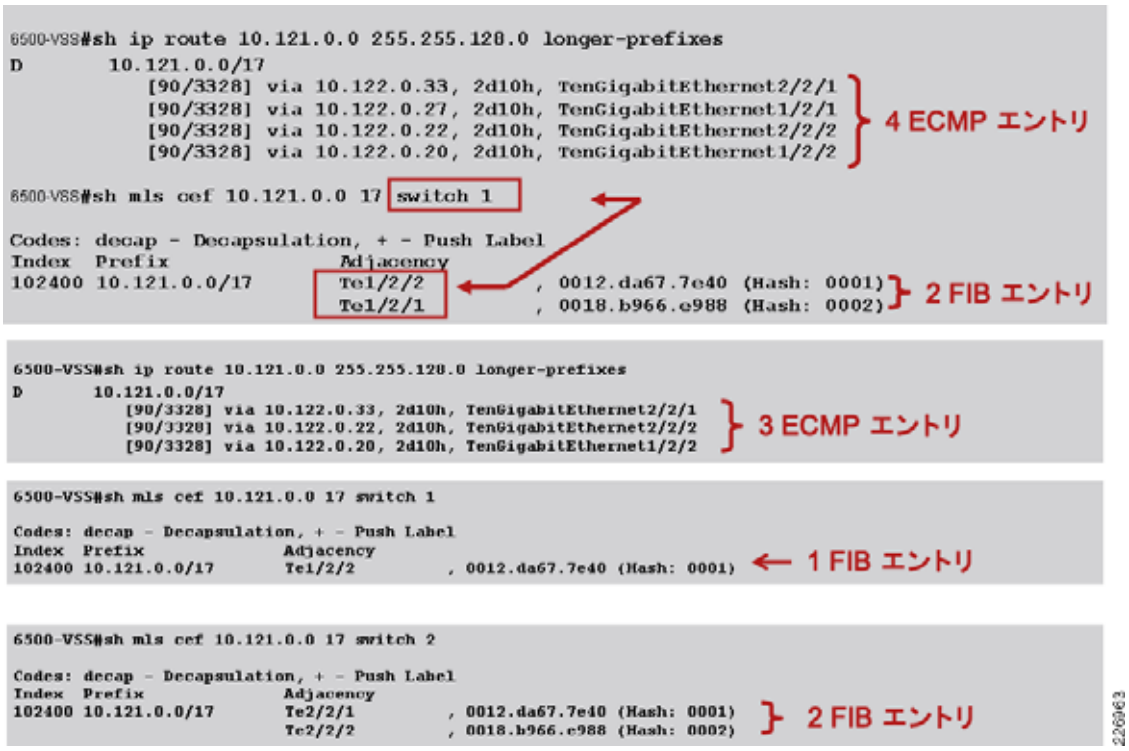


図 3-5　ECMP フォワーディング エントリ：グローバルおよびスイッチ特有



マルチキャスト トラフィック フロー

VSS はスタンドアロン Multicast Multilayer Switching（MMLS; マルチキャスト マルチレイヤ スイッチング）技術の利点と制限事項をすべて共有しています。MMLS によって、デュアル スーパーバイザを使用したマルチキャスト フォワーディング冗長化がイネーブルになります。マルチキャスト フォワーディング ステートには、(*,g) および（s,g）が含まれます。これは、特定のマルチキャスト フローの着信および発信インターフェイス リストがアクティブ スーパーバイザ Policy Feature Card（PFC; ポリシー フィーチャ カード）上で Multicast Entries Table（MET; マルチキャスト エントリ テーブル）でプログラムされていることを示します。このテーブルはホットスタンバイ スーパーバイザで同期化されます。スイッチオーバー中は、マルチキャスト データ フローがハードウェアに転送されます。コントロール プレーンはそのネイバーとの Protocol Independent Multicast（PIM）ネイバー関係を復旧して再確立します。ハードウェアでの複製が必要なユーザ データ トラフィック フローは、VSS フォワーディングに関しては、ユニキャストと同じルールに従います。VSS は常にローカル リンクを優先して、レイヤ 2 ドメインのマルチキャスト トラフィックを複製します。レイヤ 2 関連設計は、「マルチキャスト トラフィックおよびトポロジ デザインにおける考慮事項」（P.3-43）で説明しています。レイヤ 3 ドメインの VSS とのマルチキャスト インタラクションには、マルチキャスト ツリーを構築するときのマルチキャスト コントロール プレーンの動作と、ECMP および MEC ベース トポロジの フォワーディングの差異が含まれます。レイヤ 3 およびマルチキャスト インタラクションは、「VSS を使ったルーティング」（P.3-46）で説明しています。

VSS 障害ドメインおよびトラフィック フロー

この項では、前項で説明したトラフィック フローの動作を使用します。VSS における障害時のトラフィック フローは、ローカル リンクの可用性と、VSS からコアおよびアクセス レイヤへの接続オプションに依存しています。障害の種類は、シャーシ、ノード、リンク、あるいはラインカードのいずれかが考えられます。VSS 障害は大きく次の 3 つのドメインのいずれかになります。

- VSS メンバ障害
- Core と VSS 間の障害（リンク、ラインカード、あるいはノードを含む）
- アクセス レイヤと VSS 間の障害（リンクまたはラインカードを含む）

この項では、MEC ベースのエンドツーエンドである優先接続方式、およびコア、ディストリビューション（VSS）、およびアクセス レイヤにまたがる障害ドメインを使用します。一部の障害は、さまざまなレイヤで複数の異なる復旧をトリガーします。たとえば、VSS のラインカード障害は、コアまたはアクセス レイヤで EtherChannel 復旧をトリガーする場合がありますが、ディストリビューション レイヤ（VSS）で VSL 再ルーティングもトリガーします。このようにして、アップストリーム トラフィックとダウンストリーム トラフィック復旧は、特定の障害に対して非対称になる場合があります。並列でトリガーできる復旧の種類は次のとおりです。

- EtherChannel ベースの復旧
- ECMP またはローカル CEF ベースの復旧
- VSL 上での再ルーティング（VSL バンドル上でのトラフィックの再ルーティングをトリガーする障害）

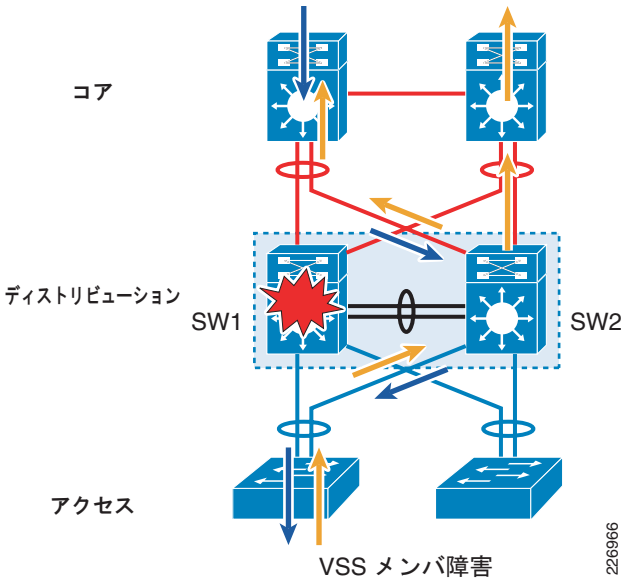
この項では、MEC ベースのトポロジを使用した場合のエンドツーエンド トラフィック フロー動作を説明します。コンバージェンスの項（マルチレイヤおよびルーティング）では、ECMP ベースのトポロジとパケット損失という点での障害の影響を説明します。

VSS メンバ障害

EtherChannel 障害は、VSS に隣接するノードまたは VSS 自体で発生する可能性があります。いずれのケースにおいても、ハードウェアによるリンク停止の検出と、障害が発生したメンバから残りの EtherChannel メンバへのフローの再ハッシュ計算に基づいて復旧が行われます。障害によっては、EtherChannel 障害（復旧）を隣接するノードに限定し、VSS には影響しないようにできます。

図 3-6 に、VSS ノード障害を示します。コア デバイスもアクセス デバイスも MEC 経由で VSS に接続しているため、復旧は EtherChannel に基づいて行われます。双方向のトラフィック（アップストリームおよびダウンストリーム）は、各レイヤの残りの EtherChannel メンバに対してハッシュ計算され、VSS スイッチに転送されます。VSS スイッチがハードウェアのトラフィックを転送し、障害が発生した VSS メンバがアクティブであった場合には、VSS コントロール プレーンが復旧を行います。SSO が利用できる場合には、VSS ではこの処理を行いません。スイッチにはハードウェア ベースの CEF があるため、次のホップとスイッチのリンクが直接隣接ノードに接続していることを把握できます。障害が発生した VSS のメンバがアクティブなスイッチではない場合は、復旧は単純にコアおよびアクセス レイヤの EtherChannel 検出に基づいて行われます。

図 3-6　VSS メンバ障害

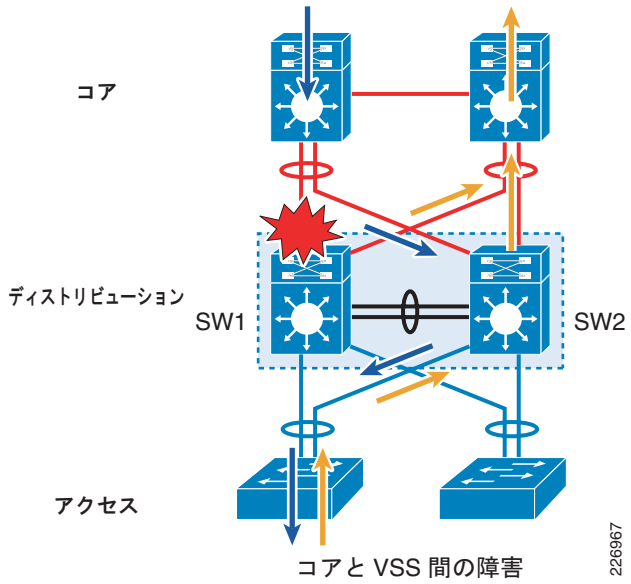


コアと VSS 間の障害

図 3-7 に、VSS とコア ルータ間のポート チャネルの 1 つのリンク メンバで発生した障害例を示します。ダウンストリーム トラフィック復旧は、コア ルータの残りのメンバに対するフローの再ハッシュ計算に基づきます。これは、EtherChannel ベースの復旧です。アップストリーム トラフィック復旧は、VSS でトリガーされるローカル CEF スイッチングのある ECMP に基づいて行われます。図 3-7 で示されている設計では、アップストリーム トラフィック フローが使用する宛先に対して 2 つのルートを通知する 2 つのポート チャネル パス（各コア ルータから 1 つずつ）があります。VSS（シングル論理ルータ）の観点から、ポート チャネル インターフェイスのリンク メンバ障害が発生すると、ルーティング プロトコル コンフィギュレーションによっては、利用可能なルーティング パスが変更されない場合があります。つまり、VSS の観点では、まだ宛先に対して 2 つのルートがあり、それぞれのルートが各ポート チャネルを使用しているということになります。ただし、メンバ スイッチの観点では、ローカル CEF スイッチングがトリガーされます。つまり、ポート チャネル内のリンク損失が代替パスの再選択につながるということになります（たとえば、図 3-7 の各スイッチには 2 つの論理ルート決定されたポート チャネル インターフェイスがあります）。これは、ポート チャネルのいずれかがそのメンバ スイッチへのローカル リンクを持たないために発生します。物理スイッチ（SW1）に代替のローカル接続されたリンクがある場合は、そのパスがパケット フォワーディングに使用され、コンバージェンスは ECMP パスのローカル CEF 隣接更新に基づきます。そうでない場合は、メンバ スイッチは VSL リンク上でトラフィックを再ルーティングします。前述のとおり、復旧はルーティング プロトコル コンフィギュレーションに依存しています。このような依存と VSS とコア間の設計に関する設計選択肢については、「VSS を使ったルーティング」（P.3-46）を参照してください。

VSS メンバのいずれかからコア レイヤへのすべての接続が失敗する場合（ラインカード障害または両方のリンクがディセーブルになっている場合）は、VSS メンバのいずれかから利用できるローカル パスはありません。これによって、コアから VSS へのトラフィックが VSL を横断します。「VSL バンドルのキャパシティ プランニング」（P.3-12）を参照してください。

図 3-7　コアと VSS 間の障害

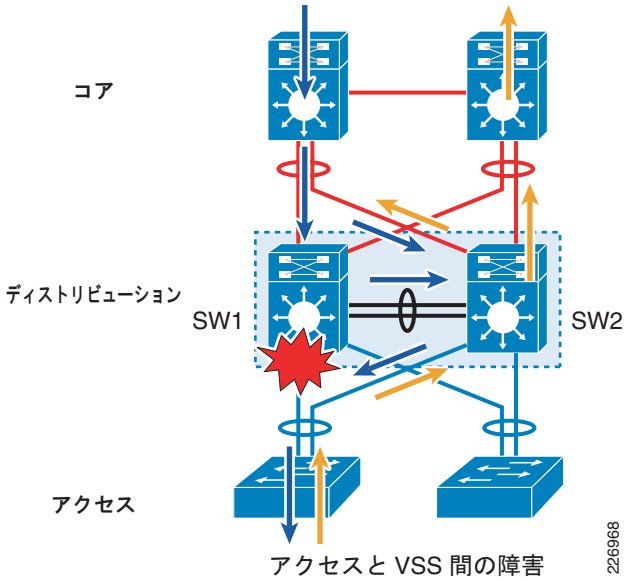


アクセス レイヤと VSS 間の障害

通常は、MEC ベースのトポロジは VSL バンドル上のトラフィックを回避します。ただし、複数の障害状況により、最終手段のパスとして、トラフィックが VSL バンドルを横断することがあります。これは、*オーファン デバイス再ルーティング*と呼ばれます。

コアまたはアクセス レイヤへの接続全体の障害により、VSL 上でのトラフィックの再ルーティングが発生します。また、アクセス レイヤ スイッチから VSS へのリンク障害の場合も、VSL リンク上のトラフィック再ルーティングが発生します。この障害は、図 3-8 で示されています。この場合は、コア ルータはアクセス レイヤでリンク障害が発生していることを認識していません。コア ルータは特定の VSS メンバ（SW1）にダウンストリーム トラフィックを送信し続けます。VSS コントロール プレーンは、アクセス レイヤ スイッチに接続しているローカル リンクに障害が発生したことを検出します。VSS はレイヤ 2 MEC 接続にはまだ SW2 に接続しているメンバが 1 つあることを認識しています。VSS のソフトウェアはこのようなフローを再プログラムして、トラフィックが SW2 に向かって VSL バンドル上を通過し、最終的にはアクセス レイヤ スイッチに到達するようにします。アップストリーム トラフィック復旧は、アクセス レイヤの EtherChannel に基づきます。

図 3-8　VSS へのアクセス障害



アクセスと VSS 間の障害

ある VSS のメンバからアクセス スイッチへのすべての接続で障害が発生した場合は、ダウンストリーム トラフィック復旧として、VSL バンドル再ルーティングが行われます。VSS 上のラインカード全体の障害については、「VSL バンドルのキャパシティ プランニング」（P.3-12）を参照してください。

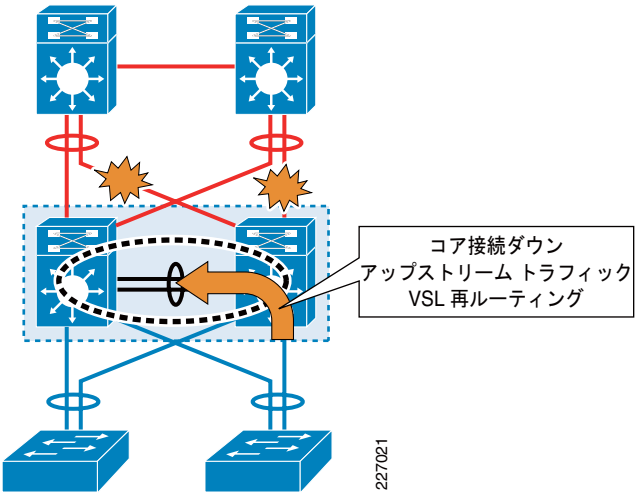
VSL バンドルのキャパシティ プランニング

通常の状態では、VSL バンドルのトラフィック負荷は、ネットワーク コントロール プレーンとシャーシ内コントロール プレーン トラフィックから構成されています。通常の状態では、両方の種類のトラフィック負荷は非常に軽く、厳密な優先度に従って送信されます。VSS のキャパシティ プランニングとリンク サイジングは、障害状態において 2 つのノード間のリンクが計画されたキャパシティと同じトラフィック負荷を伝送できるはずである従来のマルチレイヤ デザインとほとんど同じです。

次のように、2 つの障害点が VSL リンクの最小帯域幅要件を決定します。

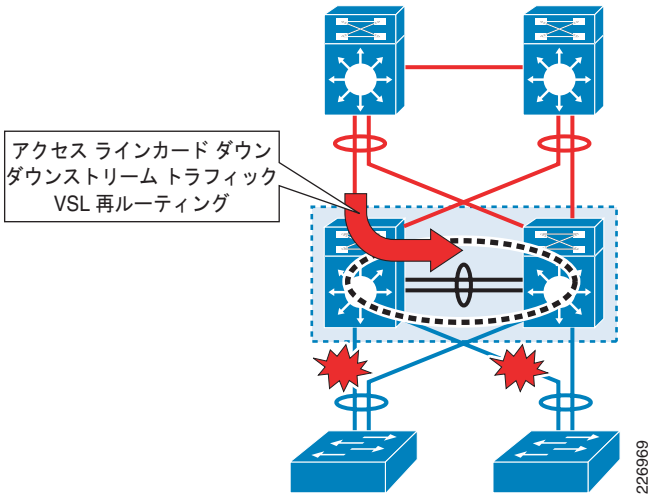
- コアへの VSS のメンバに接続したすべてのアップリンクの障害（図 3-9）。この障害では、すべてのアップストリーム トラフィックが VSL バンドルを横断します。アップリンクの数と速度によって、VSL 上を通過できる最大トラフィック数が制限されます。従来のフルメッシュ型デザインでは、各スイッチ メンバが 20 ギガビットの帯域幅（2 つの 10 ギガビット リンク）を使用します。このため、2 つのリンクを持つレジリエンシー デザインの最小 VSL バンドルで十分になります。

図 3-9　コアへのすべてのアップリンクの障害



- あるスイッチ メンバからアクセス レイヤ スイッチへのすべてのダウンストリーム リンクの障害（図 3-10）。この障害では、すべてのダウンストリームとインターアクセス トラフィックが VSL バンドルを横断します。コア方面へ向かうトラフィックは、アクセス レイヤの EtherChannel メンバ経由で復旧されます。アクセス レイヤ リンクは、コアへの接続で障害が発生していない健全な VSS メンバに接続しているため、VSL を横断する必要はありません。アクセス スイッチからの帯域幅および接続要件は、企業アプリケーションのニーズによって異なります。障害中の本当のトラフィック キャパシティを判断することは難しいことです。ただし、一般的に、すべてのアクセス レイヤ スイッチは、同時にライン速度でトラフィックを送信しません。このため、通常では、インターアクセスのオーバーサブスクリプションがシングル VSS スイッチのアップリンク キャパシティを超えることはありません。主な理由は、一般的に、インターアクセス レイヤの方向の場合よりも、コアの方向（WAN、インターネット、あるいはデータセンター指向）の場合に、アクセス スイッチからのトラフィック フローが高くなるためです。

図 3-10　アクセス レイヤへのすべてのダウンストリーム リンクの障害



いずれの場合も、各スイッチからのキャパシティを伝送する通常のトラフィックは、各スイッチからコアに接続しているリンクによって決まります。これは、各スイッチはローカル接続されたインターフェイスからのトラフィックだけを転送できるためです。このため、最小 VSL バンドルの帯域幅は、少なくとも 1 つの物理スイッチに接続しているアップリンクと同じにするべきです。

次の考慮事項により、VSL リンクの詳細なキャパシティ プランニングが必要です。

- シングルホームのデバイス接続を使用するネットワーク（MEC を使用しない）のデザインでは、少なくとも半分のダウンストリーム トラフィックが VSL リンク上をフローします。この種類の接続を行わないことを強く推奨します。
- あるスイッチ メンバから別のスイッチ メンバへのリモート SPAN。SPAN トラフィックはシングル フローであると見なされるため、トラフィックは特定のリンクのオーバーサブスクリプションにつながりうるシングル VSL リンク上でだけハッシュ計算を行います。トラフィック配信の確立を上げる唯一の方法は、VSL リンクを追加することです。リンクを追加すると、別のリンク上で送信できる SPAN トラフィックを伝送する同じリンク上で、ハッシュ計算された通常のトラフィックを配信する機会が多くなります。
- VSS が FWSM、WiSM、IDS などのサービス ハードウェアを使用している場合は、サービス ブレード経由で渡される予定のすべてのトラフィックを VSL 上で伝送できます。サービス ブレードのキャパシティ プランニングは、このデザイン ガイドの範囲には含まれていないため、このデザイン ガイドでは説明していません。

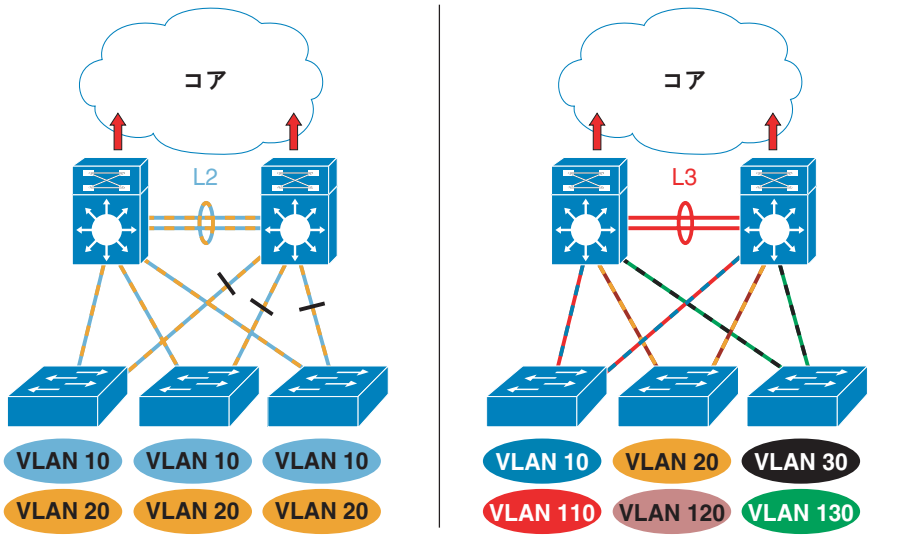
VSS を使用したマルチレイヤ デザインのベスト プラクティス

「ディストリビューション ブロックの VSS」（P.1-3）では、このデザイン ガイドの範囲を説明し、キャンパス ネットワークで最も一般的なマルチレイヤ デザインの概要を説明します。シスコのキャンパス デプロイメントのための高可用性ソリューション オプションの開発は、多くの設計およびチューニングの選択肢（および妥協点）につながっています。この主要な要因は、Voice over IP（VoIP）のサポートが必要なキャンパス ネットワークの常に変化するニーズと、多くのリアルタイム処理アプリケーションの出現です。ネットワークの設計者が自身の状況を評価し、さまざまなデプロイメントの選択と妥協をすると、一般的に 2 つの基礎モデルのうちの 1 つの選択に絞られます。ループ型モデルとループフリー モデルです。これらのモデルは次の項で説明しています。また、これらのモデルを使用して、この設計マニュアルのディストリビューション ブロックの VSS のアプリケーションを例示します。

マルチレイヤ デザイン最適化と制限事項の概要

図 3-11 に、ループ フリーおよびループ型マルチレイヤ設計の比較を示します。

図 3-11　ループ型およびループ フリー マルチレイヤ設計の比較



ループ型トポロジ
すべての VLAN がすべてのアクセス スイッチをスパン

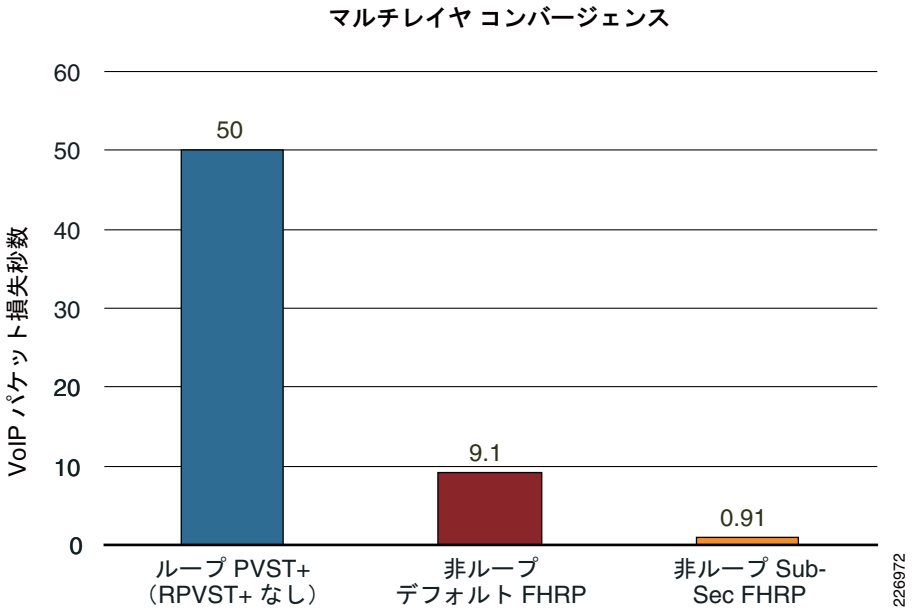
非ループ型トポロジ
VLAN = サブネット = クローゼット

表 3-1 に、ループ型および非ループ型デザイン環境の概要を示します。いずれのデザイン方式も複数の制御プロトコルを使用し、チューニングおよびコンフィギュレーション オプションを一貫した方法で適用し、レジリエンシー設計を行います。また、表 3-1 で説明し、図 3-12 で例示しているコンバージェンスでは、Sub-Sec コンバージェンスでは、First Hop Routing Protocol（FHRP）、HSRP/GLBP/VRRP タイマー チューニング、および VLAN が複数のアクセス スイッチをまたがないようにするトポロジ制約が必要になることを示しています。さらにチューニングが必要となる追加の注意事項およびプロトコル動作もあります。いずれかの設計において、Sub-Sec コンバージェンスでは、すべてのプロトコルとレイヤが連携して動作するように密接に関係した設計を行う必要があります。

表 3-1　ループ型および非ループ型マルチレイヤ設計の概要

ループ型トポロジ	非ループ型トポロジ
少なくとも一部の VLAN が複数のアクセス スイッチをまたがっています レイヤ 2 ループ ディストリビューション間のリンク上で実行されるレイヤ 2 および 3 ブロックされたリンク	各アクセス スイッチには 一意の VLAN があります レイヤ 2 ループはありません ディストリビューション間のレイヤ 3 リンク ブロックされたリンクはありません
アプリケーション	
ユーザ アプリケーションでは、アクセス スイッチ全体でレイヤ 2 接続が必要です。 新しいビジネスの課題を解決する NAC、ゲスト ワイヤレスといった最新の技術を採用 追加および変更の動作における柔軟性 効率的なサブネットの使用	高可用性アプリケーション要件：VoIP、トレーディング フロア ループの暴露を排除 HSRP 経由のコンバージェンスの制御 副作用の削減
最適化要件	
HSRP およびルート マッチング 手動 STP トポロジ メンテナンスによる負荷分散 ユニキャスト フラッディングの緩和：MAC および ARP タイマー チューニング コンフィギュレーション チューニング：トランキング、EtherChannel など STP：RPVST+ および MST STP ツールキット：Root Guard、Loop Guard、BPDU Guard、ポート セキュリティ ブロードキャスト制御 STP ツールキット：Root Guard、Loop Guard、BPDU Guard、ポート セキュリティ ブロードキャスト制御	基本 STP 保護：BPDU Guard、ポート セキュリティ HSRP タイマー チューニング FHRP グループ経由での負荷分散 トランク コンフィギュレーション チューニング レイヤ 3 集約コンフィギュレーション
コンバージェンス	
PVST：最大 50 秒 RPVST + FHRP（デフォルト タイマー）：10 ～ 11 秒 その他のバリエーションもあります	FHRP デフォルト：10 秒 FHRP でチューニングされたタイマー：900 ミリ秒 その他のバリエーションもあります

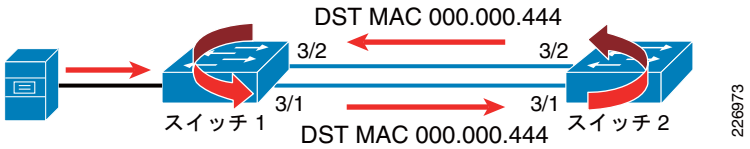
図 3-12　マルチレイヤ コンバージェンス比較



スパニング ツリー プロトコルによるループ ストーム状況

Spanning Tree Protocol（STP; スパニング ツリー プロトコル）は BPDU を使用して代替パスをブロックするため、ループ フリー トポロジをイネーブルにできます。ただし、STP がブロックするポートを判断できず、結果としてループ フリー トポロジを決定できないということになる可能性があります。通常、この問題は、BPDU が見つからなかったり、破損していたりするために発生します。結果として、多くのデバイスがアクティブとなり（リンクがフォワーディング ステートに移行）、ループ フリー パスを検出します。BPDU イベントの損失が解決されると、トポロジ検出処理が終了します。ただし、BPDU の損失が続く場合は、BPDU が継続的に循環し、各 STP 対応ポートがループ フリー パスを検出しようとする状態を停止する固有のメカニズムはありません。図 3-13 を参照してください。

図 3-13　ループ状況の一般的な例



BPDU ストームを停止する唯一の方法は、ネットワーク デバイスを 1 台ずつシャットダウンし、注意しながら一度にデバイスを再導入してネットワークをバックアップにすることです。ループはアクセス レイヤのユーザ活動によって発生する可能性があるため、ループはループ型トポロジでも、非ループ型トポロジでも発生する可能性があります。ただし、ループ型設計の場合にこの問題が発生しやすい主な理由は、特定の VLAN トポロジで利用できる論理パスの数が多いためです。

次の問題は、STP がブロックできない可能性があるループを示しています。

- BPDU の不足につながる障害状態のハードウェア（GBIC、ケーブル配線、CRC など）
- 高 CPU 利用を引き起こし、BPDU 処理を妨げる障害状態のソフトウェア
- BPDU ブラック ホールを引き起こすフォワーディング リンク上の BPDU フィルタなどのコンフィギュレーション エラー

- 標準以外のスイッチ実装（BPDU を吸収しても送信しない、あるいは BPDU の廃棄）
- BPDU の不足につながるラップトップ ネットワーク経由でユーザが作成したトポロジ

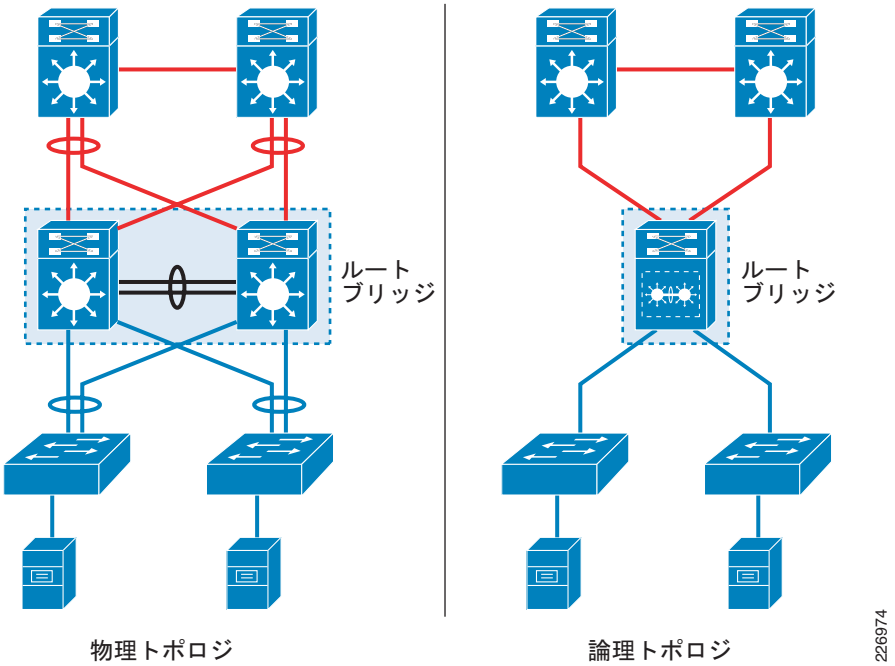
VSS の利点

マルチレイヤ キャンパスのディストリビューション レイヤにおける VSS アプリケーションは、従来のマルチレイヤ設計と比較して次の明らかな利点を持つトポロジを作成できるように設計されています。

- MEC および統合化コントロール プレーンを使用したループ フリー トポロジ
- デフォルト ゲートウェイ（HSRP、GLBP、または VRRP）のコンフィギュレーションおよび Sub-Sec コンバージェンスを実現するためのチューニング要件が不要
- EtherChannel を使用したトラフィック フローによるビルトインの最適化
- シングル コンフィギュレーション管理：ノードの統合
- レイヤ 2 ベースの接続を必要とするサービス統合の実現
- チューニングやコンフィギュレーションによる複雑な処理が不要な Sub-Sec コンバージェンス

VSS はディストリビューション ブロックで適用され、物理および論理トポロジは、図 3-14 で示されています。シングル論理ノードと MEC の組み合わせにより、STP ではスター型のトポロジを利用できます。このトポロジには代替パスがなく、ループ フリー 設計が作成されるため、デュアル リンク、デュアル ノード設計の冗長性に影響しません。詳細は、第 2 章「Virtual Switching System 1440 アーキテクチャ」を参照してください。

図 3-14　　物理および論理トポロジ



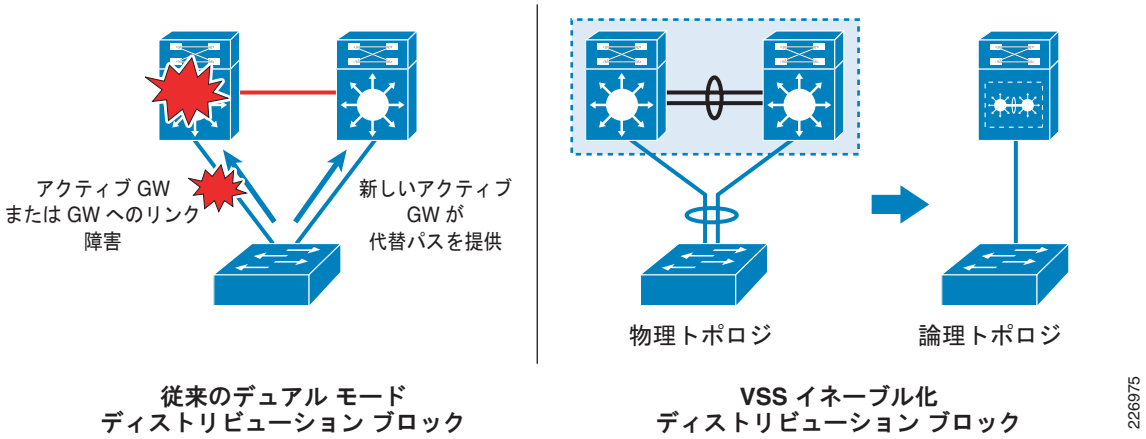
FHRP コンフィギュレーションの排除

図 3-14 に示すように、VSS トポロジはディストリビューション レイヤで 2 つの論理ノードを置き換えます。このトポロジによって、デフォルト ゲートウェイ冗長性の要件がなくなります。これは、デフォルト ゲートウェイが、インターフェイス VLAN IP アドレスが物理シャーシの両方で利用できるシングル論理ノードで置き換えられるためです。デフォルト ゲートウェイ冗長性のコンバージェンス動作は、SSO と EtherChannel で置き換えられます。したがって、FHRP 最適化や Sub-Sec チューニングにおける複雑性が必要ではなくなります。

VSS はエンド ステーションに対する 1 つの復元力に優れたデフォルト ゲートウェイまたはファースト ホップ アドレスとなります。非 VSS 環境では、FHRP プロトコルが冗長化ツールとして機能し、ディストリビューション スイッチまたはアクセス レイヤ リンク障害を含む複数障害に対する保護を提供します。非 VSS トポロジでは、Cisco Unified Communications の Sub-Sec コンバージェンス要件に対応するために、FHRP の最適化が必要になります。Sub-Sec コンバージェンスと負荷分散要件に対応するために、HSRP、GLBP、および VRRP コンフィギュレーションをチューニングする場合は、HSRP、GLBP、および VRRP コンフィギュレーションがかなり複雑になる可能性があります。コンバージェンスの向上に必要な最適化は次のとおりです。

- FHRP hello の Sub-Sec タイマー コンフィギュレーション
- プリエンプトおよびスタンバイ ディレイ コンフィギュレーション
- ループ型トポロジにおける STP コンバージェンスへの依存
- プラットフォーム依存と FHRP の Sub-Sec タイマーを処理する CPU キャパシティ

図 3-15　デフォルト ゲートウェイ冗長性としての FHRP の排除



さらに、FHRP を使用してアップストリーム トラフィックの負荷分散を最適化する場合は、次が必要になります。

- 各ディストリビューション ノードで定義された HSRP グループと、2 台のルータ上のディストリビューションによるアクティブ FHRP およびセカンダリ FHRP の調整。
- GLBP を使用することで、自動アップリンク ロード バランシングが容易になります。ただし、デフォルト ゲートウェイに対する代替 MAC アドレス割り当てが行われるため、ループ型トポロジにおいては最適な方法ではありません。

キャンパスに VSS を導入すると、上記のすべての必須コンフィギュレーションの複雑性が排除されます。デフォルト ゲートウェイ冗長性として使用される HSRP/GLBP の排除に関して、実際上の問題が発生します。デフォルト ゲートウェイ IP アドレスの MAC アドレスは一意であり、HSRP/GLBP と一致しています。VSS 対応のキャンパスでは、VLAN インターフェイス IP がデフォルト ゲートウェイとなります。デフォルト ゲートウェイ IP アドレスは同じ（エンド ホストへの変更を避けるため）で、

一般的には VLAN インターフェイスに伝送されます。VLAN インターフェイス MAC アドレスは HSRP あるいは GLBP MAC アドレスと同じではありません。VLAN インターフェイス MAC はシステムで生成されたアドレスです（詳細については、「MAC アドレス」（P.2-45）を参照してください）。一般的に、gratuitous ARP が発行され、の IP アドレスは変わりませんが、MAC アドレスは変更されます。この MAC アドレスの変更により、エンド ホストが動作していない場合や、エンド ホストにデフォルト ARP エントリの更新を禁止するコンフィギュレーションがある場合に、トラフィックが停止する可能性があります。一般的に、エンド ホストはデフォルト ゲートウェイの ARP テーブル エントリを 4 時間キャッシュに保存します。

この問題に対する考えられる解決策は、ネイバーのない VSS に HSRP/GLBP コンフィギュレーションを伝送することです。デフォルト ゲートウェイの MAC アドレス更新を回避するためだけに HSRP/GLBP を保持することは、理想的なベスト プラクティスではありません。これは、アクティブなスイッチの障害時のデフォルト ゲートウェイ復旧は、HSRP/GLBP が SSO ベースの復旧中に初期化できる速度に依存しているためです。したがって、考えられる代替案の 1 つは、次に示すように、VLAN インターフェイス上のデフォルト ゲートウェイ IP アドレスを使用し、一時的に同じグループ ID を使用して HSRP/GLBP コンフィギュレーションを設定することです。

```
Interface Vlan200
 ip address 10.200.0.1 255.255.255.0 <-- old HSRP IP
 standby 200 ip 10.200.0.2 <--- old HSRP group id#, but new IP address
```

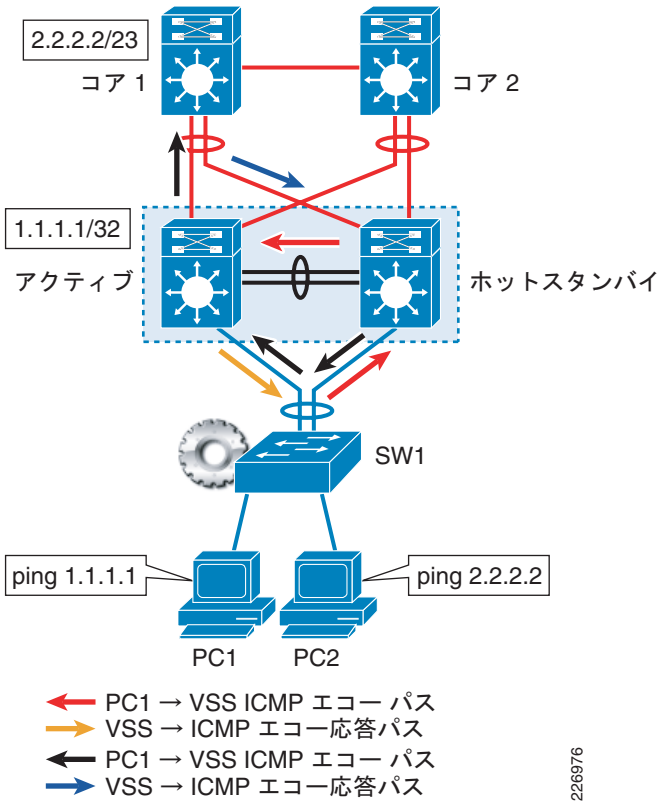
上記のコンフィギュレーションによって、Vlan200 SVI は HSRP グループ 200 MAC アドレスを所有できるようになり、HSRP グループがデフォルト ゲートウェイ IP アドレスにリンクしていないため、HSRP グループへの依存は作成されません。VSS への移行後、ホストはフレームを HSRP MAC アドレスに送信し続けます。時間が経過するにつれ、パケットはこれらのホスト宛ての VSS を入力するため、VSS は関連するホストに ARP 要求を送信します。この ARP 要求は、デフォルト ゲートウェイ IP アドレスに対するホストの ARP エントリを固定するため、新しい MAC アドレスを指すようになります。ARP 更新をトリガーできるように VSS からのトラフィックがないホストの場合でも、4 時間以内に ARP テーブルが更新されるため、VSS の新しい IP アドレスが決定されます。

約 4 時間が経過したら、まだ古い MAC アドレスを使用しているホストがある可能性は非常に低いため、すべての SVI から安全に HSRP コンフィギュレーションを削除できます。安全のため、この時間を延長したり、各サーバの ARP テーブルを確認するスクリプトを顧客側で実行してから、HSRP/GLBP コンフィギュレーションを削除したりできます。

デフォルト ゲートウェイへのトラフィック フロー

図 3-16 に、デフォルト ゲートウェイ経由でのエンド ホストからの ping のフローを示します。ICMP トラフィックのアップストリーム パスは、ハッシュ決定に基づいてアクセス レイヤのスイッチで選択されます。ハッシュ結果がホットスタンバイ スイッチに接続しているリンク選択となった場合は、パケットは VSL リンクを横断して、アクティブなスイッチに到達して応答します。ユーザ データ トラフィック フォワーディングの場合、VSS は常にローカル パスを優先するため、VSS からの応答には必ずローカル リンクを使用します。ping が VSS から実行された場合は、まず VSS はローカル パスを選択し、次に応答側がいずれかのリンクを選択する可能性があります。VSS を横断する ping またはデータ トラフィックは、「VSS 対応キャンパスのトラフィック フロー」（P.3-5）に示されているように、通常のフォワーディングに従います。

図 3-16　ICMP エコー応答トラフィック フロー



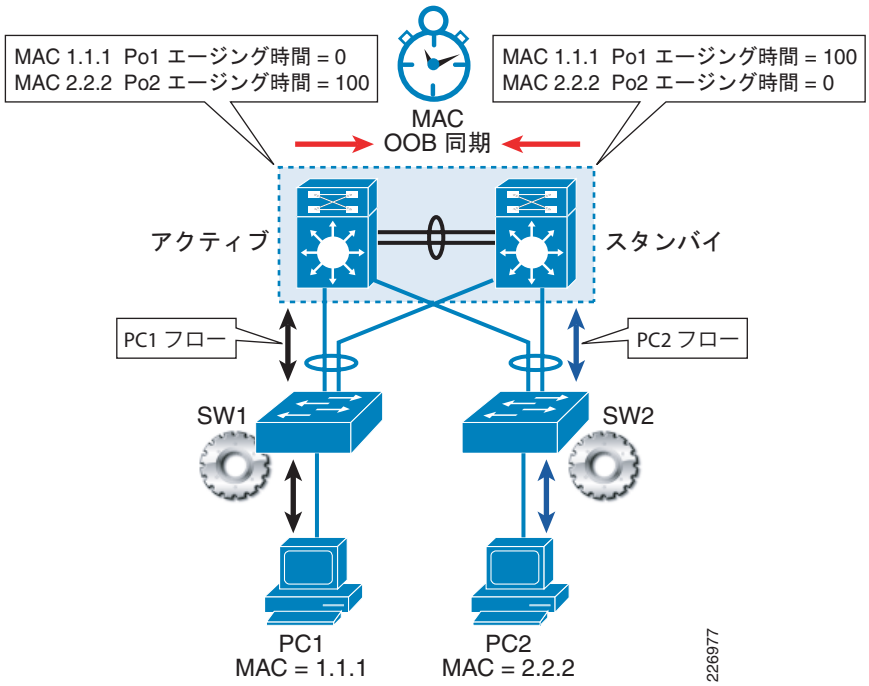
MEC トポロジを使用した VSS におけるレイヤ 2 MAC 学習

スタンドアロンの導入においては、VSS スイッチ メンバは独立してハードウェア ベースの送信元 MAC アドレス学習が採用されます。VSS ではマルチシャーシ分散型フォワーディングも可能です。分散型スイッチングでは、Distributed Feature Card（DFC）は独自の Content-Addressable Memory（CAM; 連想メモリ）テーブルを保持します。これは、各 DFC が MAC アドレスを学習し、その特定のエントリの CAM のエージング時間とトラフィック一致に基づいて、MAC アドレスの期間を決定します。アクティブなタイマーとエージング（アイドル）タイマーを保持するために、VSS はスタンドアロン導入時と同じタイマーに従います。フォワーディング テーブルの動的 MAC アドレス エントリには次のモードがあります。

- *アクティブ*: スイッチが同じ送信元 MAC アドレスからのネットワーク上で、アクティブにトラフィックを切り替えているときには、スイッチは動的 MAC エントリを *アクティブ*なエントリと見なします。スイッチは特定の送信元 MAC アドレスからトラフィックを受信するたびに、エージング タイマーを 0 秒に設定します。
- *アイドルまたはエージング*: この MAC エントリはフォワーディング テーブルに格納されますが、MAC に対するアクティブなフローはありません。デフォルトでは、300 秒後に、アイドルの MAC エントリがレイヤ 2 のフォワーディング テーブルから削除されます。分散型スイッチングでは、通常、スーパーバイザ エンジンに特定の MAC アドレスがしばらく表示されません。したがって、エントリは時間切れになります。現在では、異なるフォワーディング エンジン間で CAM テーブルの一貫性を維持するメカニズムが 2 つあります。ライン モジュールにある DFC とスーパーバイザ モジュールにある PFC です。

- *Flood-to-Frame（FF）*：このハードウェア ベースの学習方式は、新しい MAC アドレスがラインカード付与されるたびにトリガーされます。MAC アドレスは分散化方式でフォワーディング テーブルに追加されます。最初に MAC アドレスを学習するラインカードまたはポートは、*プライマリ エントリ*または*送信元ラインカード*と呼ばれます。
- *MAC Notification（MN）*：これは非プライマリ エントリ ラインカード上の MAC アドレスを追加または削除して、システム内のユニキャスト 継続的なフラッディングを回避するためのハードウェア ベースの方式です。MAC 宛先へのトラフィックが DFC ラインカード レベルで示される場合は、システム内の MAC アドレスの場所に関する情報がないため、まずそのフレームをシステム全体にフラッディングさせます。プライマリ エントリ ラインカードがフラッディングしたフレームを受信するとすぐに、+MN を送信して、このトラフィックの送信元の DFC ラインカード上に MAC エントリを追加します。DFC ラインカードのエントリが時間切れになった場合は、-MN を使用して、DFC ラインカードからエントリを削除します。図 3-17 を参照してください。

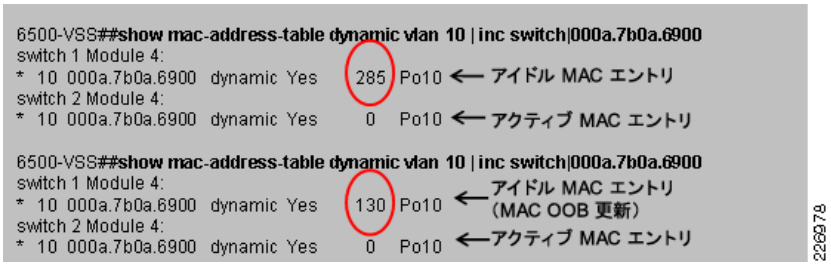
図 3-17　　MAC 通知



- *MAC Out-of-Band Sync（OOB）*：通常の状態では、トラフィックはシャーシ単位で VSS に入り、出て行きます（「VSS 対応キャンパスのトラフィック フロー」（P.3-5）を参照）。つまり、一般的なフローは 1 台のシャーシの MAC エントリだけを更新することになります。図 3-17 に、アクセス レイヤでハッシュ計算を行う EtherChannel に基づいて、PC1 フローが SW1 を選択する例を示します。PC1 MAC エントリのエージングが SW2 で始まります。同様に、PC2 MAC エントリも SW1 で時間切れになります。アイドル時間になると、MAC アドレスが非プライマリ ラインカード、ピア シャーシ PFC、およびそのラインカード上で時間切れとなります。このようなラインカードにトラフィックがある場合は、システム全体にフラッディングする必要があります。また、MEC（VSS の必須コンポーネント）が分散 EtherChannel モードで動作している可能性があり、MAC アドレスがさまざまなラインカードで時間切れになる確率が上がります。DFC または PFC のエントリの時間切れを防止するために、MAC OOB ソフトウェア プロセス メカニズムは、MAC アドレスに対するトラフィックがない場合でも、すべてのラインカードと PFC において定期的にアクティブな MAC エントリを更新します。MAC OOB は、アクティブな MAC エントリが VSS（およびスタンドアロン システム）内のいかなる場所でも時間切れにならないように設計されています。プライマリ エントリ モジュールだけがアクティブな MAC エントリを同期化します。アイドルの MAC エントリは同期

化されず、独立して時間切れになります。図 3-18 に、MAC エージングと MAC OOB で更新されたエントリを表示するために必要な CLI を示します。最初の CLI 出力で示されているように、SW2 モジュール 4 のエージング時間がゼロになるときに、アクティブな MAC エントリがあります。MAC がデフォルト タイマーの 480 秒に対してエージングしている図 3-18 に示されるように、フローは SW2 に対してハッシュ計算されるため、SW1 上の同じ MAC エントリのエージングが開始します。2 番目の CLI 出力は、OOB プロセスが、SW1 モジュール 4 の MAC エントリがリセットした MAC エントリを同期化した後に行われます。OOB がない場合は、SW1 モジュール 4 上の MAC エントリが時間切れになり、一時的なユニキャスト フラッディングを引き起こす可能性があります。

図 3-18　　MAC OOB Synchronization



（注）　仮想スイッチ ノード間の MAC 同期化処理は、VSL EtherChannel 上および特に VSL 制御リンク上で行われます。

Out-Of-Band Synchronization コンフィギュレーションの推奨

次の CLI は OOB をイネーブルにするために使用されます。デフォルトの MAC OOB 間隔は 160 秒です。MAC OOB Synchronization は、3 つのアクティビティ間隔で、すべてのモジュールでアクティブな MAC エントリのエージング時間を更新するようにプログラムされています。アイドルの MAC エージング タイマーは 480 秒（MAC OOB 間隔 × 3 つのアクティビティ間隔）に設定する必要があります。

```
VSS(config)# mac-address-table synchronize activity-time ?
  <0-1275>  Enter time in seconds <160, 320, 640>
% Current activity time is [160] seconds
% Recommended aging time for all vlans is at least three times the activity interval

6500-VSS# show mac-address-table synchronize statistics
MAC Entry Out-of-band Synchronization Feature Statistics:
-----------------------------------------------------------
Switch [1] Module [4]
---------------------
Module Status:
Statistics collected from Switch/Module   : 1/4
Number of L2 asics in this module         : 1

Global Status:
Status of feature enabled on the switch   : on
Default activity time                     : 160
Configured current activity time          : 480
```

MAC OOB Synchronization アクティビティ間隔設定は、システム全体で適用されます。ただし、各モジュールは独立して個々のエージングを管理します。

注意　Cisco IOS Release 12.2(33)SXI よりも前は、Cisco Catalyst 6500 システムで MAC OOB Synchronization がイネーブルになっている場合は、RP 上のデフォルトのアイドル MAC エージング タイマーが 300 という不正なエージング時間を示しています。ただし、SP および DFC モジュールでは正しい値の 480 秒を示しています。この問題は、これより後のリリースでは、ソフトウェア バグ（CSCso59288）によって解決されました。

（注）　WS-6708-10G が VSS システムにある場合は、MAC 同期化は自動的にイネーブルになります。そうでない場合は、MAC 同期化を手動でイネーブルにする必要があります。

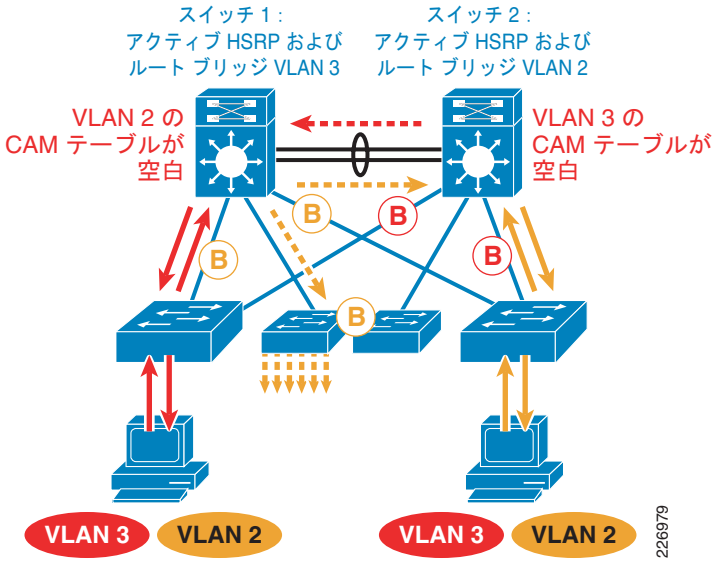
（注）　デフォルトでは、6708 モジュールを含む Cisco Catalyst 6500 システム上の動的 MAC エントリ エージング時間は 480 秒に設定されています。非 6708 DFC モジュールの Cisco Catalyst 6500 がスイッチにある場合は、手動で MAC エージング タイマーを 300 秒から 480 秒に変更する必要があります。IOS Release 12.2(33) SXI 以降は、デフォルト アイドル MAC エージェント タイマーは自動的に 480 秒に設定されます。

ヒント　**シスコでは、デフォルト MAC OOB Synchronization アクティビティ間隔の 160 秒（設定可能最小値）と、デフォルト MAC OOB Synchronization アクティビティ間隔の 3 倍の値のアイドル MAC エージング タイマー（480 秒）をイネーブルにして、保持することをお勧めします。**

非対称フォワーディングおよびユニキャスト フラッディングの排除

アップストリームおよびダウンストリーム フローに非対称フォワーディング パスが含まれる場合には、不明なユニキャスト フラッディングが発生します。非対称フォワーディング パスはスタンドアロン設計で作成されます。この場合、特定の送信元 MAC に対するアップストリーム トラフィックは、必ずデフォルト ゲートウェイへと伝送されます。一方で、ダウンストリーム トラフィックは、ディストリビューション レイヤのゲートウェイに到達するコア レイヤのルータによって負荷分散を行います。送信元 MAC がデフォルト ゲートウェイに対する最初の ARP 検出を送信し始めると、両方のディストリビューション ルータは MAC を学習し、ARP と MAC 間のマッピングが作成されます。CAM タイマーは 5 分で時間切れになり、APR エントリは 4 時間で時間切れになります。アップストリーム トラフィックはディストリビューション ノードのいずれかでだけルーティングされるため、その MAC の CAM タイマーは ARP エントリが *スタンバイ* ディストリビューション ルータにあるときに時間切れになります。スタンバイ ルータがその MAC アドレス宛てのトラフィックを受信するときに、ARP エントリはレイヤ 2 のカプセル化を提供しますが、対応する CAM エントリは CAM テーブルに存在しません。すべてのレイヤ 2 デバイスにおいて、トラフィックは、VLAN のすべてのポートに対してフラッディングする必要のある不明なユニキャストとして認識されます。この問題は、図 3-19 で例示されています。ここでは、2 台のディストリビューション ルータに、対応する VLAN の空の CAM テーブルがあります。VLAN 3 および VLAN 2 デバイスが相互に通信する場合は、デフォルト ゲートウェイ経由で行われます。各デフォルト ゲートウェイでは、このトラフィックは不明なユニキャストとして処理されます。非ループ型トポロジでは、1 つのフレームだけがディストリビューション ルータ間のリンク上でフラッディングします。ただし、ループ型トポロジの場合は、関連する VLAN が存在するすべてのアクセス レイヤのスイッチに、この不明なフレームを送信する必要があります。伝送量が多いフローの種類の場合（FTP、ビデオなど）は、これによってエンド デバイスに負荷がかかり、基幹アプリケーションの応答時間が極度に遅くなる可能性があります。大部分のネットワークでは、この症状は存在しますが、ネットワーク レベルでは何も示されないため、ネットワーク運用時には認識されません。

図 3-19　空の CAM テーブルの例

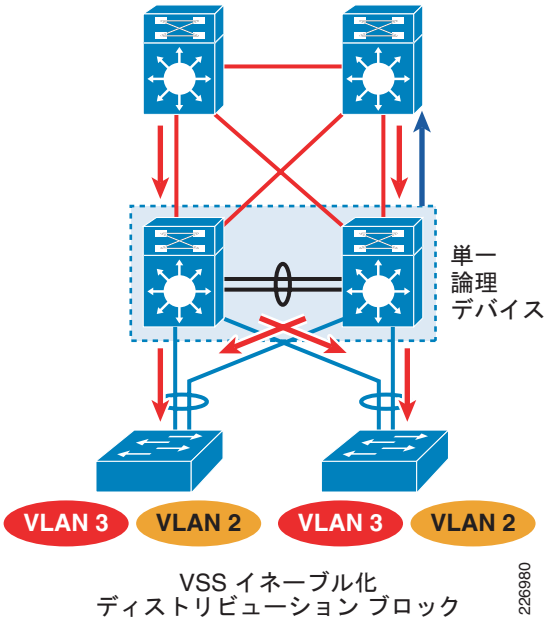


ユニキャスト フラッディングは、VLAN が複数のアクセス レイヤ スイッチをまたがる場合に、より顕著になります。VLAN トラフィックが存在するすべてのポートで、トラフィックのフラッディングを停止する固有の方法はありません。一般的には、ユニキャスト フラッディングを低減するための 3 つの方法があります。

- アクセス スイッチごとに一意の音声およびデータ VLAN がある非ループ型トポロジを使用します。それでもユニキャスト フラッディングが発生する可能性がありますが、VLAN がローカルに置かれるためユーザへの影響が抑えられます。
- ARP タイマーを 270 秒に調整し、CAM タイマーをデフォルトの 5 分にします。このようにすると、必ず ARP が CAM タイマーの前にタイムアウトし、対象のトラフィックの CAM を更新します。10,000 を超える ARP エントリを含むネットワークの場合は、CAM タイマーを選択して、デフォルト ARP タイマーと一致するようにタイマーの値を上げます。これで、ARP ブロードキャストによって発生する CPU スパイクが低減されます。
- 前述のアプローチでは、トポロジまたは追加コンフィギュレーションの選択が必要になります。VSS にはビルトインのメカニズムがあり、このような制限事項に影響されずに、CAM のタイムアウトに関連するユニキャスト フラッディングを回避するようになっています。VSS は 1 つの論理ルータ トポロジをイネーブルにして、図 3-20 に示されるように、デュアルホーム接続トポロジにおけるフラッディングを回避します。ユニキャスト フラッディングを回避するために、両方のメンバ スイッチが SSO 経由で ARP テーブルを同期化し続けます。これは ARP が SSO を認識しているためです。VSS の両方のメンバにおける MAC アドレスの同期化の場合は、VSS は 3 つの異なる方式を使用します。
  - *フラッディングからフレーム*：ハードウェアにおける明示的な送信元学習
  - *MAC 通知*：DFC ラインカードに対する MAC エントリの +MN および –MN 更新
  - *Out-of-band sync*：160 秒ごとに MAC アドレスをグローバルに同期化

これらの方式については、「MEC トポロジを使用した VSS におけるレイヤ 2 MAC 学習」（P.3-21）を参照してください。MAC アドレスの到達可能性は両方のメンバ経由で実現され、MEC トポロジが最適なローカル リンク フォワーディングを可能にするため、ユニキャスト フラッディングを防止します。スイッチ間の一時的なフラッディングが発生し、MAC アドレスの再学習が実行される場合がありますが、これはユーザのアプリケーション パフォーマンスには影響しません。

図 3-20　VSS 対応のディストリビューション ブロック



マルチレイヤ デザイン、ベスト プラクティスのチューニング

キャンパスのディストリビューションで VSS をデプロイすることにより、従来のコンフィギュレーション ベスト プラクティスの一部が変更される場合があります。このデザイン ガイドは、マルチレイヤ キャンパス設計で利用可能あるいは一般的に使用されているコンフィギュレーション オプションすべてを評価しないという点に留意することをお勧めします。キャンパス マルチレイヤ コンフィギュレーションで推奨されるベスト プラクティスへの変更は、高可用性ネットワーク設計および VSS デプロイという点に限り、このガイドで説明しています。

トランキング構成のベスト プラクティス

スタンドアロン スイッチを特徴とする従来のマルチレイヤ設計においては、Dynamic Trunking Protocol（DTP; ダイナミック トランキング プロトコル）と 802.1Q または Inter-Switch Link（ISL; スイッチ間リンク）ネゴシエーションがイネーブルの場合に、ノードまたはインターフェイスを復旧するときのトランク設定のネゴシエーションにかなりの時間がかかる場合があります。ネゴシエーション中は、リンクはレイヤ 2 から見て動作しているため、トラフィックは廃棄されます。最大 2 秒の損失が発生する可能性があります。これはトランク インターフェイスが呼び出された場所によって異なります。ただし、この構成では、DTP はトランクのステートを積極的に監視していないため、不適切に構成されているトランクは簡単に特定されません。高速コンバージェンスと、構成と変更を管理する能力の間でのバランスがあります。

VSS では、*desirable* あるいは *undesirable* ポート チャネル インターフェイスのトランク モードは、スタンドアロン ノードの動作を示しません。VSS では、各アクセス レイヤはポート チャネル（MEC）経由で接続され、リンク メンバがオンラインになるときには、ネゴシエーションが個別に行われず、EtherChannel グループに追加されます。各ノードに個別のトランキング イベントをネゴシエートする個別のコントロール プレーンがあるスタンドアロン デュアル ノード設計と比較すると、ノード関連の復旧時の損失も問題ではありません。VSS では、ノードが復旧するときには、リンクアップ イベントは MEC の追加メンバ リンクであり、トランク インターフェイスではありません。図 3-21 を参照してください。

図 3-21　コンバージェンス損失比較

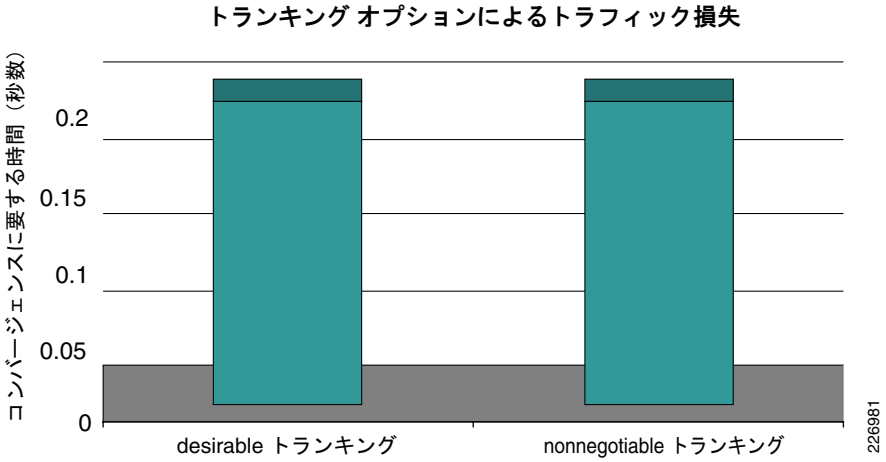


図 3-21 に、*desirable* トランク モードあるいは *nonnegotiable*（*desirable* ではない）トランク モードに関連するコンバージェンス損失を比較します。いずれの構成でも、損失は 200 ミリ秒以下です。*desirable* トランク モードを実行するもう 1 つの利点は、トランキングしたインターフェイスでは運用が効率的になることです。VSS では、複数の VLAN が複数のアクセス レイヤのスイッチにまたがることができます。つまり、トランキングされる VLAN の数が多くなるため、変更管理中に発生するエラーの可能性も高くなり、次に VSS ドメインを停止する可能性があります。構成が一致せず、このオプション設定によってある不一致状況で syslog メッセージが生成される場合は、トランキングが形成されないため、*desirable* オプションはトラフィックのブラック ホールを低減し、運用スタッフに障害の診断結果を示します。

ヒント　**シスコでは、VSS 対応設計において、インターフェイスの両端のトランクについて、desirable-desirable または auto-desirable オプションを使用して構成することをお勧めします。**

トランク上の VLAN 構成

ループのない VSS 対応レイヤ 2 設計においては、かなり直感的な方法で、VLAN の拡大と任意のアクセス レイヤのスイッチからの VLAN へのアクセスが可能になります。制御されていない VLAN の拡張とアクセス ポリシーを制限することが一般的に行われているベスト プラクティスです。トランキングされたポート チャネル上で **switchport trunk allowed vlan** コマンドを使用して、表示される対象のスイッチに転送される VLAN を制限するべきです。これによって、移動、追加、および変更の影響を低減し、大規模な VLAN ドメインの問題を解決するときにより明確になります。

トランク単位で VLAN を制限するもう 1 つの利点は、ラインカードとスイッチごとに、STP 論理ポート機能の使用を最適化できることです。STP 論理ポートのキャパシティは、CPU が数多くの STP BPDU per-VLAN per-physical を処理し、トポロジ変更中に物理ポートごとの STP BPDU を送信する効率によって決まります。論理 STP 対応ポート キャパシティの数は、特定の STP ドメインのラインカードと全体のシステム機能によって決まります。これらの制限は、次の URL にあるリリース ノートに記載されています。

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/release/notes/ol_14271.html#wp26366

VSS の観点からは、論理ポート制限はシャーシ単位ではなく、システム単位で適用されます。これは、両方のシャーシ インターフェイスを管理するコントロール プレーンが 1 つしかないためです。ラインカードごとの STP 論理ポートの最大数は、使用するラインカードの種類によって異なります。VSS でサポートされるラインカードの種類は、最大 1,800 の論理ポートを構成できる WS-X67xx シリーズだけですが、Cisco IOS Release 122.(33)SXI1 ではこの制限はありません。ラインカードごとの STP 論理制限を超えないようにするには、次の 2 つの方法があります。

- トランクごとに許可される VLAN の数を制限する
- VSS の複数のラインカード上でアクセス レイヤの接続を分散する

次の CLI に、VSS システムがラインカードごとの STP 論理制限を超えているかどうかを判断する方法を示します。

```
6500-VSS# sh vlan virtual-port switch 1 slot 8
Slot 8 switch : 1
Port        Virtual-ports
-------------------------
Gi1/8/1             8
Gi1/8/2             8
Gi1/8/3             207
Gi1/8/4             207
Gi1/8/5             207
Gi1/8/6             207
Gi1/8/7             207
Gi1/8/8             207
Gi1/8/9             207
Gi1/8/10            207
Gi1/8/11            207
Gi1/8/12            207
Gi1/8/13            207
Gi1/8/14            207
Gi1/8/15            207
Gi1/8/16            7
Gi1/8/17            9
Gi1/8/19            1
Gi1/8/21            9
Gi1/8/23            9
Gi1/8/24            9
Total virtual ports:2751
```

前述の出力例では、合計すると STP 論理ポート制限を超える数多くのポート上で 200 を超える VLAN だけが許可されています。次の **show** コマンド出力では、この数を計算する方法を示します。

```
6500-VSS# sh int po 221 trunk

Port    Mode         Encapsulation  Status        Native vlan
Po221   desirable    802.1q         trunking      221

Port    Vlans allowed on trunk
Po221   21,121,400,450,500,550,600,900 <-

Port    Vlans allowed and active in management domain
Po221   21,121,400,450,500,550,600,900

Port    Vlans in spanning tree forwarding state and not pruned
Po221   21,121,400,450,500,550,600,900
```

トランクで許可されている VLAN インスタンス数は、使用される論理 STP ポートに等しくなります。前述の出力の場合、この数は 8 になります（21、121、400、450、500、550、600、および 900 は 8 論理ポートになります）。

ここで、次の出力例で示されている制限されていないポートを考えます。

```
6500-VSS# sh int po 222 trunk

Port    Mode         Encapsulation  Status        Native vlan
Po222   desirable    802.1q         trunking      222

Port    Vlans allowed on trunk
Po222   1-4094

Port    Vlans allowed and active in management domain
Po222   1-7,20-79,102-107,120-179,202-207,220-279,400,450,500,550,600,650,900,999 <-

Port    Vlans in spanning tree forwarding state and not pruned
Po222   1-7,20-79,102-107,120-179,202-207,220-279,400,450,500,550,600,650,900,999
```

制限されていないポートの場合、すべての VLAN が許可されるため、STP 論理ポート数が 207 へと大幅に増加します （1-7、20-79、102-107、120-179、202-207、220-279、400、450、500、550、600、650、900、および 999 は 207 論理ポートになります）。

VSS システムの STP 論理ポート数の合計は、**show vlan virtual-port** コマンドで表示できます。本来このコマンドはスタンドアロン システム用であるため、VSS CLI 出力では、システム固有の制限計算が完全に正しく表示されません。VSS の STP ポートの合計論理数は、EtherChannel ポート上で STP が実行されている場合でも、各スイッチ数を加算して計算されます。そのため、ポートの半分だけが STP 論理ポート制限に対してカウントされることになります。次の URL の VSS リリース ノートを参照してください。

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/release/notes/ol_14271.html#wp26366

ヒント　**シスコでは、必要な VLAN についてはトランク経由で転送するように明示的に設定することをお勧めします。**

Unidirectional Link Detection（UDLD; 単方向リンク検出）

通常モードの UDLD は、ポートを検出して error-disable にすることで、ケーブル配線の不一致が原因で生じるループ ブロードキャスト ストームを回避するために使用します。アグレッシブ UDLD は通常の UDLD が強化された形で、従来はリンクの整合性と障害状態のハードウェアを検出するために使用されています。UDLD プロトコルは、STP コンバージェンスが最大 50 秒かかる可能性がある PVST 環境においてかなり早い段階で、STP ループの問題を検出するために使用します。（ハードウェア障害が原因の）ループ状況を検出するためのツールとしてアグレッシブ UDLD を適用することは、VSS 対応のレイヤ 2 ドメインに制限されています。これは、VSS は本来ループ フリー トポロジであるためです。また、新しい STP プロトコル（RPVST+ および MST）プロトコルに関連する一般的なコンバージェンスは、アグレッシブ UDLD 検出時間よりも著しく速くなります。VSS 環境におけるアグレッシブ UDLD 検出の副作用は、その効果よりも大幅に影響が大きいものです。VSS（統合型コントロール プレーン）において、重要プロセスの多くは CPU による処理が必要です。一般的に、DFC ベースのラインカードの初期化には時間がかかります。障害状態のソフトウェアは、アグレッシブ UDLD プロセスが hello を処理できないように、CPU リソースを占有する場合があります。このような状況では、アグレッシブ UDLD が強制的に動作し、両側の接続が error-disable になる false-positive を引き起こす可能性があります。

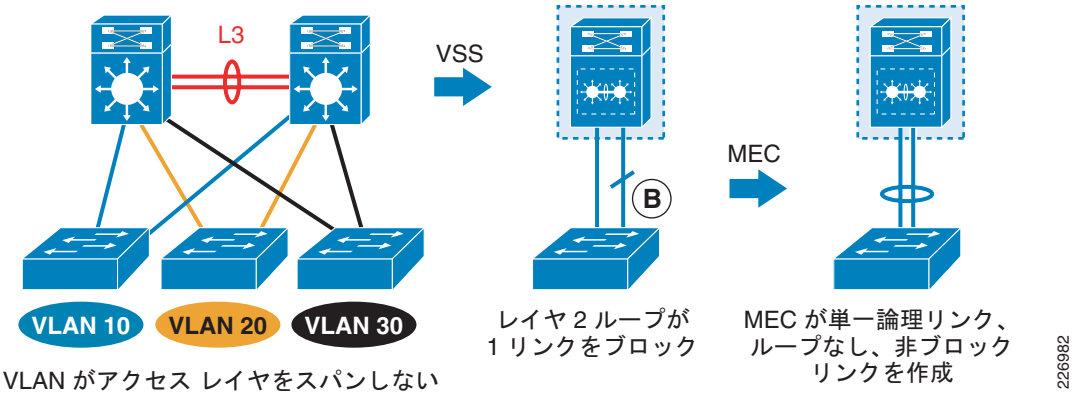
ヒント　**アグレッシブ UDLD は、リンク整合性チェックとして使用するべき *ではありません*。ケーブル配線の障害の検出や、リンク整合性のチェックには、通常モードの UDLD を使用してください。**

VSS でのトポロジの考慮事項

VSS（シングル論理デバイス）の導入は、キャンパス トポロジに多大な影響を与えます。VSS に接続したデバイス（MEC がある場合もない場合も）も重要な役割を担います。特定のトポロジのレイヤ 2 およびレイヤ 3 インタラクションは、トポロジ動作を決定し、ユーザ データ トラフィックのコンバージェンスを決定します。この項ではレイヤ 2 ドメインについて説明します。レイヤ 3 ドメインについては、「VSS を使ったルーティング」（P.3-46）で説明します。

従来は、多くのネットワークで、VLAN がクローゼットをまたがないように、最適化されたマルチレイヤ トポロジ（V 字型または U 字型）が採用されてきました。図 3-22 に示されるように、MEC を使用せずにこのようなトポロジで VSS を配置すると、STP ループがネットワークに再導入されます。同じデバイスからの 2 つのレイヤ 2 リンクが VSS に接続するときには必ず、MEC を使用する必要があります。図 3-22 に、MEC がある場合とない場合の VSS 対応の非ループ型 V 字型トポロジの動作を例示します。

図 3-22　非ループ型トポロジの動作



間接接続を特徴とするデイジー チェーン型のアクセス スイッチ トポロジの場合、次の 2 つの設計上の課題があります。

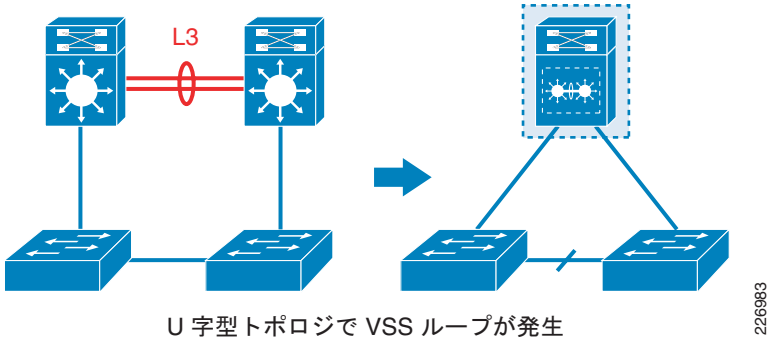
- ユニキャスト フラッディング
- ループ（ブロックされたリンク）

ディストリビューション レイヤの仮想スイッチを使用すると、ユニキャスト フラッディングの問題が生じますが、STP ブロックされたリンクでの設計においては、まだネットワークにはレイヤ 2 ループがあります。トラフィック復旧時間は、リンク障害またはノード障害時のスパニング ツリー復旧によって決まります。

図 3-23 に、VSS の U 字型トポロジを示します。2 つの望ましくない効果があります。

- ループのあるトポロジを作成します。
- STP トポロジが形成され、VSS に接続したアップリンクがブロックされる場合には、50% のダウンストリーム トラフィックが VSL リンクを横断します。

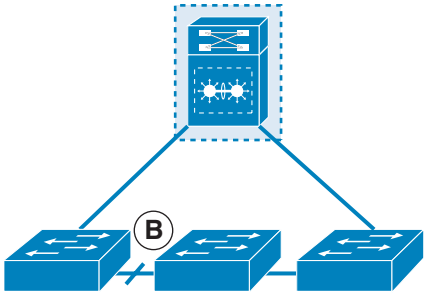
図 3-23　U 字型 VSS トポロジ ループの紹介



U 字型トポロジで VSS ループが発生

VSS が間接接続を検出できないデイジー チェーン型のアクセス トポロジでは、レイヤ 2 ループと STP ブロックされたリンクが導入されています。図 3-24 を参照してください。

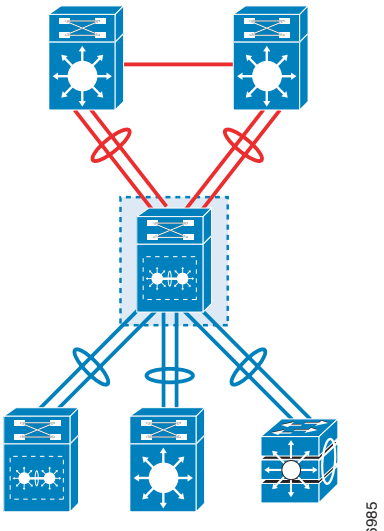
図 3-24　STP ブロックされたリンク



レイヤ 2 ループが 1 スイッチ分少ないものの依然として存在

デイジー チェーン型のトポロジの場合でも、U 字型トポロジの場合でも、2 つの解決策があります。各スイッチから MEC を配置して、アクセス レイヤへの間接接続を回避するか、交差に積み重ねて配置された EtherChannel 対応スイッチ（Catalyst 37xx スタック）をアクセス レイヤで使用します。図 3-25 を参照してください。

図 3-25　スター型トポロジ



スター型トポロジによる STP ループの回避

ヒント　**シスコでは、ループを回避し、リンクまたはノードの障害時に最良のコンバージェンスを実現するために、VSS に接続する各デバイスは MEC（レイヤ 2 およびレイヤ 3）機能を備えたスター型トポロジを常に使用することをお勧めします。**

VSS でのスパニング ツリー構成のベスト プラクティス

注意　VSS ベースの設計の利点の 1 つは、レイヤ 2 ドメイン全体で STP をアクティブにできることです。VSS は単にループ フリー トポロジを STP に提供します。ネットワーク設計者が作成したトポロジがスター型（MEC 対応）でない場合は、ループ フリー トポロジを提供する固有の方法はありません。（同じケーブルを連続して VSS メンバ シャーシに接続して）アクセス レイヤまたは VSS システム内で発生した偶発的ループを検出するために、VSS 対応設計ではスパニング ツリーをイネーブルにする必要があります。非 VSS 対応ネットワークでは、スパニング ツリー ツールのセットを使用して、ストームのループやストームの効果の減少からネットワークを保護し、修正措置を取ることができます。非 VSS マルチレイヤ設計におけるループ ストーム状態保護の詳細については、http://www.cisco.com/en/US/products/hw/switches/ps700/products_white_paper09186a00801b49a4.shtml#cg5 を参照してください。

VSS 対応ネットワークではループは発生しないため、このデザイン ガイドでは、ブロードキャスト ストームの効果を弱める STP ツールの詳細については説明しません。ただし、次の STP 関連要因については、VSS 対応キャンパスでも考慮する必要があります。

- 「STP 選択」（P.3-33）
- 「ルート スイッチと Root Guard 保護」（P.3-33）
- 「Loop Guard」（P.3-33）
- 「トランク上の PortFast」（P.3-34）
- 「PortFast および BPDU Guard」（P.3-37）
- 「BPDU フィルタ」（P.3-38）

これらの点については、次の項で簡単に説明します。

STP 選択

特定の STP プロトコル実装の選択 : Rapid per VLAN Spanning Tree Plus（RPVST+）または Multiple Instance Spanning Tree（MST; 複数インスタンス スパニング ツリー）は完全に顧客設計要件に基づいています。平均的なエンタープライズ キャンパス ネットワークでは、シスコ以外のスイッチとの相互運用性が求められる場合を除き、RPVST+ が最も一般的で適切です。非常に多数の VLAN と論理ポートをサポートする機能の利点と対比して、MST の追加の依存と要件を評価する必要があります。ただし、大多数のキャンパス ネットワークでは、RPVST+ の 12000 論理ポートキャパシティで十分です。VSS については、次のリリース ノートの URL を参照してください。

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/release/notes/ol_14271.html#wp26366

ルート スイッチと Root Guard 保護

STP のルートは必ず VSS と同じでなければなりません。スパニング ツリー ルートについては、統計的に定義され、ハード コーディングされた値を使用して、ネットワーク上の他のスイッチが特定のスパニング ツリー ドメインのルートを使用できないようにします。VSS に対するアクセス レイヤのスイッチのリンク上の Root Guard を使用するか、アクセス レイヤのスイッチのユーザ ポートで Root Guard をイネーブルにします。ただし、後者の場合は、他のユーザがアクセス レイヤのスイッチを、ルートとして引き継ぎ可能な別のスイッチと置き換えようとするときにそれを防止できません。ルート変更は非ループ型設計においてはフォワーディングに影響しない場合があります（ルート選択は代替パス（ループ）が STP に示される時にだけ問題となります）。ただし、非準拠スイッチがルートになることで生じる BPDU の損失または変化しやすさは、ネットワークにおける不安定性につながる可能性があります。

デフォルトでは、アクティブなスイッチのベース MAC アドレスは、VSS のルート アドレスとして使用されます。このルート アドレスは SSO スイッチオーバー中には変更されないため、アクセス レイヤのスイッチはルート変更を認識しません。詳細については、「VSS を使用した STP 操作」（P.3-38）を参照してください。

Loop Guard

一般的な顧客ネットワークでは、CPU 利用、障害状態のハードウェア、構成エラー、あるいはケーブル配線の問題により、BPDU が存在しなくなります。このような状況によって、代替ポートがフォワーディング モードになり、ループ ストームをトリガーします。BPDU Loop Guard は 6 秒以内にこのような状況を防止します（不足している 3 つの連続 BPDU）。通常、Loop Guard は代替ポートで稼動し、STP 対応ポートでだけ動作します。MEC のある VSS 対応設計では、ループ型トポロジを STP プロトコルに提供しません。結果として、すべてのポートが転送し、ブロックしていないため、VSS 対応ネットワークにおいて、Loop Guard が特に有用な機能ではない場合があります。

Loop Guard がトランク インターフェイスの両側でイネーブルになり、BPDU の損失が発生する場合は、アクセス スイッチの EtherChannel ポート（STP が実行されている場所）のステートは、ルート不整合に遷移します。EtherChannel 全体をディセーブルにすることは、ループが存在しない設計におけるソフト エラーを検出するための望ましい方法ではありません。

Loop Guard がイネーブルではなく、BPDU 損失が検出される場合は、アクセス レイヤのスイッチがローカル データベースで定義された VLAN のルートになります。ユーザ トラフィックは継続する場合がありますが、数秒後に UDLD または PAgP タイマーによって問題が検出され、ポートが error-disable になります。

通常モードの UDLD または通常の PAgP/LACP の hello 方式を使用して、ソフト エラーを検出してください。UDLD および PAgP/LACP は EtherChannel の個々のリンク メンバだけをディセーブルにし、アクセス スイッチの接続はアクティブに保ちます。さらに、Cisco IOS Release 12.2(33) SXI での開発により、この種類の問題を解決するためのより効果的な方法が導入されました。詳細については、次のリンクを参照してください。

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/configuration/guide/spantree.html#wp1098785

ヒント　シスコでは、VSS 対応キャンパス ネットワークでは、Loop Guard をイネーブルにしないことをお勧めします。

トランク上の PortFast

トランクに対する PortFast の実装は、待ち受けや STP の学習フェーズに入らずに、すぐに VLAN をフォワーディング ステートにします。ただし、トランクのポート port-fast ステートは、接続のリモート側からの BPDU を確認するとすぐに、通常の STP ポートになります。そのため、主要な利点は、初期化中に STP のフォワーディング ステートを高速化することです。

従来のマルチレイヤ ループ型ネットワークでは、トランクでのポート高速化機能の使用は、代替パスをブロックまたは検出するのに時間がかかる場合があるため、高メッシュ型トポロジにおける一時ループにつながる可能性があります。このリスクのため、アプリケーションが制限されてきました。VSS 対応の設計においては、トランク上でのポート高速化機能の使用は安全です。これは、VSS トポロジには本来ループがなく、トランク上のポート高速化機能によって生じる一時ループの可能性がなくなるためです。

アクセス レイヤのスイッチのデュアル ノード設計（非 VSS）では、各インターフェイスがディストリビューション レイヤで別個のノードに接続します。トランクの障害や初期化がそれぞれ独立して発生し、STP のステートまたはトランク ネゴシエーションのため、インターフェイスはトラフィック（パケット損失）を転送する準備ができていません。VSS ではアクセス レイヤが STP が動作するポート チャネルで接続されるため、この遅延がなくなります。別のインターフェイスを効果的に追加すると、別の EtherChannel メンバが追加されます。STP およびトランク ステートをネゴシエーションする必要はありません。

次の syslog 出力の例では、ポート高速化がトランクでイネーブルになっているときに、初期遅延が最大 1 秒低減されていることがわかります。

（注） ユーザ データ トラフィックにおけるこの遅延の影響を定量化することは簡単ではありません。ツールではインターフェイスが初期化されるときに正確にタイムスタンプを測定できず、再起動（**no shutdown** コマンド）とフル フォワーディングの間に失われるデータ量もわかりません。さらに、ツールはデータを送信してから、インターフェイスを再起動しなければなりません。インターフェイス初期化イベントの前に送信されたデータ間の差異を判断する方法はありません。概算実験によると、トランクで PortFast がイネーブルになっていない場合は、最大 600 ミリ秒の接続停止が発生することが示されています。

トランクでディセーブルになっている PortFast

次に、トランクでディセーブルになっている PortFast を表示する CLI の出力例を示します。

VSS Syslog

```
6500-VSS# sh log | inc 106
Oct 22 14:03:31.647: SW2_SP: Created spanning tree: VLAN0106 (5554BEFC)
```

```
Oct 22 14:03:31.647: SW2_SP: Setting spanning tree MAC address: VLAN0106 (5554BEFC) to
0008.e3ff.fc28
Oct 22 14:03:31.647: SW2_SP: setting bridge id (which=3) prio 24682 prio cfg 24576 sysid
106 (on) id 606A.0008.e3ff.fc28
Oct 22 14:03:31.647: SW2_SP: STP PVST: Assigned bridge address of 0008.e3ff.fc28 for
VLAN0106 [6A] @ 5554BEFC.
Oct 22 14:03:31.647: SW2_SP: Starting spanning tree: VLAN0106 (5554BEFC)
Oct 22 14:03:31.647: SW2_SP: Created spanning tree port Po206 (464F4174) for tree VLAN0106
(5554BEFC)
Oct 22 14:03:31.647: SW2_SP: RSTP(106): initializing port Po206
Oct 22 14:03:31.647: %SPANTREE-SW2_SP-6-PORT_STATE: Port Po206 instance 106 moving from
disabled to blocking <- 1
Oct 22 14:03:31.647: SW2_SP: RSTP(106): Po206 is now designated
Oct 22 14:03:31.667: SW2_SP: RSTP(106): transmitting a proposal on Po206
Oct 22 14:03:32.647: SW2_SP: RSTP(106): transmitting a proposal on Po206
Oct 22 14:03:32.655: SW2_SP: RSTP(106): received an agreement on Po206
Oct 22 14:03:32.919: %LINK-3-UPDOWN: Interface Vlan106, changed state to up
Oct 22 14:03:32.935: %LINEPROTO-5-UPDOWN: Line protocol on Interface Vlan106, changed
state to up
Oct 22 14:03:32.655: %SPANTREE-SW2_SP-6-PORT_STATE: Port Po206 instance 106 moving from
blocking to forwarding  <- 2
Oct 22 14:03:34.559: %PIM-5-DRCHG: DR change from neighbor 0.0.0.0 to 10.120.106.1 on
interface Vlan106
```

アクセス レイヤ スイッチ

```
Access-Switch# show logging
Oct 22 14:03:29.671: %DTP-SP-5-TRUNKPORTON: Port Gi1/1-Gi1/2 has become dot1q trunk
Oct 22 14:03:31.643: %LINK-3-UPDOWN: Interface Port-channel1, changed state to up
Oct 22 14:03:31.647: %LINEPROTO-5-UPDOWN: Line protocol on Interface Port-channel1,
changed state to up
Oct 22 14:03:31.651: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet1/1,
changed state to up
Oct 22 14:03:31.636: %EC-SP-5-BUNDLE: Interface GigabitEthernet1/1 joined port-channel
Port-channel1
Oct 22 14:03:31.644: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 6 moving from disabled
to blocking
Oct 22 14:03:31.644: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 106 moving from disabled
to blocking <-1
Oct 22 14:03:31.644: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 900 moving from disabled
to blocking
Oct 22 14:03:31.660: %LINK-SP-3-UPDOWN: Interface Port-channel1, changed state to up
Oct 22 14:03:31.660: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface
GigabitEthernet1/1, changed state to up
Oct 22 14:03:31.664: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface Port-channel1,
changed state to up
Oct 22 14:03:31.867: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet1/2,
changed state to up
Oct 22 14:03:31.748: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 900 moving from blocking
to forwarding
Oct 22 14:03:31.856: %EC-SP-5-BUNDLE: Interface GigabitEthernet1/2 joined port-channel
Port-channel1
Oct 22 14:03:31.868: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface
GigabitEthernet1/2, changed state to up
Oct 22 14:03:32.644: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 6 moving from blocking
to forwarding
Oct 22 14:03:32.644: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 106 moving from blocking
to forwarding <- 2
```

特定の VLAN のポート チャネル インターフェイスを初期化する時間は、約 1 秒です（前述の syslog 出力例のマーカーを参照）。

トランク ポート チャネルでイネーブルになっている PortFast

次に、トランクのポート チャネルでイネーブルになっている PortFast を表示する CLI 出力例を示します。

VSS Syslog

```
6500-VSS# sh log | inc 106
Oct 22 14:14:11.397: SW2_SP: Created spanning tree: VLAN0106 (442F4558)
Oct 22 14:14:11.397: SW2_SP: Setting spanning tree MAC address: VLAN0106 (442F4558) to
0008.e3ff.fc28
Oct 22 14:14:11.397: SW2_SP: setting bridge id (which=3) prio 24682 prio cfg 24576 sysid
106 (on) id 606A.0008.e3ff.fc28
Oct 22 14:14:11.397: SW2_SP: STP PVST: Assigned bridge address of 0008.e3ff.fc28 for
VLAN0106 [6A] @ 442F4558.
Oct 22 14:14:11.397: SW2_SP: Starting spanning tree: VLAN0106 (442F4558)
Oct 22 14:14:11.397: SW2_SP: Created spanning tree port Po206 (464F2BCC) for tree VLAN0106
(442F4558)
Oct 22 14:14:11.397: SW2_SP: RSTP(106): initializing port Po206
Oct 22 14:14:11.401: %SPANTREE-SW2_SP-6-PORT_STATE: Port Po206 instance 106 moving from
disabled to blocking   <- 1
Oct 22 14:14:11.401: SW2_SP: RSTP(106): Po206 is now designated
Oct 22 14:14:11.401: %SPANTREE-SW2_SP-6-PORT_STATE: Port Po206 instance 106 moving from
blocking to forwarding  <- 2
Oct 22 14:14:11.769: %LINK-3-UPDOWN: Interface Vlan106, changed state to up
Oct 22 14:14:11.777: %LINEPROTO-5-UPDOWN: Line protocol on Interface Vlan106, changed
state to up
Oct 22 14:14:13.657: %PIM-5-DRCHG: DR change from neighbor 0.0.0.0 to 10.120.106.1 on
interface Vlan106
```

アクセス レイヤ スイッチ

```
Access-switch# show logging
Oct 22 14:14:04.789: %LINK-SP-3-UPDOWN: Interface Port-channel1, changed state to down
Oct 22 14:14:05.197: %LINK-SP-3-UPDOWN: Interface GigabitEthernet1/1, changed state to
down
Oct 22 14:14:05.605: %LINK-SP-3-UPDOWN: Interface GigabitEthernet1/2, changed state to
down
Oct 22 14:14:05.769: %LINK-SP-3-UPDOWN: Interface GigabitEthernet1/1, changed state to up
Oct 22 14:14:06.237: %LINK-3-UPDOWN: Interface GigabitEthernet1/2, changed state to up
Oct 22 14:14:06.237: %LINK-SP-3-UPDOWN: Interface GigabitEthernet1/2, changed state to up
Oct 22 14:14:09.257: %DTP-SP-5-TRUNKPORTON: Port Gi1/1-Gi1/2 has become dot1q trunk
Oct 22 14:14:11.397: %LINK-3-UPDOWN: Interface Port-channel1, changed state to up
Oct 22 14:14:11.401: %LINEPROTO-5-UPDOWN: Line protocol on Interface Port-channel1,
changed state to up
Oct 22 14:14:11.401: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet1/1,
changed state to up
Oct 22 14:14:11.385: %EC-SP-5-BUNDLE: Interface GigabitEthernet1/1 joined port-channel
Port-channel1
Oct 22 14:14:11.397: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 6 moving from disabled
to blocking
Oct 22 14:14:11.397: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 6 moving from blocking
to forwarding
Oct 22 14:14:11.397: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 106 moving from disabled
to blocking    <- 1
Oct 22 14:14:11.397: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 106 moving from blocking
to forwarding  <- 2
Oct 22 14:14:11.397: %SPANTREE-SP-6-PORT_STATE: Port Po1 instance 900 moving from blocking
to forwarding
Oct 22 14:14:11.413: %LINK-SP-3-UPDOWN: Interface Port-channel1, changed state to up
Oct 22 14:14:11.913: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet1/2,
changed state to up
Oct 22 14:14:11.413: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface
GigabitEthernet1/1, changed state to up
```

```
Oct 22 14:14:11.413: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface Port-channel1,
changed state to up
Oct 22 14:14:11.901: %EC-SP-5-BUNDLE: Interface GigabitEthernet1/2 joined port-channel
Port-channel1
Oct 22 14:14:11.913: %LINEPROTO-SP-5-UPDOWN: Line protocol on Interface
GigabitEthernet1/2, changed state to up
```

前述の syslog 出力例のマーカーで示されているように、ブロックおよびフォワーディング間の時間は事実上ゼロになります。

Portfast 機能によってトランクの初期化を最適化するオプションは、追加の構成要件に対して重み付けする必要があります。初期システム起動中は、MAC の学習、ARP 検出、およびコントロール プレーン アクティビティ経由でのネットワーク デバイス検出の完了（ネイバー隣接、NSF、ルーティングなど）を行ってはじめて、システムでパケットを転送できるようになります。これらの追加処理によって、オンラインで高速化できるトランクの利点がなくなる可能性があります。しかし、スイッチまたはルータが完全に動作したときに、ポートが強制的に再起動するか稼動状態になるときには有用です。

PortFast および BPDU Guard

VSS におけるエッジ ポートの動作の保護と改善方法は、他のキャンパス設計と同じです。ホスト ポート マクロを使用してエッジ ポートを構成して、STP フォワーディング ステートにします。Topology Change Notification（TCN; トポロジ変更通知）メッセージを削減し、他のソフトウェア コンフィギュレーションが EtherChannel とトランクをチェックするときに生じる遅延をなくします。VSS はループフリー トポロジですが、エンド ユーザのアクションまたはアクセス レイヤのスイッチのケーブル誤配線によって、ネットワークでループが発生する可能性があります。ループしたネットワークは最終的に予期しないコンバージェンスにつながり、ループ ベースのブロードキャスト ストームの可能性が大幅に高まる場合があります。

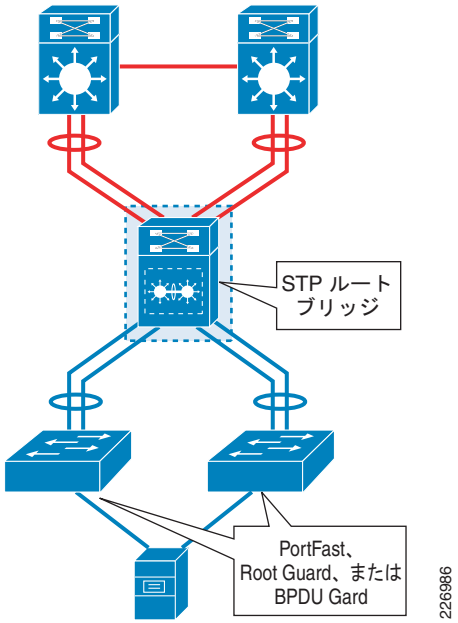
ヒント　VSS 対応ネットワークでは、エッジ ポートを STP に関与させないようにすることが非常に重要です。シスコでは、エッジ ポートで PortFast と BPDU Guard をイネーブルにすることを強くお勧めします。

グローバルでイネーブルにすると、BPDU Guard が稼動 PortFast ステートのすべてのインターフェイスに適用されます。次に、BPDU Guard をイネーブルにしたコンフィギュレーション例を示します。

```
VSS(config-if)# spanning-tree PortFast
VSS(config-if)# spanning-tree bpduguard enable
%SPANTREE-2-BLOCK_BPDUGUARD: Received BPDU on port FastEthernet3/1 with BPDU Guard
enabled. Disabling port.
%PM-4-ERR_DISABLE: bpduguard error detected on Fa3/1, putting Fa3/1 in err-disable state
```

図 3-26 に、さまざまな STP 機能のコンフィギュレーション ゾーンの例を示します。

図 3-26　PortFast、BPDU Guard、およびポート セキュリティ



BPDU フィルタ

BPDU フィルタ機能を不適切な方法で使用すると、ネットワークでループが発生する可能性があります。従来のマルチレイヤ設計と同様に、VSS 対応ネットワークでは、BPDU フィルタリングを使用しないでください。代わりに、BPDU Guard を使用してください。

VSS を使用した STP 操作

VSS は、どの仮想スイッチ メンバがアクティブ ステートになっているかには関係なく、1 つの STP ブリッジ ID と優先度をアドバタイズする 1 つの論理スイッチから構成されています。ブリッジ ID は次の出力で示されるアクティブなシャーシに基づいています。

```
6500-VSS# sh catalyst6000 chassis-mac-addresses
  chassis MAC addresses: 1024 addresses from 0008.e3ff.fc28 to 0008.e400.0027
6500-VSS# sh spanning-tree vla 450

VLAN0450
  Spanning tree enabled protocol rstp
  Root ID     Priority    25026
              Address     0008.e3ff.fc28
              This bridge is the root
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID   Priority    25026  (priority 24576 sys-id-ext 450)
              Address     0008.e3ff.fc28
              Hello Time   2 sec  Max Age 20 sec  Forward Delay 15 sec
              Aging Time 480

Interface           Role Sts Cost      Prio.Nbr Type
------------------- ---- --- --------- -------- --------------------------------
Po202               Desg FWD 3         128.1699 P2p
```

VSS では、スパニング ツリーは SSO を認識しています。SSO は、SSO スイッチオーバー（アクティブなスイッチの障害）中に、STP プロトコルのレジリエンシーを実現します。新しいアクティブなスイッチ メンバは最初にアドバタイズされた STP ブリッジの優先度と各アクセス レイヤのスイッチの ID を保持します。つまり、STP ではコンバージェンスを Sub-Sec のレベルにまで加速させるネットワークの学習処理を再初期化して実行する必要はありません。

同様に、STP が EtherChannel ポートで稼動していないため、MEC のメンバ リンク障害も TCN を生成しません。次の出力を参照してください。

```
6500-VSS# show spanning-tree vl 10 detail | inc Times|Port-channel
  Root port is 1665 (Port-channel10), cost of root path is 3
          from Port-channel10
  Times:  hold 1, topology change 35, notification 2
 Port 1665 (Port-channel10) of VLAN0010 is root forwarding
6500-VSS#show interface port-channel10 | inc Gi
  Members in this channel: Gi1/1 Gi1/2
6500-VSS# conf t
VSS(config)# int gi1/1
VSS(config-if)# shut

6500-VSS# show spanning-tree vlan 10 detail | inc Times|Port-channel
  Root port is 1665 (Port-channel10), cost of root path is 4
          from Port-channel10
  Times:  hold 1, topology change 35, notification 2
 Port 1665 (Port-channel10) of VLAN0010 is root forwarding
6500-VSS#show interface port-channel10 | inc Gi
  Members in this channel: Gi1/2
```

アクティブなスイッチが BPDU を生成します。各 BPDU フレームの送信元 MAC アドレスは、STP ポート（MEC）が終端するラインカードに基づいています。通常、MEC ポートから継承した MAC アドレスは、BPDU フレームの送信元 MAC アドレスとして使用されます。この送信元 MAC アドレスは、ノード障害、ラインまたはカード障害、ポート障害のために、動的に変更できます。BPDU は新しい送信元 MAC とともに送信されるため、アクセス スイッチでは、新しいルートなどのイベントを認識する場合があります。ただし、この障害はネットワークにおける STP トポロジの再計算は行いません。これは、ネットワークがループ フリーのネットワークであり、STP ブリッジ ID と優先度が同じであるためです。次の **debug** コマンドでは、Cisco Catalyst 3560 スイッチでこの動作を監視する方法を示します。BPDU の送信元 MAC アドレスは変更されましたが、ルート ブリッジのブリッジ ID は同じままです（VSS はシングル論理ルートであるため）。

```
3560-switch# debub spanning-tree switch  rx decode
3560-switch# debug spanning-tree switch  rx process

Apr 21 17:44:05.493: STP SW: PROC RX: 0100.0ccc.cccd<-0016.9db4.3d0e type/len 0032 <-
Source MAC
Apr 21 17:44:05.493:     encap SNAP linktype sstp vlan 164 len 64 on v164 Po1
Apr 21 17:44:05.493:     AA AA 03 00000C 010B SSTP
Apr 21 17:44:05.493:     CFG P:0000 V:02 T:02 F:3C R:60A4 0008.e3ff.fc28 00000000
Apr 21 17:44:05.493:     B:60A4 0008.e3ff.fc28 86.C6 A:0000 M:1400 H:0200 F:0F00 <- Root
Bridge ID
Apr 21 17:44:05.493:      T:0000 L:0002 D:00A4
Apr 21 17:44:05.544: %DTP-5-NONTRUNKPORTON: Port Gi0/1 has become non-trunk
Apr 21 17:44:06.030: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet0/1,
changed state to down
Apr 21 17:44:06.072: STP SW: PROC RX: 0100.0ccc.cccd<-0016.9db4.d21a type/len 0032 <- New
Source MAC
Apr 21 17:44:06.072:     encap SNAP linktype sstp vlan 20 len 64 on v20 Po1
Apr 21 17:44:06.072:     AA AA 03 00000C 010B SSTP
Apr 21 17:44:06.072:     CFG P:0000 V:02 T:02 F:3C R:6014 0008.e3ff.fc28 00000000
Apr 21 17:44:06.072:     B:6014 0008.e3ff.fc28 86.C6 A:0000 M:1400 H:0200 F:0F00 <- Same
Bridge ID
Apr 21 17:44:06.072:      T:0000 L:0002 D:0014
Apr 21 17:44:06.072: STP SW: PROC RX: 0100.0ccc.cccd<-0016.9db4.d21a type/len 0032
```

```
Apr 21 17:44:06.072:     encap SNAP linktype sstp vlan 120 len 64 on v120 Po1
Apr 21 17:42:05.939:      T:0000 L:0002 D:0016
```

次の syslog はリンク変更の兆候のように見えますが、リンク メンバの無効化に関するルート変更はありません。

```
Apr 21 17:39:43.534: %SPANTREE-5-ROOTCHANGE: Root Changed for vlan 1: New Root Port is
Port-channel1. New Root Mac Address is 0008.e3ff.fc28
```

大規模レイヤ 2 VSS 対応キャンパス ネットワーク デザインに関する考慮事項

ループ フリー設計である VSS 対応の場合、ネットワーク設計者がネットワーク設計時に受ける制限は少なくなります。VSS 実装時には、ネットワークは複数のスイッチ上の VLAN をまたぐことができるため、各スイッチへの複数の VLAN の配置をサポートできます。設計などの主な理由は、運用上の柔軟性とリソース使用の効率化（サブネット、VLAN など）です。次に、明らかな疑問点を示します。

Q. 適切な STP ドメインのサイズはどのくらいか。

Q. VSS 単位のペアで許可される VLAN 数はいくつか。

Q. VSS 単位のペアでサポートされるデバイス数はいくつか。

STP ドメインのサイジングと上記の疑問点への答えは、STP ドメインにおける非 VSS デバイスを含むさまざまな考慮事項によって異なります。STP ドメインは VSS だけではなく、STP トポロジに参加している他のデバイスからも構成されています。ただし、スパニング コンバージェンスに影響する主要な要因は、STP ドメインの範囲を決定するときに考慮する必要があります。

- 収束までの時間：実装されているプロトコルによって異なります（802.1d、802.1s、あるいは 802.1w）。
- 初期化および障害中にアドバタイズし学習する MAC アドレスの数。
- トポロジ：ループ フリー トポロジを検出する代替パスの数。トポロジの階層が深くなるほど、ループ フリー パスを検出するのに時間がかかります。
- MAC アドレスの学習：ハードウェア（高速）またはソフトウェア（低速）ベース機能にできます。
- スパニング ツリー ドメインの MAC アドレス キャパシティ：MAC アドレスの数が多くなると、コンバージェンスに時間がかかります。
- VLAN と STP インスタンスの数は、CPU で処理する必要がある BPDU の数を制御します。低キャパシティ CPU は BPDU を廃棄する可能性があるため、STP の収束には時間がかかります。
- 各リンク全体でトランキングされる VLAN の数：スイッチ CPU が BDPU を送信しなければならない STP 論理ポートの数。
- 各スイッチにおける VLAN の論理ポートの数：全体のシステム機能。

VSS は、本質的に設計によって、次をイネーブルにすることで、STP コンバージェンスに影響する前述の要因の大部分を削除します。

- ループ フリー トポロジ：スパニング ツリー プロセスで確認できるトポロジ変更はありません。したがって、応答や依存はありません。
- シングル論理スイッチ（ルート）ではないため、VSS メンバ シャーシへの障害時のルート再選択処理はありません。

- VSS はハードウェア ベースの MAC 学習をサポートします。
- STP データベース、ARP、およびポート チャネル インターフェイス ステートなどの STP 動作を制御する主要コンポーネントは、SSO を認識しています。STP に関する変更や学習は、アクティブなシャーシとホットスタンバイのシャーシの間で同期化されるようになったため、MAC ベースのトポロジ依存がなくなります。

ただし、STP ドメインは VSS だけから構成されているのではなく、アクセス レイヤのスイッチや VSS に接続するその他のレイヤ 2 デバイスも含まれています。これらの要素は、VSS だけでは対応できないため、設計における追加の制約事項となり、大規模レイヤ 2 ネットワークの設計時に考慮する必要があります。次に、関連する制約事項の例を示します。

- ローエンド スイッチ プラットフォームでサポートされている VLAN の最大数は、制約要因になる可能性があります。たとえば、Cisco 2960 とシスコ以外のスイッチの VLAN サポートは制限されています。
- STP ドメインにおける残りのスイッチの MAC アドレス学習と Ternary Content Addressable Memory（TCAM）キャパシティは、制約になる場合があります。一般的に、アクセス スイッチの TCAM 機能は制限されています。TCAM キャパシティを超えると、MAC アドレスがソフトウェアで利用できるようになるため、オーバーフロー MAC アドレス宛てのパケットがハードウェアではなく、ソフトウェアでスイッチされます。これによって、スイッチと STP ドメインがさらに不安定になる可能性があります。
- VSS 対応ネットワークで提供されている MAC アドレス変更速度は、コントロール プレーンのアクティビティによって増加する場合があります。1 秒間に移動、追加、または削除される MAC アドレスの数が非常に多く、アクティブ スイッチとホットスタンバイ スイッチを同期するコントロール プレーンの機能が影響を受ける可能性があります。
- ウイルスやその他の感染制御ポリシーに対する露出ドメインは制約になることがあります。適切なレイヤ 2 セキュリティ対策が採られていない大規模なレイヤ 2 ネットワークを使用している場合、緩和策が適用される前に、ホスト感染件数が増大する可能性があります。
- 既存のサブネット構造、および VLAN サイジングにより、複数のアクセス レイヤ スイッチにスパンする大規模なレイヤ 2 VLAN の利用を制限できます。一般的なキャンパスでは、アクセス レイヤ スイッチは、よりローエンドのスイッチング プラットフォームで使用できる物理ポートに自然にマップされる 256 個のホストのサブネット範囲（サブネット マスクは 24 ビット）を持ちます。この構造を変更するのは必ずしも簡単ではありません。VLAN がアクセス レイヤ スイッチ内に常に含まれるようにすると、より明確なトラブルシューティングとモニタリングの境界線が提供されます。

レイヤ 2 ネットワーク スパンを外れた範囲は、設計ポリシーをどのように講じるかに大きく左右されるため、すべてのエンタープライズ キャンパス ネットワークでの使用に適したサイズというものはありません。一般的には、ネットワーク管理、ゲスト アクセス、ワイヤレス、隔離、パスチャー アセスメントなどの機能的な利用に対して割り当てられた VLAN の VSS が持つスパニング機能の使用が推奨されます。音声およびデータ トラフィックは依然として 1 つのアクセス レイヤに閉じ込められたままですが、前述の設計要素を網羅した慎重な計画をもって拡張できます。コンバージェンスとスケーラビリティの観点から、このデザイン ガイドで行われる検証では、表 3-2 にまとめられている属性を持つキャンパス ネットワーク環境を使用します。

表 3-2　キャンパス ネットワークのキャパシティの概要

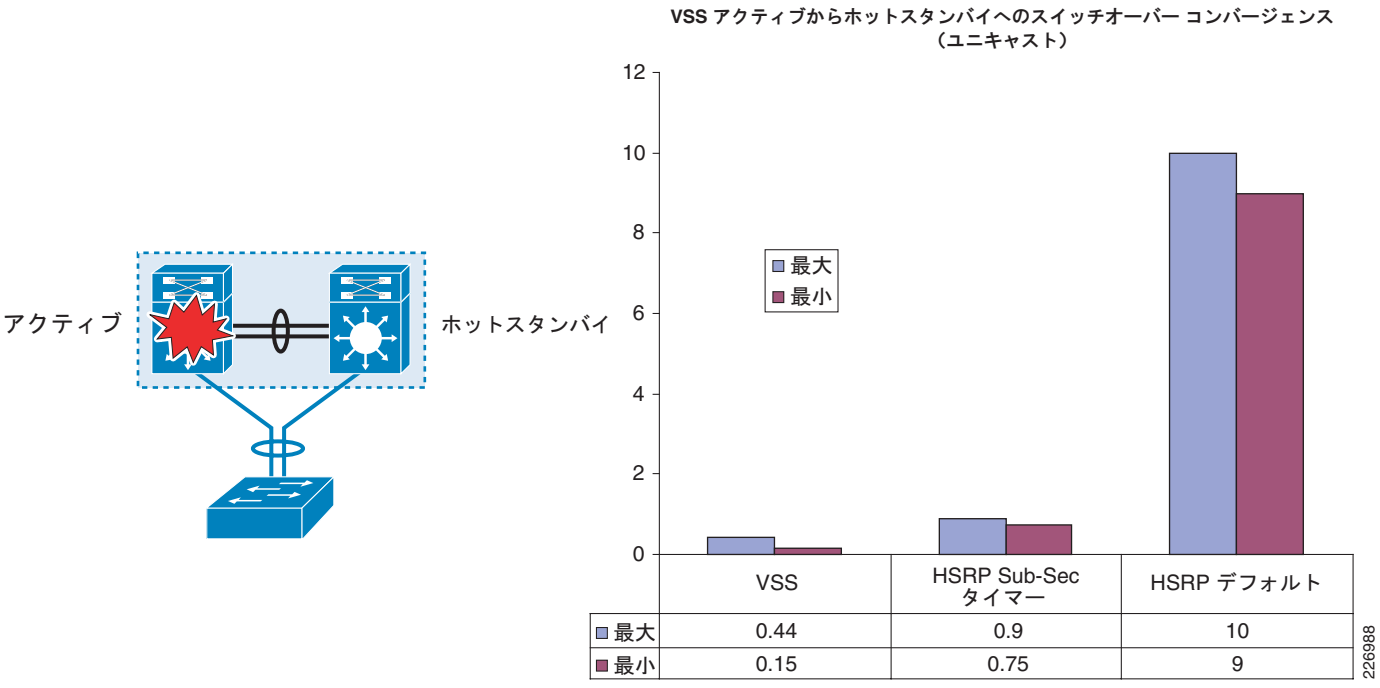
キャンパス環境	平均キャパシティおよび範囲	検証されたキャンパス環境	説明
ディストリビューション ブロック 1 つあたりの平均ネットワーク デバイス数	4K ～ 6K	～ 4500	ホスト対 MAC ごとに一意
ディストリビューション ブロック 1 つあたりの平均アクセス レイヤ スイッチ数	30 ～ 50	70	VSS 1 つあたり 70MEC

表 3-2　キャンパス ネットワークのキャパシティの概要 （続き）

キャンパス環境	平均キャパシティおよび範囲	検証されたキャンパス環境	説明
複数のスイッチにスパンされた VLAN	可変	8 つの VLAN	前述の設計要素により制約される
スパンされた VLAN の MAC アドレス	可変	720 の MAC/VLAN	
アクセス レイヤに含まれる VLAN	可変	140	アクセス レイヤ スイッチ 1 つあたりの制限された音声およびデータ VLAN

スパンされた VLAN の有無に関係なく、アクティブからスタンバイへの障害に関連するコンバージェンスに変化はありません。このデザイン ガイドの検証では、前述の ST に関連するコンバージェンス要因を除去する VSS について段落で説明した機能を明らかにします。図 3-27 は、VSS 導入時のコンバージェンスを表しています。スタンドアロン対応デフォルト コンバージェンスは、デフォルトの FHRP の場合は 10 秒で、最高の条件のコンバージェンスに調整した場合は約 900 ミリ秒です。対照的に、VSS の平均コンバージェンス時間は 200 ミリ秒ですが、複雑な調整や専用コンフィギュレーションは必要ありません。

図 3-27　スイッチオーバー コンバージェンスの比較



アクティブ スーパーバイザ障害の間に発生するトラフィック フローおよびイベントの詳細は「アクティブ スイッチ フェールオーバー」（P.4-5）を参照してください。

マルチキャスト トラフィックおよびトポロジ デザインにおける考慮事項

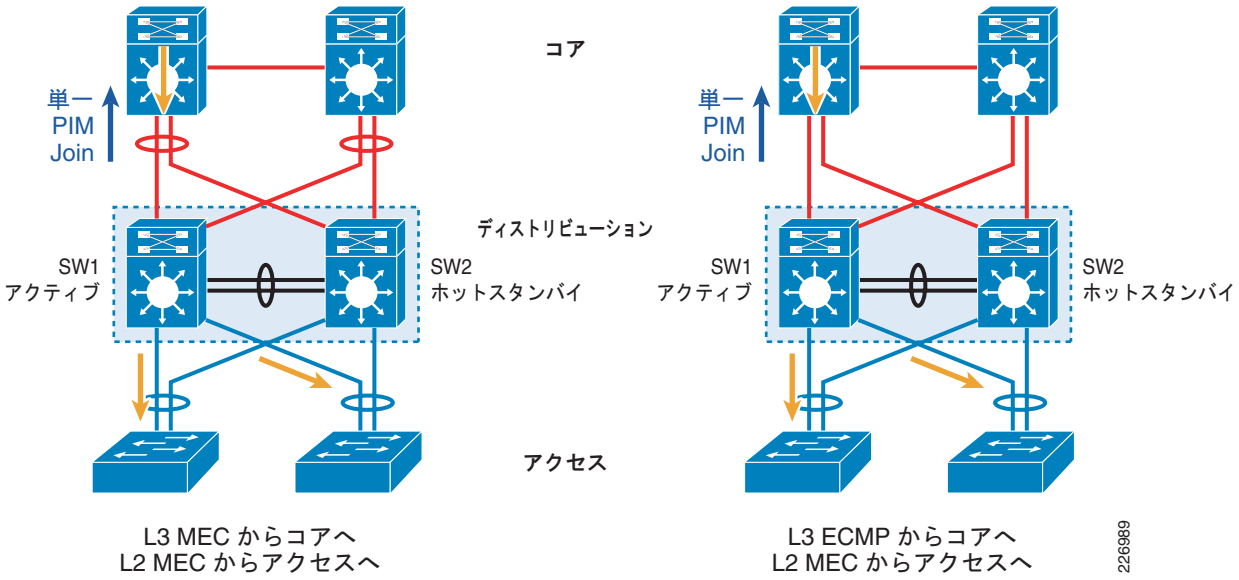
VSS は、スタンドアロン MMLS テクノロジーのメリットおよび制約事項をすべて共有します。フォワーディング ステートは、アクティブ スーパーバイザ、およびスタンバイ スーパーバイザの両方のハードウェアにプログラミングされています。スイッチオーバー中、このハードウェアはマルチキャスト データの転送を継続し、一方、コントロール プレーンは PIM ネイバー関係を回復し、再確立します。Cisco IOS リリース 12.2（18）SXF 以降が稼動している全 CEF720 ファブリック ラインカード付き Supervisor Sup720 では、マルチキャスト ingress replication（入力複製）、およびマルチキャスト egress replication（出力複製）が可能です。CFC カードによる複数の PFC ルックアップを回避するため、egress replication では、DFC 対応ラインカードの使用をお勧めします。ただし、VSS 対応コンフィギュレーションでは、マルチキャスト egress replication だけが、物理シャーシごとにサポートされています。VSS については、ingress replication モードはありません。このような制限は、VSL リンク上でマルチキャスト フローを複製する必要がある場合に、リモート ピア シャーシ上に存在するすべての発信インターフェイス リストについて、すべてのフローが複製されることを暗示しています。「レイヤ 2 MEC を使用しないマルチキャスト トラフィック フロー」（P.3-45）に説明されている、MEC をベースとしていない設計では、フローがこのように動作する可能性があります。次の CLI 出力は、このマルチキャスト機能を説明しています。

```
6500-VSS# sh platform hardware cap multicast
L3 Multicast Resources
  IPv4 replication mode: egress
  IPv6 replication mode: egress
  Bi-directional PIM Designated Forwarder Table usage: 4 total, 0 (0%) used
  Replication capability: Module                                IPv4        IPv6
                          18                                  egress      egress
                          21                                  egress      egress
                          23                                  egress      egress
                          24                                  egress      egress
                          25                                  egress      egress
                          34                                  egress      egress
                          37                                  egress      egress
                          39                                  egress      egress
                          40                                  egress      egress
                          41                                  egress      egress
  MET table Entries: Module                             Total    Used    %Used
                     18                                 65516       6       1%
                     21                                 65516       6       1%
                     34                                 65516       6       1%
                     37                                 65516       6       1%
Multicast LTL Resources
  Usage:   24512 Total, 13498 Used
```

レイヤ 2 MEC を使用したマルチキャスト トラフィック フロー

図 3-28 は、レイヤ 2 MEC ベースのネットワークにおけるマルチキャスト トラフィックの動作を表しています。レイヤ 3 接続オプションは参考です。レイヤ 3 オプションについては、「VSS を使ったルーティング」（P.3-46）で説明します。

図 3-28　レイヤ 2 MEC ベースのネットワークにおけるマルチキャスト トラフィックの動作

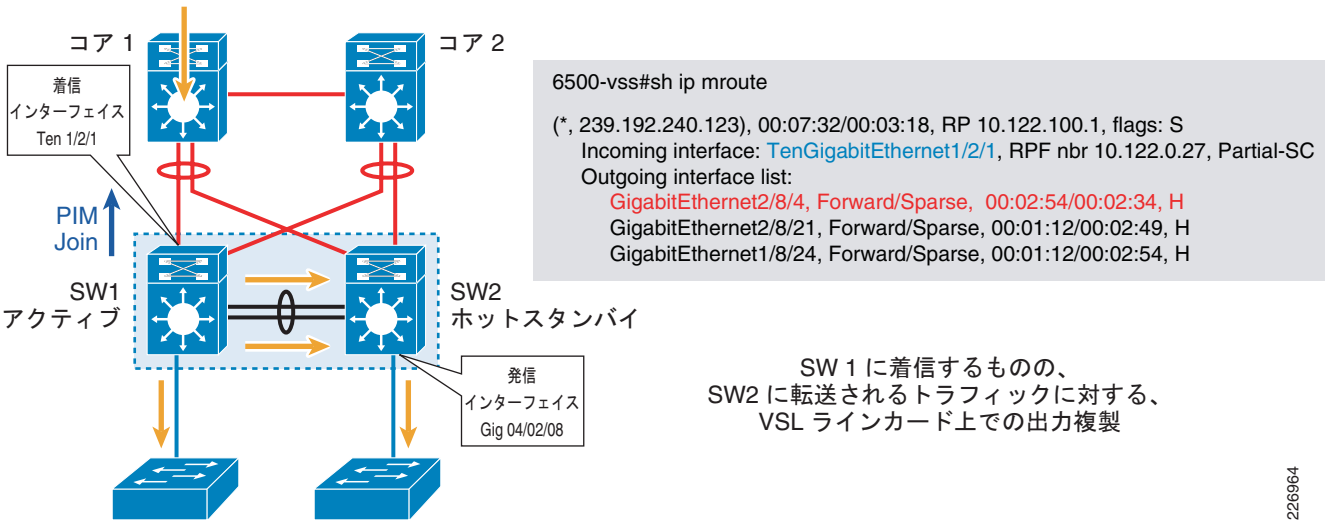


ユニキャスト データと同様、マルチキャストのコントロール プレーンは、アクティブ スイッチで処理されます。VSS は単一の論理スイッチであるため、単一の PIM Join を送信します。PIM Join の起源はアクティブ スイッチですが、ホットスタンバイ スイッチにあるインターフェイスにより送信することができます（ECMP の場合）。通常、PIM ネイバーの最大 IP アドレス（通常、マルチキャスト ストリームのソースである宛先のルーティング テーブルの先頭エントリ）が PIM Join の送信元です。着信インターフェイスの選択後、レイヤ 2 接続に基づいて、トラフィックが複製されます。MEC ベースの接続では、着信インターフェイスが存在するスイッチは、ローカルに接続された MEC メンバにマルチキャスト データを転送し、DFC ラインカードが使用できるときに出力複製を行います。

レイヤ 2 MEC を使用しないマルチキャスト トラフィック フロー

レイヤ 2 接続がレイヤ 2 MEC ベースではない（シングル ホーム接続などを使った、アクセス レイヤへの単一接続だけを持つ）、またはローカル インターフェイスの 1 つがダウンしている場合、着信および受信（複製）インターフェイスは 2 つの異なるスイッチ上に存在することができます。このタイプの接続や条件では、マルチキャスト トラフィックは VSL リンク上で複製されます。出力複製は、SW2 のすべての Outgoing Interface List（OIL; 発信インターフェイス リスト）に対する SW1 に到着するすべてのフロー（*,g および s,g）について、SW1 VSL ラインカード上で実行されます。この結果、VSL を通るデータ トラフィックの量が増えます。レイヤ 2 MEC ベースの接続は、この最適化されていないトラフィック フローを避け、VSS 対応キャンパスで MEC の必要性をさらにサポートします。図 3-29 は、レイヤ 2 MEC を使用していないマルチキャスト トラフィック フローを示します。

図 3-29　　レイヤ 2 MEC を使用しないマルチキャスト フロー



VSS：単一の論理代表ルータ

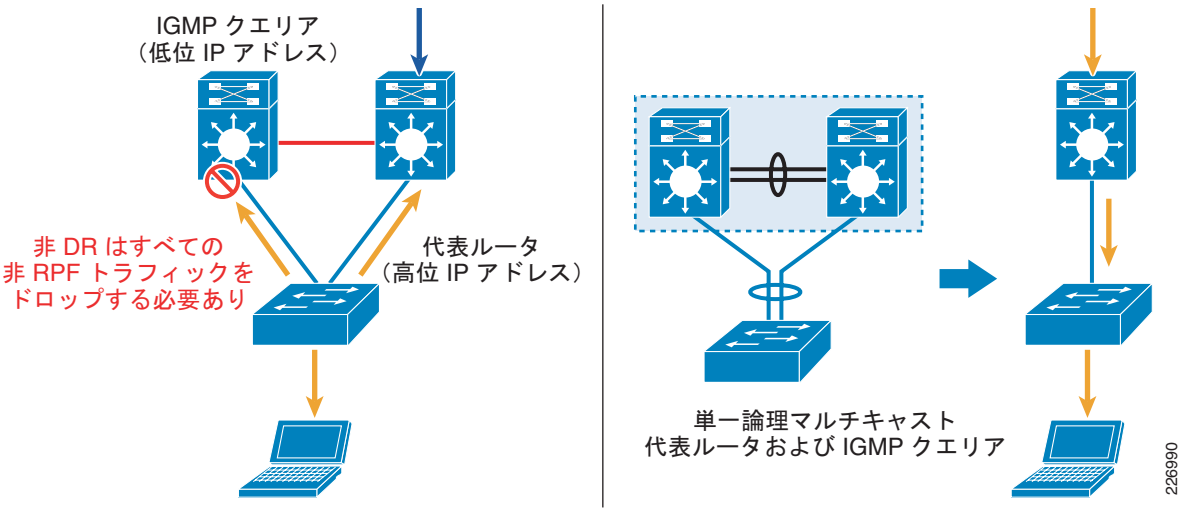
IP マルチキャストは、冗長トポロジの LAN にデータを転送するために、ルータを 1 つ使用します。複数のルータが LAN へのインターフェイスを持っている場合、データを転送するルータは 1 つだけです。LAN 上のマルチキャスト トラフィックでは、ロード バランシングは行われません。通常、マルチキャスト データの転送には、VLAN の最大 IP アドレスである Designated Router（DR; 代表ルータ）が選択されます。DR ルータで障害が発生した場合、マルチキャスト データの転送を開始するために、その他のルータ（バックアップ DR）が使用できる方法は継承されていません。転送が停止されたかどうかを判断する唯一の方法は、冗長ルータでマルチキャスト データを継続的に受信することです。つまり、アクセス レイヤ スイッチで、すべてのシングル マルチキャスト パケットをアップリンクに転送します。冗長ルータはこのデータを、LAN の発信インターフェイス上で確認します。このトラフィックは誤ったインターフェイスに到着し、Reverse Path Forwarding（RPF）チェックが失敗するため、冗長ルータはこのトラフィックをドロップする必要があります。このトラフィックは、ソースからのフローに対して反射されるため、非 RPF トラフィックと呼ばれます。IP マルチキャスト スタブ接続の詳細は、次の URL を参照してください。

http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6552/ps6592/whitepaper_c11-474791.html

非 RPF トラフィックには、2 つの副作用があります。まず、このトラフィックはアップリンク帯域幅を浪費します。次に、導入されたハードウェアに基づく適切な予防措置を講じておかなければ、CPU の利用率が高くなります。

図 3-30 に示すとおり、VSS 非対応キャンパス対 VSS 対応キャンプのマルチキャストではトポロジが変更されます。VSS は単一のマルチキャスト ルータとして扱われます。これにより、図 3-30 に示すようにマルチキャスト トポロジが簡略化されます。レイヤ 2 ドメインに接続されているノードは 1 つだけであるため、バックアップ DR は選択されません。通常の状態での VSS では、VSS スイッチ メンバ リンクの中から、どれがマルチキャスト トラフィックを受信し、着信インターフェイス リストを構築するかに基づいて、マルチキャスト フォワーダが選択されます。VSS は常にローカル フォワーディングを行うため、マルチキャスト トラフィックは、このメンバに対してローカルに接続されたリンク経由で転送されます。レイヤ 2 ドメインに接続されたリンクで障害が発生した場合、マルチキャスト データは、アクセス レイヤにデータを転送するために、VSL バンドル リンクを選択します。マルチキャスト コントロール プレーンは再収束しません。着信インターフェイス リンクで障害が発生した場合、ソースへの代替パスを検索するために、マルチキャスト コントロール プレーンは再収束する必要があります。この後で発生する障害のケースについては、「VSS を使ったルーティング」（P.3-46）で説明します。

図 3-30　　簡素化されたマルチキャスト トポロジ



VSS を使ったルーティング

このセクションでは、ディストリビューション レイヤ スイッチでの全体的な実装というコンテキストで、レイヤ 3 デバイスを使った VSS の動作および相互作用について説明します。このデザイン ガイドおよび所見は、コアおよびルーテッド アクセス設計において、同程度に適用できる VSS 実装です。このセクションの後半では、これらの VSS 設計のメリットについて説明します。通常、3 階層アーキテクチャでは、ディストリビューション レイヤにより、レイヤ 2 およびレイヤ 3 ドメインの間にバウンダリ機能が提供されます。VSS の動作は、レイヤ 3 ドメインでのスタンドアロン ノードと次のように異なります。

- ルーティング プロトコル、トポロジ、および操作
- アクティブ障害中のルーティング プロトコル相互作用

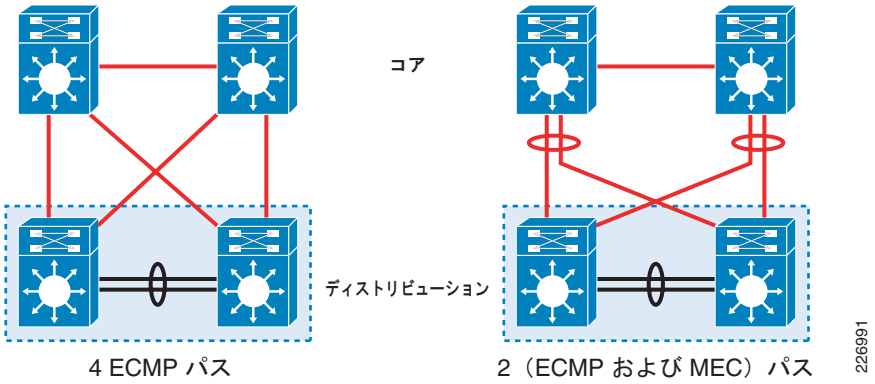
ルーティング プロトコル、トポロジ、および操作

VSS は、Enhanced Internal Gateway Routing Protocol（IGRP）、Open Shortest Path First（OSPF）、Border Gateway Protocol（BGP; ボーダー ゲートウェイ プロトコル）、Intermediate System-to-Intermediate System（IS-IS）、Routing Information Protocol（RIP）などの一般的なルーティング プロトコルをサポートしています。このデザイン ガイドで説明するのは、EIGRP、および OSPF だけです。ルーティング プロトコルと VSS コアの相互作用は、導入されているトポロジに大きく依存しています。一般に、VSS をコア デバイスに接続するには、次の 2 つの方法があります。

- Equal Cost Multipath（ECMP; 等コスト マルチパス）
- レイヤ 3 MEC

図 3-31 は、VSS を使ったフルメッシュ接続オプションを示しています。

図 3-31　フルメッシュ接続 VSS オプション



従来のベスト プラクティス キャンパス設計では、フルメッシュの導入が推奨されています。これは、リンクまたはノードでの障害は、ルーティング プロトコルに依存したトラフィックの再ルーティングを強制せず、代わりに障害復旧はハードウェアの CEF パス スイッチングに依存して行われるからです。これは、VSS デプロイにも適用されますが、その理由は異なります。（ディストリビューション レイヤに VSS と、各メンバを対応するコア デバイスに接続する 1 つの物理リンクを持つ）フルメッシュではないリンクの障害では、トラフィックは強制的に VSL バンドルへ再ルーティングされます。障害中の遅延が長くなり、VSL リンク上でトラフィックの輻輳が起こる可能性があるため、これは最適な設計の選択ではありません。フルメッシュ設計では、VSL の再ルーティングは、コアに接続するリンクが別のラインカード上で分散しておらず、ラインカードで障害が発生した場合にだけ可能です。完全冗長接続を提供するために、フルメッシュ リンクを 2 つの異なるラインカードに分散する必要があるのはこのためです。

ユニキャスト トラフィックは、ECMP ベース、および MEC ベースの両方のフルメッシュ トポロジで最適なパスをとります。各 VSS メンバは、トラフィックをローカルに使用可能なインターフェイスに転送するため、正常な動作状態では VSL バンドルを通過するトラフィックはありません。

表 3-3 は、コアでのディストリビューション ノードと従来のノードにおける VSS とのルーティング プロトコル相互作用の点で、ECMP ベースのトポロジと MEC ベースのトポロジの差をまとめたものです。

表 3-3　トポロジの比較

トポロジ	ECMP	MEC	説明
レイヤ 3 ルーテッド インターフェイス	ポイントツーポイント 4 つ	レイヤ 3 ポートチャネル 2 つ	
EIGRP または OSPF ネイバー	4 つ	2 つ	
指定された宛先に対するルーティング テーブル エントリ	4 つ	2 つ	
VSS を起源とする hello、または VSL 上のルーティング アップデート	はい（ホットスタンバイ メンバ上のネイバーに対して）	いいえ、hello を運ぶローカルに接続されたインターフェイスです	障害条件により、デフォルトの動作が変更される可能性があります
リモート デバイスを起源とする hello、または VSL 上のルーティング アップデート	はい（ホットスタンバイ メンバ上のネイバーに対して）	ハッシュの結果によっては、VSL 上を通過する可能性があります	障害条件により、デフォルトの動作が変更される可能性があります

表 3-3 に示すとおり、MEC ベースの接続により、ネイバー数が削減され、ルーティング テーブルのエントリ数も減らすことができます。これにより、ルーティング プロトコルに関連するコントロール プレーンのアクティビティに対する CPU 負荷が軽減されます。これは、コアやルーテッド アクセス設計など、多数のレイヤ 3 接続デバイスを使用し、高度にメッシュされた設計では特に便利です。

また、このトポロジにより、ネイバーと、ネットワーク上のその他のレイヤ 3 ルーテッド デバイスとの関係がどのように構築されるかが決まります。ECMP では、これは直接ポイントツーポイント接続で、ここでは、ネイバー hello は対称的に交換されます。ECMP ベースのトポロジでは、（コア デバイスと VSS の両方から）ホットスタンバイ スイッチに接続されたリンクのネイバー hello VSL リンクを通過する必要があります。これは、ネイバールータとの隣接関係は、ポイントツーポイント リンクが接続されたところから始まる必要があるからです。

MEC ベースの接続では、ネイバー hello 接続は非対称である可能性があります。MEC ベースのトポロジでは、VSS は常に hello の送信を好み、ローカルに接続されたリンク（EtherChannel の一部）から（トポロジ、または LSA を）更新します。ハッシュでは、常にローカルに使用可能なリンクが選択されます。EtherChannel 経由で VSS に接続されているリモート デバイスでも、hello やルーティング アップデートの送信時に通過するリンクを選択する際、同様のハッシュ決定プロセスが行われます。各ルーティング プロトコルについて、hello およびルーティング アップデート パケットは、宛先とは異なる IP アドレスを使用します。ルーティング プロトコル パケットのタイプに基づいて、ハッシュの結果により、アクティブ、またはホットスタンバイに接続されたリンクが選択される可能性があります。したがって、VSS のアクティブスイッチ メンバに到達するために、hello およびアップデートが異なるリンクを選択する可能性もあります。このネイバー hello およびルーティング アップデートに対するパスの選択動作は、デュアルアクティブ障害状況で、ネイバーの安定性を判断する際に、重要な役割を果たします。

ECMP および MEC トポロジを使用したデザインにおける考慮事項

この項では、VSS（ディストリビューション レイヤ）およびコアの間に導入されたトポロジの効果について説明します。トポロジの選択に影響を与える主要な設計ポイントは次の 2 つです。

- リンク障害時のコンバージェンス

- リンク障害中のフォワーディング キャパシティ（パスのアベイラビリティ）

トラフィック フローとコンバージェンス の動作については、第 4 章「コンバージェンス」を参照してください。

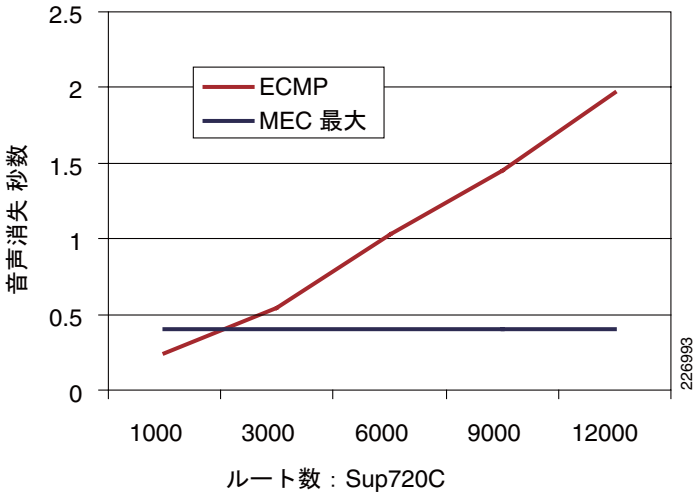
リンク障害時のコンバージェンス

図 3-32 に ECMP および MEC を使用している場合のリンク障害時のコンバージェンスを示します。ただし、ECMP ベースのトポロジ コンバージェンスはルーティング テーブルのサイズに左右されます。

ECMP

ルーティング エントリの数が大きくなると、失われる VoIP データの数が大きくなったことが測定されました（図 3-32 を参照）。これは、CEF が、障害の発生したリンク上で VoIP フローを再調整する必要があるからです。この復旧がルーティング プロトコルに左右されないとしても、宛先へのパスの再プログラミングは、ルーティング テーブルのサイズによっては、長い時間がかかります。

図 3-32　ルータの数 対 音声の損失



MEC

EtherChannel の検出はハードウェアベースで行われます。正常なメンバ リンクを通る VoIP フローのリンクと調整の障害には、一貫性があります。MEC リンク メンバ障害のワーストケース損失は約 450 ミリ秒です。平均では、Cisco Catalyst 4500 および Cisco Catalyst 3xxx スイッチング プラットフォームにより、200 ミリ秒の復旧が期待できます。

リンク障害中のフォワーディング キャパシティ（パスのアベイラビリティ）

ECMP

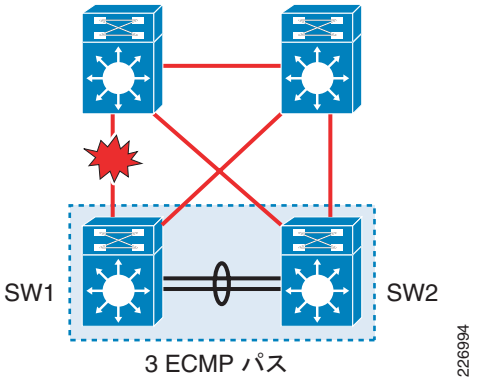
正常な動作状態では、VSS はメンバ スイッチ 1 つあたり 2 つのパス、つまり、合計 4 つのフォワーディング リンクを持ちます。次の出力例は、指定された宛先に対するルーティング テーブル エントリを示しています。

```
6500-VSS# sh ip route 10.121.0.0 255.255.128.0 longer-prefixes
D       10.121.0.0/17
```

```
[90/3328] via 10.122.0.33, 2d10h, TenGigabitEthernet2/2/1
[90/3328] via 10.122.0.22, 2d10h, TenGigabitEthernet2/2/2
[90/3328] via 10.122.0.20, 2d10h, TenGigabitEthernet1/2/2
```

どのようなシングルリンク障害でも、ECMP パスの再プログラミングという結果になります。その他の 3 つのリンクはすべて動作可能で、フォワーディングに利用できます。図 3-33 を参照してください。

図 3-33　　リンク障害時の効果



「レイヤ 3 ECMP トラフィック フロー」（P.3-7）で説明したとおり、SW1 は引き続き、トラフィック フローを、ローカルに使用可能なインターフェイスに転送します。SW2 は 2 つのリンクを持ち、トラフィックの転送を続けます。つまり、ECMP トポロジでは、トラフィックの転送に、3 つのパスをすべて使用できます。したがって、論理帯域幅のアベイラビリティは、物理リンクのアベイラビリティと変わりません。

MEC

VSS からコアへの MEC ベースのトポロジでは、論理ルーテッド ポートチャネル インターフェイスが 2 つだけ使用可能です。これは、1 つのスイッチ メンバ（シングル論理スイッチ）につき、ECMP パス 2 つだけの提供につながります。リンク障害の効果は、使用可能なパスのルーティング プロトコルとメトリックの再計算によって変わります。

auto-cost reference bandwidth を使った OSPF

キャンパス接続に対する高帯域幅インターフェイスの統合により、OSPF の参照帯域幅の調整が必要になる可能性があります。参照帯域幅を調整しなかった場合、OSPF Shorted Path First（SPF）ベースのアルゴリズムは、100 Mbps を超えるインターフェイス帯域幅のコストを区別できません。この結果、ルーティングが最適ではなくなり、予期しない輻輳が発生します。参照帯域幅は通常、100 Mbps のデフォルトから、ネットワークで使用可能な限り最高の帯域幅に調整されます。参照帯域幅は、2 つの 10 ギガビット メンバ リンクから構成される MEC インターフェイス バンドルを持つ VSS で、簡単に 20 ギガビットに到達できます。Cisco IOS では、基本となる物理リンクにより、ルーテッド リンクのメトリック（OSPF コスト）を反映させることができます。MEC のメンバ リンクで障害が発生したときに、OSPF 対応インターフェイス（ルーテッド）コストが変更されるように auto-cost reference bandwidth が構成されている場合、フォワーディング機能はパスのアベイラビリティとは異なります。このデザイン ガイドでは、VSS はポートチャネル インターフェイス経由でコア（2 つの 10 ギガビット リンク）に接続されています。MEC メンバ リンク障害は、ポートチャネル インターフェイスの 1 つでコストが上昇するきっかけとなり、この結果、指定された宛先へのルートが取り消されます。物理トポロジの観点から見ると、トラフィックを転送できるインターフェイスは 3 つあります。しかし、効果的なフォワーディング キャパシティは、シングル コアから出たたった 1 つの使用可能な論理パスに左右されます。

図 3-34　auto-cost reference bandwidth を使った OSPF

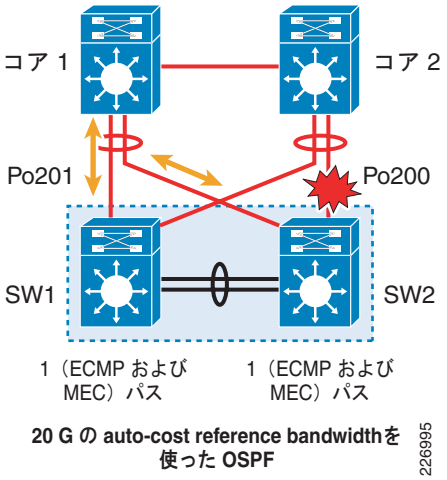


図 3-34 は、OSPF およびレイヤ 3 MEC に関連するメトリック変更動作を示しています。ポートチャネル（Po200）からの 1 つのリンクで発生した障害は、ポートチャネル全体から学習されたルートが削除される原因になります。その結果、2 つのルートの代わりに、コア ルータの 1 つからのルート（Core-1）だけが使用可能になります。VSS では、これは両方のメンバに接続された 1 つの論理パス経由で反映されます。ここで、SW1 には物理リンクが 2 つありますが、Core-1 に接続されたリンクだけが使用されていることに注意してください。論理パスのフォワーディング アベイラビリティは、図 3-34 では、黄色の矢印で示されています。次の **show** コマンドからの出力は、auto-cost reference bandwidth に関連する動作を示しています。

```
6500-VSS# show running-config | begin router ospf
router ospf 100
 router-id 10.122.0.235
 log-adjacency-changes detail
 auto-cost reference-bandwidth 20000
 nsf
 area 120 stub no-summary
 area 120 range 10.120.0.0 255.255.0.0 cost 10
 area 120 range 10.125.0.0 255.255.0.0 cost 10
 passive-interface default
 no passive-interface Port-channel200
 no passive-interface Port-channel201
 network 10.120.0.0 0.0.255.255 area 120
 network 10.122.0.0 0.0.255.255 area 0
 network 10.125.0.0 0.0.3.255 area 120
```

2 ポートチャネル インターフェイスを使用した正常な動作状態で使用できるルーティング、およびハードウェア CEF パスは、次のコマンド出力に現れています。

```
6500-VSS# sh ip route 10.121.0.0
Routing entry for 10.121.0.0/16
  Known via "ospf 100", distance 110, metric 13, type inter area
  Last update from 10.122.0.20 on Port-channel201, 00:51:31 ago
  Routing Descriptor Blocks:
  * 10.122.0.27, from 30.30.30.30, 00:51:31 ago, via Port-channel200
      Route metric is 13, traffic share count is 1
    10.122.0.20, from 30.30.30.30, 00:51:31 ago, via Port-channel201
      Route metric is 13, traffic share count is 1

6500-VSS#sh mls cef 10.121.0.0 16 sw 1

Codes: decap - Decapsulation, + - Push Label
Index  Prefix              Adjacency
```

```
108803 10.121.0.0/16       Po201           , 0012.da67.7e40 (Hash: 007F)
                           Po200           , 0012.da65.5400 (Hash: 7F80)
6500-VSS#sh mls cef 10.121.0.0 16 sw 2

Codes: decap - Decapsulation, + - Push Label
Index  Prefix              Adjacency
108802 10.121.0.0/16       Po201           , 0012.da67.7e40 (Hash: 007F)
                           Po200           , 0012.da65.5400 (Hash: 7F80)
```

次の出力リストは、2 つのポートチャネル インターフェイスを伴うメンバリンク障害中のハードウェア CEF パス アベイラビリティを示しています。この出力リストからは、使用可能な物理パスは 3 つあるのに、各スイッチの論理パスは 1 つしか使用できないことがわかります。

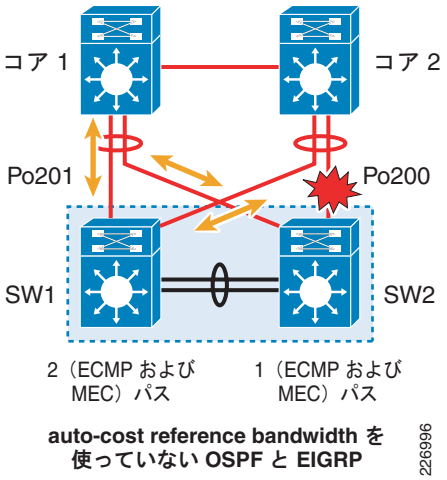
```
6500-VSS# sh ip route 10.121.0.0
Routing entry for 10.121.0.0/16
  Known via "ospf 100", distance 110, metric 13, type inter area
  Last update from 10.122.0.20 on Port-channel201, 00:51:31 ago
  Routing Descriptor Blocks:
  * 10.122.0.20, from 30.30.30.30, 00:51:31 ago, via Port-channel201
      Route metric is 13, traffic share count is 1

6500-VSS# sh mls cef 10.121.0.0 16 sw 1

Codes: decap - Decapsulation, + - Push Label
Index  Prefix              Adjacency
108803 10.121.0.0/16       Po201           , 0012.da67.7e40 (Hash: 007F)

6500-VSS# sh mls cef 10.121.0.0 16 sw 2

Codes: decap - Decapsulation, + - Push Label
Index  Prefix              Adjacency
108802 10.121.0.0/16       Po201           , 0012.da67.7e40 (Hash: 007F)


6500-VSS# sh ip os ne

Neighbor ID     Pri   State           Dead Time   Address         Interface
10.254.254.8      0   FULL/  -        00:00:36    10.122.0.20     Port-channel201
10.254.254.7      0   FULL/  -        00:00:39    10.122.0.27     Port-channel200

6500-VSS# sh run int po 200
Building configuration...

Current configuration : 378 bytes
!
interface Port-channel200
 description 20 Gig MEC to cr2-6500-1 4/1-4/3
 no switchport
 dampening
 ip address 10.122.0.26 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 logging event spanning-tree status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
end

6500-VSS#sh run int po 201
Building configuration...
```

```
Current configuration : 374 bytes
!
interface Port-channel201
 description 20 Gig to cr2-6500-1 4/1-4/3
 no switchport
 dampening
 ip address 10.122.0.21 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 logging event spanning-tree status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
end
```

auto-cost reference bandwidth を使っていない OSPF

auto-cost reference bandwidth（100 Mbps）を使って OSPF を構成しているネットワークでは、リンクメンバの障害により、ルーテッド ポートチャネル インターフェイスのコストが変わることはありません。つまり、ルートは取り消されず、システムには宛先への 2 つのルートが残されます。しかし、これにより、3 つの物理パス（SW2 に 1 つ、SW1 に 2 つ）をすべて使用することができます。図 3-35 は、auto-cost reference bandwidth を使用していない OSPF ベースの環境を示しています。

図 3-35　auto-cost reference bandwidth を使っていない OSPF



次の CLI 出力は、リンク障害*前*と*後*で、ルーティング テーブルとハードウェア CEF のステートは変わらないということを示しています。ここで示されているのは、2 つのポートチャネル インターフェイスを使用した、正常な動作状態で使用できるルーティング、およびハードウェア CEF パスです。

```
6500-VSS# sh ip route 10.121.0.0
Routing entry for 10.121.0.0/16
  Known via "ospf 100", distance 110, metric 13, type inter area
  Last update from 10.122.0.20 on Port-channel201, 00:51:31 ago
  Routing Descriptor Blocks:
  * 10.122.0.27, from 30.30.30.30, 00:51:31 ago, via Port-channel200
      Route metric is 13, traffic share count is 1
```

```
    10.122.0.20, from 30.30.30.30, 00:51:31 ago, via Port-channel201
      Route metric is 13, traffic share count is 1
6500-VSS# sh mls cef 10.121.0.0 16 sw 2

Codes: decap - Decapsulation, + - Push Label
Index  Prefix              Adjacency
108803 10.121.0.0/16       Po201           , 0012.da67.7e40 (Hash: 007F)
                           Po200           , 0012.da65.5400 (Hash: 7F80)
6500-VSS# sh mls cef 10.121.0.0 16 sw 1

Codes: decap - Decapsulation, + - Push Label
Index  Prefix              Adjacency
108802 10.121.0.0/16       Po201           , 0012.da67.7e40 (Hash: 007F)
                           Po200           , 0012.da65.5400 (Hash: 7F80)
```

次の出力例に示すように、リンク障害により、ルーティング パスはそのまま保たれ、障害の発生したリンク（SW1）のハードウェア CEF パスだけが削除されます。

```
6500-VSS# sh ip route 10.121.0.0
Routing entry for 10.121.0.0/16
  Known via "ospf 100", distance 110, metric 13, type inter area
  Last update from 10.122.0.20 on Port-channel201, 00:51:31 ago
  Routing Descriptor Blocks:
  * 10.122.0.27, from 30.30.30.30, 00:51:31 ago, via Port-channel200
      Route metric is 13, traffic share count is 1
    10.122.0.20, from 30.30.30.30, 00:51:31 ago, via Port-channel201
      Route metric is 13, traffic share count is 1
6500-VSS# sh mls cef 10.121.0.0 16 sw 2

Codes: decap - Decapsulation, + - Push Label
Index  Prefix              Adjacency
108803 10.121.0.0/16       Po201           , 0012.da67.7e40
6500-VSS#

6500-VSS# sh mls cef 10.121.0.0 16 sw 1

Codes: decap - Decapsulation, + - Push Label
Index  Prefix              Adjacency
108802 10.121.0.0/16       Po201           , 0012.da67.7e40 (Hash: 007F)
                           Po200           , 0012.da65.5400 (Hash: 7F80)
```

EIGRP

EIGRP メトリック計算は、遅延の合計と最小帯域幅をあわせたものです。メンバ リンクで障害が発生すると、EIGRP は変更された帯域幅の値を認識し、使用しますが、遅延は変更されません。メトリック計算で使用されるのは、パスの最小帯域幅であるため、これは合成メトリックに影響する可能性もしない可能性もあります。したがって、ローカル帯域幅の変更は、それがパスの最小帯域幅である場合だけ、メトリックに影響します。キャンパス ネットワークでは、コアと VSS の間で変更された帯域幅はギガビット順に提供されますが、大半のルートでは、通常、これは最小帯域幅ではありません。したがって、実用的には、EIGRP は帯域幅変更の影響を受けず、デフォルトの auto-cost reference bandwidth を使った OSPF と同じ動作になります。合成メトリックが影響を受けるような状況では、EIGRP は、auto-cost reference bandwidth が設定された OSPF と同様の動作をします。

まとめ

OSPF およびレイヤ 3 MEC トポロジを使用した設計の選択は、障害発生中に使用可能な合計帯域幅の選択で、ユーザ データ コンバージェンスに対する影響ではありません。これはパケットの損失が最小限に抑えられるからです。詳細については、「ルーティング（VSS からコアへ）コンバージェンス」（P.4-14）を参照してください。

ルーティングのループを回避するため、auto-cost reference bandwidth コンフィギュレーションは、ネットワーク全体で同一でなければなりません。ただし、VSS については、auto-cost reference bandwidth を設定しないという例外をもうけることができます。これが可能なのは、ベスト プラクティス設計では、通常、アクセス レイヤは完全スタブ エリアとして構成されているからです。このようなトポロジでは、アクセス レイヤからコア、またはコアからアクセス レイヤへのバックドア代替パスが提供されません。auto-cost のルールを緩和することには、ルーティング プロトコルやそのコンフィギュレーションに関係なく、ユーザ データ トラフィックにより使用されるパスすべてのアベイラビリティが高くなるというメリットがあります。

また、アプリケーションの応答時間もこのコンフィギュレーション オプションを選択するときに考慮すべき点です。メトリックの変更に伴い、コアと VSS の間でのフォワーディング キャパシティが低下する可能性があります。適切な QoS マーキングは、VOIP やビデオなどの重要なトラフィックに対応します。その他の重要ではないトラフィックは、帯域幅を共有することになりますが、これは十分ではない可能性があります。キャンパス ネットワークでは、通常は、レイヤ 3 MEC ベースの接続で使用されるリンクについては、デフォルトの OSPF および EIGRP 設定のままにしておくといいでしょう。

ECMP と レイヤ 3 MEC オプションのまとめ

全体的に、レイヤ 3、MEC ベースの接続は、一貫性のあるコンバージェンスとパスのアベイラビリティ オプションを提供します。したがって、コアにレイヤ 3 MEC を実装する設計が推奨されます。ECMP およびレイヤ 3 MEC オプションを比較したまとめについては、表 3-4 を参照してください。

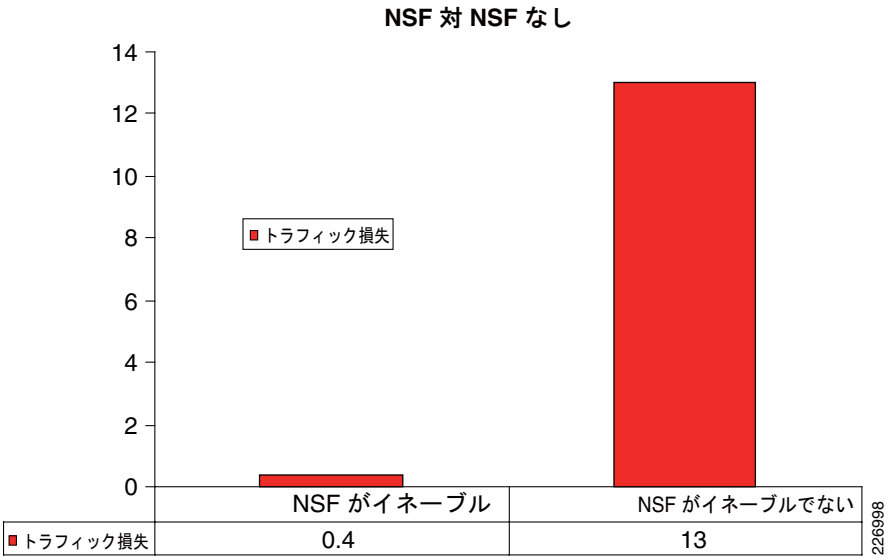
表 3-4　ECMP およびレイヤ 3 MEC オプションの比較

設計要素	ECMP からコアへ	レイヤ 3 MEC からコアへ
シングル VSS メンバでのリンク障害の復旧方法	ECMP パスをローカル メンバに切り替える	ルーティング プロトコル コンフィギュレーションに応じて、ルート取り消しを基準にしたパスの選択を行う
リンク障害中に使用可能なパス	3 つ	ルーティング プロトコル コンフィギュレーションによって、2 つまたは 3 つ

アクティブ障害中のルーティング プロトコル相互作用

「ステートフル スイッチオーバー：統合コントロール プレーンおよび分散データ転送」（P.2-23）で説明したとおり、VSS はアクティブおよびホットスタンバイの 2 つのスーパーバイザから構成されています。アクティブ スーパーバイザで障害が発生したとき、SSO ベースの同期により、SSO に対応したすべてのプロトコルの回復が支援されます。ルーティング プロトコルの復元力と回復は SSO の一部ではありません。スイッチオーバー中、ホットスタンバイ スーパーバイザは、ルーティング プロトコルを再初期化する必要があります。この結果、ネイバールータは隣接関係がリセットされたことに気づきます。これには、VSS から学習されたダウンストリーム サブネットのルートが削除されるという副作用があります。このような損失を回避するには Non-Stop Forwarding（NSF; ノンストップ フォワーディング）を使って VSS を構成し、ネイバールータを NSF に対応させる必要があります。図 3-36 は、NSF を有効化しなかった場合の影響を示しています。この図から、VSS と隣接するルーティング デバイスで NSF 機能を有効化しなかった場合、最高 13 秒分のトラフィック損失につながる可能性があることがわかります。

図 3-36　NSF と非 NSF での音声の損失



ヒント　VSS と隣接するルーティング ノードでは、NSF を実行することを強くお勧めします。

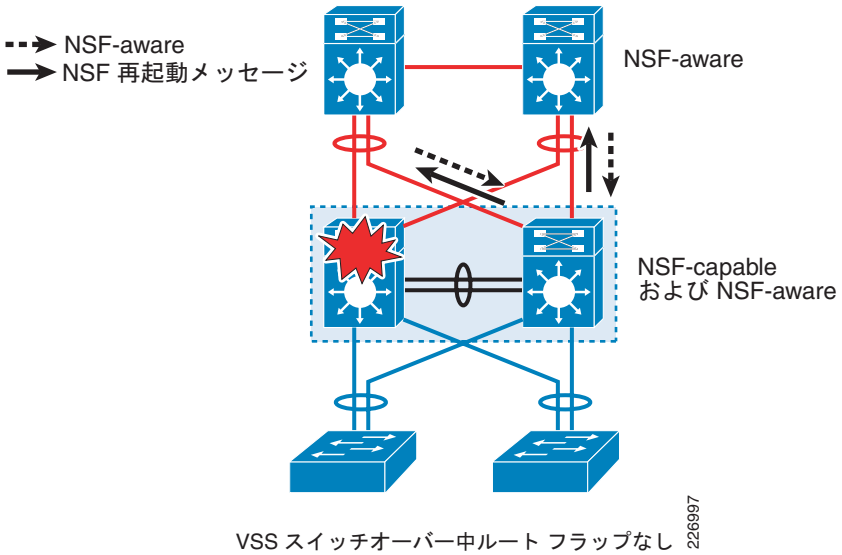
NSF 要件および復旧

NSF はルーティング プロトコルのグレースフル リスタートを提供します。これにより、ルーティング プロトコルはフェールオーバー中も復旧中のコントロール プレーンを認識し続けます。隣接性がリセットされることはありません。NSF では、ルータは既知のルートに沿ってデータの転送を続けながら、ルーティング プロトコル情報を復元することができます。SSO は、スイッチオーバー発生時に、インテリジェントなプロトコル復旧を提供します。これはスイッチオーバー中も継続的にパケットを転送するためです。しかし、ルーティング プロトコルが障害中に、障害が発生したイベントに反応を示した場合、再起動されるシステムへのパスが変更され、パケットの転送が行われなくなり、SSO の有効性が低下します。NSF は、ルーティング トポロジを維持し、ハードウェア フォワーディング テーブルをグレースフルに更新することにより、特に、スイッチオーバー中のパケット損失を軽減するように設計されています。NSF 復旧において重要なコンポーネントは次のとおりです。

- *NSF-capable ルータ*：スイッチオーバー中に引き続き転送を行う能力を持つルータは *NSF に対応*しています。NSF-capable ルータは、隣接する NSF-capable、または NSF-aware ルータからルーティング情報を再構築できます。
- *NSF-aware ルータ/NSF ヘルパー*：ネイバー ルータによる NSF 再起動の実行を支援する NSF 互換ソフトウェアが稼動しているルータです。再起動中の（NSF に対応した）ルータにトラフィックを転送し続けられる範囲のルーティング プロトコルの拡張をサポートするデバイスは *NSF-aware* です。NSF に対応した Cisco のデバイスは、NSF を認識できるデバイスでもあります。ルーティング機能をサポートする Cisco スイッチング プラットフォームはすべて、NSF-aware をサポートしています。

NSF の復旧プロセスのまとめについては、図 3-37 を参照してください。

図 3-37　NSF の復旧



NSF の復旧は、スーパーバイザ フェールオーバー中にルーティング プロトコル レベルで行われる NSF-capable、および NSF-aware ルータの相互作用によって決まります。フェールオーバー中、ルーティング プロトコルは新たにアクティブになったスーパーバイザで再起動します。図 3-37 は、NSF の復旧が行われている NSF-capable ルータを示しています。NSF の復旧は、CEF およびルーティング プロトコル拡張に左右されます。NSF リカバリ モードでは、NSF-capable ルータは CEF から Routing Information Base（RIB; ルーティング情報ベース）の接続を解除します。この接続解除により、ハードウェアでパケットの転送を続けながら、個別にコントロール プレーンを復旧できるようになります。

図 3-37 のとおり、再起動中のルータ（ホットスタンバイ）は、その NSF-aware ネイバーに、ネイバー関係を再初期化しないように通知します。再起動インジケータを受信したルータは、ネイバー ステータスをホールドダウン モードに設定します。ほとんどの場合、再起動インジケータは、hello パケットに再起動フラグを設定し、復旧処理中に、短い間隔で hello パケットを送信します（「NSF の復旧、および IGP の相互作用」（P.3-58）で説明されているとおり、この機能は、ネイバーの hello タイマー値に影響を与えます）。NSF に対応していないネイバーは再起動インジケータを無視し、隣接をディセーブルにするため、ルートが削除され、これがパケットの損失につながります。1 つのネットワークに NSF に対応しているルータと対応していないルータを混在させないことを強くお勧めします。隣接のリセットを回避するプロセスは、ルートのコンバージェンス回避にも役立ちます。これは、NSF-capable ルータ経由で通知されるネットワークでは、ルートの再計算は行われないからです。NSF の復旧中、ルーティング プロトコル ネイバーは特別なリカバリ モードに入ります。ルーティング プロトコル ネイバー交換の詳細については、URL http://www.cisco.com/en/US/tech/tk869/tk769/technologies_white_paper0900aecd801dc5e2.shtml を参照してください。

上の URL は、汎用の NSF/SSO デザイン ガイドラインにリンクしています。これ以降、キャンパス関連の設計について、詳しく説明します。

NSF の復旧、および IGP の相互作用

NSF は、コンバージェンスの回避を前提に設計されています。これは、障害の発生しているドメインをローカルにし、Interior Gateway Protocol（IGP）タイマーにより長いルート コンバージェンスが決定されるのを回避するという方針にうまく適合しています。IGP ネイバー タイマーは、使用可能な代替パスをすばやく検出し、提供するために用意されています。このため、NSF が使用可能な環境では、フェールオーバーで隣接のリセットを回避しなければならないような、IGP ネイバー デッド タイマーが検出されたかどうかを判断する必要があります。IGP デッド タイマーは、次の式の結果よりも大きくなければなりません。

SSO の復旧 + ルーティング プロトコルの再起動 + 最初の hello を送信するまでの時間

スタンバイ スーパーバイザがアクティブになると同時に、OSPF は Fast hello パケットを 2 秒間隔で送信し、スイッチオーバー後のコンバージェンス時間を早めます。EIGRP は、タイマーの復旧用に独自のメカニズムを備えています。この操作の詳細については、次の URL を参照してください。

http://www.cisco.com/en/US/tech/tk869/tk769/technologies_white_paper0900aecd801dc5e2.shtml

各イベントに直接関係する復旧時間は、hello タイマーの下限（最小値）を制御します。SSO の復旧では、コントロール プレーンが初期化され、同期されたデータベースを使って（実行ステート）プロトコルが実行されます。ルーティング プロトコルの再起動では、複数のコンポーネントが初期化されます（ルーティング プロセスの開始、接続済みネットワークの再構築、および CEF プロセスの相互作用が行われます）。最後に、再起動フラグとともに hello パケットを送信するときに、ネイバーごとに hello パケットを処理し、カプセル化するために必要な時間が計上されます。VSS では、指定された有効な環境で、この時間は 9 ～ 13 秒です。

OSPF および EIGRP で推奨される時間については、表 3-5 を参照してください。これは、ルーティング テーブルに 3000 ルートを持つコアにある Cisco Catalyst 6500 Sup720 に基づく所見です。

表 3-5　NSF に対する IGP タイマー要件

ルーティング プロトコル	標準 IOS	モジュラ IOS
EIGRP hello/ホールド秒	5/15：デフォルト	5/15 秒
OSPF hello/デッド秒	10/40：デフォルト	10/60 秒

ここで、モジュラ Cisco IOS のタイマー要件はネイティブ Cisco IOS のタイマー要件よりも大きい値になっているため、OSPF デッド タイマーにデフォルトの 40 秒よりも大きい値の設定が必要な可能性があります。

（注） BGP および IGP 相互作用のデザイン要件は、キャンパス固有のデザイン目標では評価できないため、さらに調整が必要である可能性があります。

表 3-5 にまとめられたタイマーは、ベスト プラクティス ベースのキャンパス ネットワークにおける最低要件を表しています。

ヒント **表 3-5 に記載されている値を下回る値に*調整しない*ことを強くお勧めします。その他の NSF 関連のルート タイマーはすべてデフォルト値のままにし、変更しないでください。**

OSPF

ルーティング プロセスは、アクティブ スーパーバイザだけで実行されます。スタンバイ スーパーバイザには、OSPF に関連するルーティング情報や、Link-State Database（LSDB; リンクステート データベース）は含まれません。また、ネイバーのデータ構造も保持されません。スイッチオーバーが発生す

ると、ネイバー関係は必ず再確立されます。NSF-capable ルータは、ネイバー ステートを完全に再起動する必要がありますが、NSF-aware ネイバー ルータはネイバーのリセットを回避するために、NSF-capable ルータからシグナリングされる特別な再起動ビットを使って、リカバリ モードに入ります。

ネイバー隣接交換に関する次の syslog メッセージは、NSF-aware ルータで NSF が再起動されたことを示しています。

```
%OSPF-5-ADJCHG: Process 100, Nbr 10.120.250.4 on Port-channel6 from FULL to
EXSTART,OOB-Resynchronization
%OSPF-5-ADJCHG: Process 100, Nbr 10.120.250.4 on Port-channel6 from EXSTART to EXCHANGE,
Negotiation Done
%OSPF-5-ADJCHG: Process 100, Nbr 10.120.250.4 on Port-channel6 from EXCHANGE to LOADING,
Exchange Done
%OSPF-5-ADJCHG: Process 100, Nbr 10.120.250.4 on Port-channel6 from LOADING to FULL,
Loading Done
```

OSPF NSF-aware ルータは INIT から FULL にはなりません。その代わりに、OOB-Resynchronization ステートで、FULL から EXT-START になります。ただし、NSF-capable ピアでは、6 つの OSPF ネイバー ステート交換すべてを通じて、完全な再起動シーケンスが実行されます。

EIGRP

EIGRP にも、ネイバーに NSF 再起動を知らせるための同様の方法が用意されています。しかし、この方法では、OSPF のようなステートの遷移は見られません。次の syslog メッセージは、NSF の再起動を表しています。

```
%DUAL-5-NBRCHANGE: IP-EIGRP(0) 100: Neighbor 10.120.0.211 (Port-channel2) is up: peer NSF
restarted
```

コンフィギュレーション、およびルーティング プロトコルのサポート

Cisco NSF は、EIGRP、OSPF、BGP、および IS-IS ルーティング プロトコルでサポートされています。

（注）　このデザインおよび操作ガイドは、OSPF および EIGRP だけをカバーしています。

Cisco Catalyst スイッチにおける NSF 機能のコンフィギュレーションは非常に簡単です。これは、各ルーティング プロトコル インスタンスで、**nsf** キーワードにより有効化されます。NSF 対応スイッチは、自動的に NSF を認識するようになります。NSF-aware 機能は、ルーティング プロトコルがサポートされていれば、これに組み込まれます。特別なコンフィギュレーションは必要ありません。しかし、NSF-aware 機能をサポートするソフトウェア リリースは必要です。次のコマンド例は、NSF 機能のコンフィギュレーションを表しています。

EIGRP：

```
Router(config)# router eigrp 100
Router(config-router)# nsf
```

OSPF：

```
Router(config)# router ospf 100
Router(config-router)# nsf
```

（注）　Cisco IOS は、IETF ベースのグレースフル リスタート拡張、および Cisco バージョンのグレースフル リスタートの両方をサポートしています。

NSF のモニタリング

次の **show** コマンドの例は、ルーティング プロトコルのタイプに基づき、NSF-capable および NSF-aware ルータで NSF のコンフィギュレーションおよびステートを監視する方法を示しています。

OSPF

次の **show** コマンド出力からは OSPF が NSF に対応していることがわかります。この出力文は、リンクローカル シグナリングをサポートしています。これは、このルータが NSF を認識していることも示します。この例は、NSF が有効化されているときに、最後の再起動が行われた様子を示しています。これは、NSF 再起動が完了するまでにかかった時間を示している点に注意してください。NSF 再起動は、NSF/SSO のスイッチオーバー時間ではなく、ルーティング プロトコルの再同期を意味します。データ フォワーディング コンバージェンス時間は表しません。タイマーのガイドラインについては、「NSF の復旧、および IGP の相互作用」（P.3-58）を参照してください。

```
Router# sh ip ospf
 Routing Process "ospf 100" with ID 10.120.250.4
 Start time: 00:01:37.484, Time elapsed: 3w2d
 Supports Link-local Signaling (LLS)
! <snip>
LSA group pacing timer 240 secs
 Interface flood pacing timer 33 msecs
 Retransmission pacing timer 66 msecs
 Non-Stop Forwarding enabled, last NSF restart 3w2d     ago (took 31 secs)
```

ここで、主要な出力は、Link Local Signaling（LLS）オプション出力です。OSPF NSF は、スタンバイ スーパーバイザに OSPF ステート情報を保持しないため、新たなアクティブ スーパーバイザは、その LSDB とネイバーを同期する必要があります。これは、Out-of-Band Resynchronization（OOB-Resync; アウトオブバンド再同期）により行われます。

次の出力例にある LR ビットは、このネイバーが NSF を認識し、ローカル ルータで NSF の再起動をサポートできることを示しています。ここで、最初のネイバーは、OOB-Resync の出力を使用して、NSF 復旧を行っています。OOB-Resync メッセージは 2 つ目のネイバーから欠落しています。これは、NSF 復旧が行われていないからです。

```
Router# sh ip ospf neighbor detail
 Neighbor 10.122.102.2, interface address 10.120.0.200
    In the area 120 via interface Port-channel6
    Neighbor priority is 0, State is FULL, 7 state changes
    DR is 0.0.0.0 BDR is 0.0.0.0
    Options is 0x50
    LLS Options is 0x1 (LR), last OOB-Resync 3w2d     ago
    Dead timer due in 00:00:07
Neighbor 10.122.102.1, interface address 10.120.0.202
    In the area 120 via interface Port-channel5
    Neighbor priority is 0, State is FULL, 6 state changes
    DR is 0.0.0.0 BDR is 0.0.0.0
    Options is 0x50
    LLS Options is 0x1 (LR)
    Dead timer due in 00:00:05
```

EIGRP

EIGRP も類似した復旧方式をとっています。アクティブとなるスーパーバイザはルーティング プロセスを初期化し、hello および INIT パケットに RS ビットを入れて、NSF-aware ネイバーに信号を送信する必要があります。EIGRP NSF 機能を検出するには、**show ip protocol** コマンドを使用します。次の出力は、EIGRP で NSF 機能のデフォルト タイマーが有効化されていること、また、NSF が認識されていることを示しています。タイマーのガイドラインについては、「NSF の復旧、および IGP の相互作用」（P.3-58）を参照してください。

```
Router# sh ip protocol
*** IP Routing is NSF aware ***
Routing Protocol is "eigrp 100 100"
! <snip>
EIGRP NSF-aware route hold timer is 240s
  EIGRP NSF enabled
     NSF signal timer is 20s
     NSF converge timer is 120s
```

VSS でのレイヤ 3 マルチキャスト トラフィック デザインにおける考慮事項

VSS は、さまざまなマルチキャスト機能、および関連するデザイン オプションをサポートしています。マルチキャスト機能のすべてを網羅した検証やガイダンスはこのデザイン ガイドの範囲を超えています。このデザイン ガイドでは、VSS 対応キャンパスのマルチキャスト トラフィックに影響を与える重要な設計ポイントについて説明します。マルチキャスト トラフィックの動作に影響を与える要因のうち、この設計では、Rendezvous Point（RP; ランデブー ポイント）の配置、RP フェールオーバー、RP としての VSS などには対応していません。第 4 章「コンバージェンス」では、重要な障害状況について説明します。ただし、マルチキャスト トポロジを使用した大規模な検証はこのデザイン ガイドの範囲を超えています。VSS 対応キャンパスで、レイヤ 3 でのマルチキャスト トラフィック相互作用に影響を与える重要な設計要因は次のとおりです。

- ECMP と MEC のトラフィック フローの比較
- ECMP および MEC での VSS メンバ障害の影響

ECMP と MEC のトラフィック フローの比較

VSS は 1 つのマルチキャスト ルータを表します。PIM Join は、ルーティング テーブルで使用可能な ECMP パスのうち、最も大きい PIM ネイバー IP アドレス（通常、指定された発信元の、ルーティング テーブルの先頭エントリ）に基づいて送信されます。PIM Join が送信されるスイッチと受信されるスイッチは異なるため、Incoming Interface List（IIL; 着信インターフェイス リスト）と OIL（発信インターフェイス リスト）は、結果として得られるマルチキャスト フォワーディング トポロジが、VSL リンクを通じて転送可能なマルチキャスト トラフィックに組み込まれるように、非対称的に形成されます。PIM Join が同一の物理 VSS メンバ スイッチとの間でやりとりされない場合、図 3-38 に示すとおり、マルチキャスト トラフィックを VSL リンク経由で渡すことができます。

図 3-38　VSL リンク経由で渡されるマルチキャスト トラフィック

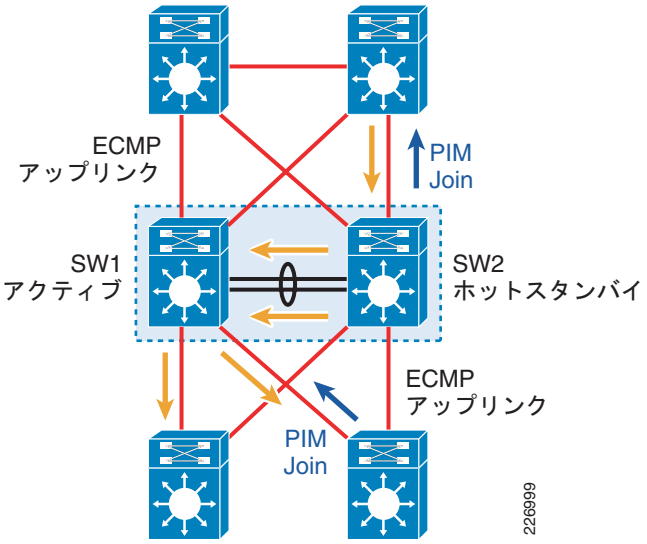
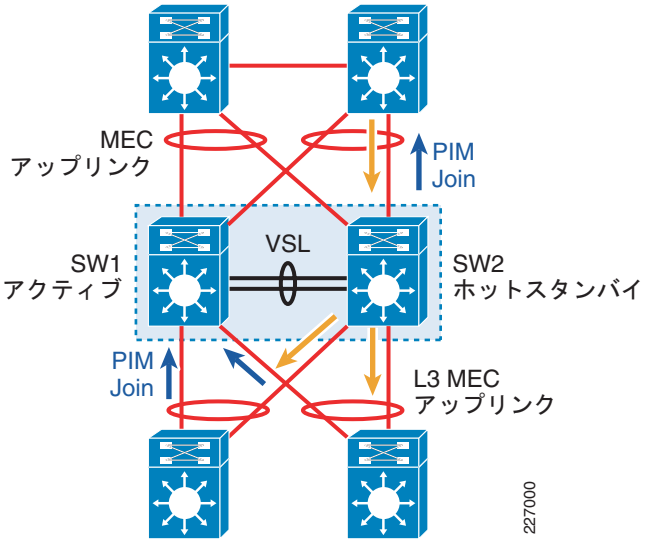


図 3-38 では、一番下にあるレイヤ 3 デバイスが、SW1 に接続されたリンク経由で、ルーティング テーブルでの最高位エントリ（最も大きい PIM IP アドレス）に基づき PIM Join を送信します。しかし、PIM Join は、SW2 に接続するリンク上の VSS により送信されます。VSS 経由で送信される PIM Join は 1 つだけであるため、マルチキャスト トラフィックの着信インターフェイスは SW2 上に構築されます。SW2 では、発信インターフェイス リストはローカルに構築されていないため、SW1 が発信インターフェイス リストを構築します。SW2 は VSS スイッチの一部であるため、ユニファイド コントロール プレーンは、VSL バンドル経由でトラフィックを複製（出力の物理的な複製）しなければならないことを知っています。この複製は、フローごと（*,g および s,g）に、個々の発信インターフェイス リスト エントリに対して行われます。この結果、VSL リンクに対し、帯域幅が膨大に要求されるとともに、マルチキャスト トラフィックの遅延がさらに長くなります。これを解決するには、図 3-39 に示すように MEC ベースの接続を使用します。

図 3-39　MEC ベースの接続オプション



MEC ベースのトポロジでは、VSS の個々の物理スイッチ上で着信インターフェイス リスト（IIL）と発信インターフェイス リスト（OIL）を使用して、非対称的な PIM Join 処理を行うことが可能です。図 3-39 に示すとおり、IIL は SW2 に構築されますが、OIL は SW1 に構築されます。しかし、IIL と OIL は両方とも、ポートチャネル インターフェイス上に構築されています。SW1 トラフィック上にやってきた PIM Join が SW2 によって転送されたとしても、マルチキャスト トラフィックは SW2 に到着します。これは、ポートチャネル インターフェイスのインスタンスが、デフォルトで、両方のスイッチ上に存在するからです。VSS は、ユニキャストおよびマルチキャスト トラフィックを転送するために、常にローカルに使用可能なインターフェイスを使用します。したがって、マルチキャスト トラフィックは、VSL バンドル経由ではなく、ローカル リンク経由で転送されます。MEC コンフィギュレーションのため、マルチキャスト トラフィックは、ハッシュの結果に基づいていずれかのリンク メンバを選択することができますが、このトポロジでは、どちらの側にも MEC が実装されているため、トラフィックは VSL バンドルを通過しません。リンクで障害が発生した場合、マルチキャスト トラフィックは VSL リンクを横切り、ローカル スイッチの複製を行います。これは、マルチティア（マルチコアや、ルーテッド アクセス設計など）のレイヤ 3 デバイスとして導入されている VSS で可能なタイプのトポロジです。マルチキャスト トラフィックを使ったルーテッド アクセス環境でのアクセスから、MEC アップリンクを使用します。

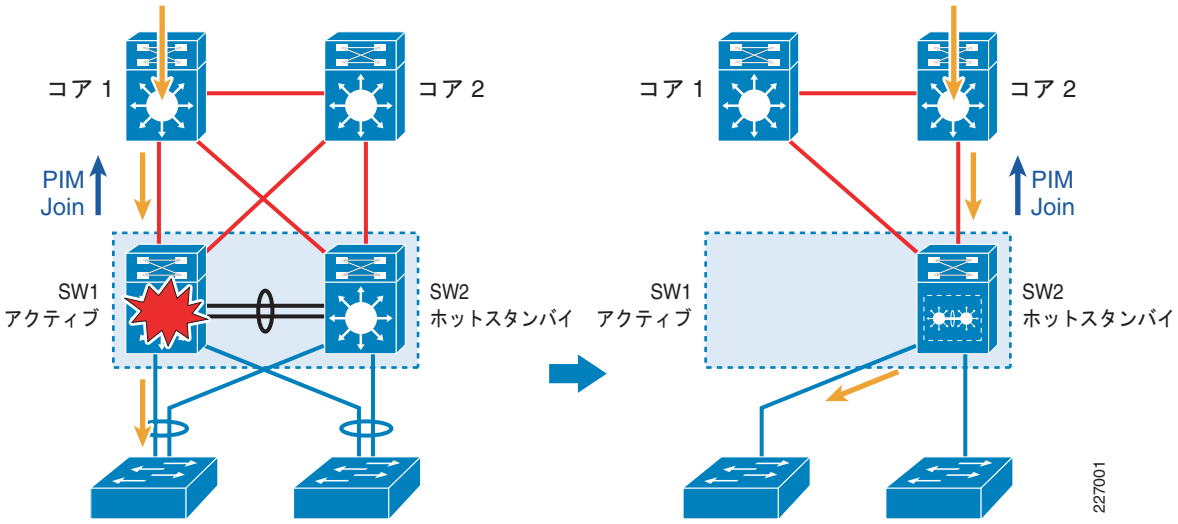
ヒント　シスコでは、レイヤ 3 の MEC ベースのトポロジを使用して、VSL バンドルでのマルチキャスト トラフィックの複製を回避し、VSL リンクでのトラフィックの再ルーティングで発生する遅延を防ぐことをお勧めします。

ECMP および MEC での VSS メンバ障害の影響

ECMP

ECMP ベースのトポロジで、PIM Join は、ルーティング テーブルで使用可能な ECMP パスのうち、最も大きい PIM ネイバー IP アドレス（通常、指定された発信元の、ルーティング テーブルの先頭エントリ）に基づいて送信されます。リンクまたはノードの障害により、PIM ネイバーが切断されると、必ず、マルチキャスト コントロール プレーンは、使用可能なインターフェイスで新たに PIM Join を発行し、マルチキャスト フォワーディング ツリーを再構築する必要があります。ECMP では、マルチキャスト コンバージェンスは着信インターフェイスの場所によって異なります。図 3-40 は、着信インターフェイスがアクティブ スイッチ上に構築されている場合のマルチキャスト トラフィック フローの動作を示しています。

図 3-40　マルチキャスト フォワーディング ツリーの再構築



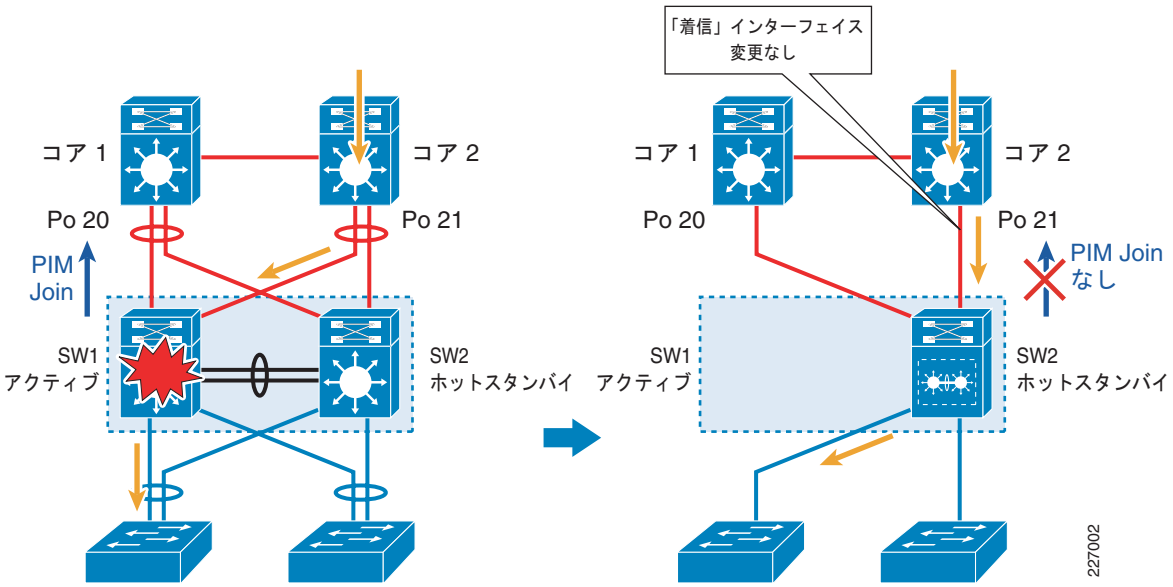
マルチキャスト ストリームの着信インターフェイスが SW1（図 3-40）上に構築されているときに、SW2 がこのマルチキャスト ストリームを転送する前に SW1 で障害が発生した場合、SW2 は新しい Shortest-Path Tree（SPT）の構築と、新たなアクティブ スイッチ（SW2）での着信インターフェイスの選択を要求します。マルチキャスト データの配信は、マルチキャスト コントロール プレーン コンバージェンス経由で新しいパスが発見されるまで停止します。このタイプの障害に関連するコンバージェンスは多様に変化し、ランデブー ポイント（RP）Reverse Path Forwarding（RPF; リバース パス フォワーディング）チェック、ユニキャスト ルーティング プロトコル コンバージェンスなど多数の要因に基づいて、その範囲は 2 ～ 3 分から、それ以上までさまざまです。

問題の発生しているスイッチが、マルチキャスト トラフィック（ここには示されていない）の着信インターフェイスを運んでいないような障害である場合、コンバージェンスの範囲は 200 ～ 400 ミリ秒です。図 3-40 では、マルチキャスト フローの着信インターフェイスは SW1 に構築されています。SW2 で障害が発生した場合、この着信インターフェイスは変更されません。したがって、マルチキャスト コントロール プレーンは新たな PIM Join を送信する必要はありません。トラフィックは、引き続き、ハードウェア内を転送されます。

MEC

MEC ベースの接続では、いずれかのメンバ スイッチで障害が発生すると、EtherChannel メンバの 1 つが使用可能なまま残されます。コア ルータは、ポートチャネル インターフェイスでマルチキャスト データを転送します。VSS スイッチ メンバに障害が発生した場合、コア ルータの観点からは、マルチキャスト トラフィックの着信インターフェイスは変更されず、そのままです。コア ルータがしなければならないことは、残りのリンク メンバへのフローの再ハッシュだけです。これは、マルチキャスト コントロール プレーンにステートの変更が報告されないことを暗示しています。MMLS テクノロジーは、マルチキャスト ステート（*,g および s,g）を同期したままにし、スイッチ ハードウェアはマルチキャスト トラフィックのスイッチングを続けます。アクティブからホットスタンバイへのスイッチオーバー中、マルチキャスト コンバージェンスの範囲は常に 200 ～ 400 ミリ秒です。図 3-41 は、MEC ベースの接続でのマルチキャスト トラフィック フローの動作を示しています。

図 3-41　マルチキャスト トラフィックに対する着信インターフェイスを持たないスイッチでの障害



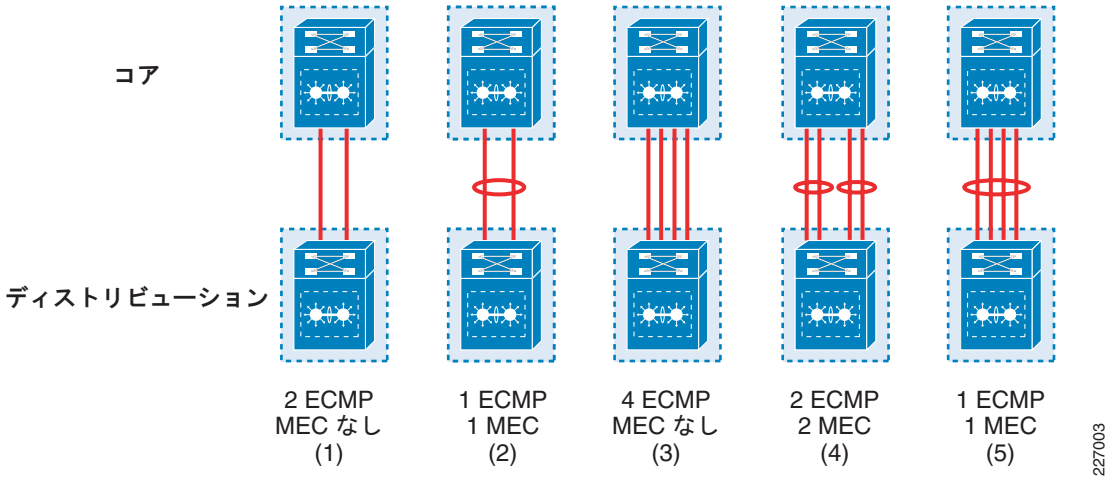
PIM は、ECMP と同じルールに従って、Join（最大の PIM ネイバー アドレス）を送信するためのルーテッド インターフェイスを選択します。しかし、レイヤ 3 MEC では、PIM Join は、ハッシュ値に基づいて、リンク メンバの 1 つを選択できます。その結果、この Join はコア ルータのいずれかに到達できます。図 3-41 は、PIM Join が Core-2 に送信され、その結果、着信インターフェイスが Core-2 に構築されたマルチキャスト トラフィック フローの動作を示しています。Core-2 は、マルチキャスト フロー発信元と宛先の IP アドレスのハッシュに基づいて、SW1 または SW2 へのマルチキャスト フローの転送を選択できます。図 3-41 は、ハッシュの結果により、SW1 に接続されたリンク メンバが選択された様子です。SW1 で障害が発生すると、コア ルータの両方で、ポートチャネルからリンク メンバが 1 つ削除されます。その一方で、SW2 はアクティブ スイッチの役割を引き継ぎます。SW2 に接続されている 1 つのリンク メンバでポートチャネル インターフェイスが依然としてアクティブであるため、Core-2 の観点から見ると、マルチキャストの着信インターフェイスは変更されていません。Core-2 は、再ハッシュ後も、マルチキャスト フローを送信し続けます。マルチキャスト フローを受信した SW2 は、このトラフィックを、古いアクティブからのマルチキャスト エントリをハードウェアベースの MMLS 同期することにより、転送します。

マルチキャスト コンバージェンスは多数の要因に左右されます。しかし、MEC ベースのトポロジを使用した VSS の主な利点は、ノード障害コンバージェンスを 1 秒未満に抑えられるところにあります。さらに、ECMP ベースの接続とは対照的に、マルチキャスト コンバージェンスは、着信インターフェイスの構築場所に依存しません。mroute（*,G および S,G）数の多いマルチキャスト ネットワークでは、特定の障害からの復旧では、コンバージェンスが 1 秒を超える可能性があります。

コアの VSS

このデザイン ガイドの主な焦点は、ディストリビューション レイヤの VSS アプリケーションにあてられていました。しかし、この項ではコアの VSS アプリケーションについて簡単に説明します。これまでに説明した設計要因はすべて、コアの VSS にも適用されます。コア レイヤの設計で考慮される要因は多数あります。コア レイヤとディストリビューション レイヤの両方が VSS を使用している場合、考慮すべき設計要因はこれらのレイヤの接続だけです。図 3-42 に示すとおり、コアおよびディストリビューション レイヤでの VSS の主な接続オプションは 5 種類あります。この図はコア、リンク、およびディストリビューション レイヤの仮想化の論理的な結果を表しています。

図 3-42　VSS コアおよびディストリビューションの接続オプション



ヒント　レイヤ 2 環境では、2 つの VSS 間に単一の論理リンク（オプション 5）を配置する方法が*唯一の*推奨トポロジです。他の接続シナリオではトポロジがループします。

VSS コアとディストリビューションに対する最適な接続オプションの決定に最も影響のあるものは、表 3-6 に示した多数の設計要因のうち、太字で強調されているものです。オプション 4 および 5 は、表 3-6 で強調されている要因のコンテキストで詳しく説明します。

表 3-6　トポロジ オプションの設定要因

	トポロジ オプション				
設計要素	**ECMP リンク 2 つ、各シャーシから 1 つずつ（1）**	**2 つのリンク、各シャーシから 1 つ、MEC から 1 つ（2）**	**4 つのリンク、フル メッシュ、ECMP（3）**	**4 つのリンク、フル メッシュ、MEC 2 つ、ECMP 2 つ（4）**	**4 つのリンク、フル メッシュ、MEC 1 つ（5）**
物理リンク総数	2	2	4	4	4
論理リンク総数	0	1	0	2	1
レイヤ 3 での合計リンク数	2	1	4	2	1
ECMP ルーティング パス	2	0	4	2	0
ルーティング オーバーヘッド	**2 倍**	**1 倍**	**4 倍**	**2 倍**	**1 倍**
ネイバー数の減少	なし	あり	なし	あり	あり
シングル リンク障害からの復旧	**VSL 経由**	**VSL 経由**	**ECMP**	**MEC**	**MEC**
マルチキャスト トラフィックの復旧	可変	一貫性がある	可変	一貫性がある	一貫性がある
CEF 負荷分散	あり	なし	あり	あり	なし
MEC 負荷分散の利点	なし	あり	なし	あり	あり
混合負荷分散 - CEF および MEC	なし	なし	なし	あり	なし

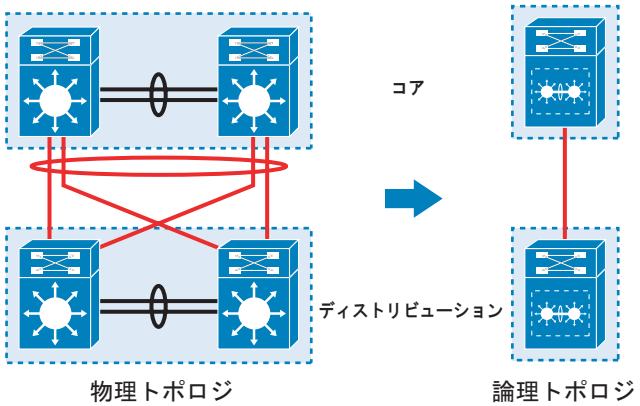
表 3-6　トポロジ オプションの設定要因 （続き）

設計要素	トポロジ オプション				
	ECMP リンク 2 つ、各シャーシから 1 つずつ（1）	2 つのリンク、各シャーシから 1 つ、MEC から 1 つ（2）	4 つのリンク、フルメッシュ、ECMP（3）	4 つのリンク、フルメッシュ、MEC 2 つ、ECMP 2 つ（4）	4 つのリンク、フルメッシュ、MEC 1 つ（5）
デュアルアクティブ トラスト サポート	なし	あり	なし	あり	あり
リンク障害に伴うメトリック変更への影響	なし	なし	なし	あり	なし
コンフィギュレーションとトラブルシューティングの複雑さ	中	高	中	中	低
単一のリンク障害でのコンバージェンス	可変	可変	可変	約 100 ミリ秒	約 100 ミリ秒
推奨されるベスト プラクティス Core ルーティング設計	推奨されない	推奨されない	推奨されない	OK	ベスト

1 つのレイヤ 3 MEC：フルメッシュ ポートチャネル インターフェイス リンク用 - オプション 5

図 3-43 は、フルメッシュ環境を対象にしたシングル レイヤ 3 MEC を示しています。

図 3-43　フルメッシュ環境用シングル レイヤ 3 MEC



この設計には、次のような長所があります。

- この設計は、ルーティング トポロジとネイバー メンテナンスの面で、ルーティング コントロール プレーンのオーバーヘッドを本質的に軽減します。
- リンク障害やノード障害はルーティング テーブルのサイズに依存しないため、復旧には一貫性があります。
- リンク障害により、トラフィックが強制的に VSL バンドルを通過することはないため。VSL リンクの遅延と輻輳が軽減されます。

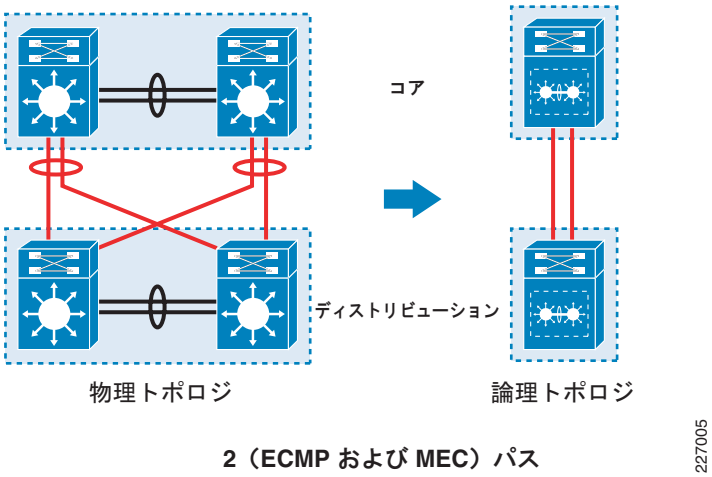
- 1 つのパス スルーが経由する論理デバイスは 1 つであるため、プロトコルが OSPF（auto-cost がデフォルトかどうか）でも EIGRP でも関係なく、パスのアベイラビリティは、ルーティング アップデート、またはメトリックの変更による影響をほとんど受けません。
- WS-6708 ハードウェアに Fast Link Notification（FLN）機能が用意されているため、リンク障害のあるコンバージェンスの平均は約 100 ミリ秒です。
- 論理インターフェイスが 1 つであるため、コンフィギュレーションやトラブルシューティングのオーバーヘッドが軽減されます。

レイヤ 3（CEF 負荷分散用）とレイヤ 2（EtherChannel ハッシュ負荷分散）が明確に使用されている場合、マルチステージでの負荷分散方式を使用できます。この設計オプションには、論理的なレイヤ 3 パスが 1 つしかなく、CEF 負荷分散は使用できません。コアのトラフィック パターンが、レイヤ 3 および レイヤ 2 負荷分散方式を使用した、より細かいレベルのトラフィック制御を必要としている場合は、この設計オプションは理想的ではない可能性があります。一般的なネットワークでは、このように細かいレベルの制御は必要とされません。したがって、通常は、コンフィギュレーションの単純さ、コントロール プレーン オーバーヘッドの低さ、および常に短いコンバージェンス時間により、レイヤ 3 MEC を 1 つ使ったソリューションが最も好まれます。

2 つのレイヤ 3 MEC：2 つのレイヤ 3（ECMP）ポートチャネル インターフェイス（それぞれ 2 つのメンバを持つ）：オプション 4

図 3-44 は、2 つの レイヤ 3（ECMP）ポートチャネル インターフェイスを備えた環境での 2 つのレイヤ 3 MEC を示しています。

図 3-44　それぞれ 2 つのメンバを持つ 2 つのレイヤ 3（ECMP）ポートチャネル インターフェイス



このオプションには、レイヤ 3 MEC を 1 つ使った接続とほとんど同じ利点がありますが、次の点も考慮してください。

- ルーティング コントロール プレーンのオーバーヘッドが大きくなります。
- コンフィギュレーションおよびトラブルシューティングのオーバーヘッドが大きくなります。
- パフォーマンスはメトリックの変更とルーティング プロトコルのコンフィギュレーションに依存します。

この設計の主な利点は、マルチステージ（レイヤ 3 CEF、およびレイヤ 2 EtherChannel 負荷分散）が使用できるため、必要に応じて、より詳しいトラフィック制御ができるところにあります。

ECMP フル メッシュ：オプション 3

このオプションについては、この項の最初で説明しました。選択肢の中では、これは通常、望ましいオプションではありません。

正方メッシュ トポロジおよびフル メッシュではないトポロジ：各シャーシから 1 つのリンク：オプション 1 および 2

これは、今までに説明したオプションと比べて欠点が多く、最も望ましくないトポロジです。オプション 1 および 2 が最も望ましくないという最大の理由は、リンク障害中に VSL リンクを通過するトラフィックの依存関係にあります。

VSS を使ったルーテッド アクセス デザインの利点

ルーテッド アクセス設計はマルチレイヤ設計に代わるものです（第 1 章「Virtual Switching Systems のデザインの概要」を参照）。ルーテッド アクセスは、単にレイヤ 3 バウンダリをアクセス レイヤに延長します。いろいろな意味で、ルーテッド アクセス設計は、VSS 対応キャンパス設計に似ています。これらのモデルは両方とも、トポロジを単純化し、リンクやノードの障害によってもたらされるトポロジ変更を減らすことにより、同じ問題を解決します。これらには、たとえば次のような共通する利点があります。

- 実装が簡単で、コンフィギュレーションを減らすことができます。
- FHRP に依存しません。
- STP/HSRP/GLBP プライオリティを照合しません。
- 単一の代表ルータ経由でレイヤ 2 / レイヤ 3 マルチキャスト トポロジに不整合はありません。
- 単一のコントロール プレーンで、デバイスと障害ドメインの管理がしやすくなっています。
- コンバージェンス時間に矛盾がなく、GLBP/HSRP の調整に依存しません。

これらの設計アプローチに共通する利点は、ルーテッド アクセス設計における VSS の採用という明らかな疑問につながります。レイヤ 2 ドメインの VSS は、VLAN をスパニングさせながら、関連するリスクを取り除いていくことにより、マルチレイヤ実装を改善するという点において、大きく貢献しています。ルーテッド アクセス設計での VSS の使用もおそらく有益でしょう。次の項では、まず、ルーテッド アクセスと VSS における重大なコンポーネント障害復旧について説明します。最後に、ルーテッド アクセス設計に VSS を導入する利点をまとめます。

ディストリビューション レイヤの復旧

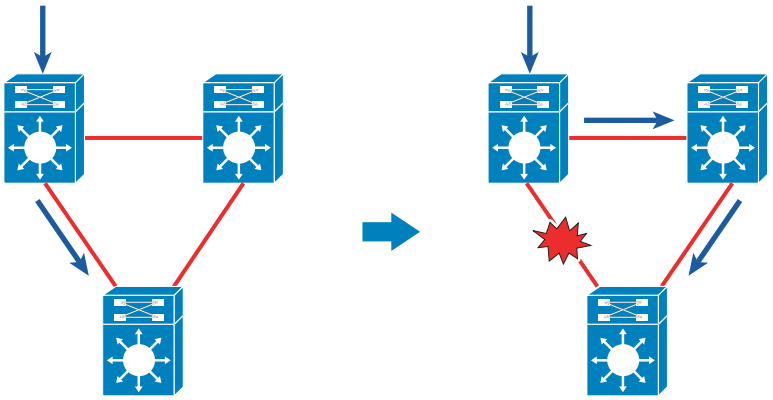
ルーテッド アクセス

ルーテッド アクセス設計の主な利点は、非常に高速なコンバージェンス（200 ミリ秒）にあります。リンク障害の場合、復旧はディストリビューション ノード間のレイヤ 3 リンクを通じた再ルーティングに依存しています。この復旧は、特に次の要因に依存しています。

- インターフェイスのダウンを検出するまでの時間
- ルーティング プロトコルが代替ルートの検索を収束するまでの時間

検出時間は、インターフェイスのタイプ（ファイバ、または銅線）、およびインターフェイス コンフィギュレーションの物理面によって変わります。制御可能な要素は、ルーティング プロトコルの動作だけです。ルーティング プロトコルがどのくらい早く障害イベントを検出、通知、および対応できるかによって、コンバージェンスの速度が決まります。図 3-45 は、一般的な復旧プロセスを表しています。

図 3-45　一般的なルーティング プロトコルの復旧プロセス



OSPF/EIGRP ダウンストリームの復旧

227006

OSPF では、コンバージェンスは、1 秒未満のレベルへの SPF および LSA タイマーの調整、アクセス レイヤ サブネットの集約、およびアクセス レイヤで定義されたエリアのタイプに左右されます。これら 3 つの重要なコンポーネントがすべて適切に構成されていれば、200 ミリ秒で収束することも可能になります。

EIGRP にはタイマーへの依存性がありません。しかし、EIGRP スタブ、および集約はコンバージェンスにとって重要です。

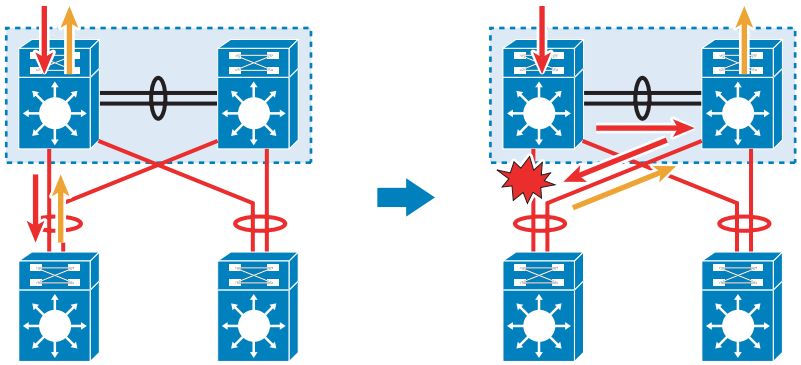
関連する障害およびコンフィギュレーションのガイダンスの詳細については、付録 B「参考資料」の『*Routed Access Design Guide*』を参照してください。

VSS 復旧

アクセス レイヤ リンクの復旧にレイヤ 3 リンク経由の再ルーティングが必要になるルーテッド アクセス設計と同様、VSS もダウンストリーム トラフィック フローの復旧に対して同様の動作をします（「VSS 対応キャンパスのトラフィック フロー」（P.3-5）を参照）。VSS シナリオでは、アクセス レイヤのリンク障害により、ダウンストリーム トラフィックが強制的に VSL 経由で再ルーティングされます。アップストリーム トラフィック フローは EtherChannel 経由で復旧されます（アクセス レイヤの残っているリンクでフローが再ハッシュされます）。

ディストリビューション時に導入された VSS により、ルーティング プロトコル コンバージェンスへの依存性は解消されます。MEC ベースのトポロジには、メンバ リンク障害が、ポートチャネル インターフェイスをダウンさせることはないという利点があります。アクセス レイヤ リンクで障害が発生した場合、ポートチャネル インターフェイスは無効化されないため、ダウンストリーム トラフィックでのルーティング アップデートは必要ありません（図 3-46 を参照）。その結果、コア ルータのルーティング トポロジではコンバージェンスは行われません。

図 3-46　ルーティングの変更を伴わないダウンストリーム復旧

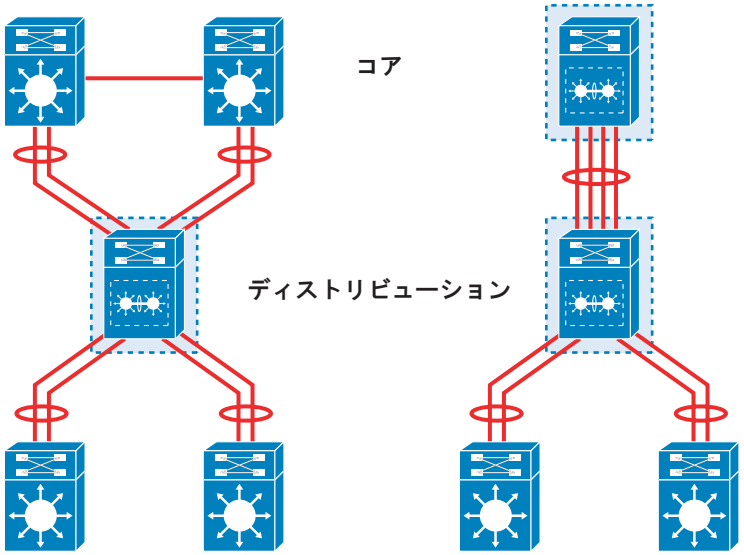


ルート変更なしの OSPF/EIGRP ダウンストリームの復旧

227007

ディストリビューション レイヤで VSS を使用することにより、よりシンプルかつ高速に、さらに高度なレベルにコンバージする環境の作成というルーテッド アクセス設計の目的がさらに強化されます。これにより、コンフィギュレーション、および障害ドメインの複雑さが低減されます。このデザイン ガイドでは、ルーテッド アクセス キャンパス設計における VSS アプリケーションについて、そのすべての側面をカバーすることはできません。しかし、前の項で説明した重要な障害復旧を検証することにより、このような設計の利点を判定します。図 3-47 は、VSS を使ったルーテッド アクセス設計の 2 つのモデルを示しています。1 つ目のモデルは、ディストリビューション レイヤでの VSS の使用を表しています。2 つ目は、レイヤ 3 MEC を使ったトポロジをさらに簡略化し、ディストリビューション レイヤ、およびコアでの VSS の使用を図にしたものです。

図 3-47　VSS 対応ルーテッド アクセス



ディストリビューション レイヤの VSS　　ディストリビューション レイヤとコア レイヤの VSS

VSS 対応ルーテッド アクセス

227008

VSS 対応ルーテッド アクセス キャンパス デザインの利点

VSS 対応ルーテッド アクセス キャンパス デザインに関連する利点は次のとおりです。

- 簡略化され、統合されたコンフィギュレーションにより、操作上の複雑度が低下し、事態が悪化する可能性も低くなります。
- ルーテッド アクセスにより、トポロジと復旧が簡略化されます。VSS では、各レイヤに 1 つずつ、エンドツーエンドで論理デバイスが提供されるため、デザインがさらに簡略化されます。1 つの論理ルータは、コントロール プレーンでのルーティングにおけるデュアル ノードの非効率性の大半を排除しますが、同時に、トポロジ データベース、ピアリング、ネイバー関係に関連するコントロール プレーンの負荷の軽減など、さまざまなよい副産物もなくなります。
- ユーザが採用したデザインで、さらなる柔軟性を実現します。ルーテッド アクセスを使った VSS では、OSPF および EIGRP 調整条件がより柔軟になります。アクセス レイヤのリンク障害では、ルートの再計算は要求されないため、コア レイヤやディストリビューション レイヤにおける Sub-Sec （1 秒未満）のタイマー コンフィギュレーションはそれほど重要な要件ではありません。これにより、コンフィギュレーションが簡略化され、キャンパス以外のデバイスにタイマーを拡張するというトポロジ的な依存関係がなくなります。
- リンクメンバの障害により、ルーテッド インターフェイスがダウンすることはありません。これにより、コア レイヤおよびそれ以降のレイヤにルーティングの変更を通知する必要がなくなります。従来のデザインでは、このようなイベントで見られた不必要なルートの変動は、ルート集約により減少しました。VSS は、ルート集約の必要性を低減します。それでも、ベスト プラクティスをベースにしたデザインでは、コントロール プレーンの不安定性を軽減し、VSS では解決できない可能性のある障害に対処するために、集約が推奨されます。エンタープライズ ネットワークでは、レガシー計画のインストール ベースおよび継承のため、IP サブネットのサマライズが困難であることがよくあります。ルーテッド アクセスを使った VSS により、集約の重大性が軽減され、既存のネットワークに柔軟性が提供されます。
- 冗長スーパーバイザは、SSO 対応プロトコル経由でレジリエンシー（復元力）を提供します。この結果、ディストリビューション レイヤでのノードのフェールオーバー中に一貫した復旧が行われます。たとえば、OSPF、または EIGRP NSF/SSO の実装により、コンバージェンスがルーティング テーブルのサイズに依存することがなくなります（図 3-32 を参照）。
- コアおよびディストリビューション レイヤの単一の論理マルチキャスト ルータにより、マルチキャスト トポロジが簡略化され、その結果、ノード障害に対して、コアおよびディストリビューション レイヤでのコンバージェンスが 1 秒未満に抑えられます。

ハイブリッド デザイン

レイヤ 2 ドメインを拡張し、レイヤ 3 ドメインに拡張機能をもたらす VSS の機能により、マルチレイヤ デザインとルーテッド アクセス デザインを完全に統合されたデザインに融合する*ハイブリッド*デザイン アプローチの作成が可能になります。それぞれのデザインの利点を生かし、テクノロジーやビジネスにおける特定の条件に役立てることができます。ハイブリッド デザインでは、複数のクローゼットへのスパニングを必要とする VLAN は VSS で定義され、スパニング VLAN を必要としない VLAN はルーティングされ、アクセス レイヤで定義されます。このためには、VSS と、スパンされた VLAN がトランキングされ、スパンされなかった VLAN がルーティングされるアクセス レイヤの間でトランキングされたコンフィギュレーションを行います。複数のアクセス レイヤ スイッチのスパニングを必要とする機能 VLAN にはたとえば、次のものがあります。

- ネットワークの仮想化（一時的な接続をサポートするゲスト VLAN、社内接続、企業合併など）
- 会議、メディア室、公共アクセス VLAN
- Network Admission Control（NAC）VLAN（隔離、パスチャライズ、およびパッチ適用）
- スパンされた VLAN を必要とする外注グループ、およびエージェンシー間リソース

- 集中管理コントローラのないワイヤレス VLAN
- ネットワーク管理およびモニタリング（SNMP、SPAN）

データおよび音声 VLAN や、アクセス レイヤ スイッチ内に限定されているその他の接続は、ルーティング可能な VLAN に含まれます。

（注）　このハイブリッド デザイン アプローチは、本デザイン ガイドのこのリリースでは実証されていません。

CHAPTER 4

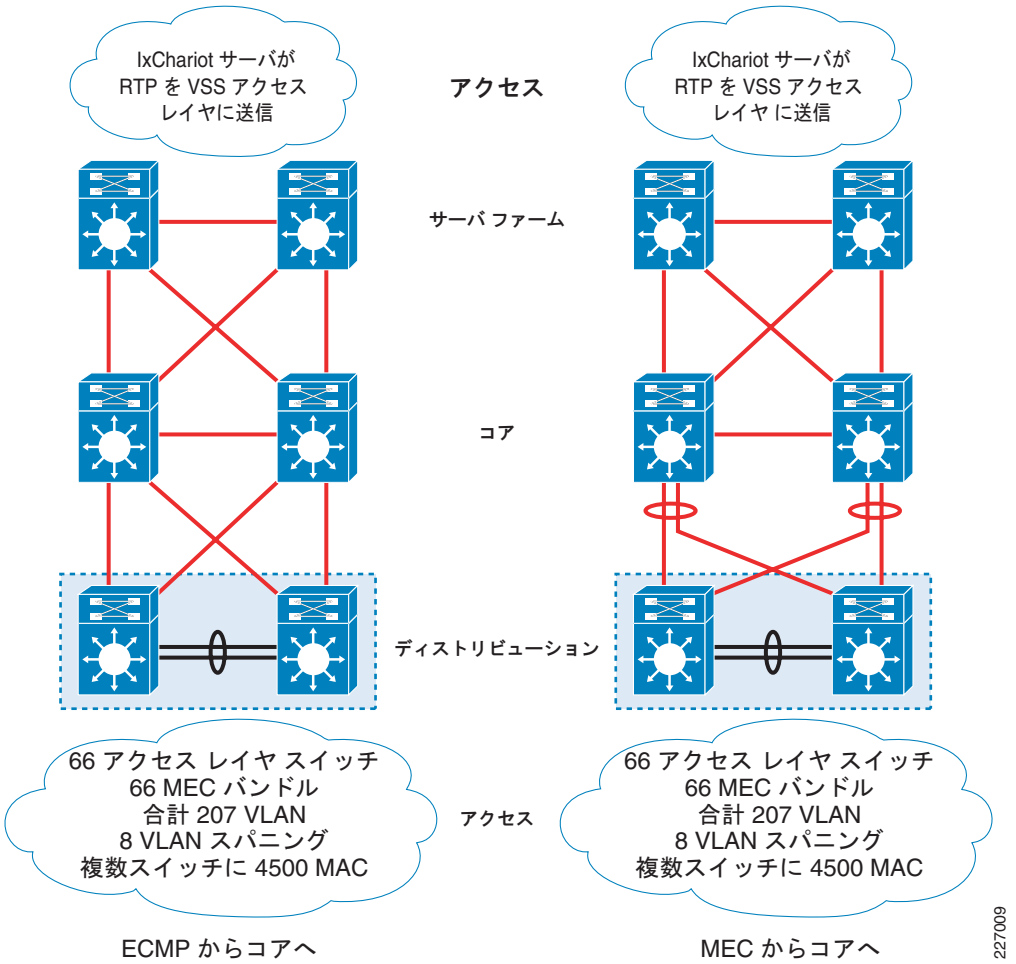
コンバージェンス

ここでは、Virtual Switching System（VSS）コンポーネントの障害の際のコンバージェンス結果とトラフィック フローについて説明します。これまでの章で詳しく説明したすべての検証済みのベスト プラクティスを使用しています。コンバージェンスのセクションでは、コンバージェンスに影響を与えるコンポーネントとテクノロジーの種類を特定する際に、「VSS 障害ドメインおよびトラフィック フロー」（P.3-9）を頻繁に使用します。検証は、ほとんどの障害の種類の組み合わせとプロトコルおよびトポロジの繰り返しに対して完了しています。ここでは、すべての障害とコンバージェンス シナリオの特性を記述することはしません。キャンパス ネットワークで重要かつ一般的な障害を扱います。

ソリューション トポロジ

VSS ソリューション検証環境は、Equal Cost Multi-Path（ECMP; 等コスト マルチパス）トポロジと MEC トポロジおよび関連するコンフィギュレーションのベスト プラクティスを扱います。図 4-1 に、一般的な VSS 環境の概要を示します。キャンパス接続のためのこの 3 つの階層は、以前の設計ガイドで確立されたベスト プラクティスを基にしています。コアよりも先の接続に関して、リファレンス トポロジは ECMP によって接続されたサーバファーム スイッチで構成されます。コンバージェンス結果のために使用されるコンフィギュレーションは、この設計ガイドを作成する際に生み出された検証済みのベスト プラクティスに基づいています。特に明記しない限り、すべてのプロトコル コンフィギュレーションで、デフォルトのタイマー設定と動作を使用します。

図 4-1　VSS ソリューション トポロジ



ソフトウェアとハードウェアのバージョン

表 4-1 に、本書が扱う VSS 環境に関連するソフトウェアとハードウェアのバージョンの要約を示します。

表 4-1　VSS ソフトウェアとハードウェアの要約

プラットフォーム	ソフトウェア リリース	ハードウェア構成	装置の役割
Catalyst 6500-E	12.2(33)SXH2(a)	Sup720-10GE	VSS DUT ディストリビューション レイヤ
		6708-10GE 6724-100/1000	
Catalyst 6500-E	12.2(33)SXH1	Sup720 6708-10GE	コア レイヤ
アクセス レイヤ			
Catalyst 6500	ネイティブ 12.2(33)SXH	Sup32-8GE 6148-GE-TX	DUT

表 4-1　VSS ソフトウェアとハードウェアの要約 （続き）

プラットフォーム	ソフトウェア リリース	ハードウェア構成	装置の役割
Catalyst 6500	CatOS 8.6	Sup32-8GE 6148-GE-TX	DUT
Catalyst 6500	モジュラ 12.2(33)SXH	Sup32-8GE 6148-GE-TX	DUT
Catalyst 4500	12.2(40)SG	SupV 10GE	DUT
Catalyst 3750	12.2(40)SE	5 メンバ スタック	DUT
Catalyst 3560	12.2(40)SE	スタンドアロン	DUT
Catalyst 3550/3560	12.2(37)SE	60 台のスイッチ	コントロール プレーンの負荷

VSS 対応のキャンパス ベスト プラクティス ソリューション環境

表 4-2 ～表 4-4 は、本書で説明するキャンパス関連の VSS 実装のベスト プラクティスの要約を示します。

表 4-2　VSS 環境

キャンパス環境	検証されたキャンパス環境	説明
VSL リンク	スーパーバイザ ポートおよび WS-X6708 上に分散	
NSF 機能を設定	はい	
トポロジ	ECMP および MEC	
ルート数	3000	
CEF 負荷分散	はい	
デフォルト VSLP タイマー	はい	
仮想 MAC を使用	はい	
ポートチャネル負荷分散型	src-dst-ip 拡張	

表 4-3　レイヤ 3 ドメイン

キャンパス環境	検証されたキャンパス環境	説明
ルーティング プロトコル	EIGRP および OSPF	
コアにおける NSF 認識	はい	
EIGRP hello タイマーおよびホールド タイマー	デフォルト	5/15
OSPF hello タイマーおよびホールド タイマー	デフォルト	10/40
マルチキャスト ルーティング プロトコル	PIM-SPARSE	
ランデブー ポイント	コアでの ANYCAST IP	
マルチキャスト グループ数	80	

表 4-3　レイヤ 3 ドメイン （続き）

キャンパス環境	検証されたキャンパス環境	説明
マルチキャスト SPT しきい値	デフォルト	
トポロジ	ECMP および MEC	
ルート数	3000	
ルート 集約	はい	
CEF 負荷分散	はい	
コア接続	WS-X6708 10G	
コア デバイス	スタンドアロン 6500	

表 4-4　レイヤ 2 ドメイン

キャンパス環境	検証されたキャンパス環境	説明
STP	RPVST+	
ディストリビューション ブロックあたりのアクセス レイヤ スイッチ数	66	VSS あたり 66 MEC
合計 VLAN	207	
VLAN スパニング	8 つの VLAN	複数スイッチ
ディストリビューション ブロックあたりのネットワーク デバイス数	～ 4500	ホスト対 MAC ごとに一意
スパンされた VLAN の MAC アドレス	720 の MAC/VLAN	
各アクセス レイヤ スイッチに限られる VLAN	140	アクセス レイヤ スイッチあたりの音声とデータ
一意の IP アプリケーション プラウ	8000 ～ 11000	
EtherChannel テクノロジー	PAgP および LACP	
EtherChannel モード : PAgP	Desirable-Desirable	
EtherChannel モード : LACP	アクティブ-アクティブ	
PAgP および LACP タイマー	デフォルト	
トランキング モード	Desirable-Desirable	
トランキング タイプ	802.1Q	
VLAN の per-trunk 制限	はい	
UDLD モード	通常	
アクセス スイッチ接続	スーパーバイザ アップリンク ポートまたはギガビット アップリンク	

コンバージェンスおよびトラフィック リカバリ

ここでは、前半で VSS に関連する障害について説明し、後半で VSS 対応のキャンパスにおけるルーティングとコア コンポーネントに関係する障害について説明します。それぞれの種類の障害には、ユニキャスト トラフィックとマルチキャスト トラフィックの両方について障害復旧を表す表が含まれています。次の簡単な説明では、VSS に関連するトラフィック パターンと復旧方法の概要を示します。

- *ユニキャスト アップストリーム トラフィック*: アクセス レイヤから送信され、サーバファーム スイッチに向かうトラフィックを指します。
- *ユニキャスト ダウンストリーム トラフィック*: サーバファーム スイッチから送信され、アクセス レイヤ スイッチに向かうトラフィックを指します。
- *マルチキャスト トラフィック*: サーバファーム スイッチに接続された送信元と、アクセス レイヤから送信される受信者の参加を指します。一般に、ユニキャスト ダウンストリーム コンバージェンスに従います。
- *EC リカバリまたはフェールオーバー*: EtherChannel リンク障害と、残りのメンバ リンクへのトラフィックの再ハッシュを指します。
- *ECMP*：等コスト マルチパス（ECMP）は、ハードウェアで負荷分散型 CEF パスを提供する、完全にメッシュ化されたルーテッド インターフェイス トポロジを指します。
- *ローカル CEF :* VSS 専用の CEF スイッチング動作。ピア スイッチ パスよりもローカル CEF パスが優先されます。
- *マルチキャスト コントロール プレーン*: マルチキャスト コンポーネントに関連するコンバージェンスを指します。PIM 復旧、mroute の再設定（**,g* および *s,g*）、および Reverse Path Forwarding（RPF）、最短パス ツリーの構築などを含みますが、これに限りません。
- *アクティブまたはホットスタンバイ上での IIL および OIL*：RPF チェックで決定される Incoming Interface List（IIL; 着信インターフェイス リスト）の場所と、特定の VSS メンバ スイッチ上でマルチキャスト トラフィックをスイッチングするために使用される Outgoing Interface List（OIL; 発信インターフェイス リスト）の場所を指します。MEC ベースのトポロジでは、特定のマルチキャスト グループに対し、IIL と OIL は常に VSS ペアの同じメンバ上にあります。通常、ルーテッド インターフェイスのステータスが変化すると、マルチキャスト コントロール プレーンが変化します。
- *Stateful Switch Over（SSO）*: Stateful Switch Over（SSO; ステートフル スイッチオーバー）は、アクティブからホットスタンバイに復旧する方法を指します。
- *Multicast Multilayer Switching（MMLS）*: Multicast Multilayer Switching（MMLS; マルチキャスト マルチレイヤ スイッチング）は、（**,g* および s,g）エントリをホットスタンバイ スーパーバイザに複製する固有の方法を指します。スーパーバイザ障害時にマルチキャスト トラフィックをハードウェアで転送することも、リンク障害時にトラフィックのリダイレクトを起動することもできます。

VSS 固有のコンバージェンス

アクティブ スイッチ フェールオーバー

アクティブ フェールオーバーは、次のいずれかの動作によって開始されます。

- redundancy force-failover コマンドの適用
- アクティブ スーパーバイザのサービスからの物理的な取り外し
- アクティブ スーパーバイザの電源切断

アクティブ フェールオーバーの上記のどの方法に対しても、コンバージェンスは同じです。VSS スイッチ メンバ間でのスイッチオーバー処理（アクティブからホットスタンバイへ）は、これまでに説明した多数の概念と設計上の留意事項によって影響を受けます。次のイベント シーケンスは、フェールオーバー コンバージェンス処理の概要を示しています。

1. スイッチオーバーは、ソフトウェア CLI、スーパーバイザの取り外し、アクティブ スイッチの電源切断、システムの開始によって引き起こされます。
2. アクティブ スイッチは統一されたコントロール プレーンを明け渡し、ホットスタンバイが SSO コントロール プレーンを初期化します。
3. アクティブ スイッチに関連付けられているすべてのラインカードは、アクティブ シャーシのリブートに従って非アクティブ化されます。
4. その間に、新しいアクティブ スイッチ（以前のホットスタンバイ スイッチ）がルーティング プロトコルを再起動し、NSF 復旧処理を開始します。
5. 同時に、コアとアクセス レイヤが、トポロジに応じてトラフィック フローを再ハッシュします。ユニキャスト トラフィックは新しいアクティブ スイッチに向けられ、そこでハードウェア CEF テーブルを使用してトラフィックが転送されます。マルチキャスト トラフィックは、トポロジ設計に従い、マルチキャスト コントロール プレーンを再構築するか、MMLS ハードウェア テーブルを使用してトラフィックを転送します。
6. NSF 機能と SSO 機能は完全に初期化され、ネイバー デバイスからのルーティング情報の学習を開始し、必要に応じてフォワーディング テーブルおよびコントロール プレーン プロトコル テーブルを更新します。

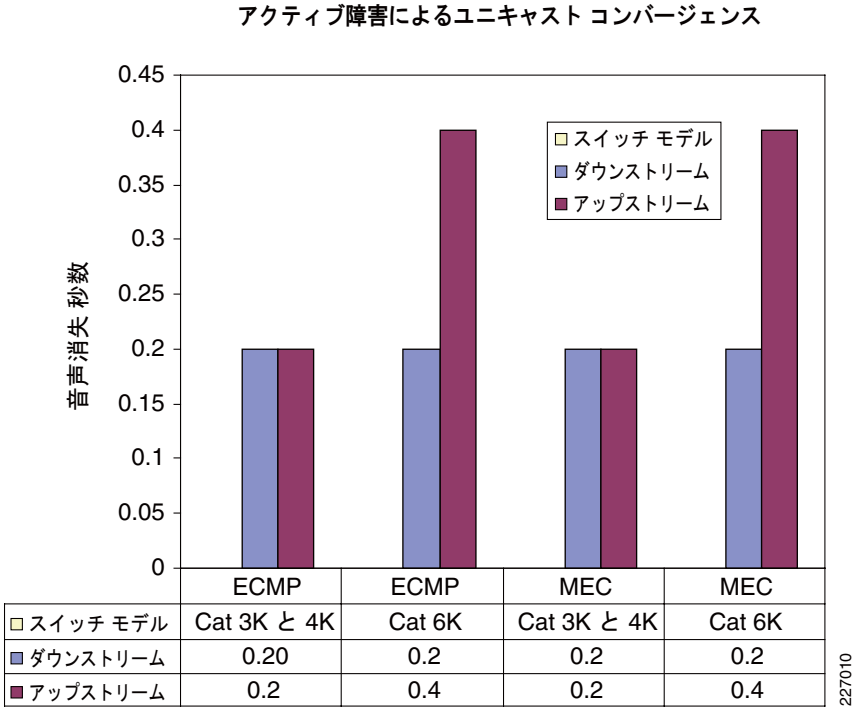
アクティブな障害コンバージェンスが、EIGRP および OSPF ルーティング プロトコルと、コアへの ECMP または MEC 接続のトポロジの組み合わせを使用して検証されます。表 4-5 に示すように、どちらのルーティング プロトコルの復旧方法も同じです。

表 4-5　アクティブ障害復旧

トポロジ	ECMP	MEC	共通の復旧
ユニキャスト復旧方式			
ユニキャスト アップストリーム	アクセスでの EC フェールオーバー	アクセスでの EC フェールオーバー	SSO
ユニキャスト ダウンストリーム	CEF	コアでの EC フェールオーバー	SSO
マルチキャスト復旧方式			
アクティブ スイッチ上の IIL	マルチキャスト コントロール プレーン	コアでの EC フェールオーバー	MMLS
ホットスタンバイ スイッチ上の IIL	MMLS	コアでの EC フェールオーバー	MMLS

コンバージェンス損失は、EIGRP と OSPF で同じです。図 4-2 は、平均コンバージェンスが、Cisco Catalyst 3xxx または Cisco Catalyst 45xx スイッチング プラットフォームで 200 ミリ秒以下、Catalyst 65xx スイッチング プラットフォームで約 400 ミリ秒であることを示しています。Cisco Catalyst 6500 で若干損失が大きい理由の 1 つは、分散したファブリックベースのアーキテクチャでは、フローを使用可能なメンバ リンクに再ルーティングする前に、依存関係を考慮する必要があるためです。

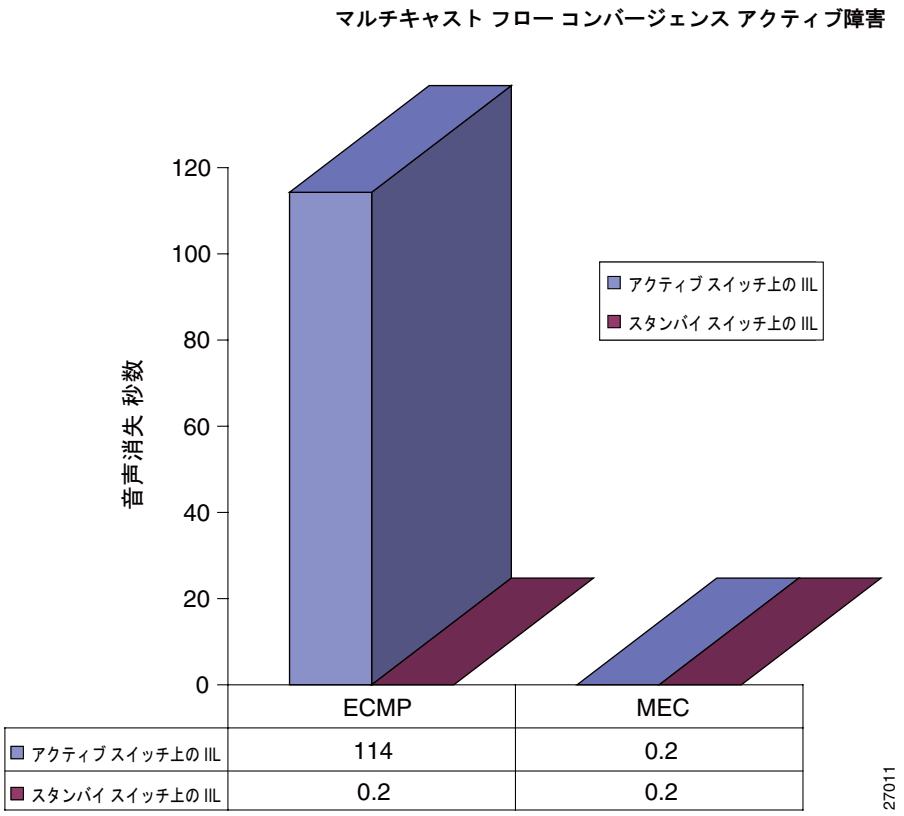
図 4-2　アクティブ障害コンバージェンス



	ECMP	ECMP	MEC	MEC
スイッチ モデル	Cat 3K と 4K	Cat 6K	Cat 3K と 4K	Cat 6K
ダウンストリーム	0.20	0.2	0.2	0.2
アップストリーム	0.2	0.4	0.2	0.4

図 4-3 に示すマルチキャスト コンバージェンスは、トポロジと、着信インターフェイス リスト（IIL）の構築場所に依存します。この設計上の選択については、「VSS でのレイヤ 3 マルチキャスト トラフィック デザインにおける考慮事項」（P.3-61）を参照してください。IIL リストの場所とスイッチ障害の組み合わせによっては、ECMP でのマルチキャスト コンバージェンスが大きくなることに注意してください。

図 4-3　マルチキャスト コンバージェンス



ホットスタンバイ フェールオーバー

ホットスタンバイ フェールオーバーでは、コントロール プレーンのコンバージェンスが起こりません。これは、さまざまなプロトコルの管理とその更新について自発的に責任を負わないためです。しかし、ECMP トポロジでは、ホットスタンバイ スイッチで接続されたネイバーはリセットされ、ホットスタンバイに接続されているリンクはダウンします。トラフィックの復旧は、SSO 初期化遅延が存在しないことを除き、アクティブの障害と同じです。表 4-6 を参照してください。

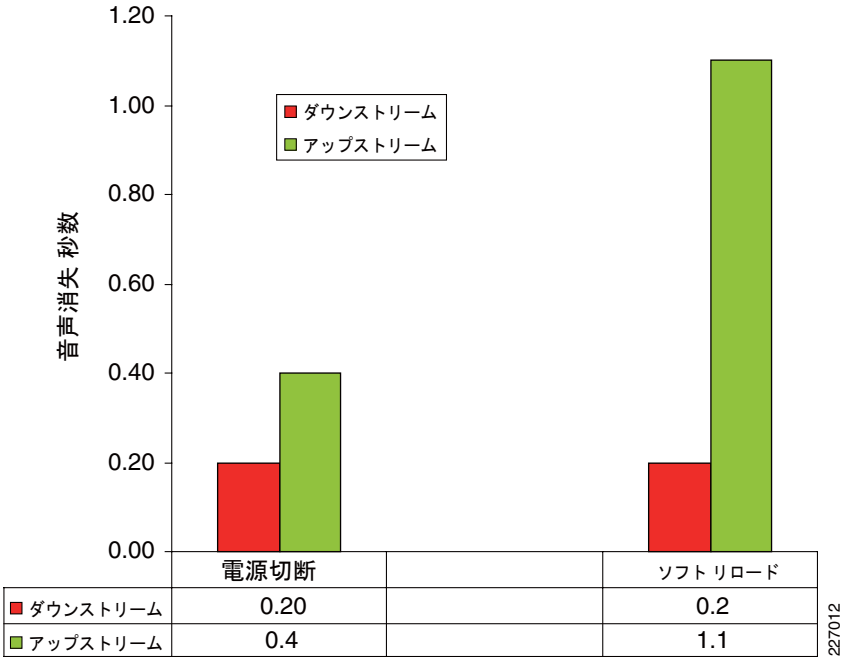
表 4-6　ホットスタンバイ障害復旧

トポロジ	ECMP	MEC	共通の復旧
ユニキャスト復旧方式			
ユニキャスト アップストリーム	アクセスでの EC フェールオーバー	アクセスでの EC フェールオーバー	
ユニキャスト ダウンストリーム	コアでの ECMP フェールオーバー（CEF）	コアでの EC フェールオーバー	
マルチキャスト復旧方式			
アクティブでの IIL	影響なし	影響なし	MMLS
ホットスタンバイでの IIL	マルチキャスト コントロール プレーン	EC	MMLS

図 4-4 に示すように、停電時のコンバージェンスは 1 秒未満であるのに対し、ソフトウェア障害ではパケット損失が若干多くなります。

図 4-4　ホットスタンバイ コンバージェンス特性の比較

ソフトウェア リロードによるホットスタンバイ障害 対 電源切断



	電源切断		ソフト リロード
ダウンストリーム	0.20		0.2
アップストリーム	0.4		1.1

図 4-4 に示すアップストリーム コンバージェンスは、ネットワーク接続の設定方法に固有です。検証済み設計では、コアを接続するアップリンクは DFC WS-X6708 ラインカード上にあります。アップストリーム コンバージェンスは可変であり、特定の接続が設定されているシャーシ内のラインカードの位置に依存します。表 4-6（P.4-8）では、ホットスタンバイ ソフトウェア リロードに関連する間欠的な損失または復旧を扱っていません。Cisco IOS ソフトウェアは、ホットスタンバイのソフトウェア リロード中に、スーパーバイザ カードを削除した後、ラインカードをスロットの降順に削除します（スロット番号が小さいものが最初に削除されます）。この動作を次の syslog の出力に示します。あるシナリオで、コアへの 10 ギガビット接続が存在するスロット 2 がオフラインになります。その間、アクセス レイヤ接続（スロット 7、8、および 9）はまだアップ状態のままであるため、アクセス レイヤ スイッチは、アップストリーム トラフィックをホットスタンバイに送信し続けます。コアへの直接のアップストリーム接続が存在しないため、このトラフィックは VSL に再ルーティングされます。これは、アップストリーム トラフィックに関連する損失が増える一因になります。コアに接続されているラインカードを最後のスロットに移動すると、アクセス ラインカードの電源が最初に切断されるため、損失は逆になります。これにより、アクセス レイヤ スイッチは、EtherChannel 上の残りのリンクにトラフィックを再ルーティングします。しかし、まだ VSS で受信されているダウンストリーム トラフィックは、ラインカードが削除されるまで VSL リンクに再ルーティングされます。そのため、この場合、ダウンストリーム損失が大きくなります。

```
Nov 14 08:43:03.519: SW2_SP:  Remote Switch 1 Physical Slot 5 - Module Type LINE_CARD
removed
Nov 14 08:43:03.667: SW2_SP:  Remote Switch 1 Physical Slot 2 - Module Type LINE_CARD
removed
Nov 14 08:43:04.427: SW2_SP:  Remote Switch 1 Physical Slot 7 - Module Type LINE_CARD
removed
Nov 14 08:43:04.946: SW2_SP:  Remote Switch 1 Physical Slot 8 - Module Type LINE_CARD
removed
```

```
Nov 14 08:43:05.722: SW2_SP:  Remote Switch 1 Physical Slot 9 - Module Type LINE_CARD
removed
Nov 14 08:47:09.085: SW2_SP:  Remote Switch 1 Physical Slot 5 - Module Type LINE_CARD
inserted
Nov 14 08:48:05.118: SW2_SP:  Remote Switch 1 Physical Slot 2 - Module Type LINE_CARD
inserted
Nov 14 08:48:05.206: SW2_SP:  Remote Switch 1 Physical Slot 7 - Module Type LINE_CARD
inserted
Nov 14 08:48:05.238: SW2_SP:  Remote Switch 1 Physical Slot 8 - Module Type LINE_CARD
inserted
Nov 14 08:48:05.238: SW2_SP:  Remote Switch 1 Physical Slot 9 - Module Type LINE_CARD
inserted
```

ホットスタンバイ復旧

トラフィック復旧は次の 2 つの要因に依存します。

- *スロットの順序*：スロットの順序が問題になるのは、ラインカードの電源がシーケンシャルに投入されるためです。
- *カードの種類*：ラインカードの種類も転送状態に影響を与えます。DFC ラインカードでは、ブートに長い時間がかかります。

コア接続が最初に復旧されると、ダウンストリーム トラフィックに複数の復旧があります。最初の復旧はコア レイヤの CEF（ECMP）ベースまたは EtherChannel ベースの復旧です。2 つ目の復旧は、トラフィックが VSS に到達したときに起こります。アクセス レイヤに接続されたラインカードはまだオンラインになっていないため、VSS は VSL リンク上で再ルーティングする必要があります。同様に、アクセス レイヤ ラインカードが最初にアップ状態になる場合、アップストリーム トラフィックには複数の復旧があります。

ECMP のマルチキャスト復旧には初期の影響はありません。これは、着信インターフェイス（アクティブ スイッチ上で構築されている場合）が変化しないためです。しかし、新しい PIM Join を新たに追加されたルート経由で送信でき（ホットスタンバイ ECMP リンクが RPF チェックの起動を復旧するため）、これによりマルチキャスト コントロール プレーンのコンバージェンスが誘発されます。MEC ベースのトポロジでは、ホットスタンバイに接続されたリンクがレイヤ 3 MEC に追加されるときに、コアでの EtherChannel ハッシュ結果に基づいて復旧されます。その後、アクセス レイヤ ラインカードのブート ステータスに基づいて VSL でトラフィックを再ルーティングできます。表 4-7 を参照してください。

表 4-7　ホットスタンバイ スイッチ復旧

トポロジ	ECMP	MEC	共通の復旧
ユニキャスト復旧方式			
ユニキャスト アップストリーム	可変	可変	上記説明を参照
ユニキャスト ダウンストリーム	可変	可変	上記説明を参照
マルチキャスト復旧方式			
アクティブでの IIL	可変、マルチキャスト コントロール プレーン	可変：EC ハッシュ ラインカードのブート状態	
ホットスタンバイでの IIL	非該当	非該当	スタンバイ復旧

上で説明した要因により、VSS ベースの環境でコンバージェンスが可変になる可能性があります。一般に、復旧損失は、700 ミリ秒 ～ 4 秒の範囲です。次の URL にあるドキュメントで説明されている ARP スロットリング動作により、これらの損失は、スタンドアロンノードの復旧よりもはるかに少なくなります。

http://www.cisco.com/en/US/docs/solutions/Enterprise/Campus/HA_recovery_DG/campusRecovery.html

VSL リンク メンバ障害

VSL バンドルは、ポートチャネル インターフェイスです。その障害コンバージェンスと復旧特性は、MEC リンク障害特性と同様です。VSL リンク障害が発生すると、残りのリンク上でトラフィックの再ハッシュが行われます（ユーザ データとコントロール リンクの両方）。レイヤ 3 およびレイヤ 2 MEC トポロジは、対称型のローカル フォワーディングを提供し、リンクが障害になっても VSL リンク上のユーザ データの再ルーティングは影響を受けません。しかし、アクセス デバイスが 1 つだけのメンバに接続されているトポロジや、アップリンク フォーム アクセス レイヤ スイッチの 1 つが障害になったトポロジでは、VSS はダウンストリーム トラフィックの半分を再ルーティングして VSL リンクを経由させます。シングルホーム接続のトポロジにおけるリンクの障害は、ユーザ データのコンバージェンスに影響を与えます。データ トラフィックのコンバージェンスは、1 秒未満から数秒に及ぶことがあります。非 MEC ベースの設計によって構成される次善のトポロジは、次善のコンバージェンスになります。VSS に接続されるすべてのデバイスに対してデュアルホーム接続の MEC 設計を実装してください。

VSS でのラインカード障害

ラインカードの障害は、次の理由で発生する可能性があります。

- *ハードウェア障害*：ハードウェアの交換が必要であり、一般には計画されたイベントです。
- *ソフトウェア障害*：ラインカードをリセットすることでこの問題が解決されることがあります。

「VSS 障害ドメインおよびトラフィック フロー」（P.3-9）に示すように、ラインカードが障害になると、基本的に VSS へのシングルホーム リンクまたは孤立した接続リンクができます。この障害では、ラインカードがアクセス レイヤとコア レイヤのどちらに接続されているかによって、アップストリーム トラフィックまたはダウンストリーム トラフィックが再ルーティングされます。

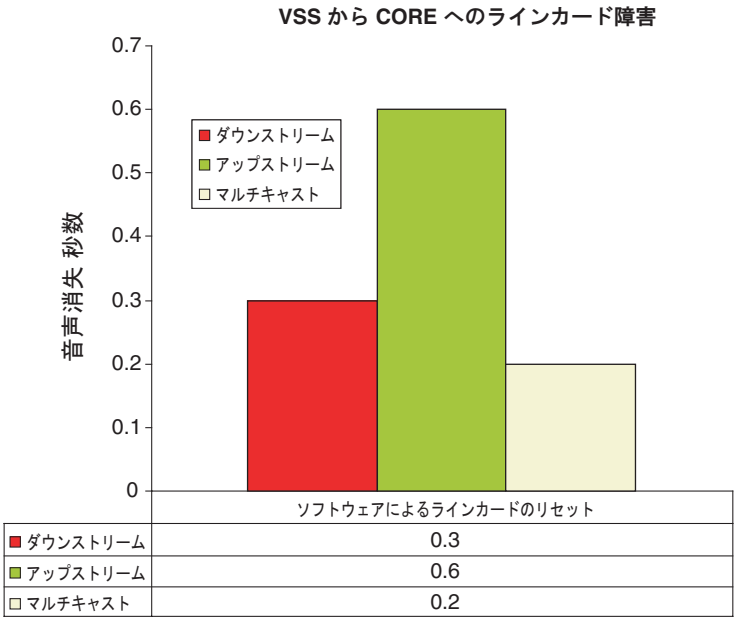
コア レイヤに接続されたラインカード

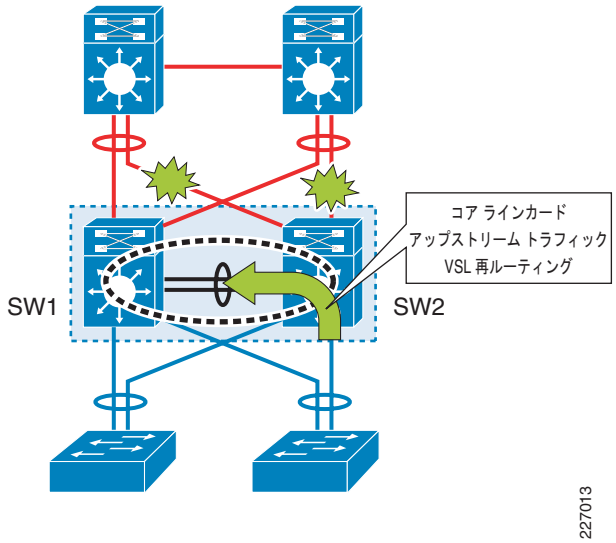
シングル ポイント障害を回避するには、コアへの接続で複数のラインカードを使用する必要があります。検証は、最悪の損失シナリオを示すために、VSS をコアに接続する単一のラインカードに適用されます。ある VSS メンバ スイッチからの接続全体がダウンするときのコンバージェンスとトラフィック フローについては、図 4-5 を参照してください。表 4-8 は、コア レイヤの接続障害と復旧の一覧を示します。

表 4-8　コア接続（ラインカード障害）の復旧

トポロジ	MEC	追加の復旧
ユニキャスト アップストリーム	VSL 再ルーティング	
ユニキャスト ダウンストリーム	コアでの EtherChannel フェールオーバー	
障害になったラインカードでのマルチキャスト ハッシュ	EtherChannel フェールオーバー	MMLS

図 4-5　VSS-コア間での単一ラインカード障害と復旧コンバージェンス





ダウンストリームに流れるマルチキャスト トラフィックの場合、コア デバイスのハッシュ処理により、そのフローのフォワーダになる VSS メンバが選択されます。図 4-5 で、SW1 に接続される残りのリンク上で再ハッシュします。トラフィックは、(s,g) ペアをピア スイッチ（図 4-5 の SW1）に同期した MMLS テクノロジーを使用して、SW1 によりハードウェアで転送されます。

注意　マルチキャスト データ コンバージェンスは、(s,g) ペア（mroute）の数と、その他いくつかのマルチキャスト コントロール プレーン機能に大きく依存します。mroute 数が多い場合、コンバージェンスにさらなる検証が必要になることがあります。

アクセス レイヤに接続されたラインカード

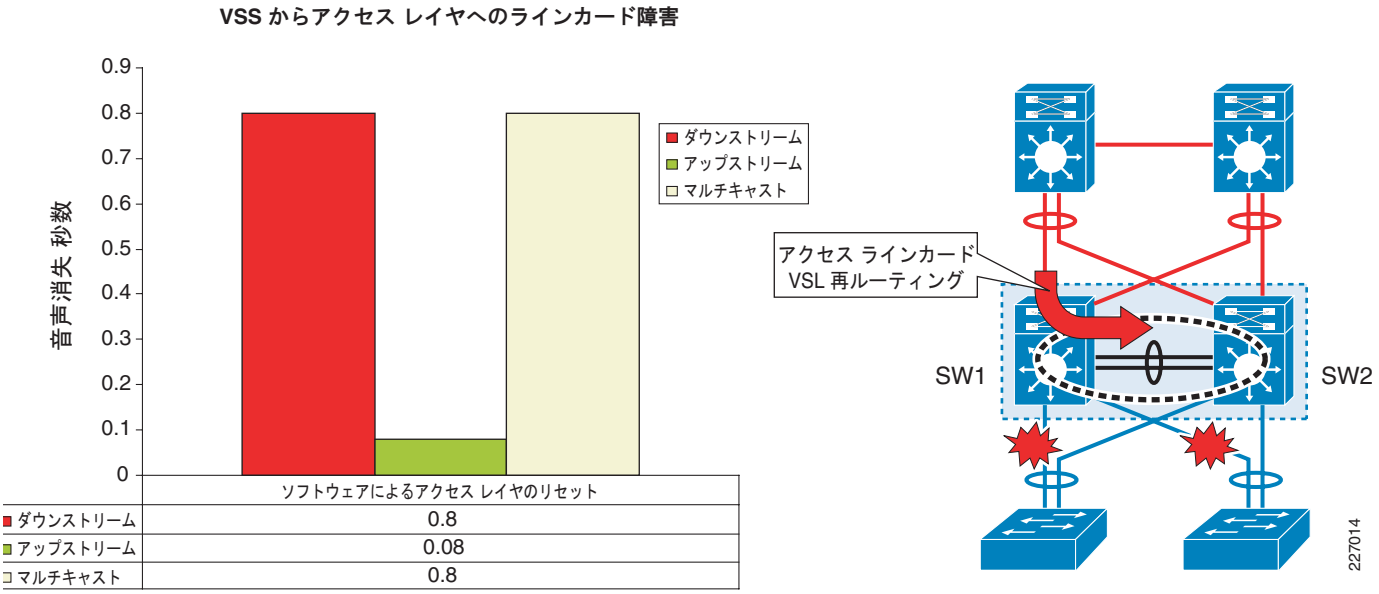
アクセス レイヤのラインカード障害では、トラフィック フローの復旧はコア ラインカード障害の逆になります。表 4-9 を参照してください。

表 4-9　アクセス レイヤに接続されたラインカードの障害復旧の要約

トポロジ	MEC	追加の復旧
ユニキャスト アップストリーム	アクセスでの EtherChannel フェールオーバー	
ユニキャスト ダウンストリーム	VSL 再ルーティング	
障害になったラインカードでのマルチキャスト ハッシュ	VSL 再ルーティング	MMLS

図 4-6 に、アクセス レイヤに接続されたラインカードの障害と復旧の要約を示します。

図 4-6　アクセス レイヤに接続されたラインカードの障害と復旧の要約



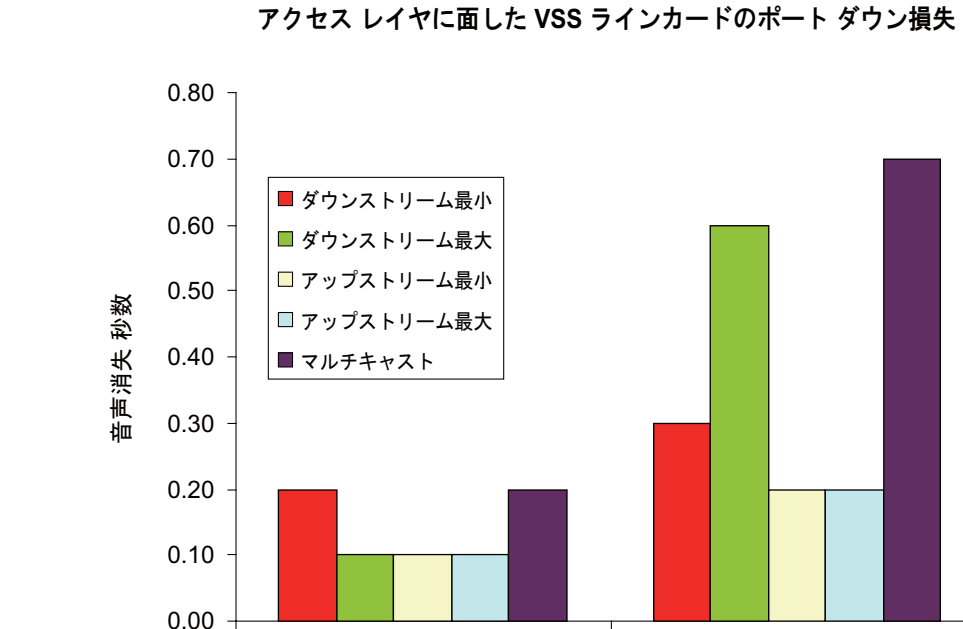
ダウンストリームに流れるマルチキャスト トラフィックの場合、コア デバイスのハッシュ処理により、そのフローのフォワーダになる VSS メンバが選択されます。アクセス レイヤ ラインカード障害では、マルチキャスト トラフィックを VSL リンク上で再ルーティングする必要があります。ピア スイッチは、アクセス レイヤ スイッチへの既存のレイヤ 2 MEC 接続経由でトラフィックを転送します。この転送で、ピア スイッチは、(s,g) ペアをピア スイッチ（図 4-6 の SW2）に同期した MMLS を使用します。

ポート障害

アクセス レイヤに接続されたポートの障害は、トラフィック フローの復旧に関して、アクセス レイヤ ラインカードの障害と同様です。ポート ステータスの変化を取り込む方法と場所によって、コンバージェンスが影響を受けます。

図 4-7 に示すコンバージェンスは、CLI によって引き起こされたポートのシャットダウンが、ファイバ接続を物理的に取り外すよりも良いコンバージェンスになることを示しています。キャリア遅延およびポート ステータスを検出するためのポートのソフトウェア ポーリングにより、ファイバ リンクを物理的に取り外した場合のコンバージェンス遅延が増えます。

図 4-7　ポート障害の復旧の比較



	ソフトウェア シャットダウン	ファイバの引き抜き
ダウンストリーム最小	0.20	0.3
ダウンストリーム最大	0.1	0.6
アップストリーム最小	0.1	0.2
アップストリーム最大	0.1	0.2
マルチキャスト	0.2	0.7

アクセス レイヤのアップリンク ポートでポートの **shutdown/no shutdown** シーケンスが発生すると、数秒間にわたってパケットが損失する恐れがあります。Cisco Catalyst 6500 システムのポート ステータス検出方法が、そのような遅延の根本原因であると考えられます。この動作は、スタンドアロン シナリオと、VSS ベースのシステムで共通です。将来の Cisco IOS リリースでは、ポート ステータス検出の最適化により、関連するパケット損失とコンバージェンス遅延が低減される可能性があります。操作上、ポート ステータスの変更を、アクセス レイヤではなく VSS で行うことをお勧めします。

ルーティング（VSS からコアへ）コンバージェンス

レイヤ 3 ドメインで VSS を使用する設計上の選択については、「VSS を使ったルーティング」（P.3-46）を参照してください。このセクションでは、レイヤ 3 MEC トポロジが、ルーティング エンティティとの VSS 相互接続を構築するための最も有効な方法であることが示されています。ここでは、この設計上の選択がさらに具体化されています。そのため、ECMP ベースのコンバージェンスについては、本書では説明しません。また、ここでは、コア デバイスが障害になったときの VSS トラフィック フローとコンバージェンスに対する影響についても詳しく説明します。

EIGRP および OSPF と MEC を使用したコア ルータ障害

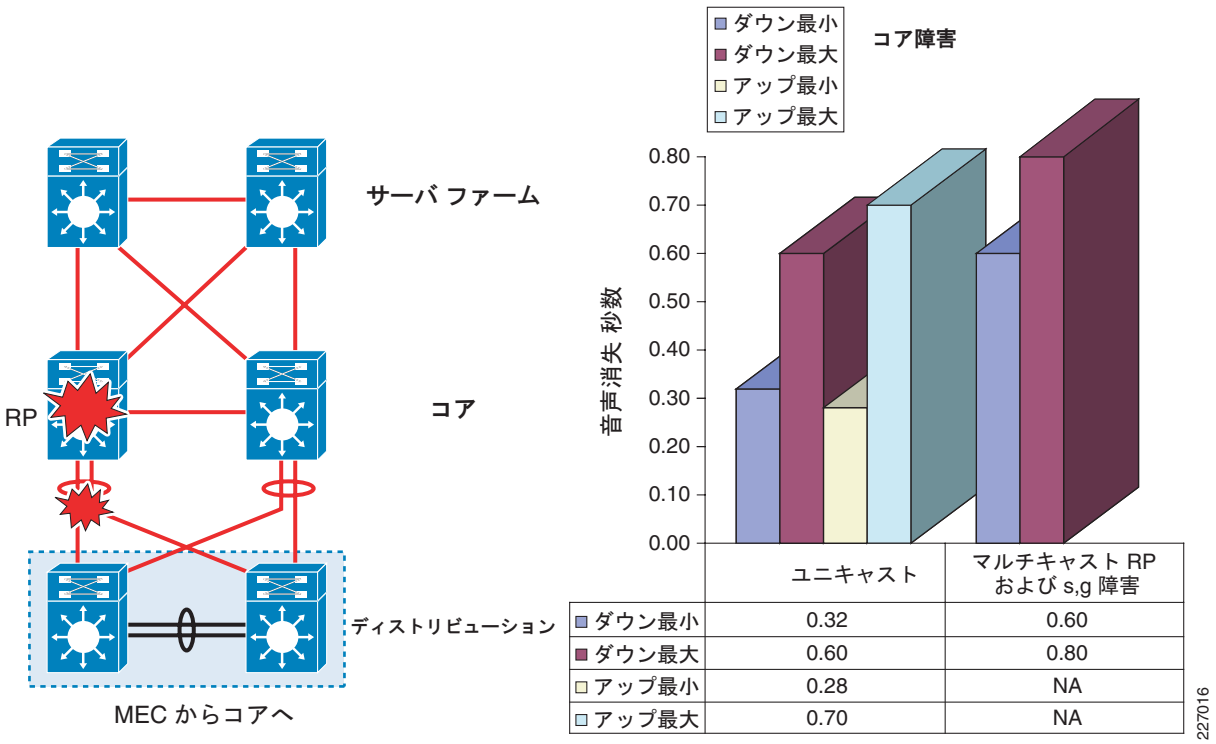
この設計ガイドは、スタンドアロン コア デバイスで検証されています。この設計ガイドでは、コアでのマルチキャスト トラフィックに対する Rendezvous Point（RP; ランデブー ポイント）の提供がイネーブルになっています。RP の配置に関する設計上の選択は、マルチキャスト アプリケーションに依存するため、ここでは扱いません。ただし、RP の障害は、コア障害における障害になったコンポーネントの説明の一部として記載されています。表 4-10 を参照してください。

表 4-10　コア ルータ障害復旧の要約

トポロジ	MEC	説明
ユニキャスト アップストリーム	VSS での ECMP	コアの Cisco Catalyst 6500 スタンドアロン
ユニキャスト ダウンストリーム	コア サーバでの ECMP	
アクティブでの IIL または障害になったコアでの OIL	マルチキャスト コントロール プレーン	
ホットスタンバイでの IIL	非該当	

OSPF および EIGRP の場合、ユニキャスト トラフィックに対するコア障害コンバージェンスは同じです。コアと VSS はデフォルトの hello タイマーおよびホールド タイマーが設定されています。LSA と SPF の OSPF タイマーはデフォルト値に設定されています。アップストリームおよびダウンストリーム コンバージェンスは ECMP に基づいています。どちらの VSS メンバにもフォワーディングで使用可能なローカル ECMP パスがあり、トラフィックは VSL リンクを経由しません。マルチキャスト トラフィックの場合、コア障害により着信インターフェイスが変化します。VSS の着信インターフェイスも変更されるように、マルチキャスト トポロジは、コンバージェンスを経て、代替ポートを通じて新しい着信インターフェイスを見つける必要があります。マルチキャスト コンバージェンスは、コアの先にあるトポロジと送信元の場所に応じて、大幅に変わります。図 4-8 に、80 個のマルチキャスト フローに対するこのコンバージェンスを示します。mroute が多い場合（多いフロー）、コンバージェンスが変わることがあります。

図 4-8　コア ルータ障害に対する復旧の比較



リンク障害時のコンバージェンス

リンク障害時の動作については、「ECMP および MEC トポロジを使用したデザインにおける考慮事項」（P.3-48）を参照してください。ベスト プラクティスに基づくコンフィギュレーションは、その説明から導き出されています。そこでは、MEC トポロジの使用が強調されています。このセクションでは、MEC トポロジ オプションだけを扱います。

OSPF を使用した MEC リンク メンバ障害

ルーティング プロトコルとメトリック変更の依存関係については、「リンク障害中のフォワーディング キャパシティ（パスのアベイラビリティ）」（P.3-49）を参照してください。MEC ベースのトポロジでは、ルーティング プロトコルと関連するコンフィギュレーションによっては、リンク障害により使用可能なフォワーディング容量が減少する可能性があります。トラフィック フロー復旧と関連する属性の要約を表 4-11 に示します。

表 4-11　MEC リンク メンバ障害の OSPF 復旧の要約

トポロジ	OSPF（auto-cost reference bandwidth 20G）	OSPF（auto-cost reference bandwidth なし）
メトリック変更	はい	なし
結果の帯域幅	2 つのパス	3 つのパス
ユニキャスト アップストリーム	安全なルートの取り消し	ローカル ハードウェア CEF パス

表 4-11　MEC リンク メンバ障害の OSPF 復旧の要約 （続き）

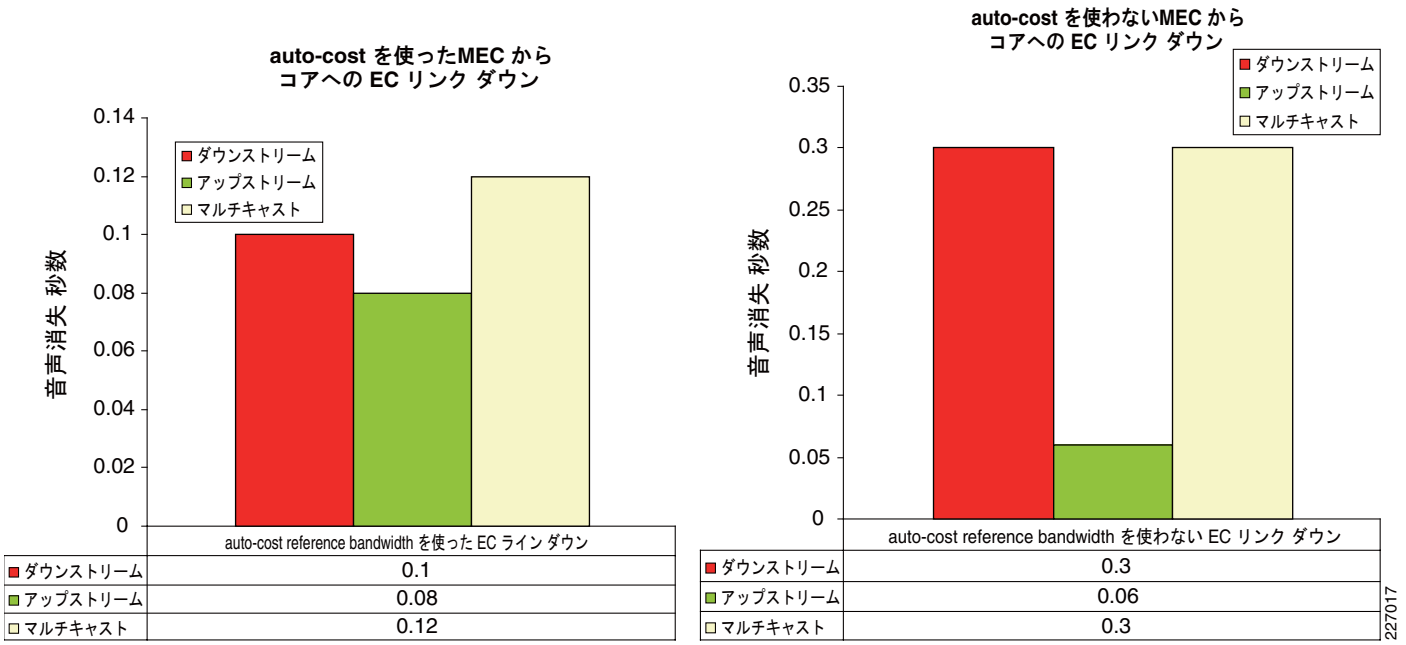
トポロジ	OSPF（auto-cost reference bandwidth 20G）	OSPF（auto-cost reference bandwidth なし）
ユニキャスト ダウンストリーム	安全なルートの取り消し	コアでの EC 復旧
非集約ネット対 集約ネット	なし	なし
マルチキャスト	マルチキャスト コントロール プレーン：ルートの取り消しによりコアで発信インターフェイス リストが変化	いくらかの影響があるか影響がない場合、マルチキャスト フローの再ハッシュ

検証トポロジには、VSS で 2 つのレイヤ 3 MEC を含みます。各コア ルータには、VSS の 2 つのメンバに接続されたメンバ リンクを持つ、単一のポートチャネルが設定されます。結果的なルーティング トポロジは、2 つの ECMP パス（各コア ルータから 1 つずつ）で構成されます。auto-cost 基準が設定されている OSPF では、リンク障害により、ルーテッド ポートチャネル インターフェイスの 1 つでメトリックが変化します。メトリックの変化の影響により、2 つの等コスト パスのいずれかから学習したルートが取り消されます。これにより、各 VSS メンバからのルーテッド リンクが、ルーティング テーブル中で 1 つだけ使用可能になります。アップストリームとダウンストリームの復旧は、安全なルートの取り消しによって変わります。ユーザ データ トラフィックへの影響はきわめて小さくなります。これは、ルートが取り消されるまでは VSS 上でまだ使用可能なリンクにトラフィックが転送され続けるためと、WS-X6708 ラインカードは FLN をサポートしているためです（リンク ステータス変更の通知はハードウェアで行われます）。「リンク障害中のフォワーディング キャパシティ（パスのアベイラビリティ）」（P.3-49）の該当する CEF 出力を参照してください。

auto-cost reference bandwidth がない OSPF では、リンク障害によりルーティング情報が変更されません。これは、リンク障害が発生しても、20 ギガビット集約 EtherChannel 帯域幅のメトリック変更が発生しないためです。ポートチャネル帯域幅が 10 ギガビットに変化しても、デフォルトの auto-cost が 100 MB であるため、コストは 1 のままになります。アップストリーム トラフィックの復旧は、VSS での CEF における単純な隣接関係の更新に基づき、ルーテッド インターフェイス全体がディセーブルになったときに引き起こされる ECMP 復旧（CEF ネクストホップ アップデート）には基づきません。このトポロジにおけるダウンストリームの影響は、コアでの EtherChannel の復旧によって変わります。「リンク障害中のフォワーディング キャパシティ（パスのアベイラビリティ）」（P.3-49）の該当する CEF 出力を参照してください。

図 4-9 に示す、auto-cost がある場合とない場合の復旧パフォーマンスの比較を参照してください。

図 4-9　auto-cost と非 auto-cost の復旧の比較



ルートの取り消しでは、ルートが同じ単一の論理ノードから学習されるため、メトリックが変化してもトポロジは変化しません。

OSPF とレイヤ 3 MEC トポロジを使用した場合、パケット損失が最小限になるため、唯一の設計上の選択は、障害時に使用可能な合計帯域幅であり、ユーザ データ コンバージェンスに対する影響ではありません。

EIGRP を使用した MEC リンク メンバ障害

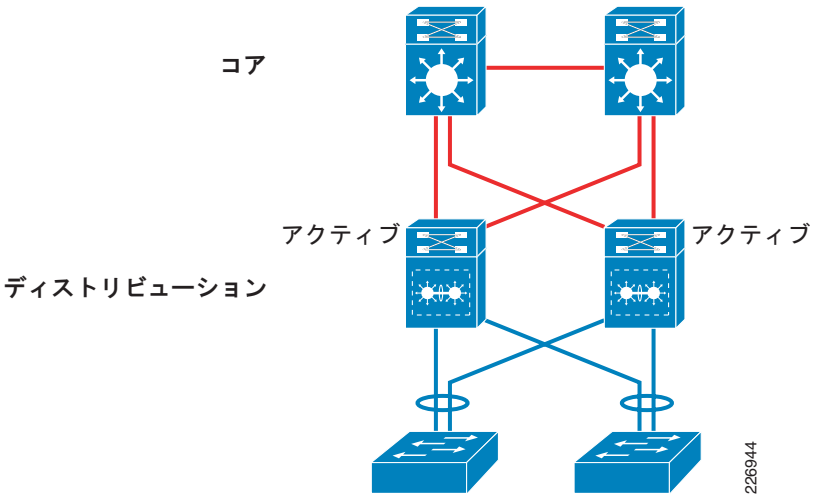
EIGRP メトリック計算は、遅延の合計と最小帯域幅をあわせたものです。メンバ リンクが障害になると、EIGRP は変更された帯域幅の値を認識して使用しますが、遅延は変化しません。これは、合成メトリックに影響を与えることもあれば、影響を与えないこともあります。というのは、パスにおける最低帯域幅がメトリック計算で使用されるため、ローカルな帯域幅の変化は、それがパス中の最低の帯域幅の場合にだけ影響するためです（合計遅延は変化していません）。キャンパス ネットワークでは、コアと VSS の間で提供される帯域幅の変化はギガビット単位です。一般に、これは、ほとんどのルートで最低帯域幅ではありません。したがって、実用的には、EIGRP は帯域幅変更の影響を受けず、デフォルトの auto-cost reference bandwidth を使った OSPF と同じ動作になります。合成メトリックが影響を受ける状況がある場合、EIGRP は、auto-cost reference bandwidth が設定されている OSPF と同じ動作になります。

VSS デュアルアクティブ スーパーバイザを使用したキャンパス復旧

デュアルアクティブ状態

上記のセクションでは、VSL バンドルが、コントロール プレーンとユーザ データ トラフィックを伝送できるシステム リンクとして機能する方法について説明しました。コントロール プレーン トラフィックは、2 つの VSS シャーシ間でステート マシンの同期を保ちます。VSL リンクの分断または通信障害が発生すると、VSS がきわめて不安定になります。「VSS での SSO 動作」（P.2-25）で説明したように、ホットスタンバイの役割を担うスイッチ メンバは、アクティブ スイッチと常に通信を保ちます。ホットスタンバイ スイッチの役割は、VSL リンクを通じてピアとの通信損失を検出したときに、すぐにアクティブの役割を引き継ぐことです。この役割の移行は、スイッチオーバーによって引き起こされるか（ユーザ起動）、アクティブ スイッチが何らかの障害になった場合に正常な動作です。しかし、障害状態の間、リモート スイッチがリブートしたのか、アクティブ スイッチとホットスタンバイ スイッチの間のリンクが機能しなくなったのかを区別する方法はありません。どちらの場合も、ホットスタンバイ スイッチはすぐにアクティブ スイッチの役割を引き継ぎます。これにより、*デュアルアクティブ*状態と呼ばれる状態になることがあります。この状態では、両方のスイッチのスーパーバイザが、コントロール プレーンに責任を持つと考え、アクティブ スーパーバイザとしてネットワークとの対話を開始します。図 4-10 に、デュアルアクティブ状態のキャンパス トポロジを示します。

図 4-10　　デュアルアクティブ トポロジ



ネットワークがデュアルアクティブ状態になるのを回避するための最善の方法は、次のベスト プラクティスを適用することです。

- 冗長なポート、ラインカード、内部システム リソースを使用して、VSL 接続を分散させる。推奨されるコンフィギュレーション オプションを「レジリエント VSL デザインの考慮事項」（P.2-19）に示します。
- 各 VSL リンクに対して多様な光ファイバ パスを使用する。単一コンジット障害の場合、デュアルアクティブ状態は発生しません。
- 通常状態と異常状態のキャパシティ プランニングを使用して、VSL リンク上で転送されるトラフィックを管理する。VSL 上でトラフィックを管理するための設計ガイダンスについては、第 3 章「VSS 対応キャンパス デザイン」を参照してください。

ベスト プラクティスに基づいて設計を実装することで、デュアルアクティブ状態になる可能性を大幅に減らすことができますが、問題がなくなるわけではありません。デュアルアクティブの問題を引き起こす一般的な原因としては、次のものがあります。

- 短い LMP タイマーと不適切な VSL ポートチャネル コンフィギュレーション。
- ユーザによって引き起こされる VSL ポートチャネルの偶発的なシャットダウン。
- CPU 使用率が高いと、VSLP hello ホールド タイマーがタイムアウトし、その結果すべての VSL リンクが VSL EtherChannel から削除される可能性がある。
- システム ウォッチドッグ タイマー障害の影響は、CPU 使用率が高い場合と同等。これらにより、VSL EtherChannel が動作しなくなる可能性があります。
- ポートチャネル インターフェイスをディセーブルにしてしまうソフトウェアの異常。
- 初期設定時または変更時に、両方のスイッチで同じスイッチ ID が偶発的に設定された場合。

デュアルアクティブ状態を回避することはきわめて重要ですが、そのような状態を検出し、復旧のための手順をすばやく実行することも重要です。以降では、次の点について説明します。

- 検出方法が存在しないネットワークに対するデュアルアクティブ状態の影響
- 使用可能な検出方法、その動作、復旧
- 特定の設計に対して予想されるコンバージェンスと適用されるベスト プラクティス オプション

検出方法が存在しないネットワークに対するデュアルアクティブの影響

デュアルアクティブ状態が発生すると、各メンバ シャーシがアクティブの役割を担います。これは、各メンバが、同じ IP アドレスと MAC アドレスを使用して、スタンドアロン デバイスとして動作することを意味します。ネットワーク コントロール プレーンの重複は、ルータ ID、STP ルート ブリッジ、ルーティング プロトコルのネイバー ルータとの隣接関係などにも影響を与えます。実稼動ネットワークにおけるデュアルアクティブ状態の影響は、2 つに分けられます。

- コントロール プレーンの分断
- ユーザ データ トラフィックの中断

特定のネットワークでの実際の動作は、トポロジ（MEC または ECMP）、展開されているルーティング プロトコル（OSPF または EIGRP）、使用している相互接続の種類（レイヤ 2 またはレイヤ 3）によって変わります。ここでは、検出方法を展開することの重要性について説明します。重要なコンポーネントとトポロジの留意事項だけを扱います。

レイヤ 2 MEC に対する影響

デュアルアクティブ状態になると、同じ STP ドメインに対し、2 つのアクティブなルートができます。両方のアクティブ シャーシが、各ラインカードから、異なる送信元 MAC アドレスを使用して、個別の STP BPDU を生成します。レイヤ 2 MEC が設定されたアクセス レイヤ スイッチは、STP ツリーの送信元であることを主張する複数の MAC アドレスを検出します。これは、PAgP により EtherChannel の不整合として検出され、最終的にアクセス レイヤ ポートチャネル インターフェイスが error-disable 状態になります。これにより、syslog メッセージが生成されます。アクセス レイヤ スイッチで出力されるメッセージは、次のソフトウェア バージョンに依存します。

Cisco IOS を使用する Cisco Catalyst 65xx、Cisco Catalyst 45xx、および Cisco Catalyst 35xx

```
%PM-SPSTBY-4-ERR_DISABLE: channel-misconfig error detected on Gi5/1, putting Gi5/1 in
err-disable state
%PM-SPSTBY-4-ERR_DISABLE: channel-misconfig error detected on Gi5/2, putting Gi5/2 in
err-disable stat
```

CATOS を使用する Cisco Catalyst 65xx

```
%SPANTREE-2-CHNMISCFG: STP loop - channel 5/1-2 is disabled in vlan/instance 7
%SPANTREE-2-CHNMISCFG2: BPDU source mac addresses: 00-14-a9-22-59-9c, 00-14-a9-2f-14-e4
ETHC-5-
PORTFROMSTP: Port 5/1 left bridge port 5/1-2
```

EtherChannel の不整合の詳細については、次の URL を参照してください。

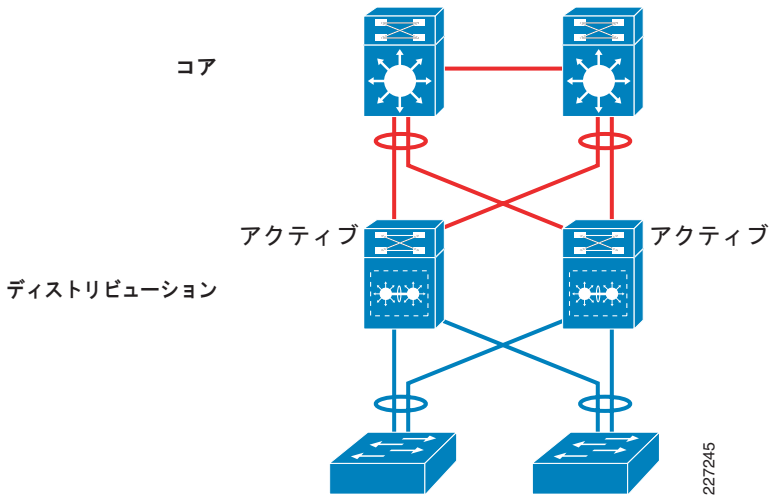
http://www.cisco.com/en/US/tech/tk389/tk213/technologies_tech_note09186a008009448d.shtml

（注）　PAgP または LACP プロトコル自体は、コアまたはアクセス レイヤに対する EtherChannel の不整合を引き起こしません。これは、両方のアクティブ ルータが共通の PAgP および LACP コントロール プレーン デバイス ID 情報を通知するためです。デュアルアクティブ状態では、PAgP が異なる MAC アドレスから送信された BPDU を検出し、ポートチャネルで error-disable 状態になります。

EIGRP および OSPF を使用したレイヤ 3 MEC

デュアルアクティブ状態の間、両方のアクティブ VSS ルータはそれぞれのレイヤ 3 MEC インターフェイスを動作可能状態に保ちます。しかし、各アクティブ ルータは相手側のシャーシに関連付けられているリンク メンバを削除します。これは、各ルータが、リモート ピアはダウンしており、そのピアに関連付けられているすべてのインターフェイスを削除する必要があると考える状況を反映しています。インターフェイスが削除されることで、コアに対するトポロジ アップデートが開始されます。しかし、各シャーシには物理的に以前のインターフェイスがすべてあります。各アクティブ ルータは、これらのインターフェイスを使用して、ネイバーおよびルーティング プロトコル アップデートを送信し続けます。図 4-11 を参照してください。

図 4-11　EIGRP および OSPF トポロジを使用したレイヤ 3 MEC のデュアルアクティブ状態



レイヤ 3 MEC ベースのトポロジでは、図 4-11 に示すトポロジでネイバーが 2 つしかありません。正常に動作するトポロジでは、コアは特定の宛先に対し 1 つの論理 ルータと 1 つのパスだけを認識します。しかし、VSS は特定の宛先に対し 2 つのパス（各コア ルータから 1 つずつ）を認識します。デュアルアクティブ状態では、ルーティング プロトコルに応じて、コア ルータが複数のルータを認識する可能性があります。EtherChannel ハッシュにより、hello メッセージ（マルチキャスト）と update（ユニキャスト）メッセージを送信するためのリンクの非対称型の選択が可能になります。コアから VSS へのフローでは、ハッシュ計算により、これらのメッセージ タイプが異なる EtherChannel リンク メンバ

上で送信される可能性がありますが、コントロール プレーン用の VSS からコアへの接続はローカル インターフェイス上にとどまります。これにより、デュアルアクティブ イベントの間、隣接関係がリセットされるか不安定になる可能性があります。

EIGRP

hello メッセージと update メッセージが、コアからいずれかの VSS アクティブ シャーシへのメンバ リンクの 1 つにハッシュされる方法に応じて、一部のルータでは隣接関係がそのままになりますが、その他のルータでは隣接関係が不安定になることがあります。不安定になると、ネイバー ホールド タイマーまたは NSF 信号タイマーのタイムアウト、および stuck-in-INIT エラーが原因で、隣接関係がリセットされることがあります。

OSPF

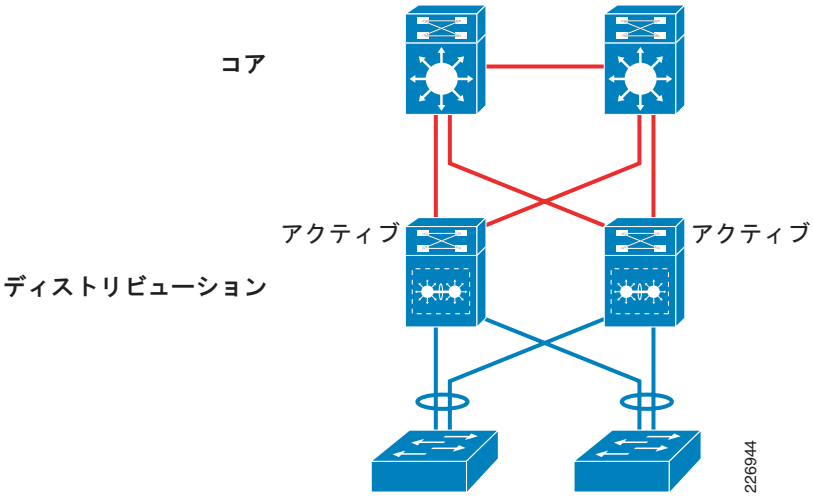
デュアルアクティブ イベントが OSPF ベースのネットワークで発生すると、どちらのアクティブ ルータでも隣接関係が安定することはありません。OSPF が隣接関係を構成するためには、プロトコルでネイバーの可用性の双方向の検証が必要です。デュアルアクティブ状態では、OSPF ネイバーは、2 台のアクティブ ルータによってハッシュされた複数のメッセージを受信します。また、コア ルータは、同じルータ ID を通知する 2 台のルータを認識します。重複するルータ ID と隣接関係のリセットの組み合わせにより、アクセス レイヤのサブネットがコア ルータの OSPF データベースから削除されます。

どちらのルーティング プロトコルでも、隣接関係のリセットまたは不安定が原因で、ルートの取り消しとユーザ トラフィックの中断が起こります。また、アクセス レイヤでのレイヤ 2 が、「レイヤ 2 MEC に対する影響」（P.4-20）で説明したように、error-disabled 状態になります。

EIGRP および OSPF を使用したレイヤ 3 ECMP

図 4-12 に示すように、通常の動作可能トポロジでは、コア ルータは、特定の宛先に対し、1 つの論理ルータと 2 つのパスだけを認識します。しかし、VSS は、特定の宛先に対し 4 つのパス（各コア ルータから 2 つずつ）を認識します。デュアルアクティブ状態では、ルーティング プロトコルに応じて、コアが複数のルータを認識する可能性があります。レイヤ 3 MEC トポロジを使用した場合と異なり、ネイバー ルータとの隣接関係と ECMP（独立パスとしての）を使用したルーティング アップデートに対するハッシュ関連の影響はありません。

図 4-12　EIGRP および OSPF トポロジを使用したレイヤ 3 ECMP



EIGRP

デュアルアクティブ状態の間、ルータは隣接関係を失いません。EIGRP には、衝突するルータ ID がなく（EIGRP がいくつかのサブネットの再配布ポイントとして使用されないかぎり）、各リンクは、コアルータまたはアクティブ VSS メンバで隣接関係の変更が発生しないようにルーティングされます。ユーザ トラフィックは、事実上ユーザ データ トラフィックへの影響なく転送され続けます。そのため、このトポロジでは、デュアルアクティブ状態によるレイヤ 3 接続への影響はありません。しかし、レイヤ 2 MEC が error-disable 状態になり、ユーザ データ トラフィックが中断される可能性があります。

OSPF

デュアルアクティブ イベント中、同じ IP ループバック アドレスを使用する 2 台のルータが重複するルータ ID を通知します。両方のアクティブ ルータは同じ LSA を通知し、コア ルータでアクセス レイヤ サブネットに対する LSA フラッディングが発生します。

検出方法

ここでは、さまざまな検出方法とその操作手順および復旧手順について説明します。デュアルアクティブ状態を検出するには次の方法があります。

- 拡張 PAgP
- fast-hello：VSLP フレームワーク ベースの hello
- Bidirectional Forwarding Detection（BFD）

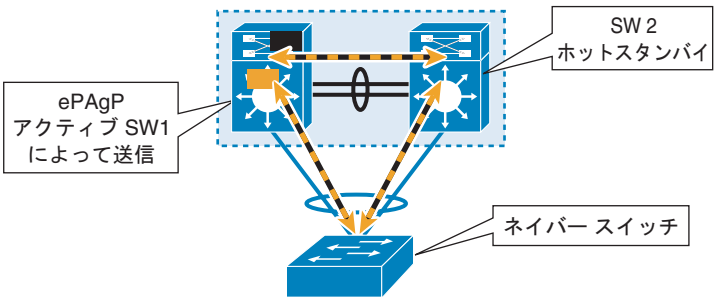
（注）　拡張 PAgP と BFD は、Cisco IOS リリース 12.2(33)SXH からサポートされ、fast-hello には Cisco IOS リリース 12.2(33)SXI が必要です。

拡張 PAgP

通常動作

拡張 PAgP は PAgP プロトコルを拡張したものです。拡張 PAgP では、新しい Type Length Value（TLV; タイプ、長さ、値）が導入されています。ePAgP メッセージの TLV には、デュアルアクティブ検出のための ID として、アクティブ スイッチの MAC アドレス（アクティブ スイッチのバックプレーンに由来します）が含まれます。通常の動作モードでは、アクティブ スイッチだけが拡張 PAgP メッセージを送信します。アクティブ スイッチは、両方の MEC リンク メンバ上で、拡張 PAgP メッセージを、30 秒ごとに送信します。ePAgP 検出は、ネイバー スイッチを、デュアルアクティブ状態を検出するための第 3 の接続として使用します （すべての検出方法では、スイッチが VSL リンクの状態を得るための第 3 の接続が必要です。これは、VSL リンクが動作していないことを検出しただけでは、どちらの側も、相手がダウンしていると見なすことができないためです）。アクティブ スイッチ ID を含む ePAgP メッセージは、ローカルに接続された MEC リンク メンバ上または VSL リンク上で、アクティブ スイッチにより送信されます。この動作を、図 4-13 の黒と黄色の四角形に示します。これは、ePAgP メッセージと、ネイバー スイッチを経由するパスを示します。ネイバー スイッチは、各アップリンク経由のこれらのメッセージの両方を単に反映します。これにより、アクティブ スイッチが VSL リンクの双方向の完全性を単独で検証します。ネイバー スイッチがこの動作を支援するためには、拡張 PAgP をサポートする Cisco IOS ソフトウェア バージョンが必要になります。図 4-13 を参照してください。

図 4-13　拡張 PAgP の通常の動作



通常モード

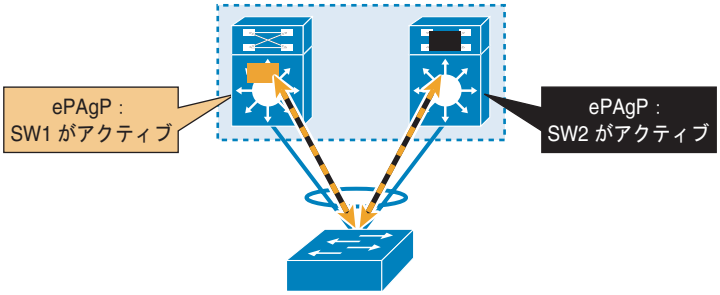
ePAgP メッセージ パス：
アクティブ SW1 → VSL リンク → ホットスタンバイ → ネイバー スイッチ → アクティブ スイッチ
アクティブ SW1 → 信頼ローカル MEC メンバ → ネイバー スイッチ →
ホットスタンバイ → VSL リンク → アクティブ スイッチ

226945

拡張 PAgP を使用したデュアルアクティブ検出

デュアルアクティブでは、VSL バンドル内のすべてのリンクが動作不能になり、ホットスタンバイ スイッチ（図 4-14 の SW2）がアクティブに遷移します（リモート スイッチのステータスは知りません）。SW2 がアクティブになると、自身のアクティブ スイッチ ID を使用して、独自の拡張 PAgP メッセージを生成し、ローカルに接続された MEC リンク メンバを経由し、ネイバー スイッチを経由して SW1 に送信します。VSL リンクがダウンしていると、以前のアクティブ スイッチ（SW1）は、自身の拡張 PAgP メッセージの受信を停止し、リモート スイッチ（以前のホットスタンバイ）によって生成された拡張 PAgP メッセージを受信します。これら 2 つのメッセージとそのパスを、図 4-14 に黒と黄色の四角形で示します。SW1 は、自身がアクティブ スイッチであったことを覚えており、以前アクティブだった SW1 だけがデュアルアクティブ状態を検出してそこから復旧できます。

図 4-14　拡張 PAgP を使用したデュアルアクティブ検出の動作



デュアルアクティブ検出

ePAgP メッセージ パス：
アクティブ SW2 → 信頼ローカル MEC メンバ リンク → ネイバー スイッチ → アクティブ SW1
アクティブ SW1 → ローカル MEC メンバ リンク → ネイバー スイッチ → アクティブ SW2

226946

SW2 からの ePAgP メッセージを以前のアクティブなスイッチ（SW1）が受信すると、SW1 は自身のアクティブ スイッチ ID（ローカル バックプレーンに由来する MAC アドレス）と新しいアクティブ スイッチ ID を比較します。受信した ID と期待される ID が異なる場合、以前のアクティブ シャーシは、デュアルアクティブ状態が VS ドメインで発生していると判断し、復旧処理を開始します。デュアルアクティブ状態は、以前のアクティブなスイッチ上で CLI を実行することでだけ表示されます。これは、検出がアクティブなのがこのスイッチであるためです。

```
6500-VSS# sh switch virtual dual-active summary
Pagp dual-active detection enabled: Yes
Bfd dual-active detection enabled: Yes

No interfaces excluded from shutdown in
 recovery mode

In dual-active recovery mode: Yes
  Triggered by: PAgP detection
  Triggered on interface: Gi2/8/19
  Received id: 0019.a927.3000
  Expected id: 0019.a924.e800
```

（注） Cisco IOS リリース（12.2(33) SXH および 12.2(33) SXI）では、以前のアクティブ スイッチと新たなアクティブ スイッチを区別する方法がありません。どちらのスイッチもアクティブになり、どちらも同じコマンド プロンプトを表示します。そのため、上記のコマンドを実行するときに、操作が困難になります。将来のリリースでは、オペレータが 2 つのアクティブなスイッチを区別できるように、以前のアクティブ スイッチのプロンプトが意味のあるものに変更される可能性があります。

デュアルアクティブ状態では、各スイッチで異なる種類の syslog メッセージが生成されます。以前のアクティブ スイッチ（SW1）では、次の syslog メッセージが表示されます。

```
%PAGP_DUAL_ACTIVE-SW2_SP-1-RECOVERY: PAgP running on Gi2/8/19 triggered dual-active
recovery: active id 0019.a927.3000 received, expected 0019.a924.e800
%DUAL_ACTIVE-SW2_SP-1-DETECTION: Dual-active condition detected: all non-VSL and
non-excluded interfaces have been shut down
```

新たなアクティブ スイッチ（SW2）では、次の syslog メッセージが表示されます。

```
%VSLP-SW1_SP-3-VSLP_LMP_FAIL_REASON: Te1/5/4: Link down
%VSLP-SW1_SP-3-VSLP_LMP_FAIL_REASON: Te1/5/5: Link down
%VSLP-SW1_SP-2-VSL_DOWN:   All VSL links went down while switch is in ACTIVE role
```

拡張 PAgP プロトコルをサポートするネイバー スイッチでも、次のデュアルアクティブ発生メッセージが表示されます。

```
%PAGP_DUAL_ACTIVE-SP-3-RECOVERY_TRIGGER: PAgP running on Gi6/1 informing virtual switches
of dual-active: new active id 0019.a927.3000, old id 0019.a924.e800
```

注意 ネイバー スイッチでは、通常のスイッチオーバー時にもこの syslog メッセージが表示されます。これは、実際に何が起きたかを知らないためです。単に複数のスイッチがアクティブであることを主張していることが検出されます。

拡張 PAgP のサポート

前のセクションで説明したように、デュアルアクティブ検出をサポートするためには、ネイバー スイッチが拡張 PAgP プロトコルを理解する必要があります。これは、拡張 PAgP が動作するためには、MEC 構成で PAgP プロトコルが動作している必要があることも意味します。PAgP をディセーブルにして拡張 PAgP を動作させることはできません。拡張 PAgP は、レイヤ 2 またはレイヤ 3 PAgP MEC メンバ上でイネーブルにできます。つまり、拡張 PAgP は、VSS とコア ルータの間で実行できます。表 4-12 を参照してください。

表 4-12　Cisco IOS バージョンの拡張 PAgP のサポート

プラットフォーム	ソフトウェア	説明
Cisco Catalyst 6500	12.2(33)SXH	Sup720 および Sup32
Cisco Catalyst 45xx および Cisco Catalyst 49xx	12.2(44)SG	
Cisco Catalyst 29xx、Cisco Catalyst 35xx および Cisco Catalyst 37xx	12.2(46)SE	Cisco Catalyst 37xx スタックにはサポートがありません。後述の文章を参照してください。
Cisco Catalyst 37xx スタック	未サポート	クロススタック EtherChannel は LACP だけをサポート

PAgP は、Cisco Catalyst 37xx スタック構成を除くすべてのプラットフォームでサポートされています。この構成では、クロススタック EtherChannel（LACP）で VSS との MEC 接続が必要です。クロススタック EtherChannel は PAgP をサポートしていないため、デュアルアクティブ検出のために拡張 PAgP を使用できません。このサポートの違いを解決するための一般的なアプローチは、同じスタック メンバからの 2 つの EtherChannel リンクを使用することです。この解決方法で、シングル ポイント障害ができるため、最適な設計上の選択ではありません。その EtherChannel を含むスタック メンバが障害になると、スタックからの接続全体が障害になります。このシングル ポイント障害の問題を解決するために、2 つのデュアルリンク EtherChannel グループを、それぞれ VSS に接続された個別のスタック メンバ上に配置できます。ただし、これによりループのあるトポロジが作成されます。ループのないトポロジでは、単一の EtherChannel バンドルを複数のメンバに分散させる必要があり、その結果、LACP が必要になります。

スタック専用のアクセス レイヤ要件には、次の 2 つの解決方法があります。

- fast-hello または BFD をデュアルアクティブ検出方法として使用します（「fast-hello（VSLP フレームワークベースの検出）」（P.4-27）または「Bidirectional Forwarding Detection」（P.4-31）を参照してください）。
- 拡張 PAgP は、レイヤ 2 またはレイヤ 3 PAgP MEC メンバ上でイネーブルにできます。つまり、拡張 PAgP を VSS とコア ルータの間で実行できます。ただし、コア ルータでは、拡張 PAgP のサポートと、VSS へのレイヤ 3 MEC トポロジの実装が必要です。

拡張 PAgP の設定とモニタリング

拡張 PAgP のデュアルアクティブ検出はデフォルトでイネーブルになっていますが、特定の MEC グループを信頼できるグループとして指定する必要があります。MEC グループを信頼できるメンバとして識別する特定の CLI は、仮想スイッチ コンフィギュレーションで必要です。すべての PAgP ネイバーで信頼をイネーブルにしない理由は、保護されていないスイッチ、意図しないベンダー接続など、不要な拡張 PAgP メンバを回避するためです。

EtherChannel 上で拡張 PAgP をイネーブルにするには、次の条件が必要です。

- 信頼を追加または削除する際に、MEC が管理ディセーブル状態になっている必要があります。そうでない場合、エラー メッセージが表示されます。
- MEC メンバで PAgP プロトコルが動作している必要があります。接続の両側で、希望するモードで PAgP を設定することをお勧めします。

pagp rate fast コマンドを使用して PAgP hello タイマーをデフォルト値の 30 秒から 1 秒に変更しても、ユーザ トラフィックのコンバージェンス時間の短縮には役立ちません。これは、デュアルアクティブ検出は、PAgP パケットがどれだけ早く送信されるかには依存せず、デュアルアクティブ検出を行うために、ホットスタンバイ スイッチが、自身のアクティブ ID を使用して拡張 PAgP メッセージをどれだけ早く生成できるかに依存するためです。図 4-15 を参照してください。

図 4-15　PAgP が動作した状態での MEC 上の信頼のイネーブル化

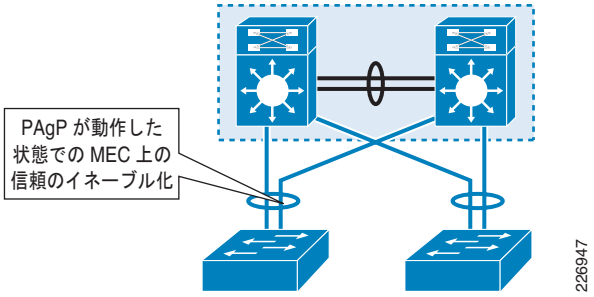


図 4-15 に、MEC メンバが拡張 PAgP ベースのデュアルアクティブ検出に参加できるようにするために必要な信頼の設定を示します。PAgP が動作した状態で MEC 上で信頼をイネーブルにするには、次のコマンドを使用します。

```
6500-VSS(config)# switch virtual domain 10
6500-VSS(config-vs-domain)# dual-active detection pagp trust channel-group 205
```

拡張 PAgP のサポートと信頼設定は、VSS スイッチと拡張 PAgP ネイバー上で、次の設定例に示すコマンドで確認できます。

VSS スイッチ：

```
6500-VSS# show switch virtual dual-active pagp
PAgP dual-active detection enabled: Yes
PAgP dual-active version: 1.1

! << Snip >>

Channel group 205 dual-active detect capability w/nbrs
Dual-Active trusted group: Yes
          Dual-Active     Partner              Partner   Partner
Port      Detect Capable  Name                 Port      Version
Gi1/8/19  Yes             cr7-6500-3           Gi5/1     1.1
Gi1/9/19  Yes             cr7-6500-3           Gi6/1     1.1
```

拡張 PAgP をサポートするネイバー スイッチ：

```
4507-Switch# show pagp dual-active
PAgP dual-active detection enabled: Yes
PAgP dual-active version: 1.1

Channel group 4
          Dual-Active     Partner              Partner   Partner
Port      Detect Capable  Name                 Port      Version
Te1/1     Yes             cr2-6500-VSS         Te2/2/6   1.1
Te2/1     Yes             cr2-6500-VSS         Te1/2/6   1.1
```

fast-hello（VSLP フレームワークベースの検出）

fast-hello は、最も新しいデュアルアクティブ検出方法であり、Cisco IOS 12.2(33) SXI 以降のリリースで使用できます。fast-hello を展開する主な理由は次のとおりです。

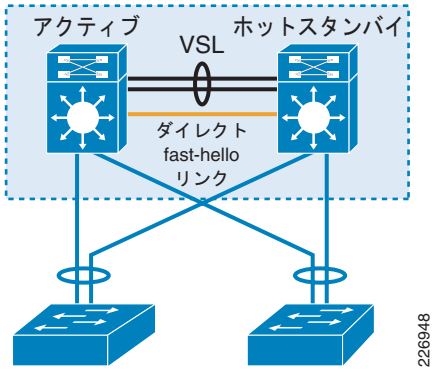
- 拡張 PAgP の展開が不可能な場合。サーバアクセス接続で、サーバが VSS に接続され、コア接続がレイヤ 3 MEC ベースでない場合など。
- インストールされている Cisco IOS のバージョンで拡張 PAgP がサポートされていない場合。
- EtherChannel グループ プロトコルが LACP の場合。

- 設定が単純であることが必要で、BFD の代わりに fast-hello を使用する場合（「Bidirectional Forwarding Detection」（P.4-31）を参照してください）。

通常動作

fast-hello は、直接接続型のデュアルアクティブ検出メカニズムです。2 つの仮想スイッチ ノード間で、セッションを確立するための専用の物理ポートが必要です。fast-hello はコネクションレス型のプロトコルであり、fast-hello 隣接関係を形成するためにいかなる種類のハンドシェイク メカニズムも使用しません。適切な TLV 情報を使用したピア ノードからの受信 fast-hello メッセージにより fast-hello の隣接関係が確立されます。図 4-16 を参照してください。

図 4-16　　fast-hello の設定



各デュアルアクティブ fast-hello メッセージには、TLV 内に次の情報が含まれます。

- *VSS ドメイン ID*：VSS 仮想スイッチ ノードでは、各 hello メッセージに共通のドメイン ID が含まれている必要があります。
- *スイッチ ID*：各仮想スイッチ ノードは、自身が送信する hello メッセージ用に、ローカルな仮想スイッチ ID を通知します。
- *スイッチの優先順位*：各仮想スイッチ ノードは、自身が送信する hello メッセージ用に、ローカルな仮想スイッチの優先順位を通知します。

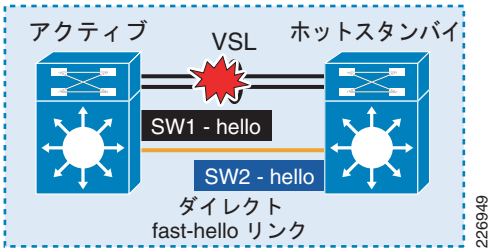
デフォルトでは、各仮想スイッチ ノードは 2 秒ごとに fast-hello パケットを送信します。デュアルアクティブ fast-hello 送信タイマーはハードコーディングされており、VSS システム内のエンドユーザに対して透過的です。ハードコーディングされた fast-hello タイマーは設定も調整もできず、**debug** コマンドを使用した確認だけが可能です。各仮想スイッチ ノードは、ピア ノードとのセッションを確立するために、fast-hello をデフォルトの間隔で送信します。確立されたデュアルアクティブ fast-hello 隣接関係は、いずれかの仮想スイッチ ノードで、hello メッセージを 5 回連続して送信しても、ピア ノードからの hello を受信しなくなった場合に解放されます。デフォルトでは、ホールドダウン タイマーは 10 秒にハード コーディングされています。問題発生や設定ミスなど、何らかの理由で隣接関係を確立するための処理が失敗すると、設定されている側は、セッションを確立するために、デフォルトの間隔で hello メッセージを送信し続けます。fast-hello サポートのために設定された専用のリンクは、コントロール プレーンやユーザ データ トラフィックを伝送できません。

fast-hello を使用したデュアルアクティブ検出

アクティブ スイッチもホットスタンバイ スイッチも、リモート ピアと VSS バンドルの障害を区別できません。アクティブ スイッチの SSO プロセスは、ホットスタンバイに対する通信の損失に対応する必要があるため、VSS コントロール プレーンに対し、fast-hello 用に設定されているリモート ポートを含め、ホットスタンバイ スイッチに関連付けられているすべてのインターフェイスとラインカードを削除するよう通知します。しかし、デュアルアクティブ中は、fast-hello を伝送するために設定され

ているリンクは動作可能であり（ホットスタンバイはまだ動作可能です）、定期的に hello を交換します。その結果、以前のアクティブ スイッチは、fast-hello リンクに関するこの矛盾する情報に気付き、リモート ノードが動作可能な場合にだけこの状況が発生すると判断します。これは、デュアルアクティブが発生したことを意味します。もしそうでなければ、以前のアクティブ スイッチは fast-hello を受信することはありません。図 4-17 を参照してください。

図 4-17　　fast-hello を使用したデュアルアクティブ検出



以前のアクティブ スイッチは、次の条件がすべて満たされている場合にデュアルアクティブ検出処理を開始します。

- VSL EtherChannel チャネル全体が動作不能であること。
- 各仮想スイッチ ノード上の fast-hello リンクが動作可能であること。
- 以前のアクティブ スイッチが、通常の 2 秒間の間隔で、少なくとも 1 回 fast-hello を送信していること。

VSL EtherChannel を失う際、以前のアクティブ スイッチは、通常の間隔で fast-hello を少なくとも 1 回送信した後、fast-hello をより高速に送信します（500 ミリ秒ごと）。この設計により、アクティブおよびホットスタンバイの過渡的なネットワーク状態で、CPU が無駄に消費されるのを防ぎます。次の **show** コマンドの出力は、デュアルアクティブ状態と、以前のアクティブ スイッチでの検出状態を示します。

```
6500-VSS# show switch virtual dual fast-hello
Fast-hello dual-active detection enabled: Yes

Fast-hello dual-active interfaces:
Port       Local State    Peer Port    Remote State
-----------------------------------------------------
Gi1/5/1    Link up        Gi2/5/1      Link up

6500-VSS# show switch virtual dual-active summary
Pagp dual-active detection enabled: Yes
Bfd dual-active detection enabled: Yes
Fast-hello dual-active detection enabled: Yes

No interfaces excluded from shutdown in recovery mode

In dual-active recovery mode: Yes
  Triggered by: Fast-hello detection
  Triggered on interface: Gi1/5/1
```

デュアルアクティブ状態が発生した場合に、以前のアクティブ スイッチ（SW1）で表示される syslog メッセージは、次のとおりです。

```
Dec 31 22:35:58.492: %EC-SW1_SP-5-UNBUNDLE: Interface TenGigabitEthernet1/5/4 left the
port-channel Port-channel1
Dec 31 22:35:58.516: %LINK-SW1_SP-5-CHANGED: Interface TenGigabitEthernet1/5/4, changed
state to down
Dec 31 22:35:58.520: %LINEPROTO-SW1_SP-5-UPDOWN: Line protocol on Interface
TenGigabitEthernet1/5/4, changed state to down
```

```
Dec 31 22:35:58.536: %VSLP-SW1_SP-3-VSLP_LMP_FAIL_REASON: Te1/5/4: Link down
Dec 31 22:35:58.540: %VSLP-SW1_SP-2-VSL_DOWN:   Last VSL interface Te1/5/4 went down
Dec 31 22:35:58.544: %LINEPROTO-SW2_SP-5-UPDOWN: Line protocol on Interface
TenGigabitEthernet1/5/1, changed state to down

Dec 31 22:35:58.544: %VSLP-SW1_SP-2-VSL_DOWN:   All VSL links went down while switch is in
ACTIVE role

! << snip >>

Dec 31 22:35:59.652: %DUAL_ACTIVE-SW1_SP-1-DETECTION: Dual-active condition detected: all
non-VSL and non-excluded interfaces have been shut down ! <- Fast-hello triggers recovery
! process and starts recovery process the old active switch.
Dec 31 22:35:59.652: %DUAL_ACTIVE-SW1_SP-1-RECOVERY: Fast-hello running on Gi1/5/1
triggered dual-active recovery
! << snip >>
Dec 31 22:36:09.583: %VSDA-SW1_SP-3-LINK_DOWN: Interface Gi1/5/1 is no longer dual-active
detection capable
```

デュアルアクティブ状態が発生した場合の、新たなアクティブ スイッチ（SW2）上の syslog メッセージは、次のとおりです。

```
Dec 31 22:35:58.521: %PFREDUN-SW2_SPSTBY-6-ACTIVE: Initializing as Virtual Switch ACTIVE
processor  Ð Starting NSF Recovery process
! << snip >>

Dec 31 22:36:09.259: %VSDA-SW2_SP-3-LINK_DOWN: Interface Gi2/5/1 is no longer dual-active
detection capable Ð Dual ACTIVE fast-hello link goes down and declares no longer
dual-active detection capable
```

fast-hello の設定とモニタリング

fast-hello の設定は単純です。次のようにして、まず仮想スイッチ ドメインでグローバルにイネーブルにし、専用のイーサネット ポートで定義します。

VSS グローバル コンフィギュレーション モードでイネーブルにします。

```
6500-VSS(config)# switch virtual domain 1
6500-VSS(config-vs-domain)# dual-active detection fast-hello
```

インターフェイス レベルで fast-hello をイネーブルにします。

```
6500-VSS(config)# int gi1/5/1
6500-VSS(config-if)# dual-active fast-hello

WARNING: Interface GigabitEthernet1/5/1 placed in restricted config mode. All extraneous
configs removed!

6500-VSS(config-if)# int gi2/5/1
6500-VSS(config-if)# dual-active fast-hello

WARNING: Interface GigabitEthernet2/5/1 placed in restricted config mode. All extraneous
configs removed!

%VSDA-SW2_SPSTBY-5-LINK_UP: Interface Gi1/5/1 is now dual-active detection capable
%VSDA-SW1_SP-5-LINK_UP: Interface Gi2/5/1 is now dual-active detection capable
```

fast-hello のサポートがイネーブルになっているリンクは、デュアルアクティブ fast-hello メッセージだけを伝送します。STP、CDP、Dynamic Trunking Protocol（DTP; ダイナミック トランキング プロトコル）、IP などのすべてのデフォルト ネットワーク プロトコルは自動的にディセーブルになり、処理されません。物理イーサネット ポートだけが fast-hello コンフィギュレーションをサポートできます。SVI やポートチャネルなどのその他のポートは、fast-hello リンクとして使用できません。冗長性

を持たせるために、複数の fast-hello リンクを設定できます。スーパーバイザが *10 ギガビット専用* モードで設定されていなければ、Sup720-10G 1 ギガビット アップリンク ポートを使用できます。fast-hello コンフィギュレーションがイネーブルになっているポートのステータス（アクティブおよびホットスタンバイ スイッチ）は、次の **show** コマンドの出力例を使用して知ることができます。

```
6500-VSS# show switch virtual dual-active fast-hello
Fast-hello dual-active detection enabled: Yes
Fast-hello dual-active interfaces:
Port       Local State    Peer Port    Remote State
-----------------------------------------------------
Gi1/5/1    Link up        Gi2/5/1      Link up

6500-VSS# remote command standby-rp show switch virtual dual-active fast-hello
Fast-hello dual-active detection enabled: Yes

Fast-hello dual-active interfaces:
Port       Local State    Peer Port    Remote State
-----------------------------------------------------
Gi2/5/1    Link up        Gi1/5/1      Link up
```

Bidirectional Forwarding Detection

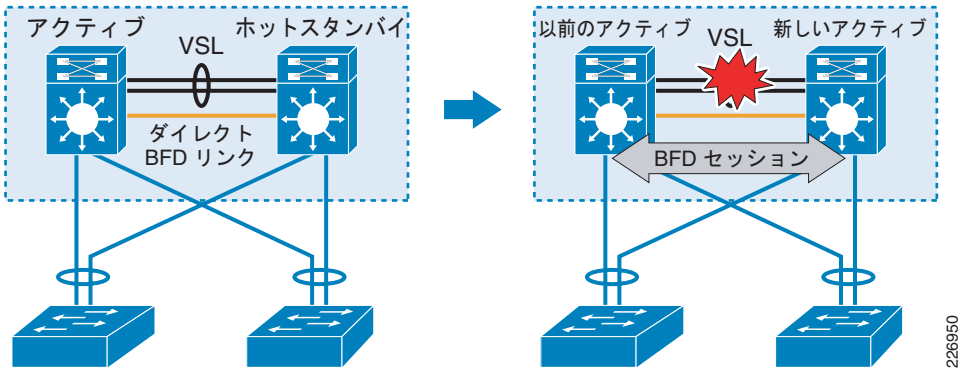
BFD は、次の理由によりデュアルアクティブ検出が不可能な場合の代替手段です。

- Cisco IOS のバージョンの制限により、拡張 PAgP または fast-hello 展開ができない場合。
- EtherChannel グループ プロトコルが LACP であり、fast-hello をサポートする Cisco IOS バージョンが使用できない場合。
- 特定のトポロジで短時間でのコンバージェンスが必要な場合。OSPF がイネーブルになっている ECMP ベースのトポロジなど。

通常動作

fast-hello 検出と同様に、BFD 検出では、VSS メンバ シャーシ間専用の第 3 の接続が必要です。VSS では、デュアルアクティブ検出のために、BFD バージョン 1 のエコー モードを使用します。一般的な BFD の情報については、cisco.com を参照してください。BFD 検出は、受動的な検出方法です。VSS が正常に動作しているときは、BFD が設定されたインターフェイスは up/up のままになります。ただし、そのリンク上で BFD セッションはアクティブになりません。図 4-18 を参照してください。

図 4-18　Bidirectional Forwarding Detection（BFD）



BFD を使用したデュアルアクティブ検出

BFD セッションの確立は、デュアルアクティブ状態が発生したことを示します。VSS は単一の論理ノードであるため、通常の状態では、VSS は自身との BFD セッションを確立できません。デュアルアクティブ イベントが発生すると、シャーシ間で BFD セッションを可能にする専用の BFD リンクを除き、2 つのシャーシが物理的に分離されます。事前に設定され、接続されたスタティック ルートが BFD セッションを確立します（BFD セッションの確立メカニズムの説明については、「BFD の設定とモニタリング」（P.4-33）を参照してください）。BFD セッションは非常に短時間（1 秒未満）だけ存在するため、BFD セッションのアクティビティは直接監視できません。BFD セッションの確立またはティアダウン ログは、BFD のデバッグ コマンドがイネーブルになるまで表示できません。ただし、以前のアクティブ スイッチでは、次の syslog が表示されます。

```
10:28:56.738: %LINK-SW1_SP-3-UPDOWN: Interface Port-channel1, changed state to down
10:28:56.742: %LINEPROTO-SW1_SP-5-UPDOWN: Line protocol on Interface
TenGigabitEthernet1/5/4, changed state to down
10:28:56.742: %VSLP-SW1_SP-3-VSLP_LMP_FAIL_REASON: Te1/5/4: Link down
10:28:56.742: %EC-SW1_SP-5-UNBUNDLE: Interface TenGigabitEthernet2/5/4 left the
port-channel Port-channel2
10:28:56.750: %VSLP-SW1_SP-2-VSL_DOWN:   Last VSL interface Te1/5/4 went down
10:28:56.754: %VSLP-SW1_SP-2-VSL_DOWN:   All VSL links went down while switch is in ACTIVE
role
```

debug ip routing コマンドの出力では、VSL リンクが動作不能になり、リモート インターフェイスに対する接続の損失（以前のアクティブ スイッチから見て）が検出されるため、ピア スイッチ（ホット スタンバイ スイッチ）のルートの削除が表示されます。

```
Jul 31 10:29:21.394: RT: interface GigabitEthernet1/5/1 removed from routing table
Jul 31 10:29:21.394: RT: Pruning routes for GigabitEthernet1/5/1 (1)
```

次の syslog 出力は、以前のアクティブ スイッチ上での復旧処理を起動する BFD を示します。

```
10:29:21.202: %DUAL_ACTIVE-SW1_SP-1-RECOVERY: BFD running on Gi1/5/1 triggered dual-active
recovery <- 1
10:29:21.230: %DUAL_ACTIVE-SW1_SP-1-DETECTION: Dual-active condition detected: all non-VSL
and non-excluded interfaces have been shut down
```

次の syslog 出力は、デュアルアクティブ中の新しいアクティブを示します。番号 2 が示す太字のタイムスタンプに注意してください。これを、以前のアクティブ スイッチにおける BFD タイムスタンプと比較してください（前の例の番号 1 を参照してください）。

```
10:28:56.738: %VSLP-SW2_SPSTBY-3-VSLP_LMP_FAIL_REASON: Te2/5/4: Link down
10:28:56.742: %VSLP-SW2_SPSTBY-2-VSL_DOWN:   Last VSL interface Te2/5/4 went down
10:28:56.742: %VSLP-SW2_SPSTBY-2-VSL_DOWN:   All VSL links went down while switch is in
Standby role
```

次の出力は、新たなアクティブ スイッチ上での復旧処理を起動する BFD を示します。

```
10:28:56.742: %DUAL_ACTIVE-SW2_SPSTBY-1-VSL_DOWN: VSL is down - switchover, or possible
dual-active situation has occurred <- 2
10:28:56.742: %VSL-SW2_SPSTBY-3-VSL_SCP_FAIL: SCP operation failed
10:28:56.742: %PFREDUN-SW2_SPSTBY-6-ACTIVE: Initializing as Virtual Switch ACTIVE
processor
```

新たなアクティブ スイッチでの次の出力は、以前のアクティブ スイッチとの BFD セッションを確立するための、接続されたルートの設定を示しています。

```
10:28:58.554: RT: interface GigabitEthernet2/5/1 added to routing table
10:29:21.317: RT: interface GigabitEthernet2/5/1 removed from routing table
10:29:21.317: RT: Pruning routes for GigabitEthernet2/5/1 (1)
```

BFD を使用したデュアルアクティブ検出には、22 ～ 25 秒かかります（上記 syslog の番号 1 と 2 が指しているタイムスタンプを参照してください）。BFD では、fast-hello による検出方法と比べて、次の理由から、以前のアクティブ スイッチをシャットダウンするのに長い時間がかかります。

- BFD セッションの確立が IP 接続に基づいている。ホットスタンバイ スイッチでは、IP 接続を開始する前に、SSO を通じたコントロール プレーンの初期化が必要です。
- IP プロセスを開始し、接続されたスタティック ルートをインストールするための時間が必要。
- BFD セッションの初期化と 2 つのシャーシ間のセッションの確立に時間が必要。

検出時間が長いことによるユーザ データ トラフィックへの影響は、それほど大きくない可能性があり、ルーティング プロトコルとトポロジによって変わります。BFD ベースの検出は、コンバージェンス時間を短縮するために、特定のトポロジで必要になります。しかし、BFD ベースの検出方法は、fast-hello の改良された hello 検出に取って代わられ、将来のソフトウェア リリースで非推奨となります。「デュアルアクティブ状態によるコンバージェンスとユーザ データ トラフィックへの影響」（P.4-39）を参照してください。

BFD の設定とモニタリング

BFD の設定では、2 つの VSS シャーシ間で、専用の、直接接続された物理イーサネット ポートが必要です。BFD ペアリングは、レイヤ 3 EtherChannel または SVI インターフェイスでイネーブルにできません。Sup720-10G 1 ギガビット アップリンク ポートは、スーパーバイザが 10 ギガビット専用モードで設定されていない場合にだけ使用できます。

デュアルアクティブ用の BFD の設定は、標準インターフェイス上での通常の BFD の設定とは異なります。スイッチ間の BFD セッションの接続は、デュアルアクティブ状態の間だけ必要です。まず、グローバル仮想スイッチモードで BFD 検出をイネーブルにします。次に、リンクの両端で、専用の BFD インターフェイスに一意の IP サブネットが必要です。通常の動作状態では、2 つの接続されたインターフェイスは同じサブネットを共有できませんが、デュアルアクティブ イベント中は、BFD ピア接続用に共有が必要です。インターフェイスがペアになったら、仮想スイッチが、ペアになったインターフェイスを使用して、接続されたルートとして 2 つのスタティック ルートを自動的にインストールします。また、インターフェイスのペアの解除時には、スタティック ルートを削除します。

BFD ベースの検出方式を設定するには、次のコマンドを実行します。

VSS グローバル コンフィギュレーション モードでイネーブルにします。

```
6500-VSS(config)# switch virtual domain 10
6500-VSS(config)# dual-active pair interface gig 1/5/1 interface gig 2/5/1 bfd
```

専用のインターフェイス上で一意の IP サブネットと BFD 間隔をイネーブルにします。

```
6500-VSS# conf t
6500-VSS(config)# interface gigabitethernet 1/5/1
6500-VSS(config)# ip address 192.168.1.1 255.255.255.0
6500-VSS(config)# bfd interval 50 min_rx 50 multiplier 3

6500-VSS(config)# interface gigabitethernet 2/5/1
6500-VSS(config)# ip address 192.168.2.1 255.255.255.0
6500-VSS(config)# bfd interval 50 min_rx 50 multiplier 3
```

上記の設定シーケンスを実行すると、必要なスタティック ルートが自動的にインストールされます。ディスプレイ コンソールに次のメッセージが表示されます。

```
Console Message:

adding a static route 192.168.1.0 255.255.255.0 Gi2/5/1 for this dual-active pair
adding a static route 192.168.2.0 255.255.255.0 Gi1/5/1 for this dual-active pair
```

スイッチ 1 のインターフェイス（上記の例では 1/5/1）上で設定されたサブネットのスタティック ルートが、スイッチ 2（2/5/1）上にあるインターフェイスを経由して使用可能であることに注意してください。この設定が必要なのは、スタティック ルートにより、デュアルアクティブ イベント中にシャーシが分離された場合に、BFD セッションの接続が確立可能になるためです。BFD プロトコル自体には、セッションを確立するためにサブネットが分離されていなければならないという制限はありません。

（注）　推奨される BFD メッセージの間隔は 50 ～ 100 ミリ秒の間で、乗数の値が 3 です。ベスト プラクティスを検証する際に、推奨値を超えてタイマー値を増やしたところ、データ損失が増えました。BFD コンフィギュレーション用に、組織に属する IP アドレス範囲のどれにも含まれていない固有のサブセットを使用してください。BFD 関連の接続されたスタティック ルートを除外するため、再配布が接続された（他の接続で必要な場合）ルートマップを使用してください。

BFD 検出のコンフィギュレーションは、次の CLI コマンドで監視できます。

```
6500-VSS# sh switch virtual dual-active bfd
Bfd dual-active detection enabled: Yes

Bfd dual-active interface pairs configured:
  interface-1 Gi1/5/1 interface-2 Gi2/5/1

6500-VSS# sh switch virtual dual active summary
Pagp dual ACTIVE detection enabled: No
Bfd dual ACTIVE detection enabled: Yes
No interfaces excluded from shutdown in recovery mode
In dual ACTIVE recovery mode: No
```

設定上の注意事項

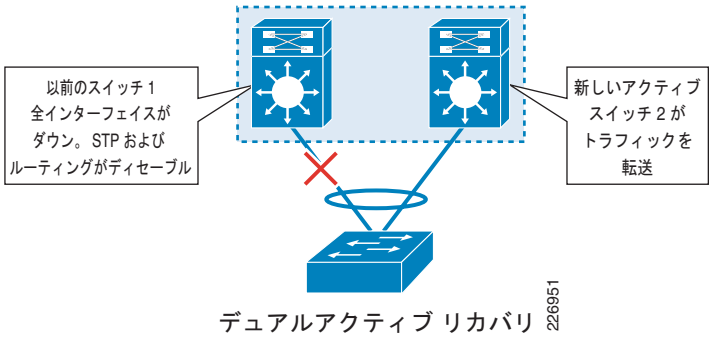
BFD hello タイマーの設定値は、両方のスイッチで同じであることが必要です。また、IP アドレスや BFD コマンドに関連するコンフィギュレーションを変更すると、グローバル仮想スイッチモードで BFD 検出コンフィギュレーションが削除され、手動で追加し直す必要があります。これは、不整合を回避するための設計です。というのは、デュアルアクティブ イベントが発生しない限りコンフィギュレーションの妥当性が検証できないためです。次の注意がコンソールに表示されます。

```
6500-VSS(config)# interface gig 1/5/1
6500-VSS(config-if)# ip address 14.14.14.14 255.255.255.0
The IP config on this interface is being used to detect dual-active conditions. Deleting
or changing this config has deleted the bfd dual-active pair: interface1: Gi1/5/1
interface2: Gi2/5/1
deleting the static route 3.3.3.0 255.255.255.0 Gi1/5/1 with this dual-active pair
deleting the static route 1.1.1.0 255.255.255.0 Gi2/5/1 with this dual-active pair
```

デュアルアクティブ復旧

検出方法によりデュアルアクティブ状態が検出されたら、復旧フェーズが始まります。復旧処理は、3 つの検出方法すべてで同じです。どの場合も、以前のアクティブ スイッチが復旧を起動します。このガイドに示す例では、SW1（元のまたは以前のアクティブ スイッチ）が、SW2 もアクティブ スイッチになったことを検出し、これによりデュアルアクティブ状態が検出されます。SW1 は、すべてのローカル インターフェイス（ループバック以外）をディセーブルにして、ネットワークが不安定になるのを回避します。SW1 は、ルーティング インスタンスと STP インスタンスもディセーブルにします。以前のアクティブ スイッチは、ネットワークから完全に除去されます。図 4-19 を参照してください。

図 4-19　デュアルアクティブ復旧



exclude interface オプションを使用して、デュアルアクティブ復旧処理中に、管理ポートなどの指定したポートを動作可能なままにできます。ただし、以前のアクティブ スイッチにはルーティング インスタンスがないため、**excluded port** コマンドにはルーティングされる接続がありません。次に、該当するコマンドの例を示します。

```
VSS(config-vs-domain)# dual-active exclude interface port_nubmer
```

（注）　SVI または EtherChannel 論理インターフェイスは、デュアルアクティブ イベント中は除外できません。

通常状態およびデュアルアクティブ状態の間は、どちらのシャーシに対してもコンソールベースでアクセスすることを強くお勧めします。

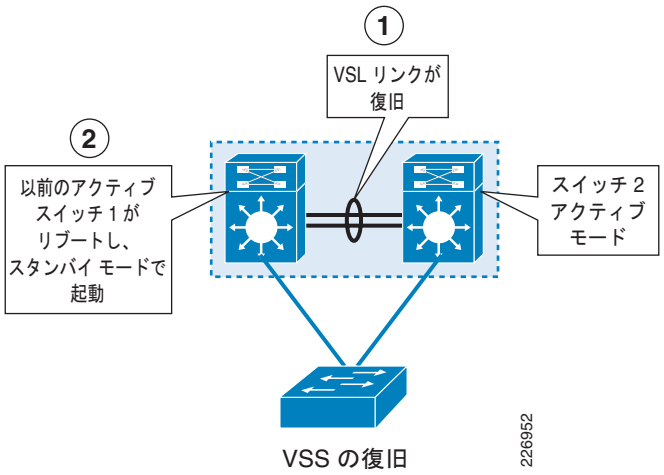
VSS 復旧

VSS 復旧処理は、VSL 接続が再確立されると開始されます。デュアルアクティブの原因になる次のイベントでは、VSS スイッチ メンバ間の VSL 接続が復旧されます。

- 光ファイバ接続の復旧。これは、ネットワークでファイバ リンクに物理的な問題がある場合に発生します。
- コンフィギュレーションの変更の反転。VSL バンドルがシャットダウンされることがあります。
- 故障しているハードウェアの復旧。これは、レジリエンシー設計を採用している場合、最も発生する可能性が低いイベントです。

図 4-20 は、VSS の復旧処理の概要を示します。

図 4-20　VSS 復旧



VSL 接続が確立されると、役割のネゴシエーション（RRP プロトコルを通じて）により以前のアクティブ スイッチ（図 4-20 の SW1）がホットスタンバイ スイッチになることが決定されます。既存の（新しい）アクティブ スイッチの役割を変更して、さらなるデータ損失を引き起こす理由はありません。ソフトウェア リセットを行わずにスイッチを直接ホットスタンバイ状態にできないため、SW1 をリブートする必要があります。コンフィギュレーションの不一致が見つからない場合、SW1 は自動的にリブートし、ホットスタンバイ モードで初期化されます。SW 1 上のすべてのインターフェイスがオンラインになり、SW1 がパケットのフォワーディングを開始し、ネットワークの完全なキャパシティが回復されます。たとえば、SW1（以前のアクティブなスイッチ）上での VSL バンドルの復旧中に、次のコンソール メッセージが表示されます。

```
17:36:33.809: %VSLP-SW1_SP-5-VSL_UP:  Ready for Role Resolution with Switch=2,
MAC=001a.30e1.6800 over Te1/5/5
17:36:36.109: %dual ACTIVE-1-VSL_RECOVERED: VSL has recovered during dual ACTIVE
situation: Reloading switch 1
! << snip >>
17:36:36.145: %VSLP-SW1_SP-5-RRP_MSG: Role change from ACTIVE to HOT_STANDBY and hence
need to reload
Apr  6 17:36:36.145: %VSLP-SW1_SP-5-RRP_MSG: Reloading the system...
17:36:37.981: %SYS-SW1_SP-5-RELOAD: Reload requested Reload Reason: VSLP HA role change
from ACTIVE to HOT_STANDBY.
```

デュアルアクティブ復旧ステージの最中にコンフィギュレーションが変更されると、復旧したシステムで、**reload** コマンドを使用し手動でコンフィギュレーションを同期することが必要になります。デュアルアクティブなどのネットワーク停止が発生した場合、多くのネットワーク オペレータは、ネットワーク停止を解決するのに役立つ追加のコマンドを探してコンフィギュレーション モードを開始する習慣を持っています。しかし、変更を行わなくても、コンフィギュレーション モードを開始および終了しただけで、コンフィギュレーションが dirty と見なされ、手動介入が必要になります。デュアルアクティブ状態は、VSL ポートチャネル インターフェイスの偶発的なソフトウェア シャットダウン時にも発生します。コンフィギュレーション同期処理は、この変更を両方のシャーシに反映させます。VSL 接続を復旧するための唯一の方法は、コンフィギュレーション モードを開始することです。これにより、VSS デュアルアクティブの手動復旧が強制的に実行されます。VSL バンドルが復旧されると、次の syslog メッセージが以前アクティブだったスイッチのコンソール出力だけに表示されます。

```
11:02:05.814: %DUAL_ACTIVE-1-VSL_RECOVERED: VSL has recovered during dual-active
situation: Reloading switch 1
11:02:05.814: %VS_GENERIC-5-VS_CONFIG_DIRTY: Configuration has changed. Ignored reload
request until configuration is saved
11:02:06.790: %VSLP-SW1_SP-5-RRP_MSG: Role change from Active to Standby and hence need to
reload
```

```
11:02:06.790: %VSLP-SW1_SP-5-RRP_UNSAVED_CONFIG: Ignoring system reload since there are
unsaved configurations. Please save the relevant configurations
11:02:06.790: %VSLP-SW1_SP-5-RRP_MSG: Use 'reload' to bring this switch to its preferred
STANDBY role
```

VSS が SSO モードで動作するためには、両方のシャーシでコンフィギュレーションがまったく同じであることが必要です。互換性を保証するためのコンフィギュレーションの確認は、VSL リンクが復旧されるとすぐに実行されます。何らかの変更があると（コンフィギュレーション モードを開始しただけでも）、コンフィギュレーション ステータスを確認するために使用するフラグが dirty とマークされます。これは、コンフィギュレーションの不一致を意味します。この不一致が発生すると、次のいずれかが必要になります。

- 両方のシャーシのコンフィギュレーションが一致するように、一致しないファイルを修正し、変更を反映する。
- コンフィギュレーションを NVRAM に保存してフラグをクリアする。

デュアルアクティブ イベント時には 2 種類のコンフィギュレーション変更が考えられます。

- 「VSL リンク以外のコンフィギュレーションの変更」（P.4-37）
- 「VSL リンク関連のコンフィギュレーションの変更」（P.4-37）

これら 2 種類のコンフィギュレーションの変更については、次のセクションで説明します。それぞれ適切な対処が必要です。

（注）　コンフィギュレーションの変更に対するシステムの動作は、搭載されている Cisco IOS のバージョンに依存します。次の動作説明は、Cisco IOS リリース 12.2(33)SXH だけに適用されます。

VSL リンク以外のコンフィギュレーションの変更

VSL バンドルに影響しないコンフィギュレーションの変更に対しては、これらの変更を適用するシャーシを決定する必要があります。変更が以前のアクティブ スイッチに対するものである場合、コンフィギュレーションを変更して手動でスイッチをリブートすることにより、スイッチがホットスタンバイ モードで復旧されます。VSL リンクを復旧する前に、以前のアクティブ スイッチで変更を保存した場合は、コンフィギュレーションの保存により dirty ステータス フラグがクリアされるため、以前のアクティブ スイッチの手動リブートは不要です。アクティブ スイッチを変更した場合、それらの変更は、デュアルアクティブ復旧アクティビティに影響しません。復旧後（VSL リンクが復旧された後）、ピア スイッチ（以前のアクティブ スイッチ）がホットスタンバイ スイッチになったときに、新たなアクティブ スイッチのコンフィギュレーションを使用して、そのコンフィギュレーションが上書きされます。アクティブ スイッチに対して行った変更は、以前のアクティブ スイッチのコンフィギュレーションに一致する必要はありません。これは、以前のアクティブ スイッチ（新たなホットスタンバイ スイッチ）のコンフィギュレーションが上書きされるためです。

VSL リンク関連のコンフィギュレーションの変更

デュアルアクティブ状態は、次のようなさまざまなイベントによって引き起こされます。

- ユーザによって引き起こされる VSL ポートチャネルの偶発的なシャットダウン
- EtherChannel に対する変更による、すべてのリンクまたは最後の動作可能な VSL リンクの切断

VSL ポートチャネルに対する変更が行われたとき、デュアルアクティブ イベントが起動される前に、変更内容は両方のシャーシに保存されます。VSL 関連のコンフィギュレーション不一致を復旧する唯一の方法は、コンフィギュレーション モードを開始して、目的のコンフィギュレーションを一致させることです。VSL リンクに関連するコンフィギュレーションを一致させず、以前のアクティブ シャーシをリブートすると、シャーシは Route Processor Redundancy（RPR）モードで起動します。以前の

アクティブ スイッチの復旧中（手動リブートの場合）に限り、VSL コンフィギュレーションの不一致に関する syslog 出力が新たなアクティブ スイッチ上に表示されます。次の syslog 出力例は、この不一致出力を示します。

```
Aug 28 11:11:06.421: %VS_PARSE-3-CONFIG_MISMATCH: RUNNING-CONFIG
Aug 28 11:11:06.421: %VS_PARSE-3-CONFIG_MISMATCH: Please use 'show switch virtual
redundancy config-mismatch' for details
Aug 28 11:11:06.421: %VS_PARSE-SW2_SP-3-CONFIG_MISMATCH: VS configuration check failed
Aug 28 11:11:06.429: %PFREDUN-SW2_SP-6-ACTIVE: Standby initializing for RPR mode
Aug 28 11:11:06.977: %PFINIT-SW2_SP-5-CONFIG_SYNC: Sync'ing the startup configuration to
the standby Router.
6500-VSS#show switch virtual redundancy | inc Opera
     Operating Redundancy Mode = RPR
```

VSL 関連のコンフィギュレーション変更は、**show switch virtual redundancy config-mismatch** コマンドで表示されます。次に出力例を示します。

```
6500-VSS# show switch virtual redundancy config-mismatch

Mismatch Running Config:
Mismatch in config file between local Switch 2 and peer Switch 1:
ACTIVE  : Interface TenGigabitEthernet1/5/4 shutdown
STANDBY : Interface TenGigabitEthernet1/5/4 not shut
In dual-active recovery mode: No
```

RPR モードでは、VSL がイネーブルになっているラインカード以外のすべてのラインカードがディセーブルになります。アクティブ スイッチでコンフィギュレーションを修正したら、**write memory** コマンドを実行すると、スタートアップ コンフィギュレーションが RPR スイッチ スーパーバイザに書き込まれます。**redundancy reload peer** コマンドを実行すると、スイッチがリブートされ、RPR モードからホットスタンバイ モードに切り替わります。この処理の設定例を次に示します。

```
6500-VSS# conf t
6500-VSS(config)# int te1/5/4
6500-VSS(config-if)# no shut
6500-VSS(config-if)# end
6500-VSS# wr mem

Aug 28 11:17:30.583: %PFINIT-SW2_SP-5-CONFIG_SYNC: Sync'ing the startup configuration to
the standby Router. [OK]
6500-VSS# redundancy reload peer
Reload peer [confirm] y
Preparing to reload peer
```

コンフィギュレーションの修正が VSL リンクの復旧前に同期されない場合、VSL のコンフィギュレーションを変更すると、広範囲の停止が発生します。VSL コンフィギュレーションの不一致が発生したかどうかは、VSL リンクの復旧後に以前アクティブだったスイッチのブート アップ後にしか判定できないためです。つまり、スイッチは 2 回リブートされることになります。最初に不一致を検出するためにリブートされ、2 回目のブートは、ホットスタンバイの役割を担うために修正されたコンフィギュレーションを使用するために必要になります。複数回のリブートを回避するには、VSL リンクが復旧される *前に* VSL コンフィギュレーションの不一致を確認します。VSL コンフィギュレーションの変更に関しては特に注意してください。

ヒント　**ベスト プラクティスの推奨事項は、VSS 環境でデュアルアクティブ イベントが発生している間は、コンフィギュレーション モードを開始 *しない*ことです。しかし、VSL リンクの偶発的なシャットダウンに対して必要なコンフィギュレーションの変更や、VSL を適切に復旧するために必要なコンフィギュレーション変更は回避することができません。**

デュアルアクティブ状態によるコンバージェンスとユーザ データ トラフィックへの影響

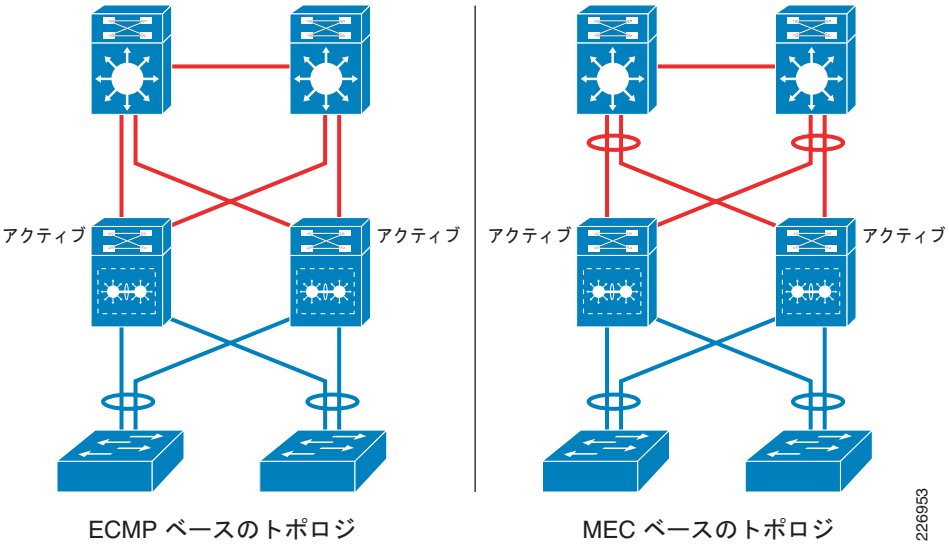
ここでは、デュアルアクティブ状態のアプリケーションおよびユーザ データ トラフィックに対する影響について説明します。検出方法ごとにデュアルアクティブ状態の検出にかかる時間が異なる点に注意してください。ただし、この問題のユーザ データ トラフィックが復旧される速度に対する影響は、多数の要因に依存します。これらは、後述するコンバージェンス要因の一覧で要約します。非常に細かなイベントの詳細（デュアルアクティブ時）と、次のコンバージェンス要因との相互作用は、本設計ガイドの範囲を超えています。全般的に、重要なことは、観察されたコンバージェンス データに基づく特定の環境での検出方法の選択です。

ユーザ トラフィック コンバージェンスは、次の要因に依存します。

- デュアルアクティブ検出方法：拡張 PAgP、fast-hello、BFD
- レイヤ 3 コアと VSS の間で設定されているルーティング プロトコル：EIGRP または OSPF
- VSS とレイヤ 3 コアの間で使用されるトポロジ：ECMP または MEC
- SSO 復旧
- NSF 復旧

図 4-21 に、デュアルアクティブ イベントとデータ トラフィックに関連付けられているコンバージェンスを使用した、すべての所見を測定した検証トポロジを示します。すべてレイヤ 2、レイヤ 3、エンドツーエンド VSS など、トポロジのさまざまな組み合わせにおいて、より良いコンバージェンスまたは最悪のコンバージェンスを実現することが完全に可能です。しかし、コンバージェンスに影響を与える一般的な原則は同じです。

図 4-21　　MEC トポロジと ECMP トポロジの比較



ECMP ベースのトポロジには、ルーティング プロトコルの隣接関係が 4 つありますが、MEC には 2 つあります。ECMP ベースのトポロジでは、VSS は、ホットスタンバイが接続されたリンク経由での hello の送信を含め、個別の hello を各リンク上で送信します。つまり、アクティブ スイッチは、ホットスタンバイを経由して接続されたリンクによって送信されるべき hello を、VSL リンク上で送信することを意味します。コア デバイスによって送信された hello は、VSS と同じパスをたどります。MEC ベースのトポロジでは、VSS から送信された hello は、常にアクティブ スイッチのローカル リンクから送信されます。しかし、コア デバイスから送信される hello では、ハッシュ結果に基づいて、ホット

スタンバイに接続されたリンクが選択されることがあります。コア デバイスからのこのリンク選択動作は、ルーティング アップデート パケットに対しても繰り返されます。その結果、ECMP トポロジと MEC トポロジの動作は、ルーティング プロトコルの隣接関係と NSF 手順において異なります。その結果、データ トラフィックのコンバージェンスがどれだけ高速に実行できるかにおいて主要な役割を果たします。

ある検出方法でデュアルアクティブ イベント時に発生するイベントの一般的なシーケンスを、背景の参照のために次に示します。これは、完全な処理の定義を構成するものではありません。

1. 最後の VSL リンクがディセーブルになります。
2. 現在アクティブなスイッチは、VSL がディセーブルになったのか、リモート ピアがリブートされたのかを知りません。現在アクティブなスイッチは、ピア スイッチと関連するインターフェイスが失われたと見なし、これを OIR イベントとして扱います。その結果、ホットスタンバイ スイッチ インターフェイスが down/down 状態になりますが、リモート スイッチ（現在のホットスタンバイ スイッチ）はまだ動作しています。その間、以前のアクティブなスイッチに接続されたローカル インターフェイスは動作可能なままとなり、コントロール トラフィックとデータ トラフィックの転送を続けます。アクティブ スイッチに接続されたインターフェイスは、このイベント中に観察されたリモート スイッチ（ホットスタンバイ）インターフェイス ステータスに関するルーティング アップデートを通知することがあります。
3. すべての VSL リンクがディセーブルになった結果、ホットスタンバイ スイッチは、リモート スイッチ（以前のアクティブなスイッチ）がリブートしたかアクティブなままかを知らずに、アクティブに遷移します。この場合、以前アクティブだったスイッチが再起動されていないため、状況はデュアルアクティブ状態として扱われます。
4. 新しいアクティブなスイッチは SSO 対応のコントロール プロトコルを初期化し、ローカル シャーシに関連付けられていたインターフェイスを取得します（これらの各インターフェイスのライン プロトコル ステータスは、SSO 復旧のために動作不能になりません）。
5. 新しいアクティブなスーパーバイザがルーティング プロトコルを再起動し、隣接ルータが NSF 対応であれば、NSF 復旧処理を実行します。そうでない場合は、ルーティング プロトコルの隣接関係の再起動が新たに開始されます （「VSS を使ったルーティング」（P.3-46）を参照してください）。
6. VSS SSO 復旧は、スタンドアロンのデュアルスーパーバイザ構成と同様に、コントロール プレーンとデータ プレーンの分離という同じ処理に従います。その結果、両方のスイッチのフォワーディング プレーンが動作可能なままになり、ユーザ トラフィックは両方のスイッチのハードウェアでスイッチングされます。このフォワーディングは、コントロール プレーンが復旧するまで続けられます。コントロール プレーンの復旧は、両方のアクティブなスイッチで実行されます。以前のアクティブなスイッチは、単にルーティング プロトコルの hello と、リモート スイッチ インターフェイス レイヤ 3 ステータスに関する一部のアップデートの送信を続けます。新たなアクティブ スイッチは、ルーティング プロトコルを再起動し、レイヤ 3 コア デバイスの隣接関係を命令しようと試みます。これらの並行するコントロール プレーン アクティビティにより、隣接関係がリセットされ、フォワーディング パスが変化し（ルーティング プロトコルとトポロジの使用に依存）、デュアルアクティブ検出が以前のアクティブ スイッチ インターフェイスのシャットダウンを引き起こす可能性があります。

実装されている場合、デュアルアクティブ検出方法により、検出が行われ、以前アクティブだったインターフェイスのシャットダウンが開始される速さが決まります。拡張 PAgP と fast-hello では、2 ～ 3 秒かかりますが、BFD では 22 ～ 25 秒かかります。しかし、検出方法だけでは、ユーザ データ トラフィック フローのコンバージェンスに影響を与えません。高速な検出方法を採用しても、ユーザ データ トラフィックに対してよい影響があったり、その逆であったりします（低速な検出方法の方がユーザ データのコンバージェンスが高速になる場合もあります）。

デュアルアクティブ イベント時のコンバージェンスの検証とトラフィック フローの特徴については、以降のセクションで、採用されているルーティング プロトコルごとに説明します。ECMP および BFD ベースの環境について一般的に説明しますが、各ルーティング プロトコルに対するデュアルアクティブ状態固有のイベントを示す、以降の詳細な説明では、MEC ベースのトポロジだけを使用します。

ヒント　デフォルトの hello タイマーとホールド タイマーを実行するには、ルーティング プロトコルを使用することをお勧めします。

EIGRP を使用したデュアルアクティブ イベントからのコンバージェンス

拡張 PAgP と fast-hello 検出

ECMP および MEC ベースのコアへの接続では、ユーザ トラフィックのコンバージェンスが 1 秒未満になります。

BFD

ECMP ベースのコア BFD 設計は、拡張 PAgP と同じコンバージェンス特性になります。MEC ベースのコアを使用した復旧はより複雑です。MEC コア設計では損失が大きくなります。EIGRP 隣接関係が不安定で、アップストリームとダウンストリームのコンバージェンスに影響するルートが削除されるためです。図 4-22 を参照してください。

図 4-22　EIGRP を使用した VSS デュアルアクティブ コンバージェンス

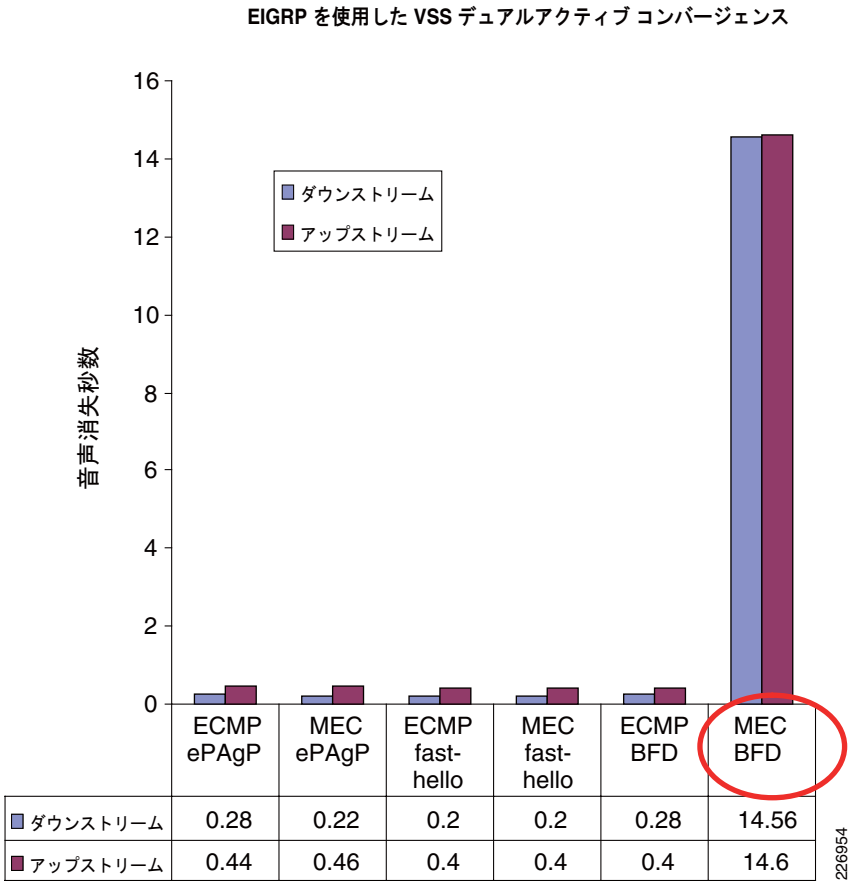


表 4-13 に、各組み合わせに対するコンバージェンスの損失と根本原因を示します。

表 4-13　コンバージェンス復旧損失と原因（EIGRP 環境）

デュアルアクティブ検出プロトコル	コアと VSS の接続	エンドツーエンド コンバージェンス	デュアルアクティブ イベント時の復旧と損失処理の要約
拡張 PAgP または fast-hello	ECMP	アップストリームおよびダウンストリーム：200 ～ 400 ミリ秒	MEC は、以前のアクティブなスイッチの観点からは 1 つのリンクを失いますが、EC リンク変更の際に、ルート取り消し通知が送信または削除されません。新しいアクティブ スイッチが隣接関係を通知する前に、以前のアクティブなスイッチがシャットダウンされます（2 ～ 3 秒）。新しいアクティブ スイッチが NSF 隣接の再起動を通知する間（7 ～ 12 秒）、再起動する NSF ピアからルートが取り消されないため、クリーンな NSF 再起動が行われます。復旧に関連して発生する唯一の損失は、以前アクティブだったスイッチによるインターフェイスのダウンです。
	MEC	アップストリームおよびダウンストリーム：200 ～ 400 ミリ秒	MEC は、以前のアクティブなスイッチの観点からは 1 つのリンクを失いますが、EC リンク変更の際に、ルートの取り消しが送信または削除されません。新しいアクティブなスイッチが隣接関係を通知する前に、以前のアクティブなスイッチがシャットダウンされます（2 ～ 1/2 秒）。クリーンな NSF が再起動されます（新しいアクティブなスイッチが NSF 隣接関係の再起動を通知する間（7 ～ 12 秒）、NSF 再起動ピアからルートが取り消されないため）
BFD	ECMP	アップストリームおよびダウンストリーム：200 ～ 400 ミリ秒	BFD 検出では、検出と以前のアクティブなスイッチのシャットダウンを完了するのにより長い時間がかかりますが、拡張 PAgP（または fast-hello）を使用した ECMP と同じ動作と結果になります。シャットダウンが開始されるまで、以前のアクティブ スイッチへのトラフィックはハードウェアで転送され続けます。以前のアクティブ スイッチは、ピアに接続されているすべてのインターフェイスを削除し、コアへの無限のメトリック値を使用してアップデートを送信します。ただし、新しいアクティブなスイッチ インターフェイスは動作可能です。コア ルータからはルートは取り消されません。これは、EIGRP が、インターフェイスを通じて問い合せられたルートに関する明示的な通知が発生するまで、ローカル トポロジ計算を実行しないためです。その間、新しいアクティブなスイッチは NSF 再起動を実行し、隣接とルートアップデートを更新します。デュアルアクティブ イベント中は、ルートまたは隣接リセットは発生しません。
	MEC	アップストリームおよびダウンストリーム：4 ～ 14 秒	EIGRP のネイバー ルータとの隣接関係の一部は、VSS に向かう IP とレイヤ 3 MEC メンバ リンク ハッシュに基づいて、アクティブ ルータの 1 つに落ち着く可能性があるため、データ損失は大きくなります。EIGRP アップデート メッセージと hello メッセージは、コア ルータとは異なるリンクにハッシュされ、隣接関係とルートの損失が発生します。EIGRP 隣接関係のいずれかが、以前のアクティブ スイッチまたは新しいアクティブ スイッチに落ち着きます。以前のアクティブ スイッチに落ち着いた場合、トラフィックの損失はより顕著になります。以前のアクティブ スイッチがシャットダウンした後、新しいアクティブ スイッチとの EIGRP 隣接関係を再確立する必要があるためです。その結果、コンバージェンスを完了するのに要する時間は変化します。

MEC ベースのトポロジを使用した BFD 検出の詳細

表 4-13 に示すように、組み合わせによってより不安定になります。検出方法は完了までに長い時間がかかり、隣接の不安定化により多くのトラフィックが中断されるためです。この不安定化の深刻さは、送信元 IP アドレスと宛先 IP アドレスのハッシュ結果と、EIGRP の hello（マルチキャスト）およびアップデート（ユニキャスト）の送信によって変わります。これらは、以前のアクティブなスイッチまたは新しいアクティブなスイッチに接続された MEC メンバ リンク上で送出されます。通常のトポロジにおける EIGRP パケット フォワーディング パスでは、VSS トポロジ内の次の組み合わせの 1 つを採用できます。

- 以前のアクティブ スイッチ上のマルチキャストと、新しいアクティブ スイッチ上のユニキャスト
- 新しいアクティブ スイッチ上のマルチキャストと、以前のアクティブ スイッチ上のユニキャスト
- 以前のアクティブ スイッチ上でのマルチキャストとユニキャスト
- 新しいアクティブ スイッチ上でのマルチキャストとユニキャスト

上記の組み合わせを考慮すると、デュアルアクティブ状態の際に、次のイベントによって EIGRP 隣接関係がリセットされます。

- コアからの hello が受信されず、ホールド タイマーがタイムアウトするような、いずれかのアクティブなスイッチ上での EIGRP 隣接関係設定。これが発生するのは、通常の動作状態では、パケットを VSL リンク経由で転送していた、ホットスタンバイ スイッチに hello を送信することになるハッシュ計算が、そうならなくなるためです（VSL リンクがダウンしているため）。その結果、以前のアクティブなスイッチは hello パケットを受信せず、隣接がタイムアウトします。これにより、VSS に接続されたアクセス レイヤに関係するコア ルータでルートが損失し、すべてのアップストリーム接続で VSS 上のルートが損失します。
- リモート ルータが新しいアクティブ スイッチからの NSF 再起動 hello に応答しなかったため、NSF 信号タイマーがタイムアウトした。これは、リモート ルータ（コア）ハッシュが hello および NSF hello-ack を以前のアクティブ スイッチに送信するために発生する可能性があります。以降、NSF タイムアウトは、新しいアクティブ スイッチによって検出され、正常な復旧が妨げられます。その結果、完全な隣接の再起動が開始されます。
- ルート アップデート処理中に、NSF 再起動処理がスタックし（たとえば、ルートのアップデートが、ユニキャスト ハッシュを使用して以前のアクティブ スイッチに送信される）、新しいアクティブ スーパーバイザが INIT 状態で隣接がスタックしたことを宣言し、完全な再起動を強制する。

デュアルアクティブ イベント後の隣接の落ち着く場所が、コンバージェンスの変動を決定します。

デュアルアクティブ状態の際に、隣接が、隣接の完全な再起動を実行する必要がない、新しいアクティブなスイッチに落ち着くように IP アドレスが設定されている場合、コンバージェンスが高速になる可能性があります。しかし、次のデュアルアクティブ状態の際には、一貫性がなくなります。隣接は以前のアクティブで落ち着き、上記の段落で説明したトリガー状態の 1 つが発生するためです。

隣接が以前のアクティブ スイッチで落ち着く場合には、22 ～ 25 秒後に、デュアルアクティブ イベントが内部シャットダウン（管理シャットダウンとは異なります）を引き起こし、これにより隣接プロセスが新しいアクティブ スイッチで再起動し、通常の隣接セットアップ（NSF の正常な再起動ではなく）が実行されます。隣接リセットにより、コアと VSS からのルートが取り消され、ダウンストリーム パスとダウンストリーム パスで以降に不定のパケット損失が発生します。

特定のネットワーク上で、部分的な症状だけが現れる可能性があります。多数の変数が関係するため、コンバージェンスによるトラフィックの中断を完全かつ一貫して特徴付けることは困難です。重要な点は、BFD 検出中にコンバージェンスに影響を与える多数の要因を適度に理解し、これらの要因が、他のどの組み合わせよりも高く、BFD 検出後のコンバージェンスの原因となることを理解することです。

OSPF を使用したデュアルアクティブ イベントからのコンバージェンス

OSPF では、本質的に、Shortest-Path-First（SPF）データベースを構築し維持するための一意の接続が必要です。これには、2 つの組み込みの検証チェックがあり、デュアルアクティブ状態でより詳しいネットワークの可視性を提供します。それは、ルータ ID とネイバー到達可能性の双方向検証です。

拡張 PAgP と fast-hello

ECMP ベースのコア設計では、高い率でのトラフィック損失が発生します。デュアルアクティブ イベント中に、OSPF によってコアでのアクセス レイヤ ルートが削除されるためです。OSPF がこれを行うのは、コア ルータによって重複するルータ ID が認識されるためです。

MEC ベースのトポロジでは、コンバージェンスははるかに短くなります。MEC ベースのコア設計は、コアにおけるルート削除の影響を受けません。これは、OSPF ルート取り消しが送信されないためです（EtherChannel インターフェイスは引き続き動作可能です）。また、デュアルアクティブの検出は、2 ～ 3 秒以内に開始され、新しいアクティブ スイッチ上でコントロール プレーンの復旧がそれに続きます。復旧中は、両方のスイッチ メンバ インターフェイスがデータ転送を続け、コンバージェンス時間が 1 秒未満になります。

BFD

ECMP ベースのコア設計では、拡張 PAgP 設計に比べて、コンバージェンスが短くなります。これは、BFD の遅延復旧動作により、少なくとも 1 つのアクセス レイヤ ルートがコアで動作可能なままになるためです。

MEC ベースのコアを使用した復旧はより複雑です。MEC コア設計では、隣接関係が不安定になり、ルートが削除されるため、トラフィック損失の割合が高くなります。これは、アップストリームとダウンストリームのコンバージェンスに影響を与えます。また、コアが隣接損失を検出し、アクセス レイヤ サブネットのルートをより早く削除するため、ダウンストリームの損失が増えます。これに対し、VSS は隣接とアップストリーム ルートを保持します。図 4-23 は、さまざまな構成オプションでのデュアルアクティブ コンバージェンスを比較したものです。

図 4-23　OSPF を使用した VSS デュアルアクティブ コンバージェンス

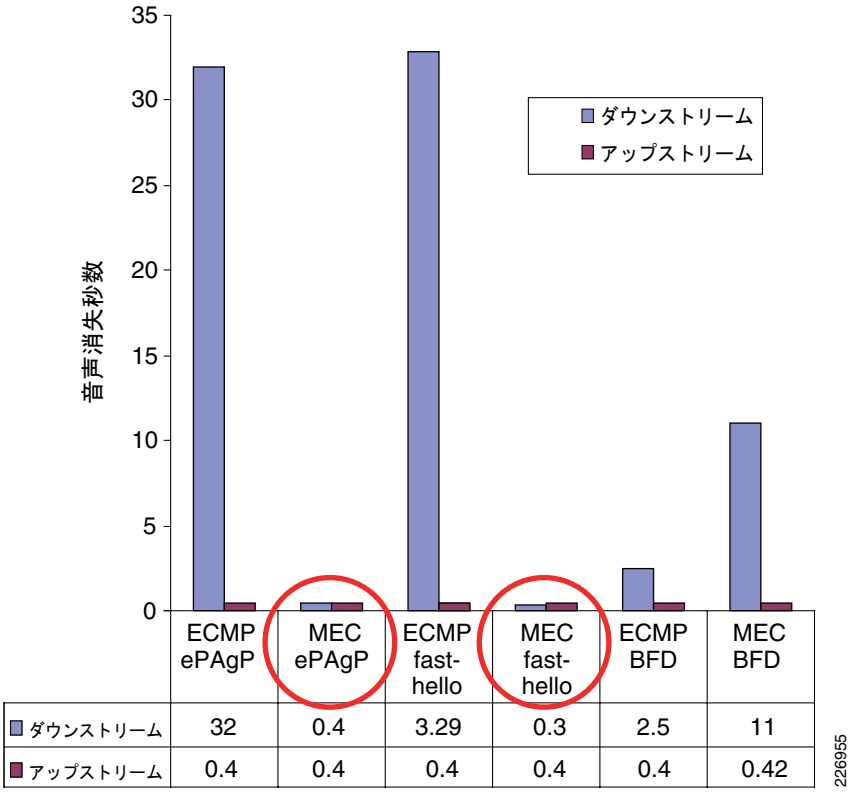


表 4-14 に、各組み合わせに対するコンバージェンスの損失と根本原因を示します。

表 4-14　コンバージェンス復旧損失と原因（OSPF 環境）

デュアルアクティブ検出プロトコル	コアと VSS の接続	エンドツーエンド コンバージェンス	デュアルアクティブ イベント時の復旧と損失処理の要約
拡張 PAgP	ECMP	ダウンストリーム：30 ～ 32 秒 アップストリーム：200 ～ 400 ミリ秒	アクセス レイヤへの 4 つのルートすべてが取り消されるため、ダウンストリーム トラフィック損失は大きくなります。VSL リンクの損失により、以前のアクティブなスイッチがホットスタンバイ スイッチ（新しいアクティブ スイッチ）上のすべてのインターフェイスをディセーブルにする際に削除します（実際には動作可能であるにもかかわらず）。このルート アップデートはコア ルータに通知され、新しいアクティブ スイッチから学習した 2 つのルートが取り消されます（動作可能であるにもかかわらず）。拡張 PAgP または fast-hello 検出は、以前のアクティブ スイッチ インターフェイスをシャットダウンし、コア ルーティング テーブルからの 2 つ目のセットのルートが取り消されます。新しいアクティブ スイッチがその NSF 再起動を完了しルートをコア ルータに送信するまで、ダウンストリーム トラフィックのブラック ホール化が発生します。VSS からは重複するルータ ID が認識されないため、アップストリーム ルートの削除は発生しません。
	MEC	ダウンストリームおよびアップストリーム：200 ～ 400 ミリ秒	EtherChannel リンクの変更中にルートは削除されません。新しいアクティブ スイッチが隣接の再起動を通知する前に、以前のアクティブなスイッチがシャットダウンされます（2.5 秒）。クリーンな NSF 再起動が行われます（新しいアクティブ スイッチが NSF 隣接の再起動を通知する前に（7 ～ 12 秒）、同じ再起動する NSF ピアに対するルートが取り消されません）。
BFD	ECMP	ダウンストリーム：2 ～ 2.5 秒 アップストリーム：200 ～ 400 ミリ秒	上記の拡張 PAgP と ECMP の場合と同様ですが、BFD では、以前のアクティブ スイッチが 22 ～ 25 秒間切断されず、アクセス レイヤ サブネット用にコアで少なくとも 1 つのルートが保持されます。そのルートを保持することで、新しいアクティブ スイッチ上で NSF 復旧処理が完了するまで、トラフィックのブラック ホール化が防止されます。
	MEC	ダウンストリーム：200 ミリ秒 ～ 11 秒 アップストリーム：200 ～ 400 ミリ秒	トラフィックは、同時に発生するいくつかのイベントによって影響を受けます。OSPF 隣接関係は、BFD が古いアクティブ インターフェイスをシャットダウンするまで安定しない可能性があります（22 ～ 25 秒）。どの VSS メンバがアクティブか、および、OSPF hello がハッシュされる場所によっては、NSF 再起動の安定性が影響を受けることがあります。NSF 再起動がクリーンに行われた場合、コンバージェンスは 1 秒未満になります。

MEC ベースのトポロジを使用した BFD 検出の詳細

前述のように、MEC を使用すると、コアから VSS への hello（マルチキャスト）メッセージとアップデート（ユニキャスト）メッセージの非対称ハッシュが、通常の動作状態とデュアルアクティブ状態の両方で可能です。VSS からコアへ、コントロール プレーンの接続はローカル インターフェイス上にとどまります。この動作の組み合わせでは、デュアルアクティブ状態で隣接のリセットが行われます。また、隣接形成のための OSPF hello プロトコルにおける双方向のネイバー可用性の検証のため、OSPF 隣接はいずれかのアクティブ VSS ルータで安定しません。

通常の動作状態では、コア ルータは VSS に対する単一のルータ ID を認識します。デュアルアクティブ中、コア ルータは同じルータ ID が 2 つのアクティブ スーパーバイザによって通知されるのを認識します。コア ルータの SPF は混乱状態となり、OSPF プロセスで詳細な隣接ロギングが有効になっていれば、重複するルータ ID を syslog に表示します。OSPF 隣接の決着に対しては、コア ルータは、実際に何が起きているかを知らずに、以前の VSS アクティブ スイッチまたは新しい VSS アクティブ スイッチからの要求に応答します。しかし、コア ルータは、ハッシュ（インターフェイス アドレスの送信元 IP と、宛先 224.0.0.5）に従って、マルチキャスト hello を以前のアクティブ スイッチまたは新しいアクティブ スイッチのいずれかに送信します。デュアルアクティブ イベント中、コアから hello が送信される、次の 2 つの可能性があります。

- 新しいアクティブ スイッチに接続されたリンク：コアが新しいアクティブ スイッチに接続されたリンクに hello を送信している間、以前のアクティブ ルータは動作し続け、通常の OSPF hello をそのローカル リンクを通じてコアに送信し続けます。同時に、新しいアクティブ ルータが隣接を確立し、RS ビットが設定された特殊な hello（NSF 再起動）を送信することで、そのコア ルータとの接続を再起動しようとします。この隣接の再起動は、NSF ビットや RS ビットが設定されていない hello が以前のアクティブなスイッチからコア ルータ経由で受信されるまで続けられる可能性があります（以前のアクティブなルータは、何が起きたのかを知らないため、動作し続けます）。これにより、コア ルータの NSF 対応の手順が混乱し、コア ルータで隣接がリセットされることがあります。その間、以前のアクティブなルータも、コアからの hello を受信せずに、タイムアウトすることが考えられます。最終的に、いずれかのアクティブ VSS スイッチ ネイバーが隣接をリセットします。隣接リセットが起動されると、コア ルータは新しいアクティブなルータとのネイバー隣接関係を確立しようとし（ハッシュにより）、FULL ADJ に到達します。その間、以前のアクティブなルータは再度 hello の送信を試みますが、以前のアクティブ ルータからの INIT hello であるため、今度はコア ルータがそれ自身の IP アドレスを受信した hello の中で認識しません。これは、コアに対し、fast-hello の送信と、新しい Database Descriptor（DBD; データベース記述子）シーケンス番号の新しいアクティブ ルータへの送信を促します（ほぼ新しいアクティブ ルータとの FULL ADJ であるため）。新しいアクティブ ルータはこれを BAD_SEQUENCE 番号と見なし、ADJ を FULL から EX-START にリセットします。
- 以前のアクティブ スイッチ上で接続されているリンク：この場合、コアが hello メッセージを以前のアクティブ スイッチに送信し、新しいアクティブ スイッチが、まず NSF 再起動で、次に INIT hello で開始するようなハッシュであることがわかります。コア ルータは以前のアクティブ ルータに応答を送信し続けるため、新しいアクティブなルータは、コア ルータから受信した応答を受信しません。その結果、最終的に、新しいアクティブ スーパーバイザによって隣接の再起動が開始されます。これは、検出手段が採用されていないか（BFD の場合）、隣接リセットにより高いパケット損失が発生する場合に、無限に続きます。

デュアルアクティブ方式の選択

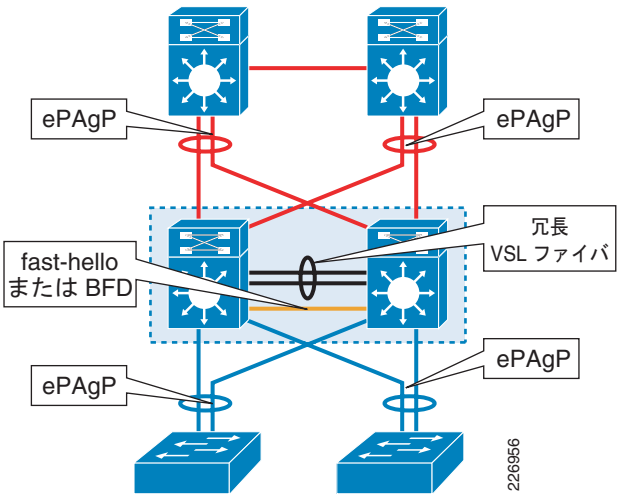
複数の方法を展開できます。複数の検出方法の展開は、復元力のある VSS リンク構成の代わりにはなりません。複数の検出方法がある場合、拡張 PAgP と fast-hello が、BFD よりも最初に検出されます。複数の検出方法がある場合、最初にデュアルアクティブを検出した方法がコンバージェンスを支配します。

複数の方法を展開する際の唯一の例外は、OSPF ルーティングが ECMP ベースのトポロジとともにイネーブルにされている場合です。このトポロジでは、推奨される唯一の検出方法は BFD です。BFD は、最善のコンバージェンスを提供する唯一の方法であるためです。他の方法を BFD とともに展開すると、BFD はデュアルアクティブを検出する最初の方法になりません（BFD は、拡張 PAgP や fast-hello と比べて検出に時間がかかります）。

拡張 PAgP 検出は、レイヤ 2 またはレイヤ 3 MEC で使用できます。拡張 PAgP 検出は、単一のネイバー上だけで実行することが必要な場合があります。しかし、すべてのインターフェイス上で拡張 PAgP を使用することで、復旧のためのパスが存在するように、ワーストケースでも、少なくとも 1 つのスイッチが同じ VSS ペアの両方のメンバに接続されます（障害状態ですべてのケーブル パスが影響

を受けない前提）。図 4-24 に、VSL リンクの可用性を確保するための複数の冗長性を備えたトポロジの概要と、デュアルアクティブ イベント状態でのトラフィックの中断を減らすのに役立つ検出ツールの配置を示します。

図 4-24　　デュアルアクティブ検出の可能性



要約と推奨事項

ここでは、該当するトポロジにおける検出方法と該当するルーティング プロトコルのさまざまな組み合わせを特徴付ける検証に基づく、推奨事項を示します。次のリストは、デュアルアクティブ検出方法の実装のための推奨事項の要約です。

- コアへのレイヤ 3 MEC トポロジでは、OSPF と EIGRP でコンバージェンスが 1 秒未満となるように、拡張 PAgP と fast-hello の検出の組み合わせを使用します。
- BFD 検出を EIGRP および OSPF とともに使用するのは避けます。VSS からのコア接続がレイヤ 3 MEC ベースの場合に復旧が複雑で可変になるためです。
- OSPF では、コンバージェンス時間が 1 秒未満になるように、可能であれば拡張 PAgP または fast-hello を使用します（どちらも MEC と対になっています）。
- 復旧が複雑にならず、トポロジの制限もなく、コンバージェンスが短くなるように、VSLP fast-hello と MEC を BFD の代わりに使用します（この推奨事項の例外は、OSPF を ECMP ベースのトポロジで実装する場合です）。

表 4-15 に、これらの推奨事項の要約を示します。「良好」は 1 秒未満の復旧を示し、「OK」は復旧が 1 秒未満でないことを示します。

表 4-15　　コンバージェンス方法ごとの復旧の比較

デュアルアクティブ	拡張 PAgP	fast-hello	BFD
EIGRP と ECMP コア	良好	良好	良好
EIGRP と MEC コア	良好	良好	OK
OSPF と ECMP コア	OK	OK	良好
OSPF と MEC コア	良好	良好	OK

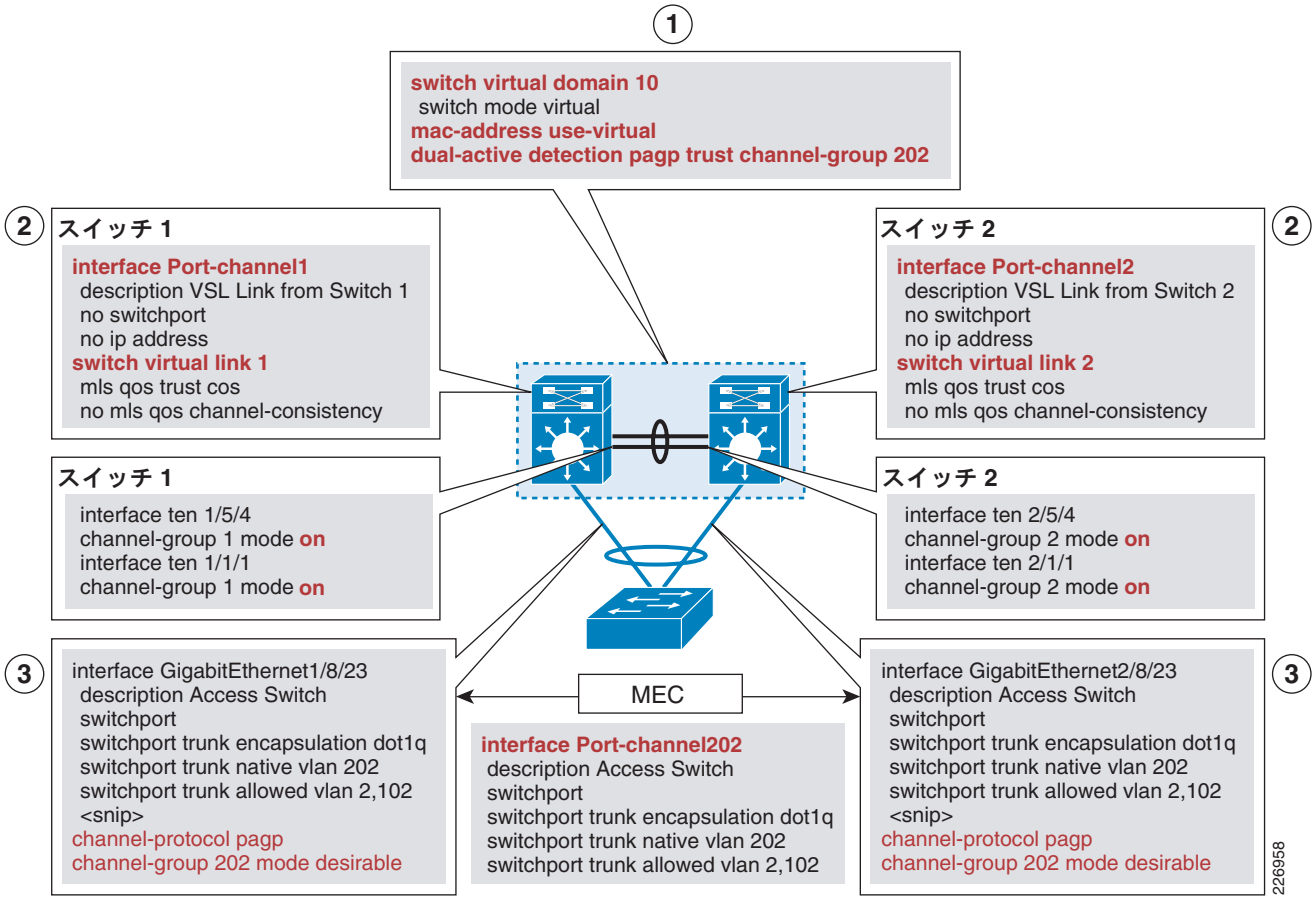
（注）　本設計ガイドにおける上記の説明および記載したその他の参考資料にあるとおり、MEC ベースのコアへの接続により、ユニキャストとマルチキャストで 1 秒未満のコンバージェンス時間が可能になります。

APPENDIX A

VSS 対応キャンパスのベスト プラクティス設定例

図 A-1 は、基本的な VSS 対応ネットワークをセットアップするために必要なベスト プラクティスの設定のベースラインを示します。丸囲みは、スタンドアロンから VSS システムを構築するために必要な重要なステップであることを示しています。VSS の設定で重要な CLI 情報は赤字で強調表示しています。コメントは*青字の斜体*で表しています。

図 A-1　　VSS 対応キャンパスのベスト プラクティス設定の概要の全体図



エンドツーエンドのデバイスの設定

エンドツーエンドのデバイスの設定は、3 種類の重要なセクションに分類されます。各セクションの設定には、ベスト プラクティスの設定の一部として必要な特定の CLI と、それに対応する説明が含まれます。

- VSS および L2 ドメイン：上記の基本設定と L2 のドメイン設定が含まれます。
- アクセス レイヤ：L2 ドメイン設定のサンプル。
- L3 ドメイン：VSS およびコア ルータ用グローバル L3 設定。この後、セクションは EIGRP と OSPF 別に、トポロジの種類（ECMP と MEC）に応じてそれぞれの設定を示します。

VSS 固有

VSS グローバル設定

```
switch virtual domain 10 ! Must configure unique domain ID
 switch mode virtual
 switch 1 priority 110 ! Not needed, helps in operational mgmt
 switch 2 priority 100 ! Not needed, helps in operational mgmt
 dual-active exclude interface GigabitEthernet1/5/3 ! Connectivity to VSS during dual
active
mac-address use-virtual ! Required for consistent MAC address
dual-active detection pagp trust channel-group 202 ! Enhanced PAgP based dual active
detection

redundancy ! Default SSO Enabled
 main-cpu
  auto-sync running-config
 mode sso
```

スイッチ 1

```
interface Port-channel1 ! Unique port-channel number for SW 1
 description VSL Link from Switch 1
 no switchport
 no ip address
switch virtual link 1 ! Defines switch ID for SW 1
 mls qos trust cos
 no mls qos channel-consistency

interface ten 1/5/4
channel-group 1 mode on ! EC mode is ON - EtherChannel Managemeent Protocol off
interface ten 1/1/1
channel-group 1 mode on
```

スイッチ 2

```
interface Port-channel2 ! Unique port-channel number for SW 1
 description VSL Link from Switch 2
 no switchport
 no ip address
switch virtual link 2 ! Defines switch ID for SW 2
 mls qos trust cos
 no mls qos channel-consistency

interface ten 2/5/4
channel-group 2 mode on ! EC mode is ON - EtherChannel Managemeent Protocoloff
```

```
interface ten 2/1/1
channel-group 2 mode on
```

レイヤ 2 ドメイン

VSS

```
udld enable
vtp domain campus-test
vtp mode transparent

spanning-tree mode rapid-pvst
no spanning-tree optimize bpdu transmission
spanning-tree extend system-id
spanning-tree vlan 2-999 priority 24576 !  STP Root

port-channel load-balance src-dst-mixed-ip-port !  Enhanced hash algorithem


vlan 400 !  VLANs spanning multiple access-layer SWs
 name L2_ Spaned_VLAN _400

vlan 450
 name L2_ Spaned_VLAN _450

vlan 500
 name L2_ Spaned_VLAN _500

vlan 550
 name L2_ Spaned_VLAN _550

vlan 600
 name L2_ Spaned_VLAN _600

vlan 650
 name L2_Spaned_VLAN_650

vlan 900
 name NetMgmt_VLAN_900

vlan 999
 name Unused_Port_VLAN_999

vlan 2
 name cr7-3750-Stack-Data-VLAN
!
vlan 102
 name cr7-3750-Stack-Voice-VLAN

interface Vlan2 ! Sample VLAN interface configuration
 ip address 10.120.2.1 255.255.255.0
 no ip redirects
 no ip unreachables
 ip flow ingress
 ip pim sparse-mode
 logging event link-status
 hold-queue 150 in
 hold-queue 150 out
!
```

VSS：Multi-Chassis EtherChannel

PAgP

```
interface GigabitEthernet1/8/23  !  Interface on SW 1
 description Access Switch Facing Interface
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 202
switchport mode dynamic desirable ! Trunk mod dynamic and desirable
 switchport trunk allowed vlan 2,102,400,450,500,550,600,650,900 !  Only allow need VLANs
for a given trunk
logging event link-status !  Logging for link status
logging event trunk-status !  Logging for trunk status
logging event bundle-status !  Logging for port-channel status
 load-interval 30
 mls qos trust dscp
channel-protocol pagp
channel-group 202 mode desirable !  Define Port-channel, PAgP mode desirable


interface GigabitEthernet2/8/23  !  Interface on SW 2
 description Access Switch Facing Interface
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 202
switchport mode dynamic desirable
 switchport trunk allowed vlan 2,102,400,450,500,550,600,650,900
 logging event link-status
 logging event trunk-status
 logging event bundle-status
 load-interval 30
 mls qos trust dscp
 load-interval 30
channel-protocol pagp
channel-group 202 mode desirable

interface Port-channel202  !  Automatically created by defining at interfaces
 description Access Switch MEC
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 202
 switchport trunk allowed vlan 2,102,400,450,500,550,600,650,900
logging event link-status
logging event spanning-tree status ! STP logging enabled on port-channel
 load-interval 30
 mls qos trust dscp
 spanning-tree portfast ! Optional - helps during initialization
 hold-queue 2000 out
```

LACP

LACP サンプル設定

```
interface GigabitEthernet1/8/23
 description Access Switch Facing Interface
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 202
switchport mode dynamic desirable
 switchport trunk allowed vlan 2,102,400,450,500,550,600,650,900
logging event link-status
```

```
logging event trunk-status
logging event bundle-status
 load-interval 30
 mls qos trust dscp
 channel-protocol lacp
 channel-group 202 mode active
 hold-queue 2000 out

interface GigabitEthernet2/8/23
 description Access Switch Facing Interface
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 202
switchport mode dynamic desirable
 switchport trunk allowed vlan 2,102,400,450,500,550,600,650,900
logging event link-status
logging event trunk-status
logging event bundle-status
 load-interval 30
 mls qos trust dscp
 channel-protocol lacp
 channel-group 202 mode active
 hold-queue 2000 out

interface Port-channel202 !  Automatically created by defining at interfaces
 description Access Switch MEC
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 202
 switchport trunk allowed vlan 2,102,400,450,500,550,600,650,900
logging event link-status
logging event spanning-tree status
 load-interval 30
 mls qos trust dscp
 spanning-tree portfast ! Optional - helps during initialization
 hold-queue 2000 out
```

アクセス レイヤ スイッチ

サンプル設定（設定はプラットフォームに応じて異なる）

```
interface GigabitEthernet0/27
 description Uplink to VSS Switch Gig 1/8/24
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 203
switchport mode dynamic desirable
 switchport trunk allowed vlan 3,103,400,450,500,550,600,650,900
 logging event link-status
 logging event trunk-status
 logging event bundle-status
 carrier-delay msec 0
 srr-queue bandwidth share 1 70 25 5
 srr-queue bandwidth shape  3  0  0  0
 priority-queue out
 mls qos trust dscp
channel-protocol pagp
channel-group 1 mode desirable

interface GigabitEthernet0/28
 description Uplink to VSS Switch Gig 2/8/24
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 203
```

```
 switchport trunk allowed vlan 3,103,400,450,500,550,600,650,900
 switchport mode dynamic desirable
 logging event link-status
 logging event trunk-status
 logging event bundle-status
 carrier-delay msec 0
 srr-queue bandwidth share 1 70 25 5
 srr-queue bandwidth shape  3  0  0  0
 priority-queue out
 mls qos trust dscp
channel-protocol pagp
channel-group 1 mode desirable


interface Port-channel1 !  Automatically created by defining at interfaces
 description EC Uplink to VSS
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 203
 switchport trunk allowed vlan 3,103,400,450,500,550,600,650,900
 switchport mode dynamic desirable
logging event link-status
logging event spanning-tree status
 carrier-delay msec 0
 spanning-tree portfast ! Optional - helps during initialization
```

レイヤ 3 ドメイン

グローバル設定

```
mls ip cef load-sharing <option> !  Apply Campus Best Practices
```

マルチキャスト

VSS

```
ip multicast-routing
ip pim rp-address 10.122.100.1 GOOD-IPMC override !  RP mapping with filter

ip access-list standard GOOD-IPMC
 permit 224.0.1.39
 permit 224.0.1.40
 permit 239.192.240.0 0.0.3.255
 permit 239.192.248.0 0.0.3.255
```

コア 1

コア RP ANYCAST：プライマリ

```
ip multicast-routing

interface Loopback0
 description MSDP PEER INT !  MSDP Loopback
 ip address 10.122.10.1 255.255.255.255

interface Loopback1
 description ANYCAST RP ADDRESS (PRIMARY) !  Anycast RP Primary
```

```
 ip address 10.122.100.1 255.255.255.255

interface Loopback2
 description Garbage-CAN RP
 ip address 2.2.2.2 255.255.255.255

interface Port-channel1 !  Core 1- Core2 L3 for MSDP
 description Channel to Peer Core Node
 dampening
 ip address 10.122.0.18 255.255.255.254
 ip pim sparse-mode
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp

ip access-list standard GOOD-IPMC
 permit 224.0.1.39
 permit 224.0.1.40
 permit 239.192.240.0 0.0.3.255
 permit 239.192.248.0 0.0.3.255

ip msdp peer 10.122.10.2 connect -source Loopback0 ! MSDP Configuration
ip msdp description 10.122.10.2 ANYCAST-PEER-6k-core-2
ip msdp cache -sa-state
ip msdp originator-id Loopback0
```

コア 2

```
ip multicast-routing

interface Loopback0
 description MSDP PEER INT
 ip address 10.122.10.2 255.255.255.255

interface Loopback1
 description ANYCAST RP ADDRESS
 ip address 10.122.100.1 255.255.255.255 ! Secondary ANYCAST RP
 delay 600

interface Loopback2
 description Garbage-CAN RP
 ip address 2.2.2.2 255.255.255.255

interface Port-channel1
 description Channel to Peer Core node
 dampening
 ip address 10.122.0.19 255.255.255.254
 ip pim sparse-mode
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp

ip pim rp-address 10.122.100.1 GOOD-IPMC override
ip access-list standard GOOD-IPMC
 permit 224.0.1.39
 permit 224.0.1.40
 permit 239.192.240.0 0.0.3.255
 permit 239.192.248.0 0.0.3.255

ip msdp peer 10.122.10.1 connect-source Loopback0
ip msdp description 10.122.10.1 ANYCAST-PEER-6k-core-1
ip msdp cache-sa-state
```

```
ip msdp originator-id Loopback0
```

EIGRP MEC

VSS

```
router eigrp 100
 passive-interface default
 no passive-interface Port-channel200
 no passive-interface Port-channel201
 network 10.0.0.0
eigrp log-neighbor-warnings
eigrp log-neighbor-changes
 no auto-summary
 eigrp router-id 10.122.102.1
 eigrp event-log-size 3000
 nsf !  Enable NSF Capability

interface Port-channel200 !  Create L3 MEC Interface first
 description 20 Gig MEC to CORE-1 (cr2-6500-1 4/1-4/3)
 no switchport
 dampening
 ip address 10.122.0.26 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip summary-address eigrp 100 10.125.0.0 255.255.0.0 5 !  Summarization for Access-subnets
 ip summary-address eigrp 100 10.120.0.0 255.255.0.0 5
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!
interface Port-channel201
 description 20 Gig to CORE-2 (cr2-6500-1 4/1-4/3)
 no switchport
 dampening
 ip address 10.122.0.21 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip summary-address eigrp 100 10.125.0.0 255.255.0.0 5
 ip summary-address eigrp 100 10.120.0.0 255.255.0.0 5
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out

interface TenGigabitEthernet1/2/1
 description 10 GigE to Core 1
 no switchport
 no ip address
 logging event link-status
 logging event bundle-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 channel-protocol pagp
 channel-group 200 mode desirable
```

```
 hold-queue 2000 in
 hold-queue 2000 out
!
interface TenGigabitEthernet1/2/2
 description 10 GigE to Core 2
 no switchport
 no ip address
 logging event link-status
 logging event bundle-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 channel-protocol pagp
 channel-group 201 mode desirable
 hold-queue 2000 in
 hold-queue 2000 out

interface TenGigabitEthernet2/2/1
 description to core 1
 no switchport
 no ip address
 logging event link-status
 logging event bundle-status
 logging event spanning-tree status
 load-interval 30
 mls qos trust dscp
 channel-protocol pagp
 channel-group 200 mode desirable
 hold-queue 2000 in
 hold-queue 2000 out

interface TenGigabitEthernet2/2/2
 description 10 GigE to Core 2
 no switchport
 no ip address
 logging event link-status
 logging event bundle-status
 load-interval 30
 mls qos trust dscp
 channel-protocol pagp
 channel-group 201 mode desirable
 hold-queue 2000 in
 hold-queue 2000 out
```

コア 1

```
router eigrp 100
 passive-interface default
 no passive-interface Port-channel1
 no passive-interface Port-channel20
 no passive-interface Port-channel221
 network 10.0.0.0
 no auto-summary
eigrp log-neighbor-warnings
eigrp log-neighbor-changes
eigrp event-log-size 3000

interface Port-channel20
 description 20 Gig MEC to VSS 1/2/1 2/2/1
 dampening
 ip address 10.122.0.27 255.255.255.254
 ip flow ingress
```

```
 ip pim sparse-mode
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

コア 2

```
router eigrp 100
 passive-interface default
 no passive-interface Port-channel1
 no passive-interface Port-channel20
 no passive-interface Port-channel221
 network 10.0.0.0
 no auto-summary
eigrp log-neighbor-warnings
eigrp log-neighbor-changes
eigrp event-log-size 3000

interface Port-channel21
 description 20 Gig to VSS 1/2/2-2/2/2
 dampening
 ip address 10.122.0.20 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

EIGRP ECMP

VSS

```
router eigrp 100
 passive-interface default
 no passive-interface TenGigabitEthernet1/2/1
 no passive-interface TenGigabitEthernet1/2/2
 no passive-interface TenGigabitEthernet2/2/1
 no passive-interface TenGigabitEthernet2/2/2
 network 10.0.0.0
 no auto-summary
 eigrp router-id 10.122.102.1
 eigrp log-neighbor-warnings
 eigrp log-neighbor-changes
 eigrp event-log-size 3000
 nsf !  Enable NSF Capability

interface TenGigabitEthernet1/2/1
 description 10 GigE to Core 1
 no switchport
 dampening
 ip address 10.122.0.26 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
```

```
 ip summary-address eigrp 100 10.120.0.0 255.255.0.0 5
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!
interface TenGigabitEthernet1/2/2
 description 10 GigE to Core 2
 no switchport
 dampening
 ip address 10.122.0.23 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip summary-address eigrp 100 10.120.0.0 255.255.0.0 5
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out

interface TenGigabitEthernet2/2/1
 description to Core 1
 no switchport
 dampening
 ip address 10.122.0.32 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip summary-address eigrp 100 10.120.0.0 255.255.0.0 5
 logging event link-status
 load-interval 30
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!
interface TenGigabitEthernet2/2/2
 description 10 GigE to Core 2
 no switchport
 dampening
 ip address 10.122.0.20 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip summary-address eigrp 100 10.120.0.0 255.255.0.0 5
 logging event link-status
 load-interval 30
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

コア 1

```
router eigrp 100
 passive-interface default
 no passive-interface TenGigabitEthernet4/1
 no passive-interface TenGigabitEthernet4/3
 network 10.0.0.0
 no auto-summary
eigrp log-neighbor-warnings
eigrp log-neighbor-changes
eigrp event-log-size 3000
```

```
interface TenGigabitEthernet4/1
 description To VSS Ten1/2/1
 dampening
 ip address 10.122.0.27 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out

interface TenGigabitEthernet4/3
 description To VSS Ten2/2/1
 dampening
 ip address 10.122.0.33 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 logging event link-status
 logging event bundle-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

コア 2

```
router eigrp 100
 passive-interface default
 no passive-interface TenGigabitEthernet4/1
 no passive-interface TenGigabitEthernet4/3
 network 10.0.0.0
 no auto-summary
eigrp log-neighbor-warnings
eigrp log-neighbor-changes
eigrp event-log-size 3000

interface TenGigabitEthernet4/1
 description To VSS Ten 1/2/2
 dampening
 ip address 10.122.0.22 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out

interface TenGigabitEthernet4/3
 description To VSS Ten 2/2/2
 dampening
 ip address 10.122.0.21 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 logging event link-status
 load-interval 30
 carrier-delay msec 0
```

```
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

OSPF MEC

VSS

```
router ospf 100
 router-id 10.122.0.235
 log-adjacency-changes detail
 auto-cost reference-bandwidth 20000 !  Optional
 nsf !  Enable NSF Capability
 area 120 stub no-summary
 area 120 range 10.120.0.0 255.255.0.0 cost 10
 area 120 range 10.125.0.0 255.255.0.0 cost 10
 passive-interface default
 no passive-interface Port-channel200
 no passive-interface Port-channel201
 network 10.120.0.0 0.0.255.255 area 120
 network 10.122.0.0 0.0.255.255 area 0
 network 10.125.0.0 0.0.255.255 area 120


interface Port-channel200
 description 20 Gig MEC to VSS (cr2-6500-1 4/1-4/3)
 no switchport
 dampening
 ip address 10.122.0.26 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!
interface Port-channel201
 description 20 Gig to VSS (cr2-6500-1 4/1-4/3)
 no switchport
 dampening
 ip address 10.122.0.21 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

コア 1

```
router ospf 100
 router-id 10.254.254.7
 log-adjacency-changes detail ! Helps in NSF Restart Activities
```

```
 auto-cost reference-bandwidth 20000 !  Optional
 passive-interface default
 no passive-interface Port-channel1
 no passive-interface Port-channel20
 network 10.122.0.0 0.0.255.255 area 0

interface Port-channel20
 description 20 Gig MEC to VSS 1/2/1 2/2/1
 dampening
 ip address 10.122.0.27 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

コア 2

```
router ospf 100
 router-id 10.254.254.7
 log-adjacency-changes detail
 auto-cost reference-bandwidth 20000 !  Optional
 passive-interface default
 no passive-interface Port-channel1
 no passive-interface Port-channel20
 network 10.122.0.0 0.0.255.255 area 0

interface Port-channel21
 description 20 Gig to VSS 1/2/2-2/2/2
 dampening
 ip address 10.122.0.20 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

OSPF ECMP

VSS

```
router ospf 100
 router-id 10.122.0.235
 log-adjacency-changes detail
 auto-cost reference-bandwidth 20000 !  Optional
 nsf !  Enable NSF Capability
 area 120 stub no-summary
 area 120 range 10.120.0.0 255.255.0.0 cost 10
 area 120 range 10.125.0.0 255.255.0.0 cost 10
 passive-interface default
 no passive-interface TenGigabitEthernet1/2/1
 no passive-interface TenGigabitEthernet1/2/2
```

```
 no passive-interface TenGigabitEthernet2/2/1
 no passive-interface TenGigabitEthernet2/2/2
 network 10.120.0.0 0.0.255.255 area 120
 network 10.122.0.0 0.0.255.255 area 0
 network 10.125.0.0 0.0.255.255 area 120


interface TenGigabitEthernet1/2/1
 description 10 GigE to Core 1
 no switchport
 dampening
 ip address 10.122.0.26 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!
interface TenGigabitEthernet1/2/2
 description 10 GigE to Core 2
 no switchport
 dampening
 ip address 10.122.0.23 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!


interface TenGigabitEthernet2/2/1
 description to Core 1
 no switchport
 dampening
 ip address 10.122.0.32 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!
interface TenGigabitEthernet2/2/2
 description 10 GigE to Core 2
 no switchport
 dampening
 ip address 10.122.0.20 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 mls qos trust dscp
```

```
 hold-queue 2000 in
 hold-queue 2000 out
```

コア 1

```
router ospf 100
 router-id 10.254.254.7
 log-adjacency-changes detail
 auto-cost reference-bandwidth 20000 !  Optional
 passive-interface default
 no passive-interface GigabitEthernet2/5
 no passive-interface TenGigabitEthernet4/1
 no passive-interface TenGigabitEthernet4/3
 no passive-interface Port-channel1
 network 10.122.0.0 0.0.255.255 area 0


interface TenGigabitEthernet4/1
 description To VSS Ten1/2/1
 dampening
 ip address 10.122.0.27 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!
!
interface TenGigabitEthernet4/3
 description To VSS Ten2/2/1
 dampening
 ip address 10.122.0.33 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

コア 2

```
router ospf 100
 router-id 10.254.254.7
 log-adjacency-changes detail
 auto-cost reference-bandwidth 20000 !  Optional
 passive-interface default
 no passive-interface GigabitEthernet2/5
 no passive-interface TenGigabitEthernet4/1
 no passive-interface TenGigabitEthernet4/3
 no passive-interface Port-channel1
 network 10.122.0.0 0.0.255.255 area 0


interface TenGigabitEthernet4/1
 description To VSS Ten 1/2/2
 dampening
```

```
 ip address 10.122.0.22 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
!
!
interface TenGigabitEthernet4/3
 description To VSS Ten 2/2/2
 dampening
 ip address 10.122.0.21 255.255.255.254
 ip flow ingress
 ip pim sparse-mode
 ip ospf network point-to-point
 logging event link-status
 load-interval 30
 carrier-delay msec 0
 mls qos trust dscp
 hold-queue 2000 in
 hold-queue 2000 out
```

APPENDIX B

参考資料

次の各マニュアルおよび参考資料のリンクは、このマニュアルで説明した Campus 3.0 VSS 設計をサポートする補足説明を提供します。

- 『*Enterprise Campus 3.0 Architecture: Overview and Framework*』
  http://www.cisco.com/en/US/docs/solutions/Enterprise/Campus/campover.html
- 『*Campus Network for High Availability Design Guide*』
  http://www.cisco.com/en/US/docs/solutions/Enterprise/Campus/HA_campus_DG/hacampusdg.html
- 『*High Availability Campus Recovery Analysis*』
  http://www.cisco.com/en/US/docs/solutions/Enterprise/Campus/HA_recovery_DG/campusRecovery.html
- 『*High Availability Campus Network Design: Routed Access Layer using EIGRP or OSPF*』
  http://www.cisco.com/en/US/docs/nsite/campus/ha_campus_routed_access_cvd_ag.pdf
- 『*Cisco Catalyst 6500 Series Virtual Switching System (VSS) 1440 - CCO White Paper on VSS technology*』
  http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps9336/white_paper_c11_429338.pdf
- 『*Best Practices for Catalyst 6500/6000 Series and Catalyst 4500/4000 Series Switches Running Cisco IOS Software*』
  http://www.cisco.com/en/US/products/hw/switches/ps700/products_white_paper09186a00801b49a4.shtml
- 『*Migrate Standalone Cisco Catalyst 6500 Switch to Cisco Catalyst 6500 Virtual Switching System*』
  http://www.cisco.com/en/US/products/ps9336/products_tech_note09186a0080a7c74c.shtml