Cisco VN-Link：仮想化対応ネットワーキング

このドキュメントの内容

最近まで、データセンター ネットワークは、各エンド ノードがネットワークのエンドオブロー スイッチのアクセス ポートに接続され、単一のイメージ（OS の単一インスタンスや特定のアプリケーションの単一インスタンス）を実行する単一のサーバに相当する、という安全な前提のもとに設計されていました。また、アプリケーションとその関連 OS は常に特定の物理サーバ上で稼働し、別の物理サーバに移行することはほとんどない、ということも前提となっていました。しかし近年、ブレード サーバ アーキテクチャとそれに続くサーバ仮想化の登場によってこれらの前提が崩れ、データセンター ネットワークの設計にいくつかの新たな課題が生まれました。このドキュメントで説明するように、Cisco® VN-Link テクノロジーは、これらの課題に対処します。

データセンター ネットワークの設計の階層

現在のデータセンター ネットワークの設計は、世界最大規模のいくつかのデータセンターでのこの数年間のテストと改善によって実証済みの階層化アプローチに基づいています。データセンター ネットワークの 3 つのレイヤは、次のとおりです。

- **コア レイヤ**：データセンターで送受信するすべてのフローを処理する、高速パケット スイッチング バックプレーン
- **アグリゲーション レイヤ**：ロード バランシング、侵入検知、ファイアウォール、SSL オフロード、ネットワーク分析といったネットワーク ホステッド サービスの統合など、重要な機能を提供
- **アクセス レイヤ**：サーバをネットワークに物理的に接続し、Access Control List（ACL; アクセス コントロール リスト）、Quality of Service（QoS）、VLAN などのネットワーク ポリシーを適用

アクセスレイヤ ネットワークのインフラストラクチャは、一般的に大型のモジュラ スイッチをサーバ列の端に配置して、その列にある各サーバに接続を提供する方法（エンドオブロー モデル）か、より小型の固定構成のトップオブラック スイッチによって単一のラックまたはいくつかの隣接ラックに接続を提供し、アグリゲーション レイヤ デバイスにアップ

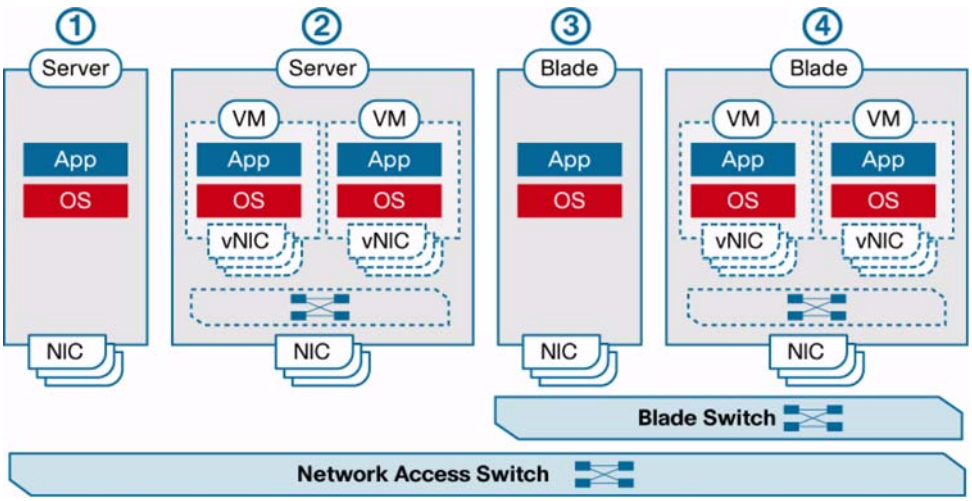
リンクする方法（トップオブラック モデル）のどちらかで実装できます。ブレード サーバ アーキテクチャでは、オプションの組み込みブレード スイッチをブレード エンクロージャ内に配置することで、アクセス レイヤが変化します。ブレード スイッチは、機能的にはアクセスレイヤ スイッチと同様で、トポロジ的にはアクセス レイヤに配置されますが、アクセスレイヤ スイッチとコンピューティング ノード（ブレード）との間にネットワークの追加レイヤとして展開されることも多く、この場合にはネットワーク設計に 4 つ目のレイヤが加わります。

仮想化の影響

サーバ仮想化では、複数の OS イメージが同じ物理サーバおよび I/O デバイスを透過的に共有できるので、データセンター ネットワーク設計における前述の前提は、いずれも変更されます。その結果、同じサーバ内の異なる仮想マシン間でのローカル スイッチングをサポートする必要があり、ネットワークのアクセス レイヤは本来の配置から遠ざけられ、各ネットワーク アクセス ポートが単一イメージを実行している単一物理サーバに相当するという前提は、無効になります。

サーバ仮想化では、イメージとネットワークの関係が静的であるという2 つめの前提も無効になります。仮想化では、ソフトウェアがハードウェアから切り離されるので、事実上、OS イメージの移行が可能になります。つまり、仮想マシンは、同じデータセンター内で、さらには複数のデータセンター間で、1 つの物理サーバから別の物理サーバへ移行できることになります。同じアクセス スイッチ内で移行できるだけでなく、同じまたは異なるデータセンターの別のアクセス スイッチにも移行できます。このこれまでにないモビリティがネットワークに与える影響は大きく、アクセス レイヤにとどまりません。たとえば、アグリゲーション レイヤで展開されるサービスも、仮想マシンのモビリティをサポートするために変更が必要になることがあります。基本的なレイヤ 2 スイッチングや接続の場合でも、VMware VMotion などの製品によって実装された仮想マシンのモビリティによって、基盤となるネットワーク インフラストラクチャ、特にアクセス レイヤにかなり厳しい要件が課されます。たとえば、送信元ホストと宛先ホストの両方が、同じレイヤ 2 ドメイン（VLAN）のグループに含まれている必要があります。したがって、特定の仮想化クラスタのスイッチ ポートはすべて、クラスタの仮想マシンが使用するすべての VLAN からのトラフィックを許可するトランク ポートとして、均一に設定する必要があります。これは、従来のネットワーク設計のベスト プラクティスとは明らかに異なります。図 1 に、いくつかのアクセス レイヤ接続オプションの比較図を示します。

図 1　アクセス レイヤ接続オプションの比較：(1) 仮想化していないラック最適化サーバ (2) 仮想化したラック最適化サーバ (3) 仮想化していないブレード サーバ (4) 仮想化したブレード サーバ



また、コンピューティングが比較的静的で、データセンター内の物理サーバはめったに移動されないことを前提としたネットワークに実装されているその他の機能にも、仮想マシンのモビリティは影響を及ぼします。たとえば、物理ポートに基づいてステート情報を保持するポート セキュリティ、IEEE 802.1x、IP ソース ガードなどの機能は、新世代のアクセスレイヤ スイッチには適用できません。なぜなら、仮想マシンは、任意の時点で移行するからです。さらに、仮想マシンは 1 つの物理サーバから他の物理サーバに移行するので、仮想マシンがネットワーク上のどこに配置されているとしても、ネットワークに定義されている仮想マシン用のすべてのネットワーク ポリシー（ACL など）が、一貫して適用されることが望ましいです。

ハイパーバイザ組み込み仮想スイッチ

仮想マシンのネットワーク化の最も簡単でシンプルな方法は、スタンドアロン ソフトウェア スイッチをハイパーバイザの一部として実装することです。これは、VMware が仮想スイッチ（vSwitch）で行ったことです。各 Virtual Network Interface Card（vNIC; 仮想ネットワーク インターフェイス カード）によって仮想マシンを vSwitch に論理的に接続し、仮想マシンがそのインターフェイス経由でトラフィックを送受信できるようにします。同じ vSwitch に接続している 2 つの vNIC 間で通信する必要がある場合は、vSwitch がレイヤ 2 スイッチング機能を直接実行するので、物理ネットワークにトラフィックを送信する必要はありません。

組み込み vSwitch を使ったアプローチの最も大きな利点は、そのシンプルさです。各ハイパーバイザには、vSwitch の 1 つまたは複数の独立したインスタンスが含まれます。ただし、各組み込み vSwitch をそれぞれ設定する必要があるので、データセンターで複数の VMware ESX サーバを展開する場合には、この利点が欠点になってしまいます。vSwitch のもう 1 つの問題は、ネットワークの一部でありながら、ネットワーク インフラストラクチャの一貫した管理の対象にならないことです。実際、ネットワーク管理者は、vSwitch にアクセスさえできないこともあります。ほとんどの実稼働環境では、vSwitch は管理対象外のネットワーク デバイスになりますが、これはもちろん、望ましい状況ではありません。適切なレベルのコンプライアンスや可視性を確保するために IT 部門がネットワークの機能を活用している、ミッションクリティカルな、または規制の厳しい環境では、なおさらのことです。このアプローチでは、IT インフラストラクチャの重要ポイントで、運用上の一貫性が失われます。サーバ管理者は、インフラストラクチャ全体に適用されているベスト プラクティス、診断ツール、管理およびモニタリング機能を使用せずに、ネットワークの一部のメンテナンスとセキュリティの責任を負う必要があります。

さらに、vSwitch には、仮想マシンのモビリティの問題を解決するための特別な手段はありません。管理者は、仮想マシンの移行によってネットワーク ポリシー違反または基本的な接続障害が起きないように、移行元と移行先の両方の VMware ESX ホスト上の vSwitch と、アップストリーム物理アクセスレイヤ ポートの設定が整合しているかどうかを、手動で確認する必要があります。vSwitch を使用して仮想マシンのネットワーキングを実行する仮想化サーバ環境でモビリティの要件をサポートするには、物理アクセスレイヤ ポートをトランク ポートとして設定することが避けられない条件になります。

組み込み vSwitch の制限を克服するために、VMware とシスコは Distributed Virtual Switch（DVS; 分散仮想スイッチ）のコンセプトを共同で作成しました。基本的には、組み込みスイッチのコントロール プレーンとデータ プレーンを切り離し、複数の独立した vSwitch（データ プレーン）を集中管理システム（コントロール プレーン）によって管理するという

ものです。VMware は、DVS の独自の実装を vNetwork Distributed Switch として製品化しており、VMware vCenter にコントロール プレーンのコンポーネントが実装されています。このアプローチにより、仮想マシンの管理者は、実質的にホストレベルのネットワーク設定から解放され、VMware ESX クラスタ レベルでネットワーク接続を管理できるようになります。

Cisco VN-Link

シスコは、DVS フレームワークを使用して、分散ハイパーバイザ レイヤ内で直接運用できるネットワーキング ソリューションのポートフォリオを用意し、シスコの他のネットワーキング製品と整合性のある機能セットおよび運用モデルを提供しています。このアプローチにより、サーバ仮想化がもたらす新しい要件を満たすためのエンドツーエンド ネットワーク ソリューションが実現します。具体的には、既存のネットワーク運用モデルに合致した方法で、仮想マシンのインターフェイスを個別に識別、設定、監視、移行、および診断できる新しい機能セットが提供されます。

これらの機能の総称が、Cisco Virtual Network Link（VN-Link）です。VN-Link は、文字どおり、仮想マシン上の vNIC と、VN-Link 対応のシスコ製スイッチとの間に論理リンクを作成します。このマッピングは、ケーブルを使用して NIC をアクセスレイヤ スイッチのネットワーク ポートに接続することを論理的に行ったものです。

仮想イーサネット インターフェイス

VN-Link 対応スイッチは、Virtual Ethernet（vEth; 仮想イーサネット）インターフェイスのコンセプトに基づいて動作します。これらの仮想インターフェイスは、ハイパーバイザ管理レイヤ（VMware vCenter など）による仮想マシンのプロビジョニングの結果として、スイッチに格納されるネットワーク ポリシーに基づいて動的にプロビジョニングされます。さらに、これらの仮想インターフェイスは、モビリティ イベントに関係なく、特定の仮想インターフェイスのネットワーク設定アトリビュート、セキュリティ、および統計情報を保持します。

仮想イーサネット インターフェイスは、物理ネットワーク アクセス ポートの仮想バージョンです。VN-Link 対応スイッチでは、1 つの物理ポートに複数の vEth インターフェイスを実装でき、各 vEth インターフェイスと、仮想マシン上の対応する vNIC 間でマッピングが行われます。vEth インターフェイスの最も重要な利点は、仮想マシンが 1 つの物理サーバから別の物理サーバに移行しても、vNIC を追跡できることです。この移行は、NetFlow、ポート統計情報、Switched Port Analyzer（SPAN; スイッチド ポート アナライザ）セッションなどのポートの設定およびステートを保持したまま実行されます。vEth インターフェイスを使用してネットワーク アクセス ポートを仮想化することにより、VN-Link は、異なる物理サーバ間および異なる物理アクセスレイヤ スイッチ間での仮想マシンの透過的なモビリティを可能にします。

ポート プロファイル

ポート プロファイルは、物理インターフェイスまたは仮想インターフェイスのどちらにも動的に適用できる、インターフェイス コンフィギュレーション コマンドの集合です。特定のポート プロファイルへの変更は、そのプロファイルに関連付けられているすべてのポートに即時に伝播されます。ポート プロファイルには、VLAN、Private VLAN（PVLAN; プライベート VLAN）、ACL、ポート セキュリティ、NetFlow 収集、レート制限、QoS マーキング、さ

らには仮想マシン単位で高度なトラブルシューティングができるリモートポート ミラーリング（Encapsulated Remote SPAN [ERSPAN] を使用）まで、きわめて高度なアトリビュートの集合を定義できます。

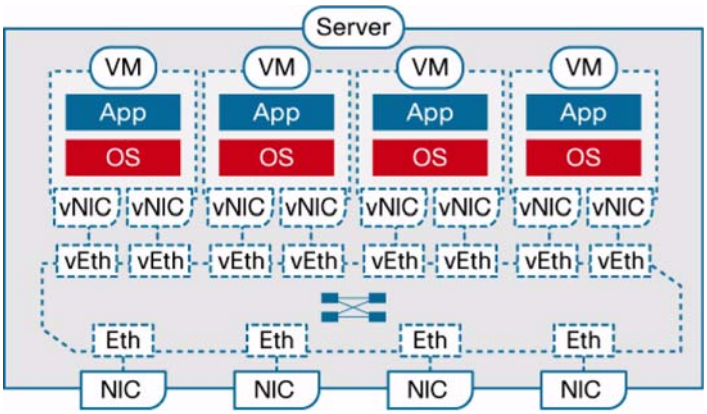
次に、ポート プロファイルの設定例を示します。

```
(config)# port-profile webservers
(config-port-prof)# switchport access vlan 10
(config-port-prof)# ip access-group 500 in
(config-port-prof)# inherit port-profile server
```

作成したポート プロファイルは、次のように、特定の vEth インターフェイスに割り当てます。

```
(config)# interface veth1
(config-if)# inherit port-profile webservers
```

図 2　VN-Link 対応スイッチ（Cisco Nexus™ 1000V シリーズ スイッチ）での仮想ネットワークと物理ネットワークの構築関係



ポート プロファイルは、仮想マシンの管理レイヤ（VMware vCenter など）に密接に統合されるので、仮想インフラストラクチャの管理が簡素化されます。ポート プロファイルは、ネットワーク管理者が設定して、管理します。仮想マシン管理レイヤへの統合を容易にするために、Cisco VN-Link スイッチは、ポート プロファイルのカタログを VMware vCenter などの仮想マシン管理ソリューションにプッシュします。ここで、ポート プロファイルは個別のポート グループとして認識されます。このような統合の結果、仮想マシンの管理者は、仮想マシンを作成する際、単純にプロファイルのメニューから選択するだけで済みます。仮想マシンの電源をオンまたはオフにすると、対応するプロファイルにより、VN-Link スイッチで vEth インターフェイスが動的に設定されます。

VN-Link を実装する方法は 2 つあります。

- ハイパーバイザ　レイヤ内で完全なソフトウェアとして　Cisco　DVS　を実行する（Cisco Nexus 1000V シリーズ）
- Network Interface Virtualization（NIV; ネットワーク インターフェイス仮想化）をサポートする新しいクラスのデバイスを使用し、ハイパーバイザ内でのソフトウェアベースのスイッチングを不要にする

Cisco Nexus 1000V シリーズを使用した既存ネットワークでの VN-Link の展開

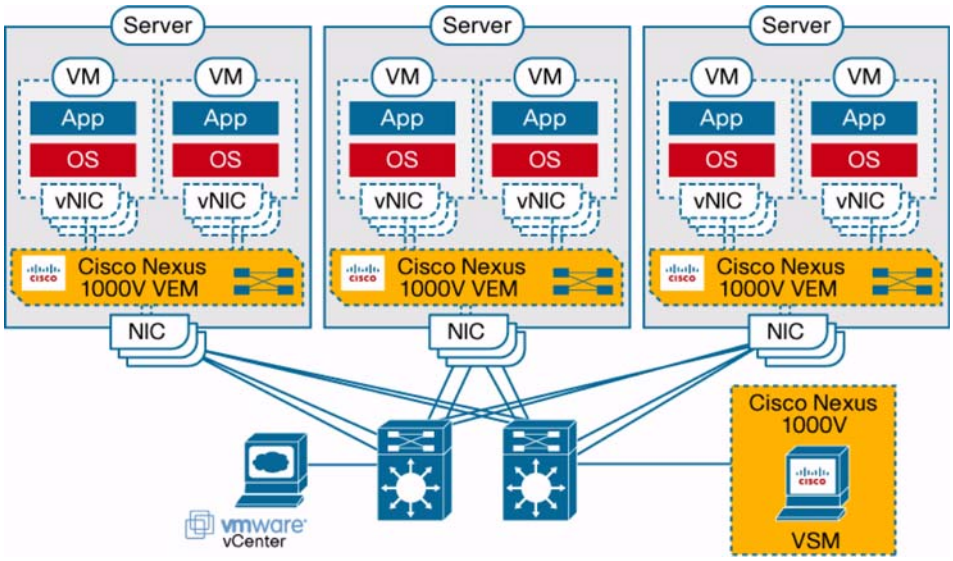
DVS フレームワークの導入により、サードパーティ ネットワーキング ベンダーは、VMware の vNetwork スイッチ API インターフェイスを使用して分散仮想スイッチの独自の実装を提供できるようになりました。シスコは VMware との密接なコラボレーションによって、これらの API を設計しました。Cisco Nexus 1000V シリーズは、仮想化管理者用の VMware vCenter を含む VMware 仮想インフラストラクチャに完全に統合できる、最初のサードパーティ DVS です。Cisco Nexus 1000V シリーズを展開すると、仮想化管理者の定期的なワークフローが維持されるだけでなく、ネットワーク管理者にとっても vSwitch およびポート グループ設定の負荷が軽減されるので、ネットワーク設定のミスが削減され、データセンター全体で一貫したネットワーク ポリシーを確実に施行できるようになります。

Cisco Nexus 1000V シリーズは、冗長スーパーバイザ機能を備えた 66 スロット モジュライーサネット スイッチを仮想的にエミュレートできる、2 つの主要コンポーネントで構成されています。

- Virtual Ethernet Module（VEM）- データ プレーン : ハイパーバイザ内で稼働する軽量のソフトウェア コンポーネントです。高度なネットワーキングおよびセキュリティ機能をサポートし、直接接続された仮想マシン間のスイッチングを実行して、他のネットワークへのアップリンク機能を提供します。実質的に vSwitch の代替になります。各ハイパーバイザに、1 つの VEM が組み込まれています。
- Virtual Supervisor Module（VSM）- コントロール プレーン : Cisco Nexus 1000V シリーズ システム（VSM と、VSM が管理するすべての VEM の組み合わせ）の設定、管理、監視、診断、および VMware vCenter との統合を実行するスタンドアロンの外付け物理または仮想アプライアンスです。単一 VSM で、最大 64 の VEM を管理できます。VSM は、アクティブ/スタンバイ モデルで配置できるので、ハイアベイラビリティを確保できます。

Cisco Nexus 1000V シリーズでは、仮想マシン間のトラフィックは、VEM の各インスタンスでローカルでスイッチングされます。各 VEM はまた、アップストリーム アクセスレイヤ ネットワーク スイッチ（ブレード、トップオブラック、エンドオブローなど）経由で、ローカル仮想マシンをネットワーク全体に相互接続する役割も担っています。VSM は、コントロール プレーン プロトコルを実行し、これに応じて各 VEM のステートを設定しますが、実際のパケットの転送にはまったく関与しません（図 3）。

図 3　Cisco Nexus 1000V シリーズの分散スイッチング アーキテクチャ



ネットワーク インターフェイス仮想化での VN-Link の展開

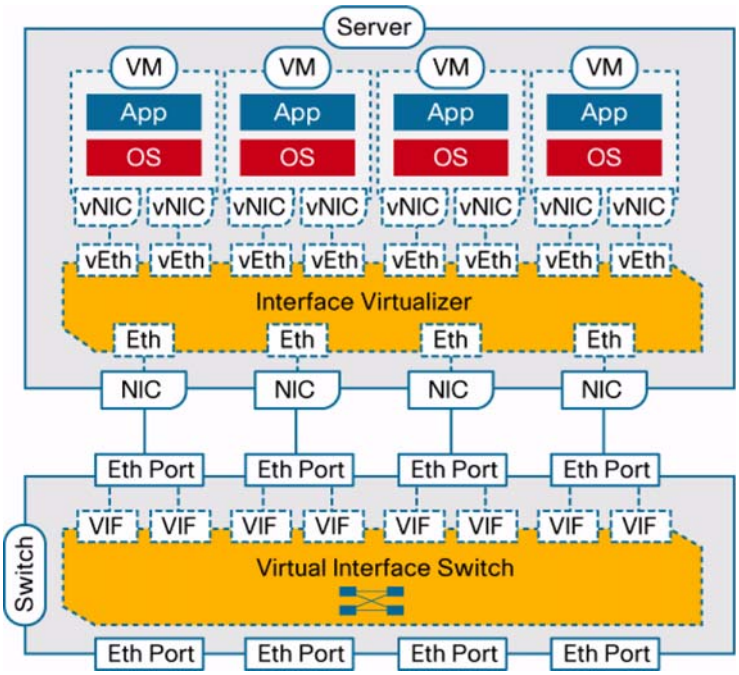
ハイパーバイザ、管理レイヤ、仮想ネットワーキング コンポーネント間の密接な統合を必要とし、ハイパーバイザ内のソフトウェアでスイッチングを実現する DVS モデルに加え、シスコはNetwork Interface Virtualization（NIV; ネットワーク インターフェイス仮想化）のコンセプトに基づくハードウェア アプローチを開発しました。NIV では、すべてのスイッチング機能がハイパーバイザから取り除かれ、サーバとは物理的に独立しているハードウェア ネットワーク スイッチによって実行されます。しかし、NIV ではホスト上にコンポーネントが必要となります。これは、インターフェイス バーチャライザと呼ばれ、ハイパーバイザ内のソフトウェアまたはインターフェイス バーチャライザ対応アダプタ内のハードウェアのどちらかで実装します。インターフェイス バーチャライザには、2 つの目的があります。

- サーバからネットワークへの発信トラフィックの場合、インターフェイス バーチャライザは送信元 vNIC を識別し、その vNIC により生成された各パケットに対して、Virtual Network Tag（VNTag）と呼ばれる固有タグを明示的に付加します。
- ネットワークからの着信トラフィックの場合、インターフェイス バーチャライザは VNTag を取り外して、そのパケットを特定の vNIC に転送します。

インターフェイス バーチャライザは、仮想マシン間のローカル スイッチングは実行しません。スイッチング プロセスは、ハイパーバイザから完全に切り離されます。これにより、仮想マシンのネット ワーキングは、実質的に物理デバイスのネット ワーキングと同等になります。

スイッチングは常時、インターフェイス バーチャライザが接続しているネットワーク スイッチによって実行されます。このスイッチは、物理ポート間のスイッチングだけでなく、スイッチのリモート vNIC に対応する仮想インターフェイス（VIF）間のスイッチングも実行するので、Virtual Interface Switch（VIS; 仮想インターフェイス スイッチ）と呼ばれます。つまり、仮想マシンの各 vNIC は VIS の VIF に対応し、スイッチングまたはポリシー施行の機能はすべて、ハイパーバイザではなく VIS 内で実行されます。VIS には、NIV をサポートしていれば、ネットワーク内の任意のアクセスレイヤ スイッチ（ブレード、トップオブラック、またはエンドオブローのスイッチ）を使用できます（図 4）。

図 4　NIV モデルのアーキテクチャの要素



NIV モデルにおいて重要なことは、VIS は単なる IEEE 802.1D 準拠イーサネット スイッチではなく、新たに定義された衛星的な関係をサポートするために、いくつかの拡張機能を実装している必要があるということです。これらの拡張機能はリンク ローカルで、スイッチおよびインターフェイス バーチャライザの両方に実装されている必要があります。仮想マシンは単一物理リンク上で多重化されるので、これらの拡張機能が実装されていないと、異なる仮想マシンに属しているトラフィックを識別できません。

また、VIS は、場合によっては送信元ポートと同じポート上にフレームを戻せる必要があります。レイヤ 2 イーサネット スイッチの動作を定義する IEEE 801.D 標準では、この規格に準拠するスイッチは、送信元と同じインターフェイス上にフレームを戻すことは許可されないと、明記されています。この対策は当初、レイヤ 2 トポロジでのループ発生を回避し、レイヤ 2 フォワーディング エンジンのハードウェア実装を比較的簡単にするために、規格に取り入れられました。現在のフォワーディング エンジン実装テクノロジーは、より高度なアルゴリズムを採用しているので、この要件はもはや必須ではありません。しかしながら、パケットを送信元と同じインターフェイスに戻すというネットワーク機能には、現在も適切なレベルの標準化が必要とされています。シスコが定義した VNTag のプロトコルは、標準化のために IEEE 802.3 タスク フォースに提起されています。

NIV は、VN-Link 運用フレームワーク内で展開するために設計された、レイヤ 2 の革新的なコンセプトです。具体的には、ポート プロファイル、vEth インターフェイス、仮想マシンのモビリティのサポート、一貫したネットワーク展開および運用モデル、ハイパーバイザ マネージャとの統合など、Cisco Nexus 1000V シリーズと同じメカニズムが含まれています。

まとめ

ブレード サーバ アーキテクチャとサーバ仮想化の導入により、データセンター ネットワークにおけるいくつかの設計上、運用上、および診断上の前提は、もはや無効になりました。サーバ仮想化では、複数の OS イメージが、同じ物理サーバおよび I/O デバイスを透過的に共有できます。その結果、同じサーバ内の仮想マシン間でのローカル スイッチングをサポートする必要性が生じます。シスコと VMware は、VMware 仮想インフラストラクチャにサードパーティのネットワーキング機能を透過的に統合できる API のセットを共同開発しました。

シスコは、VN-Link を実装するために、この機能の利点を活用した最初のネットワーク ベンダーです。VN-Link は、分散ハイパーバイザ レイヤ内で直接運用できるネットワーキング ソリューションのポートフォリオで、シスコの他のネットワーキング製品と整合性のある機能セットおよび運用モデルが提供されます。このアプローチにより、サーバ仮想化がもたらす新しい要件を満たすためのエンドツーエンド ネットワーク ソリューションが実現します。

VN-Link は、ハイパーバイザ レイヤ内で完全にソフトウェアとして実行される分散仮想スイッチ（DVS）を使用する方法（Cisco Nexus 1000V シリーズ）、またはネットワーク インターフェイス仮想化（NIV）をサポートする新しいクラスのデバイスを使用して、ハイパーバイザ内のソフトウェアベースのスイッチングを不要にする方法のいずれかで実装できます。VN-Link は、仮想マシンのネットワーキング要件を満たし、仮想化データセンターでの高度で簡素化された将来的な接続オプションの基盤となる、即効性のあるソリューションです。

関連情報

Cisco Nexus 1000V シリーズ スイッチについて詳しくは、http://www.cisco.com/jp/go/nexus1000/ を参照してください。

Cisco Data Center 3.0 の詳細については、http://www.cisco.com/jp/go/datacenter/

IEEE に提起された NIV プロポーザルの詳細については、http://www.ieee802.org/1/files/public/docs2008/new-dcb-pelissier-NIC-Virtualization-0908.pdf（英語）を参照してください。

シスコシステムズ合同会社
〒 107-6227 東京都港区赤坂 9-7-1 ミッドタウン・タワー
http://www.cisco.com/jp
お問い合わせ先：シスコ コンタクトセンター
0120-092-255（フリーコール、携帯・PHS 含む）
電話受付時間：平日 10:00 ～ 12:00、13:00 ～ 17:00
http://www.cisco.com/jp/go/contactcenter/

お問い合せ先

09.06