

ネットワークに最適なルータ スイッチング パスの選択方法

目次

[概要](#)

[プロセス スイッチング](#)

[割り込みコンテキスト スイッチング](#)

[ファストスイッチング](#)

[最適なスイッチング](#)

[Cisco Express Forwarding](#)

[どのスイッチング パスが最適か](#)

[関連情報](#)

概要

さまざまな Cisco ルータおよび Cisco IOS® リリースで使用できるスイッチング パスが数多くあります。ネットワークに最適なスイッチング パスはどれですか。これらはすべてどのように機能しますか。この解説書では、次に示す各種のスイッチング パスについて説明し、使用中のネットワークに最適なスイッチング パスを決定できるようにします。

最初に転送プロセスについて説明します。ルータを経由したパケットの転送は、3 段階の手順で成り立っています。

1. パケットの宛先に到達できることを判定する。
2. 宛先への経路のネクストホップと、そのホップへ到達できるインターフェイスを判別する。
3. パケットの Media Access Control (MAC; メディア アクセス制御) ヘッダーを書き換え、ネクストホップに正しく到達できるようにする。

これらの各手順は、パケットが宛先に到達するために必須の項目です。

注: このドキュメントでは、IP スイッチング パスを例として使用します。ここで説明する情報は、他のプロトコルの同様のスイッチング パスがある場合でも適用できます。

プロセス スイッチング

プロセススイッチングは、スイッチング パスの中では最も一般的に広まっているものであり、各バージョンの IOS および各種プラットフォームで、スイッチングの対象になるあらゆる種類のトラフィックをサポートしています。プロセススイッチングは、次の 2 つの重要な概念によって定義されます。

- パケット上の MAC ヘッダーを書き換えるために使用されるフォワーディングの判断と情報は、ルーティング テーブル (ルーティング情報ベース、または RIB) と Address Resolution Protocol (ARP; アドレス解決プロトコル) キャッシュ、あるいは、そのルータに直接接続さ

れている各ホストの IP アドレスにマップされた MAC ヘッダー情報を含む他のテーブルから得られます。

- パケットは、IOS 内で実行されている通常のプロセスによって交換されます。つまり、フォワーディングの判断は、IOS スケジューラでスケジュールされたプロセスによってなされ、ルーティング プロトコルのようにルータ上の他のプロセスと同等のものとして実行されます。ルータ上で実行されている通常のプロセスは、パケットのプロセススイッチングのために中断されることはありません。

次の図では、プロセススイッチングの処理経路について説明しています。

この図を詳しく説明します。

1. ネットワークメディア上にパケットがあることがインターフェイス プロセッサによって最初に検出され、このパケットがルータの入出力メモリ (I/Oメモリ) に転送されます。
2. インターフェイス プロセッサは受信割り込みを生成します。この割り込みの間、中央のプロセッサがパケットのタイプを判別し (ここでは IP パケットと仮定します)、必要に応じてパケットを適プロセッサ メモリにコピーします (この決定はプラットフォームによって異なります)。最終的にプロセッサによってパケットが適切なプロセスの入力キューに配置され、割り込みが解除されます。
3. 次にスケジューラが実行される時、`ip_input` の入力キューにあるパケットが確認され、このプロセスが実行されるようスケジュールされます。
4. `ip_input` の実行時には、RIB を参考にしてネクストホップと出力インターフェイスが決定されます。さらに、ARP キャッシュを参考にして、ネクストホップの正確な物理層レイヤのアドレスが決定されます。
5. 次に `ip_input` によりパケットの MAC ヘッダーが書き換えられ、このパケットが正しい送信インターフェイスの出力キューに配置されます。
6. 送信インターフェイスの出力キューから送信インターフェイスの送信キューにパケットがコピーされます。送信に関する QOS は、この 2 つのキューの間で行われます。
7. 出力インターフェイス プロセッサによって送信キュー上のパケットが検出され、ネットワーク メディアへパケットが送信されます。

ネットワーク アドレス変換 (NAT) やポリシー ルーティングなど、パケット交換に影響を与える機能のほとんどは、プロセススイッチング パスに導入されました。 [割り込みコンテキスト スイッチング](#)でも、これらの実証済みで最適化された機能が利用されることがあります。

[割り込みコンテキスト スイッチング](#)

割り込みコンテキスト スイッチングは、シスコのルータで 2 番目によく使用されているスイッチング方法です。割り込みコンテキスト スイッチングとプロセススイッチングの主な違いは次のとおりです。

- プロセッサ上で現在実行中のプロセスは、パケットの交換のために中断される。パケットは `ip_input` プロセスがスケジュールされた時点でのみ交換されるのではなく、要求があった時点で交換される。
- プロセッサは、いくつかのルート キャッシュの形式を使用して、パケットの交換に必要な全情報を収集する。

次の図に、割り込みコンテキスト スイッチングについて示します。

この図を詳しく説明します。

1. ネットワークメディア上にパケットがあることがインターフェイス プロセッサによって最初に検出され、このパケットがルータの入出力メモリ (I/Oメモリ) に転送されます。
2. インターフェイス プロセッサは受信割り込みを生成します。この割り込みの間に、中央のプロセッサによってパケットのタイプが判断され (ここでは IP パケットとして仮定します)、パケットの交換が開始されます。
3. プロセッサによってルート キャッシュが検索されます。これにより、パケットの宛先が到達可能かどうか判断され、選択する出力インターフェイス、この宛先へ送るためのネクストホップ、そして最後にパケットをネクストホップへ問題なく送るための MAC ヘッダーが決定されます。プロセッサはこの情報を使用してパケットの MAC ヘッダーを書き換えます。
4. この時点で、パケットが送信用インターフェイスの送信キューまたは出力キューのいずれかにコピーされます (さまざまな要素に応じて異なります)。受信割り込みが返され、割り込みが発生する前にプロセッサ上で動作していたプロセスの実行が継続されます。
5. 出力インターフェイス プロセッサによって送信キュー上のパケットが検出され、ネットワーク メディアへパケットが送信されます。

この説明を読んだ後に最初に浮かぶ疑問は、「キャッシュの内容は何か」です。割り込みコンテンツ スイッチングのタイプに応じて 3 通りの可能性があります。

- [ファストスイッチング](#)
- [最適なスイッチング](#)
- [Cisco Express Forwarding](#)

[ファストスイッチング](#)

ファストスイッチングでは、フォワーディング情報と MAC ヘッダーの書き換え文字列を蓄積し、高速に検索および参照できるようにバイナリ ツリーの形式を採用しています。次の図では、バイナリ ツリーの例を示しています。

ファストスイッチングでは、到達可能性に関する情報が、そのパケットの宛先に対するバイナリ ツリーにノードがあることで示されます。各宛先に対する MAC ヘッダーと送信インターフェイスは、ノード情報の一部としてツリー内に蓄積されます。バイナリ ツリーには、実際には 32 レベルを作成できます (簡潔に図示するために、上の図では省略してあります)。

バイナリ ツリーを検索するには、検索する (バイナリ) 番号の左側 (最上位桁) から開始し、その番号に基づいてツリーを右または左に分岐します。たとえば、このツリーで番号 4 に関する情報を検索するには、2 進数値の最初の桁が 1 であるため、右側に分岐します。2 進数の次の桁を見ながら、最後の桁になるまでツリーを下っていきます。

[ファストスイッチングの特性](#)

ファストスイッチングは、バイナリ ツリーの構造に、ツリーのノードの一部として MAC ヘッダーの書き換え情報を保管するため、いくつかの特性があります。

- ルーティング テーブルとファストキャッシュの内容 (MAC ヘッダーの書き換え情報など) には相関関係がないため、キャッシュ エントリの作成にはプロセススイッチング パスの中で実行されなければならないすべてのプロセスが関連します。したがって、ファストキャッシュのエントリは、パケットがプロセススイッチングされた時点で構築されます。
- ARP キャッシュの MAC ヘッダー (書き換えに使用) とファストキャッシュの構造には相関関係がないため、ARP テーブルが変更されたときには、ファストキャッシュの一部を無効化

する必要があります (パケットのプロセススイッチングにより再構築されます)。

- ファストキャッシュでは、ルーティング テーブルにある特定の宛先に対して、1 階層 (1 プレフィクス長) にだけエントリが構築されます。
- ファストキャッシュ内 (MAC ヘッダーおよびアウトバウンドインターフェイス情報はノード内にあると前提します) の 1 つのエントリから別のエントリを指定できません。よって、全てのルーティング回帰の解決はファストキャッシュエントリが構築されているうちに行わなければならない。つまり、ファストキャッシュ自体の中では再帰ルートを解決できません。

ファストスイッチングのエントリのエージング

ファストスイッチングのエントリからルーティング テーブルや ARP キャッシュとの同期が失われないようにするため、および、ファストキャッシュ内の使用されていないエントリがルータのメモリを過度に浪費しないようにするために、ファストキャッシュの内容が毎分 20 分の 1 ずつランダムに無効化されます。ルータのメモリが非常に低い水準値を下回った場合は、ファストキャッシュのエントリが毎分 5 分の 1 ずつ無効化されます。

ファストスイッチングのプレフィクス長

全ての宛先に対して 1 プレフィクス長でのみエントリが構築できる場合、ファストスイッチングではどのようなプレフィクス長でエントリを構築するでしょうか。ファストスイッチングでは、宛先とはルーティング テーブル内の 1 つの到達可能な宛先、またはメジャー ネットワークです。特定のキャッシュ エントリを作成するためのプレフィクス長を決定するルールは次のとおりです。

- ファースト ポリシー エントリを構築する場合は、常に /32 にキャッシュ
- Multiprotocol over ATM virtual circuit (MPOA VC; MPOA バーチャル サーキット) のエントリを構築する場合には、常に /32 にキャッシュ
- ネットワークがサブネット化されていない場合 (メジャー ネットワークのエントリの場合) : 直接接続の場合は、/32 を使用そうでない場合は、そのメジャー ネット マスクを使用
- スーパーネットの場合、そのスーパーネットのマスクを使用
- ネットワークがサブネット化されている場合 : 直接接続の場合は、/32 を使用サブネットへの複数の経路がある場合は、/32 を使用他のすべてのケースでは、メジャー ネットでの最長のプレフィクス長を使用

ロードシェアリング

ファストスイッチングは完全に宛先ベースであるため、ロードシェアリングは宛先ごとに発生します。ある宛先ネットワークに対して、複数の等コストパスがある場合、ファストキャッシュではそのネットワーク内にある到達可能なホストごとに 1 つのエントリがありますが、あるホストに向けたトラフィックはすべて 1 つのリンクを辿ります。

最適なスイッチング

最適なスイッチングでは、フォワーディング情報と MAC ヘッダーの書き換え情報を、256 way multiway tree (256 way mtree) に蓄積します。次の図に示すように、mtree を使用することにより、プレフィクスを調べる際に必要となる手順の数が少なくなります。

256 の分岐のうち、各ツリー レベルをたどる分岐を判別するには各オクテットが使用されます。つまり、宛先の検出で最大 4 回のルックアップが行われます。プレフィクス長が短い場合には、

必要なルックアップが 1 ~ 3 回のみという可能性があります。MAC ヘッダー書き換え情報および出カインターフェイス情報は、ツリー ノードの一部として蓄積されるため、ファストスイッチングでもキャッシュの無効化とエージングが発生します。

最適なスイッチングでは、ファストスイッチングと同様の方法で、各キャッシュ エントリに対するプレフィクス長を決定します。

Cisco Express Forwarding

Cisco Express Forwarding でも、フォワーディング情報および MAC ヘッダー書き換え情報を蓄積するために 256 way データ構造を使用しますが、ツリーは使用しません。Cisco Express Forwarding は、trie を使用します。つまり、実際の検索対象情報がデータ構造内になく、データは個別のデータ構造に蓄積され、trie は単にその構造をポイントするだけです。つまり、Cisco Express Forwarding は送信インターフェイスおよび MAC ヘッダー書き換えをツリー自体の中に蓄積せずに、この情報を隣接関係テーブルと呼ばれる別個のデータ構造に蓄積します。

到達可能性に関する情報 (CEF テーブル) とフォワーディング情報 (隣接関係テーブル) を分けることによる利点がいくつかあります。

- 隣接関係テーブルは CEF テーブルとは別に構築できるため、パケットをプロセススイッチングすることなく両方を構築できます。
- パケットの転送で使用される MAC ヘッダーの書き換えはキャッシュ エントリに保存されないため、MAC ヘッダー書き換え文字列を変更してもキャッシュ エントリを無効にする必要がありません。
- 再帰ルートを解決するため、再帰ネクスト ホップではなくフォワーディング情報を直接ポイントすることができます。

この仕組みでは基本的にすべてのキャッシュのエージングが除外され、キャッシュはルーティング テーブルと ARP キャッシュに含まれている情報に基づいて事前に構築されます。この方法では、キャッシュ エントリの構築のためにパケットをプロセススイッチングする必要がありません。

隣接関係テーブルのその他のエントリ

隣接関係テーブルには、MAC ヘッダー書き換え文字列や送信インターフェイス情報以外のエントリも含めることができます。隣接関係テーブルに含めることのできる各種のエントリには次のものがあります。

- **キャッシュ** : ある隣接ホストまたはルータに到達するために使用される、MAC ヘッダー書き換え文字列および送信インターフェイス。
- **受信** : この IP アドレスに宛てられたパケットは、ルータによって受信される。この情報には、ブロードキャスト アドレスとルータ自体に設定されたアドレスが含まれる。
- **ドロップ** : この IP アドレスに宛てられたパケットはドロップする。この情報は、アクセス リストによって拒否されたり、NULL インターフェイスに送られたトラフィック用に使用されたりする場合がある。
- **パント** : Cisco Express Forwarding はこのパケットを交換できない。次の最適なスイッチング方式 (一般的にはファストスイッチング) で処理されるように転送する。
- **グリーニング** : ネクストホップは直接接続されているが、現在使用できる MAC ヘッダー書き換え文字列がない。

グリーニング隣接関係

グリーニング隣接関係エントリは、特定のネクストホップが直接接続されている必要があるが、MAC ヘッダー書き換え情報がないことを示しています。これらの情報はどのように構築され、使用されるのでしょうか。次の図で示すように、CEF を実行し、ブロードキャスト ネットワークに接続されたルータは、デフォルトでいくつかの隣接関係テーブルを構築しています。

デフォルトで作成される 4 つの隣接関係テーブルのエントリは次のとおりです。

```
10.1.1.0/24, version 17, attached, connected
0 packets, 0 bytes
  via Ethernet2/0, 0 dependencies
    valid glean adjacency
```

```
10.1.1.0/32, version 4, receive
10.1.1.1/32, version 3, receive
10.1.1.255/32, version 5, receive
```

ここでは、4 つのエントリがあることに注意してください。3 つの受信と 1 つのグリーニングです。各受信エントリは、ブロードキャスト アドレスまたはルータ上で設定されたアドレスを表し、グリーニング エントリは、接続されたネットワークのアドレス空間の空きを示しています。ホスト 10.1.1.50 宛のパケットが受信されると、ルータはこれを交換しようと試み、パケットがグリーニング隣接関係に解決されていることを発見します。その後、ARP キャッシュ エントリが 10.1.1.50 のために必要とされていること、ARP プロセスから ARP パケットが送信されたこと、そして適切な隣接関係テーブルのエントリが新しい ARP キャッシュ情報から構築されたことを、Cisco Express Forwarding が通知します。この手順が完了すると、隣接関係テーブルに 10.1.1.50 のエントリが作成されます。

```
10.1.1.0/24, version 17, attached, connected
0 packets, 0 bytes
  via Ethernet2/0, 0 dependencies
    valid glean adjacency
10.1.1.0/32, version 4, receive
10.1.1.1/32, version 3, receive
10.1.1.50/32, version 12, cached adjacency 208.0.3.2
0 packets, 0 bytes
  via 208.0.3.2, Ethernet2/0, 1 dependency
    next hop 208.0.3.2, Ethernet2/0
    valid cached adjacency
10.1.1.255/32, version 5, receive
```

次にルータが受信する 10.1.1.50 宛のパケットは、この新しい隣接関係を使って交換されます。

ロードシェアリング

Cisco Express Forwarding では、Cisco Express Forwarding テーブルと隣接関係テーブルの分離を使用して、他の割り込みコンテキスト スイッチング モードよりも優れた方式のロードシェアリングも提供しています。ロードシェア テーブルは、次の図で示すように、Cisco Express Forwarding テーブルと隣接関係テーブルの間に挿入されます。

Cisco Express Forwarding テーブルがポイントするロードシェア テーブルには、使用可能なパラレルパス向けの隣接関係テーブル エントリを参照するポインタが含まれています。送信元アドレスと宛先アドレスはハッシュ アルゴリズムを介して渡され、各パケットに使用するロードシェア テーブル エントリを判別します。パケットごとに異なるロードシェア テーブル エントリを使用する場合、パケット単位のロードシェアリングを設定できます。

各ロードシェア テーブルには 16 個のエントリが含まれており、その中では使用可能なパスがルーティング テーブルのトラフィック共有カウンタに基づいて分類されます。ルーティング テー

ブルのトラフィック共有カウンタがすべて 1 の場合 (等コスト パスが複数ある場合)、各ネクストホップはロードシェア テーブルから同数のポインタを受信します。使用可能なパスの数を 16 に均等に割り切れない場合 (16 個のロードシェア テーブル エントリがあるため)、パスによっては他のパスよりもエントリが多くなります。

Cisco IOS ソフトウェア リリース 12.0 以降の場合、ロードシェア テーブル内のエントリ数が少なくなり、ロードシェア テーブルのエントリの数がパスごとに適切になります。たとえば、ルーティング テーブルに等コスト パスが 3 つある場合、ロードシェア テーブル エントリは 15 だけ使用されます。

どのスイッチング パスが最適か

割り込みコンテキストでのスイッチングはプロセスレベルのスイッチングよりも少なくとも 1 桁高速であるため、可能であれば常にルータの割り込みコンテキストを使用してスイッチングしようとするでしょう。Cisco Express Forwarding スwitching は、その他のスイッチング モードよりも高速かつ高品質です。実行中のプロトコルと IOS が Cisco Express Forwarding をサポートしているなら、Cisco Express Forwarding の使用をお勧めします。これは特に、トラフィックがロードシェアリングされるパラレル リンクが複数ある場合に有効です。CEF サポートに必要な IOS を確認するには、[Cisco Feature Navigator](#) ([登録ユーザ専用](#)) のページを使用してください。

関連情報

- [シスコ エクスプレス フォワーディング スwitching の確認方法](#)
- [Cisco エクスプレス フォワーディングを使用したパラレル リンク上のロード バランシングに関するトラブルシューティング](#)
- [CEF でのロード バランシング](#)
- [Cisco エクスプレス フォワーディングに関するサポート ページ](#)
- [IP ルーティングに関するサポート ページ](#)
- [Cisco IOS スwitching サービス設定ガイド、リリース 12.1](#)
- [テクニカルサポート - Cisco Systems](#)