# シスコ エクスプレス フォワーディングの実装


- シスコ エクスプレス フォワーディングの実装 （1 ページ）


# シスコ エクスプレス フォワーディングの実装

Cisco Express Forwarding（CEF）は、拡張レイヤ 3 IP スイッチング テクノロジーです。CEF によって、インターネットや、Web ベースのアプリケーションまたは対話型セッションが集中的に使用されるネットワークなどの、大規模でダイナミックなトラフィック パターンを持つネットワークのパフォーマンスおよびスケーラビリティが最適化されます。CEF は内蔵されている機能であり、イネーブルにするために設定を行う必要はありません。必要に応じて、デフォルトのルート パージ遅延とスタティック ルートを変更できます。

### コンポーネント

Cisco IOS XR ソフトウェアの CEF は常に、次の 2 種類のコンポーネントとともに CEF モードで動作します。

- 転送情報ベース（FIB）データベース：プロトコル依存の FIB プロセスは、ルート プロセッサに IPv4 および IPv6 ユニキャストの転送テーブルを保持します。各ノード上の FIB はルーティング情報ベース（RIB）を更新し、ルート解決を実行してルートプロセッサに個別に FIB テーブルを保持します。各ノード上の FIB テーブルに格納されている情報は、テーブルによって若干異なることがあります。
- 隣接関係テーブル：プロトコルに依存しない隣接情報ベース（AIB）

CEF は、Cisco IOS XR ソフトウェアのプライマリ IP パケット転送データベースです。CEF の役割は次の機能を果たすことです。

- ソフトウェア スイッチング パス
- ソフトウェアおよびハードウェア転送エンジンの転送テーブルおよび隣接関係テーブルのメンテナンス（AIB によるメンテナンス）

Cisco IOS XR ソフトウェア上の CEF では、次の機能をサポートしています。

- バンドル インターフェイスのサポート
- マルチパス サポート
- ルート整合性
- パッケージング、再起動性、リソース不足（OOR）処理などのハイ アベイラビリティ機能
- OSPFv2 SPF プレフィックス優先順位付け
- BGP 属性ダウンロード

**CEF の利点**

- パフォーマンス向上：CEF は、高速スイッチング ルート キャッシングよりも CPU を消費しません。より多くの CPU 処理能力を Quality of Service（QoS）や暗号化などのレイヤ 3 サービスに向けることができます。
- スケーラビリティ：CEF では、各ライン カードでスイッチング機能を最大限に活用できます。
- 復元力：CEF では、大規模な動的ネットワーク上で比類ないレベルのスイッチング一貫性と安定性を実現します。動的ネットワークでは、ルーティング変更のために、高速にスイッチングされるキャッシュ エントリが頻繁に無効化されます。ルーティング変更により、ルート キャッシュを使用した高速スイッチングではなく、ルーティング テーブルを使用したトラフィックのプロセス スイッチングが行われることがあります。転送情報ベース（FIB）ルックアップテーブルにはルーティングテーブルに存在する既知のルートがすべて含まれているため、ルートキャッシュのメンテナンスが不要になるほか、高速スイッチングまたはプロセス スイッチング フォワーディングのシナリオも必要ありません。CEF では、一般的なデマンド キャッシング スキームよりも効率よくトラフィックを切り替えることができます。

Cisco IOS XR ソフトウェアでは、次の CEF 転送テーブルがメンテナンスされます。

- IPv4 CEF データベース：IPv4 ユニキャスト パケット転送用の IPv4 ユニキャスト ルートが保存されます。
- IPv6 CEF データベース：IPv6 ユニキャスト パケット転送用の IPv6 ユニキャスト ルートが保存されます。
- MPLS LFD データベース：MPLS パケット転送用の MPLS ラベル テーブルが保存されます。

# CEF の確認

IPv4 または IPv6 CEF テーブルの詳細を表示するには、次のコマンドを使用します。

- show cef {*ipv4 address* | *ipv6 address*} hardware egress

IPv4 または IPv6 CEF テーブルを表示します。ネクスト ホップおよび転送インターフェイスがプレフィックスごとに表示されます。**show cef** コマンドの出力は、場所によって異なります。

```
Router# show cef 203.0.1.2 hardware egress
  203.0.1.2/32, version 0, internal 0x1020001 0x0 (ptr 0x8d7db7f0) [1], 0x0 (0x8daeedf0),
 0x0 (0x0)
 Updated Nov 20 13:33:23.557
 local adjacency 203.0.1.2
 Prefix Len 32, traffic index 0, Adjacency-prefix, precedence n/a, priority 15
   via 203.0.1.2/32, HundredGigE0/9/0/0, 3 dependencies, weight 0, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0x8cfc81a0 0x0]
    next hop 203.0.1.2/32
    local adjacency
```

- show cef {ipv4 | ipv6} summary

IPv4 または IPv6 CEF テーブルのサマリーを表示します。

```
Router#show cef ipv4 summary
Fri Nov 20 13:50:45.239 UTC

Router ID is 216.1.1.1

IP CEF with switching (Table Version 0) for node0_RP0_CPU0

  Load balancing: L4
  Tableid 0xe0000000 (0x8cf5b368), Vrfid 0x60000000, Vrid 0x20000000, Flags 0x1019
  Vrfname default, Refcount 4129
  56 routes, 0 protected, 0 reresolve, 0 unresolved (0 old, 0 new), 7616 bytes
  13 rib, 0 lsd, 0:27 aib, 1 internal, 10 interface, 4 special, 1 default routes
  56 load sharing elements, 24304 bytes, 1 references
  1 shared load sharing elements, 432 bytes
  55 exclusive load sharing elements, 23872 bytes
  0 route delete cache elements
  13 local route bufs received, 1 remote route bufs received,  0 mix bufs received
  13 local routes, 0 remote routes
  13 total local route updates processed
  0 total remote route updates processed
  0 pkts pre-routed to cust card
  0 pkts pre-routed to rp card
  0 pkts received from core card
  0 CEF route update drops, 0 revisions of existing leaves
  0 CEF route update drops due to version mis-match
  Resolution Timer: 15s
  0 prefixes modified in place
  0 deleted stale prefixes
  0 prefixes with label imposition, 0 prefixes with label information
  0 LISP EID prefixes, 0 merged, via 0 rlocs
 28 next hops
  1 incomplete next hop

 0 PD backwalks on LDIs with backup path
```

- show cef { ipv4 address | ipv6 address } detail

IPv4 または IPv6 CEF テーブルの詳細を表示します。

```
Router#show cef 203.0.1.2 detail
203.0.1.2/32, version 0, internal 0x1020001 0x0 (ptr 0x8d7db7f0) [1], 0x0 (0x8daeedf0),
 0x0 (0x0)
 Updated Nov 20 13:33:23.556
 local adjacency 203.0.1.2
```

```
 Prefix Len 32, traffic index 0, Adjacency-prefix, precedence n/a, priority 15
  gateway array (0x8d84beb0) reference count 1, flags 0x0, source aib (10), 0 backups
                [2 type 3 flags 0x8401 (0x8d99a598) ext 0x0 (0x0)]
  LW-LDI[type=3, refc=1, ptr=0x8daeedf0, sh-ldi=0x8d99a598]
  gateway array update type-time 1 Nov 20 13:33:23.556
 LDI Update time Nov 20 13:33:23.556
 LW-LDI-TS Nov 20 13:33:23.556
   via 203.0.1.2/32, HundredGigE0/9/0/0, 3 dependencies, weight 0, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0x8cfc81a0 0x0]
    next hop 203.0.1.2/32
    local adjacency
        Load distribution: 0 (refcount 2)

    Hash  OK  Interface                 Address
    0     Y   HundredGigE0/9/0/0        203.0.1.2
```

- show adjacency detail

インターフェイスごとのレイヤ 2 情報など詳細な隣接情報を表示します。show adjacency コマンドの出力は、場所によって異なります。

```
Router#show adjacency detail

-------------------------------------------------------------------------------
0/RP0/CPU0
-------------------------------------------------------------------------------
Interface                   Address                  Version  Refcount Protocol
Hu0/9/0/0                  (interface)                  13        1(    0)
                           (interface entry)
                           mtu: 1500, flags 1 4


Hu0/9/0/1                  (interface)                  31        1(    0)
                           (interface entry)
                           mtu: 1500, flags 1 4
```

# フロー単位ロード バランシング

システムは基本的に 7 タプル ハッシュ アルゴリズムをサポートしています。ロード バランシングでは、レイヤ 3（ネットワーク層）およびレイヤ 4（トランスポート層）のルーティング情報に基づいてパケットを複数のリンクに分散させるルータ機能について説明します。ルータが宛先に至るパスを複数検出した場合は、その宛先の複数のエントリでルーティングテーブルが更新されます。

フロー単位のロード バランシングでは、以下の機能が実行されます。

- 着信データ トラフィックは、複数の等コスト接続に均等に分散されます。
- 着信データ トラフィックは、バンドル インターフェイス内の複数の等コスト接続メンバリンクに均等に分散されます。
- レイヤ 2 バンドルとレイヤ 3（ネットワーク レイヤ）のロード バランシングの決定は、IPv4、IPv6、および MPLS フローで行われます。IPv4 または IPv6 のペイロードの場合は、7 タプル ハッシングが実行されます。3 つ以下のラベルが設定された MPLS ペイロードの場合は、ハードウェアが下部のペイロードを解析し、そのペイロードに IPv4 または IPv6 ヘッダーがあるかどうかを特定します。IPv4 または IPv6 ヘッダーの場合は、IP 送信元、

IP 送信先、ルータ ID、およびラベル スタックの基づいて 4 タプル ハッシングが実行されます。それ以外の場合は、MPLS ラベル ベースのハッシングが実行されます。MPLS ラベル ベースのハッシングの場合は、上位 4 ラベルがハッシュの計算に使用されます。 ただし、J2 ライン カードの場合は、すべてのラベルが MPLS ラベルベースのハッシングに使用されます。

- 7 タプル ハッシュ アルゴリズムはより細かなロード バランシングを実現し、複数の等コスト レイヤ3（ネットワーク層）パス全体でのロードバランシングに使用されます。レイヤ3（ネットワーク層）パスは、物理インターフェイスまたはバンドルインターフェイス上にあります。また、メンバ リンクに対するロード バランシングが、レイヤ 2 バンドルインターフェイス内で行われることがあります。
- 7 タプル ロードバランス ハッシュ計算には以下が含まれます。
  - 送信元 IP アドレス
  - 宛先 IP アドレス
  - IP プロトコル タイプ
  - ルータ ID
  - 送信元ポート
  - 宛先ポート
  - 入力インターフェイス

（注） 入力インターフェイスはロードバランスハッシュ計算のパラメータとして考慮されないため、J2 ラインカードでは、6 タプルのロードバランス ハッシュ計算がサポートされています。

### 宛先別ロード バランシング

再帰 MPLS パス（BGP 3107 を通じて学習したパスなど）を介して遷移するパケットには、宛先別ロード バランシングが使用されます。宛先別ロード バランシングとは、ルートの宛先に基づいてルータがパケットを分散します。同じネットワークへのパスが 2 つある場合、そのネットワーク上の宛先 1 へのすべてのパケットが最初のパスを介して移動し、そのネットワーク上の宛先 2 へのすべてのパケットが 2 番目のパスを介して移動するという具合になります。これにより、パケットの順序は保持されますが、リンクの使用が不等になる可能性があります。1 つのホストがトラフィックの大部分を受け取る場合は、すべてのパケットが 1 つのリンクを使用し、他のリンクの帯域幅は使用されないままとなります。宛先アドレスが多数ある場合は、より均等にリンクが使用されます。

## スタティック ルートの設定

ルータは、ユーザが手動で設定したルート テーブル エントリのルート情報を使用するか、またはダイナミック ルーティング アルゴリズムで計算されたルート情報を使用して、パケット

を転送します。スタティック ルートは、2つのルータ間の明示パスを定義するものであり、自動的にはアップデートされません。ネットワークに変更があった場合は、ユーザが手動でスタティック ルートを再設定する必要があります。スタティック ルートは、ダイナミック ルートに比べて使用する帯域幅が少なくなります。スタティック ルートは、ネットワーク トラフィックが予測可能で、ネットワーク設計がシンプルな環境で使用します。スタティック ルートはネットワークの変化に対応できないので、大規模でたえず変化しているネットワークでは、スタティック ルートを使用すべきではありません。大部分のネットワークは、ルータ間の通信にダイナミック ルートを使用しますが、特殊な状況でスタティック ルートを 1 つか 2 つ設定する場合があります。スタティック ルートは、最終手段としてのゲートウェイ（ルーティング不能なすべてのパケットの送信先となるデフォルト ルータ）を指定する場合にも便利です。

### 設定例

HundredGigE インターフェイスを介してルータ A とルータ B 間にスタティック ルートを作成します。宛先 IP アドレスは 203.0.1.2/32、ネクストホップ アドレスは 1.0.0.2 です。

```
Router(config)#router static address-family ipv4 unicast
Router(config-static-afi)#203.0.1.2/32 HundredGigE 0/9/0/0 1.0.0.2
Router(config-static-afi)#commit
```

### 実行コンフィギュレーション

```
Router#show running-config router static address-family ipv4 unicast
router static
 address-family ipv4 unicast
 203.0.1.2/32 HundredGigE 0/9/0/0 1.0.0.2
 !
!
```

### 確認

Next Hop Flags のフィールドに COMPLETE と表示され、設定が適切に機能していることを確認してください。

プレフィックスが更新された LPM、EXT-TCAM、LEM などのデータベースも、出力を通じて提供されます。プレフィックスのスケーリングを把握できるようになるため、ネットワークリソースを効率的に管理できます。また、デバイスに対する特定の IP アドレスの設定が失敗した理由を理解でき、デバッグが簡単になります。

```
Router#show cef 203.0.1.2/32 hardware egress details location 0/0/CPU0
Wed Nov  6 10:09:23.548 UTC
111.0.0.1/32, version 221, attached, internal 0x1000041 0x0 (ptr 0x8b00ea80) [1], 0x0
(0x8afd9768), 0x0 (0x0)
Updated Nov  6 10:08:07.424
Prefix Len 32, traffic index 0, precedence n/a, priority 2
  gateway array (0x8ae4baf0) reference count 1, flags 0x0, source rib (7), 0 backups
                [2 type 3 flags 0x40008441 (0x8af020c0) ext 0x0 (0x0)]
  LW-LDI[type=3, refc=1, ptr=0x8afd9768, sh-ldi=0x8af020c0]
  gateway array update type-time 1 Nov  6 10:08:07.423
LDI Update time Nov  6 10:08:07.423
LW-LDI-TS Nov  6 10:08:07.424
```

```
   via tunnel-ip1, 0 dependencies, recursive [flags 0x8]
    path-idx 0 NHID 0x0 [0x8ae0d728 0x0]
    local adjacency

 LEAF - HAL pd context :
 sub-type : IPV4, ecd_marked:0, has_collapsed_ldi:0
collapse_bwalk_required:0, ecdv2_marked:0,
HW Walk:
LEAF:
    PI:0x308b00ea80 PD:0x308b00eb20 rev:293 type: IPV4 (0)
    LEAF location: LEM
    FEC key: 0x1640000dc6

    LWLDI:
        PI:0x308afd9768 PD:0x308afd97a8 rev:292 p-rev:291  ldi type:IP
        FEC key: 0x1640000dc6 fec index: 0x0(0) num paths:1, bkup paths: 0

REC-SHLDI HAL PD context :
ecd_marked:0, collapse_bwalk_required:0, load_shared_lb:0

    RSHLDI:
        PI:0x308af020c0 PD:0x308af02190 rev:291 dpa-rev:2448 flag:0x1
        FEC key: 0x1640000dc6 fec index: 0x2001ffd8(131032) num paths: 1
        p-rev:
        Path:0 fec index: 0x2001ffd8(131032) DSP fec index: 0x20000001(1),
               TEP Encap Id: 0x40013801

LEAF - HAL pd context :
 sub-type : IPV4, ecd_marked:0, has_collapsed_ldi:0
collapse_bwalk_required:0, ecdv2_marked:0,
HW Walk:
LEAF:
    PI:0x308b00e558 PD:0x308b00e5f8 rev:270 type: IPV4 (0)
    LEAF location: LEM
    FEC key: 0x1240000dc6

    LWLDI:
        PI:0x308afd9128 PD:0x308afd9168 rev:269 p-rev:268 268  ldi type:IP
        FEC key: 0x1240000dc6 fec index: 0x0(0) num paths:2, bkup paths: 0

        SHLDI:
            PI:0x308af00d88 PD:0x308af00e58 rev:268 dpa-rev:2433 cbf_enabled:0
pbts_enabled:0 flag:0x0
            FEC key: 0x1240000dc6 fec index: 0x20000001(1) num paths: 2 bkup paths: 0
            p-rev:265 262
            Path:0 fec index: 0x2001ffdc(131036) DSP:0x15 Dest fec index: 0x0(0)
            Path:1 fec index: 0x2001ffdd(131037) DSP:0x16 Dest fec index: 0x0(0)

            TX-NHINFO:
                PI: 0x308c8d8298 PD: 0x308c8d8318 rev:265 dpa-rev:2431 Encap hdl:
0x308c84c350
                Encap id: 0x40010001 Remote: 0 L3 int: 1552 flags: 0x3
                npu_mask: 0x1 DMAC: f0:78:16:62:f6:a7

            TX-NHINFO:
                PI: 0x308c8d80b0 PD: 0x308c8d8130 rev:262 dpa-rev:2429 Encap hdl:
0x308c84c968
                Encap id: 0x40010000 Remote: 0 L3 int: 1551 flags: 0x3
                npu_mask: 0x1 DMAC: f0:78:16:62:f6:a6


    Load distribution: 0 (refcount 2)
```

```
Hash  OK  Interface                Address
0     Y   tunnel-ip1               10.10.10.1
```

### 関連コマンド

- router static
- show cef

# BGP 属性ダウンロード

BGP 属性ダウンロード機能を使用すると、CEF にインストールした BGP 属性を表示できます。

- **show cef bgp-attribute** コマンドは、CEF にインストールした BGP 属性を表示します。
- 属性 ID とローカル属性 ID で特定の BGP 属性を表示するには、**show cef bgp-attribute** *attribute-id* コマンドと **show cef bgp-attribute** *local-attribute-id* コマンドを使用します。

### 確認

```
Router# show cef bgp-attribute
Router ID is 216.1.1.1

IP CEF with switching (Table Version 0) for node0_RP0_CPU0

  Load balancing: L4
  Tableid 0xe0000000 (0x8cf5b368), Vrfid 0x60000000, Vrid 0x20000000, Flags 0x1019
  Vrfname default, Refcount 4129
  56 routes, 0 protected, 0 reresolve, 0 unresolved (0 old, 0 new), 7616 bytes
  13 rib, 0 lsd, 0:27 aib, 1 internal, 10 interface, 4 special, 1 default routes
  56 load sharing elements, 24304 bytes, 1 references
  1 shared load sharing elements, 432 bytes
  55 exclusive load sharing elements, 23872 bytes
  0 route delete cache elements
  13 local route bufs received, 1 remote route bufs received,  0 mix bufs received
  13 local routes, 0 remote routes
  13 total local route updates processed
  0 total remote route updates processed
  0 pkts pre-routed to cust card
  0 pkts pre-routed to rp card
  0 pkts received from core card
  0 CEF route update drops, 0 revisions of existing leaves
  0 CEF route update drops due to version mis-match
  Resolution Timer: 15s
  0 prefixes modified in place
  0 deleted stale prefixes
  0 prefixes with label imposition, 0 prefixes with label information
  0 LISP EID prefixes, 0 merged, via 0 rlocs
 28 next hops
  1 incomplete next hop

 0 PD backwalks on LDIs with backup path


VRF: default
_____________
Table ID: 0xe0000000. Total number of entries: 0
OOR state: GREEN. Number of OOR attributes: 0
```

### 関連コマンド

- show cef bgp-attribute

## プロアクティブなアドレス解決プロトコルおよびネイバー探索

CEF は、レイヤ 2 隣接関係情報がないルートをインストールすると、不完全なレイヤ 3 ネクストホップを作成してハードウェアにプログラムします。この不完全なプログラミングが原因となり、最初のパケットはソフトウェア転送パスに転送されます。次に、ソフトウェア転送ではレイヤ 2 隣接関係情報を解決するために、パケットからレイヤ 2 ヘッダーを除去して、ARP（アドレス解決プロトコル）または ND（ネイバー探索）に転送します。このようなパケットでレイヤ 2 ヘッダー内に機能固有の情報が含まれていると、ソフトウェア転送パスがレイヤ 2 ヘッダーを完全に除去できないため、ARP または ND は欠落しているレイヤ 2 隣接関係情報を解決できずに、トラフィックがドロップされる結果になります。

プロアクティブな ARP および ND 機能は、欠落しているレイヤ 2 隣接関係情報を解決するために CEF が ARP または ND をプロアクティブにトリガーするようにし、ネクストホップ情報が解決されるまで 15 秒ごとに再試行することで上記の問題を解決します。したがって、不完全なネクストホップ情報を含むスタティック ルートを設定すると、この機能によって ARP または ND の解決が自動的にトリガーされます。