概要

この章の内容は、次のとおりです。

- 概要，1 ページ
- FCoE 初期化プロトコル，2 ページ
- Data Center Bridging Exchange プロトコル，4 ページ
- ロスレス イーサネット，6 ページ

概要

FCoE を使用すると、物理的なイーサネット リンクを介してファイバ チャネル トラフィックをカプセル化できます。 FCoE フレームでは固有のイーサタイプが使用されるため、FCoE トラフィックおよび標準イーサネット トラフィックを同一リンクで伝送できます。

従来のイーサネットはベストエフォート型プロトコルです。輻輳が発生した場合、イーサネットではパケットが廃棄され、再送信など信頼性を確保するための機能は上位プロトコルに委ねられます。 ファイバ チャネル トラフィックにはロスレス トランスポート層が必要です。データ ストレージ プロトコルでは、1 つのデータ パケットが単独で消失することは認められません。 ネイティブ ファイバ チャネルでは、バッファ間クレジット システムによりトランスポート層にロスレス サービスが実装されます。

FCoE トラフィックに対しては、イーサネット リンクによりロスレス サービスを実装する必要があります。 Cisco Nexus 5000 シリーズ スイッチのイーサネット リンクでは、リンク レベル フロー制御（LL-FC）およびプライオリティ フロー制御（PFC）という 2 つのメカニズムにより、FCoE トラフィックのロスレス トランスポートが実現されます。

IEEE 802.3x リンク レベル フロー制御により、輻輳したレシーバは遠端に信号を発信し、データ送信を短時間一時停止させます。 この一時停止機能はリンク上のすべてのトラフィックに適用されます。

プライオリティ フロー制御機能は、イーサネット リンク上の特定のトラフィック クラスにポーズ機能を適用します。 これにより、たとえばFCoE トラフィックに対してはロスレス サービス、標準イーサネット トラフィックに対してはベストエフォート サービスを実現できます。 PFC は、

（IEEE 802.1p トラフィック クラスを使用して）特定のイーサネット トラフィック クラスに、さまざまなレベルのサービスを提供することができます。

Cisco Nexus シリーズ スイッチでは、10 ギガビット イーサネット インターフェイス上で T11 準拠の FCoE がサポートされています。

FCoE 初期化プロトコル

スイッチでは、FCoE Initialization Protocol（FIP; FCoE 初期化プロトコル）により、イーサネット LAN に接続された FCoE 対応エンティティの検出および初期化を実行できます。 Cisco Nexus シリーズ スイッチでサポートされている FIP には、次の 2 つのバージョンがあります。

- FIP：Converged Enhanced Ethernet Data Center Bridging Exchange（CEE-DCBX）プロトコルにより、T11 準拠の Gen-2 CNA がサポートされています。
- Pre-FIP：シスコ、Intel、Nuova の Data Center Bridging Exchange（CIN-DCBX）プロトコルにより、Gen-1 CNA がサポートされています。

Cisco Nexus シリーズ スイッチは、接続されている CNA の機能を検出したうえで、適切な FIP モードに切り替わります。

FIP 仮想リンクのインスタンス化

Cisco NX-OS Release 4.1(3)N1(1) では新たに、Cisco Nexus シリーズ スイッチで T11 準拠の FIP がサポートされています。

FIP は、デバイスの検出、初期化、およびリンクのメンテナンスを実行する際に使用されます。FIP により、次のプロトコルが実行されます。

- FIP 検出：FCoE デバイスがファブリックに接続されている場合、検出要求メッセージが送信されます。 このメッセージに対しては、Fibre Channel Forwarder（FCF; ファイバ チャネル フォワーダ）またはスイッチが、送信要求されたアドバタイズメントで応答します。このアドバタイズメントにより、それ以降のログインに使用する FCF MAC アドレスが取得されます。
- FCoE 仮想リンクのインスタンス化：FIP では、ファブリック ログイン（FLOGI）、ファブリック検出（FDISC）、ログアウト（LOGO）、および交換リンク パラメータ（ELP）の各フレーム、およびそれらに対応する応答フレームのカプセル化が定義されています。 FCoE デバイスでは、これらのメッセージに基づいて、ファブリック ログインが実行されます。
- FCoE 仮想リンクのメンテナンス：FIP では、接続が継続的に有効であることを確認するため、スイッチと CNA との間で定期的にメンテナンス メッセージが送信されます。

FCoE フレームの形式

FCoE は、固有のイーサタイプ 0x8906 を持つイーサネット パケットにファイバ チャネル フレームをカプセル化することによって実装されます。 このパケットには、4 ビットのバージョン フィールドがあります。フレーム内のその他のフィールド（送信元 MAC アドレス、宛先 MAC アドレス、VLAN タグ、およびフレーム マーカー）はすべて、標準のイーサネット フィールドです。予備ビットの追加により、FCoE フレームの長さは IEEE 802.3 の最小パケット長である 64 バイトになります。

ファイバ チャネル フレームは、36 バイトのヘッダーと最大 2,112 バイトのデータで構成され、その合計サイズは最大で 2,148 バイトになります。 カプセル化されたファイバ チャネル フレームは、標準ヘッダーがすべて含まれているため、追加の修正を施すことなくストレージ ネットワークへ渡すことができます。 FCoE フレームのファイバ チャネル フレームを最大に合わせるため、class-fcoe はデフォルトで 2240 バイトの MTU サイズに定義されます。

FCoE フレームの VLAN タギング

スイッチによってアダプタに送信されるイーサネット フレームには、IEEE 802.1Q タグを付加できます。 このタグには、PFC で使用するサービス クラス（CoS）値用のフィールドが含まれます。 また、IEEE 802.1Q タグには VLAN フィールドも含まれます。

FIP の T11 準拠 CNA から送信されるフレームの場合、Cisco Nexus シリーズ スイッチでは FCoE VLAN 用の VLAN タグが付加されているという前提で処理が行われます。 タグが正しく付加されていないフレームは廃棄されます。

pre-FIP CNA から送信されるフレームの場合、スイッチでは FCoE CoS 値を持つプライオリティ タグが付加されているという前提で処理が行われます。 ただし、この CNA から送信されるフレームは、タグが付加されていなくても廃棄されません。

FIP イーサネット フレームの形式

FIP は、固有のイーサタイプ 0x8914 を持つイーサネット パケットにカプセル化されます。 このパケットには、4 ビットのバージョン フィールドがあります。 また FIP パケットには、送信元 MAC アドレスおよび宛先 MAC アドレスのほか、FIP 動作コードや FIP 動作サブコードも含まれています。 次の表は、FIP 動作コードをまとめたものです。

表 *1*：*FIP* 動作コード

FIP 動作コード	FIP サブコード	FIP 動作
0x0001	0x01	検出要求
	0x02	検出アドバタイズメント

FIP 動作コード	FIP サブコード	FIP 動作
0x0002	0x01	仮想リンク インスタンス化要求
	0x02	仮想リンク インスタンス化応答
0x0003	0x01	FIP キープアライブ
	0x02	FIP クリア仮想リンク
0x0004	0x01	FIP VLAN 要求
	0x02	FIP VLAN 通知

Pre-FIP 仮想リンクのインスタンス化

Pre-FIP 仮想リンクのインスタンス化は、Data Center Bridging Exchange（DCBX）プロトコルによるリンクの検出、およびそれに続くファブリック ログインという 2 つのフェーズで構成されます。

Cisco Nexus シリーズ スイッチには、pre-FIP モードで動作する Gen-1 CNA に対して下位互換性があります。

（注） Pre-FIP は、シスコ、Intel、Nuova の Data Center Bridging Exchange（CIN-DCBX）プロトコルとも呼ばれます。

Data Center Bridging Exchange プロトコル

Data Center Bridging Exchange（DCBX）プロトコルは、Link Layer Discovery Protocol（LLDP）を拡張したものです。 DCBX エンドポイントは、要求および Acknowledgment（ACK; 確認応答）メッセージを交換します。 柔軟性については、パラメータは TLV フォーマットで符号化されます。

Cisco Nexus シリーズ スイッチでサポートされている DCBX には、次の 2 つのバージョンがあります。

- CEE-DCBX：Converged Enhanced Ethernet DCBX は、すべての T11 準拠 Gen-2 CNA でサポートされています。
- CIN-DCBX：シスコ、Intel、Nuova の DCBX は、Gen-1 CNA でサポートされています。 CIN-DCBX を使用すると、リンク検出を始め、さまざまな機能を実行できます。

DCBX は、Cisco Nexus シリーズ スイッチと CNA の間の物理イーサネット リンク上で実行されます。 デフォルトでは、DCBX はイーサネット インターフェイスでイネーブルです。 イーサネット インターフェイスがアップすると、スイッチでは CNA との通信が自動的に開始されます。

スイッチと CNA の間で FCoE が通常の動作をしている場合は、DCBX によりリンク エラーの検出が行われます。

また DCBX は、スイッチと CNA の間で機能についてのネゴシエーションを行い、CNA に設定値を送信する場合にも使用します。

Cisco Nexus シリーズ スイッチに接続された CNA は、スイッチから送信された設定値が適用されるようにプログラミングされています。これにより、スイッチに接続されたすべての CNA に対してスイッチから設定値を配布することが可能で、設定エラーが生じる可能性が低くなるだけでなく、CNA の管理が容易になります。

DCBX 機能のネゴシエーション

スイッチと CNA の間では、機能情報および設定値が交換されます。 Cisco Nexus シリーズ スイッチは次の機能をサポートします。

- FCoE：CNA が FCoE 機能をサポートしている場合は、FCoE パケットで使用する IEEE 802.1p CoS 値がスイッチにより送信されます。
- Priority Flow Control（PFC; プライオリティ フロー制御）：アダプタが PFC をサポートしている場合、PFC で使用できる IEEE 802.1p CoS 値がスイッチにより送信されます。
- プライオリティ グループの Type-Length-Value（TLV）
- イーサネット論理リンク アップおよびダウン信号
- pre-FIP CNA 用の FCoE 論理リンク アップ/ダウン信号

次のルールによって、ネゴシエーションの後に機能がイネーブルになるかどうか決定されます。

- スイッチと CNA との間で機能およびその設定値が一致する場合、その機能は有効になります。
- 機能は一致するがその設定値が一致しない場合は、次のようになります。
  - スイッチの設定値を受け入れるよう CNA が設定されている場合は、スイッチの値を使用して機能が有効になります。
  - スイッチの設定値を受け入れるよう CNA が設定されていない場合、機能は無効のままです。
- CNA が DCBX 機能をサポートしていない場合、その機能は無効のままです。
- CNA が DCBX を実装していない場合、すべての機能は無効のままです。

（注） Cisco Nexus シリーズ スイッチでは、アダプタとの PFC ネゴシエーションの結果を手動で上書きする CLI コマンドを使用できます。 インターフェイス単位で、機能を強制的にイネーブルまたはディセーブルにできます。

ロスレス イーサネット

標準のイーサネットは、ベスト エフォート型のメディアであるため、どのような形のフロー制御も備えていません。輻輳や衝突が発生した場合、イーサネットではパケットが廃棄されます。失われたデータの検出および廃棄されたパケットの再送信は、上位プロトコルにより行われます。

ファイバ チャネルを適切にサポートできるよう、イーサネットには Priority Flow Control（PFC; プライオリティ フロー制御）メカニズムが追加されています。

ローカル リンクのアップ/ダウン

次の拡張モジュールを使用すると、ネイティブ ファイバ チャネル ポートを介して、Cisco Nexus 5000 シリーズ スイッチを他のファイバ チャネル デバイスに接続できます。

- N5K-M1404 Cisco Nexus 5000 1000 シリーズ モジュール 4x10GE 4xFC 4/2/1
- N5K-M1008 Cisco Nexus 5000 1000 シリーズ モジュール 8xFC 4/2/1
- N5K-M1060 Cisco Nexus 5000 1000 シリーズ モジュール 6xFC 8/4/2/1

ネイティブ ファイバ チャネル リンクでは、一部の設定アクション（VSAN の変更など）で、インターフェイス ステータスをリセットする必要があります。 インターフェイス ステータスをリセットすると、そのインターフェイスはスイッチによりいったん無効化され、その直後に再び有効化されます。

イーサネット リンクにより FCoE サービスが実装されている場合は、物理リンクをリセットしないでください。リセットすると、そのリンク上のすべてのトラフィックが中断されます。

論理リンク アップ/ダウン機能を使用すると、それぞれの仮想リンクを個別にリセットできます。論理リンク ダウンは、FIP クリア仮想リンク メッセージを受けて実行されます。

pre-FIP CNA の場合は、仮想ファイバ チャネル インターフェイスだけをリセットするように CNA へ要求するための DCBX メッセージがスイッチから送信されます。

（注） 論理リンク レベル アップ/ダウン機能をサポートしていない CNA では、物理リンクがリセットされます。 この場合、イーサネット インターフェイスのすべてのトラフィックが中断されます。

DCBX ベースの FC 論理リンク ステータス シグナリングは、pre-FIP CNA への FCoE セッションにだけ適用されます。

統合型ネットワーク アダプタ

使用できる CNA には、次のようなタイプがあります。

- ハードウェア アダプタ
  - サーバ内で、既存のファイバ チャネル Host Bus Adapter（HBA; ホスト バス アダプタ）ドライバおよびイーサネット Network Interface Card（NIC; ネットワーク インターフェイス カード）ドライバとともに動作します。
  - ネットワークのサーバ オペレーティング システム表示は変更されません。CNA はオペレーティング システムに SAN インターフェイスおよび LAN インターフェイスを提供します。
- FCoE ソフトウェア スタック
  - 既存の 10 ギガビット イーサネット アダプタで動作します。

Cisco Nexus シリーズ スイッチでは、次の 2 世代の CNA がサポートされています。

- FIP アダプタ：スイッチとの間で FIP を介して、使用可能な機能に関する情報を交換するほか、設定可能な値に関するネゴシエーションを行います。
- pre-FIP アダプタ：スイッチとの間で DCBX を介して、使用可能な機能に関する情報を交換するほか、設定可能な値に関するネゴシエーションを行います。

設定エラーを抑制し管理を容易にするため、スイッチから、接続されているすべてのアダプタへ設定データが配布されます。