



Overview

This chapter contains the following sections:

- [Overview, page 1](#)
- [FCoE Initiation Protocol, page 2](#)
- [Data Center Bridging Exchange Protocol, page 4](#)
- [Lossless Ethernet, page 5](#)

Overview

Fibre Channel over Ethernet (FCoE) allows Fibre Channel traffic to be encapsulated over a physical Ethernet link. FCoE frames use a unique EtherType so that FCoE traffic and standard Ethernet traffic can be carried on the same link.

Classic Ethernet is a best-effort protocol; in the event of congestion, Ethernet will discard packets, relying on higher level protocols to provide retransmission and other reliability mechanisms. Fibre Channel traffic requires a lossless transport layer; as a data storage protocol, it is unacceptable to lose a single data packet. Native Fibre Channel implements a lossless service at the transport layer using a buffer-to-buffer credit system.

For FCoE traffic, the Ethernet link must provide a lossless service. Ethernet links on Cisco Nexus devices provide two mechanisms to ensure lossless transport for FCoE traffic: link-level flow control (LL-FC) and priority flow control (PFC).

IEEE 802.3x link-level flow control allows a congested receiver to signal the far end to pause the data transmission for a short period of time. The pause functionality is applied to all the traffic on the link.

The priority flow control feature applies pause functionality to specific classes of traffic on the Ethernet link. For example, PFC can provide lossless service for the FCoE traffic and best-effort service for the standard Ethernet traffic. PFC can provide different levels of service to specific classes of Ethernet traffic (using IEEE 802.1p traffic classes).

Cisco Nexus switches support T11-compliant FCoE on all 10-Gigabit Ethernet interfaces.

FCoE Initiation Protocol

The FCoE Initialization Protocol (FIP) allows the switch to discover and initialize FCoE-capable entities that are connected to an Ethernet LAN. Two versions of FIP are supported by the Cisco Nexus device:

- FIP—The Converged Enhanced Ethernet Data Center Bridging Exchange (CEE-DCBX) protocol supports T11-compliant Gen-2 CNAs.
- Pre-FIP—The Cisco, Intel, Nuova Data Center Bridging Exchange (CIN-DCBX) protocol supports Gen-1 converged network adapters (CNAs).

The Cisco Nexus device detects the capabilities of the attached CNA and switches to the correct FIP mode.

FIP Virtual Link Instantiation

Cisco NX-OS support the T11-compliant FIP on Cisco Nexus devices.

FIP is used to perform device discovery, initialization, and link maintenance. FIP performs the following protocols:

- FIP Discovery—When a FCoE device is connected to the fabric, it sends out a Discovery Solicitation message. A Fibre Channel Forwarder (FCF) or a switch responds to the message with a Solicited Advertisement that provides an FCF MAC address to use for subsequent logins.
- FCoE Virtual Link instantiation—FIP defines the encapsulation of fabric login (FLOGI), fabric discovery (FDISC), logout (LOGO), and exchange link parameters (ELP) frames with the corresponding reply frames. The FCoE devices use these messages to perform a fabric login.
- FCoE Virtual Link maintenance—FIP periodically sends maintenance messages between the switch and the CNA to ensure the connection is still valid.

FCoE Frame Format

FCoE is implemented by encapsulating a Fibre Channel frame in an Ethernet packet with a dedicated EtherType, 0x8906. That packet has a 4-bit version field. The other header fields in the frame (the source and destination MAC addresses, VLAN tags, and frame markers) are all standard Ethernet fields. Reserved bits pad the FCoE frame to the IEEE 802.3 minimum packet length of 64 bytes.

A Fibre Channel frame consists of 36 bytes of headers and up to 2112 bytes of data for a total maximum size of 2148 bytes. The encapsulated Fibre Channel frame has all the standard headers, which allow it to be passed to the storage network without further modification. To accommodate the maximum Fibre Channel frame in an FCoE frame, the class-foe is defined with a default maximum transmission unit (MTU) of 2240 bytes.

VLAN Tagging for FCoE Frames

The Ethernet frames that are sent by the switch to the adapter might include the IEEE 802.1Q tag. This tag includes a field for the class of service (CoS) value used by the priority flow control (PFC). The IEEE 802.1Q tag also includes a VLAN field.

The Cisco Nexus device expects frames from a FIP T11-compliant CNA to be tagged with the VLAN tag for the FCoE VLAN. Frames that are not correctly tagged are discarded.

The switch expects frames from a pre-FIP CNA to be priority tagged with the FCoE CoS value. The switch will still accept untagged frames from the CNA.

FIP Ethernet Frame Format

FIP is encapsulated in an Ethernet packet with a dedicated EtherType, 0x8914. The packet has a 4-bit version field. Along with the source and destination MAC addresses, the FIP packet also contains a FIP operation code and a FIP operation subcode. The following table describes the FIP operation codes.

Table 1: FIP Operation Codes

FIP Operation Code	FIP Subcode	FIP Operation
0x0001	0x01	Discovery Solicitation
	0x02	Discovery Advertisement
0x0002	0x01	Virtual Link Instantiation Request
	0x02	Virtual Link Instantiation Reply
0x0003	0x01	FIP Keep Alive
	0x02	FIP Clear Virtual Links
0x0004	0x01	FIP VLAN Request
	0x02	FIP VLAN Notification

Pre-FIP Virtual Link Instantiation

Pre-FIP virtual link instantiation consists of two phases; link discovery using the Data Center Bridging Exchange protocol (DCBX), which is followed by Fabric Login.

The Cisco Nexus device is backward compatible with Gen-1 CNAs that operate in pre-FIP mode.



Note

Pre-FIP is also known as the Cisco, Intel, Nuova Data Center Bridging Exchange (CIN-DCBX) protocol.

Data Center Bridging Exchange Protocol

The Data Center Bridging Exchange (DCBX) protocol is an extension of the Link Layer Discovery Protocol (LLDP). DCBX end points exchange request and acknowledgment messages. For flexibility, parameters are coded in a type-length-value (TLV) format.

The Cisco Nexus device supports two versions of DCBX:

- CEE-DCBX—The Converged Enhanced Ethernet DCBX is supported on all T11-compliant Gen-2 CNAs.
- CIN-DCBX—The Cisco, Intel, Nuova DCBX is supported on Gen-1 converged network adapters (CNAs). CIN-DCBX is used to perform link detection in addition to other functions.

DCBX runs on the physical Ethernet link between the Cisco Nexus device and the CNA. By default, DCBX is enabled on Ethernet interfaces. When an Ethernet interface is brought up, the switch automatically starts to communicate with the CNA.

During the normal operation of FCoE between the switch and the CNA, DCBX provides link-error detection.

DCBX is also used to negotiate capabilities between the switch and the CNA and to send configuration values to the CNA.

The CNAs that are connected to a Cisco Nexus device are programmed to accept the configuration values sent by the switch, allowing the switch to distribute configuration values to all attached CNAs, which reduces the possibility of configuration errors and simplifies CNA administration.

DCBX Feature Negotiation

The switch and CNA exchange capability information and configuration values. The Cisco Nexus devices support the following capabilities:

- FCoE—If the CNA supports FCoE capability, the switch sends the IEEE 802.1p CoS value to be used with FCoE packets.
- Priority Flow Control (PFC)—If the adapter supports PFC, the switch sends the IEEE 802.1p CoS values to be enabled with PFC.
- Priority group type-length-value (TLV).
- Ethernet logical link up and down signal.
- FCoE logical link up and down signal for pre-FIP CNAs.

The following rules determine whether the negotiation results in a capability being enabled:

- If a capability and its configuration values match between the switch and the CNA, the feature is enabled.
- If a capability matches, but the configuration values do not match, the following occurs:
 - If the CNA is configured to accept the switch configuration value, the capability is enabled using the switch value.
 - If the CNA is not configured to accept the switch configuration value, the capability remains disabled.

- If the CNA does not support a DCBX capability, that capability remains disabled.
- If the CNA does not implement DCBX, all capabilities remain disabled.

**Note**

The Cisco Nexus device provides CLI commands to manually override the results of the PFC negotiation with the adapter. On a per-interface basis, you can force capabilities to be enabled or disabled.

Lossless Ethernet

Standard Ethernet is a best-effort medium which means that it lacks any form of flow control. In the event of congestion or collisions, Ethernet drops packets. The higher level protocols detect the missing data and retransmit the dropped packets.

To properly support Fibre Channel, Ethernet has been enhanced with a priority flow control (PFC) mechanism.

Logical Link Up/Down

On a native Fibre Channel link, some configuration actions (such as changing the VSAN) require that you reset the interface status. When you reset the interface status, the switch disables the interface and then immediately reenables the interface.

If an Ethernet link provides FCoE service, do not reset the physical link because this action is disruptive to all traffic on the link.

The logical link up/down feature allows the switch to reset an individual virtual link. The logical link down is signaled with a FIP Clear Virtual Link message.

For pre-FIP CNAs, the switch sends a DCBX message to request the CNA to reset only the virtual Fibre Channel interface.

**Note**

If the CNA does not support the logical link level up/down feature, the CNA resets the physical link. In this case, all traffic on the Ethernet interface is disrupted.

DCBX-based FC Logical Link Status signaling applies only to FCoE sessions to pre-FIP CNAs.

Converged Network Adapters

The following types of CNAs are available:

- Hardware adapter
 - Server operating system view of the network is unchanged; the CNA presents a SAN interface and a LAN interface to the operating system.
- FCoE software stack
 - Runs on existing 10-Gigabit Ethernet adapters.

Two generations of CNAs are supported by the Cisco Nexus device:

- A FIP adapter uses the FIP to exchange information about its available capabilities and to negotiate the configurable values with the switch.
- A pre-FIP adapter uses DCBX to exchange information about its available capabilities and to negotiate the configurable values with the switch.

To reduce configuration errors and simplify administration, the switch distributes the configuration data to all the connected adapters.