Data Center Bridging

At-A-Glance

Extending Ethernet's Capabilities in the Data Center

The Data Center Bridging (DCB) architecture is based on a collection of open standards Ethernet extensions developed through the IEEE 802.1 working group to improve and expand Ethernet networking and management capabilities in the data center. It helps ensure delivery over lossless fabrics and I/O convergence onto a unified fabric. Each element of this architecture enhances the Cisco® Data Center Bridging implementation and creates a robust Ethernet infrastructure to meet data center requirements now and in the future. Table 1 lists the main features and benefits of the DCB architecture.

IEEE DCB builds on classical Ethernet's strengths, adds several crucial extensions to provide the next-generation infrastructure for data center networks, and delivers unified fabric. This document describes how each of the main features of the DCB architecture contributes to a robust Ethernet network capable of meeting today's growing application requirements and responding to future data center network needs.

Table 1. DCB Features and Benefits

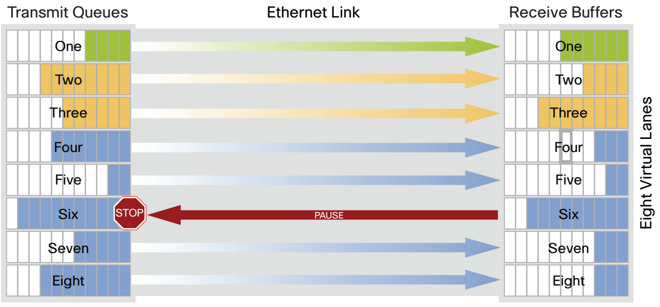
Feature	Benefit
Priority-based Flow control (PFC; IEEE 802.1 Qbb)	Provides capability to manage bursty, single traffic source on a multiprotocol link
Enhanced transmission selection (ETS; IEEE 802.1 Qaz)	Enables bandwidth management between traffic types for multiprotocol links
Congestion notification (IEEE 802.1 Qau)	Addresses the problem of sustained congestion by moving corrective action to the network edge
Data Center Bridging Exchange (DCBX) Protocol	Allows autoexchange of Ethernet parameters between switches and endpoints

Priority-Based Flow Control: IEEE 802.1Qbb

Link sharing is critical to I/O consolidation. For link sharing to succeed, large bursts from one traffic type must not affect other traffic types, large queues of traffic from one traffic type must not starve other traffic types' resources, and optimization for one traffic type must not create high latency for small messages of other traffic types. The Ethernet pause mechanism can be used to control the effects of one traffic type on another. PFC is an enhancement to the pause mechanism. PFC enables pause based on user priorities or classes of service. A physical link divided into eight virtual links (Figure 1) with PFC provides the capability to use pause on a single virtual link without affecting traffic on the other virtual links. Enabling pause based on user priority allows administrators to create lossless links for traffic requiring no-drop service, such as Fibre Channel over Ethernet (FCoE), while retaining packet-drop congestion management for IP traffic.

The current Ethernet pause option stops all traffic on a link; essentially, it is a link pause for the entire link. However, PFC creates eight separate virtual links on the physical link and allows any of these links to be paused and restarted independently. This approach enables the network to create a no-drop class of service for an individual virtual link that can coexist with other traffic types on the same interface. PFC allows differentiated quality-of-service (QoS) policies for the eight unique virtual links. PFC also plays a primary role when used with an arbiter for intraswitch fabrics, linking ingress ports to egress port resources.

Figure 1. Priority-based Flow Control



Enhanced Transmission Selection: IEEE 802.1Qaz

ETS enables optimal bandwidth management of virtual links

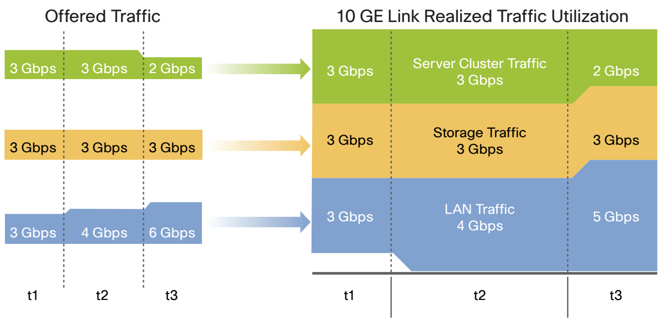
PFC can create eight distinct virtual link types on a physical link, and it can be advantageous to have different traffic classes defined within each virtual link. Traffic within the same PFC IEEE 802.1p class can be grouped together and yet treated differently within each group. ETS provides prioritized processing based on bandwidth allocation, low latency, or best effort, resulting in per-group traffic class allocation. Extending the virtual link concept, the network interface controller (NIC) provides virtual interface queues: one for each traffic class. Each virtual interface queue is accountable for managing its allotted bandwidth for its traffic group, but has flexibility within the group to dynamically manage the traffic. For example, virtual link 3 for the IP class of traffic may have a high-priority designation and a best effort within that same class, with the virtual link 3 class sharing a percentage of the overall link with other traffic classes. ETS allows differentiation among traffic of the same priority class, thus creating priority groups. Figure 2 illustrates 10 Gigabit Ethernet traffic utilization of three traffic classes with different priorities. Each class is assigned a specified bandwidth percentage. In time slot t3, the LAN traffic can use the spare bandwidth available.

Today's IEEE 802.1p implementation specifies a strict scheduling of queues based on priority. With ETS, a flexible, drop-free scheduler for the queues can prioritize traffic according to the IEEE 802.1p traffic classes and the traffic treatment hierarchy designated within each priority group. The capability to apply differentiated treatment to different traffic within the same priority class is enabled by implementing ETS (see IEEE 802.1Qaz at http://www.ieee802.org/1/pages/802.1az.html).

Data Center Bridging Exchange Protocol

DCBX is a discovery and capability exchange protocol to discover peers and exchange configuration information between DCB–compliant bridges (Figure 3).

Figure 2. Enhanced Transmission Selection



DCBX capabilities include:

- DCB peer discovery
- Mismatched configuration detection
- DCB link configuration of peers

The following parameters of IEEE DCB can be exchanged with DCBX (see http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overview-rev0.2.pdf):

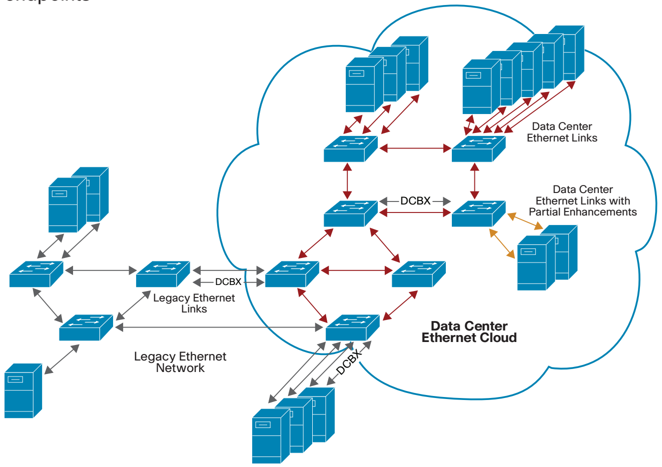
- Priority groups in ETS
- PFC
- Congestion Notification
- Applications
- Logical link-down
- Network interface virtualization

Congestion Notification: IEEE 802.1Qau

Congestion Notification is a Layer 2 traffic management system that pushes congestion to the edge of the network by instructing rate limiters to shape the traffic causing the congestion. The IEEE 802.1Qau working group accepted the Cisco proposal for Congestion Notification, which defines an architecture for actively managing traffic flows to avoid traffic jams.

Congestion is measured at the congestion point, and if congestion is encountered, rate limiting, or back pressure, is imposed at the reaction point to shape traffic and reduce the effects of the congestion on the rest of the network. In this architecture, an aggregation-level switch can send control frames to two access-level switches asking them to throttle back their traffic (Figure 4). This approach maintains the integrity of the network's core and affects only the parts of the network causing the congestion, close to the source (see IEEE 802.1Qau at http://www.ieee802.org/1/pages/802.1au.html).

Figure 3. Data Center Bridging Exchange Protocol allows autoexchange of Ethernet Parameters and Discovery Functions between switches and endpoints

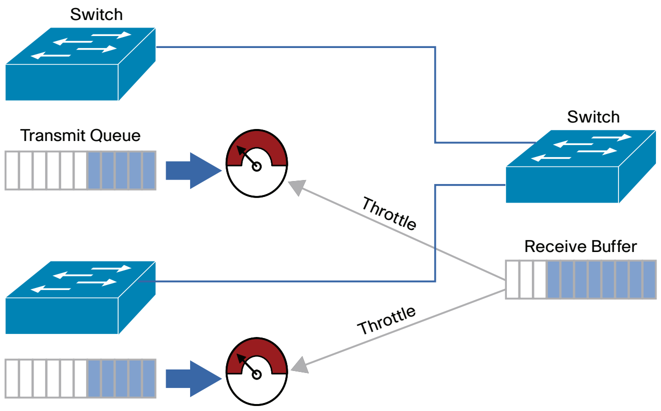


Cisco Unified Fabric Related Standards and Enhancements

In addition to IEEE DCB, Cisco Nexus® data center switches include enhancements such as FCoE and a lossless fabric to enable construction of a unified fabric.

- FCoE: FCoE transports native Fibre Channel frames over Ethernet with existing Fibre Channel management modes intact. A prerequisite for FCoE is a lossless underlying network fabric. FCoE benefits from several of the Cisco DCB extensions to manage congestion, handle bursts of traffic, and support multiple flows on the same cable to achieve unified I/O.
- Unified I/O: DCB supports the concept of running multiple traffic types (LAN, SAN, server cluster traffic, etc.) on a single network while preserving respective traffic treatments. A consolidated I/O link, or unified I/O, can deliver multiprotocol traffic to a unified fabric on a single cable. Cisco Unified Fabric is a single, multipurpose Ethernet transport that can transmit LAN and SAN traffic across a common interface and switch fabric, preserving differentiated classes of service.

Figure 4. Congestion Notification



For More Information

- DCB: http://www.cisco.com/en/US/netsol/ns783/index.html
- PFC (IEEE 802.1Qbb): http://www.ieee802.org/1/pages/802.1bb.html
- ETS (IEEE 802.1Qaz): http://www.ieee802.org/1/pages/802.1az.html
- Congestion Notification (IEEE 802.1Qau): http://www.ieee802.org/1/pages/802.1au.html
- DCBX: http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbcxp-overviewrev0.2.pdf
- IEEE DCB task group: http://www.ieee802.org/1/pages/dcbridges.html