Technical Report

Unified Connect Overview and Best Practices

Patrick Strick, Frank Pleshe, NetApp
August 2011 | TR-3888
Version 1.5

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

In data centers today, a variety of applications and technologies have been deployed to meet the growing demands of the business environment. For example, the financial department might need access to a database application that uses a server that connects to back-end disk storage using NFS. Those same users might also require access to Microsoft® Exchange servers that utilize iSCSI connections to their associated storage. Other business applications and hosts might require Fibre Channel attached storage to utilize high-performance disk drives. As each application has been deployed, different technologies and physical infrastructures have been utilized to deliver the required functionality. Some of the solutions have been able to utilize a common physical network infrastructure such as Ethernet, while others require dedicated hardware switches such as Fibre Channel.

Ethernet has long been used to carry network protocols such as NFS, CIFS, and iSCSI with TCP/IP as the transport mechanism. Both Ethernet and Fibre Channel have required companies to deploy completely separate network infrastructure to support these technologies. Deploying dissimilar technologies to serve multiple applications ultimately increases capital expenditure, operating expense, and operational management to maintain these environments.

Significant savings can be realized by consolidating resources through the use of a single network infrastructure. A new technology called Fibre Channel over Ethernet (FCoE) enables the use of a single network infrastructure that supports traditional Ethernet connections as well as enabling Fibre Channel to be carried over the same interface and Ethernet switching equipment.

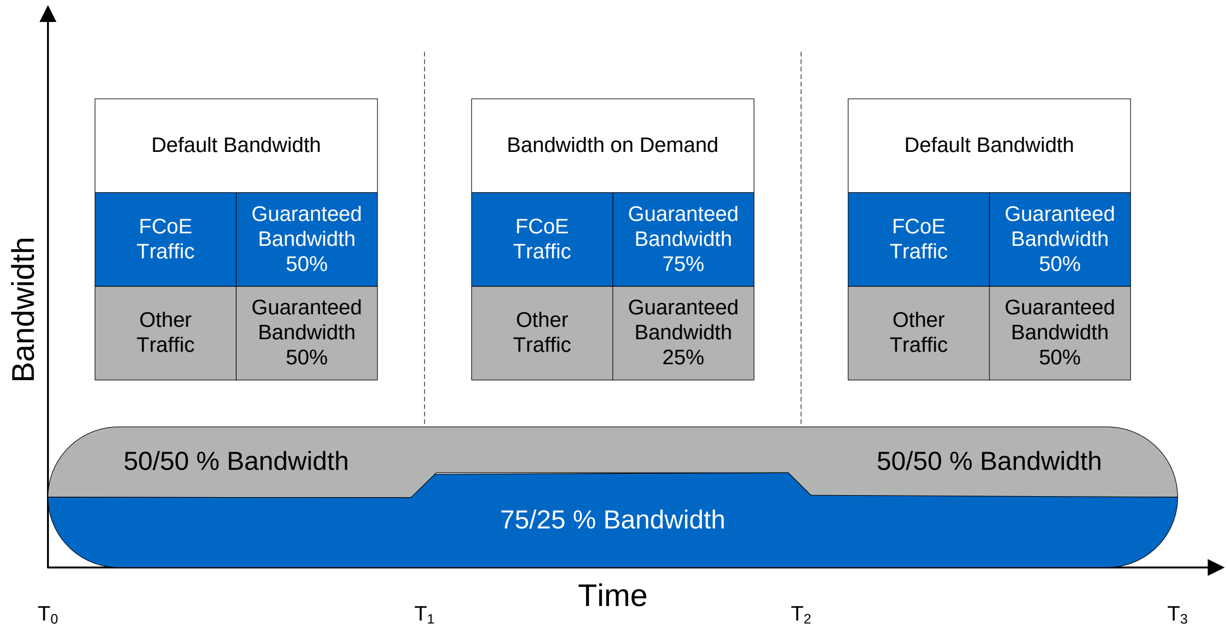
NetApp has adopted the use of this new technology by adding IP (iSCSI/NFS/CIFS) support to the existing FCoE protocol and FCoE adapter support in Data ONTAP® 8.0.1 to enable a single "unified connect" environment within data centers.

2 DATA CENTER BRIDGING

FCoE and converged Ethernet are possible due to enhancements made to the Ethernet protocol, collectively referred to as Data Center Bridging(DCB). DCB enhancements include bandwidth allocation and flow control based on traffic classification and end-to-end congestion notification. Discovery and configuration of DCB capabilities are performed using Data Center Bridging Exchange (DCBX) over LLDP.

Bandwidth allocation is performed with enhanced transmission selection (ETS), which is defined in the IEEE 802.1Qaz standard. Traffic is classified into one of eight groups (0-7) using a field in the Ethernet frame header. Each class is assigned a minimum available bandwidth. If there is competition or oversubscription on a link, each traffic class will get at least its configured amount of bandwidth. If there is no contention on the link, any class can use more or less than it is assigned.

Figure 1) ETS example.



Priority-based flow control (PFC) provides link-level flow control that operates on a per-priority basis. It is similar to 802.3x PAUSE, except that it can pause an individual traffic class. This provides a network with no loss due to congestion for those traffic classes that use PFC. Not all traffic needs PFC. Normal TCP traffic provides its own flow control mechanisms based on window sizes. Because the Fibre Channel protocol expects a lossless medium, FCoE has no built-in flow control and requires PFC to give it a lossless link layer. PFC is defined in the 802.1Qbb standard.

Figure 2) PFC with buffer near full.

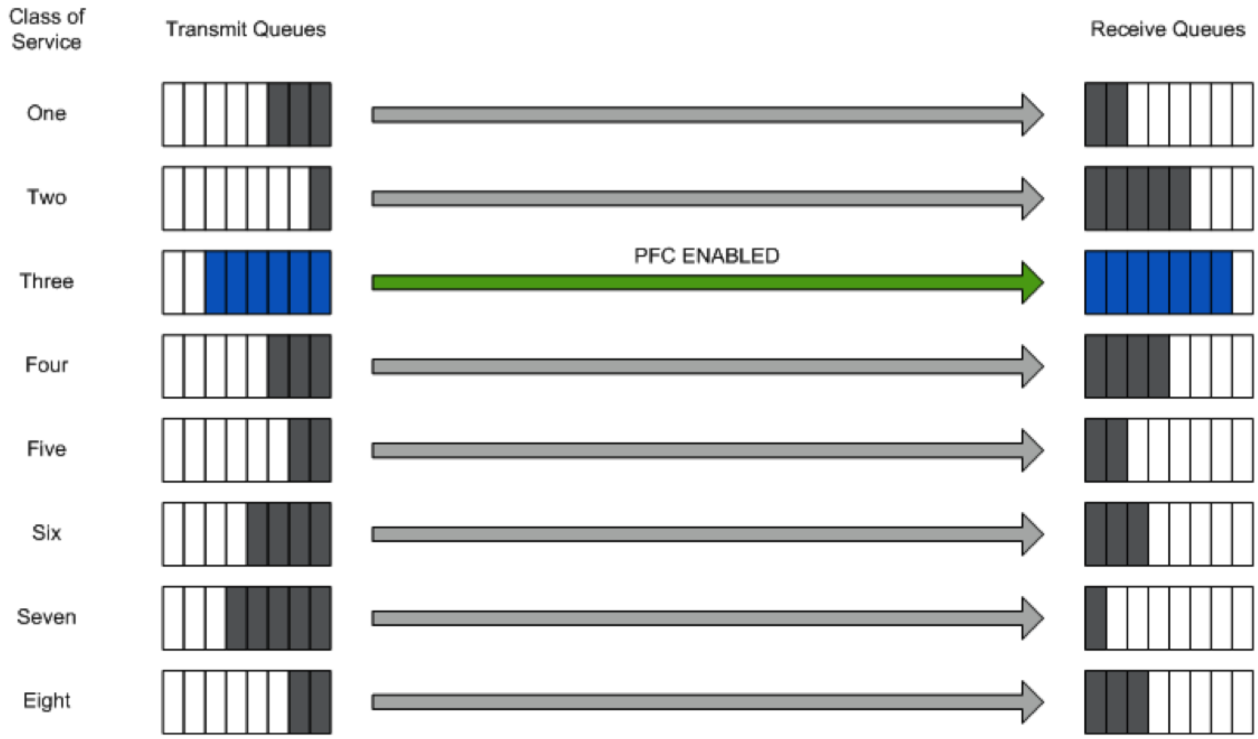
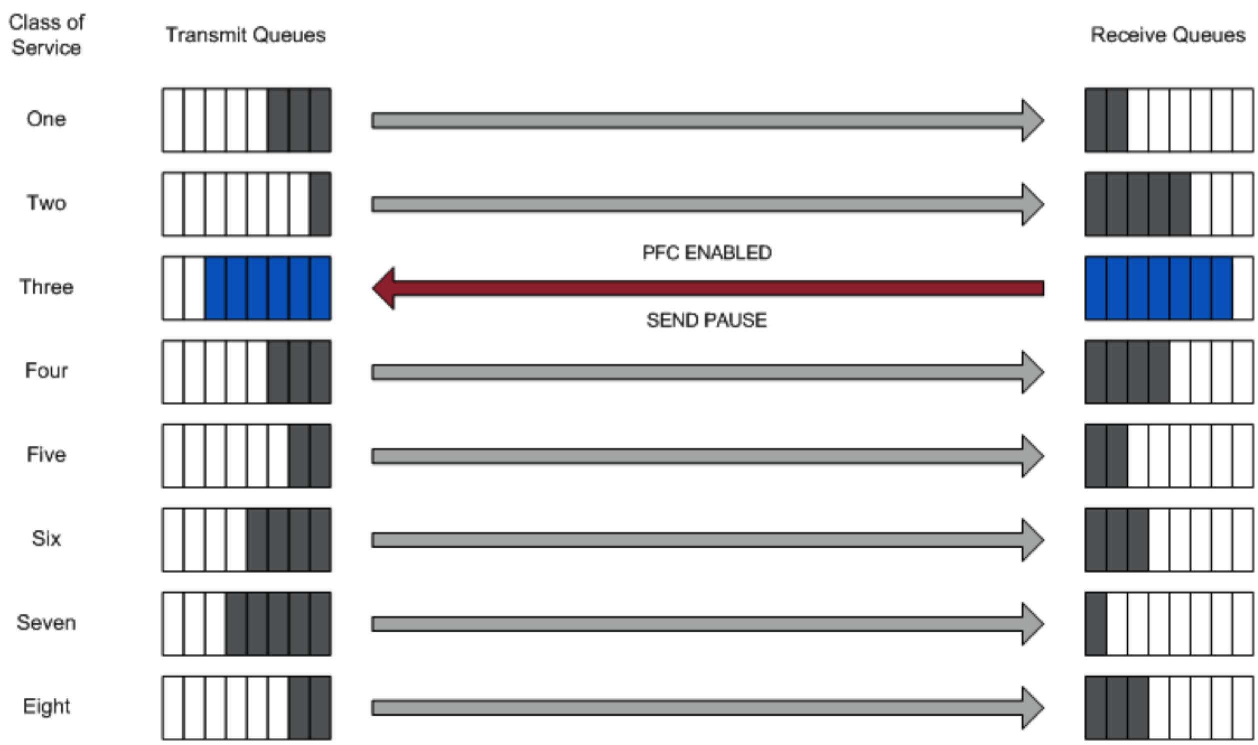


Figure 3) PFC sending pause.



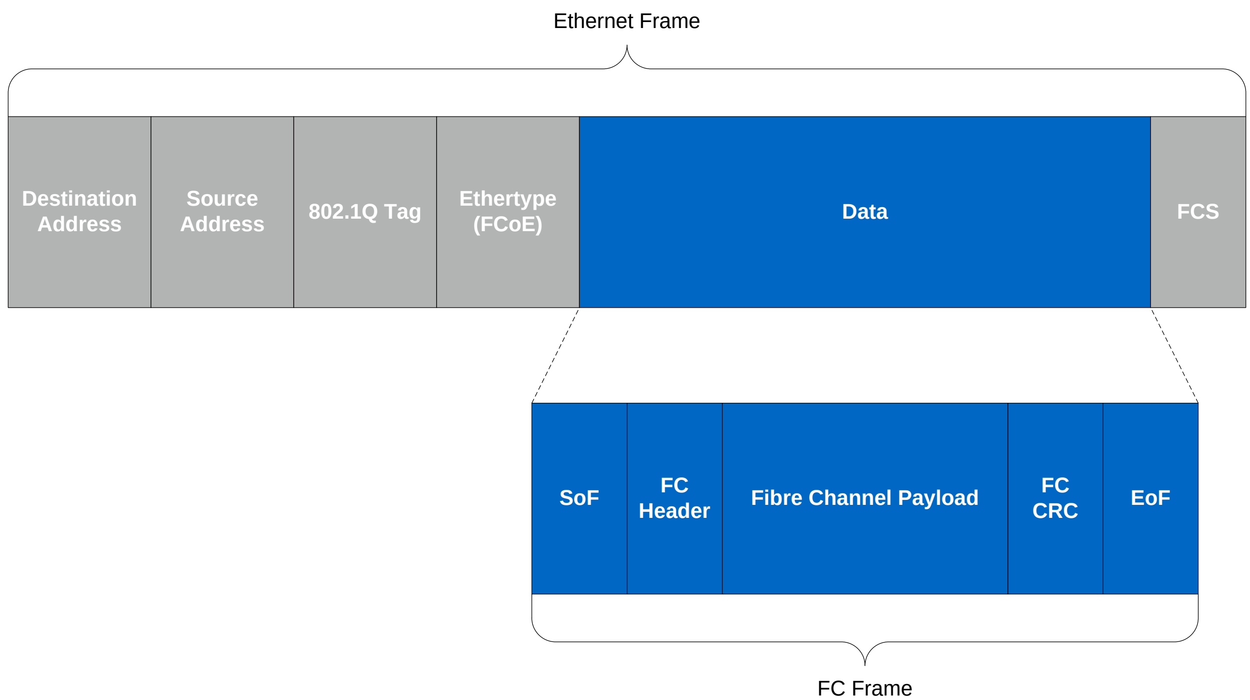
ETS and PFC values are generally configured on the DCB-capable switch and pushed out to the end nodes. For ETS, the sending port controls the bandwidth allocation for that segment of the link (initiator to switch, switch to switch, or switch to target). With PFC, the receiving port sends the per-priority pause, and the sending port reacts by not sending traffic for that traffic class out of the port that received the pause.

Congestion notification (CN) will work with PFC to provide a method for identifying congestion and notifying the source of the traffic flow (not just the sending port). The source of the traffic could then scale back sending traffic going over the congested links. This was developed under 802.1Qau, but is not yet implemented in production hardware.

3 FCOE

Fibre Channel over Ethernet (FCoE) is a SAN transport protocol that allows FC frames to be encapsulated and sent over a DCB capable Ethernet network. For this to be possible, the Ethernet network must meet certain criteria; specifically, it must support DCB.

Figure 4) FCoE frame example.



Because the FC frames are transported with the FC header all encapsulated in the Ethernet frame, movement of data between an Ethernet network and traditional Fibre Channel fabric is simple. Also, since the FC frames are being transported over Ethernet, the nodes and switches do not have to be directly connected. In fact, the FCoE standard was written to account for one or more DCB-capable switches to be in place between a node and an FCoE switch. Both of these points provide a great amount of flexibility in designing an FCoE storage solution.

A Fibre Channel frame can be up to 2,148 bytes. including the header. Consider that a standard Ethernet frame has only 1,500 bytes available for data, and it is obvious that a larger frame is needed. Luckily Ethernet frame sizes greater than 1,500 bytes have been available on many networking devices for some time now to improve performance of high-bandwidth links. For FCoE, jumbo frames are required, and all FCoE devices must support “baby jumbo” frames of 2,240 bytes. That is the maximum FC frame size plus related Ethernet overhead.

Because traditional Fibre Channel expects a highly reliable transport, the protocol does not have any built-in flow control mechanisms. In traditional FC, the transport layer with buffer-to-buffer credits handles flow control. TCP/IP traffic assumes an unreliable transport and utilizes TCP’s adjustable window size and allows retransmits to make sure that all data is transferred. Therefore, a means of making sure of the reliable transport of all FCoE frames had to be established.

Ethernet does have 802.3X PAUSE flow control (defined in 802.3 Annex 31B), but it acts on all traffic coming in on the link. The lack of granularity prevents it from being suitable for a converged network of FCoE and other traffic. The DCB working group addressed this gap with the enhancements described in the DCB section.

The general process by which FCoE is initialized is called FCoE Initialization Protocol (FIP). Before going into the process, we will first go over FCoE-specific terms:

- Converged network adapter (CNA): A unified adapter that acts as both an FCoE initiator and a standard network adapter.
- ENode: A Fibre Channel initiator or target that is able to transmit FCoE frames.
- ENode MAC address: The Ethernet MAC address used by the ENode for FIP.
- FCoE forwarder (FCF): A Fibre Channel switch that is able to process FCoE frames.
- FCoE: Fibre Channel over Ethernet.
- FIP: FCoE Initialization Protocol.
- Fabric-provided MAC address (FPMA): FPMA or SPMA is the FIP MAC address of the ENode.
- Unified target adapter (UTA): An adapter used in a NetApp® storage array that provides FCoE target ports and standard network ports.
- Virtual E_Port (VE_Port): Used to connect two FCFs using FCoE.
- Virtual F_Port (VF_Port): The port on an FCF to which a VN_Port connects.
- Virtual N_Port (VN_Port): The port on an end node used for FCoE communication.

When a node (target or initiator) first connects to an FCoE network, it does so using its ENode MAC address. This is the MAC address associated with its physical, lossless Ethernet port. The first step is DCB negotiation. Once the ETS, PFC, and other parameters are configured, the ENode sends a FIP VLAN request to a special MAC address that goes to all FCFs. Available FCFs respond indicating the VLANs on which FCoE services are provided.

Now that the ENode knows which VLAN to use, it sends a discovery solicitation to the same ALL-FCF-MACS address to obtain a list of available FCFs and whether those FCFs support FPMA. FCFs respond to discovery solicitations, and they also send out discovery advertisements periodically.

The final stage of FIP is for the ENode to log into an FCF (FLOGI). During this process, the ENode is assigned a FIP MAC address. This is the MAC address that will be used for all traffic carrying Fibre Channel payloads. The address is assigned by the FCF (FPMA).

4 CONFIGURATION RECOMMENDATIONS FOR A UNIFIED CONNECT INFRASTRUCTURE

- Evaluate bandwidth needs of all traffic sharing the converged network to determine how much is needed for FCoE traffic and how much will be needed for other types of Ethernet traffic.
- Configure ETS and PFC settings on the switches so that all nodes share the same configuration.
- When connecting multiple DCB-capable switches, configure all switches with the same DCB settings.
- Set ETS bandwidth allocation for FCoE to accommodate the minimum acceptable throughput for all SAN traffic that will utilize a link. For example, if 10 hosts connect to a single FCoE switch and SAN storage device from that switch, determine the minimum acceptable combined throughput of all 10 hosts. This will be your ETS setting for the FCoE traffic class. Because the ETS allocation value only sets the minimum available bandwidth, if more throughput is needed during spikes of traffic, it can be used as long as it is available. Likewise, if FCoE traffic is not utilizing the amount allocated, other Ethernet traffic can take advantage of the remainder.
- You must configure a dedicated VLAN for each VSAN within the FCoE-capable switch.
- A separate multiple spanning tree (MST) instance should be configured for each VSAN.
- Unified ports must be configured as IEEE 802.1Q interfaces on the DCB-capable switch.
- To use link aggregation on the NIC ports of a UTA that is also serving FCoE, each link must be to a separate switch and the switches must support a form of multi chassis link aggregation. See TR-3800 for an example.
- Use TR-3800, “Fibre Channel over Ethernet (FCoE) End-to-End Deployment Guide,” for FCoE setup and configuration.
- Follow TR-3802, “Ethernet Storage Best Practices.”

5 REFERENCES

- TR-3800, "Fibre Channel over Ethernet (FCoE) End-to-End Deployment Guide": https://fieldportal.netapp.com/viewcontent.asp?qv=1&docid=21907
- TR-3802, "Ethernet Storage Best Practices": https://fieldportal.netapp.com/viewcontent.asp?qv=1&docid=23507

NetApp provides no representations or warranties regarding the accuracy, reliability, or serviceability of any information or recommendations provided in this publication or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein are a customer’s responsibility and depend on the customer’s ability to evaluate and integrate them into the customer’s operational environment. This document and the information contained herein may be used solely in connection with the NetApp products discussed in this document.

Go further, faster®

NetApp®

www.netapp.com

© Copyright 2010 NetApp, Inc. All rights reserved. No portions of this document may be reproduced without prior written consent of NetApp, Inc. Specifications are subject to change without notice. NetApp, the NetApp logo, Go further, faster, and Data ONTAP are trademarks or registered trademarks of NetApp, Inc. in the United States and/or other countries. Microsoft is a registered trademark of Microsoft Corporation. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such. TR-3888