White Paper

Introduction to Fibre Channel over Ethernet (FCoE)

A Detailed Review

Abstract

Fibre Channel over Ethernet (FCoE) is a new storage networking option that is transitioning from standards creation to deployment in real world environments. This white paper provides an overview of FCoE, describes the hardware and software components that make up the new ecosystem, and explains how the technology is expected to mature over the next few years.

November 2008

Copyright © 2008 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com

All other trademarks used herein are the property of their respective owners.

Part Number h5916

Table of Contents

Executive summary

The question for most data center managers is not *should* they network their storage, but which of the many options available should they use and for which applications can they justify the expense. Using networked storage reduces not only capital expenses through higher utilization of resources and the ability to create highly available configurations, but also operational expenses by allowing centralized management and simplified backup and replication. While most companies are using networked storage, overall less than 20 percent[1] of all data center servers are attached to networked storage. Virtualized servers have over an 80 percent[2] network attach rate since many advanced features such as mobility and high availability cannot be utilized if these servers are not using networked storage. It is expected that as the overall trend of server virtualization increases that the percentage of servers attached to networked storage in the data center will increase.

One of the reasons why more servers do not utilize networked storage is that the storage network (Fibre Channel) is separate from the Ethernet (TCP/IP) network (see Figure 1). Fibre Channel is the predominate choice for networked storage in the data center because it allows for the creation of a highly scalable – hundreds or thousands of nodes – and reliable solution that can be centrally managed. For networking, today's servers typically have multiple 1 Gb Ethernet network interface cards (NICs). The number of NICs in a system will vary depending on the applications on the machine, and management and security requirements, but the number can be four, six, eight, or more NICs in a single machine. This high number of server adapters, cables, and separate networking and storage fabrics adds to complexity and cost. Ideally, customers would like to be able to have all applications run over a single converged network. This goal of I/O consolidation on a unified network for all traffic leads to a savings in infrastructure (cabling, reduced sparing, and so on) and a simplified management environment. Overall data center trends require solutions that take up less space, draw less power, and require less cooling; suppliers are reaching a breaking point where supporting multiple separate networks will not allow them to meet these requirements.

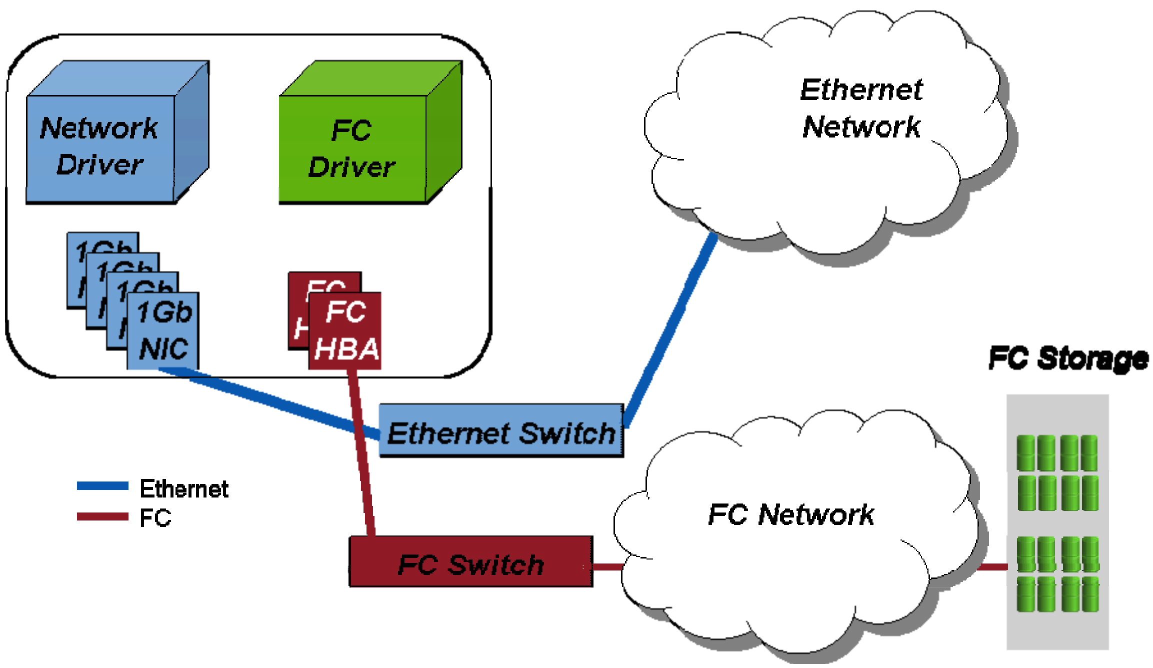

Figure 1. Existing data center with separate Ethernet and FC networks

[1] IDC, *Worldwide Fibre Channel Host Bus Adapter 2007–2011 Forecast and 2006 Vendor Shares* and *Worldwide and Regional Server 2007-2011 Forecast*
[2] Enterprise Storage Group, *Server Virtualization: The Impact on Storage*, November 7, 2007

Ethernet has been given top consideration as a potential solution for I/O consolidation due to the large install base and broad general understanding of the technology by all IT departments. While some storage applications can run comfortably with 1 Gb bandwidth available with Ethernet, many data center solutions require the higher bandwidth solutions that Fibre Channel can supply today with 2 Gb, 4 Gb, and 8 Gb options. With the maturation of 10 Gb Ethernet, however, there is an inflection point where we have the opportunity to consolidate the many existing connections in a server to a pair of 10 Gb Ethernet links. Blade servers and server virtualization solutions, as well as newer server bus architectures like PCI Express 2.0, will allow infrastructure to utilize the bandwidth of 10 Gb adapters. The challenge is in unifying storage traffic and network traffic onto the same link. Fibre Channel over Ethernet (FCoE) will allow an Ethernet-based SAN to be introduced into the FC-based data center without breaking existing administration tools or workflows. This is done by allowing Fibre Channel traffic to run over an Ethernet infrastructure. Servers and storage see FCoE as FC (Figure 2) since FCoE is simply encapsulating FC-over-the-Ethernet portions of the connectivity; to the server applications and FC SAN and FC storage, the traffic is still FC. FCoE can be easily added into existing FC environments in contrast to previous attempts at consolidation with iSCSI and InfiniBand, which required new drivers and a rip-and-replace for customers that are using Fibre Channel. Security and management best practices of having a single application per NIC will need to be revisited as part of 10 Gb Ethernet adoption.

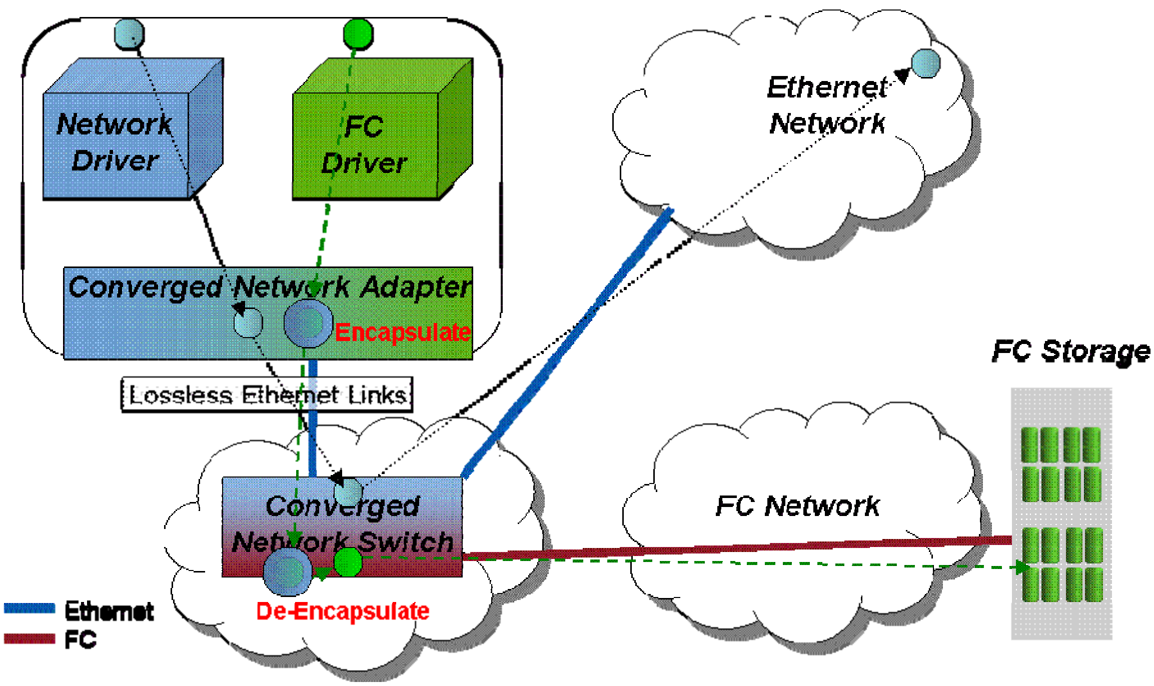

Figure 2. FCoE environment with converged adapters at the server attached to a converged switch

Introduction

Today's networks use different protocols to send information among devices. Traditional Ethernet is a family of frame-based computer networking technologies for local area networks (LANs), whereas Fibre Channel is used for storage area networking (SANs). Fibre Channel over Ethernet, or FCoE, is a new storage networking protocol that supports Fibre Channel natively over Ethernet. FCoE encapsulates Fibre Channel frames into Ethernet frames, allowing them to run alongside traditional Internet Protocol (IP) traffic. This white paper provides an overview of FCoE, describes the hardware and software components that make up the new ecosystem, and explains how the technology is expected to mature over the next few years.

Audience

This white paper is intended for network and storage administrators who want to learn more about FCoE and its benefits.

Creation of an FCoE infrastructure

FCoE requires the deployment of three new components: a Converged Network Adapter (CNA), Lossless Ethernet Links, and a Converged Network Switch (CNS). The CNA provides the functions of both a standard NIC and a FC HBA in a single adapter in the server. There are two options available for the new adapters: a "hardware" solution — where the FC stack and FCoE processes are done in hardware, or a "software" solution — where those stacks are processed in software. When obtaining a hardware solution from a vendor that manufacturers FC HBAs, the drivers will be the same for both FC and FCoE adapters; this provides the least disruption and easiest integration into existing environments. The software solutions require the development of new FC drivers that will be run in software; one of the efforts can be found at Open-FCoE.org. The CNS is the network device that will connect to existing LAN and SAN environments. The T11 standards FC-BB-5 group (http://www.t11.org/fcoe) is creating the FCoE protocol, which enables the creation of CNA and CNS devices. The link that FCoE traffic uses must provide the same level of lossless behavior that can be found with Fibre Channel environments today. The Data Center Bridging group (part of the IEEE 802.1 standards, see http://www.ieee802.org/1/pages/dcbridges.html) is investigating requirements to create an Ethernet environment that can support storage and IPC traffic. The enhancements to Ethernet will be discussed later, but new functionality is required in the devices and 10 Gb Ethernet is going to be the entry level solution. Most users have not deployed 10 Gb Ethernet in the data center yet, so the FCoE effort is poised to help mold the technology and catch the incoming wave of adoption.

FCoE has FC inside

In creating the FCoE protocol, the goal is summed up in the name — take FC and put it over an Ethernet infrastructure. To do this, the Fibre Channel frames are encapsulated — not translated or bridged — in an Ethernet frame (Figure 3). The mapping is 1:1, which means that there is no segmenting of FC frames nor are multiple FC frames put in a single Ethernet frame. It is a requirement that all devices for FCoE (adapters and switches along the path) support Jumbo Frames to allow the largest FC frame to be supported without segmentation. Flow control of FCoE is controlled by the Ethernet PAUSE or Priority Flow Control mechanisms rather than by the buffer credits of FC. The naming conventions for FCoE port types are very similar to FC:

- N_Ports (HBAs and storage) are VN_Ports (CNAs or FCoE storage).
- F_Ports (fabric ports) are VF_Ports (FCoE switch ports that attach to VN_Ports).
- E_Ports (switch to switch) are VE_Ports (ports between two FCoE Ethernet switches).

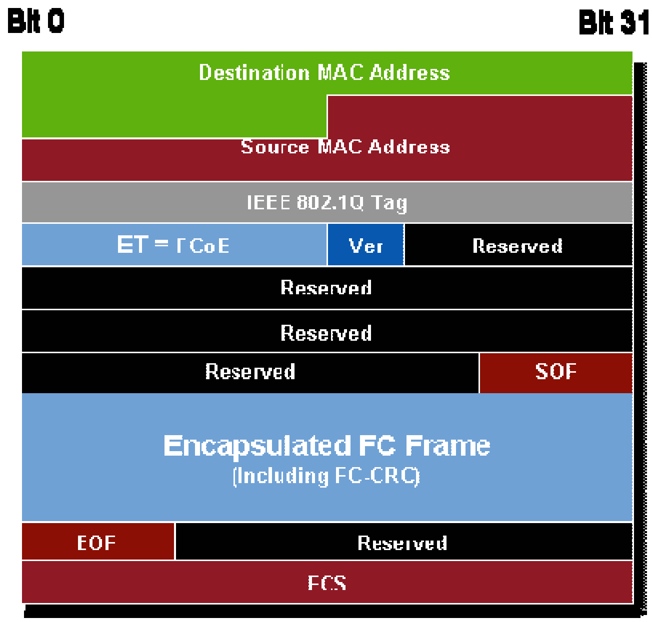

Figure 3. FCoE frame format

One of the more contentious parts of creating the FCoE standard was how addressing would be implemented. Ethernet uses MAC addresses, and there were two competing MAC addressing proposals: server provided (SPMA) and fabric provided (FPMA). After months of debate and negotiations, a compromise was reached that would allow for both methods to be valid by creating an FCoE Initialization Protocol (FIP) that would handle discovery, then login and determine whether SPMA or FPMA would be used. If products support both methods, customers will be able to choose which they want to use. Since the FIP compromise was only made in February 2008, pre-standards products will likely not have this functionality.

In Fibre Channel, security is typically not high on the list of discussion points. In the early days of FC, the small scale, physical security and optical cabling was usually sufficient for customers. The Fibre Channel Security Protocols standard (FC-SP) was approved in 2004 and addresses how to protect against security breaches. Ethernet has its own well-established security practices, so the standards were only concerned with specific threats that are not covered in a typical Ethernet environment. An issue of concern was that while FC links are always private (two ends), FCoE links can be shared (three-plus ends) due to the nature of Layer 2 Ethernet. Access Control Lists (ACLs) will be used to specify on a port basis what operations are allowed to be performed, similar to the protection provided by a firewall. Automation of this feature by the FCoE switches in the future will reduce administrative burden.

FCoE is FC over 'Lossless Ethernet'

The Ethernet infrastructure, over which FCoE will travel, must be of a lossless nature. Since the FCoE protocol does not contain TCP, any lost packets would require recovery at the SCSI layer. In a typical environment, based on existing Ethernet, this would happen much too often to be usable for storage environments. The Ethernet community has been looking at the issue of creating a lossless environment for a number of years. While the standards efforts are well underway, the terminology and list of enhancements for FCoE vary among vendors. Cisco first introduced the term "Data Center Ethernet," which they use to brand the features supporting their unified fabric based on Ethernet. IBM offered the term "Converged Enhanced Ethernet (CEE)" as an industry term for the FCoE initiative. Since the IEEE Data Center Bridging standards will take some time to become ratified, a group of CEE authors have agreed on three features that will create the "v0 CEE": Priority Flow Control, Bandwidth Management, and Configuration.

Even before FCoE was conceived, there was a PAUSE function (IEEE 802.3, Annex 31B) that would stop all traffic on a port, creating a lossless environment, when a full queue condition was approached. The problem with classical PAUSE is that you cannot specify which traffic is important, and there is no mechanism to deal with the congestion created when a PAUSE is issued. Priority Flow Control (IEEE 802.1Qbb) creates a new PAUSE function that can halt traffic according to the priority tag while allowing traffic at other priority levels to continue. Administrators use the eight lanes defined in IEEE 802.1p to create virtual lossless lanes for traffic classes like storage (that require lossless behavior) and lossy lanes for other classes.

In order to make sure that one traffic stream does not take too much of the overall *bandwidth*, the addition of Enhanced Transmission Selection (IEEE 802.1Qaz) is important for creating an environment where storage and other traffic can share the same link. A common management framework can be created for bandwidth management. High-priority traffic like storage can be prioritized and allocated bandwidth while still allowing other traffic classes to utilize the bandwidth when it is not being fully utilized.

The Data Center Bridging Exchange Protocol (DCBX) is responsible for the *configuration* of link parameters for Data Center Bridging functions. It determines which devices support the enhanced functionalities that create the "CEE" cloud where FCoE traffic can safely travel. It also allows for transparent pass-through for non-DCB traffic classes.

The enhancements discussed here are all at the link level, which starts to create a safe environment for FCoE traffic. Before these configurations can grow beyond a direct attach environment, the end-to-end traffic implications must be addressed. Congestion Notification (IEEE 802.1Qau) is a development to manage congestion beyond the link level. This is desirable when a link is reaching a PAUSE condition and the source of the traffic can be alerted and throttled appropriately, reducing congestion out of the network core. Additional enhancements to optimize Ethernet for storage environments can be expected as solutions mature.

Road to a converged fabric

Many customers are already running their storage traffic over Ethernet by utilizing iSCSI; FCoE is not for these customers and was not designed to replace iSCSI. For those customers that have an existing infrastructure and knowledge base of FC, FCoE provides a path toward reaching a converged fabric. Both FCoE and iSCSI will be able to leverage 10 Gigabit Ethernet and Lossless Ethernet enhancements (Figure 4). In general, iSCSI environments tend to be small configurations with five to 20 servers, while many FC customers are scaling into hundreds or thousands of nodes; FCoE can be plugged in to these existing environments.

VMware and FCoE

Currently, VMware server environments can choose between FC, iSCSI, and NAS for networked storage. The hypervisor has a VMkernel storage stack that presents FC (from an HBA) and iSCSI (from a NIC) traffic to individual virtual machines (VMs) as storage. There is also a virtual switch (vSwitch) in the hypervisor that sends traditional LAN traffic to the VM as network traffic. FCoE solutions using a CNA will function equivalently as existing solutions – the CNA will convert FCoE traffic to FC packets in the hardware, so the hypervisor and VMs will still work as if physical FC HBAs and NICs were installed. Software FCoE solutions with a standard NIC will require additional developments before they can be used. Currently the VMware hypervisor will not identify FCoE traffic that is still encapsulated and the vSwitch does not have Lossless Ethernet functionality, so sending traffic directly to the VM would not be reliable. FCoE will expand the available storage networking solutions options for the high-growth server virtualization market.

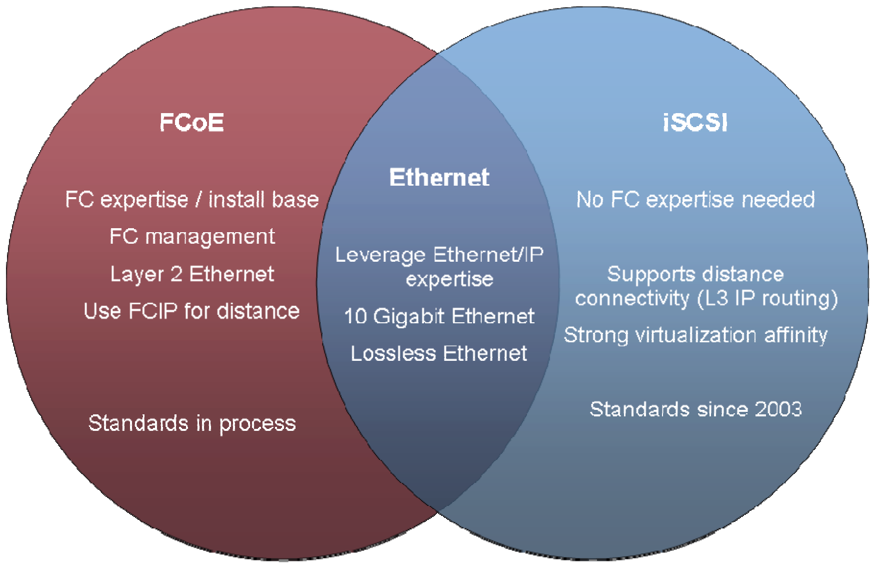

Figure 4. FCoE and iSCSI both provide paths for storage to leverage Ethernet

All new technologies take time to develop a robust ecosystem, and FCoE is no exception. Since FCoE will build on existing infrastructure, it can be rolled out in a phased manner as functionality is created and hardened. The first phase of FCoE will be the Server Phase, in which the separate network (NIC) and storage (HBA) interfaces on the server are replaced by the Converged Network Adapter (CNA) that is directly attached to the Converged Network Switch (CNS). A new rack of servers can deploy these technologies while the existing storage and network environments remain unchanged. The FCoE traffic and Lossless Ethernet environment in these early deployments will be tightly controlled, which is especially important for pre-standards deployments where some pieces discussed in T11 and IEEE standards are not in place. The first phase allows for cost savings by reducing cable, infrastructure, and power.

The second phase of FCoE will be the Network Phase, in which the CNS moves out of the rack and into a unified network. In order to create this unified network, the Ethernet enhancements must be able to create a safe environment for FCoE in the same fabric as the rest of the networking traffic. In addition to the technical concerns, there may be some internal "political" discussions for customers to decide whether the storage or networking group owns the purchase, installation, and maintenance of the CNA and CNS components. All of the enhancements to Ethernet will be new to network administrators who will need to learn these functions and will require coordination with storage administrators. The overlapping domains may compel cultural adjustments, as storage networks will no longer be dedicated and network configurations can no longer be reconfigured at-will. Network and storage administrators will each have their own management interfaces to the environments, keeping tasks separate rather than converged.

A third phase is the Storage Phase in which end-to-end FCoE is enabled with native FCoE storage. The ecosystem must be robust enough to be able to create a large network that reliably provides the same functionality that is available with FC today. Note that even in a totally Ethernet environment, the solution still maintains FC at the upper layer at both the host and storage.

No discussion of FCoE is complete without addressing the cabling infrastructure supporting the solution. Most data centers have not deployed 10 Gigabit Ethernet even though the standard has been complete and products have been available for many years. For 1 Gigabit Ethernet, the primary options for cabling are copper (1000Base-T with RJ-45 connectors) and optical (same physical cabling as FC); copper dominates this market with billions of ports installed while optical has historically been 1 percent of Ethernet ports[3]. A standard (10GBase-T) for using existing copper cabling (either CAT 6 or CAT 6a with RJ-45 connectors)

[3] IEEE Installed Cabling market data (http://ieee802.org/3/10GBT/public/jan03/flatman_1_0103.pdf)

is available, but currently the products are expensive and have high power requirements. A new copper option known as Twinax has become available for FCoE solutions of 10 Gigabit Ethernet. The option is based on the SFF-8431 standard and uses the SFP+ interface for a copper connection that is passive, low cost, and low power. The Twinax solution is limited to short distances; first-generation products are expected to support 5 meters, which is sufficient for direct attach environments in a rack. Standard multimode optical cabling will be used for environments that require longer distances such as from the rack to the core. By the time the market is ready for a unified network, it is expected that the traditional copper solutions will reach a power and price point that will allow customers to take advantage of their existing cabling plant infrastructure.

Conclusion

FCoE is a natural extension of FC networks that will allow data centers to leverage converged I/O and extend the operational efficiencies of networked storage to a new wave of servers. Server, storage, and networking vendors are all investing heavily to build this converged fabric environment. The FCoE ecosystem will develop in a phased rollout as the technologies and markets mature.