White Paper

Introduction to Fibre Channel over Ethernet (FCoE)

A Detailed Review

Abstract

Fibre Channel over Ethernet (FCoE) is a new storage networking option, based on industry standards, which is now being deployed in real world environments. This white paper provides an overview of FCoE, describes the hardware and software components that make up the ecosystem, and explains how the technology is expected to continue to mature over the next few years.

April 2011

Part Number h5916.3

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. As a result, many 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.

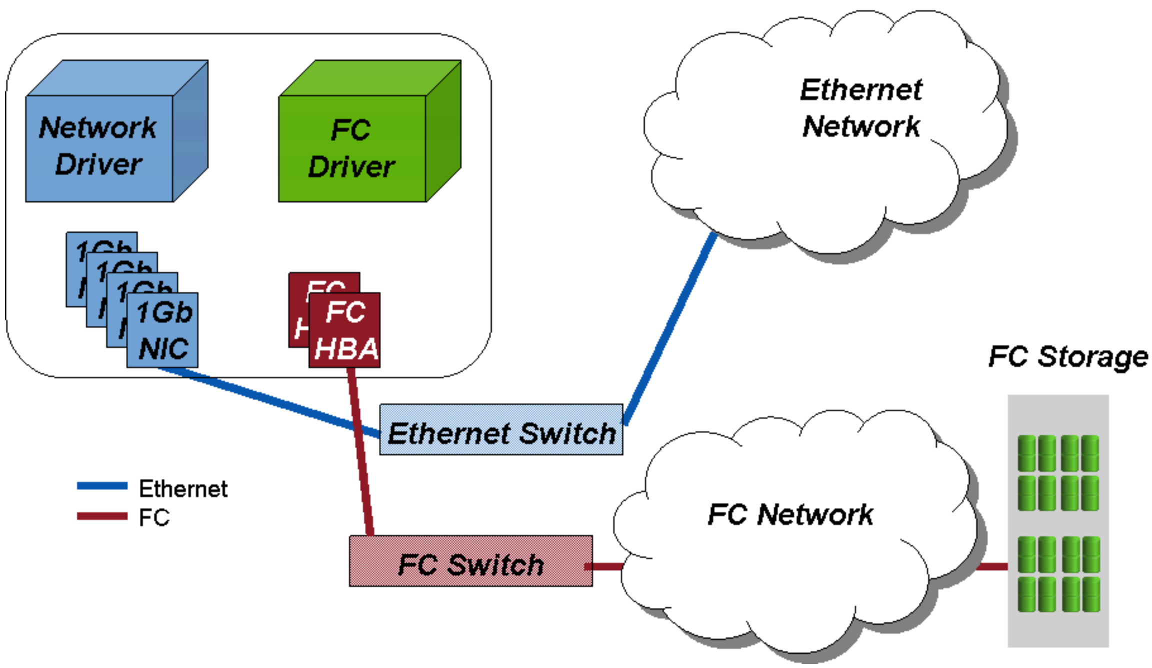

Figure 1. Existing data center with separate Ethernet and FC networks

[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 widely available with Ethernet, many data center solutions require the higher bandwidth solutions that Fibre Channel can supply today with 2 Gb, 4 Gb, 8 Gb, and 16 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, while to the server applications as well as 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. The security and management best practices of having a single application per NIC will need to be revisited as part of 10 Gb Ethernet adoption.

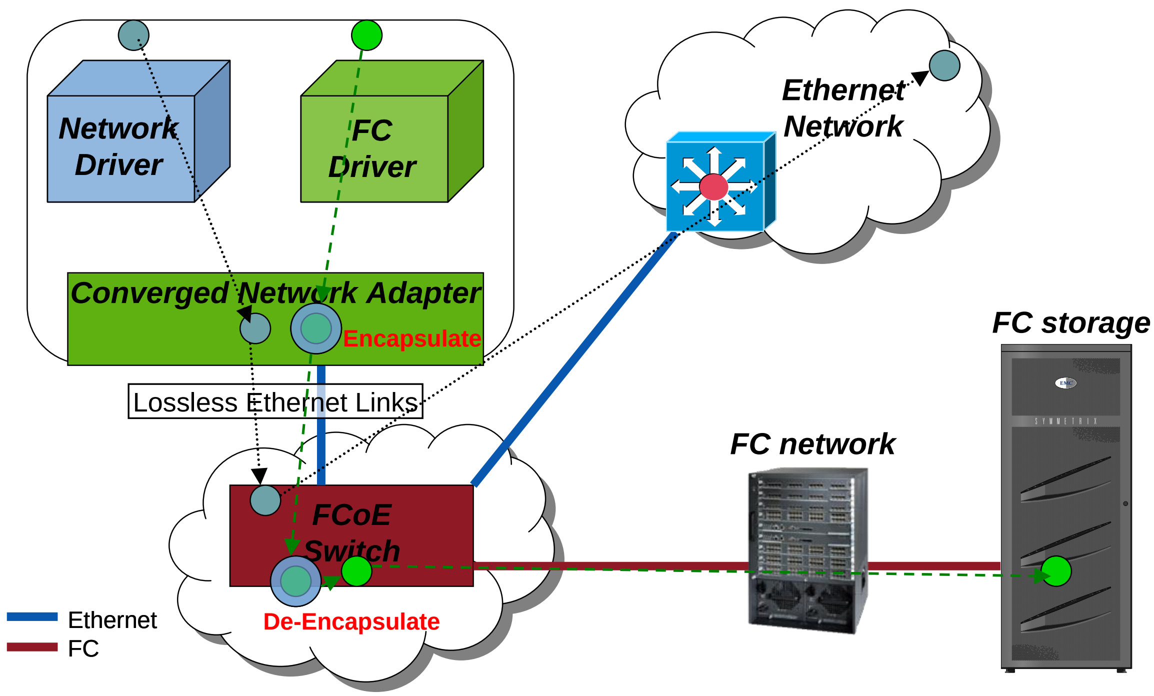

Figure 2. FCoE environment with converged adapters at the server attached to a converged switch

Introduction

Today's networks use different protocols to send information between 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 an FCoE switch. The CNA provides the functions of both a standard NIC and a FC HBA in a single adapter in the server. There are two types of CNAs available: a "hardware"-based solution — where the lower-level FC (exchange and sequence management) and FCoE functions are done in hardware — or a "software" solution — where some of those functions are performed 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 FCoE switch 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) created the FCoE protocol, which enables the creation of CNAs and FCoE switches. 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. Customers have started to introduce 10 Gb Ethernet in their data centers, 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 mini-jumbo frames to allow the largest FC frame to be supported without segmentation. Flow control of FCoE is controlled by 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).

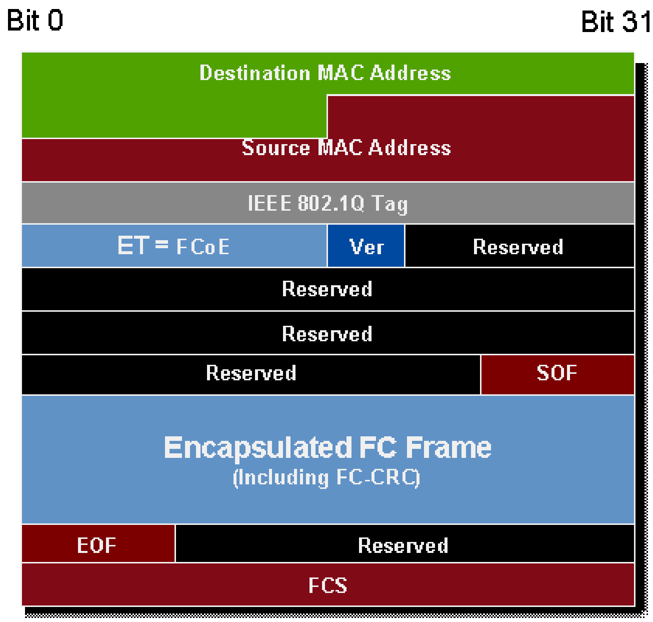

Figure 3. FCoE frame format

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 were 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. A feature called "Dynamic ACLs" is supported by the FCoE switches to reduce the administrative burden of manually configuring ACLs.

FC-BB-5, the T11 standard for FCoE was ratified in June 2009.

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. Fortunately, the Ethernet community has been looking at the issue of creating a lossless environment for a number of years. The set of features required to support lossless behavior is referred to as Data Center Bridging (DCB) and is being worked on by the IEEE DCB working group. The first set of features is link level protocols and by definition only covers the link between two devices (that is, either an adapter and switch port or two switch ports); they are Priority Flow Control (PFC) and Enhanced Transmission Selection (ETS).

Even before FCoE was conceived, there was a PAUSE function (Link Pause, IEEE 802.3, Annex 31B) that would allow for the creation of a lossless environment by stopping all traffic on a port when a full queue condition was achieved. The problem with classical PAUSE is that you cannot specify which traffic should be lossless, and as a result there is no mechanism to deal with the congestion created on all classes of traffic 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 "DCB" cloud where FCoE traffic can safely travel. It also allows for transparent pass-through for non-DCB traffic classes.

These link level enhancements start to create a safe environment for FCoE traffic. PFC, ETS, and DCBX have all been defined and therefore solutions that have these functions shipping today should be able to support FCoE.

As the configurations expand, there are additional Ethernet enhancements beyond link level that will increase performance. 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. While not necessary to support FCoE, an additional enhancement to optimize Ethernet for storage environments is Layer 2 multipathing. Ethernet solutions today use Spanning Tree Protocol (STP), which only supports a single active path between any two network nodes. The IETF standards group has created TRILL (TRansparent Interconnection of Lots of Links), which allows for increased bandwidth by allowing and aggregating multiple network paths.

Virtualization and FCoE

Currently, server virtualization environments (including VMware and Microsoft Hyper-V) can choose among FC, iSCSI, and NAS for networked storage. The hypervisor has a storage driver stack that presents FC (from an HBA) and iSCSI (from a NIC) traffic to individual virtual machines (VMs) or guests as storage. There is also a virtual switch in the hypervisor that sends traditional LAN traffic to the VM/guest 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 VM/guest 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 hypervisors will not identify FCoE traffic that is still encapsulated and virtual switches (including the Cisco Nexus 1000V in VMware® ESX®) do not have Lossless Ethernet functionality, so sending traffic directly to the VM/guest would not be reliable. FCoE will expand the available storage networking solutions options for the high-growth server virtualization market. Moving to a converged fabric will simplify mobility by reducing the number of networks that need to be moved. It also allows a standard configuration to be built, allowing for flexible storage deployment.

Road to a converged fabric

Many customers are already running their storage traffic over Ethernet by utilizing iSCSI or NAS; FCoE is not designed to replace these solutions. For customers that have an existing infrastructure and knowledge base of FC, FCoE provides a path toward reaching a converged fabric. FCoE, iSCSI, and NAS will all leverage 10 Gigabit Ethernet. In general, iSCSI environments tend to be small configurations with under 100 servers, while most FC customers are scaling into hundreds or thousands of nodes; FCoE can be plugged in to these existing environments. 10 Gigabit Ethernet will provide the common infrastructure that can support FCoE, iSCSI, and NAS (Figure 4), all of which can be used in virtualization environments, and customers will choose based upon their requirements and skill set.

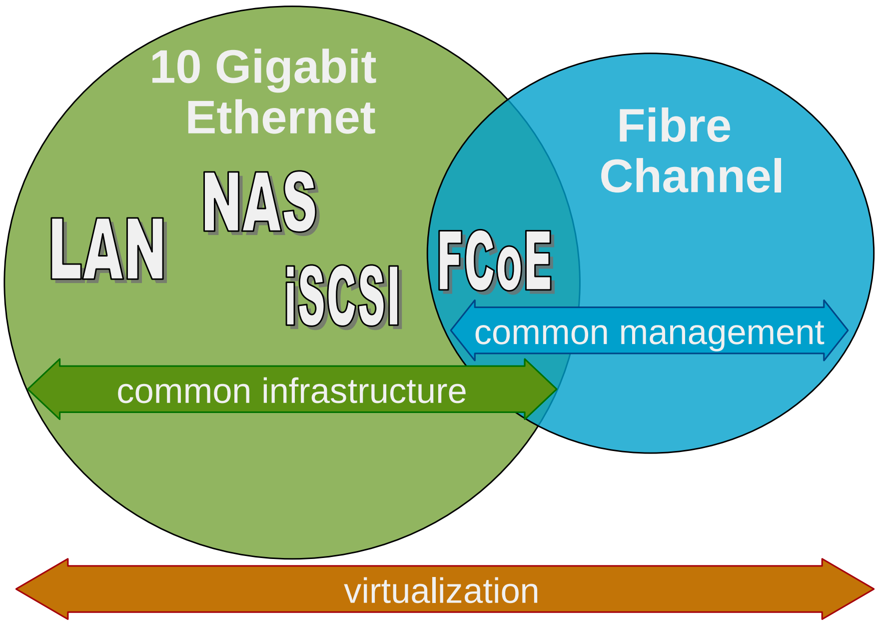

Figure 4. 10 Gigabit Ethernet common infrastructure for storage

All new technologies take time to develop a robust ecosystem, and FCoE is no exception. Since FCoE builds on the existing infrastructure, it can be rolled out in a phased manner. Today FCoE can support convergence at the server, in which the separate network (NIC) and storage (HBA) interfaces on the server are replaced by the CNA that is directly attached to the FCoE switch. A new rack of servers can deploy these technologies, while the existing storage and network environments remain unchanged. FCoE can also support convergence at the storage port and on switch-to-switch links, allowing for an end-to-end (FCoE multi-hop) configuration to be created.

Recently, FCoE solutions started to expand beyond a single switch at the access layer into multi-hop configurations. While the industry continues to work on the technical issues of expanding FCoE solutions, customers should review the structure of their storage and networking groups to make sure that there are not any "political" issues over who owns the purchase, installation, and maintenance of the CNA and FCoE switch components in a converged network environment. Some 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.

As director-class products that support FCoE are released into the market, building FCoE multi-hop configurations that include native FCoE storage begins to make sense.

No discussion of FCoE is complete without addressing the cabling infrastructure supporting the solution. When customers deploy a physical cabling infrastructure, they are looking for solutions that can be used for five to ten 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) is available but not expected to start shipping for FCoE solutions until 2012. 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 low cost and low power. The Twinax solution is limited to short distances, supporting between 1-10 meters, which is sufficient for server to top-

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

of-rack or end-of-row switch environments. Standard multimode optical cabling will be used for environments that require longer distances such as from the rack to the core.

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 have all invested heavily to create converged fabric technologies. There is a robust FCoE ecosystem that will continue to expand and mature over time.