Fibre Channel over Ethernet in the Data Center: An Introduction

Authors: Garry Lemasa and Silvano Gai, Cisco Systems

Introduction

Fibre Channel over Ethernet (FCoE) is a newly proposed standard that is being developed by INCITS T11. The FCoE protocol specification maps Fibre Channel natively over Ethernet and is independent of the Ethernet forwarding scheme. It allows an evolutionary approach towards I/O consolidation by preserving all Fibre Channel constructs, maintaining the same latency, security, and traffic management attributes of FC while preserving investments in FC tools, training, and SANs. FCoE recognizes FC as the dominant storage protocol in the data center while giving customers a viable I/O consolidation solution. FCoE simplifies customer environments by using Ethernet and allowing the industry to avoid creating another, separate protocol for I/O consolidation.

In today's datacenters, companies use both Ethernet for TCP/IP networks and Fibre Channel (FC) for storage area networks (SANs), each dedicated to specific purposes. Ethernet networks are generally implemented when end-users need to transfer relatively small amounts of information over both local and global distances or in clustered, low-latency compute environments. Storage area networks are implemented by companies who require access to block I/O for applications such as booting over SANs, mail servers, file servers and large databases. The benefits of deploying SANs remain: A) centralized management, security, and administration of the storage resources, B) uniform delivery of storage services like periodic backups, and C) running efficient utilization levels of storage resources.

Server virtualization is on the rise in data centers as well, spurring an increased demand for Fibre Channel connectivity to virtual hosts residing on these servers. Demand for Fibre Channel in this case is driven by the need for hypervisors to provide guest operating systems with virtualized storage, accessible through a Fibre Channel network infrastructure. (A hypervisor is a virtualization platform that allows multiple operating systems to run on a host computer at the same time.) Companies who have virtual-machine based mission critical applications in their data-centers typically configure their servers with a pair of Fibre Channel host bus adapters (HBA's) and two or more network interface cards (NICs). Some high end deployments have as many as eight Ethernet NICs in high performance servers that only have 2 CPU cores! The adapter requirements could only grow when 32 CPU cores become commonplace. Fibre Channel over Ethernet (FCoE) enables the consolidation of both SANs and Ethernet traffic onto a one common network adapter, reducing the ever growing number of adapters required.

FCoE combined with the advent of 10 Gigabit Ethernet (10 GE) fabrics will grant companies the ability to consolidate their I/O, cables, and adapters while at the same time increase the utilization of their servers.

The current proposal, as defined by the INCITS T11 standards body, leverages a lossless Ethernet fabric, maintains the Fibre Channel operational model and includes a newly approved frame format. FCoE is not tied to 10GE, and will be able to run over networks with varying interface speeds.

Why is FCoE important to the data-center? I/O Consolidation.

I/O consolidation is simple in concept: the sharing of both Fibre Channel and Ethernet traffic on the same physical cable or in cases that network isolation is desired, the flexibility to use and configure the same hardware for either type of network load. The benefits end-users will realize from this simple idea are significant. Companies that leverage I/O consolidation will be able to realize significant gains in server slot efficiencies with the use of multi-function network/storage adapters simplify their cabling scheme within a rack and reduce the amount of heat each server produces. Even better, companies will be able to accomplish these things while still utilizing their existing Fibre Channel SANs. These benefits are further detailed below.

Today, IT managers often use four, six or even eight network adapters in their mission critical servers. These adapters can be two Fibre Channel host bus adapters, plus a pair of Network Interface Cards for traffic and in the case of a Virtual Machine environment up to an additional four NICs, depending on the VM vendor's best

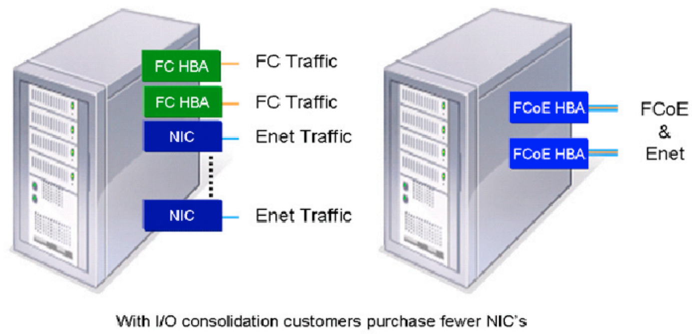

Figure 1 – FCoE I/O Consolidation Benefit – Fewer NICs per Server

practices. I/O consolidation means a customer can use multi-function network/storage adapters in place of single-function network-specific and storage-specific cards, thereby reducing the number of server slots and switch ports, as well as reducing the number of power consumed for I/O and necessary cooling. This also results in fewer points of management administrators will have to control.

A reduction in NICs through I/O consolidation has an additional important advantage. The ability to cool a set amount of heat generated per rack is the primary barrier to data-center expansion and inefficiency encountered today. If a given rack has a maximum of 12 kilowatts allotted for cooling, and that allotment is exceeded, then additional servers and switches cannot be added, regardless of the rack space available. Reducing the amount of NICs in servers can reduce the amount of heat those servers generate. In a standard PCI Express bus, each adapter has a power budget of 25 watts. If an IT manager reduced the number of adapters in the rack server from six to two, they would be potentially saving up to 100 watts per rack server (25 watts/ adapter X 4 adapters = 100 Watts).

I/O consolidation also gives companies the means to simplify their cable management. Implementing FCoE on a pair of 10 Gigabit Ethernet cables can replace the equivalent of two 4 Gb FC connections and twelve one Gigabit Ethernet connections. While still providing 20 Gb of bandwidth, this is an 86% reduction in the amount of cables to be managed.

For administrators that desire SAN/LAN isolation, FCoE allows the user to employ only FCoE multi-function network/storage adapters and without the need to employ separate hardware and cables based upon traffic types.

What is Fibre Channel over Ethernet?

Conceptually FCoE can be broken down into three components:

- Encapsulation of a Native Fibre Channel Frame into an Ethernet frame
- The extension of Ethernet to become a lossless fabric
- The replacing of a Fibre Channel link with MAC addresses in a lossless Ethernet fabric

Why is preservation of Native Fibre Channel Format so Important?

- Administration, Management and Service Delivery

Users contemplating deployment of an FCoE based network are assured of a smooth transition to FCoE because existing Fibre Channel SAN based administrative and management procedures will continue to work in an FCoE based network. Compatibility with existing Fibre Channel networks is required which will allow usage of existing FC SAN tools and management constructs.

From a Storage Administrator perspective, zoning is a basic provisioning function that is used to give

Hosts access to Storage. Zoning has a dual purpose function A) storage allocation and B) security. By virtue of zoning a Host is able to see the devices provisioned for it on the Storage subsystem. Zoning also ensures that unauthorized servers don't get access to Storage devices. FCoE based switches will continue to offer this zoning function ensuring that storage allocation and security mechanisms are unaffected.

In addition to zoning, another integral element in SANs is the Name Server Database. This database covers a list of devices currently connected to the network. It also includes information about network location and status of the connections. FCoE based SANs will continue to offer the Name Server Database function.

Beyond the basic SAN provisioning functions mentioned above, Storage Administrators also deliver services such as Disaster Recovery, Backup and Recovery. Disaster Recovery solutions typically used a protocol called FCIP. FCIP remains unchanged with FCoE. Data that needs to be sent to a remote site will hop on to a FCIP link via a gateway and similarly enter a FCoE network via a FCIP gateway. File Backups to Tape Libraries and Recovery from these devices will continue to use existing backup and recovery tools.

What about companies who have an existing Fibre Channel infrastructure? Does FCoE mean that they need to replace their existing SANs? Absolutely not. FCoE is designed to maintain the Fibre Channel model. A storage administrator logged into a server connected through an FCoE network would see Fibre Channel initiators just as if that same server were connected on a SAN. The same management tools that customers use to manage and maintain their SANs today can be used in an FCoE environment.

Finally, storage users can also be assured that new services utilizing storage virtualization applications can continue to work with FCoE because everything above the transport layer remains intact. By encapsulating native Fibre Channel in FCoE, the transparent interoperability of existing SAN networks will be possible.

Encapsulation

The encapsulation of the Fibre Channel frame occurs through the mapping of FC onto Ethernet. Fibre Channel and traditional networks have stacks of layers where each layer in the stack represents a set of functionality. The Fibre Channel stack consists of five layers, FC-0 through FC-4. Ethernet is typically considered a set of protocols in the seven-layer OSI stack that define the physical and data link layers. FCoE provides the capability to carry the FC-2 layer over the Ethernet layer. See figure 2 below. This allows Ethernet to transmit the upper Fibre Channel layers FC-3 and FC-4 over the IEEE 802.3 Ethernet layers. .

It is this FCoE mapping which allows FC traffic to pass over an Ethernet infrastructure.

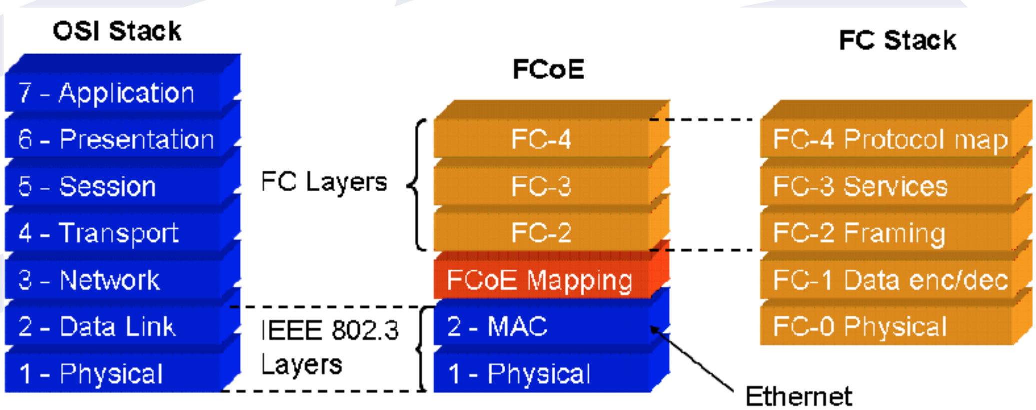

Figure 2 – FCoE mapping

Frame Format

How does FC traffic get passed over Ethernet? FCoE encapsulates a Fibre Channel frame within an Ethernet frame. Figure 3 represents the frame format as agreed to by the INCITS T11.3 standards body.

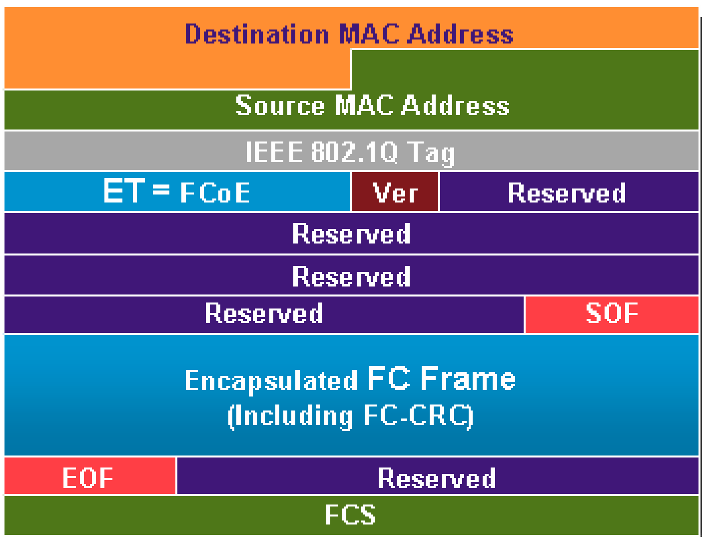

Figure 3- The FCoE frame format

The first 48-bits in the frame are used to specify the Destination MAC address and the next 48-bits specify the Source MAC Addresses. The 32-bit IEEE 802.1Q Tag provides the same function as it does for virtual LANs, allowing multiple virtual networks across a single physical infrastructure. FCoE has its own Ether type as designated by the next 16 bits, followed by the 4-bit version field. The next 100-bits are reserved and are followed by the 8-bit Start of Frame and then the actual FC frame. The 8-bit End-of Frame delimiter is followed by 24 reserved bits. The frame ends with the final 32-bits dedicated to the FCS function that provides error detection for the Ethernet frame.

The encapsulated Fibre Channel frame consists of the original 24 byte FC header and the data being transported (including the Fibre Channel CRC). The CRC is a cyclical redundancy check used for error detection. The FC header is maintained so that when a traditional FC Storage Area Network is connected to an FCoE capable switch the frame is de-encapsulated and handed off seamlessly. This capability enables FCoE to integrate with existing FC SANs without the need of a gateway.

Frame size is also a factor in FCoE. A typical Fibre Channel data frame has a 2112 byte payload, a header and FCS. A classical Ethernet frame is typically 1.5 KB or less. To maintain good performance, FCoE must utilize jumbo frames (or the 2.5 KB “baby jumbo”) to prevent a Fibre Channel frame from being split into two Ethernet frames.

Lossless Ethernet

One of the challenges with passing Fibre Channel frames over Ethernet is that FC provides a lossless transport. Congestion in a lossless network is an issue that must be managed. Traditional Fibre Channel manages congestion through the use of a link-level, credit based flow control that guarantees no loss of a frame under normal conditions. Typical Ethernet, coupled with TCP/IP, uses a packet drop flow control mechanism. Packet drop flow control is not lossless making it unacceptable for use with storage traffic. Fortunately, classical Ethernet has a PAUSE capability so that a busy receive port can send a control frame to the transmit port requesting a pause in transmission. Use of this optional IEEE 802.3x Ethernet feature enables Fibre Channel traffic to pass on an Ethernet network in a lossless fabric. See Figure 4 below.

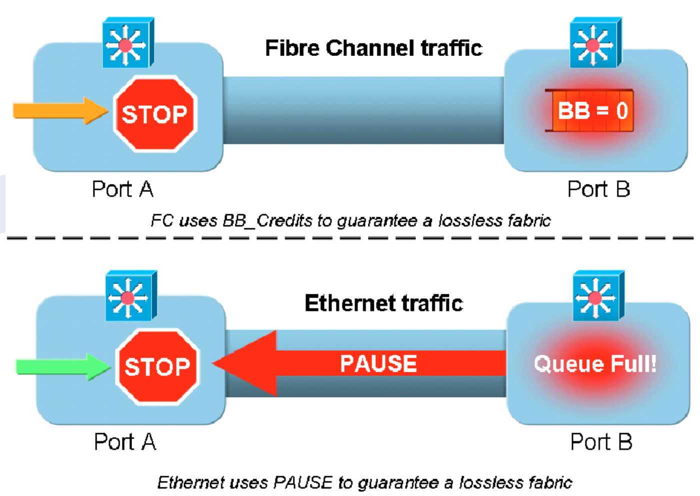

Figure 4 – Lossless Fibre Channel and lossless Ethernet

There is also a new Ethernet enhancement being developed that allows PAUSE functionality to be enabled for each user-priority supported in Ethernet. While PAUSE provides the basic functionality to make Ethernet lossless, a new proposal for Priority Flow Control will provide significant benefit for both FC and Ethernet traffic. Priority Flow Control is discussed on page 7.

Example of an FCoE switch

Figure 5 shows a functional Diagram of a FCoE capable switch as depicted by the larger, yellow field. Inside this device is both an Ethernet switch function as well as a Fibre Channel switch function. Note the Fibre Channel over Ethernet entity within the FC Switch. This entity encapsulates FC frames into FCoE Frames and de-encapsulates FCoE frames back to FC frames. FC traffic flows into one end of the FCoE entity and FCoE traffic flows out the other end.

The FCoE Entity has an Ethernet MAC address that will be used as a source or destination address when FCoE frames are transmitted through an Ethernet fabric (Further details provided below).

In figure 5, an FC frame enters from the green FC ports to the right and into the FC switch where the entity encapsulates the Fibre Channel frame into a FCoE frame. The FCoE frame is sent to an Ethernet Bridge where it is distributed to the appropriate Ethernet port on the bottom of the diagram.

The reverse is also true; FCoE traffic will flow from the red ports, through the Ethernet Bridge until it is de-encapsulated inside of the FCoE Entity, where it passes through the FC switch to the appropriate FC port.

This is just one example of how a switch can be FCoE enabled. Different vendors may choose to implement FCoE in different ways.

FC and FCoE address mapping

A third component of FCoE mentioned above is replacing of a Fibre Channel link MAC addresses in a lossless Ethernet fabric. Traditional Fibre Channel fabric switches maintain forwarding tables FC_IDs. FC switches use these forwarding tables to select the best link available for a frame so the frame reaches its destination port. Fibre Channel links are typically point-to-point and do not need an address at the link layer. An Ethernet network is different because it does not form an end-point to end-point connection in the manner of FC. This requires FCoE to rely on Ethernet MAC addresses to direct a frame to its correct Ethernet destination (the first two fields of the FCoE frame are MAC addresses, as outlined previously in figure 3).

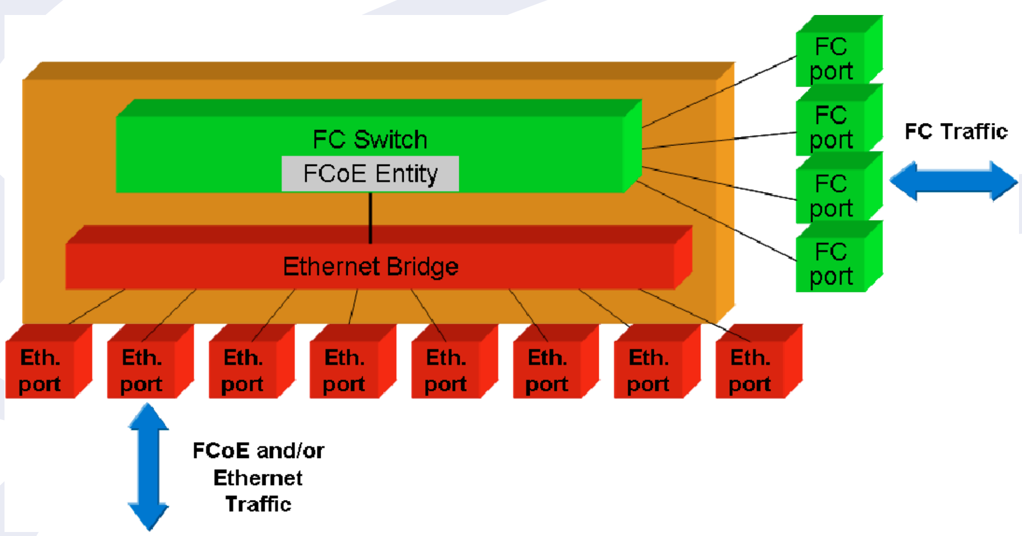

Figure 5 – Functional diagram of an FCoE switch.

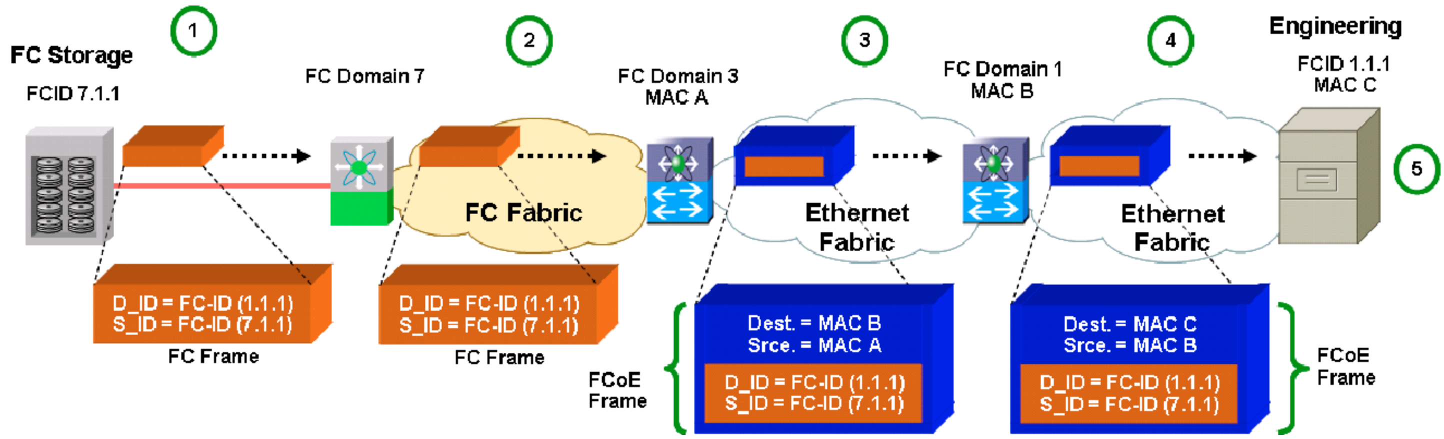

Figure 6 – FC and FCoE address mapping

Figure 6 illustrates how traditional FC addresses align with MAC addresses in an FCoE network. On the left is a storage array attached to a Fibre Channel switch labeled FC Domain 7. This storage array is in a traditional SAN and stores information for an Engineering host on an FCoE enabled Fabric. The Engineering host has both a FC_ID 1.1.1 and a FC MAC address.

Figure r Step by Step Analyses

1. The Fibre Channel N_port on the storage array sends out the FC frame, which includes the Destination FC_ID (D_ID = 1.1.1) and the Source FC_ID (S_ID = 7.1.1) in the header. For simplicity, only the header information is displayed in the diagram's frame graphic

2. The Fibre Channel switch with Domain ID 7 receives the frame. Since the destination ID (D_ID) in not in this FC domain (7), the switch looks up the destination domain ID in its forwarding table and transmits the frame on the port associated with the shortest path, as determined by the Fabric Shortest Path First algorithm (FSPF).

3. The switch with the FC Domain ID 3 receives the frame and determines that the destination ID (D_ID) is not in Domain 3 and repeats the lookup process in step 2. However, in this case the FC frame will be transmitted across an FCoE enabled Ethernet fabric. This requires the frame to be encapsulated by an FCoE entity in the switch and then transmitted on the port associated with the shortest path.

While the original FC source and destination ID's are maintained in the encapsulated FC frame, the FCoE entity will populate a new destination and source MAC address located in the Ethernet header. As mentioned above, the entity has its own MAC address. In this case, the destination is MAC address B (the MAC address of the FCoE entity in the receiving switch) and source MAC address A (the MAC address of the FCoE entity in the transmitting switch).

4. When the FCoE frame arrives at the FCoE entity with MAC address B, the frame is de-encapsulated and the switch determines that the FC frame destination is within its domain (Domain 1). The FC frame is re-encapsulated with the new destination MAC address C (which corresponds to the FC D_ID 1.1.1) and the new source MAC address B. Then the frame is transmitted out the appropriate port to the FCoE host with MAC address C.

5. When the frame is received by the FCoE host bus adapter with MAC address C, the FCoE frame is de-encapsulated and the FC frame accepted by the Engineering host with FC_ID 1.1.1.

This example demonstrates how FCoE addressing maps to traditional FC addressing. Topologies will vary depending on a customer's implementation requirements and the FCoE capable products deployed. While the topologies will change, the addressing essentials will remain constant.

Example of an FCoE topology

FCoE can be deployed in a variety of topologies, depending on the business needs of a company and the products chosen. Figure 7 depicts a simple FCoE topology where I/O consolidation is achieved at the access layer of the network. FCoE is used in this scenario to consolidate traffic from a server to FCoE enabled switches. The FCoE switches pass Fibre Channel traffic to attached SANs and Ethernet traffic to an attached Ethernet network. This deployment model targets the installed base of LAN and SAN environments in the datacenter providing a phased approach to I/O consolidation adoption.

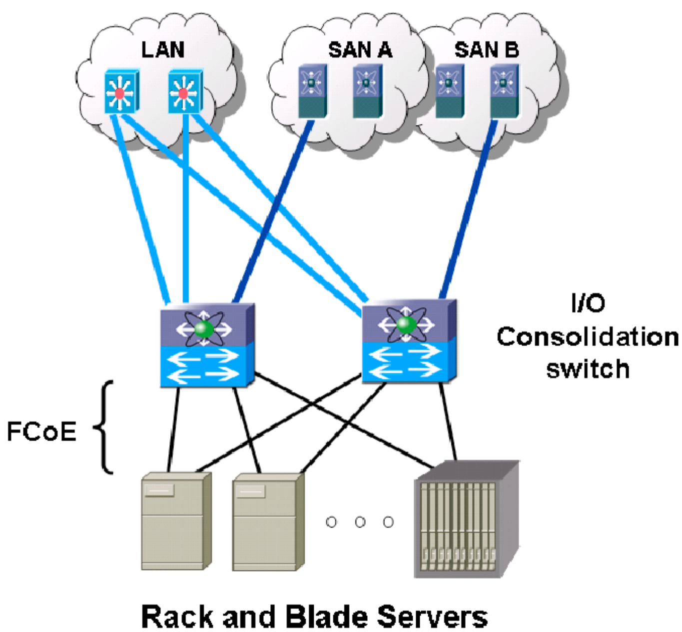

Figure 7 – A simple FCoE Topology

There are many FCoE deployment models that will be utilized by customers. Another example would be to extend the reach of FCoE into the aggregation and core layers of a network. Yet another example will be an end-to-end deployment of an entirely FCoE network. All of these topologies are provided for by the proposed T11 FC-BB-5 standard which mandates connectivity to existing infrastructures.

Ethernet technologies that enhance FCoE

Previous examples highlighted the ability for FCoE to transport FC traffic over an Ethernet infrastructure. In these examples FCoE could be viewed simply as a method for unifying I/O over Ethernet, an alternative to a pure Fibre Channel storage area network. However, there are ongoing technological developments that will make FCoE invaluable in tomorrow's data center.

Examining bandwidth requirements is a good place to start. Many customers use 1/2/4 Gbps FC connections in their datacenters today. Passing a 2 or 4 Gbps FC link over a one Gigabit Ethernet connection doesn't make a lot of sense. However, on a 10 Gigabit Ethernet link customers have the bandwidth capacity to easily pass 4 Gbps of FC. Moreover, there is plenty of bandwidth remaining in the 10 GE link to allow Ethernet traffic to pass simultaneously. Networks also could pass 8 Gbps FC over a FCoE link.

Another extension to Ethernet being developed that will support FCoE is the IEEE 802.1p Class Based Flow Control (aka CBFC, formerly called Per Priority Pause/ Priority Flow Control) functionality. The 3-bit field in IEEE 802.1p allows the creation of logical lanes on a single, physical fabric. A logical lane can be lossless when PAUSE can be implemented on a per lane basis. The per flow PAUSE function remains optional so customers could implement one or two lossless "lanes" for FC while allowing others to pass regular Ethernet traffic.

Flow Control is another Ethernet enhancement that also can improve FCoE's value in customer data centers. Fibre Channel uses a link level flow control with port based buffer-to-buffer credits to control congestion in the link. A lossless Ethernet fabric using the per flow PAUSE described above can prevent the frame drop today, but an additional option, Congestion Notification, can be utilized to enhance the network for FCoE traffic.

IEEE 802.1au Congestion Notification (CN) is a proposal which provides options for acting upon congestion in the network. CN will provide additional flow control capabilities for a network implementing the CBFC function. One option is a CN where congestion points in the network notify a rate limiter when congestion occurs, so that the traffic can be throttled back. Another, more complex CN option is a 3-point flow control scheme using in-band and out of band signaling to provide feedback to rate limiters. While a detailed review of CN is outside the scope of this document, readers interested in the proposal should refer to: http://www.ieee802.org/1/pages/802.1au.html.

As FCoE continues to evolve further into the network, technologies such as Congestion Notification will play an increasingly important role.

Conclusion

Fibre Channel over Ethernet will allow companies' data centers to optimize increases in server virtualization for the next several years. FCoE enables I/O consolidation and offers these enhancements to the data-center:

- Fewer Network Interface Cards per server
- Reduction in server power budget and reduced cooling requirements
- Significant reduction in the amount of cables to manage
- Seamless connectivity with existing Storage Area Networks and Ethernet Networks – customers can continue to leverage their existing FC tools, security model and infrastructure

FCoE positions Fibre Channel as the storage networking protocol of choice and extends the reach of Fibre Channel throughout the data center to all servers.

Readers interested in learning more about FCoE can refer to the following sources:

www.t11.org/FCoE

www.fibrechannel.org