The Role of FCoE in I/O Consolidation

Joy Jiang
Finisar
1389 Moffett Park Dr
Sunnyvale, CA 94089
+1(408)542-4259
joy.jiang@finisar.com

Claudio DeSanti
Cisco
170 W. Tasman Dr
San Jose, CA 95134
+1(408)853-9172
cds@cisco.com

ABSTRACT
Consolidation of the multiple traffic types present in current data centers over a single unified network is a big trend in the information technology industry. Extensions to 10G Ethernet are making it the most promising technology for achieving I/O consolidation. The emerging Fibre Channel over Ethernet (FCoE) protocol plays a critical role in this convergence process. This paper elaborates how FCoE leverages the ubiquity and economics of Ethernet networks while preserving the infrastructure, strengths, and tools of the existing Fibre Channel Storage Area Network (SAN) framework.

Categories and Subject Descriptors
C.2.5 [**Computer Systems Organization**]: Local and Wide-Area Networks – *Ethernet, high speed Fibre Channel.*

General Terms
Management, Measurement, Performance, Economics, Standardization.

Keywords
Fibre Channel over Ethernet, Fibre Channel, 10G Ethernet, Lossless Ethernet, Unified network, I/O consolidation, I/O convergence.

1. INTRODUCTION
I/O consolidation is the capability to consolidate the multiple traffic types that today are carried over different data center network infrastructures into a single network technology that maintains the advantages of each existing traffic type while minimizing the interactions among them. I/O consolidation provides several cost benefits in terms of reduced number of components required to build a data center and of associated simplified management structures.

Fibre Channel (FC) is the preeminent technology used today for storage networking. Among the reasons for its success we can list:

- native support to map SCSI operations over Fibre Channel protocol structures;
- native protocol support for large data transfers; and
- simplified protocol implementations due to the lossless behavior of Fibre Channel networks (see 3.2).

However, in the Local Area Network (LAN) environment, Ethernet is the dominant protocol. To consolidate LAN and SAN into one unified network, it seems intuitive to enhance Ethernet to make it able to natively carry Fibre Channel, so that the low cost Ethernet may be the technology for the converged network [1].

In order to consolidate multiple traffic types over a single network, that network has to provide enough bandwidth. 10G Ethernet (10GE) is the best placed technology for consolidated networks. However, its viability as a converged infrastructure is dependent on the ongoing cost decay in the electronic and optical components used at its physical layer. The information technology industry has continuously brought down the cost of these components and adoption of new protocols often depends on this cost curve.

In this paper we discuss I/O consolidation, its benefits, its requirements, and some technologies for achieving it, with a special attention to the emerging Fibre Channel over Ethernet (FCoE) protocol. In particular we describe FCoE, how it compares with other protocols, and what benefits it brings to the data centers in order to achieve I/O consolidation with minimum impact on existing network structures. We also evaluate the status of the FCoE protocol in terms of standardization, market interest, products and tools availability.

2. I/O CONSOLIDATION

2.1 Overview
Conventional data centers operate up to three sets of parallel networks, one for IP and local area networking (LAN) applications, based on Ethernet, one for storage area networks (SAN), based on Fibre Channel, and, in some cases, when high performance computing (HPC) is deployed, one for the HPC inter-process communications (IPC), based on Infiniband. These three networks run separately from each other because of the different application requirements they satisfy on data transfers and transmission latency. In a sense they are optimized for the specific applications they support. However, running three separate infrastructures comes with significant costs for an enterprise. Up to three different cabling systems, switching infrastructures, and management frameworks need to be maintained. As an example, it is quite common to find servers

International Conference On Advanced Infocomm Technology'08,
Jul 29–31, 2008, Shenzhen, China.

configured with four 1Gb/s Ethernet ports and two 4Gb/s Fibre Channel ports. Consider a small data center composed of 100 of these servers. There is the need of 600 cables, 400 Ethernet switch ports, and 200 Fibre Channel switch ports.

I/O consolidation, also called I/O convergence, is the ability to use a single network infrastructure to carry the three types of traffic mentioned above. With I/O consolidation servers have a common I/O subsystem that supports all three types of traffic rather than three independent I/O subsystems, as shown in Figure 1. An extremely important advantage of an I/O consolidated solution is the related reduction in cables, as well as the unification of the cabling. Considering the 100 servers data center example discussed above, if the six interfaces of each server are consolidated into two converged ports, this means 200 cables rather than 600 and 200 converged switch ports rather than 600. Moreover, all the cables are of the same type, therefore there is no need for different cabling infrastructures.

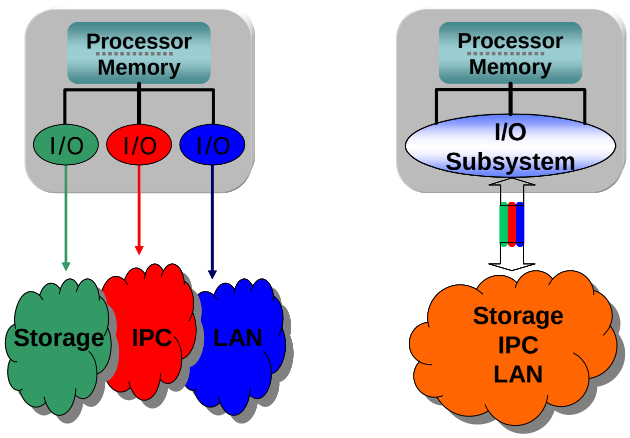

Figure 1. Conventional and Converged networks.

2.2 Benefits

As represented in Figure 2, the benefits of deploying a converged network infrastructure rather than separate networks are:

- Reduced number of hardware components. One common CNA (Converged Network Adapter) is able to replace the NIC for LAN applications, the HBA for SAN applications and the HCA for IPC applications. Fewer PCI slots are used per server. One unified network replaces the three sets of different switches.

- Reduced power requirements. Having fewer electronic components in the converged I/O subsystem means lower power consumption and less need for cooling and space.

- Simplified cabling infrastructure. Converged adapters carry all traffic on single physical links, leading to 50% or more reduction on the number of cables required in a deployment. Moreover, the cables can be of a single type, rather than having different cables for different networks.

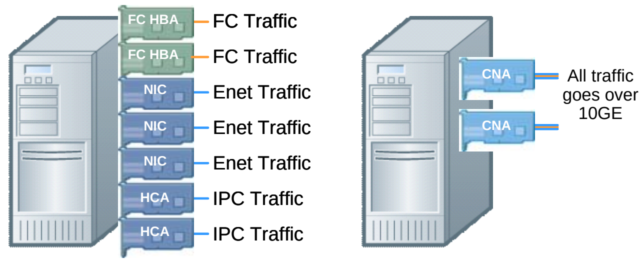

Figure 2. Reduction in Number of Adapters.

2.3 Requirements

There is a lot of investment in Ethernet in data centers and many applications assume Ethernet to be the underlying network technology. This creates the requirement for a converged network aiming at consolidating IP and LAN traffic to be based on Ethernet. A different choice would have a difficult time in being accepted.

To consolidate storage traffic, the converged network has to follow the operational model of the dominant storage networking technology, Fibre Channel.

To consolidate Inter-Process Communication (IPC) traffic, the converged network has to provide low latency and support some specialized APIs.

I/O consolidation imposes an additional requirement on the bandwidth of the converged infrastructure: in order to be able to carry all the traffics the consolidated network needs to have more bandwidth. Considering the consolidation of two 1GE ports and one 4GFC port into a converged port, at least 5.4Gb/s of raw bandwidth are needed. 10G Ethernet, for its bandwidth, is one of the best candidate technologies for I/O convergence. However, the physical layer of 10GE had until recently a very high cost, and this strongly limited the scope of its deployment. Since its inception, 10GE has been used only for switch-to-switch connectivity and for some server-to-server clustering connectivity in HPC data centers. The limited deployments are mainly due to the per-port cost of the 10G physical modules. Several types of 10G modules have been available on the market (e.g., XENPACK, X2, XFP), all much more expensive than the SFP modules used by Fibre Channel. The recent introduction of the 10G SFP+ modules brought the cost of 10G components to basically the same level of the SFP modules. Table 1 shows a comparison of the currently available 10GE physical layers.

Table 1. Comparison of 10GE Physical Layers [2].

Technology	Cable	Power	Distance	Latency
SFP+ Copper	Twinax	0.1W	10m	0.1μs
SFP+ Optical Ultra Short Reach	OM2 OM3	1W	10m 100m	0
SFP+ Optical Short Reach	62.5μm 50μm	1W	82m 300m	0
Laserwire™ Optical Active Cable	MM	0.5W	30m	0
10GBASE-T	Cat6 Cat6a/7 Cat6a/7	8W 8W 4W	55m 100m 30m	2.5μs 2.5μs 1.5μs

The cabling system most used for Ethernet is the ubiquitous Cat6 and Cat5e copper cabling. However, using this cabling for 10G is quite difficult, because of the very high power and latency numbers associated with the 10GBASE-T technology, as shown in Table 1. Other technologies seem to be better positioned to succeed in 10GE environments. In particular, SFP+ based twinax copper cables have a much lower cost than other solutions and enable pragmatic deployments of 10GE for very short distances in the data center (e.g., for intra-rack connections).

Other technological evolutions are enabling and driving the need for high bandwidth in converged network adapters.

An enabling technology is PCI Express [3]. Before PCI Express there was not enough bandwidth in the servers' bus to drive a 10Gb/s adapter. Supporting speeds from 2Gb/s to 32Gb/s, the PCI Express adoption in the industry has been and is a critical factor in moving toward I/O convergence.

Another driving technology for high bandwidth is the evolution of servers and processors. Processors are evolving toward multi-core architectures, allowing bigger and multiple workloads on a single system. The most pragmatic way of using these multi-core CPU architectures is through server virtualization. Consolidating on a single real system multiple virtual servers, each with its own bandwidth requirements, is a strong driving force toward higher network adapter speeds.

3. CONSOLIDATION TECHNOLOGIES

3.1 Overview

During recent years, several technologies attempted to achieve I/O consolidation, however they had a quite limited success. Fibre Channel has not been credible as a consolidated technology because it has been developed and pushed only for storage. Internet SCSI (iSCSI) [4], a SCSI transport protocol based on the TCP/IP suite of protocols, benefits from the ubiquity and common usage of the Ethernet LAN infrastructure. By leveraging existing deployments of Ethernet, iSCSI allows storage traffic to share commodity networking hardware with LAN traffic (e.g., NICs, LAN switches and low cost Cat5e/6 cabling), and this gives to it a cost advantage. InfiniBand [5], a 10Gb/s to 20Gb/s technology based on parallel optical or copper lanes, has been designed with the purpose of consolidation in mind, holding the promise for added flexibility and cost savings within virtualization environments. Yet both iSCSI and Infiniband have been adopted only in green field installations, because they do not provide a simple connection to existing SAN infrastructures (such as Fibre Channel based SANs). In order to connect with a Fibre Channel based SAN, both iSCSI and Infiniband require the deployment of stateful and complex gateways. In addition, iSCSI is built on top of TCP/IP and relies on the TCP processing [6] to recover from loss of frames. TCP processing is complex and not easy to implement in hardware. This reduces the iSCSI performances and increases the cost of the overall solution. A protocol designed to run over a lossless network (see 3.2) can be much simpler than a protocol that has to detect loss of packets and to recover them. Fibre Channel is much simpler than TCP and is easily and cheaply implemented in hardware.

The upcoming Fibre Channel over Ethernet (FCoE) protocol promises seamless connections between Ethernet and Fibre Channel infrastructures and is touted by many experts as the solution with the most possibilities of success for data center I/O consolidation.

The basic idea of FCoE is quite simple. Fibre Channel is easy to implement due to its lossless properties. If we find a way to extend Ethernet with a lossless behavior, then it becomes possible to define a direct mapping of Fibre Channel over Ethernet, not requiring any assistance from TCP in order to recover from frame losses. This direct mapping is achieved with a simple one-to-one encapsulation of Fibre Channel frames into Ethernet frames. Therefore FCoE is an extremely simple protocol, but requires the deployment of Ethernet extensions in order to run over the so called lossless Ethernet. Figure 3 shows a comparison of various transports for SCSI. The greater simplicity of FCoE in comparison with iSCSI and other protocols is evident.

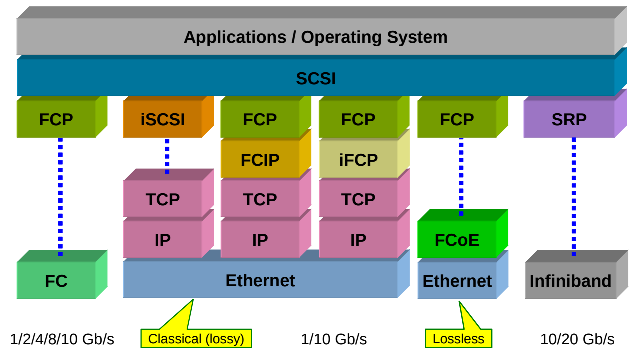

Figure 3. SCSI Transport Technologies.

An additional requirement for enabling a one-to-one mapping of Fibre Channel frames over Ethernet is the support for jumbo frames on Ethernet, because a Fibre Channel frame can be as big as 2140 bytes. Although not an IEEE 802.3 standard feature, jumbo frames are supported by most modern Ethernet implementations.

3.2 Lossless Networks

As described in [1], with lossless behavior we intend a network behavior that avoids loss of frames in presence of congestion in the network. Loss of frames due to link errors cannot be avoided, however the physical layer technologies discussed in 2.3 have a very low bit error rate, which mitigates the problem of loss of frames and allows recovering it at the upper layers.

A lossless Ethernet network is composed of full-duplex links only. For a consolidated infrastructure this is not a limitation, given that 10GE defines only full-duplex links. Even with this constraint, congestion is still possible on switches and the usual way to solve congestion in Ethernet is to lose frames. As described in [1], on a lossless network each receiving port short on buffer space is able to pace the rate of incoming frames by sending to the transmitter at the other end of the link a 'Pause' frame, asking the sender to stop transmission for a specified amount of time. The flow of frames may resume when more buffers become available or when the specified amount of time is reached. The Data Center Bridging task group of IEEE 802.1 is working on enhancements to the 'Pause' mechanism to be better suited for a consolidated network, providing different treatments for different kinds of traffic, the so called Priority-based flow control [1].

3.3 FCoE and other Protocols

Figure 4 shows the differences between FCoE and iSCSI when deployed in conjunction with a Fibre Channel fabric.

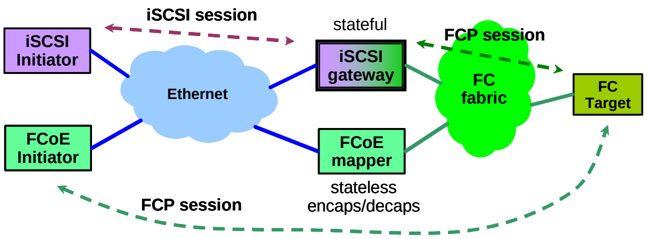

Figure 4. iSCSI and FCoE.

Both iSCSI and FCoE allow I/O consolidation on the Ethernet part of the configuration shown in Figure 4. In both cases converged adapters may be used on servers and the number of cables may be reduced. However, an iSCSI session (i.e., the mapping of a SCSI session over iSCSI) is different from a FCP session (i.e., the mapping of a SCSI session over Fibre Channel). In order to enable an iSCSI Initiator to speak with a Fibre Channel Target, the iSCSI gateway has to terminate the iSCSI session, initiate a FCP session, and convert information and maintain state between the two sessions. This has two major issues:

- the iSCSI gateway becomes a single point of failure for the infrastructure, being the only place maintaining the state mapping information;
- the iSCSI gateway does not scale when the number of iSCSI Initiators becomes large, because the gateway has to maintain mapping state per each iSCSI session. Management of all this information is a very difficult task.

Deploying Infiniband in conjunction with a Fibre Channel Fabric has the same issues.

Instead a FCoE Initiator opens a native FCP session with the Fibre Channel Target. The FCoE mapper has to perform only a stateless one-to-one encapsulation and decapsulation of Fibre Channel frames into FCoE frames, a function that does not require state and that can be easily distributed over the infrastructure. Therefore FCoE does not have scaling issues and allows a seamless integration with native Fibre Channel. This is a property that neither iSCSI nor Infiniband has.

3.4 FCoE Deployments

Figure 5 shows an example of deployment of FCoE, highlighting how seamless is the integration of a FCoE solution with an existing LAN and SAN environment. Fibre Channel achieves redundancy and high availability with two separate fabrics, called in Figure 5 SAN A and SAN B. Ethernet networks deploy high availability within a single network through redundant paths and switches. FCoE switches (i.e., Ethernet switches supporting lossless Ethernet and implementing the FCoE processing functions, including the FCoE mapper) allow the connection of consolidated servers to such an infrastructure, as shown in Figure 5, by keeping the separation of the two SANs and the single LAN. The lossless Ethernet network needs to be deployed only to connect the consolidated servers, all the remaining parts of the data center network and storage infrastructure do not need to be changed.

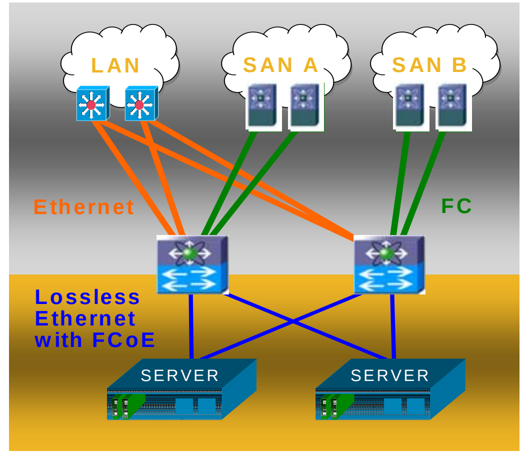

Figure 5. A Consolidated Network with FCoE.

4. FCOE BENEFITS

4.1 Overview

FCoE provides the benefits of any I/O consolidation solution, as discussed in 2.2, plus some additional advantages due to the fact that it is a mapping of Fibre Channel over Ethernet: seamless integration with existing Fibre Channel SANs and no need for stateful gateways.

4.2 Seamless Integration

As described in 3.3, there is no need for a stateful gateway to connect a FCoE device with a native Fibre Channel device. This property dramatically reduces the management overhead and improves the I/O efficiency, allowing seamless integration of FCoE devices with a Fibre Channel infrastructure.

4.3 Management Consistency

FCoE seamlessly extends the Fibre Channel storage management framework into the converged network. As shown in Figure 6, since the servers on the converged network continues to use the Fibre Channel protocol stack, existing storage resource management (SRM) software, personnel, storage management policies and processes can be easily extended across the entire consolidated infrastructure, which in turn lowers the operating cost of managing the data center.

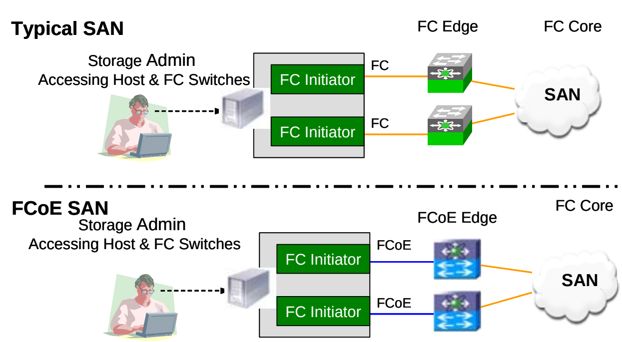

Figure 6. FCoE Maintains Consistent Management.

4.4 Investment Protection

Enterprises have made significant investments in Fibre Channel SANs. FCoE enables a seamless extension and protection of existing Fibre Channel investments by enabling servers connected to the converged network to tap into these resources using high performance FCoE switches.

4.5 Savings Case Study

Data center I/O convergence with FCoE provides big capital expenditures savings by reducing the number of cables and the number of entities that need to be managed. Figure 7 shows an example of networks with and without FCoE technology.

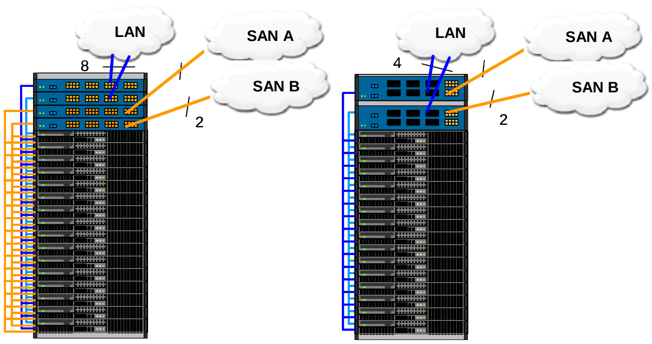

Figure 7. Example of Cables Reduction with FCoE.

Figure 7 shows two possible rack configurations for a set of 16 servers. The left rack deploys a conventional network structure, that has the SAN and the LAN completely separated. For redundancy purposes, there is the need to deploy two LAN Ethernet switches and two SAN Fibre Channel switches, all shown on the top of the rack. Each server thus has two cables for connecting to the Ethernet LAN switches and two cables for connecting to the Fibre Channel SAN switches with redundancy. The right rack deploys a consolidated network with FCoE. Only two FCoE switches which handle the converged LAN/SAN data are needed in this case, still providing full redundancy. Each server thus has two cables for connecting to the FCoE switches with redundancy. The number of cables used in these two configurations is shown in Table 2.

Table 2. Number of Cables with and without FCoE.

	Without FCoE	**With FCoE**
Cables per server	4	2
FC server cables	32	N/A
Ethernet server cables	32	N/A
FCoE server cables	N/A	32
FC uplink cables	4	4
Ethernet uplink cables	8	4
Total cables	**80**	**42**

With FCoE deployed in servers and switches, the total hardware savings include 16 switch ports, 16 adapters (in servers) and 38 cables. In addition, the number of switching platforms to be managed is also reduced, from 4 to 2.

5. FCOE STATUS

5.1 FCoE Standardization

Standardization of the FCoE protocol begun in June 2007 in the FC-BB-5 working group of the INCITS T11 Technical Committee [7, 8]. Since the beginning there has been a huge interest in this protocol and this pushed the group to reach quick agreements on all controversial topics. In August 2007 the FCoE frame format was agreed [9]. In October 2007 the FCoE functional models were agreed [10]. In February 2008 the MAC address structures were agreed [11]. In June 2008 the details of the FIP auxiliary control protocol were agreed [12]. Completion of the FC-BB-5 standard, where FCoE is defined, is expected for December 2008.

5.2 FCoE Implementations

FCoE implementations have been announced by several vendors and are coming to the market. There were two “proof of concept” demonstrations of FCoE at Storage Networking World in October 2007 and April 2008, hosted by the Fibre Channel Industry Association (FCIA). In both cases there was a huge interest in the FCoE technology and the FCoE demonstrations among the IT managers visiting the shows. The feedback provided by the visitors was that FCoE is a very promising technology because it maintains the Fibre Channel operational model while providing the benefits of I/O consolidation. Several enterprises were ready to begin testing FCoE products in their laboratories as soon as they are available. Given both the fast standard development and the interest registered at the demonstrations, we expect FCoE to have a major impact in the market for data center infrastructures.

5.3 FCoE Tools

The progress of FCoE has added new needs for protocol tools serving the purpose of protocol verification and compliance tests.

Analyzer tools may be software based or hardware based. Both solutions allow capturing data on the wire, decoding it and visualizing it. However, a software based protocol analyzer has limited capability for high speed data captures. A hardware based protocol analyzer is instead able to capture all the data on the wire, giving a complete and accurate protocol analysis.

Some features that an analyzer tool may provide specifically for FCoE are:

- Multiple CRCs monitoring. A FCoE frame encapsulates an entire FC frame, including the CRC, thus it carries two CRCs. The ability to monitor both of them is a critical functionality to catch erroneous frames.
- Time correlation testing between FC and FCoE. When FCoE is deployed in conjunction with a native Fibre Channel fabric, a FCoE switch performs the FCoE mapping function between Ethernet and Fibre Channel. The ability to correlate the time of when a FCoE frame is detected with the time of when a FC frame is seen allows a crucial monitoring capability of the FCoE mapping function.

Not many tools are commercially available for FCoE, however some products have shown the potential capability to fulfill the needs in the near future. Examples of protocol analyzers being enhanced to support FCoE are the software based Wireshark [13] and the hardware based Xgig® analyzers [14]. Xgig is a multi-platform tester supporting multi-protocols, multi-functions, multi-time-sync ports and multi-data rates. Its traffic generation tool is capable of creating FCoE and FC traffic at wire rate for the purpose of stress and performance tests. Figure 8 shows some screen captures of Xgig when used to analyze FCoE.

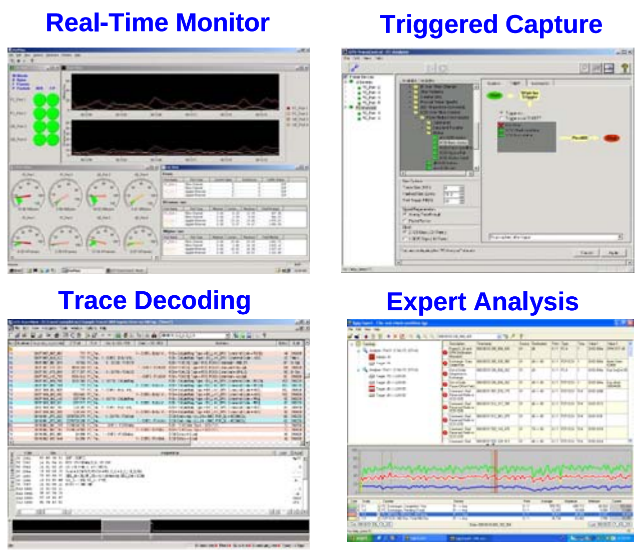

Figure 8. Some Xgig Screenshots for FCoE.

6. SUMMARY

I/O consolidation plays a critical role in current and future data center deployments, for the cost savings it promises to bring to enterprises, in terms of both reduction in equipment and simplification of management. Unified networks based on 10G Ethernet, enhanced with lossless properties, are the trend for future data center infrastructures. In this framework FCoE plays a central role, being a technology that provides the benefits of any I/O consolidation solution, plus the unique advantages of being a mapping of Fibre Channel over Ethernet: seamless integration with existing Fibre Channel SANs and no need for stateful gateways. There is a lot of interest in the market around the FCoE protocol and FCoE products are beginning to be deployed.

7. ACKNOWLEDGMENTS

Our thanks to Silvano Gai and Ralph Weber for their kind help on reviewing this paper.

8. REFERENCES

[1] C. DeSanti and J. Jiang, "FCoE in Perspective", ACM ICAIT, July 2008.

[2] S. Gai, "Data Center Networks and Fibre Channel over Ethernet (FCoE)", Lulu.com, ISBN-10: 1435714245, April 2008.

[3] PCI-SIG, http://www.pcisig.com/specifications/pciexpress/

[4] J. Satran et al., "Internet Small Computer Systems Interface (iSCSI)", RFC 3720, April 2004.

[5] Infiniband Trade Association, http://www.infinibandta.org/home/

[6] J. Postel, "Transmission Control Protocol", RFC 793, September 1981.

[7] C. Desanti, "FC-BB-5 Project Proposal", document T11/07-196v2, April 2007.

[8] C. DeSanti, et al., "Fibre Channel over Ethernet", document T11/07-303v0, June 2007.

[9] C. DeSanti, et al., "FCoE Convergence", document T11/07-479v0, August 2007.

[10] C. DeSanti, et al., "FC-BB-5 FCoE Models", document T11/560v1, October 2007.

[11] C. DeSanti, J. Jiang, et al., "FCoE: Common Addressing Structure", document T11/08-038v2, February 2008.

[12] C. DeSanti, et al. "Additional FIP Functions", document T11/08-263v1, June 2008.

[13] S. Gai, "FCoE Wireshark", document T11/07-548v0, October 2007.

[14] J. Jiang and D. Coupal, "Xgig® TraceViewer Support on FCoE", document T11/07-682v0, December 2007.