Fibre Channel over Ethernet (FCoE) Concepts

What is FCoE?

- **Fibre Channel over Ethernet (FCoE) is the transport of Fibre Channel “packets” over Ethernet**
  - Ethernet becomes the Fibre Channel physical interface
    - Ethernet NIC cards are the HBAs
    - Driver makes the NIC look like a traditional FC HBA
    - Ethernet switches make up the “Fabric”
  - Fibre Channel then becomes a transport protocol
    - There are no Fibre Channel frames (only Ethernet frames)
    - Fibre Channel frame content (“packets”) is delivered in the Ethernet frames
    - No Fibre Channel HBAs or switches are required
The Converged Ethernet Fabric

- Today, each application class has its own interface
  - Networking: Ethernet
  - Storage: Fibre Channel (or SAS or SATA)
  - Clustering: Infiniband

- This results in three different networks
  - Three different sets of hardware and cables
  - Three different tools and skill sets

- Instead, why not use a single “converged” network?
  - Fewer adapters and cables is especially important in the data center or blade servers
Converged Enhanced Ethernet (CEE)

Networking
TCP/IP
UDP
etc.

Storage
iSCSI
FCoE

Clustering
iWarp

Networking
TCP/IP
UDP
e etc.

Storage
iSCSI
FCoE

Clustering
iWarp

Ethernet NIC

Ethernet Switch(es)

FCoE Protocol “Stack”

<table>
<thead>
<tr>
<th>Application</th>
<th>Encapsulation</th>
<th>Applications</th>
<th>Transport</th>
<th>Ethernet</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCSI</td>
<td>iSCSI</td>
<td>FCP</td>
<td>IP</td>
<td>Fibre Channel</td>
</tr>
<tr>
<td>Encapsulation</td>
<td>FC/IP</td>
<td>FCP/IP</td>
<td>TCP/IP</td>
<td>Infiniband</td>
</tr>
<tr>
<td></td>
<td>TCP/IP</td>
<td>FCoE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FCoE Objectives

- Seamlessly and transparently replace the Fibre Channel physical interface with Ethernet
  - No change protocol mappings, information units, initialization steps, services, etc.
- Could be implemented totally in software using standard Ethernet NICs
  - Similar to iSCSI initiator driver
  - High-performance would require hardware assists, much as provided by existing FC HBAs
    - But no TCP/IP Offload Engines (TOEs)

Dedicated vs. Shared Storage Network

- The storage network may be dedicated to storage traffic
  - Much as Fibre Channel networks are dedicated
- Or, the network may be shared by storage and LAN traffic
  - Single adapter and interconnect for devices such as blade servers
  - Need to ensure adequate “quality-of-service” to the storage traffic
  - This can be provided by prioritizing storage (FC) traffic (e.g., as per IEEE 802.1Q)
**FCoE Converged Network**

- **Ethernet Switch(es)**
- **Ethernet Frames**
- **Ethernet Frames**
  - with FC Content (SCSI-FCP or FC-SATA)
- **Ethernet Frames**
  - (LAN Content)
- **Server**
  - (with Ethernet NIC)
- **Storage**
  - (with Ethernet NIC)

---

**Existing Mappings to Ethernet**

- There are existing storage mappings to Ethernet
  - Internet SCSI (iSCSI)
  - Internet FCP (iFCP)
  - Fibre Channel over IP (FCIP)
- All require TCP/IP
  - TCP/IP adds complexity and overhead
  - Argument is that this is not required in a local network
    - Still required for the WAN or long-haul
  - FCoE bypasses TCP/IP for efficiency and simplicity
FC over IP (FCIP) Example

- **FCIP has already defined an FC frame encapsulation method**
  - Requires use of TCP/IP
  - This introduces extra overhead in processing at
    the TCP and IP layers

- **For performance reasons, it would be nice to avoid TCP/IP altogether**
  - TCP in software is slow
  - TCP in hardware is complicated and expensive

---

**Ethernet Frame with FC Packet (Concept)**

- **Let's get rid of the TCP/IP overhead altogether**
- **Package the FC frame content directly in an Ethernet frame**
  - Less overhead = greater efficiency and performance
  - Simpler hardware can be used
Ethernet Frame with FC Packet (Concept)

- **FCoE eliminates all unnecessary overhead**
  - No TCP/IP
  - SOF and EOF are encoded into Ethernet fields
  - Ethernet frame CRC replaces FC frame CRC (same algorithm)
- **Still have FC frame header overhead**
  - Necessary for operation management and “gateway” functions

Relative Framing Overhead

- **Both FC and Ethernet have framing overhead**
  - Note that the difference in efficiency is on the order of 1 to 2%

**FC Framing Efficiency:**

\[
\text{Data} \quad 2112 = \frac{1500}{1544} = 97.24\%
\]

**Ethernet Framing Efficiency:**

\[
\text{Data} \quad 1500 = \frac{1544}{1544} = 97.15\%
\]

**FCoE Framing Efficiency (standard Ethernet Frames):**

\[
\text{Data} \quad 1500 = \frac{1550}{1550} = 95.66\%
\]

**FCoE Framing Efficiency (Ethernet Jumbo Frames):**

\[
\text{Data} \quad 2112 = \frac{2112}{2180} = 96.88\%
\]
FCoE Frame Considerations

- **FC frame content delivered in Ethernet frame**
  - 1 to 1 correspondence between FC frames and Ethernet frames
    - Each FC frame = 1 Ethernet frame
    - Multiple short FC frames are not put into the same Ethernet frame

- **FC frames can be larger than Ethernet frames**
  - FC data field maximum is 2112 bytes (+ 24 byte header + any extended headers)
  - Standard Ethernet frame data is 1500 bytes maximum
  - Options (tbd)?
    - Limit FC frame data field size during login
    - Use larger Ethernet frames ("jumbo" frames)

FCoE to FC Gateways

- **Desirable to mix FCoE and Fibre Channel in same configuration**
  - Requires a "Gateway" device
  - iSCSI gateways are complex and affect performance

- **FCoE Gateways are simple and efficient**
  - Simple frame translation with a 1:1 frame mapping
  - No need to remember state information
  - Extremely simple and low-cost
FCoE to FC Gateway

Server (with Ethernet NIC)

Ethernet Frames

Ethernet Frames with FC Content
(SCSI-FCP or FC-SATA)

FCoE to FC Gateway

Encapsulated FC Frame Content

Ethernet Frames

FC Frames

Ethernet CRC

Ethernet FO

FCoE to FC Gateway

Fibre Channel Frames
(SCSI-FCP or FC-SATA)

Storage (with Fibre Channel IF)

Lossless and Reliable Delivery

- Storage requires “reliable” delivery
- Bit Error Rate (BER)
  - Transmission errors can corrupt frames
  - Must provide an acceptable bit error rate to prevent frame corruption
- Frame Loss
  - Switches and devices must not discard frames
  - Flow control is necessary to prevent frame drop due to buffer conditions
Bit Error Rate Considerations

• Many Ethernet bit error rates are comparable to Fibre Channel
  – Bit Error Rate objective for 1 Gb and 10 Gb Ethernet is the same as for Fibre Channel (10⁻¹²)
• Some links may have higher error rates and may not be acceptable for storage
  – Cable plant may be more variable
  – Needs to be taken into consideration for FCoE usage

Fibre Channel Flow Control (Credit)

• A receiving port gives a sending port permission to send a specified number of frames
• That permission is called credit
  – When a frame is sent, the available credit is decremented (consumed)
  – When a reply is received, the available credit is incremented (replenished)
  – As long as a port has available credit, it may send additional frames
  – If the credit is exhausted, frame transmission is suspended until the credit is replenished
Ethernet “Pause” Flow Control

- Ethernet has an optional “pause” based flow control
  - Described in IEEE 802.3 Annex 31B
  - Receiver tells the sender when to pause or resume frame transmission (done in hardware, not software)
  - Receiver must send pause while there is enough buffer space to accommodate any frames in transit plus time for the pause to be received and processed

```
Receiver
Empty  Empty  Empty  Full  Full  Full  Full  Full  Full
Upper Threshold (Pause)

Lower Threshold (Resume)
```

```
Sender
Frame  Frame  Frame
Pause (time = xxx)
Pause (time = 0)
```

Whys (and Why Nots)
Why Use Ethernet?

• Why pick Ethernet as the base transport?
  – Ethernet is everywhere
    • There is a huge Ethernet infrastructure in place
    • Technology is well understood, skills and tools already in place
  – Ethernet is inexpensive
    • It offers the most “bang for the buck”
    • There is tremendous competition to drive prices down
  – Ethernet has raw speed
    • Gigabit is now mainstream
    • 10 Gbit is in its early deployment phase
    • 100 Gbit study group launched in 2006

Why Maintain Fibre Channel Content?

• Why not get rid of Fibre Channel altogether and use something else, such as iSCSI?
  – iSCSI has made inroads into storage, but the adoption has been slow
  – Often deployed where Fibre Channel is not already in use
• There is a significant install base of Fibre Channel today
  – Customers do not want to do a “rip and replace”
  – Fibre Channel is a proven technology
  – Fibre Channel supports protocols other than SCSI
    • What would be the solution for FICON without Fibre Channel?
    • FC-SATA opens opportunity for “tiered” storage environments
  – Fibre Channel will probably continue to provide the highest performance for the data center
Why not use iSCSI?

- **iSCSI is necessary for “lossy” or “out-of-order” networks**
  - e.g. many LANs, the Internet, long distance WAN solutions
- **iSCSI design was for TCP/IP networks**
  - TCP is a stateful, byte-oriented protocol
  - TCP processing adds additional overhead
    - or complexity of TCO Offload Engines
  - Memory needed for reassembly, reordering, and retransmission
  - Gateway between iSCSI and Fibre Channel is complex and expensive
    - iSCSI Information Units are different from FC Information Units
- **iSCSI provides recovery and flow control via TCP**
  - Not needed in a lossless Ethernet environment using Ethernet flow control

What About ATA over Ethernet (ATAoE)?

- **Companies have proposed (and implemented) ATA over Ethernet**
  - Map ATA commands to Ethernet frames
  - Software implementation (driver)
  - Limited products to date (e.g., Coraid)
- **ATAoE doesn’t address the SCSI command base (or other protocols)**
  - This includes most data centers and mid- to high-end servers
When?

- FCoE is in its very early stages
  - Need to develop Fibre Channel encapsulation standard
  - May need to develop Ethernet standards
  - Product development
- At this point, it is hard to pin down exact dates
Fibre Channel Standards Activity

- **Fibre Channel standards status:**
  - April 5, 2007: Initial presentations made to INCITS T11 standards body
  - June 2007: Expect a formal project proposal to develop an FCoE standard
  - 2008-2009 (?): Standard complete

- **There is much work to be done to define a complete solution**
  - Mapping FC frames to Ethernet frames is the simplest part
  - Must also address Fibre Channel functions such as:
    - Name Server
    - Fabric Controller
    - Zoning
    - State Change Notification
    - etc., etc.

Ethernet Standards

- **It's not clear if changes required to any Ethernet standards, and if so, the timeline**
  - Pause is required for reliable delivery
    - Already in in IEEE 802.3 Annex 31B)
  - VLANs (802.1Q)
    - This will provide priority and may replace zoning
  - Quality of Service considerations (802.1P)
  - Congestion Management (802.1au)
  - Routing: Transparent Interconnection of Lots of Links (TRILL)
Product Development

• No specific product dates yet
  – Early “proof-of-concept” products can be implemented in software
    • Much as the iSCSI driver is done in software
  – Later products can implement hardware assists for performance

• FCoE may end up with a “tiered” implementation structure
  – Software-based products for low cost
  – Hardware-assisted products for higher performance (at a higher cost)

Fibre Channel over Ethernet Supporters

• Multiple companies appear to be backing FCoE
• According to an article in Network World (4/5/2007) they include:
  – Brocade
  – Cisco
  – EMC
  – Emulex
  – IBM
  – Intel
  – Nuova Systems (a Cisco spinoff)
  – QLogic
  – Sun
End of “Fibre Channel over Ethernet (FCoE) Concepts”