Ethernet Technology

Allen Light, Broadcom Corp.
SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced without modification
  - The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
Abstract

Ethernet, the standard local area network (LAN) access method. A specification for "LAN," "LAN connection" or "network card" automatically implies Ethernet without saying so.

This session provides an overview of Ethernet technology, with an emphasis on the evolution of the standards from the original implementation of Ethernet on coax cable to the latest 10Gb Ethernet implementations.
Agenda

• The Original Standard
• Evolution of Ethernet
• Elements of Ethernet
  • The Frame / Addressing
  • Media Access Controller
  • Physical Media
'net-"w&rk

• A system of computers, terminals, and databases connected by communications lines

Local Area Network (LAN)

• A network of personal computers in a small area (like an office) that are linked by cable, can communicate directly with other devices in the network and can share resources

(from Merriam Webster)

• So why is this guy talking about a LAN technology at a storage networking conference?
Ethernet (IEEE 802.3)

- Ethernet (IEEE 802.3) was conceived by Bob Metcalfe in 1973.
- Formal specifications published in 1980 by DEC-Intel-Xerox
- Became IEEE 802.3 standard in 1985
  - 10 Mbps
  - Coaxial cable
- Updated to include new technologies
  - Twisted pair
  - 100 Mbps Fast Ethernet - IEEE 802.3U
- Next generation: Gigabit Ethernet - 802.3Z

*Nearly all of the traffic on the Internet either originates or terminates with an Ethernet connection*
Original Ethernet

- Single coaxial bus
- Multiple “Taps”, one per node
- Have a “shared media”
- CSMA/CD - Carrier Sense Multiple Access Collision
Shared vs. Switched

• Switched pairs communicate at same time: “point-to-point” connections

• Aggregate BW in switched network is many times shared
  – point-to-point faster since no arbitration, simpler interface

• Arbitration in Shared network?
  – CSMD/DS
  – OK if low utilization

• Modern implementations of Ethernet are Switched
Neat Facts

- Why is Ethernet capitalized?
  - The word Ethernet is capitalized to signify the official standard. Or it may be lowercased to suggest a medium without switches, routers and other intelligence.

- Why “ether”?
  - In 1973, Bob Metcalfe wrote: “While we may end up using coaxial cable trees to carry our broadcast transmissions, it seems wise to talk in terms of an ether, rather than ‘the cable’.... Who knows what other media will prove better than cable for a broadcast network: maybe radio or telephone circuits, or power wiring, or frequency-multiplexed cable TV or microwave environments, or even combinations thereof. The essential feature of our medium—the ether—is that it carries transmissions, propagates bits to all stations.”
## Ethernet Evolution

<table>
<thead>
<tr>
<th>Year</th>
<th>Standard</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>802.3</td>
<td>Ethernet (10 Mbps) Thick Coax</td>
</tr>
<tr>
<td>1986</td>
<td>802.3a</td>
<td>Ethernet Thin Coax</td>
</tr>
<tr>
<td>1991</td>
<td>802.3i</td>
<td>Ethernet Twisted Pair (TP)</td>
</tr>
<tr>
<td>1995</td>
<td>802.3u</td>
<td>Fast Ethernet (100 Mbps) Fiber, TP</td>
</tr>
<tr>
<td>1998</td>
<td>802.3z</td>
<td>Gig Ethernet (1 Gbps) Fiber</td>
</tr>
<tr>
<td>1999</td>
<td>802.3ab</td>
<td>Gig Ethernet Twisted Pair</td>
</tr>
<tr>
<td>2002</td>
<td>802.3ae</td>
<td>10G Ethernet (10 Gbps) Fiber</td>
</tr>
<tr>
<td>2006</td>
<td>802.3an</td>
<td>10G Ethernet Twisted Pair</td>
</tr>
<tr>
<td>2007</td>
<td>802.3</td>
<td>100Gig! Ethernet next data rate</td>
</tr>
</tbody>
</table>

The rapid evolution of Ethernet has taken the technology a long way from it’s original implementation.
## OSI, Fiber Channel and Ethernet / IP Communications

<table>
<thead>
<tr>
<th>Fiber Channel</th>
<th>OSI</th>
<th>Ethernet / IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC0 - Physical</td>
<td>Physical</td>
<td>Ethernet (802.3)</td>
</tr>
<tr>
<td>FC1 Encd/Decd</td>
<td>Data Link</td>
<td>TCP, UDP, etc.</td>
</tr>
<tr>
<td>FC2- Framing / Flow Co,trol</td>
<td>Transport</td>
<td>iWARP, iSCSI, FTP</td>
</tr>
<tr>
<td>FC3 and 4</td>
<td>Presentation</td>
<td>Http</td>
</tr>
<tr>
<td></td>
<td>Application</td>
<td>FireFox, etc.</td>
</tr>
</tbody>
</table>
Elements of Ethernet

- Ethernet Frame
- Medium Access Control (MAC)
- Physical Medium (PHY)

Ethernet consists of layers 1 and 2 of the Open System Interconnect (OSI) model.
Frame Format

Ethernet (IEEE802.3) format

<table>
<thead>
<tr>
<th>Destination Address 6 bytes</th>
<th>Source Address 6 bytes</th>
<th>EtherType 2 bytes</th>
<th>Message 46-1500 bytes</th>
<th>CRC-32 4 bytes</th>
</tr>
</thead>
</table>

- Ethernet addresses (48 bits) are unique and controlled by IEEE
  - 24 bit *Organizationally Unique Identifier (OUI)* by the IEEE
  - 24 bit by the organization

- **Multicast and Broadcast addresses**
  - High order address bit is 1 for multicast and broadcast
  - A destination address of only 1s is accepted by all stations
Higher level addresses

- Ethernet is a trucking system and can operate with different higher level protocols, like TCP/IP, AppleTalk, Novel etc.
- Higher level protocols have their own addressing schemes.
- They must find the right Ethernet address in order to communicate with each other.
Address Example

- Address Resolution Protocol (ARP for IPv4)
  - Station A sends a broadcast requesting the Ethernet address
    the station B that has the specific IP address
  - All stations receive the message
  - Only the station with the requested IP address reply

ARP Request Packet

<table>
<thead>
<tr>
<th>Destination Address</th>
<th>Source Address Requestor</th>
<th>EtherType 2 bytes</th>
<th>Message Request for IP Address</th>
<th>CRC-32 4 bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broadcast Requestor</td>
<td>Requestor Source Address</td>
<td>EtherType 2 bytes</td>
<td>Request for IP Address Message</td>
<td>Requested IP Address</td>
</tr>
</tbody>
</table>

ARP Response Packet

<table>
<thead>
<tr>
<th>Destination Address Requestor</th>
<th>Source Address Responder</th>
<th>EtherType 2 bytes</th>
<th>Message Response</th>
<th>CRC-32 4 bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requestor Source Address Responder</td>
<td>Responder Source Address</td>
<td>EtherType 2 bytes</td>
<td>Response Message</td>
<td>Requested IP Address</td>
</tr>
</tbody>
</table>
What is an Ethernet PHY?

- PHYsical Layer Device

- The PHY converts digital data into waveforms that can be transmitted over coax cable, twisted pair cabling, optical cabling, or backplanes and can convert received signals back to digital data

- The PHY sits between the Media Connector (BNC, RJ-45, Optical) and the repeater, switch, or MAC controller logic
Gigabit Ethernet Media Types

Two optical standards plus shielded copper cable initially. Unshielded twisted pair (CAT5) eventually.
Changes with 10GbE

- **802.3ae defines 10GbE**
  - 10Gbps Data Rates
  - Full-duplex only; no more carrier-sensing multi-access / collision detection (CSMA/CD)
  - Optical Physical Layer
    - LAN PHY and WAN PHY options
      - WAN PHY compatible with SONET
- **802.3an adds twisted pair cabling**
  - Cat 6 and Cat 7
- **802.3ap adds backplane specifications**
  - Blade servers and communications equipment
10GbE Standards

802.3ae

Media Access Controller (MAC) Full-Duplex

10 Gigabit Media Independent Interface (XGMII) or 10 Gigabit Attachment Unit Interface (XAUI)

LAN PHY (8B/10B)
Serial LAN PHY (64B/66B)
Serial WAN PHY (64B/66B + WIS)
Various Optical Technologies

10G BaseT
Cat6/7 Cable
802.3an

10G Backplane
802.3ap

802.3an and 802.3ap make 10GbE even more compelling for data center applications
10Gig PHY Technology

- Several options targeted for specific application, span length, and media
  - Fibre (Optical)
    - Single Mode Fibre
    - Multi-mode Fibre
    - Many choices
  - Copper
    - 10G Base-T
    - Serial 10G Copper Cable
  - Backplane
    - XAUI (4 x 3.125)
    - 10G Base-KR
# 10Gb Optical Choices

<table>
<thead>
<tr>
<th>Device</th>
<th>Range</th>
<th>Optics</th>
<th>PCS</th>
<th>WIS</th>
<th>Fiber</th>
</tr>
</thead>
<tbody>
<tr>
<td>10GBASE-LX4</td>
<td>300m MMF / 10km SMF</td>
<td>1310nm WWDM</td>
<td>8B/10B</td>
<td>No</td>
<td>Multimode or singlemode</td>
</tr>
<tr>
<td>10GBASE-SR</td>
<td>300 m</td>
<td>850nm</td>
<td>64B/66B</td>
<td>No</td>
<td>Multimode</td>
</tr>
<tr>
<td>10GBASE-LR</td>
<td>10 km</td>
<td>1310nm</td>
<td>64B/66B</td>
<td>No</td>
<td>Singlemode</td>
</tr>
<tr>
<td>10GBASE-ER</td>
<td>40 km</td>
<td>1550nm</td>
<td>64B/66B</td>
<td>No</td>
<td>Singlemode</td>
</tr>
<tr>
<td>10GBASE-SW</td>
<td>300 m</td>
<td>850nm</td>
<td>64B/66B</td>
<td>Yes</td>
<td>Multimode</td>
</tr>
<tr>
<td>10GBASE-LW</td>
<td>10 km</td>
<td>1310nm</td>
<td>64B/66B</td>
<td>Yes</td>
<td>Singlemode</td>
</tr>
<tr>
<td>10GBASE-EW</td>
<td>40 km</td>
<td>1550nm</td>
<td>64B/66B</td>
<td>Yes</td>
<td>Singlemode</td>
</tr>
</tbody>
</table>

PCS = Physical Coding Sub layer

WIS = WAN Interconnect Sub layer

IEEE has been working on a new 10GBASE-LRM standard for intra-building legacy MMF in the datacenter (~300M)
IEEE 10GE Standards

- IEEE P802.3ae, **10GE Working Group**
  - Definition of the 10GE MAC architecture and components
- IEEE P802.3an, **10GBASE-T Task Force**.
  - Definition of 10GE encoding over CAT 6/7 (Copper) cable
- IEEE P802.3ap, **Backplane Ethernet Task Force**.
  - Definition of 1/10GE encoding over backplanes (KX, KX4, KR)
- IEEE P802.3aq, **10GBASE-LRM Task Force**.
  - Definition of short haul MMF/1310nm LR encoding
- IEEE P802.3ar, **Congestion Management Task Force**.
  - Definition of L2 congestion discovery and avoidance protocols and more
- IEEE P802.3as, **Frame Expansion Task Force**.
  - Not “Jumbo” frames; expansion of frame envelope, not data size
- IEEE P802.3at, **DTE Power Enhancements Task Force**.
  - Definition of how to deliver power at 40-60W over 1 and 10GE
10GE Provides a New Innovation Landscape

See tutorial on “Unified Fabric” for 10GbE for more on this vision
Q&A / Feedback

• Please send any questions or comments on this presentation to SNIA: tracknetworking@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

SNIA Education Committee

Allen Light - Author
Howard Goldstein – Tutorial and Presentation Review
Walter Dey – Tutorial Review
Robert Peglar – Tutorial Review
Matthew Brisse – Tutorial Review
Questions?
CSMA/CD

- Carrier Sense Multiple Access / Collision Detect
- Listen for the media to be quiet
- If quiet, you can transmit
- Listen while you transmit to make sure it is just what you are sending
- If not, then someone else is sending at the same time – COLLISION
- Backoff and repeat process
New MSA for Next Generation 10G Optical Modules - SFP+

- The SFP+ MSA has been kicked off to develop a low cost and low power 10G serial optical interface and leverages existing SFP standard
  - The SFP+ MSA Rev 1.0 released next week and expect stability by next Rev.
  - ~30 companies actively participating in this effort
  - ~65 companies monitoring activities
- Key difference versus existing XFP optical modules
  - SFP+ removes a 10G retiming stage (saving size, cost, and power)
  - SFP+ is a smaller form factor (~46% Area Savings)
  - SFP+ Power target for up to 10km optical links (>800mW)
- Mechanical Specification – SFF-8432 improved but compatible with existing SFP
Quad 2x4 SFP+ Concept

- The demand for XFP and SFP+ are for port density and lower cost.
  - Rack spacing and port density equates to real states

**Screw attach to faceplate**
- establishes consistent compression of EMI gasket between housing & face plate of system
- Prevents oil-canning of face plate

**Optional Screw attach points**
- Secures stacked connector (assembly to host board for handling purposes)

**Optional Screw attach points**
- Secures stacked guide (assembly to base of enclosure)