Ethernet

Fulvio Risso
Politecnico di Torino
Copyright notice

- This set of transparencies, hereinafter referred to as slides, is protected by copyright laws and provisions of International Treaties. The title and copyright regarding the slides (including, but not limited to, each and every image, photography, animation, video, audio, music and text) are property of the authors specified on page 1.
- The slides may be reproduced and used freely by research institutes, schools and Universities for non-profit, institutional purposes. In such cases, no authorization is requested.
- Any total or partial use or reproduction (including, but not limited to, reproduction on magnetic media, computer networks, and printed reproduction) is forbidden, unless explicitly authorized by the authors by means of written license.
- Information included in these slides is deemed as accurate at the date of publication. Such information is supplied for merely educational purposes and may not be used in designing systems, products, networks, etc. In any case, these slides are subject to changes without any previous notice. The authors do not assume any responsibility for the contents of these slides (including, but not limited to, accuracy, completeness, enforceability, updated-ness of information hereinafter provided).
- In any case, accordance with information hereinafter included must not be declared.
- In any case, this copyright notice must never be removed and must be reported even in partial uses.
Ethernet MAC

- **CSMA/CD**
  - Multiple Access: everyone can talk (potentially at the same time)
  - Carrier Sense: listen before talking
  - with Collision Detection: listen while talking

- **MAC Protocol**
  - Defined for bus-based topologies
  - Non-deterministic
  - No upper limit for the waiting time
Collision Detection (1)

- Stations can start talking at the same time
- The "Listen while talking" mechanism is able to detect the collision
- But... propagation speed is not infinite, therefore there may be some trouble in detecting the collision
Collision Detection (2)

- Collision happens because the signals has a finite propagation speed
- In real world, collision can be undetected by the “listen while talking” mechanism
- In order to detect the collision
  - All talkers must still be active when the collision happens
  - The following entities has a strong relationship
    - Max distance between stations
    - Signal propagation speed
    - Min duration of the talk
Collision Detection (3)

A starts transmitting $t_1$ and it detects the collision at time $t_1 + t_p - \varepsilon$

B detects collision $t_1 + t_p - \varepsilon$

Worst case:
- Host B and host A are at the maximum allowed distance
- Host B starts transmitting at time $t_1 + t_p - \varepsilon$ and it detects the collision at time $t_1 + t_p - \varepsilon$
- Host A detects the collision after $2t_p - \varepsilon$
Collision Detection (4)

- **Input data**
  - Maximum distance: $D_{\text{max}}$
  - Signal propagation speed: $S_{\text{signal}}$
  - Min frame size: $F_{\text{min}}$
  - Network bandwidth: $B$

- **Some math**
  - Collision Window $t_p = \frac{D_{\text{max}}}{S_{\text{signal}}}$
  - Min Frame duration = $F_{\text{min}} / B = 2 \cdot t_p$

\[
D_{\text{max}} = \frac{F_{\text{min}} \cdot S_{\text{signal}}}{2 \cdot B}
\]

- **Example**
  - Ethernet ($B=10\text{Mbps}$, $F_{\text{min}}=64+8\text{bytes}$, $S_{\text{signal}}=2\times10^5\text{Km/s}$): 5760 m
Collision Detection (5)

- In case of collision
  - The current packet is no longer transmitted
    - Be careful: we have to guarantee the *Min duration of the talk* anyway
  - Instead, we transmit a particular jamming sequence (32 bits) until we reach the min duration of the talk

- Collision is detected by
  - Transmitting stations:
    - Coax cable: measurement of the average DC on link
    - Twisted pair, fiber: activity on both links (tx and rx)
  - Other stations (e.g., receiving ones): wrong CRC on the received frame (which should be invalid)
Back-Off

- We have to re-transmit the lost frame
- After a collision, the Truncated Binary Exponential Back-off algorithm determines when the packet will be retransmitted
- Parameters:
  - \( \tau \) = time required to transmit a 512 bits slot
  - \( n \) = number of collisions on the current frame
- Algorithm
  - Between two consecutive transmissions, we have to wait at least
    \[ T = r \times \tau \]
  - Max 16 re-transmissions on the same frame
  - \( r \) is a time value that is randomly chose between
    \[ 0 \leq r < 2^k \quad k = \min(n, 10) \]
Performances of the CSMA/CD

- Possible problems when load reaches 100%
  - Reasonable at 50% load
- Simple and distributed protocol
  - No intermediate devices, synchronization, etc
- No upper bound on delay
  - In theory not suitable for real-time communications
Ethernet DIX and IEEE 802.3

- Network
  - Data link
    - Ethernet 2.0
  - Physical
    - CSMA/CD

- LLC
- ISO 8802.2 Logical Link Control

- MAC
  - 802.3 ISO 8802.3
  - 802.5 ISO 8802.5
  - FDDI ISO 9314

- Ethernet 2.0 (DIX)
- Standard ANSI/IEEE and ISO/IEC
Ethernet frame (DIX and IEEE 802.3) (1)

Frame length: from 64 to 1518 bytes

<table>
<thead>
<tr>
<th>Preamble</th>
<th>SFD</th>
<th>MAC Dest.</th>
<th>MAC Source</th>
<th>Type</th>
<th>Data</th>
<th>FCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{≤ 1500} \]

\[ \geq 1536 \ (0x600) \]
Most protocols use direct encapsulations in Ethernet v2.0

- IP: 0x800
- ARP: 0x806
- IPv6: 0x86DD

Some (most IEEE-derived protocols such as 802.1d) use LLC
Delimiting an Ethernet frame

- **Preamble**
  - Up to 7 bytes for sync source/receiver

- **Start of Frame Delimiter**
  - Special byte for signaling the start of frame (invalid code at L1)

- **Be careful: No end-of-frame delimiter**

- **Inter-packet gap: min silence between a frame and the following**
  - 96 bit times
  - \( \rightarrow 9.6\mu s \)
  - Called “Inter-Frame Spacing” in IEEE 802.3
Ethernet physical layers

- 10 Mbps
  - 100ns bit time
- 802.3 defines different standards:
  - 10BaseT: twisted pairs (max 100 m)
- No longer in use:
  - 10Base5: thick coax (500 m)
  - 10Base2: thin coax (185 m)
  - FOIRL: fiber cable, asynchronous, for repeater-to-repeater connections (1000 m)
  - 10BaseFL: fiber cable, asynchronous, for stations or repeater-to-repeater connections (2000 m)
  - 10BaseFB: fiber cable, synchronous, for repeater-to-repeater connections (2000 m)
- Physical coding: usually Manchester
  - 10MHz
Physical layer: Coax cable

- No longer in use
  - Yellow cable (IBM)
  - “Thin” cable (RG58)
  - “Thick” cable (RG213)
Physical layer: Thick coax cable

- Characteristics
  - Max length cable: 500 m
  - Max length single clip: 117 m
  - Min distance between transceivers: 2.5 m
  - Max number of transceivers: 100
  - Max length transceiver cable: 50 m
  - Vampire taps

- Diagram:
  - Terminator (50 Ohm)
  - Transceiver
  - Barrel connector
  - Coax cable 50 Ohm (thick)

- Host
  - Transceiver cable

- Coax cable 50 Ohm (thick)
Physical layer: Thin coax cable

- Characteristics
  - Max length cable: 185 m
  - Max number of stations: 30
  - Min distance between stations: 0.5 m
  - Max length transceiver cable: 50 m

![Diagram of Thin Coax Cable Network]
Twisted pair (1)

- UTP cable (min category 3)
- Max length: 100 m
  - Configuration details are more complicated

RJ45 wall socket

RJ45 connector
Twisted pair (2)

Possible cables

- UTP (Unshielded): not shielded
- FTP (Foiled): a single shield
- STP (Shielded): global shield + a shield for each twisted pair
Fiber

- No sensitivity to electromagnetic fields
- Larger distances
- Higher costs
  - Cabling, crimping
- Less flexible
  - Fiber to the phone, in-field crimping
Ethernet topology

- Rather limited
- ~200 m diameter
  - Due to twister pair cable limits
- Larger networks with repeaters (~ 3Km)
Fast Ethernet: IEEE 802.3u (1)

- Characteristics
  - Same frames, same CSMA/CD algorithm

- Physical layers
  - 100BASE-T4 (twisted pair cable, 4 pairs)
    - 8B/6T: 37.5MHz
  - 100BASE-TX (twisted pair cable, 2 pairs)
    - 4B/5B + MLT-3: 31.25MHz
  - 100BASE-FX (fiber)
    - 4B/5B
  - TX, FX: derived from TP-PMD/PMD of FDDI (ISO 9314-3) with minor modifications
Fast Ethernet: IEEE 802.3u (2)

**Differences**

- 10x increase in speed
  - Data Rate 100Mb/s
  - Bit time 10ns
  - Interpacket gap 0.96μs
  - Slot time 5.12μs (512 bits / 64 bytes)

- /10 in distance (200m + 20m)
  - Reduced collision domain
  - Basically, Host – hub – host
  - Rather limiting
Fast Ethernet topology
Auto negotiation (1)

- Auto negotiation possibilities:
  - speed (only over copper)
  - half/full duplex (over copper and fiber optic)

- Negotiation sequence:
  - 1 Gb/s full-duplex
  - 1 Gb/s half-duplex
  - 100 Mb/s full-duplex
  - 100 Mb/s half-duplex
  - 10 Mb/s full-duplex
  - 10 Mb/s half-duplex
Auto negotiation (2)

- It requires both parties to be active, otherwise it assumes it is connected to an hub
  - Fixed setting on one side may lead to unexpected errors

- Example
  - One side: fixed 100Mbps Full Duplex
  - The other party does not receive any message and it will assume it is connected to an hub
    - It will configure the interface in 100Mbps Half Duplex

- Auto-negotiation does not work with hubs
  - Hubs operate at fixed speed
Gigabit Ethernet: IEEE 802.3z (1)

- 1Gbps (10x in speed, /10 in distance)

Characteristics

- Same frame
  - Required to maintain interoperability with other Ethernet standards

- Same CSMA/CD algorithm
  - Is becoming obsolete
    - Collision domain ~20m

- Compatibility at frame level is more important than compatibility at CSMA/CD level
Gigabit Ethernet: IEEE 802.3z (2)

- More modifications are required compared to FastEthernet
  - CSMA/CD
    - No longer used in practice
    - Full duplex is the standard operating mode
      - Introduced with FastEthernet, but initially CSMA was still largely used
      - No GE devices have been made which support CSMA/CD
  - Increased slot time, Carrier extension
  - Frame bursting

- Why Gigabit Ethernet?
  - Well, hardware is cheap
  - Market demand (and vendor offer)
  - May be useful in the server domain and for backbone links
Gigabit Ethernet Slot Time

- Increased to 512 bytes
  - ~200m diameter (star-based topology: 100m + hub + 100m)
  - Cannot increase minimum packet size (for compatibility)

<table>
<thead>
<tr>
<th></th>
<th>Ethernet</th>
<th>Fast Ethernet</th>
<th>Gigabit Ethernet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission speed</td>
<td>10 Mbps</td>
<td>100 Mbps</td>
<td>1 Gbps</td>
</tr>
<tr>
<td>Bit time</td>
<td>100 ns</td>
<td>10 ns</td>
<td>1 ns</td>
</tr>
<tr>
<td>Inter-packet gap</td>
<td>9.6 us</td>
<td>0.96 us</td>
<td>96 ns</td>
</tr>
<tr>
<td>Slot time</td>
<td>51.2 us</td>
<td>5.12 us</td>
<td>4.096 us</td>
</tr>
</tbody>
</table>
Carrier Extension

- Extends short packets to min 4096 bit times
  - Predefined sequence of symbols

<table>
<thead>
<tr>
<th>Preamble</th>
<th>SFD</th>
<th>MAC Dest.</th>
<th>MAC Source</th>
<th>Len./Type</th>
<th>Data</th>
<th>FCS</th>
<th>Extens.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>46 - 1500</td>
<td>4</td>
<td>0 - 448 (3584 bit)</td>
</tr>
</tbody>
</table>

Min packet size (64 byte, 512 bit times)

FCS coverage

Min transmission length (512 + 3584 = 4096)

Collision window
Frame Bursting (1)

- Ethernet has MTU equal to 1500 bytes
  - Due to the necessity to facilitate statistical multiplexing among sessions
  - No longer appropriate in Gigabit Ethernet
- Gigabit Ethernet allow an host to transmit several consecutive packets without releasing the channel
  - Burst-limit equal to 65536 bit (8192 bytes) + 1 packet
Frame Bursting (2)

- **Mechanism**
  - First packet must have at least min size (i.e. must be be extended if shorter than slot time)
  - Replace InterPacketGap with an appropriate Filling Extension
    - Required in order to delimit packets
  - Other hosts must wait till the packet ends (with IPG)

```
Packet 1 (+ extension)  IPG + FILL  Packet 2  IPG + FILL  IPG + FILL  Packet N  IPG
```

Burst limit (65536 bits)
Working modes

- Shared mode to be used with repeaters
  - Not used
  - Not implemented by any commercial products
- Usually used in full duplex mode
  - No carrier extension
    - Collisions does not exist
  - No burst mode
    - Contention does not exist
# Gigabit Ethernet: Physical layer

<table>
<thead>
<tr>
<th>Standard</th>
<th>Media</th>
<th>Use</th>
<th>Max leng.</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000BASE-SX</td>
<td>MMF 50/125 um (400 MHz * Km a 850nm)</td>
<td>2 fibers</td>
<td>500 m</td>
<td>FC: 8B10B</td>
</tr>
<tr>
<td></td>
<td>MMF 50/125 um (500 MHz * Km a 850nm)</td>
<td></td>
<td>550 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MMF 62.5/125 um (160 MHz * Km a 850nm)</td>
<td></td>
<td>220 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MMF 62.5/125 um (200 MHz * Km a 850nm)</td>
<td></td>
<td>275 m</td>
<td></td>
</tr>
<tr>
<td>1000BASE-LX</td>
<td>MMF 50/125 um (400/500 MHz * Km a 1300nm)</td>
<td>2 fibers</td>
<td>550 m</td>
<td>FC: 8B10B</td>
</tr>
<tr>
<td></td>
<td>MMF 62.5/125 um (500 MHz * Km a 1300nm)</td>
<td></td>
<td>550 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SMF 10/125 um</td>
<td></td>
<td>5000 m</td>
<td></td>
</tr>
<tr>
<td>1000BASE-CX</td>
<td>STP 2 pairs (jumper cable) 150 Ohm</td>
<td>25 m</td>
<td>FC: 8B10B</td>
<td></td>
</tr>
<tr>
<td>1000BASE-T</td>
<td>UTP 4 pairs</td>
<td>100 m</td>
<td>PAM5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>balanced 100 Ohm Cat. 5E</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MMF = Multi Mode Fiber
SMF = Single Mode Fiber
1000BASE-T (IEEE 802.3ab)

- Full-duplex transmission over 4 pairs
  - 250 Mb/s per pair
  - Hybrid transformers

- PAM5 Line coding (5-level Pulse Amplitude Modulation)
  - 6 binary symbols encoded in quinary symbol quadruple
  - Each symbol transmitted over a pair
    - 125 Mbaud per pair
  - Redundancy used for control codes

- Cat 5 UTP has to pass further test in addition to those provided by TIA/EIA TSB95 standard on structured wiring
1000BASE-X

- Sub Standard
  - 1000BASE-CX (copper short range)
  - 1000BASE-SX (short wavelength)
  - 1000BASE-LX (long wavelength)
- Based on Fiber Channel (FC) Physical Layer
  - Code 8B10B
  - Redundancy code: control symbol and transitions
1000BASE-CX connectors

Type 1 connector
- 1: Transmission +
- 6: Transmission -
- Shell: shield
- 5: Reception -
- 9: Reception +

Type 2 connector
- 1: Transmission +
- 3: Transmission -
- 6: Reception -
- 7: Reception +

Type 2 connector socket
1000BASE-SX and 1000BASE-LX connectors
Wave-Length and standard

<table>
<thead>
<tr>
<th>Wave-Length and standard</th>
<th>dB/km</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000BASE-SX</td>
<td></td>
</tr>
<tr>
<td>I Window 850nm</td>
<td></td>
</tr>
<tr>
<td>II Window 1310nm</td>
<td></td>
</tr>
<tr>
<td>III Window 1550nm</td>
<td></td>
</tr>
<tr>
<td>Visible Light</td>
<td></td>
</tr>
</tbody>
</table>
1000BASE-LX & multimode fiber: Mode Conditioning Patch Cord

SC connectors
B E I G E C o l o r

Equiment 1000BASE-LX port

RX

TX

SC Connector
B L U C o l o r

Optical patch panel

Junction between SMF and MMF fiber

SC Connector
B E I G E C o l o r

MMF = Multi Mode Fiber
SMF = Single Mode Fiber
Non standard products

- 1310 nm single-mode fiber: 10 Km
  - Example Cisco GBIC 1000BASE-LX/LH
- 1550 nm single-mode fiber dispersion shift: 100 Km
  - Example Cisco GBIC 1000BASE-LZ
- Interoperability between products of different vendors is not guaranteed
10 Gigabit Ethernet - IEEE 802.3ae

- IEEE 802.3 frame
- Full-duplex mode
  - No repeater
  - No CSMA/CD
  - No carrier extension
- Keep Ethernet’s good reputation
  - 10 times more efficient
  - 3 times more expensive
- Break into metropolitan network (MAN) and wide area network (WAN) markets
  - Price/Bandwidth ratio is higher than traditional solutions (SONET/SDH, Frame Relay, ATM)
WAN PHY

- Enables transport over existent MAN and WAN infrastructure
  - DWDM (Dense Wavelength Division Multiplexing)
- Enables existent MAN and WAN component reuse
  - SONET/SDH transceiver and circuitry
- Different transmission speed (9.6 Gb/s) respect to LAN PHY’s speed
- WAN PHY and LAN PHY common properties → market is waiting for components with both functionalities
  - 10GBASE-R and 10GBASE-W in particular
- WIS (WAN Interface Sublayer) tunes PCS’ signal
  - Bit scrambling
  - SONET/SDH headers
10GE frame over SONET/SDH

STS-192c = Synchronous Transport Signal – of level 192, c = concatenated

SPE = Synchronous Payload Envelope

Data stream from PCS

Path Overhead Column

Fixed Stuff (not used)
10GE and SONET/SDH

- Simplified version of SONET/SDH
  - Avoid imposed complexities required by SONET/SDH
  - Limit component cost

- Only some header’s fields are used

- High precision synchronization has been removed
  - No Stratum-1 clock ($10^{-12}$ precision)

- Frames are generated and forwarded by 10GE devices in asynchronous mode using
  - SONET/SDH framing
  - Limited SONET/SDH management functionalities
## Physical layer

<table>
<thead>
<tr>
<th>Standard</th>
<th>Fiber</th>
<th>Max length</th>
<th>Window</th>
<th>Usage</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>10GBASE-SR</td>
<td>Multimode 62.5 µm</td>
<td>26 – 33 m</td>
<td>850 nm</td>
<td>Building (horizontal</td>
<td>64B/66B</td>
</tr>
<tr>
<td></td>
<td>Multimode 50 µm</td>
<td>66 – 300 m</td>
<td></td>
<td>wiring)</td>
<td></td>
</tr>
<tr>
<td>10GBASE-LR</td>
<td>Monomode (10 µm)</td>
<td>10 Km</td>
<td>1310 nm</td>
<td>Area</td>
<td>64B/66B</td>
</tr>
<tr>
<td>10GBASE-ER</td>
<td>Monomode (10 µm)</td>
<td>40 Km</td>
<td>1550 nm</td>
<td>Metropolitan</td>
<td>64B/66B</td>
</tr>
<tr>
<td>10GBASE-LX4</td>
<td>Multimode 62.5 µm</td>
<td>300 m</td>
<td>1310 nm</td>
<td>Building (horizontal</td>
<td>FC 10G: 8B10B</td>
</tr>
<tr>
<td></td>
<td>Multimode 50 µm</td>
<td>240 – 300 m</td>
<td></td>
<td>wiring)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monomode (10 µm)</td>
<td>10 Km</td>
<td></td>
<td>Area</td>
<td></td>
</tr>
<tr>
<td>10GBASE-SW</td>
<td>Multimode 62.5 µm</td>
<td>26 – 33 m</td>
<td>850 nm</td>
<td>Building (horizontal</td>
<td>64B/66B SONET/SDH framing</td>
</tr>
<tr>
<td></td>
<td>Multimode 50 µm</td>
<td>66 – 300 m</td>
<td></td>
<td>wiring)</td>
<td></td>
</tr>
<tr>
<td>10GBASE-LW</td>
<td>Monomode (10 µm)</td>
<td>10 Km</td>
<td>1310 nm</td>
<td>Area</td>
<td>64B/66B SONET/SDH framing</td>
</tr>
<tr>
<td>10GBASE-EW</td>
<td>Monomode (10 µm)</td>
<td>40 Km</td>
<td>1550 nm</td>
<td>Metropolitan</td>
<td>64B/66B SONET/SDH framing</td>
</tr>
</tbody>
</table>
10GBASE-X

- Coding derived from 10G FC (Fiber Channel at 10 Gb/s)
- 32 bit blocks are encoded in 4 blocks of 10 bit each
- Sent over 4 lane
  - 3.125 Gbaud per lane
- Redundancy used for control codes
  - For example idle signal act as inter-frame gap
10GBASE-LX4