Ethernet evolutions

Fulvio Risso

Politecnico di Torino
An historic perspective (1)

- Ethernet was very successful
  - On the wired side, basically no competitors
- New problems soon appeared
  - Necessity of more speed
    - FDDI was used to create backbones, but expensive and not appropriate for desktops
  - Necessity to interconnect more networks
    - FDDI was not able to interconnect Token Ring networks
      - MTUs: FDDI 4352, 4Mb Token Ring 4464, 16Mb Token Ring 17,914
An historic perspective (2)

- Features not negotiable in new LAN standards
  - Keep compatibility with frame format
  - Preserve investments in human workforce
    - Hard to convert people to new technologies

- Two new specs appeared
  - 100VG-Anylan (Hewlett Packard)
    - Demand Priority Protocol, a sort of polling-based mechanism
  - Fast Ethernet (3Com)
    - Classical Ethernet, but ported to 100Mbps
An historic perspective (3)

- 100VG-Anylan disappeared soon
- FastEthernet: slow adoption at the beginning
  - FDDI better for backbone (robustness and network diameter)
  - No such need of speed at the desktop
  - Mainstream from ‘95-’96
  - No interconnection for Token Ring networks (MTU issues)
    - Killed Token Ring Technology
- Later, Gigabit Ethernet
  - Mainly for backbones, now also for desktop
- Now, 10GB Ethernet
  - Intended to cover MANs as well
- Even faster Ethernet are in the pipeline
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
    - Other standards have been defined, but never implemented
      - The existing ones worked extremely well
Fast Ethernet: IEEE 802.3u (2)

- Differences
  - 10x increase in speed
    - Data Rate 100Mb/s
    - Bit time 10ns
    - Inter-frame 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
  - Introduces “Full Duplex” mode
    - No CSMA/CD on that link
Fast Ethernet topology (1)

- Limited network size, but still usable
  - Compatible with structured cabling limits (100m)
Fast Ethernet topology (2)

- Bridges/switches becomes common in those times
  - No CSMA/CD → less limitations to the network limit
    - The limit becomes the attenuation on cables
- Very complex topologies (several bridges in cascade)
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)

- Possible only if connected to another host, or to a bridge/switch
  - Hubs operate at fixed speed; hence, cannot negotiate anything!
- If, during the procedure, the other party does not respond, the negotiating station 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
    - There may be tons of false collisions on that host
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 format, 64/1536 bytes frame size
  - In principle, same CSMA/CD algorithm
    - Although collision domain was ridiculous (~20m), if other parameter kept unchanged
  - At a first sight, FastEthernet at 10x speed

- A note: compatibility at frame level is more important than compatibility at CSMA/CD level
GbE and collision domain (1)

- A simple extension of FastEthernet originates a toy network
  - Usual rule: speed 10x, network diameter 10/
  - Collision domain was ridiculous (~20m), if other parameter kept unchanged
- Needed to define some tricks in order to increase the network diameter
  - Modification in the slot-time (presented later)

- But... was CSMA/CD still needed in GbE?
GbE and collision domain (2)

- A lesson learned from Fast Ethernet
  - Very slow adoption at the beginning
    - The limited network diameter was really a show stopper
    - Is it reasonable to have a backbone max 200m wide?
      - Most people still used FDDI, which did not have such limitations
  - FastEthernet become successful only when switches become mainstream
  - The idea of even faster Ethernet standards (e.g. Gbe) with even a smaller network diameter was known to be very stupid from the beginning

- So, why Gigabit Ethernet used CSMA/CD?
  - Not really needed from the technical point of view
    - A pure switched network does not use CSMA/CD and it works
  - Economic reasons?
GbE and collision domain (3)

- So, CSMA/CD specified anyway
  - With some additional tricks to have a reasonable network diameters
- And so were repeaters, etc.

- Nobody never implemented those specs
  - All GbE product did not use the CSMA/CD
  - Worked only in a pure switched (i.e., full-duplex) environment
**Gigabit Ethernet vs Fast Ethernet**

- In fact, GbE brings some new idea compared to FastEthernet, which is simply a 10x Ethernet
  - Increased slot time
  - Added Carrier Extension
  - Added Frame bursting
  - Full-duplex becomes, in fact, the standard operating mode (CSMA/CD no longer used in practice)

- **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: Carrier Extension (1)

- The minimum frame size of 64 bytes limits the network diameter
  - Need to increase the minimum duration of the transmission
    - Please note that “Min duration of the transmission” is different from “minimum frame size”
  - But... cannot increase minimum frame size (for compatibility)
    - If so, how can we transport a 64B Ethernet frame into GE?
  - We increase the *slot time*
Gigabit Ethernet: Carrier Extension (2)

- Pad short frames to reach at least 4096 bit times (slot time)
  - Predefined sequence of symbols
    - Valid at physical layer, but not used when transporting Eth data
- Frames bigger than 4096 bits (512 bytes) are not extended

<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></td>
<td></td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>46 - 1500</td>
<td>4</td>
<td>0 - 448</td>
</tr>
</tbody>
</table>

FCS coverage
Min frame size (64 bytes, 512 bit times)
Min transmission length (512 + 3584 = 4096 bits)
Collision window
Gigabit Ethernet: Frame Bursting (1)

- The maximum frame of 1536 bytes is obsolete
  - In Ethernet, 1536 bytes $\rightarrow$ 1.2ms channel occupancy
  - Reasonable for guaranteeing statistical demultiplexing
  - Cannot increase the maximum frame size
    - If so, how can we transport a large GE frame into Ethernet?
- We concatenate several frames one after the other: frame bursting

Network A
$F_{\text{max}} = 1.5\text{bytes}$

Network B
$F_{\text{max}} = 8\text{Kbytes}$
Gigabit Ethernet: Frame Bursting (2)

- Gigabit Ethernet allows an host to transmit several consecutive frames without releasing the channel
  - Burst-limit equal to 65536 bit (8192 bytes) + 1 frame
  - The frame currently under transmission when the 8192 limit is passed is still allowed to go on

- Advantages
  - Carrier extension (optionally) present only after the first frame
  - No “lost” time in contention after each frame (only after the burst)
  - Throughput increases especially in case of short frames
  - The maximum frame size is still 1536 bytes
    - Compatibility with older Ethernet networks
  - Very simple implementation (a counter when exceeding 8192 bytes)
Gigabit Ethernet: Frame Bursting (3)

- **Mechanism**
  - First frame must have at least min size (i.e. must be extended if shorter than slot time)
  - Inter-Frame Gap is still present, but the physical coding is different in order to distinguish this case from the “standard” IFG
    - Called “Filling Extension” (indicated as “IFG+FILL” in the picture)
    - Required in order to delimit frames
    - Always 96 bit times
  - Other hosts must wait till the frame ends (with IFG)
  - **All frames include SFD, Preamble and the actual frame**

<table>
<thead>
<tr>
<th>Frame 1 (+ extension)</th>
<th>IFG + FILL</th>
<th>Frame 2 (no ext)</th>
<th>IFG + FILL</th>
<th>IFG + FILL</th>
<th>Frame N (no ext)</th>
<th>IFG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burst limit (8192 bytes + 1 frame)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GbE: recap of most important parameters

- Slot time increased to 4096 bits (512 bytes)
  - ~200m diameter (star-based topology: 100m + hub + 100m)
- IFG kept at 96 bit times
- Bit time reduced to 1/10, hence speed increased 10x

<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-frame 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>
Working modes

- Shared mode (i.e., CSMA/CD) to be used with repeaters
  - Not used
  - Not implemented by any commercial products

- GbE usually deployed 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>Cabling</th>
<th>Use</th>
<th>Max length</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000BASE-SX</td>
<td>MMF 50/125 um (400 MHz * Km, 850nm) MMF 50/125 um (500 MHz * Km, 850nm) MMF 62.5/125 um (160 MHz * Km, 850nm) MMF 62.5/125 um (200 MHz * Km, 850nm)</td>
<td>2 fibers</td>
<td>550m 550m 220m 275m</td>
<td>8B10B</td>
</tr>
<tr>
<td>1000BASE-LX</td>
<td>MMF 50/125 um (4/500 MHz * Km, 1300nm) MMF 62.5/125 um (500 MHz * Km, 1300nm) SMF 10/125 um</td>
<td>2 fibers</td>
<td>550m 550m 5Km</td>
<td>8B10B</td>
</tr>
<tr>
<td>1000BASE-CX</td>
<td>STP (jumper cable), 150Ohm</td>
<td>2 pairs</td>
<td>25m</td>
<td>8B10B</td>
</tr>
<tr>
<td>1000BASE-T</td>
<td>UTP (balanced 100 Ohm, Cat. 5E)</td>
<td>4 pairs</td>
<td>100m</td>
<td>PAM</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)
  - A signal of 5 different levels is transmitted over the 4 pairs
    - $5^4 = 625$ possible symbols, of which only 256 are valid
    - Each pair transports 2 bits $\rightarrow$ 125 Mbaud (250Mbps) per pair
  - Redundancy used for control codes

- Cat 5 UTP has to pass additional tests compared to the ones defined by TIA/EIA ISB95,
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

Visible Light

Wave-Length and standard

Attenuation (dB/km)

Wave-Length (nm)

I Window 850nm

II Window 1310nm

III Window 1550nm

1000BASE-SX

1000BASE-LX
1000BASE-LX & multimode fiber: Mode Conditioning Patch Cord

Equipment 1000BASE-LX port

SC connectors
BEIGE Color

SC Connectors
BLU Color

Junction between SMF and MMF fiber

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 better than traditional solutions (e.g. SONET/SDH)
WAN PHY

- Enables transport over existent MAN and WAN infrastructure
  - DWDM (Dense Wavelength Division Multiplexing)
- Enables existent MAN and WAN component reuse
  - SONET/SDH transceivers 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
10GE and SONET/SDH

- Simplified version of SONET/SDH
  - Avoid imposed complexities required by SONET/SDH
  - Limit component cost
  - Keeps resiliency (SONET or DWDM rings)
- 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 wiring)</td>
<td>64B/66B</td>
</tr>
<tr>
<td></td>
<td>Multimode 50 µm</td>
<td>66 – 300 m</td>
<td></td>
<td></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 wiring)</td>
<td>FC 10G: 8B10B</td>
</tr>
<tr>
<td></td>
<td>Multimode 50 µm</td>
<td>240 – 300 m</td>
<td></td>
<td></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 wiring)</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></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

- Copper
- 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 lanes
  - 3.125 Gbaud per lane
- Redundancy used for control codes
  - For example idle signal act as inter-frame gap
10GBASE-LX4

- Fiber

WD = Wavelength Division
Conclusions

- Ethernet is the facto standard for cabled LANs
  - Widely diffused on MAN as well
  - Progressively used on WAN
- Probably will replace also other standards
  - E.g. Fibre Channel, Infiniband
  - New features (“lossless Ethernet”) being added
- CSMA/CD no longer present
  - Its influence (e.g. min frame) still present nowadays
  - Everything “switched”, Full Duplex
- Framing ins basically what remains from Ethernet DIX
- Higher speeds (40Gbps, 100Gbps) in the pipeline
  - Datacenter