Ethernet

Chapter 12 in Stallings 10th Edition
Figure 12.1  Data Center Study—Percentage of Ethernet Links by Speed
Early on...

Earliest was ALOHA

- Developed for packet radio networks
- Station may transmit a frame at any time
- If frame is determined invalid, it is ignored
- Maximum utilization of channel about 18%

Next came slotted ALOHA

- Organized slots equal to transmission time
- Increased utilization to about 37%
ALOHA

- Packet Radio
- When station has frame, it sends it!
- Station listens (for max round trip time) plus small increment
- If ACK, fine. If not, retransmit
- If no ACK after repeated transmissions, give up.
- Frame check sequence (as in HDLC)
- If frame OK and address matches receiver, send ACK
- Frame may be damaged by noise or by another station transmitting at the same time (collision)
- Any overlap of frames causes collision
- Max utilization 18%
Slotted ALOHA

- Time in uniform slots equal to frame transmission time
- Need central clock (or other sync mechanism)
- Transmission begins at slot boundary
- Frames either miss or overlap totally
- Max utilization 37%
CSMA/CD Precursors

• Carrier Sense Multiple Access (CSMA)
  — Station listens to determine if there is another transmission in progress
  — If idle, station transmits
  — Waits for acknowledgment
  — If no acknowledgment, collision is assumed and station retransmits
  — Utilization far exceeds ALOHA
Figure 12.2 CSMA Persistence and Backoff

Channel Busy

Ready

1-Persistent:
• Transmit as soon as channel goes idle
• If collision, back off

P-Persistent:
• Transmit as soon as channel goes idle with probability P
• Otherwise, delay one time slot and repeat process
• If collision, back off

Nonpersistent:
• Transmit if idle
• If busy, wait random time and repeat process
• If collision, back off

Constant or variable delay
Nonpersistent CSMA

If the medium is idle, transmit; otherwise, go to step 2

Disadvantage:
Capacity is wasted because the medium will generally remain idle following the end of a transmission even if there are one or more stations waiting to transmit

If the medium is busy, wait an amount of time drawn from a probability distribution and repeat step 1
1-Persistent CSMA

- Avoids idle channel time
- Rules:
  1. If medium is idle, transmit
  2. If medium is busy, listen until idle; then transmit immediately
- Stations are selfish
- If two or more stations are waiting, a collision is guaranteed
P-Persistent CSMA

- A compromise to try and reduce collisions and idle time
- P-persistent CSMA rules:
  1. If medium is idle, transmit with probability $p$, and delay one time unit with probability $(1-p)$
  2. If medium is busy, listen until idle and repeat step 1
  3. If transmission is delayed one time unit, repeat step 1
- Issue of choosing effective value of $p$ to avoid instability under heavy load
Value of $p$?

- Have $n$ stations waiting to send
- At end of transmission, expected number of stations is $np$
  - If $np > 1$ on average there will be a collision
- Repeated transmission attempts mean collisions are likely
- Eventually all stations will be trying to send, causing continuous collisions, with throughput dropping to zero
- To avoid catastrophe $np < 1$ for expected peaks of $n$
  - If heavy load expected, $p$ must be small
  - Smaller $p$ means stations wait longer
Description of CSMA/CD

1. If the medium is idle, transmit; otherwise, go to step 2

2. If the medium is busy, continue to listen until the channel is idle, then transmit immediately

3. If a collision is detected, transmit a brief jamming signal to assure that all stations know that there has been a collision and cease transmission

4. After transmitting the jamming signal, wait a random amount of time, referred to as the backoff, then attempt to transmit again
Figure 12.3 CSMA/CD Operation
Which Persistence Algorithm?

- IEEE 802.3 uses 1-persistent
- Both nonpersistent and p-persistent have performance problems

1-persistent seems more unstable than p-persistent

- Because of greed of the stations
- Wasted time due to collisions is short
- With random backoff unlikely to collide on next attempt to send
Binary Exponential Backoff

- IEEE 802.3 and Ethernet both use binary exponential backoff
- A station will attempt to transmit repeatedly in the face of repeated collisions
  - On first 10 attempts, mean random delay doubled
  - Value then remains the same for 6 further attempts
  - After 16 unsuccessful attempts, station gives up and reports error
- 1-persistent algorithm with binary exponential backoff is efficient over wide range of loads
- Backoff algorithm has last-in, first-out effect
Collision Detection

On baseband bus

- Collision produces higher signal voltage
- Collision detected if cable signal is greater than single station signal
- Signal is attenuated over distance
- Limit to 500m (10Base5) or 200m (10Base2)

On twisted pair (star-topology)

- Activity on more than one port is collision
- Use special collision presence signal
IEEE 802.3
Frame Format

<table>
<thead>
<tr>
<th>Layer</th>
<th>Octets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preamble</td>
<td>7</td>
</tr>
<tr>
<td>Start of Frame Delimiter</td>
<td>1</td>
</tr>
<tr>
<td>Destination Address</td>
<td>6</td>
</tr>
<tr>
<td>Source Address</td>
<td>6</td>
</tr>
<tr>
<td>Length/Type</td>
<td>2</td>
</tr>
<tr>
<td>MAC Client Data</td>
<td>46-1500 or 1504 or 1982</td>
</tr>
<tr>
<td>Pad</td>
<td>4</td>
</tr>
<tr>
<td>Frame Check Sequence</td>
<td>4</td>
</tr>
<tr>
<td>Extension</td>
<td>4</td>
</tr>
</tbody>
</table>

**Figure 12.4 IEEE 802.3 MAC Frame Format**
# Table 12.1

IEEE 802.3 10-Mbps Physical Layer Medium Alternatives

<table>
<thead>
<tr>
<th>Transmission medium</th>
<th>10BASE5</th>
<th>10BASE2</th>
<th>10BASE-T</th>
<th>10BASE-FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coaxial cable (50 ohm)</td>
<td>Coaxial cable (50 ohm)</td>
<td>Unshielded twisted pair</td>
<td>850-nm optical fiber pair</td>
<td></td>
</tr>
<tr>
<td>Signaling technique</td>
<td>Baseband (Manchester)</td>
<td>Baseband (Manchester)</td>
<td>Baseband (Manchester)</td>
<td>Manchester/on-off</td>
</tr>
<tr>
<td>Topology</td>
<td>Bus</td>
<td>Bus</td>
<td>Star</td>
<td>Star</td>
</tr>
<tr>
<td>Maximum segment length (m)</td>
<td>500</td>
<td>185</td>
<td>100</td>
<td>500</td>
</tr>
<tr>
<td>Nodes per segment</td>
<td>100</td>
<td>30</td>
<td>—</td>
<td>33</td>
</tr>
<tr>
<td>Cable diameter (mm)</td>
<td>10</td>
<td>5</td>
<td>0.4 to 0.6</td>
<td>62.5/125 µm</td>
</tr>
</tbody>
</table>
# Table 12.2

## IEEE 802.3 100BASE-T Physical Layer Medium Alternatives

<table>
<thead>
<tr>
<th></th>
<th>100BASE-TX</th>
<th>100BASE-FX</th>
<th>100BASE-T4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transmission medium</strong></td>
<td>2 pair, STP</td>
<td>2 pair, Category 5 UTP</td>
<td>2 optical fibers</td>
</tr>
<tr>
<td><strong>Signaling technique</strong></td>
<td>MLT-3</td>
<td>MLT-3</td>
<td>4B5B, NRZI</td>
</tr>
<tr>
<td><strong>Data rate</strong></td>
<td>100 Mbps</td>
<td>100 Mbps</td>
<td>100 Mbps</td>
</tr>
<tr>
<td><strong>Maximum segment length</strong></td>
<td>100 m</td>
<td>100 m</td>
<td>100 m</td>
</tr>
<tr>
<td><strong>Network span</strong></td>
<td>200 m</td>
<td>200 m</td>
<td>400 m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200 m</td>
</tr>
</tbody>
</table>
100BASE-X

- Uses a unidirectional data rate 100 Mbps over single twisted pair or optical fiber link
- Encoding scheme same as FDDI — 4B/5B-NRZI

**Two physical medium specifications**

- **100BASE-TX**
  - Uses two pairs of twisted-pair cable
  - STP and Category 5 UTP allowed
  - MTL-3 signaling scheme is used

- **100BASE-FX**
  - Uses two optical fiber cables
  - Convert 4B/5B-NRZI code group into optical signals
100BASE-X Data Rate and Encoding

- Unidirectional data rate 100 Mbps over single link
  - Single twisted pair, single optical fiber
- Encoding scheme same as FDDI
  - 4B/5B-NRZI
  - Modified for each option
100BASE-X Media

• Two physical medium specifications
• 100BASE-TX
  — Two pairs of twisted-pair cable
  — One pair for transmission and one for reception
  — STP and Category 5 UTP allowed
  — The MTL-3 signaling scheme is used
• 100BASE-FX
  — Two optical fiber cables
  — One for transmission and one for reception
  — Intensity modulation used to convert 4B/5B-NRZI code group stream into optical signals
  — 1 represented by pulse of light
  — 0 by either absence of pulse or very low intensity pulse
100BASE-T4

- 100-Mbps over lower-quality Cat 3 UTP
  - Takes advantage of large installed base
  - Does not transmit continuous signal between packets
  - Useful in battery-powered applications
- Can not get 100 Mbps on single twisted pair
  - So data stream split into three separate streams
  - Four twisted pairs used
  - Data transmitted and received using three pairs
  - Two pairs configured for bidirectional transmission
- Use ternary signaling scheme (8B6T)
100BASE-T Options

- 100BASE-T
  - 100BASE-X
    - 100BASE-TX
      - 2 Category 5 UTP
      - 2 STP
    - 100BASE-FX
      - 2 Optical Fiber
    - 100BASE-T4
      - 4 Category 3 or Category 5 UTP
Full Duplex Operation

- Traditional Ethernet half duplex
- Using full-duplex, station can transmit and receive simultaneously
- 100-Mbps Ethernet in full-duplex mode, giving a theoretical transfer rate of 200 Mbps
- Stations must have full-duplex adapter cards
- And must use switching hub
  - Each station constitutes separate collision domain
  - CSMA/CD algorithm no longer needed
  - 802.3 MAC frame format used
Mixed Configurations

• Fast Ethernet supports mixture of existing 10-Mbps LANs and newer 100-Mbps LANs

• E.g. 100-Mbps backbone LAN to support 10-Mbps hubs
  — Stations attach to 10-Mbps hubs using 10BASE-T
  — Hubs connected to switching hubs using 100BASE-T
    • Support 10-Mbps and 100-Mbps
  — High-capacity workstations and servers attach directly to 10/100 switches
  — Switches connected to 100-Mbps hubs using 100-Mbps links
  — 100-Mbps hubs provide building backbone
    • Connected to router providing connection to WAN
Gigabit Ethernet - Differences

- Carrier extension
  - At least 4096 bit-times long (512 for 10/100)
- Frame bursting
  - allow multiple frames to be transmitted consecutively, i.e. “together”
- Not needed if using a switched hub to provide dedicated media access
Gigabit Ethernet – Physical

- 1000Base-SX
  - Short wavelength, multimode fiber
- 1000Base-LX
  - Long wavelength, Multi or single mode fiber
- 1000Base-CX
  - Copper jumpers <25m, shielded twisted pair
- 1000Base-T
  - 4 pairs, cat 5 UTP
Figure 12.5  Gigabit Ethernet Medium Options (log scale)
10Gbps Ethernet

- Growing interest in 10Gbps Ethernet
  - High-speed backbone use
  - Future wider deployment
- Alternative to ATM and other WAN technologies
- Uniform technology for LAN, MAN, or WAN
- Advantages of 10Gbps Ethernet
  - No expensive, bandwidth-consuming conversion between Ethernet packets and ATM cells
  - IP and Ethernet together offers QoS and traffic policing approach ATM
  - Have a variety of standard optical interfaces
Figure 12.6  Example 10 Gigabit Ethernet Configuration
10Gbps Ethernet Distance Options (log scale)

- **10GBASE-S** (850 nm)
  - 50-µm multimode fiber
  - 62.5-µm multimode fiber

- **10GBASE-L** (1310 nm)
  - Single-mode fiber

- **10GBASE-E** (1550 nm)
  - Single-mode fiber

- **10GBASE-LX4** (1310 nm)
  - 50-µm multimode fiber
  - 62.5-µm multimode fiber

---

**Figure 12.7** 10-Gbps Ethernet Distance Options (log scale)
100-Gbps Ethernet (100GbE)

- Preferred technology for wired LAN
- Preferred carrier for bridging wireless technologies into local Ethernet networks
- Cost-effective, reliable and interoperable
- Popularity of Ethernet technology:
  - Availability of cost-effective products
  - Reliable and interoperable network products
  - Variety of vendors
Figure 12.8 Example 100-Gbps Ethernet Configuration for Massive Blade Server Site
Multilane Distribution

used to achieve the required data rates

- **Multilane distribution:**
  - Switches implemented as multiple parallel channels
    - Separate physical wires

- **Virtual lanes:**
  - If a different number of lanes are actually in use, virtual lanes are distributed into physical lanes in the PMD (physical medium dependent) sublayer
  - Form of inverse multiplexing
Figure 12.9 Multilane Distribution for 100-Gbps Ethernet

(a) Virtual lane concept

(b) Alignment block

1 0 | Frm1 | Frm2 | reserved | reserved | reserved | reserved | ~VL# | VL#
# Media Options for 40-Gbps and 100-Gbps Ethernet

## Table 12.3

<table>
<thead>
<tr>
<th></th>
<th>40 Gbps</th>
<th>100 Gbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>1m backplane</td>
<td>40GBASE-KR4</td>
<td></td>
</tr>
<tr>
<td>10 m copper</td>
<td>40GBASE-CR4</td>
<td>1000GBASE-CR10</td>
</tr>
<tr>
<td>100 m multimode fiber</td>
<td>40GBASE-SR4</td>
<td>1000GBASE-SR10</td>
</tr>
<tr>
<td>10 km single mode fiber</td>
<td>40GBASE-LR4</td>
<td>1000GBASE-LR4</td>
</tr>
<tr>
<td>40 km single mode fiber</td>
<td></td>
<td>1000GBASE-ER4</td>
</tr>
</tbody>
</table>

**Naming nomenclature:**
- Copper: K = backplane; C = cable assembly
- Optical: S = short reach (100m); L = long reach (10 km); E = extended long reach (40 km)
- Coding scheme: R = 64B/66B block coding
- Final number: number of lanes (copper wires or fiber wavelengths)
Figure 12.10  Tagged IEEE 802.3 MAC Frame Format

CFI = Canonical Format Indicator
VLAN = virtual local area network
Summary

• Traditional Ethernet
  — IEEE 802.3 medium access control
  — IEEE 802.3 10-Mbps specifications (Ethernet)

• IEEE 802.1Q VLAN standard

• High-speed Ethernet
  — IEEE 802.3 100-Mbps specifications (Fast Ethernet)
  — Gigabit Ethernet
  — 10-Gbps Ethernet
  — 100-Gbps Ethernet