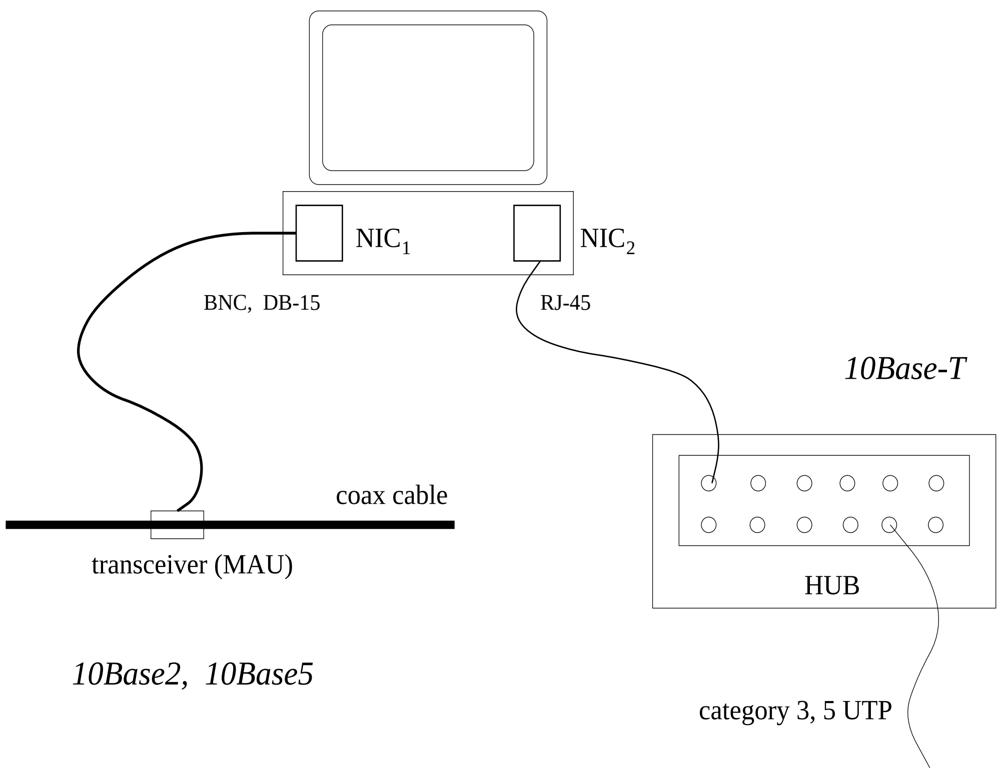
Ethernet and CSMA/CD

⟶ copper, fiber

Types:

- 10Base2 (ThinNet): coax, segment length 200 m, 30 nodes/segment
- 10Base5 (ThickNet): coax, segment length 500 m, 100 nodes/segment
- 10Base-T: twisted pair, segment length 100 m, 1024 nodes/segment
- 100Base-T (Fast Ethernet): category 5 UTP, fiber (also 100VG-AnyLAN)
- Gigabit & 10 Gbps Ethernet: fiber, category 5 UTP

Connectivity example:



- single-homed vs. multi-homed
- unique Ethernet address per NIC
- physical network: bus vs. hub vs. switch

  → very old vs. old vs. not-so-old

⟶ hub: multi-tap junction

⟶ bus and hub: logically equivalent

Wire segments can be hooked up by repeaters, bridges, hubs or switches.

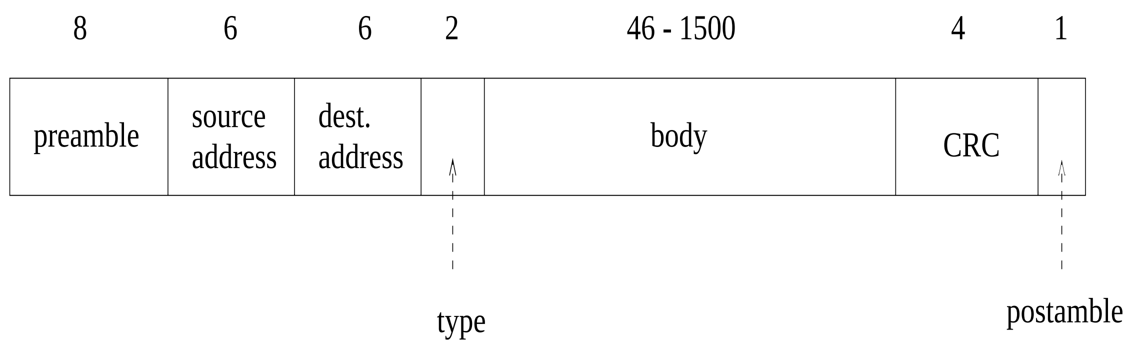
- maximum of 2 (4 for IEEE 802.3) repeaters between two hosts; 1500 m
- for Fast Ethernet, 2 repeater hops

High-speed Ethernets have shorter network diameter

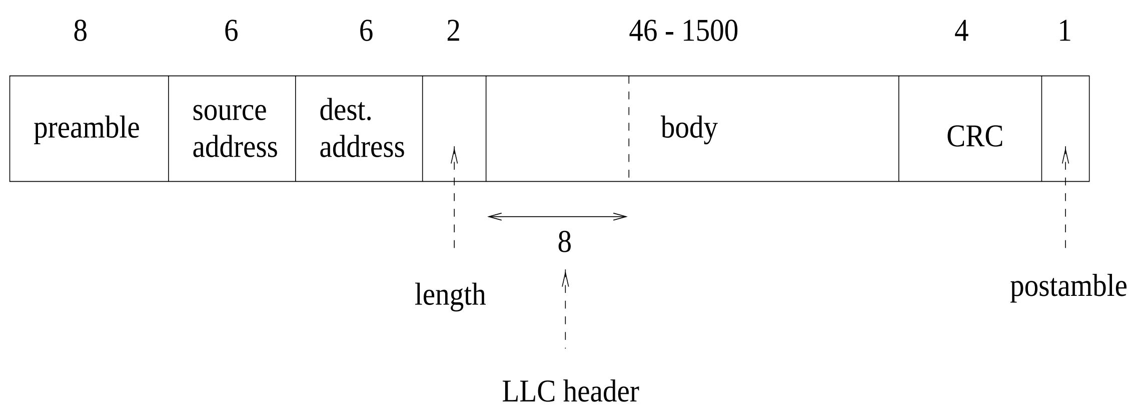
- about 2500 m for 10 Mbps Ethernet
- about 200 m for 100 Mbps Ethernet
- even shorter for 1 Gbps Ethernet

  → additional complications for medium-haul

DIX Ethernet frame:



IEEE 802.3 Ethernet frame:



$\longrightarrow$ IEEE 802.2 LLC (Logical Link Control)

$\longrightarrow$ common interface to different link protocols

Encoding: Manchester

$\longrightarrow$ recall: Ethernet is baseband

Addressing:

- 48 bit unique address
- point-to-point
- broadcast (all 1's)

Receiver: Ethernet adaptor accepts frames with "relevant" address.

- accepts only own frame address
- accepts all frames: promiscuous mode

  $\rightarrow$ NIC feature

  $\rightarrow$ sniffing

CSMA/CD MAC:

- CS (Carrier Sense): can detect if some other node is using the link

  → rule: if busy, abstein

- MA (Multiple Access): multiple nodes are allowed simultaneous access

  → rule: if channel seems silent, send

- CD (Collision Detection): can detect if collision due to simultaneous access has occured

  → rule: if collision, retry later

Wired vs. wireless media:

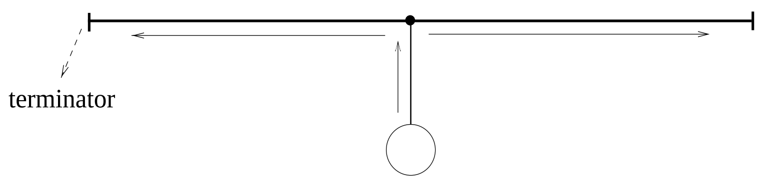
⟶ CD is key difference

⟶ diffcult to detect collision while transmitting

Signal propagation and collision:

Bi-directional propagation

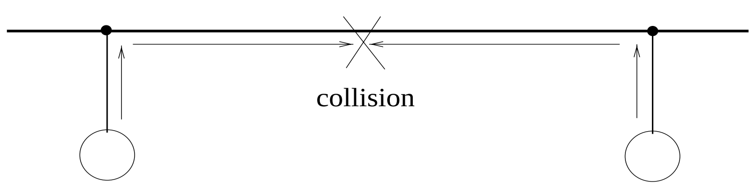
$\longrightarrow$ terminator absorbs signal: prevent bounce back



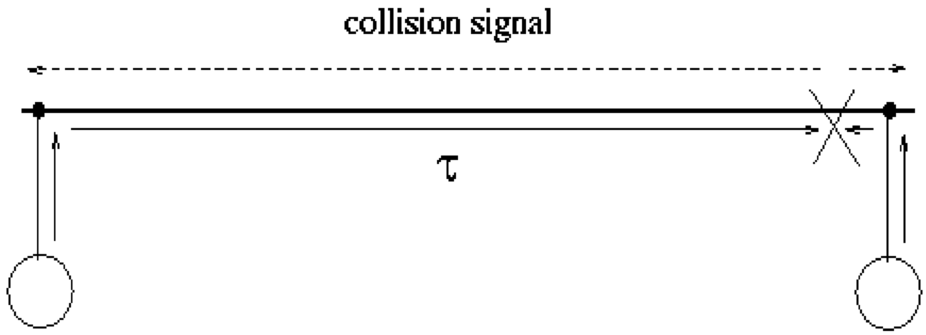
Best-case collision: 2 stations

$\longrightarrow$ meet in the middle

$\longrightarrow$ worst-case?



Worst-case collision scenario:



$\longrightarrow$ τ: one-way propagation delay

- sender needs to wait 2τ sec before detecting collision
- for 2500 m length, 51.2 μs round-trip time (2τ)
  $\rightarrow$ fact
- enforce 51.2 μs slot time
- at 10 Mbps, 512 bits; i.e., minimum frame size
  $\rightarrow$ assures collision detection

Transmit at least 512 bits

$\longrightarrow$ $6 + 6 + 2 + 46 + 4 = 64$ B $= 512$ bits

$\longrightarrow$ note: delay-bandwidth product

Retry upon collision: exponential backoff

1. Wait for random $0 \leq X \leq 51.2$ μs before first retry
2. On i'th collision, wait for $0 \leq X \leq 2^{i-1}\, 51.2$ μs before next attempt
3. Give up if $i > 16$

$\longrightarrow$ a form of stop-and-wait

$\longrightarrow$ what's the ACK?

$\longrightarrow$ guaranteed reliability?

$\longrightarrow$ pretty drastic measure: necessary?

CSMA/CD Throughput

$\longrightarrow$ approximate analysis in simplified setting

Assumptions:

- time is slotted

  $\rightarrow$ slot duration: 2τ

- k hosts; each host transmits with probability p at every slot

  $\rightarrow$ transmission behavior among hosts independent

  $\rightarrow$ transmission behavior across slots independent

New performance metric: utilization (ϱ)

$\longrightarrow$ fraction of total bandwidth attained

$\longrightarrow$ $0 \leq \varrho \leq 1$

$\longrightarrow$ captures efficiency and wastage

In slotted CSMA/CD:

$\longrightarrow$ fraction of usefully used slots

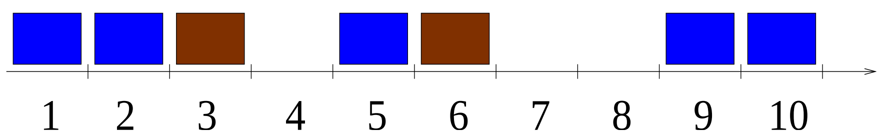
$\longrightarrow$ what are "uselessly used" slots?

Ex.: snapshot of baseband channel over 10 time slots

→ blue: successfully transmitted frames

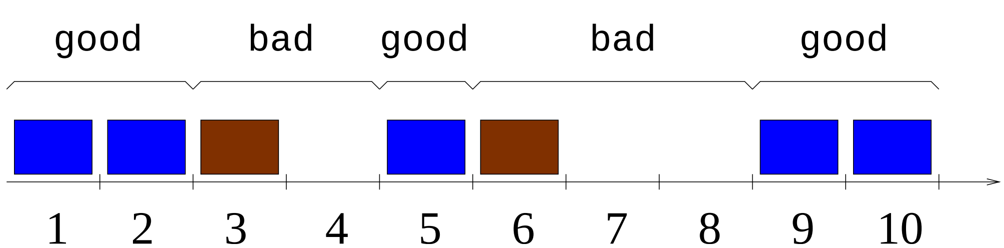
→ brown: collided frames

→ utilization ϱ?



One more viewpoint:

⟶ note: useful and useless "periods" alternate



In the long run,

$$\varrho = \frac{E[\text{good}]}{E[\text{good}] + E[\text{bad}]}$$

$\longrightarrow$ avrg. length of adjacent “good” and “bad” periods

$\longrightarrow$ formula holds under mild conditions

Next: calculate $E[\text{good}]$ and $E[\text{bad}]$

Fix time slot. Probability that a fixed host acquires the slot successfully

$$p(1-p)^{k-1}$$

Probability that some host acquires the slot

$$\eta = kp(1-p)^{k-1}$$

$\longrightarrow$ why?

Now, let's be generous and find p that maximizes η

$\longrightarrow$ upper bounding

Fact: η is maximized at $p = 1/k$. Also,

$$\lim_{k\to\infty} \eta = \lim_{k\to\infty} \left(1 - \frac{1}{k}\right)^{k-1} = 1/e.$$

$\longrightarrow$ many user assumption

$\longrightarrow$ common practice to simplify expression (valid?)

Probability bad period persists for exactly i slots

$$(1-\eta)^{i-1}\eta$$

Thefore average bad period

$$E[\text{bad}] = \sum_{i=0}^{\infty} i(1-\eta)^{i-1}\eta = 1/\eta$$

$E[\text{bad}]$ is in unit of slots. Convert to second:

$$2\tau/\eta = 2\tau e$$

Similarly calculate $E[\text{good}]$; call it γ.

Convert γ to second:

$$\gamma F/B$$

where

F: frame size (bits)

B: bandwidth (bps)

Putting everything together

$$\begin{aligned}\varrho &= \frac{E[\text{good}]}{E[\text{good}] + E[\text{bad}]} \\ &= \frac{\gamma F/B}{\gamma F/B + 2\tau e} \\ &= \frac{\gamma F/B}{\gamma F/B + 2Le/c} \\ &= \frac{1}{1 + (2e/c\gamma)BL/F}\end{aligned}$$

where

L: length of wire (meters)

c: speed of light (m/s)

What does the formula say?

For example, if B is increased, what must be done to maintain high utilization?

In practice today: switched Ethernet

- contention moved from bus to "single point"

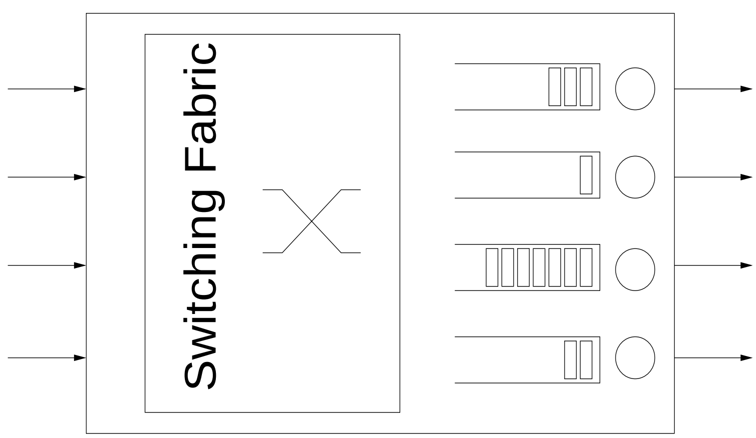
  → switch: star topology

  → analogous to old telephone switch-boards

- Ethernet frames are logically scheduled

  → includes buffering

Diagram of output-buffered switch:



⟶ interconnection networks (e.g., shuffle-exchange)

⟶ switching fabric: hardware

- Ethernet switch emulates CSMA/CD

  → backward compatibility

  → use same frame format

- upon buffer overflow: send collision signal

  → transparent to legacy host NIC

  → awkward: instituted for incremental deployment

  → Internet: new technology must respect legacy

Ex.: 10Base-T, 100Base-T, 1000Base-T and 1000Base-X

⟶ FE: 802.3u; GigE: 802.3ab and 802.3z

⟶ negotiation: e.g., full/half duplex

⟶ how can GigE overcome length limitation?

⟶ e.g., supports 200 m as in FE

Slot time extension:

- frame format remains the same
- minimum slot time extended from 64 B to 512 B

  → padding: transparent to legacy CSMA/CA

  → also called carrier extension

  → reconciliation sublayer between MAC and PHY

Packet bursting:

- slot time extension alone problematic

  → small frames: marginal increase in throughput
- allow burst of packets

  → only first packet is padded & burst limit

Longer distances?

$\longrightarrow$ e.g., 1000Base-LX

Medium-haul GigE/10GigE (802.3ae): 500m, 5km, 40km

- CSMA/CD disabled

  → purely point-to-point link

  → switch-to-switch

  → simpler

  → backward compatibility: not an issue

- flow control

  → pause frame to prevent buffer overflow

QoS: 802.3p

$\longrightarrow$ frame tagging conveys priority

$\longrightarrow$ priority classes supported at switches

FDDI (Fiber Distributed Data Interface)

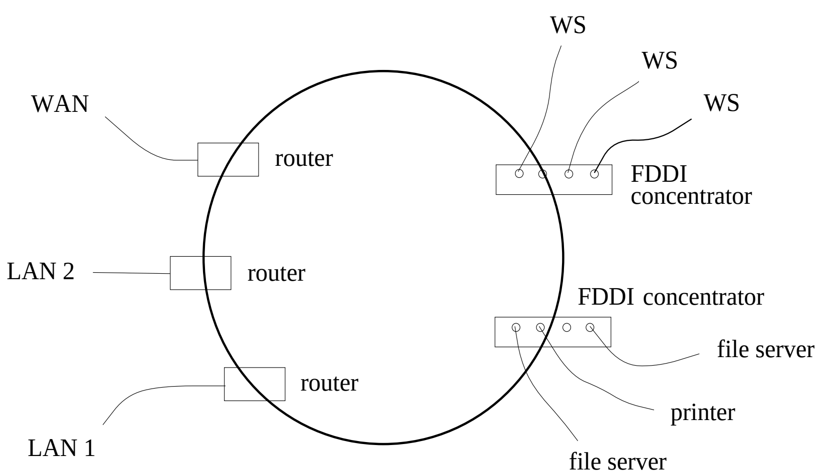
⟶ token ring architecture

High-bandwidth extension of IBM 4 Mbps token ring and 16 Mbps IEEE 802.5 token ring standard.

⟶ 100 Mbps bandwidth

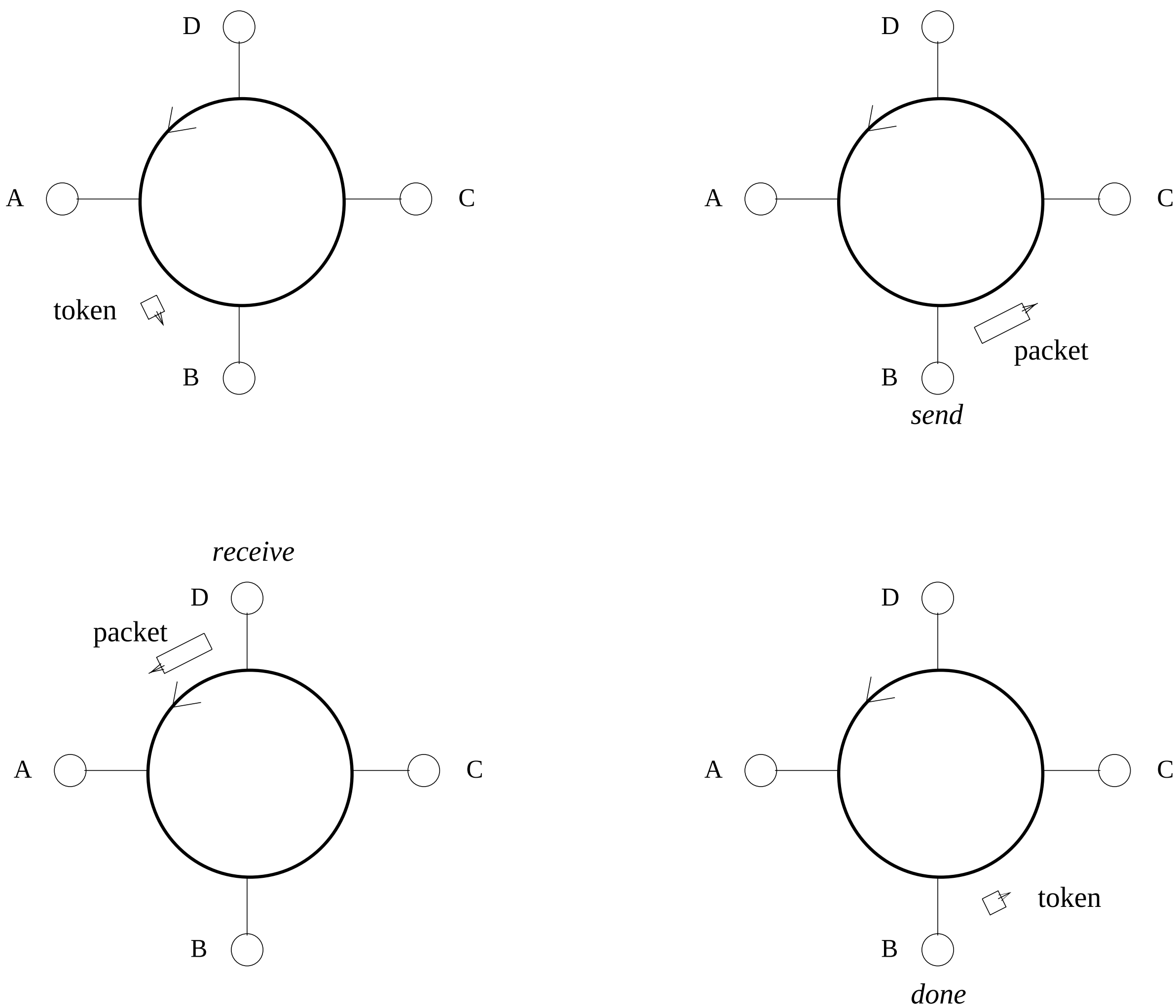
Used as high-bandwidth campus/city backbone.

⟶ metropolitan/campus distance: MAN

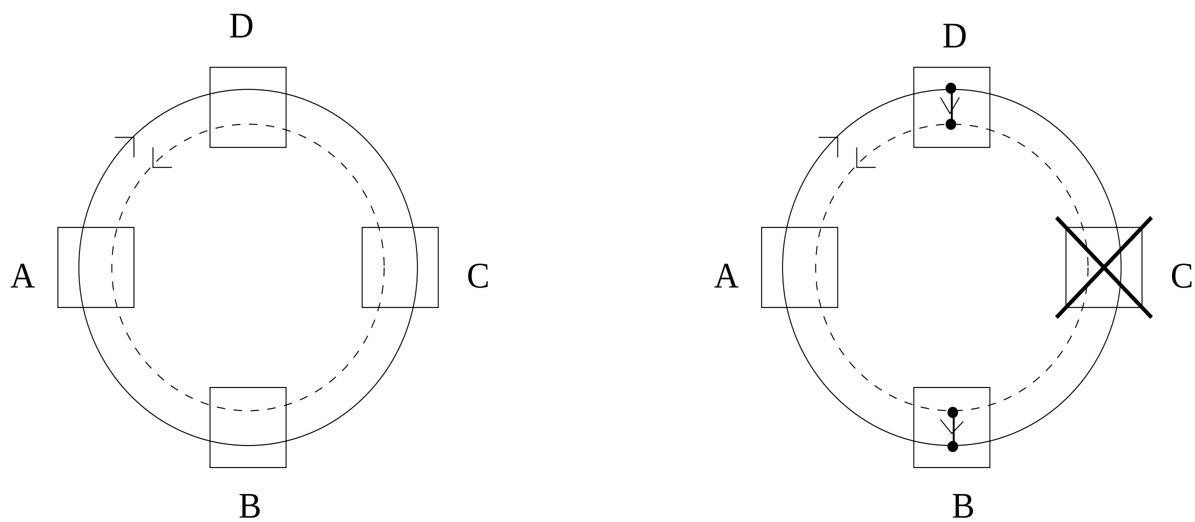


Basic operation:

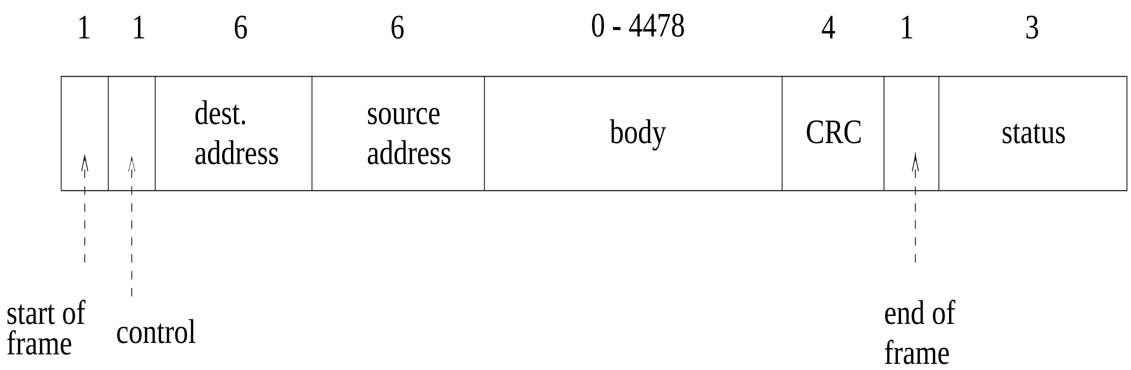
$\longrightarrow$ B wants to send to D



Fault-tolerance:



- DAS (dual attachment station)
- SAS (single attachment station)



- frame size < 4500 B
- 4B/5B encoding
- synchronous/asynchronous data
- 2 km inter-station distance
- 200 km diameter (multimode fiber); 100 km circumference

Performance issues: fairness and efficiency

- TRT (token rotation time)
- THT (token holding time)

$$\text{TRT} = \text{no. of nodes} \times \text{THT} + \text{link latency}$$

To increase efficiency: increase THT

$\longrightarrow$ let station send as much as it needs

$\longrightarrow$ same as frame size $\uparrow$

$\longrightarrow$ THT $\uparrow \implies \rho \uparrow$

To increase fairness: limit THT

$\longrightarrow$ limit station's one-time sending of data

To facilitate fairness: introduce TTRT (target token rotation time).

THT determining factor (assume TTRT is given):

- prioritized frames: synchronous/asynchronous
- Synchronous frames always get sent.
- If TRT $>$ TTRT, then late; don't send asynchronous data.
- If TRT $\leq$ TTRT, then early; send asynchronous data for max $\{$ TTRT $-$ TRT, single frame time $\}$ duration.

How to set TTRT?

$\longrightarrow$ token claim process

$\longrightarrow$ initiate when needed (e.g., start-up)

- Each station submits claim frame containing TTRT bid.
- Smaller TTRT bid overrides higher TTRT bids.
  - Compare claim frame bid against own desired TTRT.
  - If less, then reset own TTRT to lower value.
  - If larger, then put lower bid on claim frame and forward.
- Winner: same bid value when claim frame makes full circle.

$\longrightarrow$ leader election

At the end of the day, consistent TTRT value among all stations.

$\longrightarrow$ consensus problem

Compare against Ethernet's CSMA/CD.

$\longrightarrow$ round-robin reservation

$\longrightarrow$ absence of MA and collision

$\longrightarrow$ determinism vs. indeterminism

$\longrightarrow$ imperfect QoS assurance

$\longrightarrow$ performance vis-à-vis CSMA/CD?

Cooperative vs. noncooperative protocols

$\longrightarrow$ robust if some users use selfish MAC

$\longrightarrow$ could be malicious