Lecture 7
Datalink – Ethernet, Home

David Andersen
School of Computer Science
Carnegie Mellon University

15-441 Networking, Spring 2005
http://www.cs.cmu.edu/~srini/15-441/S05

1

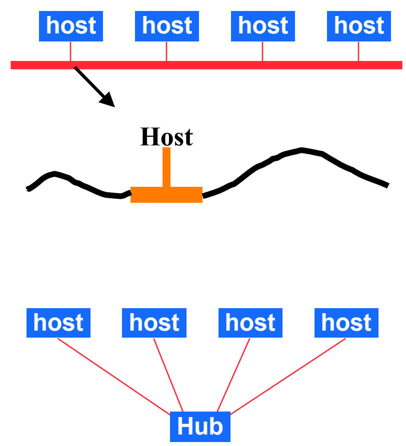
Ethernet Physical Refresher

- 10Mhz signal (baseband modulation)
- Nyquist limit says we should be able to get 20 Mbits/s from that.
- Manchester encoding solves runs of 1s and 0s problem
  - » Wastes 1/2 of the possible data rate - back to 10Mbps

2

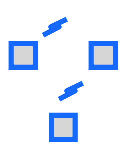
Ethernet Physical Layer

- **10Base2 standard based on thin coax.**
  - » **Thick coax no longer used**
  - » **Nodes are connected using thin coax cables and "T" connectors in a bus topology**
- **10-BaseT uses twisted pair and hubs.**
  - » **Hub acts as a concentrator**
- **The two designs have the same protocol properties.**
  - » **Key: electrical connectivity between all nodes**
  - » **Deployment is different**

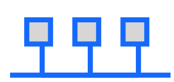


3

Ethernet over Time



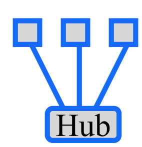
Aloha
packet radio



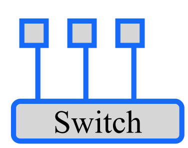
Ethernet on coax

10base-2 (thinnet)

10base-5 (thicknet)



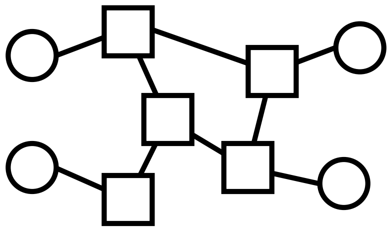
?baseT with hub
(twisted pair)

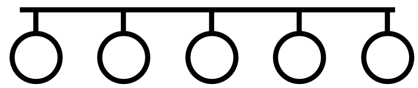


?baseT with switch
(point to point links)

4

Datalink Layer Architectures





- Packet forwarding.
- Error and flow control.

- Media access control.
- Scalability.

5

Multiple Access Protocols

- Prevent two or more nodes from transmitting at the same time over a broadcast channel.
  - If they do, we have a collision, and receivers will not be able to interpret the signal
- Several classes of multiple access protocols.
  - Partitioning the channel, e.g. frequency-division or time division multiplexing
    - With fixed partitioning of bandwidth – not flexible
  - Taking turns, e.g. token-based, reservation-based protocols, polling based
  - Contention based protocols, e.g. Aloha, Ethernet

6

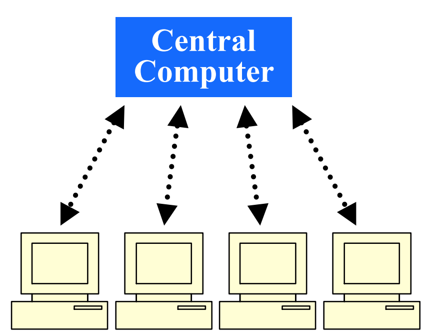
Contention-Based Protocol

- Goal: share the communication channel among multiple hosts sharing it.
- Problem: how to arbitrate between the connected hosts.
- Desired properties:
  - » High bandwidth utilization
  - » Avoid starvation, achieve fairness
  - » Simple solution
- Idea: access the channel in a random way - when collisions occur, recover.
  - » Collision: two or more nodes transmitting at the same time

7

Aloha

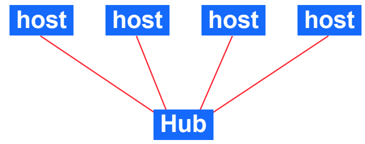
- Node sends the message when it has data to send.
- If it receives an ack, it considers the transmission completed, otherwise it retransmits after a random delay.
- Simple, distributed protocol, but not very efficient
  - » 18% maximum utilization
- Slotted Aloha: more efficient.
  - » Transmit only in specific time slot
  - » Reduces chances of collision
  - » 37% maximum utilization



8

802.3 Ethernet

Broadcast technology

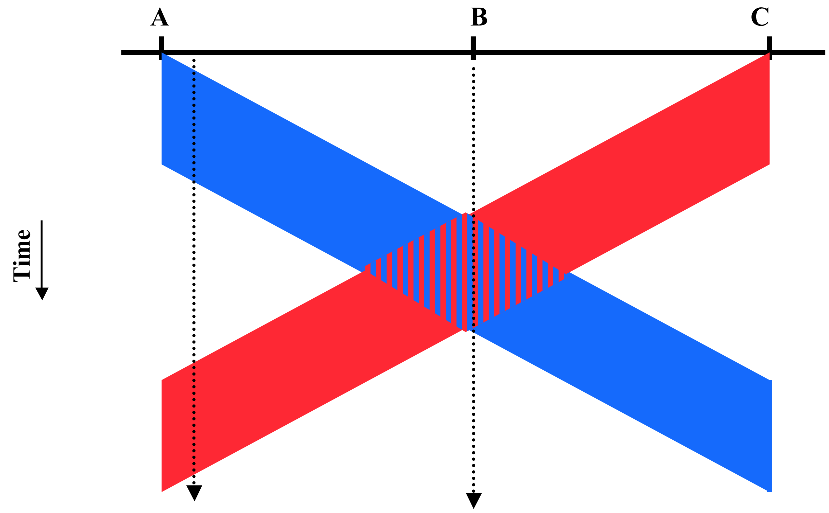


- Carrier-sense multiple access with collision detection (CSMA/CD).
  - » MA = multiple access
  - » CS = carrier sense
  - » CD = collision detection
- Base Ethernet standard is 10 Mbs.
  - » Original design was ~2 Mbs
  - » Faster versions discussed later

CSMA/CD Algorithm

- Sense for carrier.
- If carrier present, wait until carrier ends.
  - » Sending would force a collision and waste time
- Send packet and sense for collision.
- If no collision detected, consider packet delivered.
- Otherwise, abort immediately, perform “exponential back off” and send packet again.
  - » Start to send at a random time picked from an interval
  - » Length of the interval increases with every retransmission

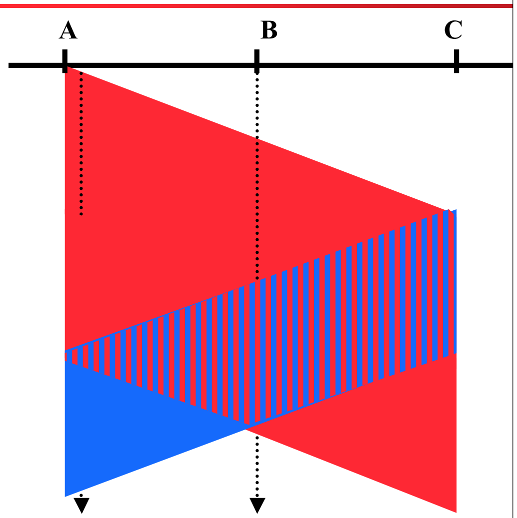
Collision Detection



11

Collision Detection: Implications

- All nodes must be able to detect the collision.
  - » Any node can be sender
- => Must either have short wires, long packets, or both.
- Can calculate length/distance based on transmission rate and propagation speed.
  - » Messy: propagation speed is media-dependent, low-level protocol details, ..
  - » Minimum packet size is 64 bytes
    - – Cable length ~256 bit times
  - » Example: maximum coax cable length is 2.5 km



12

Minimum Packet Size

- Give a host enough time to detect a collision.
- In Ethernet, the minimum packet size is 64 bytes.
  - » 18 bytes of header and 46 data bytes
  - » If the host has less than 46 bytes to send, the adaptor (pads) bytes to increase the length to 46 bytes
- What is the relationship between the minimum packet size and the size of LAN?

LAN = (min frame size) * light speed / (2 * bandwidth)

- How did they pick the minimum packet size?

13

CSMA/CD: Some Details

- Successive frames are separated by an “inter-frame” gap.
  - » Nodes must switch from “send” to “receive” mode
  - » Set to 9.6 μsec or 96 bit times
- When a sender detects a collision, it sends a “jam signal”.
  - » Make sure that all nodes are aware of the collision
  - » Length of the jam signal is 32 bit times
  - » Permits early abort - don’t waste max transmission time
- Exponential backoff operates in multiples of 512 bit times.
  - » Longer than a roundtrip time
  - » Guarantees that nodes that back off longer will notice the earlier retransmission before starting to send

14

Ethernet Frame Format

8	6	6	2			4
Preamble	Dest	Source	Type	Data	Pad	CRC

- Preamble marks the beginning of the frame.
  - » Also provides clock synchronization
- Source and destination are 48 bit IEEE MAC addresses.
  - » Flat address space
  - » Hardwired into the network interface
- Type field is a demultiplexing field.
  - » What network layer (layer 3) should receive this packet?
  - » Is actually a length field in the 802.3 standard
- CRC for error checking.

Traditional IEEE 802 Networks: MAC in the LAN and MAN

- Ethernet defined as IEEE 802.3.
- The IEEE 802.* set of standards defines a common framing and addressing format for LAN protocols.
  - » Simplifies interoperability
  - » Addresses are 48 bit strings, with no structure
- 802.3 (Ethernet)
- 802.4 (Token bus)
- 802.5 (Token ring)
- 802.6 (Distributed queue dual bus)
- 802.11 (Wireless LAN)
- 802.14 (Cable Modem)
- 802.15 (Wireless Personal Area networks - based on bluetooth)
- 802.16 (Broadband wireless access - “WiMAX”)

LAN Properties

- Exploit physical proximity.
  - » Often a limitation on the physical distance
  - » E.g. to detect collisions in a contention based network
  - » E.g. to limit the overhead introduced by token passing
- Relies on single administrative control and some level of trust.
  - » Broadcasting packets to everybody and hoping everybody (other than the receiver) will ignore the packet
  - » Token-based protocols: everybody plays by the rules
- Broadcast: nodes can send messages that can be heard by all nodes on the network.
  - » Almost essential for network administration
  - » Can also be used for applications, e.g. video conferencing
- But broadcast fundamentally does not scale.

17

How Do We Go Faster?

- How about FDDI?
  - » Too complex
- How about switching, e.g. ATM?
  - » Too expensive and complicated
- How about a faster Ethernet?

Or How about switching Ethernet?
  - » It is simple
  - » It inter-operates with a large installed base
  - » It is Ethernet
  - » Fast Ethernet and Gigabit Ethernet

18

Why Ethernet?

- Easy to manage.
  - » You plug in the host and it basically works
  - » No configuration at the datalink layer
- Broadcast-based.
  - » In part explains the easy management
  - » Some of the LAN protocols (e.g. ARP) rely on broadcast
    - – Networking would be harder without ARP
    - – Address Resolution Protocol (“who-has 18.31.0.114?” -> MAC address).
  - » Not having natural broadcast capabilities adds a lot of complexity to a LAN
    - – Example: ATM
- Drawbacks.
  - » Broadcast-based: limits bandwidth since each packets consumes the bandwidth of the entire network
  - » Distance (if shared)

19

802.3u Fast Ethernet

- Apply original CSMA/CD medium access protocol at 100Mbps
- Must change either minimum frame or maximum diameter: change diameter
- Requires
  - » 2 UTP5 pairs (4B5B) or
  - » 4 UTP3 pairs (8B6T) or
  - » 1 fiber pair
- No more “shared wire” connectivity.
  - » Hubs and switches only

20

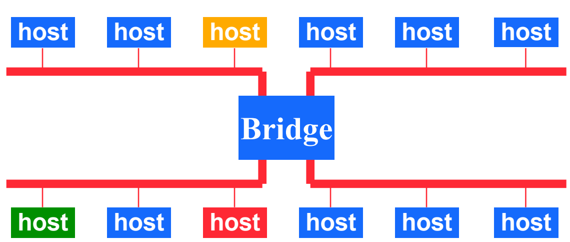
802.3z Gigabit Ethernet

- Same frame format and size as Ethernet.
  - This is what makes it Ethernet
- Full duplex point-to-point links in the backbone are likely the most common use.
  - Added flow control to deal with congestion
- Alternative is half-duplex shared-medium access.
  - Cannot cut the diameter any more (set to 200m)
  - Raise the min frame time (256 bytes), but not frame size
- Choice of a range of fiber and copper transmission media.
- Defining “jumbo frames” for higher efficiency.

21

Building Larger LANs: Bridges

- Bridges connect multiple IEEE 802 LANs at layer 2.
  - Only forward packets to the right port
  - Reduce collision domain compared with single LAN
- In contrast, hubs rebroadcast packets.



22

Transparent Bridges

- Design goals:
  - » "Plug and play" capability
  - » Self-configuring without hardware or software changes
  - » Bridge do not impact the operation of the individual LANs
- Three parts to making bridges transparent:
  - Forwarding of frames
  - Learning of addresses
  - Spanning tree algorithm

23

Frame Forwarding

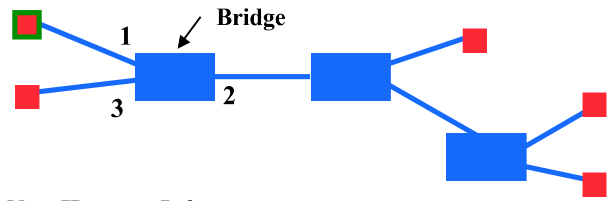
- Each switch maintains a forwarding database:

```
<MAC address, port, age>
    MAC address: host or group address
    Port: port number on the bridge
    Age: age of the entry
```

- Meaning: A machine with MAC address lies in the direction of number port of the bridge
- For every packet, the bridge "looks up" the entry for the packets destination MAC address and forwards the packet on that port.
  - » Other packets are broadcasted – why?

24

Address Lookup

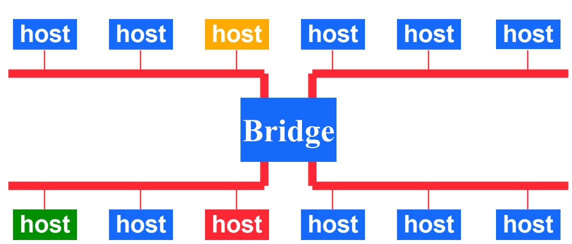


Address	Next Hop	Info
A21032C9A591	1	8:36
99A323C90842	2	8:01
8711C98900AA	2	8:15
301B2369011C	2	8:16
695519001190	3	8:11

- Address is a 48 bit IEEE MAC address.
- Next hop: output port for packet.
- Timer is used to flush old entries
- Size of the table is equal to the number of hosts.

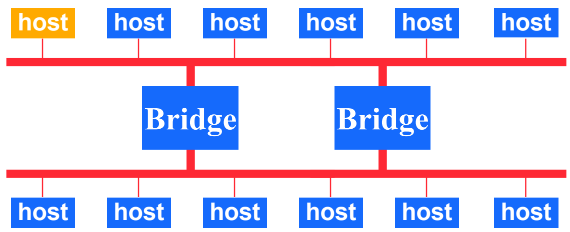
Learning Bridges

- Bridge tables can be filled in manually.
  - Time consuming, error-prone
  - Self-configuring preferred
- Keep track of source address of packets arriving on every link, showing what segment hosts are on.
  - Fill in the forwarding table based on this information



Spanning Tree Bridges

- More complex topologies can provide redundancy.
  - » But can also create loops.
- What is the problem with loops?



27

Spanning Tree Protocol Overview

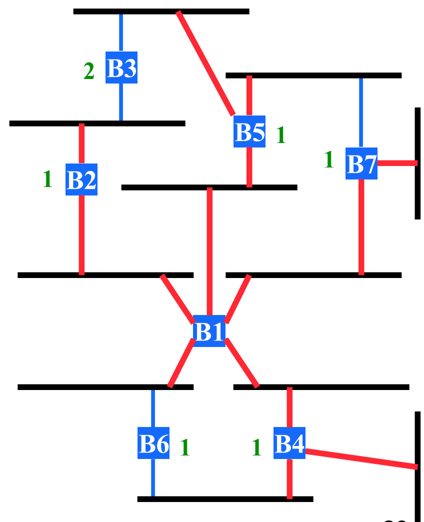
Embed a tree that provides a single unique path to each destination:

- Elect a single bridge as a root bridge
- Each bridge calculates the distance of the shortest path to the root bridge
- Each LAN identifies a *designated bridge*, the bridge closest to the root. It will forward packets to the root.
- Each bridge determines a *root port*, which will be used to send packets to the root
- Identify the ports that form the spanning tree

28

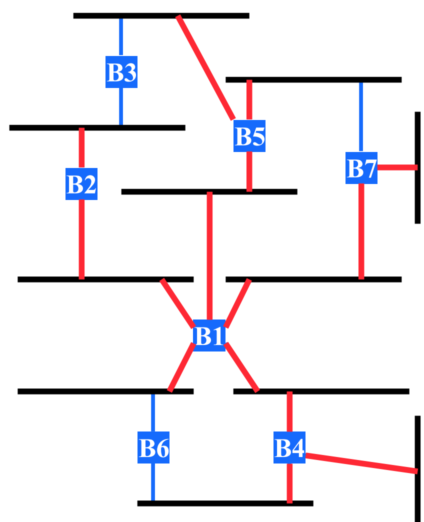
Spanning Tree Algorithm Steps

- Root of the spanning tree is the bridge with the lowest identifier.
  - All ports are part of tree
- Each bridge finds shortest path to the root.
  - Remembers port that is on the shortest path
  - Used to forward packets
- Select for each LAN the designated bridge that has the shortest path to the root.
  - Identifier as tie-breaker
  - Responsible for that LAN



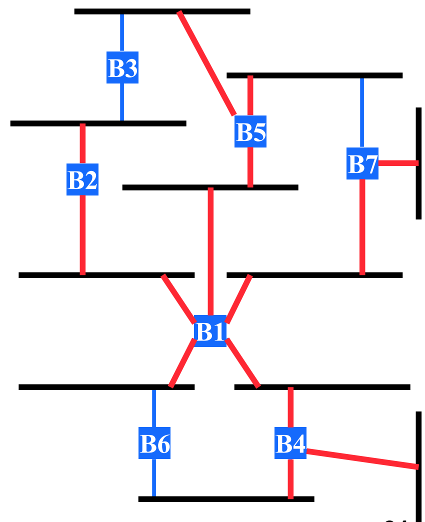
Spanning Tree Algorithm

- Each node sends configuration message to all neighbors.
  - Identifier of the sender
  - Id of the presumed root
  - Distance to the presumed root
  - E.g. B5 sends (B5, B5, 0)
- When B receive a message, it decide whether the solution is better than their local solution.
  - A root with a lower identifier?
  - Same root but lower distance?
  - Same root, distance but sender has lower identifier?
- After convergence, each bridge knows the root, distance to root, root port, and designated bridge for each LAN.



Spanning Tree Algorithm (part 2)

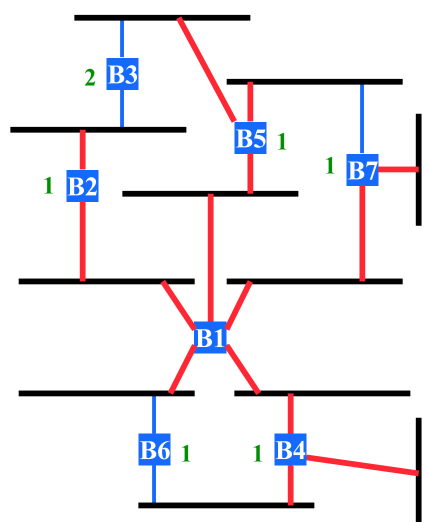
- Each bridge B can now select which of its ports make up the spanning tree:
  - B's root port
  - All ports for which B is the designated bridge on the LAN
- Bridges can not configure their ports.
  - *Forwarding state* or *blocked state*, depending on whether the port is part of the spanning tree
- Root periodically sends configuration messages and bridges forward them over LANs they are responsible for.



31

Spanning Tree Algorithm Example

- Node B2:
  - Sends (B2, B2, 0)
  - Receives (B1, B1, 0) from B1
  - Sends (B2, B1, 1) "up"
  - Continues the forwarding forever
- Node B1:
  - Will send notifications forever
- Node B7:
  - Sends (B7, B7, 0)
  - Receives (B1, B1, 0) from B1
  - Sends (B7, B1, 1) "up" and "right"
  - Receives (B5, B5, 0) - ignored
  - Receives (B5, B1, 1) - better
  - Continues forwarding the B1 messages forever to the "right"



32

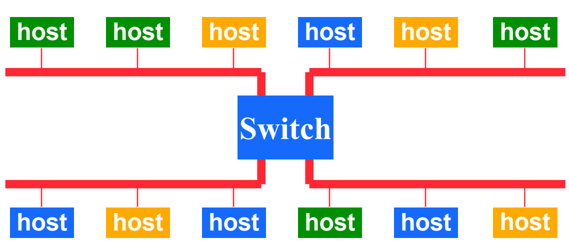
Ethernet Switches

- Bridges make it possible to increase LAN capacity.
  - » Packets are no longer broadcasted - they are only forwarded on selected links
  - » Adds a switching flavor to the broadcast LAN
  - » Some packets still sent to entire tree (e.g., ARP)
- Ethernet switch is a special case of a bridge: each bridge port is connected to a single host.
  - » Can make the link full duplex (really simple protocol!)
  - » Simplifies the protocol and hardware used (only two stations on the link) – no longer full CSMA/CD
  - » Can have different port speeds on the same switch
    - – Unlike in a hub, packets can be stored
    - – An alternative is to use cut through switching

33

Virtual LANs

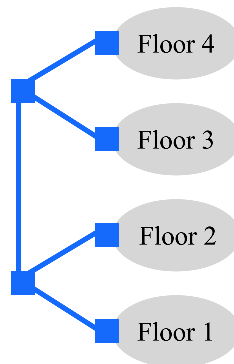
- Single physical LAN infrastructure that carries multiple “virtual” LANs simultaneously.
- Each virtual LAN has a LAN identifier in the packet.
  - » Switch keeps track of what nodes are on each segment and what their virtual LAN id is
- Can bridge and route appropriately.
- Broadcast packets stay within the virtual LAN.
  - » Limits the collision domain for the packet



34

Example LAN Configuration

- 10 or 100 Mbit/second connectivity to the desk top using switch or hubs in wiring closets.
- 100 or 1000 Mbit/second switch fabric between wiring closets or floors.
- Management simplified by having wiring based on star topology with wiring closet in the center.
- Network manager can manage capacity in two ways:
  - » speed of individual links
  - » hub/bridge/switch tradeoff



35