



Ethernet

EE 122: Intro to Communication Networks

Fall 2007 (WF 4-5:30 in Cory 277)

Vern Paxson

TAs: Lisa Fowler, Daniel Killebrew & Jorge Ortiz

<http://inst.eecs.berkeley.edu/~ee122/>

Materials with thanks to Jennifer Rexford, Ion Stoica,
and colleagues at Princeton and UC Berkeley

1

Announcements

- Solutions for Homework #1 are now out.
 - Pick up your graded homeworks in Section, or swing by Jorge's office hours
- Phase 2 of Project #1 is now out.
- No lecture next Wednesday, Oct 10
 - And I won't have office hours that day
- **Midterm** Friday Oct 12
 - Closed book, no gadgets
 - You can have **one** 8.5x11" sheet of notes (both sides)
- I will have *extra office hours* that day, 1-2:30
 - And by appointment on Monday Oct 8

2

Goals of Today's Lecture

- MAC (Media Access Control) protocols, esp.
 - **CSMA/CD**
 - o Carrier Sense Multiple Access / Collision Detection
- Ethernet: single segment
 - Frame structure
 - Length/timing constraints due to Collision Detection
- Ethernet: spanning multiple segments
 - Repeaters and hubs
 - Bridges and switches
 - **Self-learning** (*plug-and-play*)
 - Spanning trees (time permitting)

3

Three Ways to Share the Media

- Channel partitioning MAC protocols (TDMA, FDMA):
 - Share channel efficiently and fairly at high load
 - **Inefficient at low load** (where load = # senders):
 - o $1/N$ bandwidth allocated even if only 1 active node!
- “Taking turns” protocols (discussed in **Section**)
 - Eliminates empty slots without causing collisions
 - Overhead in acquiring the token
 - **Vulnerable to failures** (e.g., failed node or lost token)
- Random access MAC protocols
 - **Efficient at low load**: single node can fully utilize channel
 - High load: collision overhead

4

Key Ideas of Random Access

- **Carrier sense**
 - *Listen before speaking, and don't interrupt*
 - Checking if someone else is already sending data
 - ... and waiting till the other node is done
- **Collision detection**
 - *If someone else starts talking at the same time, stop*
 - Realizing when two nodes are transmitting at once
 - ...by detecting that the data on the wire is garbled
- **Randomness**
 - *Don't start talking again right away*
 - Waiting for a random time before trying again

5

Random Access Protocols

- When node has packet to send
 - Transmit at full channel data rate
 - No *a priori* coordination among nodes
- Two or more transmitting nodes \Rightarrow **collision**
 - Data lost
- Random access MAC protocol specifies (1) how to detect collisions, (2) how to recover from collisions
- Examples
 - ALOHA and **Slotted ALOHA** (discussed in **Section**)
 - **CSMA**, **CSMA/CD** (Ethernet), CSMA/CA (wireless)
 - o CSMA = *Carrier Sense Multiple Access*

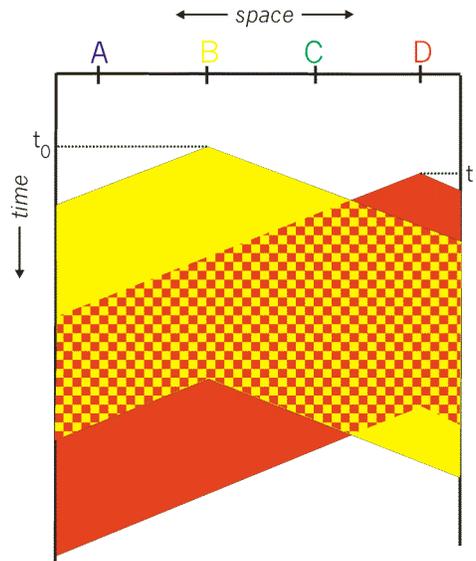
6

CSMA Collisions

Collisions *can* still occur:
propagation delay means two nodes may not hear each other's transmission in time.

At time t_1 , **D** still hasn't heard **B**'s signal sent at the earlier time t_0 , so **D** goes ahead and transmits: failure of *carrier sense*.

Collision:
entire packet transmission time wasted



CSMA/CD (Collision Detection)

- CSMA/CD: carrier sensing, deferral as in CSMA
 - Collisions detected within short time
 - Colliding transmissions aborted, reducing wastage
- Collision detection
 - Easy in wired LANs: measure signal strengths, compare transmitted, received signals
 - Difficult in wireless LANs
 - o Reception shut off while transmitting
 - o Even if on, **might not be able to hear the other sender**, even though **the receiver can**
 - o Leads to use of *collision avoidance* instead

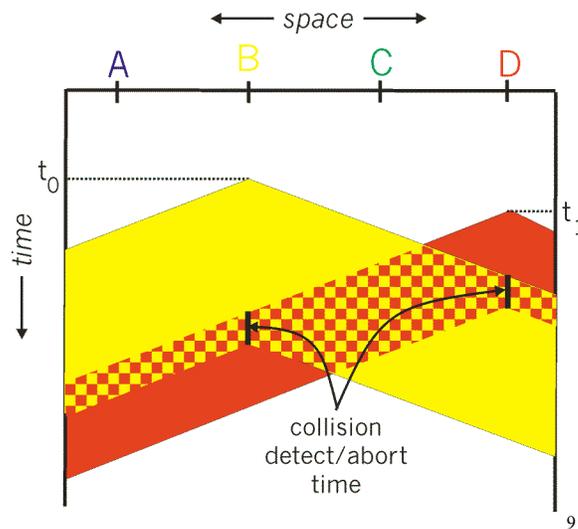
8

CSMA/CD Collision Detection

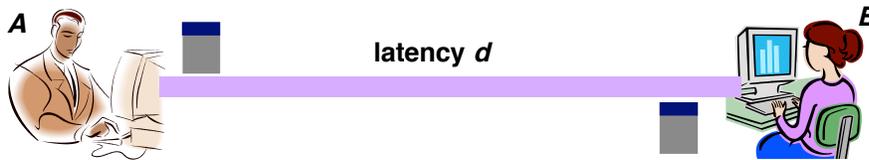
Both **B** and **D** can tell that collision occurred.

This lets them (1) know that they need to resend the frame, and (2) recognize that there's **contention** and adopt a strategy for dealing with it.

Note: for this to work, we need restrictions on **minimum frame size** and **maximum distance**

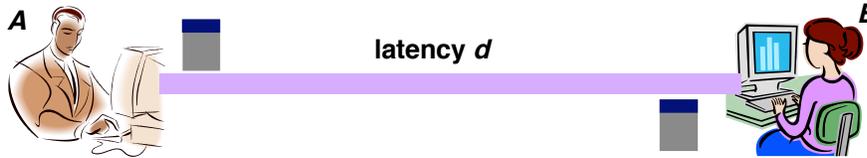


Limits on CSMA/CD Network Length



- Latency depends on physical length of link
 - Time to propagate a packet from one end to the other
- Suppose *A* sends a packet at time t
 - And *B* sees an idle line at a time just before $t+d$
 - ... so *B* happily starts transmitting a packet
- *B* detects a collision, and sends **jamming signal**
 - But *A* can't see collision until $t+2d$

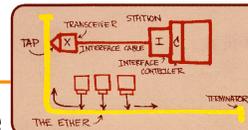
Limits on CSMA/CD Network Length



- A needs to wait for time $2d$ to detect collision
 - So, A should **keep transmitting** during this period
 - ... and keep an eye out for a possible collision
- Imposes restrictions. E.g., for 10 Mbps Ethernet:
 - **Maximum length** of the wire: 2,500 meters
 - **Minimum length** of a frame: 512 bits (64 bytes)
 - o 512 bits = 51.2 μ sec (at 10 Mbit/sec)
 - o For light in vacuum, 51.2 μ sec \approx 15,000 meters
 - vs. 5,000 meters “round trip” to wait for collision

11

Ethernet: CSMA/CD Protocol

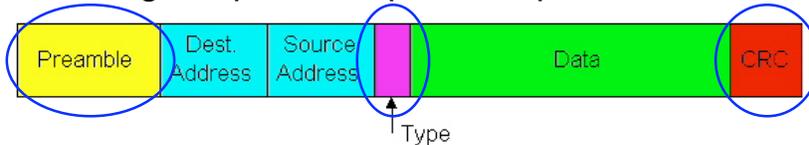


- **Carrier sense:** wait for link to be idle
- **Collision detection:** listen while transmitting
 - No collision: transmission is complete
 - Collision: abort transmission & send **jam** signal
- **Random access:** **exponential back-off**
 - After collision, wait a random time before trying again
 - After m^{th} collision, choose K randomly from $\{0, \dots, 2^m - 1\}$
 - ... and wait for $K \cdot 512$ bit times before trying again
- **The wired LAN technology**
 - **Hugely** successful: 3/10/100/1000/10000 Mbps

12

Ethernet Frame Structure

- Sending adapter encapsulates packet in frame

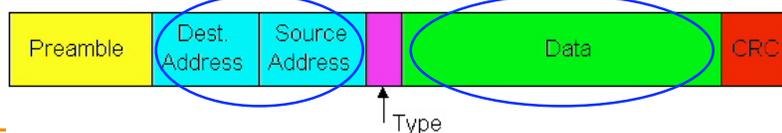


- **Preamble:** synchronization
 - Seven bytes with pattern **10101010**, followed by one byte with pattern **10101011**
 - Used to synchronize receiver & sender
- **Type:** indicates the higher layer protocol
 - Usually IP (but also Novell IPX, AppleTalk, ...)
- **CRC:** cyclic redundancy check
 - Receiver checks & simply drops frames with errors

13

Ethernet Frame Structure (Continued)

- **Addresses:** 48-bit source and destination **MAC addresses**
 - Receiver's adaptor passes frame to network-level protocol
 - o If destination address matches the adaptor's
 - o Or the destination address is the **broadcast address** (**ff:ff:ff:ff:ff:ff**)
 - o Or the destination address is a **multicast group** receiver belongs to
 - o Or the adaptor is in **promiscuous** mode
 - Addresses are **globally unique**
 - o Assigned by NIC vendors (top three **octets** specify vendor)
 - During any given week, > 500 vendor codes seen at LBNL
- **Data:**
 - **Maximum:** 1,500 bytes
 - **Minimum:** 46 bytes (+14 bytes header + 4 byte trailer = 512 bits)



14

Ethernet, con't

- Connectionless
 - No handshaking between sending and receiving adapter
- Unreliable
 - Receiving adapter doesn't send ACKs or NACKs
 - Packets passed to network layer can have gaps
 - Gaps will be filled if application is using TCP
 - Otherwise, application will see the gaps
- 2,700 page IEEE 802.3 standardization
 - <http://standards.ieee.org/getieee802/802.3.html>
- Note, “classical” Ethernet has no length field ...
 - ... instead, sender pauses 9.2 μ sec when done
 - 802.3 shoehorns in a length field

15

Benefits of Ethernet

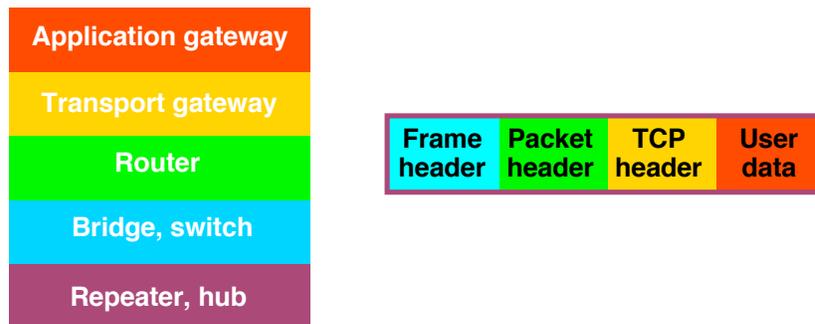
- Easy to administer and maintain
- Inexpensive
- Increasingly higher speed

- Evolved from shared media to **switches**
 - And from electrical signaling to also optical
 - Changes **everything** except the frame **format**
 - A good general lesson for evolving the Internet:
 - o The right **interface** (service model) can often accommodate **unanticipated changes**

16

Shuttling Data at Different Layers

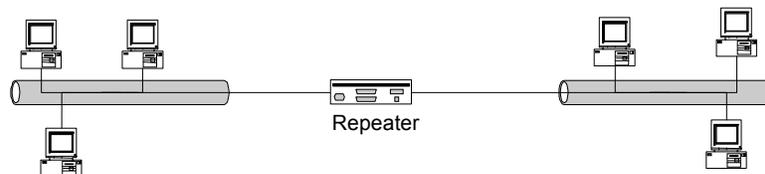
- Different devices switch different things
 - Physical layer: electrical signals (**repeaters** and **hubs**)
 - Link layer: frames (**bridges** and **switches**)
 - Network layer: packets (**routers**)



17

Physical Layer: Repeaters

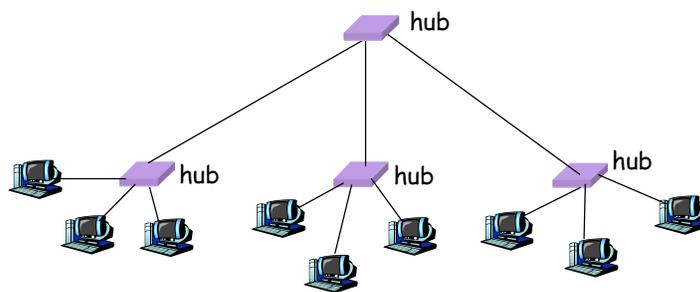
- Distance limitation in local-area networks
 - Electrical signal becomes weaker as it travels
 - Imposes a limit on the length of a LAN
 - o In addition to limit imposed by collision detection
- Repeaters join LANs together
 - Analog electronic device
 - Continuously monitors electrical signals on each LAN
 - Transmits an amplified copy



18

Physical Layer: Hubs

- Joins multiple input lines electrically
 - Do not necessarily amplify the signal
- Very similar to repeaters
 - Also operates at the physical layer



19

Limitations of Repeaters and Hubs

- One large collision domain
 - Every bit is sent everywhere
 - So, aggregate throughput is limited
 - E.g., three departments each get 10 Mbps independently
 - ... and then if connect via a hub must **share** 10 Mbps
- Cannot support multiple LAN technologies
 - Repeaters/hubs do not buffer or interpret frames
 - So, can't interconnect between different rates or formats
 - E.g., no mixing 10 Mbps Ethernet & 100 Mbps Ethernet
- Limitations on maximum nodes and distances
 - Does not circumvent limitations of shared media
 - E.g., still cannot go beyond 2500 meters on Ethernet

20

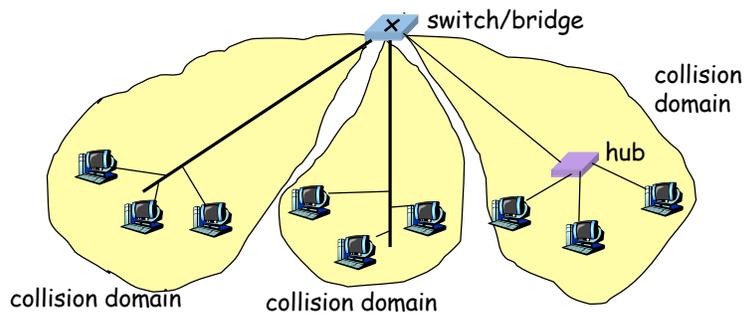
5 Minute Break

Questions Before We Proceed?

21

Link Layer: Switches / Bridges

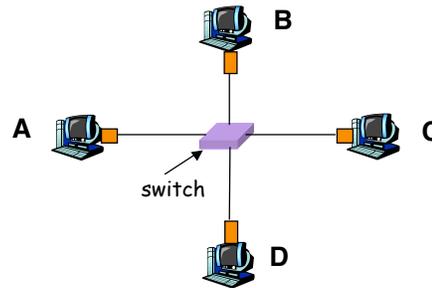
- Connect two or more LANs at the **link layer**
 - Extracts destination address from the frame
 - Looks up the destination in a table
 - Forwards the frame to the appropriate LAN segment
 - o Or point-to-point link, for higher-speed Ethernet
- Each segment is its **own** collision domain



22

Switches & Concurrent Communication

- Host A can talk to C, while B talks to D



- If host has (dedicated) point-to-point link to switch:
 - Full duplex: each connection can send in both directions
 - o At the same time (otherwise, “half duplex”)
 - Completely avoids collisions
 - o No need for carrier sense, collision detection, and so on

23

Advantages Over Hubs & Repeaters

- Only forwards frames as needed
 - Filters frames to avoid unnecessary load on segments
 - Sends frames only to segments that need to see them
- Extends the geographic span of the network
 - Separate collision domains allow longer distances
- Improves privacy by limiting scope of frames
 - Hosts can “snoop” the traffic traversing their segment
 - ... but not all the rest of the traffic
- If needed, applies carrier sense & collision detection
 - Does not transmit when the link is busy
 - Applies exponential back-off after a collision
- Joins segments using different technologies

24

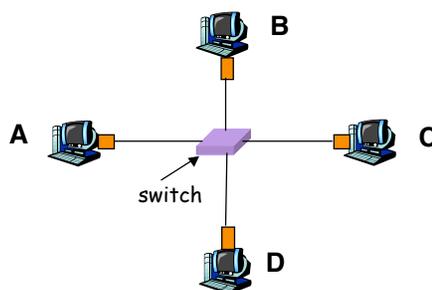
Disadvantages Over Hubs & Repeaters

- Higher cost
 - More complicated devices that **cost** more money
- Delay in forwarding frames
 - Bridge/switch must receive and parse the frame
 - ... and perform a look-up to decide where to forward
 - Introduces **store-and-forward** delay
 - o Can ameliorate using *cut-through switching*
 - Start forwarding after only header received
- Need to **learn** where to forward frames
 - Bridge/switch needs to construct a *forwarding table*
 - Ideally, without intervention from network administrators
 - Solution: **self-learning**

25

Motivation For Self Learning

- Large benefit if switch/bridge forward frames only on segments that need them
 - Allows concurrent use of other links
- Switch table
 - Maps destination MAC address to outgoing interface
 - Goal: construct the switch table automatically

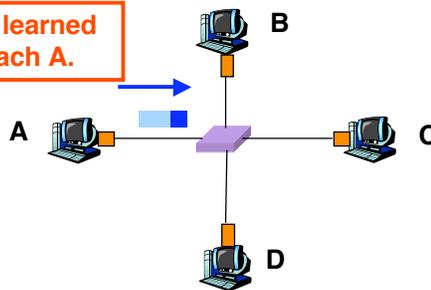


26

Self Learning: Building the Table

- When a frame arrives
 - Inspect *source* MAC address
 - Associate address with the *incoming* interface
 - Store mapping in the switch table
 - Use **time-to-live** field to eventually forget the mapping
 - **Soft state**

Switch just learned how to reach A.

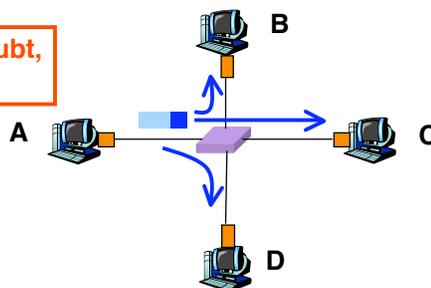


27

Self Learning: Handling Misses

- When frame arrives with unfamiliar destination
 - Forward the frame out **all** of the interfaces (“flooding”)
 - ... except for the one where the frame arrived
 - Hopefully, this case won’t happen very often
 - When destination replies, switch learns that node, too

When in doubt, shout!



28

Switch Filtering / Forwarding

When switch receives a frame:

index the switch table using MAC dest address

if entry found for destination {

if dest on segment from which frame arrived

then drop frame

else forward frame on interface indicated

}

else flood

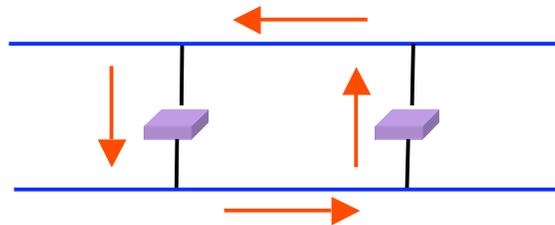
Problems?

forward on all but the interface
on which the frame arrived

29

Flooding Can Lead to Loops

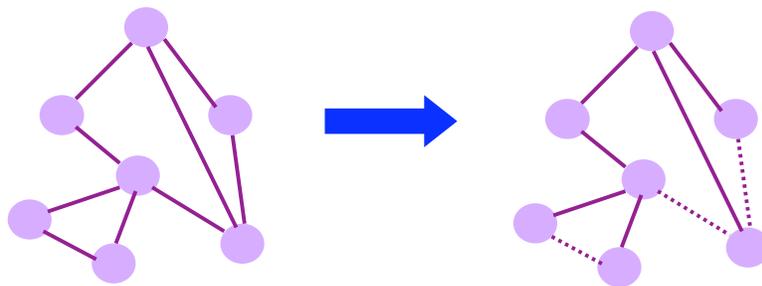
- Switches sometimes need to broadcast frames
 - Upon receiving a frame with an unfamiliar destination
 - Upon receiving a frame sent to the broadcast address
 - Implemented by flooding
- Flooding can lead to **forwarding loops**
 - E.g., if the network contains a cycle of switches
 - o Either accidentally, or by design for higher reliability
 - “Broadcast storm”



30

Solution: Spanning Trees

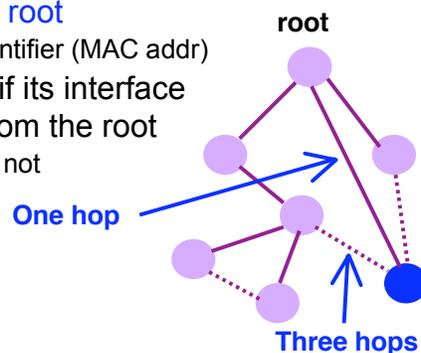
- Ensure the forwarding **topology** has no loops
 - Avoid using some of the links when flooding
 - ... to prevent loop from forming
- **Spanning tree** (K&R pp. 406-408)
 - **Sub-graph** that covers all vertices but *contains no cycles*
 - Links not in the spanning tree do not forward frames



31

Constructing a Spanning Tree

- Need a distributed algorithm
 - Switches cooperate to build the spanning tree
 - ... and **adapt automatically** when **failures** occur
- Key ingredients of the algorithm
 - Switches need to **elect a root**
 - o The switch w/ smallest identifier (MAC addr)
 - Each switch determines if its interface is on the **shortest path** from the root
 - o Excludes it from the tree if not
 - Messages (Y, d, X)
 - o From node X
 - o Proposing Y as the root
 - o And the distance is d



32

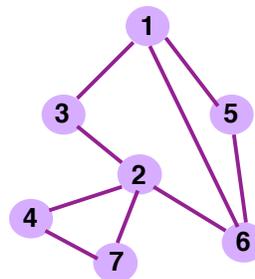
Steps in Spanning Tree Algorithm

- Initially, each switch proposes itself as the root
 - Switch sends a message out every interface
 - ... proposing itself as the root with distance 0
 - Example: switch X announces (X, 0, X)
- Switches update their view of the root
 - Upon receiving message (Y, d, Z) from Z, check Y's id
 - If new id smaller, start viewing that switch as root
- Switches compute their distance from the root
 - Add 1 to the distance received from a neighbor
 - Identify interfaces not on shortest path to the root
 - ... and exclude them from the spanning tree
- If root or shortest distance to it **changed**, flood updated message (Y, d+1, X)

33

Example From Switch #4's Viewpoint

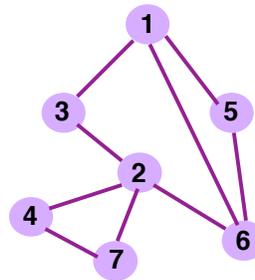
- Switch #4 thinks it is the root
 - Sends (4, 0, 4) message to 2 and 7
- Then, switch #4 hears from #2
 - Receives (2, 0, 2) message from 2
 - ... and thinks that #2 is the root
 - And realizes it is just one hop away
- Then, switch #4 hears from #7
 - Receives (2, 1, 7) from 7
 - And realizes this is a longer path
 - So, prefers its own one-hop path
 - And removes 4-7 link from the tree



34

Example From Switch #4's Viewpoint

- Switch #2 hears about switch #1
 - Switch 2 hears (1, 1, 3) from 3
 - Switch 2 starts treating 1 as root
 - And sends (1, 2, 2) to neighbors
- Switch #4 hears from switch #2
 - Switch 4 starts treating 1 as root
 - And sends (1, 3, 4) to neighbors
- Switch #4 hears from switch #7
 - Switch 4 receives (1, 3, 7) from 7
 - And realizes this is a longer path
 - So, prefers its own three-hop path
 - And removes 4-7 link from the tree



35

Robust Spanning Tree Algorithm

- Algorithm must react to **failures**
 - Failure of the root node
 - o Need to elect a new root, with the next lowest identifier
 - Failure of other switches and links
 - o Need to recompute the spanning tree
- Root switch continues sending messages
 - Periodically reannouncing itself as the root (1, 0, 1)
 - Other switches continue forwarding messages
- Detecting failures through timeout (**soft state**)
 - Switch waits to hear from others
 - Eventually times out and claims to be the root

36

Moving From Switches to Routers

- Advantages of switches over routers
 - Plug-and-play
 - Fast filtering and forwarding of frames
- Disadvantages of switches over routers
 - Topology restricted to a spanning tree
 - Large networks require large ARP tables
 - Broadcast storms can cause the network to collapse
 - Can't accommodate non-Ethernet segments (why not?)

37

Comparing Hubs, Switches & Routers

	<u>hubs</u>	<u>switches</u>	<u>routers</u>
traffic isolation	no	yes	yes
plug & play	yes	yes	no
optimized routing	no	no	yes
cut-through	yes	yes*	no*

38

Summary

- Ethernet as an exemplar of link-layer technology
- Simplest form, single segment:
 - *Carrier sense, collision detection*, and random access
- Extended to span multiple segments:
 - Hubs & repeaters: physical-layer interconnects
 - Bridges / switches: link-layer interconnects
- Key ideas in switches
 - Self learning of the switch table
 - Spanning trees
- Next time: midterm review