Hubs, Switches, and Routers, Oh My!

Mike Freedman

COS 461: Computer Networks

http://www.cs.princeton.edu/courses/archive/spr20/cos461/

Today: Hubs, Switches, and Routers, Oh My!
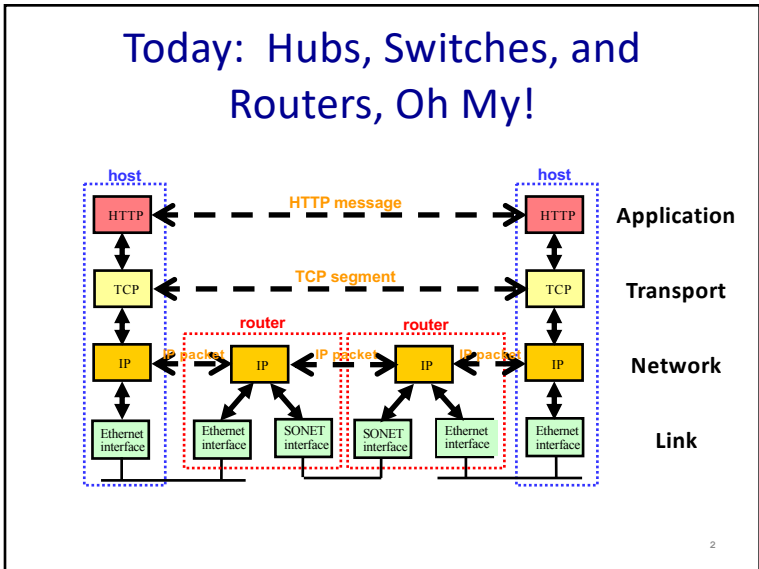
OSI Model

Developed in late 1970s / early 1980s

Data	Layer	
Data	**Application** — Network Process to Application	**Layer 7**
Data	**Presentation** — Data representation and Encryption	
Data	**Session** — Interhost communication	
Segments	**Transport** — End-to-End connections and Reliability	**Layer 4**
Packets	**Network** — Path Determination and IP (Logical addressing)	**Layer 3**
Frames	**Data Link** — MAC and LLC (Physical addressing)	**Layer 2**
Bits	**Physical** — Media, Signal and Binary Transmission	

Host Layers: Application, Presentation, Session, Transport
Media Layers: Network, Data Link, Physical

---

Terminology

- **Hubs and Repeaters**
  - Connect machines on same "layer 2" LAN
  - Broadcast: All frames are sent out all physical ports

- **Switches and Bridges**
  - Connect machines on same "layer 2" LAN
  - Only send frames to selected physical port based on destination MAC address

- **Routers**
  - Connect between LANs at "layer 3", e.g., wide area
  - Only send packet to selected physical port based on destination IP address

1

## "Layer 2"
Hubs and Switches

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
- Repeaters join LANs together
  - Analog electronic device
  - Continuously monitors electrical signals
  - Transmits an amplified copy
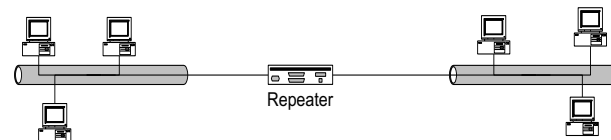
Repeater

Physical Layer: Hubs

- Joins multiple input lines electrically
  - Designed to hold multiple line cards
  - Do not necessarily amplify the signal
- Very similar to repeaters
  - Also operates at the physical layer
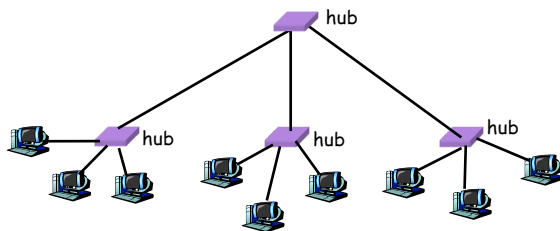
Limitations of Repeaters and Hubs

- One large shared link
  - Each bit is sent everywhere
  - So, aggregate throughput is limited
- Cannot support multiple LAN technologies
  - Does not buffer or interpret frames
  - Can't interconnect between different rates/formats
- Limitations on maximum nodes and distances
  - Shared medium imposes length limits
  - E.g., cannot go beyond 2500 meters on Ethernet

Link Layer: Bridges

- Connects 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 segment
- Each segment can carry its own traffic
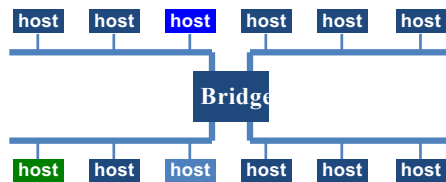
Link Layer: Switches

- Typically connects individual computers
  - A switch is essentially the same as a bridge
  - … though typically used to connect hosts
- Supports concurrent communication
  - Host A can talk to C, while B talks to D
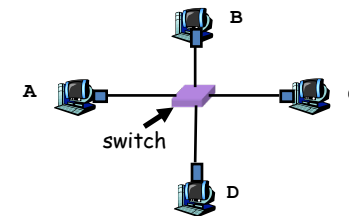
## Bridges/Switches: Traffic Isolation

- Switch filters packets
  - Frame only forwarded to the necessary segments
  - Segments can support separate transmissions
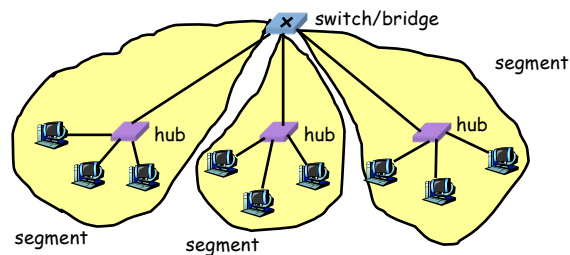
Self Learning: Building the Table

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

Switch learns how to reach A
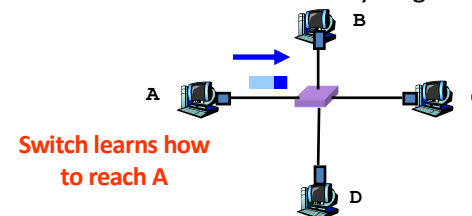
Self Learning: Handling Misses

- When frame arrives with unfamiliar destination
  - Forward the frame out all of the interfaces
  - … except for the one where the frame arrived
  - Hopefully, this case won't happen very often!

When in doubt, shout!

13
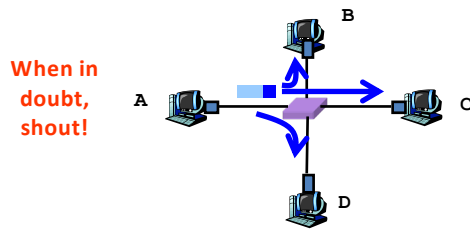
Switches vs. Hubs

- Compared to hubs, Ethernet switches support
  - (Y) Larger geographic span
  - (M) Similar span
  - (C) Smaller span

- Compared to hubs, switches provides
  - (Y) Higher load on links
  - (M) Less privacy
  - (C) Heterogenous communication technologies
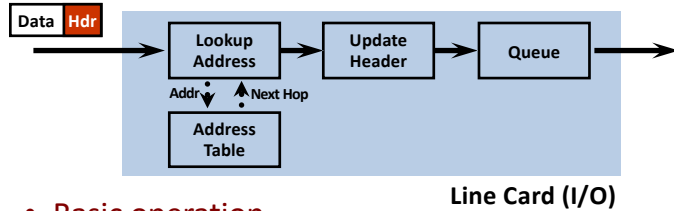
14

Looking closer…

15

Basic Router Architecture

- Each switch/router has a forwarding table
  - Maps destination address to outgoing interface

- Basic operation
  1. Receive packet
  2. Look at header to determine destination address
  3. Look in forwarding table to determine output interface
  4. Modify packet header (e.g., decr TTL, update chksum)
  5. Send packet to output interface

16

Basic Router Architecture

| Data | Hdr |

Lookup Address → Update Header → Queue →

Addr / Next Hop
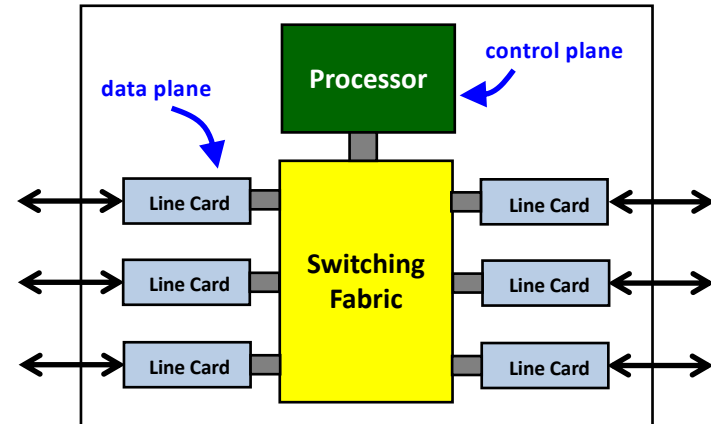
Address Table

Line Card (I/O)

- Basic operation
    1. Receive packet
    2. Look at header to determine destination address
    3. Look in forwarding table to determine output interface
    4. Modify packet header (e.g., decr TTL, update chksum)
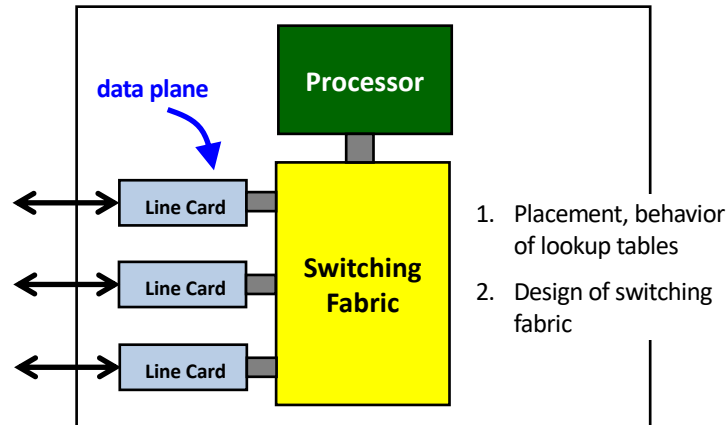   5. Send packet to output interface

17

Router

Processor

data plane   control plane

Line Card   Line Card

Line Card   **Switching Fabric**   Line Card

Line Card   Line Card

18

Router

data plane

Processor

Line Card

Line Card   **Switching Fabric**

Line Card

1. Placement, behavior of lookup tables
2. Design of switching fabric

19

Lookup algorithm depends on protocol

Protocol	Mechanism	Techniques
Ethernet (48 bits) MPLS ATM	Exact Match	• Direct lookup • Associative lookup • Hashing • Binary tree
IPv4 (32 bits) IPv6 (128 bits)	Longest-Prefix Match	• Radix trie • Compressed trie • TCAM

20

5

Longest Prefix Match

- Each packet has destination IP address
- Router looks up table entry that matches address

68.211.6.120

| Prefix	Output
68.208.0.0/12	1
68.211.0.0/17	1
68.211.128.0/19	2
68.211.160.0/19	2
68.211.192.0/18	1

21

Longest Prefix Match

- Each packet has destination IP address
- Router looks up table entry that matches address

- Benefit of CIDR allocation and LPR
  - Efficiency: Prefixes can be allocated at much finger granularity
  - Hierarchical aggregation: Upstream ISP can aggregate 2 contiguous prefixes from downstream ISPs to shorter prefix

22

How? Address lookup using trie

- Prefixes "spelled out" by following path from root
- To find the best prefix spell out address in trie

	Prefixes
A	111*
B	10*
C	1010*
D	10101
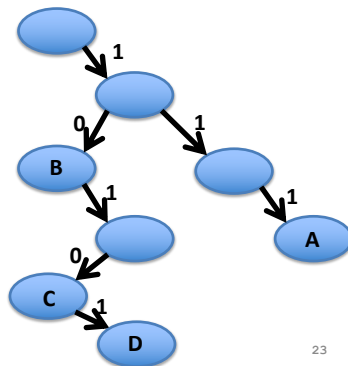
23

How? Address lookup using trie

- Prefixes "spelled out" by following path from root
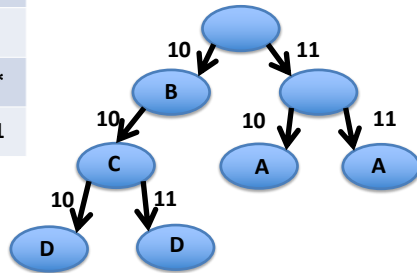- To find the best prefix spell out address in trie

- 1 per / lvl = max 32 lookups!
- Too slow:
  - E.g., "Optical Carrier 48" line (2.5 Gbps) requires 160ns lookup … or 4 memory accesses

24

6
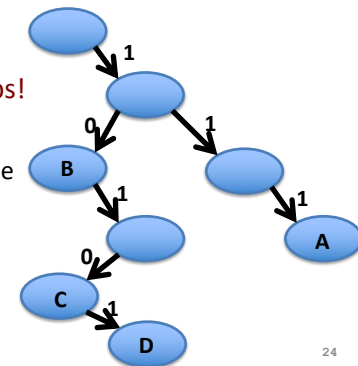
K-ary trie (k=2)

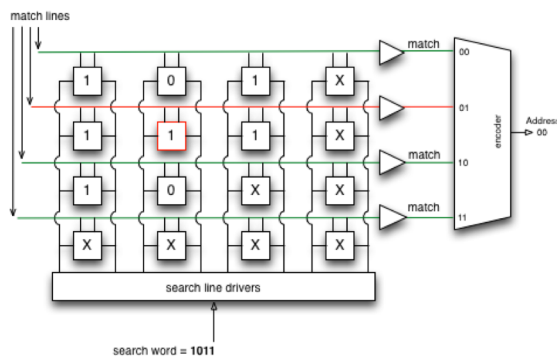|	Prefixes
A	111*
B	10*
C	1010*
D	10101

25

Alternatives to trie for LPM

- Content-Address Memory (CAM)
  - Input: tag (address)
  - Output: value (port
  - Exact match, but O(1) in hardware
- Ternary CAM
  - 0, 1, *
  - "value" memory cell and "mask" (care / don't care) cell
- LPM via TCAM
  - In parallel, search all prefixes for all matches
  - Then choose longest match
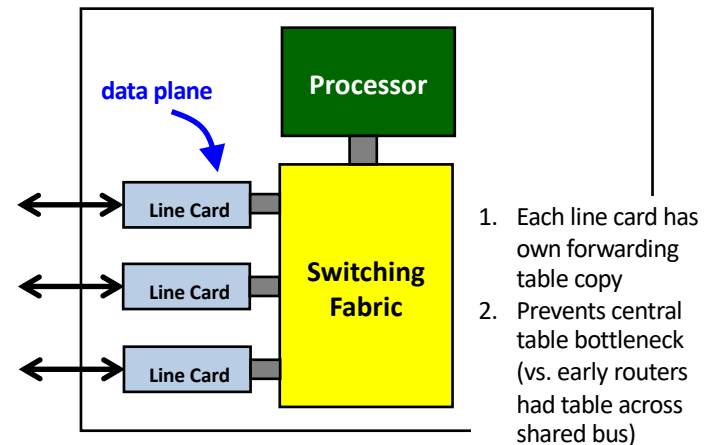   - Trick: choose first match, but already sorted by prefix length
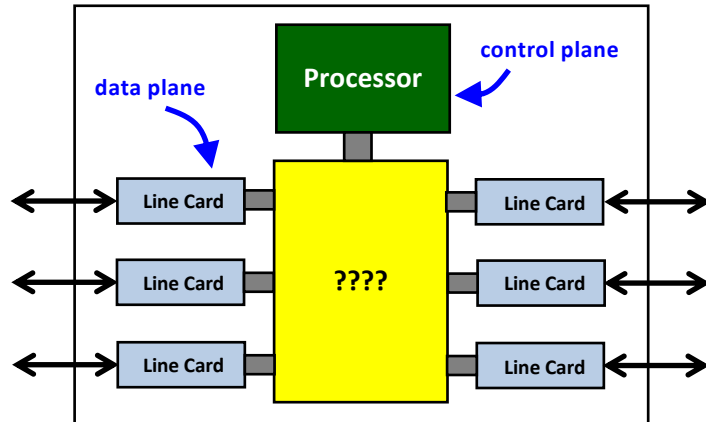
26

Alternatives to trie for LPM

search word = 1011

27

Decision: Forwarding table per line card

1. Each line card has own forwarding table copy
2. Prevents central table bottleneck (vs. early routers had table across shared bus)

7

Decision: Crossbar switch

Processor

control plane

data plane

Line Card | ???? | Line Card
Line Card | | Line Card
Line Card | | Line Card

29

Decision: Crossbar switch

- Shared bus
  - Only one input can speak to one output at a time

- Crossbar switch / switched backplane
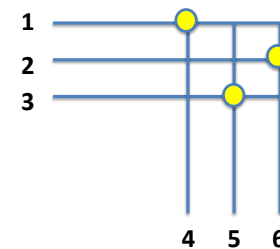 - Input / output pairs that don't compete can send in same timeslot

30

Crossbar switching

- Every input port has connection to every output port
- In each timeslot, each input connected to zero or more outputs


1 | 4
2 | 5
3 | 6

31

Crossbar switching

- Every input port has connection to every output port
- In each timeslot, each input connected to zero or more outputs

1
2
3
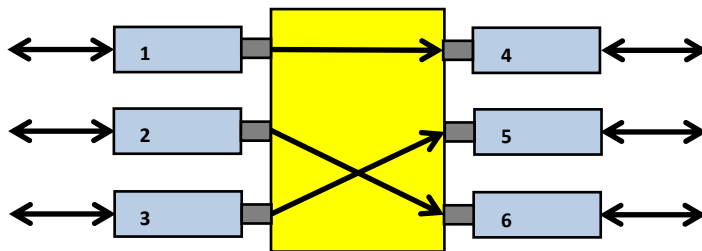
- Good parallelism

- Needs scheduling

4 5 6

32

8

Everything gets complicated…

- Problem: Head-of-line blocking
  - The packet in front of queue blocks packets behind it from being processed
  - Say first packet at input 1 wants to go to output 5; second packet at 1 that wants 4 is still blocked
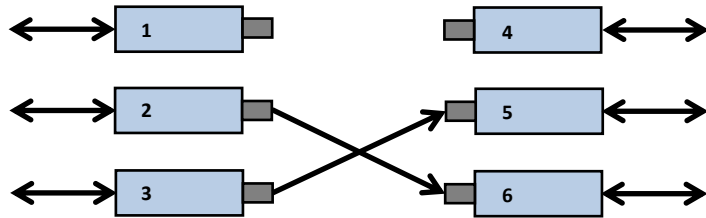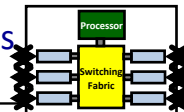

33

Everything gets complicated…

- Solution: Virtual output queues
  - One queue at input per output port
  - So avoids head-of-line blocking during crossbar scheduling
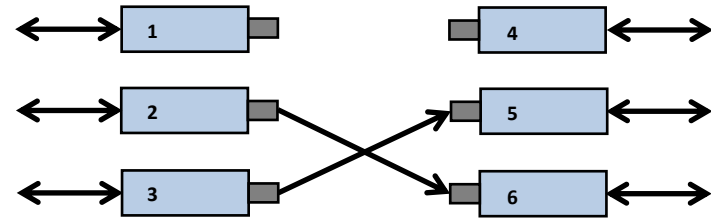
34

Data, Control, & Management Planes

	Data	Control	Management
Time-scale	Packet (ns)	Event (10 ms to sec)	Human (min to hours)
Tasks	Forwarding, buffering, filtering, scheduling	Routing, signaling	Analysis, configuration
Location	Line-card hardware	Router software	Humans or scripts

35

Cisco 8000 Series Routers

- Up to 648 400 GbE
- 260 Tbps backplane

36

Conclusions

- Physical devices sharing L2 & L3 networks have many common features
  - Forward table lookups
  - Queueing and backplane switching
  - Fast vs. slow paths
    - Switches and routers separate routing decisions (control plane) from forwarding actions (data plane)
- High speed necessitates innovation
  - Specialized hardware
  - Software algorithms

37