COS 461 Computer Networks
Lecture 4: Hubs, Switches, and Routers

Kyle Jamieson
Fall 2020
Today: Hubs, Switches, and Routers, Oh My!

Application
Transport
Network
Link
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
“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
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!
Switches vs. Hubs

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

• Compared to hubs, switches provide
  (Y) Higher load on links
  (M) Less privacy
  (C) Traffic isolation
Routers: Looking closer...
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
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
Router

1. Placement, behavior of lookup tables
2. Design of switching fabric
Lookup algorithm depends on protocol

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Mechanism</th>
<th>Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethernet (48 bits)</td>
<td>Exact Match</td>
<td>• Direct lookup</td>
</tr>
<tr>
<td>MPLS</td>
<td></td>
<td>• Associative lookup lookup</td>
</tr>
<tr>
<td>ATM</td>
<td></td>
<td>• Hashing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Binary tree</td>
</tr>
<tr>
<td>IPv4 (32 bits)</td>
<td>Longest-Prefix Match</td>
<td>• Radix trie</td>
</tr>
<tr>
<td>IPv6 (128 bits)</td>
<td></td>
<td>• Compressed trie</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• TCAM</td>
</tr>
</tbody>
</table>
Longest Prefix Match (LPM)

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

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>68.208.0.0/12</td>
<td>1</td>
</tr>
<tr>
<td>68.211.0.0/17</td>
<td>1</td>
</tr>
<tr>
<td>68.211.128.0/19</td>
<td>2</td>
</tr>
<tr>
<td>68.211.160.0/19</td>
<td>2</td>
</tr>
<tr>
<td>68.211.192.0/18</td>
<td>1</td>
</tr>
</tbody>
</table>

68.211.6.120
LPM: Motivation

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

• Benefits of CIDR allocation and LPM
  – **Efficiency:** Prefixes can be allocated at much finer granularity
  – **Hierarchical aggregation:** Upstream ISP can aggregate 2 contiguous prefixes from downstream ISPs to shorter prefix
Software LPM lookup using trie

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

<table>
<thead>
<tr>
<th>Prefixes</th>
<th>Prefixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>111*</td>
</tr>
<tr>
<td>B</td>
<td>10*</td>
</tr>
<tr>
<td>C</td>
<td>1010*</td>
</tr>
<tr>
<td>D</td>
<td>10101</td>
</tr>
</tbody>
</table>

Diagram:

```
A  1
  /  
B  0   C
 /    /  
1    1    0
```
Software LPM lookup using trie

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

- 1 lookup per level → max 32 lookups/address!
- Too slow:
  - E.g., “Optical Carrier 48” line (2.5 Gbps) requires 160ns lookup ... or 4 memory accesses
Software LPM lookup: k-ary trie (k=2)

<table>
<thead>
<tr>
<th></th>
<th>Prefixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>111*</td>
</tr>
<tr>
<td>B</td>
<td>10*</td>
</tr>
<tr>
<td>C</td>
<td>1010*</td>
</tr>
<tr>
<td>D</td>
<td>10101</td>
</tr>
</tbody>
</table>

Diagram:

```
  11
  / 
 10 10
 /   /
B 10 C 10
 |   |   |
D   A   A
```
Hardware for LPM lookup

- **Content-Address Memory (CAM)**
  - Input: tag (address)
  - Output: value (port)
  - Exact match, but O(1) in hardware

- **Ternary CAM**
  - Can have wildcards: 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
Example: LPM with a TCAM

Prefix | TCAM format
-------|-------------
101/3  | 101X
111/3  | 111X
10/2   | 10XX
0/0    | XXXX
Decision: Forwarding table per line card

1. Each line card has its own forwarding table copy.
2. Prevents central table bottleneck (vs. early routers had table across a shared bus).
Decision: Crossbar switch

- Processor
- Data plane
- Control plane
- Line Card
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
Crossbar switching

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

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

- Good parallelism
- Needs scheduling
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
Everything gets complicated...

• **Solution: Virtual output queues**
  
  – One queue at input, *per output port* (for all inputs)
  
  – So avoids head-of-line blocking during crossbar scheduling
## Data, Control, & Management Planes

<table>
<thead>
<tr>
<th>Time-scale</th>
<th>Data</th>
<th>Control</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet (ns)</td>
<td>Event (10 ms to sec)</td>
<td>Human (min to hours)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Data</th>
<th>Control</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forwarding, buffering, filtering, scheduling</td>
<td>Routing, signaling</td>
<td>Analysis, configuration</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Data</th>
<th>Control</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line-card hardware</td>
<td>Router software</td>
<td>Humans or scripts</td>
<td></td>
</tr>
</tbody>
</table>
Cisco 8000 Series Routers

- Up to 648 400 GbE
- 260 Tbps backplane
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