CSE/EE 461 – Lecture 9

David Wetherall
djw@cs.washington.edu

Last Time

- Focus
  - How do we make routing scale?
- Topics
  - IP Addressing (ARP, DHCP, CIDR, subnets)
  - Inter-domain routing (EGP, BGP)

Application
Presentation
Session
Transport
Network
Data Link
Physical

This Time

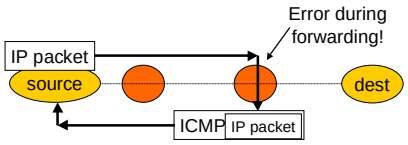
- Focus
  - How do we build a router?
- Topics
  - First an aside (ICMP)
  - A brief look at Queuing and Switching

Application
Presentation
Session
Transport
Network
Data Link
Physical

An Aside – ICMP

- What happens when things go wrong?
  - Need a way to test/debug a large, widely distributed system
- ICMP = Internet Control Message Protocol (RFC792)
  - Companion to IP – required functionality
- Used for error and information reporting:
  - Errors that occur during IP forwarding
  - Queries about the status of the network

ICMP Generation



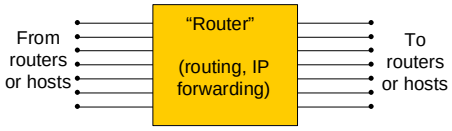
Common ICMP Messages

- Destination unreachable
  - "Destination" can be host, network, port or protocol
- Redirect
  - To shortcut circuitous routing
- TTL Expired
  - Used by the "traceroute" program
- Echo request/reply
  - Used by the "ping" program
- ICMP messages include portion of IP packet that triggered the error (if applicable) in their payload

ICMP Restrictions

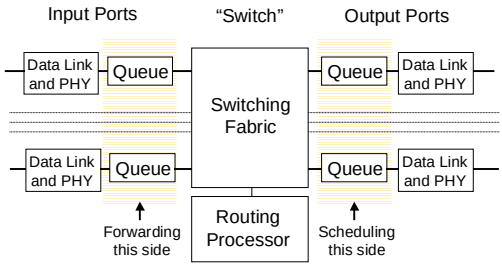
- The generation of error messages is limited to avoid cascades ... error causes error that causes error!
- Don't generate ICMP error in response to:
  - An ICMP error
  - Broadcast/multicast messages (link or IP level)
  - IP header that is corrupt or has bogus source address
  - Fragments, except the first
- ICMP messages are often rate-limited too.

What's in a Router?



- By convention, draw input ports on left, output on right. But a single physical port handles both directions.

Model of a Router



Where do packets get dropped?

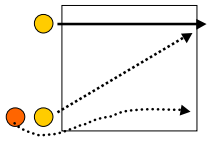
- Queuing can occur at both input and output stages
  - Packets are discarded when a queue overflows
- But:
  - If we can process packets at the line rate
  - If switching fabric is at least N times the line rate (N ports)
  - Then queue overflow (loss) won't happen at input
- However:
  - That's expensive ... so we may have input queuing

Switching Fabrics

- There's a network within a router!
- Possibilities:
  - Shared memory (imagine a PC with line cards)
  - Shared bus (PC with transfers between line cards)
  - Interconnection network
- Interconnection networks
  - Crossbar (imagine a matrix)
  - Knockout Switch, Banyan, Batcher-Banyan, ...

Head of Line Blocking

- Consequence of input queuing when there is contention for output ports



- Orange packet is blocked behind the yellow head of line packet, even though it's output port is free
  - Limits crossbar utilization to < 60% with uniformly distributed traffic

Scheduling

- What order do we sent queued packets?
  - Order called a scheduling discipline
- FIFO scheduling (Drop Tail routers)
  - Simple and widely used in practice
- WFQ = Weighted Fair Queuing
  - Some notion of "give each user their share"
  - More complex, may be used in the future
- We return to this topic later: Quality of Service (QOS)

Key Concepts

- ICMP helps us test/debug IP networks
- Router internals
  - Queuing (packet loss)
  - Switching fabrics (performance)
  - Scheduling (QOS)