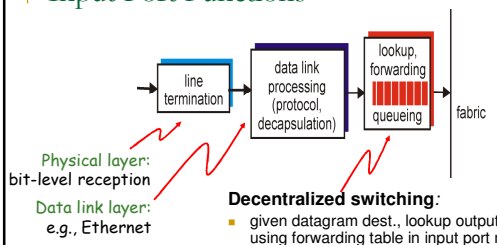
Routers: Forwarding

EECS 122: Lecture 13

Department of Electrical Engineering and Computer Sciences
University of California
Berkeley

Input Port Functions



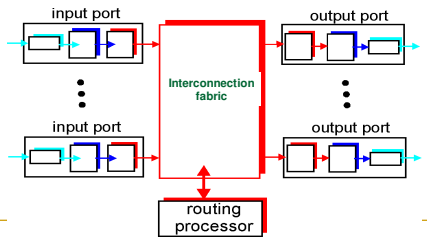
Decentralized switching:

- given datagram dest., lookup output port using forwarding table in input port memory
- goal: complete input port processing at 'line speed'
- queuing: if datagrams arrive faster than forwarding rate into switch fabric

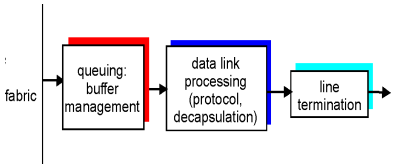
Router Architecture Overview

Two key router functions:

- run routing algorithms/protocol (RIP, OSPF, BGP)
- *forwarding* datagrams from incoming to outgoing link



Output Ports

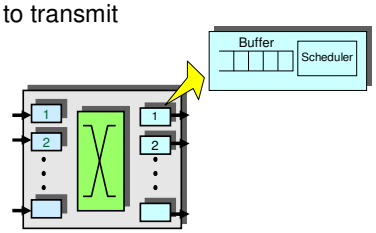


Today: Focus on Forwarding Datagrams

- Input Ports
- Output Ports
- Interconnection Fabric
- Forwarding Process
  - Datagrams: Lookups
  - (Virtual Circuit next lecture)
- Examples of Routers

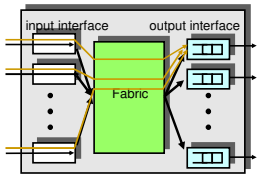
Queuing Functions

- Buffer management: decide when and which packet to drop
- Scheduler: decide when and which packet to transmit

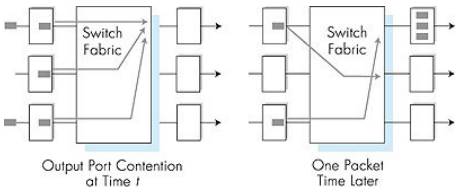


Output Queued Router

- Only output interfaces store packets
- Advantage
  - Easy to design algorithms: only one congestion point

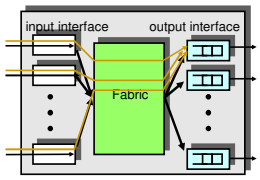


Output Queued Routers



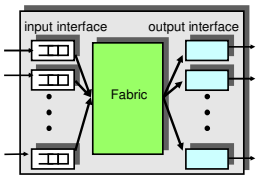
Output Queued Router

- Only output interfaces store packets
- Advantage
  - Easy to design algorithms: only one congestion point
- Disadvantage
  - Requires a speedup of a factor of N, where N is the number of interfaces → not feasible for large N



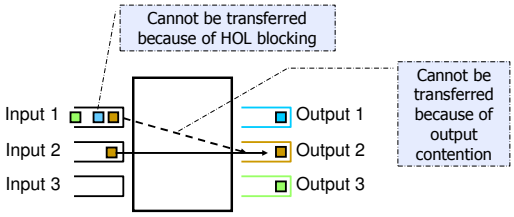
Input Queues

- Only input interfaces store packets
- Advantages
  - Easy to build
  - Simple algorithms



Input Queues: Head-of-line Blocking

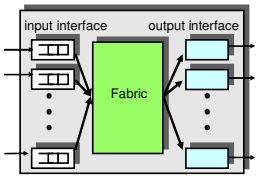
- The packet at the head of an input queue cannot be transferred, thus blocking the following packets



Wastes router capacity

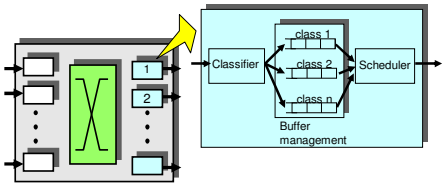
Input Queues

- Only input interfaces store packets
- Advantages
  - Easy to build
  - Simple algorithms
- Disadvantages
  - HOL Blocking
- Need a speedup that eliminates HOL but does not create output queues…
  - About 2 suffices…

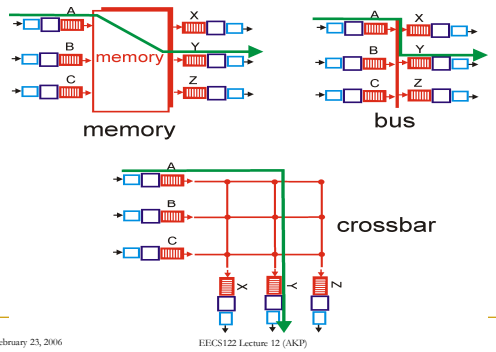


Advanced Queuing Functions

- Packet classification: map each packet to a predefined class
  - use to implement more sophisticated services (e.g., QoS)
- Flow: a subset of packets between any two endpoints in the network

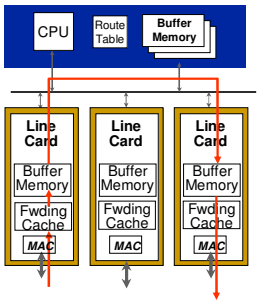


Three types of switching fabrics



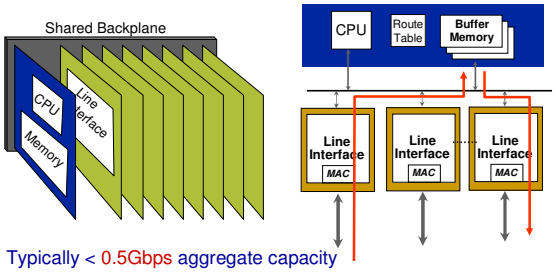
Shared Bus Fabrics

•Typically < 5Gb/s aggregate capacity
•Limited by shared bus
•1 Gbps bus, Cisco 1900: sufficient speed for access and enterprise routers (not regional or backbone)



Slide by Nick McKeown

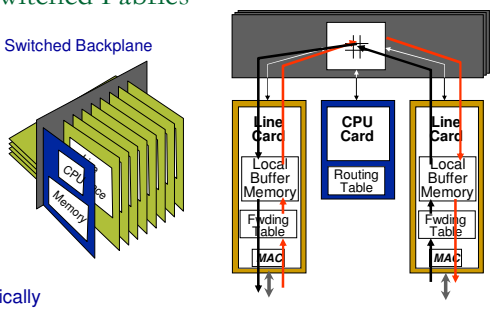
Shared Memory Based Fabrics



Typically < 0.5Gbps aggregate capacity
Limited by rate of shared memory

Slide by Nick McKeown

Switched Fabrics



Typically
< 50Gbps aggregate capacity

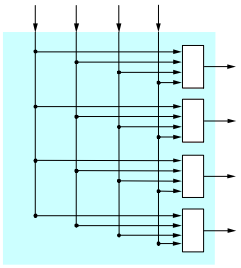
Slide by Nick McKeown

Switched Fabrics

- Overcome bus bandwidth limitations
- Banyan networks, other interconnection nets initially developed to connect processors in multiprocessor
- Advanced design: fragmenting datagram into fixed length cells, switch cells through the fabric.
- Cisco 12000: switches Gbps through the interconnection network

Example: Crossbar Switches

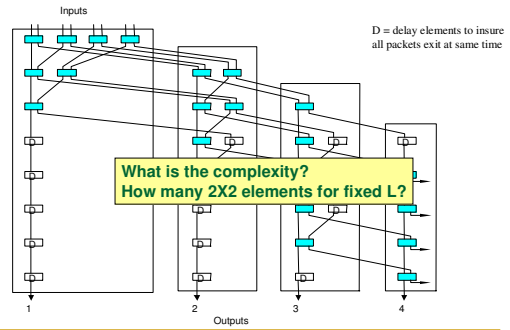
- Basic Idea
  - N^2 switching points
    - not scalable
- Engineering Idea
  - It is very unlikely that more than L packets will want to go to the same output port simultaneously
  - How many switching points can we save for fixed L?



The Knockout Concentrator

- Goal: If there are greater than L packets that want to go to the same destination, pick L in a fair manner.
- Organize the switching elements as if they are implementing a multi-round tournament
  - A game consists of two players and the winner is selected at random (at a switching element)
  - The winner moves on to the next round, while the loser plays a "consolation" rank
  - The top L players are selected.

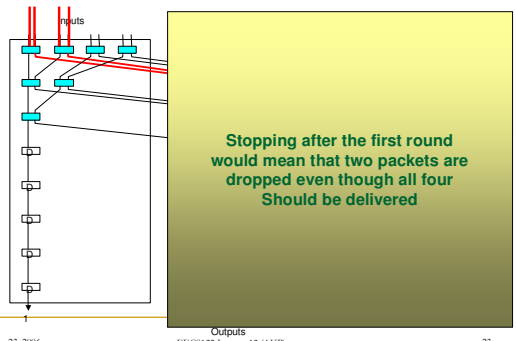
Knockout Switch Concentrator



Why have multiple rounds?

- Example: N=8, L=4
- After one round, there are four winners and four losers
- Why not just pick the winners and drop the losers?

Example



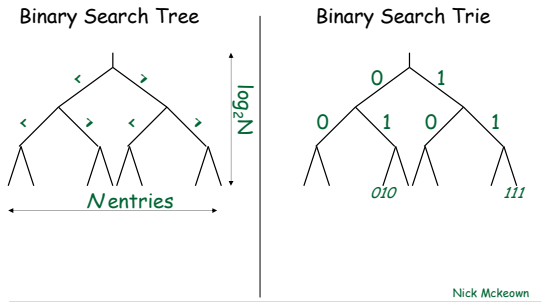
The Forwarding Decision Process

- Datagram Routing: Each packet is independently forwarded at each router
  - Must look up IP address ranges
- Virtual Circuit Routing:
  - call setup, teardown for each call *before* data can flow
  - each packet carries VC identifier (not destination host address)
  - *every* router on source-dest path maintains "state" for each passing connection
  - link, router resources (bandwidth, buffers) may be *allocated* to VC (dedicated resources = predictable service)

Datagram Route Lookup

- Longest Prefix Match
  - Not easy to do at line speeds!
- It is useful to think of the search process as a traversal of a special kind of labeled tree called a Trie

Trees and Tries

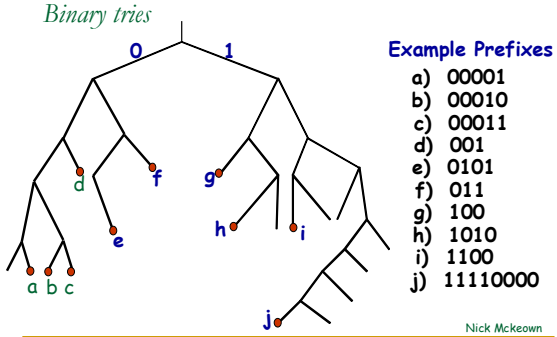


Simple Tries and LPM

- The routing table entry is a variable length prefix
  - E.g. 01111111 00001111 0000100100 for 128.23.9.0/26
  - A balanced tree won't work
  - Variable number of steps required

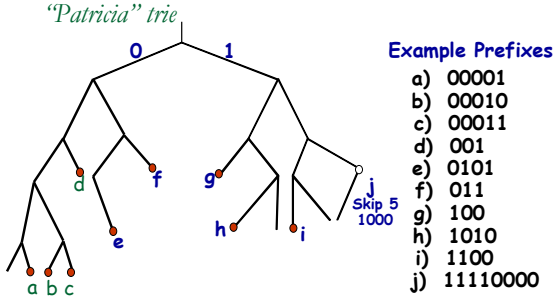
LPM in IP Routers

Binary tries



LPM in IP Routers

"Patricia" trie

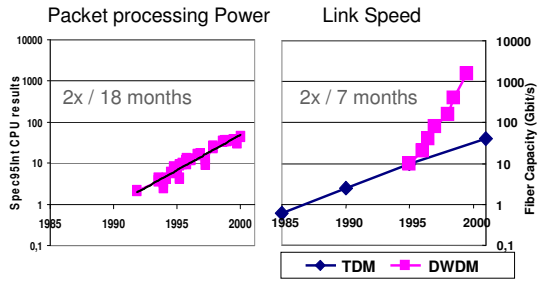


Router Performance

- Goal: To work at line speed
  - Depends on interfaces
- Throughput is difficult to quantify
  - Depends on traffic flow
  - Traffic flow is hard to model
- Packets per second hard to quantify
  - IP packets are of variable size
- Routers at different parts of the network have different characteristics

Why we Need Faster Routers

1: *To prevent routers from being the bottleneck*

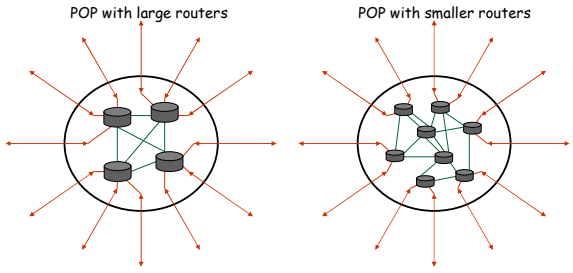


Source: SPEC95Int & David Miller, Stanford.

Slide by Nick McKeown

Why we Need Faster Routers

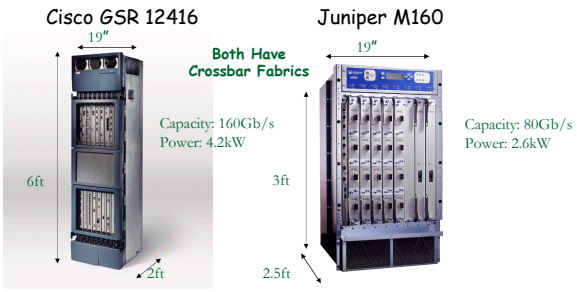
2: *To reduce cost, power & complexity of Data Centers*



- Ports: Price >$100k, Power > 400W.
- It is common for 50-60% of ports to be for interconnection.

Slide by Nick McKeown

Example: Wide Area Routers



Slide by Nick McKeown

Examples: Cisco 7600

- MAN-WAN Router
- Up to 128 Gbps with Crossbar Fabric
- 10Mbps – 10Gbps LAN Interfaces
- Various WAN Interfaces
- Many QoS features and interfaces

Examples: Cisco cat 6500

OC-n: Optical carrier n = 51.84Mbs

- From LAN to Access
- 48 to 576 10/100 Ethernet Interfaces
- 10 GigEth, OC-3, OC-12, OC-48, ATM
- QoS, Security
- Load Balancing; VPN
- Up to 128Gbps (with crossbar)
- L4-7 Switching
- IP Telephony (E1, T1, inline-power Ethernet)
- SNMP, RMON