Master Course
Computer Networks
IN2097

Prof. Dr.-Ing. Georg Carle
Christian Grothoff, Ph.D.
Stephan Günther

Chair for Network Architectures and Services
Department of Computer Science
Technische Universität München
http://www.net.in.tum.de

Outline

- Feedback on project VMs
  - own AS was not reachable, because of problems with own (open source) routers
  - who did not submit project Milestone 2 but want to continue using VMs, please send mail to Stephan Günter <guenther@net.in.tum.de>
- Exam

  The exam is scheduled for **Saturday, February 16, 2013, from 9:00 to 10:00am in MI HS1**. The exam will be **closed book**, i.e., no supplemental material is allowed (you won't even need a pocket calculator - but you should have a precision of 1E-03 built-in (;).
- Homework
  - Solution sketches
  - today: solution sketch of Homework 1 will be made available

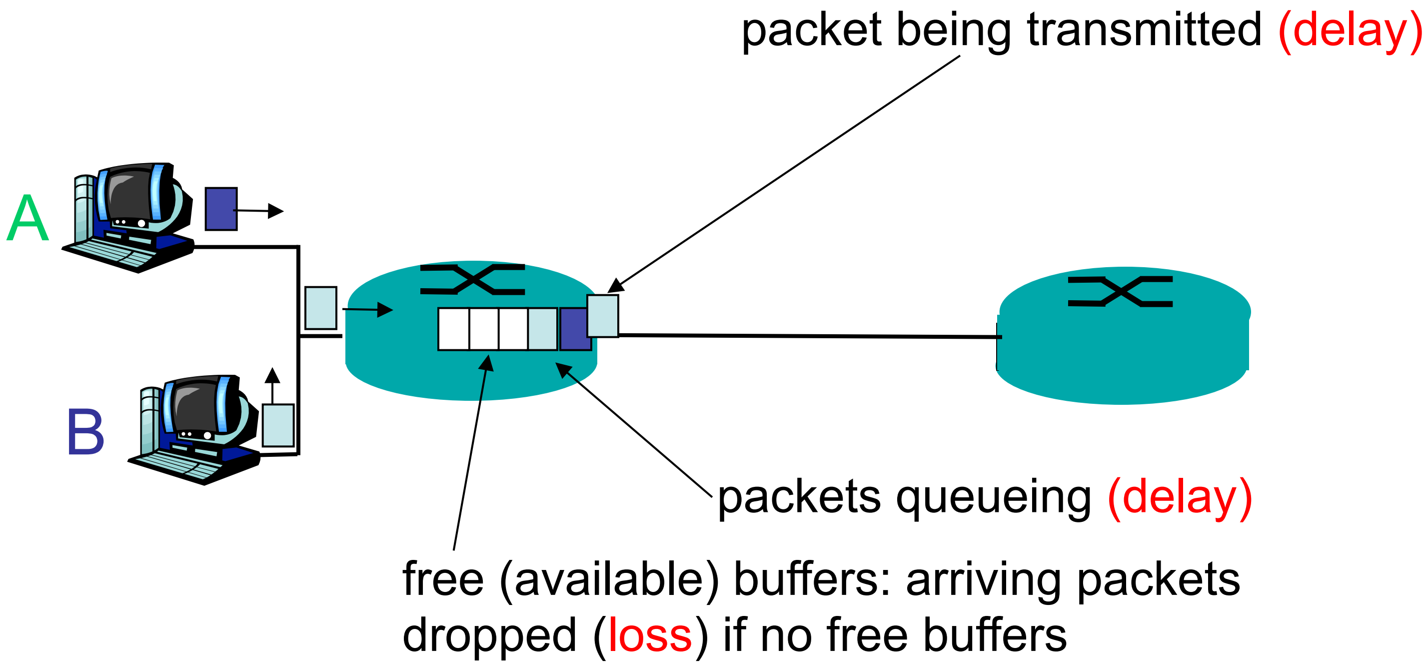
- Lecture
  - No lecture this week, Friday 11.1.2013
  - ⇨ time for you to work on the project

Node Forwarding Performance

Reasons for delay and loss

packets *queue* in router buffers

- packet arrival rate to link exceeds output link capacity
- packets queue, wait for turn



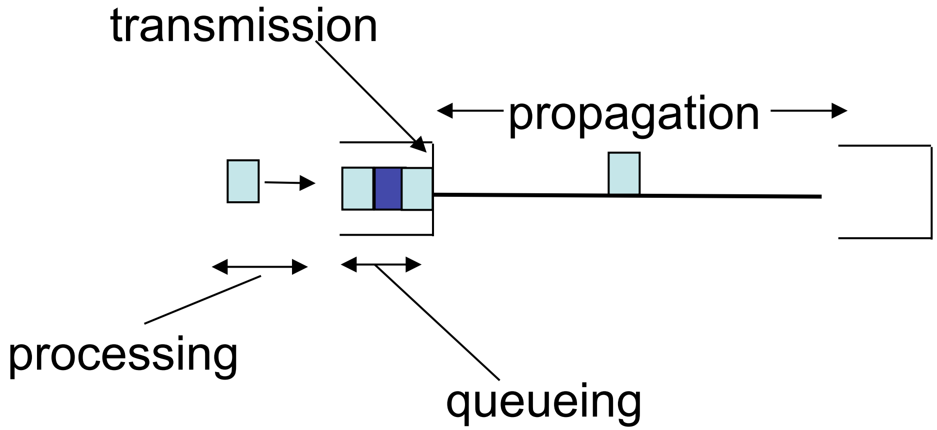
Background: Sources of packet delay

1. Processing delay:
   - Sending: prepare data for being transmitted
   - Receiving: interrupt handling

2. Queueing delay
   - time waiting at output link for transmission

3. Transmission delay:
   - L=packet length (bits)
   - R=link bandwidth (bps)
   - time to send bits into link = L/R

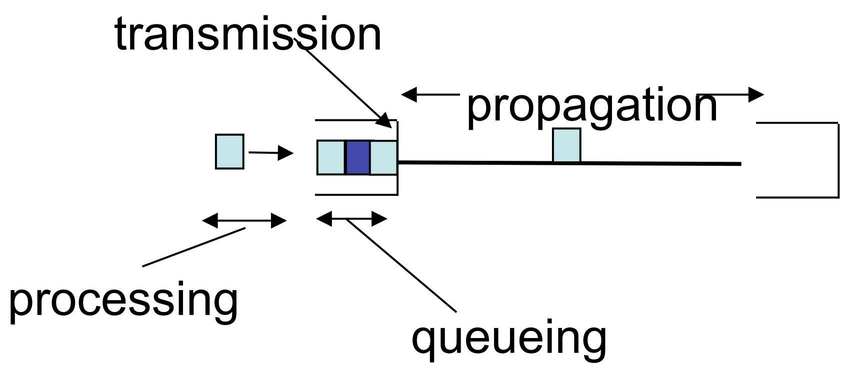
4. Propagation delay:
   - d = length of physical link
   - s = propagation speed in medium (~2x10^8 m/sec)
   - propagation delay = d/s



Nodal delay

- d_{proc} = processing delay
  - typically small - a few microseconds (μs) or less
- d_{queue} = queuing delay
  - depends on congestion - may be large
- d_{trans} = transmission delay
  - = L/R, significant for low-speed links
- d_{prop} = propagation delay
  - a few microseconds to hundreds of msecs

$$d_{nodal} = d_{proc} + d_{queue} + d_{trans} + d_{prop}$$



Impact Analysis: Advances in Network Technology

Data rate	Delay (1bit)	Length (1bit)	Delay (1kbyte)	Length (1kbyte)
1 Mbit/s	1 us	200 m	8 ms	1600 km
10 Mbit/s	100 ns	20 m	0,8 ms	160 km
100 Mbit/s	10 ns	2 m	80 us	16 km
1 Gbit/s	1 ns	0,2 m	8 us	1600 m
10 Gbit/s	100 ps	0,02 m	0,8 us	160 m
100 Gbit/s	10 ps	0,002 m	80 ns	16 m

- Assessment
  - Transmission delay becomes less important
    ⇨ over time; in the core
  - Distance becomes more important
    ⇨ matters for communication beyond data center
  - Network adapter latency less important
    ⇨ Latency of communication software becomes important

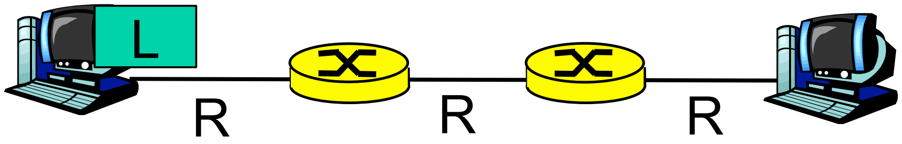
Propagation Delay

- Propagation speed: 2×10^8 m/sec
- Transmission of 625 byte (= 5000 bit): t= L/R=5000 / 1Gbit/s = 5 us

Distance	Propagation Delay	equivalent Transmission Delay (625 byte)	CPU cycles per packet (1 GHz)	CPU cycles per byte (1 GHz)
100 m	500 ns	10 Gbit/s	500	<1
1 km	5 us	1 Gbit/s	5.000	8
10 km	50 us	100 Mbit/s	50.000	80
100 km	500 us	10 Mbit/s		800
1.000 km	5 ms	1 Mbit/s		8.000
10.000 km	50 ms	100 Kbit/s		80.000

- Suggestion for home exercise: plot graphs

Store-and-Forward vs. Circuit Switching



- Transmission delay:
  L=packet length (bits)
  R=link bandwidth (bps)
  time to transmit packet of L bits on to link with R bps = L/R
- Store and forward: entire packet must arrive at router before it can be transmitted on next link:
- Total transmission delay = 3L/R

Example: Large Message L

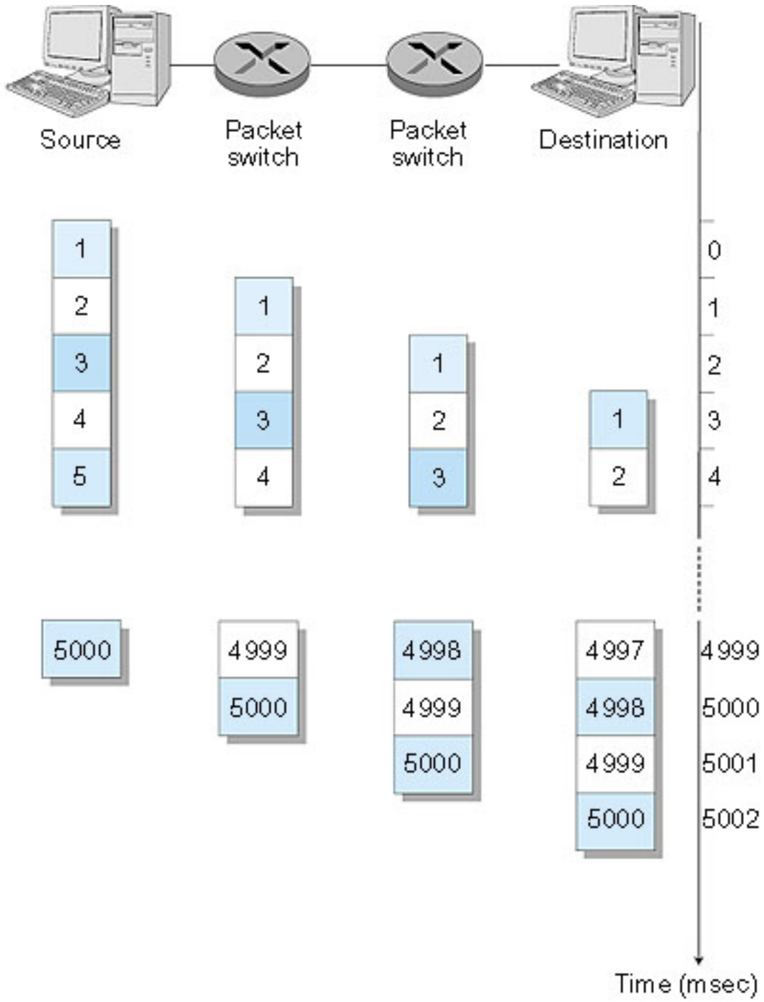
Store-and-Forward:

- L = 7.5 Mbit
- R = 1.5 Mbit/s
- Transmission delay = 15 s

Circuit Switching:

- L = 7.5 Mbit
- R = 1.5 Mbit/s
- Transmission delay = 5 s

Packet Switching: Message Segmenting



Now break up the message into 5000 packets

- Each packet 1,500 bits
- 1 msec to transmit packet on one link
- *pipelining:* each link works in parallel
- Delay reduced from 15 sec to 5.002 sec (as good as circuit switched)
- Advantages over circuit switching?
- Drawbacks (of packet vs. Message)

Discussion

- What is the role of header lengths?
- What is the role of header compression?
- What is the cost of tunneling?
- What are the benefits of overprovisioning?
- Can you „imagine“ a visualisation of packets being transmitted over different types of links?

Questions

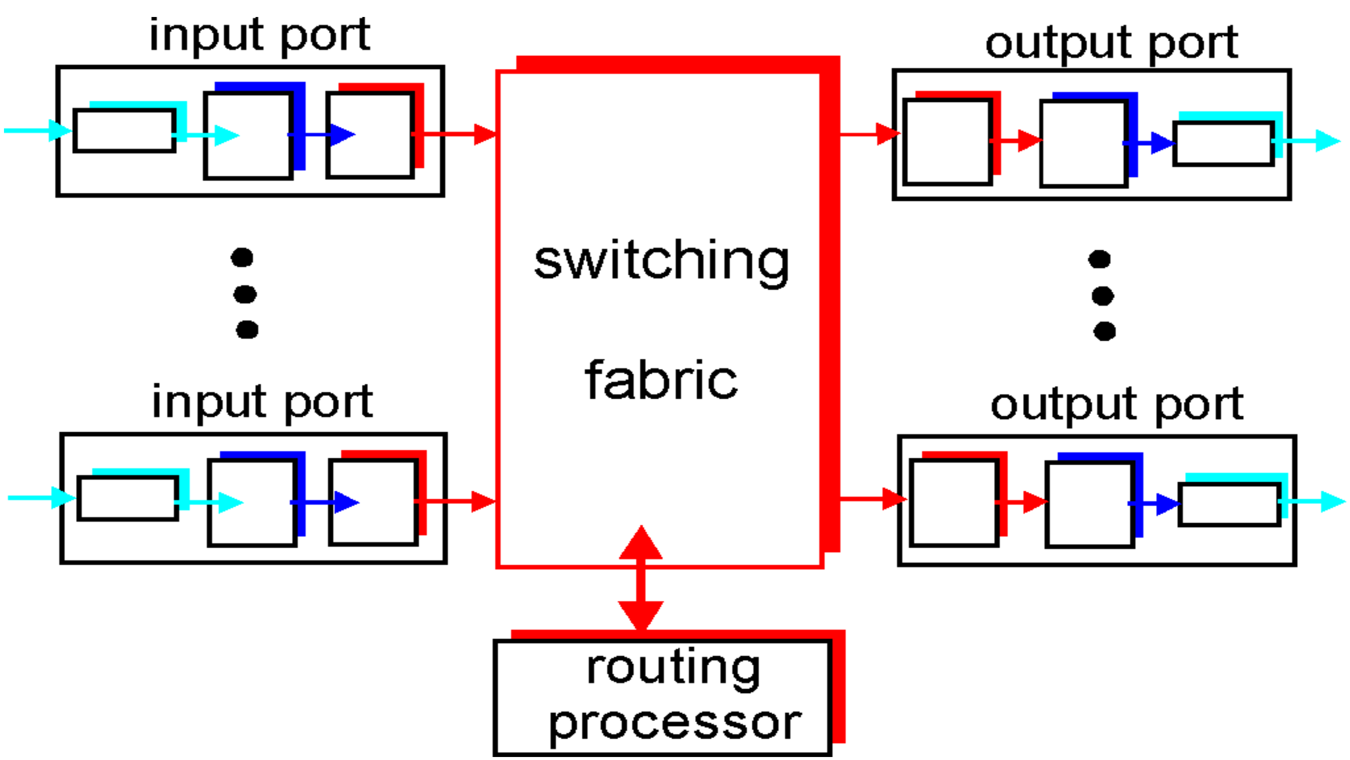
- Why/when is circuit switching expensive?
- Why/when is packet switching cheap?
- Is best effort packet switching able to carry voice communication?
- What happens if we introduce "better than best effort" service?
- How can we charge fairly for Internet services:
  by time, by volume, or flat?

Node Architectures

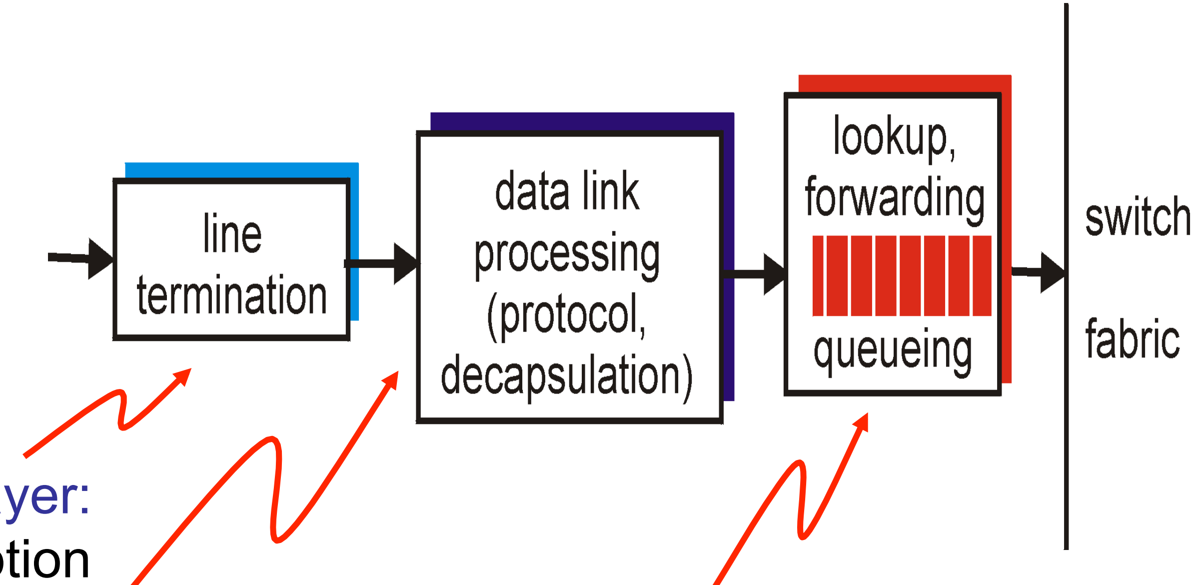
Router Architecture Overview

Two key router functions:

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



Input Port Functions



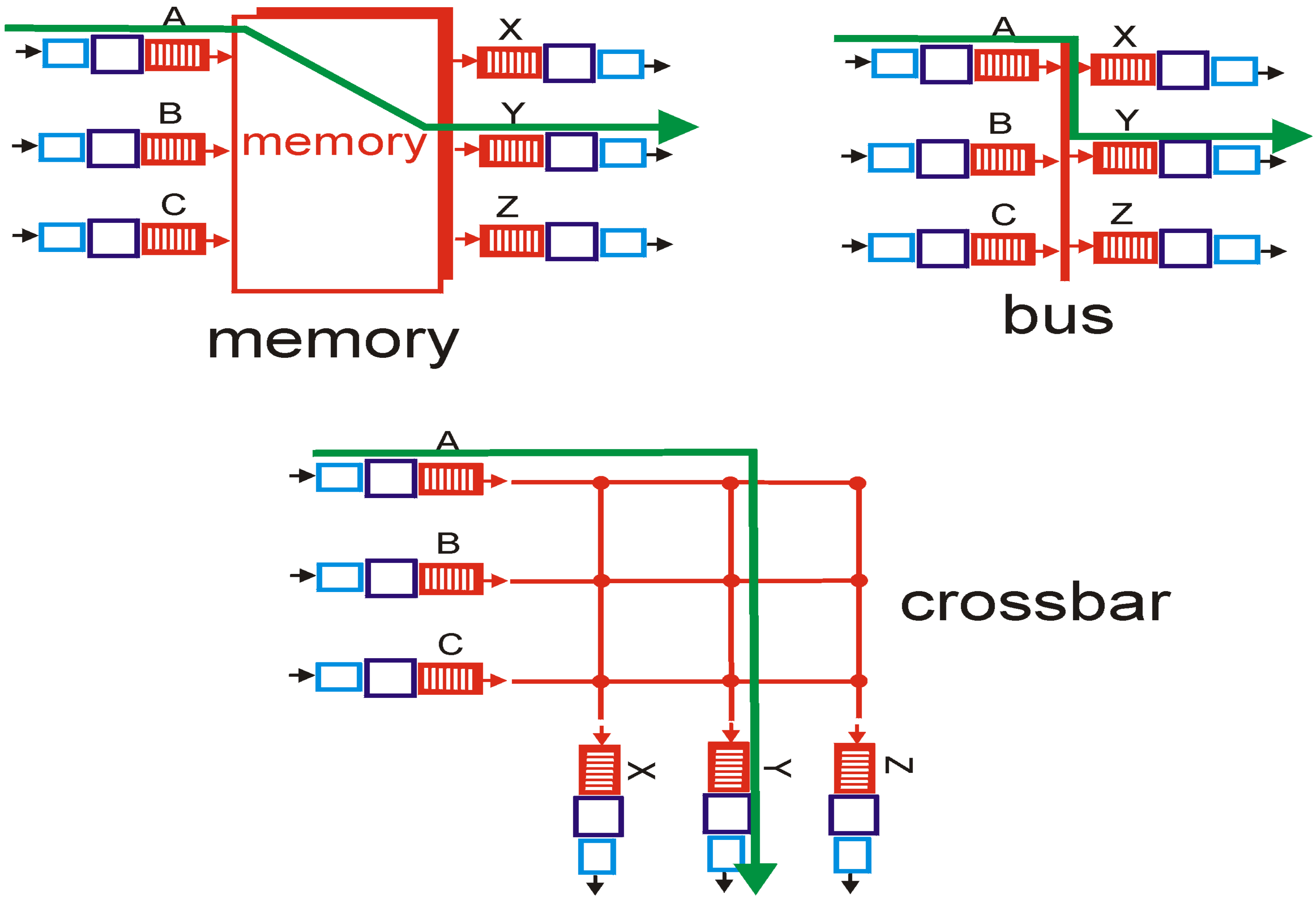
Physical layer:
bit-level reception

Data link layer:
e.g., Ethernet

Decentralized switching:

- given datagram destination, 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

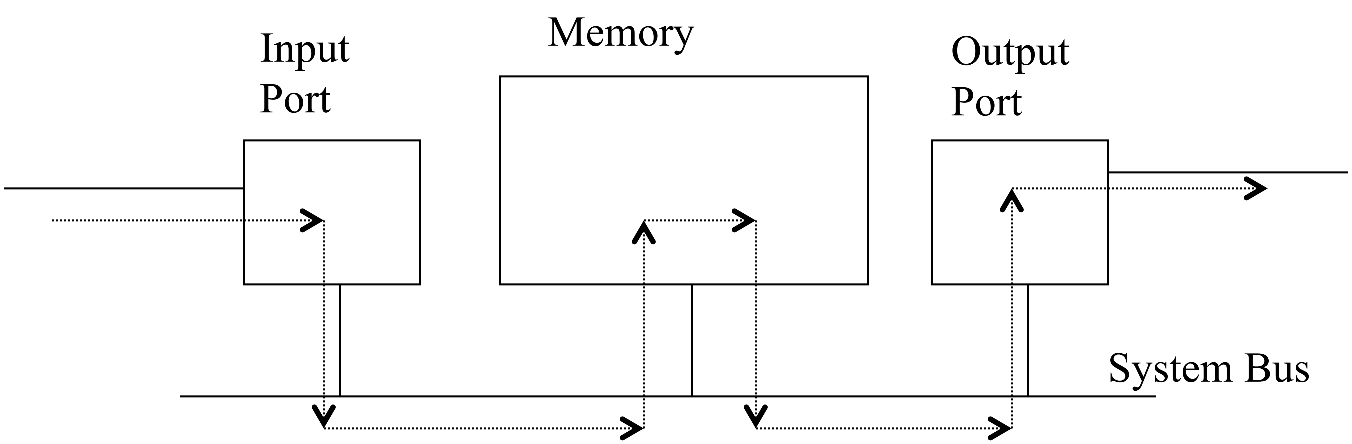
Three types of switching fabrics



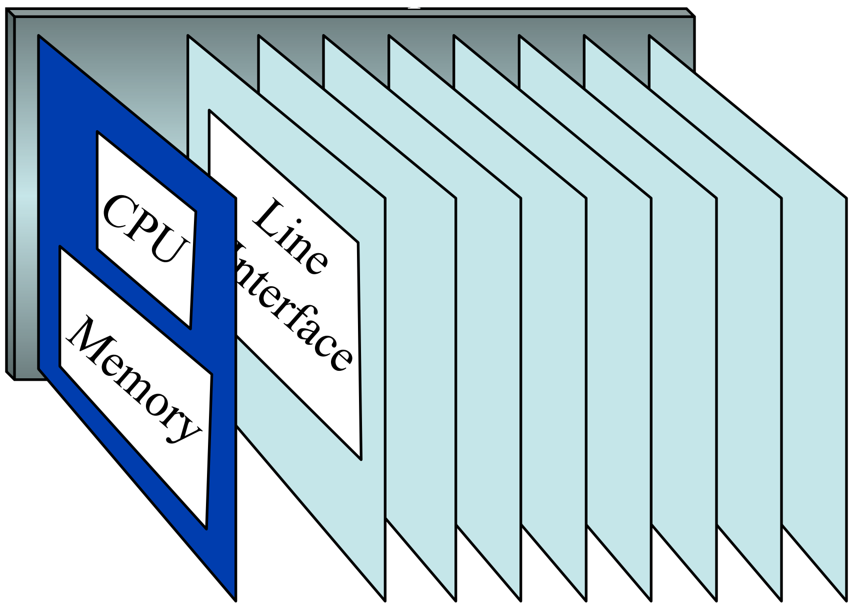
Switching Via Memory

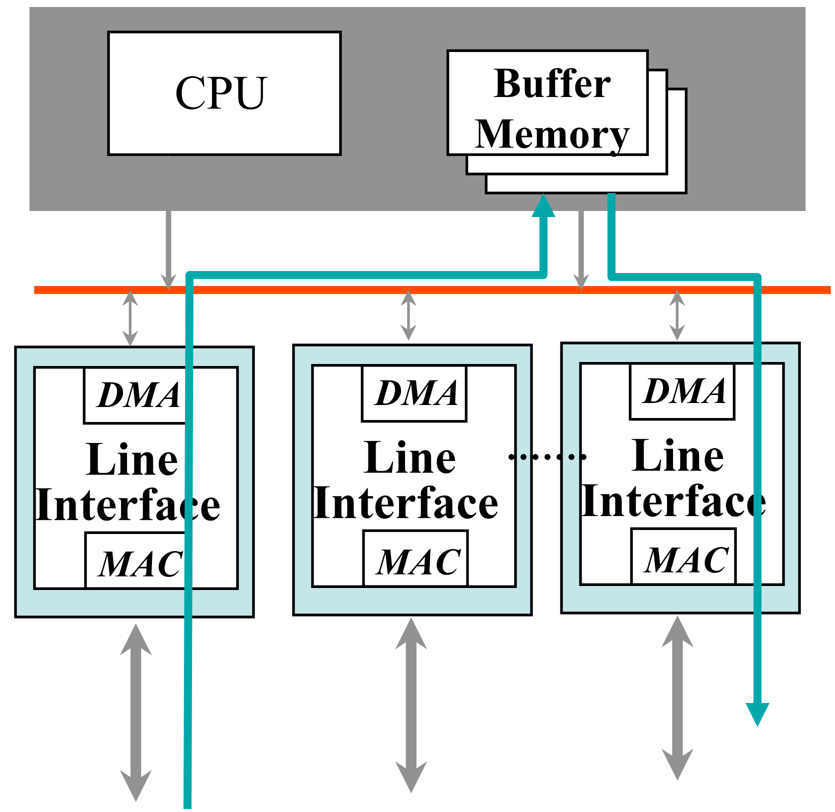
First generation IP routers:

- traditional computers with switching under direct control of CPU
- packet copied to system' s memory
- speed limited by memory bandwidth (2 bus crossings per datagram)

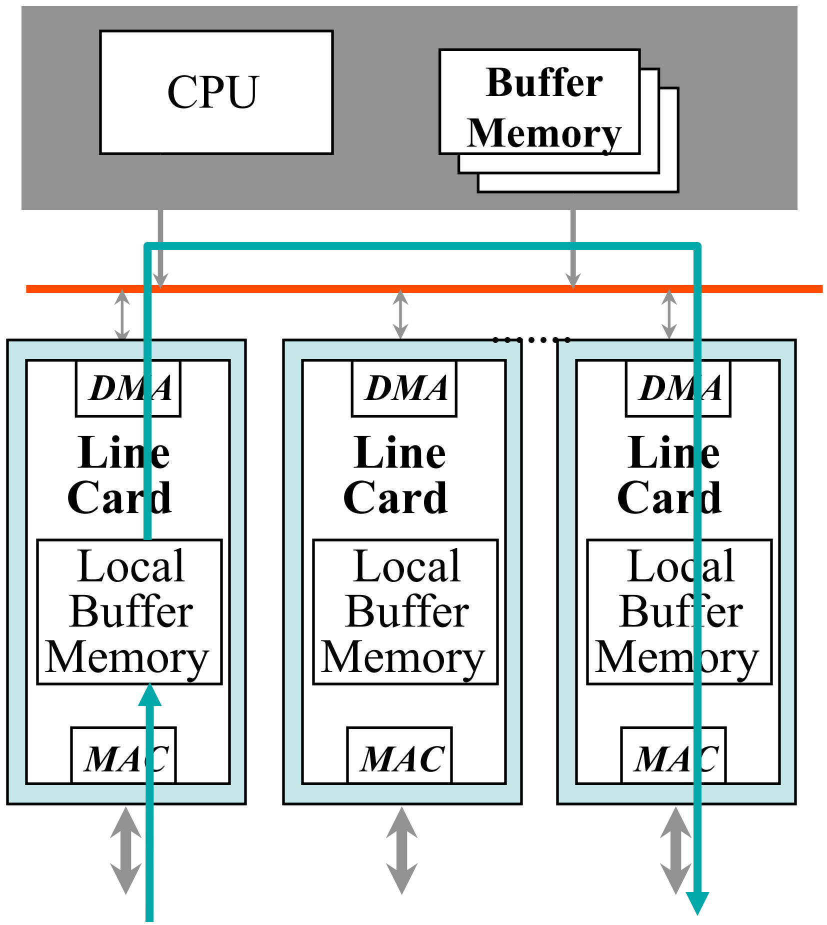


First-Generation IP Routers

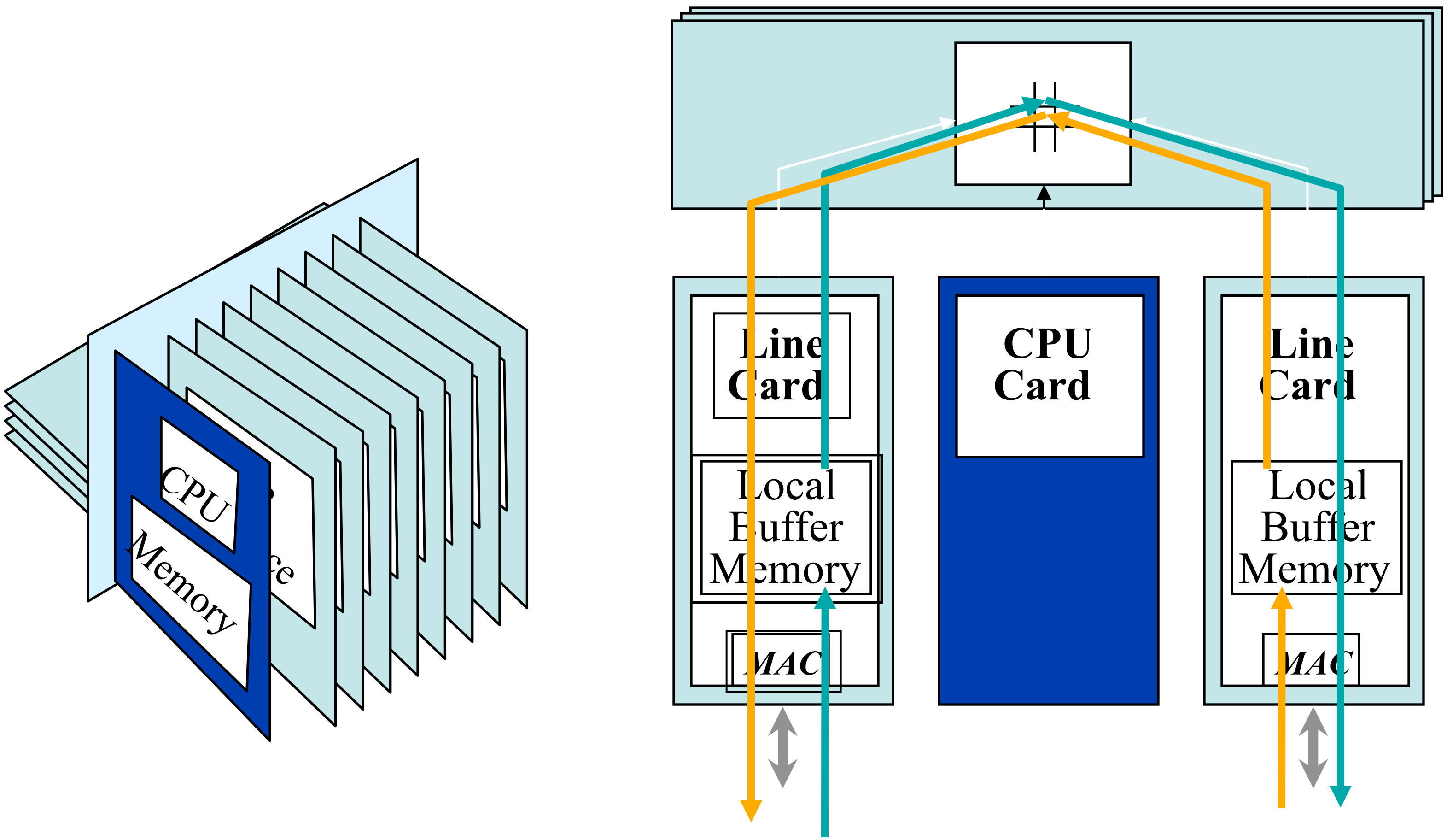




Second-Generation IP Routers

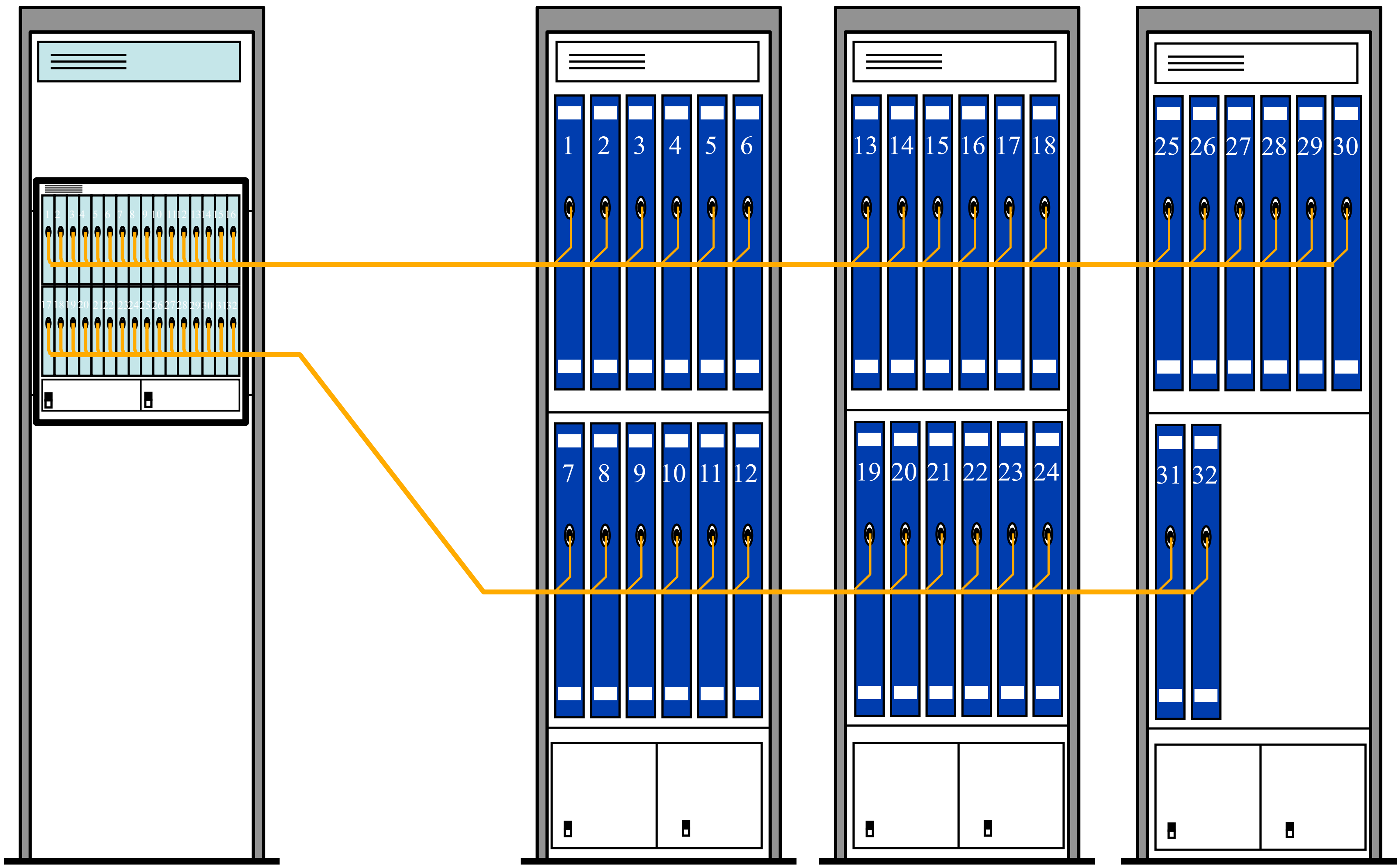


Third-Generation Switches/Routers



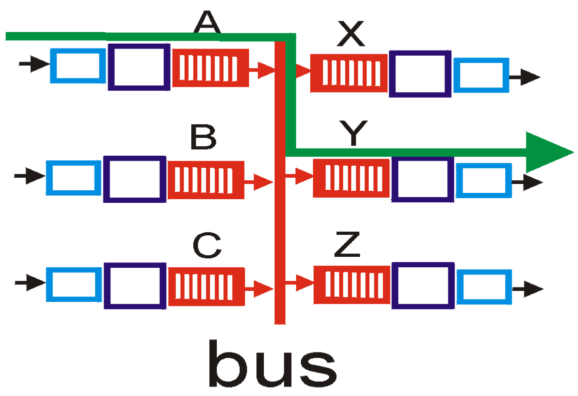
Fourth-Generation Switches/Routers

Clustering and Multistage



Switching Via a Bus

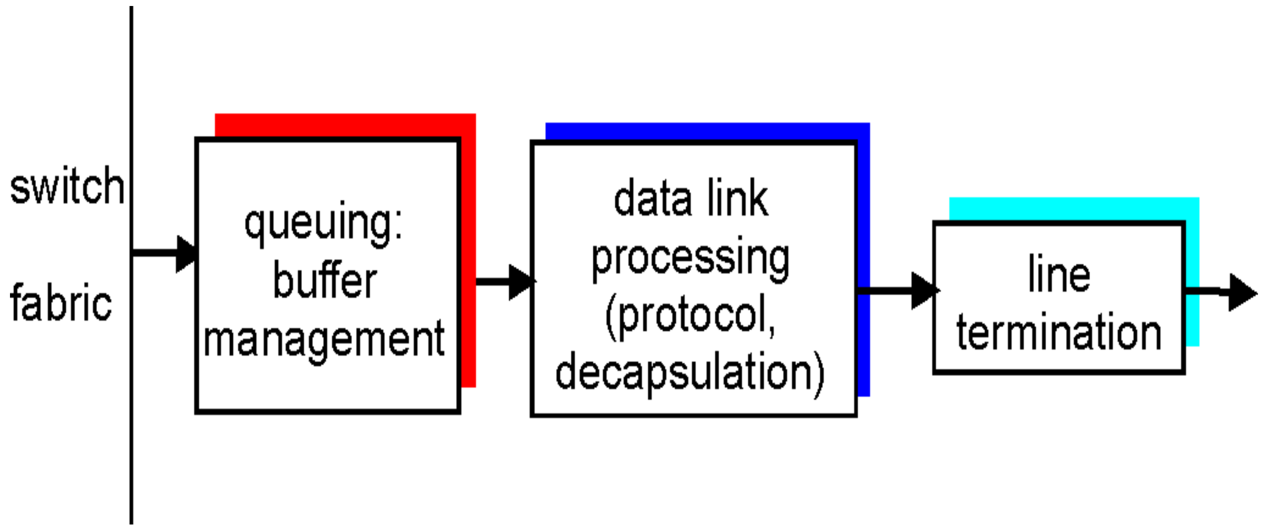
- datagram from input port memory to output port memory via a shared bus
- bus contention: switching speed limited by bus bandwidth
- 32 Gbps bus, Cisco 5600: sufficient speed for access and enterprise routers



Switching Via An Interconnection Network

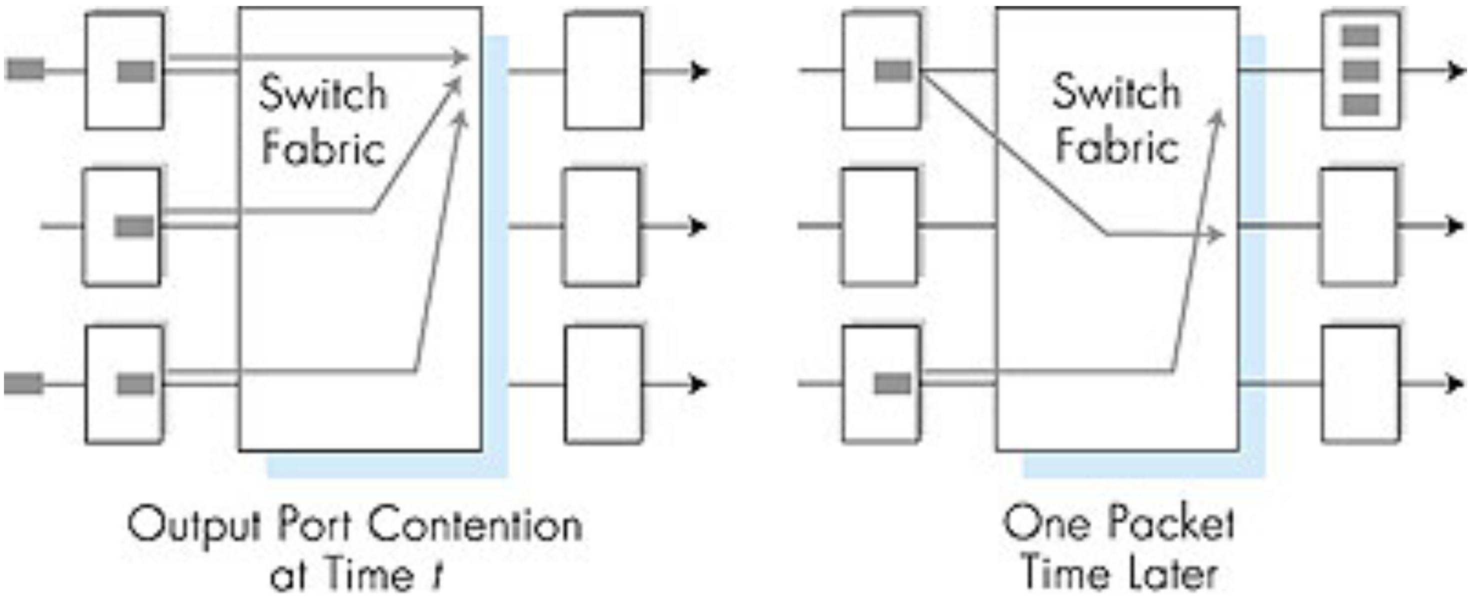
- overcome bus bandwidth limitations
- Banyan networks, other interconnection nets initially developed to connect processors in multiprocessor
- advanced design: fragmenting datagrams into fixed length cells, switch cells through the fabric.
- Cisco 12000: switches 60 Gbps through interconnection network

Output Ports



- *Buffering* required when datagrams arrive from fabric faster than the transmission rate
- *Scheduling discipline* chooses among queued datagrams for transmission

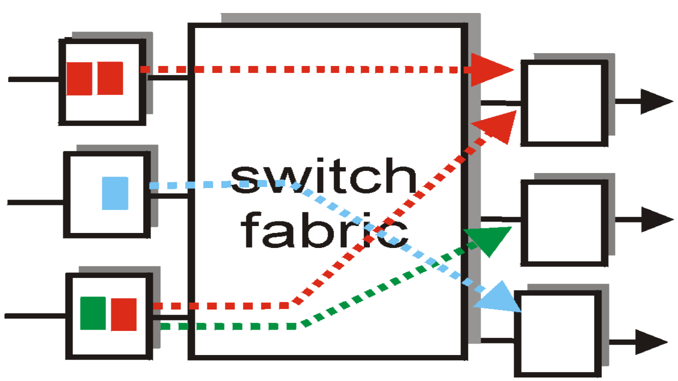
Output port queueing



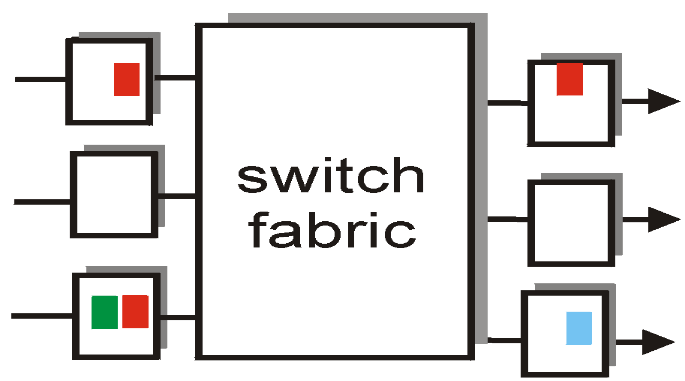
- buffering when arrival rate via switch exceeds output line speed
- *queueing (delay) and loss due to output port buffer overflow!*

Input Port Queuing

- Fabric slower than input ports combined ⇨ queueing may occur at input queues
- Head-of-the-Line (HOL) blocking: queued datagram at front of queue prevents others in queue from moving forward
- *queueing delay and loss due to input buffer overflow!*



output port contention
at time t - only one red
packet can be transferred



green packet
experiences HOL blocking

How much buffering?

- RFC 3439 (2002) rule of thumb: average buffering equal to "typical" RTT times link capacity C
  - e.g., RTT= 250 msec, C = 10 Gps link: 2.5 Gbit buffer
- More recent recommendation
  - Guido Appenzeller, Isaac Keslassy and Nick McKeown: Sizing Router Buffers, ACM SIGCOMM 2004
  - with *N* flows, buffering equal to

$$\frac{RTT \cdot C}{\sqrt{N}}$$

e.g., C = 10 Gps link: 2.5 Gbit buffer

100 flows ⇨ sqrt(N) = 10 ⇨ 250 Mbit buffer

1.000 flows ⇨ sqrt(N) ≈ 30 ⇨ ~ 100 Mbit buffer

10.000 flows ⇨ sqrt(N) =100 ⇨ 25 Mbit buffer