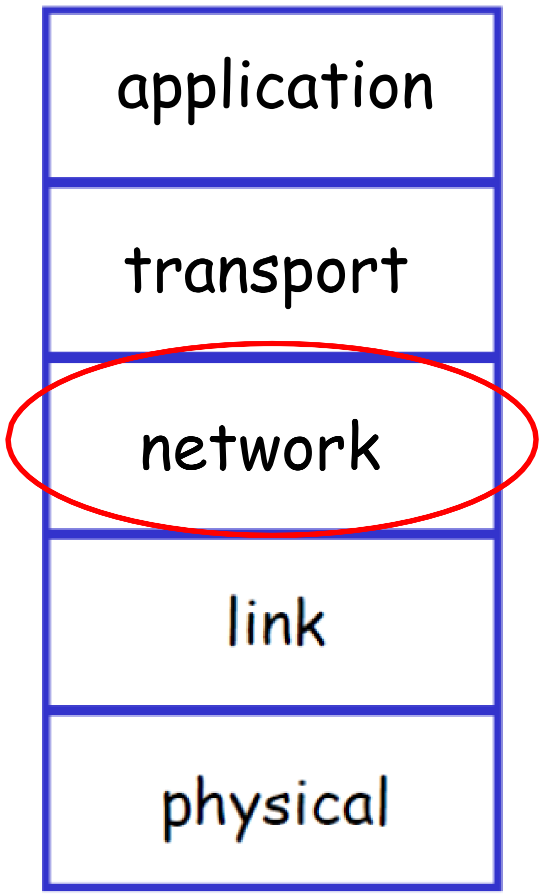
Packet-switching networks

Packet-switching networks



Packet-switching networks
Outline

- Context/overview
- Basic approaches to operating a packet-network: datagrams and virtual circuits
- Network layer functions: Routing and forwarding
- Overview of Network layer: data plane and control plane
- Network layer: The Data Plane
    - What's inside a router
    - The Internet Protocol (IPv4, DHCP, NAT, IPv6)
    - Generalized Forward and SDN
- Network layer: The Control Plane
    - Overview of routing in packet networks
    - The SDN control plane
    - ICMP: The Internet Control Message Protocol

Packet-switching networks

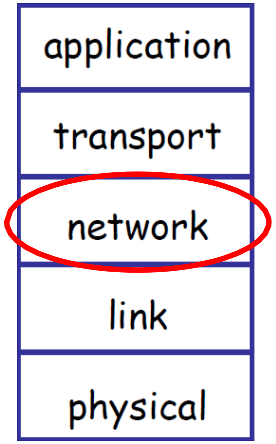
Context/Overview

- Packet switching networks and the network layer
- The network layer and its relevance to the network edge and the network core
- Packet switching networks, and how they differ from circuit-switching networks
- Applications and their link to and their use of packet switching networks

(Recall)

Internet protocol stack

Network Layer



- When transferring application messages, the transport layer provides process-to-process communication service whereas the network layer provides host-to-host communication service
- Transfers ***packets or datagrams*** from one host to another (passing through several routers!).
- The transport layer protocol (TCP or UDP) in a source host passes a segment and a destination address (got from the user inputs) to the network layer. The network layer then provides the service of delivering the segment to the transport layer in the destination host.
- The Internet's network layer provides "connectionless service" which is "best-effort connectionless packet transfer".
- The Internet"s network layer includes the key protocol, IP (Internet Protocol).
  - IP defines the fields in the packet/datagram and how the end systems and routers act on these fields.
  - There is only one IP protocol and all Internet components which have a network layer must run the IP protocol.
  - The Internet"s network layer also contains „routing protocols" that determine the routes/paths that packets/datagrams take between sources and destinations.
  - Although the network layer contains both the IP protocol and many routing protocols, it is often simply referred to as the IP layer (since **IP is the one which glues different networks in the Internet together!**)

Note: The term "packets" might, sometimes, be used to refer to any data chunks.

Packet-switching networks
Outline

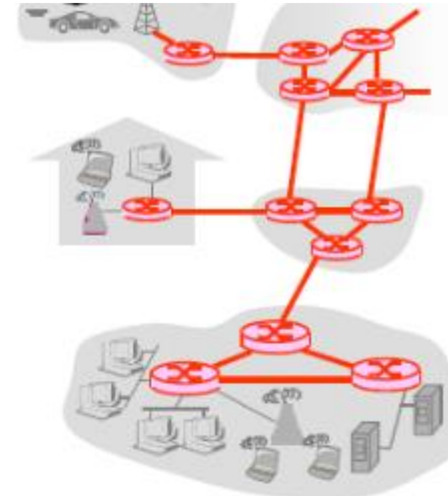
- Context/overview
- Basic approaches to operating a packet-network: datagrams and virtual circuits
- Network layer functions: Routing and forwarding
- Overview of Network layer: data plane and control plane
- Network layer: The Data Plane
    - What's inside a router
    - The Internet Protocol (IPv4, DHCP, NAT, IPv6)
    - Generalized Forward and SDN
- Network layer: The Control Plane
    - Overview of routing in packet networks
    - The SDN control plane
    - ICMP: The Internet Control Message Protocol

(Recall)

Network Core: Packet Switching

each end-end data stream divided into *packets*

- user A, B packets *share* network resources
- each packet uses full link bandwidth
- resources used *as needed*

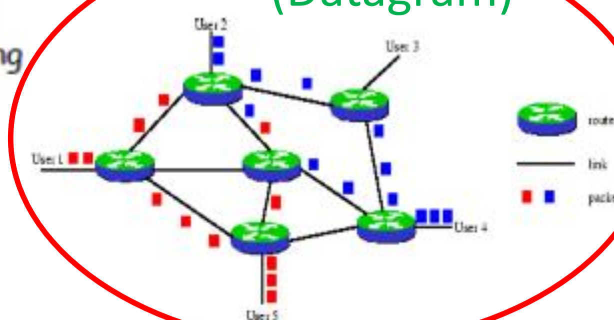


resource contention:

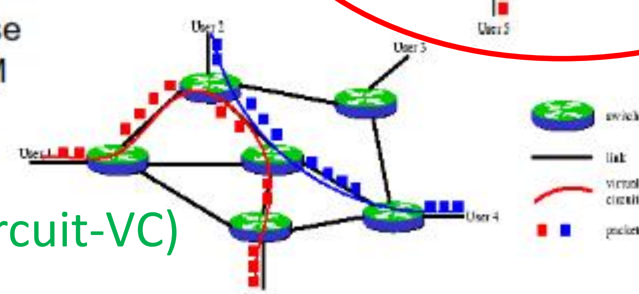
- aggregate resource demand can exceed amount available
- congestion: packets queue, wait for link use
- store and forward: packets move one hop at a time
  - Node receives complete packet before forwarding

~~Bandwidth division into "pieces"~~
~~Dedicated allocation~~
~~Resource reservation~~

(Datagram)



- Could be **connection-less service (IP networks)** (Different packets of the same message may traverse different routes) or connection oriented service (ATM networks)



(Virtual circuit-VC)

- The Internet is a datagram network

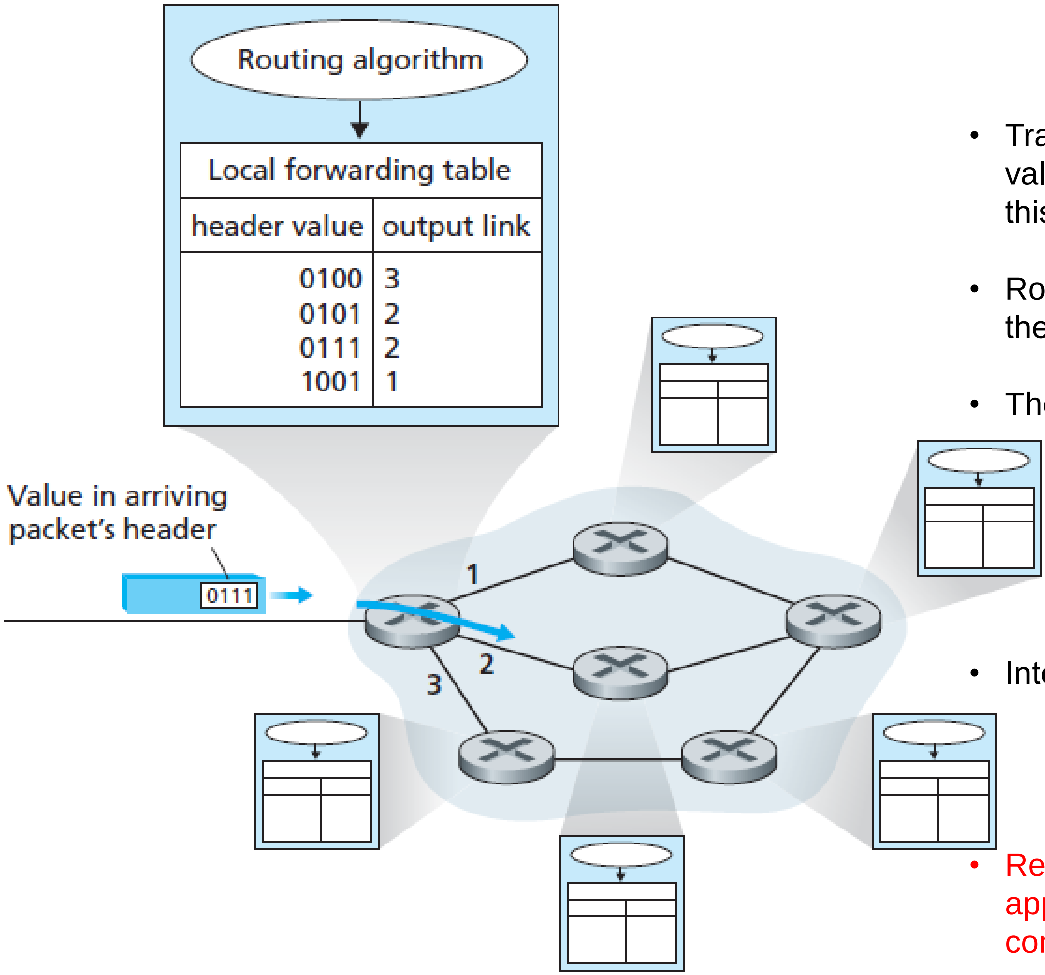
Two Key Network-Layer Functions

- **forwarding:** move packets from router's input to appropriate router output (router-local action)

- **routing:** determine routes/paths taken by packets from source to dest. (network-wide process).
  - 'routing algorithms' calculate these paths

(another important function in ‘virtual circuit networks’: ‘connection setup’)

Interplay between routing and forwarding



- Every router has a forwarding table
- 'Forwarding table' and 'routing table' are sometimes used interchangeably (there are some differences).
  - Routing tables are generally not used directly for packet forwarding in modern router architectures; instead, they are used to generate the information for a smaller forwarding table
- Traditionally, a router forwards a packet by examining the value of a field in its header (address/identifier), and then using this header value to index into the router's forwarding table.
- Routing algorithm determines the values that are inserted into the routers' forwarding tables.
- The routing algorithm may be
  - centralized (e.g., with an algorithm executing on a central site and downloading routing information to each of the routers) or
  - decentralized (i.e., with a piece of the distributed routing algorithm running in each router).
- Internet's routing algorithms/protocols (distributed):
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - BGP: Border Gateway Protocol
- Recent development: the Software Defined Network (SDN) approach uses a logically centralized controller (server) to compute routes

Packet-switching networks
Outline

- Context/overview
- Basic approaches to operating a packet-network: datagrams and virtual circuits
- Network layer functions: Routing and forwarding
- Overview of Network layer: data plane and control plane
- Network layer: The Data Plane
    - What’s inside a router
    - The Internet Protocol (IPv4, DHCP, NAT, IPv6)
    - Generalized Forward and SDN
- Network layer: The Control Plane
    - Overview of routing in packet networks
    - The SDN control plane
    - ICMP: The Internet Control Message Protocol

Network layer: data plane, control plane

Data plane

- local, per-router function
- determines how datagram arriving on router input port is forwarded to router output port
- forwarding function

Control plane

- network-wide logic
- determines how datagram is routed among routers along end-end path from source host to destination host (+ other management)

Two approaches to structuring network control plane:

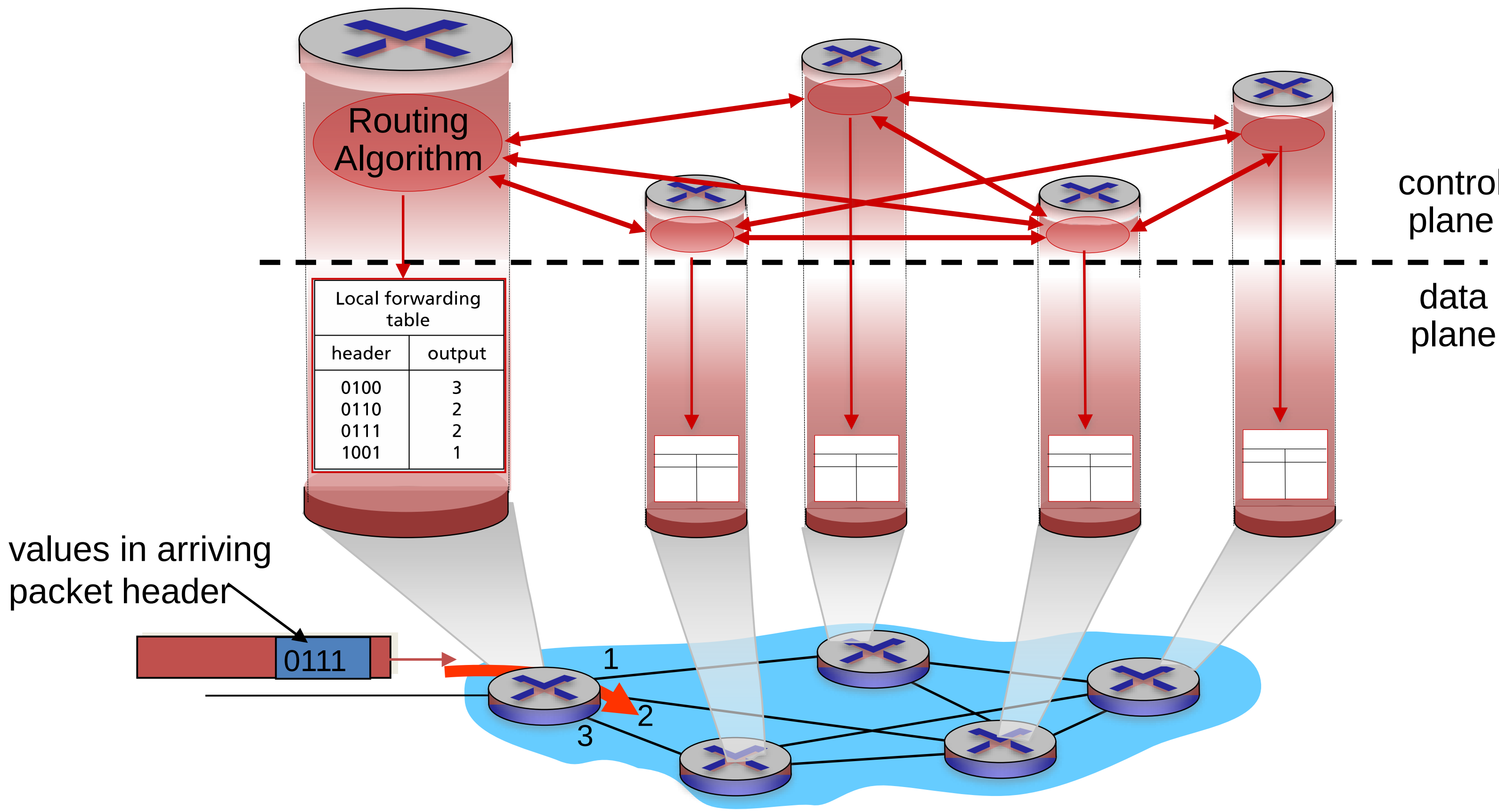
- per-router control (traditional)- implemented in routers
- logically centralized control (software defined networking-SDN)- implemented in (remote) servers

Recall: two network-layer functions:

- *forwarding:* move packets from router's input to appropriate router output — *data plane*
- *routing:* determine route taken by packets from source to destination — *control plane*

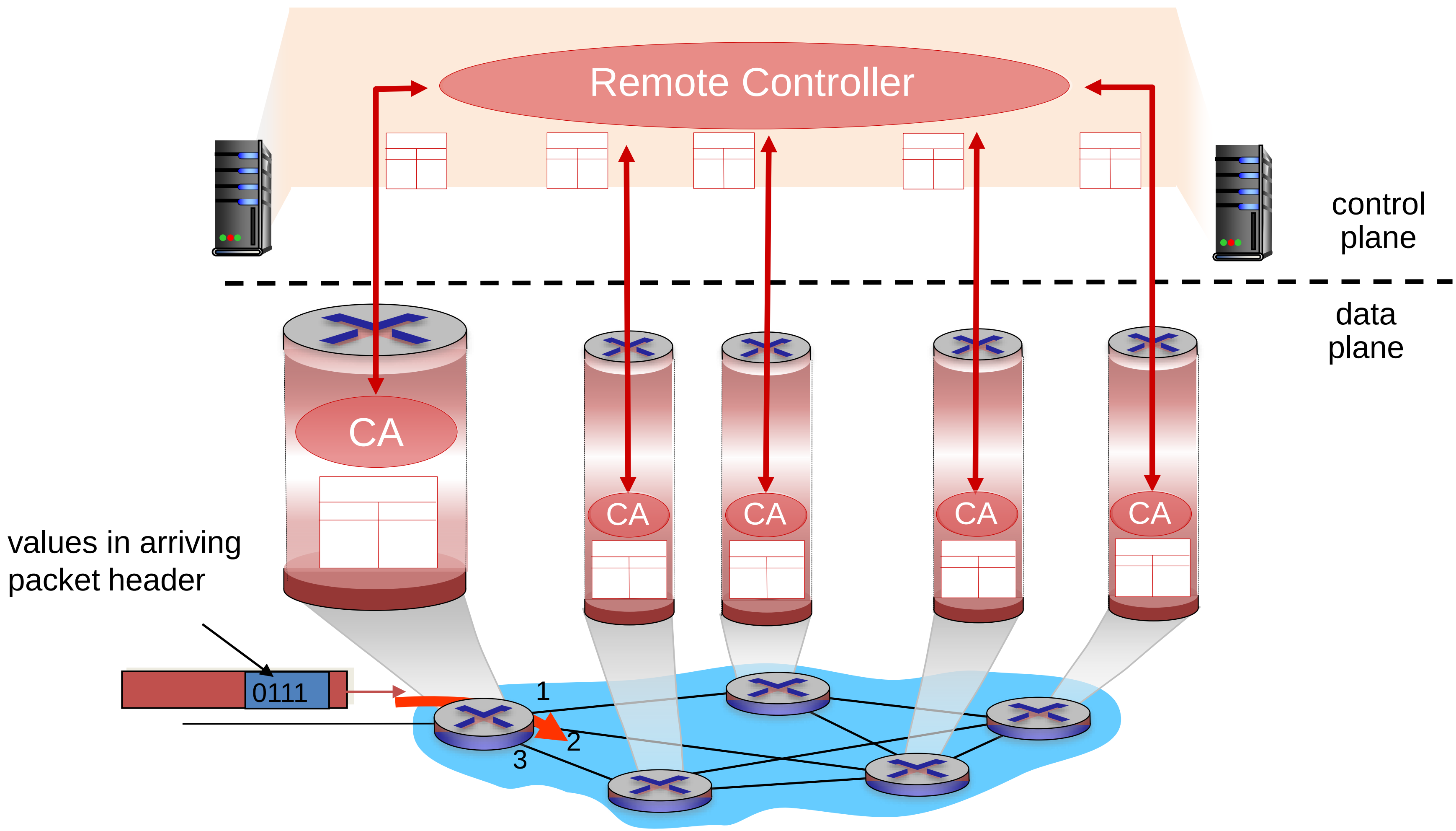
Per-router control plane

Individual routing algorithm components *in each and every router* interact in the control plane



Logically centralized control plane

A distinct (typically remote) controller interacts with local control agents (CAs)



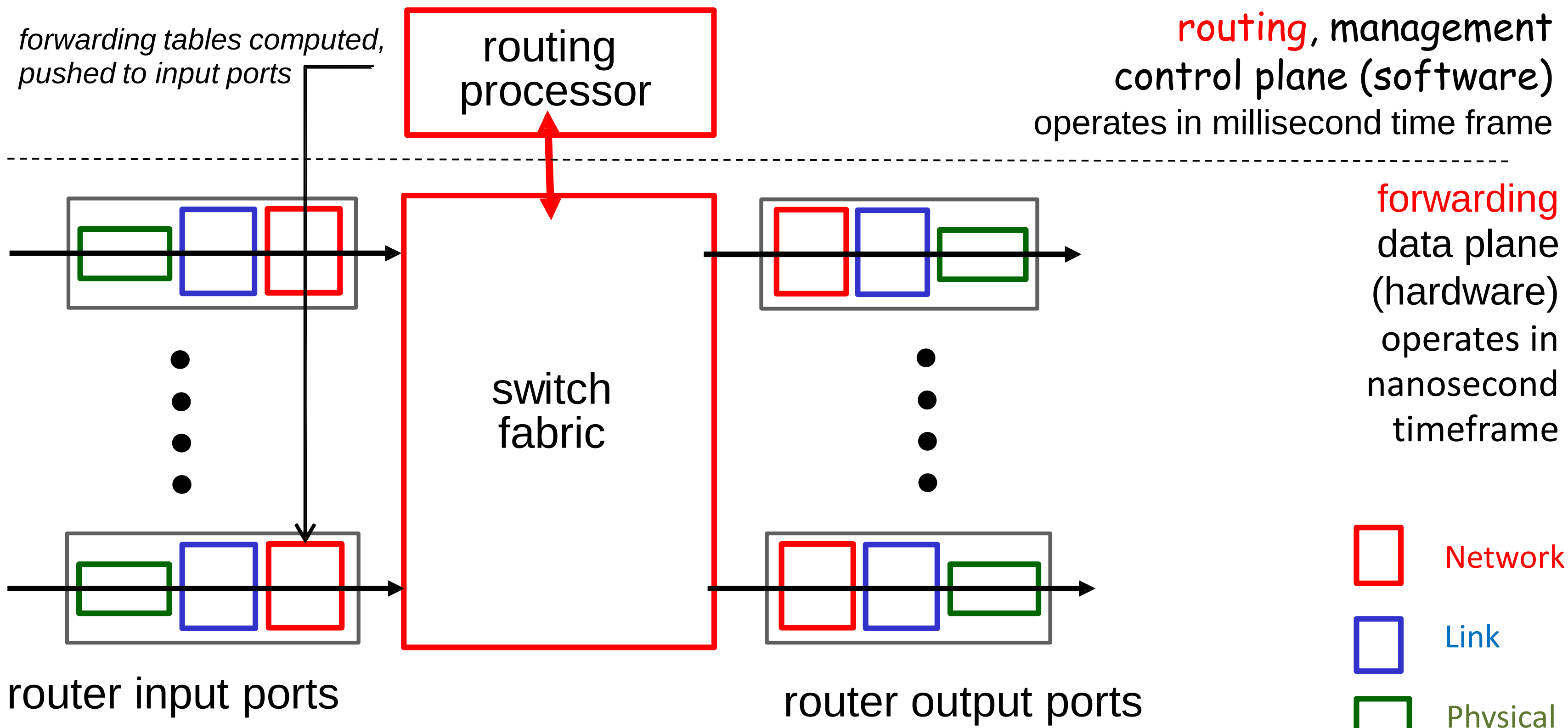
Packet-switching networks
Outline

- Context/overview
- Basic approaches to operating a packet-network: datagrams and virtual circuits
- Network layer functions: Routing and forwarding
- Overview of Network layer: data plane and control plane
- Network layer: The Data Plane
    - What's inside a router
    - The Internet Protocol (IPv4, DHCP, NAT, IPv6)
    - Generalized Forward and SDN
- Network layer: The Control Plane
    - Overview of routing in packet networks
    - The SDN control plane
    - ICMP: The Internet Control Message Protocol

Router architecture overview

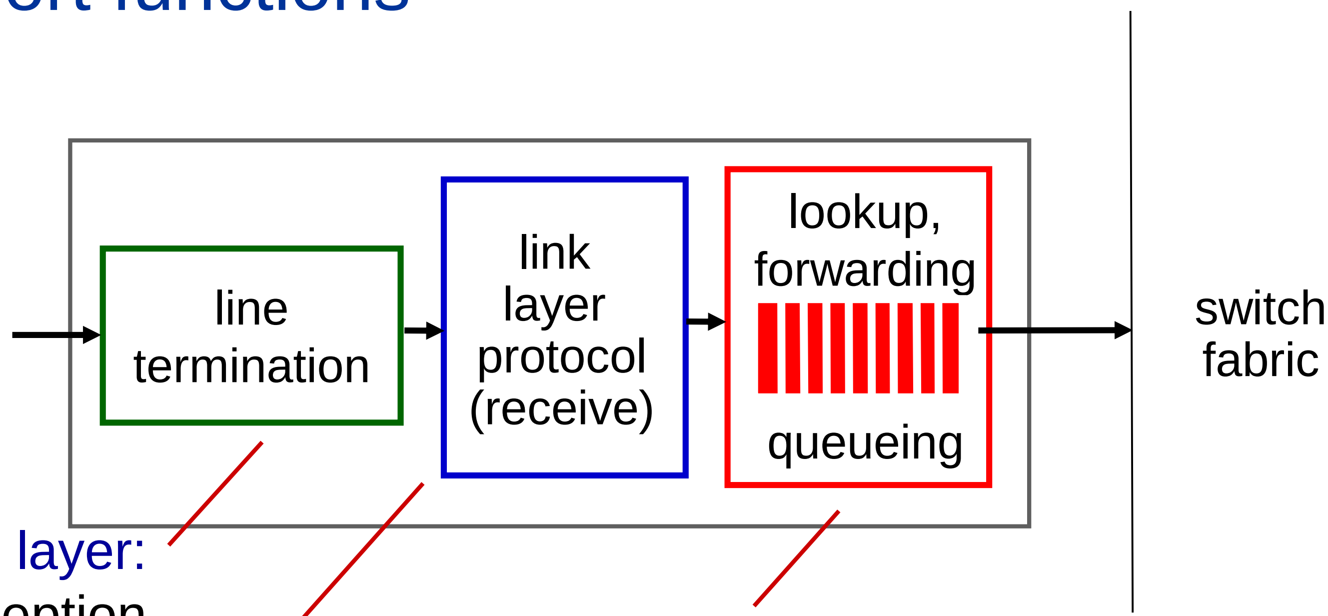
two key router functions:

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



Here, we learn more about 'forwarding' or 'switching'

Input port functions



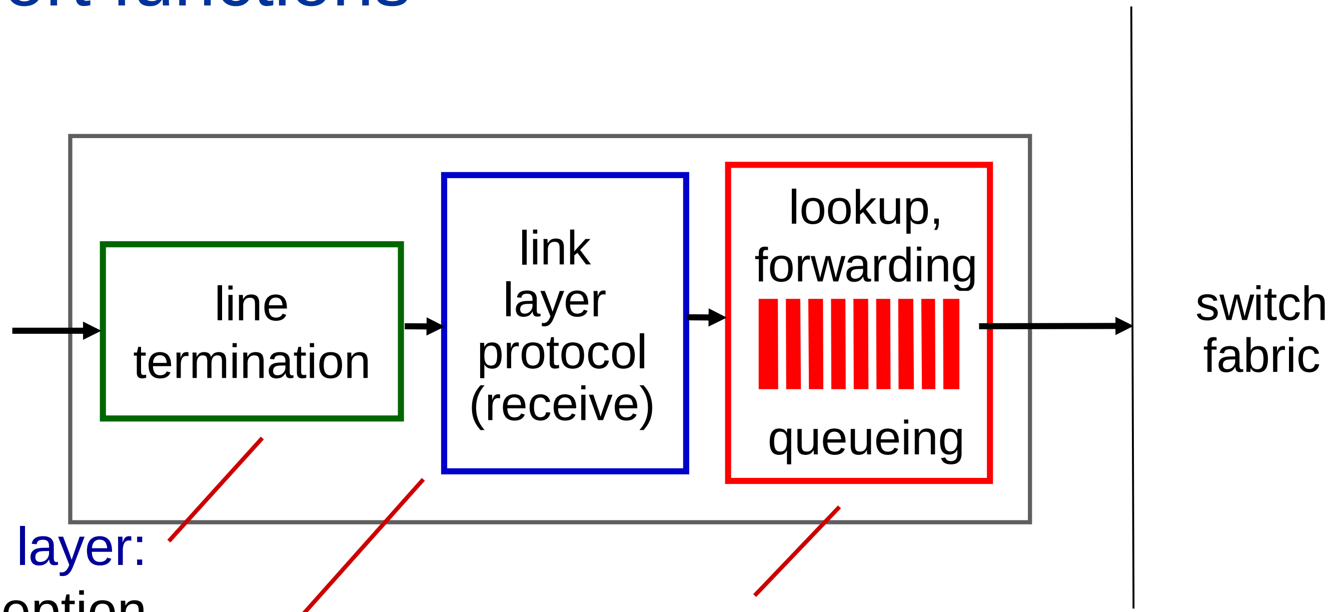
physical layer:
bit-level reception

data link layer:
e.g., Ethernet

decentralized switching*:*

- using header field values, lookup output port using forwarding table in input port memory *("match plus action")*
- goal: complete input port processing at 'line speed'
- queuing: if datagrams arrive faster than forwarding rate into switch fabric
- Other functions: checking the packet's version number, checksum (rewrite) and time-to-live field (rewrite). Update counters used for network management

Input port functions



physical layer:
bit-level reception

data link layer:
e.g., Ethernet

decentralized switching*:*

- using header field values, lookup output port using forwarding table in input port memory *("match plus action")*
- *destination-based forwarding:* forward based only on destination IP address (traditional)
- *generalized forwarding:* forward based on any set of header field values

Destination-based forwarding

Destination Address Range	Link Interface
11001000 00010111 00010000 00000000 through 11001000 00010111 00010111 11111111	0
11001000 00010111 00011000 00000000 through 11001000 00010111 00011000 11111111	1
11001000 00010111 00011000 00000000 through 11001000 00010111 00011111 11111111	2
otherwise	3

Longest prefix matching

Prefix	Link Interface
11001000 00010111 00010	0
11001000 00010111 00011000	1
11001000 00010111 00011	2
otherwise	3

A packet arrives at a router with the destination address (DA):

DA: 11001000 00010111 00010110 10100001 Which interface?

DA: 11001000 00010111 00011000 10101010 Which interface?

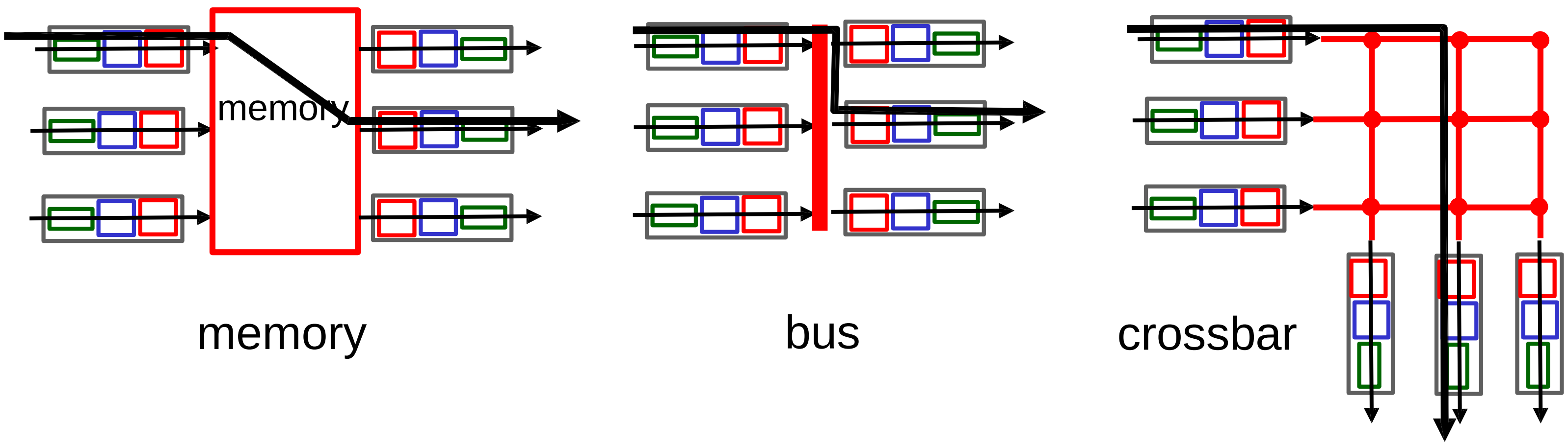
When there are multiple matches, the router uses the **longest prefix matching rule**; it finds the longest matching entry in the table and forwards the packet to the link interface associated with the longest prefix match

A forwarding table in an Internet core router might have more than 400,000 IP prefixes

- Longest prefix matching: often performed using ternary content addressable memories (TCAMs)
  - *content addressable:* present address to TCAM, retrieve address in one clock cycle, regardless of table size
  - Cisco Catalyst: can up ~1M routing table entries in TCAM

Switching fabrics

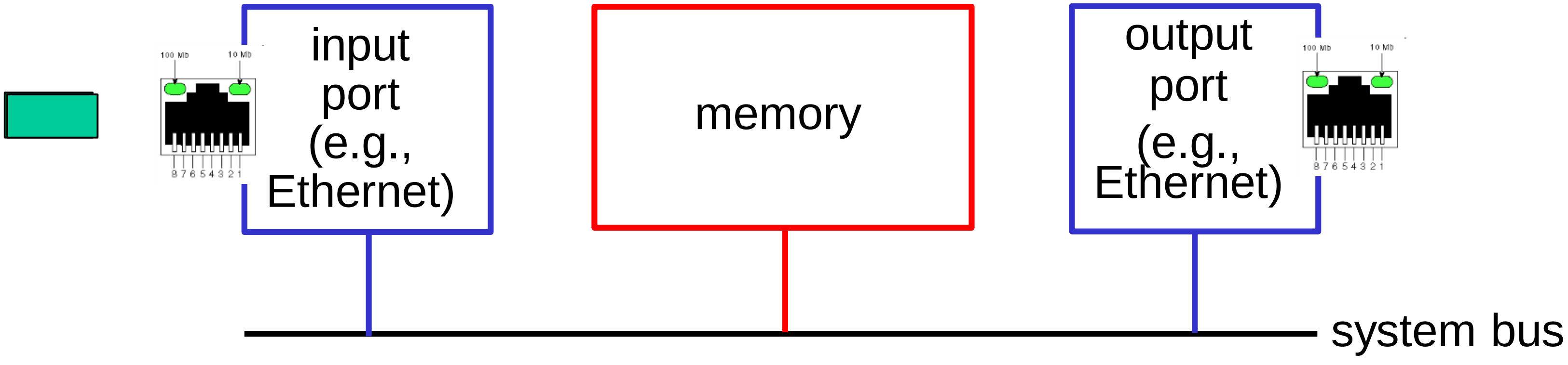
- transfer packet from input (buffer) to appropriate output (buffer)
- switching rate: rate at which packets can be transferred from inputs to outputs
  - often measured as multiple of input/output line rate
  - N inputs: switching rate N times line rate desirable
- three types of switching fabrics



Switching via memory

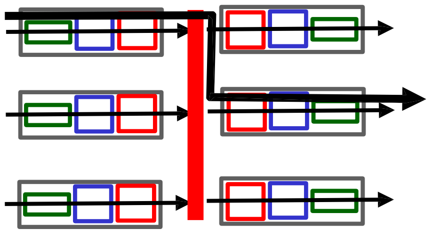
first generation routers:

- 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)



Switching via a bus

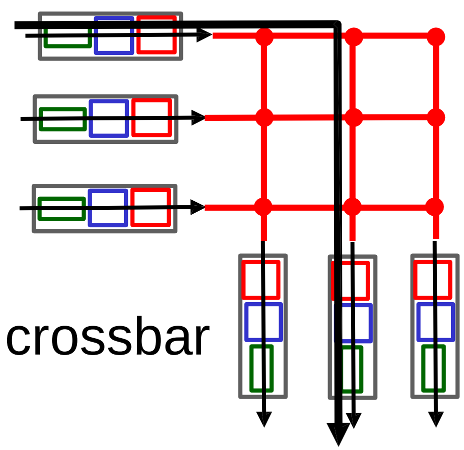
- datagram from input port memory to output port memory via a shared bus
- bus contention: switching speed limited by bus bandwidth
- 32 Gbps bus, Cisco 5600: sufficient speed for access and enterprise routers



bus

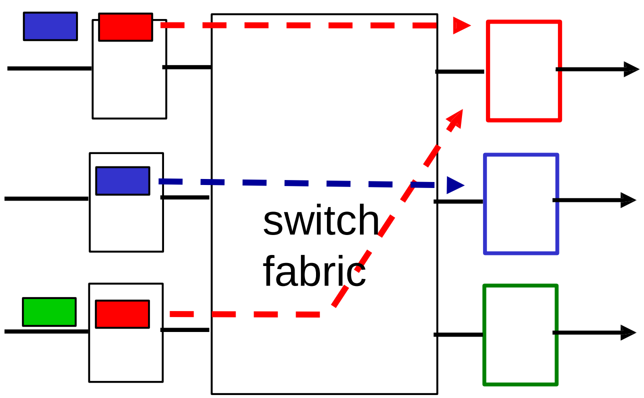
Switching via interconnection network

- overcome bus bandwidth limitations
- banyan networks, crossbar, 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 60 Gbps through the interconnection network

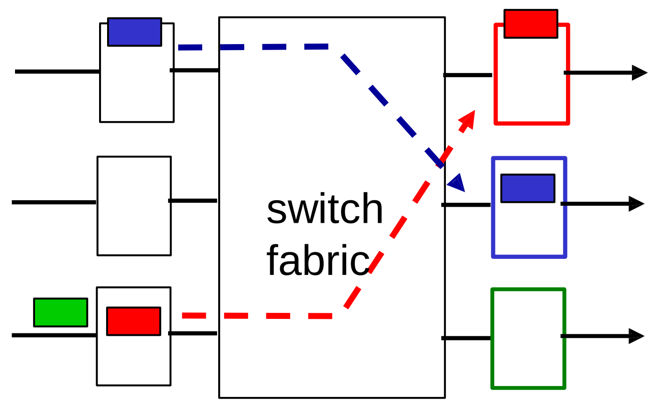


Input port queuing

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

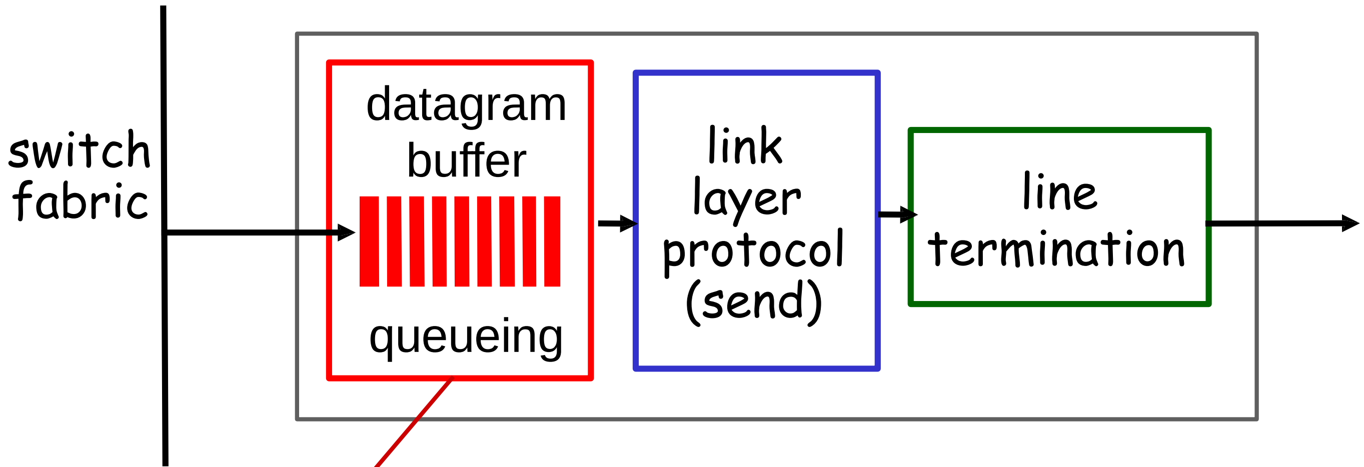


output port contention:
only one red datagram can be transferred.
lower red packet is blocked



one packet time later:
green packet experiences HOL blocking

Output ports

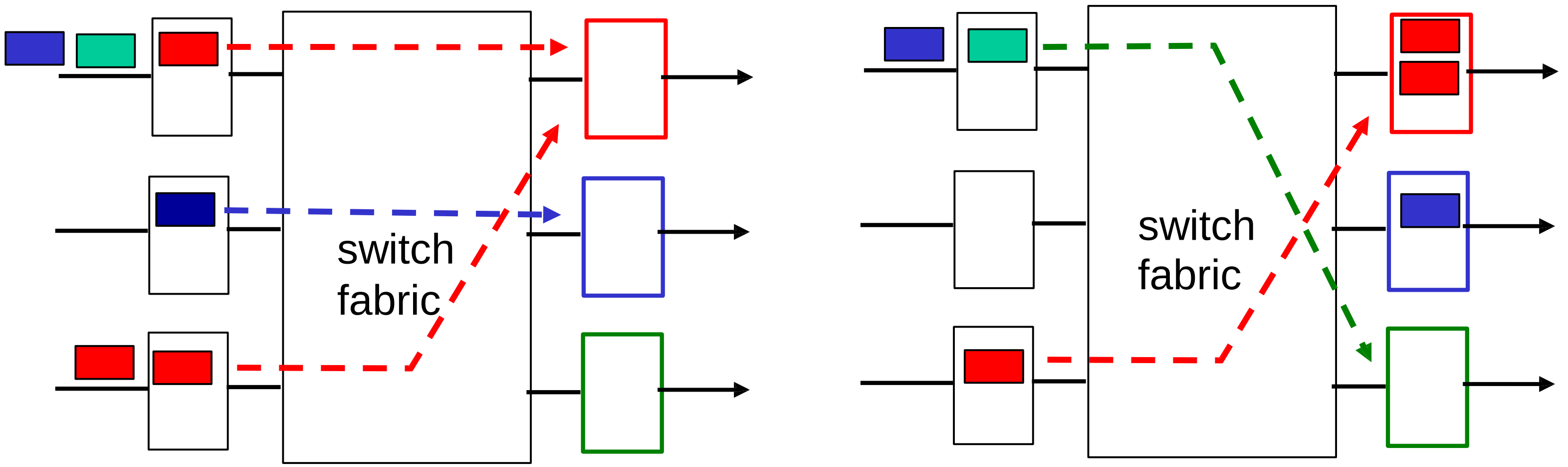


- buffering required when datagrams arrive from fabric faster than the transmission rate
- scheduling discipline chooses among queued datagrams for transmission

Datagram (packets) can be lost due to congestion, lack of buffers

Priority scheduling – who gets best performance, network neutrality

Output port queueing



at *t,* packets move
from input to output

one packet time later

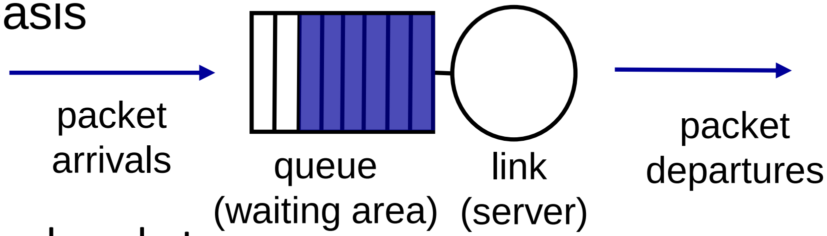
- buffering when arrival rate via switch exceeds output line speed
- queuing (delay) and loss due to output port buffer overflow!

How much buffering?

- RFC 3439 rule of thumb: average buffering equal to "typical" RTT (say 250 msec) times link capacity C
  - e.g., C = 10 Gpbs link: 2.5 Gbit buffer
- recent recommendation: with N flows, buffering equal to $\dfrac{\text{RTT} \cdot C}{\sqrt{N}}$

Scheduling mechanisms

- *Scheduling:* choose next packet to send on link
- *FIFO (first in first out) scheduling:* send in order of arrival to queue
  - *discard policy:* if packet arrives to full queue: who to discard?
    - *tail drop:* drop arriving packet
    - *priority:* drop/remove on priority basis
    - *random:* drop/remove randomly



- *priority scheduling:* send highest priority queued packet
  - multiple *classes*, with different priorities. Class may depend on marking or other header info, e.g. IP source/dest, port numbers, etc.
- *Round Robin (RR) scheduling:*
  - multiple classes
  - Cyclically scan class queues, sending one complete packet from each class (if available)
- *Weighted Fair Queuing (WFQ):*
  - generalized Round Robin
  - each class gets weighted amount of service in each cycle

Active Queue Management (Buffer Management): RED (Random Early Detection)