CS3640

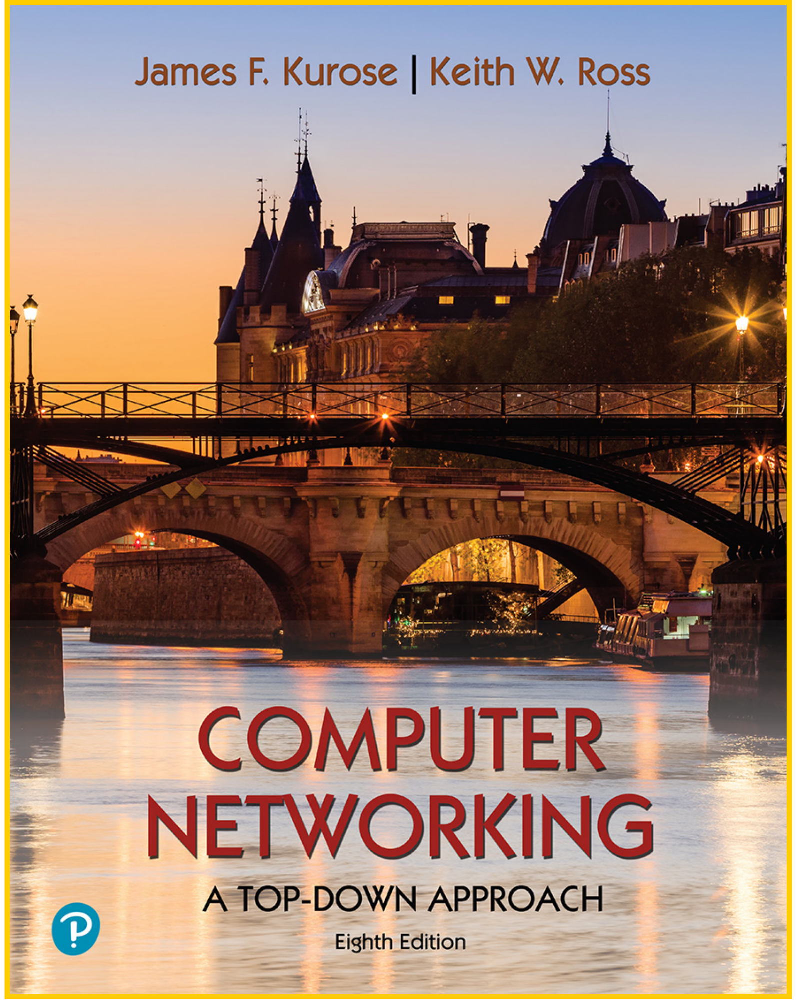
Network Layer (1): Routers

Prof. Supreeth Shastri
Computer Science
The University of Iowa

Lecture goals

understanding the network layer services and how they are implemented in a router

- *routing and forwarding*
- *design and architecture of a router*



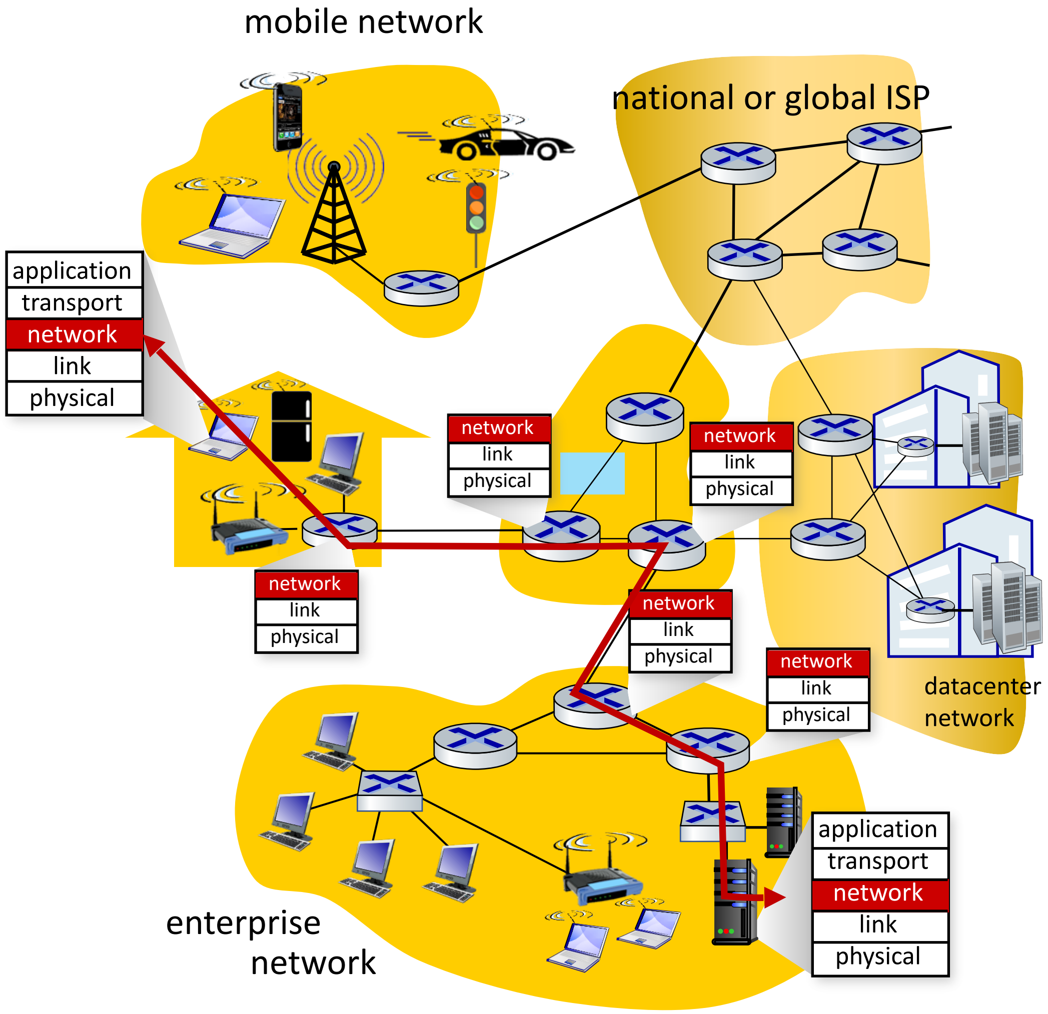
Chapters 4.1 - 4.2

An Overview of Network Layer

Network layer: logical communication b/w **network hosts**

Transport layer: logical communication b/w **processes**

Network layer functionalities exist in each and every Internet device!



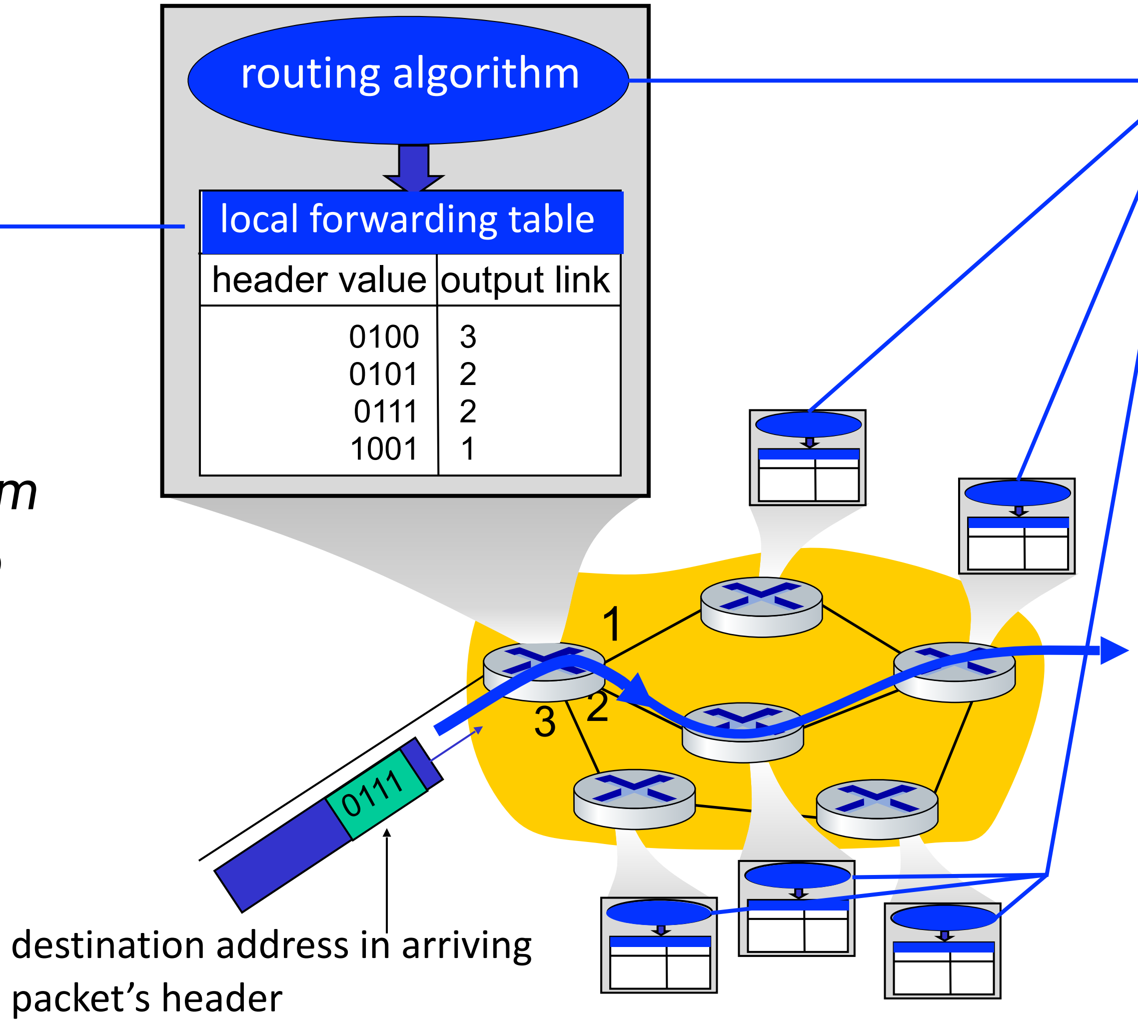
What does a router do?

- examines header fields in all IP datagrams passing through it
- transfers datagrams from input ports to output ports in order to move them along an end-end path

Two Essential Functions of the Network Layer

Forwarding

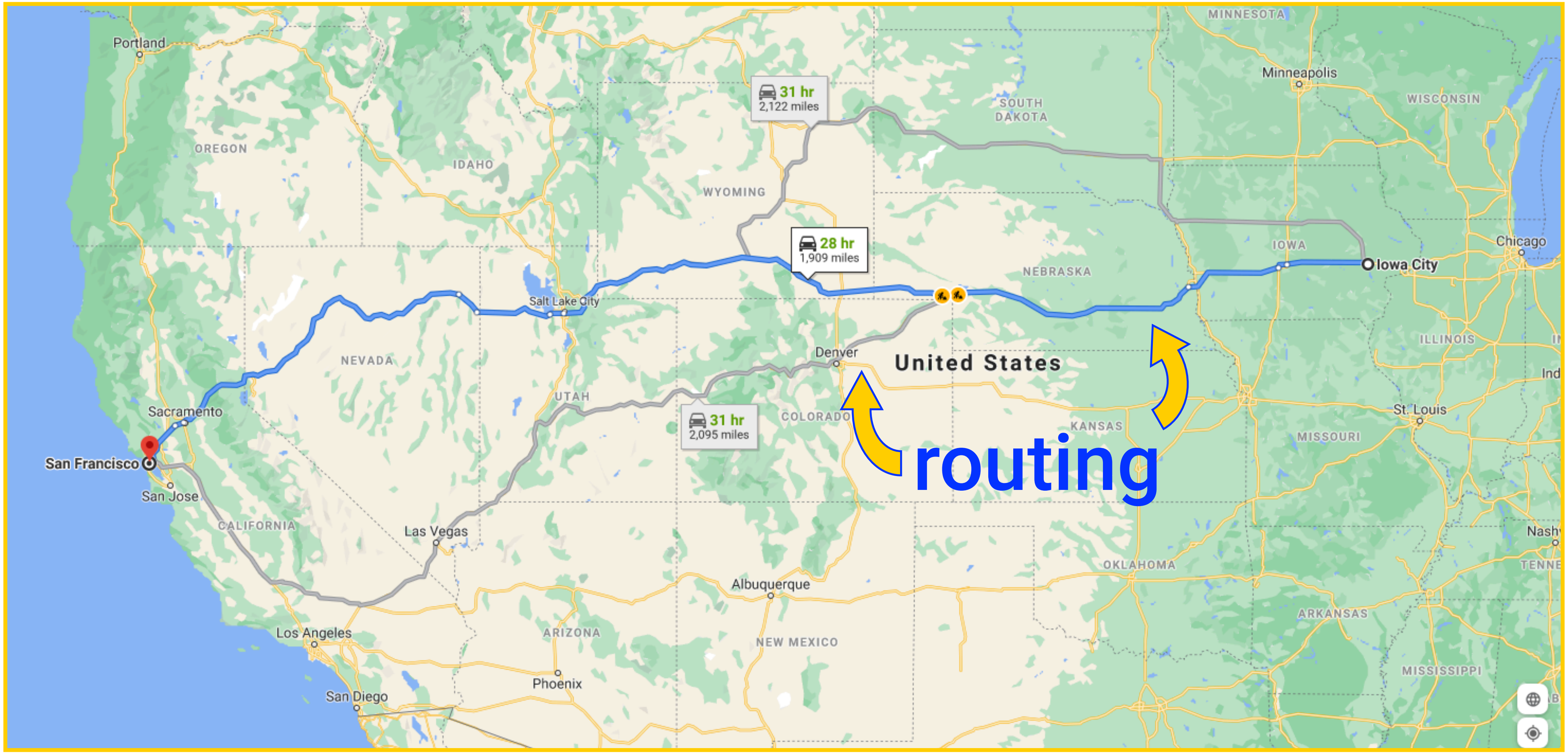
- *aka "switching"*
- ***local action**: move arriving packets from router's input link to appropriate router output link*



Routing

- ***global action**: determine source-destination paths taken by packets*
- *routing algorithms*

Routing vs. forwarding

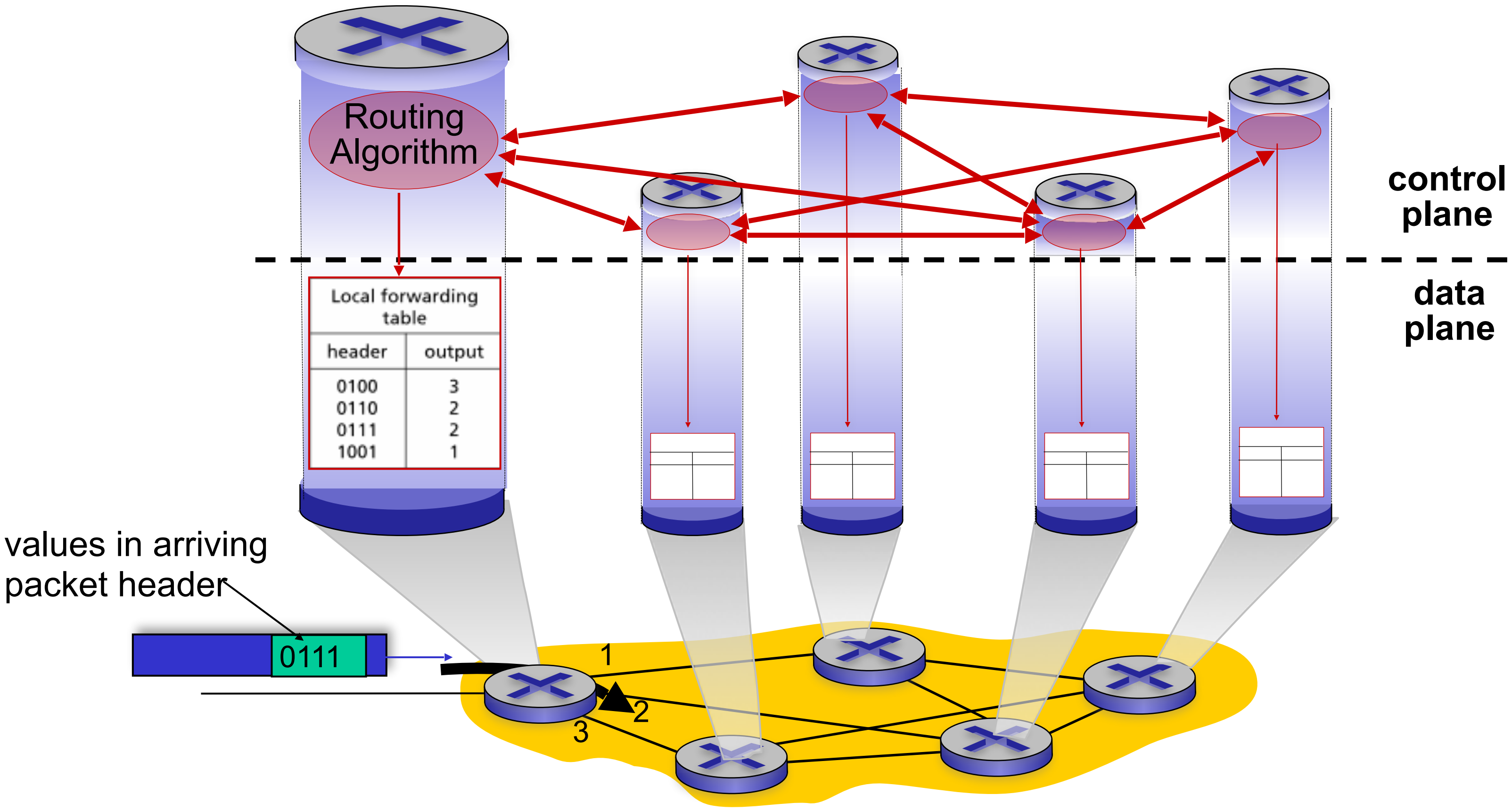


Routing vs. forwarding

Data Plane vs. Control Plane

Forwarding component handles the transfer of datagrams in real time → **data plane**

Routing component determines network-wide logic and paths → **control plane**



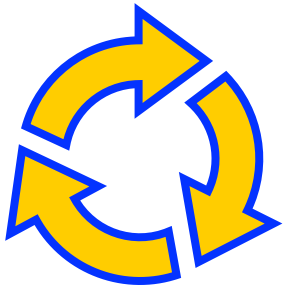
Network Layer Service Models

	Service Model	Bandwidth	Lossless	In-order	Timing
Internet	Best effort	none	no	no	no
Internet	IntServe (RFC 1633)	yes	yes	yes	yes
Internet	DiffServe (RFC2475)	yes	possible	possible	no
ATM	Available bit rate	Guaranteed min	no	yes	no

Reflections on Best-effort Service

Simplicity of the mechanism
has allowed the Internet to be widely deployed and adopted

Sufficient provisioning of bandwidth
has allowed applications to be *good enough* for *most of the time*



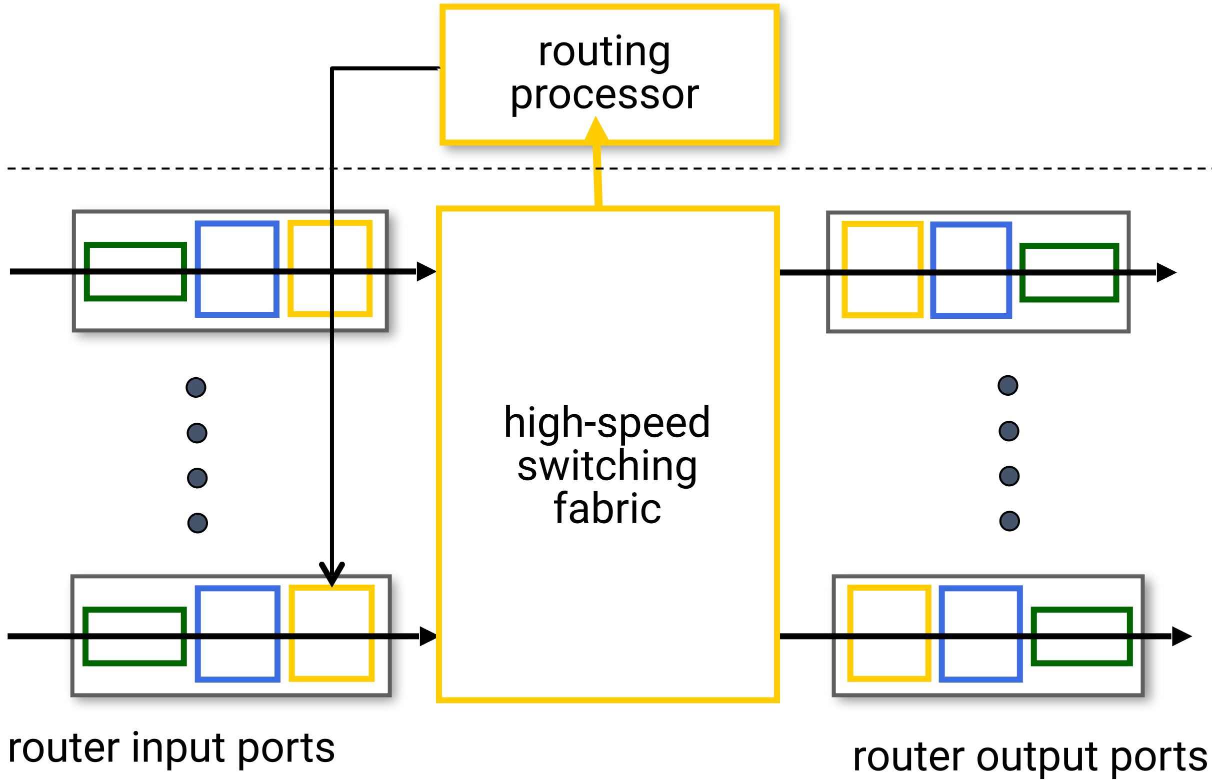
Enhanced application architectures
have allowed services to be reliable, scalable, and economical

It's hard to argue with success of best-effort service model

Cisco ASR 9922 family

Router: Design and Operation

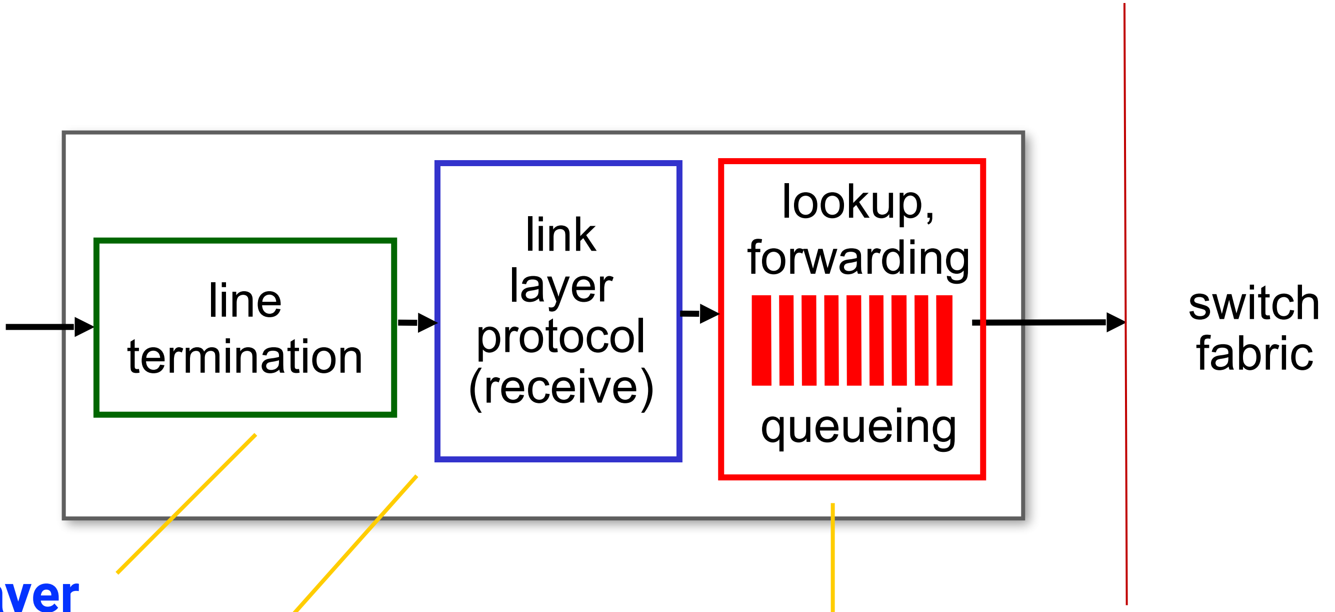
Generic Architecture of a Router



Routing / Control-plane
implemented in software
operates in millisec timeframe

Forwarding / Data-plane
implemented in hardware
operates in nanosec timeframe

Input Port



physical layer
bit-level reception

link layer
e.g., Ethernet

Network layer forwarding

- using header field values, lookup output port using forwarding table in input port memory (“match plus action”)
- destination-based forwarding: forward decision is based only on destination IP address (traditional)
- input port queuing: if datagrams arrive faster than forwarding rate into switch fabric

Destination-based forwarding

forwarding table

Destination Address Range	Link Interface
11001000 00010111 00010000 00000000 through	0
11001000 00010111 00010000 00000100 through 11001000 00010111 00010000 00000111	3
11001000 00010111 00011000 11111111	
11001000 00010111 00011001 00000000 through 11001000 00010111 00011111 11111111	2
otherwise	3

Q: What happens if ranges don't divide up so nicely?

Destination-based forwarding

longest prefix match

when looking for forwarding table entry for given destination address, use longest address prefix that matches destination address.

Destination Address Range	Link interface
11001000 00010111 00010*** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011*** ********	2
otherwise	3

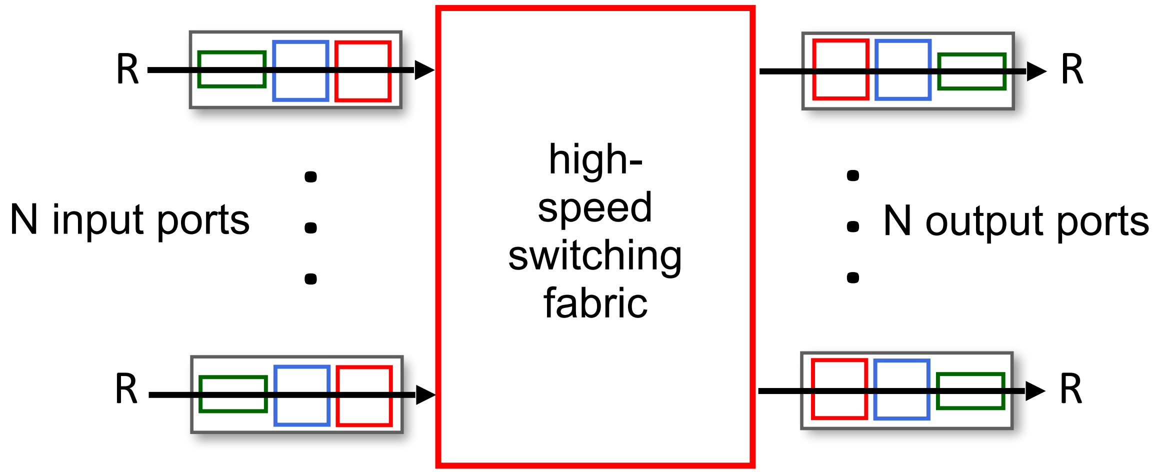
match: 24 bits

match: 21 bits

11001000 00010111 00011000 10101010 which interface?

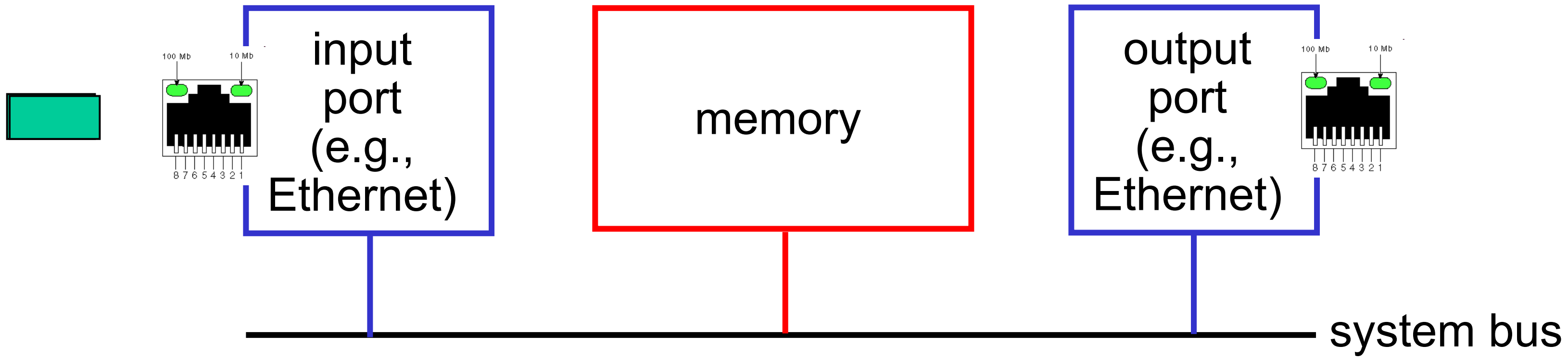
Switching Fabrics

- **Goal**: transfer packets from input link to the appropriate output link
- **switching rate**: aggregate rate at which packets can be transferred
- Measured in multiples of input/output line rates; Non-blocking rate = N*R
- Three types of switching fabrics: *memory-based, bus-based*, and *interconnection network based*



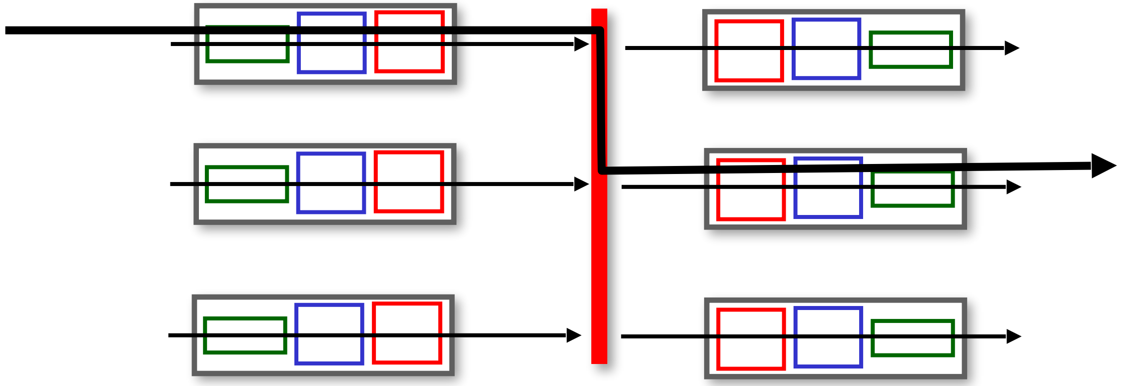
Switching via Memory

- traditional computers switching under the direct control of CPU
- packet are copied to and from system's memory
- speed limited by memory bandwidth (2 bus crossings per datagram)
- First generation of Internet routers (circa 1980)



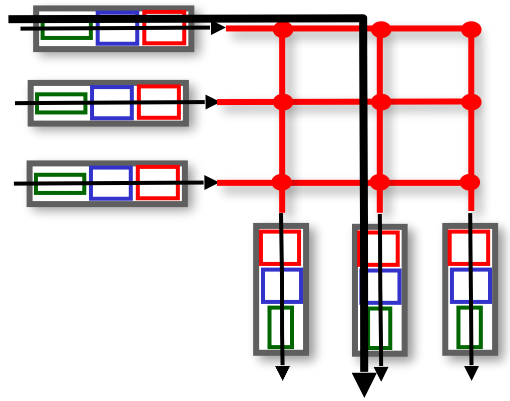
Switching via Bus

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



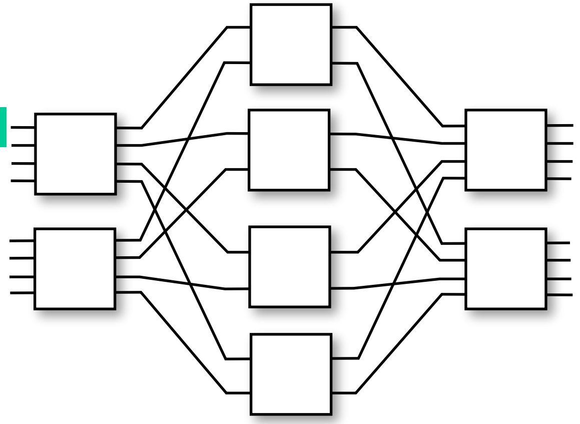
Switching via Interconnection Network

- Interconnection networks such as Crossbar and Clos, initially developed as multiprocessor interconnects, are adapted for packet switching
- Allows parallel flows, thereby increasing throughput



3x3 crossbar

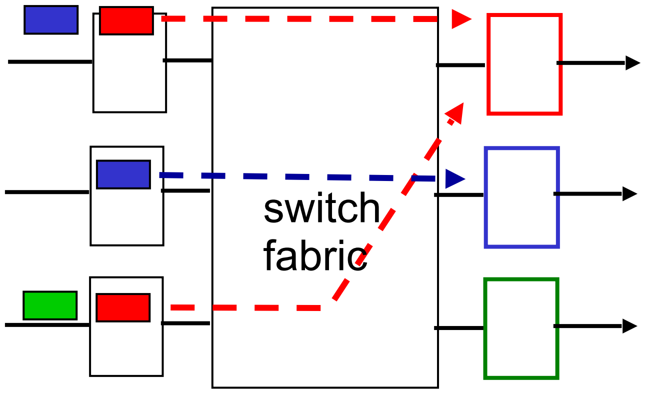
- **Multi-stage switching**: n*n switch built from multiple smaller switches connected in stages
- Exploits parallelism: a datagram could be fragmented into fixed length cells, and then multiple cells could be switched in parallel and reassembled at exit



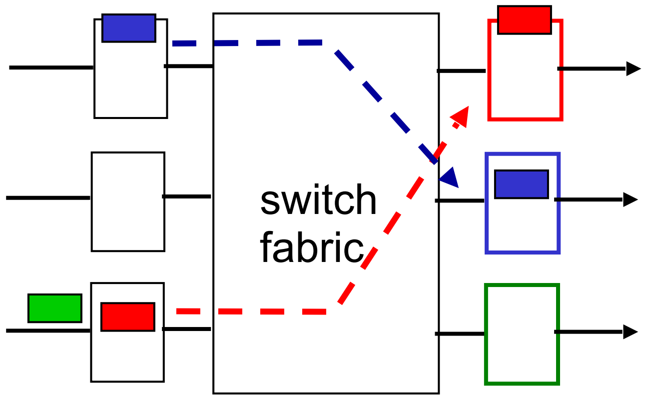
8x8 multistage

Queuing at Input Port

- if switching fabric is slower than input ports combined ⇒ input port queueing will occur
- this could result in increased packet latency or loss due to input buffer overflow
- **Head-of-the-Line (HOL) blocking**: datagram at front of queue prevents others in queue from moving forward even though their destination port and switching fabric capacity is available



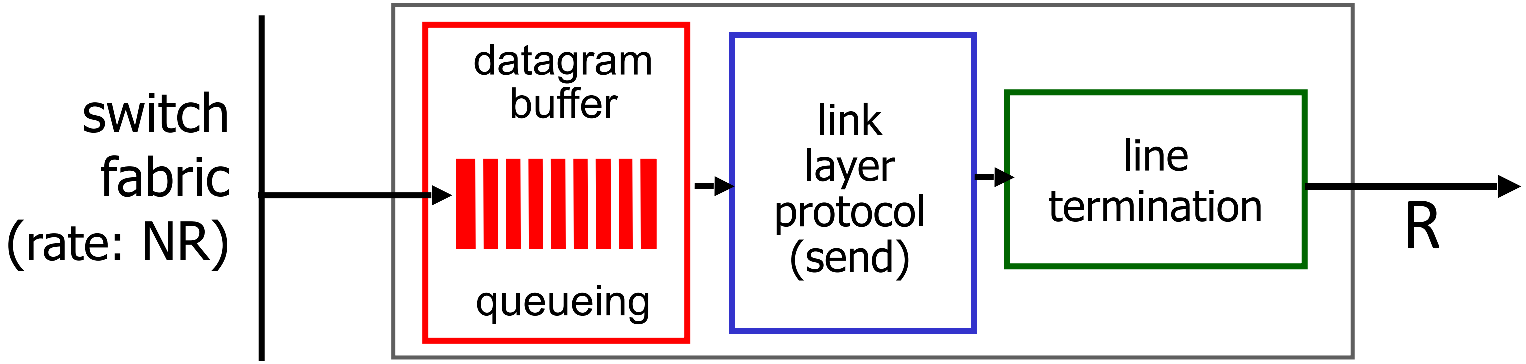
output port contention
only one red datagram can be transferred.
Lower red packet is *blocked*



one packet time later: green packet experiences **HOL blocking**

Queuing at Output Port

- **Output buffering** required when datagrams arrive from the switching fabric faster than the link's outgoing transmission rate
- **Scheduling policy**: how to choose from amongst the queued datagrams for transmission?
- **Drop policy**: which datagrams to drop if no free buffers?

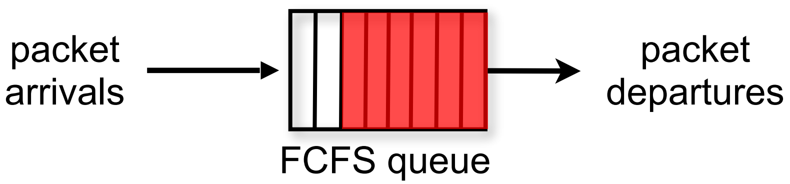


Q: Would network performance improve if routers had unlimited buffers?

Scheduling Policies

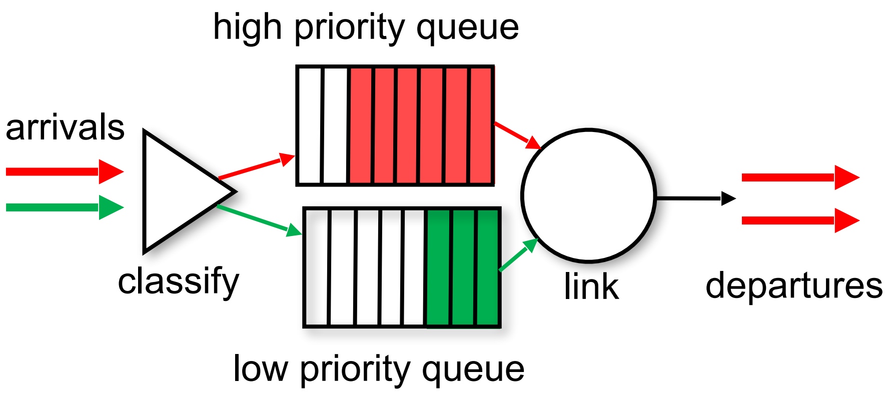
First Come First Serve (FCFS)

- packets transmitted in order of arrival
- simple implementation



Priority scheduling

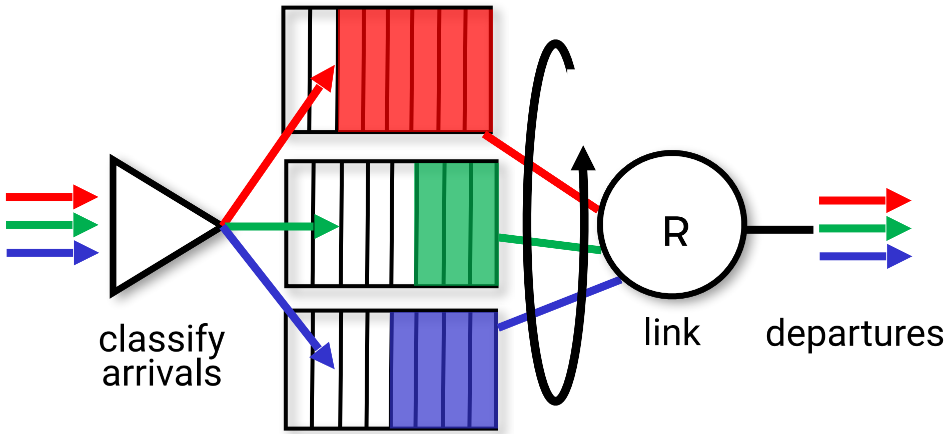
- arriving traffic classified into multiple priority queues;
- always transmit packets from the highest priority queue



Scheduling Policies

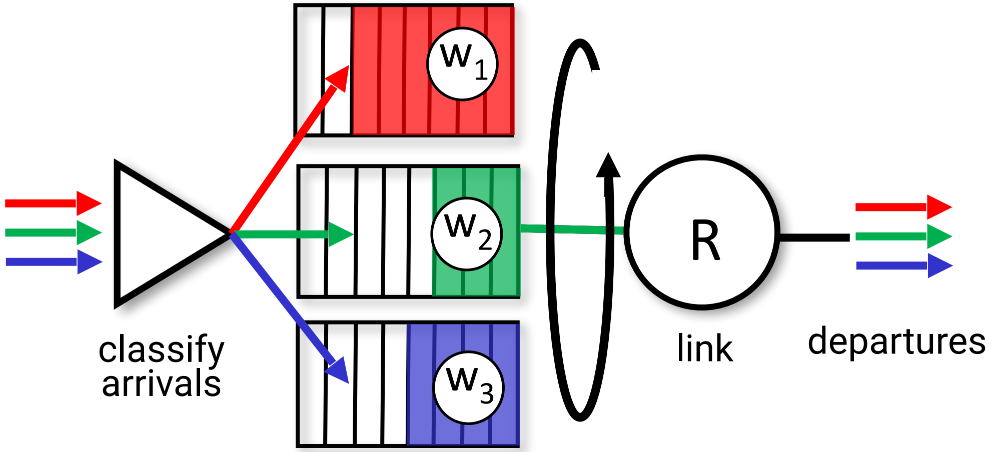
Round Robin (RR)

- arriving traffic classified into multiple classes
- scan class queues cyclically, sending one complete packet from each class (if available) in turn



Weighted Fair Queue

- generalized form of Round Robin
- each class i, has weight w(i) and gets weighted amount of service in each cycle



both policies can offer minimum bandwidth guarantee per traffic class

Spot Quiz (ICON)