Internet Protocols EBU5403
The Network Layer Part 1

Module organiser: Richard Clegg
(r.clegg@qmul.ac.uk)
Michael Chai (michael.chai@qmul.ac.uk)
Cunhua Pan(c.pan@qmul.ac.uk)

	Part 1	Part 2	Part 3	Part 4
Ecommerce + Telecoms 1	Richard Clegg		Cunhua Pan	
Telecoms 2	Michael Chai			

Structure of course

- Part A
  - Introduction to IP Networks
  - The Transport layer (part I)
- Part B
  - The Transport layer (part II)
  - The Network layer (part I)
  - Class test
- Part C
  - The Network layer (part II)
  - The Data link layer (part I)
  - Router lab tutorial (assessed lab work after this week)
- Part D
  - The Data link layer (part II)
  - Network management and security
  - Class test

Network Layer: outline

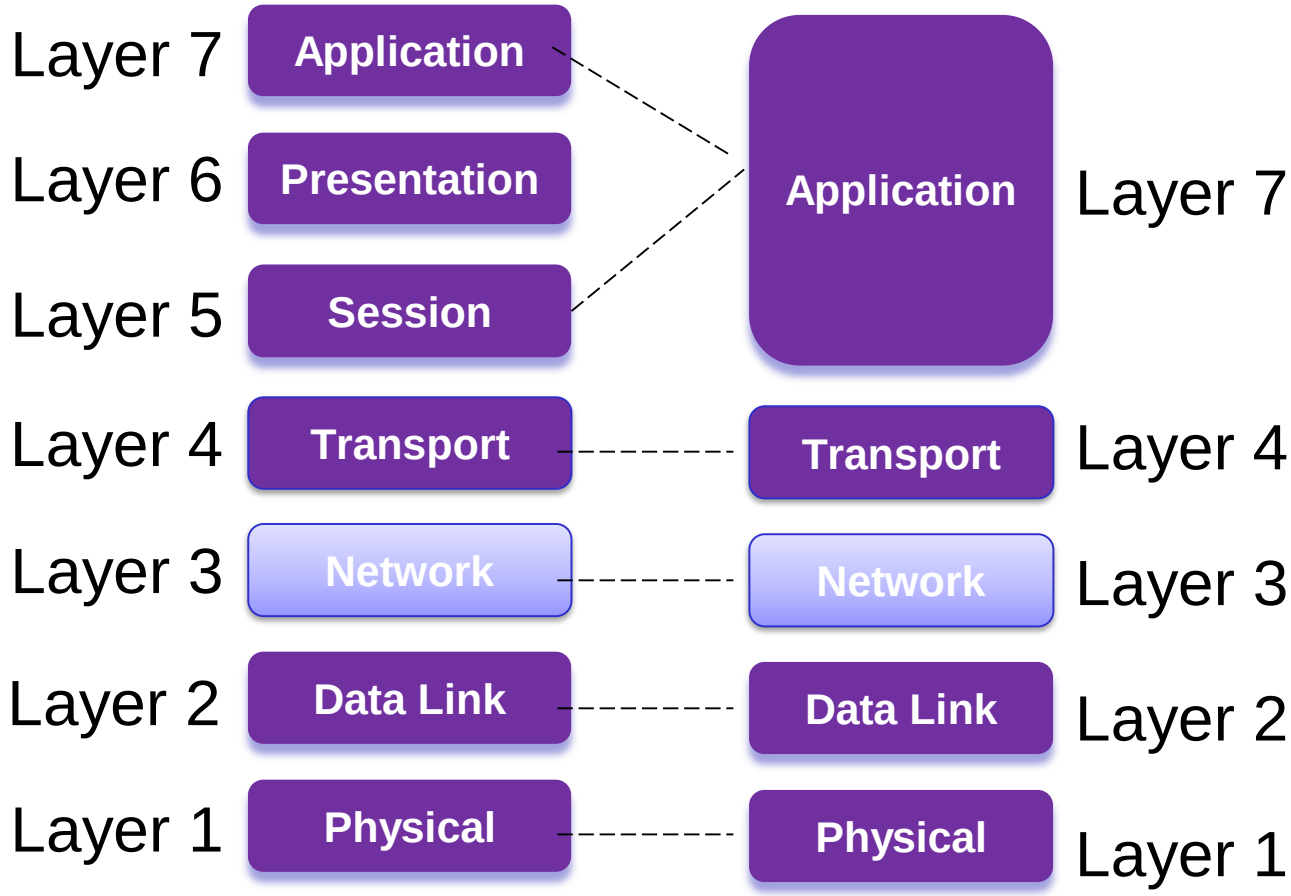
4.1 Overview of Network layer
- data plane
- control plane

4.2 What's inside a router

4.3 IP: Internet Protocol
- datagram format
- fragmentation
- IPv4 addressing
- network address translation
- IPv6

4.4 Generalized Forward and SDN
- match
- action
- OpenFlow examples of match-plus-action in action

Network Layer



Network layer

Goals:

- understand principles behind network layer services, focusing on data plane:
  - network layer service models
  - forwarding versus routing
  - how a router works
  - generalized forwarding
- instantiation, implementation in the Internet

Two key network-layer functions

network-layer functions:

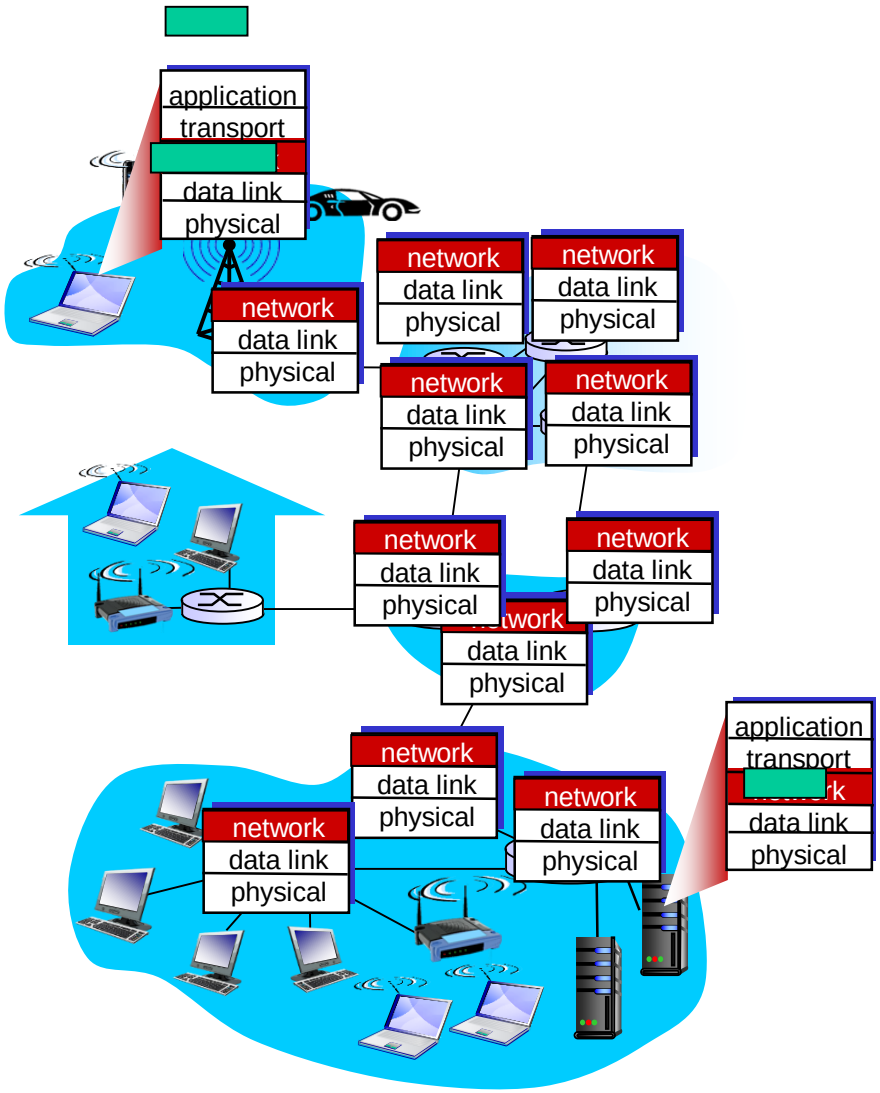
- *forwarding:* move packets from router's input to appropriate router output
- *routing:* determine route taken by packets from source to destination
  - *routing algorithms*

analogy: taking a trip

- *forwarding:* process of getting through one road junction
- *routing:* process of planning trip from source to destination

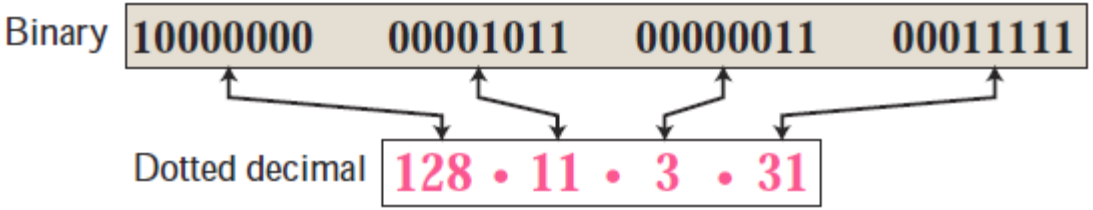
Network layer

- transport segment from sending to receiving host
- on sending side encapsulates segments into datagrams
- on receiving side, delivers segments to transport layer
- network layer protocols in *every* host, router
- router examines header fields in all IP datagrams passing through it



IPv4 address notation

- There are three common notations to show an IPv4 address:
  - binary notation
  - dotted-decimal notation (most commonly used)
  - hexadecimal notation



Testing your understanding

Convert the following IPv4 addresses from binary to dotted-decimal notation.

a. 10000001 00001011 00001011 11101111

b. 11000001 10000011 00011011 11111111

c. 11100111 11011011 10001011 01101111

d. 11111001 10011011 11111011 00001111

Testing your understanding

Convert the following IPv4 addresses from binary to dotted-decimal notation.

a. 10000001 00001011 00001011 11101111

129.13.13.239

b. 11000001 10000011 00011011 11111111

193.131.27.255

c. 11100111 11011011 10001011 01101111

231.219.139.111

d. 11111001 10011011 11111011 00001111

249.155.251.15

Converting binary to/from decimal (8 bits)

Convert 10100010 to decimal

Factor	128	64	32	16	8	4	2	1
Binary	1	0	1	0	0	0	1	0
Decimal	128	0	32	0	0	0	2	0

Add together 128 + 32 + 2 = 162

Convert decimal to binary – if number is bigger than or equal to factor subtract factor and put 1 in binary column. Example 155

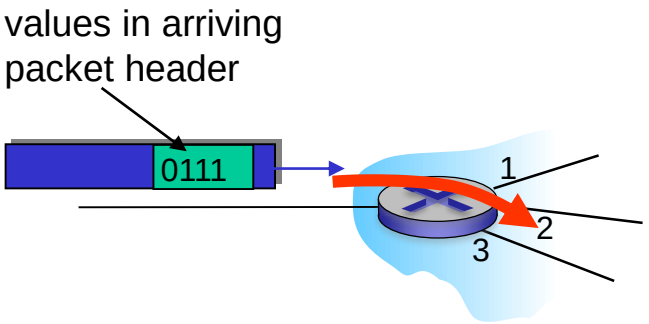
Factor	128	64	32	16	8	4	2	1
Decimal	155-128	27	27	27-16	11-8	3	3-2	1
Binary	1	0	0	1	1	0	1	1

155 in decimal is 10011011

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

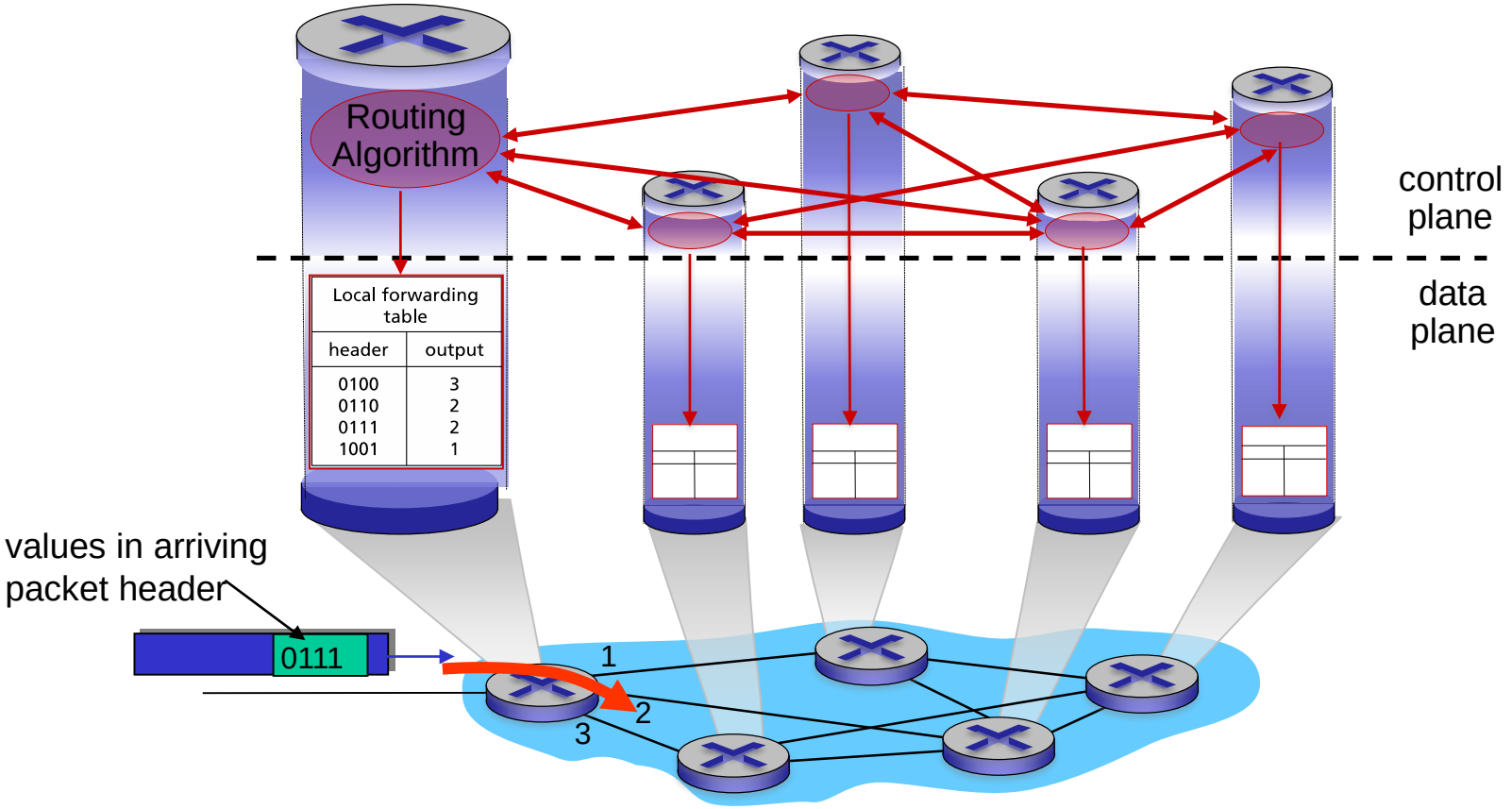


Control plane

- network-wide logic
- determines how datagram is routed among routers along end-end path from source host to destination host
- two control-plane approaches:
  - *traditional routing algorithms:* implemented in routers
  - *software-defined networking (SDN)*: implemented in (remote) servers

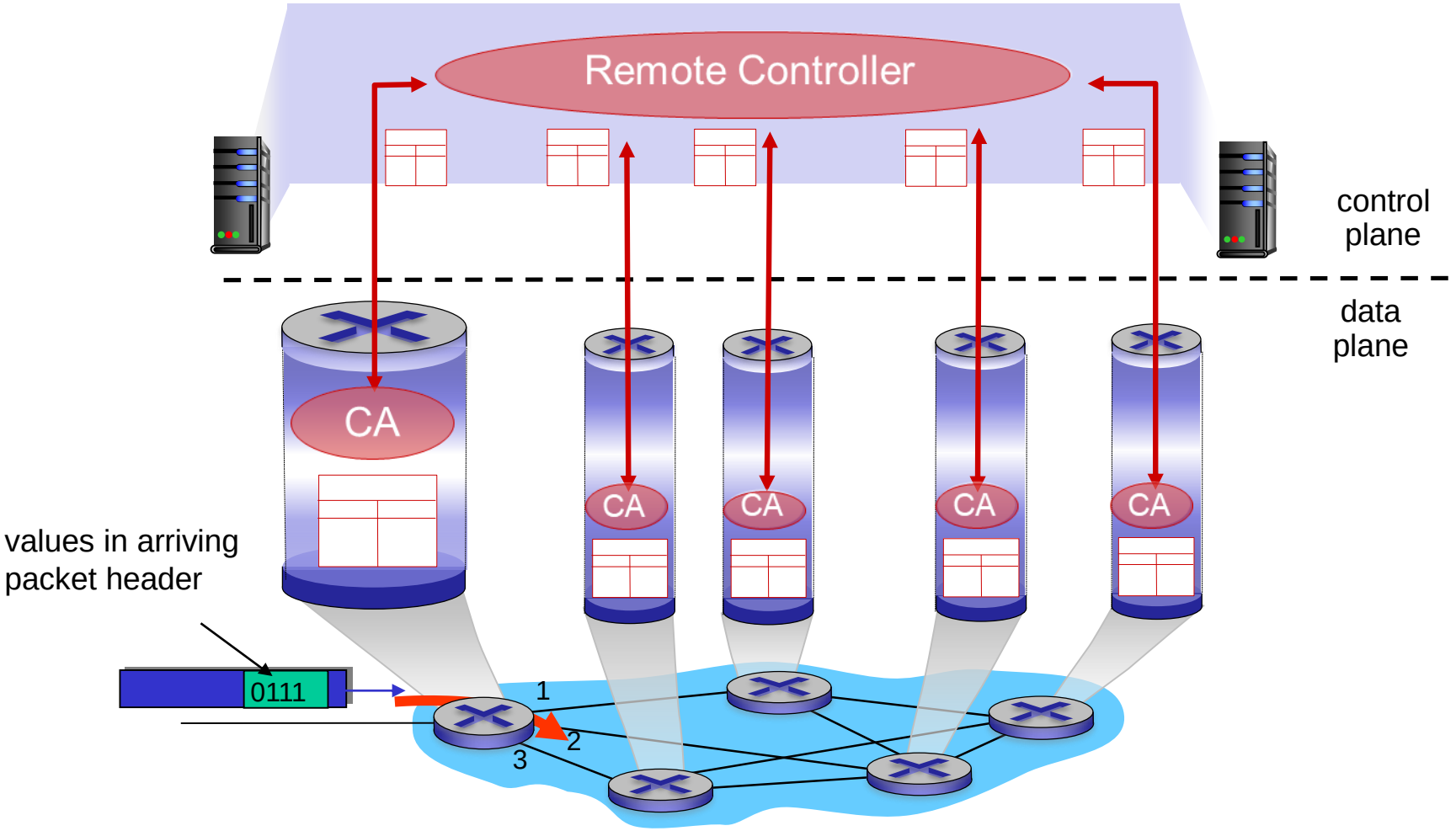
Per-router control plane

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)



Network service model

Q: What is *service model* for "channel" transporting datagrams from sender to receiver?

example services for individual datagrams:

- guaranteed delivery
- guaranteed delivery with less than 40 msec delay

example services for a flow of datagrams:

- in-order datagram delivery
- guaranteed minimum bandwidth to flow
- restrictions on changes in inter-packet spacing

Network layer service models:

Network Architecture	Service Model	Guarantees ?				Congestion feedback
		Bandwidth	Loss	Order	Timing	
Internet	best effort	none	no	no	no	no (inferred via loss)
ATM	CBR	constant rate	yes	yes	yes	no congestion
ATM	VBR	guaranteed rate	yes	yes	yes	no congestion
ATM	ABR	guaranteed minimum	no	yes	no	yes
ATM	UBR	none	no	yes	no	no

Network Layer (Data): outline

4.1 Overview of Network layer
- data plane
- control plane

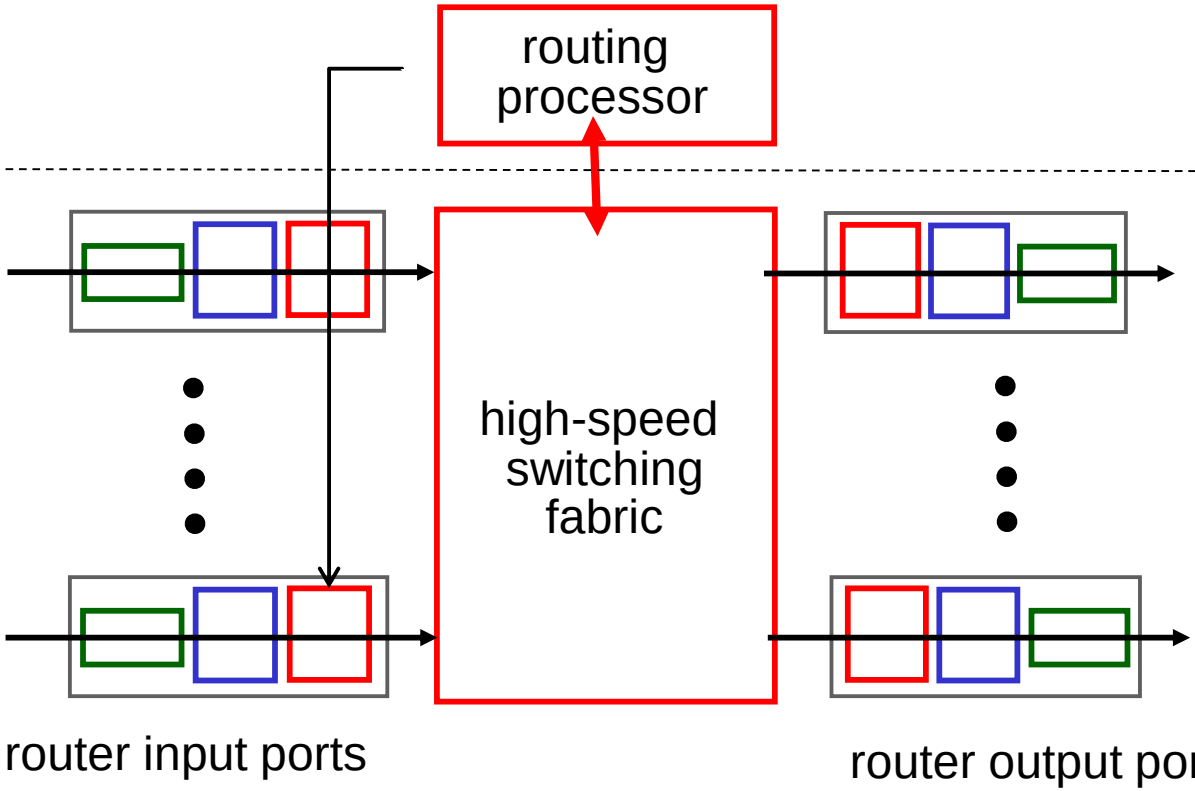
4.2 What's inside a router

4.3 IP: Internet Protocol
- datagram format
- fragmentation
- IPv4 addressing
- network address translation
- IPv6

4.4 Generalized Forward and SDN
- match
- action
- OpenFlow examples of match-plus-action in action

Router architecture overview

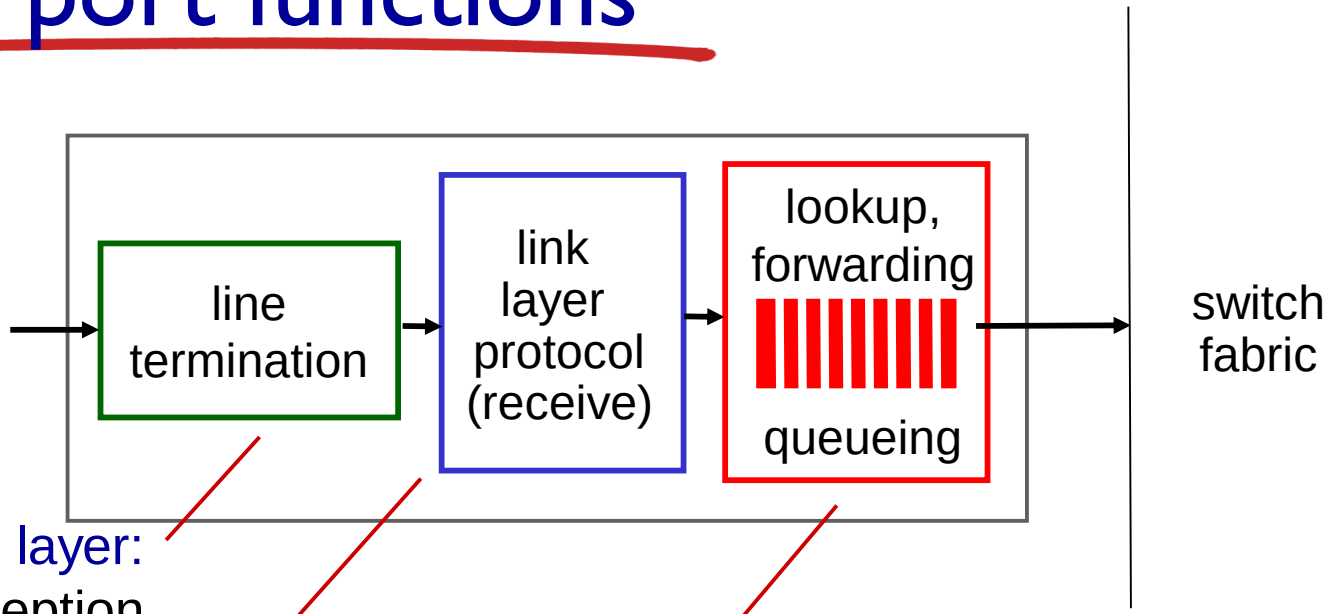
- high-level view of generic router architecture:



routing, management control plane (software) operates in millisecond time frame

forwarding data plane (hardware) operates in nanosecond timeframe

Input port functions



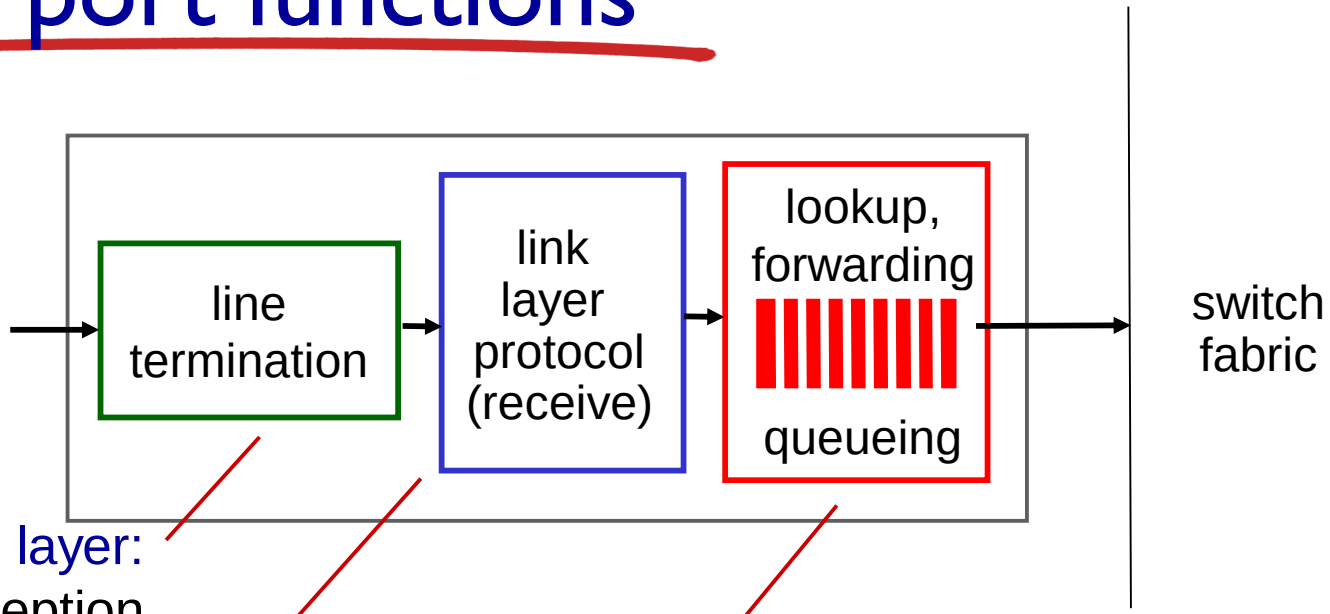
physical layer:
bit-level reception

data link layer:
e.g., Ethernet
see chapter 5
of course text

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

Input port functions



physical layer:
bit-level reception

data link layer:
e.g., Ethernet
see chapter 5
of course text

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

Longest prefix matching

longest prefix matching

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

examples:

DA: 11001000 00010111 00010110 10100001 which interface?

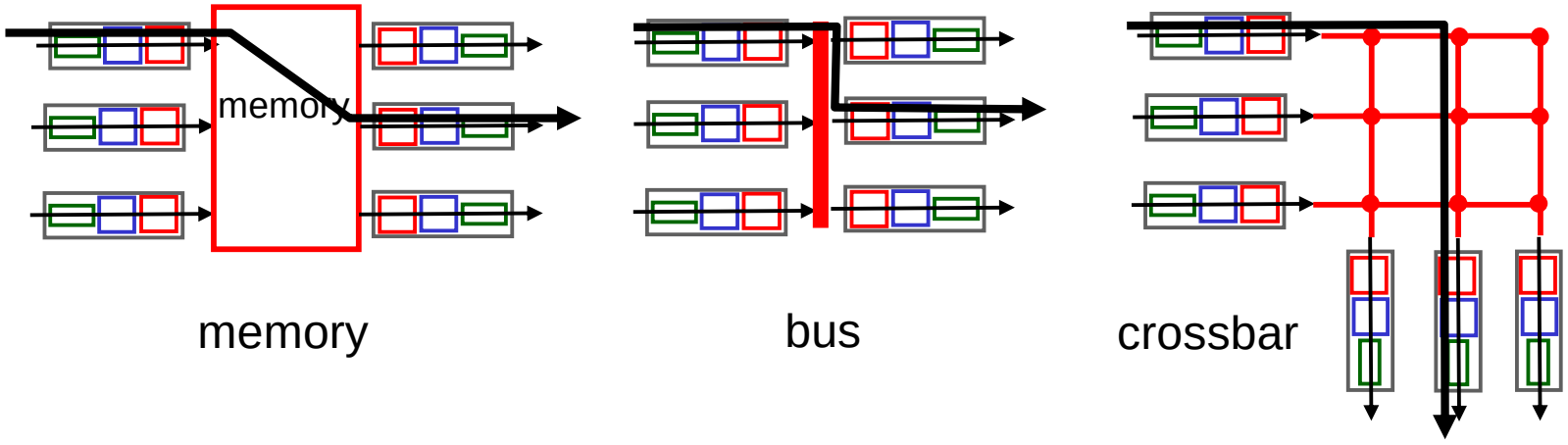
DA: 11001000 00010111 00011000 10101010 which interface?

Longest prefix matching

- we'll see *why* longest prefix matching is used shortly, when we study addressing
- longest prefix matching: often performed using ternary content addressable memories (TCAMs) specialised very high speed memory
  - *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

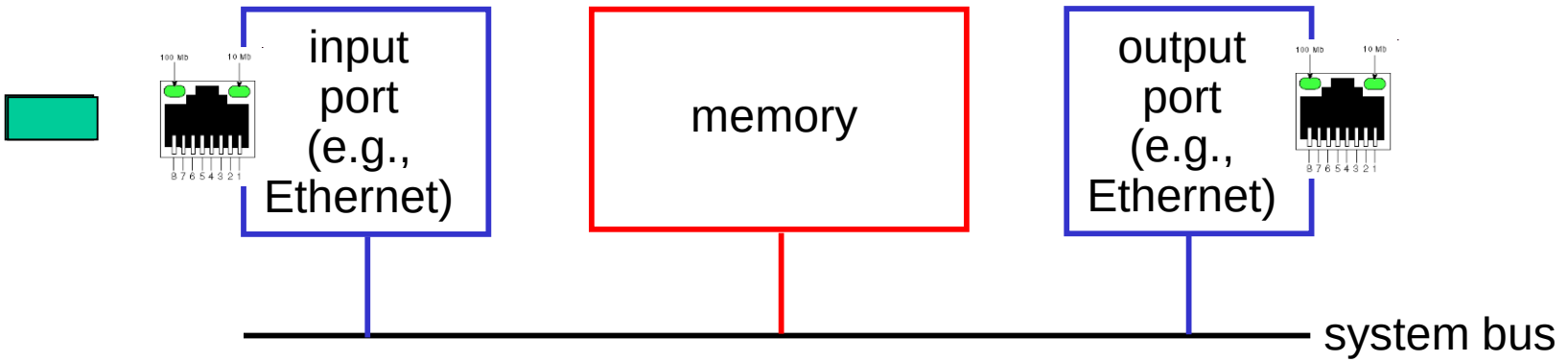
- transfer packet from input buffer to appropriate output buffer
- switching rate: rate at which packets can be transfer from inputs to outputs
  - often measured as multiple of input/output line rate
  - N inputs: switching rate N times line rate desirable
- three types of switching fabrics



Switching via memory

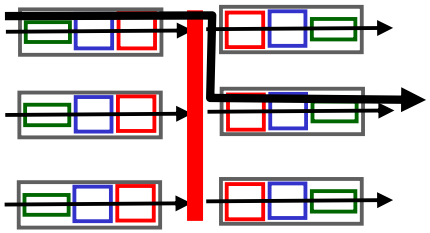
first generation routers:

- 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

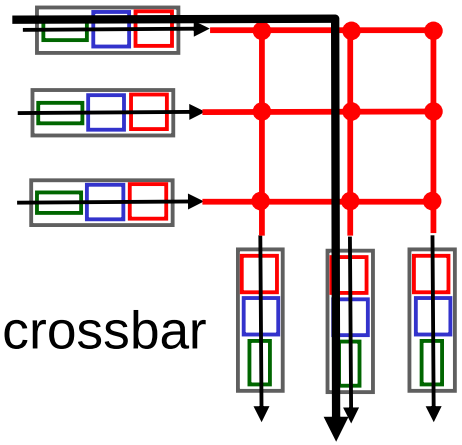
- datagram from input port memory to output port memory via a shared bus
- *bus contention:* switching speed limited by bus bandwidth
- 32 Gbps bus, Cisco 5600: sufficient speed for access and enterprise routers



bus

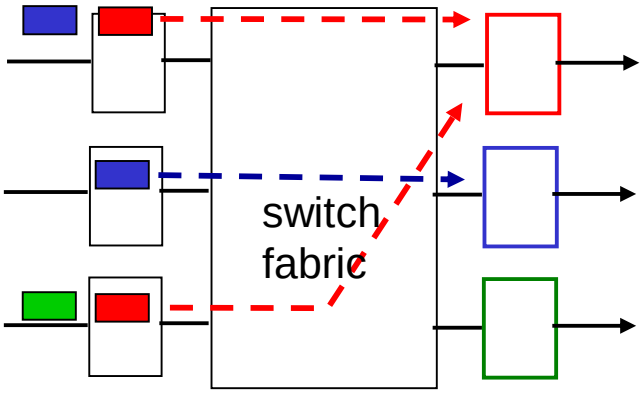
Switching via interconnection network

- 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

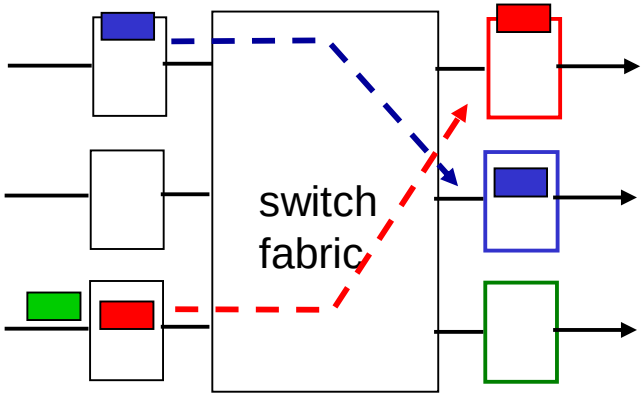


Input port queuing

- fabric slower than input ports combined -> queueing may occur at input queues
  - *queueing 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

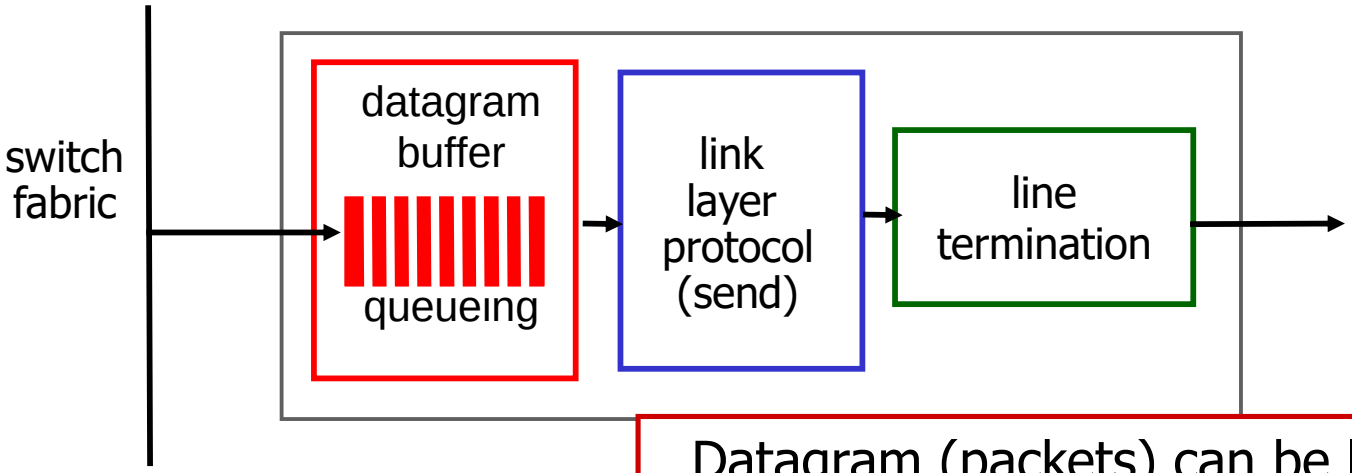


output port contention:
only one red datagram can be transferred.
lower red packet is blocked



one packet time later:
green packet
experiences HOL
blocking

Output ports

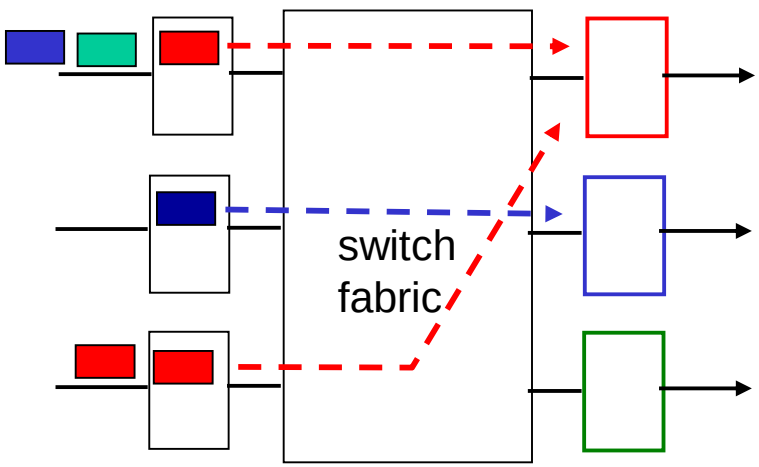


Datagram (packets) can be lost due to congestion, lack of buffers

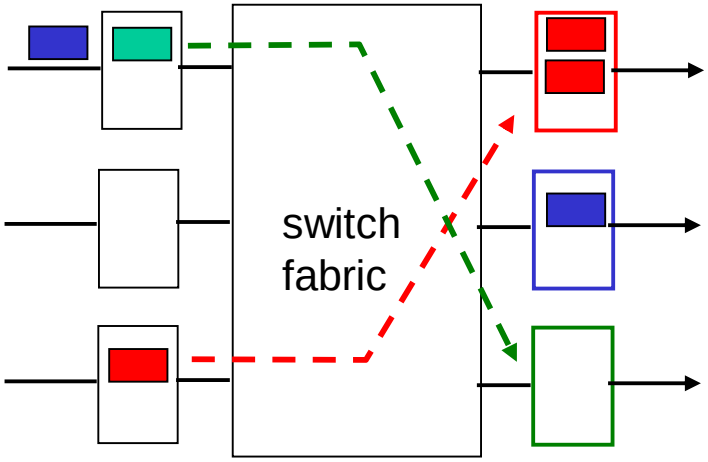
- *buffering* required when datagrams arrive from fabric faster than the transmission rate
- *scheduling discipline* chooses among queued datagrams for transmission

Priority scheduling – who gets best performance, network neutrality

Output port queueing



at *t*, packets more from input to output



one packet time later

- buffering when arrival rate via switch exceeds output line speed
- *queueing (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

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

- Why? Why not simply have very large buffers so no loss?

Scheduling mechanisms

- *scheduling:* choose next packet to send on link.
- FIFO scheduling – first in first out
  - Like an orderly queue of people, no pushing in.
  - If queue is full last packets are dropped.
- Priority scheduling
  - Some packets are more important
  - Example: You need live video packets now, email could wait.
- Round robin scheduling
  - If your queue is from several inputs ports treat them fairly
  - Pick a packet from input port 1, then 2, then 3, then 4
  - Port which is sending lots of traffic doesn't block others.
  - Weighted Fair Queue (like this but give some queues a little more priority -- give a little more traffic to port 1)

Network Layer: outline

4.1 Overview of Network layer
- data plane
- control plane

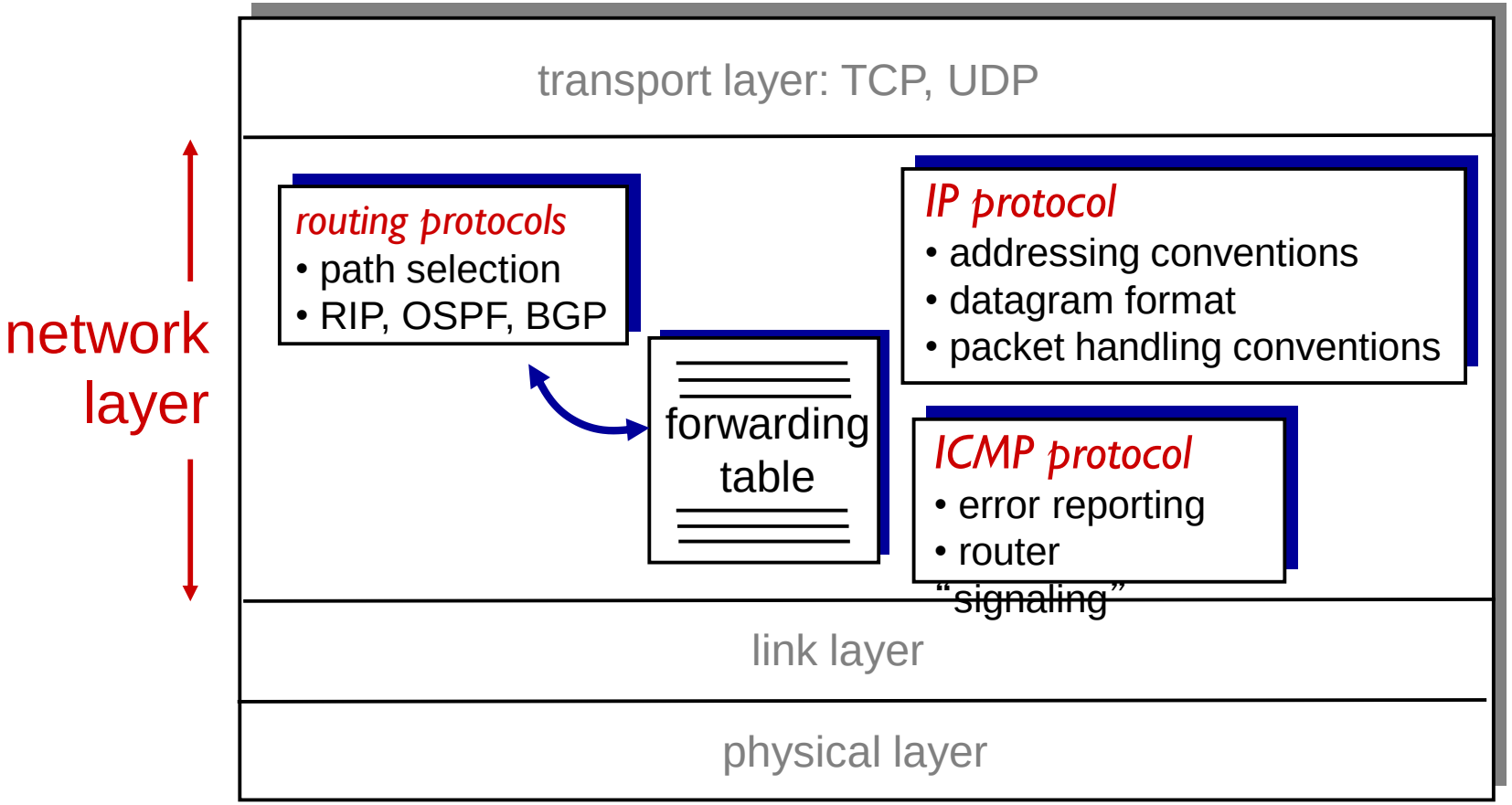
4.2 What's inside a router

4.3 IP: Internet Protocol
- datagram format
- fragmentation
- IPv4 addressing
- network address translation
- IPv6

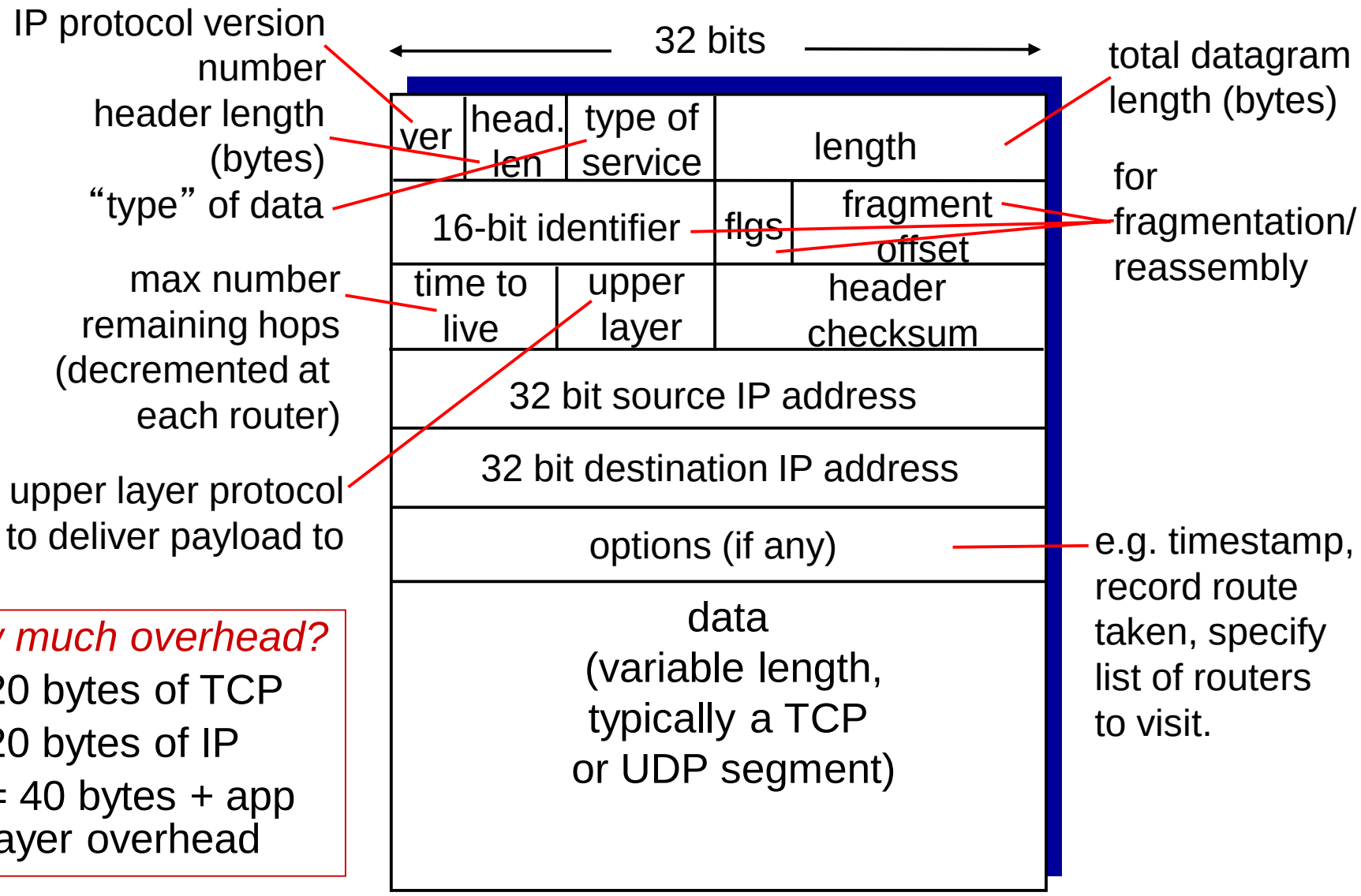
4.4 Generalized Forward and SDN
- match
- action
- OpenFlow examples of match-plus-action in action

The Internet network layer

host, router network layer functions:



IP datagram format

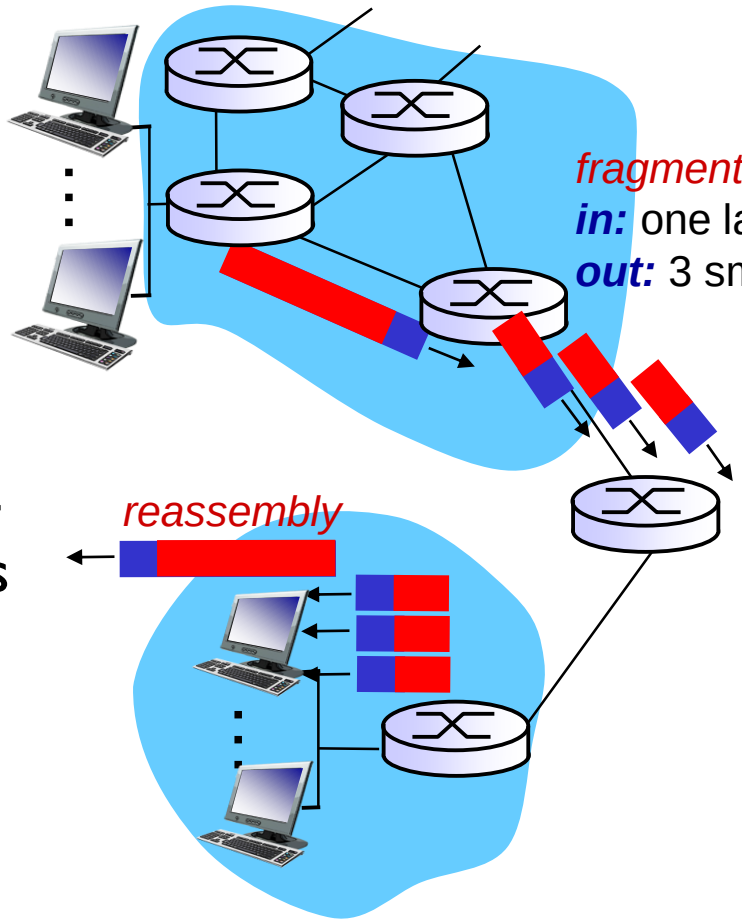


how much overhead?

- 20 bytes of TCP
- 20 bytes of IP
- = 40 bytes + app layer overhead

IP fragmentation, reassembly

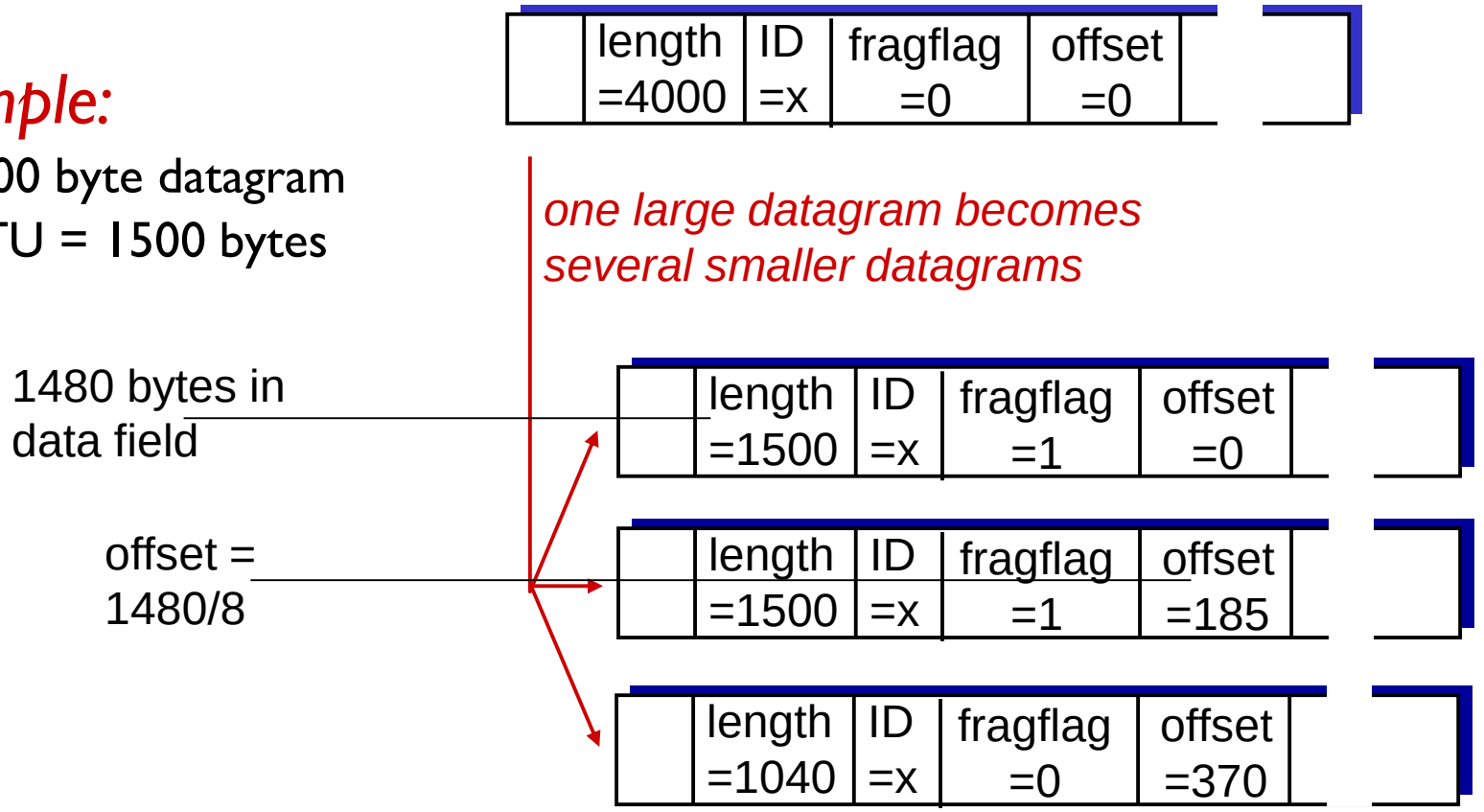
- network links have MTU (max.transfer size) - largest possible link-level frame
  - different link types, different MTUs
- large IP datagram divided ("fragmented") within net
  - one datagram becomes several datagrams
  - "reassembled" only at final destination
  - IP header bits used to identify, order related fragments



IP fragmentation, reassembly

example:

- 4000 byte datagram
- MTU = 1500 bytes



Test your understanding

- An IP packet has 4620 bytes in total of which 20 bytes are the IP header. It enters a network that has a Maximum Transfer Unit of 1900 bytes.
  i) How many fragments will this packet be broken into?
  ii) What are the values of the IP header fields for the offset
  iii) What are the values of the fragmentation flag("More Fragments" bit) for each fragment?

Test your understanding

i) How many fragments will this packet be broken into?

Three fragments.

ii) What are the values of the IP header fields for the offset

The values of the IP header fields for the offset for the three fragments are

0, 235, 470

iii) What are the values of the fragmentation flag("More Fragments" bit) for each fragment?

1, 1, 0

What have we learned?

- Network layer moves data from "source" all the way to the "destination".
- Control plane makes decisions about routes taken by packets it spans a network.
- Data plane implements the decisions and gets the data from place to place. It is local to a router.
- Longest prefix matching on a router is used to pick where to send data.
- Queuing at routers can have huge effects.
- Scheduling polices can benefit some traffic over other traffic.
- Fragmentation and reassembly can be used to split and rejoin IP packets.