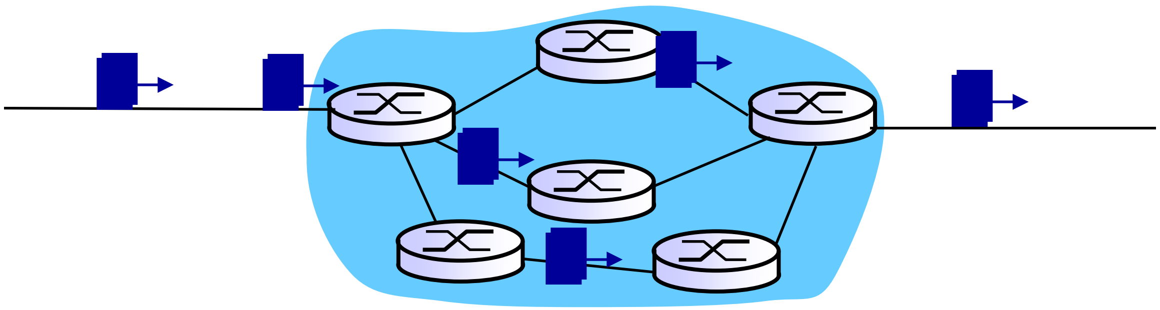
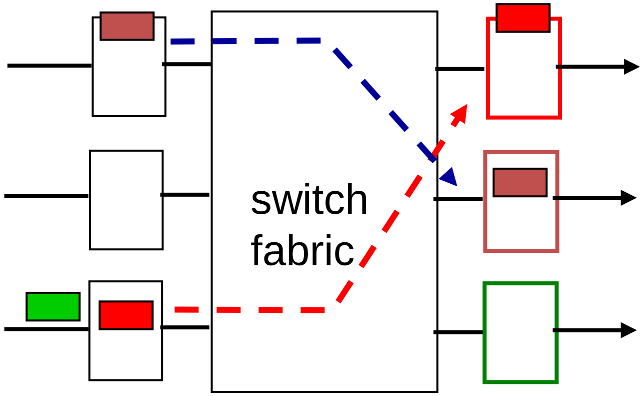
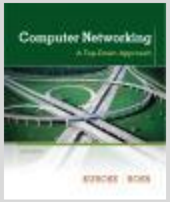
Datagram networks, routers, IP





Computer Networking: A Top Down Approach
6th edition
Jim Kurose, Keith Ross
Addison-Wesley



Chapter 4: outline

4.1 Introduction

4.2 Virtual circuit and datagram networks

4.3 What's inside a router

4.4 IP: Internet Protocol

- Datagram format
- IPv4 addressing
- ICMP
- IPv6

4.5 Routing algorithms

- Link state
- Distance vector
- Hierarchical routing

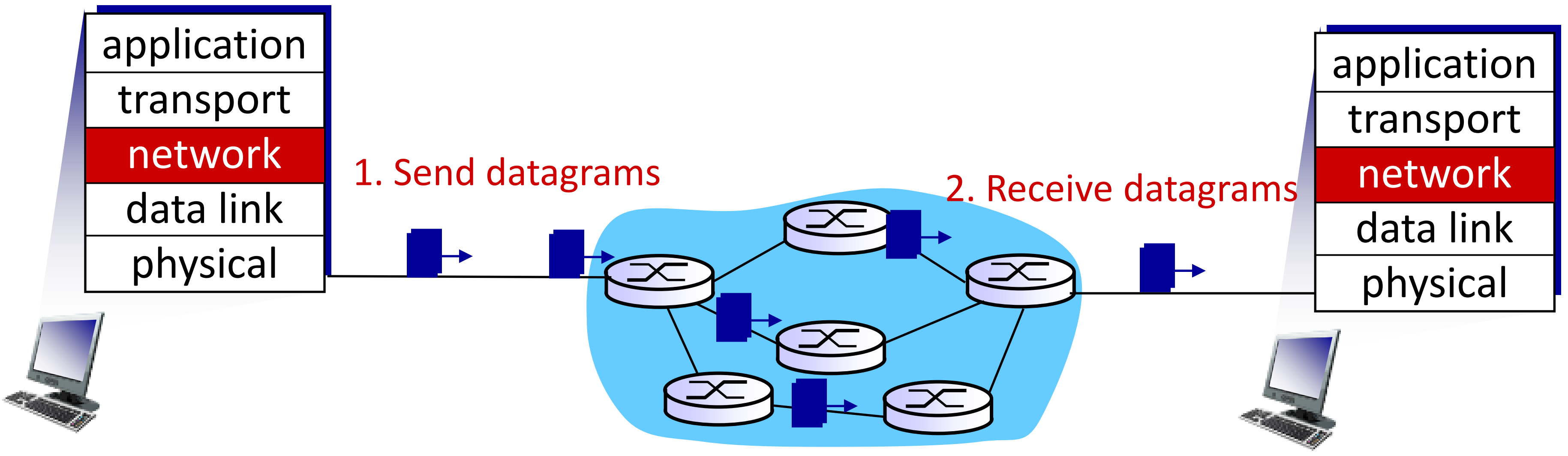
4.6 Routing in the Internet

- RIP
- OSPF
- BGP

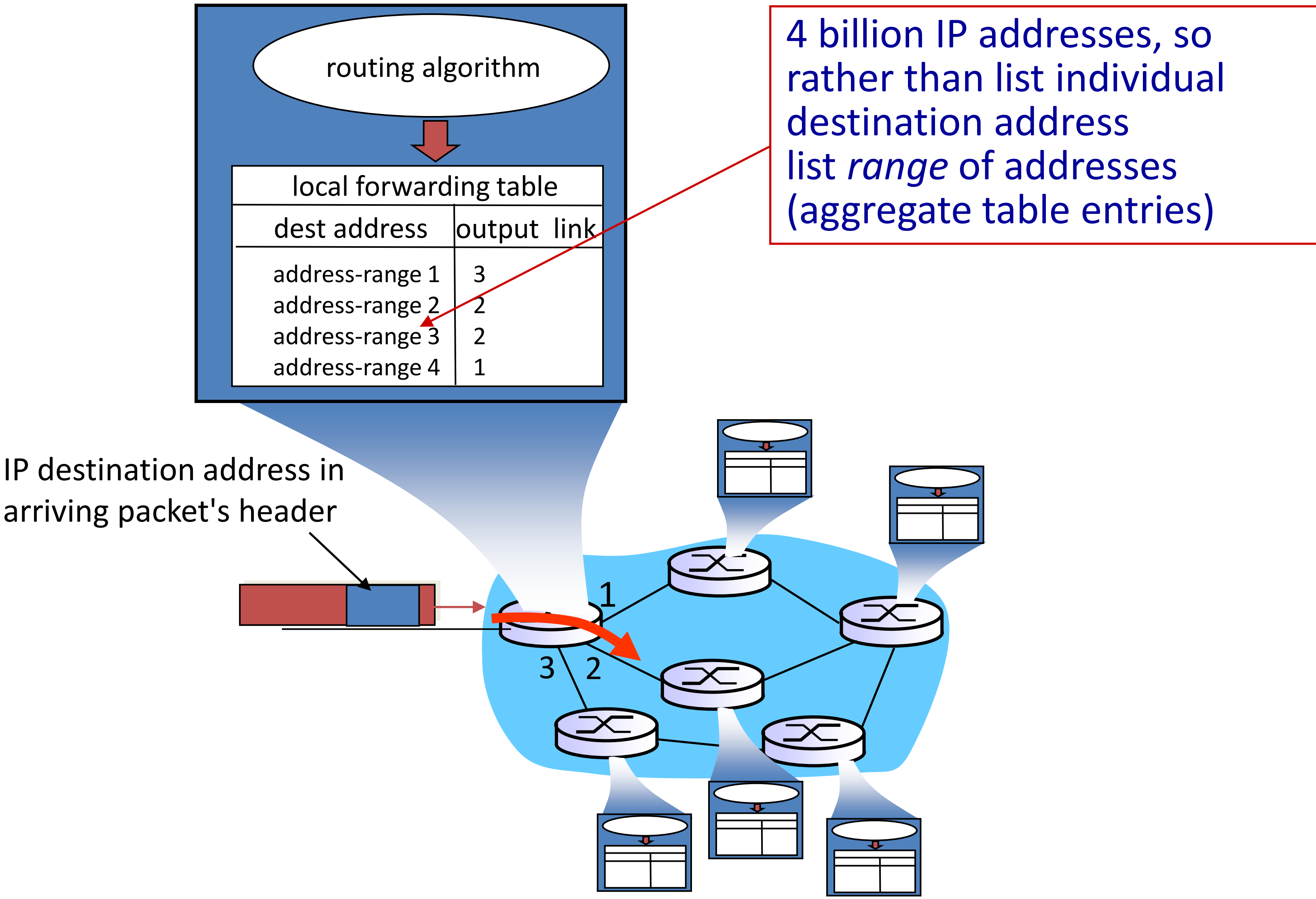
4.7 Broadcast and multicast routing

Datagram networks

- No call setup at network layer
- Routers: no state about end-to-end connections
  - No network-level concept of "connection"
- Packets forwarded using destination host address



Datagram forwarding table



Datagram forwarding table

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 00011001 00000000 through **11001000 00010111 00011111 11111111**	2
otherwise	3

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

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?

Datagram or VC network: why?

Internet (datagram)

- Data exchange among computers
  - Elastic service, no strict timing requirements
- Many link types
  - Different characteristics
  - Uniform service difficult
- Smart end systems (computers)
  - Can adapt, perform control, error recovery
  - ***Simple inside network, complexity at edge***

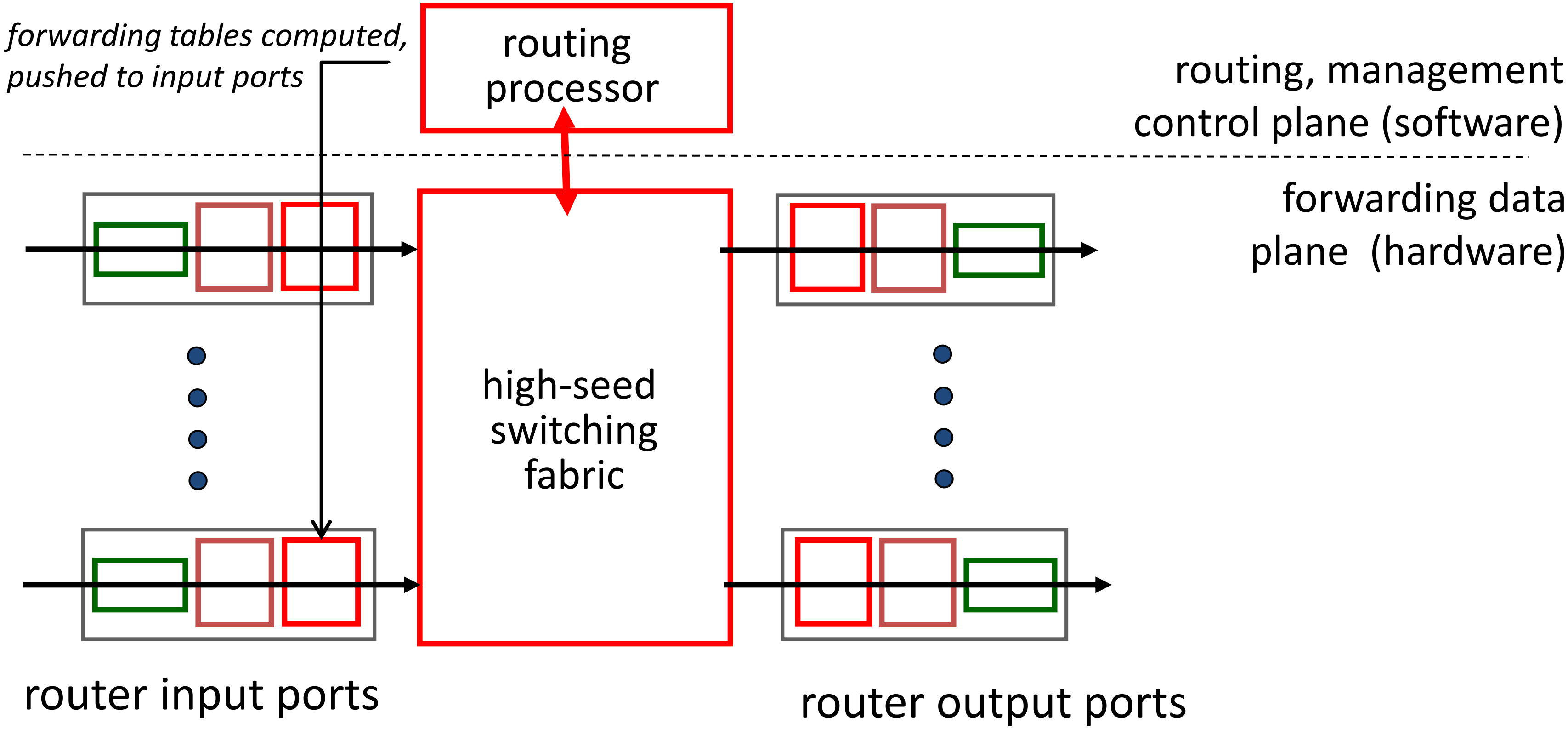
ATM (VC)

- Evolved from telephony
- Human conversation:
  - Strict timing, reliability requirements
  - Need for guaranteed service
- Dumb end systems
  - Telephones
  - ***Complexity inside network***

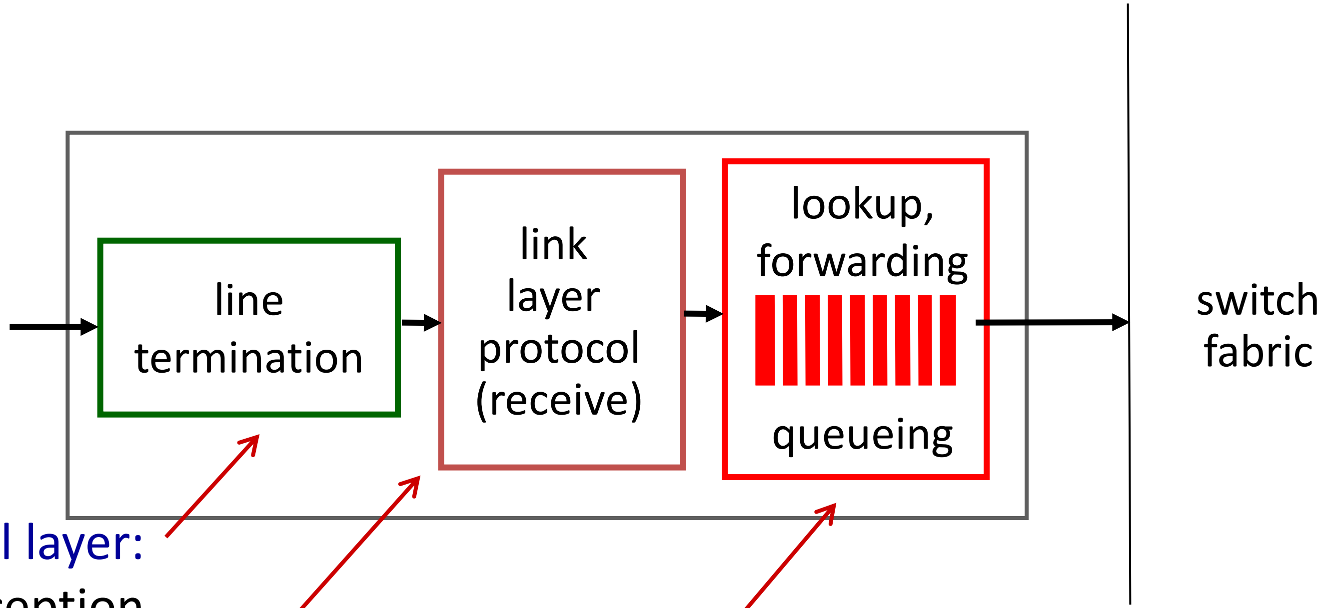
What's inside a router?

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

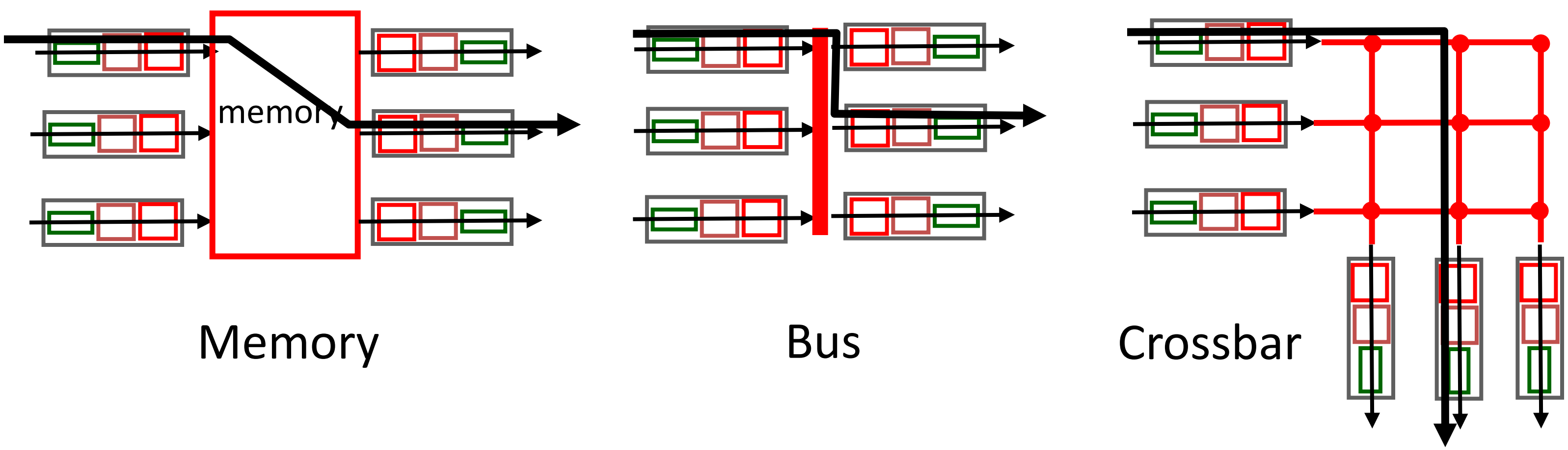
Data link layer:
e.g., Ethernet
see chapter 5

Decentralized switching:

- Given datagram destination, 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

Switching fabrics

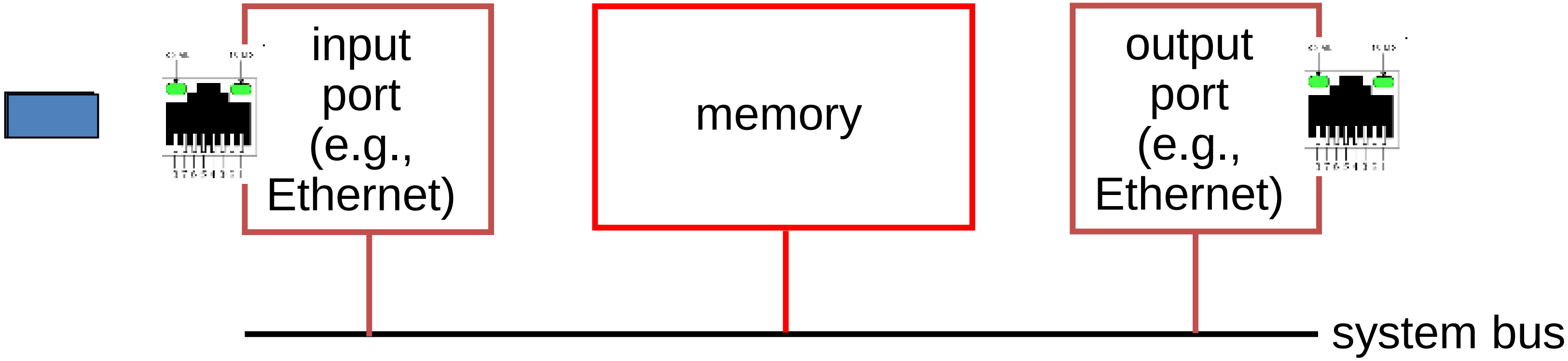
- Transfer packet from input buffer to appropriate output buffer
- Switching rate: rate at which packets can be transfer from inputs to outputs
  - Often measured as multiple of input/output line rate
  - N inputs: switching rate N times line rate desirable
- Three types of switching fabrics:



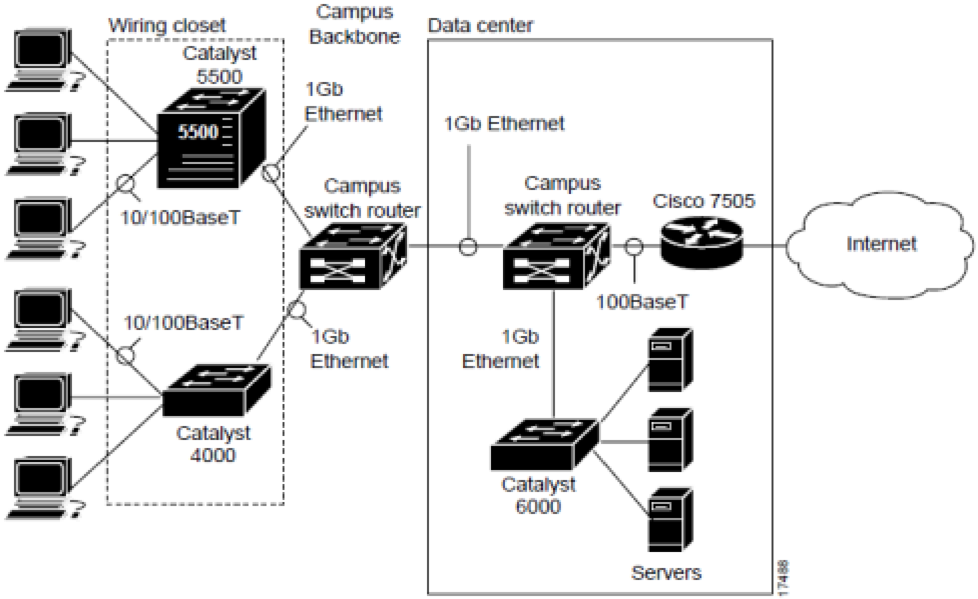
Switching via memory

Switching via memory:

- Traditional computer with switching under CPU control
- Packet copied to system's memory
- Speed limited by memory bandwidth
  - 2 bus crossings per datagram
- e.g. Cisco Catalyst 8500 campus switch router



Cisco Catalyst 8500



$549.95 refurbished
$580.35 with est. tax & shipping
Neobits.com
★★★★★ 147 reviews

Distributed Hardware Forwarding

The Catalyst 8500 campus switch router employs a distributed architecture in which the control path and data path are relatively independent. The control path code, such as routing protocols, runs on the route processor, whereas most of the data packets are forwarded by the Ethernet line module and the switching fabric.

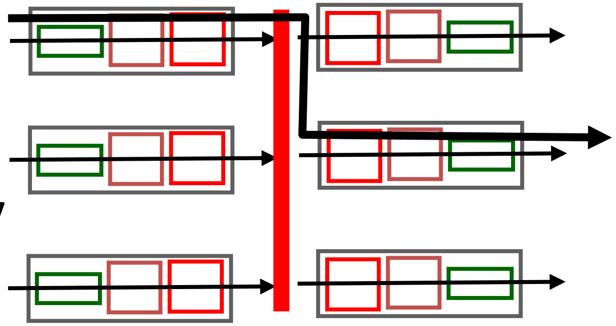
Each line module includes a microcoded processor that handles all packet forwarding. The main functions of the control layer between the routing protocol and the firmware datapath microcode include:

- Managing the internal data and control circuits for the packet forwarding and control functions
- Extracting the other routing and packet forwarding related control information from the Layer 2 and Layer 3 bridging and routing protocols and the configuration data, and then conveying the information to the line module to control the datapath
- Collecting the datapath information, such as traffic statistics, from the line module to the route processor
- Handling certain data packets sent from the Ethernet line modules to the route processor

http://www.cisco.com/univercd/cc/td/doc/product/l3sw/8540/rel_12_0/w5_6f/softcnfg/1cfg8500.pdf

Switching via a bus

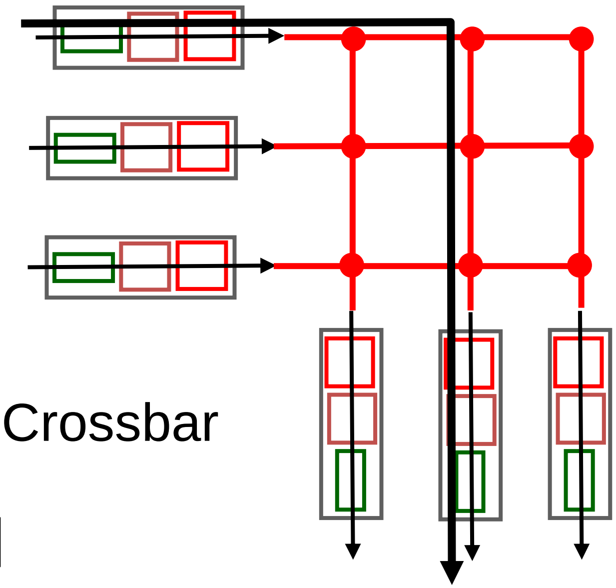
- Shared bus
  - Datagram from input port memory to output port memory via a shared bus
- Bus contention:
  - Switching speed limited by bus bandwidth
- 32 Gbps bus, Cisco 5600
  - Sufficient speed for access and enterprise routers



Bus

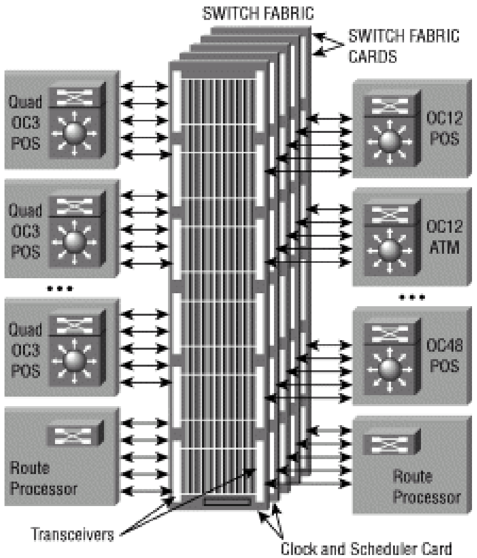
Switching via interconnection network

- Overcome bus bandwidth limit
- Interconnection nets
  - e.g. Banyan networks, crossbar
  - Initially developed to connect processors in multiprocessor
- Advanced design:
  - Fragmenting datagram into fixed length cells, switch cells in fabric
- Cisco 12000:
  - Switches 30-1280 Gbps through the interconnection network



Crossbar

Cisco 12000 Series Internet Router



Switch Fabric

At the heart of the Cisco 12000 Series Internet Router is a multi−gigabit crossbar switch fabric that is optimized to provide high capacity switching at gigabit rates. The crossbar switch enables high performance for two reasons:

- Connections from the line cards to a centralized fabric are point−to−point links that can operate at very high speeds
- Multiple bus transactions can be supported simultaneously, increasing the aggregate bandwidth of the system. The Switch Fabric Card (SFC) receives the scheduling information and clocking reference from the Clock Scheduler Card (CSC), and performs the switching functions. You can imagine the SFC as an NxN matrix where N is the number of slots.

This architecture allows multiple line cards to transmit and receive data simultaneously. The CSC is responsible for selecting which line cards transmit and which line cards receive data during any given fabric cycle.

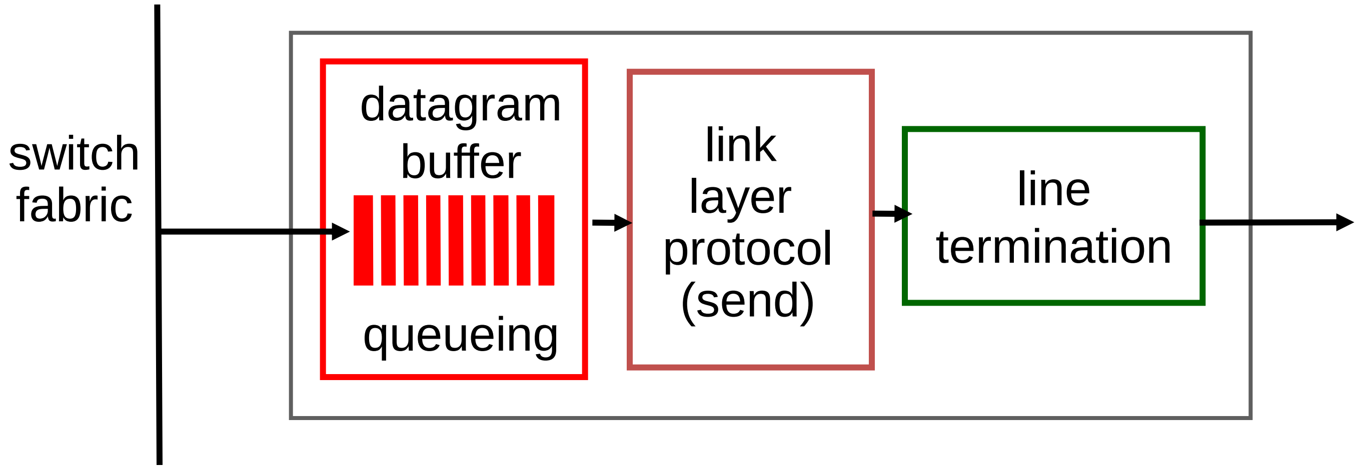
Cisco Cells

The unit of transfer across the crossbar switch fabric is always fixed−size packets, also referred to as Cisco cells, which are easier to schedule than variable−size packets. Packets are broken into cells before being placed on the fabric, and are reassembled by the outbound LC before they are transmitted. Cisco cells are 64 bytes long, with an 8−byte header, a 48−byte payload, and an 8−byte cyclic redundancy check (CRC).

$11,232
from 3 stores
Compare prices

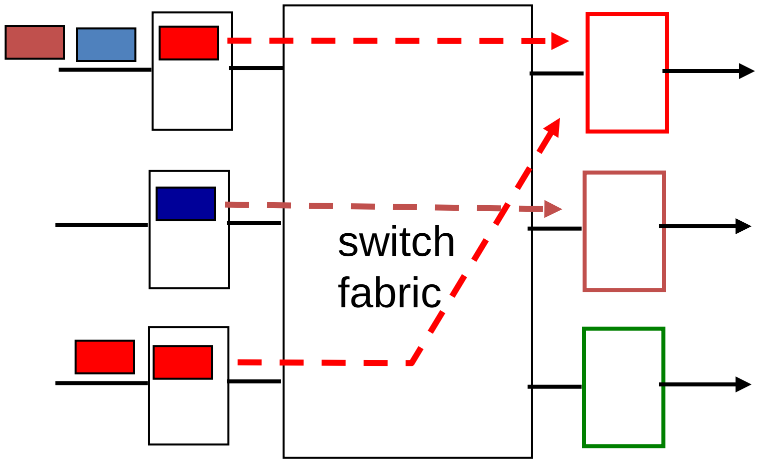
http://www.cisco.com/image/gif/paws/47240/arch12000-swfabric.pdf

Output ports

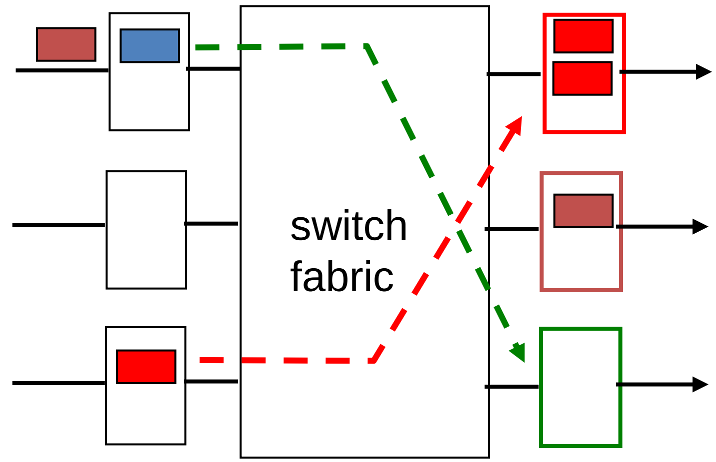


- *Buffering*
  - Required when datagrams arrive from fabric faster than the transmission rate
- *Scheduling discipline*
  - Chooses among queued datagrams for transmission

Output port queueing



At *t*, packets more
from input to output



OIne packet time later

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

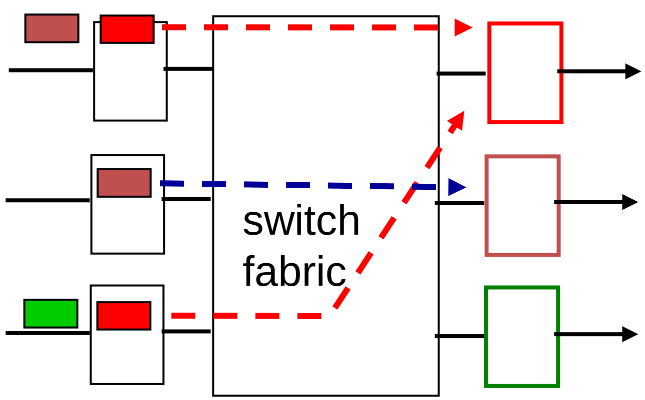
How much buffering?

- 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

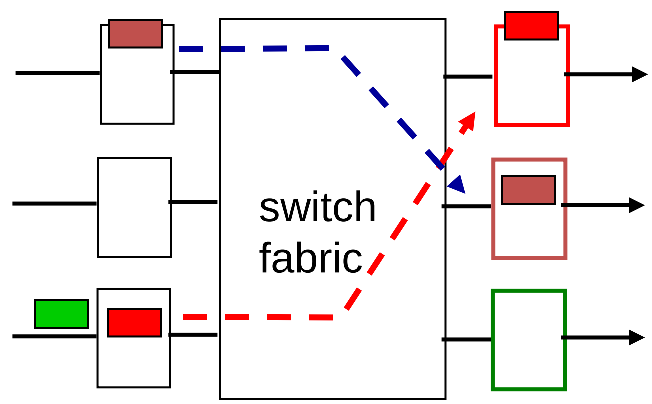
$$\frac{RTT \cdot C}{\sqrt{N}}$$

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



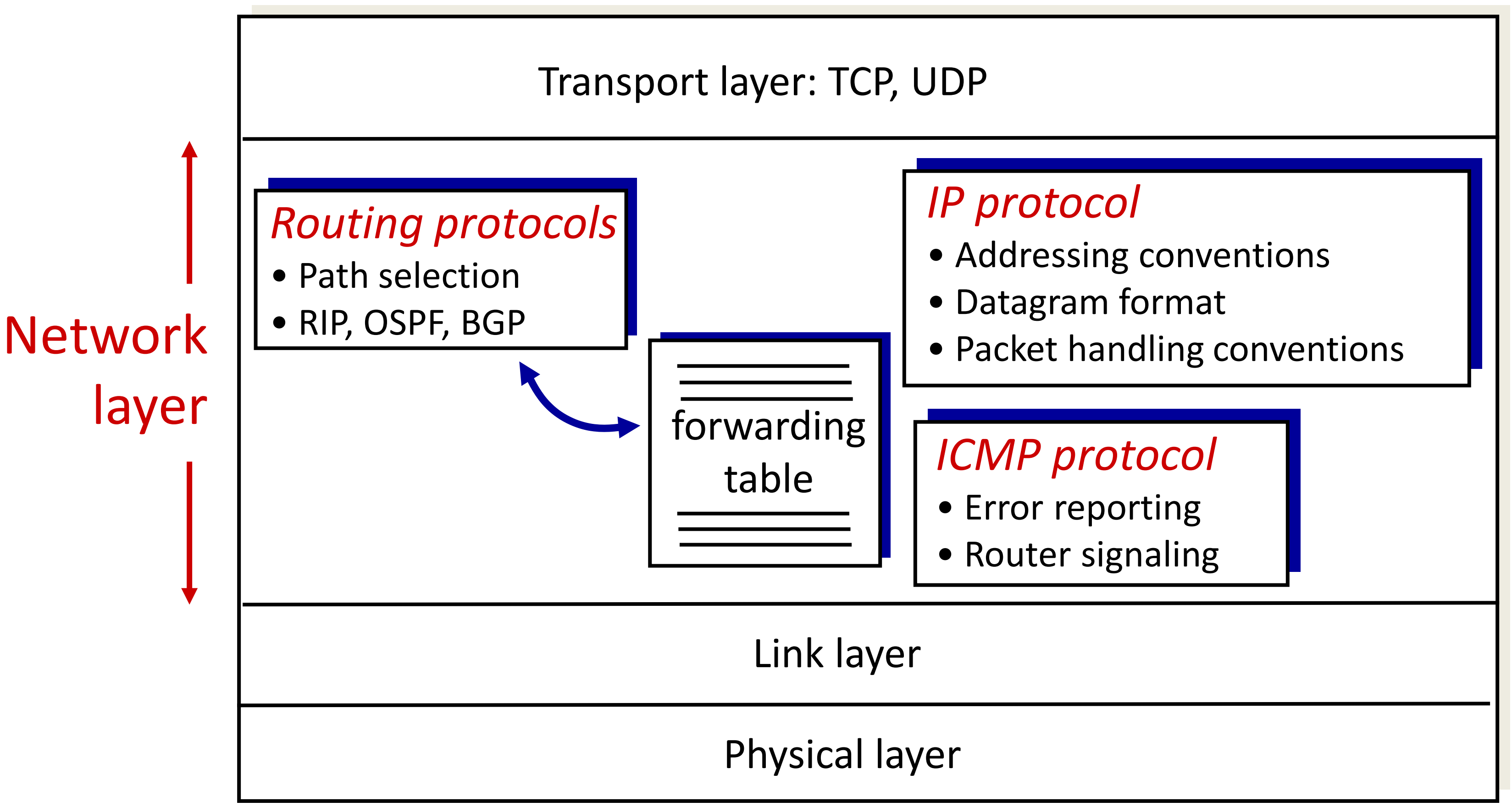
Output port contention:
only one red datagram can be transferred.
Lower red packet is blocked



One packet time later: green packet experiences HOL blocking

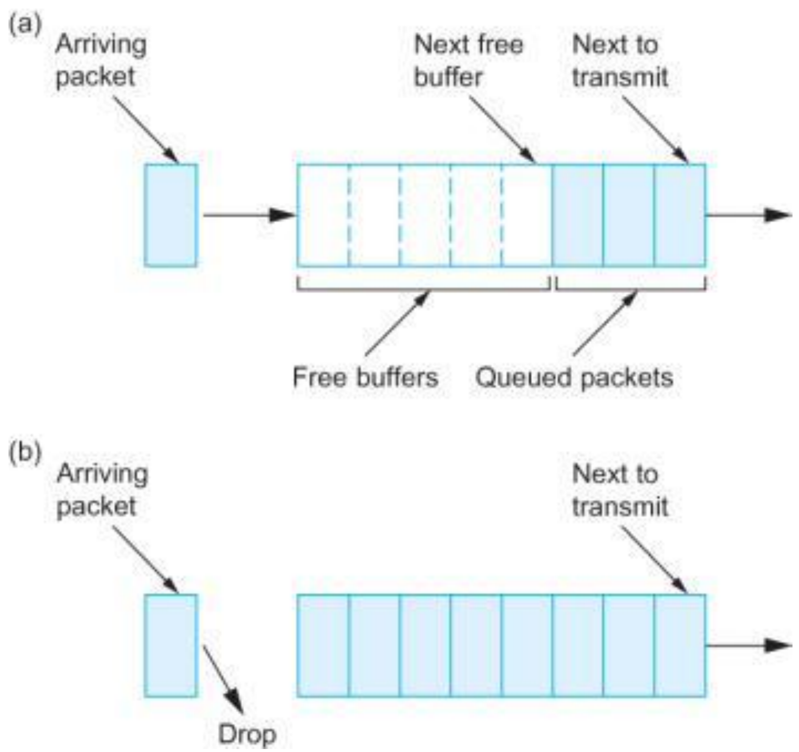
The Internet network layer

Host, router network layer functions:



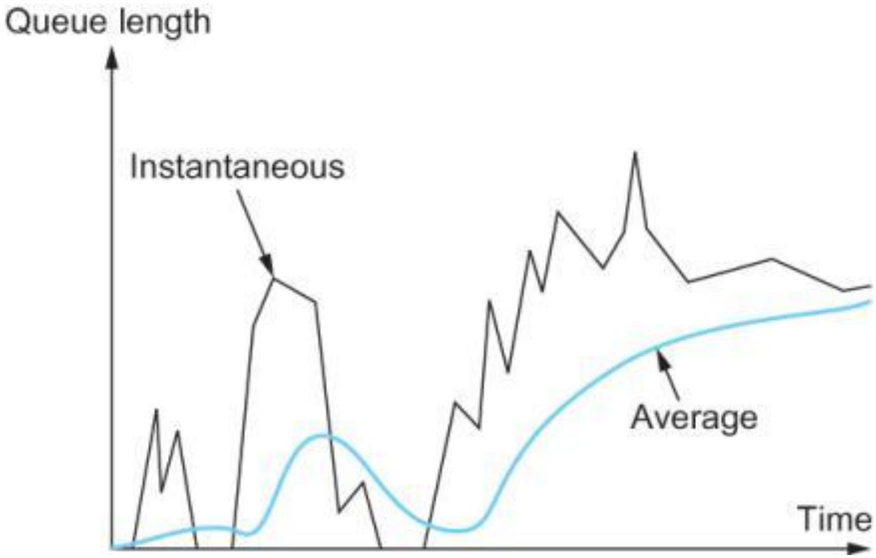
What routers do

- Too many packets arrive too quickly
  – Which packets should we drop?
- First-in first-out (FIFO) with tail drop
  – Simple, drop the new guy that doesn't fit in buffer



Active queue management (AQM)

- Random early detection (RED)
  - If router close to congestion: drop a random packet
  - Source detects packet loss and can adjust send rate
  - Randomness approximates fairness since more likely to signal host sending lots of packets
  - Various parameters controlling drop behavior



Internet Protocol (IP)

- Packet delivery model
  - Connectionless
  - Best-effort (unreliable)
    - Packets may be lost
    - Packets may arrive out of order
    - Duplicate packets may occur
    - Packet may get delayed
- Global addressing scheme
  - How do we identify hosts on the network?

IP history and goals

- Internet Protocol (IP)
  - 1974 Cerf and Kahn propose common layer hiding network differences
  - Eventually split into TCP and IP
  - IP foundation of the modern Internet
  - Awarded 2004 Turing Award

A Protocol for Packet Network Intercommunication

VINTON G. CERF AND ROBERT E. KAHN,
MEMBER, IEEE

Abstract — A protocol that supports the sharing of resources that exist in different packet switching networks is presented. The protocol provides for variation in individual network packet sizes, transmission failures, sequencing, flow control, end-to-end error checking, and the creation and destruction of logical process-to-process connections. Some implementation issues are considered, and problems such as internetwork routing, accounting, and timeouts are exposed.

of one or more *packet switches*, and a collection of communication media that interconnect the packet switches. Within each HOST, we assume that there exist *processes* which must communicate with processes in their own or other HOSTS. Any current definition of a process will be adequate for our

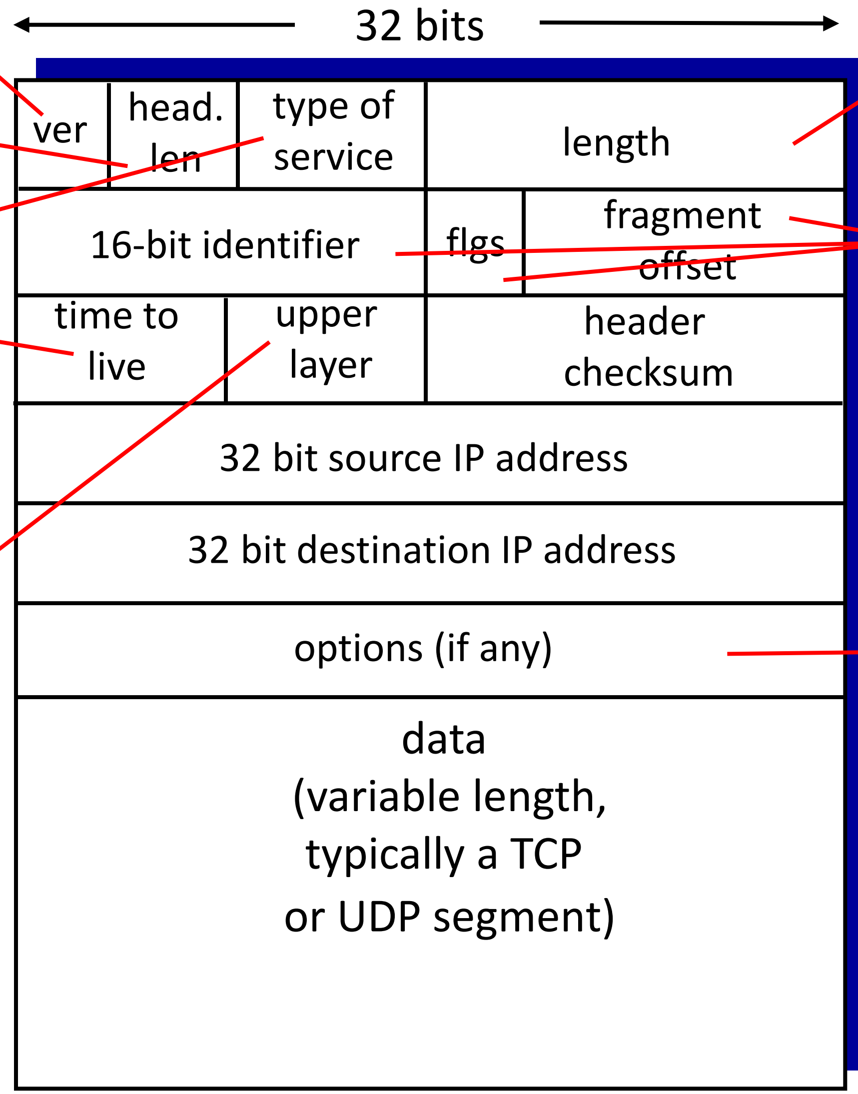
IP history and goals

- Connect existing networks
  - Multiplex existing links such as radio networks
- Motivating application
  - Remote login to servers
    - Inherently bursty traffic, long silence periods
- Robust to failures
  - Survive equipment failure or attack
  - Traffic routes around trouble

IP history and goals

- Support multiple types of services
  - Differing requirements for speed, latency, reliability
- Heterogeneous networks
  - Minimal assumptions about underlying network
- Distributed management of resources
  - Node managed by different institutions
- Cost effective
  - Packet switched, share links via multiplexing

IP datagram format



How much overhead?

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

Summary

- Datagram networks
  - No call setup, no fixed path
  - Forwarding via longest prefix match
- Inside a router
  - Runs routing algorithms
  - Forward datagrams from incoming to outgoing link
  - Switching fabric, buffering
- Introduction to IP protocol
  - History and goals