Chapter 4:
Network Layer Data Plane

Chapter objectives:

- understand principles behind network layer services, with a focus on the data plane:
  - network layer service models
  - forwarding versus routing
  - Addressing
  - generalized forwarding
  - Internet architecture
- implementation in the Internet
  - IP protocol
  - NAT, middleboxes

Lecture 19

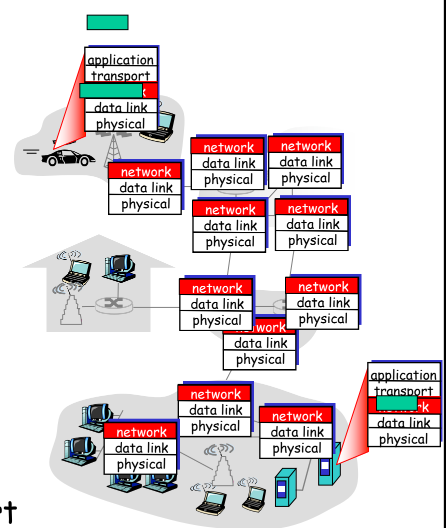
- Sections 4.1, 4.1.1, 4.1.2, 4.2, 4.2.1, 4.2.2
- Overview of network layer
  - Forwarding and control planes
  - Network service models
- What is inside a router
  - Input port processing & destination-based forwarding
  - Switching

Chapter 4: Network Layer

4. 1 Introduction

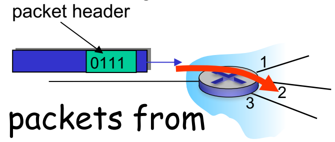
Network layer

- network layer protocols in *every* host, router
- data unit is packet or datagram (independent from other data units)
- transports segment from sending to receiving host
- sending side: encapsulates segments into packet
- network routers: examines header fields in IP packet & sends packet on an output port
- rcving side: delivers segments to transport layer



Two Key Network-Layer Functions

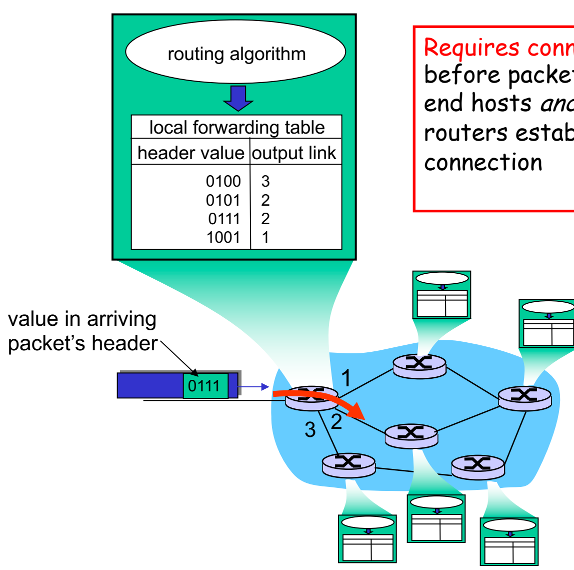
- *forwarding:* move packets from router's input to appropriate router output
  - router (or switch) does not determine route for each packet -> route is determined for all packets in a flow at the beginning, and router forwards packets on predetermined route



- *routing:* determine route taken by packets from source to destination
  - *routing algorithms* are used by every router to determine route -> route may be computed for each packet independently

5

Interplay between routing and forwarding



Requires connection setup: before packets flow, two end hosts *and* intervening routers establish virtual connection

6

Network service model

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

example services for individual packets:

- 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

Channel may not offer any guarantees – best effort service

Network layer service models:

Network Architecture	Service Model	Guarantees ? Bandwidth	Loss	Order	Timing	Congestion feedback
Internet	best effort	none	no	no	no	no (inferred via loss)

Internet "best effort" service model

No guarantees on:

i. successful datagram delivery to destination
ii. timing or order of delivery
iii. bandwidth available to end-end flow

Chapter 4: Network Layer

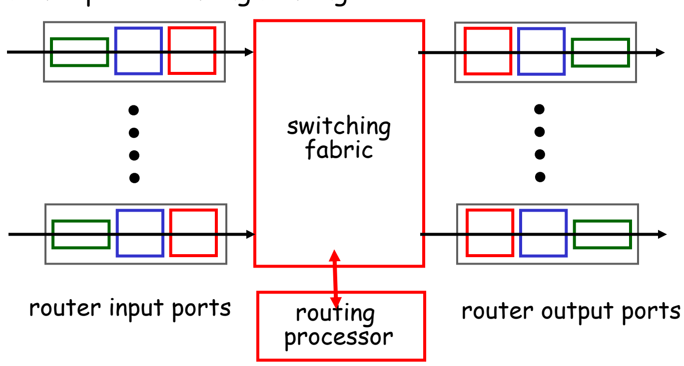
4. 1 Introduction

4.2 What's inside a router?

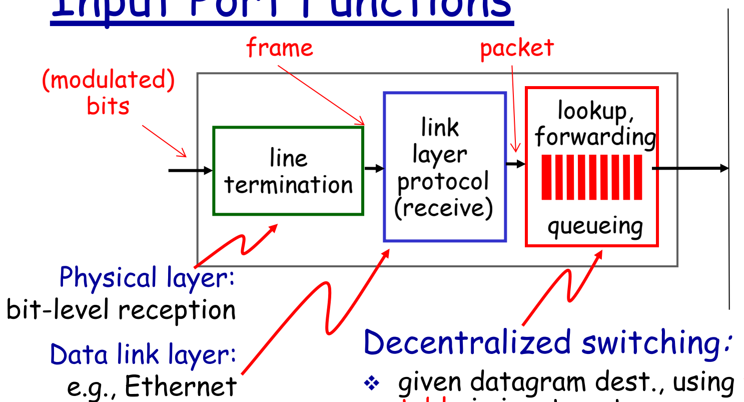
Router Architecture Overview

two key router functions:

- run routing algorithms/protocol (RIP, OSPF, BGP):
  - construct routing tables: which port leads to destination
- *forwarding* datagrams from incoming to outgoing link:
  - route packets using routing table



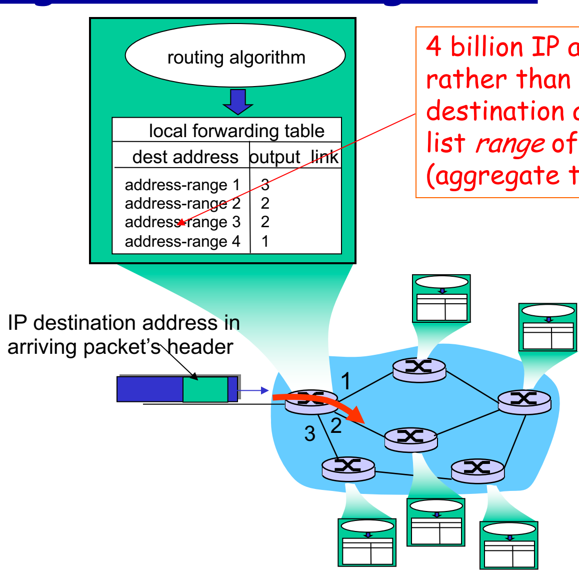
Input Port Functions



Decentralized switching:

- given datagram dest., using forwarding table in input port memory, lookup output port
- goal: complete input port processing at 'line speed'
- queuing: if datagrams arrive faster than forwarding rate into switch fabric

Datagram Forwarding table



4 billion IP addresses, so rather than list individual destination address list *range* of addresses (aggregate table entries)

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

13

Longest prefix matching

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

examples:

11001000 00010111 00010110 10100001 which interface?

11001000 00010111 00011000 10101010 which interface?

14

Longest prefix matching

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!

examples:

11001000 00010111 00010110 10100001 which interface?

11001000 00010111 00011000 10101010 which interface?

15

Longest prefix matching

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!

examples:

11001000 00010111 00010110 10100001 which interface?

11001000 00010111 00011000 10101010 which interface?

16

Longest prefix matching

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!

examples:

11001000 00010111 00010110 10100001	which interface?
11001000 00010111 00011000 10101010	which interface?

17

Longest prefix matching

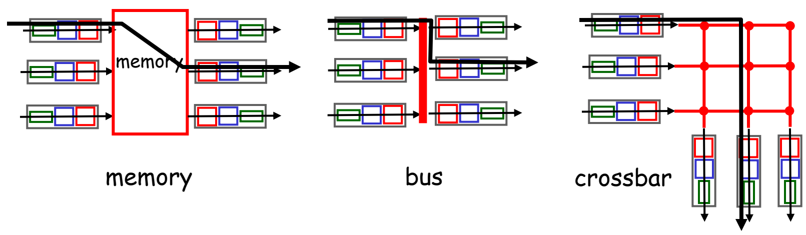
- we'll see *why* longest prefix matching is used shortly, when we study addressing
- longest prefix matching: often performed using ternary content addressable memories (TCAMs)
  - *content addressable:* present address to TCAM: retrieve address in one clock cycle, regardless of table size
  - Cisco Catalyst: ~1M routing table entries in TCAM

18

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

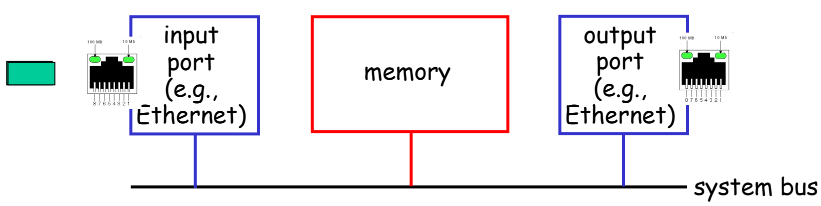
  e.g., N inputs: switching rate N times line rate desirable
- three types of switching fabrics



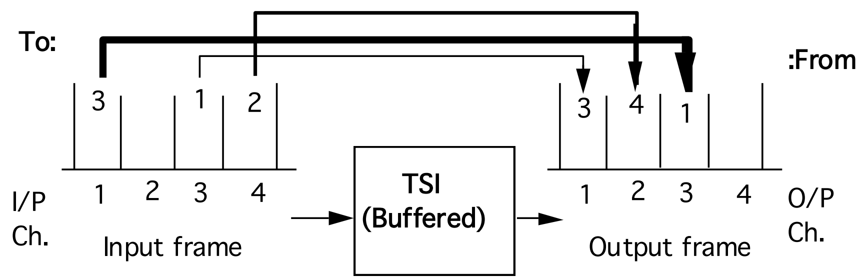
Switching Via Memory

First generation routers:

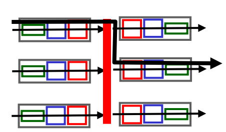
- traditional computers with switching under direct control of CPU
- packet copied to system's memory
- speed limited by memory bandwidth (2 bus crossings per datagram)



Switching via memory: Time-Slot Interchanging (TSI)



Switching Via a Bus

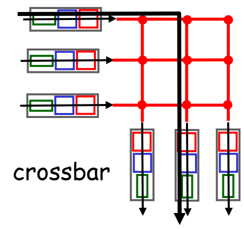


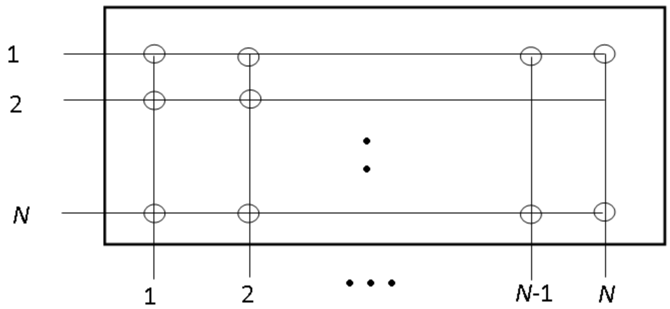
bus

- datagram from input port memory to output port memory via a shared bus
  - input ports contend in transmitting packets on the shared bus, after adding output port numbers to packets
  - each output port listens to packets transmitted on bus, and extracts those with matching output port numbers
- bus contention: switching speed limited by bus bandwidth

Switching Via An Interconnection Network (space swtiching)

- overcome bus bandwidth limitations
- Banyan networks, crossbar, other interconnection nets initially developed to connect processors in multiprocessor





Crossbar switch