DTS205TC High Performance Computing

Lecture 4 Network 2

Di Zhang, Spring 2024

Objectives of today's lesson
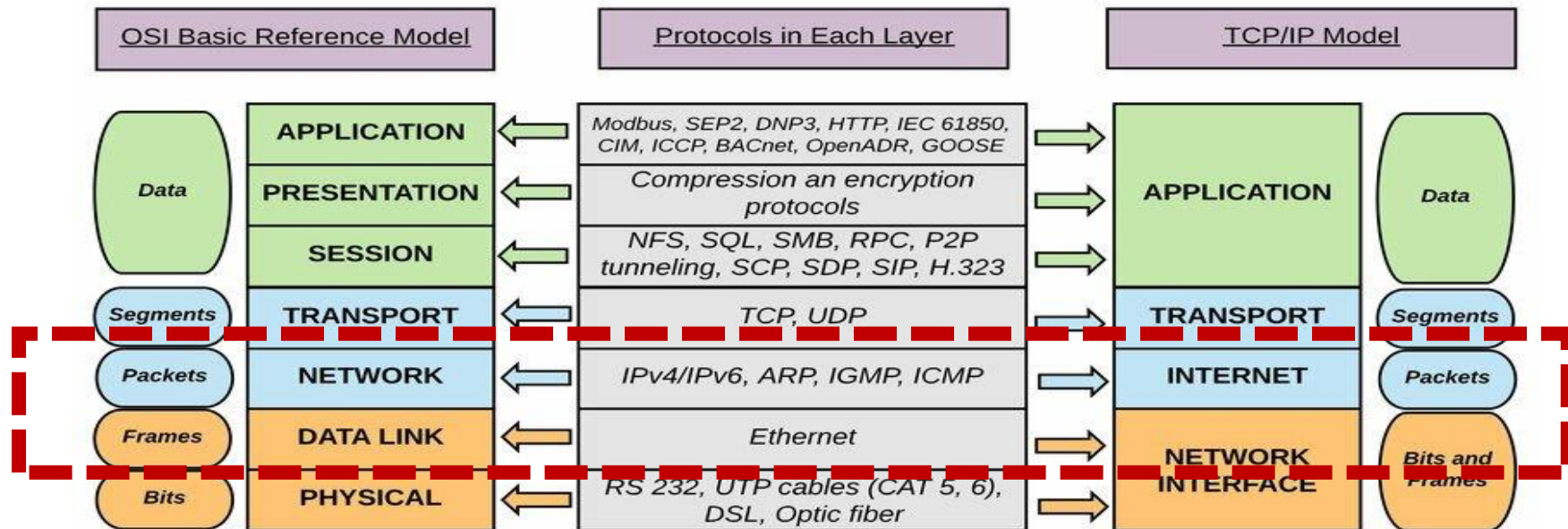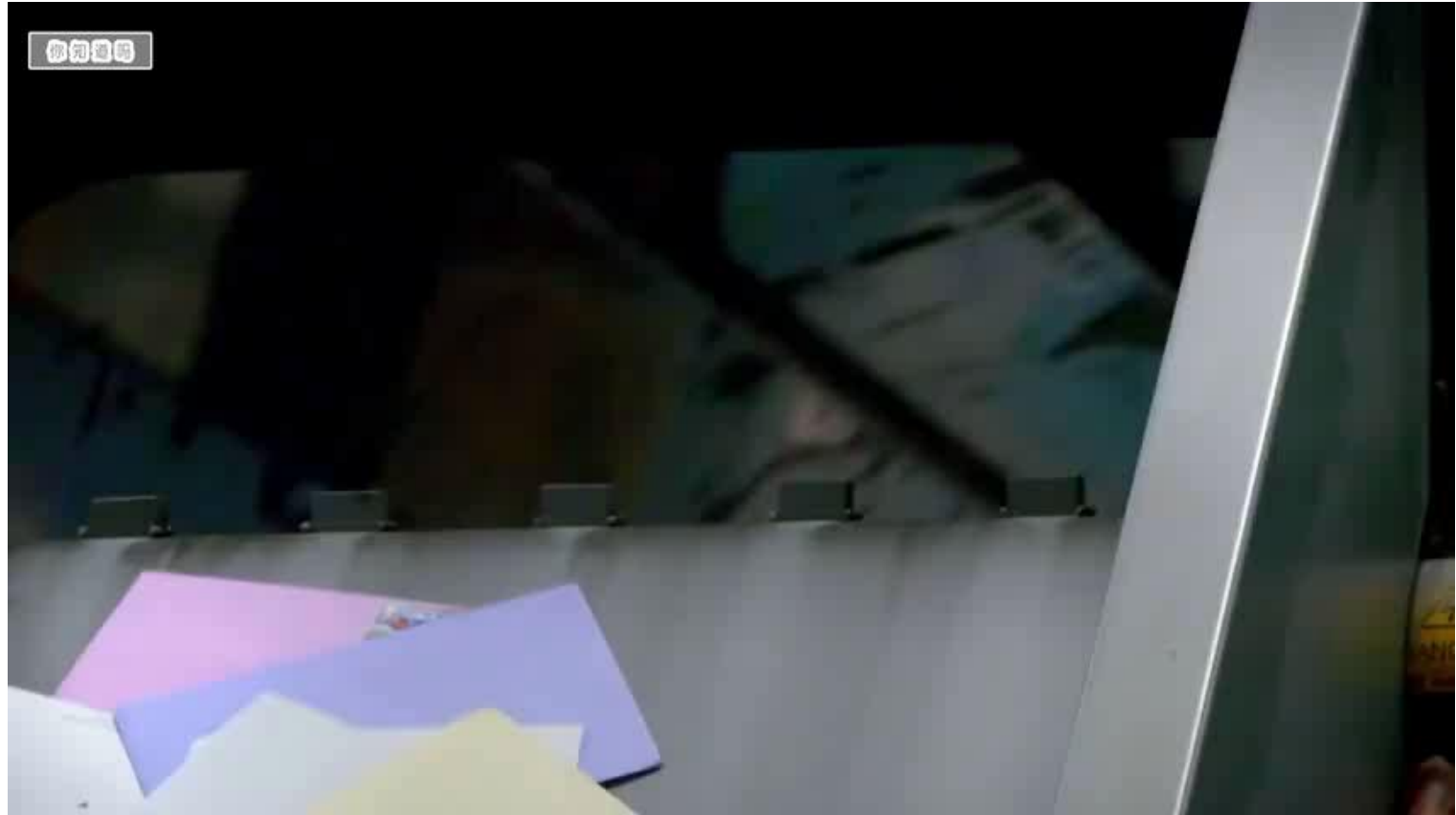
OSI Basic Reference Model		Protocols in Each Layer	TCP/IP Model	
Data	APPLICATION	Modbus, SEP2, DNP3, HTTP, IEC 61850, CIM, ICCP, BACnet, OpenADR, GOOSE	APPLICATION	Data
	PRESENTATION	Compression an encryption protocols		
	SESSION	NFS, SQL, SMB, RPC, P2P tunneling, SCP, SDP, SIP, H.323		
Segments	TRANSPORT	TCP, UDP	TRANSPORT	Segments
Packets	NETWORK	IPv4/IPv6, ARP, IGMP, ICMP	INTERNET	Packets
Frames	DATA LINK	Ethernet	NETWORK INTERFACE	Bits and Frames
Bits	PHYSICAL	RS 232, UTP cables (CAT 5, 6), DSL, Optic fiber		

Three networking methods to know

| Type	Device	Purpose	Scope	Entity	Address	Delivery Method
Wide Area Network	**Router**	**Connected to Internet**	**Global**	**Host with an IP address**	**IP**	**Forward along network**
Local Area Network	Switch	Company/School	Building	Node with a net card	MAC	Forward along network
Local Area Network	Hub	Family	Home	Node with a net card	MAC	Broadcast

- Analogy between WAN and postal network

  - Both sender and receiver must have an address

  - Hierarchical Address/Zip Code

  - Collect messages from streets and cities to provinces, and then disperse to cities and streets

Three networking methods to know

Type	Device	Purpose	Scope	Entity	Address	Delivery Method
Wide Area Network	Router	Connected to Internet	Global	Host with an IP address	IP	Forward along network
Local Area Network	**Switch**	**Company/School**	**Building**	**Node with a net card**	**MAC**	**Forward along network**
Local Area Network	Hub	Family	Home	Node with a net card	MAC	Broadcast

- Analogy between LAN by switch and Mailbox

  - The receiving and sending parties only need to have their names (assuming there are no duplicate names)

  - Address format: only one level

  - Forwarding, usually up to two levels

Three networking methods to know

Type	Device	Purpose	Scope	Entity	Address	Delivery Method
Wide Area Network	Router	Connected to Internet	Global	Host with an IP address	IP	Forward along network
Local Area Network	Switch	Company/School	Building	Node with a net card	MAC	Forward along network
Local Area Network	**Hub**	**Family**	**Home**	**Node with a net card**	**MAC**	**Broadcast**

- Analogy between LAN by hub and loudspeaker

  - Broadcast by name without any forwarding

Section

- Network layer

How to design an Internet?

- The Inter-net is a global network that comprises many voluntarily interconnected autonomous networks.

- Key: TCP/IP Protocol
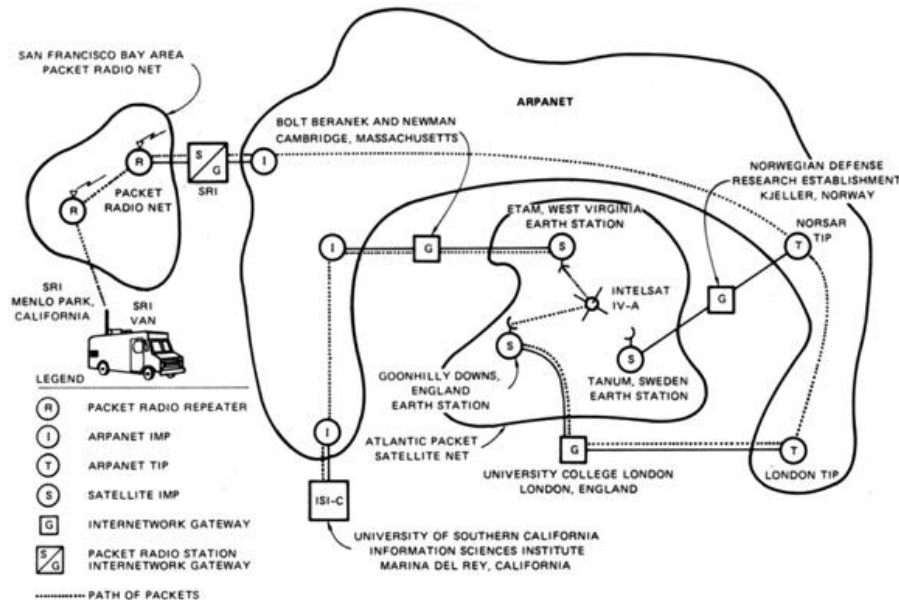
the first internetworked connection

Design goals

- Functionality

  - TCP: connection, connectionless

  - IP: forwarding and routing

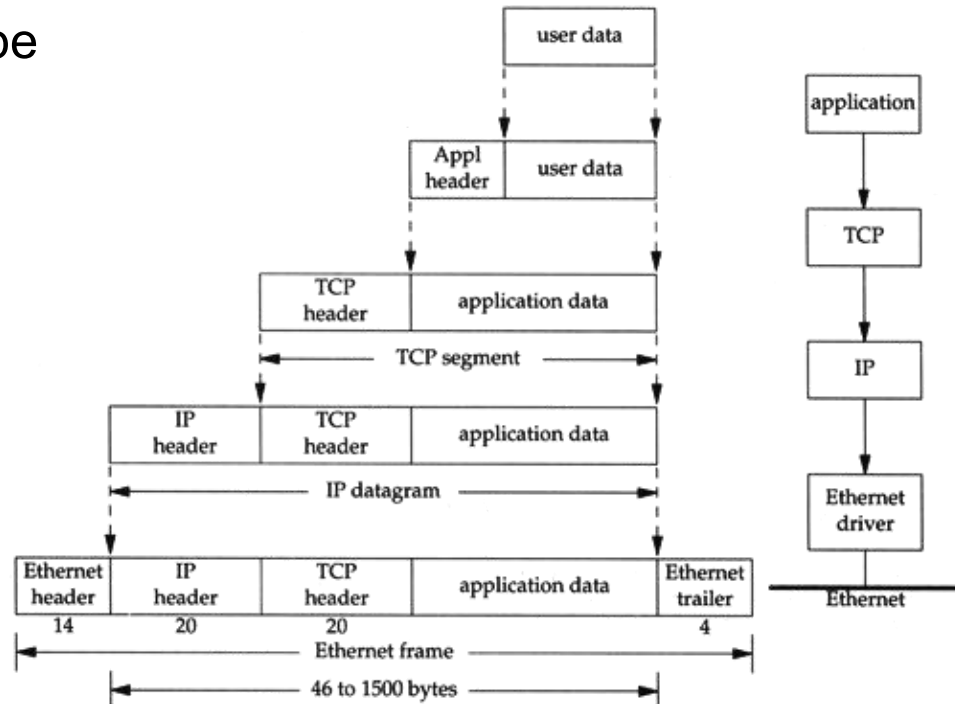 - ICMP: control

  - IGMP: multicast
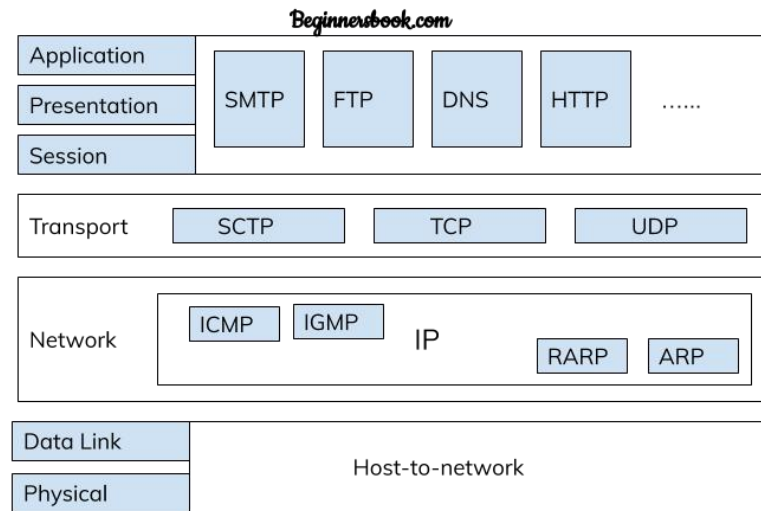
Design goals

- Non-Functionality (DFX)

  - Simplicity：

    - Only has necessary functions for communication

    - Does not involve hardware, application

Design goals

- Extensibility

  - In theory, TCP and IP can be independent of each other

Design goals

- Scalability：

  - Support billions of devices

- Performance

  - Fast Processing
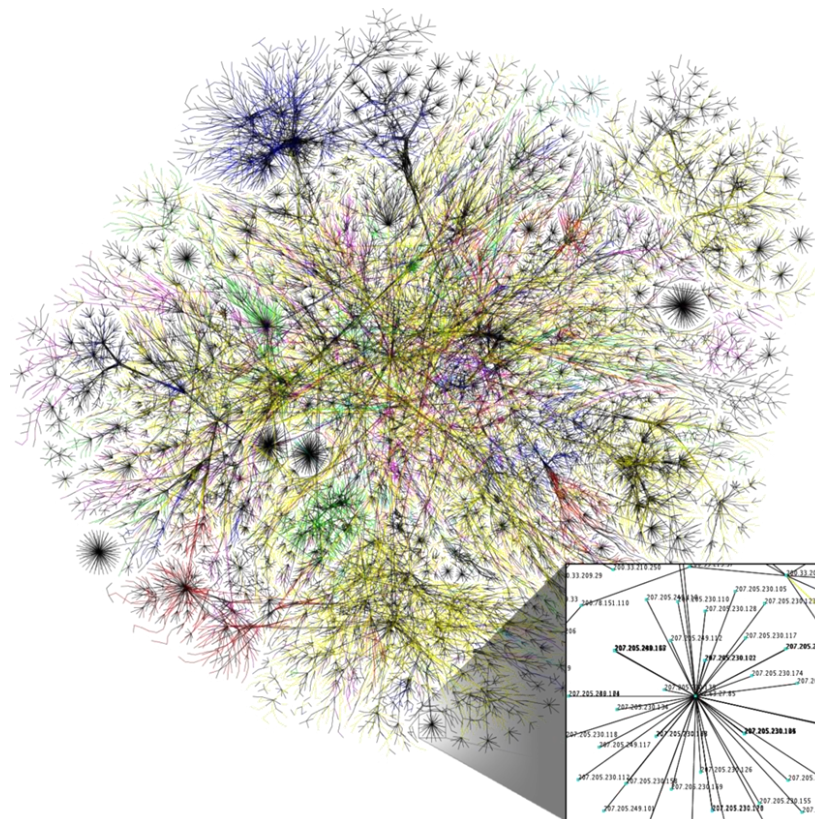
Visualization of Internet routing paths

Network layer: forwarding and routing

- Differences from postal mail:
Digitization

  - The address and letter format must be standardized
  Large number of messages and high speed required (80Tbps for a typical router)
  Address allocation order, which does not imply a geographic hierarchy
  The address changes dynamically
  Routing paths are diverse and dynamic

IPv4 Address Map of Year 2016

EUROPE NORTH AMERICA SOUTH AMERICA OCEANIA

ASIA AFRICA ANTARCTICA UNASSIGNED

RESERVED
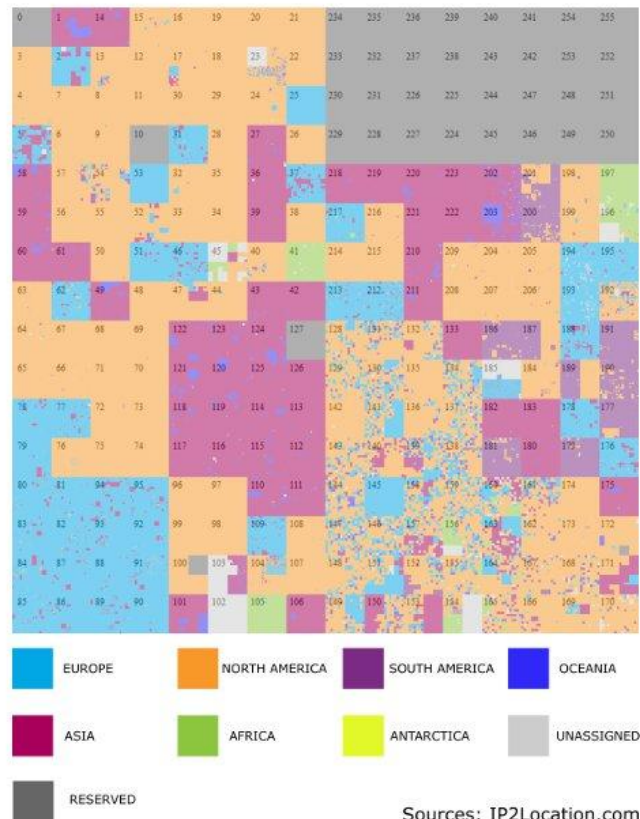
Sources: IP2Location.com

IP Datagram format

IP protocol version number

header length(bytes)

"type" of service:
- diffserv (0:5)
- ECN (6:7)

TTL: remaining max hops (decremented at each router)

upper layer protocol (e.g., TCP or UDP)

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

32 bits

ver	head. len	type of service	length
16-bit identifier		flgs	fragment offset
time to live	upper layer	header checksum	
source IP address			
destination IP address			
options (if any)			
payload data (variable length, typically a TCP or UDP segment)			

total datagram length (bytes)

fragmentation/ reassembly

header checksum

32-bit source IP address

32-bit destination IP address
e.g., timestamp, record route taken

Maximum length: 64K bytes
Typically: 1500 bytes or less

- Why not use XML as the format?

```xml
<?xml version="1.0"?>
<quiz>
 <qanda seq="1">
  <question>
   Who was the forty-second
   president of the U.S.A.?
  </question>
  <answer>
   William Jefferson Clinton
  </answer>
 </qanda>
 <!-- Note: We need to add
 more questions later.-->
</quiz>
```

XML

Forwarding table

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

Hierarchical addressing: route aggregation

Organization 0
200.23.16.0/23

Organization 1
200.23.18.0/23

Organization 2
200.23.20.0/23

Organization 7
200.23.30.0/23

Fly-By-Night-ISP

"Send me anything with addresses beginning 200.23.16.0/20"

ISPs-R-Us

"Send me anything with addresses beginning 199.31.0.0/16"

Internet
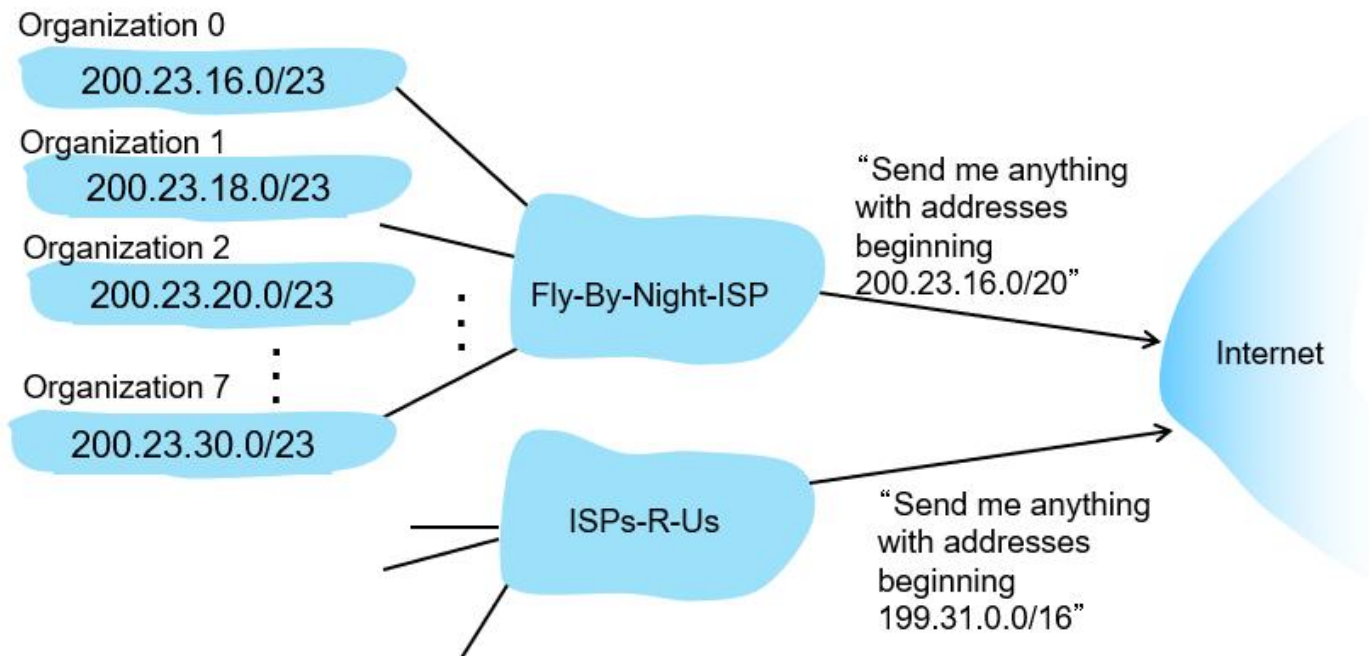
Implementation of router

- Knowing the above principle, you can even implement a router with a PC However, as mentioned in Lec. 1, the general von Neumann architecture has slower performance

Routers in reality
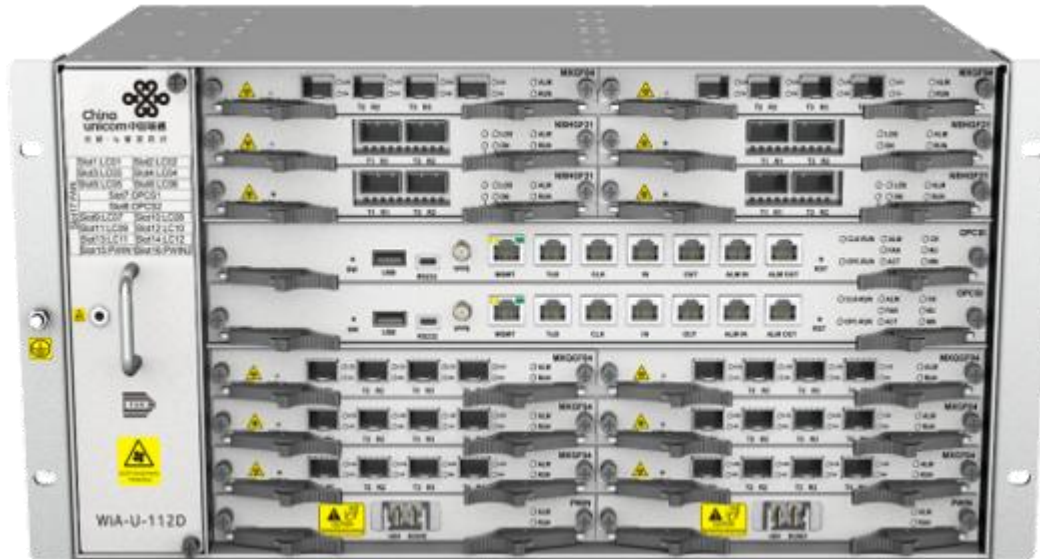
- Note that the wireless router in daily life is actually a link-layer device, not the router mentioned here

Routers in reality

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

forwarding data plane
(hardware) operates
in nanosecond
timeframe

routing processor

high-speed switching fabric

router input ports

router output ports
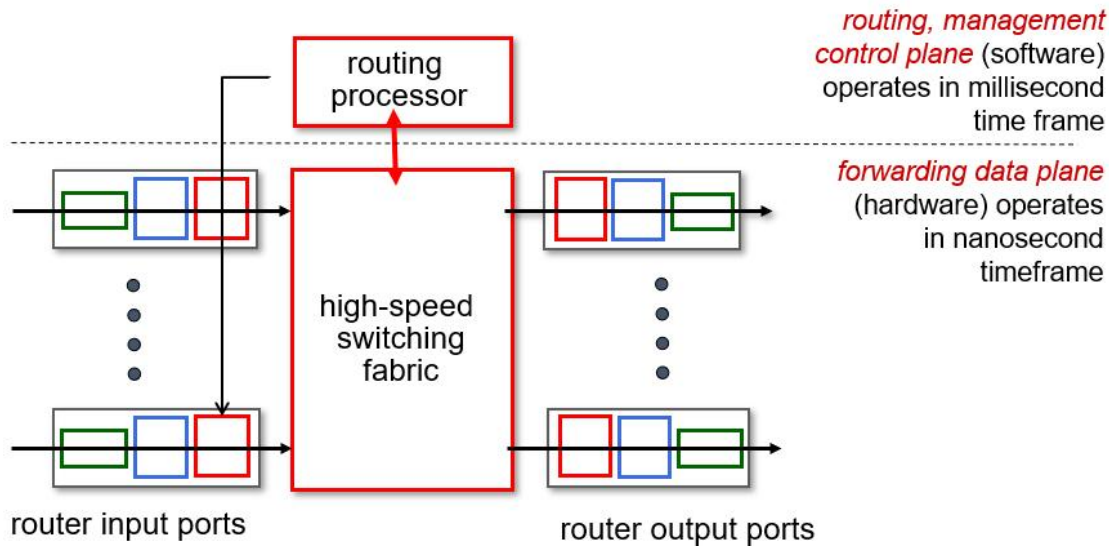
- Input Port: Parses packets and determines the forwarding destination
  Switch fabric：Message transmission

- Output Port: Packet encapsulation

Switching fabrics

memory

bus

interconnection network
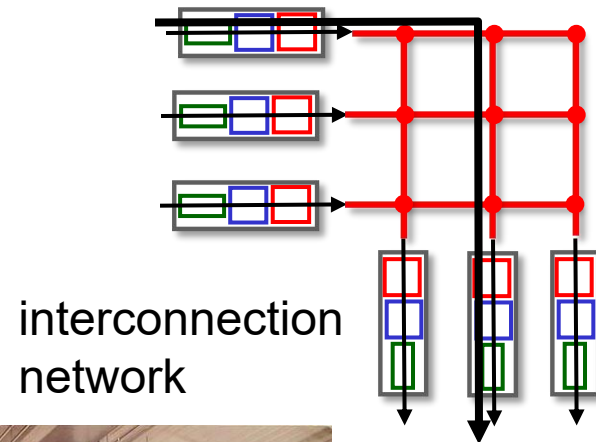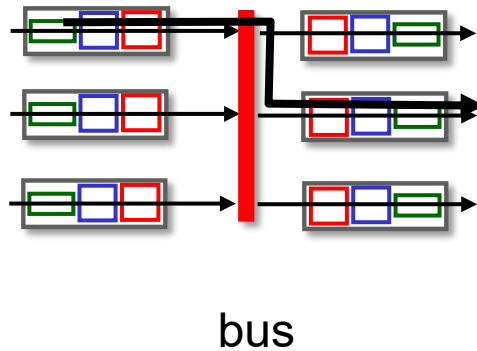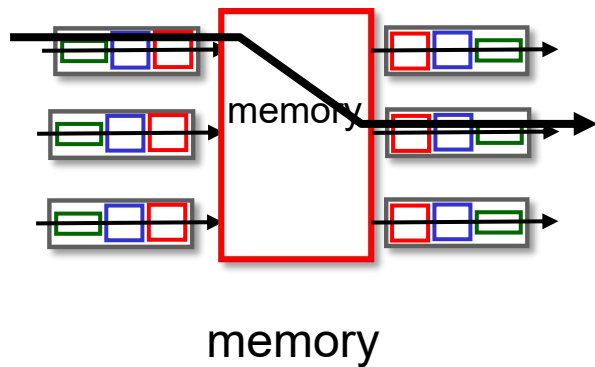
- Network inside Network
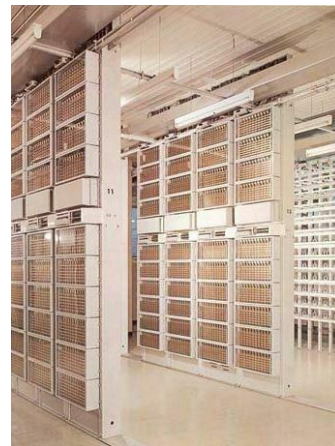
Mechanical Crossbar

Generation of forwarding table

- By exchanging information between routers
  By a centralized controller
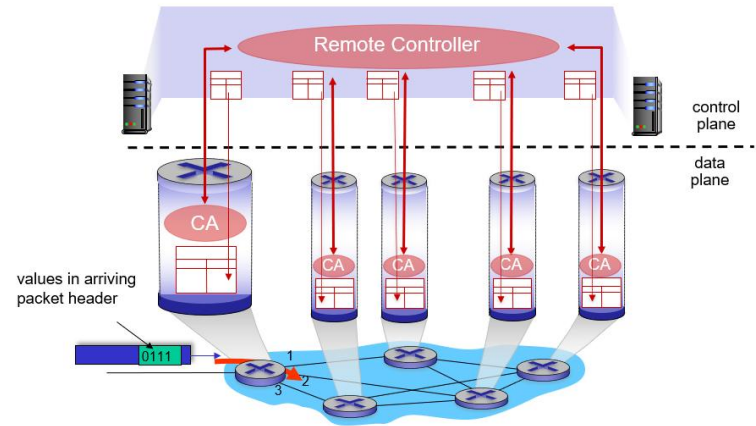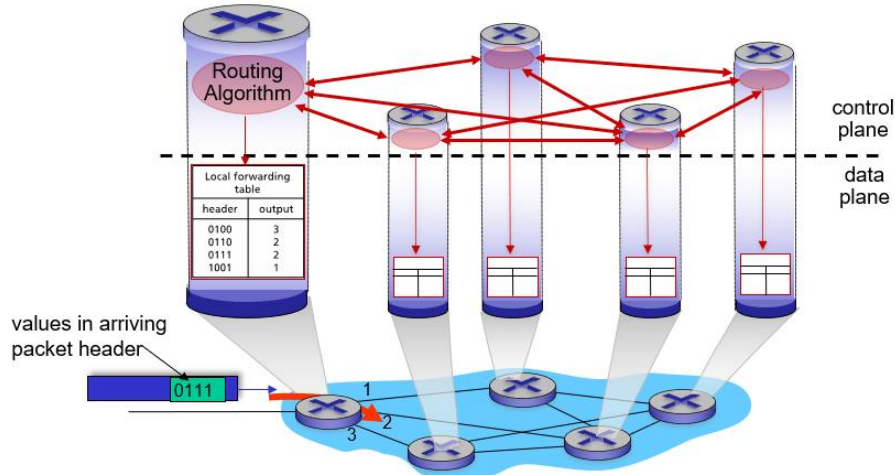
- The forwarding table in each router must provide a forwarding destination for all IP addresses; That is, the forwarding table provides a holistic picture of the Internet's connectivity based on a local perspective

Section

- Link layer

- There are three phases:

  - Take responsibility that IP protocol leave blank, but needed to drive the hardware

    - Error detection, retransmission, etc.

  - Multiple nodes share a link

    - Provide a simple, inexpensive way

  - Link layer switches

    - Straightforward and easy to use

MAC Addresses are physically attached to a hardware device.

Ethernet datagram

type

| preamble | dest. address | source address | | data (payload) | CRC |
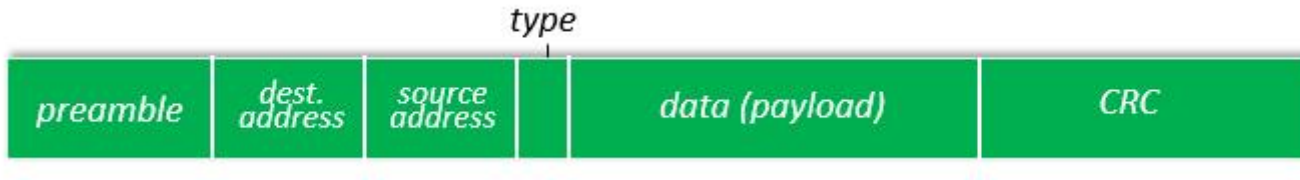
- Synchronization, sending and receiving addresses, verification

Multiple access links

- broadcast (shared wire or medium)

  - old-fashioned Ethernet

shared wire (e.g., cabled Ethernet)

shared radio: 4G/5G

shared radio: WiFi
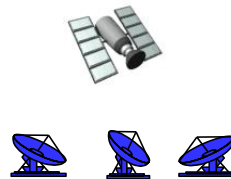
shared radio: satellite

humans at a cocktail party (shared air, acoustical)

CSMA/CD

- Carrier Sense Multiple Access with Collision Detection
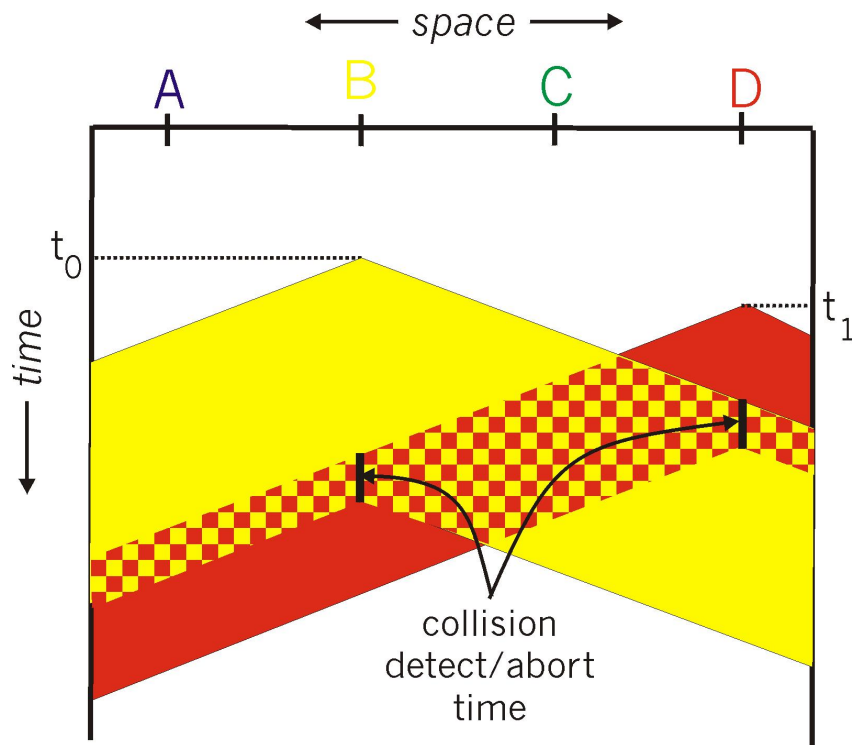
- Transfer -> Detect Collision -> Wait a random period -> retry

Link layer switch

Address	Interface	Time
62-FE-F7-11-89-A3	1	9:32
7C-BA-B2-B4-91-10	3	9:36
....	....	....

- Forwarding table

  - Strictly matches the destination address one by one
    Self-learning capability, no configuration required

Comparison of devices

- Comparison of the typical features of popular interconnection devices

	Hubs	Routers	Switches
Traffic isolation	No	Yes	Yes
Plug and play	Yes	No	Yes
Optimal routing	No	Yes	No

Topology of usual networks
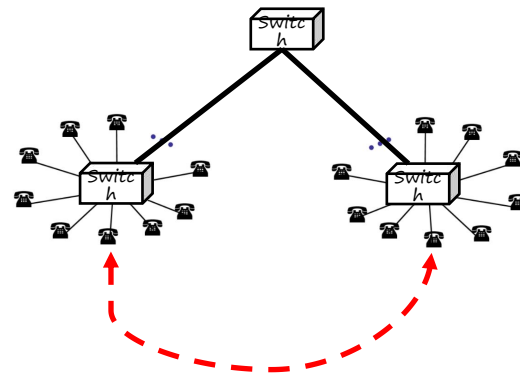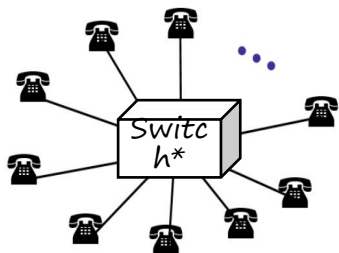
- Topology

  - describes how elements of a set relate spatially to each other (wiki)

- Star: The simplest structure (* -- linked by a switch with an internal bus)

- Tree: Direct contact between two leaves will be slow. For general needs, such as surfing Web, direct contact is rarely required, but it is not for HPC.
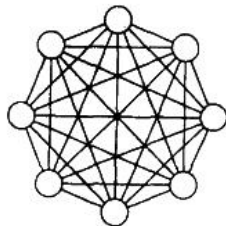
Other possible network structures
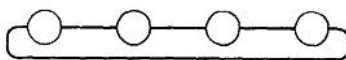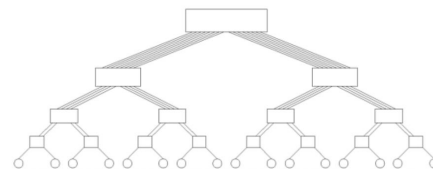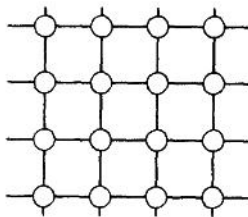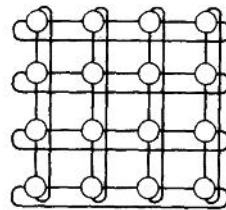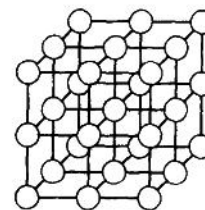

Linear   Loop   Fat Tree

Fully Connected   2-D mesh   2-D wraparound mesh   3-D mesh

- Balance between cost and performance:

  - More connections always make communication faster

  - The higher the number of connections, the higher the cost

- Obviously, some structures look more "reasonable" and "connected" than others

  - How to quantify this concept?