Network Layer: Internet Protocol

2021/22 COMP3234B & ELEC3443B

Contents

- Overview of services in Network Layer
- Internet Protocol (IP)
  - Packet header
  - IP fragmentation and reassembly
  - IP addressing
    - CIDR – Classless InterDomain Routing
    - Subnetting and Supernetting

Learning Outcomes

- **[ILO2 - Technologies and Protocols]** be able to describe the working principles behind key network technologies and protocols used in modern computer networks.
  - ILO 2b - **Network Layer:** comprehend and explain the following network-layer technologies and concepts : **Internet protocol, IP addressing**, ICMP, NAT, DHCP and IPv6

- **[ILO5 - Practicability]** be able to plan for IP networks and properly assign IP addresses to interfaces in given networks – *IP addressing & forwarding*

Reading

- Chapter 4 of Computer Networking – A Top-Down Approach Featuring the Internet, 7th edition by J. Kurose et. Al
  - Sections 4.1, 4.2, 4.2.1, 4.3, 4.3.1, 4.3.2, 4.3.3

Good Reference

- TCP/IP Guide
  - TCP/IP Internet Layer (OSI Network Layer) Protocols
  - http://www.tcpipguide.com/free/t_NetworkLayerProtocols.htm

Network Layer

- To support **host-to-host** communication
  - The most important service is to **find the path (i.e., determine the route) to deliver the packets** from the source host to the destination host
  - Challenge 1: need to deal with very large scale of networks and how can packets find the most efficient route through the network
    - billions of end-systems in geographically distributed networks
  - Challenge 2: need to deal with heterogeneity
    - connectivity between hosts could involve different physical technologies
      - E.g., one end uses WiFi and the other end on 5G network
      - involving different physical addressing schemes, different packet sizes, service model…

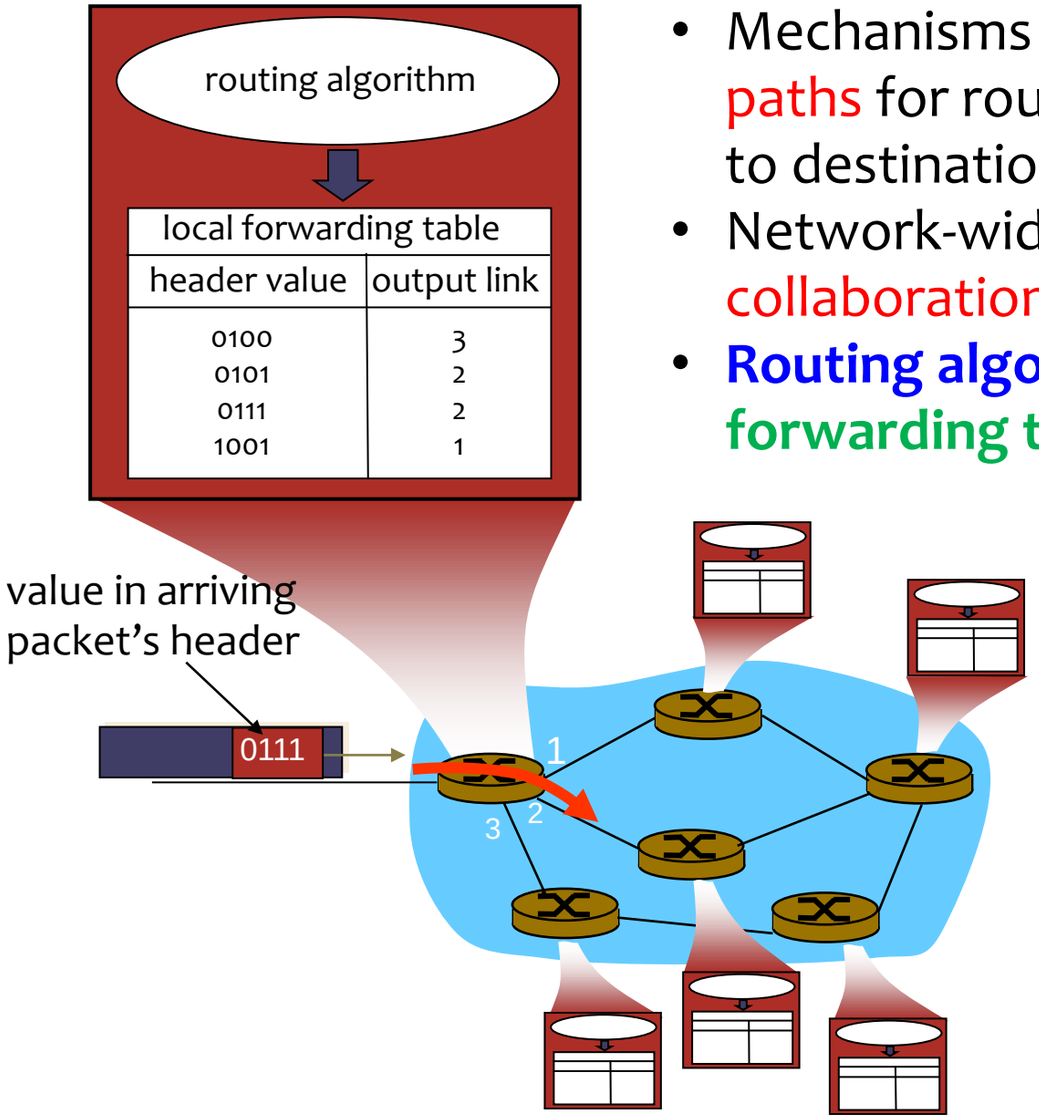
Network Layer

- Network layer functions/services
  - Forwarding & Routing – Challenge 1 & 2
  - Fragmentation & reassembly – Challenge 2
    - Handling of the different packet sizes used by different underlying networks
  - Other features (not going to cover in this course)
    - Priority & Scheduling – QoS (Quality Of Service), i.e., service guarantee
    - Security – e.g., IPSec
    - Multicast support

Routing and Forwarding

Forwarding or switching function

- forwards packets from input port to appropriate output port by dynamically connecting the input to output
- **port selected based on "addressing" information in packet header and lookup the forwarding table**
- Takes place at very short timescales



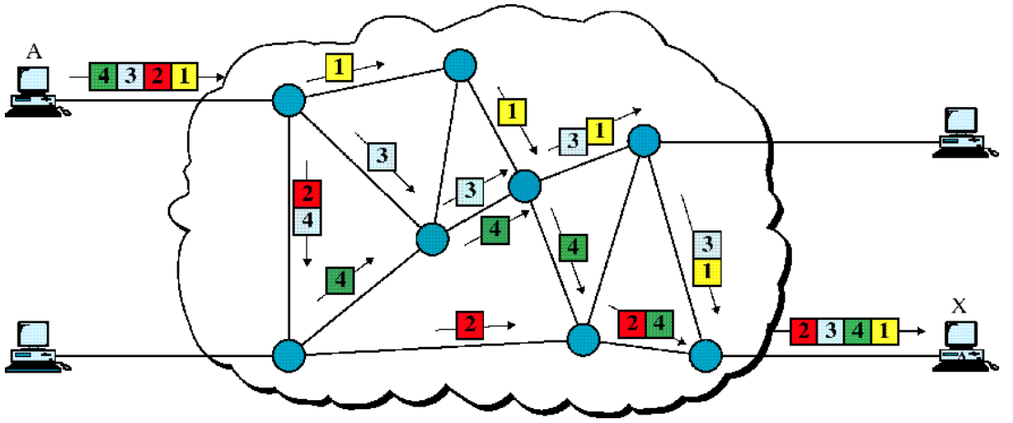
Routing function

- Mechanisms for determining the best paths for routing packets from source to destination
- Network-wide: require the collaboration of routers
- **Routing algorithm determines the forwarding tables of each routers**

Example: Datagram Packet Networks

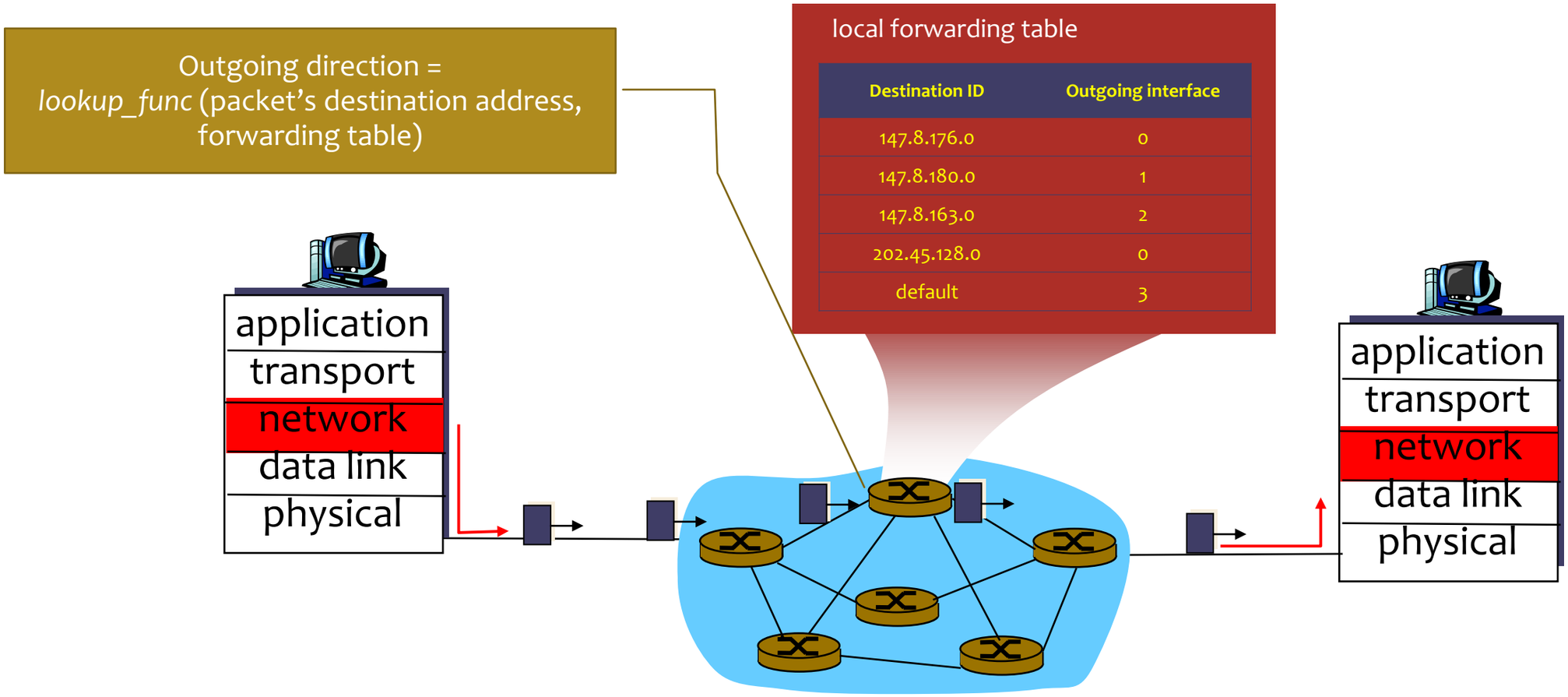
- Each time a host wants to send a packet, it places the destination host's address in the packet's header
- Each packet is being routed independently
- **Using destination address information** for the forwarding decision
  - Each router has a forwarding table that **maps destination address to outgoing interface**
- Packets between same source-destination pair (**in principle**) may take different routes
  - may result in out-of-order arrival
- Individual packet may be missing

- Responsibility of the receiver
  - re-order packets
  - handle packet losses



Forwarding Table

- Each router maintains a forwarding table

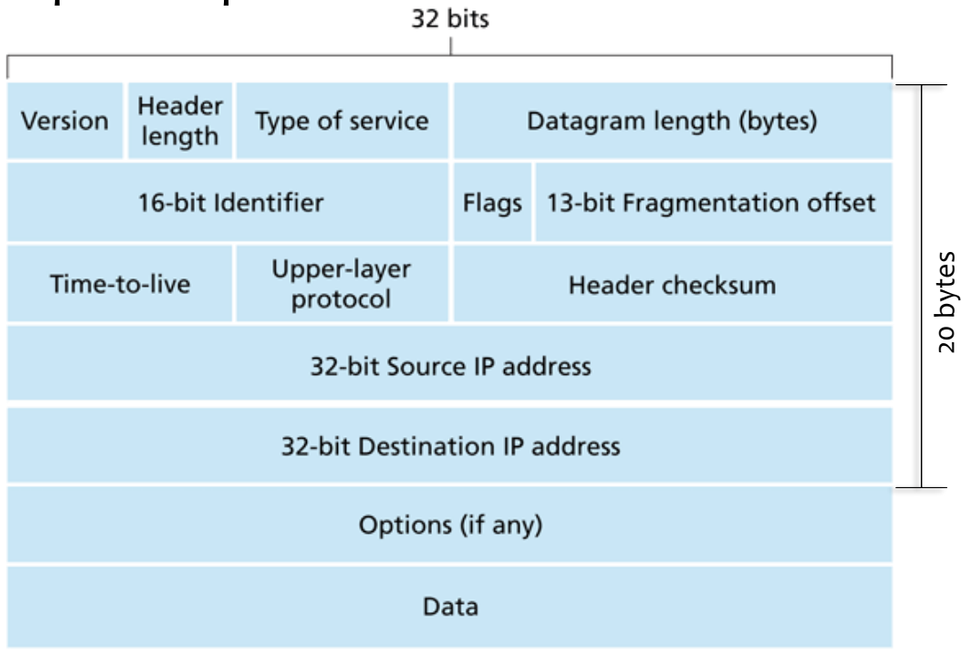


Internet Protocol

Internet Protocol

- Network layer protocol
  - Provides best effort, connectionless packet delivery
  - Responsibility
    - Forwarding
      - Addressing schemes
      - Path selections
    - Datagram management
      - Datagram format
      - IP Fragmentation & Re-assembly
    - Error control
      - Error reporting -- ICMP

- IP datagram format
  - Header part + Data part
  - Header format
    - a **20 byte-fixed part** and a variable length optional part



Packet Header

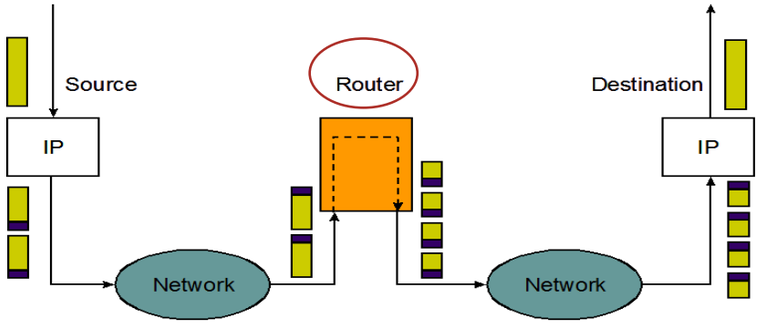
- Version (4-bit)
  - IP protocol version number,
  - e.g. IPv4, IPv6
- Header length (4-bit)
  - indicates the length of header in units of 4 bytes
  - 4-bit field ⇒ max 60 bytes
- Type of service (8-bit)
  - Distinguish between different classes of service, e.g. real-time data flow
  - To specify level of service
- Datagram length (16-bit)
  - The **whole datagram** length in bytes (header plus data)
  - 16 bits ⇒ max 65535 bytes
- Identifier, Flags, Offset
  - Used for fragmentation and reassembly

Packet Header (2)

- Time To Live (8-bit)
  - Specifies the max. no. of routers (hops) the packet is allowed to traverse through
  - Reduces one unit at each router
  - Router discards datagram when TTL reaches zero
    - Sends back an ICMP packet to source
  - This ensures packets will not be in the network forever
- Protocol (8-bit)
  - Indicates the specific transport layer or other protocol, e.g. TCP or UDP, which is carried in the datagram
- Header Checksum (16-bit)
  - Internet Checksum
  - Must be recomputed per hop. Why?
- Source & destination IP address (32-bit)
- Options
  - at most 40 bytes
  - Use for extending the IP header for options that are rarely used and experimental purpose, e.g. source routing

IP Fragmentation

- Physical links have MTU (Maximum Transfer Unit)
  - A limitation on how many bytes of data in the payload field of a Link-layer frame can carry
  - Different link-layer technologies have different MTUs
- Packets may pass through different link-layer protocols along the path on Internet from the source to destination



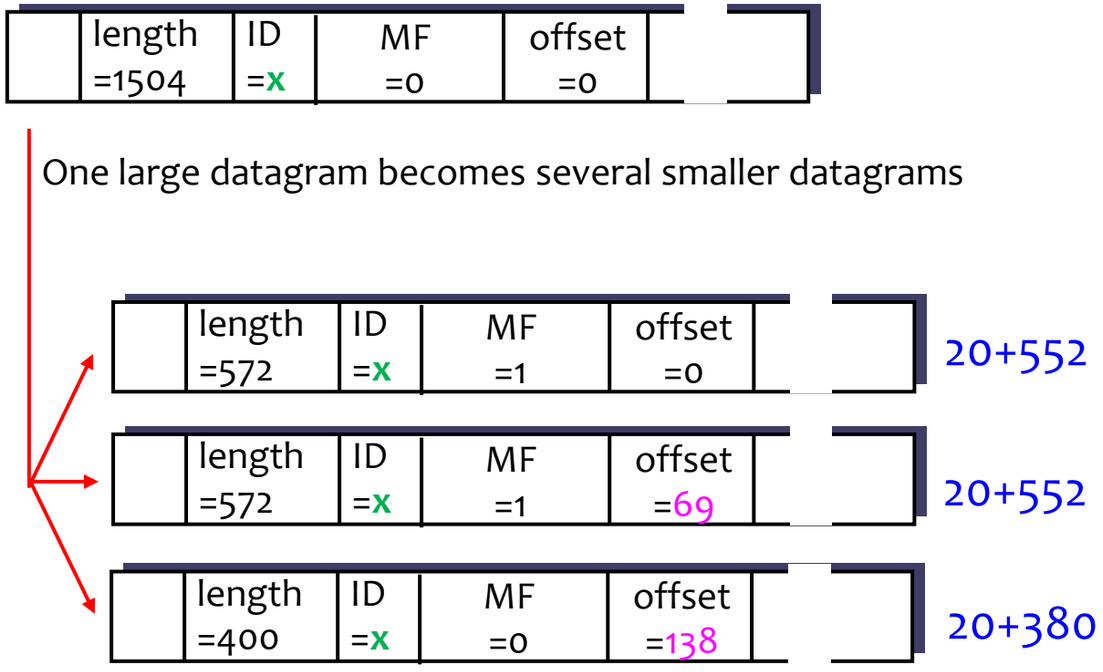
- Large IP datagram
  - one IP datagram (e.g., 10000 bytes) could be **fragmented** into several smaller IP datagrams by the router
  - **Reassembled** at the final destination
  - We need to identify individual pieces of the original datagram, order them, and reassemble them to get back the original datagram before passing it to upper layer
    - IP header fields: identifier, flag, offset are used for identifying and reassembling related fragments

Example

- A packet is to be forwarded to a network with **MTU of 576 bytes**. The packet has an IP header of **20 bytes** and a data (payload) part of **1484 bytes**.
- Available data length per fragment = 576 - 20 = 556 bytes

- We set maximum data length to 552 bytes so as to make it in **multiple of 8**.

ID - all fragments of the same IP datagram have same identifier
MF – more fragment bit, one of the bits in the flags field
1 = more to come
0 = last fragment
offset - where this fragment starts; in units of 8-byte.



Avoid Fragmentation

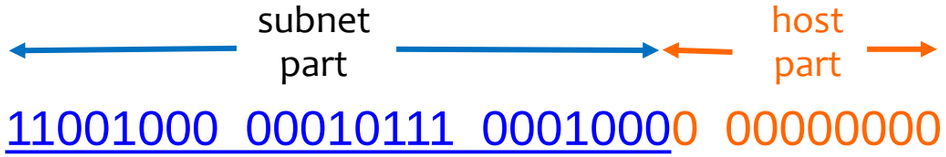
- Path MTU discovery
  - Defined in RFC 1191
  - A mechanism which allows a host to **detect a path MTU** smaller than its interface MTU
  - This **avoids** the needs to have fragmentations at intermediate routers
  - Two components are the keys to this mechanism
    - Don't Fragment (DF) bit of the IP header Flags field, which prevents a router from performing fragmentation
      - A host sets the DF bit of the outgoing packets when sending to the destination; any device along the path whose MTU is smaller than the packet will drop the packet
    - A specific ICMP (Internet Control Message Protocol) message is generated to report the error to source which includes the MTU of the link necessitating fragmentation

IPv4 Addressing

- 32-bit identifier
  - To **uniquely identify an interface** in the global network
  - Not refer to a host, but to indicate a network interface of a router/host
- Interfaces
  - routers have multiple ports
    - that means multiple interfaces
    - Not all ports are assigned with IP addresses; only active port that attaches to a network is assigned with an IP address **associated with the attached network**
  - hosts can have multiple interfaces too - "multi-homed"
    - IP address is assigned to each active interface
- Dotted-Decimal Notation:
  - int_1. int_2. int_3. int_4 where int_j = integer value of j^{th} octet
  - IP address of `11011111 00000001 00000001 00000001`
    - is `233.1.1.1` in dotted-decimal notation

IPv4 Addressing

- An IP address has two parts
  - **subnet part** (high order bits)
    - What's a "subnet"? From IP address perspective
      - A subnet is also called an IP network
      - **all interfaces** under the same IP network have the **same subnet part**
        - All interfaces in same subnet can physically reach each other without intervening router
          - means they are not using network layer info (IP address) to locate the destination interface; instead, they use link layer info (MAC address) to locate the destination interface



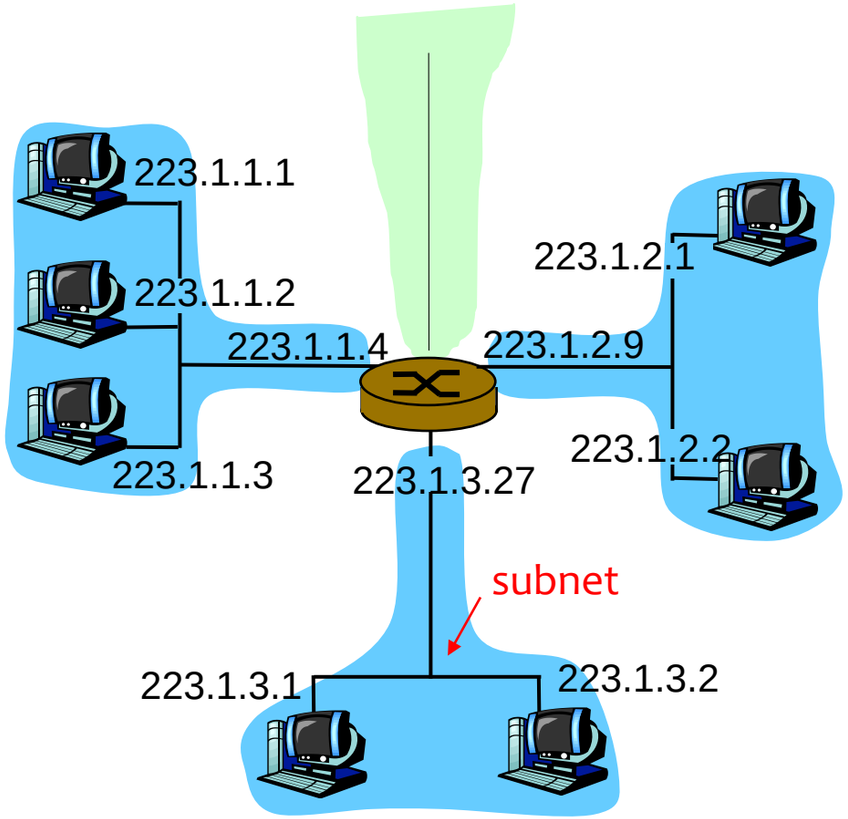
  - **host part** (low order bits)
    - identify an individual "interface" in that subnet
    - can be assigned locally by network admin without global coordination

  - The combination is **globally unique** (except those private IP addresses)

Subnet Address [Network Address]

- The example network consists of 3 subnets
  - 223.1.1.0/24
    - 223.1.1.1, 223.1.1.2, 223.1.1.3, 223.1.1.4
  - 223.1.2.0/24
    - 223.1.2.1, 223.1.2.2, 223.1.2.9
  - 223.1.3.0/24
    - 223.1.3.1, 223.1.3.2, 223.1.3.27

223.1.1.0 – this IP address represents an IP Subnet



- /24 notation is known as subnet mask
  - also refer to as the network prefix of the address
  - we can express it as 255.255.255.0
  - indicates the **left most** 24 bits of an IP address is used to represent the subnet (network) address of that IP address

Exercise

- Given the IP address 201.14.78.65 and the subnet mask 255.255.255.224, what is the address of this subnet? (sometimes we call it subnet address or network address) Give the range of IP address in this subnet?

No. of bits in the mask = 8+8+8+3 = 27
Extract the high-order 27 bits from the address:

	201. 14. 78. 65	11001001.00001110.01001110.01000001
AND	255.255.255.224	11111111.11111111.11111111.11100000
	201. 14. 78. 64	11001001.00001110.01001110.01000000

So the address of this subnet is 201.14.78.64/27.

The no. of IP addresses in this subnet is indicated by the host part, which has 5 bits in this case; that means we can have 2^5 addresses.
The range of IP address in this subnet is 201.14.78.**64**-201.14.78.**95**.
Please note that 201.14.78.64 and 201.14.78.95 are reserved IP addresses in this subnet, which represent the subnet address and broadcast address of this subnet respectively; the two addresses cannot be assigned to any hosts.

Current Internet's Address Assignment Strategy - CIDR

- CIDR – Classless InterDomain Routing
  - subnet portion of IP address is of arbitrary length
    - As compared to the old strategy – Classful scheme which has the subnet portion is of fixed length for each class
  - An organization is typically assigned a block of contiguous addresses
    - the no. of IP addresses in this range **must be** in power of two
    - All addresses within this range can be represented by the same network prefix

11001000 00010111 00010000 00000000

200.23.16.0/23 range: 200.23.16.0 – 200.23.17.255

Subnetting

- After getting the range of (consecutive) IP address, **within** the organization
  - further **subnetting** can be applied to divide this range to a few subranges to ease the administration
    - Introduces another **hierarchical level**
    - Splits a network into several smaller subnets for internal use
      - take away **some high-order bits from the host part** to identify the subnets **within** organization
      - the no. of IP addresses in each smaller subnet must be in power of two
  - The created subnets can be **visible only** within site (the organization)

- Example
  - Organization has network address 150.100.0.0/16
  - Create subnets with up to 100 hosts each
    - 7 bits sufficient to represent hosts in each subnet
  - Subnet addresses are in the form of 150.100.x.y/25
  - How to find the subnet address for this ip addr: 150.100.12.176

    - IP addr = 10010110 01100100 00001100 10110000
    - netmask = 11111111 11111111 11111111 10000000
    - AND = 10010110 01100100 00001100 10000000
    - Subnet = 150.100.12.128 /25 [150.100.12.128 - 150.100.12.255]

  - Subnet addresses used by routers **within** organization

Subnetting

- An important consequence of subnetting is that different parts of the internet see the world differently

- View from the backbone router (**outside** the organization)
  - external view the organization network as a single network; those internal subnets are not visible
  - thus, **a single routing entry** of the form a.b.c.d/x will be sufficient to **summarize all destination hosts within** the organization
    - an entry with 150.100.0.0./16 in the forwarding table

- View from the routers within the organization
  - Need to be able to route packets to the right subnet within the organization
    - Each subnet has an entry in the forwarding tables of the internal routers

- This is one of the advantages of the CIDR scheme

CIDR Routing

- Forwarding
  - Forwarding table holds routing entries of the form (Subnet address, Subnet Mask, NextHop).
  - To perform forwarding, the router scans the table entry by entry
    - masking the destination IP address field of the incoming packet with the subnet mask of the entry
      - Destination network = bitwise AND (packet's destination address, subnet mask)
    - compare to the Subnet address of current entry for match

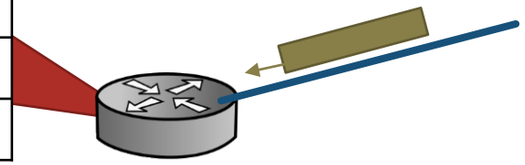
- Longest Prefix Match
  - Multiple entries may match a given destination IP address
    - Packet must be routed **using the more specific route**, that is, the longest prefix match
      - In the coming example, the packet will forward to interface 2

  - Several fast longest-prefix matching algorithms are available

Forwarding Table (A simplified version)

- Each router has forwarding table that maps destination addresses to (outgoing) link interfaces
- Example:

Destination network	Link interface
200.23.16.0/21	0
200.23.24.0/21	1
200.23.24.0/24	2
default	3



- Consider an incoming packet with this destination address **200.23.20.114**, which outgoing interface will be selected?

Prefix length - 21

```
11001000.00010111.00010100.01110010
11111111.11111111.11111000.00000000
11001000.00010111.00010000.00000000
     200.      23.      16.       0
```

```
11001000.00010111.00010100.01110010
11111111.11111111.11111111.00000000
11001000.00010111.00010100.00000000
     200.      23.      20.       0
```

Prefix length - 24

- Consider another incoming packet with destination address **200.23.24.123**, which outgoing interface will be selected?

Prefix length - 21

```
11001000.00010111.00011000.01111011
11111111.11111111.11111000.00000000
11001000.00010111.00011000.00000000
     200.      23.      24.       0
```

```
11001000.00010111.00011000.01111011
11111111.11111111.11111111.00000000
11001000.00010111.00010100.00000000
     200.      23.      24.       0
```

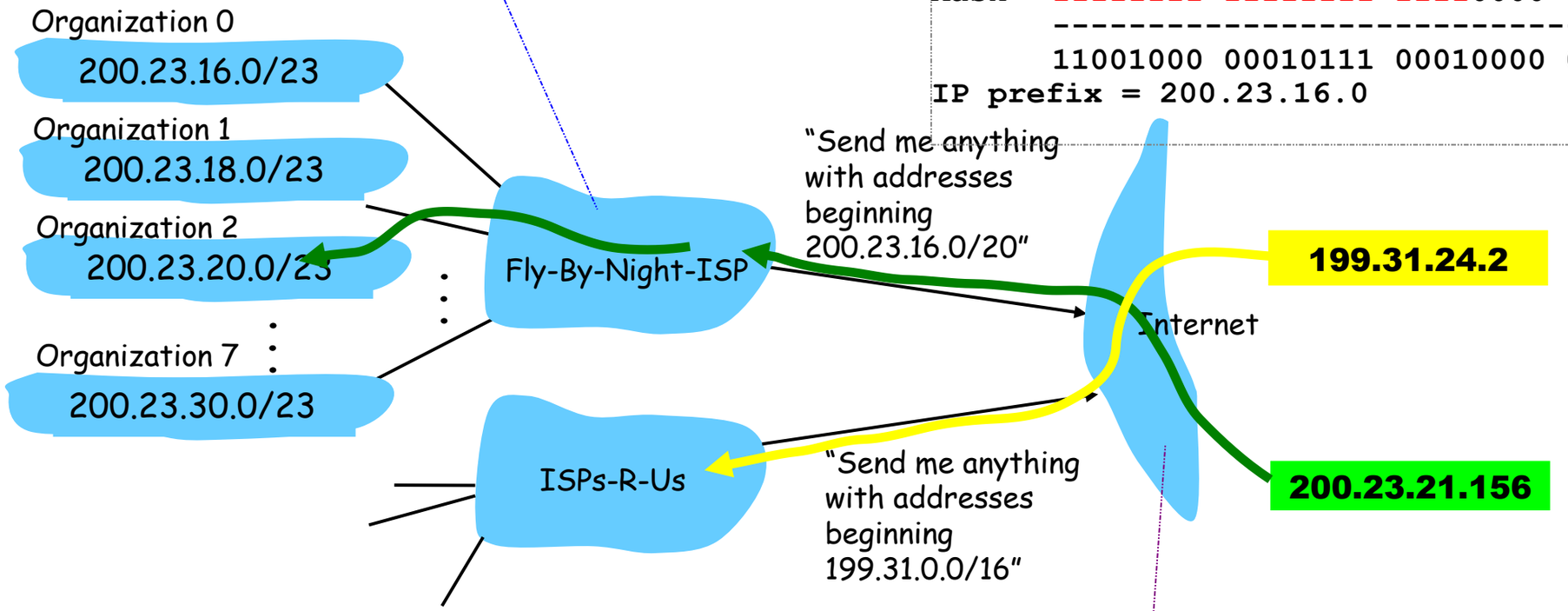
Prefix length - 24

Supernetting

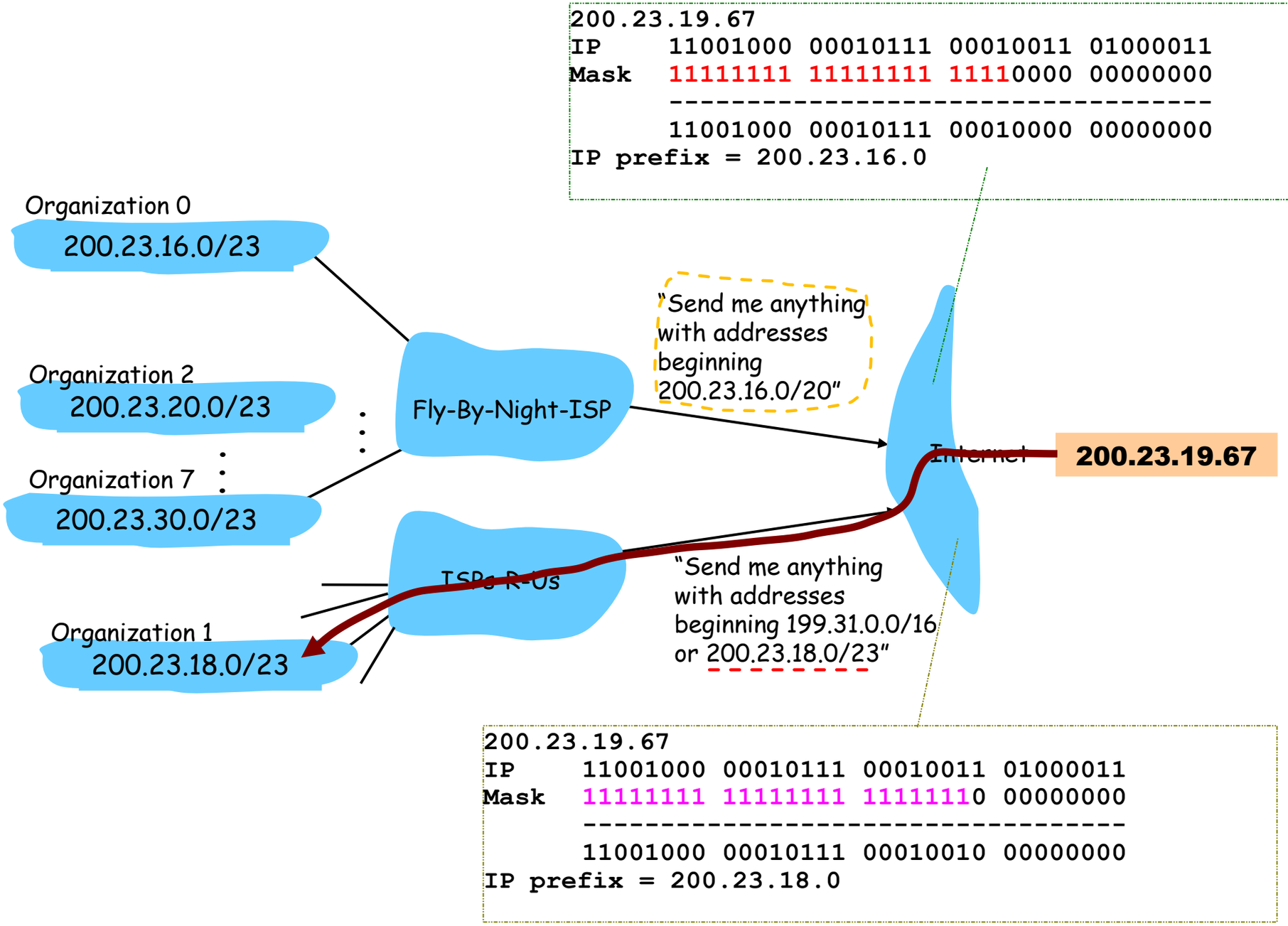
- Subnetting has a counterpart, sometimes called Supernetting
  - Allows us to coalesce several subnet addresses into a single/bigger "supernet"
  - Instead of having a forwarding entry for each subnet, subnetting aggregate them as a single entry in the forwarding table for reaching to that group of subnets
  - Those smaller subnets must share a common network prefix
- Example
  - An Internet service provider network assigns IP addresses to the customers in a way that many different customer networks share a common, shorter network prefix.
  - Then the provider can aggregate those customer networks as a single route to the external world

```
200.23.21.156
IP     11001000 00010111 00010101 10011100
Mask   11111111 11111111 11111110 00000000
       ----------------------------------
       11001000 00010111 00010100 00000000
IP prefix = 200.23.20.0
```

```
200.23.21.156
IP     11001000 00010111 00010101 10011100
Mask   11111111 11111111 11110000 00000000
       ----------------------------------
       11001000 00010111 00010000 00000000
IP prefix = 200.23.16.0
```



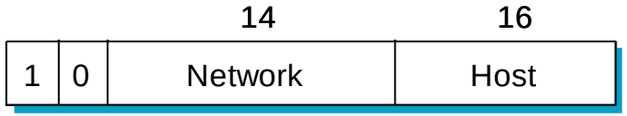

```
199.31.24.2
IP     11000111 00011111 00011000 00000010
Mask   11111111 11111111 00000000 00000000
       ----------------------------------
       11000111 00011111 00000000 00000000
IP prefix = 199.31.0.0
```


200.23.19.67
IP 11001000 00010111 00010011 01000011
Mask 11111111 11111111 11110000 00000000
11001000 00010111 00010000 00000000
IP prefix = 200.23.16.0
Organization 0
200.23.16.0/23
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"
Internet
200.23.19.67
ISPs-R-Us
"Send me anything with addresses beginning 199.31.0.0/16 or 200.23.18.0/23"
Organization 1
200.23.18.0/23
200.23.19.67
IP 11001000 00010111 00010011 01000011
Mask 11111111 11111111 11111110 00000000
11001000 00010111 00010010 00000000
IP prefix = 200.23.18.0

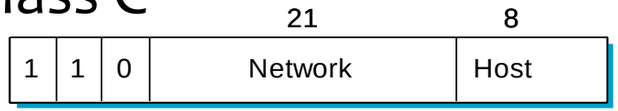
Classful Addressing - Historical

- Divide IP addresses to 5 classes
  - Class A

    - 126 networks with up to 16 million hosts
    - 0.0.0.0 to 127.255.255.255
  - Class B



    - 16384 networks with up to 65 thousand hosts
    - 128.0.0.0 to 191.255.255.255
  - Class C



    - 2 million networks with up to 254 hosts
    - 192.0.0.0 to 223.255.255.255
  - Class D
    - start 1110
    - 28 bits for the multicast addresses
    - 224.0.0.0 to 239.255.255.255
  - Class E
    - Start 1111

Special addresses

- Special addresses
  - 0.0.0.0 – represents this host (used when booting up) on this network
  - subnet part is all zeros with a non-zero host part – a host in this network
  - 255.255.255.255 – broadcast on the local network (subnet)
  - host part is all ones with a non-zero subnet part – broadcast on the targeted remote network
  - 01111111 (127) – loopback
- Private IP Addresses
  - Are restricted to private networks
  - routers in public Internet **discard packets** with these addresses
    - Range 1: 10.0.0.0 to 10.255.255.255
    - Range 2: 172.16.0.0 to 172.31.255.255
    - Range 3: 192.168.0.0 to 192.168.255.255
  - Network Address Translation (NAT) used to convert between private & global IP addresses

ConcepTest 5

Summary

- In datagram packet networks, routers use the packet's destination address to forward the packet to the next hop.
- Fragmentation is a mechanism that allows the IP network to interconnect between different physical networks, which have different frame payload sizes for carrying the IP packet.
- IP addresses are designed with a hierarchical structure. A portion of the address indicates the network (netid), a portion indicates the subnet, and a portion indicates the host on the network.

Summary (2)

- CIDR has two main features: (1) the allocation of the IP addresses is in variable-sized block and there is no concept of classes; (2) the change to classless addressing scheme makes the need to change in routing table organization and searching.
  - Because of variable-sized subnetting, ISP can aggregate a few small contiguous subnets into one single prefix, this significantly reduces the size of the routing table.
  - The use of variable-length prefixes requires that the routing tables be searched to find the longest prefix match.