

Network/Internet Layer Protocols

Peerapon S.

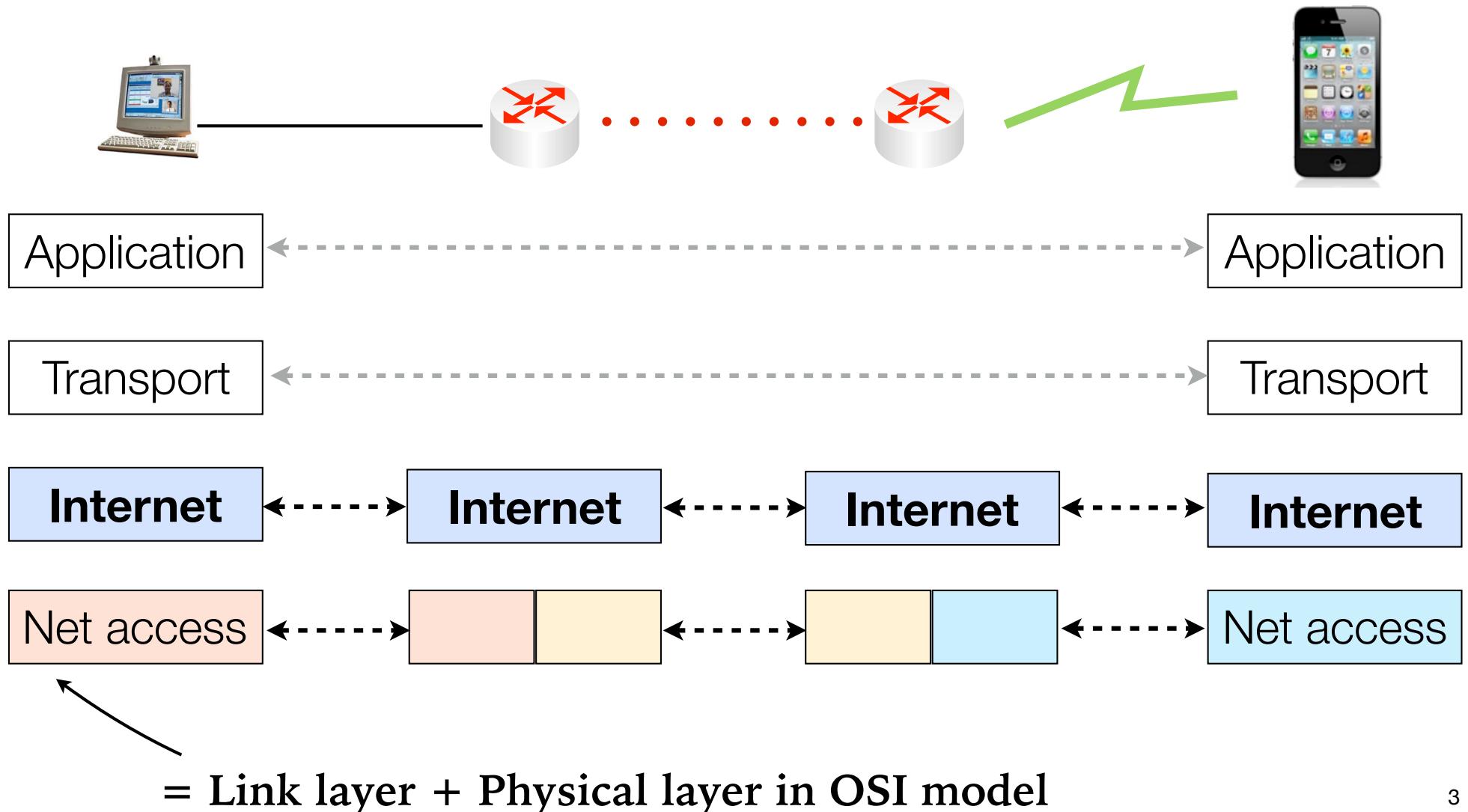
CPE 314: Computer Networks (2/61)

Topics

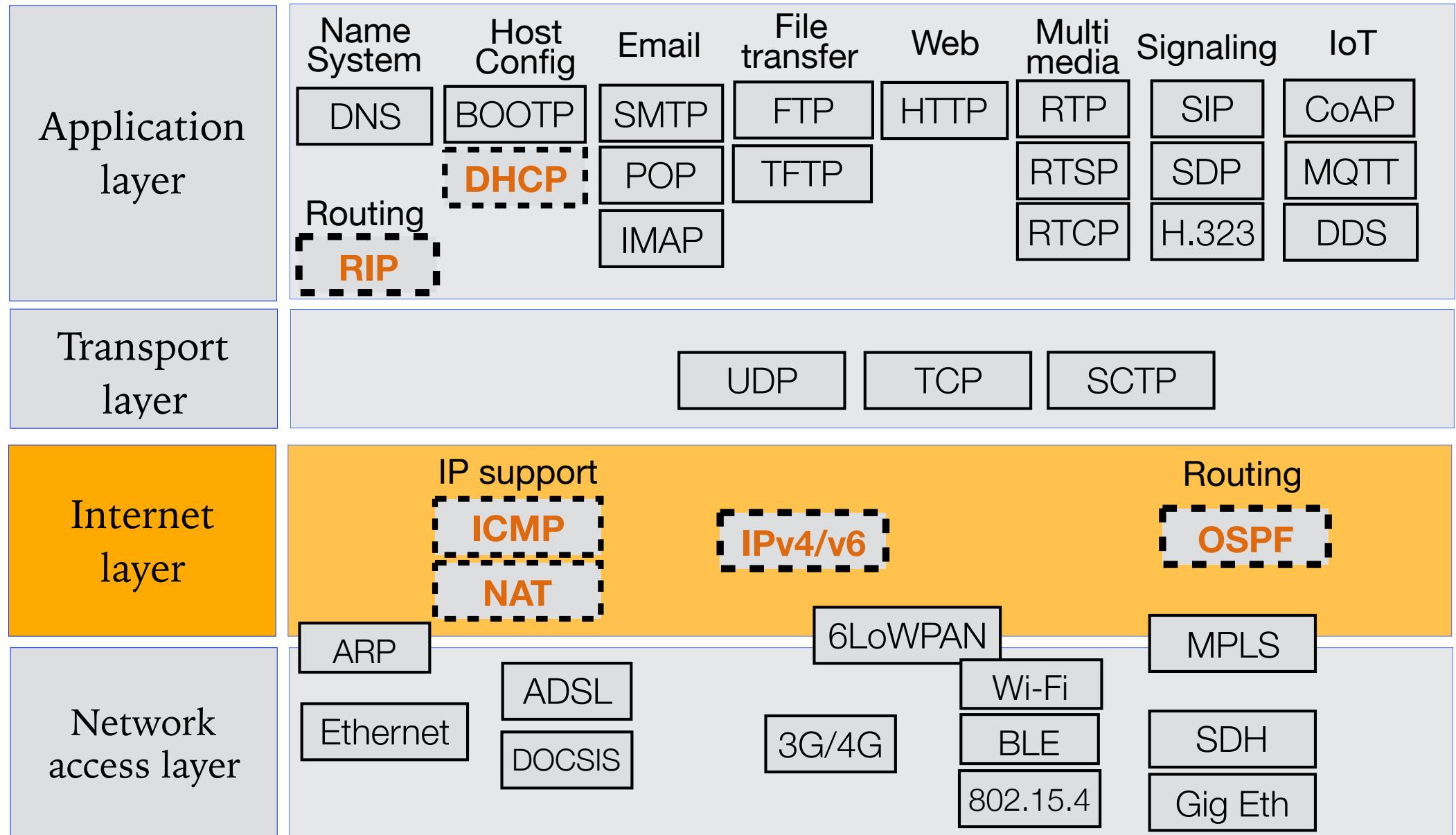
- What does the Internet/Network layer do ?
- Internet Protocol (IPv4)
 - Logical addressing
 - Packet forwarding
- IP address assignment with DHCP
- Network troubleshooting with ICMP
- Network address translation mechanism
- Network congestion and TCP congestion control

Internet/Network Layer Service

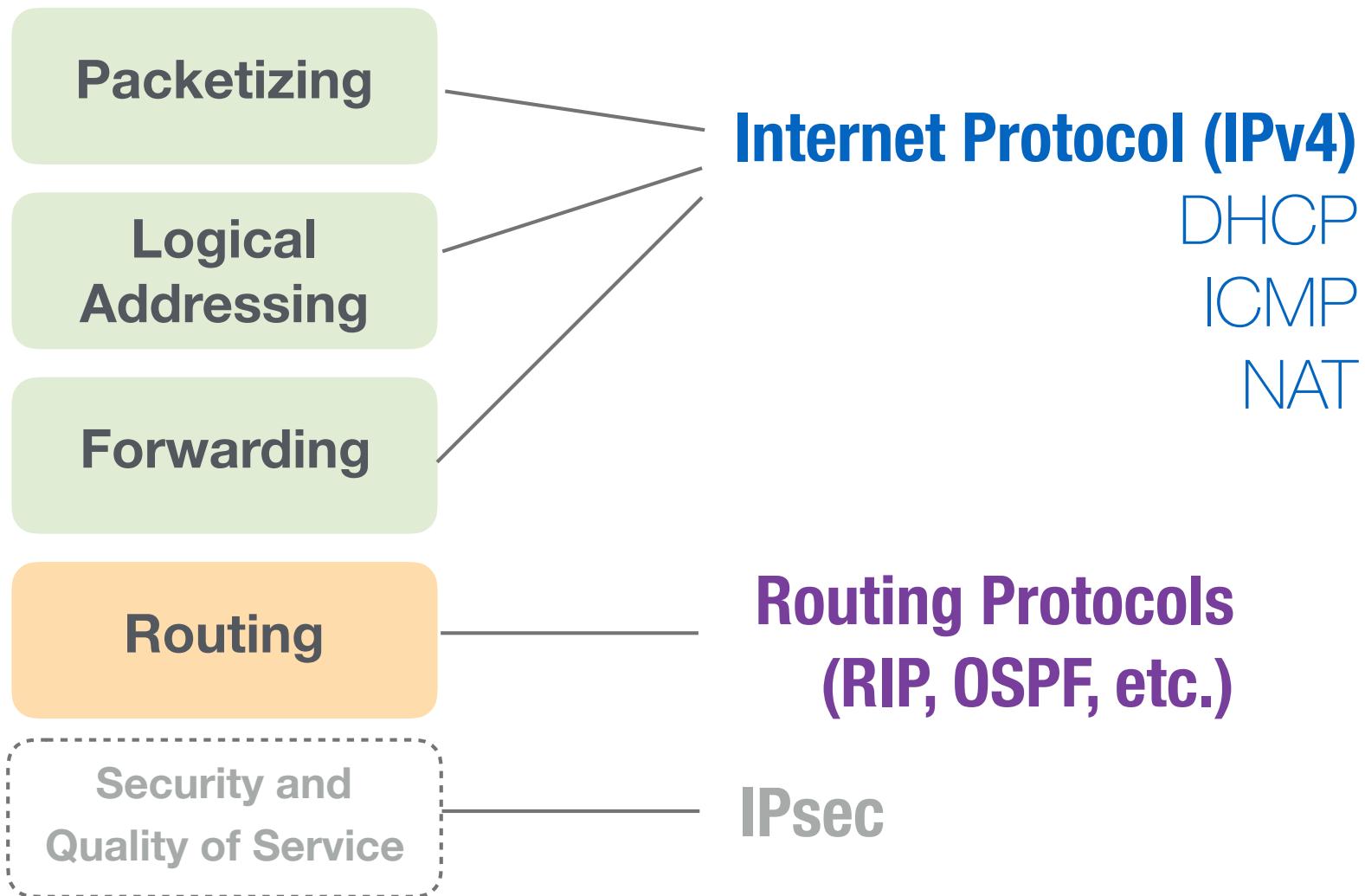
- *Host-to-host packet delivery service (over an internetwork)*

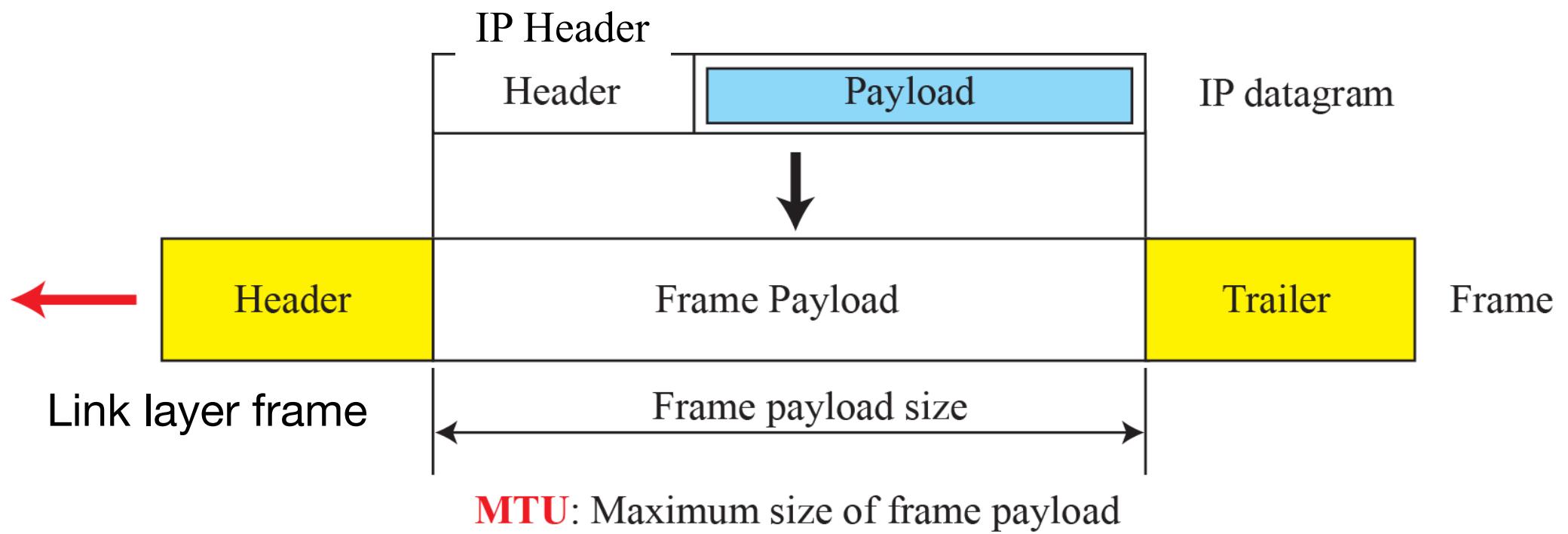
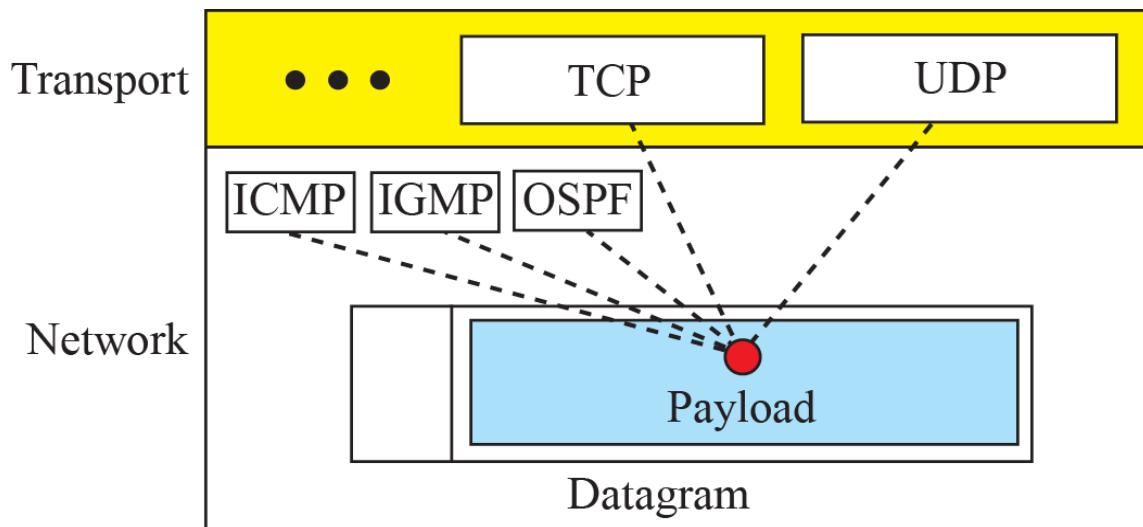


Protocols in TCP/IP Suite

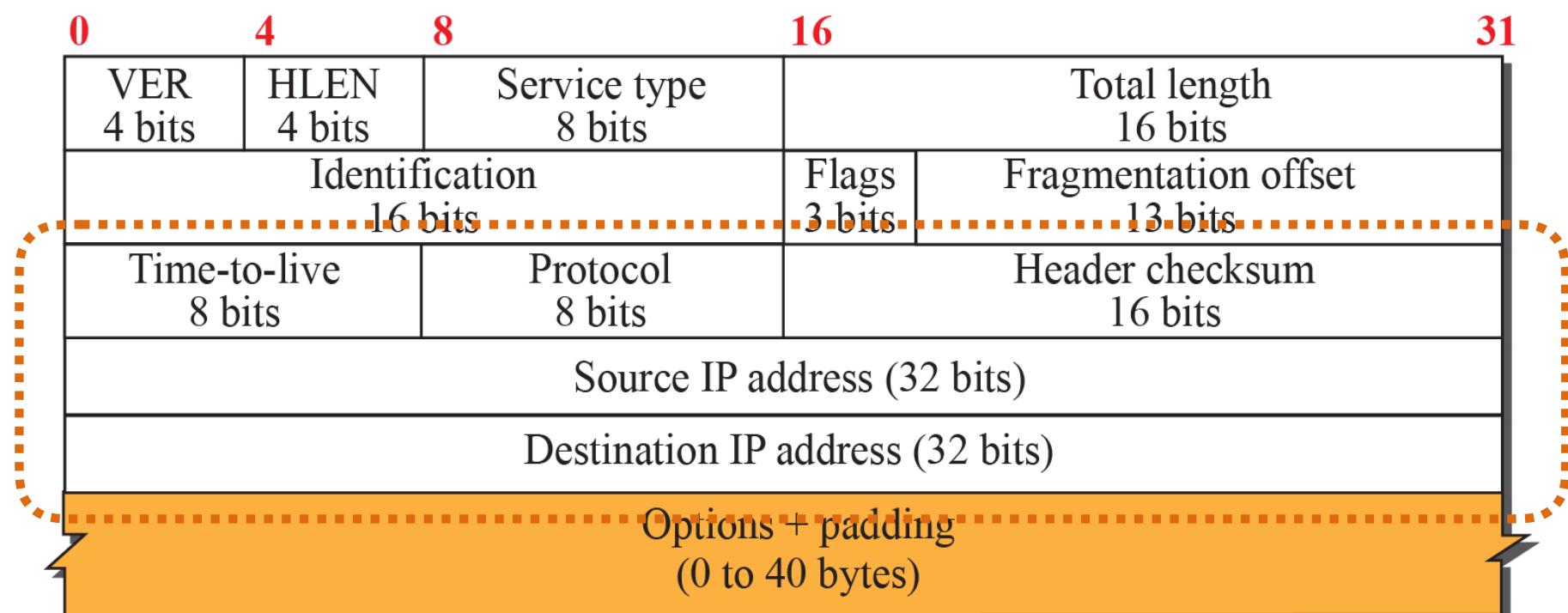
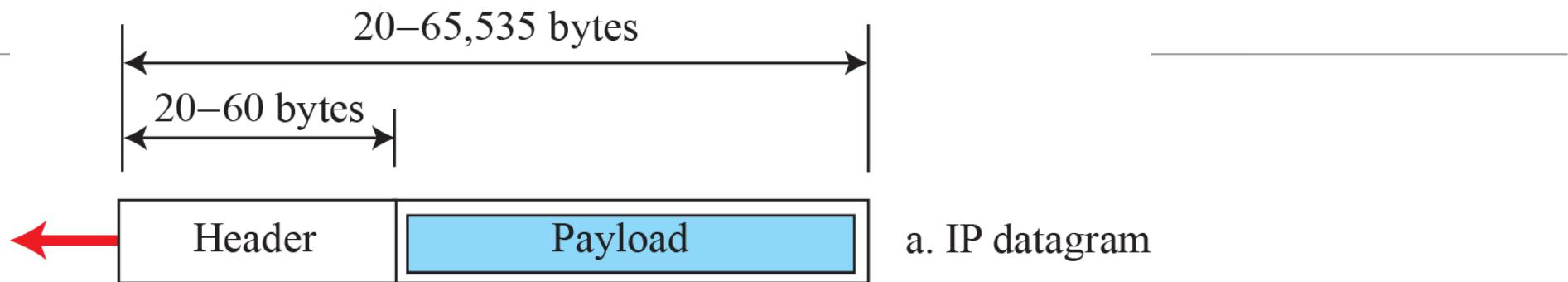


Network Layer Functions





IP Datagram



b. Header format

Capturing from Wi-Fi: en0

Apply a display filter ... <⌘/>

No. Time Source Destination Protocol Length Info

47	1.110498	17.248.154.209	192.168.1.54	TCP	1514	[TCP Out-Of-Order] 443 → 58019 [ACK] Seq=5667 Ack=13162 Win=
48	1.110561	192.168.1.54	17.248.154.209	TCP	66	58019 → 443 [ACK] Seq=5667 Ack=13162 Win=
49	1.127918	192.168.1.54	17.248.154.209	TLSv1...	1484	Application Data
50	1.127918	192.168.1.54	17.248.154.209	TLSv1...	276	Application Data
51	1.272802	17.248.154.209	192.168.1.54	TCP	66	443 → 58019 [ACK] Seq=13162 Ack=7085 Win=
52	1.272806	17.248.154.209	192.168.1.54	TCP	66	443 → 58019 [ACK] Seq=13162 Ack=7295 Win=

► Frame 1: 66 bytes on wire (528 bits), 66 bytes captured (528 bits) on interface 0

► Ethernet II, Src: Humax_5f:52:86 (2c:08:8c:5f:52:86), Dst: Apple_5b:89:c0 (8c:85:90:5b:89:c0)

▼ Internet Protocol Version 4, Src: 17.248.154.209, Dst: 192.168.1.54

0100 = Version: 4
.... 0101 = Header Length: 20 bytes (5)
► Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
Total Length: 52
Identification: 0x57f5 (22517)
► Flags: 0x02 (Don't Fragment)
Fragment offset: 0
Time to live: 48
Protocol: TCP (6)
Header checksum: 0x8427 [validation disabled]
[Header checksum status: Unverified]
Source: 17.248.154.209
Destination: 192.168.1.54
[Source GeoIP: Unknown]
[Destination GeoIP: Unknown]

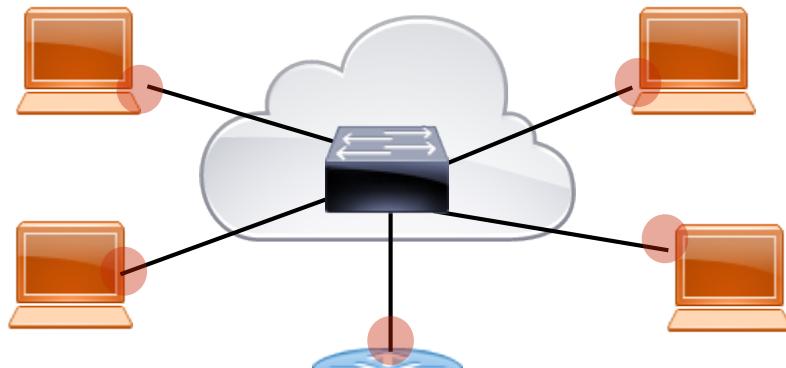
► Transmission Control Protocol, Src Port: 443, Dst Port: 58019, Seq: 1, Ack: 1, Len: 0

0000	8c 85 90 5b 89 c0 2c 08	8c 5f 52 86 08 00 45 00	...[...,_R..E.
0010	00 34 57 f5 40 00 30 06	84 27 11 f8 9a d1 c0 a8	.4W.@.0. .'.....
0020	01 36 01 bb e2 a3 5f 78	ef cb 0c dd a7 06 80 10	.6....._x
0030	03 ab 02 23 00 00 01 01	08 0a b8 9b 7d 1b 4f ed	...#....}..0.
0040	95 1c		..

Internet Protocol Version 4 (ip), 20 bytes

Packets: 1696 · Displayed: 1696 (100.0%) · Profile: Default

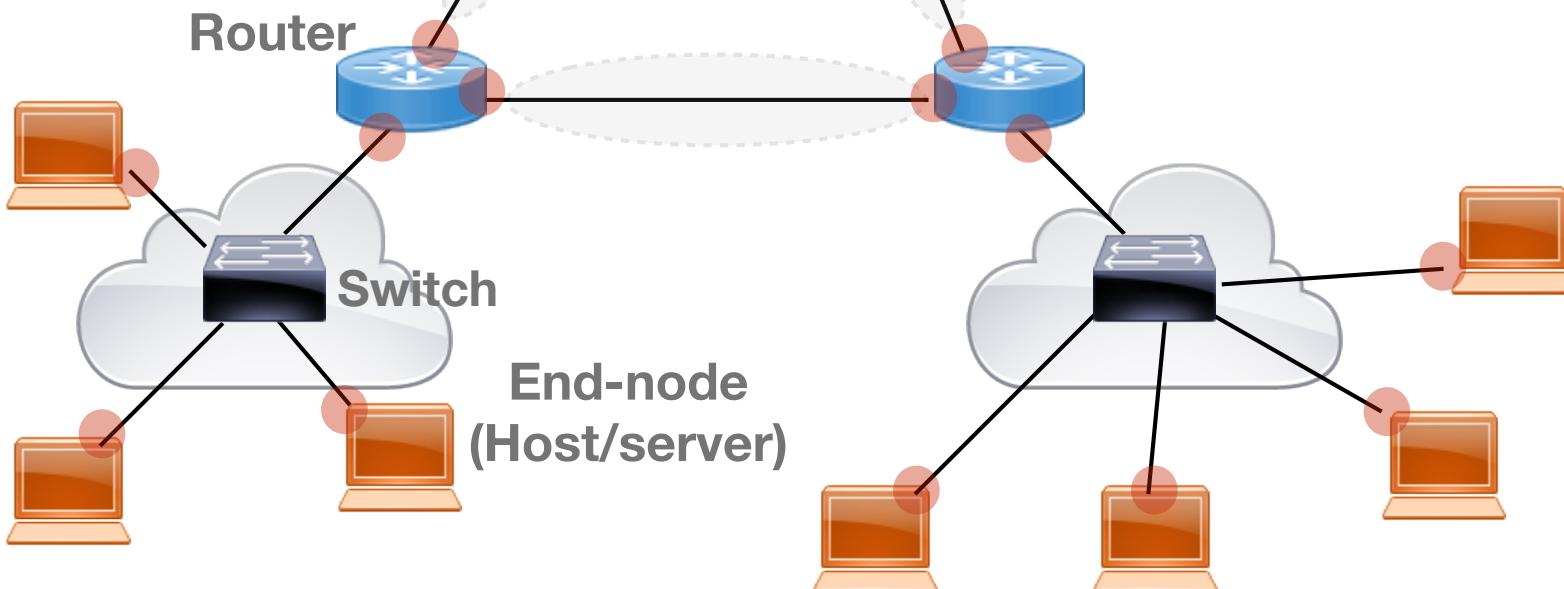
Local Network or Subnet(work)



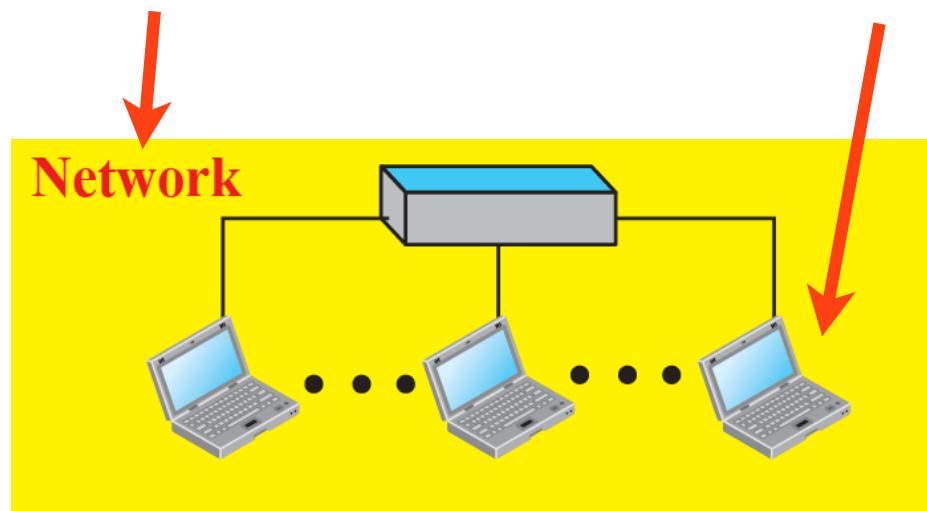
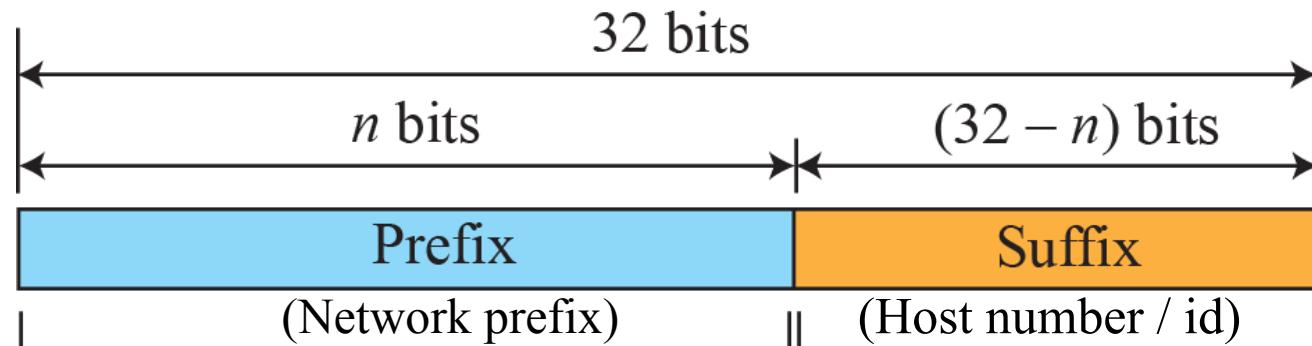
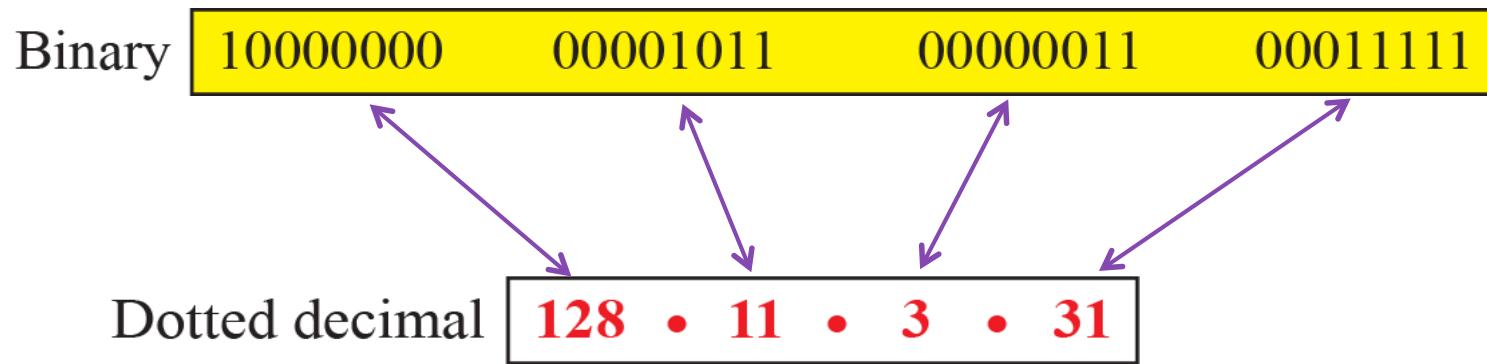
20 network interfaces(Routers + End-nodes)
One unique id per interface

6 local networks or subnets

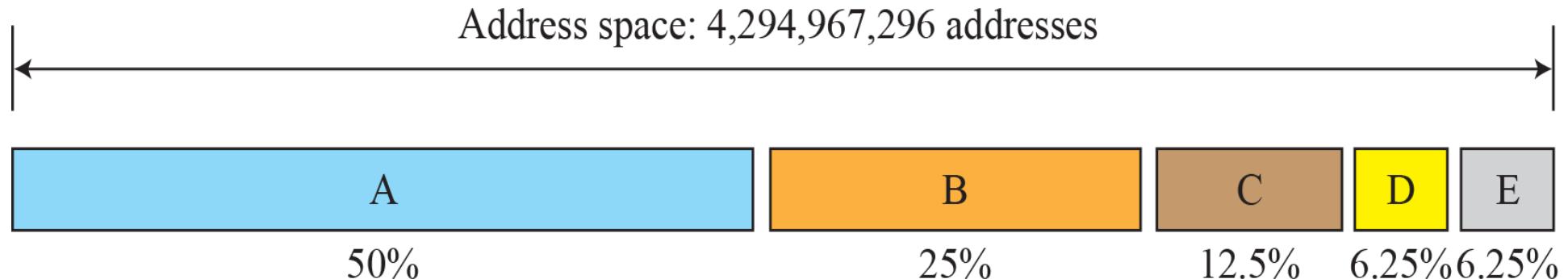
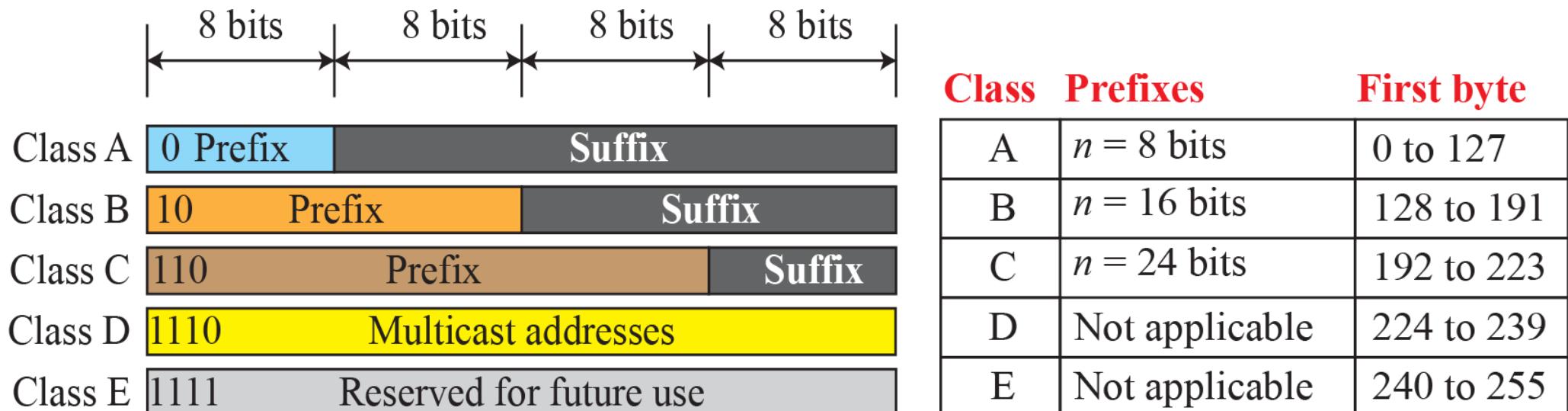
Two interfaces are in different subnetworks
if one or more routers are in between.



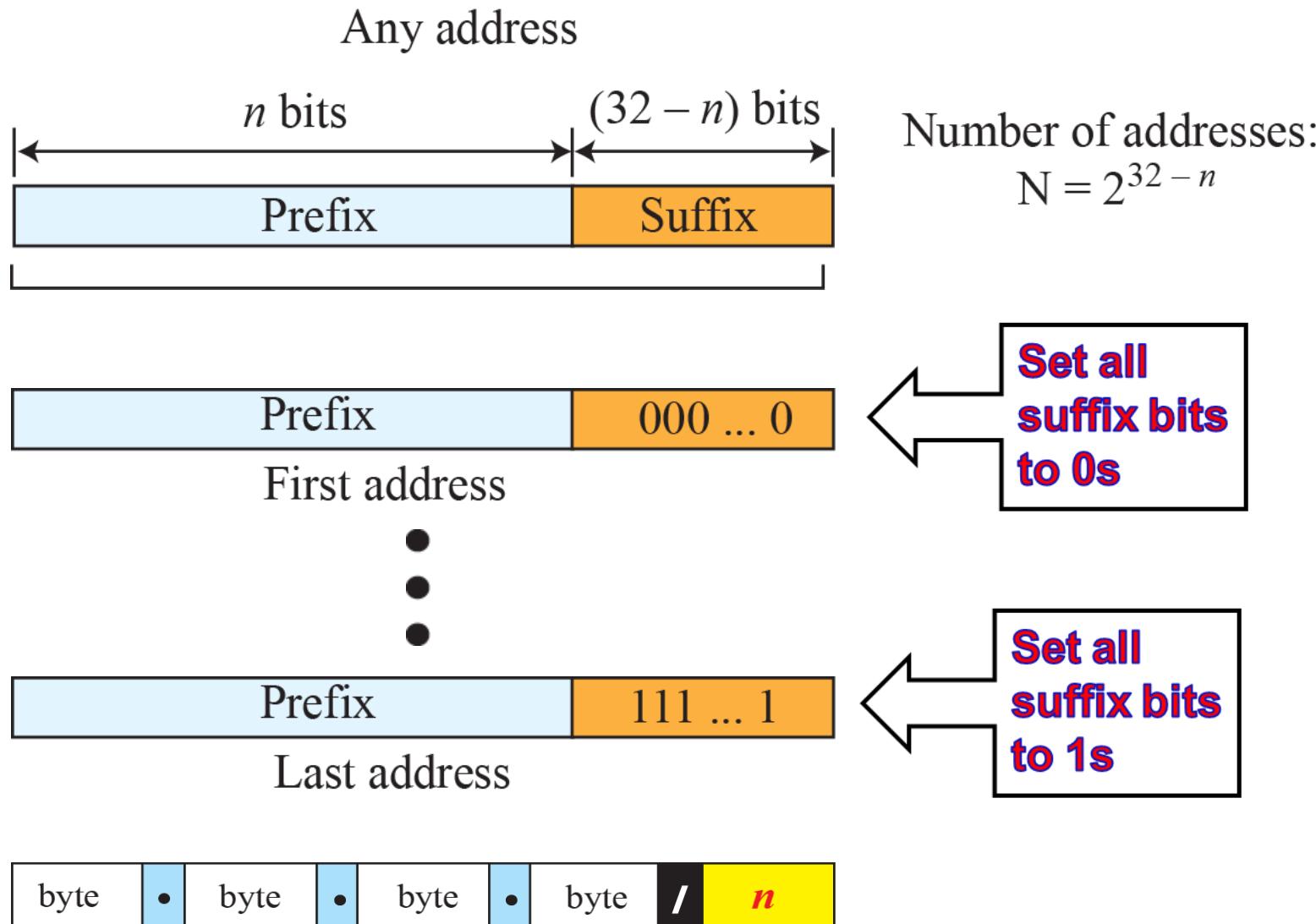
IPv4 Addressing



Classful Addressing (Obsolete)



Classless Addressing

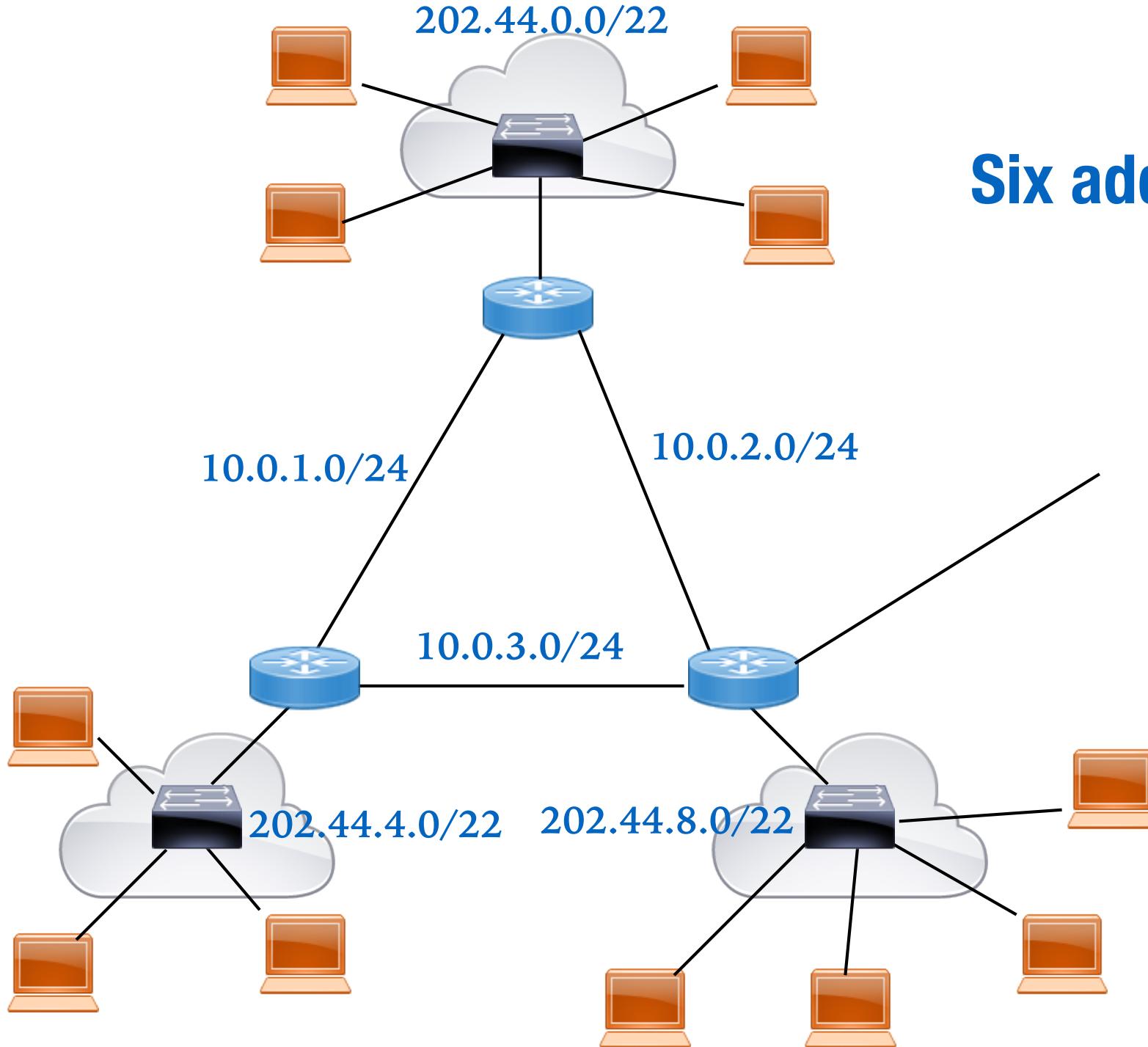


Classless InterDomain Routing (CIDR) or Slash notation

Dotted Decimal Notation

Network prefix	Host number
10100111 11000111 10101010 010	10010
Dotted decimal notation:	
167.199.170.82 / 27 or 167.199.170.82 with netmask 255.255.255.224	
	
First: 167.199.170.64/27	10100111 11000111 10101010 010 <u>00000</u> (Network address)
Last: 167.199.170.95/27	10100111 11000111 10101010 010 <u>11111</u> (Broadcast address)

Six address blocks



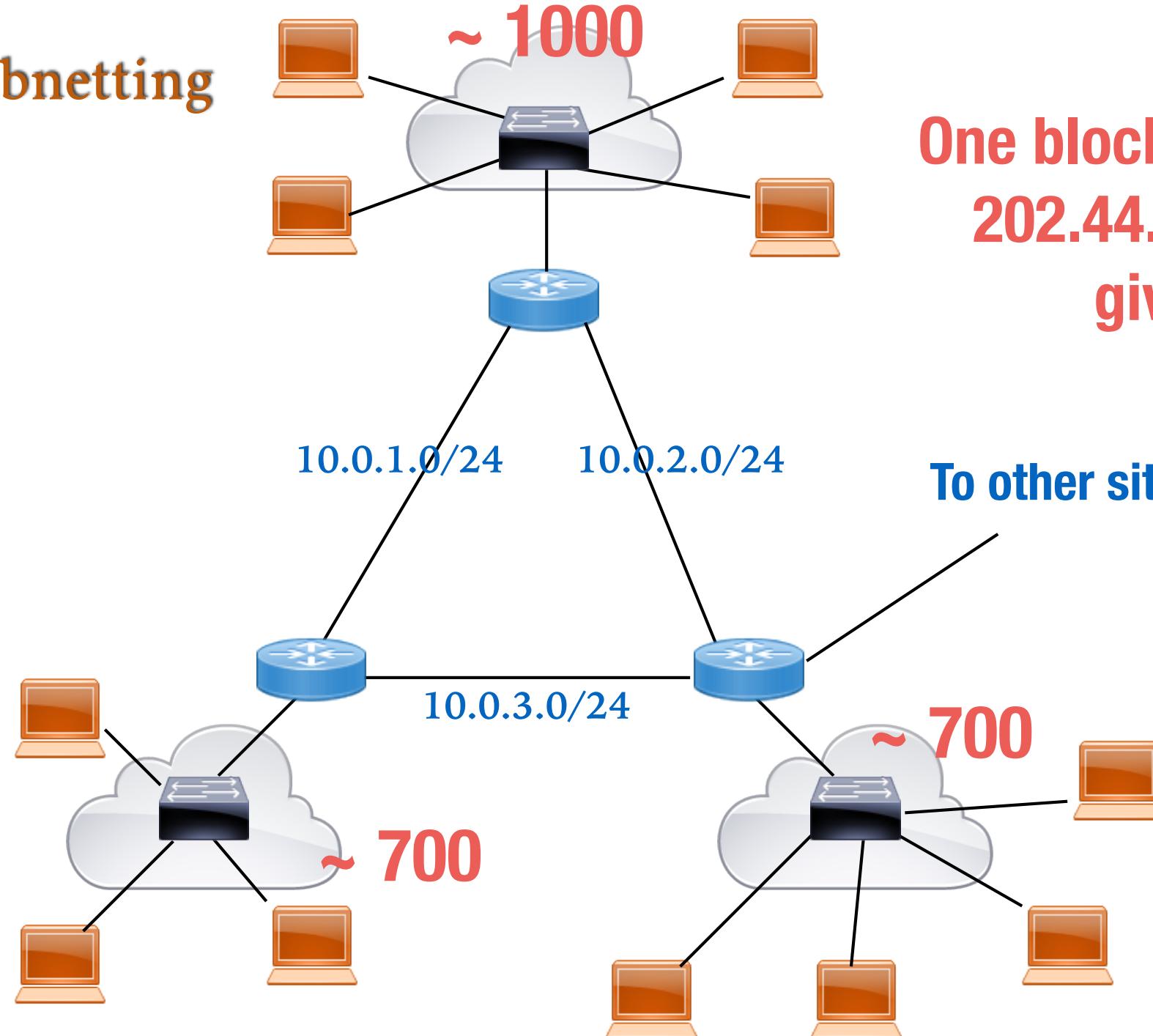
Self Test

- What is the maximum number of network interfaces that can be configured in an address block 202.44.12.0 / 26 ?
 - A. 30
 - B. 32
 - C. 62
 - D. 64

Some Special IP Address Blocks

- 127.0.0.0/8 are loopback addresses (only 127.0.0.1 in MacOS)
 - (Virtual) interface lo0 with IP address 127.0.0.1.
 - For debugging network program locally
- Private network addresses
 - 10.0.0.0/8 \approx 16M addresses
 - 172.16.0.0/12
 - 192.168.0.0/16

Subnetting



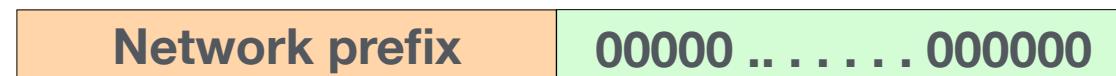
One block or prefix
202.44.12.0/20
given

- Address block can be divided and allocated to smaller subnetworks.

First: 202.44.0.0/20 11001010 00101100 0000**0000** 00000000

Last: 202.44.15.255/20 11001010 00101100 0000**1111** 11111111

Original block

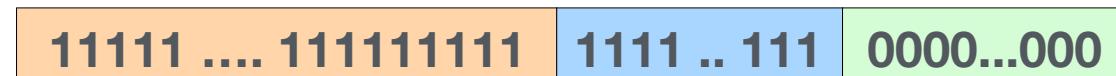


←———— Subnet prefix —————→

Subnet address



Subnetmask
or Netmask
or Address mask



ควรเป็น $2^{11} \rightarrow 2048$
ไว้การณ์ขนาดมันขยายขึ้น

ขนาดของblockต้องเป็น $2^n \rightarrow$ address
ที่ใกล้ที่สุดคือ $2^{10} \rightarrow 1024$

Faculty	Addresses needed	Addresses Allocated	Subnet blocks
Engineering	1000	1024	11001010 00101100 0000
Industrial Edu.	700	1024	11001010 00101100 0000
IT	700	1024	11001010 00101100 0000

Original block: 202.44.0.0/20

11001010 00101100 00000000 00000000

Addresses Allocated	Bits in host address	
1024	10	11001010 00101100 0000 0000 00000000
1024	10	11001010 00101100 0000 0100 00000000
1024	10	11001010 00101100 0000 1000 00000000

เลข subblock ที่ขึ้นมา



Original block: 202.44.0.0/20

11001010 00101100 00000000 00000000

Faculty		Address block
Engineering 0000 00000000	202.44.0.0/22
Industrial Edu. 0100 00000000	202.44.4.0/22
IT 1000 00000000	202.44.8.0/22

Original block: 202.44.0.0/20

11001010 00101100 00000000 00000000

Remaining block: 202.44.12.**0**/22

11001010 00101100 ~~0000111100~~ 00000000

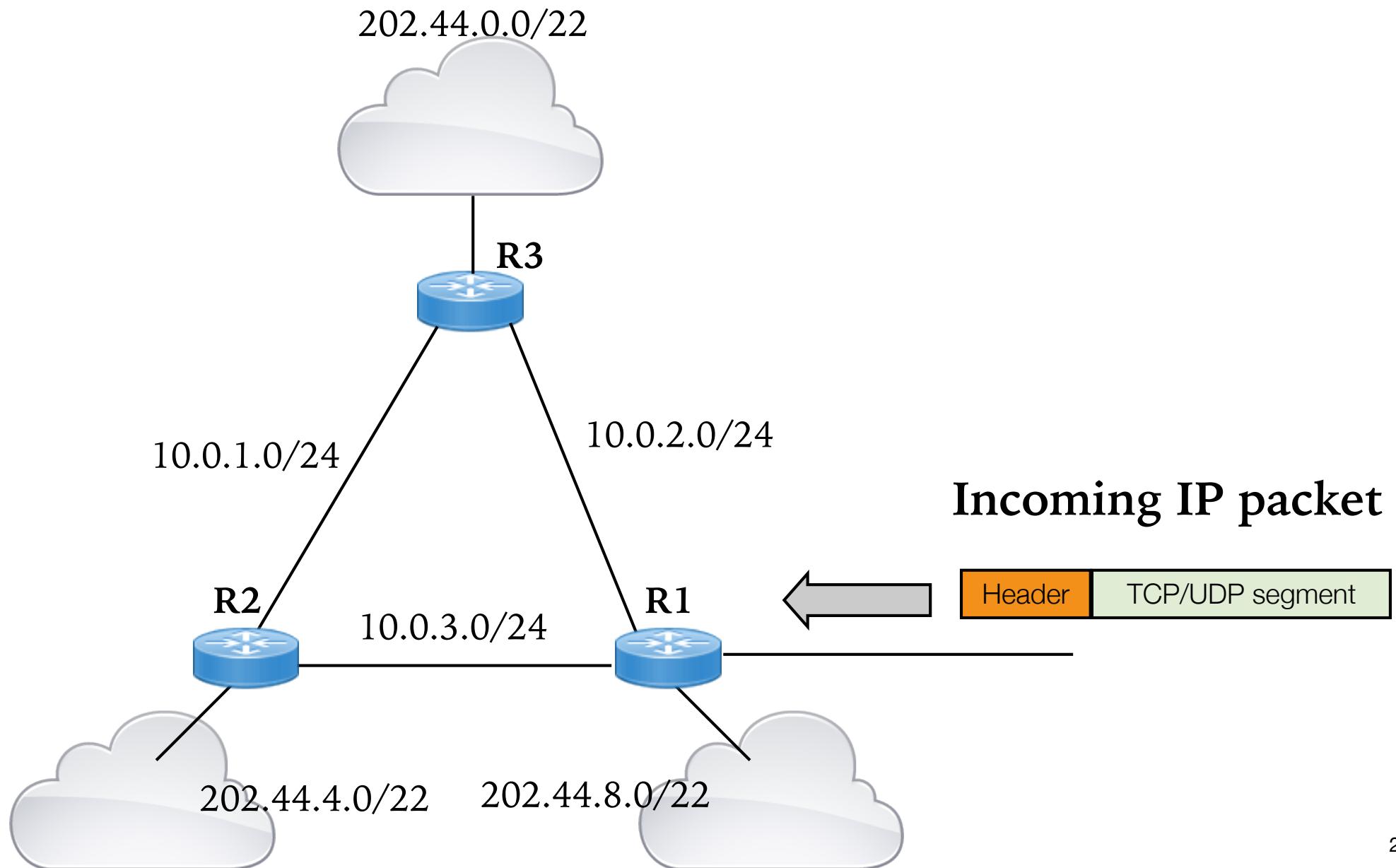
can be further split if necessary

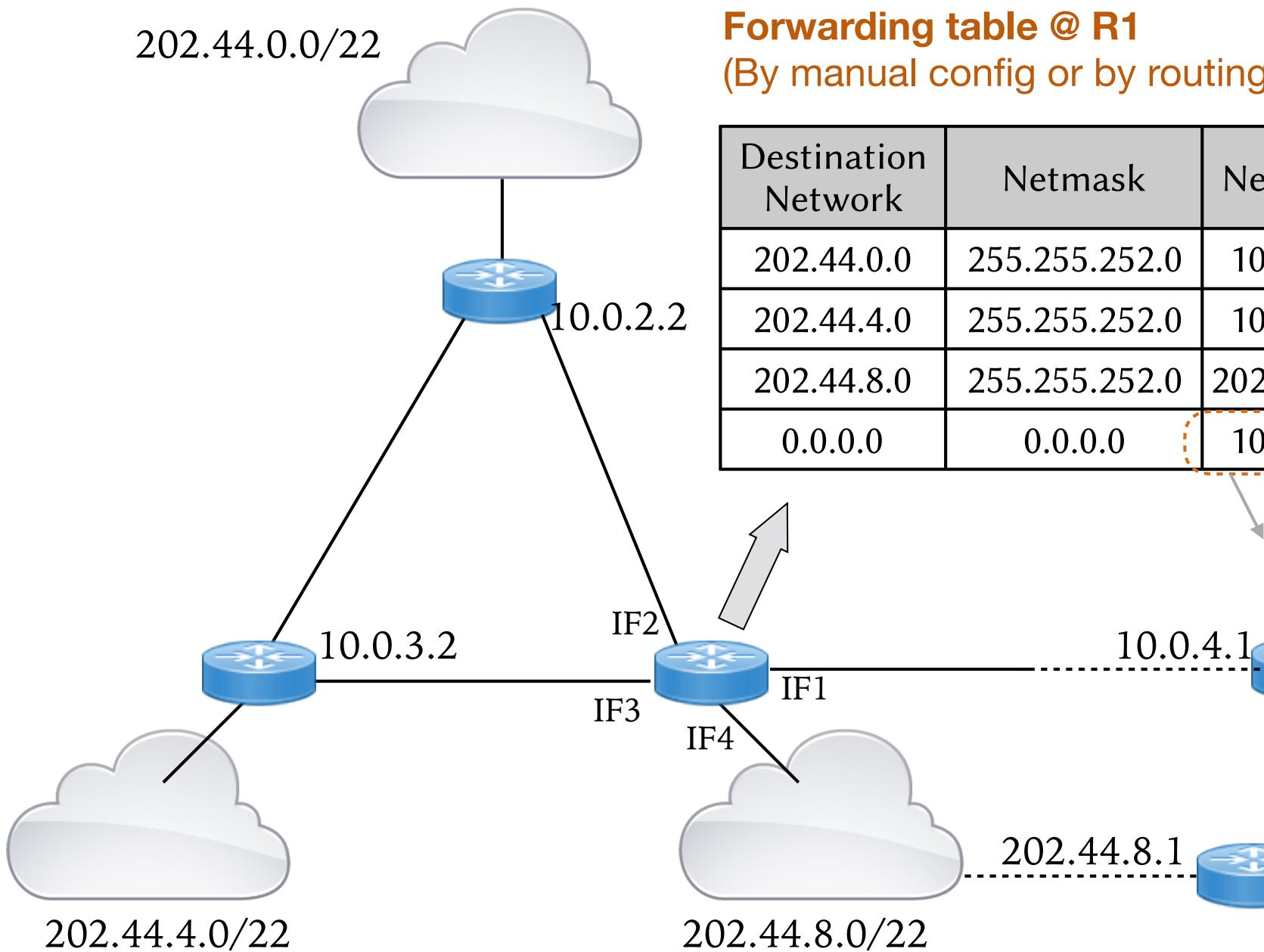
0000 1100

Why Subnetting?

- Performance
 - Broadcast traffic limit
 - Load separation
- Administration
 - Access control list
 - Policy enforcement
- Security
 - Servers better protected if hosts are compromised.

IP Packet Forwarding





Incoming IP packet

Header	TCP/UDP segment
--------	-----------------



Destination IP address
202.44.8.5

Forwarding table @ R1

Destination Network	Netmask	Nexthop	Interface
202.44.0.0	255.255.252.0	10.0.2.2	IF2
202.44.4.0	255.255.252.0	10.0.3.2	IF3
202.44.8.0	255.255.252.0	202.44.8.1	IF4
0.0.0.0	0.0.0.0	10.0.4.1	IF1

AND 255.255.252.0 = 202.44.8.0 (not matched)
AND 255.255.252.0 = 202.44.8.0 (not matched)
AND 255.255.252.0 = 202.44.8.0 (matched)
AND 0.0.0.0 = 0.0.0.0 (matched)

Longest-Prefix Matching

Incoming IP packet



Destination Network	Netmask	Nexthop	Interface
137.204.57.64	255.255.255.192	37.48.5.24	IF1
137.204.57.0	255.255.255.0	155.148.27.34	IF2
0.0.0.0	0.0.0.0	67.28.141.18	IF3

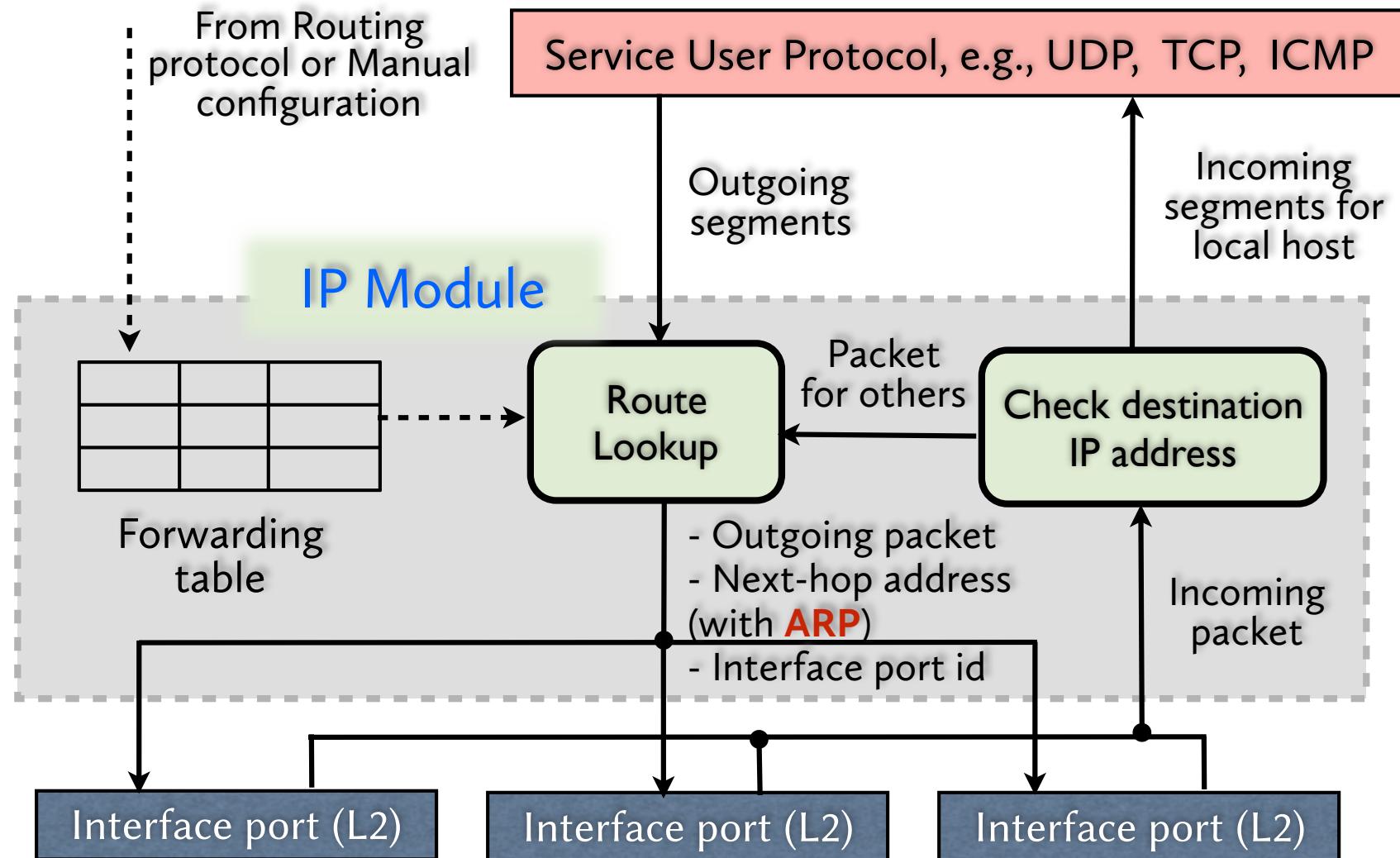
Destination IP address
137.204.57.210

AND 255.255.252.192 = 137.204.57.192 (not matched)

AND 255.255.255.0 = 137.204.57.0 (matched)

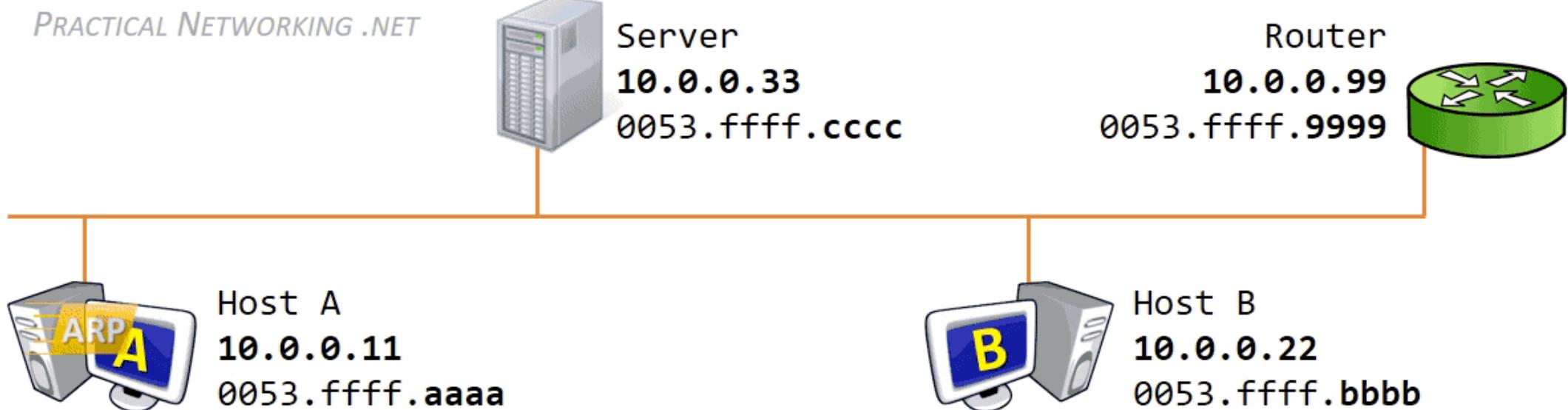
AND 0.0.0.0 = 0.0.0.0 (matched)

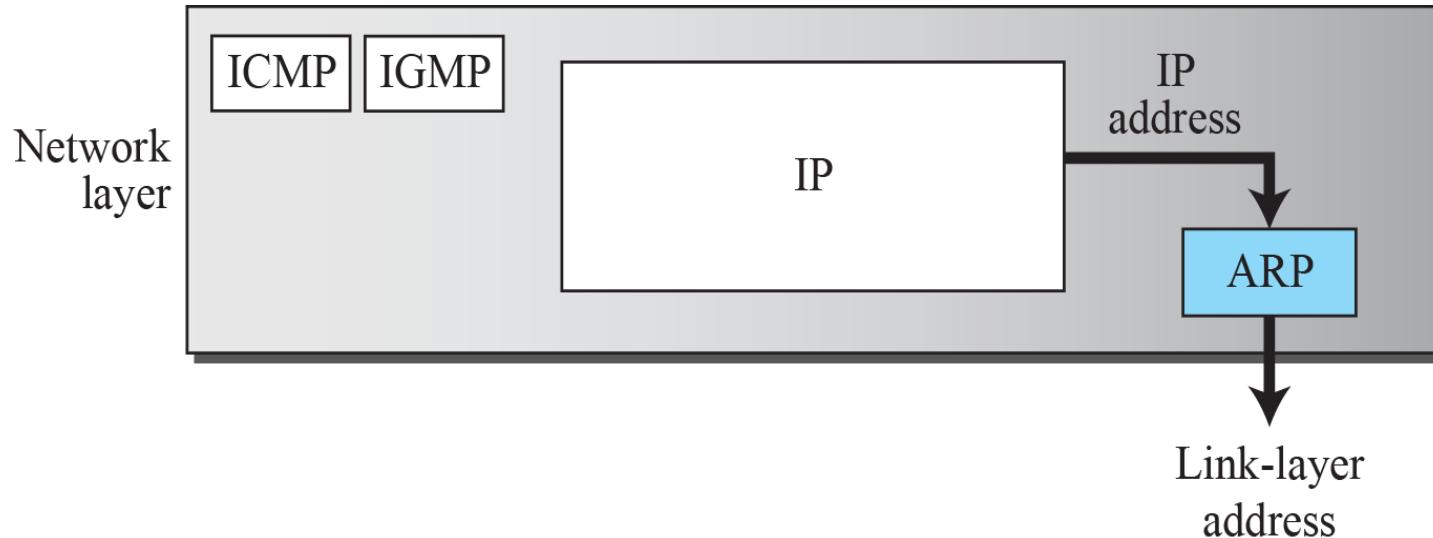
Processing of IP Datagram



Address Resolution Protocol (ARP)

- Resolve a hardware address from a given IP address





0	8	16	31
Hardware Type		Protocol Type	
Hardware length	Protocol length	Operation Request:1, Reply:2	
Source hardware address			
Source protocol address			
Destination hardware address (Empty in request)			
Destination protocol address			

Hardware: LAN or WAN protocol
Protocol: Network-layer protocol

ARP Request

Ethernet: en0

arp

No.	Time	Source	Destination	Protocol	Length	Info
13	3...	Apple_ca:ac:b8	Broadcast	ARP	42	Who has 10.35.21.2? Tell 10.35.21.148
14	3...	3comEuro_74:13:01	Apple_ca:ac:b8	ARP	60	10.35.21.2 is at 00:1e:c1:74:13:01

► Frame 13: 42 bytes on wire (336 bits), 42 bytes captured (336 bits) on interface 0

▼ Ethernet II, Src: Apple_ca:ac:b8 (98:5a:eb:ca:ac:b8), Dst: Broadcast (ff:ff:ff:ff:ff:ff)

- Destination: Broadcast (ff:ff:ff:ff:ff:ff)
- Source: Apple_ca:ac:b8 (98:5a:eb:ca:ac:b8)
- Type: ARP (0x0806)

▼ Address Resolution Protocol (request)

- Hardware type: Ethernet (1)
- Protocol type: IPv4 (0x0800)
- Hardware size: 6
- Protocol size: 4
- Opcode: request (1)
- Sender MAC address: Apple_ca:ac:b8 (98:5a:eb:ca:ac:b8)
- Sender IP address: 10.35.21.148
- Target MAC address: 00:00:00_00:00:00 (00:00:00:00:00:00)
- Target IP address: 10.35.21.2

0000	ff ff ff ff ff ff 98 5a eb ca ac b8 08 06 00 01Z
0010	08 00 06 04 00 01 98 5a eb ca ac b8 0a 23 15 94Z#..
0020	00 00 00 00 00 00 0a 23 15 02# ...

Frame (frame), 42 bytes

Packets: 17 · Displayed: 2 (11.8%)

Profile: Default

ARP Reply

Ethernet: en0

No.	Time	Source	Destination	Protocol	Length	Info
13	3...	Apple_ca:ac:b8	Broadcast	ARP	42	Who has 10.35.21.2? Tell 10.35.21.148
14	3...	3comEuro_74:13:01	Apple_ca:ac:b8	ARP	60	10.35.21.2 is at 00:1e:c1:74:13:01

Frame 14: 60 bytes on wire (480 bits), 60 bytes captured (480 bits) on interface 0

Ethernet II, Src: 3comEuro_74:13:01 (00:1e:c1:74:13:01), Dst: Apple_ca:ac:b8 (98:5a:eb:ca:ac:b8)

Destination: Apple_ca:ac:b8 (98:5a:eb:ca:ac:b8)

Source: 3comEuro_74:13:01 (00:1e:c1:74:13:01)

Type: ARP (0x0806)

Padding: 00

Address Resolution Protocol (reply)

Hardware type: Ethernet (1)

Protocol type: IPv4 (0x0800)

Hardware size: 6

Protocol size: 4

Opcode: reply (2)

Sender MAC address: 3comEuro_74:13:01 (00:1e:c1:74:13:01)

Sender IP address: 10.35.21.2

Target MAC address: Apple_ca:ac:b8 (98:5a:eb:ca:ac:b8)

Target IP address: 10.35.21.148

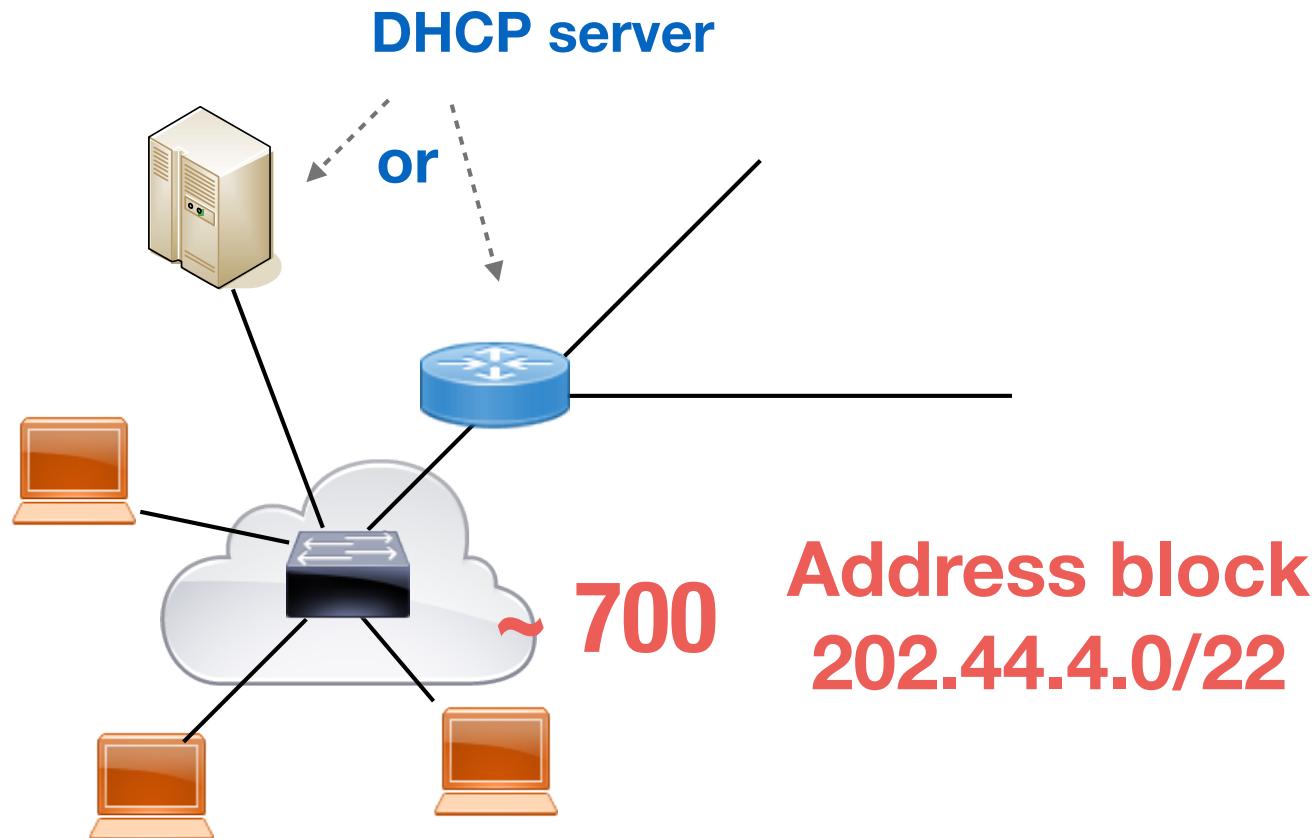
0000	98	5a	eb	ca	ac	b8	00	1e	c1	74	13	01	08	06	00	01	.Z.....	.t.....
0010	08	00	06	04	00	02	00	1e	c1	74	13	01	0a	23	15	02t....#...
0020	98	5a	eb	ca	ac	b8	0a	23	15	94	00	00	00	00	00	00	.Z.....#
0030	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00

Address Resolution Protocol (arp), 28 bytes

Packets: 17 · Displayed: 2 (11.8%)

Profile: Default

Dynamic Host Configuration Protocol (DHCP)



Network Setup (LAN)

Gateway IP

Local IP Address:

192 . 168 . 1 . 1

Subnet Mask:

255 . 255 . 255 . 0

Warning: Changes to LAN IP network settings may require reconfiguration of all attached devices. Some network devices may be out of service until the change is detected.

Network Address Server Settings (DHCP)

DHCP Server:

Enable Disable

[Connected Devices Summary](#)

[Pre-assigned DHCP IP Addresses](#)

Starting IP Address:

192.168.1. 41

Maximum Number of
DHCP Users:

128

Client Lease Time:

1440 minutes (0 means one day)

LAN 1 Static DNS 1:

0 . 0 . 0 . 0

LAN 1 Static DNS 2:

0 . 0 . 0 . 0

LAN 1 Static DNS 3:

0 . 0 . 0 . 0

DHCP Operations

Client



IP Address: ?

DHCPDISCOVER

Transaction ID: 1001
Lease time:
Client address:
Your address:
Server address:
Source port: 68 Destination port: 67
Source address: 0.0.0.0
Destination address: 255.255.255.255.

IP Address: 181.14.16.170

Server



Legend

Application
UDP
IP

DHCPREQUEST

Transaction ID: 1001
Lease time: 3600
Client address: 181.14.16.182
Your address:
Server address: 181.14.16.170
Source port: 68 Destination port: 67
Source address: 181.14.16.182
Destination address: 255.255.255.255.

DHCPOFFER

Transaction ID: 1001
Lease time: 3600
Client address:
Your address: 181.14.16.182
Server address: 181.14.16.170
Source port: 67 Destination port: 68
Source address: 181.14.16.170
Destination address: 255.255.255.255.

ARP/ICMP requests
sent before offering

DHCPACK

Transaction ID: 1001
Lease time: 3600
Client address:
Your address: 181.14.16.182
Server address: 181.14.16.170
Source port: 67
Source address: 181.14.16.170
Destination address: 255.255.255.255.

X dhcp.pcapng [Wireshark 1.10.4 (SVN Rev 54184 from /trunk-1.10)]

File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help

Filter: bootp Expression... Clear Apply Save

No.	Time	Source	Destination	Protocol	Length	Info
49	13.546210000	192.168.2.152	192.168.2.1	DHCP	342	DHCP Release - Transaction ID 0x8ab8e6f6
50	13.546730000	0.0.0.0	255.255.255.255	DHCP	342	DHCP Discover - Transaction ID 0xb47dc89
56	13.583942000	192.168.2.1	255.255.255.255	DHCP	342	DHCP Offer - Transaction ID 0xb47dc89

Bootstrap Protocol

- Message type: Boot Request (1)
- Hardware type: Ethernet (0x01)
- Hardware address length: 6
- Hops: 0
- Transaction ID: 0xb47dc89
- Seconds elapsed: 0

Bootp flags: 0x0000 (Unicast)

- Client IP address: 0.0.0.0 (0.0.0.0)
- Your (client) IP address: 0.0.0.0 (0.0.0.0)
- Next server IP address: 0.0.0.0 (0.0.0.0)
- Relay agent IP address: 0.0.0.0 (0.0.0.0)
- Client MAC address: Apple_33:c9:54 (b8:e8:56:33:c9:54)
- Client hardware address padding: 00000000000000000000000000000000
- Server host name not given
- Boot file name not given
- Magic cookie: DHCP

Option: (53) DHCP Message Type

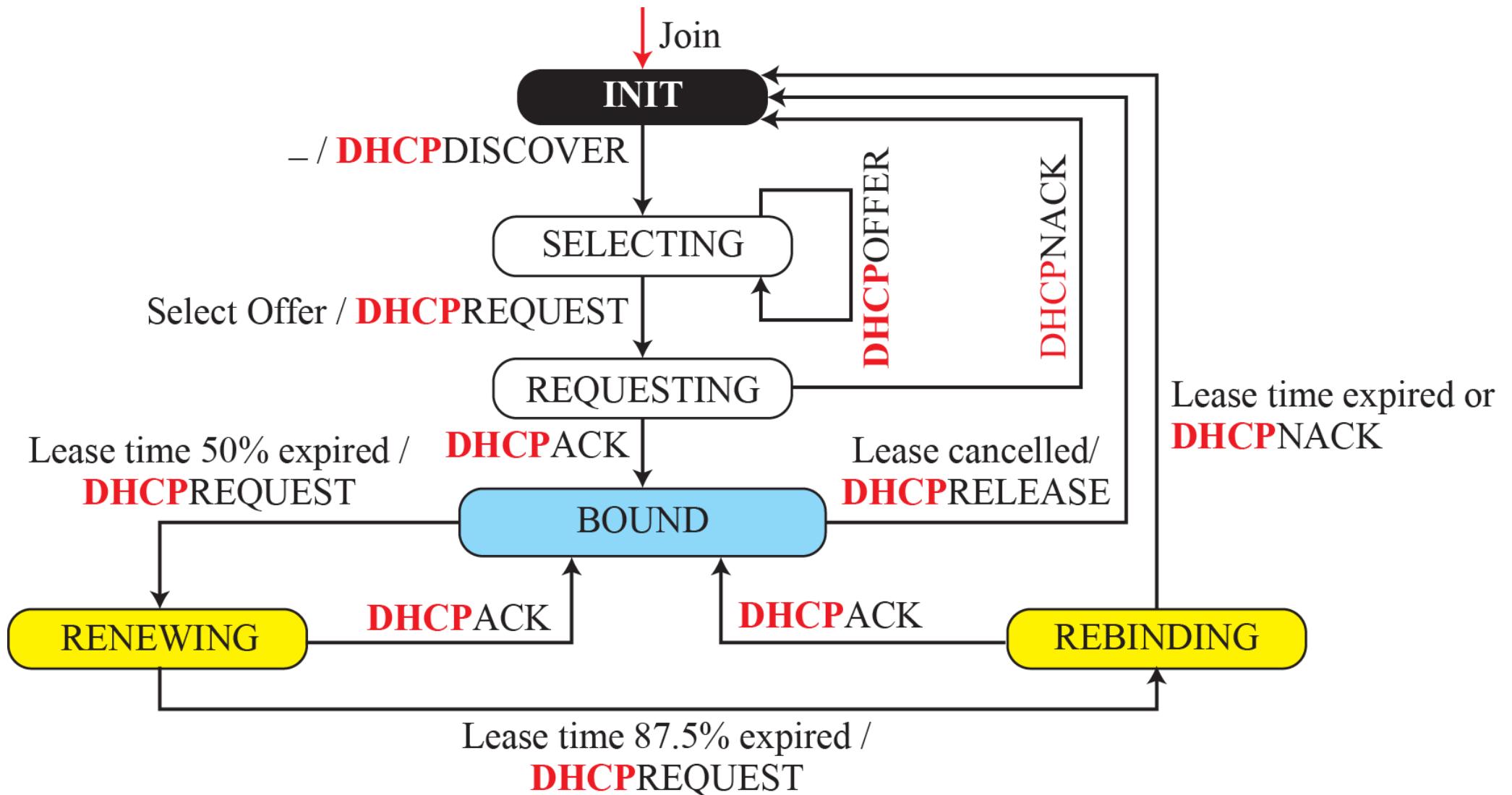
- Length: 1
- DHCP: Discover (1)

Option: (55) Parameter Request List

0000	ff ff ff ff ff b8 e8	56 33 c9 54 08 00 45 00 V3.T..E.
0010	01 48 4f e0 00 00 ff 11	6a c5 00 00 00 00 ff ff	.HO..... j.....
0020	ff ff 00 44 00 43 01 34	eb 12 01 01 06 00 0b 47	...D.C.4G
0030	dc 89 00 00 00 00 00 00	00 00 00 00 00 00 00 00
0040	00 00 00 00 00 00 b8 e8	56 33 c9 54 00 00 00 00 V3.T....

File: "/Users/peerapon/Dropbox..." Packets: 5618 · Displayed: 18 (0%) Profile: Default

Finite State Machine (FSM) of DHCP

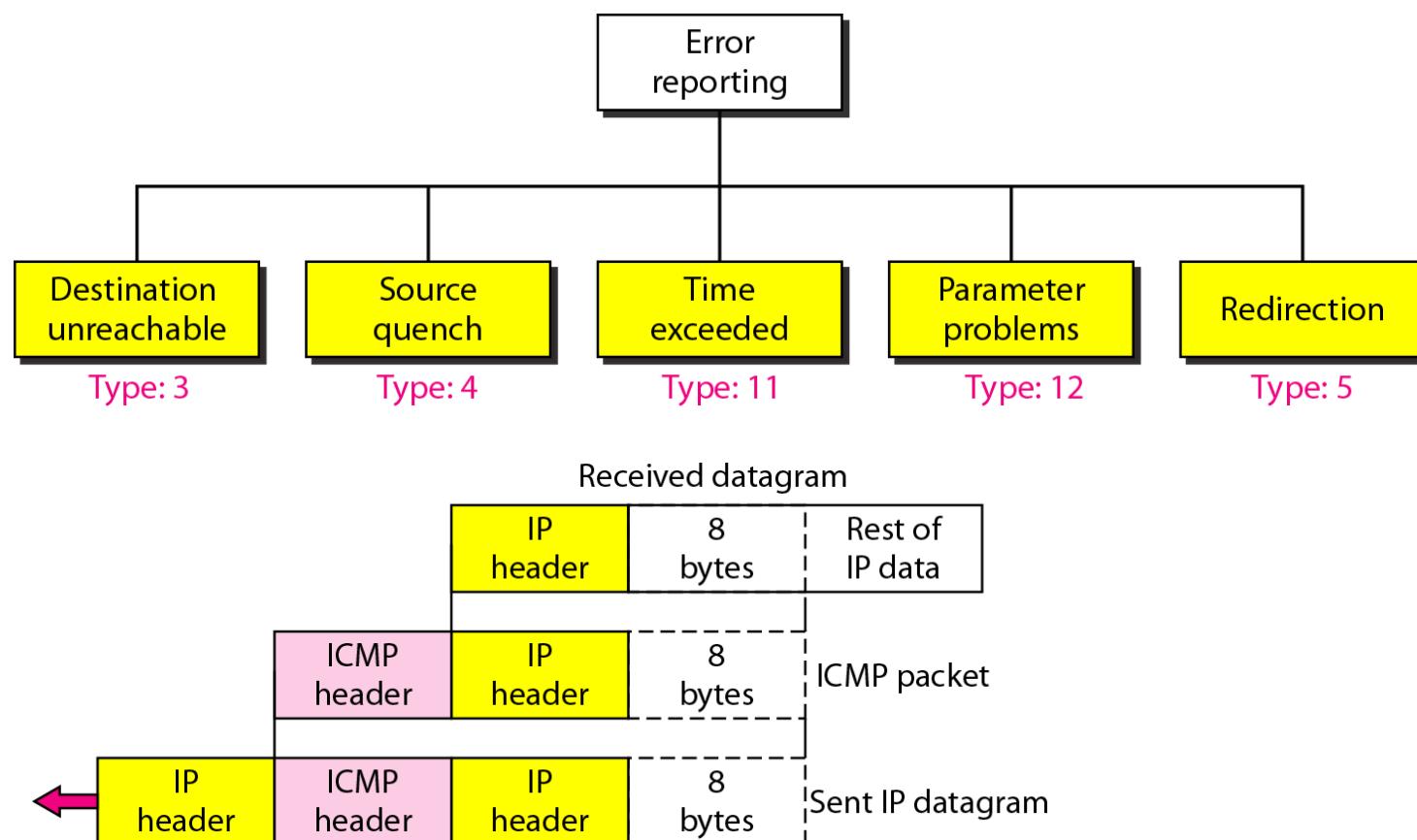


Internet Control Message Protocol (ICMP)

- Supplementary L3 protocol to report problems and communicate network-layer information.
- Two broad categories of ICMP messages
 - **Error reporting**: unreachable host, network, port, protocol
 - **Query**: diagnose network problems, connectivity, interface status.
- Network utility tools based on ICMP
 - Ping
 - Traceroute (MS Windows)

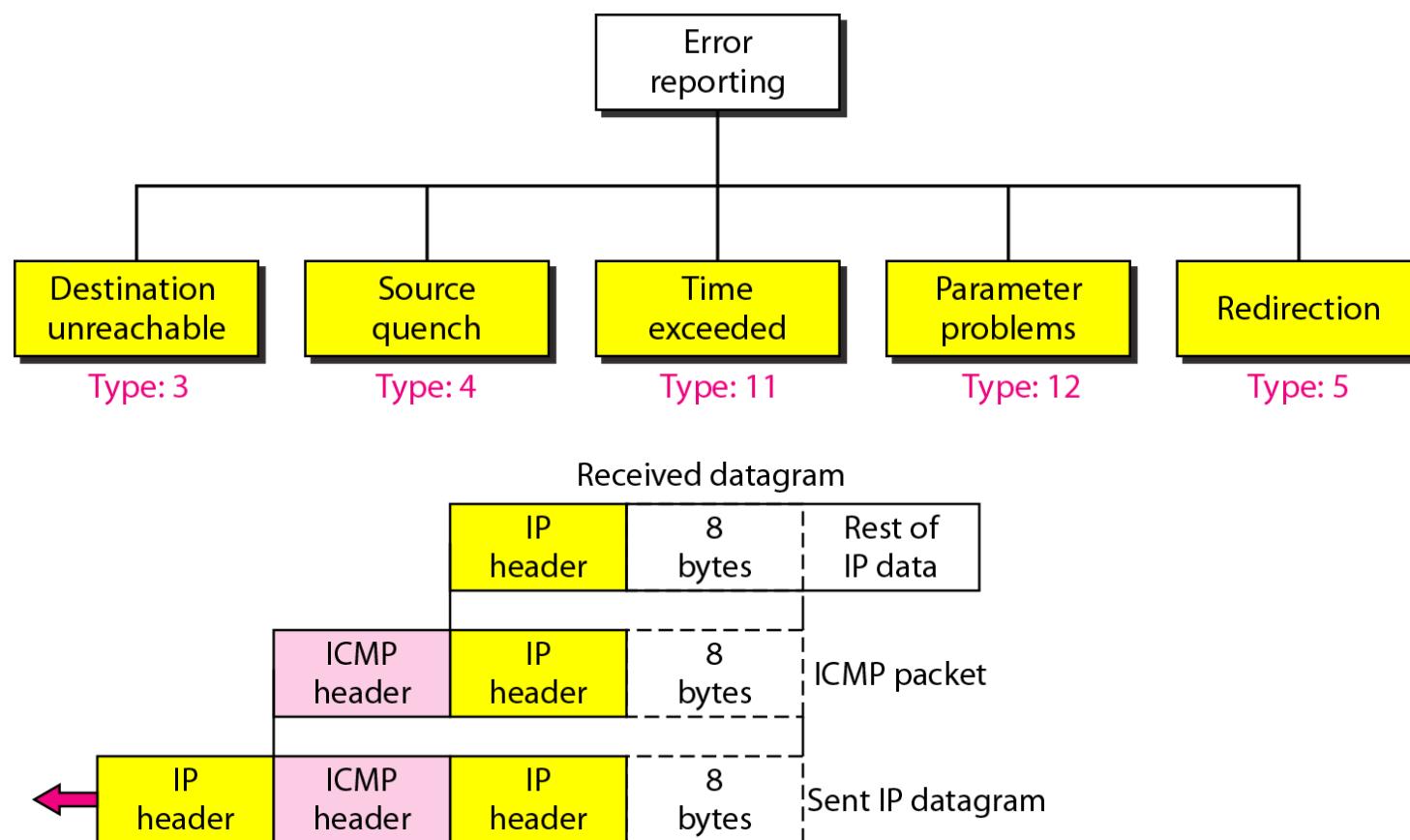
ICMP Error Report Messages

- ICMP Error report message contains “type”, “code”, and first 8 bytes of the IP datagram causing error.



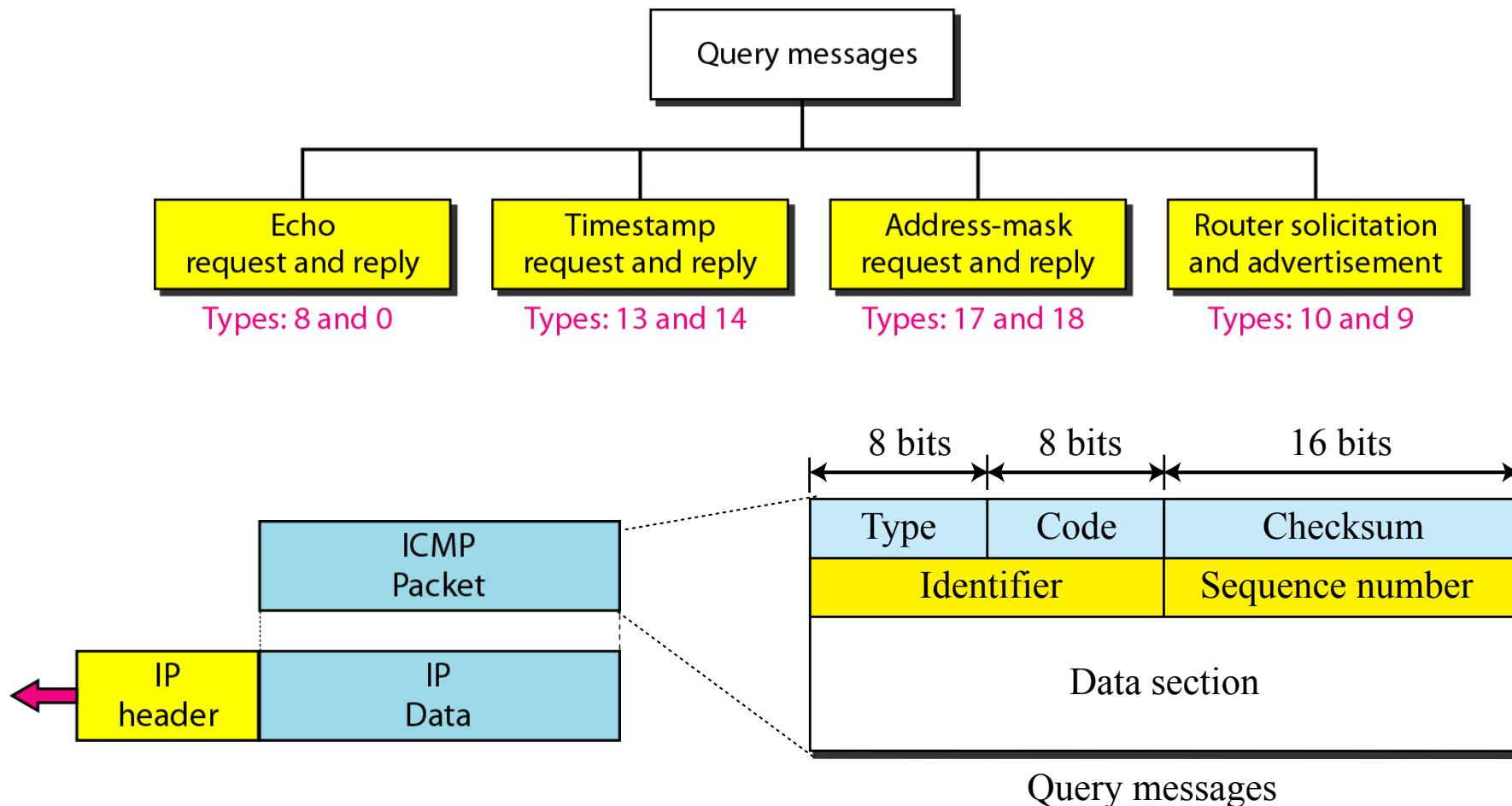
ICMP Error Report Messages

- ICMP Error report message contains “type”, “code”, and first 8 bytes of the IP datagram causing error.



ICMP Query Messages

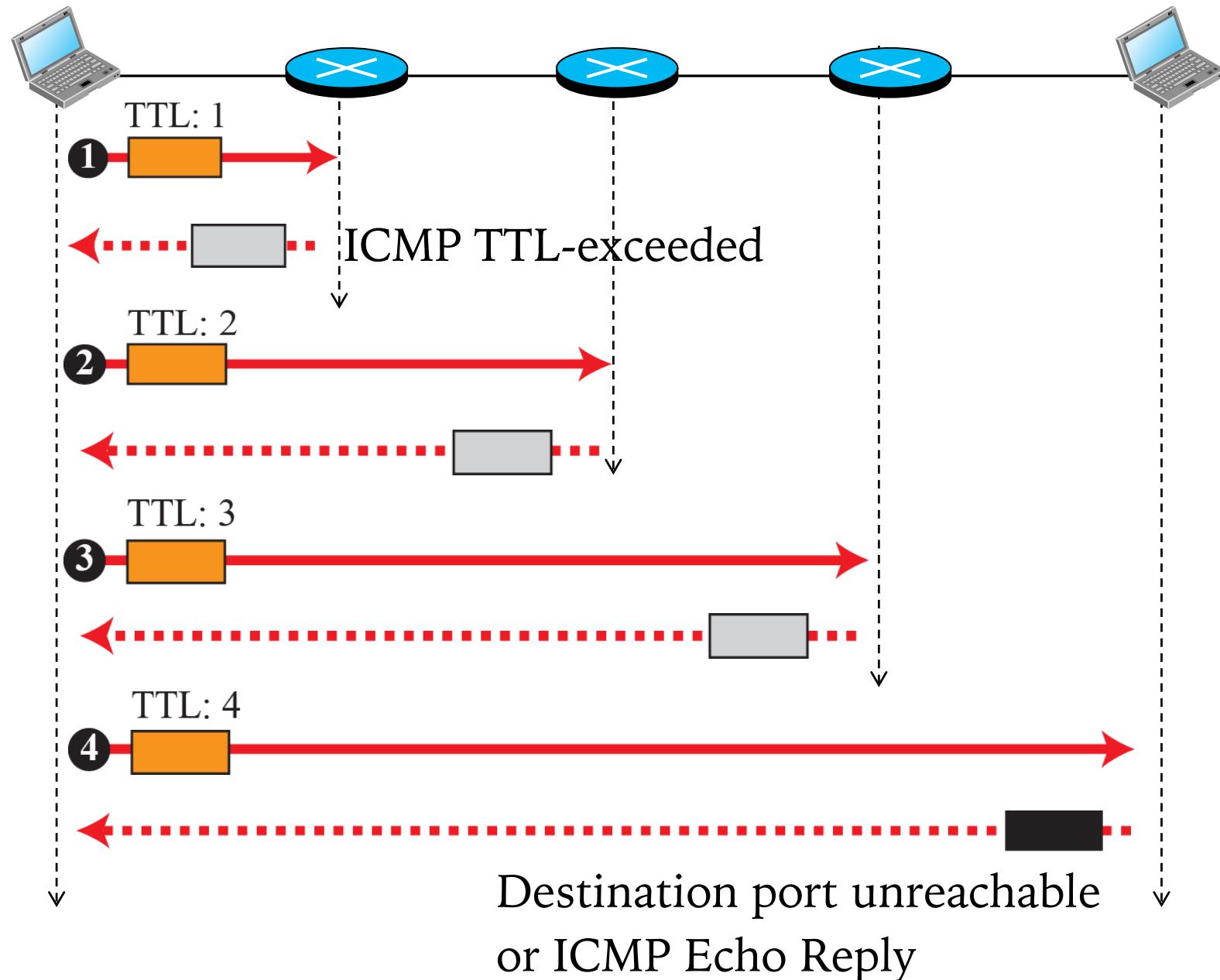
- The ping utility uses Echo request and reply to check the network connectivity.



Traceroute Utility

Message types

- Traceroute
- Time-exceeded
- Destination-unreachable



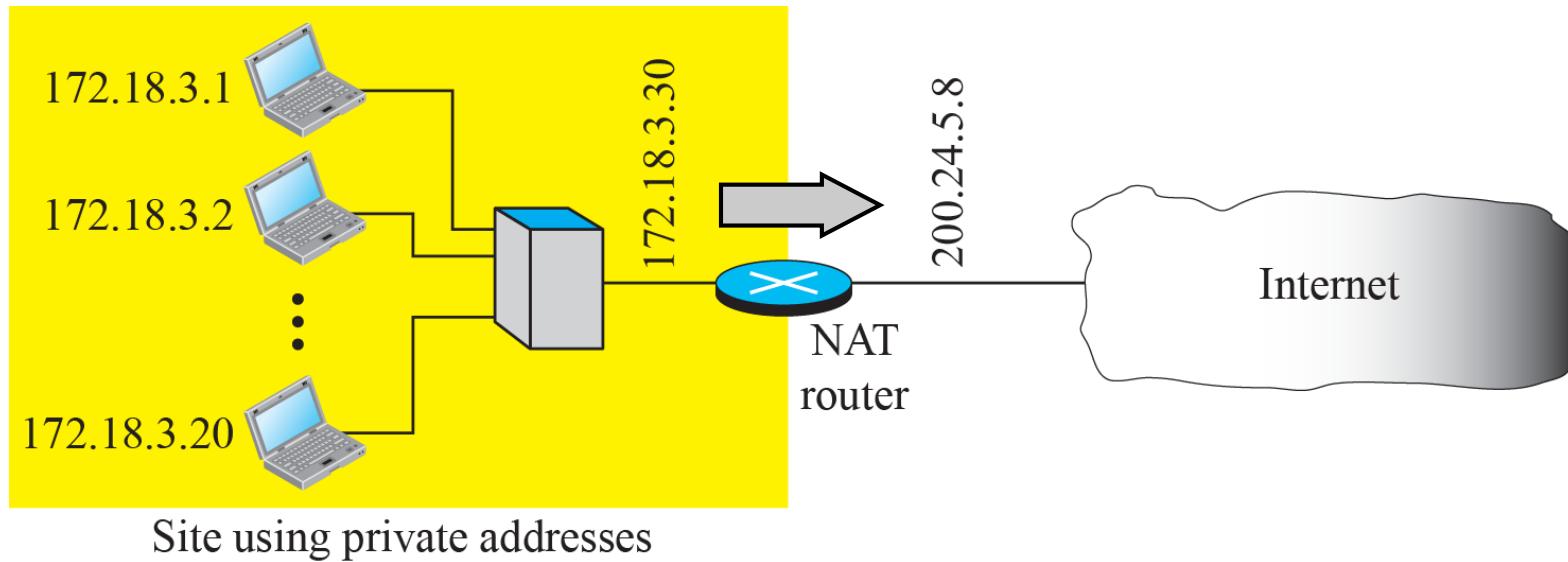
Some ICMP Types and Codes

Type	Code	Description
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable (UDP)
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

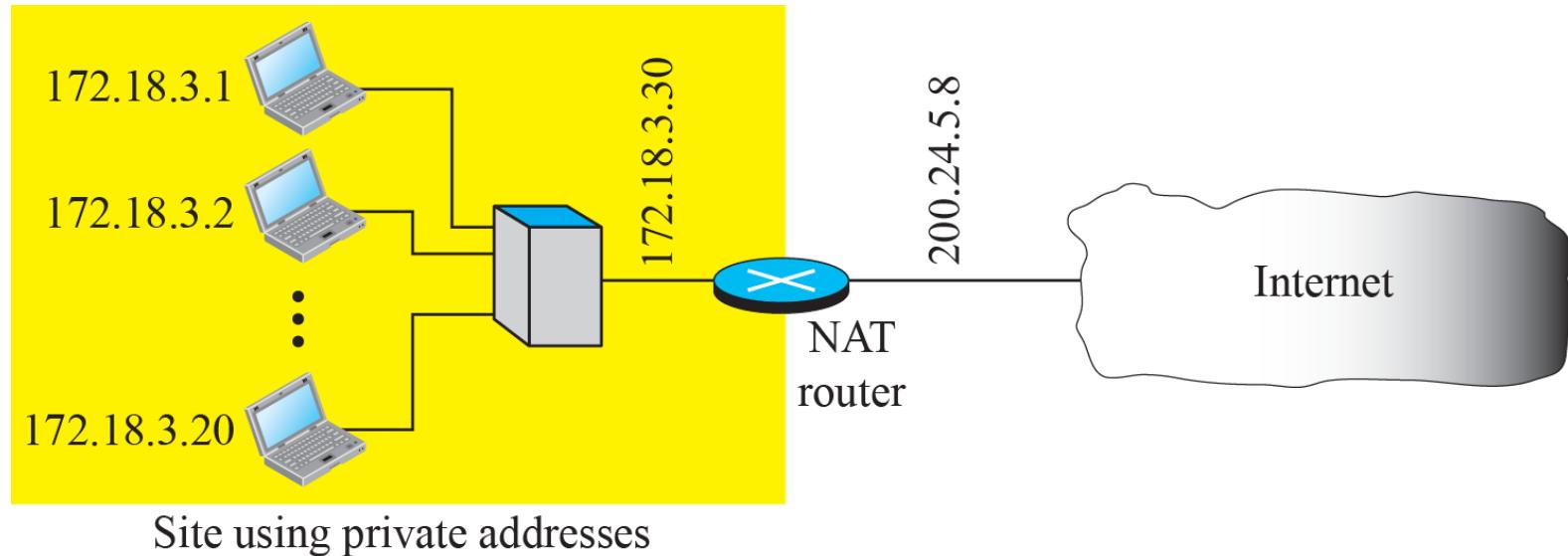
Network Address Translation^{RFC 3022}

- Too many interfaces, too few IP addresses
- Allow many private hosts to share a public IP address
 - 10.0.0.0 to 10.255.255.255 (10.0.0.0/8) ~ 16M
 - 172.16.0.0 to 172.31.255.255 (172.16.0.0/12) ~ 1M
 - 192.168.0.0 to 192.168.255.255 (192.168.0.0/16, in home LANs) ~ 64k

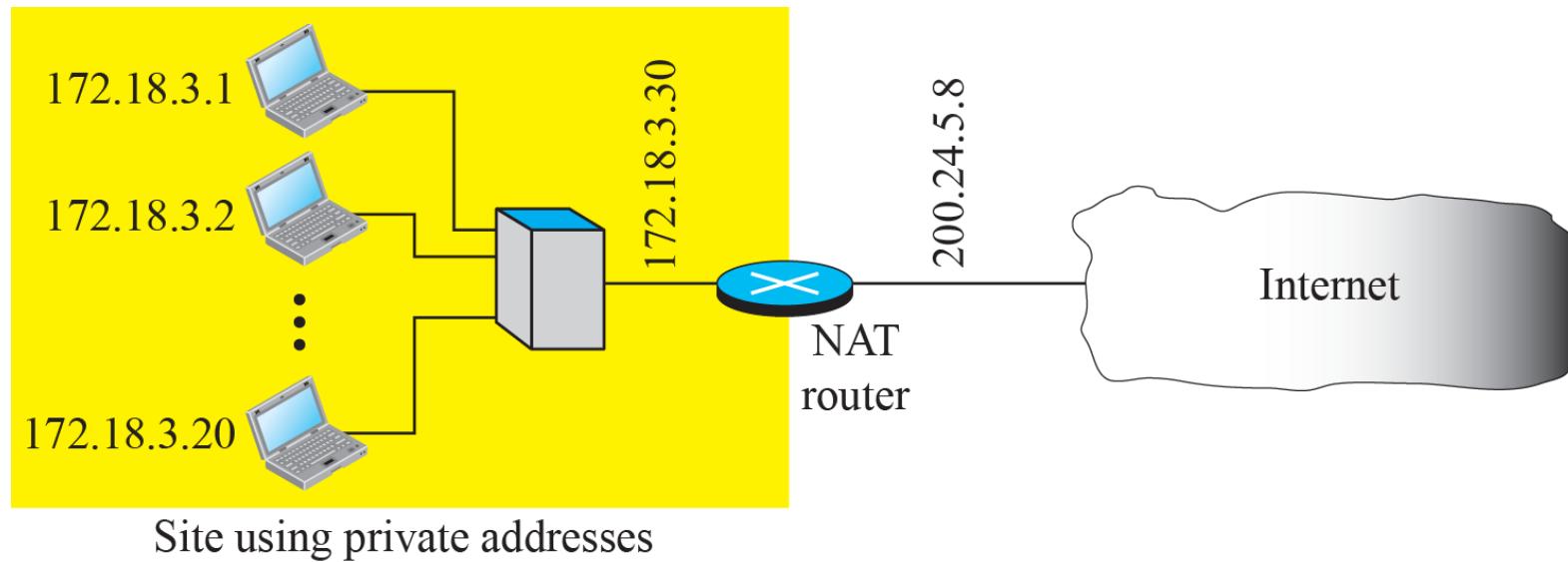
Network Address (and Port) Translation: NAT (NAPT)



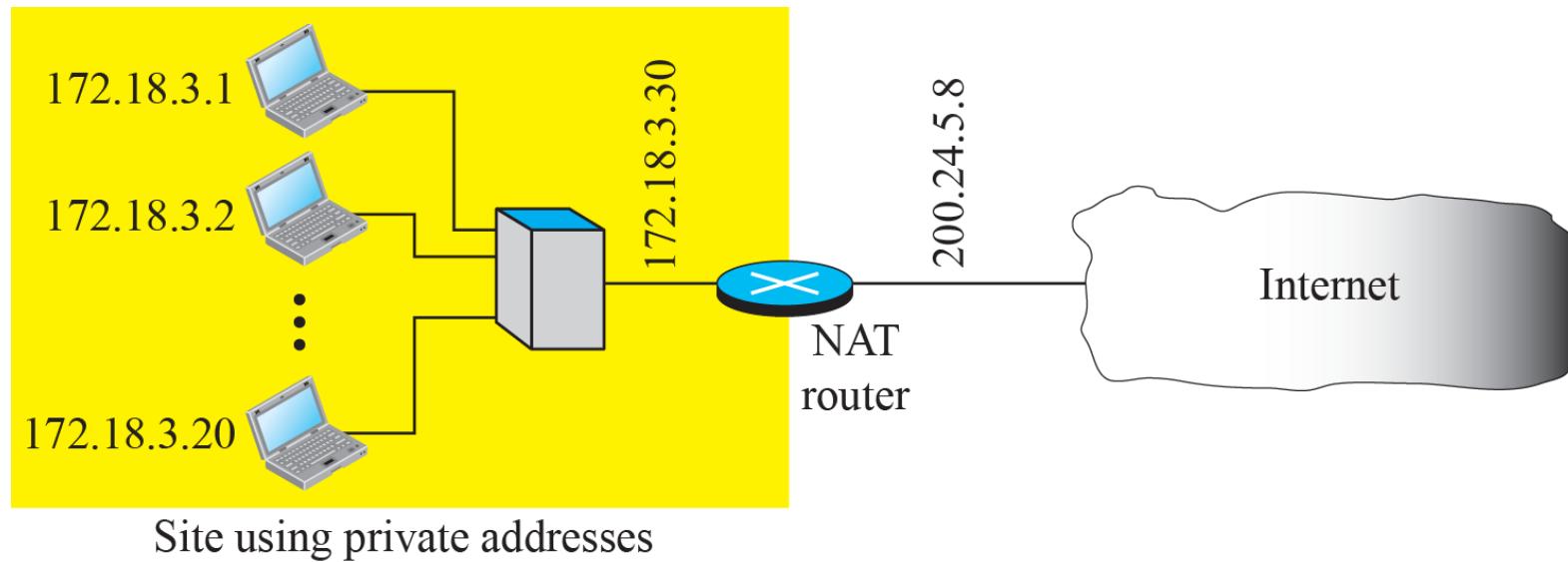
Private network		External network	
IP address	Port number	IP address	Port number



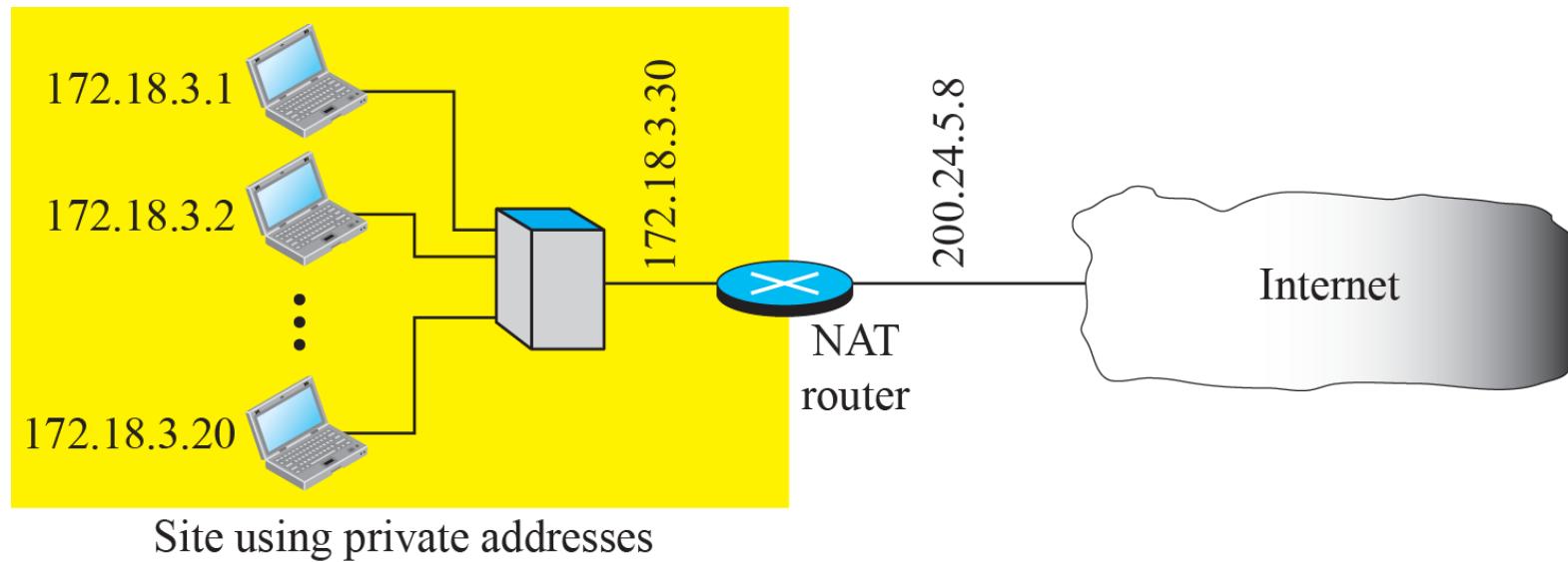
Private network		External network	
IP address	Port number	IP address	Port number
172.18.3.1	1400		



Private network		External network	
IP address	Port number	IP address	Port number
172.18.3.1	1400	200.24.5.8	2567



Private network		External network	
IP address	Port number	IP address	Port number
172.18.3.1	1400	200.24.5.8	2567
172.18.3.2	1401		

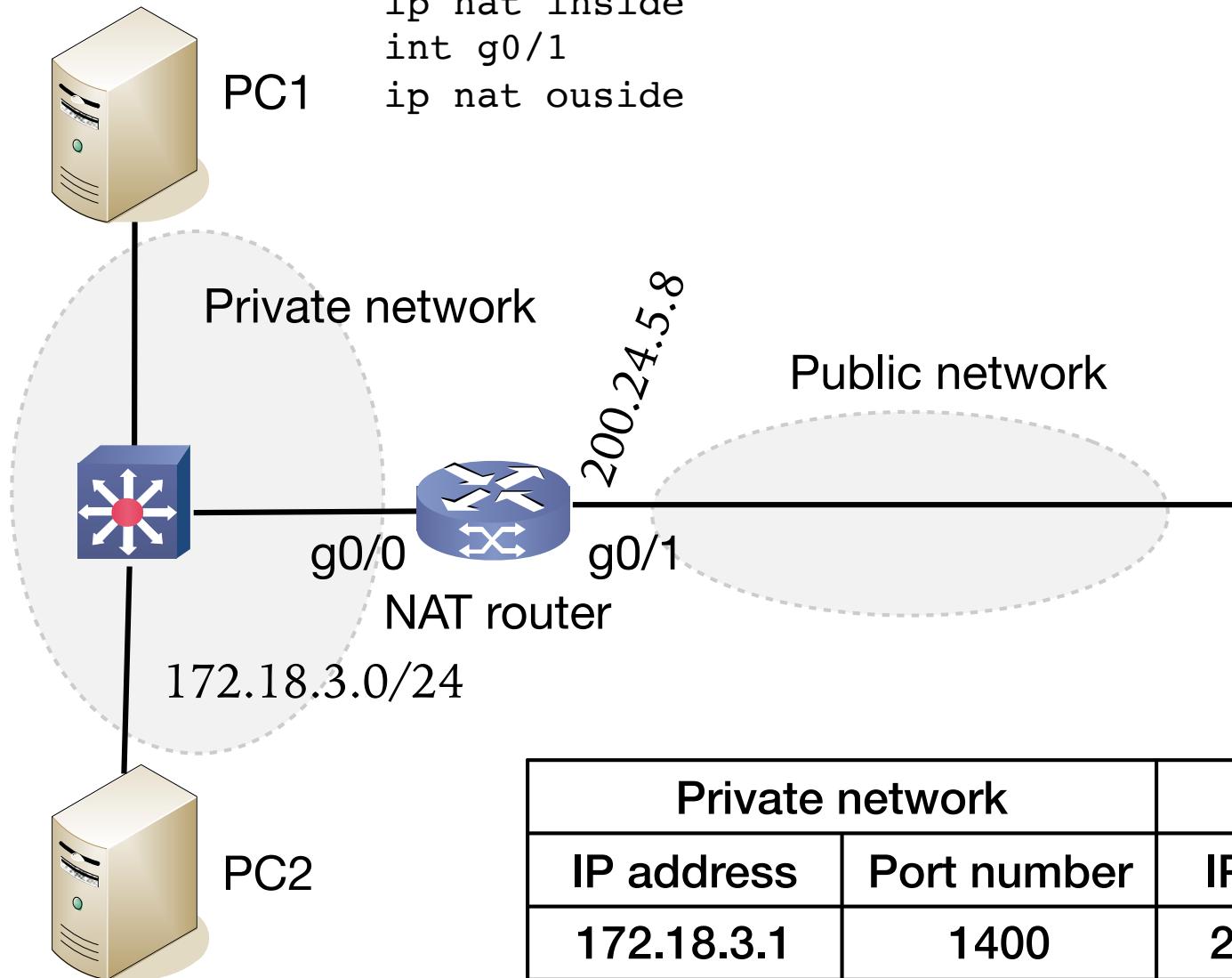


Private network		External network	
IP address	Port number	IP address	Port number
172.18.3.1	1400	200.24.5.8	2567
172.18.3.2	1401	200.24.5.8	3345

```

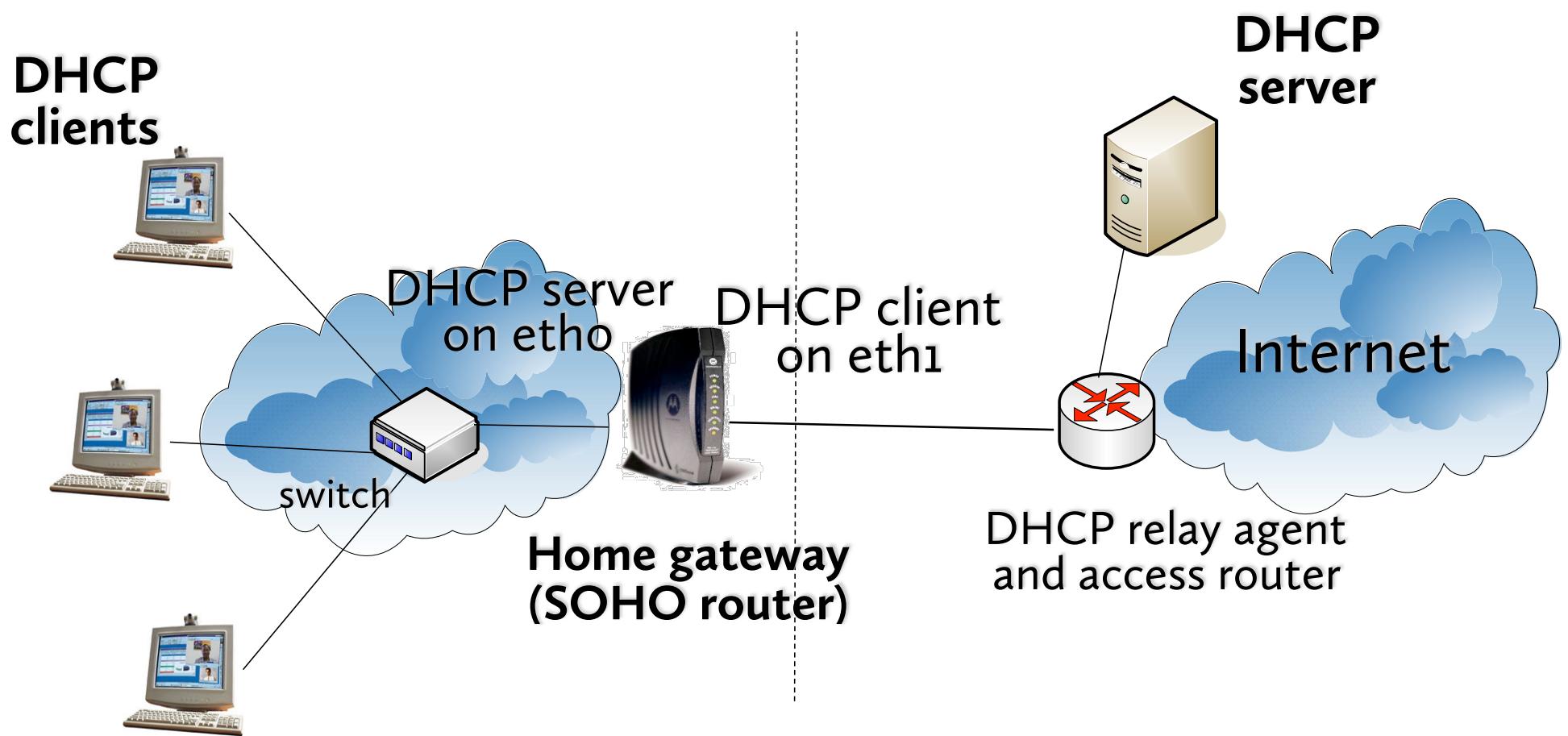
ip nat pool net-172 200.24.5.8 200.24.5.223 prefix-length 24
access-list 1 permit 172.18.3.0 0.0.0.255
ip nat inside source list 1 pool net-172
int g0/0
ip nat inside
int g0/1
ip nat outside

```



Private network		External network	
IP address	Port number	IP address	Port number
172.18.3.1	1400	200.24.5.8	2567
172.18.3.2	1401	200.24.5.8	3345

Example: SOHO Network



Notes

□ Advantages of NAT

- Address space saving, Independent of ISPs
- Local reconfigurable without side effect
- Security

□ Issues

- ICMP packets have no port number
- Server also runs behind a NAT -- **Port redirection**
- Some apps encode IP address/port in app layer

Port Redirection/Forwarding

- Allow to run servers behind NAT
- Request for a given port number is mapped to a specific internal (private) IP address and internal port number.

External		Internal				
Start Port	End Port	IP Address		Start Port	End Port	Protocol
0	to	0	0.0.0.0	0	to	0
0	to	0	0.0.0.0	0	to	0
0	to	0	0.0.0.0	0	to	0
0	to	0	0.0.0.0	0	to	0
0	to	0	0.0.0.0	0	to	0

Pre-assigned DHCP IP Addresses

External	Internal	Protocol	Enable
0 to 0	0.0.0.0	0 to 0	TCP
0 to 0	0.0.0.0	0 to 0	UDP
0 to 0	0.0.0.0	0 to 0	Both
0 to 0	0.0.0.0	0 to 0	TCP
0 to 0	0.0.0.0	0 to 0	TCP

TCP
 UDP
 Both

Pre-assigned DHCP IP Addresses

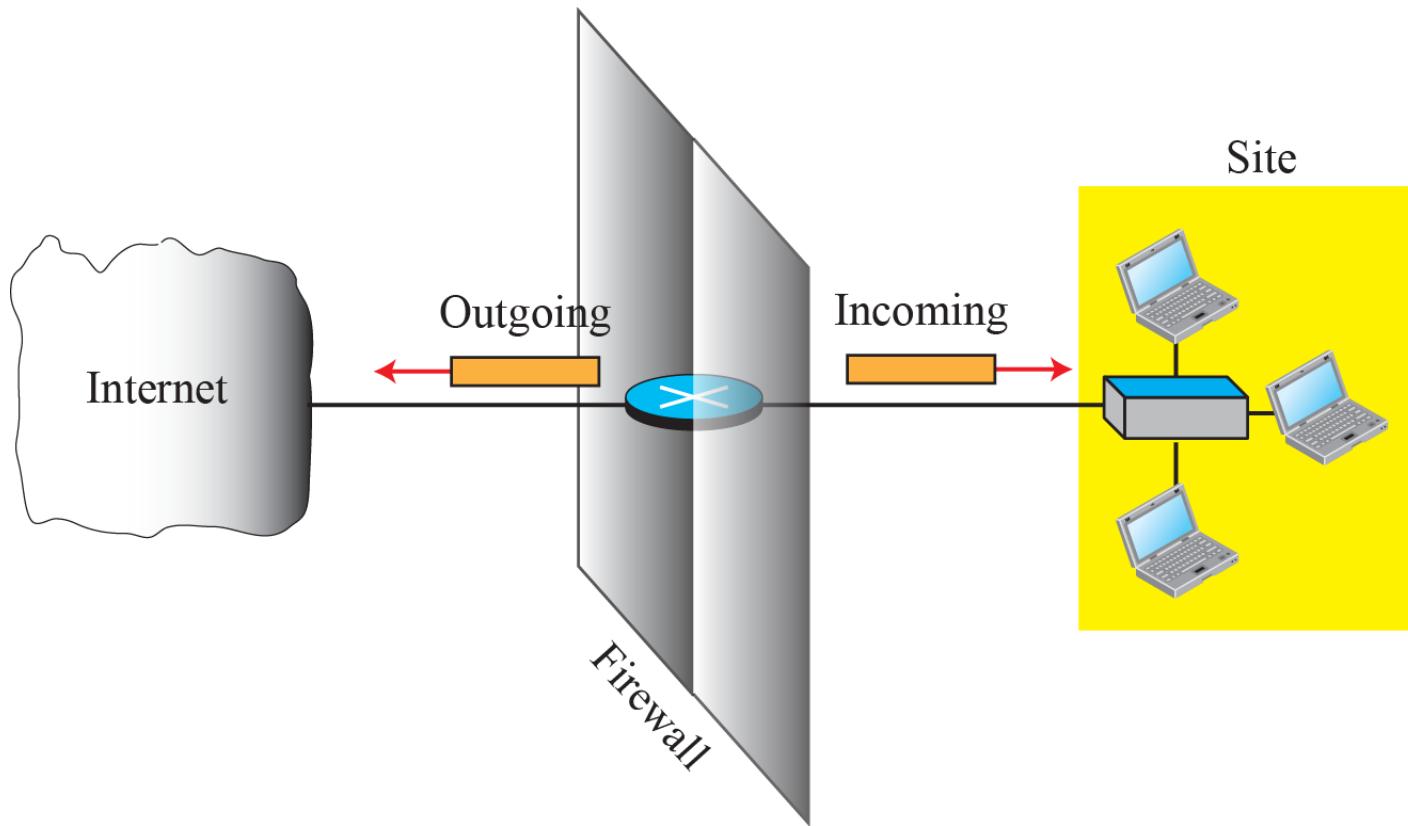
MAC Address: 00:00:00:00:00:00

Assign to IP: _____

_____ . _____ . _____ . _____

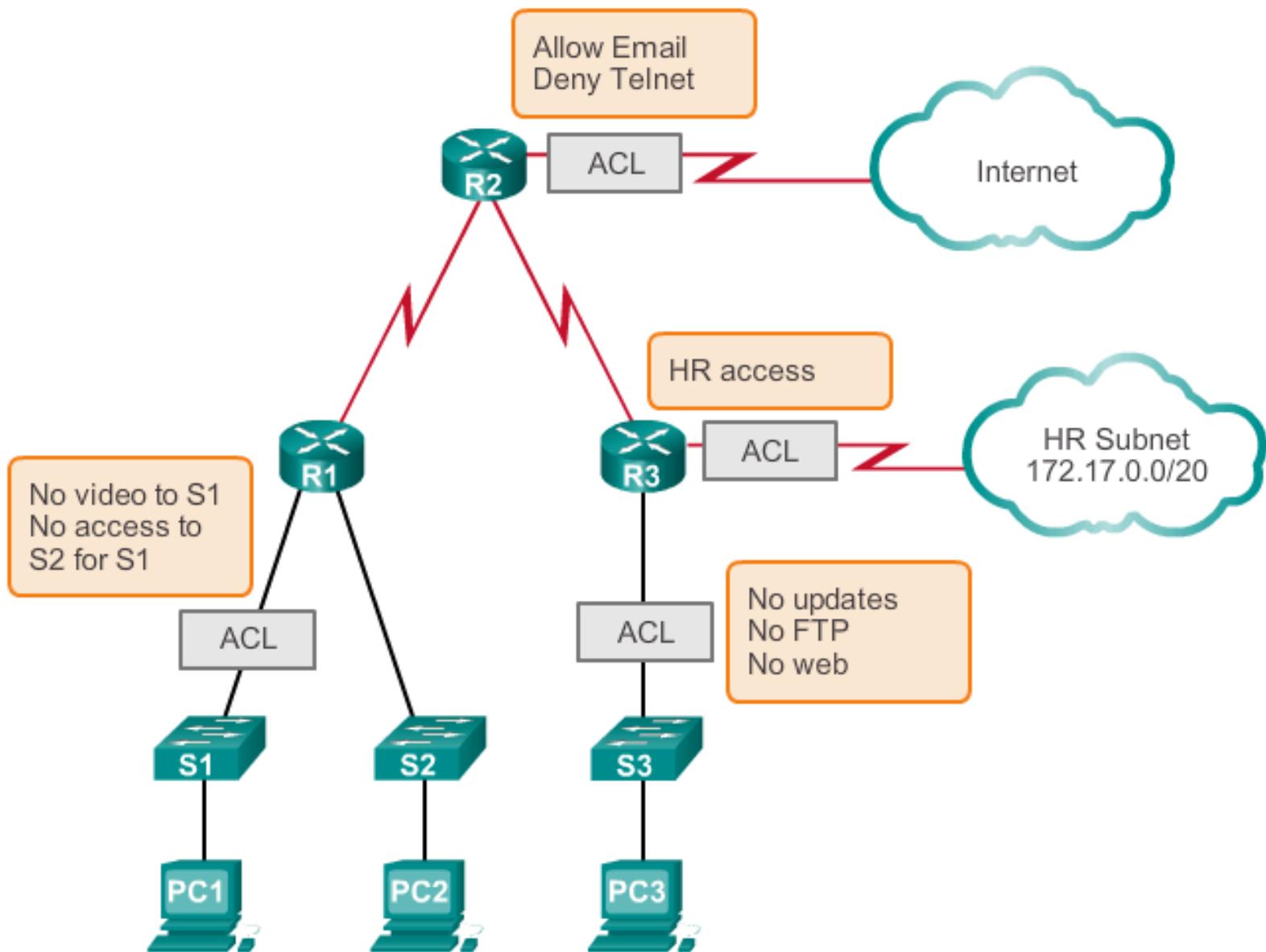
Add Static IP

Packet Filtering



Interface	Source IP	Source port	Destination IP	Destination port
1	131.34.0.0	*	*	*
1	*	*	*	23
1	*	*	194.78.20.8	*
2	*	*	*	80

Access Control List (ACL)



Cisco ACL Implementation

Standard ACLs

```
access-list 10 permit 192.168.30.0 0.0.0.255
```

- Range (1-99) and (1300-1999)
- Based on source addresses only

Extended ACLs

```
access-list 103 permit tcp 192.168.30.0 0.0.0.255 any eq 80
access-list 103 deny ip any 192.168.0.0 0.0.255.255
access-list 103 permit any any
```

- Range (100-199) and (2000 -2699)
- Based on addresses, ports, protocol types.

Iptables (Netfilter Project)

- Command line utility for Linux firewall in kernel.

TABLE 1

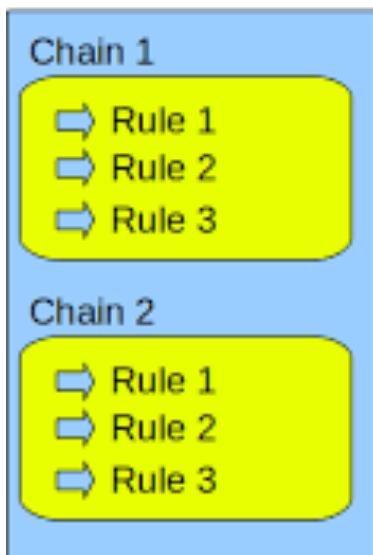
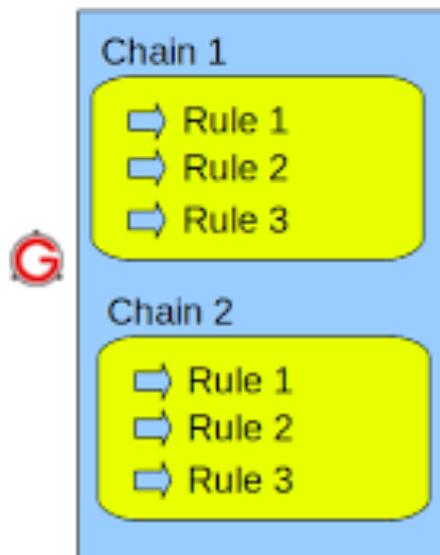


TABLE 2

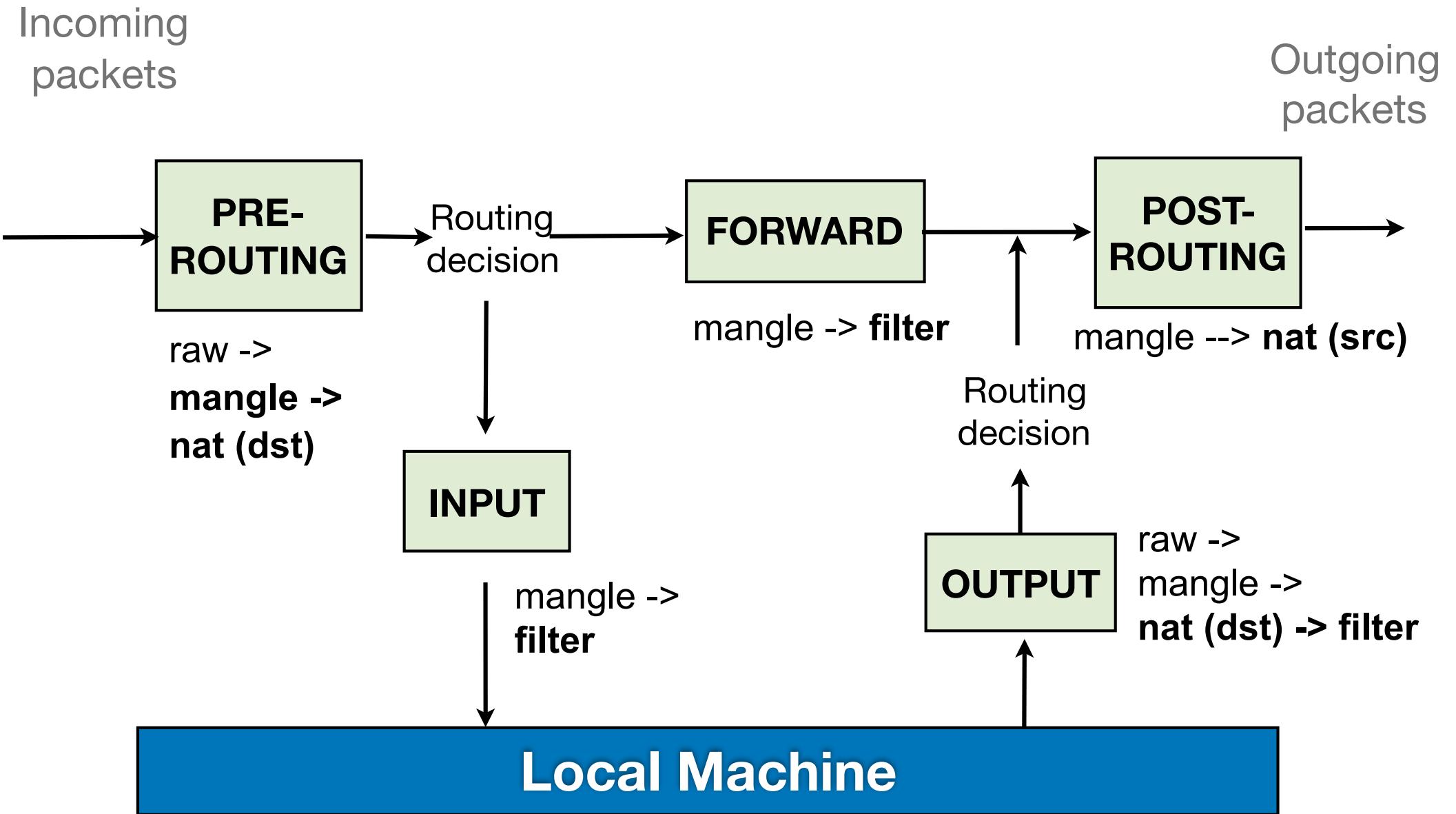


Rule = Condition + Action to apply

Source IP	ACCEPT
Dest IP	DROP
Source port	DNAT
Dest port	SNET
Incoming intf	
Outgoing intf	

Tables: raw, mangle, filter, nat

Chains: PREROUTING, INPUT, OUTPUT, FORWARD, POSTROUTING



```
iptables -A INPUT -i eth1 -s 10.0.0.0/8 -j DROP
```

Add rule to **INPUT** chain in **filter** table

Drop packet with source ip address 10.0.0.0/8 coming at eth1

```
iptables -A INPUT -i eth1 -d 10.0.0.0/8 -j DROP
```

Add rule to **INPUT** chain in **filter** table

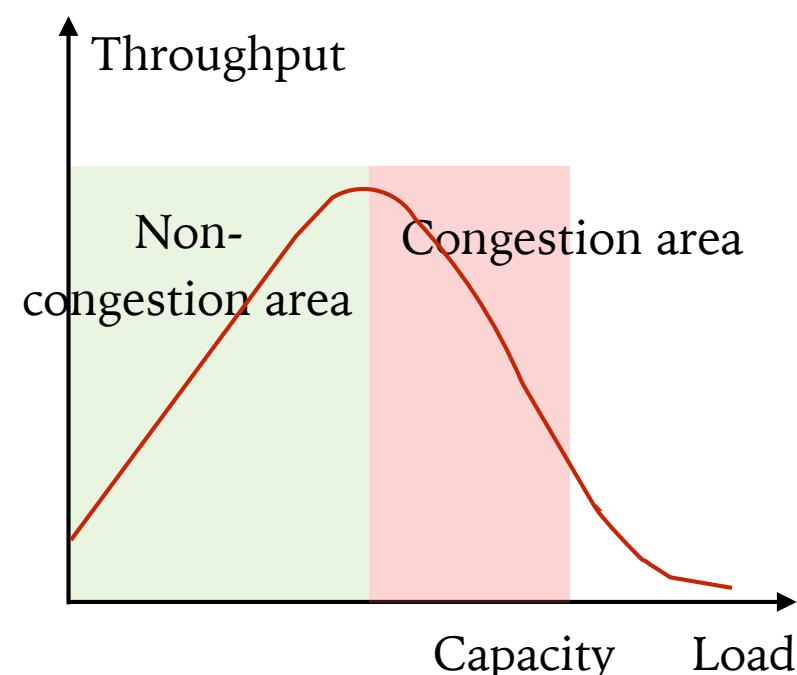
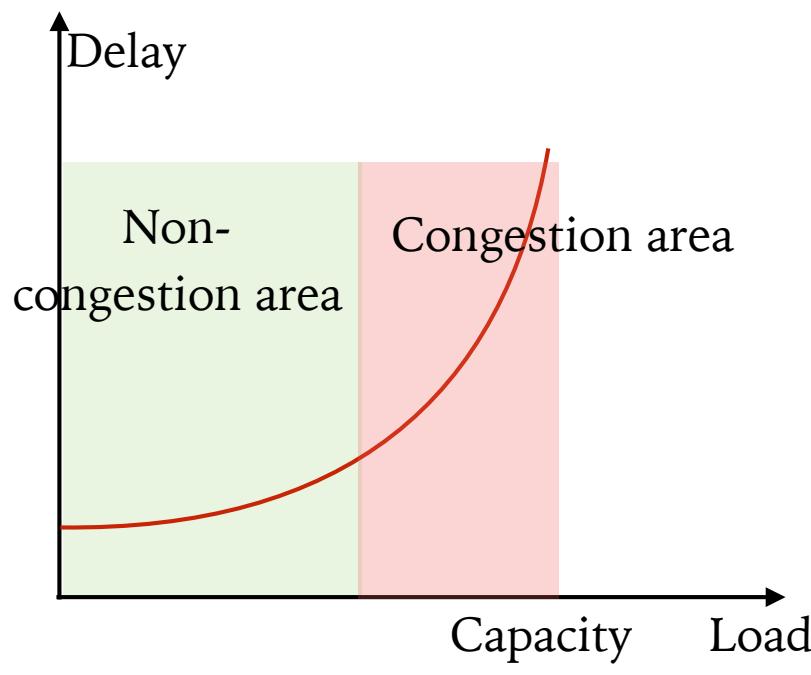
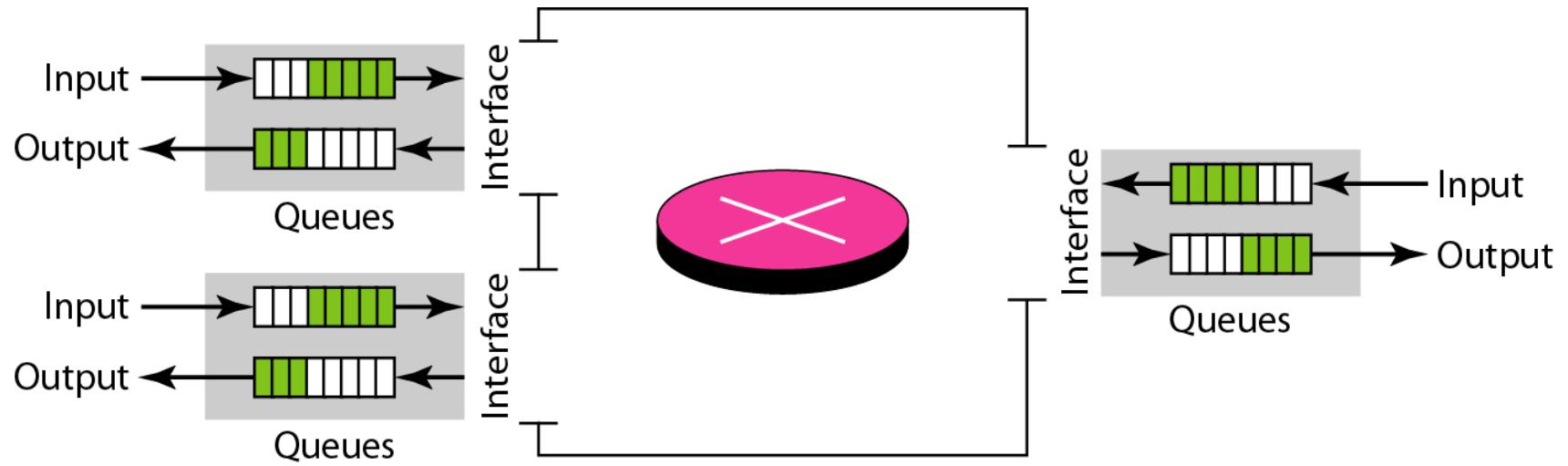
Drop packet with dest ip address 10.0.0.0/8 coming at eth1

```
iptables -t nat -A POSTROUTING -j MASQUERADE -s 10.0.1.0/24
```

Add rule to **POSTROUTING** chain in **nat** table

NAT source ip address 10.0.0.1/24 to that of the outgoing interface.

Network Congestion



TCP Congestion Control RFC 2581

- Sender limits transmission by maintaining that:

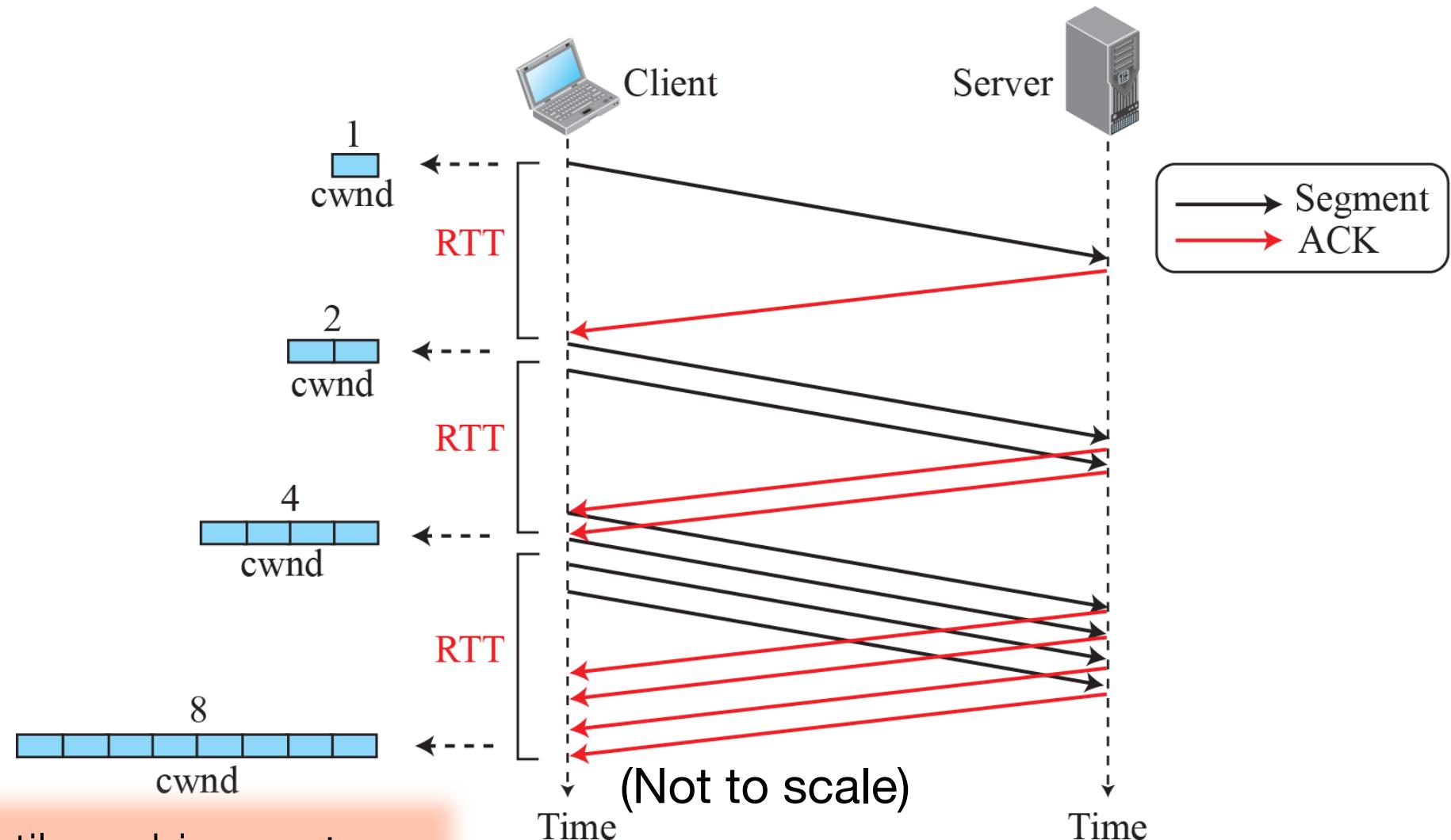
$$\text{LastByteSent} - \text{LastByteAcked} \leq \min(\text{cwnd} * \text{MSS}, \text{rwnd})$$

- **MSS** : maximum segment size (in bytes)
- **cwnd** (Congestion Window) : Dynamic fn. of perceived network congestion (unit MSS).

- Assuming a very large receiver buffer,

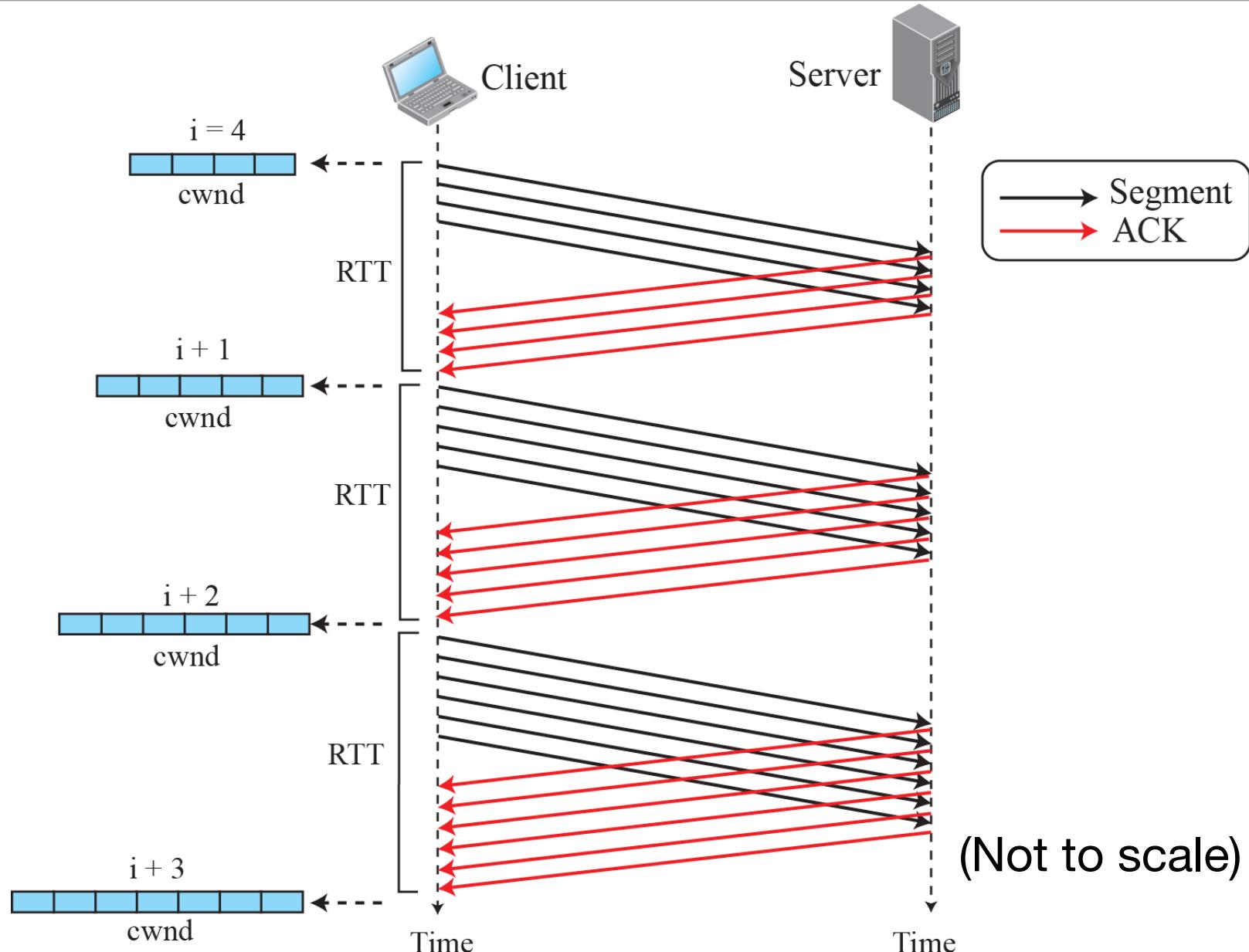
Avg. Rate ❤	$\frac{\text{Avg cwnd} * \text{MSS}}{\text{RTT}}$	Bytes/sec
--------------------	---	-----------

TCP Slow Start (Bandwidth Probing)

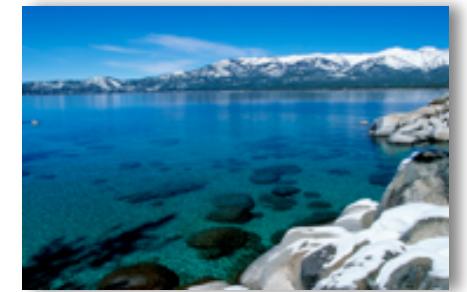


Until reaching ssthresh
(Slow start threshold)

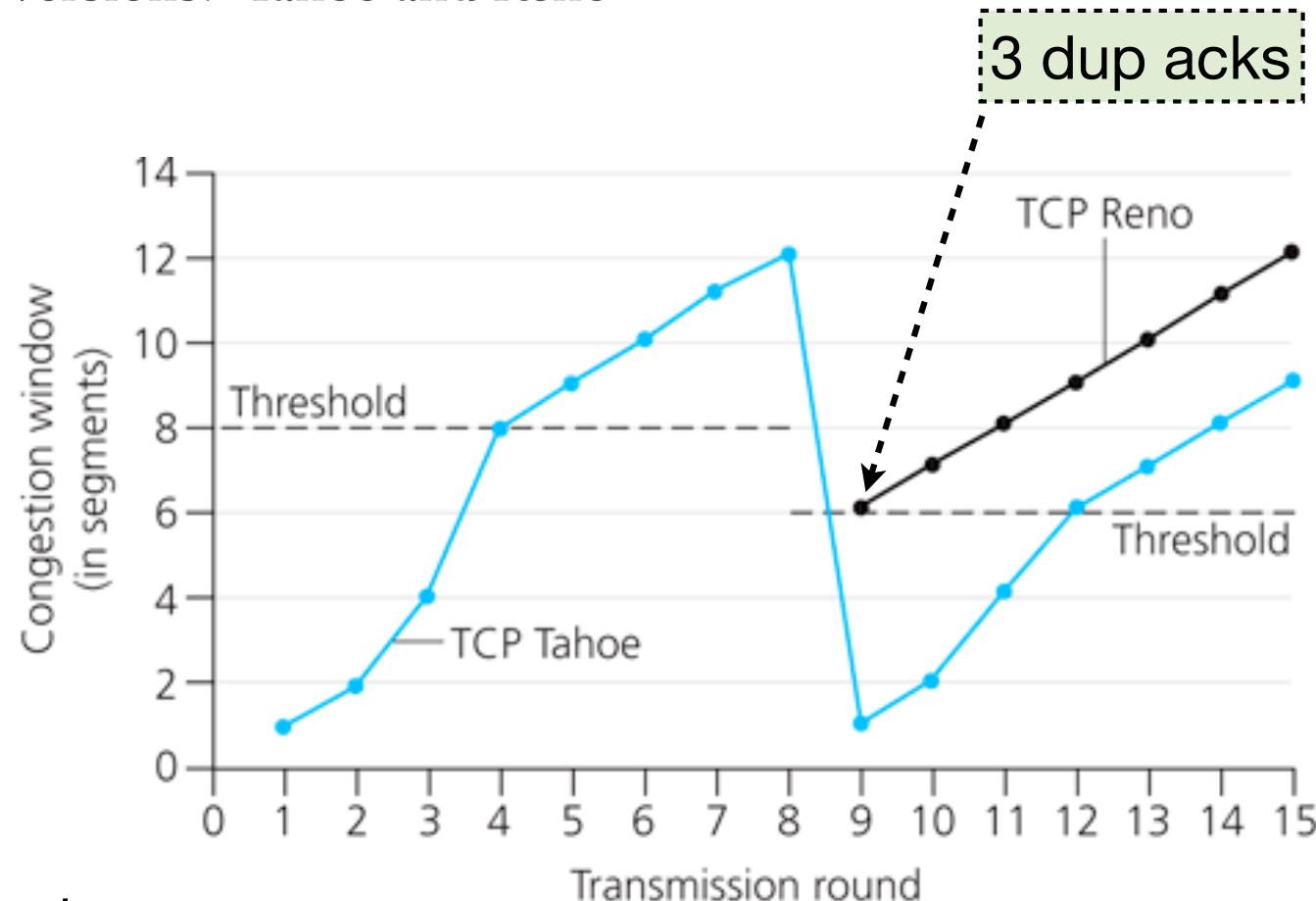
Congestion Avoidance : Additive Increase (AI)



Congestion Avoidance: Multiplicative Decrease (MD)



- Decrease cwnd when a “loss event” occurs.
 - Two versions: Tahoe and Reno

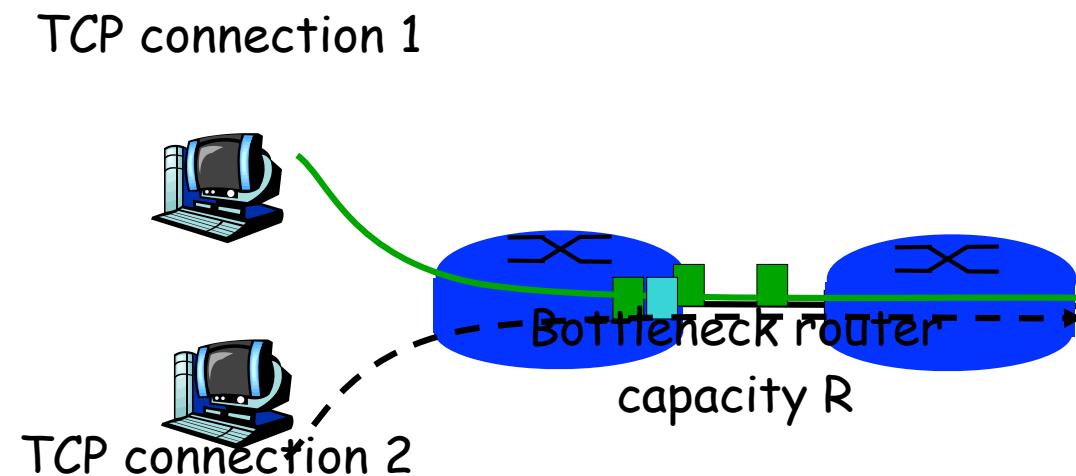


Kurose text

Figure 3.53 ♦ Evolution of TCP’s congestion window (Tahoe and Reno)

TCP Fairness

- **Fairness goal:** if K TCP sessions share the same bottleneck link of bandwidth R , each should get an average rate of R/K .



Why is TCP fair?

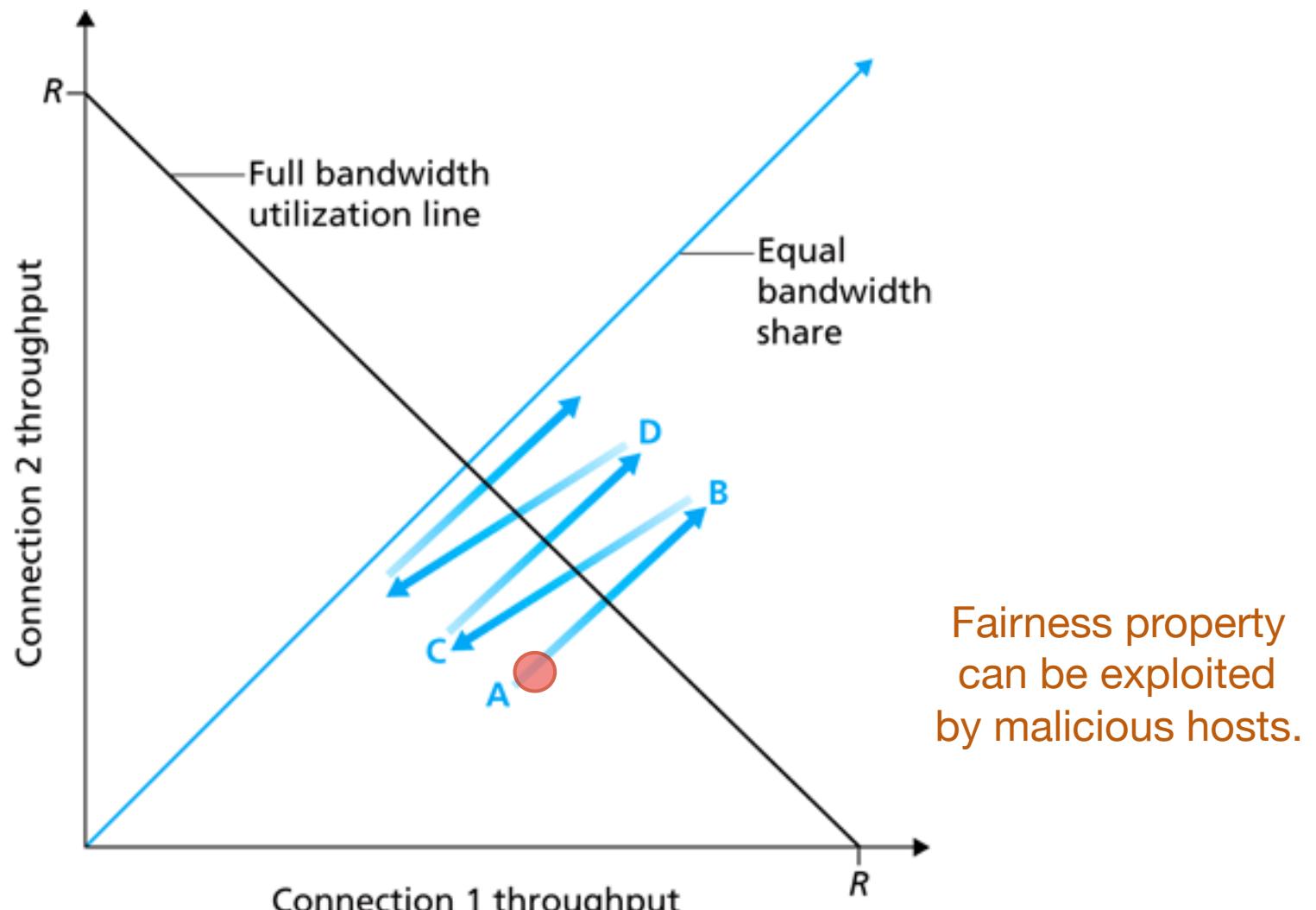


Figure 3.55 ♦ Throughput realized by TCP connections 1 and 2

Highly Simplified TCP throughput

- Macroscopic observation
 - Long-run approximate behavior
 - Ignore slow start (very short)

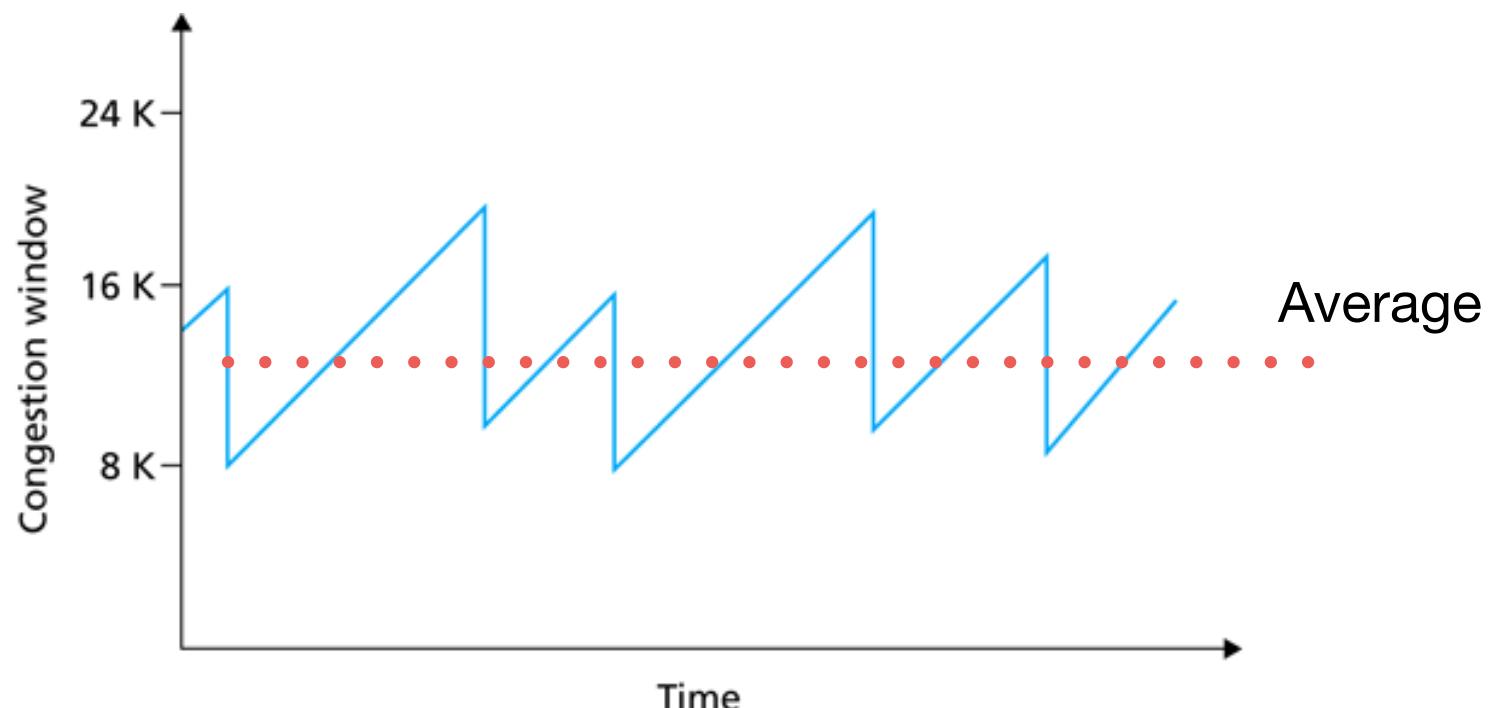


Figure 3.51 ♦ Additive-increase, multiplicative-decrease congestion control

GOOGLE ประกาศใช้ TCP BBR ในระบบเครือข่ายของ GCP เพิ่ม throughput สูงสุด 2,700 เท่า

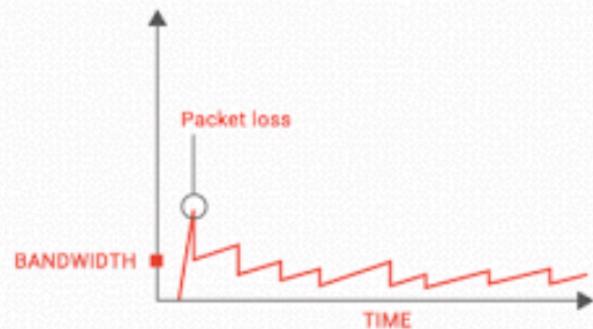
⌚ July 21, 2017

📁 Cloud and Systems, Cloud Services, Google, IT Knowledge, IT Researches, Networking, Products, Switch and Router

Google ออกมาประกาศถึงการนำ TCP BBR ซึ่งเป็น Congestion Control Algorithm ใหม่สำหรับระบบเครือข่ายมาใช้งานภายใน Google Cloud Platform (GCP) ซึ่งช่วยให้ประสิทธิภาพการทำงานของระบบในบางแห่งมุ่งสูงขึ้นถึง 2,700 เท่าเลยทีเดียวเมื่อเทียบกับ Algorithm แบบก่อนๆ

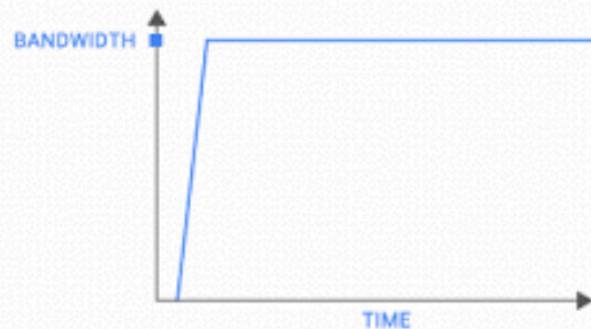
TCP before BBR

Today's Internet is not moving data as well as it should. TCP sends data at lower bandwidth because the 1980s-era algorithm assumes that packet loss means network congestion.



TCP BBR

BBR models the network to send as fast as the available bandwidth and is 2700x faster than previous TCPs on a 10Gb, 100ms link with 1% loss. BBR powers google.com, youtube.com, and apps using Google Cloud Platform services.



<https://cloudplatform.googleblog.com/2017/07/TCP-BBR-congestion-control-comes-to-GCP-your-Internet-just-got-faster.html>

Summary

- Internet Protocol (IP) for (hierarchical) addressing and forwarding in the network layer.
- Supplementary protocols
 - DHCP for address allocation
 - ICMP for error reporting and query
- Network address translation
 - Share public IP address to many private hosts.
 - Often implemented with port forwarding and access control list
- Network congestion control using TCP
 - Prevent congestion collapse
 - Also result in fairness among flows.

Midterm Exam

- Closed-book, calculator NOT allowed
- Multipart exam
 - Questions for short answers.
 - T/F questions
 - Multiple-choice questions
 - Questions for detailed solutions.