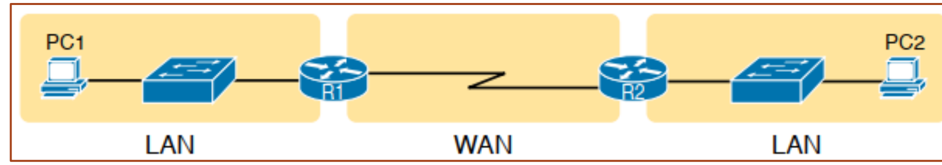
523353 – Computer Networks

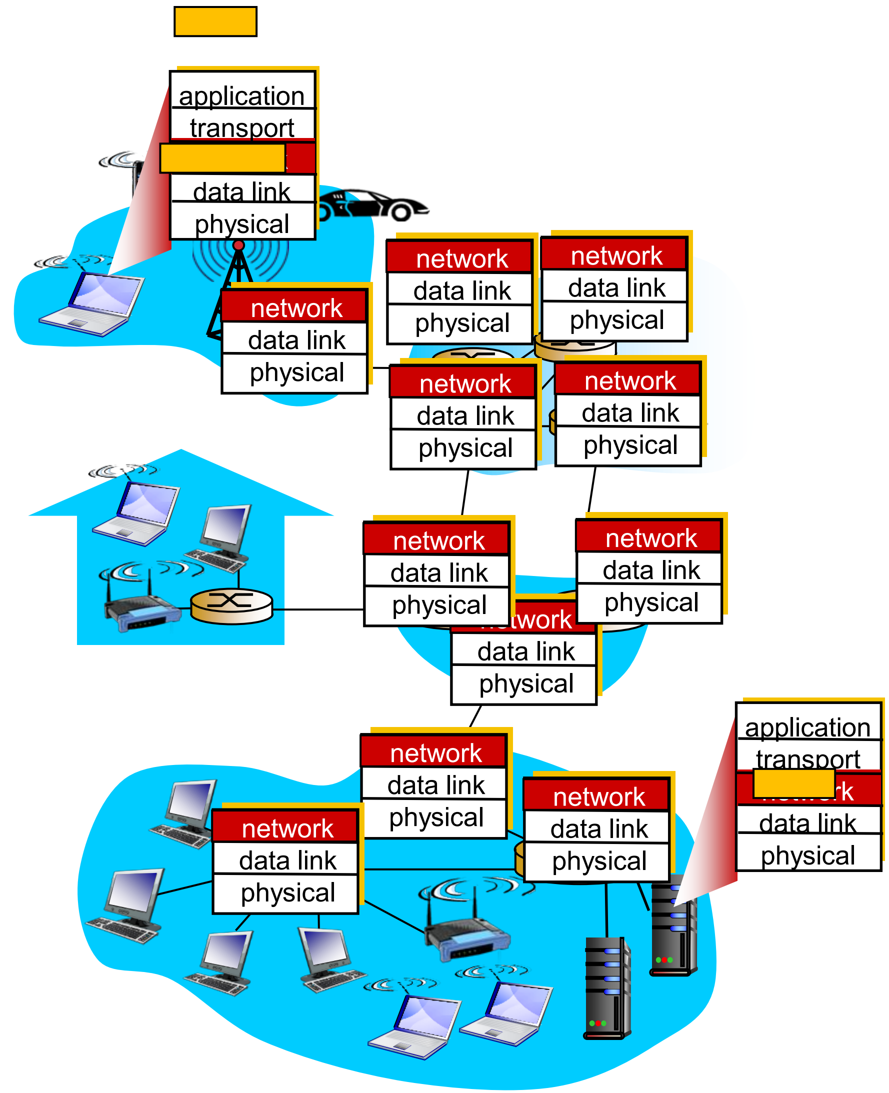
Lecture 5: Network Layer Part1

Dr. Parin Sornlertlamvanich
CompTIA Security+
Huawei Certified HCIA-Datacom
parin.s@sut.ac.th

Network layer



- The internetwork uses a router connected to each LAN, with a WAN (Wire-Area Networks) link between the routers
- Transport segment from sending to receiving host
  - On sending side encapsulates segments into datagrams
  - On receiving side, delivers segments to transport layer
- Network layer protocols in *every* Host, Router
  - Router examines header fields in all IP datagrams passing through it



Two key network-layer functions

Network-layer functions:

- *forwarding:* move packets from router's input to appropriate router output

- *routing:* determine route taken by packets from source to destination
  - *routing algorithms*

Analogy: taking a trip

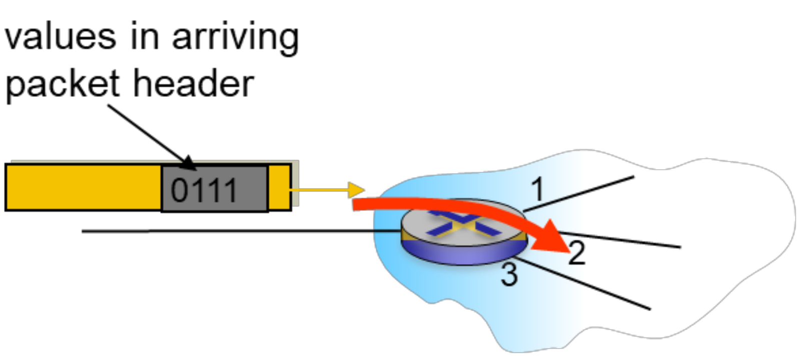
- *forwarding:* process of getting through single interchange

- *routing:* process of planning trip from source to destination

Network layer: data plane, control plane

Data plane

- Local, per-router function
- Determines how datagram arriving on router input port is forwarded to router output port
- Forwarding function

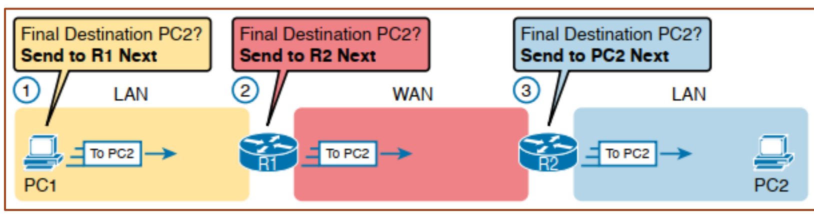


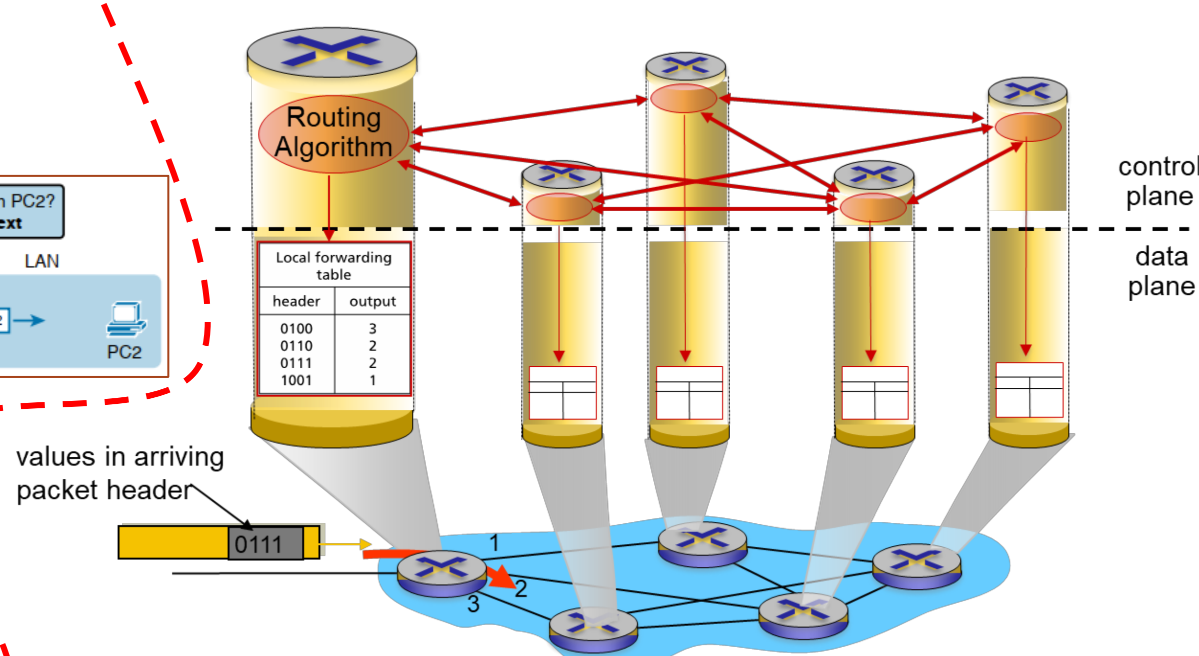
Control plane

- Network-wide logic
- Determines how datagram is routed among routers along end-end path from source host to destination host
- Two control-plane approaches:
  - *Traditional routing algorithms:* implemented in Routers
  - *Software-defined networking (SDN):* implemented in (remote) Servers
    - Controller

Per-router control plane

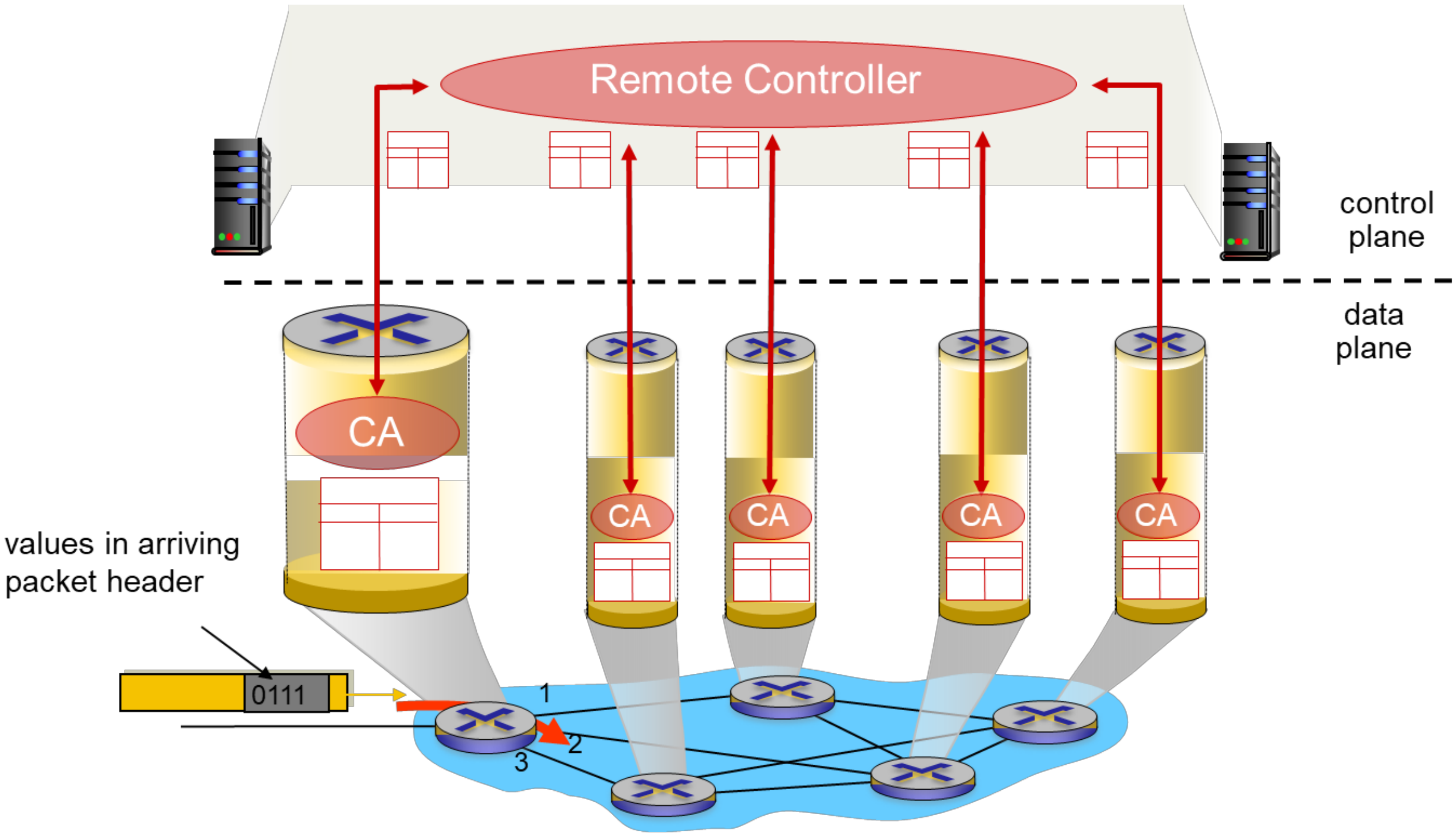
- Individual routing algorithm components in each and every router interact in the control plane





Logically centralized control plane

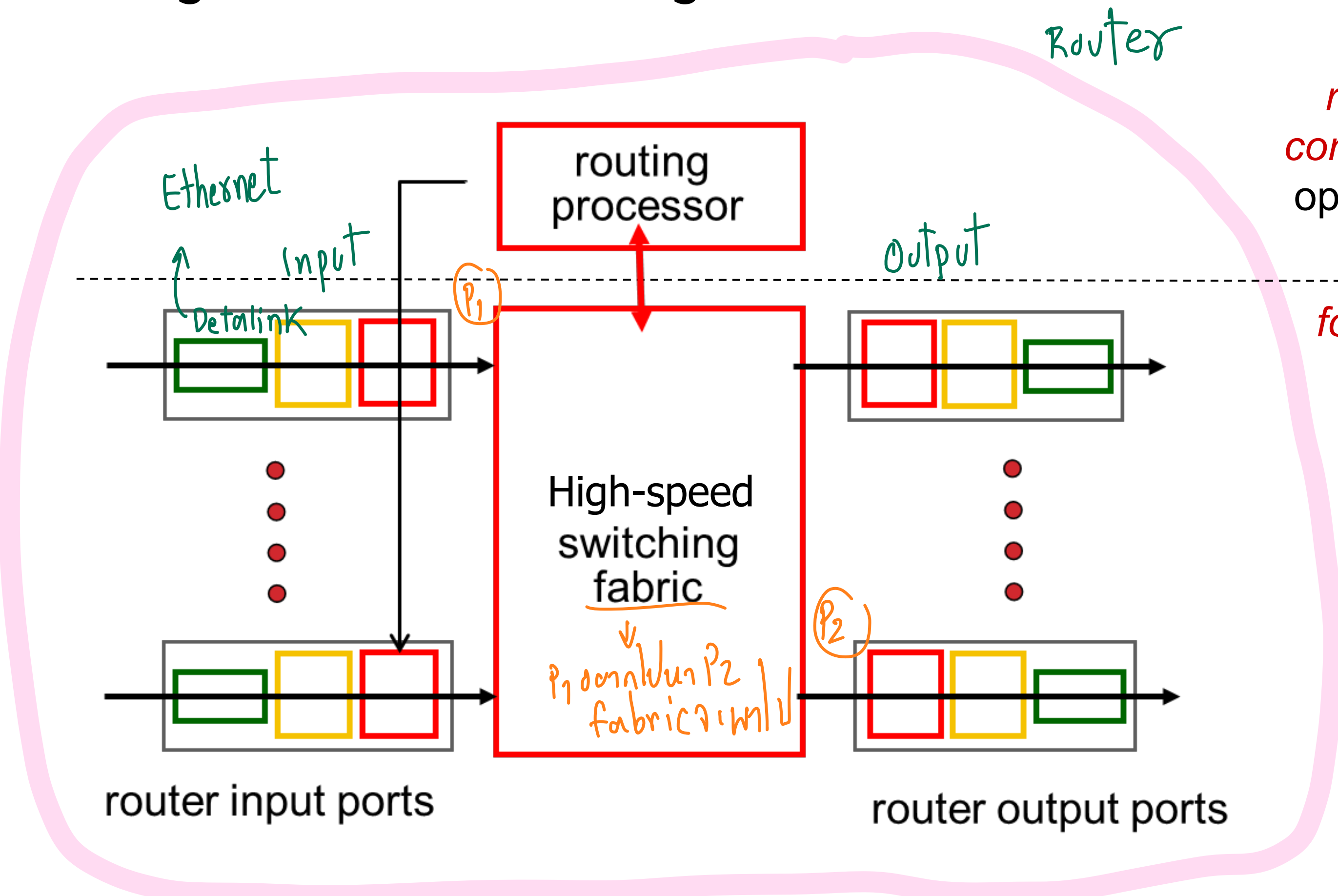
- A distinct (typically remote) controller interacts with local Control Agents (CAs)



Router

Router architecture overview

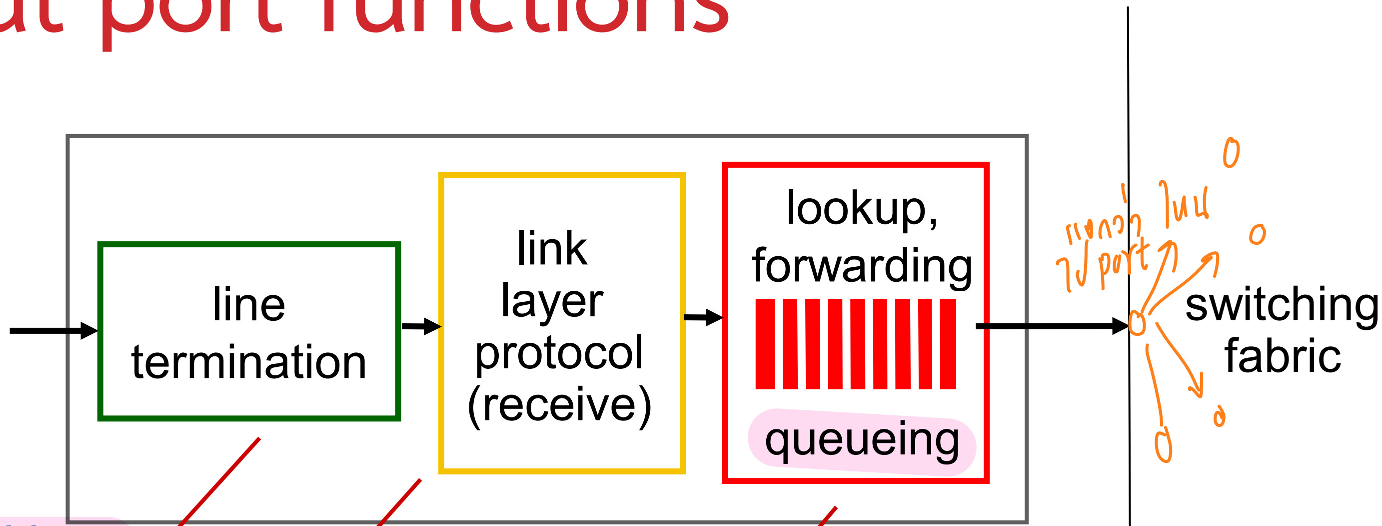
- High-level view of generic router architecture:



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

forwarding data plane
(hardware) operates
in nanosecond
timeframe

Input port functions



physical layer:
bit-level reception

data link layer:
e.g., Ethernet

(CSMA/CD
(เรื่องของการชน)

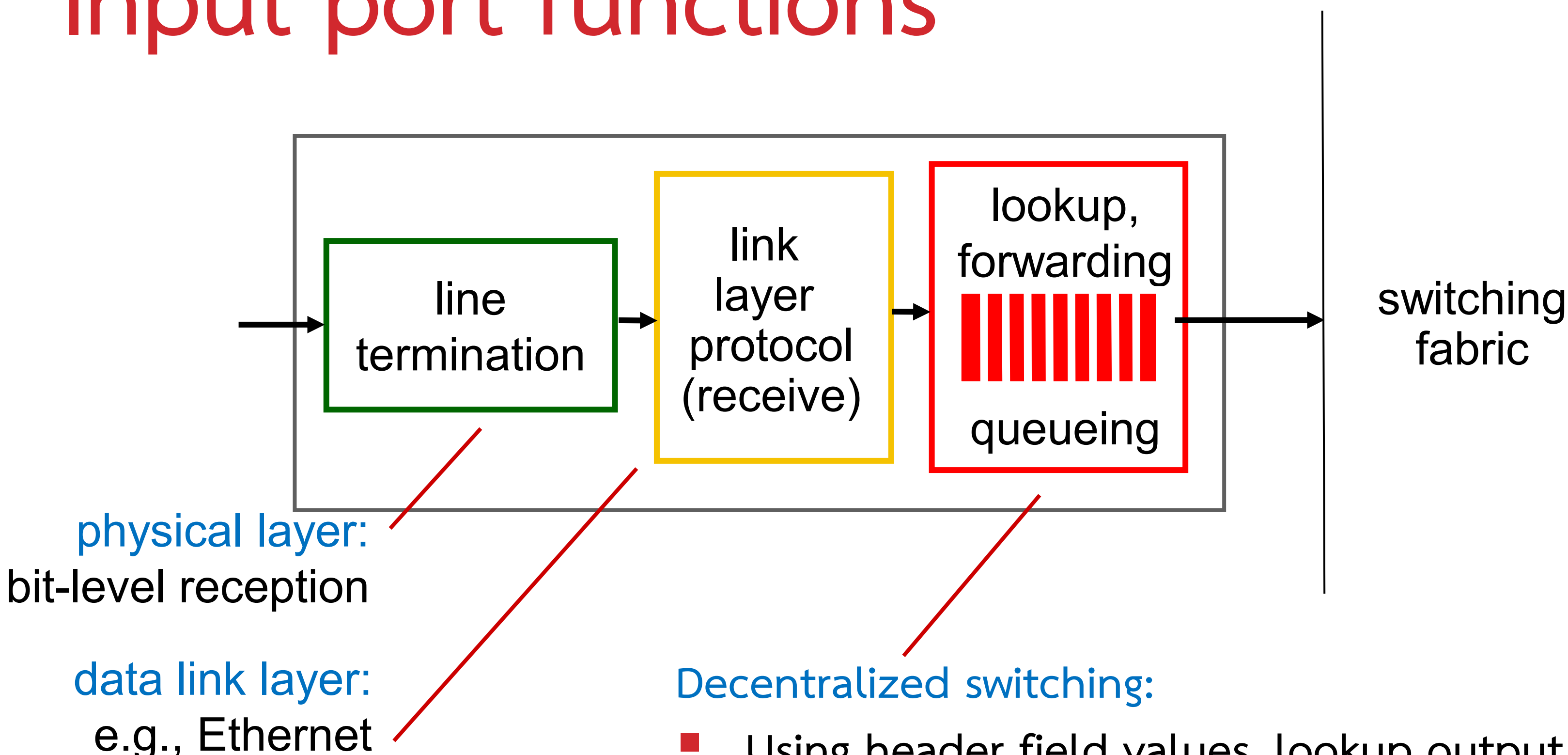
ของ router จ้า

Decentralized switching:

ดู IP Address

- Using header field values, lookup output port using forwarding table in input port memory ("match plus action") ดูว่าปลายทางจะไปออก port ไหน
- Goal: complete input port processing at 'line speed'
  ข้อมูลที่เข้ามาเร็วกว่าตัวด้านใน (มันแจกไม่ทัน)
- Queuing: if datagrams arrive faster than forwarding rate into switch fabric ก็เลยเอาข้อมูลมาเก็บไว้ในนี้

Input port functions



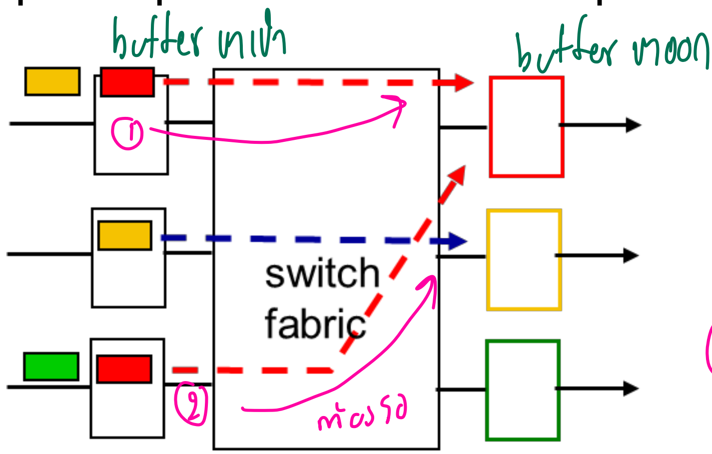
Decentralized switching:

- Using header field values, lookup output port using forwarding table in input port memory ("match plus action")
  ดู IP ปลายทางเป็นหลัก
- Destination-based forwarding: forward based only on destination IP address (traditional)
- Generalized forwarding: forward based on any set of header field values

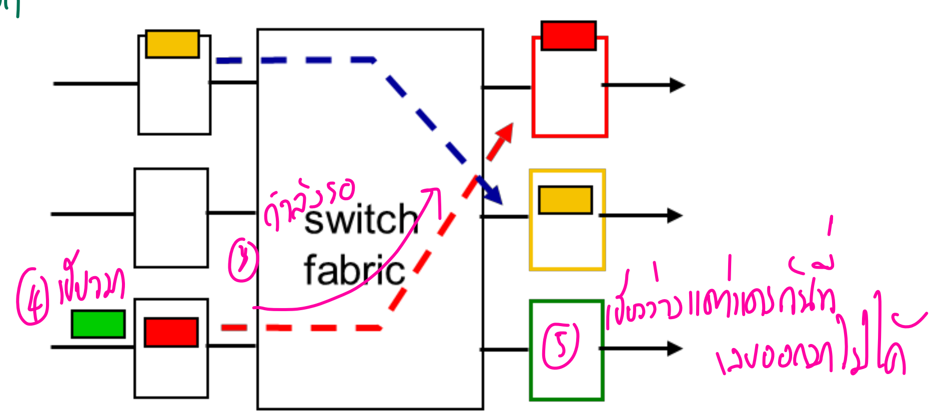
Input port queuing

- Fabric slower than input ports combined -> queueing may occur at input queues ถ้า Q เต็มเกิด (ดูไลฟ์แล้วภาพหาย, แลค)
  - 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

คือตัวเองไปต่อไม่ได้แล้วไปขวางทางคนที่ทำได้



output port contention:
only one red datagram can be transferred.
lower red packet is blocked

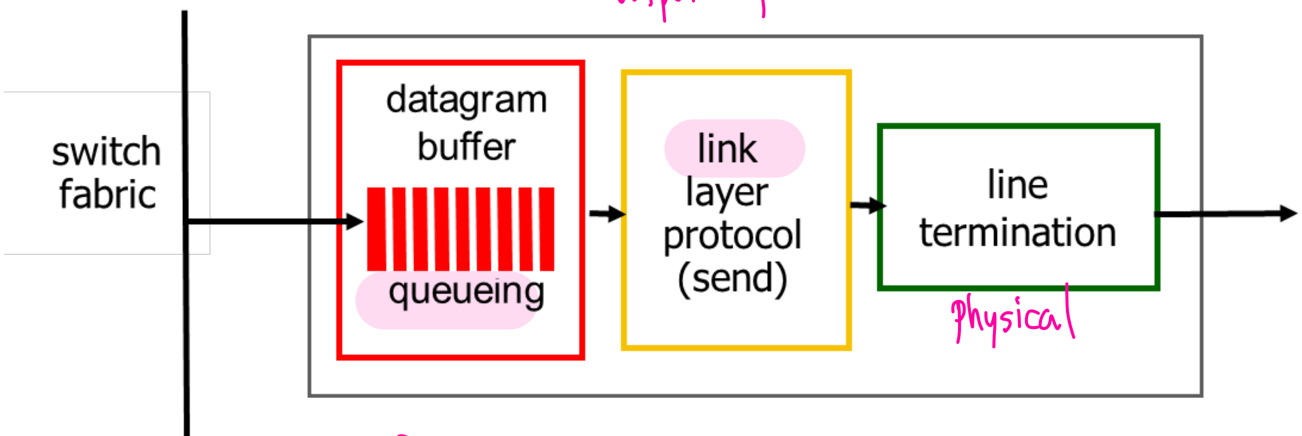


one packet time later:
green packet experiences HOL blocking

ทำให้เกิดปัญหา Delay
แม้ Q ไม่เต็ม

Output ports

output port



Buff จะเก็บเมื่อ

- **Buffering** required when datagrams arrive from fabric faster than the transmission rate
  เร็วกว่า ส่งไปข้างนอก

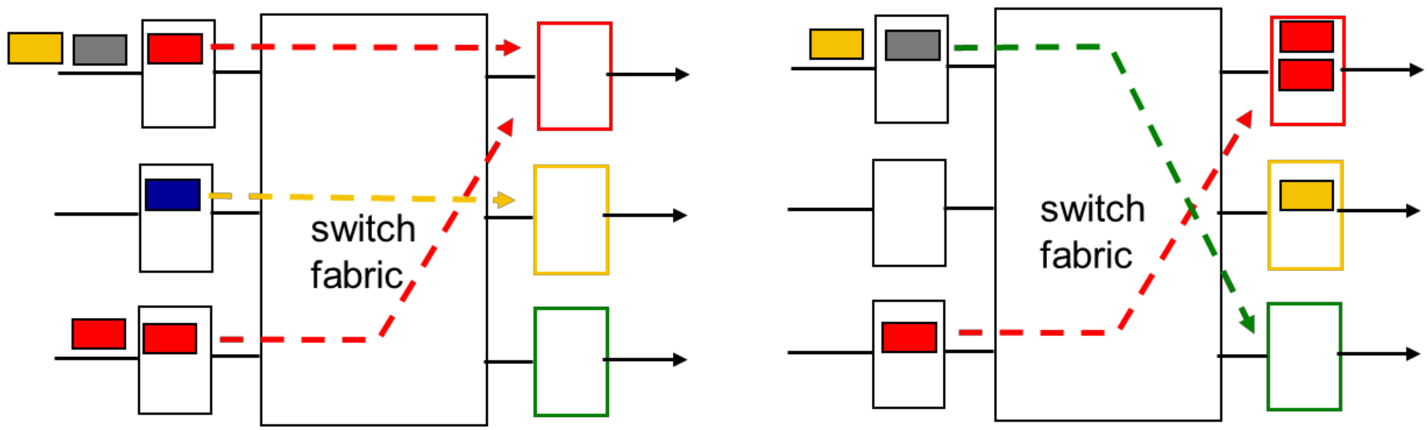
Datagram (packets) can be lost due to congestion, lack of buffers

ป้องกัน Q เต็มโดยเข้ามาจัดการ โดย

- **Scheduling** discipline chooses among queued datagrams for transmission

Priority scheduling – who gets best performance, network neutrality

Output port queueing



at *t*, packets more
from input to output

one packet time later

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

ล้นเกิน

คำนวณ buffering

How much buffering?

S P

มีต้นทาง-ปลายทาง

- RFC 3439: average buffering (B) equal to "typical" RTT (say 250 msec) times link capacity C ค.เร็วของลิงค์
  - e.g., C = 10 Gpbs link: 2.5 Gbit buffer
    - B = RTT x C
    - 2.5 Gbit = 10 Gpbs x 250 msec

- Recent recommendation: a large number of TCP flows (N) passing through a link, buffering equal to

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

Scheduling mechanisms

↓ วิธีจัดการ เมื่อ buffer เต็ม

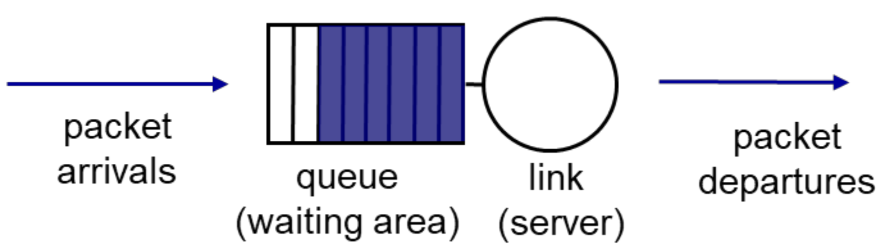
- **Scheduling:** choose next packet to send on link

  วิธีแก้ง่ายสุด (ก็เอาอันที่เข้ามาแรก ออกไป)

- **FIFO (first in first out) scheduling:** send in order of arrival to queue
  - Discard policy: if packet arrives to full queue: who to discard?

    อะไรที่ล่าสุดเข้ามา (หลังเต็ม) ก็ดรอปไปเลย

    - Tail drop: drop arriving packet
    - Priority: drop/remove on priority basis ดู Priority
    - Random: drop/remove randomly สุ่มอันออกเพื่อให้ อันที่มาเข้าได้



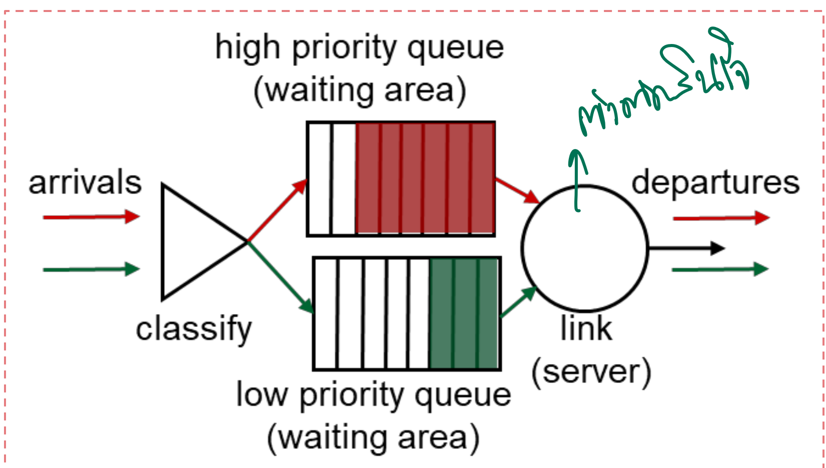
ถ้าไม่ได้มีแค่ FIFO เพราะมี

Scheduling: priority

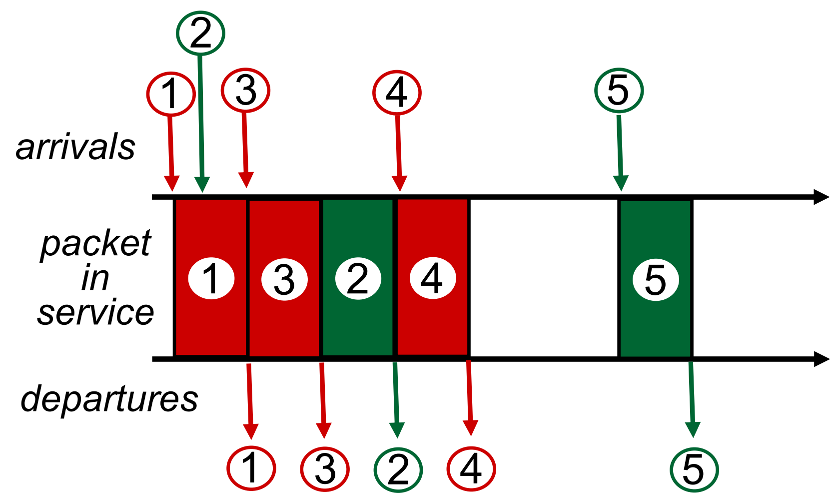
Not FIFO?

Priority scheduling: send highest priority queued packet

- multiple classes, with different priorities
  - class may depend on marking or other header info, e.g. IP source/dest, port numbers, etc.
  - real world example?



ต่างตรงนี้ใจ



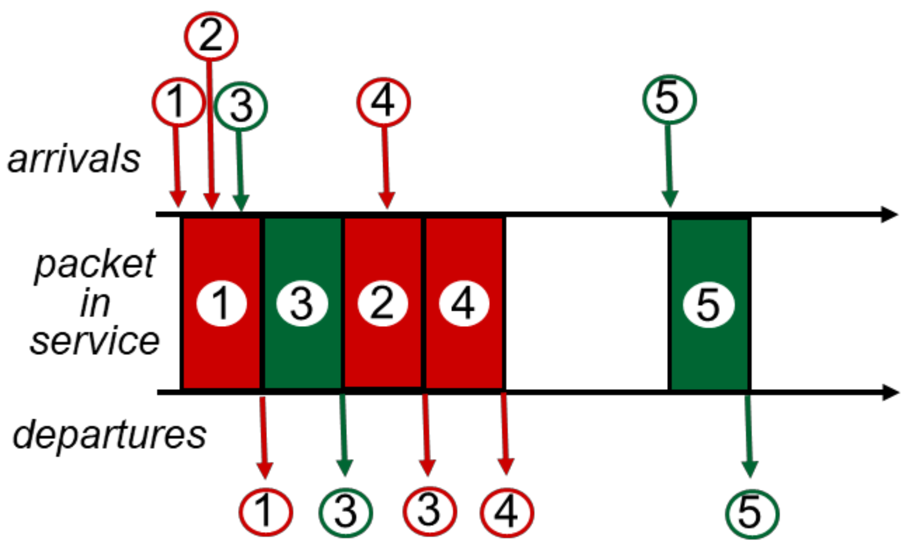
Scheduling: still more (1)

Piority ต่ำ ก็ถูกทำ ไม่ได้ถูกทิ้ง

Round Robin (RR) scheduling:

(ผลัดกันทำ ถ้า piority สูงๆ ส่งแล้วหมดมาต่ำ)

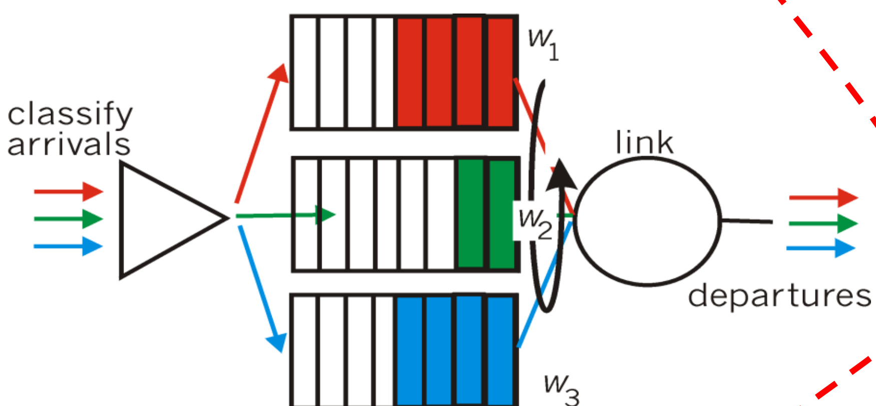
- Multiple classes
- Cyclically scan class queues, sending one complete packet from each class (if available)

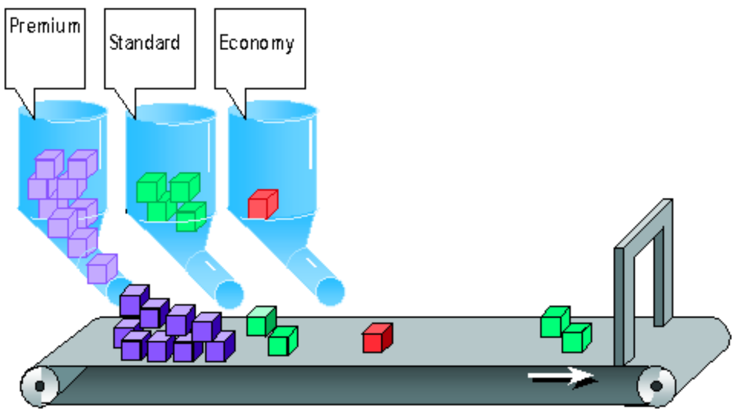


Scheduling: still more (2)

Weighted Fair Queuing (WFQ):

- Generalized Round Robin
- Each class gets weighted amount of service in each cycle





Destination-based forwarding

รู้ได้ไงว่า เราจะต้องไปไหน

↗ แระ Routing 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 00011000 00000000 through 11001000 00010111 00011111 11111111	2
otherwise	3

ล็อก 21 bits (ด้านหลังใส่ไรก็ได้)

200.23.16.0

24 bit

200.23.24.0

200.23.24.255

ล็อก 21 bits

200.23.24.0

11001000 00010111 00010000 00000000 = 200.23.16.0
11001000 00010111 00011000 00000000 = 200.23.24.0

Longest prefix matching

เทียบ prefix จน.เลขบิทเหมือนมากสุด

longest prefix matching

when looking for forwarding table entry for given destination address, use *longest* address prefix that matches destination address.

Prefix	Link interface
/21 ipaddress	
11001000 00010111 00010*** *********	0
11001000 00010111 00011000 *********	1
11001000 00010111 00011*** *********	2
Otherwise = ?	3

examples:

1. เอา Prefix มาเทียบอันไหนเหมือนสุด

DA: 11001000 00010111 00010110 10100001 which interface?

DA: 11001000 00010111 00011000 10101010 which interface?

DA: 11001000 00010111 00011001 10101010

เอาเหมือนกัน และเอาอันที่ยาวสุด

เทียบเหมือนกันเป๊ะ ๆ มากกว่า

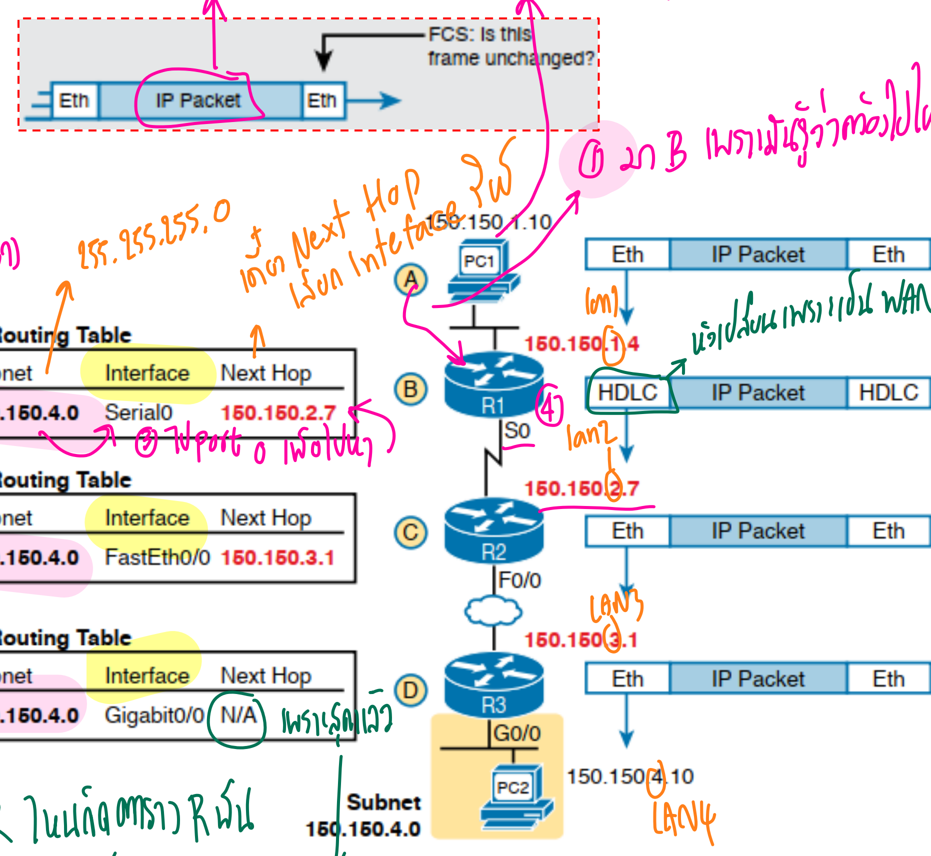
Sending IP datagrams (1)

- **Step1:** Use the data-link Frame Check Sequence (FCS) field to ensure that the frame had no errors; if errors occurred, discard the frame.

- **Step2:** To discard the old data-link header and trailer, leaving the IP packet.

- **Step3:** Compare the IP packet's destination IP address to the routing table, and find the route that best matches the destination address

- **Step4:** Encapsulate the IP packet inside a new data-link header and trailer, and forward the frame.



R1 Routing Table

Subnet	Interface	Next Hop
150.150.4.0	Serial0	150.150.2.7

R2 Routing Table

Subnet	Interface	Next Hop
150.150.4.0	FastEth0/0	150.150.3.1

R3 Routing Table

Subnet	Interface	Next Hop
150.150.4.0	Gigabit0/0	N/A

Sending IP datagrams (2)

กรณีมีทางแยก

3 จุดแรกตรงกัน แสดงว่า Subnet, VLAN เดียวกัน ยิงตรง, ไม่ไปต่างฝั่ง

Routing table

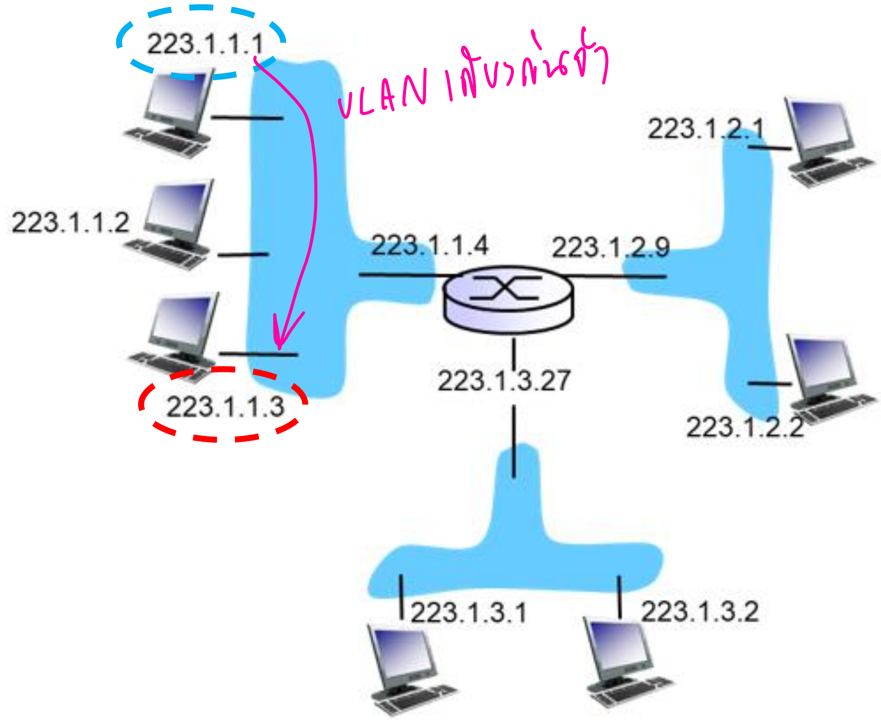
ต้นทาง

➤ Source: 223.1.1.1 /24

ปลายทาง

➤ Destination: 223.1.1.3 /24

- Look up network address
  - (D) is on same subnet
- Link Layer will send this datagram directly to (D) รู้แค่ 2 ตัวนี้ก็ยิงได้
  - In Ethernet frame
  - Using ARP table



Sending IP datagrams (3)

- Source: 223.1.1.1 /24
- Destination: 223.1.2.1 /24

อยู่คนละ subnet, ข้าม VLAN

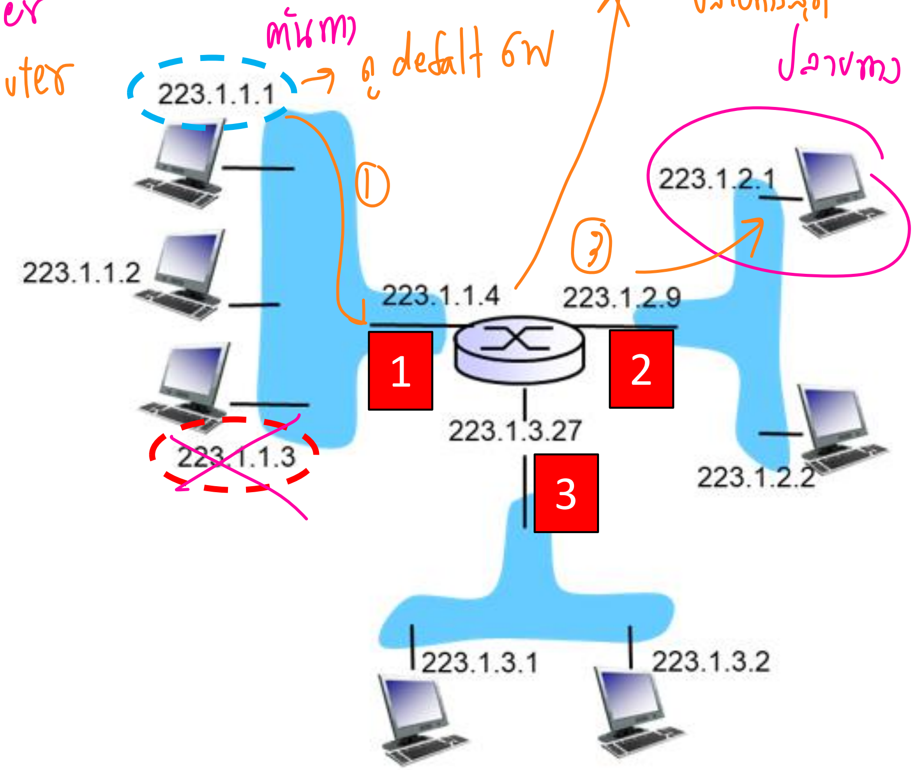
ต้องใช้ network layer

ต้องมี Router

Dest. Network	Next Router	Interface
223.1.1.0/24		1
223.1.2.0/24		2
223.1.3.0/24		3

ดูอันไหนเหมือน ปลายทางสุด

- Look up network address
  - (D) is on different network
- Router checks the routing table
  - (D) on subnet directly attached to Interface 2
- Link Layer will send this datagram directly to (D)
  - Interface 2 (223.1.2.9)



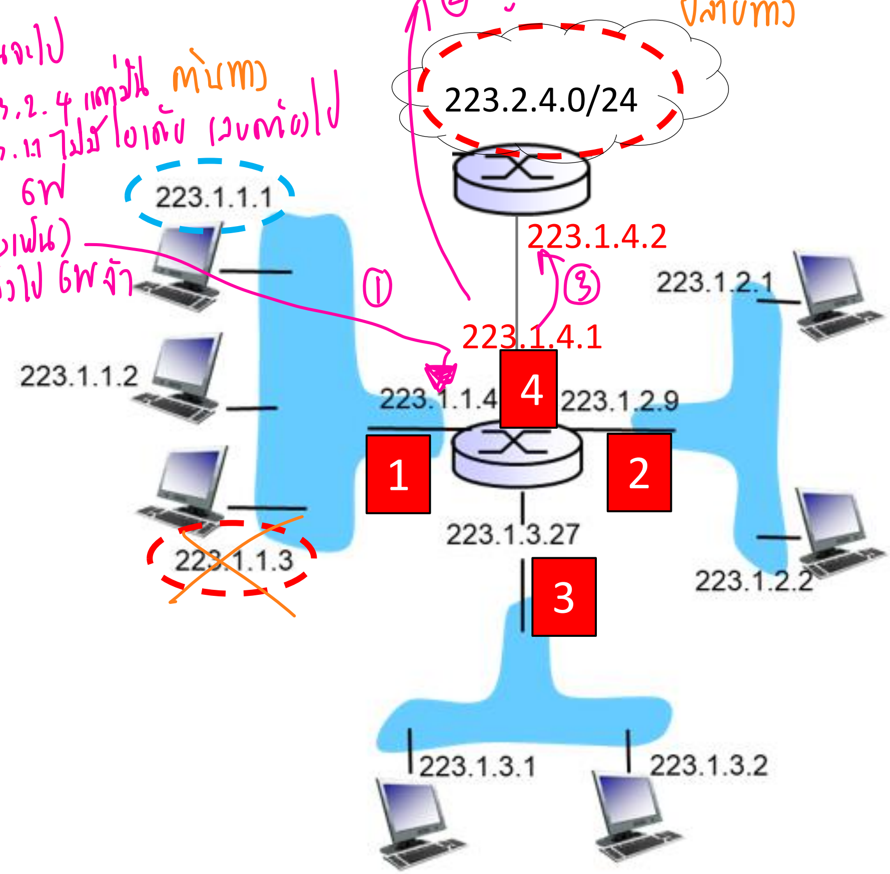
ฝั่ง Router

Sending IP datagrams (4)

Dest. Network	Next Router	Interface
223.1.1.0/24		1
223.1.2.0/24		2
223.1.3.0/24		3
223.1.4.0/24		4
223.2.4.0/24	223.1.4.2	4

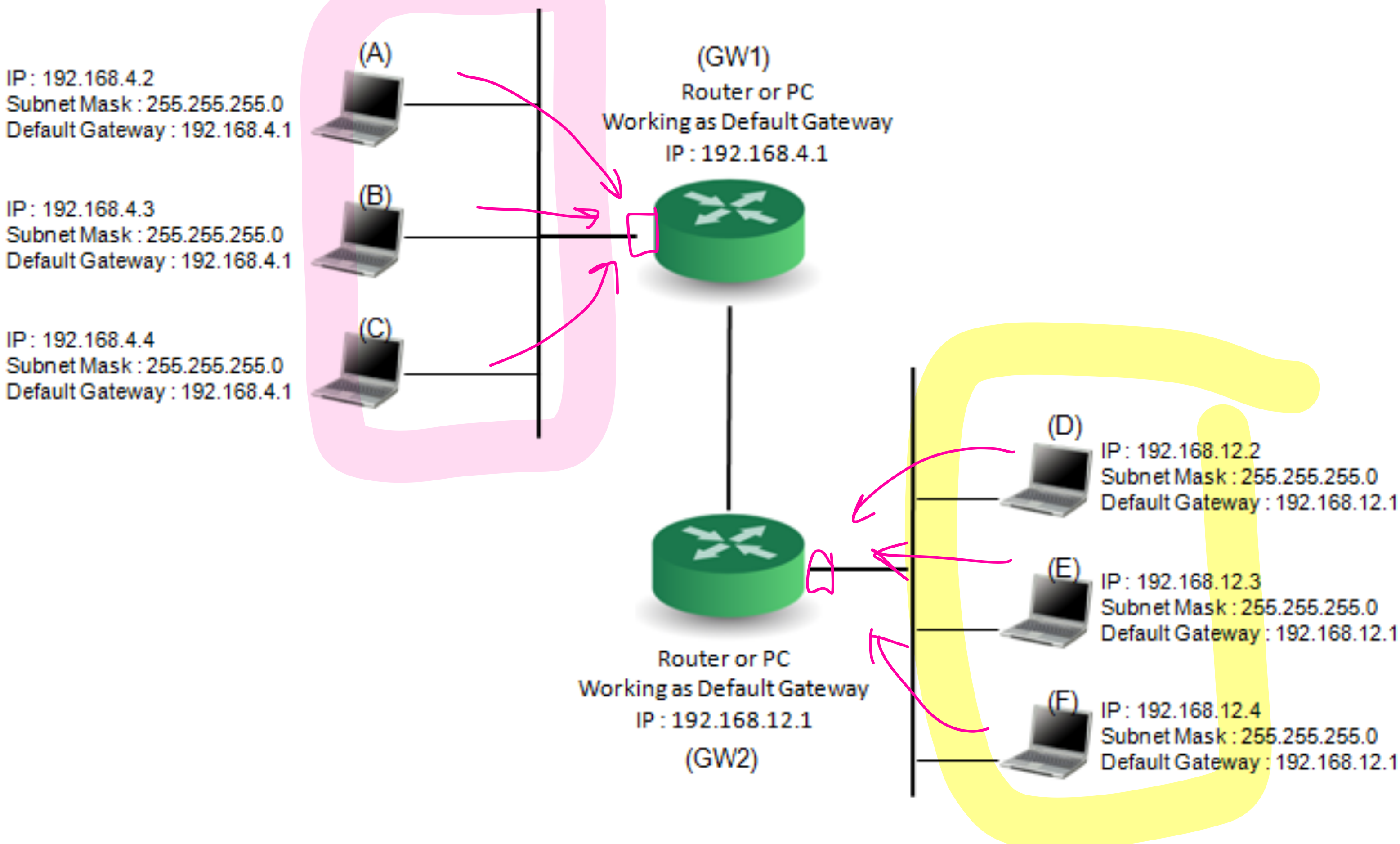
- ➢ Source: 223.1.1.1 /24
- ➢ Destination: 223.2.4.77 /24

- Look up network address
  - (D) is on different network
- Router checks the routing table
  - (D) on subnet directly attached to Interface 4
- Link Layer will send this datagram directly to (D)



Sending IP datagrams (5) – Default Gateway

- All the clients (PCs) on a network point to a default gateway that routes their traffic
  - It's generally the router interface address attached to the local network



Sending IP datagrams (6) – Default Route

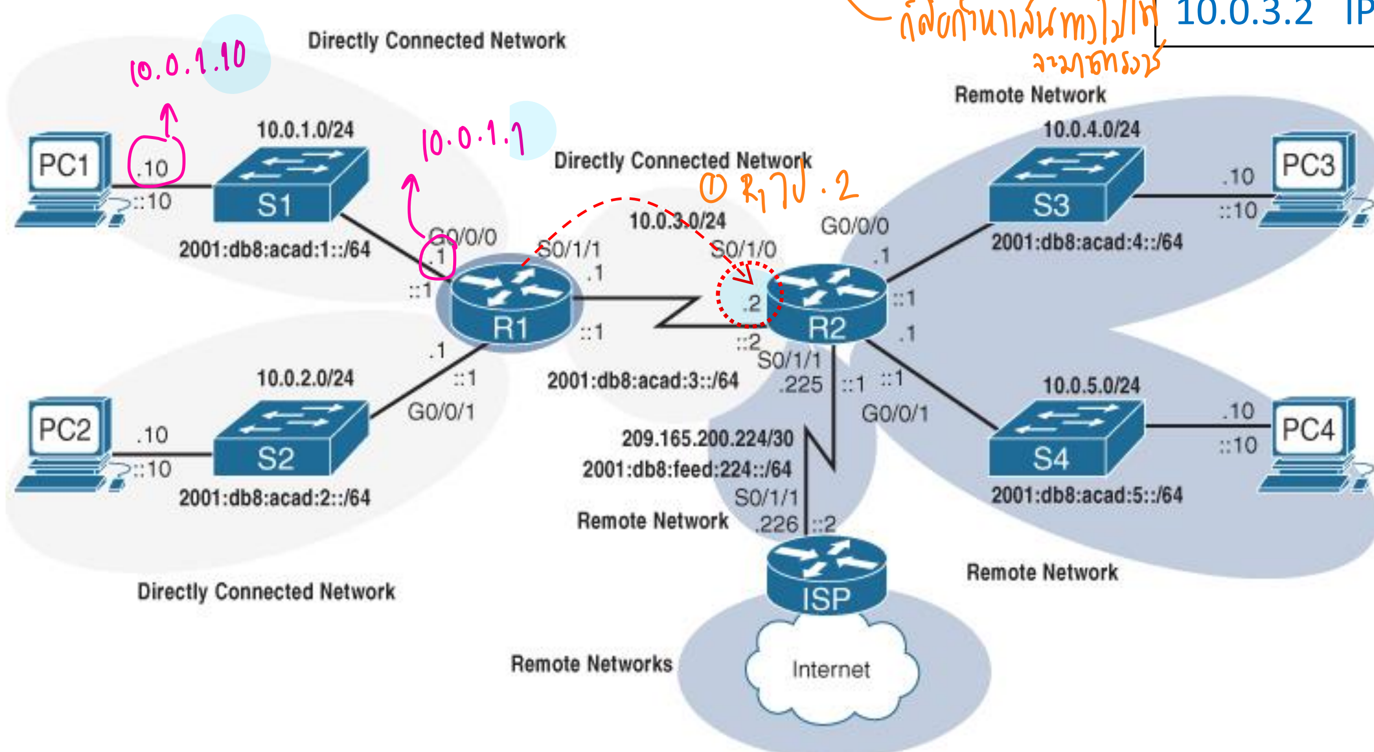
ไม่มีในเส้นทางไหน ปลายทางนั้น จะไปที่ ↗ (เหมือนทางแรกที่ไป Default GW น่ะ)

- For Router
  - There is no better (longer) match in the Routing table
  - Ex. R1 IPv4 routing table has a default route to forward all packets to R2 for any remote network for which it does not have a more explicit route
    - ② พิมพ์
    - R1 (config)# ip route 0.0.0.0 0.0.0.0 10.0.3.2
    - If its's not local, send it to 10.0.3.2 (R2) ③ 10.0.3.2

ปลายทาง

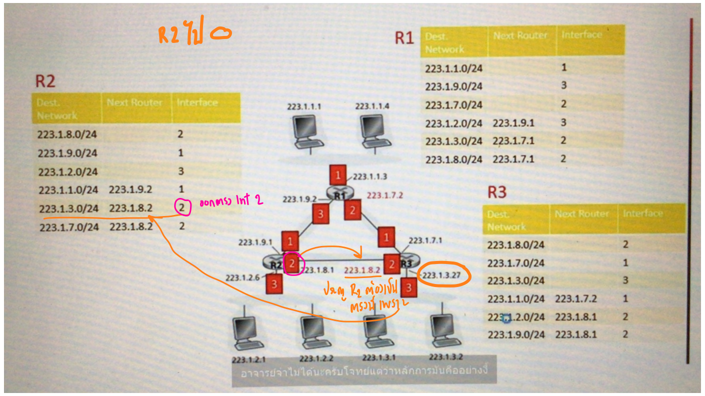
0.0.0.0	Network Address
0.0.0.0	Subnet Mask
10.0.3.2	IP Address

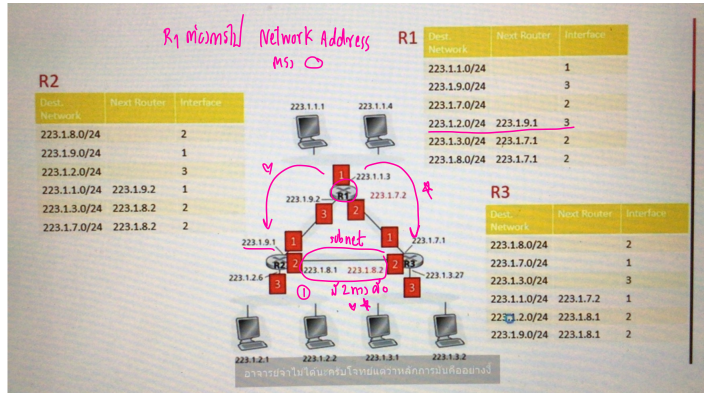
ก็ส่งต่อมาเส้นทางไปที่ จะมาถึงตรงนี้



IP (Internet Protocol)

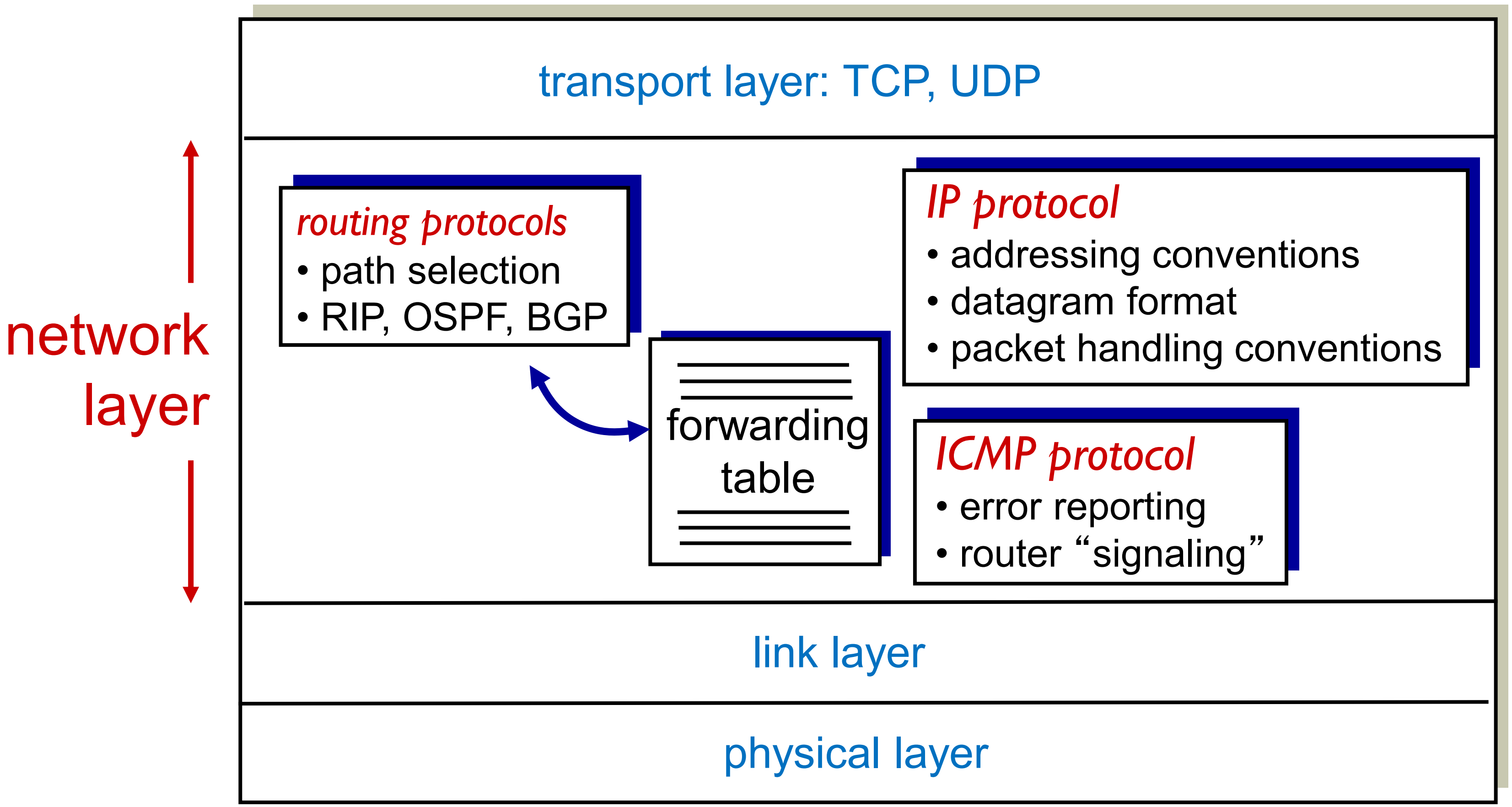
ทบทวน



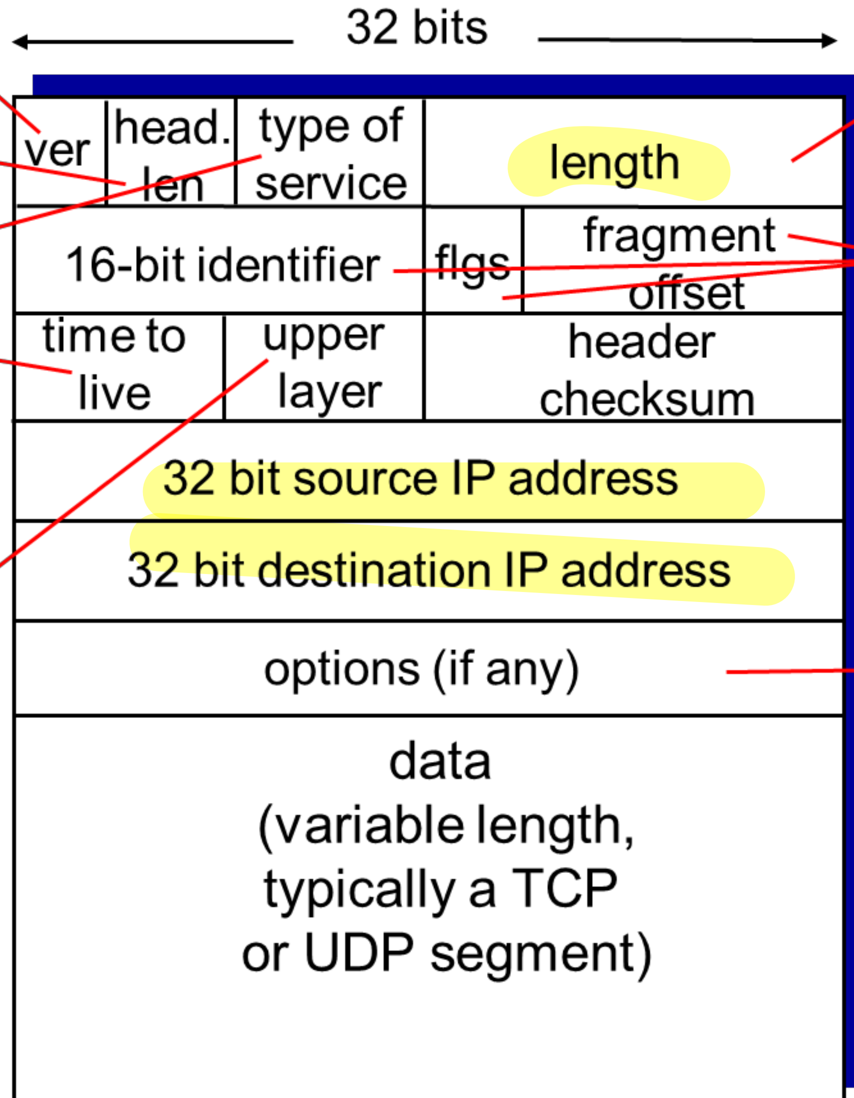


The Internet network layer

Host, Router network layer functions:



IP Datagram format



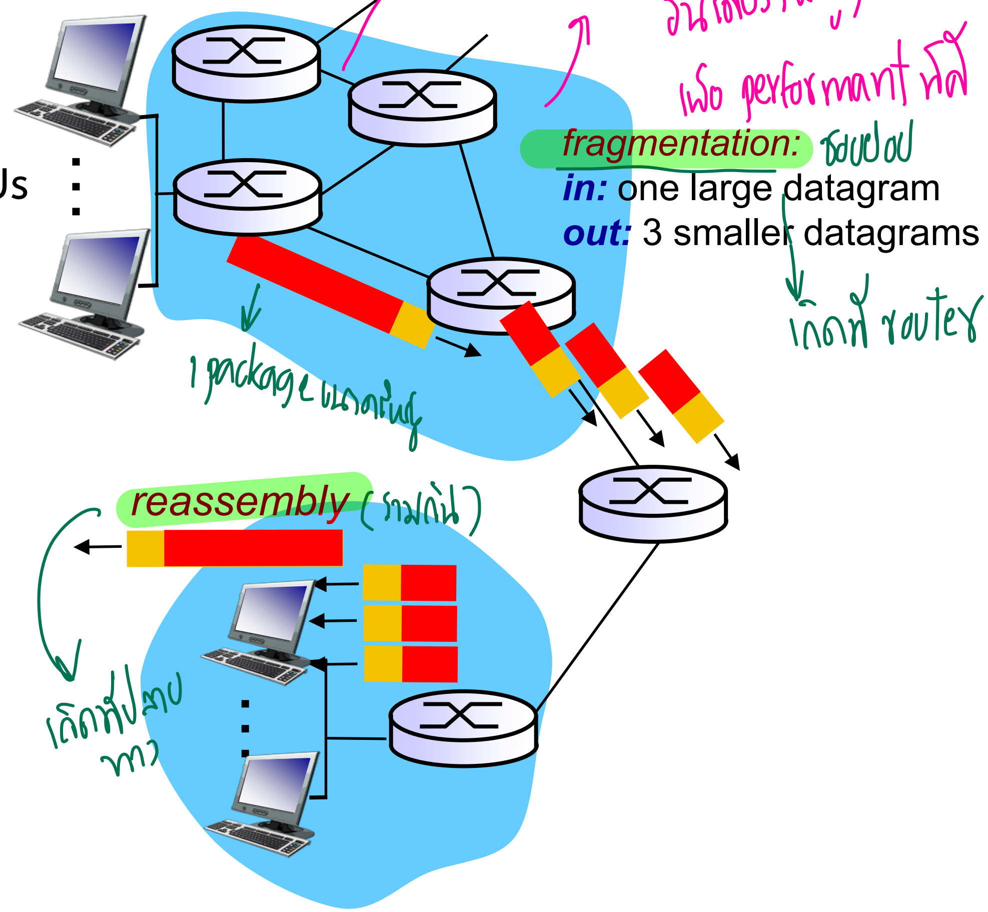
how much overhead?
- 20 bytes of TCP
- 20 bytes of IP
- = 40 bytes + app layer overhead

IP Fragmentation, reassembly

อันใหญ่ จะบอกว่าส่งได้ใหญ่สุดเท่าไร

- Network links have MTU (max transfer size) - largest possible link-level frame
  - Different link types, different MTUs
- Large IP datagram divided
  อันใหญ่ถูกซอยย่อย
  ("fragmented") within net
  - One datagram becomes several datagrams
  - "Reassembled" only at final destination
  - IP header bits used to identify, order related fragments

บัตรระบุขนาดของท่อ ว่ารับ package ใหญ่สุดเท่าไร



ตรงคำนี้ๆ

IP Fragmentation, reassembly

จะบอกขนาดที่ส่งได้

IP Header + Data
(20 byte) (อีก 3980 byte)

① แต่ถ้าโจทย์บอก MTU=1500 4000 ส่งไม่ได้จะไปซอยย่อย (fragment)

Example:

- Length = 4000 bytes
  - ② เอาแค่ data
  - Data = 3980 bytes
- MTU = 1500 bytes ③

-20 ออกไปค่า header = 1480

	length	ID	fragflag	offset	
	=4000	=x	=0	=0	

one large datagram becomes several smaller datagrams

④ เอามา 1480 ก่อน เป็น package แรก

1480 bytes in data field

⑥ ปรับ offset ใหม่ โดยเอามาหาร 8 = 185

offset = 1480/8

⑤

frag ment ①

	length	ID	fragflag	offset		
	=1500	=x	=1	=0		1480
	=1500	=x	=1	=185		1480
	=1040	=x	=0	=370		1020

บวก + header อีก 20 ไปแล้ว

⑦ 3980 - 1480 = 2500

⑧ offset ต่อไป x2

⑨ 3980 - 1480 = 1020

คือ 185×2 = 370
หรือเอา 1480/8 แล้ว
= 185 แล้วเอาไปบวกอันก่อนหน้า 185+185 = 370

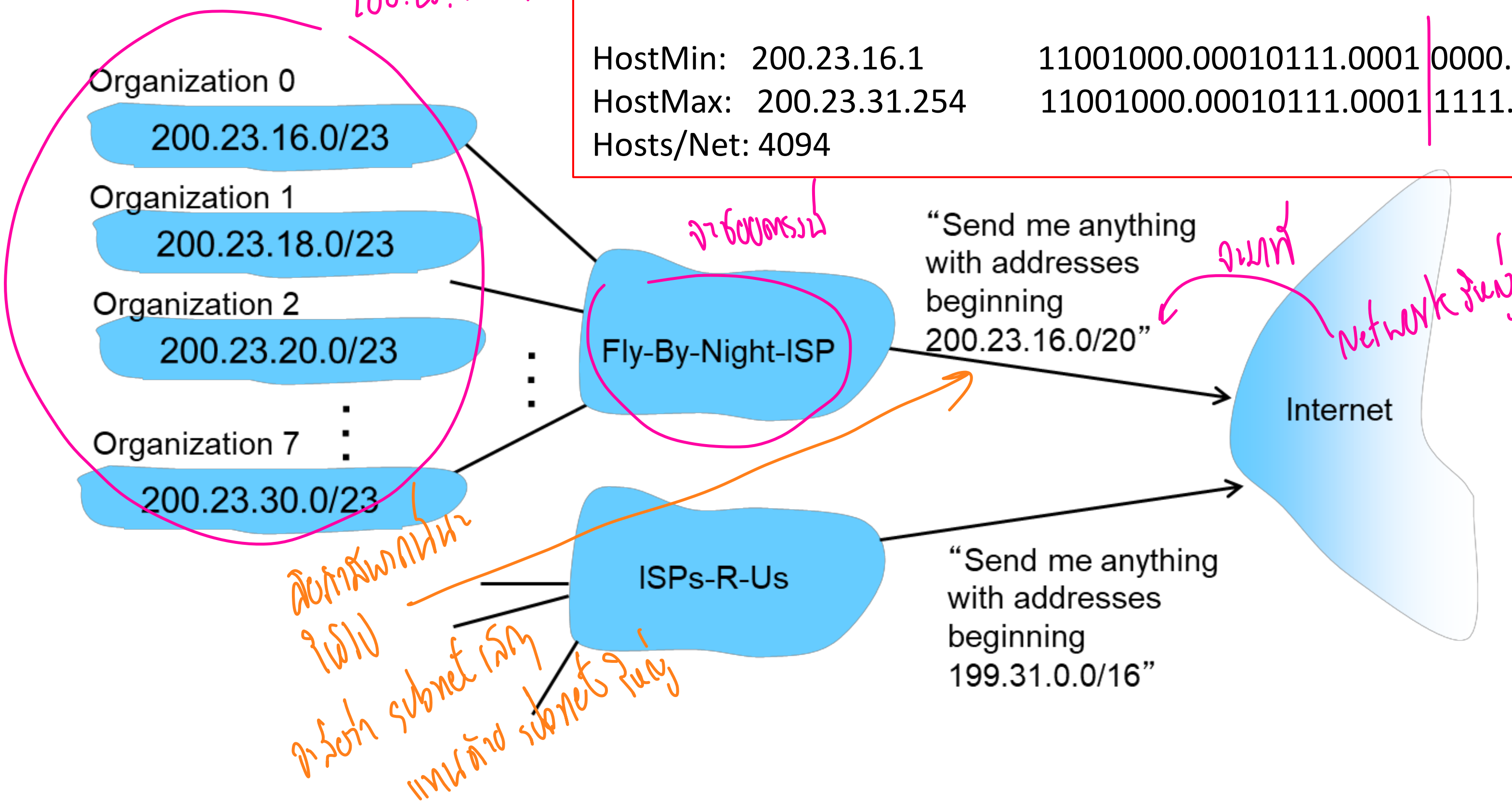
⑩ รวมทั้งหมด 3980 ✓

Hierarchical addressing: route aggregation

รวม subnet เล็กๆ เป็น subnet ใหญ่

- Hierarchical addressing allows efficient advertisement of routing information:

Network:	200.23.16.0/20	11001000.00010111.0001 0000.00000000
Netmask:	255.255.240.0	11111111.11111111.1111 0000.00000000
HostMin:	200.23.16.1	11001000.00010111.0001 0000.00000001
HostMax:	200.23.31.254	11001000.00010111.0001 1111.11111110
Hosts/Net:	4094	

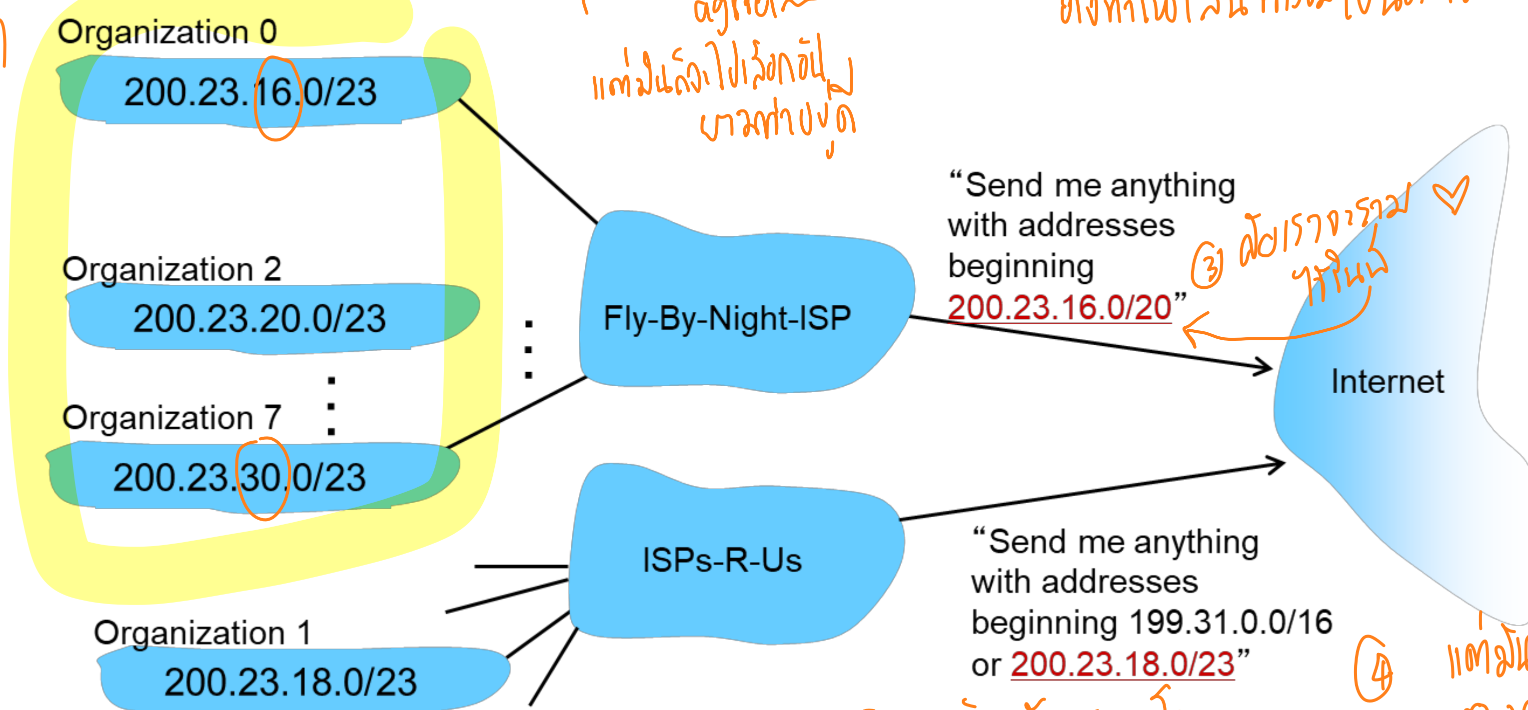


Hierarchical addressing: more specific routes

- ISPs-R-Us has a more specific route to Organization 1
  - Routers in the Internet will use longest prefix matching

↳ คือดูว่าอันไหนตรงมากๆที่สุด ต้องทำให้เส้นทางไม่เป็นหัวใจ

(5) สรุปต้องจะทำ rout agrret~ แต่มันก็จะไปเลือกอันที่ยาวที่สุด



(3) ถ้าเราจะรวม ทำไมนะ

(1) ถ้ามี 200.23.18.0 เข้ามา

(2) แล้วเราจะมีตารางคือ
200.23.16.0/20
200.23.18.0/23

(4) แต่มันดันตรงกับ มากกว่า เลยทำให้เลือก 200.23.18.0/23

Reference

- CCNA 200-301 Official Cert Guide, Volume 1 (2019)
  - By Wendell Odom
- Computer Networking Problems and Solutions (2017)
  - By Russ, Ethan Banks
- Computer Networking: A Top-Down Approach, Global Edition (2016)
  - By Keith Ross James Kurose