



Thiết kế mạng IP

Bài 1: Kết nối liên mạng (Inter-networking)

Phạm Huy Hoàng

SoICT/HUST

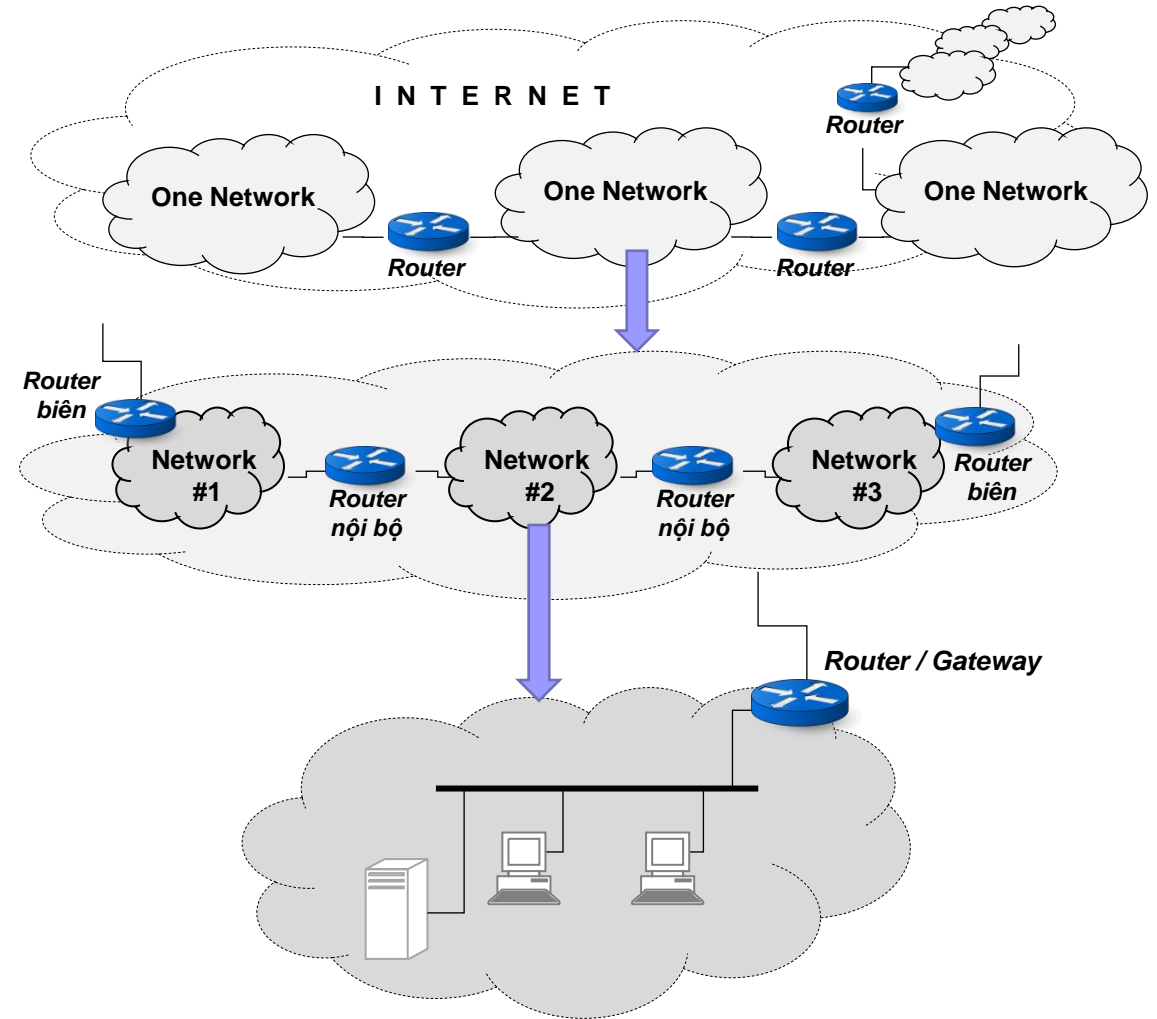
hoangph@soict.hust.edu.vn

Nội dung

- Khái niệm kết nối liên mạng & kết nối IP end-to-end
- Internet backbone & các mạng backbone khác
- Kết nối mạng business vào mạng backbone
- Khái niệm Gateway & bảng routing của Gateway/Router
- Hai giải thuật chuyển gói tin IP từ điểm cuối đến điểm cuối (iner-network & inter-network)
- Routing tĩnh & Routing động
- RIP
- OSPF
- BGP

Khái niệm kết nối liên mạng (internetworking)

- Internet là mạng “lớn nhất” kết nối tất cả các mạng trên thế giới bằng các thiết bị định tuyến (router) ở tầng IP
- “zoom out” một mạng có thể thấy nó được xây dựng bằng nhiều mạng “nhỏ hơn” kết nối với nhau bằng các router nội bộ (internal). Ngoài ra có (một hoặc một số) router kết nối với các mạng bên ngoài khác, gọi là các router biên (border)
 - ➔ IGP [1] & BGP [2]
- Đơn vị “nhỏ nhất” trong hệ thống kết nối liên mạng internetworking là các mạng LAN bao gồm các trạm kết nối (máy tính, máy chủ, các thiết bị IoT, v.v.) và một router đóng vai trò cửa ngõ (gateway) của mạng LAN “đi ra” bên ngoài
- Kết nối liên mạng “phân cấp” thực tế chuyển thành kết nối liên mạng “phẳng” (flat): gói tin IP được chuyển tiếp (store & forward) lần lượt trên các router để đi từ mạng gửi đến mạng nhận

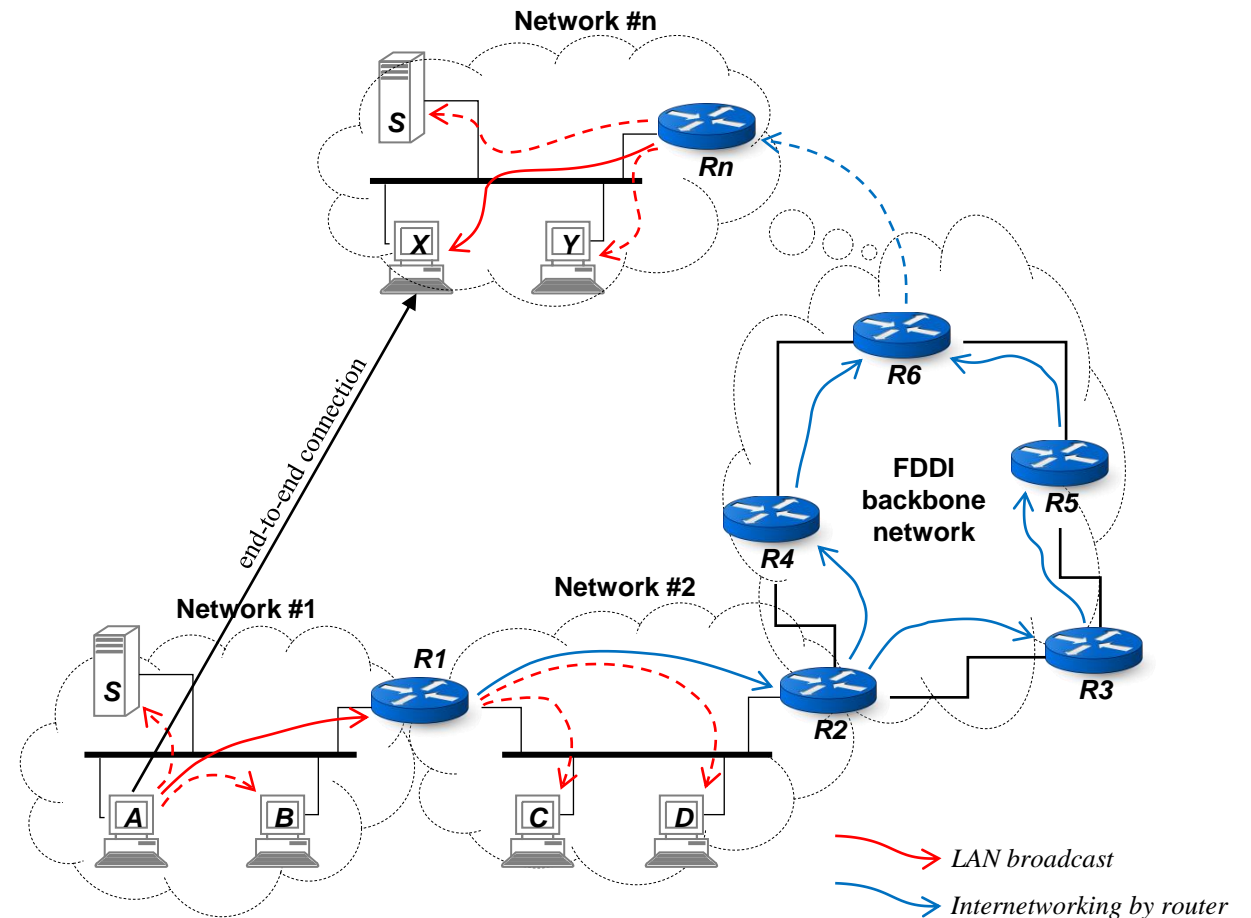


[1] Interior Gateway Protocol: https://en.wikipedia.org/wiki/Interior_gateway_protocol

[2] Border Gateway Protocol: https://en.wikipedia.org/wiki/Border_Gateway_Protocol

Kết nối IP điểm cuối đến điểm cuối (end-to-end)

- Mô hình kết liên mạng (internetworking) giải quyết bài toán chuyển gói tin IP giữa các mạng. Thực tế gói tin IP cần chuyển từ điểm cuối đến điểm cuối (ví dụ các trạm làm việc): end-to-end connection
- Mô hình kết nối IP end-to-end:
 - LAN broadcast connection: trạm truyền → gateway
 - Inter-networking (router store & forward): gateway → mạng đích
 - LAN broadcast connection: gateway mạng đích → trạm nhận
- Hoạt động:
 - Các trạm trong cùng mạng nội bộ liên kết trực tiếp với nhau qua đường truyền vật lý
 - Gói tin IP khi chuyển xuống tầng 2 được broadcast trên đường truyền đến tất cả các trạm trong mạng LAN (trong đó có máy gateway)
 - Gateway/Router sử dụng thuật toán tìm đường (dựa trên bảng routing của mình và địa chỉ IP trong gói tin IP) để xác định router tiếp theo cần chuyển tiếp gói tin

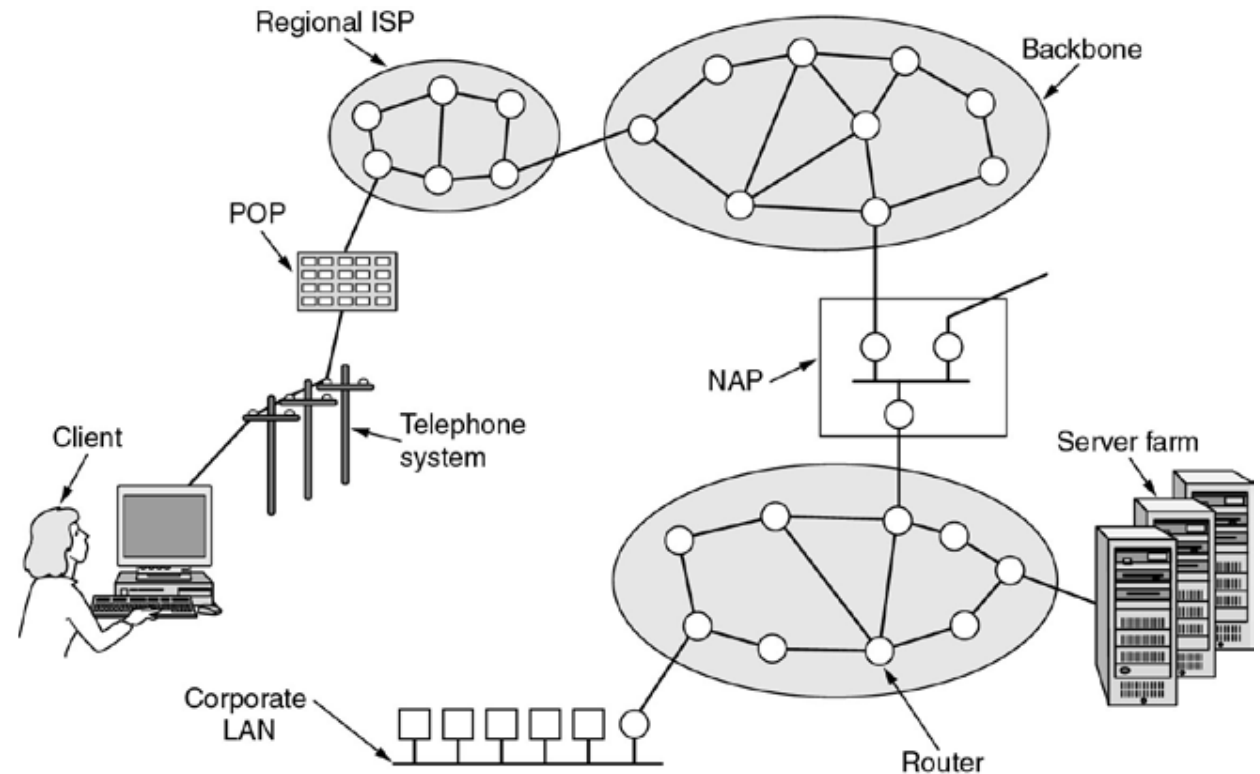


$A \rightarrow X = [A \rightarrow B, S, R1] \& [R1 \rightarrow C, D, R2] \& [R2 \rightarrow R4] \& [R4 \rightarrow R6] \& [R6 \rightarrow \dots] \& [\dots \rightarrow Rn] \& [Rn \rightarrow Y, S, X]$

Mô hình kết nối liên mạng - 2003

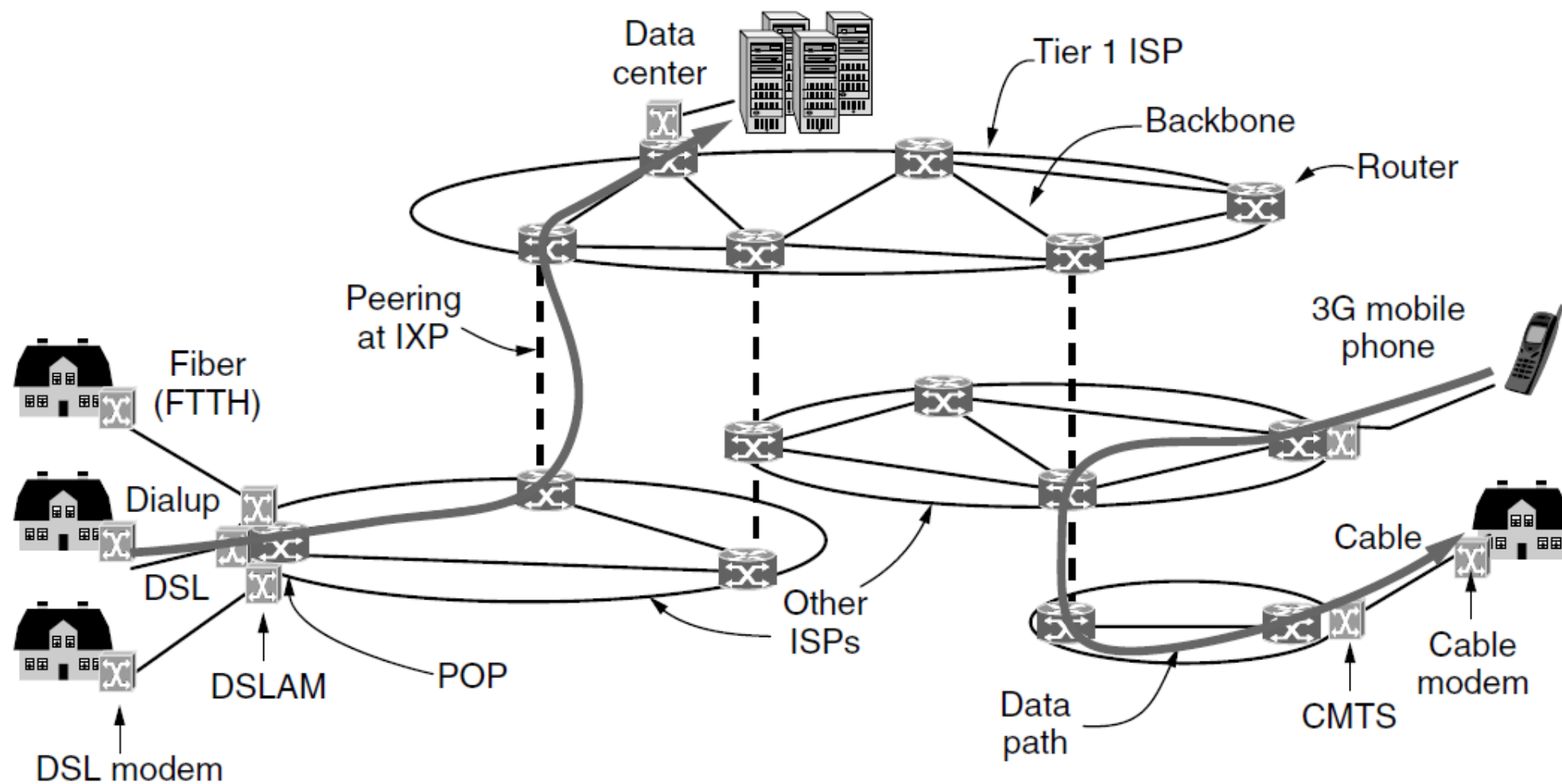
■ Bài tập tại lớp:

- Sử dụng kinh nghiệm thực tế sử dụng/cài đặt kết nối Internet, bạn hãy nêu tên & chức năng các thành phần trong mô hình kết nối thực tế (hình vẽ bên)
- Thành phần nào trong hình vẽ đã được phát triển (tiến hóa) & cần cập nhật lại trong hình vẽ?
- Nêu tên các tổ chức thực tế quản lý các thành phần kết nối mạng
- Xác định các mạng & các mạng con



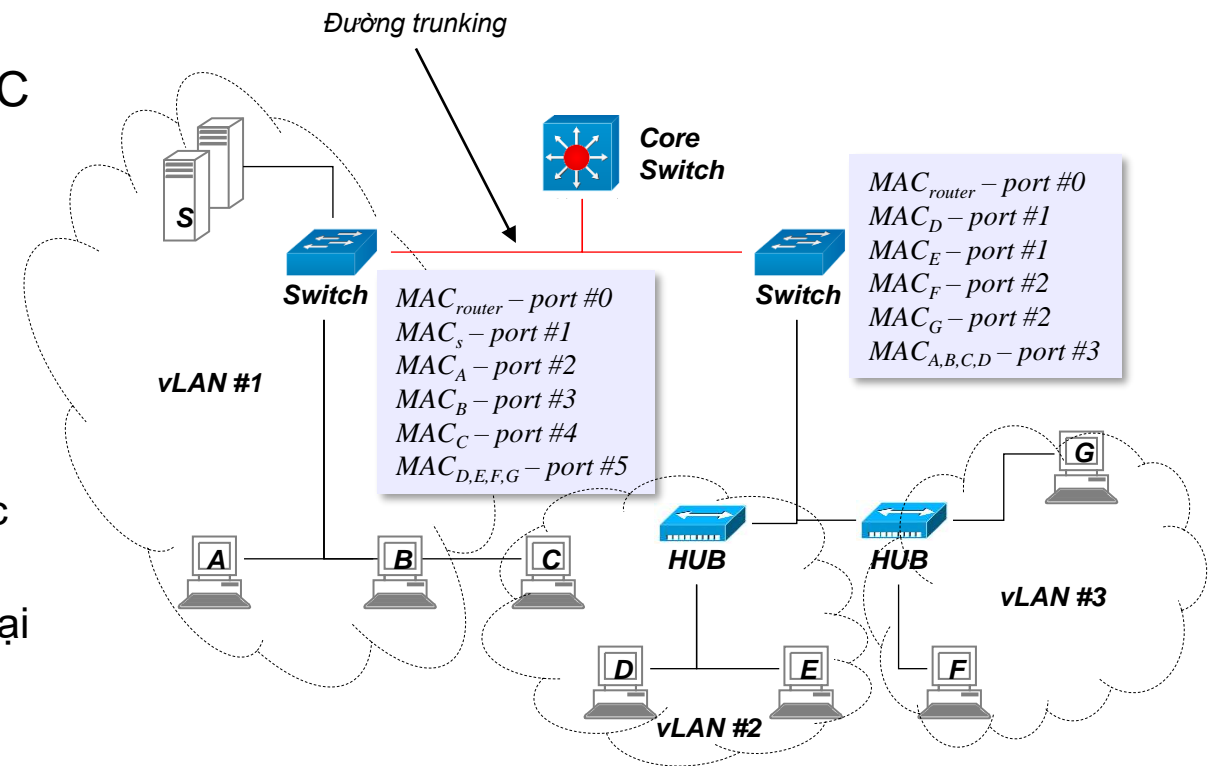
A. S. Tanenbaum, Computer Networks, 4th edition, 2003

Mô hình kết nối liên mạng - 2020



Broadcast zone tầng 2 & Virtual LAN

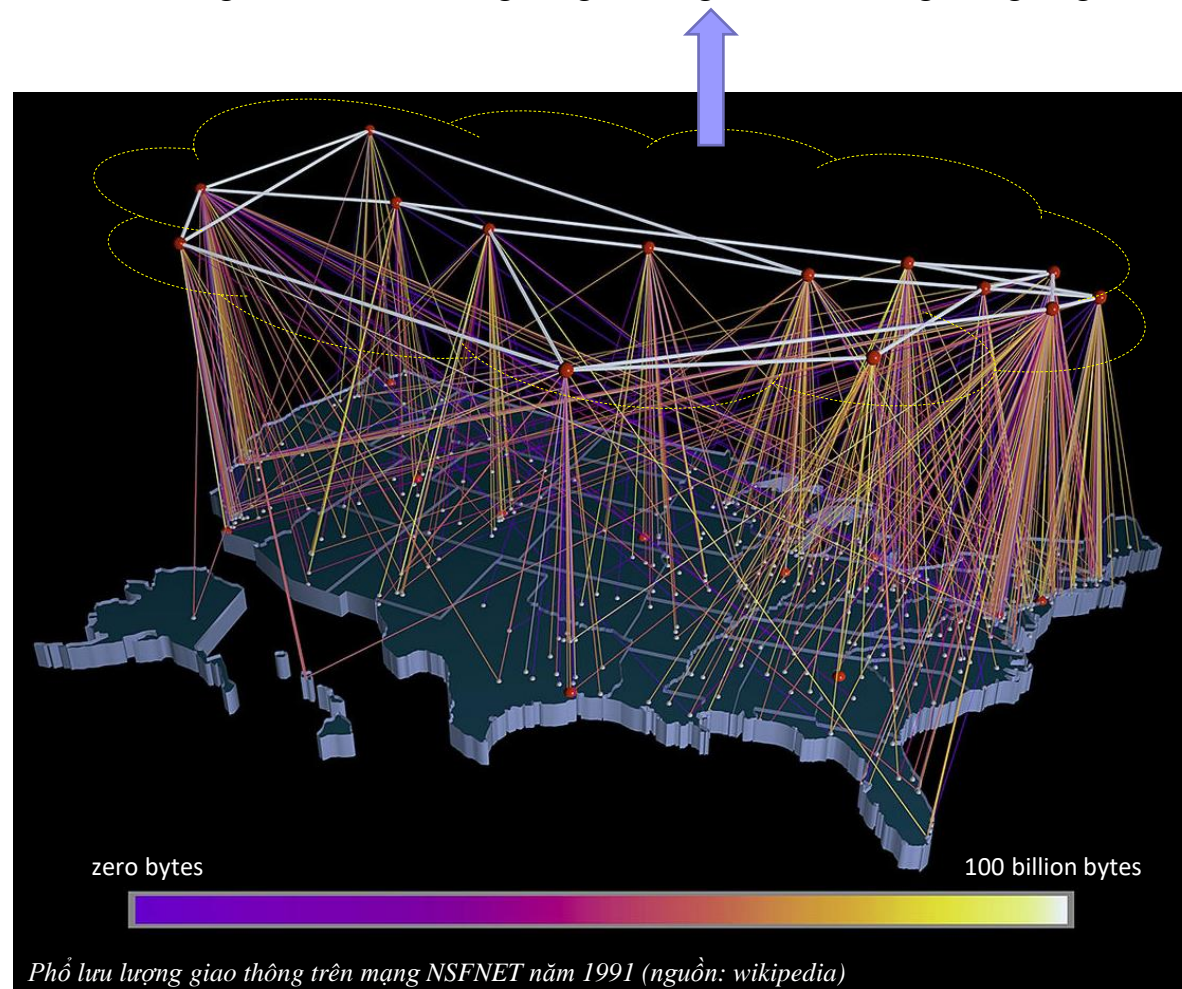
- Switch hoạt động ở tầng 2 (tổng hợp tín hiệu nhận trên đường truyền thành frame dữ liệu ở tầng 2) → xác định được địa chỉ MAC của trạm nhận → lựa chọn cổng switch phù hợp để “copy frame”.
- Switch duy trì bảng ARP mapping giữa địa chỉ MAC máy tính với cổng switch mà nó kết nối
- Broadcast zone: các trạm nhận được data frame gửi broadcast
- Virtual LAN:
 - Xây dựng với mạng Switch, thay vì có 1 broadcast zone (toàn bộ mạng LAN) thì tổ chức thành nhiều broadcast zone. Mỗi broadcast zone có thể gồm các máy kết nối phân tán trên nhiều switch
 - Các broadcast zone được quản lý tập trung (ví dụ tại core switch) theo từng mã số riêng, và được ánh xạ đến từng cổng switch trong LAN
 - Các switch xử lý broadcast frame theo mã broadcast zone và các cổng có cùng mã này



Internet backbone > < Local network

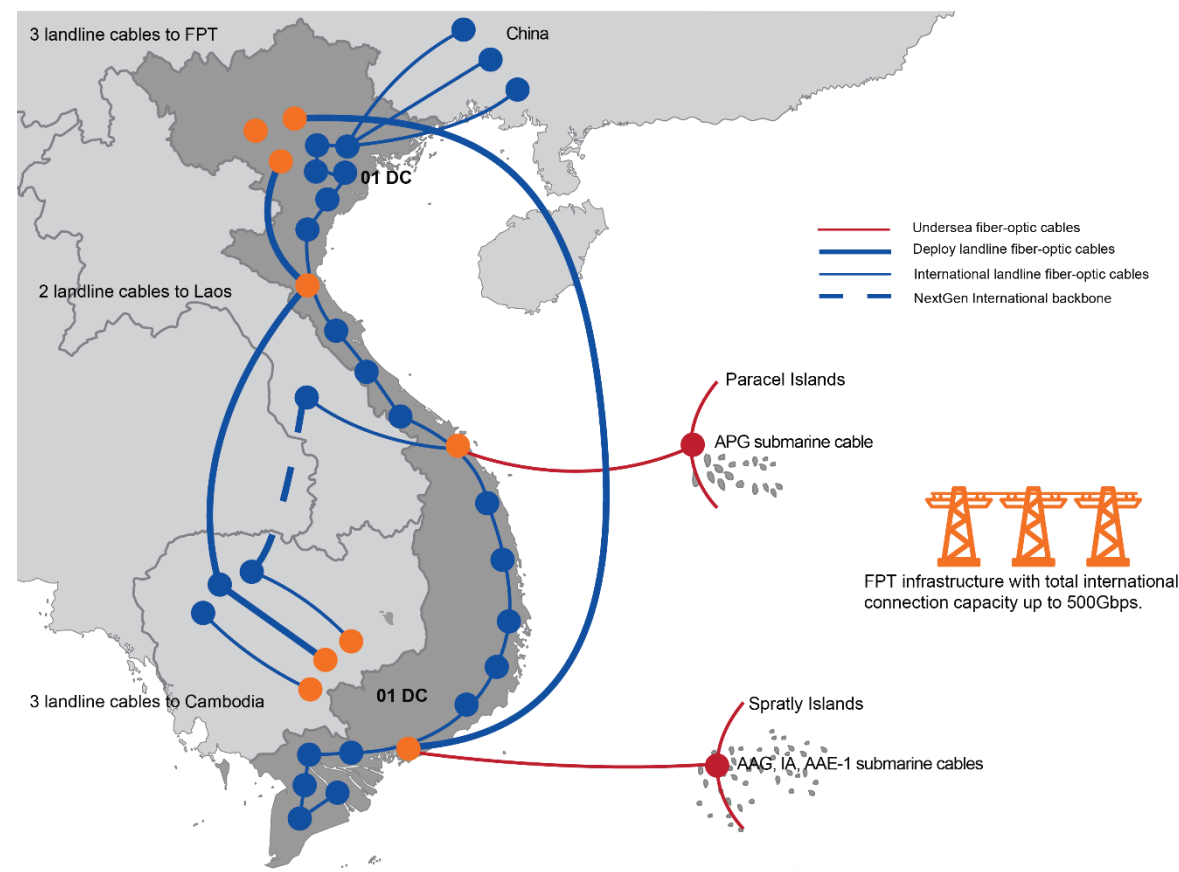
- Giao thức IP cho phép đồng nhất (về mặt lý thuyết) việc truyền gói tin trong mạng nội bộ, mạng diện rộng và mạng Internet.
- Internet = 1 mạng, nhưng tải lưu lượng trên các kênh truyền là rất khác nhau
- Lưu lượng đường truyền mạng LAN ~ nhu cầu trao đổi thông tin nội bộ
- Lưu lượng đường truyền giữa VN với mạng quốc tế ~ nhu cầu trao đổi thông tin giữa VN với bên ngoài
- → Tổ chức một số đường truyền đặc biệt để đảm bảo đáp ứng nhu cầu lưu lượng

“mạng con” đặc biệt, gồm các kênh truyền có lưu lượng cao, cùng với các thiết bị mạng cấu hình mạnh, có nhiệm vụ đảm bảo các hoạt động quan trọng của toàn bộ hệ thống mạng → mạng backbone (mạng xương sống)



Các mạng backbone

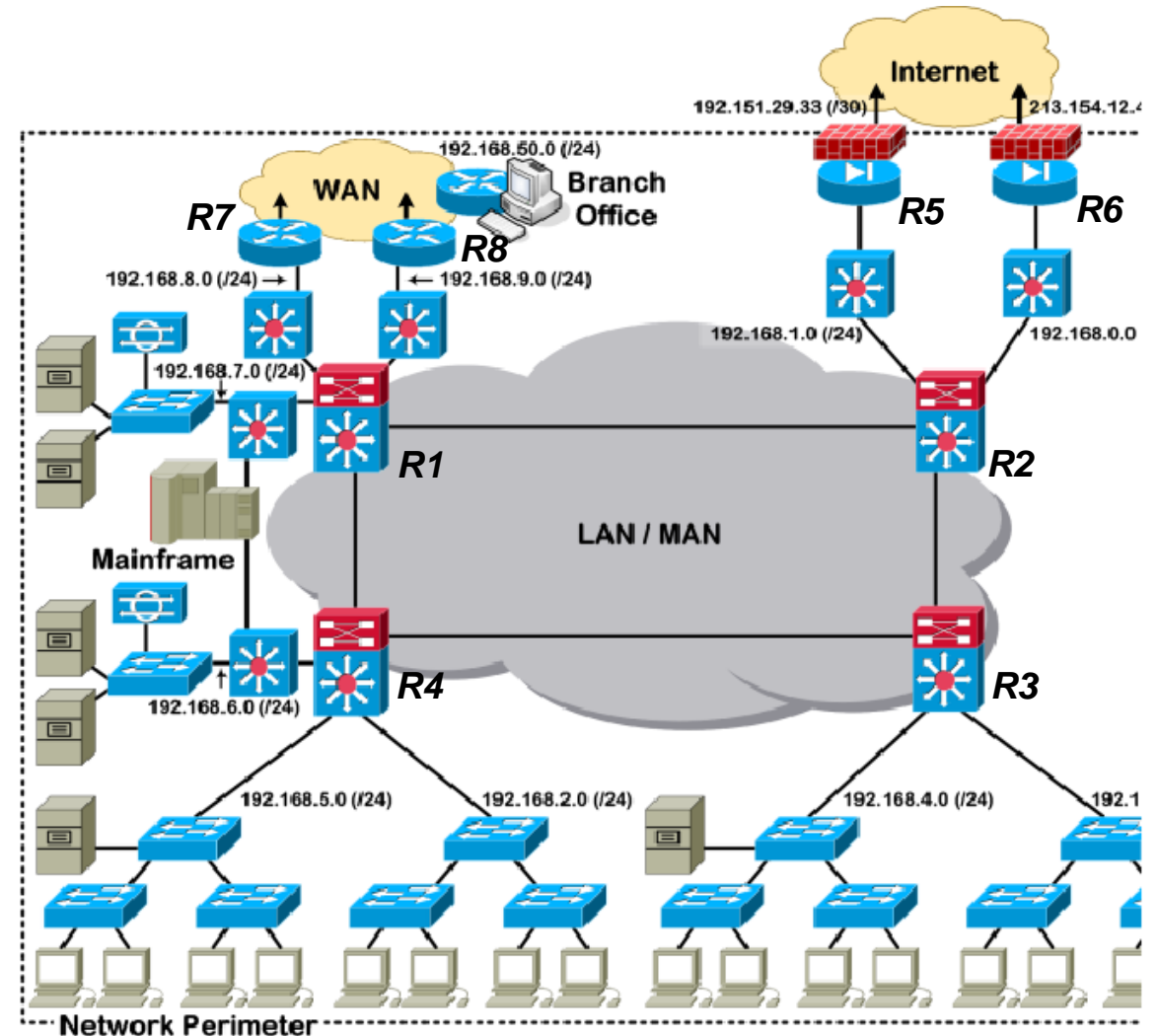
- Các tổ chức kinh doanh, chính phủ, giáo dục, v.v.. có nhu cầu riêng về lưu lượng đường truyền giữa các trạm/mạng của tổ chức
- Nhu cầu đảm bảo kết nối mạng 24/7 cũng được xem xét bên cạnh yêu cầu về đáp ứng lưu lượng đường truyền
- → Mỗi tổ chức có thiết kế hệ thống mạng riêng với hệ thống mạng backbone riêng.
- → Mạng backbone có vai trò đảm bảo tính sẵn sàng trong kết nối đường truyền và đảm bảo hỗ trợ lưu lượng truyền dữ liệu trong nội bộ tổ chức
- → Thông thường, hệ thống mạng backbone riêng này có điểm kết nối với hệ thống mạng backbone bên ngoài để đảm bảo kết nối với bên ngoài



<https://fpt.vn/en/business/services/internet-leased-line.html>

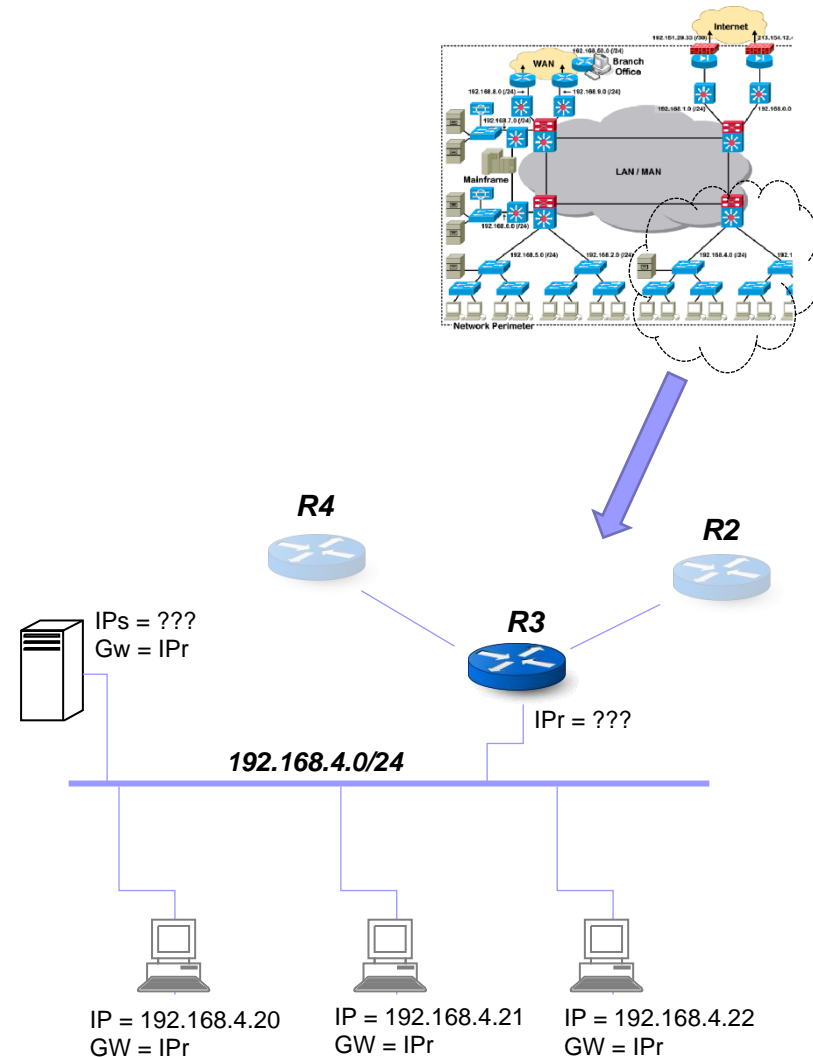
Kết nối các mạng business vào mạng backbone

- Mạng business: kết nối các trạm làm việc của người dùng để thực hiện các tác vụ business của tổ chức
- Khi kết nối nhiều mạng business của tổ chức, hoặc các trạm làm việc có vị trí “xa nhau”, hoặc cần kết nối mạng business với mạng bên ngoài → cần sử dụng mạng backbone
- Bài tập tại lớp: phân tích sơ đồ mạng bên cạnh
 - Đây là mạng backbone?
 - Kết nối với mạng bên ngoài như thế nào?
 - Các mạng business của tổ chức nằm ở đâu?
 - Kết nối các mạng business này vào mạng backbone được thực hiện thế nào?
 - Xác định các trạm đặc biệt (gateway) có vai trò chuyển tiếp các gói tin IP từ trong mạng business ra ngoài?



“Zoom in” một mạng business

- Địa chỉ IP gán cho 1 mạng business
- Địa chỉ các trạm trong mạng tương thích với địa chỉ IP của mạng (chú ý phân biệt địa chỉ trạm làm việc và máy chủ)
- Xác định trạm Gateway & liệt kê các mạng business vào bảng routing
- Các thông số cấu hình địa chỉ cho các trạm trong mạng
- Hoạt động gửi gói tin IP tại một trạm:
 - Xác định gói tin IP nội bộ hay gửi ra ngoài
 - Nội bộ → chuyển đổi địa chỉ IP trạm nhận thành địa chỉ MAC của trạm nhận & gửi xuống tầng 2 để broadcast lên đường truyền
 - Ra ngoài → gửi xuống tầng 2 với địa chỉ MAC của trạm nhận là Gateway
- Hoạt động nhận gói tin IP tại một trạm:
 - Trạm làm việc hoặc máy chủ: tầng 2 kiểm tra gói tin (frame) vừa nhận & so sánh địa chỉ MAC trong gói tin với MAC của mình để quyết định xử lý tiếp (chuyển lên tầng 3) hoặc loại bỏ
 - Trạm Gateway: nhận gói tin IP, trích xuất thông tin địa chỉ IP đích từ gói tin, so sánh với bảng tìm đường (routing table), chuyển tiếp gói tin đến router tương ứng được mô tả sẵn trong bảng tìm đường
- → Làm sao Gateway liệt kê hết được các mạng trên Internet để phục vụ tìm đường?



2 giải thuật chuyển tiếp gói tin IP

■ Hoạt động gửi gói tin IP tại trạm truyền:

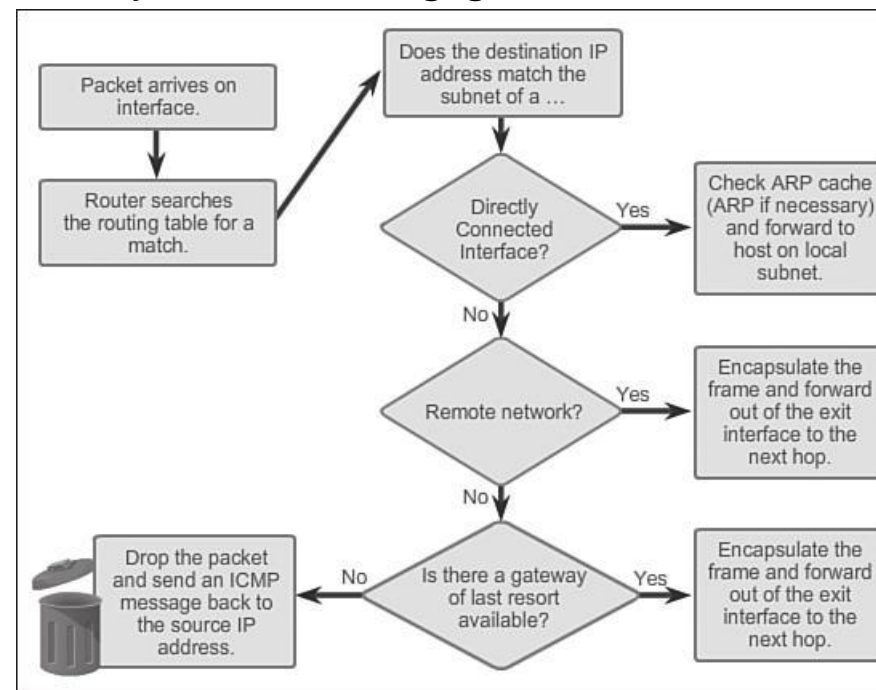
Hoạt động gửi gói tin IP:

- Xác định gói tin IP nội bộ hay gửi ra ngoài
- Nội bộ → chuyển đổi địa chỉ IP trạm nhận thành địa chỉ MAC của trạm nhận & gửi xuống tầng 2 để broadcast lên đường truyền
- Ra ngoài → gửi xuống tầng 2 với địa chỉ MAC của trạm nhận là Gateway

■ Bài tập: chuyển thành sơ đồ mô tả giải thuật

- ☐ Cách xác định gói tin IP gửi nội bộ hay ra ngoài?
- ☐ Cách chuyển đổi địa chỉ IP thành địa chỉ MAC?

■ Giải thuật chuyển tiếp gói tin IP tại gateway/router trung gian



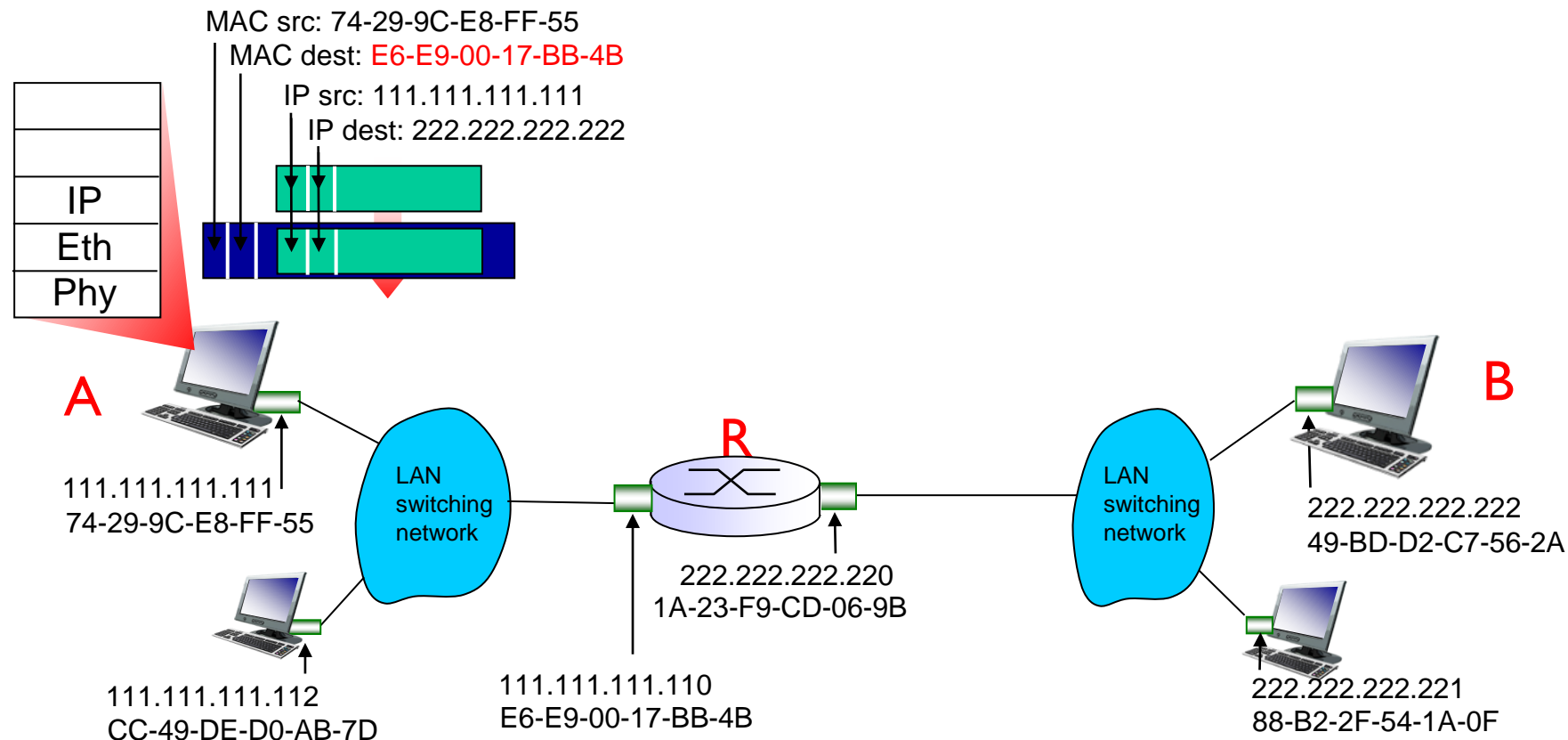
Tìm hiểu về địa chỉ IP classless và CDIP [1]

[1] https://en.wikipedia.org/wiki/Classless_Inter-Domain_Routing

Chuyển tiếp dữ liệu trong LAN & giữa các LAN

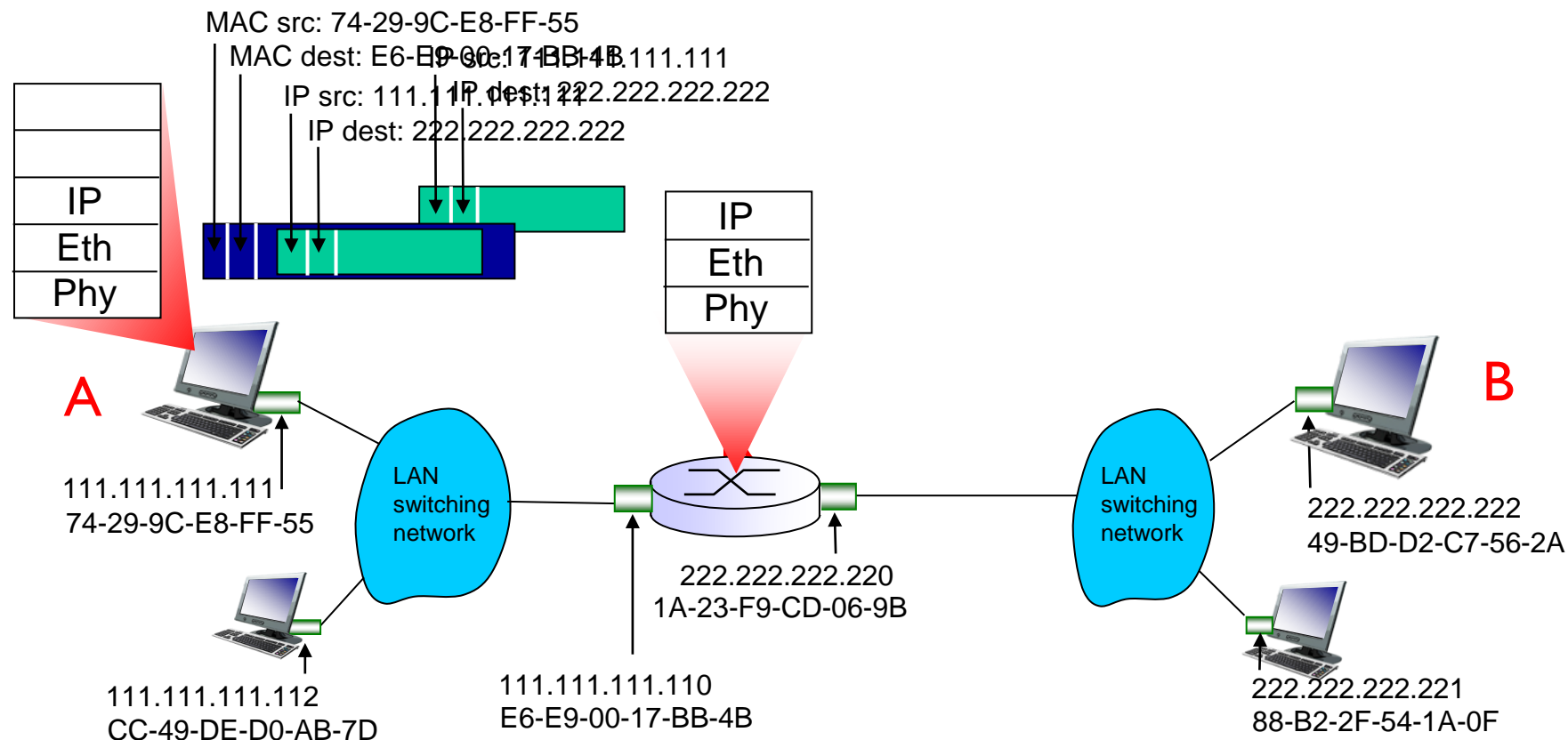
Ví dụ: **Gửi dữ liệu từ A tới B qua router R**

- A soạn một gói tin IP với địa chỉ nguồn là A, địa chỉ đích là B
- A xác định B không nằm trong cùng mạng → gói tin IP chuyển xuống tầng Data Link được đóng gói Data Frame với địa chỉ MAC nguồn là A, **địa chỉ MAC đích là R**



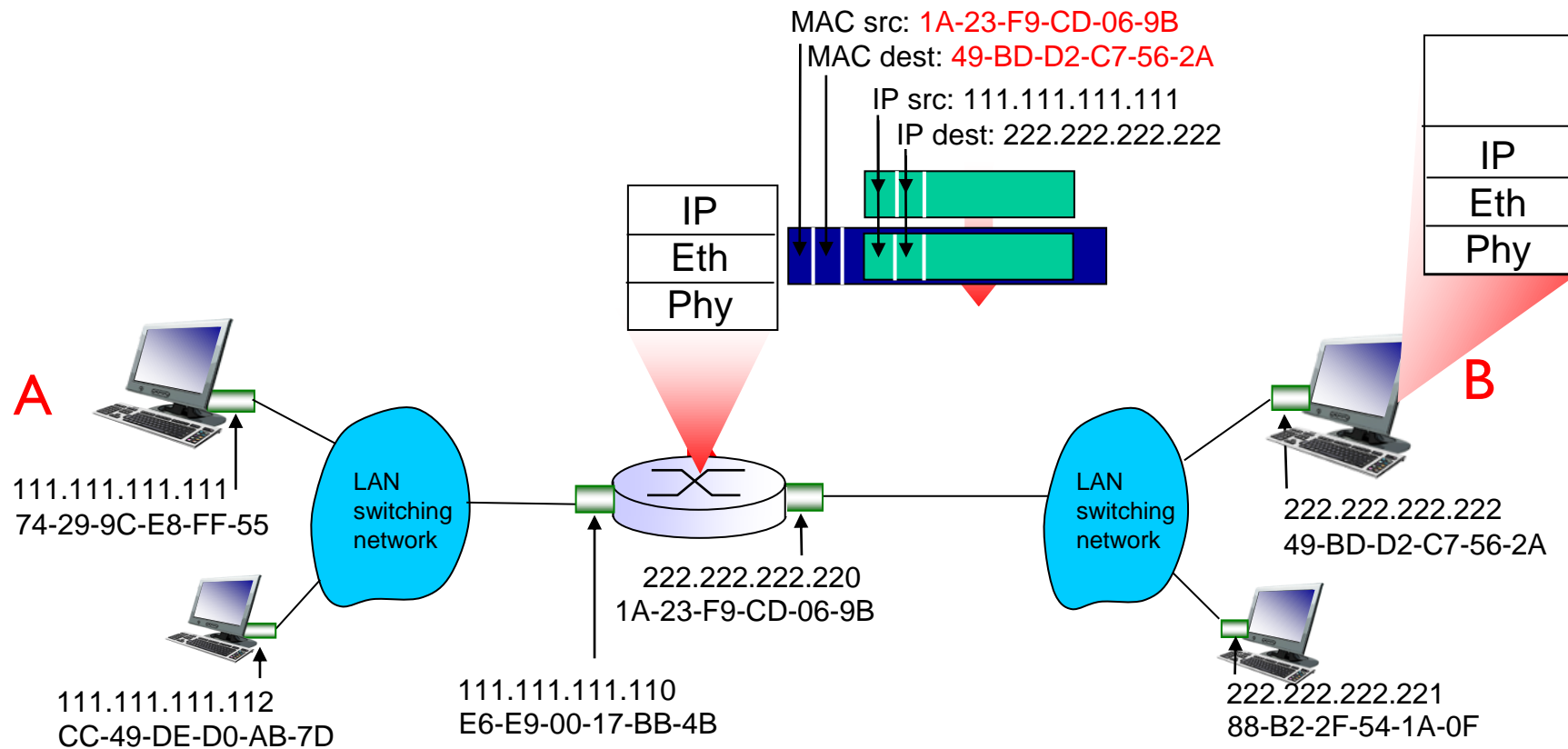
Chuyển tiếp dữ liệu trong LAN & giữa các LAN

- ❖ Data Frame được chuyển từ A tới R trong mạng switch với cơ chế *broadcast zone*
- ❖ Tại R: Data Frame được tách phần data (thành gói tin IP) & chuyển lên cho tầng IP



Chuyển tiếp dữ liệu trong LAN & giữa các LAN

- ❖ Sử dụng giải thuật chuyển tiếp gói tin với bảng routing, R chuyển gói tin IP từ mạng LAN A sang mạng LAN B, gói tin vẫn có địa chỉ IP nguồn là A, IP đích là B
- ❖ Gói tin chuyển xuống tầng Data Link, với địa chỉ MAC nguồn là R, địa chỉ MAC đích là B & được chuyển từ R tới B trong mạng switch với cơ chế *broadcast zone*

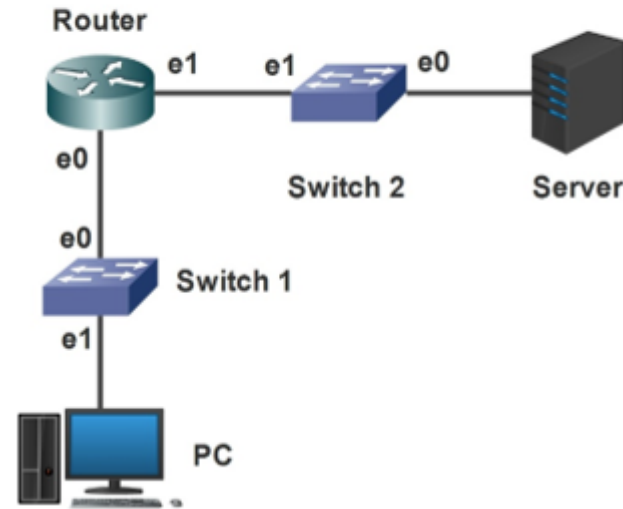


Câu 13.

Cho sơ đồ mạng như hình dưới đây và địa chỉ các nút mạng cho trong bảng.

→

Interface	Địa chỉ MAC	Địa chỉ IP
PC	cc-cc-cc-11-11-11	192.168.1.10
Switch1-e0	11-11-11-e0-e0-e0	
Switch1-e1	11-11-11-e1-e1-e1	
Router-e0	cc-cc-cc-e0-e0-e0	192.168.1.1
Router-e1	bb-bb-bb-e1-e1-e1	10.0.0.1
Switch2-e0	22-22-22-e0-e0-e0	
Switch2-e1	22-22-22-e1-e1-e1	
Server	bb-bb-bb-22-22-22	10.0.0.20



Giả sử máy trạm PC gửi một thông điệp ICMP tới máy chủ Server.

a. Hãy cho biết các thông số địa chỉ trên gói tin được gửi đi từ PC?

Trả lời: (0.5 điểm)

- Địa chỉ MAC nguồn: cc-cc-cc-11-11-11
- Địa chỉ MAC đích: cc-cc-cc-e0-e0-e0
- Địa chỉ IP nguồn: 192.168.1.10
- Địa chỉ IP đích: 10.0.0.20

b. Hãy cho biết các thông số địa chỉ trên gói tin khi tới Server?

Trả lời: (0.5 điểm)

- Địa chỉ MAC nguồn: bb-bb-bb-e1-e1-e1
- Địa chỉ MAC đích: bb-bb-bb-22-22-22
- Địa chỉ IP nguồn: 192.168.1.10
- Địa chỉ IP đích: 10.0.0.20

Routing table tại Gateway & Router

Luật: $\langle IP \text{ mạng đích} \rangle \rightarrow \langle \text{next hop} \rangle$

■ R1:

- Network #1 \rightarrow direct
- Network #2 \rightarrow direct
- Network #n \rightarrow R2

■ R2:

- Network #1 \rightarrow R1
- Network #2 \rightarrow direct
- Network #n \rightarrow R3 | R4

■ R3:

- Network #1 \rightarrow R2
- Network #2 \rightarrow R2
- Network #n \rightarrow R5

• R4:

- Network #1 \rightarrow R2
- Network #2 \rightarrow R2
- Network #n \rightarrow R6

• R5:

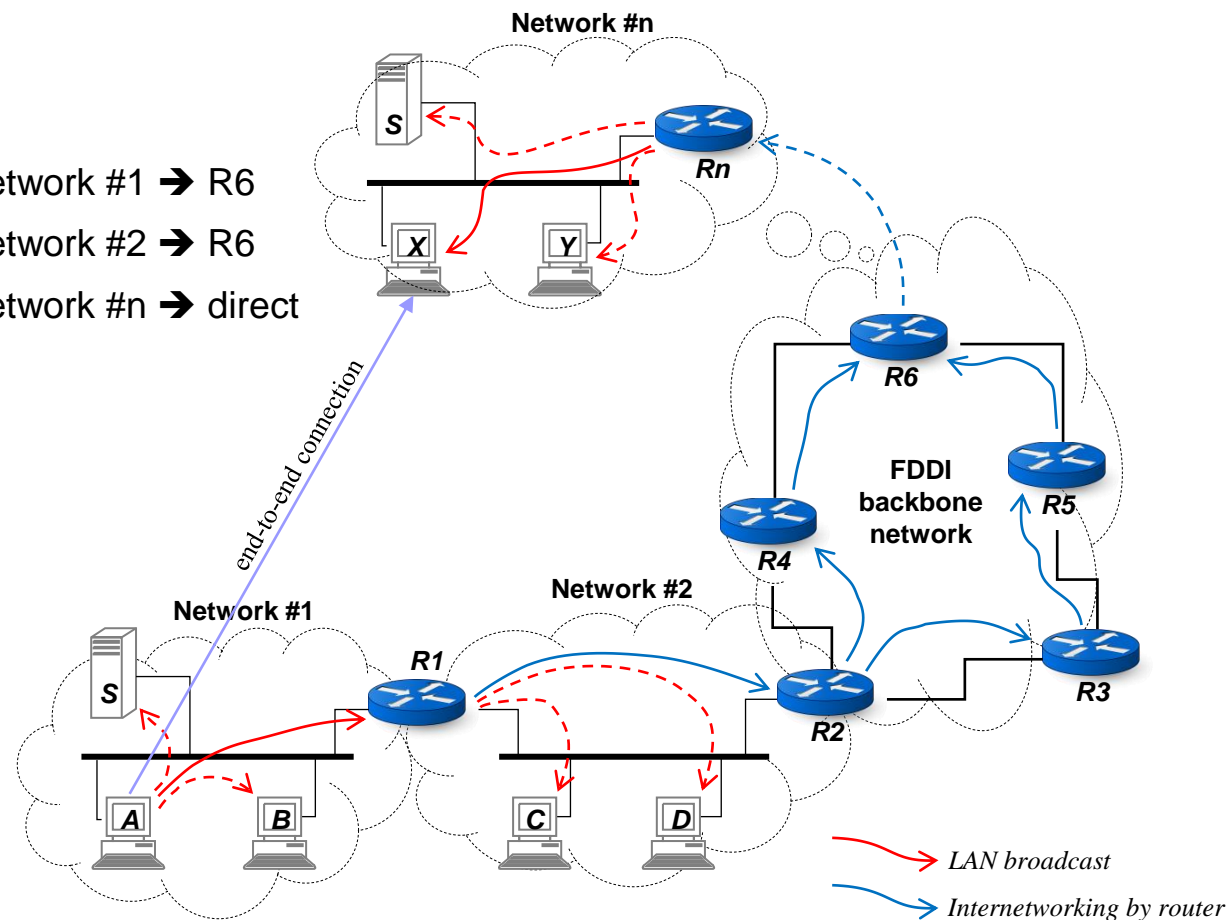
- Network #1 \rightarrow R3
- Network #2 \rightarrow R3
- Network #n \rightarrow R6

• R6:

- Network #1 \rightarrow R4 | R5
- Network #2 \rightarrow R4 | R5
- Network #n \rightarrow Rn

• Rn:

- Network #1 \rightarrow R6
- Network #2 \rightarrow R6
- Network #n \rightarrow direct



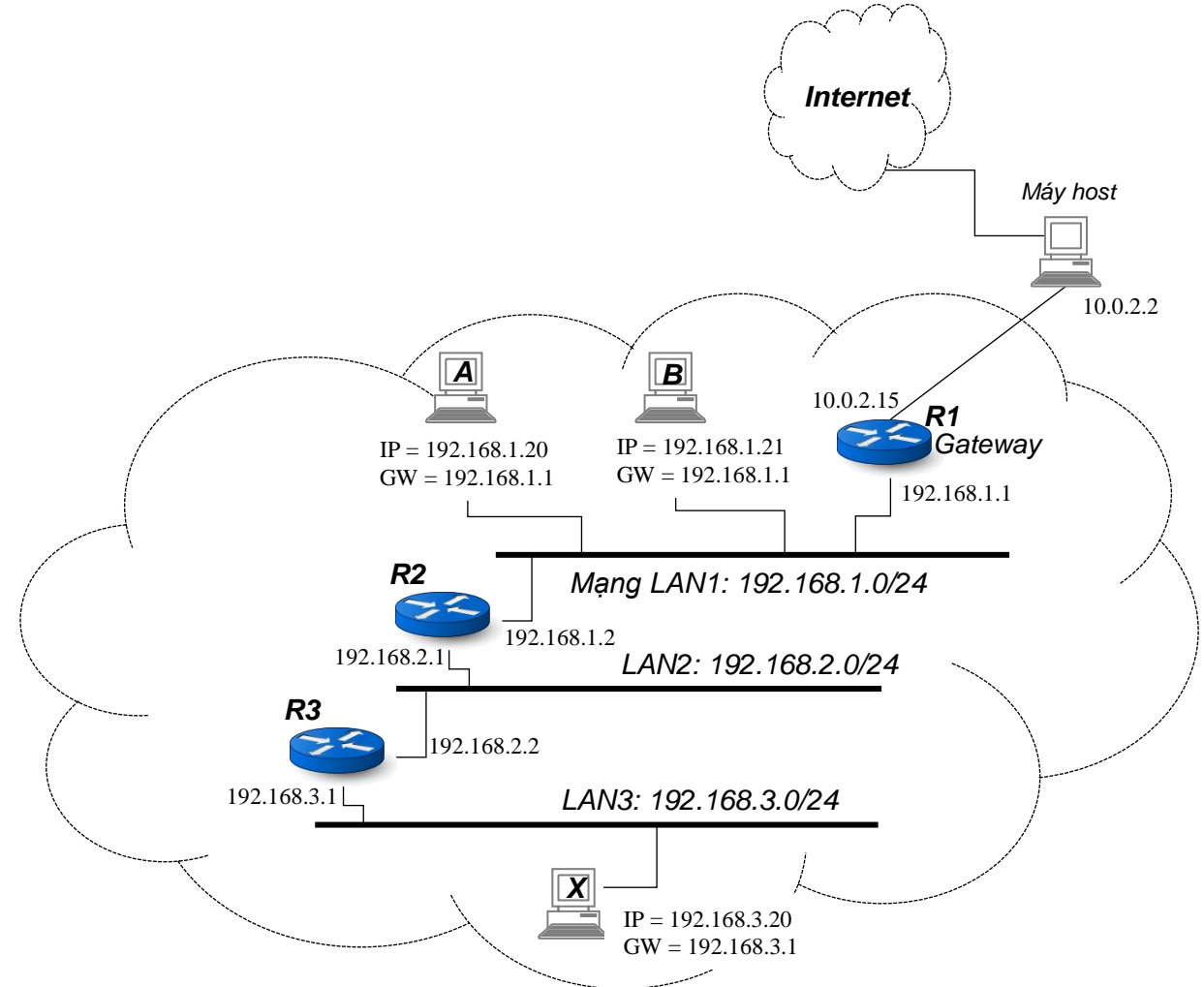
Gói tin IP ($A \rightarrow X$) = header[$IP_a \rightarrow IP_x$] & data = header[($IP_{net_a} + IP_{host_a}$) \rightarrow ($IP_{net_x} + IP_{host_x}$)] & data

Bài toàn routing với classless IP?

thực hành:

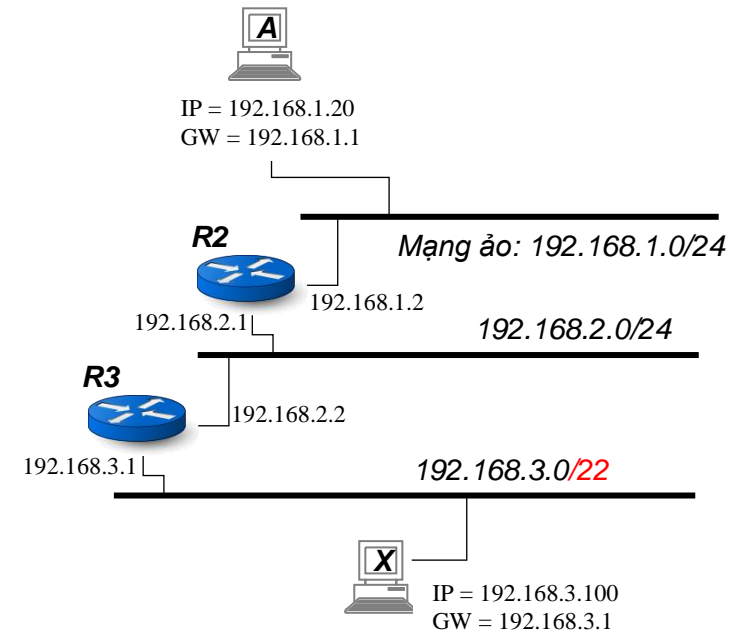
Thiết lập môi trường giả lập mạng & kết nối Internet thông qua máy host

- Cài đặt Virtualbox, các máy ảo & mạng ảo
- Thiết lập kết nối giữa máy gateway ảo R với máy host (máy thật) để gateway kết nối được ra Internet
- Thiết lập mạng ảo với các máy trạm ảo A & B, gateway mạng ảo là R để A & B cũng kết nối được ra Internet
- Tạo thêm 2 mạng ảo 192.168.2.0 và 192.168.3.0 kết nối với mạng 192.168.1.0 qua các router R2 & R3
- Cấu hình các bảng routing trên R1, R2, R3 để toàn bộ hệ thống kết nối được với nhau và Internet



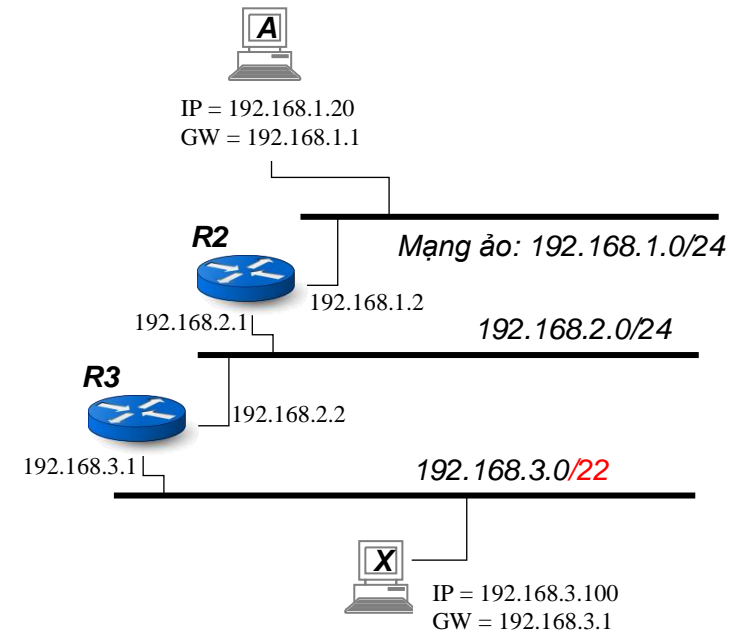
Một số chú ý với CIRP

- CIRP: Classless Inter-Domain Routing [1]
- Chồng lấn dải địa chỉ IP (IP address overlapping)
- Ví dụ:
 - A ping R2,R3: ok (?)
 - R3 ping R2,A: ok
 - X ping R3: ok
 - X ping R2,A: **not ok**
- Debug:
 - Enable *iptables* trên các router R2, R3
> *service iptables start*
 - Xóa luật tường lửa cấm gói tin đi qua, trả lời bằng ICMP prohibited:
> *iptables -L -line-number*
> *iptables -D FORWARD <#>*
 - Kiểm tra log các gói tin ICMP đi qua router
> *iptables -t mangle -A FORWARD -j LOG*
> *tail -f /var/log/message | grep ICMP*



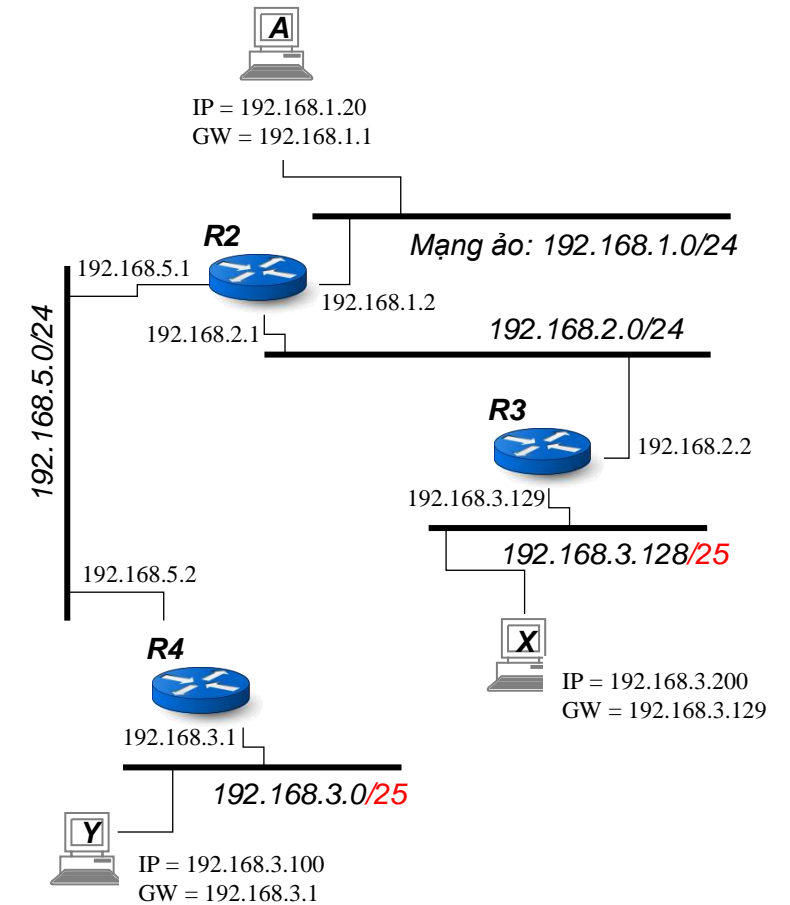
Một số chú ý với CIRP (2)

- R2 routing:
 - Routing table: 192.168.3.0/22 (~192.168.0.0) → 192.168.2.2
 - A ping R3: [192.168.1.20] ping 192.168.3.100
 - Thuật toán: tìm IPnet của dest (192.168.3.0/24) trong routing table → không khớp. Sao chạy?
 - Subnet match: khớp với 192.168.0.0/22 (subnet)
- Dải địa chỉ 192.168.3.0/22 chồng lấn (overlap) dải địa chỉ 192.168.1.0/24
- Thuật toán routing trong mạng LAN:
 - X ping A: [192.168.3.100/22] ping 192.168.1.21
 - So sánh IPnet của source & dest: đều là 192.168.0.0/22
 - → không gửi gói tin ra gateway
- Vấn đề tránh được khi qui hoạch địa chỉ IP cho các mạng thuộc cùng một zone quản trị (Autonomous Zone/System)



Một số chú ý với CIRP (3)

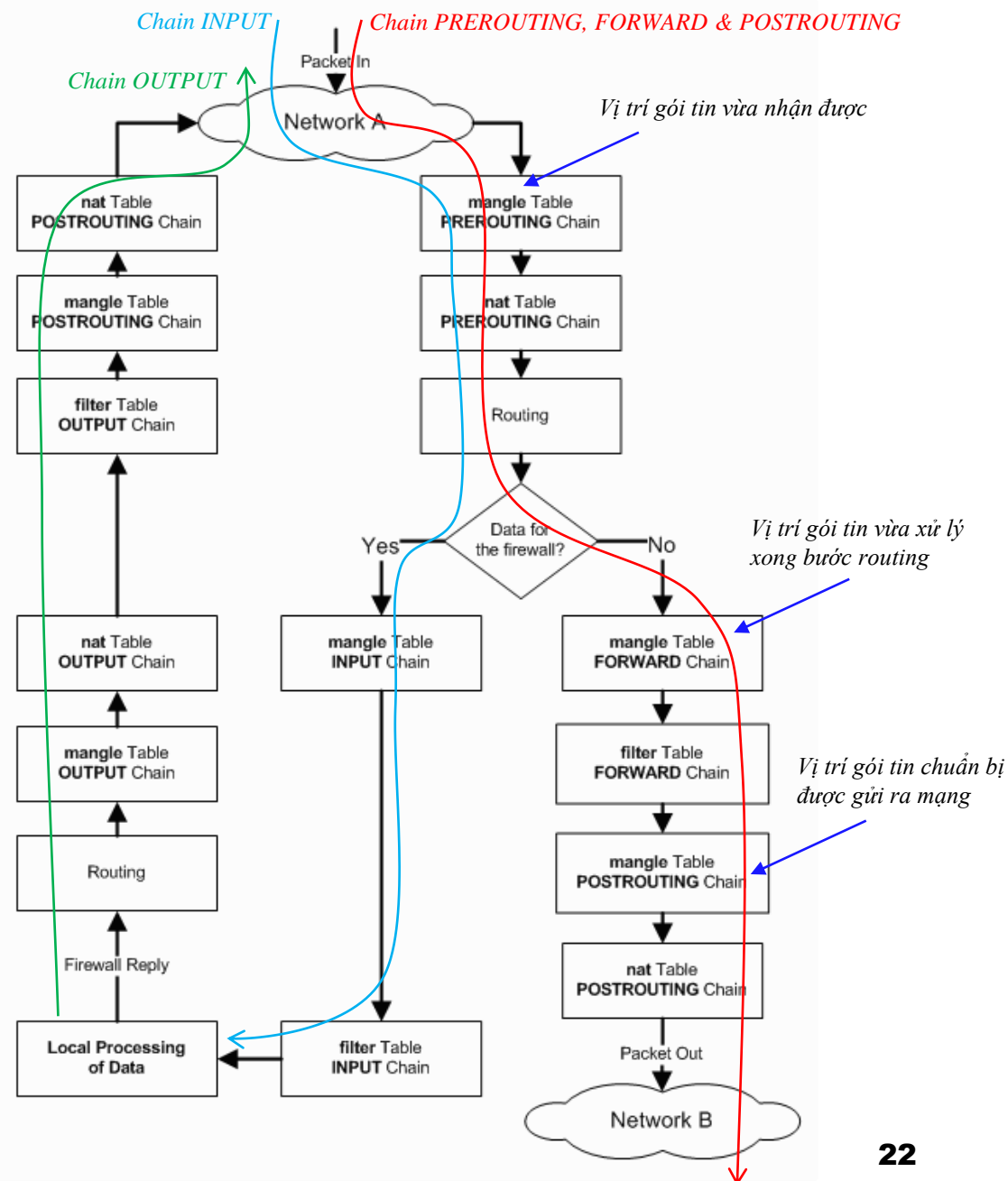
- R2 routing table:
 - 192.168.3.0/25 → 192.168.5.2
 - 192.168.3.128/25 → 192.168.2.2
- A ping X: [192.168.1.20] ping 192.168.3.200
 - Thuật toán: tìm IPnet của dest (192.168.3.0/24) trong routing table
 - Khớp với cả 2 dòng
 - Áp dụng classless netmask theo từng dòng trong routing table (/25) → tìm được 1 dòng
- → Các classless netmask mỗi dòng trong các bảng routing phải không tạo ra lỗi (admin vẫn có thể tạo nhiều dòng để route đến 1 địa chỉ mạng)



Linux *iptables*

- Cài đặt sẵn trong kernel Linux, xử lý gói tin theo dòng (chain) đi vào card mạng #1 & đi ra card mạng #2
- Chain:
 - INPUT, OUTPUT
 - PREROUTING, FORWARD, POSTROUTING
- Các luật (rule) được khai báo tại các vị trí trên chain để áp dụng để xử lý gói tin
- Rule có thể là cấm (reject), ghi lại thông tin (log), sửa đổi địa chỉ IP – NAT, v.v...
- Các rule được gộp với nhau theo mục đích sử dụng tạo thành các bảng (table)
- Table:
 - filter
 - nat
 - mangle

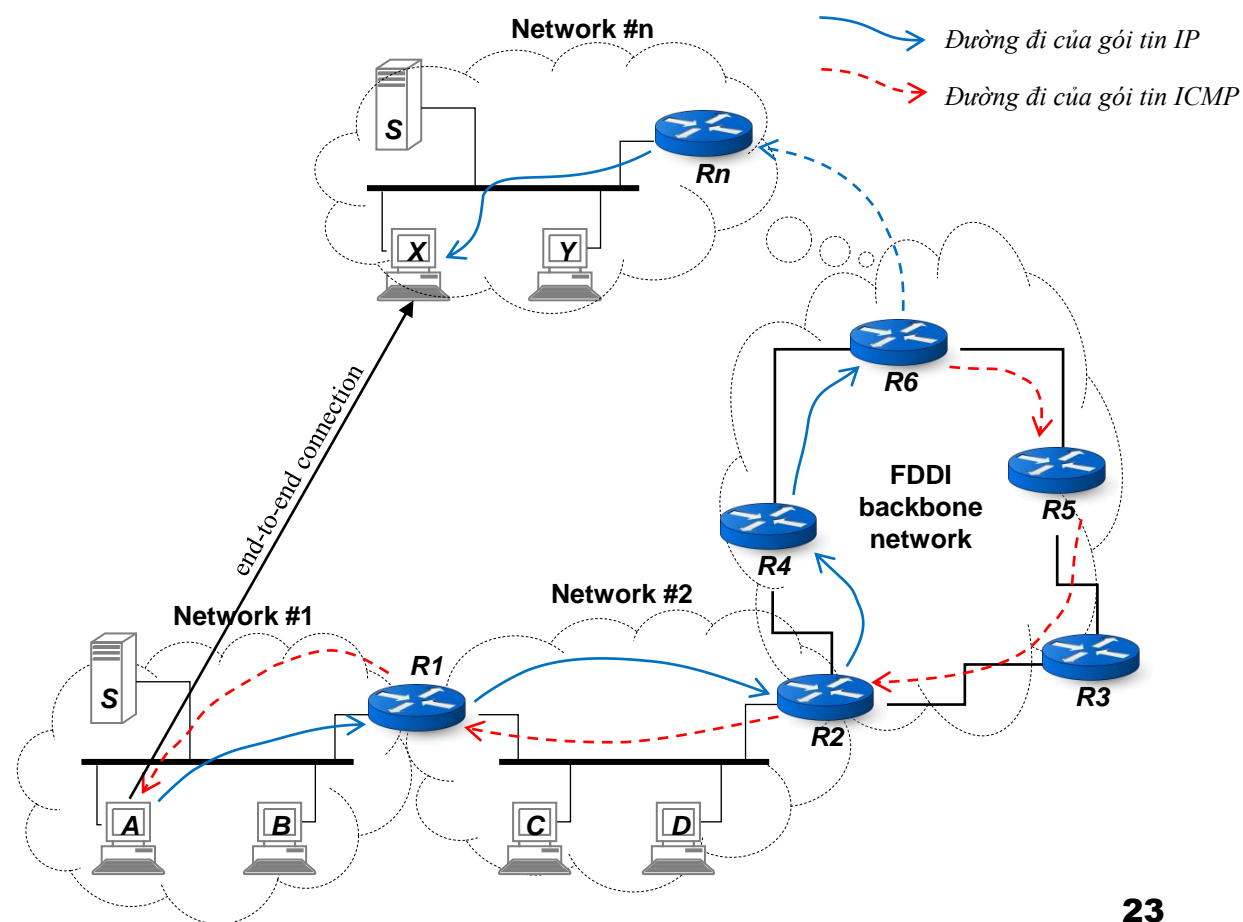
Kiểm tra log các gói tin ICMP đi qua router
> `iptables -t mangle -A FORWARD -j LOG`
> `tail -f /var/log/message | grep ICMP`



IP & ICMP

- IP: truyền dữ liệu không kết nối (connectionless), không tin cậy, nhiều khả năng không đến được đích
- ICMP
 - Type & Code
 - Hỗ trợ IP trong một số trường hợp để thông báo lỗi truyền gói IP
 - ICMP hoạt động trên IP (giống TCP/UDP) – ICMP được đóng gói trong IP ở phần payload data
- Do cấu hình routing table các router
→ đường đi 1 chiều thông không đảm bảo chiều ngược lại thông
- Nhiều trường hợp ICMP gửi về nhưng bị mất giữa chừng → trạm truyền không nhận được
- Ping:
 - Request timed out
 - Destination host unreachable
 - Destination prohibited
 - Transmit failed, error code #

	Bits 0–7	Bits 8–15	Bits 16–23	Bits 24–31
Header (20 bytes)	Version/IHL	Type of service	Length	
	Identification		flags and offset	
	Time To Live (TTL)	Protocol	Header Checksum	
	Source IP address			
	Destination IP address			
ICMP Header (8 bytes)	Type of message	Code	Checksum	
	Header Data			
ICMP Payload (optional)	Payload Data			

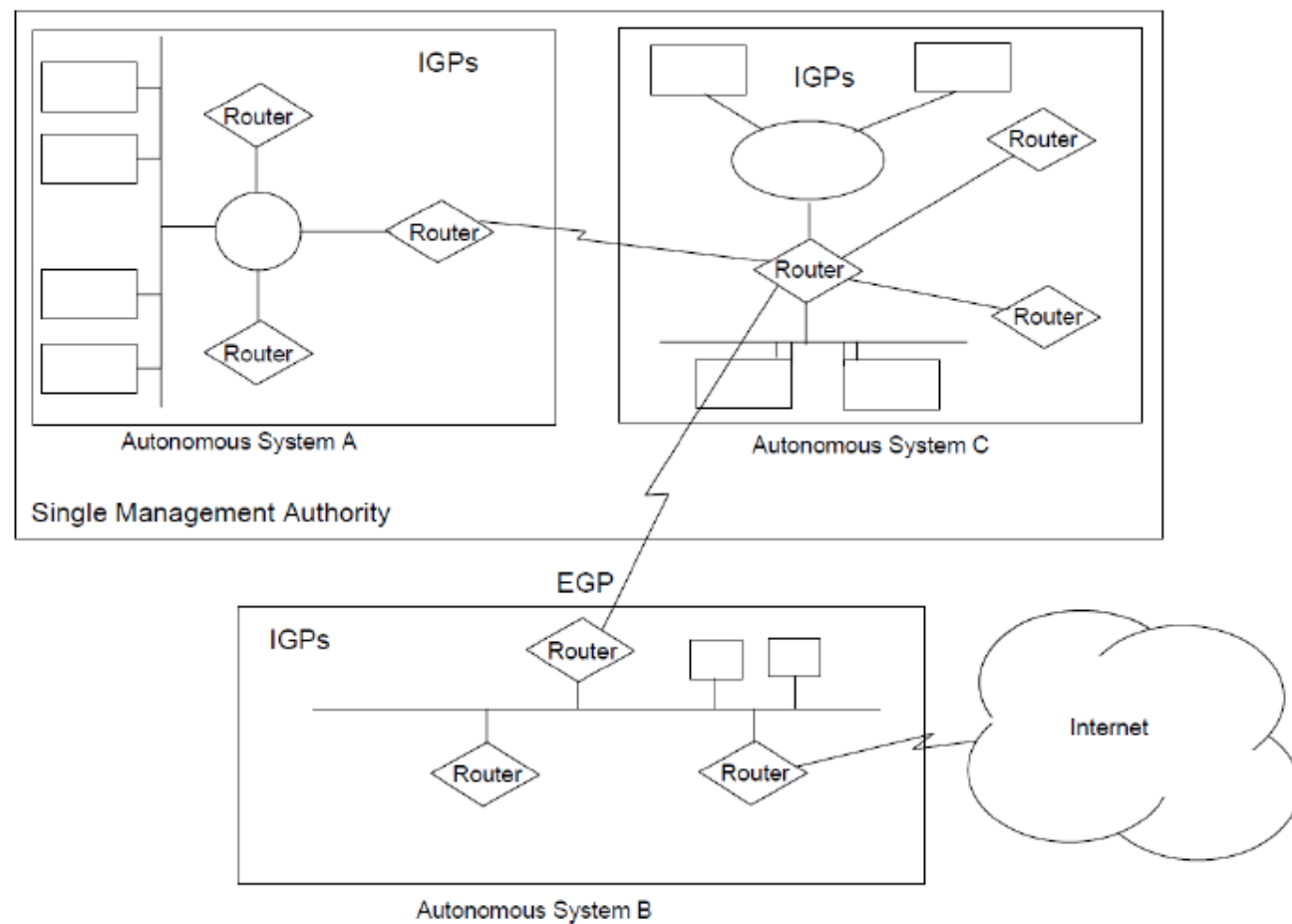


ICMP Type & Code

<i>TYPE</i>	<i>CODE</i>	Mô tả
0: Echo Reply	0	Sử dụng để cài đặt chương trình <i>ping</i> . Đây là thông điệp trả lời cho thông điệp ICMP số 8 (Echo Request)
3: Destination Unreachable	0	Thông báo lỗi không tìm được mạng đích
	1	Thông báo lỗi không tìm được máy đích
	2	Thông báo lỗi không tìm được protocol đích
	3	Thông báo lỗi không tìm được cổng đích
8: Echo Request	0	Sử dụng để cài đặt chương trình <i>ping</i> . Yêu cầu trả lời bằng thông điệp ICMP số 0 (Echo reply)
11: Time Exceeded	0	Thông báo lỗi hết thời gian TTL của gói tin IP
	1	Thông báo lỗi hết thời gian khi thực hiện ghép mảnh

Autonomous System

- Khái niệm quản lý nhưng ảnh hưởng đến hành vi của các thiết bị định tuyến
- Hệ thống tự trị (Autonomous System – AS [1]) là tập hợp kết nối một số mạng IP mà được quản lý định tuyến dưới sự kiểm soát của một thực thể hành chính
- Do được quản lý chung bởi một tổ chức, cấu hình và sơ đồ kết nối mạng trong một AS là xác định
- Tổ chức quản lý một AS không nắm được qui hoạch mạng của một AS khác
- Internal Gateway Protocols (IGPs): phương pháp định tuyến nội bộ, dựa trên các thông tin qui hoạch mạng tổng thể của một AS
- Exterior Gateway Protocols (EGPs): phương pháp định tuyến kết nối giữa các AS



[1] [https://en.wikipedia.org/wiki/Autonomous_system_\(Internet\)](https://en.wikipedia.org/wiki/Autonomous_system_(Internet))

Các phương pháp định tuyến

- Tĩnh (static routing): các đường định tuyến (giữa các mạng nghiệp vụ) được xác định sẵn (theo qui hoạch mạng tổng thể của AS) và được nhân viên quản trị mạng cấu hình sẵn trong các bảng routing.
- Động (dynamic routing): các thuật toán cài đặt trong router phép các chúng liên lạc với nhau để tự động phát hiện và duy trì thông tin về các đường định tuyến.
- Định tuyến động có thể được áp dụng bên trong một AS hoặc giữa các AS (cần chú ý vấn đề bảo mật)



Routing Information Protocol (RIP)

RIP¹ Introduction

- Phương pháp véc tơ khoảng cách (distance vector): số đo khoảng cách từ vị trí của mình đến mỗi điểm đến (mạng đích) – số router trung gian trên đường truyền (giống khái niệm số đo TTL trong gói IP)
- Router sử dụng cột Metric trong bảng routing để thể hiện khoảng cách để đi từ nó tại đến mạng đích tương ứng
- Khoảng cách càng ngắn
→ đường đi càng nhanh (tương đối)
- Giải thuật routing:
tìm thấy nhiều đường đi có thể →
chọn đường đi ngắn nhất (giá trị Metric bé nhất)
- Router được cài đặt giao thức RIP để tự xây dựng và duy trì (cập nhật) bảng routing với thông số Metric
- Các router trao đổi thông tin theo kiểu lan truyền, mỗi router chỉ có thể kết nối với router láng giềng (neighbor) – có kế nối mạng trực tiếp

```
> route -n
```

Kernel IP routing table				
Destination	Gateway	Genmask	Flags	Metric
192.168.2.0	0.0.0.0	255.255.255.128	U	0
192.168.2.128	192.168.2.2	255.255.255.128	UG	2
192.168.3.0	192.168.2.2	255.255.255.0	UG	3
192.168.1.0	0.0.0.0	255.255.255.0	U	0

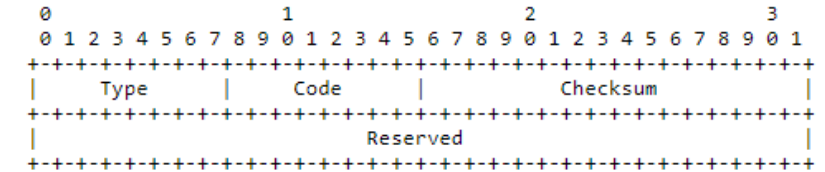
[1] https://en.wikipedia.org/wiki/Routing_Information_Protocol

ICMP Router Discovery Protocol

- ICMP Internet Router Discovery Protocol (IRDP) [1]
- Hỗ trợ các router (và các host) phát hiện router kết nối trực tiếp (neighbor)
- Sử dụng 2 thông điệp ICMP [2]:
 - ICMP Router Solicitation:
 - type = 10
 - Gửi từ host để tìm router
 - ICMP Router Advertisement
 - type = 9
 - Gửi từ router để thông báo sự có mặt của mình
- Gửi đi đâu:
 - Solicitation: Multicast/Broadcast
 - Advertisement: Multicast/Broadcast/Unicast
- Gửi khi nào:
 - Khi bắt đầu kết nối
 - Theo chu kỳ refresh dữ liệu

Sử dụng iptables xem log các gói tin ICMP khi router liên lạc với nhau bằng RIP

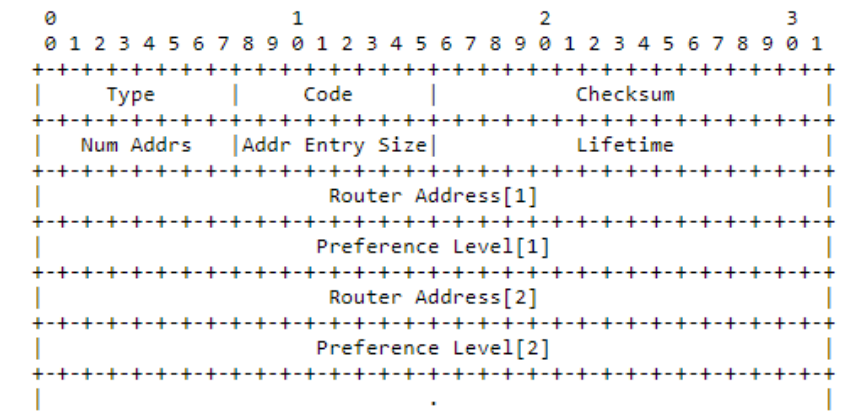
ICMP Router Solicitation Message



IP Fields:

- Source Address: An IP address belonging to the interface from which this message is sent, or 0.
- Destination Address: The configured SolicitationAddress.
- Time-to-Live: 1 if the Destination Address is an IP multicast address; at least 1 otherwise.

ICMP Router Advertisement Message



IP Fields:

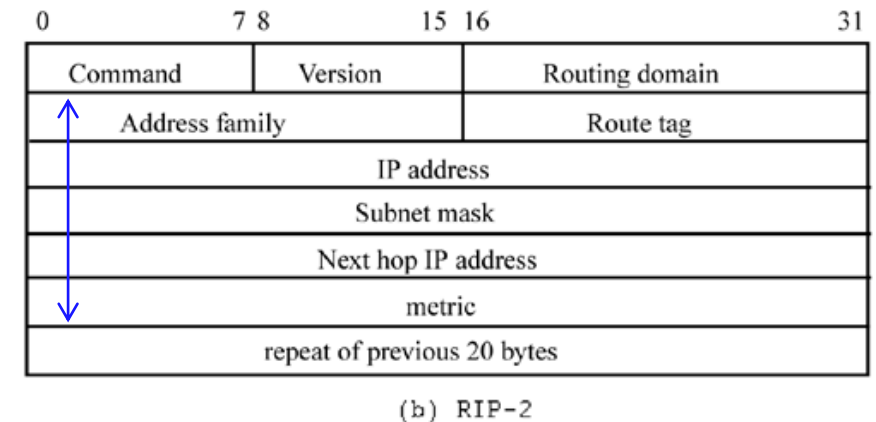
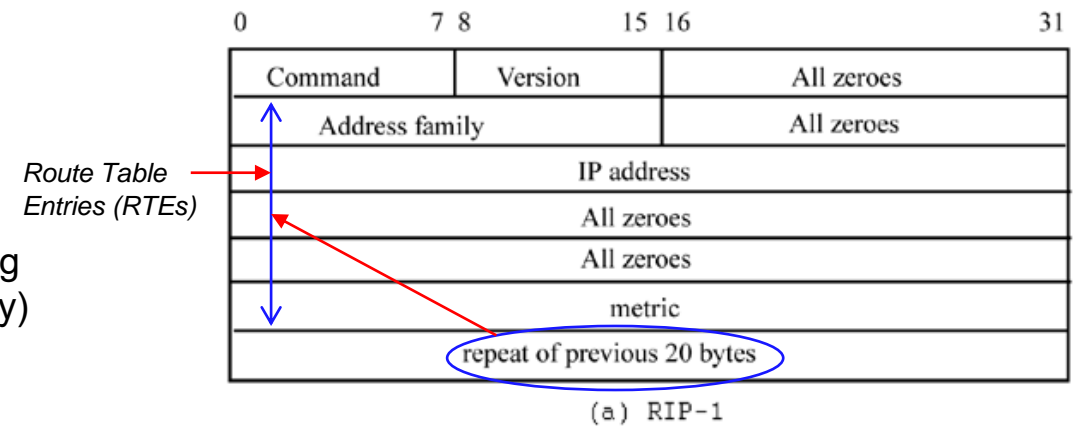
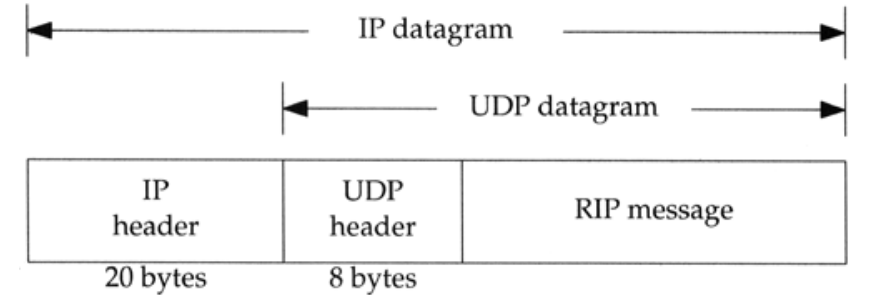
- Source Address: An IP address belonging to the interface from which this message is sent.
- Destination Address: The configured AdvertisementAddress or the IP address of a neighboring host.
- Time-to-Live : 1 if the Destination Address is an IP multicast address; at least 1 otherwise.

[1] https://en.wikipedia.org/wiki/ICMP_Router_Discovery_Protocol

[2] <https://tools.ietf.org/html/rfc1256>

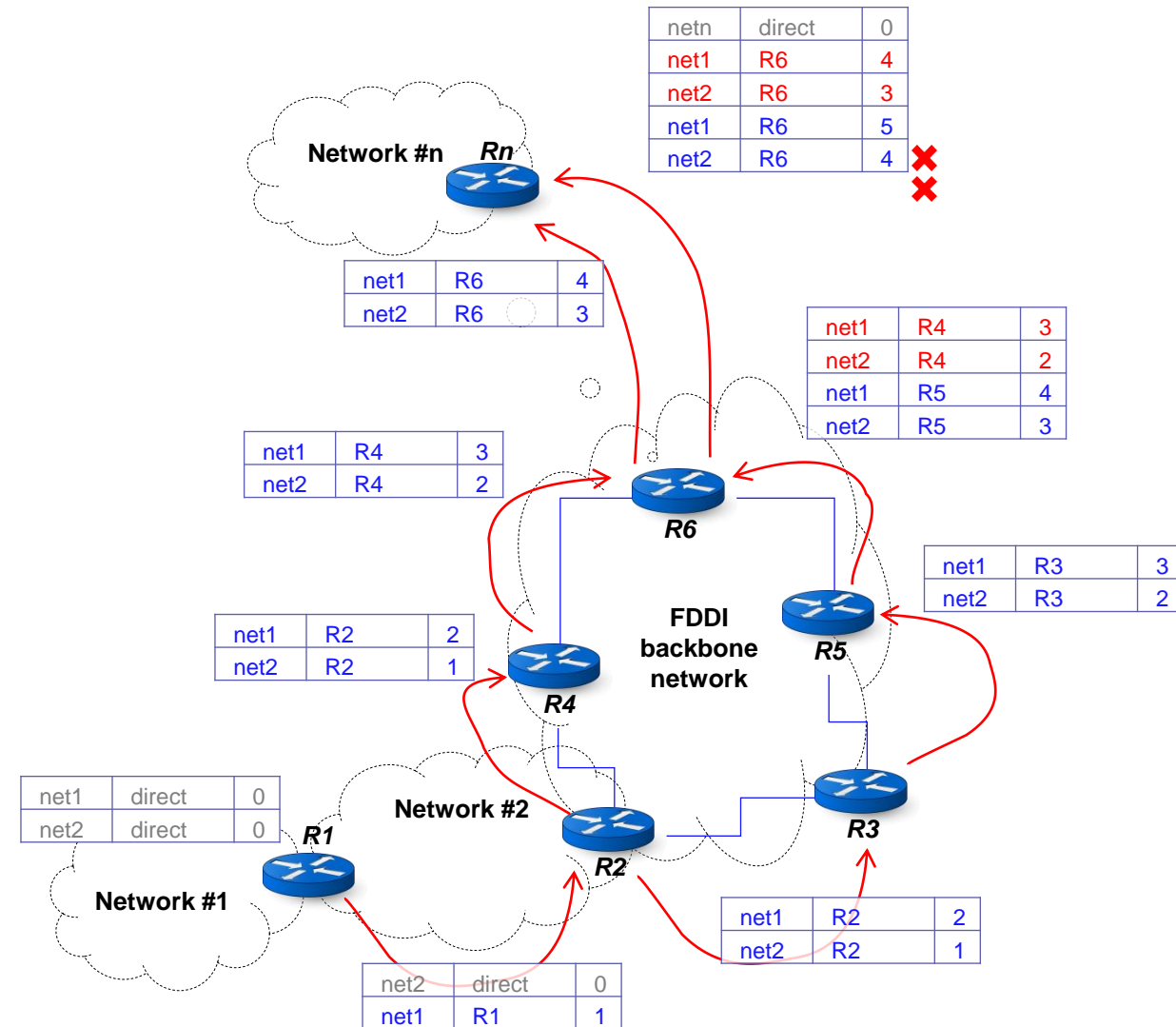
RIP Message Format

- RIPv1 (1988), RIPv2 (1993), RIPv6 (1997 – IPv6)
- Cài đặt trên UDP cổng 520
- RIPv1 sử dụng broadcast, RIPv2 sử dụng multicast 224.0.0.9 để yêu cầu cập nhật thông tin từ router láng giềng.
- 2 loại message:
 - Request: yêu cầu router neighbor gửi routing table.
 - Response: gửi routing table của mình cho router neighbor. Mỗi dòng trong bảng routing được đóng gói bằng một RTE (Route Table Entry) trong gói tin RIP
- RIPv1 chỉ hỗ trợ classfull IP address, RIPv2 hỗ trợ classless bằng trường *subnet mask*



Hoạt động của RIP

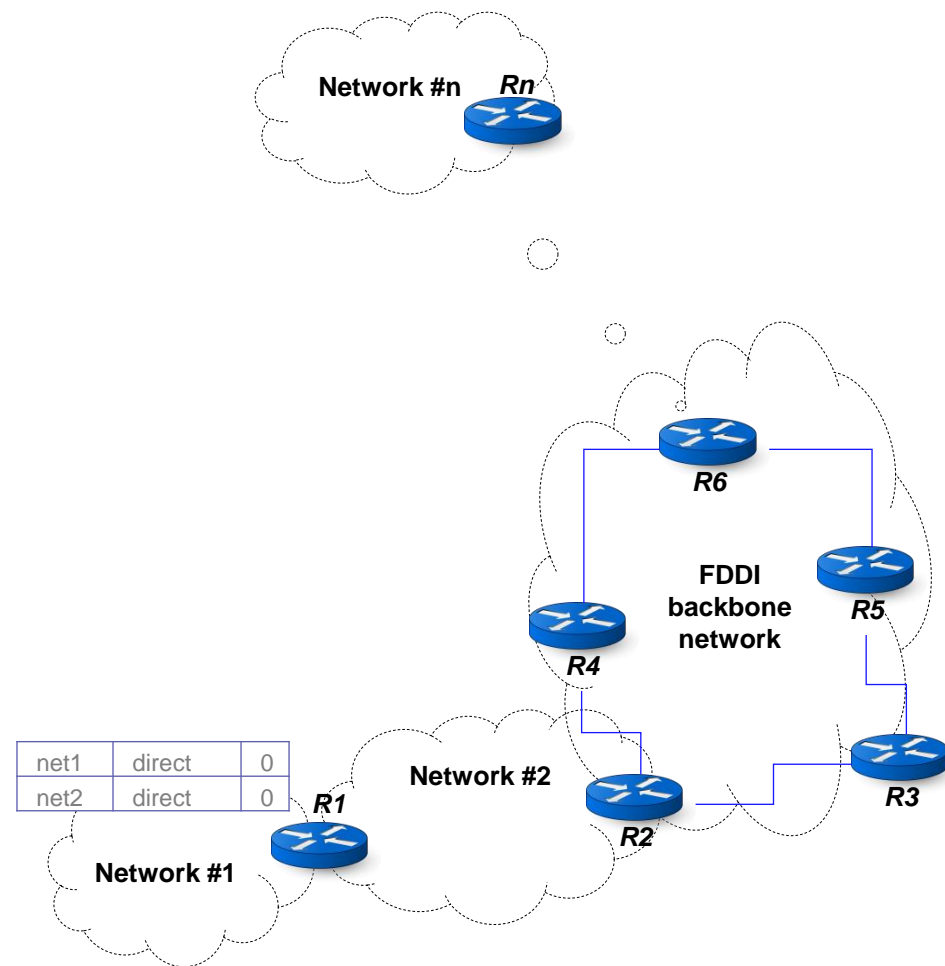
- Sử dụng Router Discovery Protocol để xác định router láng giềng rồi gửi RIP request yêu cầu cập nhật bảng routing (thực tế có thể broadcast, bỏ qua bước Router Discovery)
- Router nhận được yêu cầu sẽ trả lời bằng RIP Response message với nội dung là toàn bộ bảng routing hiện tại của mình (thông qua các Route Table Entries - RTEs)
- Xử lý nhận được RIP Response:
 - Trích xuất từng RTE, so sánh với RTE đang có trong bảng routing của mình
 - Nếu đã có (trùng cả network & gateway), so sánh giá trị Metric để loại đi dòng có Metric cao hơn
 - Nếu chưa có, thêm dòng routing mới
 - Thiết lập gateway của dòng mới là router láng giềng đã gửi RTE
 - Thiết lập giá trị Metric của dòng mới = Metric (RTE) + 1
- Quá trình lan tỏa bảng routing thông qua RIP message sẽ đi đến một điểm hội tụ (convergence) mà tất cả các mạng nghiệp vụ đã được cập nhật vào bảng routing của tất cả router → hoàn thành.



RIP Demonstration

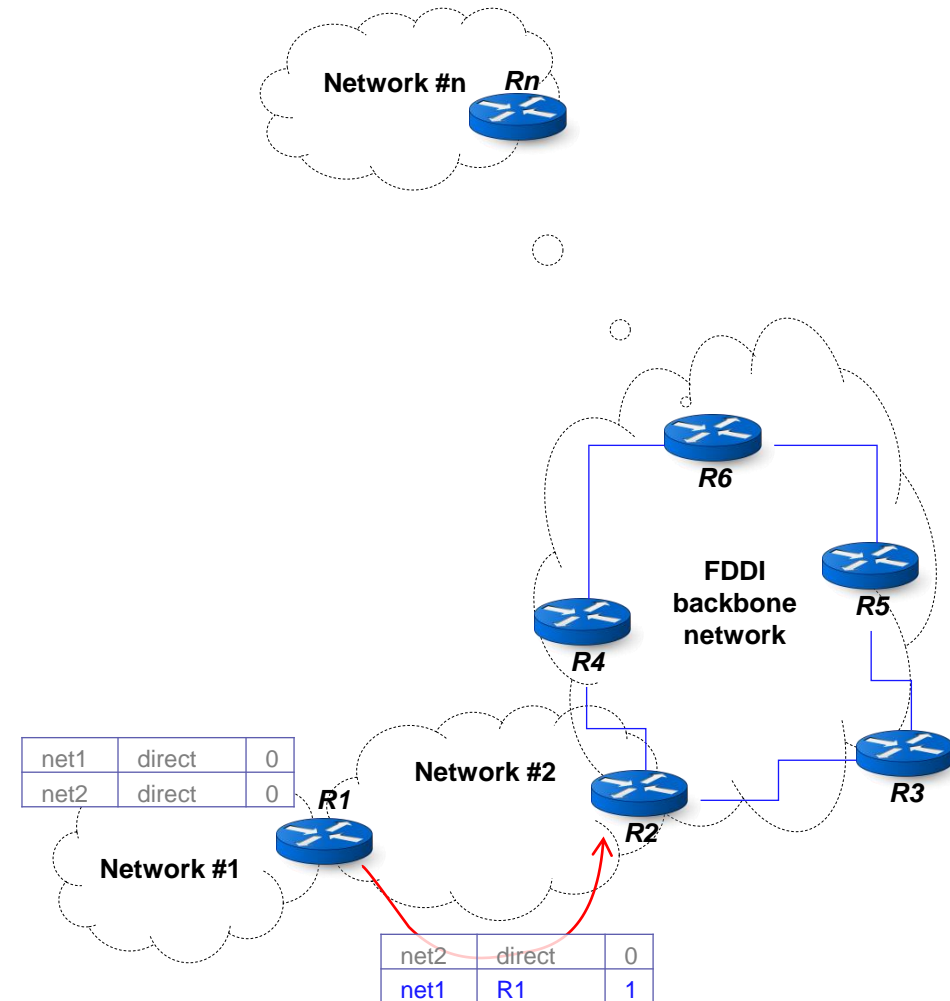
■ R1: tự động khởi tạo routing table:

- net1, direct, 0
- net2, direct, 0



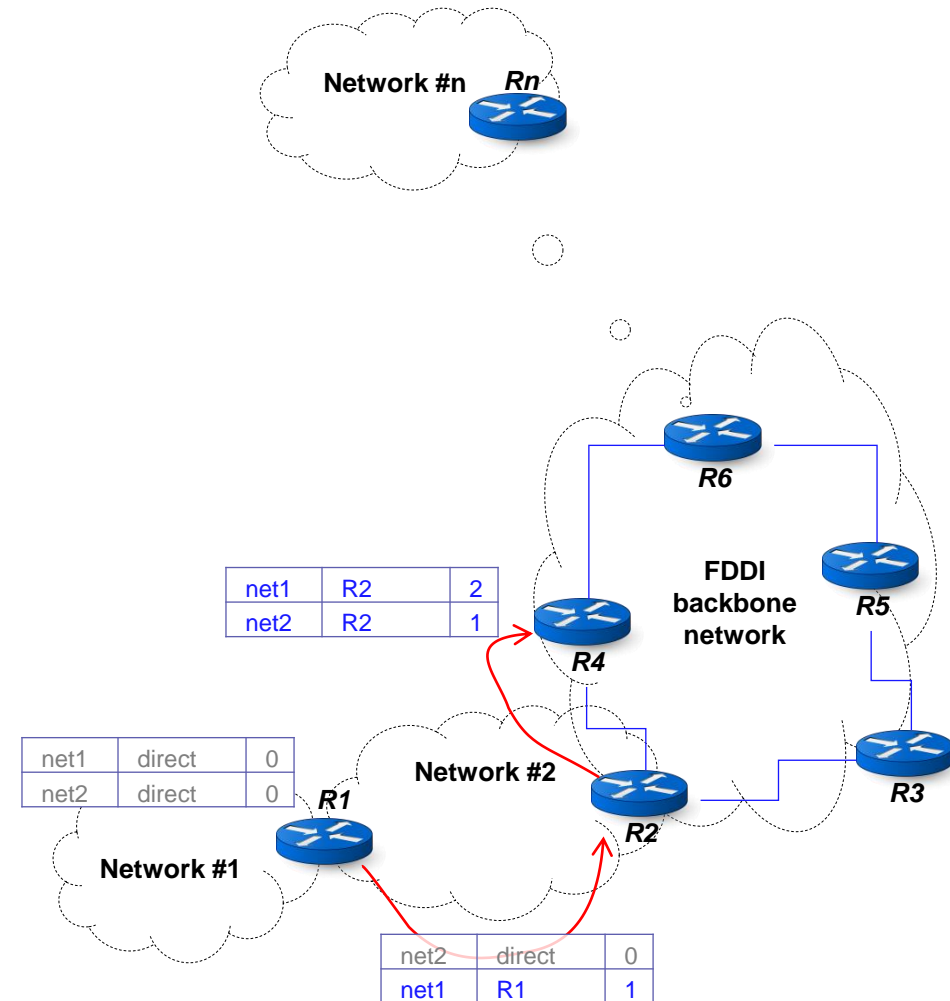
RIP Demonstration

- R1: tự động khởi tạo routing table:
 - net1, direct, 0
 - net2, direct, 0
- R1 → R2:
 - RTE: net1, direct, 0 → (+) net1, R1, 0+1
 - RTE: net2, direct, 0 → (-) net2, direct, 0



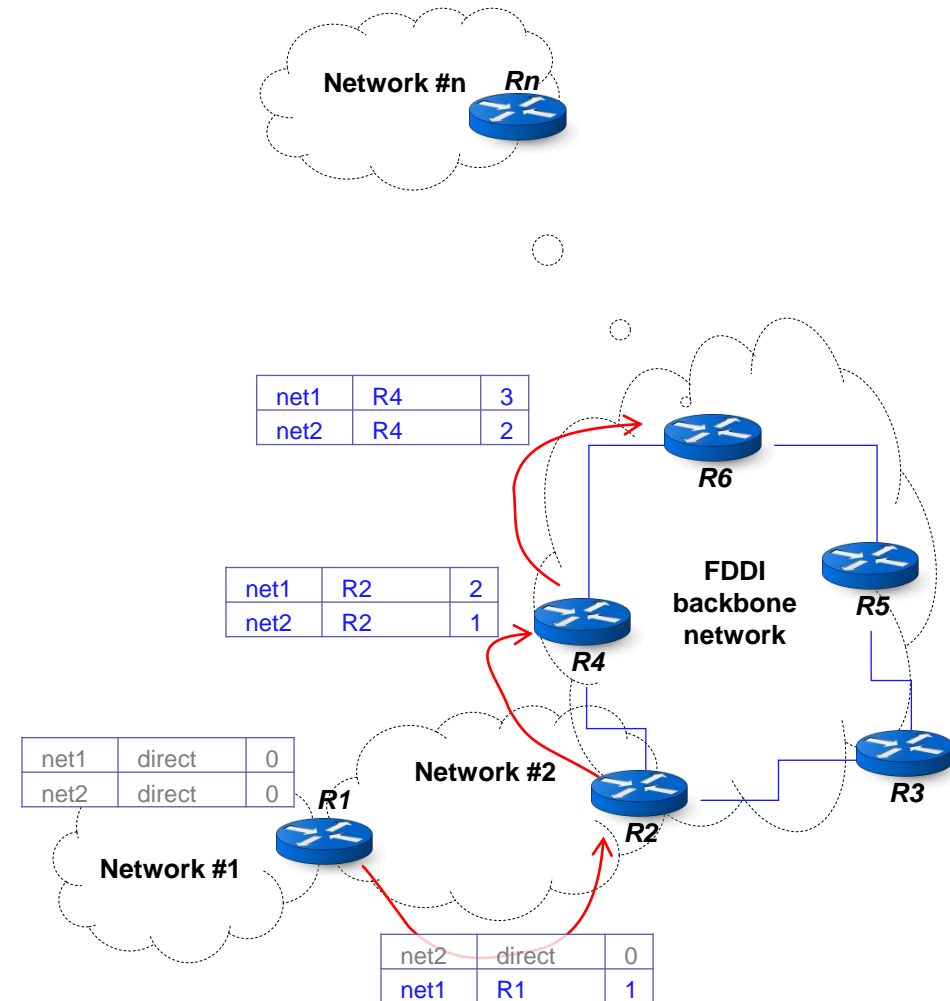
RIP Demonstration

- R1: tự động khởi tạo routing table:
 - net1, direct, 0
 - net2, direct, 0
- R1 → R2:
 - RTE: net1, direct, 0 → (+) net1, R1, 0+1
 - RTE: net2, direct, 0 → (-) net2, direct, 0
- R2 → R4:
 - RTE: net1, R1, 1 → (+) net1, R2, 1+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1



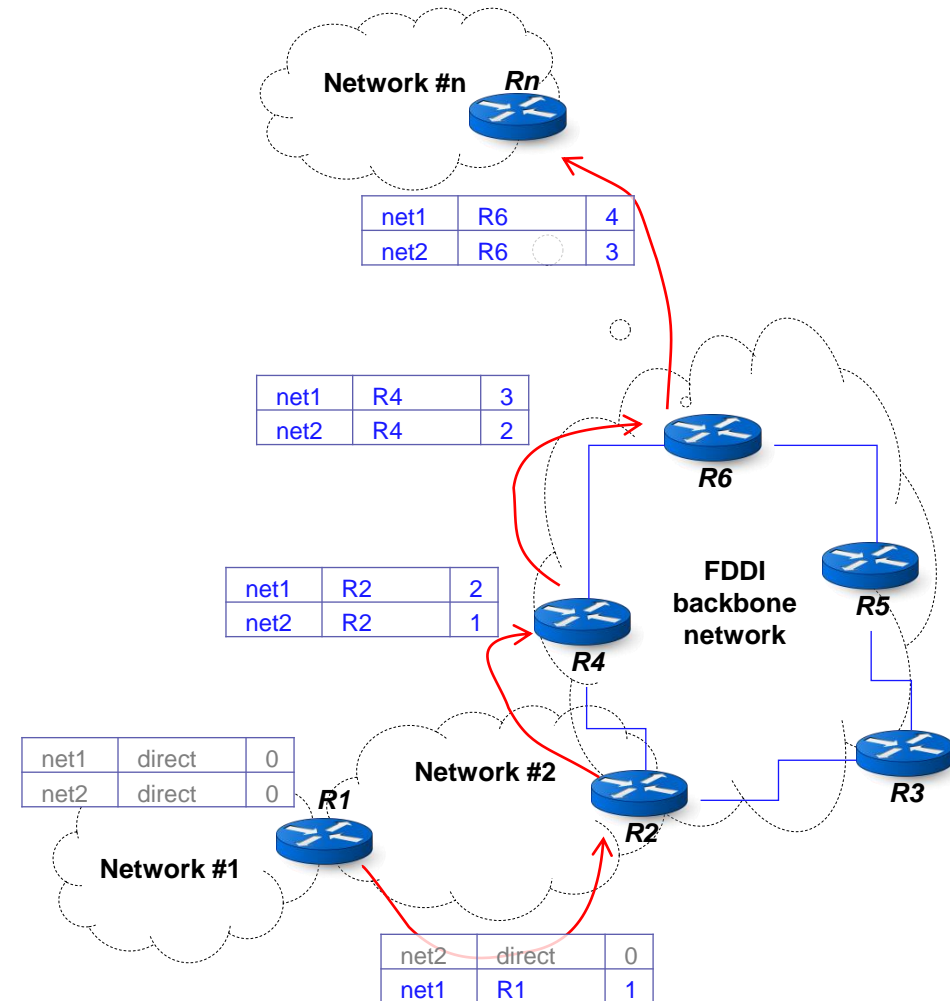
RIP Demonstration

- R1: tự động khởi tạo routing table:
 - net1, direct, 0
 - net2, direct, 0
- R1 → R2:
 - RTE: net1, direct, 0 → (+) net1, R1, 0+1
 - RTE: net2, direct, 0 → (-) net2, direct, 0
- R2 → R4:
 - RTE: net1, R1, 1 → (+) net1, R2, 1+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1
- R4 → R6:
 - RTE: net1, R2, 2 → (+) net1, R4, 2+1
 - RTE: net2, R2, 1 → (+) net2, R4, 1+1



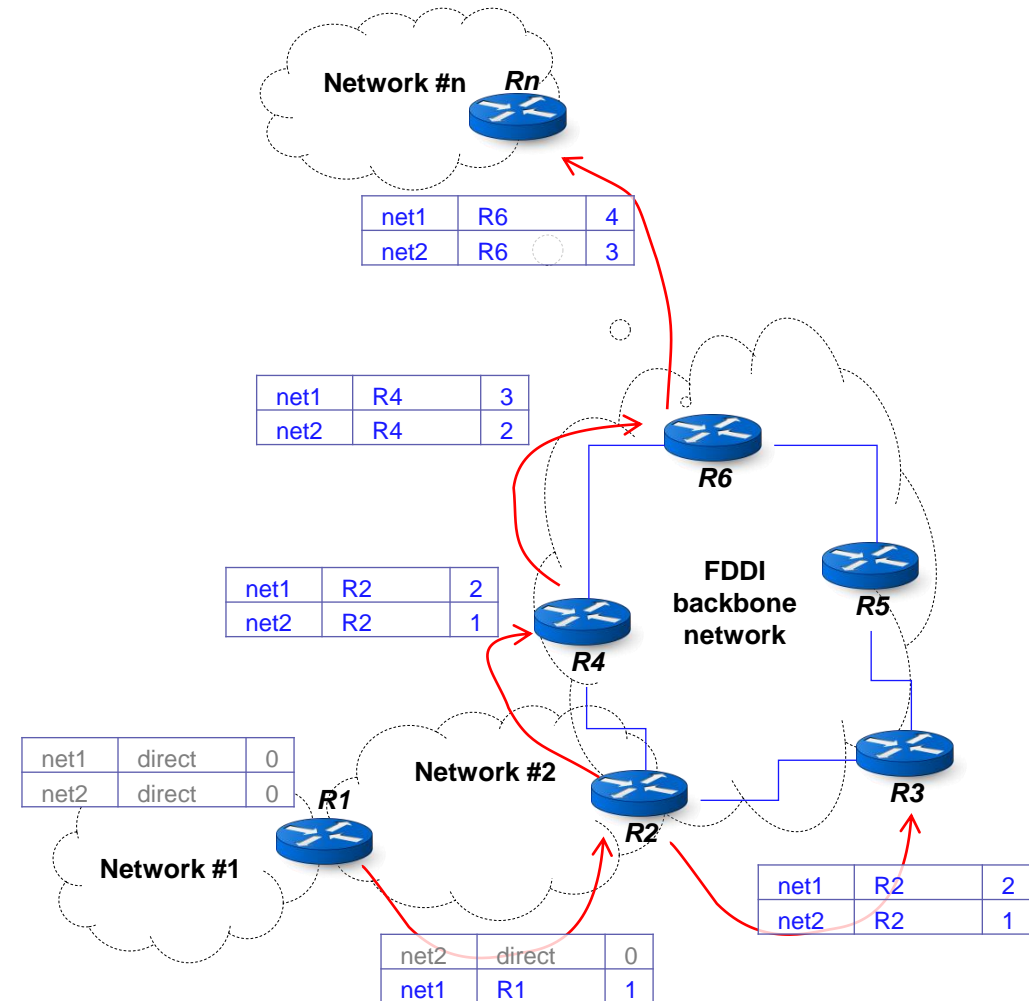
RIP Demonstration

- R1: tự động khởi tạo routing table:
 - net1, direct, 0
 - net2, direct, 0
- R1 → R2:
 - RTE: net1, direct, 0 → (+) net1, R1, 0+1
 - RTE: net2, direct, 0 → (-) net2, direct, 0
- R2 → R4:
 - RTE: net1, R1, 1 → (+) net1, R2, 1+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1
- R4 → R6:
 - RTE: net1, R2, 2 → (+) net1, R4, 2+1
 - RTE: net2, R2, 1 → (+) net2, R4, 1+1
- R6 → Rn:
 - RTE: net1, R4, 3 → (+) net1, R6, 3+1
 - RTE: net2, R4, 2 → (+) net2, R6, 2+1



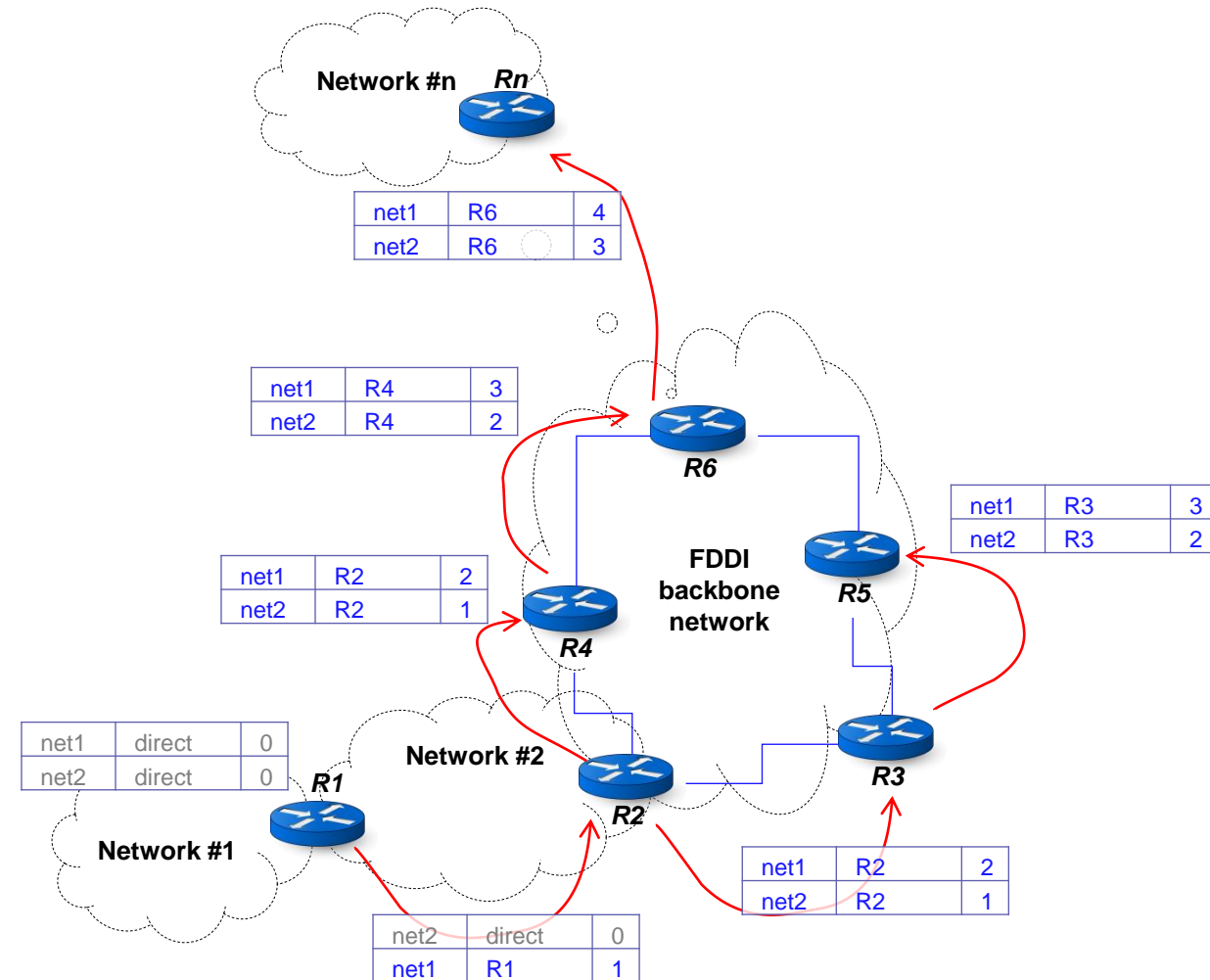
RIP Demonstration

- R1: tự động khởi tạo routing table:
 - net1, direct, 0
 - net2, direct, 0
- R1 → R2:
 - RTE: net1, direct, 0 → (+) net1, R1, 0+1
 - RTE: net2, direct, 0 → (-) net2, direct, 0
- R2 → R4:
 - RTE: net1, R1, 1 → (+) net1, R2, 1+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1
- R4 → R6:
 - RTE: net1, R2, 2 → (+) net1, R4, 2+1
 - RTE: net2, R2, 1 → (+) net2, R4, 1+1
- R6 → Rn:
 - RTE: net1, R4, 3 → (+) net1, R6, 3+1
 - RTE: net2, R4, 2 → (+) net2, R6, 2+1
- R2 → R3:
 - RTE: net1, direct, 0 → (+) net1, R2, 0+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1



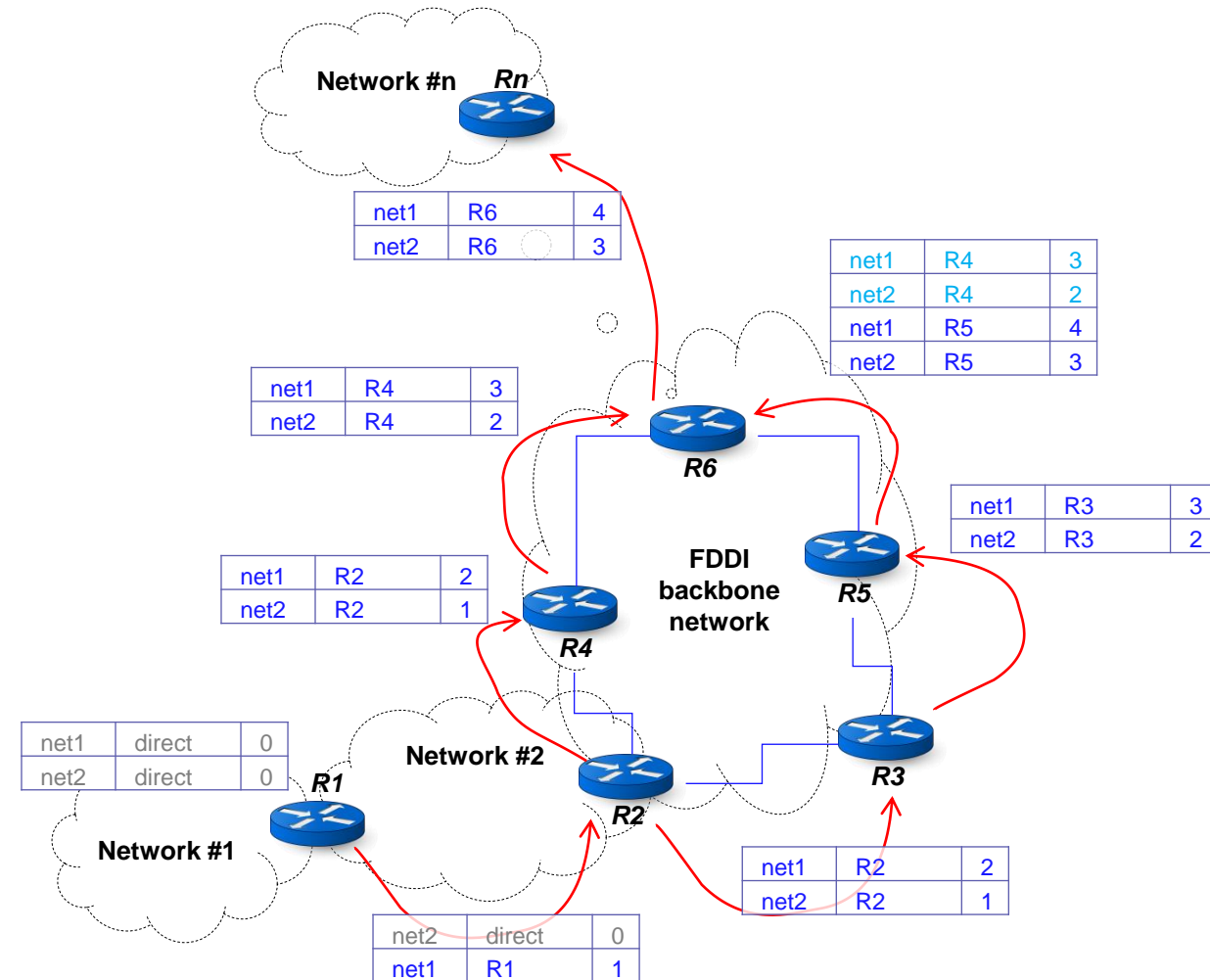
RIP Demonstration

- R1: tự động khởi tạo routing table:
 - net1, direct, 0
 - net2, direct, 0
- R1 → R2:
 - RTE: net1, direct, 0 → (+) net1, R1, 0+1
 - RTE: net2, direct, 0 → (-) net2, direct, 0
- R2 → R4:
 - RTE: net1, R1, 1 → (+) net1, R2, 1+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1
- R4 → R6:
 - RTE: net1, R2, 2 → (+) net1, R4, 2+1
 - RTE: net2, R2, 1 → (+) net2, R4, 1+1
- R6 → Rn:
 - RTE: net1, R4, 3 → (+) net1, R6, 3+1
 - RTE: net2, R4, 2 → (+) net2, R6, 2+1
- R2 → R3:
 - RTE: net1, R1, 1 → (+) net1, R2, 1+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1
- R3 → R5:
 - RTE: net1, R2, 2 → (+) net1, R3, 2+1
 - RTE: net2, R2, 1 → (+) net2, R3, 1+1



RIP Demonstration

- R1: tự động khởi tạo routing table:
 - net1, direct, 0
 - net2, direct, 0
- R1 → R2:
 - RTE: net1, direct, 0 → (+) net1, R1, 0+1
 - RTE: net2, direct, 0 → (-) net2, direct, 0
- R2 → R4:
 - RTE: net1, R1, 1 → (+) net1, R2, 1+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1
- R4 → R6:
 - RTE: net1, R2, 2 → (+) net1, R4, 2+1
 - RTE: net2, R2, 1 → (+) net2, R4, 1+1
- R6 → Rn:
 - RTE: net1, R4, 3 → (+) net1, R6, 3+1
 - RTE: net2, R4, 2 → (+) net2, R6, 2+1
- R2 → R3:
 - RTE: net1, R1, 1 → (+) net1, R2, 1+1
 - RTE: net2, direct, 0 → (+) net2, R2, 0+1
- R3 → R5:
 - RTE: net1, R2, 2 → (+) net1, R3, 2+1
 - RTE: net2, R2, 1 → (+) net2, R3, 1+1
- R5 → R6:
 - RTE: net1, R3, 3 → (+) net1, R5, 3+1
 - RTE: net2, R3, 2 → (+) net2, R5, 2+1



RIP Demonstration

■ R1: tự động khởi tạo routing table:

- net1, direct, 0
- net2, direct, 0

■ R1 → R2:

- RTE: net1, direct, 0 → (+) net1, R1, 0+1
- RTE: net2, direct, 0 → (-) net2, direct, 0

■ R2 → R4:

- RTE: net1, R1, 1 → (+) net1, R2, 1+1
- RTE: net2, direct, 0 → (+) net2, R2, 0+1

■ R4 → R6:

- RTE: net1, R2, 2 → (+) net1, R4, 2+1
- RTE: net2, R2, 1 → (+) net2, R4, 1+1

■ R6 → Rn:

- RTE: net1, R4, 3 → (+) net1, R6, 3+1
- RTE: net2, R4, 2 → (+) net2, R6, 2+1

■ R2 → R3:

- RTE: net1, R1, 1 → (+) net1, R2, 1+1
- RTE: net2, direct, 0 → (+) net2, R2, 0+1

■ R3 → R5:

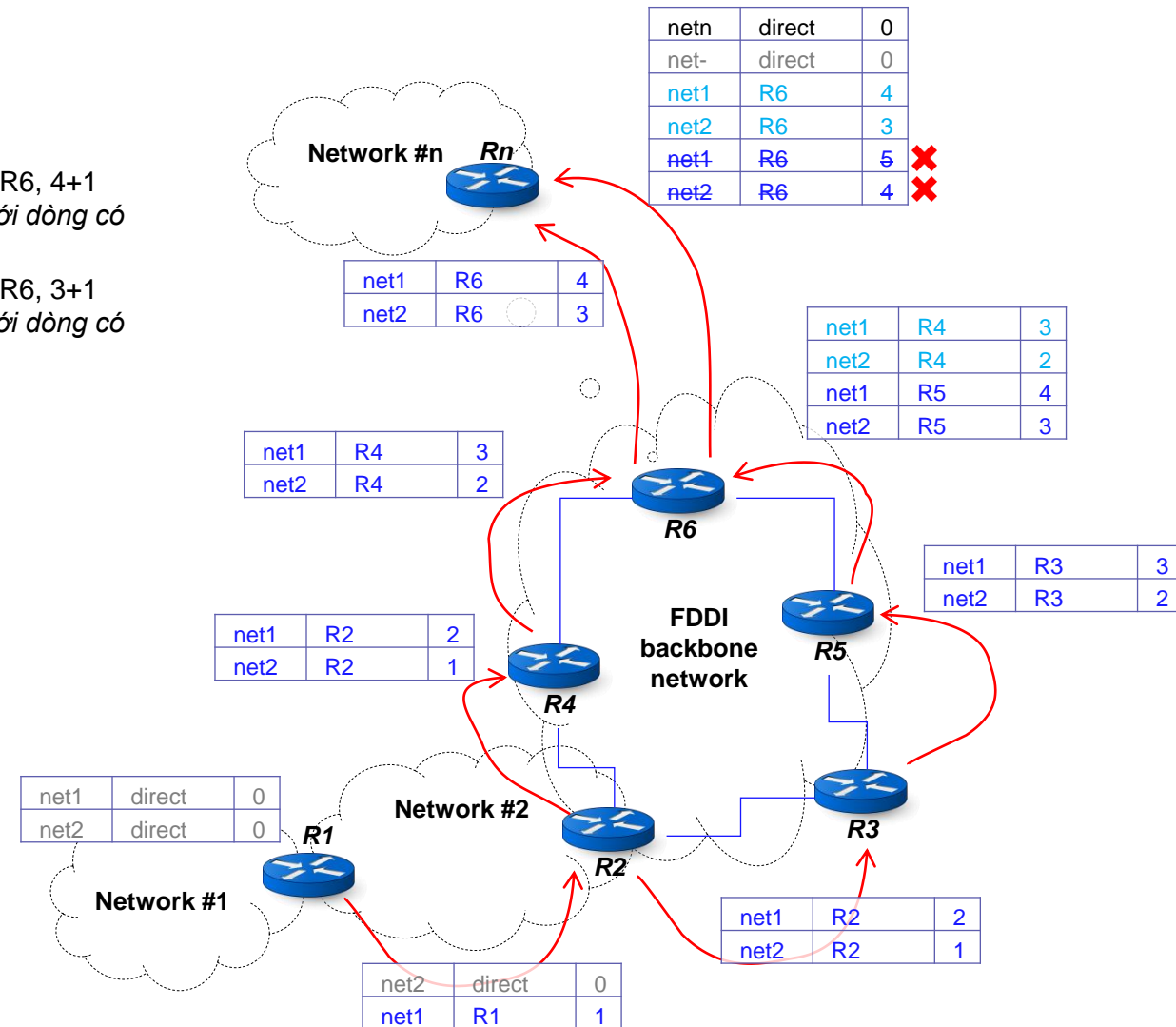
- RTE: net1, R2, 2 → (+) net1, R3, 2+1
- RTE: net2, R2, 1 → (+) net2, R3, 1+1

■ R5 → R6:

- RTE: net1, R3, 3 → (+) net1, R5, 3+1
- RTE: net2, R3, 2 → (+) net2, R5, 2+1

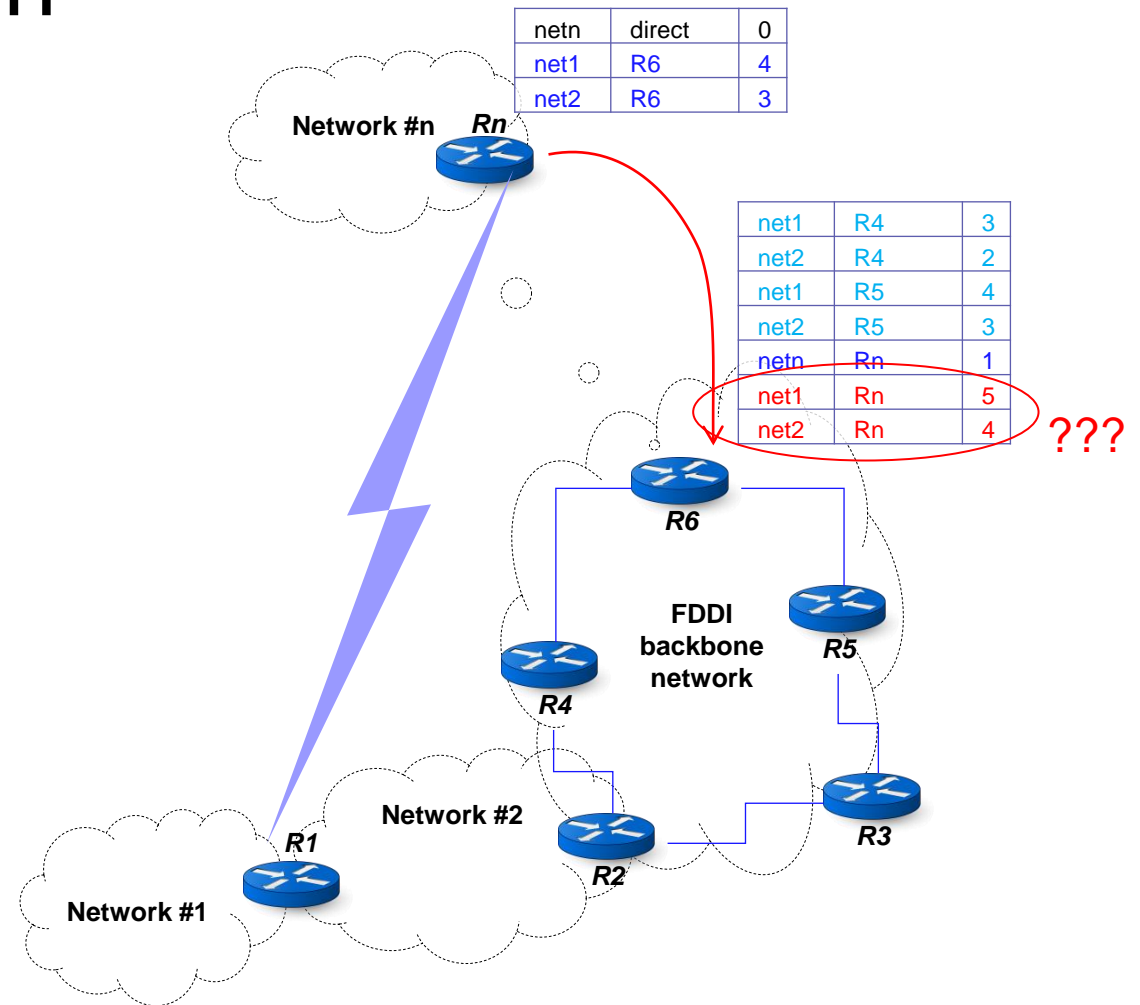
■ R6:

- net1, R5, 4 → (-) net1, R6, 4+1
(trùng net & gateway với dòng có metric bé hơn)
- net2, R5, 3 → (-) net2, R6, 3+1
(trùng net & gateway với dòng có metric bé hơn)



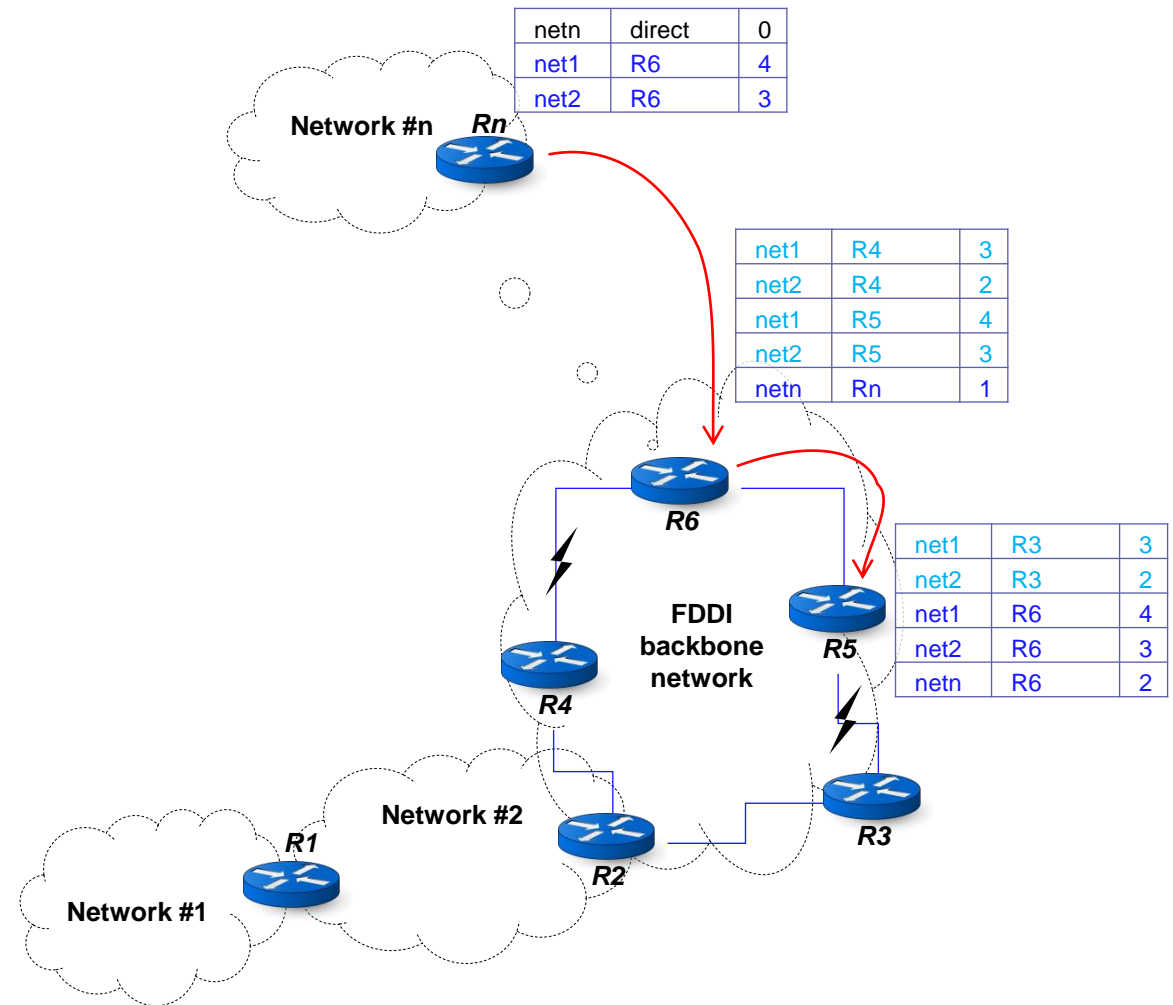
Loop Routing & Split Horizon

- Rn:
 - netn, direct, 0
 - net1, R6, 4
 - net2, R6, 3
- R6 → Rn:
 - RTE: netn, direct, 0 → (+) netn, R6, 0+1
 - RTE: net1, R6, 4 → (?) net1, Rn, 4+1
 - RTE: net2, R6, 3 → (?) net2, Rn, 3+1
- Khả năng có thể có một kết nối khác từ Rn đến network #1 (mà không phải đi qua R6), và cần duy trì nhiều đường đi để backup khi có sự cố mạng → R6 chấp nhận 2 RTE đi đến net1 & net2 được gửi đến từ Rn
- Tiềm ẩn vòng lặp routing giữa Rn & R6:
 - Gói tin cần gửi từ Rn đến net1 hoặc net2
 - Rn → R6
 - Lựa chọn Metric bé nhất, R6 → R4
 - Kết nối đến R4 và R5 có lỗi, R6 cập nhật trạng thái → còn duy nhất các RTE net1 & net2 đi qua Rn → loop routing giữa R6 & Rn
- Giải pháp “Split Horizon” [1]: không gửi RTE mà đã nhận từ chính router láng giềng này



Route Poisoning

- Vòng lặp routing vẫn xuất hiện khi các router kết nối dạng loop
- Rn → R6 có xử lý “Split Horizon”: ok
- R6 → R5 có xử lý “Split Horizon”: tránh được các RTE đã gửi từ R5 nhưng vẫn chấp nhận các RTE gửi từ R4:
 - net1, R6, 4
 - net2, R6, 3
- Tình huống loop routing lại xuất hiện giữa R6 & R5 đối với net1 & net2, khi link (R6,R4) và (R5,R3) cùng xảy ra sự cố
- Giải pháp “Route Poisoning” [1]: phát hiện link down, router gửi thông báo Metric = ∞ đến router láng giềng nhằm tránh route qua đường này:
 - Link (R4, R6) down. R6 → R5: (net1, R6, ∞)
 - R5 phải update RTE này (ngược với qui tắc lấy Metric bé !!!)
 - R5 không route gói tin net1 đến R6 nữa
 - Tương tự khi link (R3,R5) down, R6 cũng sẽ không route gói tin net1, net2 qua R5 nữa.



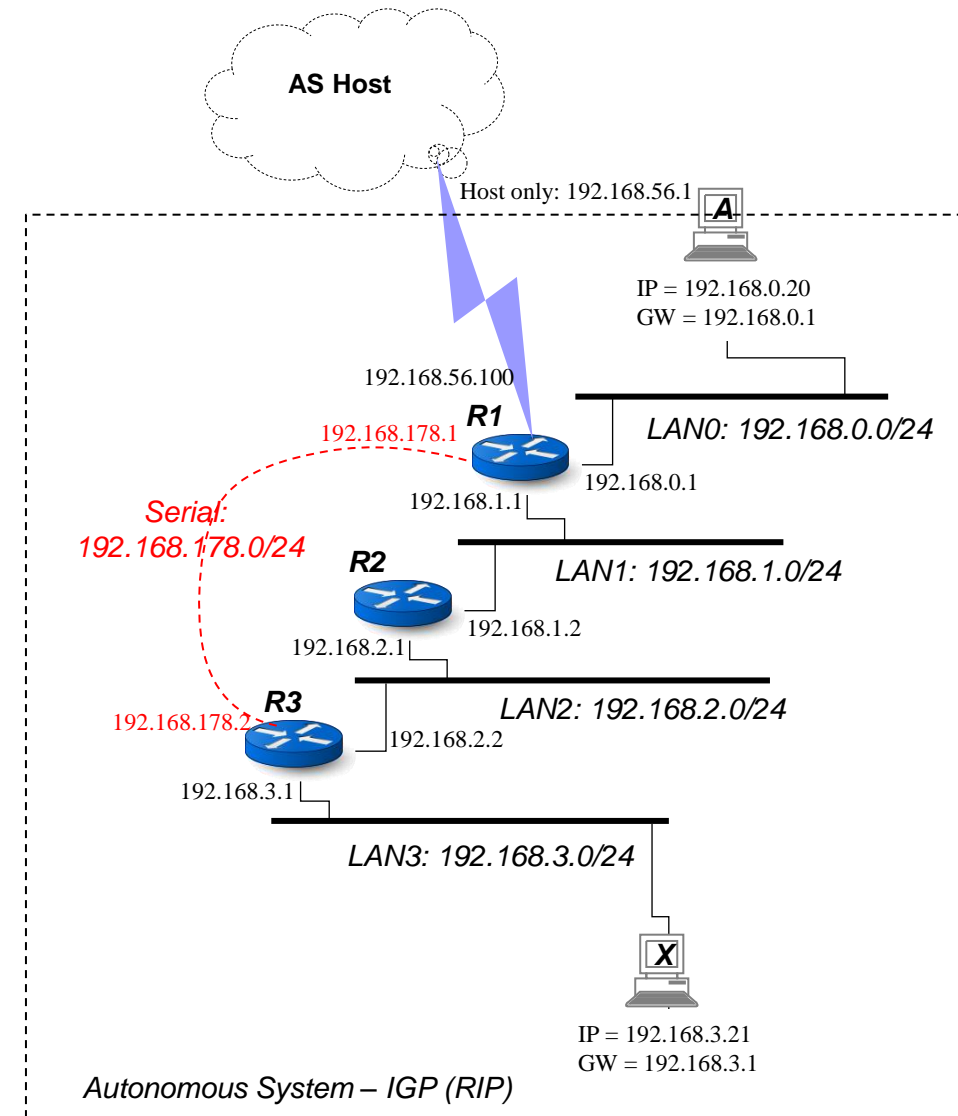
Tối ưu RIP

- Holddown timer: Khi router nhận được “poisoned RTE”, nó khởi tạo Holddown timer (180 giây) để ignore tất cả các request update RTE này từ các router khác, ngoại trừ router đã gửi “poisoned RTE”. Nếu trong khoảng thời gian Holddown timer này, đường link được phục hồi, router khởi tạo “poisoned RTE” cần gửi lại yêu cầu update RTE với Metric được phục hồi.
- Trong giai đoạn quảng bá thông tin giữa các router, cần một khoảng thời gian (gọi là khoảng thời gian hội tụ - convergence time) để tất cả các thiết bị router trong mạng tính toán và đi đến thiết lập bảng định tuyến thống nhất. Trên qui mô lớn, các mạng phức tạp sử dụng thuật toán vector khoảng cách có thể đòi hỏi khoảng thời gian hội tụ quá lớn. Trong thời gian các bảng định tuyến chưa được hội tụ, hệ thống mạng dễ bị định tuyến theo các đường đi không phù hợp. Điều này có thể gây ra các vòng lặp định tuyến hoặc các tình huống chuyển tiếp gói tin không ổn định.
- Để giảm thời gian hội tụ, một giới hạn thường được đặt ra dựa trên số lượng tối đa các bước nhảy trên một đường định tuyến đến mạng đích. Điều này dẫn đến tính hướng các đường định tuyến hợp lệ nhưng vượt quá giới hạn này lại không thể được sử dụng trong các mạng sử dụng định tuyến theo vector khoảng cách.

thực hành:

RIP

- Cài đặt *quagga* cho các router để hỗ trợ RIP
- Tạo môi trường ảo gồm 4 mạng (192.168.x.0/24) kết nối với nhau thông qua 3 router R1, R2, R3
- Cấu hình RIP (zebra & ripd) trên các router
- Vận hành RIP & kiểm tra các bảng routing trên router
- Kiểm tra tính đáp ứng khi hệ thống mạng thay đổi
 - Thêm kết nối serial R1-R3 theo địa chỉ mạng 192.168.178.0/24
 - Kiểm tra đường đi của gói tin giữa A & X trước và sau khi có kết nối mới bằng *traceroute*
 - Ngắt kết nối serial R1-R3 & kiểm tra đường đi của gói tin giữa A & X
- Kết nối với bên ngoài Autonomous System
 - AS được thiết lập gồm các mạng trên và kết nối với máy host bằng network interface Host only của router R1 (mạng 192.168.56.0/24)
 - Máy Host đóng vai trò là một hệ thống AS khác cần được kết nối thông suốt với AS hiện tại → cần lan tỏa kết nối mạng Host only (192.168.56.0/24) đến các router R2, R3
 - Khai báo “redistribute connected” trong file ripd.conf của R1
- Bắt & phân tích các gói tin RIP





Open Shortest Path First (OSPF)

OSPF Introduction

- Version 1 - RFC 1131 (1989), V2 - RFC 2328 (1998), V3 - RFC 5340 (2008) hỗ trợ IPv6. Cài đặt trên router các hãng khác nhau
- Dựa trên trạng thái đường truyền (link state) thay vì khoảng cách như RIP → khắc phục nhược điểm đường đi ngắn nhưng băng thông hẹp.
- Tương tự RIP, dựa trên kết nối láng giềng nhưng không lan truyền bảng routing mà lan truyền toàn bộ topo mạng → mọi router đều lưu giữ topo mạng đầy đủ dưới dạng một đồ thị (đỉnh đồ thị là các router & các mạng nghiệp vụ, cạnh là các kết nối mạng của router)
- Giá (cost) của các cạnh đồ thị thể hiện trạng thái kết nối mạng (network interface) của router.

$$\text{Cost} = 10^8 / \text{interface bandwidth (Mbps)}$$

56-kbps serial link = 1785

64-kbps serial link = 1562

128-kbps serial link = 781

T1 (1.544-Mbps serial link) = 64

E1 (2.048-Mbps serial link) = 48

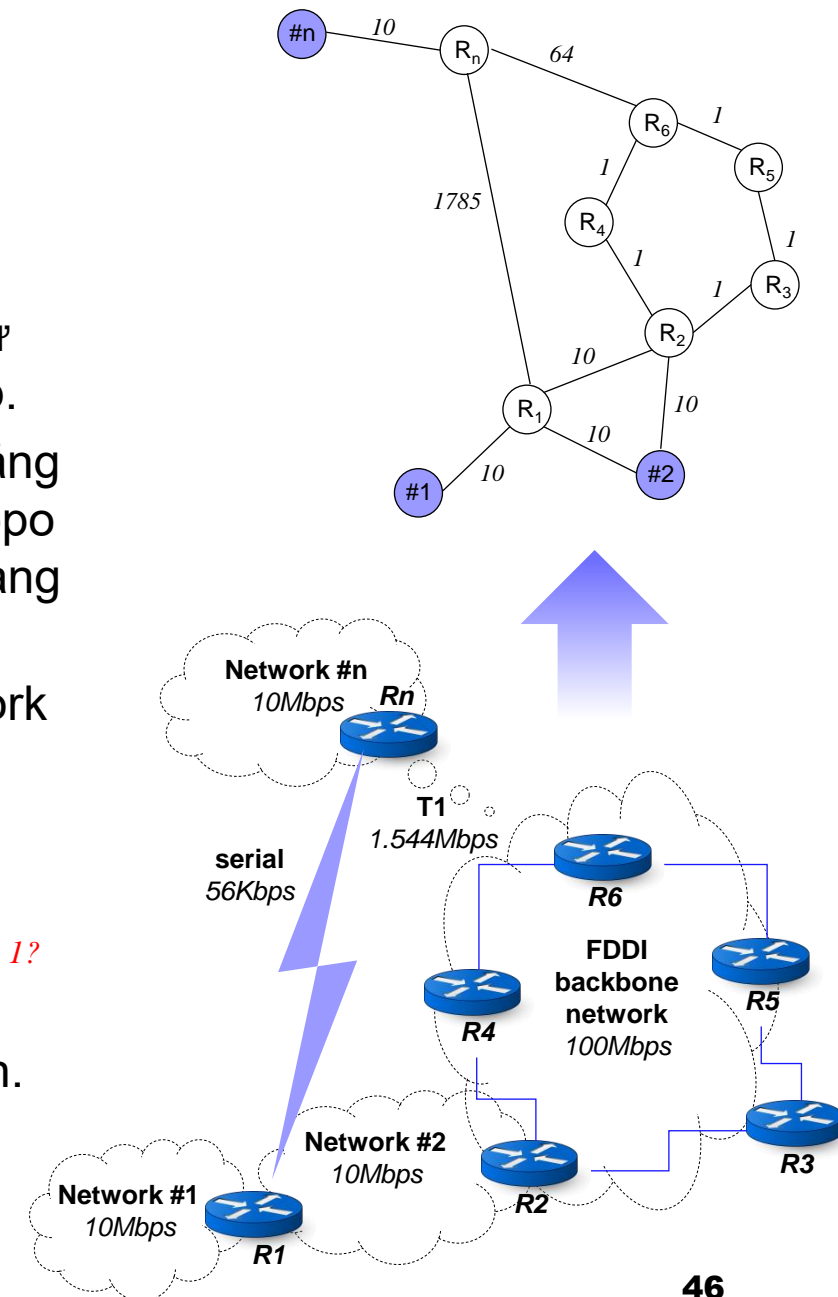
Token Ring (4-Mbps) = 25

Ethernet = 10

Fast Ethernet = 1

Problem: Gigabit Ethernet and faster = 1?

- Mỗi router OSPF sử dụng thuật toán SFP (Shortest Path First¹ – Dijkstra) để tính toán đường đi có tổng cost ngắn nhất đến mạng đích.
- **Các router đều giữ bản đồ topo giống nhau → đường đi có tổng cost ngắn nhất không tạo loop routing**



[1] https://en.wikipedia.org/wiki/Link-state_routing_protocol

Link State, LS Database & Advertisement

■ Link State

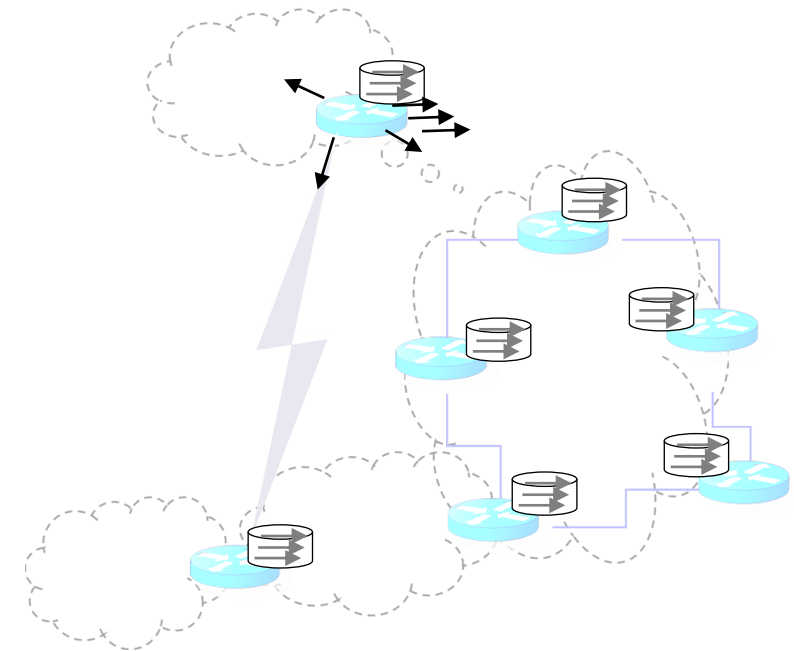
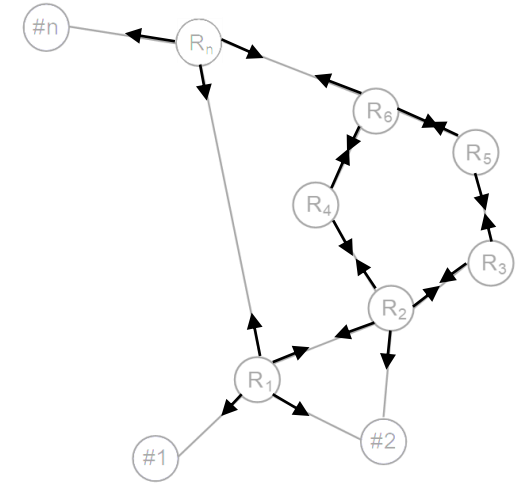
- Link: kết nối mạng của mỗi router
- Link State: router giám sát trạng thái kết nối mạng của mình để phát hiện có sự thay đổi

■ Link State Database

- Tập các Link State của toàn bộ mạng
- Mỗi router lưu giữ một bản copy đầy đủ của LS Database
 - ➔ có thể “tái tạo” đồ thị topo mạng đầy đủ
- Cho phép tìm đường đi có tổng cost nhỏ nhất giữa 2 nút
- Đường đi “ngắn nhất” loại bỏ chu trình (tránh loop routing)

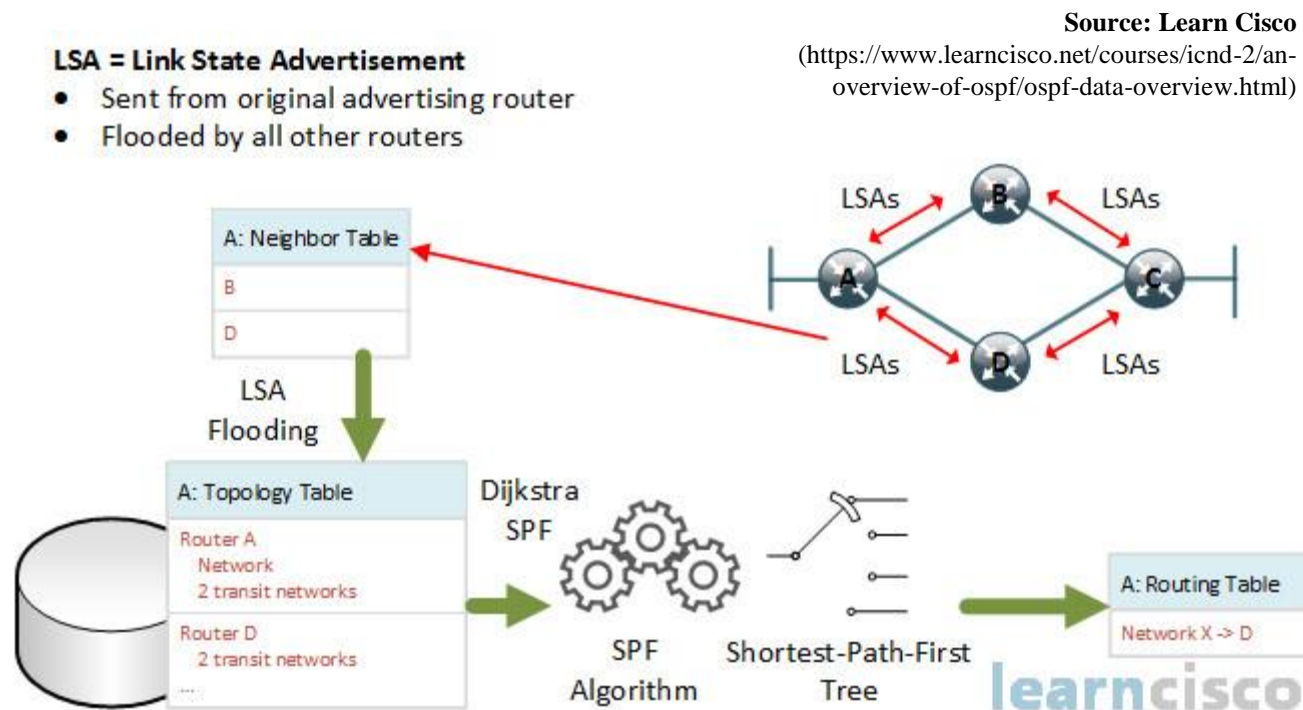
■ Link State Advertisement

- Router phát hiện thay đổi Link State của mình và gửi cho láng giềng
- Router láng giềng tiếp tục lan truyền Link State đến các router khác
 - ➔ quá trình lan truyền Link State (flooding propagation)
- Router kích hoạt quá trình lan truyền gọi là Advertisement Router



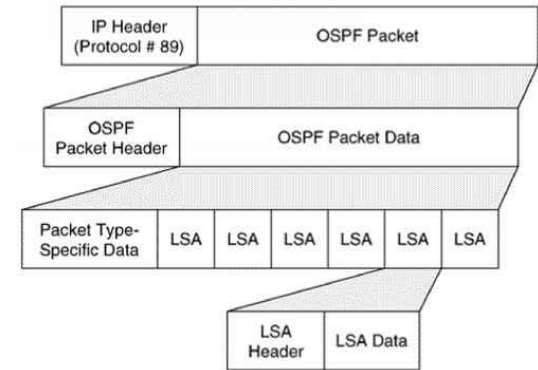
Hoạt động chung

- Khi router được kết nối mạng, nó chạy “hello protocol” để thiết lập quan hệ láng giềng
 - Gửi bản tin Hello đến các router láng giềng yêu cầu cung cấp thông tin
 - Nhận bản tin Hello & thiết lập danh sách láng giềng (neighbor).
- Khi có thay đổi trong mạng (làm topo mạng thay đổi hoặc link state thay đổi)
→ router gửi thông tin về trạng thái liên kết (link state) cho láng giềng bằng bản ghi LSA (Link State Advertisement).
- LSA được tiếp tục lan truyền (flooding propagation) trên toàn vùng mạng để thống nhất mọi router đều cập nhật trạng thái liên kết mới này vào đồ thị topo mạng mà nó đang lưu giữ (dưới dạng một cơ sở dữ liệu trạng thái liên kết – Link State Database).
- Các router láng giềng cũng thường xuyên đồng bộ LS Database bằng cách gửi nhau các bản tin Database description, mỗi bản tin chứa một tập các LSA.
- Router có thể chủ động yêu cầu cập nhật LS Database bằng cách gửi LSA request cho láng giềng.
- Sau khi cập nhật LS Database, giải thuật Dijkstra SPF được chạy để tính toán đường đi có cost nhỏ nhất đến tất cả các mạng trong hệ thống & cập nhật vào bảng routing.



OSPF Packet Format

- OSPF chạy trực tiếp trên IP (Protocol number = 89)
- Trường Type xác định mục đích & nội dung (data) packet:
 - 1: Hello
 - 2: Database Description
 - 3: Link-State Request
 - 4: Link-State Update
 - 5: Link-State Acknowledgment
- Định danh trong topo mạng:
 - Area ID: tối ưu thời gian hội tụ (convergence time), chia AS thành nhiều vùng (Area), mỗi vùng được định danh bằng một Area ID phân biệt.
 - Router ID: định danh cho router, được gán tự động theo địa chỉ IP cao nhất của router, hoặc được admin cấu hình theo giá trị tùy ý
 - Link State ID: định danh của LSA.
- LSA: đơn vị dữ liệu OSPF
 - LS age: rất giống TTL trong gói tin IP nhưng tính bằng giây, hỗ trợ lưu trữ & loại bỏ trong Link-State DB. Có giá trị bằng 0 khi vừa được tạo, tăng lên trong mỗi bước lan truyền (flooding propagation hop) & bị loại bỏ khi đến giá trị MaxAge.
 - LS Sequence number: dùng để xác định phiên bản LSA. Có giá trị InitialSequenceNumber (0x80000001) khi được khởi tạo & tăng lên 1 sau mỗi lần update
 - Advertising Router: router chịu trách nhiệm giám sát trạng thái LSA và kích hoạt quá trình lan truyền LSA. Adv Router duy trì phiên bản hiện tại của LSA và tăng lên 1 mỗi khi kích hoạt lan truyền một phiên bản mới của LSA này.
 - LSA là “các mảnh ghép dữ liệu” của toàn bộ hệ thống (lưu trữ & lan truyền)
→ quan trọng nhất & cũng phức tạp nhất của hệ thống



OSPF Packet header:

0	1	2	3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1			
Version #	Type	Packet length	
Router ID			
Area ID			
Checksum		AuType	
Authentication			
Authentication			

LSA header:

0	1	2	3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1			
LS age	Options	LS type	
Link State ID			
Advertising Router			
LS sequence number			
LS checksum		length	

Các loại OSPF Packet

Version#	1	Packet length	
Router ID			
Area ID			
Checksum		AuType	
Authentication			
Authentication			
Network mask			
Hello interval		Options	Rtr Pri
Router dead interval			
Designated router			
Backup designated router			
Neighbor			

Hello

Link-state Request

Version#	3	Packet length
Router ID		
Area ID		
Checksum	AuType	
Authentication		
Authentication		
LS type		
Link-state ID		
Advertising router		

Database Description

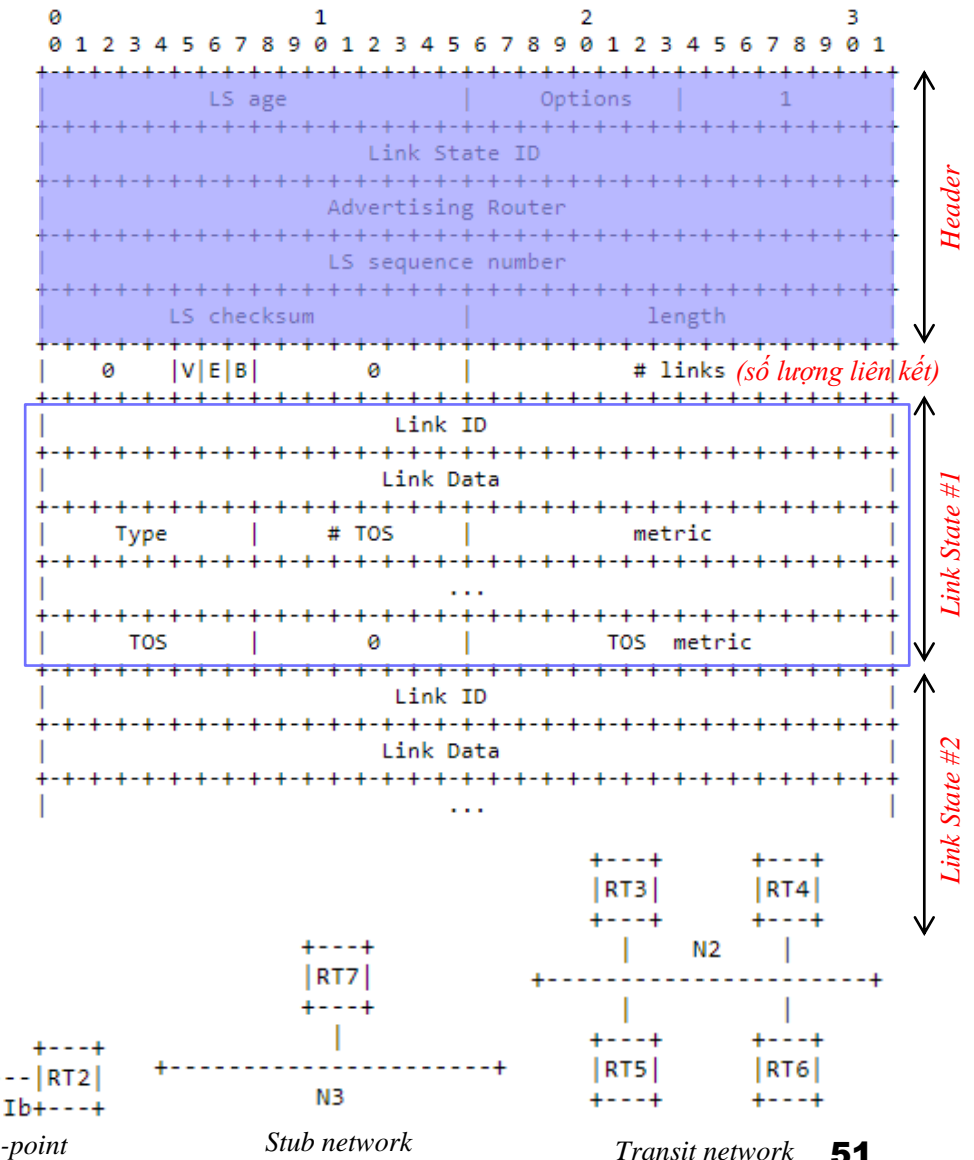
Version#	2	Packet length													
Router ID															
Area ID															
Checksum				AuType											
Authentication															
Authentication															
Interface MTU				Options		0	0	0	0	0	0	0	1	M	MS
DD sequence number															
LSA header															

Link-State Update

Version#	4	Packet length
Router ID		
Area ID		
Checksum	AuType	
Authentication		
Authentication		
No. of Advertisements		
List of LSAs		

Router-LSA & Router Network Interface

- Có nhiều loại LSA để phục vụ các tình huống lan truyền và lưu trữ khác nhau (trong vùng, giữa các vùng, v.v.). Router-LSA là LSA cơ bản nhất (LS Type=1), được dùng để mỗi router thông báo về các liên kết mạng của nó đến các router khác trong hệ thống.
- Link State ID = Router ID (router khởi tạo LSA này).
- # links: số kết nối mạng của router, cũng là số bản ghi Link State được gửi trong phần Data của packet.
- Các trường bit:
 - bit V (virtual link): router kết nối với một liên kết ảo (virtual link – kết nối trực tiếp 2 Area mà không đi qua Area 0)¹
 - bit E (external): router biên của AS (có kết nối với một AS khác - ASBR router)
 - bit B (border): router biên của một Area (ABR router)
- Router kết nối mạng bằng các network interface thuộc 1 trong 4 kiểu (xác định bằng trường Type):
 - 1: kết nối point-to-point với một router khác. Link ID là địa chỉ IP router láng giềng.
 - 2: kết nối vào mạng vận chuyển (transit network) cùng nhiều router khác. Link ID là IP của router được chỉ định (Desinated Router²) trong transit network này.
 - 3: kết nối vào mạng nghiệp vụ (OSPF gọi là stub network – deadend network) – không có router nào khác kết nối vào mạng này. Link ID là địa chỉ IP của mạng stub.
 - 4: Virtual link. Link ID là địa chỉ IP của router láng giềng trong virtual link này



[1] <https://tools.ietf.org/html/rfc2328#section-15>: Virtual Links

[2] <https://tools.ietf.org/html/rfc2328#section-7.3>: The Designated Router

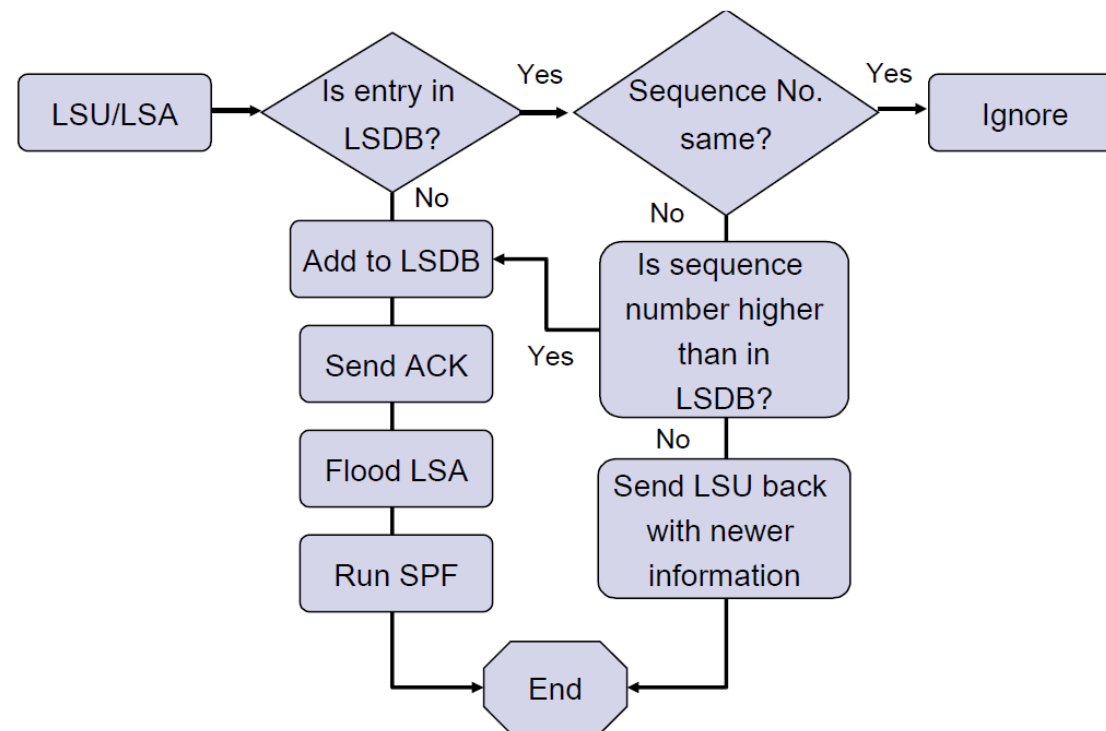
Các loại LSA

Link-state type	OSPF function
1	Router link states
2	Network link states
3	Summary link states
4	ASBR link state
5	External link advertisement
7	NSSA external link state
8	External attributes for BGP
9, 10, 11	Opaque LSA



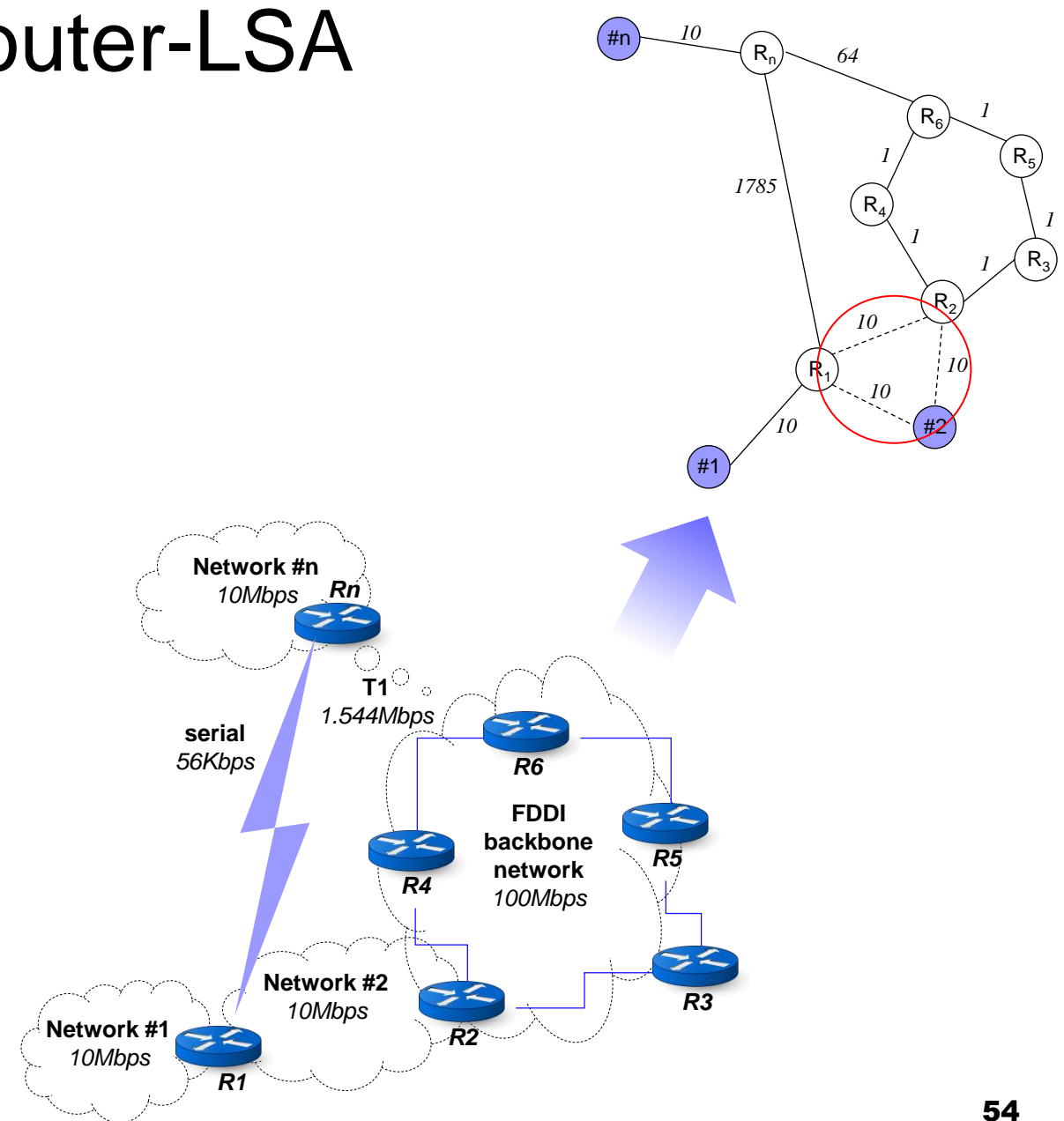
Thuật toán cập nhật Link-state DB theo LSA

1. Router nhận OSPF Packet type=4 (Link State Update) hoặc các loại packet khác, lấy LSA từ trường Data
2. Kiểm tra Link State ID để xác định LSA đã có trong link-state database?
3. Nếu chưa có, cập nhật ngay vào database, gửi ACK (OSPF Type=5) và lan truyền (flood) LSA này tiếp cho các router láng giềng. Chạy thuật toán Shortest Path First để cập nhật bảng routing
4. Nếu đã có, kiểm tra Sequence Number để xác định LSA vừa nhận là cũ hay mới. Nếu LSA nhận được cũ hơn trong database, trả lời bằng OSPF Link State Update (type=4) với LSA mới (lấy trong link-state database).
5. Nếu LSA nhận được mới hơn trong database, cập nhật vào database và xử lý tiếp như bước 3.



Link-state Database by Router-LSA

- R1-LSA: #link = 3
 - Interface #1 – stub network #1: cost=10, #seq
 - Interface #2 – transit network #2: cost=10, #seq
 - Interface #3 – point-to-point Rn: cost=1785, #seq
- R2-LSA: #link = 3
 - Interface #1 – transit network #2: cost=10, #seq
 - Interface #2 – point-to-point R4: cost=1, #seq
 - Interface #3 – point-to-point R3: cost=1, #seq
- R3-LSA: #link = 2
 - Interface #1 – point-to-point R2: cost=1, #seq
 - Interface #2 – point-to-point R5: cost=1, #seq
- R4-LSA: #link = 2
 - Interface #1 – point-to-point R2: cost=1, #seq
 - Interface #2 – point-to-point R6: cost=1, #seq
- R5-LSA: #link = 2
 - Interface #1 – point-to-point R3: cost=1, #seq
 - Interface #2 – point-to-point R6: cost=1, #seq
- R6-LSA: #link = 3
 - Interface #1 – point-to-point R4: cost=1, #seq
 - Interface #2 – point-to-point R5: cost=1, #seq
 - Interface #3 – point-to-point Rn: cost=64, #seq
- Rn-LSA: #link = 3
 - Interface #1 – stub network #n: cost=10, #seq
 - Interface #2 – point-to-point R6: cost=64, #seq
 - Interface #3 – point-to-point R1: cost=1785, #seq



Network-LSA & Designated Router

■ Topo mạng transit

- Nhiều router cùng kết nối vào một mạng (broadcast/NBMA) để forward IP packet
- Nếu áp dụng Router-LSA: các router đôi một láng giềng & kết nối trực tiếp trong mạng transit → đồ thị topo hiểu là kết nối đầy đủ point-to-point (sai !!!)
- Mô hình hóa bằng nút đồ thị đặc biệt, đại diện cho mạng transit & kết nối với các nút router
→ **ai chịu trách nhiệm giám sát & kích hoạt lan truyền LSA của transit network?**

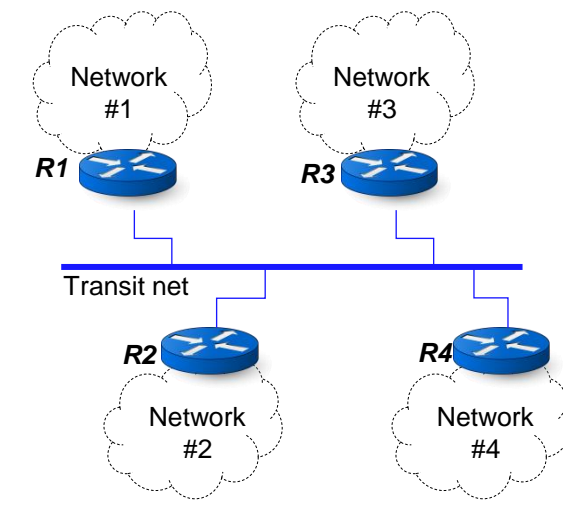
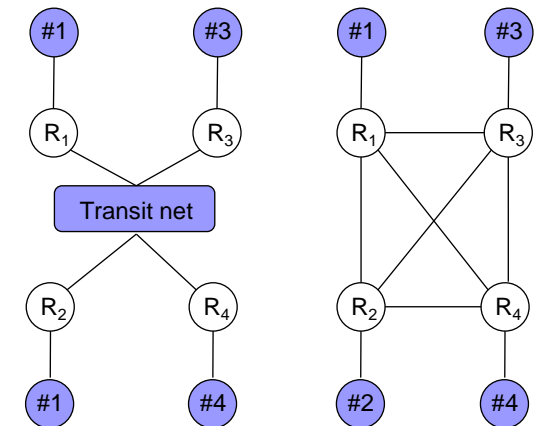
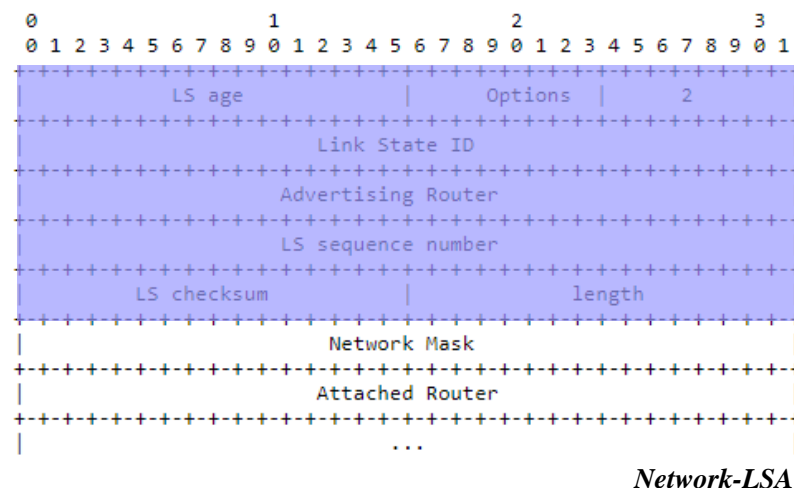
■ Designated Router (router được chỉ định)

- Chịu trách nhiệm giám sát & kích hoạt quá trình lan truyền LSA của transit network
- Designated router được chỉ định bằng giao thức OSPF Hello giữa các router kết nối trong transit network

■ Network-LSA (Type=2)

- Đóng gói thông tin của transit net phục vụ lan truyền và lưu trữ
- Link State ID = IP kết nối vào mạng transit của Designated Router (prefix địa chỉ mạng được xác định bằng trường Network Mask)
- Adv Router = ID của Designated Router
- Danh sách các router trong transit net
- (không quan tâm đến metric)

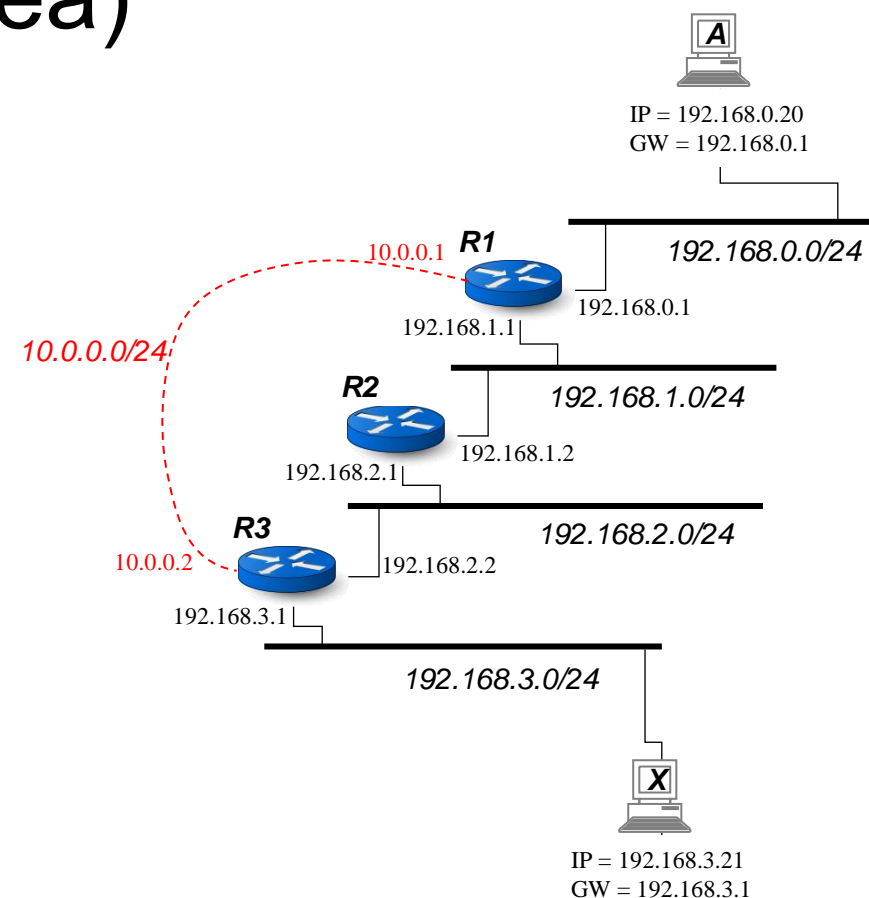
- Các router sử dụng Network-LSA để hiểu đúng về topo mạng & phục vụ thuật toán SPF. Cost (metric) vẫn sử dụng giá trị trong các Router-LSA



thực hành:

OSPF cơ bản (Single Area)

- Cấu hình *ospf* cho các router (đã cài *quagga*)
- Sử dụng môi trường mạng & router như đã thực hành với RIP
- Vận hành OSPF & kiểm tra các bảng routing trên router
- Kiểm tra tính đáp ứng khi hệ thống mạng thay đổi
 - Thêm kết nối serial R1-R3 theo địa chỉ mạng 10.0.0.0/24, đặt Cost của link là 10 (giống các kết nối khác)
 - Kiểm tra đường đi của gói tin giữa A & X trước và sau khi có kết nối mới bằng *traceroute*
 - Thay đổi Cost của kết nối serial R1-R3 thành 64 (đường T1) & kiểm tra đường đi của gói tin giữa A & X
- Sử dụng *telnet* vào *ospfd* (cổng 2604) để xem các thông tin link-state database Router-LSA và Network-LSA
 - telnet 127.0.0.1 2604
 - show ip ospf database router
 - show ip ospf database network



Các loại LSA

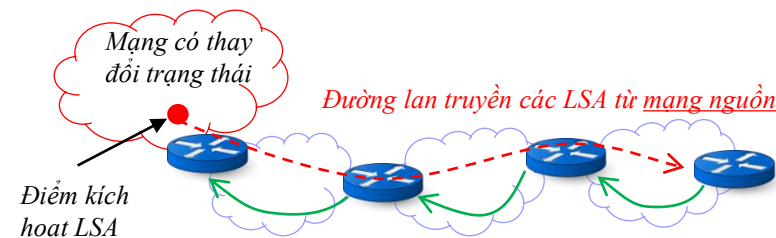
Link-state type	OSPF function
1	Router link states
2	Network link states
3	Summary link states
4	ASBR link state
5	External link advertisement
7	NSSA external link state
8	External attributes for BGP
9, 10, 11	Opaque LSA



Dùng để lan truyền trạng thái từng kết nối mạng (link state) nhằm xây dựng một bản đồ mạng thống nhất trên toàn bộ các router & hỗ trợ tìm đường cost thấp nhất



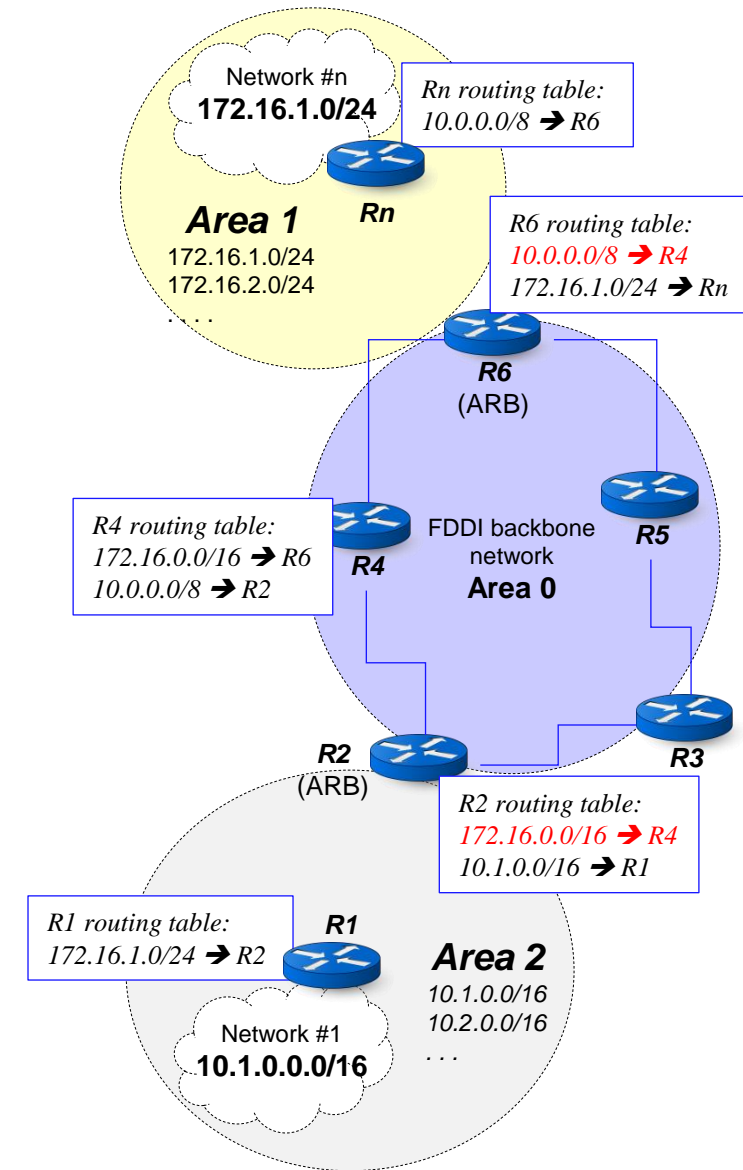
Dùng để thông báo sự có mặt của một mạng đích và xây dựng đường đi đến nó từ mỗi router trong hệ thống.



- Đường routing được xây dựng từ các router đến mạng đích bằng các RTE “học” được khi router lan truyền LSA
- Khác với RIP, đường routing không nhất thiết trùng với đường lan truyền LSA

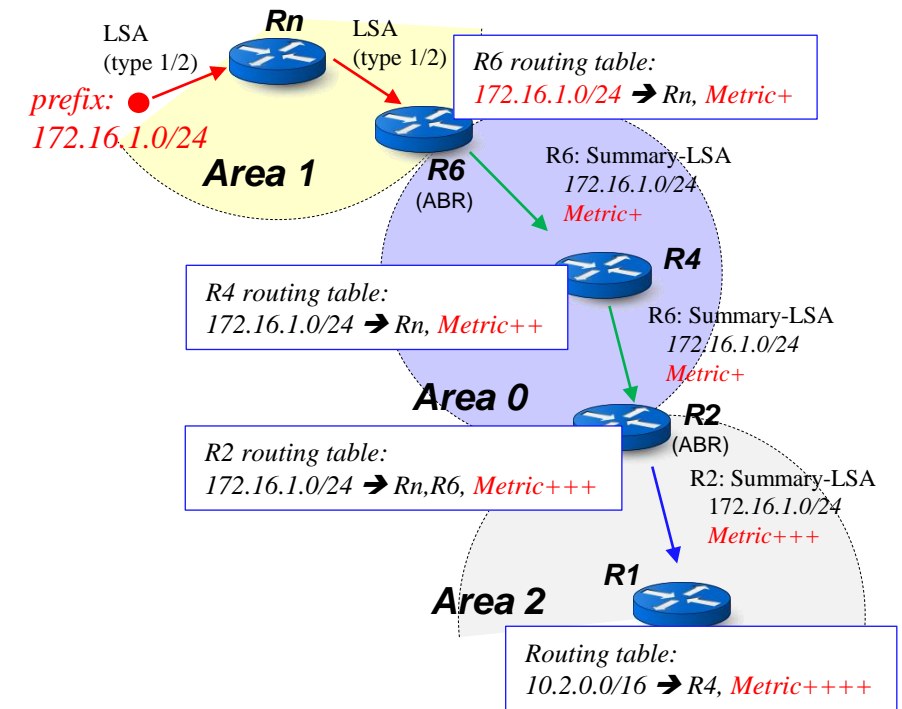
OSPF Multi Area

- Điều gì xảy ra khi số router lớn (hàng trăm hoặc hàng nghìn?)
 - Dữ liệu LS Database quá lớn, vượt khả năng lưu trữ tại các router
 - Thời gian lan truyền LSA lâu.
 - Nhiều LSA ở mạng xa hoặc mạng riêng, không liên quan đến router nhưng khi nhận được trạng thái mới của LSA này, router vẫn phải chạy lại thuật toán SPF
 - Bảng routing dài (trăm hay nghìn dòng) làm giảm hiệu suất routing, mặc dù nhiều dòng routing sẽ không dùng đến
- OSPF lãng phí tài nguyên LSA trong nhiều qui hoạch mạng IP:
 - 2 vùng mạng (1&2) sử dụng 2 dải địa chỉ (10.0.0.0/8 & 172.16.0.0/16) chia thành các subnet nội bộ trong vùng (10.x.0.0/16 & 172.16.x.0/24) → trong mỗi vùng, các router sử dụng OSPF để lan truyền các LSA ở cấp subnet để tính cost routing nội vùng.
 - Các router “gateway vùng” kết nối nhau hoặc qua backbone → không cần quan tâm subnet, chỉ quan tâm IP lớp mạng lớn khi cần chuyển gói tin giữa các vùng.
 - **Việc lan truyền LSA (subnet) trong một vùng sang vùng khác là không cần thiết**
- Giải pháp phân vùng OSPF
 - Mạng được phân chia thành nhiều vùng, mỗi vùng được cập nhật bản đồ đầy đủ của riêng vùng đó. Vùng khác nhau không có bản đồ của nhau mà chỉ có thông tin định tuyến.
 - Area 0 (area backbone): chứa các backbone router (R2, R3, R4)
 - Các vùng phải kết nối vào Area 0, và không có 2 Area nào kết nối với nhau mà không qua Area 0
 - Area border router (ABR): R2, R4
 - Internal area router: R1, R5



Summary-LSA

- Router-LSA và Network-LSA chứa các thông tin để xây dựng đồ thị topo mạng → không được lan truyền ra ngoài Area.
- Summary-LSA được sử dụng để lan truyền giữa các Area về sự có mặt của các mạng bên trong một Area (giúp cho việc định tuyến, không dùng để xây dựng đồ thị topo mạng).
- Cấu trúc gói tin Summary-LSA
 - Type = 3
 - Link State ID = IP mạng khởi phát
 - Giống Network-LSA, net prefix mạng nguồn = IP mạng khởi phát AND Network Mask
 - Metric: giá của đường đi ngắn nhất từ adv router đến mạng khởi phát
- Hoạt động:
 - Vùng khởi phát: có thay đổi LSA → lan truyền Router-LSA hoặc Network-LSA. ABR nhận LSA trong vùng, cập nhật LS Database & chạy thuật toán SPF để tính cost cho bảng routing. Kích hoạt Summary-LSA (mới hoặc update với #sequence +1) với Link State ID = IP mạng khởi phát và lan truyền vào Area 0.
 - Vùng backbone: các router trong Area 0 nhận được Summary-LSA, cập nhật LS Database, update routing table với cost mới đến mạng khởi phát. Lan truyền tiếp Summary-LSA này trong Area 0.
 - Vùng khác: khi Summary-LSA đi đến ABR để chuyển từ Area 0 vào một vùng khác, ABR không tiếp tục lan truyền Summary-LSA này mà kích hoạt Summary-LSA mới: adv router là chính nó, link state ID giữ nguyên (là IP mạng khởi phát), Metric là cost từ ABR đến network khởi phát và lan truyền vào vùng riêng



Summary-LSA

0										1										2										3											
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1
LS age										Options										3 or 4																					
Link State ID																																									
Advertising Router																																									
LS sequence number																																									
LS checksum																				length																					
Network Mask																																									
0										metric																															
TOS										TOS metric																															
...																																									

Tối ưu OSPF liên vùng với Summary-LSA

■ Hoạt động của ABR router:

- Chặn các LSA type 1&2 để chỉ lan truyền trong nội vùng, giúp các router xây dựng đồ thị topo mạng của vùng.
- Kích hoạt Summary-LSA (để tóm tắt sự có mặt của một mạng & cost đi đến mạng này từ ABR) & lan truyền từ trong vùng ra ngoài hoặc từ ngoài vùng vào trong (tùy thuộc mạng cần thông báo nằm ở vị trí nào trên toàn bộ “liên vùng”)
- ➔ Các router không biết topo chi tiết của vùng khác, nhưng có thông tin routing đến các mạng trong liên vùng (giống RIP)

■ Summary-LSA không đòi hỏi chạy lại thuật toán SPF

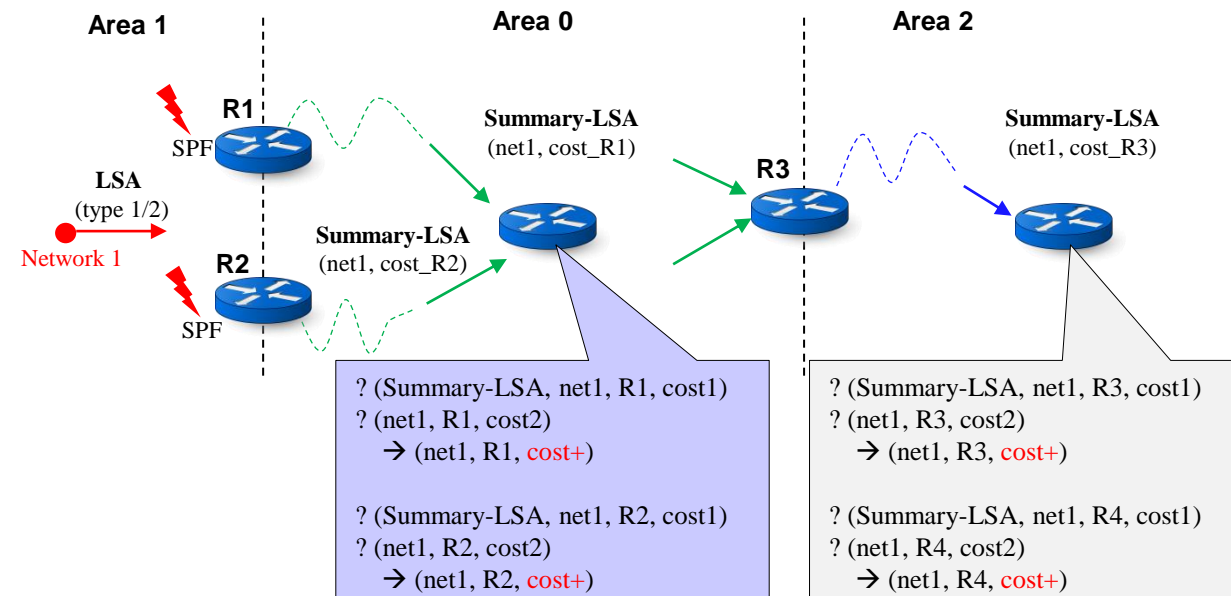
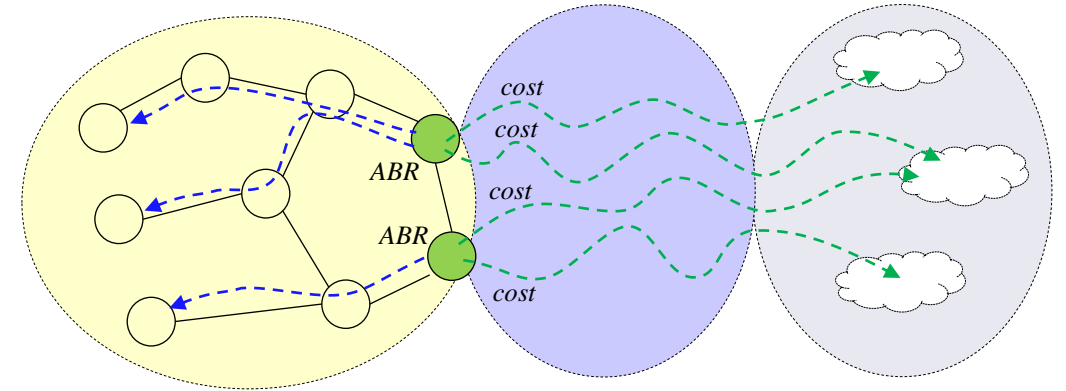
- Mạng khởi phát link state: Network 1
- Lan truyền LSA (type 1,2) trong vùng, đến các ABR (R1, R2)
- ABR chạy thuật toán SPF để tính cost mới đi đến mạng khởi phát (tương ứng là $cost_R1$ & $cost_R2$) và kích hoạt Summary-LSA vào Area 0

- Router trong Area 0 nhận được các Summary-LSA, chạy thuật toán update routing:

$$cost+ = cost2 - cost1 + cost_R1$$

- Lan truyền Summary-LSA trong Area 0 đến các ABR (R3, R4)
- Router ABR update routing như các router khác trong Area 0, kích hoạt Summary-LSA vào vùng riêng
- Router trong vùng riêng nhận được các Summary-LSA, chạy thuật toán update routing:

$$cost+ = cost2 - cost1 + cost_R3$$



thực hành: Summary-LSA

- Các router được tổ chức thành 3 vùng:
 - Area 0: R1, R2, R3, R4
 - ABR: R1 với Area 1, R3, R4 với Area 2
- Kiểm tra LS database của R2 lọc theo Summary-LSA: thấy R3 và R4 đều kích hoạt Summary-LSA cho mạng 192.168.3.0/25 và lan truyền vào Area 0 nhưng với Metric khác nhau

```
R2> show ip ospf database summary adv-router 3.3.3.3

OSPF Router with ID (2.2.2.2)

Summary Link States (Area 0.0.0.0)

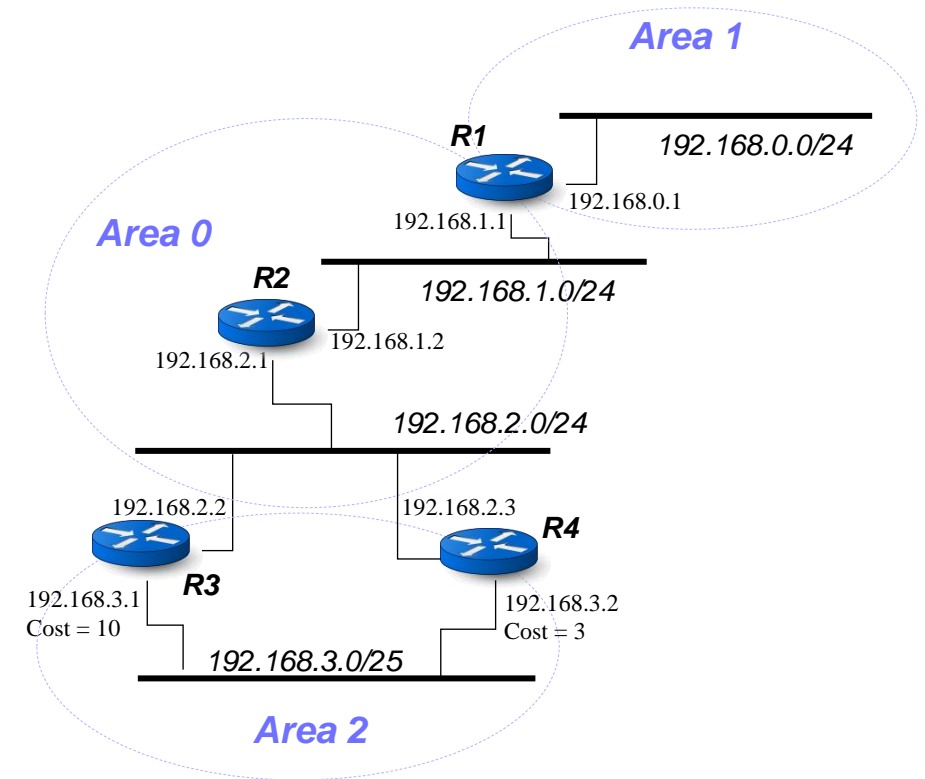
LS age: 519
Options: 0x2 : *i-l-l-l-l-l-E!*
LS Flags: 0x6
LS Type: summary-LSA
Link State ID: 192.168.3.0 (summary Network Number)
Advertising Router: 3.3.3.3
LS Seq Number: 80000015
Checksum: 0x7dc8
Length: 28
Network Mask: /25
TOS: 0 Metric: 10
```

```
R2> show ip ospf database summary adv-router 4.4.4.4

OSPF Router with ID (2.2.2.2)

Summary Link States (Area 0.0.0.0)

LS age: 728
Options: 0x2 : *i-l-l-l-l-l-E!*
LS Flags: 0x6
LS Type: summary-LSA
Link State ID: 192.168.3.0 (summary Network Number)
Advertising Router: 4.4.4.4
LS Seq Number: 80000002
Checksum: 0x3f1d
Length: 28
Network Mask: /25
TOS: 0 Metric: 3
```



- Kiểm tra LS Database của R1 trong Area 1, thấy R1 kích hoạt Summary-LSA cho các network trong hệ thống. Với network 192.168.3.0/25, lấy cost bằng đường đi qua R4 (nhỏ hơn đi qua R3)

```
LS age: 1033
Options: 0x2 : *i-l-l-l-l-l-E!*
LS Flags: 0x3
LS Type: summary-LSA
Link State ID: 192.168.1.0 (summary Network Number)
Advertising Router: 1.1.1.1
LS Seq Number: 80000015
Checksum: 0xcc04
Length: 28
Network Mask: /24
TOS: 0 Metric: 10
```

```
LS age: 212
Options: 0x2 : *i-l-l-l-l-l-E!*
LS Flags: 0x3
LS Type: summary-LSA
Link State ID: 192.168.2.0 (summary Network Number)
Advertising Router: 1.1.1.1
LS Seq Number: 80000002
Checksum: 0x4c8c
Length: 28
Network Mask: /24
TOS: 0 Metric: 20
```

```
LS age: 1164
Options: 0x2 : *i-l-l-l-l-l-E!*
LS Flags: 0xb
LS Type: summary-LSA
Link State ID: 192.168.3.0 (summary Network Number)
Advertising Router: 1.1.1.1
LS Seq Number: 80000002
Checksum: 0x62f1
Length: 28
Network Mask: /25
TOS: 0 Metric: 23
```

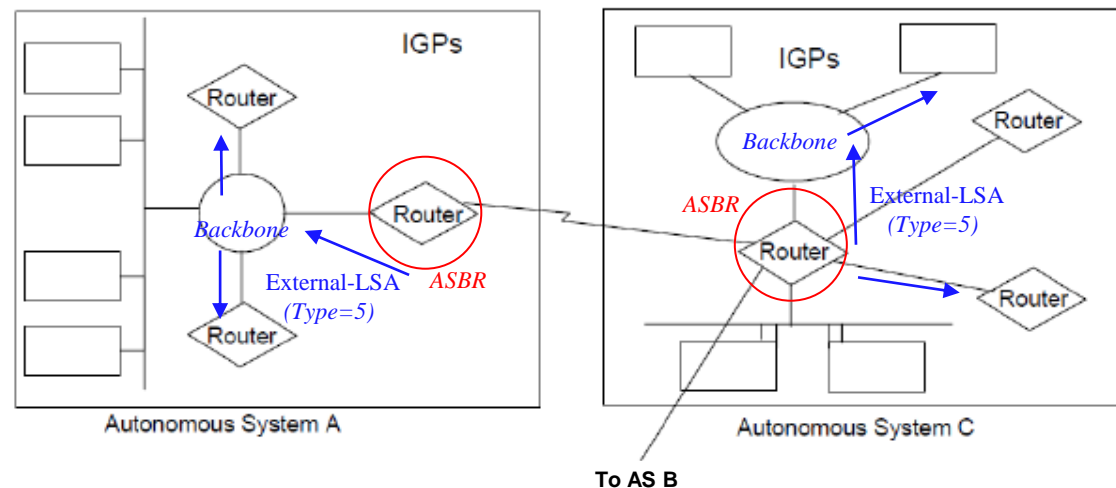

Autonomous System & External-LSA

■ Autonomous System Border Router (ASBR)

- Có thể thuộc Backbone Area hoặc một Area khác.
- Có kết nối trực tiếp với một router thuộc AS khác.
- Chuyển tiếp gói tin từ bên trong AS ra mạng ngoài (external network)

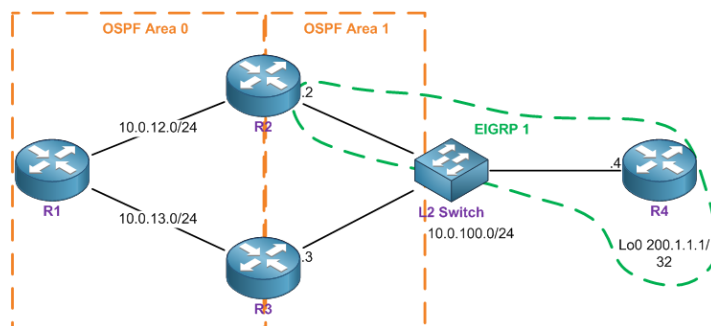
■ Hoạt động của External-LSA

- ASBR nhận thông tin routing từ AS khác (theo giao thức EGP giữa AS), kích hoạt External-LSA và lan truyền vào trong AS, qua các Area để đến tất cả router.
- Mục đích lan truyền External-LSA (chứa thông tin external net) vào bên trong AS không phải để các router cập nhật LS Database mà để “học” một RTE đi đến external net thì cần forward cho ASBR này.



■ Cấu trúc External-LSA

- Type=5
- Link State ID = IP mạng ngoài (kết hợp Network Mask để xác định network prefix của mạng ngoài)
- Bit E: Metric Type
 - = 0: hệ thống Metric mạng ngoài giống mạng trong
 - = 1: hệ thống Metric mạng ngoài khác mạng trong
- Forwarding address: gateway đi ra mạng ngoài. Nếu = 0.0.0.0 thì gateway chính là adv router



■ Ví dụ sử dụng Forwarding address:

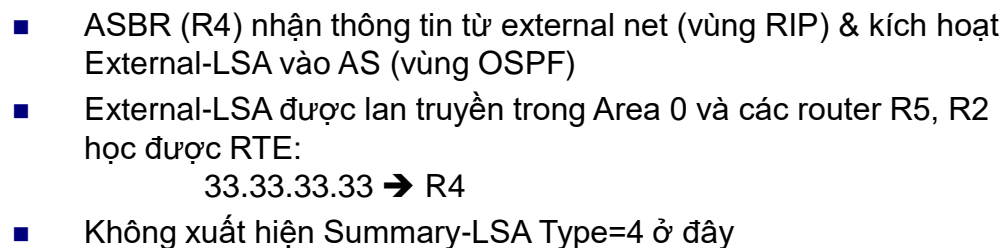
- AS = OSPF Area 0 + Area 1. Mạng ngoài: 200.1.1.1
- ASBR: R2
- Gateway ra mạng ngoài có thể là R2 (10.0.12.2) hoặc Switch (10.0.100.1).

0	1	2	3																										
0 1 2 3 4 5 6 7 8 9 0	1 2 3 4 5 6 7 8 9 0	1 2 3 4 5 6 7 8 9 0	1 2 3 4 5 6 7 8 9 0 1																										
LS age										Options										5									
Link State ID																													
Advertising Router																													
LS sequence number																													
LS checksum															length														
Network Mask																													
E	0				metric																								
Forwarding address																													
External Route Tag																													
E	TOS				TOS metric																								
Forwarding address																													

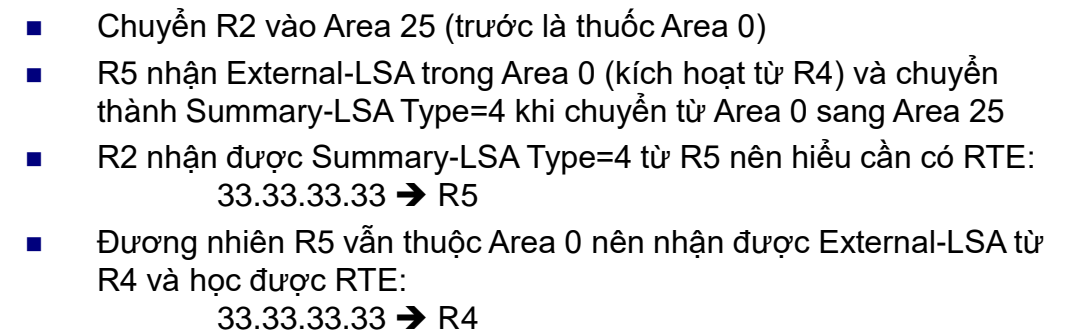
External-LSA

Summary-LSA Type 4

- Có thể hiểu Summary-LSA Type=3 dùng để báo cho router nhận (receiver) biết route đến một network đích (nơi đã khởi tạo LSA này) thì cần đi qua router đang gửi LSA này (Advertising Router). Điều này giống với mục đích của External-LSA, chỉ khác là network đích là một external network (nằm ngoài AS)
- Summary-LSA Type=4 được ABR (không phải ASBR) kích hoạt vào bên trong Area khi cần thông báo đường đi đến external net cần được forward qua ASBR.
- Cấu trúc giống Summary-LSA Type=3, khác Link State ID:
 - Type 3: Link State ID = IP của mạng nguồn (đã khởi tạo LSA)
 - Type 4: Link State ID = Router ID của ASBR
- ➔ Trùng lặp chức năng & hoạt động của External-LSA và Summary-LSA Type 4(?)¹



Thực nghiệm trên Quagga OSPF cho thấy Type 5 vẫn đi vào các Area riêng (không chỉ Area 0) và không xuất hiện Type 4



thực hành:

External-LSA

- Kết nối mạng ngoài bằng kiểu Host-only Adapter
- R2: ASBR
- R2 kích hoạt External-LSA để lan truyền vào Area 0 và các area khác

```
R2> show ip ospf database external

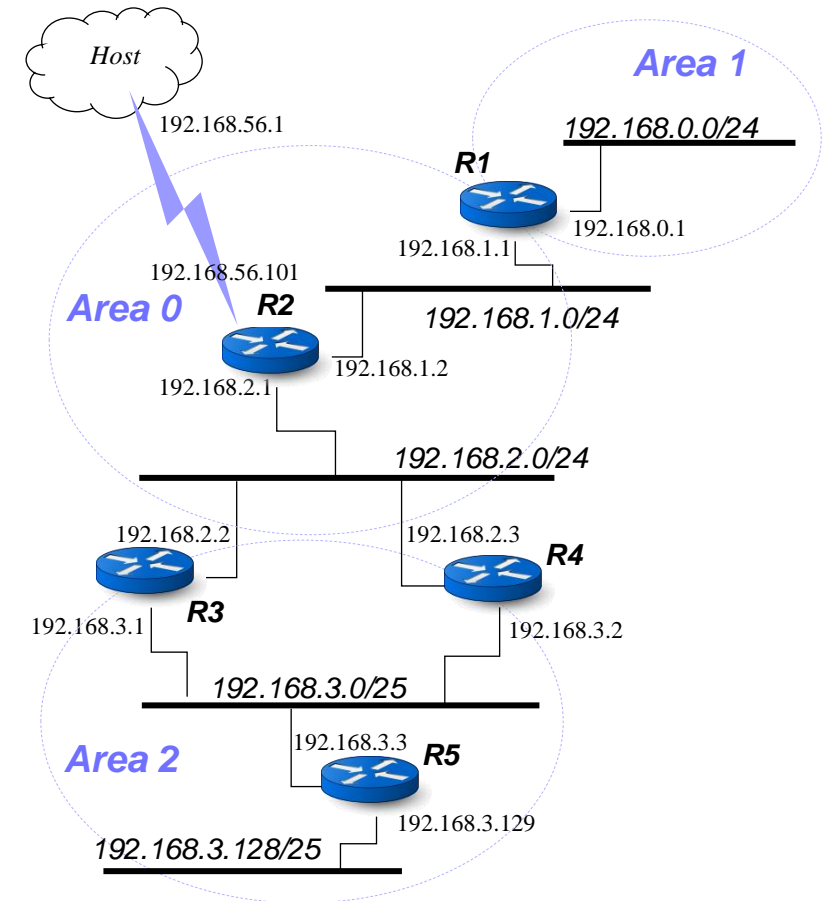
OSPF Router with ID (2.2.2.2)

        AS External Link States

LS age: 76
Options: 0x2 : *i--i--i--iE!*
LS Flags: 0xb
LS Type: AS-external-LSA
Link State ID: 192.168.56.0 (External Network Number)
Advertising Router: 2.2.2.2
LS Seq Number: 80000004
Checksum: 0x1dc4
Length: 36
Network Mask: /24
    Metric Type: 2 (Larger than any link state path)
    TOS: 0
    Metric: 20
    Forward Address: 0.0.0.0
    External Route Tag: 0
```

- Các router nhận được External-LSA tạo đường route đi ra mạng ngoài dựa trên thông tin forward address

```
[root@R3 ~]# route -n
Kernel IP routing table
Destination Gateway Genmask Flags Metric Ref Use Iface
192.168.3.0 0.0.0.0 255.255.255.128 U 0 0 0 eth2
192.168.3.128 192.168.3.3 255.255.255.128 UG 20 0 0 eth2
192.168.2.0 0.0.0.0 255.255.255.0 U 0 0 0 eth1
192.168.1.0 192.168.2.1 255.255.255.0 UG 20 0 0 eth1
192.168.0.0 192.168.2.1 255.255.255.0 UG 30 0 0 eth1
192.168.56.0 192.168.2.1 255.255.255.0 UG 20 0 0 eth1
```



Các loại Area & Tối ưu vùng OSPF

■ Stub Area (vùng kết thúc – nhớ lại “stub network”):

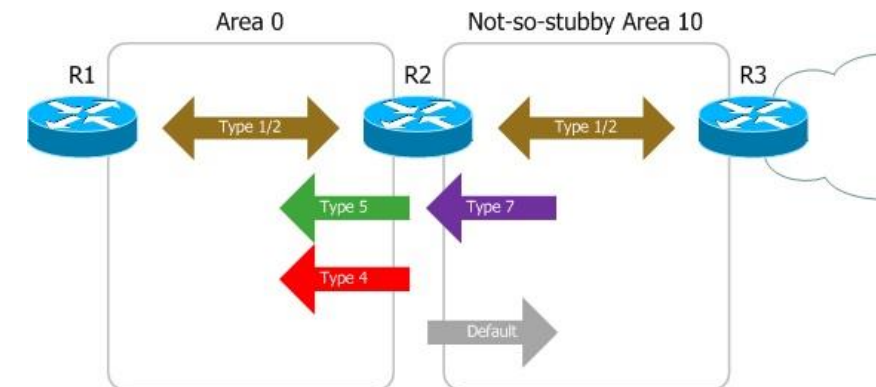
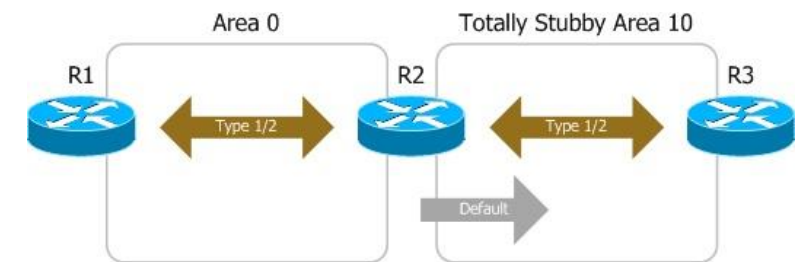
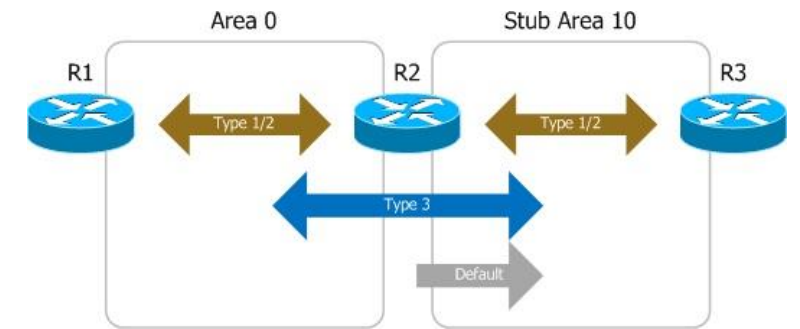
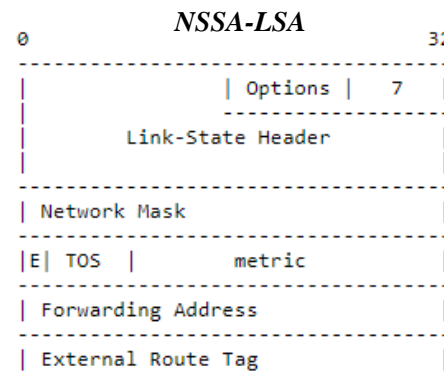
- Trong đồ thị mạng của một AS, Stub Area có vị trí là “lá” của đồ thị. Tức là chỉ có duy nhất một ABR kết nối vùng này với Area 0.
- Đối với các route để đi ra mạng ngoài (được lan truyền bằng External-LSA), đa phần không quan tâm đến Metric thì Stub Area có thể xử lý bằng RTE Default Gateway.
- Tuy nhiên, Stub Area vẫn cần lan truyền các Summary-LSA trong AS để có LS Database thống nhất với các router khác (bên ngoài Stub Area)
- ➔ ABR chặn External-LSA và kích hoạt Summary-LSA với Link State ID = 0.0.0.0 để báo các router bên trong Stub Area áp dụng qui tắc default gateway cho các route đi ra mạng ngoài

■ Totally Stub Area

- Stub Area mà áp dụng cơ chế default gateway không chỉ với các route ra mạng ngoài mà cả các route đi ra các Area khác (tính “totally” ở chỗ này).
- ➔ ABR chặn cả External-LSA và Summary-LSA, đồng thời kích hoạt Summary-LSA với Link State ID = 0.0.0.0 để báo các router bên trong Stub Area áp dụng qui tắc default gateway cho các route đi ra mạng ngoài & các Area khác


■ Not so stubby area (NSSA):

- Stub Area nhưng có kết nối mạng ngoài - RFC3010 (2003)¹
- Do Stub Area đã cấm lan truyền External-LSA (type=5) nên cơ chế External Network này không áp dụng được
- ➔ Đưa ra NSSA-LSA Type=7 để xử lý.
Cấu trúc gói tin và hoạt động của NSSA-LSA giống như External-LSA



[1] The OSPF Not-So-Stubby Area (NSSA) Option: <https://tools.ietf.org/html/rfc3101>

Các loại LSA

Link-state type	OSPF function	
1	Router link states	✓
2	Network link states	✓
3	Summary link states	✓
4	ASBR link state	✓
5	External link advertisement	✓
7	NSSA external link state	✓
8	External attributes for BGP	} 
9, 10, 11	Opaque LSA	

thực hành:

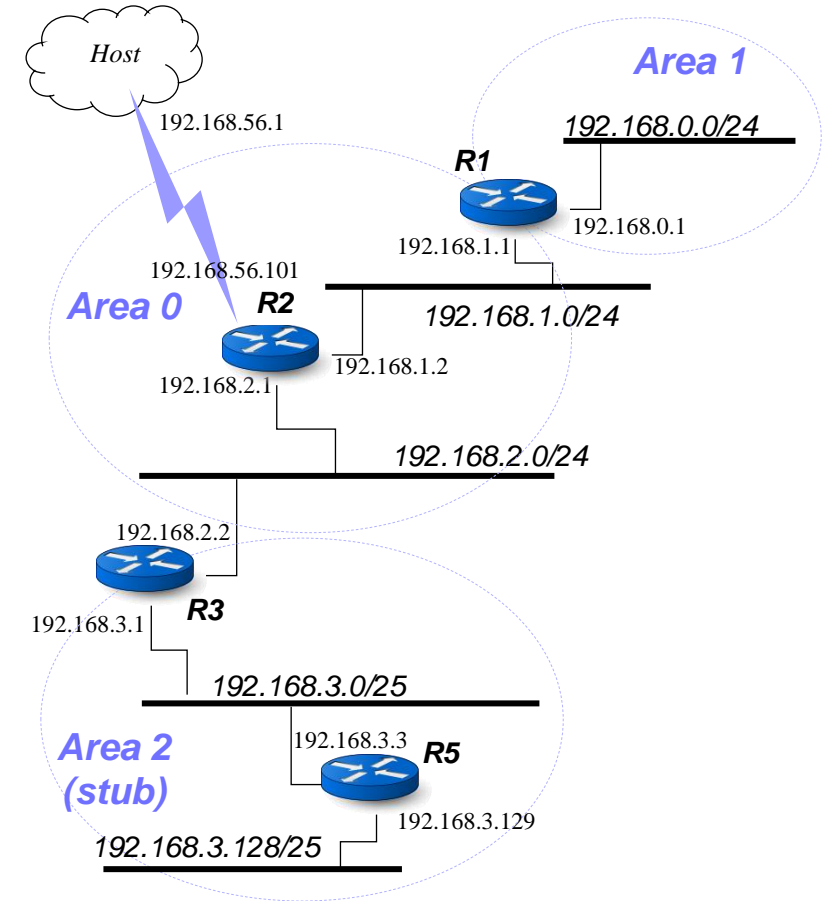
OSPF Multi Area & Tối ưu vùng

- Thiết lập để Area 2 có duy nhất router R3 kết nối ra Area 0, thiết lập Area 2 là Stub Area
- Kiểm tra bảng routing trên R5 thấy xuất hiện default gateway cho route ra mạng ngoài

```
[root@R5 ~]# route -n
Kernel IP routing table
Destination Gateway Genmask Flags Metric Ref Use Iface
192.168.3.0 0.0.0.0 255.255.255.128 U 0 0 0 eth1
192.168.3.128 0.0.0.0 255.255.255.128 U 0 0 0 eth2
192.168.2.0 192.168.3.1 255.255.255.0 UG 20 0 0 eth1
192.168.1.0 192.168.3.1 255.255.255.0 UG 30 0 0 eth1
192.168.0.0 192.168.3.1 255.255.255.0 UG 40 0 0 eth1
0.0.0.0 192.168.3.1 0.0.0.0 UG 11 0 0 eth1
```

- Thiết lập Area 2 là Totally Stub
- Kiểm tra bảng routing trên R5 thấy default gateway được áp dụng cho tất cả các route đi ra ngoài Area (cả mạng ngoài lẫn mạng trong liên vùng)

```
[root@R5 ~]# route -n
Kernel IP routing table
Destination Gateway Genmask Flags Metric Ref Use Iface
192.168.3.0 0.0.0.0 255.255.255.128 U 0 0 0 eth1
192.168.3.128 0.0.0.0 255.255.255.128 U 0 0 0 eth2
0.0.0.0 192.168.3.1 0.0.0.0 UG 11 0 0 eth1
```





Border Gateway Protocol (BGP)

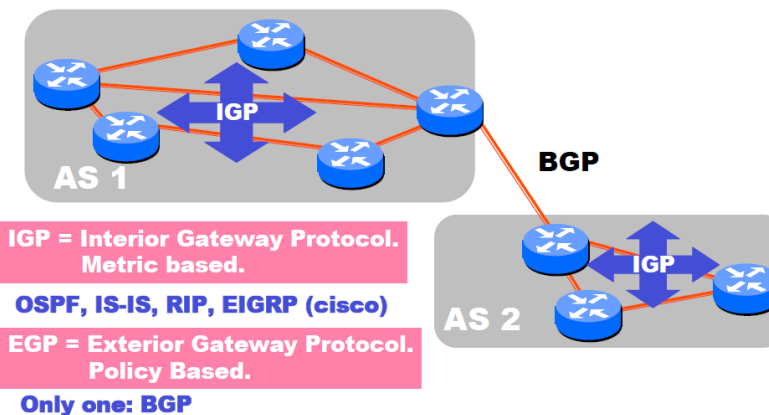
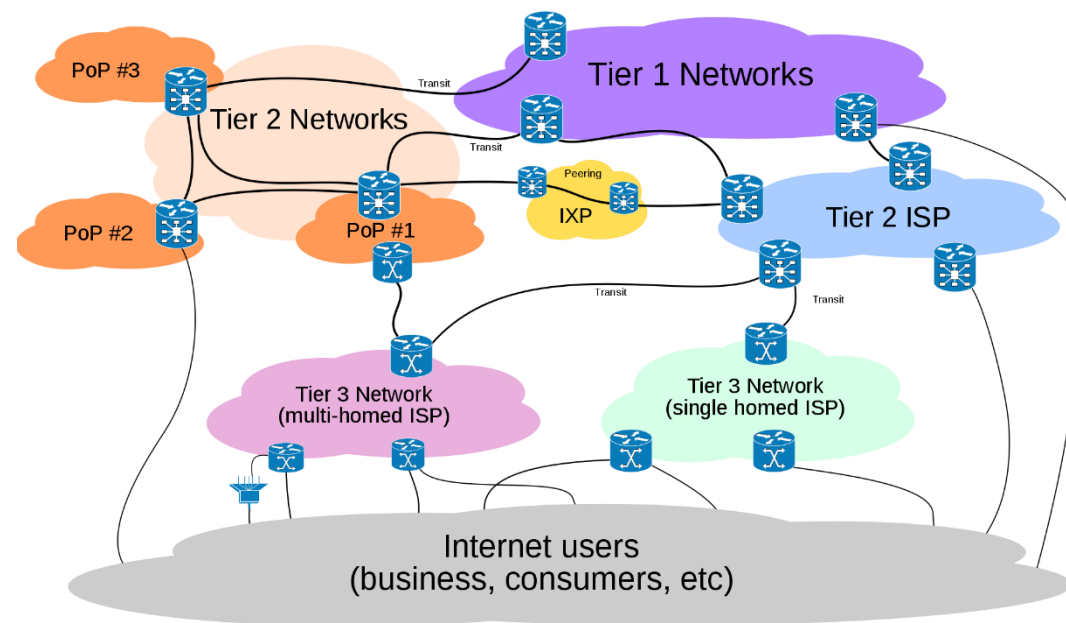
BGP Introduction

- Là “chất keo” kết dính toàn bộ hệ thống Internet hiện nay
- Ra đời năm 1994, phiên bản hiện nay là version 4 chuẩn hóa năm 2006 (RFC4271¹)
- Hỗ trợ CIDR
- Định tuyến theo policy, hơn là theo đường đi ngắn nhất
- “Đơn vị” routing là Autonomous System - tìm đường đi theo các kết nối giữa các AS
- Kết hợp với các IGP (RIP, OSPF, v.v..) trong AS để tạo nên giải pháp dynamic routing hoàn chỉnh trên toàn bộ hệ thống Internet

[1] <https://tools.ietf.org/html/rfc4271>

Tier 1 Networks

- Tier 1 Networks = cấu trúc lõi của Internet, gồm các backbone router mạnh, kết nối nhau bằng đường truyền tốc độ rất cao & phủ sóng toàn cầu.
- Định nghĩa (theo Wikipedia¹): là một mạng IP mà có thể kết nối với tất cả các mạng khác trên Internet.
- Tier 1 Network thường được vận hành bởi một công ty mạng có tiềm lực tài chính rất mạnh.
- Business với Internet Traffic (mua/thuê kết nối)
 - Liên kết Tier 1 khác (peering) để mở rộng vùng phủ sóng. Liên kết peering ở mức Tier 1 dựa trên Peering Policy và Default-free.
 - Kết nối xuống Tier 2 để “bán buôn” traffic. Tier 2 tiếp tục kết nối xuống các Tier 3 (thường là ISP) – down link/up link
 - ISP cuối cùng bán đường truyền kết nối Internet cho end-user
- Không có admin con người nào có thể kiểm soát cấu hình kết nối của các router trong một hệ thống lớn và phức tạp như vậy → BGP đảm nhiệm chức năng kết nối các router của các mạng Tier này, kết hợp với các IGP (RIP, OSPF, v.v..) vận hành bên trong các Tier.



The Routing Domain of BGP is the entire Internet

[1] https://en.wikipedia.org/wiki/Tier_1_network

Danh sách các mạng Tier 1

Name	Headquarters	AS number	CAIDA AS Rank ^[10]	Fiber Route Miles	Fiber Route km	Peering Policy
AT&T^[11]	United States	7018	23	410,000	660,000 ^[12]	AT&T Peering policy
CenturyLink (formerly Level 3 formerly Global Crossing) ^{[13][14]}	United States	3549	12	750,000	885,139 ^{[15][16]}	North America; InternationalLevel 3 Peering Policy
CenturyLink (formerly Level 3) ^{[13][14]}	United States	3356	1	750,000	885,139 ^{[15][16]}	North America; InternationalLevel 3 Peering Policy
Deutsche Telekom Global Carrier^[17]	Germany	3320	20	155,343	250,000 ^[18]	DTAG Peering Details
GTT Communications, Inc.	United States	3257	3	144,738	232,934 ^{[19][20]}	GTT Peering Policy
Liberty Global^{[21][22]}	United Kingdom^[23]	6830	31	500,000	800,000 ^[24]	Peering Principles
NTT Communications (America) (formerly Verio) ^[25]	Japan	2914	5	?	?	North America
Orange (OpenTransit) ^[26]	France	5511	18	?	?	OTI peering policy
PCCW Global	Hong Kong	3491	9	?	?	Peering policy
Sprint (SoftBank Group) ^[27]	Japan	1239	27	26,000	42,000 ^[28]	Peering policy
Tata Communications (formerly Teleglobe) ^[29]	India	6453	6	435,000	700,000 ^[30]	Peering Policy
Telecom Italia Sparkle (Seabone) ^[31]	Italy	6762	8	347,967	560,000	Peering Policy
Telia Carrier^[34]	Sweden	1299	2	40,000	65,000 ^[35]	TeliaSonera International Carrier Global Peering Policy
Telxius (Subsidiary of Telefónica) ^[32]	Spain	12956	14	40,000	65,000 ^[33]	Peering Policy
Verizon Enterprise Solutions (formerly UUNET) ^[40]	United States	701	22	500,000	805,000 ^[41]	Verizon UUNET Peering policy 701, 702, 703
Zayo Group (formerly AboveNet) ^[42]	United States	6461	10	122,000	196,339 ^[43]	Zayo Peering Policy

Nguồn: Wikipedia (https://en.wikipedia.org/wiki/Tier_1_network)

Hoạt động chung của BGP

■ Dựa trên hoạt động láng giềng: 2 loại láng giềng BGP

- Giữa 2 AS: BGP router gửi message trực tiếp cho nhau → eBGP
- Bên trong một AS: BGP router gửi message dựa trên các IGP → iBGP
- BGP láng giềng được khai báo (cấu hình) chứ không phải qua thủ tục tìm kiếm. Các BGP bên trong một AS được khai báo là láng giềng của nhau.

■ AS number (được gán cho AS theo thủ tục đăng ký)

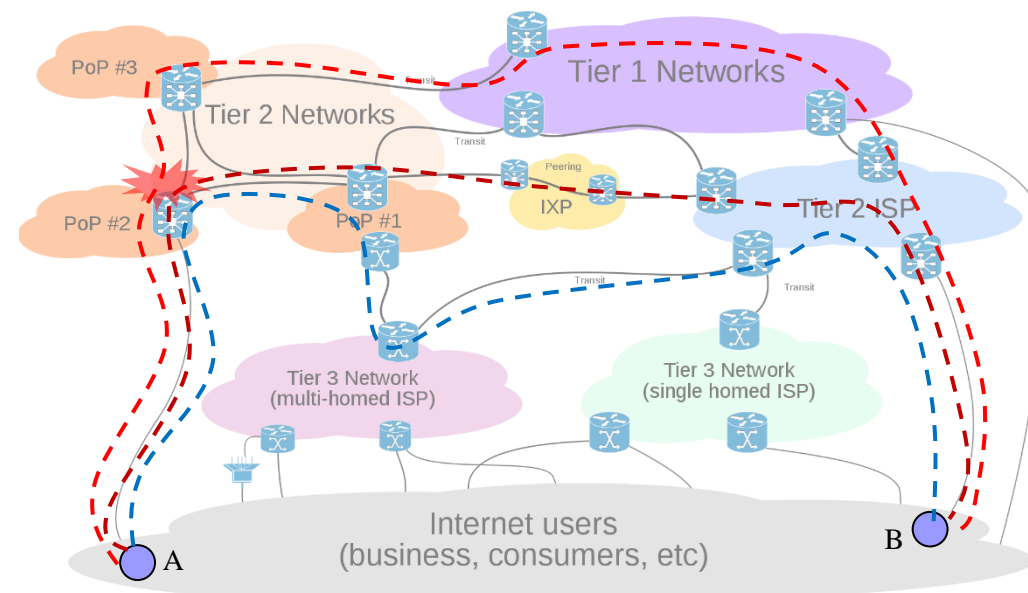
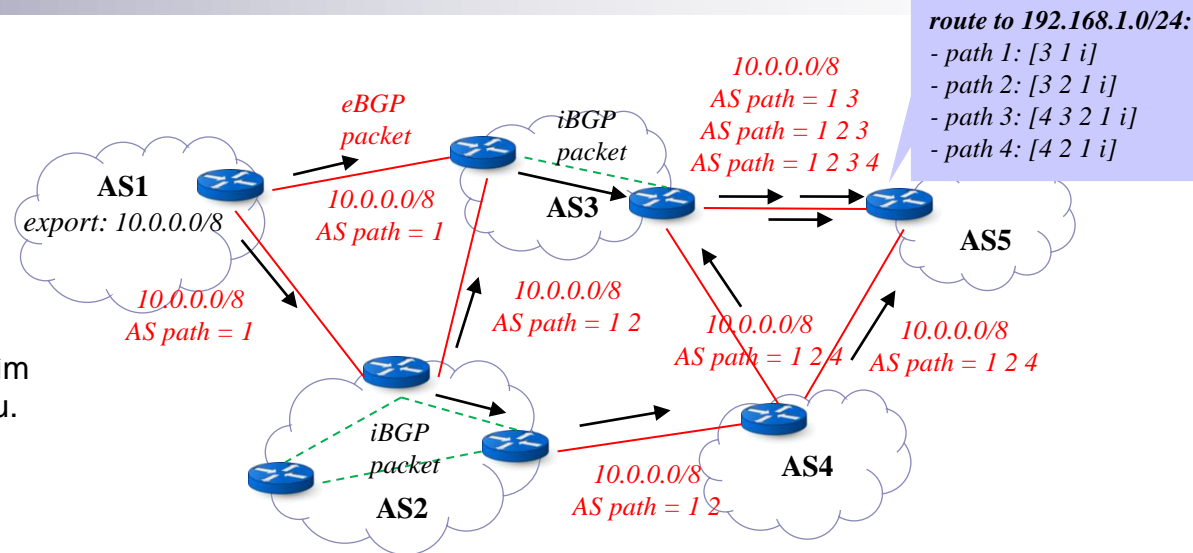
- 16 bit nhị phân. Dải number 64512-65535 được quy hoạch cho private
- Check AS number: <https://www.ultratools.com/tools/asnInfo>
- FPT: AS45894 (Date: 2009-08-28, Owner: FPTONLINE-AS-VN)
- Viettel: AS7552 (Date: 2002-10-08, Owner: VIETTEL-AS-AP Viettel Group)
- Route = AS path (AS1, AS2, AS3, v.v...)

■ BGP speaker & propagation process:

- Sử dụng kênh TCP (cổng 179) để kết nối láng giềng
- Loan báo (speak) các BGP packet đến láng giềng cung cấp khả năng kết nối (reachability) đến một network được "export", đồng thời xây dựng AS path để đi về "exported network" đó
- BGP hoạt động giống RIP ở khâu quảng bá các network prefix và xây dựng đường routing về gốc, tuy nhiên mục đích là loan báo sự tồn tại (khác với mục đích của RIP là hội tụ tạo đường đi có distance vector nhỏ nhất)

■ Quyết định lựa chọn AS-Path theo Policy

- Không nhất thiết là đường đi ngắn nhất
- BGP cho phép xác định nhiều AS-Path để route từ A đến B
- Chọn AS-Path nào là do các mạng Tier áp dụng policy riêng (băng thông, kinh tế, chính trị, v.v..)

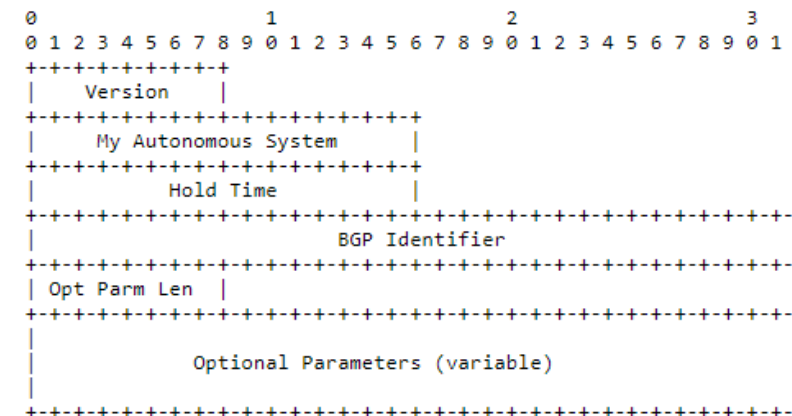


BGP packet format

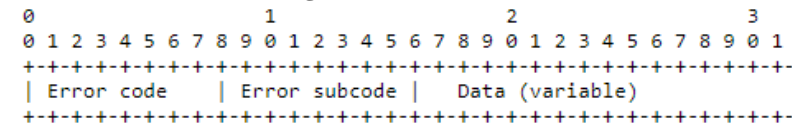
- Cấu trúc: *<Marker(16-octet), Length(2-octet), Type(1-octet)><Data>*
 - Type: 1 - OPEN, 2 - UPDATE, 3 - NOTIFICATION, 4 - KEEPALIVE
- OPEN: được gửi từ mỗi BGP router ngay khi kết nối TCP được thiết lập. KEEPALIVE được trả về nếu router chấp nhận kết nối.
 - *My Autonomous System*: AS number
 - *Hold Time*: Timeout, tính bằng giây. Duy trì liên kết khi chưa nhận được KEEPALIVE hoặc UPDATE hoặc NOTIFICATION.
 - *BGP Identifier*: tương tự Router ID trong OSPF (admin cấu hình hoặc lấy theo IP)
- KEEPALIVE: không có data, chỉ có header với Type=4. Được gửi trước khi Hold time hết hạn, để thông báo duy trì kết nối
- NOTIFICATION: dùng để thông báo lỗi
- UPDATE: quảng bá các đường route có khả năng được thiết lập giữa các router BGP
 - Có thể được sử dụng để thông báo hủy bỏ một số route:
 - *Withdrawn Router Length*: độ dài trường Withdrawn Routers ngay sau
 - *Withdrawn Routers*: danh sách các route cần hủy - *List<length, network prefix>*
 - Cùng một UPDATE message có thể thông báo hủy một số route và thêm một số route có thể được bổ sung hoặc cập nhật:
 - *Total Path Attribute Length*: độ dài trường Path Attributes ngay sau
 - *Path Attributes*: danh sách dạng *List<attribute type, attribute length, attribute value>*
 - *Network Layer Reachability Information*: danh sách địa chỉ IP mạng (cũng dưới dạng bộ đôi *List<length, network prefix>*) mà BGP router có thể route đến (độ dài trường này được tính bằng độ dài gói tin trừ đi độ dài các trường trên)



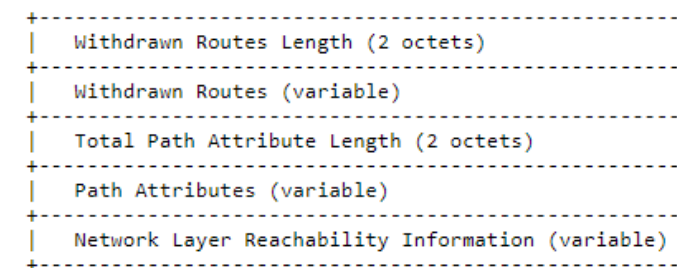
OPEN Message



NOTIFICATION Message



UPDATE Message



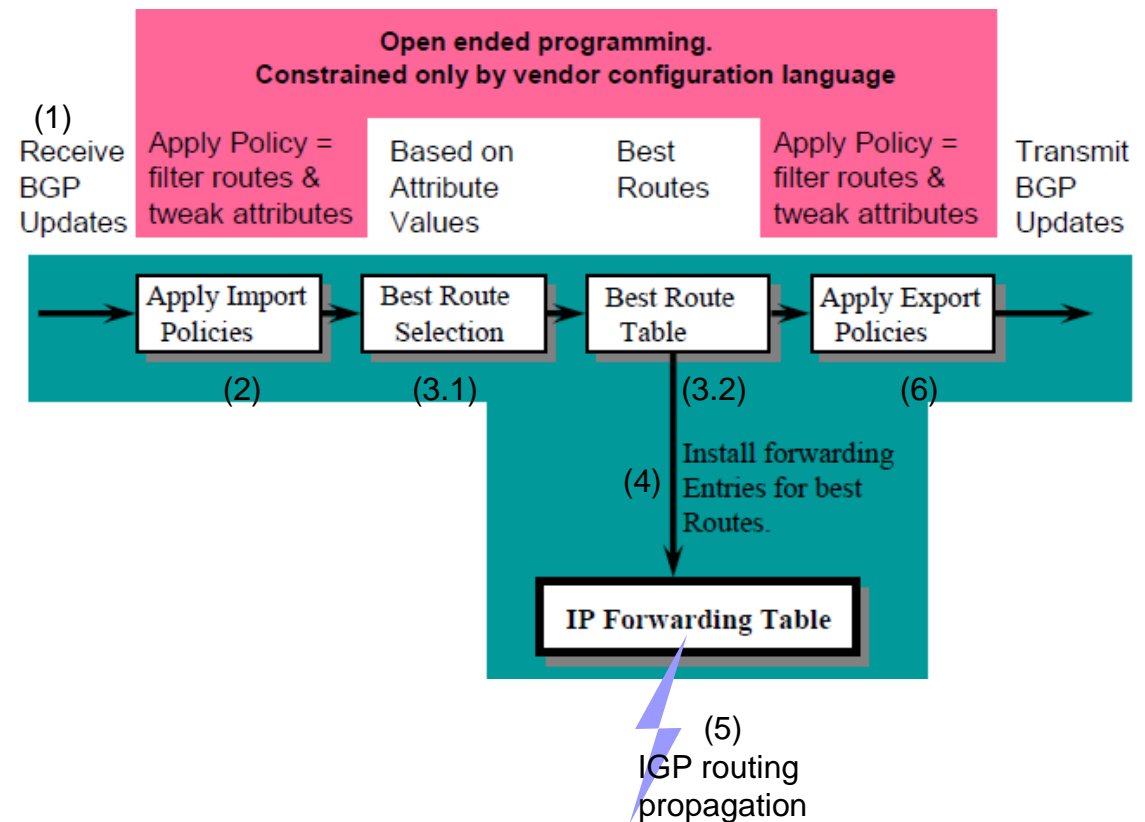
BGP Policy: Decision Process¹

- 1. Weight check: prefer higher local weight routes to lower routes.
- 2. Local preference check: prefer higher local preference routes to lower.
- 3. Local route check: Prefer local routes (statics, aggregates, redistributed) to received routes.
- 4. AS path length check: Prefer shortest hop-count AS_PATHs.
- 5. Origin check: Prefer the lowest origin type route. That is, prefer IGP origin routes to EGP, to Incomplete routes.
- 6. MED check: Where routes with a MED were received from the same AS, prefer the route with the lowest MED. See BGP MED.
- 7. External check: Prefer the route received from an external, eBGP peer over routes received from other types of peers.
- 8. IGP cost check: Prefer the route with the lower IGP cost.
- 9. Multi-path check: If multi-pathing is enabled, then check whether the routes not yet distinguished in preference may be considered equal. If `bgp bestpath as-path multipath-relax` is set, all such routes are considered equal, otherwise routes received via iBGP with identical AS_PATHs or routes received from eBGP neighbours in the same AS are considered equal.
- 10. Already-selected external check: Where both routes were received from eBGP peers, then prefer the route which is already selected. Note that this check is not applied if `bgp bestpath compare-routerid` is configured. This check can prevent some cases of oscillation.
- 11. Router-ID check: Prefer the route with the lowest router-ID. If the route has an ORIGINATOR_ID attribute, through iBGP reflection, then that router ID is used, otherwise the router-ID of the peer the route was received from is used.
- 12. Cluster-List length check: The route with the shortest cluster-list length is used. The cluster-list reflects the iBGP reflection path the route has taken.
- 13. Peer address: Prefer the route received from the peer with the higher transport layer address, as a last-resort tie-breaker.

[1] BGP Decision Process: <https://tools.ietf.org/html/rfc4271#section-9.1>

Xử lý bên trong router BGP

1. Kết nối TCP lắng giềng và nhận gói tin BGP (chứa thông tin export một mạng nội bộ nào đó) → incoming Routing Information Base (RIB-in)
2. Kiểm tra import policy để quyết định có xử lý update RIB-in hay không
3. Xử lý BGP routing process:
 1. Trích xuất thông tin từ RIB-in để cập nhật AS path từ BGP router hiện tại đến mạng nội bộ mà đang được lan tỏa trong RIB-in.
 2. Lựa chọn AS path “phù hợp nhất theo policy”, tính toán next hop và các thông số khác để chuẩn bị cập nhật bảng routing. Lưu ý là “next hop” của AS path khác với next hop đường định tuyến (trường hợp iBGP)
4. Cập nhật tuyến đường đã chọn trong bước 3 cùng với next hop và các thông số liên quan (ví dụ RIP metric hay OSPF cost) vào bảng định tuyến để áp dụng cho giao thức IP – routed protocol
5. Kích hoạt tiến trình lan truyền IGP để quảng bá đến các router nội bộ AS cập nhật thông tin tuyến đường mới vừa được BGP đưa vào bảng định tuyến
6. Áp dụng export policy để xác định RIB-out, phục vụ update cho BGP lắng giềng kế tiếp



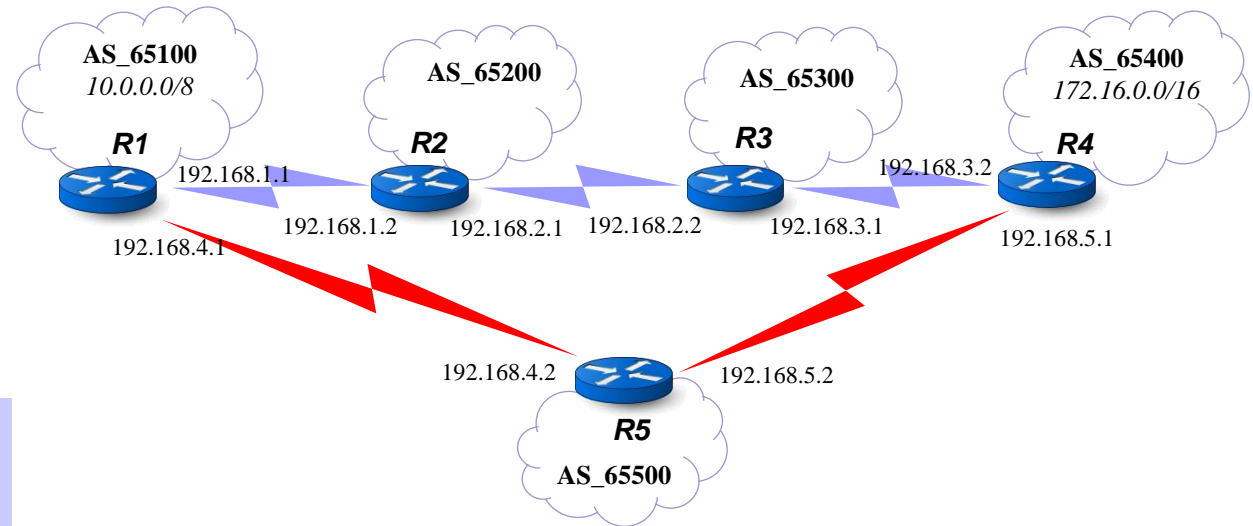
thực hành:

BGP

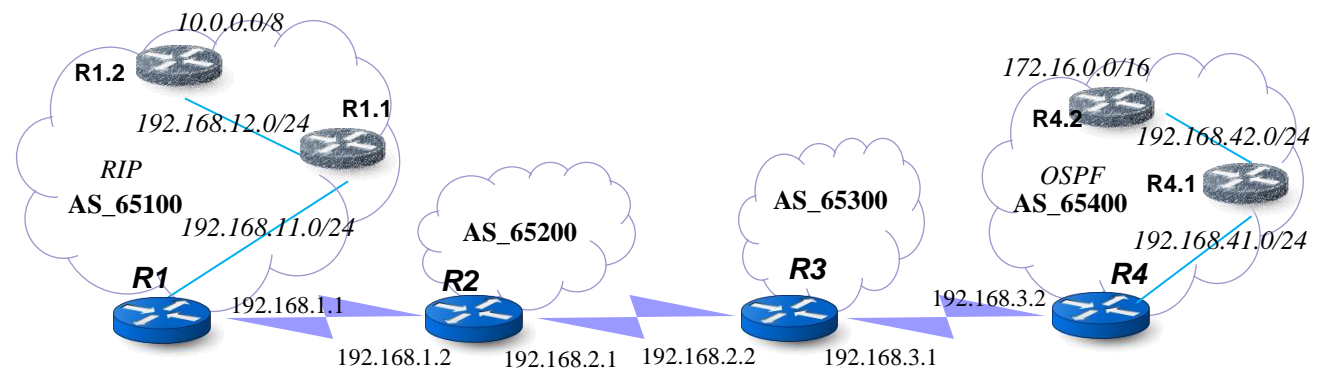
- Các BGP router R1, R2, R3, R4 nối serial với nhau để kết nối các AS 65100, 65200, 65300, 65400
- Tại AS 65100 có mạng 10.0.0.0/8 và tại AS 65400 có mạng 172.16.0.0/16
- Sau khi chạy BGP, các mạng trên đã xuất hiện trong các routing table của tất cả các router

```
R2> route -n
Kernel IP routing table
Destination    Gateway      Genmask      Flags Metric Ref    Use Iface
10.0.0.0       192.168.1.1  255.0.0.0    UG    20     0      0 enp0s8
172.16.0.0     192.168.2.2  255.255.0.0  UG    20     0      0 enp0s9
192.168.1.0    0.0.0.0     255.255.255.0 U    0     0      0 enp0s8
192.168.2.0    0.0.0.0     255.255.255.0 U    0     0      0 enp0s9
```

- Bổ sung AS 65500 với BGP R5 nối R1 & R4. Đường đi từ R1 đến 172.16.0.0/16 được đổi sang R5: do policy chọn AS path bé nhất.
- Thay đổi weight của kết nối R1-R2: 101, kết nối R1-R5: 55. Đường đi được chuyển sang qua R2: do policy chọn weight lớn nhất có mức ưu tiên cao hơn policy AS path nhỏ nhất
- Tích hợp BGP với các IGP (RIP, OSPF) trong AS 65100 và 65400



Bài tập lớn: thêm các router trong từng AS và sử dụng RIP hoặc OSPF để cấu hình IGP trong các AS này. Kiểm tra các đường route được AS học theo BGP đã được tự động đưa vào bảng routing của các router trong AS



eBGP và iBGP

■ IGP (RIP)

- Loan báo thông tin network lần lượt qua các router đôi một láng giềng
- Kết nối routing giữa các network
- Routing next hop = router
- RIP dùng đường loan báo xác định “khoảng cách” và đường routing (bằng cách đi ngược lại đường loan báo) → router RIP láng giềng trực tiếp

■ BGP

- Loan báo thông tin network lần lượt qua các router đôi một láng giềng
- Kết nối routing giữa các AS
- Routing next hop = AS
- BGP loan báo trên kênh kết nối TCP cổng 179 để thiết lập AS path → không cần láng giềng kết nối trực tiếp
- BGP giữa 2 AS: kết nối trực tiếp → eBGP
- BGP bên trong AS: kết nối gián tiếp (sử dụng IGP nội vùng AS để chuyển gói tin giữa 2 router BGP → iBGP

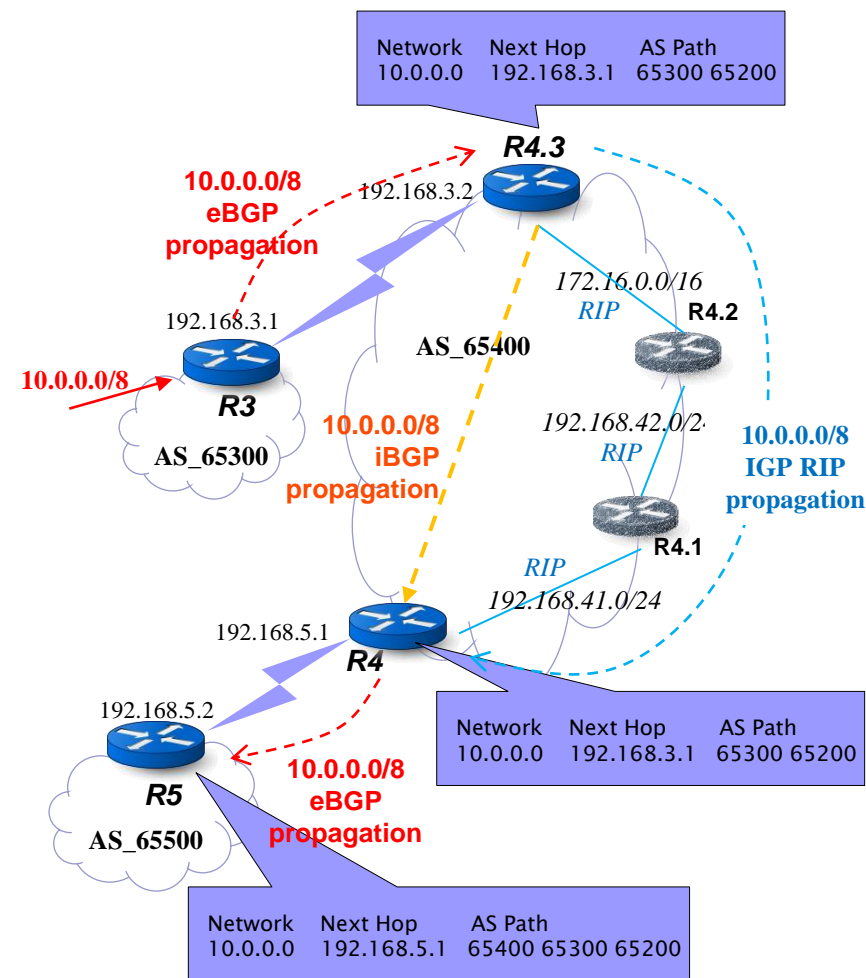
□ eBGP và iBGP đều là giao thức BGP

□ Xử lý phù hợp với môi trường inter-AS và intra-AS

□ Routing inter-AS: next hop (BGP) = next hop (routing)

□ Routing intra-AS: next hop (BGP) != next hop (routing)

□ BGP routing qua một AS (AS_65400 trong hình vẽ) có khả năng không thành công nếu lan tỏa BGP đi trước IGP: thông tin routing cho mạng 10.0.0.0/8 được cập nhật trong BGP R5 trong khi các IGP trong AS 65400 chưa kịp lan tỏa thông tin về mạng này → **vấn đề synchronization!**



thực hành:

eBGP & iBGP

- Tiếp tục bài thực hành BGP, bổ sung thêm trong AS 65400 có 2 BGP router: R4.3 và R4
- Cấu hình kết nối láng giềng giữa các BGP inter-AS như bài trước (là eBGP)
- Cấu hình kết nối láng giềng giữa các BGP R4.3 với R4 bên trong AS 65400 (là iBGP)
- Kiểm tra AS Path đến mạng 10.0.0.0/8 được export từ AS_65100 trên R4.3 và R4: thấy giống nhau (cùng nhận R3 là next hop)
- Cấu hình IGP cho AS 65400 bằng OSPF như bài trước, kiểm tra lan tỏa 10.0.0.0/8 từ bên ngoài AS vào bên trong AS
- Kiểm tra lan tỏa 10.0.0.0/8 từ R4 đến R5

