

Chương 4. Tầng mạng

Tổng quan
Giao thức IP
Định tuyến



1

1. Tổng quan về tầng mạng

Tổng quan
Giao thức IP
Định tuyến



2

Tầng mạng trên kiến trúc phân tầng



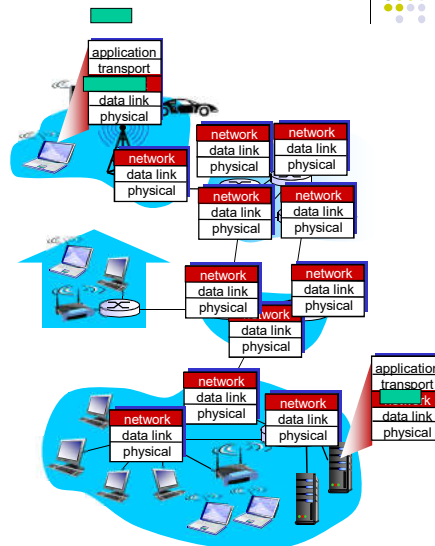
Application (HTTP, Mail, ...)	Hỗ trợ các ứng dụng trên mạng
Transport (UDP, TCP ...)	Điều khiển truyền dữ liệu giữa các tiến trình của tầng ứng dụng
Network (IP, ICMP...)	Điều khiển truyền dữ liệu giữa các nút mạng qua môi trường liên mạng
Datalink (Ethernet, ADSL...)	Hỗ trợ việc truyền thông cho các thành phần kế tiếp trên cùng 1 mạng
Physical (bits...)	Truyền và nhận dòng bit trên đường truyền vật lý

3

Tầng mạng



- Truyền dữ liệu từ host-host
- Cài đặt trên mọi hệ thống cuối và bộ định tuyến
- Đơn vị truyền: datagram
- Bên gửi: nhận dữ liệu từ tầng giao vận, đóng gói
- Bên nhận: mở gói, chuyển phần dữ liệu trong payload cho tầng giao vận
- Bộ định tuyến: định tuyến và chuyển tiếp



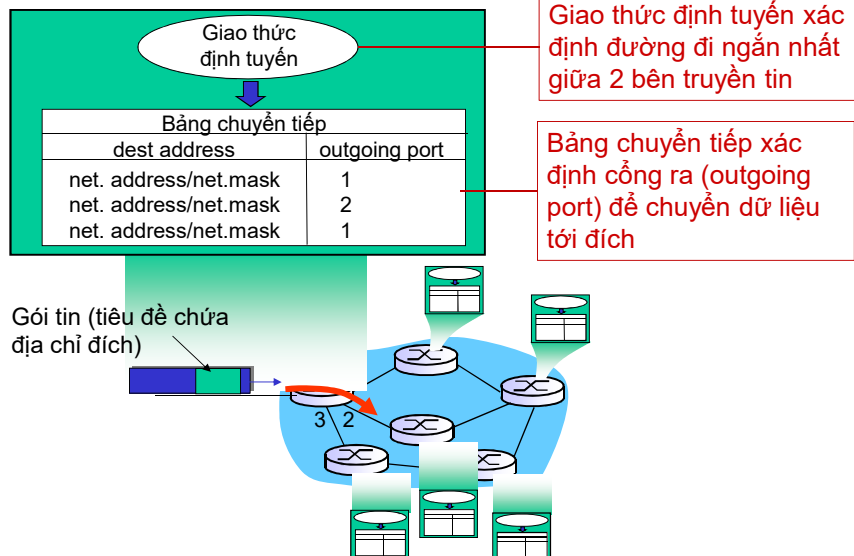
4

Chức năng chính

- **Định tuyến (Routing)**: Tìm tuyến đường (qua các nút trung gian) để gửi dữ liệu từ nguồn tới đích
- **Chuyển tiếp (Forwarding)**: Chuyển gói tin trên cổng vào tới cổng ra theo tuyến đường
- **Định địa chỉ (Addressing)**: Định danh cho các nút mạng
- **Đóng gói dữ liệu (Encapsulating)**: Nhận dữ liệu từ giao thức ở trên, thêm tiêu đề mang thông tin điều khiển quá trình truyền dữ liệu từ nguồn tới đích
- **Đảm bảo chất lượng dịch vụ (QoS)**: đảm bảo các thông số phù hợp của đường truyền theo từng dịch vụ

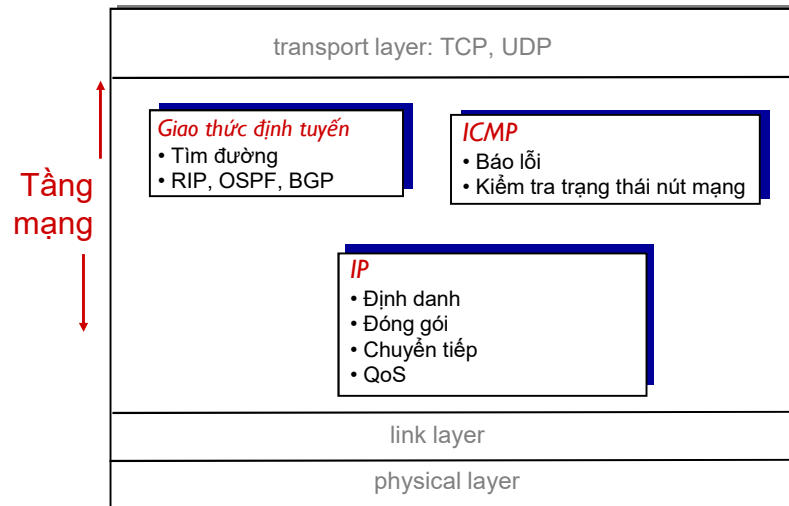
5

Định tuyến và chuyển tiếp



6

Các giao thức tầng mạng



7

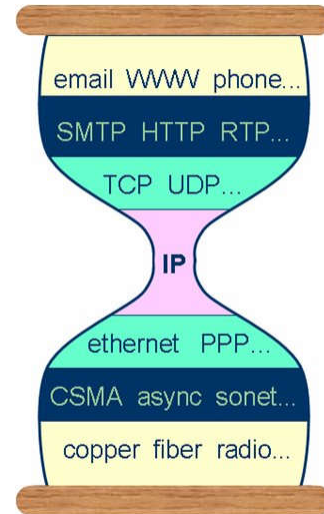
2. Giao thức IP



8

2.1. Đặc điểm giao thức

- Là giao thức cơ sở của tầng mạng
 - Kết nối liên mạng
- Là giao thức được định tuyến (**routed protocol**)
 - Đòi hỏi phải có các giao thức định tuyến để xác định trước đường đi cho dữ liệu.
- Giúp ứng dụng tầng trên không phụ thuộc vào tầng dưới



9

Đặc điểm của giao thức IP

- Giao thức hướng không liên kết
- Các gói tin được xử lý độc lập
- Không tin cậy / nhanh
 - Truyền dữ liệu theo phương thức “*best effort*”
 - IP không có cơ chế phục hồi nếu có lỗi
 - Khi cần, ứng dụng sẽ sử dụng dịch vụ tầng trên để đảm bảo độ tin cậy (TCP)

10

Chức năng cơ bản của IP



- Định địa chỉ: địa chỉ IP
- Đóng gói dữ liệu
 - Dồn kênh/Phân kênh
- Chuyển tiếp: theo địa chỉ IP (sẽ đề cập trong phần sau)
- Đảm bảo chất lượng dịch vụ

11

2.2. Địa chỉ IPv4

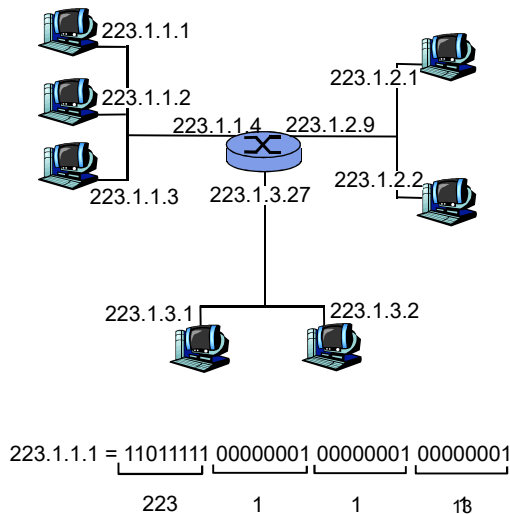
Lớp địa chỉ IP
CIDR – Địa chỉ IP không phân lớp
Mạng con và mặt nạ mạng
Các địa chỉ IP đặc biệt



12

Địa chỉ IP (IPv4)

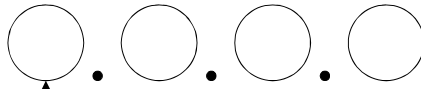
- **Địa chỉ IP:** Một số 32-bit để định danh cổng giao tiếp mạng trên nút đầu cuối (PC, server, smart phone), bộ định tuyến
- Mỗi địa chỉ IP được gán cho một cổng duy nhất
- Địa chỉ IP có tính duy nhất trong mạng



Làm thế nào để cấp phát địa chỉ IP?

- Cấp phát cố định(Static IP):
 - Windows: Control Panel → Network → Configuration → TCP/IP → Properties
 - Linux: /etc/network/interfaces
- Cấp phát tự động: **DHCP- Dynamic Host Configuration Protocol**

Biểu diễn địa chỉ IPv4



8 bits
0 – 255 integer

Ví dụ:

203.178.136.63

o

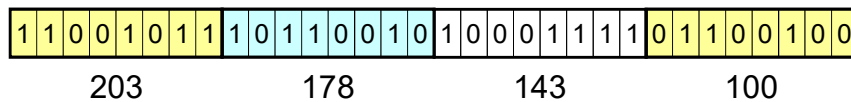
259.12.49.192

x

133.27.4.27

o

Sử dụng 4 phần 8 bits để miêu tả một địa chỉ 32 bits

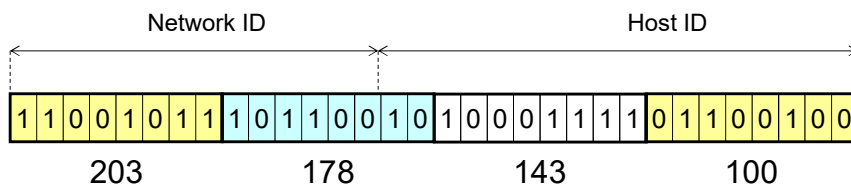


15

Địa chỉ IPv4



- Địa chỉ IP có hai phần
 - Host ID – phần địa chỉ máy trạm
 - Network ID – phần địa chỉ mạng



- Làm thế nào biết được phần nào là cho máy trạm, phần nào cho mạng?
 - Phân lớp địa chỉ
 - Không phân lớp – CIDR

16

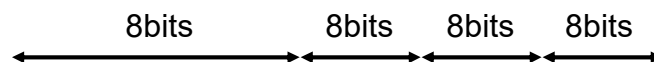
Các dạng địa chỉ



- Địa chỉ mạng (Network Address):
 - Định danh cho một mạng
 - Tất cả các bit phần HostID là 0
- Địa chỉ quảng bá (Broadcast Address)
 - Địa chỉ dùng để gửi dữ liệu cho tất cả các máy trạm trong mạng
 - Tất cả các bit phần HostID là 1
- Địa chỉ máy trạm (Unicast Address)
 - Gán cho một cổng mạng
- Địa chỉ nhóm (Multicast address): định danh cho nhóm

17

Phân lớp địa chỉ IP(Classful Addressing)



Class A	0	7 bit	H	H	H
Class B	1 0	6 bit	N	H	H
Class C	1 1 0	5 bit	N	N	H
Class D	1 1 1 0	Multicast			
Class E	1 1 1 1	Reserve for future use			

	# of network	# of hosts/1 net
Class A	128	$2^{24} - 2$
Class B	16384	65534
Class C	2^{21}	254

18

Hạn chế của việc phân lớp địa chỉ



- Lãng phí không gian địa chỉ
 - Việc phân chia cứng thành các lớp (A, B, C, D, E) làm hạn chế việc sử dụng toàn bộ không gian địa chỉ

Cách giải quyết ...

- CIDR: **C**lassless **I**nter **D**omain **R**outing
 - Classless addressing
 - Phần địa chỉ mạng sẽ có độ dài bất kỳ
 - Dạng địa chỉ: **m1.m2.m3.m4 /n**, trong đó n (mặt nạ mạng) là số bit trong phần ứng với địa chỉ mạng

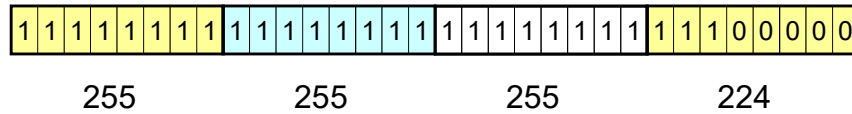
19

Mặt nạ mạng



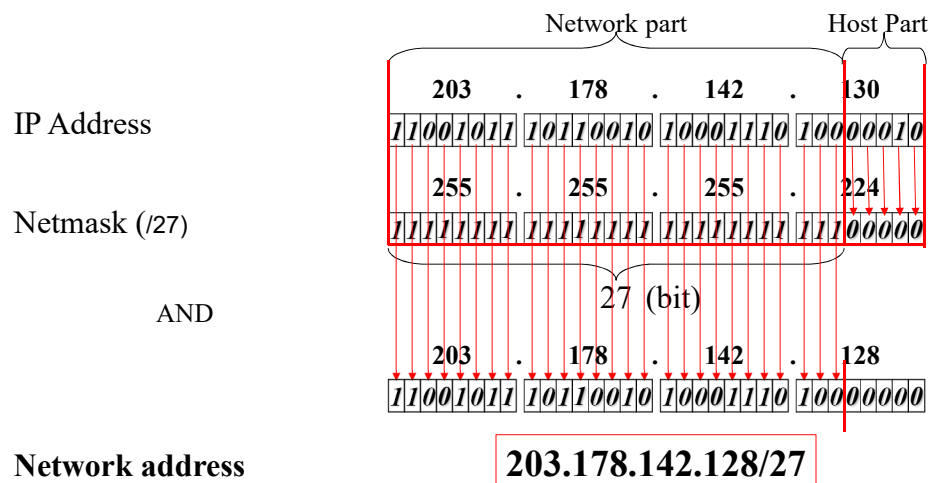
- Mặt nạ mạng chia một địa chỉ IP làm 2 phần
 - Phần ứng với máy trạm
 - Phần ứng với mạng
- Dùng toán tử AND
 - Tính địa chỉ mạng
 - Tính khoảng địa chỉ IP

20



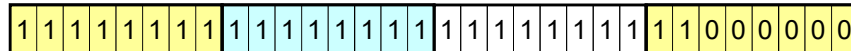
- Thập phân có chấm 255.255.255.224
- Prefix: /27
- Hexa: 0xFFFFFe0

21



22

Mặt nạ mạng và kích thước mạng



255

255

255

192

- Kích thước
 - Theo lũy thừa 2
- [RFC1878](#)
- Trong trường hợp /26
 - Phần máy trạm = 6 bits
 - $2^6=64$
 - Dải địa chỉ có thể gán:
 - 0 - 63
 - 64 - 127
 - 128 - 191
 - 192 - 255

23

Địa chỉ IP và mặt nạ mạng

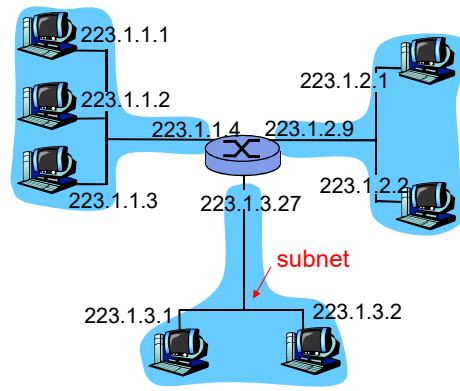


- Địa chỉ nào là địa chỉ máy trạm, địa chỉ mạng, địa chỉ quảng bá?
 - (1) 203.178.142.128 /25
 - (2) 203.178.142.128 /24
 - (3) 203.178.142.127 /25
 - (4) 203.178.142.127 /24
- Lưu ý: Với cách địa chỉ hóa theo CIDR, địa chỉ IP và mặt nạ mạng luôn phải đi cùng nhau

24

Mạng con - subnet

- Là một phần của một mạng nào đó
 - ISP thường được gán một khối địa chỉ IP
 - Một vài mạng con sẽ được tạo ra
- Tạo subnet như thế nào
 - Sử dụng một mặt nạ mạng dài hơn



Mạng với 3 mạng con

25

Ví dụ: Chia làm 2 subnets

11001000 00010111 00010000 00000000
200. 23. 16. 0 /24



11001000 00010111 00010000 00000000 /25
200. 23. 16. 0
11001000 00010111 00010000 10000000 /25
200. 23. 16. 128

26

Ví dụ: Chia làm 4 subnets



- Mạng với mặt nạ /24
- Cần tạo 4 mạng con
 - Mạng với 14 máy tính
 - Mạng với 30 máy tính
 - Mạng với 31 máy tính
 - Mạng với 70 máy tính



/28
/27
/26
/25



27

Không gian địa chỉ IPv4



- Theo lý thuyết
 - Có thể là 0.0.0.0 ~ 255.255.255.255
 - Một số địa chỉ đặc biệt
- Địa chỉ IP đặc biệt ([RFC1918](#))

Private address	10.0.0.0/8
	172.16.0.0/16 → 172.31.0.0/16
	192.168.0.0/24 → 192.168.255.0 /24
Loopback address	127.0.0.0 /8
Multicast address	224.0.0.0
	~239.255.255.255

- Địa chỉ liên kết nội bộ: 169.254.0.0/16 (tự động cấu hình)

28

Quản lý địa chỉ IP công cộng



- Internet Corporation for Assigned Names and Numbers (ICANN): quản lý toàn bộ tài nguyên địa chỉ IP
- Regional Internet Registries: quản lý địa chỉ IP theo vùng (châu Á-Thái Bình Dương, châu Âu và Trung Đông, châu Phi, Bắc Mỹ, Nam Mỹ)
- Cơ quan quản lý quốc gia
 - Việt Nam: VNNIC
- Nhà cung cấp dịch vụ (ISP)
- Cơ quan, tổ chức
- Ví dụ

ICANN → APNIC → VNNIC → HUST

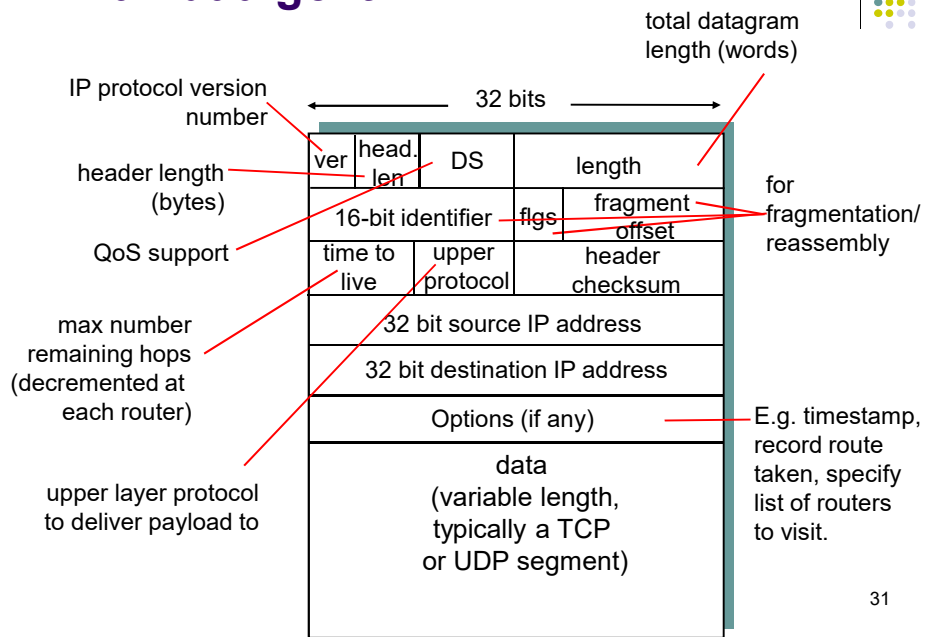
29

2.3. Khuôn dạng gói tin IP



30

Phần đầu gói tin IP



IP header (1)

- Version: Phiên bản giao thức (4 bits)
 - IPv4
 - IPv6
- Header length: Độ dài phần đầu: 4bits
 - Tính theo từ (4 bytes)
 - Min: 5 x 4 (byte)
 - Max: 15 x 4 (byte)
- DS (Differentiated Service : 8bits)
 - Tên cũ: Type of Service
 - Hiện tại được sử dụng trong quản lý QoS

32

IP header (2)



- Length: Độ dài toàn bộ, tính cả phần đầu (16 bits)
 - Theo bytes
 - Max: 65536
- Identifier – Số hiệu gói tin (16 bit)
 - Dùng để xác định một chuỗi các gói tin của một gói tin bị phân mảnh
- Flag – Cờ báo phân mảnh(3 bit)
- Fragmentation offset – Vị trí gói tin phân mảnh trong gói tin ban đầu (13 bit)

33

IP header (3)



- TTL, 8 bits – Thời gian sống
 - Độ dài đường đi gói tin có thể đi qua
 - Max: 255
 - Router giảm TTL đi 1 đơn vị khi xử lý
 - Gói tin bị hủy nếu TTL bằng 0
- Upper protocol – giao thức tầng trên
 - Giao thức giao vận phía trên (TCP, UDP,...)
 - Các giao thức tầng mạng khác (ICMP, IGMP, OSPF) cũng có trường này
 - Sử dụng để dẫn kênh/phân kênh

34

IP header và trường Protocol



Ver	HLEN	DS	Total Length	
Identification			Flags	Fragmentation offset
TTL	Protocol	Header Checksum		
Source IP address				
Destination IP address				
Option				

Protocol:

1: ICMP
2: IGMP
6: TCP
17: UDP
89: OSPF

Có thể xem số hiệu giao thức tại

/etc/protocols
C:\WINDOWS\system32\drivers\etc\protocols

35

IP header (4)

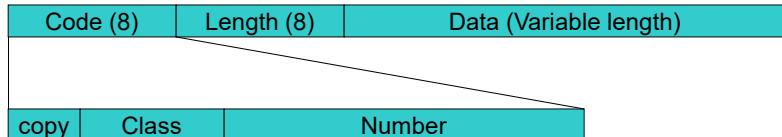


- Checksum – Mã kiểm soát lỗi
- Địa chỉ IP nguồn
 - 32 bit, địa chỉ của trạm gửi
- Địa chỉ IP đích
 - 32 bit, địa chỉ của trạm đích

36

IP header (5)

- Options: Dùng để thêm vào các chức năng mới.
- Có thể tới 40 bytes



Copy:

- 0: copy only in first fragment
- 1: copy into all fragment

Class:

- 00: Datagram control
- 01: Reserved
- 10: Debugging and measurement
- 11: Reserved

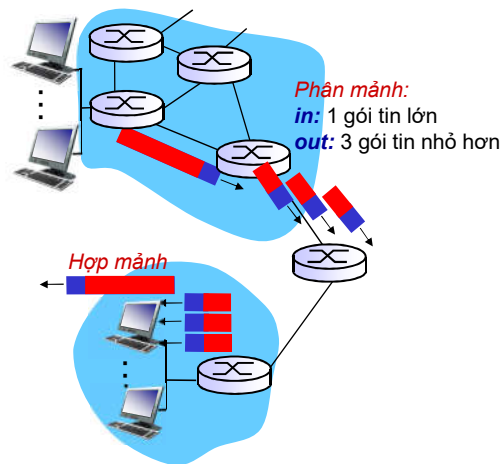
Number:

- 00000: End of option
- 00001: No operation
- 00011: Loose source route
- 00100: Timestamp
- 00111: Record route
- 01001: Strict source route

37

Phân mảnh gói tin (1)

- Đường truyền có một giá trị MTU (Kích thước đơn vị dữ liệu tối đa)
- Các đường truyền khác nhau có MTU khác nhau
- Một gói tin IP có kích thước lớn quá MTU sẽ bị
 - Chia làm nhiều gói tin nhỏ hơn
 - Được tập hợp lại tại trạm đích



38

Phân mảnh (2)



- Trường Identification
 - ID được sử dụng để tìm các phần của gói tin
- Flags – cờ (3 bits)
 - 1st bit: Dự phòng
 - 2nd bit: Không được phép phân mảnh
 - 3rd bit: Còn phân mảnh
- Độ lệch - Offset
 - Vị trí của gói tin phân mảnh trong gói tin ban đầu
 - Theo đơn vị 8 bytes

39

Phân mảnh (3)



Ví dụ:

- ❖ Gói tin: 4020 byte (header: 20 byte)
- ❖ MTU = 1500 bytes

1480 bytes in data field

offset =
1480/8

length	ID	fragflag	offset
=4020	=x	=0	=0

length	ID	fragflag	offset
=1500	=x	=1	=0

length	ID	fragflag	offset
=1500	=x	=1	=185

length	ID	fragflag	offset
=1060	=x	=0	=370

40

2.4. Chuyển tiếp gói tin IP



41

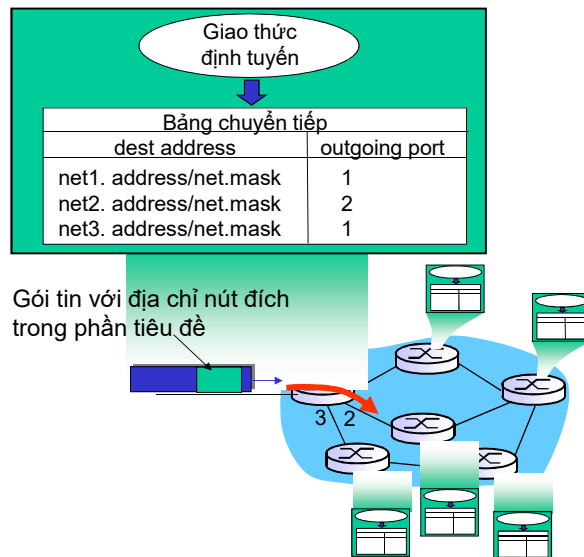
Chuyển tiếp gói tin IP



- Mỗi nút mạng sử dụng bảng chuyển tiếp (Forwarding Table)
 - Là một phần của bảng định tuyến (Routing Table)
- Các thông tin:
 - Đích đến (Destination): Địa chỉ mạng/Mặt nạ (/n)
 - Sử dụng địa chỉ 0.0.0.0/0 đại diện cho một đích bất kỳ chưa biết → lỗi ra mặc định
 - Cổng ra (Outgoing port): địa chỉ của cổng ra trên router để chuyển tới nút kế tiếp trong đường đi

42

Bảng chuyển tiếp



43

Ví dụ - Bảng chuyển tiếp trên máy trạm

C:\Documents and Settings\tungbt>netstat -r
Route Table

Interface List

0x1 ... Destination ... MS TCP Loopback interface
0x2 ... Outgoing port ... Realtek RTL8139 Family PCI Fast Ethernet NIC -

Network	Netmask	Gateway	Interface	Metric
0.0.0.0	0.0.0.0	192.168.1.1	192.168.1.34	20
127.0.0.0	255.0.0.0	127.0.0.1	127.0.0.1	1
192.168.1.0	255.255.255.0	192.168.1.34	192.168.1.34	20
192.168.1.34	255.255.255.255	127.0.0.1	127.0.0.1	20
192.168.1.255	255.255.255.255	192.168.1.34	192.168.1.34	20
224.0.0.0	240.0.0.0	192.168.1.34	192.168.1.34	20
255.255.255.255	255.255.255.255	192.168.1.34	192.168.1.34	1

Default Gateway: 192.168.1.1

44

Ví dụ - Bảng chuyển tiếp trên router (rút gọn)



Router# **show ip route**

Destination Outgoing port

O 203.238.37.0/24 via 203.178.136.14, FastEthernet0/1

O 203.238.37.96/27 via 203.178.136.26, Serial0/0/0

C 203.238.37.128/27 is directly connected, Serial0/0/0

O 192.68.132.0/24 via 203.178.136.14, FastEthernet0/1

C 203.254.52.0/24 is directly connected, FastEthernet0/1

C 202.171.96.0/24 is directly connected, Serial0/0/1

45

Kết hợp đường đi



- Điều kiện:
 - Đường đi có chung Outgoing Port và k bit đầu tiên (prefix) trong phần NetworkID của địa chỉ đích giống nhau
 - Điều gì xảy ra nếu kết hợp không đầy đủ các mạng con được chia từ một mạng lớn hơn. Ví dụ kết hợp 1.1.0.0/16 và 1.2.0.0/16 thành 1.0.0.0/14?
- Mục đích: giảm kích thước bảng chuyển tiếp, tăng tốc quá trình tìm kiếm đường ra

Destination	Outgoing Port
1.1.0.0 /16	Se0/2
1.2.0.0 /16	Se0/2
1.3.0.0 /16	Se0/1
171.8.0.0 /19	Se0/1
171.8.32.0 /19	Se0/1

Destination	Outgoing Port
1.1.0.0 /16	Se0/2
1.2.0.0 /16	Se0/2
1.3.0.0 /16	Se0/1
171.8.0.0 /18	Se0/1

46

Nguyên tắc tìm kiếm



- So sánh n bit đầu tiên trên địa chỉ đích gói tin với các bit tương ứng trên địa chỉ mạng đích
 - /n: Mặt nạ mạng đích
 - Nếu có mạng đích khớp chuyển ra cổng tương ứng
 - Nếu không có mạng đích nào khớp, chuyển ra cổng mặc định (nếu có)
- Quy tắc “longest matching”: nếu có nhiều mạng đích thỏa mãn, chuyển tiếp tới mạng đích có mặt nạ lớn nhất

Địa chỉ đích của gói tin:
11.1.2.10

Destination	Outgoing Port
11.0.0.0 /8	Se0/1
11.1.0.0 /16	Se0/2
11.1.2.0/24	Se0/3

47

Chuyển tiếp gói tin trên router



- B1 : Nếu TTL = 1(hoặc TTL = 0), hủy gói tin và báo lỗi. Kết thúc.
- B2 : Nếu TTL >1, lấy địa chỉ đích DA của gói tin. Mặt nạ hóa địa chỉ đích của gói tin với các mặt nạ của mạng đích
- B3 : So sánh kết quả mặt nạ hóa với địa chỉ mạng đích tương ứng. Nếu có mạng đích khớp chuyển tới cổng ra tương ứng, giảm TTL.
- B4 : Nếu không có mạng đích khớp, kiểm tra cổng ra mặc định (tương ứng với đích 0.0.0.0 /0)
 - Có cổng mặc định : chuyển gói tin tới cổng mặc định, giảm TTL
 - Không có : hủy gói tin, báo lỗi.

48

Ví dụ



- Cho bảng chuyển tiếp của một router trong bảng 1. Hãy cho biết router xử lý như thế nào khi nhận được các gói tin có các thông số trong bảng 2.

Destination	Outgoing Port
45.45.160.0 /19	E3
45.45.64.0 /19	E1
45.45.96.0 /20	E3
45.64.0.0 /10	E2
45.45.96.128 /26	E1
0.0.0.0 /0	E1

Destination Address	TTL
45.45.100.64	0
45.45.100.64	5
45.45.145.144	15

49

NAT

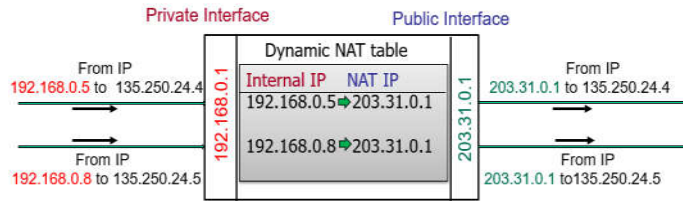


- Dữ liệu chuyển tiếp từ mạng LAN(sử dụng địa chỉ cục bộ) sang mạng Internet(sử dụng địa chỉ công cộng) và ngược lại cần được chuyển đổi địa chỉ
- Network Address Translation
- PAT : Port Address Translation
 - NAT with overloading sử dụng thêm số hiệu cổng ứng dụng trong quá trình chuyển đổi
- Trên thực tế, có thể sử dụng NAT để chuyển đổi địa chỉ IP từ mạng LAN này sang mạng LAN khác

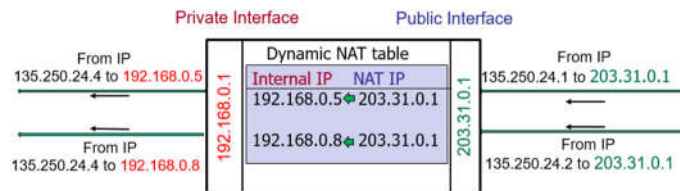
50

Hoạt động của NAT

- Gói tin đi từ trong mạng ra ngoài



- Gói tin đi từ ngoài vào trong mạng



51

2.5. Giao thức IPv6



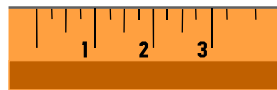
52

Giao thức IPv6



- Xuất phát từ nhu cầu thực tế: địa chỉ IPv4 cạn kiệt, không đủ để cấp phát
- Cải tiến trên IPv6:
 - Mở rộng không gian địa chỉ
 - Sử dụng địa chỉ có độ dài 128 bit
 - Phân vùng địa chỉ(scope)

Không gian địa chỉ IPv4



1mm

Không gian địa chỉ IPv6



84.000 lần đường kính của thiên hà

53

Cải tiến trên IPv6(tiếp)



- Tăng tốc độ
 - Khuôn dạng header đơn giản hơn
 - Ít trường hơn
 - Độ dài cố định
 - Bỏ checksum
 - Không phân mảnh gói tin
- Hỗ trợ QoS tốt hơn
- Về an toàn an ninh
 - Sử dụng IPSec như một chuẩn
- Tự động cấu hình
 - Chuẩn hóa cơ chế tự động cấu hình

54

Địa chỉ IPv6



55

Cách thức biểu diễn



- 128 bit, biểu diễn bởi số hệ 16
- Phân cách “:” giữa các nhóm gồm 4 số hexa
 - 3ffe:501:100c:e320:2e0:18ff:fe98:936d
- Bỏ qua chuỗi liên tiếp các số 0
 - 3ffe:501:100c:e320:0:0:0:1 →
3ffe:501:100c:e320::1
- Sử dụng ký hiệu mặt nạ mạng /n

56

Cấu trúc địa chỉ IPv6



- 2 phần: Network prefix và Interface ID
- Network prefix: 64 bit
- Host ID: 64 bit
 - Chuẩn EUI-64 (*extended unique identifier*)
 - Trong trường hợp mạng Ethernet, Host ID được xác định từ địa chỉ MAC

57

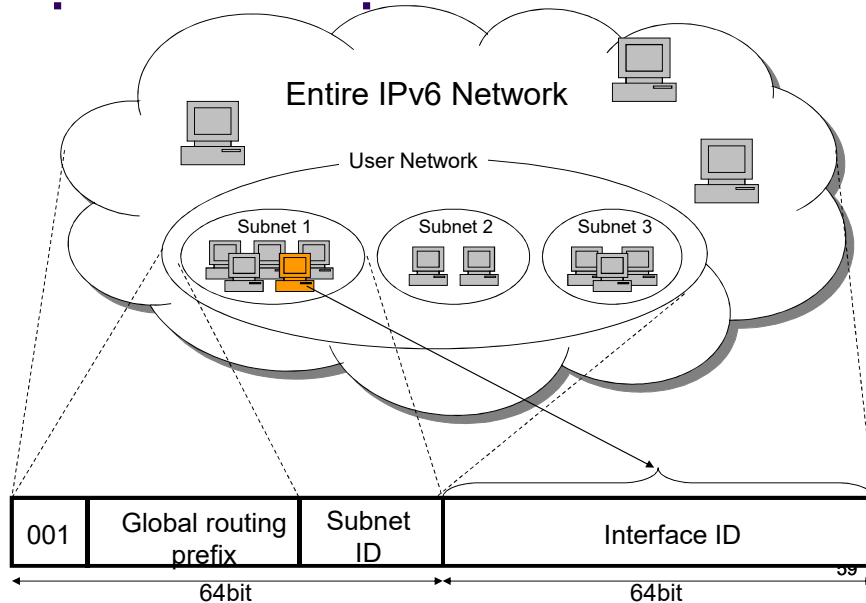
Phân vùng địa chỉ



- Địa chỉ toàn cục: Network prefix có
 - Global routing prefix: 48 bit với 3 bit đầu là 001
 - SubnetID: 16 bit
- Địa chỉ liên kết nội bộ: Network prefix có
 - 10 bit đầu là 1111 1110 10
 - 54 bit còn lại mang giá trị 0
- Địa chỉ cục bộ: Network prefix có
 - 10 bit đầu là 1111 1110 11
 - 38 bit kế tiếp mang giá trị 0
 - SubnetID: 16 bit

58

Địa chỉ toàn cục



Các dạng địa chỉ IPv6

- Địa chỉ Unicast Address: gán cho một cổng giao tiếp mạng
- Địa chỉ Anycast Address: gán cho một nhóm cổng giao tiếp mạng
 - Gói tin gửi tới địa chỉ anycast addr. được chuyển tiếp cho nút gần nhất xác định bởi giao thức định tuyến
- Địa chỉ Multicast Address: gán cho một nhóm cổng giao tiếp mạng trong một scope
 - Bắt đầu bởi 1111 1111
 - Gói tin gửi tới địa chỉ multicast addr. được chuyển tới tất cả các nút trong nhóm

Khuôn dạng gói tin IPv6



61

Khuôn dạng gói tin IPv6



Version (4 bit)	Traffic Class (8 bit)	Flow Label (20 bit)	
Payload Length (16 bit)		Next Header (8 bit)	Hop Limit (8 bit)
Source address (128bit)			
Destination address (128bit)			
Payload(gồm tiêu đề mở rộng nếu có và gói tin của giao thức tầng trên)			

Các trường đổi tên từ IPv4

62

Khuôn dạng gói tin IPv6(tiếp)



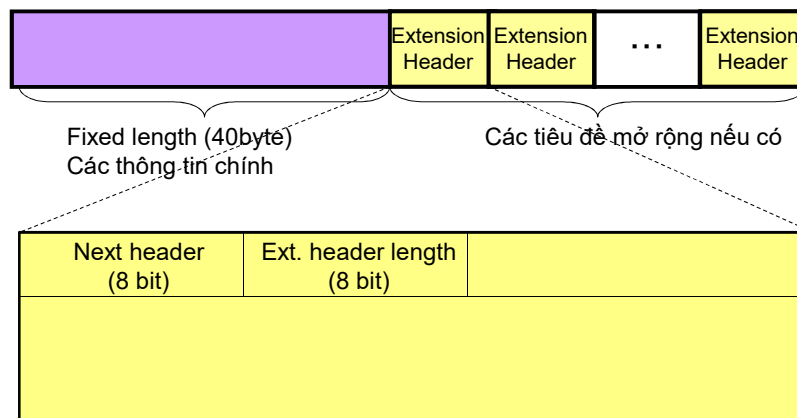
- Version: Phiên bản giao thức(=110)
- Traffic Class: số hiệu giao thức tầng trên
- Flow Label: Điều khiển QoS
- Payload Length: kích thước phần dữ liệu
- Next header: Tiêu đề tiếp theo mở rộng tiếp theo
- Hop limit: tương tự TTL

63

IPv6 header



- Trong IPv6, tiêu đề mở rộng đặt trong phần payload



64

3. Internet Control Message Protocol

Tổng quan
Khuôn dạng gói tin
Ping và Traceroute



65

3.1. Tổng quan về ICMP (1)



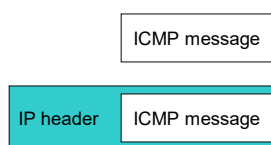
- IP là giao thức không tin cậy, không liên kết
 - Thiếu các cơ chế hỗ trợ và kiểm soát lỗi
- ICMP được sử dụng ở tầng mạng để trao đổi thông tin
 - Báo lỗi: báo gói tin không đến được một máy trạm, một mạng, một cổng, một giao thức.
 - Thông điệp phản hồi

66



Tổng quan về ICMP (2)

- Cũng là giao thức tầng mạng, song “phía trên” IP:
 - Thông điệp ICMP chứa trong các gói tin IP
- **ICMP message**: Type, Code, cùng với 8 bytes đầu tiên của gói tin IP bị lỗi

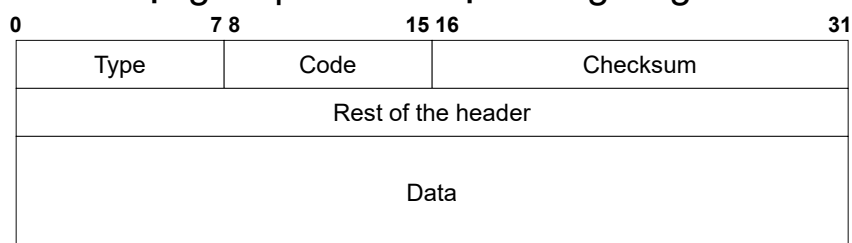


67



Khuôn dạng gói tin ICMP

- Type: dạng gói tin ICMP
- Code: Nguyên nhân gây lỗi
- Checksum
- Mỗi dạng có phần còn lại tương ứng



68

Một số dạng gói tin ICMP



ICMP Message Type	Error-reporting messages	3	Destination Unreachable
		4	Source quench
		5	Redirection
		11	Time exceeded
		12	Parameter problem
	Query messages	8 or 0	Echo reply or request
		13 or 14	Time stamp request or reply
		17 or 18	Address mask request or reply
		9 or 10	Router advertisement or solicitation

69

3.2. ICMP và các công cụ debug



- ICMP luôn hoạt động song song trong suốt với người sử dụng
- NSD có thể sử dụng ICMP thông qua các công cụ debug
 - ping
 - traceroute

70

Ping và ICMP



- ping
 - Sử dụng để kiểm tra kết nối
 - Gửi gói tin “ICMP echo request”
 - Bên nhận trả về “ICMP echo reply”
- Mỗi gói tin có một số hiệu gói tin
- Trường dữ liệu chứa thời gian gửi gói tin
 - Tính được thời gian đi và về - RTT (round-trip time)

71

Ping: Ví dụ



```
C:\Documents and Settings\admin>ping www.yahoo.co.uk

Pinging www.euro.yahoo-eu1.akadns.net [217.12.3.11] with 32 bytes of data:

Reply from 217.12.3.11: bytes=32 time=600ms TTL=237
Reply from 217.12.3.11: bytes=32 time=564ms TTL=237
Reply from 217.12.3.11: bytes=32 time=529ms TTL=237
Reply from 217.12.3.11: bytes=32 time=534ms TTL=237

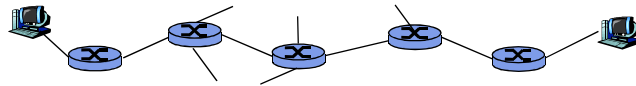
Ping statistics for 217.12.3.11:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
    Approximate round trip times in milli-seconds:
        Minimum = 529ms, Maximum = 600ms, Average = 556ms
```

72

Traceroute và ICMP



- Bên gửi truyền gói tin cho bên nhận
 - Lượt thứ nhất có TTL = 1
 - Lượt thứ 2 có TTL=2, ...
- Khi gói tin thứ n đến router thứ n:
 - Router hủy gói tin
 - gửi một gói tin ICMP (type 11, code 0)
 - có chứa tên và địa chỉ IP của router
- khi nhận được gói tin trả lời, bên gửi sẽ tính ra RTT



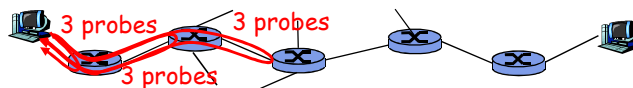
73

Traceroute và ICMP



Điều kiện kết thúc

- Gói tin đến được đích
- Đích trả về gói tin ICMP “time exceeded” (type 11)
- Khi nguồn nhận được gói tin ICMP này sẽ dừng lại
- Mỗi gói tin lặp lại 3 lần



74

Traceroute: Ví dụ



```
C:\Documents and Settings\admin>tracert www.jaist.ac.jp
```

Tracing route to www.jaist.ac.jp [150.65.5.208]
over a maximum of 30 hops:

1	1 ms	<1 ms	<1 ms	192.168.1.1
2	15 ms	14 ms	13 ms	210.245.0.42
3	13 ms	13 ms	13 ms	210.245.0.97
4	14 ms	13 ms	14 ms	210.245.1.1
5	207 ms	230 ms	94 ms	pos8-2.br01.hkg04.pccwbtn.net [63.218.115.45]
6	*	403 ms	393 ms	0.so-0-1-0.XT1.SCL2.ALTER.NET [152.63.57.50]
7	338 ms	393 ms	370 ms	0.so-7-0-0.XL1.SJC1.ALTER.NET [152.63.55.106]
8	402 ms	404 ms	329 ms	POS1-0.XR1.SJC1.ALTER.NET [152.63.55.113]
9	272 ms	288 ms	310 ms	193.ATM7-0.GW3.SJC1.ALTER.NET [152.63.49.29]
10	205 ms	206 ms	204 ms	wide-mae-gw.customer.alter.net [157.130.206.42]
11	427 ms	403 ms	370 ms	ve-13.foundry2.otemachi.wide.ad.jp [192.50.36.62]
12	395 ms	399 ms	417 ms	ve-4.foundry3.nezu.wide.ad.jp [203.178.138.244]
13	355 ms	356 ms	378 ms	ve-3705.cisco2.komatsu.wide.ad.jp [203.178.136.193]
14	388 ms	398 ms	414 ms	c76.jaist.ac.jp [203.178.138.174]
15	438 ms	377 ms	435 ms	www.jaist.ac.jp [150.65.5.208]

Tại sao trễ có đột biến?

Trace complete.

75

4. Định tuyến



76

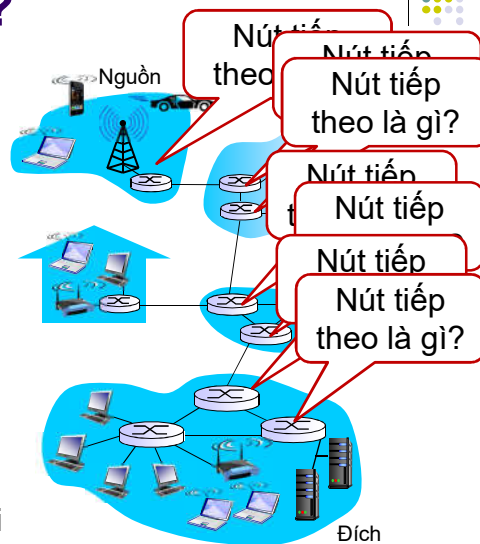
4.1. Khái niệm cơ bản về định tuyến



77

Định tuyến là gì?

- **Tìm đường đi** tốt nhất để **chuyển tiếp** dữ liệu tới nút đích
- Trong mạng IP chuyển mạch gói:
 - Các gói tin được chuyển tiếp độc lập
 - Định tuyến trên từng chặng: Mỗi nút không chỉ ra toàn bộ các chặng trên đường đi tới đích



78

Các thành phần của định tuyến



- Thông tin định tuyến: Thông số của đường truyền được sử dụng làm độ đo tính toán chi phí đường đi
- Bảng định tuyến: Lưu thông tin đường đi đã tìm được tới các mạng đích
- Giải thuật, giao thức định tuyến: cách thức tìm đường đi và trao đổi thông tin định tuyến giữa các nút mạng

79

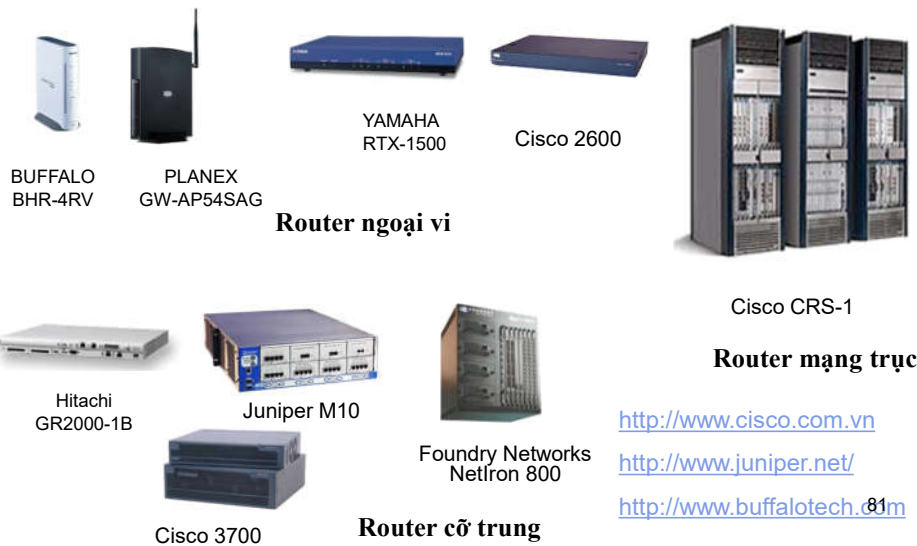
Bộ định tuyến (router)



- Thiết bị chuyển tiếp các gói tin giữa các mạng
 - Là một máy tính, với các phần cứng chuyên dụng
 - Kết nối nhiều mạng với nhau
 - Chuyển tiếp gói tin dựa trên bảng định tuyến
- Có nhiều cổng kết nối mạng
- Phù hợp với nhiều dạng lưu lượng và phạm vi của mạng

80

Một số ví dụ...



4.2. Bảng định tuyến

Destination	Outgoing Port	Next hop	Cost

- Destination : địa chỉ mạng đích
 - Định tuyến classless: Sử dụng địa chỉ không phân lớp
 - Định tuyến classful: Sử dụng địa chỉ phân lớp
- Outgoing Port : cổng ra cho gói tin để tới mạng đích
- Next hop : địa chỉ cổng nhận gói tin của nút kế tiếp
- Cost : chi phí gửi gói tin từ nút đang xét tới đích

Vấn đề cập nhật bảng định tuyến

- Sự thay đổi cấu trúc mạng: thêm mạng mới, một nút mạng bị mất kết nối
- Sự cần thiết phải cập nhật bảng định tuyến
 - Cho tất cả các nút mạng (về lý thuyết)
 - Thực tế, chỉ một số nút mạng phải cập nhật
- Làm thế nào để cập nhật ?
 - Định tuyến tĩnh: Các mục trong bảng định tuyến được sửa đổi thủ công bởi người quản trị
 - Định tuyến động: Tự động cập nhật bảng định tuyến các giao thức định tuyến

83

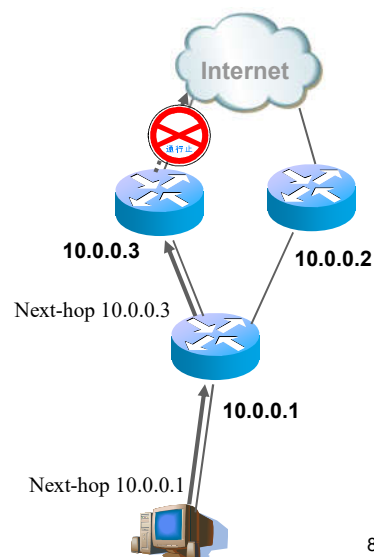
Định tuyến tĩnh

- Khi có sự cố:
 - Không thể nối vào Internet kể cả khi có tồn tại đường đi dự phòng
 - Người quản trị mạng cần thay đổi

Bảng định tuyến của 10.0.0.1 (1 phần)

Prefix	Next-hop
0.0.0.0/0	10.0.0.3

Kết nối bị lỗi



84

Định tuyến động

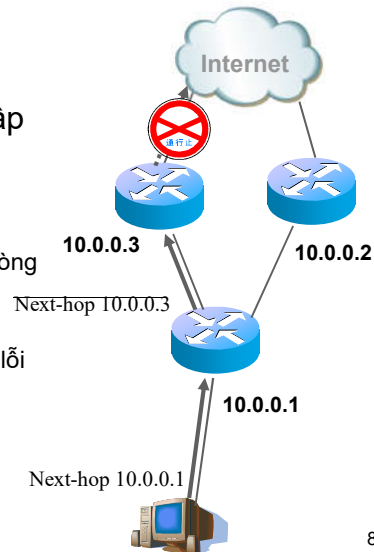
- Khi có sự cố:
 - Đường đi thay thế được cập nhật một cách tự động

Bảng định tuyến của 10.0.0.1 (1 phần)

Prefix	Next-hop
0.0.0.0/0	10.0.0.2
0.0.0.0/0	10.0.0.3

Kết nối dự phòng

Kết nối bị lỗi



85

Đặc điểm của định tuyến tĩnh

- Ưu
 - Ổn định
 - An toàn
 - Không bị ảnh hưởng bởi các yếu tố tác động
- Nhược
 - Cứng nhắc
 - Không thể sử dụng tự động kết nối dự phòng
 - Khó quản lý

86

Định tuyến động



- Ưu
 - Dễ quản lý
 - Tự động sử dụng kết nối dự phòng
- Nhược
 - Tính an toàn
 - Các giao thức chọn đường phức tạp và khó hiểu

87

4.3. Các giải thuật và giao thức định tuyến

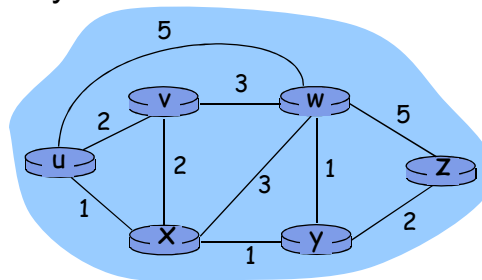
Giải thuật Dijkstra và Bellman-Ford
Giao thức dạng link-state và dạng
distance-vector



88

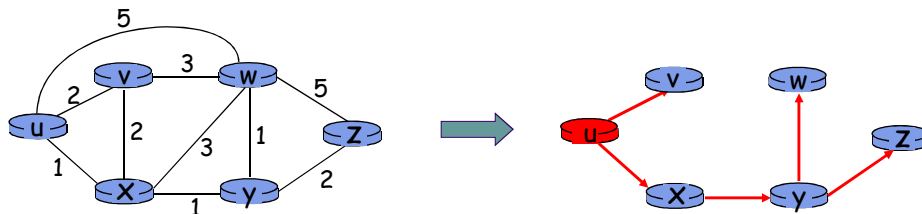
Biểu diễn mạng bởi đồ thị

- Đồ thị với các nút (bộ định tuyến) và các cạnh (liên kết)
- Chi phí cho việc sử dụng mỗi liên kết $c(x,y)$
 - Băng thông, độ trễ, chi phí, mức độ tắc nghẽn...
- Giải thuật định tuyến: Xác định đường đi ngắn nhất giữa hai nút bất kỳ



89

Cây đường đi ngắn nhất - SPT



- SPT – Shortest Path Tree
- Các cạnh xuất phát từ nút gốc và tới các lá
- Đường đi duy nhất từ nút gốc tới nút v, là đường đi ngắn nhất giữa nút gốc và nút v
- Mỗi nút sẽ có một SPT của riêng nút đó

90

Tập trung hay phân tán?



- Tập trung
 - Thu thập thông tin kết nối của toàn bộ mạng
 - Sử dụng các **giải thuật tìm đường đi trên đồ thị (thường sử dụng giải thuật dạng link-state)**
 - Phân bổ băng định tuyến từ nút trung tâm tới các nút
- Phân tán
 - Mỗi nút tự xây dựng bảng chọn đường riêng
 - **Giải thuật định tuyến:** Link-state hoặc distance-vector
 - Được sử dụng phổ biến trong thực tế

91

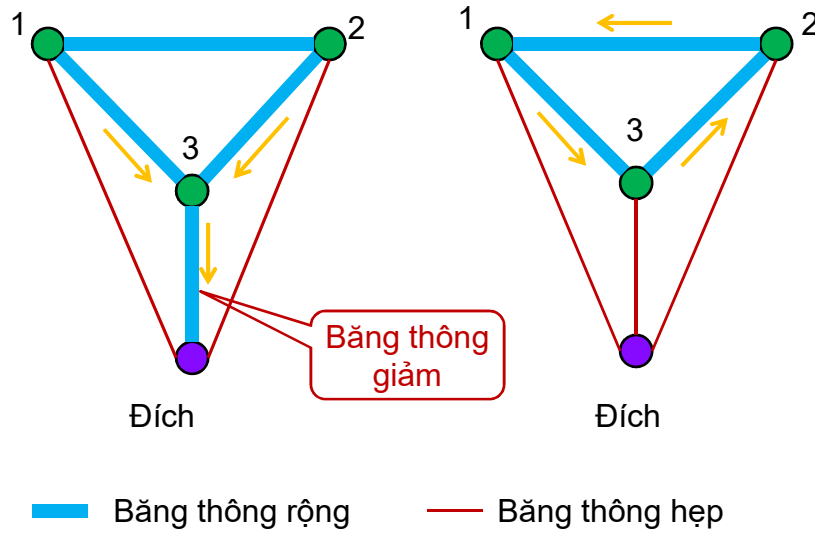
Chọn độ đo (metric) nào?



- Để lựa chọn đường đi tốt nhất cần đánh giá chi phí đường đi dựa trên các độ đo
- Các thông số đường truyền có thể sử dụng làm độ đo:
 - Độ trễ
 - Mức độ tắc nghẽn
 - Băng thông
 - Độ tin cậy
 - Tải
 - Độ mất gói tin...

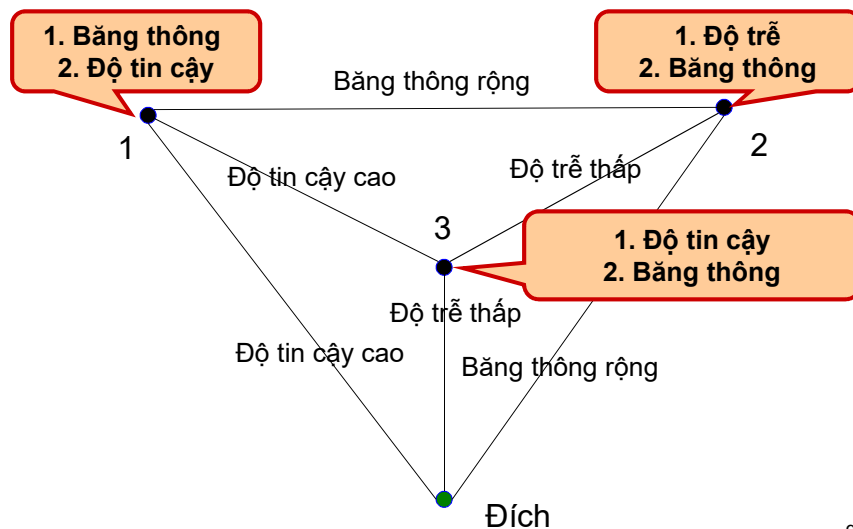
92

Thử chọn độ đo là băng thông



93

Các nút mạng không dùng chung độ đo



94

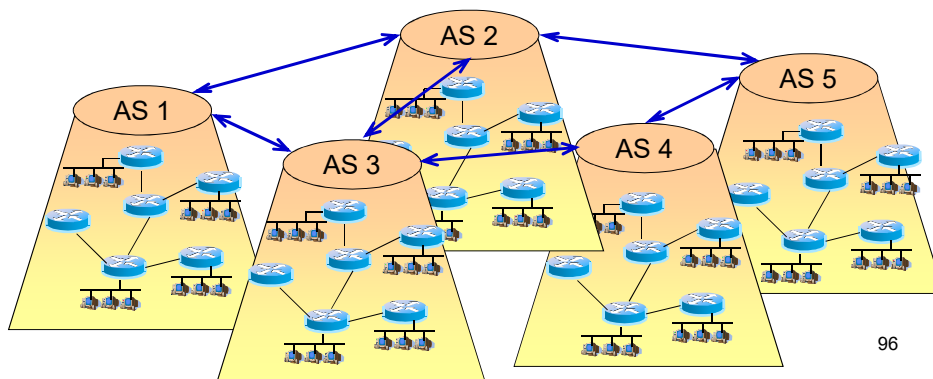
Vấn đề chọn độ đo

- Các nút mạng khi trao đổi thông điệp định tuyến
 - Cùng độ đo
 - Độ đo phải tăng nếu đi theo vòng lặp
 - Quá trình phải hội tụ
- cần giao thức định tuyến
 - Giải thuật nào?
 - Cách thức trao đổi thông tin định tuyến

95

Phân cấp trong định tuyến

- Internet = Mạng của các mạng
- Các mạng được quản lý bởi các tổ chức khác nhau, có chính sách rất khác nhau, trong đó có định tuyến
- Mỗi mạng như vậy có thể gọi là một hệ tự trị - Autonomous System (AS)



96

Khái niệm hệ tự trị - AS



- Góc nhìn định tuyến: Tập hợp các nút mạng có cùng chính sách chọn đường (Giao thức, quy ước chi phí...)
- Các ASes được nối kết thông qua các router hay gateway
- Mỗi hệ tự trị có một số hiệu riêng – AS number (ASN - 16 bits hay 32 bits).

[2914](#) NTT-COMMUNICATIONS-2914 - NTT America, Inc.
[3491](#) BTN-ASN - Beyond The Network America, Inc.
[4134](#) CHINANET-BACKBONE No.31,Jin-rong Street
[6453](#) GLOBEINTERNET Teleglobe America Inc.
[24087](#) VNGT-AS-AP Vietnam New Generation Telecom
[24066](#) VNNIC-AS-VN Vietnam Internet Network Information Center
[17981](#) CAMBOTECH-KH-AS ISP Cambodia
.....

97

Source: <http://www.cidr-report.org>

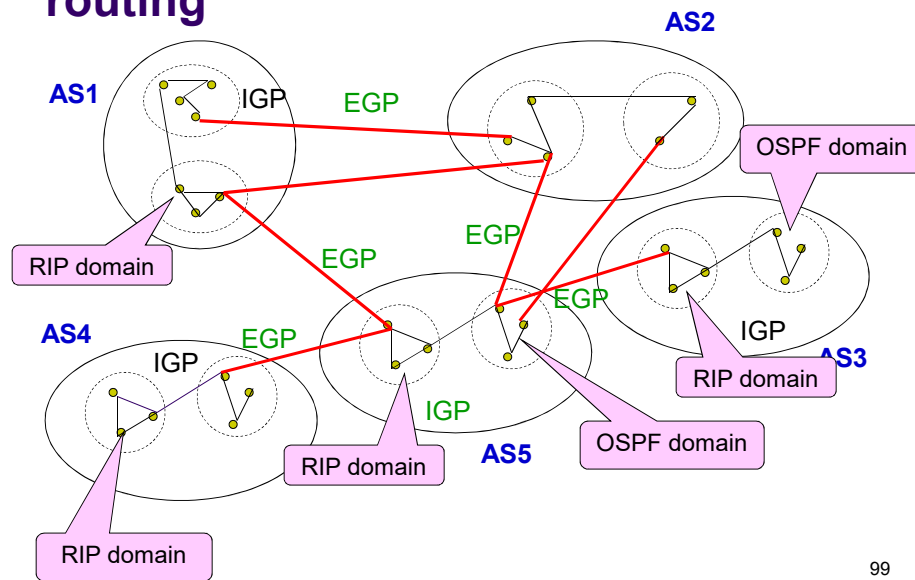
Phân cấp giao thức định tuyến



- **Trong một hệ tự trị:** Giao thức định tuyến nội vùng
 - **IGP:** *Interior Gateway Protocol*
 - RIP: Routing Information Protocol
 - OSPF: Open Shortest Path First
 - IS-IS, IGRP, EIGRP (Cisco)...
- **Giữa các hệ tự trị:** Giao thức định tuyến liên vùng
 - **EGP:** *Exterior Gateway Protocol*
 - BGP (v4): Border Gateway Protocol

98

Extra-domain và Inter-domain routing



99

Các yêu cầu với giao thức định tuyến

- Đường đi có chi phí tối thiểu theo độ đo đã lựa chọn
- Hội tụ
- Khả năng mở rộng
- Bảo mật
 - Điều gì xảy ra nếu có một nút “nói dối” mà không phát hiện được?
- Hỗ trợ chính sách

100

4.3.1. Giải thuật dạng distance-vector



Phương trình Bellman-Ford (quy hoạch động)

Định nghĩa

$d_x(y)$: chi phí của đường đi ngắn nhất từ x tới y

$c(x,v)$: chi phí từ x tới hàng xóm v

Giải phương trình:

$$d_x(y) = \min_v \{c(x,v) + d_v(y)\}$$

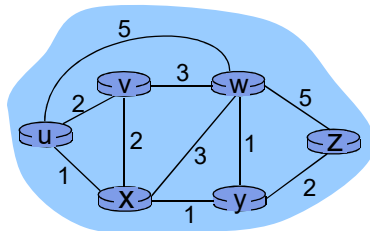
cho tất cả các v là hàng xóm của x

101

Minh họa Bellman-Ford Eq.



Dễ thấy, $d_v(z) = 5$, $d_x(z) = 3$, $d_w(z) = 3$



B-F eq. cho ta biết:

$$\begin{aligned} d_u(z) &= \min \{ c(u,v) + d_v(z), \\ &\quad c(u,x) + d_x(z), \\ &\quad c(u,w) + d_w(z) \} \\ &= \min \{ 2 + 5, \\ &\quad 1 + 3, \\ &\quad 5 + 3 \} = 4 \end{aligned}$$

Nút nào làm giá trị trên nhỏ nhất → Lựa chọn là nút kế tiếp trong bảng định tuyến

102

Giải thuật dạng distance-vector (2)

ý tưởng cơ bản:

- DV: Vector khoảng cách, tạm coi là đường đi ngắn nhất của từ một nút tới nút khác
- Mỗi nút tính toán lại DV nếu:
 - Liên kết thay đổi
 - Nhận được DV từ hàng xóm
- Thông báo cho hàng xóm nếu DV thay đổi
 - Hàng xóm sẽ thông báo cho hàng xóm của chúng

Mỗi nút:

Chờ (Thay đổi trong DV của nút bên cạnh)

Tính lại ước lượng DV

Nếu DV thay đổi, *Báo* cho nút bên cạnh

103

$$D_x(y) = \min\{c(x,y) + D_y(y), c(x,z) + D_z(y)\}$$

$$= \min\{2+0, 7+1\} = 2$$

nút x

	x	y	z
x	0	2	7
y	∞	∞	∞
z	∞	∞	∞

nút y

	x	y	z
x	∞	∞	∞
y	2	0	1
z	∞	∞	∞

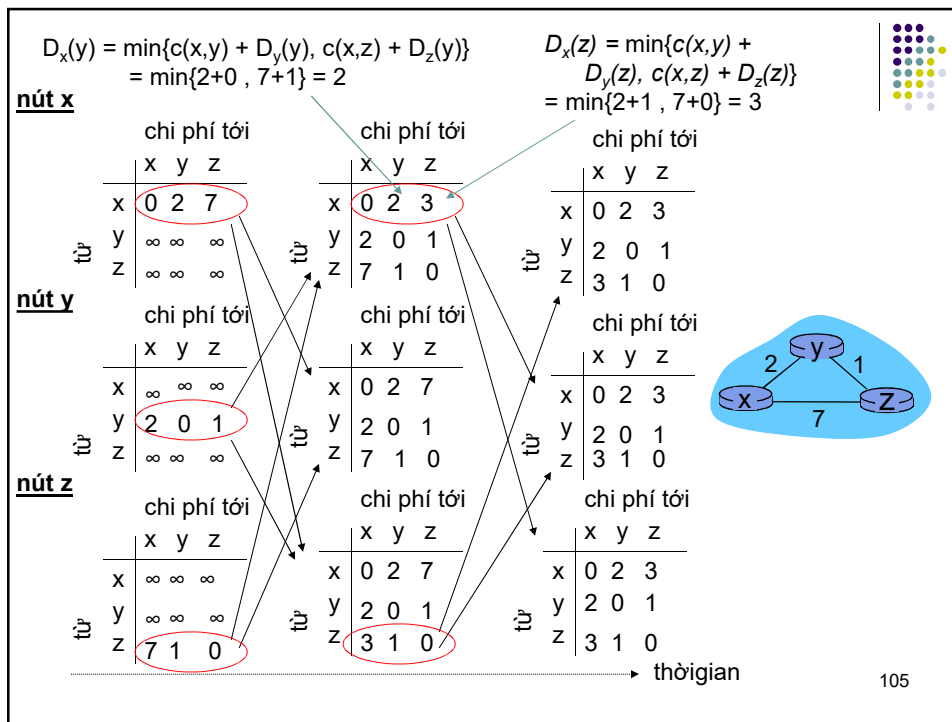
nút z

	x	y	z
x	∞	∞	∞
y	∞	∞	∞
z	7	1	0

nút x $D_x(z) = \min\{c(x,y) + D_y(z), c(x,z) + D_z(z)\}$
 $= \min\{2+1, 7+0\} = 3$

thời gian

104



105

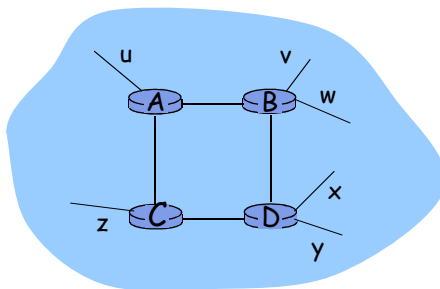
Vấn đề đếm tới vô cùng

- Count-to-infinity
- Xảy ra khi chi phí một liên kết tăng lên so với ban đầu
- Mỗi router không biết đường đi đến router khác lại đi qua chính nó
- Sau mỗi lần trao đổi và cập nhật bảng định tuyến, khoảng cách ước lượng sẽ tăng dần đến vô cùng

106

RIP (Routing Information Protocol)

- IGP
- RIP v.1, phiên bản mới RIP v.2
- Giao thức dạng vector khoảng cách
- Chọn đường đi theo số nút mạng đi qua



Từ nút A:

<u>Đích đến</u>	<u>số nút</u>
u	1
v	2
w	2
x	3
y	3
z	2

107

RIP: Trao đổi thông tin

- Trao đổi bảng chọn đường
- Định kỳ
 - Các vector khoảng cách được trao đổi định kỳ - 30s
 - Mỗi thông điệp chứa tối đa 25 mục
 - Trong thực tế, nhiều thông điệp được sử dụng
- Sự kiện
 - Gửi thông điệp cho nút hàng xóm mỗi khi có thay đổi
 - Nút hàng xóm sẽ cập nhật bảng chọn đường của nó

108

RIP: Đề tranh lỗi lặp vô hạn



- Giới hạn số hop tối đa
 - v1 : 15
 - v2 : 30
- “Split horizon”
 - Thông tin chọn đường không được quay về nút nguồn
- “Poison reverse”
 - Khi liên kết bị lỗi, gửi giá trị của chi phí là 16 (hoặc 31)
 - Liên kết chuyển sang trạng thái hold-down

109

4.3.2. Giải thuật dạng link-state



- Mỗi nút thu thập thông tin từ các nút khác để xây dựng topo của mạng
- Áp dụng giải thuật tìm đường đi ngắn nhất tới mọi nút trong mạng
- Khi có sự thay đổi bảng định tuyến, mỗi nút gửi thông tin thay đổi cho tất cả các nút khác trong mạng.

110

Ký hiệu



- $G = (V, E)$: Đồ thị với tập đỉnh V và tập cạnh E
- $c(x, y)$: chi phí của liên kết x tới y ; $= \infty$ nếu không phải 2 nút kề nhau
- $d(v)$: chi phí hiện thời của đường đi từ nút nguồn tới nút đích. v
- $p(v)$: nút ngay trước nút v trên đường đi từ nguồn tới đích
- T : Tập các nút mà đường đi ngắn nhất đã được xác định

111

Các thủ tục



- **Init():**
Với mỗi nút v , $d[v] = \infty$, $p[v] = \text{NIL}$
 $d[s] = 0$
- **update(u, v)**, trong đó (u, v) là một cạnh nào đó của G
if $d[v] > d[u] + c(u, v)$ then
 $d[v] = d[u] + c(u, v)$
 $p[v] = u$

112

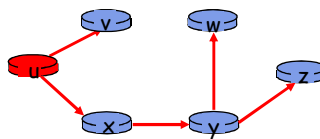
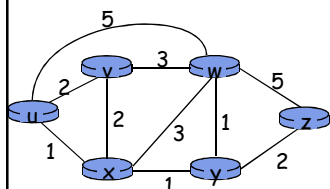
Dijkstra's Algorithm

1. **Init()** ;
2. $T = \Phi$;
3. **Repeat**
4. $u: u \notin T \mid d(u)$ là bé nhất ;
5. $T = T \cup \{u\}$;
6. for all $v \in \text{neighbor}(u)$ và $v \notin T$
7. $\text{update}(u,v)$;
8. **Until** $T = V$

113

Dijkstra's algorithm: Ví dụ

Step	T	$d(v), p(v)$	$d(w), p(w)$	$d(x), p(x)$	$d(y), p(y)$	$d(z), p(z)$
0	u	2, u	5, u	1, u	∞	∞
1	ux	2, u	4, x		2, x	∞
2	uxy	2, u	3, y			4, y
3	uxyv		3, y			4, y
4	uxyvw					4, y
5	uxyvwz					



SPT của u:

Bảng định tuyến của u:

destination	link
v	(u,v)
x	(u,x)
y	(u,x)
w	(u,x)
z	(u,x)

OSPF: Open Shortest Path First



- IGP
- **Open**: Chuẩn mở của IETF (phiên bản 3, định nghĩa trong [RFC 2740](#))
- **Shortest Path First**: Cài đặt giải thuật Dijkstra.
- Thông tin về trạng thái liên kết - LSA (link state advertisement) được quảng bá “tràn ngập” trên toàn AS

115

Một số đặc điểm của OSPF



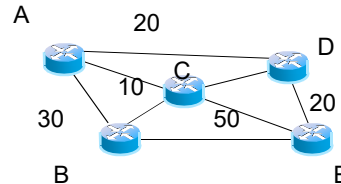
- **An toàn**: thông điệp OSPF được bảo vệ
- Với các AS lớn: OSPF **phân cấp**
- **Địa chỉ không phân lớp** (Variable-Length Subnet Masking -VLSM)
- Mỗi link sẽ có nhiều giá trị về **chi phí** khác nhau dựa trên TOS (tuy nhiên hơi phức tạp và chưa được sử dụng)

116

Thông tin chọn đường?



- Link-State Advertisement (LSA): Chỉ ra một nút được nối tới nút nào (link) và chi phí (cost) tương ứng
- Ví dụ: nút A
 - link to B, cost 30
 - link to D, cost 20
 - link to C, cost 10
- Ví dụ: nút D
 - link to A, cost 20
 - link to E, cost 20
 - link to C, cost 50



117

Chi phí trong giao thức OSPF - metric



- Giá trị mặc định
100Mbps / bandwidth of interface
 - Hiện nay người quản trị có thể gán giá trị này
- Khi tính toán bảng chọn đường
 - Chọn đường đi chi phí nhỏ nhất
- Chi phí bằng nhau
 - Có thể thực hiện cân bằng tải

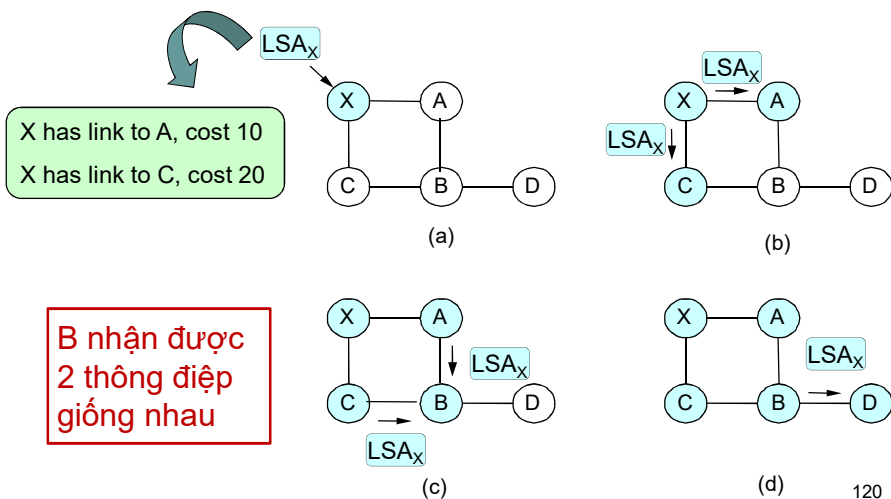
118

Chi phí mặc định của OSPF

Link Bandwidth	Default OSPF cost
56Kbps serial link	1785
64Kbps serial link	1562
T1 (1.544Mbps) serial link	65
E1 (2.048Mbps) serial link	48
4Mbps Token Ring	25
Ethernet	10
16Mbps Token Ring	6
FDDI or Fast Ethernet	1
Gigabit Ethernet / 10G network	1

119

Quảng bá thông tin LSA-Flooding



Phân cấp OSPF

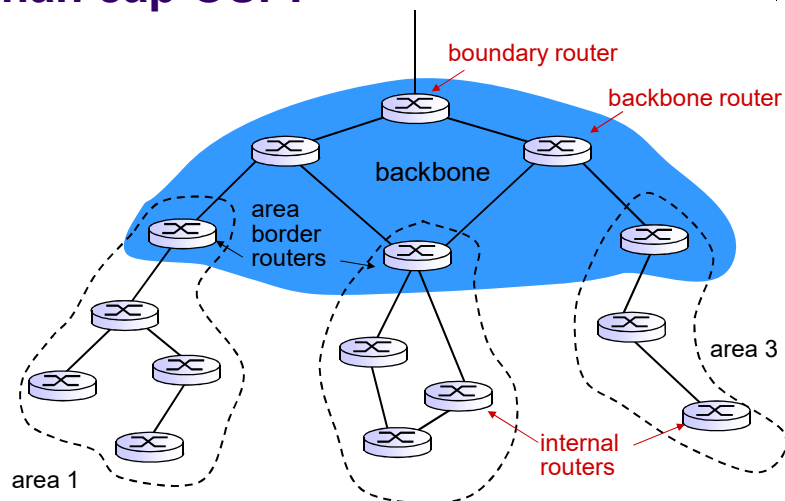
- Trong việc chọn đường, tại sao phải chia mạng thành các vùng nhỏ hơn?
- Nếu có quá nhiều router
 - Thông tin trạng thái liên kết được truyền nhiều lần hơn
 - Phải liên tục tính toán lại
 - Cần nhiều bộ nhớ hơn, nhiều tài nguyên CPU hơn
 - Lượng thông tin phải trao đổi tăng lên
 - Bảng chọn đường lớn hơn



- Vùng
 - Nhóm các router có cùng thông tin LSA

121

Phân cấp OSPF



122

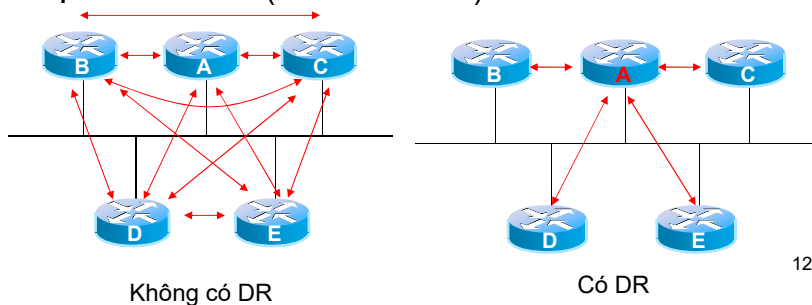
Các dạng router

- **ABR - Area border routers:** Quản lý 1 vùng và kết nối đến các vùng khác
- **ASBR - Autonomous system boundary router:** Nối đến các AS khác
- **BR - backbone routers:** thực hiện OSPF routing trong vùng backbone
- **Internal Router** – Thực hiện OSPF bên trong một vùng

123

Router đại diện - DR

- Để tăng hiệu quả của việc quảng bá LSA trong mạng quảng bá đa truy nhập
- Mỗi router phải lập quan hệ với router đại diện - designated router (DR)
 - Trao đổi thông tin thông qua DR
 - DR dự phòng
- Chọn DR và BDR(Backbone DR)?



124

So sánh các giải thuật LS và DV



Thông điệp trao đổi

- LS: n nút, E cạnh, $O(nE)$ thông điệp
- DV: Chỉ trao đổi giữa các hàng xóm
 - Thời gian hội tụ thay đổi

Tốc độ hội tụ

- LS: Thuật toán: $O(n^2)$ cần $O(nE)$ thông điệp
- DV: Thay đổi

Sự chắc chắn: Giải sử một router hoạt động sai

LS:

- nút gửi các chi phí sai
- Mỗi nút tính riêng bằng chọn đường -> có vẻ chắc chắn hơn

DV:

- DV có thể bị gửi sai
- Mỗi nút tính toán dựa trên các nút khác
 - Lỗi bị lan truyền trong mạng

125

RIP vs. OSPF



	RIP	OSPF
Đặc điểm	<ul style="list-style-type: none"> • Router bình đẳng • Cấu hình dễ dàng • Mạng cỡ nhỏ 	<ul style="list-style-type: none"> • Phân cấp • Cấu hình phức tạp • Mạng cỡ vừa và lớn
Khả năng mở rộng	X	O
Độ phức tạp tính toán	Nhỏ	Lớn
Hội tụ	Chậm	Nhanh
Trao đổi thông tin	Bảng chọn đường	Trạng thái liên kết
Giải thuật	Distant vector	Link-state
Cập nhật hàng xóm	30s	10s (Hello packet)
Đơn vị chi phí	Số nút mạng	Bảng thông

126

4.4. Định tuyến liên vùng



127

4.4.1. Giới thiệu chung về BGP



128

Yêu cầu mới với các giao thức định tuyến liên vùng



- Mỗi AS có chính sách riêng về định tuyến
 - Nhà cung cấp X không muốn truyền dữ liệu của họ qua mạng nhà cung cấp Y
 - Nhà cung cấp X không muốn dữ liệu của nhà cung cấp Z truyền qua hệ thống mạng của họ
 - Không thể thể hiện những chính sách này trong khái niệm “đường đi ngắn nhất”
- Mỗi AS muốn “tự trị”: tự chọn các giao thức định tuyến nội vùng và các chính sách định tuyến đi kèm
- Mỗi AS muốn tính “riêng tư”: tự chọn topology, các chính sách điều phối hoạt động

129

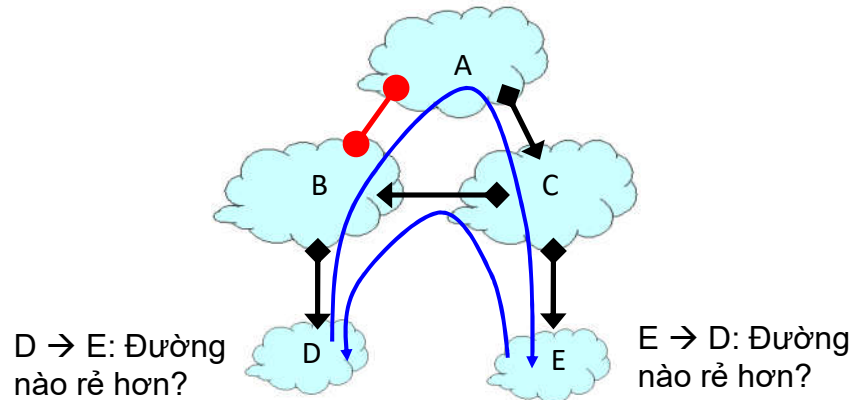
Yêu cầu mới với các giao thức định tuyến liên vùng



- Xuất hiện các mối quan hệ “kinh doanh” trên topology và chính sách của mạng:
 - AS X có thể là nhà cung cấp (NCC) của AS Y
 - AS X có thể là khách hàng (KH) của AS Y
 - AS X và AS Y có thể là đối tác ngang hàng (Peer)
- Quan hệ kinh doanh
 - Khách hàng trả tiền cho nhà cung cấp
 - Đối tác ngang hàng không cần trả tiền

130

Định tuyến theo chi phí (Routing follows the Money)



Quan hệ giữa các AS

NCC ↔ KH
peer ●—● peer

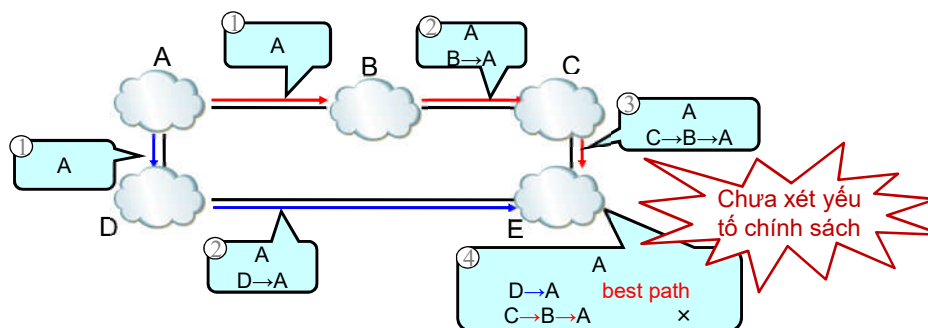
Quan hệ kinh doanh

- KH trả tiền cho NCC
- Đối tác ngang hàng: miễn phí

Chọn giải thuật định tuyến nào?



- Giữa các AS nên dùng giao thức nào?
 - Khó có một chính sách và đơn vị chi phí chung
 - LS: Chi phí không đồng nhất, CSDL quá lớn
 - DV: Mạng quá rộng, khó hội tụ
- Giải pháp: BGP chọn đường theo path-vector



132

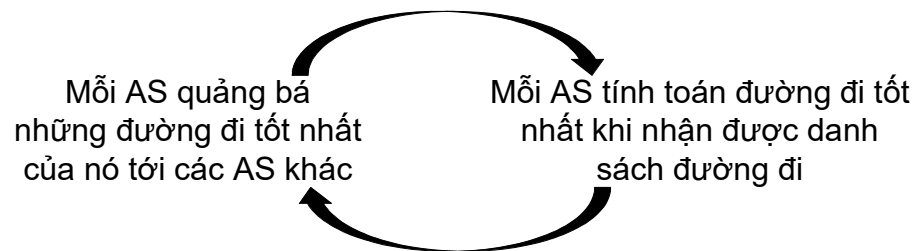
BGP – Border Gateway Protocol



- Yếu tố gắn kết của Internet, kết nối các hệ tự trị
- Trao đổi thông tin đường đi NLRI (Network Layer Reachability Information)
 - Cho phép một AS biết được thông tin đi đến AS khác
 - Gửi thông tin này vào bên trong AS đó
 - Xác định đường đi tốt nhất dựa trên thông tin đó và các chính sách chọn đường
- Cho phép thiết lập các chính sách
 - Chọn đường ra
 - Quảng bá các đường vào

133

BGP: Ý tưởng tìm đường



Có vẻ như BGP sử dụng lại ý tưởng nào đó!

134

BGP sử dụng ý tưởng của DV



- Quảng bá thông tin về đường đi tới đích
- Không chia sẻ thông tin về topology của mạng
- Lặp liên tục cho đến khi hội tụ (tìm thấy đường đi tốt nhất)
- Với 4 điểm khác biệt

135

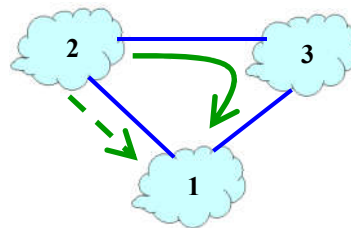
BGP khác DV (1)

Không chọn đường đi ngắn nhất



- Đường đi ngắn nhất (qua ít AS nhất) là mục tiêu rất quan trọng trong định tuyến liên vùng (vì sao?) nhưng không phải là ưu tiên hàng đầu.
- BGP ưu tiên chọn đường theo chính sách trước

Nút 2 có thể chọn
“2, 3, 1” thay vì “2, 1”

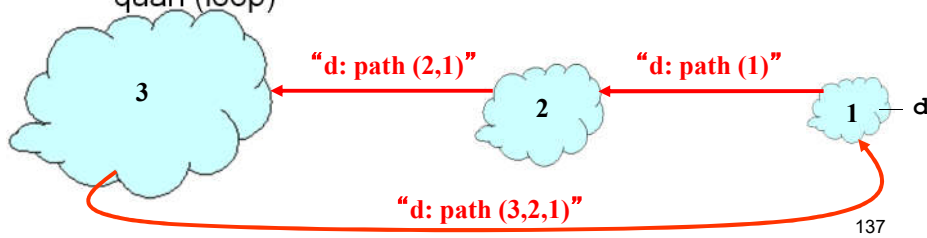


136

BGP khác DV (2)

Định tuyến theo vector đường đi

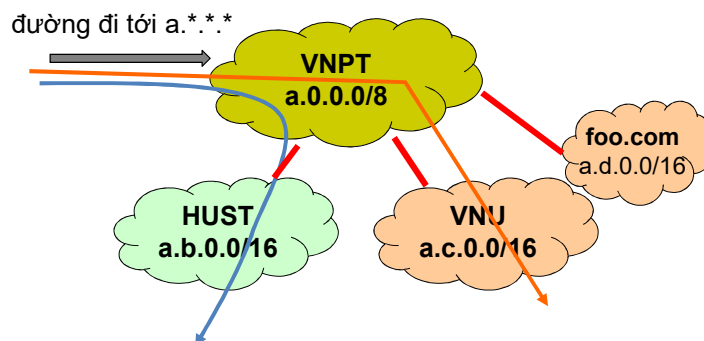
- Ý tưởng chính: quảng bá toàn bộ các chặng trên đường đi
 - Distance vector: chọn đường dựa trên chi phí của đường đi tới các đích → có thể xuất hiện đường đi quẩn do vấn đề đếm tới vô cùng (count-to-infinity)
 - Path vector: chọn đường dựa trên các chặng của đường đi tới đích → dễ dàng phát hiện các đường đi quẩn (loop)



BGP khác DV (3)

Kết hợp đường đi

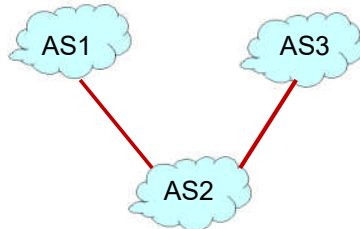
- BGP có khả năng kết hợp các đường đi tới các mạng con không đầy đủ



BGP khác DV (4)

Quảng bá có chọn lựa

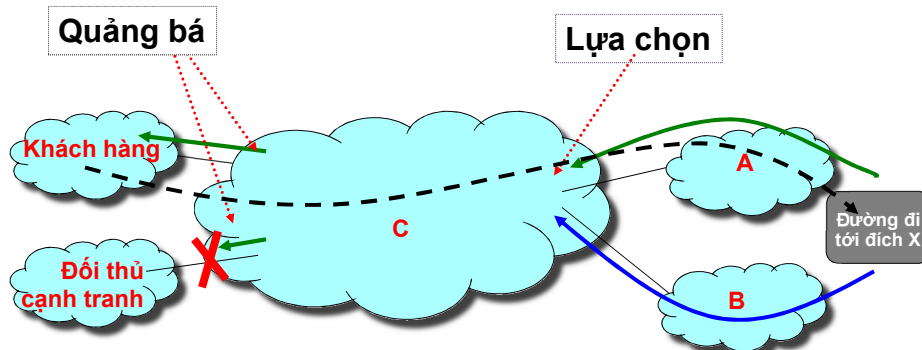
- Một AS có thể chọn để không quảng bá đường đi tới một đích nào đó
- Nói một cách khác, một AS có đường đi tới AS đích nhưng không đảm bảo sẽ vận chuyển mọi lưu lượng mạng tới đó



Ví dụ: AS2 không muốn vận chuyển thông tin từ AS1 tới AS3

139

Quảng bá có chọn lựa



- **Lựa chọn:** Đường đi nào được dùng để chuyển dữ liệu tới AS đích?
 - Kiểm soát thông tin ra khỏi AS
- **Quảng bá:** Quảng bá cho AS khác đường đi nào?
 - Kiểm soát thông tin vào AS

140

Quảng bá có chọn lựa: Một số chính sách điển hình



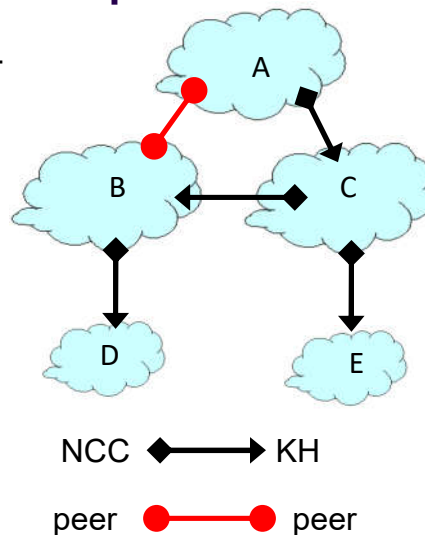
- Thay đổi mục tiêu khi định tuyến:
 - Tối thiểu hóa chi phí, tối đa hóa lợi nhuận
 - Tối đa hóa hiệu năng (đường đi qua ít AS nhất)
 - Tối thiểu hóa lưu lượng qua AS (định tuyến kiểu “hot potato”)
 - ...
- BGP có cơ chế gán thuộc tính cho các tuyến đường để thực hiện các mục tiêu trên

141

Quảng bá có chọn lựa Ví dụ: Định tuyến theo chi phí



- Quan hệ ngang hàng (peer-peer)
 - B quảng bá đường đi của AS nào cho A?
 - B quảng bá đường đi của A cho những AS nào?
- Quan hệ NCC-KH (provider-customer)
 - B quảng bá đường đi của D cho những AS nào?
 - B quảng bá đường đi của C cho những AS nào?



142

4.4.2. Hoạt động của BGP



143

eBGP, iBGP và IGP

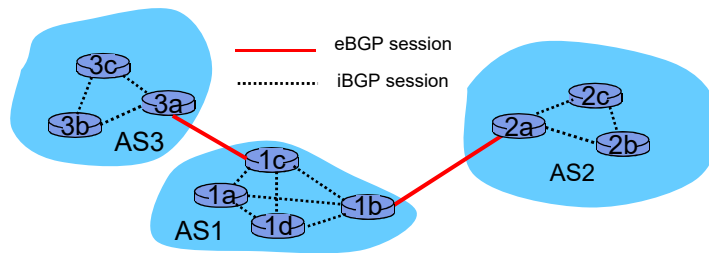


- BGP cài đặt trên các router biên của AS(kết nối tới các AS khác): 2 phiên hoạt động
 - External BGP (eBGP): thực hiện trao đổi thông điệp với các router biên trên AS khác để tìm đường đi tới đích nằm ngoài AS của nó
 - Internal BGP (iBGP): trao đổi thông điệp với các router biên và router nội vùng cùng AS để quảng bá đường đi tới đích nằm ngoài AS của nó
- IGP : Interior Gateway Protocol = Intra-domain Routing Protocol
 - Cài đặt trên router nội vùng
 - Tìm đường đi tới đích nằm trong vùng AS
 - Dữ liệu tới đích ngoài AS sẽ được chuyển tới router biên

144

eBGP và iBGP – Ví dụ

- Quảng bá thông tin đường đi
 1. 3a gửi tới 1c bằng eBGP
 2. 1c gửi thông tin nội bộ tới (1b, 1d, ...) trong AS1 bằng iBGP
 - 1b: Router biên cài BGP
 - 1a, 1d: Router nội vùng cài IGP
 3. 2a nhận thông tin từ 1b bằng eBGP



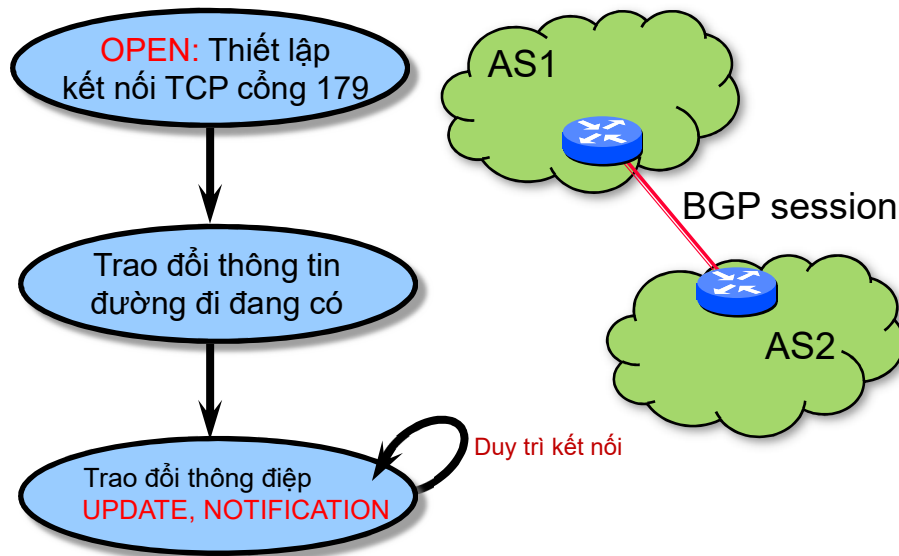
145

Các thông điệp BGP cơ bản

- OPEN:
 - Thiết lập phiên trao đổi thông tin đường đi
 - Sử dụng TCP, cổng 179
- NOTIFICATION: thông báo các sự kiện bất thường
- UPDATE:
 - Thông báo về đường đi mới
 - Thông báo về đường đi không còn khả dụng
- KEEPALIVE: thông báo duy trì kết nối TCP
 - Kết nối của TCP cung cấp cho BGP là nửa duy trì (semi-persistent)

146

Phiên trao đổi thông tin của BGP

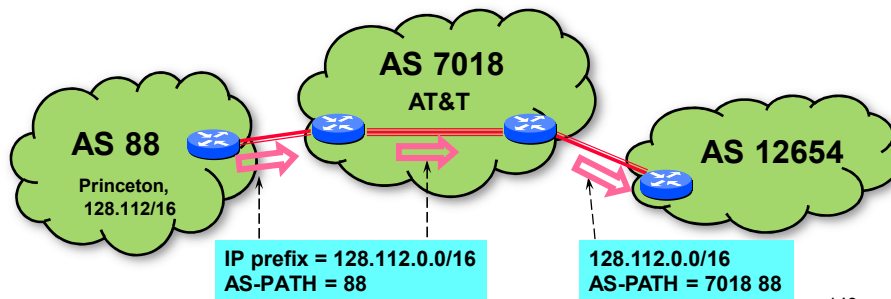


UPDATE = <IP prefix: Thuộc tính>

- IP prefix: địa chỉ của đích
- Thuộc tính gán cho đường đi: sử dụng cho mục đích lựa chọn/quảng bá đường đi nào:
 - Các thuộc tính nội bộ: chỉ dùng cho các thông điệp trao đổi trong AS. Ví dụ: LOCAL-PREF
 - Các thuộc tính sử dụng cho eBGP: ORIGIN, AS-PATH, NEXT-HOP, MED
 - Các thuộc tính khác: ATOMIC_AGGREGATE, AGGREGATOR, COMMUNITY...

Thuộc tính(1): AS-PATH

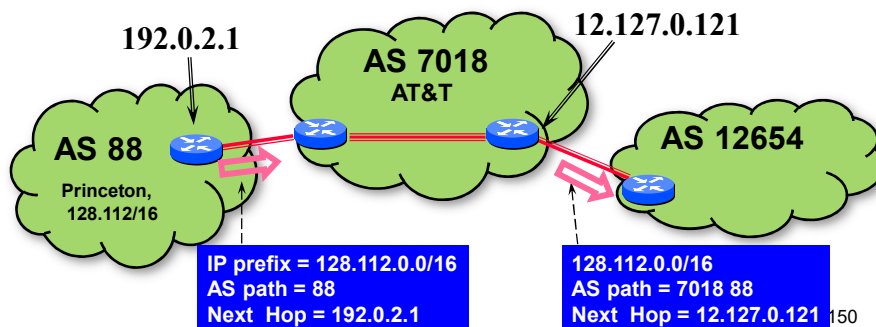
- Thông tin về đường đi tới một đích (IP prefix)
- Liệt kê số hiệu các AS trên đường đi tới đích (theo thứ tự gần đích tới xa đích)



149

Thuộc tính(2): NEXT-HOP

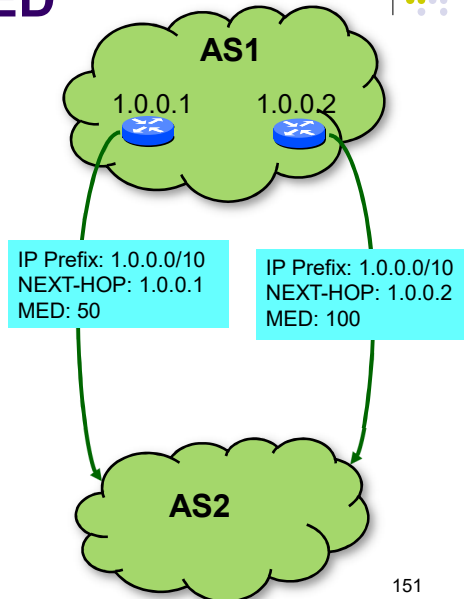
- Địa chỉ IP của router tiếp theo trên đường đi tới đích
- Cập nhật trên thông điệp UPDATE ra khỏi AS



150

Thuộc tính(3): MED

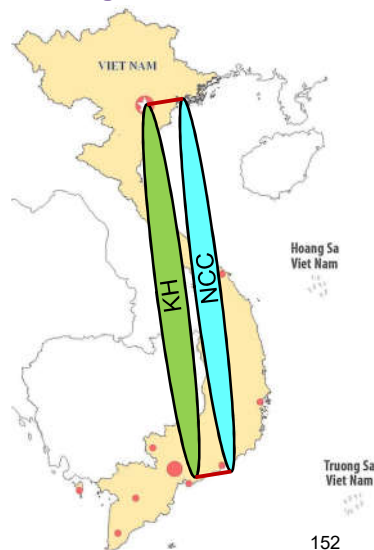
- Multi-Exit Discriminator
 - Dùng cho eBGP
 - Sử dụng trong trường hợp một AS có nhiều liên kết tới một AS khác:
 - AS quảng bá đường đi với giá trị MED khác nhau qua các liên kết khác nhau
 - AS nhận thông tin sẽ chọn đường đi có MED nhỏ hơn
- điều khiển lưu lượng vào



151

Thuộc tính(3): MED Routing follow the Money

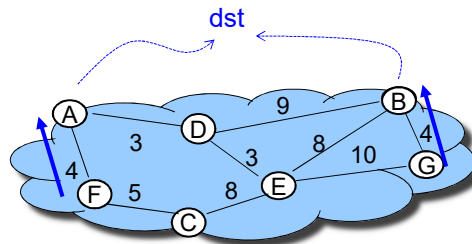
- Nhà cung cấp (NCC) có 2 điểm cấp quang quốc tế cập bờ ở Hải Phòng và Vũng Tàu
- Khách hàng(KH) đầu nối với NCC tại 2 điểm trên
- Giả sử có lưu lượng Internet quốc tế tới AS của NCC tại Vũng Tàu và đích là nút mạng đặt tại Hà Nội thuộc AS của KH
- Đặt MED như thế nào?
- Nếu quan hệ là ngang hàng?



152

Định tuyến kiểu “hot potato”

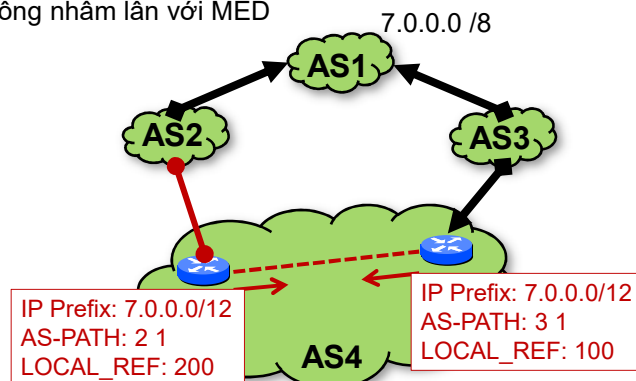
- AS chuyển dữ liệu ra khỏi nó nhanh nhất có thể
- Là chiến lược mặc định của các giao thức định tuyến nội vùng (IGP)
 - Router nội vùng tìm router biên gần nó nhất để chuyển dữ liệu ra ngoài AS
 - Có thể mâu thuẫn với MED → khi cần thì chuyển qua chiến lược “cold potato”



153

Thuộc tính(4): LOCAL_REF

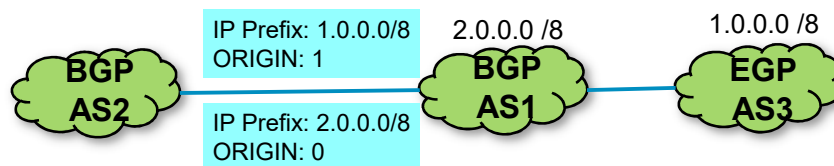
- Local reference
- Trao đổi trên các thông điệp iBGP
- Gán cho các đường đi tới cùng đích
 - Chọn đường đi có LOCAL_REF lớn hơn → điều khiển lưu lượng ra
 - Không nhầm lẫn với MED



154

Thuộc tính (5): ORIGIN

- Chỉ ra nguồn gốc của thông tin về đường đi
- Sử dụng 1 trong 3 giá trị:
 - 0-IGP: thông tin đường đi học được từ trong AS qua giao thức IGP
 - 1-EGP: thông tin đường đi học được từ ngoài AS qua giao thức EGP (Exterior Gateway Protocol) [RFC904]
 - ?-INCOMPLETE: đường đi học được từ nguồn không xác định (thường do định tuyến tĩnh)



155

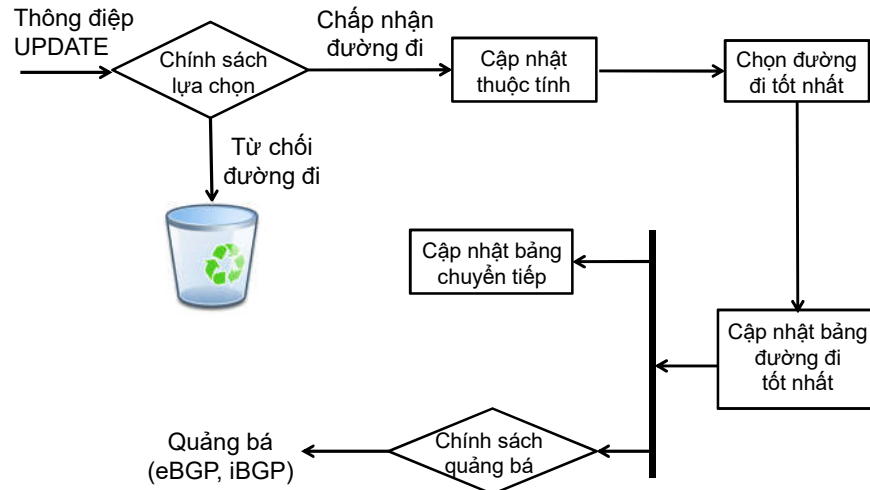
Sử dụng các thuộc tính

- Thứ tự ưu tiên khi chọn đường

Độ ưu tiên	Tiêu chí	Mục tiêu
1	LOCAL PREF	Cao nhất
2	ASPATH	Qua ít AS nhất
3	MED	Thấp nhất
4	eBGP > iBGP	Chọn đường đi học từ AS khác
5	iBGP path	Đường đi tới router biên gần nhất
6	Router ID	Địa chỉ IP nhỏ nhất

156

Quá trình xử lý thông điệp UPDATE



157

Một số vấn đề tồn tại của BGP

- An toàn bảo mật:
 - Tấn công vào BGP có thể gây thiệt hại lớn
- Không đảm bảo hiệu năng
 - Vì ưu tiên tìm đường theo chính sách trước tìm đường ngắn nhất
- Hội tụ chậm
 - Dưới 35% router có thời gian hoạt động 99.99%
 - Khoảng 10% có thời gian hoạt động dưới 95%
 - 40% số đường bị lỗi cần 30 phút để cập nhật xong
 - May mắn là hầu hết đường đi đều ổn định.
- Phức tạp khi cần triển khai các chính sách

158

Còn rất nhiều vấn đề về tầng mạng



- Giao thức IPv6
- Mobile IP
- Định tuyến trên mạng cáp quang
- Định tuyến trên mạng không dây
- Broadcast và Multicast
- An toàn bảo mật thông tin tầng mạng
- An toàn bảo mật cho các giao thức định tuyến...

159

Tài liệu tham khảo



- Keio University
- “Computer Networking: A Top Down Approach”, J.Kurose
- “Computer Network”, Berkeley University

160