Introduction to Routing

Srinath Beldona

Senior Technical Specialist

srinath@apnic.net

Agenda

- Pre-requisites
- OSI & TCP/IP Layering architecture
- IP Addressing
- The need to route
- Static Routing
- Dynamic Routing
- Distance Vector Routing Protocols and Limitations
- Link State Routing Protocols (OSPF)
- BGP as an Inter AS Routing Protocol

Pre-requisites

APNIC

Pre-requisites

- Good knowledge of computers
- Binary System of numbers
- Hexadecimal representation of numbers
- Thorough understanding of OSI Layers
- Good knowledge Data Link layer concepts and protocols
- Good knowledge of TCP/IP

OSI Layers

APNIC

Communication Architecture

- Strategy for connecting host computers and other communicating equipment.
- Defines necessary elements for data communication between devices.
- A communication architecture, therefore, defines a standard for the communicating hosts.
- A programmer formats data in a manner defined by the communication architecture and passes it on to the communication software.
- Separating communication functions adds flexibility, for example, we do not need to modify the entire host software to include more communication devices.

Layered Architecture

- Layered architecture simplifies the network design.
- It is easy to debug network applications in a layered architecture network.
- The network management is easier due to the layered architecture.
- Network layers follow a set of rules, called protocol.
- The protocol defines the format of the data being exchanged, and the control and timing for the handshake between layers.

Who created OSI reference model ?

- International standard organization (ISO) established a committee in 1977 to develop an architecture for computer communication.
- Open Systems Interconnection (OSI) reference model is the result of this effort.
- In 1984, the Open Systems Interconnection (OSI) reference model was approved as an international standard for communications architecture.
- Term “open” denotes the ability to connect any two systems which conform to the reference model and associated standards.

International Standards Organization

- Established in 1947, the *International Standards Organization (ISO)* is a multinational body dedicated to worldwide agreement on international standards. Almost three-fourths of countries in the world are represented in the ISO. An ISO standard that covers all aspects of network communications is the *Open Systems Interconnection (OSI)* model. It was first introduced in the late 1970s.

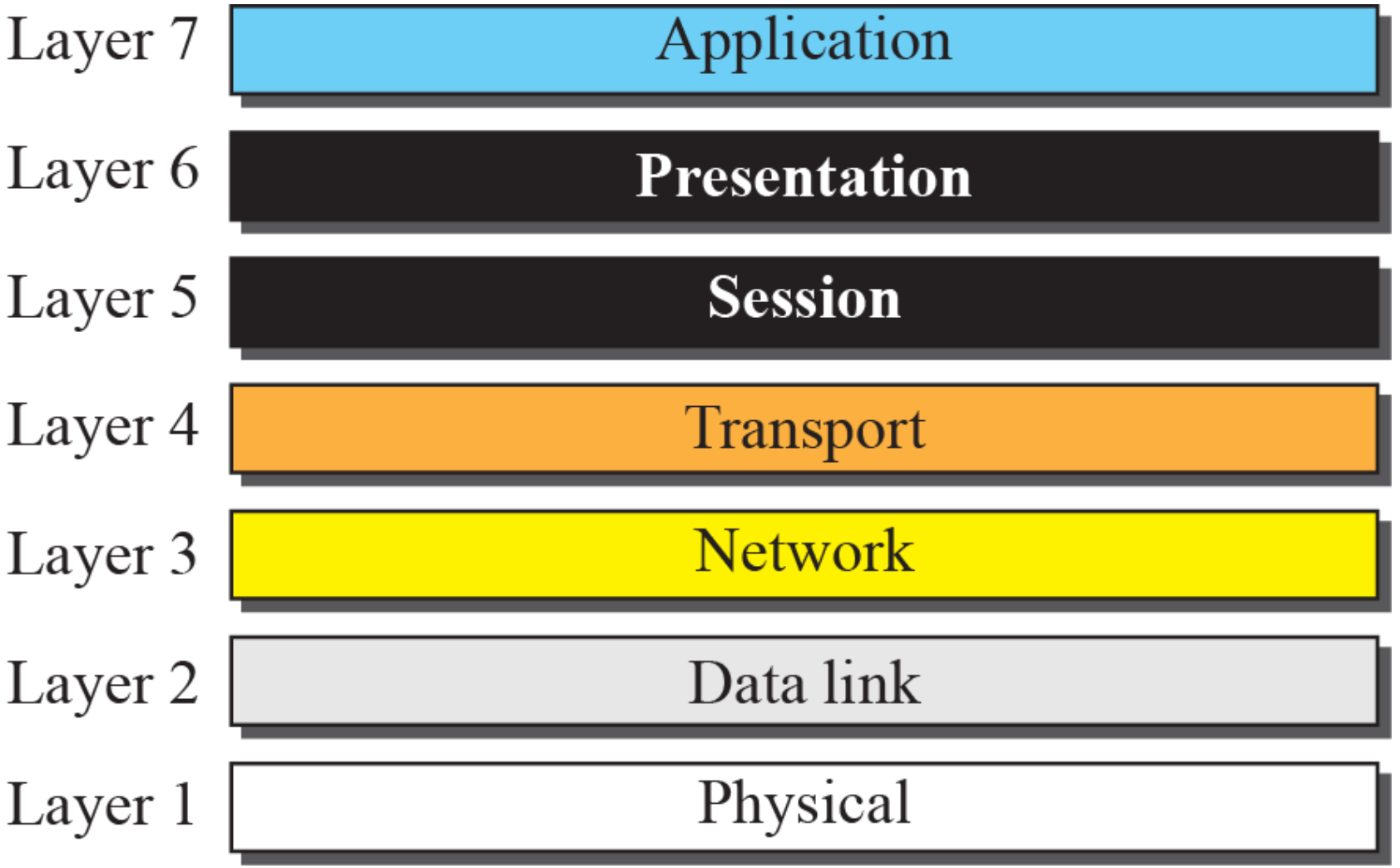
OSI Reference Model

- The OSI model is now considered the primary Architectural model for inter-computer communications.
- The OSI model describes how information or data makes its way from application programs (such as spreadsheets) through a network medium (such as wire) to another application program located on another network.
- The OSI reference model divides the problem of moving information between computers over a network medium into SEVEN smaller and more manageable problems .
- This separation into smaller more manageable functions is known as layering.

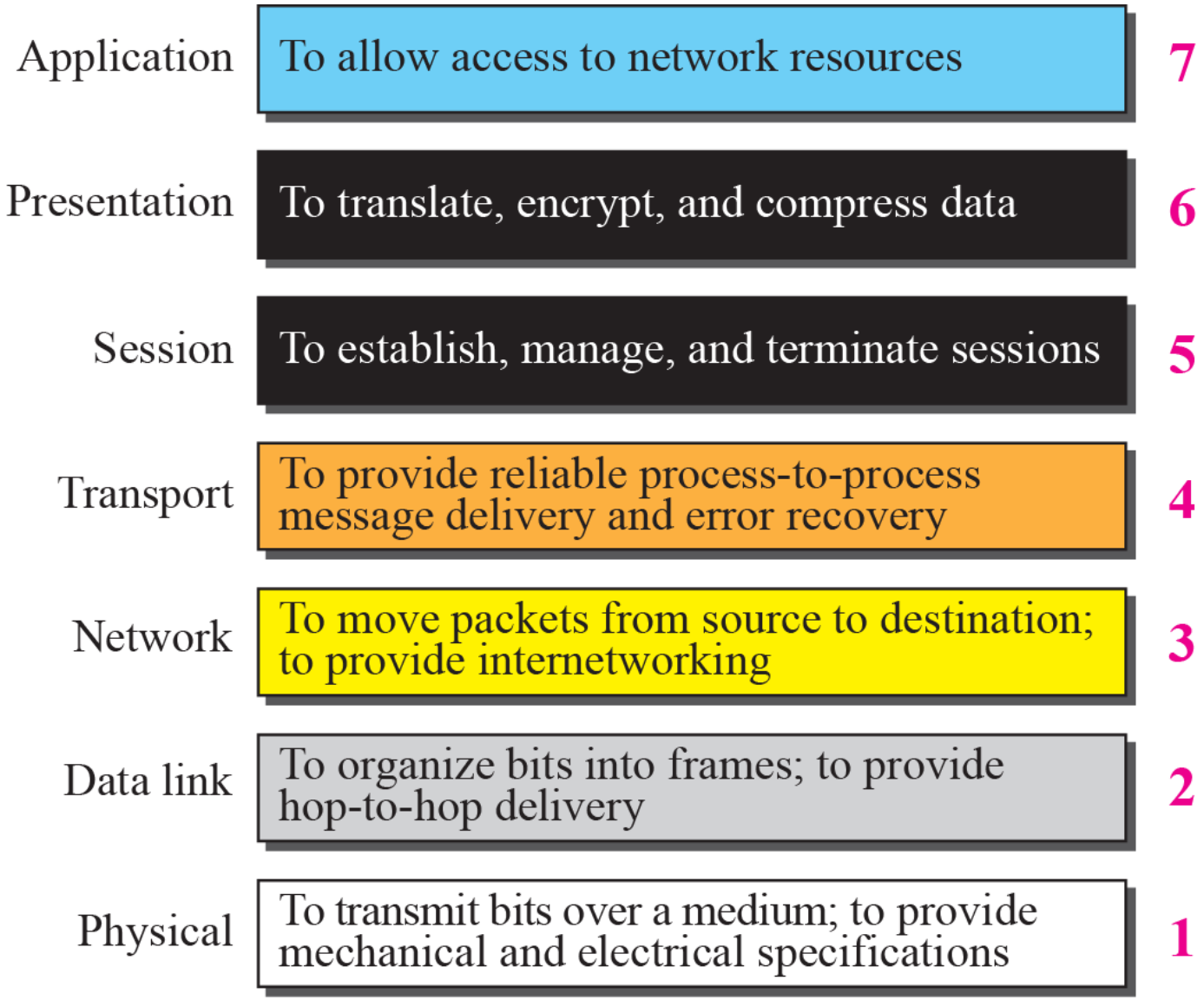
Objectives of OSI Layering

- To discuss the idea of multiple layering in data communication and networking and the interrelationship between layers.
- To discuss the OSI model and its layer architecture and to show the interface between the layers.
- To briefly discuss the functions of each layer in the OSI model.
- To introduce the TCP/IP protocol suite and compare its layers with the ones in the OSI model.
- To show the functionality of each layer in the TCP/IP protocol with some examples.
- To discuss the addressing mechanism used in some layers of the TCP/IP protocol suite for the delivery of a message from the source to the destination.

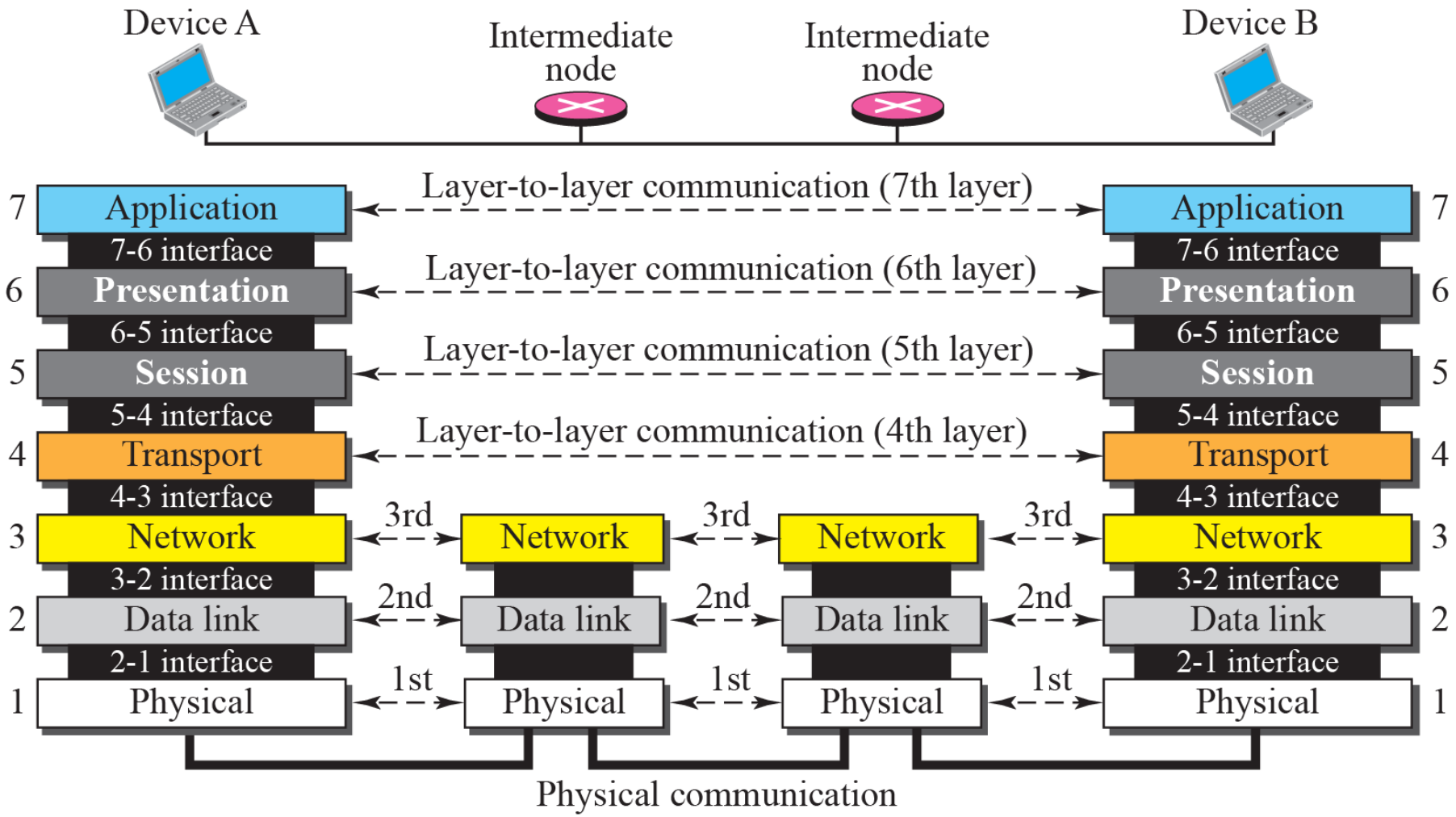
OSI Layers


Layer 7
Application
Layer 6
Presentation
Layer 5
Session
Layer 4
Transport
Layer 3
Network
Layer 2
Data link
Layer 1
Physical

Functional Overview of OSI Layers

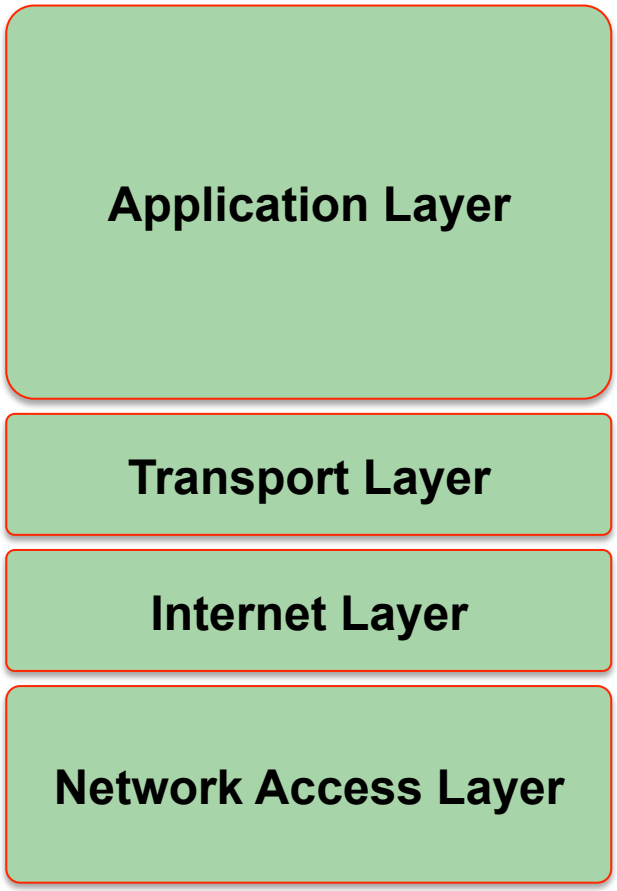


Communications in OSI Architecture



TCP/IP History

- Department of Defense (DoD) in the US created the TCP/IP reference model on the objective creating a resilient network.
- Few of the layers in the TCP/IP model have the same name as layers in the OSI model.



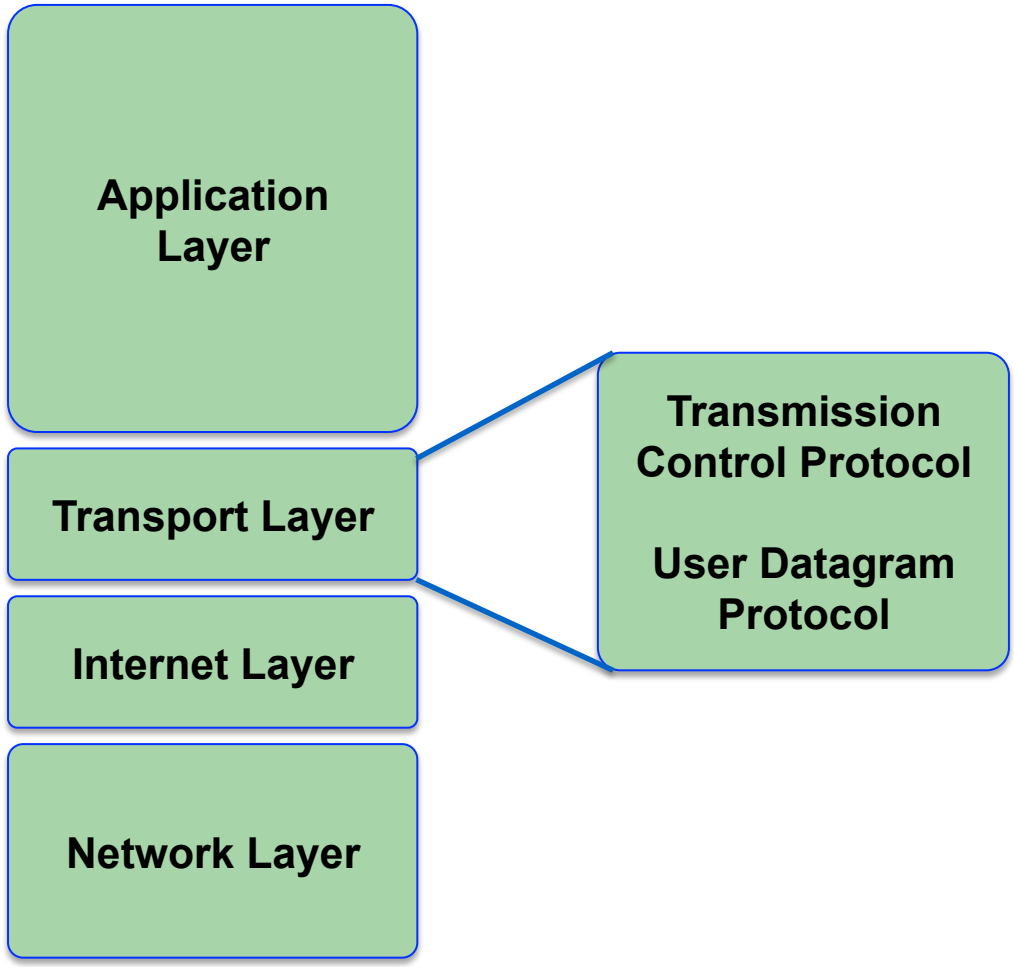
Application Layer functions & Examples

- Application Layer Functions:
  - Handles high-level protocols, issues of representation, encoding, and dialog control.
  - The TCP/IP protocol suite combines all application related issues into one layer and ensures this data is properly packaged before passing it on to the next layer.
- Application Layer Examples:
  - Telnet – Provides the capability to remotely access another computer
  - File Transfer Protocol – Download or upload files
  - Hypertext Transfer Protocol – Works with the World Wide Web

Transport Layer

Five basic services:

- Segmenting upper-layer application data
- Establishing end-to-end operations
- Sending segments from one end host to another end host
- Ensuring data reliability
- Providing flow control



Internet Layer = Network layer of OSI

- The purpose of the Internet layer is to send packets from a network node and have them arrive at the destination node independent of the path taken.
- Internet layer protocols:
  - Internet Protocol (IP)
  - Internet Control Message Protocol (ICMP)
  - Address Resolution Protocol (ARP)
  - Reverse Address Resolution Protocol (RARP)

Internet Layer = Network layer of OSI

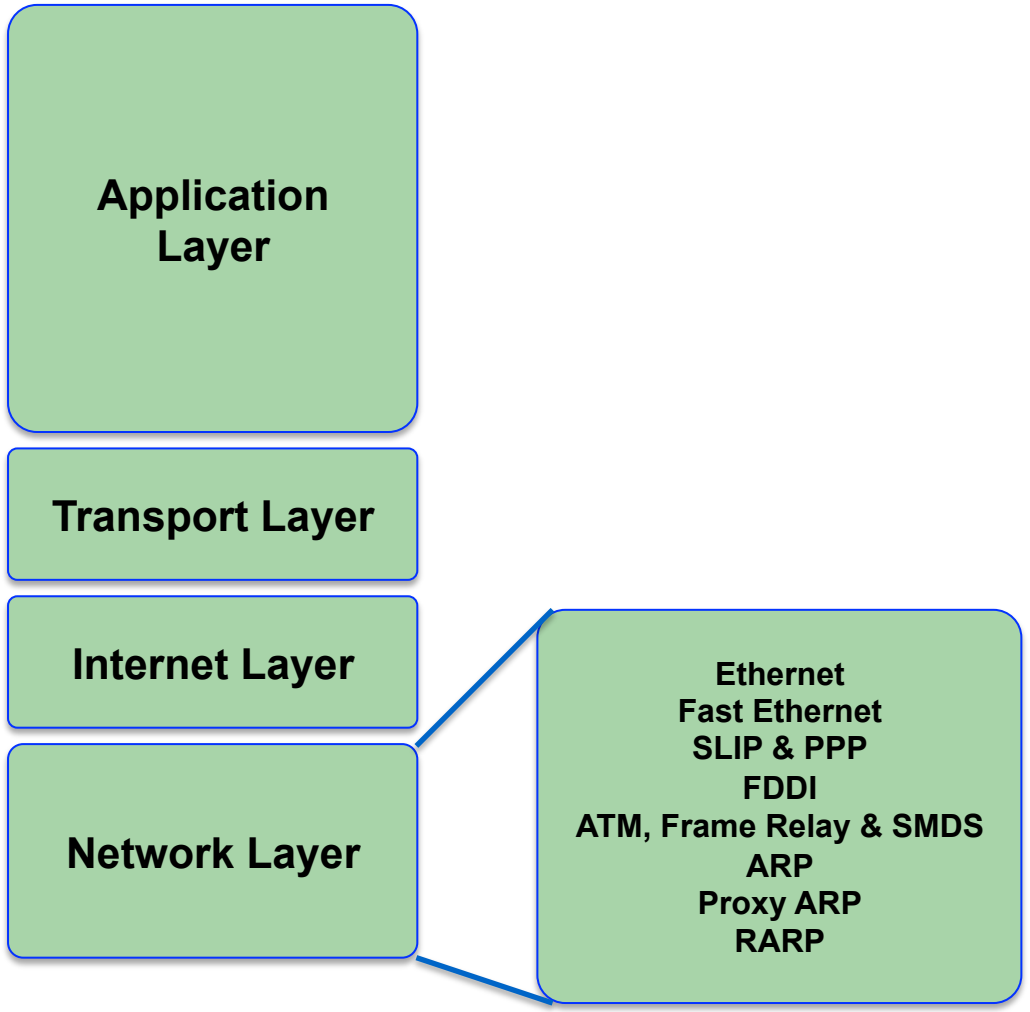
- Internet layer protocols functions:
  - IP header encapsulation for data from upper layer
  - Internet Control Message Protocol (ICMP) troubleshooting and supporting operations and maintenance
  - Address Resolution Protocol (ARP) is used to identify the physical layer addresses
  - Reverse Address Resolution Protocol (RARP) is used identify the IP address when a host/network element knows its link layer address such as the mac address.

Internet Layer = Network layer of OSI

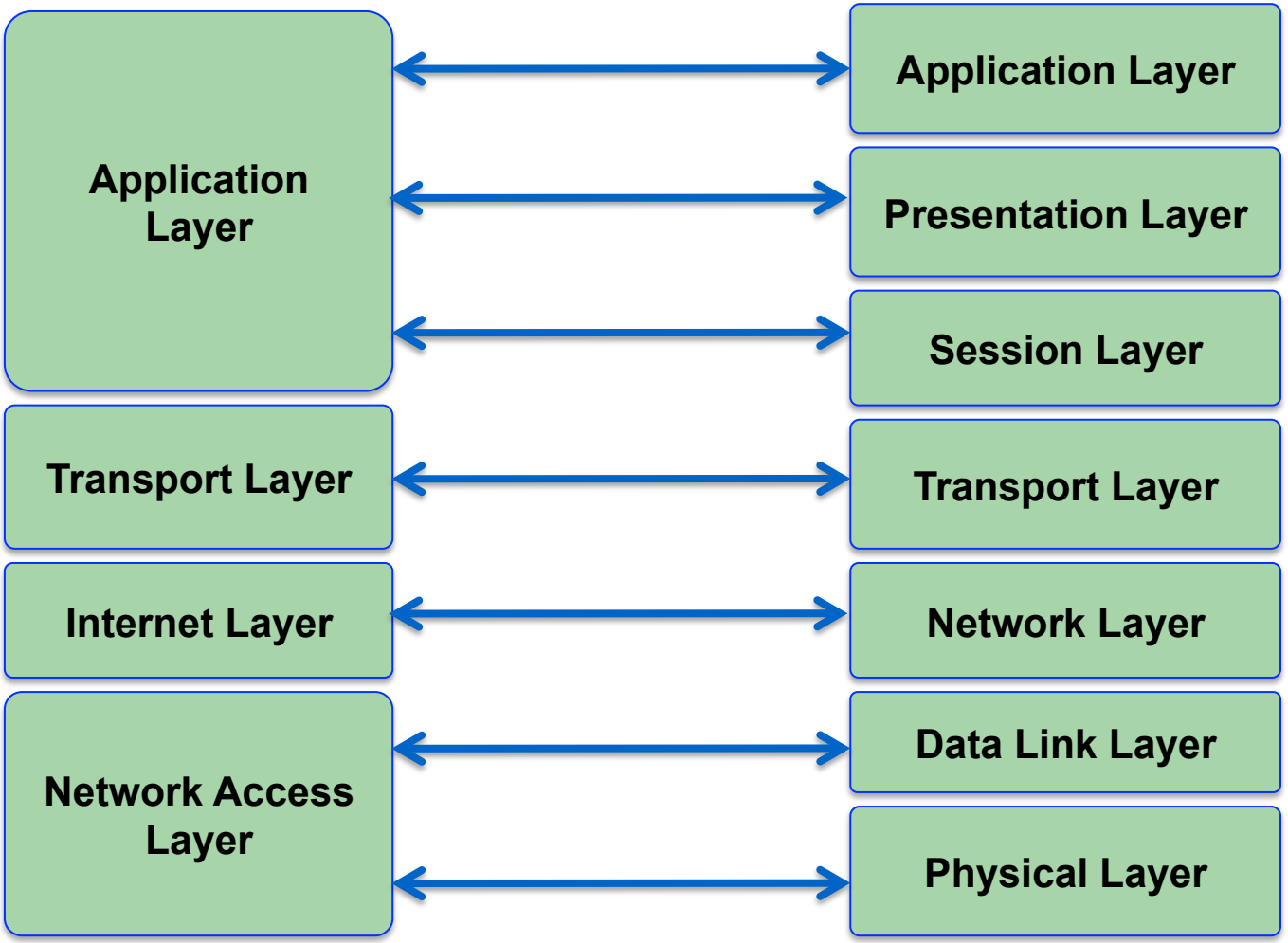
- Implements routing of frames (packets) through the network.
- Defines the most optimum path the packet should take from the source to the destination
- Defines logical addressing so that any endpoint can be identified.
- Handles congestion in the network.
- Facilitates interconnection between heterogeneous networks (Internetworking).
- The network layer also defines how to fragment a packet into smaller packets to accommodate different media.

Network Access Layer

- The network access layer is concerned with all of the issues that an IP packet requires to actually make a physical link to the network media.
- It includes the LAN and WAN technology details, and all the details contained in the OSI physical and data link layers.

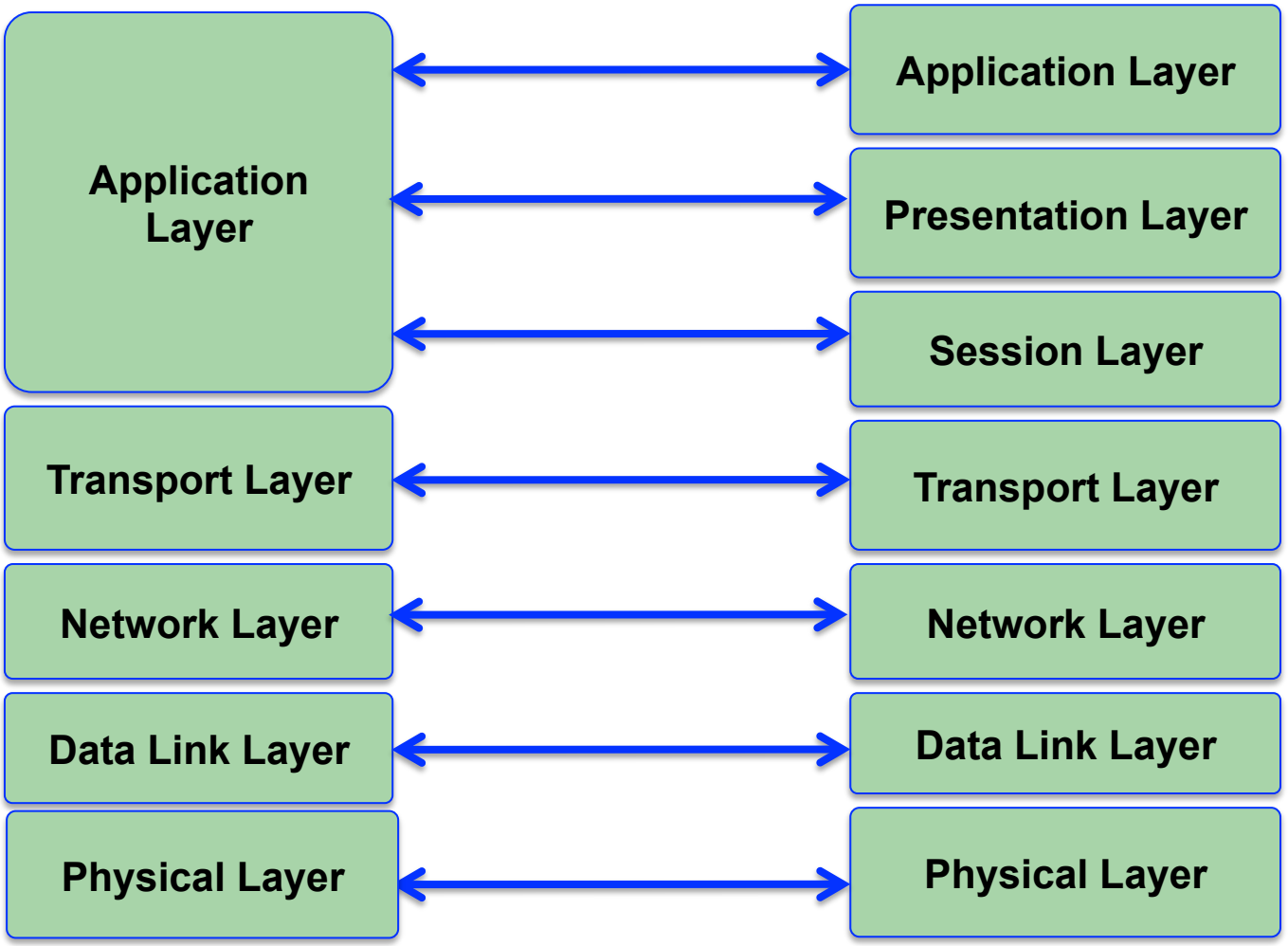


Comparing the OSI & TCP/IP Model



Originally as defined

Comparing the OSI & TCP/IP Model



Architecture used in this presentation

Similarities of the OSI & TCP/IP models

- Both have layers.
- Both have application layers, though they include very different services.
- Both have comparable transport and network layers.
- Packet-switched, not circuit-switched, technology is assumed.
- Networking professionals need to know both models.

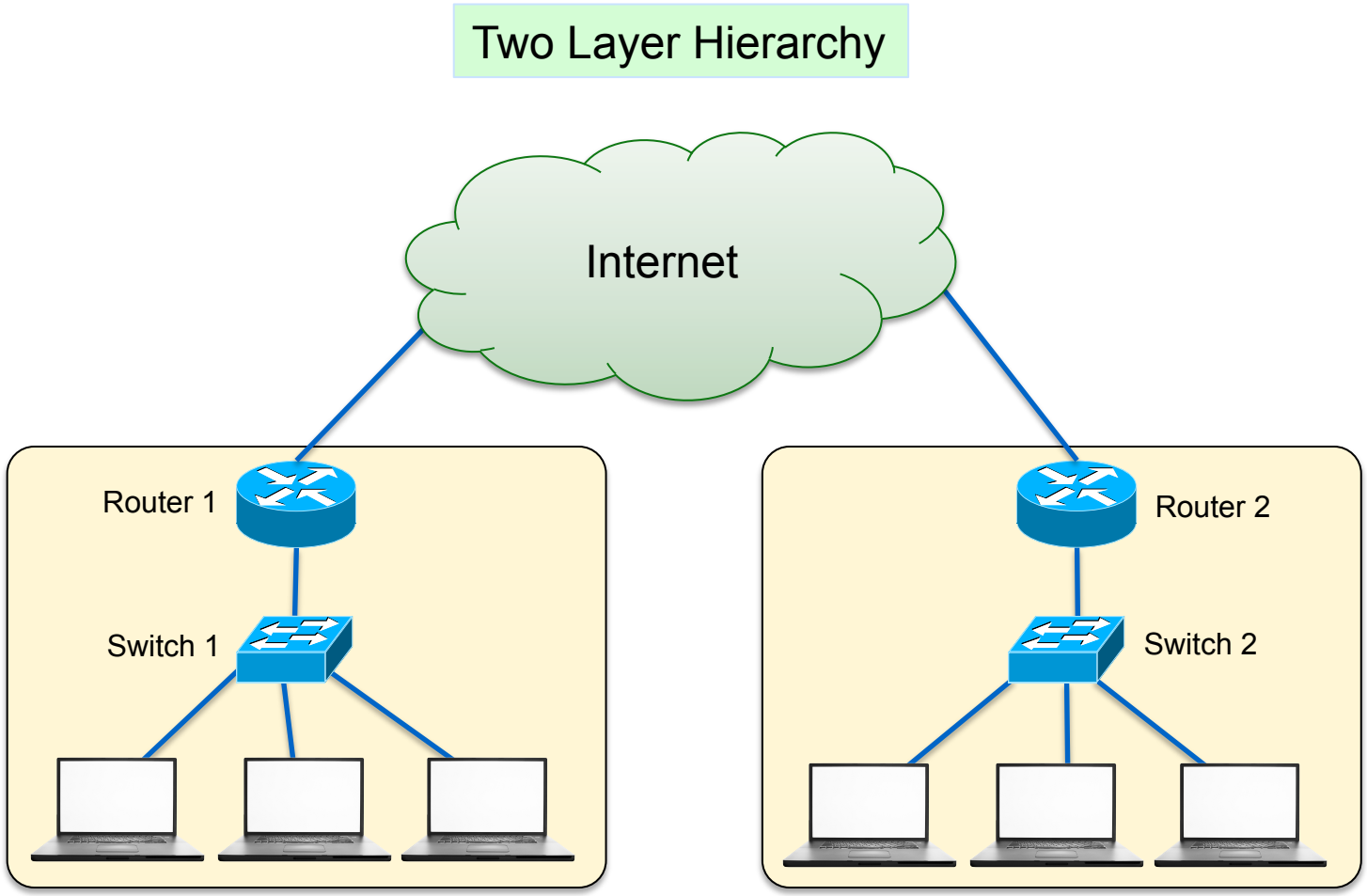
Differences of the OSI & TCP/IP models

- TCP/IP combines the presentation and session layer into its application layer.
- TCP/IP combines the OSI data link and physical layers into one layer.
- TCP/IP appears simpler because it has fewer layers.
- TCP/IP transport layer using UDP does not always guarantee reliable delivery of packets as the transport layer in the OSI model does.

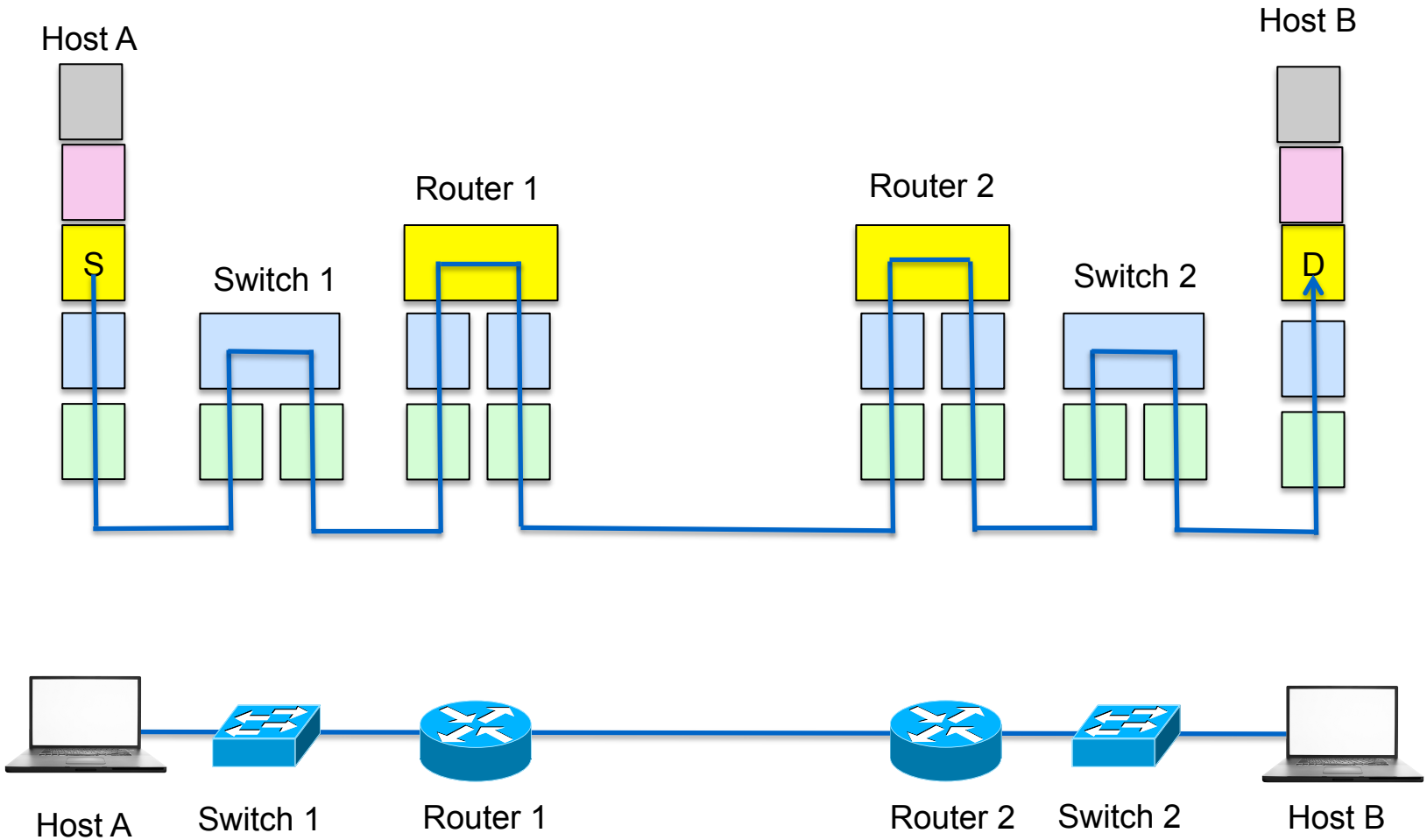
Internet Architecture

- Two computers, anywhere in the world, following certain hardware, software, protocol specifications, can communicate, reliably even when not directly connected.
- LANs are no longer scalable beyond a certain number of stations or geographic separation.

Internet Architecture: Idea of Routing



Typical communication in a Network



Network or Internetwork Layer Functions

- Design Goals
  - ‘Independent’ of layer 1 & 2 implementations
  - Hide layer 1 & 2 details from upper layers
- Architecture
  - Connection oriented
  - Connectionless
  - (where should reliability be done?)
- Services
  - Routing (Path selection)
  - Adaptation to different lower layers

Objectives for Routing Algorithms

- Goals
  - Optimality
  - Fairness
  - Stability
  - Robustness
  - Correctness
  - Simplicity
- Adaptive versus Static
- Congestion Control

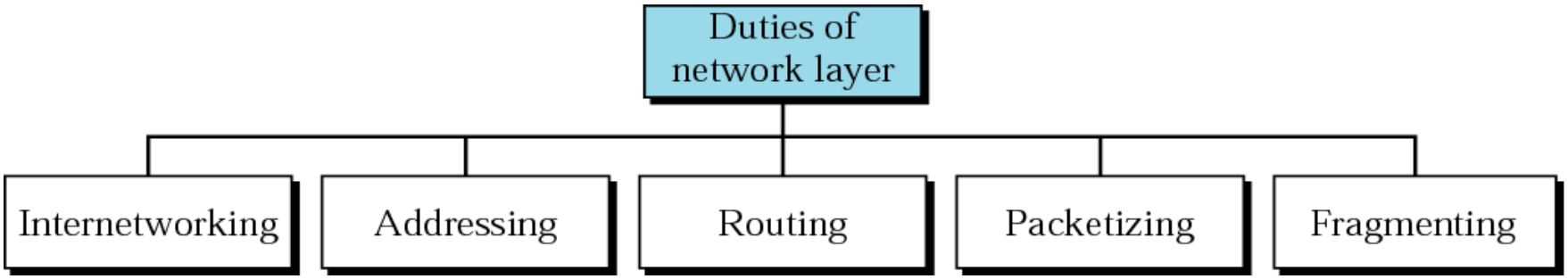
Adaptive Routing

- Centralized
- Isolated
- Distributed

Distributed Routing

- Metric - Vector Algorithms
  - sometimes called shortest path
  - Bellman-Ford most famous
  - Knowledge of immediate neighbors
  - Result is “first step” in path to ultimate destination
- Link State Algorithms
  - OSPF {Open Shortest Path First}
  - Knowledge of network layer map (connectivity)

Summary of Network layer functions



IP Addressing and its role in Routing

APNIC

Why do we need IP addresses?

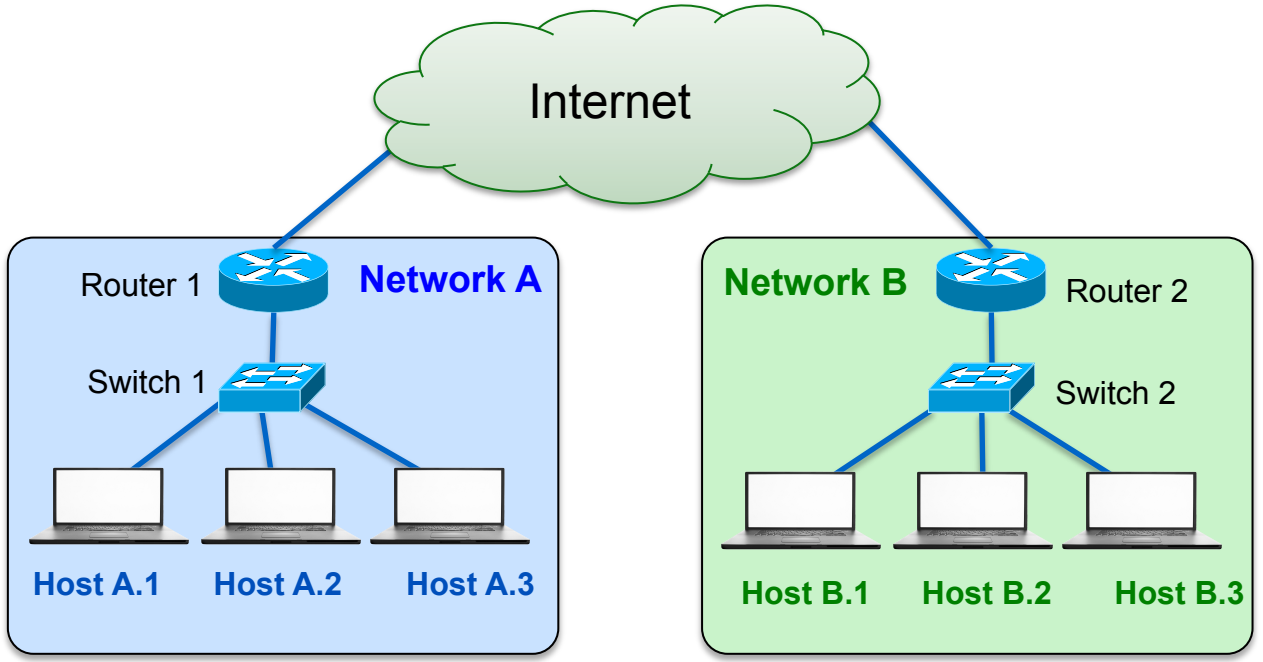
- Impossible to engineer a universal network from a single network technology because no single network suffices for all uses
- Some technologies such as Ethernet LANs are good for high speed connection within a Local Area only.
- Some technologies such as Serial Interfaces were very good for use within in a Wide Area environment
- Critical need for internetworking between various LAN and WAN technologies
- Different technological approaches create issues such as Physical addresses used in one technology vs another. Example Ethernet Mac address, Addressing Token ring, FDDI Addressing etc.

Why do we need IP addresses?

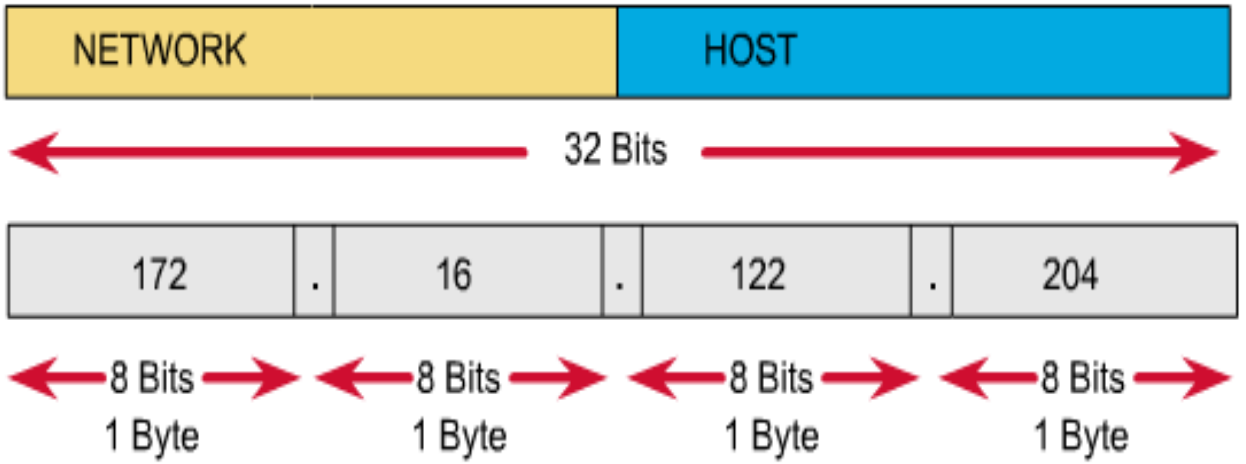
- Networks commonly built using disparate technologies
- Internetworking using Logical Addressing helps connectivity of disparate networks
- This approach hides the details of the underlying L2 and L1 technologies completely
- The primary goal is a system that hides the details of underlying network hardware while providing universal communication services.
- Two fundamental observations about the design of communication systems:
  - Network-Level Interconnection a No single network hardware technology can satisfy all constraints.
  - Users desire universal interconnection
- The need for all computers to communicate using a universal set of machine identifiers. Unique IP Addresses become a key requirement

IP Addresses in a Network

- Every host in the Internet requires a unique IP address for communication. This is a key architectural requirement
- An IP address uniquely identifies a network interface on a host.
- A host may have many interfaces
- Each IP Address consists of a network portion and a host portion.
- Combining network and host portion of the IP address, all machines on the internet can be uniquely identified on the internet.

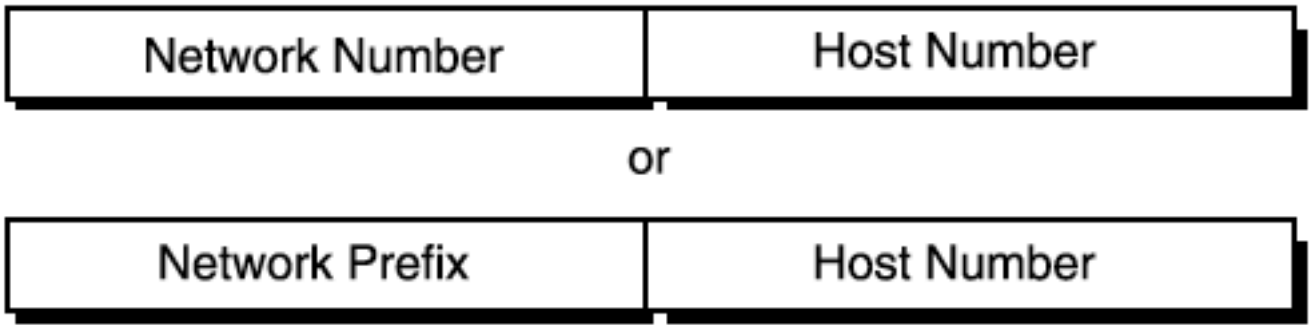


IP Address: 32-Bit Binary Number



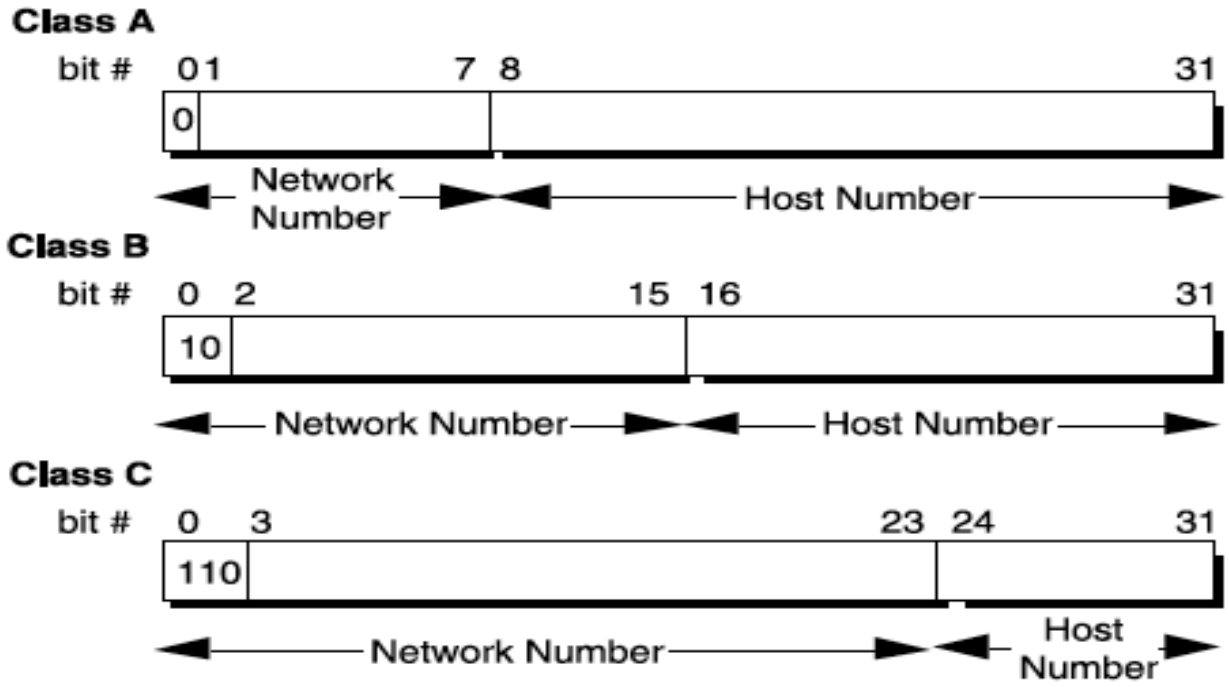
History of IP Addressing

- Classful IP Addressing
  - When IP was first standardized in September 1981, the specification required that each system attached to an IP-based Internet be assigned a unique, 32-bit Internet address value.
- Systems that have interfaces to more than one network require a unique IP address for each network interface.
- The first part of an Internet address identifies the network on which the host resides, while the second part identifies the particular host on the given network.



Classful IP Addressing

- To provide the flexibility required to support networks of varying sizes, the Internet designers decided that the IP address space should be divided into three address classes-Class A, Class B, and Class C. This is often referred to as Classful addressing.



Dotted-Decimal Notation

Address Class	Dotted-Decimal Notation Ranges
A (/8 prefixes)	1.xxx.xxx.xxx through 126.xxx.xxx.xxx
B (/16 prefixes)	128.0.xxx.xxx through 191.255.xxx.xxx
C (/24 prefixes)	192.0.0.xxx through 223.255.255.xxx

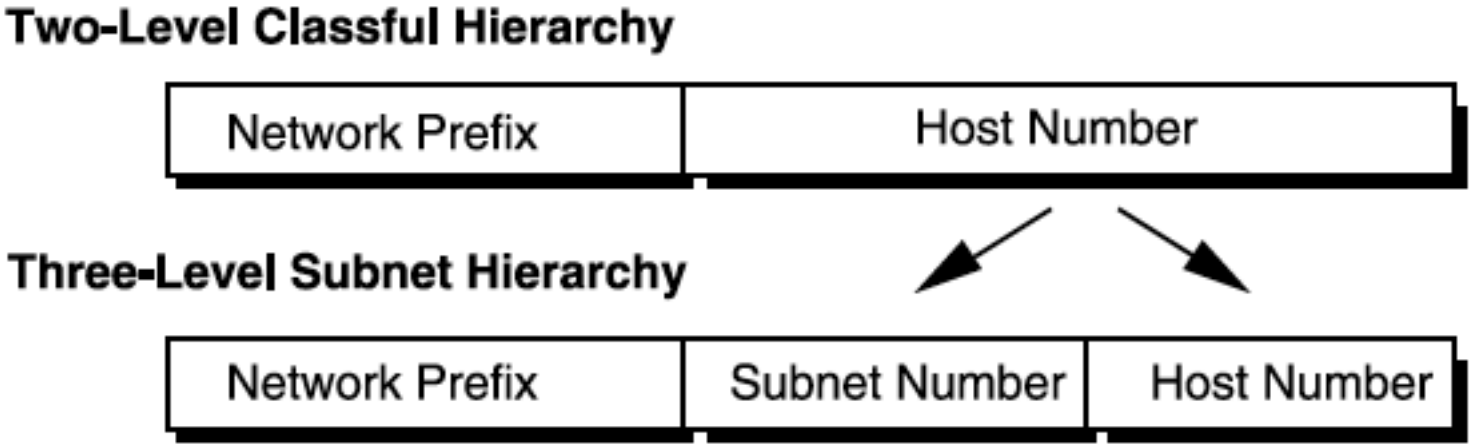
To make Internet addresses easier for people to read and write, IP addresses are often expressed as four decimal numbers, each separated by a dot. This format is called “dotted-decimal notation.”

Limitations to Classful Addressing

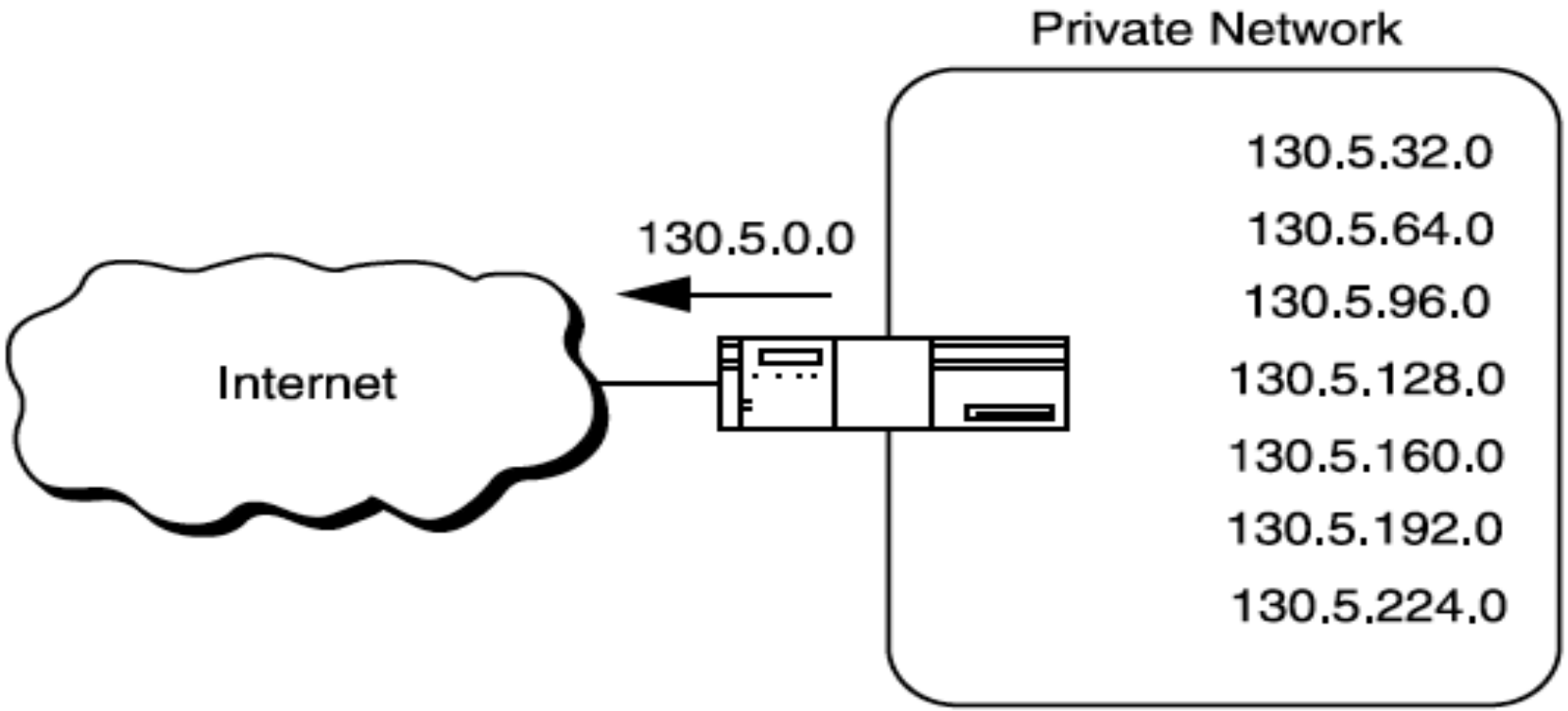
- IP Addresses were allocated to an organization based on its request rather than its need.
- The decision to standardize on a 32-bit address space meant that there were only 232 (4,294,967,296) IPv4 addresses available.
- The Classful A, B, and C octet boundaries were easy to understand and implement, but they did not foster the efficient allocation of a finite address space

Subnetting

- In 1985, RFC 950 defined a standard procedure to support the subnetting, or division, of a single Class A, B, or C network number into smaller pieces.
  - Internet routing tables were beginning to grow.
  - Local administrators had to request another network number from the Internet addressing agency before a new network could be installed at their site.
- Resolved using a Subnet architecture



Subnetting & Routing Requirements



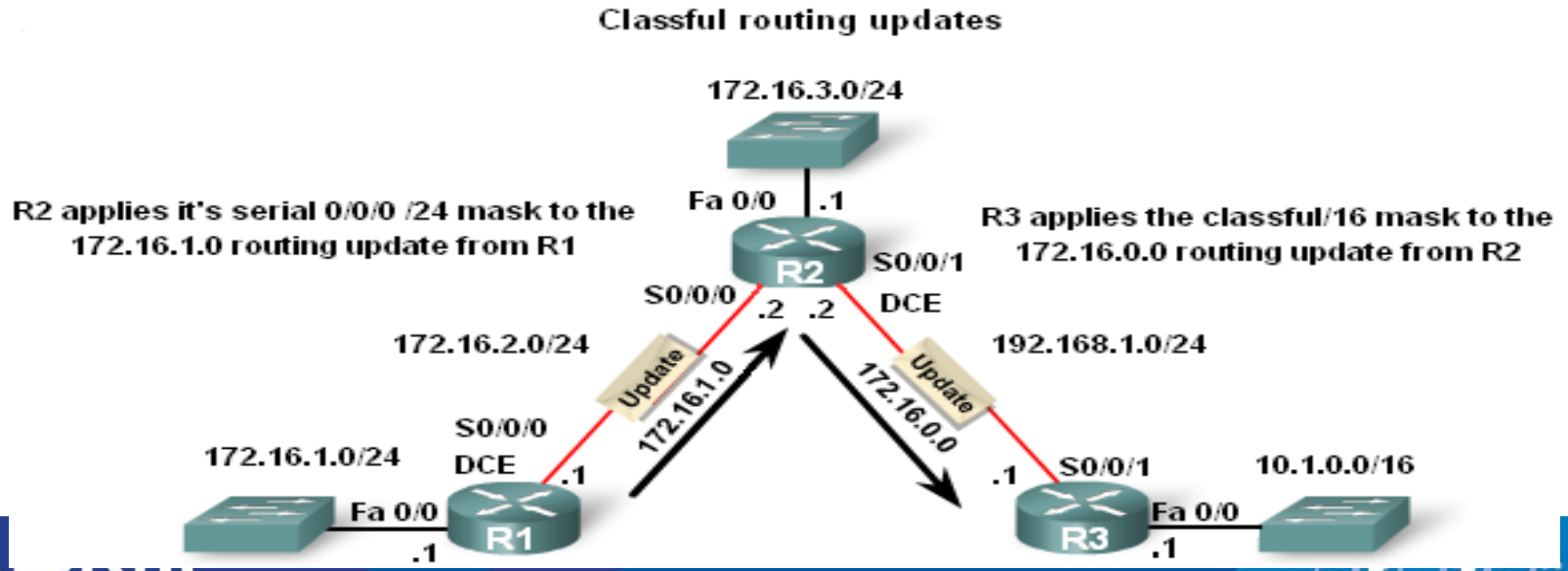
The router accepts all traffic from the Internet addressed to network 130.5.0.0, and forwards traffic to the interior sub networks based on the third octet of the Classful address.

Advantages of Subnetting

- The size of the global Internet routing table does not grow
- Net Administrator has the flexibility to deploy additional subnets without obtaining a new network number from the Internet agency
- Route flapping (that is, the rapid changing of routes) within the private network does not affect the Internet routing table as they need not know about the reachability of each subnet

Routing Protocols & Classful Addressing

- Classful Routing Updates
  - Classful routing protocols (i.e. RIPv1, IGRP & BGP v3) do not send subnet masks in their routing updates.
  - The routing update could determine the subnet mask simply by examining the value of the first octet in the network address



Classless Inter-domain Routing (CIDR – RFC 1519)

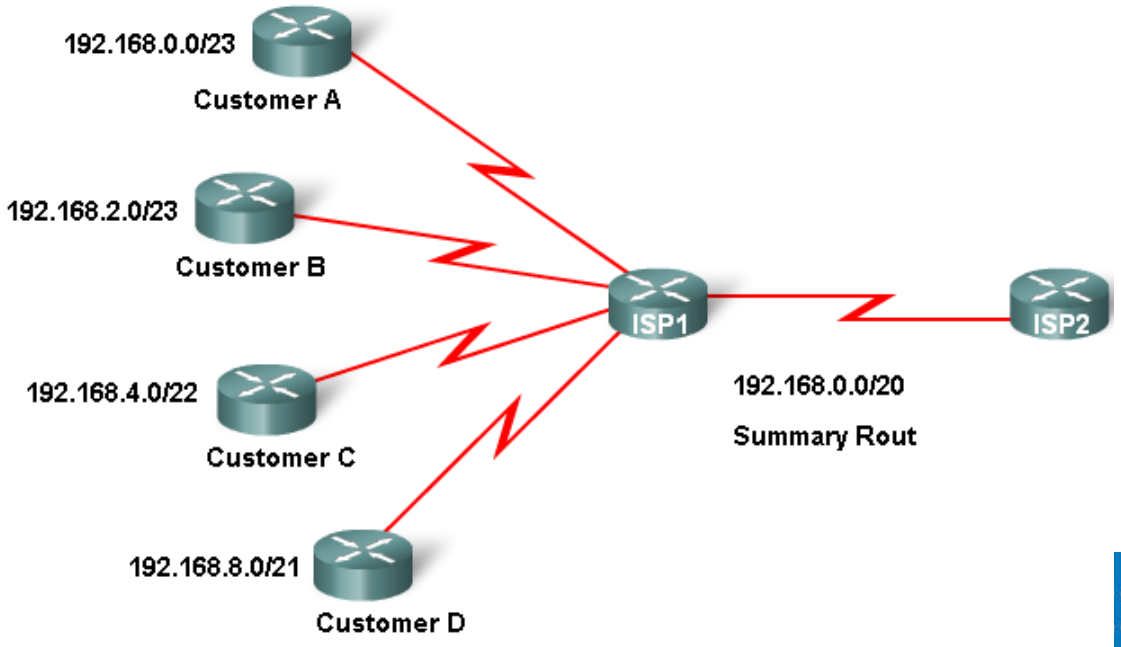
- Advantages of CIDR
  - More efficient use of IPv4 address space
  - Route summarization
    - reduce routing table size
    - reduce routing update traffic

Classless Inter-domain Routing (CIDR – RFC 1519)

- Requires subnet mask to be included in routing update because address class is meaningless
- The network portion of the address is determined by the network subnet mask, also known as the network prefix, or prefix length (/8, /19, etc.).
- The network address is no longer determined by the class of the address
- Blocks of IP addresses could be assigned to a network based on the requirements of the customer, ranging from a few hosts to hundreds or thousands of hosts.

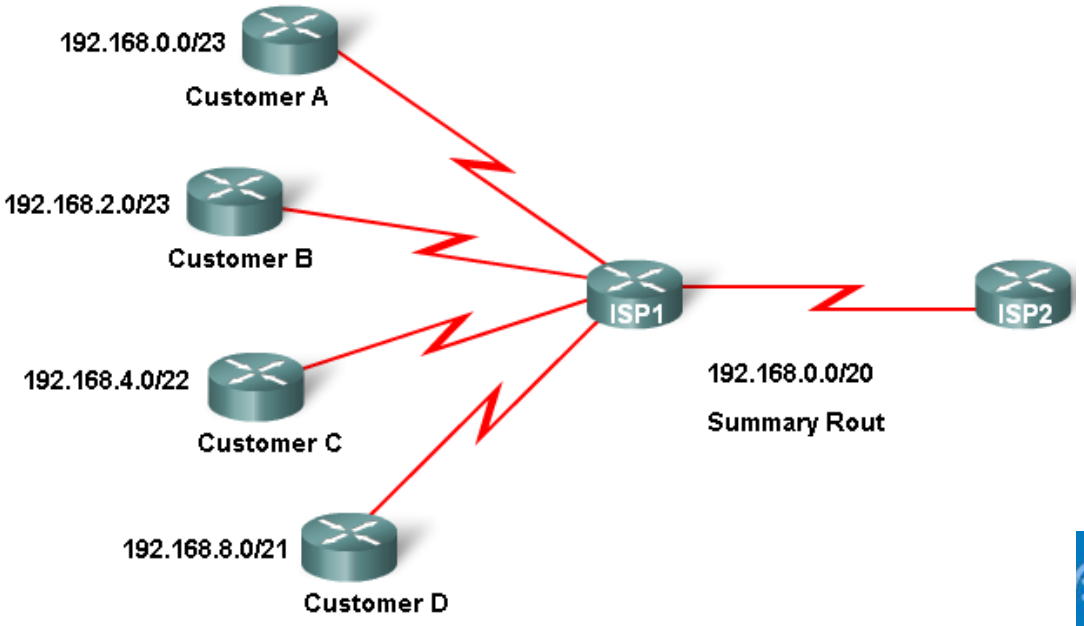
Classless IP Addressing

- Classless IP Addressing
- CIDR & Route Summarization
  - –Variable Length Subnet Masking (VLSM)
  - –Allows a subnet to be further sub-netted
    - according to individual needs
  - –Prefix Aggregation a.k.a. Route Summarization
  - –CIDR allows for routes to be summarized as a single route



Classless IP Addressing and Routing

- For Route Summarization
  - Propagating VLSM and supernet routes requires a classless routing protocol, because the subnet mask can no longer be determined by the value of the first octet.
  - Classless routing protocols include the subnet mask with the network address in the routing update.
  - RIPv2, EIGRP, IS-IS, OSPF and BGP



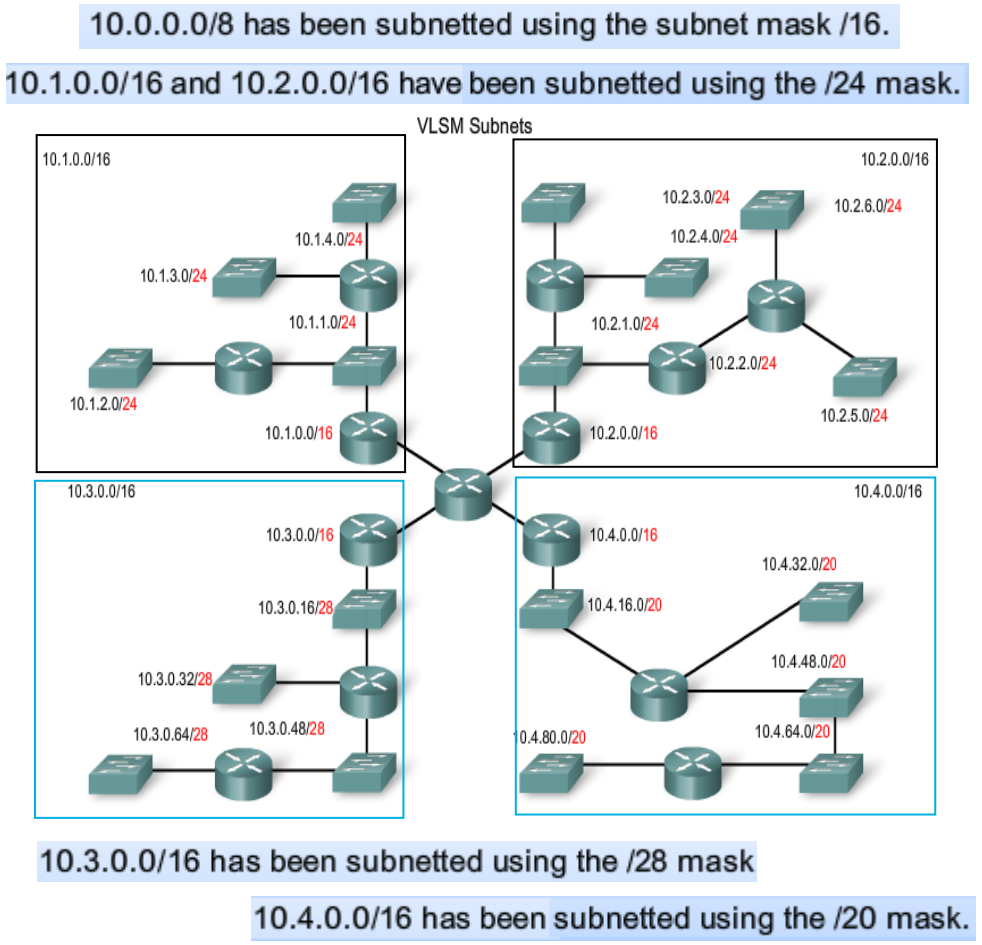
VLSM

- Classful routing
  - -only allows for one subnet mask for all networks
- VLSM & classless routing

  -This is the process of subnetting a subnet

  -More than one subnet mask can be used

  -More efficient use of IP addresses as compared to classful IP addressing

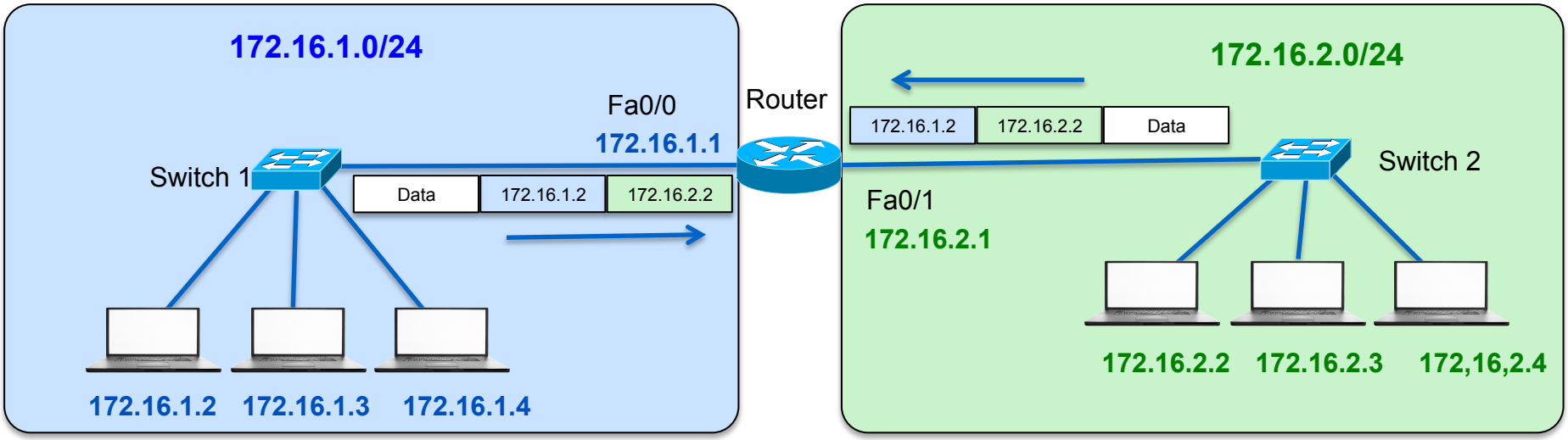


The need to route

Router as a Computer

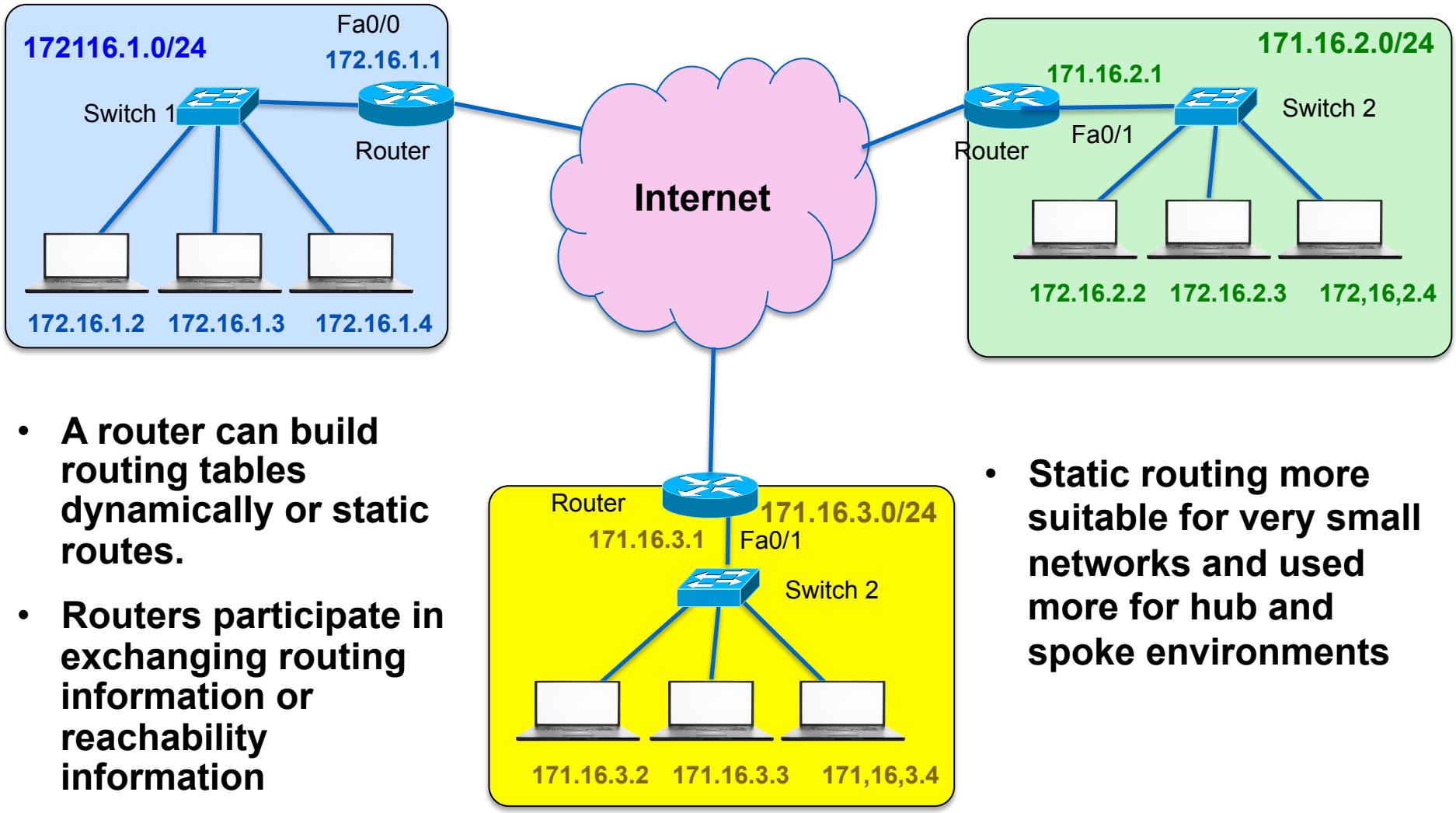
- Router Interface is a physical connector
- Each interface connects to a separate network
- Consist of socket or jack found on the outside of a router
- Types of router interfaces:
  - –-Ethernet
  - –-Fast Ethernet
  - –-Serial
  - –-DSL
  - –-ISDN
  - –-Cable

The concept of Routing



- Every IP packet originated by the hosts contains a source ip address and a destination IP address.
- The router uses this information to forward packets based on the destination IP address
- For example: Packets arriving on Fa0/0 with Source IP Address 172.16.1.2 and destination IP Address of 172.16.2.2 will be forwarded out on Fa0/1 and Vice Versa
- This technique helps to expand to multiple routers involved in routing to scale connectivity between hosts on different networks.

Scaling connectivity requires Routing



- **A router can build routing tables dynamically or static routes.**
- **Routers participate in exchanging routing information or reachability information**

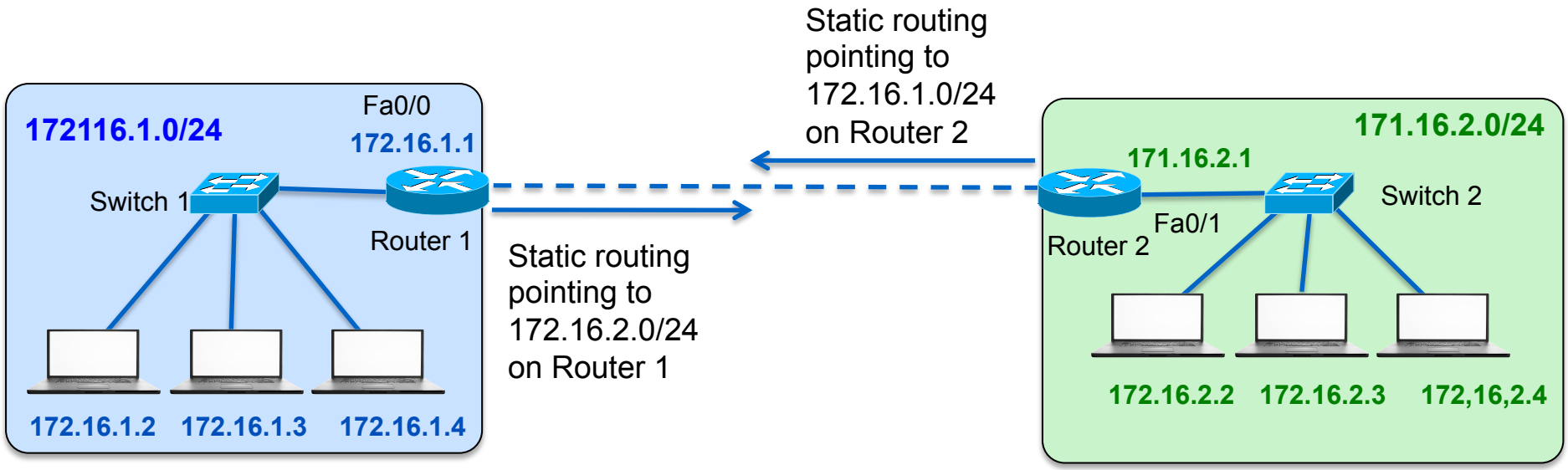
- **Static routing more suitable for very small networks and used more for hub and spoke environments**

Static Routing

APNIC

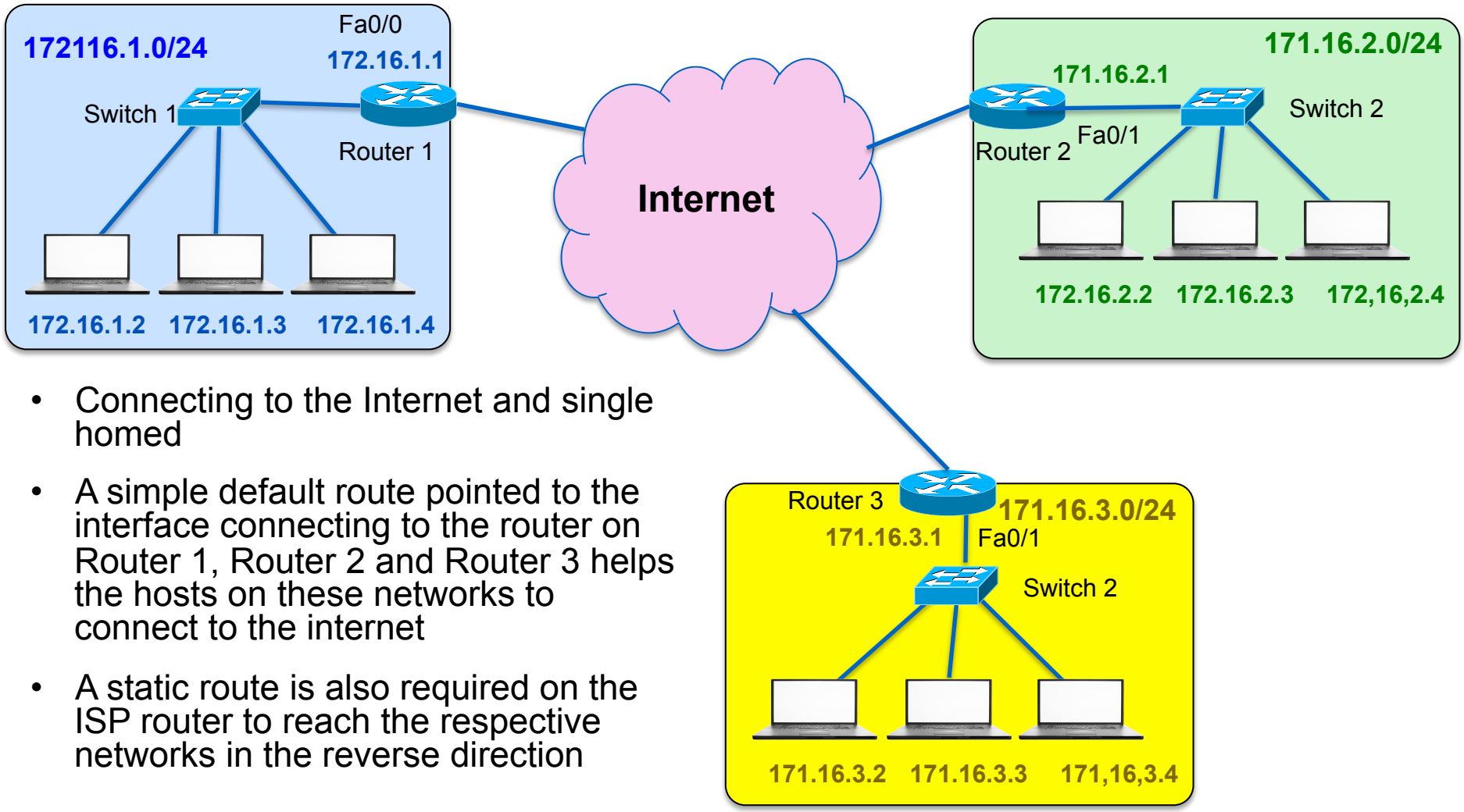
Static routing

- A manually configured route on a router to reach a specific destination network



- Useful for small networks
- Mostly used in hub and spoke networks
- Connecting to the Internet and single homed

Static Routing Scenarios



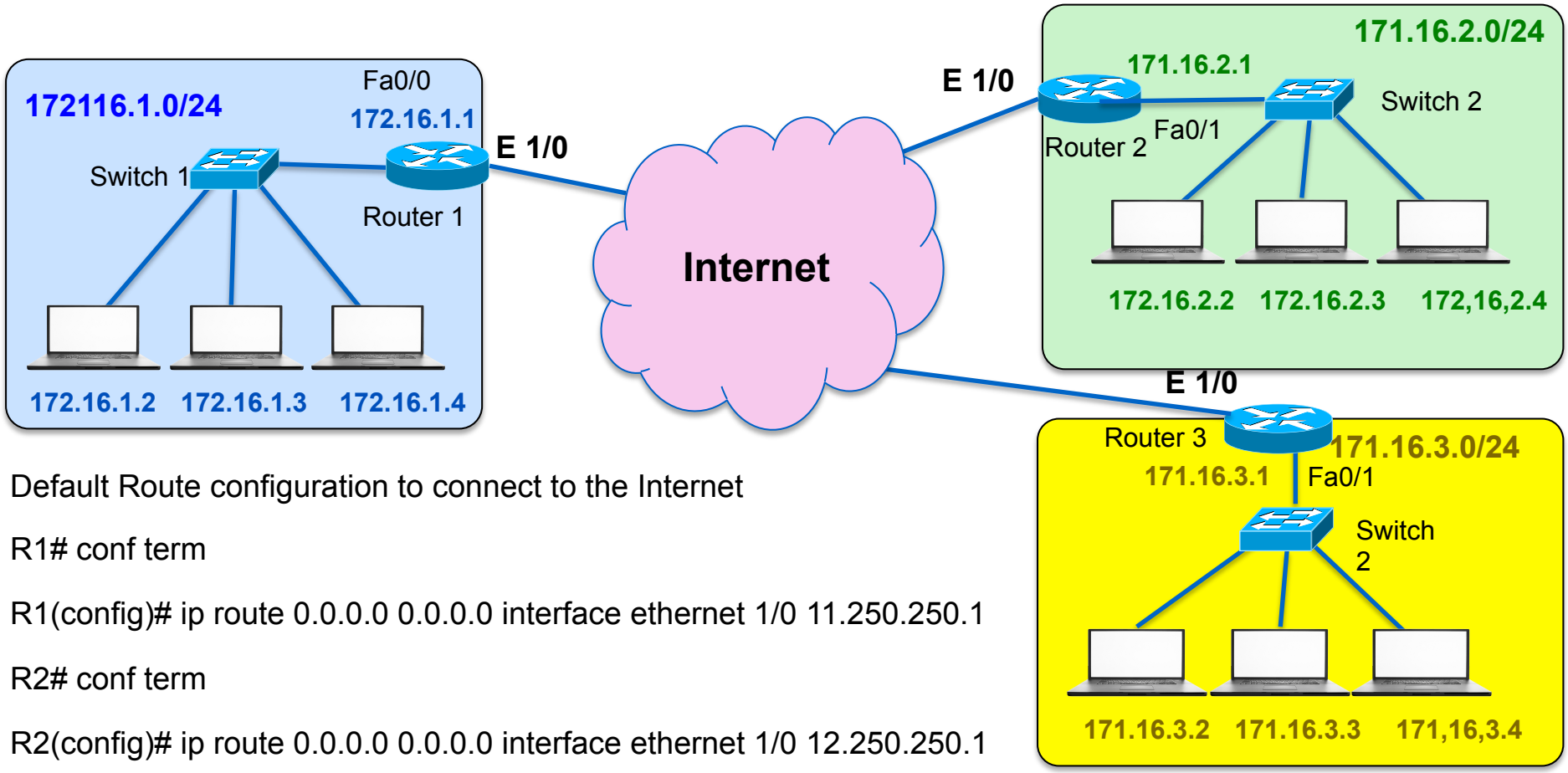
- Connecting to the Internet and single homed
- A simple default route pointed to the interface connecting to the router on Router 1, Router 2 and Router 3 helps the hosts on these networks to connect to the internet
- A static route is also required on the ISP router to reach the respective networks in the reverse direction

Static Route Configuration

```
Router(config)# ip route network-address subnet-mask
{ip-address | exit-interface }
```

Parameter	Description
network-address	Destination network address of the remote network to be added to the routing table.
subnet-mask	Subnet mask of the remote network to be added to the routing table. The subnet mask can be modified to summarize a group of networks.
ip-address	Commonly referred to as the next-hop router's IP address.
exit-interface	Outgoing interface that is used to forward packets to the destination network.

Static routing configuration Example



Default Route configuration to connect to the Internet

R1# conf term

R1(config)# ip route 0.0.0.0 0.0.0.0 interface ethernet 1/0 11.250.250.1

R2# conf term

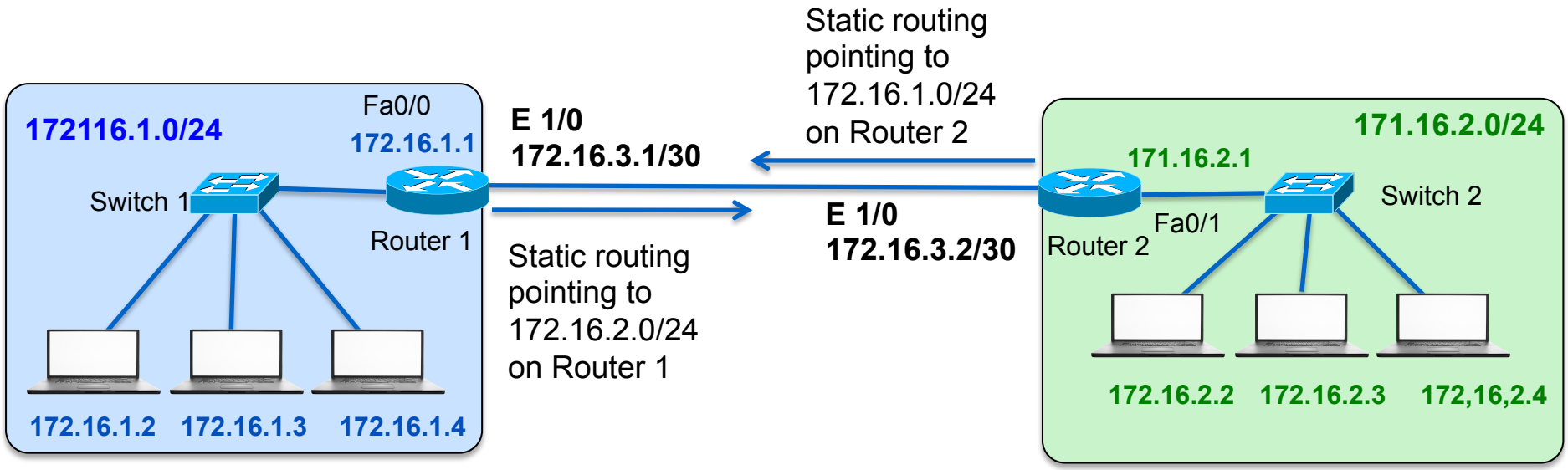
R2(config)# ip route 0.0.0.0 0.0.0.0 interface ethernet 1/0 12.250.250.1

R3# conf term

R3(config)# ip route 0.0.0.0 0.0.0.0 interface ethernet 1/0 13.250.250.1

Static routing

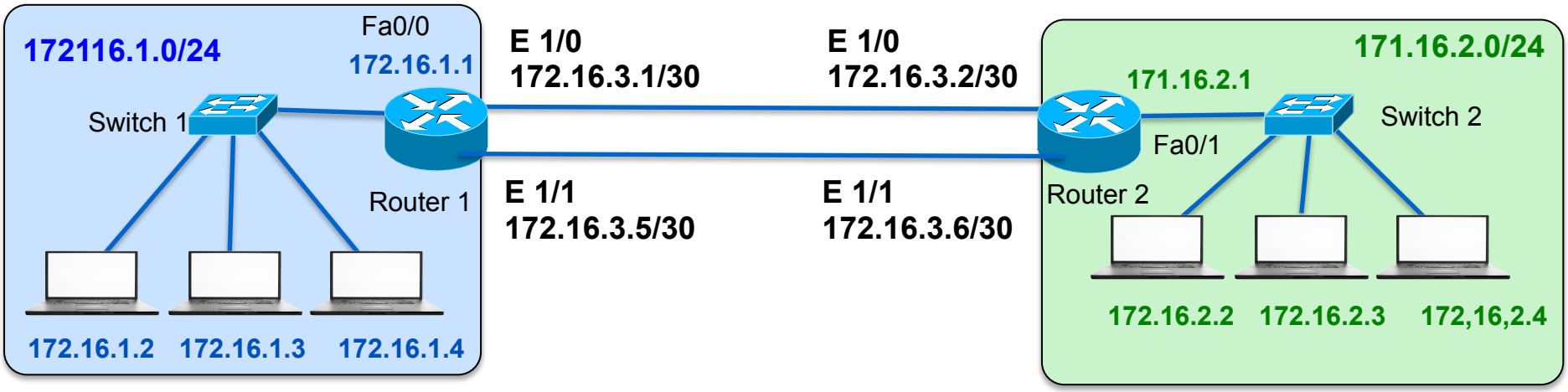
- A manually configured route on a router to reach a specific destination network



Router1# conf term
Router1(config)# ip route 172.16.2.0 255.255.255.0 ethernet 1/0 172.16.3.2

Router2# conf term
Router2(config)# ip route 172.16.1.0 255.255.255.0 ethernet 1/0 172.16.3.1

Load sharing using static routing



Router1# conf term

Router1(config)# ip route 172.16.2.0 255.255.255.0 ethernet 1/0 172.16.3.2

Router1(config)# ip route 172.16.2.0 255.255.255.0 ethernet 1/1 172.16.3.6

Router2# conf term

Router2(config)# ip route 172.16.1.0 255.255.255.0 ethernet 1/0 172.16.3.1

Router2(config)# ip route 172.16.1.0 255.255.255.0 ethernet 1/1 172.16.3.5

Pros and Cons of Static routing

- Advantages of static routing
  - It can backup multiple interfaces/networks on a router
  - Easy to configure
  - No extra resources are needed
- -More secure
- Disadvantages of static routing
  - Network changes require manual reconfiguration
  - Does not scale well in large topologies

Dynamic Routing

APNIC

Characteristics of Dynamic Routing

- Dynamic routing protocols fulfill the following functions
  - Dynamically share information between routers
  - Automatically update routing table when topology changes
  - Determine best path to a destination
- Routing protocols are grouped as either
  - Interior gateway protocols (IGP)Or
  - Exterior gateway protocols(EGP)

Terminology

- **Dynamic routing protocols** fulfill the following functions
  - –-Dynamically share information between routers
  - –-Automatically update routing table when topology changes
  - –-Determine best path to a destination
- **Routing protocols are grouped as either**
  - –-Interior gateway protocols (IGP)**Or**
  - –-Exterior gateway protocols(EGP)
- **Types of IGPs include**
  - –-Classless routing protocols - these protocols include subnet mask in routing updates
  - –-Classful routing protocols - these protocols do not include subnet mask in routing update

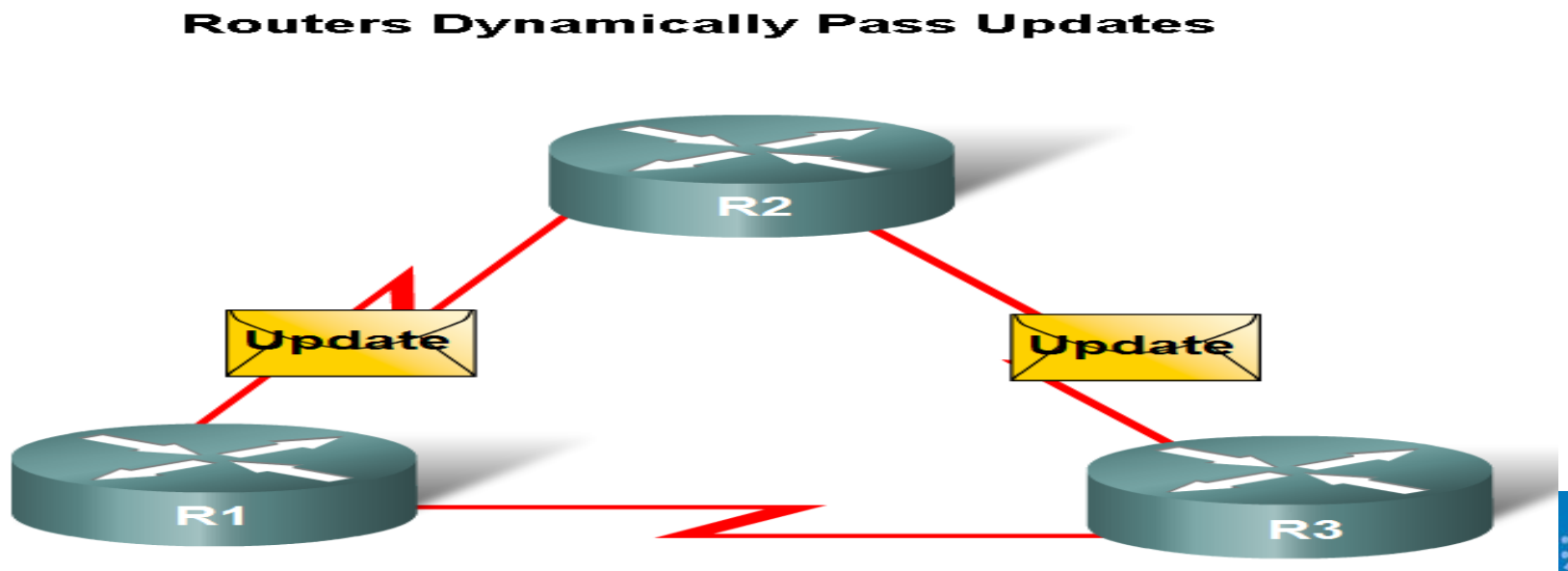
Terminology

- **Metrics** are used by dynamic routing protocols to calculate the best path to a destination.
- **Administrative distance** is the feature that routers use in order to select the best path when there are two or more different routes to the same destination from two different routing protocols. Administrative distance defines the reliability of a routing protocol. Each routing protocol is prioritized in order of most to least reliable (believable) with the help of an administrative distance value.”
- **Components of a routing table** include:
  - –-Route source
  - –-Administrative distance
  - –-Metric

Dynamic Routing Protocols

Function(s) of Dynamic Routing Protocols:

- –-Dynamically share information between routers.
- –-Automatically update routing table when topology changes.
- –-Determine best path to a destination.



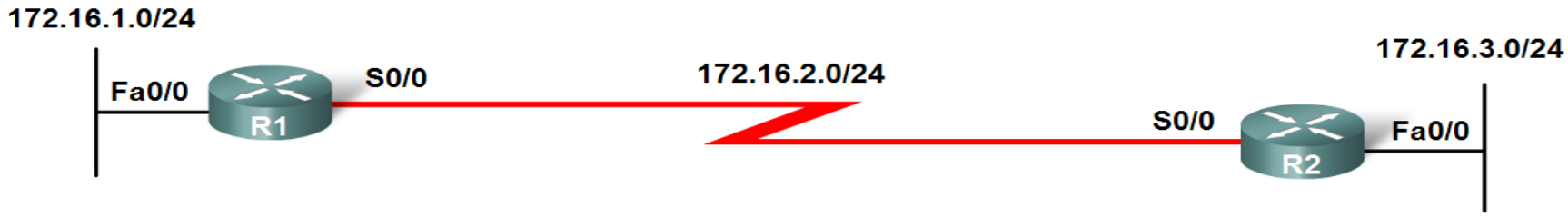
Dynamic Routing Protocols

The **purpose of a dynamic routing protocol** is to:

- –-Discover remote networks
- –-Maintaining up-to-date routing information
- –-Choosing the best path to destination networks
- –-Ability to find a new best path if the current path is no longer available

Routing Protocol Operation

Routing protocols are used to exchange routing information between the routers.



Dynamic Routing Protocols

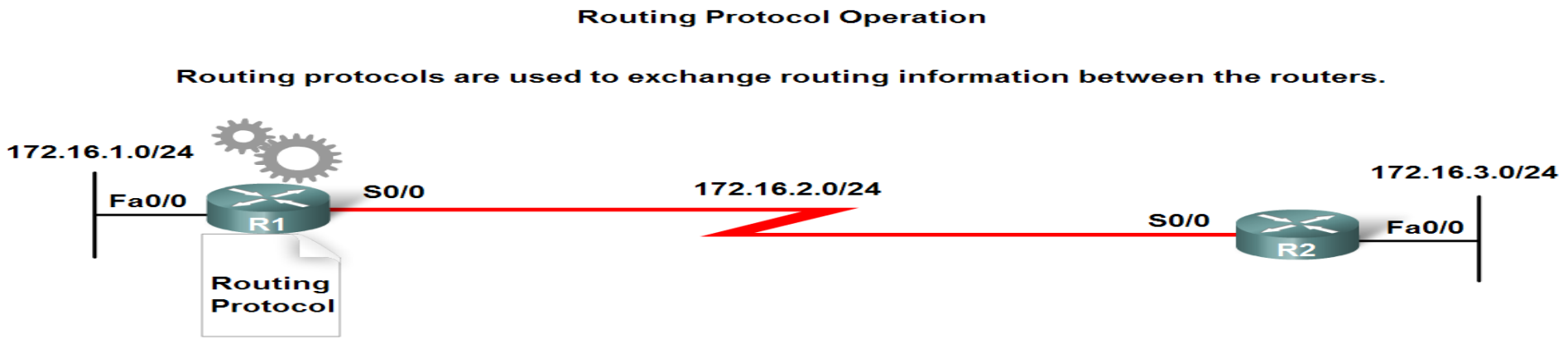
Components of a routing protocol

–Algorithm

- In the case of a routing protocol algorithms are used for facilitating routing information and best path determination

–Routing protocol messages

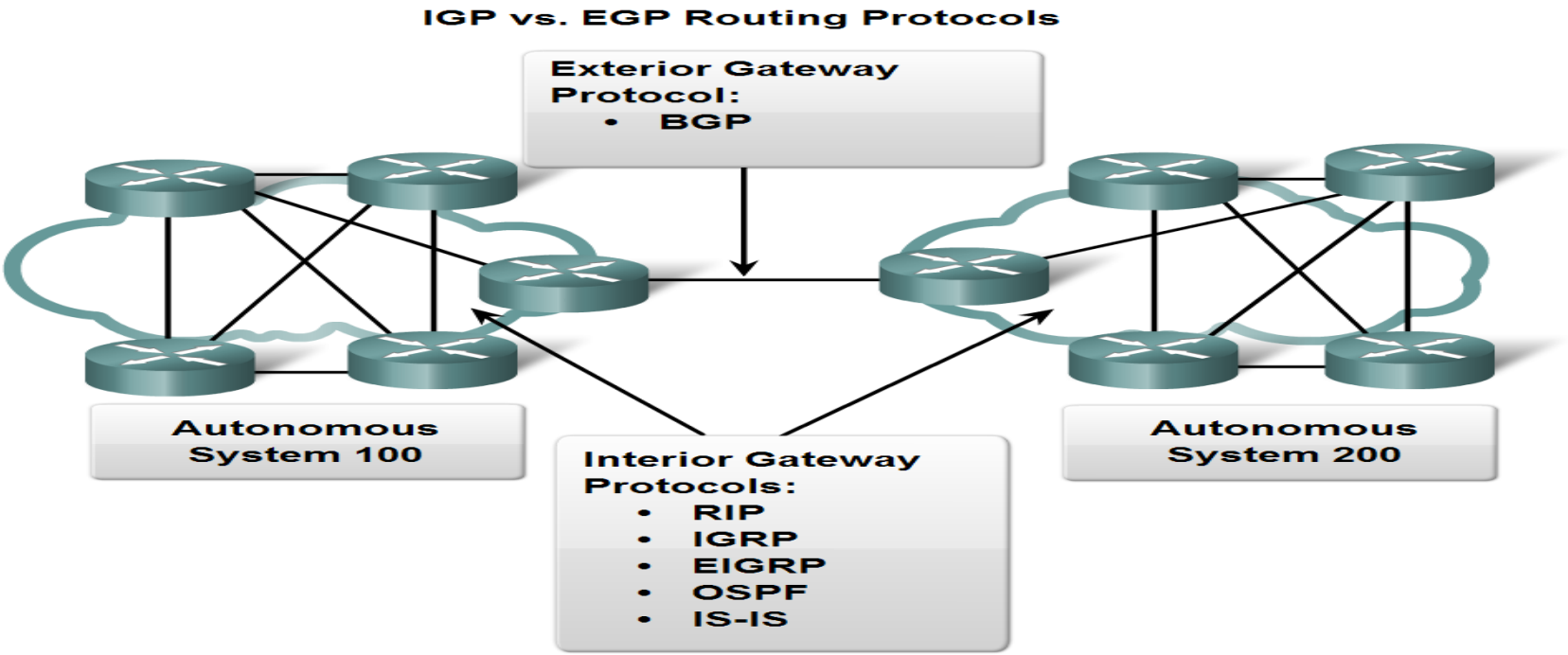
- These are messages for discovering neighbors and exchange of routing information



Classifying Routing Protocols

Types of routing protocols:

- –**Interior Gateway Protocols** (IGP)
- –**Exterior Gateway Protocols** (EGP)
- –

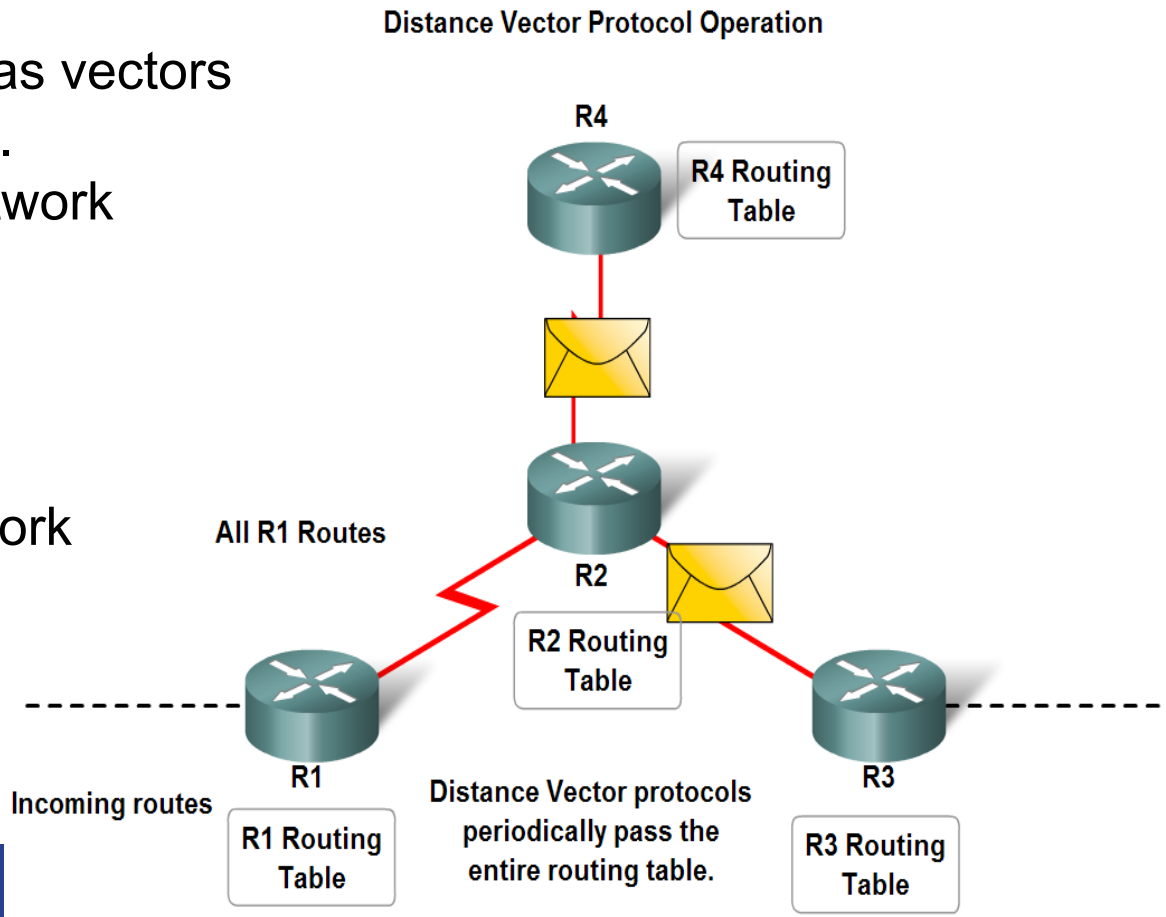


Classifying Routing Protocols

- **Interior Gateway Routing Protocols (IGP)**
  - –-Used for routing inside an autonomous system & used to route within the individual networks themselves.
  - –-Examples: RIP, EIGRP, OSPF
- **Exterior Routing Protocols (EGP)**
  - –-Used for routing between autonomous systems
  - –-Example: BGPv4

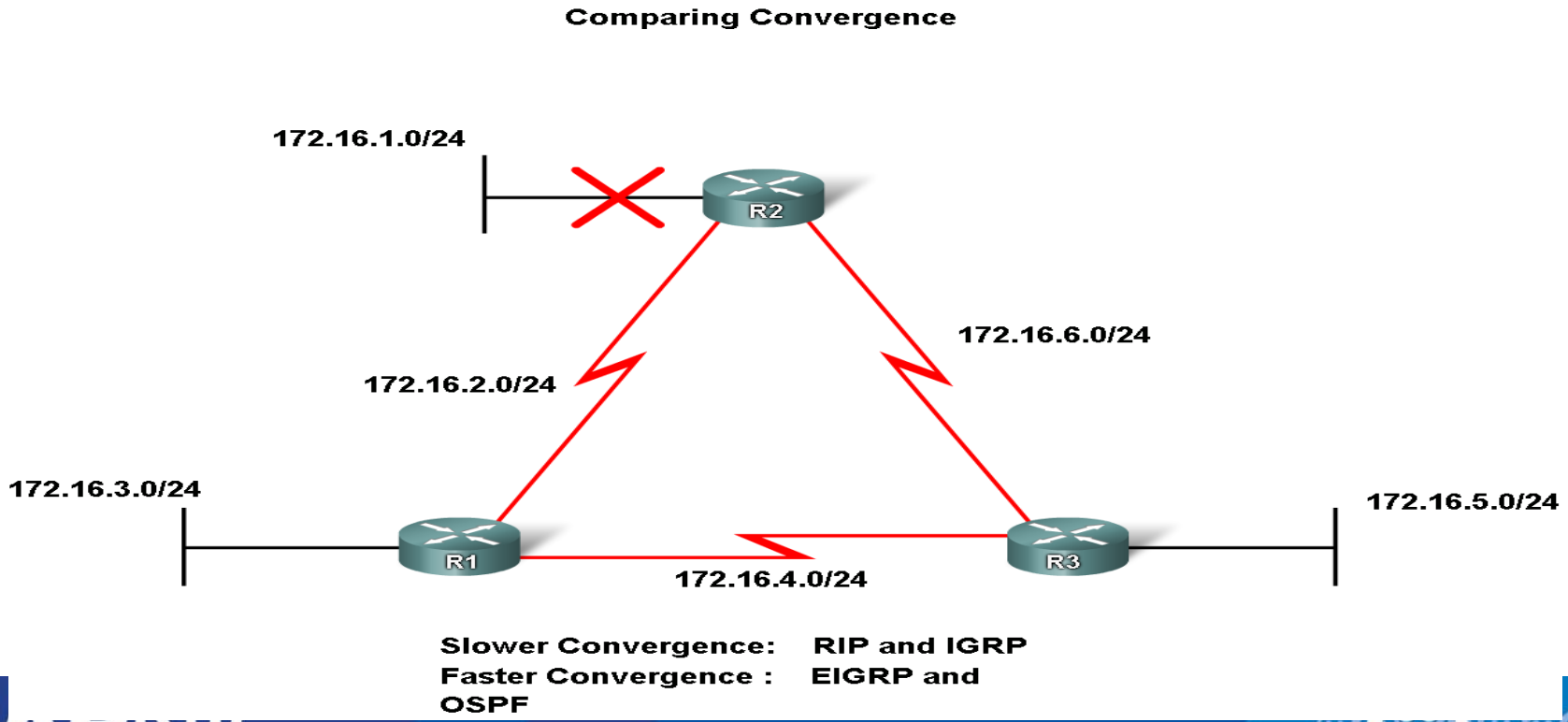
Classifying Routing Protocols

- IGP: **Comparison of Distance Vector** & **Link State** Routing Protocols
  - –**Distance vector**
  - – routes are advertised as vectors
  - – of distance & direction.
  - – incomplete view of network
  - – topology.
  - –Generally, periodic
  - – updates.
  - –**Link state**
  - – complete view of network
  - – topology is created.
  - – updates are not
  - – periodic.
  - – Faster Convergence



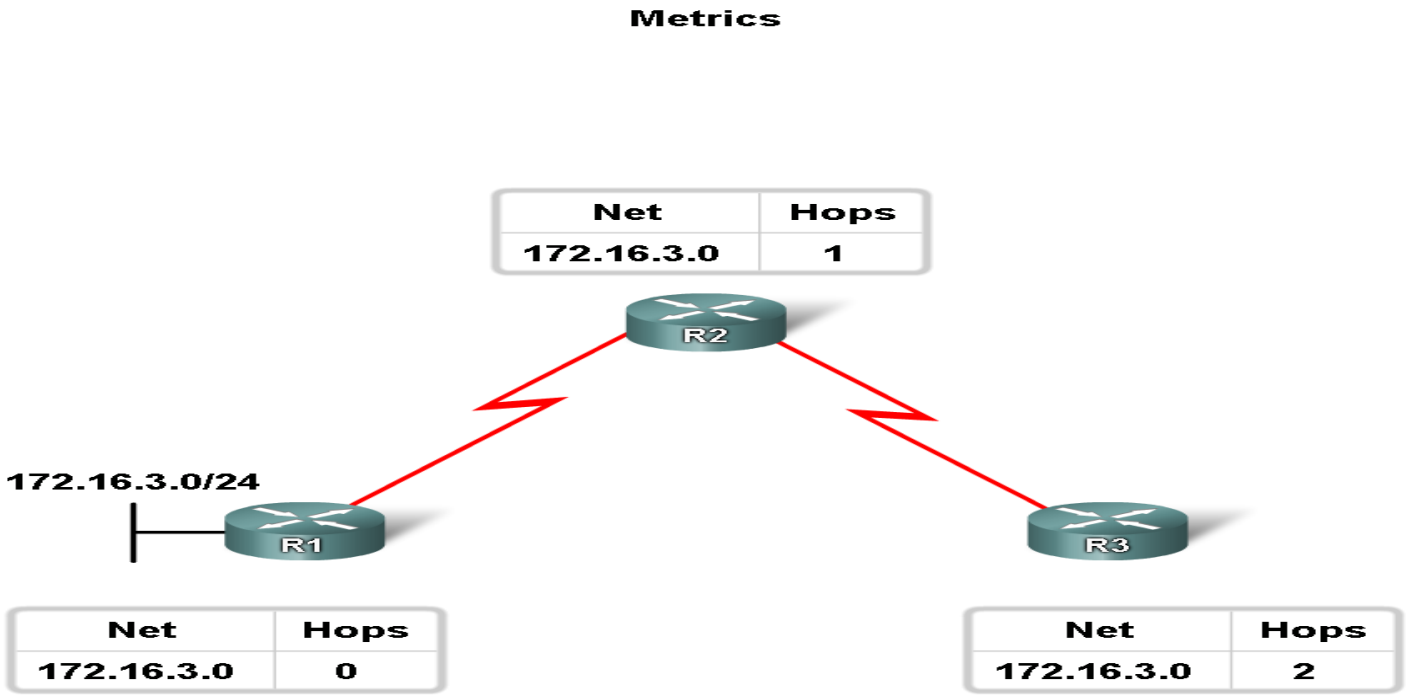
Classifying Routing Protocols

- **Convergence** is defined as: when all routers' routing tables are at a state of consistency



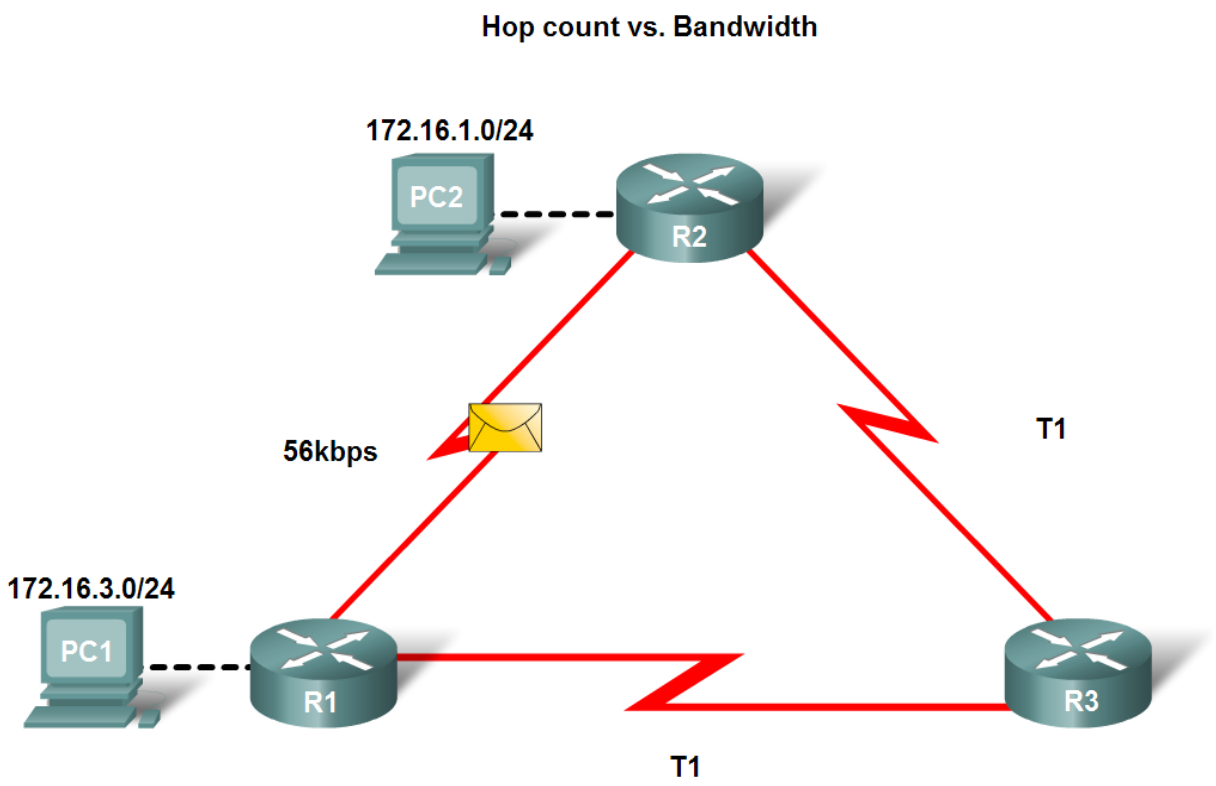
Routing Protocols Metrics

- **Metric**
  - –A value used by a routing protocol to determine which routes are better than others.



Routing Protocols Metrics

- **Metrics used in IP routing protocols**
  - –-Bandwidth
  - –-Cost
  - –-Delay
  - –-Hop count
  - –-Load
  - –-Reliability



Routing Protocols Metrics

- The Metric Field in the Routing Table
- **Metric** used for each routing protocol
  - –-RIP - hop count
  - –-IGRP & EIGRP - Bandwidth (used by default), Delay (used by default), Load, Reliability
  - –-IS-IS & OSPF – Cost, Bandwidth (Cisco's implementation)

–RIP-Routing Information Protocol

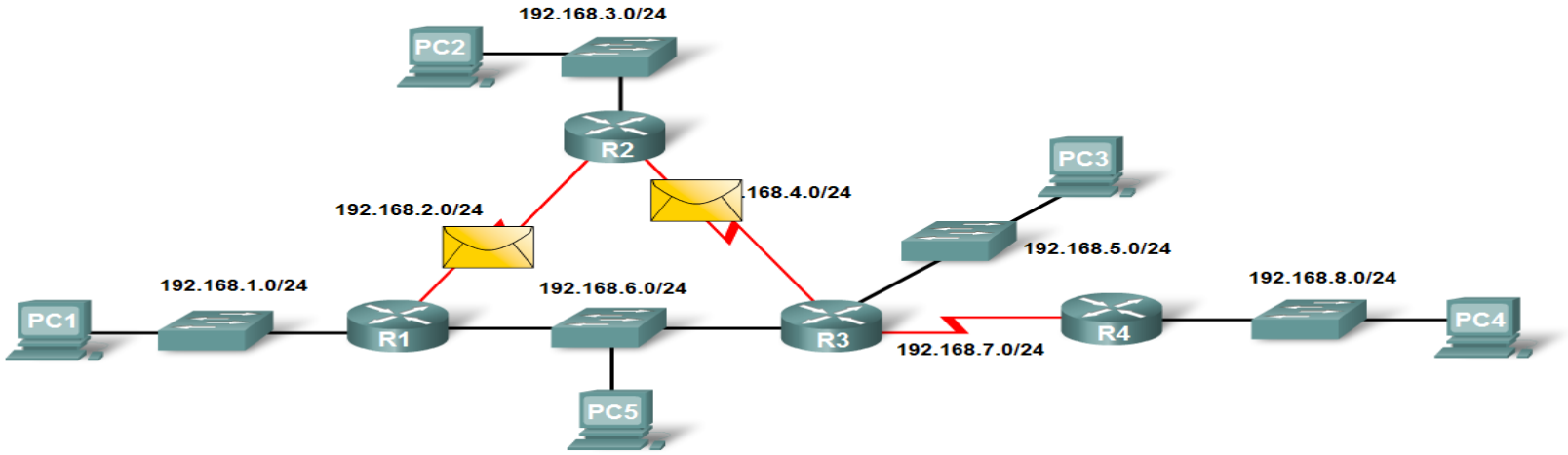
–IGRP-Interior Gateway Routing Protocol

–EIGRP-Enhanced Interior Gateway Routing Protocol

–IS-IS - Intermediate System to Intermediate System

Routing Protocols Metrics

- **Load balancing**
  - –This is the ability of a router to distribute packets among multiple same cost paths



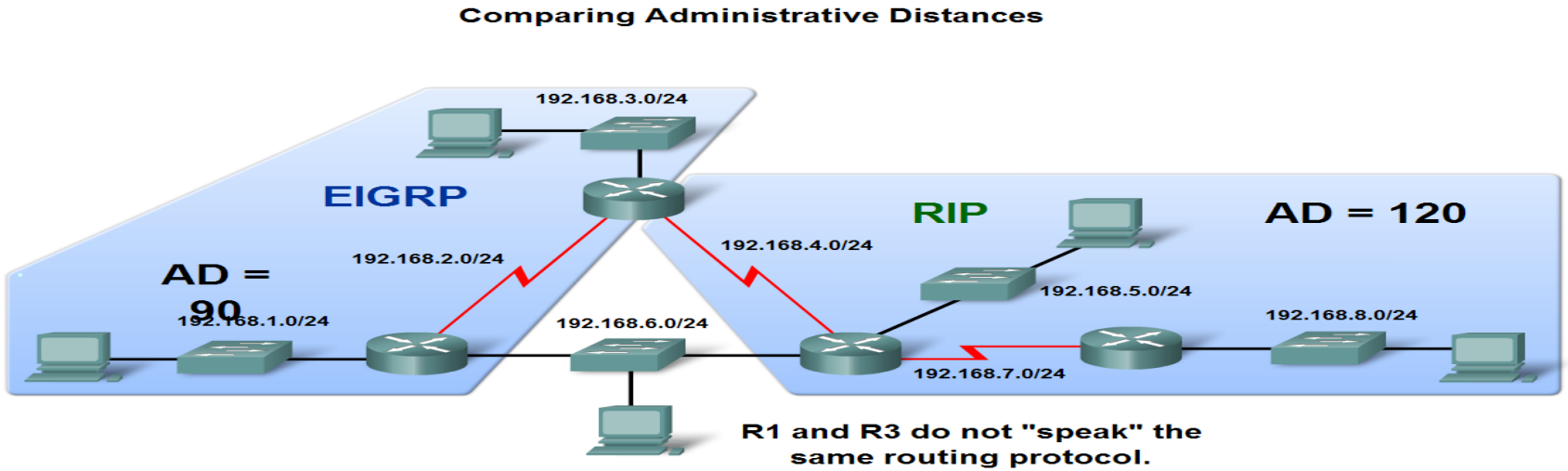

```
R2#show ip route
<output omitted>

R     192.168.6.0/24 [120/1] via 192.168.2.1, 00:00:24, Serial0/0/0
                     [120/1] via 192.168.4.1, 00:00:26, Serial0/0/1
```

Load Balancing Across Equal Cost Paths

Administrative Distance of a Route

- **Purpose of a metric**
  - –It's a calculated value used to determine the best path to a destination
- **Purpose of Administrative Distance**
  - –It's a numeric value that specifies the preference of a particular route



Administrative Distance of a Route

Identifying the Administrative Distance (AD) in a routing table

–It is the first number in the brackets in the routing table

```
R2#show ip route
<output omitted>

Gateway of last resort is not set

D     192.168.1.0/24 [90/2172416] via 192.168.2.1, 00:00:24, Serial0/0/0
C     192.168.2.0/24 is directly connected, Serial0/0/0
C     192.168.3.0/24 is directly connected, FastEthernet0/0
C     192.168.4.0/24 is directly connected, Serial0/0/1
R     192.168.5.0/24 [120/1] via 192.168.4.1, 00:00:08, Serial0/0/1
D     192.168.6.0/24 [90/2172416] via 192.168.2.1, 00:00:24, Serial0/0/0
R     192.168.7.0/24 [120/1] via 192.168.4.1, 00:00:08, Serial0/0/1
R     192.168.8.0/24 [120/2] via 192.168.4.1, 00:00:08, Serial0/0/1
```

```
R2#show ip rip database
192.168.3.0/24    directly connected, FastEthernet0/0
192.168.4.0/24    directly connected, Serial0/0/1
192.168.5.0/24
    [1] via 192.168.4.1, Serial0/0/1
192.168.6.0/24
    [1] via 192.168.4.1, Serial0/0/1
192.168.7.0/24
    [1] via 192.168.4.1, Serial0/0/1
192.168.8.0/24
    [2] via 192.168.4.1, Serial0/0/1
```

Administrative Distance of a Route

Dynamic Routing Protocols

Default Administrative Distances

Route source	Default AD
Connected interface	0
Static	1
EIGRP summary route	5
eBGP	20
EIGRP (Internal)	90
IGRP	100
OSPF	110
IS - IS	115
RIP	120
EIGRP (External)	170
iBGP	200
Unknown	255

Administrative Distance of a Route

- **Directly connected routes**
  - –Have a default **AD of 0**
- **Static Routes**
  - –Administrative distance of a static route has a **default value of 1**

```
R2#show ip route 172.16.3.0
Routing entry for 172.16.3.0/24
Known via "static", distance 1, metric 0 (connected)
  Routing Descriptor Blocks:
  * directly connected, via Serial0/0/0
      Route metric is 0, traffic share count is 1
```

Administrative Distance of a Route

Directly connected routes

–-Immediately appear in the routing table as soon as the interface is configured

```
R2#show ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter area
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route

Gateway of last resort is not set

     172.16.0.0/24 is subnetted, 3 subnets
C       172.16.1.0 is directly connected, FastEthernet0/0
C       172.16.2.0 is directly connected, Serial0/0/0
S       172.16.3.0 is directly connected, Serial0/0/0
C    192.168.1.0/24 is directly connected, Serial0/0/1
S    192.168.2.0/24 [1/0] via 192.168.1.1
```

Routing Basics

ISP Workshops

Routing Concepts

- IPv6
- IPv4
- Routing
- Forwarding
- Some definitions
- Policy options
- Routing Protocols

IPv6

- Internet is starting to use IPv6
  - Addresses are 128 bits long
  - Internet addresses range from 2000::/16 to 3FFF::/16
  - The remaining IPv6 range is reserved or has “special” uses
- IPv6 address has a network portion and a host portion

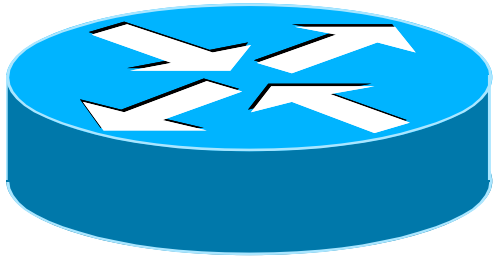
IPv4

- Internet still uses IPv4
  - (legacy protocol)
  - Addresses are 32 bits long
  - Range from 1.0.0.0 to 223.255.255.255
  - 0.0.0.0 to 0.255.255.255 and 224.0.0.0 to 255.255.255.255 have “special” uses
- IPv4 address has a network portion and a host portion

IP address format

- Address and subnet mask
  - IPv4 written as
    - 12.34.56.78 255.255.255.0 *or*
    - 12.34.56.78/24
  - IPv6 written as
    - 2001:db8::1/128
  - mask represents the number of network bits in the address
  - The remaining bits are the host bits

What does a router do?



A day in a life of a router

find path

forward packet, forward packet, forward packet, forward packet…

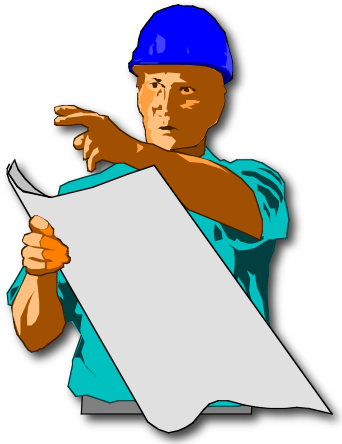
find alternate path

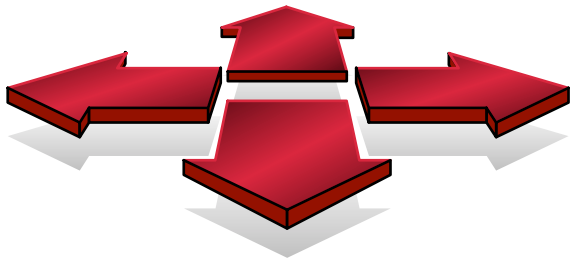
forward packet, forward packet, forward packet, forward packet…

repeat until powered off

Routing versus Forwarding

- Routing = building maps and giving directions
- Forwarding = moving packets between interfaces according to the “directions”





IP Routing – finding the path

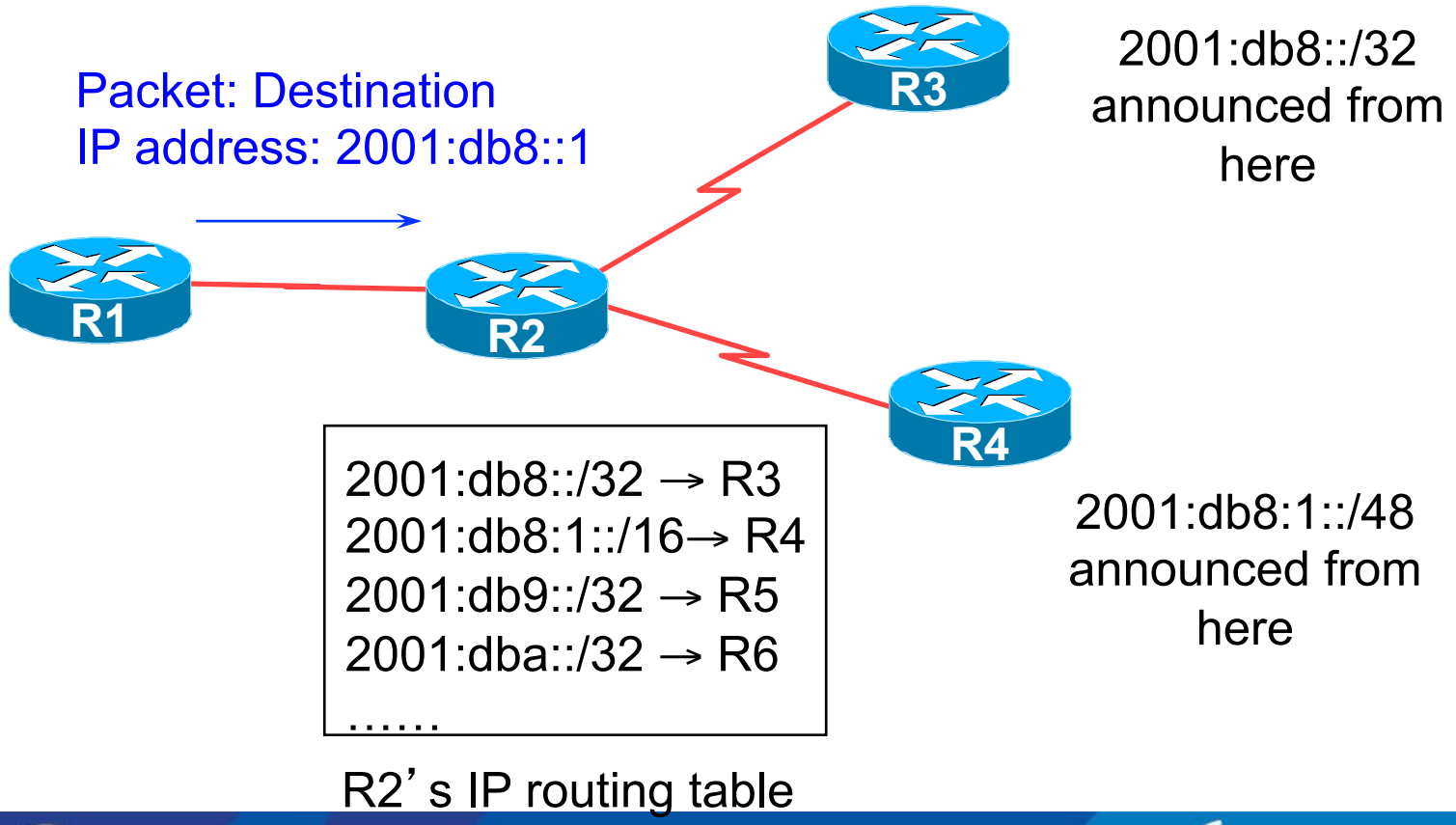
- Path derived from information received from a routing protocol
- Several alternative paths may exist
  - best path stored in **forwarding** table
- Decisions are updated periodically or as topology changes (event driven)
- Decisions are based on:
  - topology, policies and metrics (hop count, filtering, delay, bandwidth, etc.)

IP route lookup

- Based on destination IP address
- “longest match” routing
  - More specific prefix preferred over less specific prefix
  - Example: packet with destination of 2001:db8::1/128 is sent to the router announcing 2001:db8:1::/48 rather than the router announcing 2001:db8::/32.

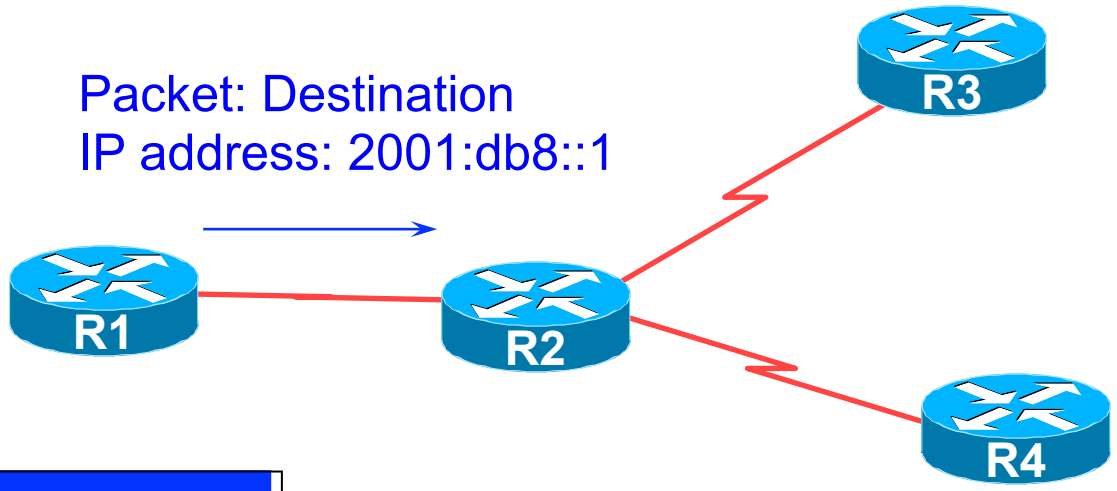
IP route lookup

- Based on destination IP address



IP route lookup: Longest match routing

- Based on destination IP address



2001:db8::/32 → R3
2001:db8:1::/48 → R4
2001:db9::/32 → R5
2001:dba::/32 → R6
……

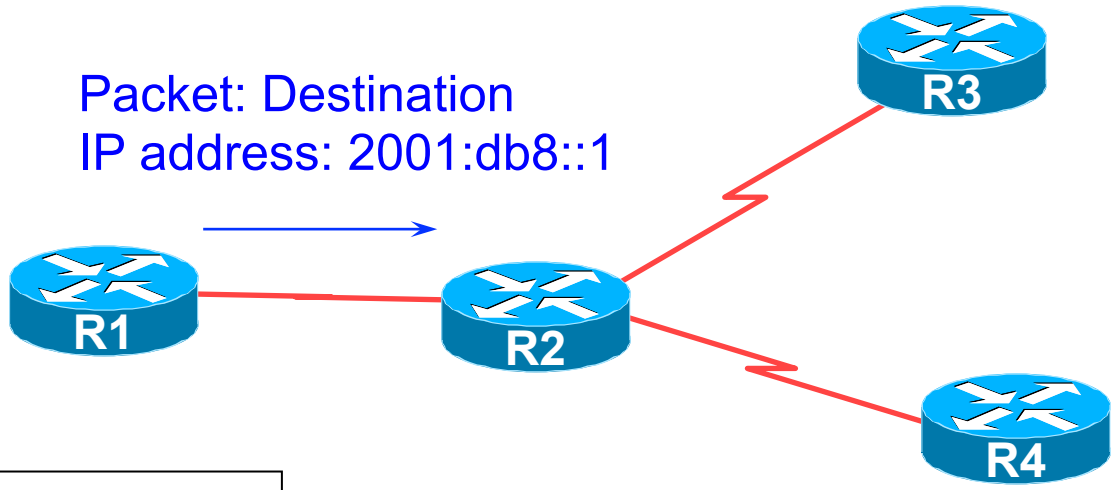
R2's IP routing table

2001:db8::1 && ffff:ffff::
vs.
2001:db8:: && ffff:ffff::

Match!

IP route lookup: Longest match routing

- Based on destination IP address



2001:db8::/32 → R3
2001:db8:1::/48 → R4
2001:db9::/32 → R5
2001:dba::/32 → R6
……

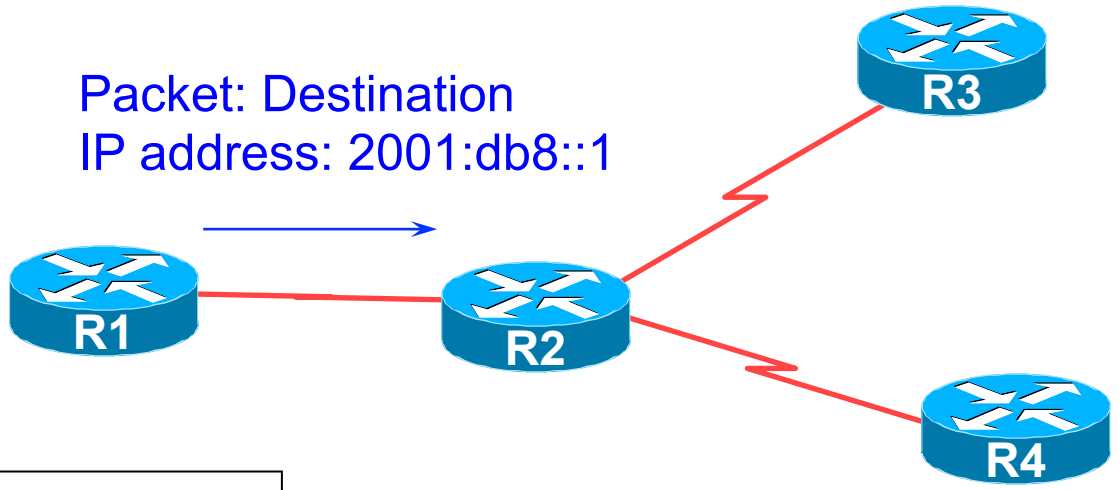
R2's IP routing table

2001:db8::1 && ffff:ffff:ffff::
vs.
2001:db8:1:: && ffff:ffff:ffff::

Match as well!

IP route lookup: Longest match routing

- Based on destination IP address



2001:db8::/32 → R3
2001:db8:1::/48 → R4
2001:db9::/32 → R5
2001:dba::/32 → R6
……

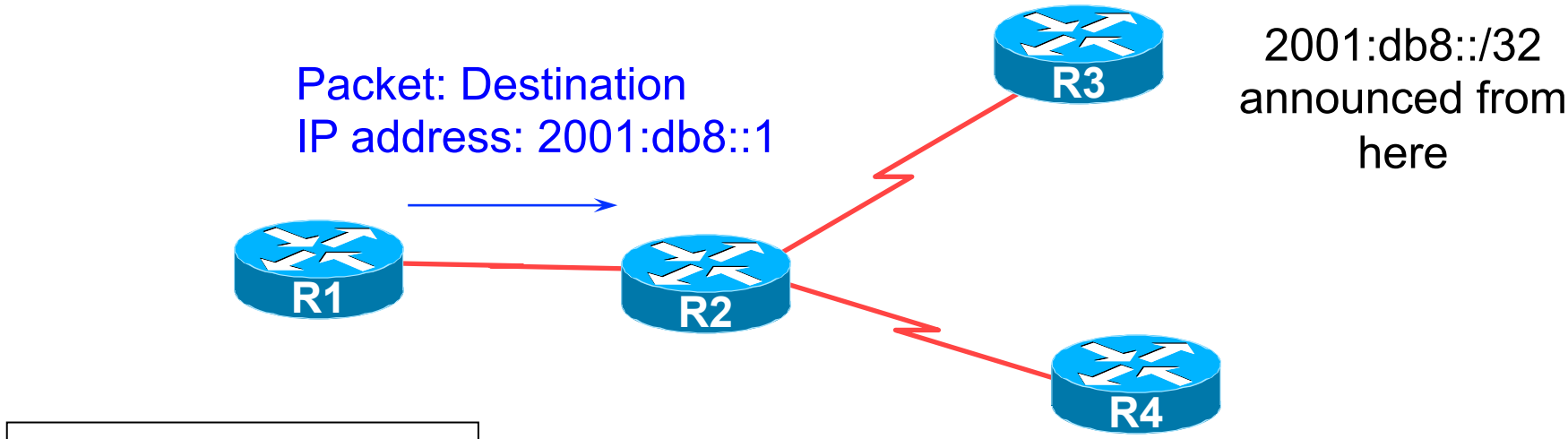
R2's IP routing table

2001:db8::1 && ffff:ffff::
vs.
2001:db9:: && ffff:ffff::

Does not match!

IP route lookup: Longest match routing

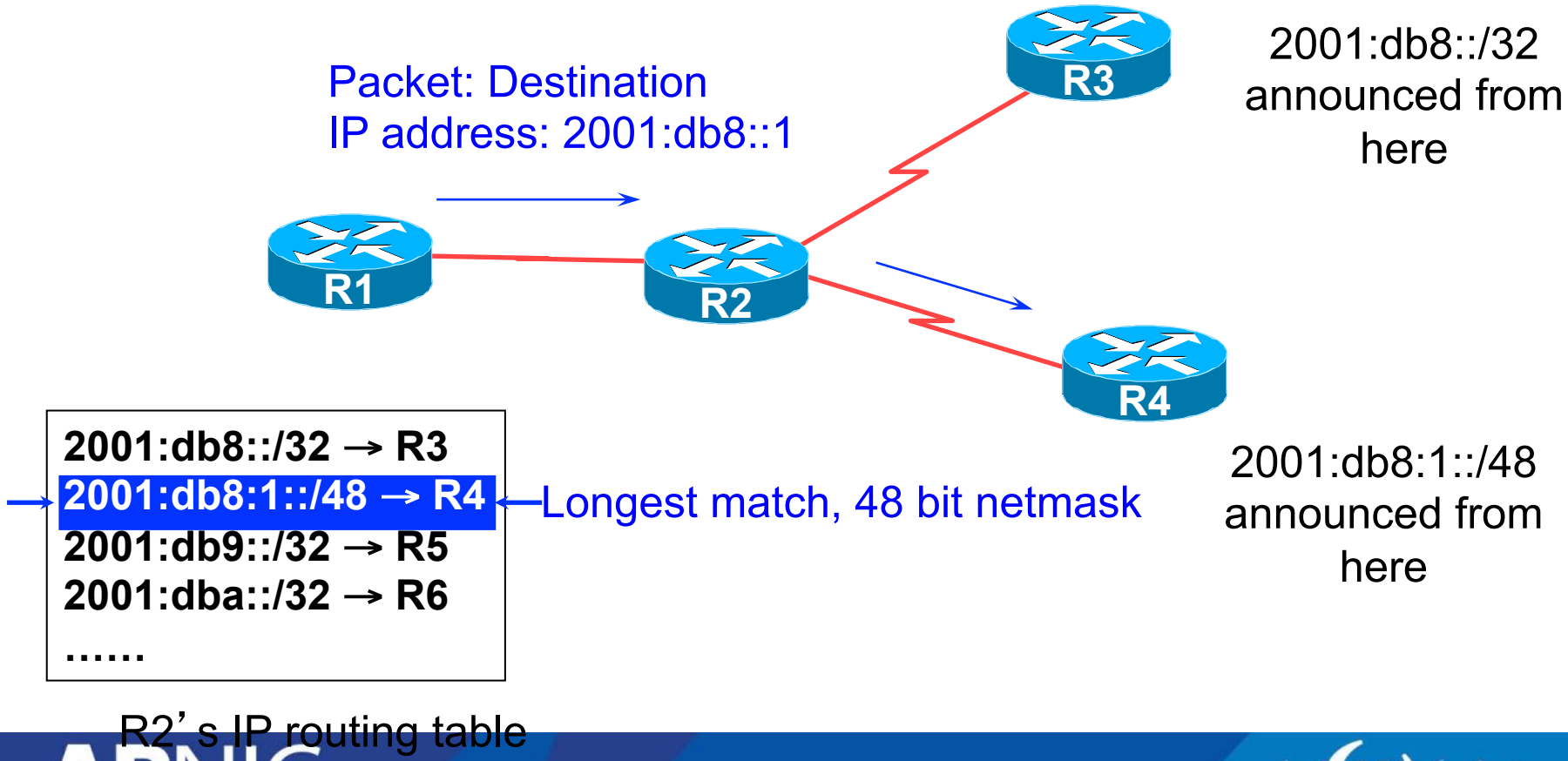
- Based on destination IP address



R2's IP routing table

IP route lookup: Longest match routing

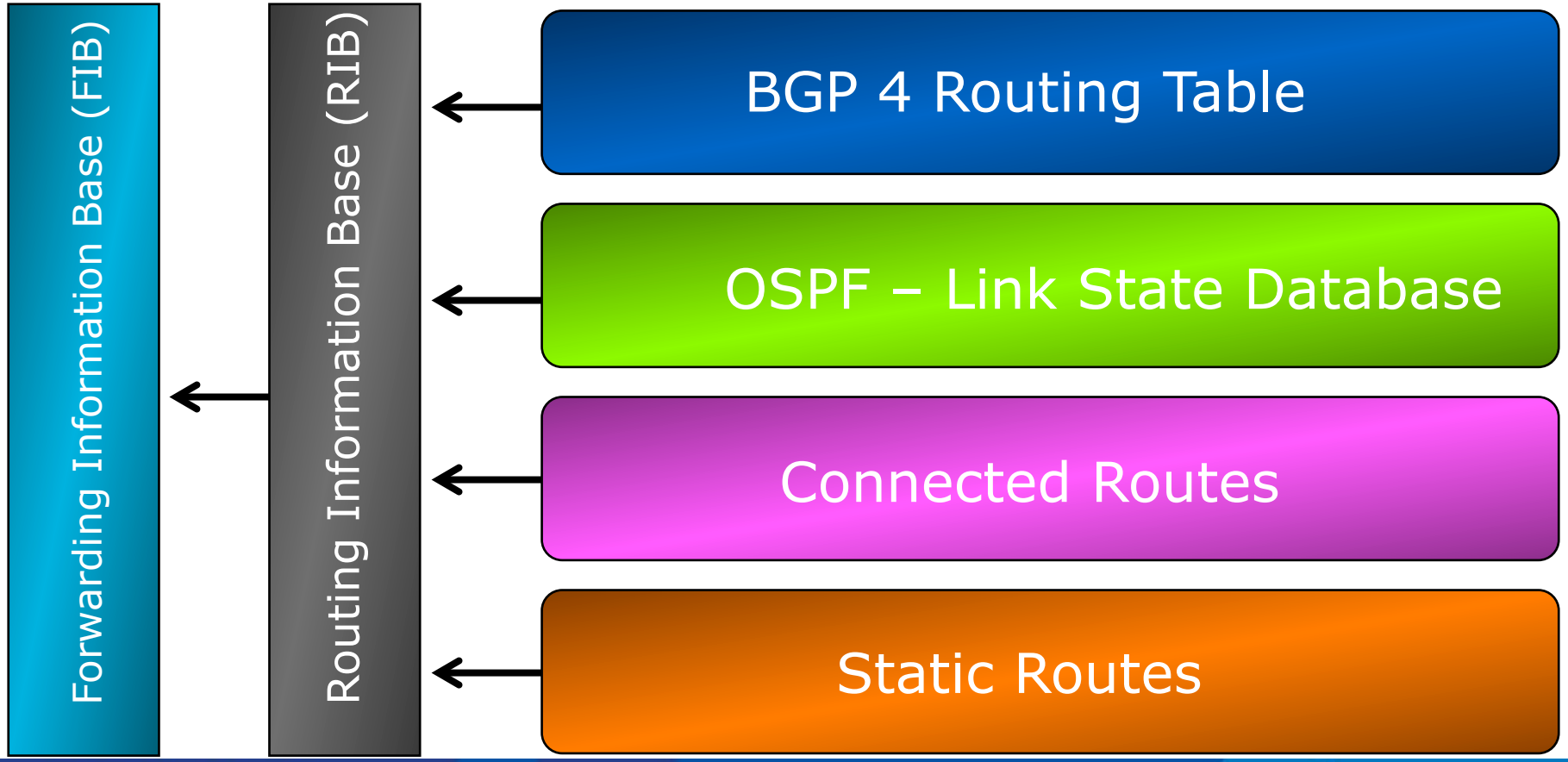
- Based on destination IP address



IP Forwarding

- Router decides which interface a packet is sent to
- Forwarding table populated by routing process
- Forwarding decisions:
  - destination address
  - class of service (fair queuing, precedence, others)
  - local requirements (packet filtering)
- Forwarding is usually aided by special hardware

Routing Tables Feed the Forwarding Table



RIBs and FIBs

- FIB is the Forwarding Table
  - It contains destinations and the interfaces to get to those destinations
  - Used by the router to figure out where to send the packet
  - Careful! Some people still call this a route!
- RIB is the Routing Table
  - It contains a list of all the destinations and the various next hops used to get to those destinations – and lots of other information too!
  - One destination can have lots of possible next-hops – only the best next-hop goes into the FIB

Explicit versus Default Routing

- Default:
  - simple, cheap (cycles, memory, bandwidth)
  - low granularity (metric games)
- Explicit (default free zone)
  - high overhead, complex, high cost, high granularity
- Hybrid
  - minimise overhead
  - provide useful granularity
  - requires some filtering knowledge

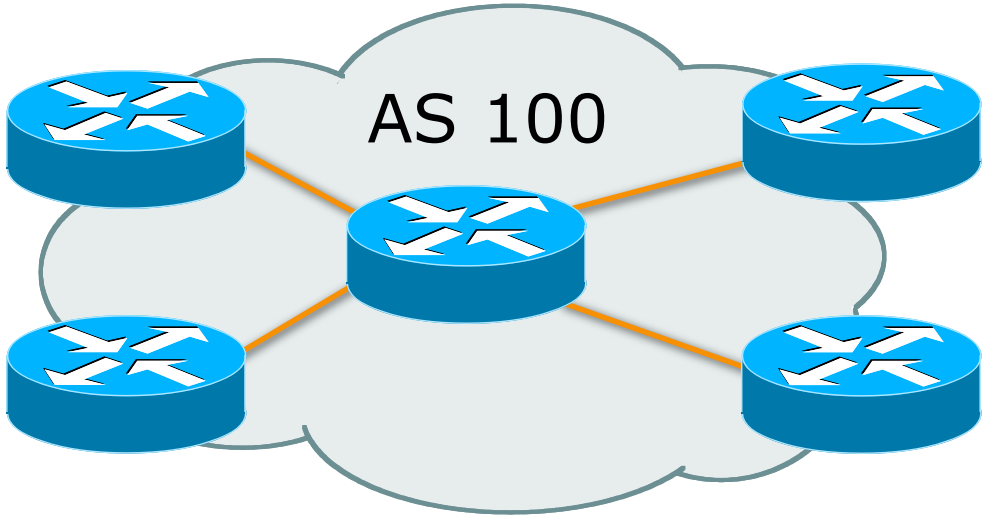
Egress Traffic

- How packets leave your network
- Egress traffic depends on:
  - route availability (what others send you)
  - route acceptance (what you accept from others)
  - policy and tuning (what you do with routes from others)
  - Peering and transit agreements

Ingress Traffic

- How packets get to your network and your customers' networks
- Ingress traffic depends on:
  - what information you send and to whom
  - based on your addressing and AS's
  - based on others' policy (what they accept from you and what they do with it)

Autonomous System (AS)

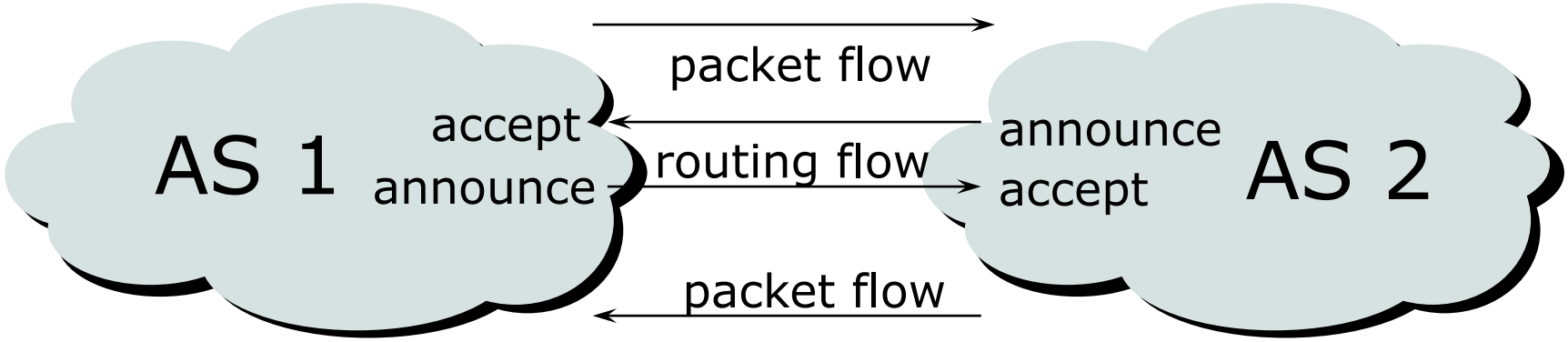


- Collection of networks with same routing policy
- Single routing protocol
- Usually under single ownership, trust and administrative control

Definition of terms

- Neighbours
  - AS's which directly exchange routing information
  - Routers which exchange routing information
- Announce
  - send routing information to a neighbour
- Accept
  - receive and use routing information sent by a neighbour
- Originate
  - insert routing information into external announcements (usually as a result of the IGP)
- Peers
  - routers in neighbouring AS's or within one AS which exchange routing and policy information

Routing flow and packet flow



For networks in AS1 and AS2 to communicate:

AS1 must announce to AS2

AS2 must accept from AS1

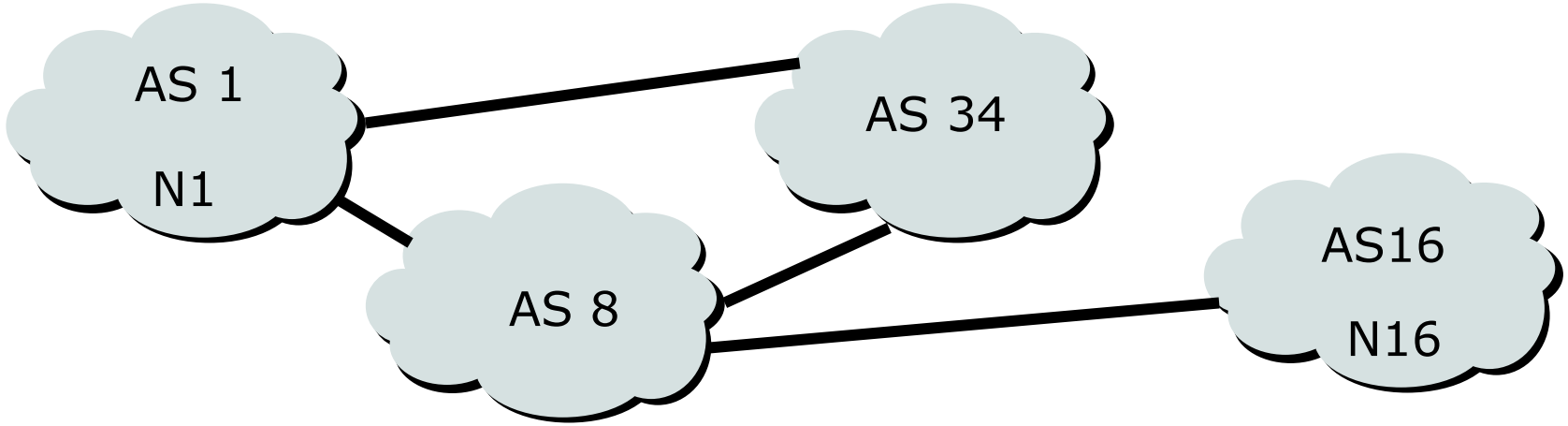
AS2 must announce to AS1

AS1 must accept from AS2

Routing flow and Traffic flow

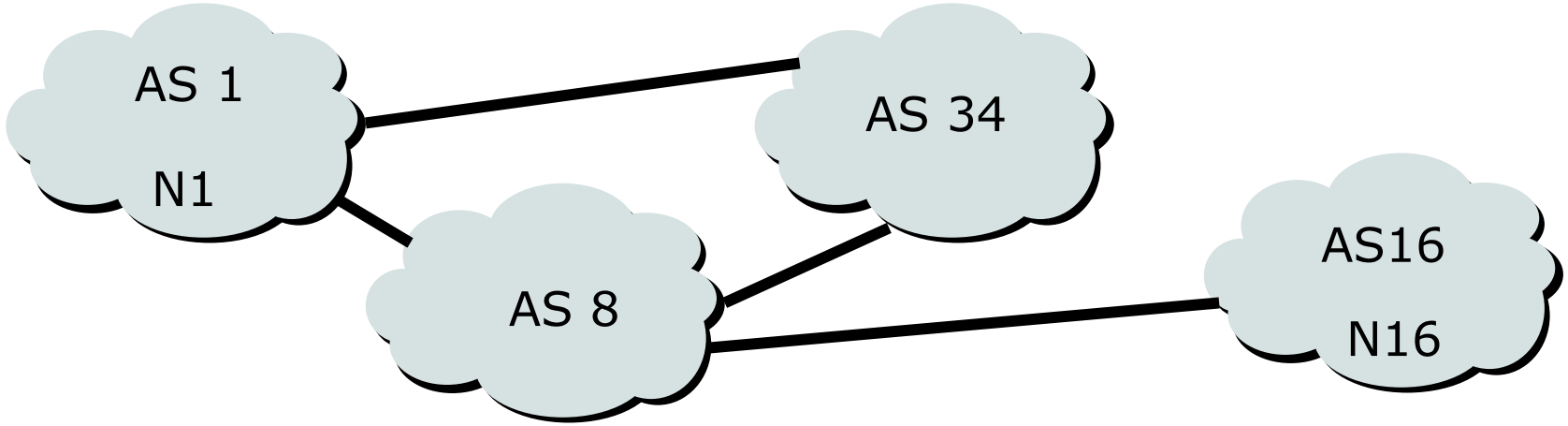
- Traffic flow is always in the opposite direction of the flow of Routing information
  - Filtering outgoing routing information inhibits traffic flow inbound
  - Filtering inbound routing information inhibits traffic flow outbound

Routing Flow/Packet Flow: With multiple ASes



- For net N1 in AS1 to send traffic to net N16 in AS16:
  - AS16 must originate and announce N16 to AS8.
  - AS8 must accept N16 from AS16.
  - AS8 must announce N16 to AS1 or AS34.
  - AS1 must accept N16 from AS8 or AS34.
- For two-way packet flow, similar policies must exist for N1

Routing Flow/Packet Flow: With multiple ASes

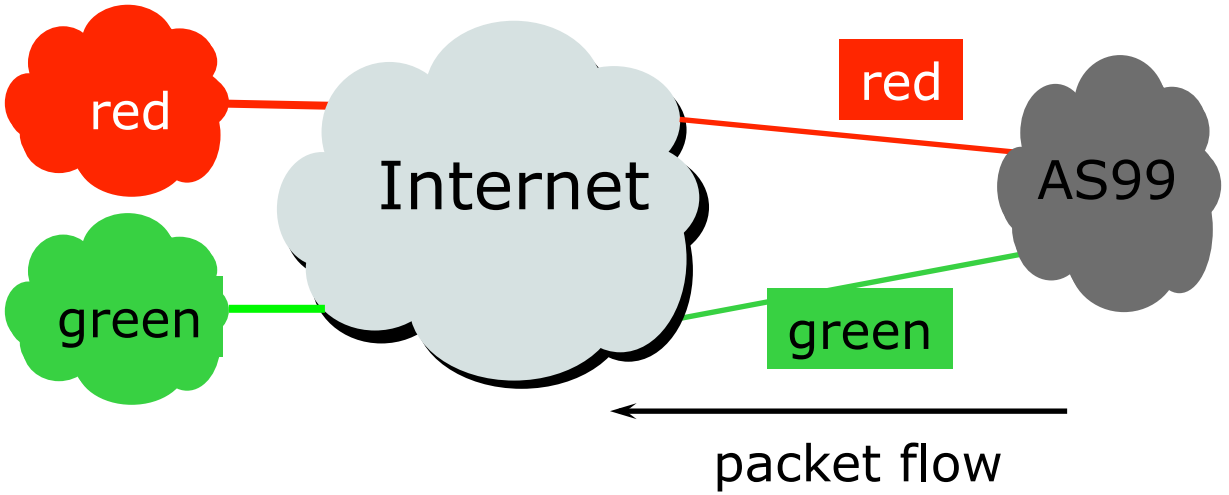


- As multiple paths between sites are implemented it is easy to see how policies can become quite complex.

Routing Policy

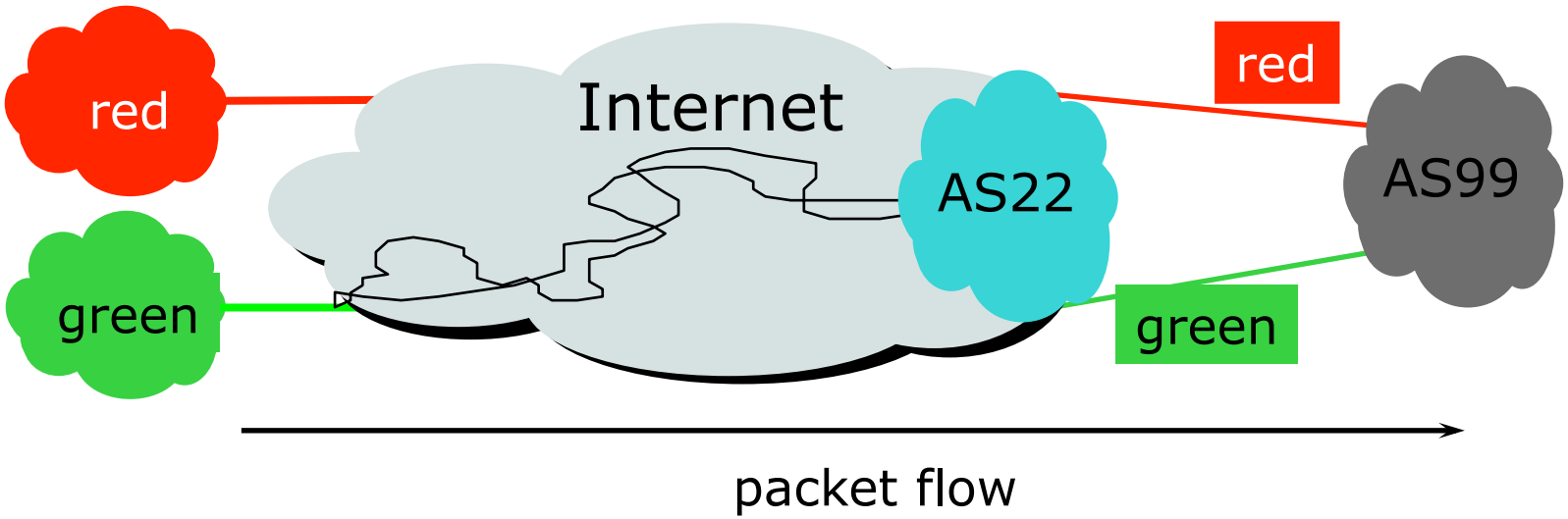
- Used to control traffic flow in and out of an ISP network
- ISP makes decisions on what routing information to accept and discard from its neighbours
  - Individual routes
  - Routes originated by specific ASes
  - Routes traversing specific ASes
  - Routes belonging to other groupings
    - Groupings which you define as you see fit

Routing Policy Limitations



- AS99 uses red link for traffic to the red AS and the green link for remaining traffic
- To implement this policy, AS99 has to:
  - Accept routes originating from the red AS on the red link
  - Accept all other routes on the green link

Routing Policy Limitations



- AS99 would like packets coming from the green AS to use the green link.
- But unless AS22 cooperates in pushing traffic from the green AS down the green link, there is very little that AS99 can do to achieve this aim

Routing Policy Issues

- April 2013:
  - 12900 IPv6 prefixes & 460000 IPv4 prefixes
    - Not realistic to set policy on all of them individually
  - 44500 origin AS's
    - Too many to try and create individual policies for
- Routes tied to a specific AS or path may be unstable regardless of connectivity
- Solution: Groups of AS's are a natural abstraction for filtering purposes

Routing Protocols

We now know what routing means…

…but what do the routers get up to?

And why are we doing this anyway?

1: How Does Routing Work?

- Internet is made up of the ISPs who connect to each other's networks
- How does an ISP in Kenya tell an ISP in Japan what customers they have?
- And how does that ISP send data packets to the customers of the ISP in Japan, and get responses back
  - After all, as on a local ethernet, two way packet flow is needed for communication between two devices

2: How Does Routing Work?

- ISP in Kenya could buy a direct connection to the ISP in Japan
  - But this doesn't scale – thousands of ISPs, would need thousands of connections, and cost would be astronomical
- Instead, ISP in Kenya tells his neighbouring ISPs what customers he has
  - And the neighbouring ISPs pass this information on to their neighbours, and so on
  - This process repeats until the information reaches the ISP in Japan

3: How Does Routing Work?

- This process is called “Routing”
- The mechanisms used are called “Routing Protocols”
- Routing and Routing Protocols ensures that the Internet can scale, that thousands of ISPs can provide connectivity to each other, giving us the Internet we see today

4: How Does Routing Work?

- ISP in Kenya doesn't actually tell his neighbouring ISPs the names of the customers
  - (network equipment does not understand names)
- Instead, he has received an IP address block as a member of the Regional Internet Registry serving Kenya
  - His customers have received address space from this address block as part of their "Internet service"
  - And he announces this address block to his neighbouring ISPs – this is called announcing a "route"

Routing Protocols

- Routers use “routing protocols” to exchange routing information with each other
  - IGP is used to refer to the process running on routers inside an ISP’s network
  - EGP is used to refer to the process running between routers bordering directly connected ISP networks

What Is an IGP?

- Interior Gateway Protocol
- Within an Autonomous System
- Carries information about internal infrastructure prefixes
- Two widely used IGPs:
  - OSPF
  - ISIS

Why Do We Need an IGP?

- ISP backbone scaling
  - Hierarchy
  - Limiting scope of failure
  - Only used for ISP's infrastructure addresses, not customers or anything else
  - Design goal is to minimise number of prefixes in IGP to aid scalability and rapid convergence

What Is an EGP?

- Exterior Gateway Protocol
- Used to convey routing information between Autonomous Systems
- De-coupled from the IGP
- Current EGP is BGP

Why Do We Need an EGP?

- Scaling to large network
  - Hierarchy
  - Limit scope of failure
- Define Administrative Boundary
- Policy
  - Control reachability of prefixes
  - Merge separate organisations
  - Connect multiple IGPs

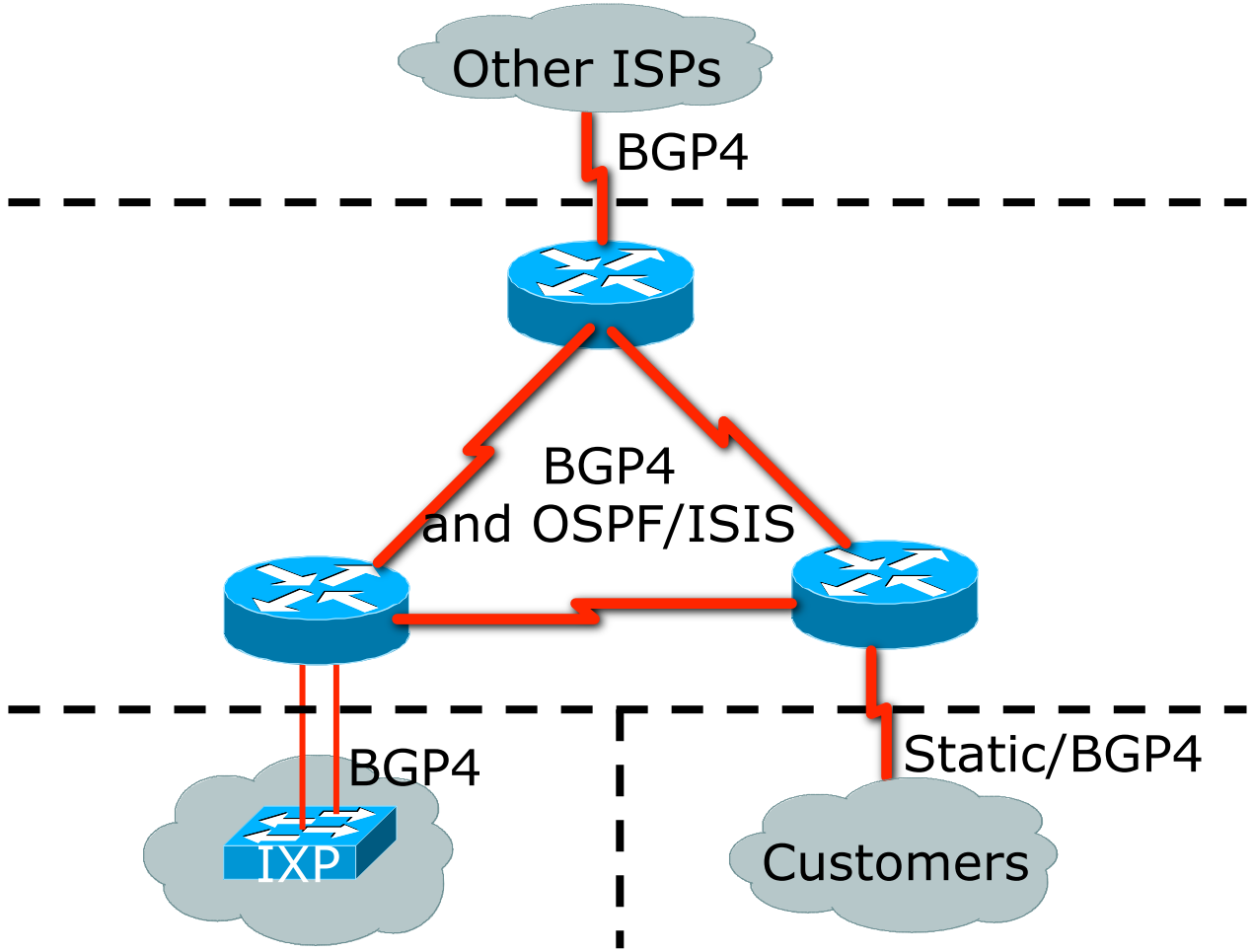
Interior versus Exterior Routing Protocols

- Interior
  - automatic neighbour discovery
  - generally trust your IGP routers
  - prefixes go to all IGP routers
  - binds routers in one AS together
- Exterior
  - specifically configured peers
  - connecting with outside networks
  - set administrative boundaries
  - binds AS's together

Interior versus Exterior Routing Protocols

- Interior
  - Carries ISP infrastructure addresses only
  - ISPs aim to keep the IGP small for efficiency and scalability
- Exterior
  - Carries customer prefixes
  - Carries Internet prefixes
  - EGPs are independent of ISP network topology

Hierarchy of Routing Protocols



FYI: Cisco IOS Default Administrative Distances

Route Source	Default Distance
Connected Interface	0
Static Route	1
Enhanced IGRP Summary Route	5
External BGP	20
Internal Enhanced IGRP	90
IGRP	100
OSPF	110
IS-IS	115
RIP	120
EGP	140
External Enhanced IGRP	170
Internal BGP	200
Unknown	255

Routing Basics

ISP Workshops

Distance Vector routing protocols

APNIC

Distance Vector Routing Protocols

Examples of Distance Vector routing protocols:

- Routing Information Protocol (**RIP**)
- Interior Gateway Routing Protocol (**IGRP**)
- Enhanced Interior Gateway Routing Protocol (**EIGRP**) -- hybrid

Distance Vector Routing Protocols

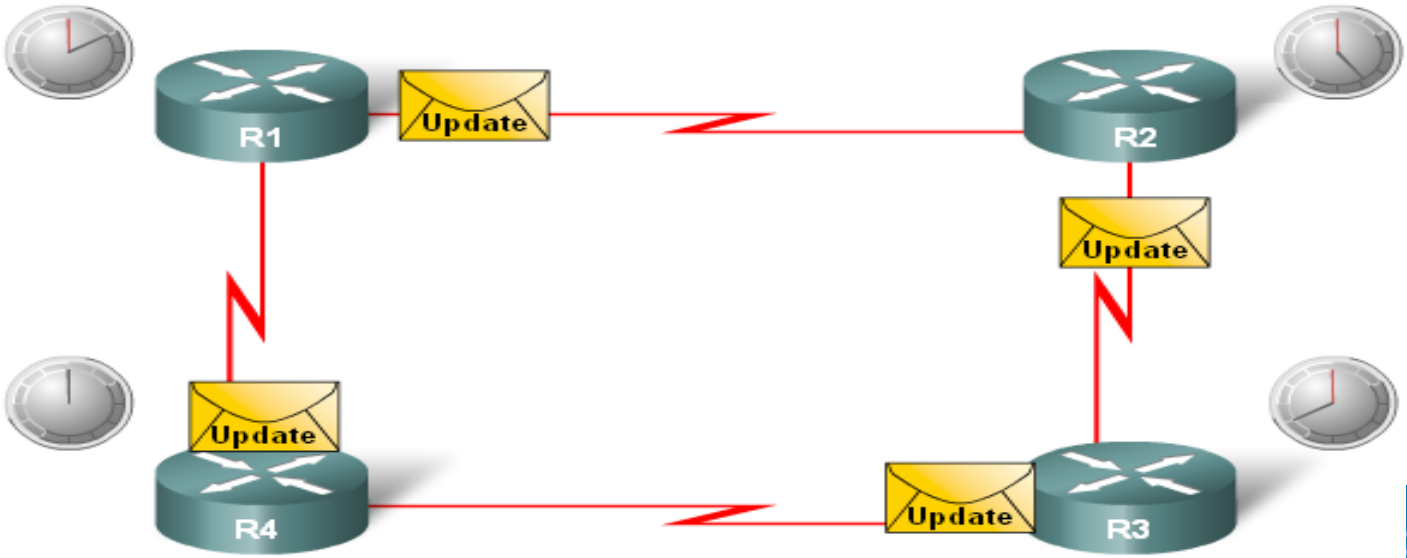
Distance Vector Technology

- **The Meaning of Distance Vector**:
  - A router using distance vector routing protocols knows 2 things:
    - **Distance** to final destination
    - **Vector, or direction**, traffic should be directed

Distance Vector Routing Protocols

Characteristics of Distance Vector routing protocols:

- Periodic updates
- Neighbors
- Broadcast updates
- Entire routing table is included with routing update



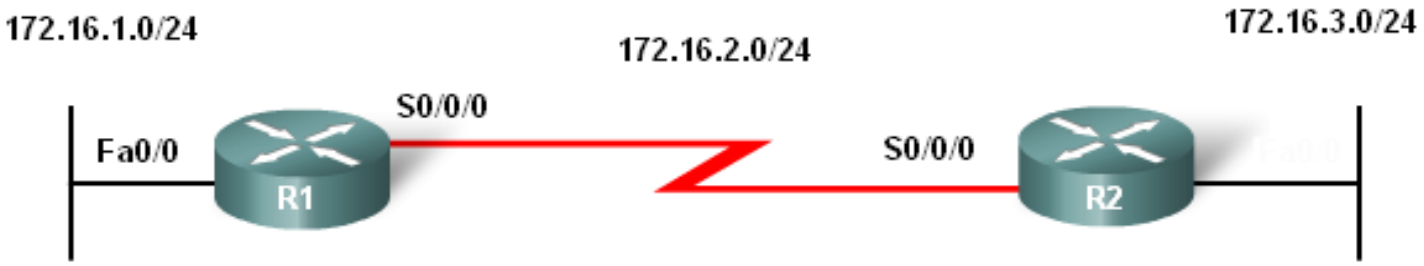
Distance Vector Routing Protocols

Routing Protocol Algorithm:

-Defines a procedure for accomplishing a certain task

Purpose of Routing Algorithms

1. Send and Receive Updates
2. Calculate best path; install routes
3. Detect and react to topology changes



Network	Interface	Hope
172.16.1.0/24	Fa0/0	0
172.16.2.0/24	S0/0/0	0
172.16.3.0/24	S0/0/0	1

Network	Interface	Hope
172.16.3.0/24	Fa0/0	0
172.16.2.0/24	S0/0/0	0
172.16.1.0/24	S0/0/0	1

Distance Vector Routing Protocols

Routing Protocol Characteristics

–Criteria used to compare routing protocols includes

- -Time to convergence
- -Scalability
- -Resource usage
- -Implementation & maintenance

Distance Vector Routing Protocols

Advantages & Disadvantages of Distance Vector Routing Protocols

Advantages:	Disadvantages:
Simple implementation and maintenance. The level of knowledge required to deploy and later maintain a network with distance vector protocol is not high.	**Slow convergence.** The use of periodic updates can cause slower convergence. Even if some advanced techniques are used, like triggered updates which are discussed later, the overall convergence is still slower compared to link state routing protocols.
Low resource requirements. Distance vector protocols typically do not need large amounts of memory to store the information. Nor do they require a powerful CPU. Depending of the network size and the IP addressing implemented they also typically do not require a high level of link bandwidth to send routing updates. However, this can become an issue if you deploy a distance vector protocol in a large network.	**Limited scalability.** Slow convergence may limit the size of the network because larger networks require more time to propagate routing information.
	Routing loops. Routing loops can occur when inconsistent routing tables are not updated due to slow convergence in a changing network.

Link State routing protocols

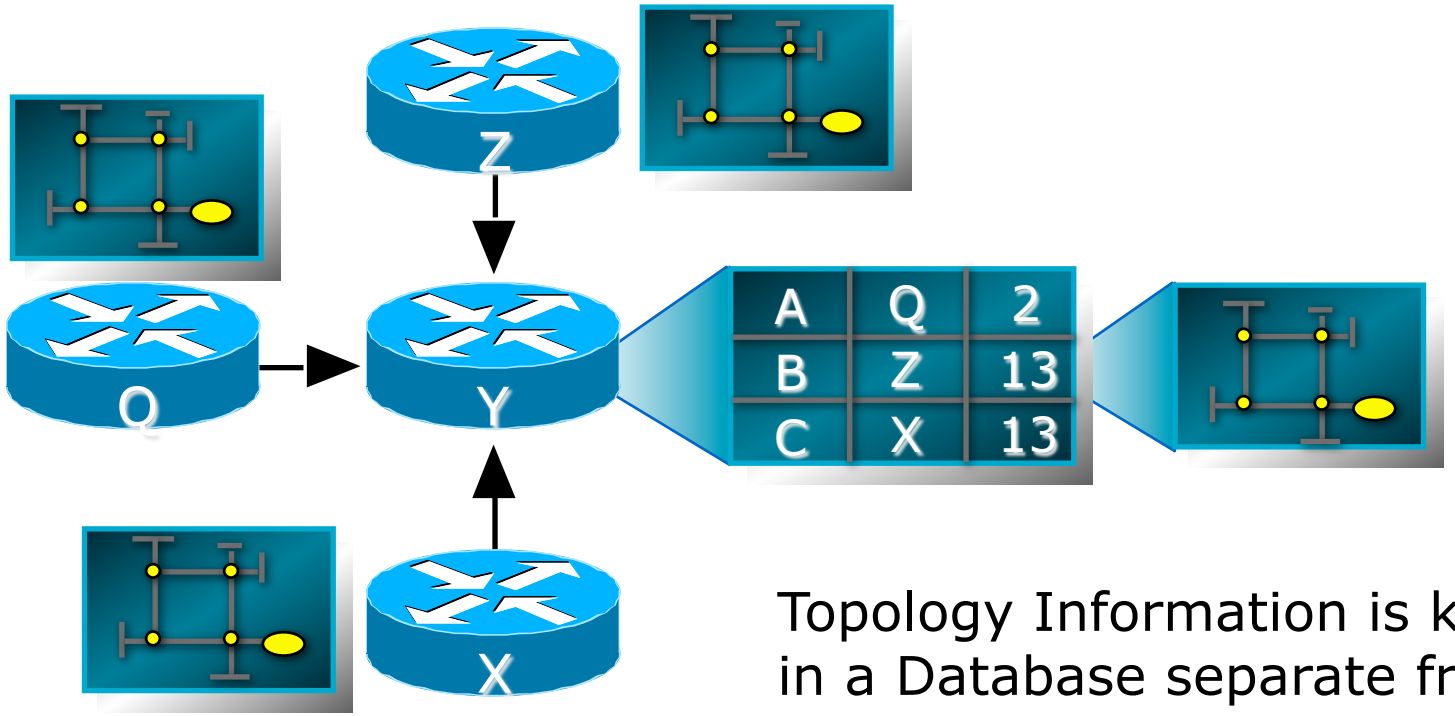
APNIC

OSPF

- Open Shortest Path First
- Link state or SPF technology
- Developed by OSPF working group of IETF (RFC 1247)
- OSPFv2 standard described in RFC2328

- Designed for:
  - TCP/IP environment
  - Fast convergence
  - Variable-length subnet masks
  - Discontiguous subnets
  - Incremental updates
  - Route authentication
- Runs on IP, Protocol 89

Link State

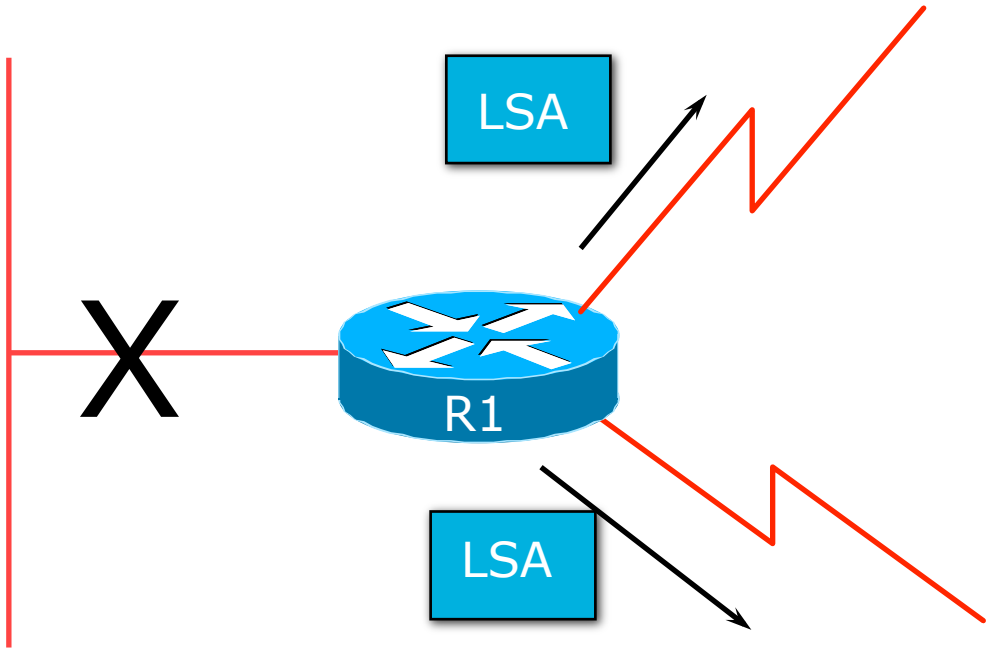


Topology Information is kept in a Database separate from the Routing Table

Link State Routing

- Neighbour discovery
- Constructing a Link State Packet (LSP)
- Distribute the LSP
  - (Link State Announcement – LSA)
- Compute routes
- On network failure
  - New LSPs flooded
  - All routers recompute routing table

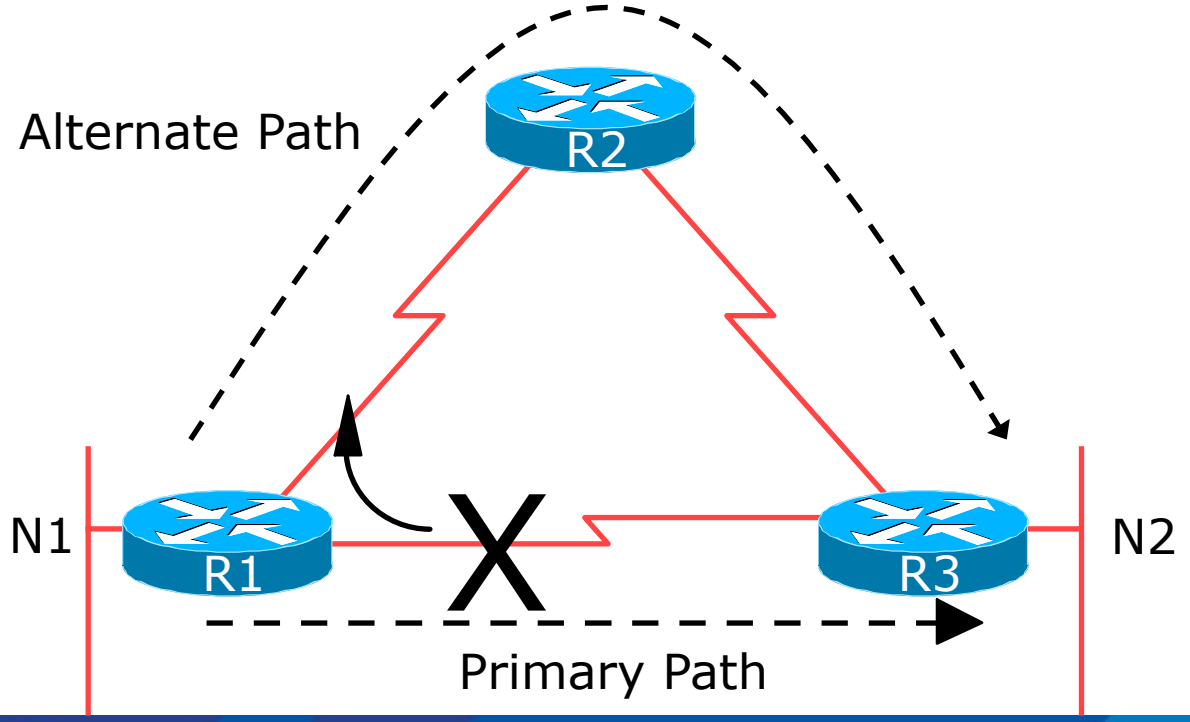
Low Bandwidth Utilisation



- Only changes propagated
- Uses multicast on multi-access broadcast networks

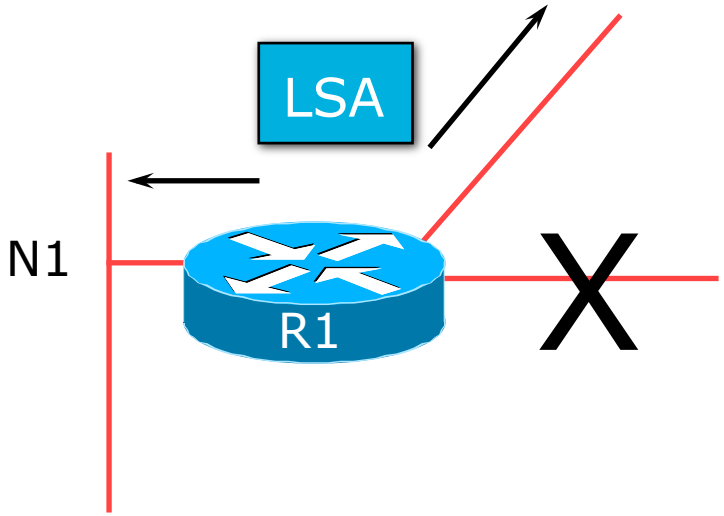
Fast Convergence

- Detection Plus LSA/SPF
  - Known as the Dijkstra Algorithm



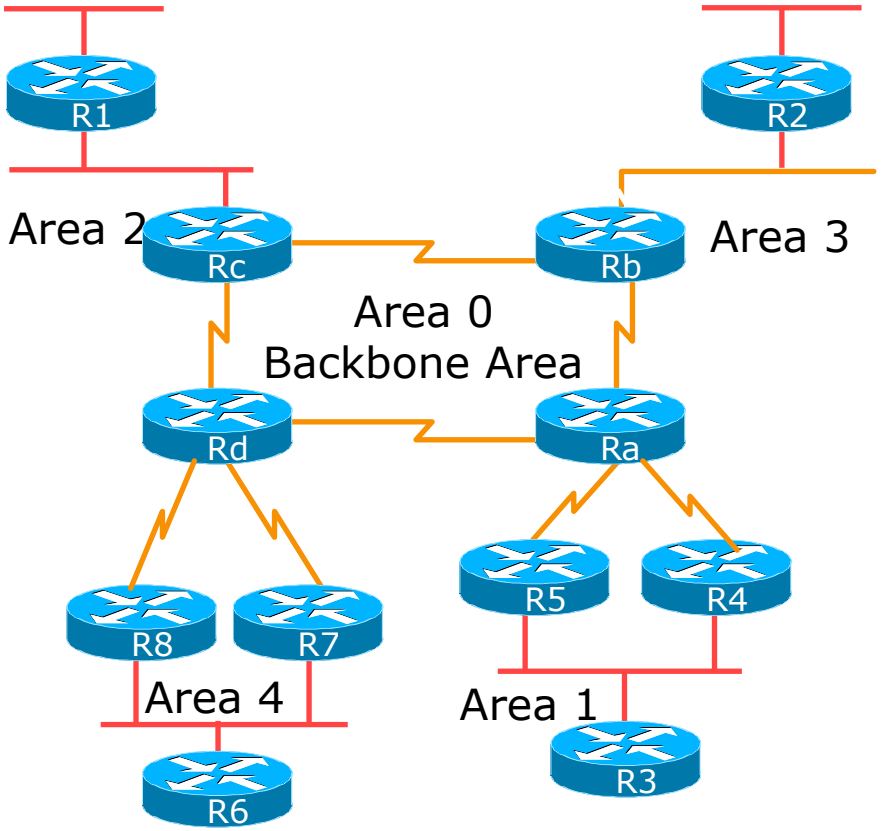
Fast Convergence

- Finding a new route
  - LSA flooded throughout area
  - Acknowledgement based
  - Topology database synchronised
  - Each router derives routing table to destination network



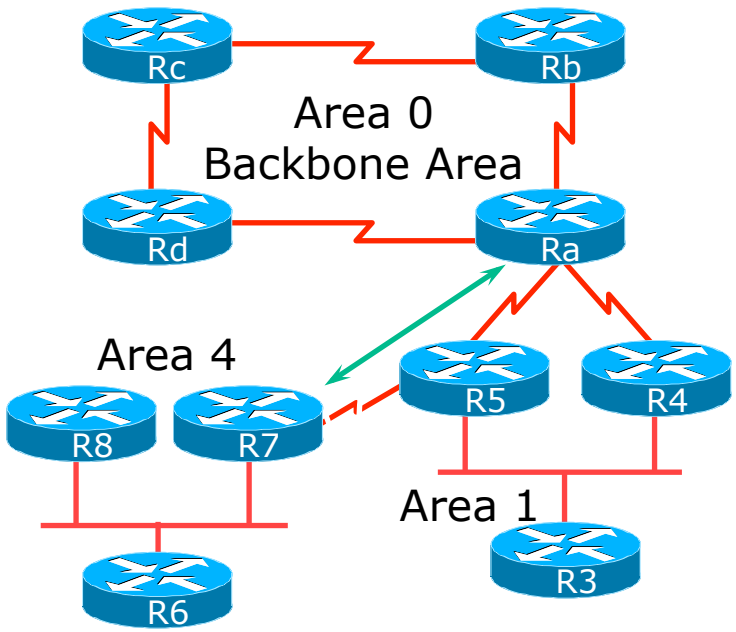
OSPF Areas

- Area is a group of contiguous hosts and networks
  - Reduces routing traffic
- Per area topology database
  - Invisible outside the area
- Backbone area **MUST** be contiguous
  - All other areas must be connected to the backbone

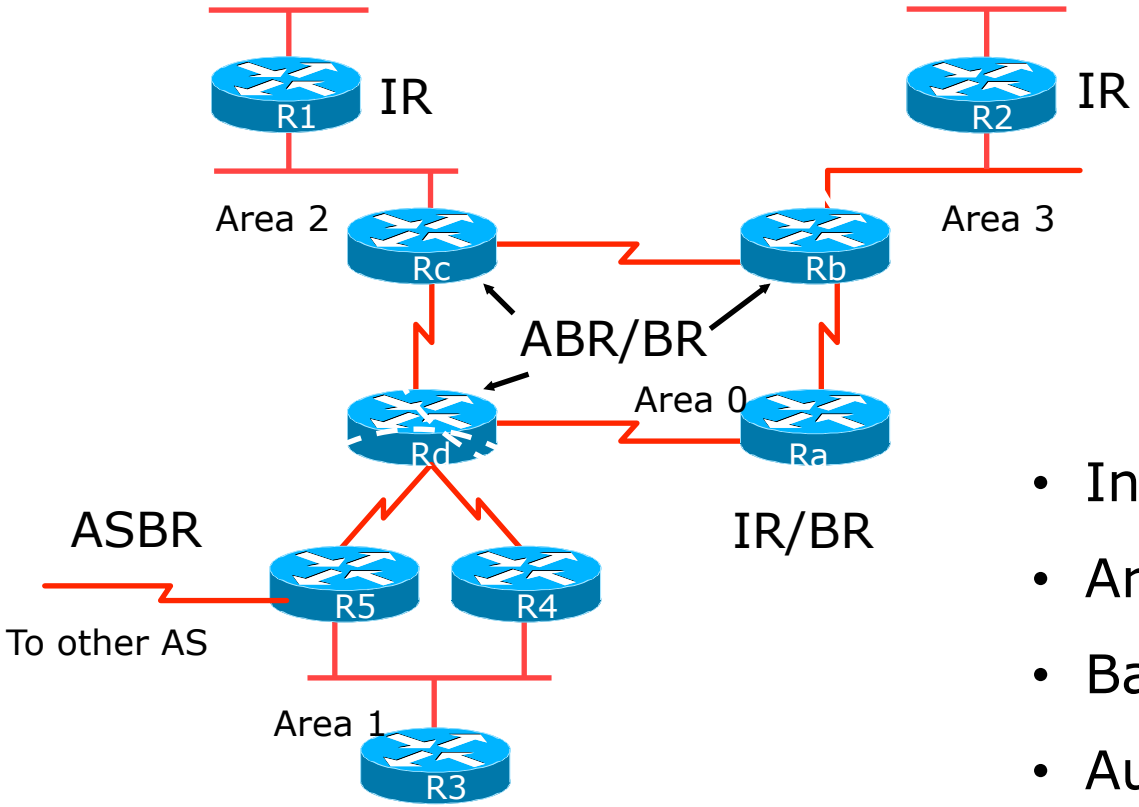


Virtual Links between OSPF Areas

- Virtual Link is used when it is not possible to physically connect the area to the backbone
- **ISPs avoid designs which require virtual links**
  - Increases complexity
  - Decreases reliability and scalability

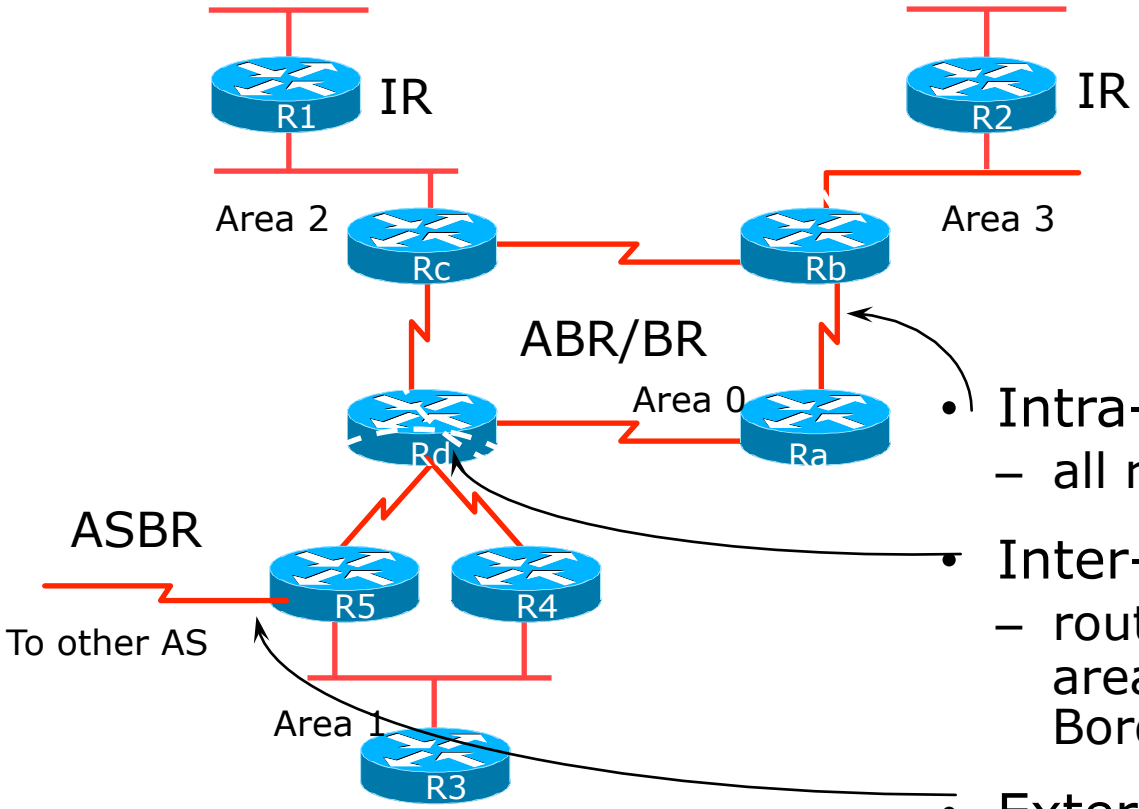


Classification of Routers



- Internal Router (IR)
- Area Border Router (ABR)
- Backbone Router (BR)
- Autonomous System Border Router (ASBR)

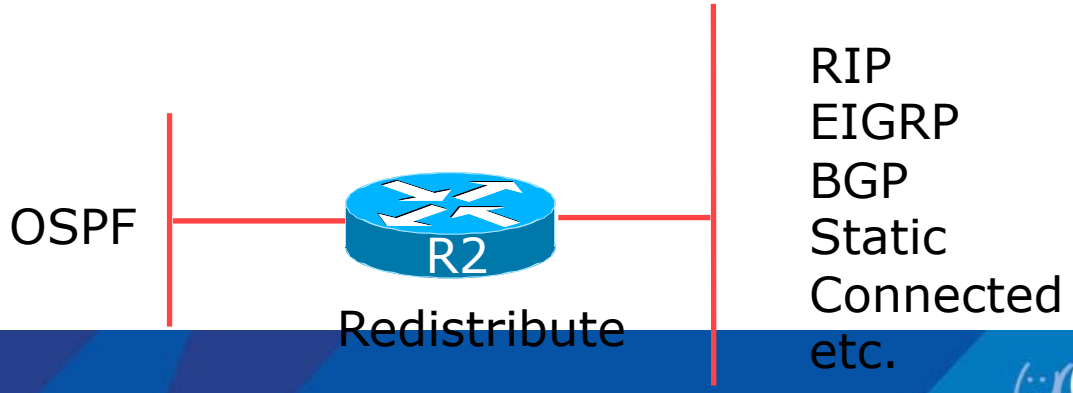
OSPF Route Types



- Intra-area Route
  - all routes inside an area
- Inter-area Route
  - routes advertised from one area to another by an Area Border Router
- External Route
  - routes imported into OSPF from other protocol or static routes

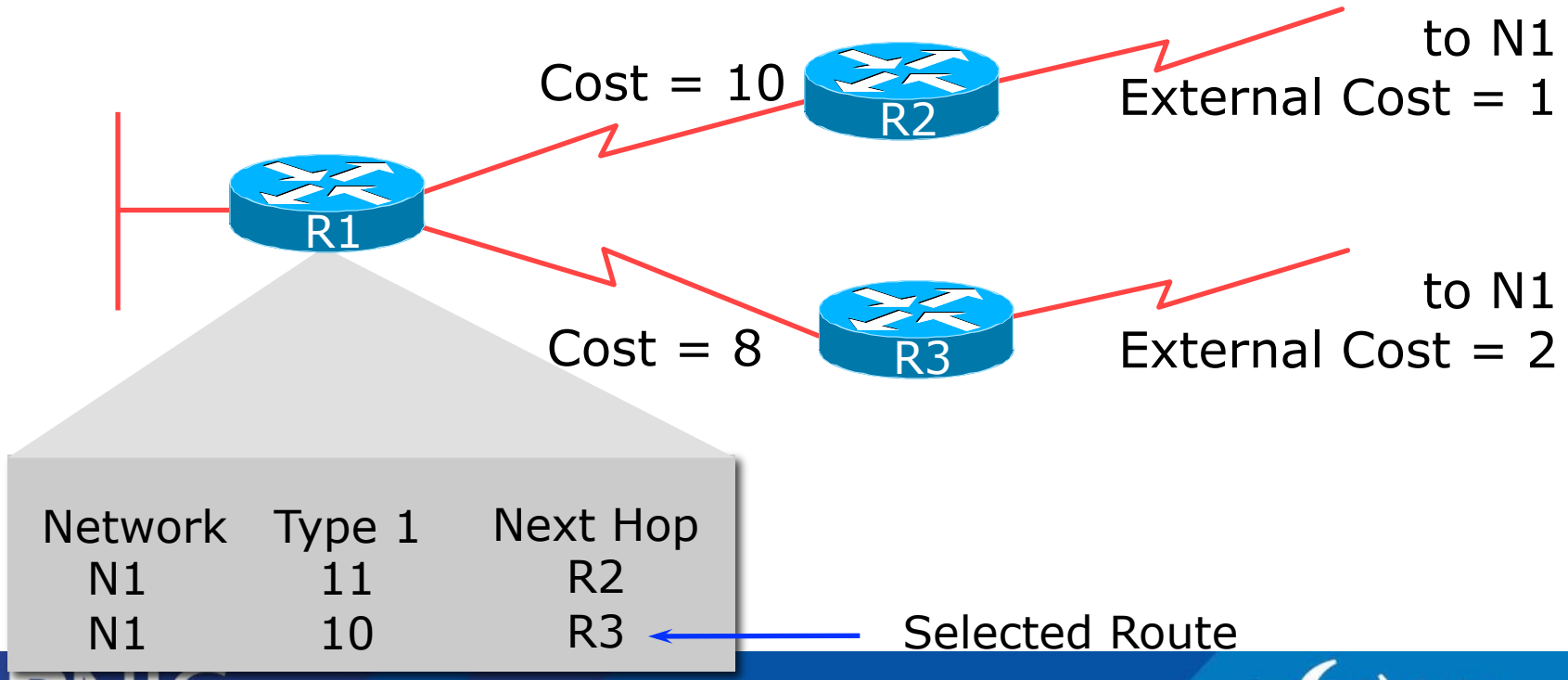
External Routes

- Prefixes which are redistributed into OSPF from other protocols
- Flooded unaltered throughout the AS
  - **Recommendation: Avoid redistribution!!**
- OSPF supports two types of external metrics
  - Type 1 external metrics
  - Type 2 external metrics (Cisco IOS default)



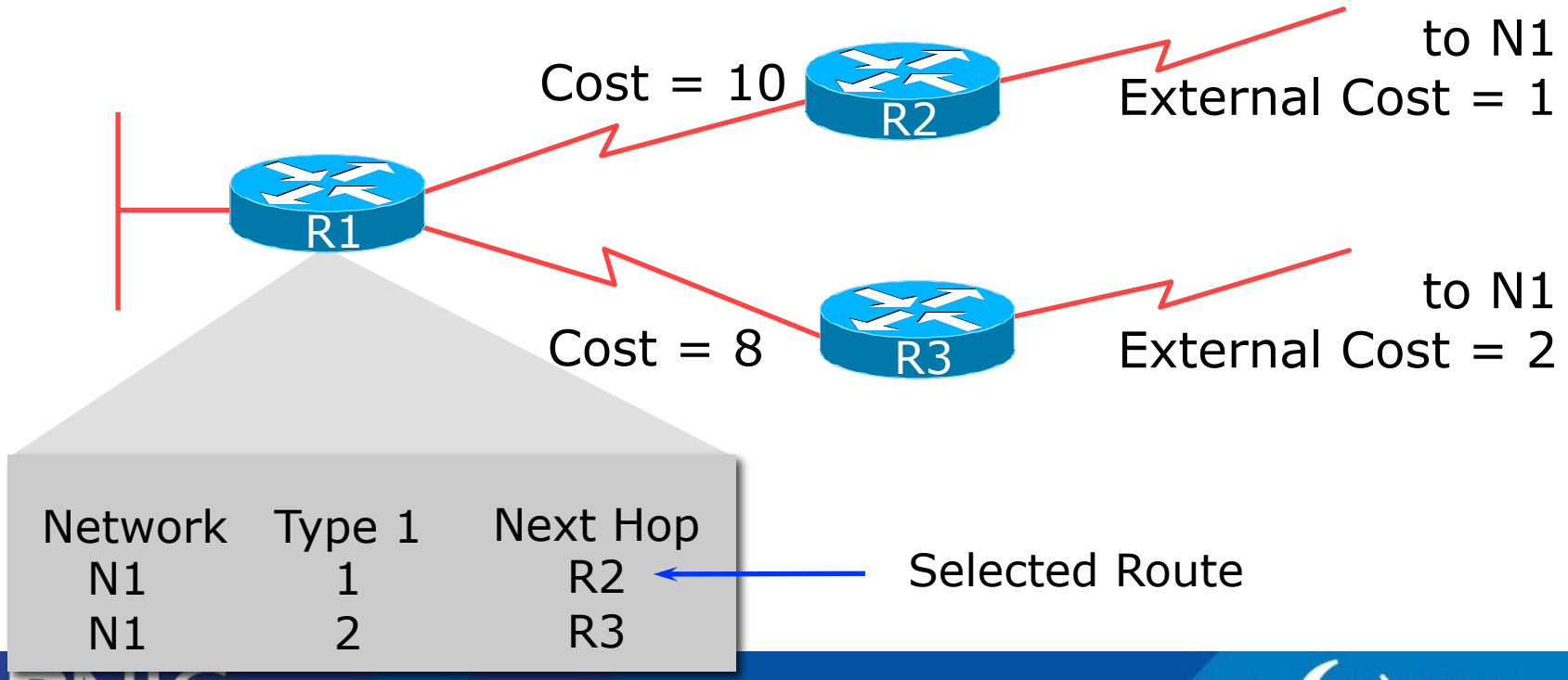
External Routes

- Type 1 external metric: metrics are added to the summarised internal link cost



External Routes

- Type 2 external metric: metrics are compared without adding to the internal link cost

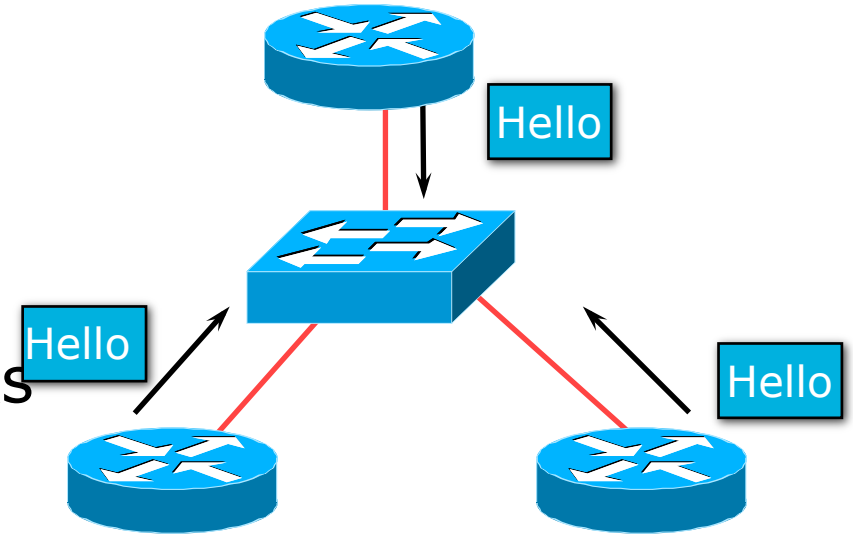


Topology/Link State Database

- A router has a separate LS database for each area to which it belongs
- All routers belonging to the same area have identical database
- SPF calculation is performed separately for each area
- LSA flooding is bounded by area
- Recommendation:
  - Limit the number of areas a router participates in!!
  - 1 to 3 is fine (typical ISP design)
  - >3 can overload the CPU depending on the area topology complexity

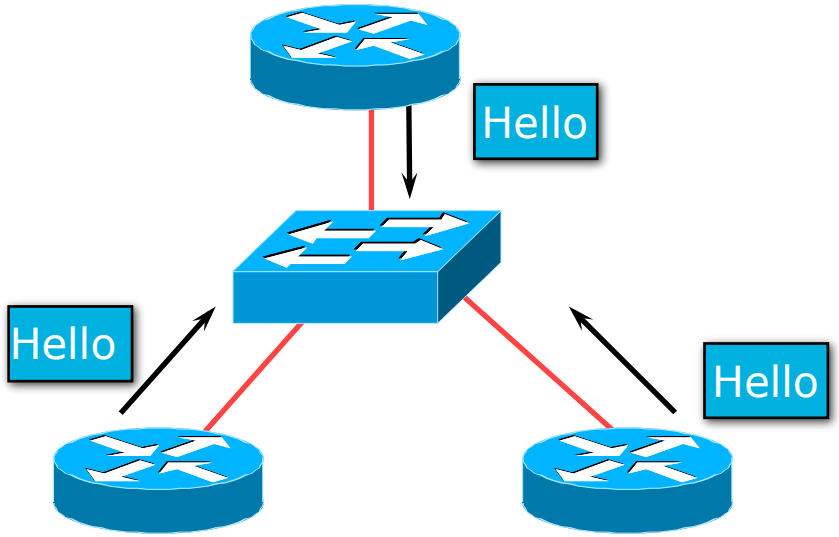
The Hello Protocol

- Responsible for establishing and maintaining neighbour relationships
- Elects designated router on multi-access networks



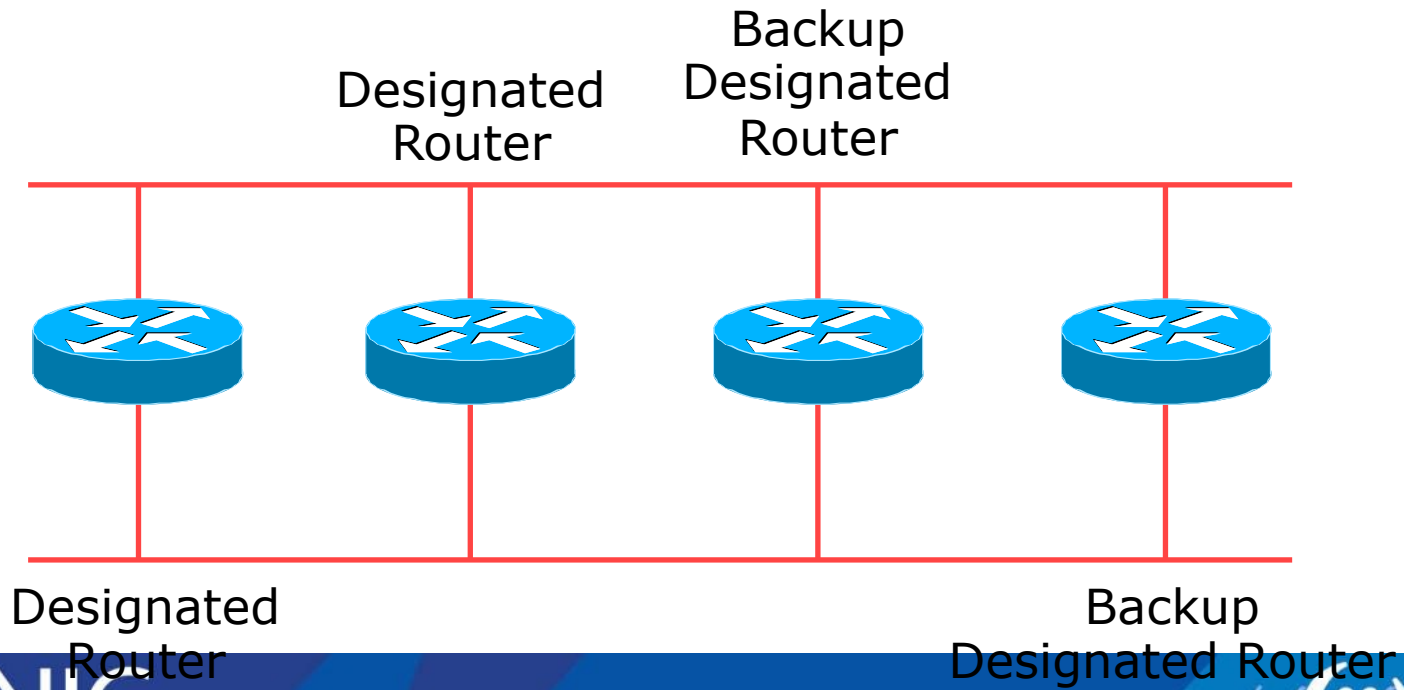
The Hello Packet

- Contains:
  - Router priority
  - Hello interval
  - Router dead interval
  - Network mask
  - List of neighbours
  - DR and BDR
  - Options: E-bit, MC-bit,… (see A.2 of RFC2328)



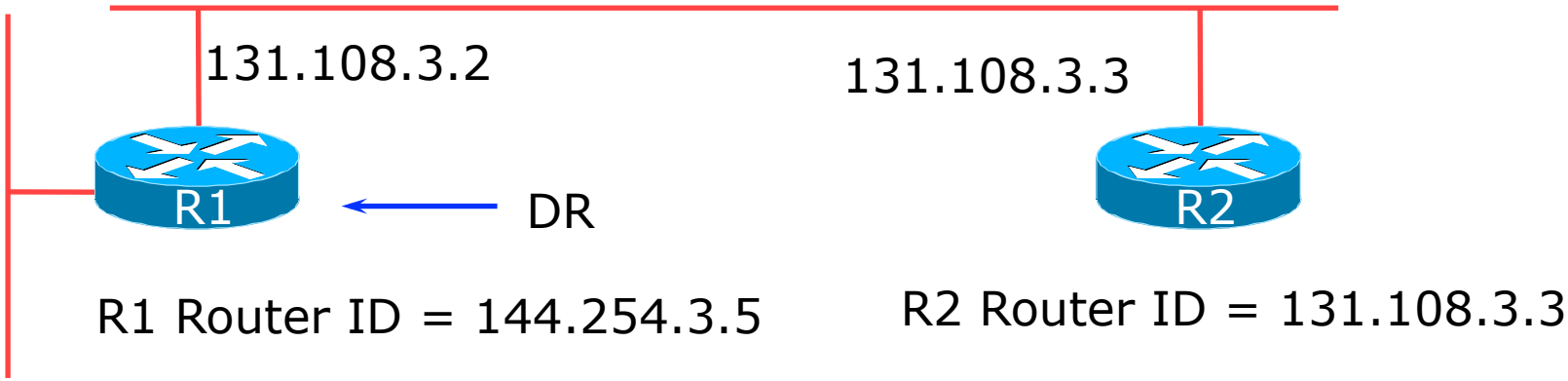
Designated Router

- There is ONE designated router per multi-access network
  - Generates network link advertisements
  - Assists in database synchronization



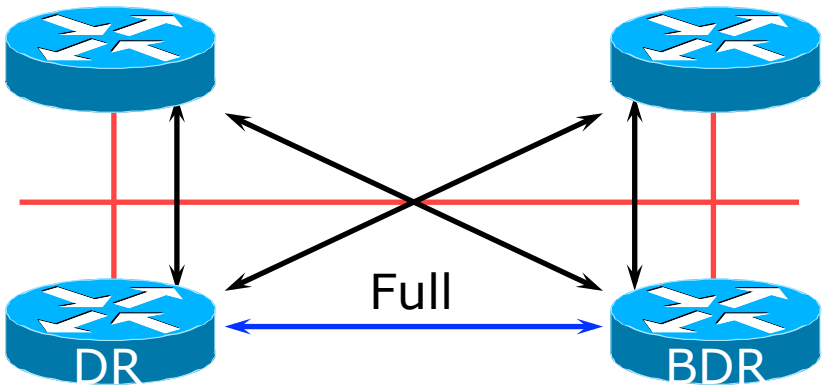
Designated Router by Priority

- Configured priority (per interface)
  - ISPs configure high priority on the routers they want as DR/BDR
- Else determined by highest router ID
  - Router ID is 32 bit integer
  - Derived from the loopback interface address, if configured, otherwise the highest IP address



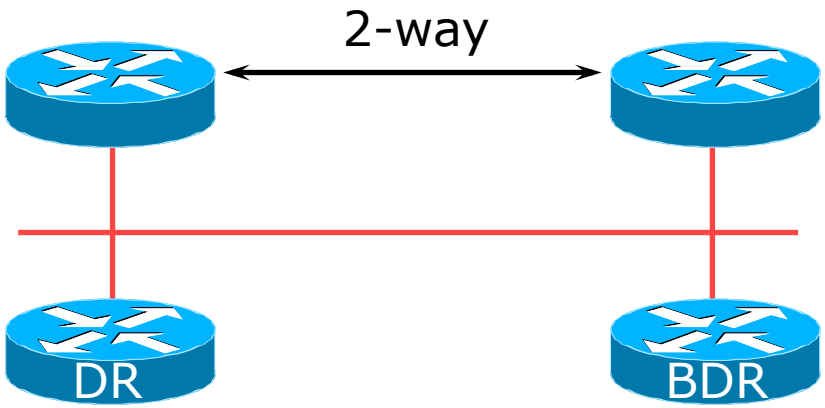
Neighbouring States

- Full
  - Routers are fully adjacent
  - Databases synchronised
  - Relationship to DR and BDR



Neighbouring States

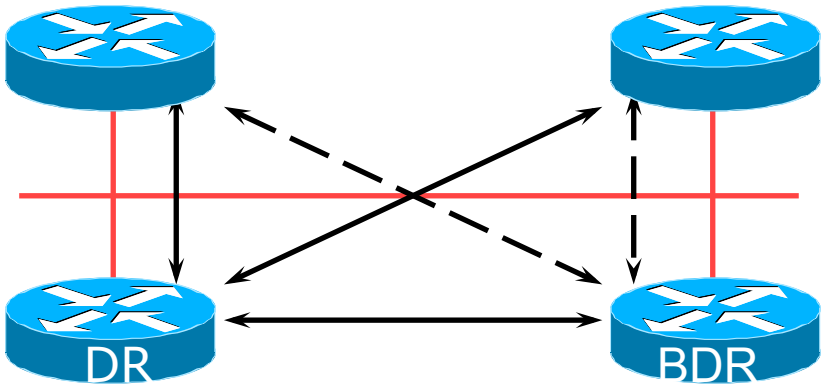
- 2-way
  - Router sees itself in other Hello packets
  - DR selected from neighbours in state 2-way or greater



When to Become Adjacent

- Underlying network is point to point
- Underlying network type is virtual link
- The router itself is the designated router or the backup designated router
- The neighbouring router is the designated router or the backup designated router

LSAs Propagate Along Adjacencies



- LSAs acknowledged along adjacencies

Broadcast Networks

- IP Multicast used for Sending and Receiving Updates
  - All routers must accept packets sent to AllSPFRouters (224.0.0.5)
  - All DR and BDR routers must accept packets sent to AllDRouters (224.0.0.6)
- Hello packets sent to AllSPFRouters (Unicast on point-to-point and virtual links)

Routing Protocol Packets

- Share a common protocol header
- Routing protocol packets are sent with type of service (TOS) of 0
- Five types of OSPF routing protocol packets
  - Hello – packet type 1
  - Database description – packet type 2
  - Link-state request – packet type 3
  - Link-state update – packet type 4
  - Link-state acknowledgement – packet type 5

Different Types of LSAs

- Six distinct type of LSAs
  - Type 1 : Router LSA
  - Type 2 : Network LSA
  - Type 3 & 4: Summary LSA
  - Type 5 & 7: External LSA (Type 7 is for NSSA)
  - Type 6: Group membership LSA
  - Type 9, 10 & 11: Opaque LSA (9: Link-Local, 10: Area)

Router LSA (Type 1)

- Describes the state and cost of the router's links to the area
- All of the router's links in an area must be described in a single LSA
- Flooded throughout the particular area and no more
- Router indicates whether it is an ASBR, ABR, or end point of virtual link

Network LSA (Type 2)

- Generated for every transit broadcast and NBMA network
- Describes all the routers attached to the network
- Only the designated router originates this LSA
- Flooded throughout the area and no more

Summary LSA (Type 3 and 4)

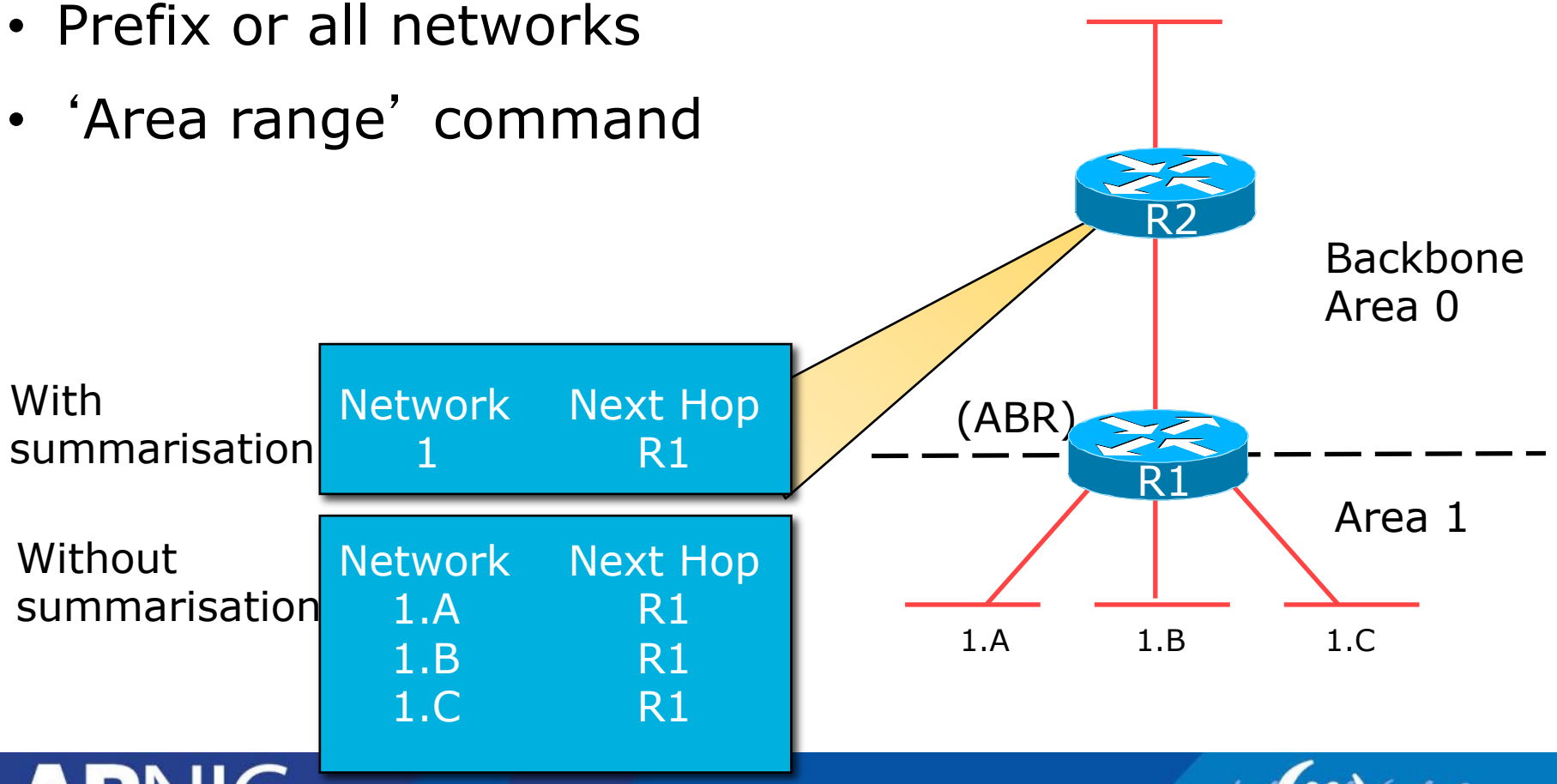
- Describes the destination outside the area but still in the AS
- Flooded throughout a single area
- Originated by an ABR
- Only inter-area routes are advertised into the backbone
- Type 4 is the information about the ASBR

External LSA (Type 5 and 7)

- Defines routes to destination external to the AS
- Default route is also sent as external
- Two types of external LSA:
  - E1: Consider the total cost up to the external destination
  - E2: Considers only the cost of the outgoing interface to the external destination
- (Type 7 LSAs used to describe external LSA for one specific OSPF area type)

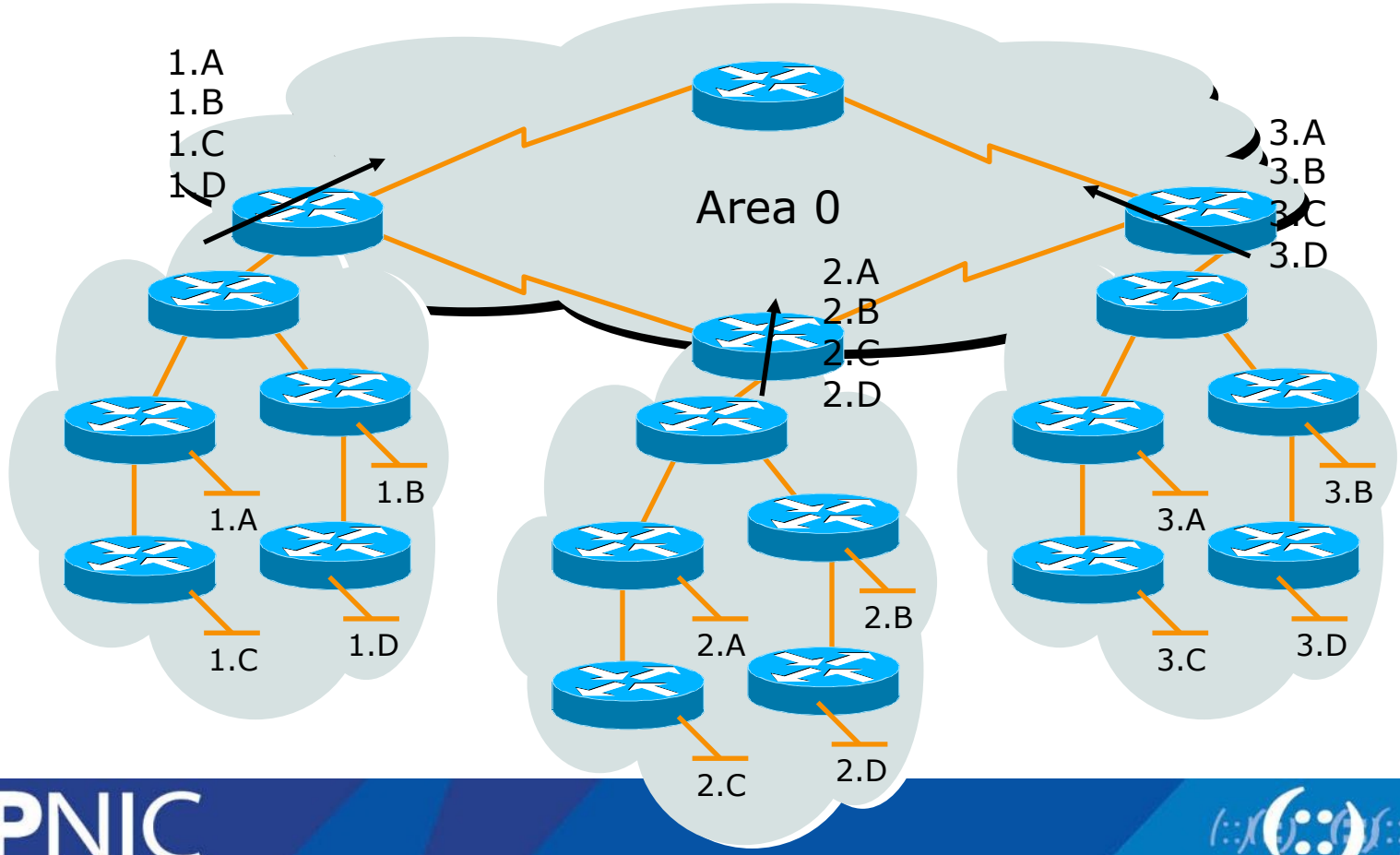
Inter-Area Route Summarisation

- Prefix or all subnets
- Prefix or all networks
- ‘Area range’ command



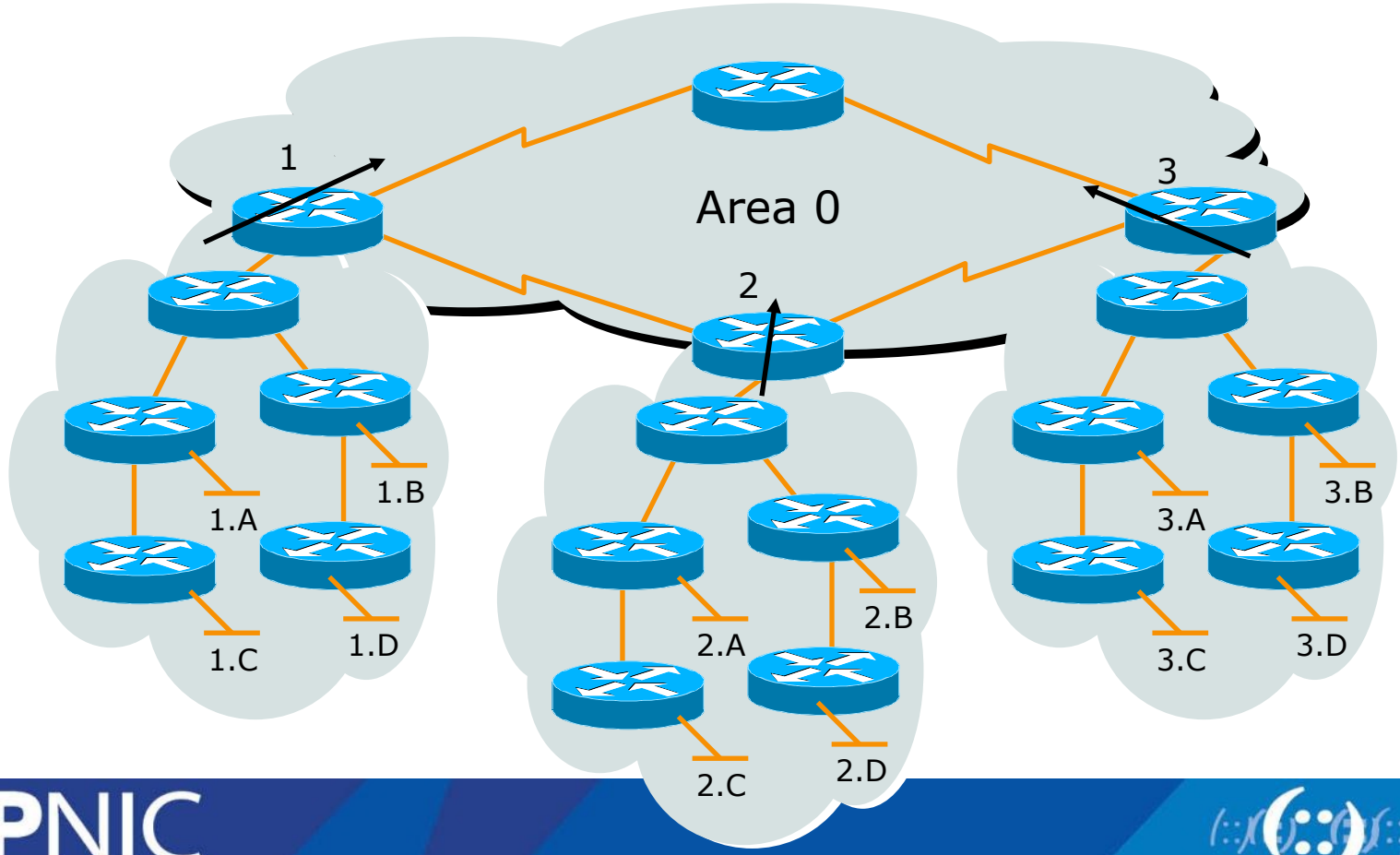
No Summarisation

- Specific Link LSA advertised out of each area
- Link state changes propagated out of each area



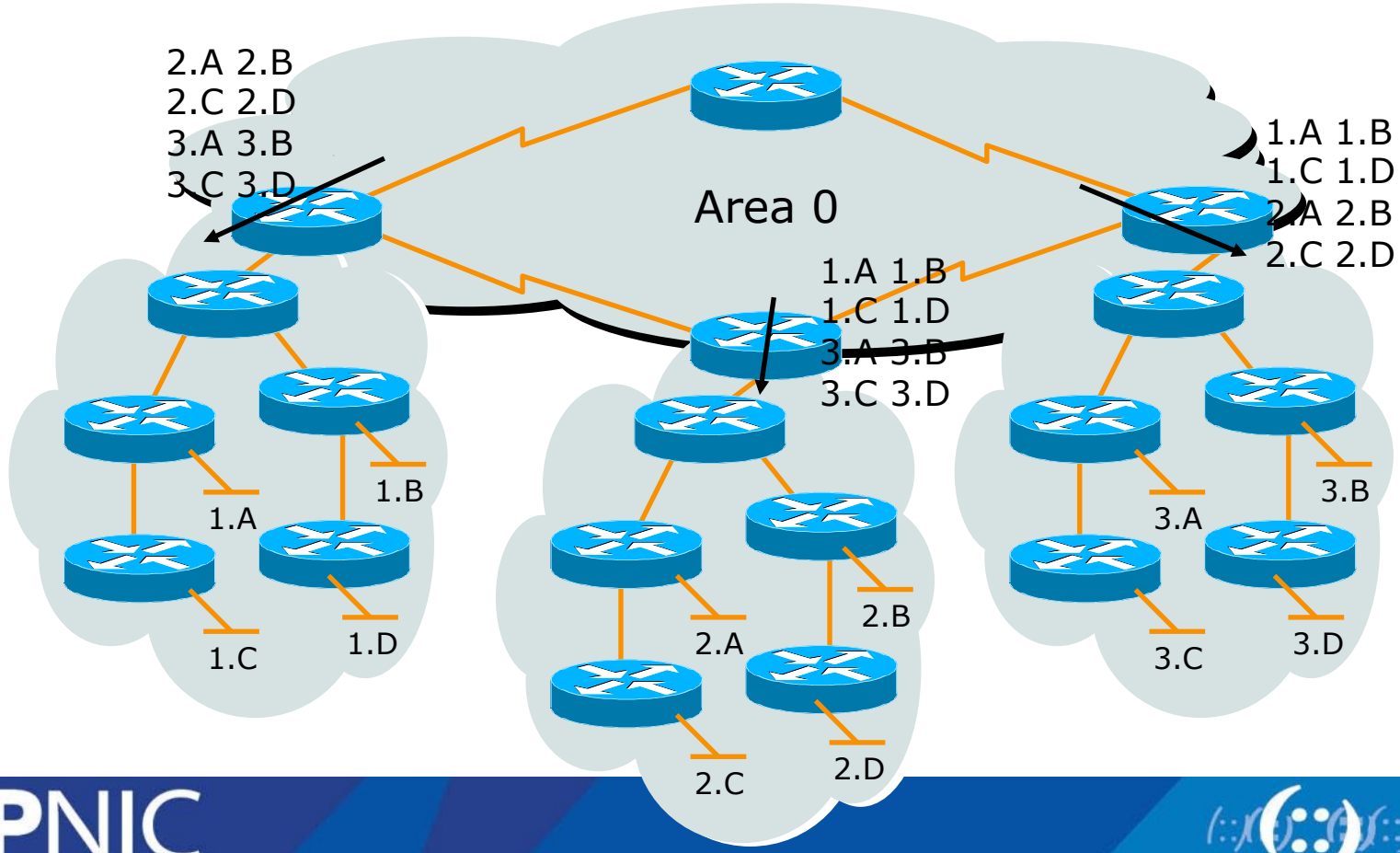
With Summarisation

- Only summary LSA advertised out of each area
- Link state changes do not propagate out of the area



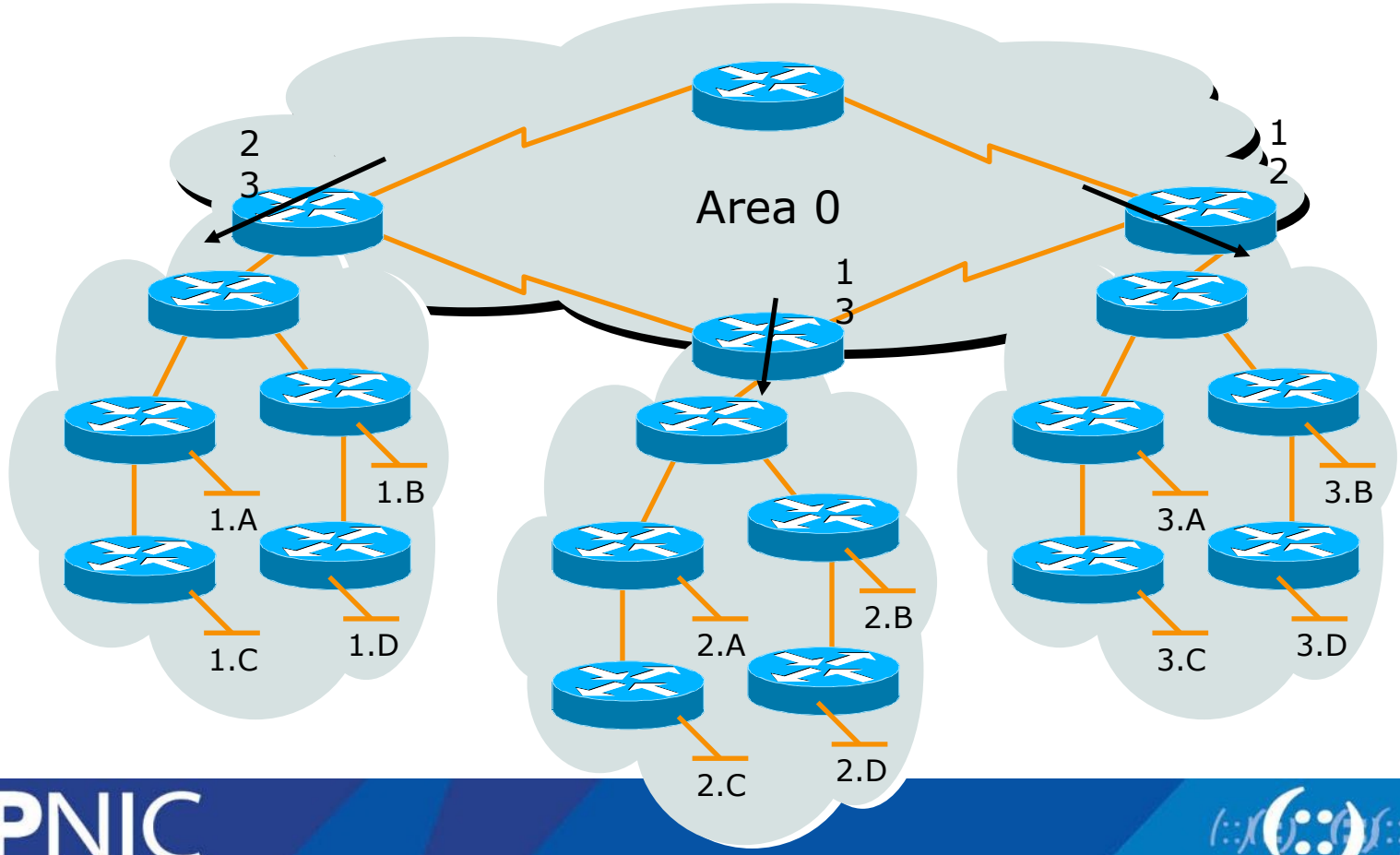
No Summarisation

- Specific Link LSA advertised in to each area
- Link state changes propagated in to each area



With Summarisation

- Only summary link LSA advertised in to each area
- Link state changes do not propagate in to each area

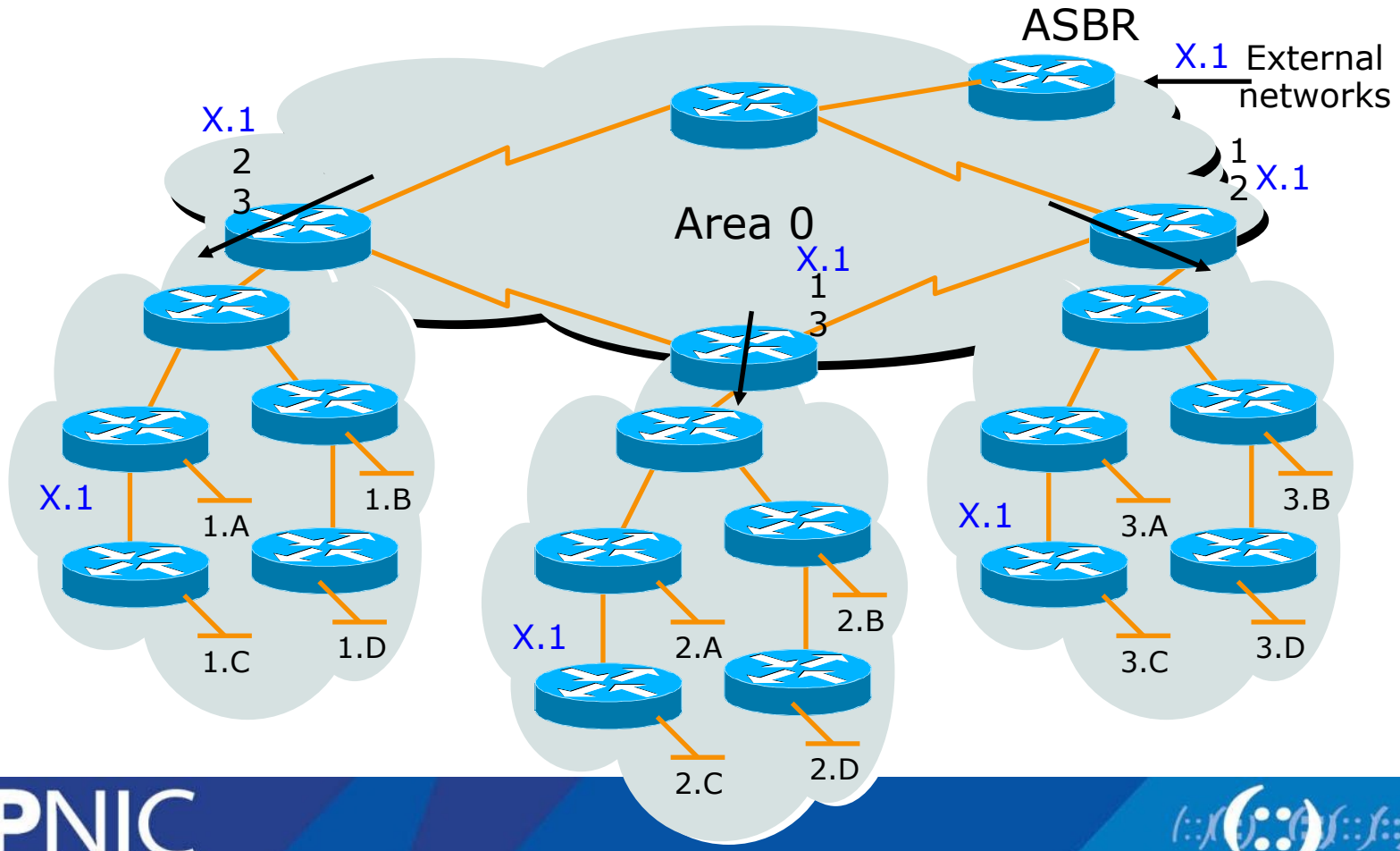


Types of Areas

- Regular
- Stub
- Totally Stubby
- Not-So-Stubby
- **Only "regular" areas are useful for ISPs**
  - Other area types handle redistribution of other routing protocols into OSPF – ISPs don't redistribute anything into OSPF
- The next slides describing the different area types are provided for information only

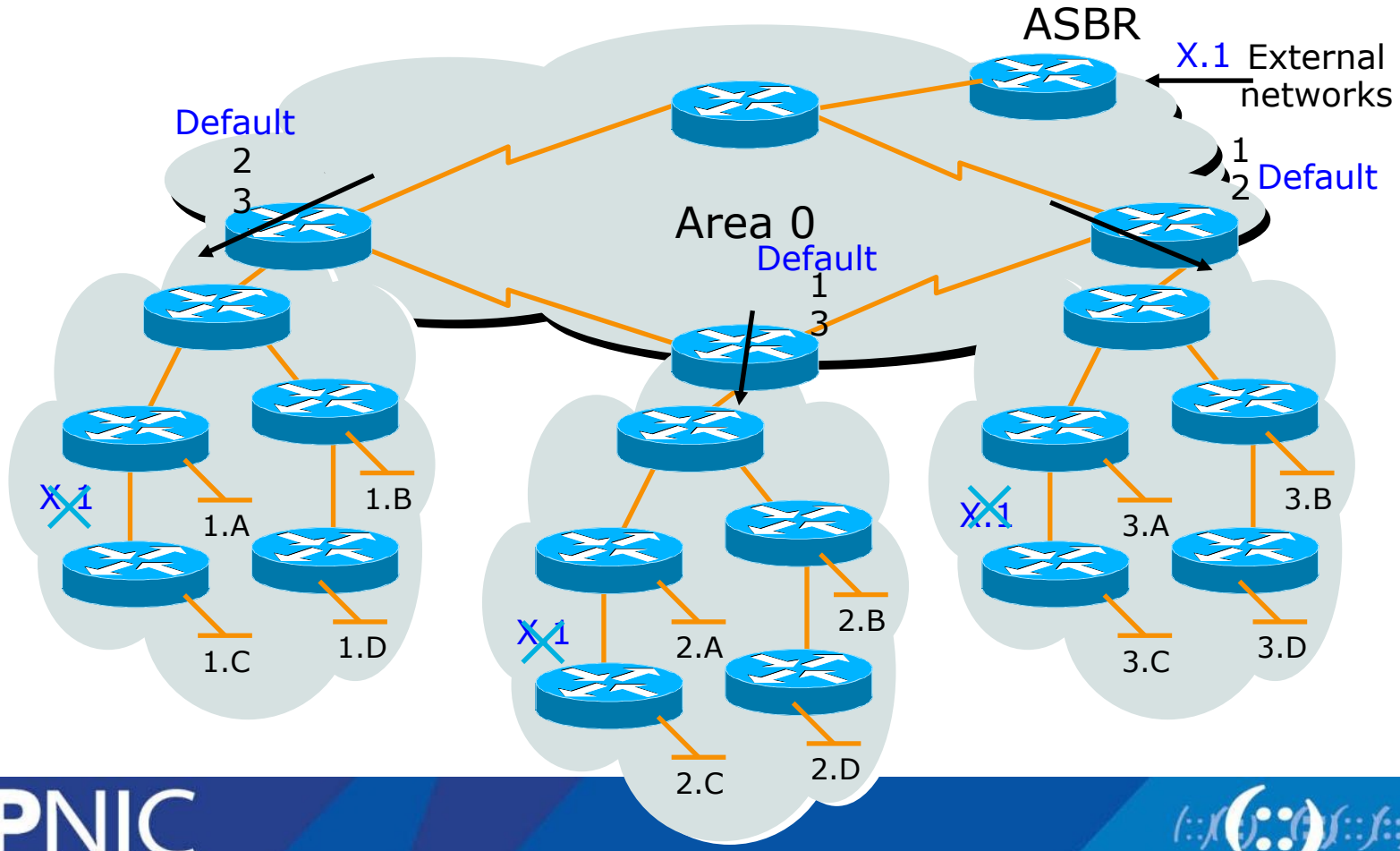
Regular Area (Not a Stub)

- From Area 1's point of view, summary networks from other areas are injected, as are external networks such as X.1



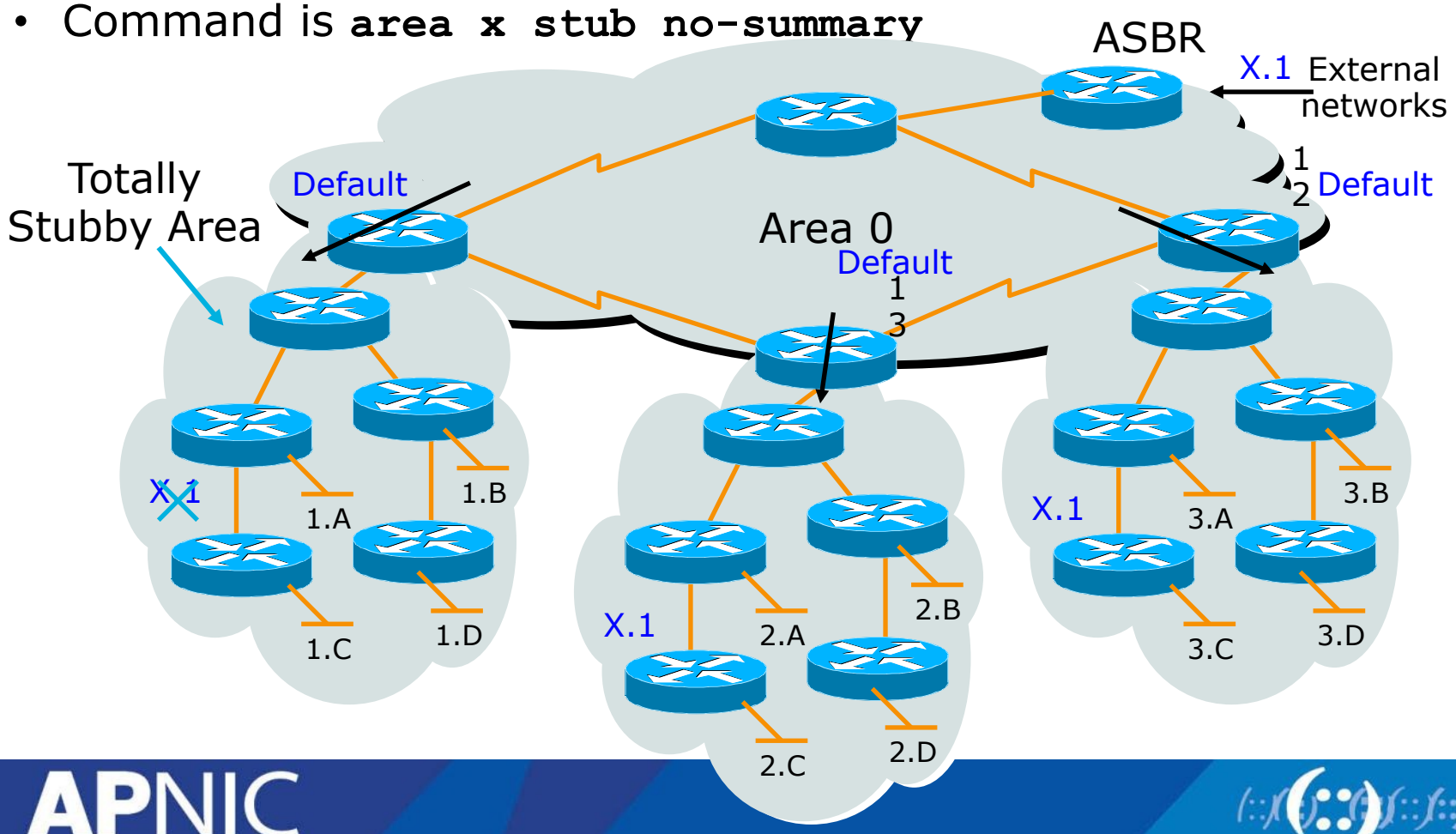
Normal Stub Area

- Summary networks, default route injected
- Command is **`area x stub`**



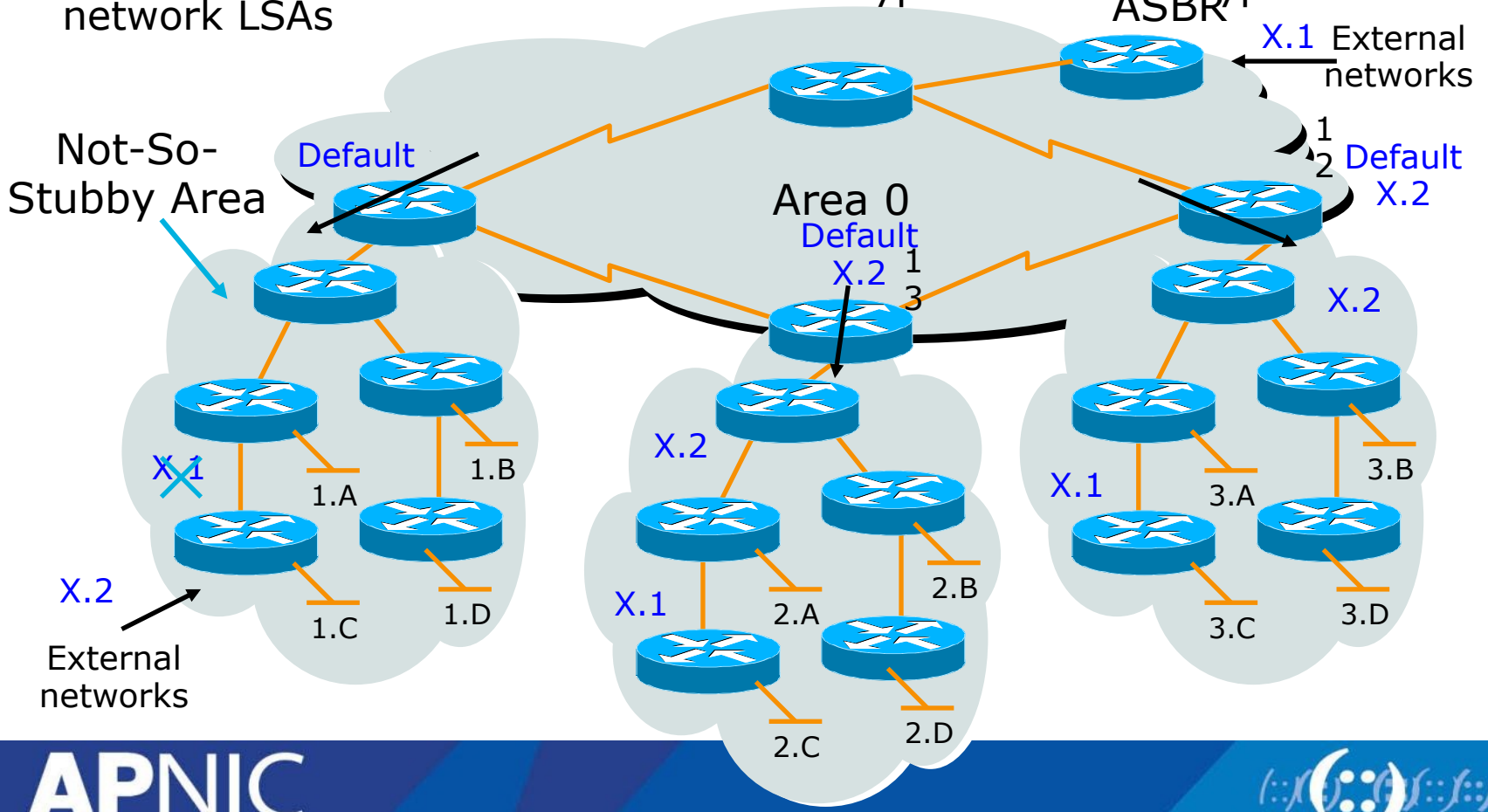
Totally Stubby Area

- Only a default route injected
  - Default path to closest area border router
- Command is **`area x stub no-summary`**



Not-So-Stubby Area

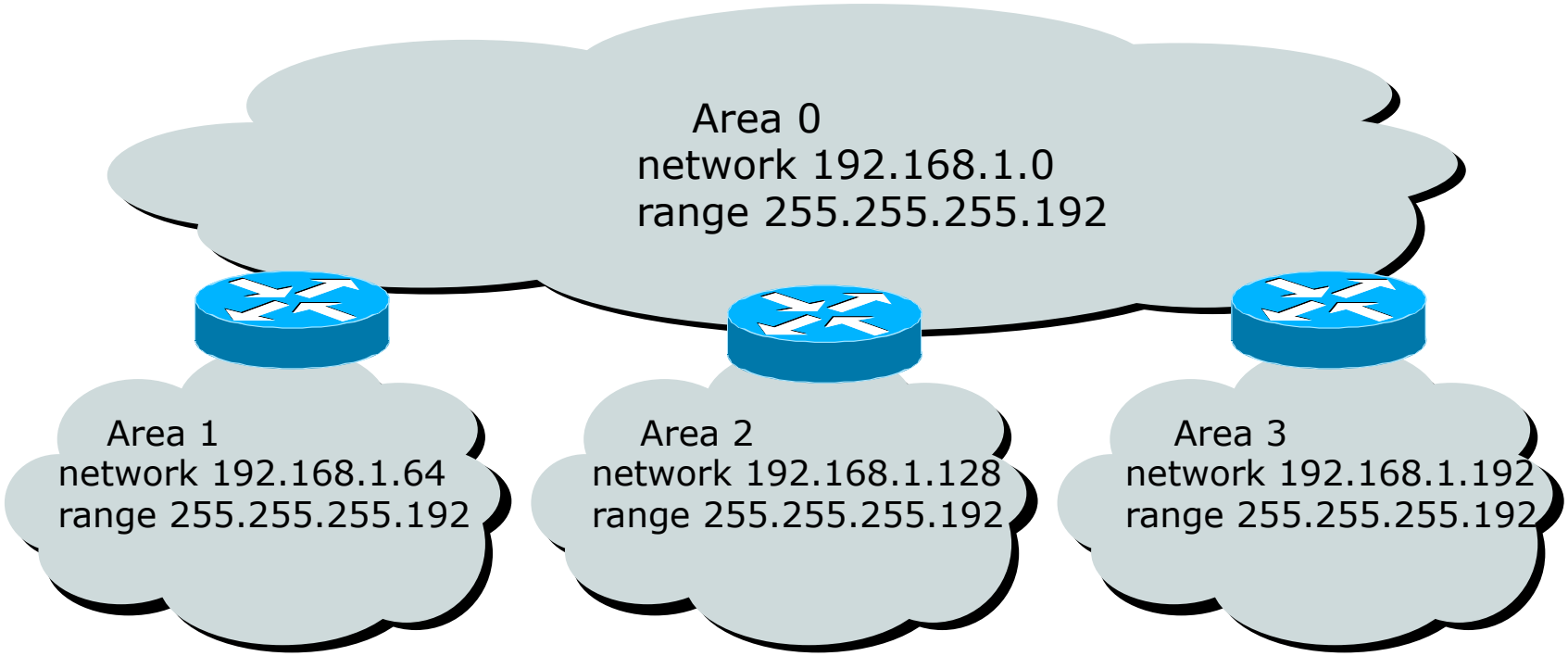
- Capable of importing routes in a limited fashion
- Type-7 LSA's carry external information within an NSSA
- NSSA Border routers translate selected type-7 LSAs into type-5 external network LSAs



ISP Use of Areas

- ISP networks use:
  - Backbone area
  - Regular area
- Backbone area
  - No partitioning
- Regular area
  - Summarisation of point to point link addresses used within areas
  - Loopback addresses allowed out of regular areas without summarisation (otherwise iBGP won't work)

Addressing for Areas



- Assign contiguous ranges of subnets per area to facilitate summarisation

Summary

- Fundamentals of Scalable OSPF Network Design
  - Area hierarchy
  - DR/BDR selection
  - Contiguous intra-area addressing
  - Route summarisation
  - Infrastructure prefixes only

BGP as an Inter AS Routing protocol

APNIC

Border Gateway Protocol

- A Routing Protocol used to exchange routing information between different networks
  - Exterior gateway protocol
- Described in RFC4271
  - RFC4276 gives an implementation report on BGP
  - RFC4277 describes operational experiences using BGP
- The Autonomous System is the cornerstone of BGP
  - It is used to uniquely identify networks with a common routing policy

BGP

- Path Vector Protocol
- Incremental Updates
- Many options for policy enforcement
- Classless Inter Domain Routing (CIDR)
- Widely used for Internet backbone
- Autonomous systems

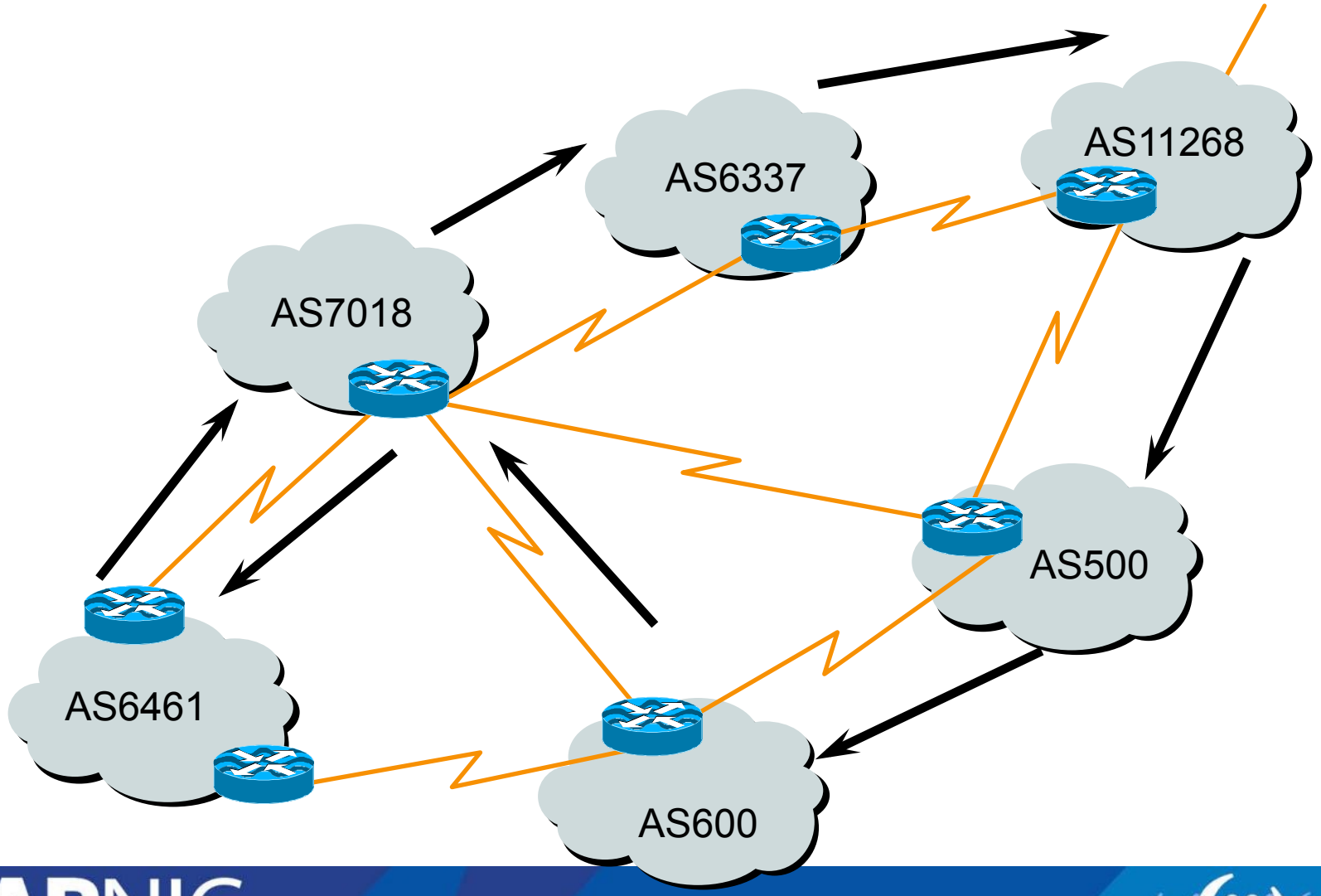
Path Vector Protocol

- BGP is classified as a *path vector* routing protocol (see RFC 1322)
  - A path vector protocol defines a route as a pairing between a destination and the attributes of the path to that destination.

```
12.6.126.0/24 207.126.96.43  1021  0 6461 7018 6337 11268 i
```

AS Path

Path Vector Protocol



Definitions

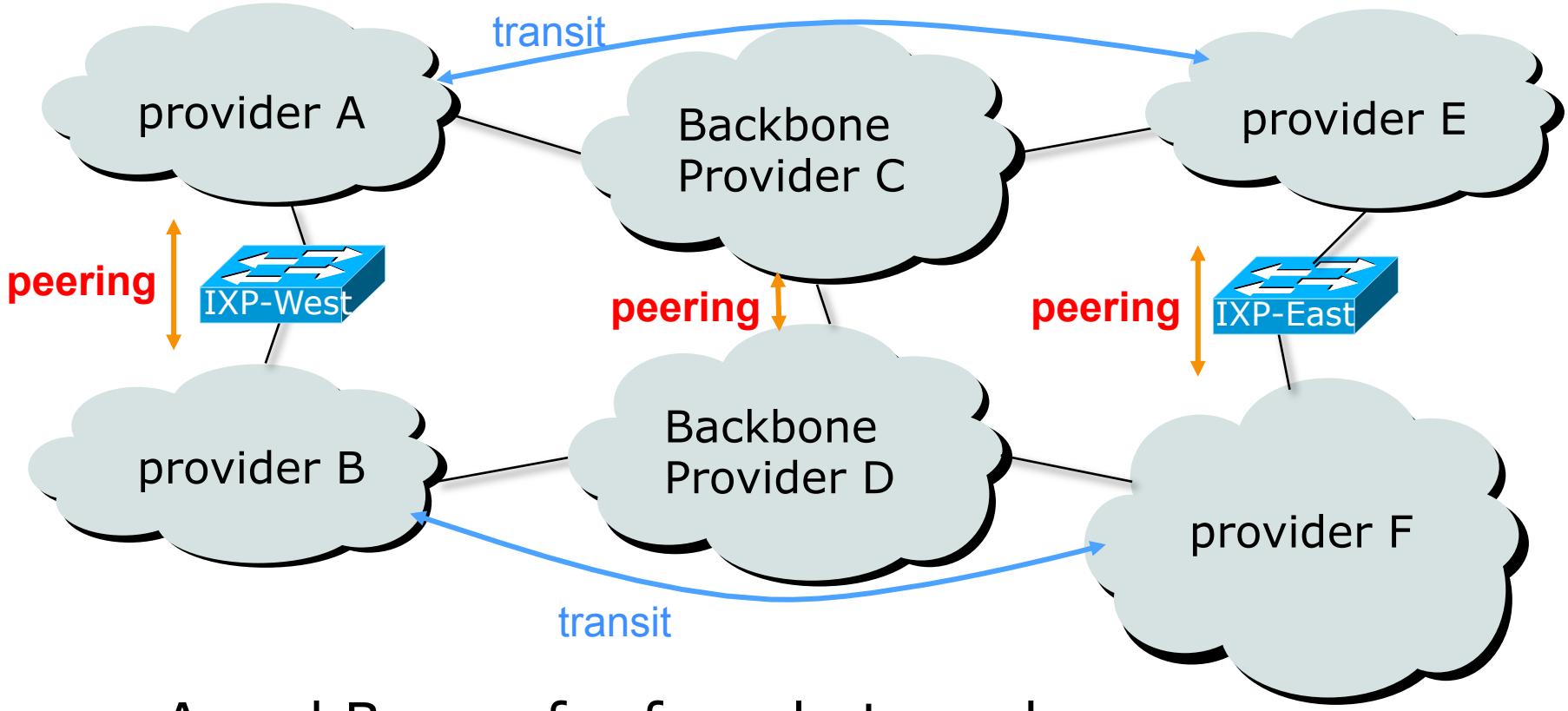
- Transit – carrying traffic across a network, usually for a fee
- Peering – exchanging routing information and traffic
- Default – where to send traffic when there is no explicit match in the routing table

Default Free Zone

The default free zone is made up of Internet routers which have explicit routing information about the rest of the Internet, and therefore do not need to use a default route

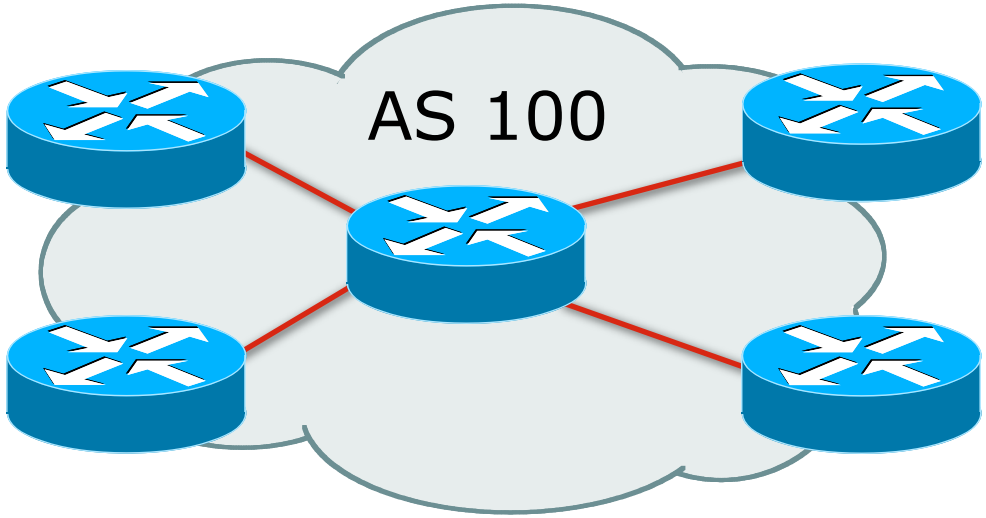
NB: is not related to where an ISP is in the hierarchy

Peering and Transit example



A and B peer for free, but need transit arrangements with C and D to get packets to/from E and F

Autonomous System (AS)



- Collection of networks with same routing policy
- Single routing protocol
- Usually under single ownership, trust and administrative control
- Identified by a unique 32-bit integer (ASN)

Autonomous System Number (ASN)

- Two ranges

0-65535	(original 16-bit range)
65536-4294967295	(32-bit range – RFC6793)

- Usage:

0 and 65535	(reserved)
1-64495	(public Internet)
64496-64511	(documentation – RFC5398)
64512-65534	(private use only)
23456	(represent 32-bit range in 16-bit world)
65536-65551	(documentation – RFC5398)
65552-4199999999	(public Internet)
4200000000-4294967295	(private use only)

- 32-bit range representation specified in RFC5396
  - Defines “asplain” (traditional format) as standard notation

Autonomous System Number (ASN)

- ASNs are distributed by the Regional Internet Registries
  - They are also available from upstream ISPs who are members of one of the RIRs
- Current 16-bit ASN assignments up to 63487 have been made to the RIRs
  - Around 44500 are visible on the Internet
  - Around 1500 left unassigned
- Each RIR has also received a block of 32-bit ASNs
  - Out of 4800 assignments, around 3700 are visible on the Internet
- See www.iana.org/assignments/as-numbers

Configuring BGP in Cisco IOS

- This command enables BGP in Cisco IOS:

```
router bgp 100
```

- For ASNs > 65535, the AS number can be entered in either plain or dot notation:

```
router bgp 131076
```

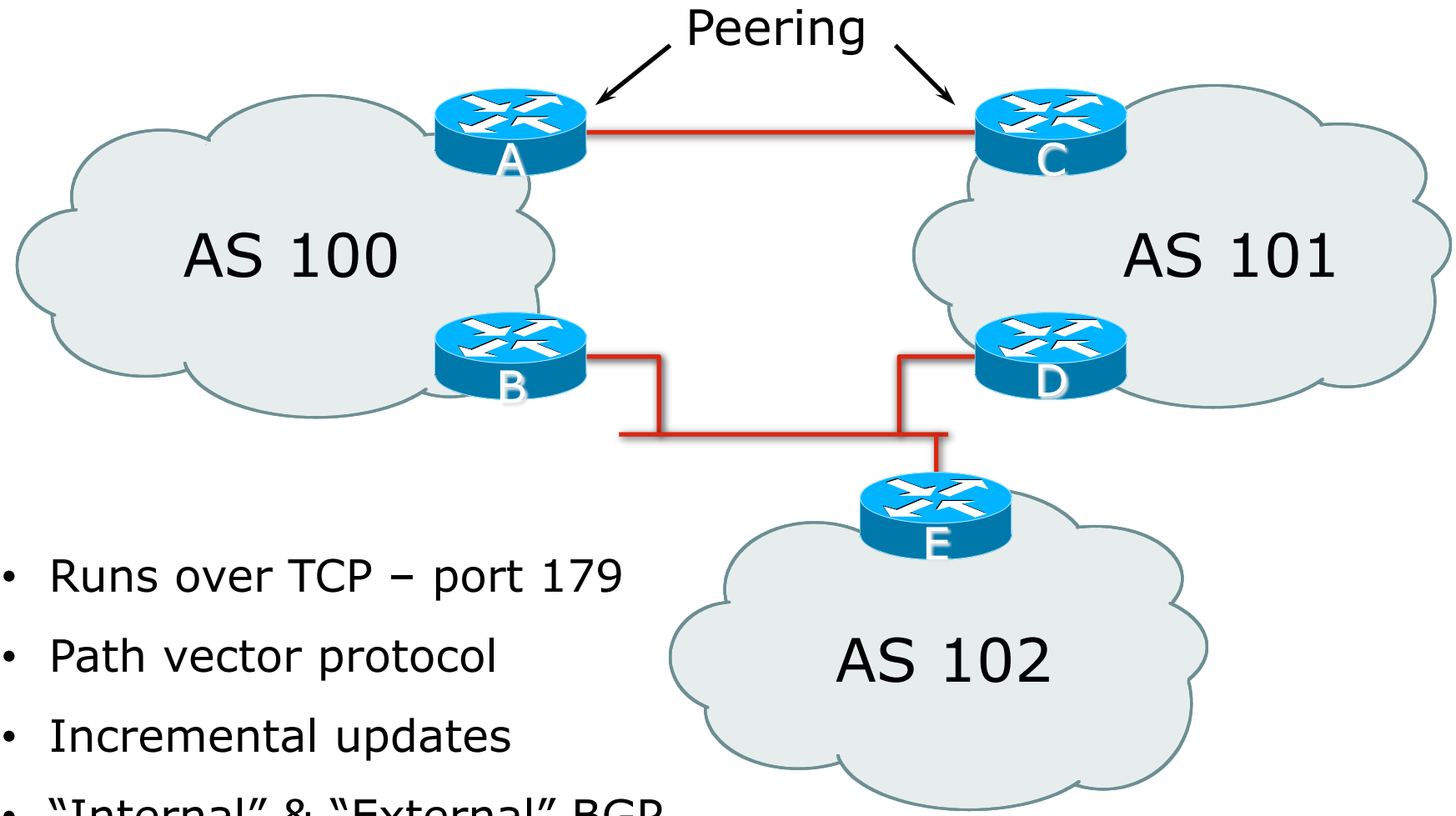
or

```
router bgp 2.4
```

- IOS will display ASNs in plain notation by default
  - Dot notation is optional:

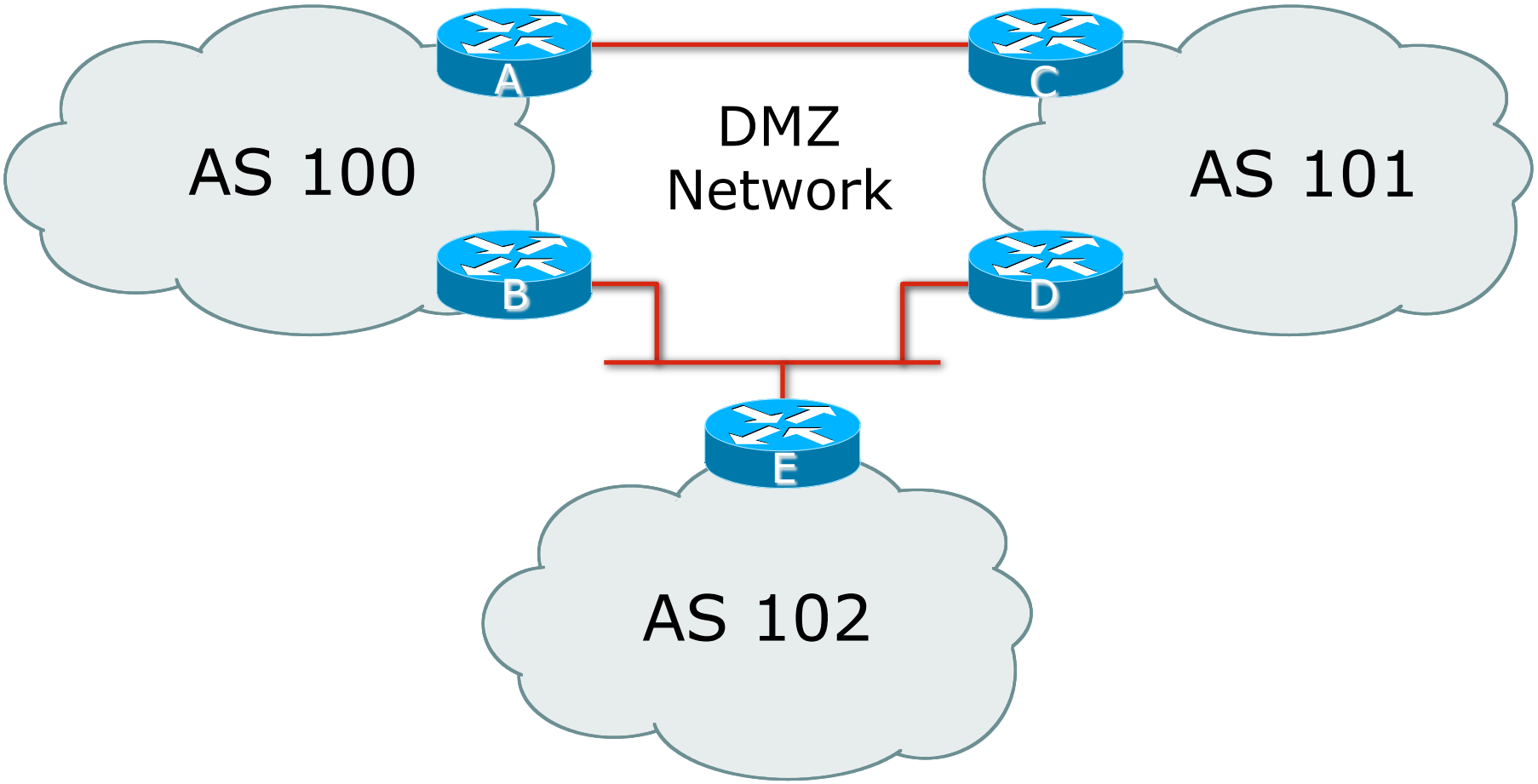
```
router bgp 2.4
 bgp asnotation dot
```

BGP Basics



- Runs over TCP – port 179
- Path vector protocol
- Incremental updates
- "Internal" & "External" BGP

Demarcation Zone (DMZ)



- DMZ is the link or network shared between ASes

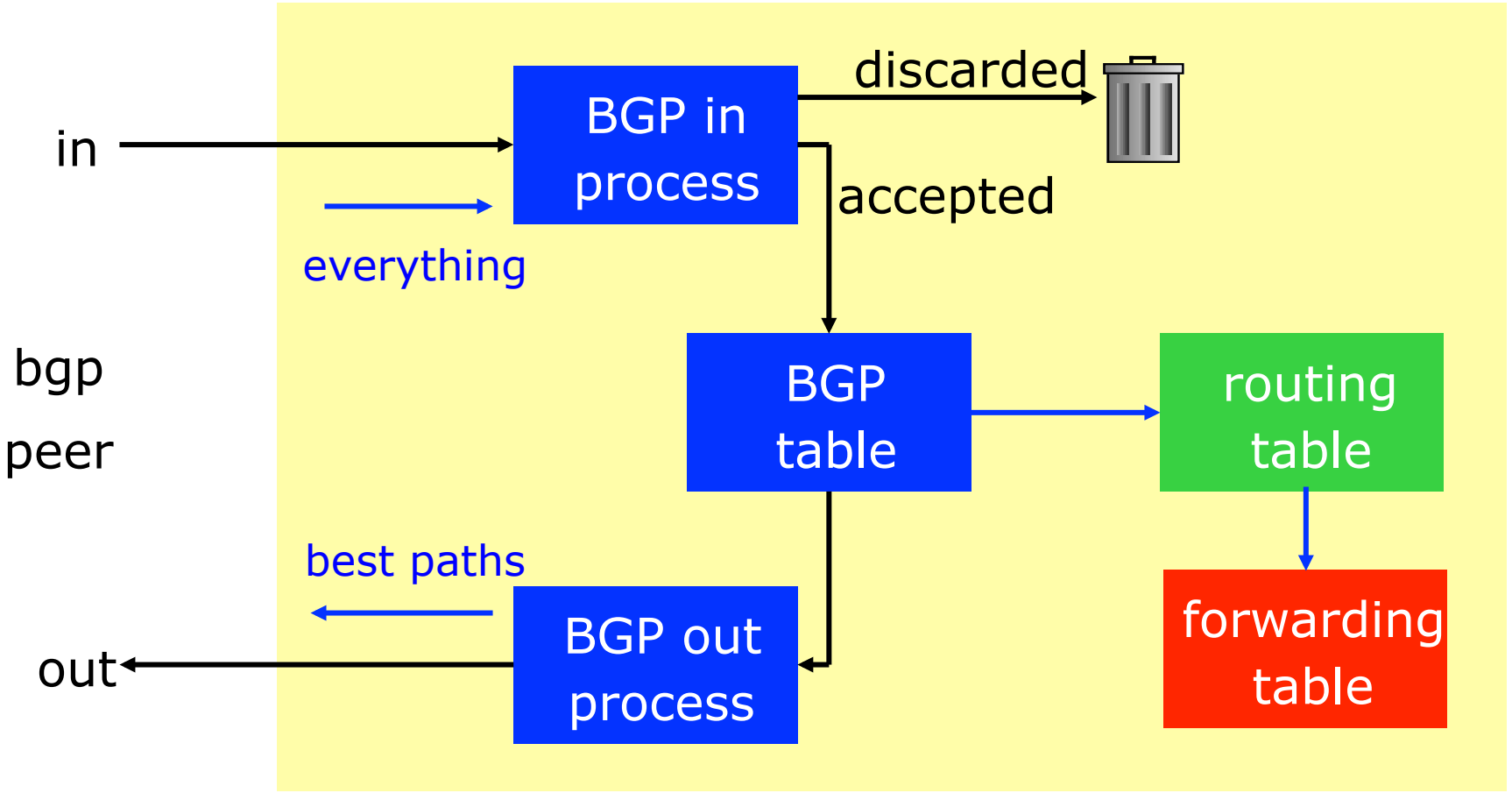
BGP General Operation

- Learns multiple paths via internal and external BGP speakers
- Picks the best path and installs it in the routing table (RIB)
- Best path is sent to external BGP neighbours
- Policies are applied by influencing the best path selection

Constructing the Forwarding Table

- BGP “in” process
  - receives path information from peers
  - results of BGP path selection placed in the BGP table
  - “best path” flagged
- BGP “out” process
  - announces “best path” information to peers
- Best path stored in Routing Table (RIB)
- Best paths in the RIB are installed in forwarding table (FIB) if:
  - prefix and prefix length are unique
  - lowest “protocol distance”

Constructing the Forwarding Table

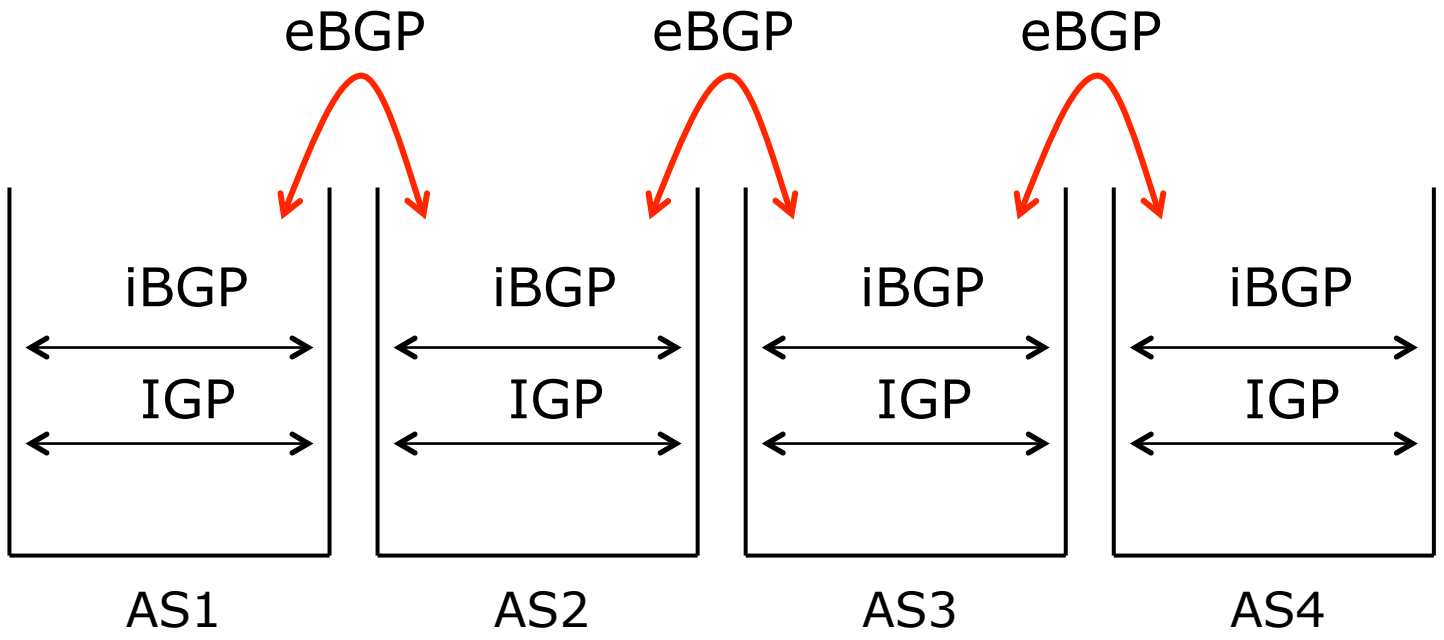


eBGP & iBGP

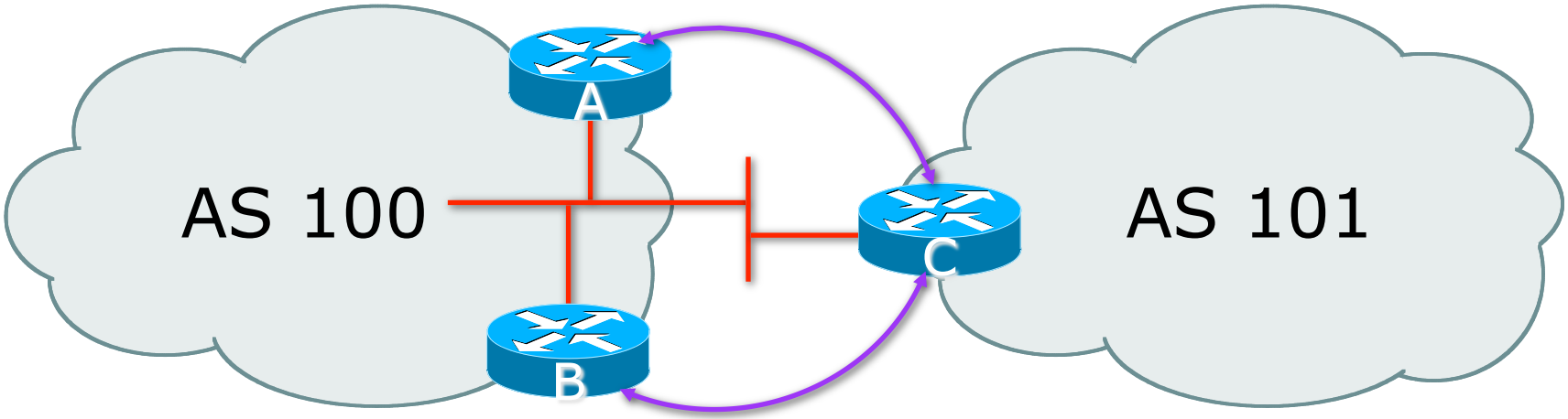
- BGP used internally (iBGP) and externally (eBGP)
- iBGP used to carry
  - Some/all Internet prefixes across ISP backbone
  - ISP's customer prefixes
- eBGP used to
  - Exchange prefixes with other ASes
  - Implement routing policy

BGP/IGP model used in ISP networks

- Model representation



External BGP Peering (eBGP)

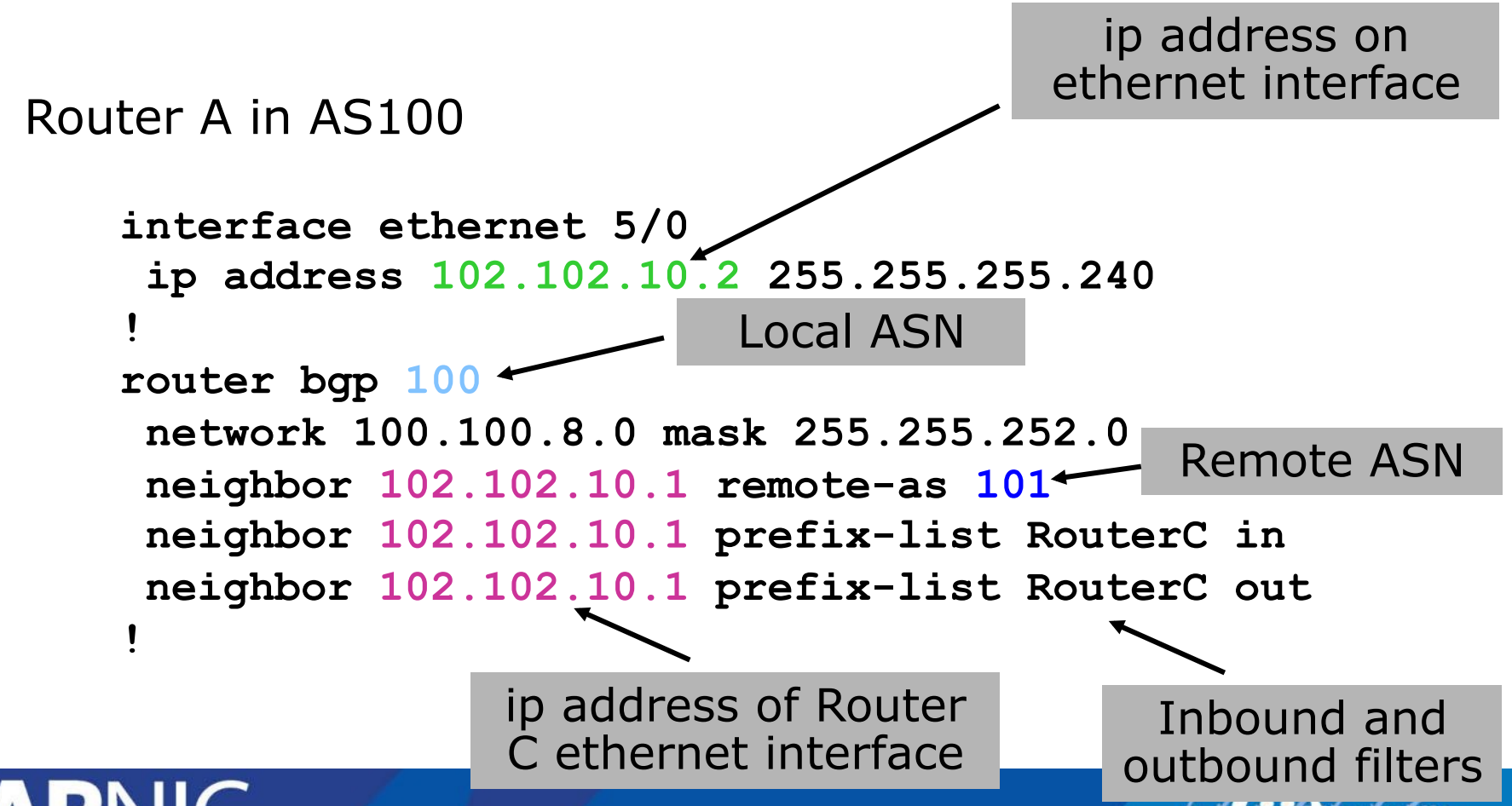


- Between BGP speakers in different AS
- Should be directly connected
- **Never** run an IGP between eBGP peers

Configuring External BGP

Router A in AS100

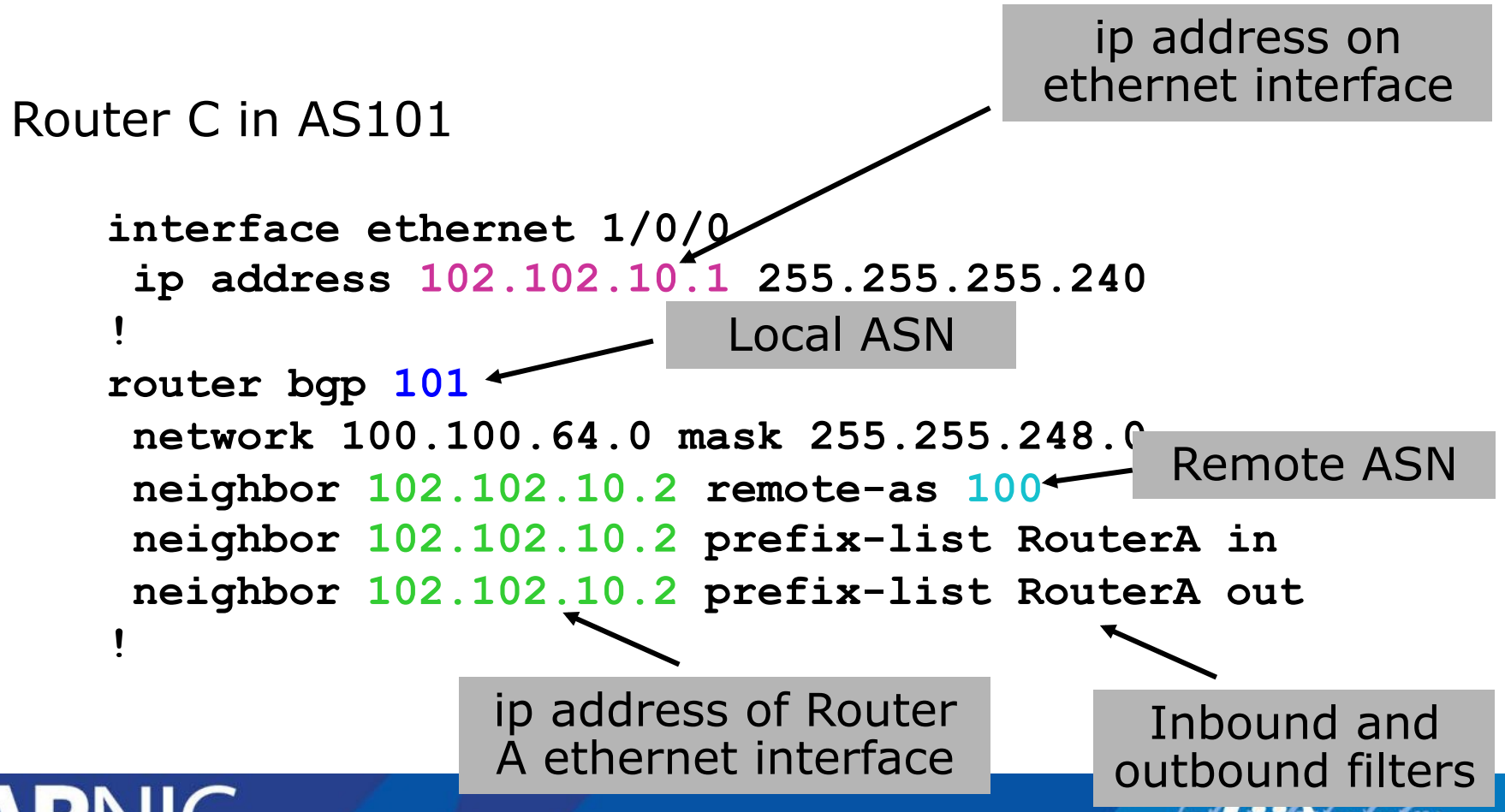
```
interface ethernet 5/0
 ip address 102.102.10.2 255.255.255.240
!
router bgp 100
 network 100.100.8.0 mask 255.255.252.0
 neighbor 102.102.10.1 remote-as 101
 neighbor 102.102.10.1 prefix-list RouterC in
 neighbor 102.102.10.1 prefix-list RouterC out
!
```



Configuring External BGP

Router C in AS101

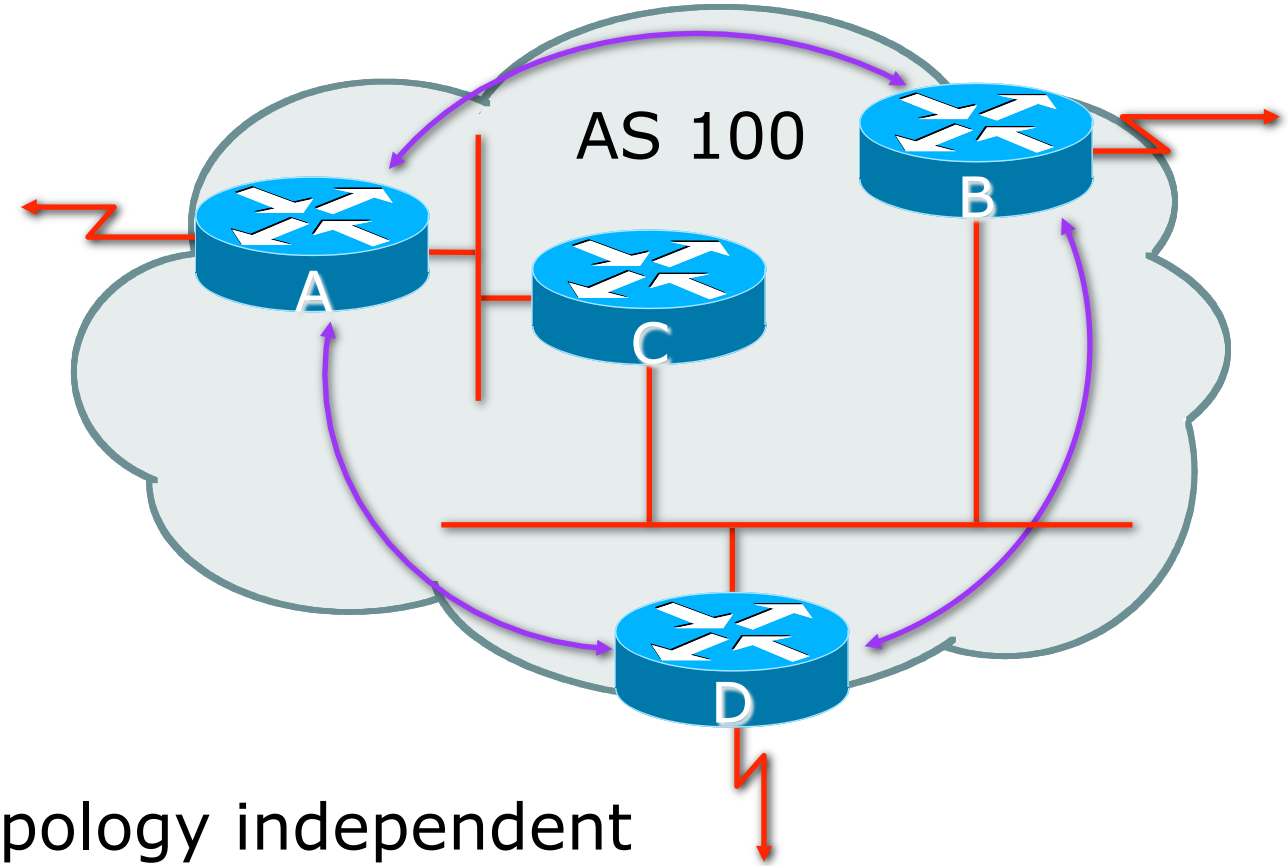
```
interface ethernet 1/0/0
 ip address 102.102.10.1 255.255.255.240
!
router bgp 101
 network 100.100.64.0 mask 255.255.248.0
 neighbor 102.102.10.2 remote-as 100
 neighbor 102.102.10.2 prefix-list RouterA in
 neighbor 102.102.10.2 prefix-list RouterA out
!
```



Internal BGP (iBGP)

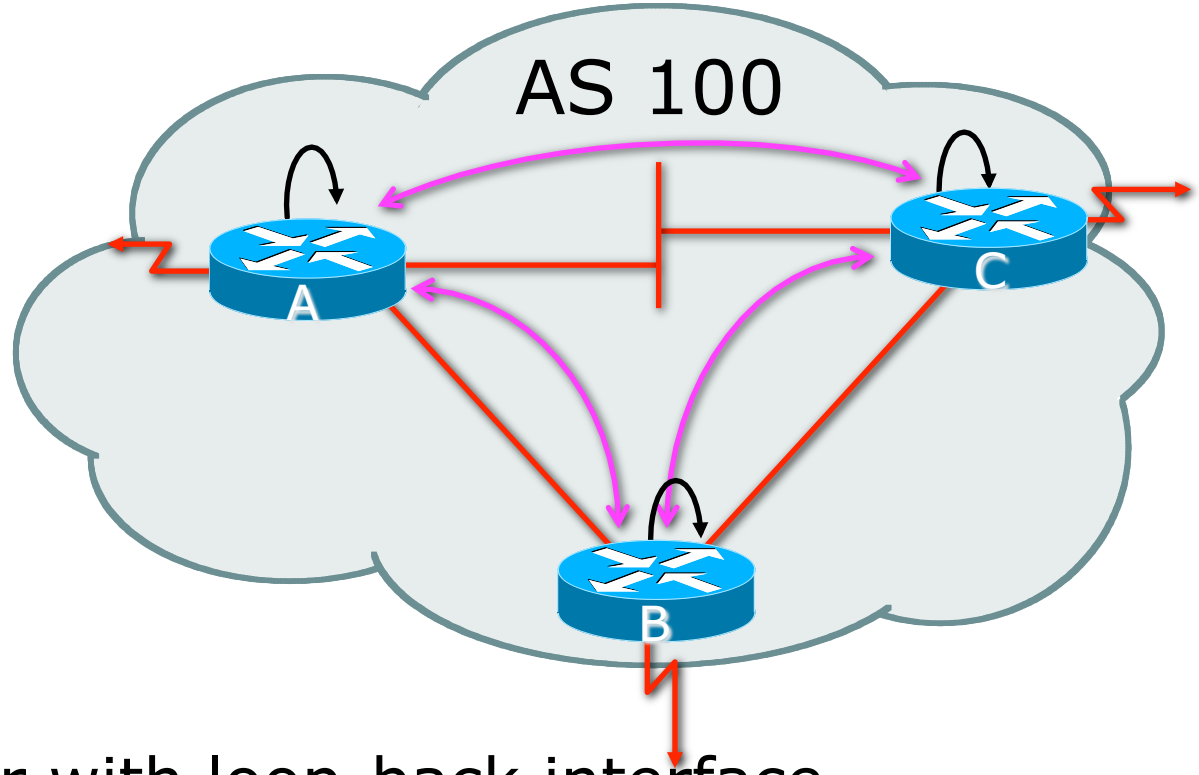
- BGP peer within the same AS
- Not required to be directly connected
  - IGP takes care of inter-BGP speaker connectivity
- iBGP speakers must be fully meshed:
  - They originate connected networks
  - They pass on prefixes learned from outside the ASN
  - They do not pass on prefixes learned from other iBGP speakers

Internal BGP Peering (iBGP)



- Topology independent
- Each iBGP speaker must peer with every other iBGP speaker in the AS

Peering between Loopback Interfaces

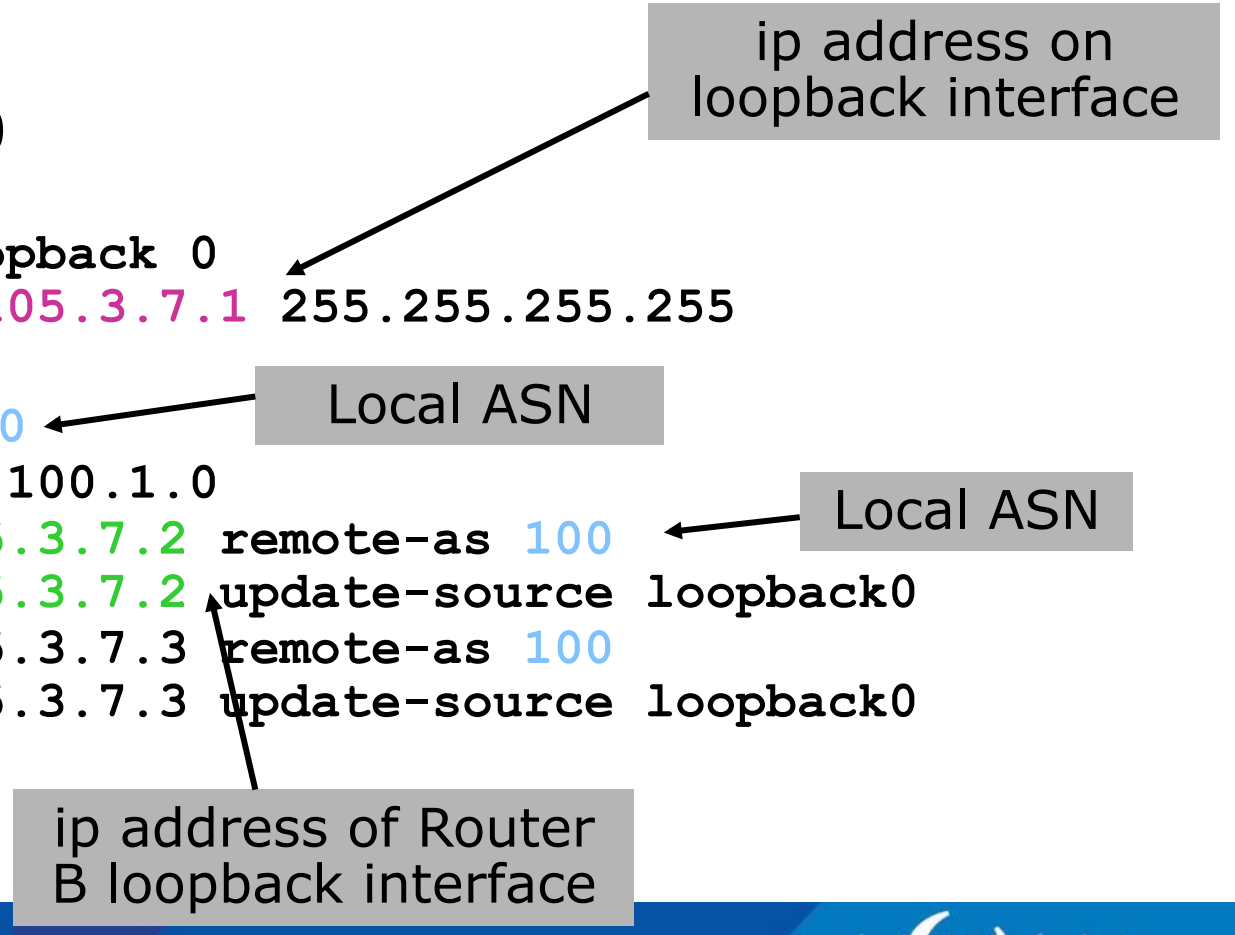


- Peer with loop-back interface
  - Loop-back interface does not go down – ever!
- Do not want iBGP session to depend on state of a single interface or the physical topology

Configuring Internal BGP

Router A in AS100

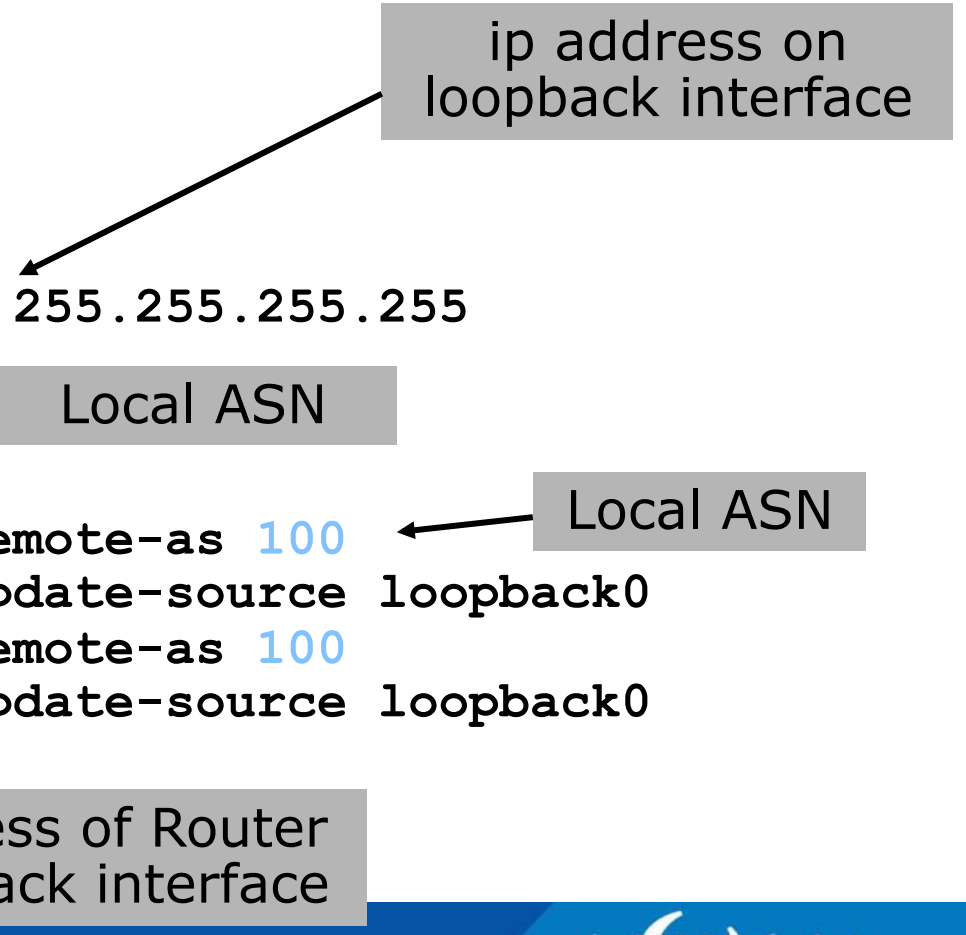
```
interface loopback 0
 ip address 105.3.7.1 255.255.255.255
!
router bgp 100
 network 100.100.1.0
 neighbor 105.3.7.2 remote-as 100
 neighbor 105.3.7.2 update-source loopback0
 neighbor 105.3.7.3 remote-as 100
 neighbor 105.3.7.3 update-source loopback0
!
```



Configuring Internal BGP

Router B in AS100

```
interface loopback 0
 ip address 105.3.7.2 255.255.255.255
!
router bgp 100
 network 100.100.1.0
 neighbor 105.3.7.1 remote-as 100
 neighbor 105.3.7.1 update-source loopback0
 neighbor 105.3.7.3 remote-as 100
 neighbor 105.3.7.3 update-source loopback0
!
```



Inserting prefixes into BGP

- Two ways to insert prefixes into BGP
  - **`redistribute static`**
  - **`network`** command

Inserting prefixes into BGP – redistribute static

- Configuration Example:

```
router bgp 100
 redistribute static
ip route 102.10.32.0 255.255.254.0 serial0
```

- Static route must exist before redistribute command will work
- Forces origin to be “incomplete”
- Care required!

Inserting prefixes into BGP – redistribute static

- Care required with redistribute!
  - `redistribute <routing-protocol>` means everything in the <*routing-protocol*> will be transferred into the current routing protocol
  - Will not scale if uncontrolled
  - Best avoided if at all possible
  - **`redistribute`** normally used with “route-maps” and under tight administrative control

Inserting prefixes into BGP – network command

- Configuration Example

```
router bgp 100
 network 102.10.32.0 mask 255.255.254.0
ip route 102.10.32.0 255.255.254.0 serial0
```

- A matching route must exist in the routing table before the network is announced
- Forces origin to be “IGP”

Configuring Aggregation

- Three ways to configure route aggregation
  - **`redistribute static`**
  - **`aggregate-address`**
  - **`network`** command

Configuring Aggregation

- Configuration Example:

```
router bgp 100
 redistribute static
ip route 102.10.0.0 255.255.0.0 null0 250
```

- Static route to “null0” is called a pull up route
  - Packets only sent here if there is no more specific match in the routing table
  - Distance of 250 ensures this is last resort static
  - Care required – see previously!

Configuring Aggregation – Network Command

- Configuration Example

```
router bgp 100
 network 102.10.0.0 mask 255.255.0.0
ip route 102.10.0.0 255.255.0.0 null0 250
```

- A matching route must exist in the routing table before the network is announced
- Easiest and best way of generating an aggregate

Configuring Aggregation – aggregate-address command

- Configuration Example:

```
router bgp 100
 network 102.10.32.0 mask 255.255.252.0
 aggregate-address 102.10.0.0 255.255.0.0 [summary-only]
```

- Requires more specific prefix in BGP table before aggregate is announced
- summary-only keyword
  - Optional keyword which ensures that only the summary is announced if a more specific prefix exists in the routing table

Summary
BGP neighbour status

```
Router6>sh ip bgp sum

BGP router identifier 10.0.15.246, local AS number 10

BGP table version is 16, main routing table version 16

7 network entries using 819 bytes of memory

14 path entries using 728 bytes of memory

2/1 BGP path/bestpath attribute entries using 248 bytes of memory

0 BGP route-map cache entries using 0 bytes of memory

0 BGP filter-list cache entries using 0 bytes of memory

BGP using 1795 total bytes of memory

BGP activity 7/0 prefixes, 14/0 paths, scan interval 60 secs


Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd

10.0.15.241     4    10       9       8       16    0    0 00:04:47        2

10.0.15.242     4    10       6       5       16    0    0 00:01:43        2

10.0.15.243     4    10       9       8       16    0    0 00:04:49        2

...
```

BGP Version

Updates sent and received

Updates waiting

Summary
BGP Table

```
Router6>sh ip bgp
BGP table version is 16, local router ID is 10.0.15.246
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
              x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

     Network          Next Hop            Metric LocPrf Weight Path
 *>i 10.0.0.0/26      10.0.15.241              0    100      0 i
 *>i 10.0.0.64/26     10.0.15.242              0    100      0 i
 *>i 10.0.0.128/26    10.0.15.243              0    100      0 i
 *>i 10.0.0.192/26    10.0.15.244              0    100      0 i
 *>i 10.0.1.0/26      10.0.15.245              0    100      0 i
 *>  10.0.1.64/26     0.0.0.0                  0         32768 i
 *>i 10.0.1.128/26    10.0.15.247              0    100      0 i
 *>i 10.0.1.192/26    10.0.15.248              0    100      0 i
 *>i 10.0.2.0/26      10.0.15.249              0    100      0 i
 *>i 10.0.2.64/26     10.0.15.250              0    100      0 i
...
```

Summary

- BGP4 – path vector protocol
- iBGP versus eBGP
- stable iBGP – peer with loopbacks
- announcing prefixes & aggregates

Thank you!

End of Session

APNIC